

М. Ф. БОНДАРЕНКО, канд. техн. наук, Э. М. БУЗНИЦКАЯ,  
Ю. В. ЛОПУХИН, Н. К. СВИНАРЬ

## МОРФОЛОГИЧЕСКАЯ КЛАССИФИКАЦИЯ ИМЕН ПРИЛАГАТЕЛЬНЫХ РУССКОГО ЯЗЫКА

В настоящей работе рассматриваются алгоритмы морфологической классификации имен прилагательных русского языка с последующей их машинной реализацией.

Задача в общем виде может быть сформулирована следующим образом. Пусть имеется конечное множество  $M = \{a_1, a_2, \dots, a_n\}$ , где каждый элемент  $a_i$  есть имя прилагательное русского языка. Требуется построить разбиение этого множества на классы эквивалентности относительно различных грамматических категорий на основании заданного признака  $P_i \in P$ , где  $P = \{P_1, P_2, \dots, P_k\}$  — множество всех формальных или структурных признаков, которые заложены в самом рассматриваемом слое. Такими признаками являются наличие заданного символа или совокупности символов на определенной позиции слова, длина слова и т. д.

Грамматическая информация об имени прилагательном, входящем в определенное предложение, состоит из набора элементарных грамматических информационных (род, число, падеж и т. д.). Часто определить элементарную грамматическую информацию к словоформе удастся лишь учитывая соседние словоформы или целое предложение, в которое входит данная словоформа.

В случае морфологического анализа, т. е. при рассмотрении словоформы отдельно от контекста, в который она входит, каждому анализируемому слову должна быть поставлена в соответствие грамматическая информация, являющаяся дизъюнкцией элементарных грамматических информационных.

Основной проблемой рационального построения алгоритма морфологического анализа является такое расположение его частей, при котором соблюдена правильная последовательность выявления грамматических признаков. Это означает, что место конкретных правил в общей системе алгоритма определяется тем, что одни признаки могут быть выведены из совокупности других.

К сожалению, в отдельных точках алгоритмов не удается избежать использования так называемых словарей «исключений». Существование такого рода словарей объясняется нерегулярностями, отражающими характер русского языка. Использование таких словарей вполне правомерно, так как попытка унифицировать формальные признаки абсолютно всех имен прилагательных приводит к усложнению алгоритмов, а следовательно, к нерациональному увеличению затрат машинного времени на данный фрагмент алгоритма. В случае сравнения анализируемого слова со словом определенного словаря ответ о значении требуемого грамматического признака дается без дальнейшего анализа.

Морфологический анализ представляется целесообразным начать с разбиения всего исходного множества  $M$  на непересекающиеся подмножества  $M_1, M_2, M_3$ .

Класс  $M_1$  характеризуется наличием противопоставлений по роду, числу и падежу (это соответствует полным именам прилагательным и превосходной степени сравнения — аналитической).

Класс  $M_2$  — наличием противопоставлений по роду и числу (соответствует кратким формам имен прилагательных).

Класс  $M_3$  — наличием нулевой парадигмы (сюда войдут неизменяемые прилагательные — слова иностранного происхождения и компаратив).

Лингвистическими предпосылками организации формальных классов  $M_1, M_2, M_3$  могут служить следующие соображения. В парадигме формального класса должны быть учтены все противопоставления вариантов, свойственные членам этого класса. Поэтому парадигма может быть построена лишь на основе сопоставления парадигм отдельных лексем и выявления таким образом всех противопоставлений, встречающихся между элементами внутри данных частных парадигм.

У имен прилагательных русского языка легко выделяется парадигма по категориям рода, числа и падежа. Можно показать, что все другие категории, приписываемые обычно прилагательному, не входят в состав парадигм этого класса слов.

Решение задачи разбиения имен прилагательных на формальные классы позволяет производить дальнейший анализ имен прилагательных для каждого класса отдельно.

