

ЛЕКСИЧЕСКИЕ ИЗМЕНЕНИЯ ПОД ВЛИЯНИЕМ ИИ: ОТ КОРПУСА К КОММУНИКАЦИИ

AI-DRIVEN LEXICAL CHANGE: FROM CORPORA TO COMMUNICATION

**E. Belozerova
A. Dekanov**

Summary: The article analyzes the impact of generative artificial intelligence models on the lexical system of the English language. Particular attention is paid to the formation of an AI-induced lexicon, which includes archaisms, stylistically marked vocabulary, and specialized terminology. The study examines such phenomena as cognitive metaphorization, passive lexical archiving, and lexical standardization. It provides an overview of domestic and international publications (2017–2024) addressing lexical transformations, cognitive metaphor theory, language standardization, and the influence of algorithmic models on speech practices. The work emphasizes the need for an interdisciplinary approach to the study of linguistic shifts in the age of algorithmic communication.

Keywords: artificial intelligence, language model, lexicology, cognitive metaphor, standardization.

Белозёрова Екатерина Юрьевна

кандидат филологических наук, доцент, Финансовый университет при Правительстве Российской Федерации, (г. Москва)

eybelozyorova@fa.ru

Деканов Андрей Евгеньевич

Финансовый университет при Правительстве Российской Федерации, (г. Москва)

deckandr@yandex.ru

Аннотация: Статья посвящена анализу влияния генеративных моделей искусственного интеллекта на лексическую систему английского языка. Особое внимание уделяется формированию ИИ-индуцированного лексикона, включающего архаизмы, стилистически маркированную лексику и специализированные термины. Рассматриваются такие феномены, как когнитивная метафоризация, «пассивная архивация» лексем и лексическая стандартизация. Проводится обзор публикаций отечественных и зарубежных исследователей (2017–2024 гг.), посвящённых лексическим трансформациям, когнитивной метафоре, стандартизации языка и влиянию алгоритмических моделей на речевые практики. Работа подчёркивает необходимость междисциплинарного подхода к изучению языковых сдвигов в эпоху алгоритмизированной коммуникации.

Ключевые слова: искусственный интеллект, языковая модель, лексикология, когнитивная метафора, стандартизация.

Активное внедрение генеративных моделей искусственного интеллекта (ИИ) оказывает системное воздействие на лексическую систему английского языка, инициируя качественные сдвиги в способах порождения, распространения и употребления лексических единиц. Особое внимание исследователей привлекает формирование ИИ-индуцированного лексикона, включающего архаизмы, специализированные термины, стилистически маркированную лексику, а также когнитивные и операциональные метафоры. Эти процессы сопровождаются тенденцией к лексической стандартизации и отражают переход к новой языковой норме, формируемой вне рамок традиционных лингвистических институтов.

Актуальность исследования обусловлена тем, что ИИ всё чаще становится активным участником языкового процесса, влияя на лексико-семантическую структуру текста в профессиональных, академических и массовых регистрах. Объект исследования — лексическая трансформация английского языка в условиях цифровизации и алгоритмизированной коммуникации. Предмет исследования — механизмы влияния ИИ на лексику, включая архаизацию, неологизацию, когнитивную метафоризацию и стандартизацию. Цель исследования — выявить

и систематизировать ключевые направления лексических изменений под воздействием генеративного ИИ на основе анализа актуальных научных публикаций. Материалом исследования выступают работы отечественных и зарубежных исследователей за 2017–2024 гг., посвящённые цифровым языковым трансформациям. Методология основана на обзорно-аналитическом подходе с междисциплинарной опорой на прикладную лингвистику, когнитивную семантику и философию языка. Научная новизна заключается в комплексной оценке лексических изменений, вызванных ИИ, с учётом взаимодействия лингвистических, когнитивных и технологических факторов.

Расширение словарного запаса под влиянием ИИ

Одним из значимых следствий активного внедрения искусственного интеллекта (ИИ) в сферу письменной коммуникации и автоматизированной генерации текста является трансформация лексической системы английского языка. На современном этапе это влияние проявляется в интенсивной актуализации ранее малоупотребляемых единиц, генерации новых терминов, а также в репрофилировании существующих лексем посредством операциональных метафор. В совокупности данные

процессы свидетельствуют о становлении специфического ИИ-индуцированного лексикона, который оказывает устойчивое воздействие на лексико-семантический уровень языка.

