

DOI: 10.33184/dokbsu-2026.3.23

## **Сопоставительный анализ дискурсивных стратегий в научном медицинском тексте: методология графосемантического моделирования**

**Л. А. Линник**

*Башкирский государственный медицинский университет  
Россия, Республика Башкортостан, 450008 г. Уфа, ул. Ленина, 3.*

*Email: lyubovlinnik07@gmail.com*

В фокусе статьи – разработка методологии сопоставительного анализа дискурсивных стратегий в русско- и англоязычных медицинских аннотациях на базе метода графосемантического моделирования. В отличие от традиционного структурно-семантического подхода, акцентирующего когезию и когерентность, предлагаемая модель рассматривает текст как проекцию когнитивных сценариев автора, регулирующих степень экспликации субъекта и терминологическую плотность текста. Выявлены две базовые стратегии: статусно-ориентированная (русскоязычный дискурс) и диалогово-ориентированная (англоязычный дискурс). Инструментарий графосемантического моделирования позволяет визуализировать данные стратегии в виде графосемантических моделей, что открывает перспективы для типологического изучения научного дискурса.

**Ключевые слова:** медицинский дискурс, аннотация, графосемантическое моделирование, когнитивно-коммуникативная стратегия, сопоставительное языкознание, научный текст, авторская интенция.

Состояние современной лингвистики характеризуется переходом от линейного изучения текстовых параметров к многомерному моделированию дискурсивной деятельности. Текст перестает восприниматься лишь как сумма когезийных и когерентных связей [1]; на передний план выходит его интерпретация как динамической самоорганизующейся системы, в которой пересекаются когнитивные, социальные и собственно языковые факторы [2–3]. Особую значимость такой ракурс приобретает в сопоставительных исследованиях научного дискурса, где различия в языковом кодировании отражают не столько грамматические расхождения, сколько несовпадающие традиции организации приращения и трансфера знания и взаимодействия с адресатом [4].

Цель данной работы – обосновать применение метода графосемантического моделирования для сопоставительного конструирования дискурсивных стратегий в русско- и англоязычных медицинских аннотациях. Выбор аннотации в качестве эмпирического материала не случаен: будучи вторичным, компрессированным жанром, она

в концентрированном виде манифестирует те когнитивные сценарии, которые автор считает релевантными для вхождения в дискурсивное пространство [5]. Материалом послужили корпуса аннотаций медицинских статей с наибольшим индексом цитирования, который принимается нами как индекс степени интеграции в научную коммуникацию, опубликованных в ведущих российских медицинских журналах, собранных в базах данных e-library и PubMed.

Традиционная лингвистика текста, сформировавшаяся в последней четверти XX в., наделила текст набором онтологических признаков, в ряду которых центральными признаны связность (когезия) и содержательная цельность (когерентность) [1]. Под когезией понимается сеть поверхностных лексико-грамматических отношений, обеспечивающих континуум текста, тогда как когерентность выводит на глубинный уровень смыслового единства, реализуемого через тема-рематические прогрессии, посредством которых и осуществляется трансфер и приращение знания. Эти механизмы, подробно описанные в отечественном языкознании, легли в основу деривационного подхода, трактующего порождение текста как последовательное инкорпорирование смысловых блоков, где каждая предыдущая конструкция выступает темой для последующей [6–7].

Однако при сопоставительном изучении текстов, функционирующих в разных лингвокультурных средах, структурно-семантического инструментария оказывается недостаточно. Фиксация различий в синтаксической сложности, частотности пассивных конструкций или объеме терминологического слоя не отвечает на вопрос о том, какие когнитивные и коммуникативные стратегии управляют отбором этих средств. Для ответа на этот вопрос мы обращаемся к методологии графосемантического моделирования [5], которая трактует текст как многоплановую распределенную систему. В ее рамках содержание не просто членится на микротемы, а репрезентируется как иерархия субцельностей, связанных с речевыми тактиками автора [3].

