

УДК 81'38

doi 10.17072/2073-6681-2025-4-81-90

<https://elibrary.ru/lqzslv>

EDN LQZSLV



Функциональная стилистика научного текста М. Н. Кожиной и современные корпусные исследования по идентификации искусственно сгенерированного контента

Рябцева Надежда Константиновна

д. филол. н., ведущий научный сотрудник отдела прикладной лингвистики

Институт языкоznания Российской академии наук

125009, Россия, г. Москва, Большой Кисловский пер., д. 1, стр. 1. nadia_riabceva@mail.ru

SPIN-код: 8966-9640

ORCID: <https://orcid.org/0000-0002-2042-4615>

ResearcherID: S-7138-2016

Scopus Author ID: 55218430500

Статья поступила в редакцию 26.09.2025

Одобрена после рецензирования 29.09.2025

Принята к публикации 01.10.2025

Информация для цитирования

Рябцева Н. К. Функциональная стилистика научного текста М. Н. Кожиной и современные корпусные исследования по идентификации искусственно сгенерированного контента // Вестник Пермского университета. Российская и зарубежная филология. 2025. Т. 17, вып. 4. С. 81–90. doi 10.17072/2073-6681-2025-4-81-90. EDN LQZSLV

Аннотация. В статье подчеркивается значимость научной деятельности выдающегося отечественного ученого, профессора Пермского государственного университета М. Н. Кожиной и ее особый вклад в лингвистику в целом и в современную стилистику в особенности. Показана взаимосвязь современных компьютерных и корпусных статистических исследований с направлениями работы пермской стилистической школы, основанной М. Н. Кожиной. На материале последних развернутых статистических исследований в области идентификации искусственно сгенерированных текстов продемонстрирована преемственность и эффективность квантитативного анализа языка, речи и коммуникации, а также его значимость для развития науки в целом. Так, данные последних развернутых исследований в этой области свидетельствуют об устойчивом росте использования общедоступного инструмента класса «Искусственный интеллект» ChatGPT-4 в научных публикациях почти сразу после его выпуска (30 ноября 2022 г.). Причем самый большой и быстрый рост искусственно сгенерированного контента отмечается в публикациях по компьютерным наукам – до 17,5 %. Основным принципом установления данного факта являлось системное масштабное статистическое сравнение свыше 950 900 научных статей, опубликованных на английском языке с января 2020 г. по февраль 2024 г. в ведущих мировых научных журналах по различным академическим дисциплинам – написанных до выпуска ChatGPT-4 и после него. При этом частотность слов в аннотациях по компьютерным наукам за последние 14 лет (2010–2024 гг.) оказалась непропорционально большой у четырех слов – *realm, intricate, showcasing, pivotal*. Причем именно после поступления ChatGPT-4 в свободный доступ они показали внезапный всплеск употребления. Таким образом, использование ChatGPT-4, особенно в области научной коммуникации, создает принципиально новые проблемы, в том числе лингвистические и стилистические, по идентификации искусственно сгенерированного контента.

Ключевые слова: функциональная стилистика; научный текст; корпусные исследования; частотность; ChatGPT-4; искусственно сгенерированный контент.

М. Н. Кожина и современная лингвистика

Маргарита Николаевна Кожина как выдающийся ученый и заслуженный деятель науки РФ, как создатель пермской школы функциональной стилистики, известной в России и за рубежом, а также как выдающийся специалист в области изучения родного языка и его истории, социолингвистики, риторики и прагматики, внесла своими трудами неоценимый вклад в развитие лингвистики в целом. Этот выдающийся вклад был по достоинству оценен многочисленными коллегами, соратниками и учениками М. Н. Кожиной, которые в своих публикациях с большой благодарностью и уважением выражают ей признательность как своему учителю, наставнику и соратнику. Так, опубликованный в 1966 г. фундаментальный и новаторский труд М. Н. Кожиной «О специфике художественной и научной речи в аспекте функциональной стилистики» [Кожина 1966], по сути, предвосхитил дальнейшее развитие лингвистической науки в изучении речи, дискурса и коммуникации, особенно научной. В книге, в частности, особо отмечается, что познавательная оценка, выраженная в научном тексте, – это его когнитивный стержень, объединяющий все варианты компонентов знания и подчиняющий их взаимодействие цели научного текста – фиксировать получение нового научного знания. Эти и многие другие положения данного исследования, а также всех последующих нашли позитивный отклик в большом количестве публикаций, рецензий, отзывов и комментариев как отечественных, так и зарубежных лингвистов (см., например: [Гайда 2010; Данилевская 2005] и мн. др.).

В фундаментальном труде «Очерки истории научного стиля русского литературного языка XVIII–XX вв.» в своих статьях М. Н. Кожина [Кожина 1996, 1998] указывает на особенности стилевой дифференциации разных литературных языков с учетом своеобразия их истории, а также на перспективы развития и важность сопоставительной стилистики, учета стилевой дифференциации разных литературных языков, взаимодействия функциональных стилей, стилистической организации целого речевого произведения и мн. др., что стало объектом особого содержательного рассмотрения в статье [Салимовский 2013: 20–22]. В ходе глубокого и тонкого лингвистического анализа классических произведений русской литературы М. Н. Кожина «предложила новое понимание конструктивного принципа речевой организации художественного текста, назвав эту стилевую черту художественно-образной речевой конкретизацией», что также отмечается в статье [там же: 14], в которой подчеркивается, что особый вклад М. Н. Кожиной в

стилистику текста заключается также в раскрытии междисциплинарности стилистики и в ее функционально-коммуникативной интерпретации текста [там же: 10], в характеризации текста как явления культуры, в выделении стилистически значимых средств коммуникации, в идентификации «функциональных семантико-стилистических категорий» (ФССК) текста, в определении коммуникативной организации научного текста и мн. др. (см.: [Кожина 1996; 1998]).

