

УДК: 81.13

РУССКИЙ МУЛЬТИКАНАЛЬНЫЙ ДИСКУРС. ЧАСТЬ II. РАЗРАБОТКА КОРПУСА И НАПРАВЛЕНИЯ ИССЛЕДОВАНИЙ¹

© 2018 г. А. А. Кибрик

*Федеральное государственное бюджетное учреждение науки Институт языкознания РАН;
125009 Москва, Б. Кисловский пер. 1, Россия.*

*Филологический факультет МГУ имени М.В. Ломоносова; 119991, Москва, Ленинские горы, Россия.
Доктор филологических наук, директор Института языкознания РАН. E-mail: aakibrik@gmail.com*

Поступила 29.04.2017

Аннотация. В первой части статьи, опубликованной в номере 1 “Психологического журнала” за 2018 год, было обосновано понятие мультимодального дискурса как нового и перспективного объекта исследования. В отличие от традиционного представления о языке как вербальной структуре, было предложено более объемлющее представление, при котором в язык включается широкий спектр коммуникативных каналов. Вербальный канал и просодический канал, включая многочисленные его компоненты, реализуются в рамках вокальной (слуховой) модальности. Кинетическая (зрительная) модальность представлена в виде большого набора коммуникативных каналов, включая движения глаз, различные типы жестов и проксемику. Были рассмотрены принципы аннотирования основных коммуникативных каналов, включая вербальный, просодический, мануально-жестовый и окуломоторный. В данной (второй) части обсуждается ресурс “Рассказы и разговоры о грушах”, включающий серию записей естественной коммуникации между несколькими участниками, в совокупности с вокальной и кинетической аннотацией этих записей. Ресурс сформирован в рамках комплексного теоретико-методологического подхода к изучению русского мультимодального дискурса. Записи были выполнены при помощи оригинальных технических решений, включающих высококачественную аудиозапись и видеозапись, а также регистрацию движений глаз. На этой основе формулируется программа мультимодальных исследований, включая общие понятия теории коммуникации и проблему взаимодействия между каналами. Дается обзор основных направлений актуальных исследований в области мультимодальной коммуникации. В заключении подводятся итоги обеих частей статьи и формулируются дальнейшие научные перспективы. Подход, представленный в статье, предлагается со стороны лингвистики, но при этом является междисциплинарным и ориентирован на психологическую аудиторию.

Ключевые слова: коммуникация, дискурс, мультимодальность, мультимодальность, просодия, жесты, движения глаз.

DOI: 10.7868/S0205959218020083

СТРУКТУРА МУЛЬТИКАНАЛЬНОГО КОРПУСА

В настоящее время исследовательская группа, базирующаяся в Институте языкознания РАН, создает мультимодальный ресурс (корпус) “Рассказы

и разговоры о грушах”. Этот корпус состоит из отдельных коммуникативных эпизодов (так называемых записей), в каждом из которых участвует по четыре человека. Название корпуса обусловлено тем, что он основан на известном в лингвистике стимульном материале — шестиминутном “Фильме о грушах”, созданном в 1970-е гг. в Беркли исследовательской группой под руководством У. Чейфа [45]. В этом фильме, не включающем никаких вокальных действий персонажей, рассказывается история о взаимодействии ряда лиц, в том числе садовника, собирающего груши, и мальчика, крадущего корзину с грушами. Фильм представляет собой хорошо продуманную цепь физических и социальных событий и давно зарекомендовал

¹ Исследование выполнено при поддержке РНФ, проект № 14-18-03819 “Язык как он есть: русский мультимодальный дискурс”, см. www.multidiscourse.ru. Автор выражает признательность своим коллегам по проекту, совместная работа с которыми позволила сформулировать содержащиеся в этой статье идеи. Особая благодарность О.В. Федоровой, сделавшей ряд ценных замечаний. Автор искренне благодарит Ю.И. Александрову за заинтересованное обсуждение этой статьи, а также анонимных рецензентов журнала, чьи полезные комментарии позволили улучшить текст работы.



Рис. 1. Общий дизайн коммуникативной ситуации.

себя как превосходный способ получать компактные и сравнимые между собой пересказы.

На предварительном этапе фильм просматривается двумя участниками коммуникации, которые в дальнейшем выполняют роли Рассказчика и Комментатора. Затем наступает этап коммуникации как таковой, в этом участвуют Рассказчик, Комментатор и Пересказчик. На рисунке 1 можно видеть общий дизайн ситуации, в которой происходит коммуникация.

Сначала Рассказчик рассказывает Пересказчику содержание фильма в режиме монолога. Затем следует интерактивная фаза, в течение которой Комментатор дополняет или уточняет рассказ, а Пересказчик задает вопросы обоим собеседникам, видевшим фильм. После этого появляется Слушатель, и Пересказчик пересказывает ему фильм, опять в режиме монолога. В конце Слушатель должен письменно записать еще один, вторичный пересказ фильма. Два последних этапа необходимы для того, чтобы мотивировать всех участников к полноценной и осмысленной коммуникации.

Вся коммуникация записывается несколькими устройствами. Во-первых, ведется качественная аудиозапись на многоканальный рекордер, к каждому участнику подведен индивидуальный петличный микрофон. Во-вторых, каждый из трех основных участников снимается при помощи индивидуальной фронтальной камеры. Для этого были использованы промышленные высокоскоростные камеры, снимающие с частотой 100 кадров в секунду и дающие возможность записывать

видеопоток в формате *mjpeg* (без межкадрового сжатия). В-третьих, осуществляется видеозапись общего плана (см. рис. 1). В-четвертых, на Рассказчика и Пересказчика надеты легкие айтрекеры в виде очков (производитель — *Tobii*, Швеция), каждый из которых оборудован двумя камерами: одна из них снимает сцену, на которую смотрит коммуникант, а вторая снимает движения глаз; затем при помощи специального программного обеспечения точка, обозначающая зрительный фокус, накладывается на видеоизображение сцены. Более подробно технологические решения, позволяющие нам осуществлять качественную регистрацию вокального, жестового и окулomotorного поведения, описаны в [14].

Летом 2015 года было осуществлено 24 записи, в которых в общей сложности приняли участие 96 человек от 18 до 36 лет, включая 34 мужчин и 62 женщины. Средняя длительностью одной записи — 23 минуты, с разбросом от 12 до 38 минут. Общая длительность записей составила около 9 часов. В результате последующей обработки каждая запись была представлена в виде комплекта из 15 аудио- и видеофайлов. В ходе этой обработки необходимо было решить большое число технических проблем, в частности связанных с синхронизацией видеозаписей, изначально созданных с различной частотой (что было обусловлено техническими характеристиками использованных устройств). Помимо самих записей, корпус включает аннотации вокальных и кинетических каналов, выполненные при помощи соответствующего программного обеспечения; см. www.multidiscourse.ru.

