

УДК 004.89:004.93

А. В. Ниценко, В. Ю. Шелепов, С. А. Большакова

Государственное учреждение «Институт проблем искусственного интеллекта», г. Донецк
83048, г. Донецк, ул. Артема, 118-б

ОБ АВТОМАТИЧЕСКОМ СНЯТИИ ОМОНИМИИ ПРЕДИКАТИВНЫХ СЛОВСОЧЕТАНИЙ. РЕЗУЛЬТАТЫ РАБОТЫ С НАЦИОНАЛЬНЫМ КОРПУСОМ РУССКОГО ЯЗЫКА

A. V. Nicenko, V. Ju. Shelepov, S. A. Bolshakova

Public institution «Institute of Problems of Artificial intelligence», c. Donetsk
83048, Donetsk, str. Artema, 118-b.

ON AUTOMATIC DISAMBIGUATION OF PREDICATIVE WORD COMBINATIONS. RESULTS OF WORK WITH THE NATIONAL CORPUS OF THE RUSSIAN LANGUAGE

А. В. Ниценко, В. Ю. Шелепов, С. А. Большакова

Державна установа «Інститут проблем штучного інтелекту», м. Донецьк
83048, м. Донецьк, вул. Артема, 118-б.

ПРО АВТОМАТИЧНЕ ЗНЯТТЯ ОМОНИМІЇ ПРЕДИКАТИВНИХ СЛОВСПОЛУЧЕНЬ. РЕЗУЛЬТАТИ РОБОТИ З НАЦІОНАЛЬНИМ КОРПУСОМ РОСІЙСЬКОЇ МОВИ

В статье предложены методы автоматического снятия омонимии словосочетаний, которые могут выступать в роли предикативов. Результаты получены и проверены на материале Национального корпуса русского языка. Они реализованы в экспериментальном программном обеспечении для снятия омонимии.

Ключевые слова: автоматический анализ текста, методы снятия омонимии, предикативное словосочетание, предложная группа, национальный корпус.

The article proposes methods for automatic disambiguation for word combinations that can act as predicatives. The results were obtained and verified on the material of the National Corpus of the Russian Language. They are implemented in experimental disambiguation software.

Key words: automatic text analysis, disambiguation methods, predicative word combination, prepositional group, national corpus.

У статті запропоновано методи автоматичного зняття омонімії словосполучень, які можуть виступати в ролі предикативу. Результати отримано та перевірено на матеріалі Національного корпусу російської мови. Вони реалізовані в експериментальному програмному забезпеченні для зняття омонімії.

Ключові слова: автоматичний аналіз тексту, методи зняття омонімії, предикативне словосполучення, прийменникова група, національний корпус.

Введение

Из недавних работ об автоматическом снятии омонимии отметим работы [1-13].

В статье [14] авторами предложен механизм снятия омонимии предикатива, выраженного словосочетанием. Как предикатив словосочетание представлено в словаре [11], с которым мы работаем, одной словарной единицей. То же имеет место, когда оно выступает в функции наречия, междометия, частицы. Так его и должна воспринимать программа. Однако возможны ситуации, когда на самом деле мы имеем дело с набором отдельных слов, например с предложной группой. Тогда программа должна разделять его на слова, помещаемые в отдельных строках.

Одним из ключевых моментов указанного механизма является использование описанного в [14] файла *Предл гр.txt*. Там же отмечено, что этот файл может и должен модифицироваться и пополняться за счет работы с текстовыми корпусами, в частности, с Национальным корпусом русского языка [15]. Ввиду того, что указанных словосочетаний сравнительно немного, в настоящее время эта работа близка к завершению. Покажем ее результаты на примере ряда словосочетаний. При этом мы не будем приводить формальное содержание соответствующих разделов файла *Предл гр.txt*, а опишем их в словесной форме.

1 Правила снятия омонимии для словосочетаний «в состоянии» / «не в состоянии»

Словосочетания «в состоянии» и «не в состоянии» могут быть интерпретированы либо как предикатив, либо как группа слов, содержащая существительное в предложном падеже и предлог (предложная группа). Для определения правильного варианта анализируется отрезок текста между ближайшими знаками препинания, содержащий словосочетание. При этом используются следующие правила.

1. Если за словосочетанием следует несклоняемое слово, то словосочетание является предложной группой.

Пример: **В состоянии** шоссе мы не заметили ничего необычного.