В условиях ограниченного доступа к специализированным платформам для графосемантического моделирования (где узлами графа выступают концептуальные доминанты, а ребрами – отношения обусловленности, уточнения или альтернативы) в качестве операционального инструментария был использован комплекс Voyant Tools. Он позволяет реализовать принципы семантического картирования на уровне лексико-семантической структуры текста. Так, модуль Cirrus (облако слов) визуализирует концептуальные доминанты как ядро терминологического поля; WordsTree (дерево слов) реконструирует локальные семантические связи, соответствующие отношениям уточнения и обусловленности; а Terms Radio (динамика частотности слов) фиксирует динамику актуализации ключевых смыслов по сегментам текста, что позволяет проследить распределение субцельностей и смену речевых тактик автора в рамках общей иерархии интенций.

Преимущество этого подхода в сопоставительном аспекте заключается в возможности визуализировать не только то, что сообщается, но и то, как организуется дискурсивное взаимодействие. Визуализируются не просто термины, а стоящие за ними когнитивные сценарии: стратегия дистанцирования автора от утверждения, стратегия прямого диалога с читателем, стратегия иерархизации доказательств и др.

Мы использовали сервис Voyant Tools, поскольку именно он позволяет исследовать тексты малых размеров и по количественным характеристикам (частотность слов, длина предложений, лексическая плотность и т.д.), и по качественным (позволяет выявить семантические паттерны и тематические кластеры). В самом сервисе мы использовали функции резюме (сбор количественных характеристик текста) и инструменты визуализации, описанные нами выше.

Приведем пример анализа аннотаций двух статей с высоким индексом научного цитирования: «Сахарный диабет в российской федерации: динамика эпидемиологических показателей по данным федерального регистра сахарного диабета за период 2010–2022 гг.» (2023 г., индекс цитирования 525 на июнь 2026 г.) и «Global cancer statistics 2024 г: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries» (2024 г., индекс цитирования 28 601 на июнь 2026 г.).

Количественные характеристики текстов представим в *табл. 1*.

Таблица 1

Сопоставление русскоязычной и англоязычной аннотаций по данным функции сервиса Voyant Tools «Резюме»

| Параметр                                                    | Русскоязычная аннотация                          | Англоязычная аннотация                                         |
|-------------------------------------------------------------|--------------------------------------------------|----------------------------------------------------------------|
| Общее количество уникальных словоформ                       | 215                                              | 163                                                            |
| Лексическая плотность                                       | 0,591                                            | 0,578                                                          |
| Readability Index <sup>1</sup><br>(индекс «читабельности»): | 17,345                                           | 14,395                                                         |
| Среднее количество слов в предложении                       | 24,3                                             | 35,3                                                           |
| Самые часто встречающиеся слова в корпусе                   | сд2 (11); сд1 (10); тыс (8);<br>сд (8); года (8) | cancer (15); million (7); rates (4);<br>100,000 (4); women (3) |

<sup>1</sup> Automated Readability Index (ARI) — метрика, калиброванная под английский язык, и её значения для русского текста носят ориентировочный характер.

Лексическая плотность почти одинаковая (0,591 и 0,578), оба текста содержат сопоставимую долю знаменательных слов, ни один из текстов не «разбавлен» служебными словами сильнее другого. Наибольшая разница заключается в количестве словоформ, длине предложений и, что обращает на себя особое внимание, индексе «читабельности», который характеризует сложность и степень открытости текста как пространства дискурсивного взаимодействия к включению большего количества участников научной коммуникации. Чем выше индекс, тем сложнее и более «закрыт» для коммуникации текст. Большее уникальных словоформ в русскоязычной аннотации свидетельствует о большей лексической вариативности, которая связана в научном дискурсе с использованием узкоспециальных терминов и их форм. Существенно отличается длина предложений: русскоязычная аннотация тяготеет к использованию аббревиатур и сокращений, англоязычная – к развернутым формулировкам и полной номинации терминов. Различие в индексах читабельности маркирует не только доступность текста, но и разную степень терминологической компрессии и их различную прагматическую направленность: категоричность и компактность русскоязычного и осторожное представление результатов англоязычного при сопоставимой информационной насыщенности (лексической плотности). Также и самые частые слова русскоязычной аннотации преимущественно аббревиатуры и сокращения, в англоязычной – полные тематические термины и числительные.

