

ФИЛОЛОГИЧЕСКИЕ НАУКИ / PHILOLOGICAL SCIENCES



Научная статья

УДК 81'42

<https://doi.org/10.37493/2409-1030.2024.2.21>

КОМПОЗИЦИОННО-СТРУКТУРНЫЕ, СЕМАНТИЧЕСКИЕ И ПРЕСУППОЗИЦИОНАЛЬНО-ПРАГМАТИЧЕСКИЕ ПАРАМЕТРЫ И ДЕФЕКТЫ СГЕНЕРИРОВАННЫХ КОРОТКИХ ТЕКСТОВ В ЯЗЫКОВОЙ НЕЙРОСЕТИ GIGACHAT

Сергей Викторович Гусаренко¹, Марина Константиновна Гусаренко^{2*}

^{1, 2} Северо-Кавказский федеральный университет (д. 1, ул. Пушкина, Ставрополь, 355017, Российская Федерация)

¹ Доктор филологических наук, профессор
sgusarenko@mail.ru; <https://orcid.org/0009-0000-9245-2255>

² Кандидат филологических наук, доцент
mkgusarenko@mail.ru; <https://orcid.org/0009-0005-9312-8621>

* Автор, ответственный за переписку

Аннотация. Введение. Актуальность выполненного исследования состоит прежде всего в том, что все более активное обращение самых широких кругов пользователей к генерации текстов разных жанров, свойств и объёмов с помощью так называемых LLM (Large Language Model, Большая Языковая Модель, БЯМ) порождает необходимость изучения собственно лингвистических возможностей этих генеративных моделей, в том числе необходимости лингвистического исследования порождаемых ими текстов. Немаловажным также представляется тот факт, что большое количество специалистов из разных областей деятельности активно используют генеративные возможности языковых нейросетей в своих сугубо профессиональных целях, из чего следует, что качество генерируемых текстов приобрело статус производственного фактора и непосредственным образом влияет на успех в профессиональной деятельности, что также указывает на актуальность предпринятого нами исследования. **Материалы и методы.** В качестве объекта исследования были избраны тексты коротких рассказов, поскольку в определенном отношении такого рода тексты могут рассматриваться как архетипические структуры, лежащие в основе текстов некоторых функциональных типов (сценарии, рекламные тексты, рассказы). **Анализ.** Такое положение дел потребовало изучения сгенерированных рассказов в том числе в аспекте репрезентации в нём фреймовых структур, поскольку это позволяет составить представление о том, как устроены тексты этого типа с точки зрения стандартной языковой семантики, включая синтаксическую семантику. **Результаты.** Установлено, что при всей грамматической, структурно-семантической и композиционной адекватности,

сгенерированные в GigaChatPro тексты коротких рассказов, созданные сетью на определенную тему или по определенному фрейму, могут содержать нарушения в глобально-семантической организации, порожденные как пресуппозиционно-прагматическими нарушениями, так и нарушениями текстовой референции. Отмечены также нарушения в общей функциональной прагматике сгенерированных рассказов: выраженная назидательность, длина предложений в них – 5-7 слов свойственны рассказам для детей младшего школьного возраста, при том, что темы этих рассказов изначально не соответствуют их возрастным интересам.

Ключевые слова: текст, генерация, LLM, языковая модель, нейросеть, семантический, прагматический, дефект, галлюцинация.

Для цитирования: Гусаренко С.В., Гусаренко М.К. Композиционно-структурные, семантические и пресуппозиционно-прагматические параметры и дефекты сгенерированных коротких текстов в языковой нейросети GigaChat // Гуманитарные и юридические исследования. 2024. Т. 11. № 2. С. 368–379. <https://doi.org/10.37493/2409-1030.2024.2.21>

Конфликт интересов: автор заявляет об отсутствии конфликта интересов.

Вклад авторов: все авторы внесли равный вклад в подготовку публикации.

Статья поступила в редакцию: 19.12.2023.

Статья одобрена после рецензирования: 20.02.2024.

Статья принята к публикации: 27.04.2024.

Research article

COMPOSITIONAL-STRUCTURAL, SEMANTIC AND PRESUPPOSITIONAL-PRAGMATIC PARAMETERS AND DEFECTS OF GENERATED SHORT TEXTS IN THE GIGACHAT LANGUAGE NEURAL NETWORK

Sergey V. Gusarenko¹, Marina K. Gusarenko^{2*}

^{1, 2} North-Caucasus Federal University (1, Pushkina St., Stavropol, 355017, Russian Federation)

¹ Dr. Sc. (Philology), Professor
sgusarenko@mail.ru; <https://orcid.org/0009-0000-9245-2255>

² Cand. Sc. (Philology), Associate Professor
mkgusarenko@mail.ru; <https://orcid.org/0009-0005-9312-8621>

* Corresponding author

Abstract. Introduction. The relevance of the study lies primarily in the fact that the increasingly active appeal of the widest circles of users to the generation of texts of different genres, properties and volumes using the so-called LLM (Large Language Model) creates the need to study the actual linguistic capabilities of these generative models, including the need for linguistic research of the texts they

generate. It is also important that a large number of specialists from different fields actively use the generative capabilities of linguistic neural networks for purely professional purposes, which means that the quality of generated texts has acquired the status of a production factor and directly affects success in professional activities, which also indicates the relevance of our research. **Materials and Methods.** The

© Гусаренко С. В., Гусаренко М. К., 2004

texts of short stories were chosen as the object of study, since in a certain respect, such texts can be considered as archetypal structures that underlie texts of certain functional types (scripts, advertising texts and stories). **Analysis.** This state of affairs requires studying the generated stories, including in the aspect of the representation of frame structures in it, since this will allow us to get an idea of how texts of this type are structured from the point of view of standard linguistic semantics, including syntactic semantics. **Results.** It has been established that, despite all grammatical, structural-semantic and compositional adequacy, the texts of short stories generated in GigaChatPro, created by the network on a certain topic or according to a certain frame, may contain violations in global semantic organization, generated by both presuppositional-pragmatic violations and violations in text reference. Violations in the general functional pragmatics of the generated stories were also noted: pronounced edification and the length of sentences up to 5-7 words are characteristic of stories for children of primary

Введение. Все более активное обращение самых широких кругов пользователей к генерации текстов самых разных жанров, свойств и объёмов с помощью так называемых LLM (Large Language Model, Большие Языковые Модели) порождает естественную необходимость изучения генеративных возможностей таких моделей, в том числе необходимость лингвистического исследования порождаемых ими текстов. Таким образом, цель статьи – рассмотреть композиционно-структурные и семантические параметры текстов, генерируемых языковой нейросетью GigaChatPro, созданной специалистами Сбера [13], при этом особое внимание уделялось анализу лингвистических дефектов в сгенерированных текстах. В ходе изучения композиционно-структурных параметров полученных текстов использовался также анализ с позиций фреймовой семантики нарратива [2; 12].

Большое количество специалистов из разных областей деятельности активно используют генеративные возможности языковых нейросетей в своих сугубо профессиональных целях, из чего следует, что качество генерируемых текстов приобрело статус производственного фактора и самым непосредственным образом влияет на успех в профессиональной деятельности, что указывает на актуальность предпринятого нами исследования.