М. Н. Кожиной принадлежат также такие программные для пермской школы стилистики исследовательские положения, как «разработка методики применения статистики для изучения воздействия на характер речи того или иного конкретного из ряда действующих экстралингвистических факторов – так называемый метод срезов, на основе использования которого можно решать вопросы стилевой дифференциации речи на уровне более частных факторов и различные задачи социальной и коммуникативной лингвистики» [Штайн 2004: 34].

В целом «функционально-стилистическая теория текста М. Н. Кожиной знаменовала качественный скачок отечественного языкоznания в понимании коммуникативной природы текста и закономерностей его лингвостилистической организации» [Баженова 2013: 80]. М. Н. Кожина также особо подчеркивала, что развернутые вариативные повторы в научном тексте поддерживают его динамику, задают его организацию и воплощают введение нового научного знания: «в создании стиля участвует фактор частоты употребления языковых средств» (см.: [Кожина 2003; 2004; 2020]). Так, «деятельностная концепция,ложенная в основание теории интерпретации научного текста» позволила М. Н. Кожиной, ее ученикам и соратникам-коллегам «прийти к выводу об огромной организующей роли развернутых вариативных повторов (РВП) в научном тексте» [Штайн 2004: 21].

Следует подчеркнуть, что проблеме языковой и речевой частотности вообще и повторяемости в частности, как особо актуальной и значимой в лингвистике, постоянно посвящается значительное количество разнообразных лингвистических исследований (а также разного рода словарей; см., например, [Засорина 1977; Шайкевич, Андрющенко, Ребецкая 2016; Divjak 2019; Беляева 2021] и мн. др.). Так, частотность как лексикографически значимое явление «имеет большую прагматическую и научную значимость в синхронном и диахронном языковом описании. Употребительность слов учитывается при составлении словарей-минимумов, используемых для обучения иностранному и родному языку; в сфере стилистики она считается показателем

идиостиля автора и его эволюции» (см., например: [Арутюнова 1996]); частотность в аспекте исторической лексикографии «позволяет судить о динамике лексики, о формировании узульной нормативности в прошлом и о глубинных истоках современного словарного состава языка» и мн. др. [Глинкина 2011, с. 7].

Поэтому неудивительно, что с развитием цифровых технологий частотность становится всё более актуальной и информативной лингвистической категорией. Одним из таких, можно сказать, неожиданных эффектов в области изучения языковой частотности современными компьютерными и корпусными методами оказалось сравнение «естественных» научных текстов с текстами, в генерировании которых был использован принципиально новый цифровой работающий общедоступный инструмент класса «Искусственный интеллект» (ИИ), основанный на больших языковых моделях (Large Language Models, LLM) – ChatGPT-4: LLM-modified texts. «GPT – Generative Pretrained Transformer – трансформер, обученный для генерации текста» [Козловская 2023: 71].

ChatGPT-4 и особенности научной коммуникации

С момента появления нового общедоступного ИИ-инструмента ChatGPT-4, 30 ноября 2022 г., в научной литературе всё более активно обсуждается проблема его использования в порождении всех типов текстов, особенно в образовании, науке и средствах массовой информации (см., например: [Кашук 2024; Liang et al. 2023a; Poel, Gasiorek 2024; Yang et al. 2025] и мн. др.). Главная проблема заключается в том, что порожденные этим инструментом тексты с большим трудом можно отличить от текстов, написанных человеком. Так, например, в научных текстах по медицине большинство аннотаций, сгенерированных при помощи ChatGPT-4 – LLM-modified, не поддаются распознаванию как «искусственные». В средствах массовой информации, в свою очередь, было обнаружено свыше 700 новостных сайтов на 15 языках, информационные сообщения которых вызывают сомнение в их достоверности и соответствии действительности [NewsGuard 2023; Cantor 2023; Liang et al. 2024a].

К настоящему времени разработан целый ряд методов класса «GPT-Detector» для различия «естественного» текста и текста, сгенерированного искусственно (LLM-modified). Однако их эффективность и надежность подвергаются сомнению [Kelly 2023 и мн. др.]. Особо актуальной данная задача является в области научной коммуникации, поскольку здесь использование искусственного инструмента порождения текста

угрожает «экологичности» научной коммуникации. Так, в работе W. Liang с соавторами отмечается, что «стремительное освоение генеративных языковых моделей принесло значительный прогресс в цифровой коммуникации, но, в то же время, породило подозрения в возможном сомнительном использовании ИИ-сгенерированного контента» (“The rapid adoption of generative language models has brought about substantial advancements in digital communication, while simultaneously raising concerns regarding the potential misuse of AI-generated content”) [Liang et al. 2023a, с. 1]. В связи с этим подчеркивается, что результаты исследования, проведенного W. Liang и соавторами, «свидетельствуют о наличии определенных этических проблем в данной области и о необходимости предостережения авторов относительно использования ChatGPT в образовательном и квалификационном контексте» (“Our results call for a broader conversation about the ethical implications of deploying ChatGPT and caution against their use in evaluative or educational settings”) [ibid.: 1], а также о важности дальнейшего исследования способов идентификации ИИ-сгенерированного текста.

