

УДК 81'322

DOI: 10.18413/2313-8912-2026-12-3-0-7

Колмогорова А. В.¹ **AI как писатель глазами читателя: результаты психолингвистического эксперимента по оценке сгенерированных художественных текстов любительской прозы**

¹ НИУ «Высшая школа экономики» – Санкт-Петербург
наб. канала Грибоедова, 123, Санкт-Петербург, 197343, Россия
E-mail: akolmogorova@hse.ru
ORCID: 0000-0002-6425-2050

*Статья поступила 10 марта 2026 г.; принята 15 сентября 2026 г.;
опубликована 30 сентября 2026 г.*

Аннотация: Большие языковые модели выполняют все больше разнообразных задач, в том числе генерируют длинные тексты, претендующие на художественность. **Цель статьи** – сравнить субъективные оценки качества художественных текстов, созданных большими языковыми моделями (БЯМ), с текстами, написанными людьми в жанре любительской прозы по мотивам романов о Гарри Поттере. **Материалом** для оценки послужили шесть текстов на английском языке: три, сгенерированные искусственным интеллектом с использованием разных подходов, и три, написанные реальными авторами. Исследование проведено с помощью **метода** психолингвистического эксперимента. Восемнадцать информантов оценивали тексты по восьми критериям, отражающим ключевые категории художественности: когерентность (целостность), смысловая и лексико-грамматическая связность (когезия), ритм повествования, эмоциональность, субъективная модальность сомнения, сложность синтаксиса и выразительность (идиоматичность, креативность, сильные позиции текста). Обработка результатов включала статистический анализ объективных параметров лингвистической сложности текстов, полученных от информантов субъективных оценок, а также кластеризацию векторизованных пересказов текстов, предоставленных информантами.

Результаты показали, что разрыв между человеческими и машинными текстами неодинаков в разных категориях. Благодаря специальным методам генерации, синтетические тексты смогли приблизиться к естественным по параметрам общей когерентности и синтаксической сложности. Однако наиболее значимые отличия проявились в категориях ритма, смысловой связности и присутствия субъективной модальности сомнения, которые читатели оценили в синтетических текстах значительно ниже. Тем не менее, один из сгенерированных текстов продемонстрировал качество, сопоставимое с написанными текстами, по таким параметрам, как идиоматичность, креативность и способность маркировать сильные позиции. Кластеризация пересказов подтвердила, что этот текст, в отличие от двух других сгенерированных, вызвал у читателей более единообразное понимание сюжета.

Делается вывод, что главным вызовом для ИИ остается не столько построение иерархической структуры связного текста, сколько создание динамичного, психологически достоверного и ритмически организованного повествования.

Категории смысловой связности, ритма и модальности сомнения являются на данный момент той границей, которая отделяет искусственный интеллект от человеческого в сфере литературного творчества.

Ключевые слова: Большая языковая модель; Генерация текстов; Художественный текст; Психолингвистический эксперимент; Критерии художественности

Благодарности: Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ (HSE-BR-2025-032).

Информация для цитирования: Колмогорова А. В. AI как писатель глазами читателя: результаты психолингвистического эксперимента по оценке сгенерированных художественных текстов любительской прозы // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 3. С. 170-199. DOI: 10.18413/2313-8912-2026-12-3-0-7

УДК 81'322

DOI: 10.18413/2313-8912-2026-12-3-0-7

Anastasia V. Kolmogorova¹ 

**AI as a literary author from the reader's perspective:
findings from a psycholinguistic experiment assessing
synthetic amateur prose**

¹ Higher School of Economics National Research University (HSE University) – St. Petersburg,
123 Griboyedov Canal Emb., St. Petersburg, 197343, Russia

E-mail: akolmogorova@hse.ru

ORCID: 0000-0002-6425-2050

Received 10 March 2026; accepted 15 June 2026; June 30 March 2026

Abstract: Large language models (LLMs) are performing an increasing variety of tasks, including generating long texts that aspire to literary quality. **The aim** of this article is to compare subjective assessments of the quality of literary texts created by large language models with human-written texts in the genre of fan fiction based on the Harry Potter novels. **The material** for the evaluation consisted of six texts in English: three generated by AI using different approaches, and three written by human authors. The present study utilizes a psycholinguistic experimental design. Eighteen informants evaluated the texts according to eight criteria reflecting key categories of literariness: coherence (integrity), semantic and lexico-grammatical cohesion, narrative rhythm, emotionality, subjective modality of doubt, syntactic complexity, and expressiveness (idiomaticity, creativity, strong textual positions). The data processing included statistical analysis of objective parameters of linguistic complexity in the texts, subjective evaluations obtained from informants, as well as clustering of vectorized retellings of the texts provided by the informants.

The results showed that the gap between human and AI-generated texts varies across different categories. Thanks to specific generation methods, the synthetic texts managed to approach human-written ones in terms of overall coherence and syntactic complexity. However, the most significant differences emerged in the categories of rhythm, semantic cohesion, and the presence of the subjective modality of doubt, which readers rated significantly lower in the synthetic texts. Nevertheless, one of the generated texts was of a quality comparable to that of a human writer in terms of idiomaticity, creativity and the ability to take a strong position on a topic. Clustering of summaries confirmed

that this text, unlike the other two generated ones, led to a more uniform understanding of the plot among readers.

It can be concluded that the main challenge for AI is not so much the construction of a coherent, hierarchical text structure, but rather the creation of a dynamic, psychologically authentic and rhythmically organised narrative. The categories of semantic cohesion, rhythm, and modality of doubt currently represent the frontier separating artificial intelligence from human intelligence in the realm of literary creativity.

Keywords: Large language model; Text generation; Literary text; Psycholinguistic experiment; criteria of literariness

Acknowledgements: The article is an output of a research project HSE-BR-2025-032 implemented as part of the Basic Research Project at HSE University.

How to cite: Kolmogorova, A. V. (2026). AI as a literary author from the reader's perspective: findings from a psycholinguistic experiment assessing synthetic amateur prose, *Research Result. Theoretical and Applied Linguistics*, 12 (3), 170–199. DOI: 10.18413/2313-8912-2026-12-3-0-7

1. Введение

Появление Больших языковых моделей (далее – БЯМ) существенно изменило не только диапазон функций, которые отныне человек может делегировать моделям, но и заставило перейти к размышлениям о том, что составляет специфику когнитивной системы и речевой деятельности Homo Sapiens.

Когда в 2021 г. начали внедряться в практику модели типа Generative Pre-trained Transformer (GPT) стало очевидно, что функция порождения осмысленных текстов им вполне по силам. Однако на тот момент моделям явно не хватало дополнительных знаний, они не обладали экспертностью в специфических областях профессиональной деятельности, например, в машиностроении, медицине, креативных индустриях и не могли перейти от рутинных задач на уровень созидательности в процессе генерации текстов. К 2026 году многие из вышеупомянутых лакун оказались заполнены.

Так, благодаря технологии Retrieved Augmented Generation (RAG) (Lewis et al., 2021), современные БЯМ легко решают задачи генерации текстов, вербализующих экспертную оценку: чтобы предлагать пользователю достоверную информацию

из авторитетных источников, модель обращается к специальным базам знаний, подготовленным специалистами в той или иной сфере.

БЯМ могут выступать в качестве внешнего эксперта, оценивая свои собственные результаты – в этом случае применяется подход «*LLM as a judge*» (Zheng et al., 2023). Модели эффективно используются для анализа мнений и эмоциональной тональности текстов (Vatolin, 2025; Kolmogorova et al., 2025), тестируются в задаче эмоциональной оценки стимульных слов (Iaroshenko, Louckachevitch, 2025). Иными словами, осталось ограниченное число задач в области порождения и понимания текстов, в которых естественные когнитивные агенты (люди) сохраняют свою исключительность и уникальность. Одна из таких сфер – создание художественных текстов, обладающих оригинальным содержанием, системой смыслов и собственной эстетикой.

В данной статье представлен один из аспектов работы, проведенной группой исследователей по генерации длинных текстов. Поскольку на сегодняшний момент так называемое контекстное окно (т.е. то количество токенов, которое БЯМ может удерживать в оперативной памяти) большинства моделей достаточно

ограничено, модель не может сгенерировать связный и тематически единый текст объемом 40 000 знаков и более по одному запросу. В этой связи «длинная генерация» представляет собой исследовательскую и техническую проблему, которой и было посвящено соревнование, организованное в рамках Международной научной конференции по текстовой генерации INLG-2024¹. Однако в фокусе нашей статьи находится другой аспект – не технический, а лежащий в области психолингвистики текста. После того, как команды-участники соревнования предоставили организаторам получившиеся синтетические фанфикшн-тексты, мы перемешали их с подобными текстами, созданными реальными авторами, пишущими в данном жанре. Сформулировав ряд категорий, важных для характеристики художественного текста, мы разработали внутри каждой категории критерии для оценивания и предложили информантам оценить по ним шесть текстов, три из которых были сгенерированными, а другие три – написанными.

Получившиеся результаты, на наш взгляд, требуют филологического и психолингвистического осмысления, поскольку в них интересно не только то, как данные категории поддаются синтетическому воспроизведению, но и то, как читатели мыслят «художественность». Таким образом, цель данной статьи состоит в том, чтобы сравнить полученные в эксперименте субъективные оценки качества художественных текстов, созданных БЯМ и написанных людьми, и оценить на материале данной конкретной выборки, какие текстовые категории 1) важны для субъективного канона “художественного текста”, сформированного в сознании студентов-филологов, принимавших участие в

эксперименте; 2) уже присутствуют в синтетических текстах, позволяя последним определенным образом конкурировать с написанными текстами.

2. Техники генерации длинных текстов непрофессиональной прозы

Главная проблема генерации длинных текстов состоит в том, что БЯМ пока не способны строить иерархически организованный текст. Основное математически обоснованное отличие текстов сгенерированных от текстов написанных состоит в том, что значения автокорреляции² у первых уменьшаются по экспоненте, скользя резко вниз по мере удлинения текста, а у вторых – по степенному закону, т.е. постепенно (Базилевский, 2018). Иными словами, модель быстро «забывает» то, что было написано ею ранее – чем дальше от актуального «окна» текстопорождения отодвигается контекст, тем меньше вероятность того, что модель каким-то образом свяжет его с текстом, генерируемым в данный момент (Takahashi and Tanaka-Ishii, 2019). Более плавное же снижение значений автокорреляции, по мнению Н. Э. Михайловского (Mikhaylovskiy, 2025), обусловлено иерархичностью организации художественного текста, написанного писателем. Так, в романе «Война и мир» Л. Н. Толстого можно насчитать семь уровней текстовой иерархии: весь роман – том – часть – глава – параграф – строка – буква (Там же). При этом, реализуя такую иерархичность, писатель может легко переходить между уровнями – оттолкнувшись от некоторой детали, создать образ персонажа, действия и поступки которого составят основу сюжетной канвы всего произведения. Иллюстративный пример

¹ Сайт Международной научной конференции International Natural Language Generation Conference. URL: <https://2025.inlgmeeting.org/> (дата обращения: 25.05.2026).

² Статистическая мера, которая показывает, насколько значения одного и того же сигнала (например, последовательности слов в тексте) коррелируют друг с другом на разных временных или пространственных расстояниях.

структурирования такого рода иерархии на этапе писательского замысла можно найти в автобиографических заметках Ю. Олеша «Ни дня без строчки» (Олеша, 1961). В одной из заметок можно проследить, как писатель в режиме реального времени намечает контуры замысла будущей пьесы об учителе пения Цветкове, который хочет познакомить всех учеников школы с творчеством У. Шекспира и расстраивается, когда намеченная читка текстов классика превращается в нелепую сцену из-за организационной оплошности его лучшей ученицы Тони. После такой, первой, идеи Ю. Олеша формулирует основную мысль будущей пьесы: «каждый может ознакомиться с Шекспиром, важно самому быть героем» и переносит дальнейшее действие в годы Великой Отечественной войны, где Тоня героически погибает от рук фашистов. Иными словами, мысль писателя движется от первой возникшей в его воображении сцены к общему замыслу пьесы, чтобы затем вновь вернуться к развертыванию дальнейшего сюжета. Кроме того, в случае кольцевой структуры романа, писатель в конце текста может предложить читателю вновь обратиться к некоторым важным ориентирам, оставленным им в начале текста для того, чтобы полнее переосмыслить метаморфозы, которые пережил герой (Шкловский, 2025: 58). Таким образом, именно присутствие иерархичности на уровне художественной организации позволяет признавать некоторую совокупность фраз естественного языка художественным текстом (Лотман, 1998: 38).

