

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 специальностям и отрасли науки, по которым она представлена к защите

На основании формулировок объекта, предмета и цели исследования, поставленных задач, использованной методологии и полученных результатов исследования, а также в соответствии с паспортом специальности 10.02.21 – прикладная и математическая лингвистика можно заключить, что диссертация Меньшениной Ирины Александровны «Автоматическая смысловая компрессия публикаций научного медицинского дискурса (на материале английского языка)» в полном объеме соответствует названной специальности и отрасли науки – филологические науки.

2. Актуальность темы диссертации

Актуальность темы диссертационного исследования Меньшениной И.А. обусловлена насущной необходимостью ускорения доступа специалистов в Республике Беларусь к новейшим данным в различных областях научного и профессионального знания в условиях перманентного увеличения объемов информации и количества узкоспециализированных публикаций на иностранных языках при недостаточной продуктивности и точности имеющихся способов и инструментов автоматической обработки и компрессии текстов научного и профессионального содержания.

В этой связи диссертация Меньшениной И.А. безусловно актуальна в плане разработки лингвистического обеспечения и создания компьютерной программы для автоматической компрессии научных медицинских статей на английском языке, направленной на существенное уменьшение их объема с адекватным сохранением основного содержания.

Исследование Меньшениной И.А. является актуальным еще и потому, что в нем ставится и решается в прикладном аспекте на материале англоязычных узкоспециализированных текстов научного медицинского дискурса целый ряд важных вопросов современной когнитивной и дискурсивной лингвистики: (1) какие лингвистические особенности научного дискурса затрудняют или облегчают понимание его содержания для иноязычного адресата; (2) возможно

ли существование унифицированной композиционной схемы изложения материала в узкоспециализированных научных текстах как основы для формализации и последующей компрессии его содержания; (3) какие языковые элементы и параметры структурной организации иноязычного научного текста являются наиболее значимыми при его автоматическом распознавании и реферировании.

Диссертация Меньшениной И.А. видится весьма актуальной и в плане активизации теоретического изучения структурно-смысловой и формально-языковой организации научного медицинского дискурса, а также решения прикладных задач в области разработки способов, средств и инструментов повышения скорости и точности автоматической компрессии содержания узкоспециализированных текстов на иностранном языке.

Актуальность проведенного Меньшениной И.А. диссертационного исследования заключается еще и в том, что оно позволит в рамках прикладной лингвистики решить проблему эффективного доступа научных работников медицинской сферы и практикующих врачей к новейшей профессионально значимой информации в англоязычных научных журналах, имеющих высокий рейтинг в международных наукометрических базах данных, что позволит существенно повысить уровень медицинской науки и качество оказания медицинской помощи в Республике Беларусь.

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

Полученные в диссертации Меньшениной И.А. выводы и результаты, вынесенные на защиту положения являются бесспорно новыми и фактически свидетельствуют об успешной реализации прикладного подхода при изучении параметров структурно-семантической компрессии научного текста. Это подтверждается, прежде всего, как оригинальностью и новизной цели исследования, сложностью задач, решаемых в диссертации, так и значимостью объекта и предмета исследования, что важно для объективной интерпретации структурно-смысловой организации узкоспециализированных текстов англоязычного научного медицинского дискурса и разработки компьютерной программы для автоматического сжатия исходного текста без утраты его основного содержания.

К однозначно новым можно отнести следующие научно значимые результаты диссертации:

- 1) определение и классификация дискурсивных характеристик англоязычного узкоспециализированного научного медицинского текста (статьи), позволяющих выделить его содержательно значимые элементы для их формализации и последующей компьютерной автоматической обработки;

2) определение понятий, выявление организации и особенностей взаимосвязи субструктуры, суперструктуры, микроструктуры и макроструктуры научной медицинской статьи в рамках лингвистического обеспечения автоматизированного процесса структурно-смысловой компрессии исходного текста;

3) выявление организации субструктуры научной медицинской статьи и ее моделирование в форме базовой унифицированной поверхностной схемы, обеспечивающей идентификацию начальной и конечной точек смысловых частей исходного текста, предназначенного для автоматической компрессии;

4) выявление организации суперструктуры научной медицинской статьи и ее моделирование в форме глубинного логического каркаса, обеспечивающего релевантность информации, полученной в результате автоматической структурно-смысловой компрессии исходного текста;

