

Метод выделения главных членов предложения в виде предикативных структур, использующий минимальные структурные схемы

Харламов А.А., доктор технических наук

Ермоленко Т.В., кандидат технических наук

Дорохина Г.В., младший научный сотрудник

Гнитько Д.С., магистрант

В статье дан краткий обзор подходов, используемых при синтаксическом анализе предложений естественного языка, приведено обоснование выбора синтаксического представления предложения в виде предикатной структуры. Для формального описания базовой структуры простого предложения в работе используется предикатная конструкция, реализованная на атрибутивном уровне описания своих составляющих, включающая актанта, объединённые с предикатом системой отношений. Выявление предикативно связанных грамматических субъекта и предиката в простом предложении осуществляется с помощью минимальных структурных схем предложений. На основе минимальных структурных схем предложений строятся соответствующие им шаблоны и далее проводится их поиск в предложении.

• синтаксический анализ • семантический анализ • грамматический предикат • грамматический субъект • актанта • валентность предиката • атрибутивный уровень описания • минимальная структурная схема предложения • морфологическая информация • копула.

The paper presents a short overview of approaches, used in syntactic analysis of sentences in the natural language. It also presents an argumentation of the choice of a syntactic representation of a sentence in the form of Subject-Predicate structure. To describe the basic structure of the simple sentence on a formal level the authors use a predicate structure described as a network of its constituents, including actants, being part of the predicate system of relations. Identification of grammatical subject and predicate in the simple sentence is performed using the minimal structural schemes of the sentence. On this basis corresponding templates are built, and then their automatic extraction is carried out.

• prosodic speech features • melodic contour • syntagma • intonational structure • kernel of intonational structure • pitch • microprosody • description of intonational accentual units.

Введение

В последнее время активно используются различного рода интеллектуальные информационные системы, выполняющие обработку текстов на естественном языке (далее ЕЯ). Один из ключевых элементов таких систем — лингвистический процессор. Классиче-



ская структура лингвистического процессора содержит три последовательных блока морфологического, синтаксического и семантического анализа ЕЯ-текста [1].

Морфологический анализ текста на ЕЯ не представляет серьёзных трудностей для программной реализации. Сложность создания механизмов синтаксического и семантического анализа обусловлена в значительной степени отсутствием единой теории языкового общения, охватывающей все аспекты взаимодействия коммуникантов: грамматика ЕЯ принципиально недетерминирована и неоднозначна, синтаксис ЕЯ весьма разнообразен, сложен и произволен. Поскольку полной и строгой формальной модели ни для одного естественного языка пока не создано, при разработке средств общения конечных пользователей используется ограниченный ЕЯ.

Для автоматической обработки трудны такие вполне допустимые в ЕЯ явления, как эллипсис (пропуск обязательных фрагментов предложения в силу возможности их восстановления из предыдущего контекста) и анафора (отношение между словами или словосочетаниями, при котором в смысл одного выражения входит отсылка к другому, ранее упомянутому, языковому выражению). Кроме того, при синтаксическом анализе текста на ЕЯ одна из основных проблем — разрешение неоднозначностей [1, 2]. При разработке синтаксического анализатора существуют два подхода: формально-грамматический и вероятностно-статистический [3].

Методы первого подхода направлены на создание сложных систем правил, которые позволяли бы в каждом конкретном случае принимать решение в пользу той или иной синтаксической структуры. Правила представляются в виде грамматик, задающих синтаксис языка [4, 5]. Хотя такой подход может обеспечить высокую точность анализа, возникают сложности в связи с сильной зависимостью от конкретной грамматики языка. Создание анализатора структурного типа — весьма сложный процесс. Наиболее трудоёмкую часть работы (создание системы правил) выполняет лингвист высокой квалификации.

