

Структурно-смысловая организация публикаций научного медицинского дискурса в свете задач прикладной лингвистики

Меньшенина И.А.

Учреждение образования «Витебский государственный медицинский университет», Витебск

Экспоненциальный рост количества научных публикаций в медицинской сфере препятствует их эффективной обработке и анализу специалистами, что может привести к упущению ценных результатов исследований. Решение задачи ускорения доступа к необходимой информации посредством автоматической системы смысловой компрессии требует выявления структурно-смысловой организации научных медицинских публикаций, что и является целью данной работы.

Материал и методы. Практическим материалом исследования послужили 100 научных статей, опубликованных в ведущих и наиболее цитируемых британских и американских научных медицинских журналах в период с 2017 по 2022 г.

В процессе работы были использованы такие лингвистические виды анализов и методы, как структурный, контекстуальный, когнитивно-семантический, сопоставительно-параметрический, дискурс-анализ, классификационный метод и метод тематического моделирования.

Результаты и их обсуждение. Анализ, проведенный на основе субъектно-ориентированного и когнитивно-информационного подходов, позволил выявить 4 иерархически действующих типа структур в англоязычных научных публикациях медицинского профиля. Установленные, они функционируют на поверхностном и глубинном уровнях организации текста, обеспечивая возможность получения прикладного результата: формализацию процесса смысловой компрессии.

Заключение. Полученные результаты позволяют решить прикладные задачи оперативного создания рефератов медицинских научных публикаций, необходимые специалистам медицинского профиля для оптимизации научного поиска.

Ключевые слова: научная медицинская статья, структурно-смысловая организация, субструктура, суперструктура, микроструктура, макроструктура.

(Ученые записки. – 2024. – Том 40. – С. 158–162)

The Structural and Semantic Organization of Scientific Medical Discourse Publications in View of Applied Linguistics Tasks

Menshenina I.A.

Education Establishment "Vitebsk State Medical University", Vitebsk

The exponential growth in the number of scientific publications in the medical field hinders their effective processing and analysis by specialists, which can lead to the loss of valuable research findings. Solving the problem of accelerating access to necessary information through an automatic semantic compression system requires identifying the structural and semantic organization of scientific medical publications, which is the purpose of this work.

Material and methods. The practical material of the study was 100 scientific articles published in the leading and most cited British and American scientific medical journals in the period from 2017 to 2022. In the course of the work, such linguistic methods as discourse analysis, structural, contextual, cognitive-semantic analysis, classification method, comparative parametric analysis, and thematic modeling method were used.

Findings and their discussion. The analysis carried out on the basis of subject-oriented and cognitive-informational approaches allowed us to identify 4 hierarchically acting types of structures in English-language scientific publications of the medical profile. They function at the surface and deep levels of text organization, providing the opportunity to obtain an applied result: formalization of the process of semantic compression.

Conclusion. The results obtained make it possible to solve applied problems of the operational creation of medical scientific publications abstracts necessary for specialists to optimize scientific research.

Key words: scientific medical article, structural and semantic organization, substructure, superstructure, microstructure, macrostructure.

(Scientific notes. – 2024. – Vol. 40. – P. 158–162)

Научная медицинская публикация представляет особую значимость для развития медицинской науки с целью поиска путей оказания качественной профессиональной помощи пациенту. Однако экспоненциальный рост количества публикуемых статей, а также сложность и времязатратность их обработки для иноязычного адресата-специалиста препятствуют использованию результатов новейших исследований в медицинской практике. Ускорение доступа специалистов к новейшей профессиональной информации возможно посредством разработки системы автоматической смысловой компрессии, обеспечивающей получение 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.

Анализ текстов научных медицинских публикаций на основе субъектно-ориентированного и когнитивно-информационного подходов позволяет выявить элементы их структурно-смысловой организации, а также определить порядок иерархического взаимодействия выявленных типов структур, необходимых для реализации смысловой компрессии.

