

На правах рукописи

Савина

САВИНА Ольга Юрьевна

**МОДЕЛЬ ГЕНЕРАЦИИ
ТЕКСТООРИЕНТИРОВАННОГО
ЛЕКСИЧЕСКОГО МИНИМУМА**

**Специальность 10.02.21 — Прикладная
и математическая лингвистика**

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата филологических наук

Тюмень
2017

Работа выполнена на кафедре французской филологии Института филологии и журналистики федерального государственного автономного образовательного учреждения высшего образования «Тюменский государственный университет».

Научный руководитель: **Лыкова Надежда Николаевна**,
доктор филологических наук, профессор,
профессор кафедры французской филологии
ФГАОУ ВО «Тюменский государственный
университет»

Официальные оппоненты: **Щипицина Лариса Юрьевна**,
доктор филологических наук, доцент,
профессор кафедры немецкой и французской
филологии ФГАОУ ВО «Северный
(Арктический) федеральный университет
имени М.В. Ломоносова» (г. Архангельск)

Софронова Татьяна Марковна,
кандидат филологических наук, доцент, доцент
кафедры английской филологии ФГБОУ ВО
«Красноярский государственный педагогический
университет им. В.П. Астафьева»

Ведущая организация: **ФГАОУ ВО «Южно-Уральский
государственный университет
(национальный исследовательский
университет)»**

Защита диссертации состоится 22 ноября 2017 года в 12.30 на заседании совета Д 212.274.15 по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук при ФГАОУ ВО «Тюменский государственный университет» по адресу: 625003, г. Тюмень, ул. Республики, 9, ауд. 211.

С диссертацией можно ознакомиться в Информационно-библиотечном центре ФГАОУ ВО «Тюменский государственный университет» по адресу: 625003, г. Тюмень, ул. Семакова, 18, а также на официальном сайте ТюмГУ: diss.utmn.ru.

Автореферат разослан «___» сентября 2017 г.

*Ученый секретарь
диссертационного совета Д 212.274.15
кандидат филологических наук, доцент*

 **Д.В. Шапочкин**

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Реферируемая диссертационная работа посвящена проблеме моделирования процесса генерации текстоориентированного лексического минимума, обладающего вариативностью и антропоцентрическим характером и построенного по лексикографическому принципу.

Выбор направления исследования обусловлен несколькими тенденциями развития современной лингвистики и компьютерной лингводидактики, подтверждающими его *актуальность*.

Так, для компьютерной лингводидактики важными являются вопросы определения и отбора лексического минимума в дидактических целях и необходимость составления лексических минимумов для разных этапов обучения и разных целевых групп. Современное распространение компьютерных технологий и развитие корпусной лингвистики позволяет по-новому взглянуть на вопросы отбора лексики в лингводидактических целях.

В области автоматизации лингвистических исследований существует потребность в разработке инструментальных средств интегративного характера, предназначенных для извлечения информации из лингвистических баз данных и способных усовершенствовать автоматизированное рабочее место лингвиста. В настоящее время компьютерный арсенал специалиста по лингвистике включает довольно много прикладных и инструментальных программ для создания дидактических материалов. Их целенаправленное использование позволяет значительно повысить эффективность труда преподавателя за счет автоматизации рутинных операций нетворческого характера. Однако множество задач по-прежнему решаются только в режиме ручного неэффективного труда. Настоящее исследование призвано заполнить одну из таких лакун.

Степень разработанности темы исследования определяется его теоретической базой. В области лингвистики текста учеными были определены понятие текста и его основные характеристики: И.Р. Гальперин (1981), О.И. Москальская (1981), З.Я. Тураева (1986), В.Б. Касевич (1988), В.Е. Чернявская (2014), Р. де Богранд и В.У. Дресслер (R. De Beaugrande, W.U. Dressler, 1981); в области лингвистического моделирования описаны понятие модели, их виды и этапы моделирования: Ю.Д. Апресян (1966), И.И. Ревзин (1967),

В.Д. Табанакова (2009) и др.; в области теоретической лексикографии сформулированы параметры описания и отбора лексических единиц, параметры словарей и структура словарной статьи: Ю.Н. Караулов (1981), В.Н. Крупнов (1987), Г.И. Кустова и Е.В. Падучева (1994), И.С. Кудашев (2007) и др.; в области компьютерной лексикографии представлены принципы использования лингвистических баз данных: Г.Г. Белоногов (2004), Л.Н. Беляева (2010, 2016) и др.; в области учебной лексикографии определено понятие лексического минимума и принципы его формирования, описаны особенности учебных лексикографических продуктов: Б.А. Лapidус (1986), Р. Ладó (1989), Л.И. Журомская (1998), А.А. Миролюбов (2002), В.В. Морковкин (2006), А.Н. Щукин (2008), Э.Г. Азимов (2009), Л.И. Кузнецова (2010), Е.И. Маркина (2011) и другие ученые.

Лингводидактические вопросы организации лексических единиц в учебных целях и специфики лексических минимумов для разных целевых групп были исследованы В.Л. Скалкиным (1979), Л.И. Журомской (1998), А.А. Миролюбовым (2002), Л.К. Раицкой (2009), К.А. Митрофановой (2009), А.Н. Шамовым (2011), А.Н. Щукиным (2011), Н.Д. Гальсковой и Н.И. Гез (2013), В.Э. Матвеевко (2013). По вопросам компьютерной лингводидактики основу исследования составили работы А.Р. Арутюнова, Н.Б. Музрукова и П.Г. Чеботарева (1981), М.А. Бовтенко (2005), А.Д. Гарцова (2007), О.В. Нагель (2008), Р.К. Потаповой (2002), А.В. Зубова (2012) и других ученых, посвященные использованию компьютерных баз данных в лингводидактических целях.