Главная особенность методов вероятностного типа — отсутствие жёсткой системы синтаксических правил, для создания которой, собственно, и требовалось участие лингвиста. Вместо системы синтаксических правил используется обширный набор примеров предложений, разобранных человеком вручную, для получения статистики встречаемости различных структур в похожем контексте. Этот набор примеров используется для «обучения» статистического распознавателя, опирающегося на известный метод дерева принятия решений [6]. Затраты на разработку вероятностных анализаторов могут быть существенно ниже, чем на создание исчерпывающих структурно-грамматических моделей естественного языка. Однако для функционирования вероятностно-статистических методов необходим представительный банк синтаксических структур, полученный в результате «ручного» синтаксического разбора. Для достижения приемлемой точности анализа их могут потребоваться тысячи. Разновидностью статистических систем синтаксического анализа являются анализаторы, которые используют описание языка в виде моделей управления. Они настроены на работу в заданной предметной области и получены в результате предварительного анализа корпуса текстов этой предметной области. Каждой модели управления приписывается частотность, характеризующая вероятность использования этой модели управления для новых текстов данной области [7].

Таким образом, разработка методов синтаксического анализа ЕЯ-текстов без привязки к конкретному языку и легко адаптируемых под нужды конкретной

предметной области представляет собой важную научную задачу и имеет существенное практическое значение.

Особое и обязательное свойство предложения — предикативность — соотнесённость сообщаемого с действительностью. Предикатная структура простого предложения обуславливается общими принципами воссоздания действительности и не зависит от конкретного языка. Этот вывод имеет далеко идущие последствия как для структурирования речевого материала в общем (общий структурный синтаксис), так и для вопросов автоматической обработки текста (структура базы знаний, формирование лингвистического процессора и т.п.).

Данная статья посвящена решению задачи синтаксического анализа, которая заключается в получении синтаксической структуры входного предложения в виде предикатных структур на основе использования морфологической информации о словоформах, полученной на этапе морфологического анализа.

Предикатная структура как первооснова предложения

Представление о предикатной структуре как первооснове предложения возникло ещё в античный период, когда предложение и суждение строго не разграничивались, их компоненты зачастую отождествлялись. И в предложении, и выражаемой им мысли друг другу противопоставлялись субъект и предикат, поэтому в традиционной грамматике прочно утвердилось представление о двусоставности как важнейшем признаке предложения. Субъектно-предикатная структура предложения играет огромную роль в языке.

В контексте языкознания предикат обозначает то, что высказывается (утверждается или отрицается) о субъекте. Предикат находится в предикативном отношении к субъекту, способном принимать отрицание и различные модальные значения. К понятию предиката предъявляются определённые семантические требования, а именно, предикат — не всякая информация о субъекте, но указание на признак субъекта, его состояние и отношение к другим предметам.

В ряде современных направлений логики понятие «предиката» заменено понятием «пропозициональная функция», аргументы которой представлены актантами (термами) — субъектом и объектами [8].

Грамматический субъект (подлежащее) — ещё одна конститутивная знаковая единица в составе предложения [9, 10]. Его означаемым является, прежде всего, один из семантических актантов с присущей ему ролевой нагрузкой. Наряду с этим, его означаемым часто оказывается логический субъект как представление об исходном предмете мысли. Субъект обеспечивает идентификацию носителя признака. Наложение на функцию одного из нескольких семантических актантов (если их в пропозиции более одного) функции логического субъекта придаёт суждению (и выражающему его предложению) свойство ориентированности и по отношению к предикату, и по отношению к объекту/дополнению (или объектам/дополнениям). Тем самым маркируется выдвигание одного из актантов на роль первого, главенствующего в логическом плане среди равных. Субъект задаёт грамматико-смысловую перспективу предложения.

Формальными признаками грамматического субъекта могут быть его начальная позиция в линейной структуре предложения, а в языках с развитой системой словоизменения — падежные флексии. Так, в языках номинативного строя подлежащее, в основном, представляется именной частью речи в именительном падеже, реже — инфинитивом, который является формальным субъектом. В языках эргативного строя выбор падежа для подлежащего зависит от переходности или непереходности глагола. В языках активного строя для подлежащего при сказуемом со значением действия используется активный падеж, а при глаголах со значением состояния — инактивный падеж.



