В. А. ЛОВИЦКИЙ, канд. техн. наук

СТРУКТУРНЫЙ ПОДХОД К РЕШЕНИЮ МОРФОЛОГИЧЕСКИХ ЗАДАЧ

Построение искусственных систем, понимающих естественный язык, связано с реализацией процессов грамматической обработки отдельных слов языка. Понимание слов и их значений является первым из трех условий, выполнение которых обеспечивает понимание естественноязыкового высказывания [1]. Условно можно выделить два подхода к решению морфологических задач (МЗ): алгоритмический [2] и логический [3]. В первом случае МЗ решаются путем составления алгоритмов, с привлечением эвристической информации, во втором — описание процесса решения МЗ осуществляется на языке обобщенных логических функций. Основной недостаток этих подходов состоит в том, что описание процессов решения МЗ связано с конкретным естественным языком и отражает его особенности. Этот недостаток привел, напри-

мер, к необходимости разрабатывать различные системы дл морфологической классификации глагольных форм [4] и име прилагательных русского языка [5]. Более того, система дл морфологической классификации глагольных форм русского язык не может быть использована для решения таких же МЗ в англи ском языке без существенного изменения программного обесп чения системы, отражающего особенности естественного языка

В то же время мозг ребенка легко усваивает любой язы выделяя закономерности его построения на основании анали элементов «речевой» среды. Этот общеизвестный факт позволя утверждать, что мозг ребенка располагает определенными мех низмами и структурами, обеспечивающими усвоение особенност любого естественного языка. Поскольку закономерности язык выявляются в результате анализа относительно небольшого мн жества конкретных примеров, логично предположить, что чел веческий мозг использует индуктивные механизмы формирова понятий [6]. В самом деле, с помощью индуктивного метода лег определить, например, понятие «существительное среднего ро единственного числа именительного падежа», как слова, оканч вающиеся на «О», «Е» или «МЯ». Но с помощью данного поня можно решать только метазадачи, позволяющие определить пр надлежность любого входного слова заданному классу. А след вательно, решение задач преобразования слов (именно к это классу относится большинство МЗ) требует подхода, отлично от индуктивного формирования понятий функциональным пут Назовем его структурным методом формирования индуктивн понятий. Суть этого метода состоит в том, что в результа анализа небольшого числа слов, заданного конечным множеств X, строится не функция, включающая в себя существенные зн чения признаков всех слов множества X [6], а структура, котор содержит все слова множества X и отражает закономерност объединяющую слова $x_i \in X$ в одно множество.

Продемонстрируем работу алгоритма формирования индукти ных понятий структурным методом на конкретном пример Структуру будем строить в виде графа, вершины которого пре ставлены буквами слова, а дуги указывают на связь букв в слов Будем считать, что на множестве X задано разбиение его подмножества X_i, каждое из которых характеризуется некотор именем понятия $\mathrm{CNP}(<X_i>)$ [6]. Для рассматриваемого приме $\bigcup_i X_i = X$, а $\bigcap_i X_i = \varnothing$. Итак, пусть $X =$ {торчать, трещать, б жать, торчит, торчат, торчал, торчала, торчало, торчали}, прич $X_1 =$ {торчать, трещать, бежать}, $X_2 =$ {торчит}, $X_3 =$ {торчат $X_4 =$ {торчал}, $X_5 =$ {торчала}, $X_6 =$ {торчало}, $X_7 =$ {торчали а $\mathrm{CNP}(<X_1>) =$ {< что делать >}, $\mathrm{CNP}(<X_2>) =$ {< ч делает >, < муж. род >, < ед. число >}, $\mathrm{CNP}(<X_3>)$ = {< что делают >, < множ. число >}, $\mathrm{CNP}(<X_4>) =$ {< ч делал >, < муж. род >, < ед. число >}, $\mathrm{CNP}(<X_5>)$

= {< что делала >, < жен. род >, <ед. число>}, CNP (<X_6>)=
= {<что делало>, <ср. род>, <ед. число>}, CNP (<X_7>)=
= {<что делали>, <множ. число>}. Приведенные характеристики X_i ни в коей мере не претендуют на полноту и указаны только с целью демонстрации работы алгоритма. Идея работы алгоритма основана на построении древовидной структуры (ДС) [7]. Если ДС строится на основании элементов множества X, то будем ее называть прямой ДС (ПДС). Если же в качестве элементов множества взять обращенные слова, то получим X^R = = {ЬТАЧРОТ, ЬТАЩЕРТ, ЬТАЖЕЬ, ТИЧРОТ, ТАЧРОТ, ЛАЧРОТ, АЛАЧРОТ, ОЛАЧРОТ, ИЛАЧРОТ}, а ДС, построенную на основании этих слов, назовем обратной ДС (ОДС). Совместив эти две структуры соответствующим образом, получим ТВ-структуру (сокращение от английского There and Back: «туда и обратно»).

