



ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И (ПОСТ)СТРУКТУРНАЯ СЕМАНТИКА¹

Диана Эдиковна Гаспарян – кандидат философских наук, доцент. Национальный исследовательский университет Высшая школа экономики, факультет философии, кафедрa онтологии, логики и теории познания.
E-mail: anaid6@yandex.ru.

В современных исследованиях по философии сознания, ориентированных на отождествление или максимальное сближение человеческого и машинного интеллекта, выделяются две главные трудности: проблема машинной эмуляции значения и проблема машинной эмуляции смысла. В настоящем исследовании анализ этих проблем ведется с точки зрения положений структурной и постструктурной лингвистики, практически не задействованных философией сознания. Посредством обращения к базовым определениям знака, значения и смысла, принятым в структурализме и постструктурализме, раскрывается принципиальное различие между возможностями машинной и собственно психической обработки знака. Широко обсуждаемое в современной философии сознания различие между синтаксическими и семантическими возможностями интеллекта получает в данном исследовании новое, возможно, наиболее наглядное подтверждение.

Ключевые слова: искусственный интеллект, сильный искусственный интеллект, слабый искусственный интеллект, синтаксис, семантика, значение, смысл, структурализм, постструктурализм, нарратив.

ARTIFICIAL INTELLIGENCE AND (POST)-STRUCTURAL SEMANTICS

Diana Gasparyan – Candidate of philosophical sciences, Associate Professor. National Research University Higher School of Economics, Department of Philosophy.



There are two problem areas associated with modern studies of philosophy of mind focusing on identification and convergence of human and machine intelligence. One problem is machine simulation of meaning and the other – machine simulation of sense. In the present study the analysis of the stated problems is carried out based on the concepts of structural and post-structural linguistics almost entirely ignored by philosophy of mind. If we refer to the basic definitions of “sign” and “meaning” found in structuralism and post-structuralism, we will see a fundamental difference between the capabilities of a machine and the human brain engaged in processing of a sign. The present study will exemplify and provide more and probably stronger evidence to support distinction between syntactic and semantic aspects of intelligence, the issue widely discussed by adepts of contemporary philosophy of mind.

Key words: Artificial Intelligence, Strong Artificial Intelligence, Weak Artificial Intelligence, Meaning, Sense, Reference, Semantics, Syntaxes, Structural Semantics, (Post)-Structural Semantics, Narrative.

¹ Выполнено в рамках проекта Центра фундаментальных исследований НИУ ВШЭ «Следование правилу: рассуждение, разум, рациональность», 2014.



Введение

Бурный рост философских исследований по проблематике сознания, имеющий место в наши дни, был во многом подготовлен развитием таких отраслей науки, как когнитивная психология, нейропсихология и даже нейрофизиология. Программы изучения мозга являются довольно перспективными и успешно поддерживаются на самых разных уровнях [Nilsson, 2005: 36–38]. Вместе с тем специфика современных исследований по философии сознания заключается в заметной дисциплинарной замкнутости, а именно слабой ориентации на некоторые контексты континентальной философии. В частности, ведущиеся исследования по одному из самых активно обсуждаемых направлений современной философии сознания, связанному с возможностью отождествления работы человеческой психики и машинного интеллекта, практически не принимают во внимание определения знака, значения и смысла, развиваемых такими разделами философии, как структурализм и постструктурализм. Речь идет о том, что в существующих на сегодняшний день дискуссиях по философии сознания довольно четко обозначена проблема несовпадения синтаксической работы интеллекта, отвечающей за установления значений, и семантической работы интеллекта, ответственной за фиксацию смысла. При этом данная постановка проблемы претерпела ряд серьезных изменений за последние 15–20 лет [Duch et al., 2008].

Ранние исследования по философии сознания практически не различали функцию установления значения и функцию установления смысла. Эта неразличенность в свою очередь вызывала бурный рост программ *искусственного интеллекта* (ИИ) и укрепляла оптимизм исследователей в скорейшем разрешении определенных проблем сознания посредством создания технического (компьютерного) аналога человеческого интеллекта [Brachman, 2005]. Позднее благодаря усилиям ряда исследователей и в первую очередь американского философа Дж. Сёрла удалось показать принципиальное различие между синтаксической (машинной) и семантической (собственно психической) обработкой информации и тем самым объяснить истинную причину неудач создания ИИ [Searle, 1991]. Однако спустя некоторое время подход Сёрла претерпел серьезную критику не только сторонников ИИ [Denett, 1998], но и противников [Bringsjord, 1999], так как при детальном изучении аргументации Дж. Сёрла и его последователей обнаружилось, что она не вполне ослабляет позиции сторонников ИИ, но, указывая на невозможность создания машинного аналога разума, скорее апеллирует к интуиции, чем к серьезным логическим аргументам против такой возможности. Поскольку создание машинного интеллекта по-прежнему представляется логически выполнимым,



большая часть существующей на сегодняшний день критики не выходит за пределы апелляции к здравому смыслу или интуиции.

