

УДК 519.766.2

М. Ф. БОНДАРЕНКО, В. М. БОНДАРЕВ

**МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СЛОВОИЗМЕНЕНИЯ ИМЕН  
СУЩЕСТВИТЕЛЬНЫХ. Сообщение I**

Связь между словом  $X$ , его формой  $Y$  и грамматическими характеристиками формы  $Z$ , объективно существующая в языке, может быть описана в виде тернарного отношения  $L(X, Y, Z)$ , названного морфологическим [1]. При этом  $X, Y \in S$  — множеству всевозможных последовательностей конечной длины, составленных из букв русского алфавита, символов дефиса и ударения;  $Z \in M$  — множеству наборов значений словоизменятельных грамматических категорий.

Будем говорить, что морфологическое отношение  $L(X, Y, Z)$  выполняется тогда и только тогда, когда: 1)  $Y$  — словоформа парадигмы слова  $X$ ; 2)  $Y$  выражает значение грамматических категорий  $Z$ ; 3) значение категорий  $Z$  допустимо для данного слова  $X$ . Таким образом, выполнимость морфологического отношения целиком сводится к выполнимости трех бинарных отношений  $l_1(X, Y)$ ,  $l_2(Y, Z)$ ,  $l_3(X, Z)$ , что справедливо не для всякого тернарного.

Языковая практика подсказывает нам, что результат проверки каждого из трех условий должен быть однозначным, а это влечет однозначность выполнимости морфологического отношения.

Поясним сказанное примером. Пусть задано слово *стол*, словоформа *столами* и следующие грамматические категории: падеж *творительный*, число *множественное*. Чтобы установить, состоят ли эти данные в морфологическом отношении, следует убедиться, что: 1) *столами* есть форма слова *стол*; 2) словоформа *столами* стоит в *именительном* падеже *множественного* числа; 3) для слова *стол* имеют смысл такие категории, как падеж и число, и такие их значения, как *творительный* и *множественное*. Так как все условия в данном случае выполняются, выполняется и морфологическое отношение.

Может показаться, что морфологическое отношение сводится только к выполнению двух первых требований. Но тогда надо согласиться, что тройка  $\langle \textit{тереть}, \textit{три}, \text{часть речи} — \textit{числительное} \rangle$  состоит в морфологическом отношении, так как: 1) *три* является формой глагола *тереть*; 2) *три* — *числительное*. Отношение, определенное подобным образом, не представляет интереса для математического описания языка.

Если бы удалось получить такое описание морфологического отношения, которое позволило эффективно определять неизвестные компоненты кортежа  $\langle X, Y, Z \rangle$  по заданным, возникла бы возможность решать задачи анализа, синтеза, нормализации изолированных словоформ. Задача синтеза требовала бы отыскания компоненты  $Y$  по заданным  $X$  и  $Z$ , задача анализа — отыскания  $Z$  по заданной компоненте  $Y$ , а нормализация заключалась бы в получении  $X$  по известной  $Y$ . Целью данной работы и является построение математической модели словоизменения имен существительных русского языка путем формального описания соответствующего морфологического отношения. Описание должно позволять достаточно эффективно отыскивать значения неизвестных компонент кортежа  $\langle X, Y, Z \rangle$  по заданным.

Основой для математического описания морфологии может послужить изучение процесса синтеза словоформ. Решить задачу синтеза — значит указать множество словоформ определенного слова, обладающих заданными значениями грамматических категорий. При этом значения грамматических категорий должны иметь смысл для данного слова. Совокупность всех троек  $\langle X, Y, Z \rangle$ , где  $X$  и  $Z$  — исходные данные для синтеза, а  $Y$  — одна из словоформ, полученных в результате синтеза, задает тернарное отношение. Из определения морфологического отношения непосредственно следует, что отношение, задаваемое синтезом, и является морфологическим. Необходимо заметить, что в основу математической модели может быть положена процедура анализа или нормализации. Выбор синтеза продиктован его наибольшей разработанностью [2].