Наша программа, используя словарь [11], создает файл *res.txt*:

```
(в | предл пр !
в | предл вин)
состоянии | сущ неод ед ср пр
шоссе | сущ неод ед ср
мы | мест сущ мн им
не | част
заметили | гл сов пер/не прош мн
(ничего | нар опред спос
ничего | част
ничего | предик
ничего | мест сущ род !)
(необычного | прл ед муж род !
необычного | прл ед муж вин одуш
необычного | прл ед ср род)
```

Здесь программа приводит либо очередное слово предложения с разметкой из словаря [11], либо соответствующую группу омонимов, которая заключена в скобки, отмечая выбираемый ею омоним восклицательным знаком.

Ниже, мы с целью экономии места будем вместо групп омонимов оставлять лишь выбранные строки, оформляя результат в виде таблицы. В ее первом столбце – анализируемое предложение, во втором – разметка, создаваемая программой.

2. Если после словосочетания есть существительное в родительном падеже, то оно – предложная группа. Перед существительным могут быть одно или несколько прилагательных, причастий или наречий, союзы «и», «или», «как бы», частица «не».

Пример:

| | |
|-------------------------|---|
| <i>эти</i> | <i>мест прил мн вин неод</i> |
| <i>массы</i> | <i>сущ неод мн вин</i> |
| , | |
| <i>давно</i> | <i>нар обст врем</i> |
| <i>находившиеся</i> | <i>прч сов непер воз прош мн вин неод</i> |
| в | <i>предл пр</i> |
| <i>состоянии</i> | <i>сущ неод ед ср пр</i> |
| <i>брожения</i> | <i>сущ неод ед ср род</i> |
| , | |
| <i>можно</i> | <i>предик</i> |
| <i>было</i> | <i>гл 2вид непер прош ед ср</i> |
| <i>успокоить</i> | <i>гл сов перех инф</i> |
| <i>только</i> | <i>част</i> |
| <i>ощутительными</i> | <i>прил мн тв</i> |
| <i>для</i> | <i>предл род</i> |
| <i>всех</i> | <i>мест сущ мн род</i> |
| <i>мерами</i> | <i>сущ неод мн тв</i> |
| <i>справедливости</i> | <i>сущ неод ед жен род</i> |
| <i>и</i> | <i>союз</i> |
| <i>общей</i> | <i>прил ед жен род</i> |
| <i>пользы</i> | <i>сущ неод ед жен род</i> |

3. Если после словосочетания есть прилагательное или причастие в предложном падеже (сразу за ним, либо после наречия, либо после словоформы слова «быть»), то оно – предложная группа.

Пример:

| | |
|-------------------------|--------------------------------------|
| <i>небольшая</i> | <i>прил ед жен им</i> |
| <i>радость</i> | <i>сущ неод ед жен им</i> |
| <i>видеть</i> | <i>гл несов пер/не инф</i> |
| в | <i>предл пр</i> |
| <i>состоянии</i> | <i>сущ неод ед ср пр</i> |
| <i>болезненном</i> | <i>прил ед ср пр</i> |
| <i>человека</i> | <i>сущ одуш ед муж вин</i> |
| , | |
| <i>которого</i> | <i>союз сл ед муж вин одуш</i> |
| <i>я</i> | <i>мест сущ ед им</i> |
| <i>знал</i> | <i>гл несов пер/не прош ед муж</i> |
| в | <i>предл пр</i> |
| <i>положении</i> | <i>сущ неод ед ср пр</i> |
| <i>цветущем</i> | <i>прч несов непер наст ед ср пр</i> |

4. Если предыдущее правило относится к причастию, то между ним и сочетанием может стоять запятая.

5. Если сразу за словосочетанием следует союз «и»/«или» и существительное, которое может употребляться в предложном падеже, то выбирается предложная группа.

Пример:

| | |
|----------------------|---------------------------------|
| <i>ни</i> | <i>част</i> |
| <i>в</i> | <i>предл пр</i> |
| <i>величине</i> | <i>сущ неод ед жен пр</i> |
| <i>листа</i> | <i>сущ неод ед муж род</i> |
| , | |
| <i>ни</i> | <i>част</i> |
| <i>в виде</i> | <i>предл род</i> |
| <i>кожи</i> | <i>сущ неод ед жен род</i> |
| , | |
| <i>ни</i> | <i>част</i> |
| <i>в</i> | <i>предл пр</i> |
| <i>нежности</i> | <i>сущ неод ед жен пр</i> |
| <i>дерева</i> | <i>сущ неод ед ср род</i> |
| , | |
| <i>ни</i> | <i>част</i> |
| <i>в</i> | <i>предл пр</i> |
| <i>состоянии</i> | <i>сущ неод ед ср пр</i> |
| <i>и</i> | <i>союз</i> |
| <i>расположении</i> | <i>сущ неод ед ср пр</i> |
| <i>сучьев</i> | <i>сущ неод мн род</i> |
| <i>не</i> | <i>част</i> |
| <i>находил</i> | <i>гл сов перех прош ед муж</i> |
| <i>я</i> | <i>мест суц ед им</i> |
| <i>ничего</i> | <i>мест суц род</i> |
| <i>сомнительного</i> | <i>прл ед муж род</i> |

