УДК 519.762.2

М. Ф. БОНДАРЕНКО., канд. техн. наук, *Н. В. ШАРОНОВА*

ЗАДАЧА ФРАГМЕНТАЦИИ СУФФИКСОВ ИМЕН СУЩЕСТВИТЕЛЬНЫХ

В пределах употребительной лексики словообразовательная система русского языка складывается из трех основных подсистем: корнетипы знаменательных слов, приставки, суффиксы. Из трех типов морфем складываются модели простых и сложных слов по определенным семантическим схемам. Смысловые ассоциации наиболее сложно организованы, поскольку значение слова не есть простая сумма значений составляющих его частей. Составление слова происходит по ступеням, одновременно присоединяются друг к другу только две части.

В русском языке около двухсот разных суффиксов, из них две трети составляют производные суффиксы, которые образовались путем наращивания звуков на исходные морфемы [1]. Наиболее разветвленную подсистему образуют именные суффиксы. Каждая часть речи имеет свой инвентарь суффиксов, самая разнообразная истема суффиксов у имени существительного.

Образованию суффиксальных основ русского языка свойственны элементы агглютинации. Простые суффиксы относительно немногочисленны, наиболее разнообразны цепочки суффиксов, из которых исторически возникали производные суффиксальные морфемы. Значимость осложненных, вторичных суффиксов определяется исходными морфемами. Следует отметить, что, наряду с очевидным разнообразием и многообразием суффиксов, в их составе часто встречаются одинаковые кусочки, исторически

восходящие к простым суффиксам, а также одинаковые (в смысле сочетания букв) наращения исходных морфем. Например, можно заметить, что в 170 различных (по грамматике [2]) суффиксах имен существительных часто встречаются около 30 таких кусочков, как *ец, ик, ек, ок, ан, ян* и т. п., которые сами по себе могут быть простыми суффиксами, а иногда встречаются и в качестве наращений на простые, неосложненные суффиксы. На основании этих фактов нам представляется целесообразным разделение суффиксов на такие кусочки, наиболее часто встречающиеся и являющиеся как бы «кирпичиками» при построении простых и сложных суффиксов. Такие кусочки будем в дальнейшем называть фрагментами суффиксов, полагая, что фрагмент суффикса несет в себе ту же семантику, что и суффикс, в который он входит, т. е. самостоятельным значением фрагмент суффикса обладает лишь тогда, когда он совпадает с простым суффиксом. Суффиксы могут состоять из одного и более фрагментов. Суффикс, состоящий из одного фрагмента, будем называть *однофрагментным*, в других случаях — *двух-*, *трех-* и т. д. *фрагментным* соответственно. Максимальное количество фрагментов в суффиксе имени существительного равно пяти.

Разбиение на фрагменты суффиксов производим следующим образом. Существует набор фрагментов-эталонов. Их значительно меньше общего количества суффиксов, что позволяет делать более экономичную формальную запись. Еще больше сокращает количество фрагментов то обстоятельство, что некоторые из них можно рассматривать как варианты одного и того же фрагмента. Например, фрагменты *ак-ач, як-яч, ок-оч* и т. п., то есть при определенных условиях фрагменты *ак, як, ок, ик* переходят в *ач, яч, оч, ич* (*бедняк — беднячка, дождик — дождичек, грибок — грибочек*).

В результате исследований, проведенных с помощью обратного словаря [3], выяснилось, что определенные фрагменты суффиксов чаще всего стоят на определенных местах в слове. Например, фрагменты с опорными согласными буквами *к, ц, ч, ш, л, г* могут находиться лишь в конце слова, перед окончанием, либо перед строго ограниченным количеством фрагментов суффиксов, таких как *-к(а), ок-ек-ёк.* Таких фрагментов большинство, и они составляют основной инвентарь простых суффиксов. Другие фрагменты суффиксов, такие как *ов* (*ев*), *ин, ен* и т. п., в большинстве случаев встречаются перед основными фрагментами и чаще всего являются теми наращениями, которые усложняют суффиксы. Разбиение суффиксов на фрагменты позволяет приблизительно втрое сократить количество деривационного материала, требующего формального представления. Кроме того, мы можем выделить процессы, протекающие на границах фрагментов суффиксов, т. е. непосредственно в самом сложном суффиксе, а также на стыках суффикса с основой и окончанием.

Для описания этих и других словообразовательных процессов нами будет использоваться алгебра конечных предикатов, описанная в [4, 5] и примененная при описании процессов словоизменения [6, 7].

