

Н. В. ШАРОНОВА, Э. М. БУЗНИЦКАЯ, канд. техн. наук

О СТРУКТУРЕ СИСТЕМЫ ПРИЗНАКОВ ПРИ МОДЕЛИРОВАНИИ СЛОВООБРАЗОВАНИЯ

Одним из самых важных этапов при решении задач переработки естественной языковой информации является введение и обоснование системы признаков, адекватно отображающей связи и отношения в языке. Признаки — это формализованное представление тех смысловых значений, которые передает та или иная часть текста. Если в морфологии (словоизменении) достаточно много сделано в области выявления и формирования признаков (род, число, падеж — эти категории общеизвестны и узаконены), то в словообразовании все известные работы носят фрагментарный, специальный характер, т. е. нет таких исследований, где была бы предложена общая система достаточно формализованных семантических и грамматических признаков, позволяющая решать задачи анализа и синтеза словоформ на уровне словообразования.

Этот факт — лишнее свидетельство сложности поставленной задачи. Однако, чтобы решать задачи еще более высокого уровня, связанные с семантикой словосочетаний и предложений, невозможно обойтись без решения задач внутри слова, связанных со словообразованием. Здесь предлагается система признаков, общая для образования всех отглагольных существительных русского языка, позволяющая описывать средствами алгебры конечных предикатов [1] более 150 суффиксальных словообразовательных типов, даются рекомендации по расширению возможностей этой системы при описании всей русской суффиксации.

лучше и точнее описать систему более мелких отношений, затем собрать их в общее морфологическое отношение с применением теоремы о разложении.

Введенная нами система признаков состоит из тридцати семантических и традиционно грамматических признаков, позволяющих описать все отглагольные существительные. Рис. 1 отражают структуру введенных признаков. Видно, что эти признаки определенным образом группируются. В первую группу (рис. 1) входят традиционные, «классические» признаки: x_2 — вид морфемы (корень, суффикс, префикс, окончание); x_3 — часть

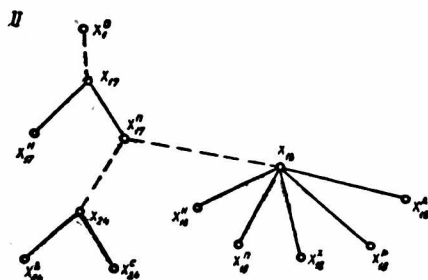


Рис. 2

речи (существительное, прилагательное, глагол, числительное, наречие); x_4 — род (мужской, женский, средний); x_{25} — число (единственное, множественное); x_{21} — современность (современный, устаревший); x_7 — происхождение (русский, иностранный), а также су

губо словообразовательный признак x_6 — часть речи мотивирующей основы (существительное, прилагательное, глагол и др.). Возглавляет эту группу признак x_1 — вид морфологических преобразований с значениями словоизменение, словообразование и формообразование. В этой группе связи между признаками не носят характера строгой иерархии, имеется некоторая свобода перестановки признаков. Подчиненность на рис. 1 указана пунктиром.

В группу II (рис. 2) входят признаки, связанные со словообразовательной продуктивностью. Эта группа признаков свойственна лишь словообразованию, поэтому ее истоком является x_1^0 — вид морфологических преобразований со значением «словообразование». Словообразовательный тип может быть продуктивным (x_{17}^P) или непродуктивным (x_{17}^N), при этом если продуктивный — то с высокой (x_{24}^B) или слабой (x_{24}^C) продуктивностью и в каких текстах: x_{18}^H — научно-техническая терминология, x_{18}^N — наименования профессий, x_{18}^X — художественно-публицистическая речь, x_{18}^P — разговорная речь, x_{18}^D — просторечие, x_{18}^C — специальные термины.