Теоретические положения в области автоматизированного рабочего места основываются на исследованиях Л.Н. Беляевой (2004, 2015), П.И. Образцова (2012), А.А. Виландеберк и О.В. Буевич (2016) и других.

По вопросам корпусной лингвистики базу составили исследования видов, основных характеристик и областей применения лингвистических корпусов В.П. Захарова (2003, 2005), Ю.Д. Апресяна, И.М. Богуславского и Б.Л. Иомдина (2005), А.Н. Баранова (2007), В.А. Плунгяна (2008), Е.Г. Соколовой, И.С. Кононенко и Ю.А. Загорюлько (2008), Р.К. Потаповой (2009), Ю.Н. Марчука (2007, 2009), Л.А. Нефедовой (2015), Р. Филера и П. Вагенера (R. Fiehler, P. Wagener, 2005), Р. Дуфнера, А. Камбера и А. Нэфа (R. Duffner, A. Kamber, A. Näf, 2009), Л. Лемницера и Х. Цинсмайстера

(L. Lemnitzer, H. Zinsmeister, 2010), а также работы по автоматическому анализу текста и экспериментальной лингвистике Г.Г. Белоногова (2004), Н.Н. Леонтьевой (2006), Е.А. Сидоровой (2008), В.Ф. Хорошевского (2009), К.И. Белоусова и Н.А. Блазновой (2011), Е.И. Большаковой (2011) и других исследователей.

При решении вопросов оценки качества программных продуктов были использованы работы И.В. Сегаловича (2002), А.И. Башмакова и И.А. Башмакова (2003), Т.М. Зубковой (2004), С.Ф. Сергеева, С.В. Синицина и Н.Ю. Налютинина (2006), В.В. Кулямина (2007), В.В. Головача (2010), К. Снайдер (C. Snyder, 2003), Я. Нильсена (J. Nielsen, 2014) и др.

Методология исследования определена с учетом рекомендаций Ю.С. Степанова (2005), В.О. Сафонова (2007), Н.И. Кузнецова (2009), К.И. Белоусова и Н.А. Блазновой (2011) и З.И. Комаровой (2012).

С учетом современного состояния указанных теоретических направлений были сформулированы основные параметры настоящего исследования.

Цель исследования — разработка модели генерации текстоориентированного лексического минимума, формируемого с помощью сетевых лингвистических информационных ресурсов, и создание на основе этой модели алгоритма автоматизированного отбора и упорядочения лексикографических параметров на материале немецкоязычных текстов.

Сформулированная таким образом цель исследования предполагает решение следующих **задач**:

- 1) определить основные параметры моделирования компонентов автоматизированного рабочего места лингвиста;
- 2) проанализировать существующие инструментальные программы по автоматизации операций по обработке текстов;
- 3) изучить существующие точки зрения по вопросу создания лексического минимума и предложить определение текстоориентированного лексического минимума;
- 4) представить структуру единицы лексического минимума, основанную на лексикографических параметрах, отражаемых в наиболее авторитетных немецкоязычных лингвистических сетевых информационных ресурсах;
- 5) разработать структурно-функциональную модель генерации лексического минимума и построить алгоритм ее функционирования;

6) разработать техническое задание по созданию прототипа инструментальной программы для генерации лексического минимума и провести тестирование прототипа;

7) проанализировать потенциал программы и проверить эффективность предлагаемого алгоритма путем сравнения результатов его функционирования с результатами, получаемыми при применении традиционных методов.

Объект диссертационного исследования — лексический минимум как совокупность лексических единиц, подлежащих усвоению за определенный промежуток учебного времени, и лексикографические параметры этих единиц.

Предмет исследования — процесс генерации лексического минимума, ориентированного на текст и имеющего антропоцентрический характер.

Гипотеза исследования заключается в том, что структурно-функциональная модель генерации лексического минимума, в основу которой положен лексикографический принцип, реализованная в виде алгоритма, позволит автоматизировать рутинные операции по формированию текстоориентированного лексического минимума, и, таким образом, расширить возможности автоматизированного рабочего места лингвиста и повысить эффективность лингводидактического и исследовательского труда.

Новизна исследования заключается в том, что

— впервые определено понятие текстоориентированного лексического минимума;

— впервые разработаны структурно-функциональная модель и алгоритм генерации текстоориентированного лексического минимума, использующие в качестве источников лексикографических параметров сетевые лингвистические информационные ресурсы;

— впервые предлагается ввести в автоматизированное рабочее место лингвиста инструмент для генерации текстоориентированного лексического минимума.

Теоретическая значимость исследования заключается в развитии положений прикладной лингвистики и компьютерной лингводидактики, в частности, в систематизации критериев отбора лексических единиц для лексического минимума, в разработке типизированной структуры лексического минимума, в определении специфики текстоориентированного лексического минимума. Отдельные положения ра-

боты могут применяться в соответствующих курсах по прикладной лингвистике в программах бакалавриата и магистратуры.

Практическая значимость работы состоит в том, что на основе предлагаемой модели разработан алгоритм и создан прототип инструментальной программы, позволяющей пользователю автоматизировать процесс генерации лексического минимума (на материале немецкого языка). Результаты исследования имеют прикладное значение для функционального совершенствования автоматизированного рабочего места лингвиста и способствуют повышению эффективности его труда.

