

Глава 3

КОНСТРУКТИВНО-СИНТАКСИЧЕСКИЙ АНАЛИЗ ПРЕДЛОЖЕНИЯ

СЕМАНТИКО-СИНТАКСИЧЕСКИЙ СЛОВАРЬ

Понятие семантико-синтаксического варианта

Конструктивно-синтаксический анализ¹ предложения начинается с обращения к семантико-синтаксическому словарю с целью извлечения соответствующей информации в отношении входящих в предложение буквенных токенов. В итоге к значениям графических и морфологических признаков, приписанных им на предшествующих этапах, добавляется соответствующая семантико-синтаксическая информация.

В описываемом словаре лексические единицы представлены в виде набора семантико-синтаксических вариантов (ССВ). ССВ — это, по сути, аналог значения (или оттенка значения) в толковом словаре, с той разницей, что существенное внимание при выделении этих вариантов уделяется их синтаксической реализации. Вследствие этого набор семантико-синтаксических вариантов слова, как правило, отличается от совокупности его значений и оттенков значений, выделяемых в традиционных словарях.

С одной стороны, то, что в толковом словаре выступает в качестве единого значения, в нашем описании может быть разбито на несколько ССВ. Чаще всего это обусловлено тем, что в традиционных толковых словарях характеристика подчинительных связей слова дается общим списком. К примеру, глагол *просить* в значении ‘до-

¹ Вслед за И. П. Распоповым [1970: 31], М. В. Никитиным [2007: 503] и Я. Г. Тестельцом [2001: 438] мы считаем целесообразным рассматривать синтаксис как многослойное образование и выделяем «конструктивный синтаксис» и «коммуникативный синтаксис». При этом, в отличие от М. В. Никитина [2007: 504], мы не ставим знак равенства между конструктивным и формальным синтаксисом: конструктивно-синтаксический анализ в описываемом лингвистическом процессе задействует информацию о значении лексических единиц из специально разработанного семантико-синтаксического словаря, которому и посвящена настоящая глава.

бываться чего-л. у кого-л.; склонять к исполнению желаемого' имеет следующие пометы: «(кого-что), кого-чего, о чем, с инф., с придат. предл.». Понятно, что такое перечисление едва ли пригодно для автоматической обработки языка. Требуется преобразовать этот список в структурированное описание валентностей глагола и способов их реализации. Глагол *просить* имеет три валентности: субъекта, контрагента и объекта. Субъект и контрагент имеют стандартное выражение (соответственно подлежащим и прямым дополнением), в то время как валентность объекта (содержание просьбы) может быть реализована существительным в родительном падеже (*хлеба, слова*), предложно-падежной формой «о S_{предл.}» (*о помощи*), инфинитивом (*помочь*) либо изъяснительным придаточным, вводимым союзом *чтобы* (*чтобы помогли*). В итоге одно словарное значение в нашем словаре представлено в виде четырех ССВ.

С другой стороны, количество ССВ, выделяемых у той или иной леммы, может быть меньше, чем число словарных значений. Во-первых, есть случаи, когда правомерность выделения того или иного значения в толковом словаре вызывает сомнения. Например, у прилагательного *светлый* Большой толковый словарь русского языка (БТС 2009), вслед за Словарем русского языка в четырех томах (СРЯ 1981–1984), выделяет значение 'блестящий, излучающий блеск', которое иллюстрируется, в частности, речением *светлый кинжал* (в четырехтомном словаре — цитатой из стихотворения М. Ю. Лермонтова «Кинжал», которая была преобразована в речение). С нашей точки зрения, отдельного значения здесь нет (ср. *светлое лезвие*, *светлый нож*, *светлые украшения*), есть лишь авторское окказиональное употребление.

Во-вторых, семантико-синтаксический словарь более ориентирован на описание денотата слова, чем его сигнификата. Такой выбор неизбежен в контексте автоматической обработки языка. Напротив, традиционная лексикография при описании семантики слова руководствуется общностью сигнификативного компонента отдельных значений. Различие между этими подходами отчетливо видно на примере многозначных прилагательных типа *тяжелый, мягкий, темный, высокий* и др. Вот как словарь² описывает семантику слова *высокий* (конец словарной статьи, содержащий фразеологические обороты и производные слова, опущен):

² Здесь и далее источником лексикографической информации служит (БТС 2009).

1. Большой по протяженности снизу вверх; находящийся далековерху; превосходящий обычную, среднюю высоту (противоп.: низкий). *В-ая гора. В-ие облака. В-ая трава. В. каблук. В. потолок. В. прыжок. В. лоб.*
2. Значительный по количеству, интенсивности, степени проявления и т. п.; превышающий средний уровень, среднюю норму. *В-ая активность. В-ая производительность труда. В. урожай. В-ие темпы. В-ие цены. В-ое напряжение, давление. В-ая концентрация вещества. В-ая чувствительность. В. уровень жизни. В-ие требования. В-ое мастерство.*
3. Очень хороший; отличный. *В-ие оценки. В-ие достоинства. Показать в. класс. Товары высокого качества. Быть высокого мнения о себе. Высокой пробы кто-, что-л.*
4. Выдающийся по значению; почетный, важный. *В-ая награда. В. пост. В-ая честь. В-ое звание. В-ое общественное положение. Оправдать в-ое доверие. Приветствовать высокого гостя. Вращаться в высоких сферах. В-ие договаривающиеся стороны.*
5. Возвышенный, благородный, лишенный корыстных помыслов. *В. порыв. В-ое чувство. В-ое искусство. В. идеал. В-ое предназначение. В-ая цель. Говорить о высоких материях (обычно ирон.).*
6. Приподнятый, торжественный (о языке, стиле). *В. слог.*
7. Вызываемый колебаниями большой частоты; тонкий и звонкий (о звуке, голосе). *В-ая нота. В. дискант.*

Из этих значений лишь шестое и седьмое денотативно ориентированы и без труда выделяются в семантико-синтаксическом словаре в качестве отдельных вариантов. Описание остальных значений с точки зрения семантических требований к управляющему существительному затруднено. Как видно из примеров, в первом значении (и соответствующем оттенке) прилагательное *высокий* может сочетаться и с конкретными, и с абстрактными существительными. Что касается второго, третьего и пятого значений, все предложения демонстрируют сочетаемость с абстрактными существительными, причем без возможности дальнейшей специализации, а если специализация невозможна, то нет и самостоятельных ССВ. Наконец, четвертое значение, как и первое, не предполагает никаких, даже самых грубых (в терминах «конкретное — абстрактное»), ограничений на тип существительного.

Подводя итог, мы констатируем, что число ССВ, выделяемых у того или иного слова, как правило, не совпадает с количеством его словарных значений.

**От морфологической атрибуции словоформы
к ее семантико-синтаксическому варианту**

Как известно, «в последние десятилетия проблема многозначности наиболее остро стоит в компьютерной лингвистике» [Июдин 2014: 89]. Основное назначение семантико-синтаксического словаря состоит в том, чтобы способствовать снятию лексико-грамматической неоднозначности: графических и морфологических характеристик, приписанных токенам на предыдущих этапах анализа, для этого, как правило, недостаточно. Благодаря специально разработанному формату описания, словарь представляет в структурированном виде информацию об особенностях реализации слова (точнее, каждого из его ССВ). Сравнение словарной информации с текстовыми данными позволяет снизить (а зачастую и снять) грамматическую омонимию словоформы в тексте, а также сократить ее лексическую многозначность, уменьшив количество допустимых интерпретаций.

Как организована информация в данном словаре? Выше было сказано, что словарная статья состоит из списка ССВ. При каждом из этих вариантов имеется фильтр, содержащий графические и морфологические требования к его реализации (фильтр может включать несколько ограничений). Например: «с прописной буквы», «только 3 л.», «только полн.». Если никаких графических и морфологических ограничений для данного варианта нет, фильтр пуст. Задача фильтра заключается в том, чтобы в ходе синтаксического анализа отсеивать те ССВ, которые не соответствуют информации, приписанной соответствующему токеноу на предшествующих этапах анализа.

Более подробно эта процедура может быть представлена следующим образом. Как правило, словоформа имеет несколько вариантов морфологической атрибуции (а иногда и несколько лемм). В семантико-синтаксическом словаре по лемме ищется соответствующая словарная статья, представляющая собой список ССВ. Если лемма не одна, берутся все ССВ для всех лемм. Далее нужно «примерить» каждый «морфологический вариант» на каждый «семантический вариант». Происходит последовательное сопоставление морфологической и графической информации, приписанной каждому из результатов атрибуции, с фильтром каждого из ССВ. Как правило, в ряде случаев такое сравнение выявляет противоречия. Следовательно, итоговое число вариантов с приписанной им графической, морфологической и семантико-синтаксической информацией обычно оказывается меньше того, что может быть получено сугубо комбинатор-

но. В целом, однако, вариативность на этом этапе резко возрастает; ее снижение происходит позднее — при выявлении синтаксических связей между токенами и их объединении в словосочетания.

Поясним сказанное на примере. Допустим, в тексте встретилась словоформа *леса*. Результат атрибуции данного токена выглядит следующим образом:

лес	1	сущ; неод; мр	род; ед
лес	2	сущ; неод; мр	им; мн
лес	2	сущ; неод; мр	вин; мн
леса	1	сущ; неод; жр	им; ед
леса	2	сущ; неод; жр	им; ед ³
леса	2	сущ; мн ⁴ ; неод	им; мн
леса	2	сущ; мн; неод	вин; мн

Итак, у нас есть семь вариантов морфологической атрибуции. Обратимся теперь к словарю. Совокупно по леммам *лес* и *леса* мы находим пять ССВ, смысл которых для простоты изложения мы сформулируем при помощи обычных словарных дефиниций, ср.:

1. Множество дикорастущих деревьев, расположенных на большом пространстве; пространство, обильно поросшее деревьями.
2. Большое количество, множество (о возвышающихся предметах).
3. (собр.) Срубленные деревья как строительный и промышленный материал; бревна.
4. То же, что *леска*.
5. (мн) Временное сооружение, устанавливаемое у стен здания для строительных и ремонтных работ.

Как видно из помет (которые, напомним, в семантико-синтаксическом словаре входят в состав фильтра), третий вариант может реализоваться только при условии, что слово *лес* стоит в единственном числе, а пятый, — напротив, ограничен множественным числом, так как слово *леса* в данном значении принадлежит к существительным *pluralia tantum*.

³ В значении 'леска' данная словоформа имеет вариативное ударение.

⁴ Включение «мн» в перечень неизменяемых морфологических характеристик (с дублированием этой пометы в значениях словоизменительных категорий) — способ обозначения существительных *pluralia tantum*, «унаследованный» от проекта АОТ.

Осталось наложить семантико-синтаксические варианты на результаты морфоанализа, принимая во внимание фильтры. В итоге остаются девять вариантов: пять — для леммы *лес* (три единственного числа и два множественного числа), два — для единственного числа существительного женского рода *леса* (с вариантами ударения) и два — для существительного *pluralia tantum леса*. Как и ожидалось, в результате использования семантико-синтаксического словаря число вариантов возросло, однако увеличение оказалось не столь значительным (с семи до девяти), что объясняется наличием грамматических ограничений у двух из пяти значений.

Условия реализации

Продолжим рассмотрение того, как устроено описание лексики в семантико-синтаксическом словаре. Прежде всего следует подчеркнуть, что в нем нет привычных словарных дефиниций наподобие тех, что были приведены в примере выше для удобства изложения. Вместо этого используется специально разработанный формат описания, состоящий из двух частей — «условий реализации» и «собственного значения». В совокупности эти два компонента позволяют осуществлять такое семантическое описание, которое выявляет не только глубинные аспекты значения лексической единицы, но и совокупность семантических и грамматических требований к сочетающимся с ней словам.

Условия реализации слова включают указание на его частеречную принадлежность и комплексное описание семантических и грамматических требований к другим словам, способным сочетаться с ним в синтаксических конструкциях (прежде всего — в рамках сильного управления). Комплексное описание может включать вплоть до восьми условий, каждое из которых представляет собой группу признаков со своими значениями. Для разных частей речи условия реализации имеют различное наполнение. Наиболее сложными со структурной точки зрения являются условия реализации у глаголов — для них предусмотрена возможность заполнения всех восьми групп признаков (для потенциальных восьми актантов⁵).