Группа III — группа признаков имени существительного, IV — глагола. Объединяет эти группы то, что рассматриваются отглагольные существительные, т. е. образования с признаками как глагола, так и имени. В группе III (рис. 3, а) происходит конкретизация значения имени существительного по признакам: x_{28} — отношение к процессуальному признаку (носитель процессуального признака или отвлеченный процессуальный признак), далее, если

ситель — то конкретный (x_8^k) или абстрактный (x_8^a), если конкретный — то предмет (x_{10}^n) или вещество (x_{10}^b), одушевленный (x_5^o) или неодушевленный (x_5^h), один (x_{23}^o) или совокупность (x_{23}^c) и т. д. группе IV (рис. 3, б) происходит конкретизация глагольных ойств: x_{14} — признак отношения к действию (x_{14}^c — субъект, x_{14}^o — объект действия, x_{14}^n — средство для производства действия, x_{14}^b — результат действия, x_{14}^m — помещение для его производства); x_5 — склонность к действию (склонный — не склонный).

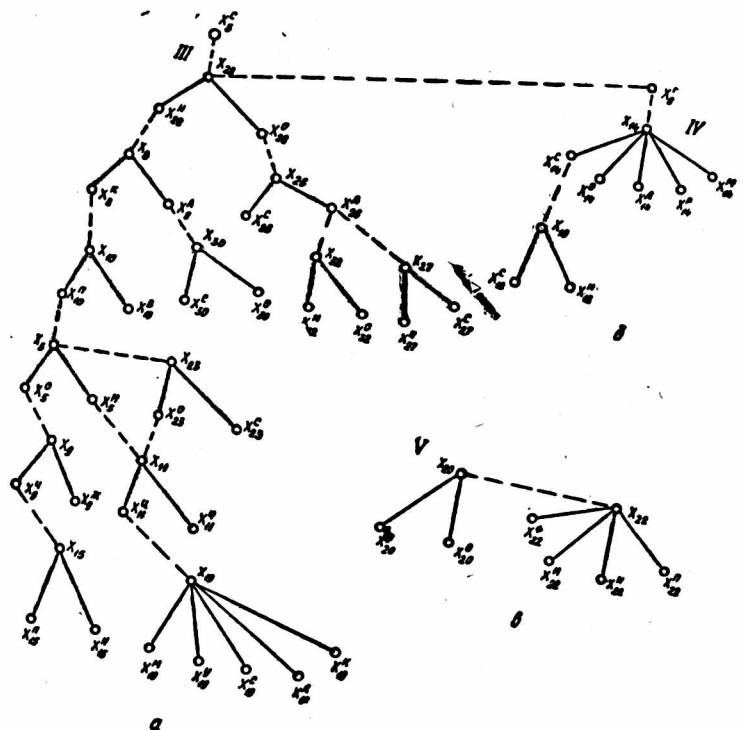


Рис. 3

Пятую группу составляют признаки экспрессивной окраски (рис. 3, в). Для отглагольной группы существительных эти признаки не играют существенной роли, они значительно более разиты у модификационных существительных, мотивированных существительными.

Предлагаемая система признаков обладает определенной универсальностью. При рассмотрении словообразования существительных, мотивированных не глаголами, а какой-либо другой частью речи, изменения потребовалось бы внести лишь в IV и, возможно, в V группы признаков. Глагольная группа признаков

заменялась бы признаками мотивирующей части речи, причём тоже наблюдалось бы пересечение с группой признаков существительного по отношению к предметному (мотивация имен существительным), непроцессуальному (мотивация прилагательным) или количественному (мотивация числительным) признакам. Таким образом, введенная система признаков нуждается лишь в дополнении при переходе к другой мотивации. При смене части речи групповая структура признаков сохраняется, но сами признаки в группах III, IV должны соответствовать тем частям речи, к которым относятся мотивируемое и мотивирующее слова.

