

АВТОМАТИЗИРОВАННЫЕ МЕТОДЫ ОЦЕНКИ ЭМОЦИОНАЛЬНЫХ СУЖДЕНИЙ В ИНТЕРНЕТЕ

А.А. КОТОВ, Л.Я. ЗАЙДЕЛЬМАН, А.А. ЗИНИНА, Н.А. АРИНКИН
Национальный исследовательский центр «Курчатовский институт», Москва

Эмоциональное поведение традиционно является объектом изучения психологии. Вместе с тем общение людей в диалоге становится объектом лингвистики, а влияние эмоций на жизнь человека рассматривается в различных прикладных дисциплинах: в политологии, маркетинге, рекламе, эргономике и дизайне. Целый ряд современных прикладных исследований связан с моделированием эмоций: это системы автоматического определения эмоций текста, создание эмоциональных интерфейсов – роботов и «аватаров». В статье описана одна из прикладных систем автоматического анализа текста и перспективы ее использования для различных направлений психологии. Данная система (парсер) выделяет из текста смысл (строит семантическое представление) и позволяет исследователю анализировать множество этих смыслов, проводить поиск и выполнять количественные подсчеты. Предпринята попытка воспроизвести процессы понимания текста на естественном языке. В области психологии эта система может помочь в решении целого ряда задач: а) анализ структуры понятия, стереотипных образов и культурных представлений, б) анализ структуры категории: прототипа и периферии, в) поиск суждений с заданным смыслом, г) кластеризация и классификация выделенных смыслов с целью автоматического пополнения знаний системы об описываемом мире. В прикладных областях система применима для информационного поиска, выделения из текстов Интернета эмоциональных оценок заданного продукта (товара, торговой марки, образа политика), а также для автоматического анализа текста роботом или компьютерным агентом, поддерживающим диалог на естественном языке.

Ключевые слова: Интернет, коммуникация, эмоциональный агент, речевой интерфейс, определение тональности текста, понимание текста, семантика, синтаксис, автоматический анализ текста.

Традиционно в рамках психологии изучаются особенности психологического функционирования личности и группы. Прикладные области психологии применяют выявленные в исследованиях закономерности, например в консультировании, образовании, при разработке методик поведения или общения и др. Одним из основных современных направлений психологии по праву считается инженерная психология. Именно в этой области исследуется взаимодействие человека со сложными техническими (информационными) устройствами. Однако на сегодняшний

день вызов современных технологий для психологии является беспрецедентным. С одной стороны, этот вызов состоит в разработке интерфейсов, распознающих эмоции человека; например, это модули анализа мимики пользователя сотовым телефоном, системы, демонстрирующие эмоции в мимике и поведении: интерактивные персонажи компьютерных игр или анимированные «смайлики» в сотовых телефонах и мессенджерах. С другой стороны, это системы взаимодействия с пользователем с помощью текстов: создание профиля пользователя на основе его поведения в Сети, адаптация поиска документов под конкретного пользователя, генерация текстов «по требованию», например, автоматический перевод теста пользователя. На такую конкуренцию

Исследование выполнено при финансовой поддержке Министерства науки и высшего образования Российской Федерации, соглашение № 14.601.21.0017 (УИР RFMEF160117X0017).

подходов можно посмотреть и с противоположной стороны, если представить, что компьютерные методы могут стать для психологии инструментом изучения эмоций, процессов мышления и понимания текста (Величковский, 2006). Компьютерные методы могут использоваться в психологии для фиксации материала (создания корпусов), анализа эмоций в текстах Интернета, а также для компьютерного моделирования естественных когнитивных функций человека.

Прежде всего, компьютерные технологии создали крупные хранилища эмоциональных текстов и видеозаписей эмоционального поведения. Эмоциональные жесты и мимика плохо приспособлены для регистрации: хотя киноленты и телепередачи представляют собой естественные хранилища эмоциональной экспрессии человека, только с развитием компьютерных технологий этот материал стало удобно организовывать в корпуса и сопровождать разметкой. Что же касается текстов, то здесь мы получили уникальное хранилище данных – Интернет. Личное общение обычно происходит через частные закрытые каналы: системы обмена текстовыми сообщениями – SMS, Telegram, WhatsApp. Общественно значимые темы, как правило, обсуждаются на дискуссионных площадках: на форумах (message-boards) или в блогах. Автор блога стремится привлечь больше читателей, получить общественное одобрение и, как следствие, доходы от рекламы. Это заставляет его открывать блог для всех желающих, обсуждения становятся свободно доступными в Сети. У исследователя даже отпадает необходимость фиксировать материал: по сути, содержание веб-страниц уже является корпусом, готовым для анализа¹. Анализ такого ма-

териала открывает огромные возможности для изучения коллективных эмоций и общественных настроений, для исследования репрезентации значимых событий и явлений в общественном сознании. На основе материалов, доступных в Сети, становится возможным выявление значимых социальных угроз, а также перспективных направлений их преодоления. Путем автоматического извлечения фактов из большого массива информации также можно изучать обыденные представления о какой-либо предметной области.

Глобальная задача, на которой мы остановимся подробнее, состоит в разработке методов автоматического понимания текста и в создании роботов, которые реагировали бы на входящие тексты подобно человеку. Вопросы о том, что является «сутью процесса понимания» и возможно ли понимание текста компьютером, вызывают большие споры в современной науке². В лингвистике *понимание* означает конструирование смысла текста, его семантического представления³. Описанная ниже компьютерная модель решает задачу конструирования смысла текста для аналитических целей и для управления роботом. Система предоставляет доступ ко всем промежуточным репрезентациям текста и помогает анализировать большие массивы текстов: изучать структуру понятий и типовые суждения, включая эмоциональные высказывания.

гвистической разметкой (обзор см.: Резникова, Коптев, 2005).

² Прежде всего это дискуссия вокруг аргумента «китайской комнаты» Дж. Сирла (Сирл, 2003).

³ Со стороны наблюдателя «понимание» выглядит как релевантная реакция человека на услышанное. Поэтому семантическое представление должно быть достаточным, чтобы позволять автоматической системе выполнять с текстом все те операции, которые человек делает с «понятным» текстом: отвечать на вопросы по тексту, делать выводы или извлекать мораль, отвечать на эмоцию текста, выбирать ответ – тот ответ, который мог бы предложить человек, услышав текст от собеседника.

¹ Содержание сайтов форумов и блогов все же изменчиво, для надежной фиксации такого материала создаются лингвистические корпуса, которые дублируют тексты блогов и сопровождают их лин-

СУЩЕСТВУЮЩИЕ МЕТОДЫ АНАЛИЗА

Методы компьютерной лингвистики позволяют извлекать из текстов, написанных на естественном языке, информацию о мнениях и эмоциях авторов сообщений – эти методы разрабатываются в областях контент-анализа или анализа тональности (sentiment analysis) текста. На эмоциональность текста влияет множество факторов: эмоциональная и оценочная лексика, пунктуация, смайлики и эмодзи эмодзи (язык пиктограмм, используемый в электронных сообщениях и на веб-страницах). Кроме того, воздействие оказывают личность автора, общая тема и контекст. Даже для человека определение эмоции текста не всегда очевидно по нескольким причинам: а) авторы устных текстов и блогов стремятся уйти от слов со стертым значением и выразить эмоции через сленговые выражения и неологизмы: *кривое объяснение* вместо *плохое объяснение* (Карпова и др., 2011); б) текст может выражать эмоции неявно, имплицитно: с помощью иронии, сарказма, намеков, недосказанности и неоднозначности (Грайс, 1985; Санников, 1999); в) при рассказе историй автор стремится реализовать сюжетные отношения – перейти от негативной ситуации к позитивной, отразить противоположную точку зрения, чтобы рассказ не звучал прямолинейно (Пропп, 1928; Schank, 2000). Вместе с тем многие тексты стремятся выразить эмоциональную оценку прямо, чтобы воздействовать на адресата: прежде всего это относится к рекламе и публицистике. Пользователи Сети часто однозначно выражают свои эмоции в отзывах на кинофильмы и книги, на рестораны и выставки, они сообщают о своих политических взглядах, ругают и хвалят продавцов магазинов и работников банков. Такие посты обычно и становятся объектами исследований по определению тональности: компаниям интересно и важно понимать, что о них говорят и думают.

Часто для оценки эмоциональности текста используют списки эмотивной лексики, т.е. перечни слов и словосочетаний с приписанным им уровнем эмоциональности. В простейшем случае этот уровень выражается бинарно: слово несет в себе или отрицательную, или положительную оценку: *ненавидеть* – отрицательное, *обожаю* – положительное. Шкала оценки может быть не бинарной, а более дифференцированной: в работе М.В. Клековкиной и Е.В. Котельникова (2012) авторы используют словарь из 60 эмоциональных слов, каждое из которых оценено по шкале от -5 до -1 для отрицательных слов и от $+1$ до $+5$ – для положительных. Кроме оценочных слов авторы используют словомодификаторы, которые могут как увеличить, так и уменьшить эмоциональность стоящего рядом оценочного слова (*хорошо* имеет меньшую оценку, чем *очень хорошо* или *действительно хорошо*, и большую оценку по сравнению с *не хорошо*). Эмоциональность всего текста складывается из входящих в этот текст оценочных слов и их модификаторов. Для исследований в области оценки эмоциональности текстов необязательно составлять свой словарь эмотивной лексики с нуля, существуют большие готовые словари. Одним из них является словарь EmoLex (Mohammad, Turney, 2010); в нем содержатся 14 тыс. слов, которым приписаны тональность (положительная или отрицательная) и конкретные эмоции (злость, отвращение, радость и пр.).