### Работа алгоритма

1. Проверить, входит ли анализируемое слово в словарь 1 (неизменяемые слова), если «да» — слово из  $M_3$ .
2. Сравнить последнюю букву анализируемого слова с а, я, у, ю, о, е, ы, и, ь, й в случае несовпадения выполнять п. 6.
3. Если последняя буква а, о, ы, отбросить ее, иначе — п. 8.
4. Сравнить две последние буквы слова с ын, ин, ов, ев, если ответ отрицательный — на выходе  $M_2$ .
5. Сравнить входное слово со словарем 2, в случае совпадения —  $M_2$ , иначе —  $M_1$ .
6. Сравнить две последние буквы слова с ом, ем, их, ым, в случае несравнения выполнять п. 4.
7. Если входное слово входит в словарь 5, то оно принадлежит классу  $M_2$ , в противном случае —  $M_1$ .
8. Если последняя буква анализируемого слова не я, и, то выполняется п. 10.
9. Сравнить две последние буквы слова с ья, ви, ая, яя, в случае, если произошло сравнение, то слово из  $M_1$ , в противном случае — из  $M_2$ .
10. Если последняя буква анализируемого слова не е, то перейти к п. 16.
11. Сравнить две последние буквы слова с ье, ые, ое, ие, в случае положительного ответа — слово из  $M_1$ .
12. Если слово оканчивается на ударное ее, то ответ  $M_3$ , иначе —  $M_1$ .
13. Сравнить две последние буквы слова с ще, в случае несовпадения выполнять п. 15.
14. Проверить вхождение анализируемого слова в словарь 3, если слово из словаря, то  $M_3$ , иначе —  $M_2$ .

15. Проверить принадлежность входного слова словарю 4, если слово из словаря, то  $M_2$ , в противном случае —  $M_3$ .

16. Если слово имеет ударное окончание ей, то оно принадлежит  $M_3$ , иначе —  $M_1$ .

Конец работы алгоритма.

### СЛОВАРИ ИСКЛЮЧЕНИЙ

#### Словарь 1

ампир	дезабилье
апаш	диез
аплике	ин-кварти
беж	ин-октаво
бемоль	ин-фолио
бордо	кlesh
брутто	комильфо
буфф	коми
валансьен	люкс
волан	майонез
галифе	максимум
гала	минимум
гротеск	маренго
гольф	массака
демисезон	мокко
декольте	мажор

минор
модерн
моль
неглиже
пикколо
плаке
плиссе
пик
перманент
реглан
рондо
сомон
sote
трико
томат
фантази

#### Словарь 2

нов
здоров
суров
готов

#### Словарь 3

гуше
проше
слаше
толще
чаще
чище

#### Словарь 4

свеже
несвеже
сине

#### Словарь 5

лих
нем
тих
хром

Определение грамматической категории числа у имен прилагательных русского языка осуществляется с помощью приведенного ниже алгоритма.

1. Если последняя буква анализируемого слова *ы, и*, ответ: «множественное число».

2. Сравнить две последние буквы входного слова с *ые, не, ых*, *их*, в случае отрицательного ответа перейти к п. 4.

3. Если результат сравнения входного слова с «тих», «лих» положителен, ответ — «единственное число», иначе — «множественное число».

4. Сравнить две последние буквы входного слова *ым, им*, в случае сравнения ответ: «единственное/множественное число» («единственное или множественное число»), в противном случае — «единственное число».

Решение данной задачи позволяет представить множество изменяемых прилагательных как  $M_{12} = M_{\text{мн}} \cup M_{\text{ед}}$ , куда входят прилагательные соответственно единственного и множественного числа. Это кажется целесообразным при определении грамматического признака рода, так как для  $M_{\text{мн}}$  происходит нейтрализация всех трех граммем рода, т. е. категория рода во множественном числе считается избыточной.

#### Алгоритм определения рода полных имен прилагательных

1. Если последняя буква анализируемого слова *а, я, у, ю, о, е, й*, перейти к п. 3.

2. Сравнить две последние буквы входного слова с *ым, им, ом, ем*, в случае совпадения ответ: «мужск./ср. род», в случае отрицательного ответа на выходе «мужск. род».

3. Сравнить последнюю букву слова с о, е, в случае несравнения выполнять п. 5.

