

В.И. Шестопалова, Т.И. Петрова, М.А. Болгов

РЕГИОНАЛЬНЫЙ ВАРИАНТ ЖИВОЙ РУССКОЙ РЕЧИ КАК ОБЪЕКТ КОРПУСНОЙ ЛИНГВИСТИКИ¹

Введение

К числу признаков, существенных для любого из компьютерных корпусов, относится репрезентативность (представительность), определяемая не только объемом корпуса, но и сбалансированностью состава текстов: в корпус должны быть включены «по возможности все типы письменных и устных текстов, существующих на данном языке в данном временном отрезке», причем все эти тексты должны входить в корпус «по возможности пропорционально их доле в языке соответствующего периода»². Определенную трудность представляет решение проблемы жанровой сбалансированности корпуса – проблемы, которая «не только не решена, но даже и не поставлена в сколько-нибудь явном виде»³. Еще более сложной является проблема социологической сбалансированности, наиболее значимой для подкорпуса устной речи. По замечанию Е.А. Гришиной, «вопрос о какой бы то ни было социологической сбалансированности будущего устного подкорпуса НКРЯ в данный момент не ставится по экстралингвистическим причинам»⁴ (имеются в виду финансовые и органи-

¹ Исследование осуществляется при поддержке РГНФ: проект «Создание электронной базы данных «Живая речь дальневосточников» (№ 08-04-12103в).

² *Плунгян В.А., Резникова Т.И., Сичинава Д.В.* Национальный корпус русского языка: общая характеристика // НТИ. Сер. 2. 2005. № 3. С. 9–13.

³ *Вербицкая Л.А., Казанский Н.Н., Касевич В.Б.* Некоторые проблемы создания национального корпуса русского языка // НТИ. Сер. 2. 2003. № 6. С. 2–8.

⁴ *Гришина Е.А.* О принципах размещения устных текстов в Национальном корпусе русского языка // НТИ. Сер. 2. 2005. № 3. С. 31–38.

зационные трудности, связанные с получением материала, сбалансированного по социологическим характеристикам информантов. «Заведомое отсутствие социологической сбалансированности», как считает Е.А. Гришина, может быть компенсировано «достаточным объемом и относительно полным набором жанров». Признавая правомерность такой позиции для начального этапа формирования корпуса живой русской речи, все же считаем необходимой организацию речевого материала на основе именно целенаправленного (по социологическим характеристикам) отбора информантов.

Требование представительности предполагает включение в Национальный корпус и различных территориальных вариантов устной речи¹. В таком случае правомерно говорить о территориальной – или региональной – сбалансированности корпуса. В НКРЯ включены тексты живой речи не только москвичей и петербуржцев, но и жителей Саратова, Ульяновска, Таганрога, Екатеринбурга, Норильска, Воронежа, Новосибирска и мн. др. Дополнить этот ряд может и речевой материал, относящийся к наиболее удаленному российскому региону – Дальнему Востоку. Целью нашего исследования стало рассмотрение особенностей подготовки речевого материала для его включения в базу данных «Живая речь дальневосточников». Прежде всего, представляется уместным уточнить содержание понятия «живая речь», что, на наш взгляд, имеет непосредственное отношение к проблеме отбора речевого материала и его размещения в корпусе.

«Живая речь» и проблемы ее изучения

Традиционно «живой» называют звучащую естественную речь в ситуациях неподготовленного неофициального общения. Исследования живой речи органично связаны с темой «язык

¹ Плунгян В.А., Резникова Т.И., Сичинава Д.В. Национальный корпус русского языка: общая характеристика...