6. Если сразу за словосочетанием следует слово, состоящее только из латинских букв, то выбирается предложная группа. Пример:

| | |
|---------------------|---|
| <i>аскет</i> | <i>суц одуш ед муж им</i> |
| , | |
| <i>находящийся</i> | <i>прч несов непер воз наст ед муж им</i> |
| <i>в</i> | <i>предл пр</i> |
| <i>состоянии</i> | <i>сущ неод ед ср пр</i> |
| <i>indhi</i> | <i>нет в словаре</i> |
| , | |
| <i>поднимается</i> | <i>гл несов непер воз наст ед 3-е</i> |
| <i>на</i> | <i>предл вин</i> |
| <i>воздух</i> | <i>суц неод ед муж вин</i> |
| <i>с</i> | <i>предл тв</i> |
| <i>такою</i> | <i>мест прил ед жен тв</i> |
| <i>же</i> | <i>част</i> |
| <i>простою</i> | <i>суц неод ед жен тв</i> |
| <i>и</i> | <i>союз</i> |
| <i>с</i> | <i>предл тв</i> |
| <i>таким</i> | <i>мест прил ед ср тв</i> |
| <i>же</i> | <i>част</i> |
| <i>мальм</i> | <i>прл ед ср тв</i> |
| <i>усилием</i> | <i>суц неод ед ср тв</i> |
| <i>воли</i> | <i>суц неод ед жен род</i> |
| , | |
| <i>как</i> | <i>союз</i> |
| <i>обыкновенный</i> | <i>прл ед муж им</i> |
| <i>человек</i> | <i>суц одуш ед муж им</i> |
| <i>прыгает</i> | <i>гл несов непер наст ед 3-е</i> |

7. Если рассматриваемый отрезок текста содержит глагол в начальной форме, то словосочетание – предикатив. Пример:

| | |
|---------------------------|-----------------------------------|
| <i>тем не менее</i> | <i>част</i> |
| <i>одна</i> | <i>мест прил ед жен им</i> |
| <i>она</i> | <i>мест сущ ед жен им</i> |
| <i>была</i> | <i>гл 2 вид непер прош ед жен</i> |
| <i>бы</i> | <i>част</i> |
| <i>в состоянии</i> | <i>предик</i> |
| <i>разрешить</i> | <i>гл сов пер/не инф</i> |
| <i>весьма</i> | <i>нар опред степ</i> |
| <i>многие</i> | <i>прл мн вин неод</i> |
| <i>важные</i> | <i>прл мн вин неод</i> |
| <i>вопросы</i> | <i>сущ неод мн вин</i> |

8. В остальных случаях выбирается предикатив.

Результаты получены и проверены на материале Национального корпуса русского языка [15] (ниже «корпус»), который содержит 6451 предложение со словосочетанием «в состоянии» и 4601 предложение со словосочетанием «не в состоянии».

2 Правила снятия омонимии для словосочетания «без интереса»

Словосочетание «без интереса» может быть как предикативом, так и предложной группой.

При автоматическом снятии омонимии программа использует следующие правила для определения правильного варианта:

1. Если на анализируемом отрезке текста между двумя соседними знаками препинания есть существительное или местоимение, существительное в дательном падеже, то словосочетание является предикативом. Пример:

| | |
|----------------------------|----------------------------|
| <i>ты</i> | <i>мест сущ ед им</i> |
| <i>фермы</i> | <i>сущ неод мн вин</i> |
| <i>хоть</i> | <i>част</i> |
| <i>языком</i> | <i>сущ ед муж тв</i> |
| <i>вылижи</i> | <i>гл сов перех пов ед</i> |
| - | |
| <i>ему</i> | <i>мест сущ ед муж дат</i> |
| <i>без интереса</i> | <i>предик</i> |

