

УДК 004.912

А. В. Ниценко, В. Ю. Шелепов, С. А. Большакова, К. С. Ивашко
Государственное учреждение «Институт проблем искусственного интеллекта», г. Донецк
83048, г. Донецк, ул. Артема, 118-б

О СЛОВЕСНЫХ ЗАМЕНАХ, СОХРАНЯЮЩИХ СМЫСЛ РУССКОГО ПРЕДЛОЖЕНИЯ

A. V. Nicenko, V. Ju. Shelepov, S. A. Bolshakova, K. S. Ivashko
Public Institution «Institute of Problems of Artificial Intelligence»
83048, Donetsk, str. Artema, 118-b.

ABOUT WORD-REPLACEMENTS CONSERVING THE SENSE OF THE RUSSIAN SENTENCE

А. В. Ніценко, В. Ю. Шелепов, С. А. Большакова, К. С. Івашко
Державна установа «Інститут проблем штучного інтелекту», м. Донецьк
83048, м. Донецьк, вул. Артема, 118-б.

ПРО ЗАМІНИ СЛІВ, ЩО ЗБЕРІГАЮТЬ ЗМІСТ РОСІЙСЬКОГО РЕЧЕННЯ

Статья посвящена проблеме автоматического выполнения в русскоязычном предложении синонимических замен, сохраняющих смысл и правильный синтаксис. Метод использует специальную базу синонимов, в основе которой лежит понятие синонимического ряда, начинающегося заглавным словом – доминантой (см. [13]). Член синонимического ряда, встретившийся в тексте, заменяется соответствующей доминантой, что является элементом адаптации текста. Чтобы автоматическая замена не приводила к нелепому результату, смыслы, выражаемые недоминантным членом синонимического ряда, не должны выходить за пределы смыслов, выражаемых доминантой. Проведен связанный с этим анализ синонимических рядов, а также максимизация доминанты по частоте употребления. Для обеспечения в полученном предложении правильного синтаксиса в базе синонимов вводятся специальные обозначения.

Ключевые слова: адаптация текста, база синонимических рядов, доминанта, синтаксис, пометы в базе.

The article is devoted to the problem of automatic synonymic replacements in the Russian sentence. They must preserve the meaning and the correct syntax. The method employs a special synonym database, based on the concept of a synonymic series (see [13]). A member of the synonymic series that occurs in the text is replaced by the corresponding dominant. This is an element of adaptation of the text. In order for the replacement not to lead to an absurd result, the meanings expressed by the non-dominant member of the synonymic series should not go beyond the meanings expressed by the dominant. In this regard, an analysis of synonymous series and also maximization of the dominant in frequency of use were performed. To ensure the correct syntax in the received sentence, special notations were introduced in the synonyms database.

Keywords: text adaptation, synonym database, dominant, syntax, litters in the database.

Статтю присвячено проблемі автоматичного виконання в російськомовному реченні синонімічних заміни, які зберігають зміст і правильний синтаксис. Метод використовує спеціальну базу синонімів, в основі якої лежить поняття синонімічного ряду, який починається з головного слова - домінанти (див. [13]). Член синонімічного ряду, що зустрівся в тексті, замінюється відповідною домінантою, що є елементом адаптації тексту. Щоб автоматична заміна не приводила до безглузлого результату, зміст, що виражений недомінантним членом синонімічного ряду, не повинен виходити за межі змісту, що виражений домінантою. Проведено пов'язаний з цим аналіз синонімічних рядів, а також максимізація домінанти за частотою вживання. Для забезпечення в отриманому реченні правильного синтаксису в базі синонімів вводяться спеціальні позначення.

Ключові слова: адаптація тексту, база синонімічних рядів, домінанта, синтаксис, позначення в базі.

Введение

Проблеме адаптации (упрощения), в том числе автоматической адаптации, русскоязычных текстов посвящены работы [1-10].

Среди элементов адаптации, выполняемой человеком, могут быть выделены следующие:

1. Отказ от предложений или целых абзацев, которые могут быть опущены без существенных потерь для общего содержания текста.
2. Отказ от отдельных слов в предложениях на тех же условиях.
3. Замена предложений синтаксически более простыми.
4. Лексическое упрощение, а именно, замена отдельных слов и словосочетаний их более общими или более употребительными синонимами.

Настоящая статья посвящена автоматической реализации процедур, связанных с последним пунктом приведенного списка. Работа выполнена и программно реализована авторами.