Существует целый ряд методов, помогающих в создании словарей эмотивной лексики. Во-первых, можно оттолкнуться от небольшого списка эмоциональных слов (seed words) и распространить эмоциональную оценку каждого слова на его синонимы и на слова, встречающиеся в похожих с ним контекстах; для этого можно использовать толковые словари, словари синонимов, тезаурусы и корпуса текстов. Этим способом пользуются, например,

в работе (Kim, Novy, 2004): из тезауруса WordNet (Miller et al., 1990) авторы получали синонимы и антонимы для исходных слов, ставили им в соответствие положительную или отрицательную оценку (в зависимости от оценки исходного слова) и расширяли таким образом свой словарь эмотивной лексики. Во-вторых, можно оттолкнуться от общей эмоциональной разметки текста (это разметка, проведенная экспертом, или число «звездочек», которые сам автор поставил фильму рядом с текстом своей рецензии): можно разметить все слова, которые преимущественно характерны для негативных текстов как негативные, и аналогичным образом поступить с позитивными текстами. Таким методом формирования словаря пользуются Е.В. Тутубалина и соавт. (2015), которые сформировали два корпуса текстов: негативный, куда были помещены фрагменты из отзывов, озаглавленные как «Недостатки», и позитивный, озаглавленный как «Преимущества» (платформа otzovik.com, использовавшаяся в данной работе, предлагает пользователям оставлять отзывы определенной структуры на любой товар или услугу). Недостатки учитывались только из крайне негативных отзывов (с оценкой 1 или 2), а преимущества — только из хвалебных (с оценкой 5). В каждом из корпусов внутри каждой предметной области были выделены наиболее частотные слова разных частей речи — как наиболее ярко характеризующие данную сферу с позитивной или негативной стороны.

Система LIWC (Linguistic Inquiry and Word Count), созданная психологами Техасского университета в Остине (Tausczik, Pennebaker, 2010), читает текст и с помощью словаря, в котором все слова отнесены к одной из 80 категорий, подсчитывает, сколько слов из этого текста отражают различные эмоции, стили мышления, социальные проблемы. Слова в словаре разнесены по категориям таким образом, чтобы по тексту определить социальное и психо-

логическое состояние его автора. Авторы системы признаются, что LIWC не умеет распознавать сарказм, иронию и идиомы. Однако предложенный подход интересен тем, что при определении эмоциональной оценки текста учитываются не только эмотивные слова, в нем содержащиеся. Каждому слову из текста, включая местоимения и артикли, ставятся в соответствие категории (одна или несколько): языковые (глаголы, артикли и пр.) и оценочные (грусть, негативные эмоции и пр.).

В простом случае эмоциональность текста рассчитывается как среднее значение всей эмоциональной лексики текста. Более сложные подходы используют эмоциональные словари совместно с лингвистическим анализом. Так, в системе анализа А.Г. Пазельской и А.Н. Соловьева (2011) производится синтаксический анализ: в предложении выделяются субъект, объект, предикат (глагол) и зависимые обороты, что позволяет определить объект оценки на основании его словарных характеристик и положения в синтаксической структуре, а также рассчитать его эмоциональную оценку.

Одним из самых современных подходов к решению задачи оценки тональности является подход с применением нейронных сетей. Особенность этого подхода состоит в том, чтобы автоматически настроить алгоритм-классификатор, который для каждого входящего текста возвращает его тональность⁴. Во время «обучения» на вход классификатора подаются тексты с известной тональностью (например, размеченной

⁴ Нейросети основываются на коннекционистском подходе к обработке информации, предложенном в рамках психологии Дж. Макклелландом и Д. Румельхартом (McClelland, Rumelhart, 1981). В рамках этого подхода предполагается, что обработка информации производится сетью из множества узлов со связями между ними. Обучение сети состоит в том, чтобы автоматически настроить веса на связях, так, чтобы сеть с минимальным числом ошибок классифицировала элементы обучающей выборки.

экспертом). При неправильном определении тональности классификатор корректируется (обучается) так, чтобы для всего множества текстов снизить число ошибочных оценок. После такого «обучения» сеть может оценить новый текст, похожий по своей структуре на тексты из обучающей выборки. Данный метод избавляет разработчиков от ручного написания правил и словарей, но требует сбора и экспертной оценки огромной коллекции текстов для обучения сети. Р. Сочер и соавт. (Socher et al., 2013) использовали нейросеть для определения тональности текстов отзывов пользователей на кинофильмы. Для формирования обучающей выборки брались 12 тыс. размеченных по тональности предложений из реальных отзывов, для каждого предложения было сформировано его синтаксическое представление в виде набора вкладывающихся друг в друга фрагментов — *составляющих*. Например, предложение *Я посмотрел скучный фильм* можно разделить на составляющие так: [[Я] [[посмотрел] [[скучный] [фильм]]]]. Все такие составляющие также получили экспертную оценку по шкале от крайне негативной до очень позитивной (—, —, 0, +, ++). Таким образом, у нейросетевого алгоритма для обучения была информация об эмоциональной оценке 12 тыс. предложений, включая оценки всех их составляющих. Такой подход, по утверждению авторов, позволяет правильно оценить эмоциональный настрой различных по сложности предложений, включая предложения с разными видами отрицаний и противопоставлений.

Неудобство нейросетевых алгоритмов состоит в том, что они формируют репрезентацию текста в виде массивов чисел, а эти данные непригодны для прямой интерпретации человеком. Для исследовательских целей может быть полезна система, которая строит репрезентации текста (например, семантическое представление) в таком виде, который понятен специалисту.

ПРЕДЛАГАЕМЫЙ МЕТОД АНАЛИЗА

Картины проявлений эмоций в текстах исключительно разнообразны: мировая литература дает нам множество примеров сложно устроенных эмоциональных текстов. Вместе с тем простые тексты — публицистика, реклама и бытовые диалоги — специально ориентированы на то, чтобы передать адресату конкретную эмоцию, поэтому эмоциональные высказывания в них обладают более строгой структурой. Мы опираемся на предложенную нами ранее модель эмоциональных высказываний (Котов, 2003). Согласно этому подходу, под воздействием эмоции человек строит эмоциональные высказывания типа *Я никому не нужен!*, а при попытке оказать воздействие на другого человека — высказывания сходной структуры: *Ты никому не нужен!* В позитивных контекстах при «самопрезентации» человек может говорить: *На меня все обращают внимание!*, при выражении восхищения: *На него все обращают внимание!*, а в ситуации комплимента: *На тебя все будут смотреть!* Таким образом, в разных эмоциональных диалогах высказывания обладают сходной синтаксической схемой и сходным смыслом — семантикой. Отличия состоят в том, как роли этого высказывания (семантические валентности) распределены между участниками коммуникации: «На меня обращают внимание» или «На тебя обращают внимание». Многие подходы к анализу эмоций текста подчеркивают роль речевой фонетики или графики. К примеру, в стихотворном тексте эмоциональное воздействие может обеспечиваться стихотворным размером: ритмом и рифмой. Целый ряд риторических фигур связан с повтором звуков (аллитерация, ассонанс) или с изменением букв в слове (метатеза, анаграмма). Попытки смоделировать воздействие этого уровня текста делаются в целом ряде работ по фоносемантике (Журавлев, 1974). Мы сконцентрируемся на более традиционном

представлении о том, что важнейшим уровнем передачи смысла (и эмоций) в лингвистической модели является семантика. Чтобы соотносить смысл высказывания с теми или иными эмоциями, нужно автоматизировать конструирование семантического представления, а также описать структуру эмоциогенных высказываний — выделить те смыслы, которые выражают эмоции и регулярно выступают в качестве стимулов для эмоций.

Высказывание *На тебя все смотрят!* достаточно эмоционально. В негативном контексте оно может указывать на неадекватность внешнего вида адресата и будет преследовать цель застыдить его: *Что ты сделал? На тебя все смотрят! Тебе должно быть стыдно!* В позитивных контекстах оно может представлять собой комплимент адресату: *Ты сшила потрясающее платье! На тебя все смотрят!* С точки зрения лингвистической теории, на эмоциональный характер этого высказывания указывает как общая семантическая структура ('X обращает внимание на Y'), так и конкретное преувеличение в местоимении *все*. Есть целый ряд признаков, изменение которых влияет на эмоциональность высказывания. Так, фраза *Куда ты засунул мой рюкзак?* отличается от высказывания *Куда ты положил мой рюкзак?* признаком интенсивности глагола: *засунул* (≈ «разместил интенсивно») vs. *положил* (≈ «разместил»). Признак «интенсивно» может описывать ситуацию реального мира, если адресат действительно положил куда-то рюкзак «с силой», но часто этот признак обозначает именно эмоцию: говорящий недоволен и поэтому преувеличивает интенсивность обозначаемого действия (Апресян, 2003; Гловинская, 2004). Можно сказать, что здесь проявляется известное в психологии явление активизации эмоции «сверху-вниз» (top-to-bottom), когда возникающая эмоция меняет репрезентацию исходной ситуации (Clore, Ortony, 2000). Реальная ситуация «положить рюкзак»

представляется говорящему как «интенсивно положить рюкзак»: это отличие может быть вызвано эмоцией говорящего, оно проявляется в тексте и позволяет слушающему распознать исходную эмоцию. С учетом таких ключевых признаков для распознавания позитивной эмоциональной тональности в высказывании *На тебя все смотрят!* можно создать шаблон, который указывает, что:

- глагол должен включать смысл «обращать внимание»;
- подлежащее (агенса⁵) должно обозначать людей, число которых может быть преувеличено;
- прямое дополнение (пациенса) должно обозначать человека, испытывающего эмоцию: это говорящий при самопрезентации⁶ или адресат в случае комплимента или воздействия (рекламы)⁷.

