

УДК 510.62

Н. В. ШАРОНОВА

**ФОРМАЛИЗАЦИЯ РАСПРЕДЕЛЕНИЯ МОРФОВ
В МОРФЕМЕ**

Чтобы успешно решить задачу математического описания процессов словообразования русского языка, для работы на семантическом уровне прежде всего надо решить задачу формализации на более низком уровне — на этапе формирования морфемы и распределения морфов в морфеме. Наша цель — дать формализованное представление распределения суффиксальных морфов в суффиксе.

В русском языке более пятисот различных суффиксальных морфов [1, с. 755—759]. Морфы объединяются в морфемы (суффиксы) не произвольно, а по определенным правилам. В одну морфему могут войти лишь такие морфы, которые обладают фонематической близостью при тождестве значения, характеризующиеся позиционным распределением либо вариативностью [1, с. 124—126]. Таким образом, морф — линейная единица языка, а морфема — нелинейная, обобщенная единица, представителями которой являются морфы. Например, морфема $\langle -ств/o/ \rangle$ состоит из морфов $-ств/o/$, $-есть/o/$, $тельств/o/$, $-овств/o/$. Распределение морфов внутри морфемы обусловлено чисто формальными признаками, идущими от основы: последняя буква основы или сочетание нескольких букв, ударение, для отглагольных существительных — иногда тип мотивирующего глагола и т. д.

В рамках постановки задачи описания функций морфем с использованием алгебры конечных предикатов [2] морфы объединяются в морфему при фиксации признаков, идущих от дальнего текста, и рассмотрении только признаков, идущих от самого слова, что не противоречит описанной выше ситуации.

При решении задачи формализации распределения морфов в морфеме мы выделяем следующие подзадачи: 1) рассматриваются и анализируются любые чередования в суффиксах, причем сами чередования рассматриваются шире, чем просто мена согласных или гласных; 2) группировкой морфов в морфемы достигается группировка сегментов в сегменты, т. е. множества сегментов, переходящих друг в друга при чередованиях, с одной и той же семантикой; 3) изучается, как окружение сегмента слева и справа управляет чередованием; 4) законы этого управления записываются уравнениями, связывающими буквы сегмента с буквами его окружения.

Рассмотрим конкретный пример описания распределения морфов в морфеме $\langle -аци|j| \rangle$. В эту морфему со значением процессуального признака входят следующие морфы (в орфографическом варианте): $-ация$ (*деградировать* — *деградация*), $-яция$ (*компилировать* — *компиляция*), $-иция$ (*репетировать* — *репетиция*), $-енция$ (*конкурировать* — *конкуренция*), $-ция$ (*абстрагировать* — *абстракция*), $-ия$ (*амнистировать* — *амнистия*), причем распределение морфов можно записать в виде следующих условий.

1. Морф $-яция$ следует после мягкого л'.
2. Морф $-иция$ выступает после мягких д', т', з'.
3. Морф $-енция$ следует после р', ц.
4. Морф $-ция$ идет после согласных к, б, п.
5. Морф $-ия$ может следовать после з', с', т', м', ц.

6. Морф *-ация* выступает после парнотвердых согласных, кроме л, а также после слабой гласной фонемы [а] и после гласной у.

Эти правила можно записать на языке алгебры конечных предикатов.

Введем следующие переменные: y_1 — последняя буква основы со значениями из множества букв русского алфавита; s_1, s_2, s_3, s_4, s_5 — буквы суффиксальных морфов морфемы <ация> (таблица).

