

НЕПОЛНОТА ПАРАДИГМ И СИММЕТРИЯ

Необходимость отражения развивающегося и изменяющегося мира средствами языка приводит к тому, что активно происходит процесс моделирования новых слов. Любое слово первоначально возникает только в одной грамматической форме, причем далеко не всегда в начальной. И только в том случае, если слово путем повторения в разных грамматических позициях приобретает относительно стабильные грамматические характеристики, можно считать, что это слово вошло в систему языка.

Критерии, определяющие быстроту освоения нового слова, могут быть как экстралингвистическими, так и лингвистическими. К первым относится актуальность понятия, отражаемого словом, для настоящего момента жизни общества и, как следствие, повторяемость в общении и СМИ (для сравнения: отсутствие фиксации и популяризации, давление нормы лишает носителей языка возможности пользоваться шедеврами, например, детского словотворчества). Что касается лингвистических «требований» к новым словам, то, во-первых, многое зависит от того, на какой производящей основе создано слово. Основа родного или близкородственного языка в большинстве случаев гарантирует стандартность суффиксально-флективной части слова, и, следовательно, возможность приписывания всей парадигмы по аналогии. Во-вторых, большую роль играет удобопроизносимость слова, отсутствие нехарактерных для данного языка звукосочетаний. В-третьих, должна существовать возможность синтаксической адаптации, т. е. употребления в разных контекстах (не только в виде цитаты) и в разных синтаксических позициях, что выявит морфологические характеристики нового слова при возможных вариативных способах морфологической интерпретации.

Язык отражает мир настолько полно, насколько это необходимо для понимания говорящего слушающим. Следовательно, если у говорящего и слушающего заложены одинаковые языковые знания и подобные пресуппозитивные знания, не всегда есть необходимость в полноте их экспликации. Редко когда создание нового слова в его полной парадигме является самоцелью. Чаще всего словоформы создаются при отражении определенной ситуации. Форма отражения логики ситуации и будет грамматикой – морфологией и синтаксисом. А логическая ограниченность ситуаций и отношений между объектами приводит к тому, что грамматика (больше морфология, чем синтаксис) становится замкнутой системой в рамках данного языка, в отличие от лексики.

Анализ полноты / неполноты лингвистического универсума на морфологическом уровне требует решения следующих задач: 1) объяснить причины необходимости системной полноты на морфологическом уровне; 2) проанализировать вероятность употребления всех форм парадигмы

конкретного слова в зависимости от его частотности; 3) доказать, что большинство словоформ моделируется в процессе речевой деятельности по аналогии, т. е. симметрично-асимметричным способом; 4) показать на примерах функциональную неполноту парадигм существительного и выявить причины этой неполноты.

Из почти 40 тыс. слов, включенных в Частотный словарь русского языка [1], только 9044 встречаются в текстах с частотой не менее 10. Если учесть, что в среднем каждый класс имеет по 10 форм изменения (а на самом деле у глагола с учетом именных форм их насчитывается значительно больше), то вероятность того, что хотя бы по одному разу встретится каждая форма, равна единице менее чем для 25 % слов. Представительный объем текстов выборки – 1 млн. словоупотреблений – показывает, что лишь четвертая часть лексики теоретически может быть освоена в грамматическом отношении непосредственно из восприятия. Так, не зная русского языка, ничего нельзя сказать о парадигме существительного *клубника* (частота 1) или глагола *захиреть* (1), в какой бы форме они ни встретились: если не «встроена» система склонения существительных на *-ика* и спряжения глаголов на *-еть*, то отсутствуют формы для сравнения и, следовательно, идентификации грамматических категорий. Таким образом, 75 % слов требуют моделирования большей части своей парадигмы по аналогии, которая основана на принципе симметрии / асимметрии.

Частотный словарь Э.А.Штейнфельдт [2] приводит данные по частоте встречаемости форм парадигмы для большинства классов. Всего из отобранных для анализа текстов общим объемом 400 тыс. словоупотреблений получено 24 224 разных слова. В словарь включены только те 2 500, которые встретились в 14 текстах и более, т. е. с частотой не менее 14. Зная минимальную частоту и количество слов в полной парадигме одного класса, можно рассчитать вероятность встречаемости отдельной словоформы. Так, для существительного *игрушка* (частота 15) она равна 1,25 (при 12 формах в парадигме); для глагола *понадобиться* (частота 14) – 1,077 (при 13 формах в парадигме, без учета причастий), для прилагательного *приятный* (частота 14) – 0,1 (81) (при 77 формах в парадигме, с учетом простых форм степеней сравнения).

