

М. Ф. БОНДАРЕНКО, канд. техн. наук, С. И. МАЛЕНЧЕНКО

О ВВЕДЕНИИ БУКВЕННЫХ ПЕРЕМЕННЫХ ДЛЯ МАТЕМАТИЧЕСКОЙ ЗАПИСИ РУССКИХ СЛОВ

Для того чтобы получить возможность моделирования морфологии русского языка на базе формального аппарата теории интеллекта [1, 2], необходимо прежде научиться записывать русские слова в виде уравнений алгебры конечных предикатов. Это, в свою очередь, требует введения набора буквенных переменных, принимающих в качестве своих значений различные русские слова. Наиболее простой способ введения искомого набора состоит в следующем. Вводится набор переменных $\langle x_1, x_2, \dots, x_n \rangle$, каждая из которых задана на русском алфавите. Число n выбираем с таким расчетом, чтобы оно совпало с числом букв самого длинного слова, указанного в орфографическом словаре русского языка.

Для математической записи какого-нибудь слова, например *стол*, полагаем $x_1 = \text{с}$, $x_2 = \text{т}$, $x_3 = \text{о}$, $x_4 = \text{л}$, $x_5 \div x_n = -$, где $-$ знак пробела. Описанный способ математической записи слов прост, однако он имеет существенный недостаток. Дело в том, что при такой записи полностью игнорируется внутренняя структура русских слов. Как известно, слово членится на отдельные морфы, например, *пере-дел-ыва-ть*, в роли которых выступают префиксы, корни, суффиксы и флексии. При описанном способе формального представления слов однотипные морфы окажутся на разных местах (например, корень *дел-* в словах *перedelывать* и *деловой*), а это в конечном итоге чрезвычайно усложнит и затруднит математическое моделирование морфологии русского языка.

Представляется гораздо более логичным поступить иначе: предусмотреть последовательность наборов переменных

$\langle x_{11}, x_{12}, \dots, x_{1n} \rangle, \langle x_{21}, x_{22}, \dots, x_{2n} \rangle, \dots, \langle x_{m1}, x_{m2}, \dots, x_{mn} \rangle$

с тем, чтобы каждым набором представлять тот или иной морф слова. В данной статье рассматривается вопрос о представлении подобным способом суффиксов и флексий. Представлении префиксов и корней здесь не затрагивается. Прежде всего необходимо решить вопрос о выборе чисел m и n . Число n определяется максимальной длиной суффикса или окончания. Суффиксы и флексии берем из перечней, указанных в книге [3, с. 755—760]. Самое длинное окончание имеет 3 буквы (например, *-ами*), самый длинный суффикс имеет 8 букв (*-истическ-*). Наибольшее число суффиксов в слове — пять (например, *вероятно-остро-н-ый*), окончаний — одно. Таким образом, можно принять $n=8, m=6$. Под числом m понимаем число букв в слове после его корня, причем имеются в виду только слова с одним корнем.

Теперь необходимо рациональным способом расположить буквы отдельно взятого i -го морфа в последовательности $\langle x_{i1}, x_{i2}, \dots, x_{in} \rangle$. Если нумеровать буквы морфа просто слева направо, то немедленно обнаруживается, что однотипные части морфа оказываются на различных местах (например, часть *ов* в морфах *ов, ова, ствова, изирова, овь, овск, овит*). Это свидетельствует о том, что отдельно взятый морф, как и все слово в целом, имеет свою собственную структуру, которую мы пока что не учли. Вскрывшееся обстоятельство вынуждает продолжить дробление слова и ввести понятие о части слова более мелкой, чем морф, но более крупной, чем отдельная буква слова. Назовем такие части морфа морфоидами (морфоподобными частями слова). Будем считать, что если морфоид содержит гласную букву, то она всегда стоит в нем на первом месте. Далее полагаем, что любой морфоид содержит не более одной гласной буквы. Однако мы будем допускать морфоиды, вообще не содержащие ни одной гласной буквы. Такие морфоиды могут получаться, например, в результате выпадения беглой гласной буквы.

