

## Формализация раздела «Обсуждение» научной медицинской статьи на английском языке

**Меньшенина И.А.**

Учреждение образования «Витебский государственный ордена Дружбы народов  
медицинский университет», Витебск

*Стремительный рост объемов научной медицинской документации обуславливает необходимость создания системы автоматического реферирования научных медицинских текстов, которая позволит ускорить процесс обработки информации специалистами, предоставляя возможность просматривать публикации в сжатом варианте. Наиболее релевантным для разработки системы компрессии медицинского текста видится комбинированный подход, включающий сочетание статистических и лингвистических методов для формализации содержания научной медицинской статьи на английском языке при помощи выделения основных смысловых компонентов ее разделов, а также определения лексических и грамматических маркеров, вербализующих данные компоненты в тексте.*

*Цель данной работы – провести анализ структурно-семантической организации раздела «Обсуждение» текстов научной медицинской статьи на английском языке, раскрыть коммуникативные задачи каждого из компонентов данного раздела, а также способы их маркирования.*

**Материал и методы.** *Практическим материалом исследования послужили научные статьи, опубликованные в ведущих и наиболее цитируемых британских и американских научных медицинских журналах в период с 2017 по 2020 год. В процессе работы были использованы такие лингвистические методы, как сравнительный, контекстуальный контент- и дискурс-анализ, статистические методы.*

**Результаты и их обсуждение.** *Определение закономерностей структурно-семантической организации раздела «Обсуждение» научной медицинской статьи позволило конкретизировать наиболее важные смысловые компоненты данного типа текстов, а также выделить лексические и грамматические маркеры, при помощи которых смысловые сегменты вербализуются в тексте. В ходе анализа были выделены 5 прототипических компонентов суперструктуры раздела «Обсуждение» научной медицинской статьи на английском языке с выявлением наиболее употребительных языковых маркеров для каждого из них.*

**Заключение.** *Полученные результаты предоставляют возможность формализации распознавание семантических компонентов раздела «Обсуждение» научной медицинской статьи на английском языке в целях их последующего сокращения и преобразования для построения реферата.*

**Ключевые слова:** *научная медицинская статья, формализация, суперструктура, автореферирование, лингвистические маркеры.*

(Ученые записки. – 2021. – Том 34. – С. 151–155)

## Formalization of the “Discussion” Section of the Scientific Medical Article in English

**Menshenina I.A.**

Education Establishment “Vitebsk State Order  
of Peoples' Friendship Medical University”, Vitebsk

*The rapid growth in the volume of scientific medical documentation necessitates the creation of an automatic summarization system for scientific medical texts, which will speed up the process of information processing by specialists, providing the opportunity to read publications in a compressed version. The most relevant for the development of a medical text compression system is a combined approach, including a combination of statistical and linguistic methods to formalize the content of a scientific medical article in English by highlighting the main semantic components of its sections, as well as determining lexical and grammatical markers that verbalize these components in the text.*

The purpose of this work is to analyze the structural and semantic organization of the "Discussion" section of the texts of a scientific medical article in English, to reveal the communicative tasks of each of the components of this section, as well as ways of marking them.

**Material and methods.** The practical material of the study was scientific articles published in the leading and most cited British and American scientific medical journals in the period from 2017 to 2020.

In the course of the work, such linguistic methods as comparative, contextual content and discourse analysis were used. In addition, statistical methods were used.

**Findings and their discussion.** The determination of the regularities of the structural and semantic organization of the "Discussion" section of a scientific medical article made it possible to specify the most important semantic components of this type of texts, as well as to identify lexical and grammatical markers with which semantic segments are verbalized in the text. During the analysis, 5 prototypical components of the superstructure of the "Discussion" section of a scientific medical article in English were identified with the identification of the most commonly used language markers for each of them.

**Conclusion.** The obtained results provide an opportunity to formalize the recognition of semantic components of the "Discussion" section of a scientific medical article in English in order to reduce and transform them for the construction of an abstract.

