

ОСМЫСЛЯ ПРОСТРАНСТВО: КАК ИСПОЛЬЗУЮТ СХЕМЫ ТОПОЛОГИЧЕСКОЙ СЕМАНТИКИ ЕСТЕСТВЕННЫЕ И ИСКУССТВЕННЫЕ КОГНИТИВНЫЕ АГЕНТЫ (НА МАТЕРИАЛЕ ОПИСАНИЙ КАРТИН ИЗ КОЛЛЕКЦИИ ЭРМИТАЖА)*

Колмогорова Анастасия Владимировна – доктор филологических наук, профессор, заведующий лабораторией языковой конвергенции. НИУ «Высшая школа экономики – Санкт-Петербург». Российская Федерация, 190068, г. Санкт-Петербург, наб. канала Грибоедова, д. 119–121; e-mail: akolmogorova@hse.ru

Налобина Полина Алексеевна – стажер-исследователь лаборатории языковой конвергенции. НИУ «Высшая школа экономики – Санкт-Петербург». Российская Федерация, 190068, г. Санкт-Петербург, наб. канала Грибоедова, д. 119–121; e-mail: pnalobina@edu.hse.ru

Статья посвящена описанию различий в концептуализации пространства, наблюдаемых у информантов, генеративных языковых моделей и моделей компьютерного зрения, способных генерировать текст описания «увиденного». Мы используем понятие когнитивного агента и обосновываем различие «естественный vs искусственный когнитивный агент»: под первым понимается человек, под вторым – модель ИИ, способная принимать решения и выполнять задачи адекватно заданной ситуации. Цель исследования – сравнить способы осмысления местоположения объекта в пространстве, в том числе относительно других объектов, у естественных когнитивных агентов (ЕКА) и искусственных когнитивных агентов (ИКА) двух типов: больших языковых моделей, способных генерировать текст, и моделей, созданных для решения задачи Image to Text. Основными методами выступили метод лингвистического эксперимента и метод семантического описания на основе теории топологической семантики Л. Талми. В качестве стимульного материала использовались шесть картин из коллекции Государственного Эрмитажа, разделенные на три группы: портреты, монофигурные картины на мифологическую или религиозную тематику, многофигурные композиции. Участниками экспериментов выступили: 63 информанта (Mean age = 19,1, 48 девушек, 15 юношей), 5 мультязычных генеративных моделей, 6 моделей Image to Text, основанных на технологии компьютерного зрения и способных генерировать описания распознанных изображений на английском языке. Используя типологию конфигурационных топологических схем и схем типа «фигура – фон» в концепции топологической семантики Л. Талми, мы сравнили способы осмысления пространства, на которые опираются а) большие языковые модели; б) модели Image to Text и в) информанты. В результате мы сформулировали ряд выводов, главный из которых заключается в том, что ЕКА отличаются от ИКА своей способностью интегрировать процесс концептуализации объекта в пространстве в другие когнитивные процессы: распознавания сущности и категоризации, механизма внимания, осознания причинно-следственных связей. ИКА только учатся такой интегративности и взаимной координации, например, когда генеративные модели концептуализируют те объекты, в которых они не уверены, поскольку это продукты галлюцинирования,

* Статья подготовлена по материалам проекта «Текст как Big Data: методы и модели работы с большими текстовыми данными», выполняемого в рамках Программы фундаментальных исследований НИУ ВШЭ в 2024 г.



как объекты с нечеткими границами, а модели Image to Text объединяют в единый неоднородный объект человека и наиболее яркую оригинальную деталь его окружения, поскольку «считают», что это самое важное для задач описания изображения.

Ключевые слова: концептуализация пространства, топологическая семантика, искусственный когнитивный агент, естественный когнитивный агент, фигура и фон, большие языковые модели



CONCEPTUALIZING THE SPACE: HOW NATURAL AND ARTIFICIAL COGNITIVE AGENTS USE TOPOLOGICAL SEMANTICS SCHEMES (BASED ON DESCRIPTIONS OF PAINTINGS FROM THE HERMITAGE COLLECTION)

Anastasia V. Kolmogorova –
DSc in Philology, Professor,
Head of the Laboratory
of Language Convergence.
HSE Campus in Saint
Petersburg.

119–121 nab. Griboyedov
Canal, St. Petersburg 90068,
Russian Federation;
e-mail: akolmogorova@hse.ru

Polina A. Nalobina –
Researcher at the Laboratory
of Language Convergence.
HSE Campus in Saint
Petersburg.

119–121 nab. Griboyedov
Canal, St. Petersburg 90068,
Russian Federation;
e-mail: pnalobina@edu.hse.ru

The article is devoted to the description of the differences in the conceptualization of space observed in informants, large language models and computer vision models capable of generating a text describing what they “saw”. We use the concept of a cognitive agent and substantiate the distinction between “natural vs artificial cognitive agent”: the first is understood as a person, the second is an AI model capable of making decisions and performing tasks adequately in a given situation. The aim of the study is to compare the ways of understanding the location of an object in space in natural cognitive agents and artificial cognitive agents of two types: large language models and models created for Image to Text task. The main methods are the method of linguistic experiment and the method of semantic description based on the theory of topological semantics by L. Talmi. As an incentive material, six paintings from the collection of the State Hermitage Museum were used, divided into three groups: portraits, monofigure paintings on mythological or religious themes, and multi-figure compositions. The participants of the experiments were: 63 informants (Mean age = 19.1, 48 females, 15 males), 5 LLMs, 6 Image to Text models based on computer vision technology and capable of generating descriptions of recognized images



in English. Using the typology of configurational topological schemes and “figure – background” type schemes, we compared the ways of understanding space that the models rely on. As a result, we have formulated a number of conclusions, the most important of which is that natural cognitive agents differ from artificial cognitive agents in its ability to integrate the process of conceptualization of an object in space into other cognitive processes: entity recognition and categorization, attention mechanisms, awareness of cause-and-effect relationships. Artificial cognitive agents are only learning such integrativity and mutual coordination, for example, when generative models conceptualize those objects in which they are not sure, since these are products of hallucination, as objects with fuzzy boundaries, and Image to Text models combine into a single heterogeneous human object and the most striking original detail of its environment, because they “believe” that this is the most important thing for description tasks.

Keywords: conceptualization of space, topological semantics, artificial cognitive agent, natural cognitive agent, figure and background, large language models

1. Введение

Данное исследование появилось в рамках большого прикладного проекта, связанного с дигитализацией музейных практик Государственного Эрмитажа. Работа с несколькими видами компьютерных языковых моделей, с одной стороны, и экспериментальная работа с информантами – с другой, натолкнули на мысль о необходимости сравнения тех топологических семантических моделей, которые применяют ИИ и человек, осмысляя и вербализуя сцены, локализованные в пространстве картин.

У музейной институции есть замечательный цифровой «двойник» – Цифровая коллекция Эрмитажа, включающая в себя цифровые изображения более чем миллиона экспонатов, находящихся как в выставочных пространствах, так и в запасниках музея. Такие экспонаты имеют информационное сопровождение: указаны годы создания и приобретения, автор, номер зала, если работа выставлена, а также некоторая другая информация. Ресурс мог бы быть полезен для тех, кто хочет заранее выстроить маршрут, включив в него картины, которые давно хочется увидеть воочию. Однако для этого нужно точно знать, как называется картина. Далеко не все пользователи, потенциальные посетители залов Эрмитажа, владеют такой информацией – скорее, они хранят в памяти стереотипный образ того, что на картине изображено, например, «Женщина с каким-то зверем на плечах» («Дама с горностаем»). Однако существующая поисковая система Цифровой коллекции не поддерживает такие «неточные» запросы.



Задача большого проекта, которым мы занимаемся, сводится как раз к созданию системы автоматического поиска картин по слабо формализованным запросам пользователей.

В ходе реализации данного проекта нам потребовалось достаточно много тестировать существующие Большие языковые модели (далее – БЯМ), работающие с русским языком, а также модели, распознающие объекты на изображении и генерирующие их краткое описание (модели Image to Text). Кроме того, мы провели эксперимент с информантами, которые выполняли задание по формированию запросов по поиску картин.

