

«Рассказы и разговоры о грушах»: принципы вокальной аннотации

(Версия 10.01.2019)

Н. А. Коротаев¹

Вокальная аннотация корпуса включает в себя разметку двух коммуникативных каналов — вербального и просодического (см. [Кибрик 2018]). В процессе аннотирования решаются две крупные группы задач:

- (i) Выполняется сегментация речевого потока на значимые единицы различных уровней (элементарные дискурсивные единицы, слова, заполненные паузы и др.); для всех выделенных единиц указываются их точные временные границы. Аналогичному анализу подвергаются и разнообразные неречевые вокальные действия участников, смех, свист, цоканье и т. д.
- (ii) Размечаются свойства речевых единиц, наиболее существенные для описания локальной дискурсивной структуры устного текста: просодические характеристики, иллюкутивно-фазовые значения, способы преодоления речевых сбоев и проч.

С содержательной точки зрения используемая разметка наследует принципам, описанным в коллективной монографии [Кибрик, Подлесская 2009 (ред)], см. также <http://www.spokencorpora.ru/>, [Kibrik et al. 2019 in print]. Вместе с тем при аннотировании корпуса выработан ряд новых решений и формальных приемов — обоснованных, прежде всего, диалогическим и мультиканальным характером размечаемых записей. Цель настоящего текста — дать краткое описание общих принципов и этапов разметки, а также пояснить все обозначения, используемые в вокальной аннотации. Более подробный содержательный анализ обозначаемых явлений можно обнаружить в вышеуказанных работах, а также в других публикациях нашей научной группы.

1. Процедура аннотирования и формат индивидуальных аннотаций

1.1. Общие замечания

Вокальное поведение каждого из трех основных участников записи (Рассказчика, Комментатора, Пересказчика) аннотируется *по отдельности*. (Ниже, в разделе 7, описывается формат единого представления вокальных действий всех трех участников записи, т. н. партитурная нотация. Однако такие транскрипты не редактируются в процессе аннотирования, а создаются уже после реализации индивидуальных разметок.)

Соответственно, главным материалом для аннотирования являются индивидуальные аудиофайлы, записанные с микрофонов отдельных участников (см. <http://multidiscourse.ru/solutions/>). В качестве вспомогательных материалов также используются общий аудиофайл и видеофайлы с наложенным звуком. При прослушивании общего аудиофайла может стать более понятным контекст коммуникативного обмена: если в каждом индивидуальном аудиофайле вклад одного говорящего может заглушать вокальные действия других участников, то в общем аудиофайле громкость речи трех говорящих выровнена. Сравнение индивидуальных аудиофайлов с общим особенно актуально при анализе диалогических фрагментов. Кроме того, общий аудиофайл содержит полезную информацию об интегральных характеристиках записи (см. ниже, подраздел 1.2). Обращение

¹ При ссылке на данный документ, пожалуйста, используйте следующий формат:

Коротаев Н.А. «Рассказы и разговоры о грушах»: принципы вокальной аннотации. Версия 10.01.2019.
<http://multidiscourse.ru>

к индивидуальным видеофайлам также помогает точнее интерпретировать вокальные действия участников: некоторые недостаточно разборчивые сегменты становятся понятными, если проанализировать движения губ и общее кинетическое поведение; на видео легче определить источник ряда звуков, не относящихся к вокальному поведению (почесывания, случайные прикосновения к микрофону, хлопки ладонями и т. д.). Тем не менее безусловно приоритетным методом аннотирования является перцептивный и инструментальный анализ индивидуальных аудиофайлов.

Для прослушивания аудиофайлов и анализа их акустических характеристик мы рекомендуем использовать свободно распространяемую программу Praat. Программа снабжена подробной справочной системой; читателям, не знакомым с общими принципами работы в Praat и / или с особенностями создания и редактирования текстгридов (см. ниже, разделы 2-3), мы предлагаем самостоятельно ознакомиться со справкой или посетить сайт разработчиков <http://www.fon.hum.uva.nl/praat/>.

1.2. Этапы разметки

Процедура вокальной разметки разбивается на несколько этапов, точное наполнение которых в существенной степени зависит от количества аннотаторов и соотношения их квалификации. Ниже описана схема, при которой сначала с записью работают менее опытные разметчики, а затем их результаты обрабатывают и дополняют более опытные аннотаторы.

Предварительный этап: первое знакомство с записью

До начала непосредственной аннотации мы рекомендуем прослушать целиком общий аудиофайл. Это самый удобный способ ознакомиться с общим содержанием записи, манерой речи говорящих, с местами, содержащими наложения, и т. д. Манера речи может быть связана с общими коммуникативными установками участников, поэтому полезно также полностью или частично просмотреть общий видеофайл. Наблюдения, сделанные при таком предварительном знакомстве с записью, играют важную роль в дальнейшем аннотировании. Поскольку основная разметка основана на анализе относительно коротких речевых отрезков, представление об интегральных характеристиках всей записи и ее крупных фрагментов позволяет рассматривать конкретные явления не изолированно друг от друга, а как проявления общих стратегий конкретных говорящих и / или особенностей взаимодействия участников на конкретном этапе записи.

Первый этап: первичная расшифровка, временная привязка, предварительная сегментация на ЭДЕ

На первом этапе аннотирования в индивидуальных аудиофайлах выделяются отдельные слова, заполненные паузы и некоторые виды неречевых вокальных действий. Для всех этих единиц, которые мы в дальнейшем будем называть единицами *нижнего уровня сегментации*, устанавливаются их временные границы, т. е. точки начала и конца звучания относительно начала звукового файла. Особое внимание необходимо уделить тому, чтобы каждая единица была приписана правильному говорящему, а также тому, чтобы ни одна единица не была пропущена. Для этого аннотаторам необходимо внимательно и целиком прослушивать все индивидуальные аудиофайлы — в том числе и те их фрагменты, в которых соответствующие участники преимущественно молчат по условиям записи. Так, например, Пересказчики, которых перед началом записи просили не перебивать Рассказчиков на этапе Рассказа, в действительности регулярно подают сигналы обратной связи, которые необходимо фиксировать в разметке. В сложных случаях мы рекомендуем обращаться к общему аудиофайлу и / или общему и индивидуальным видеофайлам. Конкретные правила записи единиц нижнего уровня сегментации и общие принципы их временной привязки рассматриваются ниже в разделе 2.

Кроме того, уже на первом этапе аннотирования производится предварительная сегментация речевого потока на *элементарные дискурсивные единицы* (далее — ЭДЕ). ЭДЕ является центральным понятием применяемой нами системы дискурсивной транскрипции. Каждая ЭДЕ представляет собой минимальный шаг в развитии дискурса; подробнее см. раздел 3 и упоминаемые там ссылки на литературу. С точки зрения используемой в проекте аннотационной схемы (см. документ «Схема мультисканальной аннотации в ELAN», http://multidiscourse.ru/data/ann/pears_multichannel_annotation.pdf) ЭДЕ являются единицами *верхнего уровня сегментации*.

Второй этап: уточнение сегментации на ЭДЕ, подробная просодическая и дискурсивная разметка

На втором этапе уточняются границы ЭДЕ и прочих единиц верхнего уровня сегментации, (см. раздел 3), при необходимости вносятся коррективы в состав и временные границы единиц нижнего уровня сегментации, а также реализуется подробная просодическая и дискурсивная разметка ЭДЕ и входящих в них слов (раздел 4). Кроме того, дополнительно выявляются накладывающиеся на речь вокальные явления: смех, улыбка, скрипучий голос (раздел 5).

Как уже было отмечено выше, приведенная схема актуальна для ситуации, при которой работа по аннотированию разделяется между более и менее опытными разметчиками. Если процесс разметки не подразумевает такого разделения, то первый и второй этапы могут производиться не последовательно, а параллельно. В то же время результат первого этапа (запись всех единиц нижнего уровня сегментации, указание их временных границ и их предварительное объединение в ЭДЕ) уже можно рассматривать как неполную, но внутренне непротиворечивую разметку.

1.3. Итоговый формат индивидуальных аннотаций

Итоговая разметка вокального поведения каждого из трех участников записи сохраняется в двух форматах.

- (i) Информация о временных границах каждой единицы хранится в файлах формата textgrid (далее — *текстгриды*), создавать и редактировать которые можно в программе Praat.
- (ii) Подробная просодическая и дискурсивная аннотация создается и редактируется в формате doc, реализованном в среде MS Word. Возможна также и работа в аналогичном свободно распространяемом текстовом редакторе, однако оптимальные результаты достигаются именно в Word.

На рис. 1 приведен скриншот текстгрида, соответствующего фрагменту разметки вокальных действий Пересказчицы в записи 22.

Информация в текстгриде разнесена по трем слоям:

- (i) В слое EDU размечаются границы ЭДЕ и прочих единиц верхнего уровня сегментации. Здесь же, при помощи отдельной нумерации, аннотируются абсолютные паузы между единицами. Подробнее о содержательных принципах разметки этих единиц см. раздел 3.
- (ii) В слое Words размечаются границы слов, заполненных пауз и прочих единиц нижнего уровня сегментации (раздел 2).
- (iii) В слое Collat размечаются границы неречевых вокальных явлений, сопутствующих речи (раздел 5).