4. Если предпоследняя буква слова г, на выходе ответ: «мужск./ср. род», в противном случае — «ср. род».

5. Если последняя буква слова я, ю, ответ: «женский род».

6. Если последняя буква слова а, то на выходе ответ: «мужск. род/женск. род/ср. род».

7. Совпадение с окончаниями ый, ий дает ответ: «мужской род».

8. Если две последние буквы анализируемого слова ой, на выходе ответ: «мужск. род/женск. род», иначе — «женский род».

Конец работы алгоритма.

*Алгоритм определения рода кратких имен прилагательных*

1. Если последняя буква анализируемого слова а, я, то на выходе «женский род».

2. Если анализируемое слово оканчивается на о, е, то на выходе «средний род», отрицательный ответ дает признак «мужской род».

Конец работы алгоритма.

Задача определения падежа имени прилагательного русского языка может быть представлена в следующем виде:

1. Если последняя буква анализируемого слова ы, я, на выходе ответ: «именительный падеж».

2. Если входное слово оканчивается на ю, то это «винительный падеж».

3. Сравнить последнюю букву слова с у, в случае отрицательного ответа перейти к п. 5.

4. Если вторая от конца буква слова м, то ответ: «дательный падеж», иначе — «дательный/винительный падеж».

5. Сравнить последнюю букву анализируемого слова с о, е, в случае несравнения перейти к п. 7.

6. Если вторая от конца буква слова г, то на выходе «родительный/винительный падеж», в противном случае — «именительный/винительный падеж».

7. Сравнить последнюю букву слова с и, если это не и, перейти к п. 9.

8. Проверить на совпадение вторую от конца букву слова с м, если это м, на выходе «творительный падеж», иначе — «именительный падеж».

9. Если анализируемое слово оканчивается на а, на выходе «именительный/родительный/винительный падеж».

10. Если слово оканчивается на ом, ем, ответ: «предложный падеж».

11. Если две последние буквы слова ым, им, на выходе «дательный/творительный падеж».

12. В случае совпадения двух последних букв анализируемого слова с ый, ий, ответ: «именительный/винительный падеж».

13. Если слово оканчивается на ей, на выходе «родительный, дательный/творительный/предложный падеж».

14. Совпадение двух последних букв слова с ых, их дает ответ: «родительный/винительный/предложный падеж».

15. Если слово оканчивается на ой, на выходе «именительный/одительный/дательный/предложный падеж», иначе — «имени-ельный/винительный падеж».

Конец работы алгоритма.

Приведенные алгоритмы были реализованы на ЭВМ «Урал-14Д» с использованием алгоритмического языка АЛГОЛ-ЦЭМИ. Выбор именно этого языка объясняется наличием ряда расширений (по сравнению с эталонным языком АЛГОЛ-60), позволяющих производить обработку текстовой информации [2].

Опишем процедуру выделения окончаний входных слов на языке АЛГОЛ-ЦЭМИ, используемую во всех программах:

```
procedure B (v, w, n);  
value n, text v, w; integer n;  
begin  
integer i;  
for i:=1 step 1 until 128 do  
if v from i= then  
begin  
w:=v from (i-n) thru i-1;  
goto exit  
end;  
end;  
exit: end.
```

В описании процедуры  $v$  — входное слово (длина его определяется);  $w$  — окончание слова длиной  $n$  литер.

Программы, написанные на языке АЛГОЛ-ЦЭМИ, обладают высокой надежностью поиска динамических ошибок, однако малая скорость обработки информации не позволяет использовать язык на ЭВМ «Урал-14Д» в качестве базового для создания системы обработки языковой информации.

#### СПИСОК ЛИТЕРАТУРЫ

1. Грамматика русского языка. Т. I. Под ред. В. В. Виноградова. М., изд. АН СССР, 1960. 720 с.
2. Транслятор АЛГОЛ-ЦЭМИ для ЭВМ «Урал-14». Инструкция. М., 1971. 24 с. Авт.: К. С. Кузьмин, М. Р. Левинтон, И. В. Максимова, А. В. Юнисова.

Поступила 10 ноября 1974 г.