

Человек способен решать задачи морфологической классификации, т. е. классифицировать части речи по некоторым признакам. Нас интересуют математические модели поведения идеально грамотного человека* (и. г. ч.) при решении таких задач классификации, например, при классификации глаголов по признаку лица. И. г. ч. предъявляется глагол (например, входящий в словарь русского языка [2]) в любой форме, представляющей собой одно слово. На основании только этой информации и. г. ч. должен принять решение, к какому лицу отнести предложенную глагольную форму (если последняя имеет признак лица) или ответить, что глагольная форма признака лица не имеет (если у данной формы глагола его нет). Таким образом, человек должен принять решение, какой из четырех признаков (1-е лицо, 2-е лицо, 3-е лицо, отсутствие лица) необходимо поставить в соответствие любому данному глаголу. Глагол предъявляется и. г. ч. без контекста. Поэтому когда предложенный глагол является омографом и без контекста его классификация может быть различной, верным ответом при выборе решения следует считать набор признаков лица, соответствующих данному глаголу.

Математическая модель способности и. г. ч. принимать решения при классификации глаголов по признаку лица в виде блок-схемы алгоритма приведена на рисунке. Множество входных сигналов $X = \{x_1, \dots, x_n\}$ составляют все глаголы русского языка, входящие в словарь [2], и те их формы, которые представляют собой одно слово. Множество выходных сигналов $Y = \{y_1, \dots, y_m\}$ — это множество X , классифицированное по признаку лица, т. е. выходной сигнал y_i — это входной сигнал x_i с добавлением признака лица: Π_1 — если глагол стоит в 1-м лице; Π_2 — если глагол во 2-м лице; Π_3 — если в 3-м и $\neg\Pi$ — если предложенная глагольная форма не имеет признака лица. Если же классификация глагола при отсутствии контекста окажется неоднозначной, то приписываемый глаголу признак объединит все признаки лица, соответствующие данному глаголу, например, $\Pi_1 \wedge \Pi_2$ или $\Pi_2 \wedge \neg\Pi$ и т. д.

Подавая на вход B алгоритма какое-либо слово, на его выходе B получаем то же слово с добавлением признака лица. После этого на вход алгоритма можно подавать новое слово.

Алгоритм составлен на основании только формальных признаков. Он использует анализ одной или нескольких последних букв слова, кроме случаев, когда это может привести к ошибкам. То-

* В данном случае имеется в виду человек, который решает предложенную задачу классификации точно в соответствии с правилами грамматики русского языка [1].

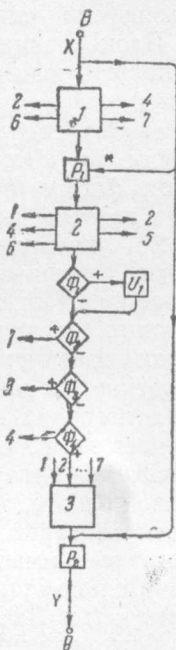
гда допускается минимально возможный словарь исключений. В нашем алгоритме обращение к словарям осуществляется в блоках 1 и 2.

Алгоритм включает в себя элементарные блоки и составные, которые, в свою очередь, состоят из элементарных. Рассмотрим работу элементарных блоков нашего алгоритма: распознавателей и операторов. Распознавателем Φ_j назовем блок, который проверяет, удовлетворяет ли слово, поданное на его вход, какому-либо условию, причем в случае выполнения данного условия слово выходит из блока по стрелке, отмеченной знаком плюс; в случае невыполнения — по стрелке со знаком минус. Блок U_i , который изменяет поступающий на его вход сигнал, т. е. совершает над этим сигналом определенную операцию, назовем оператором.

Распознаватели, входящие в состав алгоритма (рисунок), выполняют проверку следующих условий: Φ_1 проверяет две последние буквы слова на «ся» или «сь»; Φ_2 — последнюю букву слова на «у», «ю» или «м»; Φ_3 — ту же букву на «т», а Φ_4 — на «б», «й» или «е». Оператор U_1 отбрасывает две последние буквы слова.

