

Е. А. СОЛОВЬЕВА, канд. техн. наук, О. А. ЖУКОВ

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ГРАММАТИЧЕСКОЙ ОБРАБОТКИ ГЛАГОЛОВ ТИПА 13

В работе исследуются процессы грамматической обработки глагольных форм русского языка и строится их действующая модель с применением алгебры конечных предикатов Ю. П. Шабанова-Кушнаренко [1—4].

Рассмотрим морфологическое отношение $L(X, Y, Z)$, где X — основная форма слова, Y — словоформа, Z — набор грамматических признаков. $L(X, Y, Z) = 1$, если форма Y образована от исходной формы X и признаков в соответствии с нормами грамматики и $L(X, Y, Z) = 0$ — в противном случае. Например, $L(\text{читать}, \text{читаю}, \text{наст. время, ед. число} \dots \text{1-е лицо}) = 1$, а $L(\text{читать}, \text{читал}, \text{наст. время, ед. число} \dots \text{1-е лицо}) = 0$.

Для автоматической обработки словоформ целесообразно построить модель отношения L в виде системы уравнений 1, 2, ..., алгебры конечных предикатов. Применяя действующую программу решения уравнений, получим, в зависимости от исходных данных решение задачи автоматического анализа ($Y \rightarrow Z$), синтеза ($X \rightarrow Z \rightarrow Y$), нормализации ($Y \rightarrow X$) и др. Построим модель на основании работ [5, 6].

Математическое описание отдельных фрагментов обработки прилагательных и глаголов рассмотрено в статьях [7, 8]. В данной работе предложена модель морфологического отношения на примере глаголов 13-го типа. Известно [5], что к таким глаголам относятся глаголы *давать* (и производные) и глаголы на *знавать*, *ставать*. Тип спряжения выбран произвольно, однако он отражает основные закономерности обработки глагольных форм.

Введем форму слова $Y = \langle y_1, y_2, \dots, y_n \rangle$, где y_1, y_2, \dots соответственно 1, 2, ..., n -я буквы словоформы; $n = 30$ достаточно для размещения любой глагольной формы. Введем основную (каноническую) форму слова $X = \langle x_1, x_2, \dots, x_{20} \rangle$, окончание $E = \langle e_1, e_2, \dots, e_{10} \rangle$, приставку $P = \langle p_1, p_2, \dots, p_{10} \rangle$, корень $B = \langle b_1, b_2, b_3 \rangle$. Переменные y_i, x_i, e_i, p_i, b_i принимаю

значения из алфавита, состоящего из всех русских букв и знаков: пробела — «□», ударения — «/», дефиса — «—».

Под окончанием, следуя А. А. Зализняку [5], понимаем последние буквы слова, независимо от его членения на морфемы. Для выделения флексии, суффикса и постфикса достаточно добавить ряд уравнений.

Таблица 1

В табл. 1 приведены категории и значения глагольных форм и их обозначения в виде переменных z_i из Z .

Через z_{11} обозначим признак типа спряжения. z_{11} принимает значение 1, если слово относится к 1-му типу спряжения, 2 — ко 2-му, ..., 16 — к 16-му, 17 — если слово не является глаголом. Признак архаичности z_{12} введен для различения параллельных форм деепричастий на b (современная c) и $вши$ (архаичная a), а признак экспрессивности z_{13} — для различных форм повелительного наклонения с частицей $ка$ ($z_{13} = д$) и без нее ($z_{13} = н$). Изменение причастий по признакам прилагательных целесообразно рассматривать при описании морфологического разряда прилагательных [7].

Введем вспомогательные переменные $T = \langle t_1, t_2, t_3 \rangle$, которые позволяют упростить описание. Через t_1 обозначим признак

Грамматическая категория	Переменная	Грамматическое значение	Значение переменной
Наклонение	z_1	Изъявительное	и
		Повелительное	п
		Сослагательное	с
Время	z_2	Настоящее	н
		Прошедшее	п
Число	z_3	Будущее	б
		Единственное	е
Лицо	z_4	Множественное	м
		Первое	п
Род	z_5	Второе	в
		Третье	т
		Мужской	м
Вид	z_6	Женский	ж
		Средний	с
		Совершенный	с
Репрезентация	z_7	Несовершенный	н
		Деепричастие	д
		Инфинитив	и
Залог	z_8	Личная форма	л
		Причастие	п
		Действительный	д
Возвратность	z_9	Страдательный	с
		Возвратная (форма)	в
Переходность	z_{10}	Невозвратная (форма)	н
		Переходный	п
		Непереходный	н

ударения со значениями y , если знак ударения присутствует в словоформах и $б$ — в противном случае. Рассмотрим наиболее часто встречающийся на практике случай, когда словоформы поступают на обработку без знака ударения, т. е.

t_1^6 .

