

На правах рукописи



БАШМАКОВА Анастасия Юрьевна

**СЕМАНТИЗАЦИЯ
ИСТОРИКО-КУЛЬТУРНОГО КОМПОНЕНТА
В УЧЕБНОМ МУЛЬТИМОДАЛЬНОМ СЛОВАРЕ
(НА МАТЕРИАЛЕ РУССКОГО /
АНГЛИЙСКОГО ЯЗЫКОВ)**

**Специальность 10.02.20 — Сравнительно-историческое,
типологическое и сопоставительное языкознание**

**АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата филологических наук**

**Тюмень
2022**

Работа выполнена на кафедре английской филологии и перевода Института социально-гуманитарных наук федерального государственного автономного образовательного учреждения высшего образования «Тюменский государственный университет».

Научный руководитель: **Дрожащих Наталия Владимировна**, доктор филологических наук, доцент, заведующий кафедрой английской филологии и перевода ФГАОУ ВО «Тюменский государственный университет»

Официальные оппоненты: **Питина Светлана Анатольевна**, доктор филологических наук, профессор, профессор кафедры теоретического и прикладного языкознания ФГБОУ ВО «Челябинский государственный университет»

Бочегова Наталья Николаевна, доктор филологических наук, доцент, профессор кафедры «Зарубежная филология, лингвистика и преподавание иностранных языков» ФГБОУ ВО «Курганский государственный университет»

Ведущая организация: **ФГАОУ ВО «Казанский (Приволжский) федеральный университет»**

Защита состоится «18» июня 2022 года в 12:00 на заседании диссертационного 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. Ленина, 23, ауд. 516.

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

Автореферат разослан «16» апреля 2022 года.

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



Д. В. Шапочкин

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Реферируемое диссертационное исследование посвящается изучению особенностей и принципов семантизации историко-культурного компонента значения лексемы в учебном мультимодальном словаре на материале корпуса новостных текстов про образование на русском и английском языках. **Проблематика** работы лежит на стыке учебной компьютерной лексикографии, лингводидактики, лингвокультурологии, контрастивной лингвистики.

В современном мире происходит активное интегрирование информационных технологий и лингвистического знания (А.А. Бонч-Осмоловская, Д. Журавски, Н.В. Лукашевич, Р. Митков, В.П. Селегей). Данный процесс влияет на появление новых стилей взаимодействия человека с компьютером, в которых важную роль играет естественный язык. Формируется новое поколение пользователей — *digital natives* (М. Prensky, 2001), которые получают информацию посредством цифровых каналов. Появился новый феномен — *электронный (цифровой) текст*, который представлен как в традиционной линейной форме, так и в формах, предназначенных для интерактивного взаимодействия и поиска (*гипертекст, лингвистические корпуса, компьютерные словари*). Трансформируются подходы в преподавании иностранного языка, появляются новые инструменты для его изучения (*компьютерные приложения, программы и словари*), активно разрабатываемые в лингвокультурологии, лингводидактике и лингвострановедении (О.К. Ансимова, О.Н. Иванищева, Н.А. Лукьянова). Данные ресурсы являются репозиторием лексического состава языка, аккумулирующего знания о культуре и истории народа-носителя языка. Для изучения лингвокультурологических особенностей лексической семантики и поиска историко-культурного следа в текстах используются традиционные эмпирические методы анализа, требующие значительных временных затрат и вовлеченности большого количества экспертов. В последние десятилетия в лингвистике и гуманитарных науках более востребованы компьютерные, квантитативные и корпусные подходы (М.А. Большакова, А.А. Зализняк, А.Б. Кутузов, М. Anandarajan, L. Borin, E. Lieberman, J.V. Michel). Они позволяют автоматически извлекать сведения о языке и культуре. С помощью их методологического аппарата ученые исследуют употребления лексем и коллокаций в текстовых массивах для выявления лингвокультурных и социолингвистических закономерностей. В этой связи важное значение приобретает **проблема семантизации слова как достояния**

лингвокультуры в диахроническом (историко-хронологическом) аспекте. В современной науке о языке данная проблема малоизучена.

В контексте трансформации формата представления современного лингвистического знания лингводидактика ставит перед собой задачу создания новых, более эффективных учебных ресурсов, в том числе компьютерных словарей, которые способствуют успешному изучению иностранного языка и культуры. Недостатком существующих учебных словарей является отсутствие мультимодального компонента, этимологической и лингвокультурологической информации о слове, системной организации справочных материалов, демонстрирующих экстралингвистические связи лексемы. В настоящей диссертации представлен учебный мультимодальный словарь, созданный на основе тематически отобранных данных из лингвистических корпусов, исторических словарей и тезаурусов, который выступает альтернативой традиционному учебному словарю. Он совмещает разные мультимодальные средства, становится интерактивным образовательным ресурсом, фиксирующим и визуализирующим диахронические изменения в культуре, и удовлетворяет запросы нового типа пользователей. Меняется формат процесса семантизации слова, позволяющий визуально, более репрезентативно по сравнению с традиционными словарями, организовать лингвокультурные и социолингвистические данные о лексеме, показать ее концептуальные и контекстуальные связи. Этот формат является главным преимуществом *компьютерного учебного мультимодального словаря историко-культурологического типа.*

Степень разработанности проблемы. Термин *семантизация* используется не только по отношению к процессу знакомства с новой лексикой во время занятий по иностранному языку, но и для обозначения особенностей описания значения слова в словарной статье. Изучением семантизации занимаются как специалисты в области лингводидактики, так и лексикографы. Значимый вклад в развитие этого направления внесли работы отечественных и зарубежных исследователей, которые изучали *особенности семантизации лексических единиц*: в рамках *психолингвистического эксперимента* на примере тематически связанных слов (У.М. Трофимова, 2005); с точки зрения *развития когнитивных характеристик межкультурной компетенции* (А.В. Соболева, О.А. Обдалова, 2015); в аспекте *обучения чтению оригинальных иноязычных текстов* (Т.А. Дакукина, 2015); в вопросе *осмысления и изучения нового иноязычного материала* (В.В. Мальцев, 2018). Ученые предлагают различные подходы и способы

семантизации иноязычной лексики: путем *анализа морфемных составляющих и словообразовательных элементов* (О.К. Атрашевская, 2012); путем *анализа содержательной и формальной стороны понятия* (И.В. Левенталь, 2014).