5) выявление организации микроструктуры научной медицинской статьи и ее моделирование в форме сложного комплекса дискурсивных маркеров (повторяющихся в разных текстах лексических единиц и их сочетаний), указывающих на основное содержание и используемых для распознавания компонентов суперструктуры исходного текста в процессе его автоматической компрессии;

6) выявление и моделирование макроструктуры как эмерджентной проекции целостного смыслового представления научной медицинской статьи на глубинном уровне организации, образуемой в результате взаимодействия компонентов суперструктуры с лингвистическими опорами микроструктуры исходного текста в процессе его автоматической компрессии;

7) разработка алгоритма автоматизации структурно-смысловой компрессии научной медицинской статьи на основе иерархического взаимодействия на поверхностном и глубинном уровнях четырех типов структур, последовательно чередующихся в процессе автоматического извлечения и преобразования основного содержания исходного текста;

8) создание и тестирование автоматической компьютерной программы «MedCompress» («Medical Compression»), предназначенной для машинной обработки содержания, выделения структур глубинного и поверхностного уровней, выявления ключевых концептов, структурно-смысловой компрессии и генерации рефератов научных медицинских статей на английском языке.

Перечисления научно значимых результатов, полученных соискателем, вполне достаточно, чтобы характеризовать диссертацию Меньшениной И.А. как актуальное и оригинальное научное исследование, основные выводы которого содержат новые положения, важные для прикладной лингвистики и для практической разработки на основе компьютерных технологий систем автоматического реферирования научных медицинских текстов, позволяющих суще-

ственно ускорить получение новейшей информации из авторитетных иностранных источников и значительно повысить профессиональный уровень научных и практикующих работников в сфере здравоохранения Республики Беларусь.

4. Обоснованность и достоверность выводов и рекомендаций, сформулированных в диссертации

Обоснованность и достоверность сформулированных в диссертации Меньшениной И.А. выводов и рекомендаций обеспечивается, прежде всего, достаточной полнотой и разнообразием фактологической базы исследования. Языковым материалом послужили 100 научных медицинских статей (количественно сбалансированных в трех ключевых областях современной медицины – 36 по кардиологии, 32 по пульмонологии и 32 по гастроэнтерологии), опубликованных в период с 2017 по 2022 гг. в англоязычных научных журналах «The Lancet», «The British Medical Journal (BMJ)», «The New England Journal of Medicine», «The Journal of the American Medical Association (JAMA)», которые имеют в международных наукометрических базах данных самые высокие в мире показатели (квартиль и импакт-фактор).

Анализ фактического материала последовательно осуществлялся в диссертации на основании комплексного подхода, включающего использование целого ряда методов и приемов: структурного, контекстуального, когнитивно-семантического, компонентного, сравнительного, количественного анализа, анализа словарных дефиниций и др.

Теоретическая база исследования, обеспечивающая высокую степень обоснованности и достоверности выводов, отражена в списке научных работ отечественных и зарубежных специалистов по теме диссертации (всего 136 позиций, в том числе 25 источников на иностранном языке – английском).

Отмеченное позволяет оценить выводы и рекомендации, предложенные соискателем, как объективные, достоверные и обоснованные.

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

Научная значимость диссертации Меньшениной И.А. заключается как в теоретическом аспекте, так и в прикладном плане. Выявлена когнитивно-дискурсивная организация научных медицинских статей (субструктура, суперструктура, микроструктура, макроструктура), которая использована в качестве лингвистического обеспечения для разработки на основе компьютерных технологий автоматизированных процессов компрессии содержания и генерации реферата англоязычного научного узкоспециализированного текста.

Практическая значимость диссертационного исследования очевидна и бесспорна. Соискателем разработана автоматическая компьютерная программа «MedCompress», которая благодаря своей многомодульности и полифункциональности позволяет по заложенным в нее алгоритмам выполнять компрессию содержания научного медицинского текста в жанре статьи с сохранением основного смысла. Автоматизация процесса генерации реферата позволила получить экономию времени в 240 раз при экспериментально доказанном качестве содержания с информативностью 90%. Сравнительный анализ АКП «MedCompress» с другими компьютерными программами, имеющими аналогичные функции, показал ее значительные преимущества в соответствии с производственными требованиями, предъявляемыми пользователями к системам автоматической компрессии текстов.

Материалы и выводы диссертации Меньшениной И.А. используются в Белорусском государственном медицинском университете (БГМУ) и Витебском государственном медицинском университете (ВГМУ). Получено 7 актов внедрения в научно-исследовательскую работу и учебный процесс на кафедрах иностранных языков (БГМУ), клинической фармакологии (БГМУ), организации фармации (БГМУ), клинической микробиологии (ВГМУ), педиатрии ПК и ПК (ВГМУ), травматологии и ортопедии (ВГМУ).

