

ЕСТЬ ЛИ ЭЛЕМЕНТАРНЫЕ ДИСКУРСИВНЫЕ ЕДИНИЦЫ В РАЗМЕТКАХ НЕПОДГОТОВЛЕННЫХ АННОТАТОРОВ?

Н.А. Коротаев

Российский государственный гуманитарный университет, Институт языкознания РАН
n_korotaev@hotmail.com

I. Постановка задачи

Одна из базовых характеристик устного дискурса заключается в его квантованности. Устная речь порождается не единым потоком, а отдельными шагами, которые, с одной стороны, обусловлены фундаментальными свойствами человеческого поведения (Chafe 1994), с другой стороны, обладают набором формальных — в первую очередь, просодических — свойств (Degand, Simon 2009; Barth-Weingarten 2016; и др.). В работах научной группы под руководством А.А. Кибрика для обозначения таких квантов закрепился термин *элементарная дискурсивная единица* (ЭДЕ). Определение границ ЭДЕ в реальных текстах — это экспертная процедура, при реализации которой аннотаторы опираются на набор интонационных и (в меньшей степени) содержательных критериев: наличие единого интонационного контура и акцентных пиков, типические скоростные и громкостные паттерны, способность выступать в качестве минимального узла локальной дискурсивной структуры (Кибрик, Подлесская (ред.) 2009; Kibrik et al. 2020). В целом, выявлению и применению подобного рода критериев посвящена обширная литература; см., в частности, обзор в Izre'el et al. 2020: 7–13. При этом наряду с подходом, который можно условно назвать «экспертно-ориентированным», существует и альтернативная традиция — основанная на оценке разметок, выполненных «наивными» носителями языка, см. Himmelmann et al. 2018.

В настоящем исследовании представлена попытка совместить эти две перспективы. На протяжении нескольких лет я предлагал неподготовленным разметчикам — студентам Института лингвистики РГГУ — выявить сегментацию на элементарные кванты в одном и том же устном тексте. Их разметки были сопоставлены как между собой, так и с «эталонным» транскриптом, выполненным подготовленными аннотаторами. Исследовательские вопросы

тут таковы: (а) способны ли неподготовленные разметчики выявлять ЭДЕ достаточно надежным (согласованным) образом; (б) на какие свойства устного текста они при этом опираются; (в) чем их разметки и используемые ими критерии отличаются от эталона, основанного на эксплицитной инструкции.

II. Материал и протокол исследования

Материалом исследования послужили разметки одного устного рассказа, включенного в корпус «Истории о подарках и катании на лыжах» (<http://spokencorpora.ru/>). Продолжительность аудиофайла — 36 секунд, прослушать его можно на странице http://spokencorpora.ru/showtrans.py?file=03pands/rus/Pic-RUS_03-m_Ski-T&type=1. Обновленная версия транскрипта, использовавшаяся в качестве эталонной разметки, содержится в Приложении 2. В качестве «наивных» разметчиков выступили студенты Института лингвистики РГГУ. В 2020–2023 годах они выполняли неопениваемое задание в рамках курсов «Общая теория текста и дискурса» и «Анализ устной речи». Им предлагалось 2-3 раза целиком прослушать аудиофайл и расставить границы ЭДЕ в текстовой расшифровке, содержащей все слова и заполненные (но не абсолютные!) паузы рассказа без каких-либо дополнительных транскрипционных обозначений; см. Приложение 1.

При этом ни понятие ЭДЕ, ни — тем более — содержательные и формальные критерии обнаружения границ ЭДЕ в устных текстах до выполнения задания со студентами не обсуждались. В начале занятия разбиралось лишь самое общее представление о наличии некоторых минимальных шагов, при помощи которых говорящие реализуют свой сложный коммуникативный замысел — подобно тому, как мы естественным образом членим на элементарные участки движения то или иное сложное перемещение в пространстве. Задание формулировалось устно следующим образом: «Укажите в расшифровке границы между фрагментами текста, соответствующими минимальным шагам в порождении речи»¹.

¹ Похожее задание выполняется и в рамках курса «Общая теория дискурса», который читает на ОТиПЛе МГУ А.А. Кибрик. На одном из занятий студенты также слушают аудиофайл и расставляют границы ЭДЕ, после чего их разметки сравниваются между собой. Однако имеются и важные отличия. Во-первых, студенты МГУ слушают рассказ на незнакомом языке — таким образом, они лишены доступа к семантико-синтаксической структуре текста и могут ориентироваться только на просодические факторы. Во-вторых, до выполнения задания с ними обсуждаются возможные формальные критерии сегментации.