Как и в работе [11], мы используем словарь [12] русских словоформ объемом более 4 миллионов единиц. В *txt*-формате он представляет собой совокупность строк – словоформ с сопровождающей грамматической информацией, собранных в парадигмы, каждая из которых начинается леммой. Леммы расположены по алфавиту. Парадигмы разделены пробельными строками. Представление словаря в виде дерева позволяет осуществлять в нем практически мгновенный поиск любой словоформы. Кроме того, его формат позволяет также быстро получать соответствующую лемму: достаточно сместиться вверх до ближайшей строки, следующей за пробельной строкой. Однако леммой для причастия договоримся считать не инфинитив соответствующего глагола, а форму этого причастия в именительном падеже единственного числа мужского рода. Тогда леммой для слова «искавшую» является слово «искавший», а леммой для «ищущим» является «ищущий». В этом смысле причастия трактуются как самостоятельные части речи – аналоги прилагательных. Это достаточно естественно ввиду обилия у них словоформ и наличия процедуры склонения. Леммой же для деепричастия служит инфинитив глагола.

Отметим также следующее. Парадигма прилагательного включает формы его превосходной степени. Например,

быстрый | прл ед муж им
быстрого | прл ед муж род
.....
быстрейший | прл прев ед муж им
быстрейшего | прл прев ед муж род
.....
наибыстрейший | прл прев ед муж им
наибыстрейшего | прл прев ед муж род

Для наших целей удобно считать леммой для какой-либо словоформы превосходной степени прилагательного форму превосходной степени именительного падежа, мужского рода, единственного числа. Она определяется путем движения по списку от лемматизируемой словоформы в направлении снизу вверх.

1 Лингвистическая основа

Лингвистической основой для разработки интересующей нас программы могут служить словари синонимов [13-18], а также частотные словари русского языка [19], [20]. Мы будем использовать «Словарь синонимов русского языка» З. Е. Александровой [13] и Электронную версию <http://dict.ruslang.ru/freq.php> издания: О. Н. Ляшевская, С. А. Шаров. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009 [20].

В основе организации словаря [13] лежит понятие синонимического ряда. Как указано в предшествующем словарию описании, «Синонимический ряд начинается заглавным словом (доминантой). За ним следуют синонимы – члены ряда. Доминанта должна быть стилистически нейтральной, семантически прозрачной и емкой, по сравнению с другими членами ряда иметь большую употребительность и главное – наиболее широкую сочетаемость, общую со всеми членами ряда. За неимением нейтрального слова, отвечающего этим требованиям, доминантой может оказаться книжное, разговорное или даже устарелое слово. Иногда доминантой является словосочетание (например, ЖЕЛЕЗНАЯ ДОРОГА, СТАРАЯ ДЕВА)».

Идея обсуждаемой компьютерной программы кратко формулируется так: член синонимического ряда, встретившийся в тексте, должен быть заменен соответствующей доминантой.

В соответствии с этим создается файл *База.txt*, представляющий собой последовательность групп, каждая из которых начинается соответствующей доминантой (отмечается в файле восклицательным знаком), а затем содержит отрезок синонимического ряда. Пример такой группы:

безграничность !
безбрежность
безмерность
бесконечность
бескрайность
беспредельность
необозримость
необъятность

Для удобства группы разделены пробельными строками. Вместе с каждым словом в базу можно было бы включить все его словоформы. Это увеличило бы ее на порядки, но, представив ее в виде дерева, можно было бы обеспечить в ней, как и в словаре [12], практически мгновенный поиск. Однако для того, чтобы база была легко обозримой с точки зрения семантики, мы включаем в нее только начальные формы слов и используем лемматизацию исходного предложения.

2 О программе в целом

Итак,

1. Программа лемматизирует предложение.
2. Затем она выделяет в нем слово, содержащееся в Базе.
3. Если оно не является доминантой, то программа заменяет его соответствующей доминантой.
4. Наконец, она восстанавливает в полученном предложении правильный русский синтаксис.

Отметим сразу, что при создании программы автоматической замены нельзя использовать доминанты и синонимические ряды словаря [13] совершенно формально, ибо этот словарь создан как справочник для *людей*, пишущих на русском языке. Так для доминанты АВТОМОБИЛЬ словарь предлагает следующий синонимический ряд: автомобиль ! авто, автомашина, железный конь, железка, жестянка, кар, колеса, машина, мотор, тачка

Для человека, пользующегося словарем, это набор ценных подсказок, позволяющий найти и использовать стилистически подходящий синоним. Но для компьютерной программы, которая *всегда* будет заменять член ряда доминантой, слова *железка, жестянка, колеса, машина, мотор, тачка* неприемлемы, ибо в каких-то ситуациях они будут обозначать совсем не автомобиль.

Обратим внимание на то, что слово *железка* даже в том же словаре может иметь другую доминанту ЖЕЛЕЗНАЯ ДОРОГА. Для программы, использующей принцип автоматической замены члена синонимического ряда соответствующей доминантой, такая доминанта, очевидно, должна быть единственной. Поэтому мы, используя материал словаря [13], пока вынуждены ограничиться словами, для каждого из которых словарь предлагает единственную доминанту. Что же касается феномена только что проиллюстрированной семантической неоднозначности, то в связи с ней для наших целей проведен анализ и сокращение предлагаемых синонимических рядов. Это большая работа филологического характера. Для того, чтобы замена не приводила к нелепому результату, смыслы, выражаемые недоминантным членом синонимического ряда, не должны выходить за пределы смыслов, выражаемых доминантой. В то же время доминанта может обладать и другими смыслами.

