

УДК 004.912:61:81'373.45

Евчик Надежда Семеновна
доктор филологических наук,
профессор кафедры речеведения
и теории коммуникации
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Nadejda Evchyk
Doctor of Philology,
Professor of the Department of Speech
Studies and Communication Theory
Belarusian State University
of Foreign Languages
Minsk, Belarus
nadejda.evchyk@gmail.com

Меньшенина Ирина Александровна
кандидат филологических наук,
старший преподаватель кафедры
иностранных языков
Витебский государственный
медицинский университет
г. Витебск, Беларусь

Irina Menshenina
PhD in Philology,
Senior Lecturer of the Department
of Foreign Languages
Vitebsk State Medical University
Vitebsk, Belarus
irinaslav23@rambler.ru

НАУЧНЫЙ ДИСКУРС И ТЕХНОЛОГИИ СМЫСЛОВОЙ ИНТЕРПРЕТАЦИИ В УСЛОВИЯХ ЦИФРОВОЙ ТРАНСФОРМАЦИИ: ОТ ЯЗЫКОВОГО БАРЬЕРА К ПРОФЕССИОНАЛЬНОМУ ВЗАИМОПОНИМАНИЮ

Статья посвящена анализу проблем восприятия англоязычных медицинских публикаций специалистами, для которых английский язык не является родным, и описанию возможностей их преодоления с помощью разработанной системы MedCompress. В ходе трехэтапного анкетирования были выявлены ключевые лингвистические барьеры: профессионально-детерминированный языковой инвентарь, спорадическое использование проформ, особенности авторского самовыражения, синтаксическая сложность англоязычных текстов. Для их преодоления создана интеллектуальная программа MedCompress, обеспечивающая автоматическую смысловую компрессию публикаций, генерацию рефератов и построение частотных словарей. Система поддерживает как классический, так и интерактивный режимы, что позволяет персонализировать результаты обработки под конкретные исследовательские задачи. Экспериментальное внедрение в медицинских университетах Беларуси показало рост продуктивности анализа публикаций более чем в четыре раза при сохранении качества понимания. Полученные результаты свидетельствуют о том, что цифровые технологии способны не только автоматизировать работу с текстами, но и формировать новое пространство профессионального взаимопонимания, снижая языковые барьеры и способствуя развитию научного диалога в условиях цифровой трансформации.

Ключевые слова: научный дискурс; профессиональная коммуникация; языковые барьеры; смысловая компрессия; автоматическое реферирование; система MedCompress.

SCIENTIFIC DISCOURSE AND TECHNOLOGIES OF SEMANTIC INTERPRETATION IN THE CONTEXT OF DIGITAL TRANSFORMATION: FROM LANGUAGE BARRIER TO PROFESSIONAL MUTUAL UNDERSTANDING

The article addresses the challenges faced by non-native English-speaking specialists in comprehending medical publications and demonstrates how these barriers can be overcome through the use of the MedCompress system. A three-stage survey identified key linguistic difficulties: professionally determined lexical inventory, sporadic use of preforms, specifics of authorial expression, syntactic complexity of English texts. To address these challenges, the MedCompress program was developed, providing automatic semantic compression of publications, generation of abstracts, and construction of frequency dictionaries. The system supports both classical and interactive modes, enabling personalized processing results tailored to specific research tasks. Experimental implementation in Belarusian medical universities demonstrated more than a fourfold increase in the efficiency of analyzing publications while maintaining comprehension quality. The findings confirm that digital technologies can not only automate text processing but also foster a new space of professional understanding, reducing linguistic barriers and supporting the development of scientific dialogue in the context of digital transformation.

Key words: scientific discourse; professional communication; language barriers; semantic compression; automatic summarization; MedCompress system.

Современные тенденции цифровой трансформации научной коммуникации все более четко очерчивают приоритеты смысловой интерпретации и компрессии знаний. Актуальность этого процесса особенно значима в области медицины, где объем англоязычных научных публикаций растет экспоненциально [1], а эффективное профессиональное взаимодействие между специалистами различных стран требует сокращения когнитивных и языковых барьеров [2, 3]. Эти барьеры препятствуют полноценному участию в международном научном диалоге и затрудняют использование последних достижений в клинической практике и исследовательской работе.

В научном пространстве ключевой проблемой становится не просто доступ к англоязычным публикациям, а возможность глубокого понимания их содержания специалистами, для которых английский не является родным языком. Профессиональные читатели – врачи и ученые – нуждаются в точном и полном осмыслении текста, сохраняющем логику, доказательность и новизну авторского замысла. В этом контексте задача интеллектуального анализа выходит за рамки традиционного реферирования, предполагающего обобщение и упрощение. Напротив, необходимо формировать условия для смысловой интерпретации, обеспечивающей сохранение глубины, последовательности и эпистемологической целостности научного изложения.

