

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

О СНЯТИИ ОМОНИМИИ СЛОВСОЧЕТАНИЙ, КОТОРЫЕ МОГУТ БЫТЬ ПРЕДИКАТИВАМИ

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 THE DISAMBIGUATION OF WORD COMBINATIONS THAT MAY BE PREDICATIVES

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 ЗНЯТТЯ ОМОНИМІЇ СЛОВСПОЛУЧЕНЬ, ЯКІ МОЖУТЬ БУТИ ПРЕДИКАТИВАМИ

В статье предложены методы автоматического снятия омонимии словосочетаний, которые могут выступать в роли предикативов. Приведено описание соответствующих алгоритмов, рассмотрены примеры их работы. Результаты реализованы в экспериментальном программном обеспечении для снятия омонимии.

Ключевые слова: автоматический анализ текста, методы снятия омонимии, неделимое словосочетание, предложная группа.

The article proposes methods for automatic disambiguation for word combinations that can act as predicatives. The description of the corresponding algorithms is given, examples of their work are considered. The results are implemented in experimental disambiguation software.

Key words: automatic text analysis, disambiguation methods, indivisible collocation, prepositional phrase.

У статті запропоновано методи автоматичного зняття омонімії словосполучень, які можуть виступати в ролі предикативів. Наведено опис цих алгоритмів, розглянуто приклади їх роботи. Результати реалізовані в експериментальному програмному забезпеченні для зняття омонімії.

Ключові слова: автоматичний аналіз тексту, методи зняття омонімії, неподільне словосполучення, прийменникова група.

Введение

Тематика работы связана с задачей обучения компьютера умению разбираться с проблемой омонимии. Процедуры снятия омонимии являются важным и необходимым этапом для качественной обработки и анализа текстов. Они активно исследуются в настоящее время [1-10].

Данная статья посвящена автоматическому снятию омонимии в группах омонимов, содержащих словосочетания, которые могут играть роль предикативов. Они могут также выступать в роли наречий, местоимений, частиц и вводных словосочетаний. Во всех этих случаях словосочетание интерпретируется как единое целое. С другой стороны оно может в ряде случаев требовать разбиения на отдельные слова. Здесь выделяются случаи предложной группы (см. разделы 5 и 6), разделимой непреложной группы (пометка (соч), см. раздел 4) и того, что естественно именовать отрицанием (см. раздел 3). Раздел 2 посвящен выбору омонима в случае, если словосочетание трактуется как единое целое. Раздел 1 описывает организацию программы, в которой реализуются подходы, развитые в последующих разделах статьи.

1 О программе снятия омонимии

Все результаты работы программно реализованы. В качестве лексической базы мы используем словарь [11] (подробнее по этому поводу см. работу [12]) и прежде всего, те его словосочетания, которые входят в него с пометкой «предик». Если может потребоваться разбиение словосочетания на отдельные слова, мы включаем его в словарь с одной из пометок «предл гр», «соч», «отриц». Мы используем представление этого словаря в виде дерева, что позволяет очень быстро осуществлять в нем поиск. Мы реализовали также вариант словаря, в котором следом за каждой словоформой приведена ее лемма (начальная форма слова). Это позволяет с той же сверхвысокой скоростью осуществлять лемматизацию слов.

Отметим также, что в ряде случаев мы ограничиваемся лишь словами, вошедшими в частотный словарь [13] Ляшевской и Шарова, созданный на основе корпуса русского языка. На рис. 1 представлено основное окно программы:

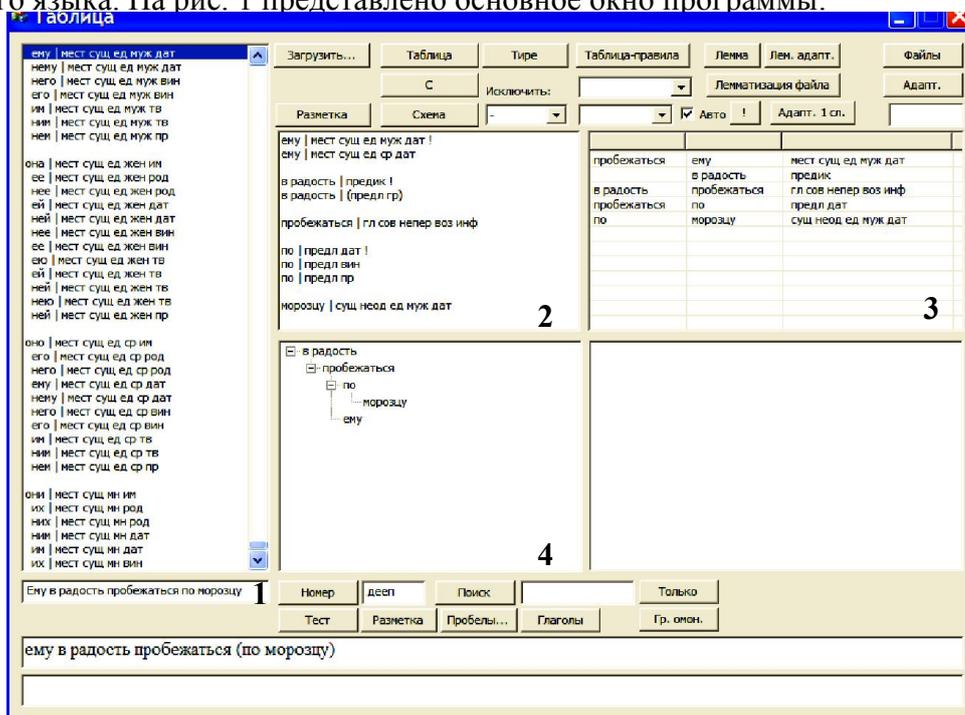


Рисунок 1 – Основное окно программы для снятия омонимии

Анализируемое предложение вводится в поле 1. Поле 2 содержит последовательность групп омонимов. В каждой из них программа отмечает выбранный ею омоним восклицательным знаком. Отметим, что одновременно она создает таблицу подчинения (поле 3) и подчинительное дерево (поле 4). Возможны случаи, когда предложение может нести более одного смысла. В этом случае однозначный выбор омонима компьютером невозможен. По умолчанию восклицательный знак ставится, учитывая наиболее частый вариант. При необходимости можно скорректировать выбор программы вручную, щелкнув мышкой на нужной строке, и получить новый результат работы программы, нажав кнопку «С».

2 Неделимые словосочетания

Словосочетания:

в ажуре, в долгу, в охотку, в самый раз, в тягость, в ходу, гуд бай, как без рук, на выданье, на загляденье, на слуху, на сносях, не вправе, не для чего, не до смеха, не до шуток, не замай, не к спеху, не надо, не надобно, не под силу, не под стать, не поздоровится, не тут-то было, не худо бы, ни гугу, ни при чем, ни шиша, по нутру, под силу, под стать, пустяк пустяком

никогда не разделяются на отдельные слова.

Для данных словосочетаний группа омонимов будет содержать предикатив и омонимы из числа следующих: наречие, междометие и вводное словосочетание. Мы предлагаем следующие правила, определяющие выбор омонима:

- 1) если предложение состоит только из указанного словосочетания, то для него выбирается значение предикатива;
- 2) если словосочетание выделено запятыми, то это вводное словосочетание;
- 3) если словосочетание выделено восклицательным знаком, то это междометие;
- 4) если в предложении есть глагол в личной форме, то для словосочетания выбирается значение наречия. Оно подчиняется этому глаголу;
- 5) если в предложении есть глагол в инфинитиве, то в рассматриваемой группе выбирается предикатив;
- 6) если в предложении нет глагола, то в рассматриваемой группе выбирается предикатив.