Пример, когда замена допустима, хотя доминанта обладает и другими смыслами:

подтягивать !

подпевать

Здесь «подтягивать», помимо «подпевать», может использоваться иначе: «подтягивать гайку», «подтягивать штаны» и т.д. В частности, доминанта одного синонимического ряда может быть недоминантным членом другого синонимического ряда.

Кроме того, с помощью словаря [20] для каждой группы проанализирована частотность доминанты и членов синонимического ряда, и при возможности в качестве доминанты выбран синоним с наибольшей частотностью.

Итак, в данной работе мы ограничиваемся заменой одиночных слов, для каждого из которых словарь [13] дает единственную доминанту.

На рис. 1 представлено окно программы. Подлежащее обработке предложение вводится в поле 1 и лемматизируется. Полученные леммы ищутся в Базе. В поле 2 скобками выделены слова исходного предложения, леммы которых обнаружены в Базе и не являются доминантами. В поле 3 представлен результат замены этих слов на доминанты с последующим восстановлением правильного синтаксиса.

3 Первоначальная работа с синтаксисом

В работе [11], реализована автоматическая процедура определения грамматических признаков введенных словоформ путем поиска их в словаре [12]. Здесь она реализована как функция, вызываемая по кнопке «Таблица» (рис. 1), результат в поле 4.

Для примера, содержащегося в рис. 1, имеем:

они | мест сущ мн им
 купили | гл сов перех прош мн
 новую | прл ед жен вин !
 автомашину | сущ неод ед жен вин

В [11] описаны также правила создания подчинительного дерева, представленного в поле 5 на рис. 1. Сейчас это дерево нас интересовать не будет.

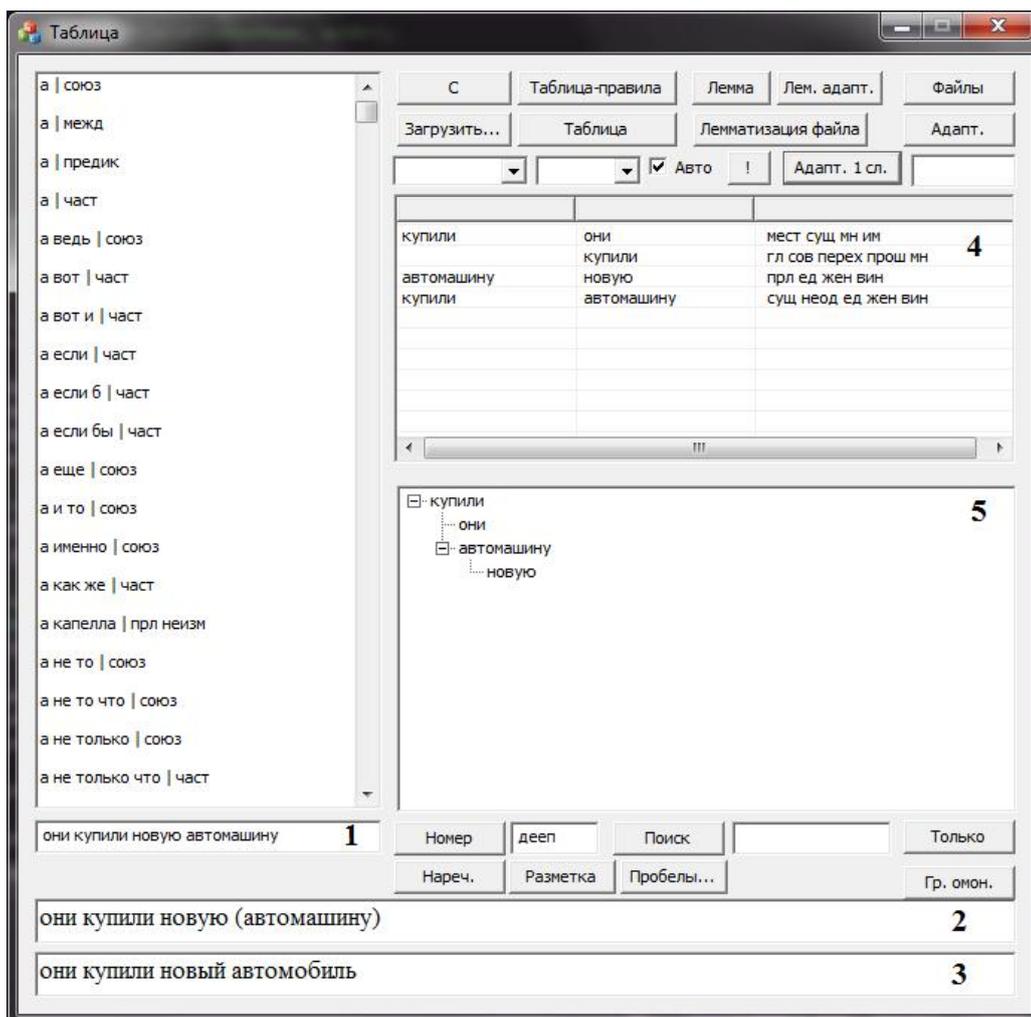


Рисунок 1 – Окно программы с примером замены в предложении

При построении результата описываемая программа первоначально сохранит все слова за пределами скобок в тех формах, в которых они были в исходном предложении. Если заменяемое и заменяющее слова имеют одинаковые грамматические характеристики, то заменяющее слово ставится в той же форме, в которой находилось заменяемое. Если это существительные, отличающиеся родом, то программа меняет форму относящегося к нему прилагательного, местоимения-прилагательного или причастия, согласуя его с новым существительным. То же делается по отношению к глаголу. Так, например, фраза АВТОМАШИНА ТРОНУЛАСЬ преобразуется во фразу АВТОМОБИЛЬ ТРОНУЛСЯ.

4 Об изменении синтаксического управления

Зачастую для получения в замененном предложении правильного синтаксиса требуется изменить синтаксическое управление. Для этого используются следующие обозначения, проставляемые при необходимости в базе справа от соответствующих членов синонимического ряда.

| – означает, что последующая запись относится к заменяемому слову;

/ – означает, что последующая запись относится к слову, которое непосредственно предшествует заменяемому;

\ – означает, что последующая запись относится к слову, которое непосредственно следует за заменяемым;

Остальное разъясним на конкретном примере. Запись

\ (с) тв-[на] вин

означает, что, если существительное, стоит в творительном падеже с предлогом «с» следом за заменяемым словом, то за заменяющим словом оно должно следовать в винительном падеже с предлогом «на». Отдельные элементы в этой записи могут отсутствовать. Например, запись

\ (с) тв-дат

означает, что творительный падеж с предлогом «с» должен быть заменен на винительный падеж без предлога. Пример: «Раскланялся с соседями» программа превратит в «Поклонился соседям».

Конечно, предлоги и падежи могут быть другими. Полный список помет, относящихся к синтаксическому управлению, приведен в работе [25].

5 Другие пометки в базе

1. Пометка «1» следом за доминантой означает, что при замене доминанта используется только в единственном числе. Пример:

поведение ! 1

поступок

«Он озадачил меня своими поступками» преобразуется в «Он озадачил меня своим поведением».

2. Пометка «2» следом за доминантой означает, что при замене доминанта используется только во множественном числе. Пример:

аплодисменты ! 2

орация

рукоплескание

«Концерт закончился орацией» преобразуется в «Концерт закончился аплодисментами».

3. Пометка «2» следом за членом синонимического ряда означает, что этот член заменяется доминантой только тогда, когда стоит во множественном числе. Пример:

артиллерия ! 1

орудие 2

пушка 2

«Орудия стреляют» преобразуется в «Артиллерия стреляет». «Орудие стреляет» программа оставляет без изменения.

4. Пометка /форма/ около члена синонимического ряда означает, что в процессе обработки слова оно не подвергается лемматизации. Пример:

способен !

горазд /форма/

«Он горазд плясать» преобразуется в «Он способен плясать».

5. Пометка /форма,/ (с запятой) около члена синонимического ряда означает, что в процессе обработки слова оно не подвергается лемматизации лишь в случае, когда после него стоит запятая или точка. Пример:

нет!

дудки /форма,/

«Дудки, ничего не выйдет!» преобразуется в «Нет, ничего не выйдет!».

6. Пометка {в-во} означает, что при замене предлог «в» заменяется предлогом «во». Пример:

время ! 2 {в-во}

година

«В году испытаний» преобразуется в «Во времена испытаний».

7. Пометка /сущ/ означает, что замена происходит только в том случае, когда член синонимического ряда определен в предложении функцией «Таблица» (см. п. 4) как существительное. Пример:

бюллетень ! /сущ/

больничный

Если бы пометы не было, то программа преобразовала бы предложение «Прошли больничный коридор» в «Прошли бюллетень коридор».

6 О словах в функциях наречия и существительного

Определенная проблема возникает, если группа, в которой доминанта является наречием, содержит слово, которое может выступать как в функции наречия, так и в функции существительного. Пример:

отдельно !

