

В. И. Подлеская
V. I. Podlesskaya

СИНТАКСИС И ПРОСОДИЯ САМОИСПРАВЛЕНИЙ ГОВОРЯЩЕГО ПО ДАННЫМ КОРПУСА С ДИСКУРСИВНОЙ РАЗМЕТКОЙ¹

SYNTAX AND PROSODY OF SELF-REPAIRS: A CORPUS APPROACH

Аннотация. Качественный и количественный анализ около двухсот самоисправлений говорящего в спонтанной монологической речи произведен на материале экспериментального корпуса «Веселые истории из жизни» из коллекции «Корпуса звучащей речи» (spokencorpora.ru). Разметка дискурсивной информации в корпусе позволила выделить основные стратегии преодоления речевых затруднений, затрагивающих коммуникативную, просодическую и грамматическую организацию устного дискурса.

Abstract. The paper investigates self-initiated self-repairs in «Funny stories» – a subcorpus of the Prosodically Annotated Corpus of Spoken Russian (spokencorpora.ru), a unique resource where speech disfluencies in Russian natural discourse are systematically registered. I present results of a qualitative and a preliminary quantitative analysis of about 200 repairs and explore main repair strategies.

1. Постановка вопроса

Корпусное исследование нарушений гладкости речи – сравнительно новая область, и если для английского, шведского, китайского, японского и некоторых других языков уже получены серьезные результаты, то для русского языка эта работа только начинается. В докладе исследуются самоисправления говорящего в экспериментальном корпусе «Веселые истории из жизни» из коллекции «Корпуса звучащей речи», доступной в пилотном

¹ Работа выполнена при поддержке РГНФ (грант №12-04-00258) и Программы стратегического развития РГГУ.

режиме на сайте **spokencorpora.ru** (разрабатывается коллективом исследователей при участии автора). Основные принципы разметки дискурсивной информации в корпусах этой коллекции были сформулированы в работе [Кибрик, Подлесская 2009]². В этой же работе были сформулированы базовые подходы к систематизации речевых сбоев и к их аннотированию. Для последующего изложения существенны следующие сведения о типах дискурсивной информации и способах их аннотирования в «Корпусах звучащей речи». (1) В транскриптах произведено деление на элементарные дискурсивные единицы (ЭДЕ). ЭДЕ – минимальный фрагмент дискурса, который способен вступать в смысловые отношения с другими фрагментами в структуре дискурса. Просодически, ЭДЕ – это автономная единица, интонационно организованная вокруг одного речитативного (фразового) акцента; с синтаксической точки зрения ЭДЕ в сильной степени коррелирует с клаузой. (2) Транскрипты снабжены просодической разметкой – в частности, указаны локализация акцентов и тональный тип акцента (нотируется перед словом иконически с помощью косых черт), размечены абсолютные и заполненные паузы³. Ударный слог в слове – носителе фразового акцента подчеркивается. (3) Речевой сбой маркируется знаком «==» на границе ЭДЕ, и знаком «||» внутри ЭДЕ.

Экспериментальный корпус «Веселые истории из жизни» состоит из 40 рассказов о смешном происшествии в жизни респондентов (возраст от 18 до 60 лет). Общая длительность – около 1 часа 10 минут; объем – около 7 тысяч словоупотреблений. При сплошной ручной проверке в корпусе было

² Кибрик А.А., Подлесская В.И. (Ред.). (2009). Рассказы о сновидениях: Корпусное исследование устного русского дискурса. М.: ЯСК.

³ В приводимых ниже примерах из корпуса числовые значения пауз удалены из соображений экономии места и нотированы лишь соответствующим числом точек или литер э/м (для заполненных пауз) – две при длительности паузы до 0,5 сек, три – от 0,5 сек до 1 сек, четыре – более 1 сек.

обнаружено 194 случая самоисправлений говорящего. Они были подвергнуты анализу по четырем параметрам, которые рассмотрены в следующем разделе.