В настоящее время все чаще встречается мнение о том, что концепция так называемого сильного ИИ, согласно которой наличие сознания эквивалентно выполнению программы, практически не имеет шансов на подтверждение [Cohen, 2005], в то время как концепция слабого ИИ, согласно которой действия человеческого интеллекта могут успешно эмулироваться на компьютере, имеет хорошие позиции [Duch, 2008]. Однако противники ИИ планировали демонстрацию невозможности любой эмуляции, в том числе и слабой, и в этом смысле не добились поставленных целей. Решающий аргумент против способности программы к осуществлению семантических операций так и не был предъявлен [Duch, 2008]. Кроме того, те аргументы, которые продолжают предлагаться, лежат в области демонстрации эмпирической невозможности машинного интеллекта, в то время как наиболее сильным аргументом могла бы стать демонстрация логической (аподиктической) невозможности. Вместе с тем можно попробовать продемонстрировать именно логическую проблематичность осуществления машиной семантических процедур. Эта демонстрация наиболее наглядна в случае обращения к неклассической (структурной и постструктурной) теории знака, значения, смысла и в целом языка.

Речь идет о том, что если принять во внимание принципиальную несводимость семантического приращения к конфигурации значений, а также трудности формализации смысла вообще, то можно указать на закономерный характер неудачных попыток создать машинные модели обработки информации, имитирующие работу человеческой психики, объяснив попутно причины и механизм подобных неудач.

Исходя из сказанного я попробую прояснить следующие вопросы:

- ◇ показать, что в случае как сильной, так и слабой версий ИИ недооцененным оказывается понятие смысла, заменяемое понятием значения;
- ◇ продемонстрировать, что существуют некоторые принципиальные трудности формализации смысла;
- ◇ показать, что с помощью *(пост)структурного определения* смысла можно заметно усилить аргументацию противников ИИ.

Трудности синтаксиса и семантики

Для начала определим, как понимается сознание с точки зрения философских программ ИИ. Согласно данным программам, сознание есть функциональный (и, следовательно, формализуемый) процесс, который может быть воспроизведен на любом носителе [Turing,



1950]. В свою очередь под концепцией сильного ИИ понимается такое представление о сознании, где компьютерная программа является не просто моделью сознания, но в полном смысле самим сознанием. Адекватно функционирующая программа реализует сознание в том же смысле, в котором оно присуще человеку. Напротив, согласно концепции слабого ИИ, программа есть лишь инструмент, позволяющий решать те или иные задачи, которые не требуют полного спектра человеческих познавательных способностей.

Чтобы понять источник оптимизма сторонников ИИ, необходимо вкратце вскрыть предысторию вопроса – почему с некоторых пор идея о том, что машина может «думать», перестала быть метафорой. Причиной тому послужило успешное создание программ, которые весьма эффективно эмулировали человеческое мышление. «Общение» с такой программой подчас не позволяло отличить машину от живого человека – машина давала не менее адекватные, резонные и любопытные ответы, чем реальный собеседник из плоти и крови. Такой успех позволил некоторым философам сделать простой вывод – если машина ведет себя так же, как человек, то на каком основании мы продолжаем утверждать, что машинный интеллект отличен от человеческого? Напротив, у нас есть все основания их отождествить, так как, кроме критерия внешнего проявления (даются ли осмысленные ответы, последовательно ли и логично ли рассуждение, есть ли приращение информации и т.д.), у нас больше ничего нет [McDermott, 1997]. Толчком для такого сближения послужили знаменитые тесты Тьюринга, в которых перед человеком (Экспертом), разговаривающим с машиной, ставилась задача определить, имеет ли он дело с другим человеком по ту сторону монитора или с Программой, написанной программистами специально для того, чтобы ввести в заблуждение Эксперта.

Приведем реальные примеры такого общения, из которых можно увидеть, насколько подчас убедительными могут выглядеть ответы программы.

Эксперт 1: Считаете ли вы, что не история делает политика, а политик историю?

Программа 1: По-моему, это не более чем игра слов.

Эксперт 1: Полагаете, большинство людей с вами согласится?

Программа 1: А что вы подразумеваете, когда спрашиваете о большинстве? На мой взгляд, большинство людей со мной согласится [Cohen, 2005: 67].

Эксперт 2: Я говорю. Я люблю. Я живу. А ты умеешь что-нибудь из этого?

Программа 2: Интересно, ты когда-нибудь думаешь? А то у меня такое чувство, что эксперт здесь я! [Cohen, 2005: 61]

Как видим, ответы программы выглядят довольно естественно, и если что-то и возбуждает наше подозрение, то уличить «отвечаю-



щего» в том, что он Программа, а не человек, будет вдвое сложнее, если принять во внимание, что по условиям теста реальный человек в качестве отвечающего может пытаться сбить с толку Эксперта. Но поскольку задача Эксперта состоит в том, чтобы продемонстрировать, что он в любом случае отличит машинный интеллект от человеческого, любая его ошибка или даже замешательство будет указывать на отсутствие надежных критериев по их различию. Кроме того, можно утверждать, что если ответы укладываются в рамки нормы и Эксперт не наблюдает очевидных погрешностей и нестыковок, то машинный интеллект справляется с теми же задачами, с которыми справляется человек [Turing, 1950]. Вывод, который из этого делают сторонники ИИ: нет никаких препятствий к тому, чтобы утверждать единство и универсальность интеллекта, одинаково присущего как машинам, так и людям [Denett, 1998].