Материал и методы. Практическим материалом исследования послужили 100 научных статей, опубликованных в ведущих и наиболее цитируемых британских и американских научных медицинских журналах в период с 2017 по 2022 г.: «The Lancet», «The British Medical Journal (BMJ)», «The New England Journal of Medicine», «The Journal of the American Medical Association (JAMA)».

В процессе работы были использованы такие лингвистические виды анализов и методы, как структурный, контекстуальный, когнитивно-семантический, сопоставительно-параметрический, дискурс-анализ, классификационный метод и метод тематического моделирования.

Результаты и их обсуждение. Применение субъектно-ориентированного подхода позволяет 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, чтобы сжато излагать суть текста автоматическим способом. В результате изучения когнитивных аспектов порождения и понимания профессионального дискурса в медицинской сфере в сочетании с его лингвистическим анализом была выявлена структура организации смысла научной медицинской публикации, включающая 4 типа иерархически взаимодействующих друг с другом структур поверхностного и глубинного уровней (рис. 1).

1. Субструктура – инициальный выделенный исследователем элемент – представляет собой универсальную поверхностно-наблюдаемую схему научных медицинских статей, публикуемых в англоязычных журналах. С позиций ведущего для формализации признака функциональности разделов субструктура включает: реквизитные – используемые для формальной идентификации первичного документа, содержательные – представляющие информацию непосредственно по обсуждаемой проблеме, и акцессорные – являющиеся вспомогательно-справочным аппаратом публикации. Из них в качестве основных для выполнения автоматической смысловой компрессии выступают содержательные разделы, заключающие в себе смысл медицинской публикации: Introduction ‘Введение’, Methods ‘Методы’, Results ‘Результаты’, Discussion ‘Обсуждение’. Выделение субструктуры публикаций обеспечивает техническую возможность инициации в создании программного обеспечения для формализации условий первичной ориентации программы в имеющемся объеме научной информации статьи, а также служит основой

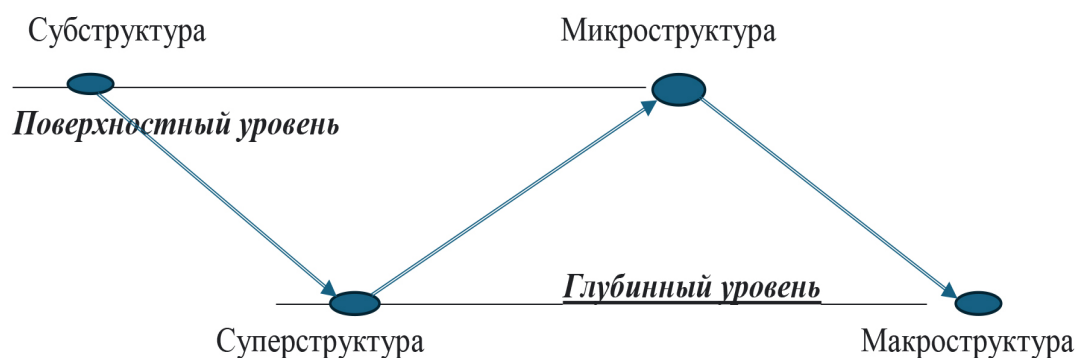


Рис. 1. Уровневая процедуральность реализации иерархически организованных типов структур англоязычных научно-медицинских публикаций при автоматическом выполнении их смысловой компрессии

для выявления следующего иерархического уровня – суперструктуры, которая поверхностно не наблюдается, поскольку глубинно связана с изложением сути публикации.

2. Суперструктура, выделенная исследователем на втором этапе анализа, представляет собой жанровую схему научной медицинской публикации, определяющую ее тематическую организацию, основным назначением которой является отражение универсальной смысловой модели профессиональных знаний адресата и адресанта.

Суперструктура состоит из 16 компонентов, которые распределяются по 3–5 на каждый из 4-х исходно выделенных 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.

