

УДК 81:004.032.26

DOI 10.17726/philIT.2022.1.2



Слово в техногенном многомерном пространстве

Быльева Дарья Сергеевна,

кандидат политических наук, доцент,

Санкт-Петербургский политехнический университет

Петра Великого,

г. Санкт-Петербург, Россия

bylieva_ds@spbstu.ru

Аннотация. Сегодня искусственный интеллект активно осваивает естественные языки, становясь собеседником и партнером человека в разных аспектах деятельности. Однако символичный подход, подразумевающий передачу правил и логики, потерпел фиаско, а количество правил и исключений языка не позволяет провести его формализацию, поэтому современное «глубокое обучение» искусственных нейронных сетей подразумевает самостоятельный поиск закономерностей в обширных базах данных. В ходе обучения искусственный интеллект ставит слово в предложение, чтобы синтагматические отношения были максимально приближенными к таковому у целевого слова в базе, учитывая как семантические связи слов, так и отношения между словами в последовательности изложения. «Язык» информационных технологий цифровой. При работе с естественным языком слова представляются в векторной форме как последовательность чисел. Идея представлять слова математически знакома людям и ассоциируется, как правило, с логической непротиворечивостью. Визуализация положения слов в многомерном пространстве, созданном искусственным интеллектом, демонстрирует ряд закономерностей, очевидных семантических и синтаксических взаимосвязей, однако суть других отношений между словами неочевидна. Математическое представление слов, созданное искусственным интеллектом, может позволить взглянуть на язык с новой, нечеловеческой точки зрения.

Ключевые слова: искусственный интеллект; язык; машинный перевод; слово; word2vec.

Word in technogenic multidimensional space

Bylieva Daria,

PhD, associated professor,

*Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russia*

bylieva_ds@spbstu.ru

Abstract. Today, artificial intelligence is actively mastering natural languages, becoming an interlocutor and partner of human in various aspects of activity. However, the symbolic approach, which implies the transfer of rules and logic, has failed, the number of rules and exceptions of the language does not allow its formalization, so modern «deep learning» of artificial neural networks involves an independent search for patterns in extensive databases. During training, artificial intelligence puts a word into a sentence so that the syntagmatic relationships are as close as possible to those of the target word in the base, taking into account both the semantic relationships of words and the relationships between words in the sequence of presentation. The «language» of information technologies is digital. During natural language processing, words are represented in vector form as a sequence of numbers. The idea of representing words mathematically is familiar to people and is usually associated with logical consistency. Visualization of the position of words in a multidimensional space created by artificial intelligence demonstrates a number of patterns, obvious semantic and syntactic relationships, but the essence of other relationships between words is not obvious. The mathematical representation of words, created by artificial intelligence, can allow you to look at the language from a new non-human point of view.

Keywords: artificial intelligence; language; Machine translate; word; word2vec.

Введение

Технологии изменяют жизнь человека во множестве аспектов. Цифровизация трансформирует все формы и методы взаимодействия в области работы, развлечения, обучения, покупок и т.п. Однако технологии искусственного интеллекта (ИИ) не только способствуют решению широкого круга практических задач, но и выдвигают многочисленные вопросы, требующие осмысления. Активно вступая во взаимодействие с человеком, становясь работ-

ником и партнером, искусственный интеллект заставляет по-другому взглянуть на традиционные философские проблемы и ставит вопрос о понятиях: что такое интеллект, сознание, любовь, творчество и т.п.

Интересным для осмысления является вопрос освоения искусственным интеллектом естественного языка. Активное становление ИИ как собеседника, языковая имитация человеческих, межличностных отношений [1] изменяет привычные социальные практики, но также позволяет взглянуть на естественный язык с новой, «нечеловеческой», точки зрения [2].

Естественный и математический языки

Семиотические системы являются важным средством отражения мира и передачи информации. Они устанавливают связи между знаками и внешним миром. Естественный язык – открытая динамическая знаковая система, наиболее близкая человеку. Существуют не только социокультурные, но и социобиологические версии происхождения языка. Так, Ноам Хомский утверждал, что синтаксические знания, по крайней мере частично, являются врожденными, так как уже ребенок способен понимать и производить потенциально бесконечное количество новых комбинаций из известных компонентов [3].

