

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ  
«МИНСКИЙ ГОСУДАРСТВЕННЫЙ ЛИНГВИСТИЧЕСКИЙ УНИВЕРСИТЕТ»

Объект авторского права  
УДК 811.111'371 (043.3)

МЕНЬШЕНИНА  
Ирина Александровна

**АВТОМАТИЧЕСКАЯ СМЫСЛОВАЯ КОМПРЕССИЯ  
ПУБЛИКАЦИЙ НАУЧНОГО МЕДИЦИНСКОГО ДИСКУРСА  
(на материале английского языка)**

Автореферат  
диссертации на соискание ученой степени  
кандидата филологических наук

по специальности 10.02.21 – прикладная и математическая лингвистика

Минск, 2025

Научная работа выполнена в учреждении образования «Минский государственный лингвистический университет»

Научный руководитель

**Евчик Надежда Семеновна,**

доктор филологических наук, профессор, профессор кафедры, учреждение образования «Минский государственный лингвистический университет», кафедра речеведения и теории коммуникации

Официальные оппоненты:

**Иванов Евгений Евгеньевич,**

доктор филологических наук, профессор, заведующий кафедрой, учреждение образования «Могилевский государственный университет им. А. А. Кулешова», кафедра теоретической и прикладной лингвистики

**Степанова Дарья Валерьевна,**

кандидат филологических наук, доцент, доцент кафедры, учреждение образования «Минский государственный лингвистический университет», кафедра теории и практики английской речи

Оппонирующая организация

учреждение образования «Гродненский государственный университет им. Янки Купалы»

Защита состоится «1» апреля 2025 года в 11:30 на заседании совета по защите диссертаций К 02.22.02 при учреждении образования «Минский государственный лингвистический университет» по адресу: 220034, г. Минск, ул. Захарова, 21, ауд. Б-303, e-mail: info@mslu.by, тел. ученого секретаря: 289-46-43.

С диссертацией можно ознакомиться в библиотеке учреждения образования «Минский государственный лингвистический университет».

Автореферат разослан «28» февраля 2025 г.

Ученый секретарь

совета по защите диссертаций,  
кандидат филологических наук, доцент

Н. В. Михалькова

## ВВЕДЕНИЕ

Характерный для современного социума экспоненциальный рост объемов профессионально значимой информации можно проследить по возрастанию числа научных журналов в мире. Согласно данным Международной ассоциации научных, технических и медицинских издателей (STM), на 2023 год зарегистрировано 42 000 рецензируемых журналов, в которых ежегодно печатается свыше 3 000 000 научных статей. При этом наибольшее число публикаций издается по естественным и медико-биологическим наукам, превышая показатели гуманитарных дисциплин более чем в 2 раза.

Для решения задачи по ускорению доступа специалистов к информации в настоящее время широко применяются автоматические системы, моделирующие процесс смысловой компрессии. Однако, несмотря на 70-летнюю историю развития и разнообразие предложенных методов обработки текста, они не обеспечивают достаточную точность, столь востребованную в медицинской сфере.

Такое положение делает очевидной и назревшей необходимость разработки новых подходов к решению проблемы реализации смысловой компрессии научных медицинских публикаций, способных обеспечить создание вторичного текста предельной точности и терминологической представленности без искажения смысла, заложенного адресантом, а также исчерпывающей полноты в изложении профессионально значимой информации, необходимой адресату, для ее безошибочной интерпретации.

Актуальность настоящего исследования обусловлена тем, что решение данной проблемы осуществляется посредством применения субъектно-ориентированного подхода, позволяющего проводить когнитивный анализ внутреннего построения научной публикации с учетом заинтересованной включенности разных участников медицинского дискурса: 1) автора публикации, владеющего дискурсивными стратегиями передачи нового медицинского знания, 2) профессионального читателя статьи, заинтересованного в получении актуальной научной информации, 3) исследователя, обладающего лингвистическими методами структурно-семантического анализа текста и прикладным инструментарием для формализации его обработки автоматическим способом. Изучение когнитивных аспектов порождения и понимания профессионального дискурса в медицинской сфере в сочетании с его лингвистическим анализом позволяет обнаружить и представить многослойную структуру организации смысла научной медицинской публикации. Это выводит на возможность точного

выделения семантически наиболее важной информации для осуществления формализации процесса компрессии текста.

Отличительную особенность настоящего исследования составляет то, что теоретическое изучение структурно-смысловой организации англоязычной научной медицинской публикации выполняется с позиций перспектив его прикладного использования для разработки инструмента автоматической смысловой компрессии, более эффективного относительно существующих с аналогичными функциями. Такой подход продиктован насущной проблемой обеспечения неограниченного доступа белорусских врачей и научных работников медицинской сферы к новой международно-значимой профессиональной информации с тем, чтобы повысить качество медицинской практики в Республике Беларусь и странах СНГ.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Связь работы с научными программами (проектами), темами.** Работа входит в программу научных исследований, проводимых в Минском государственном лингвистическом университете в рамках госбюджетных НИР: «Структурно-семантический анализ текста с использованием методов корпусной лингвистики (на материале английских, немецких, французских, белорусских и русских текстов)» (2009–2013 гг., № ГР 20090428), «Статистические данные об употребительности в исследуемых текстах различных текстовых единиц (информационные технологии извлечения информации из текстов и формальное представление этой информации)» (2014–2018 гг., № ГР 20142330), «Коммуникативно-прагматические категории дискурса в лингвокультурном аспекте языковой реализации (на материале белорусского, русского и английского языков)» (2016–2020 гг., № ГР 20160986), «Лингвистическая специфика компьютерного опосредования коммуникации» (2019–2023 гг., № ГР 20191959).

**Цель, задачи, объект и предмет исследования.** Целью исследования является разработка лингвистического обеспечения и основанной на нем системы автоматической компрессии научных медицинских публикаций на английском языке, направленной на существенное уменьшение их объема с сохранением основного авторского смысла.

Достижение данной цели потребовало решения следующих **задач** теоретического и прикладного характера:

1) определить лингвистические особенности научных медицинских публикаций на английском языке с учётом восприятия адресата – специалиста-

неносителя данного языка для их дальнейшего учёта при отборе ключевой информации;

2) выделить базовую структурную схему научных медицинских публикаций, необходимую для разработки системы автоматической смысловой компрессии текста;

3) раскрыть структурно-смысловую организацию текста научной медицинской публикации на английском языке;

4) разработать лингвистическое обеспечение системы автоматической компрессии текста, позволяющее сохранить основной смысл публикации;

5) реализовать и протестировать систему автоматической смысловой компрессии публикаций научного медицинского дискурса.

**Объектом** исследования являются научные медицинские публикации на английском языке. **Предмет исследования** составляет структурно-смысловая организация научных медицинских публикаций на английском языке, служащая для автоматического сжатия текста без утраты исходного смысла.

