УДК 510.62

М. Ф. БОНДАРЕНКО, д-р техн. наук, *Д. Э. СИТНИКОВ, Н. В. ШАРОНОВА,* канд. техн. наук, *Е. В. ЯВТУШЕНКО*

О МАТЕМАТИЧЕСКОМ ОПИСАНИИ МЕЖМОРФЕМНЫХ ОТНОШЕНИЙ

Важнейшей компонентой автоматизированных информационных систем является подсистема информационно-лингвистического обеспечения, под которой обычно понимается совокупность информационных структур, объединенных в базу данных, алгоритмов их обработки, а также моделей и языков для доступа к базе данных. Наиболее удобное средство для общения человека с ЭВМ — естественный язык, использование которого требует наличия в автоматизированной системе элементов интеллекта для понимания (интерпретации) естественно-языкового запроса и формирования адекватного сообщения.

Ключевой проблемой при создании автоматизированных информационных систем является проблема понимания смысла сообщения, которая неизбежно приводит к решению проблемы понимания смысла слова. По мнению многих ведущих специалистов в области языкознания и прикладной лингвистики, смысл слова не является в общем случае суммой смыслов составляющих его морфем, это даже не функция, это отношение более общего вида. В связи с этим становится актуальной задача исследования и математического моделирования межморфемных отношений, т. е. тех связей, которые существуют в слове между префиксами и корнями, корнями и суффиксами, основами и окончаниями (флексиями).

Пусть M — множество морфем одного типа (например, префиксов). На множестве M введем систему предикатов S так, чтобы любой предикат $P(\tilde{t}) \in S$ обращался в 1 на множестве морфем с какой-либо определенной семантической ролью и был равен 0 в противном случае. Таким образом, множество предикатов S можно отождествить с множеством семантических ролей префиксальных морфем. Каждому элементу A из M соответствует множество предикатов из S, дающих 1 при подстановке A. Следовательно, каждому $A \in M$ взаимно-однозначно соответствует некоторый одноместный подстановочный предикат $A(t)$, где $t \in S^*$. Мы получили множество S семантических ролей с определенным на нем множеством M предикатов-морфем.

Рассмотрим теперь два множества семантических ролей S_1 и S_2 и соответственно два множества предикатов-морфем M_1 и M_2. Операция соединения двух морфем $P_1(t_1)$ и $P_2(t_2)$ из M_1 и M_2 характеризуется согласованием семантических ролей, присущих данным множествам морфем. Результатом такого соединения на-

* Шабанов-Кушнаренко Ю. П. Теория интеллекта. Проблемы и перспективы. Х., 1987. С. 39.

зовем множество связей между семантическими ролями, т. е. множество пар семантических ролей, характеризующих две стоящие рядом морфемы.

Такое множество представляет собой некоторый бинарный предикат $P(t_1, t_2)$, причем $P(t_1, t_2) \to P_1(t_1)\ P_2(t_2)$. Предположим возможность согласования семантических ролей не зависит от того, к каким морфемам они относятся, т. е. любые две роли из множеств S_1 и S_2 либо согласуются, либо нет. Тогда на декартовом произведении множеств $S_1 \times S_2$ можно задать предикат $\lambda(t_1, t_2)$, принимающий значение 1, если роли t_1 и t_2 можно согласовать, и значение 0 в противном случае.

Результатом соединения двух морфем $P_1(t_1)$ и $P_2(t_2)$ является логическое умножение $P_1(t_1) \cdot P_2(t_2)$, что может означать возможность согласования любой роли морфемы $P_1(t_1)$ с любой ролью морфемы P_2. Однако словарная реализация не подтверждает этого утверждения: часто некоторые семантические роли рядом стоящих морфем не согласуются. В общем случае операция соединения (*) морфемных семантических ролей запишется так:

$$P_1(t_1) * P_2(t_2) = \lambda(t_1, t_2)\, P_1(t_1)\, P_2(t_2). \tag{1}$$

Действительно, логическое произведение $P_1(t_1)\ P_2(t_2)$ исчерпывает все связи между семантическими ролями морфем P_1 и P_2, а $\lambda(t_1, t_2)$ исключает часть нереализованных связей. Язык использует не все возможные подмножества семантических ролей. Однако для удобства математического описания без особого огрубления задачи будем считать, что множества M_1 и M_2 совпадают с множествами всех предикатов, определенных соответственно на множествах S_1 и S_2. Тогда булевы алгебры предикатов из множеств M_1 и M_2 являются подалгебрами булевой алгебры бинарных предикатов, заданных на $S_1 \times S_2$. Докажем теорему, позволяющую аксиоматически вести операцию соединения морфем.