(1)

Для случая $t_1 = y$ система уравнений строится аналогичным образом. Признак формы t_2 принимает значения c (синтетическая) и a (аналитическая) и признак основы t_3 — значения n (полная) и k (краткая).

Проведенные исследования позволили разбить глаголы 13-го типа на 5 подтипов. Введем словарный признак s со значениями 1, если слово принадлежит к 1-му подтипу глаголов 13-го типа, 2 — ко 2-му, ..., 5 — к 5-му, 6 — если слово не входит в 13-й тип. В табл. 2 приведены основные формы (в данном случае — невозвратные инфинитивы) глаголов 13-го типа.

Таблица 2

i	s_i
1	Наставать, переставать, преставать, приставать, недоставать, отставать, уставать, вставать, привставать, представать, восставать, приотставать, приуставать
2	Повставать
3	Наподдавать, передавать
4	Давать, задавать, обдавать, наддавать, поддавать, предавать, раздавать, издавать, переиздавать, воздавать, создавать, пересоздавать, воссоздавать, придавать, додавать, недодавать, подавать, преподавать, продавать, запрашивать, перепродавать, допродавать, распродавать, сдавать, пересдавать, отдавать, выдавать, признавать, познавать, прознавать, вызнать, опознавать, распознавать, сознать, осознать, узнать, предугадывать, разузнавать, заставить, отставать, доставать
5	Надавать, насдавать

Для каждой переменной z_i , t_j , s справедливы уравнения (1)–(17), характеризующие ее область определения, например

$$z_7^n \vee z_7^a \vee z_7^c \vee z_7^d. \quad (8)$$

Связь вспомогательных переменных с основными выразим уравнениями (18)–(25), например, в аналитической форме из двух слов, разделенных одним пробелом:

$$\overline{y_1 y_2} \dots \overline{y_i y_{i+1} y_{i+2}} \dots \overline{y_i y_{i+1} y_{i+2}} = t_2^a \quad (i = \overline{4, 15}; j = \overline{i + 3, 22}) \quad (20)$$

или в синтетической форме из одного слова:

$$\overline{y_1 y_2} \dots \overline{y_i y_{i+1} y_{i+2}} = t_2^c \quad (i = \overline{3, 17}). \quad (21)$$

Уравнение сочетаемости значений позволяет из множества комбинаций значений введенных переменных удалить запрещенные

$$t_2^c (z_7^n (z_1^n (t_3^k (z_9^n z_8^d (z_6^n z_2^n (s^1 \vee s^3 \vee s^4) \vee z_6^c \wedge \wedge z_2^c (s^2 \vee s^3 \vee s^5)) \vee z_8^c z_6^n z_2^n z_4^b z_9^b (s^1 \vee s^3 \vee s^4) \vee \vee z_9^b z_2^n (s^1 \vee s^3 \vee s^4) \vee t_3^n z_2^n z_{11}^{13}) \vee z_1^n z_8^d \wedge z_4^n z_3^n z_{11}^{13}) \vee$$

$$\begin{aligned} & \vee z_7^n t_{32}^n z_{11}^{13} \vee z_7^n t_{32}^k z_6^n z_2^n z_8^n \wedge (s^1 \vee s^3 \vee s^4) \vee \\ & \vee z_7^n z_8^n ((s^1 \vee s^3 \vee s^4) z_6^n z_2^n \vee z_2^n \wedge z_{11}^{13}) t_3^n \vee \\ & \vee t_2^n z_7^n t_3^n (z_6^n (s^1 \vee s^3 \vee s^4) z_2^n z_1^n \vee z_7^n z_{11}^{13}). \end{aligned} \quad (26)$$