**Материалом** для исследования исходно послужили 100 научных медицинских статей, опубликованных в ведущих англоязычных журналах за 2017–2022 гг.: «The Lancet», «The British Medical Journal (BMJ)», «The New England Journal of Medicine», «The Journal of the American Medical Association (JAMA)». Эти издания характеризуются самыми высокими в мире показателями наукометрии, включая квартиль и импакт-фактор.

**Эмпирическую базу исследования** составили статьи из трёх ключевых областей современной медицины в международной практике: кардиология (36 текстов), пульмонология (32 текста), гастроэнтерология (32 текста).

**Методы исследования.** Для достижения цели исследования использован комплексный подход, включающий следующие **методы**: структурный, контекстуальный и когнитивно-семантический анализ, метод компонентного анализа, метод сравнительного анализа, метод анализа словарных дефиниций, анкетирование, классификационный метод, количественный анализ.

**Научная новизна** данного исследования заключается в следующем:

– определены языковые барьеры, усложняющие декодирование содержания англоязычных медицинских публикаций для иноязычных адресатов-специалистов – белорусских врачей и научных работников медицинской сферы деятельности;

– разработана классификация дискурсивных разделов научной медицинской публикации на английском языке, позволяющая выделить содержательно-значимые элементы для автоматической обработки;

– уточнены понятия субструктуры, суперструктуры, микроструктуры и макроструктуры применительно к созданию лингвистического обеспечения системы автоматической смысловой компрессии научных медицинских публикаций на английском языке, а также их взаимосвязи в процессе автоматической компрессии текста;

– представлена универсальная поверхностная схема научной медицинской публикации, выступающая в качестве базовой субструктуры, обеспечивающей идентификацию начальной и конечной точек смысловых частей текста, предназначенных для компрессии;

– выявлена логика построения суперструктуры научных медицинских публикаций как глубинного каркаса, организующего смысловое наполнение текста;

– определены дискурсивные маркеры как лингвистические опоры, составляющие поверхностную микроструктуру текста и указывающие на ключевые элементы содержания;

– проведена классификация дискурсивных маркеров в зависимости от их функций в компонентах суперструктуры при формировании макроструктуры текста;

– раскрыто понятие макроструктуры как эмерджентной проекции целостного смыслового представления текста публикации на глубинном уровне его организации, создаваемого через взаимодействие компонентов суперструктуры с лингвистическими опорами микроструктуры;

– разработано лингвистическое обеспечение для автоматической смысловой компрессии публикаций научного медицинского дискурса;

– автоматически создаваемый реферат научной медицинской публикации представлен как результат иерархического взаимодействия четырех типов структур, функционирующих на поверхностном и глубинном уровнях, последовательно чередующихся для извлечения в ней смысла, заложенного автором;

– создана автоматическая компьютерная программа MedCompress, являющаяся продуктом прикладной лингвистики, которая доказала свою эффективность в существенном сокращении времени обработки научных медицинских публикаций на английском языке при достижении высокого качества содержания вторичного текста и очевидных преимуществ по сравнению с существующими программами, имеющими аналогичные функции.

#### **Положения, выносимые на защиту.**

1. Обработка русскоязычным адресатом содержания публикаций научного медицинского дискурса осложняется факторами двух типов:

а) лингвистическими: профессионально детерминированным языковым инвентарем, постоянное обновление которого исключает его использование для алгоритмов компрессии текста, и конструкциями, осложненными собственно англоязычными грамматическими структурами, требующими упрощения; б) субъектно-маркированными: спорадическим использованием проформ и языковыми особенностями авторски-индивидуального самовыражения, не несущими важной для понимания смысла публикации информации и подлежащими элиминации из лингвистической базы данных, предназначенных для автоматической компрессии.

2. Научные медицинские публикации на английском языке обладают унифицированной базовой схемой — субструктурой, элементы которой по признаку функциональности разделяются на 1) реквизитные — используемые для формальной идентификации первичного документа; 2) содержательные — представляющие информацию непосредственно по обсуждаемой проблеме; 3) акцессорные — являющиеся вспомогательно-справочным аппаратом публикации. На поверхностном уровне анализа субструктура предоставляет возможность приведения статей к единому формату, применимому ко всему исследуемому корпусу с целью формализации.

3. С опорой на субструктуру, составляющую поверхностный базис для выделения содержательной части текста, в структурно-смысловой организации публикаций научного медицинского дискурса выделяется суперструктура — глубинный логический каркас содержания публикации. Суперструктура определяет работу анализирующего модуля, в котором идёт поиск информации в исходном тексте, релевантной в порождаемом вторичном.

4. Алгоритмическое распознавание компонентов суперструктуры научной медицинской публикации на английском языке осуществляется с помощью элементов микроструктуры — дискурсивных маркеров, которыми выступают единицы второстепенной семантики. Дискурсивные маркеры по их функциональному назначению классифицируются на четыре категории: 1) маркирующие особенности авторского замысла; 2) указывающие на значимость характеристик содержания; 3) направляющие на логику изложения авторской мысли; 4) темпоральные и пространственные дейктические средства. Дискурсивные маркеры сочетаются в зависимости от компонента суперструктуры и, повторяясь в разных публикациях, формируют микроструктуру текста.

5. Смысловую компрессию англоязычных публикаций научного медицинского дискурса реализует автоматическая компьютерная программа MedCompress (Medical Compression) с шестью модулями: 1) контролируемый выбор пользователем информации для выполнения запросно-

ориентированного реферата и алфавитно-частотного словаря; 2) предварительная обработка текста; 3) выделение структур поверхностного и глубинного уровней; 4) выделение ключевых концепций смысла публикации; 5) генерация реферата; 6) построение частотного словаря. Лингвистическую базу знаний АКП MedCompress составляют: 1) установленные в ходе исследования элементы субструктуры, суперструктуры и микроструктуры, иерархическое взаимодействие которых используется для автоматического создания макроструктуры – эмерджентного образования нового уровня организации, содержащего проекцию целостного смыслового представления текста; 2) разработанный комплекс правил, используемых для обработки полученной макроструктуры с целью построения реферата – краткого изложения основного смысла публикации.

Программа обладает модульностью, эмерджентностью и полифункциональностью, обеспечивающими высокое качество автоматически создаваемого вторичного текста научной медицинской публикации.

**Личный вклад соискателя ученой степени.** Диссертация является самостоятельным и завершенным научным исследованием. Теоретические выводы, положения, выносимые на защиту, и практические результаты получены соискателем лично и являются итогом самостоятельной исследовательской работы, выполненной на основе создания теоретической концепции, ее прикладной реализации с использованием значительного фактического материала и широкой апробации полученного прикладного продукта в научной медицинской среде Республики Беларусь.

