

# О КОРПУСЕ ТЕКСТОВ ЖИВОЙ РЕЧИ: НОВЫЕ ПОСТУПЛЕНИЯ И ПЕРВЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

*Н. В. Богданова*

*[nvbogdanova\\_2005@mail.ru](mailto:nvbogdanova_2005@mail.ru)*

*Филологический факультет СПбГУ;*

*Санкт-Петербург, Россия*

*2010 г.*

# Звуковой корпус русского языка (ЗКРЯ)

- В основе - *принцип ковчега* («каждой твари по паре»);
- Сбалансированный состав информантов;
- Сбалансированная лингвистическая программа:
  - чтение;
  - пересказ;
  - описание изображения;
  - свободный рассказ на заданную тему.
- Экспериментальная речь.

# Пополнение корпуса новыми текстами

- **MED**
  - Речь медицинских работников
  - 32 диктора, 210 текстов
  - 6 час. звучания
- **JUR**
  - Речь юристов
  - 40 дикторов, 322 текста
  - 16 час. звучания
- **RKI**
  - Речь преподавателей РКИ
  - 20 дикторов, 70 текстов
  - 3,5 час. звучания
- **STUD**
  - Речь студентов
  - 5 разных блоков
  - 7 час. звучания
- **COMP**
  - Речь «компьютерщиков»
  - 14 дикторов, 56 текстов
  - Более 2 час. звучания
- **RIA**
  - Интерферирующая русская речь американцев
  - 68 дикторов, 204 текста
  - 10 час. звучания
- **RIK**
  - Интерферирующая русская речь китайцев
  - 2 диктора, 4 текста
  - 10 мин.т звучания

# Пересечение материала двух блоков (речь одного информанта)

- Полная лингвистическая программа ЗКРЯ, записанная 4-хкратно, с интервалом в 1 мес., в течение 3-х лет (всего 28 текстов)
- Один речевой день (ОРД) этого информанта, его общение с коммуникантами:
  - мама,
  - друг,
  - подруга,
  - директор фирмы,
  - зам. директора,
  - кассир в авиакассе,
  - бабушка (по телефону).

# Результаты анализа существующего материала

Внутриязыковая интерференция профессиональной и бытовой речи:

## MED

- **медицинская терминология (1,0%)**

*высокий УРК - атеросклероз сосудов головного мозга, сердечная недостаточность, энергозатраты, генотип;*

*низкий УРК - боль, больной, медсестра*

# Переключение на медицинскую тему (только высокий УРК):

- *если бы не Макдоналдс я бы там наверное не выжила потому что японцы едят совершенно изумительную пищу // э-э опять-таки медицина // ну ха-ха вам придется на это сделать сноску // э-э они где-то лет двадцать назад стали вымирать от инсультов потому что у них произошла американизация и появилась американская пицца сеть Макдоналдс все вот эти Кэрролс и так далее // японский организм оказался к этому непривычным и они стали / э-э стали болеть атеросклерозом причем вот именно японская нация стала болеть атеросклерозом сосудов головного мозга // поэтому количество инсультов превзошло превзошло все возможные ожидания они увеличились на порядок по отношению к первоначальному / после этого они вернулись к традиционной пище // традиционную пищу есть могут только японцы [Инф. 20, жен., высокий УРК].*

# JUR

- **Элементы официально-деловой речи («стилевая разноголосица»):**
  - *отдых в выходные связан / непосредственно с нахождением дома;*
  - *основополагающим моментом моих выходных это должно быть конечно хорошая погода;*
  - *сам процесс отдыха основная составляющая которого / заключается в том что кататься на лыжах;*
  - *лёгкий просмотр телевизора с приёмом завтрака;*
  - *в индивидуальном порядке в номере занимаюсь иногда / лечебной такой гимнастикой.*