⁵ Указанное число является произвольным и взято «с запасом». С одной стороны, наибольшее число валентностей у глаголов в русском языке, отмеченное у глагола *командировать*, равняется шести [Апресян 1995а: 137], с другой стороны, в ряде случаев граница между обязательными и факультативными валент-

Если у слова выделяется несколько ССВ, то каждый из них описывается через присущие ему условия реализации и собственное значение. Обоснованность выделения того или иного ССВ обеспечивается прежде всего спецификой его условий реализации.

Условия реализации, как и фильтр (см. выше), направлены на то, чтобы сократить лексико-грамматическую неоднозначность. Но, в отличие от фильтра, они ориентированы не «назад» (к уже имеющейся графической и морфологической информации), а «вперед» — к тому, что выявляется в ходе синтаксического анализа. К примеру, наречие в русском языке может быть синтаксически связано с глаголом, прилагательным, наречием или существительным, быть способным или неспособным выступать в функции предикатива, иметь или не иметь предпочтительную (или даже закрепленную) позицию по отношению к главному слову, причем эта позиция может располагаться как перед главным словом (*очень красивый, глубоко порядочный*), так и после него (*форель по-царски, дорога домой*). Сопоставление записанных в словаре условий реализации конкретного наречия с текстовыми данными позволяет отбросить неподходящие варианты и в итоге сократить число возможных интерпретаций.

Рассмотрим по очереди структуру условий реализации для слов, принадлежащих к разным частям речи и лексико-грамматическим разрядам. Как уже говорилось, для глагольных ССВ предусмотрена возможность заполнения восьми групп признаков — соответственно для восьми потенциальных актантов в тексте⁶. Каждый актант характеризуется следующими признаками: «тип синтаксического заполнения» (с такими, например, значениями, как ‘именная группа в родительном падеже без предлога’; ‘инфинитив’; ‘придаточное предложение, вводимое союзом *чтобы*’), «тип семантического заполнения» (с семантическими признаками: ‘человек’, ‘транспортное средство’, ‘абстрактная сущность’, ‘контейнер’ и пр.), а также «семантическая валентность» (субъект, инструмент, начальная

ностями неоднозначна (см., например, [Апресян 1995а: 119–163; Семиотика и информатика 1998; Тестелец 2001: 156–228]).

⁶ Если та или иная семантическая валентность глагола является нереализуемой (например, инкорпорированной), она не указывается. Это решение имеет свои недостатки, связанные, в частности, с возможностью нереализуемых валентностей «выходить на поверхность» на более обширных отрезках текста. В дальнейшем планируется усложнить формат описания, сделав его более «семантически ориентированным», что позволит последовательнее отражать диатезы глагола и различные актантные деривации (см. [Тестелец 2001: 411–436]).

точка / конечная точка и т. д.). Заметим, что в семантико-синтаксическом словаре понятие видовой пары не используется, и каждый глагол описывается отдельно в совокупности всех его ССВ. Это решение обусловлено разнородностью видовых пар (разным семантическим соотношением между соответствующими глаголами несовершенного и совершенного вида), а также тем, что их члены нередко обладают разными системами лексических значений⁷. Соответственно, причастия и деепричастия «привязываются» к конкретному глаголу.

Все компоненты этого описания значимы для дифференциации ССВ глагола: при автоматическом анализе предложения они способствуют его правильной интерпретации. Хорошо известно, что тождество синтаксической конструкции не означает тождества реализуемой валентности. К примеру, глаголы *гордиться* и *забивать* (в знач.: ‘ударами вбить, вогнать во что-л.’) управляют существительным в творительном падеже без предлога, при этом реализуя разные валентности, соответственно, их описание будет различаться, ср.:

гордиться (условия реализации) глагол; первый актанта: <...>; второй актанта: падеж = творительный / предлог = без предлога / сем. признаки: неспецифицир. / валентность = объект

забивать (условия реализации) глагол; первый актанта: <...>; второй актанта: <...>; третий актанта: падеж = творительный / предлог = без предлога / сем. признаки: базовый = предмет; агрегат. сост. = твердый / валентность = инструмент

Более того, одна и та же предложно-падежная форма при наличии единственного глагола в предложении также не гарантирует тождества реализуемой валентности, ср. *пить чай с вареньем / с женой / с удовольствием*. Впрочем, в данном случае мы имеем дело с так называемым слабым управлением, которое не входит в актантажную

⁷ Например, у глагола *встречаться* есть два значения (‘время от времени иметь свидания, быть в близких отношениях’ и ‘бывать, попадаться’), отсутствующие у глагола *встретиться*; у глагола *обвинять* есть значение ‘поддерживать обвинение, выступать в качестве судебного обвинителя, прокурора’, которого нет у соотносительного по виду глагола *обвинить*, и т. д. Помимо этого, видовые соотношения у многозначных глаголов нередко формируются на уровне их отдельных значений, ср. *мочить* — *намочить*, *замочить* (‘делать мокрым, влажным’) и *вымочить* (‘пропитывать влагой для придания каких-л. свойств’). Такое положение вещей диктует необходимость создания индивидуального описания для каждого глагола.

рамку глагола и не влияет на его значение: во всех трех предложениях реализуется один и тот же ССВ глагола *пить*.

Однако нередко бывает так, что разные ССВ глагола характеризуются одинаковым набором валентностей, каждая из которых выражается регулярным образом (к примеру, субъект — именной группой в именительном падеже, объект — прямым дополнением и т. д.). Иными словами, мы имеем и тождественные наборы семантических валентностей, и тождественные синтаксические реализации этих валентностей. Тогда единственный способ их разграничить — это учитывать семантические признаки и их значения⁸. К примеру, глагол *жарить* в трех (из четырех) значений имеет валентности субъекта и объекта, ср.:

- 1) что. Приготовлять (мясо, рыбу, картофель и т. п.), подвергая действию жара без воды. *Ж. котлеты. Ж. картошку на масле;*
- 2) кого-что. Разг. Обжигать лучами, палить (о солнце). *Солнце жарит нещадно.* □ безл. *Ну и жарит сегодня! Спину так и жарит!*
- 3) что. Разг. Жарко топить (печку, баню и т. п.). *Зачем целый день ж. печку? Уже и так жарко.*

Несложно заметить, что семантические валентности у глаголов в данных ССВ заполняются существительными с разными значениями семантических признаков. Убрав (для краткости) из условий реализации сведения о валентностях и их синтаксической реализации, получим:

- 1) (условия реализации) глагол; первый актанта: сем. признаки: базовый: живое существо; человек = +⁹; второй актанта: сем. признаки: базовый: предмет; продукты питания = +;
- 2) (условия реализации) глагол; первый актанта: сем. признаки: базовый: предмет; второй актанта: сем. признаки: базовый: живое существо / часть тела;

⁸ Такая ситуация характерна, в частности, для глаголов физического действия. Так, в [Скребцова 1996] было показано, что изменение семантического типа субъекта и/или объекта (по отношению к ситуации, описываемой основным значением соответствующего глагола) является важным индикатором его употребления в производном значении.

⁹ Здесь и далее некоторые семантические признаки обозначаются фактически через их значения, а в качестве значения стоит «+».

- 3) (условия реализации) глагол; первый актант: сем. признаки: базовый: живое существо; человек = +; второй актант: сем. признаки: базовый: предмет.

Прилагательные также снабжаются семантической характеристикой существительных, способных с ними сочетаться, что позволяет в ходе синтаксического анализа сокращать лексическую неоднозначность. Так, прилагательное *жесткий* в своем прямом значении согласуется с предметными существительными (*кровать, корка* и т.д.), а в производных — с абстрактными (*характер, тон; срок, требования*). Дальнейшее разграничение производных значений ('суровый, резкий' vs. 'не допускающий отклонений') требует более тонкой «настройки» ожидаемых значений признаков.

Актантной структурой обладают также отглагольные имена со значением действия (*переезд, раздача, отправление*) и некоторые существительные, образованные от прилагательных (*возможность, доступность, синева*). Помимо таких очевидных номинализаций, есть целый ряд других существительных, которые, как правило, имеют при себе зависимые слова, ср. *цена, доход, разница, довод, отец, вера, жажда, ревность*. При выделении подобных существительных мы опираемся не столько на отношения производности, сколько на данные толковых словарей (сведения о подчинительных связях слова, дефиницию, речения и цитаты), а также на материалы НКРЯ.

У многих предметных существительных (точнее, их ССВ) условия реализации являются пустыми, т.е. не содержат ни одной характеристики. Эта ситуация характерна для основных значений таких существительных, как *стол, дом, синица, сосна, чашка, автомобиль* и пр., где отчетливо преобладает денотативный компонент значения, а производные реализуются в условиях синтаксических ограничений. Так, у существительных *стол* и *дом* есть производные значения, связанные с названием различного рода учреждений, ср. *стол заказов (находок), Дом творчества (культуры, архитектора, офицеров* и др.). В толковых словарях эти значения сопровождаются грамматической пометой «чего» или «кого/чего». В нашем семантико-синтаксическом словаре эта информация отражена в условии реализации, где указано, что данный ССВ требует зависимую именную группу в форме родительного падежа без предлога с определенными семантическими признаками и их значениями. (Справедливости ради следует сказать, что слово *дом* в указанном значении (и только в нем) пишется с заглавной буквы, поэтому выбор нуж-

ного ССВ может произойти за счет информации, содержащейся в фильтре, — правда, при условии, что данное слово не занимает инициальную позицию в предложении или прямой речи.)

Для предлогов в условиях реализации заполняются падеж управляемого существительного и значения его семантических признаков. Благодаря такому описанию наглядно разграничиваются значения предлога *в*, реализующиеся в следующих предложно-падежных формах:

в <июне>: (условия реализации) предлог; падеж = предложный / сем. признаки: базовый: время

в <лесу>: (условия реализации) предлог; падеж = предложный / сем. признаки: базовый: предмет; пространств. = +

в <лес>: (условия реализации) предлог; падеж = винительный / сем. признаки: базовый: предмет; пространств. = +

Составные предлоги (которые ранее — в ходе графематического анализа — были найдены в словаре устойчивых словосочетаний и соответственно помечены) также могут иметь более одного значения, ср.:

в течение <июня>: (условия реализации) предлог; падеж = родительный / сем. признаки: базовый: время

в течение <лекции>: (условия реализации) предлог; падеж = родительный / сем. признаки: базовый: действие

Для словосочетаний, которые могут функционировать и в качестве свободных, и в качестве связанных (ср. *до лампочки, по барабану*), предусмотрен учет обоих вариантов в ходе синтаксического анализа. В первом случае из семантико-синтаксического словаря будет получена информация, касающаяся отдельно предлога и существительного, во втором — информация, связанная с устойчивым сочетанием.

Для ССВ наречий указывается частеречная характеристика главного слова и значения его семантических признаков, а также фиксированность / нефиксированность позиции наречия по отношению к главному слову (и если позиция фиксирована, то какая — препозиция или постпозиция), способность / неспособность выступать в предикативной функции.

В отношении других частей речи (местоимений, числительных, союзов, частиц, вводных слов и оборотов) следует сказать, что, помимо снятия лексико-грамматической омонимии внутри предложе-

ния, они играют важную роль в обеспечении локальной и глобальной когеренции текста. Соответственно, наряду с описанием в семантико-синтаксическом словаре, они также рассматриваются на этапе коммуникативного синтаксиса, подразумевающим учет структурных, прагматических, риторических особенностей построения текста; об этом речь пойдет в следующей главе.

Собственное значение

Обратимся теперь ко второму компоненту описания семантико-синтаксических вариантов — к их «собственному значению». Оно складывается из значений соответствующих семантических признаков.

Заметим, что выше речь шла об использовании семантических признаков в условиях реализации, где они служили для характеристики слов, способных выступать в сочетании с тем или иным ССВ данного слова. Тот же самый набор семантических признаков с их значениями задействуется также и для описания самих ССВ как таковых, отдельно от их контекстуальных возможностей.

При разработке списка семантических признаков и их значений мы учитывали результаты, опубликованные в ряде источников, в том числе Русский семантический словарь (1998–2007), словари под редакцией Л. Г. Бабенко (БТСРГ 2007; БТСРС 2008), электронную онтологию проекта UNL¹⁰, список лексико-семантических классов в Национальном корпусе русского языка¹¹, отдельные идеографические словари (Hallig, Wartburg 1963; Баранов 1995) и некоторые другие. Прежде всего мы выделяем так называемый базовый семантический признак (со значениями 'живое существо', 'предмет', 'время', 'действие', 'качество', 'количество' и др.), благодаря которому осуществляется самая приближительная, грубая характеристика ССВ.

