

УДК 801.55.001.57

Е. А. СОЛОВЬЕВА, канд. техн. наук, О. Т. БАНАРЬ

**ИСПОЛЬЗОВАНИЕ АЛГЕБРЫ КОНЕЧНЫХ ПРЕДИКАТОВ ДЛЯ
АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ГЛАГОЛЬНЫХ ФОРМ**

В работе делается попытка формально описать процессы обработки глагольных форм на языке алгебры конечных предикатов предложенном Ю. П. Шабановым-Кушнарено [1—3].

Предикативные глагольные формы характеризуются словоизменительными категориями — наклонения, времени, числа, лица и рода [4]. Например, глагольная форма *читал* имеет признаки изъявительного наклонения прошедшего времени единственного

числа мужского рода, а форма *вычислите* — признаки повелительного наклонения множественного числа 2-го рода.

Введем переменные: z_1 — наклонение со значениями: и — изъявительное, п — повелительное; z_2 — время со значениями: н — непрошедшее, п — прошедшее, б — нейтральное время; z_3 — число со значениями: е — единственное, м — множественное; z_4 — лицо со значениями: п — первое, в — второе, т — третье, б — нейтральное; z_5 — род со значениями: м — мужской, ж — женский, с — средний, б — нейтральный.

Значения «б» переменных z_2, z_4, z_5 введены в соответствии с грамматикой русского языка (например: «Родовые формы лишены показателей значения лица» [4, с. 366]; формы прошедшего времени не имеют показателей лица и являются неличными» [4, с. 362]) и вследствие того, что они позволяют значительно упростить запись большого числа уравнений, моделирующих процессы обгати глаголов.

Для рассмотренных переменных справедливы уравнения:

$$z_1^H \vee z_1^P; z_2^H \vee z_2^P \vee z_2^B; z_3^E \vee z_3^M; \\ z_4^P \vee z_4^B \vee z_4^T \vee z_4^B; z_5^M \vee z_5^J \vee z_5^C \vee z_5^B.$$

Сочетаемость значений введенных переменных для глагольных форм расширенной парадигмы опишем уравнением

$$z_1^H (z_2^H (z_3^E \vee z_3^M) (z_4^P \vee z_4^B \vee z_4^T) z_5^B) \vee z_2^P z_4^B (z_3^E (z_5^M \vee z_5^J \vee z_5^C) \vee \\ \vee z_3^M z_5^B) \vee z_1^P z_2^B z_5^B (z_3^M z_4^P \vee (z_3^E \vee z_3^M) z_4^B).$$

Введем форму слова $Y = \langle y_1, y_2, \dots, y_m \rangle$, где y_1, y_2, \dots, y_m — соответственно 1, 2, ..., m -я буквы формы слова (принимаем $m = 50$). Введем также основу слова $\xi = \langle x_1, x_2, \dots, x_n \rangle$, окончание $E = \langle e_1, e_2, \dots, e_e \rangle$, постфикс $H = \langle h_1, h_2, \dots, h_t \rangle$, частицу $K = \langle k_1, k_2, \dots, k_k \rangle$, префикс $P = \langle p_1, p_2, \dots, p_p \rangle$, суффикс $S = \langle s_1, s_2, \dots, s_s \rangle$, корень $B = \langle b_1, b_2, \dots, b_e \rangle$, в которых буквы морфов расположены в порядке возрастания их номеров. Число элементов в кортежах окончания, постфикса, частицы, префикса и суффикса выбираем в соответствии с нормами грамматики [4] настолько большим, чтобы в них можно было записать любой глагольный морф: $e = 3$; $t = 4$; $k = 3$; $p = 8$; $s = 9$; $n = m - 9 = 41$; $r = n - 15 = 26$. Форму, основу и другие морфы записываем, начиная с первой левой позиции, а позиции, оставшиеся неиспользованными, заполняем знаками пробела. Например, в окончании $-ym$ $e_1 = y, e_2 = m, e_3 = \square$; в частице $-ка$ $h_1 = -, h_2 = к, h_3 = а$.