Пусть имеется конечная булева алгебра M и две подалгебры $M_1 \subseteq M$, $M_2 \subseteq M$.

Теорема. *Пусть функция f: $M_1 \times M_2 \to M$ удовлетворяет свойствам:*

$$f(x, y)\, xy = f(x, y);$$

$$f(x_1 \vee x_2, y) = f(x_1, y) \vee f(x_2, y);$$

$$f(x, y_1 \vee y_2) = f(x, y_1) \vee f(x, y_2),$$

тогда $f(x)$ можно представить в виде

$$f(x, y) = \lambda xy,$$

где λ — фиксированный элемент из M. Обратно, если $f(x, y)$, можно представить в виде (5), то выполняются свойства (2)—(4).

Доказательство. Вторая часть теоремы очевидна. Докажем первую часть. Пусть $\{x_i\}_{i=1}^m$ — система минимальных ненулевых элементов булевой подалгебры M_1, $\{y_j\}_{j=1}^n$ — система минимальных ненулевых элементов подалгебры M_2. Пусть также для функции f выполняются свойства (2)—(4). Положим $f(0, y)=0$ для любого y, $f(x, o)=0$ для любого x. Покажем, что для любых ненулевых x, y

$$f(x, y)=\left[\bigvee_{i=1}^{m}\bigvee_{j=1}^{n} f(x_i, y_j)\right] xy. \tag{6}$$

Любой ненулевой элемент x подалгебры M_1 может быть представлен в виде дизъюнкции некоторых минимальных элементов этой подалгебры:

$$x=x_{i_1}\vee x_{i_2}\vee\ldots\vee x_{i_k}.$$

Аналогично $y=y_{j_1}\vee y_{j_2}\vee\ldots\vee_{j_e}$. Следовательно,

$$\left[\bigvee_{i=1}^{m}\bigvee_{j=1}^{n} f(x_i, y_j)\right] xy=\left[\bigvee_{i=1}^{m}\bigvee_{j=1}^{n} f(x_i, y_j)\right](x_{i_1}\vee x_{i_2}\vee\ldots\vee$$

$$\vee x_{i_k})\wedge(y_{j_1}\vee y_{j_2}\vee\ldots\vee y_{j_e})=\left[\bigvee_{i=1}^{m}\bigvee_{j=1}^{n} f(x_i, y_j)\right](x_{i_1}, y_{j_1}\vee$$

$$\vee X_{i_1}y_{j_2}\vee X_{i_2}Y_{j_3}\vee\ldots\vee x_{i_1}y_{j_e}\vee x_{i_2}y_{j_1}\vee x_{i_2}y_{j_2}\vee\ldots\vee x_{i_2}y_{j_e}\vee\ldots\vee$$

$$\vee x_{i_k}y_{j_1}\vee x_{i_k}y_{j_2}\vee\ldots\vee x_{i_k}y_{j_e}=\left[\bigvee_{i=1}^{m}\bigvee_{j=1}^{n} f(x_i, y_j)\, x_i y_j\right]\wedge$$

$$\wedge(x_{i_1}y_{j_1}\vee x_{i_1}y_{j_2}\vee\ldots\vee x_{i_1}y_{j_e}\vee x_{i_2}y_{j_1}\vee x_{i_2}y_{j_2}\vee\ldots\vee$$

$$\vee x_{i_2}y_{j_e}\vee\ldots\vee x_{i_k}y_{j_1}\vee x_{i_k}y_{j_2}\vee\ldots\vee x_{i_k}y_{j_e}).$$