Следует признать, что пока БЯМ не могут в полной мере выстроить подобную иерархию и свободно перемещаться между ее уровнями, подобно человеку (Ippolito et al., 2022; Jin et al., 2022), довольствуясь ролью ассистента писателя (Singh et al., 2023; Zhong et al., 2023; Zhou et al., 2024) или поэта (Chakrabarty et al., 2022). На поиск возможных путей преодоления этих трудностей и была нацелена коллективная

работа по генерации длинных художественных текстов в рамках INLG.

Общая методологическая рамка экспериментов с БЯМ была выстроена следующим образом.

Задачей участников было получить от БЯМ, выбор которых не ограничивался, тексты любительской прозы по мотивам романа Джоан Роулинг «Гарри Поттер». Важным условием было то, что текст должен быть написан на английском языке, а его объем не должен был быть меньше, чем 40 000 знаков. В качестве некоторой начальной стимульной инструкции для БЯМ предлагались тексты объемом 1500 знаков, написанные студентами-лингвистами 4 курса бакалаврской программы «Лингвистика» отделения иностранных языков Пермского государственного национально-исследовательского университета и представляющие собой зачины историй на основе сюжета романа.

Большинством команд-участников соревнования была применена близкая методология.

Так, в (Seredina, 2024) использовался подход, при котором на первом этапе БЯМ Mistral-7B-Instruct-v0.2-GPTQ³ генерировала структуру нарратива: общее описание контекста, в котором происходит действие; завязку; кульминацию и развязку истории. Затем с помощью нового промпта и той же модели генерировались названия глав и их краткое содержание. На третьем этапе проводилась тонкая настройка модели Mistral-7B-Instruct-v0.2-GPTQ, которая дообучалась гибридным способом на основе методов обучения с учителем (Supervised Learning) и самообучения (Self-Supervised Learning). Для тонкой настройки модели использовался датасет, который состоял из текстов 500 книг на английском языке, размещенных на платформе проекта Гуттенберг⁴. Каждая

³ <https://huggingface.co/TheBloke/Mistral-7B-Instruct-v0.2-GPTQ>

⁴ <https://www.gutenberg.org/>

книга разбивалась на фрагменты по 1500 знаков, для которых затем БЯМ давала краткий пересказ. Таким образом, модели обучалась сопоставлять пересказ и развернутый оригинальный текст. После такой тонкой настройки Mistral-7B-Instruct-v0.2-GPTQ получала промпт на английском языке такого содержания: «Представь, что ты великий писатель и ты пишешь текст, следуя сюжетной канве <описание содержания главы, полученное на предыдущем этапе>. Опиши не только действия персонажей, но и их чувства, поведение и мотивы, добавь столько деталей, сколько нужно». В результате реализации такой структуры был получен текст требуемого объема в 40 000 знаков. Данный текст в нашей выборке обозначен номером 6.

В (Boriskin, Galimzianova, 2024) описан сходный, но не идентичный метод: авторы не использовали тонкую настройку, а просили модель генерировать сначала общее содержание истории, затем – названия глав и краткое содержание каждой главы, а потом последовательно подавали на вход модели Llama 3.1⁵ (70 миллиардов параметров) промпты, которые содержали в себе краткий пересказ содержания предыдущей главы. В нашей выборке данный текст будет иметь номер 3.

Похожий подход применялся в модели, которая была предложена организаторами соревнования⁶. Сначала на вход БЯМ подавалась инструкция на английском языке следующего содержания: «Представь, что ты великий писатель и ты хочешь написать новую книгу о Гарри Поттере. Используй следующий зачин для начала истории <один из зачинов, предложенных студентами>. Раздели свое описание сюжета на 5 частей по образцу: “Содержание: Часть №: описание содержания части”. Не добавляй никаких

комментариев». Аналогичным образом генерировались описания отдельных глав, а затем на вход модели подавались краткие содержания глав, чтобы получить собственно текст главы. Полученный текст будет фигурировать в нашей выборке под номером 5 (Migal et al., 2024).

Таким образом было получено три синтетических текста любительской прозы на английском языке по мотивам романа Джоан Роулинг «Гарри Поттер».

Общая описательная статистика по синтетическим и естественным текстам, участвовавшим в психолингвистическом эксперименте по оценке качества генерации представлена в разделе «Материал и методы».

3. Основные категории художественного текста

В исследованиях по созданию синтетических художественных текстов уже сложилась некоторая критериальная сетка, которая используется при оценке качества генерации такого текста информантами. Например, в (Lee et al., 2024) сгенерированные тексты оцениваются по 8 категориям: когерентность (соответствуют ли заголовки тексту); логичность развертывания сюжета; разнообразие тематических линий в процессе развертывания нарративной схемы; повторяемость тем (оценивается, насколько успешно в сгенерированном тексте удастся избежать семантических повторов); «полезность информации», добавляемой в частях, конкретизирующих общую нарративную схему; «степень улучшения» (оценивает, насколько конкретизирующий текст отдельных глав и частей лучше раскрывает детали и смыслы общей нарративной схемы). В работах (Kryscinski et al., 2019; Shen et al., 2023; Guan et al., 2022) в качестве основных категорий используются: релевантность тем; соответствие частей (связь частей текста между собой); отсутствие повторяющихся сюжетов, отсутствие

⁵ <https://llama3-1.com/>

⁶ <https://lsgc.vercel.app/baseline>

хаотических сцен, беглость (естественность текста каждого предложения).

В нашем эксперименте мы обратились к лингвистической традиции описания художественного текста и предложили информантам оценить тексты по следующим категориям: когерентность, смысловая связность, лексико-грамматическая связность, динамичность ритма повествования, присутствие субъективной модальности сомнения, эмоциональность, выразительность.

Обоснуем выбор категорий.

В немецкой школе анализа текстуальности (например, в работах Р.-А. Богранда и В. Дресслера) когерентность понимается как когнитивные взаимосвязи между компонентами текста, которые пытаются установить продуцент и реципиент текста (Филиппов, 2007: 127). Во французской традиции П. Шародо и Д. Мэнгно конкретизируют данную категорию: когерентность – это такой принцип организации текста, который позволяет читателю заключить, что последовательность частей текста не является произвольной (Charaudeau, Maingueneau, 2002; Charolles, 1978; Combettes, 1978). Представляется, что данный универсальный принцип также следует использовать для оценки синтетических художественных текстов.

Что касается когезии, то И. Р. Гальперин отмечает, что в художественном тексте настолько переплетены психологические, логические и формально-структурные средства когезии, что трудно дать их строгую типологическую характеристику (Гальперин, 2007: 74).

Мы выбрали два аспекта когезии: смысловую связность и лексико-грамматическую связность. Важность первого аспекта проистекает из того, что общий смысл текста выясняется адресатом постепенно, на основе накопления и анализа информации из его минимальных

смысловых частей, при этом логика рассуждения автора и его логика изложения могут не всегда совпадать (Матвеева, 1990: 23). Для оценки компетенций AI как писателя в задаче длинной генерации чрезвычайно важно оценить, сможет ли такой «синтетический» автор справиться с задачей нелинейной организации хода мысли.

Лексико-грамматическая связность является необходимой предпосылкой создания связности смысловой, но действует преимущественно на уровне микроконтекстов.

Следующая категория – это ритм повествования в художественном тексте, который охватывает разные его уровни и включает в себя: различного рода повторы, контрасты тех или иных мотивов, образов и ситуаций, закономерности сюжетного движения, соотношения различных композиционно-речевых единиц, определенный порядок в разворачивании системы образов-характеров и каждого из них в отдельности, мелодичность фразы, специфика построение сверхфразовых единств (Гиршман, 1974: 128). Нас интересовало, удастся ли в синтетических текстах увидеть движение ритма, его смену, обусловленную художественным замыслом.

В (Zhang et al., 2022) ставится проблема, связанная с тем, что БЯМ, в отличие от человека, “не осознают” ограниченности своих знаний, поэтому им не свойственно сомнение. Оттолкнувшись от данного тезиса, мы решили ввести в число текстовых категорий, подлежащих оцениванию, «присутствие субъективной модальности сомнения», подразумевая, что данная модальность может присутствовать как в авторском повествовании в случае гомодиегетического текста, так и в репликах героев или в их нарративах в случае гетеродиегетического текста.

Мы также включили в число оцениваемых категорий синтаксическую

структуру предложений и выразительность. Первую – поскольку при сложном синтаксисе БЯМ еще иногда генерируют слабо согласованные конструкции, а во вторую категорию мы включили оценку креативности и разнообразия стилистических приемов, использование идиом, которые традиционно считаются слабым местом моделей, а также степень выделенности сильных позиций текста⁷, в соответствии с концепцией стилистики декодирования (Арнольд, 2019).

Решено было также использовать категорию эмоциональности, поскольку «художественная литература отражает жизнь как картину мира в субъективно-эмоциональном преломлении конкретного автора» (Шаховский, 2010: 64). В этом отношении интересно было посмотреть, будет ли эмоциональность «выдавать» присутствие или отсутствие автора-человека.

4. Материал и методы психолингвистического эксперимента

Материал для эксперимента составили 6 текстов, 3 из которых были сгенерированы (в таблицах 1–2 имеют постфикс _gen, а затем на протяжении всего текста они употребляются под постоянными номерами – текст № 3, текст № 5, текст № 6), а еще 3 – написаны реальными авторами-англофонами, в течение нескольких лет пишущими в фанатских сообществах Джоан Роулинг. Объем каждого текста составил от 40000 до 45000 знаков без пробелов, а суммарный итоговый корпус составил 267 000 знаков. В таблице 1 представлены значения описательных статистик для данных текстов, а в таблице 2 – значения ряда метрик лингвистической сложности. Так, мы можем наблюдать, что один текст (№ 6) из 3 сгенерированных текстов имеет наибольшие в целом по выборке значения средней длины слова и среднего количества слогов в слове, при этом все сгенерированные тексты имеют большие значения средней длины предложения в

Таблица 1. Описательная статистика по текстам
Table 1. Descriptive statistics for the texts

№ Текста	Средняя длина слова	Средняя длина предложения	Общая длина в предложениях	Среднее количество слогов на слово	Общая длина в словах
1	4.53	9.34	2503	1.55	23385
2	4.38	12.13	3233	1.5	39203
3_gen	4.44	15.31	3283	1.51	50179
4	4.52	11.97	3590	1.55	42954
5_gen	4.48	13.46	3040	1.51	40926
6_gen	5.98	14.63	3308	2.01	48278

⁷ Сильные позиции текста – это элементы, которые играют ключевую роль в раскрытии основной темы, иерархии образов и идей художественного произведения. Они служат ориентирами для формирования стратегии восприятия и предварительного прогнозирования. К сильным позициям текста И. В. Арнольд относит: эпиграф, заголовок, начало текста и его конец (Арнольд, 2019).

Метрика Флеша-Кинкейда рассчитывается для английского языка по формуле:

$$0.39 \times (\text{общее кол} - \text{во слов} / \text{общее кол} - \text{во предложений}) + 11.8 \times (\text{общее кол} - \text{во слогов} / \text{общее кол} - \text{во слов}) - 15.59$$
 [23].

По данной метрике значения всех сгенерированных текстов в разной степени превышают значения для написанных текстов. Наибольшие отличия обнаруживает текст № 6.

Метрика MATTR (Moving-Avantage Type-Token-Ratio) высчитывается по формуле:

$$\frac{\text{количество уникальных слов} / \text{общее количество слов}}{A} \times 100\%$$

$$\frac{F}{A} * 100\%$$

где F – количество уникальных слов, A – общее количество слов.

Однако делается это на отрезках установленной исследователем длины (у нас это – 100 слов), а затем после

обработки всех сегментов значения TTR усредняются.

Отмечаем (таблица 2), что у двух из трех сгенерированных текстов значение метрики меньше, чем у текстов естественных, то есть тексты 3 и 5 содержат меньше, чем написанные тексты 1, 2, 4, уникальных лемм. Исключение составляет текст № 6, для которого метрика лексического разнообразия имеет самые высокие значения в выборке. Нарав_legomena отражает количество слов в тексте, встречающихся единожды, поэтому логично, что распределение ее значений в целом повторяет распределение значений MATTR.

Таблица 2. Значения метрик сложности текстов

Table 2. Values of text complexity metrics

№ Текста	Индекс Флеша-Кинкейда	MATTR	harax_legomena
1	6.31	0.143	1622
2	6.79	0.129	2627
3_gen	8.21	0.083	1739
4	7.39	0.143	3250
5_gen	7.5	0.097	1787
6_gen	13.84	0.184	3790

В качестве основного метода мы использовали метод психолингвистического эксперимента, в котором информантам (N=18) предлагалось прочитать текст и оценить его по разным категориям, используя шкалы от 1 до 5, а также высказать свою точку зрения относительно текста в свободной форме или пересказать его. Данные об информантах представлены в таблице 3: в выборке доминируют

женщины, средний возраст информантов 20 лет и шесть месяцев, подавляющее большинство владеет английским языком на уровнях B1-B2, по регулярности чтения длинных текстов на английском языке информанты разделились на три примерно равные группы – читаю 1 раз в неделю/ в месяц/ в полгода. Все информанты являются студентами 3 курса ОП “Филология” НИУ ВШЭ – Санкт-Петербург.