С лексикографической точки зрения особенности семантизации рассматриваются в контексте *развития речевой лексической компетенции в ситуациях иноязычного профессионального общения* на примере учебного терминологического словаря (Т.С. Серова, Г.Р. Чайникова, 2015). В терминологическом аспекте представлены исследования семантизации на материале: *терминологии лингводидактики* (И.И. Ярославская, Е.В. Астремская, 2014); *терминологии высшего образования в русском языке* (Е.А. Айсакова, Т.А. Филипская, 2018); в *юридическом дискурсе* (П.В. Сысоев, В.В. Завьялов, 2018; К.Г. Чикнаверова, 2019).

Вопросы *мульти-modalности* изучаются в рамках теории социальной семиотики (М.Т. Маубури, 1999; J. Bezemer, G. Kress, 2015), лингвистики (А.А. Кибрик, 2010), теории дискурса и педагогики (И.П. Хутыз, 2016; Ю.В. Сорокина, 2017), лексикографии (Е.П. Иванова, 2007; R. Lew, 2010; T. Sato, 2016; А.Е. Гулевская, С.Л. Мишланова, 2017). Развитие теоретических основ и положений компьютерной и мульти-modalной лексикографии оказало влияние на разработку словарей различной направленности: *лингвокультурологических, идеографических, переводных, иллюстративных, терминологических, сопоставительных* и других (Н.Г. Кантышева, 2011; Н.В. Шафтельская, 2013; И.С. Волошина, 2015; Г.Ф. Богачёва, А.И. Ольховская, М.К. Парамонова, 2017; Т.Н. Хомутова, М.Г. Мамонтова, 2018; Д. Помаролли, 2018; Г.Р. Сафиуллина, 2021).

Вместе с тем ряд вопросов еще не получил должного освещения. К ним, в частности, относится *использование мульти-modalных средств представления информации для семантизации лингвокультурологического и этимологического компонента лексики и создание компьютерного учебного лексикографического ресурса*. Такой ресурс позволяет сохранить и визуально представить культурное наследие, фиксирующее изменения в мышлении и культурном сознании носителей языка. Изучение мульти-modalных инструментов семантизации лексемы и разработка *модели компьютерного учебного мульти-modalного словаря историко-культурологического типа* способствует более глубокому осмыслению языковых и культурных изменений. Предлагаемый словарь — это попытка создания современного учебного словаря нового типа.

Актуальность темы диссертации определяется:

- стабильным вниманием научного сообщества к проблематике взаимосвязи языка и культуры народа;
- недостаточной разработанностью способов репрезентации историко-культурного (диахронического) компонента значения лексемы в учебных словарях;
- сменой научных парадигм и ориентацией современной лингвистики на цифровизацию языка и текста;
- применением методов компьютерной лингвистики к анализу семантических изменений в диахронии.

Особую актуальность приобретает изучение различий в лексике тематической сферы образования в русском / английском языках в условиях глобализации общества. Наблюдается увеличение количества иностранных студентов в российских вузах и русскоговорящих студентов в зарубежных вузах. Создаваемый словарь становится важным образовательным ресурсом для таких студентов. Он направлен на репрезентацию историко-культурного своеобразия лексического состава русского / английского языков, обусловленного экстралингвистическими факторами.

В качестве **гипотезы** исследования выступает предположение о том, что принципы и технологии моделирования в современной мультимодальной лексикографии позволяют создать учебный компьютерный словарь нового типа, визуализирующий диахронические изменения в языке и семантизирующий историко-культурные особенности лексики типологически разных языков.

Объект исследования — моделирование компьютерного учебного мультимодального словаря историко-культурологического типа для изучающих иностранные языки.

Предмет исследования — способы семантизации историко-культурных изменений лексем тематического поля «Образование / Education» в русском и английском языках.

Цель настоящего исследования заключается в выявлении принципов и особенностей семантизации историко-культурного компонента лексики посредством разработки модели компьютерного учебного мультимодального словаря. Достижение поставленной цели предполагает решение следующих **задач**:

1) описать основные направления исследований семантизации в области учебной, компьютерной и мультимодальной лексикографии на примере существующих технологических решений;

2) составить список характеристик компьютерного учебного мультимодального словаря и способов семантизации историко-культурного компонента значения в словарной статье;

3) разработать модель компьютерного учебного мультимодального словаря и подобрать инструменты для ее реализации;

4) выявить состав тематического поля «Образование / Education» в русском и английском языках, описать его семантические характеристики с целью выявления историко-культурных изменений;

5) представить пилотную модель словаря для визуализации результатов исследования тематического поля «Образование / Education», описать этапы и технологию его разработки.

Методологической и теоретической основой исследования послужили положения российских и зарубежных исследователей: *о связи языка и культуры при изучении иностранного языка*: Е.М. Верещагин, В.Г. Костомаров (1971), О.Н. Иванищева (2005), О.Л. Петрушова (2009), О.К. Ансимова (2012, 2017), Н.А. Агапова, Н.Ф. Картофелева (2014), Н.Н. Белозерова, С.Н. Антонова (2017), В.Л. Ибрагимова (2019), Н.Н. Бочегова (2019), О.Л. Крамаренко, О.Ю. Богданова, Ю.В. Мошкина (2020, 2021); *основных принципах составления учебных и компьютерных словарей*: Л.В. Щерба (1974), В.В. Морковкин (1985), Y. Wilks, В.М. Slatog, L.M. Guthrie (1996), В.В. Дубичинский (1998), В.Д. Табанакова (2001, 2009), О.С. Рублева (2010), Л.Ю. Щипицина (2013), А.В. Палкова (2015); *использовании современных компьютерных средств в процессе обучения иностранному языку*: К.Р. Пиотровская (1991), М. Levy (1997), М.А. Бовтенко (2000), Э.И. Клейман, Е.Ю. Мамонова (2019); *особенностях семантизации иноязычной лексики* А.Н. Шамов (2008), А.М. Юсупова (2017), Е.О. Кузьмина (2019), Л.О. Бутакова (2021); *взаимодействии различных вербальных / невербальных кодов и модулей в тексте*: R. Lew (2010), G. Kress (2010), Е.В. Рябцев, Н.В. Халина, Л.М. Комиссарова, Т.Н. Внучкова (2018), С.А. Песина (2020), С.А. Пятина, А.Д. Щербов (2020); *особенностях исторического развития языка и семантических изменений* К. Reisig (1839), А.А. Потенбня (1864), А.И. Смирницкий (1952), М.М. Покровский (1959), Г.П. Пауль (1960), В.В. Виноградов (1999), А.А. Зализняк (2001), Н.Н. Лыкова (2005, 2015), О.Б. Пономарева (2011), Н.В. Дрожжих (2012, 2018), А.Б. Кутузов (2020);

