

Стратегии членения спонтанной речи на синтаксические единицы¹

Strategies of delimitation of syntactic units in spontaneous speech

Рыко А. И. (aryko@mail.ru), **Степанова С. Б.** (stsvet_2002@mail.ru)

Факультет филологии и искусств Санкт-Петербургского государственного университета, Санкт-Петербург, Россия

На материале Звукового корпуса русского языка проведен перцептивный эксперимент по членению спонтанных монологов на предложения. Анализируются стратегии экспертов, принимавших участие в эксперименте.

1. Введение

Нет ничего проще, чем выделить предложение в письменном тексте. Знаки препинания — точка, вопросительный, восклицательный знаки, многоточие — почти однозначно отмечают конец предложения. И дальше можно определять его тип: сложное или простое, сложносочиненное или сложноподчиненное, с союзной связью или бессоюзное, односоставное или двусоставное и т. п., выявлять связи между членами предложения, анализировать интонацию чтения того или иного типа предложений и т. д. Казалось бы, те же процедуры легко проделает любой лингвист и со спонтанной звучащей речью. Однако, как показывает практика и как пишут исследователи, занимающиеся живой, «непричесанной», записанной в полевых условиях речью, выделить в ней предложение (фразу, высказывание — в данном случае для нас это синонимы) — задача весьма сложная, а может быть, даже невыполнимая.

Именно поэтому многие исследователи отказываются от этого понятия при работе со спонтанной речью.

Так, при аннотировании корпуса устных рассказов детей о сновидениях создатели корпуса сегментируют речь на элементарные дискурсивные единицы (ЭДЕ), определяя их, вслед за У. Чейфом, как некие «минимальные кванты устного дискурса» [Подлеская, Кибрик 2004: 1], «вербализующие "фокус сознания" говорящего, т. е. совокупность информации, которую селективное человеческое сознание может одновременно удерживать в активном состоянии» [Кибрик и др. 2008: 1]. И хотя авторы пишут,

что «типичная ЭДЕ совпадает по объему с предикацией (элементарным предложением, clause)» [там же], примеры показывают, что ЭДЕ московских ученых ближе к тому, что мы назвали бы «синтагмой».

В другом исследовании С.В. Андреева предлагает в качестве наименьшей структурной единицы синтаксиса устной речи принять конструктивно-синтаксические единицы (КСЕ), которые, с ее точки зрения, позволяют «охватить как предложенческие, так и непредложенческие, как предикативные, так и непредикативные речевые факты» [Андреева 2005: 9]. С. В. Андреева не разрабатывала специально принципы выделения КСЕ, лишь указала, что «при вычленении вспомогательных единиц речевой коммуникации учитывались следующие моменты: 1) коммуникативное, интонационное их выделение и образование синтагмы, 2) возможность при определенных условиях быть самостоятельной единицей» [там же: 45], что опять же говорит о близости этой единицы к синтагме.

Некоторые исследователи, определяя границы между предложениями в спонтанном тексте, прибегают к методу «пунктирования»: большая группа носителей русского языка ставит точки в орфографической расшифровке устных монологов, в которой убраны как знаки препинания, так и знаки пауз [Богданова, Бродт 2007]. Этот метод вполне применим тогда, когда исследователи интересуются прежде всего синтаксическими связями слов внутри предложений.

Мы же предполагали, что наличие звука поможет экспертам точнее и единодушнее определить границы такой привычной для всех единицы лингвистического анализа, как предложение. В связи с этим нас

¹ В настоящее время работа проводится при поддержке РГНФ — проект 07-04-12163в «Разработка информационной среды для мониторинга устной русской речи».

заинтересовал сам разброс мнений экспертов, принимавших участие в нашем эксперименте, мы попытались установить причины, вызывающие различия в стратегиях, и найти те факторы, которые определяют однозначное выделение границ предложения.

2. Материал и методика

Материалом в нашем эксперименте послужили спонтанные монологи и реплики из записей Звукового корпуса русского языка. Расшифрованные и представленные в орфографической форме без каких-либо знаков пауз или препинания, тексты были разосланы шести коллегам-лингвистам, имеющим достаточно большой опыт работы со спонтанной речью. Они могли неоднократно прослушивать записи. Задание было сформулировано так: поставить знаки конца предложений.

Выбор именно лингвистов на первом этапе проведения эксперимента обусловлен тем, что мы предполагали наличие у них сходного понимания термина «предложение» (в том числе и предложение в устной речи) и ожидали более-менее однотипных способов членения потока речи. Однако, как видно из таблицы 1, при большом количестве совпадений достаточно много и границ, поставленных одним-двумя испытуемыми.

