

**О НЕКОТОРЫХ ВОПРОСАХ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ  
СМЫСЛА СУФФИКСАЛЬНЫХ МОРФЕМ**

Решение задачи естественной языковой коммуникации человека и ЭВМ невысказимо без автоматизированной обработки текстовой информации. Настоящая статья посвящена задаче моделирования смысла имен существительных со значением «лица», получаемых в процессе словообразования от всех частей речи.

Как известно, существительные со значением лица могут быть образованы путем аффиксации от глаголов (*писатель*), прилагательных (*толстяк*), существительных (*химик*). Имеют место также образования, семантически мотивированные словосочетаниями с мотивирующими прилагательными в качестве определения (*буровой рабочий — буровик*). Будем придерживаться мнения [1], что производные существительные со значением лица, соотносимые со словосочетаниями, удобнее рассматривать как семантически образованные от существительных (глаголов), мотивирующих соответствующие прилагательные. Целесообразность данного подхода подтверждается, например, следующими доводами [1]: только таким образом может быть поставлена на практике задача синтеза значения слова на основе значений составляющих его морфов. Так, в слове *буровик* нет никаких следов значения «мастер», зато налицо морфемы, обозначающие лицо и характер действия, определяющего это лицо; некоторые производные существительные со значением лица могут рассматриваться и как семантически мотивированные соответствующими существительными или глаголами. Например, *портовый рабочий — портовик; рабочий порта — портовик; работник рыбной промышленности — рыбак; специалист по рыбе — рыбак*. Остановимся на ряде обстоятельств в поддержку изложенного подхода и дополняющих техническую сторону вопроса.

При решении задачи анализа деривата часто приходится сталкиваться с многослойными производными структурами, т. е. с дериватами, получаемыми в результате многократной суффиксации. Русский язык, как известно, весьма богат подобными словообразовательными средствами. Для определения смысла такого рода сложной структуры дериват необходимо препарировать, причем начинать процесс с аффикса, прикрепленного последним. В результате проведенных действий (отсечение словообразующего морфа) мы должны получить мотивирующую основу и словообразующий аффикс. Для определения дополнительной семантики, возникающей в результате подключения этого аффикса, необходимо знать семантику основы и ее частеречную принадлеж-

ность. Поскольку основа сложная, не совпадающая с корнем (рассматривается многослойный дериват), то в машинном словаре корневых морфов она отсутствует, и процесс препарации придется продолжить. В этом случае возможна некоторая минимизация описанного алгоритма, а, следовательно, и всей стратегии анализа смысла. Например, рассмотрим две равноправные дефиниции: «портовый работник», «работник порта», за которыми, тем не менее, стоят разные формально-семантические связи с производными а) *портов-ик* и б) *порт-овик* соответственно. Основа (б) *порт-* совпадает с корневым морфом и ее можно найти в машинном словаре. С основой (а) *портов-* дело обстоит иначе, тут придется продолжить процесс препарации. Отсекаем еще один аффикс *порт-ов*. Суффикс *-ов/-ев*, соединяясь с основами существительных (в данном случае *порт-*), имеет значение «принадлежащий тому, что названо мотивирующим словом». Суффикс *-ик*, соединяясь с основой (*портов-*), раскрывает смысл принадлежности. Итак, имеем полный смысл «лицо, профессионально занимается, место деятельности». Рассмотрим теперь дефиницию «работник порта». Она соответствует словообразовательной структуре *порт-овик* и заключает в себе семантическую трансформацию *порт* — *портовик*. Во втором случае имеем *порт* — *портовый* — *портовик*, т. е. налицо лишняя словообразовательная ступень при одинаковом толковании.

