

Временная координация между речевыми и жестовыми единицами в мультимодальной коммуникации¹

О.В. Фёдорова, А.А. Кибрик, Н.А. Коротаев,
А.О. Литвиненко, Ю.В. Николаева

Диалог, 2 июня 2016 г.

¹Язык как он есть: русский мультимодальный дискурс (РНФ 14-18-03819)

«Рассказы и разговоры о грушах»

Общие сведения

Мультимодальный корпус «Рассказы и разговоры о грушах» был собран летом 2015 г. по мотивам «Фильма о грушах» У. Чейфа (Chafe ed. 1980).

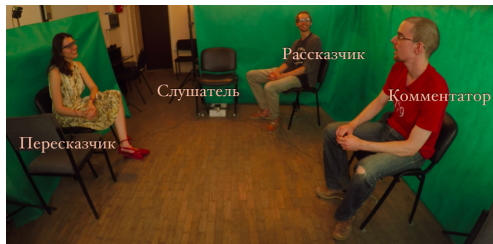
- 24 записи (длительностью от 12 до 38 минут)
- общая длительность 10 часов
- 110 тыс. словоупотреблений
- 11 медиафайлов для каждой записи
- 24 письменных пересказа

«Рассказы и разговоры о грушах»

Новая методика сбора материала

Четыре участника с заранее распределенными ролями, 4 этапа:

- **Рассказ:** Рассказчик рассказывает сюжет Пересказчику
- **Обсуждение:** Комментатор дополняет, Пересказчик уточняет
- **Пересказ:** Пересказчик пересказывает сюжет Слушателю
- **Письменный пересказ:** Слушатель записывает свой пересказ



«Рассказы и разговоры о грушах»

Оборудование

- рекордер ZOOM H6 Handy Recorder (96 kHz / 24 bit, wav)
- три видеокамеры JAI GO-5000M (100 к/с и 1392x1000, mjpeg)
- видеокамера GoPro Hero 4 (50 к/с и 2700x1500, mp4)
- две пары очков Tobii Glasses II Eye Tracker (50 Hz и 1920x1080)



Речь и жестикуляция: временная координация

Общие сведения

Один из ключевых вопросов мультимодальных исследований – временная координация между иллюстративными мануальными жестами и элементарными дискурсивными единицами

- **Иллюстративные мануальные жесты:** жесты рук, сопровождающие речь, носящие спонтанный характер и не имеющие закрепленной формы и фиксированной связи между означаемым и означающим (Николаева и др. 2015).
- **Элементарные дискурсивные единицы (ЭДЕ):** базовая единица локальной структуры дискурса, выделяющаяся на основании преимущественно просодических критериев и прототипически соответствующая одной клаузе (Кибрик, Подлесская ред. 2009; Kibrik 2015)

Речь и жестикуляция: временная координация

Опережение, синхронизация, отставание

Возможно три варианта времени начала / конца жеста по отношению ко времени начала / конца ЭДЕ:

- опережение
- синхронизация
- отставание

Популярная тематика: специализированные воркшопы и конференции



Речь и жестикуляция: временная координация

Гипотеза Д. МакНилла (McNeill 1992):

Время начала жеста обычно опережает время начала речи

Гипотеза подтверждена на материале трех корпусов:

- английский корпус (Loehr 2012)
- французский корпус (Ferré 2010)
- польский корпус (Karpiński et al. 2009)

Одно из возможных объяснений:

Общий когнитивный источник одновременно запускает активацию как абстрактных семантических репрезентаций, так и более конкретных моторных. Однако время моторного поиска оказывается меньше времени, необходимого для семантического поиска.

Речь и жестикуляция: временная координация

Задача исследования:

Проверить гипотезу об опережении жестов на материале корпуса «Рассказы и разговоры о грушах»

Преимущество нашего корпуса:

Возможность аннотировать видеозаписи mjpeg с частотой 100 к/с

Подкорпус 24 мин.:

- четыре фрагмента по 6 мин. каждый
- все фрагменты были взяты из стадии обсуждения
- аннотации речи и жестов были произведены независимо
- в программе PRAAT было выделено 1673 ЭДЕ
- в программе ELAN было выделено 614 жестов

Медиафайлы и иллюстративный фрагмент

Медиафайлы

В данном исследовании мы использовали 7 медиафайлов из 11:

- один звуковой файл для каждого из трех участников
- одно индивидуальное видео для каждого из трех участников
- одно видео общего плана

Насколько тонкие различия мы изучаем

Можно ли заметить на глаз рассинхронизацию жестов и речи?