Грамматический предикат (сказуемое) — вторая конститутивная знаковая единица в составе предложения и может характеризоваться определённым местом в линейной структуре предложения. В его позиции чаще всего выступает глагол [9]. В языках с развитой системой глагольного словоизменения в словоформе глагола выражается набор самых разных граммем, принадлежащих к формоизменительным категориям времени, вида, наклонения, залога, отрицания, вопросительности, а также к согласовательным категориям лица, числа, иногда рода и т.д. Сказуемое может быть также представлено другими предикатными словами (прилагательное, наречие, предикатив, неличные формы глагола), а также существительным (со связкой или без неё). Возможны различные способы усложнения сказуемого. И граница между сложным членом глагольным сказуемым как целостным членом предложения и сочетанием сказуемого с другими компонентами часто устанавливается произвольно.

В содержательно-ориентированных теориях синтаксиса особо подчёркивается, что на сказуемое, выступающее в качестве ядра, вокруг которого организуется ближайшее окружение, или же на предикативное отношение, связывающее сказуемое с подлежащим, ложится функция актуализации предложения в модально-временном плане, отнесения его содержания к описываемой ситуации действительности, утверждения или отрицания существования этой ситуации. И сама связь между сказуемым и предикатом, и отнесённость предложения в целом к действительности (независимо от наличия или отсутствия в нём подлежащего) характеризуются в терминах «предикация» и «предикативность». Только совокупность средств выражения предикации и референции (пространственно-временной локализации) обеспечивают привязку предложения к действительности, его актуализацию [8].

В связи с вышеизложенным, наличие предикативно связанных грамматических субъекта и предиката многие исследователи считают обязательным свойством предложения. Эти члены предложения квалифицируются как главные, поскольку они формируют предикативную основу предложения, его конструктивный минимум. В конструкциях с безобъектными, непереходными (в широком смысле) глаголами позиция дополнения представлена нулём, т.е. отсутствует. Точно так же может отсутствовать и позиция подлежащего, когда предложение развёртывается на основе бессубъектного глагола (*темнеет, морозит*). Субъектная позиция здесь также представлена нулём. В ряде языков появляется нечто вроде формального подлежащего. Бессубъектными следует признать предложения, ядром которых являются событийные имена и имена состояний (*война, пожар, мороз, морозно*).

В рамках данной работы используется модель языка, в которой на синтаксическом уровне предикат — ядерная структура, включающая в свой состав *п* актантов. В общем случае, само ядро — глагольная конструкция, актанты объединяются с ядром системой отношений [10]. Узлами в этой конструкции являются имена (существительное, местоимение, числительное) в их атрибутивной форме. Актанты могут быть представлены или в виде отдельных объектов, или в форме конкретных характеристик предикатора, представленных наречиями (*вчера, сегодня, там, здесь* и т.п.). Изложим описание предикатной структуры предложения более подробно.

Формальное описание базовой структуры простого предложения

Предикатная структура реализуется на объектном уровне, где каждую её составляющую (объект — Obj, субъект — Subj, действие — Pred) человек всегда воспринимает как некоторую целостность, которая всегда реализуется через совокупность своих признаков. Язык имеет средства для представления

этих признаков, с помощью которых разделяются объекты одного класса. Обозначим подобные языковые средства как Attr(Obj). Совокупность средств Obj и Attr(Obj) позволяет задавать полное описание объекта уже на уровне фиксации отдельных признаков.

Иначе говоря, предикатная конструкция, реализованная на атрибутивном уровне описания своих составляющих, является эталонной структурой описания отдельной ситуации окружающей действительности. Эту базовую конструкцию речевой деятельности будем считать простым предложением.