Эта работа прикладного характера и обсуждение ее результатов с разработчиками и тестировщиками компьютерных моделей привели к осознанию того, что естественные когнитивные агенты (люди) и искусственные когнитивные агенты (модели) выполняют операции по концептуализации объектов в пространстве по-разному. Для того чтобы увидеть отличия, потребовалась некая единая концептуальная рамка – ею послужила концепция топологической семантики, предложенная Леонардом Талми и хорошо известная отечественным когнитивным лингвистам [Рахилина, 2002; Скребцова, 2018].

Данная проблематика также тесно связана с философским пониманием субъекта сознания (и познания). В условиях появления искусственного интеллекта в научном сообществе остро ощущается потребность в уточнении границ и сущностных характеристик данного фундаментального понятия [Касавин, 2024].

Одной из стартовых точек для дискуссии в кругах специалистов по компьютерным и когнитивным наукам стало определение сознания, данное Томасом Нагелем: «Организм имеет сознательные ментальные состояния тогда и только тогда, когда есть что-то, каково это *быть* этим организмом – что-то, каково это для организма» [Nagel, 1974, p. 436]. Продолжая мысль Т. Нагеля, отечественный физиолог К.В. Анохин отмечает, что «сознание имеет особые качественные свойства и специфические операциональные особенности. И те и другие – атрибуты когнитивного агента и не могут быть поняты без понимания устройства последнего...» [Анохин, 2021, с. 48].

Специалисты в области когнитивных и нейронаук активно обращаются к вопросу о том, обладают ли нейросетевые модели субъективным опытом – одним из онтологических признаков сознания. В фундаментальной работе [Ji et al., 2023] идеологов глубинного обучения проводится параллель между функциями аттракторов в организации рабочей памяти у человека и механизмами внимания, используемыми в моделях-трансформерах в глубинном обучении. При этом утверждается, что оба механизма являются практически эквивалентными вариантами «сжатия информации» при попытке передачи познающим субъектом многообразия своего субъективного опыта. Другие специалисты в области обучения нейросетевых моделей



[Lillicrap et al., 2020] считают, что приобретение и структурирование субъективного опыта в кортексе человека работает точно так же, как механизм обратного распространения ошибки, широко используемый при обучении нейросетей. Таким образом, отмечают исследователи, последний можно использовать для объяснения нейрокогнитивных процессов у человека.

Иными словами, новое знание, даже достаточно локальное, об отдельных и общих свойствах естественных и искусственных когнитивных агентов может быть рассмотрено как шаг на пути к философскому осмыслению природы сознания – так называемому «главному вопросу» в том числе и эпистемологии науки.

В дальнейшем текст будет построен следующим образом: 1) сначала мы обратимся к основным понятиям и конструктам топологической семантики в работах Л. Талми и его последователей; 2) затем представим цель, материал и методы, примененные нами в исследовании; 3) потом опишем и прокомментируем полученные результаты; 4) в заключение покажем перспективы исследования и его ограничения.

2. Основные понятия и методологические принципы семантической топологии

Лингвистическая концепция Л. Талми, основные постулаты которой изложены в [Talmy, 2000a; 2000b; 2018], построена на допущении о том, что в каждом языке можно выделить две группы систем: открытые классы слов (open-class words) и закрытые классы слов (closed-class words) [Talmy, 2000a, p. 8–14]. Первые – это лексическая подсистема (лексико-семантические группы слов, лексико-семантические ряды, поля и т.д.), вторые – преимущественно грамматическая подсистема: союзы, частицы, артикли, предлоги, грамматические категории, грамматические значения, аффиксы с грамматическим значением. Если первые отражают содержание концептуальной системы человека, то вторые – концептуальные структуры, в которые это содержание «упаковывается».

На основе того, что структуры мышления и языка, используемые для описания пространства, места и местоположения объекта, легко проецируются на другие сферы, формирующие человеческое восприятие (например, деятельность или отдельный ее акт, имеющие протяженность во времени, могут представляться как объект, т.е. занимать некоторое место в пространстве: *Я позволю нескольким людям может легко превратиться в Я сделаю несколько звонков*), Л. Талми высказал гипотезу о том, что в каждом языке количество таких структур конечно, они глубоко фундированы в когнитивном



(преимущественно сенсомоторном) опыте человека, а потому – легко выполняют функцию структурирования всей той информации, которая поступает в когнитивную систему человека благодаря восприятию [Talmy, 2000a, p. 21–23].

Подобные схемы исследователь назвал топологическими, поскольку топология – специальный раздел геометрии – изучает такие характеристики формы объектов, которые сохраняются, оставаясь узнаваемыми, даже после его сильной деформации [Прасолов, 2012]. Для описания объектов с точки зрения топологии не важны расстояния между их частями или размеры – важно, из скольких частей они состоят, много их или объект один, с чем он соединен, что находится рядом [Talmy, 2000a, p. 47–48]. Именно языковая топология позволяет нам одинаково успешно сказать, что «Пятно на платье слева больше, чем то пятно на платье, которое справа» и «Территория Китая меньше, чем территория России», хотя размеры пятна и территорий стран несравнимо различны. Л. Талми подчеркивает, что закрытые грамматические классы слов отражают не Евклидово пространство, а относительные, субъективно воспринимаемые Наблюдателем¹ паттерны пространства (расстояния, формы, размеры) – их топологические характеристики и схемы [Ibid., 48–49].

Для описания топологических характеристик объектов Л. Талми предложил целый ряд конструкторов и категорий.

Во-первых, это понятие схематизации. Под схематизацией понимается процесс, при котором Наблюдатель системно отбирает такие аспекты референтной ситуации, которые в его описании будут репрезентировать всю наблюдаемую сцену целиком, а другие аспекты и характеристики как бы выпадают из поля зрения [Ibid., p. 225]. Сцена – это воспринимаемая Наблюдателем ситуация, прошедшая через «горнило» схематизации, когда посредством элементов закрытых классов (грамматической системы) и структуры предложения все наблюдаемое делится на сегменты (portions), между которыми устанавливаются связи [Ibid., p. 229].

Каждая референциальная сцена структурируется Наблюдателем (говорящим) и реструктурируется его партнером по коммуникации, целью которого становится воссоздать систему элементов, составляющих сцену, и их взаимосвязи максимально близко к тому, что описывал говорящий.

Для того чтобы осмыслить, а затем описать сцену, Наблюдатель опирается на некоторые схематические структуры, находящиеся в системных отношениях между собой и объективируемые при помощи определенных групп языковых средств из номенклатуры слов

¹ Мы используем термин «Наблюдатель» вслед за У Матураной и Ф. Варелой [Maturana, Varela, 1997] для обозначения говорящего и познающего субъекта в одном лице – «Все сказанное сказано Наблюдателем» [Ibid.].



закрытых классов (дейктики, аспектуальные и временные формы глаголов, маркеры числа у именных и глагольных классов и т.д.).

Л. Талми выделяет пять таких схематических систем, но в данной публикации мы остановимся только на так называемых конфигурационных структурах и схеме «фигура – фон».

Существует три основных типа конфигурационных структур, призванных реализовывать соответствующие категории, используемые для того, чтобы осмысливать местоположение объекта: комплексность (plexity), ограниченность (boundedness) и отдельность (dividedness).

Под комплексностью исследователь понимает категорию, строящуюся на оппозиции «единственность – множественность», аналогично категории числа у имен и глаголов. Однако Л. Талми придает данной категории статус концептуального континуума, что на практике означает ее расширение за счет включения не только грамматических, но и лексических средств, с разной степенью интенсивности способных передавать идею количества применительно к осям и пространства, и времени [Talmy, 2000a, p. 57]. Так, на последней из упомянутых осей понятие действия, т.е. активности Х, длящейся во времени, получает возможность реализовать идею количества посредством использования временных и видовых форм глаголов. Например, так называемые «точечные» глаголы актуализируют идею единичности, или, в терминах Л. Талми, униплексности (*он поперхнулся* – это единичное действие), а итеративные – множественности, или мультиплексности (*он попивал кофе, сидя у окна* – он делал не один, а много глотков, один за одним, с некоторым интервалом).