Продемонстрируем возможности системы признаков, описанные уравнениями отглагольные имена существительные мужского рода. В этих уравнениях присутствуют общие для групп суффиксы множители:

$$T_1 = x_1^0 x_2^c x_3^c x_4^m x_5^r x_6^k x_7^p x_8^p; \quad T_2 = x_1^0 x_2^c x_3^c x_4^m x_5^r x_6^p x_7^k x_8^p; \quad T_3 = x_1^0 x_2^c x_3^c x_4^m x_5^r x_6^h x_7^k x_8^p$$

смысл которых сводится к следующему: рассматривается следующее образование (x_1^0) суффиксальное (x_2^c) имен существительных (мужского рода (x_4^m) , отглагольных (x_5^r) , конкретных (x_6^k) , русского (x_7^p) для T_1 и T_2) и иностранного (x_7^h) для T_3) происхождения продуктивных (x_1^p) для T_1 и T_3) и непродуктивных (x_1^h) для T_2 типов.

Запишем уравнение:

$$T_1 (x_5^0 x_4^c (x_1^p x_18^p \vee x_15^h x_18^h) \vee x_5^h (x_14^c x_18^h \vee x_14^c x_18^h)) \supset \langle \text{тель} \rangle.$$

Оно означает, что морфема $\langle \text{тель} \rangle$ появляется в одушевленных именах существительных, обозначающих субъект действия, при этом образование таких имен в значении «профессия» продуктивно в профессиональной терминологии, а «временное занятие» (не профессия — x_{15}^h) — в художественно-публицистической речи; кроме того, морфема $\langle \text{тель} \rangle$ появляется в неодушевленных именах существительных для обозначения орудия выполнения действия с продуктивностью типа в научно-технической литературе и неодушевленного субъекта действия с продуктивностью там же. Распределение морфов в этой и других морфемах описано в [2] с использованием чисто формальных признаков.

Уравнение

$$T_1 (x_5^0 x_4^c (x_9^c x_18^p \vee x_9^h x_18^p) \vee x_5^h ((x_14^c \vee x_14^0) x_18^p \vee x_14^c x_18^p)) \supset \langle \text{ник} \rangle$$

означает, что морфема $\langle \text{ник} \rangle$ появляется в одушевленных именах существительных, обозначающих человека — субъекта действия с продуктивностью типа в профессиональной терминологии или животное — субъекта действия с продуктивностью разговорной речи, а также в неодушевленных именах, обозначающих субъект или объект действия (продуктивны в разговорной речи) или орудие, средство для выполнения действия (п

продуктивны в научно-технической терминологии, в сфере названий орудий и механизмов). Примеры: *заточник, клеветник, ощупник* (жук), *окучник, зимовник, мученик*.

Для морфемы <щик> запишем следующее уравнение:

$$T_1(x_5^0 x_{14}^c (x_{15}^n x_{18}^n \vee x_{15}^n x_{18}^p \vee x_5^n (x_{14}^c \vee x_{14}^n) x_{18}^n)) \tau_1 \supset \langle \text{щик} \rangle. \quad (3)$$

Уравнение (3) очень похоже на уравнение (1), так как морфема <щик> передает те же общие значения, что и морфема <тель>. Разница же заключается в условии τ_1 в уравнении (3), обозначающем возможность двойной мотивации в случае с морфемой <щик>. Известно, что большинство слов, образованных с помощью морфемы <щик>, мотивируется одновременно существительными с суффиксом $-к(a)$, называющими действие и мотивированными теми же глаголами: *наладчик* → *наладить* и *наладка*, *забастовщик* → *забастовать* и *забастовка* и др. У морфемы <тель> такой двойственной мотивации нет:

$$T_1 x_5^0 x_{14}^c (x_{18}^n \vee x_{18}^p) \supset \langle \text{льщик} \rangle \quad (4), \quad T_1 x_5^n x_{14}^n (x_{18}^n \vee x_{18}^p) \supset \langle \text{льник} \rangle \quad (5).$$

Уравнение (4) описывает появление в одушевленных именах существительных, обозначающих субъект действия и продуктивных в профессиональной терминологии и разговорной речи, морфемы <льщик>.

Уравнение (5) описывает появление морфемы <льник> в неодушевленных именах существительных, обозначающих орудие, средство для выполнения действия и продуктивных в научно-технической терминологии и разговорной речи.

