

А. А. Зинина, Л. Я. Зайдельман, Н. А. Аринкин, А. А. Котов
НИИЦ «Курчатовский институт»
(Россия, Москва)
zinina_aa@nrcki.ru

ПОДХОДЫ К КЛАССИФИКАЦИИ КОММУНИКАТИВНЫХ ЖЕСТОВ: ТЕОРИЯ И ПРИЛОЖЕНИЕ К РОБОТАМ-КОМПАНЬОНАМ*

В работе представлен подход к классификации коммуникативных жестов — описаны жесты, у которых означаемое может быть выражено с помощью различных означающих. К классификации жестов мы подходим через выделение коммуникативных функций (означаемого) и инварианта выполнения жеста (означающего). Определенная коммуникативная функция может быть выражена жестами с различным набором исполнительных органов: с помощью мимики, движений головы, рук и тела. Путем выделения коммуникативных функций решается задача моделирования сложного коммуникативного поведения роботом-компаньоном. В работе обосновывается необходимость выделения коммуникативных функций при анализе реального поведения человека в эмоциональных ситуациях (на основе мультимодального корпуса REC), а также описываются контексты и инварианты коммуникативных функций. На примере коммуникативной функции *ожидание обратной связи* описывается инвариант «акцентирование внимания к ответной реакции собеседника» и различные варианты соответствующих жестов: (1) акцентирование взгляда на адресате, (2) демонстрация собственного бездействия при ожидании адресата, (3) демонстрация готовности перейти к последующим действиям, (4) кивки в подтверждение собственного тезиса или вопросительные движения с запросом реакции собеседника. Также описывается управляющая архитектура робота, с помощью которой моделируются основные особенности естественной коммуникации: коммуникативные функции и жесты. Описание инварианта и жестов для определенной коммуникативной функции позволяет более тонко настраивать коммуникативные реакции робота на входящее высказывание.

Ключевые слова: мультимодальный корпус, жестовая коммуникация, иконические жесты, робот-компаньон.

* Исследование выполнено при поддержке проекта РНФ № 19-78-00113 «Разработка эмоциональной модели робота-компаньона для решения обучающих задач с детьми».

Современная корпусная лингвистика развивается сразу в двух самостоятельных сферах: с одной стороны, исследования в области жестикуляционной лингвистики обладают самостоятельной теоретической значимостью, с другой — широким практическим потенциалом — их результаты могут быть полезны при синтезе невербального коммуникативного поведения человекоподобным роботом. Более того, только в единстве «теории» и «практики» открывается возможность создания такого робота-компаньона, который будет способен формировать положительное впечатление у пользователя, поддерживать коммуникацию, длительное время оставаться интересным и привлекательным для собеседника.

В области социальной робототехники проводится множество исследований социального взаимодействия между роботом и человеком. Известно, что часть пользователей устанавливает эмоциональный контакт даже с бытовыми роботами-пылесосами [Sung et al. 2007]. Поэтому робот не обязательно должен выглядеть как человек, наоборот, чрезмерная схожесть робота с человеком вызывает у пользователей эффект «зловещей долины» [Morì 1970]. Наш подход заключается в том, что робот должен обладать минимальной конфигурацией (наиболее простой кинематической моделью), но при этом проявлять максимум эмоций, чтобы эффективно действовать в качестве робота-компаньона. Такие роботы обычно имеют голову, лицо, иногда — ручки или уши.

Результаты эксперимента [Leite, Pereira et al. 2008] показали, что именно эмоциональное поведение повышает эффективность взаимодействия между физическим роботом и пользователем. Например, робот Kismet обладает эмоциональной моделью и может выражать имитируемые эмоции с помощью ограниченного набора движений элементов лица (глаз, бровей, рта, ушей), а также с помощью речи, включая речевую просодию [Breazeal 2003]. Робот iCat также обладает эмоциональной моделью, позволяющей либо ориентироваться при ответе на эмоции адресата, либо предпочитать собственные реакции [Hindriks et al. 2012]. Этот робот имеет подвижную голову, глаза, брови и рот. Приведенные исследования показывают, что робот должен обладать не только развитой эмоциональной моделью, но и широким спектром эмоциональных выразительных средств — жестов и мимики. Исследователи отмечают существенную значимость гибкого комбинирования жестов и речи при «интуитивном» человеко-машинном взаимодействии [Beuter, Spexard et al. 2008].

Чтобы иметь возможность выполнять ориентированные жесты (например, поворачивать голову к собеседнику или указывать направление), а также общаться с несколькими собеседниками одновременно, роботу необходимо обладать компьютерным зрением. Кроме того, для лучшего установления контакта между роботом и собеседником робот должен имитировать поведение живого человека [Breazeal, Scassellati 2002]. В исследовании [Klamer 2011] отмечается значимость для робота речевых и невербальных средств поддержания коммуникации при установлении долгосрочного контакта. Согласно психологическим исследованиям, привлекательность робота обусловлена тем, что его коммуникативное поведение должно быть максимально приближено к коммуникативному поведению

человека. Очевидно, что жестовое поведение человекоподобного робота должно быть сложным и разнообразным, поэтому при разработке такого поведения необходимо ориентироваться на теоретические представления о природе жестового общения, в частности на существующие классификации жестов.

Движения человека доступны для наблюдения другими людьми и могут передавать информацию о состоянии человека, его целях и эмоциях. С этой точки зрения движения могут являться знаками, передающими информацию наблюдателю. Вместе с тем жесты не составляют такого же словаря знаков, как слова естественного языка. Означающие и означаемые жестов достаточно размыты, это затрудняет создание классификации. Жесты разнородны как по своей произвольности (намеренности выполнения), так и по типу отношений между означающим и означаемым. С одной стороны, произвольные жесты контролируются говорящим и направлены на передачу конкретного значения: *махать рукой* выражает 'приветствие', а ладонь с растопыренными пальцами может обозначать число 'пять' — говорящий сознательно выполняет эти жесты с целью транслировать сообщение адресату. С другой стороны, трясущиеся руки могут обозначать напряжение говорящего, прикрывание лица рукой может вызываться смущением — говорящий может не контролировать эти движения, что тем не менее не препятствует передаче адресату информации о состоянии говорящего.

