

## ПРОГНОСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПИСЬМЕННОГО НАУЧНОГО МЕДИЦИНСКОГО ТЕКСТА В ЦИФРОВОЙ СРЕДЕ

Л.В. Ягенич

<sup>1</sup>Крымский федеральный университет имени В. И. Вернадского,  
г. Симферополь

В статье рассматривается проблема прогностического моделирования с точки зрения функционирования научного медицинского текста в цифровой среде. Цель исследования заключается в выявлении и апробации критериев прогностического моделирования письменных научных текстов медицинского содержания. Выполнен историографический анализ научной литературы; статистический анализ, метод сопоставительного количественного анализа. Материалами исследования послужили научные статьи, монографии, авторефераты диссертаций на английском языке, которые были обработаны с помощью программного обеспечения Neural Designer, степень востребованности текстовых примеров была оценена с точки зрения выполнения прогностической функции для пользователя в цифровой среде. Выявлена прямая положительная корреляция между уровнем когнитивной сложности и прогностической функцией письменного научного медицинского текста.

**Ключевые слова:** прогностическая функция текста, цифровая среда, моделирование, медицинский научный текст, компьютерная лингвистика.

В эпоху цифровизации растет количество специалистов медицинского профиля, которые ищут научную информацию в Интернете. Всего, по оценкам Министерства здравоохранения Российской Федерации, 80% россиян именно там ищут информацию на медицинские темы [5]. По оценкам исследователей, 45% населения нашей страны используют Интернет для поиска профессиональной медицинской помощи, включая онлайн-консультации врачей и запись на прием к специалистам [7:63]. Цифровая среда в данном случае рассматривается исследователями как форма распространения и хранения медицинской информации [4: 88]. Данная информация может быть представлена в текстах различных жанров, в том числе и в жанре научного текста (статьи, монографии, диссертации, авторефераты диссертаций и т.п.).

Научный текст обладает определенными жанровыми и структурными характеристиками [14:17]: монологичность; наличие общенаучной и специальной терминологии; обобщенно-отвлеченный характер изложения; подчеркнутая логичность; строгая последовательность изложения; смысловая точность; информативная насыщенность.

Главной особенностью медицинского научного текста является использование общенаучных и специальных терминов. В связи с глобальным характером языка медицины (т.е. его ориентация на латынь и греческий, а в последние десятилетия и на английский) на русском языке можно найти точные аналоги или калькированные термины.

По мнению исследователей [10:17], медицинскую лингвистику можно определить как прикладную отрасль филологических наук, изучающую функционирование языка и использование языковых ресурсов в области медицины на основе исследований лингвистических аспектов профессиональной научной коммуникации в области медицины и здравоохранения [10:18]. Область интересов медицинской лингвистики базируется на междисциплинарном подходе.

Можно также выделить особое направление – компьютерная медицинская лингвистика [6:10]. Прогностическое моделирование медицинских текстов в цифровой среде можно определить как процесс создания текстов с учетом поисковых предпочтений, уровня образования и наиболее популярных тем о здоровье и медицинской помощи для пользователей в цифровой среде. Под цифровой средой в данном контексте следует понимать «новый тип среды взаимодействия между медицинскими специалистами, в которой функционируют научные медицинские тексты» [15:231].

Цель исследования заключается в формулировке и апробации критериев прогностического моделирования письменных научных медицинских текстов на английском языке, функционирующих в цифровой среде.

В настоящее время проблема прогностического моделирования для обслуживания научной коммуникации в медицинской отрасли в целом и письменной научной коммуникации между специалистами посредством письменных медицинских текстов получила развитие в целом ряде научных исследований. В частности, на современном этапе для такой коммуникации используются научные базы и цифровые платформы, на которых размещаются научные медицинские тексты для специалистов в различных областях медицины [11:591].

Медицинский письменный научный текст в Интернет-среде как самостоятельный вид научного текста рассматривается в работе Я. Тяжлова и А. Белоедовой, которые предлагают следующие определения основных понятий: «Научный текст — это сообщение, обладающее смысловой и структурной завершенностью и связанностью всех входящих в него самостоятельных предложений и информативных частей — абзацев, глав, разделов, параграфов. Научные тексты содержат научную информацию» [12:452].