В качестве объекта исследования избраны короткие рассказы, поскольку в определенном отношении они могут рассматриваться как архетипические структуры, лежащие в основе текстов некоторых функциональных типов (сценарии, рекламные тексты, рассказы). Из сказанного следует, что необходимо изучать сгенерированный рассказ с точки зрения репрезентации в нём фреймовых структур, поскольку это позволит составить представление о том, как устроены тексты этого типа с точки зрения стандартной языковой семантики, включая синтаксическую семантику.

Материалы и методы. Специалисты в области создания и применения LLM уделяют много внимания вопросам порождения правильных, семантически полноценных нарративных текстов. Так, в статье MaxProb: Controllable Story Generation from Storyline (MaxProb: Управляемое генерирование рассказов по сюжетной линии) приведены виды контроля, применяемого при порождении текстов с заданными параметрами, а также представ-

school age, despite the fact that the themes of these stories initially do not correspond to their age interests.

Keywords: text, generation, LLM, Large Language Model, neural network, semantic, pragmatic, defect, hallucination.

For citation: Gusarenko SV, Gusarenko MK. Compositional-structural, semantic and presuppositional-pragmatic parameters and defects of generated short texts in the GigaChat language neural network. *Humanities and law research*. 2024;11(2):368-379. (In Russ.). <https://doi.org/10.37493/2409-1030.2024.2.21>

Conflict of interest: the authors declare no conflicts of interests.

Contribution of the authors: the authors contributed equally to this article.

The article was submitted: 19.12.2023.

The article was approved after reviewing: 20.02.2024.

The article was accepted for publication: 27.04.2024.

лен авторский метод разработки контролируемой генерации текстов [19]. Много внимания уделяется настройке параметров генеративной языковой модели [14], разработке эффективных методов управления генерацией текста [18], свойствам генерируемых художественных текстов [8; 4], качеству в том числе и коротких текстов [7] и др.

Особый интерес у исследователей вызывают сгенерированные нейросетями тексты, в которых содержатся разного рода семантические дефекты, или галлюцинации, как их называют в среде специалистов по генеративным языковым моделям. Под галлюцинациями понимаются определенные нарушения в семантическом устройстве сгенерированных текстов, такие, как несоответствие между входными данными и полученным результатом, фактологические отклонения, нерелевантные фрагменты в полученных текстах [16; 15]. Михайловский Н., Чурилов И. указывают на такую особенность сгенерированных текстов, как неспособность выстраивать долгосрочные семантические зависимости в тексте, что иногда приводит к вырождению текста [17]; Остякова Л., Петухова К. и др. отмечают подобные отклонения и приходят к выводу, что генеративные языковые модели необходимо контролировать в процессе их реального общения с пользователями [18]. Снижение качества генерируемых текстов при увеличении длины последовательности отмечают Кривошеев Н.А., Иванова Ю.А., Спицын В.Г. в статье «Автоматическая генерация коротких текстов на основе применения нейронных сетей LSTM и SecGAN» и предлагают решения, которые позволили бы повысить качество генерации коротких текстов [7].

Более подробно о проблемах порождения когерентного текста пишут Козловский А. В., Мельник Я. Э., Волощук В. И. в работе «О подходе для автоматической генерации сюжетно связанного текста», где отмечают, что при более или менее подробном анализе выявляются изъяны, допущенные языковой моделью, и выделяют такой недостаток, как учёт только семантики слов, что «создаёт проблемы с логичностью повествования, поэтому для этого необходимо будет принимать во внимание и другие характеристики (персонажей и предметов их окружающих). Таким образом должны устраниться проблемы с потерей контекста и

общей сути происходящего на продолжительном промежутке времени» [6, с. 161].

Качество текстов, генерируемых большими лингвистическими моделями, неуклонно повышается, и те дефекты, о которых писалось в статьях даже двухлетней давности, уже во многом устранены, тем не менее работа с практическим материалом показала, что такие отклонения, как нарушения логичности, непоследовательность, семантическая аномальность сгенерированных текстов (в терминах проблем LLM), все еще присутствуют, что обусловило проблематику, материал и методы нашего исследования.

Как отмечено выше, в качестве материала исследования избраны короткие рассказы, сгенерированные языковой нейросетью GigaChatPro, размещенном в сети Telegram. Как следует из функционального описания языковой нейросети GigaChat, размещенного на сайте разработчика, этот ресурс предназначен в том числе для порождения текстов самой разной функциональной направленности: рекламные слоганы, сценарии мультфильмов и компьютерных игр, рассказы, рекламные тексты, рейтинговые списки продуктов, научные доклады и т.п. [10].

В качестве основных избраны методы композиционно-семантического, структурно-семантического и фреймового анализа текстов коротких рассказов, что позволило выявлять и описывать дефекты сгенерированных текстов, порожденные нарушениями в их пресуппозициональной и референциальной семантике. В ходе работы также привлекались элементы стилистического и фактологического анализа.

Следует сказать, что многократные верификации полученного результата при работе с языковыми нейросетями не всегда возможны, поскольку на одно и то же задание нейросеть может выдать разные по содержанию, структуре и объёму тексты, тем не менее по мере возможности такие верификации регулярно выполнялись. То же можно сказать о повторах однотипных, но разных по содержанию заданий: верификации не всегда возможны, поскольку сеть может выдать совершенную бессмыслицу.

В результате проведенного исследования были выявлены два основных типа дефектов в сгенерированных текстах коротких рассказов – нарушения прагматических пресуппозиций [11] и референциальные нарушения макроструктур [1; 3]. В контексте нашего исследования под нарушениями прагматических пресуппозиций (или пресуппозиционально-прагматические нарушения) мы имеем в виду включение в сгенерированные тексты языковых выражений, содержание которых противоречит стандартным представлениям о положении дел в мире. Референциальные нарушения макроструктур трактуются в нашей статье как нарушения референциальной связности, порождающие дефекты макроструктур текста. Пресуппозиционально-прагматические нарушения следует

рассматривать в тесной связи с нарушениями фреймовой семантики текста, вызванными несоответствием какого-либо фрагмента текста фрейму как стандартной языковой модели, имеющей явно конвенциональный характер [2].

Анализ. Лингвистические модели обучаются на большом количестве текстов, имеют свою базу текстовых данных, что естественным образом отражается на качестве порождаемых текстов, и для иллюстрации сказанного можно поставить нейросети задачу построить микротекст без указания каких-либо конкретных параметров. Рассмотрим ответ нейросети GigaChat в Telegram на вопрос «Что говорил Паниковский, заглядывая в рот коню на постоялом дворе?» без указания источника (рис. 1).



Рисунок 1. Короткий ответ на запрос без конкретных параметров / Figure 1. Short response to a request without specific parameters

Как видим, нейросеть приписала Паниковскому слова, которых он не говорил: как известно, Паниковский в тексте романа «Золотой теленок» И. Ильфа и Е. Петрова [5] произносит фразу «Добрый жеребец», глядя в зубы доброй кобыле, – и сделала его персонажем романа «Двенадцать стульев», что может быть объяснено тем обстоятельством, что в базе данных GigaChat отсутствуют тексты романов «Золотой теленок» и «Двенадцать стульев», но при этом есть публицистические или другие тексты, в которых присутствуют имена Паниковский и «Двенадцать стульев».

Первый абзац представляет собой свободное изложение несуществующего фрагмента текста романа «Двенадцать стульев», остальная часть ответа – интерпретация нейросетью его же сгенерированного высказывания Паниковского, а также его действий, описанных в запросе.