В работе были использованы следующие **методы**:

1) *теоретико-аналитический метод*, включающий обзор и анализ научной литературы;

2) *методы компьютерной и квантитативной лингвистики* (веб-скрейпинг, контент-анализ, частотный анализ, векторизация текста, машинное обучение, нейронные сети), направленные на компьютеризированные сбор корпуса новостных статей, предобработку текста и выявление ключевых слов;

3) *метод семантико-дефиниционного анализа*, применяемый для анализа лексем и выявления расхождений на семантическом уровне;

4) *метод этимологического анализа*, способствующий объяснению исходных словообразовательных связей;

5) *метод сопоставительного анализа*, позволивший выявить общие и дифференциальные признаки лексем русского и английского языка;

6) *описательно-аналитический метод*, позволивший обобщить, описать и представить полученные результаты;

7) *метод моделирования в компьютерной лингвистике*, заключающийся в разработке структуры компьютерного словаря и программировании интерактивного веб-приложения для его оболочки.

Материалом исследования является корпус новостных статей в тематике «Образование / Education», собранных с сайтов образовательного онлайн-портала «*EDU-Inform*» (русский язык) и журнала «*Education Today Magazine*» (английский язык). Корпус разделен на две части: *русскаяязычная часть* насчитывает 1119 новостных статей (779793 словоупотребления); *англияязычная часть* состоит из 534 новостных статей (285320 словоупотреблений). При сборе частей корпуса использован весь опубликованный текстовый материал. Для сохранения однородности данных, стиля и формата публикаций принято решение не добавлять и не использовать дополнительно другие англоязычные источники. Различия в объеме русскоязычной и англоязычной частей корпуса рассматриваются как индивидуальные характеристики и особенности источников материала.

В качестве дополнительного систематизирующего материала использовались *толковые, учебные, этимологические, ассоциативные, идеографические и энциклопедические словари, интерактивные тезаурусы и корпуса* для русского и английского языков: Национальный корпус русского языка (НКРЯ), Онлайн-тезаурус «Карта слов», Тезаурус RuWordNet, WordNet, Google Books Ngram Viewer, The Visual Thesaurus, Fine Dictionary, Oxford English Dictionary, Linguazza.com и др. На данный момент модель словаря включает в себя 20 словарных статей (10 русских и 10 английских лексем из тематического поля «Образование / Education»).

Положения, выносимые на защиту:

1. *Компьютерный учебный мультимодальный словарь историко-культурологического типа* — это словарь для нового типа пользователя, который представляет собой современный многофункциональный образовательный ресурс. Его преимуществом является *новый формат семантизации слов с использованием мультимодальных ресурсов*, который позволяет визуально, более репрезентативно по сравнению с традиционными словарями, организовать культурологические и социально-исторические данные о лексеме, а также показать ее концептуальные, контекстуальные и экстралингвистические связи в языке.

2. *Историко-культурный (диахронический) компонент* значения лексемы в мультимодальном словаре является совокупностью информации об истории семантического развития слова, национально-культурных коннотациях, контексте и особенностях его употребления в определенную историческую эпоху, которая описывает слово как постоянное определенной лингвокультуры.

3. Для *семантизации историко-культурного компонента* в словарной статье компьютерного учебного мультимодального словаря используются следующие элементы: *описательное вербальное толкование слова на иностранном языке; краткая этимологическая справка; экстралингвистический комментарий; мультимодальные ресурсы; наглядная визуализация языкового окружения лексемы; гипертекстовый доступ.*

4. Базовыми компонентами модели учебного мультимодального словаря историко-культурологического типа являются его макроструктура (*вводная часть, облака́ слов, словник*) и микроструктура, элементами которой являются *заголовочное слово; грамматическая помета; фонетическая транскрипция; аудиопрезентация слова; современные дефиниции слова с примерами употребления; этимологическая справка; историко-культурный комментарий; информация о частотности употреблений лексемы в диахронии; примеры употребления слова в корпусе; иллюстрации; видео-контент; разбор слова по составу (для русского языка); информация об уровне владения языком по системе CEFR для каждого значения лексемы (для английского языка); визуализация хронологии развития семантических изменений в значениях слов в формате «*timeline*» (для английского языка).*

5. Ключевая лексика *тематического поля «Образование / Education»* на материале текстов образовательных СМИ на русском и английском языках имеет отличительные характеристики, обусловленные *историческими*

(ключевая лексика английского языка входит в употребление раньше, чем лексемы русского языка, что объясняет ее многозначность и наличие слов-когнатов), *социально-политическими* (прослеживается значительное влияние государства на стиль публикаций на русском языке, в обоих языках доминирует тематика пандемии коронавируса) и *национально-культурными* (тематика благотворительности в английском языке, вопросы образовательного менеджмента в обоих языках) особенностями выбранных языков.

Научная новизна исследования определяется:

— комплексным описанием принципов и приемов беспереводной семантизации историко-культурного компонента лексемы в словарной статье с использованием мультимодальных инструментов представления информации;

— обоснованием необходимости создания учебного словаря нового типа, способного интегрировать в себе описание лексической единицы как динамичного образования, отражающего лингвокультурологические процессы в диахронии и синхронии;

— изучением и сопоставительным анализом диахронических семантических изменений ключевых слов образовательного дискурса русского и английского языков цифровыми методами лингвистического исследования;

— попыткой репрезентировать и визуализировать формализованную модель знаний о лексике тематического поля «Образование / Education» в русском и английском языках;

— построением модели компьютерного учебного мультимодального словаря историко-культурологического типа, основанной на технологиях компьютерной лексикографии (написание программного кода на языках программирования Python 3 и R).

Теоретическая значимость исследования обусловлена:

— уточнением и систематизацией принципов беспереводной семантизации историко-культурных изменений семантики лексемы в словарной статье в рамках теории мультимодальности;

— изложением методики компьютеризированного выявления ключевой лексики текстов образовательного дискурса и алгоритма ее визуализации;

— изучением и сопоставлением особенностей семантики, этимологии и частотных характеристик ключевых слов тематического поля «Образование / Education» в русском и английском языках в рамках сравнительно-исторического подхода;

— конкретизацией методики построения словарной статьи за счет использования мультимодальных инструментов для семантизации лексики.

