

Компьютерная лингвистика и интеллектуальные технологии:
по материалам международной конференции «Диалог 2019»

Москва, 29 мая — 1 июня 2019 г.

ЕДИНАЯ МУЛЬТИКАНАЛЬНАЯ АННОТАЦИЯ КАК ИНСТРУМЕНТ АНАЛИЗА ЕСТЕСТВЕННОЙ КОММУНИКАЦИИ¹

Кибрик А. А. (aakibrik@gmail.com)

Институт языкознания РАН,
МГУ им. М. В. Ломоносова, Москва, Россия

Коротаев Н. А. (n_korotaev@hotmail.com)

РГГУ, Институт языкознания РАН, Москва, Россия

Федорова О. В. (olga.fedorova@msu.ru)

МГУ имени М. В. Ломоносова,
Институт языкознания РАН, Москва, Россия

Евдокимова А. А. (arochka@gmail.com)

Институт языкознания РАН, Москва, Россия

Данная статья вносит вклад в область мультимедийного анализа дискурса. Мультимедийный анализ дискурса исследует многочисленные каналы, задействованные в естественной коммуникации, такие как вербальная структура, просодия, жестикация, движения головы, взгляд, положения туловища и т. д., рассматривая их как части единого процесса. Для изучения того, как взаимодействуют между собой коммуниканты и как устроена координация между различными коммуникативными каналами, мы ввели понятие единой мультимедийной аннотации. Единая мультимедийная аннотация была реализована в среде ELAN. В частности, мы рассмотрели три конкретных примера: (1) временную координацию между речью коммуникантов и их мануальной жестикацией; (2) распределение зрительного внимания коммуникантов между речью и мануальной жестикацией их собеседников; (3) взаимодействие между положением тела и движениями головы коммуникантов.

Ключевые слова: мультимедийный дискурсивный анализ, единая аннотация, речь, мануальные жесты, движения головы, взгляд, положения туловища

¹ Работа выполнена при финансовой поддержке РФФИ, проект 19-012-00626.

UNIFIED MULTICHANNEL ANNOTATION: A TOOL FOR ANALYSING NATURAL COMMUNICATION

Kibrik A. A. (aakibrik@gmail.com)

Institute of Linguistics RAS, Lomonosov Moscow State
University, Moscow, Russian Federation

Korotaev N. A. (n_korotaev@hotmail.com)

RSUH, Institute of Linguistics RAS, Moscow, Russian Federation

Fedorova O. V. (olga.fedorova@msu.ru)

Lomonosov Moscow State University, Institute of Linguistics
RAS, Moscow, Russian Federation

Evdokimova A. A. (arochka@gmail.com)

Institute of Linguistics RAS, Moscow, Russian Federation

This paper contributes to the research field of multichannel discourse analysis. Multichannel discourse analysis explores numerous channels involved in natural communication, such as verbal structure, prosody, manual gesticulation, head movements, eye gaze, torso postures, etc., and treats them as parts of an integrated process. For the purposes of investigating the way participants interact with one another and the way different communication channel correlate, we introduce the notion of an integrated multichannel annotation created with ELAN software. In particular, we consider three topics: (1) temporal alignment between participants' speech and manual gesticulation; (2) distribution of participants' visual attention as they watch their interlocutors talking and gesticulating manually; (3) interrelationship between participants' torso postures and head movements.

Key words: multichannel discourse analysis, unified annotation, speech, manual gesticulation, head movements, eye gaze, torso postures

1. Введение. Мультиканальная лингвистика

Мультиканальная (мультимодальная) лингвистика изучает все реальное многообразие «живой» коммуникации между людьми: слова, интонацию, жестикуляцию, направление взгляда, мимику, см. **рис. 1**. Интерес к изучению мультиканальности возник еще в древности, см., напр., [Квинтилиан 1834: XI], однако современная лингвистика, берущая начало в первые десятилетия XX в., долгое время занималась исключительно письменными текстами, то есть вербальным каналом [Linell 1982]. В конце XX в. ситуация начала меняться

в сторону изучения устного дискурса, то есть к вербальному каналу прибавился просодический, см., в частности, работы У. Чейфа [Chafe ed. 1980]; [1994]. Наконец, в XXI в. на наших глазах формируется новый мультиканальный подход [Kress 2010]; [Кибрик 2010]; [Knight 2011]; [Adolphs, Carter 2013]; [Müller et al. eds. 2013–2014]; [Кибрик 2018], принимающий во внимание все каналы общения между людьми, в том числе кинетические [Efron 1941]; [Крейдлин 2002]; [Kendon 2004]; [Бутовская 2004]; [McNeill 2005]; [Гришина 2017].