На оси пространства возникает понятие места – количества пространства, занимаемого некоторой сущностью, становящейся при таком ракурсе объектом. Идея единичности такого объекта задается у существительных формами неопределенного артикля (в языках аналитического строя), числительными в препозиции (*один дом*) или может быть реализована и на фоне множественности за счет, например, конструкции *один из*:

- (1) *В комнату зашли трое, один из них был одет в гимнастерку, остальные – в гражданском.*

Если *трое* и *остальные* – лексикализованные способы выражения идеи множественности, то конструкция *один из* – способ выделить единичность из этой множественности. Таким образом, комплексность возникает как некая универсальная категория для выражения идеи количества на пересечении категорий времени и пространства, становящейся возможной в языке благодаря тому, что последний может достаточно легко концептуализировать и то, и другое как сущность.



Ограниченность – категория, базирующаяся на различии, с одной стороны, понятия некоторого количества материального субстрата, границы которого не определены, а лексема-вербализатор понятийного содержания не имеет эксплицитного признака конечности, финитности [Talmy, 2000a, p. 50], и, с другой стороны, понятия некоторого количества материального субстрата, границы которого либо четко заданы контекстом, либо имплицированы семантикой языковой единицы-вербализатора. Например, для глаголов характерными формами, маркирующими ограниченность действия, понимаемого как сущность, являются все варианты перфектных форм, терминативные глаголы (ex.g. *упасть*, *входить*), а для существительных, в семантике которых имплицирована идея границ, примерами могут быть *море* (в отличие от *вода* – неисчисляемого существительного), *пляж* (в отличие от *песок*), *отпуск* (в отличие от *отдых*).

Отдельность – следующая категория, которая строится на противопоставлении признаков континуальности некоторой сущности и ее внешней или внутренней дискретности (сегментации) [Ibid., p. 56]. Так, формы *листья* и *вышагивать* имплицировывают идею отдельности каждого элемента множества (листочка или шага), а *листва* и *шагать* – идею имплицитной дискретности, существующей внутри множества (листва состоит из листьев, действие *шагать* – из шагов). Напротив, ни одна из этих идей не присутствует в семантике форм *пыль* (от листьев, например) или *идти*, что позволяет классифицировать данные формы как не обладающие признаком отдельности.

Наконец, самое, пожалуй, известное положение теории семантической топологии состоит в том, что всякую сцену Наблюдатель воспринимает как комбинацию фигуры (чего-то подвижного, с четкими контурами, небольшого по размеру) и фона (неподвижного, с нечеткими границами, невыделенного). Классический пример – *автомобиль стоит у дома*, где *автомобиль* – фигура, а *дом* – фон.

Мы будем опираться на вышеописанные положения в дальнейшем тексте.

3. Цель исследования

Цель работы – сравнить типы топологических семантических схем, используемых для описания «сцены», изображенной на картине, естественными когнитивными агентами (людьми; далее – ЕКА) и искусственными когнитивными агентами – Большими языковыми моделями и генеративными моделями, обученными для задачи Image to Text (далее – ИКА).

Уточним содержание трех терминов, которыми будем пользоваться в дальнейшем.



Хотя понятие интеллектуального или когнитивного агента восходит к работам отечественного теоретика кибернетики Д.А. Поспелова [Поспелов, 1986], интерес к его содержанию возродился относительно недавно в связи с созданием систем искусственного интеллекта.

Когнитивный агент определяется как интеллектуальная система, способная выбирать оптимальную линию поведения в любой ситуации [Russell & Norvig, 2000; Schatten et al., 2021], т.е. умеющая воспринимать среду, воздействовать на нее, обучаться, хранить опыт взаимодействия с ней и на его основе предсказывать реакцию среды на свои действия и планировать их последовательность [Витяев и др., 2020, с. 8].

Естественный когнитивный агент – это интеллектуальная система, которая способна все это делать на основе биологических, химических и других ресурсов, предоставляемых ее телесностью.

Искусственный когнитивный агент извлекает все необходимые ресурсы из собственной базы данных и знаний, получаемых от человека в процессе диалога «человек – искусственный агент», а также от сенсоров системы, основанной на грануляции и градуировании входной информации [Калуцкая, 2010, с. 6].

Проанализировав на основе методологического аппарата топологической семантики результаты решения задачи концептуализации того, что изображено на картинах, мы можем получить представление о различиях, существующих в типичных паттернах осмысления пространства и объектов в нем естественными и искусственными агентами.

4. Материал и методы

Материал исследования составили три корпуса данных.

Первый корпус составили примеры поисковых запросов, предложенных в ходе эксперимента 63 информантами (Mean age = 19,1, 18 девушек, 15 юношей), студентами вузов Санкт-Петербурга для шести картин из коллекции Государственного Эрмитажа. Для всех информантов русский язык является родным. Эксперимент проводился онлайн посредством платформы Google Forms, участники не были ограничены по времени ответа. Получено 378 реакций респондентов.

Информантам была сформулирована следующая инструкция: «Посмотрите на картину (визуальный стимул был представлен в виде изображения в формате jpg внутри опросника), сфокусируйтесь на том, что на ней изображено. Представьте себе, что вы хотите ее найти в Интернете, но не помните ни автора, ни названия. Как бы вы



описали то, что на ней изображено, формулируя поисковой запрос?» Время ответа не контролировалось. Визуальные стимулы были представлены в опроснике, поэтому информант при желании мог вновь рассмотреть их.

Информантам были предложены шесть визуальных стимулов:

1. Т. Гейнсборо «Портрет дамы в голубом», Государственный Эрмитаж, Санкт-Петербург (инв. ГЭ-3509)

<https://www.hermitagemuseum.org/wps/portal/hermitage/digital-collection/01.+paintings/38759>

2. Караваджо «Юноша с лютней», Государственный Эрмитаж, Санкт-Петербург (инв. ГЭ-45)

<https://www.hermitagemuseum.org/wps/portal/hermitage/digital-collection/01.+paintings/31511>

3. Джорджоне «Юдифь», Государственный Эрмитаж, Санкт-Петербург (инв. ГЭ-95)

<https://www.hermitagemuseum.org/wps/portal/hermitage/digital-collection/01.+paintings/32196>

4. Тициан «Кающаяся Мария Магдалина», Государственный Эрмитаж, Санкт-Петербург (инв. ГЭ-117)

<https://www.hermitagemuseum.org/wps/portal/hermitage/digital-collection/01.+paintings/32191>

5. А. Ватто «Затруднительное предложение», Эрмитаж, Санкт-Петербург (инв. ГЭ-1150)

<https://www.hermitagemuseum.org/wps/portal/hermitage/digital-collection/01.+paintings/37466>

6. Д. Веласкес «Завтрак», Государственный Эрмитаж, Санкт-Петербург (инв. ГЭ-389)

<https://www.hermitagemuseum.org/wps/portal/hermitage/digital-collection/01.+paintings/32729>

Все картины-стимулы можно условно объединить в три группы по критерию характера сцены, изображенной на них и становящейся объектом концептуализации: 1) представлен один человек, сцена не осложнена дополнительными символическими и культурно-специфическими смыслами (картины № 1, 2); 2) представлен один человек, но сцена осложнена дополнительными символическими и культурно-специфическими смыслами (картины № 3, 4); 3) в сцене участвуют несколько человек (картины № 5, 6). В дальнейшем специфика концептуализации сцен будет анализироваться в соответствии с данными группами.

В дальнейшем анализ данного подкорпуса будет сопровождаться пометой ЕКА (естественные когнитивные агенты).

Второй и третий корпуса данных составили ответы моделей ИИ двух типов. Во-первых, анализировались ответы 5 генеративных языковых моделей: русскоязычная Яндекс GPT-3 <https://ya.ru/ai/gpt-3> и мультиязычные Mixtral <https://console.mistral.ai/> (2 версии: 8x7B



и 8x22B), Llama-3.1(70B) <https://huggingface.co/meta-llama/Llama-3.1-70B> и ChatGPT-4.o <https://chatgpt.org/>.

Промпты для них подавались на русском языке, и ответы генерировались тоже на русском языке.