Покажем неполноту парадигм существительных на примере двух грамматических категорий: числа и падежа.

Число имен существительных – это грамматико-семантическая категория. При существовании в современном русском языке двух форм числа (ед. и мн.) возможно существование четырех и только четырех подсистем:

1) ++ существительные, имеющие формы ед. и мн. числа. В этой подсистеме затруднений с идентификацией и моделированием форм числа не возникает;

2) + – существительные *singularia tantum*. Традиционно к ним относятся лексико-семантические группы существительных, обозначающих процессы (обычно с суффиксами *-ни(е) / -ени(е), -б(а), -к(а)*, нулевым суффиксом),

химические элементы и вещества, признаки (с суффиксами *-от(a)*, *-ость*), названия наук, теорий и течений в науке, политике и т.п. (с суффиксом *-изм*) и некоторые другие;

3) – + существительные *pluralia tantum*: *ножницы, литавры, брюки* и т.п.;

4) – – неизменяемые существительные, число которых можно определить только синтаксически: *нов-ое пальто* и *нов-ые пальто*.

Вводя дополнительные признаки в классификацию (например, двойственное число или собирательную форму), мы получим соответственно классификацию из 8-и или 16-и подсистем.

Подсистемы не изолированы друг от друга, представляют собой целостную связанную систему, а примеры из текстов демонстрируют переход из одной системы в другую. Развитие семантики и разделение понятия хотя бы на два значения ведет к появлению форм мн. числа: например, *смол-а* ранее была только естественным веществом, но появились синтетические *смол-ы* разного состава, а также *клеи, допинги, глины, порохи и пороха, мяса...* Все они отмечены в банке данных НИЛ теоретической и прикладной лингвистики БГУ. Приведем примеры существительных и из других лексико-семантических групп, чтобы показать, что процесс «достройки» парадигмы до полноты очень активен: *реальности, виноватости, неразрешимости; борения, слушания, подбрасывания; застои, затишья, жары, любви, пользы* и даже *злы от зло* (Э. Успенский). В просторечии, у А. Белого, М. Булгакова и К. Чуковского зафиксирована форма *польта* и *польты* (от *пальто*), демонстрирующая переход из подсистемы 4 в подсистему 1. В этом случае несклоняемые существительные обретают парадигму и становятся склоняемыми, например: *Сразу видно – НИИ. Прошлый раз тоже из ниёв приехали...* («Литературная газета»); *Много хлопот было с нашими мадамами...* (В. Конецкий). Возможен переход и из подсистемы 3 в подсистему 1, хотя таких примеров меньше. Так, форма ед. числа появилась у существительных *ресурсы – ресурс, промтовары – промтовар* (в последнем случае аббревиатура *промтовары* сразу возникла как *pluralia tantum*, несмотря на то что мотивирована словом *товар*, имеющим обе формы числа).

В Частотном словаре [2] приведено распределение существительных по формам числа. Уже в первой сотне частотных слов обнаруживается неполнота по этому признаку: во-первых, отдельно приводятся супплетивные формы *люди, ребята, дети*; во-вторых, 6 слов не употребляются в форме мн.ч. по семантике (*помощь, правда, мир, солнце, внимание, молодежь*). Общее количество таких существительных составляет 187 из 1011. При общей тенденции преобладания форм ед.ч. у большинства существительных, некоторые существительные «тяготеют» ко мн. числу. Так, соотношение форм ед. и мн. числа у существительного *глаз* равно 39 : 419, *нога* – 51 : 148, *пора* 63 : 104, *рабочий* – 76 : 195, *гость* – 37 : 112. Можно выделить четкие семантические подгруппы: названия парных частей тела, названия людей, отражающие наш коллективизм (*пионеры, комсомольцы, ученики*), названия

интервалов времени (*секунды, минуты, года*). Если при анализе учитывать еще и максимумы падежных форм, то можно предполагать наиболее вероятные контексты. Так, пик употреблений слова *пора* приходится на форму род.п. мн.ч., следовательно, это могут быть сочетания *с тех пор, до сих пор*.

В отличие от категории числа, категория падежа имеет более сложную структуру. Употребление падежных форм также ограничивается логической семантикой, однако значения большинства русских падежей неоднозначны, и, следовательно, можно ожидать более равномерного распределения словоформ в парадигме. Однако, анализируя данные Частотного словаря [2], приходим к выводу, что в универсуме частотное распределение по падежам неодинаково как внутри парадигмы отдельного слова, так и в целом по словарю.