Инструменты Cirrus и TermsRadio использовались для качественного анализа лексики, данные визуализации представлены на *рис. 1–2*.



оба текста ориентированы во времени (в ядре указаны годы и даже даты) и имеют указания на пациентов и на их количество. Однако русскоязычная аннотация рассматривает пациентов более обобщенно (акцент на лексеме *население*), англоязычная – более персонифицировано (акцент на *women, men*, в отличие от русского сокращенного и уменьшенного *муж., жен.*).

TermsRadio позволил оценить распределение этих маркеров по тексту и оценить, насколько они сконцентрированы или рассредоточены, что отражает разные прагматические стратегии подачи информации. Анализ графика дистрибуции частотных лексем по сегментам аннотации дает возможность установить, где именно локализуется высокая лексическая плотность, создать пространственно-временную карту терминологической нагрузки текста, что, в свою очередь, демонстрирует, какие элементы текста влияют на индекс «читабельности» и делают коммуникативное пространство аннотации более закрытым.

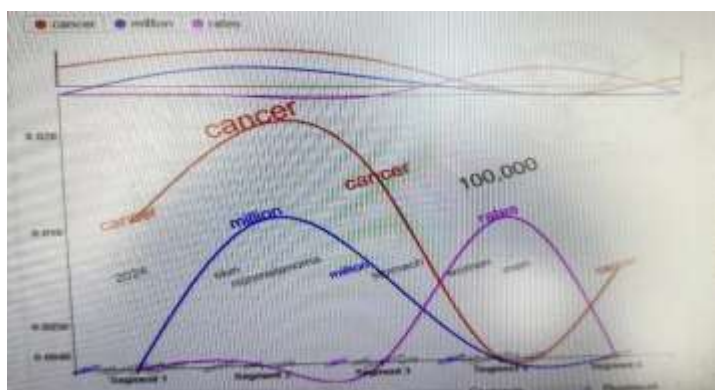
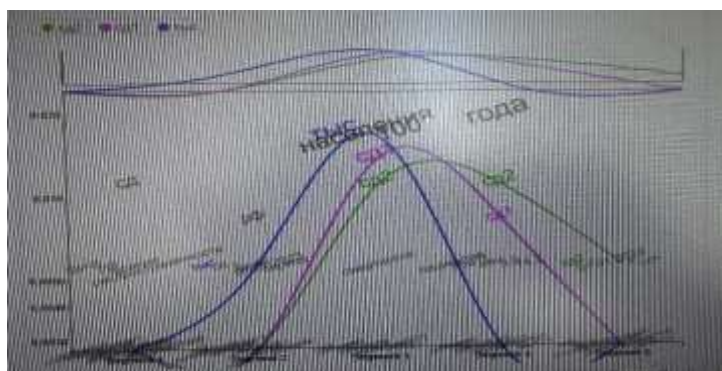


Рис. 2. Визуализация TermsRadio сервиса Voyant Tools.

Визуализация выявляет тактику «терминологических всплесков» в русскоязычной аннотации. Резкий пик частоты слова *года* (сегмент 7) и высокая концентрация сокращения *тыс* (сегмент 5) демонстрируют, что авторы намеренно группируют наиболее насыщенные цифровыми данными фрагменты в отдельных смысловых блоках. В англоязычной аннотации, напротив, частотные слова (*million, rates*) распределены равно-

мерно, а использование полной номинации терминов снижает воспринимаемую сложность текста, отражая установку на широкую доступность данных и открытость коммуникативного пространства научного дискурса.

Семантические маркеры в русском тексте работают как «якоря» данных: термины *сд1* и *сд2* локализованы в секторе представления результатов, маркируя переход к конкретным цифрам. В английском тексте семантические маркеры (*cancer, incidence*) выполняют функцию «скелета» повествования, равномерно структурируя прогнозные данные на протяжении всего текста. Эти различия в топологии терминологической нагрузки подтверждают гипотезу о разных прагматических стратегиях: категоричное утверждение факта в русском тексте против вероятностного представления данных в английском. Сопоставление данных визуализации представим в *табл. 2*.

