

амонімы, антонімы, паронімы, амографы, амафоны)», «Слоўніка фразеалагізмаў», «Славянскага асоцыятыўнага слоўня: русскага, беларускага, болгарскага, украінскага».

У выніку праведзенага даследавання атрымалася не толькі апісаць сінтагматычныя ўласцівасці атрыбутыўных адзінак ў розных дыскурсіўных атачэннях (у прыватнасці, мастацкім і спартыўным), але і стварыць **алгарытм адбору** лінгвістычнага матэрыялу ў камбінаторны слоўнік прыметніка, які да сёння не складзены ў айчынным мовазнаўстве. Былі вылучаны тры асноўныя этапы, якія вызначаюць паслядоўнасць працы лінгвіста над структурай і ключавымі элементамі слоўнікавага артыкула новага тыпу:

- па-першае, трэба правесці пошук кантэкстных ужыванняў прыметнікаў даследуемай тэматычнай групы;
- па-другое, трэба супаставіць словазлучэнні, выкарыстаныя ў мастацкіх тэкстах, з ілюстрацыйнымі прыкладамі слоўнікавых артыкулаў ужо існуючых лексікаграфічных выданняў;
- і нарэшце, з ліку канструкцый, не зафіксаваных слоўнікамі, неабходна адабраць найбольш частотныя ўжыванні прыметніка з назоўнікам.

Такі алгарытм працы па адборы матэрыялу для камбінаторнага слоўніка дазваляе з усёй паўнатай апісаць спалучальнасць ўласцівасці прыметніка ў межах мініяцюрнага жанру – слоўнікавага артыкула.

Таксама быў праілюстраваны феномен набывання прыметнікам у кантэксте новага спецыфічнага значэння. У абедзвюх мовах быў вылучаны асобны клас прыметнікаў, спецыфічную семантыку якіх акадэмічныя тлумачальныя слоўнікі не фіксуюць увогуле. Колькасць такіх ужыванняў прыметніка: у рускім спартыўным кантэксте – **133** атрыбутыўныя словазлучэнні ( $\approx 4\%$ ): *лицевая линия* (у значэнні ‘лінія, якая абмяжоўвае поле для гульні’), *зрячий удар* (‘удар, які падрыхталі, дакладны, эфектыўны’) і інш.; у беларускім спартыўным кантэксте – **30** словазлучэнняў ( $\approx 2\%$ ): *стыкавы матч* (у значэнні ‘дадатковы матч паміж камандамі, якія занялі другія месцы ў сваіх групах падчас адборачнага турніру на буйное спаборніцтва, што дае права ўдзельнічаць у асноўнай частцы турніру ў выпадку перамогі’), *мірны вынік* (‘нічыіны вынік’) і інш.

#### Літаратура

1. Тлумачальны слоўнік беларускай мовы : У 5 т. / Пад агул. рэд. акад. К.К. Атраховіча (К. Крапівы). – Мінск, 1977–1984. Т. 1–5.
2. *Шчэрбін, В.К.* Спалучальнасць / В.К. Шчэрбін // Беларуская мова : Энцыклапедыя. – Мн.: БелЭн, 1994. – С. 543–544.
3. *Юдина, Н.В.* Лексическая сочетаемость в когнитивном аспекте (на материале конструкции «прилагательное + существительное») : автореф. дис. д-ра фил. наук : 10.02.19 / Н.В. Юдина ; Тамбовский гос. ун-т им. Г.Р. Державина. – Москва, 2006. – 40 с.

©МГЛУ

## АВТОМАТИЧЕСКИЙ АНАЛИЗ АНГЛОЯЗЫЧНЫХ СИНТАКСИЧЕСКИ ОСЛОЖНЕННЫХ СТРУКТУР

**В.В. КОСТЮК, И.И. ЗУБОВА**

The article deals with the problem of automatic generation of English Internet text context based abstracts. The formal model is created with this aim. It consists of the following parts: 1) the computer data base including English accessory words – prepositions, articles and conjunctions; 2) the algorithm. The Internet text sense compression is based on the statistic-distributional method. The computer experimental results don't contain any errors in case the user query includes one or two key words

Ключевые слова: анализ, вложение, многозначность, модель, синтаксически осложненная структура

В современном языкознании существует два основных подхода к изучению текста: от целого текста к его составляющим и от составляющих текста к его единому целому. Большая часть лингвистических исследований выполнена в русле первого направления. Они позволяют определить наиболее типичные для текстов разных стилей и жанров структуры предложений, типы словосочетаний и слов, способы связи предложений и другие факты. Изучение синтаксической организации текста приводит к выводу о том, что в языке существует целый ряд синтаксически осложненных структур, вызывающих трудности при понимании текста, как человеком, так и компьютером. В то же время автоматический синтаксический анализ является важным компонентом систем автоматической обработки текста.