Отмеченное подтверждает социальную и экономическую значимость диссертации Меньшениной И.А., поскольку материалы диссертационного исследования могут быть использованы как в производстве для решения прикладных задач при реферировании научных медицинских статей на иностранном языке, так и в образовательной практике в Республике Беларусь в сферах лингвистического и медицинского образования. Материалы диссертации будут способствовать оптимизации финансовых затрат как на подготовку учебных и справочных материалов, так и на дальнейшую разработку систем автоматической обработки узкоспециализированных научных текстов, информационно-поисковых систем и систем автоматического реферирования научных статей в сфере здравоохранения.

6. Опубликованность результатов диссертации в научной печати

Основные положения диссертации Меньшениной И.А. нашли отражение в 26 единоличных публикациях (из них 3 за рубежом) общим объемом 7,6 авт. л., в том числе 8 научных статьях в научных рецензируемых изданиях, что полностью соответствует п. 19 Положения о присуждении ученых степеней и ученых званий в Республике Беларусь.

На публикации по теме диссертации сделаны ссылки в заключении диссертации, где изложены основные результаты, на основании которых сформулированы положения, выносимые на защиту. Все положения, выносимые на

защиту, нашли отражение в публикациях. Автореферат соответствует содержанию диссертации.

7. Соответствие оформления диссертации требованиям ВАК

Оформление диссертации и автореферата в целом соответствует требованиям Высшей аттестационной комиссии Республики Беларусь, содержащимся в пп. 24–26 Положения о присуждении ученых степеней и присвоении ученых званий в Республике Беларусь, а также в Инструкции по оформлению диссертации и автореферата. При изложении содержания диссертации использован научный стиль в основном с правильным и уместным применением метаязыка современной теоретической и прикладной лингвистики.

8. Недостатки диссертации

Вероятность и обоснованность основных положений и выводов диссертации Меншениной И.А. не вызывает сомнений, бесспорна также научная новизна работы и ее практическая и социально-культурная значимость. Корректность и объективность фактического материала, научный уровень и полнота его интерпретации высокие. Поэтому приведенные ниже вопросы, замечания и комментарии свидетельствуют не столько о недостатках диссертации, сколько являются размышлениями дискуссионного характера относительно отдельных положений и фактов, требующих дополнительных пояснений автора этой интересной и содержательной исследовательской работы.

1. Автор диссертации обосновывает актуальность своего исследования тем, что существующие автоматические системы, обеспечивающие смысловую компрессию текста, не удовлетворяют современным требованиям в части скорости, точности и полноты реферирования научных медицинских текстов на английском языке. В результате исследования создана автоматическая компьютерная программа, позволяющая получить при реферировании англоязычных текстов научных медицинских статей экономию времени в 240 раз при экспериментально доказанном качестве содержания с информативностью 90%. Это, безусловно, весьма значимый прикладной результат, который найдет свое применение в медицинской науке.

Однако возникает вопрос, насколько новый программный продукт может широко и надежно использоваться в медицинской практике, если большинство практикующих врачей либо недостаточно хорошо владеют английским языком для специальных целей, либо не владеют им вовсе, а потеря 10% информативности в случае поиска нужных сведений по автоматически сгенерированным рефератам научных статей может спровоцировать пропуск весьма важной информации для достаточно часто непредсказуемой врачебной

практики (в рамках медицинской статистики 1 из 10 – это все-таки очень много).

В этой связи требуется пояснить, насколько конкурентноспособным будет новый продукт на рынке ИКТ, во-первых, в сравнении с существующими автоматическими системами поиска и перевода информации в сфере медицины, а во-вторых, в сравнении со всё более изощренными системами искусственного интеллекта, стремительно проникающими во все информационные сферы, в том числе и в сферу реферирования и перевода англоязычных узкоспециальных научных текстов на другие языки (в том числе на русский язык)? Не устареет ли полученный диссертантом продукт ИКТ (и, соответственно, научное обоснование его разработки) еще до того момента, как будет широко внедрен в Республике Беларусь? Какой в этом плане соискатель может дать прогноз относительно сохранения научной значимости и практического применения результатов диссертации?