изолированно

особняком

В предложении «Он стоял особняком» слово «особняком» – наречие и должно заменяться словом «отдельно». В предложении же «Он обзавелся особняком» это слово выступает в функции существительного и указанной замены быть не должно. Если слово может выступать как в функции наречия, так и в функции существительного, то, как правило, последнее стоит в творительном падеже. Если ему предшествует соответствующий предлог, то слово заведомо является существительным. Вот список таких предлогов:

в комплекте с, в лад с, в погоне за, в ряду с, в связи с, в согласии с, в сообществе с, в соответствии с, в сопоставлении с, в соприкосновении с, в сочетании с, в сравнении с, в унисон с, в уровень с, не считаясь с, по аналогии с, вдогон за, вдогонку за, обок с, вкуче с, вместе с, вплотную с, вразрез с, вровень с, врозь с, вслед за, наедине с, наравне с,

наряду с, невысоко над, несообразно с, по сравнению, рядом с, рядом с, следом за, совместно с, совокупно с, согласно с, созвучно с, сообразно с, сообща с, соответственно с, соразмерно с, сравнительно с, кончая, над, надо, начиная, перед, передо, по-за, по-над, по-под, пред, предо.

В остальных случаях для ответа на вопрос программа ориентируется на содержащийся в предложении глагол. Для этого создается файл *Глаг-нар.txt*, который содержит последовательность групп вида:

авансом # (нар), брать, взыскать, внести, вносить, выдавать, выдать, выплатить, выполнить, делать, заплатить, засчитывать, начислять, платить, получать, получить. (1)

Если глагол содержится в группе вида (1), то первое слово (помеченное значком #) программа однозначно интерпретирует как наречие и заменяет соответствующей доминантой (в данном примере это «заранее»). Если глагол не содержится в группе вида (1), то нет гарантии, что указанное слово должно однозначно интерпретироваться как наречие, и поэтому замене оно не подлежит.

Для поиска нужных глаголов использовался Национальный Корпус русского языка [21], а также источники [22-25]. Полностью содержание файла *Глаг-нар.txt*, а также список глаголов, обеспечивающих однозначную интерпретацию слова как существительного, будут опубликованы в работе [26].

7 О леммах в базе

Подчеркнем в заключение, что замена синонима осуществляется только тогда, когда он, не являясь доминантой, стоит в группе в начальной форме. Например, имея в Базе группу

десерт !
пирожное
сладкое (2)

программа преобразует предложение «Подадим сладкое» в «Подадим десерт», а предложение «Подадим сладкое печенье» оставит без изменения. Причина в том, что в первом случае функция «Таблица» (см. п. 4) идентифицирует «сладкое» как существительное, для которого в группе есть лемма (2), а во втором – как словоформу прилагательного «сладкий», для которого в группе леммы нет.

Заключение

Проведенные эксперименты показали, что разработанная база синонимов позволяет успешно осуществлять замену отдельных слов в предложении их синонимами с сохранением правильного синтаксиса и семантики предложения. Всего база синонимов насчитывает 26 070 слов, содержащихся в 7 381 синонимических рядах. Полученные результаты могут использоваться при автоматическом анализе и обработке текста для задач адаптации, поисковой оптимизации и автоматического реферирования текстов.

Список литературы

1. Karpov N. Corpus-Based Text Retrieval and Adaptation for Learning System [Текст] / N. Karpov // International Conference on Advances in Computing and Information Technology (ACIT 2014). – 2014. – Pp. 60–65.

2. Karpov N. Single-sentence Readability Prediction in Russian / N. Karpov, F. Vitugin, J. Baranova // 3rd International Conference on Analysis of Images, Social networks, and Texts. – 2014. – ch. 436. – мpp. 91–100.
3. Karpov N. Towards Automatic Text Adaptation In Russian [Текст] / N. Karpov, V. Sibirtseva // Higher School of Economics Research Paper No. WP BRP. – 2014. – Т. 16.
4. Karpov N. Development of modern electronic textbook of Russian as a foreign language: content and technology [Текст] / N. Karpov, V. Sibirtseva // Series WP «Working Papers of Humanities». – 2012. – No. 2012-6.
5. Сравнение трёх систем семантической близости для русского языка [Текст] / Н. В. Арефьев, А. И. Панченко, А. В. Луканин и др. // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции Диалог (Москва, 27-30 мая 2015 г.). – Вып. 14 (21). – Т. 2. – Москва : РГГУ, 2015. – С. 116–128.
6. Баранова Ю. Н. Создание вспомогательного информационного ресурса для анализа учебных текстов на русском языке [Текст] / Ю. Н. Баранова, Т. С. Елипашева // Человек в информационном пространстве. – Ярославль : ЯГПУ, 2014. – С. 232–246.
7. Леонтьева Н. Н. О методах смысловой компрессии текста [Текст] / Н. Н. Леонтьева // Труды Всероссийской объединенной конференции «Информационное общество – Интернет и современное общество». – СПб, 2007. – С. 269–273.
8. Сибирцева В. Г. Автоматическая адаптация текстов для электронных учебников [Текст] / В. Г. Сибирцева, Н. В. Карпов // Новая русистика. – 2014. – № 7. – С. 19–33.
9. Сибирцева В. Г. Национальный корпус русского языка как основа новаторских электронных учебников [Текст] / В. Г. Сибирцева, А. Ю. Хоменко, Ю. Н. Баранова // Образовательные технологии и общество. – 2013. – Т. 16, № 3. – С. 508–520.
10. Первухина С. В. Адаптированный художественный текст: способы повышения понятности [Текст] / С. В. Первухина // Вестник Челябинского государственного университета. – 2011. – № 25. – С. 130–134.
11. Ниценко А. В. Об автоматическом построении дерева синтаксического подчинения [Текст] / А. В. Ниценко, В. Ю. Шелепов // XII Мультиконференция по проблемам управления (МКПУ-2019): Материалы XII мультиконференции (г. Геленджик, 23 – 28 сентября 2019 г.). – Т. 1. Ростов н/Д.: ЮФУ, 2019. – С. 119–121.
12. Хаген М. Полная парадигма. Морфология [Электронный ресурс]. – URL : <http://www.speakrus.ru/dict/#morph-paradigm> (дата обращения: 10.02.2020).
13. Александрова З. Е. Словарь синонимов русского языка: Практический справочник [Текст] / З. Е. Александрова. – М. : Рус. яз., 2001. – 568 с.
14. Абрамов Н. Словарь русских синонимов и сходных по смыслу выражений [Текст] / Н. Абрамов – М. : Русские словари. 1999. – 433 с.
15. Словарь синонимов русского языка [Текст] / Л. Г. Бабенко, Ю. В. Казарин, Н. А. Дьячкова и др. – М. : Аст-Пресс, 2008. – 688 с.
16. Ушакова О. Д. Синонимы и антонимы: Словарик Школьника [Текст] / О. Д. Ушакова. – СПб. : Литера, 2008. – 96 с.
17. Гаврилова А. С. Словарь синонимов и антонимов современного русского языка [Текст] / А. С. Гаврилова. – М. : «Аделант», 2014. – 800 с.
18. Жуков В. П. Словарь фразеологических синонимов русского языка: около 730 синоним. рядов [Текст] / В. П. Жуков, М. И. Сидоренко, В. Т. Шкляров. – М : Рус. Яз., 1987. – 448 с.
19. Засорина Л. Н. (ред.). Частотный словарь русского языка: Около 40 000 слов [Текст] / Л. Н. Засорина. – М. : Русский язык, 1977. – 936 с.
20. 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.
21. Национальный корпус русского языка [Электронный ресурс] – URL: <http://ruscorpora.ru/new/index.html>. (дата обращения: 10.02.2020).
22. Карта слов и выражений русского языка. Онлайн-тезаурус с возможностью поиска ассоциаций, синонимов, контекстных связей и примеров предложений к словам и выражениям русского языка [Электронный ресурс]. – URL : <https://kartaslov.ru/> (дата обращения: 10.02.2020).
23. Фундаментальная электронная библиотека «Русская литература и фольклор» (ФЭБ) [Электронный ресурс]. – URL: <http://feb-web.ru/feb/mas/MAS-abc/23/ma464911.htm> (дата обращения: 10.02.2020).
24. Словари и энциклопедии на Академики. [Электронный ресурс]. – URL : <https://dic.academic.ru/> (дата обращения: 10.02.2020).

25. Русский Викисловарь — многофункциональный многоязычный словарь и тезаурус [Электронный ресурс]. – URL : <https://ru.wiktionary.org/wiki> (дата обращения: 10.02.2020).
26. Русское синтаксическое управление при словесных заменах. Слова с функциями наречия и существительного. [Текст] / А. В. Ниценко, В. Ю. Шелепов, С. А. Большакова, К.С. Ивашко // Проблемы искусственного интеллекта. – 2020. – № 2(17).

References

1. Karpov N. Corpus-Based Text Retrieval and Adaptation for Learning System. *International Conference on Advances in Computing and Information Technology (ACIT 2014)*. 2014. pp. 60-65.
2. Karpov N., Vitugin F., Baranova J. Single-sentence Readability Prediction in Russian. 3rd *International Conference on Analysis of Images, Social networks, and Texts*. 2014. ch. 436. pp. 91-100.
3. Karpov N., Sibirtseva V. Towards Automatic Text Adaptation In Russian. *Higher School of Economics Research Paper No. WP BRP*. 2014. Vol. 16.
4. Karpov N., Sibirtseva V. Development of modern electronic textbook of Russian as a foreign language: content and technology. *Series WP "Working Papers of Humanities"*. 2012. No. 2012-6.
5. N.V. Arefev, A.I. Panchenko, A.V. Lukanin. [Comparison of three semantic proximity systems for the Russian language]. *Komp'yuternaja lingvistika i intellektual'nye tehnologii: Po materialam ezhegodnoj Mezhdunarodnoj konferencii "Dialog"* [Computational Linguistics and Intelligent Technologies: Proceedings of the annual International Conference "Dialogue"]. Iss. 14 (21). Vol. 2. Moscow: RGGU. 2015. pp. 116-128. (In Russ.).
6. Baranova Ju.N., Elipasheva T.S. [Creation of an auxiliary information resource for the analysis of educational texts in Russian]. *Chelovek v informacionnom prostranstve* [Man in the information space]. Jaroslavl': JaGPU. 2014. pp. 232-246. (In Russ.).
7. Leont'eva N.N. [About methods of semantic text compression]. *Trudy Vserossijskoj ob'edinennoj konferencii "Informacionnoe obshhestvo – Internet i sovremennoe obshhestvo"* [Proceedings of the All-Russian Joint Conference "Information Society - Internet and Modern Society"]. SPb. 2007. pp. 269-273. (In Russ.).
8. Sibirceva V.G., Karpov N.V. [Automatic texts adaptation for electronic textbooks]. *Novaja rusistika - New Russian Studies*. 2014. No. 7. pp. 19-33. (In Russ.).
9. Sibirceva V. G., Homenko A. Ju., Baranova Ju. N. [The national corpus of the Russian language as the basis for innovative electronic textbooks]. *Obrazovatel'nye tehnologii i obshhestvo - Educational Technology and Society*. 2013. Vol. 16. No. 3. pp. 508-520. (In Russ.).
10. Pervuhina S.V. [Adapted artistic text: ways to improve clarity]. *Vestnik Cheljabinskogo gosudarstvennogo universiteta - Bulletin of Chelyabinsk State University*. 2011. №25. pp. 130-134. (In Russ.).
11. Nicenko A.V., Shelepov V.Ju. [About automatically building a syntax dependency tree]. *XII Mul'ti-konferencija po problemam upravlenija (MKPU-2019): Materialy XII mul'tikonferencii* [XII Multi-Conference on Management Problems (MKPU-2019): Proceedings of the XII Multi-Conference]. Vol. 1. Rostov-On-Don:JuFU. 2019. pp. 119–121. (In Russ.).
12. Hagen M. Polnaja paradigma. Morfologija [The complete paradigm. Morphology]. Available at: <http://www.speakrus.ru/dict/#morph-paradigm> (accessed : 10.02.2020). (In Russ.).
13. Aleksandrova Z.E. *Slovar' sinonimov russkogo jazyka: Prakticheskij spravocchnik* [Dictionary of Russian Synonyms: A Practical Guide]. Moscow: Russian Language, 2001. 568 p. (In Russ.).
14. Abramov N. *Slovar' russkih sinonimov i shodnyh po smyslu vyrazhenij* [Dictionary of Russian synonyms and similar expressions]. Moscow: Russian Dictionaries. 1999. 433 p. (In Russ.).
15. Babenko L.G., Kazarin Ju.V., D'jachkova N.A.. *Slovar' sinonimov russkogo jazyka* [Dictionary of Russian Synonyms]. Moscow: Ast-Press, 2008. 688 p. (In Russ.).
16. Ushakova O. D. *Sinonimy i antonimy: Slovarik shkol'nika* [Synonyms and Antonyms: Schoolboy Dictionary]. SPb.: Litera, 2008. 96 p. (In Russ.).
17. Gavrilova A.S. *Slovar' sinonimov i antonimov sovremennogo russkogo jazyka* [Dictionary of synonyms and antonyms of the modern Russian language]. Moscow:«Adelant», 2014. 800 p. (In Russ.).
18. Zhukov V. P., Sidorenko M.I., Shkljarov V.T. *Slovar' frazeologicheskikh sinonimov russkogo jazyka* [Dictionary of phraseological synonyms of the Russian language]. Moscow: Russian Language, 1987. 448 p. (In Russ.).
19. Zazorina L.N. (ed.). *Chastotnyy slovar' russkogo jazyka: Okolo 40 000 slov* [Frequency Dictionary of the Russian Language: About 40,000 words], М ., Russian language, 1977, 936 p.

