

К вопросу о моделировании процесса понимания текста с помощью систем искусственного интеллекта

Т. В. Кружилина¹ ✉

¹ Юго-Западный государственный университет
ул. 50 лет Октября, д. 94, г. Курск 305040, Российская Федерация

✉ e-mail: t.kruzhilina@yandex.ru

Резюме

Целью исследования является проведение сопоставительного анализа текстов, сгенерированных системой искусственного интеллекта в процессе обработки исходного текста на естественном языке, и встречаемых текстов, являющихся результатом понимания человеком исходного художественного текста.

Методы. Для проведения сопоставительного анализа денотатных структур встречаемых текстов автором была применена экспериментальная методика. Участники эксперимента (7 студентов 4 курса факультета дополнительного образования «Переводчик в сфере профкоммуникации», 3 доцента кафедры иностранных языков ЮЗГУ) оценивали 10 встречаемых текстов разной природы – сгенерированные ИИ и человеком на предмет успешности воссоздания смысловой структуры текста.

Результаты эксперимента свидетельствуют в пользу того, что полнота смыслового содержания сгенерированного текста не зависит от структуры исходного текста. Современные методы семантической обработки текста ИИ системой делают возможным получение на выходе полноценных текстовых произведений, созданных с учётом правил и норм естественного языка. ИИ системы успешно воссоздают денотатную структуру текста и реконструируют синтаксическую структуру.

Закключение. Доступ к большим базам данных позволяет обучать нейронную сеть на больших корпусах текстов, что отражается на повышении точности и вариативности используемых лексических единиц и конструктов. Точность передачи смыслового содержания текста варьируется. Она зависит от степени сжатия текста – чем она выше, тем меньше может быть точность, т.к. нейронная сеть не в состоянии классифицировать денотатные связи на предмет значимости для основного смысла. Степень точности передачи смыслового содержания определяется успешностью / неуспешностью понимания глубинного скрытого смысла, которые обусловлены пониманием лингвистического и экстралингвистического контекста. Способность опознать модель ситуации, воссоздаваемую в исходном тексте, является залогом понимания скрытого смысла. ИИ система может довольно корректно и точно воссоздать поверхностную денотатную структуру текста, но на данном этапе развития не способна конструировать модель ситуации

Ключевые слова: когнитивное моделирование; психолингвистика; нейронная сеть; обработка естественного языка; понимание текста; встречаемый текст.

Конфликт интересов: Автор декларирует отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Для цитирования: Кружилина Т. В. К вопросу о моделировании процесса понимания текста с помощью систем искусственного интеллекта // Известия Юго-Западного государственного университета. 2023; 27(4): 98-116. <https://doi.org/10.21869/2223-1560-2023-27-4-98-116>.

Поступила в редакцию 03.11.2023

Подписана в печать 24.11.2023

Опубликована 21.12.2023

On Text Understanding Modeling by Means of Artificial Intelligence Systems

Tatiana V. Kruzhilina¹ ✉

¹ Southwest State University
50 Let Oktyabrya str. 94, Kursk 305040, Russian Federation

✉ e-mail: t.kruzhilina@yandex.ru

Abstract

Purpose of research is to conduct a comparative analysis of texts generated by an artificial intelligence system in the process of processing source text in natural language and counter texts that are the result of human understanding of the source literary text.

Methods. To achieve the goal and objectives of the study, the author used an experimental technique to conduct a comparative analysis of the denotational structures of counter texts. Participants in the experiment (7 4th year students of the Faculty of Additional Education "Translator in the field of professional communications", 3 associate professors of the Department of Foreign Languages of South- West State University) assessed the success of recreating the semantic structure of the text of 10 counter texts of different nature - generated by AI and humans.

Results. The results of the experiment indicate that the completeness of the semantic content of the generated text does not depend on the structure of the source text. Modern methods of semantic text processing by an AI system make it possible to obtain the output of full-fledged text works created taking into account the rules and norms of natural language. AI systems successfully recreate the denotational structure of the text and reconstruct the syntactic structure.

Conclusion. Access to large databases allows you to train a neural network on large text corpora, which results in an increase in the accuracy and variability of the lexical units and constructs used. The accuracy of conveying the semantic content of the text varies. It depends on the degree of text compression - the higher it is, the less accuracy can be, because the neural network is unable to classify denotational connections for relevance to the underlying meaning. The degree of accuracy in conveying semantic content is determined by the success / failure of understanding the deep hidden meaning, which is determined by the understanding of the linguistic and extralinguistic context. The ability to recognize the situation model recreated in the source text is the key to understanding the hidden meaning. The AI system can recreate the surface denotational structure of the text quite correctly and accurately, but is not able to construct a model of the situation at this stage of development.

Keywords: cognitive modeling; psycholinguistics; neural network; natural language processing; understanding of the text; counter text.

Conflict of interest. The author declare the absence of obvious and potential conflicts of interest related to the publication of this article.

For citation: Kruzhilina T. V. On Text Understanding Modeling by Means of Artificial Intelligence Systems. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta = Proceedings of the Southwest State University*. 2023; 27(4): 98-116 (In Russ.). <https://doi.org/10.21869/2223-1560-2023-27-4-98-116>.

Received 03.11.2023

Accepted 24.11.2023

Published 21.12.2023

Введение

Стремительное распространение технологий, задействующих системы искусственного интеллекта (ИИ) для обработки естественного языка, наблюдаемое нами в последние десятилетия, позволяет говорить о революции в методах обработки текстов и изображений. Компьютерная лингвистика, сфера науки, изучающая автоматическую обработку текстов, сделала большой шаг в исследовании применения нейронных сетей для работы с текстом и глубинного обучения нейронных сетей. Именно этим вопросам были посвящены последние конференции ассоциации компьютерной лингвистики (ACL). Пристальное внимание к вопросу развития возможностей нейронных сетей обусловлено перспективностью их применения в самых разных сферах деятельности, предполагающей повышение качества и скорости решения стандартных рутинных задач по классификации и обработке / переработке текстов; снижение трудоёмкости в работе с большими объёмами текстов; решение прикладных задач, например, создание чат-ботов, голосовых помощников, ассистентов-переводчиков и пр. При этом, учитывая тот факт, что технологические перспективы применения новых систем ИИ блестящи и список возможностей этого применения обширен, существует целый ряд ограничений и во-

просов, требующих детального рассмотрения с точки зрения самых разных дисциплин и парадигм. Статья продолжает ряд публикаций автора, посвящённых вопросам понимания текста. В предыдущих работах нами были рассмотрены факторы, обуславливающие процесс формирования и становления способности понимания текста; психолингвистические особенности деятельности понимания; феномен функциональной неграмотности как фактор снижения успешности понимания как записанного текста, так и звучащей речи; стратегии понимания текста и их видоизменение на современном этапе. В работе планируется сравнить результаты понимания текста человеком и нейронной сетью и выявить характерные особенности построения встречного текста человеком и нейронной сетью на настоящем этапе её развития. Для решения задач исследования необходимо рассмотреть процесс понимания текста человеком и процесс воссоздания смыслового содержания исходного текста нейронной сетью.