Полученные в ходе исследования выводы вносят вклад в развитие компьютерной, учебной и мультимодальной лексикографии в рамках лингводидактического подхода, а применяемая методика автоматического выявления и изучения ключевой лексики может быть реализована при проведении исследований текстов других жанров и типов дискурса, что расширяет общетеоретическую значимость выдвинутых положений.

Практическая значимость работы заключается в том, что результаты проведенного исследования могут быть использованы для анализа особенностей и принципов семантизации лексики, а также характера её историко-культурных изменений в процессе преподавания лексикологии, лингвокультурологии, лингводидактики, компьютерная лексикография и др. Разработанная *модель компьютерного учебного мультимодального словаря историко-культурологического типа* может быть применена для улучшения существующих компьютерных учебных словарей и создания аналогичных лексикологических ресурсов для пары русского / английского языков и других.

Степень достоверности результатов исследования определяется опорой на авторитетные теоретические положения отечественных и зарубежных ученых, применением методов анализа, соответствующих целям и задачам исследования, привлечением обширного фактического материала, согласованностью теоретических выводов с результатами практического исследования и возможностью их верификации на другом материале.

Апробация результатов исследования проводилась в форме отчетных выступлений на заседаниях кафедры английской филологии и перевода и кафедры английского языка Тюменского государственного университета. Основные положения исследования излагались в виде докладов и сообщений на международных и всероссийских научных конференциях: *VII всероссийская молодежная научно-практическая конференция «Множественность интерпретаций: верификация гуманитарного знания»*, ТюмГУ, Тюмень, 2018 г.; *Онлайн-конференция 68th Studentische Tagung Sprachwissenschaft (StuTS)*, Берлин, 2020 г.; *Всероссийская междисциплинарная молодежная научная конференция с международным участием «IX Информационная школа молодого ученого»*, ЦНБ УрО РАН, Екатеринбург, 2021 г.; *III Международная конференция «Синергия языков и культур»*:

междисциплинарные исследования», СПбГУ, Санкт-Петербург, 2021 г.; *Международная научно-практическая конференция «Язык. Культура. Перевод»*, МГИМО, Одинцово, 2022 г.; доклады на заседаниях лаборатории обработки естественного языка (NLP Lab) в Государственном университете Монтклера (Montclair State University), г. Монтклер, США, во время научно-исследовательской стажировки по программе У. Фулбрайта для выпускников вузов и аспирантов (Fulbright Foreign Student Program), 2021 г. Основное содержание диссертации отражено в 6 публикациях автора, 4 из которых в ведущих рецензируемых изданиях.

Объем и структура диссертации. Диссертация состоит из введения, четырех глав с выводами по каждой главе, заключения и библиографического списка, включающего в себя 273 источника, из них 92 источника на английском языке. Общий объем работы составляет 197 страниц печатного текста. Работа иллюстрирована таблицами, схемами и рисунками. Дополнительные материалы исследования и программный код на языках программирования Python 3 и R расположены в репозитории GitHub в сети Интернет по ссылке¹.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** приведены области соответствия настоящей работы паcпорту специальности **10.02.20 «Сравнительно-историческое, типологическое и сопоставительное языкознание»**, обосновывается актуальность темы диссертации, указывается степень разработанности проблемы мультимодальности и семантизации в лингводидактическом и сравнительно-историческом аспектах; формулируются объект, предмет, цель и задачи исследования; описывается материал исследования; определяются научная новизна, теоретическая и практическая значимость исследования; выдвигаются гипотеза и положения, выносимые на защиту; перечисляются методы, использованные при выполнении работы; приводятся сведения о степени достоверности и апробации результатов исследования; представляется общая структура и объем работы.

В первой главе «Мультимодальность как современный принцип представления информации в словаре» рассматриваются теоретические предпосылки настоящей работы на материале исследований

¹ Репозиторий GitHub, URL: <https://github.com/nastuk/dissertation>

в области *компьютерной лексикографии и теории мультимодальности*. В **разделе 1.1** «*Компьютерный словарь в современной лексикографии*» обозначена роль компьютерной лексикографии в современной лингвистике, описываются основные направления исследований, дается определение понятия *компьютерного словаря* с рассмотрением его технологических и структурных особенностей. В **разделе 1.2** «*Типология компьютерных лексикографических ресурсов*» представляются основные характеристики и типология компьютерных лексикографических ресурсов в обзоре работ отечественных и зарубежных лингвистов (*традиционный словарь, тезаурус, словарь лексической сочетаемости, двуязычный словарь, стилистический словарь, конкордансы*). В **разделе 1.3** «*Эволюция компьютерных лексикографических ресурсов*» описывается история формирования компьютерной лексикографии как современного научного направления начиная с 60-х годов прошлого века до настоящего момента. Представляются этапы развития компьютерных лексикографических ресурсов с особым вниманием на форматы репрезентации семантических отношений в них (*WordNet, FrameNet, генеративный лексикон, корпусные словари*). Далее в **разделе 1.4** «*Мультимодальный словарь как словарь для нового пользователя*» рассматриваются основные положения теории мультимодальности и особенности мультимодального представления информации в компьютерных словарях (*вербальной и невербальной*). На примере существующих технологических решений для русского и английского языков (*тезаурус PyTez, RuWordNet, тезаурус «Карта слов», The Visual Thesaurus, Fine Dictionary*) анализируются особенности использования в словаре элементов различных модальностей, таких как *текст, изображения, аудио- и видеоматериалы*, которые способны облегчить процесс изучения иностранного языка, а также повысить интерес и вовлеченность обучающихся. Результаты изучения теоретического материала и анализа существующих компьютерных лексикографических ресурсов показывают, что требования современного общества и цифровизация ставят перед лексикографией новые задачи по созданию компьютерных мультимодальных ресурсов, включающих различные форматы представления информации, такие как *аудио- и видеозаписи, иллюстрации, интерактивные элементы* и другие.