Автоматический анализ должен создавать для каждого высказывания такое семантическое представление, которое можно сравнить с имеющимися шаблонами: например, чтобы семантика высказывания *На тебя все смотрят!* прямо совпадала с приведенным шаблоном. Вот как эта задача решается в разрабатываемой нами системе автоматического анализа текста — парсере. В соответствии с уровневой моделью языка (Мельчук, 1999), парсер обрабатывает текст на морфологическом и синтаксическом уровнях, а также строит семантическое представление (смысл) текста. Смысл передается в компонент сце-

⁵ Мы используем инвентарь валентностей или семантических ролей на основе типологии (Fillmore, 1968).

⁶ Говорящий сам испытывает эмоцию, что подталкивает его к суждениям типа *Я отлично выгляжу! На меня все смотрят!*

⁷ Говорящий в этом случае намерен вызвать эмоцию адресата с помощью суждений вида *Если вы купите эту машину, вы всегда будете в центре внимания!* Адресат рекламы редко поддается речевому воздействию, однако высказывание сконструировано именно таким образом, чтобы вызывать эмоцию адресата.

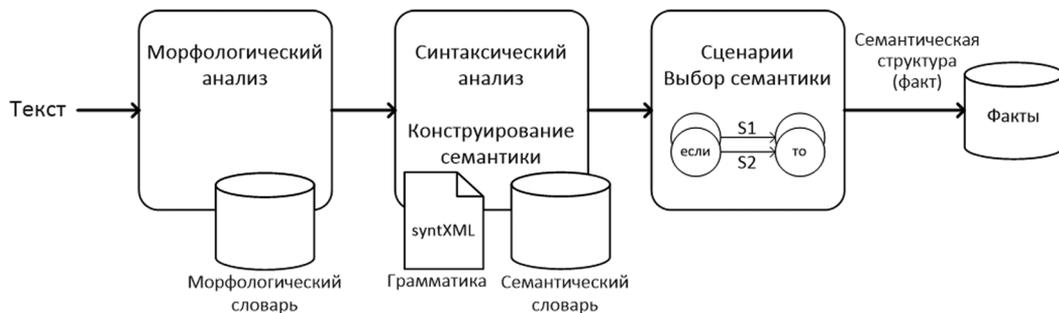


Рис. 1. Общая структура системы автоматической обработки текста – парсера

нариев, где разрешается многозначность: удаляются альтернативные семантические представления, если они были построены ранее. Если парсер работает в составе робота, поддерживающего диалог (Зинина и др., 2018), то сценарии могут выбрать коммуникативный ответ робота: пакет, состоящий из жестов, мимики и речи, описанный на языке BML – Behavior Markup Language (Kopp et al., 2006; Vilhjálmsson et al., 2007). Если же парсер действует отдельно от робота, то он сохранит семантическое представление обработанных текстов в базу данных для последующего анализа исследователем (рис. 1).

Парсер анализирует входящий текст по предложениям, а каждое предложение – по словам. Он может принимать на вход текстовые файлы, например художественную литературу, а также может получать тексты текущих новостей и блогов через систему RSS. Если парсер управляет роботом, то он может получать звуковые сообщения от пользователя и преобразовывать их в текст с помощью стороннего сервиса распознавания речи. В морфологическом компоненте парсер определяет часть речи и грамматические характеристики каждого слова. В синтаксическом компоненте парсер устанавливает синтаксические связи между словами текста; при этом используется описание грамматики русского языка на языке syntXML (Kotov et al., 2017). Итоговая задача этого компонента в том,

чтобы построить синтаксическое дерево, т.е. объединить синтаксическими связями все слова предложения *На тебя все смотрят!* в древесный граф (рис. 2). Глагол, как правило, является синтаксической вершиной предложения, поэтому слово *смотрят* занимает позицию вершины в синтаксическом дереве. Глагол обладает валентностями, с помощью которых он может присоединить другие слова предложения: местоимение *все* и предложную группу *на тебя*. Если какая-то словоформа разбрана несколькими способами или если синтаксические связи могут быть проведены по-разному, возникает омонимия⁸. Например, словоформа *все* получает второй разбор как форма множественного числа слова *весь*, но этот разбор не сможет объединиться с другими словами в общее синтаксическое дерево, поэтому он отбрасывается⁹. Также словоформа *все* получает еще один разбор как форма наречия *всё*, этот разбор дает другое, омонимичное

⁸ Относительно анализа типов омонимии см.: Гладкий, 1985.

⁹ Слово *все* является местоимением-существительным – оно действует как существительное: *Люди смотрят. Все смотрят*. Слово *весь* является местоимением-прилагательным – оно действует как прилагательное: *Внимательные люди смотрят. Все люди смотрят*. Эти два слова рассматриваются как разные лексемы. Словоформа *все* может относиться к каждой из этих лексем, т.е. эта словоформа омонимична.

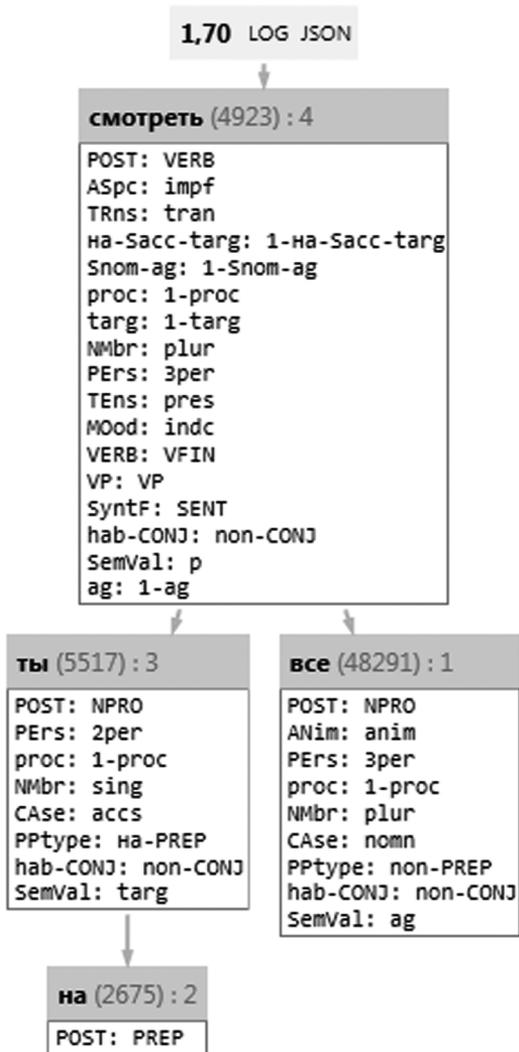


Рис. 2. Синтаксическое дерево предложения
Все на тебя смотрят

синтаксическое дерево *Все смотрят на тебя!* В результате анализа на выход из синтаксического компонента поступят оба дерева, но второе дерево далее проиграет первому при анализе смысла.

При семантической обработке для каждого предложения строится его семантическое представление: после разрешения омонимии одно из них будет использовано для выбора роботом коммуникативной реакции в компоненте *сценариев*.

Мы считаем, что эти процессы связаны: если робот находится в негативном состоянии, то под действием эмоционального процесса «сверху-вниз» он выберет более «негативное» и соответствующее его эмоции семантическое представление. В конструкции парсера мы воспроизводим этот естественный эффект: сценарий, обладающий большей активизацией, способствует выбору наиболее близкого к нему семантического представления. Если парсер управляет роботом, имитирующим негативное состояние, то высказывание *Все на тебя смотрят* парсер будет интерпретировать как «против тебя что-то замышляют»¹⁰ (Kotov, Zinina, Filatov, 2015).

Семантическое представление — это множество признаков, распределенных по семантическим валентностям. Мы используем инвентарь из 606 признаков на основе набора базовых понятий (Вежицкая, 1999) и семантического словаря (Шведова, 1998). Семантические признаки приписаны слову. Слово может обладать омонимией, в этом случае признаки в его значении разделены на группы. В нашем словаре каждому значению слова может быть приписано несколько признаков: некоторые из них отражают семантический класс слова, а другие могут быть важны для эмоциональной обработки. Например, глагол *смотреть* будет относиться к семантическим полям ментальных предикатов («*пм*») и обладать признаками *пм-думать*, *пм-видеть*, *пм-внимание* («обращать внимание»), а также периферийным признаком *п-бездействовать*, проявляющимся в контекстах типа *Чего ты смотришь?* [и ничего не делаешь]¹¹. При конструировании семантического представления признаки слова копируются в ту валентность, которую слово заняло в синтаксической

¹⁰ Данная интерпретация соответствует сценарию ПЛАНИР, см. Приложение.

¹¹ Такие контексты обрабатываются сценарием БЕЗДЕЙСТВ, см. Приложение.

p : 4	ag : 1	targ : 2
1 1 3 пм-думать (31)	1 1 9 ч-некто (1)	1 1 3 ч-некто (1)
1 1 3 пм-видеть (35)	1 1 3 чс-много (6)	1 1 1 ч-эгоцентрик (10)
1 1 9 п-бездействовать (160)	1 1 1 чс-все (74)	1 1 3 о-об (14)
1 1 3 пм-внимание (187)		
1 1 3 п-направлять-внимание (542)		

Рис. 3. Семантическое представление высказывания *Все на тебя смотрят*. Выделились валентности *p* (предикат), *ag* (агенса) и *targ* (конечная точка)

структуре. Мы используем инвентарь из 23 валентностей на основе работы (Fillmore, 1968): глагол занимает валентность *предикат*, подлежащее — валентность *агенса*, дополнение — валентность *пациенса* и т.д. Семантическую структуру простого предложения можно представить как двухуровневое дерево, где валентность *предикат* находится в вершине и подчиняет другие валентности, но для компактности эту структуру также можно записать в виде таблицы (рис. 3).