Channel	Annotation
N-vEDU [30]	N-v269
N-vV [30]	у мальчика на велосипеде была шляпа
N-vPunct [27]	Quest
N-vSegm [81]	мальчи велосипе был шляпа разве
N-vAccents [32]	R
N-vMainAccent [28]	Main
N-vLength [4]	Len
N-vPhon [88]	мал'чик в'эльс'ип был шл'апл раз'в'и
N-mChain [6]	N-mc055
N-mGesture [25]	N-mg316
N-mHandedness [25]	Rt
N-mPhaseStructure	P-S
N-mLtPhases [24]	S
N-mRtPhases [33]	S
N-cGesture [15]	N-cg205
N-cTags [15]	jerk
N-tChain [13]	N-tc108
N-tGesture [17]	N-tg173
N-pState [9]	N-p122
N-pPosition [9]	Towards Commentator
N-fMouthzone [36]	Lips
N-fEyezone [33]	Blinking
N-oFixations [100]	N-o0667
N-oInterlocutor [100]	C
N-oLocus [100]	Face

Рис. 2. Образец мультиканальной аннотации.

Аннотация различных каналов может выполняться в отдельных средах, а может и в единой среде. На рис. 2 представлена объединенная мультиканальная аннотация с большим количеством слоев, выполненная в среде ELAN (tla.mpi.nl/tools/tla-tools/elan).

Группа слоев N-vEDU содержит информацию о структуре элементарных дискурсивных единиц (ЭДЕ) и их свойствах [4]. Группа N-vSegm включает разметку отдельных слов и их признаков. Слой N-mChain и группа N-mGesture описывают мануальные жесты [5]. Слои N-cGesture, N-tGesture и N-pState содержат сведения о жестах головы, жестах тела и проксемике соответственно. Слои N-fMouthzone и N-fEyezone связаны с мимикой говорящего, а группа N-oFixations – с окулomotorным поведением. Эта многослойная аннотация показывает, сколь много различных типов осмысленного поведения одновременно выполняет человек, находящийся в режиме коммуникации с другим. Фактически говорящий человек подобен симфоническому оркестру, в котором одновременно играет множество различных инструментов, каждый со своей партией, и только учет всех этих инструментов одновременно дает представление о целом. Отмечу, что в аннотации на рис. 2 отражены коммуникативные действия лишь одного собеседника – Рассказчика (префикс N, т.е. Narrator). Общая картина носит значительно более сложный характер.

На опыте, полученном при сборе корпуса “Рассказы и разговоры о грушах”, будет базироваться следующая стадия исследования, в ходе которой мы намерены записывать и анализировать обычные бытовые диалоги с неподготовленным окружением (например, группа людей за обеденным столом) и с неструктурированными коммуникативными ролями. Этап не полностью естественного взаимодействия в рамках “Рассказов и разговоров о грушах” необходим для отладки методологии и технологии сбора данных.

В мировой науке существует ряд мультимедийных ресурсов, от которых мы отталкивались при формировании собственного корпуса. Среди русскоязычных ресурсов следует особенно отметить корпус МУРКО, созданный Е.А. Гришиной (www.ruscopora.ru/murco-authors.html; [3]), и Русский эмоциональный корпус А.А. Котова (www.harpia.ru/rec; [7]). Информацию о корпусах речи на других языках можно найти в работах [18; 26; 40; 42], а обзоры существующих корпусов – в работах [15; 16; 33], а также на сайтах www.multimodal-corpora.org, emotion-research.net, sspnet.eu. К числу главных особенностей нашего ресурса, отличающего его от прочих, относятся в первую очередь: большое количество учитываемых явлений (включая просодию и глазодвигательное поведение); высокие характеристики видеозаписи; единая системная методология аннотации данных, относящихся к различным коммуникативным каналам; многоцелевой характер создаваемого

корпуса; оригинальный теоретический подход, ранее выработанный применительно к вокальной модальности.

НАПРАВЛЕНИЯ ИССЛЕДОВАНИЙ В РАМКАХ МУЛЬТИКАНАЛЬНОГО ПОДХОДА

В этом разделе я остановлюсь на некоторых направлениях исследований, которые представляются особенно актуальными в контексте мультимодального подхода к коммуникации — от сугубо теоретических до конкретно-прикладных.

Теория языка и теория коммуникации. Теоретические работы в области человеческого языка и даже в области коммуникации чаще всего исходят из униканального подхода: предполагается, что в ходе коммуникации люди всего лишь обмениваются вербальными структурами. Иногда звучащая речь рассматривается более комплексно, с включением и просодического канала, но сопровождающие речь кинетические сигналы во внимание, как правило, не принимаются. Переход к бимодальной и мультимодальной перспективе вынуждает пересмотреть обычные представления в этой области.

Разграничение между порождением (говорением) и пониманием (слушанием). Стандартные представления о языковой коммуникации предполагают четкое разграничение двух процессов — порождения (говорения) и понимания (слушания). Соответственно различаются базовые коммуникативные роли говорящего (адресанта) и слушающего (адресата). Согласно этим представлениям, в каждый момент времени тот или иной участник коммуникации является либо говорящим, либо слушающим. При мультимодальном подходе это противопоставление неизбежно становится менее жестким. Адресант, порождающий речь в данный момент, одновременно отслеживает кинетическое поведение слушающего (жесты, направление взора), то есть в некотором смысле является также и адресатом, а слушающий, выполняя такое кинетическое поведение, отчасти выполняет роль адресанта. Таким образом, коммуникативные действия участников образуют сложную и неоднородную сеть, в которой необходимо признать возможность одновременной передачи значимого сигнала в разные стороны.

Чередование реплик. В связи с предыдущим, традиционное униканальное (и унимодальное) представление о речи в качестве одного из базовых понятий использует понятие реплики (*turn*), см. обзор в [17, с. 577–589]. Изредка реплики

могут накладываться, но нормальной предполагается ситуация, в которой за репликой *n* участника А следует реплика *n + 1* участника Б. При мультимодальном подходе дискретное понятие реплики в значительной мере размывается. Оно должно быть заменено более общим понятием коммуникативного действия, и разные участники коммуникации могут выполнять разные действия одновременно.

Паузация. Еще одно ключевое понятие в унимодальном подходе к речи — пауза. Это интервал между периодами вокализации, отсутствие речевого действия [10, раздел 4.4]. Но если в течение такого периода происходит кинетическое коммуникативное поведение того или иного участника, то это всего лишь пауза в вокализации, но не в коммуникации.