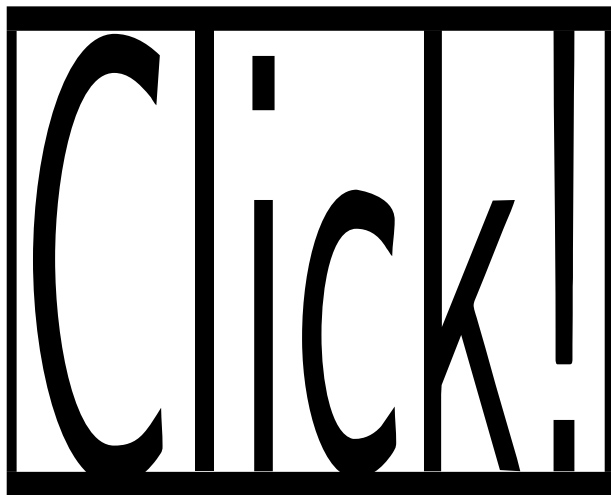
Медиафайлы и иллюстративный фрагмент

Фрагмент 18 секунд, Рассказчик, 22 запись

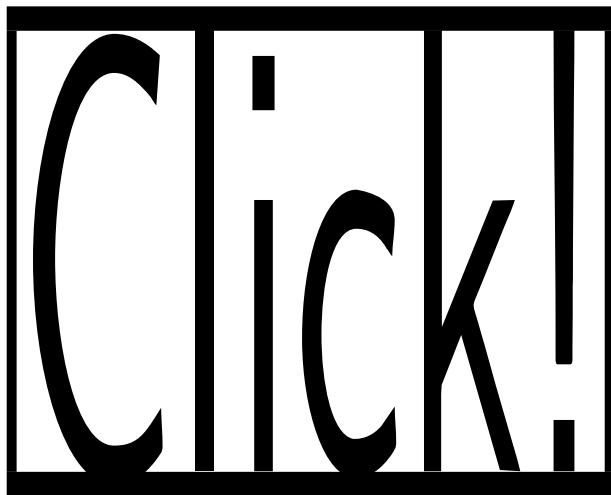
- звуковой файл, 17 ЭДЕ
- индивидуальное видео, 18 жестов
- видео общего плана

Play

Медиафайлы и иллюстративный фрагмент



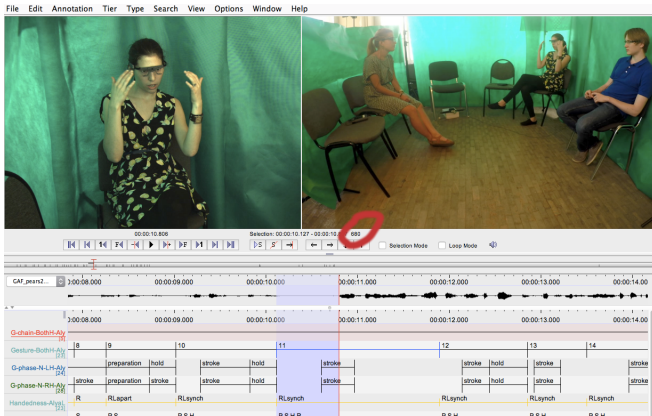
Медиафайлы и иллюстративный фрагмент



Медиафайлы и иллюстративный фрагмент

Аннотация в программе ELAN (А.О. Литвиненко)

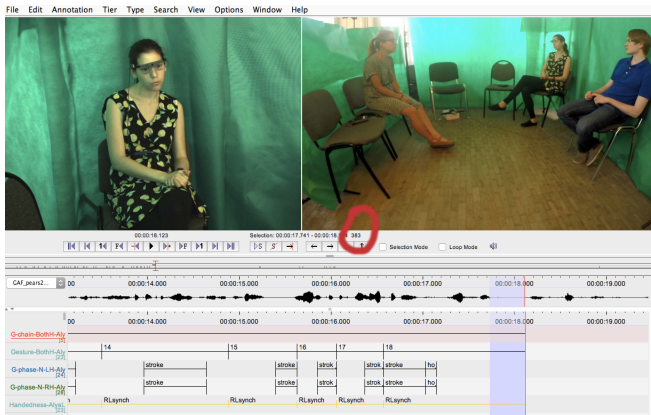
Начало жеста №11 («шляпа») опережает начало ЭДЕ на 680 мс



