

УДК 510.62

*М. Ф. БОНДАРЕНКО, д-р техн. наук, А. С. ЛЕВИЦКИЙ,
О. А. ЛИХАЧЕВА*

**О МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ СЕМАНТИКИ
ПРОИЗВОДНЫХ С МОДИФИКАЦИОННЫМИ ЗНАЧЕНИЯМИ**

Одной из важнейших задач деривационного анализа, входящего в технологию формализации естественного языка, является определение семантики дериватов, содержащих более одного аффикса. В статье в качестве объекта исследования выбраны существительные со значением лица, получаемые в результате деривационного процесса от всех частей речи. Производные со значением «лицо» выбраны в качестве объекта исследования не случайно. Это прекрасный материал для иллюстрации способов деривационного анализа многоаффиксных (многосуффиксных) лексических единиц, во-первых, потому, что многие существительные со значением лица образованы вследствие двойной деривации (звон — звонить — звонарь), во-вторых, это материал для последующей

деривации, поскольку существительные со значением лица могут семантически модифицироваться семей «детскость», оттенками субъективной оценки и семей «женскость», а любая семантическая модификация находит свое морфологическое отражение в виде очередного аффиксального слоя (учи-тель-ница).

Остановимся на некоторых вопросах, возникающих при попытке формализации смысла многослойных производных. Деривационный анализ любой производной лексической единицы начинается с проведения морфемных швов. Следующим шагом в акте определения смысла деривата является установление словообразовательного значения морфа, следующего сразу за корнем. Если суффикс один, то на этом акт анализа заканчивается. На языке алгебры конечных предикатов [1] вычисление семантической роли морфа может выглядеть так [2]:

$$M^i \cdot K_i^l \sim r^i, \quad (1)$$

M^i — морфема i , к которой принадлежит исследуемый морф; K_i^l — класс основ, принадлежащий морфеме M^i ; r^i — семантическая роль морфа, т. е. его конкретное словообразовательное значение:

$$r^i = \bigwedge_{l=1}^n x_l^i, \quad (2)$$

x_l^i — l -й семантический оттенок, где $i \in \{a, б, в \dots\}$ — значения оттенков смысла. Если перед нами цепочка суффиксальных морфов, то предикатом (1) мы воспользуемся лишь в конце акта анализа деривата.

После морфной препарации многослойного деривата предстоит вычислить коэффициент семантического словообразовательного класса, к которому относится первая производная основа, полученная в результате соединения корня слова с первым словообразующим аффиксом

$$M^i K_i^l \sim K_p^q. \quad (3)$$

Если следующий аффикс, с которым соединяется основа, последний, то можно воспользоваться предикатом (1) и вычислить словообразовательное значение, мотивируемое этим морфом:

$$M^p K_p^q \sim r^q. \quad (4)$$

В противном случае опять воспользуемся предикатом (3) и вычислим словообразовательный класс, в котором окажется вторая (по отношению ко второму морфу) основа.

Приведем примеры реальных машинных уравнений, записанных на языке алгебры конечных предикатов, иллюстрирующих все сказанное выше. Пусть в деривате исследова-тель-ниц(а) проведены морфемные швы. Основа исследова- принадлежит к семантическому классу K_5^6 морфемы <тель> (M^5). Тогда на языке алгебры конечных предикатов можно записать следующее:

$$M^5 K_5^6 \sim r^6 \cdot K_{31}^{80}, \quad (5)$$

$$r^1 \vee r^2 \vee \dots \vee r^{80} = 1, \quad (6)$$

где r^6 — семантическая роль 6 морфа, относящегося к морфеме M^5 . Предикат (5) описывает также условия попадания дериватов, образующих следствие соединения основ с морфами морфемы M^5 , в класс K_{31}^{80} :

$$r^6 = x_1^1 x_2^2 x_3^3 x_{11}^c x_{12}^{\phi} x_{13}^{\phi}, \quad (7)$$

$$x_1^1 \vee x_1^* \vee x_1^n = 1, \quad (8)$$

$$x_2^2 \vee x_2^n \vee x_2^0 = 1, \quad (9)$$

$$x_3^M \vee x_3^* \vee x_3^0 = 1, \quad (10)$$

$$x_{11}^0 \vee x^c = 1, \quad (11)$$

$$x_{12}^n \vee x^c \vee x^1 \vee x^1 \vee x^v \vee x^{\phi} = 1, \quad (12)$$

$$x_{13}^c \vee x_{13}^{\phi} = 1, \quad (13)$$

x_1^1 — лицо; x_2^0 — характеризуется действием; x_3^M — муж. пол., x_{11}^c — субъект действия; x_{12}^{ϕ} — профессионально производит действие; x_{13}^{ϕ} — профессия.

k_{31}^{80} - (предикат 5) — условие принадлежности деривата (в нашем примере исследователь) к семантическому словообразовательному классу K_{31}^{80} :

$$M^{31} k_{31}^{80} \sim x_5^*. \quad (14)$$

Этот предикат может иметь следующую содержательную интерпретацию. Если основа (производная или производная) принадлежит к семантическому словообразовательному классу K_{31}^{80} морфемы M^{31} <ниц(а)>, то ее соединение с морфемой M^{31} означает модификацию семантики деривата семей «женскость»:

$$x_5^* \vee x_5^2 \vee x_5^H = 1, \quad (15)$$

x_5^* — женскость; x_5^2 — жена лица; x_5^H — женскость + неодобрение.

Отметим, что предлагаемый алгоритм строился на допущении, что все семантические модификации производящей единицы формально отражены в морфемной структуре производной лексической единицы, в виде морфов. Или иными словами, семантика производной лексической единицы (содержание) однозначно связана с формой (морфемной структурой).

Рассмотрим семантический класс, основы которого соединяясь с морфом -к(а), модифицируются семей «женскость». Тут могут находиться производные лексические единицы (сосед-ка, шпионка), а также и производные (отглагольные — эмигрант-ка, отсубстантивные производные — мотоцикл-ист-ка). Этот пример хоро-

шо иллюстрирует и то обстоятельство, что в один семантический словообразовательный класс могут входить как непрямые, так и производные основы, причем последние могут быть мотивированы разными частями речи. Но главное то, что всем дериватам в нашем примере характерно формально-семантическое единство, т. е. материализация словообразовательной семантики морфов и словообразовательных шагов в виде морфемных цепочек. Эмигрант- (лицо, муж. пол., характеризуется отношением к действию, объект) -к(а)-женскость; мотоцикл- -ист-(лицо, муж. пол., характеризуется отношением к предмету) -к(а)-(женскость), наличие единства формы и содержания—соответствие смысла словообразовательной и морфологической структурам. Однако подобное соответствие в многослойных производных единицах соблюдается далеко не всегда.

Приведем пример класса основ иллюстрирующего «диссонанс» формы и содержания. Пусть это будет семантический словообразовательный класс, основы которого, соединяясь с суффиксом -ц(а) модифицируются семьей «женскость» (A_{39}^E). В этом классе могут содержаться непрямые лексические единицы (мастер-, фельдшер-, царь, и т. д.). К этому классу принадлежат также производные основы, мотивированные морфами -ец, -лец (чтец, певец, страдалец) и морфами -ик, -ник, -чик, -щик (ударник, работник, закройщик, буфетчик). Соединим предлагаемые основы с морфом -ц(а). Имеем: мастер-иц(а), цар-иц(а) — в первом случае. Во втором случае — чт-иц(а), страда-л-иц(а); в третьем случае — ударн-иц(а), работн-иц(а), закрой-щ-иц(а), буфет-ч-иц(а). Во втором и третьем случае легко видеть отсутствие единства формы и содержания. Наряду с различного рода усечениями ранее соединенных морфов (закрой-щ-иц(а)) можно наблюдать и полное отсутствие морфов, определяющих категориальную семантику мотивирующей основы (чт-иц(а), пев-иц(а)).

Предложенный пример можно отнести не к исключениям, а скорее к языковой норме, поскольку в результате исследований установлено, что одни морфы, мотивирующие «женскость», могут совсем не усекают предыдущий словообразующий аффикс (если он имеется), другие — в одних случаях усекают, а в других вовсе исключают его. В процессе моделирования семантики многослойных производных авторам показалось удобным придерживаться следующего правила в акте деривационного анализа. Если словообразующий (у нас модифицирующий) морф не исключает и не усекает предыдущий аффикс, то его удобно рассматривать в качестве однооттеночного модификатора семантики (у нас «женскость»). В противном случае этот морф считать категоризатором семантики основы и относить его к разряду неоморфов. Последнее было особенно удобно делать в том случае, когда предыдущий морф исключается вовсе. Тогда морф-модификатор можно отнести к некоторой морфеме (если это возможно), а семантическую роль, соответствующую исключенному морфу, актуализировать с соот-

ветствующей модификацией в случае присутствия нашего морфа-модификатора. На языке алгебры конечных предикатов [1] по этому поводу можно записать следующее высказывание:

$$M^a k_i^j \sim r^j x_s^k, \quad (16)$$

где M^a — морфема, к которой можно отнести морф-модификатор; k_i^j — семантический класс j , относящийся к морфеме M^i , один из морфов которой исключается в результате соединения основы с аффиксом морфемы M^a ; r^j — семантическая роль исключенного морфа, присвоенная аффиксу-модификатору.

Вернемся к рассмотрению морфов, усекающих предыдущий аффикс. Обычно после усечения в производной лексической единице может остаться одна или две буквы от усекаемого морфа. Нам было удобно отнести их к следующему (усекающему) аффиксу, и эту новую конструкцию считать морфом (неоморфом), относясь к нему, как и в предыдущем случае, к категоризатору семантики, однако содержащем в себе и модифицирующий оттенок. Для вычисления словообразовательной семантики неоморфов можно пользоваться предикатом (16), однако морфема M^a в этом случае должна содержать в своем составе и неоморф (неоморфы).

Остановимся на примерах. Рассмотрим фрагмент класса основ, относящегося к морфеме <ик>. Пусть это будут основы, актуализирующие совместно с морфом -ик, следующую семантику: «лицо, муж. пол., характеризуется отношением к предмету-явлению, профессионально занимается, ученый (специалист по...)»

хим- k_{11}^{56} ,

физ- k_{11}^{56} ,

электр- k_{11}^{55} ,

механ- k_{11}^{56} .

Понятно, что многие из предлагаемых лексических единиц могут иметь и какую-то другую семантику, кроме «ученый специалист», химик — «работник химической промышленности», механик — «лицо, наблюдающее за работой машины» и т. д. Но поскольку вопрос многозначности сейчас нас не интересует, то оставим для наглядности примера против каждой основы по одному коэффициенту класса k_i^j , т. е. будем считать, что эти основы могут совместно с аффиксом -ик актуализировать лишь по одной семантической роли, что вовсе не будет ущербно для изложения.

Пусть на вход системы поступил дериват химичка. Проведем морфемные швы следующим образом: хим-ичк(а). Дело в том, что конструкцию ичк(а) удобно считать неоморфом, способным категоризировать семантику основы. Этот морф можно отнести к морфеме -к(а). Тогда на языке алгебры конечных предикатов имеем:

$$S_{11}^- S_{12}^- S_{13}^- S_{21}^u S_{23}^r S_{31}^- S_{32}^k S_{33}^- S_{41}^- S_{42}^- S_{43}^- \sim M^{23}, \quad (17)$$

где S_{ij} — разряд регистра сегментированных суффиксов [3]; i — номер сегмента; j — номер позиции в сегменте. Тогда реальное словообразовательное значение, актуализируемое этим морфом, можно определить из уравнения

$$M^{23}k_{11}^{56} \sim r^{58} \cdot x_5^{\text{ж}}. \quad (18)$$

К достоинствам предлагаемого способа можно отнести то, что с помощью предиката (18) можно будет определить и смысл потенциальных (не принятых узусом), но понятных носителям языка лексических единиц (механичка, электр-ичка).

Предложенный способ позволяет несколько упростить алгоритм определения смысла многоаффиксных дериватов, так как идентификация усеченных морфов нам представляется несколько затруднительной.

Список литературы: 1. Шабанов-Кушнарченко Ю. П. Теория интеллекта: Математические средства. Х., 1984. 144 с. 2. Шаронова Н. В. Математические модели суффиксального словообразования и их использование для автоматической обработки отглагольных имен существительных в текстах русского языка. Автореф. дис. ... канд. техн. наук. Х., 1984. 222 с.

Поступила в редколлегию 15.03.89