Таблица 1

	T1	T2	T3	T4	T5	T6	T7	Все тексты
Количество границ	4	14	14	6	11	16	91	156
6 совпадающих границ		1	1	2	4	4	23	31
5 совп. гр.	1	4	1	2		2	14	24
4 совп. гр.			4			4	12	20
3 совп. гр.		1	2		1	2	9	15
2 совп. гр.	2	4	3	2	3	3	9	26
1 совп. гр.	1	4	3		3	1	24	36

3. Обсуждение результатов

3.1. Стратегии испытуемых

При анализе полученных результатов выяснилось, что испытуемые, которых мы привлекали к членению спонтанного текста на предложения, реализуют различные стратегии, которые можно условно назвать «максималистской» и «минималистской».

При реализации «максималистской» стратегии в качестве предложений выделяются длинные, многосинтагменные структуры. В этом варианте членения текста в большом количестве представлены бессоюзные предложения, сложносочиненные и сложноподчиненные предложения. При следовании «максималистской» стратегии испытуемый ориентируется на письменный текст, стремится увидеть аналог «литературного» предложения, отталкивается в первую очередь от синтаксической структуры, зачастую игнорируя при этом интонацию. «Минималистская» стратегия позволяет провести границу везде, где только это можно сделать: два и более простых предложений вместо одного бессоюзного сложного, ряд простых предложений с эффектом парцелляции вместо сложноподчиненных и сложносочиненных. Очевидно, что «минималистская» стратегия предполагает ориентацию на интонацию текста, допуская расширение набора возможных синтаксических структур. Фактически любая затянувшаяся межсинтагменная пауза в рамках такой стратегии — повод для проведения границы между предложениями. Любопытно, что один и тот же испытуемый придерживается то одной, то другой стратегии — ни один из них не был последователен на протяжении всего эксперимента.

«Максималистская» стратегия:

- (1) я вот так смотрю на такую вот машину сразу думаю не директор ли мой едет у него по-моему «Вольво» //

«Минималистская» стратегия:

- (2) я вот так смотрю на такую вот машину // сразу думаю не директор ли мой едет // у него по-моему «Вольво» //

Проведение некоторых границ определенно вызывает различную интерпретацию разными испытуемыми. Так, «пограничные маркеры» (*вот, как бы, ага, ну и вот*) выделяются только с одной стороны, присоединяясь к предложению справа или слева, то с двух сторон, т. е. как отдельные предложения.

- (3) *приходила когда я уже позавтракаю поиграю [3] вот [4] оставляла мне еду в термосе: приходила когда я уже позавтракаю поиграю вот // оставляла мне еду в термосе горячую vs приходила когда я уже позавтракаю поиграю // вот // оставляла мне еду в термосе горячую*

На наш взгляд, не имеет особого значения, с одной или с двух сторон выделяется такой маркер — его пограничный характер очевиден в любом случае, и, по большому счёту, испытуемые, считаю-

шие такой маркер отдельной единицей или служебным средством, включенным в другую единицу, друг другу не противоречат.

По-разному могут трактоваться и случаи обрывов или самокоррекции: наличие синтаксического сбоя (перестройки синтаксической структуры) приводит к тому, что испытуемые воспринимают его как границу между предложениями:

- (4) *у тебя нету этого ничего [6] те-е [2] тебе не надо ничего продавать понимаешь*
- (5) *с ним кане... [4] ну блин [1] ты помнишь как я парилса-то с этим*

Очевидно, что не все участки спонтанного текста членятся на предложения с одинаковой степенью легкости. Чем больше количество обрывов, слов-паразитов, пауз, чем труднее уловить смысл сказанного, тем больше усилий затрачивается на членение такого фрагмента на предложения и тем больше возникает разногласий у испытуемых.

Ср. «легкий» фрагмент:

- (6) *ну у нас тоже на лекциях я вот помню что все сидели вообще никто ничего не понимал [6] самые примерные просто успевали списывать там (нрзб) [6] ну я на практике еще что-то как-то понимала [2] а на теории так это вообще [1] никак [6] Котов у нас был этот прикольный*

и «трудный» фрагмент:

- (7) *так хорошо [4] просто Вик(?) а что это знаешь [3] а она что-то [2] такая [2] всем понятно [1] такая*

По-видимому, это можно считать аргументом в пользу того, что предложение в восприятии испытуемых по большей части является единицей семантического плана: если смысл от слушающего ускользает, то не принципиально, где именно провести границу и проводить ли ее вообще.