Отметим разнотилевые выражения в контекстно расположенных предложениях: «сила», «грация» и «природная красота» в одном и «состояние полости рта» в другом. Также можно отме-

тить, что второй и третий абзацы построены по одной композиционной модели: первая часть – интерпретация фразы Паниковского, вторая часть – интерпретация его действий по отношению к лошади. Между вторым и третьим абзацем наблюдается семантическое расширение – от «любви и уважении Паниковского к лошадям» до «интереса к животным» вообще. Последний абзац имеет обобщающий характер, что вполне свойственно развернутому ответу в научном и публицистическом стилях общения, при этом названные нейросетью черты характера «общая чувствительность и забота об окружающем мире» представляют собой искусственные приращения, хотя не аномальные.

Порожденный нейросетью текст не имеет целостного стилистического образа, хотя и вполне правилен в структурно-семантическом отношении. В то же время сама структура текста – от исходного текста (тезиса) к его последующей интерпретации с последовательным расширением смысла свойственна научным исследовательским текстам, публицистике, юридическим толкованиям. Такое положение дел можно объяснить тем, что, с одной стороны, модель обучалась во многом на медицинских, юридических и экономических текстах; с другой – запрос не содержал четко поставленной задачи, или промпта, какой именно текст следовало сгенерировать. В целом, ответ нейросети можно охарактеризовать следующим образом: перед нами грамматически и композиционно-семантически правильный (когерентный) текст, построенный по модели тематического фрагмента текста или краткого ответа в диалоге в одном из названных стилей; при этом в глобальном плане текст семантически аномален – в первую очередь потому, что основной его элемент – высказывание Паниковского не соответствует прагматическим presuppositions дискурса [3]. Следует также сказать, что на четыре других фактологических вопроса по романам И. Ильфа и Е. Петрова нейросеть дала четыре разных по композиции и структуре, а также по характеру семантических аномалий текста-ответа.

Иную картину можно наблюдать, если поставить нейросети задачу с точным указанием темы и жанра генерируемого текста. Проведенный анализ позволил сделать вывод, что при всей лексической, грамматической, структурно-семантической и композиционной правильности, сгенерированный текст может содержать аномалии в глобальной семантической организации. Обновленной версии GigaChatPro был сформулирован промпт с указанием темы ПОСЕЩЕНИЕ ТИРА, объема и жанра – рассказ (рис. 2).

Нейросеть выдала результат (см. рис. 2) – когерентный, грамматически и структурно-семантически правильный текст объемом 135 слов, содержащий, тем не менее, нарушения в глобальной связности и прагматике адресата. В данном случае микротекст построен по модели рассказа-притчи с соблюдением пространственно-

временной последовательности события ПОСЕЩЕНИЕ ТИРА.



Рисунок 2. Короткий рассказ на тему ПОСЕЩЕНИЕ ТИРА / Figure 2. Short story on the topic VISITING A SHOOTING GALLERY

Первый абзац – зачин: вводится главный персонаж с указанием имени, дается краткая характеристика его личностных качеств, ввод предложения о событии – решение посетить тир и причина этого решения; второй абзац: прибытие персонажа к месту основного действия, начало основного действия, физическая характеристика персонажа; третий абзац: развитие главного действия – от неудачи к успеху через настойчивое повторение действия с описанием сопутствующих психолого-волевых и когнитивных состояний; четвертый абзац: кульминация основного действия, описание сопутствующих психологических состояний и соответствующих действий второстепенных персонажей; пятый абзац – финал: поучительный вывод, наставление, «правильное» продолжение действий главного персонажа (ср. финал в сказке: стали жить-поживать и добра наживать).

Рассказ представляет собой в высокой степени когерентное структурно-семантическое образование – за исключением одного референциального нарушения глобальной связности: главный персонаж Никита пришел в тир один, но после успешного выстрела его поздравляют друзья. Можно также считать нарушением следующее presuppositionальное несоответствие: финал рассказа сообщает, что главный персонаж «продолжил тренироваться, чтобы стать еще лучше» – при том, что в описанном событии он проявил себя как посредственный стрелок, однако цель – стать еще лучше можно отнести к общей характеристике персонажа.

Следует отметить грамматическую, стилистическую, пунктуационную правильность сгенерированного текста. Длина предложений, в том числе простых предложений в составе сложных, составляет 5 – 7 слов, что соответствует рассказам для детей младшего школьного возраста. В этом отношении можно отметить нарушения прагматического характера: сочинение нравоучительных рассказов для младших школьников на таком материале – ПОСЕЩЕНИЕ ТИРА не соответствует их возрастным интересам.

Сгенерированный текст характеризуется высоким уровнем семантической локальной связности текста: к примеру, указанные в зачине качества главного персонажа – *смелый* и *отважный* непосредственным образом объясняют мотив посещения тира – проверить свою *меткость* и *ловкость*; в третьем абзаце сентенция *меткость приходит с опытом* объясняет действия главного персонажа *продолжал тренироваться и стрелять* из следующего предложения; в последнем абзаце за сентенцией *нужно быть терпеливым, настойчивым* следуют действия персонажа *продолжил тренироваться*.

Высокая семантическая локальная связность текста, его строгая логичность, выраженная назидательность позволяют предположить, что обучающие тексты данной семантики или тексты, к которым модель имела доступ в ходе генерации, относятся к одной жанрово-тематической группе – рассказы для детей. Это подтверждает и нарушение глобальной связности: тексты такого рода предполагают участие «правильных» друзей, и модель их ввела в рассказ по статистическим параметрам, но в нарушение глобальной связности.

В ходе работы была отмечена еще одна особенность: сгенерированные нейросетью GigaChatPro три рассказа на тему ПОСЕЩЕНИЕ ТИРА (два повторяются) завершаются поучительным выводом, при том, что сгенерированные этой же нейросетью другие микрорассказы такого финала не имеют.

С точки зрения фреймовой семантики, рассказ о посещении тира также представляет собой дефектное образование – в нем не содержится одного из слотов в терминалах высшего уровня, а именно слота ОРУЖИЕ. При этом ядерные пропозиции с соответствующими ядерными предикатами [Мы, Токарева] в полной мере присутствуют: *Никита посещает тир, Никита стреляет, Никита промахивается, Никита попадает в цель, Никита тренируется*.

Естественным продолжением исследования представляется постановка задания по определенному семантико-синтаксическому фрейму [2; 12]. На начальном этапе работы языковой нейросети GigaChatPro была поставлена задача сформулировать понятие семантико-синтаксического фрейма. В неполной мере, но нейросеть справилась с поставленной задачей, представив средней полноты дефиницию (рис. 3).

