

О НЕКОТОРЫХ ЛЕКСИЧЕСКИХ «ОТКРЫТИЯХ»
НА МАТЕРИАЛЕ РУССКОЙ СПОНТАННОЙ РЕЧИ
(корпусное исследование)

SOME LEXICAL “DISCOVERIES” ON THE
MATERIAL OF RUSSIAN SPONTANEOUS SPEECH
(corpora research)

Н.В.Богданова, Н.А.Осьмак
Санкт-Петербург - Москва
«Диалог» – 2011 г.

Звуковой корпус русского языка

- Способ мониторинга и фиксации звукового материала естественного языка
- Организация информационной среды и программного инструментария для нужд интегрального моделирования речи
- Корпусный подход к описанию «лингвистической истории большого города» (Б. А. Ларин)
- Источник новых учебных материалов и база для изучения и преподавания языка в его звуковой форме
- Два разных блока – САТ и ОРД

Блок САТ

(Сбалансированная Аннотированная Текстотека)

- запись по «принципу ковчега»
- естественная речь человека в рамках типичных речевых сценариев (**лингвистическая балансировка**):
 - чтение
 - пересказ
 - описание зрительного ряда
 - свободный рассказ
- учет социальных характеристик говорящего (**социолингвистическая балансировка**)
- учет психологических характеристик говорящего (**психолингвистическая балансировка**)
- речь 5-ти профессиональных групп информантов (медики, юристы, «компьютерщики», филологи, философы), и нескольких групп студентов
- около 700 текстов, около 40 часов звучания

ОРД - «ОДИН РЕЧЕВОЙ ДЕНЬ»

- повседневная речь носителей языка
- запись по «принципу невода»
- методика 24-часовой записи
- 47 информантов
- + около 600 коммуникантов
- 400 часов звучания
- 40 часов расшифровано и размечено в программе ELAN
- около 300 тыс. словоупотреблений

Одна из возможностей использования материала Звукового корпуса

Создание лексикографического описания бытовой спонтанной звучащей речи

Словарь
русской
бытовой
разговорной
речи

Словарь
контекстных
экспрессем
русской
разговорной
речи

Словарь
хезитаци-
онных
конструкций

Словарь
редуцирован-
ных форм
русской речи

Материалы к словарю русской бытовой разговорной речи

- Словник – по дифференциальному принципу:
 - Новые лексемы
 - Новые семантемы
 - Новые особенности функционирования
- Проверка по Национальному корпусу и сети Интернет
- Обращение к ментальному лексикону носителя языка (особенно – стилистические пометы)

ХЛЕБ

МАС:

- 1) *только ед.ч.* Пищевой продукт, выпекаемый из муки. *Ржаной хлеб.*
- 2) *только ед.ч.* Зерно, из которого готовится мука для выпечки такого продукта. *Везти хлеб на элеваторы.*
- 3) (мн. хлеба□), *перен., разг.* Пища, пропитание. *Противен мне хлеб ваш.*
- 4) *перен., только ед.ч., разг.* Средства к существованию. *Добывать хлеб трудом.*

ОРД:

- 16 словоупотреблений
- Место в частотном словнике - от 1643 до 1836
- 5 употреблений (31,3 %) не укладываются в словарные определения:
 - *хлеб намазывала маслом*
- Много вопросов:
- *хлеб / это тоже живое существо / # да // # а вы его ножом;*
- *булки нет в доме / есть только хлеб;*
- *я (...) пакеты как хлеб покупаю на работу // *П у меня в одном (...) пакет кармане / и второй;*
- *буду как в детстве сейчас / есть сосиски с хлебом;*
- *хлеба надо тебе ?*

бабушкин хлеб - ?

