ЯЗЫК
И ИССЛЕДОВАТЕЛЬСКИЕ ПРОЕКТЫ БУДУЩЕГО

О.Ю. Крючкова, В.Е. Гольдин, А.П. Сдобнова

ЭЛЕКТРОННЫЙ ДИАЛЕКТНЫЙ КОРПУС КАК НОВЫЙ ИСТОЧНИК ИЗУЧЕНИЯ РУССКИХ НАРОДНЫХ ГОВОРОВ[14]

Саратовский государственный университет

В последнее десятилетие существенно меняется круг и характер научных лингвистических источников. В лингвистических исследованиях все более широкое применение находят электронные ресурсы различных типов: электронные словари, базы данных, текстовые корпусы. Однако таких источников еще явно недостаточно, и их создание является насущной задачей прикладной лингвистики.

Наличие электронных машинообрабатываемых лингвистических баз данных не только значительно ускоряет и оптимизирует трудоемкий процесс сбора языкового материала, но и, как верно отмечено акад. А.П. Ершовым на первой Всесоюзной конференции по проблемам создания машинного фонда русского языка (1983 г.), ведет к смене научной парадигмы в лингвистике. По мнению А.П. Ершова, автоматизированная система, которая “должна быть адекватной и равнообъемной живому организму языка, но в то же время… анатомически отпрепарированной, разъятой, доступной для наблюдения, изучения и изменения” [1: 12], позволила бы русистике перейти “от наблюдательного периода к измерительному” [1: 9].

Все это имеет особую значимость для диалектологии, ограниченной до последнего времени и кругом источников, и, соответственно, доступной для изучения проблемной областью. Основным источником диалектологии до недавнего времени оставался вторичный материал, т.е. материал, собранный по специальным вопросникам и представленный в картотеках, словарях, атласах. Специфика материала определяла особый, периферийный, статус диалектологии в составе русистики, значительно сужала круг решаемых диалектологией задач, ограничивая диалектологическую проблематику в основном сферой системно-структурного своеобразия русских

[14] Работа выполнена при поддержке Российского фонда фундаментальных исследований (РФФИ), проект № 06-06-80428-а.

народных говоров. Отсутствие доступного широкому кругу русистов репрезентативного первичного диалектного материала – текстов – не позволяет составить целостного представления о диалектной коммуникации, построить симметричное литературному языку описание диалектной речи.

Наиболее надежной формой хранения диалектных текстов и оптимальной источниковой базой является программно обеспеченный электронный текстовый корпус. Электронная форма представления диалектных текстов повышает сохранность этого уникального материала, создает возможность для более свободного доступа лингвистов различной специализации к первичному диалектному материалу, позволяющему наблюдать реальные отношения между единицами в потоке диалектной речи. Программное обеспечение корпуса позволяет каждому исследователю при минимальных затратах усилий самостоятельно создавать на основе корпуса полные базы данных в соответствии со своими исследовательскими задачами, классифицировать материал на основании отдельных параметров и их комплексов.

Мысль о необходимости создания машинного фонда диалектных текстов была аргументирована еще в 1980-х г. А.С. Гердом и В.Е. Гольдиным [1: 72; 2]. В 1980-2000-е гг. в разных лингвистических центрах страны осуществлялось активное накопление и обработка текстового диалектного материала. Именно тексты стали осознаваться как основной информационный ресурс диалектологии. Вместе с этим развивалось понимание того, что русские народные говоры целесообразно изучать не только в лингвогеографическом аспекте, но и как самостоятельные коммуникативные системы, представляющие особый тип речевой культуры (см., напр., работы В.Е. Гольдина, томских диалектологов), стало оформляться особое направление – “коммуникативная диалектология” [3]. Наметилось смещение интереса диалектологов с междиалектных различий на то, что характеризует общую специфику “диалектов как диалектов” [см., напр.: 4; 5].

Однако доступ к текстовому диалектному материалу пока еще ограничен местом его локализации, а в случае его хранения в твердой форме оказывается затрудненным полное использование эвристического потенциала этого ценнейшего источника. Изменить это положение может создание и открытие для научной общественности электронных корпусов диалектных текстов. В настоящее время ведется работа в этом направлении: создаются диалектный подкорпус в составе Национального корпуса русского языка (ИРЯ РАН) и Саратовский диалектный корпус (СГУ). Названные проекты различаются своими целями, принципами организации баз данных, методами обработки (разметки) текстов, включаемых в состав соответствующих электронных корпусов, подробнее об этом см. в [6; 7].