Таблица 3. Общая характеристика информантов

Table 3. General characteristics of the informants

Гендер %			Средний возраст		Уровень английского языка %			Регулярность чтения на английском языке		
м	ж	другое	STD	Mean age	B1	B2	C1	1 раз в неделю	1 раз в месяц	1 раз в полгода
22.2	72.2	7.6	2.13	20.6	55.56	38.89	5.56	33.3	33.3	34.4

В приложении Google Forms был создан опросник, структура которого представлена ниже (таблица 4). Инструкция сформулирована таким образом, чтобы информант мог оценить категории лингвистики текста, взятые нами за основу категорий оценки, в том числе без опоры на профессиональные филологические знания – предполагалось, что необходимым и достаточным является наличие читательского опыта. Кроме того, с информантами была проведена тренировочная сессия разметки, в ходе которой на других текстах обсуждались критерии и особенности оценивания. Опросник предваряла следующая общая инструкция:

Дорогие друзья, спасибо, что продолжили участвовать в нашем исследовании. Вам на электронные почтовые ящики были отправлены три (пока, возможно потом добавится еще два) текста. Это любительские прозаические тексты по мотивам «Гарри Поттера». Ваша итоговая задача – оценить естественность данных текстов по тем шкалам, которые вы предложили

сами, а мы просто отобрали нужно и структурировали. Вот несколько общих правил, которыми мы предлагаем вам руководствоваться: 1) сначала ознакомьтесь со шкалами нашего опросника (он будет состоять из четырех разделов – по одному разделу для каждого текста, в каждом разделе шкалы будут повторяться, но начнете вы с так называемой «паспорттики», т.е. общих сведениях об информанте); 2) затем внимательно прочитайте по частям или единым рывком первый текст и оцените его в первом разделе; 3) аналогичным образом поступите со следующими двумя текстами; 4) если есть необходимость, вы можете вернуться к каждому из разделов и изменить оценки; 5) после каждого раздела мы будем предлагать вам в открытой форме записать какие-то свои рефлексии на тему. Ну, давайте начнем. Сначала небольшой социолингвистический блок, чтобы мы знали о вас немного больше, чем в первом опросе, на электронную почту по-прежнему нужно добавить для отчетности...C'est la vie

Таблица 4. Структура опросника

Table 4. Questionnaire structure

Блок	Вопросы
0. Паспорттика	1. Сколько Вам лет? 2. К какому гендеру Вы себя относите? 3. На каком уровне Вы знаете английский язык? 4. С какой периодичностью Вы читаете длинные тексты (более 25 страниц) на английском языке? (никогда не читаю, раз в год, раз в полгода, раз в месяц, раз в неделю, каждый день)
1. Когерентность	1.1. Оцените смысловое соответствие названия фанфика дальнейшему тексту по следующей шкале, где 1 – не соответствует совсем (вообще непонятно, почему так назвали), 3 – связано с общей темой текста, 5 – вскрывает глубинные смыслы текста, обладает метафоричностью;

	1.2. Насколько гармонично сочетаются названия глав и подглав с общим стилем текста и содержанием глав и подглав? Используйте шкалу ниже, где 1 – не чувствую ни гармонии, ни соответствия; 3 – кажется к общей теме названия подходят, но стилистика не та или названия замысловатые и вписываются в общий стиль, но как будто не совсем про то; 5 – чувствуется и общий индивидуальный стиль, и, посмотрев на названия, можно представить себе о чем пойдет речь и не ошибиться; 1.3. Оцените крепость художественной связи, которая существует между всеми элементами текста по шкале, где 1 – "повсюду развешаны ружья, которые не стреляют", 3 – большинство персонажей, действий и деталей взаимообуславливают существование друг друга, но есть какие-то странные вещи, которые, появившись, исчезают в никуда; 5 – все ружья выстреливают; 1.4. Оцените присутствие в тексте единой идеи по шкале, где 1 – я ее не уловил (а); 3 – какие-то обрывочные мысли вспыхивают, но объединить во что-то единое трудно; 5 – однозначно есть
2. Смысловая связность	2.1. Сможете ли вы пересказать основной сюжет текста? Оцените по шкале, где 1 – ни за что, там так все запутано и бессмысленно; 3 – я вообще плохо пересказываю, но попробую; 5 – конечно, перескажу и моим слушателям будет интересно, но не потому, что у меня талант рассказывать, а потому, что сюжет увлекательный; 2.2. Перескажите текст в десяти предложениях
3. Лексико-грамматическая связность	3.1. Оцените характер повторов слов, форм в тексте по шкале, где 1 – повторы есть и они ничем не оправданы, просто как будто автор не знает, что такое синонимы; 3 – повторы встречаются и на лексическом, и на грамматическом уровне, но где-то кажется, что это для связности, а где-то думаешь, что это скудность словарного запаса; 5 – повторы служат либо обеспечению связности, либо имеют художественную нагруженность
4. Ритм	4.1. Оцените общий ритм развертывания сюжета, где 1 – все происходит всегда очень быстро или всегда очень медленно (ритм остается одним и тем же); 3 – ритм неровный, но объяснить почему и интерпретировать сложно; 5 – ритм меняется вместе с развертыванием сюжета от завязки к кульминации и развязке
5. Субъективность модальность сомнения	5.1. Оцените эмоциональность текста по шкале, где 1 – в тексте есть лишь одна из трех эмоциональных тональностей (негативная/позитивная/нейтральная); 3 – эмоции меняются, но как-то странно и резко; 5 – эмоции есть и выражаются они не только прямо (эмотивами "радостно", "грустно"), но и косвенно: через жесты, голос, мимику героев
6. Эмоциональность	6.1. Оцените эмоциональность текста по шкале, где 1 – в тексте есть лишь одна из трех эмоциональных тональностей (негативная/позитивная/нейтральная); 3 – эмоции меняются, но как-то странно и резко; 5 – эмоции есть и выражаются они не только прямо (эмотивами "радостно", "грустно"), но и косвенно: через жесты, голос, мимику героев
7. Синтаксис	7.1. Оцените синтаксис текста по шкале, где 1 – в тексте нет сложных оборотов (вставочных конструкций, придаточных разного типа, деепричастных оборотов, безличных или неопределенно-личных предложений, эллипсов); 3 – в тексте есть умеренное их использование; 5 – в тексте используются многие перечисленные сложные конструкции
8. Выразительность	8.1. Оцените использование идиом в текстах по следующей шкале, где 1 – идиомы в тексте не встречаются; 3 – встречаются, но употреблены в неуместных контекстах; 5 – встречаются и употреблены там, где нужно, обогащая смысл текста 8.2. Оцените, насколько выражены так называемые сильные позиции в тексте – его начало и конец. Используйте шкалу, в которой 1 – у текста ни

	начала, ни окончания; 3 – начало интригует, привлекает, а конец разочаровывает – мутный; 5 – это, действительно, две сильные позиции в тексте 8.3. Оцените уровень креативности текста по шкале, где 1 – в тексте нет никаких стилистических средств, кроме банальных сравнений; 3 – в тексте есть узуальные метафоры, сравнения, известные клишированные оксюмороны, но не более того; 5 – в тексте есть оригинальные приемы, чтобы привлечь внимание читателя: игра слов, литоты или гиперболы, авторские метафоры, встречаются окказионализмы
--	--

Вышеприведенная структура вопросов повторялась для каждого текста, иными словами, она была воспроизведена 6 раз.

Информантам давалось 14 дней для того, чтобы прочитать тексты и ответить на вопросы.

Общий план обработки результатов:

1) подсчет средних значений и стандартных отклонений значений, полученных по шкалам для каждого вопроса;

2) визуализация значений, полученных по каждому вопросу, с помощью диаграмм «скрипка», чтобы увидеть специфику распределения «сырых» значений;

3) векторизация с помощью модели из семейства SentenceTransformer и последующая кластеризация кратких пересказов содержания текстов, данных информантами (для каждого теста в отдельности). Цель данной операции – сравнить конфигурации распределения по кластерам для написанных и сгенерированных текстов.

При обработке результатов мы пользовались стандартными функциями Excel для вычисления средних значений и стандартного отклонения для значений, отмеченных информантами по шкалам, для подсчета характеристик лингвистической

сложности текстов использовалась библиотека ruTS⁸. Для кластеризации использовался пакет cluster из библиотеки sklearn, для анализа качества кластеров – пакет metrics из той же библиотеки.

5. Результаты

При ответе на вопрос 1.1 о соответствии названия текста дальнейшему сюжету, информанты выше всего оценили текст № 2 (естественный) – среднее значение 3.2 (таблица 5), наибольшее количество оценок поместилось в интервале от 2,5 до 4,5 (Рисунок 1). Автор назвал свой текст «The Magical Friendship of Liza Turpin».

Интересно, что особое положение среди естественных текстов занимает текст № 4 (Рисунок 1). Для этого текста получено самое низкое среднее значение (1,2 – таблица 5, п. 1.1) по тому же критерию при наименьшем стандартном отклонении (0,9 – таблица 5, п. 1.1). Среди сгенерированных текстов наибольшее среднее значение для вопроса 1.1 имеет текст № 6. Тем не менее, диаграмма (Рисунок 1) демонстрирует, что «скрипка» для текста № 6 имеет очень вытянутую форму, что говорит о наличии совершенно разных мнений у информантов, а небольшое утолщение наблюдаем в районе единицы – в данной зоне присутствует скопление ответов.

8

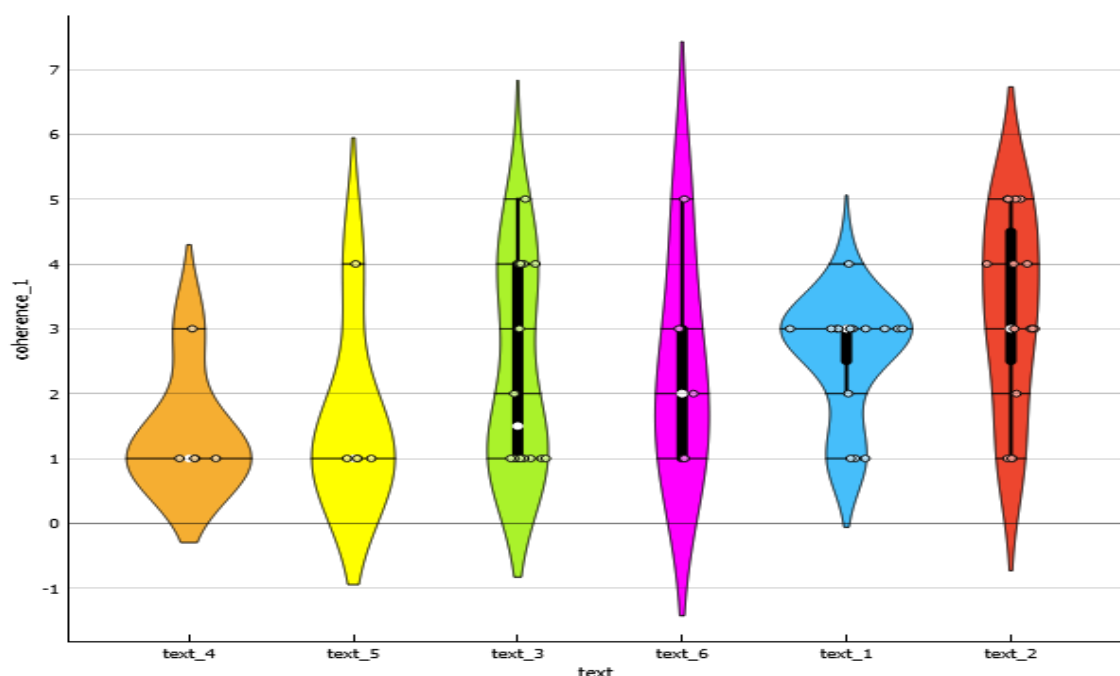
<https://github.com/SergeyShk/ruTS?ysclid=mpjsjevttoy50356446>

Таблица 5. Средние значения и стандартное отклонение в категории «Когерентность»
Table 5. Mean values and standard deviation in the category «Coherence»

Номер вопроса из категории «Когерентность»/ номер текста		текст 1	текст 2	текст 3_gen	текст 4	текст 5_gen	текст 6_gen
1.1	mean	2,6	3,2	2,3	1,2	1,6	2,4
	std	0,9	1,5	1,5	0,9	1,3	1,6
1.2	mean	2,9	3,9	2,8	2	3,4	3,5
	std	1,2	1,4	1,1	1,4	1,1	0,9
1.3	mean	2,9	3,5	2,1	3,5	1,9	2,8
	std	1,2	1,2	1,1	0,9	1,1	1,5
1.4	mean	3,8	4	2,3	3,9	2,2	3
	std	1,3	1,4	1,4	1,1	1,3	1,5

У текста № 3 также видим скопление оценок в зоне 1 и менее значительное – в

зоне 4. Текст № 5 имеет наибольшее количество оценок в зоне 1.