Особенность приведённой конструкции — иерархическая зависимость между лексическими составляющими, поступающими на вход логических схем формирования описания элементов внешнего мира. Здесь явно прослеживаются три уровня формирования описания:

ядро конструкции — имя Obj, Subj или Pred;
атрибутивный уровень — список атрибутов (Attr1(Obj), Attr2(Obj) и т.п.);
уровень меры признака (перечень элементов Attr(Attr)).

Совокупность этих трёх уровней полностью определяет атрибутивный уровень описания объекта или действия и полностью представляет языковую деятельность человека.

Конструкцию, являющуюся атрибутивным уровнем описания объекта или действия, в контексте данной работы будем называть группой существительного или глагола соответственно. Например, в конструкции *очень быстрая ходьба* объект Obj — *ходьба*, атрибут Attr(Obj) — *быстрая*, уровень меры признака Attr(Attr) — *очень*.

Простое предложение — это прежде всего двухсоставная конструкция

Subj — R0 — Pred,

где Subj — активный субъект, который инициирует использование предиката Pred; R0 — отношение «быть субъектом».

Если раскрыть все характеристики предиката (его валентности), то структура простого предложения будет иметь вид:

Subj — R0 — Pred — Ri — Obj_i, i = 1, n,

где Ri — предикативные отношения, n — количество актантов.

Последнее выражение определяет монопредикатную структуру описания отдельной ситуации. Простое предложение — это двусоставная конструкция отображения произвольной ситуации, объединяющая субъект с определённым предикатом, которые синтаксически соотносятся с главными членами предложения. Анализ сложных синтаксических конструкций и текста основывается на возвращении к принципам построения простых предложений [10]. Следовательно, для проведения эффективного синтаксического анализа, в первую очередь, необходимо разработать алгоритм выделения главных членов простого предложения, позволяющий представить их в виде двусоставной конструкции.

Выделение предикативной основы простого предложения

Приведём алгоритм поиска главных членов простого предложения. Для этого введём несколько обозначений, приведённых в таблице 1.

Таблица 1

Формы слов/групп, входящих в предикатные структуры предложений

| Форма слова/группы | Обозначение |
|--|-------------|
| 1. Показатели предикативности | |
| группа спрягаемой формы глагола (не инфинитив) | V(f) |
| спрягаемые формы связки — служебных слов <i>быть, стать, являться, значит</i> и т.д. | Сop(f) |

Таблица 1 (окончание)

| | |
|---|-------------|
| копула (тире, тире + это и т.п.) | Cop |
| группа инфинитива глагола, или связки | Inf |
| группа спрягаемой формы глагола 3-го лица единственного числа | V(sn,3) |
| группа спрягаемой формы глагола 3-го лица множественного числа | V(pl,3) |
| 2. Имена и наречия | |
| группа имени (существительного, личного местоимения, количественного числительного, прилагательного, для которого нет согласованного с ним существительного) в i-том падеже | NI |
| группа предложной формы i-того падежа, способная сочетаться со связкой | Nip |
| группа беспредложной и предложной формы косвенного падежа, способная сочетаться со связкой | N2...p |
| группа именительного и творительного падежа прилагательных и страдательных причастий | Adj1 и Adj5 |
| группа кратких форм и компоративов прилагательных и страдательных причастий | Adj(f) |
| наречия, способные сочетаться со связкой (предикативы) | Adv_pr |

Главное слово в группе будем обозначать так: <обозначение группы>_1.

Следует обратить внимание на возможные варианты групп V(f) и Inf. В случае наличия в предложении нескольких групп Inf (*крайне не хотеть заставить себя прилежно учиться*) без копулы между ними, они объединяются в одну. Тогда Inf_1 — конструкция из нескольких инфинитивов (для словосочетания *крайне не хотеть заставить себя прилежно учиться* Inf_1=*не хотеть заставить учиться*).