2. При разработке лингвистического обеспечения автоматизированной компьютерной программы по структурно-смысловой компрессии научных медицинских статей автор диссертации использует когнитивную теорию дискурса, разработанную Т.А. Ван Дейком, в рамках которой выделяется три типа структур в построении дискурса (глобальные макроструктуры, формальные суперструктуры, локальные микроструктуры). Автор диссертации дополняет эту когнитивную модель структуризации дискурса еще и уровнем субструктуры, подробно анализирует полученную полиструктурную модель дискурсивной организации научного медицинского текста и далее интерпретирует ее в качестве процедурного компонента и части формального алгоритма автоматического процесса структурно-смысловой компрессии текста. Ср. в положениях, выносимых на защиту: «субструктура предоставляет возможность <...> формализации», «суперструктура определяет работу анализирующего модуля», «алгоритмическое распознавание компонентов суперструктуры», «модуль <...> 3) выделение структур поверхностного и глубинного уровней», «автоматическое создания макроструктуры», «обработки полученной макроструктуры» (стр. 10).

Таким образом, в диссертации явно выделяются различные модели данных структур (когнитивная, лингвистическая, алгоритмическая, процедурная). Разграничение разных видов моделей принципиально важно как в методологическом плане, так и в теоретическом аспекте (в том числе метаязыковом). В диссертации такое разграничение прямо нигде не объясняется. При этом сам термин *модель* в диссертации используется всего несколько раз в различных смыслах, ср. «продуктивная модель, образец текстового построения» (стр. 31),

«алгоритмическая модель смысловой компрессии» (стр. 55), «модель референтной ситуации» (стр. 17, 59), «универсальная модель жанра [статьи]» (стр. 60).

В этой связи требуется пояснить, как именно соотносятся в понимании соискателя понятия когнитивной модели, лингвистической модели, формальной (алгоритмической) модели, процедурной модели в отношении дискурсивной структуры научного текста и автоматизированного процесса его компрессии?

3. В диссертации отсутствует описание технической стороны процесса разработки АКП «MedCompress» («Medical Compression»), не излагаются физические и математические принципы ее действия, нет образцов перевода информации, содержащейся в неформализованном виде в тексте, на формализованный искусственный язык на основе определенного математического аппарата, что обычно должно присутствовать в диссертационном исследовании в случае, если создание компьютерной программы является основным результатом диссертации по прикладной и математической лингвистике и отражается в положениях, выносимых на защиту.

В этой связи от автора диссертации требуется дополнительное пояснение технических параметров и математического обоснования созданной компьютерной программы «MedCompress» («Medical Compression»). Если же машинный код программы был использован в уже готовом виде, то следует указать его характеристики и сообщить источник его получения.

4. Хорошо известно, что следование терминологической точности в научном описании во многом предопределяет как объективность процесса исследования, так и достоверность его результатов.

В любой диссертации как квалификационной работе такую точность задает ее название, в котором должна быть непротиворечиво и безупречно сформулирована главная научная проблема исследования, что весьма важно для адекватного восприятия его содержания. В этом смысле название диссертации является, на наш взгляд, не вполне корректным в плане научной точности и использованной терминологии.

Так, термин *публикация* в строгом смысле не означает (т. е. означает не только) ‘научная статья’, поскольку *публикация* может быть и *доклада*, и *монографии*, и *научного отчета*, и *медицинского протокола*, и *инструкции* и т. д. В этом смысле название диссертации «Автоматическая смысловая компрессия публикаций научного медицинского дискурса (на материале английского языка)» не вполне точно (или даже совсем не точно) отражает действительный объект исследования – *тексты научных статей медицинского содержания*,

структура и содержание которых существенно отличается от структуры и содержания иных, упомянутых выше, публикаций научного медицинского дискурса. Термин *публикация* используется также и в формулировках цели, задач, объекта и предмета исследования, а также в положениях, выносимых на защиту. Однако в описании материала для исследования прямо называется его объект – «научные медицинские статьи, опубликованные в ведущих англоязычных журналах» (стр. 8). В данном случае возникает противоречие между реальным объектом исследования и его терминологическим обозначением (термины *публикация* и *научная статья* соотносятся как гипероним и гипоним), чего не должно быть в научном исследовании.

В этом случае соискатель должен пояснить, насколько и как именно распространяется использование разработанной автоматической компьютерной программы на материал медицинских публикаций иных жанров, кроме проанализированных в диссертации научных статей?

