

УДК 004.89:004.93

DOI 10.24412/2413-7383-2025-2-37-13-20

А. В. Ниценко, В. Ю. Шелепов
ФГБНУ «Институт проблем искусственного интеллекта»,
283048, г. Донецк, ул. Артема, 118-б

МЕТОД СНЯТИЯ НЕОДНОЗНАЧНОСТИ ВИНИТЕЛЬНОГО И ТВОРИТЕЛЬНОГО ПАДЕЖА ДЛЯ РЯДА СУЩЕСТВИТЕЛЬНЫХ РУССКОГО ЯЗЫКА

A. V. Nicenko, V. Ju. Shelepov
Federal Budgetary State Scientific Institution «Institute of Artificial Intelligence Problems»,
Donetsk, Artema str., 118-b.

A METHOD OF ACCUSATIVE AND INSTRUMENTAL CASES DISAMBIGUATION FOR A NUMBER OF RUSSIAN NOUNS

В статье предложен метод автоматического снятия омонимии винительного и творительного падежа для ряда существительных с использованием данных о глагольном управлении, полученных из Национального корпуса русского языка. Результаты реализованы в экспериментальном программном обеспечении для снятия омонимии.

Ключевые слова: обработка естественного языка, снятие омонимии, существительное, винительный падеж, творительный падеж, глагольные формы.

The article proposes a method for automatic disambiguation of accusative and instrumental cases for a number of nouns using verb control data obtained from the Russian National Corpus. The results are implemented in experimental software for disambiguation.

Keywords: natural language processing, disambiguation, noun, accusative case, creative case, verb forms.

Введение

Явление омонимии в русском языке создает значительные сложности при автоматической обработке текста, поэтому необходимым этапом является ее снятие. Проблема снятия омонимии актуальна для множества задач: машинный перевод (например, в [1] показано влияние разрешения омонимии на точность перевода), автоматическое извлечение информации, автоматический анализ содержания, обработка речи и текста [2-13]. Для текстов на русском языке проблема заключается не только в частеречной, но и в морфологической омонимии. В частности, большую трудность представляет омонимия падежных форм существительных, при которой словоформы в разных падежах совпадают по написанию.

Целью данной работы является разработка метода снятия омонимии винительного и творительного падежа существительного, основанного на использовании глагольных форм, связанных с рассматриваемыми омонимами.

О снятии омонимии винительного и творительного падежа существительных

При анализе текста мы работаем с морфологическим словарем [14], содержащим более четырех миллионов русских словоформ (с полной грамматической разметкой), представляя его в виде дерева. Последнее дает возможность почти мгновенного поиска нужной словоформы и, в частности, позволяет легко выделить все омонимы для каждого слова в тексте.

В используемом нами морфологическом словаре насчитывается 85 словоформ существительных с данной омонимией. В настоящей работе мы ограничимся следующими 21-ой омонимичными словоформами:

блудней, глушней, долбней, дусей, емелей, земелей, злыдней, зюзей, квашней, клянчей, мозглей, мямлей, офеней, нюней, писклей, простофилей, растерей, рохлей, сводней, тихоней, тютей.

Излагаемые ниже результаты реализованы в виде программного модуля в экспериментальном программном обеспечении для морфологической разметки текста [15], [16]. При вводе русского предложения программа выводит в соответствующем окне вместо словоформ столбец всех соответствующих омонимов, снабжая выбранные омонимы восклицательными знаками (рисунок 1).

Сразу оговорим следующие используемые нами принципы:

- 1) все формулируемые ниже правила применяются только к отрезку между двумя ближайшими знаками препинания, содержащему интересующий нас омоним,
- 2) если программа однозначно определила на упомянутом отрезке наличие существительного в винительном падеже (без омонимии), то для интересующего нас омонима винительный падеж не выбирается.