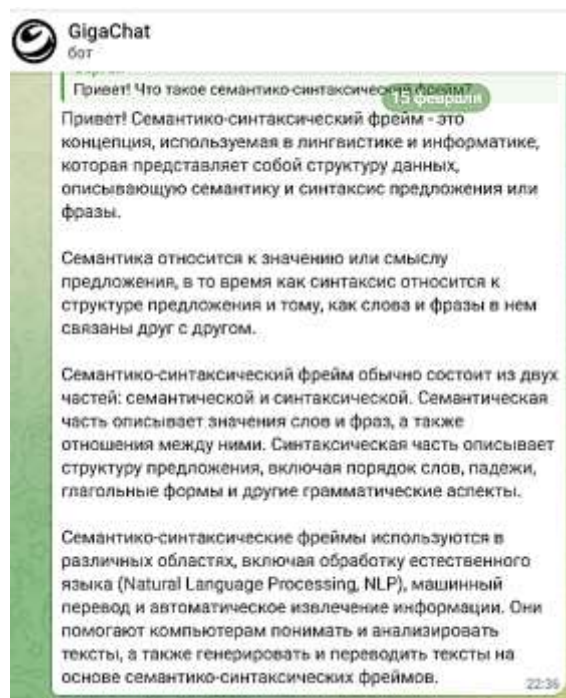


Рисунок 3. Определение семантико-синтаксического фрейма / Figure 3. Definition of a semantic-syntactic frame

Как видим, дано в большей мере научно-популярное определение семантико-синтаксического фрейма, тем не менее оно содержит основные признаки определяемого объекта, хотя данное определение и не содержит указаний на два существенных свойства фрейма: во-первых, не отражает его иерархическую организацию, что существенно для определения значимости его элементов и поисковых процедур; во-вторых, не отражает горизонтальные и вертикальные связи между строевыми элементами фрейма как сложной семантической структуры, включая временную связь в последовательности событий фрейма (сценарий).

Для проверки соответствия представленной нейросетью структуры фрейма и ее способностью порождать тексты по этой модели, нейросети была поставлена задача построить по фрейму ПОСЕЩЕНИЕ ТИРА рассказ, заданная длина рассказа была увеличена до 200 слов для расширения возможностей генерации, и получен следующий результат (рис. 4).

В жанрово-стилистическом отношении сгенерированный текст представляет собой короткий рассказ с назидательным финалом, в содержательно-композиционном плане данный рассказ повторяет первый рассказ – рассказ на тему ПОСЕЩЕНИЕ ТИРА: приход в тир – неудачная стрельба – успешная стрельба – дальнейшая тренировка. В обоих случаях присутствуют лица, не относящиеся к первым термам фрейма – друзья и сотрудник тира, во втором случае в качестве главного персонажа выступает сам рассказчик, то есть нейросеть по какой-то причине пе-

решла к повествованию от первого лица. Рассказ выдержан в содержательном плане, не содержит ненужных деталей и подробностей.

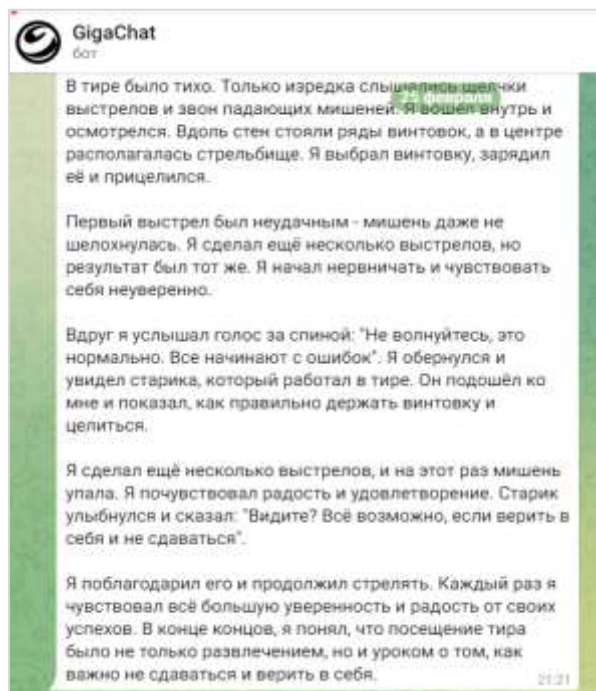


Рисунок 4. Короткий рассказ по фрейму ПОСЕЩЕНИЕ ТИРА / Figure 4. Short story based on the frame VISITING A SHOOTING GALLERY

По последнему рассказу была построена упрощенная модель семантико-синтаксического фрейма ПОСЕЩЕНИЕ ТИРА с замещенными позициями в слотах (таблица 1). В квадратных

скобках приведены имплицитные пропозиции, представляющие фрейм в тексте рассказа.

Как видим, семантико-синтаксический фрейм ПОСЕЩЕНИЕ ТИРА весьма подробно представлен поверхностными структурами в сгенерированном рассказе. Иерархизация фрейма выполнена по принципу наибольшей значимости пропозиции для выражения основного события (основных событий). Обращает на себя внимание количественное преобладание пропозиций четвертого уровня – самого отдаленного от ядра фрейма, что может быть объяснено относительной свободой содержания жанра рассказа. Отметим, что в первом рассказе содержится три пропозиции, описывающие психологические состояния.

Рассказ содержит ряд отклонений в плане его соответствия прагматическим пресуппозициям более или менее грамотного читателя. Во-первых, рассказчик слышит звон падающих мишеней, что свойственно скорее стрельбищу на биатлонной трассе, чем закрытому тиру (*Я вошел внутрь...*); во-вторых, рассказчик увидел в тире в одном пространстве ряды винтовок вдоль стен и стрельбище, что при прочих нормальных условиях невозможно: комнаты для хранения оружия, если они есть, в тире отделены от стрелкового павильона; в-третьих, рассказчик само место для стрельбы (стрелковый павильон, огневой рубеж) называет стрельбищем, между тем как стрельбище – это обширное открытое пространство, чаще за пределами населенного пункта, подготовленное для проведения стрельб.

Таблица 1 / Table 1

Семантико-синтаксическая модель фрейма ПОСЕЩЕНИЕ ТИРА по данным из рассказа / Semantic-syntactic model of the frame VISITING A SHOOTING GALLERY based on data from the story

Уровень терминала	Слоты	Пропозиции	Наполнение слотов по рассказу
Первый уровень	<i>А – первый терм</i> <i>Тир – место события</i>	Ядерная пропозиция: <i>А стреляет в тире</i>	<i>Я сделал ещё несколько выстрелов [в тире]</i>
	<i>Цель / мишень</i>	Ядерная пропозиция: <i>А попадает в цель</i>	<i>На этот раз мишень упала [потому что я попал в мишень]</i>
		Ядерная пропозиция: <i>А промахивается</i>	<i>Первый [мой] выстрел был неудачным [я промахнулся]</i>
Второй уровень	<i>Оружие</i>	Пропозиция: <i>А берет оружие</i>	<i>Я выбрал винтовку</i>
		Пропозиция: <i>А держит оружие</i>	<i>[А держит винтовку]</i>
		Пропозиция: <i>А заряжает оружие</i>	<i>Я зарядил её</i>
	<i>Цель / мишень</i>	Пропозиция: <i>А прицеливается [в мишень]</i>	<i>Я прицелился [в мишень]</i>
Третий уровень	<i>Сопутствующие звуки</i>	Пропозиция: <i>От оружия исходит звук</i>	<i>Слышались щелчки выстрелов</i>
		Пропозиция: <i>От поражённой мишени исходит звук</i>	<i>Слышался звон падающих мишеней</i>
Четвертый уровень	<i>S - психологические состояния</i>	Пропозиция: <i>А испытывает S</i>	<i>Я начал нервничать</i>
			<i>Я начал чувствовать себя неуверенно</i>
			<i>Я почувствовал радость</i>
			<i>Я почувствовал удовлетворение</i>
			<i>Я чувствовал уверенность</i>

Из логических нарушений можно отметить несоответствие следствия исходной посылке: *Я сделал ещё несколько выстрелов, и на этот раз мишень упала*: рассказчик сделал несколько последовательных выстрелов, но мишень упала только единожды – *на этот раз*. В случае с адекватным рассказчиком-человеком такого рода отклонения были бы маловероятны.