3 Словосочетания с отрицанием

Индикатор разделения: после словосочетания стоит противопоставление, начинающееся союзом «а», или после словосочетания просто стоит запятая. В скобках приведены примеры предложений:

не грех (Это не грех, а достоинство); *не дело* (Это не дело, а развлечение); *не беда* (Это не беда, а неприятность); *не впрок* (Это не впрок, а хлеб насущный); *не лучше* (Это не лучше, а хуже); *не мудрено* (Это не мудрено, а совсем просто); *не разгонишься* (Там не разгонишься, а только успевай тормозить); *не редкость* (Это не редкость, а обычная вещь); *не резон* (Это не резон, а не понятно что); *не смешно* (Это не смешно, а грустно); *не удивительно* (Это не удивительно, а закономерно); *не худо* (Это не худо, а хорошо); *не чета* (Это не чета, а любовная пара); *не в ходу* (Эта электричка не в ходу, а в ремонте); *не к лицу* (Прижимай не к лицу, а к груди); *не к масти* (Он привязывается не к масти, а к силе лошади); *не к месту* (Он привязан не к месту, а к делу); *не по душе* (Он грустит не по душе ее, а по телу); *не по мне* (Иди

не по мне, а рядом); *не по нутру* (Это удар не по нутру его, а по душе); *не по силам* (Он действует не по силам, а по хотению); *не при деньгах* (Он остался не при деньгах, а только при славе); *не до того* (Нам не до того, чтобы развлекаться).

Еще индикаторы разделения: наличие в предложении дательного падежа существительного (местоимения-существительного), предлога «без» или наречий «зря», «впустую», «напрасно», «понапрасну»:

не прочь (Не прочь другому горя. Не прочь без нужды худа. Не прочь результата зря (впустую, напрасно, понапрасну)).

Словосочетание «*ничего себе*» включено в файл «*Предл гр.txt*» (см. раздел 5). Индикатором разделения для него является наличие слова «не» в предложении. Кроме того, словосочетание разделяется, если ему непосредственно предшествует либо следует за ним союз «и». С этой целью в файле «*Предл гр.txt*» соответствующая строка содержит запись «и+».

4 Правила для делимых словосочетаний, которые не являются предложными группами

Данные словосочетания введены нами в словарь [11] с пометкой (соч).

куда там – делится на отдельные слова, если предложение вопросительное и в нем есть одно из следующих слов: *встать, стать, становиться, идти, пойти, зайти, ехать, поехать, заехать, скакать, летать, лететь, полететь, лечь, прилечь, залечь, класть, положить, вкладывать, вложить, заложить, разложить, расположить, ставить, поставить, писать, написать, вписать, вписывать*. С этой целью словосочетание включено в файл «*Предл гр.txt*» и слова-индикаторы в нем помечены знаком «?». Данное словосочетание интерпретируется как предикатив, если дальше есть инфинитив глагола, либо как частица, если инфинитива нет. То же относится к словосочетанию «*куда тут*».

вероятнее всего – если после «*вероятнее всего*» идет одно из слов *добиться, достигнуть* и дальше дееприч. оборот или оборот, начинающийся словами *при условии* или союзом *если* или союзом *когда*, то словосочетание делится. В противном случае это предикатив или вводное (в случае выделения запятыми).

все равно – если дальше есть числительное в дат. падеже, то словосочетание делится. В противном случае необходимо сделать выбор между частицей и предикативом вручную.

и все тут – если после словосочетания стоит знак препинания, то это – предикатив (по умолчанию) или выбор правильного варианта вручную. Если знака препинания нет, словосочетание делится.

как раз – программа выберет (соч) и разделит словосочетание в случае предложения без подлежащего. Пример: *Как раз случилось в декабре*. Если в предложении есть подлежащее и есть глагол в личной форме, то словосочетание – частица. Пример: *Как раз случилось наводнение*. Если в предложении есть подлежащее, но нет глагола в личной форме, то необходимо выбрать вручную между предикативом и частицей. Примеры: *Пиджак как раз. Как раз поезд*.

