

М. Ф. БОНДАРЕНКО., канд. техн. наук, Н. В. ШАРОНОВА

О МАТЕМАТИЧЕСКОМ ОПИСАНИИ ПРОЦЕССОВ СЛОВООБРАЗОВАНИЯ

Для эффективного использования возможностей электронной вычислительной техники необходимо, чтобы ЭВМ могла воспринимать и обрабатывать информацию, представленную на естественном языке. В связи с этим возникла настоятельная необходимость в создании действующих моделей языка. Исследование языка и поиски способов формализации языковых структур необходимы не только для машинного перевода, но и для других более общих задач переработки информации с помощью ЭВМ.

В ряде практических задач (таких, как машинный перевод, машинное реферирование и аннотирование, перевод с естественного языка на формально-логические и т. п.) необходимо дать систему явно выраженных правил и записать эти правила с помощью специального аппарата так, чтобы ЭВМ могла воспринимать и обрабатывать вводимую языковую информацию.

Современное состояние лингвистических исследований таково, что «дальнейшее отсутствие практики, т. е. действующих систем машинного перевода (МП), тормозит не только его развитие, но и разработку самой лингвистической теории. Поэтому и необходимо рациональное сочетание теории с практикой и усиление прикладных лингвистических исследований в интересах реализации МП и решения других актуальных задач, связанных с автоматизацией информационных процессов в народном хозяйстве страны» [1, с. 49].

Одним из наименее исследованных с точки зрения формализации аспектов языка является словообразование.

Словообразование — один из самых важных и сложных разделов языкознания. Словарный состав языка чувствителен к человеческой деятельности и изменяется почти непрерывно. Непрерывный процесс появления новых слов осуществляется путем заимствования из других языков и путем собственно словообразования различными способами, причем значение и удельный вес последнего пути образования новых слов неизмеримо выше первого, хотя и многочисленного по составу разрядов слов. Словарный состав языка обогащался ранее и обогащается в настоящее время преимущественно путем деривации, т. е. образования новых слов на основе и с использованием того материала, который в языке уже имеется. Именно этим и определяется огромное значение словообразования в общей системе языка [2]. Суффиксальное словообразование относится к морфологическому словопроизводству и является «наиболее продуктивным, творческим и моделированным» [2, с. 253].

Одна из основных трудностей при попытках формального описания словообразовательных процессов заключается в тесной связи словообразования и семантики. «Тут непочатый край работы, необходимый для продвижения вперед в деле формализации лингвистических фактов и дальнейшей автоматизации их анализа и синтеза» [3, с. 52].

В самих словообразовательных процессах заложены предпосылки для успешного хотя бы частичного их формального описания. Например: «Несмотря на все многообразие лексических значений отдельных слов, входящих в словообразовательные типы, можно выделить повторяющиеся в ряде слов, наиболее распространенные значения, наслаивающиеся на значения типа. Их можно назвать частными словообразовательными значениями, они составляют более узкие единицы — семантические подтипы в рамках типа» [4, с. 40]. Важную роль при этом в развитии словообразовательной системы русского языка советской эпохи играют процессы увеличения регулярности словообразовательных типов. Особо важную роль играют два обстоятельства: 1) стандартность значения аффиксов в сочетании с определенным типом основ; 2) строгая закономерность сочетания аффикса с определенным типом основ (т. е. ограничения в сочетаемости аффикса, если они существуют, должны подчиняться строгим законам).

Изучению и описанию процессов словообразования в русском языке, в частности суффиксального словообразования имен существительных методами традиционного языкознания, посвящено немало работ. Однако результаты этих работ имеют лишь относительную ценность для формального описания словообразовательных процессов [5—7]. Основная причина непригодности таких грамматик для «обучения» ЭВМ естественному языку состоит в том, что традиционные описания языка адресованы человеку, исходят из его фактического владения языком, исполь-

зуют его интуицию, способность самостоятельно и легко выявлять аналогии и т. п. Традиционные грамматические описания имеют ценность лишь при наличии математического аппарата, способного перевести эти описания на язык, доступный ЭВМ. Таким математическим аппаратом, на наш взгляд, является алгебра конечных предикатов, предложенная в [8, 9] и успешно используемая при описании процессов словоизменения [10, 11].