Однако на смену подобному оптимизму довольно быстро пришло разочарование. Немалую роль в переломе настроений сыграла критика идеи машинного интеллекта, предпринятая Дж. Сёрлом в его знаменитом аргументе «Китайская комната» [Searle, 1991]. Аргумент представляет собой вывод из следующего мыслительного эксперимента. Представим, что перед нами ставится задача говорить на том языке, которого мы не знаем (например, китайском). Как кажется, это невозможно. Однако, проявив находчивость, мы сможем сделать вид, что владеем китайским, и даже ввести в заблуждение китайцев. Для этого нам будет достаточно обзавестись учебником на понятном нам языке, в котором приводятся правила сочетания символов китайского языка, причем правила эти можно применять, зная лишь форму символов, понимать их значение совсем необязательно. Например, правила могут гласить: «Возьмите такой-то иероглиф из корзинки номер один и поместите его рядом с таким-то иероглифом из корзинки номер два». Тогда люди, говорящие по-китайски и находящиеся, скажем, за дверью той комнаты, куда помещен наш испытуемый вместе со всеми учебниками, словарями и справочниками, передавая в комнату наборы символов, будут получать в ответ вполне осмысленные ответы. При этом человек, запертый в комнате, будет лишь манипулировать символами согласно правилам и передавать обратно неясные ему, но ясные конечным адресатам наборы символов. Например, люди, находящиеся снаружи, могут передать непонятные сидящему в комнате человеку символы, означающие: «Какой цвет вам больше всего нравится?» В ответ, выполнив предписанные правилами манипуляции, человек, запертый в комнате, выдаст символы ему хоть и непонятные, но означающие, что его любимые цвета синий и зеленый. Таким образом, он легко выдержит тест Тьюринга на понимание китайского языка, при этом не понимая ни слова по-китайски. В этом примере книга правил будет аналогом компьютерной программы,



люди, написавшие ее, – программистами, человек, запертый в комнате, – компьютером, наборы символов, передаваемых в комнату, – вопросами, а наборы, выходящие из комнаты, – ответами [Searle, 1991].

Сёрловский мыслительный эксперимент показывает, что тот факт, что машина может вести с человеком вполне осмысленную беседу, вовсе не означает, что машина может думать. Согласно аргументу Сёрла, программа будет успешно справляться с задачей внятной коммуникации, если только будет ставить в соответствие одним символам (вопросам) другие (ответы). Программа исключительно *синтаксична*, т.е. в ней принимается во внимание лишь начертание символов. *Значение их никак не раскрывается*. Человек, работающий в качестве такой программы, видит непонятные значки и ставит им в соответствие другие значки. При этом «китайские» собеседники полагают, что он ведет с ними вдумчивый разговор, на деле же он лишь механически соединяет загадочные иероглифы. Однако реальное человеческое понимание предполагает не только синтаксические правила, но и владение *значением* слова, т.е. по сути знание того, как выглядит денотат. Программа должна была бы обладать опытом «синего и зеленого цвета» – в этом случае ее «знание» терминов «синий» и «зеленый» было бы семантическим и подобным человеческому. Но поскольку на это машинный интеллект не способен, то он не обладает реальным пониманием. Таким образом, даже если компьютер проходит тест Тьюринга, т.е. демонстрирует вербальное поведение, неотличимое от человеческого, из этого еще не следует, что он обладает подлинным интеллектом. Компьютерные программы работают исключительно с синтаксическими операциями и потому носят сугубо формальный характер. Такой вывод делает Сёрл [Searle, 1991].

Семантика: работа не только со значением, но и со смыслом

Между тем, согласившись с выводами Сёрла, можно попробовать показать, что «интеллект» программ не может быть приравнен к интеллекту человека не только по критерию невосприимчивости к значению, но в первую очередь по критерию невосприимчивости к *смыслу*. По Сёрлу, сознание понимает что-либо потому, что апеллирует к знанию денотата. Иными словами, «понимать» значит устанавливать значения (денотаты). Можно, однако, показать, что сознание человека в первую очередь восприимчиво к смыслу, т.е. знанию коннотата. Это важно потому, что коннотативное измерение интеллекта затрудняет его (интеллекта) формализацию еще больше и делает более очевидным разрыв между классом задач, с которыми может справиться



машина, и задачами, доступными человеческому интеллекту. Иными словами, важность проведения различия между значением и смыслом представляется резонной в силу того, что понимание скорее относится к установлению смысла, нежели к установлению значения.

Смысл я буду понимать в трех его интерпретациях – в версии, предложенной Г. Фреге, и версиях, предложенных структурной и соответственно постструктурной лингвистикой. И фрегевское и структурное определения смысла восходят еще к стоикам. Именно они впервые предложили различать смысл и значение в трехстороннем разделении знака на означающее (то, что слышат эллины и варвары), обозначаемое (то, что видят эллины и варвары) и мысль («лектон» – смысл) (то, что эллины понимают, а варвары не понимают) [Секст Эмпирик, 1975; Edelstein, 1966].

В наиболее оформленном и отрефлексированном виде принцип разграничения смысла (Sinn) и значения (Bedeutung) проводится Фреге, которого принято считать родоначальником оппозиции смысла и значения. У Фреге значение есть то, что соответствует сказанному (референт или денотат), а смысл является способом (репрезентацией), каким задается значение, т.е. «способом языковой данности». Или по-другому, референт – это *то, о чем* сообщается (содержание), а смысл – *то, что* сообщается о референте (выражение). Тогда смысл есть ситуация несовпадения того, что сказано, с тем, о чем это сказано. Например, два высказывания «Утренняя звезда» и «Вечерняя звезда» равно имеют своим референтом «звезду Венеру», но выражают разный смысл. Также выражения «полководец, одержавший победу при Аустерлице» и «полководец, потерпевший поражение при Ватерлоо», имеют разный смысл при одинаковом референте – Наполеоне Бонапарте. Смысл входит в структуру *Знака* (материального носителя нематериального смысла, отсылающего к материальному значению, например след чернил на бумаге при написании слова или колебание воздуха при его произнесении) и составляет одно из двух оснований так называемого треугольника Фреге [Frege, 1948]:





В рамках структурной лингвистики и возникшего из нее структурализма фрегеанское разделение на смысл и значение сохраняется, однако несколько меняется представление о самом смысле. Справедливости ради отметим, что структурализм никогда не ставил перед собой задачи обосновать или опровергнуть возможности создания машинного интеллекта. С помощью своего метода структурализм обещал подвергнуть эффективной «расшифровке» смысл любых культурных явлений – человеческих ритуалов, обычаев, традиций, текстов, коллективных и индивидуальных действий, верований, мифов и проч. [Косиков, 1998]. Между тем то, каким образом в структурализме вводится смысл, может помочь пролить свет на ряд положений, существенных в вопросах обсуждения перспектив создания ИИ.

Согласно структурному методу, если мы хотим понять смысл какого-либо повествования, то нам просто следует обнаружить глубинную структуру, которая обслуживает функционирование данного повествования на поверхностном уровне. С позиции структурализма, если некоторая история может быть рассказана разными способами, то у всех этих способов будет один смысл [Lévi-Strauss, 1977]. Таким образом, всякое знаковое образование рассматривается как некая конфигурация, созданная формальными отношениями между элементами различного уровня и порядка [Levin, 1974]. И тогда само «формальное» (т.е. структура) и есть то, что порождает смысл, по сути с ним совпадая. В такой интерпретации важно, что, несмотря на разность конечных версий повествования, человек, как носитель сознания и часть культуры, в состоянии увидеть общее – единство идеи, т.е. по сути единство смысла. Однако происходит это не потому, что мы знакомы с денотатом, как сказал бы Сёрл, мы знаем, что такое синий, поскольку имеем представление о значении слова «синий». Единство смысла человеческого сознание усматривает потому, что изначально является носителем глубинных структур-смыслов, и разность повествовательных рядов не может сбить его с толку – в основе разнородного материала сознанию открывается ядро смысла.

В свою очередь постструктуралистский подход не разделяет взглядов структуралистской программы. Сведение смысла к некой базовой структуре, «морали» повествования постструктурализму представляется совершенно недопустимым, так как утрачивается своеобразие каждой из рассказанных историй, представляющих свой собственный смысл. Например, для структурного подхода истории о Гамлете, Дон-Кихоте или князе Мышкине будут просто манифестациями одной базовой истории «о рефлектирующем герое, мучительно переживающем социальную несправедливость и моральную небезупречность человеческого существования». Напротив, для постструктурного подхода смыслом окажется то, как рассказана каждая конкретная история, каковы ее детали и подробности художественно-



го исполнения. Если в повествовании о Дон-Кихоте какой-то элемент не получается удачно вписать в структуру, то эти «случайные» атрибуты повествования, ускользающие из-под диктата структурного объяснения, и будут смыслами – самопроизвольными проявлениями знаковых систем.

Кроме того, согласно постструктурализму, смысл никак нельзя сводить к структуре, так как трудности формализации смысла имеют гораздо более глубокие основания [Derrida, 2001]. Структурализм предлагает редукционистскую модель объяснения смысла, где смысл отождествляется со структурой. Эффективность этой процедуры обеспечивается тем, что агенты коммуникации используют существующие знаковые сообщения, смысл которых уже накоплен в языке. Остается только найти подходящую модель (структуру), которая обеспечивает циркуляцию смыслов в цепи означающих. Так, если мы хотим понять смысл сказки о Красной Шапочке, чьи герои совершают определенные действия и отпускают определенные реплики, то нам просто следует обнаружить глубинную структуру, которая обеспечивает функционирование данного повествования на поверхностном уровне. Таким образом, всякое знаковое образование рассматривается как некая конфигурация, созданная *формальными* отношениями между элементами различного уровня и порядка. И тогда само «формальное» (т.е. структура) и есть то, что порождает смысл, по сути с ним совпадая. В свою очередь отношения и иерархия элементов и уровней понимается как нечто существующее независимо от структурной процедуры дешифровки. Исследователь (или любой агент коммуникации) может только зафиксировать смысл, который организует то, что проявлено на поверхностном уровне.

Постструктурализм сделал многое для того, чтобы ослабить оптимизм структуралистской программы. С позиции структурализма, если история про Красную Шапочку может быть рассказана разными способами, то у всех этих способов будет один смысл – тот, который залегает на уровне структуры. Постструктурализму это представляется недопустимым, так как утрачивается своеобразие каждой из рассказанных историй, представляющих каждый раз свой собственный смысл. Чтобы обозначить свое расхождение со структурализмом, постструктурализм предпочтет говорить о *событии*, впрочем, избегая напрямую противопоставлять его структуре [Косиков, 1998]. Эта терминология практически позаимствована у самого структурализма, который предпочитал говорить о «структуре» как об инварианте, а о «событии» как ее (структуры) манифестации. Например, такие сказки, как «Золушка» или «Морозко», есть просто манифестации одной протосказки «О гонимой падчерице»², а, скажем, злая мачеха – манифестация структур-

² По указателю сюжетов фольклорной сказки Аарне-Томпсона [Thompson, 1973].