На этом основании было проведено одно из наиболее развернутых и фундаментальных исследований в данной области. Оно заключалось в последовательном масштабном статистико-стилистическом сравнении научных текстов, написанных до и после выпуска ChatGPT-4, и состояло из трех этапов [Liang et al. 2023b; 2024a; 2024b]. На первом из них, под названием “Can Large Language Models provide useful feedback on research papers? A large-scale empirical analysis” («Могут ли большие языковые модели дать полезную обратную связь по исследовательским работам? Масштабный эмпирический анализ») [Liang et al. 2023b: 1–39], изучался вопрос о том, насколько полезным может быть использование аппарата ChatGPT-4 в научной коммуникации, в первую очередь в процессе порождения рецензий на тексты научных докладов, представляемые на конференции по компьютерным наукам. При этом сравнивались рецензии, написанные специалистами (до выпуска ChatGPT-4) и предположительно сгенерированные при помощи ChatGPT-4. В ходе исследования был установлен целый ряд их отличий, а также положительные свойства созданных с помощью ChatGPT-4 рецензий. В частности, было высказано мнение, что «естественно» созданные рецензии на тексты научных докладов и искусственно сгенерированные – LLM feedback – в некотором смысле дополняют друг друга, особенно на предварительном этапе подготовки рукописи к печати, тем более что в некоторых об-

ластях, в первую очередь в компьютерных науках, наблюдается «лавинообразный» рост публикаций, и не всегда нужный специалист-рецензент находится в зоне доступа автора. Но это ни в коем случае не отменяет строгого подхода «живого» рецензента (human reviewer / expert) – специалиста и авторитета в данной области – и его участия в издательском процессе (“Our results suggest that LLM and human feedback can complement each other. While human expert review is and should continue to be the foundation of rigorous scientific process, LLM feedback could benefit researchers, especially in earlier stages of manuscript preparation before peer-review”) [Liang et al. 2023b: 2]. «Автоматическое рецензирование ни в коем случае не отменяет рецензирование, выполненное специалистом, и которое составляет неотъемлемое звено в продвижении научного знания» (“Automatically generating reviews without thoroughly reading the manuscript would undermine the rigorous evaluation process that forms the bedrock of scientific progress”) [Liang et al. 2023b: 7]. Следовательно, необходимо различать написанные специалистами и ИИ-сгенерированные рецензии.

На втором этапе – “Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews” («Масштабный мониторинг ИИ-модифицированного контента: Изучение влияния ChatGPT на экспертные заключения на конференциях по ИИ») [Liang et al. 2024a: 1–46] – была разработана специальная компьютерная дистрибутивная методика количественной оценки частотности лексики (distributional GPT quantification framework for estimating the fraction of AI-modified content in a corpus), которая позволяет наблюдать в LLM-сгенерированном тексте тенденции на уровне корпуса, слишком тонкие для обнаружения на индивидуальном текстовом уровне (“We thus can observe corpus-level trends in LLM-generated text which may be too subtle to detect at the individual level”) [ibid.: 1].

Данная методика, разработанная на материале рецензий на тексты докладов, представленных на ведущие конференции по искусственному интеллекту (ИИ; Artificial Intelligence, AI) и машинному обучению (Machine Learning, ML): ICLR (International Conference on Learning Representations), NeurIPS (Neural Information Processing Systems), EMNLP (Conference on Empirical Methods in Natural Language Processing) и CoRL (Conference on Robot Learning), позволяет продемонстрировать, что хотя и небольшая, но вполне значимая их часть, представленная на соответствующие конференции после выпуска инструмента ChatGPT-4 в свободный доступ, является ИИ-модифицированной (AI-modified).

Так, на основе специально полученных эмпирических статистических данных о том, что частотность употребления таких специфических прилагательных, как *commendable* («достойный похвалы»), *meticulous* («тщательный») и *intricate* («усложненный»), неожиданно увеличивается в ICLR-рецензиях, опубликованных после появления ChatGPT-4 в свободном доступе, то есть после 30 ноября 2022 г., было принято решение разработать соответствующие системные дистрибутивные принципы установления растущего использования LLM в корпусе научных рецензий и оценки в нем доли ИИ-модифицированного контента.

Указанные прилагательные при их более подробном статистическом изучении показали ощутимый рост в их использовании в рецензиях на материалы конференции ICLR именно после выпуска ChatGPT-4, в 2024 г., соответственно в 9,8, 34,7 и 11,2 раз (“We find a significant shift in the frequency of certain tokens in ICLR-2024, with adjectives such as “commendable”, “meticulous”, and “intricate” showing 9.8, 34.7, and 11.2-fold increases in probability of occurring in a sentence”) [ibid.: 2]. Подчеркивается, что аналогичные результаты можно получить на материале других частей речи.

В качестве исходной посылки в указанном исследовании, как уже отмечалось, послужило положение о том, что современные предполагаемые ИИ-модифицированные тексты (рецензий) можно сравнивать с текстами, написанными до появления ChatGPT-4 (“We assume that we have access to a collection of reviews which are known to contain only human-authored text”) [Liang et al. 2024a: 4]. В качестве одного из исходных предположений было принято также отмеченное в литературе наблюдение о том, что в ИИ-сгенерированном тексте вряд ли будут цитаты из других публикаций и, соответственно, цепочки вида “et al.”, отражающие их присутствие (“LLMs are less likely to include scholarly citations, as highlighted by recent studies [Walters, Wilder 2023]; we thus hypothesize that reviews containing scholarly citations might indicate lower LLM usage. To test this, we use the occurrence of the string “et al.” as a proxy for scholarly citations in reviews”) [Liang et al. 2024a: 8].

