

1. СОСТАВ КОРПУСА

Сейчас в состав корпуса входят тексты архангельских, курских, брянских, тульских, орловских, псковских, новгородских, саратовских, волгоградских, вологодских и некоторых других говоров. Сравнительно в небольшом объеме представлены забайкальские говоры— благодаря текстам, полученным от В. Л. Кляуса.

Тексты различны по тематике и устройству. Среди них есть рассказы о конкретных случаях из жизни (тексты И. И. Исаева). Напротив, база А. Л. Мороза сконцентрирована большей частью на бычаях и традициях, а не на происшествиях из жизни конкретных носителей.

Каждый тип текста позволяет ставить свой круг исследовательских задач. Так, очень ценны длинные тексты, описывающие жизнь одного человека (например, рассказы Миньковой и Лопатиной из собрания Л. Л. Касаткина)— они позволяют проследить словупотребление и строение текста в зависимости от темы разговора. Более короткие тексты А. Л. Мороза выявляют различия в словупотреблении и грамматике между носителями одной и той же группы говоров.

2. НОВЫЕ ПОМЕТЫ В СИСТЕМЕ АННОТАЦИИ ДИАЛЕКТНЫХ ТЕКСТОВ

В связи с пополнением корпуса новыми текстами мы столкнулись с рядом диалектных явлений, не учтенных в предыдущей версии разметки. Их отражение в корпусе потребовало введения новых помет. Ниже мы кратко охарактеризуем каждую из этих помет.

2.1. Iter

В новой версии разметки появилась помета *iter* (итератив). Она используется для производных глаголов несовершенного вида типа *заганивать* ‘загонять’, соответствия которым в литературном языке нет. Точные видовые свойства этих глаголов, возможно, различаются для разных глагольных лексем. По-видимому, ближе всего они находятся к хабитуалису (постоянно повторяющейся ситуации) в прошедшем времени, см., например, *А она и пахать не няла и не рабливала* (Хрестоматия: Архангельск); *В порог нож не втыкают.*— *А под подушку?*— *Под подушку кладывала* (Мороз); *Я сам*

долго нашивал лапти (Хрестоматия: Вологда); *Каждый день ходила за ягодам, по корзине насбирывала* (Исаев: Владимир). Имеется в виду, что ситуация повторялась — более того, была постоянным свойством субъекта — а её положение во времени точно не определено. См. также обсуждение форм такого типа в (Пожарицкая 2005: 151–153), где указывается, что «временное значение данных форм близко к абсолютному плюсквамперфекту» (Пожарицкая 2005: 153).

Иногда значение повторяемости утрачивается, и производный глагол обозначает просто постоянное свойство предмета, например, способность: *Ак я ишшо тоже... сидела на пелёнках, не говаривала* (Исаев: Вологда). Заметим, что в тех немногочисленных случаях, когда эта форма употребляется в русском литературном языке, она всегда подразумевает повторяемость ситуации, например, невозможно **Ребёнок ещё не говаривал* при допустимом *Мой дядя говаривал, что родился в необычной семье*.

Эти формы описаны в работах по диалектологии, но корпусной подход позволяет выявить дополнительные ограничения на их употребление. По нашим данным, они сочетаются только с субъектом-человеком: казалось бы, семантически абсолютно нормальны сочетания типа *Скот на поле не хаживал*, однако в корпусе конструкции такого рода не встречаются. С одной стороны, это можно связать с общими особенностями диалектных текстов: как правило, они рассказывают именно о жизни людей, об их обычаях, а не о животных или неживой природе. С другой стороны, можно считать, что это семантическая особенность рассматриваемой формы: значение обычного состояния фиксируется именно для людей. Более того, в подавляющем большинстве случаев эта форма употребляется для обозначения обычного действия рассказчика или связанных с ним лиц. Наконец, в корпусе еще не зафиксировано употреблений хабитуалиса в настоящем времени.