Значимость слов в человеческом сообществе очень высока, хотя представление о роли языка может варьироваться, его срединная позиция между сознанием и миром не позволяет дать однозначные интерпретации. Как отмечает Дэвид Кристал, два диаметрально противоположных представления о словах мирно сосуществуют в нашем отношении к языку: одно состоит в том, чтобы рассматривать слова как неадекватное представление мыслей и опасное отвлечение от эмпирической реальности, другое заключается в том, чтобы рассматривать их как незаменимые для выражения мыслей и как средство наведения порядка в мире [4, р. 24].

В то же время общепринятым является представление, что в основе научного знания и технологического прогресса лежит язык математики. Как отмечает С. А. Лебедев, математика обеспечивает формальное единство, разрабатывая для всех областей и видов научного знания количественный язык описания научных объектов любого рода [5]. Однако язык математики сложен,

как подчеркивал Гуго Штейнгауз, в математике «несравненно отчетливее, чем в других науках, проявляется то, как растянут прогресс человечества. Сегодня одновременно с нами на Земле живут люди, с точки зрения понимания математики принадлежащие эпохе более древней, чем египетские пирамиды, и их значительное большинство; небольшой процент добрался до средних веков, до XVIII века не дошел и один из тысячи» [6, р. 375]. «Язык» информационных технологий цифровой. Процесс цифровизации затронул огромные пласты человеческой культуры. Аналоговые данные переводятся в серии единиц и нулей, которые могут считываться и подвергаться компьютерной обработке. Компьютерные программы представляют собой формальные математические объекты со строгой, однозначной семантикой. Но в отличие от других математических структур исполняемость программы помещает ее непосредственно в физический мир [7]. «Общение» с компьютерами на соответствующем математическом языке доступно только специалистам, поэтому возникла задача перевода естественного языка на математический язык современных технологий для обеспечения гладкого человеко-машинного взаимодействия, что породило область обработки естественного языка (NLP), складывающуюся из понимания естественного языка (NLU) и генерации естественного языка (NLG). Задача работы с текстами на естественном языке включает в себя распознавание речи (звуковые волны нужно сегментировать на слова, проверить и исправить ошибки), анализ семантической паутины, декомпозицию естественного языка на семантические части, синтез речи, диалоговые системы, анализ тональности и эмоций, анализ информации о пользователе и т.д.

Естественный язык является прекрасно разработанной системой, служащей многообразным человеческим целям. Синтаксис и семантика естественного языка очень сложны, существует огромное количество правил и исключений, отсутствует единообразие в представлении – предложения, несущие одинаковое значение, могут быть представлены во множестве различных синтаксисов, наблюдается множество двусмысленностей, слова и предложения могут иметь много разных значений, и эти значения зависят от контекста. Отдельной проблемой для ИИ являются представления о реальном мире. Хотя в языке заложены представления о мире, тем не менее языковая и физическая реальность далеко не идентичны. И последняя передается в языке искаженно, в ней отсутствуют наи-

более очевидные и явные явления и отношения в мире. Семантика включает знания, относящиеся к здравому смыслу (обычно приобретаемые в детстве), и общие знания (накапливаемые в течение всей жизни). Э. Камбрия и Б. Уайт приводят в качестве примера общих знаний о мире выражение «стул – это тип мебели», а в качестве примера здравого смысла такие очевидные или общепринятые вещи, которые люди знают о мире, но которые не упоминаются в дискурсе, например убеждение в том, что предметы падают вниз (а не вверх) и люди улыбаются, когда они счастливы [8]. Более подробно эти не представленные в явном виде в естественном языке знания включают числа (численные операции и операции над множествами), пространство (геометрия и навигация), физику (неодушевленные объекты и механика) и психологию (агенты и группы). Исследователи предлагают различать интуитивную физику и интуитивную психологию [9] video games, and board games, achieving performance that equals or even beats that of humans in some respects. Despite their biological inspiration and performance achievements, these systems differ from human intelligence in crucial ways. We review progress in cognitive science suggesting that truly human-like learning and thinking machines will have to reach beyond current engineering trends in both what they learn and how they learn it. Specifically, we argue that these machines should (1. Даже совсем маленькие дети ожидают, что неодушевленные объекты будут следовать принципам постоянства, непрерывности, сплоченности и прочности (что объекты должны двигаться по плавным траекториям, а не мигать и исчезать, не проникать друг в друга и не действовать на расстоянии) [10] и что агенты будут действовать целенаправленно, эффективно и социально обоснованно [11].