Перейдем к формальному описанию известных грамматических правил и выявленных в результате проведенных исследований закономерностей обработки глагольных форм. Слово (здесь и далее имеется в виду синтетическая глагольная форма расширенной

парадигмы) в 1-м лице единственного числа непрошедшего времени изъявительного наклонения оканчивается на \tilde{y} , $\tilde{ю}^*$.

Постфикс *-ся (-сь)* не влияет на рассматриваемые в работе значения грамматических категорий, поэтому удобно ввести выражения:

$$A_1 = h_1^c h_2^a \vee h_1^- h_2^-; A_2 = h_1^c h_2^b \vee h_1^- h_2^-.$$

В уравнениях (1) — (3) учтены глаголы *есть*, *дать* и их производные, имеющие в формах единственного числа особые флексии: 1-го лица $\tilde{м}$, 2-го лица $\tilde{шь}$, 3-го лица $\tilde{т}$, которые присоединяются к гласной основе [4, с. 412]:

$$z_2^h z_3^e z_4^п = (e_1^y \vee e_1^ю \vee (b_1^d b_2^a \vee b_1^e) e_1^м) A_2. \quad (1)$$

Слово во 2-м лице единственного числа непрошедшего времени оканчивается на $\tilde{ешь}$, $\tilde{ишь}$:

$$z_2^h z_3^e z_4^a = (e_1^e e_2^ш e_3^b \vee e_1^и e_2^ш e_3^b \vee (b_1^e \vee b_1^d b_2^a) e_1^ш e_2^b) A_1. \quad (2)$$

Слово в 3-м лице единственного числа непрошедшего времени оканчивается на $\tilde{ет}$, $\tilde{ит}$:

$$z_2^h z_3^e z_4^т = (e_1^e e_2^т \vee e_1^и e_2^т \vee (b_1^e \vee b_1^d b_2^a) e_1^e e_2^т) A_1. \quad (3)$$

Слово в 1-м лице множественного числа непрошедшего времени оканчивается на $\tilde{ем}$, $\tilde{им}$:

$$z_2^h z_3^м z_4^п = (e_1^e e_2^м \vee e_1^и e_2^м) A_1.$$

Слово во 2-м лице множественного числа непрошедшего времени оканчивается на $\tilde{ете}$, $\tilde{ите}$:

$$z_2^h z_3^м z_4^a = (e_1^e e_2^e e_3^e \vee e_1^и e_2^e e_3^e) A_2.$$

Слово в 3-м лице множественного числа непрошедшего времени оканчивается на $\tilde{ат}$, $\tilde{ят}$, $\tilde{ут}$, $\tilde{ют}$:

$$z_2^h z_3^м z_4^т = (e_1^a e_2^т \vee e_1^и e_2^т \vee e_1^y e_2^т \vee e_1^ю e_2^т) A_1.$$

Первая буква основы совпадает с c_{11-l} , вторая — с c_{12-l} , и т. д., последняя — с c_{10} , где l — длина основы, изменяющаяся в пределах от 1 до 10. Слово в мужском роде единственного числа прошедшего времени изъявительного наклонения имеет

* Знак \sim введен для упрощения записи. Он означает, что имеется в виду не только проведенный признак, но и полученный из него приписыванием постфикса *-ся*, *(-сь)* за буквой, отмеченной сверху волнистой чертой. Например, в данном случае \tilde{y} , $\tilde{ю}$ обозначают формальные признаки *-у*, *-усь*, *-ю*, *-юсь*.