**Апробация диссертации и информация об использовании ее результатов.** Основные результаты диссертации были представлены в форме докладов на XLIII Международной филологической конференции (Санкт-Петербург, 2014 г.), на XLV Международной филологической конференции (Санкт-Петербург, 2016 г.), на Международной научной конференции «Молодые ученые в инновационном поиске» (Минск, 2019 г., 2021 г.), на LI Международной научной филологической конференции имени Л. А. Вербицкой (Санкт-Петербург, 2024 г.), на V Международной научной конференции по инженерной и прикладной лингвистике «Пиотровские чтения 2024» (С.-Петербург, 2024 г.). Автоматическая программа смысловой компрессии научных медицинских публикаций MedCompress прошла апробацию на базе 2 высших медицинских учреждений образования Республики Беларусь (2021–2024 гг.). Вся перечисленная деятельность подтверждается 7 актами о внедрении и практическом использовании полученных в исследовании результатов (2021–2024 гг.).



**Опубликованность результатов диссертации.** Основные положения и результаты диссертационного исследования отражены в 26 публикациях автора: 8 статьях в рецензируемых журналах (общим объемом 3,6 авт. л.), 6 статьях в сборниках научных статей (общим объемом 1,4 авт. л.) и 12 публикациях в виде материалов научных конференций (общим объемом 2,6 авт. л.).

**Структура и объем диссертации.** Работа состоит из введения, общей характеристики, трёх глав, заключения, библиографического списка и приложений. Общий объем диссертации составляет 203 с., из которых основной текст занимает 135 с., библиографический список – 20 с., приложения – 48 с. Библиографическая часть включает список использованной литературы на английском и русском языках (136 позиций), список публикаций соискателя (26 позиций), список источников иллюстративного материала (40 позиций). Таблицы (26) и рисунки (11) находятся непосредственно в тексте диссертации и занимают 26 страниц. Приложения содержат иллюстрацию исходного текста оригинальной исследовательской статьи (А); иллюстрацию вторичного текста, созданного с помощью АКП MedCompress (Б); инвентарь дискурсивных маркеров микроструктуры, функционирующих в компонентах суперструктуры (В); блок-схемы, отражающие работу АКП MedCompress по созданию реферата и алфавитно-частотного словаря научной медицинской публикации на английском языке (Г); данные об условиях включения предложений в реферат (Д); данные об условиях невключения предложений в реферат (Е); сведения о внедрении и практическом использовании результатов исследования (Ж).

## ОСНОВНАЯ ЧАСТЬ

Первая глава **«Проблема обработки содержания публикаций научного медицинского дискурса для смысловой компрессии»** состоит из пяти разделов, в которых представлена теоретическая база исследования.

В *разделе 1.1* рассматриваются подходы к интерпретации ключевого в современной лингвистике понятия *дискурс*. В результате анализа представленных в научной литературе точек зрения, отраженных в концептуально значимых теоретических дискуссиях (Н. Ф. Алефиренко, Н. Д. Арутюнова, Э. Бенвенист, В. З. Демьянков, В. И. Карасик, В. В. Красных, В. Г. Костомаров, Е. С. Кубрякова, Э. М. Пешё, П. Серио, М. Фуко, В. Е. Чернявская, М. Agar и др.), в данном исследовании выработано определение медицинского дискурса, под которым понимается

совокупность профессиональных стереотипов, выступающих в качестве параметров текстообразования и детерминирующих способ концептуализации проблемы, обсуждаемой его участниками.

Для успешного анализа публикаций научного медицинского дискурса с целью смысловой компрессии необходимым является применение субъектно-ориентированного подхода, раскрывающего дискурс с позиций трёх субъектов: 1) с позиций адресанта как когнитивно обусловленный контекст передачи нового знания в социально-актуальном контексте профессионального общения; 2) с позиций адресата как модель референтной ситуации, наличие которой в сознании позволяет реконструировать содержание заложенного в текст смысла; 3) с позиций исследователя как текст, подвергнутый анализу для применения лингвистических знаний с целью решения прикладных задач: выявления схем построения текста, его смылосодержащих структур и используемых для этого языковых средств.

*Раздел 1.2* посвящен изучению научной статьи как особого жанра медицинского дискурса, динамично распространяющего результаты новейших исследований по оказанию качественной профессиональной помощи пациенту (Н. В. Карацева, Ж. Н. Макушева, Л. А. Абраменко, М. А. Косычева, И. И. Торубарова, И. Ф. Шамара и др.).

Приведены результаты двухэтапного анкетирования 25 отечественных специалистов медицинского профиля с целью выявления трудностей при смысловой обработке текста русскоязычным адресатом. В результате выявлены факторы двух порядков, препятствующие пониманию авторского смысла в англоязычной публикации: 1) лингвистические (профессионально детерминированный языковой инвентарь и конструкции, осложненные собственно англоязычными грамматическими структурами) и 2) субъектно-маркированные (спорадическое использование проформы и языковые особенности авторски-индивидуального самовыражения).

В содержании *раздела 1.3* представлена параметрическая классификация англоязычных научных медицинских статей, публикуемых в четырёх ведущих и наиболее цитируемых журналах, послуживших источниками материала для настоящего исследования.

Полученные данные демонстрируют существенные различия 6 типов медицинских публикаций, что диктует необходимость отбора для дальнейшего изучения оригинальных исследовательских статей (*original research article*), как отражающих ключевые этапы новых клинических исследований с подробным описанием проводимых экспериментов.

В разделе 1.4 показана необходимость выделения базисной поверхностной схемы построения оригинальных исследовательских статей для формализации условий первичной ориентации в имеющемся объеме научной информации. Полученная в результате работы *субструктура* научной медицинской публикации с позиций ведущего для формализации признака функциональности разделов включает: реквизитные – используемые для формальной идентификации первичного документа, содержательные – представляющие информацию непосредственно по обсуждаемой проблеме, и аксессуарные – являющиеся вспомогательно-справочным аппаратом публикации. *Субструктура* обеспечивает техническую возможность создания программного обеспечения для автоматизации процесса установления содержательных разделов, подлежащих смысловой компрессии.

Раздел 1.5 посвящен установлению характеристик вторичного текста, соответствующего задачам выполнения автоматической смысловой компрессии научной медицинской публикации. В результате проведенного анализа выявлена необходимость построения вторичного текста в виде реферата, характеристики которого отвечают следующим восьми критериям: 1) по средствам, используемым при обработке текста, – *машинный*; 2) по процедуре создания – *экстрактивный*; 3) по виду представления информации – *текстовый*; 4) по тематической ориентации – *ориентированный на предметную область* медицины; 5) по отношению к интересам пользователя – *как общий*, так и *запросно-ориентированный*; 6) по характеру и детальности содержания создаваемого документа – *информативный*; 7) по количеству входных документов – *однодокументный*; 8) по степени сжатия – *расширенный*.

Вторая глава «**Научная медицинская публикация как объект анализа с позиций её структурной организации**» состоит из трёх разделов, где выявляется структурно-смысловая организация текста научной медицинской публикации в перспективе реализации автоматической компрессии содержания и определяются способы языковых средств маркирования такой организации в англоязычном дискурсе.