2. Исключениями являются случаи, когда такое существительное является частью предложной группы с предлогами «к» или «по». В этих случаях словосочетание «без интереса» интерпретируется как предложная группа. Пример:

| | |
|------------------------|-----------------------------------|
| <i>она</i> | <i>мест сущ ед жен им</i> |
| <i>встревоженно</i> | <i>нар опред кач</i> |
| <i>отвела</i> | <i>гл сов перех прош ед жен</i> |
| <i>глаза</i> | <i>сущ неод мн вин</i> |
| , | |
| <i>без</i> | <i>предл род</i> |
| <i>интереса</i> | <i>сущ неод ед муж род</i> |
| <i>скользнула</i> | <i>гл сов непер прош ед жен</i> |
| <i>ими</i> | <i>мест сущ мн тв</i> |
| <i>по</i> | <i>предл дат</i> |
| <i>проходящим</i> | <i>сущ одуш мн дат</i> |
| <i>и</i> | <i>союз</i> |
| <i>начала</i> | <i>гл сов пер/не прош ед жен</i> |
| <i>прощаться</i> | <i>гл несов непер воз инф</i> |

3. Во всех остальных случаях словосочетание «без интереса» – предложная группа. Результаты получены и проверены на материале Корпуса, который содержит 114 предложений со словосочетанием «без интереса».

3 Правила снятия омонимии для словосочетаний «в диковинку» и «в новинку»

Словосочетания «в диковинку» и «в новинку» являются предложными группами в случае, если отрезок текста содержит один из глаголов: «вглядываться», «вглядеться», «вкладывать», «вложить», «вложиться», «всматриваться», «всмотреться». Во всех остальных случаях эти словосочетания являются предикативами.

Примеры:

| | |
|--------------------|------------------------------------|
| <i>это</i> | <i>мест суц ед ср им</i> |
| <i>тебе</i> | <i>мест суц ед дат</i> |
| в диковинку | предик |
| , | |
| <i>потому что</i> | <i>союз</i> |
| <i>ты</i> | <i>мест суц ед им</i> |
| <i>простой</i> | <i>прл ед муж им</i> |
| <i>человек</i> | <i>суц одуш ед муж им</i> |
| , | |
| <i>никогда</i> | <i>мест нар</i> |
| <i>не</i> | <i>част</i> |
| <i>видел</i> | <i>гл несов пер/не прош ед муж</i> |
| <i>порядочных</i> | <i>прл мн вин одуш</i> |
| <i>людей</i> | <i>суц одуш мн вин</i> |

Неудивительно, что он решил вложиться в новинку эту.

| | |
|----------------------|---------------------------------|
| <i>неудивительно</i> | <i>предик</i> |
| , | |
| <i>что</i> | <i>союз</i> |
| <i>он</i> | <i>мест суц ед муж им</i> |
| <i>решил</i> | <i>гл сов перех прош ед муж</i> |
| <i>вложиться</i> | <i>гл сов непер воз инф</i> |
| в | предл вин |
| новинку | суц неод ед жен вин |
| <i>эту</i> | <i>мест прил ед жен вин</i> |

Результаты получены и проверены на материале Корпуса, который содержит 146 предложений со словосочетанием «в новинку» и 220 предложений со словосочетанием «в диковинку».

4 Правила снятия омонимии для словосочетания «в сборе»

Словосочетание «в сборе» является предложной группой в случае:

– если за ним следует предлог «с» с существительным в творительном падеже.

Пример:

| | |
|-------------------|---------------------------------------|
| <i>кронштейны</i> | <i>суц неод мн им</i> |
| <i>меняются</i> | <i>гл несов непер воз наст мн 3-е</i> |
| в | предл пр |
| сборе | суц неод ед муж пр |
| <i>с</i> | <i>предл тв</i> |
| <i>рычагами</i> | <i>суц неод мн тв</i> |

– если за ним следует существительное в родительном падеже. Пример:

| | |
|----------------------|------------------------------------|
| <i>она</i> | <i>мест суц ед жен им</i> |
| <i>принимала</i> | <i>гл несов пер/не прош ед жен</i> |
| <i>энергичное</i> | <i>прл ед ср вин</i> |
| <i>участие</i> | <i>суц неод ед ср вин</i> |
| в | предл пр |
| <i>сборе</i> | <i>суц неод ед муж пр</i> |
| <i>пожертвований</i> | <i>суц неод мн род</i> |
| <i>на</i> | <i>предл вин</i> |
| <i>построение</i> | <i>суц неод ед ср вин</i> |
| <i>царицынского</i> | <i>нет в словаре</i> |
| <i>монастыря</i> | <i>суц неод ед муж род</i> |

