

М. Д. Воейкова

Проблемы использования подкорпуса устной разговорной речи *(на примере анализа русских диминутивов)*

I. ВВЕДЕНИЕ

В первой части статьи рассматриваются
проблемы использования подкорпуса устной
разговорной речи на примере анализа
русских диминутивов. В работе дан
описание подкорпуса устной разговорной
речи, собранного в рамках проекта
«Корпус устной разговорной речи»
Института лингвистических исследований
РАН. В статье описаны основные
проблемы использования подкорпуса
устной разговорной речи для лингвистических
исследований. В частности, описаны
проблемы, связанные с наличием в
подкорпусе устной разговорной речи
разных диалектных вариантов
слов, а также с наличием в нем
некорректных с точки зрения
грамматики и лексикой оборотов.

первой части статьи рассматриваются
некоторые проблемы расшифровки
записей устной разговорной речи
на основе трехлетнего опыта сбора
данных устной речи в ИЛИ

РАН (Санкт-Петербург), вторая—иллюстративная—часть посвящена анализу функционирования диминутивов от имен существительных на материале этого подкорпуса. Как показывает практика, нкря в течение ряда лет используется как уникальная по объему база данных, снабженная мощной системой поиска, позволяющая за короткое время собрать статистически обрабатываемый путем сплошной выборки материал. Создание и пополнение подкорпуса устной разговорной речи (а также диалектного подкорпуса) относится к одной из приоритетных задач, в первую очередь потому, что русская разговорная речь обладает рядом существенных отличий от речи письменной.

Благодаря исследованиям группы Е. А. Земской, работам М. Я. Гловинской, Е. В. Красильниковой, О. А. Лаптевой, О. Б. Сиротиной, основные особенности русской разговорной речи внимательно анализировались с конца 70-х гг. XX столетия. Детальное описание этих особенностей позволило включать русский язык, наряду с немецким и английским, в сопоставительные работы по синтаксису устной речи [Miller, Weinert 1998]. Разговорная разновидность русского языка попадает в поле внимания лингвистов, в первую очередь, из-за особенностей ее «фрагментарного синтаксиса» — не встречающихся в письменной речи конструкций с эллипсисом опорных именных компонентов типа *С зелеными балконами/ это ваш, Молоко разносит/ не приходила еще? У окна лежала/ капризная очень...* [Земская 1973: 227] или конструкций с «конденсацией относительного местоимения»: *У тебя нет чем писать? Папе надо кресло сидеть, Возьми на чем сидеть, Где здесь касса платить?* (примеры из [Земская и др. 1981]). Обращает на себя внимание то обстоятельство, что сфера имени оказывается более «уязвимой» для эллипсиса, чем сфера глагола: название предмета с определенными свойствами становится излишним, ненужным и поэтому выводится за рамки эксплицитной информации. Особенности употребления имен существительных в разговорной речи связаны с намерениями говорящего. Нейтрально имена существительные употребляются в номинативной функции, т.е. называют участников ситуации или объекты. Особая интенциональность в употреблении имен связана с подчеркиванием их морфологических признаков, или деталей номинации. Когда в фокусе внимания оказывается морфологическая информация, говорящий прибегает к запретным или стоящим на грани нормы формам имен существительных. Примерами этого могут служить окказиональные или инновационные формы, привлекающие внимание собеседника и придающие высказыванию неформальный или иронический оттенок, ср. примеры из записей спонтанной речи или из работ: *Я остаюсь в *казинах Лихтенштейна, А меня не приковывали к *пианине, и теперь я об этом жалею* (примеры из [Гловинская 2007: 180]), *Здесь *дошираки продаются, Хочу стены расписать. *Матами?* [Горбова и др. 2006: 26], *Пригласим*

хороших *девушков (Записи 2004). Свобода в обращении с формами существительных создает непринужденность, ощущение опыта в обращении с объектами номинации.

Намеренное подчеркивание деталей самой номинации связано с тем, что говорящий применяет не общепринятое название лица или предмета, а окказиональное или экспрессивно-оценочное, например: *И не хочу там мешать *тупизмом своим* (пример из [Горбова и др. 2006]), *Вот тебе *крестик, я сидел и ждал твоего звонка, Я не понимаю, с чем связаны *замороки по оформлению с той машиной, которая есть* (Записи 2006). В экспрессивном употреблении говорящий или изобретает новое слово по существующей модели, или употребляет уменьшительно-ласкательный суффикс, или даже, как в последнем примере, образует от ложного диминутива несуществующий симплекс (*замороки от заморочки). Такие употребления наводят на мысль о том, что использование диминутивов в речи не всегда связано с особенностями их семантики, но зависит от контекста и ситуации общения. Употребляемые окказиональные образования показывают, что в подходящей ситуации могут нарушаться как семантические, так и морфонологические запреты на образование диминутивов или антидиминутивов (ложных симплексов). Такие примеры свидетельствуют о прагматической направленности в употреблении диминутивов (см. ниже, подпункт 5).

Использование слов с уменьшительно-ласкательными суффиксами относится к идиоэтническим феноменам: в каждом языке сфера их употребления определяется индивидуально. Функционирование таких номинаций в разговорном русском языке может быть адекватно описано только с помощью материала НКРЯ. В данной работе формулируются задачи поиска и анализа диминутивов в устной речи и характеризуются первые результаты такого описания.