ного персонажа «Вредитель». Итак, если изначально, т.е. в рамках самого структурализма, «событием» называлась реализация структуры на поверхностном уровне, то в последующем, в постструктурализме под «событием» в общем виде понимается все то, что не укладывается в рамки структурного объяснения. Точно так же, если в повествовании о Красной Шапочке какой-то элемент (допустим, возраст Красной Шапочки или то, что она направляется именно к бабушке, а не к дедушке) не получается удачно вписать в структуру, то эти «случайные» атрибуты повествования и есть «события» – самопроизвольные проявления знаковых систем [Косиков, 1998].

Впрочем, проблема невозможности исчерпывающим образом связать поверхностный и глубинный уровни повествования не была обнаружена исключительно постструктуралистами. Структурализм вполне отдавал себе отчет в наличии этих проблем. В любом повествовании, детально разобранном по «функциям» и «действующим лицам», сохранялся значительный «остаточный» *семантический* материал («характерь», «мотивация» и «взаимодействие» персонажей), ответственный за поддержание *конкретного сюжета*, но не поддающийся структурному объяснению. Так, если сказка о Красной Шапочке может быть проанализирована с точки зрения схематического ранжирования «Красная Шапочка – Адресант, Бабушка – Адресат, Волк – Противник, Охотник – Помощник»³, то при всей убедительной лаконичности данной модели необъяснимым остается главное – почему сказка имеет именно такое, а не другое конкретное исполнение (сюжет)? Последовательный структуралист, возможно, ответит, что структурный анализ осознанно абстрагируется от подобной конкретики – место (в структуре) Красной Шапочки мог бы занимать Мальчик в Желтой Бейсболке, место Волка – Лиса, а место Охотника – Добрая Лесная Фея. Однако именно этот уровень конкретной персонификации относится к *сюжетному*, который действительно актуализирует логические отношения на уровне структуры, но делает это определенным образом [Косиков, 1998].

Если мы вернемся к вопросу о смысле и спросим, к какому уровню – сюжетному (поверхностному) или структурному (глубинному) принадлежит *смысл*, то обнаружим те трудности формализации смысла, которые имеют непосредственное отношение к созданию моделей ИИ. Структурализм занимает здесь противоречивые позиции (на что потом и укажет постструктурализм). С одной стороны, он указывает, что за смысл ответственна структура. Так, смысл сказки о Золушке в том, что гонимая падчерица награждается за кротость, трудо-

³ В данном примере мы воспользовались моделью А. Греймаса, согласно которой любое произведение может быть формализовано с помощью 6 «актант» (Субъект–Объект, Адресант–Адресат, Помощник–Противник), связанных между собой отношениями модальности («желать», «знать», «мочь») и выполняющих 20 возможных функций [Греймас, 1996].



любие и терпение, а родные дочери, лишённые этих качеств, но старающиеся получить ту же награду, терпят фиаско. С другой стороны, он готов признать, что именно на уровне сюжета происходит семантизация структурных категорий, в противном случае нам бы не удалось отличить сказку о Золушке от сказки о Морозко. Прояснить эту двусмысленность как раз и позволяет концептуализация смысла, предложенная Фреге. У Фреге, как мы видели, ясно продемонстрировано, что смысл принадлежит уровню языковой данности (значения). Придерживаясь этой логики, можно сказать, что сюжет по сути есть не что иное, как *способ структурной данности*. Позиция структурализма в этом случае просто переворачивается – отныне структура предстает своего рода значением (референтом или денотатом), а ее репрезентация – смыслом. Проблемы же структурализма как раз и проистекали из старого метафизического понимания смысла как синонима значения. Именно в связи с этим недоразумением структурализму не удавалось дать исчерпывающее объяснение семантике некоторого повествования, которая не дедуцировалась из структуры напрямую. Если же дополнить структурализм новаторской методологией Фреге (хотя постструктуралисты предпочитают ссылаться на стоиков, действительно впервые отделивших смысл от значения), то можно сказать, что сюжетная семантика обладает некой автономией становления, т.е. собственной и независимой от формально-структурного уровня стихийностью порождения смыслов.

Если смысл отделяется от структуры, то становится очевидной невозможность редукционизма по отношению к литературному произведению, мифологии, явлению культуры и в целом любому нарративу. Как тогда можно формализовать смысл? На простейший вопрос о том, что обеспечивает смысл некоего повествования, оказывается не так просто ответить. Достаточно ли ответа, данного Фреге? В действительности, именно Фреге последовательно демонстрирует, что смысл не может быть локализован ни на уровне сигнификации (или означающего), ни на уровне денотации (или означаемого). Смыслом предложения «ученик Сократа и учитель Аристотеля» не является количество букв (знаков), составляющих это предложение. Также и на уровне денотации (референции) локализация смысла невозможна. «Философ Платон» – есть только значение (денотат, референт) этого предложения, но не его смысл. Что тогда означает формула «способ, каким задается значение» – принадлежит ли сам этот способ отношению сигнификации или денотации? Если смысл это не знак (сигнификация) и не вещь (денотация), но способ, каким первое говорит о втором, тогда мы должны сказать, что смысл – это некое *идеальное, нематериальное образование*, в то время как *вещь* и *знак* суть материальные объекты. Это и не акт понимания, так как смысл есть скорее то, что понимается, а не само понимание.