**Key words:** scientific medical article, formalization, superstructure, autosummarization, linguistic markers.

(Scientific notes. – 2021. – Vol. 34. – P. 151–155)

**П**роблема информационной перегрузки, характерная для современного мира, затрагивает все профессиональные области, в том числе и сферу медицины. Для обеспечения надлежащего информационного обслуживания специалиста (врача) представляется необходимым разработать систему автоматического реферирования научных медицинских текстов, которая позволит ускорить процесс обработки информации, предоставляя возможность просматривать публикации в сжатом варианте. Наиболее релевантным для разработки системы компрессии медицинского текста видится комбинированный подход, включающий сочетание статистических и лингвистических методов для формализации содержания научной медицинской статьи на английском языке при помощи выделения основных смысловых компонентов ее разделов, а также определения лексических и грамматических маркеров, вербализующих данные компоненты в тексте. В статье рассматривается структурно-семантическая организация раздела «Обсуждение» текстов научной медицинской статьи на английском языке, раскрываются коммуникативные задачи каждого из компонентов суперструктуры данного раздела, а также способы их маркирования.

Цель работы – анализ структурно-семантической организации раздела «Обсуждение» текстов научной медицинской статьи на английском языке, раскрыть коммуникативные задачи каждого из компонентов данного раздела, а также способы их маркирования.

**Материал и методы.** Практическим материалом исследования послужили научные статьи, опубликованные в ведущих и наиболее цитируемых британских и американских научных медицинских журналах в период с 2017 по 2020 год. В процессе работы были использованы такие лингвистические методы, как сравнительный, контекстуальный контент- и дискурс-анализ, статистические методы.

**Результаты и их обсуждение.** Раздел «Обсуждение» представляет собой наиболее творческую и вместе с тем сложную для написания часть научной медицинской статьи. Для данного раздела характерно использование композиционно-речевой формы «рассуж-

дение» в отличие от «описания» и «повествования» в других разделах [1]. Такая особенность обусловлена определенными коммуникативными функциями: обобщить и проанализировать полученные в ходе эксперимента данные, сопоставить результаты с предыдущими исследованиями, сформулировать выводы, продемонстрировав сильные стороны своей работы, а также, как правило, указать на вопросы, требующие дальнейшей проработки. Помимо приведения достоинств исследования, в задачи раздела входит обсуждение всех возможных недостатков, систематических ошибок и ограничений.

Анализ смысловой структуры данного раздела показал, что она включает 5 основных компонентов:

1. Ответ на гипотезу.
2. Сравнение с другими исследованиями.
3. Ограничения исследования.
4. Внешняя валидность.
5. Перспективы исследования.

В 62% проанализированных статей (95% из которых опубликованы в британских изданиях) раздел «Обсуждение» начинается с ответа на гипотезу или задачу, поставленную во «Введении», либо с повторения в сжатом виде (чаще всего 1 предложение) цели или гипотезы исследования с последующим ответом на них.

Соответственно, маркерами подкомпонента суперструктуры «Ответ на гипотезу» будут выступать следующие выражения: *protocol was effectively delivered* 'протокол 'был успешно выполнен'; *we have achieved this aim* 'Мы достигли этой цели'; *we also noted* 'Мы также отметили'; *we noted* 'мы отметили'; *Data shows* 'Данные показывают'; *The findings from our trial provide conclusive evidence that* 'Результаты нашего исследования убедительно свидетельствуют о том, что'; *our results suggest that* 'наши результаты предполагают, что' и др.

*Our very early mobilisation protocol was effectively delivered, leading to an earlier, more frequent, and higher dose of out-of-bed sitting, standing, and walking activity than usual care.* 'Наш очень ранний протокол мобилизации был эффективно реализован, что привело к более ранней, более частой и более продолжительной

способности находится в положении сидя, стоя, а также активно ходить, чем при обычном уходе’.