2. Особенности кодировки данных разговорной речи

Наши материалы готовились в соответствии с общим форматом представления данных устной разговорной речи в НКРЯ [Гришина

2005]. Собирая данные¹, мы пользовались внутренней инструкцией НКРЯ, однако были вынуждены расширить ее, так как в процессе работы расшифровщики записей постоянно сталкивались с такими особенностями звучащей речи, передача которых на письме нуждается в обсуждении. Большинство решений было предложено членом нашей группы сотрудником ИЛИ РАН И. А. Магиным, имеющим опыт полевой работы с диалектным материалом, и обсуждалось далее с другими участниками, которые были заняты практической расшифровкой записей. Так, стяженные формы типа *щас* вместо *сейчас* сохранялись только в тех случаях, когда они приобретают особое значение, например: —*Так я и побежала! Щас!* (в значении нежелания совершить действие). В других случаях форма *сейчас* давалась в обычной орфографии. Принятое решение продиктовано тем, что сохранение единообразия написания желательно даже при передаче неформальной устной речи. Его возможный минус состоит в том, что, заменяя стяженные формы на полные, мы не сможем проследить развития потенциальных новых значений в будущем. Именно поэтому нам представляется необходимым создание архива звуковых записей текстов разговорной речи, собираемых для НКРЯ, которые могли бы в дальнейшем использоваться для специального анализа. Такой архив, не обязательно открытый для всех пользователей, позволил бы и в других случаях уточнить принятые в НКРЯ параметры расшифровки, например, восстановить фонетические паузы, которые отсутствуют в современном формате. Архивы звучащей речи существуют в некоторых международных корпусах (например, в системе обмена данными по детской речи CHILDES [MacWhinney 2000, Sokolov, Snow 1995], <http://childes.psy.stu.edu>). Удобное для использования оформление такого архива предполагает специальную обработку данных, после которой части расшифрованного текста связываются с вырезанным куском зву-

¹ Сбор данных разговорной речи поддерживался программой Президиума РАН "Русский язык, литература и фольклор в информационном обществе ИМЛИ ЗОИФ. Дальнейшая обработка осуществляется в рамках проекта «Семантическая и формальная избыточность текста в современном русском литературном языке» по Программе фундаментальных исследований ОИФН РАН СЛЯ «Текст во взаимодействии с социокультурной средой: уровни историко-литературной и лингвистической интерпретации».

кового файла таким образом, что при помощи простой команды можно немедленно послушать интересующую реплику или более крупный фрагмент. Эта работа, даже при имеющемся объеме данных устной речи, потребует больших финансовых вложений и займет много времени. Не случайно подкорпус устной речи в любом корпусе является одной из наиболее трудоемких частей. По оценкам создателя Международной системы обмена данными детской речи CHILDES Б. МакУинни, только расшифровка магнитофонной записи (без разметки и дополнительной обработки) требует в 12 раз больше времени, чем длится сам разговор. Однако, думая о будущем нкря, необходимо предусмотреть и дополнительные, в том числе и дорогостоящие, пути его обогащения.

Далее приводятся те особенности звуковой речи, которые пока не отражаются в текстах нкря, но могут быть учтены в дальнейшем или восстановлены с помощью архива звуковых данных.

2.1. Комментарии к репликам включают те невербальные средства коммуникации, которые полноправно участвуют в диалоге, но не имеют кодифицированной формы, например: Информант@ *знаешь, есть такие вещи #вертит пальцем у виска#*. Эти жесты существенно дополняют разговор и могут даже изменить смысл звучащей реплики: *разводит руками, показывает пальцем* и т.д.

При первичной обработке данных такие комментарии записываются так, как показано в примере (между знаками *диеза*—«решетками» #). Так же оформляются важные для понимания коммуникации действия говорящего, например: *#выбегает#* или жесты, имеющие звуковую сторону: *смеётся, щёлкает по кадыку, фыркает, хлопает в ладоши, щёлкает пальцами, чихает, стучит по столу*.

Заполненные паузы *хезитации* и *междометия* должны быть унифицированы. В противном случае при любом статистическом анализе одно и то же *междометие*, записанное разными способами (например, *gm* и *xm*) будет означать новое слово. До сих пор мы применяли следующий примерный список пауз *хезитации* и *междометий*: *мм, ага, угу, э* (для обозначения согласия и заполнения своей реплики (*turn-taking*)), *ц-ц* (*цоканье* для выражения скепсиса, в значении 'всё не так-то просто, есть некоторые затруднения'), *хм* (*скептическое*), *бэ* (*знак отвращения, имитирующий тошноту*), *ф-ф*

(выдыхание, обозначает затруднение). Специально отмечаются те случаи, когда собеседник начал говорить, но его перебили или он задумался и не закончил реплику, например: *Информант@ Это же +*. Этот случай отличается от «приёма умолчания» или произнесения слова одними губами (см. далее).

2.2. Необходимо учитывать также некоторые особенности произношения при имитации чужой речи. Если в речи встречаются иноязычные вставки, то их, насколько возможно, следует записывать на языке оригинала. Выбор между русской записью и иностранной можно сделать, следуя намерению говорящего, т.е. произносит ли он «*битлс*» или на английский манер «*beatles*». Во всех случаях, когда произнесение слова искажается, в расшифровке вслед за произносимой формой под звездочкой указывается стандартное написание слова. Так, если в речи встречается имитация иностранного акцента, её записывают так: *Информант@ Так сказать, FIN: тоска, но не тот тоска, который терефо{*дерево}, а тот, который крусть{*грусть}:FIN*. Согласно нашей конвенции, эта запись означает, что говорящий имитирует «финский» акцент (оглушение звонких согласных). Двойная запись слов, содержащих имитацию акцента, позволяет при необходимости включить их в морфологическую разметку. При этом также разграничиваются иноязычные вставки и пародирование иноязычного акцента. До сих пор в практических целях мы применяли следующий список акцентов: ANG (английский), ASI (центрально азиатский, азербайджанский, таджикский, узбекский, турецкий, восточный вообще), DEU (немецкий), EST (эстонский), FIN (финский), FRA (французский), HEB (еврейский), GEO (грузинский), KAV (кавказский, если непонятно, что изображает информант—ASI или GEO), RUS (нарочито русский, т.е. содержащий имитацию диалектной речи, окания и т.п.), UKR (украинский), INO (иностранный вообще). В наших записях случаи имитации иностранного акцента были очень редки, и в этих ситуациях ни одного диминутива не встретилось. Этих употреблений явно недостаточно, чтобы сделать вывод о том, что, имитируя «чуждость» речи, говорящие сознательно избегают тех словообразовательных средств, которые призваны обозначать «свой» микромир [Апресян 1995: 145; Протасова 2001а, 2001б], однако эту гипотезу можно будет проверить в дальнейшем. В записях встречаются так-