В разделе 2.1 рассматриваются подходы к изучению структурно-смысловой организации текста (Е. А. Баженова, Т. М. Дридзе, Н. И. Жинкин, Б. А. Маслов, А. И. Новиков, И. П. Севбо, П. А. Бех). Для реализации смысловой компрессии текста особый интерес представляет теоретическая концепция Т. А. ван Дейка, согласно которой в построении дискурса с учетом положений когнитивной науки следует выделять три типа структур: 1) глобальные макроструктуры, отражающие содержательное наполнение текста; 2) формальные суперструктуры, представляющие собой схемы

построения текстов определенного дискурса; 3) локальные микроструктуры структуры, функционирующие на различных лингвистических уровнях.

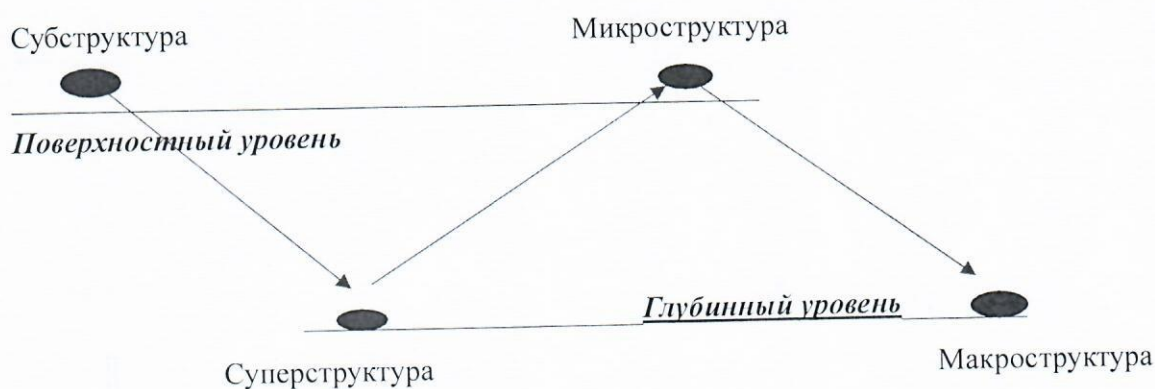
Продуктивность концепции проявилась в том, что она в своей части формальной суперструктуризации была применена к разработке когнитивно-дискурсивного подхода (Т. П. Карпилович), который «предполагает всестороннее освещение языковых явлений с точки зрения их когнитивных и коммуникативных функций». На протяжении двух последних десятилетий эта идея успешно применялась в ряде лингвистических исследований, в задачи которых входило воплощение когнитивно-дискурсивного подхода к выявлению особенностей структурирования текстов разных жанров (Т. П. Карпилович, Н. С. Сычевская, Т. А. Сысоева, В. Ю. Якубова, А. К. Шевцова, П. П. Глазко, Н. В. Егоров). При всей важности полученных в них результатов весь потенциал концепции до сих пор остается не задействованным. Прикладные цели настоящего исследования требуют выделения не только суперструктуры научных медицинских публикаций, но и исследования микро- и макроструктурной составляющих, дополнения их инвентаря и уточнения распределения типов структур по рангам иерархии.

В *разделе 2.2* выявляется суперструктура научной медицинской публикации на английском языке с опорой на субъектно-ориентированный подход: анализ иллюкативных интенций автора, проведенный исследователем дискурса; изучение исследователем инструкций редакционных комитетов англоязычных научных журналов, рекомендуемых адресантам дискурса; привлечение исследователем к определению структурных частей текста научной медицинской публикации адресатов дискурса – отечественных специалистов научно-медицинского профиля. По результатам анализа в 4 содержательных разделах субструктуры с названиями «Введение», «Методы», «Результаты» и «Обсуждение» выделено от 3 до 5 субординативных компонентов суперструктуры, фиксирующих глубинные в плане раскрытия содержания подразделы статьи.

Проведенный в *разделе 2.3* анализ способов англоязычной вербализации компонентов суперструктуры научной медицинской публикации показал, что каждый из них вербализуется в тексте дискурсивными маркерами – элементами *микроструктуры*, которые являются поверхностными опорами, указывающими на глубинный смысл статьи. Дискурсивные маркеры классифицируются по критерию их функционального назначения на 4 группы, реализуя коммуникативные задачи эксплицируемого компонента суперструктуры.

Третья глава «Реализация автоматической смысловой компрессии научной медицинской публикации» посвящена воплощению прикладной задачи исследования.

В разделе 3.1 с использованием иерархически организованных типов структур англоязычных научных медицинских публикаций разработана процедура выполнения автоматической смысловой компрессии (см. рисунок 1).



**Рисунок 1 – Уровневая процедуральность реализации иерархически организованных типов структур англоязычных научно-медицинских публикаций при автоматическом выполнении их смысловой компрессии**

Рисунок 1 иллюстрирует процедуру компрессии, которая инициируется субструктурой, реализованной на поверхностном уровне анализа публикации. Внутри 4 разделов субструктуры на глубинном уровне выстраивается суперструктура – логический каркас текста, включающий 16 компонентов. Вербализацию компонентов суперструктуры на поверхностном уровне осуществляют 243 дискурсивных маркера микроструктуры – дейктические единицы второстепенной семантики.

Иерархическое взаимодействие суб-, супер- и микроструктур является процессом, ведущим к выполнению макроструктуры – эмерджентного образования глубинного уровня, содержащего проекцию целостного смыслового представления текста.

Реферат строится на основе макроструктуры статьи с учетом 27 разработанных условий сжатия текста. Каждая фраза реферата отражает в сокращенном виде представленные в макроструктуре ключевые аспекты содержания, обеспечивая тем самым понимание адресатом основного смысла исходной публикации.

В разделе 3.2 описывается автоматическая компьютерная программа Medical Compression (АКП MedCompress), созданная на базе разработанного лингвистического обеспечения для автоматической реализации смысловой

компрессии научных медицинских публикаций на английском языке с возможностью составления лексического минимума специальности. Производственный процесс реализации работы АКП MedCompress может быть представлен последовательно реализуемыми модулями перцептивной активности, каждый из которых отвечает за выполнение конкретной задачи:

1. Интерактивный модуль.
2. Модуль предварительной обработки текста.
3. Модуль анализа структуры.
4. Модуль выделения ключевых концепций.
5. Модуль сжатия и генерации реферата.
6. Модуль создания частотного словаря.

Взаимодействие пользователя с автоматической программой происходит в интерактивном модуле, где производится регистрация, загрузка требующих обработки текстов, введение списка минус-слов для исключения их из получаемого словаря и определение типа генерируемого реферата.

Модули 2–6 представляют непосредственный автоматический анализ текста – процесс автоматического производства реферата и составления его алфавитно-частотного словаря.