– если отрезок текста содержит одно из слов: *свобода, заключаться, помеха, мешать, помешать, помощь, помогать, помочь, соотношение, состоять, участие, неучастие, участвовать*. Пример:

| | |
|------------------|----------------------------------|
| <i>все</i> | <i>мест прил ед ср им</i> |
| <i>ваше</i> | <i>мест прил ед ср им</i> |
| <i>дело</i> | <i>суц неод ед ср им</i> |
| <i>будет</i> | <i>гл 2 вид непер буд ед 3-е</i> |
| <i>состоять</i> | <i>гл несов непер инф</i> |
| в | предл пр |
| <i>сборе</i> | <i>суц неод ед муж пр</i> |
| <i>и</i> | <i>союз</i> |
| <i>пересылке</i> | <i>суц неод ед жен пр</i> |
| <i>ко</i> | <i>предл дат</i> |
| <i>мне</i> | <i>мест суц ед дат</i> |
| <i>оброка</i> | <i>суц неод ед муж род</i> |

В остальных случаях словосочетание «в сборе» является предикативом.

Результаты получены и проверены на материале Корпуса, который содержит 1 783 предложения со словосочетанием «в сборе».

5 Правила снятия омонимии для словосочетания «без ума»

Словосочетание «без ума» может быть интерпретировано либо как предикатив, либо как наречие, либо как предложная группа.

1. Словосочетание является наречием, если отрезок текста содержит один из следующих глаголов: *бежать, влюбиться, влюбляться, врезаться, любить, полюбить, предаваться, предаться*

Пример: *Она первая заставила его понять, что можно без ума влюбиться в женщину бесспорно некрасивую.*

| | |
|-------------------|---------------------------------|
| <i>она</i> | <i>мест суц ед жен им</i> |
| <i>первая</i> | <i>прл ед жен им</i> |
| <i>заставила</i> | <i>гл сов перех прош ед жен</i> |
| <i>его</i> | <i>мест суц ед муж вин</i> |
| <i>понять</i> | <i>гл сов пер/не инф</i> |
| , | |
| <i>что</i> | <i>союз</i> |
| <i>можно</i> | <i>предик</i> |
| без ума | нар опред спос |
| <i>влюбиться</i> | <i>гл сов непер воз инф</i> |
| в | предл вин |
| <i>женщину</i> | <i>суц одуш ед жен вин</i> |
| <i>бесспорно</i> | <i>нар опред кач</i> |
| <i>некрасивую</i> | <i>прл ед жен вин</i> |

2. Словосочетание является предикативом если анализируемый отрезок текста содержит предлог «от» с существительным в родительном падеже.

Пример: *Рядом с ним сидела старушка, видать, без ума от сына.*

| | |
|-----------------------|------------------------------------|
| <i>рядом с</i> | <i>предл тв</i> |
| <i>ним</i> | <i>мест суц ед муж тв</i> |
| <i>сидела</i> | <i>гл несом пер/не прош ед жен</i> |
| <i>старушка</i> | <i>суц одуш ед жен им</i> |
| , | |
| <i>видать</i> | <i>ввод</i> |
| , | |
| <i>без ума</i> | <i>предик</i> |
| <i>от</i> | <i>предл род</i> |
| <i>сына</i> | <i>суц одуш ед муж род</i> |

Исключение: в случае предлога «от» со словами «злости», «пьянства», «ревности», «хмеля» имеем предложную группу.

Результаты получены и проверены на материале Корпуса, который содержит 447 предложений со словосочетанием «без ума».

3. В остальных случаях словосочетание «без ума» является предложной группой.

6 Правила снятия омонимии для словосочетания «в обтяжку»

Словосочетание «в обтяжку» может быть интерпретировано либо как предикатив, либо как наречие, либо как предложная группа.