Уравнение для морфа *-енция* будет записано в виде

$$y_1^{p'} \vee y_1^n \supset s_1^e s_2^n s_3^c s_4^n s_5^a; \quad (1)$$

для морфа *-ация* — в следующем виде:

$$y_1^n \vee y_1^b \vee y_1^f \vee y_1^s \vee y_1^t \vee y_1^a \vee y_1^c \vee y_1^s \vee y_1^k < y_1^r \vee y_1^x \vee y_1^m \vee y_1^n \vee y_1^p \vee y_1^a \vee y_1^i \vee y_1^y \supset s_1^a s_2^u s_3^u s_4^n s_5^a; \quad (2)$$

для морфа *-яция*:

$$y_1^n \supset s_1^a s_2^u s_3^u s_4^n s_5^a; \quad (3)$$

для морфа *-иция*:

$$y_1^n \vee y_1^r \vee y_1^z \supset s_1^a s_2^u s_3^u s_4^n s_5^a; \quad (4)$$

для морфов *-ция* и *-ия*:

$$y_1^k \vee y_1^b \vee y_1^f \supset s_1^u s_2^u s_3^u s_4^n s_5^a \quad (5); \quad y_1^z \vee y_1^c \vee y_1^t \vee y_1^m \vee y_1^n \supset s_1^u s_2^u s_3^u \times s_4^n s_5^a. \quad (6)$$

Однако одного признака последней буквы основы оказывается недостаточно для решения этих уравнений. Как видно из условий 1—6, множества букв здесь пересекающиеся. Например, после буквы ц может идти и морф *-енция*, и морф *-ия*. Для снятия неоднозначности левые части уравнений, содержащие одинаковые буквы индексов y_1 , необходимо домножить на признаки, свойственные лишь определенным морфам. В данном примере таким признаком может быть признак ударности-безударности суффиксального морфа. При одних и тех же значениях признака последней буквы основы (например, морфы *-иция* и *-ция* могут следовать после общих букв т', з') один морф всегда под ударением (*репетиция*, *транспозиция* и т. п.), а другой морф всегда безударен (*ревизия*, *амниция* и др.).

Введем переменные признаки: x_1 — признак твердости-мягкости буквы со значениями x_1^m — твердая, x_1^n — мягкая; x_2 — при-

знак ударности морфа со значениями xy — ударный, x_2^6 — безударный и перепишем уравнения (1)–(6) с использованием этих признаков:

$$y_1^p x_1^m \vee y_1^n x_2^y \supset s_1^e s_2^u s_3^u s_4^u s_5^a;$$

$$(y_1^k \vee y_1^n \vee y_1^6) x_1^t x_2^y \vee (y_1^f \vee y_1^a \vee y_1^t \vee y_1^u \vee y_1^c \vee y_1^3 \vee y_1^r \vee y_1^x \vee y_1^m \vee y_1^v \vee y_1^p) x_1^t \vee y_1^{a|} \vee y_1^y \supset s_1^a s_2^u s_3^u s_4^u s_5^a; \quad (7)$$

$$y_1^l x_1^m \supset s_1^a s_2^u s_3^u s_4^u s_5^a;$$

$$(y_1^a \vee y_1^t \vee y_1^3) x_1^m x_2^y \supset s_1^u s_2^u s_3^u s_4^u s_5^a;$$

$$(y_1^k \vee y_1^6 \vee y_1^n) x_1^t x_2^6 \supset s_1^u s_2^u s_3^u s_4^u s_5^a;$$

$$-((y_1^3 \vee y_1^c \vee y_1^t \vee y_1^m) x_1^m \vee y_1^n) x_2^6 \supset s_1^u s_2^u s_3^u s_4^u s_5^a.$$

Система уравнений (7) и есть формальное представление распределения морфов в морфеме <ация>.

Как видно из этого примера, у всех морфов одной морфемы имеется общая часть и совсем не обязательно ее каждый раз записывать. Если воспользоваться фрагментированной формой записи (3), то можно описать только чередования в первом или первых двух фрагментах, не рассматривая неизменяющихся фрагментов текста.