Исследование проводится с применением комплексного подхода, предполагающего использование различных **методов**. На отдельных этапах настоящего исследования используются методы теоретического и эмпирического исследования, а также методы метатеоретического уровня.

Основу исследования составляет интуитивный прогностический **метод моделирования** процесса генерации лексического минимума. В процессе проведения исследования создается идеальная (основанная на мысленной аналогии ручного и автоматизированного процесса отбора лексикографических параметров) знаковая (схематическая) модель процесса генерации лексического минимума. Построение концептуальной модели подразумевает вербальное описание процесса формирования лексического минимума и приведение его к этапам модели. Конструирование формализованной модели включает схематическое представление процесса сбора лексикографических параметров. Моделирование процесса генерации лексического минимума позволяет алгоритмизировать его и создать прототип инструментальной программы. Кроме того, привлекаются отдельные методы прикладной лингвистики, алгоритмизации и некоторые методы оценки качества программного обеспечения.

На защиту выносятся следующие **положения**:

1. **Текстоориентированный** лексический минимум обладает вариативностью, так как отражает лексический состав конкретного текста, и имеет антропоцентрический характер, так как учитывает индивидуальные потребности обучающихся, которые определяют набор лексикографических параметров, включаемых в такой минимум.

2. Структурно-функциональная модель генерации текстоориентированного лексического минимума, элементами которой являются

текст, словоупотребления, лемматизатор, лексикографические параметры, характеризующие каждую лексическую единицу, и сетевые лингвистические информационные ресурсы, позволяет интегрировать разноуровневые лексикографические параметры лексических единиц.

3. Предлагаемая модель реализована в алгоритме, автоматизирующем процесс формирования текстоориентированного лексического минимума: поисковые запросы к сетевым ресурсам, сбор и упорядочение искомых лексикографических параметров к отбираемым из текста словоупотреблениям. В результате исполнения алгоритма генерируется текстоориентированный лексический минимум.

4. Разработанный алгоритм генерации текстоориентированного лексического минимума положен в основу прототипа инструментальной программы. Его верификация подтверждает жизнеспособность предложенной модели. Эффективность исследовательского и лингводидактического труда повышается за счет автоматизации рутинных процессов. Созданный инструмент может являться компонентом автоматизированного рабочего места лингвиста, способствующим оптимизации его деятельности.

Практическим материалом исследования послужили сетевые немецкоязычные лингвистические ресурсы: поисковая лексическая система «Wortschatz» объемом в 100 тысяч лемм, грамматическая сетевая база данных «Canoo.net» объемом в три миллиона словоформ, интернет-словарь современного немецкого языка «Digitales Wörterbuch der deutschen Sprache», основанный на корпусе объемом более полутора миллиардов словоупотреблений, онлайн-словарь «Duden», основанный на корпусе объемом более двух миллиардов словоформ, и онлайн-словарь «PONS». Их выбор обусловлен признанным авторитетом указанных баз данных и словарей.

Достоверность и научная обоснованность теоретических и практических результатов исследования обеспечивается:

— междисциплинарной методологической базой исследования (положения компьютерной лексикографии и корпусной лингвистики, а также методы разработки и оценки качества программного обеспечения);

— применением количественных методов анализа в экспериментальной части исследования;

— положительными результатами верификации разработанной модели.

Апробация результатов исследования осуществлялась в выступлениях на заседании кафедры французской филологии ТюмГУ (2013), на научно-методическом семинаре кафедры немецкой филологии ТюмГУ (2010, 2014), на международных конференциях «Естественный и виртуальный дискурс: когнитивные, категориальные и семиолингвистические аспекты», ТюмГУ (2009); Международном конгрессе учителей и преподавателей немецкого языка «Internationale Deutschlehrertagung», Боцен, Италия (2013).

Основные результаты и положения изложены в шести публикациях, из них четыре — в журналах, рекомендованных ВАК, одна — в зарубежном рецензированном сборнике материалов конференции. Разработанная программа зарегистрирована в Федеральной службе по интеллектуальной собственности (патент № 2015619379).

Диссертационная работа, общим объемом 151 страница, состоит из введения, трех глав, каждая из которых завершается выводами, заключения, списка принятых сокращений, списка терминов, библиографического списка, списка иллюстративного материала, а также четырех приложений.

Библиографический список включает 175 источников, из них 121 научная работа (в т.ч. 12 на иностранных языках), 30 программ и 13 баз данных.

Приложение содержит документы, отражающие процесс разработки и верификации модели, алгоритма и прототипа инструментальной программы, а также носитель (компакт-диск) с прототипом программы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во Введении обоснована актуальность исследования, сформулированы его цель и задачи, обозначены объект и предмет. Здесь же отражена новизна, теоретическая и практическая значимость работы. Кроме того, во Введении указаны использованные методы исследования, охарактеризован его материал.

Первая глава «Моделирование компонентов автоматизированного рабочего места» отражает современное состояние исследований в области автоматизированного рабочего места специалиста:

представлены основы моделирования компонентов автоматизированного рабочего места лингвиста и проведен анализ инструментальных программ по автоматизации операций по обработке текстов.