- *хлебушек есть // но он такой //*
**П назыв... он назывался /*
бабушкин хлеб //
- *а / почему бабушкин ?*
- *не знаю // так называется //*
бабушкин хлеб почему-то //
- **Н так называется? *П ёшкин /кот!*
- *ну вот // ну давай же шить !*

ДЕВОЧКА и ДЕВЧОНКА

- **МАС:** ‘ребенок или подросток женского пола’
- **ОРД:** ‘лицо одного возраста с говорящим’ (мн.ч.)
- *я жалею / что я не пошла с **девочками** / **девочки** как раз ходили на концерт / а я чем-то была развлечена другим;*
- *я вот этой (...) по совету **девчонок** / они же опытные / всё знают // купила вчера капли анти... антибиотические // специальный антибиотик;*
- *а мне господи / банально просто не с кем // не знаю может с **девчонками** куда-нибудь сегодня пойдём // но уже надоело с **девчонками**-то ходить // как бы это всё равно не то.*

мальчик / мальчики

ОРД: 'лицо одного возраста с говорящим или младше', только из уст женщин:

- так вот у нас был **мальчик** / который работал / у нас все работали () в равных;
- там сели поехали / собралась / там как бы / компания там / четыре девчонки из Питера там / **мальчик** потом подсоединился из Москвы // ну вот как бы / тусили // а / прикольно // нормальные ребята такие;
- охренеть какой классный фильм // ну разочек можно посмотреть **мальчикам**

В устах мужчин в подобном значении:

Младшие:

- – а есть полтаха* ? по полтахам разменять ?
– вон спроси / может **парни** разменяют ?
- **пацаны**-то рубят же в этом деле.

Старшие:

- и мы же **мужики** / друг друга понимаем // # бежит к тебе / знает // *П так значит там лучше;
- в общем / супер-техника // там / **мужики** в котельне / да вот / раньше-то был только уголь // а теперь-то с газом / и так не отойдёшь.

Тетя Мотя

В толковых словарях нет

ОРД: ‘незаметная, обыкновенная, может быть, слегка нелепая женщина, обыватель’:

- *и они вот задрали эту цену // а **тётя Мотя** тут / они сейчас критикуют / вон тут в подъезде / вон встал утром / вон президент сказал / чтоб на музее не повышали.*

Аналоги в НКРЯ и в Интернете:

- *Тетя Клава, тетя Маша, тетя Соня (одесск.), дядя Вадя, дядя Вася, дядя Петя, дядя Ваня (mín)*

Специфические черты новых идиом

Семантика

- обобщенный и неопределенный образ:
 - *тётя Мотя с улицы*
 - *тётя Клава из Тьмутаракани*
 - *тетя Маша с соседнего подъезда*
 - *Всегда найдется **какая-нибудь тетя Маша**, которая видит, или спасутся пятнадцать, два, один, которые свидетельствуют*
 - *но быть готовым, что потом за спиной **какая-нибудь тетя Маша** по этому поводу проедется))))*
- типичный, обычный и близкий всем образ:
 - *Обычный дядя Петя.*
 - *Вот перед нами **наша тетя Клава***
 - *Но все равно **наша тетя Маша** с какой-нибудь Залесской птицефабрики своей американской коллеге Мэри может только позавидовать*

Грамматика

- употребление во мн. числе или замена местоимением мн. числа:
 - *Привет вам, тётки Моти и дяди Петя*
 - *А многочисленные дяди Васи ... с пеной у рта требовали крови.*
 - *а тётя Мотя тут / они сейчас критикуют / вон тут в подъезде...*
- употребление с отрицанием:
 - *...причем не тётке Моте, а мне самой...*
 - *чтобы программа могла удостовериться, что именно я даю ей некоторые данные, а не дядя Петя*
 - *а нам зиртек 2 раза в день прописали пить... доктор прописал, а не тетя Маша с соседнего подъезда*

Грамматика

- **встраивание в ряд подобных:**
 - Конечно, это вам не «Цыганочка» в исполнении шестипудовых **тети Маши с тётей Клавой**, но тоже впечатляет;
 - **Подростки на лавочке просят дядю Петю и дядю Серегу** на соседней лавочке за чекушку принести им немного пива. На что вторые радостно соглашаются. Криминала нет.
 - **Идея проекта такова: если у вас, тетя Маша или дядя Петя**, есть некие рационализаторские предложения, которые помогут людям жить, то вы можете прийти и рассказать об этом.