В Саратовском диалектном корпусе текстов (СДК) каждый отдельный говор образует самостоятельный подкорпус и представлен значительным по объему и разнообразным текстовым материалом, соотносимым с мно-

гообразной нелингвистической информацией (фотографии, видеоиллюстрации, схемы, карты, сведения исторического, социо-культурного характера, демографические, этнографические, географические данные). Единицей хранения является “запись” – расшифровка магнитофонной фиксации непрерывного фрагмента общения, приводимая в символьной записи, близкой к орфографической. Регулярные фонетические явления (например, характер безударного вокализма) в символьных расшифровках не отображаются, отражение получают лишь лексикализованные фонетические особенности (типа ***топерь, кстить, Рожество, Паска***). Однако в корпусе учитывается значимость фонетической информации; она может быть получена из первоисточника – включаемых в корпус звуковых файлов, от которых возможен переход к символьным записям и наоборот. Такое представление фонетической информации в корпусе делает ее наиболее объективной и пригодной для использования в диалектологических исследованиях.

Электронная база корпуса позволяет осуществлять запросы, касающиеся грамматических, лексических, словообразовательных языковых явлений, извлекать многообразную нелингвистическую информацию и соотносить ее с лингвистической. С этой целью проводится пословная лексико-морфологическая разметка, а также многоаспектная метаразметка расшифрованных записей диалектной речи. Грамматический поиск (поиск с целью получения грамматической информации) в текстовом корпусе может вестись по словоформе, по маске, по начальной форме, по любому из классификационных и словоизменительных признаков слова.

Параметрами метаразметки в СДК являются сведения об информантах, о времени, месте записи, о конкретной ситуации общения, об адресатах речи, упоминаемых лицах и их отношении к информанту, о времени событий в повествовании. Проводится также разбиение диалектного текста на тематические и жанровые фрагменты. Поисковые запросы могут осуществляться в соответствии с каждым из параметров метаразметки.

Возможности, предоставляемые исследователю текстовым электронным корпусом, сегодня еще не могут быть осознаны в полной мере. Однако совершенно ясно, что изучение диалектной речи на материале текстовых корпусов дает уникальную возможность получить принципиально новую информацию о диалектной коммуникации: о строе диалектной речи, о соотношении различных жанров в составе диалектной коммуникации, об особых приемах раскрытия темы, о когнитивных особенностях диалектного общения, об особой картине мира, реализуемой в этом общении и т.д.

Так, например, на основе электронной текстовой базы данных в работах саратовских лингвистов ставились и решались следующие вопросы:

– об общих принципах диалектной коммуникации, характеризующих диалект как особый тип речи [3; 5];

– о путях развития говоров – через их внутреннюю стратификацию или через варьирование структурных элементов единой микросистемы (о соотношении дискретности и непрерывности диалектного существования) [8];

– о жанровой типологии деревенского общения [9];

– о системе и функциональном распределении синтаксических средств связи – союзов и предлогов – в одном говоре [10; 11; 12];

– об использовании метатекстовых средств в диалектной речи [13];

– о функционировании в говоре диминутивных производных [14];

– о соотношении типов знания, выражаемых в диалектной речи [15];

– о характере речевой рефлексии и реализуемой в речи культурной идентичности носителей диалекта [16; 17];

– о месте и роли различных предметных областей в когнитивно-дискурсивном поле носителей диалекта [18].

Текстовый диалектный корпус позволяет не только ставить и решать новые для диалектологии задачи, но и дает возможность по-новому рассмотреть традиционные диалектологические проблемы. Новизна подхода к традиционным проблемам определяется полнотой выборки, получаемой на основе электронного текстового корпуса, отражением в корпусе реальных отношений между единицами потока диалектной речи, возможностью адекватного выявления ее количественных характеристик. Количественное измерение диалектной речи является важным условием создания объективного представления о диалектной коммуникации, инструментом ее моделирования. Значимость учета количественных характеристик диалектного общения обусловлена такими специфическими его чертами, как активное варьирование языковых средств, параллельное использование собственно диалектных, просторечных и литературных единиц, внутренней дифференциацией диалекта (например, различием между “традиционным” и «передовым» его слоями).

Текстовый диалектный корпус позволяет осуществить переход от констатирующего описания диалекта к функциональному. Так, например, текстовый анализ морфологических и лексических диалектизмов [19; 20] дает возможность определить их коммуникативную значимость, соотношение собственно диалектных образований с их литературными и просторечными вариантами, оценить их регулярность, выявить факторы, стимулирующие или, напротив, ограничивающие употребление диалектных форм. Наблюдения за коммуникативной спецификой диалектизмов открывают новую исследовательскую перспективу, указывают на целесообразность изучения функциональной дифференциации диалектизмов (от общедиалектных до идиолектных).

Таким образом, электронные корпусы диалектных текстов являются принципиально новым источником, способствующим приобщению диалектологи к современной научной лингвистической парадигме и созданию

такой русистики, в которой изучение основных языковых страт (литературной и диалектной речи) находилось бы в необходимой и корректной корреляции.

Библиографический список

1. Машинный фонд русского языка: идеи и суждения. – М., 1986.

2. Гольдин, В.Е. К проекту текстового диалектологического подфонда Машинного фонда русского языка / В.Е. Гольдин // Доклады Третьей Всероссийской конференции по созданию Машинного фонда русского языка. – М., 1990.

