

ЗВУКОВОЙ КОРПУС РУССКОГО ЯЗЫКА «ОДИН РЕЧЕВОЙ ДЕНЬ»: ПУТИ ПОПОЛНЕНИЯ И ПЕРВЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

*Богданова Н. В. (nvbogdanova_2005@mail.ru),
Асиновский А. С. (a.s.asinovsky@gmail.com),
Маркасова Е. В. (markasovaelena@yandex.ru),
Степанова С. Б. (stsvet_2002@mail.ru),
Супрунова А. В. (nsuprunova@mail.ru),
Шерстинова Т. Ю. (sherstinova@gmail.com)*

Филологический факультет СПбГУ;

Санкт-Петербург, Россия

2010 г.

Звуковой корпус русского языка (ОРД)

- В основе - *принцип невода*;
- Методика 24-часовой записи;
- Цель – фиксация живой речи в ее свободном функционировании, не ограниченном ни лабораторными условиями записи, ни конкретными речевыми заданиями для информантов.
- Естественная речь.

Пополнение корпуса новыми записями

- Выравнивание гендерного состава информантов (11 м + 19 ж → 20 м + 20 ж);
- Отсутствие всякой иной балансировки;
- Пересечение материала двух блоков ЗКРЯ (речь одного информанта в естественных и экспериментальных условиях);
- Запись ОРД от информанта-иностранца, живущего в данное время в России.

Количественные показатели ОРД

- 40 информантов (20 мужчин и 20 женщин);
- 320 часов звучания;
- Звукозаписи переформатированы, убраны длительные (больше 5 мин.) шумовые фрагменты, не содержащие речи;
- Получено 994 файла – коммуникативного эпизода (268 час.);
- Расшифровка и многоуровневое аннотирование в профессиональной программе ELAN - 34 часа звукозаписей;
- 244 075 словоупотреблений на уровне Phrase (реплики говорящих) - 125 эпизодов.

Уровни аннотирования

- Phrase (реплики говорящих),
- Speaker (код говорящего),
- Events (невербальные аудиособытия),
- Voice (качество голоса говорящего),
- PhonetCom (фонетический комментарий),
- PhraseComment (фразовый комментарий),
- Notes (общий комментарий),
- Episode (мини-эпизод речевой коммуникации).

- Синтагматическое и фразовое членение материала;
- Частотный словник всех использованных в речи словоформ;
- Единица описания - не только графическое слово («от пробела до пробела»), но и символы, обозначающие паралингвистические явления (<смех>, <кашель>, <вздых> и под.);
- 32 201 фраза:
 - 23 494 утвердительных;
 - 6 627 вопросительных;
 - 2 080 восклицательных.

- Осуществлена сегментация 14 884 слов для 5 информантов (S01, S19, S24, S35, S37) (147 минут звучания).
- Получена выборочная реальная транскрипция для 1460 словоизменяемых морфем из речи 6 информантов.
- Морфемное аннотирование включает три уровня – орфографическую запись морфемы, её функциональный тип и реальную фонетическую транскрипцию.
- Работа продолжается...

КОНКОРДАНСЫ, ЧАСТОТНИКИ

- Самое частотное слово – *я* (5398 из 244075 словоупотреблений)
- *Ты* – только 15 место по частоте встречаемости (1674)
- *Вот , ну, да* (3, 4, 5 место) – 4800, 4727, 4370
- Первый полноценный глагол в частотном словнике - *знаю* (608; 40 место)
- Ни одного полноценного существительного в списке 150 самых частотных слов не обнаружилось:
 - *Блин* – 305 (85 место)
 - *Типа* – 199 (118,5 место)
 - *Время* – 197 (121 место)
- На 116 месте – *бл...ь* (206)

Блин

- ❑ *нам надо было **блин** ! где Мойка ? как проще вернуться ? здесь ? туда ? или с этой стороны ?*
- ❑ *вот дело в том что знаешь / это в общем короче да / и муж / вот этой подруги взвился говорит / ты вообще там идиотка **блин** / и я это всё видел;*
- ❑ *ну я сдуру **блин** / думал / быстрее пройдёт / заварил колдрекса это ...;*
- ❑ *хачики // то же самое что **блин** неграми называют.*

Типа

1) частица с весьма неопределенным значением:

- да не кто-то над тобой начальник / а **типа** ты сам будешь;
- ну вот он / **типа** въехал в тему / ну а потом уволился на кафедре;

2) частица, вводящая чужую речь (= мол):

- слушай / может быть он и заказ чёткий не сделал ? может он просто спросил / **типа** а можно ?
- и спросил / ну **типа** умеешь вообще кататься или нет ? сидела за рулем ?

3) заполнение паузы хезитации:

- Дора / **типа** привет / у меня к тебе шкурная тема Коля / и какая у тебя шкурная тема ко мне ? ну **типа** / причем разговаривает в каком- то таком стиле // красавица с накладными мозгами;

Типа

4) **сравнительный союз в составе конструкции, характеризующей тот или иной объект, обычно незаконченной – N_1 типа ($N_{1, 2}$):**

а какая-нибудь типа Ирочка Скорбеж это умеет делать ?