суффикс \tilde{l} или основу на \tilde{c} , $\tilde{з}$, $\tilde{б}$, \tilde{n} , \tilde{x} , $\tilde{р}$, $\tilde{к}$, $\tilde{г}$ (кроме слова *ляг* и его производных) при нулевой флексии:

$$z_5^M = (S_1^l \vee C_{10}^c \vee C_{10}^з \vee C_{10}^б \vee C_{10}^n \vee C_{10}^x \vee C_{10}^p \vee C_{10}^k \vee C_{10}^g \vee C_8^{-l} C_9^{-n} C_{10}^{-r}) \tilde{l}_1^- A_1.$$

Слово в женском роде единственного числа прошедшего времени изъявительного наклонения имеет суффикс \tilde{l} с флексией a :

$$z_5^K = S_1^l e_1^a A_2.$$

Слово в среднем роде единственного числа прошедшего времени изъявительного наклонения имеет суффикс \tilde{l} с флексией o :

$$z_5^o = S_1^l e_1^o A_2.$$

Слово прошедшего времени изъявительного наклонения во множественном числе имеет суффикс \tilde{l} с флексией $-u$:

$$z_2^p z_3^M = S_1^l e_1^u A_2.$$

Слово в 1-м лице множественного числа повелительного наклонения оканчивается на $\tilde{итс}$, $\tilde{ем-ка}$, $\tilde{им-ка}$, $\tilde{ите-ка}$:

$$z_1^p z_3^M z_4^p = e_1^M h_1^T h_2^e A_3 \vee ((e_1^e e_2^M \vee e_1^e e_2^M) A_1 \vee e_1^M h_1^T h_2^e A_3) k_1^- k_2^k k_3^a,$$

где

$$A_3 = h_3^c h_4^b \vee h_3^m h_4^-.$$

Во 2-м лице единственного числа повелительного наклонения слово оканчивается на $\tilde{-u}^*$ или основа оканчивается на $\tilde{-b}$ (с нулевой флексией), $\tilde{-u}$ или $\tilde{-ляг}$:

$$z_1^p z_3^e z_4^b = ((e_1^M \vee C_{10}^b e_1^m \vee C_{10}^n) A_1 \vee C_8^l C_9^r C_{10}^p A_1) A_4,$$

где

$$A_4 = k_1^- k_2^k k_3^a \vee k_1^- k_2^- k_3^m.$$

Слово стоит во 2-м лице множественного числа повелительного наклонения, если оно имеет постфикс $\tilde{-те}$ и 2-я буква флексии не m :

$$z_1^p z_3^M z_4^b = \tilde{l}_2^M h_1^T h_2^e A_4.$$

Приведенные уравнения описывают автоматическую классификацию глагольных форм. Процессы выделения основ, окончаний

* Признак \tilde{u} означает $\tilde{-u}$, $\tilde{-u-ка}$ или $\tilde{-u}$, $\tilde{-ись}$, $\tilde{-u-ка}$, $\tilde{-ись-ка}$.

и других морфов моделируются на базе информации, полученной в работе [5]. Полученные уравнения описывают ряд соотношений между грамматическими и формальными признаками глагольных форм при условии, что структура слов известна. Информация о структуре слова облегчает описание (например, если буква $л$ перед окончанием $-и$ является в слове суффиксом, то оно относится к изъявительному наклонению, если же частью основы — к повелительному), но эта информация должна содержаться в других уравнениях. Для определения структуры словоформы необходимы уравнения, связывающие формальные признаки (буквы) и их роли в слове (окончание, суффикс и т. д.), а также уравнения для формирования морфов. Приведем примеры таких уравнений.

Основа слова по крайней мере не короче формы слова:

$$Y_i^- \supset X_i^- . \quad (4)$$

Форма слова не более, чем на 9 букв длиннее основы слова:

$$X_i^- \supset Y_{i+9}^- . \quad (5)$$

Последняя буква основы совпадает с C_{10} предпоследняя — с C_9 , ..., 1-я — с C_{11-l} :

$$\overline{X_i^-} X_{i+1}^- = \bigwedge_{j=1}^{i-l} (X_{l-j+1} = C_{10-j+1}) \bigwedge_{j=1}^{10-l} C_j^- \bigwedge_{j=l+1}^{50} X_j^- . \quad (6)$$