Во-вторых, анализировались ответы шести англоязычных моделей, обученных для задач Image to Text, ответы которых мы будем анализировать в оригинальной англоязычной версии и в переводе на русский язык:

1. <https://huggingface.co/spaces/hysts/image-captioning-with-blip>
2. <https://huggingface.co/spaces/hysts/BLIP2>
3. <https://huggingface.co/Salesforce/xgen-mm-phi3-mini-instruct-interleave-r-v1>.
4. <https://huggingface.co/spaces/adept/fuyu-8b-demo>
5. <https://huggingface.co/spaces/pharmapsychotic/CLIP-Interrogator>
6. <https://huggingface.co/nlpconnect/vit-gpt2-image-captioning>

Такой выбор моделей обусловлен отсутствием моделей Image-to-Text с хорошими бенчмарками, работающими с русским языком, и лидерством по показателям качества распознавания и ответов на визуальные вопросы среди англоязычных моделей семейства моделей BLIP [Xue et al., 2024].

Для больших языковых моделей использовался промпт на русском языке «Представь, что ты обычный молодой человек 19–21 года, студент или студентка, хочешь найти в Интернете картину [название картины и ее автор], но ты не помнишь ни название, ни автора. Единственное, что ты помнишь, – это то, что на ней изображено. Как бы ты описала картину в поисковом запросе?». Отметим, что в формировании промпта мы использовали технику создания «атмосферы», когда модели предлагается примерить на себя некоторую социальную или ситуативную роль.

Англоязычным моделям на основе механизмов компьютерного зрения Image to Text на вход подавалась картина (изображение в формате jpg), а на выходе мы получали ее описание на английском языке. В дальнейшем текст данные описания приводятся в оригинале и сопровождаются переводом, полученным от Яндекс-переводчика и скорректированным двумя профессиональными переводчиками-экспертами.

Таким образом, объем второго корпуса составил 30 ответов генеративных языковых моделей и 36 от моделей Image to text.

Полученный таким образом материал анализировался сначала с использованием метода когнитивного моделирования (для каждой картины выявлялись типы и содержание топологических схем, лежащих в основе концептуализации объектов на картине), а затем при помощи тройственного сопоставительного анализа выявлялась специфика концептуализации «сцены» на картине естественными когнитивными агентами и искусственными двух типов: генеративными



языковыми моделями, компилирующими текстовые данные, опираясь на формат запроса, заданный в промпте (далее – ИКА_gen), и моделями, генерирующими текстовое описание на основе преобразования результатов работы алгоритмов компьютерного зрения (далее – ИКА_image).

5. Результаты и обсуждение

5.1. Группа картин № 1 (Т. Гейнсборо «Портрет дамы в голубом» и Караваджо «Юноша с лютней»)

Рассмотрим топологические схемы в сравнении для ЕКА и ИКА.

5.1.1. Конфигурационные схемы

Комплексность. Ограниченность. Отдельность.

ЕКА. В данной группе комплексность объективируется в основном за счет значений единственного и множественного числа существительных. Преобладает единичность – *девушка, лютня, скрипка, стол*:

(2) *Девушка играет лютне, скрипка лежит рядом на столе.*

Однако часты переходы между единичностью и множественностью для концептуализации одного и того же объекта в силу действия принципов селективности (из многого – один) и группировки (многое – в одно) (визуально данный принцип представлена на рис. 1). На картине Караваджо, например, изображены два музыкальных инструмента – скрипка и лютня. Но одни информанты фокусируются только на лютне (*музыкальный инструмент* – единичность, пр. 3), а другие «видят» сразу всю группу и используют категорию множественности (*музыкальные инструменты*, пр. 4):

(3) *Девушка за столом, музыкальный инструмент;*

(4) *Грустная девушка с музыкальными инструментами.*

Информанты могут осмыслять нотную тетрадь через категорию множественности – *ноты*, а могут использовать единичность, чтобы концептуализировать объект как одиночный, но неоднородный объект – *нотный стан*; аналогично: *цветы* – *букет цветов*.

Изображено три объекта: при действии принципа селективности внимание сосредотачивается на одном объекте, и он концептуализируется как единичный, например *скрипка*; при действии принципа группировки все три объекта объединяются в группу и концептуализируются как множественный объект, например *музыкальные инструменты*.

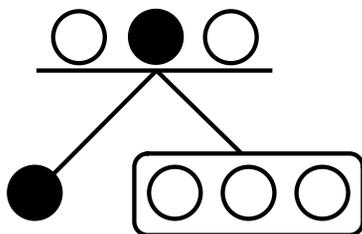


Рис. 1. Принципы селективности и группировки в категории сложности у ЕКА

Что касается категории ограниченности, то большинство сущностей на картинах осмысляются как объекты с четкими границами, согласно денотативным образам номинирующих их существительных: *платье, накидка, шляпка, лента, брошь, скрипка, ваза* и т.д. Но границы объекта могут осмысляться как нечеткие в тех случаях, 1) если объект воспринимается как фоновый (*дама в голубом, в чем-то пышном, дама с драпировкой*); 2) если объект, которым может стать вся сцена или временной интервал, не распознается точно Наблюдателем (*девушка за столом играет на каком-то инструменте*) или плохо припоминается: *что-то связанное с игрой на музыкальных инструментах в XVI–XVIII вв.*

Вербальными маркерами топологической схемы «объект с нечеткими границами» выступают неопределенные местоимения *что-то, какой-то*, использование так называемых аппроксимативов – предложных конструкций перед именными формами в среднем роде (*в белом, в голубом* и т.д.).

Описывая картины, информанты склонны объединять человека и соположенные с ним объекты посредством топологической схемы единого отдельного, но неоднородного внутри объекта (рис. 2) по схеме «X с Y, где X – человек, а Y – то, что находится рядом с X». Так, за счет коммутативной конструкции с предлогом *с* прическа предстает как продолжение, составная часть дамы: *портрет женщины с пышной прической; женщина с помпезной прической* и т.д. Конфигурация этого неоднородного объекта может быть совершенно разной: человек и то, что он держит (*портрет юноши с лютней*); человек и то, что он держит + его руки (*девушка с музыкальным инструментом в руках*), + просто рядом находящиеся объекты (*девушка в белом со скрипкой и букетом в вазе*):

Эти же объекты, например прическа, могут легко концептуализироваться и как отдельные единичные (схема такой концептуализации представлена на рис. 3):

- (5) *Дама, высокая прическа и шляпка* – три единичных объекта.



Примечательно, что глаза также могут осмысляться как отдельный объект (пр. 6) в ряду объектов – происходит извлечение и отчуждение части из целого:

- (6) *Музыкальный инструмент, скрипка, ноты, цветы, глаза.*

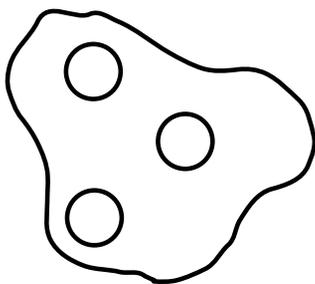


Рис. 2. Вариант концептуализации сцены
«единый неоднородный объект»

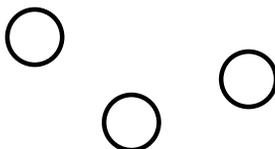


Рис. 3. Вариант концептуализации сцены
«множество отдельных объектов»

ИКА_gen. Множественность отражена в формах множественного числа существительных: *ноты, деревья, волосы, манжеты*. Переход множественности в единичность встречается один раз: *волосы – в прическу*. Не только все сущности концептуализируются как объекты с четкими границами, но и используются лексические маркеры дополнительного подчеркивания границ объектов (пр. 7):

- (7) *женщина – в полный рост* (ограничена фигурой); *волосы – уложены в прическу* (пассив: им придали форму); *стол покрыт скатертью* (пассив: границы стола подчеркнуты скатертью); *свет проникает через окно* (активная семантика глагола, наличие в нем семантической пресуппозиции ‘наличие преграды’, усиливаемое предлогом *через*: окно ограничивает свет).



В качестве объектов, не имеющих границ, осмысляются только не существующие на картине объекты, являющиеся результатом галлюцинирования моделей. Они номинируются абстрактными существительными, например: *свет, музыка, небо, пейзаж*.

подавляющее большинство объектов осмысляются как отдельные, в том числе фиксируем частые случаи отчуждения части тела или характеристики внешности человека от своего «носителя»: *женщина имеет осанку* – осанка рассматривается как отдельный объект, которым субъект обладает; аналогично – *голова наклонена, волосы уложены*.

Отмечены случаи концептуализации группы объектов как единого неоднородного объекта (см. рис. 2) при помощи схемы, отражающей объективные системные отношения целого и части «X с Y, где X – человек, а Y – часть X»: *в голубом платье с белым кружевным воротничком и манжетами; на фоне пейзажа с деревьями и небом*.