С использованием соответствующих соображений при сравнении текстов рецензий, написанных до и после выпуска инструмента ChatGPT-4, было установлено, что примерно от 7 до 15 % предложений из текстов рецензий, представленных на конференции по машинному обучению в 2024 г. (Machine Learning, ML), являются существенно ИИ-модифицированными (“Applying this method to conference reviews written before and after the release of ChatGPT shows evi-

dence that roughly 7–15% of sentences in ML conference reviews were substantially modified by AI” [ibid.: 9]. В целом полученные данные свидетельствуют, что в ИИ-модифицированных текстах существенно снижается разнообразие используемых лингвистических средств выражения, а также их «эпистемическое разнообразие» (“AI-generated texts... appear to compress the linguistic variation and epistemic diversity that would be expected in unpolluted corpora”) [ibid.: 9].

На третьем этапе исследования – “Mapping the Increasing Use of LLMs in Scientific Papers” («Отображение растущего использования LLM в научных публикациях») [Liang et al. 2024b: 1–27] – проводился анализ больших корпусных данных, включавших 950 965 научных статей, опубликованных на английском языке с января 2020 г. по февраль 2024 г. в ведущих мировых научных журналах – от вычислительной техники до медицины. Он опирался на разработанную ранее компьютерную дистрибутивную методику количественной оценки частотности лексики (distributional GPT quantification framework for estimating the fraction of AI-modified content in a corpus), которая, как уже упоминалось, позволяет наблюдать в LLM-сгенерированном тексте тенденции на уровне корпуса [Liang et al. 2024a: 1]. Сначала анализировались аннотации и введения к соответствующим статьям, поскольку они представляют собой наиболее показательные элементы научных публикаций. Устанавливалось при этом, не является ли данный (вероятно, LLM-модифицированный) текст публикации «развертыванием» ранее опубликованного текста или генерацией текста на основе заданного инструменту ChatGPT-4 плана [Liang et al. 2024b: 3].

При обучении модели была установлена частотность слов в научных текстах, написанных до выпуска ChatGPT-4 и после его выпуска. Для проверки точности модели использовались 3000 статей, вышедших из печати с 1 января 2022 года по 29 ноября 2022 года, т.е. непосредственно перед выпуском ChatGPT-4. Полученные статистические данные свидетельствуют об устойчивом росте использования LLM в научных публикациях после выпуска инструмента ChatGPT-4. Причем самый большой и быстрый рост наблюдался в публикациях по компьютерным наукам – до 17,5% (тогда как публикации по математическим наукам показали наименьшую LLM-модификацию – менее 6,3%) (подробнее см.: [Рябцева 2025]).

Полученные таким образом данные говорят о резком увеличении LLM-модифицированного (искусственно сгенерированного) контента в научных текстах публикаций уже спустя примерно пять месяцев с момента выпуска ChatGPT-4. При этом частотность слов в аннотациях по компью-

терным наукам за последние 14 лет (2010–2024 гг.) оказалась непропорционально большой у четырех слов: *realm* («область, сфера»), *intricate* («сложный, запутанный»), *showcasing* («демонстрация»), *pivotal* («решающий, основной»). И если в период с 2010 по 2022 г. эти понятия использовались в аннотациях статей по данному направлению с незначительной частотой, то с 2023 г., спустя примерно пять месяцев после поступления ChatGPT-4 в свободный доступ, они показали внезапный всплеск употребления [Liang et al. 2024b: 3]. Результаты статистического исследования к тому же свидетельствуют о более тесной связи между разными публикациями, содержащими LLM-модификации. Она проявляется в том, что «LLM-сгенерированный текст сглаживает стилистическое разнообразие лингвистических средств выражения в научной коммуникации» [ibid.: 4].

Таким образом, статистические исследования повторов и частотности лексики в научных текстах, написанных как «естественным» образом, так и предположительно ИИ-модифицированных, в разных научных направлениях и разного времени издания позволяют определить уровень «искусственности» – «модифицированности» современных научных публикаций. При этом дальнейшие попытки «измерить» степень использования LLM в порождении научных публикаций должны помочь выявить риски для «экосистемы» научных публикаций в целом. Специалисты также подчеркивают, что исследования по идентификации компьютерно-модифицированных научных текстов должны способствовать повышению прозрачности, эпистемичности и независимости научной коммуникации [Liang et al. 2024b: 9].

Заключение

В заключении следует особо отметить, что с выпуском нового общедоступного ИИ-инструмента ChatGPT-4 появился принципиально новый типа дискурса – «компьютерно-модифицированный», в результате чего наступила новая эра когнитивных, коммуникативных, стилистических, статистических и всех аналогичных исследований языка и речи, особенно в сфере научной коммуникации. В результате в настоящее время, с развитием цифровых технологий вообще, в том числе компьютерной и корпусной лингвистики в особенности, появляются разнообразные возможности взглянуть на различные коммуникативные и стилистические явления языка и речи с новой точки зрения, в том числе на особенности повторов и их частотность в различных типах текстов и на причины их появления. Как свидетельствуют соответствующие

современные цифровые статистические исследования, которые вполне правомерно назвать «компьютерной стилистикой», частотность способна выступать в качестве важнейшего критерия в установлении «естественности» или «искусственности» текстов научных рецензий и научных публикаций. Причем «искусственный» тип текста характеризуется повышенной частотностью некоторых типов, в первую очередь оценочной лексики, а также меньшей эпистемической вариативностью.