Морфологическое отношение, как и всякое другое, целиком и полностью определяется своей характеристической функцией  $L_f(X, Y, Z)$ , которая задается следующим образом:

$$L_f(X, Y, Z) = \begin{cases} 1, & \text{если } L(X, Y, Z) \text{ выполняется;} \\ 0 & \text{в противном случае.} \end{cases}$$

Аргументы  $X$ ,  $Y$  и  $Z$  заданы на множествах  $S$ ,  $S$  и  $M$  соответственно. Функцию  $L_f(X, Y, Z)$  назовем морфологической функцией. Последняя представляет собой предикат, определенный на конечных предметных областях, и является более удобным объектом для описания, чем отношение. Вновь введенное понятие позволяет уточнить постановку задачи. А именно, под математической моделью словоизменения будем понимать определенным образом заданную морфологическую функцию. Цель дальнейшего исследования — разработка способа задания и отыскание вида функции для имен существительных субстантивного склонения.

Морфологическая функция дает возможность интерпретировать задачи анализа, синтеза, нормализации как решение уравнений типа  $L_f(X, Y, Z) = 1$ . Каждая из задач содержит в своем условии значения некоторых переменных. Например, синтез словоформы *родительного падежа, множественного числа слова стол* эквивалентен отысканию корня уравнения  $L_f(\text{стол}; Y; \text{множественное число, родительный падеж}) = 1$ . Анализ словоформы *волной* сводится к решению уравнения  $L_f(X; \text{волной}; Z) = 1$ . Если уравнение имеет более одного корня, это говорит о неоднозначном решении данной морфологической задачи. Например, корнями уравнения  $L_f(\text{тенор}; Y; \text{множественное число, именительный падеж}) = 1$  являются:  $Y_1 = \text{теноры}$  и  $Y_2 = \text{тенор'а}$ ; оба корня — формы именительного падежа множественного числа слова *тенор*. Если же уравнение не имеет корней, у данной морфологической задачи отсутствует решение; например, уравнение  $L_f(X; \text{куздре}; \text{единственное число, дательный падеж}) = 1$ , соответствующее задаче отыскания лексемы словоформы *куздре*, не имеет решения, так как в русском языке такой лексемы не существует.

Известно, что всякая словоформа целиком определяется своим буквенным составом и местом ударения. Существует линеализированный способ записи словоформ с ударением, а именно, символ ударения помещается в строке непосредственно перед ударной гласной буквой. Таким образом, словоформа представима в виде конечной цепочки символов некоторого алфавита  $A$ . В этот алфавит входят все русские буквы, знак дефиса, так как существуют слова с дефисом, и символ ударения. Введем множество переменных, общей областью изменения которых является алфавит  $A$ , и назовем их буквенными переменными. Название придиктовано тем, что элементы  $A$  в основном — буквы, однако в дальнейшем будем называть буквенной любую переменную со значениями из конечного множества символов.

Переменную  $Y$  можно подменить фиксированной совокупностью буквенных переменных  $y_1, y_2, \dots, y_n$ , причем значением переменной  $y_i$  служит  $i$ -я по счету буква словоформы. Противоречие между фиксированным количеством буквенных переменных и различным числом букв в словоформах русского языка можно

разрешить следующим образом. Во-первых, потребовать, чтобы число переменных  $n$  было не меньше числа букв в самой длинной словоформе, во-вторых, дополнить каждую словоформу справа особыми символами так, чтобы она состояла ровно из  $n$  символов. Это не единственный способ разрешить упомянутое противоречие, но он является приемлемым для наших целей.