Трудности формализации смысла заключаются также в том, что смысл никогда не удастся зафиксировать как нечто предметное [Leont'eva, 1981]. Если нам дано некоторое предложение, то его смыслом будет другое предложение, которое в свою очередь затребует еще одного для прояснения своего собственного смысла, и так до бесконечности. Для того чтобы нечто *осмысленное* было высказано, оно должно быть сформулировано в виде предложения, состоящего из слов, для понимания которых мы должны сформулировать предложения, раскрывающие их смысл. Тем самым мы вынуждены ввести новые слова, требующие экспликации в новых предложениях, и т.д. и т.п. Если принять предложение за некое имя, то ясно, что каждое имя, обозначающее объект, само может стать объектом нового имени, обозначающего его смысл: n_1 отсылает к n_2 , n_2 отсылает к n_3 и т.д. Когда мы высказываем что-то, мы никогда не проговариваем непосредственный смысл того, о чем идет речь. Смысл того, о чем мы говорим, можно сделать разве что объектом следующего предложения, смысл которого также не проговаривается. В подобной ситуации имеет место своего рода бесконечный регресс того, что подразумевается [Deleuze, 1990].

Последнее обстоятельство напрямую вытекает из сделанных выше наблюдений – смысл не локализуется на уровне предложения. Эта проблема может показаться несколько надуманной, но мы отнесемся к ней с большим вниманием, если учтем, что она направлена не к практическому положению дел, где каким-то образом люди способны что-то понимать, а к чистой теории вопроса – что такое сам «смысл»? Ведь именно в этом случае мы могли бы задать формализацию смысла на машинном уровне. Трудности искомой формализации обнаруживаются сразу же, как только мы задаемся простым вопросом: что же понято нами в том или ином тексте? Ответить на него значит построить новый текст (в действительности любое знаковое образование – картину, жест, мелодию и проч.), который призван выразить то же самое, что было сформулировано в исходном тексте. Таким образом, смысл доступен нам только через вторичное означение. Однако нечто, сказанное по поводу другого, не является тождественным тому, о чем оно высказывается. Вторичные толкования также должны быть *поняты*, а это как раз и приводит к проблеме бесконечного регресса. Кроме того, при таком понимании, где смысл текста № 1 есть текст № 2, притом что другого понимания у нас просто нет, «смысл» отождествляется с лингвистическим понятием *синонимии*.

Когда на вопрос, в чем смысл исходного текста, отвечают вторым текстом, то, как кажется, имплицитно исходят из того, что смысл тождествен ситуации лингвистического перефразирования или синонимии. Но благодаря Фреге мы знаем, что разные способы репрезентации значения вовсе не обязательно должны состоять друг с другом в отношениях синонимического родства. Так, два высказывания «пол-



ководец, одержавший победу при Аустерлице» и «полководец, потерявший поражение при Ватерлоо», выражая единое значение, высказывают прямо противоположный смысл. Если же мы попробуем синонимировать не разные смыслы одного значения, а сам смысл со своим значением, то мы получим еще более удивительные результаты. Если мы говорим, что первое высказывание это смысл, а «Наполеон Бонапарт» – значение, то высказывание «Наполеон является полководцем, одержавшим победу при Аустерлице» как будто приравнивает две части предложения. Но если они *равны*, то все предложение должно быть тавтологией, что не соответствует действительности. Если же они *не равны*, то высказывание должно быть ложным, что, разумеется, тоже не так. Получается, что смысл, выраженный первым высказыванием, не означает то же самое, что «Наполеон», но он также и не означает что-то другое, заставляя нас заподозрить, что он вообще ничего не означает. Эту проблему впервые описал Б. Рассел в рамках своей теории «семантической пустоты», призванной избавиться от «лишней сущности» (смысла), открытой по сути Фреге. Согласно этой теории, определенные дескрипции (по Фреге – способы, какими задается предмет, т.е. С) суть «неполные символы», они ничего не обозначают.

Доказательство Рассела выглядит следующим образом: «Если бы “автор Веверлея” обозначало что-то другое, нежели “Скотт”, то “Скотт является автором Веверлея” должно было быть ложью, но это не так. Если бы “автор Веверлея” обозначало “Скотт”, то “Скотт является автором Веверлея” было бы тавтологией, но это не так. Следовательно, “автор Веверлея” не обозначает ни “Скотт”, ни что-либо еще, т.е. “автор Веверлея” не обозначает ничего» [Russell, 1959]. Возможно, этот вывод излишне радикален⁴, но он по меньшей мере призван

⁴ Расселовское доказательство было неоднократно подвергнуто критике сторонниками Фреге (П. Строссоном, У. Куйаном), указывавшими на неявное использование в самом доказательстве противопоставления смысла и значения. Так, термин «обозначать» (to mean) используется в доказательстве в одном случае как фрегеанский Sinn, а в другом как Bedeutung. Если meaning использовать как Bedeutung, то в первой посылке приводимая Расселом пропозиция «Скотт является автором Веверлея» действительно должна быть ложной (так как два выражения в данном случае обозначают различные объекты), и указание на то, что это не так, неверно. Но верно, что вторая посылка, когда речь идет о том, что «Скотт является автором Веверлея», не является тавтологией (поскольку два выражения при участии одного Bedeutung не обязательно дают тавтологию). Если же meaning использовать как Sinn, то верна первая посылка Рассела – обсуждаемое предложение действительно не является ложным (так как два выражения с различными Sinn не обязательно ведут к ложности). Однако в этом случае неверна будет вторая посылка, где, согласно Расселу, предложение не есть тавтология (так как два выражения с одним и тем же Sinn в обязательном порядке ведут к тавтологии). В итоге доказательство семантической пустоты может состояться лишь в том случае, когда в первой посылке под meaning подразумевается Sinn, а во второй под meaning подразумевается Bedeutung. Элиминирование смысла как лишней сущности в контексте создания ИИ имеет большое значение: если бы это элиминирование удалось, то шансы на создание непротиворечивой логической модели ИИ значительно бы выросли, и наоборот.