Завершающим этапом исследования, описанным в *разделе 3.3*, выступило тестирование АКП MedCompress с привлечением 23 специалистов медицинского профиля, владеющих английским языком. Реализация эксперимента для расчета эффективности автоматической обработки статей продемонстрировала увеличение производительности труда в 4 раза при получении высокоинформативных рефератов. Выяснено, что скорость обработки содержания исходного текста публикации при помощи АКП MedCompress в 240 раз выше, чем при ручном экспертном анализе статей. Информативность рефератов, полученных при помощи разработанной программы компрессии, составила 90%. Сравнительный анализ системы автоматического реферирования MedCompress с другими компьютерными программами, имеющими схожие функции, в соответствии требованиям, предъявляемым пользователями – белорусскими врачами и научными работниками медицинской сферы, продемонстрировал её очевидные преимущества: загрузка и одновременная обработка до 5 научных медицинских статей, построение лексического минимума специальности, перевод реферата исходной научной медицинской публикации на язык специалиста-носителя другого языка, возможность скачивания и сохранения полученных данных.

## ЗАКЛЮЧЕНИЕ

### Основные научные результаты диссертации

Проведенное исследование проблемы разработки лингвистически обоснованного подхода к созданию системы автоматической смысловой компрессии научных медицинских публикаций на английском языке позволяет сделать следующее заключение.

1. Процесс осмысления адресатом пространства текста англоязычной научной медицинской публикации с целью обнаружения новой профессионально ценностной для него информации осложняется факторами двух порядков, выявленными в ходе анкетирования специалистов-носителей русского языка:

1) лингвистическими, представленными профессионально-детерминированным языковым инвентарем: трудно декодируемыми аббревиатурами, сложносоставными терминологическими выражениями, профессионально-специфичными единицами, подверженными разного вида усечениям (100% респондентов), а также конструкциями, осложненными собственно англоязычными грамматическими структурами (98% респондентов);

2) субъектно-маркированными, выраженными спорадическим использованием проформы (87% респондентов) и широким наличием языковых особенностей авторски-индивидуального самовыражения (82% респондентов).

На этом основании выявлена актуальность разработки системы автоматической смысловой компрессии научных медицинских публикаций с учётом следующих особенностей: невозможности использования ключевого словарного состава в силу его избыточности и постоянного пополнения, необходимости упрощения предложений, осложненных собственно англоязычными грамматическими структурами, элиминации из лингвистической базы субъектно-маркированных факторов как не имеющих непосредственной связи с содержанием публикации [1; 2; 7; 13; 14; 17; 20; 21; 22; 23; 26].

2. В структурно-семантической организации научных медицинских публикаций на английском языке выявлена унифицированная базовая схема — субструктура, которая на поверхностном уровне анализа позволяет привести статьи к единому формату, применимому ко всему исследуемому корпусу текстов. С учетом функционального критерия элементы субструктуры подразделяются на три группы: 1) реквизитные, выполняющие функцию

формальной идентификации первичных документов; 2) содержательные, передающие информацию по рассматриваемой проблеме; 3) акцессорные, которые служат вспомогательным и справочным материалом текста публикации.

Основное назначение субструктуры заключается в том, что она, выступая в качестве первичной структуризации публикации, обеспечивает техническую возможность исключения из дальнейшей обработки реквизитных и акцессорных элементов, а в оставленной для автоматической смысловой компрессии содержательной части позволяет определить начальные и конечные точки разделов «Введение», «Методы», «Результаты», «Обсуждение» [3; 8; 11; 12; 16; 18; 19].

3. Логический каркас содержания на глубинном уровне представлен суперструктурой, включающей 16 компонентов, которые распределяются от 3 до 5 по каждому из четырех содержательных разделов субструктуры: 1. *Введение*: 1.1 Определение поля исследования, 1.2 Обзор опубликованных работ по теме, 1.3 Постановка проблемы, 1.4 Представление цели исследования; 2. *Методы*: 2.1 Характеристика типа исследования, 2.2 Описание материала исследования, 2.3 Описание процедур, 2.4 Представление анализа эксперимента; 3. *Результаты*: 3.1 Основные исходы исследования, 3.2 Дополнительные исходы исследования, 3.3 Нежелательные явления; 4. *Обсуждение*: 4.1 Резюме основного результата, 4.2 Сравнение с другими исследованиями, 4.3 Ограничения исследования, 4.4 Внешняя валидность, 4.5 Перспективы исследования [5; 6; 8; 15; 24; 25].

4. Выявленные в содержательных разделах публикации 243 элемента микроструктуры (58, 81, 26, 78 соответственно) тесно связаны с выделением автором значимых фрагментов в изложении содержания и классифицируются на четыре группы:

1) маркирующие особенности авторского замысла (*to our knowledge* ‘насколько нам известно’, *inclusion criteria* ‘критерии включения’);

2) указывающие на значимость характеристик содержания (*primary endpoint* ‘первичные результаты лечения’, *randomized* ‘рандомизированный’);

3) направляющие на логику изложения авторской мысли (*first* ‘во-первых’, *second* ‘во-вторых’);

4) темпоральные и пространственные дейктические средства (*currently* ‘в настоящее время’, *in Europe* ‘в Европе’).

Повторяемость в текстах публикаций и конечное количество вариантов позволяют использовать инвентарь элементов микроструктуры для алгоритмического распознавания компонентов суперструктуры научной



медицинской публикации на английском языке с целью отбора информации, наиболее семантически важной для построения реферата [5; 6; 8; 24; 25].

5. Действенным инструментом смысловой компрессии научных медицинских публикаций на английском языке является разработанная и апробированная автоматическая компьютерная программа MedCompress (Medical Compression), предоставляющая возможность получения машинного, текстового, информативного, общего или запросного, ориентированного на предметную область медицины реферата, а также составления лексического минимума областей медицины «Кардиология», «Пульмонология», «Гастроэнтерология».

Основой реализации смысловой компрессии является последовательное иерархическое взаимодействие структур поверхностного и глубинного уровней, где выявленные исследователем в ходе анализа элементы субструктуры, суперструктуры и микроструктуры задаются в базу знаний АКП MedCompress для автоматического создания макроструктуры – эмерджентного образования глубинного уровня организации. Макроструктура содержит проекцию целостного смыслового представления текста, соответствующего ходу развертывания 16 компонентов его суперструктуры. Последующее применение к полученной макроструктуре разработанного комплекса правил дополнительного сжатия обеспечивает создание реферата – краткого изложения основных смысловых элементов обрабатываемой публикации.

Функциональность процесса реализации программы представлена шестью последовательно реализуемыми модулями, каждый из которых отвечает за выполнение конкретной задачи: 1) контролируемый выбор пользователем информации для выполнения запросно-ориентированного реферата и алфавитно-частотного словаря; 2) предварительная обработка текста; 3) выделение структур поверхностного и глубинного уровней; 4) выделение ключевых концепций смысла публикации; 5) генерация реферата; 6) построение частотного словаря.