В сфере словоизменения понятие слова имеет, по крайней мере, два аспекта. Во-первых, слово — единица словаря, текст, выражающий определенный смысл. С этой точки зрения для обозначения слов пригодны любые попарно различные символы, например, натуральные числа. Во-вторых, слово служит базой для образования своих форм. В самом деле, русские словоформы одной словоизменительной парадигмы имеют много общего, и было бы удобно, если бы это общее содержалось в обозначении слова. Существует лексикографическая традиция использовать для обозначения слова одну из его форм. Для существительных обычно это форма единственного числа именительного падежа. Возникающие случаи омонимии разрешаются с помощью цифры счета омонимии, например, 1 лев (животное) и 2 лев (денежная единица). Такой способ обозначения слов отвечает двум аспектам понятия слова, о которых говорилось выше, поэтому мы им воспользуемся.

В силу избранного способа обозначения слова переменную  $X$  можно заменить множеством переменных  $x_1, x_2, \dots, x_n$  подобно тому, как это было сделано с переменной  $Y$ . Здесь также  $i$ -я буква в строке символов, представляющей слово, служит значением переменной  $x_i$ . Помимо переменных  $x_1, x_2, \dots, x_n$  введем буквенную переменную  $x_0$ , означающую цифру счета омонимии. В числе ее возможных значений примем первые  $k$  членов натурального ряда, где  $k$  — максимально возможное число слов, которые имеют одинаковый буквенный состав и ударение, но выражают различный смысл. Каждому смыслу сопоставим свое значение переменной  $x_0$ , например для слова лев (животное)  $x_0 = 1$ , а для лев (денежная единица)  $x_0 = 2$ . Словам, не имеющим омонимов, припишем  $x_0 = 1$ .

Для того чтобы определить состав грамматических категорий создаваемой модели, сформулируем два требования к самой модели. Первое: описание предиката  $L_f(X, Y, Z)$  должно быть таково, чтобы задача синтеза имела, по возможности, единственное решение. Это приблизит функционирование модели в режиме синтеза к тому, как говорит или пишет человек, а он делает это однозначно. Второе: морфологическая модель должна естественным образом состыковываться с моделью синтаксического уровня обработки текста.

Чтобы удовлетворить первому требованию, в состав грамматических категорий  $Z$  следует включить помимо словоизменительных (для существительных — число и падеж) такие категории, которые «управляют» вариантностью образования форм. Это можно сделать, если удастся установить закономерность в образовании

вариантов. Например, известно, что у некоторых слов существуют современные и архаичные варианты форм, так *рукой* — современный вариант, *рукою* — архаичный, поэтому целесообразно включить в состав  $Z$  категорию со значениями: *современный*, *архаичный*. Иногда варианты форм выражают более тонкие различия смысла, например, *лес'а* и *лѣса* (обе формы в значении рыболовной снасти). В этом случае бывает трудно указать такую категорию, значения которой определяли бы выбор той или иной формы, и мы условимся считать результатом синтеза все множество вариантов формы слова.

Второе требование к морфологической модели побуждает нас включить в состав грамматических категорий  $Z$  род и одушевленность, определяющие согласование существительного с другими словами в предложении.

Для формального описания грамматической категории воспользуемся буквенной переменной с областью определения, состоящей из всевозможных значений данной категории.  $Z$  представим в виде множества буквенных переменных  $z_1, z_2, \dots, z_m$ , каждая из которых отражает одну грамматическую категорию существительного.

Сделанные замечания позволяют сформулировать итоговую постановку задачи. Последняя сводится к формальному описанию функции  $L_f(x_0, x_1, \dots, x_n, y_1, y_2, \dots, y_n, z_1, z_2, \dots, z_k)$ , все аргументы которой — суть буквенные переменные. Функция  $L_f$  должна служить моделью словоизменения русских имен существительных субстантивного склонения.

Список литературы: 1. Шабанов-Кушнарченко Ю. П. Применение метода нуля органа в лингвистике. — В кн: Проблемы бионики. Харьков, 1979, вып. 21, с. 109—112. 2. Зализняк А. А. Грамматический словарь русского языка. М., Русский язык, 1977. 880 с.

Поступила 26 марта 1979 г.