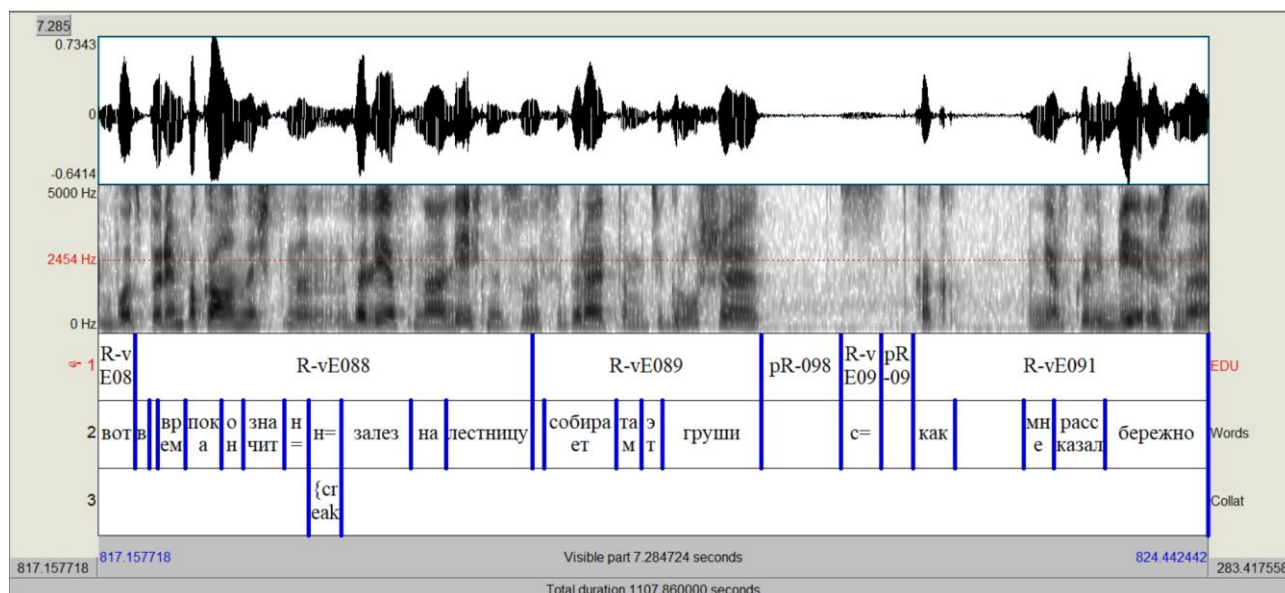


Рисунок 1. Образец текстгрида (фрагмент записи 22, Пересказчица)

В текстовых вокальных транскриптах информация о единицах, размечаемых в слоях EDU и Words, представлена в табличном формате. В приложении 2 приведен фрагмент индивидуального транскрипта Пересказчицы в записи 22, начинающийся с места, зафиксированного рис. 1. Каждая строка таблицы соответствует единице верхнего уровня сегментации, т. е. аналогична интервалу слоя EDU в текстгриде. Для каждой строки в отдельных столбцах указывается:

- (i) Время начала соответствующего интервала с точностью до сотых долей секунды (столбец Time).
- (ii) Кодовый номер единицы верхнего уровня сегментации (столбец EDU).
- (iii) Транскрипционная запись единиц нижнего уровня сегментации, входящих в данную единицу верхнего уровня (столбец Transcription; подробнее см. раздел 4 и приложение 1).
- (iv) Комментарии в свободной форме (столбец Comments).

Обозначения, используемые в транскриптах, обсуждаются ниже в разделах 2–4, а также суммированы в таблице в приложении 1.

1.4. Технические детали

Как можно заметить из приведенных примеров, в текстгридах и транскриптах используются специальные правила нумерации единиц верхнего уровня сегментации (подробнее см. подраздел 3.3). Кроме того, для поддержания консистентности данных жизненно необходимо, чтобы информация в текстгридах и транскриптах полностью соответствовала друг другу. Для более быстрого и надежного решения этих, а также некоторых других задач в рамках проекта разработаны дополнительные программные продукты, в частности (см. подразделы 3.3 и 3.4):

- (i) скрипт Praat и макрос VBA, выполняющие полуавтоматическую нумерацию единиц верхнего уровня сегментации соответственно в текстгридах и текстовых транскриптах;
- (ii) скрипт Praat, проверяющий, что все границы интервалов в слое EDU совпадают с теми или иными границами интервалов в слое Words;
- (iii) макрос VBA, проверяющий консистентность данных в текстгридах и текстовых транскриптах, а также экспортирующий в транскрипт данные о временных характеристиках единиц.

Кроме того, средства автоматизации используются также для конвертации вокальной разметки в единую структуру мультисканальной аннотации в ELAN (см. раздел 6) и при формировании партитурных вокальных транскриптов (см. раздел 7).

2. Нижний уровень сегментации: базовая разметка

В этом разделе рассматриваются правила базовой разметки слов и других единиц нижнего уровня сегментации. Мы выделяем следующие типы таких единиц:

- слова и их аналоги;
- заполненные паузы;
- смех и прочие неречевые вокальные действия;
- паузы, заполненные шумными вдохами;
- абсолютные паузы.

Далее приводятся правила записи этих единиц и обсуждаются принципы их временной привязки в Praat.

2.1. Правила записи

Правила записи единиц нижнего уровня сегментации действуют при оформлении интервалов слоя Words в текстгридах, а также при первичном заполнении столбца Transcription в текстовых транскриптах. Внимание: на последующих этапах аннотирования запись единиц в транскриптах может модифицироваться — см. раздел 4!

2.1.1. Запись слов

1. Слова записываются в стандартной орфографической форме (но см. пункты 6-7). Редкими исключениями являются особенно частотные редуцированные произнесения, резко расходящиеся с орфографической нормой: *щас*, *здрасьте* и т. д. Однако и эти единицы лучше записывать в стандартной орфографии, снабжая их комментариями в транскрипте. То же верно и для случаев окказионального нестандартного произнесения (здесь и далее приводится условная «псевдотранскрипционная» запись; настоящий, табличный, вид текстового транскрипта представлен в приложении 2):

он пытается их пересчитать //произносит «пересчиДать»

Также используется конвенциализованная запись для сигналов подтверждения: *угу* и *ага*. В том случае если говорящий произносит эти сигналы не обычным образом, а «полно» (т. е. как если бы они действительно состояли из звуков, соответствующих этому условному написанию), это указывается в столбце комментариев:

ага //произносит полно: [ага]

2. Во всех случаях, где это допускается стандартной орфографией, обязательно используется буква Ё:

берёт, своём, неё и т. д.

3. Вместе с тем все формы слов записываются так, как они произносятся участниками — например, в «неправильных» падежах, числовых, видо-временных формах и т. д. Не допускается никакой подгонки материала под представления аннотатора о том, «как должно быть». Например (здесь и далее выделение красным используется для привлечения внимания, это НЕ транскрипционное обозначение!):

и он (э) вторую точнее дополняет,
д= (э) от= оставшихся грушами,

4. Как отдельные слова фиксируются все повторы, в том числе и неполные:

на на свой ба= <какой-то> багажник?

то есть он пере= пересчитывает сначала

5. Оборванные фрагменты не произнесенных до конца слов также нотируются как отдельные единицы; наличие обрыва обозначается знаком равенства без пробела:

он он он какой-то пож= он пожилой допустим?

При этом если аннотатор с достаточной уверенностью полагает, что говорящий планировал произнести какое-то конкретное слово, оборванный фрагмент записывается по правилам стандартной орфографии:

поэтому кажется сначала,

ч= чт= что её там нет,

Если же аннотатор не может с уверенностью понять, что именно оборвано, фрагмент записывается «как слышится»:

Юра правильно сказал ш= про свист мальчиков,

См. также запись оборванных слов в строках R-vE088, R-vE090 и R-vE098 транскрипта в приложении 2.

6. Если при написании слова в стандартной орфографии используется дефис и при этом каждая из частей этого сложного слова произносится с самостоятельным ударением, части признаются отдельными словами — в текстгриде им соответствуют отдельные интервалы в слое Words, а в транскрипте они разделяются пробелами. При этом дефисы в написании сохраняются:

Мужчина

где-то лет сорока- пятидесяти

7. Если, наоборот, последовательность орфографических слов произносится настолько слитно, что попытка провести между ними границу, сохранив орфографическую форму, не имеет смысла, вся последовательность признается одним словом и записывается через нижнее подчеркивание:

в_смысле да,

8. Если слово произносится настолько нечетко, что аннотатор не уверен, что произнесено именно оно, эта единица заключается в <угловые скобки>:

что Юра <вот> как раз рассказал,

Если неуверенность касается последовательности слов, в угловые скобки заключается каждое из слов этой последовательности:

<всё> <там>.

9. Если аннотатор четко слышит сегментный состав слова, но не может его содержательно проинтерпретировать (даже в качестве оборванного фрагмента), такая единица заключается в >обратные угловые скобки<:

а потом >пст< увидела кошку,

Это же обозначение используется для тех случаев, когда нестандартное произнесение поддается интерпретации, но при этом явно случайным — и слишком заметным — образом отклоняется от обычного:

а эти трое ребят пошли в_сторону >фервела<. //Очевидно, имеется в виду «фермера».

10. Для полностью неразборчивых фрагментов используется обозначение <НРЗБ>:

потому что он как-то пытается за= сесть на этот <НРЗБ>

2.1.2. Запись звукоподражаний

Звукоподражания мы условно считаем подклассом слов. Они заключаются в #дизеы# и записываются «как слышатся»:

мальчик едет на велосипеде,

#ты-дыш# #ты-дыш# #ты-дыш# #ты-дыш#!

Если звукоподражание не поддается сегментной записи, в дизеы заключается описательное выражение:

#рычание#

#блеянье#

Также в дизеы заключаются не поддающиеся орфографической записи неречевые вокальные явления, используемые в функции лексических единиц:

и он такой #свист#

2.1.3. Запись заполненных пауз

Заполненные паузы — это вокализованные отрезки, которыми говорящий заполняет время хезитации. Жесткой границы между этими единицами и «нормальными» словами, видимо, не существует (см., в частности, работу [Clark, Fox Tree 2002], в которой аналогичные элементы английского языка рассматриваются как полноценные лексические единицы), однако в любом случае заполненные паузы образуют отдельный класс — относительно небольшой по объему, но существенный с точки зрения частоты использования. Каждый говорящий заполняет время хезитации немного по-своему, но с известной долей условности можно выделить четыре типа простых заполненных пауз:

Тип паузы	Обозначение	Код символа в скобках (Юникод)
пауза, заполненная мэканьем	(ш)	026F
пауза, заполненная эканьем	(э)	0259
пауза, заполненная аканьем	(е)	0250
пауза, заполненная гортанным скрипом	(?)	02C0

Таблица 1. Обозначения заполненных пауз

Кроме того, возможны различные комбинации при заполнении одного вокализованного отрезка хезитации: (шэ), (эш) и т. д. В этих случаях один звук переходит в другой и между разными частями нет никакого промежутка и / или прерывания.

Примечание. В итоговых транскриптах обозначения заполненных пауз сопровождаются указанием на их длительность: (ш 0.23), (еш 0.40) и т. д. Это результат автоматического экспорта данных из текстридов. То же верно для рассматриваемых ниже смеха и прочих неречевых вокальных действий, абсолютных пауз и пауз, заполненных шумными вдохами.

2.1.4. Запись смеха и прочих неречевых вокальных действий

Помимо слов и заполненных пауз, участники могут производить разнообразные неречевые вокальные действия: цокать языком, чмокать и т. д. Все эти явления записываются в {фигурных скобках} и кодируются следующим образом:

Тип неречевого вокального действия	Обозначение
откашливание	{exp}
цоканье языком	{cl}
чмоканье губами	{sm}
вздых	{sg}
сглатывание	{gp}
свист	{wh}
фырканье	{st}
кашель	{cg}
шмыганье носом	{sf}
хмыканье	{hm}
смех	{laugh}

Таблица 2. Обозначения неречевых вокальных действий

Смех несколько выделяется в ряду неречевых вокальных действий, поскольку может как присутствовать в виде отдельного сегмента, так и накладываться на речь или говорящего. В первом случае смех записывается так, как указано в табл. 2. Во втором случае смех фиксируется в слое Collat текстгрида и не размечается в транскрипте (см. раздел 5).