Для морфемы <ец> используется общий множитель T_2 (тип непродуктивен):

$$T_2(x_5^0 x_{14}^c \vee Q_1 x_5^n x_{14}^c \vee Q_2 x_{14}^p) \supset \langle \text{ец} \rangle, \quad (6)$$

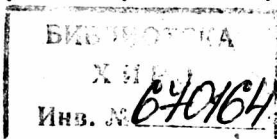
Это уравнение будет описывать появление непродуктивного типа морфемы <ец> в именах существительных одушевленных, называющих деятеля: *творец, пловец, писец*, а также соединяющегося с основами из списка Q_1 {руб-, став-, постав-, рез-, свет-, пада-, стла-, шлепа-, оста-} со значением «неодушевленный субъект действия» и с основами из списка Q_2 {питом-, скоп-, руб-} со значением «объект действия или его результат» (*рубец*):

$$T_1 x_5^0 x_{14}^c (x_{18}^p \vee x_{18}^x) \supset \langle \text{лец} \rangle. \quad (7)$$

Существительные с суффиксом $-лец$ называют лицо, производящее действие, названное мотивирующим словом: *страдалец, нескиталец, владелец* и т. п. Тип проявляет продуктивность в разговорной и художественной речи:

$$T_2(x_5^n Q_3 \vee x_5^0 Q_4) \supset \langle \text{ак} \rangle, \quad (8)$$

где Q_3 {держ-, леж-, лиз-, рез-, черп-, тес-, сто-, торч-, толл-}, Q_4 {вод(ж)}; $T_1(x_5^0 x_{14}^c (x_{18}^p \vee x_{18}^n) \vee x_5^n x_{14}^n x_{18}^p) \supset \langle \text{ач} \rangle \quad (9)$.



Существительные с суффиксом *-ач* называют предмет, служащий для выполнения действия (эти образования продуктивны в разговорной речи и просторечии), или лицо, производящее действие (продуктивность в разговорной речи):

$$T_1 (x_5^o x_{14}^c (x_9^a x_{18}^p \vee x_9^* Q_5) \vee x_5^h x_{14}^c Q_6) x_{16}^c \supset \langle \text{ун} \rangle. \quad (10)$$

Существительные с суффиксом *-ун* называют человека, производящего действие (продуктивность в разговорной речи): *бегун, прыгун, опекун, врун, вертун* и т. п., животное из списка Q_3 {грыз-, пест-, полоск-, рев-, скак-, бод-, писк}, орудия и части механизмов из списка Q_6 {кол-, шат-, полз-}, все это с оттенком «склонный к действию»:

$$T_2 (x_5^o x_9^a x_{14}^c \vee x_5^h (x_{14}^d \vee x_{14}^c \vee x_{14}^m)) \supset \langle \text{ок I} \rangle;$$

$$T_1 (x_5^o (x_9^a \vee x_9^*) \vee x_5^h) x_{14}^p x_{18}^p \supset \langle \text{ок II} \rangle. \quad (11)$$

Эти уравнения описывают два подтипа суффикса *-ок*, причем этом второй подтип продуктивен в разговорной речи и обозначает результат действия (человек, животное или неодушевленный предмет): *подарок, обрубок, переросток* и т. п.; первый подтип непродуктивен и описывает несколько слов со значением а) лицо — производитель действия (*едок, седок, игрок*); б) орудие, средство для выполнения действия (*скребок, манок* и др.); в) неодушевленный предмет — субъект действия (*росток, поплавок, движок*); г) место действия (*каток*):

$$T_2 x_5^o x_9^a Q_7 \supset \langle \text{арь} \rangle. \quad (12)$$

Существительные этого непродуктивного типа описывают лицо, производящее действие, причем основы берутся из списка Q_8 {пек-, сек-, звон-, лек-, пах-, пис-, ток-}: $T_2 x_5^o x_9^a Q_8 \supset \langle \text{аль} \rangle$ (13)

То же значение, что и в предыдущем типе, но основы берутся из списка Q_8 {вр-, кат-, стриг-, ков-, мак-, строг-}. Тип также непродуктивный.