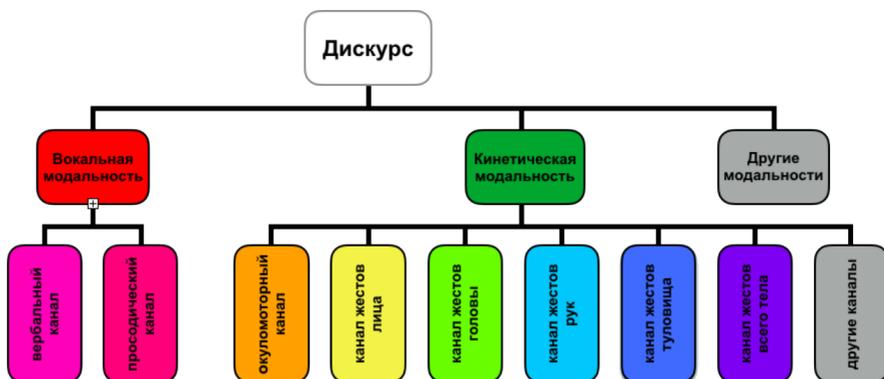


Рис. 1. Модель мультиканального дискурса

Одно из наиболее важных и перспективных направлений развития мультиканальной лингвистики — это разработка и создание мультиканальных ресурсов (корпусов). В отличие от моноканальных и мономодальных ресурсов, уже имеющих свою историю и традицию, мультиканальные ресурсы в настоящий момент находятся на этапе становления. Мультиканальный ресурс — это сочетание двух компонентов. Во-первых, это совокупность медиафайлов, фиксирующих коммуникацию между людьми. Во-вторых, это аннотация коммуникативных событий, содержащихся в медиафайлах. Данная работа посвящена описанию принципов единой мультиканальной аннотации ресурса «Рассказы и разговоры о грушах», проект по разработке которого осуществляется в Институте языкознания РАН (сайт multidiscourse.ru). В разделе 2 изложены основные характеристики собранного ресурса. Раздел 3 содержит описание принципов мультиканальной аннотации, осуществляемой в рамках проекта. В разделе 4 предложены три конкретных примера применения разработанной аннотационной схемы на практике.

2. Ресурс «Рассказы и разговоры о грушах»

Ресурс состоит из отдельных коммуникативных эпизодов (так называемых записей), в каждом из которых участвуют четыре человека — Рассказчик, Комментатор, Пересказчик и Слушатель. Название корпуса обусловлено тем, что двое из участников вначале смотрят известный в лингвистике стимульный

материал — шестиминутный «Фильм о грушах», созданный в 1970-е гг. в Беркли исследовательской группой под рук. У. Чейфа (linguistics.ucsb.edu/faculty/chafe/pearfilm.htm). В этом фильме, не включающем никаких речевых действий персонажей, показана история о взаимодействии ряда лиц, в том числе садовника, собирающего груши, мальчика, крадущего корзину с грушами, и еще нескольких. Фильм представляет собой хорошо продуманную цепь физических и социальных событий и давно зарекомендовал себя как эффективный способ получать компактные и сравнимые между собой образцы устного дискурса.

Фильм просматривается участниками, которые в дальнейшем выполняют роли Рассказчика и Комментатора. На **рис. 2** можно видеть общий дизайн коммуникативной ситуации. Сначала Рассказчик в режиме монолога рассказывает Пересказчику содержание фильма (этап рассказа). Затем наступает интерактивный этап, в ходе которого Комментатор дополняет или уточняет рассказ, а Пересказчик задает вопросы обоим собеседникам, видевшим фильм (этап разговора). После этого появляется Слушатель, и Пересказчик пересказывает ему фильм, опять в режиме монолога (этап пересказа). В конце Слушатель письменно записывает еще один, вторичный, пересказ фильма. Два последних этапа необходимы для того, чтобы мотивировать участников к полноценной и осмысленной коммуникации.



Рис. 2. Общий дизайн коммуникативной ситуации

Ресурс «Рассказы и разговоры о грушах» состоит из двух частей. Первая часть была собрана летом 2015 г.; она включает 24 записи длительностью около 9 часов и объемом около 100 тыс. словоупотреблений; в записях приняли участие 96 человек от 18 до 36 лет, в том числе 34 мужчины и 62 женщины. Вторая часть корпуса была собрана летом 2017 г.; она включает 16 записей длительностью около 6 часов и объемом около 60 тыс. словоупотреблений; в записях приняли участие 64 человека от 18 до 36 лет, в том числе 16 мужчин и 48 женщин.