Элементарная дискурсивная единица. В первой части статьи [5] было разъяснено понятие элементарной дискурсивной единицы (ЭДЕ) — кванта или минимального шага в порождении вокального сигнала. В случае принятия мультимодального подхода это понятие также требует пересмотра, ведь дискурс реализуется не только в виде вокальных, но и в виде кинетических действий. Значимость ЭДЕ как базового блока вербально-просодической структуры вряд ли подлежит сомнению, но было бы более правильно говорить об элементарных вокальных единицах. Или, если их выделение основывается на чисто просодических основаниях, о просодических единицах.

Статика и динамика. При разработке системы аннотации мануальных жестов нами был сделан вывод о том, что помимо динамических событий (собственно жестов) кинетический сигнал содержит также статические ситуации — положения, или позы [8; 13]. Говорящий на некоторое время принимает ту или иную позу рук (например, руки на коленях; руки у груди; руки на весу), и динамические жесты производятся из этой позы. В случае мануального компонента роль поз относительно скромна, так как жесты нагляднее поз, а руки являются высокоактивным органом. В других случаях соотношение иное. Так, жесты тела сравнительно немногочисленны и ограничены, а роль статических поз существенно возрастает. Важность различия между статикой и динамикой в кинетических коммуникативных каналах подталкивает к тому, чтобы поискать аналоги в каналах вокальных. В частности, говорящие на протяжении некоторых временных отрезков могут производить вокализацию в сниженном или повышенном тональном регистре, и локальные тональные акценты на этих участках реализуются от маркированной

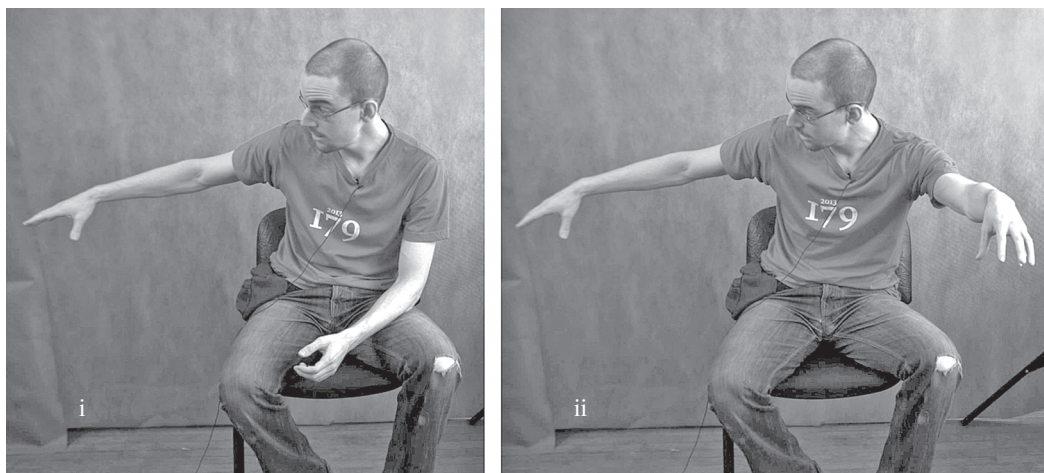


Рис. 3. Жесты, сопровождающие речь из фрагмента (2).

точки отсчета. Нередко встречаются особые фонации, например “скрипучий” голос, который представляет собой своеобразную артикуляционную позу, в рамках которой как сегментные, так и просодические единицы реализуются специфическим образом.

Индивидуальные различия. На протяжении последних ста лет в теоретической лингвистике популярно противопоставление между языком (жесткой системой, общей для всех носителей языка) и речью (реализацией языковых умений в конкретных случаях). Сейчас уже понятно, что вариативность и индивидуальные различия проявляются и в вербальном канале [1; 22]. А чем больше мы расширяем границы языка, тем более заметна вариативность используемых разными людьми сигналов. Так, в области русской просодии выделяется противопоставление двух разных интонационных фигур — финальное падение (так называемая интонация точки) и нефинальное падение (запятая с падением) [4]. Однако реализация этого противопоставления зависит от индивидуальных характеристик частотного диапазона голоса у конкретных говорящих, и для его осмысленного применения необходимо построение просодического портрета говорящего. Далее, в нашей работе по исследованию сопровождающей речь жестикуляции были отмечены большие индивидуальные различия и в частотности жестов, и в их качественных характеристиках (степень оформленности движения, темп и т.д.), и во внутренней структуре (например, использование удержаний) [8]. По-видимому, в дальнейшем придется ввести понятие жестикуляционного портрета коммуниканта. Одним из элементов такого портрета может стать нейтральное положение рук, предпочитаемое данным коммуникантом. Создание жестикуляционных

портретов позволит более адекватно интерпретировать конкретные жесты, выполняемые данным коммуникантом.

Взаимодействие между каналами. Коль скоро коммуникация носит мультимедийный характер, встает важнейший вопрос о том, как различные каналы, одновременно используемые коммуникантом, взаимодействуют между собой. Здесь есть ряд различных аспектов, в том числе семантическое взаимодействие, координация во времени, а также вопрос о наличии единого управляющего центра, из которого порождаются единицы разных каналов.

Смысловая кооперация между каналами. Как только мы начинаем обращать внимание на сопровождающую речь жестикуляцию, становится очевидно, что смысл передается не только при помощи слов, но и при помощи жестов. Рассмотрим пример, позволяющий продемонстрировать кооперацию между мануальными жестами и вербальным компонентом в конструировании смысла. Участник разговора произнес фрагмент (2), состоящий из двух ЭДЕ, а также одновременно выполнил два жеста, показанные на рис. 3.

(2) i. (0.20) у него есть багажник /сзади,

ii. (0.40) и у него есть багажник \спереди.

Жест (i) скоординирован по времени с ЭДЕ (i) (точнее, он начинается через 30 мс после ее начала), а жест (ii) практически совпадает по времени с ЭДЕ (ii) (отмечу, что жест (i) удерживается говорящим в течение всего времени, пока он производит ЭДЕ (ii) и выполняет жест (ii)). Жесты позволяют рассказчику описать смысловой компонент, который никак не содержится в речи. А именно, при помощи жестов рассказчик сообщает о том, как велосипед был ориентирован

в пространстве. В пересказываемом фильме велосипед был направлен в левую сторону; соответственно, местоположение заднего багажника было показано при помощи правой руки, а переднего — при помощи левой руки. Таким образом, пространственная семантика распределенно выражается словами и жестами. Вопрос о том, как может осуществляться подобная кооперация, представляет собой большое поле исследований (см. работы [17, с. 182–202]; [17, с. 689–707], а также [23; 29; 39]). В серии наших работ [6] было показано, что вербальный, просодический и визуальный компоненты информационно нагружены в сравнимой степени, и каждый из них несет существенную долю сведений, передаваемых в естественном дискурсе.

