

КОГНИТИВНАЯ НАУКА В МОСКВЕ: НОВЫЕ ИССЛЕДОВАНИЯ

2013

**МАТЕРИАЛЫ
КОНФЕРЕНЦИИ**



Под ред. Е.В. Печенковой, М.В. Фаликман

СТРАТЕГИИ АВТОКОРРЕКЦИИ ПРИ РЕЧЕВОМ СБОЕ В НЕПОДГОТОВЛЕННОМ МОНОЛОГИЧЕСКОМ ДИСКУРСЕ

Подлесская В.И.

podlesskaya@ocrus.ru

РГГУ

1. Данная работа ставит перед собой две взаимосвязанные цели (1) оценить частотность самоисправлений говорящего в русской спонтанной монологической речи в сравнении с имеющимися в литературе данными по другим языкам; и (2) проанализировать количественно и качественно основные стратегии самоисправлений. Материалом исследования послужила коллекция «Корпуса звучащей речи», содержащая аудиофайлы с синхронизированными транскриптами (доступна в пилотном режиме на сайте www.spokencorpora.ru, разрабатывается коллективом исследователей при участии автора). Основные принципы разметки дискурсивной информации в экспериментальных корпусах этой коллекции были сформулированы в работе (Кибрик, Подлесская 2009). В этой же работе были сформулированы базовые подходы к систематизации речевых сбоев и к их аннотированию¹.

2. Частотность самоисправлений оценивалась в трех корпусах коллекции: «Рассказы о сновидениях» (129 монологов, респонденты от 7 до 17 лет); «Рассказы сибиряков о жизни» (17 монологов, респонденты от 19 до 70 лет) и «Веселые истории из жизни» (40 монологов, респонденты от 18 до 60 лет). Самоисправления говорящего фиксировались путем сплошной ручной проверки. Общее число самоисправлений в корпусах и

¹ Для последующего изложения существенны следующие сведения о типах дискурсивной информации и способах их аннотирования в корпусе. (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 рематического (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 паузы. Ударный слог в слове – носитель фразового акцента подчеркивается. (3) Речевой сбой маркируется знаком «==» на границе ЭДЕ, и знаком «||» внутри ЭДЕ.

их частотность в приведении к числу слов и времени звучания представлены в Таблице 1:

Таблица 1. Частотность самоисправлений

Название корпуса	Объем корпуса (слов)	Общая длительность звучания (мин)	Общее число самоисправлений в корпусе	Число самоисправлений в минуту	Число самоисправлений на 100 слов
Рассказы о сновидениях	14000	120	412	3.4	2.9
Рассказы сибиряков о жизни	5000	40	132	3.3	2.6
Веселые истории из жизни	7000	70	194	2.8	2.8

Несмотря на различие в возрасте респондентов, разброс в частотности по трем исследованным нами корпусам оказался весьма незначительным. Зарегистрированная частотность самоисправлений — 2.8–3.4 в минуту, 2.6–2.9 на 100 слов — хорошо согласуется с данными по другим языкам. В целом, частотность самоисправлений обычно фиксируется в примерном интервале от одного до семи в минуту; в диалогах — выше, чем в монологах, в бытовой речи — выше, чем в официальной. Так, для китайских диалогов приводятся сведения о 5.4 случаях самоисправлений в минуту (Tseng 2006); для английских устных пересказов — 1.9–3.7 на 100 слов (Fraundorf and Watson 2008); для венгерских диалогов — 3.8 в минуту (Németh 2012); для японских монологов — 1.2 на 100 слов (Matsuyama and Sano 2006). Таким образом, полученные нами данные позволяют предположить, что межъязыковое, межжанровое и межвозрастное варьирование частотности самоисправлений находится в пределах интервала, отражающего универсальные тенденции речепроизводства.

3. Для одного из корпусов коллекции — «Веселых историй из жизни» — мы провели более подробное исследование, позволяющее выявить основные стратегии говорящего при самоисправлении. Обнаруженные в данном корпусе 194 случая самоисправлений были оценены по следующим двум параметрам: (1) наличие vs. отсутствие речевого сбоя, т. е. нарушений плавности развертывания дискурса и (2) степень изоморфизма фрагмента, забракованного говорящим, и фрагмента, являющегося откорректированным коррелятом «брака».

3.1. Наличие речевого сбоя признавалось в тех случаях, когда обнаруживались симптомы разрушения лексико-грамматической или коммуникативно-просодической когерентности дискурса. Важнейшим симптомом является появление фрагмента, который при самоисправлении подлежит «стиранию», т. е. материал до этого фрагмента и после него, сомкнувшись, мог бы образовать когерентную и удовлетворяющую говорящего последовательность. Так, в (1) забракованным фрагментом является *тряс=*, его удаление дает цельнооформленную последовательность *поезд тряхануло*.

(1)

... ээ Ну поезд тряс= /-тряхануло,,,

В 80 случаях из 194 (41 %) в точке сбоя имеется обрыв слова — как в (1), реже — обрыв ЭДЕ, как в (2) в первой и третьей строках:

(2)

... Начал ==

... Просек эту /фишку,

.. и начал ==

они ж коты \чего делают,

они .. —писают,

... начал описывать им ... —тапочки например какие-нибудь,,,
--

Однако самоисправление возможно и без речевого сбоя. Так в (3) первая ЭДЕ исправлена во второй, при этом никаких фрагментов, подлежащих отбраковке, обрывов или иных внешних симптомов сбоя нет, обе ЭДЕ полностью и правильно сформированы — и синтаксически, и просодически:

(3)

ну у нас там у \многих жили кстати в общегаге-е ээ \коты.