Согласно Ч. Пирсу, отношение между означающим и означаемым можно характеризовать через базовые типы знаков: иконические, индексные и символные. Если символные знаки совсем не похожи на обозначаемый ими предмет — их действие основано на установленной по соглашению связи означающего и означаемого, у индексных знаков означающее и означаемое находятся в отношениях пространственной и временной смежности, то у иконических знаков означающее подобно означаемому. Жесты также занимают в этой классификации различные позиции. Жесты «фига» или «ОК» имеют символную природу — означаемое в них связано с передаваемым смыслом «по договору». Китайские жесты для обозначения числительных пальцами одной руки различаются: числа от 1 до 5 — иконические: человек просто демонстрирует нужное число пальцев, а обозначения от 6 до 10 — символные знаки, они отчасти похожи на соответствующие иероглифы, но догадаться об обозначаемом числе по самим жестам — невозможно. В качестве примеров индексных знаков приведем трясущиеся руки или прикрывание лица рукой: эмоциональное состояние адресанта вызывает соответствующее движение, в понимании Пирса можно говорить о том, что между означаемым и означающим имеется физическая связь. Это затрудняет классификацию означаемых жеста: означаемым может быть и заданный смысл, как у слов естественного языка, и эмоциональное состояние говорящего — в этом случае смысл вычленяется адресатом коммуникации, но не формируется говорящим.

Такое разнообразие жестов человека усложняет создание классификации: жест может являться как рефлекторной реакцией человека, так и намеренным знаковым действием с целью передать адресату заданный смысл. Сложность представляет и классификация означающих жеста. Слова естественного языка при написании

делятся на сегменты пробелами, знаками пунктуации, а при произнесении — фонетическими параметрами: тактами вдоха и выдоха, переходами голосового аппарата от гласных к согласным звукам. По контрасту с речевым сигналом жест является движением человека в пространстве и с трудом поддается сегментации. Более того, движение человека в пространстве является сложным комплексом, который может совмещать несколько движений: жест «пять», прикрывание лица рукой, трясущиеся руки и т. д. Отсутствие для жестов заданного набора сегментов, аналогичных алфавиту, словарю или набору фонем, — также затрудняет создание классификации. Данные трудности по-разному решаются в разных теоретических подходах.

Д. Эфрон [Efron 1941/1972] предлагает разделять следующие классы жестов. Во-первых, он выделяет жесты-эмблемы, кодирующие смысл независимо от лексического контекста, имеющие строго закрепленные за ними условия употребления, — например жест «ОК». Во-вторых, жесты-иллюстраторы, которые сопровождают определенный речевой фрагмент и связаны с его содержанием. В-третьих, жесты-регуляторы, управляющие ходом коммуникативного процесса, то есть устанавливающие, поддерживающие или завершающие коммуникацию [Григорьева и др. 2001].

Согласно классификации Д. МакНила [McNeill 2005], жест обладает некоторой размерностью: иконичностью, метафоричностью, деиктичностью, ритмичностью. Под иконическими понимаются такие жесты, которые дополняют речевое высказывание, — это изображения конкретных вещей или действий. Метафорические изображают абстрактные, а не конкретные предметы, деиктичные — указывают на определенный предмет или на местоположение какой-либо абстракции в пространстве, причем такие жесты могут выполняться не только рукой, но и головой, носом, локтем, ногой, глазами и др. Ритмические жесты подчеркивают значимые фрагменты речи.

А. Кендон различает жесты по трем параметрам: обязательность сопровождающей речи, наличие или отсутствие системных языковых черт в жестикуляции, а также степень регулярности жеста. Согласно его классификации, все жесты могут быть распределены на континууме по степени обязательности сопровождающей речи — системная организация в сочетании с произвольностью отношений между означаемым и означающим. Крайняя левая точка такого континуума будет описывать «речевые» жесты, которые не могут быть однозначно интерпретированы без сопровождающего речевого высказывания [Kendon 1988].

Е. А. Гришина обобщает, что на сегодняшний день нет единой и общепринятой классификации жестов [Гришина 2017: 12]. На практике исследователи сталкиваются с невозможностью отнести какой-либо жест к одной определенной категории, поэтому, руководствуясь специализированной практической задачей — задачей моделирования коммуникативного поведения человекоподобным роботом-компаньоном, — мы разрабатываем собственную классификацию жестов.

Чтобы объединить произвольные и спонтанные знаки в одну классификацию, мы используем следующий подход. Адресант может переживать различные состояния и проявлять их в своем поведении. Пусть состояние *C* (*condition*) адресанта

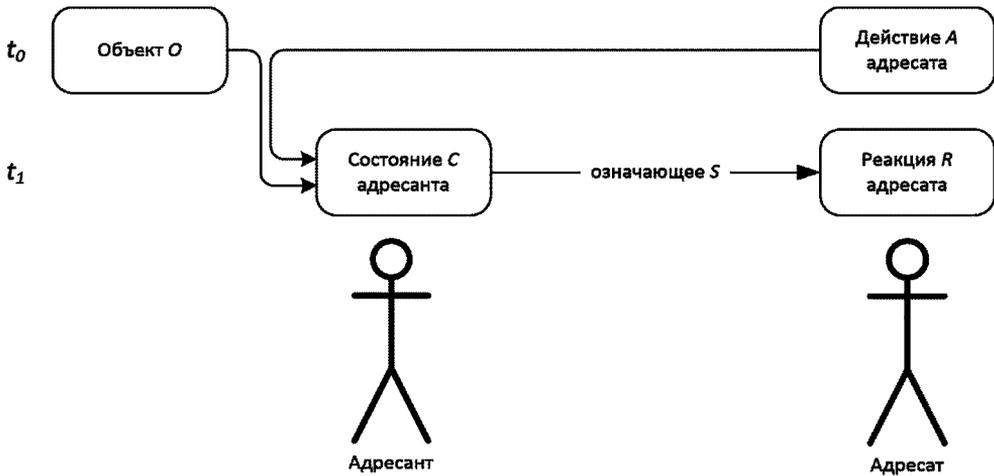


Рис. 1. Схема коммуникации для описания жестов

заставляет его выполнить действие S (*sign*) — (рис. 1). Например, адресант загрустил (C) и глубоко вздохнул (S). В этом случае C влечет S : имеет место $C \Rightarrow S$ — грусть вызывает вздох. Адресат может наблюдать поведение адресанта и интерпретировать S как знак внутреннего состояния C , в этом случае S будет индексным знаком (означающим), указывающим на C (означаемое). Мы можем считать S сообщением в коммуникации, даже если S носит спонтанный характер и не контролируется адресантом. Интерпретируя знак S , адресат реконструирует состояние C адресанта, спросит того, как дела, и утешит его. В этом случае имеет место причинно-следственная связь $S \Rightarrow R$, где R — реакция адресата.