Сравнимость единиц кинетического и вокального поведения. Только что рассмотренный пример подводит нас к следующему вопросу. Что является аналогом (“дублером”, “компаньоном”, англ. *affiliate*) жеста (i) с рисунка 3 в вокальном сигнале — слово *багажник*, или словосочетание *багажник сзади*, или вся клауза *у него есть багажник сзади*? Или, может быть, более корректно считать, что аналогом жеста является не собственно вербальная единица, а ЭДЕ? Понятно, что этот вопрос имеет общий характер и может быть задан относительно любого жеста и его возможных аналогов. Наиболее общий ответ на этот вопрос таков: возможны самые разные варианты. Указательные жесты по своей функции ближе всего к именным группам вербальной структуры; и те, и другие выполняют функцию референции, то есть указания на лица или предметы [30, ch.15]. Жесты-удары в значительной мере скоординированы с просодическими выделениями — акцентами [20]. А изобразительные жесты, такие как на рисунке 3, описывают ситуации, и в этом смысле аналогичны клаузам (которые в свою очередь коррелируют с ЭДЕ). Жестовые цепочки являются аналогами предложений или даже эпизодов [17, с. 1098–1125], то есть групп ситуаций. Такова картина в первом приближении, и она должна быть конкретизирована в дальнейших исследованиях.

Временная координация единиц разных каналов. В работе [14] мы исследовали временную координацию единиц поведения, относящихся к разным каналам (вербальному, просодическому, жестовому). В качестве базовой единицы вербального канала были использованы клаузы, просодического — ЭДЕ, мануально-жестового — отдельные жесты (*G-phrases* в терминах Kendon [28; 29]). Согласно широко распространенной точке зрения, жесты производятся несколько раньше, чем соответствующие им

вокальные единицы [28; 41; 43]. В качестве возможной причины такого опережения иногда выдвигается идея о том, что подготовка моторного акта занимает меньше времени, чем концептуальная обработка, необходимая для производства вербальной структуры [46]. Опережающий характер жестов не раз был показан на материале английской речи [37; 38], а также и других языков, например польского [26]. Мы рассмотрели этот вопрос на русском материале, опираясь на подкорпус из четырех 6-минутных фрагментов с разными наборами участников. Аннотация вокального и жестового материала была произведена независимо, в каждом из этих типов аннотации были выделены элементарные единицы (ЭДЕ и жесты, соответственно), с точностью до сотой доли секунды были определены их границы. Замечу, что практическое выделение единиц принципиально зависит от ряда методологических решений, в частности, от того, включается ли в состав ЭДЕ начальная пауза, и от того, рассматривается ли жест целиком или только в объеме его функционально центральной части (маха). Мы разделили жесты на девять логических типов в зависимости от взаимного расположения левой и правой границ жеста и соответствующей ЭДЕ, с точностью измерения 200 мс, 100 мс и 50 мс. При точности 200 мс лишь 35% жестов оказались опережающими, а при точности 100 мс и 50 мс размер этой группы увеличился, но все равно не достиг 50%. В целом, на нашем материале гипотеза об опережающем производстве жестов не подтвердилась [14]. Вопрос о координации разных коммуникативных каналов оживленно обсуждается в современной литературе [46]. Но конкретные исследования в этой области, как правило, основаны на небольших данных, специально собранных для изучения этого вопроса. До сих пор готового многоцелевого ресурса, который можно было бы использовать для этой задачи, в распоряжении специалистов не было. Создаваемый нашим коллективом ресурс, вероятно, может оказаться весьма полезен для эмпирически обоснованного изучения широкого круга вопросов, касающихся координации между каналами.

Существует ли единый управляющий источник речи и жестов? Есть серьезные основания подозревать, что такой единый управляющий центр существует, причем не только при условии четкой синхронизации жестов и речевых фрагментов, но и в том случае, если жесты стандартно имеют опережающий характер [35]. Теория “точки роста” Д. МакНилла предполагает единый общий источник жестов и речи [38]. В работе [41] выдвигается идея о едином досемантическом центре, связанном с коммуникативными намерениями. В работе [34] была предложена когнитивная модель мультимодального речепроизводства,

предполагающая формирование единых концептуализаций в рабочей памяти. Иногда выдвигается единая теория моторной активности, от целых ручных жестов до отдельных фонетических сегментов [17, с. 785–792]. Некоторые авторы склоняются к раздельному планированию жестов и речи, как моторных и пропозициональных действий соответственно [32]. Обзор различных гипотез в этой области представлен в унифицированном формате в статье [46]. Есть основания полагать, что объемный мультимодальный корпус может вывести суждение этой проблемы на новый уровень.

Просодия как интерфейс между языком в узком смысле и жестами. Исследователи мультимодальной коммуникации [38] нередко отмечают, что вербальный сигнал (язык в узком смысле) представляет собой дискретную структуру, в отличие от жестов, которые менее дискретны как в смысле своего репертуара, так и в смысле своих границ в потоке. Следует сказать, что дискретность вербальной структуры традиционно сильно преувеличена [31]. Но еще важнее другое — существует важный интерфейс между вербальной и кинетической структурой, а именно просодия. Просодия непосредственно сопрягается с вербальной структурой, фактически накладываясь на нее, и при этом имеет много сущностных сходств с жестикуляцией. Жесты-удары (*beats*) сильно напоминают дискурсивные акценты, ритмически организующие речь. Как просодические, так и жестовые средства часто служат выделению наиболее существенных информационных компонентов [44]. При исследовании жестов оказываются важными те же физические характеристики, что давно используются в просодическом анализе, в том числе скорость, ускорение, интенсивность. Как отмечалось выше, в жестикуляции, подобно просодии, наблюдается высокая степень индивидуального варьирования, и стоит задача создания жестикуляционных портретов. Детальное понимание сходств (и различий) между просодией и жестикуляцией еще в будущем, однако очевидно, что совместное исследование жестов и речи будет продуктивным лишь с учетом просодии.

Связь между направлением взгляда и речью. Еще в 1960-е гг., когда современные средства регистрации мультимодального поведения еще не были доступны, А. Кендон выдвинул ряд тезисов о зависимости зрительных фиксаций от коммуникативной роли говорящего [27]. В частности, он утверждал, что слушающий смотрит на собеседника чаще, чем говорящий, и что фиксации слушающего на собеседнике длиннее, чем у говорящего. Исследование О.В. Федоровой, основанное на

анализе “Рассказов и разговорах о грушах”, подтвердили первое, но не второе утверждение [12]. Более того, в этом исследовании предложено различие между несколькими типами стратегий окулomotorного поведения. Во-первых, существуют общие стратегии: так, фиксации на лице на порядок длиннее фиксаций на руках. Во-вторых, есть контекстно-зависимые стратегии: например, при интерактивной коммуникации фиксации на окружении встречаются реже, чем при режиме монолога. В-третьих, имеются индивидуальные стратегии: в частности, доля фиксаций на руках собеседника различается между испытуемыми в несколько раз. Эти наблюдения возвращают нас в том числе и к базовым вопросам теории коммуникации, затронутым в предыдущем разделе.