При записи корпуса были использованы оригинальные технические решения. *Аудиозапись* велась при помощи шестиканального рекордера ZOOM Handy Recorder, что обеспечивало автоматическую синхронизацию (параметры записи 96 kHz / 24 bit). Для *видеозаписи* были использованы три индивидуальные промышленные видеокамеры JAI GO-5000M с частотой 100 к/с и разрешением 1392×1000, которые записывали крупным планом каждого из трех основных коммуникантов. Для последующей аналитической работы важно, что эти камеры позволяют получить запись в формате tjpeg; данный формат выгодно отличается от остальных отсутствием межкадрового сжатия, что является необходимым условием для дальнейшего кадрового аннотирования. Частота записи 100 к/с позволяет проводить анализ собранного видеоматериала с точностью до 10 мс, что является показателем высокой точности аннотирования. Насколько нам известно, оборудование этого типа для лингвистических исследований было использовано впервые. Кроме того, общая видеокамера GoPro Hero с частотой 50 к/с и разрешением 2700 × 1500 (для записей 2015 г.) и частотой 120 к/с и разрешением 1920 × 1080 (для записей 2017 г.) записывала общий план. Для *регистрации движений глаз* были использованы две пары очков-айтрекеров фирмы Tobii Glasses II с частотой 50 Гц и разрешением видеокамеры 1920 × 1080. В очки-айтрекер вмонтирована миниатюрная видеокамера, регистрирующая движения глаз испытуемого, вторая видеокамера снимает окружающую обстановку.

3. Мультимедийная аннотация

При аннотировании корпуса мы различали два вида разметки: базовую и дополнительную. Базовая разметка включает в себя аннотирование вокального компонента (вербальный и просодический каналы), мануального (жестов рук), цефалического (жестов головы) и окуломоторного (направления взгляда) каналов. *Вокальная* аннотация, выполненная в программах Praat (fon.hum.uva.nl/praat) и MS Word, состоит в членении речевого потока на значимые фрагменты (элементарные дискурсивные единицы — ЭДЕ, слова, заполненные и незаполненные паузы, неречевые звуки), а также в приписывании свойств ЭДЕ и отдельным их частям [Кибрик, Подлесская ред. 2009]; [Kibrik et al. 2019]. Для аннотирования мануального канала применяется новая оригинальная методика, разработанная в среде ELAN (tla.mpi.nl/tools/tla-tools/elan); она основана на сегментировании потока мануального поведения на периоды неподвижности и отдельные движения, которые затем формируют функциональные единицы — жесты, адапторы и смены позы. Для этих единиц затем указываются их характеристики — рукость жестов, фазовая структура, функциональный тип и т. д., подробнее см. [Литвиненко и др. 2017]. Аннотирование *цефалического* канала основано на принципах, разработанных для аннотирования мануальных жестов. В ходе аннотирования *окуломоторного* канала производится экспорт данных айтрекинга на видеосцену, затем с помощью программы Tobii Pro Glasses Analyzer извлекаются данные о временной развертке всех фиксации длительностью выше 100 мс, на которые потом в ручном режиме накладывается аннотационная схема с указанием направления взгляда, см. [Федорова 2017].

Дополнительная разметка включает в себя аннотирование *жестов тела*, *миимики*, а также *референциальную* аннотацию, которая содержит разметку языковых выражений с конкретной референцией. Описания принципов аннотирования коммуникативных каналов, а также образцы базовой и дополнительной разметок можно найти на сайте проекта multidiscourse.ru на вкладках «Корпус» и «Принципы аннотации».

Для изучения того, как взаимодействуют между собой участники коммуникации и как устроена координация между каналами, используется *единая мультиканальная аннотация*. В такой разметке сводятся результаты аннотаций, выполненных независимо для отдельных каналов. В базовой части единой аннотации учитываются вокальные, окуломоторные, цефалические и мануальные действия трех основных участников записи. Дополнительная часть единой аннотации включает остальные коммуникативные каналы. С технической точки зрения мультиканальная аннотация представляет собой файлы формата .eaf, используемые в программе ELAN; при этом вокальная и окуломоторная разметки, изначально реализованные в других программах, конвертируются в этот формат. Для унификации обозначений мы используем следующие соглашения.

- (1) Мультиканальное поведение каждого участника фиксируется в отдельном наборе слоев аннотации. Порядок следования слоев и их названия идентичны в каждом наборе — с точностью до начальной литеры, указывающей на роль участника. Например, в слое N-mGesture размечаются мануальные жесты Рассказчика, а в слое C-mGesture — мануальные жесты Комментатора.
- (2) Единицы мультиканального поведения (ЭДЕ, слова, мануальные и цефалические жесты, фиксации и проч.) фиксируются в виде непустых интервалов в слоях с независимой временной привязкой (в системе ELAN для этого используются т.н. «стереотипы» None и Included in). В качестве названия интервала используется стандартный идентификатор, включающий в себя указание на роль участника, кодовое обозначение канала и типа единицы, а также трех- или четырехзначный номер.
- (3) Свойства единицы (напр., вербальное наполнение ЭДЕ, рукость жеста, направление взгляда) отмечаются в интервалах зависимых слоев (слоев со «стереотипом» Symbolic Association). Значения свойств либо выбираются из закрытого списка, либо вводятся вручную по специальным правилам.