3. Каждый выделенный компонент суперструктуры воплощается в лексико-синтаксическом пространстве текста определенным лингвистическим инвентарем – элементами микроструктуры, маркирующими наиболее информативные отрезки текста статьи. Данные элементы представляют собой лексические единицы второстепенной смысловой значимости и определяются как дискурсивные маркеры. Невозможность формализации в целях автоматической компрессии единиц первостепенной смысловой значимости – терминологической медицинской лексики – объясняется тем, что анализу подвергаются статьи трех разных областей медицины: кардиологии, гастроэнтерологии, пульмонологии, каждая из которых имеет отличный от других профессионально-детерминированный языковой инвентарь. Кроме того, высокая количественная представленность терминологии в каждой области осложняется динамичностью ее состава: он постоянно меняется за счет устаревания одних терминов и появления новых.

Выполняя дейктическую функцию, дискурсивные маркеры повторяются в научных медицинских публикациях, указывая на содержание одного из компонентов

суперструктуры, каждый из которых включает ограниченное количество обеспечивающих его микроструктурных элементов.

4. Элементы микроструктуры эксплицируют смыслодержающие фрагменты, передающие суть публикуемого медицинского исследования, и, таким образом, создают его макроструктуру, представляющую обобщенный смысл статьи, что составляет 4-й этап анализа текста (рис. 2).

Как видно на рисунке 2, суперструктурные компоненты раздела «Результаты», характерные для каждой научной медицинской публикации на английском языке, в случае обработки статьи «Effects of antiplatelet therapy after stroke due to intracerebral haemorrhage (RESTART): a randomised, open-label trial»¹, опубликованной в журнале The Lancet, отражаются в плане содержания элементами присущей исключительно ей макроструктуры. Основой для выделения данных элементов послужили дейктические единицы, маркирующие основное содержание каждого из компонентов суперструктуры раздела «Результаты»:

1) для компонента **Основные исходы исследования** – дискурсивный маркер *the primary outcome* ‘основной исход’;

2) для компонента **Дополнительные исходы исследования** – *the secondary outcome* ‘дополнительный исход’, *secondary outcomes* ‘дополнительные исходы’;

3) для компонента **Нежелательные явления** – *die* ‘умирать’, *deaths* ‘смерти’, *adverse events* ‘нежелательные явления’.

При анализе выделенных элементов макроструктуры были обнаружены фрагменты, не несущие существенной информации и затрудняющие интерпретацию текста адресатом. В приведенной статье к таким фрагментам эксперты отнесли:

1) в компоненте **Основные исходы исследования** – информацию, приведенную в скобках в первом предложении, а также второе и третье предложения целиком;

2) в компоненте **Дополнительные исходы исследования** – информацию, приведенную в скобках, за исключением данных, приводимых в процентах;

3) в компоненте **Нежелательные явления** – информацию, приведенную в скобках, и последнее предложение.

Таким образом, после дополнительного сжатия, в ходе которого были исключены фрагменты, отмеченные экспертами как несущественные для интерпретации, был получен следующий вторичный текст, представляющий собой реферат раздела «Результаты» данной публикации:

For the primary outcome, 12 (4%) of 268 participants allocated to start antiplatelet therapy had recurrences of intracerebral haemorrhage compared with 23 (9%) of 268 participants who did not start therapy.

¹ «Эффекты антиагрегантной терапии после инсульта, вызванного внутримозговым кровоизлиянием (RESTART): рандомизированное открытое исследование».