2.1.5. Запись абсолютных пауз и пауз, заполненных шумными вдохами

Абсолютные паузы, т. е. отсутствие какой бы то ни было вокализации у данного говорящего, выявляются «по остаточному принципу» (но см. некоторые подробности в разделе 2.2 ниже). В текстгридах абсолютным паузам соответствуют пустые интервалы в слое Words (см. пустой интервал под единицей pR-021 слоя EDU на рис. 1), в транскриптах — пустые скобки: ().

Важно помнить, что при аннотировании вокального поведения конкретного говорящего под абсолютной паузой понимается отрезок молчания *именно этого говорящего*: другие участники записи в этот момент могут произносить свои реплики. Вопрос о выделении *совместных абсолютных пауз* решается уже при формировании партитурной нотации, см. раздел 7.

С абсолютными паузами формально и функционально сближаются паузы, заполненные шумными вдохами. Вдохи следует отличать от вздохов: при *вздохе* участник шумно набирает воздух, а затем сразу же столь же шумно выдыхает; за шумным *вдохом* дальше обычно следует речь либо простой выдох. Вдох часто свидетельствует о стремлении участника начать свою реплику или перехватить очередь хода в диалоге. Паузы, содержащие шумные вдохи, обозначаются как (ц), с символом 0265 Юникода в скобках.

2.2. Принципы временной привязки

Под временной привязкой вокальных единиц мы понимаем установление их точных границ, т. е. определение времени начала и конца соответствующих интервалов в слое Words. При установлении границ аннотаторы должны опираться на слуховое восприятие и на акустические представления звукового потока, реализованные в программе Praat:

- осциллограмму (Waveform);
- спектрограмму (Spectrum);

- тонограмму (Pitch);
- график интенсивности (Intensity).

Наиболее плодотворной является опора на слуховое восприятие и спектрограмму, но все же ни одно из упомянутых выше средств не может служить абсолютно точным индикатором границы. Тем не менее совокупность нескольких параметров чаще всего позволяет с достаточной точностью правильно определить место границы.

Ниже, на рис. 2, приведен фрагмент текстгрида с выполненной временной привязкой единиц нижнего уровня сегментации для аудиофайла Рассказчицы 22-й записи. На нем, в частности, можно заметить следующее.

- (i) Абсолютным паузам (пустым интервалам в слое Words) соответствуют практически ровные линии на осциллограмме, отсутствие выделенных областей на спектрограмме и полное прерывание тоновой кривой.
- (ii) Паузе, содержащей шумный вдох (интервал (ц) в слое Words), соответствуют минимальные колебания на осциллограмме, полное прерывание тоновой кривой и характерное заполнение спектрограммы.
- (iii) Паузе, заполненной гортанным скрипом (?), соответствует характерный прерывистый график осциллограммы и следы смычек на спектрограмме.
- (iv) Паузе, заполненной эканьем (э), соответствуют вокалические осциллограмма и спектрограмма, а также ровное движение частоты основного тона в нейтральном для данного говорящего уровне.
- (v) В большинстве интервалов, соответствующих словам, на спектрограмме обнаруживается формантная структура для гласных и сонорных согласных звуков, а также характерные симптомы взрывных и щелевых согласных.
- (vi) Для гласных и сонорных сегментов строится кривая изменения частоты основного тона.

В стандартном случае наблюдаемые на графиках формальные характеристики подтверждаются перцептивно — и это способствует воспроизводимому проведению границ.

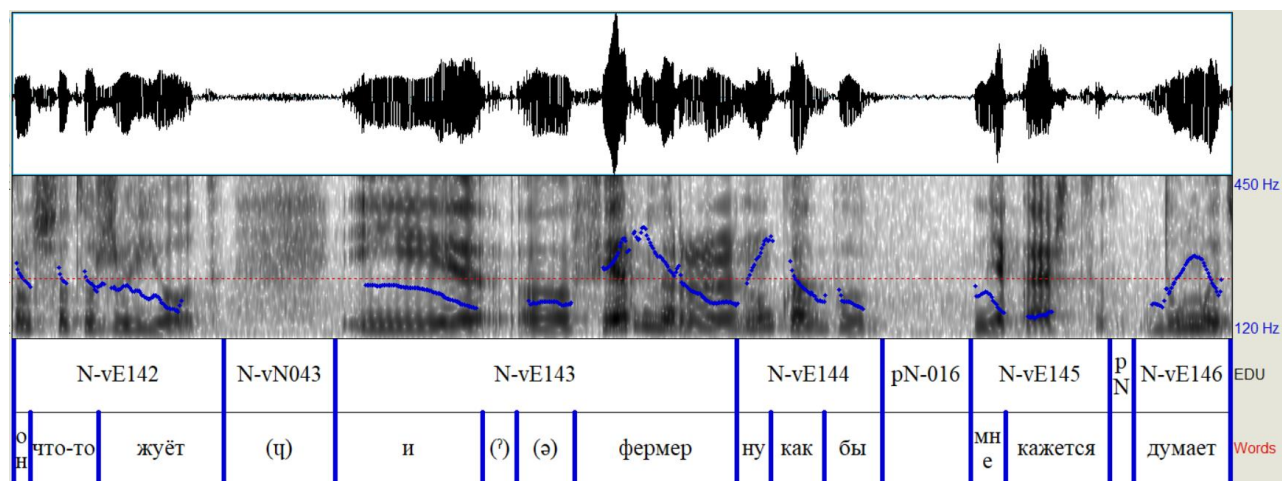


Рисунок 2. Фрагмент текстгрида записи 22 (Рассказчица) с указанием точных границ единиц нижнего уровня сегментации в слое Words

2.2.1. Некоторые важные замечания

1. Поскольку одной из базовых задач проекта является описание того, как осуществляется координация разных каналов коммуникации, устанавливать временные границы интервалов требуется с особой точностью. В оптимальной ситуации разметкой границ для одного

аудиофайла должны заниматься два или более аннотаторов — и пренебрежимое различие между их вариантами разметки не должно превышать 5 миллисекунд.

2. При проверке границ на слух практически бесполезно прослушивать очень маленькие интервалы, поскольку даже в паузах все равно присутствует небольшой шум, а микрофрагменты звуков меньше определенной длины перестают определяться на слух как фонемы. Поэтому следует прослушивать бóльшие фрагменты до и после проверяемой границы, а затем — фрагмент, в котором потенциальная граница находится примерно посередине, обращая внимание на то, как и в каком месте существенно меняется качество звука.

3. При работе со спектрограммой полезно обращать внимание не только на формантную структуру, но и на такие признаки, как, например, размыкание голосовых связок при начале говорения: достаточно часто это позволяет не пропустить начало произнесения единицы.

4. В начале и конце слов могут встречаться различные смычки, гласные призвуки, придыхания. Эти явления дополнительно аннотируются в текстовых транскриптах (см. подраздел 4.7.6), но не размечаются в текстгридах. Тем не менее они должны включаться в границы соответствующих интервалов слоя Words.

5. Абсолютные паузы между словами и прочими единицами нижнего уровня сегментации теоретически могут быть любой длительности, даже очень малой, однако чем меньше интервал, тем сложнее доказательно отличить паузу от глухой смычки или другого подобного явления. Поэтому при разметке абсолютных пауз мы предлагаем придерживаться следующей схемы.

- (i) Паузы протяженностью 0.1 секунды и более выделяются всегда.
- (ii) Паузы протяженностью менее 0.1 секунды выделяются только тогда, когда (а) пауза определяется как таковая на слух и (б) слева и справа от паузы находятся звуки, границы которых могут быть достоверно определены без сомнений.
- (iii) Если потенциальная пауза менее 0.1 секунды плохо определяется на слух, а слева и справа от нее находятся глухие согласные либо еще какие-то явления затрудняют определение ее границ (например, речь других говорящих, посторонние шумы и т. п.), то пауза не выделяется. Межсловная граница в таком случае ставится в самой левой точке, определяемой по спектрограмме и на слух.

2.2.2. Автоматизация временной привязки

Разметка временных границ слов и прочих единиц нижнего уровня сегментации может быть частично автоматизирована. В рамках проекта мы опираемся на результаты автоматического определения границ, выполненного нашими партнерами из компании STEL (<http://speech.stel.ru/>). На вход разработанной STEL процедуры подаются первичные текстовые расшифровки, на выходе порождаются файлы с указанием времени начала и конца произнесения каждого расшифрованного сегмента. При помощи дополнительного скрипта результирующие файлы преобразуются в текстгриды требуемого формата.

Частичная автоматизация временной разметки существенно ускоряет работу, но все же не отменяет необходимости дальнейшей ручной проверки. Неточности в автоматической разметке могут быть вызваны различными факторами, в том числе:

- трудностями в определении способа заполнения паузы;
- редуцированным и / или тихим произнесением;
- помехами со стороны других говорящих и посторонними шумами.

Проверка результатов автоматической разметки должна производиться в соответствии с принципами, описанными выше в настоящем разделе.

3. Верхний уровень сегментации: базовая разметка

3.1. Элементарные дискурсивные единицы

Центральным понятием используемой в проекте системы описания устной речи является *элементарная дискурсивная единица* (ЭДЕ). ЭДЕ — это базовая единица верхнего уровня сегментации; в каждой ЭДЕ реализуется один минимальный шаг в порождении дискурса. В стандартном случае элементарный характер этих единиц согласованно проявляется на различных уровнях речепроизводства: физиологическом, просодическом, семантико-синтаксическом, когнитивном — см. [Chafe 1994, Кибрик, Подлесская (ред.) 2009, Kibrik et al. in print].

В процессе вокального аннотирования корпуса для сегментации речевого потока на ЭДЕ в первую очередь учитывается набор просодических критериев:

- громкостный паттерн (начало ЭДЕ обычно произносится громче, чем завершение);
- скоростной паттерн (начало ЭДЕ часто произносится быстрее, чем завершение);
- наличие единого интонационного контура и акцентных пиков.

Наиболее значимым является последний из перечисленных выше критериев. Кроме того, в большинстве случаев полезен и коммуникативный подход, согласно которому границы ЭДЕ неслучайным образом соотносятся с границами коммуникативно-просодических составляющих, см. [Коротаев 2015].