Как и при описании процессов словоизменения можно предположить, что процесс словообразования в русском языке вводит некоторое отношение, называемое нами в дальнейшем *деривационным*, связывающее между собой мотивирующее слово X , мотивируемое слово Y (продукт деривации) и набор грамматических и семантических признаков Z , позволяющих однозначно из X получить Y . На основе знания этого отношения могут решаться многие задачи обработки словесного материала. В статье с помощью средств алгебры конечных предикатов [8, 9] делается попытка обоснования возможности построения математической модели словообразования имен существительных русского языка. Рассматривается частный случай процесса словообразования — суффиксальное словообразование имен существительных, наиболее полно представленный в русском языке по количеству разрядов слов.

Пусть $L(X, Y, Z)$ — конечный предикат такой, что $L = 1$, когда X, Y, Z находятся в деривационном отношении, и $L = 0$ — в противном случае. Математическая модель деривационного отношения будет иметь вид формулы алгебры конечных предикатов, описывающей предикат L . Располагая такой моделью и решая уравнение $L(X, Y, Z) = 1$ при различном задании исходных данных, можно будет искусственно воспроизводить те или иные процессы словообразования в русском языке.

Предикат L математически описывается следующим образом. Предполагая, что для X, Y, Z деривационное отношение выполняется, т. е. что $L(X, Y, Z) = 1$, будем записывать в виде уравнений алгебры конечных предикатов $L_1 = 1, L_2 = 1, \dots, L_n = 1$ различные связи между X, Y, Z , независимые друг от друга, существующие в русском языке при этом предположении и описывающие разнообразные словообразовательные процессы. Здесь n — общее число связей, и если бы нам удалось выявить все связи такого рода, мы могли бы записать $L = L_1 L_2 \dots L_n$, это была бы запись полной модели словообразования имен существительных русского языка. Однако все связи выявить далеко не просто. Основная трудность, возникающая при построении такой математической модели, заключается в отыскании системы признаков, в том числе и семантических, позволяющих точно и однозначно переходить от мотивирующего слова к мотивируемому, используя необходимый при этом материал словообразования (в нашем случае — суффикс). Зачастую многие суффиксы принимают участие в самых разнообразных словообразователь-

ных процессах, проявляя огромное разнообразие смысловых оттенков, описать которые очень сложно.

Рассмотрим более подробно систему признаков, грамматических и семантических, необходимых для построения математической модели суффиксального словообразования имен существительных русского языка. В связи с этим можно предположить, что в основе процессов словообразования лежит некоторая функция $Z = F(X, Y)$ зависимости выходного слова Z от слова X , участвующего в его образовании и от текста Y , в который входит слово Z . Например, текст *они сели за маленький ...* и слово *стол* однозначно определяют слово *столик*, если будут указаны соответствующие семантические признаки: уменьшительность и т. п.

Математическое описание суффиксального словообразования сводится к формированию суффикса $S = s_1 s_2 \dots s_m$, где m — число букв в суффиксе, в зависимости от набора грамматических и семантических признаков. Символы s_1, s_2, \dots, s_m обозначают соответственно первую, вторую, ..., m -ю букву суффикса. В случае отсутствия буквы ее может заменить знак пробела —.

Влияние слова X на выбор суффикса S может быть однозначно охарактеризовано набором признаков $X = (x_1, x_2, x_3, x_4)$. Здесь x_1 — признак последней буквы основы мотивирующего слова со значениями *б, в, в', г, д, д', е, ж, з, з', и, к, л, м, м', н, н', о, р, р', с, с', т, т', ф, ч, ш, щ, ь* (для всех суффиксов имен существительных), где буква со штрихом означает мягкость согласной буквы, например, *здоровый*—*здоровяк*, *левый* — *левак* — в первом случае происходит смягчение буквы *в* перед суффиксом *-а (я)к*, во втором — такого смягчения не происходит. Поэтому в первом случае последней букве основы приписывается признак мягкости.

Далее, x_2 — признак ударности основы со значениями: *у* — *ударный* (например, *овражек*), *б* — *безударный* (*бережок*). Следующий признак x_3 включает в себя семантические составляющие, характеризующие основу мотивирующего слова. Влияние мотивирующего слова может проявляться через его семантику, например, при помощи суффикса *-ец* необходимо образовать имя существительное, обозначающее лицо или предмет, характеризующийся действием, названным в основе мотивирующего слова.