Коммуникативная функция второго семантического компонента раздела «Обсуждение» заключается в критическом анализе имеющихся источников литературы и в сравнении полученных данных с результатами других авторов. Как правило, автор статьи указывает, с какими работами его выводы совпадают, от каких отличаются и почему; что было исследовано другими учеными и почему в данной работе это не освещено; что обнаружено в данной работе, а в других не исследовалось и т.д. [2].

Семантический компонент «Сравнение с другими исследованиями» вербализуется в тексте статьи при помощи маркеров *difference/differences* ‘разница / различия’; *accords with* ‘согласуется с’; *match* ‘совпадает’; *in direct contrast to* ‘в прямой противоположности’; *contradict* ‘противоречить’; *support* ‘поддерживать’; *be comparable to* ‘быть сопоставимым с’; *compare* ‘сравнивать’ и др.

*Despite different durations, the proportion of patients with bleeding events was similar between our study (12% at 9 months) and the LEADERS FREE trial (14% at 12 months).* ‘Несмотря на разную продолжительность, доля пациентов с кровотечениями была одинаковой в нашем исследовании (12% через 9 месяцев) и исследовании LEADERS FREE (14% через 12 месяцев)’.

Обсуждение результатов исследования сопровождается анализом факторов, способных повлиять на возможность применения полученных результатов за рамками эксперимента. Так называемые ограничения исследования могут касаться любого его этапа (выборки, процесса рандомизации, ослепления, применяемых методов и т.д.). Наиболее существенными по мнению ученых считаются ограничения, связанные с методами исследования, а именно условиями его проведения, объемом и качеством выборки, использованными способами рандомизации и пр. [3].

Предложения компонента «Ограничения исследования» маркируются лимитированным количеством лексем, таких как: *limitation/limitations* ‘ограничения / ограничения’, *was / were not able* ‘не мог / не могли’, *was / were / is / are unable to* ‘был / были / не в состоянии’, *could not* ‘не мог’, *cannot* ‘не может’, *was limited to* ‘был ограничен’, *was restricted by* ‘был ограничен’, *weaknesses / weakness* ‘слабые стороны / сторона’, *failed to* ‘не смог’ и т.д.

Самыми распространенными предложениями, вводящими сведения об ограничениях исследования, являются: *Our trial has several limitations.* ‘Наше исследование имеет несколько ограничений’. *Our study has some limitations.* ‘Наше исследование имеет некоторые ограничения’. *Several limitations in our study should be noted.* ‘Следует отметить несколько ограничений в нашем исследовании’. *However, several limitations are also present.* ‘Однако есть и несколько ограничений’. *Our report was limited to...* ‘Наш отчет ограничился...’ *We failed to obtain....* ‘Нам не удалось получить...’.

Предложения обычно начинаются с личного местоимения *we* ‘мы’ либо с притяжательного местоимения *our* ‘наш’.

В большинстве случаев непосредственно сами ограничения перечисляются в последующих предложениях, которые начинаются с *first* ‘во-первых’, *second* ‘во-вторых’, *third* ‘в-третьих’, *fourth* ‘в-четвертых’, *finally / last* ‘наконец, в довершение’. Следовательно, такие лексемы также являются маркерами данного семантического компонента, а предложения, которые начинаются с них, отбираются для реферата.

В том случае, если вышеприведенные лексемы не следуют за предложением, содержащим «*limitation* ‘ограничение’, т.д.», в реферат отбираются все предложения до конца абзаца.