же случаи «буквального» произношения, например: *И что ты ВУК: чувствуешь :ВУК?* или исковерканные слова типа: *Павел@ Ну хорошо, давай ВУК: пыкрасим{*покрасим} дыверь{*дверь} :ВУК #смеётся#*. Существенным для понимания реплики может оказаться и то, что некоторые слова произносятся шепотом или беззвучно, одними губами, например: *Пётр@ Он такой _осёл_, Василий@ Это полное _+_/_ эта история//*. Для темы настоящей статьи важно то, что аффиксированное произношение и в этом случае скорее препятствует использованию диминутивов.

Если говорящий переходит на язык нянь («сюсюкает», например, произносит *холосенький* вместо *хорошенький*), эта часть реплики выделяется символом ВАВ. Такой стиль речи может встречаться не только в разговоре с детьми и домашними животными: *Наталья@ Ой, у меня ВАВ: клузецька{*кружечка} :ВАВ разбилась*. Давно замечено, что речь, обращенная к детям, животным или интимным партнерам, содержит повышенный процент диминутивов, как семантических, так и эмоционально-оценочных [Dressler, Merlini-Barbaresi 1999: 5, Гаврилова 2002, Savickiene, Dressler 2007, Protassova, Voeikova 2007]. В этих ситуациях говорящие не скупаются на средства интимизации речи, используя как фонетические, так и словообразовательные и лексические маркеры. Однако эти особенные речевые регистры пока что отсутствуют в нкря.

Сбои согласования, которые часто происходят в спонтанной речи, в текстах корпуса специально не маркировались. Их можно найти, используя программу морфологического поиска. В нашей внутренней инструкции мы, однако, просили расшифровщиков отмечать такие случаи для удобства дальнейшего анализа, например: *Наташа@ В <вашем сумочке> телефон, по-моему звонит*. Отмеченные особенности текстов не входят в тот минимум разметки, которая применяется в нкря. Однако в других корпусах применяются единые знаки для кодирования ошибок, так что можно было бы ввести подобные правила и для записи текстов в нашем корпусе.

3. Структура подкорпуса устной речи.

В настоящее время подкорпус устной речи нкря включает записи спонтанной речи, интервью (устная публичная речь), расшифровки текстов кинолент (см. подробнее статью Е. А. Гришиной и С. О. Сав-

чук в наст. сборнике). Непубличная речь в целом составляет немногим более 8% всего массива текстов. Это объясняется трудоемкостью сбора данных устной речи, большими затратами времени на расшифровку текстов. Сходное соотношение объемов материала (за исключением того, что мы совсем не занимались языком кино) было и в нашем подпроекте: расшифровки устной публичной речи составляют 83% всех собранных данных. Данные разговорной речи, собираемые для НКРЯ, поступали из разных источников. Часть материала мы получили от Санкт-Петербургского Центра независимых социологических исследований. Эти файлы представляют собой расшифрованные социологические интервью, характеристики которых близки к бытовой диалогической речи. Другие записи производились специально, в большинстве случаев с ведома информантов, хотя в таких условиях мы получали в основном однородную в социальном отношении и непредставительную выборку, так как только хорошие знакомые, понимающие цели и задачи подобного сбора материала, соглашались на то, чтобы их речь была зафиксирована, и чувствовали себя непринужденно в условиях записи.

Наиболее серьезной проблемой подкорпуса устной речи остается его социальная несбалансированность: большинство записей сделано с участием студентов или лиц, имеющих высшее образование. Для большинства текстов, описываемых как разговорные, характерен высокий образовательный статус участников. В нашей выборке лишь 11% текстов записаны с участием информантов, не имеющих высшего образования, — а в эту группу попадают и люди, образовательный уровень которых нам неизвестен. Понятно, что это искажает статистическую картину функционирования устной разговорной речи в целом. Образовательный статус влияет на лексическую и синтаксическую составляющие текстов; морфология, по нашим предварительным наблюдениям, в меньшей степени зависит от социального статуса говорящих. Особенности словоупотребления также, на первый взгляд, не зависят от того, является ли разговор бытовым или публичным, и от того, насколько высок образовательный уровень участников диалога. Однако эти первые впечатления нуждаются в статистической проверке. Необходимо в целом определить, какие социальные параметры могут оказать влияние на речевые характеристики. До сих пор в Корпусе учитывались только

возраст, пол и уровень образования участников общения, а также место записи (в нашем случае—Санкт-Петербург). Действительно, простая интуиция говорит о том, что речь двух мужчин или двух женщин, находящихся в одно и то же время в Петербурге и закончивших одинаковое количество классов или имеющих высшее образование, может сильно различаться. Однако исследований того, какие еще социальные параметры (например, род занятий, уровень дохода, сфера интересов, отношение к религии и т.д.) влияют на различные речевые характеристики, на материале современной разговорной речи не проводилось. Неясно также, какие именно речевые параметры могут зависеть от социального статуса. Например, можно предположить, что возраст, пол и род занятий отражаются в особенностях произношения (естественности основного тона и соотношении ударных и редуцированных гласных), однако эти характеристики не попадают в расшифрованные данные и остаются только в виде звуковых файлов. Для нашей темы (употребление диминутивов) такие особенности речи играют только косвенную роль. Более того, улавливаемые на слух признаки «аффектированного» произношения в речи молодых девушек (повышение тона на границах синтагм, затягивание ударных слогов и нарочитое сокращение редуцированных) скорее препятствуют употреблению диминутивов. Уменьшительно-ласкательные суффиксы в общении взрослых людей чаще встречаются в дружеских и интимных беседах, лишенных аффектации.