Следующие четыре словообразовательных типа характеризуются продуктивностью в научно-технической терминологии (-атор (-ятор) -итор) и в специальной терминологии (-ор (-ер) ёр -ант (-янт) -ент, -ат (-ят)), причем суффиксы этих типов отличаются тем, что они присоединяются в подавляющем большинстве к основам иностранного происхождения (исключения редки: *ухажёр*). Для этих четырех продуктивных типов общим множителем является T_3 :

$$T_3 (x_5^o x_9^a x_{14}^c \vee x_5^h x_{14}^d) x_{18}^h \tau_2 \supset \langle \text{атор} \rangle. \quad (14)$$

Это уравнение описывает появление морфемы $\langle \text{атор} \rangle$ в существительных, называющих лицо, производящее действие или предмет, предназначенный для его осуществления, при этом тип обнаруживает продуктивность в научно-технической терминологии. Условие τ_2 означает, что существительные этого типа

одновременно мотивируются отглагольными существительными с морфемой <ация>: *организатор* → *организовать* и *организация*, *инвеститор* → *инвестировать* и *инвестиция*.

Уравнение

$$T_3 (x_5^o x_9^a x_{14}^c \vee x_5^h x_{14}^c) x_{18}^c P_1 \supset \langle \text{ор} \rangle \quad (15)$$

описывает появление морфемы <ор> в именах существительных, называющих лицо или орудие, производящее действие, названное мотивирующим словом. Тип продуктивен в специальной терминологии.

Уравнение

$$T_3 (x_5^o x_9^a x_{14}^c \vee x_5^h x_{14}^c) x_{18}^c P_2 \supset \langle \text{ант} \rangle \quad (16)$$

полностью совпадает с предыдущим уравнением (15), так как морфемы <ор> и <ант> имеют одно и то же общее значение и продуктивны в тех же текстах. Единственное, что их различает,— это возможность присоединения к разным основам: если присоединяется, например, <ант>, то невозможно присоединить <ор>, и наоборот: *консультант*, *проектант*, но *дирижер*, *тренер*. Это различие зафиксировано введением в уравнения (15), (16) переменных P_1 , P_2 соответственно, обозначающих классы основ, к которым присоединяются данные суффиксы. В простейшем варианте P_1 и P_2 — это списки основ, но в общем случае P_1 и P_2 можно выразить как определенную закономерность через признаки основы, включая семантические.

Уравнение

$$T_3 (x_5^o x_9^a (x_{14}^o \vee x_{14}^p) \vee x_5^h (x_{10}^n (x_{19}^k \vee x_{19}^d) \vee x_{10}^b)) x_{18}^c \supset \langle \text{ат} \rangle \quad (17)$$

описывает появление морфемы <ат> в существительных, называющих лицо или предмет (единичный или вещество, сочинение или документ), являющиеся результатом или объектом действия: *делегат*, *концентрат*, *фильтрат*, *экспонат*, *реферат*, *дубликат*, *аттестат* и т. п. Тип обнаруживает продуктивность в специальной терминологии, многие слова встречаются в художественно-публицистической речи.

Приведенные выше уравнения (1—17) используют около двадцати признаков системы и описывают семнадцать продуктивных и непродуктивных типов суффиксальных отглагольных существительных мужского рода. Остальные признаки используются при дальнейшем описании словообразовательной системы отглагольных имен существительных, включающей в себя конкретные и абстрактные существительные мужского, женского, среднего и общего рода и выраженной более чем ста пятьюдесятью словообразовательными типами.

Список литературы: 1. Бондаренко М. Ф., Шабанов-Кушнаренко Ю. П. О математическом описании естественного языка.— Проблемы бионики, 1981, вып. 27, с. 9—13. 2. Шаронова Н. В. Формализация распределения морфов в морфеме.— АСУ и приборы автоматики, 1981, вып. 60, с. 82—85.

Поступила в редколлегию 16.03.82.