Состояние C могло быть вызвано внешними причинами. Некоторый объект O в окружении адресанта мог вызывать его состояние C , — например пробудить интерес или напугать. Демонстрируемые при этом адресантом действия S могут служить знаком, указывающим на присутствие в окружении привлекательного или опасного объекта. Такая простая схема поведения позволяет стайным живым существам реагировать на опасность или появление еды — при этом стимулом для действий (R) живого существа становится поведение (S) сородича. Поведение S можно рассматривать как индексный знак опасности O . Такую же реакцию живое существо может демонстрировать в ответ на действия A своего оппонента. Как показано в классических работах К. Лоренца [2001], агрессия является важным семиотическим механизмом, регулирующим поведение в популяции через ритуалы и знаковое взаимодействие. Животное может продемонстрировать форму агрессивного поведения A , на которое оппонент ответит встречной агрессией S . Однако тот из оппонентов, чья мотивация ниже, может в ходе конфликта продемонстрировать знак подчинения (S), это останавливает (R) агрессию оппонента и позволяет выбрать победителя конфликта без ущерба здоровью оппонентов.

У человека можно наблюдать набор поведенческих реакций, сходный с инстинктами животных из описаний Лоренца: объекты (O) или действия (A) других могут

вызывать у нас эмоции, проявляющиеся в поведении (S). Однако человек как высшее разумное существо может контролировать собственные действия (S) и моделировать реакции оппонента (R). Пусть в простом случае запах тухлого яблока (O) вызывал у адресанта отвращение (C), что заставило адресанта закрыть нос и рот рукой (S). Это простая реакция вида $O \Rightarrow C \Rightarrow S$. Сходную поведенческую реакцию могут вызвать неприятные действия A другого человека. В этом случае также имеет место причинно-следственная связь $A \Rightarrow C \Rightarrow S$. В произвольных реакциях объекты O и действия A будут оказывать непосредственное влияние на состояние C и вызывать действия S . Теперь представим, что входящий стимул вызывает у адресанта множество вариантов интерпретации, — то есть имеется множество состояний C и множество вариантов действий S , которые адресант может контролировать. Адресант имеет возможность прогнозировать реакции адресата на каждое такое S — будем считать, что реакции R подвержены большему контролю адресанта, и обозначать их R' . Прогнозируя различные реакции R' , адресант будет выбирать такие действия S , которые приведут к наиболее выгодным R' . Контролируемые действия будем обозначать S' . С одной стороны, это позволит адресанту выполнять референцию к объекту O : адресант может продемонстрировать интерес, внимание или эмоциональную оценку (S) объекта O , чтобы нужным образом обратить на O внимание адресата (вызвать у адресата реакцию R' — внимание или эмоциональную оценку). С другой стороны, адресант сможет контролировать действия A адресата. Пусть адресат поступил некрасиво (A), что вызвало осуждение (C) адресанта. Он может проявить осуждение в форме, похожей на отвращение, — закрыть нос и рот рукой. Адресант при этом будет демонстрировать более контролируемое проявление (S') эмоций (C), направленное на воздействие (R') на адресата. В диалоге люди, как правило, стараются снизить ущерб оппоненту [Brown, Levinson 1987]: если адресат своими действиями A доставил неудобство адресанту, то он будет стараться извиниться или избежать повторения неприятной ситуации, — то есть отвечать своими реакциями R на действия S собеседника. Такое изменение поведения адресата входит в задачу намеренного использования знаков S' . Это звено является ключевым для описания перехода от неконтролируемых действий S к произвольному поведению S' . Именно на этом шаге жесты из неконтролируемых индексных знаков становятся сообщениями, обладающими иллюкутивной силой — направленными на воздействие на реакции R адресата.

Состояние C вызвано некоторым объектом O , поэтому, демонстрируя свое состояние (C), адресант транслирует оценку объекта O . При этом адресант осуществляет референцию к объекту O , вкладывая в свое сообщение как оценку O , так и указание на O . Наблюдая S , адресат будет реконструировать внутреннее состояние C адресанта, а также объект O , вызвавший это состояние. Это ключевой переход от знаков-реакций (индексных знаков, обозначающих эмоции) к знакам, обладающим референцией и отсылающим к объекту O . Если же состояние C вызвано предшествующими действиями адресата A , то знак S получает референцию к предшествующим действиям адресата. К примеру, S может сообщить адресату, что его предшествующие действия A негативно повлияли на адресанта.

В предложенной схеме опора на те или иные компоненты коммуникации прямо соответствует функциям коммуникативного акта Р. Якобсона [Jakobson 1960]. Так, ориентация говорящего на реакцию адресата *R* соответствует *конативной* функции, ориентация на объект *O* — *референтивной* функции, а ориентация на собственное состояние *S* говорящего — *эмотивной* функции. При этом в одном жесте могут совмещаться несколько функций: иконический жест может иллюстрировать объект *O*, а особенности выполнения жеста могут добавлять оценку адресанта или воздействовать на реакции адресата. Важно отметить, что мимические движения имеют те же основные функции, что и жесты рук и ног. Как замечает Крейдлин, «лицо и многие связанные с ним мимические жесты и движения, такие как поднимать брови, закрывать глаза, надуть губы, поджать губы, закусить губу, наморщить лоб, нахмуриться, улыбка, поцелуи (в лоб, в щеку, в губы), и некоторые другие не только соотносятся с конкретными эмоциями, но и выполняют определенные коммуникативные и социальные функции» [Григорьева и др. 2001: 228]. Таким образом, чтобы указать место коммуникативных жестов в общей типологии, мы будем рассматривать более широкий спектр движений человека, относя мимику и коммуникативные движения к *жестам* в расширенном понимании этого термина.