Как видно на рисунке, семантика слова *все* представлена признаками *чс-все*, *чс-много* и *ч-некто*¹² (человек или другой социальный субъект), а семантика слова *ты* представлена признаками *о-об* (физический объект), *ч-некто* и *ч-эгоцентрик* (воспринимающий субъект). Конечно, такое семантическое представление достаточно ограничено, но мы стремимся сделать его достаточным, чтобы смоделировать ключевую особенность понимания — выбор релевантной реакции.

Компонент сценариев описывает возможные реакции робота на входящий смысл. В отсутствие робота этот компонент выбирает из омонимичных вариантов то понимание (семантическое представление), которое будет сохранено в базу данных. Единицы этого компонента — *сценарии* — определяются как процедуры типа *если-то* и служат моделью для умо-

заклучений и эмоций. В работе парсера сценарии распознают эмоциогенные смыслы: сценарий активизируется, если смысл входящего текста близок к посылке (условию «если») данного сценария. Несколько сценариев могут одновременно активизироваться и конкурировать за обработку входящего смысла¹³. В работах А.А. Котова (2003, 2012) предложен инвентарь из 34 сценариев, ответственных за распознавание в текстах негативных и позитивных эмоциональных смыслов. К примеру, смысл на рис. 3 приводит к активизации сценариев ВНИМАНИЕ, ПЛАНИР, а также ОБМАН и БЕЗДЕЙСТВ (см. Приложение). Активизация сценария ВНИМАНИЕ будет означать, что парсер классифицирует входящий смысл как позитивное событие «приятно, что на меня обращают внимание». Если парсер используется в составе робота, то активизация сценария будет приводить к выполнению на работе определенного поведенческого паттерна (Зинина и др., 2018). В отсутствие робота этот сценарий будет выбирать ближайшее к нему понимание текста и сохранять его в базу данных.

Сценарии играют важную роль при разрешении многозначности (омонимии). Если при анализе предложения получено несколько синтаксических деревьев, то выбирается такое дерево, семантика которого наиболее близка к существу-

¹² Префикс «чс-» обозначает признаки числа, префикс «ч-» — признаки человека, а префикс «о-» — признаки физических объектов, включая живых существ.

¹³ В работе А.А. Котова (Kotov, 2017) высказано предположение, что конкуренция сценариев может служить ключевым элементом для моделирования сознания.

ющим сценариям. Это означает, что для анализируемых текстов используется то понимание, которое более регулярно (соответствует известным ситуациям, закрепленным в *рациональных* сценариях) или более эмоционально значимо (соответствует таким сценариям, которые отвечают за моделирование эмоций). К примеру, высказывание *Кошка пьет молоко* обладает омонимией: слово *кошка* может означать «металлические зубья, прикрепляемые к обуви», а поскольку в этом значении слово *кошка* неодушевленное, парсер будет строить отдельное синтаксическое дерево. Смысл «живое существо потребляет еду» близок к сценарию ВКУС: он правдоподобен в той картине мира, которая фиксируется сценариями. Напротив, смысл омонимичного дерева «инструмент потребляет еду» относительно далек от существующих сценариев: он менее правдоподобен с точки зрения сценариев. Сходные смыслы появляется только в виде метафор *Банкомат съел карточку*¹⁴ или *Пылесос съел носок*. Таким образом, при сравнении со сценариями будет выбираться первое дерево как обладающее более регулярным и эмоциональным смыслом. Второе дерево будет отбрасываться.

РАБОТА С РЕЗУЛЬТАТАМИ АНАЛИЗА ТЕКСТА

Результаты работы парсера сохраняются в базу данных¹⁵. Структура базы данных спланирована таким образом, чтобы хранить результаты работы каждого компонента парсера. При запуске парсера можно отдельно подключить сохранение в базу результатов морфологического анализа (варианты морфологических разборов каждого слова), синтаксических структур (как на

рис. 2) или семантического представления (как на рис. 3). И если морфологические и синтаксические данные интересны только для лингвистического анализа, то семантика текста оказывается важна именно для поиска в пространстве смысла. На основе семантических представлений множества текстов можно анализировать участие отдельного понятия в различных ситуациях, структуру категории (как элемента ситуации), а также выполнять поиск фактов, соответствующих определенному семантическому шаблону.

Анализ структуры понятия. Традиционно в психологии для исследования понятий людей просят дать определение слову. А.Р. Лурия (1979) разделяет два типа ответов на такие вопросы. В первом случае в определениях отражаются функции объекта, конкретные ситуации, в которые включен описываемый объект: *собака лает, собака охраняет дом*. Во втором случае понятие описывается путем отнесения его к определенной категории (гиперониму): *собака — это домашнее животное*. В лингвистике считается, что эти два подхода тесно связаны. Понятие может включаться в различные ситуации: «человек» может «говорить», «голосовать», «иметь избыточный вес», «выдыхать углекислый газ» и т.д.; при этом участие понятия в ситуации меняет его структуру, подчеркивая или «профилируя» те признаки, которые релевантны для конкретной ситуации (Wierzbicka, 1972; Langacker, 1991, p. 191). Например, *банк* может участвовать в ситуации «переводить деньги» и относиться при этом к классу «финансовых организаций» или участвовать в ситуации «находиться на Садовом кольце» и в этом случае — относиться к классу «зданий». В работе У. Йе, Л.У. Барсалу эти закономерности сформулированы в виде набора тезисов, регулирующих функционирование понятий в разных контекстах. Например: «Когда понятие обрабатывается в конкретной ситуации, выделяются признаки, релевантные

¹⁴ При активизации сценария ВКУС на *банкомат* переносится признак «живое существо», а на *карточку* — признак «еда». Таким образом, обработка текста «сверху—вниз» может помочь при компьютерном моделировании метафорического переноса.

¹⁵ Используется база данных PostgreSQL.

Таблица 1

Контексты слова *снег*, разделенные по семантическим валентностям

Валентность (число случаев)	Пример употребления
Инструмент (68)	Дерево полозьев облепило <i>снегом</i> .
Агенс (518)	Сухой <i>снег</i> припорошил коричневые борозды.
Пациенс (218)	Там <i>снег</i> сдуло под гребнем!
Траектория (38)	На подъемах ящик соскакивал с саней и шел по <i>снегу</i> рядом.
Целевая точка (98)	Фонарь попал в <i>снег</i> , стекло лопнуло, и свет погас.
Содержание восприятия или мышления (52)	Ташу и вижу: черный <i>снег</i> такой.
Место (125)	За пригорком из отвесной скалы среди <i>снега</i> бил ключ.
Интерпретация (63)	А тут еще и мелкий дождь превратился в первый <i>снег</i> .
Причина (38)	Кругом все белело от <i>снега</i> .
Время (27)	После <i>снега</i> во сне он твердо знал, что наутро все сладится.
Параметр или область (29)	Нам повезло со <i>снегом</i> .
<i>Примечание.</i> Приведены избранные валентности. Общее число контекстов – 2037.	

для этой ситуации» и «Когда обрабатываются конкретные признаки понятия, они активизируют соответствующую ситуацию» (Yeh, Barsalou, 2006, p. 354). Метод автоматического анализа текста позволяет быстро выбрать из текстов ситуации, в которых участвует некоторое понятие. Ситуации могут быть классифицированы на основании того, какую валентность занимает исследуемое понятие. Пример выдачи для слова *снег* приведен в табл. 1.

Пример показывает, что *снег* может описываться как активный деятель (агенс) в контекстах типа *снег припорошил*, как объект действий (пациенс), как поверхность или трехмерное пространство, где происходит действие, а также как время, «когда идет снег». Такой подход позволяет лингвистическими методами приоткрыть структуру понятия для дальнейшего изучения психологическими методами. Данный метод также позволяет исследовать восприятие некоторого понятия для целей социальной психологии; например, для оценки ситуаций, негативно воспринимаемых обществом, мы изучали контексты понятий *пенсионер*, *регион*, *враг* и т.п. Это позволило выделить классы эмоциональных суждений, связанных у авторов

текстов с этими понятиями, и проанализировать источники общественной неудовлетворенности (см. ниже).

Анализ структуры категории. Категория может рассматриваться как множество понятий, замещающих определенную валентность в ситуации, или как общее в значении всех этих понятий – их интегральный признак (Барулин, 1996, с. 39). Например, категория «напиток» может быть определена как «то, что обычно пьют», т.е. как множество пациенсов в ситуациях, где глагол включает смысл «пить». С помощью запроса к базе данных можно выбрать все контексты с такими предикатами и далее собрать статистику по заместителям валентности пациенса (табл. 2). Интересно, что в выдачу попадают различные глаголы: *пить*, *допить*, *выпить* – поиск производится по семантике слова, и результат не привязан к конкретному глаголу, как в традиционных поисковых алгоритмах.

Метод позволяет исследовать связь между понятием и категорией, т.е. предлагает альтернативу экспериментальному подходу, с помощью которого структура категорий была исследована в классических работах Э. Рош (Rosch et al., 1976).