Современные языковые модели, такие как GPT-3 и GPT-4, обучаются на многоуровневых корпусах, включающих тексты различных жанров, тематик и исторических периодов. Как указывают Т. Brown и соавторы (2020), «обучающие данные для GPT-3 включают совокупность текстов из интернет-источников, включая книги и энциклопедии, представляющих широкий спектр языкового употребления, охватывающий различные тематические области и временные отрезки» [1]. Аналогичным образом в техническом отчёте J. Achiam и соавторы (2023) подчёркивается, что «GPT-4 была обучена на наборе данных, включающем широкий спектр общедоступных и лицензированных текстов на различных языках, охватывающих разнообразные жанры и тематические области» [2].

Такое разнообразие обучающего материала обуславливает появление в генерируемых моделями текстах лексических единиц, нехарактерных для современного повседневного дискурса. В частности, наблюдается тенденция к включению архаизмов, книжной лексики и стилистически маркированных слов, ранее распространённых преимущественно в академической, юридической и художественной литературе.

Как отмечает Е. Bender (2018), современные языковые модели обучаются на корпусах, которые зачастую не являются репрезентативными с точки зрения жанрового и временного охвата. Это приводит к тому, что в генерируемых текстах нередко присутствуют лексемы, утратившие активное употребление в живом языке. По мнению исследовательницы, подобное явление можно охарактеризовать как «пассивную архивацию», при которой устаревшие слова возвращаются в современный дискурс за счёт их статистической представленности в обучающих данных [3].

D. Crystal (2019) также указывает, что статистические методы, применяемые в ИИ, способствуют возрастанию доли формально маркированной лексики, типичной для академического английского конца XIX — начала XX века. По его мнению, генеративные модели в силу своей обучающей базы склонны воспроизводить языковые паттерны, не отражающие текущую языковую норму [4].

Н.Ю. Хомякова (2021) подчёркивает, что в англоязычном научном дискурсе, созданном при участии ИИ, часто фиксируются слова, вышедшие из активного употребления, но сохраняющиеся в корпусах, используемых для обучения моделей. Она трактует этот феномен как институциональную архивацию, воспроизводимую через цифровые алгоритмы [5].

Схожие выводы делает Т.В. Скларова (2022), анализируя тексты, сгенерированные GPT-3. В своём исследовании она показывает, что частотность архаизмов (таких как *therein*, *thusly*, *forthwith*) значительно выше в ИИ-генерируемом контенте по сравнению с текстами, написанными носителями языка. Этот феномен она объясняет включением в обучающие корпуса значительно объёма классических и юридических источников [6].

Дополнительное подтверждение данной тенденции приводит В. Clarke (2024), фиксируя значительный рост частотности использования слова *delve* в англоязычных научных аннотациях после 2023 года. Исследователь связывает это с широким распространением генеративных ИИ-инструментов в академическом письме, что, по его мнению, свидетельствует о растущем влиянии языковых моделей на формирование новых стилистических норм [7].

Таким образом, совокупность современных исследований позволяет заключить, что использование многоуровневых корпусных данных при обучении языковых моделей, таких как GPT-3 и GPT-4, приводит к систематическому воспроизведению лексических единиц, нехарактерных для повседневного употребления. К числу таких единиц относятся архаизмы, книжные слова и стилистически маркированная лексика, унаследованная из научного, юридического и классического дискурсов.

Термин «пассивная архивация», используемый в настоящем исследовании как лексико-метафорическая характеристика переноса устаревших лексем в тексты, сгенерированные ИИ, не закреплён в лингвистике как официальное понятие, однако концептуально обоснован в ряде научных публикаций (Bender, Friedman, 2018; Floridi, 2019; Cabitza, 2017), посвящённых эпистемологическим и лексическим последствиям алгоритмического моделирования речи [3, 8, 9]. Этот лингвистический сдвиг, хотя и имеет скрытый характер, влияет на формирование новых норм научного и делового письма, способствует архивации дискурсивных паттернов и снижает стилистическое разнообразие речи. Системная актуализация архаизмов, таким образом, выступает не как случайный побочный эффект, а как структурно запрограммированное следствие алгоритмической природы генеративных языковых моделей.

Специализированная терминология

Показательным примером функционирования таких лексем является исследование L. Barkley и В. Merwe (2024), в котором анализируется взаимосвязь трёх ключевых понятий: *prompt engineering*, *transformer architecture* и *hallucination*. Авторы демонстрируют, как данные термины используются не только в техническом, но и в метаязыковом контексте — как средства описа-

ния, нормализации и регулирования языкового поведения моделей искусственного интеллекта. Это делает их особенно значимыми для лингвистического анализа, поскольку они выполняют не только номинативную, но и концептуально-объяснительную функцию в формирующемся лексиконе ИИ [10].