Словоформу, основную форму, приставку, корень и окончание записываем, начиная с первой левой позиции, а позиции, оставшиеся неиспользованными, заполняем знаками пробела. Например, для формы читаю: $y_1 = \text{ч}; y_2 = \text{и}; y_3 = \text{т}; y_4 = \text{а}; y_5 = \text{ю}; y_6 = \square, \dots, y_{30} = \square$. Заполнение неиспользованных позиций пробелами достигается с помощью уравнений (27)—(30):

$$\bar{y}_i \bar{y}_{i+1} = \bar{y}_i \bar{y}_{i+1} \dots \bar{y}_{30}, \quad (i = \overline{4, 23}). \quad (28)$$

Введем вспомогательные переменные $\Gamma = \langle \gamma_1, \gamma_2, \dots, \gamma_{15} \rangle$, $\Omega = \langle \omega_1, \omega_2, \dots, \omega_{15} \rangle$, которые упростят выделение глаголов 13-го типа из любых словоформ. В глаголах 13-го типа сочетания букв да, ста, зна образуют корень, а все буквы, стоящие перед ними — приставку (в синтетической форме и сослагательном наклонении), например:

$$\begin{aligned} & y_i^n y_{i+1}^a \bar{y}_i \bar{y}_{i+1} (t_2^c \vee y_{i+2}^b y_{i+3}^a y_{i+4}^a t_2^a) = (p_1 = y_1) \dots (p_{i-1} = \\ & = y_{i-1}) p_i^c b_1^a b_2^a b_3^c (\gamma_1 = y_{i+2}) \wedge \dots \wedge (\gamma_{i-i-1} = y_i) \bar{\gamma}_{i-i}, \\ & \quad (i = \overline{2, 8}; \quad j = \overline{i+2, 17}); \end{aligned} \quad (33)$$

$$\begin{aligned} & x_i^n x_{i+1}^a \bar{x}_i \bar{x}_{i+1} = (p_1 = x_1) \dots (p_{i-1} = x_{i-1}) p_i^c \wedge (\omega_1 = \\ & = x_{i+2}) \dots (\omega_{j-i-1} = x_j) \bar{\omega}_{j-i}, \quad (i = \overline{2, 8}; \quad j = \overline{i+2, 13}). \end{aligned} \quad (34)$$

Уравнение выделения глаголов 5-го подтипа 13-го типа имеет вид:

$$\begin{aligned} & s^5 = p_1^n p_2^a (p_3^c \vee p_3^c p_4^c) b_1^a b_2^a b_3^c (\gamma_1^n \gamma_2^a (\gamma_3^a ((\gamma_4^c \vee \gamma_4^c \gamma_5^k \gamma_6^a \gamma_7^c) \vee \\ & \vee \gamma_4^c \gamma_5^a (\gamma_6^c \vee \gamma_6^c \gamma_7^k \gamma_8^a \gamma_9^c) \vee \gamma_4^c \gamma_5^c (\gamma_6^c \vee \gamma_6^c \gamma_7^k \gamma_8^a \gamma_9^c) \vee \gamma_6^c \gamma_7^b (\gamma_8^c \vee \\ & \vee \gamma_3^c \gamma_9^k \gamma_{10} \gamma_{11}))) \vee \gamma_3^a (\gamma_4^c \vee \gamma_4^c \gamma_5^a \gamma_6^c \vee (\gamma_4^a \vee \gamma_4^a \vee \gamma_4^a) \times \\ & \times (\gamma_5^c \vee \gamma_5^c \gamma_6^b \gamma_7^c)) \vee \gamma_3^b (\gamma_4^c \vee \gamma_4^c \gamma_5^a (\gamma_6^c \vee \gamma_6^c \gamma_7^b \gamma_3^c \vee \\ & \vee \gamma_6^a (\gamma_7^c \vee \gamma_7^c \gamma_8^a \gamma_9^c))) \vee \gamma_1^c (\gamma_2^c \vee \gamma_2^c \gamma_3^c) \vee \gamma_1^c (\gamma_2^c \gamma_3^b \gamma_4^c \vee \gamma_2^c \gamma_3^c \vee \\ & \vee \gamma_2^c (\gamma_3^c \vee \gamma_3^c \gamma_4^c)) \vee \gamma_1^a \gamma_2^a \gamma_3^c \gamma_4^b (\gamma_5^c \vee \gamma_5^c \gamma_6^a \gamma_7^c) \vee \\ & \vee \omega_1^a \omega_2^a \omega_3^c \omega_6^b (\omega_5^c \vee \omega_5^c \omega_6^a \omega_7^c). \end{aligned} \quad (43)$$

Уравнения (39)—(42) для выделения глаголов других подтипов 13-го типа выглядят аналогично.