Структура категории «напиток»

Пациент (число случаев)	Пример употребления
Чай (226)	Он допил <i>чай</i> , отставил пиалу и начал собираться.
Вода (126)	Начальники запрещали им пить <i>воду</i> сразу и помногу.
Пиво (93)	Как это быть грузчиком и не пить <i>пива</i> !
Что (90)	<i>Что</i> будем пить?
Вино (88)	В ресторане сидели долго, пили <i>вино</i> , шутили, смеялись, разговаривали без конца
Водка (65)	<i>Водки</i> вы уже выпил.
Кофе (57)	Утром, в столовой, они пьют водянистый <i>кофе</i> , жуют черствые хлебцы.
Кровь (47)	Ты не будешь нуждаться ни в пище, ни в воде, но тебе придется пить <i>кровь</i> .
Алкоголь (24)	Богатые люди пьют дорогой <i>алкоголь</i> и меньше рискуют нарваться на подделку.
<i>Примечание.</i> Приведены наиболее частотные примеры.	

В получаемых данных видны как прототипы категорий — понятия, лучше всего репрезентирующие категорию (*чай, вода, пиво*), так и структура категории на периферии: напитки для экзотических ситуаций и воображаемых миров (*кровь*), случаи метонимии или регулярной полисемии (*чашка, бутылка*)¹⁶, метафоры или омонимичные значения глагола (*выпить таблетку*). Данный аспект метода описан в работе (Зайдельман и др., 2019). Автоматический метод также применим к социолингвистическому анализу понятия: типичные контексты использования понятия могут служить базой для выводов эксперта об отношении к соответствующему объекту, его «социальном образе» или коннотациях. К примеру, *пенсионеров* в современных новостях и блогах чаще всего *кормят* или *задерживают*¹⁷, например:

(1) *чтобы не обидеть старика, хозяин сказал, что у них акция и пенсионеров кормят бесплатно.*

Таким образом, данный метод позволяет не только делать теоретические выводы о структуре понятий, но и изучать социолинг-

вистический образ, стоящий за понятием, выделять регулярные проблемные или позитивные ситуации, связанные с понятием.

Поиск семантического шаблона. Основной репрезентацией текста, полученной в ходе семантической обработки, является семантическое представление. Это представление позволяет классифицировать факты, выбирать наиболее релевантную реакцию на данный текст (если парсер действует в составе робота), а также искать факты заданной структуры. Для этого в запросе требуется задать все признаки, которые должны присутствовать у искомого факта. Шаблон поискового запроса будет похож на семантическую структуру (рис. 3) — это также будет множество признаков, распределенных по семантическим валентностям. Например, можно выполнить поиск описанных в тексте насильственных действий, точнее, тех случаев, где агент-человек причиняет вред пациенту-человеку. Агент и пациент в искомой ситуации должны содержать признак *ч-некто* («человек»), а предикат — признак *п-кауз-вред* («предикат: каузировать¹⁸ вред»). Результат достаточно

¹⁶ Обозначение объема вещества (чай) по названию контейнера (чашка) рассматривается как метонимия или регулярная полисемия (Апресян, 1995).

¹⁷ Анализировались наиболее частые предикаты, принимающие слово *пенсионер* в качестве пациен-

¹⁸ В лингвистической семантике для описания причинно-следственных связей используется мета-языковая глагол *каузировать* (Мельчук, 1999).

Блок примеров. Контексты, в которых наносят вред гостям

В игре кубка первого канала россияне победили гостей со счетом 2:0.

Хозяева льда разгромили гостей со счетом 5:1, преврав их восьмиматчевую победную серию.

Армейцы победили гостей со счетом 2:1.

По данным следствия, из пистолета предприниматель методично расстрелял всех своих гостей, не забыв сделать контрольные выстрелы в голову.

В Москве бармен во время драки убил гостя на открытии паба.

Во Флориде 71-летний домовладелец застрелил незваного гостя, который случайно зашел через незапертую дверь, сообщает портал gainsville.com.

Мы гостей страны не обижаем!

В Косихинском районе Алтайского края мужчина зарезал гостя, который отказался покидать его дом.

Стоит отдать должное хозяевам, они не сразу шокируют гостя.

объемом и содержит 3550 случаев, поэтому он может быть сокращен или обобщен по различным критериям: в блоке примеров мы приводим только такие случаи, где вред причиняется *гостям*.

Даже из этого сокращенного множества примеров видно, что ситуации насилия дифференцируются по степени вреда: в выдачу попадают как предикаты *шокировать* и *обижать*¹⁹, так и предикаты прямого обозначения насилия: *убивать*, *расстрелять* и *зарезать*. Такие ситуации могут быть разделены с помощью уточненного запроса. Кроме того, ситуации подразделяются на насилие в реальном мире и на «разгром врага» в мире спорта — это явление обычно описывается как перенос на область спорта метафоры войны (Lakoff, Johnson, 2003).

Другим методом поиска ситуаций в семантике текста является сравнение со сценариями. Сценарии используются для разрешения омонимии на выходе из синтаксического компонента, соответственно, каждый факт в базе уже отобран как обладающий близостью с одним из сценариев. Большинство контекстов насилия

перехватываются сценарием ОПАСН, который отвечает за распознавание эмоциональной ситуации «кто-то причиняет мне вред». Для этого сценария важно, чтобы адресат идентифицировал себя с валентностью пациенса или хотя бы сочувствовал пациенсу, считал его «своим», протагонистом. Имеется симметричный сценарий МЫ • ОПАСН, который должен перехватывать ситуации, где сам адресат (протагонист) наносит вред «врагам» и получает удовлетворение от этих действий. Этот сценарий часто активизируется примерами из литературных диалогов, где агенс назван местоимением *ты*²⁰:

(2) Ты хорошо стреляешь.

(3) Ты убил фотографа?

(4) Напугала ты нас.

Как видно, сценарий активизируется ситуацией нанесения вреда, даже если это описание содержится в вопросе (3) или приведено в контексте обвинения в неадекватности (4) «ты делаешь что-то неадекватно и даже напугала нас этим». Если же в тексте нет указаний на то, кто

¹⁹ Возможно, предикат *обижать* — это литота (образное выражение, стилистическая фигура, оборот, в котором содержится художественное преуменьшение), скрывающая обозначение именно насильственных действий.

²⁰ Для более точной активизации сценариев парсер должен разделять высказывания, обращенные к нему в режиме диалога, и литературные диалоги, где собеседники обращают высказывания друг к другу. Вместе с тем моделирование эмоций собеседников (с учетом их отнесения к классам антагонистов и протагонистов) также является важным аспектом понимания художественного текста.

из участников находится ближе к читателю (т.е. активизация сценариев ОПАСН и МЫ•ОПАСН равновероятна), парсер имеет возможность выбрать сценарий в зависимости от имитируемого настроения робота; например, в негативном настроении идентифицировать²¹ себя с пациентом сценария ОПАСН: решить, что «эта ситуация мне угрожает».

Рассмотрим вариант сценария НЕАДЕКВ («Неадекватность»), который активизируется сообщениями о неадекватности действий антагониста следующими примерами:

(5) *Глупо было плевать против ветра*²².

(6) *Сначала мне показалось это невероятным*, и, естественно, я отверг ее.

(7) Впрочем, в то же время Колосков отметил, что *многие молодые люди в 19 лет также не отличаются благоразумием* относительно того, что размещают в соцсетях.

Наряду с корректными примерами, видно, что сообщения о неадекватных или затрудненных действиях могут использоваться:

(а) чтобы сообщить о преувеличенном значении некоторого признака, например, кровавая сцена убийства заставляет «обезуметь» свидетеля:

(8) когда он попал в родительскую спальню своей девушки, *он обезумел от ужаса*, пишет daily mail;

(б) чтобы сообщить о негативной ситуации, которую автор сообщения предлагает избежать:

(9) *Я вас плохо слышу*, и вы меня не слышите!

(10) Или *я плохо читаю?*

(в) контексты «отличия от других» могут указывать как на негативное, так и на позитивное качество; следующие примеры

подчеркивают позитивное качество, но оно ошибочно определено парсером как негативное:

(11) Люди, которые идут в эту сферу, *отличаются удивительным складом ума и оригинальным мышлением*.

(г) глаголы пропозициональной установки, например, *мне показалось*, позволяют ввести мнение, обозначив неуверенность; это может как указывать на ложный характер мнения (*ему показалось*), и тогда активизация сценария НЕАДЕКВ корректна, так и являться особой формой вежливости, заставляющей говорящего не высказывать свое мнение категорично, а прибегать к оговоркам (hedge) (Brown, Levinson, 1987):

(12) *Мне показалось*, что бриллиант — идеальная фамилия.

(13) *Прохожий показался мне приятным*.

Как видно из примеров, сценарии специфическим образом выделяют контексты — кандидаты на эмоциональный отклик. Вместе с тем, хотя развитие точных и полностью автоматических средств определения эмоции текста является перспективой данного метода, уже сейчас эксперт может получать результаты запросов по семантике разобранных текстов для исследования исходных источников или для анализа понятий и семантического представления ситуаций.

Объединение сходных ситуаций, кластеризация. Число сценариев ограничено теми высказываниями, которые прямо используются для провокации эмоций в публицистике и рекламе. В работе А.А. Котова (2017) выделяется 34 эмоциональных сценария, в составе парсера мы используем 88 сценариев. Многие высказывания не направлены на передачу эмоций, они описывают различные ситуации, для которых тоже целесообразно ввести сценарии для лучшего распознавания этих ситуаций на этапе перехода от синтаксического компонента к семантике. Метод анализа здесь может состоять в том, чтобы разбить все

²¹ Идентификация парсера (или робота) с героем текста состоит в добавлении к валентности этого героя признака *ч-эгоцентрик* с помощью механизма приписывания признаков «сверху—вниз».

²² Курсивом выделены фрагменты (клаузы), приводящие к активизации сценария НЕАДЕКВ.