# Два перцептивных эксперимента (чтение и аудирование)

- Эксперты-филологи (студенты и преподаватели, 20 чел.) довольно хорошо чувствуют специфику бытовой речи юристов и с высокой степенью вероятности относят предложенные им фрагменты монологов именно к речи юристов:
  - чтение – 78-89 %
  - аудирование - 73-69 %



# СОМР

## Лексические единицы профессиональной сферы компьютерщиков:

- где же т... / здес-сь чёрная лестница / думал пёс-с // в какой-то момент // ударился о-о / **рандомный** шкаф;
- он ударился здоровым боком о стекло / надеясь / что за ним находится (...) **альтернативная** дверь;
- цвет неба / одинаков // и не меняется // ни со временем ни (...) ладно не буду говорить **по координате** (...);
- из шкафа вывалилась банка **номер два** / с рыжей гадостью;
- в процессе борьбы коту удалось дотянуться до // некоего места / располагающегося выше шнурков / господина **номер три**;
- вбегают / **дополнительные** люди // первый **дополнительный** человек // не преуспевает / второй **дополнительный** человек // э-э / пользуется чем-то из шкафа.

# COMP

## Маркеры «неопределенности»:

- *пёс понял / по общему окружению // что он попал в **некую** лечебницу / возможно для собак;*
- *сюжет / двоеточие // **неназываемая** женщина / которая потом оказалась Зиной / заманивает / бедного больного пса / больной на один бок ] э-э в квартиру;*
- *Зина // просит помощи / выбегает // э-э // **неизвестное** лицо мужско... мужское лицо ;*
- *э-э // значит // **неопределённая** [нрзб] женщина // э-э // заманивает // пса / в // **некий** / длинный узкий тёмный коридор / оказавшийся больничным;*
- *облака окружены **некоторыми** змейками / они круглые / похожи на острова;*
- *ближе к полудню / эм-м / начи... начинали появляться / облака к которым казалось / прибежали **какие-то** (...) по некоторым из них иногда ка... / казалось что пробегают **какие-то** ме... мелкие серебристые змейки.*

# Чтение

## Приметы спонтанности:

- *Берестов выехал прогуляться верхом / на всякой случай взяв с собою пары <...> т<...>ри борзых;*
- *огражденный своим чином токмо <...> токмо от побоев; чьи это <...> чьи это дрожки?*
- *и рассказал все / что случи<...> рассказал все / что случилось; но он наехал на Берестова вовсе неужи<...> вовсе неожиданно.*

# Чтение

## Метакоммуникативные конструкции:

- *сии столь оклеветáнные см... а! оклеветáнные ... нет... оклеветáнные с... ладно оклеветáнные смотрители ... (инф.-м., студент-нефилолог, несюж. текст);*
- *Муромский не мог отказа...ться, ибо чувствовал себя обязанным, и таким образом Берестов возвратился домою за... со славою / затравив зайца и ведя своего противника раненым и почти ... военнопленным // мать моя (инф.-ж., студентка-филолог, сюж. текст);*
- *будет она стираемая и глаженная с двумя пристежными воротничками / бред (инф.-м., студент-филолог, экстраверт, сюж. текст).*

# Пересказ

## Различные коммуникативные стратегии:

- **Я-наблюдатель (63%):**
  - *был тёплый июльский день // солнце поднималось ] на восходе // солнце было не ] не яркое / не огненно-рыжее / оно было светлое и лучезарное // лучи его поднимаясь над землей / пронизывали (...) светлые / белые облака // облака <вдох> стоящие над землей (...) вверху на небе / лежали белыми белыми яркими ] снопами*
- **Я-читатель (32 %):**
  - *ну / начинается с того что [...] с описания ] (э-э) женщины и пса // входят они / оказываются в какой-то комнате / и реакция пса / вот на ту ситуацию которую он видит // там / какая-то лестница темная*
- **Я-исследователь (ок. 5 %):**
  - *в общем-то ] в данном тексте говорится о ] (а-а) ] как мы понимаем / о женщине / которая привела собаку / в некую комнату // при этом ] (а-а) параллельно видя ] (э-э) ситуацию глазами ] (а-а) стороннего наблюдателя и глазами собаки*