Дополнительные семантические признаки формально не подчинены базовым, поскольку в ряде случаев это удобно. К примеру, лошадь, будучи животным, одновременно может служить и транспортным средством, однако в иерархически выстроенной онтологии транспортное средство рассматривалось бы как тип предмета. Таким образом, пришлось бы выбирать между указанием на лошадь как живое существо и на лошадь как транспортное средство. Наше описание преодолевает необходимость подобного выбора, ср.:

¹⁰ См.: <http://www.undl.org/unlsys/uw/UNLONtology.html>

¹¹ См.: <http://www.ruscorpora.ru/corpora-sem.html>

лошадь: (словарное значение: 'крупное домашнее животное, используемое для перевозки людей, грузов и т. п.')

сем. признаки: базовый: животное = +; транспортное средство = +.

Однако, по сути, дополнительные семантические признаки, как правило, производны от базовых. Таковы, к примеру, признаки «агрегатное состояние» для предмета, «цвет» для качества, «физическое действие» для действия и т. д. В то же время есть и независимые признаки, ср. «наличие наблюдателя» или «релятивность», о которых речь ниже.

В некоторых случаях в качестве дополнительных семантических признаков фигурируют высоковероятные, хотя и необязательные компоненты значения. Таковым, к примеру, является признак «ювелирное украшение» в семантике слова *безделушка*, ср.:

безделушка: (собственное значение) сем. признаки: базовый: предмет; ювелирное украшение = +.

Следует подчеркнуть, что выделенные семантические признаки не имеют однозначной соотнесенности с частями речи или лексико-грамматическими разрядами. Так, значением 'действие' по базовому семантическому признаку помечаются ССВ не только глаголов, причастий и деепричастий, но и отглагольных существительных, а также образованных от причастий прилагательных (*усталый, жареный, растворимый*) и существительных (*раненый, заведующий, слабовидящий*). Это необходимо в дальнейшем для анализа свернутых пропозиций.

Приписывание значения 'действие' осуществлялось преимущественно исходя из его словарной дефиниции, точнее, из ее опорного слова. Если опорное слово дефиниции характеризуется выраженной предикатной семантикой, то рассматриваемое слово помечается семантическим признаком 'действие'. К примеру, различные значения существительного *ревность* толкуются через слова *сомнение, боязнь, усердие*, прилагательного *усталый* — через причастия *ослабевший, истомленный, потерявший, истощенный* и т. д.

Важно отметить, что способность иметь при себе сильноуправляемые члены как таковая автоматически не влечет приписывания признака 'действие'. Например, значение слова *разница* 'величина, число, являющиеся разностью между двумя числами, величинами' не имеет отношения к действию, хотя регулярно реализуется вместе с зависимыми словоформами, ср.: *Р. между старым и новым окла-*

дом. Р. между прежней и настоящей ценой товаров. Р. в десять лет. Р. в сорок рублей. Десять лет разницы между кем-л. Сорок рублей разницы между чем-л. Р. во времени между Москвой и Красноярском.

В то же время следует признать, что опора на словарные определения не всегда дает однозначный и надежный результат. В ряде случаев грань между предметным и признаковым значением очень тонкая и не всегда адекватно фиксируется лексикографами. Применительно к словам, обозначающим атмосферные осадки, этот феномен был отмечен группой исследователей [Дюндик, Ковалева, Овсянников 1985], которые обратили внимание на то, что слово *снег* в одних контекстах имеет предметное значение ('белые хлопья, кристаллики льда'), а в других — событийное ('снегопад')¹². И хотя словарь разделяет эти два смысла, словарное толкование «событийного» значения слова *снег* отсылает к понятию «твердых атмосферных осадков», а слово *осадки* определяется как 'атмосферная влага, выпадающая на землю в виде дождя, снега, росы и т.п.' Круг замкнулся, и опорные слова определений *осадки*, *влага* ситуацию не прояснили.

Подобное положение вещей не ограничивается указанной семантической группой. Так, толкование слова *взятка* в первом значении ('деньги или материальные ценности, даваемые должностному лицу как подкуп, оплата действий в интересах дающего') охватывает и предметный, и предикатный смыслы (ср. *дать / получить взятку*, но *осужден за взятку*), которые, на наш взгляд, должны быть разведены (соответственно, в семантико-синтаксическом словаре у данного слова выделяются два ССВ). Таким образом, приписывание семантического признака 'действие' само по себе оказывается непростой исследовательской задачей.

Значение 'качество' (по базовому семантическому признаку) приписывается качественным прилагательным и образованным от них наречиям (точнее, их основным ССВ), а также относительным и притяжательным прилагательным в их производных ССВ, связанных с переходом в разряд качественных прилагательных (ср. *золотой характер, железная воля, каменное лицо, стальные нервы, театральный жест; медвежья походка, собачья жизнь*). Кроме того, им снабжаются существительные со значением признака, образованные от качественных прилагательных (*белизна, грубость*), а также

¹² Кроме того, слово *снег*, наряду с некоторыми другими существительными, даже в своем предметном значении может употребляться в функции событийного имени, ср. *Он не пришел из-за мозолей / грязи / снега* [Арутюнова 1976: 144–145].

оценочные существительные, способные выступать в предикативной функции, характеризуя лицо или вещь, ср. *умница, проходимец, растяпа, лопух, орел, дрянь, прелесть* и т. д.¹³

Значение 'предмет' (по базовому семантическому признаку) получают конкретные существительные — наименования неодушевленных предметов (в том числе собирательные) и веществ, а также относительные прилагательные (*золотые украшения, крестьянская изба, машинное масло*) и притяжательные прилагательные, перешедшие в разряд относительных (*лисий воротник, медвежья шуба*).

В теоретическом плане предложенный подход коррелирует с понятием функциональной транспозиции¹⁴ у Ш. Балли [1955: 130], а также с противопоставлением двух типов семантических отношений между словообразовательно связанными единицами — синтаксической и лексической деривации [Курилович 1962]. В нашем семантическом описании, как видно из приведенных выше примеров, слова, связанные отношениями синтаксической деривации, входят в один и тот же класс, поскольку это преобразование затрагивает лишь синтаксическую функцию производящего слова (вследствие изменения его частеречной принадлежности), но не его семантику (ср. *преподавать* → *преподавание, хитрый* → *хитрость, белый* → *белизна* и т. п.).

Опора на содержательные признаки является необходимым условием снятия (или сокращения) неоднозначности. Если однокоренные слова разных частей речи встречаются в тексте рядом, то, скорее всего, из набора присущих им ССВ реализуются именно те, что характеризуются одинаковыми значениями семантических признаков (случаи каламбуров довольно редки). Ср.:

Было установлено **золотое** содержание червонца — 7,74234 грамма чистого **золота**.

7 января я **переехал** в город; жена, ввиду стоявших сильных морозов, не решилась на **переезд** и осталась еще на неделю в Царском.

И красавица, и **умница**, **умней** моего дурака раз в десять.

Приведенные примеры, заимствованные из Национального корпуса русского языка, ограничены рамками одного предложения¹⁵,

¹³ Ср. [Грамматика Пор-Рояля 1991].

¹⁴ О направлениях транспозиции на материале русского языка см. [Богданов 1977: 157–169].

¹⁵ В силу специфики корпусных данных это замечание справедливо и в отношении последующих примеров.

однако нетрудно догадаться, что на более широком контексте схожее явление наблюдается еще чаще. Следовательно, совпадение значений семантических признаков можно использовать также на следующем этапе анализа для выявления когерентных связей в тексте.

* * *

Выше мы говорили о базовом семантическом признаке и его значениях. Обратимся теперь к тому, что можно совокупно назвать дополнительными семантическими признаками. С их помощью мы пытаемся зафиксировать разнообразные аспекты содержания того или иного ССВ.

Введение различных, не упорядоченных иерархически семантических признаков обеспечивает многомерность описания, при котором каждое слово может рассматриваться как совокупность семантических формул (ср. [Тузов 2003]). Помимо самостоятельной ценности, такой подход лучше справляется с задачей снятия омонимии. Это особенно актуально для существительных и глаголов, которые характеризуются в целом более высоким индексом неоднозначности, чем слова других частей речи. Целесообразность добавления тех или иных семантических признаков напрямую зависит от их способности снимать омонимию у возможно большего числа лексических единиц. Как правило, эта мера имеет также положительный эффект и для решения дальнейших задач в рамках анализа целостного текста.

Так, у существительных собственное значение включает ряд «нетривиальных семантических признаков» [Апресян 1980], имеющих двойное назначение: дифференциацию ССВ лексемы и выявление смысловых связей в тексте. Для решения первой задачи важен, например, такой признак, как «контейнер», позволяющий учитывать регулярный метонимический перенос у предметных существительных, обозначающих разного рода емкости (ср. *стакан стоит на столе vs. выпил целый стакан*). Ввиду того что практически любой предмет может осмысляться в качестве емкости, положительное значение данного признака приписывается только словам (точнее, их ССВ), обозначающим «классические» контейнеры типа *чашка, кастрюля, бочка, бутылка, чемодан, корзина* и т. п.¹⁶

¹⁶ Метонимия в целом представляет серьезную проблему для автоматической обработки языка. Даже регулярные метонимические переносы наподобие вышеупомянутого в словарях трактуются непоследовательно (игнорируются, упо-

Признак «множественность» актуален для существительных, у которых форма единственного числа обозначает множество людей или однородных предметов¹⁷. Частично дублируя морфологический признак собирательности (ср. *студенчество, детвора, воронье, листва, горох*), он не ограничивается существительными *singularia tantum* и позволяет охватить также такие имена, как *семья, полиция, команда, армия, полк, начальство, электорат, персонал, консилиум, флот* (в знач. 'совокупность судов'), *скот, свора, табун, лес, роща, школа* (в знач. 'учителя и/или ученики'), *поселок* (в знач. 'проживающие в нем люди'), *пачка, стопка, кипа, парк (машин), библиотека* (в знач. 'собрание книг') и некоторые другие¹⁸. Многие объекты данного класса отличаются тем, что не имеют компактного расположения в пространстве или могут его утратить (например, *полк* солдат может быть расквартирован в деревне, *коллекция* картин — разделена между музеями, *табун* лошадей может разбежаться по степи, *стопка* бумаг — разлететься по полу и т. д.).

Еще один важный семантический признак для существительных связан со способностью объектов самостоятельно перемещаться и/или быть перемещаемыми. В ряде случаев это позволяет разграничить ССВ слова, ср. *лес* как 'массив деревьев' («перемещаемость» = -) vs. 'строительный материал' («перемещаемость» = +),

минаются в общей дефиниции через запятую, выделяются в качестве оттенков значения). Но охватить все «метонимические формулы» [Шмелев 1973: 224] вряд ли вообще возможно, учитывая их многообразие, а также различную степень закреплённости и контекстной обусловленности. Дополнительным осложняющим фактором являются окказиональные выражения, построенные по той или иной модели. Так, временное значение, реализуемое в сочетаниях *до чая, после армии* ('до чаепития', 'после службы в армии'), допускает различные прочтения в выражении *до телевизора* ('до покупки телевизора' или 'до просмотра телевизионных передач'). Ср. также выражение *до дяди* в приводимой Д. Н. Шмелевым [Шмелев 1973: 225–226] цитате из «Двух капитанов» В. Каверина, ср.: *Мы считали, что этих денег хватит до дяди, а там начнется увлекательная, превосходная жизнь*. В общем случае проблема заключается в том, что там, где конкретная лексика попадает в контексты, требующие имен с абстрактной семантикой, она должна получать событийное прочтение, однако содержащейся в предложении информации недостаточно для того, чтобы восстановить соответствующую пропозицию [Арутюнова 1976: 122–129; 142].

¹⁷ Ср. лексическую функцию Mult в модели «Смысл ↔ Текст» [Мельчук 1974: 88].

¹⁸ О структурных типах множества и их реализации в языке см., например, [Холодович 1979; Румянцева 2009].

земля как 'территория' («перемещаемость» = -) vs. 'почва, грунт' («перемещаемость» = +).

Для дальнейшего анализа текста представляется обоснованным ввести семантические признаки «выделенная часть» и «выделенное направление движения» (ср. [Talmy 1983]), характеризующие такие объекты, как шкаф, кресло, зеркало, дворец, автомобиль (и другие транспортные средства), очередь (см. также [Fillmore 1969; Всеволодова, Владимирский 2009: 208 и след.]). Они позволяют анализировать положения предметов в пространстве, описываемые выражениями типа *вертеться перед зеркалом, стоять за шкафом, левый флигель дворца* и т.д., что связано с более общей задачей интерпретации пространственного дейксиса. Показательно, что некоторые предложения, содержащие отсылки к объектам с выделенной частью или направлением движения, допускают двойное прочтение — дейктическое и недейктическое. В качестве иллюстрации сошлемся на пример Ю. Д. Апресяна [1995б: 635] *Девушка стояла перед машиной*, где при дейктической интерпретации девушка помещается между говорящим и машиной, а недейктическое прочтение учитывает наличие у автомобиля выделенного направления движения, и, соответственно, девушка стоит по ходу движения автомобиля, т.е. перед его капотом. При автоматическом анализе неоднозначность, присущая данному предложению, должна быть сохранена.

Данные признаки можно считать частными случаями признака «абсолютная / относительная пространственная ориентация» [Апресян 1980], ср. *верх (низ) шкафа*, но *верх (низ) кубика*. Абсолютную ориентацию по верху — низу имеют все объекты, обладающие выделенной частью, но не только они, ср. *дерево, гора*. Соответственно, объекты с выделенным направлением движения имеют абсолютную ориентацию по передней части — задней части; таковы, к примеру, транспортные средства.

Признак релятивности (ср. [Шмелев 1973: 135; Palmer 1982: 88]) призван отразить определенную дейктичность семантики слова, ее зависимость от точки отсчета. Положительное значение данного признака характерно для терминов родства, но не только для них (ср. *сосед, одноклассник, попутчик, предшественник*). В то же время для большинства наименований лиц (например, по профессии, ср. *механик, инженер, художник*) этот признак нерелевантен. Важно, что указанные два случая предполагают разные построения в информационной структуре (подробнее см. ниже).

Отметим также, что положительное значение данного признака сближает рассматриваемые существительные со словами других лексико-грамматических разрядов: личными и указательными местоимениями, наречиями, предложениями, частицами *вот* и *вон*, прилагательными типа *верхний* — *нижний*, *старший* — *младший* и другими лексическими единицами, семантика которых предполагает указание на ориентир (ср. понятие *landmark* в работе [Langacker 1988]). Очевидно, что, если бы мы в семантическом описании ограничивались только иерархически упорядоченной онтологической схемой, соответствующий компонент значения, равно как и указанная параллель, не были бы выявлены.

* * *

Большое внимание мы уделяем выделению семантических признаков для глагола, поскольку он является структурным и смысловым центром предложения¹⁹. Прежде всего, следует отметить признаки, связанные со способами глагольного действия. Руководствуясь практическими соображениями, мы использовали только те из них, которые способны служить для дифференциации значений у достаточно большого числа глаголов. Выделить такие семантические признаки и приписать их соответствующим ССВ глаголов не составляет большого труда, если способ действия представляет собой словообразовательный разряд (к тому же словарные дефиниции, как правило, строятся однотипно). В случае непоследовательно характеризованных способов действия эта задача требует ручной работы.

Перечислим некоторые способы глагольного действия, указание на которые включается в собственное значение глаголов:

- **начинательный** — *заговорить* ('начать говорить' vs. 'утомить разговором'), *заходить* ('начать ходить' vs. 'приходить куда-л.'), *пойти* ('начать идти' vs. 'вступить в брак');
- **ограничительный** — *поцарапать* ('царапать некоторое время' vs. 'оцарапать'), *пошить* ('шить некоторое время' vs. 'сшить');
- **длительно-ограничительный** — *проговорить* ('говорить некоторое время' vs. 'сказать'), *прогулять* ('гулять некоторое время' vs. 'не выйти на работу');
- **накопительный** — *налетать* ('пролететь в общей сложности какое-л. число часов, километров' vs. 'натолкнуться на преграду');

¹⁹ Более широко о регулярных семантических признаках глагола см. [Скребцова 2015а].

ду'), *сносить* ('износить' и vs. 'отнести куда-л. и принести обратно'), *вырезать* ('истребить резней' vs. 'удалить, извлечь'), *исходить* ('ходя, побывать во многих местах' vs. 'брать что-л. за основу');

- распределительный — *порубить* ('вырубить все, многое' vs. 'рубить некоторое время'²⁰), *перебить* ('убить всех, многих' vs. 'не дать договорить, прервать'), *перекатать* ('покатать всех, многих' vs. 'быстро переписать'), *переболеть* ('перенести много болезней' vs. 'перенести какую-л. болезнь');
- репродуктивный — *пересказать* ('рассказать заново' vs. 'рассказать все, многое'²¹), *пересушить* ('просушить заново' vs. 'сделать слишком сухим'²²);
- чрезмерно-нормативный — *перепечь* ('подвергнув более длительному или более сильному действию жара, лишить нужных вкусовых качеств' vs. 'испечь все, многое'²³), *перестоять* ('простоять дольше, чем нужно' vs. 'простоять где-л., пережидая').

Способы действия пронизывают систему глагольных значений в русском языке, вступая между собой в различные комбинации в семантике конкретных глаголов. Эффективность соответствующих признаков для семантико-синтаксического анализа очевидна.

Кроме указания на способ действия, для глаголов вводится признак «фаза» (со значениями 'начало'²⁴, 'продолжение', 'остановка', 'конец' и др.²⁵), который также способствует снятию неоднозначности у целого ряда глаголов. Так, семантическая структура глаголов типа *белеть*, *светлеть*, *сохнуть* характеризуется регулярной многозначностью: одно (или более) значений связано с идеей становления соответствующего свойства, а другое (другие) выражают наличие постоянного свойства, воспринимаемого неким наблюдателем. Следовательно, признак «фаза» (со значением 'становление') релевантен для первого типа значений, но не для второго.

²⁰ Ограничительный способ действия, см. выше.

²¹ Распределительный способ действия, см. выше.

²² Чрезмерно-нормативный способ действия, см. ниже.

²³ Распределительный способ действия, см. выше.

²⁴ Множество соответствующих глагольных значений в качестве своего подмножества содержит те ССВ, которые помечены признаком начинательного способа действия (см. выше).

²⁵ Понятие фазовости трактуется нами шире, чем это обычно принято, ср. также [Богданов 1985; Храковский 1987].

Характеризация глагольного значения по признаку «фаза» может оказаться полезной в дальнейшем при выявлении смысловых связей не только внутри предложения, но и текста. Возьмем в качестве примера следующую фразу:

*Остановившийся было процесс разложения **возобновился**, грозя со-
всем развалить фронт, а с ним и Россию.*

На фазу в этом примере указывают три токена: *остановившийся* и *возобновился* характеризуют стадии процесса *разложения*, в то время как деепричастие *грозя* реализует фазовое значение 'предбытие' применительно к событию, обозначенному глаголом *развалить* (*фронт*). Кроме того, данное деепричастие выражает причинно-следственную связь между пропозициями «процесс разложения возобновился» и «фронт разваливается».

В отношении деепричастия *грозя* следует также заметить, что оно (как и упоминавшиеся ранее глаголы *белеть*, *светлеть* и *сохнуть*) содержит в себе сему 'наблюдатель'²⁶. Вообще, эта сема присутствует у достаточно широкого пласта лексики (причем не только глагольной), связанной с восприятием качеств и признаков объектов (ср. *мелькать*, *шелестеть*, *сияние*, *горький* и т. д.). Выделение соответствующего семантического признака для характеристики глагольных ССВ обусловлено не столько его ролью в снятии лексической и синтаксической неоднозначности, сколько способностью восполнять имплицитные связи между пропозициями в тексте. Он позволяет связать информацию о воспринимаемом объекте с самим экспериенцером, который не входит в число актантов соответствующего глагола, но может употребляться в тексте при другом глаголе, ср.:

Там мелькали обольстительные призраки, но он не мог разглядеть их; слышались смешанные звуки — то голос славы, то любви: все это приводило его в сладкий трепет.

Еще один важный аспект глагольной семантики — признак «контролируемости / неконтролируемости» действия, позволяющий дифференцировать, к примеру, следующие ССВ глаголов²⁷, ср.:

²⁶ «Наблюдатель» понимается нами широко — как субъект не только перцепции, но и сознания (ср. [Падучева 2006]).

²⁷ Подробнее о контролируемости действий и отражении этого аспекта в языке см. [Зализняк 1992; Кустова 1992].

бренчать: 'издавать звуки, подобные ровному глухому звону' («контролируемость» = -) vs. 'производить, вызывать такие звуки' («контролируемость» = +, ср. *бренчать ключами*);

моргать: 'непроизвольно быстро опускать и поднимать веки и ресницы' («контролируемость» = -) vs. 'подавать знак движением век' («контролируемость» = +);

пройти: 'оказаться принятым, зачисленным, избранным' («контролируемость» = -, ср. *пройти по конкурсу*) vs. 'идя, совершить путь куда-л.' («контролируемость» = +).

Следует заметить, однако, что, будучи «скрытой» категорией [Филлмор 1981], контролируемость действия не обязательно выражается в том или ином контексте. Ср., например, следующие примеры из работы [Апресян 1995а: 176], где содержательная оппозиция контролируемости / неконтролируемости нейтрализуется: *Джон прострелил себе руку, Гусар загремел шпорами, Он разбил окно локтем, Большой моргнул* и т. д.

С точки зрения анализа целостного текста представляют интерес и другие семантические признаки, связанные с характером протекания действия, обозначенного глаголом. Так, по признаку, который можно назвать «предельность» или «целостность» (ср. [Виноградов 1947; Маслов 1959; Мелиг 1985; Телин 1985; Бондарко 1987]), всем глаголам совершенного вида может быть автоматически приписано значение ' +' (что, правда, в некотором роде дублирует уже имеющуюся информацию: если глагол совершенного вида, то он непременно предельный). Что же касается глаголов несовершенного вида, здесь требуется ручная работа по разделению тех из них, которые, по выражению Г. Г. Сильницкого [1983], характеризуются «потенциальной предельностью», и неопредельных глаголов.

Другой кандидат в семантические признаки — понятие кратности, которое, как правило, рассматривается совокупно с понятиями длительности и интенсивности в рамках количественной аспектуальности²⁸. В то же время иногда утверждается, что три указанных понятия представляют собой независимые семантические признаки, каждый из которых может реализоваться в виде отдельной категории [Храковский 1989: 11]. Это позволяет рассматривать кратность

²⁸ Ср.: «... в славянских языках количественная аспектуальность воплощается в различных частных видовых значениях форм СВ и НСВ, выступают в качестве способов действия (многократные глаголы типа *хаживать, есть*), фактивные — типа *стукнуть, ахнуть* и т. д.)...» [Маслов 1978: 19].

вне контекста, связанного с глагольными способами действия (ср. [Бирюлин 1989]). Если взять за основу такой подход, то данный семантический признак может быть распространен на глаголы движения, традиционно описываемые в терминах оппозиции моторно-некратные vs. моторно-кратные (*бежать* — *бегать*, *вести* — *водить*), их многочисленные приставочные дериваты, а также глаголы, обозначающие колебательное движение (*болтаться*, *шататься*) (ср. [Апресян 1995а: 72]).

Что же касается других, более тонких аспектов глагольной семантики, заинтересованно обсуждавшихся в литературе после появления знаменитой классификации З. Вендлера (ср. признаки 'актуальность', 'наличие цели', 'мгновенность'), они, как справедливо замечает Х. Мелиг [1985], отличаются высокой зависимостью от контекста. Сложные переплетения функционально-семантических полей аспектуальности, темпоральности, залоговости, модальности, субъектности / объектности и др. в русском языке (см. [Бондарко, Буланин 1967; Пупынин 1991, 1996; Бондарко 1992; Храковский 1996]) не позволяют использовать эти признаки для характеристики глаголов как единиц языка.

* * *

У прилагательных выделяются параметрические ССВ, характеризующие предмет с точки зрения размера (величины, высоты, глубины, длины, ширины, толщины), температуры, веса. По значению «параметр = +» / «параметр = -» легко дифференцировать прямые и метафорические значения соответствующих прилагательных, ср.: *высокий дом* vs. *высокие отношения*, *широкая кровать* vs. *широкие взгляды*, *глубокое озеро* vs. *глубокий анализ* и т. п. Наличие в тексте прилагательного с положительным значением данного признака нередко служит сигналом скорого появления словосочетания с количественным числительным и, следовательно, оказывается полезным для выявления смысловых связей в предложении и тексте.

Применительно к прилагательным также заслуживает внимания предложенный Ю. Д. Апресяном [1995а: 66–67] семантический признак 'предельность'. Прилагательное является предельным, если оно может сочетаться с наречиями *абсолютно*, *совсем*, *совершенно*, *идеально*: к ним относятся прилагательные цвета, формы, параметрические прилагательные (*короткий*, *узкий*, *тонкий*, *холодный* и т. д.), а также прилагательные, обозначающие соответствие / несоответствие норме (типа *здоровый*, *хороший*). Присутствие в значении со-

ответствующей семы разграничивает, например, значения прилагательного *короткий* (ср. *совсем короткий шнур* vs. **совсем короткое замыкание*) и *красный* (ср. *абсолютно красный галстук* vs. **абсолютно красный воин*).

АЛГОРИТМ КОНСТРУКТИВНО-СИНТАКСИЧЕСКОГО АНАЛИЗА

Цель конструктивно-синтаксического анализа — выявление синтаксической структуры предложения. Как и на предыдущих этапах обработки текста, мы стремимся к тщательному учету линейного контекста, чтобы максимально сократить возможную неоднозначность. В соответствии с общим принципом работы лингвистического процессора это происходит исключительно на основе языковых данных, без привлечения статистических методов. Если имеющаяся информация не позволяет сделать надежный выбор в пользу какого-нибудь одного из альтернативных вариантов синтаксического разбора предложения, все они сохраняются — вплоть до следующего этапа анализа²⁹. Если какой-то элемент не удастся синтаксически связать ни с одним членом предложения, он временно (до следующего этапа) остается в «подвешенном» состоянии.

В методологическом плане нам близок подход В.А. Тузова, который формулирует свою исходную посылку следующим образом: «Поскольку компьютер не способен понимать тексты, а может их только выполнять, то остается практически один способ формализации языка — рассматривать предложение как запись последовательности выполнимых действий. Нравится нам это или нет, компьютер не оставляет нам другого выбора» [Тузов 2003: 4]. Исходя из этого единственно верным способом представления структуры предложения автору видится суперпозиция функций, ср.: «Предложение является алгоритмом, и никакой статический тип данных не адекватен ему. Поэтому, если какая-либо грамматика приписывает какому-то предложению нечто, отличное от суперпозиции (древовидную структуру, графовидную или нечто подобное), то такая грамматика не может быть адекватной естественному языку» [Тузов 2003: 429].

²⁹ Синтаксическая неоднозначность широко встречается в публицистических текстах, особенно в заголовках информационных сообщений, ср. вынесенное в заголовок предложение *У пострадавшего от нападения балетмейстера Филина улучшилось зрение* (<http://news-w.com/62851-18021/n/62851-8631/>), которое допускает три синтаксических прочтения.

В описываемом лингвистическом процессоре синтаксический анализ работает с ближайшим линейным контекстом с целью связать между собой соседние элементы. Начиная от левого конца предложения к его правому концу идет последовательная сборка токенов в синтаксические конструкции. Как правило, этот процесс включает несколько проходов. Это обусловлено, во-первых, тем, что синтаксические правила ранжированы нами по группам. Сначала в масштабе всего предложения происходит объединение токенов, связанных сильными локальными связями, затем — более слабыми и/или более удаленными и т. д. Во-вторых, каждый проход анализатора предполагает работу с окном вплоть до пяти элементов³⁰. Очевидно, что, если предложение состоит из шести и более токенов, границы этого окна будут поэтапно смещаться от левого к правому краю предложения (1–5, 2–6, 3–7 и т. д.). В-третьих, каждый раз в рамках одной и той же пятерки элементов может осуществляться поиск релевантных связей не только на максимальном удалении друг от друга (т. е. в рамках всего окна), но и на меньшем расстоянии: если не найдено правило для пяти элементов, далее осуществляется поиск подходящего правила для первых четырех элементов, затем первых трех и, наконец, первых двух элементов. В тот момент, когда какие-то элементы (на начальном этапе — токены) объединились в конструкцию, они, в свою очередь, становятся единым элементом, и окно тем самым «растягивается». Таким образом, при укрупнении элементов мы сближаем между собой сколь угодно удаленные токены и в итоге собираем все предложение³¹.

Результат конструктивно-синтаксического анализа представляет собой последовательность групп ПСДО: подлежащего, сказуемого, дополнений (прямого и косвенных) и обстоятельственных детерминантов, относящихся ко всему предложению в целом³². При же-

³⁰ «Элемент» здесь является понятием относительным: он может представлять собой либо одиночный токен, либо ранее собранную конструкцию.

³¹ Справедливости ради следует заметить, что для сборки целого предложения пяти элементов не всегда хватает — хотя бы потому, что в русском языке есть шестивалентный глагол *командировать*, не говоря уже о некотором числе пятивалентных. Если в предложении имеется вдобавок обстоятельственный (-е) детерминант (-ы), очевидно, что оно не поместится целиком в пятиэлементное окно. Поэтому при сборке целого предложения (и только в этом случае) используется более широкое окно.

³² Детерминант понимается здесь несколько иначе, чем у автора данного термина, ср. [Шведова 1990]. Основным критерием для нас является фактор отнесенности соответствующего слова или конструкции к предложению в целом,

лании указанные группы можно легко организовать в древовидную структуру. Очевидно, что групп дополнений, так же как и групп обстоятельств, в предложении может быть больше одной. Это же верно и в отношении подлежащего, если оно представлено рядом однородных членов.

Если предложение содержит однородные сказуемые, ему соответствует более одного ПСДО (по числу сказуемых). Так же обстоит дело со сложносочиненными предложениями и сложноподчиненными предложениями с расчлененной структурой. Сложноподчиненным предложениям с присловным придаточным приписывается одна структура ПСДО. Если предложение синтаксически неоднозначно, оно тоже имеет более одного ПСДО.

Каждая из групп ПСДО может быть реализована в тексте посредством многочисленных разновидностей, которые получают формализованное описание. Так, группа подлежащего может состоять из одного или нескольких однородных существительных, местоимений, собирательных числительных, инфинитивов, а также включать разнообразные сильные и слабые управляемые члены, в том числе:

- зависимые существительные, выражающие объектные или определительные значения, с предлогом или без (*дом отца, конь в яблоках, влияние глобализации на климат Земли, дом у реки*);
- согласованное прилагательное (*красивый дом*), причастие (*разрушенный дом*), местоимение (*мой дом*) или порядковое числительное (*второй дом*);
- количественное или собирательное числительное (*два дома, оба дома, трое солдат*);
- существительное или именную группу в функции приложения (*река Москва, генерал от инфантерии Михаил Дмитриевич Скобелев*);
- определительный оборот (*белые от страха глаза*), в том числе причастный (*дом, покосившийся на один бок от старости*), в пре- или постпозиции;

а не к какому-то его отдельному члену. Вследствие этого в понятие детерминанта включаются не только собственно обстоятельственные обороты (за исключением обстоятельств образа действия, связанных более с предикатом, нежели с ситуацией в целом), но и ряд других разнородных феноменов — обращения, междометия, вводные слова и конструкции. Сюда же относятся модальные частицы, которые передают субъективное отношение автора к сообщаемому, а значит, «суперсегментны по своей сути» [Николаева 1985: 15]. Таким образом, «обстоятельственный детерминант» является для нас удобным «зонтичным» рабочим термином.

- определительное придаточное предложение (*дом, который я помнил с детства; тот, к кому я обратился с вопросом*);
- изъяснительное придаточное предложение с указательным словом и без (*слух (о том), что открылась выставка*);
- примыкающий инфинитив (*совет пойти учиться*), наречие (*взгляд исподлобья, дом напротив*), существительное или числительное (*дом номер восемь, дом восемь*).

Кроме того, возможны разнообразные комбинации перечисленных выше случаев (*два старых разрушенных дома, старый отцовский дом, который я помнил с детства* и т. д.).

Обратимся теперь к подробному рассмотрению процесса сборки предложения при помощи синтаксических правил. Правила осуществляют объединение элементов в некоторую результирующую конструкцию (в дальнейших манипуляциях она, в свою очередь, выступает в качестве элемента). При этом каждый элемент должен удовлетворять определенным условиям, которые фактически представляют собой значения признаков (графематических, морфологических, семантических и синтаксических)³³.

Допустим, что мы рассматриваем правила на два элемента. В этом случае речь может идти о конкретных значениях признаков для обоих членов правила (например, примыкание наречия с фиксированной позицией к главному слову, ср. *уха по-царски*), либо об одном конкретном значении, а другом относительном (так, род, число и падеж прилагательного (-ых) должны совпадать с соответствующими характеристиками определяемого существительного), либо, наконец, об относительных (взаимозависимых) значениях признаков для обоих элементов (например, в сочетании предлога с существительным требуется, чтобы падеж существительного совпадал с падежом, которым управляет предлог).

Результирующему объекту правило приписывает значения грамматических и семантических признаков. Так, словосочетанию

³³ То обстоятельство, что каждый токен несет на себе приписанные ему ранее значения не только морфологических, но и графематических признаков, является важным и способствующим правильной интерпретации предложения преимуществом описываемого лингвистического процессора. Если информацию, передаваемую знаками препинания, не принимать в расчет, может возникнуть ложная синтаксическая омонимия (классический пример *казнить нельзя помиловать* отнюдь не является единственным в своем роде). Подробнее о роли пунктуации в разрешении неоднозначности в системах автоматической обработки текста см. [Бердичевский, Иомдин 2007].

дубовый стол будет приписана следующая информация: «тип конструкции = существительное с согласованным определением, семантические признаки: базовый: предмет; мебель +, дерево +». Как мы видим, здесь произошло объединение значений признаков обоих слов. Однако так бывает не всегда: характеристика результирующего объекта может отличаться от характеристик членов конструкции. В качестве примера укажем на сочетания существительного с предлогом, «рождающие» новый смысл: если *стол* — это «часть речи = существительное; род = м; число = ед; падеж = им/вин», то *на стол, в стол* — это «тип конструкции = обстоятельство-места-куда»³⁴.

Синтаксические правила сгруппированы нами в соответствии с их рангом. Правила первого ранга, как было отмечено выше, — это правила объединения токенов, связанных наиболее тесными локальными связями. Правила второго ранга служат для объединения элементов (часть из которых могут представлять собой конструкции, собранные на предыдущем уровне), которые связаны менее сильными связями. Правила третьего ранга направлены в основном на присоединение обособленных оборотов и придаточных предложений к распространяемому члену (главного) предложения.

Разумеется, четкое распределение правил по уровням в известной мере идеализирует реальное положение дел: ведь для того чтобы создать жесткую последовательность их применения, необходимо исчислить возможные ситуации взаимного расположения токенов, учитывая весь набор их признаков, от графематики до семантики, но подобного описания нет, и едва ли оно возможно в принципе. Возьмем, к примеру, последовательность «существительное + существительное в родительном падеже». Если она стоит в конце предложения, можно утверждать, что мы имеем дело с генитивным оборотом, но когда она встречается в другой позиции, такой уверенности уже нет. Ср.: *На собрании выступил плотник совхоза* vs. *Плотник совхоза не мог покинуть*. Бывает так, что после выполнения правил

³⁴ Ср.: «семантический анализатор в процессе обработки исходного текста выполняет два основных действия: выбор правильного значения (предиката — формального аналога понятия) слова и построение из выбранных значений более сложного предиката. Это построение всегда носит характер функционального взаимодействия, в результате которого вычисляется новое значение, в частности — новый семантико-грамматический тип построенной конструкции» [Тузов 2003: 6].

более высокого ранга приходится возвращаться к правилам более низкого ранга.

Опишем теперь классические случаи, подпадающие под действие правил различных рангов. К правилам первого ранга относятся следующие:

- существительное + согласованное с ним одно или несколько полных прилагательных;
- предлог + существительное в соответствующей падежной форме³⁵;
- предлог + согласованное с существительным прилагательное + существительное в соответствующей падежной форме³⁶;
- частица *не* + следующее за ней знаменательное слово.