Вызывает интерес старик – работник тира, оказавший помощь главному персонажу: такой сюжетный ход – появление помощника (дарителя) свойственен волшебной сказке [9], что многое может сказать о том, на каких текстах обучалась языковая модель.

Следующим шагом в изучении коротких рассказов, сгенерированных языковой нейросетью GigaChatPro, стало усложнение задания написать рассказ по заданному фрейму: в промпт, помимо основной задачи сгенерировать рассказ по фрейму ПОСЕЩЕНИЕ ТИРА, было введено дополнительное задание включить слова *осечка, гильза, затвор, целик*, которые являясь слотами высших уровней фрейма. Таким образом преследовалась цель получить данные о том, как нейросеть внедрит эти слова в текстовое представление фрейма. В результате нейросеть обошлась минимальными структурными изменениями, а именно оставила предыдущий текст неизменным, за исключением двух фрагментов: заменила выражение во втором абзаце и дописала два абзаца в финале (рис. 5, 6).

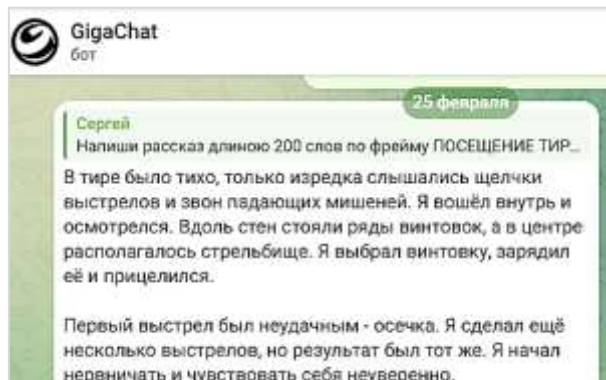


Рисунок 5. Первый изменённый фрагмент рассказа по фрейму ПОСЕЩЕНИЕ ТИРА с дополнительным заданием / Figure 5. The first modified fragment of the story based on the frame VISITING A SHOOTING GALLERY with an additional task

Как видим, во втором абзаце простое предложение *мишень даже не шелохнулась*, поясняющее предшествующее выражение, заменила словом *осечка*, которое выполняет ту же синтаксическую функцию. Отметим, что в семантико-синтаксическом отношении замена выполнена безупречно.

Как сказано выше, финал рассказа дополнен двумя абзацами (Рис. 6). Абзац *Я заметил, что...* представляет собой явно аномальное приращение: за композиционно обусловленным финалом из прежнего рассказа следует коммуни-

кативно аномальный фрагмент, включающий несущественные для сюжета детали.

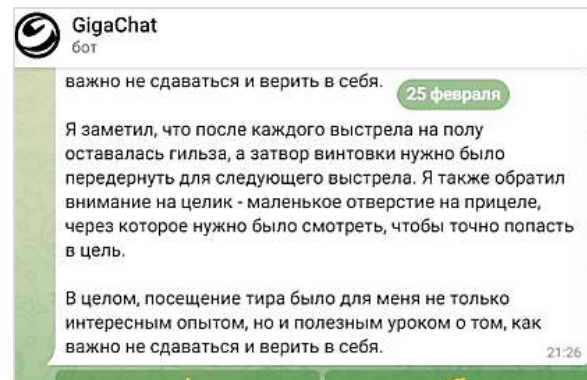


Рисунок 6. Второй изменённый фрагмент рассказа по фрейму ПОСЕЩЕНИЕ ТИРА с дополнительным заданием / Figure 6. The second modified fragment of the story based on the frame VISITING A SHOOTING GALLERY with an additional task

Первая деталь заключена в семантически аномальном высказывании *Я заметил, что после каждого выстрела на полу оставалась гильза, а затвор винтовки нужно было передернуть для следующего выстрела*, в котором за фактологическим главным предложением *Я заметил*, предполагающим ввод новых для говорящего данных, следуют две банальные детали: стреляющий неизбежно видит и знает, что после каждого выстрела на полу остаётся гильза и что затвор нужно передергивать для следующего выстрела – уже хотя бы потому, что сам с первого выстрела выполняет эту операцию. Второе предложение *Я также обратил внимание...* в этом же абзаце содержит такую же аномалию: выражение *обратил внимание* пресуппозиционно противоречит предыдущей части текста, в которой сказано, что говорящий целился, то есть изначально смотрел через целик на мишень.

Абзац *В целом, посещение...* есть не что иное, как функционально-семантический повтор последнего предложения *В конце концов, я понял...* в пятом абзаце – выводе-наставлении этого же варианта рассказа и, соответственно, финального абзаца в первом варианте рассказа.

Для верификации результатов нейросети дважды подряд было дано аналогичное задание: сначала сгенерировать короткий рассказ по определённому фрейму, затем задание сгенерировать рассказ по этому же фрейму с указанием конкретных слов-слотов этого фрейма, которые должны присутствовать в тексте. В первом случае задание было построить рассказ по фрейму ОБЕД В РЕСТОРАНЕ, во втором случае – по фрейму ПОЕЗДКА НА ПОЕЗДЕ. В результате нейросеть выдавала один и тот же текст и на задание сгенерировать текст только по фрейму, и на задание построить рассказ по фрейму с указанием обязательных слов. То есть список слов, обязательных для включения в рассказ, был проигнорирован, причем в случае с фреймом ОБЕД

В РЕСТОРАНЕ из десяти обязательных слов четыре не были использованы, в случае с фреймом ПОЕЗДКА НА ПОЕЗДЕ из десяти слов неустраиваемыми оказались восемь. Добавим, что рассказ по фрейму ОБЕД В РЕСТОРАНЕ по сюжету оказался рассказом о случайной встрече двух бывших одноклассников с неожиданным вопросом в финале *Друзья, а вы часто ходите в рестораны?* Рассказ по фрейму ПОЕЗДКА НА ПОЕЗДЕ оказался рассказом о случайной встрече рассказчицы (по окончаниям глаголов в прошедшем времени) в купе поезда со случайным пассажиром, который имел при себе пистолет, по его словам, для самообороны. В обоих случаях языковая нейросеть не связывает строго семантику фрейма в рассказе с его сюжетом, в то время как сюжет рассказа ПОСЕЩЕНИЕ ТИРА строго связан с сценарной семантике фрейма.

Следующим этапом исследования текстов, генерируемых нейросетью GigaChatPro, стала постановка задачи на создание короткого рассказа по определенному фрейму с предварительным запросом на содержание и структуру этого фрейма. Задача такого рода позволяет определить, насколько составленная языковой нейросетью модель фрейма оказывает влияние на построение текста по этому фрейму. Иными словами, посредством такого задания для нейросети можно определить, действительно ли она применяет созданные ею вспомогательные семанто-синтаксические структуры для генерации текста.

Языковой нейросети GigaChatPro задание было сформулировано следующим образом: *Составь структурную модель семанто-синтаксического фрейма ПОСЕЩЕНИЕ КОНЦЕРТА*, результат представлен на Рис. 7. Как видим, модель не содержит признаков сценария, нет связей между уровнями и слотами, приведенными в типовых пропозициях. Но в результате оказывается, что сеть «знает» все сценарии и связи между компонентами фрейма, потому что сгенерированный рассказ содержит все это.