По-разному испытуемые членят на предложения монологи, произнесенные в разных условиях. В качестве особого случая можно рассматривать речь на ходу (во время прогулки по городу) при прерывистом дыхании с длительными паузами. Здесь при последовательной реализации «минималистской» стратегии как предложение интерпретируется практически любой отрывок, произнесенный на одном выдохе и имеющий предикацию (то есть предложение фактически совпадает с синтагмой). При последовательной реализации «максималистской» (синтаксически ориентированной) стратегии паузы, как правило, игнорируются и членение «мо-

нолога на ходу» по количеству выделенных предложений не отличается от монолога, произнесенного в спокойных условиях.

3.2. Характеристика синтаксических условий границ между «предложениями»

В результате нашего эксперимента оказалось, что некоторые границы между «предложениями» независимо от приверженности той или иной стратегии испытуемые установили единодушно. Проанализируем те случаи, которые большая часть испытуемых интерпретировала как границы между предложениями (4–6 совпадений).

А. Смена плана повествования

1) Переход от общего рассуждения к рассказу о конкретных событиях

- (8) *то есть учила естественно по советскому букварю другого не было [5] и там где Красная площадь на первой странице храм Василия Блаженного она тогда меня спрашивала*

Характеристика ситуации (в широком ее понимании — вплоть до характеристики эпохи) «другого не было» > рассказ о конкретном событии (бабушка задает вопрос по конкретной картинке).

- (9) *нормальное отношение достаточно между родителями так что [2] что такого [5] мне еще сегодня предстоит (нрзб)*

Общее рассуждение (не очень понятно даже о чем оно) > сообщение о своем намерении.

- (10) *самые примерные просто успевали списывать там (нрзб) [5] ну я на практике еще что-то как-то понимала*

Сообщение обо всех > сообщение о себе.

- (11) *и там большая такая кирха [1] но это не кафедральный... собор был но какой-то большой [8¹] я зашел просто [3] там идет не служба крестины были*

Общее описание сменяется рассказом о конкретном событии.

2) «Авторское резюме»: переход от частного к общему

¹ Часть материала прослушали 8 экспертов, в этом разделе мы рассматривали границы, единодушно определенные 7–8 из них.

Как правило, включает и какую-либо оценку событий говорящим; имеет показатели — слова с обобщающим значением (*то есть, ну в общем, ну, ну там, короче, ну знаешь, получилось, ну так*).

(12) *хьям бозий [2] храм божий [смех] [4] то есть это вот всё с ее слов реально вот помню до сих пор*

Рассказ о конкретном событии > оценка его (характеристика сохранности эпизода в памяти).

(13) *и причем как бы [1] Люба в термосе картошечка там [4] ну в общем чтобы подлиннее еще было*

(14) *у нас сегодня эта твоя ссылка на мужиков в красных труселях просто весь департамент валялся [2] я всем разослала [5] ну это надо [1] это [2] вы не поймете [2] это надо видеть*

(15) *ну вот [2] ты знаешь сколько они в Омске стоят [1] по Серёгиным данным [3] самый простой шкаф вообще без всяких-всяких наворотов просто [4] ну там же ну можно до бесконечности наворачивать всё зависит от фурнитуры от всего*

(16) *ну там же ну можно до бесконечности наворачивать всё зависит от фурнитуры от всего [5] короче стоит пять тысяч погонный метр [3] рублей*

(17) *ну и вот [3] пять тысяч [1] погонный метр [1] то есть в среднем шкаф выходит там тысяча тридцать [5] ну знаешь они там обычно пятиметровые*

(18) *и они всех провожают прихожан [5] у двери стоят и провожают [7] получилось что я первый выхожу [8] он меня спросил как мне понравилась служба [4] то есть всех в лицо знают [3] они своих прихожан всех знают и спрашивают...*

«Однозначными» границами с двух сторон выделен фрагмент, который несколько выбивается из контекста повествования (перечисления последовательных действий участников общей ситуации «окончание службы»): фрагмент *получилось что я первый выхожу* включает взгляд извне ситуации, некоторое авторское резюме.

(19) *и рассказал что туда каждое первое воскресенье... каждого месяца приезжает какая-то большая община из Петербурга [3] м-м [5] (8²) ну так хорошее впечатление [5] спросил*

как впечатление [3] так ненавязчиво но тем не менее... [4] спросил поговорили [1] и тут же следующие пошли

Снова в изложение последовательных событий включается «авторское резюме» (левая его граница усилена хезитационной паузой, правая граница размыта).

3) «Смена сюжета»

Характеристика макроситуации с разных сторон; переход к повествованию о новом этапе развития сюжета. Видимо, для всех примеров этой группы характерна смена субъекта повествования. Во многих случаях подобный переход отмечается пограничными маркерами (*ага, ну и вот, так вот, причем* (= «с другой стороны»)).