Каждая ЭДЕ оформляется как отдельный интервал в слое EDU текстрида и как отдельная строка в вокальном текстовом транскрипте — см. фрагмент в приложении 2. ЭДЕ, выделяемые на просодических основаниях, неслучайным образом коррелируют с простыми клаузами. Так, в приложении 2 из 15 ЭДЕ девять соответствуют этому синтаксическому формату и лишь шесть (R-vE087, R-vE090, R-vE092, R-vE096, R-vE099, R-vE101; о принципах нумерации см. подраздел 3.3) заведомо не являются клаузами.

3.2. ЭДЕ и другие единицы верхнего уровня сегментации

В состав ЭДЕ могут входить любые единицы нижнего уровня сегментации: слова, абсолютные и заполненные паузы (см. строки R-vE093, R-vE094 в приложении 2), неречевые вокальные явления (R-vE095). Помимо ЭДЕ, мы выделяем и другие единицы верхнего уровня сегментации, которые обладают значительно менее разнообразной внутренней структурой. Такие единицы могут состоять из изолированных (т. е. не входящих ни в какие ЭДЕ) заполненных пауз или их кластеров; пауз, заполненных шумными вдохами; изолированного смеха; прочих неречевых вокальных действий. Все они также оформляются в виде отдельных строк текстового транскрипта и отдельных интервалов в слое EDU текстрида.

При проведении границ между ЭДЕ и другими единицами верхнего уровня сегментации используются следующие соглашения.

1. Каждый смех, не располагающийся *внутри* какой-либо ЭДЕ (т. е. не окруженный с *обеих сторон* элементами, входящими в одну ЭДЕ), формирует отдельную единицу верхнего уровня сегментации (см. строки R-vL024, R-vL025 в приложении 2).
2. Каждая непрерывная последовательность прочих неречевых вокальных действий и / или пауз, заполненных шумными вдохами, не располагающаяся *внутри* какой-либо ЭДЕ, также формирует отдельную единицу верхнего уровня сегментации (см. R-vN020, R-vN021, R-vN022, R-vN023).
3. ЭДЕ может начинаться с заполненной паузы или группы заполненных пауз, если непосредственно за последней паузой следует первое слово ЭДЕ (см. R-vE099). Если же

заполненная пауза отделена от первого слова ЭДЕ абсолютной паузой или какой-либо другой единицей, то эта заполненная пауза формирует отдельную единицу верхнего уровня сегментации (см. R-vF005, R-vF006).

4. Все абсолютные паузы, располагающиеся между выделяемыми по вышеуказанным правилам единицами верхнего уровня, также записываются в отдельных строках транскрипта (см. pR-098 — pR-104) и оформляются в виде непустых интервалов слоя EDU в текстгриде. О важности записи абсолютных пауз в отдельных строках см. также раздел 7, посвященный партитурной нотации.

3.3. Границы и нумерация

При сегментации на единицы верхнего уровня слой EDU в текстгриде *без остатка* членится на непустые интервалы. При этом необходимо следить за тем, чтобы границы каждой единицы верхнего уровня полностью совпадали с границами некоторых единиц нижнего уровня. Иными словами, нельзя допускать, чтобы один интервал слоя Words имел непустое пересечение более чем с одним интервалом слоя EDU. На рис. 3 представлен скриншот, на котором видно, что правая граница ЭДЕ R-vE125 не совпадает ни с одной из границ слов мужской и как. Такая ситуация является недопустимой и обязательно исправляется при технической проверке аннотации (см. подраздел 1.4).

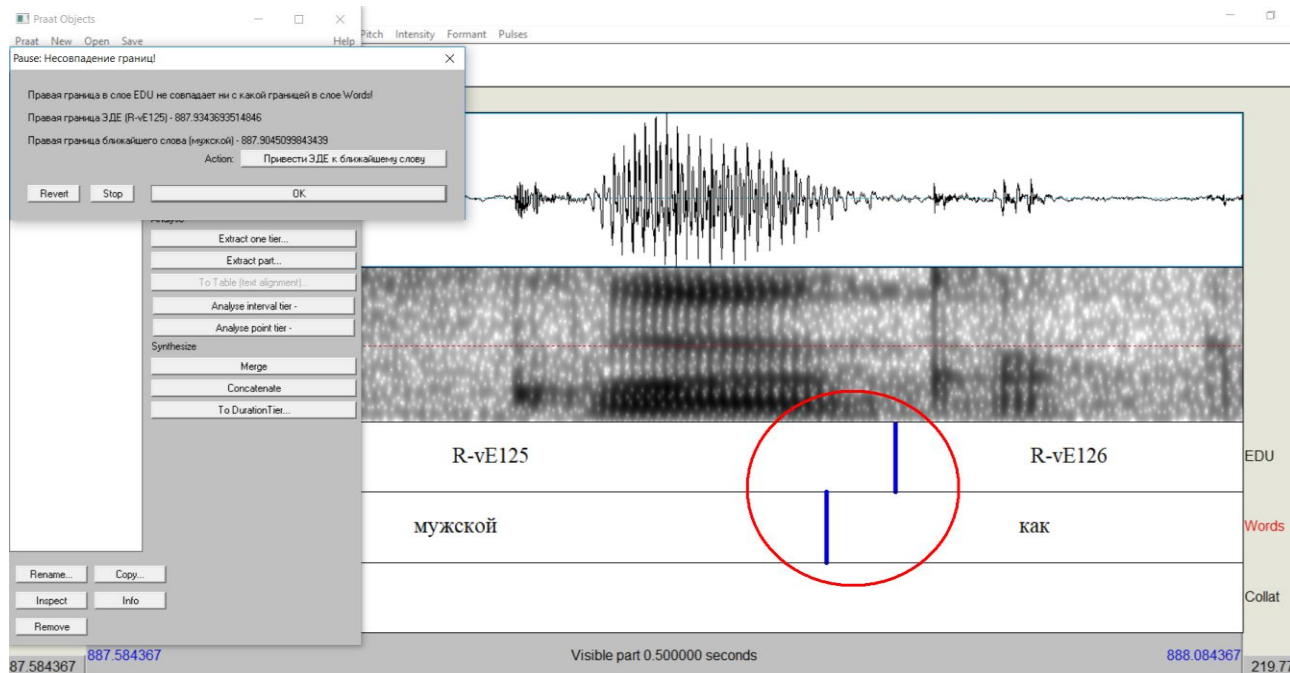


Рисунок 3. Пример ошибки, выявленной при работе скрипта по проверке границ интервалов в текстгриде

Все единицы верхнего уровня сегментации единообразно нумеруются в текстгриде и транскрипте. Для каждой единицы указывается:

- код говорящего (N для Рассказчика, С для Комментатора, R для Пересказчика);
- код вокального канала (-v);
- литеры, обозначающие тип единицы;
- трехзначный номер.

Нумерация является сквозной и непрерывной в рамках каждого типа единиц. Также непрерывную сквозную нумерацию имеют абсолютные паузы, располагающиеся между единицами верхнего уровня. В транскриптах кодовые номера некоторых единиц

дополнительно выделяются при помощи форматирования. Конкретные обозначения, используемые при нумерации, приведены в табл. 3.

Тип единицы	Общий префикс	Обозначение типа	Формат номера	Пример итогового значения	Дополнительное форматирование в транскрипте
ЭДЕ	N/C/R-v	E	000	N-vE123	
Заполненная пауза / группа заполненных пауз		F		C-vF056	Уменьшенный кегль
Смех		L		R-vL002	<i>Уменьшенный кегль, курсив</i>
Группа из прочих неречевых вокальных явлений и / или пауз со вздохом		N		R-vN096	Уменьшенный кегль, серый цвет
Абсолютная пауза между единицами верхнего уровня сегментации	pN/C/R-	pC-201			

Таблица 3. Нумерация единиц верхнего уровня сегментации и абсолютных пауз между ними

4. Разметка свойств ЭДЕ и слов

Помимо выделения единиц верхнего и нижнего уровня сегментации, вокальное аннотирование подразумевает дополнительную разметку ЭДЕ и входящих в них слов. Цель этой разметки — зафиксировать те свойства речевых единиц, которые связаны с порождением и восприятием локальной дискурсивной структуры устного текста. Разметка выполняется только в текстовом транскрипте и технически реализуется при помощи дополнительных обозначений, а в некоторых случаях — модификаций базовой записи, правила которой обсуждаются выше в разделах 2-3.

Полный список используемых обозначений приведен в приложении 1; ниже перечисляются основные обозначаемые явления с краткими пояснениями. Более подробные описания большинства рассматриваемых явлений содержатся в [Кибрик, Подлесская (ред.) 2009].

4.1. Акценты и движения тонов

Акцент — это выделенное произнесение словоформы в потоке речи. Дискурсивные функции акцентов состоят в первую очередь в выделении центрального элемента коммуникативной составляющей и в указании на способ связи этой составляющей с внешним контекстом. Вторая функция реализуется при помощи *движения тона*, ассоциированного с акцентированной словоформой (см. [Кодзасов 2009, Янко 2008]). В транскриптах наличие акцентного выделения и характер движения тона в ударном слоге обозначаются совместно — посредством иконических слэшей (/ для восходящего движения тона, \ для нисходящего и – для ровного тона) перед акцентированной словоформой. Например, слово /козой в строке R-vE093 фрагмента, приведенного в приложении 2, произносится с восходящим акцентом; слово \любовью в строке R-vE092 — с нисходящим акцентом; слово –лестницу в строке R-vE088 — с ровным акцентом. Если в рамках ударного слога реализуется сложное движение тона, используются комбинации слэшей — см. примеры в партитурном фрагменте из приложения 3: восходяще-нисходящие акценты на словах ^нет в ЭДЕ N-vE262 и ^девочки в C-vE167, восходяще-ровный акцент на слове /–мальчика в R-vE029.

В некоторых случаях для интерпретации акцента важны также и движения тона вне ударного слога — такие движения обозначаются при помощи стрелок (↑, ↓, →), располагающихся до или после слэша. Так, для интерпретации акцента на слове /↑мимо (строка R-vE097 во фрагменте из приложения 2) значимо не только восходящее движение тона на ударном слоге, но и продолжение этого движения на заударном.

В каждой ЭДЕ обычно выделяется один *главный* (или «несущий») акцент — в наибольшей степени ответственный за встраивание ЭДЕ в локальную структуру (см. также ниже, подраздел 4.2). В транскрипте несущий акцент дополнительно обозначается подчеркиванием ударной гласной акцентированного слова. В ЭДЕ могут встречаться также *вторичные* акценты; см. строку R-vE098, в которой нисходящему несущему акценту на слове \лестнице предшествуют два вторичных восходящих акцента на словах /время и /садовник. В данном случае противопоставление главного и вторичного акцентов поддерживается их коммуникативными функциями: главный акцент маркирует рему сообщения, а вторичные — тематические элементы.