Здесь мы применили свойство (2). Раскрывая скобки и используя тот факт, что конъюнкция двух различных минимальных ненулевых элементов равна нулю, получаем

$$\left[\bigvee_{i=1}^{m}\bigvee_{j=1}^{n} f(x_i, y_j)\, xy=f(x_{i_1}, y_{j_1})\, x_{i_1}y_{j_1}\vee\ldots\vee f(x_{i_k}, y_{j_e})\, x_{i_k}y_{j_e}.\right.$$

Еще раз применим свойство (2):

$$\left[\bigvee_{i=1}^{m}\bigvee_{j=1}^{n} f(x_i, y_j)\right] xy=f(x_{i_1}, y_{j_1})\vee f(x_{i_1}, y_{j_2})\vee\ldots\vee f(x_{i_1}, y_{j_e})\vee$$

$$\vee f(x_{i_2}, y_{j_1})\vee f(x_{i_2}, y_{j_2})\vee\ldots\vee f(x_{i_2}, y_{j_e})\vee\ldots\vee f(x_{i_k}, y_{j_1})\vee$$

$$\vee f(x_{i_k}, y_{j_2})\vee\ldots\vee f(x_{i_k}, y_{j_e})=f(x_{i_1}, y_{j_1}\vee y_{j_2}\vee\ldots\vee y_{j_e})\vee$$

$$\vee f(x_{\ 2}, y_{j_1}\vee y_{j_2}\vee\ldots\vee y_{j_e})\vee\ldots\vee f(x_{i_k}, y_{j_1}\vee y_{j_2}\vee\ldots\vee y_{j_e})=$$

$$=f(x_{i_1}\vee x_{i_2}\vee\ldots\vee x_{i_k}, y_{j_1}\vee y_{j_2}\vee\ldots\vee y_{j_e})=f(x, y).$$

Здесь мы применили свойства (3) и (4). Таким образом, если вместо λ взять $\left[\bigvee_{i=1}^{m}\bigvee_{j=1}^{n} f(x_i, y_j)\right]$, то получим $f(x, y)=$ $=\lambda xy$. Теорема доказана.

Проиллюстрируем соединение $p_1(t_1)$ и $P_2(t_2)$ примером взаимодействия морфем. Для этого рассмотрим взаимодействие префиксальных морфов и основ. Исследование операции соединения $P_1(t_1) * P_2(t_2)$ проведем на морфемном шве между префиксом и остальной частью слова. Пусть каждому исследуемому префиксальному морфу соответствует свое подмножество множества M_1 семантических ролей, а каждой основе, соединяющейся с данным морфом, соответствует свое подмножество множества M_2 семантических ролей.

Многозначные глагольные префиксы **вы-**, **до-**, **на-**, **пере-**, **от-** соединяются с основами **работать, рубить, лететь** и образуют слова, наделенные определенным смыслом.

В словах «переработать», «перерубить», «зарубить», «вырубить», «доработать», «вылететь», «долететь», «залететь» происходит такой выбор связей, что несколько семантических ролей префикса связаны с одной или несколькими семантическими ролями основы.

Полученное слово имеет столько смыслов, сколько существует связей между морфемами.

На конкретном примере рассмотрим образование связей между префиксальными морфемами «от-» и «у-» и основами «работать» и «лететь». Префикс «от-» обладает следующими семантическими ролями: x_1 — «отделение части от целого»; x_2 — «полнота действия»; x_3 — «обратное действие»; x_4 — «ответное действие»; x_5 — «абсолютный конец действия, длящегося неопределенный отрезок времени»; x_6 — «действие, длящееся определенный отрезок времени»; x_7 — «оттенок прошедшего времени»; x_8 — горизонтальная направленность действия от объекта в плоскости, перпендикулярной объекту-ориентиру»; x_9 — «нежелательный результат действия»; x_{10} — «результат действия».

Префиксальная морфема «у-» обладает следующим набором семантических ролей: x_{11} — «направленность действия на весь объект»; x_{12} — «уменьшение»; x_{13} — «уничтожение»; x_{14} — «умещение»; x_{15} — «сохранение прежнего состояния»; x_{16} — «достижение цели»; а также семантическими ролями x_8, x_9, x_{10}, которые являются общими для данных префиксальных морфем.