Публичная речь на современном этапе может рассматриваться как образец устной неформальной речи, так как отсутствие внешней и внутренней цензуры заметно во всех областях неформального публичного общения. Однако тексты, записанные в аудитории, отличаются малой диалогичностью: обычно участники интервью говорят значительно больше, чем журналисты и диалог становится неравноправным. В таких условиях создается меньше возможностей для эллипсиса, так как нарушается характерная для бытовой диалогической речи спаянность соседних реплик. Кроме того, специфика устной публичной речи фактически исключает разговоры на бытовые темы, что может повлиять на ее лингвистические особенности. Например, в сфере диминутивов наблюдается почти исключительное использование таких диминутивов первой сту-

пени образования, которые фактически только содержат специальные суффиксы, но не имеют диминутивной семантики (например, *книжка* или *скверик*). Функционируя наравне с диминутивами второй ступени, содержащими два и более суффикса, такими как *книжечка*, диминутивы первой ступени полностью утрачивают семантический компонент уменьшительности или развивают новое значение и постепенно лексикализируются. Например, слово *сетка* в 89% употреблений (из 34 документов, 98 контекстов) встретилась в сочетании *тарифная сетка*, *телевизионная сетка*, и лишь в 8 контекстах из фильма «Особенности национальной охоты» оно попадает в значении «рыболовная сеть». Всего один раз *сетка* употреблена в значении «верхняя часть металлической кровати», а в целом вполне живое значение «сумка для переноски продуктов или для хранения некоторых видов овощей (картошки, лука и т.д.)» не встретилось ни разу, что лишний раз показывает необходимость пополнения корпуса за счет бытовой устной непубличной речи.

4. Поиск диминутивов в НКРЯ

Диминутивы в НКРЯ представлены в семантической разметке как словообразовательная характеристика. Поиск по семантическим признакам для предметных и личных имен предусматривает возможность выбора их диминутивных вариантов. Отметим сразу, что жесткая структура поиска не предполагает выдачу диминутивов от непредметных имен, хотя теоретически они образуются от любого имени существительного. Ограничения в образовании диминутивов рассматриваются в работе [Спиридонова 1999]. Автор приводит список морфонологических запретов и подробно рассматривает, какие семантические компоненты словарного значения, а также когнитивные особенности представлений говорящих по-русски людей препятствуют образованию слов с уменьшительно-ласкательными суффиксами. Так, главными препятствиями для диминутивной деривации являются ингерентно большие и ингерентно малые размеры объектов, а также наличие у них постоянной, четко очерченной функции (ср. *гитара*, *рояль*, *холодильник*) [Спиридонова 1999: 15–16]. Ингерентно малые размеры, особенно у животных и птиц, все же не являются серьезным препятствием: в работе отмечается, что слова типа *ежик*, *синичка*, *мышка* «в плане обозначения разме-

ра вполне нейтральны», в то время как употребление симплекса скорее указывает на превышение обычных размеров [Спиридонова 1999: 15]. Критерий закрепленности функции в основном выполняется, и хотя слово *гитарка* все же один раз встретилось в контексте: [Респондент, муж] *В квартире у кого-нибудь танцуют, музыкальные, приходишь, всегда составчику каждого.* [Респондент, муж] *Там гитарка, домра... 4–5, а я всегда руководил.* [Биография (беседа лингвиста с информантом) // «Архив Хельсинкского университета», 1997], — другие существительные с отмеченным семантическим ограничением не попались ни разу.

Отметим, что некоторые морфонологические препятствия говорящие также преодолевают, если есть специальные прагматические условия. Так, ссылаясь на исследование А. К. Поливановой [Поливанова 1967], Н. Ф. Спиридонова указывает на то, что образование уменьшительно-ласкательных производных от существительных *автобус* и *бульдог* невозможны. В нкря, однако, и такие образования встречаются, правда, только по одному разу:

[Информант, муж, 51] *Автобусик так толкали / кто там есть / вытлкаивали его на улицу / на главную дорогу.* [Биография (беседа лингвиста с информантом) // «Русская разговорная речь европейского северо-востока России / Под ред. Н. С. Сергиевой и А. С. Герда, Сыктывкар, 1998», 1982]

[Коля, Никита Михалков, муж, 18, 1945] *Пока вы там богу молитесь / эта бульдожка людей кусает.* [«Георгий Данелия, Геннадий Шпаликов. Я шагаю по Москве, к/ф» 1963].

Важно, что эти примеры встретились нам только тогда, когда было набрано конкретное слово, в результате общего поиска по семантическому параметру они почему-то не попали. Таким образом, однократная встречаемость слова действительно указывает на то, что его диминутивизация затруднена, хотя и преодолима.