Последняя отличная от пробела буква машинной словоформы, совпадающая с y , $ю$, a , o , $и$ является 1-й буквой окончания:

$$y_i^- (y_{i-1}^y \vee y_{i-1}^o \vee y_{i-1}^a \vee y_{i-1}^o \vee y_{i-1}^и) = (y_{i-1} = e_1) e_2^- e_3^- . \quad (7)$$

Две последние буквы слова: $им$ или последняя $т$ и предпоследняя y , $ю$, a , $я$, e или $и$ — являются его окончанием:

$$y_i^- (y_{i-2}^y \vee y_{i-2}^o \vee y_{i-2}^a \vee y_{i-2}^e \vee y_{i-2}^и) y_{i-1}^т \vee y_{i-2}^и y_{i-1}^м) = \\ = (y_{i-2} = e_1) (y_{i-1} = e_2) e_3^- . \quad (8)$$

Три последние отличные от пробела буквы словоформы, совпадающие с $-ете$, $-ишь$, образуют окончания:

$$y_i^- (y_{i-3}^e y_{i-2}^т y_{i-1}^е \vee y_{i-3}^и y_{i-2}^ш y_{i-1}^ь) = \\ = (y_{i-3} = e_1) (y_{i-2} = e_2) \wedge (y_{i-1} = e_3) . \quad (9)$$

Две последние отличные от пробела буквы слова $-ся$ или $-ти$ образуют постфикс:

$$y_i^- (y_{i-2}^o y_{i-1}^т \vee y_{i-2}^т y_{i-1}^е) = (h_1 = y_{i-2}) (h_2 = y_{i-1}) h_3^- h_4^- . \quad (10)$$

Три последние отличные от пробела символы машинной словоформы $ка$ образуют частицу $-ка$:

$$y_{i-3}^к y_{i-2}^а y_{i-1}^а = k_1^- k_2^к k_3^а . \quad (11)$$

Четыре последние отличные от пробела буквы слова *-теть* образуют постфикс:

$$y_i (y_{i-4}^t y_{i-3}^e y_{i-2}^c y_{i-1}^b) = h_1^t h_2^e h_3^c h_4^b. \quad (12)$$

В словах, оканчивающихся на *-ка*, стоящие перед *-ка* буквы *те* и *ся* образуют постфикс:

$$\begin{aligned} \bar{y}_i \bar{y}_{i-3}^k \bar{y}_{i-2}^a \bar{y}_{i-1}^r (y_{i-5}^t y_{i-4}^e \vee y_{i-5}^c y_{i-4}^a) = \\ = (h_1 = y_{i-5}) (h_2 = y_{i-4}) h_3 \bar{h}_4. \end{aligned} \quad (13)$$

В словах, оканчивающихся на *-л*, или на *-лся*, *л* образует суффикс:

$$y_{i-3}^l (y_{i-2}^r \vee y_{i-2}^c y_{i-1}^a) y_i^r = S_1^r, S_2^r, \dots, S_9^r. \quad (14)$$

Буква *л* образует суффикс в словах, оканчивающихся на *-ла*, *-ло*, *-лась*, *-лось*:

$$\begin{aligned} \bar{y}_i (\bar{y}_{i-2}^l (\bar{y}_{i-1}^a \vee \bar{y}_{i-1}^o) \vee \bar{y}_{i-4}^l (\bar{y}_{i-3}^a \vee \bar{y}_{i-3}^o) \bar{y}_{i-2}^c \bar{y}_{i-1}^b) = \\ = S_1^l, S_2^l, S_3^l, \dots, S_9^l. \end{aligned} \quad (15)$$

Для формализации процессов обработки омографичных словоформ необходима дополнительная информация, которая для глагольных форм получена в работе [5]. Запишем примеры уравнений, учитывающих полную и частичную омографию.