ИКА_image. Доминирует категория единичности (пр. 8):

- (8) *a painting of a woman playing a lute* ('картина с женщиной, играющей на лютне').

Неопределенный артикль (*a woman, a lute*) подчеркивает единственность объекта.

Все выделяемые на картине объекты описываются как имеющие четкие границы: *portrait of a woman in a blue dress with a feathered hat* ('портрет женщины в голубом платье со шляпой с перьями').

У моделей Image to text, так же как у ЕКА, наблюдаем объединение в один объект нескольких объектов, соположенных с изображенным человеком, однако это происходит по несколько иной модели – «X с Y, где X – человек, а Y – яркий элемент его одежды»: *a woman with a bow tie* (женщина в платье с галстуком-бабочкой); *portrait of a lady with a blue hat* (портрет дамы с голубой шляпкой).

5.1.2. Фигура и фон

ЕКА. В качестве фона выступают размытые цветовые пространства (*на темном фоне*), помещения (*мальчик с балалайкой в комнате*²), одежда (*дама в платье, в кимоно*), соматические объекты

² Отметим на полях, что только у ЕКА среди ответов встречались случаи концептуализации лютниста на картине Караваджо как юноши; все же ИКА единодушно определяли его как девушку. Из этой специфической черты ЕКА, как представляется, происходит одно важное наблюдение: только ЕКА обладают способностью наблюдать «наблюдаемое не сейчас». Юноша у Караваджо, действительно, соответствует современным стандартам женской, не мужской, красоты, но ЕКА могут себе представить, что когда-то стандарты красоты были иными, и смотреть на изображаемое «глазами» Наблюдателя того времени, благодаря некоторым фоновым знаниям (благодарим за это наблюдение одного из рецензентов).



(ленточка *на шее*), а также временной отрезок, концептуализируемый как большой объект с достаточно четкими контурами (портрет дамы *в XVIII в.*).

ИКА_gen концептуализируют как фон те же объекты, что и ЕКА, за одним исключением – в функции фона никогда не используется временной период.

Для **ИКА_image** топология фигуры и фона не характерна.

5.2. Группа картин № 2 («Юдифь» Джорджоне и «Кающаяся Мария Магдалина» Тициана).

5.2.1. Конфигурационные схемы

Комплексность. Ограниченность. Отдельность.

ЕКА. Преобладает единичность: *грустная девушка с книгой смотрит в небо*. Переход от единичности к множественности при осмыслении одного и того же объекта только один: руки Марии Магдалины могут концептуализироваться либо как единичный объект (*держит руку на сердце*), либо как множественный (*полускрестив руки на груди*).

подавляющее большинство объектов концептуализируются информантами как имеющие четкие границы. Исключение составляют а) объекты, непосредственно не присутствующие в сцене, но имплицуемые – дама с мечом одолела *кого-то* (импликатура: об этом свидетельствует отрубленная голова); б) объекты, не попадающие в фокус внимания, например одежда Юдифи (*девушка в розовом с мечом и ногой на отрубленной голове*); слабо категоризируемые объекты, например *что-то странное слева* (*девушка страдает, плач, темная улица, рядом бутылка пустая, книга, рука на груди, боль, что-то странное слева*). Наконец, вся сцена может представляться как некий глобальный образ, не имеющий границ, на фоне которых выделяется единственный четкий объект – дама: *незаметный, пугающий, дама, завоевание*.

Для концептуализации головы Олоферна на картине Джорджоне используются три топологические схемы: 1) топологическая схема отчужденной части другого объекта (*отрубленная голова; оторванная голова* – имплицуется идея «отъединения» данной части от какого-то целого, которому она принадлежала); 2) неотчужденной части объекта (*наступает на голову мужчины* – при таком описании она может наступать на голову спящему совершенно «целому» мужчине); 3) описывается как отдельный объект (*валяется голова*).

Как и в предыдущих случаях, описывая картины, информанты склонны объединять человека и соположенные с ним объекты посредством топологической схемы отдельного единого, но неоднородного внутри объекта за счет атрибутивной конструкции с предлогом



с: женщина с мечом, отрубившая голову мужчине; женщина смотрит в небо с рукой у груди; очень печальная девушка с книгой на голове. Те же объекты – меч, книга, рука – могут концептуализироваться и как отдельные.

Интересны способы концептуализации ноги Юдифи (рис. 4). Нога как объект может не выделяться вовсе, но имплицироваться в объекте «тело Юдифи» (рис. 4а). Это достигается употреблением глаголов *наступить, попать*, включающих в свою семантику сему 'нога': *девушка с мечом наступает на отрубленную голову; охотница с мечом, попирающая человеческую голову.*

Нога может также осмысляться как составляющая неоднородного объекта – тела Юдифи, включающего и ногу, и меч (рис. 4б): *девушка в розовом с мечом и ногой на отрубленной голове.*

Она может концептуализироваться и как часть членимого объекта «ноги» – *одной ногой* (рис. 4в, пр. 9):

- (9) *женщина с мечом, одной ногой стоящая на человеческой голове. Рядом дуб.*

Нога может становится отдельным объектом, связь которого с Юдифью устанавливается лишь индуктивно (рис. 4г, пр. 10):

- (10) *девушка в красном халате с мечом и под ногой голова.*

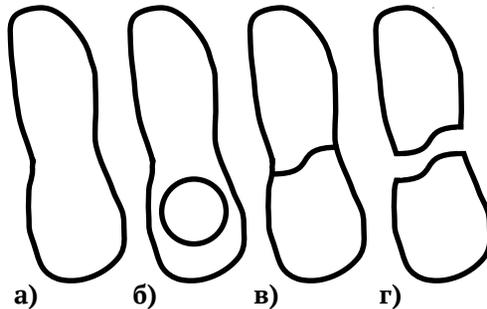


Рис. 4. Варианты концептуализации ЕКА объекта «нога» на картине Джорджоне «Юдифь»: слева направо – а) как объект не выделяется, поскольку неотделим от целого; б) составляющая неоднородного объекта «Юдифь»; в) часть объекта «Юдифь»; г) отдельный объект

При описании Марии Магдалины информанты склонны концептуализировать лицо, руки, грудь как составляющие неоднородного объекта «Мария Магдалина», о чем свидетельствует регулярное использование предлога *с*: *молящаяся некрасивая женщина со страдающим лицом; женщина смотрит в небо с рукой у груди.* Как отдельные



объекты концептуализируются ее эмоционально-волевые состояния: например, мольба (*девушка, молящая о прощении, мольба*), боль (*рука на груди, боль, что-то странное слева*). Иными словами, информанты склонны «расчленять» живой объект «Мария Магдалина», выделяя такие составляющие, как части тела, эмоции, переживания, качества личности.

ИКА_gen. Для описания картины Тициана генеративные языковые модели также используют категорию единичности, концептуализируя женщину, пещеру, череп, тень, распятие. Только два объекта концептуализируются как множественные объекты: у Яндекс GPT-3 – руки (*в руках у нее...*) и у всех моделей – волосы (*женщина с длинными волосами*). Интересно, что информанты (ЕКА) «не видели» волосы в качестве объекта ни на одной из картин. Переходов от множественности к единичности и наоборот – нет.

Модель LLama осмысляет Марию Магдалину как объект с четкими границами, выделяя только признак объекта (*полуголая*), семантика которого указывает на его принадлежность категории «человек», а грамматическая форма – на гендер: *полуголая, тень, волосы золотистые*.

Большинство моделей при описании головы Олоферна с картины Джорджоне используют преимущественно две схемы: либо голова выступает как часть неоднородного объекта «Юдифь», вследствие чего возникает когнитивный диссонанс – чья голова? (*женщина с мечом и головой*), либо как отчужденная часть другого объекта: *под ногами у нее лежит... блин, ну, голова отрубленная!* (Chat-GPT 4.0).

Генеративные языковые модели так же, как информанты, активно используют коммутативную конструкцию с предлогом *с* для концептуализации женских фигур на картинах как неоднородных объектов, включающих в себя другие объекты, например волосы или меч.