*Nevertheless, several limitations apply. First, the broad population might have included groups more likely to either respond (eg, those with minor stroke or TIA, or atherosclerotic disease) or have a major bleed (eg, those receiving thrombolysis, or having small vessel disease), which might explain the neutral results; future trials of antiplatelets might need to be more specific and focus on individuals with atherosclerotic disease. Second, the antiplatelet agents were administered in an open-label design and participants knew which drugs they were on. Third, the comparator group involved different antiplatelet agents, a situation reflecting changes in national and international guidelines that added monotherapy with clopidogrel to the existing recommendation of combined aspirin and dipyridamole. Fourth, randomly assigned treatments were given for 30 days, which might have been too long in view of the identified haemorrhage risk. Last, the trial was stopped early following recommendation by the independent data monitoring committee and the results could represent a false neutral finding related to the lower-than-planned statistical power.* ‘Тем не менее, существует несколько ограничений. Во-первых, широкая популяция могла включать группы, которые с большей вероятностью либо ответят (например, пациенты с легким инсультом или ТИА, либо атеросклеротическим заболеванием), либо имеют большое кровотечение (например, те, кто получает тромболитис или страдают заболеванием мелких сосудов), что может объяснить нейтральные результаты; возможно, будущие испытания антитромбоцитов должны быть более конкретными и сосредоточенными на лицах с атеросклеротическим заболеванием. Во-вторых, антитромбоцитарные препараты вводились открыто, и участники знали, какие препараты они принимают. В-третьих, группа сравнения включала различные антиагреганты, что отражает изменения в национальных и международных руководствах, которые добавляли монотерапию клопидогрелем к существующей рекомендации комбинированного аспирина и дипиридамола. В-четвертых, рандомизированное лечение проводилось в течение 30 дней, что могло быть слишком долгим с учетом установленного риска кровотечения. Наконец, испытание было остановлено досрочно по

рекомендации независимого комитета по мониторингу данных, и результаты могли представлять собой ложно-нейтральный вывод, связанный с более низкой, чем планировалось, статистической мощностью’.

В то же время ограничения могут быть описаны одним предложением.

*A potential limitation of our study is the short follow-up time of 12 months post-STEMI for the primary outcome of cardiac death and hospitalisation for heart failure, which might have been too short to observe any effect of ischaemic conditioning on clinical outcomes.* Потенциальным ограничением нашего исследования является короткий период наблюдения в 12 месяцев после STEMI для отслеживания первичного исхода сердечной смерти и госпитализации по поводу сердечной недостаточности, что могло быть слишком непродолжительным временем для отслеживания возможного влияния ишемической болезни на клинические исходы’.

Следующий подкомпонент раздела «Обсуждение» демонстрирует важность исследования и обобщаемость его результатов.

Уникальность работы подчеркивается превосходной, в некоторых случаях сравнительной степенью прилагательных, порядковым числительным *the first* ‘первый’:

*The longest trial ever done.* ‘Самое долгое из когда-либо проведенных испытаний’. *The largest trial done so far.* ‘Крупнейшее исследование, проведенное до сих пор. *The only and therefore the largest randomised trial.* ‘Единственное и, следовательно, самое крупное рандомизированное исследование’. *The largest trial so far for this problem.* ‘На данный момент самое крупное исследование по этой проблеме’.

Такие заявления обычно смягчаются фразами, выражающими мнение, такими как: *to our knowledge* ‘насколько нам известно’; *as far as we are aware* ‘насколько нам известно’; *to the best of our knowledge* ‘в меру наших знаний’ и т.д.

Каждое предложение включает либо личное местоимение первого лица *we* ‘мы’, либо указательные местоимения *this / these* ‘этот / эти’, либо притяжательное местоимение *our* ‘наш’:

*Our pragmatic study is the largest trial so far for this problem.* ‘Наше прагматическое исследование является крупнейшим на сегодняшний день исследованием этой проблемы’.

*These findings provide support for...* ‘Эти выводы обеспечивают поддержку’.