# Классификация эзоединиц пересказа:

- **Грамматические трансформеры** (*пощелкала пальцами – пощелкав пальцами*)
- **семантические трансформеры** (*сладкий – сладковатый*)
- **транспозиты** (*кричал – крики*)
- **синонимы** (*синева – лазурь*)
- **антонимы** (*сырость – сухость*)
- **гипонимы/гиперонимы** (*пес - животное, комната – помещение*)
- **конверсивы** (*она <...> наполнила комнату запахом – комната наполнилась запахом*)

- **свернутые/развернутые номинации** (*личность мужского пола – мужчина, облака с белыми краями – края [облаков] были белые*)
- **ассоциативные замены разного типа** (*касторка – валерьянка, небосклон – небоскребу, осколки – обломки, стекло + дверь – окно*)
- **вводящие элементы текста** (*повествуется, описывается, я прочитал отрывок из Тургенева*)
- **слова, выражающие эмоции** (*бедолага пес*)
- **слова, выражающие модальную оценку** (*скажем так, пожалуй*)
- **метакоммуникативные единицы** (*да сложновато так, насколько я помню*)
- **дискурсивные единицы** (*всё – в конце монолога*)

# Монолог-описание

## Структурные элементы

- **Собственно описание (29,0%)**

- *кусты какие-то прямо за деревьями / одно дерево какое-то кривое / с дуплом*
- *красивый летний парк с деревьями и зеленой травой / на переднем плане видна тропинка*

- **Описание с суждениями, сомнениями, домыслами, догадками говорящего (28,0%)**

- *на шестой [картинке] какой-то дядька пожарник судя по всему снимает ее с лестницей*
- *слева / небольшая лужа или пруд*

- **Оценка (7,0%)**

- *но он [Шишкин] / у нас пейзажист был / так что в общем что с него взять / рисует себе деревья / кустики там всякую траву / в общем ну странный человек конечно / лучше б людей рисовал*
- *как-то он переборщил*



- **Ощущения, впечатления (8,0%)**

- *на пикничок правда наверное неплохо / но мне кажется там холодновато*
- *достаточно приятно было бы там находиться*

- **Переключение (11,0%)**

- *вот всегда еще интересовал / вопрос / э водятся в лесу полевые мыши*
- *интересно почему еще изображают / всегда / э белку в дупле дуба / и на еще кто любит желуди а кабаны / ну если есть дубы / почему нет кабанов / ну странно*

- **Комментарии по поводу изображения (7,0%)**

- *оригинал я не видела это репродукция на открытке / хранится в Государственном Русском музее*
- *честно говоря конечно солгу что не знаю их названия / да / это дубы / там взади написано про них*