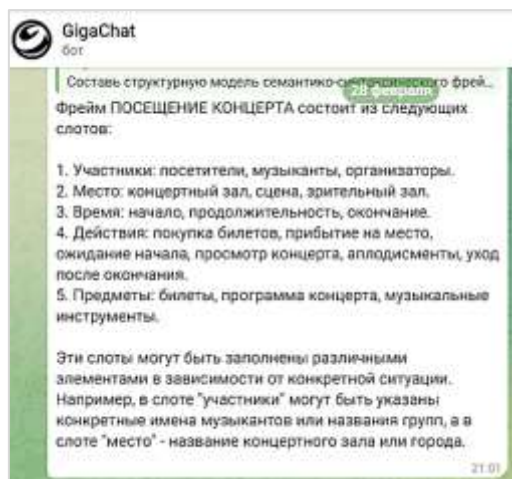


Рисунок 7. Структурная модель фрейма ПОСЕЩЕНИЕ КОНЦЕРТА / Figure 7. Structural model of the CONCERT ATTENDANCE frame

Как видим, построенная нейросетью модель фрейма имеет некоторое подобие иерархии при довольно четком определении семантических ролей, сирконстантов и предикатов первого термина. Главный недостаток этой модели состоит в том, что ядерные термины, третьи (орудийные термины) и предикаты даны разрозненно, а не в составе пропозиций. Об иерархии можно говорить в том смысле, что на первое место сеть поставила первые и вторые термины – *посетители* (первый терм), *музыканты* (первый терм), *организаторы* (второй терм). Неадекватность иерархии состоит в том, что ядерные предикаты и третьи термины поставлены на два последних места – четвертое и пятое. Модель содержит лексические отклонения: в числе первых терминов указаны посетители, а не зрители, что было бы более предпочтительным для фрейма ПОСЕЩЕНИЕ КОНЦЕРТА; в терминале МЕСТО присутствует частичный семантический повтор *концертный зал – зрительный зал*. Данные особенности составленной нейросетью модели не позволяют отнести её к семанто-синтаксическим фреймам, что свидетельствует об отсутствии в нейросети конструктивной, функциональной связи между составленным нейросетью определением семанто-синтаксического фрейма (Рис. 3) и сгенерированной ею же моделью фрейма ПОСЕЩЕНИЕ КОНЦЕРТА.

При этом короткий рассказ на тему ПОСЕЩЕНИЕ КОНЦЕРТА, сгенерированный по составленной нейросетью модели фрейма, в полной мере соответствует этой модели, хотя и не включает некоторые слоты из неё (рис. 8).

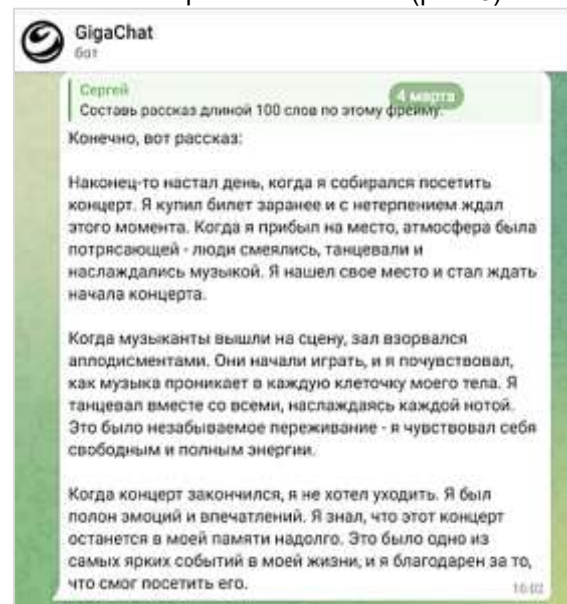


Рисунок 8. Короткий рассказ по фрейму ПОСЕЩЕНИЕ КОНЦЕРТА / Figure 8. Short story based on the frame ATTENDING THE CONCERT

Прежде всего отметим, что рассказ построен по сценарию фрейма, имеет четко выстроенную композицию: первый абзац – это действия, предшествующие собственно концерту (покупка билета

и прибытие на место); второй абзац – само концертное действие и сопутствующие психологические состояния; третий абзац – действия по окончании концерта, интеллектуально-психологические состояния. В терминале ВРЕМЯ из составленной нейросетью модели фрейма (Рис. 8) выдержанная в рассказе последовательность представлена предельно четко: «начало, продолжительность, окончание». Такое соответствие позволяет предположить, что нейросеть строила рассказ, соотнося его композицию с моделью фрейма.

Из первого терминала УЧАСТНИКИ присутствуют два слота из трех – *посетители*, *музыканты*, отсутствует слот *организаторы*; из терминала МЕСТО в рассказе присутствует слот *цена*, отсутствуют слоты *концертный зал*, *зрительный зал*; из терминала ВРЕМЯ, как уже сказано выше, присутствуют все указанные слоты; из терминала ДЕЙСТВИЯ (предикаты) присутствует пять слотов из шести: *покупка билета*, *прибытие на место*, *ожидание начала*, *аплодисменты*, *уход после окончания*, отсутствует слот *просмотр концерта*, хотя его можно идентифицировать с ядерным предикатом в предложении *Я собирался посетить концерт*, при том что сам этот предикат – *посетить концерт* отсутствует в составленной нейросетью модели; терминал ПРЕДМЕТЫ представлен в тексте слотом *билеты*, отсутствуют *программа концерта* и *музыкальные инструменты*. В то же время в сгенерированном тексте присутствует ядерная пропозиция *Музыканты играют*, предиката которой нет в составленной нейросетью модели фрейма.

О функциональной связи между построением текста рассказа и составленной ранее моделью фрейма может говорить также тот факт, что и в рассказе, и в модели присутствует слот *ожидание начала*, который вообще нельзя хотя бы в какой-то мере считать специфичным, обязательным для фрейма ПОСЕЩЕНИЕ КОНЦЕРТА, то есть нейросеть по какой-то причине ввела этот слабо релевантный слот в модель, а потом включила в рассказ в соответствии с моделью. С другой стороны, это соответствие может быть обусловлено тем, что в числе немногочисленных обучающих текстов, вероятнее всего, попались такие, в которых было описано ожидание начала концерта, что и отразилось как в модели фрейма, так и в сгенерированном рассказе.

Сгенерированный нейросетью GigaChatPro текст по фрейму ПОСЕЩЕНИЕ КОНЦЕРТА содержит ряд аномалий пресуппозиционального и референциального характера. Нарушение прагматических пресуппозиций, или несоответствие фрейму, усматривается в присутствии предложения *Люди... танцевали и наслаждались музыкой* в описании моментов, предшествующих началу концерта. К нарушению текстовой референции следует отнести предложение *...я был благодарен за то, что смог посетить его [концерт]*, поскольку в предыдущей части текста отсутствует референт адресата бла-

годарности, который сам по себе не является обязательным слотом фрейма.

На контрольный запрос – составление семантико-синтаксического фрейма ПОСЕЩЕНИЕ СПОРТИВНОГО ЗАЛА нейросеть выдала модель, которая, с одной стороны, по составу слотов отличается от модели ПОСЕЩЕНИЕ КОНЦЕРТА, с другой – представляет собой редуцированный вариант в содержательном отношении (рис. 9).