Предложения, описывающие сильные стороны исследования, маркируются следующими лексемами: *strength/ strengths, external validity, to strengthen, implication/ implications.* Как и в предыдущем подкомпоненте, перечисление достоинств исследования происходит с помощью лексем *first* ‘во-первых’, *second* ‘во-вторых’, *third* ‘в-третьих’, *fourth* ‘в-четвертых’, т.д. и *finally / last* ‘наконец, в довершение’. Например: *This trial has five major strengths.* ‘У этого испытания есть пять основных сильных сторон’. *The strengths of*

*the study, ..., are...* ‘Достоинства данного исследования это’. *The strengths of the study include...* ‘Достоинства исследования включают...’. *Our study has several broad clinical implications.* ‘Наше исследование имеет несколько широких клинических значений’.

*This trial has five major strengths. First, we included a broad and large population of patients who were representative of those referred to the cardiology clinic for assessment of suspected angina due to coronary heart disease. Second, we specifically included patients who had been excluded from previous studies of CTCA such as those with obesity, high calcium scores, and atrial fibrillation. Third, we allowed unrestricted use of further stress imaging in keeping with clinician choice and routine practice because we wanted to explore the effect of the addition of CTCA as opposed to doing a head-to-head comparison with other diagnostic approaches. Fourth, we focused on the effect of CTCA on patient-centred and clinician-centred outcomes rather than comparing diagnostic accuracy between imaging modalities or anatomic versus functional testing. Finally, we assessed the effect of this intervention on both short-term and long-term outcomes to define the impact of this additional imaging intervention to routine clinical practice.* ‘У этого испытания есть пять основных сильных сторон. Во-первых, мы включили широкую и большую популяцию пациентов, которые были репрезентативными для тех, кого направили в кардиологическую клинику для оценки подозрения на стенокардию вследствие ишемической болезни сердца. Во-вторых, мы специально включили пациентов, которые были исключены из предыдущих исследований СТСА, таких как пациенты с ожирением, высокими показателями кальция и фибрилляцией предсердий. В-третьих, мы разрешили неограниченное использование дополнительной визуализации стресса в соответствии с выбором клинициста и рутинной практикой, потому что мы хотели изучить эффект добавления СТСА, а не проводить прямое сравнение с другими диагностическими подходами. В-четвертых, мы сосредоточились на влиянии СТСА на результаты, ориентированные на пациента и клинициста, а не на сравнение точности диагностики между методами визуализации или анатомическим и функциональным тестированием. Наконец, мы оценили влияние этого вмешательства как на краткосрочные, так и на долгосрочные результаты, чтобы определить влияние этого дополнительного вмешательства с визуализацией на повседневную клиническую практику’.

Во многих статьях выводы по исследованию и его перспективы не выносятся в отдельный раздел «Заключение», а включены в последний абзац «Обсуждения» и эксплицируются в тексте статьи глаголами *to conclude* ‘заключать’, *to sum up* ‘подводить итог’, а также выражениями *in conclusion* ‘в заключение’, *in summary* ‘таким образом’ и др.

*In summary, tranexamic acid did not affect functional status at day 90, although potential benefits were seen with reductions in haematoma expansion, early death, and*

*serious adverse events*. 'Таким образом, транексамовая кислота не влияла на функциональное состояние на 90-й день, хотя потенциальные преимущества наблюдались в виде уменьшения расширения гематомы, ранней смерти и серьезных нежелательных явлений'.

Семантический компонент «Перспективы исследования» является последним, пятым компонентом раздела «Обсуждение», где ретроспективно представляется новое знание и прогнозируется его дальнейшее развитие и будущие перспективы. Здесь же, как правило, излагаются ключевые нерешенные аспекты обозначенной проблемы и указывается на необходимость дальнейших ее исследований.

В преобладающем большинстве предложений этого подкомпонента встречается будущее время глагола либо модальные глаголы *should, might, may* (реже) и такие лексемы, как: *future* 'будущий', *further* 'далее', *is / are needed* 'необходимо / необходимы', *need / needs* 'нуждается в / нуждаются в', *might be needed* 'могут потребоваться', *is / are warranted* 'оправданы', *there is a need for* 'есть необходимость', *is/are worthy of* 'достойны / необходимы', *is /are necessary* 'необходим / необходимы' и т.д.