Таким образом, «компьютерная стилистика» в целом позволяет установить новые, не отмечаемые ранее и «невидимые» ранее стилистические особенности научных текстов и рецензий на них, которые проявляются при их «компьютерном» сравнении с ИИ-сгенерированными текстами, а также проследить способы их модификации, заложенные в инструментах автоматического порождения речи. И тем самым продемонстрировать преемственность и последовательность в развитии лингвистических стилистических знаний, важный вклад в которое внесла выдающийся лингвист-стилист М. Н. Кожина и вся созданная ею пермская школа стилистических исследований.

Список литературы

Арутюнова Н. Д. Стиль Достоевского в рамке русской картины мира // Пoэтика. Стилистика. Язык и культура. Памяти Татьяны Григорьевны Винокур / отв. ред. Н. Н. Розанова. М.: Наука, 1996. С. 61–90.

Баженова Е. А. Стилистика текста М. Н. Кожиной // Векторы развития современной стилистики. Стилистика и речеведение / отв. ред. Е. П. Кучина. Пермский гос. ун-т, 2013. С. 74–81.

Беляева Т. Р. Частотность и дистрибуция единиц общенаучной (академической) лексики как маркеры дисциплинарной принадлежности дискурса // Litera. 2021. № 6. С. 164–175. doi 10.25136/2409-8698.2021.6.35902

Гайда Ст. В честь Маргариты Николаевны Кожиной. Панегирик // Стереотипность и творчество в тексте: межвуз сб. науч. тр. / под ред. М. П. Котюровой. Пермь, 2010. Вып. 14. С. 6–11.

Глинкина Л. А. Частотность как значимый регистр лексикографии и фразеографии // Проблемы истории, филологии, культуры. 2011. № 3 (33). С. 7–11.

Данилевская Н. В. Динамика формирования знания в научном дискурсе (Функционально-стилистический аспект) // Вестник ТГПУ. Серия: Гуманитарные науки (Филология). 2005. Вып. 3(47). С. 14–17.

Засорина Л. Н. Частотный словарь русского языка / под ред. Л. Н. Засорина. М.: Рус. яз., 1977. 936 с.

Кащук С. М. Теория и практика интеграции GPT-подобных языковых моделей в систему иноязычного образования // Язык и действительность. Научные чтения на кафедре романских языков им. В. Г. Гака: сб. ст. по итогам IX междунар. конф. (Москва, 25–29 марта 2024 г.). Т. 9. М.: Спутник +, 2024. С. 382–386.

Кожина М. Н. Дискурсный анализ и функциональная стилистика с речеведческих позиций // Текст – Дискурс – Стиль: сб. науч. ст. СПб., 2004. С. 9–33.

Кожина М. Н. Общая характеристика текстовых категорий в функционально-стилевом аспекте // Очерки истории научного стиля русского литературного языка XVIII–XX вв. Т. II. Ч. 2: Категории научного текста: Функционально-стилистический аспект. / отв. ред. Н. М. Кожина. Пермь, 1998. С. 3–15.

Кожина М. Н. О специфике художественной и научной речи в аспекте функциональной стилистики. Пермь, 1966. 213 с.

Кожина М. Н. Речеведение. Теория функциональной стилистики. Избранные труды. 3-е изд., стер. М.: Флинта, 2020. 623 с.

Кожина М. Н. Соотношение стилистики текста со смежными дисциплинами // Очерки истории научного стиля русского литературного языка XVIII–XX вв. Т. II. Ч. 1: Стилистика научного текста (общие параметры) / отв. ред. Н. М. Кожина. Пермь, 1996. С. 11–29.

Кожина М. Н. (отв. ред.) Стилистический энциклопедический словарь русского языка. М.: Флинта, 2003. 696 с.

Козловская Н. В. GPT и CHAT-GPT: как это по-русски? (об электронных словарных статьях в неологическом ресурсе neolex.ilinc.spb.ru) // Terra Linguistica. 2023. Т. 14, № 4. С. 67–78. doi 10.18721/JHSS.14405.

Рябцева Н. К. Когнитивные исследования дискурса и современная «компьютерная стилистика» // Когнитивные исследования языка и дискурса: материалы Всерос. науч. конф. (Москва, 30–31 октября 2025 г.). Вып. № 3(64): Ч. I / гл. ред. вып. О. К. Ирисханова. Тамбов: Изд. дом «Державинский», 2025. С. 112–119.

Салимовский В. А. Вклад М. Н. Кожиной в развитие лингвистической стилистики и становление речеведения // Векторы развития современной стилистики. Стилистика и речеведение. Пермь: Перм. гос. ун-т, 2013. С. 7–32.

Шайкевич А. Я., Андрющенко В. М., Ребецкая Н. А. Дистрибутивно-статистический анализ языка русской прозы 1850–1870-х гг. М.: Языки славянской культуры». 2016. 849 с.

Штайн К. Э. Культурное достояние России: Пермская научная школа функциональной стилистики // Стереотипность и творчество в тексте:

Межвуз. сб. науч. тр. Вып. 7 / отв. ред. М. П. Котюрова. Пермь: Перм. гос. ун-т, 2004. С. 6–57.

Cantor M. Nearly 50 news websites are ‘AI-generated’, a study says. Would I be able to tell? 2023. URL: <https://www.theguardian.com/technology/2023/may/08/ai-generated-news-websites-study> (дата обращения: 27.06.2025).

