О.Ю. Крючкова,
В.Е. Гольдин
(Саратовский государственный университет)

КОРПУС РУССКОЙ ДИАЛЕКТНОЙ РЕЧИ:

КОНЦЕПЦИЯ И ПАРАМЕТРЫ ОЦЕНКИ

«Корпус проектировался и создавался с ориентацией на… рядовых пользователей…, большинство из которых никогда в жизни не видело ни одного диалектного текста. В то время задачей было сделать своего рода «научную игрушку», которая наглядно демонстрировала бы разнообразие русского языка в его региональных вариантах»

(Е.В. Рахилина 2009)

«…пришло понимание, что этот проект полезно было бы перестроить так, чтобы он служил самим диалектологам – и как удобно организованный ресурс для учебного процесса и как инструмент для исследовательской деятельности»

(Е.В. Рахилина 2009)

Вопросы, требующие решения:

- **Принципы отбора материала для диалектного корпуса**
- **Членение речевого континуума в диалектном корпусе**
- **Параметры выдачи текстовых фрагментов**
- **Формы представления диалектных текстов**
- **Правила аннотирования текстовой базы диалектного корпуса**
- **Параметры метаразметки диалектных текстов**
- **Характер нелингвистической информации в диалектном корпусе**
- **Типы пользовательских запросов**

Проблема репрезентативности корпуса

- Стратегия ДК НКРЯ:
  простое увеличение количества любых фрагментов диалектной речи в составе корпуса.
  *(*2008 г. - *144 тыс. словоупотр.*, 2011 г. – *194 тыс. словоупотр.)*

- Общий принцип репрезентативности текстовых корпусов :
  пропорциональность языкового материала реальной коммуникации.

- Репрезентативность диалектного корпуса: **система корпусов отдельных говоров, представляющих разные диалектные типы.**

Принципы формирования текстовой базы диалектного корпуса

***Текстовая база должна отражать*:**

- **важнейшие типы диалектной речи**
- **различные формы речи**
- **разнообразную тематику сельского общения**
- **социальную дифференциацию носителей говора**
- **типы говорящих (языковые личности носителей диалекта)**
- **типы прецедентных текстов**

6

Параметр «Место записи» в ДК НКРЯ

Архангельск	– 92; 104810	Брянск	– 1; 886
Курск	– 34; 25764	Тверь	– 2; 590
Волгоград	– 10; 15919	Тула	– 1; 421
Вологда	– 17; 13101	Кострома	– 1; 277
Саратов	– 14; 9126	Воронеж	– 1; 272
Карелия	– 1; 6840	Тамбов	– 1; 229
Рязань	– 4; 5884	Псков	– 1; 189
Забайкалье	– 4; 2944	Киров	– 1; 166
Калуга	– 3; 2501	Иваново	– 1; 165
Смоленск	– 2; 1538		
Новгород	– 2; 1390		
Владимир	– 4; 1271		

Единица хранения в ДК НКРЯ (фрагменты метаразметки текстов)

Название: Порча

Тема текста: обычаи

Предложений: 11

Словоформ: 497

Название: День победы

Тема текста: война

Предложений :115

Словоформ: 677

Параметры выдачи текстовых фрагментов в ДК НКРЯ

Минимальный контекст

Купол поставили, сделали, поставили, купол, и она опеть ушла в землю вся. Утром выстали, а всё россияло уж. Вот такая церковь была.

Расширенный контекст

А потом это ночью церковь вся уйдёт опеть ниже рук. И они опеть снова кладут-кладут, опеть ниже рук, опеть, снова-снова-снова. и дошли до купола. Купол поставили, сделали, поставили, купол, и она опеть ушла в землю вся. Утром выстали, а всё россияло уж. Вот такая церковь была. [Сначала уходила в землю, а потом поднялась?] Она вся полностью, целиком уже вся пришла, вся вышла.

Параметры выдачи текстовых фрагментов в СарДК

Минимальный контекст - абзац

бывало у нас их сколько было/ моленнов-то…/ Коробово/ Доронино/ Кладовкино/ эта/ большая моленна/ Тюревска моленна/ шесть моленнов было в селе-ти// и оне были вот рядом/ вот наша-то была вон тут за горой/ она сейчас на столовой стоит/ это вот тоже их… сломали/ поразвозили/ ходили все рядышком/ все/ и были учёны/ а мы сейчас чё? самоучки// кто нас учит? нет+никто// в церкви-ти вот там учут/ ведь оне учут чего/ оне не правы// у них крещенье какое? еретическо// а у нас християнско//