Нейросетевые технологии в обработке текста

Первыми достижениями в этой области можно назвать работы Уоррена Маккаллока и Уолтера Питтса по глубинному обучению. В середине прошлого века они предложили рассмотреть человеческий мозг в качестве фор-

мальной модели для построения искусственной нейросети. Создание Френком Розенблаттом такой модели на компьютере позволило обобщить огромный объём накопленных знаний о человеческом мозге и дало начало целому направлению исследований по обучению нейронной сети [1]. Первые успехи применения алгоритма обратного распространения ошибки сопровождались сложностями в связи с ограничениями в виде необходимости применения больших объёмов данных для обучения нейросети. Исследования были приостановлены, и лишь 40 лет спустя идеи машинного обучения получили новое развитие с связи с новыми функциональными возможностями и инновациями в области обработки данных.

Архитектура нейронной сети довольно сложна для наивного пользователя. Вершинами сети, которая представляет собой «направленный граф заданной архитектуры», являются нейроны. Количество входных и выходных узлов определяется задачей. Например, решение задачи на классификацию потребует поместить на выход сети один или два нейрона (разделение на два класса) и количество нейронов n при выделении n классов. Остальные уровни называют *скрытыми слоями*¹.

¹ Хайдарова Р.Р. Архитектура и методы вычислений для распределенной сверточной нейронной сети на кластере компьютеров с ограниченными вычислительными ресурсами: дис. ... канд. техн. наук. СПб., 2020. 268 с.

Важным достижением ИИ в этой области считается изменение подхода к обработке текста. Если раньше каждый текстовый элемент, начиная с буквы, требовал описания большим количеством признаков самой разной природы, то на данном этапе развития потребность в сложных описаниях отпала с развитием *representation learning* – обучению представлению. Современная нейросеть способна самостоятельно устанавливать в исходном тексте закономерности и соответствия и использовать эти «знания» при генерировании продукта. Принципиально новым в современных ИИ системах является способность не привязываться к конструированию огромного количества признаков через нахождение внешних баз данных, концентрируясь на адекватном выборе источников баз данных и предварительной работе с текстом для последующего обучения нейросети, которое задействует большие объёмы данных, чем стандартные техники обработки текста. Задачи по автоматической обработке текста, решаемые с помощью нейронных сетей, довольно разнообразны (рис. 1).

Одним из перспективных направлений работы с текстом в этой области специалисты признают исследование дистрибутивной семантики, например, установление значения слова исходя из контекста (построение вектора слова), которое производится с помощью математической модели. Привлечение больших корпусов текстов позволяет выде-

лить для определённого слова максимальное количество контекстов коими могут считаться соседние слова и слова из одной синтаксической конструкции. В клетках матрицы фиксируется частотность употребления слова в определённом контексте или используется коэффициент PPMI (Positive Pointwise Mutual Infor-

mation – коэффициент положительной попарной взаимной информации), который показывает уровень randomness употребления слова в определённом контексте. Такого рода матрицы широко применяются для кластеризации слов и установления близких по значению групп слов [2].