Описанные выше образования мы отличаем от сходных по аффиксальному составу форм типа *отравливали*, ср. *Там детей отравливали и воровали бендеры* (Праведников): если формы типа *говаривать* образуются от глаголов несовершенного вида, то форма *отравливать* образована от глагола совершенного вида. Мы считаем, что это не хабитуальная форма, а обычная форма несовер-

шенного вида, образованная от глагола, которому в литературном языке такой дериват не свойствен.

2.2. Stem

Данная помета уже применялась в предыдущей версии разметки для случаев, когда в диалектной форме не наблюдается чередования, характерного для данной основы в литературном языке (ср. *пекёт* вместо *печёт*).

В нынешней версии разметки *stem* маркирует и ряд других диалектных явлений. К их числу относятся изменения в основе неизменяемых слов, например, *когда* вместо *когда*. Изначально конечный гласный, вероятно, был самостоятельным суффиксом временной или пространственной локализации, однако сейчас мы причисляем его к основе.

Далее, помета *stem* присваивается глаголам с изменением исхода основы, ср. *сокопляться* вместо *скапливаться*. В таких случаях исход основы отличается от литературного (-я вместо -ива). Однако тип спряжения при этом не меняется, поэтому мы приписываем эти отклонения именно основе.

В сфере местоимений *stem* обслуживает случаи употребления основы без формата *n-* при предлогах, ср. *жили до их мужчина с женщиной* (Исаев: Вологда).

В целом следует отметить, что случаи, охватываемые пометой *stem*, неоднородны. Одни из них связаны с образованием определённых форм (ср. словоформу *пекёт*, полученную от литературной глагольной основы), другие же затрагивают лексику в целом (ср. *сокопляться*, где такая основа выступает во всех формах глагола). В будущем, по-видимому, было бы целесообразным более подробно классифицировать явления, фиксируемые данной пометой и, соответственно, использовать для них несколько разных помет.

2.3. Refl и refltype

К помете *refl*, отмечающей отличие диалектного глагола от литературного возвратностью (*доспиеется* 'доспеет', Хрестоматия: Архангельск), на новом этапе работы был добавлен связанный с ней признак *refltype*. Данная помета призвана отразить вариативность

в форме возвратного показателя, широко представленную в диалектах (ср. *-се, -ся, -си*).

Необходимо отметить, что пометы *refl* и *refltype* существенно различаются по частотности и кругу охватываемых явлений. Помета *refl* встречается у ограниченного набора лексем, многие из которых по своим свойствам склонны к варьированию возвратности. Например, это приставочные глаголы типа *дожидать* (в литературном языке — *дожидаться*): здесь возвратность мотивирована не декаузативным или рефлексивным значением, а присоединением циркумфикса *до- -ся*, в котором вычленение отдельно префикса и постфикса сомнительно. В то же время отличия, связанные с возвратностью, встречаются и у бесприставочных лексем типа *гоститься* *вм. гостить, щадиться (с кем-л.) вм. щадить (кого-л.)*. Как видно, под данную помету подпадают два противоположных типа случаев: наличие возвратного постфикса, который отсутствует у данной лексики в литературном языке — и, напротив, отсутствие требуемого в ЛЯ постфикса.

Последовательное использование пометы *refl* в диалектных текстах осложнено тем, что в литературном русском языке модели, действующие возвратный постфикс, очень продуктивны. Поэтому некоторые возвратные дериваты (например, *выстираться* ‘постирать белье’) можно трактовать и как диалектизмы, и как не встречающиеся в текстах, но допустимое в литературном языке словообразование по продуктивной модели.

В отличие от *refl* помета *refltype* — одна из самых частотных: если в тексте встречаются отклонения в форме возвратного показателя, они, как правило, довольно многочисленны. При этом в одном и том же тексте может наблюдаться несколько вариантов возвратного постфикса. Помета *refltype* фиксирует как варианты, вообще не встречающиеся в литературном языке (*-се* и *-си*), так и употребление *-ся* в формах, где в ЛЯ требуется *-сь* (обратные случаи практически не встречаются). Часто формы, соответствующие норме и ей противоречащие, свободно варьируются (см. *родилась* и *родилась* в Праведников, текст 29).