Как известно, целью синтаксического анализа является распознавание синтаксической структуры предложения или отдельных его частей. В случае простого предложения, анализ сводится к определению связей между его элементами. В сложном предложении определяются границы составляющих его частей, связи между ними, а также связи между словами внутри каждой из выделенных частей. Существуют разные способы формального представления синтаксической структуры предложения. Модель представления зависит от того, на основе какой формальной грамматики строится анализ предложения. В процессе анализа устанавливается, удовлетворяет ли цепочка словоформ, из которых



тельные в притяжательном падеже были объединены в одну категорию, так как они выступают в предложении в роли определения или предикатива. В то же время в отдельную категорию был выделен герундий, так как он может выполнять в предложении различные функции, отличные от личных форм глагола. Каждой категории был приписан свой идентификатор. В качестве идентификаторов были выбраны однозначные двух- и трёхбуквенные сочетания, представляющие собой сокращенные названия различных частей речи в английском языке или их форм. Все лексические единицы совместно с их идентификаторами составили лингвистическую базу данных, которая использовалась для создания формальной модели (алгоритма) автоматического анализа англоязычных синтаксически осложненных структур.

Рассмотрим работу алгоритма на следующем примере. Предположим, в компьютер вводится многозначное предложение *They are hunting dogs*. Система обращается к лингвистической базе данных, и на ее основе каждому слову введенного предложения присваивается соответствующий идентификатор, в результате чего предложение принимает следующий вид: *They\_NN are\_VB hunting\_VBG dogs\_NN*. Поскольку предложение простое, компьютер, опираясь на идентификаторы, анализирует его слева направо с целью определения группы подлежащего и группы сказуемого. В данном случае в роли подлежащего выступает *They\_NN*, которое на составляющие не разбивается. В группу сказуемого входят все слова справа от подлежащего: *are\_VB hunting\_VBG dogs\_NN*. Далее компьютер определяет, что идентификатор *VBG* в данном контексте может выступать в различных синтаксических функциях: либо как часть сказуемого в *Present Continuous*, либо как определение стоящего справа существительного. Соответственно, группа сказуемого может быть разбита на составляющие ее элементы двумя разными способами, которые отражают синтаксическую многозначность данного контекста. Учитывая этот факт, в результате работы системы на экран будут выведены два различных синтаксических дерева, которые представляют собой возможные варианты синтаксического анализа введенного предложения (см. рисунок 1). В представлении синтаксических деревьев используются следующие условные обозначения: S (предложение); NP (группа существительного); VP (группа глагола); PP (предложная группа); AP (группа прилагательного).

Если для анализа было предложено временно многозначное предложение или предложение с элементами вложения, результатом работы программы будет одно синтаксическое дерево. В случае таких многозначных предложений как, например, *The shooting of the hunters was terrible*, когда модель его составляющих не отражает в полной мере многозначность структуры, в работу включается дополнительный модуль трансформационного анализа. Он определяет возможные ядерные структуры данного предложения, и, в дополнение к дереву составляющих, выводит эти структуры на экран в виде следующей схемы: Субъект – Действие – Объект.

Представленный выше алгоритм был реализован в виде программного продукта, написанного на языке программирования *Ruby*. Программный продукт был разработан для использования в интерфейсе командной строки большинства существующих операционных систем. Ход работы программы в командной строке среды *Cygwin* приведен на рисунке 2.

В ходе работы программы структура составляющих предложения сначала представляется в виде скобочной записи. Затем скобочная запись трансформируется в графическое синтаксическое дерево с использованием *RSyntaxTree* – свободно распространяемой библиотеки для *Ruby*. В конце своей работы программа сохраняет все варианты синтаксических деревьев в виде файлов изображений.

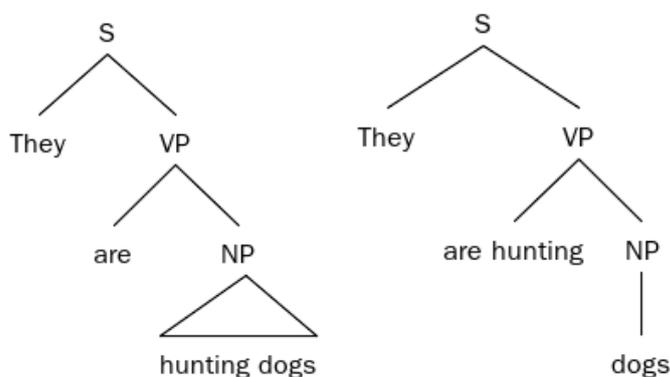
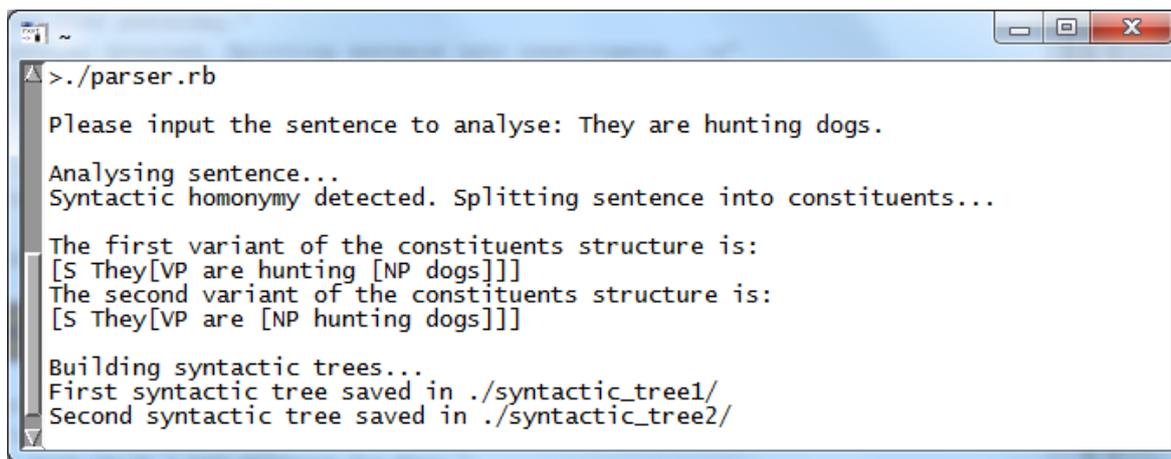