В качестве примера, иллюстрирующего возможности применения алгебры конечных предикатов для описания процессов, происходящих в фрагментах суффиксов, опишем чередования гласных и согласных букв в фрагментах суффиксов с опорной буквой *к*. Всего таких фрагментов с их вариантами пять: *ак-як, ок-ек-ёк, ик, —к, ук-юк.*

Обозначим первую и вторую буквы фрагмента s_1 и s_2 соответственно. В таком случае областью определения этих переменных будут наборы гласных и согласных букв, которые принимают s_1 и s_2, т. е. s_1 может принимать значения *а, я, о, е, ё, и, —, у, ю*, s_2 — *к, ц, ч* (здесь учитываются чередования $к \to ч$ и $к \to ц$ при определенных условиях, которые будут описаны ниже). Цель наших математических построений — определение вида функций $s_1 = f(x_1, x_2, \ldots, x_m)$, $s_2 = \varphi(x_{m+1}, \ldots, x_n)$ зависимости первой и второй букв фрагмента суффикса от системы признаков $x_1, x_2, \ldots, x_m, \ldots, x_n$.

Рассмотрим подробнее эту зависимость и сами признаки. Следует подчеркнуть, что при описании различаются три процесса, протекающие на границах фрагментов суффиксов.

1. Чередования в последней букве основы перед первой буквой фрагмента.
2. Выбор гласной s_1.
3. Чередования согласной s_2.

В первом случае определяющим признаком является присоединение фрагмента с определенной согласной буквой. Например, при присоединении большинства фрагментов с согласной буквой происходят чередования $к \to ч, г \to ж, (н, х) \to ш$ (*друг—дружок—дружочек, карман—кармашек* и т. п.). При этом выделяются три класса чередований: $\tau_1 = t_1^{е} \vee t_1^{ж}$; $\tau_2 = t_1^{к} \vee t_1^{ч}$; $\tau_3 = (t_1^{н} \vee t_1^{х}) \vee t_1^{ш}$, где t_1 — последняя буква основы со значениями: *б, в, г, д, е, з, з', к, л, л', м, н, н', о, п, р, р', с, т, ч, ш, щ, х, ь* (буква со штрихом означает мягкую согласную). Других чередований в последней букве основы перед фрагментами суффиксов с опорной буквой *к* не происходит, поэтому в этом случае справедливо равенство $\tau_1 \vee \tau_2 \vee \tau_3 = 1$.

Во втором случае, когда происходит выбор гласной s_1, в области определения переменной s_1 выделяются следующие классы чередований: $\mu_1 = s_1^{а} \vee s_1^{я}$; $\mu_2 = s_1^{о} \vee s_1^{е} \vee s_1^{ё}$; $\mu_3 = s_1^{и} \vee s_1^{ю}$; $\mu_4 = s_1^{ы}$; $\mu_5 = \overline{s_1}$, причем в пределах данного класса фрагментов $\mu_1 \vee \mu_2 \vee \mu_3 \vee \mu_4 \vee \mu_5 = 1$. Выбор класса чередований определяется тем набором семантических значений суффиксов, в которые входит данный фрагмент. Внутри каждого класса чередований (т. е. внутри каждого подкласса значений первой буквы фрагмента)

записываем правило появления того или иного значения. Например, для $\mu_1 = s_1^а \vee s_1^я$ запишутся следующие уравнения алгебры конечных предикатов: $s_1^а \supset x_1^б \vee x_1^в \vee$

$$\vee x_1^д \vee x_1^ж \vee x_1^з \vee x_1^м \vee x_1^н \vee x_1^р \vee x_1^с \vee x_1^т \vee x_1^ч \vee x_1^ш \vee x_1^щ;$$

$$s_1^я \supset x_1^{в'} \vee x_1^{д'} \vee x_1^{з'} \vee x_1^л \vee x_1^{м'} \vee x_1^н \vee x_1^0 \vee x_1^{р'} \vee x_1^{с'} \vee x_1^{т'} \vee x_1^ь.$$

Здесь x_1 — признак последней буквы основы со значениями: *б, в, ж, з, з', л, л', м, н, н', о, п, р, р', с, т, т', ч, ш, щ, ь.* Эти уравнения описывают зависимость появления первой буквы фрагмента от последней буквы основы. Мы видим, что буква *я* появляется после мягких согласных *в', д', з', м', с', т', р'*, а также после букв *л, н, о, ь*. Примеры: *рыбак, чужак*, но *здоровяк, медяк, бедняк*, а также *кривляка, гуляка* и т. п.

Для класса чередований $\mu_2 = s_1^о \vee s_1^е \vee s_1^{ё}$ существенную роль приобретает признак x_2 — признак ударности фрагмента со значениями: *у — ударный, б — безударный*;

$$s_1^о \supset (x_1^б \vee x_1^в \vee x_1^{нс} \vee x_1^д \vee x_1^з \vee x_1^л \vee x_1^м \vee x_1^н \vee x_1^п \vee x_1^р \vee x_1^с \vee$$

$$\vee x_1^т \vee x_1^ч \vee x_1^ш) x_2^у; \; s_1^{ёе} \supset (x_1^{з'} \vee x_1^{л'} < x_1^{н'} \vee x_1^{р'} \vee x_1^{ё} \vee x_1^{т'}) x_2^у;$$

$$s_1^е \supset (x_1^ж \vee x_1^ч \vee x_1^{н'} \vee x_1^ш \vee x_1^е) x_2^б.$$

Здесь на выбор первой буквы фрагмента влияет не только признак последней буквы основы, но и ударение, например, *овраг — овражек, берег — бережок*. Таким образом, можно записать $s_1 = f(x_1, x_2, \mu)$.