20. Ljashevskaja O.N., Sharov S.A. *Chastotnyj slovar' sovremennogo russkogo jazyka (na materialah Nacional'nogo korpusa russkogo jazyka* [Dictionary of the Modern Russian Language (based on the materials of the Russian National Corpus)] Moscow: Azbukovnik, 2009. 1087 p.
21. Nacional'nyj korpus russkogo jazyka [Russian National Corpus]. Available at: <http://ruscorpora.ru/new/index.html>. (accessed: 10.02.2020).
22. *Karta slov i vyrazhenij russkogo jazyka. Onlajn-tezaurus s vozmozhnost'ju poiska asociacij, sinonimov, kontekstnyh svyazej i primerov predlozhenij k slovam i vyrazhenijam russkogo jazyka* [Map of words and expressions of the Russian language. Online thesaurus with the ability to search for associations, synonyms, contextual links and examples of sentences to words and expressions of the Russian language]. Available at: <https://kartaslov.ru/> (accessed: 10.02.2020).
23. *Fundamental'naja jelektronnaja biblioteka "Russkaja literatura i fol'klor"* [The fundamental electronic library "Russian literature and folklore"]. Available at: <http://feb-web.ru/feb/mas/MAS-abc/23/ma464911.htm> (accessed: 10.02.2020).
24. *Slovary i jenciklopedii na Akademike* [Dictionaries and Encyclopedias on the Academician]. Available at: <https://dic.academic.ru/> (accessed: 10.02.2020).
25. Russkij Vikislovar' — mnogofunkcional'nyj mnogojazychnyj slovar' i tezaurus [Russian Wiktionary - multifunctional multilingual dictionary and thesaurus]. Available at: <https://ru.wiktionary.org/wiki> (accessed: 10.02.2020).
26. Nicenko A.V., Shelepov V.Ju., Bol'shakova S.A., Ivashko K.S. [Russian syntactic management with verbal replacements. About words with the functions of an adverb and a noun]. *Problemy iskusstvennogo intellekta - Problems of artificial intelligence* 2020. No. 2(17). (In Russ.)

RESUME

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

About Word-Replacements Conserving the Sense of the Russian Sentence

When dealing with any natural language, we have to distinguish between text and its meaning. Firstly, the same thing can be said in different natural languages. But while remaining within the same language, obviously, it is necessary to distinguish between the meaning of the message and its specific implementation. One and the same can be said in different ways, in different words. The task of adaptation (simplification) of texts is connected with this. As a rule, people who study foreign languages deal with adapted versions of fiction and know how useful this is. The general task of teaching a computer to work with natural languages, obviously, should keep this aspect in mind. In particular, it is desirable to have programs that automatically adapt texts in Russian. The task of adaptation, obviously, is also associated with the task of abstracting - a summary of the content of the text.

The paper deals with the problem of automatic synonym replacement in the Russian sentence on the level of one-to-one word replacement. With such replacement, the new sentence is required to have the same meaning and to be syntactically correct. The method uses the special synonym dictionary, which is based on the concept of a synonymic series. The proposed replacement method consists in the fact that a member of the synonym series that occurs in the text must be replaced by the corresponding dominant. This allows you to provide some degree of adaptation of the text.

Algorithms and software modules have been developed that are provide lemmatization, one-to-one word synonym replacement, and restoration of correct syntax after replacement.

The experiments showed that the developed synonym vocabulary allows to successfully perform one-to-one word synonym replacement in a sentence while preserving the correct syntax and semantics of the sentence. The results can be used for automatic adaptation, and summarization of a text.

РЕЗЮМЕ

*А. В. Ниценко, В. Ю. Шелепов, С. А. Большакова, К. С. Ивашко
О словесных заменах, сохраняющих смысл русского предложения*

Имея дело с любым естественным языком, приходится различать текст и его содержание. Во-первых, одно и то же можно сказать на разных естественных языках. Но и оставаясь в пределах одного языка, очевидно, следует различать содержание сообщения и его конкретную реализацию. Одно и то же можно сказать по-разному, разными словами. С этим связана задача адаптации (упрощения) текстов. Как правило, люди, занимающиеся иностранными языками, имеют дело с адаптированными вариантами художественных книг на изучаемом языке и знают, насколько это полезно. Общая задача обучения компьютера работе с естественными языками, очевидно, должна иметь в виду и этот аспект. В частности, желательно иметь программы, автоматически адаптирующие тексты на русском языке. Задача адаптации, очевидно, связана также с задачей реферирования – краткого изложения содержания текста.

Статья посвящена проблеме автоматического выполнения синонимических замен отдельных слов в русскоязычном предложении. При выполнении таких замен необходимо, чтобы новое предложение сохраняло прежний смысл, и было синтаксически правильным. Метод использует специальную базу синонимов, в основе которой лежит понятие синонимического ряда. Предлагаемый способ замены состоит в том, что член синонимического ряда, встретившийся в тексте, должен быть заменен соответствующей доминантой. Это позволяет обеспечить некоторую степень адаптации текста с сохранением основного смысла предложения.

Разработаны алгоритмы и программные модули, обеспечивающие лемматизацию, замену слов синонимами и восстановление правильного синтаксиса после замены.

Проведенные эксперименты показали, что разработанная база синонимов позволяет успешно осуществлять замену отдельных слов в предложении их синонимами с сохранением правильного синтаксиса и семантики предложения. Полученные результаты могут использоваться для задач автоматической адаптации и реферирования текстов.

Статья поступила в редакцию 24.01.2020.