Результаты семантического поиска заставляют думать, что диминутивы размечены формально, по суффиксу, поэтому в их число попадают существительные типа *носки, коробка, картошка, петрушка, кабачки*, фамилии типа *Лужков* и *Порожков*, выражения типа *в рамках*. Основанием для выделения диминутива является наличие симплекса, поэтому *бабушка* и *дедушка* попадают в их число, а *внучка* нет, хотя и понятно, что слова *баба* или *бабка* не являются

симплексами по отношению к слову *бабушка*. Здесь семантические связи между производящим и производным затемнены, диминутив можно считать отдельным словом. Разные значения одного и того же слова могут ввести в заблуждение: так, среди «диминутивов от названий продуктов питания» оказалось слово *ушки*. Хотя ушками, действительно, называют особый вид печенья (*Вот ушки берлинские*. [Разговор о пончиках (2006.11)]; [собаке] *Нора / ты берлинские хочешь пончики или берлинские ушки?* [Разговор о пончиках (2006.11)]), а в юго-западных говорах — и разновидность пельменей, тем не менее эти употребления далеко разошлись по семантике с исходным словом *ухо* и по критерию наличия симплекса не могут считаться диминутивами.

Результатом поиска в подкорпусе устной речи часто являются отдельные реплики диалога. При этом контекст «от точки до точки» оказывается недостаточным, так как может содержать только одно слово, например *пивка; блинчики тока я; еще пирожок; на Горбушке*. В целом оказалось, что из-за большого процента ошибочно найденных слов воспользоваться статистическими данными поиска невозможно. Однако первоначальный семантический поиск диминутивов-существительных на материале подкорпуса устной речи дал 1146 документов и 15950 контекстов использования диминутивов, что составляет основу для значительной выборки, если учесть, что соотношение ошибок и истинных диминутивов доходило до 9% ошибочного поиска в первых 500 произвольно взятых контекстах. Более специализированный поиск позволяет снизить число таких ошибок: так, поиск диминутивов от наименований еды и напитков в устном подкорпусе дает 391 документ с 1195 контекстами, причем процент ошибок значительно ниже (3,8%). Ошибки здесь связаны с многозначностью и омонимией найденных форм, например: [Болтянская, жен] *Пришла аннотация на себя / «Я пенсионерка / но нить / что мало платят / не буду / это и так у всех в печенках»*. [Беседа И. Хакамады со слушателями радиостанции «Эхо Москвы» // 2003–2004], [Илюхин, муж] *Что ответил Юрий Яковлевич Чайка / министр юстиции РФ / о том / что Грызлов может совмещать партийную деятельность в «Единой России» и одновременно быть министром внутренних дел*. [Беседа О. Бычковой с В. Илюхиным в эфире радиостанции «Эхо Москвы» // 2003–2004]. Количество

ошибочно найденных диминутивов понижается в различных падежных формах: так, во 2-м род. падеже процент ошибок ниже (4,8%), чем в винительном (8,2%), что объясняется частотностью образования родительного отделительного от названий продуктов питания. Ошибки и в этом случае связаны с индивидуальной омонимией отдельных словоформ ([Венедиктов, муж] *Месяц тому назад здесь был ваш заместитель / директор финансовой разведки господин Зубков*), а также с системной омонимией, ср. род. пад. ед. ч. и им. пад. мн. ч. у слов ж. р. на –а ([Венедиктов, муж] *Нам в свое время объясняли / как Чаушеску / режим строгой экономики / мы тратим только то / что мы зарабатываем / лампочки в подъездах / дежурный по подъезду / чтобы не горело / не прорывалось и так далее*. [Беседа А. Венедиктова с А. Илларионовым в эфире радиостанции «Эхо Москвы» // 2003–2004]). Интуитивно ограничивая параметры поиска, можно больше или меньше приблизиться к типичной сфере употребления диминутивов. Если более детальное обследование покажет, что ограничение параметров поиска влияет на процент «истинных» диминутивов в выдаче, это будет дополнительно свидетельствовать о том, что прагматическое объяснение употребления диминутивов (см. ниже) более соответствует картине их функционирования в русском языке, чем исходно семантическое объяснение. В последнем случае никакой зависимости от контекста быть не может. Существенно также, что многие виды общения не попадают в сферу записей: практически не встретились примеры коммуникации с детьми и животными или ситуации флирта, ухаживания, а ведь известно, что это типичные прагматические контексты употребления диминутивов.

5. Диминутивы в устной речи: семантика и функционирование

Наш интерес к диминутивам, или словам с уменьшительно-ласкательными суффиксами, в русской разговорной речи основан на наблюдениях за усвоением русского языка детьми. Распространенное в регистре общения с детьми и в речи самих детей употребление диминутивов и гипокористик (уменьшительных личных имен) является общеизвестным фактом [Гаврилова 2002: 108]. Однако эта особенность не универсальна: процент диминутивов в речи взрослых, обращенной к детям, в разных языках различен [Savickiene,

Dressler 2007]. Например, несмотря на то, что в немецком языке существуют продуктивные диминутивные суффиксы, сами уменьшительно-ласкательные наименования в языке взрослых встречаются не так часто и преимущественно в австрийском варианте немецкого [Korecky-Kröll, Dressler 2007: 207–208]. В иврите количество диминутивов в языке взрослых, обращенном к ребенку, значительно ниже, чем, например, в славянских и балтийских языках. Это заставляет предположить, что их употребление связано также и с языковой структурой и не может объясняться только семантическими и прагматическими особенностями, которые должны были бы привести к универсальному предпочтению диминутивов в ситуациях общения с детьми. Для объяснения этих фактов необходимо рассмотреть основные особенности диминутивов в русском языке.