Таким образом, препятствиями для полноценного человеко-машинного взаимодействия на естественном языке становятся как невозможность его формализации, так и сложные соотношения между языком и реальностью, которая имеет неполную или искаженную репрезентацию. Такое явление, как машинный перевод на основе правил, существовало в течение нескольких десятилетий, затем стал использоваться статистический перевод. Однако значительный прогресс в обработке естественного языка произошел только при использовании технологий искусственного интеллекта. «Глубокое обучение» не требует формулировки знаний в виде правил и закономерностей. Нейросетевой подход позволил

обучать ИИ языку не на основе представлений о логике и языке, а с помощью баз данных, значимые отношения и закономерности в которых ИИ ищет самостоятельно.

Слова как числа

Обучаясь, ИИ, как правило, ставит слово в определенное место предложения из лингвистических корпусов, так, чтобы синтагматические отношения были максимально приближены к таковому у целевого слова, роль играет близость значения и парадигматические отношения между знаками, которые могут занимать одно и то же место в сходных контекстах. Следовательно, для построения адекватной лингвистической картины нужно учитывать как семантические связи слов, так и отношения между словами в последовательности изложения. Помимо семантических словарных отношений существуют отношения в синтагме, как в линейной структуре речи, обусловленной семантико-грамматически-фонетической сочетаемостью. Положение знаков в синтагме предполагает отношение сходства или несходства между альтернативными знаками, которые имеют право занимать одно и то же положение [12, р. 44]. Нейронный перевод не имеет ничего общего с переводом одного слова или предложения с одного языка на другой. Вместо этого одна сеть кодирует слово на одном языке в векторную форму последовательности чисел, а другая декодирует последовательность чисел в слово на другом языке. Таким образом, каждое слово представлено сотнями чисел. Представить даже примерно столь многомерное векторное пространство человеку невозможно.

Тем не менее, идея представления слов математически является привлекательной для человека. Одной из самых известных является универсальный язык Лейбница, на который можно было бы переводить мысли, вычислять и проверять рассуждения. В одном из вариантов философ предлагал приписывать термину в позиции субъекта высказывания пару произвольных, не имеющих общего делителя натуральных чисел, которые ставились в соответствие паре чисел предиката так, что они образовывали правильные дроби в случае истинного высказывания, составленного из таких субъекта и предиката. В сложном понятии «характером» будет произведение чисел, обозначающих составные признаки. Например, если сопоставить термин «животное» с числом 2, а термин

«разумное» – с числом 3, то термину «человек» будет соответствовать произведение $2 \times 3 = 6$ [13]. Числовые эквиваленты слов представлялись доказательством истинности суждений, Лейбниц писал: «с помощью чисел мы сможем тотчас же судить, доказаны ли они или нет; и то, что другие [смогли сделать] с величайшим напряжением ума или случайно, мы достигаем с помощью одних лишь характеристических знаков и точного, истинно аналитического метода; поэтому то, на что в ином случае смертным потребовались бы многие тысячелетия, мы сможем совершить за столетие» [14, p. 515]. Таким образом, математическое представление языка тесно связано с идеей сделать его логически точным. Лихтенберг писал: «я уже давно мечтаю о том, чтобы существовал такой язык, на котором невозможно было бы сказать ложь, или, как минимум, в котором любая погрешность против истины была бы также и грамматической» [15]. Однако в реальности естественный язык сопротивляется математизации.

Возникшая в 1920-1930-е гг. структурная семантика утверждала, что семантический анализ языковой единицы не должен ограничиваться отдельными словами, а должен также принимать во внимание своих «соседей». Фердинанд де Соссюр считал, что язык – это система взаимосвязанных единиц и структур и что каждая единица языка связана с другими в рамках той же системы. Йост Трир и др., отказавшись от «атомистических взглядов» традиционной семасиологии, разработали тип структурной семантики списков, которая называлась теорией концептуальных («Begriffsfelder»), лексических («Wortfelder») или семантических полей («Bedeutungsfelder») [16, p. 120]. Согласно Чжоу, семантическое поле – это набор слов, которые взаимодействуют, доминируют, различаются и зависят друг от друга [17]. Слово не является обособленным носителем значения; напротив, каждое имеет значение только потому, что к нему примыкают другие [18]. П. Гарденфорс пытается теоретически представить абстрактные мысленные образы с присущей им пространственной структурой, построенные из элементарных топологических и геометрических форм. В книге «Концептуальные пространства: геометрия мысли» описывается, как информация может быть представлена точками, векторами и областями в объемных пространствах, при этом отношения подобия могут быть смоделированы в терминах расстояний [19]. Концепт рассматривается как набор областей в ряде измере-