III. Результаты

По состоянию на март 2023 года получено и обработано 70 разметок. Ниже рассматриваются результаты, позволяющие ответить на три поставленных в исследовании вопроса.

III.1. Сопоставление «наивных» разметок между собой

Первый вопрос: насколько согласованно действуют «наивные» разметчики?

Общепринятой статистической мерой для оценки степени согласованности группы аннотаторов является каппа Флейса — показатель, призванный отразить, насколько наблюдаемые сходства в разметках превышают те, которое можно было бы ожидать при случайном выборе (Fleiss 1971). Каппа Флейса варьирует от 0 до 1; для наших данных ее значение составляет 0.64. Много это или мало? Представляется, что тут можно использовать два ориентира. Во-первых, согласно предложенной в работе Landis, Koch 1977 шкале, интервал значений каппы Флейса от 0.6 до 0.8 соответствует «существенной» (*substantial*) степени согласия. (Значения выше 0.8 в этой шкале интерпретируются как признак полного или почти полного согласия.) Во-вторых, этот результат можно сопоставить с показателем, который был получен для аналогичного задания, связанного с восстановлением границ абзацев во фрагменте рассказа Б. Житкова «Над водой». Задание, суть которого описана в Кибрик 2003: 31–34, выполняли 48 из 70 студентов; каппа Флейса для этих разметок оказалась значительно ниже — лишь 0.36. Полученный результат можно трактовать таким образом, что определение границ ЭДЕ является для неподготовленных разметчиков существенно более понятной задачей, чем восстановление абзацных границ — и это притом, что абзацы, как элементы письменного текста, доступны для непосредственного наблюдения.

Таким образом, можно предположить, что носители языка обладают достаточно хорошим представлением о том, как устроена естественная сегментация устной речи. Однако полного согласия между аннотаторами нет; а значит, интересно разобраться, чем именно разметки отличаются друг от друга. Анализ показывает, что подавляющее большинство различий обусловлено степенью дробности при выполнении сегментации: более 80% несовпадений между всеми парами разметчиков заключается в строгом вложении двух или более ЭДЕ, выделенных одним разметчиком, в одну ЭДЕ, выделенную другим разметчиком; и лишь менее чем в 20% случаев наблюдается наложение границ. В целом, разметчики выбирают некоторую «стратегию дробности» и последовательно ей придерживаются. Естественным следствием этого положения дел становится то, что одни границы отмечаются явным большинством

вом разметчиков (так, восемь границ отмечены более чем в 85% аннотаций; подробнее см. Приложение 2), а другие — лишь теми из них, кто придерживается более дробной стратегии сегментации.

III.2. Сопоставление «наивных» разметок с эталонной

Второй вопрос: как разметки, выполненные студентами, соотносятся с эталонной аннотацией?

Для ответа на этот вопрос можно использовать ту же статистическую меру — каппу Флейса, но теперь уже для сопоставления каждой отдельной «наивной» аннотации с эталоном. Диаграмма размаха полученных таким образом значений приведена на рис. 1. Видно, что среднее значение составило 0.70, медианное — 0.75, а срединные 50% значений расположились в интервале от 0.62 до 0.81.

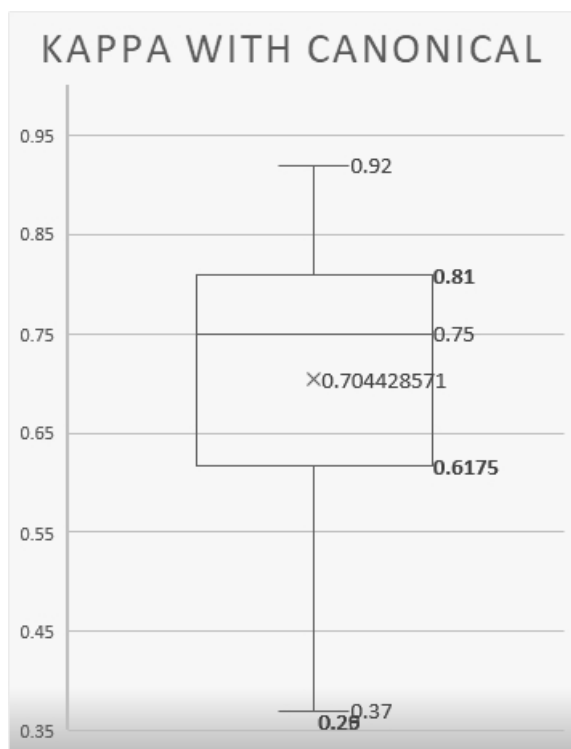


Рисунок 1. Диаграмма размаха значений кappy Флейса относительно эталонной разметки рассказа для 70 «наивных» разметок

