

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

о диссертации МЕНЬШЕНИНОЙ Ирины Александровны
«Автоматическая смысловая компрессия публикаций научного медицинского дискурса (на материале английского языка)»,
представленной на соискание ученой степени кандидата филологических наук
по специальности 10.02.21 – прикладная и математическая лингвистика

1. Соответствие диссертации специальности и отрасли науки, по которым она представлена к защите

Диссертационная работа И.А.Меньшениной направлена на совершенствование процесса автоматизации обработки и анализа текстов научных публикаций в условиях постоянного роста объема информации. Представленное исследование посвящено разработке метода, лингвистического обеспечения и алгоритма решения задачи формализации смысловой компрессии англоязычных научных медицинских публикаций с применением субъектно-ориентированного подхода. Тем самым содержание диссертационного исследования И.А.Меньшениной полностью соответствует отрасли науки (филология) и специальности 10.02.21 – прикладная и математическая лингвистика.

2. Актуальность темы диссертации.

В соответствии с Государственной программой «Цифровое развитие Беларуси» на 2021-2025 гг. основной целью цифрового развития является обеспечение внедрения информационно коммуникационных и передовых производственных технологий в отрасли национальной экономики и сферы жизнедеятельности общества. Таким образом, актуальность исследования, выполненного в русле одной из приоритетных и государственно значимых задач, обусловлена необходимостью оперативной и эффективной обработки и анализа медицинских текстов на английском языке в целях обеспечения их доступности для специалистов медицинской сферы.

Особую важность выбранное автором направление исследования приобретает благодаря разработке такого подхода к решению проблемы смысловой компрессии медицинских публикаций, который позволил обеспечить вторичный текст предельной точности, терминологической представленности и исчерпывающей полноты в изложении профессионально значимой информации в целях ее безошибочной интерпретации специалистами.

В свете названной проблемы в центре внимания исследования И.А.Меньшениной оказывается актуальным вопрос разработки лингвистической базы знаний, позволяющей сохранить основной смысл научной публикации.

Актуальность работы определяется и необходимостью решения задач прикладного характера. Это, прежде всего, создание исследовательского корпуса медицинских научных статей английского языка, опубликованных в ведущих научных журналах, выявление модели структурно-смысловой организации научной публикации в рамках субъектно-ориентированного анализа, определение инвентаря дискурсивных маркеров выделенных автором компонентов суперструктуры.

3. Степень новизны результатов, полученных в диссертации, и научных положений, выносимых на защиту.

Представленное диссертационное исследование обладает достаточной степенью новизны полученных результатов, поскольку в нем впервые ставится и успешно решается задача применения субъектно-ориентированного подхода к концептуальной разработке лингвистического и алгоритмического обеспечения системы автоматической компрессии текстов медицинских научных статей.

В работе впервые разработана методика определения структурно-функциональной организации англоязычных медицинских научных текстов с целью формирования лингвистической базы знаний системы автоматической компрессии текста. В основу разработанной методики положен субъектно-ориентированный подход, позволяющий проводить когнитивный анализ внутренней организации научного текста на поверхностном и глубинном уровнях с учетом вовлеченности трех участников медицинского дискурса: автора статьи, читателя статьи (специалиста в области медицины) и исследователя, владеющего лингвистическими методами исследования и прикладным инструментарием для формализации естественного языка.

В качестве значимых новых результатов, получивших отражение в положениях, выносимых на защиту, следует отметить теоретическое обоснование, лингвистическую базу знаний, алгоритмизацию и экспериментальное тестирование разработанной автором системы автоматической компрессии научных медицинских публикаций, отличительной особенностью которой является использование метода, позволяющего учитывать статистические и лингвистические характеристики естественного языка, что делает возможным повысить эффективность и точность созданной системы.

Таким образом, можно заключить, что полученные автором результаты исследования и научные положения характеризуются новизной.