Фонетика

– внутренняя рифмованность, эхо-прием, повтор-отзвучие, мини-стихи (**тетя Мотя, дядя Вадя**), ср.:

- *шашлык-машлык, шиворот-навыворот, ширли-мырли, шуры-муры, сикось-накось, с бухты-баракты и под.*

Графика

- *Жила-была в райцентре с трудным названием Сквородино некая **тётя-мотя** Любовь Бондарчук;*
- *Проходящая мимо **тетя-мотя** в позументах пробурчала: «Безобразие, тут иностранцы ходят, как ему разрешают это делать!»*
- *Ники в блогах Интернета: **ТетяМотя, ТеТя=МоТя.***
- **Tetyamotyа** – сайт интернет-магазина детских игрушек «Тетя Мотя»

БАБЕЦ / БАБЦА

Словари:

- ‘пожилая женщина’ (Словарь воровского жаргона)
- ‘Фам. О привлекательной женщине’ (Большой толковый словарь русского языка / ред. С.А.Кузнецов, СПб., 1998)
- ‘первоклассная девушка, женщина’ (Русский мат. Толковый словарь)

ОРД:

- ну я думаю не раз уже подворачивался / с другой стороны // смотри они сколько уже / лет шесть вместе // нда // хотя тоже там / **бабцу** какую-то / залапал;
- долго ли взирать / такие **бабцы** ?
- ну там такие **бабцы** / конечно призы... / я в шоке // откуда у них грудь ? неужели у них всех силикон ? я не понимаю.

Национальный корпус русского языка

- – ***Бабец-то**, Матрена эта самая, ничего... хе-хе!* [Д. Н. Мамин-Сибиряк. Хлеб (1895)];
- – *Нет. – Ну и **бабец**... я вам скажу! ... Она хочет, кажется, дать представление: публично расцеловать Андрея Михайлыча* [К. М. Станюкович. Жрецы (1897)].
- - ***А хорошая была бабец**, - заметил Сигизмунд Карлович* [М. Суранова. «Монстры в ящике» дурачат зрителей? // Комсомольская правда, 2005.09.12];
- ***Высокая луна-лунища. И бабец стящий*** [В. Маканин. Неадекватен // «Новый Мир», 2002];

Муж. род / Жен. род.?

- **НКРЯ и Интернет:** м.р. > ж.р. в 1,4 раза
- **Носители языка (ментальный лексикон):**
 - М.р. – 70%
 - Ж.р. – 30%
- **Зависимость от контекста (м.р. vs. ж.р.):**
 - *умный/умная* – 91% – 9%;
 - *нормальный/нормальная* – 82% – 17%;
 - *хороший/хорошая* – 81% – 18%

 - *красивый/красивая* – 49% – 51%

учитывая что

МАС: только глагол учесть/учитывать

ОРД и САТ: **учитывая что** в значении союза причины
(= потому что, поскольку, так как):

- на самом деле / я бы очень хотела себя послушать // *П я думаю / что я на работе скину (...) на компьютер // *П а потом послушаю как я говорю // # а скинь у нас тоже / *П интересно знаешь посмеяться // ну () **учитывая / что** там будут записаны / почти сутки;
- но / туда ты идешь спокойно еще / вроде не замечая дороги / но назад // это ужасно // **учитывая что** приходится обычно тащить мне две корзины / так как муж тяжести носить не может после операции (САТ);
- хотя цирк я не очень-то и люблю // но / **учитывая / что** у меня маленький ребенок шести лет естественно мы туда пошли (САТ).

Материалы к словарю контекстных экспресsem русской разговорной речи

- слушай / этот контакт / **глючит** / **глючит** просто;
- ну то есть имеет смысл / да / как бы не... @ просто я ... @ не ... не **гонево** / тебе(:) / больше не кажется / что это всё обман;
- а(:) / это (...) москвичи ко мне приезжали / я почему спрашивала что тебя до... дома бу... будешь ты или нет // *П тут вот / ночью я их **вписывала** // и они накупили / *П какого-то конопляного пива;
- они (э) дела... стараются делать для всех / то есть / невзирая на личности / да ? @ угу @ просто / что видит человека / вот **из него** можно **доить денежку** // вот например торговый представитель / определённый // **с него** можно **тянуть денежку** / определённую // вот он сейчас... / его запустить @ угу @ / **выжать из него все соки** / **получить с него максимум** / (а...) прибыли / и его можно **выкинуть**.