Рис. 1 – Синтаксические деревья, полученные в результате анализа многозначного предложения



```
>./parser.rb
Please input the sentence to analyse: They are hunting dogs.
Analysing sentence...
Syntactic homonymy detected. Splitting sentence into constituents...

The first variant of the constituents structure is:
[S They[VP are hunting [NP dogs]]]
The second variant of the constituents structure is:
[S They[VP are [NP hunting dogs]]]

Building syntactic trees...
First syntactic tree saved in ./syntactic_tree1/
Second syntactic tree saved in ./syntactic_tree2/
```

Рис. 2 – Рабочее окно системы автоматического синтаксического анализа

Синтаксический анализ англоязычных осложненных структур был проведен компьютером без ошибок. Успешные результаты эксперимента были достигнуты, в первую очередь, благодаря сравнительно небольшому объему исследованного материала. При использовании системы на большом массиве текстов можно предположить возникновение у компьютера некоторых затруднений. Например, такое явление, как инверсия, потребует доработки автоматического анализатора, поскольку он ориентирован только на обработку отдельных типов повествовательных предложений. Кроме того, все слова в предложениях, подаваемых на вход системы, были предварительно помечены указывающими на возможные функции слова в предложении идентификаторами вручную. Если все уровни синтаксического анализа будут проводиться автоматически, то ошибочная разметка слов компьютером на предыдущем уровне будет негативно отражаться и на результатах автоматического синтаксического анализа. Необходимо также отметить проблему выбора одной из нескольких синтаксических структур многозначного предложения. При использовании синтаксического анализатора в качестве отдельного модуля систем более высокого уровня, например, системы извлечения смысла из предложения, компьютер должен выбрать один вариант представления многозначного предложения, который будет подвергнут дальнейшему анализу. В качестве возможных решений этой непростой проблемы можно предложить: 1) выбор правильной структуры пользователем; 2) анализ более широкого контекста (рядом стоящих предложений или целого абзаца, если предложение является частью связного текста); 3) выбор одной из структур на основе статистических данных.

#### Литература

1. Прикладное языкознание: учебник / Л.В. Бондарко [и др.]; под общ. ред. А.С. Герда. – СПб.: Изд-во С.–Петербург. ун-та, 1996. – 528 с.
2. Васильев, В.В. Компьютерное понимание текста / В.В. Васильев. – М.: Прометей, 1988. – 258 с.

©МГЛУ

## СЕМАНТИЧЕСКИЕ ОСОБЕННОСТИ ЯЗЫКОВОГО ВЫРАЖЕНИЯ АНГЛИЙСКИХ ПРЕДРАССУДКОВ

*М.В. КОЧУРА, М.Г. БОГОВА*

The article tackles the semantics of English prejudices. The expressed meanings are classified into groups according to the fields of extralinguistic activity their referents belong to. These meanings are of interest as realisations of cultural dominants of the English speaking community. Their hierarchical arrangement is suggested based on criteria of recurrence, importance and ability to form archetypical oppositions

Ключевые слова: предрассудки, приметы, поверья, культурные доминанты

Воздействие культуры на язык проявляется в своеобразии самого процесса общения в разных культурах, что откладывается в лексическом, грамматическом ярусах языка, а также в нормативно-стилистическом укладе языка и обусловлено тем фактом, что объекты в совокупности их свойств и отношений, «не теряя своей конкретности, становятся знаками других объектов и элементов символических классификаций», позволяя посредством лексических единиц, представляющих «простейшую ориентацию человека в пространстве и восприятие контрастных ощущений», перейти к «космологическому осмыслению» [1, с. 59]. Исследование английских выражений, в которых отражены приметы и поверья, проливают свет на представления о мире этого культурно-языкового сообщества. Между тем в простонародье суеверия и предрассудки считаются преданиями предков, нацеленных на предупреждение несчастья или неудач, а также на прогнозирование будущих событий для планирования действий.