4. Обоснованность и достоверность выводов и рекомендаций, сформулированных в диссертации, обеспечиваются, прежде всего, аргументированностью концептуально-теоретических предпосылок проведения исследования и проведением всестороннего анализа применяемых методов решения поставленных задач, включая анкетирование специалистов в области медицины, обработку, анализ и интерпретацию полученных данных.

Предлагаемый в диссертации тщательный анализ зарубежных и отечественных работ по теме исследования (библиографический список представлен 136 наименованиями) позволил И.А.Меньшениной объективно оценить состояние проблемы и разработать собственную теоретическую концепцию исследования.

Достоверность полученных результатов обеспечивается:

- аутентичностью и значительным объемом экспериментального материала: исследовательский корпус текстов, в который включены отобранные автором 100 научных статей трех ключевых областей современной медицины, опубликованных в ведущих англоязычных журналах за 2017-2022 гг (при этом автор приводит критерии для отбора текстов в корпус), иерархическая схема выявленной автором структуры текста, выделенные автором элементы микроструктуры текстов (инвентарь

дискурсивных маркеров), которые эксплицируют установленные в процессе исследования компоненты суперструктуры;

- соответствием использованной методики анализа структуры и семантики англоязычных научных медицинских текстов успешно решенным задачам и сформулированной гипотезе.

Подробная и логичная интерпретация полученных экспериментальных данных по результатам проведения тестирования созданной системы MedCompress убедительно доказывает правомерность сделанных автором выводов относительно эффективности и функциональности ее работы, а также преимущество перед другими схожими по выполняемой задаче программами.

Отдельное внимание заслуживают семь приложений к диссертации, содействующие более полному раскрытию содержания и результатов работы.

Особую значимость представляет выявленная автором структурно-семантическая организация научных медицинских статей на поверхностном и глубинном уровнях, которая и послужила основой для реализации системы создания вторичного текста.

Цель работы достигнута, созданная лингвистическая база знаний и разработанная и апробированная система автоматической компрессии англоязычной медицинской научной статьи выступает эффективным инструментом уменьшения объема текста при сохранении авторского смысла публикации на 90% и, как результат, значительного сокращения времени обработки научной публикации. Программа работает автономно, порождает текст общего или запросно-ориентированного реферата.

5. Научная, практическая, экономическая и социальная значимость результатов диссертации с указанием рекомендаций по их использованию

Полученные в ходе исследования результаты весьма существенны с точки зрения их теоретической и практической значимости. Апробированная система автоматической компрессии научного медицинского текста, разработанная автором с учетом представленной теоретико-методологической базы исследования, созданной лингвистической базы данных и алгоритма реализации поставленной задачи, характеризуется высокими качественными показателями для оперативной и точной передачи смысла публикации.

Самостоятельную научную и практическую значимость имеет разработанное автором лингвистическое обеспечение системы автоматической компрессии научного текста, которое включает сформированный исследовательский корпус текстов научных статей трех ключевых областей современной медицины, иерархическая схема выявленных автором 4 структур текста и 16 компонентов суперструктуры, выделенные автором 243 элемента микроструктуры текстов (инвентарь дискурсивных маркеров), которые эксплицируют установленные в процессе исследования компоненты суперструктуры, а также условия включения и невключения предложений в реферат.

Все основные результаты промежуточных этапов работы по созданию лингвистической базы знаний и автоматической системы компрессии текста детально представлены в многочисленных таблицах, примерах, схемах и приложениях, что значительно упрощает восприятие результатов исследования.