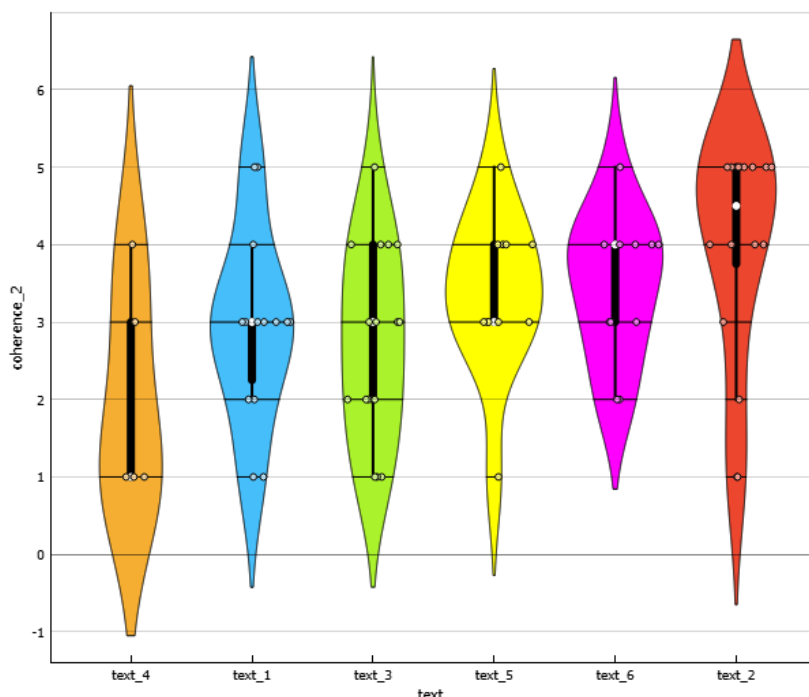
Рисунок 1. Распределение значений (вопрос 1.1.) в категории «Когерентность»
Figure 1. Distribution of values for question 1.1. in the category «Coherence»

Оценивая связь названий глав и подглав с содержанием текста (п.1.2), информанты вновь отдали предпочтение естественному тексту № 2, но другие два естественных текста, 4 и 1, оказались оценены по данному критерию низко (текст 1 в диапазоне 1–3, текст 4 в

диапазоне 2–3) (Рисунок 2). При этом оценки сгенерированных текстов по данному параметру имеют наименьшее в выборке стандартное отклонение (таблица 4) и сконцентрированы для текстов 6 и 5 в верхней части шкалы (Рисунок 2).

Рисунок 2. Распределение значений (вопрос 1.2.) в категории «Когерентность»

Figure 2. Distribution of values for question 1.2. in the category «Coherence»

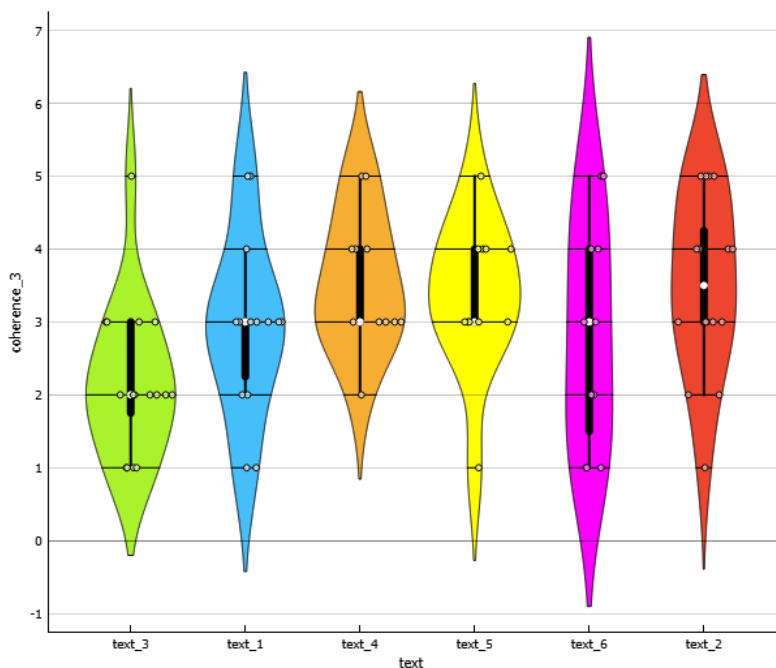


Оценивая в целом силу связи между всеми элементами художественного текста, информанты по-прежнему высоко оценили естественный текст № 2, по средним значениям (таблица 3) за ним следуют тексты № 4 и № 1 (все три относятся к

категории естественных), но сгенерированный текст № 5 имеет наибольшее скопление оценок в интервале от 3 до 4 (Рисунок 3), поэтому он по данному критерию, в общем, близок к тексту № 4.

Рисунок 3. Распределение значений (вопрос 1.3.) в категории «Когерентность»

Figure 3. Distribution of values for question 1.3. in the category «Coherence»

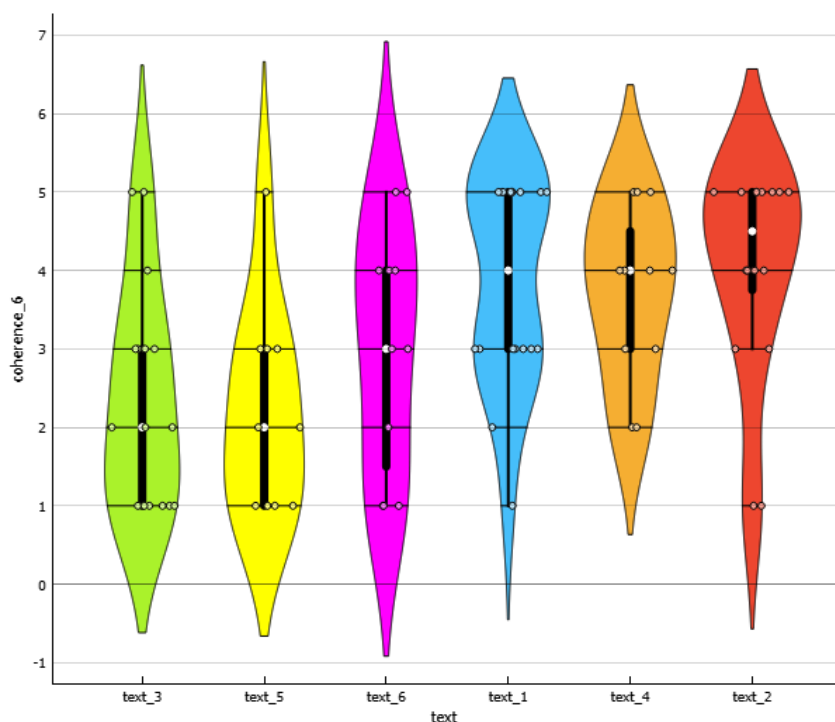


Оценивая степень выраженности в тексте единой идеи, информанты выше всего оценили естественные тексты 2, 4

и 1 – это заметно как по средним значениям (таблица 3), так и по общему распределению значений (см. «скрипки» на Рисунке 4).

Рисунок 4. Распределение значений (вопрос 1.4.) в категории «Когерентность»

Figure 4. Distribution of values for question 1.4. in the category «Coherence»



По шкале смысловой связности (п. 2.1) наибольшие средние значения имеют тексты № 2 и № 1 из числа

естественных, и текст № 6 – из числа сгенерированных (таблица 6).

Таблица 6. Средние значения и стандартное отклонение в категориях «Смысловая связность», «Лексико-грамматическая связность», «Ритм», «Сомнение», «Эмоциональность», «Синтаксис».

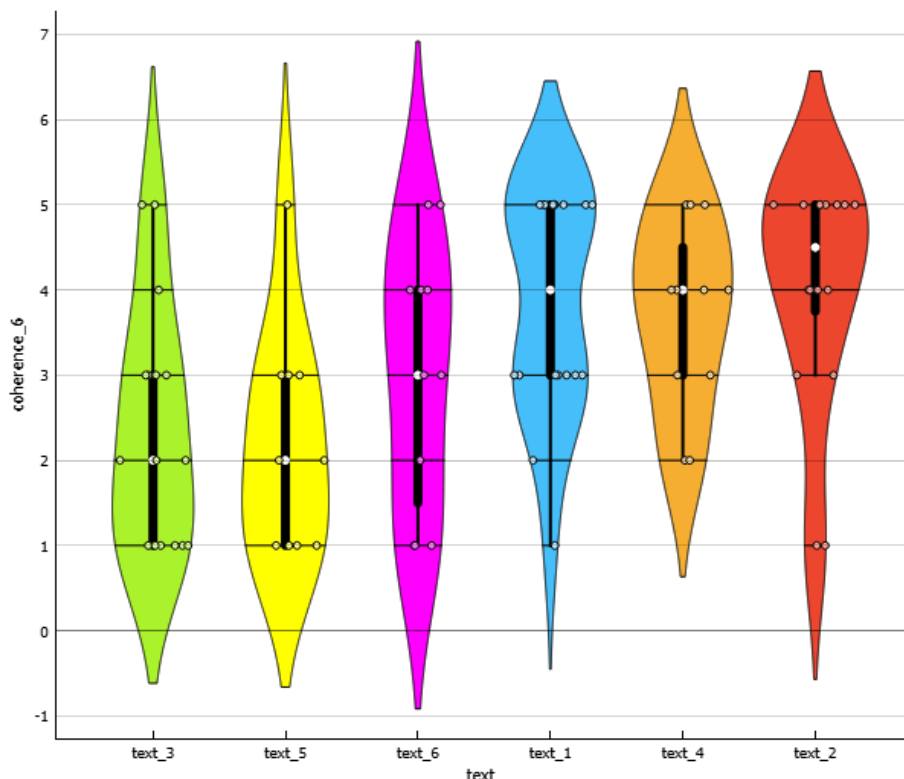
Table 6. Mean values and standard deviation in the categories “Semantic cohesion”, “Lexico-grammatical cohesion”, “Rhythm”, “Doubt”, “Emotionality”, “Syntax”

Категория/ текст	Тип значения	текст 1	текст 2	текст 3_gen	текст 4	текст 5_gen	текст 6_gen
Смысловая связность	mean	3,5	3,6	1,9	2,8	1,9	2,9
	std	1,1	1,3	1,2	0,9	1,1	1,3
Лексико-грамматическая связность	mean	3,8	3,8	2,5	4	2,6	3,4
	std	1,2	1,2	1,3	0,9	1,1	1,3
Ритм	mean	2,7	3,3	2,6	3,5	2,2	2,6
	std	1,4	1,4	1,3	1,2	1,2	1,6
Сомнение	mean	3,2	3,4	2,2	3,6	2	3,1
	std	0,9	1,1	1,1	0,9	1,1	1,1
Эмоциональность	mean	3,9	3,3	2,6	3,9	2,4	3,5
	std	1,1	1,5	1,1	1,1	1,1	1,5
Синтаксис	mean	3,6	3,5	2,8	3,6	2,8	3,8
	std	0,9	1,1	1,1	0,9	1,1	1,1

Однако если посмотреть на распределение “сырых” значений, текст № 6 окажется только на четвертом месте, после естественных текстов 1, 2, 4,

поскольку вытянутая форма “скрипки” говорит о присутствии полярных оценок и общей высокой несогласованности оценок (Рисунок 5).

Рисунок 5. Распределение значений (вопрос 2.1.) в категории «Смысловая связность»
Figure 5. Distribution of values for question 2.1. in the category “Semantic cohesion”



В задании 2.2. мы просили информантов пересказать текст в 10-ти предложениях. Соблюсти указанный объем высказывания удалось не всем. Наша гипотеза состояла в том, что если мы векторизуем тексты, а затем кластеризуем их, то пересказы более связных текстов с большей вероятностью попадут в один кластер.

Для построения подобной гипотезы у нас было несколько оснований. Согласно теории построения ментальных моделей (Johnson-Laird, 1983) связный текст требует меньше усилий для удержания в рабочей памяти, поэтому его легче пересказывать. В частности, это происходит потому, что риторическая структура (Mann and Thompson, 1988) задает явные логические связи между частями текста. При пересказе эти связи

выступают как жесткие «дискурсивные якоря», заставляя разных людей воспроизводить одни и те же смысловые паттерны. Напротив, в несвязном тексте глобальная когеренция (Graesser et al., 1994) ослаблена: слушатель вынужден самостоятельно достраивать пропущенные отношения, и каждый делает это по-своему.