Таблица 3

Примеры результатов кластеризации семантических представлений (фактов)

№ кластера	Высказывание
364	<p>Да, я имел в виду именно это. Что ты имеешь в виду? Я вовсе не то имела в виду, – пробормотала она. Ты имеешь в виду, как это мы проснулись на заре, хотя легли очень поздно? Что ты имеешь в виду? Вы представляете, что я имею в виду? Я спросил, что он имеет в виду.</p>
591	<p>Егор быстро поправлялся, и на третий день он уже ходил с медсестрой по палатам и помогал ей делать уколы. Встает из-за стола и в носках ходит по кабинету. Но разве можно не реагировать на его близость и остроумные замечания, которые он делал, пока они ходили по дому? Они пошли по коридору, где было людно. Они пошли по коридору к выходу. Он повернулся и пошел по коридору. Они пошли по коридору в направлении какой-то еще комнаты. Они пошли назад по коридорам.</p>

полученные из текстов факты на отдельные группы, предполагая, что сгруппированные факты опишут одну и ту же ситуацию. Мы задали меру близости между двумя фактами и провели кластеризацию множества фактов²³; примеры кластеров приведены в табл. 3.

Как видно из таблицы, кластеризация может объединять как примеры, относящиеся к устойчивому речевому выражению (*иметь в виду*), так и описания сходных ситуаций («ходить по помещению»). В перспективе результаты кластеризации могут использоваться в нескольких направлениях. Во-первых, для анализа текстов: кластеризация может показывать типичные ситуации, представленные в тексте; с помощью запросов можно изучать, в какие кластеры входят конкретные понятия. Во-вторых, для пополнения инвентаря

сценариев: для создания сценариев для новых ситуаций, формирующих кластер. Такой подход позволит более дифференцированно распознавать ситуации текста, приблизиться к созданию системы автоматического понимания, которая сама реконструирует типичные ситуации реального мира по прочитанным текстам.

ПРИМЕР ПРИМЕНЕНИЯ МЕТОДА В СОЦИАЛЬНЫХ ИССЛЕДОВАНИЯХ

Возможности метода автоматической обработки текста крайне важны для наук, изучающих репрезентацию значимых социальных явлений в общественном сознании: социологии и социальной психологии. С помощью предложенного метода можно исследовать обсуждения негативных ситуаций пользователями Сети, выделять источники неудовлетворенности. Далее можно предположить, что социальные меры или современные технологии, решающие эти проблемы и устраняющие неудовлетворенность, получают общественную

²³ Использовались методы *k*-средних на 1 млн фактов и иерархическая кластеризация (agglomerative clustering) на 50 тыс. фактов, для оценки близости использовалась мера Жаккара.

поддержку людей или станут успешными коммерческими решениями²⁴. Описанный метод был применен нами для детализации вызовов и угроз, в общем виде описанных в Стратегии научно-технологического развития Российской Федерации. Анализ новостей и блогов выполнялся на основе базы данных объемом свыше 42 млн словоформ, собранной за 2017–2018 гг. на основе рейтинговых новостных сайтов и блогов. В рамках аналитической обработки для каждого вызова выбирались характерные семантические шаблоны: как по участию ключевых слов, так и по структуре фактов. Мы выполняли кластеризацию фактов, а также их группировку по ключевым словам и их валентностям. В результате анализа кластеров были сделаны выводы о восприятии актуальных угроз и вызовов пользователями Интернета, а также о перспективах роста соответствующих рынков товаров и услуг.

В области экономики²⁵ в исследуемом материале неудовлетворенность респондентов вызывает непрозрачность в управлении кредитами и налогами. Слово *кредит* наряду со стандартными частотными контекстами (*взять, получить, брать, выдавать кредит*) присутствует в эмоциональных контекстах, связанных с некорректным истребованием кредита:

(14) *Сбербанк требует погасить кредит, однако я никогда не брал кредит и не был поручителем в Сбербанке.*

(15) *Максиму не дали выехать за границу. Не оплатил коммуналку? Забыл про налоги? Не вернул долги? Страшнее — не закрыл чужой кредит.*

Другая группа эмоциональных текстов посвящена обсуждению тех, кто берет микрокредиты под большие проценты — из-

за бедности или чтобы потратить деньги на необоснованные цели:

(16) *Кто берет самые глупые кредиты под огромные проценты?*

(17) *Или кредит на новый айфон под бешеные проценты — это вообще запредел!*

Структура исследования предполагает, что эксперт должен сформулировать рекомендации о возможности устранения проблем или о перспективных продуктах. В данном случае перспективы могут состоять в разработке технологий, обеспечивающих информационную прозрачность и простоту обращения с кредитами. Была подчеркнута необходимость исследовать психологические и экономические процессы, толкающие людей брать микрозаймы под существенные проценты.

В области *проблем старения* населения можно выделить большую группу текстов, где люди описывают трудности заботы о пожилым родственнике, проживающем удаленно. Можно сделать вывод, что технологии, обеспечивающие возможность семьи заботиться о пожилым родственнике (в том числе — удаленно), могут составлять перспективу развития. Также авторы блогов часто смешивают возрастные нейродегенеративные заболевания и право пожилых людей высказывать собственное мнение: мнение пожилых объявляется «старческой деменцией». Необходимость разделить две эти проблемы выделяется при анализе как важная задача общества.

В области *питания* заметны две группы текстов: люди озабочены составом еды и контролем веса. Соответственно, перспективой обладают технологии контроля состава продуктов, а также технологии контроля над собственным потреблением и весом для улучшения здоровья. Время, затрачиваемое на приготовление еды, рассматривается людьми как важный ресурс: пользователи обсуждают, что тратят слишком много времени на покупку и приготовление пищи. Соответственно, перспективой обладают

²⁴ См. в этой связи: Bourdieu, 1977.

²⁵ При выборе областей мы следовали списку «больших вызовов» из Стратегии научно-технологического развития РФ. В данной публикации представлена часть вызовов из перечисленных в Стратегии.

технологии, обеспечивающие приготовление и доставку еды к нужному времени: еда может готовиться к моменту прихода человека в кафе или доставляться к моменту прихода человека домой. В большую группу выделяются случаи отказа от еды: как социальные действия (голодовки), так и медицинские случаи (анорексия). Возможная рекомендация состоит в том, что общество должно исследовать случаи отказа от еды, поскольку такие случаи становятся способом выражения различных общественных противоречий.

Задачи *освоения пространства* часто формулируются как необходимость освоения новых территорий. Однако проживание людей «в регионах» уже сейчас вызывает жалость блогеров, а способность человека «вырваться из региона» вызывает одобрение и восхищение. Если формулировать рекомендации, то перспективные технологии и социальные меры должны предоставить людям в регионах сервис и возможности самореализации — без необходимости уехать из региона. Возможно, такие технологии могут быть связаны с развертыванием сетей связи на большинстве территорий России, а также с доставкой через средства связи услуг телемедицины, дистанционного образования, дистанционного заказа товаров, создания удаленных рабочих мест («телеворкинга») и т.д.

ЗАКЛЮЧЕНИЕ

Компьютерные системы автоматического анализа текста стремятся полностью или частично воспроизвести функцию понимания текста человеком. С одной стороны, частичный разбор текста с помощью парсера становится высокотехнологичным инструментом современной психологии, методом проведения масштабных исследований, представляющих интерес для современного российского общества, в частности, в плане поиска эффективного ответа на большие

вызовы. Промежуточные уровни анализа текста, а также семантика, выделенная из больших массивов текстов, становятся доступными для запросов к базе данных и для дальнейшей аналитической работы. С другой стороны, компьютерные системы развиваются на основе теоретических моделей, предложенных для описания мышления, коммуникации и эмоций. Разработка данных систем может продвинуть компьютерное моделирование таких функций человека, как формирование понятий, классификация ситуаций, понимание текста и выбор коммуникативной реакции.

Авторы выражают благодарность Б.М. Величковскому за неизменную поддержку этого направления когнитивных исследований и А.А. Филатову за разработку основных компонентов программного обеспечения парсера.

1. *Апресян В.Ю.* ИмPLICITная агрессия в языке // Мат-лы Междунар. конф. «Компьютерная лингвистика и интеллектуальные технологии». Протвино, 11–16 июня 2003 г. С. 32–35.
2. *Апресян Ю.Д.* Избранные труды: В 2 т. 2-е изд., испр. и доп. М.: Языки русской культуры; Восточная литература, 1995. Т. 1: Лексическая семантика.
3. *Барулин А.Н.* К построению модели синтеза русских нумеративов (глубинное и поверхностно-семантическое представление) // Моск. лингв. журн. 1996. № 2. С. 27–53.
4. *Вежицкая А.* Семантические универсалии и описание языков. М.: Языки русской культуры, 1999.
5. *Величковский Б.М.* Когнитивная наука: Основы психологии познания: В 2 т. М.: Смысл; Изд. центр «Академия», 2006. Т. 2.
6. *Гладкий А.В.* Синтаксические структуры естественного языка в автоматизированных системах общения. М.: Наука, 1985.
7. *Гловинская М.Я.* Скрытая гипербола как проявление и оправдание речевой агрессии // Сокровенные смыслы: Слово. Текст. Культура / Под ред. Ю.Д. Апресяна. М.: Языки славянской культуры, 2004. С. 69–76.
8. *Грайс Г.П.* Логика и речевое общение // Лингвистическая прагматика / Под ред. Е.В. Падучевой. М.: Прогресс, 1985. (Новое в зарубежной лингвистике; т. 16). С. 217–237.