Следует также отметить, что текст диссертации изобилует специальными лингвистическими терминами и терминологическими сочетаниями, что указывает на глубокое знание ее автором метаязыка современного языкознания, необходимого для адекватного описания анализируемых фактов и оформления полученных результатов. Вместе с тем, в тексте диссертации встречается немало случаев весьма оригинального использования как собственно лингвистических, так и общенаучных терминов, часто на грани их адекватного понимания и правильного использования. Такие случаи встречаются даже в положениях, выносимых на защиту (?!), напр.: «авторски-индивидуального самовыражения» (стр. 9), хотя существует стандартный лингвистический термин с тем же значением – «индивидуально-авторского»); «единицы второстепенной семантики» (стр. 10), хотя в лингвистике не существует ни «второстепенной», ни «первостепенной семантики», тем более что в тексте диссертации используется правильное терминологическое сочетание – «единицы второстепенной смысловой значимости» (стр. 107), в данном случае мы имеем образцовый пример смешения понятий и терминов «значение (семантика)» и «значимость».

Такое не вполне оправданное употребление научной терминологии затрудняет восприятие в тексте диссертации ее основных положений и, по нашему глубокому убеждению, не должно встречаться в научной квалификационной работе.

Высказанные замечания и вопросы, как можно видеть, носят частный и преимущественно дискуссионный характер, что не влияет на общую положительную оценку диссертации Меньшениной И.А.

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

Содержание и научно-теоретический уровень диссертации, обоснованность используемых методов, научная значимость и достоверность полученных результатов свидетельствуют о том, что научная квалификация Меньшениной Ирины Александровны соответствует требованиям, которые предъявляются к кандидату филологических наук по специальности 10.02.21 – прикладная и математическая лингвистика.

10. Заключение

Диссертация Меньшениной И.А. «Автоматическая смысловая компрессия публикаций научного медицинского дискурса (на материале английского языка)» представляет собой завершенное комплексное исследование, которое содержит новые, научно обоснованные достоверные результаты, имеет несомненную высокую теоретическую и практическую значимость. Ознакомление с диссертацией Меньшениной И.А. позволяет говорить о компетентности и научной зрелости исследователя, сумевшего решить актуальную научную проблему.

В соответствии с пп. 20 и 21 Положения о присуждении ученых степеней и присвоении ученых званий ВАК Республики Беларусь рекомендуется присудить Меньшениной Ирине Александровне ученую степень кандидата филологических наук по специальности 10.02.21 – прикладная и математическая лингвистика за:

- выявление языковых особенностей англоязычного научного медицинского дискурса, необходимых для адекватного восприятия, понимания, обработки и реферирования содержания узкоспециализированных текстов иноязычным адресатом;

- определение и описание в прикладном аспекте организации субструктуры англоязычных научных медицинских статей как совокупности элементов трех типов (реквизитных, содержательных и аксессуарных), позволяющих формализовать содержание узкоспециализированного текста в процессе его автоматической компрессии;

- выявление и характеристику в прикладном аспекте организации суперструктуры англоязычных научных медицинских статей как глубинного логического каркаса, детерминирующего релевантность информации в исходном узкоспециализированном тексте и в автоматически сгенерированном реферате;

- определение и описание в прикладном аспекте организации микроструктуры англоязычных научных медицинских статей как набора функцио-

нально дифференцированных дискурсивных маркеров, распределенных в рамках каждого компонента суперструктуры узкоспециализированного текста;

– разработку лингвистической базы данных для автоматической компрессии англоязычных научных медицинских статей в форме макроструктуры (проекции целостного смыслового представления текста из элементов его субструктуры, суперструктуры и микроструктуры) и правил ее преобразования в реферат исходного узкоспециализированного текста;

– создание автоматической компьютерной программы «MedCompress» («Medical Compression»), способной осуществлять предварительную обработку, выделение структур поверхностного и глубинного уровней, дифференциацию ключевых концептов, компрессию содержания и генерацию реферата, а также построение частотного словаря англоязычных научных медицинских статей,

что в совокупности решает актуальную для прикладной лингвистики научную проблему лингвистического обеспечения и компьютерной автоматизации процесса реферирования узкоспециализированных текстов, существенно уточняет представления о когнитивно-дискурсивной организации научной медицинской статьи, имеет важное значение в области разработки компьютерных программ, предназначенных для автоматической компрессии узкоспециализированных текстов научного медицинского дискурса на иностранном языке.

Официальный оппонент –

доктор филологических наук,
профессор

Е.Е. Иванов

20.03.2025 г.