Медиафайлы и иллюстративный фрагмент

Аннотация в программе ELAN (А.О. Литвиненко)

Конец жеста №18 («горшок») отстает от конца ЭДЕ на 383 мс



Медиафайлы и иллюстративный фрагмент

Аннотация в программе ELAN (А.О. Литвиненко)

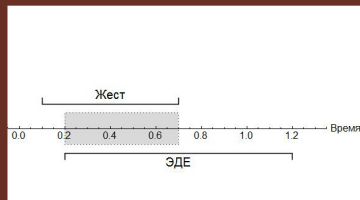
Жест №14 («лицо») полностью совпадает с ЭДЕ

The screenshot displays the ELAN software interface. The top window shows a video of three people in a room with green walls. The bottom window shows a timeline with various annotations. A red circle highlights a specific point on the timeline. The timeline includes tracks for 'CAF_pears2...', 'G-chain-BothH_Aly', 'Gesture-BothH_Aly', 'G-phase-N-LH_Aly', 'G-phase-N-RH_Aly', and 'Handedness-Aly'. The timeline is divided into segments labeled 12, 13, 14, 15, 16, 17, and 18. Segment 14 is highlighted in blue. The timeline also shows 'stroke', 'hold', and 'stroke' annotations for different gestures.

Термины и обозначения

Обозначения

- жест соответствует той ЭДЕ, с которой он имеет наибольшее пересечение и которой соответствует семантически
- длина жеста L_g , длина ЭДЕ L_{edu} , общая часть $L_g \cap L_{edu}$
- временной интервал между началами жеста и ЭДЕ **левый зазор (LGap)**, между концами жеста и ЭДЕ – **правый зазор (RGap)**

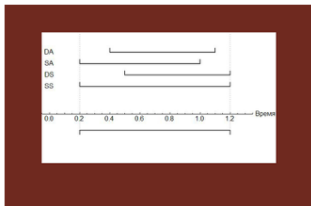


Термины и обозначения

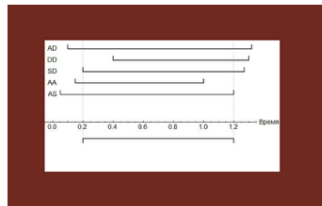
Внутренние и внешние жесты

Мы выделяем 9 логически возможных типов временной координации «жест–ЭДЕ», которые удобно разделить на две группы:

- **внутренние жесты**: жест начинается не раньше начала ЭДЕ и заканчивается не позже
- **внешние жесты**: жест начинается раньше и/или заканчивается позже начала ЭДЕ



Внутренние жесты



Внешние жесты

Термины и обозначения

Допустимая погрешность измерений

Что значит, что границы жеста (не) совпадают с границами ЭДЕ?
Мы производили данные вычисления с допустимой погрешностью
(величиной зазора, признаваемого незначимым) трех типов:

- 50 мс
- 100 мс
- 200 мс

Термины и обозначения

Точность, полнота, среднее гармоническое

Временная координация «жест–ЭДЕ» была количественно оценена при помощи трех мер:

- **Точность (precision, P)**: отношение общей части к длине жеста, $Lg \cap edu / Lg$. Точность показывает, насколько жест вписывается в границы ЭДЕ.
- **Полнота (recall, R)**: отношение общей части к длине ЭДЕ, $Lg \cap edu / Ledu$. Полнота показывает, насколько жест заполняет ЭДЕ.
- **Среднее гармоническое (harmonic mean, HM)**: рассчитывается по формуле $HM = 2PR / (P + R)$. Эту величину можно сравнить с F1-мерой, которая используется в алгоритмах извлечения информации.

Термины и обозначения

Относительная и абсолютная точность

Для проверки гипотезы об опережении жеста относительно ЭДЕ важно различать два способа измерения точности, уже упомянутых раньше:

- **относительную точность** P , вычисляемую по вышеприведенной формуле $Lg \cap edu / Lg$
- **абсолютную точность** при измерении допустимой погрешности в 200 мс, 100 мс и 50 мс

При определении координации «жест–ЭДЕ» относительная и абсолютная точность могут не совпадать. Например, если $Lg \cap edu$ мало, то значение P также будет мало, но если при этом Lg меньше заданной допустимой погрешности, то получается, что границы этого жеста и ЭДЕ совпадают. Для большей надежности результатов нужно использовать оба измерения.