Специфика медицинской публикации заключается в ее прикладной направленности: она ориентирована на практическое решение задач здравоохранения, а потому предъявляет повышенные требования к уровню понимания текста со стороны специалиста. Врач, опирающийся в своей профессиональной деятельности на данные научных исследований, должен обладать не только актуальной базой теоретических знаний и клиническими навыка-

ми, но и способностью к критическому мышлению, владеть основами доказательной медицины и уметь оценивать релевантность новой информации. Современный медицинский специалист обязан систематически знакомиться с последними достижениями в своей и смежных областях, самостоятельно искать и интерпретировать информацию по актуальным вопросам, объективно оценивать ее качество и применять в своей практике.

В условиях постоянно возрастающего объема публикаций, обработка англоязычного текста требует от читателя высокой когнитивной нагрузки, времени и специальных навыков. Как отмечается в методических рекомендациях по чтению научной медицинской литературы [4], понимание текста предполагает последовательную фиксацию ключевых идей, анализ незнакомых лексем, повторные прочтения и избирательное извлечение значимой информации, то есть по сути – ручную смысловую компрессию. Очевидно, что в реальных условиях лишь ограниченное количество публикаций может быть полноценно проанализировано медицинским специалистом с сохранением всех смысловых нюансов.

В то же время именно научные статьи обеспечивают динамическое распространение результатов новых исследований, способствуют передаче знаний и, следовательно, играют ключевую роль в развитии медицины и повышении качества медицинской помощи. Высокие требования, предъявляемые к жанру медицинской публикации, предполагают тщательное планирование исследования, сбор и верификацию данных, применение статистических методов и обеспечение внешней валидности результатов. Все это усиливает важность точного и глубокого понимания научного текста со стороны практикующего специалиста.

На этом фоне особенно актуальной становится задача выявления факторов, препятствующих адекватному восприятию англоязычных текстов. Проведенное в три этапа углубленное анкетирование 25 русскоязычных медицинских специалистов позволило квантифицировать основные лингвистические барьеры при чтении научных публикаций. Примененная прогрессивная методология, включавшая открытые вопросы, категориальный анализ и градуированную оценку, выявила четыре ключевые категории трудностей, каждая из которых характеризуется высоким уровнем сложности – от 82 % до 100 %.

Наивысший показатель сложности (100 %) зафиксирован в категории профессионально-детерминированного языкового инвентаря, отражающей многомерный терминологический вызов. Медицинская терминология, согласно лингвистическим исследованиям, насчитывает от 200 000 до 500 000 терминов, и является одной из наиболее обширных в научном дискурсе [5, с. 2]. Сложность усугубляется высокой степенью специализации: в современной медицине выделяется около 200 дисциплинарных направлений, каждое из которых развивает собственный лексикон. Дополнительные трудности создает система аббревиатур, отличающаяся как высокой степенью контекстуальной неоднозначности, так и полисемией. Особое место занимают эпонимы – собственные имена, используемые для обозначения болезней,

симптомов и методов лечения. Эта форма номинации требует от читателя знания истории медицины и культуры англоязычного медицинского сообщества, что нередко становится барьером интерпретации.

На втором уровне сложности (87 %) зафиксированы трудности, связанные со спорадическим использованием проформ и частыми лексическими повторами, которые, с точки зрения русскоязычных читателей, воспринимаются как стилистическая избыточность. Однако такая текстовая когезия объясняется структурными расхождениями между английской и русской языковыми системами: в англоязычном научном дискурсе повторы выполняют функцию поддержания логической связанности и акцентирования информации, тогда как в русскоязычной традиции предпочтение отдается разнообразию форм и средств выражения.

Следующая категория (82 %) связана с особенностями авторского самовыражения, в частности – допустимостью проявлений научной субъективности. Анализ анкет показал, что англоязычный академический стиль допускает элементы экспрессивности и аргументативного самопозиционирования, что вступает в противоречие с нормами нейтральности, характерными для русскоязычного научного письма. Это различие порождает трудности при интерпретации авторской позиции, особенно в случаях использования модальных конструкций и оценочных суждений, замаскированных под аналитическое изложение.