Более подробное описание разработанной схемы базовой мультиканальной аннотации см. на сайте проекта multidiscourse.ru на вкладке «Принципы аннотации»; фрагмент аннотации, реализованной для записи #22, представлен в Приложении.

К настоящему моменту полностью аннотированы и выложены на сайт три записи #04, 22 и 23 длительностью около 1 часа, которые составляют эталонный подкорпус. *Эталонный подкорпус* — это экспериментальная площадка, на которой тестируются различного рода гипотезы, чтобы потом

верифицировать их на более обширном материале. Как можно видеть по **табл. 1**, несмотря на то что мы не устанавливали для испытуемых временных ограничений, все три записи похожи друг на друга по длительностям этапов — на рассказ приходится около 20% от времени записи, на разговор — примерно 50%, и на пересказ — 30%.

Табл. 1. Длительность записей эталонного подкорпуса и их этапов

#	общая длительность записи (мин:сек,мсек)	рассказ (мин:сек,мсек)	разговор (мин:сек,мсек)	пересказ (мин:сек,мсек)
04	24:36,240	05:22,640 (21,9%)	12:37,920 (51,3%)	06:35,680 (26,8%)
22	18:04,960	03:37,960 (20,1%)	08:48,280 (48,7%)	05:38,700 (31,2%)
23	16:26,520	03:52,400 (23,5%)	07:41,240 (46,8%)	04:52,880 (29,7%)

Ниже в **разделе 4** описаны три конкретных примера применения разработанной единой мультимедийной аннотационной схемы; исследования **4.1** и **4.3** выполнены на материале эталонного подкорпуса, исследование **4.2** — на материале записей #04, 06, 23.

4. Примеры использования единой мультимедийной аннотации

4.1. Взаимодействие вокальной модальности и мануального канала

В исследованиях временной координации вокальных и мануальных единиц часто утверждается, что жестикация имеет опережающий характер (напр., [Kendon 1980]; [McNeill 1992]; [Loehr 2012]; [Гришина 2017]). Для проверки этой гипотезы на нашем материале мы провели анализ координации на двух уровнях: верхнем и нижнем. Единицами верхнего уровня сегментации в мануальной жестикации являются жесты, в речи — ЭДЕ; единицами нижнего уровня — соответственно маховые фазы жестов (далее — «махи») и акцентированные словоформы (далее — «акценты»). Такие соответствия объясняются как содержательными причинами (жесты и ЭДЕ выступают базовыми единицами при анализе жестикационного и речевого потока и, предположительно, отражают центральные когнитивные феномены, отвечающие за процесс коммуникации; махи и акцентированные словоформы, в свою очередь, формируют содержательные центры внутри этих базовых единиц), так и сугубо формальными характеристиками.

Табл. 2. Длительность речевых и жестикуляционных единиц в эталонном подкорпусе

	Жесты	ЭДЕ	Махи	Акценты	Все сло- воформы
Количество	2714	2980	2713	3522	9587
Длительность (среднее), мс	846	1056	453	434	305
Длительность (медиана), мс	720	922	400	420	270

Как видно из **табл. 2**, внутри пар вида «жест — ЭДЕ» и «мах — акцент» наблюдается близость средних и медианных значений длительности. Это задает необходимую для дальнейшего анализа сопоставимость рассматриваемых единиц.

В ходе работы были разработаны и внедрены три критерия соответствия жестикуляционных единиц речевым: два вида т.н. относительной точности пересечения и гармоническое среднее. Все эти критерии основаны на формальной близости единиц на общей временной шкале — см. **рис. 3**; подробнее — [Федорова и др. 2016] и [Коротаев 2018]). Было обнаружено, что *степень координации* в парах «жест — ЭДЕ» в целом выше, чем в парах «мах — акцент». Для оценки *характера* координации каждая пара была отнесена, во-первых, к одному из четырех интегральных типов, выделяемых на основании соотношения левых и правых границ (внутренние, объемлющие, ранние и поздние, см. **рис. 4**), во-вторых, к одному из трех позиционных типов, выделяемых на основании соотношения только левых границ.