Происхождение языка. Происхождение человеческой коммуникации и языка сейчас является одной из “горячих” междисциплинарных тем [2]. Неоднократно высказывалось предположение о том, что вокальной речи эволюционно предшествовала жестовая речь. Сто лет назад русский лингвист А.Л. Погодин писал: “Человеку, как говорящему существу, должен был предшествовать человек, как существо жестикулирующее, пользующееся мимикой или простейшими знаками для передачи своих несложных чувств и требований” [9, с. 121]. В настоящее время одним из наиболее активных сторонников идеи о первичности жестовой коммуникации является Майкл Томаселло, исследовавший коммуникативные процессы как в филогенезе (у высших обезьян), так и в онтогенезе (детская речь). По словам Томаселло, “вокальные конвенции изначально приобрели коммуникативное значение только за счет того, что сопровождали в качестве факультативного приложения естественные осмысленные жесты” [11, с. 269–270]. Проблему эволюционного происхождения коммуникации и языка вряд ли можно считать решенной (см. работы [17, с. 466–479]; [17, с. 480–512], а также [24, 36]). Для ее решения, опять же, необходим очень серьезный учет несегментной фонетики, т.е. просодии, которая представляет собой мост между вербальным и кинетическим поведением.

Автоматическая аннотация. Особняком от выше рассмотренных теоретических проблем стоит практическая, но также важная проблема автоматической аннотации мультимодального дискурса. Это касается и распознавания звучащей речи [25], и распознавания кинетического поведения [17, с. 857–868]. Создание мультимодальных корпусов — столь трудоемкая задача, что ее следовало бы хотя бы отчасти автоматизировать. Вручную

можно создавать лишь ограниченные по объему ресурсы. Хотя усилия в направлении автоматизации прилагаются уже давно, эффективных и доступных методов, сравнимых по качеству с ручной аннотацией, пока нет. Есть полезные вспомогательные решения. Например, в нашей работе в области вокальной аннотации мы сотрудничаем с компанией СТЭЛ (www.stel.ru), что позволяет находить точные временные координаты словоупотреблений в потоке речи (см. рис. 2, слой N-vSegm), на основе заранее сделанного вручную транскрипта. В дальнейшем это позволит исследовать координацию жестов и речи на уровне отдельных словоупотреблений. Что касается жестов и других видов кинетического поведения, то вопрос об их автоматической или полуавтоматической разметке также стоит в науке не первый год. Распространенные в настоящее время методы распознавания кинетического поведения (так называемый захват движения) чаще всего основаны на “инвазивных” технологиях, таких как прикрепление специальных маркеров к частям тела или одежде говорящего. Развиваются и другие технологии, такие как чувствительная к глубине видеозапись и распознавание жестов на основе инфракрасных лучей. Актуальна задача идентификации движений на основе ранее сделанной видеозаписи [21], но она, по-видимому, не имеет еще убедительного решения. Массовое исследование мультимедийной коммуникации зависит от решения данной технической задачи.

ЗАКЛЮЧЕНИЕ

Мультимедийный (мультимодальный) подход, сложившийся к настоящему времени, — это результат междисциплинарного взаимодействия. В исследовании реальной человеческой коммуникации участвуют не только лингвисты, но и психологи, нейрофизиологи, специалисты по искусственному интеллекту и по разным видам институционального, в том числе делового, общения. В двух статьях данного цикла я попытался представить проблему мультимедийности с лингвистической точки зрения.

В первой статье были определены теоретические понятия мультимодальности и мультимедийности. Вербальный канал и просодический канал (включая многочисленные его компоненты) реализуются в рамках вокальной (слуховой) модальности. Кинетическая (зрительная) модальность представлена в виде большого набора коммуникативных каналов, включая движения глаз, различные типы жестов и проксемику. В той же статье была

относительно подробно рассмотрена организация вербального, просодического, мануально-жестового и окулomotorного каналов. На этой основе был предложен существенно иной, по сравнению с традиционным, взгляд на человеческий язык. Узкий уникальный взгляд на язык как на набор сегментных единиц, до сих пор присущий лингвистике, неизбежно должен смениться более широким. Язык в широком смысле слова включает все каналы, которые используются людьми в реальной жизни для того, чтобы сообщить свои мысли, интенции, эмоции партнерам по коммуникации.

В настоящей статье описан корпус, ориентированный на одновременный учет максимального числа явлений мультимедийной коммуникации. Кроме того, представлены направления исследований, которые можно проводить на мультимедийных корпусах. Мультимедийный взгляд на язык влечет за собой масштабные изменения как в теории, так и в практике исследований. Если теория языка признает принципиальное равноправие различных средств, вокальных и кинетических, служащих коммуникативному обмену между людьми, то вслед за этим приходится пересмотреть такие базовые понятия, как порождение, понимание, говорящий, адресат, реплика, пауза. Кроме того, при мультимедийной перспективе еще более очевидной становится фундаментальная вариативность языка и языкового поведения. Открывается широкий круг исследовательских вопросов, связанных с взаимодействием коммуникативных каналов, включая распределение информации между ними, сравнимость единиц разных каналов, временную координацию. Все это существенно меняет общий ландшафт науки о языке.

Я хотел бы закончить упоминанием важного коммуникативного явления, обсуждению которого выше не нашлось места. Это явление — смех. Один из самых глубоких современных исследователей дискурса, американский лингвист Уоллес Чейф, посвятил смеху целую книгу [19], в которой он описал вокальные характеристики этого явления, взаимодействие смеха с речью, а также лежащую в основе смеха эмоцию “несерьезности”. Интересно, что на сайте sspnet.eu, посвященном изучению разнообразных “социальных сигналов”, на главной странице упоминается мультимедийный феномен смеха: “смех — это сложный сигнал, в состав которого входят изменения в мимике, вокализации, позе и движениях субъекта”. Похоже, что даже феномен смеха, обычно не связываемый с сущностью языка, в перспективе должен быть объединен с языковой коммуникацией единой теорией.