Эти данные можно интерпретировать следующим образом. С одной стороны, значимое большинство «наивных» аннотаторов, не прошедших специального обучения, при выявлении сегментации устного текста демонстрируют по меньшей мере существенную степень согласованности с профессиональными разметчиками, следующими эксплицитной инструкции. С другой стороны, четверть студенческих разметок довольно далеко отстоит от эталонной. Можно предположить, что эти аннотаторы недостаточно хорошо поняли поставленную перед ними задачу; при обсуждении критериев, на которые опирались студенты при выполнении задания (см. раздел III.3), такие разметки не будут приниматься во внимание.

Сводная информация о том, каким образом каждая граница ЭДЕ в эталонной разметке соотносится с границами, отмеченными «наивными» аннотаторами, представлена в Приложении 2.

III.3. Критерии сегментации

Итак, 75% «наивных» аннотаторов выполнили разметку, достаточно близкую к эталонной. Естественным образом, возникает третий вопрос: на какие критерии они при этом опирались и были ли эти критерии теми же, которые перечислены в эксплицитной инструкции, использовавшейся при создании эталонной разметки?

Для ответа на этот вопрос было введено рабочее понятие согласованности разметок по каждому слову; для его оценки использовался числовой коэффициент, варьирующий от 0.5 (ситуация, в которой ровно половина разметчиков поставила границу ЭДЕ после данного слова, а другая половина — нет) до 1 (все разметчики либо поставили, либо не поставили границу после данного слова). Диаграмма размаха значений этого показателя для входящих в рассказ 57 слов представлена на рис. 2.

Среднее значение составило 0.91, медианное — 0.98, нижние выбросы расположены ниже 0.83. Это значение было принято за пороговое: если коэффициент согласованности для слова превосходил 0.83, то считалось, что по этому слову разметчики имели согласованную позицию, в противном случае — несогласованную. При сопоставлении этого параметра с формальными свойствами слов обнаружилось, что в целом решения «наивных» разметчиков достаточно точно предсказываются двумя факторами: наличием/отсутствием после слова (i) ощутимой паузы и (ii) синтаксической границы. Модель принятия решений при этом выглядит так²:

² Немногочисленные случаи, противоречащие общей модели, выделены в Приложении 2.

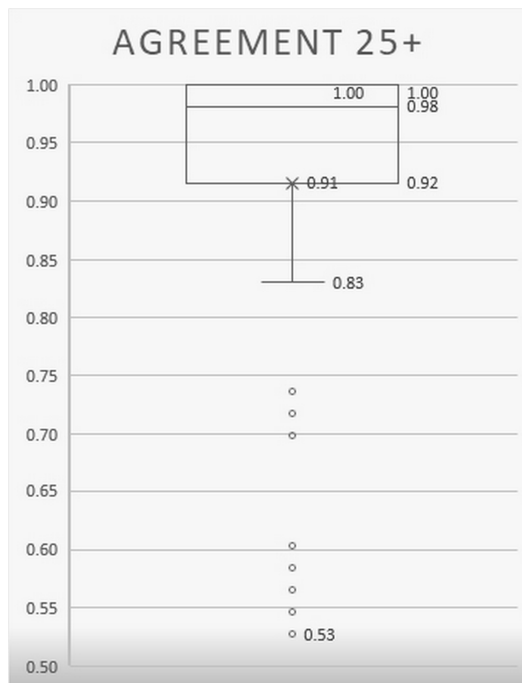


Рисунок 2. Диаграмма размаха коэффициента согласованности при (не)проставке границ после слова в 75% «наивных» разметок, наиболее согласованных с эталонной разметкой

А. Если после слова имеется (i) пауза не менее 190 мс и (ii) граница клаузы или постпозитивного добавления к клаузе, то разметчики согласованно отмечают границу ЭДЕ.

Б. Если после слова (i) нет паузы или пауза длится менее 190 мс и (ii) отсутствует граница клаузы или именной группы, то разметчики согласованно не отмечают границу ЭДЕ.