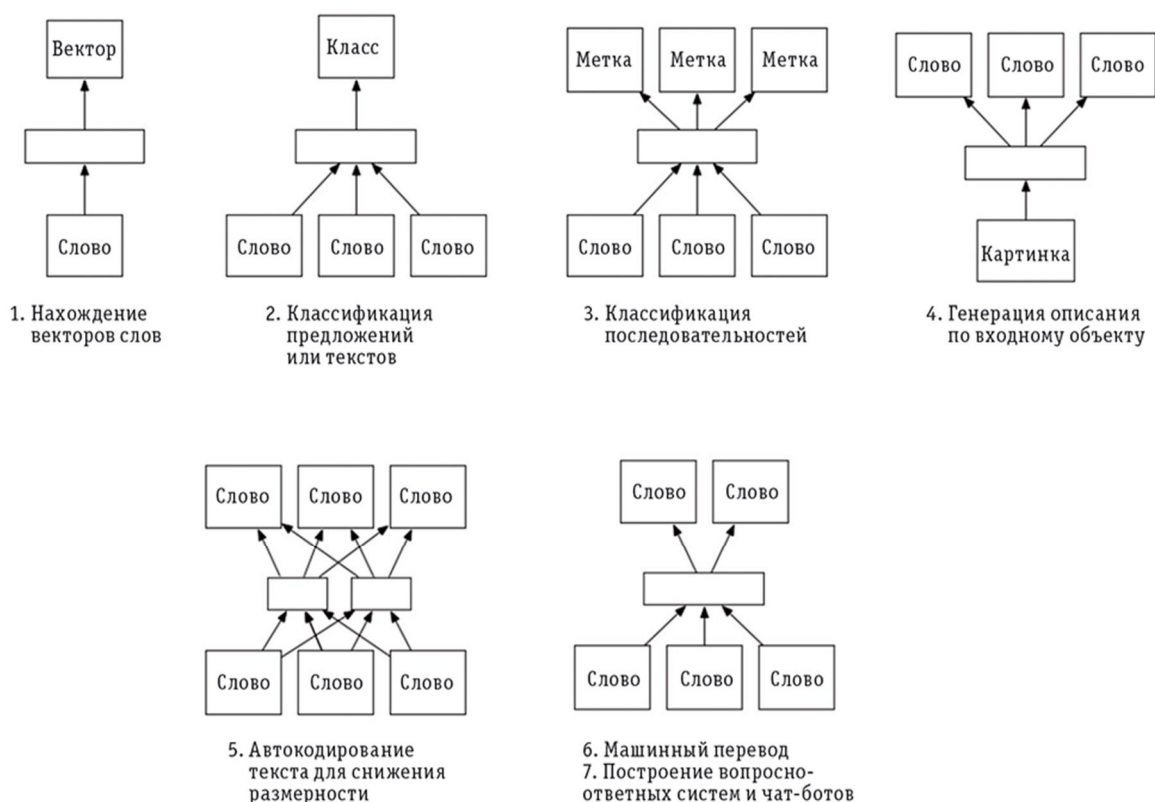


Рис. 1. Задачи по автоматической обработке текста, решаемые с помощью нейросетей

Fig. 1. Automatic text processing tasks solved with the help of neural networks

Начиная с 2013 г. технология построения векторов слов, согласно работе [3], используется при глубинном обучении нейронных сетей для установления семантической близости слов, простых логических выводов. В работе описываются две архитектуры, skip-gram и Continuous Bag of Words, объ-

единённые в общий термин word2vec. Как объясняется в [3], этот термин подразумевает функционал матрицы «слово-контекст» с весами коэффициента положительной попарной взаимной информации (PPMI). В настоящий момент word2vec относят скорее к полю дистрибутивной семантики, нежели к глу-

бинному обучению [4]. При этом векторы оказались востребованным инструментом для классификации текстов исходя из удобства представления смысла слова, подаваемого на вход глубинным нейросетям.

На сегодняшний день задача обработки текстов является одной из самых востребованных, поэтому попытки оптимизировать процесс и повысить качество генерируемого продукта инициируют новые подходы и методы. Так, для решения проблемы обработки нейронной сетью длинных предложений на входе был применён опыт использования свёрточных нейронных сетей для анализа изображения. На вход сети поступает предложение, в котором слова уже представлены векторами слов, созданными с помощью word2vec. Этот вид нейронных сетей хорошо зарекомендовал себя благодаря относительной простоте и способности к обучению с применением алгоритма обратного распространения ошибки. Для компьютерной лингвистики это эффективный и работающий метод обработки текстов, имеющий между тем свои ограничения, которые могут быть отмечены у всех ИИ систем. Речь идёт о механическом подходе к обработке языка.

Процесс генерации смысла текста нейронной сетью и человеком

Процесс восприятия текста человеком относительно хорошо исследован мировой и отечественной психологической, психолингвистической и лингви-

стической наукой [5 – 12]. Механизмы восприятия, порождения и понимания текста многократно исследованы автором данной статьи, результаты экспериментальных исследований отражены в публикациях [13, 14]. Согласно В.П. Белянину, восприятие речи человеком, «...включает в себя рецепцию слышимых или видимых элементов языка, установление их взаимосвязи и формирование представления об их значении» [15, с.154]. Понимание же есть расшифровка смысла, стоящего за воспринимаемым речевым сообщением. При этом смысл, лежащий за внешней формой, может варьироваться и зависеть от экстралингвистического контекста, от ситуации восприятия. «И во всех случаях языковые знаки являются предикатом к действительности, к разным ситуациям» [15, с. 155]. Исходя из указанной психолингвистической составляющей процесса понимания становится понятным существование феномена глубины понимания и разного качества понимания. Понимание человеком носит креативный и индивидуальный характер. Каждый субъект понимания видит в тексте определённую сугубо индивидуальную ситуацию, что-то своё, недоступное другому субъекту понимания того же объекта (текста или речевого сообщения). При этом поэтапное установление смысловых связей в совокупности рождает единое смысловое содержание. Его осмысление и даёт в итоге понимание либо непонимание (частичное понимание / недостаточно

глубокое понимание) письменного текста или речи.

Задача настоящей публикации – показать принципиальное отличие механизмов обработки текстовой информации человеком от подобного процесса, осуществляемого ИИ системой, поэтому не будем подробно останавливаться на уровнях понимания и его избирательности.

Принципиально иной подход в генерации смысла используется ИИ системой. Рассмотрим подробнее, как осуществляется семантическая обработка текста, представляющая собой анализ языковых единиц разного объёма. Это развитая нейросетями способность не просто опознавать слова, но устанавливать связи между словами в предложении и между предложениями в тексте. Для успешности семантической обработки используются следующие методы: токенизация, лемматизация, удаление стоп-слов, синтаксический анализ, семантический анализ [16].

Токенизация – это процесс выделения отдельных слов, токенов. Речь здесь не идёт о выделении денотатов, токен может быть словом любой самостоятельной части речи (кроме предлога).

Лемматизация предполагает приведение языковой единицы к начальной форме, например, к форме единственного числа у имён существительных или неопределённой форме у глагола.

Удаление стоп-слов предполагает удаление языковых единиц, не несущих большой смысловой нагрузки. Данная

операция позволяет нейронной сети сконцентрироваться на ключевой информации.

Синтаксический анализ предполагает восстановление синтаксической конструкции и предикативных связей согласно правилам построения предложения используемого естественного языка.

Семантический анализ представляет собой самый сложный этап работы, так как нейронная сеть «воссоздаёт» смысл предложений и текста с разной степенью успешности, принимая во внимание законы сочетаемости слов, ряды синонимов / антонимов, контекст и ассоциативные связи. Поправка о разной степени успешности подразумевает проблемы и ограничения, возникающие у нейронных сетей на настоящем этапе развития. Так, неспособность нейронных сетей к пониманию скрытого смысла является одним из самых больших ограничений при работе с художественными текстами наряду с «непониманием» иронии и юмора. Трудности корректного считывания контекста и полная невозможность установления экстралингвистического контекста делают семантическую обработку текста механическим препарированием текста, выполняемым с помощью довольно виртуозного жонглирования языковыми единицами. Впрочем, речь идёт о состоянии способностей нейронных сетей в настоящий момент времени. Прогноз к улучшению качества осуществляемой обработки текста вполне благоприятен, учитывая огромные ежегодные инвестиции в эту сферу.

В данной статье делается попытка проанализировать тексты, полученные в результате обработки нейронной сетью исходного художественного текста, и сравнить их со встречаемыми текстами, полученными в процессе психолингвистического эксперимента на понимание текста.