(20) *ну я на практике еще что-то как-то понимала [2] а на теории так это вообще [1] никак [5] Котов у нас был этот прикольный*

О своем восприятии учебной дисциплины > о манере преподавателя.

(21) *такая что-то сама с собой ла-ла-ла-ла-ла [4] ага [3] (8) вся аудитория просто тупо сидит так (2) о-о-о*

Описание ситуации с разных сторон, характеристика разных ее участников. Пограничный «финальный» маркер *ага* может выделяться с обеих сторон.

(22) *надо это всё узнать [4] я вот так смотрю на такую вот машину [1] сразу думаю не директор ли мой едет [3] у него по-моему «Вольво» [5] ну и вот [3] пять тысяч [1] погонный метр*

Комментарий к тому, что попало в поле зрения > возвращение к основной теме монолога. Пограничный маркер *ну и вот* часто выделяется с двух сторон.

(23) *то есть ты с людей берешь тридцатник [1] а тратишь на это допустим тысяча десять [4] ну надо только потом кого-нибудь мастеров [1] поначалу там можно самим вот допустим Вован Аримов будет собирать*

Переход к новому этапу развития сюжета.

(24) *первый шкаф конечно будет сложный [4] вот Серёга говорит что он дал рекламу [1] в Омске [1] там сейчас кому-то он что-то будет делать*

Планы на будущее > рассказ об аналогичном опыте приятеля.

² Очевидно, что все аудиторы отметили здесь границу между «предложениями». Мы сочли безразличным, где именно они ее отметили — до паузы хезитации или после нее, и посчитали это единой границей.

- (25) *там* *фурнитуры* *ну эти эти* *всякие* [1] *ролики* [1] *где там конструктор всё собирается* [5] *ну а люди которые в этом не понимают они готовы за это платить*

Характеристика ситуации с разных сторон — с технической и со стороны восприятия потенциальных клиентов.

- (26) *а потом вот ты понимаешь нанимаешь уже плотника* [2] *который* [4] *просто я в детстве работал на сборках*

Б. «Формальные» границы

4) Граница между ответом на предшествующий вопрос собеседника и собственно повествованием

- (27) *а-да-да* [4] *а потом вот ты понимаешь нанимаешь уже плотника*

- (28) *да понравилось* [7] *интересно* [1] *конечно* [2] *и разница большая между Хельсинки и Стокгольмом*

5) Как правило, с двух сторон выделяются вопросы

Это происходит благодаря специфическому интонационному оформлению и характерному синтаксическому построению высказывания. Часто граница справа оказывается «механической» (т. е. дальше идет реплика собеседника), поэтому нами не рассматривается. Границы слева и справа появляются в случае наличия риторического вопроса или при передаче вопросов диалога.

- (29) *ну я там говорил Кремль допустим* [1] *"р" не выговаривал* [1] *Кйемль* [4] *а это что* [1] *за сбор?* [4] *хьям бозий* [2] *храм божий* [смех]

- (30) *мне еще сегодня предстоит* (нрзб) [5] *можешь со мной пойти?* [4] *а это тебе не важно*

- (31) *я задумался* [1] *стал более развивать* [5] *знаешь что такое шкафы-купе?*

- (32) *тебе не надо ничего продавать понимаешь?* [4] *ты рекламу даёшь* [1] *люди тебе звонят*

- (33) *да так просто задумался* [1] *хотя мне бизнес на хер не нужен* [5] *Как у тебя с Леной?*

В. Формальные показатели границ

7) «Пограничные маркеры» (вот, как бы, ага, ну и вот)

Видимо, можно различать финальные маркеры («как бы», «вот так»), которые сигнализируют о конце предложения, начальные маркеры («ну»,

см. примеры в разделе «Авторское резюме»), которые вводят новое предложение, и пограничные маркеры («вот», «ага», «ну и вот», «и всё»), которые, как правило, отделяются значительными паузами с обеих сторон и часто интерпретируются испытуемыми как самостоятельные единицы.

- (34) *меня мама учила читать* [3] *она уходила на работу с утра* [1] *ну пока я спала и там на несколько часов* [2] *приходила когда я уже позавтракаю поиграю* [2] *вот* [4] *оставляла мне еду в термосе* [2] *горячую*

Пограничный маркер «вот» обычно «финальный», может выделяться с двух сторон, сопоставим с резюме («свернутое резюме»).