Составное глагольное сказуемое (вспомогательный глагол + инфинитив) будем относить к группе V(f). Для получения составного глагольного сказуемого последовательно анализируется группа V(f) и Inf, $V(f)_1 = V(f)_1 + Inf_1$ Так, в предложении *отец начинал сильно беспокоиться* $V(f)_1 = \text{начинал беспокоиться}$.

Введём следующие обозначения:

Subj — слово/группа, являющееся подлежащим.

Pred — слово/группа, являющееся сказуемым.

МИ — морфологическая информация словоформы.

Входные данные: простое предложение в виде

$Pr = ((W1, M1), (W2, M2), \dots, (Wn, Mn))$,

где W_i , — написание i-го слова, входящего в предложение; M_i — МИ этого слова.

Выходные данные: ядро предикатной структуры предложения в виде несимметричных пар понятий $\langle c_i, c_j \rangle$, связанных отношением R_0 (быть субъектом), где главное понятие c_i — Pred; понятие-ассоциант c_j — Subj.

Множество простых предложений русского языка задаётся перечнем минимальных структурных схем предложений (далее МСС), описывающих предикативный минимум предложения [11]. МСС — модель, отвлеченный образец, отражающий способ выражения предикативности.

Идея алгоритма заключается в поиске шаблона, соответствующего одной из МСС. МСС и соответствующие им шаблоны приведены в таблице 2, условные обозначения в шаблонах — в таблице 3. Алгоритм начинает ра-

ботать после того, как сформированы группы (атрибутивный уровень описания объекта/ субъекта и действия).

Таблица 2

Минимальные структурные схемы и шаблоны, им соответствующие

| № п/п | МСС | Шаблон МСС | Примеры предложений |
|-------|--|--------------------------|---|
| 1 | N1 V(f) | K1 | Грачи прилетели |
| 2 | N1 Cop(f) Adj1 N1 Cop(f) Adj5 N1 Cop(f) Adj(f) | K2 | Ночь тихая (тиха) Ночь тише дня |
| | | KNC_L + KCAdj | Ночь была тихая (тихой, тиха) Ночь была тише дня |
| 3 | N1 Cop N1 N1 Cop(f) N1 N1 Cop(f) N5 | K3 | Маша — красавица |
| | | KNC_L + KNC | Он был студент |
| | | KNC_L + K3_6 | Он был студентом |
| 4 | N1 Cop N2...p N1 Cop(f) N2...p N1 Cop(f) Adv_pr | KN1_P + K_P_Nobj | Дом — без лифта |
| | | K_Nom_Obj | Подарок — Васе |
| | | KN_Pred | Глаза навывкате |
| | | KNC_L + KCP + K_P_Nobj | Дом будет без лифта |
| | | KNC_L + KC_Pred | Глаза были навывкате |
| 5 | Inf V(f) | K5 | Курить строго воспрещалось Не мешало б нам встречаться чаще |
| 6 | Inf Cop(f) N5 Inf Cop N1 | KCI_Nom + K3_6 | Дозвониться было проблемой |
| | | K6 | Любить иных — тяжёлый крест |
| 7 | Inf Cop(f) Adj1 Inf Cop(f) Adj5 Inf Cop(f) Adj(f) | KCI_Nom + KCAdj | Промолчать — самое разумное Промолчать было самым разумным Промолчать — разумно |
| 8 | Inf Cop N2...p Inf Cop(f) N2...p Inf Cop(f) Adv_pr | KI_P + K_P_Nobj | Промолчать — не в его правилах |
| | | KI_Pred | Молчать некстати |
| | | KCI_Nom + KCP + K_P_Nobj | Отвечать было в его правилах |
| | | KCI_Nom + KC_Pred | Идти было трудно |
| 9 | Inf Cop Inf Inf Cop(f) Inf | K9 | Отказаться — обидеть хозяина |
| | | KCI_Nom + KCI | Отказаться было обидеть |
| 10 | Cop(pl) N2...pr Cop(pl) Adv_pr | KCP + K_P_Nobj | Дома были в слезах |
| | | KC_Pred | С ним были запросто |
| 11 | Cop(f) N1 N1 | KNC | Будет дождь. Была зима |
| | | _11 | Шепот. Робкое дыхание. Тишина |
| 12 | Cop(sn,3) Adj(f) | K12 | Ночью будет морозно |
| 13 | Cop(pl,3) Adj(f) | K13 | Результатом были довольны. Отказом были обижены |
| 14 | N(2–6)p Cop(sn,3) N2...p Cop(sn,3) Adv_pr | K_P_Nobj | На улице без осадков |
| | | KCP + K_P_Nobj | Завтра будет без осадков |
| | | KC_Pred | Было поздно. |
| 15 | V(sn,3) | _15 | Скрипело, свистало и выло в лесу. Ему нездоровится. У него кипело на сердце |
| 16 | V(pl,3) | _16 | За столом зашумели. Его обидели |
| 17 | Inf | _17 | Не нагнать тебе бешеной тройки. Быть рекам чистыми |