Приведенный здесь пример распределения морфов в морфеме — один из самых простых. Бывают случаи более сложные и запутанные, однако и для их формализации находятся признаки, позволяющие записывать довольно компактные уравнения. Для описания распределения наиболее сложных по составу морфем (например, в морфеме <к/а/ > входит 14 морфов: -к/а/, -ульк/а/, -юльк/а/, -анк/а/, -янк/а/, -ячк/а/, -авк/а/, -явк/а/, -ейк/а/, -отк/а/, -овк а/, -шк/а/, -чк/а/, -ушк/а/ — с разной степенью продуктивности) приходится составлять списки слов, включающих данные морфы, и формировать классы основ, к которым эти морфы присоединяются. Нами была проведена такая работа для суффиксов всех имен существительных русского языка, мотивированных глаголами, и получены уравнения, описывающие законы распределения суффиксальных морфов внутри этих суффиксов (морфем). Таких морфем для имен существительных со значением «носитель процессуального признака» оказалось 81 (морфов — 133), для имен существительных со значением «отвлеченный процессуальный признак» — 46 (морфов — 83), причем в это число входит значительное количество уникальных суффиксов, встречающихся в 1–2 словах, а также суффиксов, принимающих участие в образовании небольших (из двух-трех слов) групп существительных. В этих случаях мотивирующие основы задаются списками, включающими иногда только одно слово.

Приведем несколько примеров формального описания распределения морфов в морфеме.

Пример 1. Рассмотрим морфему <т-ель>, включающую морфы -тель, -итель. Существительные этого словообразовательного типа называют лицо или предмет, производящий действие или предназначенный для осуществления действия, названного мотивирующим словом. Если морфы представить в сегментированной форме, т. е. записать: <тель>: $_t_ / \text{ель} \vee \text{ит} _ / \text{ель}$, то можно задачу свести к описанию чередований в первом сегменте $s_1 s_2 s_3$, где $s_1 = \text{и} \vee _$, $s_2 = \text{т}$, $s_3 = _$.

Запишем правила появления морфов.

1. Морф -тель выступает после гласной (*испыта-ть* → *испытатель*, *храни-ть* → *храни-тель*, *владе-ть* → *владе-тель*, *жи-ть* → *жи-тель*).

2. Морф -итель выступает после согласной в следующих случаях:

а) мотивирующий глагол X класса на -е (причем -е отсекается) и VI класса на согласную (*смотре-ть* → *смотр-итель*, *спас-ти* → *спас-итель*);

б) в случае, когда мотивирующая основа на согласную совпадает с основой настоящего времени глагола (*знать*, *гон-ю* → *гон-итель*; *двигать*, *движ-у* → *движ-итель*);

в) в отдельных случаях: *сказ-ыва-ть* → *сказ-итель* (-ыва- отсекается), *повел-ева-ть* → *повел-итель* (-ева- отсекается), *блюст-и* → *блюст-итель*, *влад-еть* → *власт-итель* (редкие наращения основы за счет финали).

Если условия а, б, в обозначить предикатами, τ_1 , τ_2 , τ_3 , принимающими значение 1 в случае, когда соответствующее условие выполняется, и 0 — в противном случае, то правила появления морфов можно записать следующим образом:

$$\text{Гл}(y_1) \supset s_1^{\text{и}}; \text{Сг}(y_1) (\tau_1 \vee \tau_2 \vee \tau_3) \supset s_1^{\text{т}}$$

где y_1 — последняя буква основы; $\text{Гл}(y_1)$ и $\text{Сг}(y_1)$ — именные предикаты, обозначающие

$$\text{Гл}(y_1) = \begin{cases} 1, & \text{если } y_1 \text{ — гласная буква,} \\ 0 & \text{в противном случае.} \end{cases}$$

$$\text{Сг}(y_1) = \begin{cases} 1, & \text{если } y_1 \text{ — согласная буква,} \\ 0 & \text{в противном случае.} \end{cases}$$

Предикаты τ_1 , τ_2 , τ_3 можно попытаться выразить более точно, введя следующие признаки:

1) класс глагола $T = t_1 \vee t_2$, где t_1 — X класс на -е или VI класс на согласную, t_2 — любой другой класс;

2) признак глагольной основы $R = r_1 \vee r_2$, где r_1 — основа совпадает с основой настоящего времени глагола, r_2 — не совпадает;

3) признак, позволяющий выделить отдельные случаи редких наращений основы и отдельных образований, — выделение таких слов в отдельный список и сравнение с этим списком: Q — список исключений: $Q\{\text{сказ-}, \text{повел-}, \text{блуст-}, \text{власт-}\}$.