В словах *трусь* и *вытрусь* *-сь* может быть постфиксом или частью основы:

$$\begin{aligned} (\bar{y}_{i-6}^r \vee \bar{y}_{i-8}^b \bar{y}_{i-7}^m \bar{y}_{i-6}^n) \bar{y}_{i-5}^t \bar{y}_{i-4}^p \bar{y}_{i-3}^y \bar{y}_{i-2}^c \bar{y}_{i-1}^b \bar{y}_i^r = \\ = h_1^c h_2^b h_3^r h_4^r \vee C_9^c C_{10}^b. \end{aligned} \quad (16)$$

В словах *завись*, *обезопась*, *гундось*, *тулумбась*, *чудесь*, *торось*, *весь*, *высь*, *брось*, *квась*, *крась* и их производных две последние буквы *сь* являются частью основы:

$$\begin{aligned} \bar{y}_i \bar{y}_{i-2}^c \bar{y}_{i-1}^b (\bar{y}_{i-6}^a \bar{y}_{i-5}^a \bar{y}_{i-4}^b \bar{y}_{i-3}^n \vee \bar{y}_{i-9}^o \bar{y}_{i-8}^b \bar{y}_{i-7}^z \bar{y}_{i-6}^o \bar{y}_{i-5}^o \wedge \\ \wedge \bar{y}_{i-4}^p \bar{y}_{i-3}^z \vee \bar{y}_{i-7}^r \bar{y}_{i-6}^y \bar{y}_{i-5}^n \bar{y}_{i-4}^l \bar{y}_{i-3}^o \vee \bar{y}_{i-9}^t \bar{y}_{i-8}^y \bar{y}_{i-7}^l \bar{y}_{i-6}^y \bar{y}_{i-5}^m \wedge \\ \wedge \bar{y}_{i-4}^b \bar{y}_{i-3}^z \vee \bar{y}_{i-6}^y \bar{y}_{i-5}^l \bar{y}_{i-4}^e \bar{y}_{i-3}^z \vee \bar{y}_{i-6}^o \bar{y}_{i-5}^p \bar{y}_{i-4}^o \bar{y}_{i-3}^z \vee \\ \vee \bar{y}_{i-4}^b \bar{y}_{i-3}^z \vee \bar{y}_{i-4}^m \bar{y}_{i-3}^z \vee \bar{y}_{i-5}^b \bar{y}_{i-4}^p \bar{y}_{i-3}^o \vee \bar{y}_{i-5}^k \bar{y}_{i-4}^y \bar{y}_{i-3}^z \vee \\ \vee \bar{y}_{i-5}^k \bar{y}_{i-4}^p \bar{y}_{i-3}^a) \supset C_9^c C_{10}^b. \end{aligned} \quad (17)$$

В уравнениях (4)–(17) *i* меняется от 3 до 50, *l* — от 1 до 10. Остальные уравнения имеют аналогичный вид и получены на основании исследований в области процессов классификации [4, 5] и синтеза глагольных форм.

С помощью методики Ю. П. Шабанова-Кушнарченко разработан алгоритм решения таких уравнений, который реализуется на ЭЦВМ.

Список литературы: 1. Шабанов-Кушнарченко Ю. П. О теории интеллекта.— В кн.: Проблемы бионики. Харьков, 1977, вып. 22, с. 15—22. 2. Шабанов-Кушнарченко Ю. П. Об алгебре конечных предикатов.— В кн.: АСУ и приборы автоматики. Харьков, 1977, вып. 52, с. 21—28. 3. Шабанов-Кушнарченко Ю. П. Математическая модель склонения полных непряжательных имен прилагательных. Сб. трудов / Науч.-техн. информация. Серия 2, М., ВИНТИ, 1978, № 10, с. 20—25. 4. Грамматика современного русского литературного языка. М., Наука, 1970. 767 с. 5. Соловьева Е. А. Исследование и моделирование процессов морфологического классифицирования глагольных форм. Сб. трудов / Науч.-техн. информация. Серия 2, М., ВИНТИ, 1976, № 8, с. 34—38.

Поступила 1 апреля 1978 г.