1. Словосочетание является предложной группой, если анализируемый отрезок текста содержит один из следующих глаголов: *вкладывать, вложить, вложиться, вязаться, втянуться*. Пример:

| | |
|-----------------------|-----------------------------------|
| <i>я</i> | <i>мест суц ед им</i> |
| <i>вложил</i> | <i>гл сов перех прош ед муж</i> |
| <i>остаток</i> | <i>суц неод ед муж вин</i> |
| <i>средств</i> | <i>суц неод мн род</i> |
| <i>в</i> | <i>предл вин</i> |
| <i>обтяжку</i> | <i>суц неод ед жен вин</i> |
| <i>мягкой</i> | <i>прл ед жен род</i> |
| <i>мебели</i> | <i>суц неод ед жен род</i> |

2. Словосочетание является предикативом если в тексте существительное стоит к нему ближе, чем глагол или причастие. Пример:

| | |
|-------------------------|---|
| <i>первым</i> | <i>суц неод ед ср тв</i> |
| <i>приехал</i> | <i>гл сов непер прош ед муж</i> |
| <i>на</i> | <i>предл вин</i> |
| <i>вокзал</i> | <i>суц неод ед муж вин</i> |
| <i>шацкий</i> | <i>нет в словаре</i> |
| , | |
| <i>одетый</i> | <i>прч сов перех страд прош ед муж им</i> |
| <i>в</i> | <i>предл вин</i> |
| <i>полосатый</i> | <i>прл ед муж вин неод</i> |
| <i>костюм</i> | <i>суц неод ед муж вин</i> |
| <i>в обтяжку</i> | <i>предик</i> |
| , | |
| <i>долженствовавший</i> | <i>прч несом непер прош ед муж вин неод</i> |
| <i>изображать</i> | <i>гл несом перех инф</i> |
| <i>англичанина</i> | <i>суц одуш ед муж вин</i> |

3. Словосочетание является наречием, если в тексте глагол или причастие стоит к нему ближе, чем существительное. Кроме того, если в тексте есть глагол «сидеть», то словосочетание «в обтяжку» всегда является наречием. Пример:

| | |
|------------------|-----------------------------------|
| <i>платье</i> | <i>сущ неод ед ср им</i> |
| <i>сидело</i> | <i>гл несов пер/не прош ед ср</i> |
| <i>на</i> | <i>предл пр</i> |
| <i>ней</i> | <i>мест суц ед жен пр</i> |
| <i>в обтяжку</i> | <i>нар опред спос</i> |

Результаты получены и проверены на материале Корпуса, который содержит 161 предложение со словосочетанием «в обтяжку».

Заключение

В статье приведено описание алгоритмов автоматического снятия омонимии словосочетаний, которые могут выступать в роли предикативов, разработанных с использованием данных Национального корпуса русского языка. Алгоритмы были реализованы на языке программирования C++ в экспериментальном программном обеспечении для снятия омонимии. Полученные результаты могут быть использованы для повышения точности морфологической разметки при автоматической обработке текстов на русском языке.

Список литературы

1. Зинькина Ю. В. Разрешение функциональной омонимии в русском языке на основе контекстных правил [Текст] / Ю. В. Зинькина, Н. В. Пяткин, О. А. Невзорова // Труды международной конференции «Диалог 2005». – М. : Наука, 2005. – С. 198–202.
2. Зеленков Ю. Г. Вероятностная модель снятия морфологической омонимии на основе нормализующих подстановок и позиций соседних слов [Текст] / Ю. Г. Зеленков, И. В. Сегалович, В. А. Титов // Компьютерная лингвистика и интеллектуальные технологии. Труды международного семинара «Диалог 2005». – М. : Наука, 2005. – С. 616–638.
3. Мухамедшин Д. Р. Модуль разрешения морфологической неоднозначности: архитектура и организация базы данных [Текст] / Д. Р. Мухамедшин, Д. Ш. Сулейманов // Программные продукты и системы. – 2020. – Т. 33, № 1. – С. 38–46.
4. Бобичев В. Л. Автоматическое снятие морфологической многозначности при разметке корпуса [Текст] / В. Л. Бобичев // Труды международной конференции «Корпусная лингвистика–2008». – СПб. : СПбГУ, 2008. – С. 45–49.
5. Гатауллин Р. Р. Аналитический обзор методов разрешения морфологической многозначности [Текст] / Р. Р. Гатауллин // Электронные библиотеки. – 2016. – Т. 19, № 2. – С. 98–114.
6. Кобрицов Б. П. Снятие лексико-семантической омонимии в новостных и газетно-журнальных текстах: поверхностные фильтры и статистическая оценка / Б. П. Кобрицов, О. Н. Ляшевская, О. Ю. Шеманаева // Интернет-математика 2005. Автоматическая обработка веб-данных. – М., 2005. – С. 38–57.
7. Ширшикова А. А. О проблемах омонимии [Текст] / А. А. Ширшикова // Альманах современной науки и образования. – 2012. – № 2 (57). – С. 190–192.
8. Порохнин А. А. Анализ статистических методов снятия омонимии в текстах на русском языке [Текст] / А. А. Порохнин // Вестник астраханского государственного технического университета. – 2013. – № 2. – С. 168–174.
9. Лесько О. Н. Использование онтологии предметной области для снятия омонимии в естественно-языковых текстах / О. Н. Лесько, Ю. В. Рогушина // Проблемы програмування. – 2017. – № 2. – С. 61–71.
10. Рысаков С. В. Статистические методы снятия омонимии [Текст] / С. В. Рысаков, Э. С. Клышинский // Новые информационные технологии в автоматизированных системах. – 2015. – № 18. – С. 555–563.