Материал корпуса подтверждает, что категория возвратности, весьма сложным образом организованная в литературном языке, не менее интересна для анализа и в диалектных текстах.

2.4. Gend

Помета *gend* отвечает за отклонения от литературной нормы в сфере родовой принадлежности лексем, ср., например, *тюлью* (Мороз), *ягелью* (Азарова), *Форточка, она открыта всю лету была* (Касаткин), *Вот вам знамению даю* (Мороз), *хоть две литры носи* (Хрестоматия: Архангельск).

Определить род существительного — в том числе и выявить в этой зоне случаи диалектных отклонений от литературной нормы — можно, во-первых, на основе согласования существительного по роду с прилагательными (и с другими атрибутивами, ср. *всю лету*) или с глаголами, и во-вторых — по словоизменению самого существительного — в тех случаях, когда существительное изменяется по типу, характерному только для слов одного рода. Например, творительный на *-ью* могут иметь только существительные третьего склонения типа *мышь* — все они женского рода.

Размечая тексты, мы отдаем себе отчет в том, что отклонения по роду в некоторых случаях могут быть результатом спонтанной замены говорящим в речевом потоке изначально задуманного слова на другое слово, с иной родовой принадлежностью, ср. *приделают, какую блюдце ли, чего* ‘приделают какое-нибудь блюдце’ (возможно, исходно говорящий предполагал употребить слово *тарелка*). В подобных случаях мы тем не менее приписываем существительному помету *gend*, следуя, так сказать, принципу максимальной разметки: всегда лучше увидеть в тексте лишние диалектные черты, чем не заметить их.

2.5. Asp

Помета *asp* приписывается глагольным формам, которые отклоняются от литературного стандарта по своим аспектуальным свойствам. В первую очередь речь идет о формах, обнаруживающих в диалектном тексте иное аспектуальное значение, нежели в литературном языке. Так, например, в вологодских текстах формы типа *помирал* могут интерпретироваться как ‘помер’, то есть как формы совершенного вида.

Особенности аспектуальной семантики часто оказываются незаметны для изолированной формы, однако выявляются при сопоставлении с контекстом. Например, в песенке-колядке *Коляда, коляда, // Бабка, дай пирога, // Не даёшь пирога, // Обломаем все*

бока (Мороз) форма *даёшь* должна быть помечена *asp*: при бессоюзных условных конструкциях в литературном языке два глагола обычно имеют одну и ту же видовую характеристику, соответственно, форма *даёшь* употребляется в данном контексте, скорее всего, в значении совершенного вида. Точно так же словоформа *замерзал* характеризуется в литературном языке несовершенным видом, но в контексте *Он уже давно погиб — замерзал* (Исаев: Владимир) она имеет семантику совершенного вида — речь идёт о завершившейся ситуации, тем самым форма *замерзал* получает помету *asp*.

Второй тип случаев, охватываемых пометой *asp*, представлен отсутствующими в литературном языке видовыми коррелятами к стандартным литературным глаголам. Правда, в нашем материале случаи такого рода единичны, ср. *когда к венцу поезжают* (вм. *едут*, Мороз): в диалекте образуется видовая пара *поехать / поезжать*. Наряду с пометой *asp* таким глаголам приписывается и помета, маркирующая диалектность лексемы, — *diallex*.

Для случаев, охватываемых пометой *iter*, вторая помета — *asp* — была бы избыточна, поэтому она им не приписывается.

2.6. Flex и type

Ключевыми для диалектного корпуса являются пометы *flex* (диалектное окончание) и *type* (диалектный тип словоизменения). Ниже мы поясним разницу между ними, поскольку основания для их противопоставления не всегда очевидны.

Помета *flex* означает, что словоформа имеет показатель, вообще не употребляющийся при данной части речи в литературном языке. Такие случаи встречаются сравнительно редко. Перечислим основные типы подобных отклонений:

1) формы третьего лица единственного числа настоящего времени с окончанием без *-т* (*пойдѣ, будѣ*) или с мягким *-ть* (*значить, приходитъ*). Такого рода случаи были немногочисленны в текстах, размечавшихся до 2006 года, однако увеличение объёма корпуса позволило пополнить базу подобных примеров.