как сказать – в утвердительном предложении «*Это как сказать*» – предикатив. В ответе после вопросительного предложения – предикатив. При наличии далее текста в кавычках – словосочетание делится. Пример: *Как сказать по-английски «давно»?*

может быть – деление, если справа соседствует существительное (местоимение-существительное) или прилагательное (местоимение-прилагательное) или причастие в творительном падеже. Примеры:

Это может быть существительным. Это может быть красивым. Это может быть решенным.

Выбор вручную между делением (в случае утверждения) и предикативом (в случае предположения), если справа соседствует существительное (местоимение-существительное) в именительном падеже или прилагательное в краткой форме или причастие в краткой форме. Примеры:

Это может быть существительное. Это может быть красиво. Это может быть решено.

При выделении словосочетания запятыми оно является вводным.

проще всего – словосочетание делится, если после него идет одно из слов *остального, другого, прочего*. В противном случае – это предикатив.

так себе – если после «*так себе*» стоит знак препинания, то при наличии в предложении глагола в личной форме словосочетание есть наречие, при отсутствии – предикатив. Если знака препинания нет, словосочетание делится. Пример: *Так себе и запиши.*

5 Семантико-лексический алгоритм снятия омонимии словосочетаний, которые могут быть предложными группами

Алгоритм работает с файлом «*Предл гр.txt*», который содержит текстовые записи, определяющие, должно ли словосочетание интерпретироваться как единое целое, либо как отдельные слова. Содержимое файла приведено в табл. 1.

Таблица 1 – Содержимое файла «*Предл гр.txt*»

Словосочетание	Пометка	Слова, определяющие разделение или его отсутствие
без интереса	! (предл гр)	к
без ума	!	от
в годах	! (предл гр)	так, этак
в диковинку	! (предл гр)	вглядываться, взглядеться, всматриваться, всмотреться
в лад	!	звучать, играть, петь
в моде	! (предл гр)	авторитет, понимать, разбираться, эксперт
в новинку	! (предл гр)	вглядываться, взглядеться, вкладывать, вложить, вложиться, всматриваться, всмотреться
в обтяжку	! (предл гр)	вкладывать, вложить, вложиться, ввязаться, втянуться

Продолж. табл. 1

в ответ	! (предл гр)	вглядываться, взглядеться, вкладывать, вложить, вслушиваться, вслушаться, всматриваться, всмотреться
в расчете	! (предл гр)	неточность, ошибка, погрешность
в сборе	! (предл гр)	участие, участвовать
в состоянии	! (предл гр)	изменение, перемена
в ударе	! (предл гр)	собрать, сосредоточить, участие, участвовать
к лицу	! (предл гр)	приближать, прикладывать, прислонять
к масти	! (предл гр)	идти, подойти, подходить
к месту	!	вмешаться, встрять, выступить, высказать, выступить, говорить, сказать
к черту	! (предл гр)	валиться, идти, катиться
куда там	! (соч)	встать, стать, становиться, идти, пойти, зайти, ехать, поехать, заехать, скакать, летать, лететь, полететь, лечь, прилечь, залечь, класть, положить, заложить, разложить, расположить, ставить, поставить, писать, написать, вписать, вписывать
куда тут	! (соч)	встать, стать, становиться, идти, пойти, зайти, ехать, поехать, заехать, скакать, летать, лететь, полететь, лечь, прилечь, залечь, класть, положить, заложить, разложить, расположить, ставить, поставить, писать, написать, вписать, вписывать
к чертям	! (предл гр)	валиться, идти, катиться
на редкость	! (предл гр)	взирать, дивиться, любоваться, смотреть
не в ладах	! (предл гр)	дело, причина, проблема
не в ладу	!	с
ничего себе	! (отриц)	и+, не
не в курсе	!	дело, причина, проблема
не в ударе	! (предл гр)	дело, причина, проблема
по дороге	! (предл гр)	двигаться, перемещаться, идти, выходить, выйти, проходить, пройти, ходить, доходить, дойти, ехать, выезжать, выехать, проезжать, проехать, скакать, проскакать, катиться, прокатиться, прикатиться, летать, лететь, пролетать, пролететь, прилететь, мчаться, промчаться, примчаться, ползти, проползти, приползти