9. Журавлев А.П. Фонетическое значение. Л.: Изд-во Ленингр. ун-та, 1974.
10. Зайдельман Л.Я. и др. Формирование базы знаний робота-компаньона методами автоматического анализа текста / Зайдельман Л.Я., Аринкин Н.А., Зинина А.А., Котов А.А. // Мат-лы Всерос. междисциплинарной научно-практ. конф. с междунар. участием «Человек в информационном пространстве: понимание в коммуникации». Ярославль, 15–17 ноября 2018 г. С. 333–339.
11. Зинина А.А. и др. Разработка модели коммуникативного поведения робота ф-2 на основе мультимодального корпуса «РЕС» / Зинина А.А., Аринкин Н.А., Зайдельман Л.Я., Котов А.А. // Мат-лы Ежегод. междунар. конф. «Диалог» «Компьютерная лингвистика и интеллектуальные технологии». Москва, 30 мая – 2 июня 2018 г. С. 831–844.
12. Карпова О.С. и др. Оценочные значения ре-брендингового типа в признаковой лексике / Карпова О.С., Рахилина Е.В., Резникова Т.И., Рыжова Д. // Мат-лы Ежегод. междунар. конф. «Диалог» «Компьютерная лингвистика и интеллектуальные технологии». Бекасово, 25–29 мая 2011 г. С. 292–304.
13. Клековкина М.В., Котельников Е.В. Метод автоматической классификации текстов по тональности, основанный на словаре эмоциональной лексики // Мат-лы 14-й Всерос. науч. конф. «Электронные библиотеки перспективные методы и технологии, электронные коллекции». Переславль-Залесский, 15–18 октября 2012 г. С. 118–123.
14. Котов А.А. «Машина Оруэлла»: подходы к автоматическому созданию воздействующих текстов // Понимание в коммуникации: Человек в информационном пространстве: Сб. научных трудов: В 3 т. / Под ред. Е.Г. Борисовой, Н.В. Анискиной. Ярославль; М.: ЯГПУ, 2012. Т. 1. С. 405–418.
15. Котов А.А. Механизмы речевого воздействия в публицистических текстах СМИ: Дис. ... канд. филол. наук. М., 2003.
16. Лурия А.Р. Язык и сознание. М.: Изд-во Моск. ун-та, 1979.
17. Мельчук И.А. Опыт теории лингвистических моделей «СМЫСЛ <=> ТЕКСТ». М.: Школа «Языки русской культуры», 1999.
18. Пазельская А.Г., Соловьев А.Н. Метод определения эмоций в текстах на русском языке // Мат-лы Ежегод. междунар. конф. «Диалог» «Компьютерная лингвистика и интеллектуальные технологии». Бекасово, 25–29 мая 2011 г. С. 17.
19. Пропн В.Я. Морфология сказки. Л.: Academia, 1928.
20. Резникова Т.И., Копотев М.В. Лингвистически аннотированные корпуса русского языка (обзор общедоступных ресурсов) // Национальный корпус русского языка: 2003–2005. М.: Индрик, 2005. С. 31–61.
21. Санников В.З. Русский язык в зеркале языковой игры. М.: Языки русской культуры, 1999.
22. Сирл Д. Разум, мозг и программы // Глаз разума / Сост. и обраб. Д.Р. Хофштадтером и Д.К. Деннеттом. Самара: Бахрах-М, 2003. С. 315–331.
23. Тутубалина Е.В. и др. Тестирование методов анализа тональности текста, основанных на словарях / Тутубалина Е.В., Иванов В.В., Загулова М., Мингазов Н., Алимова И., Малых В. // Электронные библиотеки. 2015. Т. 18. № 3–4. С. 138–162.
24. Шведова Н.Ю. Русский семантический словарь. Толковый словарь, систематизированный по классам слов и значений. М.: Азбуковник, 1998.
25. Brown P., Levinson S.C. Politeness: Some universals in language usage. Studies in interactional sociolinguistics. Cambridge: Cambridge Univ. Press, 1987.
26. Bourdieu P. La production de la croyance // Actes de la recherche en sciences sociales. 1977. V. 13. N 1. P. 3–34.
27. Clore G.L., Ortony A. Cognition in emotion: Always, sometimes, or never? // Lane R.D., Nadel L. (eds). Cognitive neuroscience of emotion. Oxford: Oxford Univ. Press, 2000. P. 24–61.
28. Fillmore C.J. The case for case // Bach E.W., Harms R.T. (eds). Universals in linguistic theory. N.Y.: Holt, Rinehart & Winston, 1968. P. 1–68.
29. Kim S.-M., Hovy E. Determining the sentiment of opinions // Proceedings of the 20th International Conference on Computational Linguistics. Geneva, Switzerland, August 23–27, 2004. P. 1367.
30. Kopp S. et al. Towards a common framework for multimodal generation: The behavior markup language / Kopp S., Krenn B., Marsella S., Marshall A.N., Pelachaud C., Pirker H., Thyrisson K.R., Vilhjilmsson H. // Proceedings of the 6th International Conference “Intelligent virtual agents”. Marina Del Rey, CA; USA, August 21–23, 2006. P. 205–217.
31. Kotov A.A. A computational model of consciousness for artificial emotional agents // Psychology in Russia: State of the Art. 2017. V. 10. N 3. P. 57–73.
32. Kotov A. et al. Semantic comprehension system for F-2 emotional robot / Kotov A., Arinkin N., Filatov A., Zaidelman L., Zinina A. // Proceedings of the First International Early Research Career Enhancement School on BICA and Cybersecurity. Moscow, Russia, August 1–3, 2017. P. 126–132.

33. *Kotov A., Zinina A., Filatov A.* Semantic parser for sentiment analysis and the emotional computer agents // Proceedings of the AINL-ISMW FRUCT 2015. Saint-Petersburg, Russia, November 9–14, 2015. P. 167–170.
34. *Lakoff G., Johnson M.* Metaphors we live by. Chicago; L.: Univ. of Chicago Press, 2003.
35. *Langacker R.W.* Concept, image and symbol: The cognitive basis of grammar. Berlin: Mouton de Gruyter, 1991.
36. *McClelland J.L., Rumelhart D.E.* An interactive activation model of context effects in letter perception: Part 1. An account of Basic Findings // Psychol. Rev. 1981. V. 88. P. 375–407.
37. *Miller G.* et al. WordNet: An on-line lexical database / Miller G., Beckwith R., Fellbaum C., Gross D., Miller K. // Intern. J. Lexicography. 1990. V. 3. N 4. P. 235–244.
38. *Mohammad S.M., Turney P.D.* Emotions evoked by common words and phrases: Using mechanical Turk to create an emotion lexicon // Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text. Los Angeles, California, June 5, 2010. P. 26–34.
39. *Rosch E.* et al. Basic objects in natural categories / Rosch E., Mervis C.B., Gray W.D., Johnson D.M., Boyes-Braem P. // Cognit. Psychol. 1976. V. 8. N 3. P. 382–439.
40. *Schank R.C.* Tell me a story: Narrative and intelligence. Evanston, IL: Northwestern Univ. Press, 2000.
41. *Socher R.* et al. Recursive deep models for semantic compositionality over a sentiment treebank / Socher R., Perelygin A., Wu J., Chuang J., Manning C.D., Ng A., Potts C. // Proceedings of the 2013 Conference on empirical methods in natural language processing. Seattle, Washington, USA, October 18–21, 2013. P. 1631–1642.
42. *Tausczik Y.R., Pennebaker J.W.* The psychological meaning of words: LIWC and computerized text analysis methods // J. Language and Soc. Psychol. 2010. V. 29. N 1. P. 24–54.
43. *Vilhjálmsón H.* et al. The behavior markup language: Recent developments and challenges / Vilhjálmsón H., Cantelmo N., Cassell J., Chaffai N.E., Kipp M., Kopp S., Mancini M., Marsella S., Marshall A.N., Pelachaud C., Ruttikay Z., Thyrisson K.R., van Welbergen H., van der Werf R.J. // Proceedings of the 7th International Conference “Intelligent Virtual Agents”. Paris, France, September 17–19, 2007. P. 99–111.
44. *Wierzbicka A.* Semantic primitives. Frankfurt/M.: Athenlum-Verl., 1972.
45. *Yeh W., Barsalou L.W.* The situated nature of concepts // Am. J. Psychol. 2006. V. 119. N 3. P. 349–384.