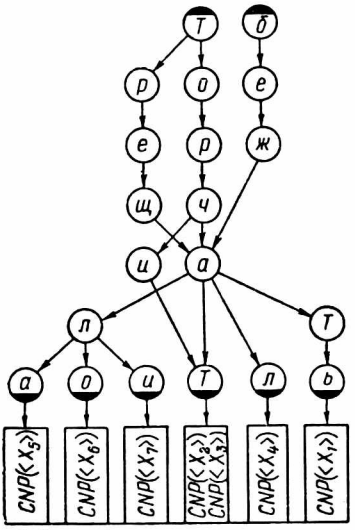
Алгоритм построения ТВ-структуры начинается с первого обращенного слова множества X_1 и продолжается до тех пор, пока не встретится «незнакомая» буква. Если этой «незнакомой» буквой является первая буква обращенного слова, то она заносится в ОДС, исключается из слова и алгоритм, используя остаток слова, переходит к построению ПДС. Так, для нашего примера построение ТВ-структуры началось со слова «торчать». Вначале в ОДС был занесен «Ь», затем в ПДС — «торчат» и наконец элемент ПДС с «Т» и элемент ОДС с «Ь» были соединены дугой. Для слова «трещать» фрагмент обращенного слова «ьта» читается в ОДС, фрагмент слова «трещ» заносится в ПДС, и элемент ПДС с «Щ» и элемент ОДС с «А» связываются между собой дугой. Таким же образом заносятся все слова множеств X_i. Иначе говоря, в ОДС заносятся только «незнакомые» последние буквы (у соответствующих им элементах (см. рисунок) заштрихована нижняя часть кружка), а в ПДС — фрагмент слова, который остался или после занесения последней буквы в ОДС, или после чтения в ОДС последовательности букв. Для рассматриваемого примера, пользуясь данным алгоритмов, получена ТВ-структура, представленная на рисунке. Вершины, соответствующие начальным буквам, заштрихованы вверху. При построении ТВ-структуры должно выполняться требование, согласно которому эти вершины не должны иметь входных связей, а вершины, отвечающие последним буквам слова — выходных связей. Названное достигается тем, что в ОДС читаются слова без начальной буквы. В процессе построения ТВ-структуры алгоритм ставит в соответствие каждому элементу указатели прямого и обратного хода, позволяющие читать слова в прямом и обратном направлениях. В ТВ-структуре, приведенной на рисунке, ориентация связей соответствует примерам рассматриваемых ниже задач.

После занесения каждого слова в ТВ-структуру, алгоритм связывает последнюю букву слова с областью, в которой хранятся

коды слов, оканчивающихся на эту букву, и их характеристик т. е. CNP ($< X_i >$).

Если построить систему, память которой представлена Т структурой, с ее помощью можно решать следующие задачи.

1. Необходимо охарактеризовать новое слово «бежала». Сло x_i называется новым, если $x_i \notin X$. Легко видеть, что в результа решения этой задачи получ CNP ($< X_5 >$).



Коды слов и их характеристики

Для глаголов, имен существи тельных и т. д. должны строить самостоятельные ТВ-структуры. Дано слово «бежала». Его необ димо так преобразовать, чтобы е соответствовала характеристика, з данная CNP ($< X_4 >$).

При решении задач данного ти система будет совершать «челове ские» ошибки. Так, преобразуя сл во «бежала» в соответствии CNP ($< X_3 >$), система выдаст сло «бежат».

Как видно из рисунка, сло оканчивающиеся буквой «Т», хара теризуются CNP ($< X_2 >$) и C ($< X_3 >$). В этом случае для ус новления взаимно-однозначного со ветствия между классами слов и характеристиками берется соответствующее число последних бу Так, для нашего примера CNP ($< X_2 >$) будет характеризов класс слов, оканчивающихся на «ИТ», а CNP ($< X_3 >$) — ок чивающихся на «АТ».

Характеристики слов связываются с начальными букв Они имеют более сложный характер и связаны с опре лением значений слов. В этом читатель сможет легко убедить построив самостоятельно ТВ-структуру для $X = \{$ХОДИ ПРИХОДИТЬ, УХОДИТЬ, ЗАХОДИТЬ, ПОДХОДИТЬ$\}$.

ТВ-структура, совместно с пирамидальной и «И/ИЛИ» стр турами [6] представляет собой память диалоговой системы, по мающей естественный язык (система ДЕСТА) и реализован в Харьковском институте радиоэлектроники.

Список литературы: 1. *Лурия А. Р.* Основные проблемы нейролингвист М., Изд-во Моск. ун-та, 1975. 253 с. 2. *Пиотровский Р. Г.* Текст, маш человек. Л., Наука, 1975. 326 с. 3. *Шабанов-Кушнаренко Ю. П.* Примене метода нуль-органа в лингвистике. — В кн.: Проблемы бионики. Харь 1978. Вып. 21, с. 109 — 112. 4. *Соловьева Е. А.* К вопросу о постро общего алгоритма морфологической классификации глагольных форм русс языка. — В кн.: Проблемы бионики. Харьков, 1975. Вып. 15, с. 143 — 5. Морфологическая классификация имен прилагательных русского язы *М. Ф. Бондаренко, Э. М. Бузницкая, Ю. В. Лопухин, Н. К. Свинарь*