мы купали в купели или в реке/ или в озере/ или где-то/ в пруду/ вон в речке/ как Исуса Христа/ Иоанн Креститель крестил/ а он [поп в церкви]?/ миропомазание/ то… то брызжет/ то ложки каке-то суёт// разве это закон? разве это божье писание?/ там этого совсем и нет// называется крещение/ а у нас по-ихнему-то крещению/ называется/ если только вы приняли крещение/ святое/ оно святое называется/ то вы после этого крещения называетесь християне/ а оне после свово этого причастия/ у них причастие оно называется/ а не крещения/ называются еретиками// а с еретиком/ тоже есть+написано/ с еретиком/ вот не пить/ и не есть/ и не вкупе богу молиться/ это в моленной вкупе// вот// и ласково слово глаголить нельзя//

Формы представления диалектных текстов в ДК НКРЯ

- Широкие юбки, рубашки свойские — лён сеяли. Хлопот за ним очень много было.
- А что Вы смотрите по телевизору? ну // какие картины такие / оне глядят и я сижу / гляжу //
- Вот это была / Никифириха-то, сама-то / Татьяна Михайловна была, маленька, вот эконька была только, шибко маленькая.
- А тогда ж батюшка ж тогда ж ведь священник был Рождество сказывал батюшка.

«Новый формат подачи» в ДК НКРЯ

- **Текст-1** – «текст в том виде, в каком он предоставлен изначально», «первоначальная транскрипция».
- **Текст-2** - «облегченная» транскрипция, основанная на принципах русской графики, с сохранением ударения»; «должен… максимально учитывать особенности транскрипции».
  «Этот текст удобнее для цитирования и для чтения широким пользователем, в том числе нелингвистом».
- **Текст-3** – «литературный подстрочник», содержащий грамматический и фонетический аналог диалектного текста».
  Напр.: У галодной куме адно на уме → У голодной кумы одно на уме,
  продавают → продают.

Аннотирование текстовой базы корпуса Морфологическая разметка в ДК НКРЯ

- Дифференциальные пометы:

✓ dialflex – диалектное окончание,
✓ dialmorph – диалектная морфология,
✓ dialstem – диалектная основа,
✓ gend – диалектные особенности в сфере рода

- Характеристика диалектных форм:

✓ другая флексия: соседы
✓ нестандартная флексия: девчонкими
✓ несловарн.: …плачь была такая дак

Способы соотнесения диалектных (и иных нелитературных) форм с литературными

- Характер лемматизации словоформ (ДК НКРЯ):

оне{они=S,мн,од=им=dialmorph,type}

- Дополнительные опции (ДК НКРЯ):

кнопка lexref, кнопка sem

- Зона литературного соответствия (СарДК):

оне{оне(они)=S,мн,од=им=*}

Тематическая и жанровая разметка в ДК НКРЯ

- быт
- война
- жизнь
- люди
- обычаи
- праздники
- природа
- работа
- религия
- семья
- фольклор

Жанрово-тематическая разметка в СарДК

- #14@1 бывало у нас их сколько было/ моленнов-то…/ Коробово/ Доронино/ Кладовкино/ эта/ большая моленна/ Тюревска моленна/ шесть моленнов было в селе-ти// и оне были вот рядом/ вот наша-то была вон тут за горой/ она сейчас на столовой стоит/ это вот тоже их… сломали/ поразвозили/ #14 #12 ходили все рядышком/ все/ и были учёны/ @1 @2 а мы сейчас чё? самоучки// кто нас учит? нет+никто// в церкви-ти вот там учут/ ведь оне учут чего/ оне не правы// у них крещенье какое? еретическо// а у нас християнско//

- мы купали в купели или в реке/ или в озере/ или где-то/ в пруду/ вон в речке/ как Исуса Христа/ Иоанн Креститель крестил/ а он [поп в церкви]?/ миропомазание/ то… то брызжет/ то ложки каке-то суёт// разве это закон? разве это божье писание?/ там этого совсем и нет// называется крещение/ а у нас по-ихнему-то крещению/ называется/ если только вы приняли крещение/ святое/ оно святое называется/ то вы после этого крещения называетесь християне/ а оне после своово этого причастия/ у них причастие оно называется/ а не крещения/ называются еретиками// @2 @8 а с еретиком/ тоже есть+написано/ с еретиком/ вот не пить/ и не есть/ и не вкупе богу молиться/ это в моленной вкупе// вот// и ласково слово глаголить нельзя// @8#12

Перечень тем для разметки в СарДК

1. История жизни
2. Семья
3. Дом и быт
4. Животноводство, птицеводство, пчеловодство
5. Полеводство, садоводство, огородничество, лесоводство
6. Охота, рыболовство, собирательство
7. Односельчане
8. Городская жизнь
9. Трудовая деятельность, служба в армии
10. Обряды, обычаи, приметы, праздники
11. Здоровье и лечение
12. Религия
13. Природа
14. История села
15. Развлечения, досуг
16. Государство, власть, политика, общественно-экономическая жизнь, СМИ
17. Мораль, поведение
18. Учеба
19. Колдовство, порча. гадание
20. Мифические существа
21. Чужие
22. Вредные привички
23. Происшествия
24. Общая оценка жизни
25. Одежда
26. Пища