Для векторизации использовалась модель семейства SentenceTransformer all-MiniLM-L12-v2⁹.

На рисунке 6 представлен результат кластеризации методом k-means на основе косинусного расстояния между векторами текстов (значение perplexity = 50, поскольку набор данных был небольшим).

⁹ <https://huggingface.co/sentence-transformers/all-MiniLM-L12-v2>

Для визуализации использовался также метод снижения размерности векторов t-SNE. Результаты кластеризации были перепроверены с помощью метода Уорда. Используемый нами метод Уорда (Ward's linkage) является стандартным подходом в иерархической кластеризации для задач, где важна компактность кластеров. В отличие от методов одиночной связи, которые склонны к «эффекту цепочки», или полной связи, чувствительной к выбросам, метод Уорда минимизирует

внутрикластерную дисперсию и дает более интерпретируемые результаты на выборках малого объема. В обоих случаях посчитаны метрики чистоты кластера (доля пересказов, попавших в «свой» кластер – чем выше процент, тем лучше), его компактности (среднее расстояние эмбединга текста до центроида класса – чем ниже значение, тем лучше) и отдельности от других кластеров (коэффициент силуэта – чем выше значение, тем более отдельный кластер).

Таблица 7. Метрики качества кластеризации для методов Уорда и k-means.

Table 7. Clustering quality metrics for Ward's method and k-means

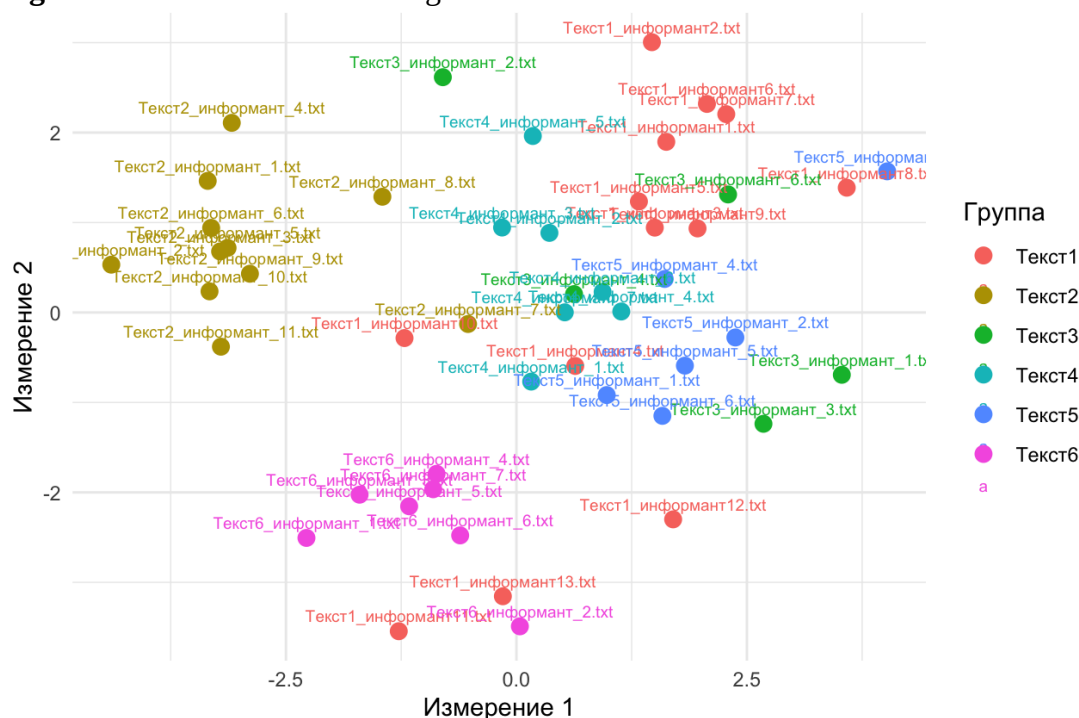
Текст	Кол-во пересказов	Ward_силуэт	Ward_чистота %	Ward_компактность	KMeans_силуэт	KMeans_чистота %	KMeans_компактность
1	13	0.206	62	0.352	0.194	62	0.355
2	11	0.196	91	0.295	0.196	91	0.295
3	10	0.210	30	0.431	0.210	30	0.431
4	7	0.078	100	0.307	0.062	86	0.311
5	6	0.046	83	0.334	0.045	67	0.336
6	7	0.096	100	0.313	0.098	100	0.313

По данным таблицы (табл. 7) мы видим, что кластеров с точностью 100 % (метрика «чистота» (Purity)) при использовании обоих методов кластеризации два – это кластер пересказов текста № 6 и кластер пересказов текста № 4. Наименьшее среднее расстояние до центроида своего кластера демонстрируют (в порядке ухудшения результатов) тексты 2

(естественный), 4 (естественный) и 6 (сгенерированный). Значения коэффициента силуэта обнаруживают зависимость от количества объектов кластеризации (чем меньше объектов, тем ниже значение коэффициента), поэтому их обсуждение пока не релевантно. Значения, зафиксированные в таблице, демонстрируют, что метрики практически не зависят от способа кластеризации.

Рисунок 6. Кластеры пересказов текстов

Figure 6. Clusters of text retellings



Приведем примеры пересказов текстов № 2 и № 6:

- (1) Текст № 2: *Лиза Турпин живет со своей матерью, которая называет себя ведьмой. Про отца она знает только из ее рассказов, он с ними не живет. Лиза получает письмо из школы Хогвартс, про которую до этого никогда не слышала. Мать радуется, что девочка будет изучать магию там. Они собирают все необходимое для школы, но они бедны, поэтому ищут вещи не в Косой аллее, а в других магловских магазинах. В Хогвартсе девочку распределяют на факультет Когтевран. В школе ей приходится нелегко, но она старается. Лиза изучает разные предметы. Ей сложно завести дружбу с другими учениками. Но в конце она решает, что Теодор, Сьюзан, Дафна и Невилл были бы хорошими друзьями для нее.*
- (2) Текст №6: *Теодор Нотт – мальчик с почти что синдромом отличника и чувством долга перед семьей. Панси*

Паркинсон – другая миссис перфект с таким же распланированным будущим. они не общаются какое-то время, но после начинают сближаться. узнают, что у них, так же как и у их семей, есть много общего и загадочного. они влюбляются друг в друга и решают создать в конце совместное будущее. конец.

Для сравнения рассмотрим пересказ естественного текста № 3 от одного из информантов:

- (3) Текст № 3: *Сначала Гарри получает письмо в Хогвартс, но в следующей главе он уже женат на Джинни и у него есть дети. Потом он опять школьник и в Министерстве магии вместе с Гермионой и Роном ищут пророчество, потом они уже в Годриковой впадине, потом у Лавгудов, потом опять какие-то приключения. Потом они предупреждают Авроров в Министерстве о Пожирателях смерти. Затем снова куча*

неразберихи и приключений. Герои пересекаются с многими другими персонажами. В конце концов Гарри работает профессором в Хогвартсе, Гермиона – Аврор, а у Рона паб в Лондоне.

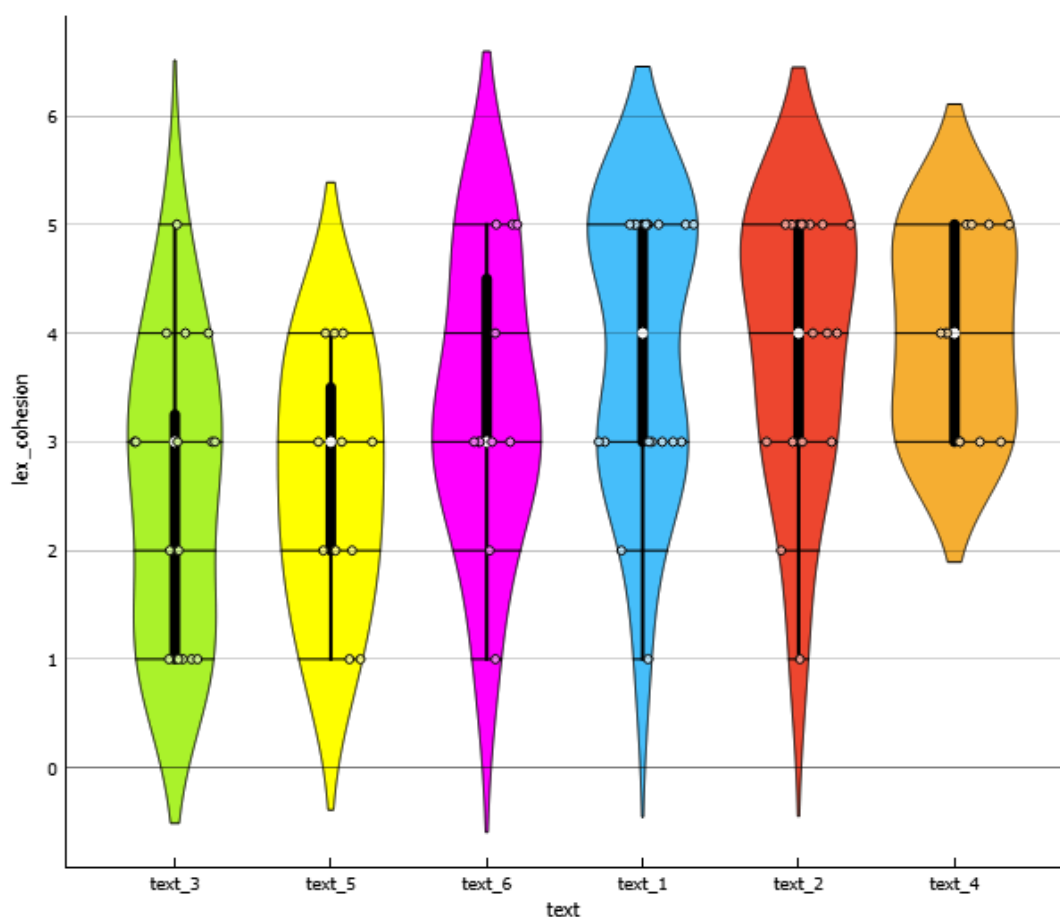
Можно увидеть, что пересказы текстов 2 и 6 сближает то, что информантам удастся не только связать прочитанное в единый сюжет, но и определить причинно-следственные связи (они бедны, поэтому ищут вещи не в Косой аллее, а в других магловских магазинах), сформировать «образ прошлого» (Про отца она знает только из ее рассказов, он с ними не живет; у их семей много общего и загадочного) и «образ будущего» героев (

и решают создать в конце совместное будущее). В пересказе текста № 3, имеющего худшие значения чистоты и компактности, наблюдается лишь попытка зафиксировать сюжет как цепочку событий, соединяя их посредством самого распространенного маркера последовательности *потом*.

Что касается лексико-грамматической связности, то лидерами как по средним значениям (таблица 4), так и по интервалам распределения значений (Рисунок 7) стали тексты 4, 2, 1 – все из категории написанных людьми. У сгенерированного текста № 6 наблюдаем наибольшее скопление оценок в зоне 3-х.

Рисунок 7. Распределение значений оценок (вопрос 3.1.) в категории «Лексико-грамматическая связность»

Figure 7. Distribution of values for question 3.1. in the category “Lexico-grammatical cohesion”



Аналогичное распределение «лидеров» по средним значениям оценки

наблюдаем и в категориях «Ритм», «Субъективная модальность сомнения».

Однако по шкале эмоциональности информанты оценили текст № 6 достаточно высоко, что позволило ему выйти на третье место (таблица 6) – среднее значение по шкале эмоциональности – 3,5.

По шкале синтаксической сложности текст № 6 имеет самые высокие средние значения оценки (среднее значение=3,9), а распределение “сырых” значений тяготеет к верхней части шкалы. Интересно то, что при этом на втором и третьем местах по параметру синтаксической сложности оказались естественные тексты № 2 и № 4 (Рисунок 8).