Основы «работать» и «лететь» характеризуются семантическими ролями $y_1 \div y_4$ и $y_5 \div y_8$ соответственно: y_1 — «находиться в действии»; y_2 — «заниматься чем-нибудь, применяя свой труд, осуществляя какую-нибудь деятельность»; y_3 — «иметь где-нибудь какое-нибудь занятие, служить»; y_4 — «обслуживать кого-нибудь своим трудом»; y_5 — «нестись, передвигаться по воздуху»; y_6 — «мчаться»; y_7 — «падать»; y_8 — «о времени: быстро проходить».

Сочетание префиксальных и корневых морфем удобно показать с помощью схемы, из которой видно, как образуются связи между данными частями слова.

Префикс «от-» и основа «работать» образуют слово «отработать», которое характеризуется шестью смысловыми значениями. Каждой связи между множеством семантических ролей префикса и множеством семантических ролей основы соответствует одно смысловое значение полученного слова: (x_4y_1) — «провести над работой, в работе определенное время»; (x_5y_1) — «перестать находиться в действии, остановиться»; (x_5y_2) — «кончить работать, занимаясь до этого чем-нибудь, применяя свой труд, осуществляя какую-нибудь деятельность»; (x_5y_3) — «окончить работу на пред-

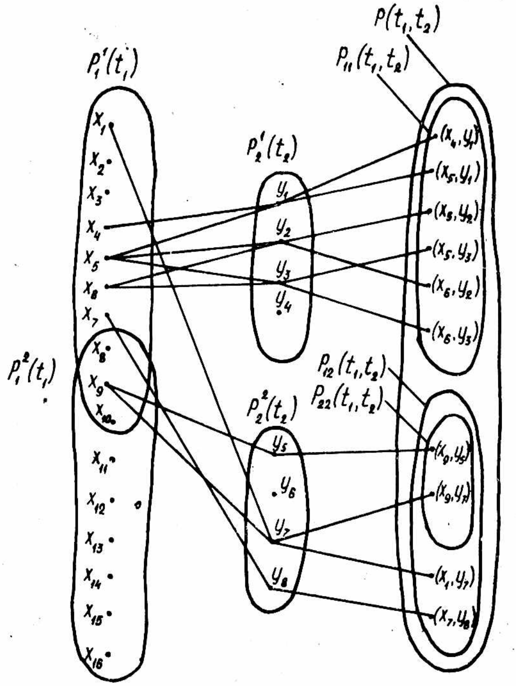

Схема сочетания префиксальных и корневых морфем

приятии, в учреждении»; (x_6y_3) — «провести над работой, в работе какое-нибудь время, занимаясь какой-нибудь деятельностью».

Слово «отлететь» получает в результате сочетания префикса «от-» и основы «лететь» четыре смысловых значения: (x_9y_5) — «летя, удалиться по отношению к объектам, которые совершают активное движение»; (x_9y_6) — «летя, удалиться и упасть по отношению к объектам, которые совершают пассивное движение»; $(x_{11}y_7)$ — «оторваться»; (x_7y_8) — «исчезнуть, пропасть».

Слово «улететь» характеризуется двумя смысловыми значениями: (x_9y_5) — «летя, удалить по отношению к объектам, которые совершают активное движение»; (x_9y_7) — «летя, удалиться по отношению к объектам, которые совершают пассивное движение».

Формула (1) позволяет описать взаимосвязи между префиксальными морфемами и основами. Для рассматриваемого примера $\lambda(t_1, t_2)$ имеет вид:

$$\lambda(t_1,\ t_2) = t_1^{x_4} t_2^{y_1} \vee t_1^{x_5} t_2^{y_1} \vee t_1^{x_5} t_2^{y_2} \vee t_1^{x_5} t_2^{y_3} \vee t_1^{x_6} t_2^{y_2} \vee t_1^{x_6} t_2^{y_3} \vee$$

$$\vee\, t_1^{x_1} t_2^{y_7} \vee t_1^{x_7} t_2^{y_8} \vee t_{x_9} t_2^{y_5} \vee t_2^{x_9} t_2^{y_7}. \qquad (7)$$