Таблица 3

Условные обозначения в МСС и шаблонах МСС

| Обозначение | Описание |
|-------------------------|---|
| K+индекс | Индекс соответствует номеру МСС, указанному в таблице 2, K означает наличие координационной связи между словами в предложении |
| _индекс | Предложение односоставное: субстантивное или с простым сказуемым |
| Cop | Наличие копулы в явном виде в предложении |
| Pred | Наличие предикатива в предикатной структуре |
| Nobj | Главное слово группы N1 в объектном падеже ($l \neq 1$) |
| Nom | Главное слово группы N1 — номинатив |
| I_Nom | Инфинитив является номинативом |
| V(pl,3) | Форма глагола множественного числа 3-его лица |
| V(sn,3) | Форма глагола единственного числа 3-его лица |
| Обозначение конструкции | Описание конструкции |
| KNC_L | Главное слово группы N1 стоит слева от копулы |
| KCAdj | Копула + зависимое слово, которое является компаративом или краткой либо полной формой прилагательного в именительном или творительном падеже |
| KNC | Копула + зависимое слово, которое является главным словом группы N1 |
| K3_6 | Копула + зависимое слово, которое является главным словом группы N5 (используется в МСС 3 и 6) |
| KN1_P | Главное слово группы N1 управляет предложом |
| K_P_Nobj | Предлог управляет главным словом группы Nobj |
| K_Nom_Obj | Главное слово группы N1 + главное слов группы Nobj |
| KN_Pred | Главное слово группы N1 + предикатив |
| KCP | Копула управляет предложом |
| KC_Pred | Копула + предикатив |
| KCI_Nom | Копула + инфинитив |
| KI_P | Инфинитив управляет предложом |
| KI_Nom_Obj | Инфинитив + группа Nobj |
| KI_Pred | Инфинитив + предикатив |

Пример работы алгоритма

Сочинять музыку — значит поручить *цапфенштетсерскому оркестру* исполнить *хор ангелов* (Т. Манн «Доктор Фаустус»).

Объединяем в одну группу Inf неразделённые Cop инфинитивы «поручить» и «исполнить», в итоге получаем Inf_1=«поручить исполнить».

Копула в явном виде (тире), до неё и после — инфинитив, получаем шаблон:
KCI_Nom + KCI, который соответствует МСС9.