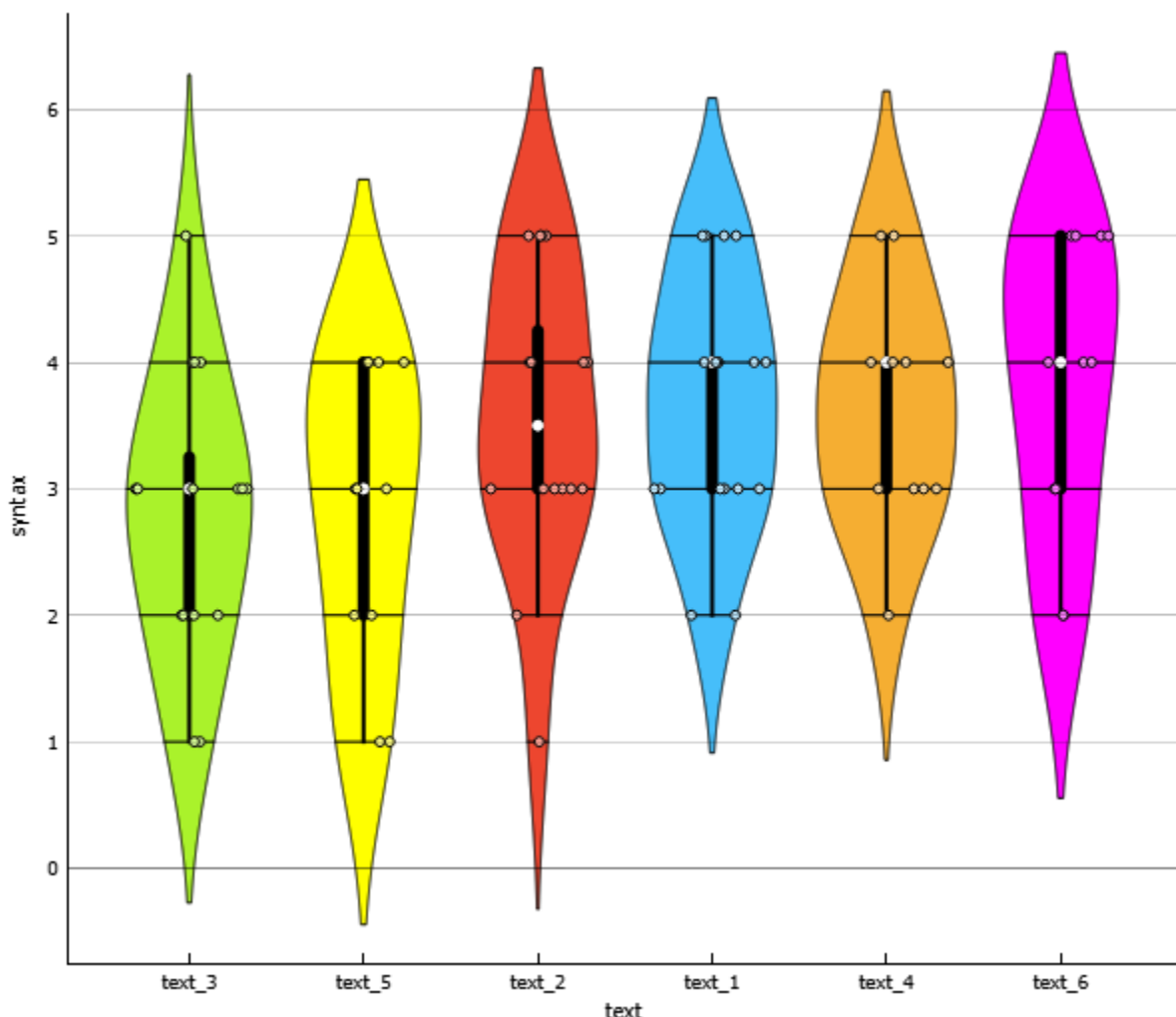
Синтаксическая канва текста № 6, действительно, содержит причастные

обороты (*dedicated to joint endeavors*, пр.4), обособленные конструкции (*their disparate focal points notwithstanding*, пр.4), перечисления (*collaborative research initiatives, interdepartmental taskforces tackling complex threats etc.*) и т.д., например:

(4) *Their disparate focal points notwithstanding, threads of convergence emerged as they pored over adjacent segments dedicated to joint endeavors: collaborative research initiatives, interdepartmental taskforces tackling complex threats, and even fledgling organizations founded by enterprising alumni addressing pressing issues plaguing the wizarding realm.*

Рисунок 8. Распределение значений (вопрос 7.1.) в категории «Синтаксис»

Figure 8. Distribution of values for question 7.1. in the category “Syntax”



Рассмотрим оценки текстов по параметру выразительности.

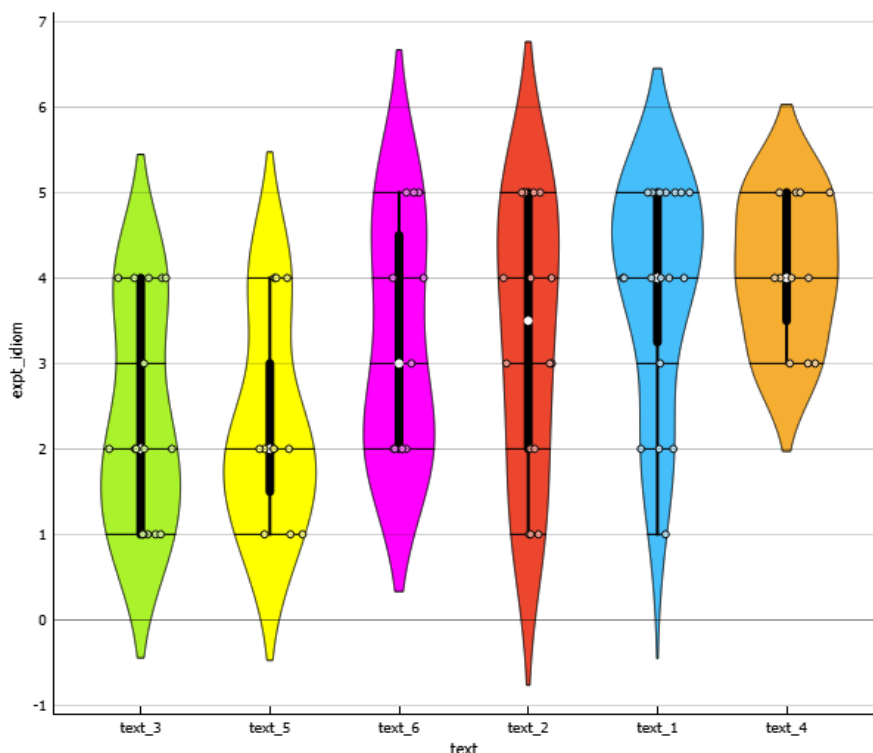
Таблица 7. Средние значения и стандартное отклонение в категории «Выразительность»
Table 7. Mean values and standard deviation in the category “Expressiveness”

Подкатегория/текст	Тип значения	текст 1	текст 2	текст 3_gen	текст 4	текст 5_gen	текст 6_gen
8.1 Идиоматичность	mean	3,8	3,3	2,3	4,1	2,3	3,2
	std	1,1	1,3	1,2	0,9	1,1	1,3
8.2 Сильные позиции	mean	3,3	3,3	2,5	3,5	2,7	3,4
	std	0,9	1,2	1,3	0,9	1,5	1,5
8.3 Креативность	mean	3,2	3,3	2,3	3,6	1,9	3,5
	std	1,3	1,5	1,3	0,8	1,2	1,3

По количеству идиоматических выражений сгенерированный текст № 6 незначительно уступает трем написанным текстам (Рисунок 9, Таблица 7). На диаграмме (Рисунок 9) видно, что при

сопоставимом с естественными текстами среднем значении, тем не менее, значительное скопление не усредненных оценок для текста № 6 наблюдаем в районе 4-х.

Рисунок 9. Распределение значений оценок (вопрос 8.1.) в категории «Выразительность»
Figure 9. Distribution of values for question 8.1. in the category “Expressiveness”



Такая достаточно высокая оценка не случайна – в тексте находим более 100 идиоматических выражений, например: *walking on eggshells* ‘вести себя крайне осторожно в деликатной ситуации’,

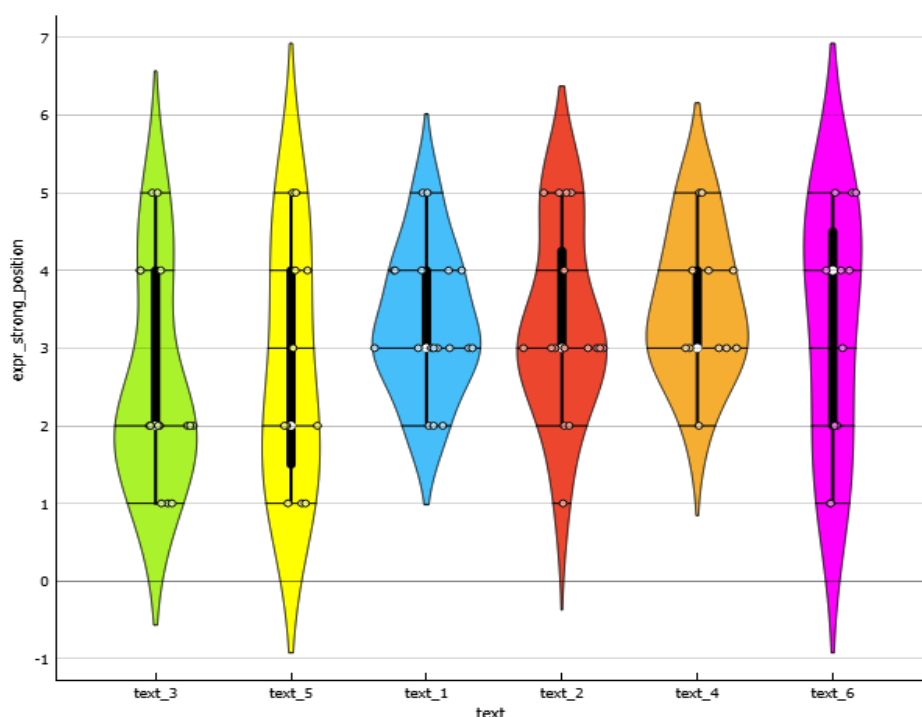
walking a tightrope ‘балансировать в сложной ситуации’, *read between the lines* ‘видеть скрытые смыслы’, *burn bridges* ‘разорвать всякие отношения’, *skeletons in the closet* ‘семейные тайны’ и т.д.

По наличию же так называемых «Сильных позиций» по средним значениям текст № 6 оказался на второй позиции после естественного текста № 4, а по общему распределению не усредненных

значений – на первой (Рисунок 10), поскольку демонстрирует незначительное «утолщение» в зоне 4-х, а максимальные значения приближаются к 5.

Рисунок 10. Распределение значений на вопрос 8.2. в категории «Выразительность»

Figure 10. Distribution of values for question 8.2. in the category “Expressiveness”



Действительно, в тексте № 6, например, в финале истории (пр.5) можем наблюдать, как усиливается эмоциональное напряжение героев, при этом в тексте концентрируются метафоры (*McGonagall's melodious voice ushered in an era of transformation; conversations mushroomed; parallel fascinations ignited Pansy's enthusiasm; recalibrating his internal ledger of favors owed and debts outstanding; timely investments in influential relationships* etc.) и метафорические сравнения (*fanning out like tender shoots bursting forth from fertile soil*), что производит конвергентный эффект и усиливает производимое на читателя впечатление:

(5) *Moments later, McGonagall's melodious voice ushered in an era of transformation, announcing the commencement of specialized coursework catering to individual strengths.*

Conversations mushroomed, exploring the vast spectrum of electives and concentrations unveiled before them. Theo listened intently, his pulse quickening as Ancient Runes and Archaeological Magic topped his list of preferred subjects. Parallel fascinations ignited Pansy's enthusiasm, her affinity for Curse-Breaking and Dark Arts Countermeasures aligning harmoniously with his own pursuits.

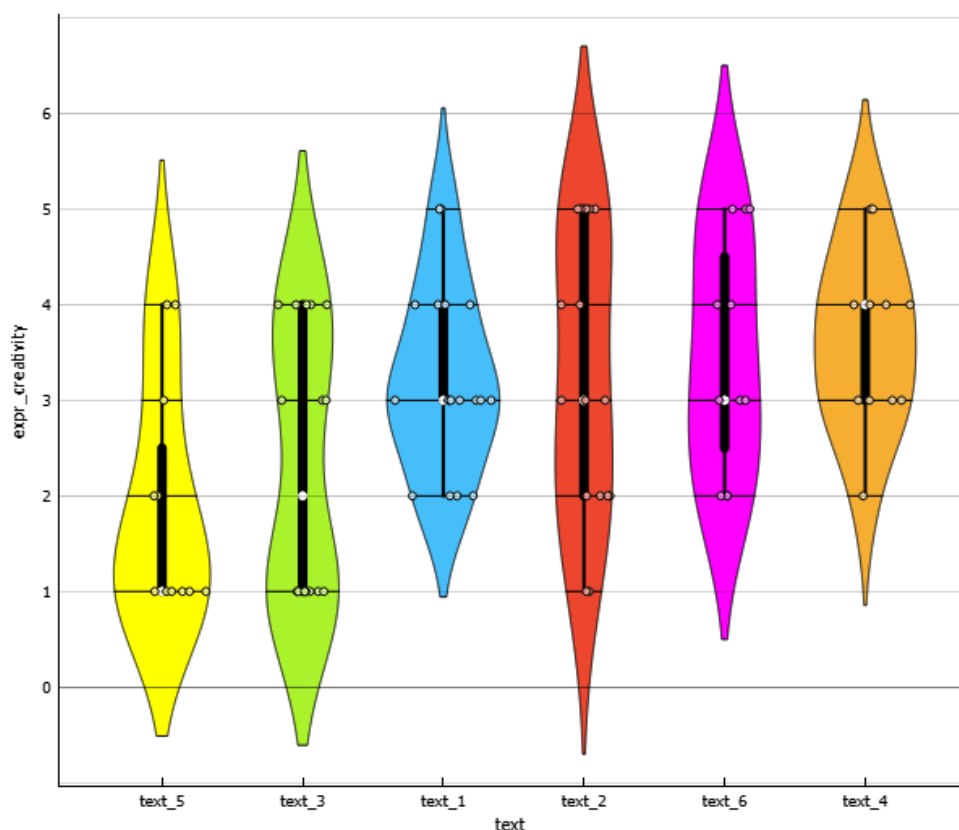
Through the cacophony of congratulatory cheers and sympathetic condolences, Blaise Zabini maintained a calculated distance, monitoring shifting allegiances and recalibrating his internal ledger of favors owed and debts outstanding. He knew full well that timely investments in influential relationships often proved decisive in securing

advantageous positions within the wizarding hierarchy.