В некоторых случаях в одной ЭДЕ может выделяться более одного несущего акцента — см. строку R-vE095. Чаще всего это происходит тогда, когда один из двух главных акцентов маркирует рематическую составляющую, а другой указывает на фазовое значение незавершенности (см. подраздел 4.2).

4.2. Иллокутивно-фазовое значение ЭДЕ

В каждой успешно завершённой ЭДЕ выражается та или иная комбинация значений, связанных с:

- (i) иллокутивной силой высказывания;
- (ii) фазовым различием завершенности и незавершенности;
- (iii) некоторым дополнительными значениями, модифицирующими (i) и / или (ii).

Для обозначения *иллокутивно-фазовых* значений ЭДЕ в транскриптах используются пунктуационные знаки в конце строк. Мы различаем знаки, (а) указывающие на иллокутивную силу высказывания, (б) выражающие тип дискурсивной незавершенности и (в) указывающие на тип модифицирующего значения.

А. Иллокутивные знаки

Завершение иллокуции *сообщения* отмечается при помощи точки (.) — см. строки R-vE091, R-vE092 и R-vE098 фрагмента в приложении 2. Чаще всего этому иллокутивно-фазовому значению сопутствует глубокое нисходящее движение тона в несущем акценте.

Общий и частный *вопрос* кодируется при помощи конечного вопросительного знака (?). Для общего вопроса типично восходящее движение тона в несущем акценте. При помощи обратного вопросительного знака (¿) обозначается *полуутверждение* — иллокуция, сочетающая в себе формальные признаки сообщения с функцией запроса на подтверждение высказанного предположения, см. [Коротаев 2018]. Иллокуция полуутверждения реализована в ЭДЕ R-vE025 и R-vE030 партитурного фрагмента в приложении 3.

Директивные иллокуции (т. е. иллокуции, содержащие в себе тот или иной призыв к совершению действия) обозначаются посредством перевернутого восклицательного знака (!); *обращения* — при помощи знака @.

Б. Знаки незавершенности

Стандартный способ выражения *незавершенности* кодируется при помощи запятой (,). Достаточно часто незавершенность выражается восходящим движением тона в несущем акценте — см. строки R-vE093, R-vE094, R-vE095 (второй акцент), R-vE097 и R-vE100 в приложении 2. Однако это далеко не единственная возможность; в частности,

незавершенности регулярно сопутствует несущий акцент, в котором реализовано нефинальное падение, см. R-vE087.

Незавершенность может сочетаться со значением *неполноты информации*, связанным с неспособностью говорящего «совершить однозначный и дискретный иллокутивный акт сообщения» [Кибрик, Подлеская (ред.) 2009: 168]. В этом случае в транскриптах используется знак «тройная запятая» (,,,) — см. строки R-vE089, R-vE101. Другое достаточно частое осложнение незавершенности связано с ее *восполнением последующим контекстом* — например, посредством прямого цитирования (см. подраздел 4.6). В таких случаях в транскриптах используется двоеточие (:).

В. Знаки модифицирующих значений

Одним из модифицирующих значений является упомянутое выше значение неполноты информации. Если оно не сочетается с незавершенностью, для него используется обозначение «троеточие» (...). Еще одно важное модифицирующее значение — это *восклицательность*; оно обозначается при помощи восклицательного знака (!), см. N-vE262 в партитурном фрагменте из приложения 3.

Рассмотренные выше явления могут комбинироваться в рамках одной ЭДЕ: например, иллокуция директива может сочетаться с незавершенностью (i,), вопрос — с восклицательностью (?!) и т. д. При этом действует ряд технических правил:

- (i) При сочетании знака незавершенности с иллокутивными или модифицирующими знаками знак незавершенности располагается в конце (!,; ĭ,,, и т. д.).
- (ii) При сочетании с другими знаками знак «точка» сокращается, например: !, а не !. и т. д.

Если ЭДЕ заканчивается иллокутивным пунктуационным знаком или сочетанием иллокутивного знака со знаком модификатора, то это сигнализирует о завершении *предложения*. Как и в стандартной письменной пунктуации, начало каждого нового предложения маркируется заглавной буквой первого слова — см., в частности, ЭДЕ R-vE099 в приложении 2.

4.3. Речевые сбои

В транскриптах размечаются речевые сбои, имеющие явную *точку прерывания* — сигнал нарушения плавности в развертывании дискурса. Одним из таких сигналов может служить обрыв слова, см. пункт 5 подраздела 2.1.1.

Чаще всего после точки прерывания следует попытка говорящего исправить неудачную вербализацию. Наиболее частотный класс случаев связан с *самоисправлениями*. В транскриптах различаются два типа самоисправлений:

- (i) «Слабый фальстарт», т. е. самоисправление, не приводящее к полному отказу от текущей ЭДЕ. Для обозначения слабых фальстартов в точке прерывания ставится знак || — см. множественные слабые фальстарты в строках R-vE088 и R-vE098 фрагмента из приложения 2. По определению, знак слабого фальстарта может располагаться только внутри ЭДЕ.
- (ii) «Сильный фальстарт», т. е. самоисправление, приводящее к отказу от построения текущей ЭДЕ и началу новой строки. Сильные фальстарты обозначаются знаком двойного равенства (==) в конце строки, см. строку R-vE090.

Подробнее о критериях разграничения слабых и сильных фальстартов см. [Кибрик, Подлеская (ред.) 2009: 187-208]; содержательная классификаций самоисправлений представлена в [Подлеская 2014].

В диалоге источником исправления может служить не только собственное коммуникативное намерение говорящего, но и *вмешательство других участников*: собеседники могут

перебивать друг друга, провоцировать реакции невербальными средствами и т. д. В тех случаях, когда есть достаточные основания полагать, что сбой вызван подобными внешними причинами, в транскрипте используются знаки ¶ (для слабого фальстарта; два символа 2E3E Юникода) и ≈ (для сильного фальстарта; два символа 2248 Юникода; см. ЭДЕ R-vE024, R-vE026, C-vE164 и C-vE168 в партитурном фрагменте из приложения 3).

Сильный фальстарт, т. е. ситуацию, при которой ЭДЕ не завершена, поскольку ее текущая форма признана говорящим неудовлетворительной, следует отличать от намеренного построения ЭДЕ, обладающей признаками формальной неполноты. Так, в примере (1) говорящая намеренно не завершает строку C-vE139, полагая, что имеющегося материала достаточно для выражения мысли о маловероятности того, что персонажи фильма могли бы носить джинсы, если бы жили в Крыму. Подобные примеры трактуются как случаи *апозиопезиса* и размечаются при помощи тильды (~).

(1) pears23: C-vE139

441.66	C-vE137	Я ещё /подумала,
442.12	C-vE138	что точно не \Крым,
442.95	C-vE139	потому что /-джинсов (0.09) –ну ~

4.4. Совместное построение реплик

Участники диалога регулярно прибегают к *совместному построению реплик*, при котором начальную часть реплики произносит один говорящий, а завершающую часть — другой, см. [Helsavuo 2004, Гренобль 2008]. Для обозначения совместного построения в транскриптах используется знак процента (%), который ставится в точках перехода от одного говорящего к другому внутри реплики — соответственно, в конце и в начале задействованных ЭДЕ. Например, во фрагменте, представленном в приложении 3, в ЭДЕ R-vE027 Пересказчица начинает строить реплику, которую в ЭДЕ N-vE264 завершает Рассказчица. В этом примере первая говорящая (Пересказчица) сразу же прекратила вербализацию, когда ее собеседница приступила к достраиванию реплики. Так, однако, происходит далеко не всегда: первый говорящий может решить продолжить, несмотря на «включение» собеседника. См. пример (2), в котором Рассказчица решает завершить начатую Комментатором реплику о корзине, но Комментатор не уступает ей очередь хода, из-за чего происходит наложение. В подобных случаях знак совместного построения (%) ставится только в начале вклада второго говорящего. Подробнее о принципах партитурной нотации диалога см. раздел 7.

(2) pears22: C-vE146 – N-vE204

446.30	447.84			C-vE145	Он /-взял самую (0.07) \первую
447.84				C-vE146	/корзину-у,
448.49		N-vE204	% ближе \всего к ним-м.		которая была с= (0.18) \полная.
	449.06				
	449.61				

4.5. Вставки

Под *вставкой* мы понимаем ЭДЕ или группу ЭДЕ, в которой говорящий временно отступает от основной линии изложения. При этом единицы, находящиеся по обе стороны вставки, демонстрируют тесную содержательную и / или формальную связь; в ряде случаев вставку можно мысленно изъять, не нарушив локальную связность текста. Довольно часто вставку формируют те или иные уточнения к содержанию основной линии. Начало и конец вставки маркируются скобками — см. строки R-vE091 и R-vE092 во фрагменте, представленном в приложении 2. Информация о том, как именно один из героев фильма собирает груши,

оформлена как вставка между подчиненной (стро́ки R-vE088 – R-vE089) и главной (начиная со строки R-vE093) частью полипредикативной конструкции.

Если вставка реализуется внутри ЭДЕ, то это дополнительно отмечается при помощи длинных тире (—) в конце первой и начале второй части разорванной ЭДЕ. Это явление мы называем *сплитом*, см. следующий пример:

(3) pears04: C-vE014 — C-vE016

359.13	C-vE014	Кстати на вот этом /мальчике —
360.63	C-vE015	(который на \велике ездил [?] .)
362.05	C-vE016	— [?] и на-а /дядьке /фермере были платки \↑одина [?] ковые!

Отметим, что части разорванной ЭДЕ оформляются как отдельные строки, каждая из которых имеет свой стандартный номер.

В некоторых случаях говорящие начинают фрагмент дискурсивной структуры как вставку, но в дальнейшем преобразовывают его таким образом, что он приобретает роль основной линии изложения. В транскрипте такие ситуации обозначаются при помощи сочетания знаков (* в начале строки, открывающей «одностороннюю» вставку).