Материалы и методы

Материалами исследования послужили встречаемые тексты, полученные в ходе психолингвистического эксперимента на понимание текста, проведённого среди студентов 1 и 2 курса специальностей «Правоохранительная деятельность», «Программная инженерия», «Нанотехнологии и наносистемы», «Электроэнергетика и электротехника» (52 участника) и преподавателей кафедры иностранных языков ЮЗГУ (3-е участников), всего – 55 человек. Исходный текст, понимание которого проверялось, представляет собой образец короткого художественного рассказа, перевод с английского языка рассказа «Appointment with Love» («Свидание с любовью») автора S. I. Kishor (Сумалит Иш-Кишор) [17]. В результате первого этапа эксперимента участниками эксперимента было записано 55 встречаемых текстов. Для этапа эксперимента, где непосредственно сравнивались результаты понимания текста человеком и ИИ системой, из 55 встречаемых текстов, записанных человеком (студентами и преподавателями – испытуемыми), было отобрано 7.

Дополнительно с помощью ИИ системы было сгенерировано 3 текста. Использовались 2 разных ресурса, предлагающих услуги по машинной обработке текста. Сгенерированные тексты представляли собой продукты разной степени сжатия. Ресурс 1 «выдал» тексты с уровнем сжатия 40% и 80%. Ресурс 2 обработал текст, сжав содержание на 40%. Целью такой обработки было, во-первых, проследить, как и насколько отличаются тексты, сгенерированные разными ресурсами; во-вторых – как и насколько отличаются тексты с разным уровнем сжатия информации.

Для проведения эксперимента на сравнение встречаемых текстов была привлечена группа из 10 человек, 7 студентов 4 курса факультета дополнительного образования ЮЗГУ «Переводчик в сфере профессиональной коммуникации» и трое экспертов, преподавателей кафедры иностранных языков ЮЗГУ (кандидаты филологических наук, доценты). Задание на проведение сопоставительного анализа встречаемых текстов (далее экспериментальные тексты не будут разделяться на встречаемые, т. е. те, что были записаны студентами как результат понимания, и тексты – результаты машинной обработки ИИ системой; далее все 10 текстов будут называться встречаемыми для упрощения идентификации) является трудоёмкой задачей, требующей определённых навыков обработки текстового материала. Для решения проблемы низкого экспертного уровня привлечённых студен-

тов были задействованы студенты 4 курса, обучающиеся по программе подготовки переводчиков, имеющие навыки работы с текстом и характеризующиеся хорошей языковой подготовкой. Перед экспериментальной группой (имеются в виду все 10 участников) не стояло задачи проведения сопоставительного анализа научными методами (в отличие от автора статьи, которая для анализа текстов использовала инструментарий лингвистики и психолингвистики). Вопросы, на которые следовало ответить, были подготовлены для «наивного пользователя», человека, умеющего читать и понимать прочитанное. Экспертная группа преподавателей (трое доцентов кафедры иностранных языков ЮЗГУ) была привлечена в качестве контрольной силы, выполняющей задание на понимание текста эталонным образом.

На начальном этапе все участники эксперимента ознакомились с содержанием рассказа «Свидание с любовью», прочитав текст на русском языке. Далее испытуемые оценивали встречные тексты различной природы и оценивали каждый текст по четырём критериям. Материалами эксперимента выступили 10 встречных текстов, результатов понимания исходного текста: 7 встречных текстов студентов и преподавателей и 3 результата обработки исходного текста ИИ системой. Приведём 4 примера использованных текстов. Тексты 1, 10 представляют собой пример обработки текста ИИ системой со сжатием 40 и 80%. Тексты 3, 7 являются встречными

текстами, результатом понимания студентом (3) и преподавателем (7). Исходный текст здесь не приводится, ссылка на него содержится в тексте данной статьи выше.

1. Лейтенант авиации, стоя на станции Гранд Централ, ожидает встречи с женщиной, чьи поддерживающие письма сопровождали его военные подвиги. Напряженность перед встречей охватывает его, несмотря на обычное чувство контроля. Вспоминая самый тяжелый воздушный бой, лейтенант вспоминает слова женщины из ее письма, которые придали ему сил и уверенности. Стремясь услышать ее настоящий голос, он внимательно наблюдает за потоком людей на станции, надеясь на скорое приближение момента встречи.

Пилот Блэндфорд обнаруживает книгу "Время страстей человеческих", на полях которой написаны размышления Холлис Мейнелл. Он находит ее адрес, пишет письмо, и начинается переписка. В течение тринадцати месяцев Холлис отвечает на его письма, даже когда они задерживаются. Блэндфорд убежден, что любит ее, и она в ответ. Однако она отказывается отправлять свою фотографию, утверждая, что внешность не имеет значения. Она предлагает встретиться лично в Нью-Йорке, чтобы решить судьбу их отношений после встречи.

Лейтенант Блэндфорд видит молодую женщину, которая привлекает его внимание своей красотой. Он шага-

ет к ней, но затем замечает Холлис Мейнелл, стоящую позади. Это женщина за сорок, но он чувствует к ней близость по духу. Он решает пригласить ее на ужин, испытывая горечь разочарования после ухода молодой женщины. Однако оказывается, что молодая красавица просила прикрепить розу к пальто Холлис и сообщить, что она ждет Блэндфорда в ресторане через дорогу, в рамках какого-то испытания. В итоге, Холлис соглашается помочь, и лейтенант отправляется на свидание с молодой женщиной.

10. Молодой лейтенант авиации пришел на вокзал за шесть минут до встречи с женщиной, с которой переписывался последние тринадцать месяцев, но ни разу не видел. Его сердце билось от волнения. Он встал возле справочной, чтобы встретить ее. Лейтенант вспомнил самый тяжелый воздушный бой, когда ее письмо с цитатой из псалма придало ему сил. Теперь он впервые услышит ее голос. За шесть минут до встречи мимо прошла девушка с цветком, но это был не тот цветок, который они условились.

Холлис Мейнелл отказывалась прислать свою фотографию, чтобы проверить искренность чувств. Лейтенант нашел ее адрес на книге "Время страстей человеческих", написал письмо и началась их переписка. Теперь он верил, что любит ее. За минуту до встречи к нему подошла красивая женщина, но позади стояла Холлис -

полная женщина за сорок в потрепанной одежде с розой на пальто.

Лейтенант колебался, но увидев сочувствие на ее лице, решился. Он поприветствовал ее, но она передала, что розу ей дала девушка в зеленом пальто, которая ждет его в ресторане напротив. Это было испытание от Холлис.

3. Однажды в Нью-Йорке молодой лейтенант авиации, только что вернувшийся с войны, стоял на станции, и пристально смотрел на часы. Как выяснилось он ждал женщину, которую ни разу не видел, но очень активно общался с ней письмами. Познакомился он с ней совершенно случайно, взяв книгу в библиотеке он заметил на полях приписки оставленные женщиной, на обложке книги было написано её имя. Он нашёл её номер по телефонному справочнику, и так началась их долгая переписка.

И вот время близится к свиданию. Они договорились, что у него в руке будет та самая книга, а у неё на плече роза. Тут мимо него прошла высокая, грациозная, красивая девушка, а за ней стояла полная женщина с розой на плече. Тут лейтенанта начало разрывать, он не знал, что ему делать, не так он всё это представлял. Ему хотелось побежать за той девушкой. Но подумав, он понял, что так поступать нельзя, и что возможно вместо любви выйдет прекрасная дружба. Он подошёл к женщине пригласил её на ужин. Но вдруг она сказала ему, что та прекрасная девушка дала мне эту розу, и сказа-

ла, что если вы пригласите меня на ужин, то она будет ждать Вас в кафе через дорогу. Это была проверка на настоящую любовь.

7. На вокзале Гранд Централ суета. Столько людей, а её нет. А будет ли она? Может. он всё придумал? Нет, невозможно такое придумать. Она писала на полях книги не для того, чтобы понравиться... Да и книга-то, книга... Не Золушка и не книга рецептов... Бремя страстей человеческих... Как женщина могла так точно видеть душу мужчины? Её пометки стали откровением, пропустить такой шанс встретить близкого тебе человека – значит быть идиотом ... А письма, её письма! Они стали его щитом, его бронёй там, в небе... Псалом 90, благословенные слова... Она сама стала его благословением. И вот. теперь он её увидит. Как жаль, что она отказалась прислать свою фотографию! У ангела должно быть лицо! Красивое лицо! Молодое лицо! Алая роза не видна нигде... Стоп, ах нет, это горошек, да и девочка юная совсем, не 29. Но... хороша. Это невыносимо, хочется курить. Она, да, точно она... Как хороша! Волосы, глаза, сама весна... Она, точно она! Нет... Или всё же? Хочешь что-то, солдатик? Ещё бы! Не уходи! Алая роза, нет сомнения... Алая роза на пиджаке. На пиджаке женщины... Полная, измученная что-ли. Как же так? Ну как же так? Живой в помощи Вышняго... Решено. Здравствуйте, Вы Холлис Мей-

нелл? Будем друзьями. Нет, не Холлис? Ах, чертовка... В ресторане напротив, значит, ждёт... Спасибо, дамочка, спасибо мамочка, здоровья твоим двум парням!

Результаты и их обсуждение

Для анализа понимания текста хорошо себя зарекомендовал денотатный анализ текста, разработанный А.И. Новиковым [18]. Суть метода заключается в качественном и количественном анализе установленных в процессе понимания денотатов и межденотатных связей, которые фиксируются либо с помощью графа, либо в таблице. Автором данной статьи денотатный анализ применялся в работе [13]. Приведём пример текста и денотатного графа, воссоздающего смысловую структуру текста. Отметим сразу, что текст, приводимый в качестве примера, был составлен для эксперимента с детьми дошкольного возраста, он примитивен с точки зрения простоты семантических и синтаксических связей. Тем не менее, он наглядно иллюстрирует, как можно проверить понимание, учитывая установленные испытуемым межденотатные связи.

Текст III

Айболит сидит за столом и выписывает рецепт. На его столе стоит телефон. Перед Айболитом на стульчике сидит мышка с перевязанной лапкой. За спиной Айболита стоит плачущая лиса. На кушетке лежит пёс с завязанным носом. Подмышкой у пса –

градусник. Около шкафа с лекарствами
стоят корова и медведь, у них болит
горло. Медведь держит в лапах банку с
витаминами.

Пример построения графа простого нарративного текста представлен на рис. 2.

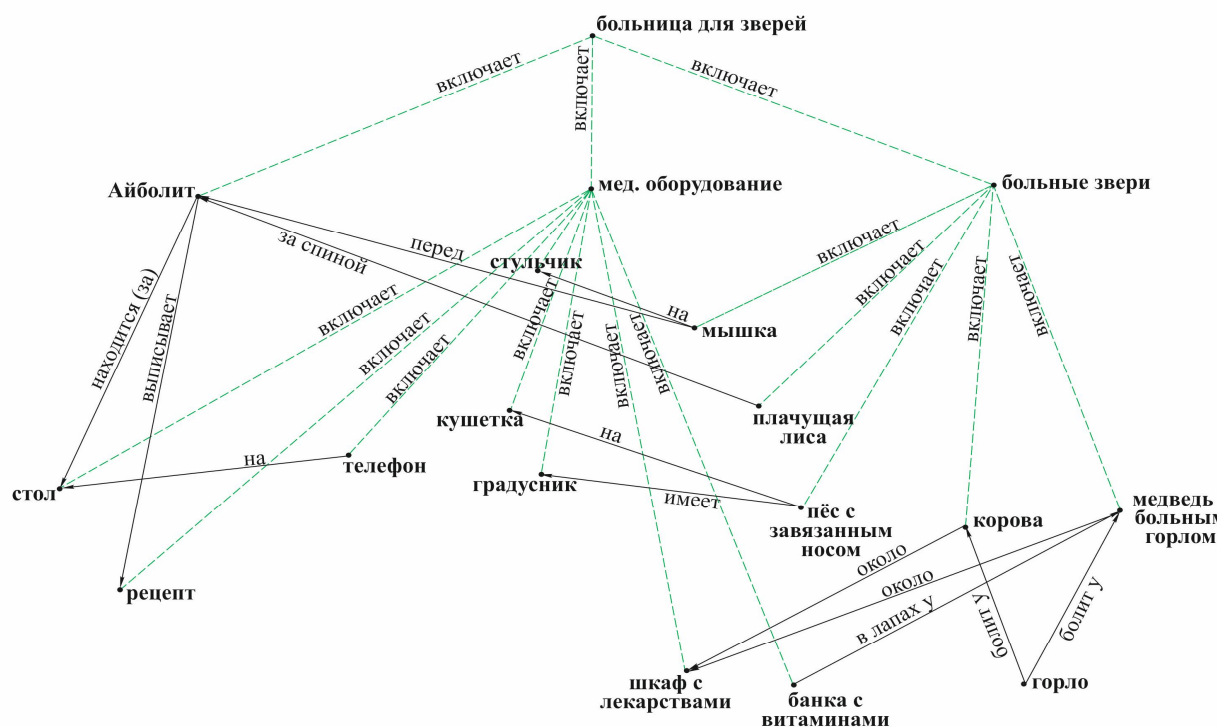


Рис. 2. Эталонный граф денотатной структуры текста III

Fig. 2. Reference graph of the denotational structure of text III

Этот метод может быть отчасти валиден при анализе текстов, сгенерированных ИИ системой. В эксперименте, который описывается в настоящей статье, в текстах 1, 10 успешно воссозданы как денотаты, так и связи между ними. Очевидно, что с помощью семантического анализа ИИ прекрасно устанавливает в тексте денотаты и межденотатные связи. Но глубинный, или скрытый, смысл (в случае нашего экспериментального текста это частично скрытый смысл), может остаться «непонятым».

Об этом говорит фраза (Текст 1): «*В итоге, Холлис соглашается помочь, и лейтенант отправляется на свидание с молодой женщиной*». ИИ система идентифицирует женщину, согласившуюся помочь, как молодую женщину. Но по замыслу автора рассказа Холлис и молодая женщина – это одно лицо, а женщина, согласившаяся помочь – это полноватая женщина за сорок с седеющими волосами. Этот факт не установлен ИИ системой, так как вывод о том, какая именно женщина является на самом де-

ле главной героиней по имени Холлис Мейнелл сложно сделать с помощью семантического анализа, используемого нейронными сетями.

Рассматривая надёжность денотатного анализа как метода исследования, следует также отметить, что он не вполне показателен в случае креативного подхода к созданию встречного текста. Это вариант текста 7. Преподаватель-участница эксперимента передала содержание рассказа от лица лейтенанта. Встречный текст 7 прекрасно воссоздаёт душевное состояние и мысли лейтенанта Блэнфорда, но не конструирует ситуацию, пропуская много деталей, и количество установленных денотатных связей гораздо меньше по сравнению с текстами 1, 10.

Обратимся непосредственно к эксперименту на сравнение встречных текстов. Перед началом экспериментальной сессии участникам объяснили значение понятий «полнота» и «точность» передачи содержания текста (первый и второй вопросы экспериментального задания). Полнота содержания предполагает упоминание максимального количества смысловых связей и отношений. Точность передачи содержания предполагает восстановление количества смысловых связей, достаточного для передачи содержания с сохранением ведущих смысловых линий и основного заложенного автором смысла. Концепция задания согласуется с теорией текста А.И. Новикова (см., напр. [18]), теорией понимания Н.И. Жинкина

(см., напр. [19]) и целой школой их последователей, к которым относит себя и автор статьи. Третий вопрос из четырёх предложенных (Можно ли по встречному тексту определить, понял ли автор встречного текста, какая из женщин – Х.М.?) может показаться чрезмерным исходя из очевидности положительного ответа. Сразу оговоримся, что, во-первых, при анализе текстов, сгенерированных ИИ системой, нельзя полагаться на веру и знания о понимании текстов человеком. Как было показано ранее в данной публикации, ИИ системы обрабатывают тексты механически, минуя стадию воссоздания ситуации, стадию, которая является не просто необходимой, но ключевой при понимании текста и которая делает возможным проникновение в глубинные смысловые структуры [20]. Интрига, «вложенная» С. Иш-Кишором, была с лёгкостью понята почти всеми участниками эксперимента (речь идёт о первом этапе эксперимента на понимание с написанием встречных текстов, в котором приняло участие 52 студента. В настоящей работе мы выбрали 7 встречных текстов из 52 имеющихся). Что касается текстов, сгенерированных ИИ системой, из трёх имеющихся текстов интрига была явно разгадана в одном, в остальных двух «понимание» интриги не носило очевидного характера. Четвёртый вопрос на установление авторства имел целью удовлетворение исследовательского интереса автора, вызванного желанием понять, имеют ли современные студен-

ты и преподаватели представление о текстах, сгенерированных ИИ, и смогут ли они отличить встречные тексты разной природы. Результаты экспериментальной сессии представлены в табл. 1. Тексты 1, 4, 10 сгенерированы ИИ си-

стемой, остальные написаны человеком (тексты 2, 3, 6, 8 – студентами, 5, 7, 9 – преподавателями). Мы предложили оценить ответ на вопрос по шкале от 1 до 10 и вычислили средний результат (табл. 1).

Таблица 1. Анализ встречных текстов, созданных человеком и сгенерированных ИИ системой

Table 1. Analysis of the counter texts created by a human and those generated by an AI system

№ текста / text №	Насколько полно передано содержание текста? / Text content completeness	Насколько точно передано содержание текста? / Text content fidelity	Понял ли автор встречного текста, какая из женщин - Х.М.?/ Has the counter text author understood which of the women is H. M.?	Авторство встречного текста – человек или ИИ система? / Is the counter text author – AI or a Human
текст 1	8,7	8,4	да – 1	человек – 5
			нет – 8	
			затрудняюсь ответить – 1	ИИ – 5
текст 2	7,2	6,3	да – 10	человек – 8
			нет – 0	
			затрудняюсь ответить – 0	ИИ – 2
текст 3	6,9	8,1	да – 10	человек – 4
			нет – 0	
			затрудняюсь ответить – 0	ИИ – 6
текст 4	4,9	6,1	да – 0	человек – 2
			нет – 7	
			затрудняюсь ответить – 3	ИИ – 8

Окончание табл. 1 / Table 1 (ending)

№ текста / text №	Насколько полно передано содержание текста? / Text content completeness	Насколько точно передано содержание текста? / Text content fidelity	Понял ли автор встречного текста, какая из женщин - Х.М.? / Has the counter text author understood which of the women is H. M.?	Авторство встречного текста – человек или ИИ система? / Is the counter text author – AI or a Human
текст 5	7,4	9,2	да – 10	человек – 9
			нет – 0	
			затрудняюсь ответить – 0	ИИ – 1
текст 6	7,9	6,9	да – 6	человек – 8
			нет – 0	
			затрудняюсь ответить – 4	ИИ – 2
текст 7	5,1	9,6	да – 10	человек – 10
			нет – 0	
			затрудняюсь ответить – 0	ИИ – 0
текст 8	4,1	6,3	да – 9	человек – 7
			нет – 1	
			затрудняюсь ответить – 0	ИИ – 3
текст 9	9,8	9,9	да – 10	человек – 9
			нет – 0	
			затрудняюсь ответить – 0	ИИ – 1
текст 10	7,7	6,4	да – 9	человек – 7
			нет – 0	
			затрудняюсь ответить – 1	ИИ – 3

Очень показательными представляются результаты анализа текста 1 (текста, сгенерированного ИИ системой со степенью сжатия 40%). Оценки полноты и точности воссоздания смыслового содержания высоки – 8,7 и 8,4 соответственно. При этом интрига автора не раскрыта, женщина-друг по переписке не установлена правильно. Это отметили 8 участников эксперимента из 10. 1 затруднился с ответом, 1 ответил неверно. Оpoznать текст, написанный нейронной сетью, смогли только 5 из 10. Можно с полной уверенностью констатировать, что на фоне эффективной семантической обработки текста ИИ система не справляется с задачей установления глубинных смысловых структур и скрытого смысла.

Определить авторство встречного текста для участников эксперимента оказалось трудновыполнимой задачей. Испытуемым намеренно не сообщалось, сколько текстов написано нейронной сетью. Как показывают результаты, в сознании наивных пользователей языка и пользователей-профессионалов не сформировано чётких критериев, по которым с точностью опознаётся текст, сгенерированный ИИ системой. Так, встречные тексты, написанные преподавателями, характеризующиеся нормативным и корректным использованием языка, правильным построением предложений, отсутствием речевых, пунктуационных и орфографических ошибок, набрали 28 голосов из 30 в пользу написания человеком. То есть среднестатистический пользователь языка не очень

высокого мнения о качестве текстов, генерируемых нейронной сетью.

Креативный подход к написанию встречного текста (текст 7) не был критичен при оценке этого текста – пересказ от лица лейтенанта не снизил точности воссозданного смыслового содержания (9,6), полнота содержания при этом составила всего 5,1. Авторство текста 7 не вызвало ни у кого сомнений, все участники эксперимента опознали в авторе человека.

Выводы

Сопоставление встречных текстов, созданных в результате понимания текста человеком и в результате обработки текста на естественном языке нейронной сетью, позволяет сделать ряд выводов. Современные методы семантической обработки текста ИИ системой делают возможным получение на выходе полноценных текстовых произведений, созданных с учётом правил и норм естественного языка. Полнота смыслового содержания сгенерированного текста не зависит от структуры исходного текста. ИИ системы успешно воссоздают денотатную структуру текста и реконструируют синтаксическую. Доступ к большим базам данных позволяет обучать нейронную сеть на больших корпусах текстов, что отражается на повышении точности и вариативности используемых лексических единиц и конструктов. Точность передачи смыслового содержания текста варьируется. Она зависит, во-первых, от степени сжатия текста – чем она выше, тем меньше может быть точность, т.к. нейронная сеть

не в состоянии классифицировать денотатные связи на предмет значимости для основного смысла. Во-вторых, степень точности передачи смыслового содержания в немалой степени определяется успешностью / неуспешностью понимания глубинного скрытого смысла. Понимание скрытого смысла тесно связано с пониманием лингвистического и экстралингвистического контекста. Способность опознать модель ситуации, воссоздаваемую в исходном тексте, является залогом понимания скрытого смысла. ИИ система может довольно корректно и точно воссоздать поверхностную денотатную

структуру текста, но скрытые детали, меняющие в корне всё понимание текста, могут оказаться упущенными. На данном этапе развития ИИ системы не способны конструировать модель ситуации. Механизмы генерирования смысла текста ИИ системой путём механической семантической обработки, в корне отличающиеся от механизмов понимания текста человеком, чётко определяют направление развития и сферу применения компьютерных систем обработки естественного языка на ближайшую перспективу.

Список литературы

1. Горбачевская Е.Н., Краснов С.С. История развития нейронных сетей. URL: <http://cyberleninka.ru/article/n/istoriya-razvitiya-neyronnyh-setey>.
2. Mikolov T. et. al. Efficient Estimation of Word Representations in Vector Space, arxiv.org. URL: <http://arxiv.org/pdf/1301.3781.pdf>.
3. Бобырь М. В. Обучение нейро-нечетких систем. М.: АРГАМАК-МЕДИА, 2017. 240 с.
4. Велихов П.Е. Машинное обучение для понимания естественного языка // Открытые Системы. СУБД. 2016. № 1. С.18–21. URL: <https://www.osp.ru/os/2016/01/13048649>.
5. Ананьев Б.Г. Психология чувственного познания. М.: Изд-во Акад. пед. наук РСФСР, 1960. 486 с
6. Блонский П.П. Память и мышление. М.: URSS: ЛЕНАНД, 2017. 204 с.
7. Бородай С. Ю. Язык и познание: Введение в пострелятивизм. М., 2019. 800 с.
8. Выготский Л. С. Мышление и речь. СПб.: Питер, 2019. 431 с.
9. Залевская А.А. Введение в психолингвистику. М.: РГГУ, 2007. 559 с.
10. Соколов А. Н. Внутренняя речь и мышление. М.: URSS, 2007. 248 с.
11. Мягкова Е.Ю. Навык чтения и проблемы понимания текста // Языковое бытие человека и этноса. Вып. 20. Материалы XV Березинских чтений. М.: ИНИОН РАН, 2018. С. 83–95.
12. Мягкова Е. Ю. Язык - эмоции - сознание - культура в теории А. Дамасио // Вопросы психолингвистики. 2022. № 4(54). С. 72-85. DOI 10.30982/2077-5911-2022-54-4-72-85.
13. Кружилина Т.В. Понимание текста в онтогенезе. Курск: Юго-Западный государственный университет, 2016. 152 с.