города», которая, по замечанию Е.А. Земской, в наше время переживает второе рождение, включившись в новый научный контекст¹. Важным является то обстоятельство, что понятие «живая речь города» имеет более широкое содержание по сравнению с близкими ему понятиями «разговорная речь» и «устная речь». Основу «языка города» составляет литературная разговорная речь. Именно такой речевой материал стал первым опытом научной публикации текстов живой речи. В полном же объеме язык города неоднороден, в связи с чем актуальным стало изучение специфики его элементов (прежде всего, городского просторечия и жаргонов), а также особенностей их взаимодействия. «Языковая жизнь» города представлена преимущественно устной речью, но не ограничивается только ею. Такое понимание интересующего нас объекта лежит в основе теоретической базы уральских исследователей², которые в понятие «языкового облика города» включают и так называемый «языковой быт»: номинации городских объектов, вывески, плакаты и т.п. Специфику именно городской речи возможно выявить только на основе сопоставления с противоположным феноменом – языком современного села. Таким образом, массив живой речи, открытой для лингвиста, шире, чем только язык города: это также и речь жителей поселков и деревень. Особую значимость в связи с этим приобретает «пространственный аспект существования языка». Как отмечает Е.В. Красильникова, при таком подходе к изучению языка города «обсуждаются две теоретические возможности: 1) литературный язык един, противопоставлены литературные и нелитературные явления, территориальные различия принадлежат к диалектной области и сферам ее влияния; 2) литературный язык существует в

¹ Земская Е.А. Разновидности городской устной речи и задачи их изучения // Земская Е. А. Язык как деятельность: Морфема. Слово. Речь. М., 2004. С. 247–289.

² *Языковой облик уральского города*: Сб. науч. тр. Свердловск, 1990.

территориальных вариантах, в определенном взаимодействии с наддиалектными и диалектными формами»¹. Изучение живой речи в этом случае базируется на «функциональном принципе», в соответствии с которым в поле зрения исследователя оказываются любые проявления естественной речи, в том числе «нелитературного характера или с нелитературными привнесениями просторечия, жаргонов, диалектных окрасок». Причем такие включения, по мнению уральских лингвистов, дают наиболее интересный материал для определения специфики городской речи в конкретном регионе². Чрезвычайно важным для исследования «локального варьирования» языка становится создание подкорпусов живой русской речи различного «географического происхождения» – своеобразных коллекций, включающих региональные варианты национального русского языка. В данной статье обсуждается вопрос о возможности создания одного из таких корпусов – живой русской речи Дальнего Востока.

К вопросу о создании подкорпуса живой речи дальневосточников

Русский язык на Дальнем Востоке функционирует в особых условиях вследствие ряда причин географического, исторического и социального характера. Детальное многоаспектное исследование особенностей современного русского языка в этом регионе возможно на основе материала, представительность которого во многом определяется отбором текстов.

¹ *Красильникова Е.В.* Язык города как лингвистическая проблема // Живая речь уральского города. Свердловск, 1988. С. 5–18.

² *Живая речь уральского города*: Сб. науч. тр. (Предисловие). Свердловск, 1988. С. 3–4.

Об отборе текстов

Представляется возможным использовать алгоритм, позволяющий отбирать речевой материал с учетом специфики данного региона.

Определение административно-территориальных единиц для сбора материала. Мы исходим из того, что исследование регионального варианта национального языка должно быть комплексным – т.е. необходим охват речевого материала на всем пространстве конкретной территории региона. В частности, коллекцию текстов живой речи Приморского края планируем организовать с выделением следующих структурных элементов: «Речь большого города» – «Речь малого города» – «Речь поселка» – «Речь села». Основу материала, безусловно, составит речь жителей Владивостока («Речь большого города»), требующая более детального представления. Выбор других административно-территориальных единиц для целенаправленной записи речи осуществляем с учетом историко-социологического и экономического факторов.

Определение круга информантов. Значимой для региональной коллекции живой речи считаем социологическую сбалансированность. Круг информантов определяем с учетом следующих факторов: возрастного, гендерного, социально-профессионального, социально-территориального. По существующей в отечественной коллоквиалистике традиции¹, выделяем три возрастных группы информантов: «старшее поколение» (старше 50 лет), «среднее поколение» (от 30 до 50 лет), «младшее поколение» (от 17 до 30 лет). Правомерным считаем выделение еще одной возрастной группы: «дети» (от 2 до 10 лет) и «подростки» (от 10 до 17 лет). Создание региональной коллекции детских текстов станет актуальным, в частности, для изучения «этнокультурных

¹ *Русская разговорная речь.* Тексты. М., 1978.