Под медицинским письменным научным текстом в Интернет-среде М. Тома и О.С. Вэйи предлагают понимать «научный текст, освещающий определенную научную медицинскую проблематику и

размещенный на научных платформах и базах для письменной коммуникации специалистов» [12:595]. Приведенные определения мы считаем наиболее актуальными и принимаем за основу в нашем исследовании.

Исследование научных письменных медицинских текстов ведется также в направлении выявления их функционирования в Интернет-среде. Например, в работе Т.В. Дубровской и Е.А. Кожемякина исследуется влияние цифровых платформ на семиотическую организацию научного текста [3:163]. Авторамиданного исследования предложен «алгоритм мультимодального исследования научных интернет-текстов» [3:164]. В исследованиях, посвященных феномену цифровой среды, также активно обсуждаются методы оптимизации поиска научных медицинских текстов на цифровых платформах [4:87].

В последние два года появились также исследования, посвященные проблеме этики и ответственности (контент, генерируемый искусственным интеллектом) в научных медицинских текстах, функционирующих в цифровой среде [1:58]. Зарубежные авторы [10:8] сосредоточены на анализе различий между медицинскими текстами, написанными экспертами в различных отраслях медицины, а также на разработке рабочих процессов машинного обучения для эффективного обнаружения и дифференциации научных медицинских текстов [14:17]. Прикладные аспекты научного текста исследуются в работах отечественных авторов [8:64].

Некоторые сформулированные исследователями определения могут быть приняты как базовые для изучения специфики научных медицинских текстов. Например, в исследовании С. Янга подробно описана прогностическая функция научного медицинского текста в интернет-среде: «прогностическая функция письменного научного медицинского текста в интернет-среде представляет собой возможность проследить закономерности подачи актуальной информации на основании письменной коммуникации между специалистами (как клинической практики, так и в сфере теоретических исследований), позволяющие устанавливать направления будущих исследований по различным направлениям медицины» [15:229].

Тем не менее, несмотря на довольно обширную литературу, в современной науке недостаточно исследован потенциал прогностического моделирования научных медицинских текстов в цифровой среде.

В качестве материалов исследования были использованы научные медицинские тексты: научные статьи, монографии, авторефераты диссертаций на английском языке. Каждый образец был представлен в объеме 1500 печатных знаков. Научные тексты отобраны со специальных цифровых платформ, на которых размещаются научные статьи, монографии и авторефераты диссертаций: ресурс научной платформы

National Library Of Medicine (<https://www.ncbi.nlm.nih.gov/>); научные базы данных: <https://www.ncbi.nlm.nih.gov/pmc/>; <https://www.ncbi.nlm.nih.gov/books/>; <https://pubmed.ncbi.nlm.nih.gov> .

Всего отобрано 626 научных медицинских текстов, каждый образец состоит из 1 печатного листа текста, на английском языке. Образцы были собраны за 2019, 2020, 2021 и 2022 годы.

Методы исследования включали: статистический анализ количественных показателей тематического распределения текстов научно-медицинского содержания. На основе анализа научной литературы *критериями* для анализа научных медицинских текстов послужили аспекты, которые исследователи, например, С. Янг, выделяют в качестве основных характеристик цифрового текста [15: 230]:

- количество слов– фактор, определяющий то, как долго пользователь сможет читать такой текст в цифровой среде (научная статья, глава монографии или автореферат на английском языке) и насколько глубоко раскрыта тема;

- когнитивная доступность – фактор, связанный с уровнем сложности синтаксической структуры научного медицинского текста, а также насыщенностью специальной терминологией, которая определяется как количественное соотношение общенаучных и специальных медицинских терминов на страницу текста.

Для поиска текстов были использованы ключевые слова, связанные с различной научной медицинской проблематикой. Вся встреченная в научных медицинских текстах терминология имеет международный статус и может быть проверена каждым медицинским специалистом по первому словарю медицинских терминов для цифровой среды, включенному в Единую Систему Медицинского Языка (англ. Unified Medical Language System, UMLS– унифицированный язык медицинских систем) – цифровой инструмент для облегчения научной письменной коммуникации при чтении текстов научного медицинского содержания специалистами разных профилей.