Особым достоинством работы является факт внедрения теоретических и практических результатов исследования в учебный (3 акта о внедрении) и научный (4 акта о внедрении) процессы (копии прилагаются):

- внедрение в учебный процесс компьютерной программы автоматического смыслового реферирования научных медицинских статей TEST10 в учреждении образования «Белорусский государственный медицинский университет», кафедра организации фармации (2022 год);

- внедрение в учебный процесс компьютерной программы автоматического смыслового реферирования научных медицинских статей TEST10 в учреждении образования «Белорусский государственный медицинский университет», кафедра клинической фармакологии (2022 год);

- внедрение в учебный процесс компьютерной программы автоматического смыслового реферирования научных медицинских статей TEST10 в учреждении образования «Белорусский государственный медицинский университет», кафедра иностранных языков (2022 год);

- внедрение в научный процесс (для проведения патентного поиска) компьютерной программы автоматической смысловой компрессии публикаций научного медицинского дискурса TEST10 в учреждении образования «Витебский государственный ордена Дружбы народов медицинский университет» (2022 год);

- внедрение в научный процесс (для научного поиска) компьютерной программы автоматической смысловой компрессии публикаций научного медицинского дискурса MedCompress в учреждении образования «Витебский государственный ордена Дружбы народов медицинский университет», кафедра педиатрии (2024 год);

- внедрение в научный процесс (для научного поиска) компьютерной программы автоматической смысловой компрессии публикаций научного медицинского дискурса MedCompress в учреждении образования «Витебский государственный ордена Дружбы народов медицинский университет», кафедра травматологии и ортопедии (2024 год);

- внедрение в научный процесс (для научного поиска) компьютерной программы автоматической смысловой компрессии публикаций научного медицинского дискурса MedCompress в учреждении образования «Витебский государственный ордена Дружбы народов медицинский университет», кафедра клинической микробиологии (2024 год);

Следует отметить, что анализ результатов работы программы, указанных в актах о внедрении, подтверждает полученные автором выводы о высокой скорости обработки данных для компрессии текста при достаточном (90%) сохранении смысла содержания. Замечаний и предложений в актах не зафиксировано, при этом отмечается эффективность программы для осуществления научного поиска, высокая информативность получаемых рефератов, не требующая обращения к полному исходному тексту статьи.

Кроме того, учитывая профиль кафедр медицинских университетов, которые внедрились компьютерную программу в учебных и научных целях, можно сделать важный вывод о том, что система автоматической компрессии, разработанная изначально на материале текстов трех областей медицины (кардиология,

пульмонология и гастроэнтерология), может применяться и для обработки медицинских текстов других областей.

6. Опубликованность результатов диссертации в научной печати

Основные положения и результаты диссертационного исследования в полной мере отражены в 26 публикациях автора, в том числе 8 статьях в рецензируемых научных журналах, 6 статьях в сборниках научных статей, 12 публикациях в сборниках материалов научных конференций. Все публикации выполнены без соавторов. В большинстве случаев (23 публикации) результаты диссертации опубликованы в изданиях общегосударственного уровня. Три публикации изданы в Российской Федерации (Нижний Новгород и Санкт-Петербург). Все публикации нашли отражение в тексте диссертации и автореферате.

Особого внимания заслуживает в перечне публикаций соискателя ученой степени указание компьютерной ссылки для доступа к разработанной системе автоматической компрессии научного текста.

Результаты исследования и положения, выносимые на защиту, нашли отражение в указанных автором публикациях. Автореферат соответствует содержанию основного текста диссертации.

7. Соответствие оформления диссертации требованиям ВАК

Диссертационная работа состоит из введения, общей характеристики работы, трех исследовательских глав, заключения, списка использованных источников, включающего библиографический список, список источников фактического материала и список публикаций соискателя, и семи приложений. Порядок изложения материала в диссертации соответствует сформулированной цели исследования. Все разделы работы характеризуются логичностью и четкостью изложения материала. Оформление диссертационного исследования и автореферата соответствует Инструкции Высшей аттестационной комиссии Республики Беларусь по оформлению диссертации и автореферата.

8. Недостатки диссертации.

Констатируя существенную научную и практическую значимость результатов проведенного исследования, обращаем внимание на некоторые дискуссионные вопросы.