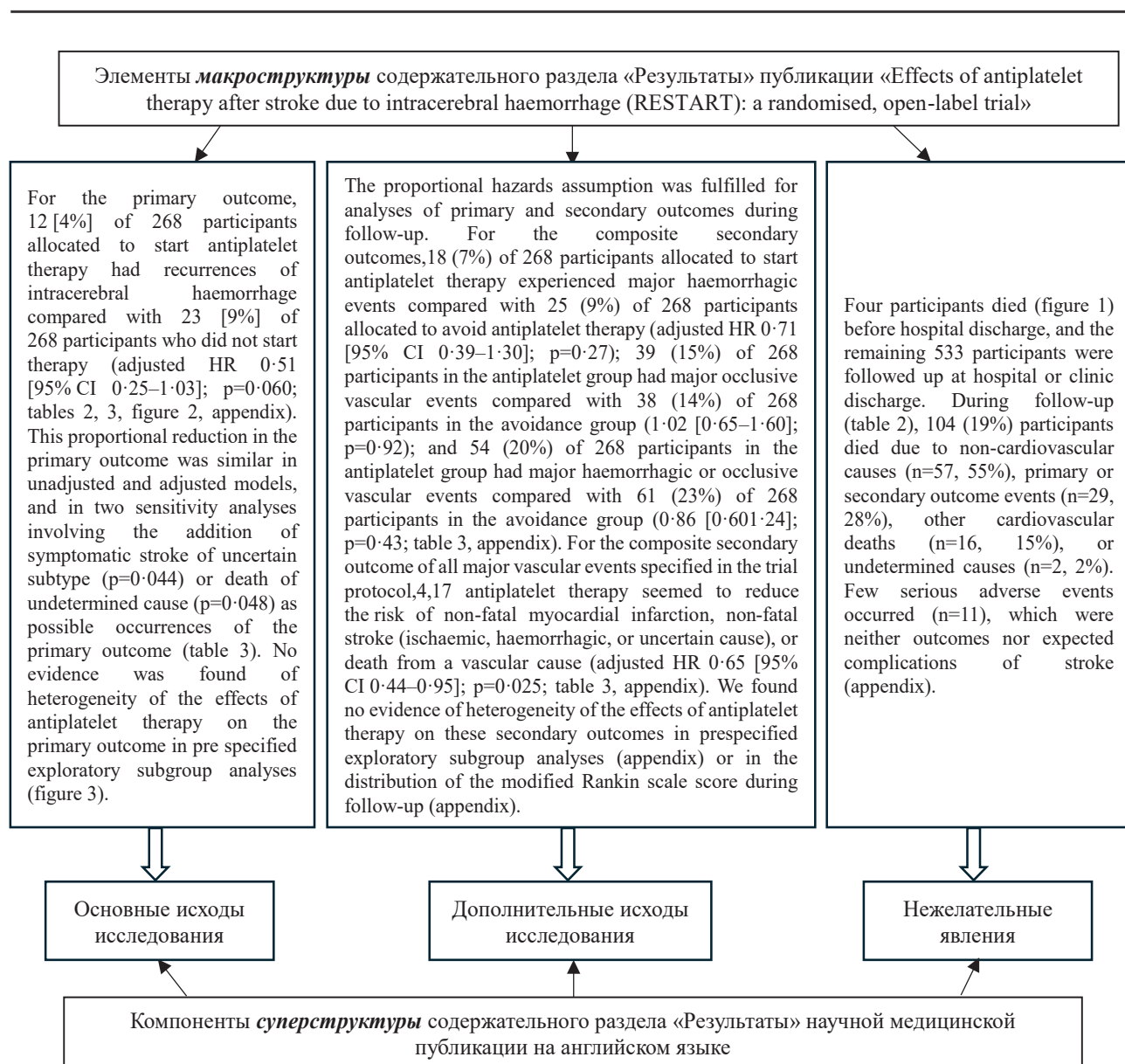


Рис. 2. Распределение макроструктурных элементов научной медицинской публикации на английском языке по компонентам суперструктуры содержательного раздела «Методы»

The proportional hazards assumption was fulfilled for analyses of **primary and secondary outcomes** during follow-up.

For the composite **secondary outcomes**, 18 (7%) of 268 participants allocated to start antiplatelet therapy experienced major haemorrhagic events compared with 25 (9%) of 268 participants allocated to avoid antiplatelet therapy, 39 (15%) of 268 participants in the antiplatelet group had major occlusive vascular events compared with 38 (14%) of 268 participants in the avoidance group, and 54 (20%) of 268 participants in the antiplatelet group had major haemorrhagic or occlusive vascular events compared with 61 (23%) of 268 participants in the avoidance group.

For the composite **secondary outcome** of all major vascular events specified in the trial protocol, antiplate-

let therapy seemed to reduce the risk of non-fatal myocardial infarction, non-fatal stroke, or death from a vascular cause.

Four participants **died** before hospital discharge, and the remaining 533 participants were followed up at hospital or clinic discharge.