особенностей речевого развития русских детей из разных регионов России» – проблемы, недостаточно исследованной в возрастной социолингвистике¹. Если говорить о социально-профессиональном аспекте отбора, то в этом случае учитываем хозяйственную специализацию региона: среди наших информантов, например, должны быть моряки, рыбаки, таежные промысловики, участники автомобильного бизнеса и т.д.

Определение круга ситуаций речевого общения. Любая коммуникативная ситуация характеризуется признаками, сгруппированными по четырем параметрам: место, время, партнеры коммуникации, тема. Определяющим назван параметр «место», в связи с чем выделяются две основных сферы коммуникации – «дом» / «вне дома»² (этот признак «чрезвычайно важен и для метаразметки устных текстов»³). Именно по локальному признаку выделяем два типа ситуаций: 1) семейное общение и 2) общение «вне дома». В зависимости от той или иной разновидности городского пространства, «вне дома» (вслед за М.В. Китайгородской и Н.Н. Розановой) выделяем ситуации общения: на улице; в объектах, связанных с целеориентированной коммуникацией (магазин, рынок, поликлиника, транспорт и пр.); в объектах, связанных с фатической коммуникацией (музей, ресторан, театр, дискотека и пр.). Кроме того, предполагаем и запись речи в ситуациях интракоммуникации (разговоры с животными, растениями, предметами и т.п.).

¹ *Доброва Г.Р.* Возрастная социолингвистика: к постановке вопроса // Проблемы онтолингвистики – 2007. СПб., 2007. С. 57–64.

² *Китайгородская М.В., Розанова Н.Н.* Современное городское общение: типы коммуникативных ситуаций и их жанровая реализация (на примере Москвы) // Современный русский язык: Социальная и функциональная дифференциация. М., 2003. С. 103–126.

³ *Гришина Е.А.* О принципах размещения устных текстов ...

О размещении материала

Устная форма, естественная для живой речи, требует особых правил размещения в корпусе. «Корпус устного употребления языка должен состоять по меньшей мере из двух компонентов – аудиозаписей речи и соответствующих этим записям транскриптов», – считают А.А. Кибрик и В.И. Подлеская, которыми принята попытка разработать транскрипцию русской речи с учетом «относительно простых и ядерных компонентов просодии»¹. Однако, исходя из практической целесообразности, создатели НКРЯ приняли решение «пренебречь на данном этапе «звучащим» аспектом устной речи»². При размещении материала в региональной базе данных, безусловно, необходимо учитывать имеющийся опыт и следовать правилам, принятым в Национальном корпусе. Целесообразным представляется использование всех трех форм подачи материала, которые названы в статье Е.А. Гришиной. Однако различной будет степень их обязательности применительно к конкретному материалу: 1) стенограмма устного текста (на основе стандартной русской орфографии) обязательна для каждого устного текста; 2) транскрипция (целого текста или его фрагмента) может сопровождать отдельные тексты (например, содержащие наиболее интересный материал для фонетистов); 3) аудиозапись (исходная форма) может быть приложением к стенограмме в том случае, если звучащая речь качественно зафиксирована на носителе.

Создание базы данных, отражающей реально существующую живую речь конкретного региона, является далеко не простой задачей, решение которой требует больших усилий и времени, для того чтобы простую электронную коллекцию текстов

¹ *Кибрик А.А., Подлеская В.И.* К созданию корпусов устной русской речи: принципы транскрибирования // НТИ. Сер. 2. 2003. № 10. С. 5–12.

² *Гришина Е.А.* О принципах размещения устных текстов...

преобразовать в полноценный корпус – представительный по объему, сбалансированный по различным параметрам, содержащий разнообразную разметку. Вследствие специфики живой речи (конситуативная обусловленность, использование невербальных средств и т.д.) первостепенную значимость приобретает метаразметка, основные правила которой сформулированы создателями НКРЯ¹. Наиболее существенным для нашего исследования является положение о необходимости социолингвистической разметки; в создаваемой нами базе она может быть соотнесена с предложенным алгоритмом отбора материала.

¹ *Савчук С.О.* Основные принципы метаразметки текстов в Национальном корпусе русского языка // НТИ. Сер. 2. 2005. № 3. С. 13–22; *Гришина Е.А.* О принципах размещения устных текстов...