Результат работы алгоритма: <<поручить исполнить», «сочинять»>

Таблица 4

Выходные данные, соответствующие найденным шаблонам МСС

| № п/п | МСС | Шаблон МСС | Выходные данные |
|-------|--|--------------------------|--|
| 1 | N1 V(f) | K1 | <V(f)_1, N1_1> |
| 2 | N1 Cop(f) Adj1 N1 Cop(f) Adj5 N1 Cop(f) Adj(f) | K2 | <Cop+Adj1, N1_1> <Cop+Adj5, N1_1> |
| | | KNC_L + KCAdj | <Cop+Adj(f), N1_1> |
| 3 | N1 Cop N1 N1 Cop(f) N1 N1 Cop(f) N5 | K3 | < Cop+N1_1, N1_1> |
| | | KNC_L + KNC | <Cop+N1_1, N1_1> |
| | | KNC_L + K3_6 | <Cop+N5_1, N1_1> |
| 4 | N1 Cop(f) N2p N1 Cop N2p N1 Cop N5p N1 Cop(f) N5p N1 Cop(f) Adv_pr | KN1_P + K_P_Nobj | <Cop+предлог+Nobj_1, N1_1> |
| | | K_Nom_Obj | < Cop+Nobj_1, N1_1 > |
| | | KN_Pred | < Cop+Pred, N1_1> |
| | | KNC_L + KCP + K_P_Nobj | <Cop+предлог+Nobj, N1_1> |
| | | KNC_L + KC_Pred | <Cop+Pred, N1_1> |
| 5 | Inf V(f) | K5 | <V(f)_1, Inf_1> |
| 6 | Inf Cop(f) N5 Inf Cop(f) N1 | KCl_Nom + K3_6 | <Cop+N5_1, Inf_1> |
| | | K6 | <Cop+N1_1, Inf_1> |
| 7 | Inf Cop(f) Adj1 Inf Cop(f) Adj5 Inf Cop(f) Adj(f) | KCl_Nom + KCAdj | <Cop+Adj_1, Inf_1> |
| 8 | Inf Cop(f) N2p Inf N2p Inf Adv_pr Inf Cop(f) Adv_pr Inf Cop Adv_pr | KI_P + K_P_Nobj | <предлог+Nobj_1, Inf_1> |
| | | KI_Nom_Obj | <предлог+N_1, Inf_1> |
| | | KI_Pred | <Cop+Pred, Inf_1> |
| | | KCl_Nom + KCP + K_P_Nobj | <Cop+предлог+Nobj_1, Inf_1> |
| | | KCl_Nom + KC_Pred | <Cop+Pred, Inf_1> |
| 9 | Inf Cop(f) Inf Inf Cop Inf | KCl_Nom + KCl | <Cop+Inf_1, Inf_1> <Cop+Inf_1, Inf_1> |
| | | | |
| 10 | Cop(pl) N2...pr Cop(pl) Adv_pr | KCP + K_P_Nobj | <Cop(pl)+N2...pr_1, 0> |
| | | KC_Pred | <Cop(pl)+Adv_pr, 0> |
| 11 | Cop(f) N1 N1 | KNC | <Cop +N1_1, 0> |
| | | _11 | <Cop+N1_1,0> |
| 12 | Cop(sn,3) Adj(f) | K12 | <Cop(sn,3)+Adj(f), 0> |
| 13 | Cop(pl,3) Adj(f) | K13 | <Cop(pl,3)+Adj(f), 0> |
| 14 | N2p Cop(sn,3) N2p Cop(sn,3) Adv_pr | K_P_Nobj | <Cop(sn,3)+предлог+ Nobj_1, 0> |
| | | KCP + K_P_Nobj | <Cop(sn,3)+предлог+ Nobj_1, 0> |
| | | KC_Pred | <Cop(sn,3)+Adv_pr, 0> |
| 15 | V(sn,3) | _15 | <V(sn,3)_1, 0> |
| 16 | V(pl,3) | _16 | <V(pl,3)_1, 0> |
| 17 | Inf | _17 | <Inf, 0> |

Использование МСС в качестве формального образца позволяет получить предикативную основу (структурную схему) простого предложения, и в дальнейшем — его предикатную



структуру. Это первый и обязательный шаг для проведения первичного семантического анализа в формировании информационного портрета текста, поскольку смысловая связь между понятиями предложения (объектом/субъектом) в общем случае может быть описана предикатом, актантами которого выступают данные понятия. Установление таких синтактико-семантических связей позволяет сформировать схему ситуации, описываемой во фразе.