During follow-up, 104 (19%) participants **died** due to non-cardiovascular causes ($n=57$, 55%), **primary or secondary outcome** events ($n=29$, 28%), other cardiovascular **deaths** ($n=16$, 15%), or undetermined causes ($n=2$, 2%).

Что касается **первичного результата**, то у 12 (4%) из 268 участников, которым была назначена антитромбоцитарная терапия, наблюдались рецидивы внутримозгового кровоизлияния по сравнению с 23 (9%) из 268 участников, которые не начинали терапию.

Предположение о пропорциональности рисков было выполнено при анализе *первичных и вторичных исходов* в ходе последующего наблюдения.

Что касается комбинированных *вторичных исходов*, то у 18 (7%) из 268 участников, которым была назначена антитромбоцитарная терапия, наблюдались серьезные геморрагические явления по сравнению с 25 (9%) из 268 участников, которым было рекомендовано избежать антитромбоцитарной терапии, у 39 (15%) из 268 участников в группе антитромбоцитарных препаратов наблюдались серьезные сосудистые осложнения по сравнению с 38 (14%) из 268 участников в группе, не принимавшей тромбоциты, и у 54 (20%) из 268 участников в группе, принимавшей антитромбоцитарные препараты, были серьезные геморрагические или окклюзионные сосудистые осложнения по сравнению с 61 (23%) из 268 участников в группе, не принимавшей тромбоциты.

Что касается комплексного *вторичного исхода* всех основных сосудистых событий, указанных в протоколе исследования, то антитромбоцитарная терапия, предположительно, снижала риск нефатального инфаркта миокарда, нефатального инсульта или смерти от сосудистой причины.

Четверо участников **умерли** до выписки из больницы, а остальные 533 участника находились под наблюдением после выписки из больницы или клиники.

За время наблюдения 104 (19%) участника **умерли** от причин, не связанных с сердечно-сосудистыми заболеваниями (n=57, 55%), *первичных или вторичных исходов* (n=29, 28%), других сердечно-сосудистых *смертей* (n=16, 15%) или от неустановленных причин (n=2, 2%) .

Исходный текст раздела «Результаты» включает 1024 слова, в то время как полученный реферат данного раздела – 255 слов. Таким образом, выделе-

ние макроструктуры текста публикации при помощи дискурсивных маркеров и последующее применение дополнительных условий сжатия привели к уменьшению объема текста в 4 раза, сохранив его смысл.

Закключение. Последовательное иерархическое взаимодействие структур поверхностного и глубинного уровней являются основой реализации смысловой компрессии текстов. Выявленные исследователем в ходе анализа элементы субструктуры, суперструктуры и микроструктуры взаимодействуют для создания макроструктуры – эмерджентного образования нового уровня организации, содержащим проекцию целостного смыслового представления текста. Полученные результаты обеспечили возможность создания инструмента смысловой компрессии научных медицинских публикаций – автоматической компьютерной программы Medical Compression (АКП MedCompress), которая предоставляет возможность получения машинного, текстового, информативного, общего или запросного, ориентированного на предметную область медицины реферата, а также составления лексического минимума узких областей медицины.

Литература

1. Дейк, Т.А. ван. Язык. Познание. Коммуникация: сб. работ / Т.А. ван Дейк; сост. В.В. Петров; пер. с англ.; под ред. В.И. Герасимова; вступ. ст.: Ю.Н. Караулов, В.В. Петров. – М.: Прогресс, 1989. – 310 с.
2. Карпилович, Т.П. Когнитивно-коммуникативная модель смысловой компрессии научного текста: дис. ... д-ра филол. наук: 10.02.19; 10.02.21 / Карпилович Татьяна Павловна; Мин. гос. лингв. ун-т. – Минск, 2005. – 223 л.
3. Effects of antiplatelet therapy after stroke due to intracerebral haemorrhage (RESTART): a randomised, open-label trial / R. Al-Shahi Salman [et al.] // Lancet. – 2019. – Vol. 393. – P. 2613–2623.

Поступила в редакцию 29.11.2024