1. На наш взгляд, нуждается в пояснении соотношение терминов, используемых в диссертации, а именно:

- смысловая компрессия/компрессия содержания текста;
- дискурсивные/содержательные разделы/разделы субструктуры;
- дискурсивные / структурные/ суперструктурные/ функциональные/ семантические компоненты.

2. В разделе 1.3. автор указывает на то, что выявленные лингвистические и субъектно-маркированные особенности «характерны для **каждой** научной медицинской статьи на английском языке». Хотелось бы уточнить, что послужило основанием для такого вывода?

3. При выделении суперструктуры каждая статья представлена автором в виде дискурсивных разделов, дискурсивных компонентов, отражающих соответствующие коммуникативные цели (стр.61). В связи с этим возникает вопрос, что позволило

автору выделить приведенные дискурсивные разделы статьи и определить коммуникативные цели автора?

4. На стр.71 автор приводит критерии для отнесения лексической единицы к классу дискурсивных маркеров, однако отсутствует указание, является ли данная классификация собственной разработкой автора.

5. Автором была проделана трудоемкая работа по определению 27 условий дополнительной обработки вторичного текста (приложения Д и Е), однако описание данного этапа работы не нашло отражения в тексте диссертации. В связи с этим, хотелось бы уточнить, как именно проводился анализ для определения данных условий?

9. Соответствие научной квалификации соискателя ученой степени, на которую он претендует.

Структура и содержание диссертации, критический анализ различных точек зрения при работе с литературными источниками, оригинальная, разносторонняя и тщательно продуманная методика решения задач теоретического и прикладного характера, логичная, обоснованная и последовательная интерпретация полученных результатов свидетельствуют о соответствии научной квалификации И.А.Меньшениной ученой степени кандидата филологических наук по специальности 10.02.21 – прикладная и математическая лингвистика.

10. Заключительная часть

Диссертационное исследование И.А.Меньшениной, подготовленное под руководством доктора филологических наук, профессора Н.С.Евчик, является законченной самостоятельно выполненной квалификационной научной работой, содержащей новые научно обоснованные и практико-ориентированные результаты, совокупность которых представляет существенный вклад в совершенствование систем автоматической обработки текста. Содержание работы соответствует специальности 10.02.21 – прикладная и математическая лингвистика.

Автор кандидатской диссертации заслуживает присвоения ученой степени кандидата филологических наук за совокупность следующих научных результатов:

- 1) выявление системы лингвистических особенностей англоязычных научных медицинских текстов, релевантных при их обработке и анализе русскоязычным адресатом-специалистом выбранной предметной области, и ключевых для создания эффективной системы компрессии текста;
- 2) определение в прикладном аспекте этапов и методики выявления структурно-смысловой организации научной публикации на английском языке в рамках субъектно-ориентированного анализа;
- 3) построение унифицированной базовой структурной схемы (**субструктуры**) англоязычных научных текстов, которая включает реквизитные, содержательные и аксессуарные элементы;
- 4) определение компонентов **суперструктуры** англоязычной научной медицинской публикации, определяющих работу анализирующего модуля системы автоматической компрессии текста;
- 5) создание инвентаря дискурсивных маркеров **микроструктуры** для распознавания компонентов суперструктуры и отбора семантически важной информации при автоматическом порождении текста реферата;

- 6) разработку лингвистической базы знаний системы автоматической компрессии текста, включающей созданный исследовательский корпус текстов, перечень установленных моделей структурно-смысловой организации научной публикации на английском языке и разработанный комплекс правил для обработки **макроструктуры** при автоматическом порождении реферата с сохранением основного смысла публикации;
- 7) разработку алгоритма автоматической компрессии научной медицинской публикации, позволяющего моделировать выделенные шесть основных задач сжатия текста;
- 8) внедрение разработанной автором компьютерной программы MedCompress по автоматическому созданию текста общего или запросно-ориентированного реферата с высокими качественными показателями в образовательный и научный процессы на уровне учреждений высшего образования медицинского профиля.

Оппонент:

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

 Д.В. Степанова

25.03.2025