продемонстрировать непродуктивность описания смысла в терминах синонимии значений.

Если теперь соединить фрегевское определение смысла с постструктурным, то станет ясно, что смысл повествования может заключаться не в том, *что* оно выражает, а в том, *как оно рассказано*. Отсюда следует, что смысл невозможно формализовать на одном только структурном или, в терминологии теоретиков ИИ, синтаксическом уровне. Вопрос о правомочности создания машинного интеллекта как аналога человеческого следовало бы тогда ставить в несколько иных терминах, а именно – способна ли машина отличить уровень *что* от уровня *как*? Являются ли для нее сказки о Спящей Красавице и Семи Богатырях и Белоснежке и Семи Гномах двумя сказками с разными значениями, но одинаковым смыслом или, напротив, одинаковым значением, но разным смыслом? Сможет ли Программа увидеть генетическое родство между двумя этими историями или расценит их как совершенно независимые друг от друга?

Кроме того, неясно, сможет ли Программа проводить такие же различия между осмысленным и неосмысленным повествованиями, которые осуществляет человеческий интеллект. Для того чтобы увидеть, в чем здесь проблема, воспользуемся следующей иллюстрацией. Представим себе, что нам показывают фильм или дают прочитать некий рассказ, где есть условно пять сцен, никак не связанных между собой.

Повествование 1:

Сцена первая. На землю прибывает группа космических пришельцев.

Сцена вторая. Двое возлюбленных наконец обрели друг друга.

Сцена третья. В крупном музее происходит громкое ограбление.

Сцена четвертая. Группа террористов захватывает самолет с заложниками.

Сцена пятая. Посреди площади большого города люди находят бесценный клад.

Как видим, все герои в этом квазиповествовании разные. Ясно, что такое изложение приведет нас в недоумение, и мы будем правы, если скажем, что эти пять сцен оказались в одном месте по чистой случайности или по недосмотру автора. Они могли сложиться в единый нарратив и в результате чьей-то злой шутки. Скажем, некто попросту изъясил эти эпизоды из других историй и произвольно сложил в своеобразный коллаж. Здесь мы имеем дело с таким случаем, когда в повествовании нет никакой связи, или если даже она есть, то настолько искусственная, что это сразу бросается в глаза. Например, если представить, что специальная компьютерная программа будет проверять различные истории с целью определения, имеем ли мы дело с бессмысленным набором сцен или с целостным художественным



произведением, то такая программа легко справится с этой задачей. Поскольку в одном повествовании должны встречаться одни и те же имена, то описанная выше история вызовет у компьютерной программы явное подозрение.

Рассмотрим другой случай. Нам предлагают познакомиться с фильмом или рассказом, где один и тот же главный герой в тех же пяти сценах производит некоторые действия.

П о в е с т в о в а н и е 2

Сцена первая. Смит влюбляется.

Сцена вторая. Смит подвергается нападению ночных грабителей.

Сцена третья. Смит узнает о краже, совершенной в музее.

Сцена четвертая. Смит встречает старого школьного друга.

Сцена пятая. Смит направляется к личному психоаналитику.

Что мы скажем, познакомившись с таким повествованием? Скорее всего, что ничего не поняли. При этом в отличие от нас компьютерная программа скорее всего не заметит подвоха, ведь была соблюдена связность сюжета, заключающаяся хотя бы в том, что различные действия производил один и тот же герой. Однако наше недоумение будет связано именно с тем, что не была соблюдена *смысловая* связность сюжета, а ведь именно отсутствие или наличие смысла заставляет нас говорить «понятно/непонятно». Нечто подобное случается с нами при просмотре рекламных роликов: зачастую мы не сразу понимаем, что за идея скрывается за серией визуальных образов. При этом *связь* кадров очевидна, но что-то важное ускользает.

Итак, машина скорее всего заметит разницу между первым повествованием и вторым. Усложним задачу и введем третий тип повествования – заметит ли машина разницу между первым, вторым и третьим повествованием, посчитает ли она третий тип повествования осмысленным?

П о в е с т в о в а н и е 3

Вчера стою около офиса, созерцаю природу. Вижу, как со стоянки вырывается дама на красной машине и, набирая скорость, несколько раз крестится. Церкви в прямой видимости нет. Церкви вообще рядом нет! Мне стало страшно...

Это третье повествование не содержит прямых имен-маркеров и выглядит довольно бессвязно. По крайней мере, связность этого повествования несколько отличается от повествования 2. Но человек мгновенно усматривает в этой истории забавное описание анекдотической ситуации, в то время как машина, возможно, отнесет повествование этого типа к повествованию 1.

Из этих примеров напрашивается очевидный вывод – не всякая связь есть связь осмысленная. Иными словами, обеспечить наличие связности (синтаксиса) еще не значит обеспечить наличие смысла



(семантики). Об этом говорят и противники ИИ. Однако их аргумент состоит в том, что программа не может «думать», так как не владеет значением, причем для того чтобы формализовать значение, следовало бы добиться, чтобы машина видела синий цвет так, как видим его мы. Мой же аргумент состоит в том, что за трудности надления машины интеллектом ответственны не столько значения, сколько смыслы. Сам смысл является трудноформализуемым эффектом, задать его для программы невозможно не только потому, что он несводим к синтаксису, но и потому, что мы пока в принципе не представляем условий его формализации.