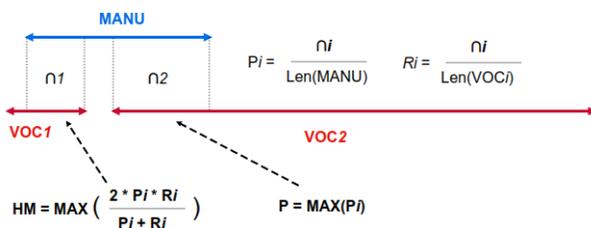


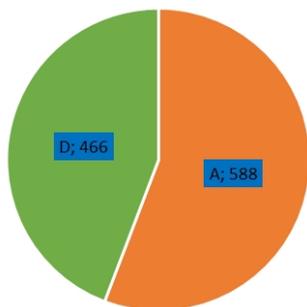
Рис. 3. Соотношение понятий точности пересечения (P) и гармонического среднего (HM; под используемым в формуле R понимается полнота пересечения)



Рис. 4. Интегральные типы жестикуляционных единиц (обозначены синим цветом) с точки зрения соотношения их границ с границами соответствующих им вокальных единиц (обозначены красным цветом)

Исходная гипотеза была подтверждена только для первой оценки: и махи, и жесты демонстрируют тенденцию к интегральному опережению соответствующих им вокальных единиц. Преимущество ранних жестикуляционных единиц над поздними во всех случаях статистически значимо (биномиальный критерий, $p < 0,001$). Однако при опоре только на левые границы картина оказалась более сложной: если левые границы махов чаще опережают левые границы соответствующих акцентов, то для жестов, напротив, выявлена тенденция к запаздыванию левых границ относительно ЭДЕ. Так, на рис. 5 показаны количественные соотношения случаев опережения (А) и запаздывания (D) левых границ махов (круговая диаграмма слева) и жестов (диаграмма справа) относительно левых границ соответствующих им речевых единиц. Представлены данные о парах жестикуляционных и речевых единиц, выявленные по критерию максимального гармонического среднего; учитывались только такие случаи опережения и запаздывания, при которых расстояния между границами превышало «дельту» в 50 мс. Наблюдаемые различия в распределении махов и жестов по двум позиционным типам статистически значимы («хи-квадрат», $p < 0,001$). При изменении критерия соответствия и / или «дельты» картина существенно не меняется.

Махи: левая граница относительно акц. слова (НМ); $\delta = 50$ мс



Жесты: левая граница относительно ЭДЕ (НМ); $\delta = 50$ мс

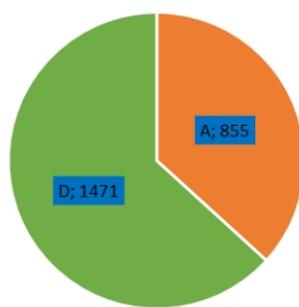


Рис. 5. Распределение махов и жестов по опережению (А) / запаздыванию (D) левых границ относительно левых границ речевых единиц, соответствующим им по критерию максимального гармонического среднего, при «дельте» в 50 мс

Кроме того, на характер распределений влияют и ингерентные свойства жестикуляционных единиц. Так, статистически значимое преимущество раннего интегрального типа над поздним наблюдается только для долгих и средних по длительности махов и средних жестов; для кратких махов, а также кратких и долгих жестов нет оснований отбрасывать нулевую гипотезу о случайности наблюдаемого распределения (биномиальный критерий, $p > 0,1$).

4.2. Взаимодействие мануального и окуломоторного каналов

Взаимодействие вокальной модальности и мануального канала, описанное в предыдущем разделе — наиболее изученная область мультимодальных исследований. Считается, что жестикуляция помогает общению коммуникантов; в частности, об этом свидетельствует мета-анализ в [Hostetter 2011]. Жестикуляция, с одной стороны, дублирует информацию, которая передается через речь, и, с другой стороны, передает дополнительную, а иногда и противоречащую, информацию. Таким образом, в процессе коммуникации собеседники принимают во внимание не только речь собеседника, но и жестикуляционный компонент. Однако так ли это на самом деле? Часто ли люди обращают внимание на жестикулирующие руки собеседника? Более точный ответ на этот вопрос дают современные исследования взаимодействия мануального и окуломоторного каналов².

На сегодняшний день проведено немного исследований, посвященных изучению зрительного внимания к жестикулирующим рукам собеседника. В пионерских работах [Gullberg, Holmqvist 1999]; [2002] было показано, что в естественной языковой коммуникации 96% всего времени слушающий смотрит

² В данном разделе мы анализируем окуломоторный канал Рассказчиков и мануальный канал Пересказчиков и Комментаторов.