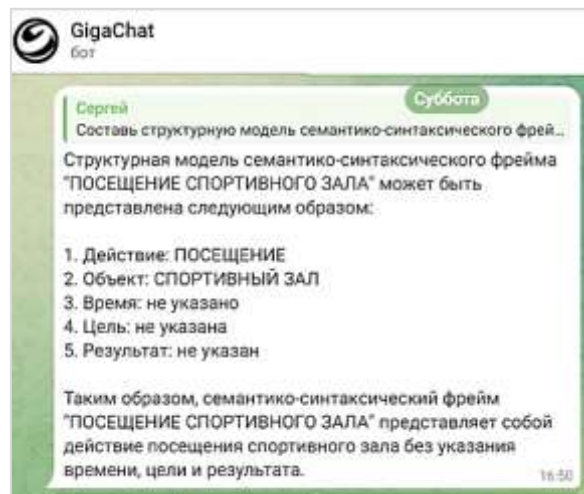


Рисунок 9. Структурная модель фрейма ПОСЕЩЕНИЕ СПОРТИВНОГО ЗАЛА / Figure 9. Structural model of the frame VISITING THE GYM

На запрос построить рассказ по этому фрейму нейросеть выдала такой же редуцированный вариант текста по отношению к тексту посещение концерта (рис. 10).



Рисунок 10. Короткий рассказ по фрейму ПОСЕЩЕНИЕ СПОРТИВНОГО ЗАЛА / Figure 10. Short story based on the frame VISITING THE GYM

В частности, из всех пяти терминалов, только два оказались отображенными в тексте. Между тем рассказ содержит едва ли не все ядерные и периферийные слоты, что и текст ПОСЕЩЕНИЕ КОНЦЕРТА: собственно спортзал и действия в нём, описание физических ощущений (*сильный и энергичный*), описание психологического состояния (*доволен*), отсутствует слот ДРУГИЕ УЧАСТНИКИ (другие первые термы). Отметим, что рассказ построен строго в соответствии со сценарием фрейма: *принятие решения – прибытие – основное действие – убытие*.

Результаты. Проведенный анализ позволил сделать следующие выводы. Прежде всего следует сказать, что сгенерированные в GigaChatPro тексты коротких рассказов в подавляющем большинстве характеризуются грамматической, структурно-семантической и композиционной правильностью. Текстам такого рода, за редкими исключениями, свойственна также семантическая локальная связность, его строгая логичность, композиционная целесообразность в соответствии с жанровыми параметрами. Предложения имеют среднюю длину 5 – 7 слов, что свойственно рассказам для детей раннего школьного возраста.

Тематическая группа коротких рассказов может иметь выраженный назидательный характер, то есть содержать эволюционное продвижение от неудачи к успеху, сопровождаемое описаниями физических и психологических состояний, завершаться поучительным выводом. Сгенерированный в GigaChatPro рассказ может также завершаться вопросом, обращенным к читателю, и такой финал приближает его к детским дидактическим рассказам.

Высокая семантическая локальная связность сгенерированных текстов, их строгая логичность и выраженная назидательность позволяют предположить, что обучающие тексты данной семантики или тексты, к которым модель имела доступ в ходе генерации, относятся к одной жанрово-тематической группе – рассказы для детей.

Сгенерированные в языковой нейросети GigaChatPro короткие рассказы могут быть построены как по сценарию фрейма, так и по микросюжету, не связанному напрямую с фреймом, в нашем случае это сюжет о неожиданной встрече.

Из повторяющихся дефектов, выявленных в коротких рассказах, сгенерированных в языковой нейросети GigaChatPro, следует отметить следующие. Прежде всего следует сказать, что короткие рассказы, созданные сетью на определенную тему или по определенному фрейму, могут содержать аномалии в глобально-семантической организации, порожденные как пресуппозиционно-прагматическими нарушениями, так и нарушениями текстовой референции. И те и другие нарушения вполне объяснимы, если предположить, что сам процесс порождения текстов и его конечный результат не подвергаются текущему контролю с целью избежания подобных дефектов.

Пресуппозиционно-прагматические нарушения в сгенерированных рассказах выражены прежде всего в том, что отдельные описываемые факты не соответствуют естественным представлениям о действительности, такие нарушения можно рассматривать также как фреймовые нарушения. Референциальные нарушения глобальной связности связаны с тем, что нейросеть может включить в текст рассказа имена и именные группы, присутствие которых не предполагается предыдущей частью текста.

Необходимо отметить нарушения в общей функциональной прагматике сгенерированных рассказов: как отмечено выше, выраженная назидательность, длина предложений в них – 5-7 слов свойственны рассказам для детей младшего школьного возраста, при этом темы этих рассказов изначально не соответствуют их возрастным интересам (посещение тира, концерта, поездка в поезде и т.п.).

Задание написать рассказ по определенному фрейму языковая нейросеть GigaChatPro воспринимает примерно так же, как и задание написать рассказ на определенную тему, при этом в случае задания генерации по фрейму рассказ либо повторял буквально рассказ, написанный на определенную тему, либо включал такое же количество слотов из верхних терминалов фрейма (4-6 слотов), что и рассказ на определенную тему. Следует отметить количественное преобладание пропозиций четвертого уровня в структуре фрейма (описание сопутствующих состояний персонажа), что может быть объяснено относительной свободой содержания жанра рассказа. Несколько иная картина наблюдается, если нейросети сначала поставить задачу составить модель семантико-синтаксического фрейма и затем поставить задачу написать рассказ по этому фрейму: в этом случае все пять терминалов, включенные нейросетью в модель фрейма, были представлены в тексте, из восемнадцати слотов в модели в текст включены одиннадцать. При этом отсутствие некоторых слотов не вызвало семантической недостаточности текста. Порождение нейросетью рассказа, в котором представлены все терминалы созданной ею же модели фрейма, позволяет говорить о существовании либо функциональной связи между создаваемыми языковой нейросетью фреймовыми моделями и порождением текстов по этим моделям, либо генеративной связи: и фреймовые модели и рассказы создаются на материале одних и тех же обучающих текстов.

Построенная нейросетью GigaChatPro структурная модель фрейма не имеет определенной внутренней иерархии, хотя семантические роли, сирконстанты и предикаты первого термина определены точно. При этом ядерные термы, орудийные термы и предикаты даны разрозненно, то есть не в составе пропозиций. Составленная нейросетью модель фрейма имеет лексическое нарушение: в состав первых термов включен слот *посетители*, хотя предпочтительнее был бы слот *зрители*. Все эти особенности и нарушения не позволяют отнести составленную нейросетью модель к семантико-синтаксическим фреймам.

Перспективы дальнейшего исследования видятся прежде всего в изучении возможностей построения нейросетью сложных текстовых структур по заданным языковым структурно-содержательным моделям типа семантико-синтаксического фрейма. Следует также отметить стилистическую «сдержанность» языковой

модели: при всей стилистической выдержанности сгенерированных ею рассказов, тексты не содержат ярких языковых выразительных средств (метафоры, сравнения, гиперболы и пр.), только от-

дельные эпитеты. Как представляется, при дальнейшем исследовании этот факт сможет многое сказать о внутренних механизмах и возможностях генеративных языковых нейросетей.