Наконец, еще одной ключевой проблемой оказалась синтаксическая сложность англоязычных текстов (98 %). Использование громоздких грамматических структур, включающих многоступенчатые подчинения, длинные конструкции и фразовую архитектуру с высокой степенью вложенности, затрудняет процесс семантического декодирования. Особенно сложными для восприятия являются субстантивные атрибутивные конструкции, характерные для научного стиля: последовательности существительных, выступающих в адъективной функции, часто не имеют прямых аналогов в русском языке. Такая языковая экономия и высокая информационная плотность требуют от читателя навыков сегментации и интерпретации смысловой иерархии компонентов высказывания.

Эмпирические данные, полученные в результате исследования, позволяют сделать важные выводы для развития технологий автоматической обработки научных медицинских текстов. Современные системы должны быть способны устранять неоднозначность терминов с полисемантической нагрузкой, адаптироваться к стилистическим и культурным особенностям дискурса, управлять синтаксической сложностью без потери семантической точности, а также сохранять эпистемологические и аргументативные нюансы оригинала. Преодоление этих трудностей требует не просто лингвистического перевода, а полноценного когнитивного и культурного посредничества. Все это обосновывает необходимость разработки интегрированных технологических решений, способных эффективно функционировать в условиях возрастающей сложности и глобализации научного медицинского дискурса.

Для решения поставленной задачи разработан программный комплекс MedCompress, который предназначен для работы с публикациями из трех ключевых областей современной медицины в международной практике: кардиологии, пульмонологии и гастроэнтерологии.

В основе архитектуры MedCompress лежит формализация процедуры генерации сжатой версии текста в форме реферата. Программа опирается на специально разработанное лингвистическое обеспечение, позволяющее выделять ключевые концепты и обеспечивать логически стройную редукцию содержания. Помимо собственно алгоритмов компрессии, в рамках проекта была разработана методика создания и ведения адаптивного специализированного медицинского словаря. Эта методика опирается на отбор лексических единиц из наиболее информативных фрагментов исходного текста, что делает словарь динамическим и способным к автоматическому расширению.

Функциональная реализация MedCompress организована в виде последовательности взаимосвязанных модулей, каждый из которых решает отдельную когнитивную задачу. В первую очередь, интерактивный модуль обеспечивает диалоговую загрузку текста и определение пользовательских параметров обработки. Пользователь может задать режим генерации реферата: классический – с автоматическим извлечением всех компонентов предметной суперструктуры или интерактивный – с выбором значимых для него разделов, таких как цели, методы или результаты. Такая возможность персонализировать выходной текст усиливает адаптивность программы к задачам конкретного специалиста.

Следующие модули обеспечивают предварительную обработку текста, структурный анализ, выделение ключевых концептов и генерацию реферата. Завершающий этап – построение алфавитно-частотного словаря на основе сгенерированного текста. Программа позволяет пользователю модифицировать словарь путем исключения так называемых "минус-слов" – второстепенных лексем, грамматических единиц, а также уже известных слов. Встроенный список стоп-слов включает артикли, предлоги, союзы и числительные. Кроме того, реализована возможность загрузки индивидуального списка исключений, что расширяет гибкость настройки результата.

MedCompress создает как полный, так и краткий частотный словарь. При этом краткий словарь исключает производные слова и сложные термины, которые могут быть поняты через знание составляющих элементов. Например, если в словаре присутствует глагол *to inflame* 'воспалить', то производные от него *inflammation* 'воспаление' и *inflammatory* 'воспалительный' могут быть исключены. Для лемматизации используется специализированная библиотека, приводящая все слова к начальной форме. Полученные словари пригодны как для учебных целей, так и для пополнения профессионального лексического минимума пользователя.

При обработке нескольких публикаций (до пяти) программа создает отдельные рефераты и словари для каждого файла, а также формирует сводные частотные словари. Эти словари сохраняются в тематически органи-

зованной базе, позволяющей накапливать и систематизировать лексические ресурсы. База может быть использована для сравнения словарей, выявления повторов и построения персонализированного корпуса терминов. Тем самым MedCompress выступает не только как средство автоматизированного реферирования, но и как инструмент создания лексико-семантического ядра научного дискурса в медицине.

В рамках пилотного внедрения программы в УО «Витебский государственный медицинский университет» и УО «Белорусский государственный медицинский университет» было обработано свыше 1000 англоязычных публикаций, как в классическом, так и в интерактивном режимах. Оценка качества программы осуществлялась методом анкетирования, в котором приняли участие 23 специалиста медицинского профиля. Анкета включала два блока: оценку удобства использования интерфейса и оценку соответствия сгенерированных рефератов заданным критериям. Полученные результаты показали высокий уровень удовлетворенности: 100 % респондентов положительно оценили функции перевода, скачивания и генерации лексического минимума; 87 % – удобство интерфейса и возможность сохранения данных. Наиболее значимыми критериями оказались точность передачи информации, доступность изложения, единый стиль и языковая четкость (100 %). За ними следовали логичность структуры (91,3 %) и полнота содержания (87 %).