Специальным образом организованное рабочее пространство позволяет специалисту более эффективно решать специфические профессиональные задачи. Совокупность предназначенных для этого аппаратных и программных средств принято называть «автоматизированным рабочим местом» (АРМ). Согласно ГОСТу подобные программно-технические комплексы могут отличаться в зависимости от специальности. Следовательно, можно вести речь об АРМ переводчика, филолога, лингвиста и т.д. Анализ существующих программ по автоматизации операций по обработке текстов показал, что арсенал компьютерных инструментов постоянно пополняется. Наиболее разработанным сегодня можно считать АРМ переводчика. Для АРМ лингвиста разработано пока достаточно небольшое количество программ — многие насущные задачи, с которыми регулярно сталкиваются лингвисты (исследователи, преподаватели, студенты), по-прежнему могут быть решены только в режиме ручного неэффективного труда. Таким образом, подтвердилась потребность в разработке специализированной инструментальной программы.

В основе создания подобных программных средств лежит моделирование исследовательской деятельности лингвиста. При этом под моделью понимается концептуальная структура продукта в виде словесного описания и схемы. Для настоящего исследования важно отграничение модели и функционального прототипа как бета-версии программного продукта, воспроизводящей все внешние параметры объекта в реальном размере и выполняющей все функции.

Во **второй главе «Специфика текстоориентированного лексического минимума»** представлены релевантные для данного исследования теоретические направления лингвистики и лингводидактики: текст как лингводидактическая категория, понятие лексического минимума, его единицы и принципы его отбора; охарактеризованы сетевые лингвистические информационные ресурсы на немецком языке, определена структура типизированного лексического минимума.

Текст как объект лингвистического исследования уже давно находится в центре внимания представителей лингвистики текста (И.Р. Гальперин, З.Я. Тураева, Н.В. Васильева, В.А. Виноградов, О.И. Москальская, В.Е. Чернявская и другие). Исследуя текст как

лингвистическую единицу, ученые определяют его как осмысленную последовательность словесных знаков и описывают основные свойства текста: связность, цельность, а также невыводимость общего смысла из простой суммы значений составляющих. В то же время лингводидактический взгляд на текст как единицу обучения шире чисто лингвистического. Так, например, возможно понимание текста как любого факта культуры иностранного языка, имеющего знаковую функцию, что позволяет использовать в лингводидактическом процессе любые образцы иноязычной культуры.

Другой важной проблемной областью лингводидактики, положенной в основу данного исследования, является проблема определения лексического минимума и принципов его отбора. Под лексическим минимумом понимают лексические единицы, которые должны быть усвоены учащимися за определенный промежуток учебного времени. С учетом вида речевой деятельности выделяют активный (продуктивный) и пассивный (рецептивный) лексический минимум, границы между которыми условны. В то же время в лингводидактическом дискурсе не определено понятие *текстоориентированного лексического минимума*. В интересах нашего исследования мы определяем его как совокупность лексических единиц (слов, словосочетаний или предложений, имеющих характер лексемы), вычлененных из одного текста, предназначенного для использования в процессе обучения иностранному языку. Такой набор лексических единиц обязательно сопровождается лексикографическими параметрами, необходимыми для семантизации каждой лексической единицы и ее закрепления в ментальном лексико-семантическом поле изучающего иностранный язык. Единицей текстоориентированного лексического минимума должно стать одно из значений слова (т.е. слово в контексте), формально зафиксированное в лексическом минимуме через лемму (т.е. слово в системе).

Вопрос о том, какие именно лексические единицы включаются в состав лексического минимума, определяется принципами отбора, т.е. измерительными признаками, на основе которых принимается решение об их (не)включении в лексический минимум. Согласно устоявшейся классификации их можно разделить на три группы:

— лингвистические: сочетаемость, стилистическая неограниченность, семантическая ценность, словообразовательная ценность, многозначность, строевая способность;

— статистические: частотность, распространенность, употребительность, необходимость (наличность);

— методические: минимизация, репрезентативность и аутентичность лексического материала, коммуникативная ценность, принцип учета интересов учащихся, информативная ценность.

Преподаватель, готовящий текстоориентированный лексический минимум для конкретной целевой группы, может и должен по возможности учитывать эти принципы в процессе отбора лингвистического материала. Вместе с тем он должен учитывать потребности своей группы, а также особенности текста, выбранного для процесса обучения.

Таким образом, лексический минимум, ориентируясь на целевой уровень владения иностранным языком, должен обладать вариативностью, т.е. отбираться из конкретного текста, и содержать лексикографические параметры в объеме, необходимом конкретной группе обучающихся.

Источниками лексикографических параметров лексических единиц, включаемых в лексический минимум, в настоящее время могут служить сетевые лингвистические информационные ресурсы. Опираясь на авторитетные исследования в области компьютерной лингвистики, мы определяем лингвистические информационные ресурсы как совокупность упорядоченных языковых и речевых данных, находящихся на машинных носителях информации, позволяющих извлекать эмпирическую информацию и предназначенных для решения задач в различных сферах лингвистических и лингводидактических исследований. При этом сетевые информационные ресурсы являются наиболее доступными и актуальными.

К наиболее авторитетным немецкоязычным лингвистическим информационным ресурсам можно отнести: онлайн-словарь немецкого языка *Duden* издательства «Библиографический институт» (Берлин) (URL: <http://www.duden.de>), Толковый интернет-словарь современного немецкого языка *DWDS* Берлинско-Брандербургской Академии наук (URL: <http://www.dwds.de>), базу данных Поисковая лексическая система *Wortschatz* Университета Лейпцига (URL: <http://www.wortschatz.uni-leipzig.de>), грамматическую сетевую базу данных *Canoo.net* Университета Базеля (URL: <http://www.canoo.net>) и некоторые другие. Все они могут служить надежным источником сведений о лексических единицах немецкого языка. Каждая из

представленных баз данных имеет собственную структуру словарной статьи и содержит, как типичные, так и уникальные лексикографические параметры.