Русский язык, подобно другим славянским (и некоторым балтийским) языкам, обладает разветвленной системой диминутивных суффиксов, позволяющих образовать уменьшительно-ласкательные наименования не только от существительных, но и от прилагательных, наречий, междометий и некоторых глаголов [Bratus 1969]. Несмотря на все разнообразие средств, можно заметить, что большинство русских диминутивных суффиксов оканчивается на -к, попадающий в позицию непосредственно перед вокалическими падежными окончаниями. Диминутивы м.р. могут содержать распространенные суффиксы -ик, -ок\-ек, -онок\-ёнок: котик, коток, котенок; гипокористики м.р. и ж.р. оканчиваются на -ша, -а: Андрюша, Маша, Боря, Люба, деда, котя; среди суффиксов диминутивов ж.р. — -ичк-, -оньк/-еньк-, -ечк-/очк-, -ушк-, -ишк-, -уш: книжка, кисонька, кисочка, хвальбуша; в ср. р. основы диминутивов оканчиваются суффиксами -ц-, -к- (-ышк-, -ушк-, -ечк-): оконце, окошко, солнышко, семечко. Отметим, что форма ж.р. *семечка в названии подсолнечного масла «*Золотая семечка» выходит за пределы литературной нормы: изменение рода диминутива по сравнению с симплексом нехарактерно для русского языка, в отличие от немецкого, где род имени существительного определяется суффиксом.

Значение диминутивов некоторые исследователи связывают с семьей «малого размера», производными от которой являются и их остальные семантические особенности: сторонники семантического подхода, такие как Д. Журавски, считают, что представле-

ние о малом размере, свойственном детям (или женщинам), лежит в основе уменьшительно-ласкательной семантики и предопределяет сферу употребления [Jurafsky 1996]. Последователи разработанной В. У. Дресслером концепции морфопрагматики, напротив, считают первичными прагматические функции и не усматривают никакого следа количественной семантики, например, в ситуациях смягчения просьбы или неформального общения. Наши данные также показывают, что употребление диминутивов в дружеском общении не может объясняться представлениями о малом размере или количестве, ср. [Шойгу, муж] *Они идут на лед ловить рыбку / получать удовольствие. [Беседа А. Венедиктова с С. Шойгу в эфире радиостанции «Эхо Москвы» // 2003–2004].* Речь, разумеется, идет не о ловле маленькой рыбы или о небольшом ее количестве, а о том, что рыбная ловля является одной из сфер неформального общения. Если ситуация в диалоге напряженная, диминутивы отсутствуют.

В русском языке возможно образование двойных и даже тройных диминутивов (ср. нормативное *книжоночка* и детское **камушек* (пример Е. Ю. Протасовой)). В таких случаях первый диминутив обычно приобретает более специальное лексическое значение по сравнению с производящим словом и не ассоциируется ни с представлениями о малом размере, ни с известными прагматическими функциями диминутивов, такими как установление фамиллярности общения или смягчение просьбы. Так, например, разорвались семантические связи между словами *сеть* и *сетка*, причем диминутив стал обозначать более узкую разновидность сетей. В противоположность этому, у слов *клеть* и *клетка* диминутив приобрел более широкую систему значений по сравнению с симплексом. В цепочке, содержащей двойные диминутивы, типичные для этого класса слов прагматические или семантические особенности характерны лишь для второй или третьей ступени диминутивизации (ср. триады: *сеть—сетка—сеточка*, *дочь—дочка—доченька*, в которых лишь третий компонент имеет собственно диминутивный смысл). Н. Ф. Спиридонова приводит целый ряд диминутивов с утраченным симплексом (исходным членом пары): *клетка*, *кусок*, *платок*, *пятка*, *пучок*, *шматок*, *хорек* [Спиридонова 1999: 15]. Обращает на себя внимание тот факт, что все утраченные симплексы обладали редким и непривычным для русского языка звучанием, особенно

в финальной части: все слова мужского рода односложные, и их конечная часть — *-ус, -ат, -ук, -орь* — является редкой, по крайней мере с точки зрения *type frequency*: так, в словаре А. А. Зализняка отмечено 7 односложных слов на *-ус*, 17 односложных слов на *-ат*, в том числе такие редкие, как *пшат* (буквальный смысл текста — из иврита) или непонятный *там*, 13 слов на *-ук*, два односложных слова на *-орь*, включая искомое *хорь*. Можно предположить, что помимо морфонологических запретов, отмеченных А. К. Поливановой, семантических ограничений, установленных Н. Ф. Спиридоновой, и с учетом прагматических условий употребления диминутивов, отчасти снимающих указанные запреты (см. работы В. У. Дресслера), в русском языке существуют еще и морфонологические предпочтения, заставляющие выбирать более «благозвучную» и распространенную или в большей степени отвечающую языковому типу форму. Не случайно в рамках естественной морфологии стремление словоформы как минимум к двусложности считается более естественным, чем односложность. В силу такого свойства, как «системная адекватность» [Dressler 1999: 137], некоторые (например, германские) языки предпочитают неестественно короткие односложные формы, для русского же языка односложные слова нехарактерны, и язык пытается избавиться от них, в том числе и путем вытеснения этих коротких форм их диминутивными соответствиями. Наиболее интересным продолжением этой темы могло бы быть выяснение другого вида частотности — *token frequency* (частота встречаемости), вычисляемой не по словарю, а по количеству употреблений слов с определенным морфонологическим обликом в текстах, однако построить поиск с такими параметрами нам пока не удалось. Для осуществления подобного анализа требуется, чтобы запрос можно было задать списком слов, которые встречаются в контекстах не все вместе, на каком-либо расстоянии, а по отдельности (т. е. не «по и», а «по или»). Для этого в лексико-грамматическом поиске должна быть опция «ввести список слов» или «искать все слова из файла». Пока что такая опция нами не обнаружена.