4.6. Цитирование

Поскольку стимульным материалом для записей корпуса послужил фильм о грушах, персонажи которого не разговаривают между собой, различные способы *передачи чужой речи* регистрируются здесь значительно реже, чем в некоторых других устных текстах. Тем не менее принципы обозначения прямого и полупрямого цитирования входят в ядро используемой в проекте системы нотации, см. [Кибрик, Подлеская (ред.) 2009: 288-309]. Прямая и полупрямая цитация заключаются в кавычки, см. следующий пример из корпуса «Весёлые истории из жизни», в котором используется техника полупрямого цитирования (дейктические и интонационные характеристики прямой речи сочетаются с наличием союза что; подробнее см. [Подлеская 2018]):

(4) FS_28: 12-15

18.74	10.	и /стоя [?] ли,
19.29	11.	и \ны [?] ли:
19.84	12.	() что «Как не /-хо-о [?] чется,,,
21.92	13.	как /-здо-о [?] рово,,,
23.05	14.	вот бы /опозда [?] ть,
24.12	15.	вот бы здесь \оста [?] ться.»

4.7. Прочие явления

4.7.1. Темп

При оценке темпа речи мы опираемся на представления о нормальном для данного говорящего темпе — и отмечаем случаи ускоренного и замедленного произнесения. Минимальной единицей, для которой указывается скоростная характеристика, является слово. Ускоренное произнесение отмечается при помощи *курсива*, замедленное — при помощи разрядки. Пример ускоренного произнесения представлен в строке R-vE091 приложения 2 (слова *мне рассказали*). Характерно, что с повышенным темпом тут

произносится часть вставочной ЭДЕ: ускоренное произнесение является одним из типичных просодических свойств вставок.

4.7.2. Редукция

В качестве значимой *редукции* отмечаются случаи утраты части фонем, которые присутствуют при нейтральном произнесении слова (см. [Кибрик, Подлеская (ред.) 2009: 350]). Мы различаем полную редукцию, действующую на всем слове, и частичную редукцию, затрагивающую лишь часть слова. В транскриптах редукция отмечается при помощи выделения серым цветом — см. слово *проходит* в строке R-vE093 фрагмента из приложения 2.

4.7.3. Удлинения

При *удлиненной реализации* фонем в транскриптах модифицируется форма слова: буквы, соответствующие удлиненным фонемам, дублируются с использованием знака - (символ 02D7 Юникода). При этом для йотированных гласных используются обозначения вида я-а, ё-о (см. ещё-о в строке R-vE093 приложения 2), е-е, й-я, й-ю и т. д. Одной из причин удлинения является хезитация — см. примеры в строках R-vE090, R-vE095, R-vE098. О других функциях удлинения см. [Кибрик, Подлеская (ред.) 2009: 344-349].

4.7.4. Эмфаза

Под *эмфатическим* произнесением мы понимаем дополнительное выделение слова или группы слов с целью выразить эмоциональное отношение говорящего к содержанию его высказывания [Кибрик, Подлеская (ред.) 2009: 353-354]. В транскриптах эмфаза отмечается при помощи **полужирного шрифта**, см. слово **бережно** в строке R-vE091.

4.7.5. Регистр

Иногда говорящие прибегают к произнесению слов или групп слов в *смещенном тональном регистре*. В этих случаях кривая частоты основного тона временно смещается к нижнему или верхнему уровню стандартного для данного говорящего диапазона или даже выходит за его пределы. В транскриптах это явление отмечается при помощи уменьшенного кегля, а для сниженного регистра (эта модификация встречается существенно чаще, чем повышенный регистр) дополнительно используется смещение шрифта вниз. Типичный контекст сниженного тонального регистра — это вставка, см. следующий пример:

(5) pears04: C-vE230

1050.95	C-vE229	А /потом уже —
1051.58	C-vE230	(в /следующий кадр,)
1052.41	C-vE231	— он \лежит,

4.7.6. Смычки и прочие явления в начале и конце слов

Началу и концу произнесения слова могут сопутствовать дополнительные вокальные явления: придыхания, гласные призвуки, смычки. Некоторые из них функционально нагружены, другие скорее отражают индивидуальные особенности говорящих, см. [Кибрик, Подлеская (ред.) 2009: 354-356]. В транскриптах, в виде верхних индексов, размечаются:

- гласный шва-призвук в начале или конце слова (литера ə; см. ещё-о^ə в ЭДЕ R-v093 приложения 2);
- гортанная смычка в начале или конце слова (литера ʔ; например, ʔони);
- губная смычка в конце слова (литера ʷ; например, идти^ʷ);

- придыхание в конце слова (литера *h*; например, *груши^h*).

4.7.7. Нестандартное / вариативное ударение

В словах с вариативным лексическим ударением (*творóг vs. твóрог*), а также в случаях нестандартного ударения (*ранчó*) или омографии, разрешаемой посредством ударения (*зáмок vs. замóк*), ударный гласный дополнительно размечается при помощи символа 0301 Юникода.

4.7.8. Комментарии

В столбце *Comments* транскрипта отмечаются дополнительные характеристики строки, касающиеся характера произнесения (например, «тихо», «шепотом», «громко» и т. д.), содержательной интерпретации (см. мотивацию наличия двух главных акцентов в строке R-vE095 в приложении 2), ссылки на общий коммуникативный контекст, не выводимый непосредственно из вокального сигнала (например, «обращается к Комментатору»), и т. д. Заполнение столбца *Comments* является факультативным и не подчиняется строго регламентированным правилам.

5. Сопутствующие речи вокальные явления

На речь говорящего могут накладываться его же неречевые звуковые действия, сопутствующие основной вокализации. Такие явления аннотируются *только в текстгриде* — в отдельном слое *Collat*. Мы размечаем три типа сопутствующих явлений:

- смех (обозначение `{laugh}`);
- улыбка (`{smile}`);
- скрипучий голос (`{creaky}`).

В целом мы придерживаемся принципа, согласно которому границы таких сопутствующих явлений размечаются независимо от границ интервалов слоя *Words*. Например, если слышно, что со скрипучим голосом произносится только часть слова, это нужно соответствующим образом отмечать в слое *Collat*, — см. рис. 4, где на спектрограмме достаточно хорошо заметно, что скрип локализуется в конце заполненной паузы (э) и в начале союза и.

Но в некоторых случаях установить такие независимые границы невозможно. Например, не всегда легко понять, где начался смех, наложенный на слово. В таких случаях мы условно ставим границу смеха там же, где уже стоит граница слова.

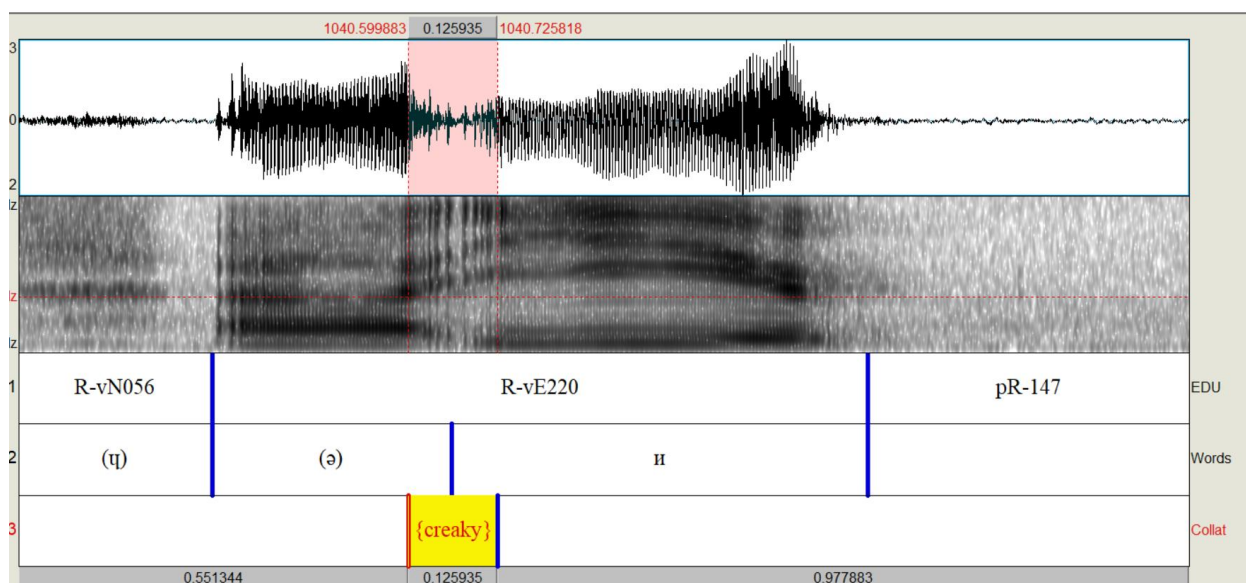


Рисунок 4. Независимые границы интервалов в слоях *Words* и *Collat*

6. Конвертация в разметку ELAN

Для совместного анализа вокальных данных и данных других каналов вокальная разметка, выполненная в Praat и Word, автоматически преобразуется в формат ELAN. Ниже описываются основные принципы конвертации данных, содержащихся в текстгридах и вокальных транскриптах, в структуру ELAN. Подробнее о схеме мультисканальной аннотации см. документ «Схема мультисканальной аннотации»

(http://multidiscourse.ru/data/ann/pears_multichannel_annotation.pdf), в котором, в частности, содержится список всех слоев и набор возможных значений в интервалах каждого слоя.

1. Единицы верхнего уровня сегментации с их временными границами и стандартной нумерацией переносятся в слой *-vLine (где * принимает одно из значений N, C и R — в зависимости от роли участника записи); единицы нижнего уровня сегментации (за исключением абсолютных пауз) переносятся в слой *-vSegm; интервалы, соответствующие вокальным явлениям, сопутствующим речи, переносятся в слой *-vCollat.
2. Абсолютные паузы фиксируются и дополнительно нумеруются в отдельном слое *-vPause, для каждой паузы в зависимом слое *-vPauseInOutEDU указывается, входит ли она в ту или иную ЭДЕ или располагается между единицами верхнего уровня сегментации.
3. В слоях *-vLineType и *-vSType указывается тип единицы соответственно верхнего и нижнего уровня сегментации: ЭДЕ, слово, заполненная пауза и др. В слоях *-vSForm и *-vCollatForm указывается конкретная форма соответственно единицы нижнего уровня сегментации и сопутствующего речи вокального явления. В слое *-vLineVerbatim содержится пословная запись единиц нижнего уровня сегментации, входящих в состав данной единицы верхнего уровня.
4. Свойства слов и ЭДЕ, рассмотренные в разделе 4, преобразуются из транскрипционных обозначений в значения интервалов в зависимых слоях. Для каждого свойства используется отдельный слой — в том числе:
 - слой *-vIllocPhase и *-vCombIllocPhase для указания на иллокутивно-фазовые значения ЭДЕ (см. подраздел 4.2);
 - слой *-vCoConstr для указания на роль ЭДЕ в совместном построении реплик (подраздел 4.4);
 - слой *-vParenth и *-vInSplit для указания на роль ЭДЕ в конструкциях со вставкой и со сплитом (4.5);
 - слой *-vCitation для указания на роль ЭДЕ в конструкциях с (полу)прямым цитированием (4.6);
 - слой *-vAccents и *-vMainAccent для обозначения движений тона в акцентированных словах и статуса акцентов как главных / вторичных (4.1);
 - слой *-vInterrupt для указания на тип речевого сбоя, связанного с данной точкой прерывания (4.3);
 - слой *-vTempo и *-vRegister для указания на отклонения от стандартного темпа (4.7.1) и регистра (4.7.5) речи;
 - слой *-vReduction и *-vEmph для указания на наличие редукции (4.7.2) и эмфатическое произнесение (4.7.4);
 - слой *-vLength и *-vStress для указания на наличие удлиненной реализации фонем (4.7.3) и нестандартного / неочевидного ударения (4.7.7);
 - слой *-vStops для указания на тип смычки / прочего явления на границе слова (4.7.6).