Diyjak D. Frequency in Language: Memory, Attention and Learning. Cambridge: Cambridge University Press, 2019. 328 p. doi 10.1017/9781316084410

Kelly S. M. ChatGPT creator pulls AI detection tool due to ‘low rate of accuracy’ // CNN Business. 2023. 25 July. URL: <https://www.cnn.com/2023/07/25/tech/openai-ai-detection-tool/index.html>. (дата обращения: 28.06.2025).

Liang W. et al. GPT Detectors Are Biased Against Non-native English Writers / W. Liang, M. Yuksekgonul, Y. Mao, E. Wu, J. Y. Zou // arXiv:2304.02819v3 [cs.CL]. 2023a 10 Jul. doi 10.48550/arXiv.2304.02819

Liang W. et al. Can Large Language Models Provide Useful Feedback on Research Papers? A Large-scale Empirical Analysis / W. Liang, Y. Zhang, H. Cao, B. Wang, D. Ding, X. Yang, K. Vodrahalli, S. He, D. Smith, Y. Yin, D. McFarland, J. Zou // arXiv:2310.01783. 2023b. doi 10.48550/arXiv.2310.01783

Liang W. et al. Monitoring AI-modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews / W. Liang, Z. Izzo, Y. Zhang, H. Lepp, H. Cao, X. Zhao, L. Chen, H. Ye, Sh. Liu, Z. Huang, D. A. McFarland, J. Y. Zou. // International Conference on Machine Learning (ICML). 2024a. doi 10.48550/arXiv.2403.07183

Liang W. et al. Mapping the Increasing Use of LLMs in Scientific Papers / W. Liang, Y. Zhang, Z. Wu, H. Lepp, W. Ji, Z. Zhao, H. Cao, S. Liu, S. He, Z. Huang, D. Yang, C. Potts, C. D. Manning, J. Y. Zou. // First Conference on Language Modeling COLM-2024. (Published as a conference paper at COLM-2024). 2024b. P. 1–27 doi 10.48550/arXiv.2404.01268

NewsGuard. Tracking AI-enabled Misinformation: 713 ‘Unreliable AI-Generated News’ Websites (and Counting), Plus the Top False Narratives Generated by Artificial Intelligence Tools. 2023. URL: <https://www.newsguardtech.com/special-reports/ai-tracking-center/> (дата обращения: 28.06.2025).

Van de Poel K., Gasiorek J. Using AI to Expand the “Toolbox” for EAP Writing Instruction: Student Experiences and Perceptions of ChatGPT’s Instructional Potential // AILA Review. 2024. doi 10.1075/aila.24029.van

Walters W. H., Wilder E. I. Fabrication and Errors in the Bibliographic Citations Generated by ChatGPT // Scientific Reports. 2023. Vol. 13. Iss. 14045. doi 10.1038/s41598-023-41032-5

Yang W., Chen J., Lin Y., Wen J. DeepCritic: Deliberate Critique with Large Language Models // arXiv:2505.00662v1 [cs.CL]. May 2025. URL: <https://arxiv.org/pdf/2505.00662.pdf> (дата обращения: 28.06.2025).

References

Arutyunova N. D. Stil' Dostoevskogo v ramke russkoy kartiny mira [Dostoevsky's style within Russian worldview]. *Poetika. Stilistika. Yazyk i kul'tura. Pamyati Tat'yany Grigor'evny Vinokur* [Poetics. Stylistics. Language and Culture. In Memory of Tatiana Grigorievna Vinokur]. Ed. by N. N. Rozanova. Moscow, Nauka Publ., 1996, pp. 61–90. (In Russ.)

Bazhenova E. A. *Stilistika teksta M. N. Kozhinoj* [Text stylistics of M. N. Kozhina]. *Vektory razvitiya sovremennoy stilistiky. Stilistika i rechevedenie* [Vectors of Development of Modern Stylistics. Stylistics and Speech Studies]. Ed. by E. P. Kuchina. Perm State University Press, 2013, pp. 74–81. (In Russ.)

Belyaeva T. R. Chastotnost' i distributsiya edinits obshchenauchnoy (akademicheskoy) leksiki kak markery distsiplinarnoy prinadlezhnosti diskursa [Frequency and distribution of the units of general scientific (academic) lexicon as the markers of disciplinary affiliation of a discourse], *Litera*, 2021, issue 6, pp. 164–175. doi 10.25136/2409-8698.2021.6.35902. (In Russ.)

Gayda St. V chest' Margarity Nikolaevny Kozhinoj. Panegirik [In honor of Margarita Nikolaevna Kozhina. Encomium]. *Stereotipnost' i tvorchestvo v texte* [Stereotypes and Creativity in Text: an interuniversity collection of scientific papers]. Ed. by M. P. Kotyurova. Perm, 2010, issue 14, pp. 6–11. (In Russ.)

Glinkina L. A. Chastotnost' kak znachimyy registr leksikografii i frazeografii [Frequency as an important characteristic of lexicography and phraseography]. *Problemy istorii, filologii, cultura* [Journal of Historical, Philological and Cultural Studies], 2011, issue 3 (33), pp. 7–11. (In Russ.)

Danilevskaya N. V. Dinamika formirovaniya znaniya v nauchnom diskurse (Funktional'no-stilisticheskiy aspekt) [Dynamics of knowledge formation in scientific discourse (Functional and stylistic aspect)]. *Vestnik TGPU. Seriya: Gumanitarnye nauki (Filologiya)* [Tomsk State Pedagogical University Bulletin. Series: Humanities (Philology)], 2005, issue 3 (47), pp. 14–17. (In Russ.)

Zasorina L. N. *Chastotnyy slovar' russkogo yazyka* [Frequency Dictionary of the Russian Language]. Ed. L. N. Zasorin. Moscow, Russian Language Publ., 1977. 936 p. (In Russ.)