Дополнительно был проведен эксперимент по сравнению эффективности анализа публикаций вручную и с использованием MedCompress. Две группы экспертов выполняли одну и ту же задачу – составление обзора литературы на основе 10 релевантных статей. Первая группа анализировала тексты вручную, затрачивая в среднем 28 минут на одну статью (2,14 статьи в час). Вторая группа использовала рефераты, сгенерированные программой: среднее время анализа составило 7 минут на статью, что соответствует 8,6 статьи в час. Таким образом, автоматизация процесса позволила повысить производительность труда более чем в четыре раза без утраты качества смысловой интерпретации.

Полученные результаты подтверждают высокую прикладную значимость интеллектуальной системы MedCompress как инструмента смысловой компрессии научных текстов. Ее функциональные возможности способствуют поддержке профессиональной коммуникации и развитию специализированной лексической базы, что особенно актуально в условиях стремительного развития цифровых технологий и трансформации научного дискурса. Важнейший вывод, вытекающий из проведенного исследования, заключается в том, что задачи, стоящие перед современной наукой, выходят за рамки простого перевода или механической обработки текста: они требуют интерпретации, способной учитывать как когнитивные, так и культурные особенности восприятия информации специалистами разных языковых и научных традиций.

Сформулированный на основе эмпирических данных подход демонстрирует, что вектор развития научной коммуникации все более уверенно сме-

щается от пассивного потребления информации к ее интеллектуальной переработке. Разработка и внедрение специализированных цифровых решений, таких как MedCompress, отражают становление новой парадигмы в научном дискурсе, в которой технологии выступают не как автономные инструменты, а как посредники между текстом и профессиональным пониманием.

В условиях экспоненциального роста англоязычных научных публикаций, особенно в таких критически важных сферах, как медицина, проблема языкового и когнитивного барьера приобретает системный характер. В этом контексте задача цифровизации научного знания заключается не столько в автоматизации чтения, сколько в создании механизмов, способствующих глубокой смысловой интерпретации, адаптированной к нуждам конкретного пользователя – специалиста, принимающего решения на основе анализа и синтеза сложной информации.

Практический опыт использования MedCompress показывает, что переход от традиционного линейного чтения к персонализированным формам смыслового восприятия текста позволяет существенно повысить эффективность междисциплинарной и межъязыковой коммуникации. Такие технологии открывают новые горизонты для профессионального взаимодействия, способствуют формированию единых понятийных оснований и обеспечивают более точную, оперативную и контекстно релевантную передачу научного знания.

Таким образом, результаты исследования позволяют утверждать, что научная коммуникация в условиях цифровой трансформации становится все более ориентированной на создание среды смыслового диалога. От языкового барьера мы переходим к формированию устойчивого пространства профессионального взаимопонимания – на основе современных технологий анализа, интерпретации и компрессии текста. Подобный сдвиг представляется ключевым для будущего международного научного сотрудничества, особенно в высокостандартизированных и социально значимых областях, таких как медицина и здравоохранение.

ЛИТЕРАТУРА

1. Johnson R. The STM report: an overview of scientific and scholarly publishing. 1968–2018. Celebrating the 50th anniversary of STM : fifth ed., Oct. 2018 [Electronic resource] // Oxford, UK : STM: International Association of Scientific, Technical and Medical Publishers, 2018. URL: <https://niascomm.in/wp-content/uploads/2019/03/STM-Report-2018.pdf> (date of access: 17.12.2024).
2. Торубарова И. И. Лингвокультурная специфика научного медицинского текста (на материале англоязыч. и русскоязыч. ст.) : дис. ...канд. филол. наук : 10.02.20. Тверь, 2021. 141 л.
3. Науменко Ю. Н. Маркеры научных категорий исследования: сопоставительный анализ способов вербализации : дис. ...канд. филол. наук : 10.02.20. Тверь, 2021. 167 л.

4. Критический анализ медицинской литературы : метод. рекомендации / [Г. Р. Касян, А. В. Говоров, К. Б. Колонтарев, Д. Ю. Пушкарь] ; Правительство Москвы, Департамент здравоохранения города Москвы. – М. : АБВ-пресс, 2017. 29 с.
5. Кириллова Т. С. Пути формирования и лексикологические особенности английской терминологии подъязыка медицины : (дерматология-венерология) : автореф. дис. ... канд. филол. наук : 10.02.04. Пятигорск, 1990. 17 л.