References in Russian:

1. *Apresyan V.Yu.* Implicitnaya agressiya v yazyke [Implicit aggression in language] // Mat-ly Mezhdunar. konferencii Dialog “Komp’yuternaya lingvistika i intellektual’nye tehnologii”. Protvino, 11–16 iyunya 2003 g. S. 32–35.
2. *Apresyan Yu.D.* Izbrannye trudy. Leksicheskaya semantika [Selected Works. Lexical semantics]. V 2 t. T. 2. 2-e izd., ispr. i dop. M.: Yazyki russkoj kul’tury; Vostochnaya literatura, 1995.
3. *Barulin A.N.* K postroeniyu modeli sinteza russkix numerativov (glubinnoe i poverxnostno-semanticheskoe predstavlenie) [Towards the description of deep and surface semantic structure of Russian numeratives] // Moskovskij lingvisticheskij zhurnal. 1996. N 3. S. 27–53.
4. *Wierzbicka A.* Semanticheskie universalii i opisanie yazykov [Semantic universals and language description]. M.: Yazyki russkoj kul’tury, 1999.
5. *Velichkovskij B.M.* Kognitivnaya nauka: Osnovy psixologii poznaniya [Cognitive science: Foundations of epistemic psychology]. V 2 t. T. 2. M.: Smysl; Izdatel’skij centr «Akademiya», 2006.
6. *Gladkij A.V.* Sintaksicheskie struktury estestvennogo yazyka v avtomatizirovannyx sistemax obshheniya [Syntactic structures of natural language in computer-aided communication systems]. M.: Nauka, 1985.
7. *Glovininskaya M.Ya.* Skrytaya giperbola kak proyavlenie i opravdanie rechevoj agressii [Hidden hyperbole as a manifestation and justification of verbal aggression] // Sokrovennye smysly. Slovo. Tekst. Kul’tura / Pod red. Yu.D. Apresyana. M.: Yazyki slavyanskij kul’tury, 2004. S. 69–76.
8. *Grice H.P.* Logika i rechevoe obshhenie [Logic and speech communication] // Novoe v zarubezhnoj lingvistike. T. 16. Lingvisticheskaya pragmatika. M., 1985. S. 217–237.
9. *Zhuravlev A.P.* Foneticheskoe znachenie [Phonetic meaning]. L.: Izd-vo Leningr. un-ta, 1974.
10. *Zajdel’man L.Ya.* et al. Formirovanie bazy znaniy robota-kompan’ona metodami avtomaticheskogo analiza teksta [Building a knowledge base through natural language text analysis methods] / Zajdel’man L.Ya. Arinkin N.A., Zinina A.A., Kotov A.A. // Mat-ly Vserossijskoj s mezhdunarodnym uchastiem mezhdisciplinarnaya nauchno-prakticheskaya konferenciya «Chelovek v informacionnom prostranstve: ponimanie v kommunikacii». Yaroslavl’, 15–17 noyabrya 2018 g. S. 333–339.
11. *Zinina A.A.* et al. Razrabotka modeli komunikativnogo povedeniya robota f-2 na osnove mul’timodal’nogo korpusa «REC» [Developing of communicative behavior model for F-2 robot basing on the “REC” multimodal corpora] / Zinina A.A., Arinkin N.A., Zajdel’man, L.Ya.. Ko-

- тов А.А. // Mat-ly Ezhegodnoj mezhdunarodnoj konferencii «Dialog» «Komp'yuternaya lingvistika i intellektual'nye texnologii». Moskva, 30 maya – 2 iyunya 2018 g. S. 831–844.
12. *Karpova O.S.* et al. Ocenochnye znacheniya rebrandingovogo tipa v priznakovoj leksike [Meaning of estimation in semantic shifts of rebranding type in adjectives and adverbs] / Karpova O.S., Raxilina E.V., Reznikova T.I., Ryzhova D. // Mat-ly Ezhegodnoj mezhdunarodnoj konferencii «Dialog» «Komp'yuternaya lingvistika i intellektual'nye texnologii». Bekasovo, 25–29 maya 2011 g. S. 292–304.
 13. *Klekovkina M.V., Kotel'nikov E.V.* Metod avtomaticheskoy klassifikacii tekstov po tonal'nosti, osnovannyj na slovare e'mocional'noj leksiki [The method of automatic texts classification by tonality, based on the emotional vocabulary] // Mat-ly 14-j Vserossijskoj nauchnoj konferencii «E'lektronnye biblioteki perspektivnye metody i texnologii, e'lektronnye kollekicii». Pereslavl'-Zaleskij, 15–18 oktyabrya 2012 g. S. 118–123.
 14. *Kotov A.A.* Mexanizmy rechevogo vozdejstviya v publicisticheskix tekstax SMI [Mechanisms of speech influence in media texts]: Dis. ... kand. filol. nauk. M., 2003.
 15. *Kotov A.A.* «Mashina Orue'lla»: podxody k avtomaticheskomu sozdaniyu vozdejstvuyushix tekstov [Orwell's machine: approaches to automatic creation of influencing texts] // Ponimanie v kommunikacii: Chelovek v informacionnom prostranstve: sb. nauchnyx trudov. V 3 tt. / Pod red. E.G. Borisovoj, N.V. Anis'kinoj. Yaroslavl'–Moskva: YaGPU, 2012. T. 1. S. 405–418.
 16. *Luriya A.R.* Yazyk i soznanie [Language and Cognition]. M.: Izd-vo Mosk. un-ta, 1979.
 17. *Mel'uk I.A.* Opyt teorii lingvisticheskix modelej «SMYSL <=> TEKST» [The theory of linguistic models Meaning-Text]. M.: Shkola «Yazyki russkoj kul'tury», 1999.
 18. *Pazel'skaya A.G., Solov'ev A.N.* Metod opredeleniya e'mocij v tekstax na russkom yazyke [A method of sentiment analysis in russian texts] // Mat-ly Ezhegodnoj mezhdunarodnoj konferencii «Dialog» «Komp'yuternaya lingvistika i intellektual'nye texnologii». Bekasovo, 25–29 maya 2011 g. S. 17.
 19. *Propp V.Ya.* Morfologiya skazki [Fairy tale morphology]. Leningrad: Academia, 1928.
 20. *Reznikova T.I., Kopotev M.V.* Lingvisticheski annotirovannye korpusa russkogo yazyka (obzor obshhedostupnyx resursov) [Linguistically annotated corpora of the Russian language] // Nacional'nyj korpus russkogo yazyka: 2003–2005. M.: Indrik, 2005. S. 31–61.
 21. *Sannikov V.Z.* Russkij yazyk v zerkale yazykovoj igry [Russian language in the mirror of the language game]. M.: Yazyki russkoj kul'tury, 1999.
 22. *Searle J.* Razum, mozg i programmy [Minds, Brains and Science] // Glaz razuma / Sost. i obrab. D. R. Xofshtadterom i D. K. Dennettom. Samara: Bahrah-M, 2003. S. 315–331.
 23. *Tutubalina E.V.* et al. Testirovanie metodov analiza tonal'nosti teksta, osnovannyx na slovaryax [Testing of dictionary-based text sentiment analysis methods] / Tutubalina E.V., Ivanov V.V., Zagulova M., Mingazov N., Alimova I., Malyx V. // E'lektronnye biblioteki. 2015. T. 18. N 3–4. S. 138–162.
 24. *Shvedova N.Yu.* Russkij semanticheskij slovar'. Tolkovyj slovar', sistematizirovannyj po klassam slov i znachenij [The Russian Semantic Dictionary]. M.: Azbukovnik, 1998.

Приложение

**Инвентарь сценариев, ответственных за распознавание эмоциональных смыслов
(Котов, 2003, 2012)**

	Название сценария	Пример
Негативные сценарии	1. ОПАСН – Опасность	Вокруг враги!
	2. ОГРАНИЧ – Ограничение	Меня все ограничивают.
	3. ПРИСВ – Присвоение ресурса	У меня отобрали самое ценное!
	4. ОБМАН – Обман, скрытность, хитрость	Меня все обманывают.
	5. МАНИП – Манипулирование	Они контролируют наше сознание!
	6. ПЛАНИР – Намерение, планирование	Они что-то задумали.
	7. НЕАДЕКВ – Неадекватность	Они не соображают, что делают.
	8. НЕПОСЛЕД – Непоследовательность	Как можно говорить одно, а потом – противоположное?

	Название сценария	Пример
	9. ЭМОЦ – Эмоциональность 10. СУБЪЕКТ – Субъективность, импульсивность, произвол 11. БЕЗДЕЙСТВ – Бездействие 12. ТЩЕТН – Тщетность 13. НЕНУЖН – Ненужность	Не нужно так психовать! Ты думаешь только о себе! Ты не можешь даже мне помочь? Все бесполезно. Я никому не нужен.
Позитивные сценарии	1. ВКУС – Восхищение вкусом еды, желание насытиться 2. ВИД – Красота, эстетическая привлекательность 3. ЗВУК – Удовольствие от восприятия музыки, красивого звука 4. КОМФОРТ – Приятные тактильные ощущения, восприятие собственного тела, домашний уют 5. СЕКС 6. ПОЛУЧ – Коллекционирование, получение ресурсов 7. ВЕСЕЛЬЕ – Хорошее настроение, веселая компания, кураж 8. КОНТРОЛЬ – Контроль над ситуацией и управление действиями других 9. МЫ • ОПАСН – Готовность победить контрагента, обладание оружием 10. МЫ • ОГРАНИЧ – Ограничение контрагента, лишение его свободы 11. МЫ • ПРИСВ – Уничтожение, присвоение или использование чужого ресурса 12. ВООДУШ – Воодушевление, противостояние опасности 13. СЛУЖЕНИЕ – Эффективность работы, служение цели, вклад в общее дело 14. ЗАЩИТА – Контрагент защищает субъекта 15. ЗАБОТА – Контрагент заботится о субъекте 16. МЫ • ЗАБОТА – Субъект заботится о контрагенте 17. НЕОБЫЧН – Открытие нового, экзотика, экстрим 18. СВОБОДА – Свобода в действиях и выражении чувств 19. ТВОРЕНИЕ – Осуществление мечты, решение проблемы, творчество 20. ПРЕВОСХ – Превосходство, одобрение 21. ВНИМАНИЕ – Внимание и понимание	Это очень вкусно. Это очень красиво. Какая красивая музыка! Мне очень удобно на этом диване. Ум – это сексуально. Собери свою коллекцию! Как здорово собрать веселую компанию. Ты полностью контролируешь поведение автомобиля на дороге. Мы уничтожим террористов. Мы не дадим врагам свободно вздохнуть. Мы отберем богатства у олигархов. Нам противостоит сильнейший враг. Мы посвятим всю свою жизнь науке. Специалисты защитят нашу экономику от потрясений. Как приятно, когда исполняют твои желания. Я покупаю кошке самое лучшее. Перед нами – огромный и необычный мир. Я могу делать, что захочу. Воплоти свою мечту. Ты – настоящий мужчина! Все обращают на тебя внимание.