Перечень жанров для разметки в СарДК

1. **Рассказ-повествование**
2. **Рассуждение**
3. **Описание**
4. **Сказка**
5. **Песня**
6. **Частушка**
7. **Пословицы, поговорки**
8. **Цитаты, аллюзии**
9. **Произведения религиозного содержания**
10. **Фатическое общение**

Структура метаразметки в ДК НКРЯ

- Название Порча
- Дата создания 1999
- Диалектолог (запись) В.Е. Гольдин, О.Ю. Крючкова
- Регион записи Саратов
- Место записи село Белогорное Вольского района Саратовской области
- Сфера функционирования бытовая
- Тема текста обычаи
- Стиль диалектный
- Предложений 11
- Словоформ 497
- Возраст аудитории н-возраст
- Уровень образования аудитории н-уровень
- Размер аудитории личная
- Источник Саратовский государственный университет
- Подкорпус ПК диалектных текстов

- Автор Вяч. Вс. Иванов
- Пол автора муж
- Дата рождения автора 1929
- Название Задачи и перспективы наук о человеке
- Дата создания 2007
- Сфера функционирования учебно-научная
- Тема текста наука и технологии
- Тип текста лекция
- Стиль нейтральный
- Предложений 621
- Словоформ 10636
- Возраст аудитории н-возраст
- Уровень образования аудитории высокий
- Размер аудитории большая
- Тип носителя книга
- Источник http://www.polit.ru/lectures/2007/09/17/ivanov.html
- Дата публикации 2007
- Подкорпус ПК письменных текстов

Параметры метаразметки текста в СарДК

1. Сведения об информантах
2. Сведения о времени и месте записи
3. Конкретная ситуация общения
4. Адресаты речи
5. Упоминаемые лица
6. Время описываемых событий
7. Перечень тем (цифровое обозначение)
8. Перечень жанров (цифровое обозначение)

Фонетический комментарий в метаразметке текста («Новый стандарт подачи» в ДК НКРЯ)

- особенности реализации «старого ятя» и <а> после мягких согласных под ударением;
- тип безударного вокализма;
- характер произношения <г> (взрывной или фрикативный);
- наличие цоканья

Нелингвистическая информация в ДК НКРЯ (Фрагмент метаразметки)

- Дата создания 1999
- Диалектолог (запись) В.Е. Гольдин, О.Ю. Крючкова
- Регион записи Саратов
- Место записи село Белогорное Вольского района Саратовской области
- Тема текста обычаи
- Предложений 11
- Словоформ 497

Биография диалектоносителя

Межевая Мария Матвеевна, 1918 г. рожд. По документам год рождения - 1917; так записали, спутав со старшей сестрой, умершей в детстве.

Девичья фамилия – Поликарпова.

Окончила 3 класса, после чего около пятнадцати лет работала в школе уборщицей; затем семь лет – поваром; потом – на сенокосе, возила на быках солому; в войну – на рытье окопов; после войны – скотницей.

Имеет 3 медали.

Воспитывала оставшуюся после смерти сестры племянницу, которая живёт в этом селе.

Было четверо братьев. Два брата погибли на фронте, похоронены в братской могиле; один пришел с фронта без ноги. Никого из братьев и сестер в живых нет.

Была два раза замужем. Первый муж был из старообрядческой общины. Вышла замуж за старообрядца, потому что в селе мало было «мирских» (православных) – всего двадцать семей. Первый муж погиб на фронте. От второго мужа родила восьмимесячного ребёнка. Преждевременные роды произошли из-за того, что пришлось поднимать тяжёлые мешки.

Муж погиб, попав под лошадь.

Сын живёт в этом же селе.

Мария Матвеевна живёт с внучкой, которая работает поваром в столовой.

С каждым текстом соотнесены модули справочного характера:
- метаописание текста,
- модуль с биографией диалектоносителя,
- модуль с иллюстративными материалами.

25

С каждым текстом соотнесены модули справочного характера:
метаописание текста, - модуль с биографией диалектоносителя,
модуль с иллюстративными материалами.

С каждым текстом соотнесены модули справочного характера:

метаописание текста,

- модуль с биографией диалектоносителя,

модуль с иллюстративными материалами.

С каждым текстом соотнесены модули справочного характера:
метаописание текста, - модуль с биографией диалектоносителя,
модуль с иллюстративными материалами.

С каждым текстом соотнесены модули справочного характера:
 метаописание текста,
- модуль с биографией диалектоносителя,
 модуль с иллюстративными материалами.

29

Система пользовательских запросов ДК

- Должна удовлетворять:
- потребностям уровневых описаний диалектной речи,
- потребностям в области коммуникативного, лингвокогнитивного и лингвокультурологического изучения диалектного общения.

Спасибо

за внимание!