2) употребление формы дательного падежа местоимения *она* (*ей*) в контексте родительного или винительного, например, *ей доволочил* ‘доволочил её’ (Мороз). Хотя такие случаи можно было бы трактовать как нестандартное употребление дательного падежа,

более естественно всё же считать, что *ей* является особой формой винительного падежа: вариант *ей* в такого рода контекстах встречается даже у информантов, обычно употребляющих падежи стандартным способом.

3) употребление формы *оне* в значении формы множественного числа именительного падежа *они*.

4) формы творительного падежа прилагательных типа *хороши-ма* и творительного падежа существительных типа *дорóгими* (вм. *дорогами*). Первая форма вообще не встречается в русском литературном языке, вторая характерна только для прилагательных и не засвидетельствована у существительных.

К менее значимым случаям, охватываемым пометой *flex*, относятся а) нелитературное окончание нумеративов типа *обеи* (вм. *обе*): в литературном языке в женском роде у нумеративов встречается только окончание *-е* (*две, обе*); б) изменение окончания в творительном падеже единственного числа (*прикладим* вм. *прикладом*) и в) фонетическое окончание *-я* в императиве множественного числа (*сымайтя* вм. *снимайте*).

Из всех перечисленных случаев наиболее частотной специфически диалектной флексией является окончание с мягким *-ть* в третьем лице глаголов — этот феномен может встречаться более десяти раз на двух страницах текста. Корпус позволяет установить, что чаще всего подобное изменение флексии связано с формами *значить* и *может* ‘может быть’, причем с их употреблениями в качестве вводного слова, а не в позиции финитного сказуемого.

Подчеркнем, что помета *flex* маркирует случаи, важные с точки зрения анализа грамматических особенностей диалектных текстов: они свидетельствуют о том, что различия между диалектами и литературным языком не исчерпываются перегруппировкой типов словоизменения, — диалекты характеризуются и более значимой грамматической спецификой.

В отличие от *flex* помета *type* употребляется довольно широко. Отклонение от стандартного словоизменительного типа может затрагивать, во-первых, всю лексему в целом, т.е. проявляться во всех ее формах, ср., например, *купелю принесут* вм. *купель* (Мороз), *Ленин сделал жизнь правильную* вм. *жизнь* (Касаткин). Такие переходы обусловлены влиянием более продуктивного типа склоне-

ния на *-а / -я*. Во-вторых, изменение типа может прослеживаться лишь в некоторых грамматических формах лексемы — остальные ее формы образуются в соответствии с литературной нормой. С фонетическими причинами связан, например, частый переход прилагательных в класс *плохой* (ср. *хорошой* вм. *хороший*).

Об изменении типа, затрагивающем отдельные формы, можно говорить и в случаях использования падежной флексии второго родительного или предложного падежей при тех лексемах, которым в литературном языке они не свойственны, например, *острову, на камню*. Хотя существительные, допускающие род2 и пр2, не составляют типа склонения в общепринятом смысле, вполне естественно объединить их в одну группу и считать диалектное расширение этой группы за счет новых лексем изменением флексивного типа этих лексем.

Необходимо отметить, что иногда мы используем помету *type* и в тех случаях, когда вариативность словоизменительного типа наблюдается и в литературном языке. Например, в материале архангельских говоров регулярно встречается форма *годов* в генитиве множественного числа (ср. литературное *лет*). Форма *годов*, по видимому, не запрещена в литературном языке, но приписываемая в данном случае помета призвана показать, что в диалектах основным становится вариант формы, редкий для литературного языка.