ний вместе с информацией о том, как соотносятся области в разных измерениях [20, р. 3]. П. Гарденфорс, Ф. Ценкер утверждают, что концептуальные пространства подходят для представления различных видов отношений подобия: чем ближе два объекта расположены в концептуальном пространстве, тем больше они похожи; «зеленый», например, ближе к «синему», чем к «красному». Кроме того, если предполагается, что измерения имеют метрику, можно говорить о расстояниях в концептуальном пространстве, так что расстояния представляют собой степени сходства между объектами, представленными в пространстве [21, р. 4]. Концептуальное пространство состоит из ряда качественных измерений. Примерами качественных измерений являются температура, вес, яркость, высота звука и сила, а также три обычных пространственных измерения: высота, ширина и глубина. Например, яблоко может быть оценено по ряду параметров: цвет, форма, текстура, вкус и т.д. (таблица 1).

Таблица 1

**Вербальное описание положения «яблока»
в многомерном пространстве**

Домен	Область
цвет	красно-желто-оранжево-красный
форма	Округлый
текстура	Гладкий
вкус	области кисло-сладкого измерения
пищевая ценность	показатели содержания сахара, клетчатки, витаминов и т.д.
особенность фрукта	особенности структуры семян, мякоти, типа кожуры и т.д.

Визуально можно представить концептуальные пространства в трехмерном измерении. По нескольким параметрам можно условно обозначить положение в виде точки, другие представлены в взаимозависимых измерениях, которые не могут быть логически разделены в перцептивном пространстве, например «цвет» определяется многомерно (оттенок, насыщенность, яркость) (рисунок 1).

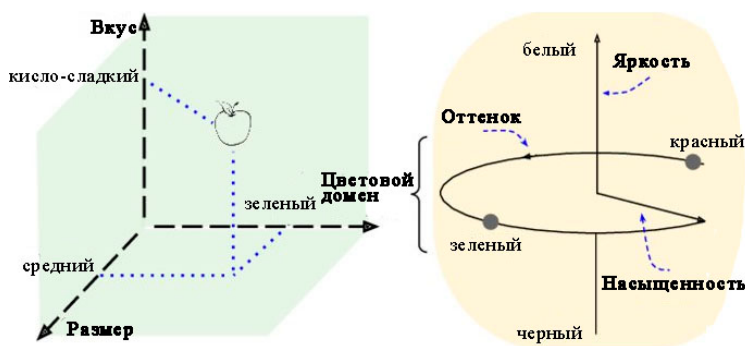


Рисунок 1. Пример концептуального пространства для представления концепта «яблоко» [22]

В таком пространстве можно представить сходные понятия, сравнив их между собой (например, на рисунке 1 можно расположить другие фрукты). Геометрическое представление концептуальных пространств дает возможность математической оценки сходства и различия понятий с помощью мер расстояний.

Числовая репрезентация слов ИИ может быть осуществлена и представлена различными способами. Один из первых методов обработки языка, появившийся в 2013 году, Word2vec представляет каждое отдельное слово определенным списком чисел, называемым вектором. Векторы построены таким образом, чтобы простая математическая функция (косинусное сходство между векторами) указывала на уровень семантического сходства между словами, представленными этими векторами. Эти отношения устанавливают значение знаков в системе языка, определяя их взаимное сходство и несходство, согласно разным порядкам отношений [23].

Искусственный интеллект присваивает слову сотни числовых значений, но их можно визуализировать в низкоразмерном пространстве. На рисунке 2 триста векторов значений, соответствующих токенам слов, были уменьшены до трех измерений с помощью PCA (анализ основных компонентов). Такое представление позволяет человеку с помощью положения (расстояния и направления) в векторном пространстве увидеть семантические связи.

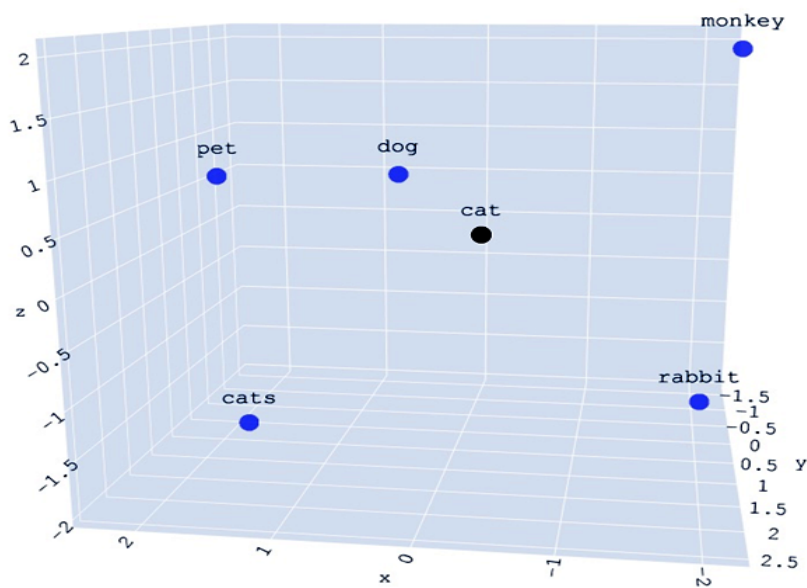


Рисунок 2. Пять ближайших соседей к токенту «cat» [12, р. 47]

Маломерное – трехмерное пространство дает возможность увидеть, как ИИ группирует семантически сходные элементы вместе, а непохожие – далеко друг от друга (рисунок 3). Исследование непрерывного языкового пространства методом векторного смещения (малоразмерные представления на уровне слов) показывает одинаковость векторных смещений между словами, различающимися по роду или форме единственного/множественного числа, то есть векторы демонстрируют семантические и синтаксические отношения между словами.

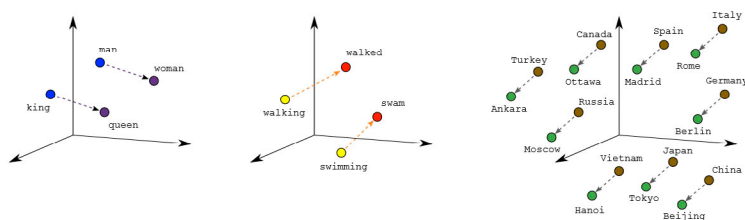


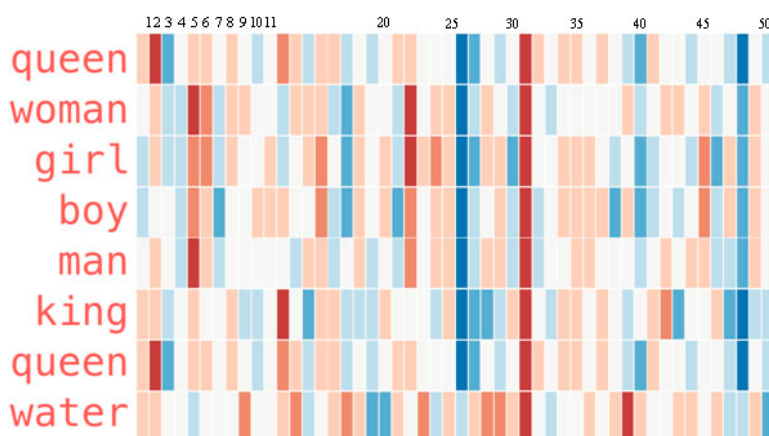
Рисунок 3. Представления в трехмерном пространстве [24]

В синтаксически насыщенных моделях перевода с английского на немецкий язык исследователи обнаруживают соответствующие закономерности. Например, слово «Fater» (с нем. «отец») относится в основном к соседним словам «his» и «father» в базовой модели, а также к более отдаленным словам «Bwelle» (человек) и «escorting» в синтаксической модели [25] for instance, is an illustrative example that generates abstract representations of tokens inputted to an encoder based on their relationships to all tokens in a sequence. Recent studies have shown that although such models are capable of learning syntactic features purely by seeing examples, explicitly feeding this information to deep learning models can significantly enhance their performance. Leveraging syntactic information like part of speech (POS). Так как мы имеем дело с числами, то возможно осуществлять математические операции со словами. Например, сложение подчас может давать осмысленные результаты: *vec* («Россия») + *vec* («река») близко к *vec* («река Волга»), а *vec* («Германия») + *vec* («столица») близко к *vec* («Берлин») [26].

Таким образом, анализируя выстраиваемое ИИ числовое представление слов, исследователи обнаруживают понятные связи, убеждаясь в «правильности» представления языка. Можно ли сказать, что «ученик» выучил урок хорошо, если может группировать слова по смыслу и знает столицы государств? Подобные вопросы хорошо вписываются в тесты на интеллект для младшего школьника. Однако, в отличие от школьника, ИИ не получал никаких представлений о том, что такое страна, река, кошка и животное, и не имел таблички соответствий государств и столиц. Символьный подход, состоящий в обучении ИИ на основании человеческих представлений о собственном когнитивном процессе, то есть на высокоуровневых символических (читаемых человеком) представлениях о задачах, логике и поиске, потерпел фиаско. Продуктивным оказался коннекционистский подход, базирующийся не на предложенной человеком лингвистической логической картине мира, а на биологической стороне мышления (передаче сигнала от дендритов к аксонам). Семейство искусственных нейронных сетей представляет особый интерес, поскольку их гибкая структура позволяет модифицировать их для самых разных контекстов во всех типах машинного обучения. Имитирующие принципы обработки информации в биологических системах, нейронные сети состоят из математически связанных блоков обработки, называемых ис-

кусственными нейронами. Подобно синапсам в мозгу, каждое соединение между нейронами передает сигналы, сила которых может быть усилена или ослаблена весом, который постоянно регулируется в процессе обучения. Сигналы обрабатываются последующими нейронами только в том случае, если превышен определенный порог, определяемый функцией активации. Таким образом, не человек «объясняет» ИИ словами и символами представления о мире и языке. ИИ нужен только большой объем данных, и он строит свои схемы без явного программирования. Поэтому числовое представление о словах заслуживает более внимательного рассмотрения, чем удовлетворение от того, что ИИ «понял» семантические связи языка без явного обучения. Действительно ли в рамках цифрового представления языка выстроились столь эффектно демонстрируемые математические связи между словами?

Чтобы более четко представить числовую репрезентацию слов искусственным интеллектом, можно использовать тепловую карту (на рисунке 4 ячейки окрашены цветом в зависимости от их значений: красный – если они близки к 2, белый – если они близки к 0, синий – если они близки к -2).



**Рисунок 4. Тепловая карта
50-мерного представления слов [27]**

Тепловая карта позволяет увидеть некоторые математические закономерности. Все выбранные слова (queen (королева), woman

(женщина), girl (девочка), boy (мальчик), man (мужчина), king (король), water (вода)) имеют одно общее измерение – яркую красную линию (№ 31). Ярко синяя линия (№ 26) в середине соединяет все слова, обозначающие людей (то есть исключая water). Видны полосы, объединяющие woman, girl, boy, man (4, 18, 25, 48, 49) и демонстрирующие разделение между ними: так, woman и girl объединены слоями 2, 3, 6, 7, 8, 10, 12, 21, 22, 23, 28, boy и man – слоями 6, 8, 9, 16, 24, 32, girl и boy – слоями 1, 5, 11, 13, 15, 20, 34, 40, 41, 43, woman и man – слоями 1, 5, 15, 20. Хорошо видно объединяющие особ королевской крови (для этого queen повторена после king) – 5, 8, 10, 16, 17, 29, 39, 50. Из чего делаются выводы о том, что нейросеть справилась с выявлением признака живого, признаками королевской власти, возраста, пола и т.п. Это вызывает восторг у исследователей. Однако можно представить ситуацию по-другому: когда не ИИ, а человек оказывается в роли экзаменуемого. Общее между мальчиком, женщиной и мужчиной (7, 24, 27, 46, 30) или нечто, отличающее женщину от всех остальных обозначений людей (35), уже не кажется нам очевидным, хотя можно подобрать несколько более или менее правдоподобных вариантов. А необходимость отыскать объединяющее короля, королеву и воду (1, 26) против остального набора слов может вовсе поставить в тупик. По крайней мере, от традиционных вопросов для проверки интеллектуального развития детей («найди лишнее слово среди перечисленных») мы перешли к задачкам «с подвохом», вроде «чем отличается пьяница от дерева».

Заключение

Вывод, который мы можем сделать, пользуясь имеющимися способами представления векторного значения слов, которые применяются ИИ при оперировании языком, заключается в том, что искусственным интеллектом выделены определенные закономерности, некоторые из которых для людей также являются важными с точки зрения представления о языке и мире, другие нам неочевидны. Сама идея представления слов в виде чисел доступна человеку и связана с семантикой и логикой, хотя словам не хватает четкости, определенности и рациональности, позволившей бы людям полноценно сопоставить их с числами. Результаты, которые демонстрируют нейронные сети последнего поколения, показыва-

ют, что применение присущей им математической репрезентации слов позволяет достаточно эффективно пользоваться языком.

В данной статье рассматривался только один аспект нейросетевой обработки языка, связанный с векторным представлением слов, которое в некотором роде позволяет приоткрыть «черный ящик» ИИ. В то время как существуют и другие аспекты машинного обучения, применяемые для других целей в рамках языковой аналитики. Значительный прогресс в обработке языка искусственным интеллектом означает много больше, чем повышение удобства человеко-машинного интерфейса, так как язык тесно связан с человеком и миром. Прогресс в языке, с одной стороны, выводит отношения человека и ИИ на новый уровень, с другой – расширяет возможности применения ИИ в разных областях деятельности. Как отмечают К. Черч и М. Либерман, ИИ строит прогнозы на основе нашего текущего понимания мира и сообщает, как изменить это понимание; учитывая, что мир редко бывает таким, каким мы его понимаем в настоящее время, это может сыграть важную роль в сокращении разрыва между нашими убеждениями и реальностью [28].

Литература

1. *Ullmann L.* The quasi-other as a Subject // *Technology and Language*. – 2022. – № 1(3). – P. 76-81. – URL: <https://doi.org/10.48417/technolang.2022.01.08>.
2. *Bylieva D.* Language of AI // *Technology and Language*. – 2022. – № 1(3). – P. 111-126. – URL: <https://doi.org/10.48417/technolang.2022.01.11>.
3. *Chomsky N.* Syntactic Structures. – Berlin: Mouton, 1957. – 116 p. – URL: <https://doi.org/10.1515/9783112316009>.
4. *Crystal D.* The lure of words // *The Oxford handbook of the word*. – Oxford: Oxford University Press, 2015. – P. 23-28.
5. Лебедев С. А. Математика и технические науки – основа целостности современного научного знания // *Гуманитарный вестник*. – 2018. – № 72(10). – С. 22-48. (*Lebedev S. A.* Mathematics and technical sciences are the basis of the integrity of modern scientific knowledge // *Gumanitarnyj vestnik*. – 2018. – Vol. 72(10). – S. 22-48.)
6. Штейнгауз Г. Задачи и размышления. – М.: Мир, 1974. – 168 с. (*Shteingauz G.* Tasks and reflections. – M.: Mir, 1974. – 168 s.)
7. *Hähnle R.* Program and Code // *Technology and Language*. – 2022. – № 3(2). – URL: <https://doi.org/10.48417/technolang.2022.02.06>.
8. *Cambria E., White B.* Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article] // *IEEE Computational Intelligence Magazine*. – 2014. – № 2(9). – P. 48-57. – URL: <https://doi.org/10.1109/MCI.2014.2307227>.

9. *Lake B. M., Ullman T.D., Tenenbaum J.B., Gershman S.J.* Building machines that learn and think like people // *Behavioral and Brain Sciences*. 2017. – (40). – e253. – URL: <https://doi.org/10.1017/S0140525X16001837>.
10. *Spelke E. S., Gutheil G. & Van de Walle G.* The development of object perception // *An invitation to cognitive science: vol. 2. Visual cognition*. 2nd ed. Bradford, 1995. – P. 297-330.
11. *Spelke E. S., Kinzler K.D.* Core knowledge // *Developmental Science*. 2007. № 1(10). P. 89-96. – URL: <https://doi.org/10.1111/j.1467-7687.2007.00569.x>.
12. *Capone L.* Which Theory of Language for Deep Neural Networks? Speech and Cognition in Humans and Machines // *Technology and Language*. – 2021. – № 4(2). – P. 29-60. – URL: <https://doi.org/10.48417/technolang.2021.04.03>.
13. *Кузнецов В.Г.* Интенциональная силлогистика Г.В. Лейбница и ее роль в истории логики // *Вестник Московского университета. Серия 7. Философия*. – 2017. – № 4. – С. 3-18. (*Kuznecov V. G.* I Intensional syllogistics of G. V. Leibniz and its role in the history of logic // *Vestnik Moskovskogo universiteta. Serii 7. Filosofii*. – 2017. – Vol. 4. – S. 3-18.)
14. *Лейбниц Г.В.* Сочинения в четырех томах: Т. 3. – М.: Мысль, 1984. – 734 с. (*Lejbnic G. V.* Works in four volumes: Vol. 3. – M.: Mysl', 1984. – 734 s.)
15. *Нордманн А.* Языковое мышление и мышление языка у Георга Кристофа Лихтенберга: «где ... любая погрешность против истины была бы также и грамматической» // *Семиотические исследования*. – 2021. – № 4. – С. 29-38. – URL: <https://doi.org/10.18287/2782-2966-2021-1-4-29-38>. (*Nordmann A.* Linguistic thinking and language thinking in Georg Christoph Lichtenberg: “where ... any error against the truth would also be grammatical” // *Semioticheskie issledovaniia*. – 2021. – Vol. 4. – S. 29-38.)
16. *Fritz G.* Theories of meaning change: An overview // *Semantics – Typology, Diachrony and Processing*. De Gruyter, 2019. – P. 113-146.
17. *Faraj G. A. K.* Semantic Field of Utterances in «‘Healthy Living Guide’» // *International Journal on Humanities and Social Sciences*. – 2022. – № 32. – P. 186-197. – URL: <https://doi.org/10.33193/IJoHSS.32.2022.400>.
18. *Trier J.* Der deutsche Wortschatz im Sinnbezirk des Verstandes. Von den Anfängen bis zum Beginn des. Jahrhunderts. – Heidelberg: Carl Winter Universitätsverlag, 1973.
19. *Gärdenfors P.* Conceptual spaces: The geometry of thought. – Cambridge, MA: MIT Press, 2009.
20. *Gärdenfors P.* Semantics Based on Conceptual Spaces // *Logic and Its Applications. ICLA 2011. Lecture Notes in Computer Science*, vol 6521. Cham: Springer, 2011. – P. 1-11. – URL: https://doi.org/10.1007/978-3-642-18026-2_1.
21. *Gärdenfors P., Zenker F.* Conceptual Spaces at Work // *Applications of Conceptual Spaces*, Synthese Library 359. – Cham: Springer, 2015. –

- P. 3-13. – URL: https://doi.org/10.1007/978-3-319-15021-5_1.
22. *Banaee H., Loutfi A.* Using Conceptual Spaces to Model Domain Knowledge in Data-to-Text Systems // Proceedings of the 8th International Natural Language Generation Conference (INLG). – Cham: Springer, 2014. – P. 11-15. URL: <https://doi.org/10.3115/v1/W14-4403>.
 23. *Mikolov T., Chen K., Corrado G., Dean J.* Efficient estimation of word representations in vector space // 1st International Conference on Learning Representations, ICLR2013 – Workshop Track Proceedings. ICLR, 2013. – ArXiv ID: 1301.3781.
 24. Prepare Your Data. – URL: <https://developers.google.com/machine-learning/guides/text-classification/step-3>.
 25. *Sundararaman D., Subramanian V., Wang G., Si S., Shen D., Wang D., Carin L.* Syntax-Infused Transformer and BERT models for Machine Translation and Natural Language Understanding. 2019. – ArXiv ID: 1911.06156.
 26. *Mikolov T., Sutskever I., Chen K., Corrado G.S., Dean J.* Distributed Representations of Words and Phrases and their Compositionality // Advances in Neural Information Processing Systems 26 (NIPS2013). Neurips, 2013. – P. 3111-3119. – URL: <https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf>.
 27. *Alammar J.* The Illustrated Word2vec. – 2019. – URL: <https://jalammar.github.io/illustrated-word2vec>.
 28. *Church K., Liberman M.* The Future of Computational Linguistics: On Beyond Alchemy // Frontiers in Artificial Intelligence. – 2021. – Vol. 4. – URL: <https://doi.org/10.3389/frai.2021.625341>.