11. Хаген М. Полная парадигма. Морфология [Электронный ресурс] / М. Хаген // Форум «Говорим по-русски» [сайт]. 2018. – Режим доступа: <http://www.speakrus.ru/dict/#morph-paradigm> (дата обращения: 10.06.2019).
12. Ниценко А. В. О подчинительном дереве для простого распространенного русского предложения [Текст] / А. В. Ниценко, В. Ю. Шелепов, С. А. Большакова // Проблемы искусственного интеллекта. – 2019. – № 2(13). – С. 94–103.
13. Ляшевская О. Н. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка) [Текст] / О. Н. Ляшевская, С. А. Шаров. – М. : Азбуковник, 2009. – 1087 с.
14. Ниценко А. В. О снятии омонимии словосочетаний, которые могут быть предикативами [Текст] / А. В. Ниценко, В. Ю. Шелепов, С. А. Большакова // Проблемы искусственного интеллекта. – 2021. – № 1(20). – С. 53–63.
15. Национальный корпус русского языка [Электронный ресурс]. – URL: <http://www.ruscorpora.ru/> (дата обращения: 01.05.2021).

References

1. Zin'kina Yu. V., Pyatkin N. V., Nevzorova O. A. Razresheniye funktsional'noy omonimii v russkom yazyke na osnove kontekstnykh pravil [Resolution of functional homonymy in Russian based on contextual rules]. *Trudy mezhdunarodnoy konferentsii «Dialog 2005»* [Proceedings of the international seminar "Dialogue 2005"], M., Nauka, 2005, S. 198–202.
2. Zelenkov Yu. G., Segalovich I. V., Titov V. A. Veroyatnostnaya model' snyatiya morfologicheskoy omonimii na osnove normalizuyushchikh podstanovok i pozitsiy sosednikh slov [Probabilistic model for removing morphological homonymy based on normalizing substitutions and positions of neighboring words]. *Komp'yuternaya lingvistika i intellektual'nyye tekhnologii. Trudy mezhdunarodnogo seminar «Dialog 2005»* [Computer linguistics and intellectual technologies. Proceedings of the international seminar "Dialogue 2005"], M., Nauka, 2005, S. 616–638.
3. Mukhamedshin D. R., Suleymanov D. Sh. Modul' razresheniya morfologicheskoy neodnoznachnosti: arkhitektura i organizatsiya bazy dannyakh [Modul' razresheniya morfologicheskoy neodnoznachnosti: arkhitektura i organizatsiya bazy dannyakh]. *Programmnyye produkty i sistemy* [Software products and systems], 2020, T. 33, № 1, S. 38–46.
4. Bobichev V. L. Avtomaticheskoye snyatiye morfologicheskoy mnogoznachnosti pri razmetke korpusa [Automatic removal of morphological ambiguity when marking the corpus]. *Trudy mezhdunarodnoy konferentsii «Korpusnaya lingvistika–2008»* [Proceedings of the international conference "Corpus linguistics-2008"], SPb., SPbGU, 2008, S. 45–49.
5. Gataullin R. R. Analiticheskiy obzor metodov razresheniya morfologicheskoy mnogoznachnosti [Analytical review of methods for resolving morphological polysemy]. *Elektronnyye biblioteki* [Electronic libraries], 2016, T. 19, № 2, S. 98–114.
6. Kobritsov B. P., Lyashevskaya O. N., Shemanayeva O. Yu. Snyatiye leksiko-semanticheskoy omonimii v novostnykh i gazetno-zhurnal'nykh tekstakh: poverkhnostnyye fil'try i statisticheskaya otsenka [Removing lexical-semantic homonymy in news and newspaper and magazine texts: surface filters and statistical evaluation]. *Internet-matematika 2005. Avtomaticheskaya obrabotka veb-dannykh* [Internet mathematics 2005. Automatic processing of web data], M., 2005, S. 38–57.
7. Shirshikova A. A. O problemakh omonimii [On the problems of homonymy]. *Al'manakh sovremennoy nauki i obrazovaniya* [Almanac of modern science and education], 2012, № 2 (57), S. 190–192.
8. Porokhnin A. A. Analiz statisticheskikh metodov snyatiya omonimii v tekstakh na russkom yazyke [Analysis of statistical methods for removing homonymy in Russian texts] *Vestnik astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta* [Bulletin of the Astrakhan State Technical University], 2013, № 2, S. 168–174.
9. Les'ko O. N., Rogushina Yu. V. Ispol'zovaniye ontologii predmetnoy oblasti dlya snyatiya omonimii v yestestvenno-yazykovykh tekstakh [Using the domain ontology for removing homonymy in natural language texts]. *Problemi programuvannya* [Problems of the program], 2017, № 2, S. 61–71.
10. Rysakov S. V., Klyshinskiy E. S. Statisticheskiye metody snyatiya omonimii [Statistical methods for removing homonymy]. *Novyye informatsionnyye tekhnologii v avtomatizirovannykh sistemakh* [New information technologies in automated systems], 2015, № 18, S. 555–563.
11. Khagen M. Polnaya paradihma. Morfologiya [Complete paradigm. Morphology]. *Forum «Govorim po-russki»* [Forum "We speak Russian"], 2018, Rezhim dostupa: <http://www.speakrus.ru/dict/#morph-paradigm> (data obrashcheniya: 10.06.2019).