С учетом введенных обозначений можно записать условия распределения морфов в морфеме $\langle\text{-тель}\rangle$:

$$\Gamma л (y_1) \supset s_1^{\pi}; \text{Сг} (y_1) (t_1 \vee r_1 \vee Q) \supset s_1^{\pi}. \quad (8)$$

Кроме того, может оказаться полезным при решении задачи синтеза то обстоятельство, что морф -итель всегда под ударением, а морф -тель безударен. Далее, когда говорится, что морф -тель может идти после гласной, то это не значит, что он идет после любой гласной. То же самое — в случае с согласными. Можно ввести дополнительные ограничения, например, первое уравнение системы (8) заменить уравнением

$$(y_1^{\pi} \vee y_1^{\epsilon} \vee y_1^{\eta} \vee y_1^{\theta}) x_r^{\epsilon} \supset s_1^{\pi}.$$

Эти ограничения вызваны словарной реализацией.

Пример 2. Рассмотрим морфему $\langle\text{-ец}\rangle$, состоящую из морфов -ец , -нец , -енец , -омец . Существительные с этим суффиксом называют лицо или предмет, характеризующийся действием, названным мотивирующим словом (*льстец*, *продавец*, *выходец*, *сорванец*, *швец*). Правила распределения морфов следующие: морф -ец следует после отсекаемой гласной инфинитивной основы, морф -нец — после неотсекаемой гласной, морф -енец имеется в единственном слове *беженец*, морф -омец — в одном слове *питомец*. В словах *гон-ец*, *жн-ец*, *остан-ец* выступает глагольная основа настоящего времени. В словах *шв-ец*, *пев-ец*, *бо-ец*, *чт-ец* нерегулярные соотношения основ.

Запишем: $\langle\text{-ец}\rangle$: $\text{ец} _ \vee _ \text{н} _ / \text{ец} _ \vee \text{е.л} _ / \text{ец} _ \vee \text{ом} _ / \text{ец}$. Чередования происходят лишь в первом сегменте $s_1 s_2 s_3$, второй сегмент не меняется. Уравнения, описывающие законы этих изменений, следующие:

$$\text{Сг} (y_1) \vee \text{Сг} (y_1) (r_1 \vee Q_1) \vee \Gamma л (y_1) Q_1 \supset s_1^{\pi} s_2^{\pi} s_3^{\pi}; \Gamma л (y_1) \supset s_1^{\pi} s_2^{\pi} s_3^{\pi} \\ Q_2 \supset s_1^{\epsilon} s_2^{\eta} s_3^{\eta}; Q_3 \supset s_1^{\theta} s_2^{\eta} s_3^{\eta}, \quad (9)$$

где Q_1 , Q_2 , Q_3 — списки нерегулярных и единичных основ (исключений): $Q_1\{\text{шв-}, \text{пев-}, \text{бо-}, \text{чт-}\}$, $Q_2\{\text{беж-}\}$, $Q_3\{\text{пит-}\}$.

Система уравнений (9) не только описывает распределение морфов в морфеме, но и позволяет выделить чередующиеся в пределах одной семантики сегменты, т. е. объединить их в сегменту. Это множество сегментов, переходящих друг в друга при чередованиях, с одной и той же семантикой.

Список литературы: 1. *Русская грамматика*. Т. 1.— М.: Наука, 1980.— 759 с.
2. *Шабанов-Кушнаренок Ю. П.* Об универсальной алгебре конечных предикатов.— АСУ и приборы автоматики, 1980, вып. 55, с. 69—74. 3. *Бондаренко М. Ф., Шаронова Н. В.* Моделирование фрагментированных суффиксов имен существительных.— Деп. ВИНТИ, 1981, № 964.— 60 с.

Поступила в редколлегию 25.12.81.