Так, лексемы *prompt engineering*, *transformer architecture* и *hallucination* демонстрируют типичную для ИИ-дискурса модель терминологической эволюции: они иллюстрируют переход ИИ-терминов из технического регистра в общее употребление, сохраняя семантическую сложность и выполняя метаязыковую функцию. Они не только описывают процессы, но и формируют нормы взаимодействия с ИИ, становясь элементами когнитивного инструментария цифровой эпохи.

Когнитивная метофоризация

ИИ-парадигма актуализирует процесс когнитивной метофоризации, в рамках которого термины, заимствованные из таких областей, как нейробиология (*neural network*), педагогика (*training, learning*) и когнитивные науки (*attention, memory*), подвергаются трансформации и приобретают специализированное техническое значение. Это явление свидетельствует не только о метафорическом обогащении языка, но и о смещении понятийных границ между человеком и машиной.

Классическая теория метафоры, представленная в работе G. Lakoff и M. Johnson *Metaphors We Live By* (1980), демонстрирует, как операционные метафоры (например, «ум — это компьютер», «обучение — это наполнение») структурируют восприятие абстрактных концептов и способствуют формированию новых категориальных рамок [11]. Эти положения приобретают особую актуальность в условиях цифровой коммуникации, где технические процессы всё чаще описываются через призму человеческого опыта.

Е. Cabitza и соавторы (2017) в междисциплинарном исследовании подчеркивают, что использование метафор *doctor-like AI* и *learning machines* не только способствует лексической экспансии, но также искажает границы ответственности и понимания технологий, особенно в чувствительных сферах — таких как медицина и образование [9]. Это создаёт риск переинтерпретации функций ИИ, приписывания ему свойств субъективности и автономности.

М. Mitchell (2019) подробно анализирует термины *neural network, training, attention, memory*, указывая на их метафорическую природу [12]. Она подчёркивает, что данные термины не следует интерпретировать буквально: несмотря на «человеко-подобный» лексический облик, они отражают исключительно вычислительные процессы.

L. Floridi (2019) вводит понятие информационно-онтологической редукции, объясняя, как чрезмерно «очеловеченная» терминология приводит к эпистемологическим искажениям и препятствует адекватному пониманию ИИ как несубъектного инструмента [8]. По его мнению, необходима деметафоризация языка описания технологий.

Схожую позицию выражают Е. Bender и А. Koller (2020), вводящие термин «обманчивая метофоризация» (*misleading anthropomorphism*) — ситуацию, в которой термины *understanding, knowledge, intelligence* применяются к алгоритмам, тем самым вводя в заблуждение как специалистов, так и широкую публику относительно границ возможностей современных языковых моделей [13].

В совокупности вышеприведённые исследования демонстрируют, что антропоморфные и образовательные метафоры упрощают понимание технологий, но одновременно формируют языковые нормы и концепты, требуя критического осмысления в цифровой лингвистике и философии языка.

Фактор лексической стандартизации

Влияние генеративных языковых инструментов на лексику пользователей приобретает устойчивый и системный характер. Такие платформы, как ChatGPT, Jasper, Writesonic, GrammarlyGO и другие, не ограничиваются функцией текстовой поддержки — они фактически становятся лексическими модераторами, формирующими устойчивые речевые паттерны и влияющими на стилистические предпочтения авторов. Как отмечает В. Clarke (2024), регулярное использование ИИ-программ ведёт к репликации ограниченного круга статистически предпочтительных лексем и синтаксических конструкций, что проявляется, в частности, в широком распространении выражений типа *delve into, foster innovation, it is important to note* [7]. Исследования Е. Mollick (2023) и J. Vincent (2023) подтверждают, что пользователи начинают бессознательно перенимать речевые модели ИИ, что приводит к снижению лексического разнообразия и стилистической оригинальности [14].

Публикация Е. Bender и В. Friedman (2018) подчёркивает, что подобная стандартизация речи является не только следствием автоматизированной генерации, но и результатом интернализации языковых стратегий, предложенных алгоритмом [3]. Пользователь, взаимодействующий с ИИ, всё чаще воспринимает предложенную лексику как «языковую норму», что влияет на последующее самостоятельное письмо.