ИКА_image. При описании картины Джорджоне модели не выделяют отрубленную голову в качестве объекта, а при описании картины Тициана – «не видят» волосы. Множественность или единичность при концептуализации рук зависит от картины: концептуализируя Юдифь, модели выделяют единственную руку – ту, что держит меч (*a woman with a sword in her hand*), концептуализируя Марию Магдалину – две руки, потому что они держат книгу (*a woman with her hands over her chest, holding a book in her hands...*). Переходов между единичностью и множественностью при осмыслении одного и того же объекта нет.

Модели концептуализируют все сущности как объекты, имеющие четкие границы: *a painting of a woman holding a sword; a painting of a woman holding a book* (изображение женщины, держащей меч; изображение женщины, держащей книгу). При этом для того, чтобы



осмыслить степень отдельности объектов, они используют типичную и для обработки других изображений схему концептуализации неоднородного объекта: «X с Y, где Y – то, что X держит в руках», поэтому голова Олоферна и нога Юдифи на картине Джорджоне не попадают в поле концептуализации моделей.

5.2.2. Фигура и фон

ЕКА. Описывая данную группу картин, информанты используют не только статический фон, например элемент рельефа (гора) или временной отрезок (эпоха Возрождения) (*картина с девушкой, которая держит руку на сердце в горе; девушка с отрубленной головой эпоха Возрождения*), но и динамический: это явления природы, на фоне которых разыгрывается сцена (сумерки: *очень печальная девушка с книгой на черепе. Смеркалось*; шторм: *женщина, раздираемая внешними муками в шторм*).

ИКА_gen. Генеративные языковые модели используют только статичный фон: лес (*картина женщина с головой мужчины на блюде в лесу*), пещера (*женщина с длинными волосами в пещере*), пейзаж (*вокруг нее – пустынный пейзаж*) и др.

ИКА_image не используют схему «фигура – фон».

5.3. Группа картин № 3 (А. Ватто «Затруднительное предложение» и Д. Веласкеса «Завтрак»)

5.3.1. Конфигурационные схемы

Комплексность, ограниченность, отдельность.

ЕКА. В ответах информантов, концептуализирующих людей на картинах, категория комплексности в том смысле, как ее понимает Л. Талми, реализована во всех возможных ее вариантах и образует градуированную шкалу:

а) множественность, маркированная употреблением числительных в препозиции к существительному: *Картина, на которой три барышни и два кавалера сидят в лесу*;

б) комплексность множества единичных объектов (сыновья): *сыновья с отцом сидят за столом с едой*;

в) множественность неоднородного объекта с четкими границами (*пятеро*, которое, в отличие от числительного *пять*, обозначает концептуализацию этих пяти единиц как единого объекта): *пятеро людей на берегу*;

г) множественность неоднородного объекта с нечеткими границами (дворяне – это номинация очень широкой сословной категории): *дворяне на природе*;

д) комплексность неопределенного множества, извлеченного из другого неоднородного объекта с нечеткими границами (несколько



людей – люди представляют собой номинацию чрезвычайно широкой категории, границы которой определить пока трудно даже семантически, но при помощи неопределенного местоимения *несколько* эта категория резко сужается, из нее извлекается часть, которая, однако, не имеет четких границ): *несколько людей на опушке*;

е) комплексность подразумеваемого множества со скрытыми границами (*разговоры и застолье* только подразумевают наличие людей, но не дают указания ни на количество, ни на границы): *разговоры на природе; застолье в темноте*.

ИКА_gen. Генеративные языковые модели предлагают меньше вариантов концептуализации людей на анализируемых двух картинах:

а) единичность, передаваемая номинациями единичных объектов (*дама и кавалер*, хотя на самом деле на картине изображено пять фигур): *картина с изображением дамы и кавалера в саду*;

б) множественность неоднородного объекта с четкими границами (*парочка*): *парочка в необычном положении среди листвы*;

в) единичность, выделенная из множественности неоднородного объекта с четкими границами (*один из них...*): *группа людей за круглым столом, один из них смотрит на зрителя*;

г) комплексность подразумеваемого множества со скрытыми границами (аналогично информантам, модели используют номинации событий, которые не могли бы состояться без участия людей, но эта категория выражена имплицитно, а границы ее не заданы – *флирт и игра*): *изображение сцены флирта и игры в парке*.

ИКА_image. Модели данного типа оперируют тремя видами концептуальных структур, сочетающих в себе признаки комплексности и ограниченности:

а) множественность, маркированная употреблением числительных в препозиции к существительному: *a painting of three men sitting at a table* (<https://huggingface.co/spaces/hysts/image-captioning-with-git>) (изображение трех мужчин, сидящих за столом);

б) множественность неоднородного объекта с четкими границами: *a painting of a group of people on a bench* (<https://huggingface.co/nlpconnect/vit-gpt2-image-captioning>) (изображение группы людей на скамейке).

Если мы сделаем матрицу соответствия вариантов концептуализации, полученных в трех выборках, то получим следующее (табл. 1): для ЕКА характерны все виды выражения множественности, но отсутствуют случаи концептуализации группы персонажей на картине через категорию единичности; ИКА_image при осмыслении пространства многофигурной картины все-таки тяготеют к категории единичности или множественности неоднородного объекта, но с четкими границами; ИКА_gen занимают промежуточное положение – они так же, как ИКА_image,



тяготееют к единичности даже при осмыслении многофигурных композиций, но, с другой стороны, демонстрируют способность к актуализации категории множественности в двух ее разновидностях (множественность неоднородного объекта с четкими границами и множественность подразумеваемого множества со скрытыми границами).

Таблица 1. Реализация различных комбинаций комплексности, ограниченности, отдельности тремя типами когнитивных агентов для концептуализации объектов на картине группы 3 (группа людей)

Вид плексности/тип выборки	ЕКА	ИКА_gen	ИКА_image
Единичность (например, <i>мужчина</i>)	-	+	+
Комплексность множества единичных объектов (например, <i>сыновья</i>)	+	-	-
Множественность неоднородного объекта с четкими границами (например, <i>пятеро</i>)	+	+	+
Множественность неоднородного объекта с нечеткими границами (например, <i>дворяне</i>)	+	-	-
Комплексность подразумеваемого множества со скрытыми границами (например, <i>застолье</i>)	+	+	-
Комплексность неопределенного множества, извлеченного из другого неоднородного объекта с нечеткими границами (например, <i>несколько людей</i>)	+	-	-
Единичность, выделенная из множественности неоднородного объекта с четкими границами (например, <i>один из группы</i>)	-	+	-

Что касается ограниченности, то еще добавим, что на картине Веласкеса информанты концептуализируют темную фигуру, стоящую за столом, как объект, не имеющий четких границ: страшная картина с отцом и двумя сыновьями и *каким-то чудищем* посередине за столом. ИКА_gen эту фигуру не концептуализируют: (ChatGPT-4.o) *На заднем плане вообще ничего особенного нет, просто стена.*

Интересен пример, где сама по себе эмоциональная оценка ситуации представляется как сущность, границы которой, однако, трудно



провести: Посиделка, обычная пятница, мрак, застолье, пьянка, *вроде весело, а вроде плакать хочется*.

У моделей такой неопределенной концептуализации не встречается.

Для осмысления отдельности объектов и информанты, и модели пользовались схемой «X с Y, где Y – то, что находится рядом с X, а X – одушевленная или неодушевленная сущность»: *человек с бутылкой; мальчик с колбой; за столом с едой; небо с облаками; подносы с едой* (ChatGPT-4.0); *a painting of four people sitting on the ground with trees in the background* (<https://huggingface.co/spaces/hysts/InstructBLIP>) (изображение четырех людей, сидящих на земле, с деревьями на фоне).

5.3.2. Фигура и фон

ЕКА. В качестве фона для концептуализируемой сцены выступают как природные объекты, так и временные отрезки: *картина XVII в. на природе; разговоры на природе XVII в.*

ИКА_gen. Генеративные языковые модели концептуализируют в качестве фона как натурфакты, так и артефакты: *парочка в необычном положении среди листвы; ...архитектурные элементы на фоне.*

ИКА_image. Модели не используют схему «фигура – фон».

Обсуждение

Последовательно рассмотрим сначала специфику реализации конфигурационных структур у трех типов когнитивных агентов, а затем – особенности реализации схемы «фигура – фон».