Обусловленный валентностью предиката семантико-синтаксический уровень анализа конструкций, не соответствующий узкому собственно формально-синтаксическому подходу, даёт возможность даже из набора неправильных форм (посредством приведения их к начальным формам) с помощью заполнения валентных гнезд определить схему предложения.

Семантико-синтаксический анализ предложения предусматривает создание электронного словаря валентности глаголов. При этом для каждого глагола (около 20 тысяч в русском языке) необходимо указать, какими падежами и с какими предлогами он может управлять, а также в каких семантических ролях (семантических падежах) выступают актанта глагола. Разработкой такого словаря для русского языка авторы планируют заняться в ближайшем будущем.

Литература

1. Волкова И.А. Введение в компьютерную лингвистику. Практические аспекты создания лингвистических процессоров. М.: Издательство ВМиК МГУ, 2006.
2. Ножов И.М. Морфологическая и синтаксическая обработка текста (модели программы). М.: Наука, 2003.
3. Евдокимова И.С. Естественно-языковые системы: курс лекций. Улан-Удэ: Издательство ВСГТУ, 2006.
4. Ахо А., Сети Р., Ульман Дж. Компиляторы: принципы, технологии и инструменты. М.: Вильямс, 2001.
5. Волкова И.А., Руденко Т.В. Формальные грамматики и языки. Элементы теории трансляции. М.: Изд-во МГУ, 1999.
6. Андреев А.М., Берёзкин Д.В., Брик А.В., Кантонистов Ю.А. Вероятностный синтаксический анализатор для информационно-поисковой системы [Электронный ресурс]. http://www.inteltec.ru/publish/articles/textan/1kx5_9.shtml.
7. Волкова И.А., Мальковский М.Г., Одинцев Н.В. Адаптивный Синтаксический анализатор // Диалог 2003: Труды Международного семинара. М., 2003, Т. 1. С. 401–406.
8. Сусов И.П. Введение в языкознание. М.: Восток-Запад, 2006.
9. Загнітко А.П. Теоретична граматики української мови: Синтаксис: Монографія. Донецьк: ДонНУ, 2001.
10. Загнітко А.П. Теоретична граматики української мови. Морфологія. Донецьк: ДонДУ, 1996.
11. Современный русский язык: Учебник для филологических специальностей высших учебных заведений / В.А. Белошапкова, Е.А. Брызгунова, Е.А. Земская и др.; Под ред. Белошапковой. 3-е изд., испр. и доп. М.: Азбуковник, 1997.

Сведения об авторах

***Харламов Александр Александрович* —**

доктор технических наук, старший научный сотрудник Института высшей нервной деятельности и нейрофизиологии РАН. Область научных интересов: нейроинформатика, распознавание речи, анализ текстов, распознавание изображений, семантические представления, искусственные нейронные сети.

***Ермоленко Татьяна Владимировна* —**

кандидат технических наук, научный сотрудник отдела распознавания речевых образов Института проблем искусственного интеллекта МОНМС и НАН Украины. Распознаванием и обработкой речевых сигналов занимается с 2002 года. К области интересов также относится автоматическая обработка ЕЯ-текстов.

***Дорохина Галина Владимировна* —**

младший научный сотрудник Института проблем искусственного интеллекта МОНМС и НАН Украины. Область научных интересов: распознавание образов, автоматический морфологический и синтаксический анализ текстов, ассоциативная память, искусственный интеллект.

***Гнитько Дмитрий Сергеевич* —**

магистрант Института информатики и искусственного интеллекта Донецкого национального технического университета. Область научных интересов: автоматический синтаксический анализ текстов, искусственный интеллект, формально-грамматический метод.