СПИСОК ЛИТЕРАТУРЫ

1. *Беликов В.И., Крысин Л.П.* Социоллингвистика. Учебник для бакалавриата и магистратуры. 2-е изд. М.: Юрайт, 2015.
2. *Бурлак С.А.* Происхождение языка: Факты, исследования, гипотезы. М.: Corpus, 2011.
3. *Гришина Е.А.* Русская жестикуляция с лингвистической точки зрения (корпусные исследования). М.: ЯСК, 2017.
4. *Кибрик А.А.* Когнитивный анализ дискурса: локальная структура // Язык и мысль: современная когнитивная лингвистика / Под ред. А.А. Кибрика, А.Д. Кошелева, А.В. Кравченко, Ю.В. Мазуровой, О.В. Федоровой. М.: ЯСК, 2015. С. 595–634.
5. *Кибрик А.А.* Русский мультимодальный дискурс. Часть I. Постановка проблемы // Психологический журнал. 2018. Т. 39. № 1. С. 71–81.
6. *Кибрик А.А., Молчанова Н.Б.* Каналы мультимодальной коммуникации: относительный вклад в понимание дискурса // Мультимодальная коммуникация: теоретические и эмпирические исследования: материалы семинара. Москва, 15 ноября 2013 г. / Под ред. О.В. Федоровой, А.А. Кибрика. М.: Буки Веди, 2014. С. 99–114.
7. *Котов А.А., Зинина А.А.* Функциональный анализ невербального коммуникативного поведения // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции “Диалог”. М.: РГГУ, 2015. С. 308–320.
8. *Литвиненко А.О., Николаева Ю.В., Кибрик А.А.* Аннотирование русских мануальных жестов: теоретические и практические вопросы // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции “Диалог” (Москва, 31 мая – 4 июня 2017 г.). Вып. 16 (23). М.: Изд-во РГГУ, 2017. С. 271–286.
9. *Погодин А.Л.* Язык как творчество (психологические и социальные основы творчества речи): Происхождение языка. Харьков, 1913.
10. Рассказы о сновидениях: корпусное исследование устного русского дискурса / Под ред. Кибрика А.А., Подлесской В.И. М.: ЯСК, 2009.
11. *Томаселло М.* Истоки человеческого общения. М.: ЯСК, 2011.
12. *Федорова О.В.* Распределение зрительного внимания собеседников в естественной коммуникации: 50 лет спустя // Когнитивная наука в Москве: новые исследования. Материалы конференции 15 июня 2017 г. / Под ред. Е.В. Печенковой, М.В. Фаликман. М.: БукиВеди, ИППИП. С. 370–375.
13. *Федорова О.В.* Вклад кинетической составляющей в мультимодальную коммуникацию, или Еще раз о стратегиях описания танграмм // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции “Диалог” (Москва, 31 мая – 4 июня 2017 г.). Вып. 16 (23). М.: Изд-во РГГУ, 2017. Т. 2. С. 118–133.
14. *Федорова О.В., Кибрик А.А., Коротаев Н.А., Литвиненко А.О., Николаева Ю.В.* Временная координация между жестовыми и речевыми единицами в мультимодальной коммуникации // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции “Диалог” (Москва, 1–4 июня 2016 г.). Вып. 15 (22). М.: Изд-во РГГУ, 2016. С. 159–170.
15. *Abuczki A., Esfandiari B.G.* An overview of multimodal corpora, annotation tools and schemes // Argumentum. 2013. V. 9. P. 86–98.
16. *Allwood J.* Multimodal Corpora // Corpus linguistics. An international handbook / Eds. Lüdeling A., Kytö M. Berlin: Mouton de Gruyter, 2008. P. 207–225.
17. *Body–language–communication: An international handbook on multimodality in human interaction / Eds. Müller C., Cienki A., Fricke E., Ladewig S.H., McNeill D., Teßendorf S.* Berlin: Mouton de Gruyter, 2013. V. 1, 2.
18. *Brône G., Oben B.* InSight interaction. A multimodal and multifocal dialogue corpus // Language resources and evaluation. 2015. V. 49 (1). P. 195–214.
19. *Chafe W.* The importance of not being earnest: The feeling behind laughter and humor. Amsterdam: Benjamins, 2007.
20. *Fernández-Baena A., Montaña R., Antonijoan M., Roversi A., Miralles D., Alias F.* Gesture synthesis adapted to speech emphasis // Speech communication. 2014. V. 57. P. 331–350.
21. *Gairola P., Kumar S.* Hand gesture recognition from video // International journal of science and research. 2014. V. 3. Is. 4. P. 154–158.
22. *Geeraerts D., Kristiansen G.* Variationist linguistics // Handbook of Cognitive Linguistics / Eds. E. Dąbrowska, D. Divjak. Berlin: de Gruyter Mouton, 2015. P. 366–389.
23. *Goldin-Meadow S.* The role of gesture in communication and thinking // Trends in cognitive sciences. 1999. V. 3 (11). P. 419–429.
24. *Goldin-Meadow S.* What modern-day gesture can tell us about language evolution // The Oxford handbook of language evolution / Eds. M. Tallerman, K.R. Gibson. Oxford: Oxford University Press, 2012. P. 545–557.
25. *Huang X., Baker J., Reddy R.* A historical perspective of speech recognition // Communications of the ACM. 2014. V. 57. № 1. P. 94–103.
26. *Karpiński M., Jarmołowicz-Nowikow E., Malisz Z.* Aspects of gestural and prosodic structure of multimodal utterances in Polish task-oriented dialogues // Speech and language technology. 2009. V. 11. P. 113–122.
27. *Kendon A.* Some functions of gaze direction in social interaction // Acta psychologica. 1967. V. 26. P. 22–63.

28. *Kendon A.* Gesticulation and speech: Two aspects of the process of utterance // *The relationship of verbal and nonverbal communication* / Ed. M.R. Key. Berlin: Walter de Gruyter, 1980. P. 207–227.
29. *Kendon A.* *Gesture. Visible action as utterance.* Cambridge: Cambridge university press, 2004.
30. *Kibrik A.A.* Reference in discourse. Oxford: Oxford university press, 2011.
31. *Kibrik A.A.* The problem of non-discreteness and spoken discourse structure // *Computational linguistics and intellectual technologies.* 2015. V. 14 (21). P. 225–233.
32. *Kita S., Özyürek A.* What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking // *Journal of memory and language.* 2003. V. 48 (1). P. 16–32.
33. *Knight D.* The future of multimodal corpora // *Revista brasileira de linguística aplicada.* 2011. V. 2. P. 391–415.
34. *Kopp S., Bergmann K., Kahl S.* A spreading activation model of the semantic coordination of speech and gesture // *Proceedings of the 35th annual meeting of the Cognitive science society.* Berlin, 2013.
35. *Leonard Th., Cummins F.* The temporal relation between beat gestures and speech // *Language and cognitive processes.* 2010. V. 26 (10). P. 1457–1471.
36. *Levinson S.C., Holler J.* The origin of human multimodal communication // *Philosophical transactions of the Royal society B.* 2014. V. 369. P. 20130302.
37. *Loehr D.* Temporal, structural, and pragmatic synchrony between intonation and gesture // *Laboratory phonology.* 2012. V. 3 (1). P. 71–89.
38. *McNeill D.* *Hand and mind: what gestures reveal about thought.* Chicago: University of Chicago Press, 1992.
39. *Melinger A., Levelt W.J.M.* Gesture and the communicative intention of the speaker // *Gesture.* 2004. V. 4.2. P. 119–141.
40. *Mondada L.* Bodies in action // *Language and dialogue.* 2014. V. 4 (3). P. 357–403.
41. *Morrel-Samuels P., Krauss R.M.* Word familiarity predicts temporal asynchrony of hand gestures and speech // *Journal of experimental psychology: human learning and memory.* 1992. V. 18 (3). P. 615–622.
42. *Multimodal corpora: From models of natural interaction to systems and applications* // Eds. Kipp M., Martin J.-C., Paggio P., Heylen D. Berlin, Heidelberg: Springer, 2009.
43. *Schegloff E.A.* On some gestures' relation to talk // *Structures of Social Action* / Eds. Atkinson J.M., Heritage J. Cambridge: Cambridge university press, 1984. P. 266–298.
44. *Swerts M., Krahmer E.* 2008. Facial expressions and prosodic prominence: comparing modalities and facial areas // *Journal of phonetics.* 2008. V. 36 (2). P. 219–238.
45. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* / Ed. Chafe W.L. Norwood, NJ: Ablex, 1980.
46. *Wagner P., Malisz Z., Kopp S.* Gesture and speech in interaction: An overview // *Speech Communication.* 2014. V. 57. P. 209–232.