Из сказанного выше можно сделать вывод, что если деривату присуща множественность словообразовательной структуры, то его следует толковать производящим структурно менее сложным. И действительно, ведь морфы *-овик*, *-евик*, *-ник*, *-ик* можно объединить в одну морфему, ибо все перечисленные морфы способны актуализировать одинаковую дополнительную семантику. Можно привести пример уравнения алгебры конечных предикатов [2], описывающего формальную модель морфемы <ник> с помощью лингвистического регистра [3]:

$$s_{11}^* s_{12}^- s_{13}^- s_{21}^- s_{22}^- s_{23}^- s_{31}^* s_{32}^- s_{33}^- s_4^- \vee s_{11}^- s_{12}^- s_{13}^- s_{21}^- s_{22}^- s_{23}^- s_{31}^* s_{32}^- \wedge s_{33}^- s_4^- \vee \vee s_1^- s_{21}^- s_{22}^- s_{23}^- s_{31}^* s_{32}^- s_{33}^- s_4^- \sim M^3. \quad (1)$$

Таким образом, система должна быть ориентирована на узнавание суффикса, как можно большей длины. Остановимся еще на некоторых аспектах проблемы толкования смысла существительных со значением лица.

Основу называния производного слова «...составляют главные ономаσιологические категории языка, и среди них категории предметности, процессуальности и признаковости» [4]. Можно привести примеры работ [3, 5], в которых авторы успешно членили главные понятийные ономаσιологические категории на субкатегории, еще раз подтверждая этим то обстоятельство, что словообразовательное значение — упорядоченная и сложно организованная структура.

Наша задача состоит в том, чтобы на основе предлагаемых уровней абстракции сконструировать модель словообразовательного значения того или иного деривата; рассмотреть дериваты, ономаσιологический базис которых приходится на формантную (аффиксальную) часть мотивированного слова, т. е. на тот или иной суффиксальный морф. Последний, соединяясь с основой, выполняет задачу категоризации семантики деривата. Считая словообразовательное значение продуктом взаимодействия смыслов основы и словообразующего аффикса, авторы все же из чисто практических соображений основной объем словообразовательного значения переложили на аффикс. В результате такой операции суффикс перестал нести одну лишь чисто категориальную нагрузку (ономаσιологический базис) и стал вместилищем смысла, учитывающим все субкатегориальные членения (той или иной категории). Все дополнительные (субкатегориальные) элементы смысла, характерные лишь для конкретной словообразовательной модели, удобно было назвать дополнительной семантикой. Смысл этого понятия очевиден из формул

$$C_{сл} = C_{осн} + C_{суф} + C_{доп}; \quad r = C_{суф} + C_{доп}, \quad (2)$$

где  $C_{сл}$ ,  $C_{осн}$  — смысл слова и основы;  $C_{суф}$  — инвариантный смысл суффикса;  $C_{доп}$  — дополнительный элемент смысла, продукт семантического взаимодействия основы и аффикса;  $r$  — полная семантика суффикса, семантическая роль.

Ранее [6] была получена формула для определения полной семантики словообразовательной морфемы  $M^i k^j \sim r^l$  (3), где  $r^l = \bigwedge_{i=1}^n x_i^{\sigma} e$ ,  $x_i^{\sigma}$  —  $l$ -й оттенок смысла в значении  $\sigma$ ,  $k_i^j$  — коэффициент, идентифицирующий  $j$ -й семантический класс основ для  $i$ -й словообразовательной морфемы  $M^i$ .

Рассмотрев 49 деривационных аффиксов, образующих существительные со значением лица, авторы обнаружили, что для многих суффиксов в качестве мотивирующих могут выступать основы двух (*арест-ант*, *музык-ант*) и даже трех (*бег-ун*, *толст-ун*, *горб-ун*) частей речи. Последнее обстоятельство значительно влияет на категориальную принадлежность деривата. Все исследуемые производные, имея общий элемент смысла в зависимости от частеречной принадлежности основы, разделились на три большие группы (признак  $x_1$ ):  $x_1^1$  — «лицо, характеризуемое отношением к признаку»,  $x_1^0$  — «лицо, характеризуемое отношением к предмету»,  $x_1^2$  — «лицо, характеризуемое отношением к действию». Область определения этого признака запишем с помощью уравнения  $x_1^1 \vee x_1^0 \vee x_1^2 = 1$  (4).