14. Кружилина Т. В. Факторы, определяющие особенности речевого развития у детей // Известия Юго-Западного государственного университета. 2012. № 4–1(43). С. 216–219.
15. Белянин В.П. Психолингвистика. М.: Флинта, 2013. 416 с.
16. Manning Ch. Computational linguistics and deep learning. Computational Linguistics, 2016. URL: http://www.mitpressjournals.org/doi/full/10.1162/COLI_a_00239#.WQH8MBhh2qA.
17. Иш- Кишор С. Свидание с любовью. URL: http://sacralis.com/biblioteka/proza/2605-svidanie_s_lubovju.html?ysclid=ls0tcuhl7u170031708.
18. Новиков А.И. Семантика текста и ее формализация. М.: Наука, 1983. 215 с
19. Жинкин Н.И. О кодовых переходах во внутренней речи. М.: Культура, Акад. Проект, 2006. С. 251– 263.
20. Звягинцева В.В. Обращение в английском семейном дискурсе (экспериментальное исследование) // Известия Юго-Западного государственного университета. Серия: Лингвистика и педагогика. 2012. № 2. С. 48–51.

References

1. Gorbachevskaya E.N., Krasnov S.S. *Istoriya razvitiya neyronnyh setej* [History of Neural Networks Development] [Elektronnyj resurs]. Available at: <http://cyberleninka.ru/article/n/istoriya-razvitiya-neyronnyh-setey>.
2. Mikolov T. et. al. *Efficient Estimation of Word Representations in Vector Space*, arxiv.org. URL: <http://arxiv.org/pdf/1301.3781.pdf> (data obrashcheniya: 24.01.2024).
3. Bobyr' M. V. *Obuchenie nejro-nechetkih sistem* [Training of Neuro-Fuzzy Systems]. Moscow, ARGAMAK-MEDIA, 2017. 240 p.
4. Velihov P.E. Mashinnoe obuchenie dlya ponimaniya estestvennogo yazyka [Machine Learning for Natural Language Understanding]. *Otkrytye Sistemy.SUBD*, 2016, no. 1, pp.18–21. Available at: <https://www.osp.ru/os/2016/01/13048649>.
5. Anan'ev B.G. *Psihologiya chuvstvennogo poznaniya* [Psychology of sensual cognition]. Moscow, 1960. 486 p.
6. Blonskij P.P. *Pamyat' i myshlenie* [Memory and cognition]. Moscow, URSS: LENAND Publ., 2017. 204 p.
7. Borodaj S. Yu. *Yazyk i poznanie: Vvedenie v postrelyativizm* [Language and cognition: Introduction to postrelativism]. Moscow, 2019. 800 p.
8. Vygotskij L. S. *Myshlenie i rech'* [Cognition and speech]. Saint Petersburg: Piter Publ., 2019. 431 p.
9. Zalevskaya A.A. *Vvedenie v psiolingvistiku* [Introduction to Psycholinguistics]. Moscow, RGGU Publ., 2007. 559 p.
10. Sokolov A. N. *Vnutrennyaya rech' i myshlenie* [Inner speech and cognition]. Moscow, URSS Publ., 2007. 248 p.

11. Myagkova E.YU. Navyk chteniya i problemy ponimaniya teksta [Reading skill and text understanding]. *Yazykovoe bytie cheloveka i etnosa. Vyp. 20. Materialy XV Berezinskih chtenij.* = *The linguistic existence of a person and an ethnic group. Issue 20. Materials of the XV Berezinsky readings.* Moscow, INION RAN Publ., 2018, pp. 83–95.
12. Myagkova E. Yu. Yazyk - emocii - soznanie - kul'tura v teorii A. Damasio [Language-emotion-recognition-culture in the theory of A. Damasio]. *Voprosy psiholingvistiki* = *Questions of Psycholinguistics*, 2022, no. 4(54), pp. 72-85. DOI 10.30982/2077-5911-2022-54-4-72-85.
13. Kruzhilina T.V. *Ponimanie teksta v ontogeneze* [Text Understanding in Ontogenesis]. Kursk, Yugo-Zapadnyj gosudarstvennyj universitet Publ., 2016. 152 p.
14. Kruzhilina T. V. Faktory, opredelyayushchie osobennosti rechevogo razvitiya u detej [Factors that Determine the Characteristics of Speech Development in Children]. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta* = *Proceedings of the Southwest State University*, 2012, no. 4–1(43), pp. 216–219.
15. Belyanin V.P. *Psiholingvistika* [Psycholinguistics]. Moscow, Flinta Publ., 2013. 416 p.
16. Manning Ch. Computational linguistics and deep learning. *Computational Linguistics*, 2016. Available at: http://www.mitpressjournals.org/doi/full/10.1162/COLI_a_00239#.WQH8MBhh2qA
17. Ish- Kishor S. *Svidanie s lyubov'yu* [Appointment with Love]. Available at: http://sacralis.com/biblioteka/proza/2605-svidanie_s_lubovju.html?ysclid=ls0tcuh17u170031708
18. Novikov N.A. *Semantika teksta i ee formalizaciya* [Text Semantics and its Formalization]. Moscow, Nauka Publ., 1983. 215 p.
19. ZHinkin N.I. *O kodovyh perekhodah vo vnutrennej rechi* [About code transitions in internal speech]. Moscow, Kul'tura Publ., Akad. Proekt, 2006, pp. 251– 263.
20. Zvyaginceva V.V. Obrashchenie v anglijskom semejnomo diskurse (eksperimental'noe issledovanie) [Address in English family discourse (experimental study)]. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Lingvistika i pedagogika* = *Proceedings of the Southwest State University. Series: Linguistics and Pedagogics*, 2012, no. 2, pp. 48–51.

Информация об авторе / Information about the Author

Кружилина Татьяна Владиславовна, кандидат филологических наук, доцент, доцент кафедры иностранных языков, Юго-Западный государственный университет, г. Курск, Российская Федерация, e-mail: t.kruzhilina@yandex.ru, ORCID: <https://orcid.org/0000-0001-5714-146X>, SPIN-код: 8142-0010, Researcher ID: 738898

Tatiana V. Kruzhilina, Cand. of Sci. (Phylologic), Associate Professor, Associate Professor of the Foreign Languages Department, Southwest State University, Kursk, Russian Federation, e-mail: t.kruzhilina@yandex.ru, ORCID: <https://orcid.org/0000-0001-5714-146X>, SPIN-код: 8142-0010, Researcher ID: 738898