Заключение

Интеграция генеративных моделей искусственно-

го интеллекта в практики письменной коммуникации инициирует масштабные трансформации лексической системы английского языка. Одним из главных проявлений этого воздействия становится формирование ИИ-индуцированного лексикона, включающего архаизмы, стилистически маркированную лексику и специализированные термины, ранее характерные преимущественно для академического, юридического и художественного дискурсов. ИИ-лексикон, в котором ключевую роль играют высокоплотные, функционально мобильные лексемы и операциональные метафоры, не только упрощает понимание технологических процессов, но и формирует нор-

мативные и концептуальные рамки цифрового дискурса.

Таким образом, искусственный интеллект выступает в роли нового лексикографического агента, способного переопределять границы употребления и значимости слов в различных регистрах — от научного до быденного. Осмысление этих процессов требует междисциплинарного подхода, объединяющего прикладную лингвистику, философию языка, когнитивную семантику и социоинформатику, и открывает перспективы для формирования новой парадигмы исследований в условиях алгоритмизированной коммуникации.

ЛИТЕРАТУРА

1. Brown T., Mann B., Ryder N. и др. Language Models are Few-Shot Learners [Электронный ресурс]. 2020. URL: https://www.researchgate.net/publication/341724146_Language_Models_are_Few-Shot_Learners (дата обращения: 25.03.2025). DOI: 10.48550/arXiv.2005.14165.
2. Achiam J., Adler S., Agarwal S. и др. GPT-4 Technical Report [Электронный ресурс]. OpenAI, 2023. URL: https://www.researchgate.net/publication/383739523_GPT-4_Technical_Report (дата обращения: 20.02.2025). DOI: 10.48550/arXiv.2303.08774.
3. Bender E.M., Friedman B. Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science // Transactions of the Association for Computational Linguistics. 2018. Vol. 6. P. 587–604. DOI: 10.1162/tacl_a_00041.
4. Crystal D. The Future of Englishes // Cambridge Encyclopedia of the English Language. 3rd ed. Cambridge, 2019. P. 582.
5. Хомякова Н.Ю. Формирование профессиональной лексики в англоязычном научном дискурсе под влиянием ИИ // Известия РАН. Серия литературы и языка. 2021. Т. 80, № 6. С. 89–96.
6. Слярова Т.В. Архаизация в цифровом английском: типология и прагматика // Вопросы прикладной лингвистики. 2022. №3(43). С. 49–57.
7. Clarke B.J. Evolving English: Globalisation and AI [Электронный ресурс]. 2024. URL: <https://www.benjclarke.me/p/evolving-english-globalisation-and-ai> (дата обращения: 25.01.2025).
8. Floridi L. The Logic of Information: A Theory of Philosophy as Conceptual Design. Oxford: Oxford University Press, 2019. 312 p.
9. Cabitza F., Rasoini R., Gensini G.F. Unintended Consequences of Machine Learning in Medicine [Электронный ресурс] // Journal of Medical Internet Research. 2017. Vol. 19, № 11. P. 354. URL: <https://www.jmir.org/2017/11/e354/> (дата обращения: 10.04.2025).
10. Barkley L., Merwe B. Investigating the Role of Prompting and External Tools in Hallucination Rates of Large Language Models [Электронный ресурс]. 2024. URL: https://www.researchgate.net/publication/385291821_Investigating_the_Role_of_Prompting_and_External_Tools_in_Hallucination_Rates_of_Large_Language_Models (дата обращения: 18.04.2025). DOI: 10.48550/arXiv.2410.19385.
11. Lakoff G., Johnson M. Metaphors We Live By. Chicago: University of Chicago Press, 1980. 242 p.
12. Mitchell M. Artificial Intelligence: A Guide for Thinking Humans. Farrar, Straus, and Giroux, 2019. 336 p.
13. Bender E.M., Koller A. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data [Электронный ресурс] // ACL. 2020. URL: https://www.researchgate.net/publication/343302041_Climbing_towards_NLU_On_Meaning_Form_and_Understanding_in_the_Age_of_Data (дата обращения: 20.04.2025).
14. Mollick E., Mollick L. How to Use AI to Do Stuff: An Opinionated Guide [Электронный ресурс] URL: <https://www.oneusefulthing.org/p/how-to-use-ai-to-do-stuff-an-opinionated> (дата обращения: 04.04.2025).

© Белозёрова Екатерина Юрьевна (eybelozyorova@fa.ru), Деканов Андрей Евгеньевич (deckandr@yandex.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»