Здесь следует сделать оговорку. С целью сокращения числа правил связывания некоторые части речи и/или лексико-грамматические разряды могут объединяться в более крупные классы слов в соответствии с их синтаксическим поведением. Так, согласование с определяемым существительным по роду, числу и падежу, помимо прилагательных, наблюдается у причастий, порядковых числительных, некоторых типов местоимений. Все эти разряды слов в некоторых правилах совокупно считаются «прилагательными». С другой стороны, есть такие местоимения (так называемые местоимения-существительные (Русская грамматика 1980)), которые ведут себя в предложении подобно существительным. В формулировке некоторых правил термин «существительное» также понимается именно в таком расширенном смысле.

К правилам первого ранга относятся также те, что направлены на сборку аналитических форм глагола, а также глаголов с фазовыми и модальными операторами, прилагательных и наречий в сравнительной и превосходной степенях, составных числительных, Ф. И. О. в полном виде (в том числе с буквенными сокращениями типа *тов.*, *проф.*, *д-р*, *г-н* в препозиции), аппозитивных именных групп (*генерал Иванов, отличник Петя, газета «Известия»*). Таким образом, выполнение правил первого ранга дает нам то, что по ана-

³⁵ Информация о том, какого падежа требует тот или иной предлог, поступает из Семантико-синтаксического словаря.

³⁶ Данное правило фактически является суммой двух предыдущих; в качестве самостоятельного оно вводится для того, чтобы не упустить возможную интерпретацию токенов типа *больной* в определенных контекстах (пример см. ниже).

логии с графематическим анализом мы можем назвать «синтаксическими атомами».

Сборка токенов в ту или иную конструкцию может быть затруднена из-за специфики синтаксического контекста. Так, для вышеупомянутого правила объединения существительного и согласованного с ним прилагательного представляет сложность ситуация, в которой прилагательное «зажато» между двух существительных и формально согласовано с ними обоими. Она хорошо иллюстрируется известной строкой из стихотворения А. С. Пушкина *Роняет лес багряный свой убор*³⁷, где прилагательное *багряный* может относиться к любому из существительных *лес* и *убор*. Этот вид неоднозначности неразрешим в рамках конструктивно-синтаксического анализа предложения.

Другая пушкинская фраза *Друзья кровавой старины народной чаяли войны* (поэма «Полтава») также формально неоднозначна, ср.:

- 1) *друзья (кровавой старины)* — генитивная конструкция, где зависимое существительное имеет при себе согласованное определение;
народной (чаяли) войны — нарушенная проективность в сочетании сказуемого с группой прямого дополнения;
- 2) *друзья (кровавой старины народной) чаяли войны*.

Второй вариант интерпретации, однако, отклоняется в силу введенного нами ограничения, согласно которому все согласованные с существительным полные прилагательные должны располагаться либо слева от него, либо справа. Нахождение одного из них в препозиции, а другого в постпозиции возможно разве что в разговорной речи, причем, как правило, в конце предложения, ср.: *Купил карамельки и шоколадных конфет вкусных*.

Следует заметить, что автоматический анализ структуры пушкинских строк даже при условии сохранения только первого варианта все равно затруднен вследствие разрыва синтаксической конструкции «прилагательное + существительное» (*народной чаяли войны*).

³⁷ Хотя описываемый лингвистический процессор не предназначен для автоматического анализа поэтической речи (см. введение), в изложении мы используем отдельные примеры из стихотворных произведений в силу их широкой известности; сконструировать структурно идентичный прозаический пример не составляет труда.

Трактовка последнего примера демонстрирует важный аспект используемых нами правил, а именно их сугубо эмпирический характер, обусловленный внимательным наблюдением над языковым материалом. Собственно говоря, у нас не было другого пути — как отмечалось выше, готового перечня интерпретаций для всевозможных линейных комбинаций из слов различных частей речи просто не существует. В связи с этим приведем еще одно наблюдение, касающееся обработки сочетания «существительное + согласованное с ним полное прилагательное». Нами было установлено, что если между прилагательным и существительным «вклинивается» предложно-падежная форма, она является несогласованным определением к тому же существительному, ср. *черная в горошек блузка*. Иными словами, предложно-падежная форма в таком контексте не нарушает проективности — в отличие от глагола (*народной чаяли войны*). Последовательность обработки такого словосочетания следующая: сначала собирается синтаксическая группа *в горошек*, а затем она включается в конструкцию «согласованное определение + несогласованное определение + определяемое существительное»³⁸.

Грамматическая омонимия слов типа *больной, военный, пожарный, слепой* потребовала введения упомянутого выше правила «предлог + согласованное с существительным прилагательное + существительное в соответствующей падежной форме», которое обеспечивает успешный анализ предложений типа *К хромому мужику подошли двое*. Если бы мы располагали исключительно правилами на два элемента «предлог + существительное» и «прилагательное + существительное», то было бы ошибочно собрано словосочетание *к хромому (хромой как существительное)*, а словоформа *мужику* оказалась бы «подвешенной», и предложение не было бы собрано. Наличие данного правила обеспечивает сборку всей именной группы *к хромому мужику*, которая в дальнейших операциях выступает в качестве единого элемента.

Данный пример также демонстрирует преимущество выбранного нами размера окна (подробнее об этом см. ниже). Если бы окно было размером в два элемента, правил на три элемента просто не существо-

³⁸ Вся эта процедура повторяется, когда подобные конструкции вложены одна в другую, ср.: *Он проглатывает слюну и смотрит на меня поверх очков **небольшими**, с нависшей над веками кожей глазами*» (В. Некрасов. В окопах Сталинграда).

вало бы, и нам пришлось бы обходиться вышеупомянутыми правилами на два элемента. Тогда также произошла бы ошибочная сборка словосочетания *к хромоту*, не позволяющая в итоге собрать предложение целиком. Окно же в пять элементов позволяет не только вместить именную группу целиком, но и увидеть следующий за ней глагол *погоди*, который, собственно говоря, и сигнализирует ее конец.

Грамматическая омонимия слов типа *больной, хромой, слепой* может приводить к синтаксической неоднозначности, ср.: *Про больного доктора говорили, что он вряд ли поправится*. В целях ее сохранения параллельно применяются соответствующие правила на два и три элемента, что приводит к сборке следующих структурных вариантов предложения:

- 1) *про больного* (предлог + существительное) *доктора* (подлежащее) *говорили...*;
- 2) *про больного доктора* (предлог + прилагательное + существительное) *говорили...* (неопределенно-личное предложение).

Оба эти варианта оказываются допустимыми. Это может показаться парадоксальным, но сохранение присущей естественно-языковому тексту синтаксической омонимии также является важной задачей автоматического анализа текста, хотя сталкиваться с ней приходится существенно реже, чем с обратной проблемой — снятием неоднозначности.

Размер окна в пять элементов выбран в качестве оптимального сугубо эмпирически. Для языка со свободным порядком слов очевидна необходимость учета как левого, так и правого контекста, что в качестве минимума дает окно в три элемента. В принципе, можно было бы обойтись и этим размером, но во многих случаях он тормозит выявление синтаксических конструкций. Возьмем, к примеру, последовательность из трех токенов *красные и белые*. Можно предположить, что это однородные члены предложения. Но, во-первых, мы не можем определить, прилагательные это или существительные. Во-вторых, мы не знаем, весь ли это однородный ряд целиком или за рамками окна остались еще какие-то его члены (ср. *зеленые, красные и белые шарики*). Следовательно, мы не можем его сразу собрать. Заметим, что окна в пять элементов оказалось бы достаточно, чтобы вместить весь этот однородный ряд целиком. При этом очевидно, что три однородных члена в предложении встречаются довольно часто — в отличие от пяти и более

членов. Наконец, вовсе не обязательно мы имеем здесь дело с однородными прилагательными, ср.: *кубики красные и белые шарик*и. В конечном итоге даже при окне в три элемента эту конструкцию удастся собрать, но скорость обработки существенно снизится. Еще одним фактором в пользу расширения окна являются частые нарушения проективности в русском языке. В то же время чрезмерное увеличение размера окна (до семи и более элементов) негативным образом сказывается на быстродействии анализатора. За счет многопроходности окно в пять элементов позволяет собрать все синтаксические группы (ПСДО) предложения³⁹.

Как уже говорилось, поиск подходящего правила в рамках текущего окна всегда начинается для максимально возможного числа элементов. Иными словами, сначала ищется правило на пять элементов (текущее + следующие четыре справа от него). Такое правило у нас одно-единственное, а именно «предлог + три полных прилагательных, разделенных пробелами (без знаков препинания) и согласованных с определяемым существительным + существительное в соответствующей падежной форме», ср. *Мы подошли [к большому белому деревянному дому]*. Когда соответствующая конструкция собрана, она становится единым элементом, который в дальнейших манипуляциях выступает в качестве цельного блока. Если подобной ситуации в тексте не обнаружено, окно сужается на один элемент и ищутся правила, которые могут быть применимы к текущей последовательности из четырех элементов. Прежде всего, это правило аналогично тому, что приведено выше для пяти элементов, с тем исключением, что вместо трех неоднородных прилагательных в нем фигурируют два. Соответственно, для трех элементов существует правило «предлог + прилагательное + существительное», для двух элементов — «предлог + существительное» (о них уже шла речь выше). Разумеется, имеются и другие правила для сборки конструк-

³⁹ При условии правильно построенного предложения. Как бы ни было условно и растяжимо это понятие, все же оно имеет границы, и их нарушение ощущается носителями языка. Время от времени в печати появляются предложения, анализ которых представляет сложность даже для человека, не говоря уже о компьютере, ср.: *Как сообщили в ГУ МВД РФ по СПб и ЛО, минувшей ночью, около 00.10, у дома № 2 по Торфяной дороге неустановленный автомобиль совершил наезд на переходившего проезжую часть по пешеходному переходу полицейского управления вневедомственной охраны* (петербургская интернет-газета «Фонтанка.ру» 18.11.2012; газета «Метро» 18.11.2012).

ций из двух, трех и четырех элементов — некоторые из них были упомянуты выше.

Прежде чем рассмотреть пример сборки предложения посредством правил первого ранга, следует заметить, что, помимо «ранжированных» правил, есть еще особые правила, направленные на проверку полноты предложения, т. е. на их соответствие определенному структурному шаблону. Они выполняются перед запуском каждого следующего ранга правил, в том числе и в самом начале анализа, перед правилами первого ранга. Если структурная схема предложения соответствует тому или иному шаблону, применение правил сборки вообще не требуется.

Так, предложение *Мама мыла раму* соответствует шаблону: «существительное в именительном падеже + переходный глагол в личной форме, согласованный с существительным + существительное в винительном падеже без предлога», — что позволяет сразу собрать его, не привлекая правила каких бы то ни было рангов. Кроме того, это уберегает нас от ошибочной сборки *мама мыла* как «существительное в именительном падеже + существительное в родительном падеже» (правило второго ранга, см. ниже) и, следовательно, «лишней» омонимии. Шаблон допускает перестановку элементов, ср.: *Мама раму мыла* и т. д.

В структурно схожем примере *Директор школы не любил* проверка полноты предложения первоначально не дает результата. После применения правила первого ранга для объединения частицы *не* с глаголом *любил* в единый элемент проверка осуществляется вновь и на этот раз приводит к сборке предложения в соответствии с тем же шаблоном, что был использован в случае *Мама мыла раму*⁴⁰. Примечательно, что автоматический анализ здесь не допускает ошибки, обыкновенно совершаемой носителем языка при обработке подобных предложений (которые в англоязычной лингвистической литературе принято называть *garden-path sentences*).

Шаблон, актуализированный в рассмотренных примерах, на самом деле является частным случаем общего рабочего принципа, связанного с проверкой заполнения актантных позиций глагола. Если предложение состоит исключительно из глагола и именных

⁴⁰ Наличие отрицания при глаголе существенно в том смысле, что допускает употребление дополнения как в форме беспредложного винительного, так и родительного падежей. В данном случае следствием является неоднозначность формы *школы*: либо родительный падеж единственного числа, либо винительный падеж множественного числа.

групп, морфологическая форма и семантическое содержание которых полностью соответствует условиям реализации какого-либо из глагольных ССВ, то предложение собирается сразу, ср. *Петя забил гвоздь в стену молотком*. Такая ситуация, правда, встречается нечасто.