Комплексность. У ЕКА наблюдаем большую гибкость переходов от единичности к множественности для концептуализации одного и того же объекта, чем у ИКА_gen и ИКА_image: если первые с известной степенью регулярности могут осмыслять объект, как вычленив его из группы (принцип селективности), так и объединяя его с другими (принцип группировки), то ИКА_gen способны производить такие трансформации лишь в редких случаях, а ИКА_image предпочитают концептуализировать всякий объект как единичный.

Ограниченность. ЕКА могут концептуализировать как всю сцену, так и любой объект в ней, в том числе – временные интервалы, мыслимые как сущности, в качестве объектов с нечеткими границами. Для того чтобы объект был концептуализирован как тот, у которого нечеткие границы, достаточно, чтобы он рассматривался ЕКА как «неважный», «незнакомый», «неопознанный». Степень размытости границ может варьировать: от практически отсутствующих,



но имплицитных границ (*флирт* – значит где-то есть категория с размытыми границами «люди») до эксплицитно нечетких границ (*что-то странное*). При этом нередко объекты концептуализируются как слабо очерченные составляющие неоднородного объекта с нечеткими границами (*несколько людей*) – получается некая «пирамида неопределенностей», которая тем не менее попадает «в кадр» концептуализации.

ИКА_gen прибегают к концептуализации объекта на картине как объекта с нечеткими границами в редких случаях: когда модели галлюцинируют и описывают то, чего на изображении нет, и когда пытаются определить сцену целиком (*сцена флирта*). В остальных же случаях ИКА_gen концептуализируют объекты как имеющие четкие, а в ряде случаев – подчеркнута четкие границы.

Аналогичная характеристика справедлива и для ИКА_image, которые в нашем материале ни разу не использовали концептуализацию сущности как объекта с нечеткими границами.

Отдельность. Все три выборки – ЕКА, ИКА_gen и ИКА_image – объединяет то, что они склонны объединять несколько соположенных в пространстве объектов в один, представляемый как неоднородный (*женщина с мечом, ваза с цветами*). Однако только ЕКА способны объединять сущности в один объект по принципу ситуативной субъективной близости – они могут находиться довольно далеко друг от друга в пространстве картины, но попадать в одно «окно внимания» Наблюдателя в силу того, что Наблюдатель «связал» их субъективно значимой связью. ИКА_gen концептуализируют как единый неоднородный объект целое и его части, а ИКА_image – человека и отличительную яркую деталь его гардероба или человека и предмет, который он держит в руках.

Другие важные особенности концептуализации пространства когнитивными агентами разных типов становятся заметны, когда речь идет о концептуализации объектов, находящихся друг с другом не только в пространственных, но и в причинно-следственных отношениях (например, Юдифь сначала убила Олоферна и теперь попирает его голову). В этом непростом случае только ЕКА демонстрируют бóльшую гибкость в способах концептуализации головы Олоферна как следствия убийства и ноги Юдифи – как символа победы убийцы. Когда Наблюдатель фиксируется на голове как следствии убийства, она концептуализируется как часть, отчужденная от другого объекта (*отрубленная голова – ее отрубили от тела*); когда такой фиксации нет – как неотчужденная часть объекта (голова мужчины) или как отдельный объект (*валяется голова*). Нога же может вообще не выделяться, ни как объект, ни как его часть (Юдифь попирает...), быть составляющей неоднородного объекта «Юдифь» (женщина с ногой на голове...), частью объекта «Юдифь» (нога женщины на голове...), отдельным объектом (нога на голове).



Можно предположить, что в первом случае Наблюдатель осознает ответственность мифологического персонажа за убийство, в каждом последующем случае – все в меньшей степени, и наконец – в последнем случае не связывает причину и следствие вовсе.

Интересно, что ИКА_image голову и ногу в этом случае вообще не «замечают», фиксируясь на мече, который женщина держит. ИКА_gen хотя и используют несколько способов концептуализации отрубленной головы, как правило, не имплицитуют причинно-следственные связи. Только одна модель из пяти (ChatGPT-4.o) концептуализировала голову как отчужденную от некоторого другого объекта, имплицитовав идею убийства.

ЕКА активно пользуются концептуальной схемой «фигура – фон», где в качестве фона могу выступать не только натурфакты (*лес, небо, пещера* и т.д.), но и артефакты (*комната*), а также временные отрезки (*XVIII в., эпоха Возрождения*) и динамические состояния природы (*смеркалось*). ИКА_image вообще не задействуют схему «фигура – фон», ИКА_gen используют в качестве фона только натурфакты и артефакты, никогда – время или динамическое состояние природы, ситуацию.

Если мы сравним между собой отличия в топологических схемах, используемых для концептуализации трех групп картин, то можно отметить следующее.

Для концептуализации картин, изображающих отдельную фигуру в некотором предметном окружении (группа 1), все группы когнитивных агентов используют наиболее разнообразные из доступных им схемы концептуализации объектов, связанные с различением «фигура и фон».

Для концептуализации сущностей на картинах, изображающих отдельную фигуру, включенную в сложную символически нагруженную ситуацию (группа 2), две из трех групп когнитивных агентов (ЕКА и ИКА_gen) фокусируются на схемах концептуализации степени *отдельности* объектов, но наибольшим разнообразием концептуальных схем этого типа отличаются естественные когнитивные агенты.

Для концептуализации объектов в многофигурных композициях все виды когнитивных агентов активнее, чем в других группах, используют топологические схемы *комплексности* и *ограниченности*.

Заключение

В целом мы наблюдаем наибольшее отличие между способами осмысления пространства, используемыми ЕКА и ИКА_image, что, как представляется, связано с тем фактом, что ЕКА и ИКА_image являются двумя разными типами Наблюдателей. ЕКА обладает



когнитивной системой интегративного типа, глубоко фундированной в телесности (т.е. воплощенной, см., например, [Chemero, 2011]) и одновременно способной благодаря языку формировать вокруг себя поле «вторичных консенсуальных взаимодействий» [Kravchenko, 2007; Кравченко, 2022], т.е. взаимодействий с объектами, которые отсутствуют в поле непосредственного восприятия, но их как бы можно «опредметить» и поместить в это поле, манипулируя соответствующими языковыми знаками. Кроме того, ЕКА обладают развитыми навыками когнитивной и социальной эмпатии [Filimon et al., 2007]. ИКА_image также берут на себя все полномочия Наблюдателя, но не обладают при этом ни телесностью, ни способностью к социальному предвосхищению, ни интегративностью. Их удел пока – это «наблюдение» того, что «вложено» разработчиками в выборки данных, использованные для обучения моделей, а также типовые фреймы «восприятия», от которых модели отклониться не могут.

Иной случай – ИКА_gen. Это модели, которые не берут на себя роль Наблюдателя – они просто интерпретируют связи и паттерны, заключенные в текстах, порожденных ЕКА и отражающих вербализованный опыт осмысления пространства ЕКА, и на этой основе порождающие свои, по сути, вторичные концептуализации пространства. Именно поэтому, как представляется, они достигают большего успеха, чем ИКА_image.

Тем не менее в своих попытках осмыслить пространство ни ИКА_image, ни ИКА_gen не обнаруживают ряда признаков, свойственных ЕКА и, по-видимому, представляющих собой косвенные свидетельства того, что Т. Нагель называл ощущением «каково это – быть данным организмом». Среди специфических для ЕКА признаков осмысления объекта в пространстве – модальность неуверенности, проистекающая от субъективного ощущения «незнания»; субъективная перцептивность в объединении разных объектов в случайную конфигурацию «единый, но неоднорный объект» по принципу «бриколажа», хорошо описанному в штудиях по мифопоэтическому мышлению; понимание зависимости восприятия от времени и эпохи (юноша/девушка с лютней); способность к опредмечиванию времени и эмоций; способность имплицировать непосредственно не данные в изображении связи между объектами. Иными словами, именно перечисленные признаки в текстовой продукции ЕКА демонстрируют на нашем локальном материале свидетельства присутствия познающего субъекта, обладающего сознанием.

Перспективу исследования составляет дообучение предобученной для русского языка Большой языковой модели на датасете описаний картин, полученных от информантов, для дальнейшего использования в задаче генерации описаний картин из коллекции Государственного Эрмитажа.