на лицо говорящего собеседника, в то время как на жестикулирующие руки он смотрит совсем редко, всего в 0,5% случаев, что покрывает только 7% жестикуляции собеседника. В работе Beattie et al. 2010 оказалось, что на лицо собеседника слушающий смотрит в 85% времени, а на руки приходится 2,1%. В разных исследованиях высказывались разные предположения о том, в каких именно случаях слушающий переводит взгляд с лица собеседника на его руки. Основное объяснение, предложенное в процитированных работах, связано с тем, что в большинстве случаев, когда испытуемый смотрит на лицо, периферическим зрением он видит и жестикулирующие руки — при стандартном расстоянии между собеседниками в 1,5–2 м и жестикуляции на уровне груди угол зрения оказывается около 8–10°, что в общем случае позволяет видеть движения рук без перевода взгляда, подробнее см. [Барабанщиков, Жегалло 2013].

При нашем дизайне Рассказчик также мог видеть руки Пересказчика периферическим зрением (так называемая ближняя периферия 8,5°), не переводя на них прямой взгляд с его лица. Поэтому, с одной стороны, айтрекинг не дает нам точного ответа на вопрос, сколько времени Рассказчик суммарно смотрел (фокальным или периферическим зрением) на жестикулирующие руки собеседников³. С другой стороны, однако, факультативность прямых взглядов на руки предоставляет нам возможность изучить индивидуальные различия между Рассказчиками — о чем говорят те случаи, когда Рассказчик мог бы и не переводить взгляд на руки Пересказчика, а тем не менее его перевел?

Были обнаружены значимые индивидуальные различия между Рассказчиками. Оказалось (см. табл. 3), что при одинаковых исходных условиях Рассказчик #23 минимально смотрит на жестикулирующие руки собеседников на протяжении всей коммуникации; внимание Рассказчика #04 привлечено к жестикулирующим рукам только на этапе пересказа, а внимание Рассказчика #06 привлечено к ним как на этапе разговора, так и на этапе пересказа. Данные различия не могут быть объяснены предлагаемыми ранее факторами, а являются следствием действия различных закономерностей распределения зрительного внимания участников коммуникации — общих, контекстно-(не)зависимых и индивидуальных, подробнее см. [Федорова, Жердев 2019].

Табл. 3. Доля зрительного внимания Рассказчика, направленного на жестикулирующие руки Пересказчика и Комментатора (по этапам, в % относительно суммарной жестикуляции)

#	рассказ	разговор		пересказ, Пересказчик
		Пересказчик	Комментатор	
04	0	3	0,8	25,5
06	0	25,6	28,4	19,7
23	0	0	1,2	8,1

³ Что связано с отсутствием информации о нефокальных движениях взгляда.

4.3. Положение туловища и движения головы коммуникантов в зависимости от их коммуникативной роли

При аннотировании жестов туловища мы выделяем плечевую зону, зону спины и зону ног⁴. Положение туловища показывает степень заинтересованности собеседника и/или его вовлеченности в разговор, ср. [Kendon 1970]; [Schefflen 1973]; [Alibalia et al. 2001]. В частности, при анализе зоны спины мы рассматриваем движения корпусом, которые могут носить как коммуникативный (наклон вперед, когда коммуникант слушает, и отклонение на спинку стула, когда он сам говорит), так и выразительный характер (подача корпуса вперед или вперед-назад для усиления своей речи) (см. подробнее [Birdwhistell 1971]; [Bobick 1997]; [Frey, von Cranach 1973]; [Kendon 1973]; [Mehrabian 1968]).

Если рассматривать канал жестов туловища не изолированно, а во взаимодействии с другими каналами, можно выявить много важных закономерностей. Так, при взаимодействии с цефалическим каналом значимым оказывается расположение туловища относительно стула. Одни участники независимо от их роли в коммуникации опираются на спинку стула, другие опираются локтями на ноги или наклоняются вперед — иногда с дополнительной опорой на стул руками. Это различие влияет на количество так называемых «эховых» движений головы, т.е. движений, в том или ином смысле инициированных движениями в других каналах⁵.

Как видно из **табл. 4**, в записи #22 наибольшее количество эховых движений головы зафиксировано у Пересказчика; при этом плечи и спина этого участника не опираются на спинку стула, а ноги убраны под стул. Меньше всего эховых движений у Комментатора, который опирается на спинку стула и спиной, и (время от времени) плечами. Наконец, Рассказчик опирается на спинку стула спиной, но не плечами — и частота эховых движений у этого участника ниже, чем у Пересказчика, но выше, чем у Комментатора. В записях #4 и #23 у всех участников наблюдаются другие закономерности в зонах туловища, что объясняется их индивидуальным стилем поведения.