- **Дискурсивные элементы (4,0%)**

- *значит на этой картинке*
- *вот чего тут*

- **Метакоммуникация (2,0%)**

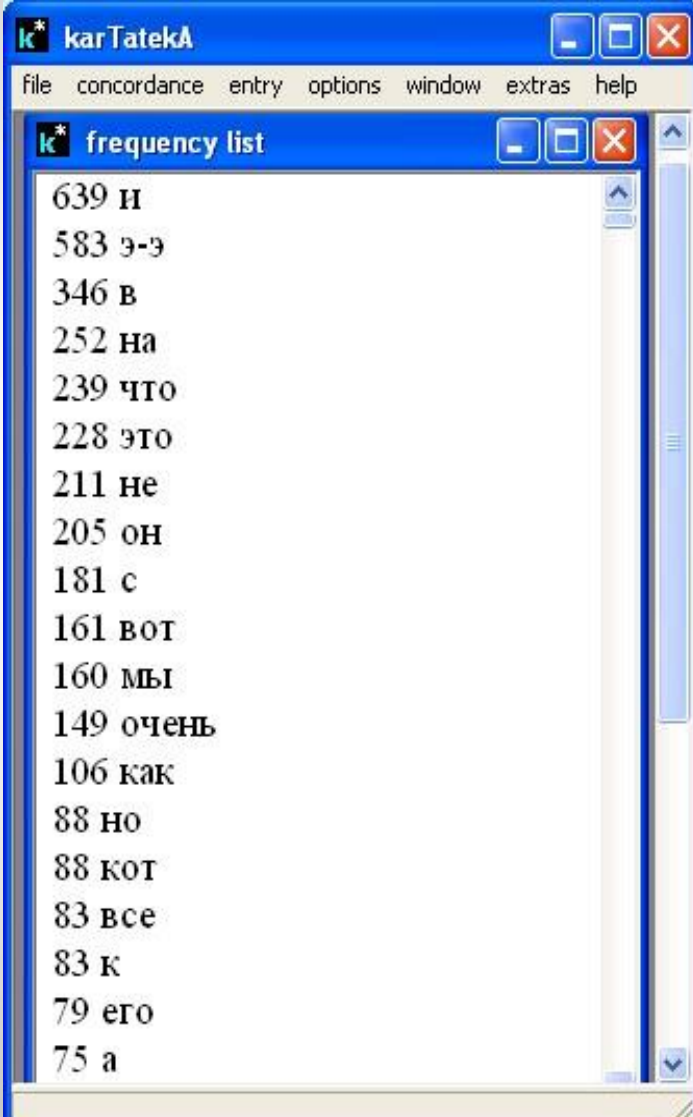
- *я много не умею говорить / могу красиво молчать*
- *ну конечно / так как / описать картины наверное / описывается (нрзб.) сначала передний план / за... / задний план / э-э-э передний фон*

- **Метакоммуникация дискурсивного характера (4,0%)**

- *слушай я не знаю что говорить просто*
- *вот я думаю все*
- *что-то я так больше ничего не могу придумать еще*

# Конкордансы, частотники

- «Словарный запас» носителей языка (MED):
  - высокий УРК – 4 тыс. слов
  - средний УРК – 3 тыс.
  - низкий УРК – 2 тыс.



The screenshot shows a software window titled 'karTatekA' with a menu bar containing 'file', 'concordance', 'entry', 'options', 'window', 'extras', and 'help'. Below the menu bar is a sub-window titled 'frequency list' which displays a list of Russian words and their corresponding frequencies. The list is as follows:

639	и
583	э-э
346	в
252	на
239	что
228	это
211	не
205	он
181	с
161	вог
160	мы
149	очень
106	как
88	но
88	кот
83	все
83	к
79	его
75	а

# Монологи-описания студентов:

- Среди 5 самых частых слов – частица НЕ;
- В описаниях несюжетного изображения, наряду с НЕ, ВОТ, НУ, ЭТО, ТАМ, - НАВЕРНОЕ;
- Единственное полноценное слово в «верхушке» словника - наречие ОЧЕНЬ, использованное девушками-нефилологами.

- **Свободные рассказы информантов-юристов**

- *э-э* (271)
- *а-а* (122)
- *вот* (492)
- *ну* (256)
- первый (наиболее частотный) глагол - глагол-связка *быть* (256)

# Публикация материалов ЗКРЯ

- *Русская спонтанная речь. Свободные монологи-рассказы на заданную тему. Тексты. Лексические материалы / Сост. В.В.Куканова / Отв. ред. и автор предисловия Н.В.Богданова. СПб., 2008.*
- *Русская спонтанная речь. Монологи-репродуктивы. Тексты. Лексические материалы / Сост. В.В.Куканова / Отв. ред. и автор предисловия Н.В.Богданова. СПб., 2010.*
- *Русская спонтанная речь. Монологи-описания. Тексты. Лексические материалы / Сост. В.В.Куканова / Отв. ред. и автор предисловия Н.В.Богданова. СПб., 2011 (в печати).*

Еще работать и работать...

Спасибо за  
внимание