Литература

1. Арутюнова Н. Д. Предложение и его смысл: логико-семантические проблемы: монография. М.: Наука, 1976. 383 с.
2. Гусаренко С. В., Гусаренко М. К. Предикатно-фреймовое описание линейной семантики художественного нарратива // Филологические науки. 2020. № 3. С. 18–30.
3. Дейк Т. А. ван, Кинч В. Стратегии понимания связного текста // Новое в зарубежной лингвистике: Когнитивные аспекты языка. 1988. Вып. XXIII. С. 153–211.
4. Демченков С. А., Федяева Н. Д. Вербальный хеппенинг: о некоторых моделях компьютерной генерации текстов в аспекте авангардной поэтики // Humanitarian Vector. 2020. Том 15. № 1. С. 27–37.
5. Ильф И., Петров Е. Двенадцать стульев. Золотой теленок: Романы. Записные книжки. М.: Эксмо, 2007. 640 с.
6. Козловский А. В., Мельник Я. Э., Волощук В. И. О подходе для автоматической генерации сюжетно связанного текста // Известия ТулГУ. 2022. Вып. 9. С. 160–167.
7. Кривошеев Н. А., Иванова Ю. А., Спицын В. Г. Автоматическая генерация коротких текстов на основе применения нейронных сетей LSTM и SecGAN // Вестник Томского государственного университета. 2021. № 57. С. 118–130.
8. Кучина С. А. Электронные генеративные художественные тексты // Вестник ТГПУ (TSPU Bulletin). 2017. Вып. 6 (183). С. 49–55.
9. Пропп В. Я. Морфология волшебной сказки: монография. М.: Лабиринт, 2001. 192 с.
10. GigaChat – русскоязычная нейросеть от Сбера. URL: <https://developers.sber.ru/gigachat/login> (дата обращения: 26.02.2024)
11. Столнейкер Р. Прагматика // Новое в зарубежной лингвистике: Лингвистическая прагматика. Вып. XVI. 1985. С. 419–438.
12. Токарева И. А. Структурно-семантическая модель семантико-синтаксического фрейма // Казанский лингвистический журнал. Т. 5. 2022. № 4. С. 547–558.
13. GigaChatPro. URL: <https://web.telegram.org/a/#6218783903> (accessed: 15.12.2023).
14. Goloviznina V., Fishcheva I., Peskicheva T., Kotelnikov E. Aspect-based Argument Generation in Russian // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue”. (2023). Issue 22. URL: <https://www.dialog-21.ru/digest/2023/articles/> (accessed: 15.02.2023).
15. Ji Z., Lee N., Frieske R., Yu T., Su D., Xu Y., Ishii E., Bang Y., Chen D., Chan H.S., Dai W., Madotto A., Fung P. Survey of hallucination in natural language generation // ACM Computing Surveys. Vol. 1. No. 1. Article. URL: <https://www.semanticscholar.org/search?q=Survey%20of%20hallucination%20in%20natural%20generation&sort=r=relevance> (accessed: 15.02.2023)
16. Maynez J., Narayan S., Bohnet B., McDonald R. T. On Faithfulness and Factuality in Abstractive Summarization // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 1906–1919.
17. Mikhaylovskiy N., Churilov I. Autocorrelations Decay in Texts and Applicability Limits of Language Models // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue”. (2023). Issue 22. URL: <https://www.dialog-21.ru/digest/2023/articles/> (accessed: 15.02.2023).
18. Ostyakova L., Petukhova K., Smilga V., Zharikova D. Linguistic Annotation Generation with ChatGPT: a Synthetic Dataset of Speech Functions for Discourse Annotation of Casual Conversations // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue” (2023). Issue 22. URL: <https://www.dialog-21.ru/digest/2023/articles/> (accessed: 15.02.2023).
19. Vychezhninin S., Kotelnikova A., Sergeev A., Kotelnikov E. MaxProb: Controllable Story Generation from Storyline // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue”. (2023). Issue 22. URL: <https://www.dialog-21.ru/digest/2023/articles/> (accessed: 15.02.2023).

References

1. Arutyunova, ND. The sentence and its meaning: logical-semantic problems: monograph. Moscow: Nauka; 1976. 383 p. (In Russ.).
2. Gusarenko SV, Gusarenko MK. Predicate-frame description of the linear semantics of artistic narrative. *Filologicheskie nauki*. 2020;3:18-30. (In Russ.)
3. Deik TA van, Kinch V. Strategies for understanding a coherent text. *Novoe v zarubezhnoi lingvistike: Kognitivnye aspekty yazyka*. 1988;XXIII:153-211. (In Russ.).
4. Demchenkov SA, Fedyaeva ND. Verbal happening: about some models of computer generation of texts in the aspect of avant-garde poetics. *Humanitarian Vector*. 2020;15;1:27-37. (In Russ.).
5. Il'f I, Petrov E. The twelve Chairs. The Golden Calf: Novels. Notebooks. Moscow: Eksmo; 2007. 640 p. (In Russ.).
6. Kozlovskii AV, Mel'nik YaE, Voloshchuk VI. About the approach for automatic generation of plot-related text). *Izvestiya TulGU*. 2022;9:160-167. (In Russ.).
7. Krivosheev NA, Ivanova YuA, Spitsyn VG. Automatic generation of short texts based on the use of LSTM and SecGAN neural networks. *Vestnik Tomskogo gosudarstvennogo universiteta*. 2021;57:118-130. (In Russ.).
8. Kuchina SA. Electronic generative literary texts. *Vestnik TGPU (TSPU Bulletin)*. 2017;6(183):49-55. (In Russ.).
9. Propp VYa. Morphology of a fairy tale. Moscow: Labirint; 2001. 192 p. (In Russ.).
10. GigaChat – Russian-language neural network from Sber. URL: <https://developers.sber.ru/gigachat/login> (accessed: 26.02.2024). (In Russ.).
11. Stolneiker R. *Pragmatics*. *Novoe v zarubezhnoi lingvistike: Lingvisticheskaya pragmatika*. 1985;XVI:419-438. (In Russ.).
12. Tokareva IA. *Structural-semantic model of semantic-syntactic frame*. *Kazanskii lingvisticheskii zhurnal*. 2022;5;4:547-558. (In Russ.).

13. GigaChatPro. URL: <https://web.telegram.org/a/#6218783903> (accessed: 15.12.2023).
14. Goloviznina V, Fishcheva I, Peskischeva T, Kotelnikov E. Aspect-based Argument Generation in Russian in Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue". (2023). Issue 22. URL: <https://www.dialog-21.ru/digest/2023/articles/> (accessed: 15.02.2024).
15. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, Bang Y, Chen D, Chan HS, Dai W, Madotto A, Fung P. Survey of hallucination in natural language generation in ACM Computing Surveys. Vol. 1. No. 1. Article. URL: <https://www.semanticscholar.org/search?q=Survey%20of%20hallucination%20in%20natural%20language%20generation&sort=relevance> (accessed: 15.02.2024).
16. Maynez J, Narayan S, Bohnet B, McDonald RT. On Faithfulness and Factuality in Abstractive Summarization in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 1906-1919.
17. Mikhaylovskiy N, Churilov I. Autocorrelations Decay in Texts and Applicability Limits of Language Models in Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue". (2023). Issue 22. URL: <https://www.dialog-21.ru/digest/2023/articles/> (accessed: 15.02.2024).
18. Ostyakova L, Petukhova K, Smilga V, Zharikova D. Linguistic Annotation Generation with ChatGPT: a Synthetic Dataset of Speech Functions for Discourse Annotation of Casual Conversations in Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue" (2023). Issue 22. URL: <https://www.dialog-21.ru/digest/2023/articles/> (accessed: 15.02.2024).
19. Vychezhninin S, Kotelnikova A, Sergeev A, Kotelnikov E. MaxProb: Controllable Story Generation from Storyline in Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue" (2023). Issue 22. URL: <https://www.dialog-21.ru/digest/2023/articles/> (accessed: 15.02.2024).