Гужбанская девка

В словарях нет

ОРД:

- *приводит / просто какую-то (...) девку // *П какую-то такую гужбанскую* какую-то пэтэушницу**

НКРЯ:

- *Четырнадцать лет, а вид гужбанский прямо купец приволжский какой-то [Г. Белых, Л. Пантелеев. Республика ШКИД (1926)]*

ИСТОКИ: *гужевой транспорт → гужбан → гужбанский пирог (для извозчиков)*

Интернет

- Оттачиваю **гужбанский лексикон**
- Парень, я ведь с тобой вежливо разговариваю, не стоит и тебе на **гужбанский слэнг** переходить
- Из отвратнейшего Саперави делали глинтвейн, чтобы хоть как-то протолкнуть эту жижу на пути к **гужбанскому кайфу**
- Аа ну ка, сиэтловчане или как вас там, расскажите как там щас вам, как там погода, самый **гужбанский район u-district**, где самые милые парни в США!

Словарь хезитационных конструкций

Причины интереса к объекту исследования:

- Заполненные *паузы хезитации* – типичное свойство спонтанной речи (временной дефицит)
- *Хезитационные конструкции (ХК)* – один из распространенных способов вербального выражения колебаний говорящего
- Вариативность и полифункциональность большинства таких конструкций

Типы хезитационных конструкций

Восходящие к значимым фрагментам текста (десемантизация)

▫ (я) не знаю

- много конечно есть примет // **не знаю** / хм-хм / связанные с погодой // не знаю / что сказать / даже // ну как // то туман стеле... / то самое (ОРД);
- так // ну / что сказать // о рыжем коте / который у хозяев всё воровал / э-э ну **я не знаю** // однажды все-таки он попался (ОРД);

▫ (я) (не) думаю (что)

- и когда рисовал Шишкин / **я думаю что** / это настолько было реально для всех поколений что в общем-то / производит впечатление конечно / очень глубокое (САТ);
- я Сашу / Са... / с... с... с... / Сашу знаю / вот / а **думаю что** там / ты говорит учти / что ты думай / что если он там узнает / что у него объёмный процесс / типа того / думай что он будет / лететь он будет / да / вот он говорю / да (ОРД).

Типы хезитационных конструкций

Восходящие к незначимым фрагментам текста (словам-паразитам)

• *(не) (это) самое*

- *знаешь / это(:) самое / *Н сидят на месте / значит надо требовать / да / жалобы писать (ОРД);*
- *там невозмо... там даже вот // с... самое / пирс // это всё это / пирс ... / *П там всё живое / там даже дотрагиваться // там даже () водоросли дотронешься // там ожоги(?) какие-то есть / сазакки # серьёзно ? @ угу // там настолько море / дикое (ОРД);*
- *да да да да @ надо что-то ... @ чтобы не очень претензионно было @ да @ и в то же время как бы ... *П @ что-то надо @ нарядное # а оно не это самое / не долбанёт ? (ОРД)*

«Кандидаты» в словник этого словаря:

- *Знаешь(...)* – 5204 употребления
- *На самом деле*
- *Как говорится*
- *Короче говоря*
- *Короче* – 197 употреблений
- *Собственно говоря*
- *Собственно* – 21 употребление
- *(Так) скажем*
- *Так сказать*
- *Понимаешь* – 164 употребления
- *Что называется*
- *Значит* – 447 употреблений

и др.

Функции ХК *это самое*

1) **маркер поиска** (наиболее распространенная функция, допускающая дальнейшую систематизацию):

- *нам кстати / *П нам вот **эту самую** (...) стенку / где дверь висит / надо новый(?) рулон там где-то // *П красивые / но дорогие (ОРД);*
- *т... ты сама ... # а сейчас с флэшкой всё таки / ага / сейчас // он ушёл в этот самый @ он ушёл @ в (э-э) ... бельё понёс / я ему сказала / он сейчас бельё отнесёт и придёт (ОРД);*

2) **маркер смены темы** в речевом фрагменте:

- *разрежь половину ягодкой @сейчас @ а то ему ... **это самое** / а половину можно бабушке отнести // @ оно полезное ! (ОРД)*