В литературе присутствуют многочисленные указания на то, что употребление диминутивов приводит к упрощению системы склонения [Olmsted 1994], облегчает усвоение родовой принадлежности существительных [Kempe et al. 2001, 2003, 2005], помогает

сегментации речевого потока [Kemper et al. 2007]. Все это привело нас к мысли о том, что особенную роль в усвоении языка детьми играют созвучия конечной части слов [Protassova, Voeikova 2007: 63]. Именно они позволяют грамматически группировать существительные, легче вычленять их в речевом потоке и быстрее осваивать систему их словоизменения. Более пристальный взгляд на соотношение слов с уменьшительно-ласкательными суффиксами и соответствующих им производящих слов не позволяет сделать однозначного вывода о том, что частое использование диминутивов делает систему склонения более простой для усвоения. С одной стороны, их употребление сводит к минимуму набор слов непродуктивного 3-го женского склонения на согласный (наблюдение Х. Олмстеда), ср. пары *дверь—дверка*, *мышь—мышка*. При образовании диминутивов от слов среднего рода часто отмечается переход ударения с окончания на основу (*яйцо—яичко*, *колесо—колесико*), что позволяет соотнести эти безударные окончания с окончаниями продуктивного 2-го женского склонения (*птичка—яичко*), ср. марку растительного масла, ошибочно названную «Золотая семечка». Эти тенденции, действительно, можно рассматривать как упрощение системы склонения. С другой стороны, уменьшительно-ласкательные образования от слов мужского рода с наиболее продуктивными суффиксами отличаются большей сложностью при склонении по сравнению с производящими словами. Мужские личные имена при этом получают «женское» окончание *-а*, ср. *Михаил—Миша*, *Владимир—Вова*. У имен нарицательных при склонении нерегулярно утрачивается гласный в суффиксе, ср. *ключика—замочка*, к тому же в уменьшительном варианте они удлиняются на целый слог, что создает дополнительные произносительные трудности. Что же заставляет взрослых и детей использовать уменьшительные наименования мужского рода так же часто, как и диминутивы женского рода?

В некоторых случаях производящие слова даже не встречаются в записях спонтанной речи, т. е. уменьшительные варианты на время становятся единственным способом номинации объекта. Примерами таких слов мужского рода могут служить слова *пупок*, *зайчик* (или *зайка*) и *мышка*. В этих существительных обращает на себя внимание то, что все они происходят от слов с редкими

финалями (*луп, заяц, медведь*). По данным «Грамматического словаря русского языка» А. А. Зализняка, слов, оканчивающихся на *-уп* и *-яц*, меньше десятка, а слов, оканчивающихся на *-едь*, около двух десятков. Используя уменьшительную форму, говорящие интуитивно избегают редких и неблагозвучных финалей, стремясь к определенному «единообразию концов слов», которое, вероятно, помогает членить речевой поток [Protassova, Voeikova 2007: 67–68]. Наблюдения Н. Ф. Спиридоновой дают более богатый материал для этого утверждения, так как оказывается, что в большинстве утраченных симплексов присутствовало сочетание непривычной финали с односложностью.

6. ПРЕДВАРИТЕЛЬНЫЕ ВЫВОДЫ

Первичное обследование уменьшительно-ласкательных наименований от имен существительных предметной семантики показывает большие потенциальные возможности подкорпуса устной речи нкря. Корпус служит уникальным источником данных, которые можно собрать за считанные минуты. Сопоставимый по количеству отбор материала в прежние годы потребовал бы нескольких месяцев работы, а во многом оказался бы невозможным. Однако само по себе количество примеров не может быть положено в основание статистических подсчетов из-за того, что в выборку попадают омонимы, многозначные слова и личные имена. Дальнейший анализ выбранных примеров, сделанный вручную, позволяет оценить «степень чистоты выборки», которая может оказаться важным параметром для описания функционирования диминутивов. Перспективной является также комбинация грамматического и семантического поиска: отдельные формы диминутивов оказываются «более чистой» выборкой по сравнению с недифференцированным поиском. Материалы нкря, таким образом, дают возможность разработки новых методов лингвистического анализа данных. Даже самые первые шаги в этом увлекательном обследовании текстов позволяют уточнить и проверить выводы предшественников и сформулировать новые гипотезы, объясняющие загадочную склонность некоторых языков к диминутивам.

ЛИТЕРАТУРА

- Виноградов 1947—Виноградов В. В. Русский язык: грамматическое учение о слове. М.—Л., 1947.
- Гаврилова 2002—Гаврилова Т. О. Регистр общения с детьми: структурный и социолингвистический аспекты (на материале русского языка). Диссертация на соискание ученой степени кандидата филологических наук. СПб., СПбГУ, 2002.
- Гловинская 2007—Гловинская М. Я. Язык интернета как средство обнаружения неустойчивых участков языка // Ш Международный конгресс исследователей русского языка «Русский язык: исторические судьбы и современность». Труды и материалы. Москва, МГУ 20–23 марта 2007.—С. 180.
- Горбова и др. 2006—Горбова Е. В., Слепокурова Н. А., Черниговская Т. В., Комовкина Е. П., Матвеева Т. В., Риехакайнен Е. И. и А. С. Романова. Предварительные результаты мониторинга современной русской устной спонтанной речи. Современная русская речь: состояние и функционирование. СПб. Филфак СПбГУ. 2006.—С. 7–30.
- Гришина 2005—Гришина Е. А. Устная речь в Национальном корпусе русского языка // Национальный корпус русского языка 2003–2005. Результаты и перспективы. М.: Индрик, 2005.—С. 94–110.
- Земская 1973—Земская Е. А. Русская разговорная речь. М.: Наука, 1973.
- Земская и др. 1981—Земская Е. А., Китайгородская М. В., Ширяев Е. Н. Русская разговорная речь. Общие вопросы. Словообразование. Синтаксис. М., Наука, 1981.
- Красильникова 1990—Красильникова, Е. В. Имя существительное в русской разговорной речи. М., Наука, 1990.
- Поливанова 1967—Поливанова А. К. Образование уменьшительных существительных мужского рода // Русский язык в национальной школе, №4, 1967.
- Протасова 2001а—Протасова Е. Ю. Овладение категорией «свое–чужое» в детской речи // Теоретические проблемы функциональной грамматики. СПб.: Наука, 2001.—С. 238–246.
- Протасова 2001б—Протасова Е. Ю. Роль диминутивов в современном русском языке // Русский язык: система и функционирование. Tartu: Tartu ülikool, 2001.—С. 72–88.
- PPP 1978—Земская Е. А. (Отв. ред.) Русская разговорная речь. Тек-