Данная система позволяет повысить когнитивную доступность научных медицинских текстов при межотраслевой медицинской письменной коммуникации специалистов разных отраслей медицины. Разработка данной системы была начата еще в 1986 году в Национальной медицинской библиотеке США (National Library of Medicine, NLM) и в настоящее время включает всю мировую медицинскую терминологию на латинском, греческом и английском языках [13]. Мы использовали версию 2005-AB в качестве единого метатезауруса для выявления названий различных специальностей в медицине и включили все словари, которые не нуждаются в дополнительной лицензии. Набор выявленных в текстах терминов составил 1 230 321 единицу, которые были сопоставлены с 814503 понятиями. В настоящее время ресурс UMLS является самым популярным и надежным медицинским ресурсом,

который фактически может заменить более сложные и длительные изыскания в рамках корпусных исследований.

В качестве примеров собранных образцов можно привести следующие фрагменты научных медицинских текстов (таблица 1):

Таблица 1.

№ п/п	Тип письменного научного медицинского текста	
	Научные статьи	Выходные данные и ресурс размещения источника
1	«Acute gastrointestinal dysfunction is a common and important complication of sepsis. as no exiting formal definition and classification of gastrointestinal dysfunction, most of the treatment strategies for gastrointestinal dysfunction are not based on clinical evidence, but on their own clinical experience»	Tyszko M., Lemańska-Perek A., Śmiechowicz J., Tomaszewska P., Biecek P., Gozdzik W., Adamik B. Citrulline, Intestinal Fatty Acid-Binding Protein and the Acute Gastrointestinal Injury Score as Predictors of Gastrointestinal Failure in Patients with Sepsis and Septic Shock. <i>Nutrients</i> . 2023 Apr 27;15(9):2100. Doi: 10.3390/nu15092100. PMID: 37432225; PMCID: PMC10180779 Научная платформа National Library Of Medicine <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10180779/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10180779/</a>
3	«This study contributes valuable findings in the usage of CAM among refugees and determines that CAM is commonly used among refugees resettled in the USA for pain management. Originality/value By better understanding CAM, physicians can address a direct need for the refugee population-seeking health care in the USA and other countries that host refugee resettlement».	Rathi, N., Mu Pye. The use of complementary and alternative medicine and pain relievers among refugees resettled in the USA. <i>International Journal of Migration, Health and Social Care</i> . 2021. Vol. 3. DOI:10.1108/IJMHS-10-2020-0092 Научная платформа National Library Of Medicine: <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5870092/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5870092/</a>

4	<p><i>«Patients with rheumatologic conditions, such as eosinophilic granulomatosis with polyangiitis (EGPA), rheumatoid arthritis (RA), and systemic lupus erythematosus (SLE), often face being immunocompromised due to their disease state, or the immunosuppressive effect of their treatments. Managing immunocompromised rheumatologic patients can be challenging and complex».</i></p>	<p>Winthrop, K., Calabrese, L. Holistic Care of the Immunocompromised Rheumatologic Patient – Case Analysis. <i>Rheumatology and Therapy</i>, 2023, 10(6). DOI:10.1007/s40744-023-00549-x          Научная платформа National Library Of Medicine: <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7185396/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7185396/</a></p>
5	<p><i>«Calcium, magnesium, potassium, and sodium oxybates (low-sodium oxybate; LXB) is an oxybate formulation with 92% less sodium than sodium oxybate (SXB; a treatment for narcolepsy) and the same active moiety. LXB is approved in the US for treatment of cataplexy or excessive daytime sleepiness (EDS) in patients 7 years of age or older with narcolepsy, and idiopathic hypersomnia in adults. In Phase 3 clinical trials, LXB exhibited a safety profile consistent with that of SXB in narcolepsy. Besides continued efficacy in treating symptoms, potential benefits of long-term LXB treatment include flexible optimization of dosing and regimen, improvement of QoL».</i></p>	<p>Lee-Iannotti, K., Bogan, R.K. Long-Term Treatment of Narcolepsy and Idiopathic Hypersomnia with Low-Sodium Oxybate. 2023; 15: 663–675. Published online 2023 Aug 19. DOI: 10.2147/NSS.S412793PMCID: PMC10445641          Научная платформа National Library Of Medicine: <a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10445641/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10445641/</a></p>

Все собранные письменные медицинские научные тексты репрезентативной выборки были проанализированы в программе расширенной статистической аналитики NuralDesigner по трем выделенным критериям анализа: объем текста, когнитивная доступность, возможность комментирования. Программа все данные приводит по 10-ти балльной шкале: 1-3 – низкий уровень критерия; 4-6 – средний уровень критерия, 7-10 – высокий уровень критерия. Данная шкала позволяет оперативно и наглядно представить данные по масштабным текстовым выборкам без потери точности. Программный анализ данных по научным медицинским текстам осуществлялся в период с 1 сентября по 20 декабря 2023 года.