Семантика обладает некой автономией продуцирования, т.е. собственной и независимой от формально-структурного уровня стихийностью порождения смыслов (Derrida, 2001). Если смысл отделяется от структуры, то становится очевидной невозможность редукционизма по отношению к сложным семантическим комплексам.

Но тогда не вполне ясно, как можно задать машине алгоритм понимания смыслов. При обсуждении перспектив создания ИИ следует принимать во внимание трудности определения смысла как образования, не совпадающего ни с референцией, ни с синтаксической линейной последовательностью повествования, в связи с чем программное моделирование смысла оказывается практически невыполнимым. Главная проблема программ, эмулирующих работу искусственного интеллекта, будет состоять в сложности фиксации «семантического окружения» конкретных значений, поскольку действие чисто синтаксических эмуляторов никак не связано не только с регистрацией значения, но в первую очередь с регистрацией смысла [Bringsjord, 1996]. Недостижимость семантической активности машинного интеллекта скорее связана с тем, что смысл не удастся свести к чему-то разложимому и символически формализуемому, вследствие чего попытки создания программного аналога семантической активности сознания серьезно осложняются.

Заключение

Итак, подытожить сказанное можно следующими выводами.

1. Благодаря *структурной* и *постструктурной интерпретациям* смысла можно с наибольшей наглядностью показать трудности *сильной* версии ИИ.

2. В силу названных трудностей формализации смысла, в том числе и для человеческого понимания, можно поставить под сомнение и *слабую* версию ИИ, согласно которой какие-то из задач, решаемых человеком, могут эмулироваться машиной.



Библиографический список

- Греймас, 1996 – *Греймас А.Ж.* Размышления об актантных моделях // Вестник Московского университета. Сер. 9. Филология. 1996. № 1.
- Косиков, 1998 – *Косиков Г.К.* От структурализма к постструктурализму (проблемы методологии). М., 1998.
- Секст Эмпирик, 1975 – *Секст Эмпирик.* Соч. В 2 т. М., 1975; 1976.
- Brachman, 2005 – *Brachman R.* Getting Back to “The Very Idea” // AI Magazine. 2005. Vol. 26, № 4.
- Bringsjord, 1996 – *Bringsjord S., Ferrucci D.* Computationalism is Dead; Now What? – Response to Fetzer’s “Minds Are Not Computers: (Most) Thought Processes Are Not Computational”, 1996.
- Bringsjord, 1999 – *Bringsjord S., Noel R.* Real Robots and the Missing Thought Experiment in the Chinese Room Dialectic, 1999.
- Cohen, 2005 – *Cohen P.R.* If Not Turing’s Test, Then What? // AI Magazine. 2005. Vol. 26, № 4.
- Deleuze, 1990 – *Deleuze G.* The Logic of Sense. Columbia University Press, 1990.
- Denett, 1998 – *Denett D.* Can Machines Think? // Brainchildren: essays on designing minds. Cambridge, 1998. P. 5.
- Derrida, 2001 – *Derrida J.* Structure, Sign and Play in the Discourse of the Human Sciences, Writing and Difference. L. : Routledge, 2001.
- Interview, 1997 – Interview with Drew McDermott // New York Times. 1997. May 14.
- Duch, 2008 – *Duch W., Oentaryo R.J., Pasquier M.* Cognitive Architectures: Where Do We Go from Here // Frontiers in Artificial Intelligence and Applications : Proc. 1st AGI Conference. 2008. Vol. 171.
- Edelstein, 1966 – *Edelstein L.* The Meaning of Stoicism. Cambridge (Mass.), 1966.
- Frege, 1948 – *Frege G.* Sense and Reference // The Philosophical Review. 1948. № 57 (3).
- Levin, 1974 – *Levin Yu.I.* On Semiotics of Truth Distortion // Informatsioimye Voprosy Semiotiki, Lingvistiki i Mashinnogo Perevoda (Information Problems in Semiotics, Linguistics, and Computer Translation). M. : VINITI, 1974. Vol. 4.
- Leont’eva, 1981 – *Leont’eva N.N.* Semantics of a Coherent Text and Units of Information Analysis // Naicli.-Tekh. Inf. Ser. 2. 1981. № 17.
- Lévi-Strauss, 1977 – *Lévi-Strauss C.* Structural Anthropology. Harmondsworth : Penguin, 1977.
- Nilsson, 2005 – *Nilsson N.J.* Reconsiderations // AI Magazine. 2005. Vol. 26, № 4.
- Russell, 1959 – *Russell B.* My Philosophical Development. N.Y. : Simon and Schuster, 1959.
- Russell, 1927 – *Russell B.* Selected Papers of Bertrand Russell. N.Y. : Modern Library, 1927.
- Saussure, 1933 – *Saussure F.* Course in General Linguistics. N.Y. : Columbia University Press, 1933.
- Searle, 1991 – *Searle J.* Minds, Brains, and Programs. The Nature of Mind ; ed. by D. Rosental. N.Y., 1991.
- Thompson, 1973 – *Thompson S.* The Types of the Folktale. Helsinki, 1973.
- Turing, 1950 – *Turing A.* Computing Machine and Intelligence // Mind. 1950. № 59.