Kashchuk S. M. Teoriya i praktika integratsii GPT-podobnykh yazykovykh modeley v sistemuyu inoyazychnogo obrazovaniya [Theory and practice of integrating GPT-like language models into the foreign language education system]. *Yazyk i deystvitel'nost'*. Nauchnye chteniya na kafedre romanskikh yazykov im. V.G. Gaka [Language and Reality. Scientific Readings at the Department of Romance Languages named after V. G. Gak: Proceedings of the IX International Conference (Moscow, March 25–29, 2024)]. Moscow, Sputnik + Publ., 2024, vol. 9, pp. 382–386. (In Russ.)

Kozhina M. N. Diskursnyy analiz i funktsional'naya stilistika s rechevedcheskikh pozitsiy [Discourse analysis and functional stylistics from a speech studies perspective]. *Tekst – Diskurs – Stil'* [Text – Discourse – Style: a collection of scientific articles]. St. Petersburg, Nauka Publ., 2004, pp. 9–33. (In Russ.)

Kozhina M. N. Obshchaya kharakteristika tekstovykh kategoriy v funktsional'no-stilevom aspekte [General characteristics of text categories in the functional-stylistic aspect]. *Ocherki istorii nauchnogo stilya russkogo literaturnogo yazyka XVIII–XX vv.* [Essays on the History of Scientific Style of the Russian Literary Language in the 18th–20th Centuries]. Ed. by M. N. Kozhina. Perm, Perm State University Press, 1998, vol. II, pt. 2. Kategorii nauchnogo teksta: Funktsional'no-stilisticheskiy aspekt [Categories of scientific text: Functional and stylistic aspects], pp. 3–15. (In Russ.)

Kozhina M. N. *O spetsifike khudozhestvennoy nauchnoy rechi v aspekte funktsional'noy stilistiki* [On the specifics of artistic and scientific speech in the aspect of functional stylistics]. Perm, Perm State University Press, 1966. 213 p. (In Russ.)

Kozhina M. N. *Rechevedenie. Teoriya funktsional'noy stilistiki: Izbrannye trudy* [Speech Studies. Theory of Functional Stylistics: Selected works]. 3rd stereot. ed. Moscow, Flinta Publ., 2020. 623 p. (In Russ.)

Kozhina M. N. Sootnoshenie stilistiki teksta so smezhnymi distsiplinami [Correlation of text stylistics with related disciplines]. *Ocherki istorii nauchnogo stilya russkogo literaturnogo yazyka XVIII–XX* [Essays on the History of Scientific Style in the Russian Literary Language of the 18th–20th Centuries]. Ed. by M. N. Kozhina. Perm, Perm State University Press, 1996, vol. II, pt. 1. Stilistika nauchnogo teksta (obshchie parametry) [Stylistics of scientific texts (general parameters)], pp. 11–29. (In Russ.)

Stilisticheskiy entsiklopedicheskiy slovar' russkogo yazyka [Stylistic Encyclopedic Dictionary of the Russian Language]. Ed. by M. N. Kozhina. Moscow, Flinta Publ., 2003. 696 p. (In Russ.)

Kozlovskaya N. V. GPT i CHAT-GPT: kak eto po-russki? (Ob elektronnykh slovarnykh stat'yakh v neologicheskem resurse neolex.iling.spb.ru) [GPT and

CHATGPT: How is it in Russian? (on electronic dictionary entries in the neological database neolex.iling.spb.ru)]. *Terra Linguistica*, 2023, vol. 14, issue 4, pp. 67–78. doi 10.18721/JHSS.14405. (In Russ.)

Riabtseva N. K. Kognitivnye issledovaniya diskursa i sovremennoy 'komp'yuternoy stilistiki' [Cognitive discourse studies and contemporary 'computer stylistics']. *Kognitivnye issledovaniya yazyka i diskursa* [Cognitive Studies of Language and Discourse: Proceedings of the All-Russian scientific conference. Moscow, 30–31 October 2025]. Issue 3(64): Pt. 1. Ed. by O. K. Iris Khanova. Moscow, Tambov, 2025, pp. 112–119. (In Russ.)

Salimovskiy V. A. Vklad M. N. Kozhinoi v razvitiye lingvisticheskoy stilistiki i stanovlenie rechevedeniya [The contribution of M. N. Kozhina to the development of linguistic stylistics and the formation of speech studies]. *Vektory razvitiya sovremennoy stilistiki. Stilistika i rechevedenie* [Vectors of Development of Modern Stylistics. Stylistics and Speech Studies]. Perm, Perm State University Press, 2013, pp. 7–32. (In Russ.)

Shaykevich A. Ya., Andryushchenko V. M., Rebetskaya N. A. *Distributivno-statisticheskiy analiz yazyka russkoy prozy 1850–1870-kh gg.* [Distributional and Statistical Analysis of the Language of Russian Prose of the 1850s–1870s]. Moscow, LRC Publishing House Publ., 2016. 849 p. (In Russ.)

Shtayn K. E. Kul'turnoe dostoyanie Rossii: Permskaya nauchnaya shkola funktsional'noy stilistiki [Cultural heritage of Russia: Perm Scientific School of Functional Stylistics]. *Stereotipnost' i tvorchestvo v tekste* [Stereotypes and Creativity in Text: an interuniversity collection of scientific papers]. Ed. by M. P. Kotyurova. Perm, Perm State University Press, 2004, issue 7, pp. 6–57. (In Russ.)