Объект исследования

В лаборатории нейрокогнитивных технологий Курчатовского комплекса НБИКС-технологий мы исследуем коммуникативное поведение людей в реальных эмоциональных ситуациях и стремимся предложить такое описание, которое может использоваться для управления поведением робота-компаньона. — При этом мы используем робота Ф-2¹ [Kotov et al. 2018; Зинина и др. 2018]. Классификацию коммуникативного поведения мы выполняем на основе мультимодального корпуса REC (Russian Emotional Corpus), содержащего размеченные в программе ELAN [Brugman, Russel 2004] видеозаписи различных эмоциональных диалогов на университетских экзаменах (295 фрагментов), в муниципальной службе одного окна (510 фрагментов), а также диалогов с информантами, которые занимаются каким-либо видом искусства, например хореографией или рисованием (10 фрагментов). Сбор и разметка корпуса осуществляются в нашей лаборатории на протяжении десяти лет [Kotov, Vudyanskaya 2012]. В корпусе вручную размечаются речевые высказывания участников диалога. Для информанта (студента, клиента, респондента) размечаются движения его глаз, рук и губ. Мимикой глаз считаются коммуникативно значимые движения глаз и их окружения: век, бровей, мышц вокруг глаз, носа, например движения глазами вверх, вбок, расширение глаз, частые моргания и др. К движениям губ относятся улыбки, облизывания, прикусывания и др. Разметка движений рук выполняется на четырех слоях: это способ выполнения жеста, активный и пассивный органы, а также траектория движения. К жестам рук относятся всевозможные почесывания, поглаживания, манипуляции, иконические

¹ Сайт робота Ф-2: <http://f2robot.com/>.

знаки и др. На сегодняшний день в корпусе насчитывается более 304 000 аннотаций. При разметке видеозаписей мы специально отслеживаем согласованность экспертов — каждую видеозапись просматривают не менее двух человек, после чего обсуждаются спорные или неоднозначные моменты.

С точки зрения приведенного основания классификации для нас представляют интерес (а) означающие жестов — какие внутренние состояния S или цели R вызывают появление того или иного жеста, а также (б) означаемые жестов — какие формы выражения S позволяют выразить определенное означающее. Означающее жестов мы будем описывать через **коммуникативные функции**, а означаемые — через **инварианты** коммуникативных действий.

Означаемое: коммуникативные функции

На основании результатов анализа базовых элементов разметки мы вводим дополнительный уровень разметки, который задает коммуникативную функцию мимического движения или жеста, если эта функция может быть определенно установлена. Коммуникативные функции приписывают новые параметры аннотациям из базовой разметки корпуса, также классифицируют движения головы и тела человека, поскольку базовая разметка у этих элементов в корпусе отсутствует. Например, перпендикулярному движению руки или повороту головы может быть приписана функция *отрицание*. Это значит, что, двигая руку или поворачивая голову, адресант выражает смысл 'нет'. Мы выделяем 35 коммуникативных функций [Котов, Зинина 2015], которые частично уже перенесены на робота Ф-2.

Данная классификация является поисковой: мы продолжаем ее развитие на основе корпуса. Разработка классификации ведется по следующим принципам. Имеются коммуникативные функции, непосредственно связанные со структурой диалога (иллюквативными целями) и ориентированные на воздействие на говорящего, например *понимание-согласие-одобрение*, *апелляция*, *побуждение* и т. д. Для этих коммуникативных функций наиболее важен компонент коммуникации R — реакция адресата. Вторая группа коммуникативных функций связана с выражением внутреннего состояния адресата: *я-хезитация*, *я-смущение-фрустрация*, *я-радость* и т. д. Наиболее существенно для этой группы внутреннее состояние адресата S , это подчеркивается префиксом «я». Однако ключевой компонент коммуникации может меняться: часто говорящий демонстрирует фрустрацию, чтобы спровоцировать адресата на помощь, то есть в коммуникативной функции *я-смущение-фрустрация* ключевым компонентом становится реакция адресата R' , а внешнее выражение этой коммуникативной функции будет лучше контролироваться адресантом — S' . Третья группа коммуникативных функций связана с проявлением состояний (S), преимущественно ориентированных на воздействие на адресата (R'): *ты-пренебрежение*, *ты-позитивен-к-объекту*, *ты-попытка-успокоить*. Ключевым компонентом коммуникации здесь будет реакция адресата — R' , однако в схожих случаях адресант сам может «наплеватьски» относиться к некоторому объекту и выражать это в своих жестах (*ты-пренебрежение*). Четвертая

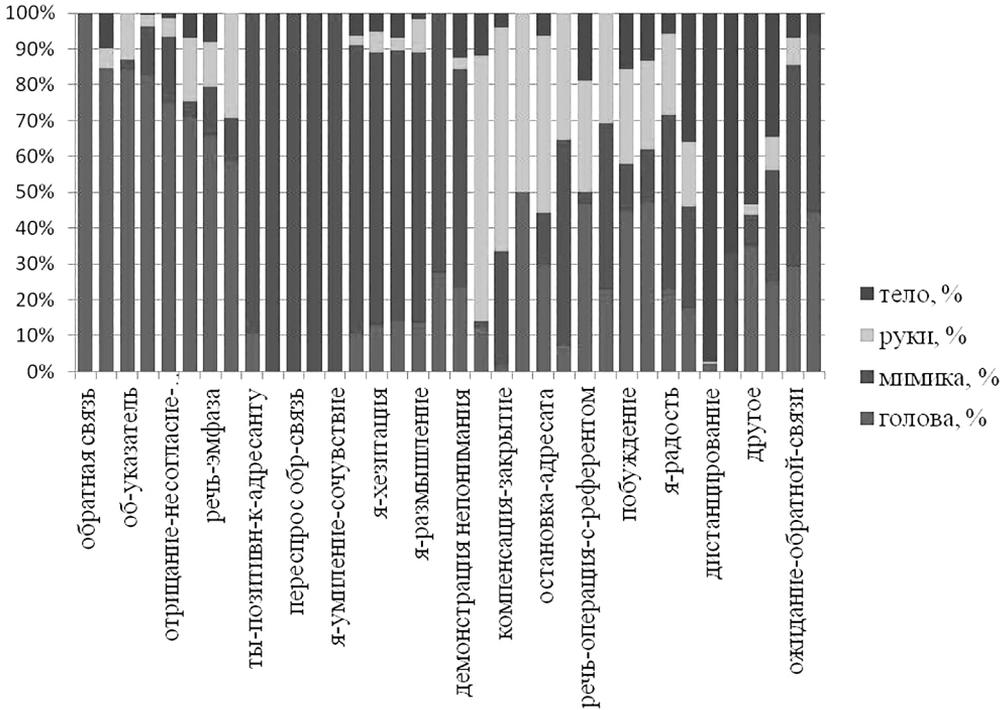


Рис. 2. Выражение коммуникативных функций

группа коммуникативных функций связана с обозначаемым объектом — реальным или воображаемым: *речь-референт*, *выбор-варианта* и т. д. Ключевым компонентом здесь будет объект *O*. Как видно из классификации, намерения выразить свое внутреннее состояние *C*, выполнить референцию к объекту *O* или воздействовать на адресата *R'* — могут в разном соотношении влиять на появление той или иной коммуникативной функции. Если же сумма этих составляющих превышает некоторый порог, коммуникативная функция считается активизировавшейся и далее может проявиться через те или иные жесты.