3. Гольдин, В.Е. Теоретические проблемы коммуникативной диалектологии / В.Е. Гольдин. – Саратов, 1997.

4. Гольдин, В.Е. Доминанты традиционной сельской культуры речевого общения / В.Е. Гольдин // Аванесовский сборник. – М., 2002.

5. Гольдин, В.Е. Машиннообрабатываемые корпусы диалектных текстов и проблема типологии русской речи / В.Е. Гольдин // Русистика сегодня. – 1995. – №3.

6. Крючкова, О.Ю. Электронный корпус русской диалектной речи и принципы его разметки / О.Ю. Крючкова // Известия Саратовского университета. – Новая серия. Филология. Журналистика. – 2007. – Т. 7. – Вып. 1.

7. Летучий, А.Б. Корпус диалектных текстов: задачи и проблемы / А.Б. Летучий // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. – М., 2005.

8. Гольдин, В.Е. Диалектологический текстовый машинный фонд говора и исследование диалектных изменений / В.Е. Гольдин // Современные процессы в русских народных говорах. – Саратов, 1991.

9. Дубосарь, Н.П. Жанры диалектного общения: дипл. соч. / Н.П. Дубосарь. – Саратов, 1988.

10. Крючкова, О.Ю. Изучение служебных частей речи на основе машинного текстового диалектного фонда / О.Ю. Крючкова, Т.Н. Медведева, С.П. Цепова // Русистика сегодня. – 1995. – №4.

11. Крючкова, О.Ю. Сочинительные союзы в мегорском говоре Вологодской области и методика их изучения / О.Ю. Крючкова, С.П. Цепова // Филология: научн. сб. памяти А.М. Богомолова. – Саратов, 1996.

12. Медведева, Т.Н. Система предлогов, предложно-падежных сочетаний в диалектной и литературно-разговорной речи (функциональный аспект): дисс. … канд. филол. наук / Т.Н. Медведева. – Саратов, 1999.

13. Крючкова, О.Ю. Метаязыковая функция в диалектной речи / О.Ю. Крючкова // Языковые средства в системе, тексте и дискурсе. – Самара, 2002. – Ч. 1.

14. Крючкова О.Ю. Вторичные диминутивы одного говора в историческом контексте / О.Ю. Крючкова // Предложение и слово. – Саратов. 2002.

15. Гольдин, В.Е. Текст и знание / В.Е. Гольдин, О.Ю. Крючкова // Исследования и материалы по русской диалектологии. – М., 2007.

16. Крючкова, О.Ю. Оценки речи как выражение специфики народного речевого сознания (в печати) / О.Ю. Крючкова.

17. Балаян, Э.В. Межконфессиональные отношения в селе Белогорное Вольского района Саратовской области по рассказам местных жителей (в печати) / Э.В. Балаян.

18. Гольдин, В.Е. Тематическая разметка и тематический анализ диалектного текстового корпуса / В.Е. Гольдин, О.Ю. Крючкова // Языковая личность – текст – дискурс: теоретические и прикладные аспекты исследования. – Самара, 2006. – Ч. 1.

19. Крючкова, О.Ю. Морфологическое своеобразие среднерусской диалектной речи по данным диалектных текстов (в печати) / О.Ю. Крючкова, В.Е. Гольдин.

20. Крючкова, О.Ю. Лексическое своеобразие среднерусской диалектной речи по данным диалектного текстового корпуса (в печати) / О.Ю. Крючкова, В.Е. Гольдин, А.П. Сдобнова.

В.Е. Гольдин, О.Ю. Крючкова, А.П. Сдобнова

АССОЦИАТИВНЫЙ СЛОВАРЬ СЕЛЬСКИХ ШКОЛЬНИКОВ[15]

Саратовский государственный университет

Словесные ассоциации сельских школьников – почти не изучены. Из современных ассоциативных словарей только электронный “Ассоциативный словарь школьников Саратова и Саратовской области” (АСШС) включает соответствующие материалы, но до настоящего времени и они не подвергались специальному изучению. Остается неизвестным, в какой мере специфично речевое сознание сельских школьников, как оно соотносится с речевым сознанием взрослых сельских жителей, с русской диалектной речью, с просторечием, с молодёжным и другими жаргонами, с устной и письменной литературной речью, как совершается формальное и семантическое развитие речи сельских школьников, как и в какой мере происходит освоение учащимися общих ассоциативных норм русского языка. Поэтому представляется полезным создание ассоциативного словаря сельских школьников или даже серии таких словарей в качестве научного источника и особого направления изучения русской народно-речевой культуры, а также культуры русской речи в целом. Однако для решения данной задачи необходим ряд предварительных исследований. Результаты одного из них, выполненного на материале АСШС, излагаются ниже.

К середине 2007 г. в базе данных АСШС, кроме 7587 анкет учащихся школ Саратова и 4507 анкет детей, обучающихся в школах районных цен-

[15] Работа выполнена при поддержке РГНФ (грант № 05-04-04142а) и РФФИ (грант № 06-06-80428-а).