Далее, для распознавания интересующих нас омонимов используется текстовый управляющий файл *Вин-тв.txt*, состоящий из групп, разделенных пробельными строками. Будем эти группы также называть управляющими. К слову *блудней* относятся две группы. Вот содержимое первой из них:

блудней !(тв) (1)
<гл-непер>
вести + себя
величать |тв-вин
выглядеть |тв-вин

глядеть
назвать |тв-вин
называть |тв-вин
нарекать |тв-вин
наречь |тв-вин
обозвать |тв-вин
обзывать |тв-вин
окрестить |тв-вин
повести + себя
представить
предстать
прожить
прозвать |тв-вин
прозывать |тв-вин
расти
сделать |тв-вин
смотреть
стать
счесть |тв-вин
считать |тв-вин
чувствовать + себя

С глаголами этой группы для омонима *блудней* программой выбирается творительный падеж. При этом строка <гл-непер> подразумевает любой непереходный глагол (подходящий по смыслу и, следовательно, возможный в тексте). Метка |*тв-вин* указывает на то, что с данным глаголом возможен также омоним винительного падежа (это не метка для программы, а информация для человека). Пример: *Он величал блудней(вин) разбойниками.*

Слово *назвать* может иметь смысл *созвать*.

Для глаголов, не попавших в группу (1) программа по умолчанию автоматически выбирает винительный падеж, кроме случая, когда глагол представлен формой краткого страдательного причастия, относящегося к нашему омониму. В последнем случае всегда выбирается творительный падеж.

Пример: *Мясо со стола украдено этой блудней.*

В управляющем файле для омонима *блудней* присутствует еще одна группа:

блудней !(вин-тв) (2)

выставить
заменить
заменять
звать
изображать
изобразить
именовать
кликать
обругать

С глагольными формами этой группы программа по умолчанию выбирает винительный падеж (если это не противоречит принципу 2), запрещающему два винительных падежа), но возможен также творительный падеж.

Примеры: *Он обругал блудней(вин). Он обругал кошку блудней(тв).*

Изменение выбора омонима осуществляется пользователем с использованием кнопки «С» (correct) в окне программы.

Аналогично и в основном теми же группами глаголов (1) и (2) управляется алгоритм выбора омонимов для следующих существительных:

глушней, долбней, дусей, емелей, земелей, злыдней, зюзей, квашней, клянчей, мозглей, мямлей, офеней, нюней, писклей, простофилей, растерей, рохлей, сводней, тихоней, тютей.

При этом для слов *глушней, долбней, злыдней, нюней, размазней, растерей, рохлей, тихоней, тютей* в группу вида (1) добавляется глагол *сидеть*, для слова *сводней* добавляется глагол *служить*, для слова *тютей* – глаголы *кутать, укутать*. В национальном корпусе русского языка [17] встречается сочетание *платить размазней(тв)*. Поэтому для слова *размазней* в группу вида (1) добавляется глагол *платить*.

Что касается слова *офеней*, то для него из соображений семантики из группы вида (1) удаляются глаголы *глядеть, нарекать, наречь, расти, смотреть*, а из группы вида (2) удаляется глагол *обругать*. Это вызвано, в частности, тем, что слово *офеней*, в отличие от слова *блудней*, не является поведенчески-оценочной характеристикой.

Аналогично для слова *земелей* из группы вида (1) удаляются глаголы *обозвать, обзывать, расти, сделать*. Для слова *тихоней* из группы вида (1) удаляются глаголы *обозвать, обзывать*, а из группы (2) удаляется глагол *обругать*.

Слово *дуся* интерпретируется не как женское имя, а как обозначение для милого, душевного человека, мужчины или женщины. Из соответствующей группы вида (1) для него из соображений семантики должны быть удалены глаголы *обозвать, обзывать*. Для глаголов *прозвать, прозывать, счесть и считать* удаляются комментирующие метки *|тв-вин*. Из соответствующей группы вида (2) должны быть удален глагол *обругать*.

Слово *емеля* также интерпретируется не как имя, а как характеристика, основным синоним которой является слово *пустомеля*.

На рисунке 1 приведен пример программного снятия омонимии для предложения «*Он искренне чувствовал себя наивным простофилей*».