Каждая коммуникативная функция может выражаться с помощью различных движений — одно означаемое может иметь различные означающие. Более того, отдельное означающее может быть по-разному распределено между несколькими исполнительными органами — выражаться преимущественно с помощью головы, рук, мимики или тела (рис. 2). Например, такая функция, как *апелляция*, в 46,1% случаев выражается с помощью движений головы, в 24,5% — жестами рук, в 15,8% — с помощью мимики и в 13,6% — движениями тела. При этом *компенсация-закрытие* жестами рук выражается в 62,7% случаев, с помощью мимики — в 31,6% случаев, движениями тела — в 3,8%, а движениями головы — лишь в 1,9% случаев. *Я-хезитация*, наоборот, выражается преимущественно мимическими движениями — в 75,7% случаев, с помощью движений головы — в 13,1%, в 6% случаев — с помощью тела и лишь в 2,7% — движениями рук. Коммуникативная

Сочетаемость коммуникативных функций

	Функция	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	понимание-согласие-одобрение	15	0	2	7	2	2	1	27	1	5	0	10	7	0	10
2	отрицание-несогласие-возражение	0	12	1	4	3	0	4	9	1	19	1	4	10	0	7
3	апелляция	2	1	48	7	5	3	1	7	0	6	2	8	27	0	14
4	ожидание-обратной-связи	7	4	7	41	1	1	2	17	1	4	2	15	35	0	12
5	остановка-адресата	2	3	5	1	7	0	1	1	1	9	0	5	0	0	5
6	отсутствие-невозможность	2	0	3	1	0	10	0	12	0	15	1	4	3	1	6
7	демонстрация-непонимания	1	4	1	2	1	0	10	6	0	7	1	6	9	0	4
8	я-смущение-фрустрация	27	9	7	17	1	12	6	28	0	3	0	31	34	1	14
9	я-хезитация	1	1	0	1	1	0	0	0	3	1	1	8	18	5	4
10	я-размышление	5	19	6	4	9	15	7	3	1	58	3	20	47	13	7
11	ты-пренебрежение	0	1	2	2	0	1	1	0	1	3	5	9	12	0	1
12	дистанцирование	10	4	8	15	5	4	6	31	8	20	9	8	32	3	17
13	речь-эмфаза	7	10	27	35	0	3	9	34	18	47	12	32	82	2	34
14	стимулирование	0	0	0	0	0	1	0	1	5	13	0	3	2	7	3
15	компенсация-закрытие	10	7	14	12	5	6	4	14	4	7	1	17	34	3	111

функция может выражаться комплексно — через мимику и движение нескольких элементов тела либо редуцированно — с помощью движения всего одним органом [Зинина и др. 2018]. Из таблицы 1 видно, что коммуникативные функции, как правило, выражаются с помощью комплексного паттерна, например функция *компенсация-закрытие* в 111 случаях выполняется двумя или более исполнительными органами². Это означает, что мимическое выражение *компенсации-закрытия* — сжатие губ, облизывание и др. — в реальном поведении сопровождается жестовым выражением — закрывающими жестами, манипулированием, почесываниями и др.

В реальном коммуникативном поведении наблюдаются случаи, когда информант одновременно выражает сразу несколько коммуникативных функций, в этом случае мы говорим об их наложении. Пересечения основных коммуникативных функций приведены в таблице 1. Всего таких пересечений в корпусе насчитывается 1810. При качественном анализе подобных пересечений можно выявить интересные тенденции, например, функция *я-размышление*, как правило, пересекается с такими функциями, как *речь-эмфаза*, *дистанцирование*, *отрицание-несогласие-возражение*, *отсутствие-невозможность*, *стимулирование*. Причем *я-размышление* в таких примерах выражается преимущественно с помощью мимических паттернов, а другие коммуникативные функции — с помощью движений рук, головы и тела. Говорящий при этом может смотреть в сторону, демонстрируя *я-размышление*, но с помощью рук продолжать взаимодействовать с собеседником: демонстрировать отрицание или эмфазу. Функция *компенсация* пересекается со следующими

² Руки рассматриваются здесь как один исполнительный орган.

функциями: *отрицание-несогласие-возражение, я-размышление, понимание-согласие-одобрение, ожидание-обратной-связи, апелляция, я-смущение-фрустрация, дистанцирование и речь-эмфаза*. Говорящий, к примеру, может продемонстрировать отрицание головой и при этом закрываться рукой, как бы компенсируя ущерб социальному лицу адресата в терминах теории вежливости [Brown, Levinson 1987].

Означающее: инварианты коммуникативных действий



Рис. 3,

а. 20081213-ург-а6 00:11.914³

б. 20081213-ург-а6 00:13.134

Означающее лингвистических знаков допускает существенную вариативность: буква может выражаться через множество графических вариантов, а фонема — через множество аллофонов. Означающее жестов еще более разнообразно, однако на этом множестве можно выделить общую основу, которую мы будем называть **инвариантом** жеста. Инвариант — это общий принцип выполнения жеста, который необходим для передачи сообщения адресату и может проявляться в движениях разных исполнительных органов человека.

Рассмотрим понятие инварианта на примере *ожидания-обратной-связи*, где адресант выполнил некоторое действие и надеется получить обратную связь от адресата. Если обратная связь отсутствует, адресант старается подтолкнуть адресата на ответную реакцию. Инвариант действий адресанта можно описать так: адресант намерен

получить от адресата некоторое сообщение для выполнения дальнейших действий и акцентирует это намерение. Этот инвариант может выражаться с помощью четырех групп жестов.

1. Адресант демонстрирует, что воспринимает реакцию адресата визуально — он хочет увидеть, что в ответ сделает адресат. В этом случае адресант может перевести взгляд на адресата в конце своего высказывания. В следующем примере адресант произносит ответ (в это время его голова и взгляд опущены), а затем поднимает голову и смотрит на адресата (рис. 3).³

Говорящий может смотреть в сторону и переводить взгляд на адресата из положения сбоку. Если к началу движения взгляд говорящего уже направлен на адресата, то говорящий вынужден акцентировать свое внимание по-другому. В следующем примере информант демонстративно поднимает брови (рис. 4).