Порядок работы и результаты

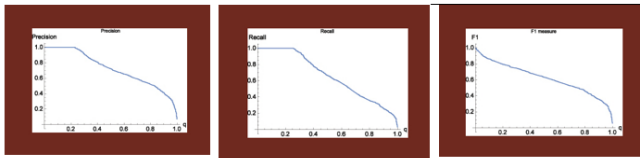
Порядок работы

- для каждой из 1673 выделенных ЭДЕ в программе PRAAT и для каждого из 614 жестов в программе ELAN с точностью до сотой доли секунды были определены время начала и время конца
- для определения временной координации была разработана специальная методика, реализованная в программе **Wolfram Mathematica**
- для каждой пары «жест–ЭДЕ» были вычислены **L-Gap**, **R-Gap**, границы $L_g \cap \text{edu}$, **P**, **R** и **NM**

Порядок работы и результаты

Результаты: меры временной координации

- **относительная точность:** 25% жестов точно входят в границы ЭДЕ, а 85% входят в границы ЭДЕ с P больше 0.5, т.е. не менее половины жеста попадает внутрь ЭДЕ
- **полнота:** 25% жестов полностью покрывают ЭДЕ, а 65% имеют $R = 0.5$, т.е. заполняют более половины ЭДЕ
- **гармоническое среднее:** 20% жестов имеют $HM = 0.8$, а 80% жестов имеют $HM = 0.5$

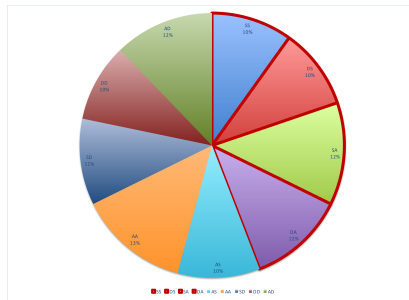


по оси x отложены квантили, по оси y – значения соответствующей меры в диапазоне от 0 до 1

Распределение пар «жест-ЭДЕ» по типам

Допустимая погрешность 200 мс

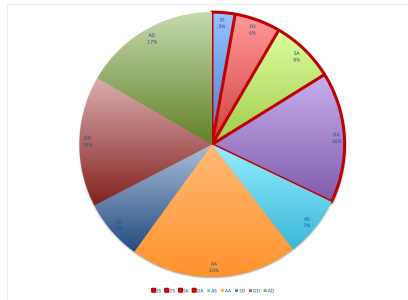
Все типы представлены примерно в равных долях, а число внутренних жестов (SS, DS, SA и DA, в сумме 44%) не намного меньше числа внешних жестов



Распределение пар «жест-ЭДЕ» по типам

Допустимая погрешность 100 мс

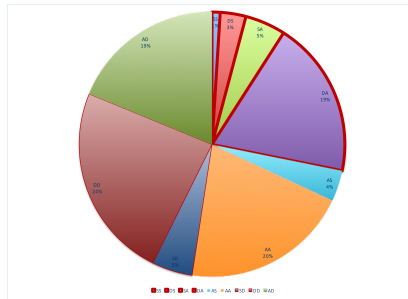
Распределение предсказуемым образом сдвигается в сторону преобладания тех типов, для которых не требуется строгое соответствие границ жеста и ЭДЕ, а количество внутренних жестов уменьшается до **33%**



Распределение пар «жест-ЭДЕ» по типам

Допустимая погрешность 50 мс

Количество внутренних жестов уменьшается до 28%, что уже сравнимо с Р, равной 25%. Как с абсолютной, так и с относительной точностью только четверть жестов вписывается в границы ЭДЕ. Можно предположить, что остальные жесты опережают ЭДЕ.