Блок P_k выполняет одновременно функции оператора и блока памяти. Каждое слово, поданное на вход блока P_1 , который отмечен звездочкой, запоминается в нем (при этом стирается находящаяся в блоке P_1 информация). На выходе блока P_1 слово, поданное на отмеченный вход, появится только в том случае, если следующий сигнал будет подан на неотмеченный вход. Блок P_2 запоминает поданный на его вход сигнал, а затем приписывает к нему поданный через некоторое время признак; полученный сигнал y_i поступает на выход алгоритма, после чего на вход алгоритма можно подавать новое слово.

Блок 3 является составным оператором формирования признаков лица. Он представляет собой параллельное соединение элементарных операторов, каждый из которых может заменять входное слово на какой-либо признак лица. Каждому элементарному оператору соответствует один пронумерованный вход блока 3, а выходы элементарных операторов соединены, образуя выход блока 3. При подаче слова на вход 1 слово заменяется признаком Π_1 ; при подаче на вход 2 — признаком Π_2 ; 3 — Π_3 ; 4 — $\neg \Pi$; 5 — $\Pi_1 \wedge \Pi_2$; 6 — $\Pi_2 \wedge \neg \Pi$; 7 — $\Pi_3 \wedge \neg \Pi$. Цифра на выходе любого блока означает, что этот выход соединяется с таким же образом пронумерованным входом блока 3.



Блоки 1 и 2 классифицируют небольшую часть глагольных форм, которые, обладая одинаковыми формальными признаками, являются омографами или же представляют собой исключения из общих правил. В состав блоков 1 и 2 входят распознаватели и операторы, принцип действия которых описан выше, а также распознаватели типов $\Phi_{И_e}$ и $\Phi_{С_m}$. Распознаватель типа $\Phi_{И_e}$ проверяет, совпадает ли конец входного слова с каким-либо одним из слов словаря $И_e$, а $\Phi_{С_m}$ — совпадает ли входное слово с одним из слов словаря $С_m$.

Блок 1 классифицирует (в данном случае — направляет на соответствующий пронумерованный выход) глаголы, оканчивающиеся на «ть», «чь», «ти» («тсья», «чься», «тись»). Этот блок по своему содержанию аналогичен алгоритму, предложенному в работе [3], который выполняет классификацию глаголов по признаку формы (неопределенной или личной). Блок 2 классифицирует глаголы, которые в 1-м лице изъявительного наклонения и во 2-м лице повелительного являются омоформами; глаголы, которые оканчиваются на «ли» («лись»), «-ка», а также глаголы типа «ляг». При решении данной задачи входным алфавитом является русский, расширенный дефисом (дефис может встречаться в глаголах повелительного наклонения, оканчивающихся на «-ка»). Введение во входной алфавит знака ударения в единичных случаях позволит избежать двойного ответа. Поскольку число таких случаев мало, а расширение входного алфавита знаком ударения усложняет алгоритм, то подобное расширение не проводилось в данной работе. Решая поставленную задачу, мы стремились свести объем словарей исключений к минимально возможному. В результате суммарный объем словаря в процессе отладки алгоритма значительно уменьшился. Количество входных слов алгоритма более чем в 1000 раз превышает число всех слов-исключений.

Метод решения поставленной задачи [3] предполагает получение точного решения для любого слова, входящего в выбранное множество входных слов. Это подтвердил анализ составленной модели, а также многочисленные эксперименты.

Предложенная математическая модель принятия решений в процессе классификации глаголов по признаку лица может оказаться полезной при составлении других подобных моделей, а также представить самостоятельный интерес.

ЛИТЕРАТУРА

1. Грамматика современного русского литературного языка. М., «Наука» 1970. 767 с.
2. Орфографический словарь русского языка, изд. 2-е. М. «Сов. энциклопедия», 1971. 520 с.
3. Бондаренко М. Ф., Соловьева Е. А. Методы решения задач морфологической и субморфологической классификации. Сб. «Проблемы бионики», вып. 10. Изд-во Харьковск. ун-та, 1973, с. 145—150.