Nearby, Ron Weasley pumped Hermione's hand aloft, basking in the glory of her triumphant entry into elite Magical Law Enforcement internship programs. Michael Corner chuckled knowingly, attributing her success to razor-sharp intellect and diligent preparation. Friendship warmed the atmosphere, reinforcing bonds tested by communal struggles and victories.

Discreetly withdrawing from the melee, Theo and Pansy commandeered a vacant study area, immersing themselves in meticulous analyses of recommended curricula. Visions of interdisciplinary collaboration sprouted, fanning out like tender shoots bursting forth from fertile soil. Convergent paths merged seamlessly, forecasting trail-blazing innovations primed to revolutionize the wizarding community.

Рисунок 11. Распределение значений на вопрос 8.3. в категории «Выразительность»
Figure 11. Distribution of values for question 8.3. in the category “Expressiveness”



Что касается креативности, то текст № 4 оценивается выше других, демонстрируя скопление оценок в районе 4-х, текст 1 – в районе 3-х, а тексты 2 и 6, несмотря на разный источник продуцирования (человек – для текста № 2 и модель – для текста № 6), сходный паттерн оценивания – низкую согласованность оценок, на что указывает вытянутая форма «скрипок» (Рисунок 11).

6. Обсуждение

Поскольку в экспериментах с БЯМ команды компьютерных лингвистов преследовали цель добиться иерархичности повествования, используя при этом разные техники, в категории «Когерентность» мы наблюдаем наименьший разрыв между написанными и сгенерированными текстами. Более значительные отличия видятся там, где речь идет о связности, ритме, эмоциях и модальности сомнения – сгенерированные

тексты оцениваются по данным критериям ниже, чем написанные. В категориях же сложности синтаксиса и выразительности различия вновь смягчаются, поскольку, например, сгенерированный текст № 6 оказался в числе лидеров и по сложности синтаксических структур, и по выделенности сильных позиций, и по креативности стиля. С данными наблюдениями отчасти согласуются и формальные метрики, предварительно измеренные для всех текстов – текст № 6 оказался самым сложным по метрике Флеша-Кинкейда, с самым высоким числом уникальных слов и самыми длинными словами и предложениями.

Таким образом, именно текст № 6, сгенерированный БЯМ после дообучения пересказу на датасете художественной литературы из проекта Гуттенберг, оказался наиболее близок по большинству метрик к естественным текстам. Представляется важным, что именно этот текст показал наиболее четкий из всех сгенерированных текстов кластер векторов пересказов, что свидетельствует о его приближении к паттернам связности. Текст № 6 также продемонстрировал хороший уровень генерации эмотивности в тексте для создания определенной эмоциональной тональности, ср. фрагмент из текста:

(6) *Through the quiet intensity of their partnership, boundaries softened. Conversations flowed freely, touching on everything from favored Quidditch matches to childhood memories. Shared laughs erased residual awkwardness, allowing them to settle comfortably into an easy camaraderie. Without realizing it, they discovered solace in each other's company, finding strength in their differences rather than weaknesses.*

Синтетическому автору удастся воссоздать позитивную эмоциональность, эмпатически передать читателю атмосферу комфортного дружеского общения:

границы стирались, разговоры текли свободно, общий смех растворял остающуюся неловкость, дружеская близость (пр.6).

Однако, главный вывод, который мы можем сделать, на наш взгляд, состоит в том, что первичная посылка организаторов соревнования о том, что все дело в иерархичности нарративной схемы, в общем, не оправдала себя. Более важными свойствами художественного текста любительской прозы оказались наличие динамичного меняющегося ритма повествования, его подчиненность логике причинно-следственных связей, присутствие сомнения, выраженного модальными языковыми средствами либо оценочными лексемами. Синтетическим авторам пока не удастся создать не только сюжет, но и микромир, атмосферу повествования. Работая с разнообразными стилистическими приемами, модели, тем не менее, пока не могут предвосхитить ожидания читателей и, например, обмануть их, что заставляет читателей трактовать сюжет как “скучный”, не обнаруживающий психологизма:

(7) *Сложные предложения, нагруженные непростыми словами, не несут за собой художественной ценности, а лишь создают шум вокруг скучного сюжета, который не до конца ясен. Частые переходы в повествовании от одного персонажа к другому не позволяют погрузиться в сюжет и представляют собой мимоходное повествование, в котором отсутствует психологизм*
(комментарий респондента по тексту № 6).

Справедливости ради отметим, что аналогичные упреки артикулировались информантами и в отношении естественных текстов 1, 2, 4, написанных авторами-любителями. В этом смысле, несомненно, стоит принимать во внимание тот факт, что сам жанр любительской прозы по мотивам известного

произведения не позволяет в полной мере предъявлять к таким текстам требования эстетической ценности и психологической глубины.

Таким образом, если мы говорим о некотором каноне восприятия художественного текста нашими информантами, то проведенное пилотное исследование демонстрирует, что когерентность и когезия, например, находятся среди периферийных признаков “настоящего” художественного текста для наших информантов. Центральное же место, по-видимому, занимает ритм, легкое переключение между жанрами и нарративными позициями, эмпатичность и эмоциональность повествования, свидетельствующая о присутствии нарратора. Приведем фрагмент комментария одного из информантов к тексту № 4, который по результатам оценивания имел низкие баллы за параметры, связанные с когерентностью и связанностью, но высокие – за ритм, эмоциональность и выразительность:

(8) *В тексте частое подробное описание вещей, людей, создаётся «естественное» ощущение рассказа от человека с детальным погружением. Обозначение громкости и долготы речи также способствует созданию естественности. Резкий переход с сюжетного повествования к эпистолярному также указывает на человеческий фактор. Также стоит заметить свободное переключение от общего фона повествования в третьем лице к монологическому без потери логичности, что очеловечивает наррацию.*

Методологический вклад результатов проведенного на уникальных данных эксперимента состоит в том, что они демонстрируют необходимость введения новых метрик для оценки качества длинной генерации художественных текстов, часть из которых была нами апробирована – это ритм, креативность, модальность сомнения и др.. Сложившийся в среде исследователей в

области компьютерных наук перечень существующих метрик (мы его привели в теоретической части статьи) демонстрирует свою ограниченность в силу фиксации на четкости логических отношений между событиями и персонажами, между разными уровнями схемы повествования. При этом за рамками исследований остаются собственно филологические и психологические аспекты рецепции текста, хорошо изученные лингвистами на материале естественных текстов как в теоретическом, так и в экспериментальном ключе. Прием, связанный с кластеризацией пересказов текстов как показателем целостности и связности, представляется нам перспективным. Дальнейшие наши эксперименты, возможно, будут связаны с его валидацией.

Несомненно, исследование имеет ряд значительных ограничений. Во-первых, достаточно специфический и узкий круг информантов, принимавших участие в эксперименте. Однако это связано с большой хронозатратностью задания: информанту нужно было прочитать около 100 страниц текста на неродном языке, а затем ответить на опросник по каждому тексту отдельно, написать комментарии и пересказы. В данных есть пробелы – например, не все информанты написали пересказы для всех текстов, поэтому наблюдается неравномерность в их распределении, что важно, например, для качества кластеризации. Мы намеренно приняли решение пригласить в качестве информантов студентов-филологов как опытных читателей, привыкших внимательно относиться к слову. Очевидно, что привлечение к эксперименту другой группы информантов привело бы к получению несколько иных результатов.

7. Заключение

Проведенные эксперименты показали, что благодаря специальным методам, сочетающим в себе промптинг и дообучение, можно решить задачу

автоматической генерации длинного текста с развернутой иерархически организованной нарративной структурой.

В то же время, автоматическая генерация пока не позволяет получить тексты, отвечающие критериям наличия динамически меняющегося ритма повествования и развертывания сюжета; таким синтетическим текстам пока не свойственна субъективная модальность сомнения.

Однако применение метода дообучения модели для задачи пересказа художественных произведений позволило получить текст, сопоставимый с естественными текстами, написанными непрофессиональными авторами, по параметрам идиоматичности, креативности, способности маркировать сильные позиции в тексте, выражать эмоции и создавать эффект эмоциональной тональности.

Кроме того, важной в методологическом смысле представляется сама постановка вопроса о критериях оценки художественности текста в условиях конкуренции авторов-людей и авторов-моделей.

В данном исследовании мы оставили «за скобками» категории эстетической и этической ценности текста, но сосредоточились на таких категориях, которые обозначают границу между естественным и искусственным интеллектом: когерентность, смысловая и лексико-грамматическая когезия, ритм текста, эмоциональность (тональность), присутствие субъективной модальности сомнения, сложность синтаксиса и выразительность (наличие сильных позиций, использование идиом и стилистических приемов).

Психолингвистический эксперимент продемонстрировал, что в нашей выборке дифференцирующую функцию в наибольшей степени выполняют оба вида связности, ритм и субъективная модальность. Остальные категории нашли свою реализацию лишь в одном из

сгенерированных текстов, что подтверждает тот факт, что данные категории составляют горизонт ближайшего развития для генеративного ИИ.

Заявление об использовании генеративного ИИ: Автор заявляет, что при проведении исследования и подготовке настоящей рукописи генеративный искусственный интеллект и ИИ-ассистированные технологии, подлежащие раскрытию в соответствии с политикой журнала, не использовались.

Statement on the Use of Generative AI: The author declares that no generative artificial intelligence or AI-assisted technologies requiring disclosure under the Journal's policy were used in conducting the research or preparing this manuscript.

Список литературы

- Арнольд И. В. Семантика. Стилистика. Интертекстуальность: сборник научных трудов / И. В. Арнольд; научный редактор П. Е. Бухаркин. М.: ФЛИНТА, 2019. 448 с.
- Базилевский М. П. Исследование новых критериев для обнаружения автокорреляции остатков первого порядка в регрессионных моделях // Математика и математические модели. 2018. С. 13–25. DOI: 10.24108/MATHM.0318.0000102
- Гальперин И. Р. Текст как объект лингвистического исследования. М.: URSS, 2007. 137 с.
- Гиршман М. Ритм и целостность прозаического художественного произведения / М. Гиршман // Вопросы литературы. 1974. № 11. С. 128–150.
- Лотман Ю. М. Структура художественного текста // Лотман Ю. М. Об искусстве. СПб.: «Искусство – СПб», 1998. 383 с.
- Матвеева Т. В. Функциональные стили в аспекте текстовых категорий: Синхронно-сопоставительный очерк. Свердловск: Изд-во Урал. ун-та, 1990. 172 с.
- Шаховский В. И. Эмоции: Долингвистика, лингвистика, лингвокультурология. М.: Либроком, 2010. 128 с.

Шкловский В. Строение рассказа и романа // О теории прозы. М.: Изд-во «Круг», 1925. С. 56–69.

Boriskin A., Galimzianova D. The LSG ChallengeWorkshop at INLG 2024: Prompting Techniques for Crafting Extended Narratives with LLMs // Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges, Tokyo, Japan. Association for Computational Linguistics. P. 118–122.

Chakrabarty T., Padmakumar V., He H. Help Me Write a Poem: Instruction Tuning as a Vehicle for Collaborative Poetry Writing. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022). Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022. P. 6848–6863.

Charolles M. Introduction aux problèmes de la cohérence des textes // Langue française. 1978. Vol. 38. P. 7–41.

Charaudeau P., Maingueneau D. Dictionnaire d'Analyse du Discours. Paris: Seuil, 2002. 661 p.

Combettes B. Thématization et progression thématique dans les récits d'enfants // Langue française. 1978. Vol. 38. Pp. 74–86.