2. Выявленные типы самоисправлений

2.1. Он-лайн коррекции vs. офф-лайн коррекции

Эти два важнейших класса самоисправлений выделяются в зависимости от того, насколько оперативно говорящий реагирует на обнаруженную проблему. В режиме он-лайн коррекции говорящий реагирует на обнаруженную проблему немедленно, поэтому возникает наблюдаемая точка прерывания и речевой отрезок до точки прерывания обладает незавершенностью по совокупности лексических, грамматических и просодических критериев. После прерывания говорящий заменяет забракованный фрагмент на другой или повторяет первоначально забракованный фрагмент и продолжает развертывать речь таким образом, чтобы забракованный фрагмент мог быть безболезненно «стерт» и материал до забракованного фрагмента и материал после него, сомкнувшись, образовали бы правильную, т. е. удовлетворяющую говорящего последовательность. Так в следующем примере вторая ЭДЕ обрывается и «аннулируется» говорящим:

(1)

•• В итоге меня на скорой увезли ночью в \больн <u>и</u> цу.
••• ээ /Приезжаем ==
•• Привозят в /больн <u>и</u> цу,

Однако возможен и менее «авральный» режим преодоления обнаруженной проблемы: при офф-лайн коррекциях уже готовый и транслированный фрагмент дискурса редактируется постфактум, т. е. уже по завершении проблемного отрезка слушающий информируется о том, что этот отрезок подлежит уточнению или исправлению. Эпизод редактирования может сопровождаться эксплицитованным признанием ошибки и уточняющими выра-

жениями – *вернее, точнее, то есть* и т.п. Так в следующем примере первая ЭДЕ редактируется в двух следующих, при этом никаких обрывов или иных внешних симптомов сбоя нет, все четыре ЭДЕ полностью и правильно сформированы – и синтаксически, и просодически:

(2)

•• Но-о /он кстати был очень \прилич ^н ный молодой человек.
Ну в смысле что он был /молодой,
и-и •• ^ну-у •• как бы /внешне не \выг ^л ядел как фрик.

Редактирование часто эксплуатирует стандартные способы выражения противопоставления, как в третьей-четвертой ЭДЕ в (3):

(3)

я когда \маленькая,
на балкон вот помню один раз /вышла,
(Вернее не /помню,
мне потом \мама рассказала.)
•• –вот,
и меня эти соседи стали \выспрашивать всё это.

При отсутствии сегментных средств функцию сигнала о наличии редактирования может полностью взять на себя просодия. Так в примере (4) в последней строке в этой функции используется эмфатическое произнесение слова *пятого* с контрастным акцентом в значении ‘[не шестого, а] пятого’:

(4)

•• И вот папа … шестого числа •• значит /доложил,
•• {ЦОКАНЬЕ} •• \пятого числа папа /доложил,
•• и шестого пошёл бегать •• ээ по всяким /авиакассам,

В некоторых случаях говорящий может сменить режим самоисправления по ходу произнесения – сначала обрывает текущую ЭДЕ, как если бы за забракованным фрагментом предполагался откорректированный коррелят, однако не справляется с задачей и вместо искомого слова/словосочетания

прибегает к косвенным, описательным способам передачи смысла. Так в (5) оборвана третья ЭДЕ, однако он-лайн коррекция не удалась – говорящий так и не смог назвать театральный жанр и переключился на режим редактирования в следующих семи ЭДЕ:

(5)

{СМЕХ} Короче мы пошли в /театр,
•• {ЦОКАНЬЕ} ээ сели на первый /ряд,
на какой-то та-ам-м ==
ну это был не мюзикз= не /мюзикл,
а-а ==
и-и н-не /опера,
а \такая в общем \обычная,
там-м они /пою-ут,,,
и-и /тексты говорят,,,
и всё \такое.

При аннотировании коррекций такие случаи относились нами к «смешанному типу» по параметру «он-лайн vs. офф-лайн».

2.2. Контактные коррекции vs. дистантные коррекции

Эти два класса самоисправлений выделяются в зависимости от того, имеется ли между забракованным фрагментом и его откорректированным коррелятом материал, не затронутый самоисправлением.