Исходный текст		
Он искренне чувствовал себя наивным простофилей.		
Словарная разметка		Разметка со снятой омонимией
он мест сущ ед муж им		он мест сущ ед муж им
искренне нар опред кач		искренне нар опред кач
чувствовал гл несов перех прош ед муж		чувствовал гл несов перех прош ед муж
себя мест сущ род		себя мест сущ вин
себя мест сущ вин !		наивным прл ед муж тв
наивным прл ед муж тв !		простофилей сущ одуш ед общ тв
наивным прл ед ср тв		
наивным прл мн дат		
простофилей сущ одуш ед общ тв !		
простофилей сущ одуш мн род		
простофилей сущ одуш мн вин		
.		

Рисунок 1 – Результат снятия омонимии

Заключение

В статье приведено описание методов автоматического снятия омонимии винительного и творительного падежа некоторых существительных в сочетаниях с глагольными формами. Разработанные алгоритмы были реализованы с использованием языка 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. Титова, Е. В. Проблема синтаксической омонимии в системах машинного перевода. *Сборник работ 70-й научной конференции студентов и аспирантов Белорусского государственного университета*, 15-18 мая 2013 г., Минск: В 3 ч. Ч. 2. - Минск, 2013. - С. 161-164.
2. Сорокин, А. Б. Автоматическая морфологическая разметка текста / А. Б. Сорокин, Л. М. Железняк. *Научно-технический вестник Поволжья*. 2020. № 2. С. 44-46.
3. Литвинов, М. И. Метод повышения качества снятия омонимии с использованием статистики совместного употребления слов / М. И. Литвинов. *Качество. Инновации. Образование*. 2010. № 11(66). С. 53-58.
4. Литвинов, М. И. Модели управления на основе сочетаний из трёх слов и глагольное управление для поверхностного синтаксического анализа / М. И. Литвинов. *Новые информационные технологии в автоматизированных системах*. 2010. № 13. С. 107-112.
5. Бручес, Е. П. Гибридный подход к снятию частеречной омонимии на примере русского языка / Е. П. Бручес. *Материалы 54-й Международной научной студенческой конференции МНСК-2016: Прикладная лингвистика*, Новосибирск, 16–20 апреля 2016 года. Новосибирск: Новосибирский национальный исследовательский государственный университет, 2016. С. 13-14.
6. Хорошилов А. А. Контекстное разрешение омонимии на основе центроидно-контекстной модели / А.А. Хорошилов, Ю.В.Никитин, А.В. Кан, Я.Д.Козловская, Е.А.Евдокимова. *Труды ИСП РАН*. 2022. т. 34, вып. 5. С. 171-182.
7. Каневский, Е. А. К вопросу о снятии омонимии при компьютерном анализе текстов / Е. А. Каневский, К. К. Боярский. *Проблемы преобразования и регулирования региональных социально-экономических систем* / Под редакцией С.В. Кузнецова. Вып. 45. Санкт-Петербург : Санкт-Петербургский государственный университет аэрокосмического приборостроения, 2019. С. 37-41.
8. Зарипова, Д. А. Подходы к автоматическому разрешению многозначности на основе неравномерности распределения значений слов в корпусе / Д. А. Зарипова, Н. В. Лукашевич. *Вестник Московского университета. Серия 9: Филология*. 2023. № 6. С. 40-51.