Таблица 2

Сопоставление русскоязычной и англоязычной аннотаций по данным визуализации сервиса Voyant Tools TermsRadio

| Параметр                                | Русскоязычная аннотация                                                                                                                                                          | Англоязычная аннотация                                                                                                                                                                                                 |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Тип номинации                           | Аббревиатуры и сокращения ( <i>сд2, тыс</i> ). На графике это проявляется как резкие пики и высокая частота условных обозначений.                                                | Полная номинация и стандартные числительные ( <i>million, 100,000</i> ) являются общеупотребительными и не повышают сложность, несмотря на обилие цифр.                                                                |
| Динамика распределения                  | Локализованная концентрация. Термины актуализируются в конкретных сегментах (сегмент 5 для <i>тыс</i> , сегмент 7 для <i>года</i> ), создавая «островки» высокой плотности.      | Равномерное распределение. Лексемы <i>cancer</i> и <i>rates</i> активны на протяжении всего текста, не создавая резких всплесков частоты.                                                                              |
| Семантическая нагрузка                  | Категоричность и факт. Термины привязаны к жестким данным регистра ( <i>ФРСД</i> ). Предикаты рядом с ними ( <i>составила, установлено</i> ) усиливают эффект констатации факта. | Прогноз и вариативность. Частотные слова связаны с вероятностными оценками ( <i>predicted, rates</i> ) и географической вариативностью ( <i>regions</i> ), что направлено на снижение перцептивной «жесткости» текста. |
| Влияние на коммуникативное пространство | Повышение сложности, закрытости. Резкие пики частоты аббревиатур и узкоспециальных сокращений делают текст специализированным.                                                   | Понижение сложности, закрытости. Равномерное распределение общеупотребимой лексики и полных номинаций делает текст алгоритмически «более простым», даже если он длиннее.                                               |

Для выявления лексико-семантических связей и контекстуальных паттернов, релевантных для анализа прагматических стратегий, мы использовали визуализацию «дерево слов», представленную на *рис. 3*.

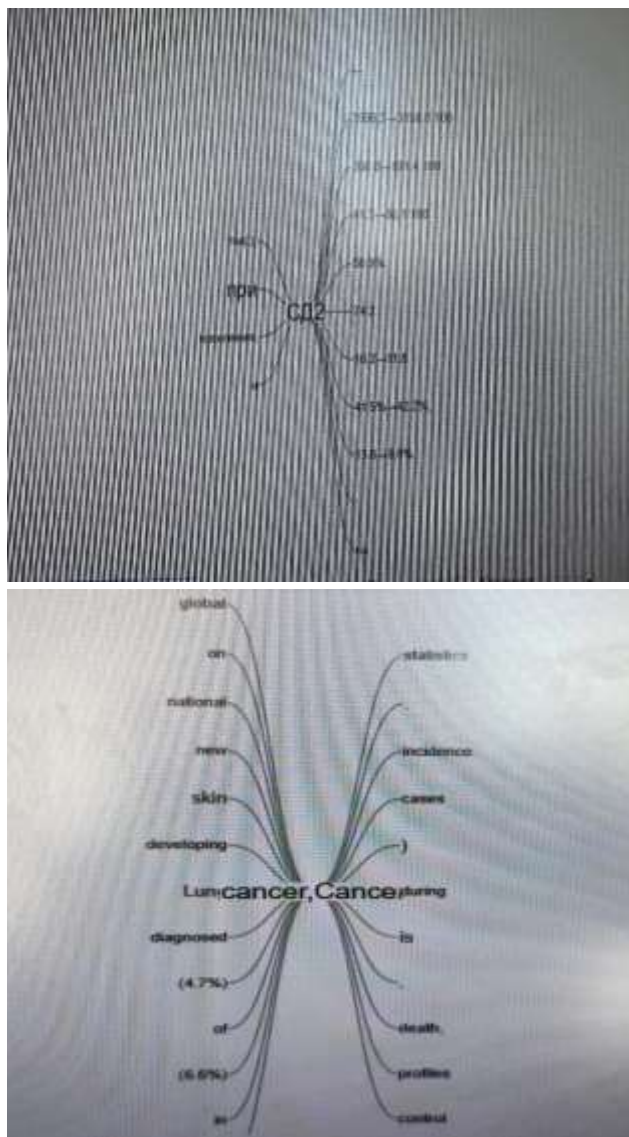


Рис. 3. Визуализация Дерево слов сервиса Voyant Tools.