Уравнения (44)—(91) связывают формальные признаки синтетической словоформы и формы сослагательного наклонения с грамматическими категориями. Например, глагол 13-го типа в синте-

тической форме, оканчивающиеся на *ю*, характеризуется признаками *ед. числа 1-го лица, невозвратности* и др.:

$$t_2^c y_i^{\bar{ю}} y_{i+1}^{\bar{ю}} = t_3^k z_1^{\bar{ю}} z_7^{\bar{ю}} z_8^{\bar{ю}} z_9^{\bar{ю}} (z_6^{\bar{ю}} z_2^{\bar{ю}} (s^1 \vee s^3 \vee s^4) \vee \vee z_2^6 z_6^c (s^2 \vee s^3 \vee s^5)) z_3^e z_4^{\bar{ю}}, \quad (i = \overline{3,10}); \quad (144)$$

сослагательное наклонение состоит из частицы *бы* и формы прошедшего времени:

$$t_2^a y_i^{\bar{ю}} y_{i+1}^{\bar{ю}} y_{i+2}^{\bar{ю}} y_{i+3}^{\bar{ю}} y_{i+4}^{\bar{ю}} = t_3^{\bar{ю}} z_7^{\bar{ю}} z_1^c (z_6^{\bar{ю}} (s^1 \vee s^3 \vee s^4) \vee \vee (s^2 \vee s^3 \vee s^5) z_6^c) z_4^e z_5^{\bar{ю}} z_9^{\bar{ю}}, \quad (i = \overline{4,12}). \quad (145)$$

Уравнения (92)—(136) описывают отношение между грамматическими категориями и окончаниями. Например, невозвратный глагол 13-го типа в настоящем времени, единственном числе, втором лице, характеризующийся рядом дополнительных признаков, имеет окончание *ёшь*:

$$t_3^k z_1^{\bar{ю}} z_7^{\bar{ю}} z_8^{\bar{ю}} z_9^{\bar{ю}} (z_6^{\bar{ю}} z_2^{\bar{ю}} (s^1 \vee s^3 \vee s^4) \vee z_2^6 z_6^c (s^2 \vee s^3 \vee s^5)) \times \times t_2^e z_3^e z_4^{\bar{ю}} = e_1^e e_2^{\bar{ю}} e_3^{\bar{ю}} e_4^{\bar{ю}} e_5^{\bar{ю}}. \quad (146)$$

Инфинитив глаголов в синтетической форме и сослагательном наклонении образуется отбрасыванием окончания и присоединением *ть* (*тбся*) или *вать* (*ватбся*), например:

$$\overline{y_i} \overline{y_{i+1}} \overline{y_{i+2}} e_i \overline{e_{i+1}} \overline{e_{i+2}} t_3^{\bar{ю}} z_{11}^{13} (t_2^c \vee z_1^c) = (x_1 = y_1) \dots (x_{i-j} = y_{i-j}) x_{i-j+1}^{\bar{ю}} x_{i-j+2}^{\bar{ю}} \wedge (x_{i-j+3}^{\bar{ю}} \vee x_{i-j+4}^{\bar{ю}} x_{i-j+5}^{\bar{ю}} z_9^{\bar{ю}}), \quad (i = \overline{3,17}; \quad j = \overline{1,8}). \quad (137)$$

Синтетическая форма и сослагательное наклонение образуется отбрасыванием сочетаний *ть* (*тбся*) или *вать* (*ватбся*) и присоединением окончания, например:

$$x_i^{\bar{ю}} x_{i+1}^{\bar{ю}} (t_2^c \vee z_1^c) t_3^{\bar{ю}} z_{11}^{13} = (y_1 = x_1) \wedge \dots \wedge (y_{i-2} = x_{i-2}) (y_{i-1} = e_1) \dots (y_{i+8} = e_{10}), \quad (i = \overline{6,13}). \quad (138)$$