Вторая глава «Мультимодальный словарь в лингводидактике» посвящена изучению актуальных направлений и результатов исследований в области *компьютерной лингводидактики и учебной лексикографии* в рамках *мультимодального подхода*. **Раздел 2.1**

«Компьютерная лингводидактика» представляет собой теоретический анализ литературы, описывающий основные тенденции в разработке компьютерных образовательных ресурсов для изучения иностранного языка, в том числе рассматриваются особенности междисциплинарного взаимодействия компьютерной лексикографии и компьютерной лингводидактики. В **разделе 2.2** «Учебный словарь для изучения иностранного языка» обозначена роль компьютерных словарей в процессе обучения языку и знакомства с новой иноязычной лексикой, в том числе описывается история разработки словарей для изучающих русский и английский языки как иностранные. В **разделе 2.3** «Создание и разработка учебного компьютерного словаря» в рамках обзора комплексно описываются вопросы создания, задачи и этапы разработки учебных словарей для изучающих иностранный язык на примере методологии различных отечественных и зарубежных научных лексикографических школ. Так, при разработке лексикографических ресурсов ученые выделяют *факторы*, которые необходимо учитывать перед началом работы по созданию концепции словаря, *ключевые элементы* словаря, *компоненты его структуры*, *особенности базы данных* словаря, его *функции*, а также *технические требования* и *инструментарий*. Необходимо отметить, что до настоящего времени не сложилось единой универсальной целостной системы, описывающей особенности и этапы моделирования словаря. Более того, вопрос совмещения мультимодальных инструментов, визуализации и историко-культурной информации в словаре остается неразработанным. По итогам главы делается вывод, что компьютерный словарь, включающий в себя различные мультимодальные форматы представления данных (*текстовый, визуальный, аудио, видео* и т.д.) с опорой на лингвистические корпуса, может стать образовательным инструментом, способствующим интеграции и использованию информационных технологий в процессе обучения иностранному языку.

В третьей главе «**Мультимодальный словарь историко-культурологического типа как новый тип учебного словаря**» определяется значимость вопроса *семантизации историко-культурного компонента лексемы* при создании компьютерного учебного мультимодального словаря для изучающих иностранный язык как нового типа словаря. В главе описываются основные направления исследований в области *лингвокультурной лексикографии, семасиологии, цифровой гуманитаристики, культурологии*. **Раздел 3.1** «Лингвокультурная лексикография в современном сопоставительном языкознании» анализирует значимость лингвокультурологической информации

о слове при разработке учебного словаря для иностранных учащихся, а также важность ее использования при сопоставительном исследовании языка. **Раздел 3.2** «Историко-культурный (диахронический) компонент лексики: история вопроса» рассматривает историографию диахронического изучения лексики с целью исследования историко-культурных особенностей ее значения. В разделе проводится анализ классических и современных компьютерных подходов исследования семантики (*культуромика, дистрибутивная семантика*). В результате формулируется определение **историко-культурного (диахронического) компонента** значения лексемы как *совокупности информации об истории семантического развития слова, национально-культурных коннотациях, контексте и особенностях его употребления в определенную историческую эпоху, которая описывает слово как достояние определенной лингвокультуры*. В **разделе 3.3** «Способы семантизации лексики в учебном словаре» представлены различные способы семантизации лексики в учебном словаре для иностранных учащихся с целью описания историко-культурных особенностей ее значений. Рассматриваются существующие переводные и беспереводные способы семантизации, которые включают в себя: *описательное вербальное толкование значения слова на иностранном языке; наглядную визуализацию (иллюстрации, графики, диаграммы); словообразовательный анализ формы слова; этимологический анализ лексемы; анализ контекстного употребления; группировку лексики на основе семантических отношений (лексико-семантические и тематические поля, синонимические и антонимические ряды, родовидовые понятия, однокоренные слова); демонстрацию устойчивых ассоциаций, связанных со словом; использование лингвострановедческой информации*. В завершении анализа формулируются **принципы и особенности семантизации историко-культурного компонента**, такие как *описательное вербальное толкование слова на иностранном языке, краткая этимологическая справка; экстралингвистический комментарий; мультимодальные элементы; наглядная визуализация языкового окружения лексемы; гипертекстовый доступ*. В **разделе 3.4** «Моделирование компьютерного учебного мультимодального словаря историко-культурологического типа» обозначаются основные содержательные и структурные компоненты моделирования компьютерного учебного мультимодального словаря, направленного на изучение и визуализацию историко-культурного (диахронического) компонента значения лексемы в рамках лингводидактического подхода.

Опираясь на положения, изученные в теоретических главах настоящей работы, мы выделяем основные *характеристики*, которыми должен обладать современный компьютерный учебный словарь для изучающих иностранный язык:

- структура, которая стремится показать концептуальные связи между лексемами в формате лексических / семантических полей и групп;
- грамматическая и морфологическая информация;
- информация об употреблении лексемы с предлогами, другими словами, словосочетаниями или конструкциями;
- дополнительная информация, помогающая студенту погрузиться в контекст употребления лексемы с ее историческими и культурологическими особенностями, а также декодировать новые встречающиеся слова с помощью контекстных подсказок;
- структура и содержание словаря, мотивирующие учащегося к дальнейшему изучению языка и формированию собственных лексических связей и семантических сетей.

Далее в разделе определяются три основных концептуально различных направления работы над моделью компьютерного словаря: (1) *моделирование концепции и структуры словаря*, (2) *работа с языковым материалом-основой* и (3) *разработка программного продукта*. Представляется развернутая схема-инфографика, отражающая *содержательные и структурные компоненты процесса моделирования* компьютерного учебного мультимодального словаря для изучающих иностранный язык, разрабатываемого в настоящем исследовании (схема 1). Процесс моделирования разделяется на шесть последовательных этапов, часть из которых включают в себя несколько равных по значимости компонентов. Последовательность их разработки определяется исследователем в каждом конкретном случае в зависимости от различных внешних и внутренних факторов.

В конце главы, делается вывод о том, что создание компьютерного учебного словаря, включающего в себя различные мультимодальные форматы представления данных (текстовый, визуальный, аудио, видео и т.д.) с опорой на лингвистические корпуса и формальные модели языковой семантики, позволяет говорить о словаре будущего, востребованном в современной теории образования. Такой словарь является инновационным инструментом преподавания иностранного языка в рамках компьютерной лингводидактики. Он направлен на нового пользователя и содержит богатый экстралингвистический материал, имеющий большую культурологическую ценность.