В. В прочих случаях разметчики действуют несогласованно.

Примечательно, что, хотя разметки, выполненные 75% «наивных» аннотаторов, в существенной степени согласованы с эталонной разметкой, при реализации эталонной аннотации в качестве основных использовались другие критерии сегментации. Ни паузация, ни особенности синтаксической структуры не входят в число базовых критериев выделения ЭДЕ, упомянутых в инструкции. Так, согласно Kibrik et al. 2020, основными критериями являются наличие у ЭДЕ общего интонационного контура и акцентных пиков, а также следование стандартным паттернам в распределении громкости и скорости; см. также раздел 1 выше. На мой взгляд, полученные результаты можно ин-

терпретировать как еще одно свидетельство высокой степени координации различных аспектов речепорождения, которая проявляется, в том числе, и при сегментации речевого потока, см. Кибрик, Подлеская (ред.) 2009: 56–57.

IV. Заключение

Объектом настоящего исследования стали разметки устного текста, выполненные неподготовленными аннотаторами — носителями русского языка. 70 студентов Института лингвистики РГГУ, не получив предварительного инструктирования, обозначили границы минимальных шагов в порождении устной речи, ориентируясь на звуковой файл и словесную расшифровку одного из рассказов корпуса «Истории о подарках и катании на лыжах». Анализ выполненных таким образом разметок показал, что «наивные» аннотаторы действуют с достаточно высокой степенью согласованности — как между собой, так и относительно эталонной разметки, осуществленной подготовленными экспертами. При этом формальные критерии, на которые, по всей видимости, опираются «наивные» разметчики, не совпадают с теми, которые приведены в инструкции для профессионального аннотирования.

Как представляется, полученные результаты примечательны с двух точек зрения. С теоретической точки зрения — отвечая на вопрос, вынесенный в заголовок данной статьи — можно предположить, что неподготовленные аннотаторы в целом обладают довольно устойчивым интуитивным представлением о характере естественной сегментации устной речи. Это, в свою очередь, позволяет надеяться на то, что теоретические рассуждения о природе ЭДЕ базируются на прочном психолингвистическом фундаменте. С практической точки зрения, можно сделать вывод о том, что сегментацию речевого потока при аннотировании устной речи все же не получается доверить полностью неподготовленными разметчикам; однако есть основания надеяться, что при минимальном дополнительном обучении результаты их аннотации приблизятся к эталонным.

Литература

- Кибрик А.А. 2003. *Анализ дискурса в когнитивной перспективе*. Диссертация, представленная на соискание ученой степени доктора филологических наук (10.02.19). Институт языкознания РАН.
- Кибрик А.А., Подлеская В.И. (ред.) 2009. *Рассказы о свидениях: корпусное исследование устного русского дискурса*. М.: Языки славянских культур.
- Barth-Weingarten D. 2016. *Intonation units revisited. Cesura in talk-in-interaction*. Amsterdam: John Benjamins.

- Chafe W. 1994. *Discourse, consciousness, and time*. Chicago: University of Chicago Press.
- Degand L., Simon A.C. 2009. Mapping prosody and syntax as discourse strategies: How basic discourse units vary across genres. In: D. Barth-Weingarten, N. Dehé, and A. Wichmann (eds.) *Where prosody meets pragmatics*. Bingley: Emerald, 81–107.
- Fleiss J.L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, v. 76, n. 5, 378–382.
- Himmelman N.P., Sandler M., Strunk J., Unterladstetter V. 2018. On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology*, v. 35, n. 2, 207–245.
- Izre'el S., Mello H., Panunzi, A., Raso T. 2020. Introduction: In search of a basic unit of spoken language. In: S. Izre'el, H. Mello, A. Panunzi, and T. Raso (eds.) *In search of basic units of spoken language: A corpus-driven approach*. Netherlands: John Benjamins Publishing Company, 1–32.
- Kibrik A.A., Korotaev N.A., Podlesskaya V.I. 2020. Russian spoken discourse: Local structure and prosody. In: S. Izre'el, H. Mello, A. Panunzi, and T. Raso (eds.) *In search of basic units of spoken language: A corpus-driven approach*. Netherlands: John Benjamins Publishing Company, 35–76.
- Landis J.R., Koch G.G. 1977. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, n. 1, 159–174.

Расширенная версия этой статьи находится по ссылке:

