УДК 62.506.2

В. М. БОНДАРЕВ

О МОРФОЛОГИЧЕСКОМ ОТНОШЕНИИ НА МНОЖЕСТВЕ МОРФЕМ

В работах [1, 2] сформулирована трактовка задач морфологической обработки слов как решение уравнений вида

$$L(x,\ y,\ z_1,\ z_2,\ \ldots,\ z_n) = 1, \tag{1}$$

где x — лексема; y — форма слова; z_1, z_2, ..., z_n — грамматические признаки. В каждом частном случае могут быть заданы значения переменных, причем набор аргументов функции, значения которых известны, определяет тип решаемой морфологической задачи. Например, известны x и y, а надо определить z_1, z_2, ..., z_n — это задача грамматического анализа.

Левая часть уравнения (1) фиксирует некоторую взаимосвязь лексемы, словоформы и грамматических признаков, другими словами, ту же самую информацию, что и раздел грамматики, именуемый морфологией.

В правилах морфологии широко используются понятия «основа слова» и его «окончание», поэтому можно рассчитывать, что морфологическая функция [1] будет проще и компактнее, если в число ее переменных войдут такие морфемы, как основа и окончание.

Поскольку любая словоформа целиком определяется составляющими ее морфемами, аргумент y должен быть исключен из числа независимых переменных морфологической функции. Новая морфологическая функция примет следующий вид: $L(x, y_1, y_2, z_1, z_2, \ldots, z_n)$, где y_1 — основа словоформы, y_2 — ее окончание. Морфологической функции L' будет соответствовать уравнение

$$L'(x, y_1, y_2, z_1, z_2, \ldots, z_n) = 1. \qquad (2)$$

Однако практика требует, чтобы в условие ряда морфологических задач, например задачи анализа, входило значение переменной y — словоформы, а не значения переменных y_1 и y_2 — основы и окончания. Возникает противоречие: практически необходимо решать уравнение (1), а грамматические сведения удобно фиксировать в виде морфологической функции, зависящей от основы и окончания, и извлекать их путем решения морфологического уравнения (2).

Ниже делается попытка изложить и обосновать простую процедуру, позволяющую искать корни уравнения (1) путем решения уравнения (2). Для удобства изложения будем называть полным решением морфологического уравнения такое множество значений всех его независимых переменных, которое обращает уравнение в тождество. Условимся называть частичным решением любое подмножество полного решения. Омонимичные значения разных переменных считаем разными элементами решения. Например, полным решением уравнения вида (2) может быть множество {стол$_x$, стол$_{y_1}$, ом$_{y_2}$ единственное, творительный}. Нижний индекс указывает на принадлежность значения определенной переменной. Очевидно, всякое полное решение уравнения (1) состоит ровно из $n + 2$ элементов, а всякое полное решение уравнения (2) — из $n + 3$ элементов, причем все нижние индексы у них различны.

Заметим, что решение любой задачи морфологической обработки может быть приведено к единой схеме нахождения таких полных решений морфологического уравнения, которые вкл чали бы в себя наперед заданное частичное решение. Например, задача синтеза словоформы родительного падежа множественного числа лексемы «книга» может быть сведена к нахождению множества {книга$_x$, книг$_y$, множественное$_{z_1}$, родительный$_{z_2}$}, включающего в себя заданное подмножество {книга$_x$, множественное$_{z_1}$, родительный$_{z_2}$}. Формализуем сделанное замечание. Для этого обозначим через M множество всех полных и частичных решений морфологического уравнения (1). Символом ξ обозначим произвольное решение уравнения (1). Выделим во множестве M все полные решения и будем помечать их нулевым нижним индексом, например $\xi_0 \cdot \xi$ и ξ_0 являются элементами множества M, но по определению сами состоят из элементов — значений аргументов морфологической функции L. Поэтому на M можно определить отношение «быть подмножеством полного решения» $<A, M>$. Будем говорить, что выполняется соотношение $\xi A \xi_0$, если $\xi \subseteq \xi_0$.

Теперь задачу определения корней морфологического уравнения можно сформулировать так: найти множество таких полных решений, чтобы для каждого из них выполнялось соотношение $\xi A \xi_0$, где ξ — наперед заданное частичное решение; ξ_0 — элемент искомого множества. Если обозначить это множество через m, то, согласно [4], можно записать

$$m = \mathrm{E}\,\{\xi_0 \in M \mid \xi A \xi_0\}. \tag{3}$$

Распространим введенную систему обозначений на уравнение (2). Множество всех решений обозначим через N, под η будем понимать произвольный элемент множества N, нулевой нижний индекс обозначает полное решение уравнения (2). По аналогии с отношением $<A, M>$ определим отношение $<B, M>$ как вхождение частичного решения уравнения (2) в полное. Корни морфологического уравнения (2) составят множество n:

$$n = \mathrm{E}\,\{\eta_0 \in N \mid \eta B \eta_0\}. \tag{4}$$

Сопоставим множества M и N и отношения на них. Определим отображение α множества N на множество M следующим образом:

$$\xi = \alpha(\eta) = \begin{cases} \eta, \text{ если } \tilde{y}_1 \notin \eta \text{ и } \tilde{y}_2 \notin \eta; \\ \eta \cup \{\tilde{y}\} \setminus \{\tilde{y}_1, \tilde{y}_2\}, \text{ если } \tilde{y}_1 \in \eta, \tilde{y}_2 \in \eta. \end{cases} \tag{5}$$

Знак тильда указывает на конкретные значения тех переменных, над символами которых он поставлен. Фигурные скобки служат для обозначения множества, $\cup$ и $\setminus$ — знаки операций суммы и разности множеств. Условимся, что порядок выполнения этих операций соответствует последовательности их знаков в формуле. Отображение α не определено на тех частичных решениях уравнения (2), которые включают в себя либо одну основу $\tilde{y}_1$ некоторой словоформы, либо одно окончание — $\tilde{y}_2$.

Нетрудно убедиться, что отображение α действительно является функцией, так как последовательность операций $\eta \cup \{\tilde{y}\} \setminus \{\tilde{y}_1, \tilde{y}_2\}$ сводится к замене основы и окончания определенной словоформы самой словоформой. Такую замену можно сделать всегда однозначно. Заметим, что отображение α сюръективно, так как всякая словоформа допускает разбиение на основу и окончание (в частном случае окончание может не содержать ни одной буквы, т. е. быть «пустым»).

Установим некоторые свойства отображения α. Из того, что выполняется отношение B для прообразов, следует, что выполняется отношение A для образов

$$\eta B \eta_0 \Rightarrow \alpha(\eta) A \alpha(\eta_0). \tag{6}$$

Действительно, если η является прообразом некоторого $\xi = \alpha(\eta)$, то либо $\tilde{y}_1 \in \eta$ и $\tilde{y}_2 \in \eta$, либо $\tilde{y}_1 \notin \eta$, $\tilde{y}_2 \in \eta$. Рассмотрим первое. По определению отображения α $\alpha(\eta) = \eta \cup \{\tilde{y}\} \setminus \{\tilde{y}_1, \tilde{y}_2\}$, $\alpha(\eta_0) = \eta_0 \cup \{\tilde{y}\} \setminus \{\tilde{y}_1, \tilde{y}_2\}$. Выполнение соотношения $\eta B \eta_0$ означает, что $\eta \subseteq \eta_0$. Отсюда и из двух предыдущих равенств непосредственно следует $\alpha(\eta) \subseteq \alpha(\eta_0)$ или, что то же самое, $\alpha(\eta) A \alpha(\eta_0)$. Во втором случае $\alpha(\eta) = \eta$, $\alpha(\eta_0) = \eta_0$, справедливость свойства (6) очевидна. Следуя [3], будем называть отображение α эпиморфизмом отношения $\langle B, N \rangle$ в отношение $\langle A, M \rangle$.

Выполнение отношения A для пары образов влечет выполнение отношения B, хотя бы для одной пары их прообразов:

$$\xi A \xi_0 \Rightarrow (\exists \eta \exists \eta_0)[\eta B \eta_0]. \tag{7}$$

Из сюръективности α следует, что найдется хотя бы один прообраз элемента ξ_0, принадлежащий множеству N. Пусть это будет η_0. Тогда $\xi_0 = \eta_0 \cup \{\tilde{y}\} \setminus \{\tilde{y}_1, \tilde{y}_2\}$ по определению α. Из этого равенства находим $\eta_0 = \xi_0 \cup \{\tilde{y}_1, \tilde{y}_2\} \setminus \{\tilde{y}\}$, где $\tilde{y}_1$ и $\tilde{y}_2$ — основа и окончание словоформы $\tilde{y}$.

Вопрос о том, каким образом определенная словоформа членится на основу и окончание, не всегда решается однозначно. Например, неакцентуированная форма записи слова «пробой» позволяет выделить в нем основу «проб» и окончание «ой» (женский род, творительный падеж) или основу «пробой» и пустое окончание (мужской род, именительный падеж). Даже если необходимые признаки словоформы известны, основа может определяться в ней неодинаковым образом разными авторами. Для нас важно, что по меньшей мере одним способом разделить слово на основу и окончание можно всегда.

Рассмотрим элемент η, определенный следующим образом:

$$\eta = \begin{cases} \xi, & \text{если } \tilde{y} \in \xi \\ \xi \cup \{\tilde{y}_1, \tilde{y}_2\} \setminus \{\tilde{y}\}, & \text{если } \tilde{y} \in \xi. \end{cases} \tag{8}$$

Пусть $\eta = \xi$. Поскольку ξ при этом не содержит элемента $\tilde{y}$, то $\xi \subset \xi_0 \cup \{\tilde{y}_1, \tilde{y}_2\} \setminus \{\tilde{y}\}$. Но, как установлено выше, правая часть включения равна η_0, а левая — η. Значит, $\eta \subset \eta_0$. Отсюда следует, что, во-первых, $\eta \in N$ как частичное решение уравнения (2), во-вторых, $\eta B \eta_0$. Поскольку ξ как элемент M не может содержать элементов $\tilde{y}_1$ и $\tilde{y}_2$, не содержит их и η, так как $\eta = \xi$. Из (5) вытекает, что $\xi = \alpha(\eta)$, т. е. определенный равенством (8) элемент η является прообразом ξ и свойство (7) выполняется.