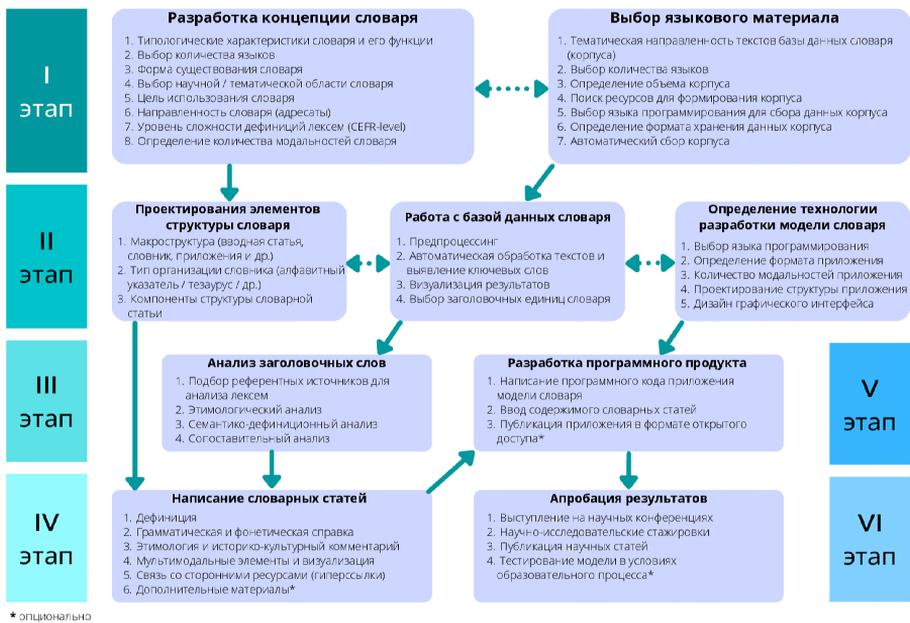


Схема 1. Содержательные и структурные компоненты процесса моделирования словаря²

Четвертая глава «Разработка модели учебного мультимодального словаря историко-культурологического типа» посвящается практической части настоящего исследования и состоит из описания этапов и технологии разработки модели компьютерного учебного мультимодального словаря для визуализации результатов диссертационного исследования и изучения практических аспектов состава тематического поля «Образование / Education» в русском и английском языках. В качестве финального продукта разрабатывается пилотная модель компьютерного учебного мультимодального словаря историко-культурологического типа для пары языков русский-английский, анализируется его макро- и микроструктура и языковой материал тематического поля «Образование / Education».

² Схему полномасштабного размера можно найти по ссылке: https://github.com/nastuk/dissertation/blob/4fa37847a6bca5fcbb15301ea5a9ad584241bb80/modelling_steps.pdf

Раздел 4.1 «Формирование концепции словаря» рассматривает вопросы содержательного описания концепции моделируемого в настоящем исследовании *компьютерного учебного мультимодального словаря историко-культурологического типа*, который представляет собой формализованную модель знаний о лексике тематического поля «Образование / Education» в русском и английском языках. Настоящий словарь комбинирует типологические и функциональные признаки различных словарей: *идеографического, толкового, учебного, частотного, двуязычного, компьютерного, мультимодального*. Словарь реализован в формате веб-ресурса, не требует инсталляции, доступ осуществляется через сеть Интернет по ссылке³. Спроектированный словарь реализует базовые функции словарей: *учебную, справочную, систематизирующую и нормативную*. В словаре представлены два языка: *русский и английский*, при этом применяются принципы *беспереводной семантизации* лексем, поэтому в словарной статье преобладает использование только *одного языка* (русского или английского соответственно). Дефиниция лексемы, справочные комментарии, дополнительные аудио- и видеоматериалы оформлены, исходя из этого условия. При выборе источника дефиниции мы руководствовались принципами *полноты, достаточности, минимальности, четкости*, а также *простоты и доступности изложения. Уровень объяснения* лексемы: вводное объяснение для учащихся, студентов, с дополнительной информацией энциклопедического характера. Дефиниции терминов взяты из одноязычных *толковых, этимологических, исторических, энциклопедических и специализированных* словарей. Так как моделируемый словарь является полифункциональным, он может быть предназначен для широкого круга пользователей. *Адресатами* словаря могут стать пользователи, изучающие русский или английский языки как иностранный на продвинутом уровне, лингвисты, занимающиеся исследованием этимологии слов, а также студенты-специалисты в области педагогики и студенты-лингвисты. Основной *целью использования словаря* может стать изучение особенностей употребления слов тематического поля «Образование / Education», их семантики и этимологии.

Раздел 4.2 «Тематическое поле «Образование / Education» в русском и английском языках: создание базы данных словаря» состоит из шести подразделов, которые описывают *этапы создания корпуса текстов* — лингвистической базы разрабатываемого словаря по-

³ Модель учебного мультимодального словаря. URL: https://na-stuk.shinyapps.io/multimodal_dict/

В связи с этим у некоторых лексем изменены части речи, например прилагательное «студенческий» заменено на существительное «студент». Такие замены мотивируются преимущественно большим количеством значений слова у существительных, чем у прилагательных, так как для моделируемого словаря в первую очередь важно представить наибольшее разнообразие значений лексем. Задача дальнейшего анализа — проследить, как исторически менялась семантика ключевых лексем, выявленных автоматизированными методами по материалам современных новостных статей в сфере образования.

В результате комплексного этимологического, семантико-дефиниционного и сопоставительного анализа языкового материала тематического поля «Образование / Education» в русском и английском языках выявляются следующие особенности:

— статистические показатели исследовательского корпуса новостных текстов на русском и английском языках подтверждают, что исследуемые русскоязычные образовательные СМИ публикуют большее количество и объем новостных материалов, чем англоязычные СМИ;

— исследуемые тексты образовательных СМИ на русском и английском языках отражают культурные особенности сферы образования двух стран. Тексты новостных статей на русском языке подвержены значительному влиянию со стороны государства, что представлено в обилии канцеляризмов, имен собственных, сложных прилагательных и нейтральной лексики. В публикациях на английском языке частотной становится тематика благотворительности, которая является культурно-специфической для Великобритании. В последние годы в обеих странах становятся актуальными вопросы образовательного менеджмента и отмечается преобладание лексики, тематически связанной с освещением организации образовательного процесса в условиях пандемии коронавируса;

— анализируемая ключевая лексика английского языка начала входить в употребление значительно раньше (XIV–XVI вв.), чем лексика русского языка (XVIII–XIX вв.). Данную особенность можно объяснить историческим развитием системы образования Европы и России, что также может стать причиной заимствований и наличия слов-когнатов, имеющих общую этимологию в русском и английском языках;

— сопоставительный анализ зеркального списка ключевых слов тематического поля «Образование / Education» в русском и английском языках показывает, что немногие слова сохранили стабильное значение на протяжении всей своей истории. Исследуемая лексика английского языка более многозначна, чем лексика русского языка. Такая особенность представляет сложность при изучении новой лексики иностранного языка

учащимися. При этом она отражает ценную информацию об историческом развитии семантики слов и является основой для анализа культурных трендов в сфере образования России и Великобритании.