RUSSIAN MULTICHANNEL DISCOURSE. PART II. A CORPUS DEVELOPMENT AND AVENUES OF RESEARCH¹

A. A. Kibrik

Federal State Institution of Science Institute of Linguistics of the Russian Academy of Sciences; 125009 Moscow, B. Kislovskij per. 1, Russia.

*Philological faculty, Lomonosov Moscow State University; 119991, Moscow, Leninskie gory, Russia.
Ph.D., Director of the Institute of Linguistics RAS. E-mail: aakibrik@gmail.com*

Received 29.04.2017

Abstract. In the first part of this paper, published in issue 1, 2018, the notion of multichannel discourse was introduced as a novel and promising object of investigation. In contrast to the traditional view of language as verbal structure, a more encompassing view was proposed, in which language involves a wide range of communication channels. The verbal channel and the prosodic channel, the latter including numerous components, are realized in the vocal (auditory) modality. The kinetic (visual) modality is represented as a

¹ This study was supported by RSF, project No. 14-18-03819 “Language as it is: Russian multimodal discourse”, see www.multidiscourse.ru. The author expresses gratitude to his coworkers in the project, the cooperative work with whom allowed him to formulate the ideas contained in this article. Special gratitudes go to Olga V. Fedorova, who made a number of valuable comments. The author sincerely thanks Yuri I. Alexandrov for an interested discussion of my research, as well as anonymous reviewers of the journal, whose useful comments allowed to improve the text of this paper.

large set of communication channels, including eye movements, various kinds of gestures, and proxemics. The principles of annotating the major communication channels were considered, including the verbal channel, the manual gestures, and the oculomotor channel. In the present part of the paper the resource “Russian Pear Chats and Stories” is discussed. The resource consists of the recordings of natural communication between several participants, along with the vocal and kinetic annotation. The resource has been created within the framework of a integrated theoretical and methodological approach towards multichannel discourse. The recordings were made with the help of original technical solutions, including high quality audio and video capture and eye tracking. This forms a basis for a program of multichannel studies, including the general notions of communication theory and the problem of interaction between channels. A review is provided of the main avenues of contemporary research in the domain of multichannel communication. The concluding section summarizes the findings of both parts of the paper and suggests the future research perspectives. The approach taken in this paper is proposed from the side of linguistics, but at the same time is of an interdisciplinary nature and is addressed to psychological readership.

Keywords: communication, discourse, multimodality, multichannel studies, prosody, gesticulation, eye movements, psychology of speech.

REFERENCES

1. *Belikov V.I., Krysin L.P.* Sociolingvistika. Uchebnik dlja bakalavriata i magistratury. 2-e izd. Moscow: Jurajt, 2015. (In Russian)
2. *Burlak S.A.* Proishozhdenie jazyka: Fakty, issledovanija, gipotezy. Moscow: Corpus, 2011. (In Russian)
3. *Grishina E.A.* Russkaja zhestikuljacija s lingvisticheskoj točki zrenija (korpusnye issledovanija). Moscow: YaSK, 2017. (In Russian)
4. *Kibrik A.A.* Kognitivnyj analiz diskursa: lokal'naja struktura // Jazyk i mysl': sovremennaja kognitivnaja lingvistika / Red. A.A. Kibrika, A.D. Kosheleva, A.V. Kravchenko, Ju.V. Mazurovoj, O.V. Fedorovoj. Moscow: YaSK, 2015. P. 595–634. (In Russian)
5. *Kibrik A.A.* Russkij mul'tikanal'nyj diskurs. Chast' I. Postanovka problemy // Psichologičeskij žurnal. 2018. V. 39. № 1. P. 71–81. (In Russian)
6. *Kibrik A.A., Molchanova N.B.* Kanaly mul'timodal'noj komunikacii: otnositel'nyj vklad v ponimanie diskursa // Mul'timodal'naja komunikacija: teoreticheskie i jempiricheskie issledovanija: materialy seminarov. Moskva, 15 nojabrja 2013 g. / Ed. O.V. Fedorovoj, A.A. Kibrika. Moscow: Buki Vedi, 2014. P. 99–114. (In Russian)
7. *Kotov A.A., Zinina A.A.* Funkcional'nyj analiz neverbal'nogo komunikativnogo povedenija // Komp'juternaja lingvistika i intellektual'nye tehnologii: Po materialam ezhegodnoj Mezhdunarodnoj konferencii “Dialog”. Moscow: RGGU, 2015. P. 308–320. (In Russian)
8. *Litvinenko A.O., Nikolaeva Ju.V., Kibrik A.A.* Annotirovanie russkix manual'nyx zhestov: teoreticheskie i praktičeskie voprosy // Komp'juternaja lingvistika i intellektual'nye tehnologii: Po materialam ezhegodnoj Mezhdunarodnoj konferencii “Dialog” (Moskva, 31 maja – 4 ijunja 2017 g.). Vyp. 16 (23). Moscow: Izd-vo RGGU, 2017. P. 271–286. (In Russian)
9. *Pogodin A.L.* Jazyk kak tvorčestvo (psichologičeskie i social'nye osnovy tvorčestva reči): Proishozhdenie jazyka. Har'kov, 1913. (In Russian)
10. *Rasskazy o snovidenijah: korpusnoe issledovanie ustnogo russkogo diskursa* / Red. Kibrika A.A., Podlesskoj V.I. Moscow: YaSK, 2009. (In Russian)
11. *Tomasello M.* Istoki čelovečeskogo obščhenija. Moscow: YaSK, 2011. (In Russian)
12. *Fedorova O.V.* Raspredelenie zritel'nogo vnimanija sobesednikov v estestvennoj komunikacii: 50 let spustja // Kognitivnaja nauka v Moskve: novye issledovanija. Materialy konferencii 15 ijunja 2017 g. / Red. E.V. Pechenkovej, M.V. Falikman. Moscow: BukiVedi, IPPiP. P. 370–375. (In Russian)
13. *Fedorova O.V.* Vklad kinetičeskoj sostavljajushhej v mul'timodal'nuju komunikaciju, ili Eshhe raz o strategijah opisanija tangramm // Komp'juternaja lingvistika i intellektual'nye tehnologii: Po materialam ezhegodnoj Mezhdunarodnoj konferencii “Dialog” (Moskva, 31 maja – 4 ijunja 2017 g.). Vyp. 16 (23). Moscow: Izd-vo RGGU, 2017. V. 2. P. 118–133. (In Russian)
14. *Fedorova O.V., Kibrik A.A., Korotaev N.A., Litvinenko A.O., Nikolaeva Ju.V.* Vremennaja koordinacija meždu zhestovymi i rečevymi edinicami v mul'timodal'noj komunikacii // Komp'juternaja lingvistika i intellektual'nye tehnologii: Po materialam ezhegodnoj Mezhdunarodnoj konferencii “Dialog” (Moskva, 1–4 ijunja 2016 g.). Vyp. 15 (22). Moscow: Izd-vo RGGU, 2016. P. 159–170. (In Russian)
15. *Abuczki A., Esfandiari B.G.* An overview of multimodal corpora, annotation tools and schemes // Argumentum. 2013. V. 9. P. 86–98.
16. *Allwood J.* Multimodal Corpora // Corpus linguistics. An international handbook / Eds. Lüdeling A., Kytö M. Berlin: Mouton de Gruyter, 2008. P. 207–225.
17. *Body–language–communication: An international handbook on multimodality in human interaction* / Eds. Müller C., Cienki A., Fricke E., Ladewig S.H., McNeill D., Teßendorf S. Berlin: Mouton de Gruyter, 2013.
18. *Brône G., Oben B.* InSight interaction. A multimodal and multifocal dialogue corpus // Language resources and evaluation. 2015. V. 49 (1). P. 195–214.
19. *Chafe W.* The importance of not being earnest: The feeling behind laughter and humor. Amsterdam: Benjamins, 2007.
20. *Fernández-Baena A., Montaña R., Antonijoan M., Roversi A., Miralles D., Alías F.* Gesture synthesis adapted to speech emphasis // Speech communication. 2014. V. 57. P. 331–350.
21. *Gairola P., Kumar S.* Hand gesture recognition from video // International journal of science and research. 2014. V. 3. Is. 4. P. 154–158.
22. *Geeraerts D., Kristiansen G.* Variationist linguistics // Handbook of Cognitive Linguistics / Eds. E. Dąbrowska, D. Divjak. Berlin: de Gruyter Mouton, 2015. P. 366–389.
23. *Goldin-Meadow S.* The role of gesture in communication and thinking // Trends in cognitive sciences. 1999. V. 3 (11). P. 419–429.