Полученные данные были обработаны программой и представлены в виду итоговой диаграммы (рисунок 1). Прогностическое моделирование было применено к письменным научным медицинским текстам по приведенной в диаграмме тематике, которые были выявлены в цифровой среде: ресурс научной платформы National Library Of Medicine (<https://www.ncbi.nlm.nih.gov>; научные базы данных: <https://www.ncbi.nlm.nih.gov/pmc>; <https://www.ncbi.nlm.nih.gov/books/>; <https://pubmed.ncbi.nlm.nih.gov>, согласно трем критериям анализа.

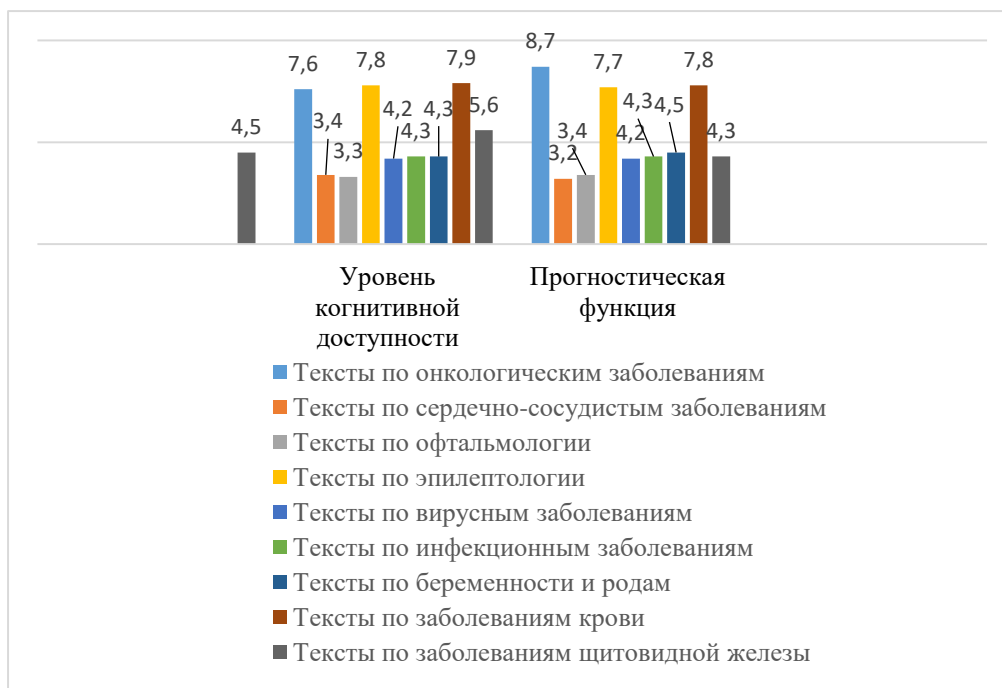


Рисунок 1. Результаты оценки тематических текстов научно-медицинского содержания в цифровой среде по критерию уровня когнитивной доступности и прогностической функции (составлено автором с помощью программы NeuralDesigner)

Было также проведено сопоставительное исследование на базе словаря UMLS[13] для уточнения количественного распределения медицинской терминологии (греческий, латинский и английский языки) и для выявления критерия «когнитивная доступность» на основе соотношения терминов в тех же научных текстах (рисунок 2).

Наибольшую когнитивную сложность составляют тексты, связанные с онкологическими заболеваниями, по эпилептологии, а также по гематологии (науке о заболеваниях крови).