Свойственные АКП MedCompress характеристики – модульность, полифункциональность, эмерджентность – обеспечивают сокращение времени обработки текста в 240 раз по сравнению с ручным методом, уменьшение объема текста в 4 раза, ускорение обработки адресатом вторичного текста по сравнению с исходным в 4 раза при точности достигаемого результата – понимании смысла публикации на 90 %, тем самым обеспечивая четырехкратное увеличение производительности труда. Удостоверительный анализ реферирования англоязычных публикаций научного медицинского дискурса путем сравнения работы АКП MedCompress

с другими компьютерными программами, имеющими аналогичные функции, продемонстрировал её очевидные преимущества в соответствии требованиям, предъявляемым научными работниками медицинской сферы при обращении к новой международно-значимой профессиональной информации с целью ее использования для повышения качества медицинской практики в Республике Беларусь и странах СНГ [4; 5; 6; 8; 9; 10; 24; 25].

#### **Рекомендации по практическому использованию результатов**

1. Материалы и выводы диссертационного исследования могут быть использованы для решения прикладных задач при создании рефератов других типов медицинских научных публикаций.

2. Материалы и выводы проведенного исследования будут востребованы при создании учебных курсов и написании учебных пособий по дискурсу-анализу, когнитивной лингвистике, лингвопрагматике, жанроведению, функциональной стилистике английского языка.

3. Полученные результаты и материал исследования могут быть востребованы в лексикографической практике для регулярного обновления лексических минимумов и составления словарей узких медицинских специальностей.

4. Разработанная методика исследования структуры и семантики англоязычных научных медицинских публикаций может быть применена для анализа других научных жанров медицинского дискурса при написании курсовых и дипломных работ, магистерских и кандидатских диссертаций в аспекте совершенствования профессиональной подготовки специалистов медицинского профиля Беларуси и СНГ.

5. Полученные результаты анализа подязыка медицины могут использоваться для разработки систем автоматической обработки текста, информационно-поисковых систем и системы автоматического перевода медицинских текстов.

6. Материалы и выводы проведенного исследования используются в образовательном процессе в Витебском государственном медицинском университете (г. Витебск, Республика Беларусь), Белорусском государственном медицинском университете (г. Минск, Республика Беларусь), что подтверждается 7 актами об их внедрении в научную работу и учебный процесс на следующих кафедрах: иностранных языков (БГМУ), клинической фармакологии (БГМУ), организации фармации (БГМУ), клинической микробиологии (ВГМУ), педиатрии ПК и ПК (ВГМУ), травматологии и ортопедии (ВГМУ) (см. приложение Ж).

## СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ УЧЕНОЙ СТЕПЕНИ

### Статьи в рецензируемых научных журналах

1. Меньшенина, И. А. Информационная специфика англоязычных медицинских текстов / И. А. Меньшенина // Вестник МГЛУ. Серия 1, Филология. – 2012. – № 5 (60). – С. 26–32.
2. Меньшенина, И. А. Лингвистическая специфика медицинского текста как жанра научного стиля / И. А. Меньшенина // Вестник МГЛУ. Серия 1, Филология. – 2013. – № 5 (66). – С. 33–37.
3. Меньшенина, И. А. Специфика организации и коммуникативно-речевая классификация медицинского англоязычного текста / И. А. Меньшенина // Вестник ВГМУ. – 2014. – Т. 13, № 2. – С. 151–156.
4. Меньшенина, И. А. Принципы отбора лексического материала для учебников по иностранным языкам в медицинских учреждениях высшего образования / И. А. Меньшенина // Вестник МГЛУ. Серия 1, Филология. – 2018. – № 1 (33). – С. 33–37.
5. Меньшенина, И. А. Структура компонента «Введение» в научных медицинских статьях на английском языке / И. А. Меньшенина // Вестник МГЛУ. Серия 1, Филология. – 2020. – № 4 (107). – С. 123–129.
6. Меньшенина, И. А. Формализация раздела «Обсуждение» научной медицинской статьи на английском языке / И. А. Меньшенина // Ученые записки ВГУ имени П. М. Машерова. – 2021. – Т. 34. – С. 151–155.
7. Меньшенина, И. А. Лингвистические особенности научной медицинской статьи на английском языке / И. А. Меньшенина // Ученые записки ВГУ имени П. М. Машерова. – 2022. – Т. 35. – С. 167–172.
8. Меньшенина, И. А. Структурно-смысловая организация публикаций научного медицинского дискурса в свете задач прикладной лингвистики / И. А. Меньшенина // Ученые записки ВГУ имени П. М. Машерова. – 2024. – Т. 40. – С. 158–162.

### Статьи в сборниках научных статей

9. Меньшенина, И. А. О создании лексического минимума англоязычных медицинских текстов / И. А. Меньшенина // Аспирант : сб. науч. тр. аспирантов / Нижегород. гос. лингвист. ун-т им. Н. А. Добролюбова; редкол.: Е. С. Гриценко [и др.]. – Нижний Новгород, 2011. – Вып. 9. – С. 80–84.
10. Меньшенина, И. А. Статистика как основной метод создания лексического минимума / И. А. Меньшенина // Исследования молодых ученых : сб. ст. аспирантов МГЛУ / Мин. гос. лингвист. ун-т ; редкол.: А. М. Горлатов (отв. ред.) [и др.]. – Минск, 2012. – С. 167–170.

11. Меньшенина, И. А. Специфика организации информации в медицинских текстах / И. А. Меньшенина // Проблемы современной прикладной лингвистики : сб. науч. ст. / Мин. гос. лингвист. ун-т ; редкол.: А. В. Зубов (отв. ред.) [и др.]. – Минск, 2014. – С. 134–139.

12. Меньшенина, И. А. Анализ структурных особенностей медицинских текстов / И. А. Меньшенина // Исследования молодых ученых : сб. ст. аспирантов / Мин. гос. лингвист. ун-т ; редкол.: А. М. Горлатов (отв. ред.) [и др.]. – Минск, 2014. – С. 63–65.

13. Меньшенина, И. А. Место статистических методов анализа текста в лингвистической методологии / И. А. Меньшенина // Исследования молодых ученых : сб. ст. аспирантов / Мин. гос. лингвист. ун-т ; редкол.: А. М. Горлатов (отв. ред.) [и др.]. – Минск, 2016. – С. 85–89.

14. Меньшенина, И. А. О подходе к отбору лексического минимума для обучения специальности / И. А. Меньшенина // Актуальные проблемы современной прикладной лингвистики : сб. науч. ст. / Мин. гос. лингвист. ун-т ; редкол.: И. В. Совпель (отв. ред.) [и др.]. – Минск, 2017. – С. 234–236.

#### **Материалы и тезисы докладов научных конференций**

15. Меньшенина, И. А. Семантико-синтаксическая организация медицинских учебных и научных текстов на материале английского языка / И. А. Меньшенина // Достижения фундаментальной, клинической медицины и фармации : материалы 67-й науч. сес. сотр. ун-та, 2–3 февр. 2012 г., г. Витебск / Витеб. гос. мед. ун-т ; редкол.: В. П. Дейкало [и др.]. – Витебск, 2012. – С. 483–484.