В **разделе 4.3** «Проектирование компьютерного приложения модели учебного мультимодального словаря историко-культурологического типа» подробно описывается завершающий этап разработки модели словаря, включающий в себя дизайн пользовательского интерфейса, подбор дополнительных справочных источников и мультимодального контента, а также написание программного кода на языке программирования R. При создании компьютерного веб-приложения модели учебного мультимодального словаря историко-культурологического типа используются современные технологии и программные продукты (*Python 3, R, Shiny Dashboards, HTML и CSS*). Для реализации графического интерфейса словаря выбран формат интерактивной платформы *Shiny Dashboard*⁴ на языке программирования R, представленный в виде информационной панели, состоящей из трех частей (рис. 2): *заголовка* (header), *боковая панель* (sidebar) и *тело сайта* (body). Основные концепции, принятые при проектировании графической оболочки, включают: *однооконный интерфейс* без использования дополнительных диалоговых окон, организацию переходов по разделам словаря при помощи *гиперссылок*.

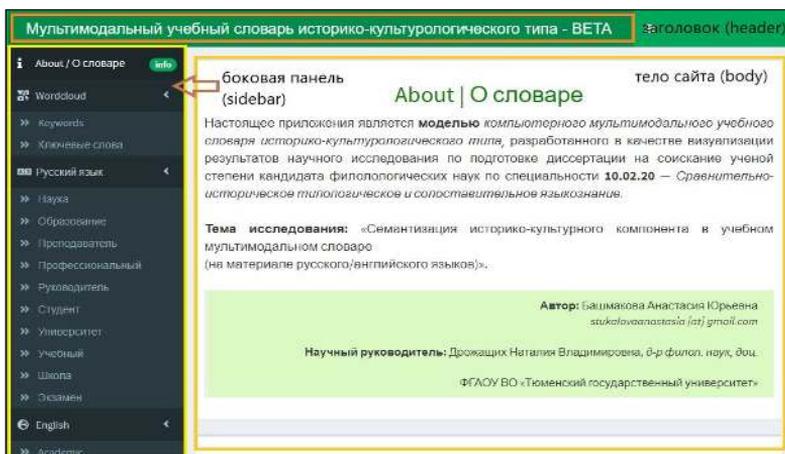


Рисунок 2. Графический пользовательский интерфейс словаря

⁴ Shiny Dashboard, URL: <https://shiny.rstudio.com/>

Интерфейс словаря построен максимально удобно для обычных пользователей и специалистов, не имеет лишних «окон» и «модулей», что обеспечивает высокую продуктивность работы.

Макроструктура словаря обеспечивает *одноканальный вход* в словарную статью по алфавитному указателю списка слов в боковой панели и включает в себя: (1) *вводную часть*: раздел «About / О словаре»; (2) *облака слов*: раздел «Wordcloud», в котором представлены облака слов ключевой лексики тематического поля «Образование / Education», при наведении курсором на слово в облаке появляется рамка, выделяющая саму лексему, и информация о ее частотности в списке ключевых слов; (3) *словник*: разделы «Русский язык» и «English», включающие в себя словарные статьи.

На уровне **микроструктуры** словарные статьи для русского (рис. 3) и английского языков (рис. 4) имеют следующие элементы:

- 1) *заголовочное слово*;
- 2) *грамматическая помета*;
- 3) *фонетическая транскрипция* (на основе фонетической системы международного фонетического алфавита IPA);
- 4) *аудиопрезентация слова* (запись женского или мужского голоса; требуется воспроизведение аудио файла вручную);
- 5) *современные дефиниции слова с примерами употребления*;
- 6) *этимологическая справка*;
- 7) *историко-культурный комментарий*;

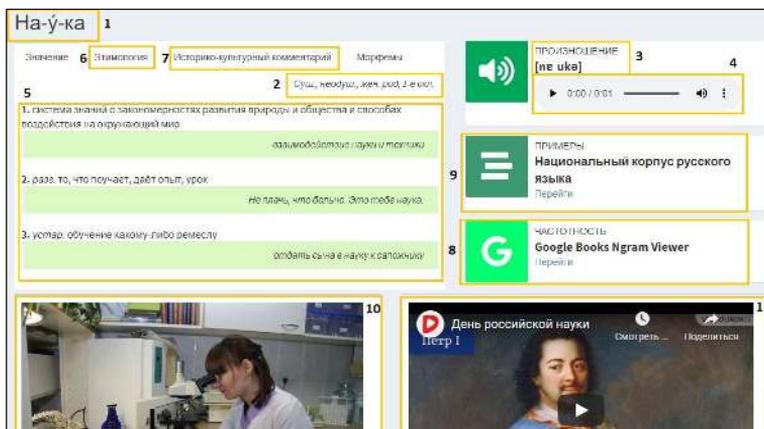


Рисунок 3. Словарная статья лексемы «Наука»



Рисунок 4. Словарная статья лексемы «Academic»

8) информация о частотности употреблений лексемы в диакронии (гиперссылка на лексикографический ресурс Google Books Ngram Viewer);

9) примеры употребления слова в корпусе (гиперссылка на НКРЯ для русского языка или на онлайн-сервис Linguazza.com для английского);

10) иллюстрации (опционально);

11) видео-контент (встроенный виджет на видео с платформы YouTube; опционально);

12) раздел «Морфемы» (только для русского языка);

13) CEFR Levels (только для английского языка);

14) блок «Timeline» (только для английского языка).

В статьях представлены различные мультимодальные элементы (аудио- и видеозаписи, иллюстрации) и дополнительные справочные материалы, которые способствуют семантизации историко-культурного компонента значения лексем, представленных в словаре, и повышают привлекательность интерфейса моделируемого приложения. Изображения подобраны индивидуально для статей на русском и английском языках, чтобы проиллюстрировать реалии каждой из стран (здания школы и университета, схемы организации систем образования России и Великобритании). Технические особенности компьютерного приложения моделируемого мультимодального словаря позволяют включить в словарную статью все элементы, описанные в ее микроструктуре. Не каждая лексема имеет достаточный набор справочных данных или может быть проиллюстрирована изображением или схемой. В связи с этим не все словарные статьи обладают полным списком мультимодальных элементов.

Блок «Timeline» состоит из *текстовой части* (рис. 5), которая описывает хронологию развития семантических изменений в значениях слов, и ее *визуализации* (рис. 6). Визуализация текстовой части представлена в виде интерактивной панели в формате временной шкалы «Timeline» и расположена ниже под самим текстом. Информация в текстовой части вкладки «Timeline» выделяется разными цветами для каждой части речи (зеленый и синий). Номера значений в текстовой части соотносятся с номерами блоков, расположенных на временной шкале «Timeline». В блоках значения слова окрашены в разные цвета для каждой части речи соответственно.