- (35) *чтобы я читала да* [2] *как бы* [4] *причем писала крупными буквами*

«Как бы» — финальный «маркер неуверенности», предполагающий, видимо, затяжную паузу.

- (36) *ну даже дело не в закалке наверно* [2] *как то ну у них это все объясняется как бы* [4] *нормальное отношение достаточно между родителями так что* [2] *что такого*

- (37) *такая что-то сама с собой ла-ла-ла-ла-ла* [4] *ага* [3] *вся аудитория просто тупо сидит так* (2) *о-о-о*

- (38) *я вот так смотрю на такую вот машину* [1] *сразу думаю не директор ли мой едет* [3] *у него по-моему «Вольво»* [5] *ну и вот* [3] *пять тысяч* [1] *погонный метр*

- (39) *говорят что зимой даже лучше чем летом* [5] *вот так*

4. Интонационный комментарий

Не будет преувеличением сказать, что интонационное оформление речи — самое вариативное из всех лингвистических явлений. Особенно это касается спонтанной речи. Во многом именно этим объясняется то разнообразие стратегий, которое продемонстрировали наши эксперты при членении речевого потока на предложения.

Возможно, у каждого носителя русского языка есть представление об «интонации точки» — интонации завершения высказывания. Акустически «интонация точки» обычно выражается в падение тона ниже среднего, за которым часто следует пауза. Если эти условия соблюдены и с семантико-синтаксической точки зрения граница представля-

ется вполне уместной, то в этом случае эксперты единодушно ставят ее. Например:

- (40) оставляла мне еду в термосе [2] горячую [4] а термос прятала [6] а оставляла записку на столе где она прятала термос

На слове *прятала* наблюдается падение ЧОТ с 240 Гц до 177 Гц на ударном гласном и падение продолжается до 100 Гц на заударных слогах, затем следует пауза длительностью в 433 мс. Эта «интонационная» граница» поддерживается возможностью постановки здесь и синтаксической границы. Все эксперты ее отметили.

Практически все «уверенные» границы в эксперименте характеризуются аналогичными изменениями ЧОТ и наличием после них либо паузы, либо смены говорящего.

Сомнения начинаются тогда, когда те же самые просодические особенности появляются в местах, где с синтаксической точки зрения точки быть не должно. В примере (40) падение ЧОТ на слове *в термосе* не меньше, чем в слове *прятала*, пауза после него — 1776 мс, однако слово *горячую* явно связано со словом *еду*, поэтому только двое экспертов увидели в данном контексте эффект парцелляции и выделили *горячую* в отдельное предложение, остальные сочли такую длительную паузу чисто хезитационной.

Не меньше сомнений вызывают и случаи, когда синтаксическая граница вполне возможна, однако

вместо падения тона слышен ровный тон или даже его повышение. Именно так часто оформляются следующие друг за другом простые предложения:

- (41) *слушай Славик [3] я тут придумал вид бизнеса [3] мне кажется достаточно перспективный [2] и не такой геморный [6]*

Кроме последнего слова все остальные выделенные слова характеризуются повышением тона, что вызывает разночтение у наших испытуемых.

5. Основные итоги

Наш эксперимент показал, что профессиональные лингвисты, сталкиваясь при анализе звучащей спонтанной речи с неполной согласованностью семантических, синтаксических и интонационных характеристик отрезков звучания, прибегают к разным стратегиям вычленения синтаксических единиц. Некоторые предпочитают интерпретировать отдельные синтагмы как самостоятельные простые предложения, некоторые объединяют их в сложные.

Анализ «уверенных» границ помог выявить те формальные признаки, которые помогут в дальнейшем расшифровщикам спонтанной речи единообразно выделять в устных текстах предложения.

Литература

1. Андреева С. В. Элементарные конструктивно-синтаксические единицы устной речи и их коммуникативный потенциал: Автореф. дис. ... докт. филол. наук // Саратов, 2005.
2. Кибрик А. А., Подлеская В. И., Дараган Ю. В., Ефимова З. В., Коротаев Н. А., Литвиненко А. О., Цуканова В. Л. Проблема сегментации устного дискурса и когнитивная система говорящего // www.ksu.ru/ss/cogsci04/science/cogsci04/107.doc
3. Богданова Н. В., Бродт И. С. Спонтанный монолог: синтаксические характеристики текста и уровень речевой культуры говорящего // Вестник Санкт-Петербургского государственного университета. Серия 9 (Филология). Вып. 1 (Часть I). Март 2007 г. С. 35–43.
4. Подлеская В. И., Кибрик А. А. Транскрипция устного дискурса для нужд корпусных исследований // <http://www.dialog-21.ru/Archive/2004/Podlesskaja.pdf>