Продолж. табл 1

по пути	! (предл гр)	двигаться, перемещаться, идти, выходить, выйти, проходить, пройти, ходить, доходить, дойти, ехать, выезжать, выехать, проезжать, проехать, скакать, проскакать, катиться, прокатиться, прикатиться, летать, лететь, пролетать, пролететь, прилететь, мчаться, промчаться, примчаться, ползти, проползти, приползти
не к чему	! (предл гр)	привязаться, придраться

Опишем работу с этим файлом на конкретном примере. Пусть в поле 1 программы введено предложение «*Нынче в моде соломенные шляпки*». По нажатию кнопки «Таблица» программа ищет словосочетания из файла «*Предл гр.txt*» во введенном предложении. Определив, что предложение содержит словосочетание «*в моде*», она ищет в предложении словоформу какой-либо из лемм, относящихся к данному словосочетанию (см. третий столбец таблицы 1). Обнаружив отсутствие таковых, она интерпретирует «*в моде*» как предикатив, выводит словосочетание в соответствующей строке поля 2 (см. рис. 1) целиком и отмечает восклицательным знаком.

Пусть теперь в окно программы введено предложение «*Девушка разбирается в моде*». Действуя аналогично вышеописанному, программа обнаруживает в предложении словоформу слова «*разбираться*». В результате она интерпретирует «*в моде*» как предложную группу и выводит в поле 2 (см. рис. 1) предлог и существительное в отдельных строках (то есть разбивает словосочетание).

Иная ситуация со словосочетанием «*в лад*», где пометка (предл гр) отсутствует. Здесь обнаружение словоформы одной из лемм, стоящих в последней ячейке, приведет к интерпретации словосочетания «*в лад*» как неразделимого и, следовательно, как наречия или предикатива. В предложении без глагола в личной форме программа определит предикатив. В предложении с глаголом в личной форме она определит наречие. Файл «*Предл гр.txt*» будет пополняться за счет работы с Национальным корпусом русского языка.

Назовем слова в последней колонке табл. 1 «индикаторами разделения словосочетаний». В случае, когда таких индикаторов заведомо не слишком много, описанный семантико-лексический метод снятия омонимии является эффективным. В других случаях может оказаться эффективным другой метод.

6 О семантико-синтаксическом алгоритме снятия омонимии в словосочетаниях с предложными группами

Опишем суть этого алгоритма на примере словосочетания «*в радость*». Здесь индикатором нераздельности является наличие в предложении существительного или местоимения-существительного в дательном падеже. Пример: *Ему в радость пробежаться по морозцу*.

Результат работы программы:

```
ему | мест сущ ед муж дат !
ему | мест сущ ед ср дат

в радость | предик !
в радость | (предл гр)

пробежаться | гл сов непер воз инф

по | предл дат !
по | предл вин
по | предл пр

морозцу | сущ неод ед муж дат
```

При отсутствии существительного или местоимения-существительного в дательном падеже рассматриваемое словосочетание обрабатывается как предложная группа. Пример: Они верили в радость грядущих дней.

Результат работы программы:

```
они | мест сущ мн им

верили | гл несов непер прош мн

в | предл пр
в | предл вин !

радость | сущ неод ед жен им
радость | сущ неод ед жен вин !

грядущих | прл мн род !
грядущих | прл мн вин одуш
грядущих | прл мн пр
грядущих | прич несов непер наст мн род
```

Другие примеры применения семантико-синтаксического алгоритма содержатся в разделе 4. Дополнительный материал по этому поводу будет в последующих публикациях.