24. *Goldin-Meadow S.* What modern-day gesture can tell us about language evolution // *The Oxford handbook of language evolution* / Eds. M. Tallerman, K.R. Gibson. Oxford: Oxford University Press, 2012. P. 545–557.
25. *Huang X., Baker J., Reddy R.* A historical perspective of speech recognition // *Communications of the ACM*. 2014. V. 57. № 1. P. 94–103.
26. *Karpiński M., Jarmołowicz-Nowikow E., Malisz Z.* Aspects of gestural and prosodic structure of multimodal utterances in Polish task-oriented dialogues // *Speech and language technology*. 2009. V. 11. P. 113–122.
27. *Kendon A.* Some functions of gaze direction in social interaction // *Acta psychologica*. 1967. V. 26. P. 22–63.
28. *Kendon A.* Gesticulation and speech: Two aspects of the process of utterance // *The relationship of verbal and nonverbal communication* / Ed. M.R. Key. Berlin: Walter de Gruyter, 1980. P. 207–227.
29. *Kendon A.* *Gesture. Visible action as utterance*. Cambridge: Cambridge university press, 2004.
30. *Kibrik A.A.* *Reference in discourse*. Oxford: Oxford university press, 2011.
31. *Kibrik A.A.* The problem of non-discreteness and spoken discourse structure // *Computational linguistics and intellectual technologies*. 2015. V. 14 (21). P. 225–233.
32. *Kita S., Özyürek A.* What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking // *Journal of memory and language*. 2003. V. 48 (1). P. 16–32.
33. *Knight D.* The future of multimodal corpora // *Revista brasileira de linguística aplicada*. 2011. V. 2. P. 391–415.
34. *Kopp S., Bergmann K., Kahl S.* A spreading activation model of the semantic coordination of speech and gesture // *Proceedings of the 35th annual meeting of the Cognitive science society*. Berlin, 2013.
35. *Leonard Th., Cummins F.* The temporal relation between beat gestures and speech // *Language and cognitive processes*. 2010. V. 26 (10). P. 1457–1471.
36. *Levinson S.C., Holler J.* The origin of human multi-modal communication // *Philosophical transactions of the Royal society B*. 2014. V. 369. P. 20130302.
37. *Loehr D.* Temporal, structural, and pragmatic synchrony between intonation and gesture // *Laboratory phonology*. 2012. V. 3 (1). P. 71–89.
38. *McNeill D.* *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press, 1992.
39. *Melinger A., Levelt W.J.M.* Gesture and the communicative intention of the speaker // *Gesture*. 2004. V. 4.2. P. 119–141.
40. *Mondada L.* Bodies in action // *Language and dialogue*. 2014. V. 4 (3). P. 357–403.
41. *Morrel-Samuels P., Krauss R.M.* Word familiarity predicts temporal asynchrony of hand gestures and speech // *Journal of experimental psychology: human learning and memory*. 1992. V. 18 (3). P. 615–622.
42. *Multimodal corpora: From models of natural interaction to systems and applications* // Eds. Kipp M., Martin J.-C., Paggio P., Heylen D. Berlin, Heidelberg: Springer, 2009.
43. *Schegloff E.A.* On some gestures' relation to talk // *Structures of Social Action* / Eds. Atkinson J.M., *Heritage J.* Cambridge: Cambridge university press, 1984. P. 266–298.
44. *Swerts M., Krahmer E.* 2008. Facial expressions and prosodic prominence: comparing modalities and facial areas // *Journal of phonetics*. 2008. V. 36 (2). P. 219–238.
45. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production* / Ed. Chafe W.L. Norwood, NJ: Ablex, 1980.
46. *Wagner P., Malisz Z., Kopp S.* Gesture and speech in interaction: An overview // *Speech Communication*. 2014. V. 57. P. 209–232.