В принципе пометы *type* и *flex* могут относиться к одной и той же словоформе, хотя на практике такие комбинации редки. Например, глагол, перешедший в другой класс словоизменения, может иметь и окончание *-ть* в третьем лице. Чаще, однако, встречаются случаи другого рода — это своего рода промежуточные явления, для которых выбор между пометами *type* и *flex* неоднозначен. К примеру, форма предложного падежа (*на*) *обех* '(на) обеих' должна трактоваться как диалектная флексия, если считать нумеративы особым классом, отличным от местоимений — и как диалектный тип, если считать нумеративы и местоимения одним классом (поскольку такую же флексию имеет местоимение *том*). С одной стороны, в корпусе литературного языка нумеративы противопоставлены местоимениям, с другой — эти грамматические классы довольно близки, соответственно, в подобных случаях мы приписываем словоформе обе пометы.

2.7. Case

Помета *case* описывает употребление (с точки зрения литературного языка) одной падежной формы вместо другой (например, см. выше *ходила по ягодам* вм. *по ягоды*). Здесь происходит унификация управления при предлоге *по* (нестандартное управление *по* + Вин. меняется на наиболее частотное для этого предлога *по* + Дат.). Эта помета, в отличие от большинства остальных, фиксирует не нестандартное слово- или формообразование, а нестандартное употребление определенной — в данном случае падежной — формы (похожие случаи описаны для пометы *asp*). В целом *case* применяется достаточно редко: как правило, падежи в текстах корпуса используются в соответствии с литературной нормой.

К явлениям, описываемым пометой *case*, относятся в частности глаголы с нелитературной моделью управления, ср. *И кормят кусок этого хлеба ей* (корове) (Мороз), где глагол *кормить* управляет названием пищи в винительном падеже и пациенсом в дательном.

В целом применение пометы *case* к диалектным текстам осложнено вариативностью падежей в литературном русском языке. Иными словами, отнесение конструкции к диалектным или к литературным, но свойственным разговорной речи, часто спорно. Например, в диалектных текстах встречаются сочетания типа *всё хватало* 'всего хватало', не соответствующие норме. Однако и в разговорной недиалектной речи при предикатах типа *нужно, хватать* наблюдаются колебания в падежном маркировании имени, например, при поиске в системе Google среди результатов, помимо нормативного *всего хватает*, выдаётся разговорное *всё хватает*. Мы полагаем, что в будущем в систему разметки следует включить разграничение между диалектными и общеразговорными явлениями, однако пока сочетания типа *всё хватает* относятся к диалектным. Аналогичный случай представлен словосочетанием *и что только нету* (Мороз): здесь пример из диалектного текста иллюстрирует вариативность падежей при отрицании, которая свойственна и литературному языку, — только в данном случае она распространяется на предикат *быть*, который в литературном языке не входит в круг глаголов с вариативностью кодирования субъекта при отрицании.

Следующий класс проблем, связанный с использованием пометы *case*, обнаруживается при анализе существительных редких типов склонения (например, *время*): в сочетаниях вида *через несколько время* (Праведников) слову *время* можно приписать как помету *case* (фиксирует употребление формы именительного/винительного вместо родительного), так и помету *type* (отражает переход слова в разряд неизменяемых). Выбрать правильную трактовку данного явления на материале небольшого текста часто невозможно, поэтому в подобных случаях мы приписываем словоформе обе пометы.

Наконец, анализ падежного кодирования имен в диалектных текстах неизбежно сталкивает исследователя с проблемой эллипсиса — синтаксического явления, чрезвычайно распространенного как в диалектах, так и в разговорной речи. В отрывке *Укажного свой рожок. И кажного по-разному играё* (Мороз) родительный падеж второй словоформы *кажного* не соответствует литературной норме. Это отклонение обусловлено, видимо, синтаксическими различиями между диалектами и литературным языком: во второй фразе опущено существительное *рожок* (*кажного рожок*, ‘рожок каждого человека’), т. е. имя, управляющее другим именем в генитиве, а такого рода опущения невозможны в литературном языке. Таким образом, строго говоря, употребление генитива в данном контексте отвечает литературной норме, а диалектное своеобразие возникает за счет нестандартного эллипсиса. Тем не менее подобные случаи мы маркируем пометой *case*, следуя общей стратегии НКРЯ: отсутствие синтаксической аннотации в корпусе частично компенсируется средствами морфологической разметки.