Таким образом, если инвариант жеста не может быть выражен через направление взгляда (взгляд уже направлен на собеседника), адресант пытается выразить

³ Корпус является закрытым, однако в научных целях можно ознакомиться с указанными примерами, если связаться с авторами работы.



Рис. 4

а. 20081230-b4-furekl-m
08:43.630, б. 20081230-b4-
furekl-m 08:44.565



Рис. 5

а. 20080717-f-psy-m 05:15.570,
б. 20080717-f-psy-m 05:01.890



Рис. 6

а. 20080717-c16-psy-m
04:12.920, б. 20080717-c16-
psy-m 04:14.440

инвариант — акцентировать свое намерение — другими средствами, в данном случае — через движение бровей.

Акцентируя свое внимание на реакции собеседника, адресант может «присматриваться» и «прислушиваться»: прищуриваться, хмурить брови, сжимать губы и наклоняться немного вперед (рис. 5, а). Также говорящий может существенным образом сокращать дистанцию с собеседником (рис. 5, б). Мы предполагаем, что мимическое выражение (S — прищуривание глаз), характерное для рассматривания (C) плохо различимого объекта, используется в данном случае как контролируемое действие S' для запроса обратной связи (R) у адресата.

2. В ожидании реакции собеседника говорящий может демонстрировать свое бездействие: 'я сделал все, что нужно, и теперь жду тебя'. К примеру, продолжая рассматривать собеседника, говорящий может перекидывать голову с одного плеча на другое, как бы демонстративно меняя позу при бездействии (рис. 6).

Демонстрируя бездействие, говорящий также может дистанцироваться от адресата (рис. 7).

3. В инварианте указано, что обратная связь необходима адресанту, чтобы перейти к следующему действию. Стремясь продемонстрировать свою готовность, говорящий может останавливать жест в середине выполнения (рис. 8)



Рис. 7

а. 20081220-zhurn-c2 00:51.357,
б. 20081220-zhurn-c2 00:51.357



Рис. 8. 20081219-zhurn-a03
01:45.363



Рис. 9. 20081219-zhurn-a0.10
00:41.065

или «группироваться», например складывать руки (рис. 9), показывая возможность перейти к действию.

4. Акцентируя свое намерение получить обратную связь, адресант может демонстрировать коммуникативные действия, близкие к подтверждению или вопросу. После своего высказывания говорящий может несколько раз кивать вперед или вбок⁴, как бы подтверждая адресату свое высказывание. Говорящий также может выполнять короткие движения головой вверх («кивки» вверх⁵), характерные для вопроса, но в данном случае — запрашивающие реакцию адресата.

Таким образом, анализ размеченных данных позволяет выделить спектр типичных представителей отдельной коммуникативной функции и их инвариант. Такое описание позволяет хранить для жеста несколько вариантов. При синтезе жестов несколько коммуникативных функций могут одновременно выполняться на разных исполнительных органах робота, что существенно разнообразит его коммуникативное поведение.

Система управления роботом Ф-2

В рамках нашей работы мы стремимся сделать робота, коммуникативное поведение которого максимально приближено к коммуникативному поведению человека. Робот должен иметь возможность реализовывать основные особенности естественной коммуникации: выполнять комплексные и редуцированные жесты, гибко комбинировать мимические и жестовые движения, характерные для различных коммуникативных функций, а также фиксировать основные контексты для подбора релевантных ситуации мимических и жестовых реакций.

⁴ Например, 20081219-zhurn-a08 01:34.

⁵ Например, 20081219-zhurn-a09 2:03.

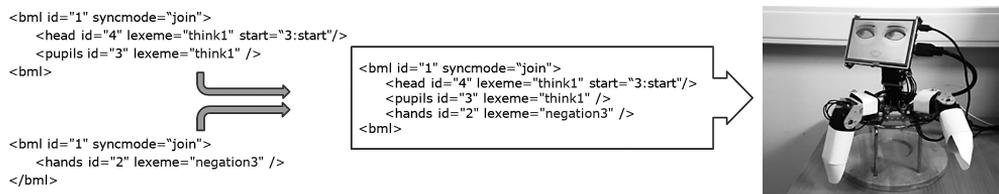


Рис. 10. Комбинирование тегов из различных пакетов BML в режиме join

Проект робота Ф-2 разрабатывается в Курчатовском институте [Kotov et al. 2018; Зинина и др. 2018]. Робот обладает гибкой эмоциональной моделью, основанной на работе 13 негативных сценариев (Опасность, Обман, Неадекватность антагониста, Эмоциональность антагониста и т. д.) и 23 позитивных (Забота, Контроль над ситуацией, Служение и т. д.) [Kotov 2003, 2012]. В составе робота важную роль играет компонент понимания текста — семантический парсер [Kotov et al. 2018]. С помощью парсера восстанавливается семантическая структура (смысл) входящего высказывания, чтобы на ее основе робот смог выбрать ответную коммуникативную реакцию. Кроме того, проект робота Ф-2 включает разработку системы управления коммуникативным поведением для жестовых и мимических реакций робота на входящие тексты. На основании эмоциональной модели строится коммуникативное поведение робота-компаньона.

Для практической реализации этой модели мы разрабатываем такую архитектуру, которая в зависимости от активации определенного сценария (коммуникативной функции) продуцирует коммуникативные реакции, вызывающие BML-пакеты [Kopp, Krenn et al. 2006; Vilhjálmsson, Cantelmo et al. 2007]. В зависимости от активации определенного сценария робот может порождать различные ответы — использовать речь и комбинации движений разных исполнительных органов. Выражение коммуникативной функции снижает активизацию соответствующего сценария, однако, если эта активизация все еще находится выше порогового значения, робот может дополнительно выразить коммуникативную функцию с помощью других элементов поведения.

Для каждой коммуникативной функции в базе хранится несколько BML-пакетов, которые кодируют различные способы ее выражения. На сегодняшний день в базе сохранено 577 BML-пакетов. BML-пакет включает набор тегов, связанных с коммуникативной функцией (рис. 10). С помощью тега описываются исполнительные органы (глаза, голова, речь, левая и правая руки), задействованные в жесте, соответствующем определенной коммуникативной функции.