12. Nitsenko A. V., VShelepov. Yu., Bol'shakova S. A. O podchinitel'nom dereve dlya prostogo rasprostranennogo russkogo predlozheniya [On the subordinate tree for a simple common Russian sentence]. *Problemy iskusstvennogo intellekta* [Problems of artificial intelligence], 2019, № 2(13), S. 94–103.
13. Lyashevskaya O. N., Sharov S. A. Chastotnyy slovar' sovremennogo russkogo yazyka (na materialakh Natsional'nogo korpusa russkogo yazyka) [Frequency dictionary of the modern Russian language (based on the National Corpus of the Russian language)], M., Azbukovnik, 2009, 1087 s.
14. Nitsenko A. V., Shelepov V. Yu., Bol'shakova S. A. O snyatii omonimii slovosochetaniy, kotoryye mogut byt' predikativami [On removing the homonymy of phrases that can be predicatives]. *Problemy iskusstvennogo intellekta* [Problems of artificial intelligence], 2021, № 1(20), S. 53-63.
15. *Natsional'nyy korpus russkogo yazyka* [National corpus of the Russian language]
URL: <http://www.ruscorpora.ru/> (data obrashcheniya: 01.05.2021).

RESUME

A. V. Nitsenko, V. Yu. Shelepov, S. A. Bolshakova
On Automatic Disambiguation of Predicative Word Combinations.
Results of Work with the National Corpus of the Russian Language

The problem of disambiguation is one of the most important in natural language automatic processing. This problem is especially relevant for the Russian language since the presence of an extremely large number of homonyms. Therefore, the disambiguation is an important and necessary stage for high-quality text processing and analysis.

The article describes the algorithms of automatic disambiguation in Russian text for word combinations that can act as predicatives. These algorithms were developed using data from the Russian National Corpus.

The developed algorithms were implemented using the C++ programming language in experimental disambiguation software.

The results obtained can be used to improve the accuracy of morphological tagging in automatic processing of Russian texts.

РЕЗЮМЕ

А. В. Ниценко, В. Ю. Шелепов, С. А. Большакова
Об автоматическом снятии омонимии предикативных словосочетаний.
Результаты работы с национальным корпусом русского языка

Проблема разрешения неоднозначности слов является одной из важнейших в задачах автоматической обработки естественного языка. Для русского языка эта проблема особенно актуальна, поскольку количество омонимов очень велико из-за наличия словоформ. Поэтому процедура снятия омонимии является важным и необходимым этапом для качественной обработки и анализа текстов.

В статье приведено описание алгоритмов автоматического снятия омонимии словосочетаний, которые могут выступать в роли предикативов. Алгоритмы разработаны с использованием данных Национального корпуса русского языка.

Алгоритмы были реализованы на языке программирования C++ в экспериментальном программном обеспечении для снятия омонимии.

Полученные результаты могут быть использованы для повышения точности морфологической разметки при автоматической обработке текстов на русском языке.

Статья поступила в редакцию 09.06.2021.