Cantor M. *Nearly 50 news websites are 'AI-generated', a study says. Would I be able to tell?* 2023. Available at: <https://www.theguardian.com/technology/2023/may/08/ai-generated-news-websites-study> (accessed 27 June 2025). (In Eng.)

Diyjak D. *Frequency in Language: Memory, Attention and Learning*. Cambridge, Cambridge University Press, 2019. 328 p. doi 10.1017/9781316084410. (In Eng.)

Kelly S. M. ChatGPT creator pulls AI detection tool due to 'low rate of accuracy'. *CNN Business*. 2023, 25 July. Available at: <https://www.cnn.com/2023/07/25/tech/openai-ai-detection-tool/index.html>. (accessed 28 June 2025). (In Eng.)

Liang W. et al. GPT detectors are biased against non-native English writers. *arXiv:2304.02819v3* [cs.CL]. 2023a, 10 Jul. doi 10.48550/arXiv.2304.02819. (In Eng.)

Liang W. et al. Can large language models provide useful feedback on research papers? A large-

- scale empirical analysis. *arXiv:2310.01783*, 2023b. doi 10.48550/arXiv.2310.01783. (In Eng.)
- Liang W. et al. Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. *International Conference on Machine Learning (ICML)*, 2024a. doi 10.48550/arXiv.2403.07183. (In Eng.)
- Liang W. et al. Mapping the increasing use of LLMs in scientific papers. *First Conference on Language Modeling COLM-2024*. (Published as a conference paper at COLM-2024), 2024b, pp. 1-27. doi 10.48550/arXiv.2404.01268. (In Eng.)
- Tracking AI-enabled misinformation: 713 ‘Unreliable AI-generated news’ websites (and counting), plus the top false narratives generated by artificial intelligence tools. *NewsGuard*, 2023. Available at: <https://www.newsguardtech.com/special-reports/ai-tracking-center/> (accessed 28 June 2025). (In Eng.)
- van de Poel K., Gasiorek J. Using AI to expand the ‘Toolbox’ for EAP writing instruction: Student experiences and perceptions of ChatGPT’s instructional potential. *AILA Review*, 2024. doi 10.1075/aila.24029.van. (In Eng.)
- Walters W. H., Wilder E. I. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, 2023, vol. 13, issue 14045. doi 10.1038/s41598-023-41032-5. (In Eng.)
- Yang W., Chen J., Lin Y., Wen J. DeepCritic: Deliberate critique with large language models. *arXiv:2505.00662v1 [cs.CL]*. May 2025. Available at: <https://arxiv.org/pdf/2505.00662.pdf> (accessed 28 June 2025). (In Eng.)

M. N. Kozhina's Functional Stylistics of Scientific Text and Current Corpus-Based Studies in Detecting Artificially Generated Content

Nadezhda K. Riabtseva

Leading Researcher in the Department of Applied Linguistics

Institute of Linguistics of the Russian Academy of Sciences

1, bld. 1, Bolshoy Kislovskiy pereulok, Moscow, 125009, Russia. nadia_riabceva@mail.ru

SPIN-code: 8966-9640

ORCID: <http://orcid.org/0000-0002-2042-4615>

ResearcherID: S-7138-2016

Scopus Author ID: 55218430500

Submitted 26 Sep 2025

Revised 29 Sep 2025

Accepted 01 Oct 2025

For citation

Riabtseva N. K. Funktsional'naya stilistika nauchnogo teksta M. N. Kozhinoj i sovremennye korpusnye issledovaniya po identifikatsii iskusstvenno sgenerirovannogo kontenta [M. N. Kozhina's Functional Stylistics of Scientific Text and Current Corpus-Based Studies in Detecting Artificially Generated Content]. *Vestnik Permskogo universiteta. Rossiyskaya i zarubezhnaya filologiya* [Perm University Herald. Russian and Foreign Philology], 2025, vol. 17, issue 4, pp. 81–90. doi 10.17072/2073-6681-2025-4-81-90. EDN LQZSLV (In Russ.)

Abstract. The paper emphasizes the significance of scientific works by the outstanding Russian scientist, Professor at Perm State University M. N. Kozhina, as well as her contribution to linguistics in general and contemporary stylistics in particular. The paper shows the connections between current computer- and corpus-based statistical studies and those of Perm stylistics school of thought, founded by M. N. Kozhina. It also presents the latest corpus-based statistical research results in detecting artificially generated texts and demonstrates the continuity and effectiveness of quantitative studies in language, speech, and communication, as well as their significance for the development of science in general. Statistical data show a steady increase in the use of the open-source Artificial Intelligence tool ChatGPT-4 in scientific publications that started almost immediately after its release on November 30, 2022. The largest and fastest growth in artificially generated content has been noted in publications on computer sciences – up to 17.5%. This fact was established as a result of a systemic large-scale statistical comparison across more than 950,900 papers published in English in leading scientific journals from various academic fields between January 2020 and Feb-

ruary 2024, i.e. before and after the release of ChatGPT-4. In abstracts published in computer science journals over the last 14 years (2010-2024), four words demonstrate disproportionately high frequency of use – *realm*, *intricate*, *showcasing*, *pivotal*. Indicatively, a sudden surge in the use of these words occurred in 2023, about five months after ChatGPT-4 became freely available, while in the period from 2010 to 2022 their use was consistently low. The paper highlights the challenges posed by the use of ChatGPT-4, especially in scientific communication, and the principles to be used for addressing them.

Key words: functional stylistics; scientific text; corpus-based studies; frequency; ChatGPT-4; artificially generated content.