При составлении текстоориентированного лексического минимума может понадобиться интегрированная информация из разных баз данных (в зависимости от целевой группы). Сбор такой информации вручную — трудоемкий и неэффективный процесс. В рамках настоящего исследования предлагается автоматизировать этот процесс, что позволит существенно сэкономить время пользователя и высвободить его для более интеллектуальной работы.

Представление лексической единицы, интегрирующее разные типы лексикографических параметров, может иметь следующий вид (рис. 1).

1. Орфографические и фонетические особенности.
2. Грамматическая информация (данные о части речи, словоформы).
3. Семантическая информация (например, дефиниция).
4. Область тезауруса.
5. Парадигматические связи лексемы (синонимы, антонимы, гиперонимы, гипонимы).
6. Частотность употребления в современном немецком языке.
7. Стилистические данные (сферы употребления) и особые указания по употреблению лексической единицы.
8. Коллокации (дистрибуция).
9. Фразеология.
10. Словообразовательная информация (морфологическая структура, сложные и производные слова).
11. Примеры использования (в т.ч. аутентичные).
12. Перевод на родной язык учащихся.
13. Страноведческий комментарий (возможно, с иллюстрацией).
14. Этимология.
15. Принадлежность к уровню владения иностранным языком (по Общеввропейской шкале уровней).

Рис. 1. Лексикографические параметры лексической единицы в структуре лексического минимума

Третья глава «Моделирование процесса генерации текстоориентированного лексического минимума» посвящена практической части исследования. В ней описаны процесс разработки модели генерации текстоориентированного лексического минимума и алго-

ритма на основе предлагаемой модели, представлен функционал прототипа инструментальной программы, реализующей эту модель. Глава завершается описанием проверки эффективности разработанного алгоритма.

В процессе настоящего исследования была создана идеальная (основанная на мысленной аналогии ручного и автоматизированного процесса поиска лингвистической информации) знаковая (схематическая) модель процесса генерации (автоматизированного отбора и упорядочения) текстоориентированного лексического минимума на основании существующих лингвистических информационных ресурсов. При этом «автоматизированный отбор информации» подразумевает большую или меньшую степень ручной работы (в отличие от «автоматического отбора», когда программа выполняет его без участия человека). В основу предлагаемой структурно-функциональной модели генерации текстоориентированного лексического минимума положен лексикографический принцип — лексикографические параметры лексических единиц, предлагаемые сетевыми лингвистическими информационными ресурсами, отбираются с учетом потребностей конкретной целевой группы.

Концептуальная модель содержит вербальное описание процесса дидактизации текстов и приведение его к этапам модели. Традиционно процесс формирования текстоориентированного лексического минимума включает следующие шаги: 1) выбор лексем, подлежащих включению в лексический минимум, из текста для чтения; 2) отбор лингвистической информации по каждой из лексических единиц из разных баз данных; 3) приведение собранной информации в дидактически удобный вид. При автоматизации этого процесса пользователь может выполнять вручную только первый шаг: вводить текст, отмечать в нем лексические единицы, для которых должна быть собрана лингвистическая информация, и указывать необходимые типы лингвистической информации. Два следующих шага могут быть автоматизированы: программа выполняет запрос к одной / нескольким базам данных по каждой из лексических единиц, маркированных в исходном тексте, копирует отмеченные типы информации, соединяет их в единый документ и приводит в стандартный вид.

Конструирование формализованной модели включает схематическое представление процесса сбора лингвистической информации (рис. 2).

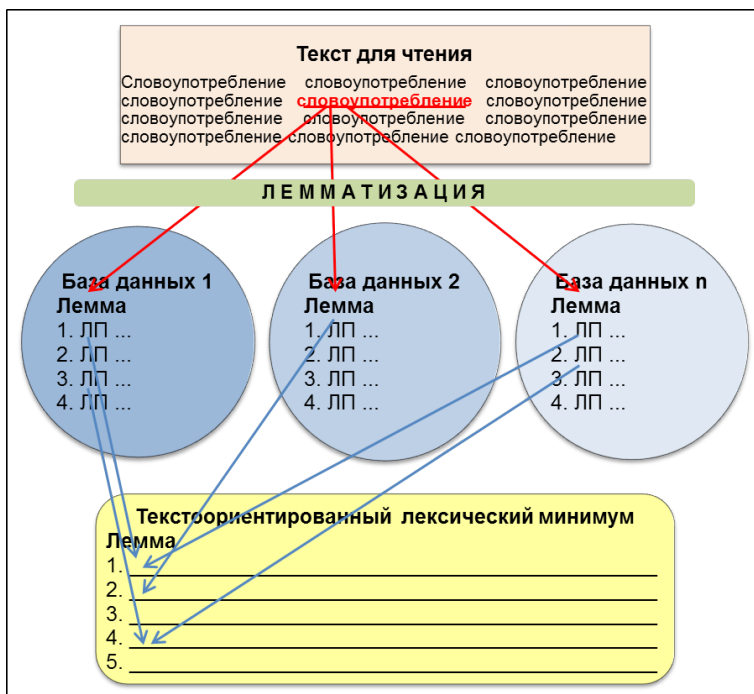


Рис. 2. Структурно-функциональная модель генерации текстоориентированного лексического минимума (ЛП — лексикографические параметры)

Таким образом, структурно-функциональная модель генерации текстоориентированного лексического минимума включает следующие элементы: текст, словоупотребления, лемматизатор, лексикографические параметры, характеризующие каждую лексическую единицу, и базы данных (сетевые лингвистические информационные ресурсы). Предлагаемая модель позволяет интегрировать разноуровневые лексикографические параметры лексических единиц.