На примере *case* можно показать, что диалектные морфологические пометы применимы даже к словоформам, не имеющим аналогов в литературном языке, например, к именам собственным. То или иное грамматическое значение в диалектных текстах иногда выражается формами, которые в литературном языке не могут соотноситься с данным значением, к какому бы типу склонения лексема ни принадлежала. Ср., например, словоформу *Умбы* в предложении *мы в Умбы жили* (Азарова): какой бы ни была исходная форма для данной словоформы — *Умба*, *Умбы*, *Умб* — в предложном падеже слово не может иметь окончания *-ы*, соответственно, мы фиксируем в данном контексте диалектную падежную форму. Здесь

произошла унификация окончания *-ы* в предложном (а также и дательном) падеже 1 склонения (*у жены, дать жены, о жены*) — явление, характерное прежде всего для западной диалектной зоны.

Итак, мы рассмотрели новые морфологические пометы, которые дополнили систему разметки диалектных текстов, и охарактеризовали стоящие за ними грамматические явления. Понятно, что добавление в корпус новых текстов повлечет за собой дальнейшее расширение системы разметки. Так, совсем недавно мы начали использовать помету *plen* — полное прилагательное вместо краткого — для случаев типа *Он был раненым* ‘он был ранен’ (Праведников).

3. НЕКОТОРЫЕ ПРОБЛЕМЫ МОРФОЛОГИЧЕСКОЙ РАЗМЕТКИ КОРПУСА ДИАЛЕКТНЫХ ТЕКСТОВ

3.1. Диалектная разметка и стандартные литературные формы

До сих пор речь шла о разметке словоформ, отличающихся от литературных по модели образования или по правилам употребления. Однако пометы, фиксирующие такого рода диалектные явления, заставляют задуматься и над анализом стандартных литературных форм. Так, если в некотором тексте мы встречаем слово *дедушко* вм. *дедушка* (напомним, этот случай маркируется пометой *type*), то как нам следует интерпретировать форму *дедушки* в словосочетании *наши дедушки*, встретившуюся в том же тексте? Как пометить исходную лексему — *дедушко* или *дедушка*? Форма множественного числа не отличается от литературной, но, сохраняя принцип последовательности в разметке, было бы логично приписать всем словоформам данной лексемы (в том числе и не отклоняющимся от литературной нормы) одну и ту же исходную форму (т.е. *дедушко*) и, соответственно, помету *type*.

Однако обычно мы сознательно этого не делаем. Причин две. Во-первых, это экономия места и времени: разметка словоформ, которые формально не отличаются от литературных, как диалектных значительно замедлила бы процесс аннотирования. Сейчас технически, в основном, достаточно проверить словоформы, не опознанные разметчиком (хотя для случаев вроде *ходить* ‘ходит’ такой подход не

срабатывает), а если бы мы следовали принципу одинаковой лемматизации всех словоформ, проверять требовалось бы гораздо больше.

Во-вторых, вариативность в диалектах настолько высока, что в одном тексте могут встретиться разные исходные формы одной и той же лексемы, ср. *ён* и *он*, *дедушко* и *дедушка*, — а значит, принцип последовательности вообще не применим к данному типу текстов. И все же в нашем материале мы сталкиваемся и с явлениями, которые не обнаруживают вариативности в пределах одного текста — например, употребление слова *ягель* в женском роде в (Азарова) — *ягелью*. В этих случаях мы маркируем как диалектные и формы, совпадающие с литературными (так, в примере выше мы приписываем помету *gend* и словоформам именительного и винительного падежей, ср. *ягель даёт*).