16. Меньшенина, И. А. Коммуникативно-речевая классификация медицинских текстов / И. А. Меньшенина // Материалы ежегодной научной конференции преподавателей и аспирантов университета, 25–26 апреля 2012 г., г. Минск : в 5 ч. / Мин. гос. лингвист. ун-т ; редкол.: Н. П. Баранова (отв. ред.) [и др.]. – Минск, 2012. – Ч. 3. – С. 199–201.

17. Меньшенина, И. А. Подходы к отбору лексики для обучения студентов-медиков / И. А. Меньшенина // Достижения фундаментальной, клинической медицины и фармации : материалы 69-й науч. сес. сотр. ун-та, 29–30 янв. 2014 г., г. Витебск / М-во здравоохранения Республики Беларусь, Витеб. гос. мед. ун-т ; редкол.: В. П. Дейкало [и др.]. – Витебск, 2014. – С. 400–401.

18. Меньшенина, И. А. Особенности организации информации в медицинских текстах / И. А. Меньшенина // Материалы XLIII Междунар. филолог. конф., 11–16 марта 2014 г., г. Санкт-Петербург / Санкт-Петербург. гос.

ун-т, филол. фак.; редкол.: М. В. Хохлова (отв. ред.) [и др.]. – Санкт-Петербург, 2014. – С. 247–248.

19. Меньшенина, И. А. Некоторые особенности организации медицинского англоязычного текста / И. А. Меньшенина // Контрастивные исследования и прикладная лингвистика : материалы междунар. науч. конф., 29–30 окт. 2014 г., г. Минск : в 2 ч. / Мин. гос. лингвист. ун-т ; редкол.: А. В. Зубов (отв. ред.) [и др.]. – Минск, 2015. – Ч. 1. – С. 72–77.

20. Меньшенина, И. А. Статистический метод оценки лексического наполнения англоязычных медицинских учебных текстов / И. А. Меньшенина // Достижения фундаментальной, клинической медицины и фармации : материалы 71-й науч. сес. сотр. ун-та, 27–28 янв. 2016 г., г. Витебск / М-во здравоохранения Республики Беларусь, Витеб. гос. мед. ун-т ; редкол.: А. Т. Щастный [и др.]. – Витебск, 2016. – С. 415–416.

21. Меньшенина, И. А. Лингвостатистический метод оценки лексического наполнения учебных текстов // тез. докл. XLV Междунар. филолог. конф., 14–21 марта 2016 г., г. Санкт-Петербург / Санкт-Петербург. гос. ун-т, филол. фак.; редкол.: М. В. Хохлова (отв. ред.) [и др.]. – Санкт-Петербург, 2016. – С. 643–644.

22. Меньшенина, И. А. Квантитативный анализ учебников иностранного языка / И. А. Меньшенина // Достижения фундаментальной, клинической медицины и фармации : материалы 72-й науч. сес. сотр. ун-та, 25–26 янв. 2017 г., г. Витебск / М-во здравоохранения Республики Беларусь, Витеб. гос. мед. ун-т ; редкол.: А. Т. Щастный [и др.]. – Витебск, 2017. – С. 570–571.

23. Меньшенина, И. А. Принципы отбора лексического минимума для формирования рецептивного словаря у студентов медицинского вуза / И. А. Меньшенина // Достижения фундаментальной, клинической медицины и фармации : материалы 73-й науч. сес. ВГМУ, 29–30 янв. 2018 г., г. Витебск : в 2 ч. / М-во здравоохранения Республики Беларусь, Витеб. гос. мед. ун-т ; редкол.: А. Т. Щастный [и др.]. – Витебск, 2018. – Ч. 2. – С. 763–766.

24. Меньшенина, И. А. Формализация распознавания компонента «Методы» в научных медицинских статьях на английском языке / И. А. Меньшенина // Молодые ученые в инновационном поиске : тез. докл. VIII Междунар. науч. конф., 29–30 мая 2019 г., г. Минск / Мин. гос. лингвист. ун-т ; редкол.: Т. П. Карпилович (отв. ред.) [и др.]. – Минск, 2020. – С. 211–216.

25. Меньшенина, И. А. Формализация распознавания компонента «Результаты» в научных медицинских статьях на английском языке / И. А. Меньшенина // Молодые ученые в инновационном поиске : тез. докл. X Междунар. науч. конф., 11–12 марта 2021 г., г. Минск / Мин. гос. лингвист.

ун-т ; редкол.: Т. П. Карпилович (отв. ред.) [и др.]. – Минск, 2022. – С. 280–285.

26. Меньшенина, И. А. Сложности англоязычной медицинской публикации с позиций её смысловой обработки адресатом-носителем русского языка / И. А. Меньшенина // Достижения фундаментальной, клинической медицины и фармации : материалы 79-й науч. сес. ВГМУ, 24–25 янв. 2024 г., г. Витебск / М-во здравоохранения Республики Беларусь, Витеб. гос. мед. ун-т ; редкол.: А. Н. Чуканов (гл. ред.) [и др.]. – Витебск, 2024. – С. 352–353.

**Компьютерная ссылка для доступа к АКП MedCompress**  
<https://medcompress.of.by>

## РЕЗЮМЕ

Меньшенина Ирина Александровна

### **Автоматическая смысловая компрессия публикаций научного медицинского дискурса (на материале английского языка)**

**Ключевые слова:** научный медицинский дискурс, субъектно-ориентированный анализ, структурно-смысловая организация, типы смысловых структур, функции дискурсивных маркеров, смысловая компрессия, автоматическое реферирование

**Целью** исследования является разработка лингвистического обеспечения и основанной на нем системы автоматической компрессии научных медицинских публикаций на английском языке, направленной на существенное уменьшение их объема с сохранением основного авторского смысла.

**Методы исследования:** структурный, контекстуальный и когнитивно-семантический анализ, метод компонентного анализа, метод сравнительного анализа, метод анализа словарных дефиниций, анкетирование, классификационный метод, количественный анализ.

**Полученные результаты и их новизна.** В результате теоретического анализа выявлена структурно-смысловая организация научной медицинской публикации: субструктура, суперструктура, микроструктура, составившие лингвистическое обеспечение для разработки автоматической компьютерной программы MedCompress. Модульность и полифункциональность программы позволяют по заложенным в нее алгоритмам выполнять компрессию текста публикации с сохранением основного авторского смысла. Выделенная автоматическим способом эмерджентно-возникающая макроструктура публикации служит основой для вербального построения получаемого реферата с экономией времени в 240 раз при экспериментально доказанном качестве содержания с информативностью 90%. Сравнительный анализ АКП MedCompress с другими компьютерными программами, имеющими аналогичные функции, продемонстрировал её значительные преимущества в соответствии требованиям, предъявляемым пользователями.