В третьем случае, при описании чередований согласной буквы s_2, выделяем два класса чередований: $\varepsilon_1 = s_2^к \vee s_2^ч$; $\varepsilon_2 = s_2^к \vee s_2^ц$. На лингвистическом уровне такие чередования происходят в трех случаях: 1) при образовании имен существительных женского рода путем присоединения суффикса *-к (а)* (*рыбачка*); 2) при образовании уменьшительно-ласкательных форм путем присоединения суффиксов *-ек, -ок, -ик*: *рыбачок, беднячок*; 3) при образовании других частей речи (*рыбак — рыбачить, рыбацкий*).

Поскольку нами рассматривается словообразование только имен существительных, то интересовать нас будут лишь первые два случая. Можно сказать, что на формирование второй буквы s_2 фрагмента суффикса влияют следующие признаки: x_3 — наличие или отсутствие следующего фрагмента с опорной буквой *k*, x_4 — род со значениями: *м — мужской, ж — женский, с — средний*. Таким образом, зависимость второй буквы фрагмента суффикса от признаков запишется в следующем виде: $s_2 = \varphi(x_3, x_4, \varepsilon)$, где ε — классы чередований, которые в данном случае зависят от последующего фрагмента суффикса: $\varepsilon_1 \vee \varepsilon_2 = 1$.

Самое важное условие при решении данной задачи — соблюдение принципа однозначности, т. е. данный набор признаков должен однозначно определять фрагмент текста, хотя возможно,

что определенному фрагменту соответствует не один набор признаков. При такой постановке задачи считается, что буквы не связаны друг с другом непосредственно, они зависят лишь от признаков, которые можно описать системой уравнений алгебры конечных предикатов.

Список литературы: 1. Образование употребительных слов русского языка/Под ред. Л. Н. Засориной.— М.: Русский язык, 1979.—278 с. 2. Грамматика современного русского литературного языка/Под ред. Н. Ю. Шведовой.— М.: Наука, 1970.—767 с. 3. Обратный словарь русского языка.— М.: Сов. энциклопедия, 1974.—944 с. 4. *Шабанов-Кушнаренко Ю. П.* О конечных предикатах.— Проблемы бионики, 1980, вып. 24, с. 3—8. 5. *Шабанов-Кушнаренко Ю. П.* О теории интеллекта.— Проблемы бионики, 1979, вып. 22, с. 15—22. 6. Математическое описание процесса склонения имен прилагательных./*Ю. П. Шабанов-Кушнаренко, М. Ф. Бондаренко, В. М. Бондарев, З. Ю. Шабанова-Кушнаренко.*— Проблемы бионики, 1980, вып. 24, с. 22—27. 7. *Бондаренко М. Ф., Бондарев В. М.* О математическом описании словоизменения существительных.— Проблемы бионики, 1979, вып. 23, с. 98—104.

Поступила 25 февраля 1980 г.

УДК 519.762.2

М. Ф. БОНДАРЕНКО., канд. техн. наук, *Н. В. ШАРОНОВА*

О МАТЕМАТИЧЕСКОМ ОПИСАНИИ ПРОЦЕССОВ СЛОВООБРАЗОВАНИЯ

Для эффективного использования возможностей электронной вычислительной техники необходимо, чтобы ЭВМ могла воспринимать и обрабатывать информацию, представленную на естественном языке. В связи с этим возникла настоятельная необходимость в создании действующих моделей языка. Исследование языка и поиски способов формализации языковых структур необходимы не только для машинного перевода, но и для других более общих задач переработки информации с помощью ЭВМ.

В ряде практических задач (таких, как машинный перевод, машинное реферирование и аннотирование, перевод с естественного языка на формально-логические и т. п.) необходимо дать систему явно выраженных правил и записать эти правила с помощью специального аппарата так, чтобы ЭВМ могла воспринимать и обрабатывать вводимую языковую информацию.

Современное состояние лингвистических исследований таково, что «дальнейшее отсутствие практики, т. е. действующих систем машинного перевода (МП), тормозит не только его развитие, но и разработку самой лингвистической теории. Поэтому и необходимо рациональное сочетание теории с практикой и усиление прикладных лингвистических исследований в интересах реализации МП и решения других актуальных задач, связанных с автоматизацией информационных процессов в народном хозяйстве страны» [1, с. 49].

Одним из наименее исследованных с точки зрения формализации аспектов языка является словообразование.