Множества семантических ролей префиксов «от-» и «у-» определяются предикатами $P_1^1(t_1)$ и $P_1^2(t_1)$ соответственно:

$$P_1^1(t_1) = t_1^{x_1} \vee t_1^{x_2} \vee t_1^{x_3} \vee t_1^{x_4} \vee t_1^{x_5} \vee t_1^{x_6} \vee t_1^{x_7} \vee t_1^{x_8} \vee t_1^{x_9} \vee t_1^{x_{10}}; \quad (8)$$

$$P_1^2(t_1) = t_1^{x_8} \vee t_1^{x_9} \vee t_1^{x_{10}} \vee t_1^{x_{11}} \vee t_1^{x_{12}} \vee t_1^{x_{13}} \vee t_1^{x_{14}} \vee t_1^{x_{15}} \vee t_1^{x_{16}}. \quad (9)$$

Множества семантических ролей основ «работать» и «лететь» определяются предикатами $P_2^1(t_2)$ и $P_2^2(t_2)$ соответственно:

$$P_2^1(t_2) = t_2^{y_1} \vee t_2^{y_2} \vee t_2^{y_3} \vee t_2^{y_4}; \quad (10)$$

$$P_2^2(t_2) = t_2^{y_5} \vee t_2^{y_6} \vee t_2^{y_7} \vee t_2^{y_8}. \quad (11)$$

Опишем с помощью формулы (1) каждую конкретную реализацию. В случае взаимодействия префикса «от» и основы «работать» образуется слово «отработать»:

$$P_{11}(t_1, t_2) = \lambda(t_1, t_2) P_1^1(t_1) P_2^1(t_2) = (t_1^{x_4} t_2^{y_1} \vee t_1^{x_5} t_2^{y_1} \vee$$
$$\vee t_1^{x_5} t_2^{y_2} \vee t_1^{x_5} t_2^{y_3} \vee t_1^{x_6} t_2^{y_2} \vee t_1^{x_6} t_2^{y_3} \vee t_1^{x_1} t_2^{y_7} \vee t_1^{x_7} t_2^{y_8} \vee$$
$$\vee t_1^{x_9} t_2^{y_5} \vee t_1^{x_9} t_2^{y_7})(t_1^{x_1} \vee t_1^{x_2} \vee t_1^{x_3} \vee t_1^{x_4} \vee t_1^{x_5} \vee t_1^{x_6} \vee t_1^{x_7} \vee$$
$$\vee t_1^{x_8} \vee t_1^{x_9} \vee t_1^{x_{10}})(t_2^{y_1} \vee t_2^{y_2} \vee t_2^{y_3} \vee t_2^{y_4}) = t_1^{x_4} t_2^{y_1} \vee t_1^{x_5} t_2^{y_1} \vee$$
$$\vee t_1^{x_5} t_2^{y_2} \vee t_1^{x_5} t_2^{y_3} \vee t_1^{x_6} t_2^{y_2} \vee t_1^{x_6} t_2^{y_3}. \quad (12)$$

Предикат $P_{11}(t_1, t_2)$ является результатом операции соединения предикатов $P_1^1(t_1)$ и $P_2^1(t_2)$:

$$P_1^1(t_1) * P_2^1(t_2) = P_{11}(t_1, t_2). \quad (13)$$

В случае сочетания префикса «у-» и основы «работать» слова не образуются:

$$P_1^2(t_1) * P_2^1(t_2) = P_{21}(t_1, t_2); \quad (14)$$

$$P_{21}(t_1, t_2) = \lambda(t_1, t_2) P_1^2(t_1) P_2^1(t_2) = (t_1^{x_4} t_2^{y_7} \vee t_1^{x_7} t_2^{y_8} \vee$$
$$\vee t_1^{x_9} t_2^{y_5} \vee t_1^{x_9} t_2^{y_7} \vee t_1^{x_4} t_2^{y_1} \vee t_1^{x_5} t_2^{y_1} \vee t_1^{x_2} t_2^{y_3} \vee t_1^{x_6} t_2^{y_2} \vee t_1^{x_6} t_2^{y_3}) \times$$
$$\times (t_1^{x_8} \vee t_1^{x_9} \vee t_1^{x_{10}} \vee t_1^{x_{11}} \vee t_1^{x_{12}} \vee t_1^{x_{13}} \vee t_1^{x_{14}} \vee t_1^{x_{15}} \vee t_1^{x_{16}}) \times$$
$$\times (t_2^{y_1} \vee t_2^{y_2} \vee t_2^{y_3} \vee t_2^{y_4}) = 0. \quad (15)$$