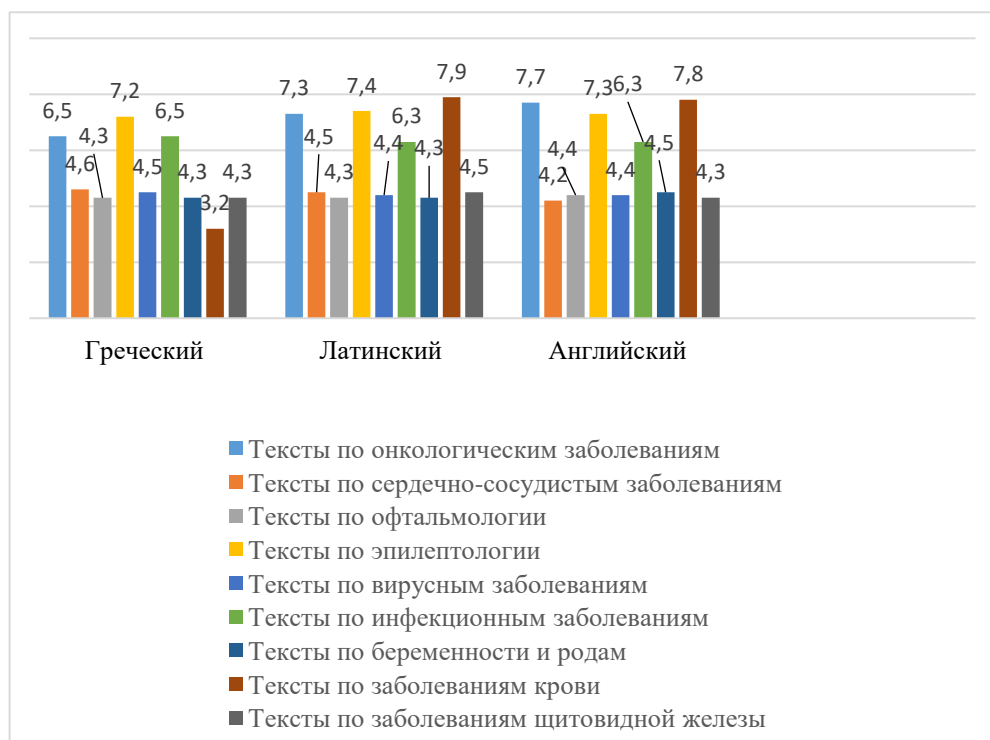


Рисунок 2. Количественное распределение медицинской терминологии (греческий, латинский и английский языки) по выявленной тематике научных медицинских текстов в цифровой среде

Проведенное исследование позволяет сформулировать следующие выводы:

1) В современной цифровой среде научный медицинский текст выполняет не только информативную, но и прогностическую функцию, что позволяет ему стать объектом прогностического моделирования в рамках относительно нового научного направления – компьютерной медицинской лингвистики.

2) Внешние (атрибутивные) и внутренние (контентные) характеристики научных медицинских текстов, которые циркулируют в цифровой среде, могут быть проанализированы на основании трех суммарных критериев: объем печатных знаков (1500 печатных знаков); когнитивная доступность – уровень сложности научного медицинского текста. Прогностическая функция, как показывает приведенный выше в диаграммах анализ, возрастает прямо пропорционально уровню когнитивной доступности. В ходе исследования была выявлена прямая положительная корреляция между уровнем когнитивной сложности и прогностической функцией научного медицинского текста. Чем выше сложность научного медицинского текста для понимания медицинскими специалистами, тем шире прогностическая функция такого текста, поскольку он фактически превращается в гипертекст в цифровой среде.