Тип отношения «лицо, характеризуемое отношением к...» конкретизируют признаки  $x_2, x_3, \dots, x_{20}$ , определяющие дополнительную семантику. Группы основ по частеречной принадлежности выделены в объеме «Грамматического словаря русского языка»

7]. Далее, внутри каждой из групп выделены классы основ с тем, чтобы любая основа класса была способна соединяться лишь с одной из 49 морфем. Всего таких классов 49. Однако среди основ встречались и такие, которые могли сочетаться с разными словообразовательными морфемами, например, *краснодерев-ец*, *краснодерев-щик*; *атом-ник*, *атом-щик* и т. д. Внутри каждого класса основ (пронумерованных в соответствии с рассматриваемым множеством суффиксов) выделены дополнительные группы по их семантическим свойствам — семантические классы. Особенностью семантического класса является то, что любая его основа способна актуализировать одну и ту же дополнительную семантику в пределах словообразовательной модели. В результате проведенных таким образом группировок основ алгоритм определения дополнительной семантики модифицировался следующим образом (остановимся на примере однослойных суффиксальных структур). 1. Исследуемая лексическая единица членится на словообразовательный аффикс и мотивирующую основу. 2. Имея текст суффиксального морфа, определяем, к какой морфеме он принадлежит. 3. По тексту морфемы находим номер группы основ, с которыми она может контактировать. 4. Пользуясь текстом основы, определяем ее принадлежность семантическому классу внутри группы. 5. Зная семантический класс основы и текст суффиксальной морфемы, вычисляем набор семантических компонентов роли. Приведенный вариант алгоритма упрощен: опущен ряд шагов, например распределение букв суффиксального морфа в лингвистическом регистре сегментированных суффиксов; шаг распределения букв корневого морфа; коррекция ошибок и т. д. Однако общая картина определения дополнительной семантики ясна.

В процессе построения математической модели авторам пришлось столкнуться с проблемой лексической многозначности дериватов, исследовать ее источники. Если производная единица обладает множественностью словообразовательной структуры, значит в процессе членения можно выделить несколько формантных частей и несколько производящих частей (см. приведенный выше пример: *портовик*). Очевидно, что варианты формального членения влекут за собой и некоторые семантические вариации деривата (назовем такие дериваты полиструктурными). Рассмотрим примеры: *туберкулез-ник* — «лицо; характеризуется отношением к предмету/явлению; профессионально занимается», либо «лицо; характеризуется отношением к предмету/явлению; находится в состоянии (болеет)»; *туберкулезн-ик* — «лицо; характеризующееся свойством (болезнь)». Явление многозначности можно наблюдать и у моноструктурных дериватов, но в последнем случае причина многозначности в множественности мотивации; *клеветать* — *клевет-ник* — «лицо; характеризуется отношением к действию; субъект действия; склонно производить действие»; *клевета* — *клевет-ник* — «лицо; характеризуется отношением к предмету; склонно заниматься предметом».

К основным источникам лексической многозначности производных единиц можно отнести множественность словообразовательной структуры, множественность мотивации, омонимию и полисемию аффиксов. Модель лексической многозначности можно получить, не затрагивая причин этого явления, а изучив их, промоделировать следствие, т. е. воспроизвести результат (многозначность) в виде предикатов, исчисляющих смыслы. Покажем на примере, как можно определить несколько наборов семантических ролей  $r^l$ . Рассмотрим фрагмент машинного словаря корневых морфем, соединяющихся с морфемой <ник>:

туберкулез-	$k_3^{54}$	$k_3^{65}$
клевает-	$k_3^1$	

Для наглядности примера будем считать, что основа туберкулез- хранится в машинном словаре, хотя на самом деле это производная единица: *туберкул-ез*.