Flesch R. A new readability yardstick // Journal of Applied Psychology. 1948. № 32 (3). Pp. 221–233.

Graesser A. C., Singer M., & Trabasso T. Constructing inferences during narrative text comprehension. Psychological Review. 1994. Vol. 101(3). Pp. 371–395.

Guan J., Huang M. UNION: An Unreferenced Metric for Evaluating Open-ended Story Generation // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, 2020. Pp. 9157–9166.

Iaroshenko P. V., Louckachevitch N. V. Large Language Models versus Native Speakers in Emotional Assessment of Russian Words // Supercomputing Frontiers and Innovations. 2025. Vol. 12. No. 3. Pp. 20–30. DOI: 10.14529/jsfi250302

Ippolito D., Yuan A., Coenen A., Burnam S. Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers. 2022. DOI: 10.48550/arXiv.2211.05030

Johnson-Laird, P. N. Mental Models: Towards a Cognitive Science of Language,

Inference, and Consciousness. Cambridge, MA: Harvard University Press, 1983. 513 p.

Jin Y., Kadam V., Wanvarie D. Plot Writing From Pre-Trained Language Models // Proceedings of the 15th International Conference on Natural Language Generation (INLG 2022). 2022. Pp. 52–67.

Kolmogorova A., Kulikova E., Lobanov V. Aspect-Based Sentiment Analysis Using Large Language Models on Museum Visitor Reviews // Supercomputing Frontiers and Innovations. 2025. Vol. 12. No. 3. Pp. 121–140. <https://doi.org/10.14529/jsfi250309>

Kryscinski W., Keskar N. S., McCann B., Xiong C., Socher R. Neural Text Summarization: A Critical Evaluation // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, 2019. Pp. 540–551.

Lee Y., Ka S., Son B., Kang P., Kang J. Navigating the Path of Writing: Outline-guided Text Generation with Large Language Models // arXiv:2404.13919. <https://doi.org/10.48550/arXiv.2404.13919>

Lewis P., Perez E., Piktus A., Petroni F., Karpukhin V., Goyal N., Küttler H., Lewis M., Yih W., Rocktäschel T., Riedel S., Kiela D. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // arXiv:2005.11401v4. 2021. DOI: 10.48550/arXiv.2005.11401.

Mann W. C., Thompson S. A. Rhetorical Structure Theory: Toward a functional theory of text organization. 1988. Text, 8 (3). Pp. 243–281.

Migal A., Seredina D., Telnina L., Nazarov N., Kolmogorova A., Mikhaylovskiy N. Overview of Long Story Generation Challenge (LSGC) at INLG 2024 // Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges. Tokyo, Japan: Association for Computational Linguistics. Pp. 47–53.

Mikhaylovskiy N. States of LLM-generated Texts and Phase Transitions between them // arXiv:2503.06330 [cs.CL]. DOI: 10.48550/arXiv.2503.06330.

Seredina D. A Report on LSG 2024: LLM Fine-Tuning for Fictional Stories Generation // Proceedings of the 17th International Natural Language Generation Conference: Generation

Challenges. Tokyo, Japan: Association for Computational Linguistics. Pp. 123–127.

Shen C., Cheng L., Nguyen X., You Y., Bing L. Large Language Models are Not Yet Human-Level Evaluators for Abstractive Summarization. 2023. <https://arxiv.org/abs/2305.1309>. DOI: 10.48550/arXiv.2305.13091

Singh N., Bernal G., Savchenko D., Glassman E. L. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence // ACM Transactions on Computer-Human Interaction. 2023. Vol. 30 (5). Pp. 1–57.

Takahashi S., Tanaka-Ishii K. Evaluating computational language models with scaling properties of natural language // Computational Linguistics. 2019. Vol. 45 (3). Pp. 481–513. DOI: 10.48550/arXiv.1906.09379

Vatolin A. Structured Sentiment Analysis with Large Language Models: A Winning Solution for RuOpinionNE-2024 // Proceedings of International Conference on Computational Linguistics and Intelligent Technologies Dialogue 2025. 2025.

Zhang Y., Sun Q., Guo Q., Wu J., Qiu X., Huang X. Do Large Language Models Know What They Don't Know? // arXiv:2210.03025. 2022. DOI: 10.48550/arXiv.2209.01515

Zheng L., Chiang W., Sheng Y., Zhuang S., Wu Z., Zhuang Y., Lin Z., Li Z., Li D., Xing E. P., Zhang H., Gonzalez J. E., Stoica I. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena // arXiv:2306.05685v4. DOI: 10.48550/arXiv.2306.05685.

Zhong W., Naradowsky J., Takamura H., Kobayashi I., Miyao Y. Fiction-writing mode: An effective control for human-machine collaborative writing // Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. 2023. Pp. 1752–1765.

Zhou D., Serman S. Ai.llude: Encouraging Rewriting AI-Generated Text to Support Creative Expression. 2024. DOI: 10.48550/arXiv.2405.17843

Материалы исследования

Олеша Ю. Ни дня без строчки. М.: ФТМ, 1961. 464 с.

Филиппов К. А. Лингвистика текста: Курс лекций. СПб.: Изд-во С.-Петерб. ун-та, 2007. 329 с.

References

Arnold, I. V. (2019). *Semantics. Stylistics. Intertextuality: Collection of Scientific Papers*, FLINTA, Moscow, 448 p. (In Russian)

Bazilevskiy, M. P. (2018). Research of new criteria for detecting first-order residual autocorrelation in regression models, *Mathematics and Mathematical Models*, 13–25. DOI: 10.24108/MATHM.0318.0000102 (In Russian)

Galperin, I. R. (2007). *Text as an Object of Linguistic Research*, URSS, Moscow, 137 p. (In Russian)

Girshman, M. (1974). Rhythm and Integrity of Prose Literary Work, *Questions of Literature*, 11, 128–150. (In Russian)

Lotman, Yu. M. (1998). Structure of the Artistic Text, in Lotman Yu. M. *On Art*, Iskustvo – SPB, St. Petersburg, 383 p. (In Russian)

Matveeva, T. V. (1990). *Functional Styles in Terms of Text Categories: A Synchronic-Comparative Essay*, Ural University Press, Sverdlovsk, 172 p. (In Russian)

Shakhovskiy, V. I. (2010). *Emotions: Pre-Linguistics, Linguistics, Linguoculturology*, Librokom Moscow, 128 p. (In Russian)

Shklovskiy, V. (1925). Structure of the Short Story and Novel, in *On the Theory of Prose*, Krug Publishing House, Moscow, 56–69. (In Russian)

Boriskina, A. and Galimzianova, D. (2024). The LSG ChallengeWorkshop at INLG 2024: Prompting Techniques for Crafting Extended Narratives with LLMs, *Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges*, Tokyo, Japan, Association for Computational Linguistics, 118–122. (In English)

Chakrabarty, T., Padmakumar, V. and He, H. (2022). Help Me Write a Poem: Instruction Tuning as a Vehicle for Collaborative Poetry Writing, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, Abu Dhabi, United Arab Emirates, Association for Computational Linguistics, 6848–6863. (In English)

Charolles, M. (1978). Introduction aux problèmes de la cohérence des textes, *Langue française*, 38, 7–41. (In French)

Charaudeau, P. and Maingueneau, D. (2002). *Dictionnaire d'Analyse du Discours*, Seuil, Paris, 661 p. (In French)

- Combettes, B. (1978). Thématisation et progression thématique dans les récits d'enfants, *Langue française*, 38, 74–86. (In French)
- Flesch, R. (1948). A new readability yardstick, *Journal of Applied Psychology*, 32 (3), 221–233. (In English)
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension, *Psychological Review*, 101(3), 371–395. (In English)
- Guan, J. and Huang, M. (2020). UNION: An Unreferenced Metric for Evaluating Open-ended Story Generation, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online, Association for Computational Linguistics, 9157–9166. (In English)
- Iaroshenko, P.V. and Louckachevitch, N.V. (2025). Large Language Models versus Native Speakers in Emotional Assessment of Russian Words, *Supercomputing Frontiers and Innovations*, 12, 3, 20–30. DOI: 10.14529/jsfi250302 (In English)
- Ippolito, D., Yuan, A., Coenen, A. and Burnam, S. (2022). Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers. DOI: 10.48550/arXiv.2211.05030 (In English)
- Jin, Y., Kadam, V. and Wanvarie, D. (2022). Plot Writing From Pre-Trained Language Models, *Proceedings of the 15th International Conference on Natural Language Generation (INLG 2022)*, 52–67. (In English)
- Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press, 513 p. (In English)
- Kolmogorova, A., Kulikova, E. and Lobanov, V. (2025). Aspect-Based Sentiment Analysis Using Large Language Models on Museum Visitor Reviews, *Supercomputing Frontiers and Innovations*, 12, 3, 121–140. DOI: 10.14529/jsfi250309 (In English)
- Kryscinski, W., Keskar, N.S., McCann, B., Xiong, C. and Socher, R. (2019). Neural Text Summarization: A Critical Evaluation, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, Association for Computational Linguistics, 540–551. (In English)
- Lee, Y., Ka, S., Son, B., Kang, P. and Kang, J. Navigating the Path of Writing: Outline-guided Text Generation with Large Language Models. arXiv:2404.13919 [cs.CL]. (In English)
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S. and Kiela, D. (2021). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv:2005.11401v4. DOI: 10.48550/arXiv.2005.11401 (In English)
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization, *Text*, 8(3), 243–281. (In English)
- Migal, A., Seredina, D., Telnina, L., Nazarov, N., Kolmogorova, A. and Mikhaylovskiy, N. (2024). Overview of Long Story Generation Challenge (LSGC) at INLG 2024, *Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges*, Tokyo, Japan, Association for Computational Linguistics, 47–53. (In English)
- Mikhaylovskiy, N. States of LLM-generated Texts and Phase Transitions between them. arXiv:2503.06330 [cs.CL]. DOI: 10.48550/arXiv.2503.06330 (In English)
- Seredina, D. (2024). A Report on LSG 2024: LLM Fine-Tuning for Fictional Stories Generation, *Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges*, Tokyo, Japan, Association for Computational Linguistics, 123–127. (In English)
- Shen, C., Cheng, L., Nguyen, X., You, Y. and Bing, L. (2023). Large Language Models are Not Yet Human-Level Evaluators for Abstractive Summarization. DOI: 10.48550/arXiv.2305.13091 (In English)
- Singh, N., Bernal, G., Savchenko, D. and Glassman, E.L. (2023). Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence, *ACM Transactions on Computer-Human Interaction*, 30 (5), 1–57. (In English)
- Takahashi, S. and Tanaka-Ishii, K. (2019). Evaluating computational language models with scaling properties of natural language, *Computational Linguistics*, 45 (3), 481–513. DOI: 10.48550/arXiv.1906.09379 (In English)
- Vatolin, A. (2025). Structured Sentiment Analysis with Large Language Models: A

Winning Solution for RuOpinionNE-2024, Proceedings of International Conference on Computational Linguistics and Intelligent Technologies Dialogue 2025. (In English)

Zhang, Y., Sun, Q., Guo, Q., Wu, J., Qiu, X. and Huang, X. (2022). Do Large Language Models Know What They Don't Know? arXiv:2210.03025. DOI:

10.48550/arXiv.2209.01515 (In English)

Zheng, L., Chiang, W., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E.P., Zhang, H., Gonzalez, J.E. and Stoica, I. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. arXiv:2306.05685v4. DOI: 10.48550/arXiv.2306.05685 (In English)

Zhong, W., Naradowsky, J., Takamura, H., Kobayashi, I. and Miyao, Y. (2023). Fiction-writing mode: An effective control for human-machine collaborative writing, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 1752–1765. (In English)

Zhou, D. and Serman, S. (2024). Ai.llude: Encouraging Rewriting AI-Generated Text to Support Creative Expression. DOI: 10.48550/arXiv.2405.17843 (In English)

Corpus Materials

Olesha, Yu. (1961). *Not a Day Without a Line*, FTM, Moscow, 464 p. (In Russian)

Filippov, K. A. (2007). *Text Linguistics: A Course of Lectures*, St. Petersburg University Press, St. Petersburg, 329 p. (In Russian)

The author has read and approved the final manuscript.

Автор прочитал и одобрил окончательный вариант рукописи.

Conflicts of interests: the author has no conflicts of interest to declare.

Конфликты интересов: у автора нет конфликтов интересов для декларации.

Анастасия Владимировна Колмогорова, доктор филологических наук, профессор, заведующий лабораторией языковой конвергенции НИУ ВШЭ Санкт-Петербург, Россия.

Anastasia V. Kolmogorova, Doctor of Philology, Professor, Head of the Laboratory of Language Convergence, National Research University Higher School of Economics, St. Petersburg, Russia.