## Список литературы

1. Веселовская Т.С. Особенности исследования цифровых учебных текстов // Ученые записки Петрозаводского государственного университета. 2022. № 2.С. 56–62.
2. Гайсина С.В. Особенности учебного текста в условиях цифровой образовательной среды // Проблемы современного педагогического образования. 2021. № 73. С. 106–110.
3. Дубровская Т., Кожемякин Е. Научный интернет-текст как семиотический объект: постановка проблемы для мультимодального анализ // Динамика медиасистем. 2023. № 1.С. 163–171.
4. Епифанова А.Г., Долдо Н.В. Функционал и дискурсивность дизайна виртуальной среды в формате культурологической интерпретации // Человек. Культура. Образование. 2021. № 2.С. 87–95. DOI: 10.34130/2233-1277-2021-2-87
5. Министерство здравоохранения Российской Федерации. Перечень медицинских организаций, подведомственных Минздраву России, в отношении которых проводится независимая оценка качества оказания услуг. URL: <https://minzdrav.gov.ru/opendata/7707778246-medorgocenkakachestva/visual> (дата обращения: 12.12. 2023).
6. Пищальникова В.А. Определение цифрового текста как методологическая проблема // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2022. № 6.С. 9–14. DOI 10.52070/2542-2197\_2022\_6\_861\_9
7. Торубарова И.И. Академичность русскоязычных научных текстов медицинского корпуса: способы реализации и особенности проявления // Языкознание. Вестник Волгоградского государственного университета. Серия 2. Языкознание. 2020. № 5.С. 61–73. DOI: <https://Doi.org/10.15688/jvolsu2.2020.5.6>
8. Шарифуллина А.Д. Компьютерная лингвистика как направление в прикладной лингвистике // Тенденции развития науки и образования. 2021. № 4.С. 64-66. DOI:10.18411/lj-06-2021-138
9. Яворская О.А., Огородникова Э.Ю. Цифровые технологии обучения переводу для специальных целей в медицинском вузе // Современные проблемы науки и образования. 2020. № 2.С. 10–14. DOI: 10.17513/spno.29638
10. Hutchinson N., Grayson, L., Megha G. Examining the Reading Level of Internet Medical Information for Common Internal Medicine Diagnoses. The American Journal of Medicine. 2016. Vol. 24. Pp. 7-10. DOI: <https://Doi.org/10.1016/j.amjmed.2016.01.008>
11. Toma M., Wei O.C. Predictive Modeling in Medicine. Encyclopedia. 2023. Vol. 3. Pp. 590–601.
12. Tyazhlov Y., Beloedova A. Scientific text on the Internet: communication and typological characteristics // Media Linguistics. 2023. Vol. 10(4). Pp. 450–477. DOI:10.21638/spbu22.2023.403
13. Unified Medical Language System (UMLS). [Electronic resource].URL: <https://www.nlm.nih.gov/research/umls/index.html> (датаобращения: 12.12.2023).

14. Wenxiong L., Zhengliang L. Differentiate ChatGPT-generated and Human-written Medical Texts. *Computer Science: Computation and Language*. 2023. Vol. 4. Pp. 15–30. DOI: arXiv:2304.11567v1
15. Yang C. Explainable Artificial Intelligence for Predictive Modeling in Healthcare. *Journal of Healthcare Informatics Research*. 2022. Vol. 6(2). Pp. 228–239. DOI: 10.1007/s41666-022-00114-1. PMID: 35194568

*Об авторе:*

ЯГЕНИЧ Лариса Викторовна, кандидат педагогических наук, доцент, заведующий кафедрой иностранных языков №4 Института филологии, Крымский федеральный университет имени В. И. Вернадского, Симферополь (295007, Республика Крым, г. Симферополь, просп. Академика Вернадского, д. 4), e-mail: yagenich@mail.ru

## **PREDICTIVE MODELING OF WRITTEN SCIENTIFIC MEDICAL TEXT IN A DIGITAL ENVIRONMENT**

**L.V. Yagenich**

V. I. Vernadsky Crimean Federal University, Simferopol

The article examines the problem of predictive modeling from the point of view of the functioning of scientific medical text in the digital environment. The research goal is to identify and test criteria for predictive modeling of written medical scientific texts. The research methodology is based on a historiography analysis; statistical analysis and method of comparative quantitative analysis. The following results were obtained: a direct positive correlation was identified between the level of cognitive complexity and the prognostic function of scientific medical text; It was found that the criterion of cognitive accessibility makes it possible to implement the predictive function of scientific medical text in the digital environment.

**Keywords:** *philology, predictive function of text, digital environment, modeling, medical scientific text, computer linguistics.*

*About the author:*

YAGENICH Larissa Viktorovna – PhD in Methodology, Ass. Professor, Head of Foreign Language Department № 4, Institute of Philology, V.I. Vernadsky Crimean Federal University, Simferopol (Prospekt Vernadskogo 4, Simferopol, Republic of Crimea, 295007), e-mail: yagenich@mail.ru

Статья поступила в редакцию 09.09.24  
Подписана в печать 25.09.24

© Ягенич Л.В. 2024