В русскоязычном тексте семантическое ядро, репрезентируемое термином-аббревиатурой, имеет наиболее плотную связь с числовыми показателями и лексемой *население*, такая конфигурация семантического поля отражает установку на точную количественную характеристику объекта в рамках национальной статистики. В англоязычном тексте семантическое ядро – общеизвестная лексема – демонстрирует лексико-семантическую связь с лексемами *global* (глобальный, мировой), *national* (национальный, по странам), *developing* (развивающиеся, здесь – страны), *diagnosed* (диагностированные), *statistics* (статистические данные, показатели), *incidence* (заболеваемость), *cases*

(случаи), *death* (летальный исход), *profiles* (профили, характеристики), *control* (контроль, борьба, здесь – с раком) и несколькими числовыми показателями и лексемами со значением органа, уточняющими виды заболевания, находящегося в центре семантического ядра, то есть с маркерами масштаба и управления (*global, national, developing, control*) и с показателями динамики (*incidence, cases, death, profiles, statistics*). Таким образом, семантическое ядро здесь ориентировано на репрезентацию проблемы в международном контексте и на управленческие аспекты ее решения.

Проведенное графосемантическое моделирование русско- и англоязычных аннотаций выявило две полярные конфигурации дискурсивных полей.

Первую конфигурацию, характерную для русскоязычного материала, мы определяем как статусно-ориентированную. Ее отличительная черта – высокая плотность лексем, репрезентирующих объективированные факты и процедурные шаги, сравним с фрагментом текста, чтобы подтвердить наш тезис – «*Общая численность пациентов с СД в РФ, состоящих на диспансерном учете, на 01.01.2023 г., по данным ФРСД, составила 4 962 762 (3,31% населения РФ), из них: СД1 – 5,58% (277,1 тыс.), СД2 – 92,33% (4,58 млн), другие типы СД – 2,08% (103 тыс.)*». Исследователь последовательно устраняется из поля зрения читателей, термины даны без пояснений как заведомо известные, маркируя текст как адресованный узкому кругу специалистов и закрывая это дискурсивное пространство для широкой аудитории.

Вторую конфигурацию, типичную для англоязычных изданий, мы квалифицируем как диалогово-ориентированную. Автор проявлен, дискурсивное пространство расширяется за счет прогнозных и рекомендательных формулировок, напрямую обращенных к лицам, принимающим решения. Это отчетливо видно в лексико-семантических связях ядра *cancer* с лексемами *control, prevention, efforts*. Такие конструкции выполняют функцию метатекстовых операторов, приглашающих к совместному осмыслению проблемы и планированию мер. Терминологические узлы (например, *Human Development Index*) вводятся в полном виде, без сокращений, что снижает порог вхождения и делает текст доступным для специалистов смежных областей. Композиционная структура аннотации (описание данных, прогноз, рекомендации) формирует метаграф, который направляет когнитивную навигацию читателя от констатации фактов к постановке задач. Подтвердим фрагментом текста аннотации: «*The authors describe national cancer incidence and mortality by world region and the Human Development Index and predict the burden in 2050 based on demographic trends*» (Авторы описывают показатели заболеваемости и смертности от рака по странам и регионам мира, а также в разрезе индекса человеческого развития, и прогнозируют бремя заболевания к 2050 г. на основе демографических тенденций).

Примечательно, что обе модели разделяют стремление к информативности через включение цифровых данных, однако их функция различается: в русскоязычных аннотациях цифра обслуживает императив объективности, в англоязычных – приглашение к верификации.

Проведенный анализ свидетельствует, что различия между русско- и англоязычной медицинской аннотацией лежат не столько в синтаксической типологии (хотя преобладание сложных конструкций в английском языке и простых осложненных в русском фиксируется стабильно), сколько в когнитивно-дискурсивной ориентации. Русскоязычный медицинский дискурс тяготеет к институциональной замкнутости, где знание транслируется как верифицированный продукт. Англоязычный дискурс, напротив, моделирует ситуацию научного диалога, куда может интегрироваться более широкий круг участников.

Метод графосемантического моделирования, апробированный на данном материале, продемонстрировал высокую объяснительную силу. Он позволяет перейти от констатации разрозненных языковых фактов к построению целостных типологических портретов дискурсивных формаций. К числу перспектив исследования отнесем сопоставление моделей аннотаций с полными текстами статей, а также экстраполяцию методики на другие институциональные дискурсы (юридический, педагогический). Подобные межпарадигмальные проекты активно разрабатываются в современной отечественной лингвистике и соответствуют актуальным направлениям, представленным в работах Уфимской лингвистической школы [8].

Предложенная в работе методологическая рамка, синтезирующая положения структурной лингвистики текста и инструментарий когнитивного картирования, открывает новое измерение в сопоставительном изучении научного дискурса. Смещение фокуса с констатации языковых расхождений на моделирование стоящих за ними когнитивных сценариев адресанта научного медицинского дискурса позволяет не только глубже понять природу межкультурной вариативности текста, но и выработать рекомендации по оптимизации русскоязычной научной коммуникации в условиях ее глобализации.

## Литература

1. Halliday M.A.K., Hasan R. Cohesion in English. London: Longman, 1976. 356 p.
2. Чернявская В. Е. Лингвистика текста и дискурс-анализ. СПб: Наука, 2006. 192 с.
3. Касымова О. П. Язык в свете основных понятий системологии. Уфа: Башкирский государственный университет, 2002. 64 с.
4. Чернявская В. Е. Интерпретация научного текста в аспекте дискурсивных стратегий // Вестник СПбГУ. Серия 9. Филология. Востоковедение. Журналистика. 2010. Вып. 2. С. 139–146.

5. Белоусов К. И., Зелянская Н. Л. Графосемантическое моделирование как метод экспликации структуры концепта // Вестник Пермского университета. Российская и зарубежная филология. 2013. №4(24). С. 5–13.
6. Гальперин И. Р. Текст как объект лингвистического исследования. М.: Наука, 1981. 139 с.
7. Карасик В. И. Языковой круг: личность, концепты, дискурс. Волгоград: Перемена, 2002. 477 с.
8. Пешкова Н. П., Таюпова О. И., Газизов Р. А. и др. Актуальные проблемы современной лингвистики через призму полипарадигмальности / ред. А. С. Самигуллина, А. Л. Федорова. М.: ООО изд-во «ТРИУМФ», 2025. 228 с.

#### URL

1. Сахарный диабет в российской федерации: динамика эпидемиологических показателей по данным федерального регистра сахарного диабета за период 2010–2022 гг. URL: <https://cyberleninka.ru/article/n/saharnyy-diabet-v-rossiyskoy-federatsii-dinamika-epidemiologicheskikh-pokazateley-po-dannym-federalnogo-registra-saharnogo-diabeta/viewer>.
2. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. URL: <https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.70090>.

Статья рекомендована к печати  
кафедрой русского языка, лингвистики  
и международной коммуникации  
ФГБОУ ВО БГМУ Минздрава РФ.

---

## Comparative Analysis of Discursive Strategies in Scientific Medical Texts: The Methodology of Graphosemantic Modeling

**L. A. Linnik**

*Bashkir State Medical University  
3 Lenin st., 450008 Ufa, Republic of Bashkortostan, Russia.*

*Email: lyubovlinnik07@gmail.com*

The article focuses on developing a methodology for the comparative analysis of discursive strategies in Russian- and English-language medical abstracts based on the graphosemantic modeling method. Unlike the traditional structural-semantic approach, which empha-

sizes cohesion and coherence, the proposed model treats the text as a projection of the author's cognitive scenarios that regulate the degree of subject explication and the terminological density of the text. Two basic strategies have been identified: status-oriented (in Russian-language discourse) and dialogue-oriented (in English-language discourse). The graphosemantic modeling toolkit makes it possible to visualize these strategies in the form of graphosemantic models, which opens up prospects for the typological study of scientific discourse.

**Keywords:** medical discourse, abstract, graphosemantic modeling, cognitive-communicative strategy, comparative linguistics, scientific text, author's intention.