9. Рыкунов, А.Н. Разработка метода обработки несловарных слов для повышения точности морфологического анализа текстов на русском языке / А. Н. Рыкунов, Е. В. Полицына, С. А. Полицын, А. С. Поречный. *Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии*. 2024. № 1. С. 166-177.
10. Кормановский, М. В. О применении методов анализа текстов на естественном языке с последующим формированием графовых структур / М. В. Кормановский, Л. Л. Волкова. *Инновационные, информационные и коммуникационные технологии : сборник трудов XXI Международной научно-практической конференции*, Сочи, 01–10 октября 2024 года. Москва: Ассоциация выпускников и сотрудников ВВИА им. проф. Жуковского, 2024. С. 88-91.
11. Красса, С. И. Снятие неоднозначности ключевых слов в экспертизе медиатекста / С. И. Красса. *Современный медиатекст и судебная экспертиза: междисциплинарные связи и экспертная оценка: Сборник научных работ по итогам Международной научно-практической конференции*, Москва, 12–13 октября 2023 года. Москва: ООО "СОЮЗКНИГ", 2023. С. 153-160.
12. Столяров, И. И. Автоматическое разграничение омографов (на материале НКРЯ) / И. И. Столяров, О. А. Митрофанова. *Международная научная конференция, посвященная 20-летию Национального корпуса русского языка : Материалы конференции, Москва, 20–21 декабря 2024 года*. Москва: Институт русского языка им. В.В. Виноградова РАН, 2024. С. 163-165.
13. Кочеткова, Н. А. Метод формирования модели глагольного управления для русского языка / Н. А. Кочеткова. *Новые информационные технологии в автоматизированных системах*. 2012. № 15. С. 199-202.
14. Большакова, С. А. О снятии омонимии «предикатив-предложная группа» для некоторых распространенных словосочетаний в русскоязычных текстах / С. А. Большакова. *Проблемы искусственного интеллекта*. 2023. № 1(28). С. 11-17.
15. 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).
16. Ниценко, А. В. О родительном падеже существительного после глагола в случае его омонимии с винительным падежом / А. В. Ниценко, В. Ю. Шелепов. *Проблемы искусственного интеллекта*. 2025. № 1(36). С. 4-11.
17. Ниценко, А. В. Об использовании семантической информации для снятия омонимии именительного и винительного падежа (как элемента создания онтологии) / А. В. Ниценко, В. Ю. Шелепов. *Проблемы искусственного интеллекта*. 2024. № 4(35). С. 16-24.
18. *Национальный корпус русского языка* [Электронный ресурс]. Режим доступа: URL: <http://www.ruscorpora.ru/> (дата обращения: 19.03.2025).

References

1. Titova, E. V. The problem of syntactic homonymy in machine translation systems. Collection of papers of the 70th scientific conference of students and postgraduates of the Belarusian State University, May 15-18, 2013, Minsk: In 3 parts. Part 2. - Minsk, 2013. - P. 161-164.
2. Sorokin, A. B. Automatic morphological tagging of text / A. B. Sorokin, L. M. Zheleznyak // *Scientific and Technical Bulletin of the Volga Region*. - 2020. - No. 2. - P. 44-46.
3. Litvinov, M. I. Method for improving the quality of disambiguation using statistics of word co-use / M. I. Litvinov // *Quality. Innovations. Education*. - 2010. - No. 11 (66). - P. 53-58.
4. Litvinov, M. I. Control models based on three-word combinations and verb control for superficial syntactic analysis / M. I. Litvinov // *New information technologies in automated systems*. - 2010. - No. 13. - P. 107-112.
5. Bruches, E. P. Hybrid approach to removing part-of-speech homonymy using the example of the Russian language / E. P. Bruches // *Proceedings of the 54th International Scientific Student Conference ISSC-2016: Applied Linguistics*, Novosibirsk, April 16–20, 2016. - Novosibirsk: Novosibirsk National Research State University, 2016. - P. 13-14.
6. Khoroshilov A. A. Contextual homonymy resolution based on the centroid-context model / A.A. Khoroshilov, Yu.V. Nikitin, A.V. Kan, Ya.D. Kozlovskaya, E.A. Evdokimova // *Proceedings of the ISP RAS*. - 2022. - Vol. 34, issue. 5. - P. 171-182.
7. Kanevsky, E. A. On the issue of removing homonymy in computer analysis of texts / E. A. Kanevsky, K.K. Boyarsky // *Problems of transformation and regulation of regional socio-economic systems* / Edited by S.V. Kuznetsov. Issue. 45. - St. Petersburg: St. Petersburg State University of Aerospace Instrumentation, 2019. - P. 37-41.