.. \Кошки.

Из 194 самоисправлений в 164 (85 %) имеется речевой сбой, в 30 случаях (15 %) — сбоя нет.

3.2. Полный изоморфизм забракованного фрагмента и его откорректированного коррелята наблюдается при повторах фрагмента (как в (2), где *начал* воспроизводится в последней строке) или в случаях, типа (1) и (3), подчиняющихся правилу сочиненной структуры Левелта (1993), согласно которому «правильное» самоисправление допускает интерпретацию через вопрос с альтернативным сочинением, ср. «поезд тряс(ло) или тряхануло?», «в общегаге жили коты или кошки?» Возможны и менее очевидные случаи изоморфизма, когда имеется лишь частичный структурный параллелизм забракованного и откорректированного фрагментов,

как в (4), где откорректированный коррелят вводится в третьей-четвертой строках не непосредственно, а через конструкцию с отрицанием:

(4)

· · обслуживали быто= ” ну всякую ==
не /бытовую,
а-а · · \копировальную технику.

Отсутствие изоморфизма наблюдается в двух основных случаях — если говорящий полностью отказывается от забракованного фрагмента, меняя линию повествования, как в (5), или если говорящий прибегает к перефразированию и детализации протяженного сегмента текста, не идентифицируя конкретный проблемный участок, как в (6):

(5)

· · У нас был · · черный потом огромный /потолок ==
\А!,
еще \больше того,
он же стал это \тушить.

(6)

И он должен был поехать в \Ташкент.
· · {ЧМОКАНЬЕ} И там · · торчать какое-то \время.
· · {ЧМОКАНЬЕ} \Вот.
То есть у него был /доклад,
а потом он еще должен был · · какое-то время там \оставаться.

Из 194 самоисправлений в 158 (81 %) имеется изоморфизм, в 31 случае (16 %) изоморфизма нет, в 5 случаях (3 %) наблюдается частичный изоморфизм.

4. Совместное применение двух рассмотренных выше параметров дает распределение самоисправлений, обобщенное в таблице 2:

Таблица 2. Количественное распределение самоисправлений по параметрам «наличие сбоя» и «изоморфизм»

	+ сбой	– сбой
+ изоморфизм (включая частичный)	149	14
– изоморфизм	16	15

Как видим, абсолютное большинство самоисправлений — 149 из 194 (77 %) — демонстрируют одновременно и наличие речевого сбоя, и изоморфизм забракованного фрагмента и откорректированного коррелята, как в (1) и (2). Это «золотой стандарт» автокоррекции, самоисправления такого типа происходят «он-лайн», в соответствии с «главным правилом прерывания» В. Левелта (1993): речь должна быть прервана незамедлительно, как только обнаруживается проблема. Полная противополож-

ность им — «офф-лайн» коррекции, в которых нет ни симптомов речевого сбоя, ни структурного изоморфизма, как в (6), их зафиксировано 15 (8 %). Представлены и промежуточные случаи — со сбоем без изоморфизма, как в (5), их 16 (8 %), или с изоморфизмом, но без речевого сбоя, как в (3), их 14 (7 %). В докладе будут подробно рассмотрены типы изоморфизма между забракованным фрагментом и его откорректированным коррелятом при он-лайн и офф-лайн коррекциях и в промежуточных случаях. В частности, будет показано, каким образом изоморфизм может затрагивать разные языковые уровни — грамматику при сохранении лексического тождества, ср. смену глагольного вида в (1); лексику при сохранении грамматического оформления, ср. замену *коты* на *кошки* в (3); фонологию при сохранении лексического и грамматического статуса, ср. сбой в артикуляторной программе при оговорках, как в (7):

(7)

и они значит вмесело весело маршируют \наверх.

Литература

1. Кибрик А.А., Подлеская В.И. (Ред.). Рассказы о сновидениях: Корпусное исследование устного русского дискурса. М.: ЯСК, 2009.
2. Tseng, S.-C, “Repairs in Mandarin conversation”, *Journal of Chinese Linguistics* 34(1): 80–120, 2006.
3. Maruyama, Takehiko and Sano, Shin'ichiro, “Classification and Annotation of Self-Repairs in Japanese Spontaneous Monologues”, *LPSS — Linguistic Patterns in Spontaneous Speech*, 283-298, Taipei, November 2006.
4. Fraundorf, S.H. and Watson, D.G. “Dimensions of variation in disfluency production in discourse”, in J. Ginzburg, P. Healey, and Y. Sato [Eds], *Proceedings of LONDIAL 2008, the 12th Workshop on the Semantics and Pragmatics of Dialogue*, 131-138. London: King's College London, 2008.
5. Levelt, Willem J.M. *Speaking: From Intention to Articulation* (ACL-MIT Series in Natural Language Processing). 1993.
6. Németh, Zsuzsanna, “Recycling and replacement self-repairs in spontaneous Hungarian conversations”, *Proceedings of the First Central European Conference in Linguistics for postgraduate Students*, 211–224, 2012.

Работа выполнена при поддержке РФФИ (грант №13-06-00179)
и Программы стратегического развития РГГУ.