При членении морфа на морфоиды полезно ввести понятие морфемоида, играющего такую же роль по отношению к морфоиду, как и морфема к морфу. Известно, что в морфемах наблюдается беглость гласных и чередование согласных. В связи с этим появляются различные графические варианты морфоида. *Морфемоидом* назовем совокупность всех возможных вариантов морфоида. Полагаем, что для записи отдельно взятого морфоида достаточно набора (x_1, x_2, x_3) , составленного из трех буквенных переменных. Первая переменная морфоида — гласная буква, если же гласная буква в морфоиде отсутствует, то в этом случае полагаем $x_1 = -$. Вторая и третья переменные

x_2 и x_3 используются для букв, следующих за гласной буквой. Это — согласные буквы, полугласная буква й и мягкий знак ь. Букву x_1 назовем *левой частью* морфоида, набор букв $\langle x_2, x_3 \rangle$ — его *правой частью*. Букву x_2 назовем *основной буквой* правой части морфоида, букву x_3 — *дополнительной*.

К числу морфоидов причисляем прежде всего все двухбуквенные суффиксальные и флексийные морфы, имеющие структуру $\langle \text{гласная, согласная} \rangle$, или $\langle \text{гласная, й} \rangle$. К ним относятся следующие морфы: об, ав, ев, ёв, ив, ов, аг, уг, ыг, ад, аж, ёж, из, оз, ой, ей, ёй, ий, ой, ый, ак, ик, ок, ук, ык, юк, як, ал, ёл, ол, ул, ам, ем, ём, им, ом, ум, ым, ям, ан, ен, ён, ин, он, ун, ар, ер, ир, ор, ур, яр, ас, ис, ус, ат, ет, ёт, ит, от, ут, ют, ят, ах, ех, ёх, их, ох, ух, ых, ях, ец, иц, аг, иг, уг, ыг, аш, иш, ош, уш, ыш, ащ, ищ, ущ. Первую букву этих морфов принимаем за x_1 , вторую — за x_2 , в качестве значения третьей буквы x_3 принимаем знак пробела —. К числу морфоидов причисляем также все трехбуквенные морфы, имеющие структуру $\langle \text{гласная, согласная, ь} \rangle$. К ним относятся морфы овь, ядь, ёжь, аль, ель, оль, ыль, ень, ынь, арь, ырь, оть, ешь, ёшь, ишь. Морфоидами считаем все однобуквенные и двухбуквенные морфы следующих структур $\langle \text{гласная} \rangle$, $\langle \text{согласная} \rangle$, $\langle \text{согласная, ь} \rangle$: а, б, в, д, е, ё, и, к, л, ль, н, нь, о, с, т, ть, у, х, ц, ч, щ, ы, ю, я. Здесь гласную берем в качестве значения первой буквы x_1 , согласную — в качестве второй буквы x_2 , мягкий знак — в качестве третьей буквы x_3 .

Остальные морфы расщепляем на морфоиды по следующему принципу: выделяем в них, где это возможно, части вида $\langle \text{гласная, согласная} \rangle$, $\langle \text{гласная, согласная, ь} \rangle$ или $\langle \text{гласная, й} \rangle$, например: ив-а, ов-а, ств-ов-а, из-ов-а. Если произведенное деление подтверждается наличием разветвления слов в этом месте при словообразовании, то выделенные части морфа принимаем за морфоиды. Например, для морфа ив-а имеем сле-

дующий случай разветвления: разбив $\begin{matrix} \text{ание} \\ \swarrow \\ \text{ка} \end{matrix}$, следовательно *ив*

принимаем за морфоид. Части морфа, полученные таким способом, считаем морфоидами только в тех случаях, когда они подходят хотя бы под одну из схем морфоидов, описанных выше, т. е. под схемы $\langle \text{гласная, согласная} \rangle$, $\langle \text{гласная, й} \rangle$, $\langle \text{гласная} \rangle$, $\langle \text{согласная} \rangle$, $\langle \text{согласная, ь} \rangle$.

Таким путем получаем следующие новые морфоиды: аб, еб, иб, яв, ж, яж, ез, ей, ен, ел, уль, онь, ень, ес, ыть, ац, еч, оч, ёш, юш, ош, ющ. Некоторые части морфов из числа полученных таким способом не подходят ни под одну из схем. Мы их расчленим на морфоиды следующим образом: с-т, ес-т, чл, инс, ург, о-ид, ар-ад, ар-н, ат-н, ль-н, ист, ль-щ, л-к, ер-к, ур-к, ек, в-ск, л-ез-ск, эз-ск, й-ск, н-ск., к-а, к-и, ш-к, изм, и-ан, л-ан, м-ан, м-ен, см-ен, з-нь, с-н, с-нь, ш-н, ль-н, -аль-н, и-ян, анс, есс, евт, м-ент, ан-ент, ем-ент, и-от, аст, ос-ть, вш-и, ог-о. В ре-