*Our next priority will be to undertake... 'Следующим нашим приоритетом будет предпринять...'. Future studies are desirable to explore... 'Желательно, чтобы будущие исследования изучили ...'. ... merits further study for... '... заслуживает дальнейшего изучения для ...'. Further studies are needed... 'Необходимы дальнейшие исследования ...'. Further evaluation of... could be warranted. Может потребоваться дальнейшая оценка.... Future research should investigate... 'В будущих исследованиях следует изучить...'. These findings are worthy of future study. 'Эти результаты заслуживают дальнейшего изучения'. ... are important areas for future investigations. '... являются важными областями для будущих исследований'. Future efforts should include... 'Будущие усилия должны включать...'.*

*Future research should investigate which subgroups of patients might benefit. Tranexamic acid is inexpensive, easy to administer, seems to be safe, and is widely available, so even a modest treatment effect could have an important impact on the global scale. Larger randomised trials are warranted. 'Дальнейшие исследования должны изучить какие подгруппы пациентов могут получить пользу. Транексамовая кислота недорогая, простая в применении, кажется безопасной, и широко доступна, поэтому даже умеренный лечебный эффект может иметь важное влияние в мировом масштабе. Требуются более крупные рандомизированные испытания'.*

Не все представленные выше маркеры могут быть безусловным признаком принадлежности предложения к тому или иному семантическому компонен-

ту суперструктуры научной медицинской статьи на английском языке. Некоторые из них являются многозначными и не могут быть однозначным идентификатором одного смыслового компонента. Наибольшую сложность для формализации представляют маркеры, состоящие из одной лексемы (*finally* 'наконец', *thus* 'таким образом', *however* 'однако' и др.).

Например, коннектор *however* 'однако' часто выделяется как маркер для опознавания пробелов предыдущих работ. Но в разделе «Обсуждение» данная лексема может быть использована автором при описания особенностей или недостатков собственного исследования:

*There was no difference in functional outcome measured using the mRS when we compared groups who continued versus stopped prestroke antihypertensive drugs. However, some secondary outcomes were worse in the group who continued treatment, although whether this finding is due to chance or is a real effect is not clear. 'Не было различий в функциональных результатах, измеренных с помощью mRS, когда мы сравнивали группы, которые продолжали и прекратили прием гипотензивных препаратов до инсульта. Все же некоторые вторичные исходы были хуже в группе, продолжавшей лечение, хотя не ясно, является ли этот результат случайным или закономерным'.*

**Заключение.** Рассмотренный подход, основанный на закономерностях структурно-семантической организации научной статьи, в сочетании с методами маркирования информации делает возможным алгоритмизацию поиска семантических компонентов раздела «Обсуждение» медицинского научного текста. Следующий этап работы предусматривает сокращение отобранных фрагментов исходного текста с последующим их преобразованием в целях построения связного текста аннотации.

## Литература

1. Шамара, И.Ф. Научный медицинский дискурс с позиций системного исследования [Электронный ресурс] / И.Ф. Шамара // Теория яз. и межкультур. коммуникация. – 2013. – № 1. – Режим доступа: <http://tl-ic.kursksu.ru/pdf/013-010.pdf>. – Дата доступа: 14.10.2021.
2. Хисамов, А.А. Как написать медицинскую статью? Структура медицинской статьи / А.А. Хисамов // Злокачественные опухоли. – 2016. – № 1. – С. 44–47.
3. Сайгитов, Р.Т. Правила и рекомендации по представлению рукописей, содержащих результаты оригинальных исследований [Электронный ресурс] / Р.Т. Сайгитов // Вопросы соврем. педиатрии. – 2015. – Т. 14, № 3. – Режим доступа: <https://publications.hse.ru/articles/154307819>. – Дата доступа: 13.10.2021.

Поступила в редакцию 13.12.2021