3.2. Частеречные характеристики

Разметка по частям речи в диалектном корпусе сопряжена с теми же трудностями, что и частеречная классификация в основном — литературном — корпусе. В частности, одна из основных проблем связана с противопоставлением служебных частей речи — союзов и частиц. В диалектных текстах эта зона оказывается еще сложнее для анализа: в отличие от письменных литературных текстов, где при разграничении между союзами и частицами можно опереться на установленные автором границы предложений, в диалектных, как и в других устных текстах, членение на предложения достаточно условно, тем самым здесь этот критерий нельзя считать надежным. Чтобы не навязывать пользователю своих исследовательских решений, мы в спорных случаях сохраняем два варианта разметки. Так, две пометы — CONJ (союз) и PART (частица) — приписывается большинству употреблений слова *ак*.

3.3. Новые проблемы лемматизации

Как оказалось, лемматизация, принятая в литературном корпусе, в некоторых случаях неприемлема для диалектного. Часто в литературном корпусе объединяются лексемы несовершенного и совершенного вида: например, словоформа *сбежали* возводится к лемме *сбегать*, и, соответственно, совершенный вид считается словоизменительной характеристикой.

В диалектных текстах одной и той же лексеме несовершенного вида могут соответствовать разные лексемы совершенного: например, *сбегли* и *сбежали* от *сбегать*. Тем самым лемматизация формы *сбегли* по лексеме *сбежать* не позволила бы отразить ее специфику: требуется показать, что в диалектном тексте используется нелитературная лексема. Поэтому в таких случаях мы используем лемму совершенного вида (в примере выше — *сбечь*).

3.4. Диалекты и просторечие

В отдельных случаях как диалектные размечаются формы и лексемы, которые встречаются и в недиалектной речи. К примерам такого рода относится глагол несовершенного вида *ложить* (вм. *класть*) и, наоборот, дериваты типа *накласть*, *покласть* вм. *наложить*, *положить*. Городское просторечие в корпусе не имеет специальной системы помет, однако, приписывая подобным явлениям диалектный статус, мы отмечаем их ненормативность в литературном языке.

Другой пример такого рода — маркирование случаев типа *у ней* (форма дательного падежа употребляется в функции родительного). Ясно, что эти формы — просторечные, тогда как примеры вроде *у мене* или *к сестры* — именно диалектные (во всяком случае, они реже встречаются вне диалектов). Тем не менее сейчас они помечаются одинаково: в системе диалекта они, как нам представляется, имеют одинаковый статус.

4. ЗАКЛЮЧЕНИЕ

Мы изложили основные особенности разметки диалектных текстов. Существенно, что во многих случаях корпусной подход позволяет уловить тенденции, которые без использования корпуса идентифицировать трудно. Среди них, в частности, ограничения на некоторые диалектные явления (например, использование имперфективов типа *говаривать* преимущественно с одушевлёнными подлежащими). Корпус позволяет выяснить и соотношение частотности диалектных явлений, например, изменения основы (помета *stem*) и изменения окончания (помета *flex*).

Впрочем, тенденции второго типа можно отметить только на больших массивах текстов. Их исследование — скорее дело будущего.

ЛИТЕРАТУРА

- Летучий А. Б. 2005. Корпус диалектных текстов: задачи и проблемы // *Национальный корпус русского языка: 2003–2005. Результаты и перспективы*. М.: Индрик.
- Пожарицкая С. К. 2005. *Русская диалектология*. М.: Парадигма—Академический проект.

ИСПОЛЬЗОВАННЫЕ ТЕКСТЫ

- Азарова—Диалектные тексты, собранные на территории Республики Карелия И. В. Азаровой.
- Мороз—Диалектные тексты, собранные на территории Архангельской области группой по изучению фольклора под руководством А. Б. Мороза.
- Исаев—Диалектные тексты, собранные на территории Владимирской и Вологодской областей И. И. Исаевым (для каждого примера отмечена группа диалектов).
- Праведников—Диалектные тексты, собранные С. П. Праведниковым на территории Курской области.
- Касаткин—Диалектные тексты, собранные Л. Л. Касаткиным на территории Волгоградской области (рассказы А. Н. Миньковой).
- Хрестоматия—Хрестоматия диалектных текстов под ред. Е. А. Нефедовой.