Более распространенными являются следующие ситуации. Первая: число именных групп в тексте меньше, чем число актантов, указанных в условиях реализации, ср. *Петя забил гвоздь в стену*. Если морфологическое оформление и семантическое содержание наличных актантов совпадает с условиями реализации, предложение также считается полным и собирается сразу. Вторая: число именных групп в тексте больше, чем число заполненных условий реализации, ср. *Петя забил гвоздь в стену молотком + утром / на даче / для развлечения / ...*. Если «лишняя» именная группа представляет собой нераспространенный обстоятельственный детерминант, относящийся к предложению в целом, оно также собирается сразу. В том случае, когда мы имеем дело с обстоятельством образа действия, относящимся непосредственно к глаголу (ср. *Петя легко / с трудом / с первой попытки забил гвоздь в стену молотком*), требуются дополнительные шаги по сборке группы данного обстоятельства (если оно распространенное) и присоединению ее к сказуемому.

Проиллюстрируем теперь последовательный процесс сборки синтаксической структуры на примере предложения *Маша купила черную в белый горошек блузку*.

0 — не найдено правило сборки полного предложения.

1.1. [*Маша купила черную в белый*] *горошек блузку* — правило не найдено.

1.2. [*Маша купила черную в*] *белый горошек блузку* — правило не найдено.

1.3. [*Маша купила черную*] *в белый горошек блузку* — правило не найдено.

1.4. [*Маша купила*] *черную в белый горошек блузку* — правило не найдено.

2.1. *Маша [купила черную в белый горошек]* *блузку* — правило не найдено.

2.2. *Маша [купила черную в белый]* *горошек блузку* — правило не найдено.

2.3. *Маша [купила черную в]* *белый горошек блузку* — правило не найдено.

2.4. *Маша [купила черную] в белый горошек блузку* — правило не найдено.

3.1. *Маша купила [черную в белый горошек блузку]* — правило не найдено.

3.2. *Маша купила [черную в белый горошек] блузку* — правило не найдено.

3.3. *Маша купила [черную в белый] горошек блузку* — правило не найдено.

3.4. *Маша купила [черную в] белый горошек блузку* — правило не найдено.

4.1. *Маша купила черную [в белый горошек блузку]* — правило не найдено.

4.2. *Маша купила черную [в белый горошек] блузку* — правило найдено. Собрана предложно-падежная группа.

5.1. [*Маша купила черную* (предложно-падежная группа) *блузку*] — правило не найдено.

.....
6.1. *Маша [купила черную* (предложно-падежная группа) *блузку*] — правило не найдено.

.....
7.1. *Маша купила [черную* (предложно-падежная группа) *блузку*] — правило найдено! Собрана группа существительного с согласованным и несогласованным определениями.

8. [*Маша купила* («существительное с определениями»)] — найдено правило сборки полного предложения. Предложение собрано: существительное в именительном падеже + переходный глагол в личной форме, согласованный с существительным + существительное в винительном падеже без предлога.

Применение правил первого ранга повторяется для предложения до тех пор, пока срабатывает хоть одно правило. Если не найдено ни одного правила, то работа с правилами первого ранга заканчивается.

* * *

Правила второго ранга имеют своей целью сборку следующих конструкций:

- однородные члены предложения (*синие, зеленые, красные и белые шарики; отличник Петя, хромой мужик и генерал Иванов*⁴¹);

⁴¹ После применения правил первого ранга каждый член однородного ряда представляет собой целостный элемент.

- существительное + существительное в родительном падеже (генитивная конструкция);
- количественное числительное + существительное в соответствующем падеже;
- даты, представленные нецифровым шаблоном (*пятнадцатому мая, 8 марта, 28-го октября, 13 апреля 2012 г.*⁴²).
- число, записанное буквами или цифрами (с буквенным окончанием или без) + буквенное сокращение + существительное в родительном падеже (*два кг яблок, 2 кг яблок, 115-й км трассы*).
- наречие + главное слово.

Обрисует вкратце проблемы, которые могут возникнуть при применении правил сборки данных конструкций.

Применительно к сочинительным конструкциям следует прежде всего отметить возможность несовпадения морфологических характеристик у однородных членов и даже их функциональной разнотипности, что очевидным образом затрудняет автоматический анализ, ср.:

Купил чаю и коробку спичек;

Впереди стоял комендант, старик бодрый и высокого роста (А. С. Пушкин. Капитанская дочка);

Прощай все и навсегда (И. С. Тургенев. Дневник лишнего человека);

В Москве и без дела он уже жить не может (С. А. Толстая);

Уехал, но ненадолго;

Уехал, и надолго.

Весьма заманчивой выглядит идея создания исчерпывающего списка сочетаний, допустимых для той или иной синтаксической функции, которая, как утверждает В. З. Санников [1989: 46], реализована им в более ранней работе [Санников 1981]. Вместе с тем широкая вариативность грамматического оформления однородных членов, даже если они выступают в одной функции (свидетельством чего могут служить как примеры из [Санников 1989], так и материал НКРЯ), вызывают сомнения в осуществимости данной задачи в пол-

⁴² Даты, записанные цифрами в соответствии с шаблонами (типа 13.02.2012), выявляются, проверяются и атрибутируются на предшествующих этапах анализа.

ном объеме, не говоря уже о случаях функциональной разнотипности.

Другая хорошо известная проблема — согласование сказуемого с сочиненными подлежащими — применительно к задаче автоматической обработки текста была внимательно рассмотрена В.З. Санниковым [1989] и Л.Л. Иомдиным [1990]. Поскольку нас интересует только анализ текста, но не его синтез, вопрос связан не с выбором нужной формы глагола (*Остался/остались девочка и мальчик*), а с вариативностью в числе сказуемого. В соответствующих синтаксических правилах приходится допускать различные варианты согласования сказуемого с сочиненными подлежащими по числу, роду и лицу, ср.:

Пришел/пришли Миша и Петя;

Почти миллион человек работал/работало/работали в сельском хозяйстве;

*Кто и что мешает (*мешают) тебе работать?;*

*Хорошее и плохое смешалось (*смешались) в памяти;*

То ли чурка, то ли бочка проплывает по реке (А. Твардовский. Василий Теркин);

*Приехала (*приехали) не только Маша, но и Петя;*

Швед, русский колет, рубит, режет (А.С. Пушкин. Полтава);

*Швед и русский поскакали (*поскакал) к реке.*

В случае употребления глагола в единственном числе дополнительная сложность связана с определением того, кореферентны ли однородные подлежащие, ср.:

Прямо у крыльца росла большая сосна и дуб (два дерева);

Выступил известный музыкант и композитор (один и тот же человек).

В предложениях наподобие последнего встает также вопрос о «сфере действия» зависимых членов. Характеризует ли согласованное определение *замечательные* в примере *Здесь работают замечательные писатели и врачи* представителей обеих профессий? К чему в предложении *На пороге показались Коля и Вася с цветами в руках* относится несогласованное определение в постпозиции —

к ближайшему подлежащему или к обоим однородным членам? В записях на формальных языках в подобных случаях используются скобки, позволяющие снимать неоднозначность, но в ЕЯ она неизбежна.

Приходится также принимать во внимание возможность семантического, а не грамматического согласования сказуемого с подлежащим по числу, ср. *Большинство родителей не смогли прийти на собрание; На тропинке оказались отец с сыном; Новый премьер-министр обратилась к собравшимся с речью.*

Еще одна трудность состоит в том, что сочиненные именные группы (так же, как и множественные именные группы) могут интерпретироваться и в объединительном, и разделительном смысле. Так, предложение *Петя и Коля получили грант в размере трех тысяч долларов* может осмысляться в двух значениях: ‘три тысячи всего’ и ‘три тысячи каждому’ [Богуславский 1996: 140–141]. Если предложение содержит несколько сочиненных групп, ситуация еще более запутывается, ср. *Вчера и позавчера Аня и Володя видели Лену и Алешу в кафе и на дискотеке*⁴³. Очевидно, что автоматический анализ, построенный на комбинаторном исчислении возможных вариантов, не способен выявить ту наиболее очевидную интерпретацию, которая сразу приходит на ум человеку.

Более того, вследствие недостаточной эксплицитности и точности языковых выражений логическое исчисление способно порождать ложные варианты. Например, из предложения *Коля и Петя купили в киоске французские, английские, немецкие газеты и журналы* вовсе не следует, что каждый из субъектов купил весь спектр печатной продукции (шесть видов). Предложение содержит утверждения о группах объектов, а не о каждом индивидуально. Однако дальнейшее развитие повествования может касаться любого из данных объектов — Коли, Пети, английских газет, немецких журналов и т. д. Поэтому на следующем этапе анализа (коммуникативно-синтаксическом) мы выделяем все восемь объектов по отдельности (Коля,

⁴³ Схожая проблема возникает и при интерпретации сочинительной связи между придаточными предложениями. Так, предложение *Давайте встретимся с избирателями, которые живут в этом районе и поддерживают «Яблоко»* предполагает конъюнкцию предикатов (*живут и поддерживают*). В отличие от него, предложение *При описании наблюдений следует избегать категоричности выводов, если число наблюдений недостаточно и если в наблюдениях имеются неясности или сомнения в точности наблюдений* допускает и конъюнкцию, и дизъюнкцию пропозиций, выраженных условными придаточными [Тискин 2016].

Петя, французские газеты, английские газеты, немецкие газеты, французские журналы, английские журналы, немецкие журналы), причем каждый вид газет и журналов связывается отношением покупки как с Колей, так и с Петей. Те или иные связи могут быть отменены при наличии опровергающей информации в последующих предложениях.

Что касается второго из перечисленных правил, связанного со сборкой генитивной конструкции, проблема заключается прежде всего в определении хозяина беспредложного генитива: ведь он может управляться как глаголом, способным присоединять соответствующий актанта, так и другим существительным. Встречаются случаи синтаксической омонимии, разрешить которую может лишь учет значений семантических признаков всех «претендентов». Иными словами, единого правила, связанного с обработкой беспредложного генитива, не существует⁴⁴. Эта ситуация, вообще говоря, вполне типична. То, что мы для краткости изложения формулируем в качестве правила сборки конструкции, на деле часто представляет собой совокупность инструкций, касающихся действий анализатора в частных случаях того или иного расположения токенов с приписанными им значениями признаков.

Если все же установлено, что мы имеем дело с генитивной конструкцией, может возникнуть еще один источник синтаксической неоднозначности, обусловленный возможностью двоякого прочте-

⁴⁴ Неоднозначность, связанная с конкуренцией между глаголом и существительным за право управлять именной группой, естественно, не ограничивается формой беспредложного родительного. Так, в предложении *Но нет уже в них того азарта и самоуверенности, и все реже и реже сбрасывают они на наши головы тучи листовок с призывами сдать и бросить надежды на идущего с севера Жукова* (В. Некрасов. В окопах Сталинграда) предложно-падежная группа «на + S_{вин.}» с формальной точки зрения может быть актантом как глагола *бросить*, так и существительного *надежды*. Однако для того чтобы она относилась к глаголу, необходимо, чтобы объектная валентность, выраженная прямым дополнением, была заполнена существительным с предметным значением (*бросить ключи на стол*). С другой стороны, предикатное существительное *надежда* может управлять практически любым существительным в данной позиции. В данном случае информация о сочетаемостных свойствах лексических единиц из Семантико-синтаксического словаря позволяет снять синтаксическую омонимию. Иначе обстоит дело со следующим заголовком информационного сообщения: *Мужчина прыгнул с моста на остров Русский во Владивостоке* (<http://rusnovosti.ru/news/255776/> (дата обращения: 24.03.2018)), — который неоднозначен в принципе.

ния — субъектного и объектного, ср. *встреча друзей, портрет художника* и т. д.

Дальнейшая работа по анализу генитивной конструкции касается ее семантической интерпретации. Как известно, конструкция с родительным падежом способна выражать широчайший спектр значений, в том числе метафорических [Кнорина 1990; Михеев 2000]. Семантический анализ генитивной конструкции в каждом конкретном случае является весьма сложной задачей, сопоставимой разве что с определением значения относительных прилагательных в сочетании с различными существительными.