Каждая основа словаря содержит набор семантических коэффициентов. Верхний индекс коэффициента класса основы определяет номер семантической роли, которую может актуализировать эта основа, соединившись с морфем (в нашем примере <ник>). Нижний индекс — это порядковый номер словообразовательной морфемы. Наличие у одного корневого морфа или основы нескольких коэффициентов говорит о лексической многозначности будущей производной единицы. Морфема <ник> имеет в своем составе следующие морфы: -ник(-еник) -ик в отглагольных образованиях; -ник/-атник/арник/-овник в отсубстантивных образованиях. Формальная модель морфемы -ник будет выглядеть следующим образом:

$$\begin{aligned}
 & S_1^- s_{21}^- s_{22}^n s_{23}^n s_{31}^n s_{32}^n s_{33}^- S_4^- \vee s_{11}^e s_{12}^- s_{13}^- s_{21}^- s_{22}^n s_{23}^n s_{31}^n s_{32}^- s_{33}^- S_4^4 \vee \\
 & \vee S_1^- S_2^- s_{31}^n s_{32}^n s_{33}^- S_4^- \vee s_{11}^n s_{12}^r s_{13}^- s_{21}^- s_{22}^n s_{23}^n s_{31}^n s_{32}^- s_{33}^- S_4^- \vee \quad (5) \\
 & \vee s_{11}^n s_{12}^p s_{13}^- s_{21}^- s_{22}^n s_{23}^- s_{31}^n s_{32}^n s_{33}^- S_4^- \vee s_{11}^0 s_{12}^n s_{13}^- s_{21}^- s_{22}^n s_{23}^n s_{31}^n s_{32}^- s_{33}^- S_4^-.
 \end{aligned}$$

Поскольку основа *туберкулез* — входит в два семантических класса, то уравнения смысла для лексемы *туберкулезник*, описывающие две семантические роли «больной», «врач», соответственно можно записать

$$M^3 k_3^{65} \sim x_0^l x_1^o x_3^m x_5^b \quad (6), \quad M^3 k_3^{54} \sim x_0^l x_1^o x_2^p x_3^m x_4^n. \quad (7)$$

Расшифруем значения оттенков смысла, фигурирующих в выражениях (6), (7):  $x_1^l$  — «лицо»;  $x_1^o$  — «характеризуется отношением к предмету/явлению»;  $x_2^p$  — «профессионально занимается»;  $x_3^m$  — «предмет, объект занятий»;  $x_3^n$  — «мужской пол»;  $x_5^b$  — «болеет».

Таким образом, построены наборы смысловых компонентов, характеризующих семантические проявления морфемы <ник> в двух различных случаях, т. е. получена модель лексической многозначности деривата. Толкование словообразовательной структуры *туберкулезн-ик* опущено, так как такое членение влечет лишнюю ступень деривационного анализа. Рассмотрим дериват *клевет-ник*. Считаем, что лексема мотивирована субстантивной основой *клевет-* по двум причинам. Во-первых, *клевет-а-ть* производное от *клевет(а)*, во-вторых, поскольку различные дефиниции заключают в себе одну и ту же семантику, нет смысла скрывать обе.

Рассмотренные приемы позволяют минимизировать обобщенный алгоритм формирования смысла суффиксальных морфов; использование предложенного варианта машинного словаря и алгоримализация языковых закономерностей дает возможность построения модели лексической многозначности.

Источники литературы: 1. Милославский И. Г. Вопросы словообразовательного анализа. М., 1980. 296 с. 2. Шабанов-Кушнарченко Ю. П. Теория интеллектуально-математические средства. Х., 1984. 144 с. 3. Шаронова Н. В. Математические модели суффиксального словообразования и их использование для автоматизированной обработки отглагольных имен существительных в текстах русского языка. Х., 1984. 213 с. 4. Улуханов И. С. Словообразовательная семантика русского языка и принципы ее описания. М., 1977. 256 с. 5. Бондаренко М. Ф. Математические модели морфологических и фонетических отношений и их применение для автоматизации обработки речевых сообщений. Х., 1984. 350 с. 6. Левицкий А. С., Шаронова Н. В., Рябова Н. В. О математическом моделировании деривационного анализа суффиксальных существительных со значением лица. К., 1988. 15 с. Деп. в УкрНИИТИ. 15.03.88. № 1313 — Укр88. 7. Зализняк А. А. Грамматический словарь русского языка. М., 1977. 880 с.

Поступила в редколлегию 11.04.88

ДК 510.62