5. Для удобства дальнейших подсчетов дополнительно создаются слои, в интервалах которых указываются значения, вычисляемые по тем или иным специальным правилам. В число дополнительных слоев, соотносящихся с ЭДЕ, входят:

- *-vWordsCount — в этом слое указывается количество слов в ЭДЕ;
- *-vPausesCount — количество абсолютных пауз внутри ЭДЕ;
- *-vFilledCount — количество заполненных пауз внутри ЭДЕ;
- *-vStartFilled — в этом слое указывается, начинается ли ЭДЕ с заполненной паузы (см. подраздел 3.2);
- *-vAccentsCount — в этом слое указывается количество акцентированных слов в ЭДЕ;
- *-vMainAccentsCount — количество слов, произносимых с главным акцентом;
- *-vMainAccents — в этом слое перечисляются движения тона, реализованные в главных акцентах ЭДЕ;
- *-vAccentsAfterMainCount — количество вторичных акцентов после (последнего) главного акцента в ЭДЕ;
- *-vInterruptCount — количество точек прерывания в ЭДЕ.

Дополнительно вычисляемые свойства единиц нижнего уровня сегментации отражаются в следующих слоях:

- *-vInOutEDU — указывается, входит ли данная единица в какую-либо ЭДЕ (см. подраздел 3.2);
- *-vNearPause — указывается, находится ли данная единица в непосредственном соседстве с абсолютной паузой;
- *-vWordNum — порядковый номер слова от начала ЭДЕ;
- *-vWordNumReversed — порядковый номер слова от конца ЭДЕ.

7. Партитурная нотация

Описанные выше принципы вокальной аннотации основаны на системе дискурсивной транскрипции, которая была разработана на материале преимущественно монологического устного дискурса. Как указано в подразделе 1.1, при аннотировании записей корпуса «Рассказы и разговоры о грушах» мы продолжаем транскрибировать вокальное поведение каждого участника по отдельности — даже на тех участках записей, где реализуется полноценное речевое взаимодействие между несколькими участниками. Этот подход оправдывается, в частности, тем, что такие базовые понятия, как ЭДЕ, акценты, иллюкутивно-фазовые значения, речевые сбои и проч., релевантны не только для монологической, но и для диалогической ситуации. В то же время можно выделить характеристики, значимо отличающие подлинный диалог от (условного) монолога. По крайней мере две из них напрямую касаются процесса вокального аннотирования и способов адекватного графического представления выполненной аннотации.

Во-первых, в диалоге существенно иной статус, чем в монологе, имеют *пограничные абсолютные паузы*. Если при транскрибировании монологического дискурса пограничные паузы уместно включать в состав следующих за ними ЭДЕ (см. [Chafe 1994, Кибрик, Подлеская 2009 (ред.): 64-66]), то при диалоге отрезки молчания невозможно приписать не только отдельной ЭДЕ (см. решение о вынесении абсолютных пограничных пауз в отдельные строки транскрипта, описанное в подразделе 3.2), но в общем случае даже и конкретному говорящему. В связи с последним замечанием также важно отличать ситуации, в которых молчит только один говорящий (именно такие отрезки размечаются в

индивидуальных транскриптах и текстгридах), от ситуаций *совместного молчания* всех активных участников. В частности, отрезки общего молчания могут указывать на особенности чередования реплик, тип организации предпочтений в смежных парах и т. д. (см. [Sacks et al. 1974] и другие работы по Анализу бытового диалога).

Во-вторых, для диалога характерны множественные случаи *наложения*, т. е. одновременного говорения более одного участника. По нашим данным, на этапе разговора наложения составляют около 15% времени; более того, без каких-либо наложений произносятся менее половины всех ЭДЕ на этом этапе. В целом, вопрос координации вокальных действий участников на единой временной шкале является одним из центральных при анализе диалогических данных.

Для наглядного отображения описанных выше свойств и явлений в проекте наряду с индивидуальными вокальными транскриптами используется единая *партитурная нотация*. Партитурная нотация автоматически формируется в среде Excel из индивидуальных вокальных транскриптов и текстгридов (для выполнения этой процедуры создан отдельный скрипт). Скриншот с фрагментом партитурной нотации 22-й записи представлен в приложении 3. При создании партитурных транскриптов реализуются следующие принципы.

1. Сохраняется сегментация речевого потока на строки, соответствующие ЭДЕ и другим единицам верхнего уровня сегментации, и нумерация этих строк (см. раздел 3). При этом вокальный вклад каждого из участников фиксируется в отдельных столбцах, обозначенных разными цветами: зеленым для Рассказчика, синим для Комментатора и темно-красным для Пересказчика.
2. Для каждой единицы верхнего уровня сегментации в столбцах TimeS и TimeE указывается соответственно время начала и конца вокализации с точностью до сотых долей секунды. Строки упорядочены по времени начала вокализации, т. е. по значениям в столбце TimeS.
3. В отдельных столбцах под заголовком Pauses, слева от столбцов TimeS и TimeE, указываются отрезки совместного молчания всех трех участников. Непустые строки в этом столбце имеют сквозную нумерацию вида vr000, для каждой общей паузы указывается ее продолжительность с точностью до сотых долей секунды. Расположение строк с общими паузами также следует основному правилу сортировки.
4. Каждая графическая строка совместного транскрипта пронумерована для удобства дальнейших ссылок (столбец Line #).
5. Если единица верхнего уровня сегментации одного участника заканчивается позже, чем начинается ближайшая единица другого участника (т. е. имеется наложение), то время завершения произнесения первой единицы указывается в дополнительной графической строке. Таким образом, значения в столбце TimeE также упорядочиваются по возрастанию; при этом ячейки в столбце с транскрипцией визуально объединяются — см., например, графические строки 0810-0813 (ЭДЕ C-vE163 Комментатора) и 0838-0842 (ЭДЕ R-vE030 Пересказчицы) во фрагменте, представленном в приложении 3.
6. Непрерывные периоды вокализации одного участника дополнительно закрашиваются цветом (по принципу, описанному в п. 1). Этими же цветами закрашиваются ячейки с указанием времени начала / конца вокализации в столбцах TimeS и TimeE. При помощи цветовой заливки наглядно выявляются отрезки наложения. Так, например, в графических строках 0810-0811 зафиксировано наложение речи Комментатора на речь Рассказчицы, в строках 0813-0815 и 0817-0818 — наложение речи Пересказчицы на речь Комментатора, а в строках 0819-0821 одновременно говорят все три участника. В строках 0830-0834 можно заметить, что наложением сопровождается совместное построение реплики: Рассказчица начинает достраивать конструкцию, начатую Пересказчицей, прежде чем Пересказчица прекращает вокализацию.

7. Также закрашиваются пустые области в столбцах TimeS и TimeE, соответствующие срединным участкам говорения одного участника. См., например, закраску в строках 0809 (Рассказчица продолжает произносить ЭДЕ N-vE261, после того как Пересказчица завершила произнесение ЭДЕ R-vE024 и в то время пока Комментатор еще не приступил к произнесению ЭДЕ C-vE163) и 0816 (Пересказчица продолжает говорить, пока Комментатор делает паузу между двумя ЭДЕ).

Литература

Гренобль Л. А. (2008) Синтаксис и совместное построение реплики в русском диалоге // Вопросы языкознания, № 1. – 25-36.

Кибрик А. А. (2018) Русский мультимедийный дискурс. Часть I. Постановка проблемы // Психологический журнал 39(1). – 70-80.

Кибрик А. А., Подлеская В. И. (ред.) (2009) Рассказы о сновидениях: Корпусное исследование устного русского дискурса. М.: ЯСК.

Кодзасов С. В. (2009) Исследования в области русской просодии. М.: ЯСК.

Коротаев Н. А. (2015) Коммуникативно-просодический подход к выявлению элементарных дискурсивных единиц в устном монологическом тексте // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Москва, 27–30 мая 2015 г.). Вып. 14 (21). М.: Изд-во РГГУ. 294–307.

Коротаев Н. А. (2018) Вопрос и полутверждение в структуре мультимедийного дискурса // VIII Международная конференция по когнитивной науке. Тезисы докладов.

Подлеская В. И. (2014) «То есть, не убили, а зарезали саблей»: самоисправления говорящего в устных рассказах // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4-8 июня 2014 г.). Вып. 13 (20). – 526-540.

Подлеская В. И. (2018) Чужая речь в свете корпусных данных // Вопросы языкознания, № 4. – 47-73.

Янко Т. Е. (2008) Интонационные стратегии русской речи в сопоставительном аспекте. М.: ЯСК.

Chafe, W. (1994) Discourse, consciousness, and time. Chicago: University of Chicago Press.

Clark, H. H., Fox Tree, J. E. (2002) Using *uh* and *um* in spontaneous speaking // Cognition, 84. - 73-111.

Kibrik, A., Korotaev, N., Podlesskaya, V. (in print) Russian spoken discourse: Local structure and prosody.

Helsavuo, M.-L. (2004) Shared syntax: the grammar of co-construction // Journal of pragmatics. 36.

Sacks, H., Schegloff, E., Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation // Language, 50. – 696-735.

Приложение 1: список обозначений, используемых в текстовых вокальных транскриптах

Ниже в таблице приведен список обозначений, используемых в текстовых вокальных транскриптах. Для каждого обозначения указывается кодируемое при его помощи явление; подраздел основного текста, в котором это явление обсуждается; а также соответствие в мультисканальной разметке в ELAN (см. раздел 6, а также *документ с описанием слов и значений в ELAN*).