3) маркер самокоррекции:

- яркая солнечная погода // говорить можно? так был ярк... § **это самое** § был § июльский день / вот / небо было чистым / безоблачным / солнце § светило () (САТ);

4) дискурсивный маркер (может быть стартовым или финальным):

- знаешь / **это(:) самое** / *Н сидят на месте / значит надо требовать / да / жалобы писать (ОРД);
- как ты переводишь ? # нет / я не могу короче (...) **это самое** (ОРД);

5) маркер хезитации без явной мотивации:

- главное / *П (э-э) десятого были / *П и... и **это самое** / следующее слушание будет (...) пятнадцатого / это два дня вот выпадает / по **это самое** / *Н // *П ну (ОРД).

Вариант - *не это самое*

- нормально ? *П ничего там тебе / **не это самое** (...);
- без фанатизма наверное ? здесь не не не / так // я тоже **не это самое.**;
- Носки // **Не это самое** / не прожжжет?
- вот ну так / ха-ха / так это сразу / **не это самое** / [нрзб] не сообразить не / сразу как-то вот / не связать;
- Вот когда нам давали эту квартиру / а нам давали квартиру **не это самое** / нам давал директор

Две идиомы, не требующие обязательного заполнения «пустой» позиции

- *А не пошли бы вы все в это самое?*
[Ю. О. Домбровский. Факультет ненужных вещей, часть 1 (1978)];
- *между прочим у французов тоже простая кухня / просто () это нам теперь здесь её () подают как это самое / как будто пре-парте какое-то (ОРД).*

Словарь редуцированных форм русской речи

- *просто не горит // а чек ты в сумку не **хошь** положить?*
- ***здрасть** / отдел кадров уже закрыт? ага / **сёння** же пятница;*
- *а-а вот мы / живем / на Смолячкова / **ничо** не знаем;*
- *мы как раз здесь побудем / за час уже выболтается всё // а он же там **буит**... это;*
- *м-м хорошо / давай // ты во **скоко** будешь дома?*

Структура словарной статьи

- 1) Исходная словарная лексема:
 - **Быть** - *бум, буюм, бует, буешь, буете*
- 2) Частеречная характеристика
- 3) Специфика функционирования :
 - **в общем** – *вощем;*
 - **в основном** – *в основном;*
 - **в самом деле** – *всамделе;*
- 4) Варианты написания:
 - **Сейчас** - *ща, щас, щаз, счас, счаз, сичас, сечас.*
- 5) Частотность (по НКРЯ или ЗКРЯ)
- 6) Примеры (НКРЯ, ЗКРЯ, публикации РР и под.)

7) Варианты произношения (в ОРД):

/šЫIZ'd'is'at/ <u>13%</u>	/šЫIS'at/ 3%
/šЫId'is'at/ 3%	/šes'at/ 3%
/šЫIZ'is'at/ 3%	/šЫIšat/ 3%
/šЫIis'at/ 3%	/šs'at/ 8%
/šЫIjs'at/ <u>43%</u>	/š:'at/ 3%
/šЫIjs'a/ <u>13%</u>	/šs'a/ 3%

8) Примечания

- Дифференциация значений форм (*жизнь - жысь*)
- Усиление фонетизации письменного представления (*жизнь → жизь → жысь; рубль → рупь; скоко → скока; моно → мона*)
- Этап графического сокращения:
 - *ЧЕСТНОЕ СЛОВО ⇒ ЧЕС. СЛОВО ⇒ ЧЕССЛОВО*
 - *ОЧЕНЬ ⇒ ОЧ. ⇒ ОЧ*
- Графико-орфографическое выравнивание:
 - *слушЬ – как отрезь, услышь.*
 - *здрасТЬ, сла-тЬ Господи*
 - *сѣдня (вместо с{ег}одня)*

В заключение:

- Все 4 типа словарей связаны друг с другом.
- Электронные версии всех предложенных словарей позволят не только дать лексикографическое описание всем единицам каждого словника, но и послушать их в различных контекстах.
- Работа в самом начале, проблем и вопросов пока больше, чем результатов, но по этому пути НАДО ДВИГАТЬСЯ.

**СПАСИБО ЗА
ВНИМАНИЕ**