На основе предложенной модели разработан алгоритм, позволяющий автоматизировать этапы лемматизации, поискового запроса и приведения к единообразному форматированию лексикографических параметров (рис. 3).

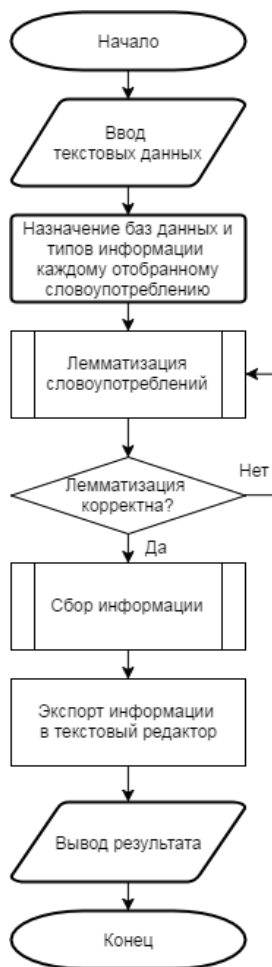


Рис. 3. Алгоритм генерации текстоориентированного лексического минимума

В результате исполнения алгоритма генерируется продукт — текстоориентированный лексический минимум, упорядоченный по лексикографическим параметрам.

Таким образом, разработанный алгоритм позволяет автоматизировать процесс формирования текстоориентирован-

ного лексического минимума: поисковые запросы к сетевым ресурсам, сбор и упорядочение искомым лексикографических параметров к отбираемым из текста словоупотреблениям.

Разработанный алгоритм реализован в виде прототипа инструментальной программы, выполняющей следующие функции: назначение лексикографических параметров к каждому отобранному пользователем словоупотреблению; поисковый запрос к базам данных; отбор и накопление лексикографических параметров; приведение полученных данных к унифицированному виду (рис. 4).

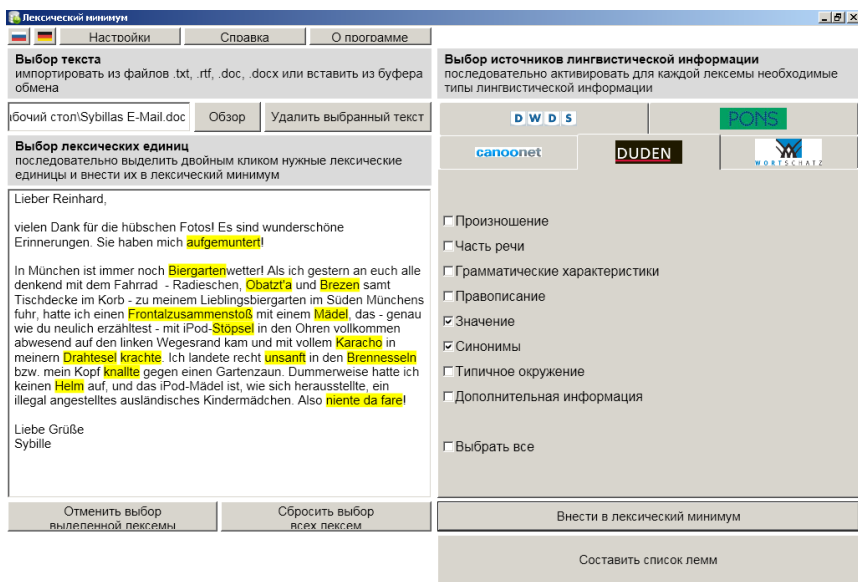


Рис. 4. Интерфейс программы генерации лексического минимума (русскоязычная версия)

Последовательность действий пользователя и программы представлена в таблице 1.

В готовом лексическом минимуме информация упорядочена по типу, указан источник информации (рис. 5).

Прототип программы прошел многоэтапную верификацию (в т.ч. тестирование на пользователях). Выявленные (непринципиальные) ошибки в функционировании были устранены.

Таблица 1

Последовательность действий пользователя и программы

Действия пользователя	Действия программы
1) запускает программу (файл <i>Vokabelliste.exe</i>);	
2) вводит в программу текст;	
3) выбирает из текста нужную лексему;	
4) активирует базы данных и необходимые типы лингвистической информации;	
5) повторяет шаги 3-4 для всех остальных нужных лексем;	
6) запускает процесс составления списка лемм;	
	7) лемматизирует отобранные лексемы и выводит список пользователю;
8) проверяет список лемм и при необходимости вносит в него изменения;	
9) подтверждает готовый список лемм и запускает процесс генерации лексического минимума;	
	10) последовательно направляет запрос по каждой лексеме к базам данных и получает запрошенные типы информации;
	11) копирует и объединяет собранную информацию;
	12) генерирует лексический минимум и приводит его в стандартный формат;
13) сохраняет готовый лексический минимум, выбрав удобный тип файла и место его расположения.	