В результате анализа эталонного подкорпуса в зависимости от ролевой характеристики и текущего статуса в коммуникации (говорящий или слушающий) выделяются следующие стратегии. Рассказчики (за исключением Рассказчика #22), предпочитают облакачиваться на спинку стула при слушании и не облакачиваться при говорении. Комментаторы чаще облакачиваются или имеют опору на руки и у них наблюдается наименьшее число эхо в цефалическом канале. Пересказчики независимо от текущего статуса чаще всего не облакачиваются на спинку стула, но имеют дополнительную опору на руки. Таким образом, у Рассказчиков и Пересказчиков не облакачивание на спинку стула является маркированным положением их текущего коммуникативного статуса.

⁴ Таким образом, в данной работе мы понимаем туловище расширительно, ср. [Bull 1987].

⁵ Принципы описания движений головы см. на сайте проекта multidiscourse.ru/annotation/.

Таблица 4. Положение туловища и эховые движения

головы в эталонном подкорпусе. Условные обозначения:

- A — опирается на спинку стула; B — не опирается на спинку стула; a — опирается на локти⁶; b — опирается руками на стул; C — ноги скрещены и вытянуты; D — ноги скрещены под стулом; E — нога на ногу; F — ноги на ширине плеч; G — ноги вместе перед собой (ср. классификацию в Bull 1987: 186–187)

#	плечи	спина	ноги	эхо в цефал. канале ⁷
Рассказчик				
04	Ba	A — когда слушает B — когда говорит	E	80 / мин.
22	B	A	C, E, D	48 / мин.
23	B	A — когда слушает B — когда говорит	D, C, G	80 / мин.
Комментатор				
04	Ba	B, A	F, D	48 / мин.
22	A, B	A	D	17 / мин.
23	Bb	B	C, D, F	17 / мин.
Пересказчик				
04	Ba	B	D — когда слушает C — когда говорит	48 / мин.
22	Bb	B	C, F, D	80 / мин.
23	Ba	A, B	F, G	17 / мин.

5. Заключение

В данной работе был предложен новый подход к анализу естественной мультимедийной коммуникации. Аннотации отдельных каналов, выполненные независимо друг от друга, были сведены в единую мультимедийную аннотацию, что позволило провести несколько конкретных исследований взаимодействия каналов. В настоящей статье было описано два примера взаимодействия базовых коммуникативных каналов (вокальной модальности vs. канала мануальных жестов; мануального vs. окуломоторного каналов), а также пример взаимодействия базового и дополнительного канала (цефалического vs. канала жестов туловища). Данная работа будет продолжена в сторону увеличения количества одновременно рассматриваемых каналов, вплоть до анализа канала жестов всего тела (см. **рис. 1**). Такой подход сделает возможным анализ таких исходно мультимедийных явлений, как, напр., смех [Chafe 2007].

⁶ Локти могут находиться на бедрах или на коленях, на любой поверхности ног, на которую можно опереться или быть уперты в подвздошные кости.

⁷ Среднее количество эховых движений в минуту.

Литература

1. *Adolphs S., Carter R.* (2013), *Spoken corpus linguistics: From monomodal to multimodal*, N.-Y.: Routledge.
2. *Alibalia M. W., Heath D. C., Myers H. J.* (2001), Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen, *Journal of Memory and Language*, Vol. 44 (2), pp. 169–188.
3. *Barabanshchikov V. A., Zhegallo A. V.* (2013), *Registration and analysis of gazes [Registratsiya i analiz napravlenosti vzora cheloveka]*. Moscow: Institut psikhologii RAN.
4. *Birdwhistell R. L.* (1971), *Kinesics and context*, London: Penguin Press.
5. *Bobick A.* (1997), *Movement, activity, and action: the role of knowledge in the perception of motion*, Royal Society Workshop on Knowledge-based Vision in Man and Machine, London.
6. *Bull P. E.* (1987), *Posture and Gesture*, New York: Pergamon Press.
7. *Butovskaya M. L.* (2004), *Body language: nature and culture (evolutionary and cross-cultural foundations for non-verbal human communication) [Yazyk tela: priroda i kul'tura (evolyutsionnye i kross-kul'turnye osnovy neverbal'noy kommunikatsii cheloveka)]*, Moscow: Nauchny mir.
8. *Chafe W.* (1994), *Discourse, consciousness, and time. The flow and displacement of conscious experience in speaking and writing*, Chicago.
9. *Chafe W.* (2007). *The importance of not being earnest: The feeling behind laughter and humor*, Amsterdam: John Benjamins.
10. *Chafe W.* (ed.) (1980), *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*, Norwood: Ablex.
11. *Efron D.* (1941), *Gesture and environment*, Oxford: King's Crown Press.
12. *Fedorova O. V., Kibrik A. A., Korotaev N. A., Litvinenko A. O., Nikolaeva Ju. V.* (2016), Temporal coordination between gestural and speech units in multimodal communication [Vremennaya koordinatsiya mezhdu zhestovymi i rechevymi edinitsami v mul'timodal'noy kommunikatsii], *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2016" [Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii]*, RGGU, Moscow, pp. 159–170.
13. *Fedorova O. V.* (2017), Distribution of the interlocutors' visual attention in natural communication: 50 years later [Raspredeleniye zritel'nogo vnimaniya sobesednikov v estestvennoy kommunikatsii: 50 let spustya], E. V. Pechenkova, M. V. Falikman (eds.) *Cognitive science in Moscow: new research. Proceedings of the conference [Kognitivnaya nauka v Moskve: novye issledovaniya. Materialy konferentsii 15 iyunya 2017]*. Moscow: BukiVedi, IPPiP, pp. 370–375.
14. *Fedorova O. V., Zherdev I. Ju.* (2019), Follow the hands of the interlocutor! (on strategies for the distribution of visual attention) [Sledi za rukami sobesednika (o strategiyakh raspredeleniya zritel'nogo vnimaniya)], *Experimental Psychology [Ekhsperimental'naya psikhologiya]*, Vol. 1.
15. *Frey S., von Cranach M.* (1973), A method for the assessment of body movement variability, M. von Cranach, I. Vine (eds.), *Social communication and movement*, pp. 389–418. London: Academic Press.