Рассмотрим случай, когда $\tilde{y} = \xi$ и $\eta = \xi \cup \{\tilde{y}_1, \tilde{y}_2\} \setminus \{\tilde{y}\}$. Как установлено выше, существует $\eta_0 = \xi_0 \cup \{\tilde{y}_1, \tilde{y}_2\} \setminus \{\tilde{y}\}$. Если $\xi \subseteq \xi_0$, то $\eta \subseteq \eta_0$ или $\eta B \eta_0$. Выясним теперь, является ли η прообразом ξ. Из определения (8) ясно, что в данном случае $\{\tilde{y}_1, \tilde{y}_2\} \subset \{y\}$. Из него же следует, что $\xi = \eta \cup \{\tilde{y}\} \setminus \{\tilde{y}_1, \tilde{y}_2\}$. Сравнивая это со второй строкой определения (5), видим, что $\xi = \alpha(\eta)$, а свойство (7) справедливо и в этом случае.

Теперь, когда предварительный анализ закончен, изложим процедуру поиска корней уравнения (1). 1. Если в условии какой-либо морфологической задачи фигурирует конкретная словоформа, определим ее основу и окончание и заменим ими словоформу в условии. Сделаем все возможные замены и получим столько частичных решений уравнений (2), сколько вариантов замены возможно. В том случае, когда значения словоформы в условии не содержится, например в задачах синтеза, переход от частичного решения уравнения (1) к частичному решению уравнения (2) происходит без изменения условия задачи. 2. Находим множество полных решений морфологического уравнения (2), включающих хотя бы по одному частичному решению, найденному в первом пункте. 3. Из основы и окончания, которые являются элементами полных решений уравнения (2), составляем словоформы, переходя таким образом к полным решениям уравнения (1).

Покажем, что, применяя эту процедуру, мы получим те и только те корни уравнения (1), которые получили бы его непосредственным решением. Для этого перепишем процедуру в формальных терминах, которые ввели ранее. 1. Находим все существующие прообразы $\eta \in N$ заданного частичного решения ξ уравнения (1). 2. Для каждого η находим все η_0, удовлетворяющие соотношению $\eta B \eta_0$, т. е. определяем множество n. 3. Для каждого множества n, найденногс в пункте 2, определяем его образ α (n) [3] во множестве M. Совокупность всех $\alpha(n)$, которую обозначим через m', и будет множеством корней уравнения (1).

Покажем, что множество m, определенное равенством (3), и m' эквивалентны. Для этого докажем справедливость двух включений: $m \subseteq m'$ и $m' \subseteq m$.

Пусть элемент $\xi_0 \in m$. Значит, выполняется соотношение $\xi A \xi_0$, где ξ — наперед заданное частичное решение. Поскольку отображение α сюръективно, во множестве B имеются прообразы элементов ξ и ξ_0. При этом в силу (7) хотя бы одна пара из них связана соотношением $\eta B \eta_0$. Поскольку в первом пункте процедуры мы находим все возможные η, а во втором —

все η_0, связанные с ними соотношением $\eta B \eta_0$, то будут найдены и элементы этой пары. Так как известно, что элемент η_0 является прообразом элемента ξ_0, то ξ_0 будет определен в третьем пункте процедуры, т. е. $\xi_0 \in m'$. Первое включение доказано.

Пусть теперь $\xi_0 \in m'$, значит, ξ_0 найден в третьем пункте процедуры. В таком случае существует элемент $\eta_0 \in B$ являющийся прообразом $\xi_0 \in A$. Этот элемент — полное решение уравнения (2)— содержит в себе такое частичное решение η, которое является прообразом заданного частичного решения уравнения (1). Это следует из содержания второго и первого пунктов процедуры.

Итак, мы установили, что $\xi_0 = \alpha(\eta_0)$, $\xi = \alpha(\eta)$, $\eta B \eta_0$. Но тогда из (6) следует, что $\xi A \xi_0$, т. е. $\xi_0 \in m$, другими словами, ξ может быть найдено путем непосредственного решения уравнения (1). Значит, $m' \subseteq m$. Таким образом, эквивалентность множеств m' и m доказана.

Мы рассуждали в предположении, что в первом пункте процедуры словоформа расчленяется лишь на две морфемы — основу и окончание. Однако можно делить словоформу на любое число частей и не обязательно на морфемы.

Список литературы: 1. *Шабанов-Кушнаренко Ю. П.* Применение метода нуль-органа в лингвистике.— В кн.: Проблемы бионики. Вып. 21. Харьков, 1978, с. 109—112. 2. *Бондарев В. М.* Алгебраическое описание задач морфологической обработки слов.— В кн.: Проблемы бионики. Вып. 21. Харьков, 1978, с. 29—34. 3. *Шрейдер Ю. А.* Равенство, сходство, порядок. М., «Наука», 1971. 256 с. 4. *Шиханович Ю. А.* Введение в современную математику. М., «Наука», 1965. 376 с.