Наибольшую нагрузку несет именительный падеж: форма им.п. не отмечена только у 22 существительных, 4 из которых несклоняемые. Все без исключения одушевленные существительные имеют наибольшую частоту употребления именно в им.п.

Нагрузка форм род.п. также велика, эти формы не отмечены только у 20 существительных. Однако лексико-семантические группы существительных, имеющих максимумы частот в род.п., иные: время (*год, день, лета, час, минута, пора, месяц*), названия организаций и предприятий (*завод, школа, цех, комитет, союз, фабрика, участок, район*) и небольшая группа названий веществ (*вода, молоко*).

Дательный падеж имеет самую низкую частоту употребления. На эту форму приходится наибольшее число минимумов, а 278 существительных вообще ее не реализуют. Максимумы в дат.п. имеют 14 слов, в большинстве своем входящих в состав устойчивых сочетаний с предлогами *к* и *по*: *пример, сожаление, мнение, мера, повод, принцип, существо* и др.

Форма вин.п. занимает третье место по частоте употребления. Отсутствует у 35 существительных: одушевленных (*молодец, зритель, журналист, токарь, дед*), названий месяцев (*март, июнь, сентябрь, ноябрь*), названий чувств (*любовь, тоска*). Максимумы употребительности отмечены у неодушевленных существительных с конкретной семантикой (*дверь, окно, рука, письмо*).

Форма тв.п. отсутствует у 150 существительных, 4 из них – несклоняемые. Во многих словах отмечена с частотой 1–2. Максимальных нагрузок немного, выделяются следующие лексико-семантические группы: названия чувств (*тревога, гордость, восторг*), названия частей тела (*спина, палец, кулак*), инструментов (*лопата*), отвлеченных существительных (*способ, метод, образ*).

Форма предл.п. отсутствует у 219 слов, подавляющее большинство из которых – одушевленные существительные (*министр, чёрт, муж*). Максимумы форм предл.п. приходятся на существительные лексико-семантической группы места (*берег, зал, угол, деревня*).

Таким образом, неполнота парадигм существительных в универсуме неизбежна и зависит от их семантики. Конкретное слово тяготеет к той или иной логической функции в тексте, а следовательно, и к определенным формам падежей.

Рассмотрев неполноту форм парадигмы существительных в универсуме, приходим к выводу о необходимости моделирования полных парадигм слов при создании машинных словарей. Однако возникает вопрос о том, следует ли моделировать неупотребительные формы. Конечно, с одной стороны, машинный словарь предназначен для автоматического анализа реальных текстов. С другой же стороны, если потенциальная форма слова возможна (т. е. имеются элементы подобного фонетико-акцентологического рисунка), то она должна быть смоделирована, поскольку предугадать, в каком тексте она может встретиться, невозможно. Так, в нашем материале есть пример предложения *За сбычу мечт!*, источник – «Литературная газета». Странной форма *мечт* выглядит только потому, что в русском языке есть всего несколько слов с подобным фонетико-акцентологическим рисунком: *мечта, тахта, тайга, карга, балда, узда, нужда* и некоторые другие. По данным «Грамматического словаря русского языка» [3], только у последнего из приведенных примеров нет пометы, что образование род.п. мн.ч. затруднительно. Именно слово *нужда* становится базой для симметричного образования формы по аналогии, поскольку является высокочастотным.

Следует заметить, что даже при одинаковом фонетическом рисунке слов можно ожидать появления вариантов:

Им.п.	судьба	мольба	резьба	борьба
Род.п. мн.ч.	судеб	–	?	?
	–	мольб	?	?

При моделировании полной парадигмы слов с затруднительным формообразованием, таким образом, следует предусмотреть все возможные потенциальные формы на основе симметрии.

Однако необходимо учесть, что машинный словарь, в котором смоделированы потенциальные формы слов, применим только в системах анализа текста. В системах, предназначенных для синтеза текста, потенциальные формы должны быть помечены как неупотребительные.

Литература

1. Частотный словарь русского языка / Под ред. Л.Н.Засориной. М., 1977.
2. Штейнфельдт Э.А. Частотный словарь русского языка. Таллин, 1963.
3. Зализняк А.А. Грамматический словарь русского языка. М., 1980.