Признак x_4 указывает грамматические характеристики основы. Часто бывает необходимо указать такие грамматические категории основы, как род, число, падеж (если основа именная) или наклонение, лицо, вид спряжения (если основа глагольная). Например, появления гласной буквы *и, е, а* перед суффиксом *-лк(а)*, обозначающим предметы женского рода, обусловлено употреблением в качестве мотивирующей основы определенных глагольных форм прошедшего времени: *кипеть* — *кипел* — *кипелка*, *гладить* — *гладил* — *гладилка*, *считать* — *считал* — *считалка*.

Влияние текста Y на суффикс S может быть однозначно охарактеризовано набором признаков $Y = (y_1, y_2, y_3, y_4, y_5)$,

где y_1 — падеж со значениями: *и* — именительный, *р* — родительный, *д* — дательный, *в* — винительный, *т* — творительный, *п* — предложный; y_2 — род со значениями: *м* — мужской, *ж* — женский, *с* — средний; y_3 — число со значениями: *е* — единственное, *м* — множественное; y_4 — признак одушевленности со значениями: *н* — неодушевленный, *о* — одушевленный; y_5 — семантический признак слова со значениями: *д* — отношение к действию, *п* — отношение к признаку, *э* — эмоциональная окраска.

Каждое из этих трех основных значений семантического признака делится на свое множество более узких значений, например, значение «отношение к действию» разбивается на следующее семейство значений: а) производитель действия; б) орудие, предназначенное для выполнения действия; в) помещение, место, где выполняется действие; г) результат действия; д) значение отвлеченного действия.

В сочетании с грамматическими признаками (род, число, одушевленность) семантические признаки дают возможность однозначно охарактеризовать продукт словообразования — слово как результат словообразовательного процесса.

Следует отметить, что система семантических признаков имеет довольно сложную иерархическую структуру. Смысловые компоненты редко выступают в «чистом» виде, чаще они смешиваются, наслаиваются друг на друга еще на нижних уровнях этой структуры, переходят на высшие уровни, в конечном результате образуя необходимый смысл. Под смыслом суффикса будем понимать набор промежуточных переменных, связывающих уравнения, описывающие данный суффикс, с уравнениями, описывающими остальную часть текста (например, предложение). Чем больший объем текста мы сможем описать уравнениями алгебры конечных предикатов, тем полнее и точнее будет описан смысл каждого фрагмента текста, так как мы сможем учесть большее число влияний текста на данный фрагмент (в нашем случае — суффикс).

Список литературы: 1. Котов Р. Г. Лингвистика и современное состояние машинного перевода в стране. — Вопросы языкознания, 1976, № 5, с. 37—49. 2. Шанский Н. М. Очерки по русскому словообразованию. — М.: Изд-во Моск ун-та, 1968. — 239 с. 3. Чхаидзе М. П. Еще раз о мифах и правде машинного перевода. — Вопросы языкознания, 1976, № 5, с. 50—57. 4. Грамматика современного русского литературного языка. — М.: Наука, 1970. — 767 с. 5. Белопольская А. Р. Выделение формальных признаков как основы для построения алгоритмов (на материале немецкого языка). Автореф. дис. ...канд. филол. наук, Л., 1971. — 23 с. 6. Волоцкая З. М., Моложная Т. Н., Николаева Т. М. Опыт описания русского языка в его письменной форме. — М.: Наука, 1964. — 186 с. 7. Кулагина О. С. Об операторном описании алгоритмов перевода и автоматизации процессов их программирования. — Проблемы кибернетики, 1959, вып. 2, с. 289—302. 8. Шабанов-Кушнарченко Ю. П. О теории интеллекта. — Проблемы бионики, 1977, вып. 22, с. 15—22. 9. Шабанов-Кушнарченко Ю. П. О конечных предикатах. — Проблемы бионики, 1979, вып. 24, с. 16—21. 10. Шабанов-Кушнарченко Ю. П. Математическая модель склонения полных неприятельных имен прилагательных. — Науч. и техн. информация, 1978, с. 9—12.