8. Zaripova, D. A. Approaches to automatic resolution of polysemy based on the uneven distribution of word meanings in a corpus / D. A. Zaripova, N. V. Lukashevich // Bulletin of Moscow University. Series 9: Philology. - 2023. - No. 6. - P. 40-51.
9. Rykunov, A. N. Development of a method for processing non-dictionary words to improve the accuracy of morphological analysis of texts in Russian / A. N. Rykunov, E. V. Politsyna, S. A. Politsyn, A. S. Porechny // Bulletin of Voronezh State University. Series: Systems Analysis and Information Technology. - 2024. - No. 1. - P. 166-177.
10. Kormanovsky, M. V. On the application of methods for analyzing texts in natural language with subsequent formation of graph structures / M. V. Kormanovsky, L. L. Volkova // Innovative, information and communication technologies: collected papers of the XXI International scientific and practical conference, Sochi, October 1-10, 2024. - Moscow: Association of graduates and employees of VVIA named after prof. Zhukovsky, 2024. - P. 88-91.
11. Krassa, S. I. Removing the ambiguity of keywords in the examination of media text / S. I. Krassa // Modern media text and forensic examination: interdisciplinary connections and expert assessment: Collection of scientific papers following the results of the International scientific and practical conference, Moscow, October 12-13, 2023. - Moscow: OOO "SOYUZKNIG", 2023. - P. 153-160.
12. Stolyarov, I. I. Automatic distinction of homographs (based on the NKRYA material) / I. I. Stolyarov, O. A. Mitrofanova // International scientific conference dedicated to the 20th anniversary of the National Corpus of the Russian language: Conference materials, Moscow, December 20-21, 2024. - Moscow: V.V. Vinogradov Institute of the Russian Language, Russian Academy of Sciences, 2024. - P. 163-165.
13. Kochetkova, N. A. Method of forming a verb control model for the Russian language / N. A. Kochetkova // New information technologies in automated systems. - 2012. - No. 15. - P. 199-202.
14. Bolshakova, S. A. On the removal of the "predicative-prepositional group" homonymy for some common phrases in Russian-language texts / S. A. Bolshakova // Problems of Artificial Intelligence. - 2023. - No. 1 (28). - P. 11-17.
15. Hagen, M. Complete paradigm. Morphology [Electronic resource] // Forum "Speak Russian" [website]. 2018. Access mode: <http://www.speakrus.ru/dict/#morph-paradigm> (date of access: 10.06.2019).
16. Nitsenko, A. V. On the genitive case of a noun after a verb in the case of its homonymy with the accusative case / A. V. Nitsenko, V. Yu. Shelepov // Problems of Artificial Intelligence. - 2025. - No. 1 (36). - P. 4-11.
17. Nitsenko, A. V. On the use of semantic information to remove homonymy between the nominative and accusative cases (as an element of creating ontology) / A. V. Nitsenko, V. Yu. Shelepov // Problems of Artificial Intelligence. - 2024. - No. 4 (35). - P. 16-24.
18. National Corpus of the Russian Language [Electronic resource]. - Access mode: URL: <http://www.ruscorpora.ru/> (date of access: 19.03.2025).

RESUME

A.V. Nitsenko, V. Yu. Shelepov

A method of accusative and instrumental cases disambiguation for a number of Russian nouns

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.

Methods for automatic removal of homonymy between the accusative and instrumental cases of some nouns in combination with verb forms are described.

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

Метод снятия неоднозначности винительного и творительного падежа для ряда существительных 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.

Описаны методы автоматического снятия омонимии винительного и творительного падежа некоторых существительных в сочетаниях с глагольными формами.

В статье приведено описание предложенных методов и 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.

Ниценко Артём Владимирович – к.т.н., старший научный сотрудник отдела распознавания речевых образов, Федеральное государственное бюджетное научное учреждение «Институт проблем искусственного интеллекта», г. Донецк. *Область научных интересов:* искусственный интеллект, обработка естественного языка, компьютерная лингвистика. Эл. почта: pav_box@mail.ru, адрес: 283048, г. Донецк, ул. Артема, д. 118 б, телефон: +7949 4315356.

Шелепов Владислав Юрьевич – д.ф.-м.н., профессор, главный научный сотрудник отдела распознавания речевых образов, Федеральное государственное бюджетное научное учреждение «Институт проблем искусственного интеллекта», г. Донецк. *Область научных интересов:* искусственный интеллект, обработка естественного языка, компьютерная лингвистика. Эл. почта: vladislav.shelepov2012@yandex.ru, адрес: 283048, г. Донецк, ул. Артема, д. 118 б, телефон: +7949 4219188.

Статья поступила в редакцию 13.06.2025.