**Рекомендации по использованию.** Практическое применение результатов исследования, подтвержденное актами внедрения, состоит в расширении их использования в научно-исследовательской работе медицинских специалистов и в учебном процессе медицинских вузов Беларуси и стран СНГ. Разработанная методика исследования научных публикаций по кардиологии, пульмонологии, гастроэнтерологии может быть применена для анализа структуры и семантики публикаций из других областей медицины, их автоматического реферирования и создания лексического минимума.

**Область применения:** прикладная лингвистика, медицина, дискурс-анализ, когнитивная лингвистика, лексикография, лингвопрагматика, жанроведение.

## РЭЗІЮМЭ

Мяньшэніна Ірына Аляксандраўна

### **Аўтаматычная сэнсавая кампрэсія публікацый навуковага медыцынскага дыскурса (на матэрыяле англійскай мовы)**

**Ключавыя словы:** навуковы медыцынскі дыкурс, суб'ектна-арыентаваны аналіз, медыцынскі дыкурс, структурна-сэнсавая арганізацыя, тыпы сэнсавых структур, функцыі дыскурсіўных маркераў, сэнсавая кампрэсія, аўтаматычнае рэфераванне

**Мэтай** даследавання з'яўляецца распрацоўка лінгвістычнага забеспячэння і заснаванай на ім сістэмы аўтаматычнай кампрэсіі навуковых медыцынскіх публікацый на англійскай мове, накіраванай на істотнае памяншэнне іх аб'ёму з захаваннем асноўнага аўтарскага сэнсу.

**Метады даследавання:** структурны, кантэкстуальны і кагнітыўна-семантычны аналіз, метады кампанентнага аналізу, метады параўнальнага аналізу, метады аналізу слоўнікавых дэфініцый, анкетаванне, класіфікацыйны метады, колькасны аналіз.

**Атрыманыя вынікі і іх навізна.** У выніку тэарэтычнага аналізу выяўлена структурна-сэнсавая арганізацыя навуковай медыцынскай публікацыі: субструктура, суперструктура, мікраструктура, якія склалі лінгвістычнае забеспячэнне для распрацоўкі аўтаматычнай камп'ютарнай праграмы MedCompress. Модульнасць і поліфункцыянальнасць праграмы дазваляюць па закладзеным у яе алгарытмах выконваць кампрэсію тэксту публікацыі з захаваннем асноўнага аўтарскага сэнсу. Выдзеленая аўтаматычным спосабам эмерджэнтна-ўзнікаючая макраструктура публікацыі служыць асновай для вербальнай пабудовы атрыманага рэферата з эканоміяй часу ў 240 разоў пры эксперыментальна даказанай якасці зместу з інфарматыўнасцю 90%. Параўнальны аналіз АКП MedCompress з іншымі камп'ютарнымі праграмамі, якія маюць аналагічныя функцыі, прадэманстравала яе значныя перавагі ў адпаведнасці патрабаванням, якія прад'яўляюцца карыстальнікамі.

**Рэкамендацыі па выкарыстанні.** Практычнае прымяненне вынікаў даследавання, пацверджанае актамі ўкаранення, заключаецца ў пашырэнні іх выкарыстання ў навукова-даследчай рабоце медыцынскіх спецыялістаў і ў навучальным працэсе ў медыцынскіх ВНУ Беларусі і краін СНД. Распрацаваная метадыка даследавання навуковых публікацый па кардыялогіі, пульманалогіі, гаэстраэнтаралогіі можа быць выкарастана для аналізу структуры і семантыкі публікацый з іншых абласцей медыцыны, іх аўтаматычнага рэферавання і стварэння лексічнага мінімуму.

**Галіна прымянення:** прыкладная лінгвістыка, медыцына, дыкурс-аналіз, кагнітыўная лінгвістыка, лексікаграфія, лінгвапрагматыка, жанразнаўства.



## SUMMARY

Iryna A. Mianshenina

### **Automatic semantic compression of scientific medical discourse publications (on the material of the English language)**

**Key words:** scientific medical discourse, subject-oriented analysis, structural and semantic organization, types of semantic structures, functions of discourse markers, semantic compression, automatic summarization

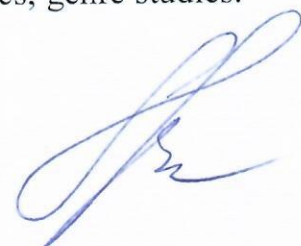
**The aim of the study** is to develop linguistic support and based on it automatic compression system for scientific medical publications in English, aimed at significantly reducing their volume while preserving the main author's information content.

**The methods of the research** include structural, contextual and cognitive-semantic analysis, componential analysis, the method of comparative analysis, analysis of dictionary definitions, questionnaires, classification method, quantitative analysis.

**The results obtained and their novelty.** As a result of the theoretical analysis, the structural and semantic organization of the scientific medical publication was revealed: Substructure, Superstructure, Microstructure, which formed the linguistic support for the development of the automatic computer program MedCompress. The modularity and multifunctionality of the program allow, according to the algorithms embedded in it, to compress the text of the publication with preservation of the main author's information content. The emergently performing macrostructure of the publication, isolated in an automatic way, serves as the basis for the verbal construction of the resulting abstract with a time saving of 240 times with experimentally proven content quality with an informative value of 90%. A comparative analysis of ACP MedCompress with other computer programs that have similar functions demonstrated its significant advantages in meeting the requirements of users.

**Recommendations on the use of the results obtained.** The practical application of the research results, confirmed by the implementation acts, is to expand their use in the research work of medical specialists and in the educational process at medical universities in Belarus and the CIS countries. The developed methodology for the study of scientific publications on cardiology, pulmonology, gastroenterology can be used to analyze the structure and semantics of publications from other fields of medicine, their automatic summarization and creation of lexical minimum.

**The sphere of application:** applied linguistics, medicine, discourse analysis, cognitive linguistics, lexicography, linguistic pragmatics, genre studies.



Научное издание

**Меньшенина Ирина Александровна**

**АВТОМАТИЧЕСКАЯ СМЫСЛОВАЯ КОМПРЕССИЯ ПУБЛИКАЦИЙ  
НАУЧНОГО МЕДИЦИНСКОГО ДИСКУРСА  
(на материале английского языка)**

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата филологических наук  
по специальности 10.02.21 – прикладная  
и математическая лингвистика

Ответственный за выпуск *И. А. Меньшенина*

Подписано в печать 28.02.2025 г. Формат 60x84 <sup>1</sup>/<sub>16</sub>.

Бумага офсетная. Гарнитура Таймс. Ризография.

Усл. печ. л. 1,86. Уч.-изд. л. 1,56. Тираж 100 экз. Заказ 209.

Издатель и полиграфическое исполнение учреждение образования  
«Витебский государственный медицинский университет»

Свидетельство о государственной регистрации издателя, изготовителя,  
распространителя печатных изданий № 3/630 от 03.04.2014.

ЛП №02330/453 от 30.12.2013

пр-т Фрунзе, 27, 210023, г. Витебск