Образование связей между множествами семантических ролей префикса «от-» и основы «лететь» дает слово «отлететь», операция соединения предикатов $P_1^1(t_1)$ и $P_2^2(t_2)$ дает множество семантических значений слова «отлететь»:

$$P_1^1(t_1) * P_2^2(t_2) = P_{12}(t_1, t_2); \quad (16)$$

$$P_{12}(t_1, t_2) = \lambda(t_1, t_2) P_1^1(t_1) P_2^2(t_2) = (t_1^{x_4} t_2^{y_1} \vee t_1^{x_5} t_2^{y_1} \vee t_1^{x_5} t_2^{y_2} \vee$$

$$\vee t^x_{15}t^y_{23} \vee t^x_{16}t^y_{22} \vee t^x_{16}t^y_{23} \vee t^x_{19}t^y_{25} \vee t^x_{19}t^y_{27} \vee t^x_{11}t^y_{27} \vee t^x_{17}t^y_{28}) \times$$

$$\times (t^x_{11} \vee t^x_{12} \vee t^x_{13} \vee t^x_{14} \vee t^x_{15} \vee t^x_{16} \vee t^x_{17} \vee t^x_{18} \vee t^x_{19} \vee t^x_{110}) \times$$

$$\times (t^y_{25} \vee t^y_{26} \vee t^y_{27} \vee t^y_{28}) = t^x_{19}t^y_{25} \vee t^x_{19}t^y_{27} \vee t^x_{11}t^y_{27} \vee t^x_{17}t^y_{28}. \quad (17)$$

Слово «улететь» образуется в результате взаимодействия предикатов $P^2_1(t_1)$ и $P^2_2(t_2)$ и характеризуется двумя смысловыми значениями:

$$P^2_1(t_1) * P^2_2(t_2) = P_{22}(t_1, t_2); \quad (18)$$

$$P_{22}(t_1, t_2) = \lambda(t_1, t_2) P^2_1(t_1) P^2_2(t_2) = (t^x_{14}t^y_{21} \vee t^x_{15}t^y_{21} \vee t^x_{15}t^y_{22} \vee$$

$$\vee t^x_{15}t^y_{23} \vee t^x_{16}t^y_{22} \vee t^x_{16}t^y_{23} \vee t^x_{19}t^y_{25} \vee t^x_{19}t^y_{27} \vee t^x_{11}t^y_{27} \vee t^x_{17}t^y_{28}) \times$$

$$\times (t^x_{18} \vee t^x_{19} \vee t^x_{110} \vee t^x_{111} \vee t^x_{112} \vee t^x_{113} \vee t^x_{114} \vee t^x_{115} \vee t^x_{116}) \wedge$$

$$\vee (t^y_{25} \vee t^y_{26} \vee t^y_{27} \vee t^y_{28}) = t^x_{19}t^y_{25} \vee t^x_{19}t^y_{27}. \quad (19)$$

При образовании слов «отлететь» и «улететь» между множествами семантических ролей префиксов «от-», «у-» и основы «лететь» возникают две одинаковые связи: x_9y_5 и x_9y_7, т. е. в значениях, которые соответствуют этим связям, данные слова являются синонимами. При сравнении предикатов $P_{12}(t_1, t_2)$ и $P_{22}(t_1, t_2)$ видно, что

$$P_{22}(t_1, t_2) \to P_{12}(t_1, t_2). \quad (20)$$

Множество семантических ролей слова «улететь» содержится в множестве семантических ролей слова «отлететь».

Приведенные в данной работе уравнения являются частью математической модели межморфемных отношений. На примере исследования отношения между префиксами и основами показаны и описаны основные закономерности такого типа отношений.

Поступила в редколлегию 18.01.90