Некоторых комментариев требует правило присоединения наречия к главному слову. Как известно, наречия могут примыкать к глаголу (включая причастие и деепричастие), прилагательному, другому наречию и существительному, так что, в принципе, может возникать ситуация конкуренции между его потенциальными «хозяевами». Однако, во-первых, сама суть синтаксического отношения примыкания предполагает контактное расположение слов в словосочетании. Во-вторых, для каждого наречия существует предпочтительная (а иногда жестко фиксированная) позиция по отношению к своему синтаксическому хозяину. И если в устной речи нарушение предпочтительной позиции «компенсируется» использованием просодических средств, то в письменном тексте это невозможно, а следовательно, она соблюдается практически всегда. К примеру, в спонтанной речи за счет определенного интонирования и паузации наречие *очень* в предложении *Девочка красивая очень смутилась и покраснела* может быть синтаксически прикреплено к прилагательному *красивая*, но в тексте однозначно воспринимается как относящееся к глаголу *смутилась*.

* * *

Правила третьего ранга имеют своей целью сборку обособленных членов предложения и их прикрепление к главному слову. Это касается одиночных причастий и деепричастий, причастных и деепричастных оборотов, обстоятельственных оборотов, выраженных именем существительным с предлогом, согласованных и несогласованных определений, приложений, пояснительных и уточняющих оборотов. Разработаны определенные правила их прикрепления к тому или иному члену результирующей синтаксической структуры предложения. Так, обособленные причастия, причастные и другие адъективные обороты могут прикрепляться к группе подлежащего

(П) или группе дополнения (Д). Некоторые сложности здесь могут возникнуть с обработкой несогласованных определений, в силу того что они имеют слабую связь с определяемым существительным, а синтаксическая интерпретация слабоуправляемых членов предложения, как будет показано ниже, не имеет универсального решения.

Обособленные деепричастия и деепричастные обороты преобразуются в самостоятельную структуру ПСДО: сказуемому соответствует деепричастие, а подлежащее дублируется из ПСДО главной пропозиции. При этом, правда, неэксплицированной остается обстоятельственная семантика деепричастий, или (в проекции на наш формат синтаксического представления) содержательная связь между ПСДО с личным глаголом и ПСДО с деепричастием. (Впрочем, это часть более общей и многосторонней проблемы о связях между пропозициями в тексте, затрагиваемой в следующей главе.) Что же касается обособленных субстантивных обстоятельственных оборотов (*во избежание падения, по причине отсутствия* и т. д.), они преобразуются в группу обстоятельственных детерминантов (О)⁴⁵.

По-видимому, не существует единого подхода к синтаксическому анализу обособленных приложений. В некоторых контекстах такое приложение содержательно относится ко всему предложению в целом [Камынина 1983: 80–81], и, следовательно, согласно принятому нами подходу, его следует квалифицировать как обстоятельственный детерминант (О). Так, в примере **Человек молчаливый и очень точный, он избегал скороспелых выводов и преждевременных обобщений** обособленный оборот не столько характеризует лицо, сколько обозначает причину того положения дел, которое описывается главной пропозицией. Однако в предложении *Через полчаса явился уездный лекарь, человек небольшого роста, худенький и черноволосый* (И. Тургенев. Уездный лекарь)⁴⁶ структурно схожее приложение фактически служит определением к слову *лекарь*, приписывая данному лицу свойство 'невысокий'. Информация, передаваемая здесь обособленным оборотом, должна быть прикреплена к группе подлежащего (П), но формальные способы дифференциации рассмотренных двух случаев пока не ясны (предположительно начальная позиция оборота в предложении марки-

⁴⁵ Неособобленное одиночное деепричастие, примыкающее справа к глаголу, входит в состав группы сказуемого (С).

⁴⁶ Примеры заимствованы из книги: [Учебно-методический комплекс... 2008: 161].

рует обстоятельственное прочтение, но эта гипотеза нуждается в дополнительной проверке).

Кроме того, правила третьего ранга осуществляют прикрепление придаточных предложений к опорному слову в главном предложении (для сложноподчиненных предложений с нерасчлененной структурой). Наиболее часто встречаются следующие два вида придаточных:

- 1) изъяснительные придаточные предложения с союзом *что*, заполняющие актантную позицию у предикатной лексики, обозначающей речевую и ментальную деятельность, эмоциональные состояния. Их прикрепление к опорному слову осуществляется с опорой на описание соответствующих слов в Семантико-синтаксическом словаре, где указаны возможности реализации объектной валентности не только посредством именных групп, но и сентенциальных актантов, вводимых различными союзами, а также прямой речи в кавычках;
- 2) определительные придаточные предложения, вводимые союзным словом *который*.

Указанные случаи достаточно легко распознаются и обрабатываются. Применительно ко второму типу, правда, следует упомянуть о ситуации, когда определительному придаточному предшествует не одиночное существительное в главном предложении, а генитивная конструкция из двух и более членов. Если существительные имеют один и тот же грамматический род, присоединение придаточного в принципе возможно к любому из них, ср.: *Директор института, который в прошлом году был награжден орденом Трудового Красного Знамени...* Нужно заметить, что данный пример является неоднозначным и для носителя русского языка⁴⁷ — в отличие от таких примеров, как:

Директор института, который расположен в городе...

Директор института, который еще недавно болел...

Однако в рамках конструктивно-синтаксического анализа эти примеры будут также интерпретироваться как неоднозначные,

⁴⁷ О психологических факторах, влияющих на интерпретацию определительной связи в подобных случаях, см. [Федорова, Янович 2004; Драгой 2006; Юдина 2006].

поскольку алгоритм не включает проверку на семантическое согласование слов *директор и болел, институт и болел* и т. д. Соответственно, будут построены различные варианты синтаксической структуры, присоединяющие придаточное определительное к одному и другому существительному. Выбор того из них, который является правильным, возможен лишь на следующем этапе анализа.

Обобщая ситуацию, связанную с автоматической обработкой сложноподчиненных предложений, можно сказать, что она происходит достаточно благополучно, если союз (или союзное слово) способен присоединять придаточное лишь какого-то одного определенного типа. Когда же связующий элемент характеризуется неоднозначностью (ср. *какой, кто, что, когда* и др.), правила сборки усложняются. Так, союзные слова *какой, кто, что* могут присоединять к главному предложению и изъяснительные, и определительные придаточные. Для дифференциации этих случаев придется брать во внимание тип опорного слова (ср. *Я спросил, кто придет.* vs. *Тот, кто придет,...; Я спросил у него, какие книги нужны* vs. *Я собрал все книги, какие нужны*). Если слово нужного типа в главном предложении отсутствует, сборка осуществляется «по дефолту».

Приведем еще один пример. Токен *когда* в расчлененных сложноподчиненных предложениях выступает в качестве союза, присоединяющего придаточное времени, а в нерасчлененных служит для прикрепления определительных или изъяснительных придаточных, ср. *Когда стемнело, все легли спать* vs. *Наступает пора, когда рано светает* vs. *Мне нравится, когда задают много вопросов*. Идентификация этих трех вариантов также базируется на поиске в главном предложении лексических единиц определенных семантических групп. Так, изъяснительное придаточное выявляется по наличию глагола эмоционального состояния *нравиться*. Определительное придаточное удастся опознать благодаря существительному со значением временного промежутка (*пора*). Что же касается придаточных времени, они не имеют специфических показателей и идентифицируются по остаточному принципу. Поэтому если в главном предложении не обнаружен ни один из вышеуказанных маркеров, придаточному приписывается синтаксическое значение 'придаточное времени'.

Существуют также специальные правила третьего ранга, направленные на сборку сложных обстоятельственных детерминантов, чьи составляющие могут быть или не быть разделены запятой, напри-

мер: на Невском пр., в д. 35, в кв. 19; в Москве на Красной площади; в прошлом году весной; в пятницу вечером.

* * *

Схематические шаблоны предложений и ранжированные правила призваны обеспечить адекватный автоматический анализ наиболее частотных, центральных синтаксических структур в русском языке, которые имеют стандартную, регулярную интерпретацию (пусть и не обязательно однозначную). Следует признать, что многие факты языка неизбежно остаются за их рамками. Прежде всего, это разнообразные периферийные феномены, перечислять которые нет ни возможности, ни необходимости в силу их окказиональности и особенности. Но между этими двумя крайностями — выраженным центром и расплывчатой периферией — расположен целый ряд достаточно регулярных структур, которые легко опознаются и выделяются в тексте, но не имеют стандартной семантико-синтаксической интерпретации. Их успешная автоматическая обработка в каждом отдельном случае требует учета разнообразных контекстуальных факторов, причем конкретный набор этих факторов и характер обусловленности нуждается в тщательной проработке. Это работа, требующая больших затрат сил и времени.

Возьмем в качестве иллюстрации феномен так называемого слабого управления [Пешковский 1956], или падежного примыкания (Русская грамматика 1980)⁴⁸. Рассмотрим грамматически тождественные предложно-падежные группы *пить чай с женой / пить чай с вареньем / пить чай с удовольствием* и покажем, что они должны получить различную семантико-синтаксическую интерпретацию.

В выражениях типа *пить чай с женой, гулять с собакой, идет дождь со снегом*, предложно-падежная группа *с + N_{тв.}* выражает идею совместности действия. Такое значение приписывается данной конструкции тогда, когда оба существительных характеризуются однотипными значениями базового семантического признака (в наших примерах это 'человек' и шире — 'живое существо'; 'природная сила'). В этом случае предложно-падежная группа должна быть поднята в ранг однородного подлежащего, ср.: *я и жена пьем чай, я и собака гуляем, идут дождь и снег*. Разумеется, данное преобразование

⁴⁸ При разработке семантической интерпретации предложно-падежных групп, помимо толковых словарей русского языка, использовался Синтаксический словарь Г. А. Золотовой (Золотова 1988).

приводит к некоторому семантическому сдвигу, но с точки зрения информационной структуры текста им можно пренебречь. Схожая ситуация наблюдается в выражениях *пригласить профессора с его учениками, положить в пенал ручки с карандашами*, где рассматриваемая конструкция должна интерпретироваться как однородное дополнение, ср.: *пригласить профессора и его учеников, положить в пенал ручки и карандаши*.

Словосочетание *пить чай с вареньем* по своей сути является неоднозначным: то ли варенье находится в чае, то ли субъект попеременно отхлебывает чай из чашки и зачерпывает из вазочки (банки и т. п.) варенье. Первая интерпретация довольно типична для слов с семантическим признаком «еда, напитки = +», ср. *каша с маслом, суп с белыми грибами, макароны с сыром* и т. д. Обоснованность второго прочтения обусловлена однородностью объектов (см. выше). В ходе конструктивно-синтаксического анализа должны быть порождены обе версии.

Наконец, в выражении *пить чай с удовольствием* предлог сочетается с абстрактным существительным, а глаголу по базовому семантическому признаку приписано значение 'действие', следовательно, данная синтаксическая группа представляет собой обстоятельство образа действия.

Вообще, сочетание предлога *с* с творительным падежом, как известно, принадлежит к числу примеров, при помощи которых А. М. Пешковский [1956: 285] иллюстрировал введенное им понятие слабого управления. Рассмотренные случаи отнюдь не исчерпывают возможные реализации данной конструкции, ср. *бутылка с молоком, девочка с косичками, мыть с мылом, выехать с рассветом* и т. д. Тем не менее их достаточно для того, чтобы показать невозможность выработки единой процедуры анализа.

Слабое управление встречается довольно часто. По-видимому, практически любая предложно-падежная группа может в том или ином контексте реализовывать сугубо грамматическую зависимость, не обусловленную семантикой главного слова. Бывает, что несколько словосочетаний находятся в отношении слабого управления по отношению к одному и тому же глаголу или существительному, ср. название известного детективного романа С. Жапризо «Дама в очках с ружьем в автомобиле».

Отсутствие непосредственной связи с семантикой главного слова затрудняет синтаксический анализ подобных выражений, поскольку их появление в тексте не предсказывается условиями реализации,

описанными в семантико-синтаксическом словаре. В ряде случаев они повышают индекс омонимии предложения за счет неоднозначности синтаксического хозяина либо из-за возможности данного словосочетания выразить как семантическую валентность, так и чисто грамматическую зависимость.

Феномен слабого управления в лингвистике изучен недостаточно, соответствующие случаи весьма многообразны, требуют индивидуального подхода и тонкого анализа. Объем работы здесь поистине огромен, и мы ни в коей мере не претендуем на исчерпывающее описание.