Второе слово глагола несовершенного вида будущего времени совпадает с инфинитивом:

$$t_2^a y_i^{\bar{ю}} y_{i+1}^{\bar{ю}} y_{i+2}^{\bar{ю}} y_i \overline{y_{i+1}} \overline{y_{i+2}} y_i \overline{y_{i+1}} = (x_1 = y_{i+2}) \wedge \dots \wedge (x_{i-i-1} = y_i) x_{i-i}^{\bar{ю}} (s^1 \vee s^3 \vee s^4) z_6^{\bar{ю}} z_2^{\bar{ю}} z_7^{\bar{ю}}, \quad (i = \overline{4,6}; \quad j = \overline{11,20}). \quad (143)$$

Формы будущего времени глаголов несовершенного вида анализируются с помощью формул (144)—(155), аналогичных

$$z_2^6 y_i^{\bar{ю}} y_{i+1}^{\bar{ю}} = t_3^a z_1^{\bar{ю}} z_7^{\bar{ю}} \wedge \wedge t_3^{\bar{ю}} (s^1 \vee s^3 \vee s^4) z_3^e z_4^{\bar{ю}} z_9^{\bar{ю}}, \quad (i = \overline{11,20}). \quad (147)$$

Уравнения (156)—(161) связывают грамматические признаки форм будущего времени глаголов несовершенного вида и окончания слова *быть*, например:

$$z_1^h z_2^b z_6^h z_7^h (s^1 \vee s^3 \vee s^4) z_3^e z_4^r = e_1^h e_2^r e_3^h e_4^h. \quad (160)$$

Будущее время у глаголов несовершенного вида образуется присоединением к основе *буд* окончания (первое слово), второе слово образуется от инфинитива:

$$\begin{aligned} x_i^h x_{i+1}^h e_1^h e_{j+1}^h e_{i+2}^h t_2^a z_9^h (s^1 \vee s^3 \vee s^4) z_6^h z_2^b &= y_1^b y_2^y y_3^h (y_4 = \\ &= e_1) \dots (y_{3+i} = e_i y_{4+i}^h \wedge (y_{5+i} = \\ &= x_i) \dots (y_{4+i+i} = x_i) y_{5+i+i}^h, \quad (i = \overline{6,13}; \quad j = \overline{1,3}). \end{aligned} \quad (162)$$

Группы уравнений (137)—(155) и (162)—(165) описывают соответствие между основной формой глагола и его инфинитивом и необходимы при синтезе и нормализации.

При решении задачи анализа используются уравнения: (44)—(91), (144)—(155); синтеза (92)—(136), (139)—(142), (156)—(165); нормализации (44)—(138), (143).

Разработанное математическое описание содержит 167 уравнений и является одним из вариантов модели грамматической обработки глаголов 13-го типа. Проводятся экспериментальные исследования различных вариантов модели на ЭВМ. Результаты исследований свидетельствуют о безошибочном действии модели.

Предложенная система уравнений позволяет выделить глагольные словоформы 13-го типа из словоформ русского языка, решить задачи их анализа, синтеза и нормализации, а также определения приставки, корня и ряд других. Полученная модель отражает специфику, основные закономерности обработки глагольных форм и используется при построении общей модели обработки глаголов и других частей речи.

- 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 автоматки. Харьков, 1977, вып. 52, с. 21—28. 3. Шабанов-Кушнарченко Ю. П. Об алгебре предикатов с отрицанием.— В кн.: АСУ и приборы автоматки. Харьков, 1977, вып. 52, с. 42—49. 4. Шабанов-Кушнарченко Ю. П. Об уравнениях теории интеллекта.— В кн.: АСУ и приборы автоматки. Харьков, 1977, вып. 53, с. 63—70. 5. Зализняк А. А. Грамматический словарь русского языка (словоизменение). М., Наука, 1977. 749 с. 6. Грамматика русского языка, т. 1 /Под ред. В. В. Виноградова и др. М., АН СССР, 1960. 719 с. 7. Шабанов-Кушнарченко Ю. П. Математическая модель склонения полных непряжательных имен прилагательных. Сб. трудов /Науч.-техн. информация. Серия 2, М., ВИНТИ, 1979. 8. 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 20 марта 1979 г.