Робот в своем поведении гибко комбинирует определенные мимические и жестовые средства, что достигается благодаря особой архитектуре. Такая архитектура позволяет по-разному формировать очередь из BML-пакетов за счет того, что у каждого BML-пакета присутствует параметр сочетания с предшествующими пакетами — `syncmode`. Установка значений данного параметра позволяет по-разному группировать BML-пакеты. Например, режим `join` объединяет теги BML-пакета, и они вместе конкурируют в очереди за возможность занять необходимые

исполнительные органы робота. Обработка тегов из пакета начинается, как только освободятся все из указанных в пакете исполнительных органов. При моделировании коммуникативного поведения на роботе-компаньоне выбор определенного способа выражения коммуникативной функции зависит от свободного исполнительного органа, — например, если с помощью рук выражается остановка адресата, то апелляция, скорее всего, будет выражаться головой или мимикой — подобная комбинация коммуникативных функций характерна для естественного поведения.

Заключение

Жесты могут быть описаны как коммуникативные знаки, распределенные от неконтролируемых рефлекторных или эмоциональных реакций (S) до контролируемых действий (S'), ориентированных на определенное воздействие (R). Жесты могут быть ориентированы на (а) выражение внутреннего состояния говорящего, (б) воздействие на адресата, в частности на регулирование ситуации коммуникации, (в) обозначение объекта реального мира или оценку этого объекта и (г) обозначение или оценку предшествующих действий адресата. По этому основанию в архитектуре компьютерного агента жесты можно представить как выражающие определенную коммуникативную функцию. Коммуникативная функция, таким образом, заменяет для жестов традиционное понятие означаемого.

Означающее жестов также имеет сложную структуру. Намеренный жест может использовать означающее, заимствованное у эмоционального бесконтрольного жеста. Жест может проявляться через разные исполнительные органы: брови, голову, руки, тело и т.д. В компьютерной модели целесообразно считать, что жест проявляется через те исполнительные органы, которые не заняты другими жестами и не заблокированы другими ограничителями. Таким образом, жест может исполняться по-разному, однако в означающем жестов можно выделить общий инвариант — при различных типах исполнения жеста говорящий стремится выполнить некоторый общий компонент движения, который объединяет различные варианты исполнения жеста.

Предложенная модель выполнения коммуникативных жестов, основанная на активации коммуникативных функций и использовании инвариантов жестов, позволяет гибко сочетать в поведении различные жесты, использовать разные способы выражения коммуникативных функций в зависимости от свободных исполнительных органов. В целом это ведет к обогащению коммуникативного поведения робота Ф-2, реализующего данную модель. Подробнее познакомиться с архитектурой и функциями робота Ф-2 можно на сайте проекта: <http://f2robot.com/>.

Литература

Григорьева С. А., Григорьев Н. В., Крейдлин Г. Е. Словарь языка русских жестов. М. : Языки русской культуры; Вена : Венский славистический альманах, 2001. 256 с.

Гришина Е. А. Русская жестикация с лингвистической точки зрения. М. : Издательский дом ЯСК, 2017. 744 с.

Зинина А. А., Аринкин Н. А., Зайдельман Л. Я., Котов А. А. Разработка модели коммуникативного поведения робота ф-2 на основе мультимодального корпуса «REC» // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог» (Москва, 30 мая — 2 июня 2018 г.). Вып. 17 (24). 2018. С. 831–844.

Котов А. А., Зинина А. А. Функциональный анализ невербального коммуникативного поведения // Компьютерная лингвистика и интеллектуальные технологии. Вып. 14. Т. 1. М. : РГГУ, 2015. С. 299–310.

Лоренц К. Агрессия (так называемое зло). СПб. : Амфора, 2001.

Beuter N., Spexard T., Lutkebohle I., Peltason J., Kummert F. Where is this? — gesture based multimodal interaction with an anthropomorphic robot: 8th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2008). 2008 P. 585–591. DOI: 10.1109/ICHR.2008.4756009.

Breazeal C. Emotive qualities in lip-synchronized robot speech // Advanced Robotics / eds. R. D. Lane, L. Nadel. 2003. Iss. 2. Vol. 17. P. 97–113.

Breazeal C., Scassellati B. Robots that imitate humans // Trends in Cognitive Sciences. 2002. Vol. 6 (11). P. 481–487.

Brown P., Levinson S. C. Politeness: Some Universals in Language Usage (Studies in Interactional Sociolinguistics). Cambridge, 1987.

Brugman H., Russel A. Annotating Multimedia Multi-modal resources with ELAN: Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004). 2004. P. 2065–2068.

Efron D. Gesture and Environment. N. Y. : King's Crown Press, 1941 (2nd ed. 1972: Gesture, Race and Culture).

Hindriks K., Neerinx M. A., Vink M. The icat as a natural interaction partner: Playing go fish with a Robot // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2012. Vol. 7068 LNAI. P. 212–231.

Jakobson R. Linguistics and poetics // Style in language. MA : MIT Press, 1960. P. 350–377.

Kendon A. How gestures can become like words // Cross-cultural perspectives in non-verbal communication / ed. F. Poyatos. Toronto, 1988. P. 131–141.

Klamer T., Allouch S. B., Heylen D. Adventures of Harvey: Use, acceptance of and relationship building with a social robot in a domestic environment // Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering. 2011. Vol. 59 LNICST. P. 74–82.

Kopp S., Krenn B., Marsella S., Marshall A., Pelachaud C., Pirker H., Thórisson K., Vilhjálmsson H. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language // Intelligent Virtual Agents. 2006. P. 205–217.

Kotov A., Budyanskaya E. The Russian Emotional Corpus: Communication in Natural Emotional Situations // Computational Linguistics and Intellectual Technologies. 2012. Iss. 11 (18). Vol. 1. P. 296–306. RSUH, M.

Kotov A. A., Zaidelman L. Y., Arinkin N. A., Zinina A. A., Filatov A. A. Frames revisited: automatic extraction of semantic patterns from a natural text // Computational Linguistics and Intellectual Technologies. 2018. Iss. 17 (24). P. 357–367.

Leite I., Pereira A., Martinho C., Paiva A. Are Emotional Robots More Fun to Play With? // Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, August 1–3, 2008. P. 77–82.

McNeill D. *Gesture and Thought*. Chicago : University of Chicago Press, 2005. 328 p.

Mori M. The uncanny valley // *Energy*. 1970. Vol. 7 (4). P. 33–35.

Sung J.-Y., Guo L., Grinter R. E., Christensen H. I. “My Roomba Is Rambo”: Intimate Home Appliances // International Conference on Ubiquitous Computing / eds. J. Krumm et al. Berlin, Heidelberg : Springer Berlin Heidelberg, 2007. P. 145–162.