- сты. М.: Наука, 1978.
- PPP 1983—Земская Е. А. (Отв. ред.) Русская разговорная речь. Фонетика. Морфология. Лексика. Жест. М.: Наука, 1983.
- Русский язык 1996—Земская Е. А. (Отв. ред.) Русский язык конца XX столетия (1985–1995). М.: Языки русской культуры, 1996.
- Спиридонова 1999—Спиридонова Н. Ф. Русские диминутивы: проблемы образования и значения // Известия АН, СЛЯ, 1999, том 58, № 2, с. 13–22.
- Bratus 1969—Bratus Boris V. The Formation and Expressive Use of Diminutives. *Studies in Modern Russian Language* 6. Cambridge University Press, 1969.
- Dressler 1994a—Dressler W.U. Diminutivbildung als nicht-prototypische Wortbildungsregel // K.-M. Köpcke (ed.) *Funktionale Untersuchungen zur deutschen Nominal- und Verbalmorphologie*. Tübingen: Niemeyer 1994. 131–148.
- Dressler 1994b—Dressler W.U. Evidence of the first stages of morphology acquisition for linguistic theory: extragrammatic morphology and diminutives // *Acta Linguistica Hafniensia* 27,1. 1994. 91–108.
- Dressler 1999—Dressler W.U. What is natural in natural morphology? // *Prague Linguistic Circle Papers*, V. 3. pp.135–144.
- Dressler, Merlini Barbaresi 1999—Dressler W.U., Merlini Barbaresi L. Morphopragmatics // J.Verschueren, J.-O. Östman, J. Blommaert and Ch. Bulcaen (eds.) *Handbook of Pragmatics*. Amsterdam: Benjamins, 1999. 1–14.
- Jurafsky 1996—Jurafsky, D. 1996. Universal tendencies in the semantics of the diminutive // *Language* 72, 533–78.
- Korecky-Kröll, Dressler 2007—Korecky-Kröll, K., Dressler W.U. Diminutives and hypocoristics in Austrian German (AG) // I., Savickiene W.U. Dressler (eds.) *The acquisition of diminutives. A cross-linguistic perspective*. Amsterdam: Benjamins, 2007. 207–230.
- Kempe et al. 2001—Kempe V., Brooks P., Pirott L. How can child-directed speech facilitate the acquisition of morphology? // M. Almgren, A. Barreña, M.-J. Ezeizabarrena, i. Idiazabal, B. MacWhinney (eds.) *Research on Child Language Acquisition. Proceedings of the 8th Conference of the International Association for the Study of Child Language*, 2001. 1234–1244.
- Kempe et al. 2003—Kempe V., Brooks P., Mironova N., Fedorova O. Di-

- minutivisation supports gender acquisition in Russian children // *Journal of Child Language* 30, 2003. 471–485.
- Kempe et al. 2005—Kempe V., Brooks P. The role of diminutives in the acquisition of Russian gender: Can elements of child-directed speech aid in learning morphology? // *Language Learning*, 55, 2005. Supplement: The Best of Language Learning. 139–176.
- Kempe et al. 2007—Kempe V., Brooks P., Gillis S., Samson G. Diminutives facilitate word segmentation in natural speech: Cross-linguistic evidence // *Memory & Cognition*, 35(4), 2007. 762–773
- MacWhinney 2000—MacWhinney B. *The CHILDES Project. Tools for Analysing Talk*. LEA Publ. Hillsdale, N.J., Hove & London, 2000 (3rd ed.).
- Miller, Weinert 1998—Miller J., Weinert R. *Spontaneous spoken language: Syntax and discourse*. Clarendon Press, Oxford, 1998.
- Olmsted 1994—Olmsted H. *Diminutive morphology of Russian children: A simplified subset of nominal declension in language acquisition* // Alexander Lipson: In memoriam. Slavica Publishers Inc. Columbus, Ohio, 1994. 165–207.
- Protassova, Voeikova 2007—Protassova E. Ju., Voeikova M. Diminutives in Russian at the early stages of acquisition // Savickiene I. & Dressler W. U. (eds.) *The Acquisition of Diminutives: A Cross-linguistic perspective*. John Benjamins. Amsterdam–Phil., 2007. 43–72
- Savickiene, Dressler 2007—Savickiene I. & Dressler W. U. (eds.). *The Acquisition of Diminutives: A Cross-linguistic perspective*. John Benjamins. Amsterdam–Phil., 2007.
- Sokolov, Snow 1995—Sokolov J., Snow C. *Handbook in Language Analysis Using CHILDES*. LEA Publ. Hillsdale, NY, 1995.