Обозначение в транскрипте	Явление	Обсуждается в подразделе	Соответствие в разметке ELAN
Деление на строки	Сегментация речевого потока на единицы верхнего уровня: ЭДЕ и др.	3.1, 3.2, 3.3	Нумерованные интервалы в слоях *-vLine
	Абсолютные паузы между единицами верхнего уровня сегментации	3.2, 3.3	Нумерованные интервалы в слое *-vPause со значением Out в зависимом слое *-vPauseInOutEDU
(0.23)	Абсолютные паузы, с указанием продолжительности	2.1.5	Нумерованные интервалы в слое *-vPause
{ц 0.73}	Паузы с шумным вдохом, с указанием продолжительности		Нумерованные интервалы в слое *-vSegm со значением HPause в зависимом слое *-vSType
{ə 0.20}, {e 0.33}, {ш 0.48}, {ʔ 0.34}, {əш 0.62} и проч.	Заполненные паузы различной природы, с указанием продолжительности	2.1.3	Нумерованные интервалы в слое *-vSegm со значением Filled в зависимом слое *-vSType
{laugh 1.02}	Смех, не накладывающийся на речь, с указанием продолжительности	2.1.4	Нумерованные интервалы в слое *-vSegm со значением Laugh в зависимом слое *-vSType
{cl 0.12}, {st 0.23}, {gr 0.18} и проч.	Прочие неречевые вокальные действия говорящего		Нумерованные интервалы в слое *-vSegm со значением Other в зависимом слое *-vSType
/ \ - / \ и т. д. (перед словом)	Наличие акцента на слове и характер движения тона на ударном слоге акцентированного слова	4.1	Непустые интервалы в слое *-vAccents
↑ ↓ →	Коммуникативно значимое движение тона вне		

Обозначение в транскрипте	Явление	Обсуждается в подразделе	Соответствие в разметке ELAN
	ударного слога акцентированного слова		
Подчеркивание ударной гласной акцентированного слова	Несущий (главный) акцент ЭДЕ	4.1	Значение Main в слове *-vMainAccent
Заглавная буква в начале строки	Начало нового предложения	4.2	
.	Завершение иллокуции сообщения		Значение Period в слове *-vIllocPhase
?	Вопрос		Значение Quest в слове *-vIllocPhase
¿	Полуутверждение		Значение Semi-St в слове *-vIllocPhase
i	Директив		Значение Dir в слове *-vIllocPhase
@	Обращение		Значение Addr в слове *-vIllocPhase
,	Стандартная иллокутивная незавершенность		Значение Comma в слове *-vIllocPhase
:	Незавершенность, восполняемая последующим контекстом		Значение Colon в слове *-vIllocPhase
...	Значение неполноты информации, совмещенное с завершением иллокуции		Значение Dots-f в слове *-vIllocPhase
'''	Значение неполноты информации, совмещенное с незавершенностью		Значение Dots-nf в слове *-vIllocPhase
!	Восклицательность	Значение Exclam в слове *-vIllocPhase	
— (в конце строки)	Начало сплита (разрыва ЭДЕ)	4.5	Значение Split в слове *-vIllocPhase; значение Enter в слове *-vParenth + значение InSplit в слове *-vInSplit

Обозначение в транскрипте	Явление	Обсуждается в подразделе	Соответствие в разметке ELAN
— (в начале строки)	Конец сплита (разрыва ЭДЕ)	4.5	Значение Return в слое *-vParenth + значение InSplit в слое *-vInSplit
()	Начало и конец вставки		Значения Start, Final, Full в слое *-vParenth
(*	Начало «односторонней» вставки		Значение *Start в слое *-vParenth
« »	Начало и конец (полу)прямой цитации	4.6	Значения Begin, End, Whole в слое *-vCitation
%	Вхождение ЭДЕ в совместно построенную реплику	4.4	Непустые интервалы в слое *-vCoConstr
	«Слабый фальстарт» (точка прерывания при самоисправлении внутри ЭДЕ)	4.3	Значение Mild в слое *-vInterrupt
= =	«Сильный фальстарт» (точка прерывания при самоисправлении на границе ЭДЕ)		Значение Fst в слое *-vIllocPhase; значение Severe в слое *-vInterrupt
⋈	«Слабый фальстарт», вызванный вмешательством других участников		Значение Mild-other в слое *-vInterrupt
≈ ≈	«Сильный фальстарт», вызванный вмешательством других участников		Значение Interrupt в слое *-vIllocPhase; значение Severe-other в слое *-vInterrupt
~	Обрыв ЭДЕ, которая расценивается говорящим как уместная и не подлежащая исправлению (апозиепезис)		Значение Tilde в слое *-vIllocPhase
=	Обрыв слова	2.1.1 (п. 5)	Значение Truncated в слое *-vTruncated
написание через нижнее_подчеркивание	Сверхслитное произнесение, соответствующее раздельному написанию в стандартной орфографии	2.1.1 (п. 7)	

Обозначение в транскрипте	Явление	Обсуждается в подразделе	Соответствие в разметке ELAN
’слово’	Гортанная смычка в начале / конце слова	4.7.6	Значения Gl-st и Gl-en в слое *-vStops
әсловоә	Гласный призвук в начале / конце слова		Значения Schw-st и Schw-en в слое *-vStops
слово ^ʷ	Губная смычка в конце слова		Значение Lab-en в слое *-vStops
слово ^h	Придыхание в конце слова		Значение Asp-en в слое *-vStops
а-а с-с я-а й-я	Удлиненная реализация фонем	4.7.3	Значение Len в слое *-vLength
ранчо́	Нестандартное / неочевидное лексическое ударение	4.7.7	Значение Stress в слое *-vStress
Выделение <i>курсивом</i>	Ускоренный темп	4.7.1	Значение Fast в слое *-vTempo
Выделение разрядкой	Замедленный темп		Значение Slow в слое *-vTempo
Выделение серым цветом	Редуцированное произнесение	4.7.2	Непустые интервалы в слое *-vReduction
Выделение полужирным	Эмфатическое произнесение	4.7.4	Значение Emph в слое *-vEmph
Уменьшенный кегль	Повышенный тональный регистр	4.7.5	Значение Hi в слое *-vRegister
Уменьшенный кегль вниз со сдвигом	Сниженный тональный регистр		Значение Lo в слое *-vRegister
#ты-дыш# #свист#	Звукоподражания / невербальные действия, равнозначные словам	2.1.2	Значение Onom в слое *-vOnom
<ВОТ>	Предположительная транскрипция неразборчивого фрагмента	2.1.1 (п. 8)	
>фервела<	Разборчивый, но неидентифицируемый фрагмент	2.1.1 (п. 9)	
<НРЗБ>	Полностью неразборчивый фрагмент	2.1.1 (п. 10)	

Приложение 2: фрагмент индивидуального вокального текстового транскрипта (запись 22, Пересказчица)

Time	EDU	Transcription	Comments
817.15	R-vE087	\Вот,	
817.40	R-vE088	в то время пока он значит н= н-н= залез на -лестницу,	
820.01	R-vE089	и собирает там эти /↓груши-и,,,	
821.51	pR-098	(0.52)	
822.03	R-vE090	с-с= ==	
822.29	pR-099	(0.21)	
822.50	R-vE091	(Как (0.45) мне рассказали \бережно.	
824.44	R-vL024	{ laugh 0.71 }	
825.15	R-vE092	И с \любовью.)	
825.94	R-vL025	{ laugh 0.77 }	
826.72	R-vF005	(в 0.41)	
827.13	pR-100	(1.64)	
828.77	R-vE093	мимо проходит (э 0.39) какой-то ещё-о ^э (0.29) другой мужчина с /козой,	
832.49	pR-101	(0.17)	
832.66	R-vN020	(ц 0.32)	
832.98	R-vE094	и коза-а (1.63) хочет видимо-о (0.19) (? 0.08) (0.20) полакомиться /грушами,	

Time	EDU	Transcription	Comments
837.89	R-vE095	но-о {sf 0.48} (ш 0.57) (0.62) её хозяин ей не \даёт этого /сделать,	Первый несущий акцент — на верификативной реме, второй — маркирует незавершенность.
841.67	pR-102	(0.26)	
841.93	R-vN021	(ц 0.11)	
842.04	R-vE096	(-вог,)	
842.20	R-vE097	они проходят /↑мимо,	
843.16	pR-103	(0.25)	
843.40	R-vN022	(ц 0.28)	
843.68	R-vF006	(э 0.34)	
844.02	pR-104	(0.72)	
844.74	R-vE098	всё это /время значит /садовник на де= н-на= на-а = на= на \лестнице.	
847.96	R-vN023	(ц 0.48)	
848.45	R-vE099	(э 0.18) –Вог,	
848.82	R-vE100	потом приезжает /мальчик,	
850.15	R-vE101	по-моему на \красном /↓велосипеде,,,	

Приложение 3: фрагмент партитурного вокального транскрипта записи 22

Line #	Pauses	TimeS	TimeE	Narrator	Commentator	Reteller
0806	вр084 (0.04)	538.63	538.67			
0807		538.67		N-vE261 У них у \всех шляпы.		
0808		538.89	539.23			R-vE024 И-и' ≈
0809						
0810		539.46			C-vE163 У них у /всех,	
0811			539.68			
0812						
0813		539.86				R-vE025 И у /мальчиков троих \тоже шляпы-ы,
0814			540.12			
0815		540.12	540.30		C-vE164 к²= ≈	
0816						
0817		540.90	541.23		C-vE165 \Нет.	
0818		541.23			C-vE166 Вот у троих мальчиков /нет,	
0819		541.26	541.68	N-vE262 /Не-ет!		
0820		541.68		N-vE263 У детей \нету.		
0821			541.74			
0822			542.27			
0823		542.27			C-vE167 и у /девочки нету.	
0824		542.48				R-vE026 То есть ≈
0825			542.58			
0826			542.69			
0827			543.19			
0828	вр085 (0.42)	543.19	543.61			
0829		543.61			C-vE168 /Шляпы ≈	
0830		543.62				R-vE027 /Шляпы только-о %
0831			544.09			
0832						
0833		544.32		N-vE264 % у \взрослых.		
0834			544.66			
0834			545.10			
0835	вр086 (0.01)	545.10	545.11			
0836		545.11	545.65			R-vE028 У /взрослых,
0837		545.65	546.31			R-vE029 и у /-мальчика,
0838		546.31				R-vE030 который на \велосипеде соответственно,
0839		547.13	547.43		C-vE169 (ш 0.11) \Да-а.	
0840						
0841		547.80			C-vE170 -Да.	
0842			547.83			
0843			548.03			
0844	вр087 (0.08)	548.03	548.10			