Vilhjálmsson H., Cantelmo N., Cassell J. E., Chafai N., Kipp M., Kopp S., Mancini M., Marsella S., Marshall A., Pelachaud C., Ruttkay Z., Thórisson K., van Welbergen H., van der Werf R. The Behavior Markup Language // *Recent Developments and Challenges in Intelligent Virtual Agents*. 2007. P. 99–111.

A. Zinina, L. Zaydelman, N. Arinkin, A. Kotov

National Research Centre “Kurchatov Institute”

(Russia, Moscow)

zinina_aa@nrcki.ru

APPROACHES TO THE CLASSIFICATION OF COMMUNICATIVE GESTURES: THEORY AND APPLICATION TO COMPANION ROBOTS

The paper presents an approach to the classification of communicative gestures. We describe the gestures such that their signified can be expressed with the help of various signifiers. We classify gestures through the identification of communicative functions (the signified) and an invariant of the performance of the gesture (the signifier). A certain communicative function can be expressed by gestures with different sets of executive organs: with the help of facial expressions, movements of the head, hands or body. By distinguishing communicative functions, we solve the task of modeling complex communicative behavior of a companion robot. In the paper we explain the need of distinguishing communicative functions in the analysis of real human behavior in emotional situations (based on the multimodal corpus REC), and describe the contexts and invariants of communicative functions. Using the case of *the feedback expectation* communicative function we describe the “focus on the interlocutor’s response” invariant and different variations of the corresponding gestures: (1) eye focus on the addressee, (2) demonstration of one’s inaction while waiting for the addressee, (3) demonstration of the readiness to make a move, (4) nods in support of one’s own thesis or the interrogative movements which are requesting the interlocutor’s reaction. We also describe the robot’s control

architecture which simulates the basic features of natural communication — communicative functions and gestures. The description of the invariant and gestures for a certain communicative function allows us to adjust the robot's communicative responses to the incoming statement.

Key words: Multimodal corpus, gestural communication, iconic gestures, robot companions.

References

Beuter N., Spexard T., Lutkebohle I., Peltason J., Kummert F. Where is this? — gesture based multimodal interaction with an anthropomorphic robot. *8th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2008)*, 2008, pp. 585–591. DOI: 10.1109/ICHR.2008.4756009.

Breazeal C. Emotive qualities in lip-synchronized robot speech. *Advanced Robotics*. Eds. R. D. Lane, L. Nadel, 2003, iss. 2, vol. 17, pp. 97–113.

Breazeal C., Scassellati B. Robots that imitate humans. *Trends in Cognitive Sciences*, 2002, vol. 6 (11), pp. 481–487.

Brugman H., Russel A. Annotating Multimedia Multi-modal resources with ELAN. *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004)*, 2004, pp. 2065–2068.

Efron D., van Veen S. Gesture, race and culture, 1972.

Grigor'eva S. A., Grigor'ev N. V., Krejdlin G. E. *Slovar' jazyka russkih zhestov* [A dictionary of the language of Russian gestures]. Moscow-Vienna, Languages of Russian culture Publ.; Viennese Slavic almanac Publ., 2001. 256 p. (In Russ.)

Grishina E. A. *Russkaya zhestikulyatsiya s lingvisticheskoi tochki zreniya (Korpusnye issledovaniya)* [Russian gestures from a linguistic perspective (A collection of corpus studies)]. Moscow, LRC Publishing House, Languages of Slavic Culture Publ., 2017. 744 p. (In Russ.)

Hindriks K., Neerinx M. A., Vink M. The icat as a natural interaction partner: Playing go fish with a Robot. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7068 LNAI, pp. 212–231.

Jakobson R. Linguistics and poetics. *Style in language*. MA, MIT Press, 1960, pp. 350–377.

Kendon A. How gestures can become like words. *Cross-cultural perspectives in non-verbal communication*. Ed. F. Poyatos. Toronto, 1988, pp. 131–141.

Klamer T., Allouch S. B., Heylen D. Adventures of Harvey: Use, acceptance of and relationship building with a social robot in a domestic environment. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, 2011, vol. 59 LNICST, pp. 74–82.

Kopp S., Krenn B., Marsella S., Marshall A., Pelachaud C., Pirker H., Thórisson K., Vilhjálmsón H. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. *Intelligent Virtual Agents*, 2006, pp. 205–217.

Kotov A. A., Zinina A. A. [Functional analysis of nonverbal communicative behavior]. *Computational Linguistics and Intellectual Technologies*. Moscow, RSUH Publ., 2015, iss. 14(1), pp. 299–310. (In Russ.)

Kotov A., Budyanskaya E. The Russian Emotional Corpus: Communication in Natural Emotional Situations. *Computational Linguistics and Intellectual Technologies*, 2012, iss. 11 (18), vol. 1, pp. 296–306. RSUH, M.

Kotov A. A., Zaidelman L. Y., Arinkin N. A., Zinina A. A., Filatov A. A. Frames revisited: automatic extraction of semantic patterns from a natural text. *Computational Linguistics and Intellectual Technologies*, 2018, iss. 17 (24), pp. 357–367.

Leite I., Pereira A., Martinho C., Paiva A. Are Emotional Robots More Fun to Play With? *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, August 1–3, 2008*, pp. 77–82.

Lorenz K. *Agressiya (tak nazываемoe zlo)* [On aggression]. St. Petersburg, Amfora Publ., 2001. (In Russ.)

McNeill D. *Gesture and Thought*. Chicago, University of Chicago Press, 2005. 328 p.

Mori M. The uncanny valley. *Energy*, 1970, vol. 7 (4), pp. 33–35.

Sung J.-Y., Guo L., Grinter R. E., Christensen H. I. “My Roomba Is Rambo”: Intimate Home Appliances. *International Conference on Ubiquitous Computing*. Eds. J. Krumm et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 145–162.

Vilhjálmsón H., Cantelmo N., Cassell J. E., Chafai N., Kipp M., Kopp S., Mancini M., Marsella S., Marshall A., Pelachaud C., Ruttkay Z., Thórisson K., van Welbergen H., van der Werf R. The Behavior Markup Language, Recent Developments and Challenges in Intelligent Virtual Agents, 2007, pp. 99–111.

Zinina A., Arinkin N., Zaydelman L., Kotov A. [Development of communicative behavior model for f-2 robot basing on «REC» multimodal corpora]. *Computational Linguistics and Intellectual Technologies*, 2018, iss. 17 (24), pp. 831–844. (In Russ.)