Типичным примером контактной коррекции является приведенный выше (1), где забракованный фрагмент *Приезжаем* непосредственно предшествует откорректированному корреляту *Привозят*. Дистантная коррекция представлена в (6), где забракован фрагмент *спустя несколько лет*, а его коррелят появляется лишь после парентезы (строки 3–4):

(6)

••• ээ Проблема сос= в /том,
что-о •• вот спустя несколько /лет ==

ээ (На самом деле устройство очень /простое,
два пластмассовых /бачка •• /и-и насосик для' ••• подачи смывной \жидкости.)
<ну> вот <к=> спустя •• несколько /лет ••• эээ •• /вдруг вышел из /строя •• вот этот \насосик,

Дистантные коррекции возникают, как правило, в тех случаях, когда говорящий преждевременно приступает к произнесению материала, понимает, что слушающему недостает фоновой информации для правильной интерпретации сказанного, достраивает необходимый участок дискурса и возвращается к основной линии повествования. Дистантными могут быть не только он-лайн коррекции, как (6), но и офф-лайн коррекции. Так, в (7) представлен случай офф-лайн коррекции, поскольку необходимость вставки строк 2–4 – не приводит к обрыву в строке 1:

(7)

•• И один раз когда мы с сестрой /спали,
•• (Кстати конечно я тоже этого не /помню,
нам рассказала /бабушка с мамой.
\Да.)
•••• когда мы /спали,

2.3. Микрокоррекции vs. макрокоррекции

Два дополнительных параметра для он-лайн коррекций связаны с объемом исправления и степенью его радикальности. При макрокоррекциях либо забракованный фрагмент, либо материал между забракованным фрагментом и его откорректированным коррелятом включают более одной ЭДЕ, как это имеет место в (6) выше, где вставка включает две ЭДЕ. При микрокоррекциях самоисправление происходит в пределах одной ЭДЕ, как в (8):

(8)

это /дерево своим || ••(0.09) своей /веткой забралó ещё один
маленький ••\кактусик.

2.4. Повторы vs. модификации vs. отмены

Степень радикальности он-лайн коррекции связана с характером производимой операции: забракованный элемент может быть повторен, частично изменен или полностью отвергнут. Частичные изменения могут быть связаны с выбором как лексики, ср. (1), так и грамматики, ср. (8). Возможен и такой специфический класс модификации, как замена так называемого маркера препаративной подстановки (*этот самый, как его*) на искомое выражение:

(9)

(Вот в \старых домах-х, •• /решетки на окнах,,,

'и-и ••• \такие || ••• типа •• \полуподвальчики такие маленькие.)

Полная отмена забракованного сегмента может быть вызвана как отказом от текущего замысла, так и сбоем в артикуляторной программе (оговоркой), ср. не мюкиз= || не /мюзикл в (5).

3. Частотность самоисправлений и некоторые обобщения

Частотность самоисправлений в корпусе – 2,8 на 100 слов (194 на 7000 слов). Этот показатель хорошо вписывается в интервал 1–6 самоисправлений на 100 слов, на который указывают большинство опубликованных источников для других языков⁴.

⁴ См. например: *Shriberg E.E.*, Preliminaries to a Theory of Speech Disfluencies. PhD thesis, University of California, Berkeley, 1994; *Eklund R.* Disfluency in Swedish human–human and human–machine travel booking dialogues. PhD thesis, Linköping University, Sweden, 2004; *Tseng S.-C.* Repairs in Mandarin conversation // *Journal of Chinese Linguistics* 34(1). P. 80–120, 2006; *Maruyama T., Sano Sh.* Classification and Annotation of

Он-лайн коррекции значительно превосходят по частотности офф-лайн коррекции (149/77% против 35/18%, смешанные – 10/5%). Контактные коррекции – значительно превосходят по частотности дистантные (146/75% против 45/25%). Внутри он-лайн коррекций микрокоррекции значительно более частотны, чем макрокоррекции (125/84% против 24/16%). Он-лайн коррекции следующим образом распределены по типу операции: повторы – 64/43%; модификации (включая замены с использованием препаративных подстановок) – 60/40%; отмены (включая оговорки) – 26/17%. В целом же, коррекции обнаруживают тенденцию употребляться кластерно – в проблемных точках дискурса говорящий обычно задействует сразу несколько техник самоисправления:

(10)

Почему-то у меня не \з-заладилось с /коржами с нашей диск= ** ммм доисторической ==
такая —
/ <u>знаете</u> ;
— *** такая *** ээ ** ну-у газ-зоявая здор-ро³ = р-рухлядь этих семидесятых брежневских —
** то есть \шестидесятых,
** ну в общем очень \старых,
— ** /го <u>дов</u> ,
вообще ни с чем не заладилось у меня в этот /д <u>ень</u> ,
\у <u>жас</u> .

Дальнейшее исследование речевых сбоев в корпусах с дискурсивной разметкой позволит установить, насколько универсальны обнаруженные нами типы самоисправлений и их количественное распределение.