Список литературы

Витяев и др., 2020 – *Витяев Е.Е., Гончаров С.С., Свириденко Д.И.* О задачном подходе в искусственном интеллекте и когнитивных науках // *Сибирский философский журнал*. 2020. № 2. С. 5–29. <https://doi.org/10.25205/2541-7517-2020-18-2-5-29>

Калуцкая, 2010 – *Калуцкая А.П.* Моделирование информационного взаимодействия когнитивного агента с внешней средой на основе псевдофизических логик и обобщенных ограничений: автореф. ... канд. тех. наук. М., 2010. 25 с.

Касавин, 2024 – *Касавин И.Т.* Эпистемический лидер в динамике культуры. Пример искусственного интеллекта // *Цифровой ученый: лаборатория философа*. 2024. Т. 7. № 1. С. 62–73. DOI: 10.32326/2618-9267-2024-7-1-62-73.

Кравченко, 2022 – *Кравченко А.В.* Экологическая парадигма в исследованиях языка: смена познавательных установок // *Лингвоэкология: проблемы и пути их решения: монография / Под ред. А.П. Сквородникова, Г.А. Копниной*. Красноярск: Сиб. федер. ун-т, 2022. С. 5–27.

Поспелов, 1986 – *Поспелов Д.А.* Ситуационное управление. Теория и практика. М.: Наука, 1986. 288 с.

Прасолов, 2012 – *Прасолов В.В.* Наглядная топология. М., 2012. 112 с.

Рахилина, 2002 – *Рахилина Е.В.* Основные идеи когнитивной семантики // *Современная американская лингвистика: Фундаментальные направления*. М., 2002. С. 370–388.

Скребцова, 2018 – *Скребцова Т.Г.* Когнитивная лингвистика: классические теории, новые подходы. М.: ЯСК, 2018. 391 с.

References

Vityaev, E.E., Goncharov, S.S., Sviridenko, D.I. “O zadachnom podkhode v iskusstvennom intellekte i kognitivnykh naukakh” [On the Problem Approach in Artificial Intelligence and Cognitive Sciences], *Sibirskii filosofskii zhurnal* [Sibirskii filosofskii zhurnal], 2020, no. 2, pp. 5–29. <https://doi.org/10.25205/2541-7517-2020-18-2-5-29>. (In Russian)

Kalutskaya, A.P. *Modelirovanie informatsionnogo vzaimodeistviya kognitivnogo agenta s vneshnei sredoi na osnove psevdofizicheskikh logik i obobshchennykh ogranichenii*: avtoref. ... kand. tekh. nauk [Modeling of Information Interaction of a Cognitive Agent with the External Environment Based on Pseudophysical Logics and Generalized Constraints: Abstract of the Candidate of Technical Sciences]. Moscow, 2010. (In Russian)

Kasavin, I.T. “Jepistemicheskiy lider v dinamike kul'tury. Primer iskusstvennogo intellekta” [Epistemic Leader in the Dynamics of Culture. Example of Artificial Intelligence], *The Digital Scholar: Philosopher's Lab*, 2024, vol. 7, no. 1, pp. 62–73. <https://doi.org/10.32326/2618-9267-2024-7-1-62-73>. (In Russian)

Kravchenko, A.V. “Ekologicheskaya paradigma v issledovaniyakh yazyka: smena poznavatel'nykh ustanovok” [The Ecological Paradigm in Language Research:



Changing Cognitive Attitudes], in: A.P. Skovorodnikov, G.A. Kopnina (eds.): *Lingvo-ekologiya: problemy i puti ikh resheniya* [Linguoecology: Problems and Solutions]. Krasnoyarsk: Sib. feder. un-t, 2022, pp. 5–27. (In Russian)

Pospelov, D.A. *Situatsionnoe upravlenie. Teoriya i praktika* [Situational Management. Theory and Practice]. Moscow: Nauka, 1986. (In Russian)

Prasolov, V.V. *Naglyadnaya topologiya* [Visual Topology]. Moscow, 2012, 112 pp. (In Russian)

Rakhilina, E.V. “Osnovnye idei kognitivnoi semantiki” [The Main Ideas of Cognitive Semantics], in: *Sovremennaya amerikanskaya lingvistika: Fundamental’nye napravleniya* [Modern American linguistics: Fundamental directions]. Moscow, 2002, pp. 370–388. (In Russian)

Skrebtsova, T.G. *Kognitivnaya lingvistika: klassicheskie teorii, novye podkhody* [Cognitive Linguistics: Classical Theories and New Approaches]. Moscow: YaSK, 2018. (In Russian)

Chemero, 2011 – Chemero, A. *Radical Embodied Cognitive Science*. MIT Press, 2011.

Filimon, Nelson, Hagler, Sereno, 2007 – Filimon, F., Nelson, J.D., Hagler, D.J., & Sereno, M.I. “Human Cortical Representations for Reaching: Mirror Neurons for Execution, Observation, and Imagery”, *NeuroImage*, 2007, no. 37 (4), pp. 1315–1328. DOI: <https://doi.org/10.1016/j.neuroimage.2007.06.008>.

Ji, Elmoznino, Deane et al. – Ji, X., Elmoznino, E., Deane, G., Constant, A., Dumas, G., Lajoie, G., Simon, J., Bengio, Y. Sources of Richness and Ineffability for Phenomenally Conscious States // arXiv:2302.064032023. <https://doi.org/10.48550/arXiv.2302.064>

Kravchenko, 2007 – Kravchenko, A.V. “Essential Properties of Language, or, Why Language Is Not a Code”, *Language Sciences*, 2007, vol. 29, no. 5, pp. 650–671. <https://doi.org/10.1016/j.langsci.2007.01.004>

Lillicrap, Santoro et al., 2020 – Lillicrap, T., Santoro, A., Marris, L., Akerman, C.J., Hinton, G. “Backpropagation and the Brain”, *Nat Rev Neurosci*, 2020, no. 21 (6), pp. 335–346. DOI: 10.1038/s41583-020-0277-3.

Maturana, Varela, 1987 – Maturana, H.R., Varela, F.J. *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston: Shambhala Publications, 1987.

Nagel, 1974 – Nagel, T. “What Is It Like to Be a Bat?”, *The Philosophical Review*, 1974, no. 83 (4), pp. 435–450.

Russell, Norvig, 2020 – Russell, S., Norvig, P. *Artificial Intelligence: A Modern Approach*, Pearson, 4th Edition, 2020.

Schatten, Đurić, Peharda, 2021 – Schatten, M., Đurić, B. O. and Peharda, T. “A Cognitive Agent for University Student Support”, *2021 IEEE Technology & Engineering Management Conference – Europe (TEMSCON-EUR)*, Dubrovnik, Croatia, pp. 1–6. DOI: <https://doi.org/10.1109/TEMSCON-EUR52034.2021.9488627>.

Talmy, 2000a – Talmy, L. *Towards a Cognitive Semantics*. Vol. 1. Concept Structuring Systems. MIT Press, 2000.

Talmy, 2000b – Talmy, L. *Towards a Cognitive Semantics*. Vol. 2. Typology and Process in Concept Structuring. MIT Press, 2000.

Talmy, 2018 – Talmy, L. *The Targeting System of Language*. MIT Press, 2018.

Touvron, Lavril, Izacard et al., 2023 – Touvron, H., Lavril, T., Izacard, G, Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N, Hambro, E., Azhar, F.,



Rodriguez, Au., Joulin, A., Grave, E., and Lample, G. “Llama: Open and Efficient Foundation Language Models”. *ArXiv*, abs/2302.13971, 2023. DOI: <https://doi.org/10.48550/arXiv.2302.13971>.

Xue, Shu, Awadalla et al., 2024 – Xue, L., Shu, M., Awadalla, A., Wang, J., Yan, A., Purushwalkam, S., Zhou, H., Prabhu, V., Dai, Y., Ryoo, M.S., Kendre, Sh., Zhang, J., Qin, C., Zhang, Sh., Chen, Ch.-Ch., Yu, N., Tan, J., Manoj, T., Heinecke, Sh., Wang, H., Choi, Y., Schmidt, L., Chen, Z., Savarese, S., Niebles, J.-C., Xiong, C., Xu, R. *xGen-MM (BLIP-3): A Family of Open Large Multimodal Models*. arXiv:2408.08872v2. 2024.