Kindermädchen		
Orthographische und phonetische Besonderheiten Орфографические и фонетические особенности		
Aussprache Произношение DUDEN	Betonung: Kindermädchen Lautschrift: [ˈkɪndɛmɛːtʃən]	
Grammatische Informationen Грамматическая информация		
Wortformen Формы слова canoonet	Wortklasse:	Nomen
	Genitiv Singular:	Kindermädchens
	Nominativ Plural:	Kindermädchen
	Artikel:	das
	Flexionsklasse:	s/-
	Besonderheiten:	n-Tilgung im Dativ Plural
Semantische Informationen Семантическая информация		
Bedeutung Значение DUDEN	[jüngere] weibliche Person, die in einer Familie zur Betreuung der Kinder angestellt ist	
Bedeutung Значение D W D S	(veraltend) in einer Familie fest angestelltes Mädchen, das die Kinder betreut das Kindermädchen spielte mit den Kindern	
Paradigmatische Beziehungen Парадигматические связи		
Paradigmatische Beziehungen Парадигматические связи D W D S	Synonymgruppe: Kinderfrau, Kindermädchen, Nanny, Tagesmutter	
Synonyme Синонимы DUDEN	Amme, Kinderpflegerin, Nanny; (veraltend) Kinderfräulein, Nurse; (veraltet) Bonne; (früher) Gouvernante	
Kollokationen Коллокации		
Typische linke Nachbarn Типичное левое окружение 	als (317,55), ein (162,61), das (145,36), einem (95,86), dem (94,22), ehemaliges (92,11), philippinische (84,63), philippinisches (70,88), Das (63,98), stachelige (59,4), ehemalige (58,19), philippinischen (45,89), früheres (45,47), neues (43,79)	
Übersetzung ins Russische Перевод на русский язык		
Übersetzung ins Russische Перевод на русский язык DUDEN	няня	

Рис. 5. Пример результата поиска лингвистической информации к одной лемме (при запросе было активировано 8 опций в пяти базах данных)

Апробация показала, что данный инструмент генерации текстоориентированного лексического минимума может быть использован для решения лингводидактических задач, а также при проведении лингвистических исследований.

Ограничения разработанного алгоритма связаны с традиционными универсальными проблемами автоматического анализа текста — многозначностью, омонимией, фразеологией, а также с морфологическими особенностями немецкого языка — отделяемыми приставками, аналитическими формами глаголов и прилагательных.

Оценка эффективности программы проведена по количеству операций, выполняемых пользователем при поиске информации вручную и поиске, автоматизированном с помощью разработанного алгоритма.

В общем виде количество операций, выполняемых вручную, вычисляется по формуле (1):

$$Q_{OP} = 2 + 3Q_{БД} + (4Q_{ЛП} Q_{БД})Q_{СВ} \quad (1),$$

где Q_{OP} — количество операций при поиске вручную; $Q_{БД}$ — количество использованных баз данных; $Q_{ЛП}$ — количество искомых лексикографических параметров; $Q_{СВ}$ — количество искомых словоупотреблений.

Количество операций при автоматизированном поиске вычисляется по формуле (2):

$$Q_{OA} = 4 + 2Q_{СВ} + Q_{СВ} Q_{БД} + Q_{СВ} Q_{БД} Q_{ЛП} \quad (2),$$

где Q_{OA} — количество операций при автоматизированном поиске; $Q_{БД}$ — количество использованных баз данных; $Q_{ЛП}$ — количество искомых лексикографических параметров; $Q_{СВ}$ — количество искомых словоупотреблений.

Если принять количественное выражение результатов в единицах собранной информации (лексикографических параметрах), а затрат — в операциях, выполняемых пользователем, то можно рассчитать эффективность поиска вручную и автоматизированного поиска как соотношение полученных результатов и потребовавшихся трудозатрат.

Данные, полученные в экспериментальной части исследования, наглядно отражают возрастание эффективности автоматизированного поиска в сравнении с эффективностью поиска вручную. На рост эффективности напрямую влияет возрастание количества искомых параметров. Полученные результаты наглядно представлены на графике (рис. 6).

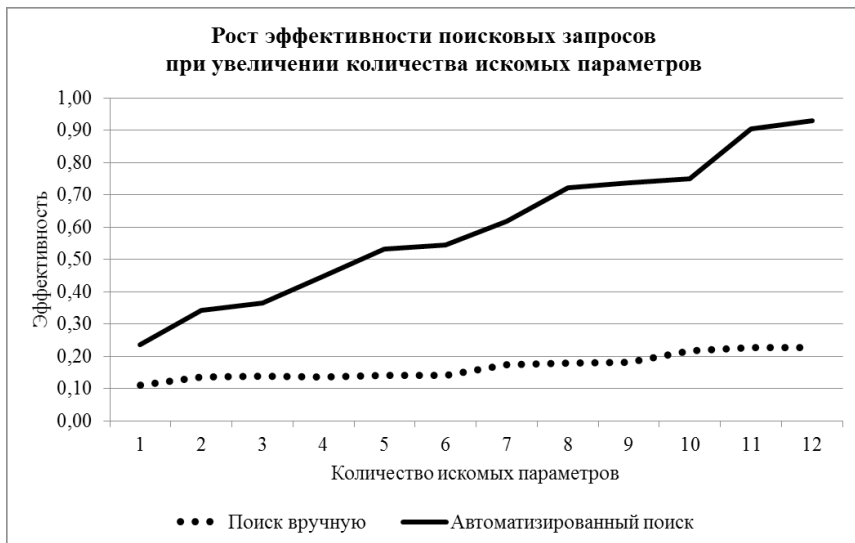


Рис. 6. Рост эффективности поисковых запросов при увеличении количества искомых параметров

Таким образом, верификация разработанного алгоритма подтверждает жизнеспособность предложенной модели. Эффективность исследовательского и лингводидактического труда повышается за счет автоматизации рутинных процессов. Это дает основания использовать созданный инструмент как компонент автоматизированного рабочего места лингвиста для оптимизации исследовательской и лингводидактической деятельности.