5) **часть составного сравнительного союза *типа того что*:**

он такой / я же сказал / надо () головой во все стороны крутить / чтобы шея / что / ну что / чтобы шея сломалась / типа того что не надо бояться;

6) **пояснительный союз (= *то есть*):**

выпили там чего-то виски ... тьпфу водку с колой / типа повеселились;

ты знаешь / кто такие эфсигип(?) ? да да да // типа / тебе пришла эсэмэска / разошли там ... пятьсот(?) эсэмэсок;

Типа

- 7) имя существительное в значении 'образец, модель или разновидность, форма, которым соответствует известная группа предметов, явлений', обычно в сочетании с указательным местоимением *такого* :
- ну такого типа где-то;*
 - что-то типа пробки там;*
- 8) имя существительное в значении 'разряд, категория людей, объединенных какими-л. характерными чертами (социальными, профессиональными, нравственными)':
- дальше мы смотрим / (а-а) ... по сути дела сангвиник / да / можем дать характеристику / как психологи / сангвинического типа темперамента / да / что (...) чем он характеризуется.*

Десятка самых частотных слов в ОРД в сравнении с другими корпусами/частотниками

Слово	Ранг (ОРД)	Ранг (Шаров)	Ранг (НКРЯ)
я	1	4	5
не	2	3	3
вот	3	9	9
ну	4	10	10
да	5	8	8
а	6	6	6
и	7	1	1
что	8	5	4
в	9	2	2
это	10	7	7

Опыт лексикографического описания бытовой спонтанной звучащей речи

- *Словарь русской бытовой разговорной речи*

ДЕВКА:

- *приводит / просто какую-то (...) **девку** // какую-то такую гужбанскую какую-то пэтэушницу;*
- *ну вот / и **девка** такая а-а-а / типа дерьмо.*

ДЕВЧОНКИ/ДЕВОЧКИ:

- *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 // ага !*

ТЕТЯ

- Тенденция к употреблению с именем:
 - *я вспомнила эту историю как Наташа / Галя ! а Юрка / **тётя** Галя ! / Наташа / Галя ! а Юра / **тётя** Галя ! и Юра только после сорока лет стал меня Галей называть // вырос / потому что / знаешь / мужик здоровый / **тётя** Галя сказать.*
- **Тетя Мотя:**
 - *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 музеи не повышали.*

Примеры новых слов

- *вот я (:)* весь в таком **раздумкивании** // что ж мне делать ?
- *стоял стоял / стоял стоял* // нету **лоховозов** / а у нас проездной //
- *тогда когда я вот (э) с Любовь Ильиничной-то* **тёрки** были у нас / с... со второго (...) класса //
- *я не могу сюда его поставить / у меня (...) нету* (...) **кряка** //

Словарь контекстных экспрессем русской разговорной речи

- она привозила с собой этого немца / который () вот **явный немец** / а косил под русского **Ваньку** // ему первые два дня было запрещено говорить // потом (...) он... **под дурачка косил** ? он был **первым парнем на деревне** // там к нему (...) / к нему все девки (...) пока не проговорился к нему все девки приставали // пока не проговорился / да / по-немецки / пока молчал // угу // это я на колонке угостил (э-э) немца // салом // выпьет водки // он долго не хотел / а потом как **подсел на это сало** / короче // и давай **один бутерброд за другим наворачивать** // просто за друг... (...) перед другими людьми уже неудобно было // так **попёрло его**

Анализ фразеологических оборотов

- **Книжные ФО** в спонтанной речи почти не употребляются (2 %):
 - *завершающий аккорд; притча во языцех; пожинать плоды; камень преткновения;*
- **Разговорных и просторечных ФО** много (68 %):
 - *челюсть отвисла; крышняк едет; вынь да положь; поджилки трясутся; дать на лапу; галопом по Европам; ни в одном глазу;*
- **Совсем новые ФО:**
 - *красавица с накладными мозгами;*
 - *обыденное корыто с сухарями;*
 - *на халяву и(:) хлорка творог /*
 - *молчат все // как рыба в пироге*

Анализ дискурсивной лексики

- «уход» вводных слов **несомненно, бесспорно, вне всякого сомнения** (оценка говорящим степени достоверности сообщаемого);
- **безусловно** – скорее, маркер манипулятивного речевого поведения личности, имеющей установку доминирования;
- «маркеров искренности» (**говоря по совести, чего скрывать, положила руку на сердце** и под.) не встретилось;
- **честно говоря** и **(если) честно** - контактоустанавливающие элементы (семантика солидаризации или интимизации общения или заполнитель паузы hesitation):
 - **я честно говоря // во-первых ну я могу конечно показать вам эти анкеты;**
 - **у меня вот честно говоря / желание / только <...> / пойти (...) влить / в себя что-нибудь ... / чтоб повеселело;**

Корпус редуцированных форм (РФ) 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?*

Варианты произношения *шестьдесят*
(37 употреблений)

/šyɪz'd'is'at/	13%	/šyɪs'at/	3%
/šyɪd'is'at/	3%	/šes'at/	3%
/šyɪz'is'at/	3%	/šyɪšat/	3%
/šyɪis'at/	3%	/šs'at/	8%
/šyɪjs'at/	43%	/š:'at/	3%
/šyɪjs'a/	13%	/šs'a/	3%

Анализ речевых сбоев, различных хезитационных явлений

- *на... н... вот наше вот это вот (э-э) вот это
вот / вот тут / тут сложнее гораздо / да //
потому что / значит / я вот вот (э-э) вот эти /
ну в принципе / значит / ну / п... по моим /
понятиям значит / я же н-не отличу так
скажем / таджика от узбека что называется
да ? и тут / важно...*



По мере сбора и обработки материалы ОРД передаются в устный подкорпус Национального корпуса русского языка («Из материалов корпуса “Один речевой день”, подготовленного группой А.С.Асиновского»)

РАБОТЫ ЕЩЕ МНОГО...

**СПАСИБО ЗА
ВНИМАНИЕ**