16. *Grishina E. A.* (2017), Russian gestures from a linguistic perspective [Russkaya zhestikulyatsiya s lingvisticheskoy tochki zreniya], Moscow: Jazyki slavyanskoj kul'tury.
17. *Gullberg M., Holmqvist K.* (1999), Keeping an eye on gestures: Visual perception of gestures in face-to-face communication, *Pragmatics and Cognition*, Vol. 7., pp. 35–63.
18. *Gullberg M., Holmqvist K.* (2002), Visual attention towards gestures in face-to-face interaction vs. on screen, I. Wachsmuth and T. Sowa (Eds.), *Gesture and Sign Language in Human–Computer Interaction*, Berlin Heidelberg: Springer, pp. 206–214.
19. *Hostetter A. B.* (2011), When do gestures communicate? A meta-analysis, *Psychological Bulletin*, Vol. 137(2), pp. 297–315.
20. *Kendon A.* (1970), Movement coordination in social interaction: Some examples described, *Acta Psychologica*, Vol. 32 (2), pp. 100–125.
21. *Kendon A.* (1973), The role of visible behaviour in the organization of social interaction, von Cranach, I. Vine (eds.), *Social communication and movement: Studies of interaction and expression in man and chimpanzee*. London: Academic Press, pp. 29–74.
22. *Kendon A.* (1980), Gesticulation and speech: Two aspects of the process of utterance, in M. R. Key (ed.), *The relationship of verbal and nonverbal communication*, pp. 207–227.
23. *Kendon A.* (2004), *Gesture. Visible action as utterance*, Cambridge.
24. *Kibrik A. A.* (2010), Multimodal linguistics [Mul'timodal'naya lingvistika], Yu. I. Aleksandrov, V. D. Solov'yev (eds.), *Cognitive studies [Kognitivnyye issledovaniya]*, Vol. IV, Institute of psychology, Moscow, pp. 134–152.
25. *Kibrik A. A.* (2018), Russian multichannel discourse. Part I. Setting up the problem [Russkiy mul'tikanal'nyy diskurs. Chast' I. Postanovka problemy], *Psikhologicheskij zhurnal*, Vol. 39 (1), pp. 70–80.
26. *Kibrik A. A., Korotaev N. A., Podlesskaya V. I.* (in press), Russian spoken discourse: Local structure and prosody, In *Search for a Reference Unit of Spoken Language: A Corpus-Driven Approach*.
27. *Kibrik A. A., Podlesskaja V. I.* (eds.), (2009), *Night Dream Stories: A corpus study of spoken Russian discourse [Rasskazy o snovideniyakh: korpusnoye issledovaniye russkogo ustnogo diskursa]*. Moscow: Jazyki slavyanskikh kul'tur.
28. *Knight D.* (2011), *Multimodality and active listenership: A corpus approach*, Bloomsbury, London.
29. *Korotaev N. A.* (2018), On temporal coordination between gestural and speech units in spontaneous spoken communication [O vremennoj koordinatsii zhestikulyatsionnyh i rechevyh edinit v nepodgotovlennoj ustnoy kommunikatsii], “Slovo i zhest” conference, Moscow, pp. 10–12.
30. *Kress G.* (2002), The multimodal landscape of communication, *Medien Journal*, Vol. 4, pp. 4–19.
31. *Kreydlin G. E.* (2002), *Nonverbal semiotics [Neverbal'naya semiotika]*, New literary review, Moscow.
32. *Linell P.* (1982), *The written language bias in linguistics*, University of Linköping.