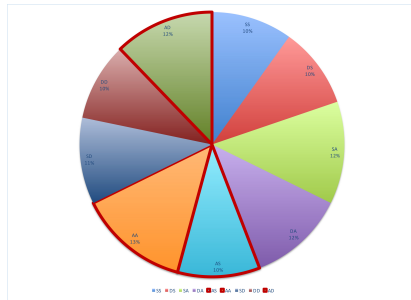


Проверка гипотезы об опережении жестов

Допустимая погрешность 200 мс: 35% жестов опережают речь

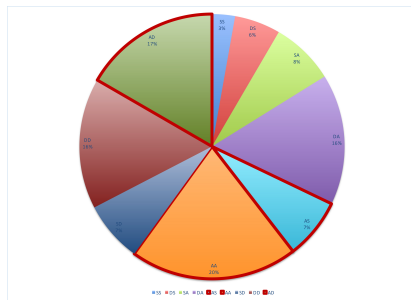
Выделенные типы были поделены на две группы:

- начало жеста опережает начало ЭДЕ (AS, AA и AD)
- начало жеста синхронизировано или отстает от начала ЭДЕ



Проверка гипотезы об опережении жестов

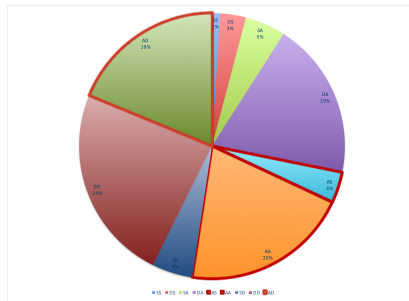
Допустимая погрешность 100 мс: 44% жестов опережают речь



Проверка гипотезы об опережении жестов

Допустимая погрешность 50 мс: 43% жестов опережают речь

Количество опережающих жестов перестает увеличиваться и не достигает даже 50%. Полученные результаты не подтверждают гипотезу об опережающем производстве жестов.



Обсуждение результатов: сравнение корпусов

В нашем корпусе опережение жеста относительно ЭДЕ:
в 35% (200 мс), 44% (100 мс) и 43% (50 мс)

- польский корпус (Karpiński et al. 2009): 69%
- французский корпус (Ferré 2010): 70%
- английский корпус (Loehr 2012): данные не приведены

Чем могут быть вызваны такие большие различия в результатах?

Обсуждение результатов: сравнение корпусов

Чем могут быть вызваны такие большие различия в результатах:

- в других корпусах аналогичные просодические единицы: Intonation Phrases в (Ferré 2010), Intermediate Phrases в (Loehr 2012) и Major Intonation Phrases в (Karpiński et al. 2009)
- размер нашего корпуса (614 жестов и 1673 ЭДЕ) больше: английский корпус включал 164с записи, французский – 244 жеста, польский – 223 жеста и 773 MajorIP
- наша частота 100 к/с в 4 раза выше, чем в других корпусах
- точность измерения погрешности сопоставимая

Обсуждение результатов: объяснение различий

Функциональные классы жестов

Иллюстративные жесты различаются по функциям, которые они выполняют в процессе коммуникации. Согласно классификации МакНилла (McNeill 1992) их можно разделить на:

- указательные жесты, выполняющие референцию к объекту
- иконические жесты, изображающие конкретные объекты или действия
- метафорические жесты, представляющие абстрактные понятия
- ритмические жесты, выделяющие фрагменты речи

Обсуждение результатов: объяснение различий

Функциональные классы жестов

Распределение жестов по функциональным классам сильно зависит от жанра дискурса и индивидуальных особенностей говорящих.

Ритмические жесты в корпусе МакНилла составляли 44.4% всех жестов (McNeill 1992), в корпусе из работы (Theune, Brandhorst 2010) – 32.1%, а в корпусе Николаевой (2013) – всего 15%. Из этих цифр следует, что ритмические жесты, обычно короткие и максимально синхронизированные с речью, могут оказывать сильное влияние на результаты временной координации по корпусу в целом. Известно также, что в диалоге количество ритмических жестов увеличивается (Bavelas et al. 1992). Таким образом, на результаты могло повлиять различие в распределении по функциональным классам жестов.

Дальнейшие исследования

Задачи развития данного исследования в будущем состоят в:

- увеличении размера подкорпуса
- временной привязке каждого слова
- учете внутренних особенностей ЭДЕ (их синтаксической, коммуникативной и интонационной структуры)
- разделении жестов на функциональные типы
- разделении жестов на подготовительную, маховую и ретракционную фазы (Kendon 1980)
- анализе временной координации не только пар «жест–ЭДЕ», но и «маховая фаза–слово»

«Рассказы и разговоры о грушах»

Речь и жестикуляция: временная координация

Медиафайлы и иллюстративный фрагмент

Термины и обозначения

Порядок работы и результаты

Обсуждение результатов и дальнейшие исследования

Обсуждение результатов

Функциональные классы жестов

Дальнейшие исследования

Спасибо за внимание!

olga.fedorova@msu.ru