			е	ё	и	о	у	ы	э	ю	я
—		а	е	ё	и	о	у	ы		ю	я
б	б	аб	еб		иб	об					
в	в вш	ав	ев евт	ёв	ив	ов овь					яз
г		аг				ог	уг	ыг			
д	д	ад			ид						ядь
ж	ж	аж	еж	ёж ежь							яж
з	з		ез		из изн	оз			эз		
й	й	ай	ей	ёй	ий	ой	уй	ый			йя
к	к	ак	ек		ик	ок	ук	ык		юк	як
л	л ль	ал аль	ел ель	ёл		ол оль	ул уль	ыль			
м		ам	ем	ём	им	ом	ум	ым			ям
н	н нъ	ан анс	ен ень ент	ён	ин инг	он онь	ун	ынь			ян
р		ар арь	ер		ир	ор	ур ург	ырь			яр
с	с ск ст	ас аст	ес есс		ис ист		ус				
т	т ть ть	ат	ет	ёт	ит	от оть	ут	ыть		ют	ят
х	х	ах	ех	ёх	их	ох	ух	ых			ях
ц	ц	ац	ец		иц						
ч	ч	ач	еч		ич	оч	уч	ыч			
ш	ш	аш	еш ешь	ёш ёшь	иш ишь	ош	уш	ыш		юш	
щ	щ	ащ			ищ	ощ	ущ			ющ	

результате получаем следующие новые морфоиды: тв, инг, ург, ид, ист, ск, эз, й, изм, з, ст, анс, есс, евт, ент, аст, ог.

Полный перечень всех выделенных морфоидов приведен в таблице. Пользуясь этой таблицей, легко произвести однозначную разбивку послекоренной части любого слова на морфоиды. Система правил, управляющая разбивкой, следующая: 1) производим членение перед каждой гласной буквой слова; 2) вычленим вхождения буквосочетания *тв*; 3) если перед *к* стоит согласная (кроме *с*), производим перед *к* членение; 4) если перед *ск* стоит согласная, производим перед *ск* членение; 5) если перед *н* стоит согласная, производим перед *н* членение; 6) в последовательности букв, включающей буквосочетание *ств*, вводим членение после *с*. Приводим примеры разбивки послекоренной части слова на морфоиды согласно этой системе правил: аккурат-н-ень-к-ий, хвощ-ев-ид-н-ый, мобил-из-ов-ан-н-ый, соц-и-ал-ист-ич-еск-ий, так-ов-ск-ий, вид-им-ос-ть.

В результате подобной разбивки предельное число морфоидов достигает восьми, что лишь немногим больше, чем предельное число морфов (6). В связи с этим представляется неэкономным вводить наборы переменных для морфов, более целесообразно проводить формальную запись слова, расчлененного на морфоиды. В этом случае достаточно ввести 8 трехбуквенных наборов переменных. Таким образом, для записи послекоренной части слова (имеются в виду слова с одним корнем) всего требуется $m \cdot n = 8 \times 3 = 24$ буквенных переменных.

В заключение перечислим некоторые из задач, ждущих решения и являющихся естественным продолжением этой работы: инвентаризация морфоидов префиксальных и корневых морфов; разработка системы правил, управляющей процессом членения префиксальной и корневой части слова на морфоиды; математическое описание процесса членения слова на морфоиды; формирование системы семантических признаков, управляющих синтезом морфов; формирование морфемоидов (системы родственных морфоидов) для каждого морфоида; изучение и математическое описание процессов беглости и чередования букв в морфемоидах; изучение и математическое описание закономерностей сочетаемости соседних морфоидов в морфах и морфов в словах; изучение и математическое описание процесса синтеза морфоидов, морфов и слов по семантическим признакам.

Список литературы: 1. Шабанов-Кушнаренко Ю. П. Применение метода нуль-органа в лингвистике.— Проблемы бионики, 1978, вып. 21, с. 109—112. 2. Шабанов-Кушнаренко Ю. П. О теории интеллекта.— Проблемы бионики, 1979, вып. 22, с. 15—22. 3. Русская грамматика, т. I.— М.: Наука, 1980.— 783 с.

Поступила в редколлегию 09.03.81.