Заключение

В статье приведено описание алгоритмов автоматического снятия омонимии в группах омонимов, содержащих словосочетания, которые могут играть роль предикативов. Разработанные алгоритмы реализованы с использованием языка программирования C++ в экспериментальном программном обеспечении для снятия омонимии. Полученные результаты могут быть использованы для автоматизации морфологической разметки естественно-языковых текстов со снятием неоднозначности, что позволит повысить точность классификации и кластеризации текстов, улучшить качество машинного перевода, информационного поиска и других приложений.

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. 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] / Бобичев В. Л. // Труды международной конференции «Корпусная лингвистика–2008». – СПб. : СПбГУ, 2008. – С. 45–49.
5. Гатауллин Р. Р. Аналитический обзор методов разрешения морфологической многозначности [Текст] / Гатауллин Р. Р. // 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 [Текст] / 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] / Ширшикова А. А. // 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] / Порохнин А. А. // 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 [Текст] / 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] // 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. 63–73.
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.

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 conference "Dialogue 2005"], M., Nauka, 2005, pp. 198–202.
2. Zelenkov Yu.G., Segalovich I.V., Titov V.A. Veroyatnostnaya model' snyatiya morfologicheskoy omonimii na osnove normalizuyushchikh podstanovok i pozitsiy soednikh slov [Probabilistic model of removing morphological homonymy based on normalizing substitutions and positions of neighboring words] *Komp'yuternaya lingvistika i intellektual'nyye tekhnologii. Trudy mezhdunarodnogo seminara «Dialog 2005»* [Computational linguistics and intellectual technologies. Proceedings of the international seminar "Dialogue 2005"], M., Nauka, 2005, pp. 616–638.
3. Mukhamedshin D.R., Suleymanov D.Sh. Modul' razresheniya morfologicheskoy neodnoznachnosti: arkhitektura i organizatsiya bazy dannykh [Module for resolving morphological ambiguity: architecture and organization of the database]. *Programmnyye produkty i sistemy* [Software products and systems] 2020, T. 33, No 1, pp. 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, pp. 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, No 2, pp. 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 [Removal of 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, pp. 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, No 2 (57), pp. 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, No 2, pp. 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, No 2, pp. 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, No 18, pp. 555-563.
11. Khagen, M. Polnaya paradigma. Morfologiya [Complete paradigm. Morphology]. *Forum «Govorim po-russki»* [Forum "We speak in Russian"], 2018, Rezhim dostupa: <http://www.speakrus.ru/dict/#morph-paradigm> (data obrashcheniya: 10.06.2019).
12. Nitsenko A.V., Shelepov V.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, No. 2 (13), pp. 63–73.
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 Materials of the National Corpus of the Russian Language)], M., Azbukovnik. 2009, 1087 pp.

RESUME

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

On the disambiguation of word combinations that may be predicative

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 principles of automatic disambiguation in Russian text for word combinations that can act as predicatives. Appropriate algorithms are proposed that use semantic-lexical and semantic-syntactic approaches to text analysis.

The article contains a description of the proposed algorithms, examples of their work are considered. The developed algorithms were implemented in experimental disambiguation software using the C++ programming language.

The results obtained can be used to automate the morphological tagging of natural language texts with disambiguation, which will improve the accuracy of text classification and clustering, improve the quality of machine translation, information search and other applications.

РЕЗЮМЕ

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

О снятии омонимии словосочетаний, которые могут быть предикативами

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.

Описан принцип автоматического снятия омонимии в тексте для словосочетаний, которые могут выступать в роли предикативов. Задача решается с помощью применения семантико-лексических и семантико-синтаксических алгоритмов, основанных на правилах.

В статье приведено описание предложенных алгоритмов, рассмотрены примеры их работы. Разработанные алгоритмы были реализованы с использованием языка программирования C++ в экспериментальном программном обеспечении для снятия омонимии.

Полученные результаты могут быть использованы для автоматизации морфологической разметки естественно-языковых текстов со снятием неоднозначности, что позволит повысить точность классификации и кластеризации текстов, улучшить качество машинного перевода, информационного поиска и других приложений.

Статья поступила в редакцию 20.01.2021.