ЗАКЛЮЧЕНИЕ

В Заключении работы подводятся основные итоги исследования, а также формулируются перспективы его развития.

Итогом настоящего диссертационного исследования стала разработка модели генерации текстоориентированного лексического минимума, на основе которой построен алгоритм автоматизации отдельных этапов деятельности специалиста. Алгоритм реализован в виде прототипа инструментальной программы, предназначенной для использования в качестве компонента автоматизированного рабочего места лингвиста.

Текстоориентированный лексический минимум характеризуется вариативностью (т.е. ориентирован на лексическое наполнение конкретного текста) и антропоцентрическим характером (т.е. учитывает потребности конкретной группы обучающихся). Типизированная структура такого лексического минимума включает лексикографические параметры, доступные для автоматизированного отбора из сетевых информационных лингвистических ресурсов.

Структурно-функциональная модель генерации текстоориентированного лексического минимума позволяет интегрировать разноуровневые лексикографические параметры лексических единиц, извлекаемые из пяти сетевых лингвистических ресурсов (для немецкого языка): онлайн-словаря *Duden*; толкового интернет-словаря современного немецкого языка *DWDS*; поисковой лексической системы *Wortschatz*; грамматической сетевой базы данных *Canoo.net* и онлайн-словаря *PONS*.

На основе предложенной модели был разработан алгоритм, позволяющий автоматизировать поисковые запросы и упорядочение искомых лексикографических параметров, обеспечивая, таким образом, генерацию текстоориентированного лексического минимума.

Разработанный алгоритм положен в основу прототипа инструментальной программы. Проведенная верификация показала работоспособность программы и успешную реализацию основных запрограммированных функций. Программа успешно генерирует текстоориентированный лексический минимум. Кроме того, она может использоваться для сбора лексикографических параметров, необходимых для проведения лингвистических исследований срав-

нительного характера. Эффективность алгоритма (определяемая по трудоемкости поисковых запросов, выполняемых при поиске вручную и при автоматизированном поиске) подтверждена. Определяющим фактором повышения эффективности является количество искомым лексикографических параметров.

Предлагаемую модель можно использовать для интеграции лингвистической информации из многих других сетевых лингвистических ресурсов. Кроме того, на базе разработанной модели возможно создание подобных инструментов для других языков, способных пополнить арсенал автоматизированного рабочего места лингвиста.

Благодарности

Автор выражает глубокую благодарность своему первому научному руководителю профессору Татьяне Вячеславовне Сотниковой за внимательное отношение и поддержку на протяжении всех этапов исследования.

Автор искренне признателен научному руководителю профессору Надежде Николаевне Лыковой за мудрые советы и неоценимую помощь на этапе завершения исследования.

Автор благодарит многих коллег за важные комментарии и полезные замечания, полученные во время обсуждений различных материалов исследования.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в рецензируемых научных изданиях,
в которых должны быть опубликованы основные научные результа-
ты диссертаций на соискание ученой степени кандидата наук,
на соискание ученой степени доктора наук

1. Савина, О.Ю. О некоторых возможностях использования культурноспецифических лингвистических информационных ресурсов при решении экстралингвистических задач / О.Ю. Савина // Вестник ТюмГУ. Изд-во ТюмГУ. 2010. № 1. С. 168-175.

2. Савина, О.Ю. Методика формирования лексического минимума с помощью конкордансера / О.Ю. Савина // Вестник ТюмГУ. Изд-во ТюмГУ. 2016. № 2. С. 92-99.

3. Савина, О.Ю. Интеграция инструментов лингвистического поиска в лингводидактических целях / О.Ю. Савина // Вопросы лексикографии. 2016. № 2 (10). С. 55-66.

4. Савина, О.Ю. Моделирование прототипа интерактивной обучающей программы по работе с трудно дифференцируемой лексикой / О.Ю. Савина // Вестник ВГУ. Серия: Лингвистика и межкультурная коммуникация. 2017. № 1. С. 151-154.

Публикации в других изданиях

5. Савина, О.Ю. Использование лингвистических информационных ресурсов в дискурсивных исследованиях / О.Ю. Савина // Лингвистическое моделирование. Коллективная монография. Тюмень : Вектор Бук, 2009. С. 15-22.

6. Savina, O. Entwicklung eines Recherchetools zur Erstellung von textbezogenen Vokabellisten // O. Savina // IDT 2013. Medien in Kommunikation und Unterricht. Hrsg. Chudak, S., Drumbl, H. Nardi, A. Zanin, R. Beitrage der XV. Internationalen Tagung der Deutschlehrerinnen und Deutschlehrer. Bozen, 29. Juli – 3. August 2013. Band 6. S. 261-272.

Программы, зарегистрированные

в Федеральной службе по интеллектуальной собственности

7. Программа генерации текстоориентированного лексического минимума «Vokabelliste» : Свидетельство о государственной регистра-

ции программы для ЭВМ № 2015619379 / О.Ю. Савина, Д.И. Волобуев; заявитель и правообладатель Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Тюменский государственный университет» (RU); заявка № 2015616392 13.07.2015; дата гос. регистрации 01.09.2015, бюллетень № 2015619379, дата публикации 20.09.2015.

Подписано в печать 18.09.2017. Тираж 120 экз.
Объем 1,0 уч. изд. л. Формат 60x84/16. Заказ 600.

Издательство Тюменского государственного университета
625003, г. Тюмень, ул. Семакова, 10
Тел./факс (3452) 59-74-68, 59-74-81
E-mail: izdatelstvo@utmn.ru