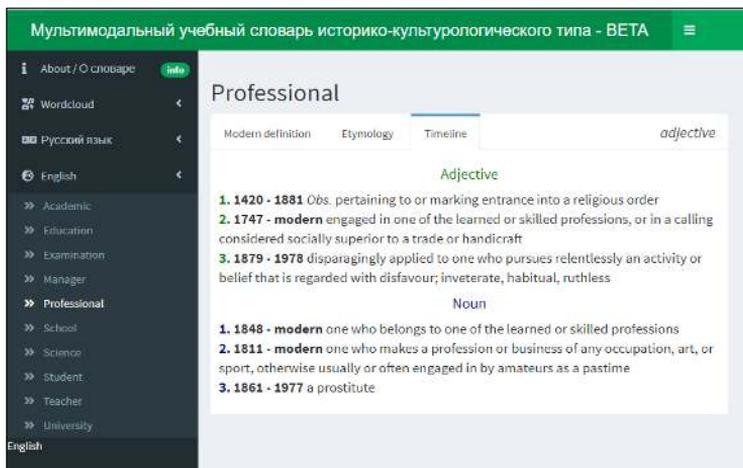


Рисунок 5. Словарная статья лексемы «Professional», текстовая часть раздела «Timeline»

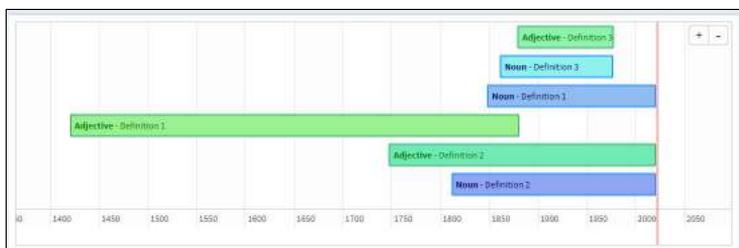


Рисунок 6. Словарная статья лексемы «Professional», визуализация раздела «Timeline»

Моделируемый в настоящем исследовании словарь является средством обучения и обширной базой данных, системным многофункциональным справочным пособием. Он соответствует характеристикам, которыми должен обладать современный компьютерный учебный словарь для изучающих иностранный язык, определенным в теоретической части исследования. В нем реализованы принципы и особенности семантизации историко-культурного компонента лексемы. Его отличительной особенностью является учет семантики слова и её исторического развития, визуально представленных на *интерактивной хронологической шкале*. *Наглядность, интерактивность, нелинейность, динамичность, антропоцентричность, культурологическая ориентированность, историчность* представляют важнейшие параметры создания и функционирования разрабатываемого словаря. Данный словарь можно охарактеризовать как словарь культурных трендов эпохи. Слово в таком словаре выступает как проводник в мир вечно меняющихся значений культуры.

В заключении диссертации обобщаются полученные результаты, формулируются итоги и значимость проведенного исследования.

Настоящее диссертационное исследование было направлено на изучение особенностей и выявление принципов семантизации историко-культурного компонента с целью разработки модели компьютерного учебного мультимодального словаря, способного визуализировать динамику семантических трансформаций лексики в диахронии на материале русского и английского языков.

Резюмируя итоги проведенного исследования, отметим, что гипотеза настоящего исследования может считаться подтвержденной. Принципы и технологии моделирования в современной мультимодальной лексикографии позволяют создать учебный компьютерный словарь нового типа, который визуализирует диахронические изменения и семантизирует историко-культурные особенности лексики типологически разных языков.

Диссертационное исследование затрагивает аспекты семантизации новой лексики иностранного языка и лексикографического моделирования в компьютерной мультимодальной учебной лексикографии — актуальном направлении современной цифровой гуманитаристики. В перспективе представленный подход к созданию словаря такого рода может найти применение при разработке аналогичных лексикографических ресурсов для различных языков.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

*Статьи в рецензируемых научных изданиях,
включенных в перечень ВАК РФ:*

1. Пономарева, Е. Ю. Реализация концепта “international student” в дискурсе британских СМИ в условиях глобализации / Е. Ю. Пономарева, А. Ю. Башмакова // Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates. — 2018. — Т. 4. — № 3. — С. 40-51. — DOI 10.21684/2411-197X-2018-4-3-40-51.

2. Дрожащих, Н. В. Учебный мультимодальный словарь историко-культурологического типа в преподавании иностранного языка / Н. В. Дрожащих, А. Ю. Башмакова // Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates. 2021. Том 7. № 1 (25). С. 43-61. — DOI: 10.21684/2411-197X-2021-7-1-43-61

3. Башмакова, А. Ю. Выявление ключевых слов тематического поля “Образование/Education” / А. Ю. Башмакова // Филологические науки. Вопросы теории и практики. — 2021. — Т. 14. — № 12. — С. 3986-3990. — DOI 10.30853/phil20210598

4. Башмакова, А. Ю. Классические и современные подходы к изучению семантических изменений / А. Ю. Башмакова // Успехи гуманитарных наук. — 2022. — № 1. — С. 80-85.

Публикации в других изданиях:

5. Башмакова, А. Ю. Культуромика в корпусном исследовании / А. Ю. Башмакова // Множественность интерпретаций: верификация гуманитарного знания : Материалы VII молодежной научно-практической конференции, Тюмень, 19 декабря 2018 года / Научный редактор Е.В. Михалькова. — Тюмень: Тюменский государственный университет, 2019. — С. 17-22.

6. Башмакова, А. Ю. Выявление ключевых слов по теме «Education» (на материале корпуса текстов онлайн-издания «Education Today Magazine») / А. Ю. Башмакова // IX Информационная школа молодого ученого: Сборник научных трудов, Екатеринбург, 20-23 сентября 2021 года / Центральная научная библиотека УрО РАН. — Екатеринбург: ООО “Издательство УМЦ УПИ”, 2021. — С. 400-407. — DOI 10.32460/ishmu-2021-9-0044.

Подписано в печать 13.04.2022. Тираж 120 экз.
Объем 1,0 уч. изд. л. Формат 60x84/16. Заказ 167.

Издательство Тюменского государственного университета
625002, г. Тюмень, ул. Осипенко, 81
Тел./факс (3452) 59-74-81, 59-75-93
E-mail: izdatelstvo@utmn.ru