УДК 62.506.2

А. И. ЧУГУН

НЕКОТОРЫЕ ВОПРОСЫ ФОРМАЛЬНОГО ОПИСАНИЯ РУССКОГО ГЛАГОЛА

В связи с невероятно быстрым ростом объема знаний, накопленных человечеством, особенно острой стала проблема создания информационных кибернетических устройств, хранящих и разыскивающих по заказу научную информацию. Решить эту проблему возможно путем широкого использования современной цифровой техники для логической, статистической, математической и другой обработки материалов. А для этого прежде всего надо «научить» машину воспринимать информацию, записанную на естественном языке. Процесс обучения ЭВМ пониманию естественного языка возможен только на основе достаточно содержательной формализации и возможности перевода формального описания этого языка на язык математики, т. е. создания его математической модели.

В настоящей работе рассматривается построение одной математической модели языкового поведения человека: моделирование процесса морфологической классификации глаголов русского языка. Глагол образует сложную систему форм, которая на основании выполняемых ими функций и свойственных им грамматических категорий подразделяется на 4 разряда: неопределенная форма, личные формы, причастия и деепричастия. Поэтому входное множество M рассматриваемой математической модели можно представить как объединение подмножеств M_i $(i=1, 2, 3, 4)$, представляющих собой названные 4 разряда глагольных форм:

$$M = \bigcup_{i=1}^{n} M_i.$$

Наиболее сложным для формализации в системе глагольных форм является подмножество M_2, включающее в себя личные спрягаемые формы глагола. Его мы и будем рассматривать в качестве исходного для предлагаемой математической модели. Так как математическая модель представляет собой некоторое множество с заданной на нем совокупностью отношений $A = \langle M, \{R_1, R_2, \ldots, R_n\} \rangle$, где $R_1, R_2, \ldots, R_n$ — отношения на множестве M, то нам необходимо определить эти отношения для исходного множества M_2. Традиционная грамматика ставит в соответствие каждому элементу данного множества следующие грамматические категории: залог, вид, наклонение, время, лицо, число и род [1]. Из перечисленных грамматических категорий целесообразно рассматривать только те из них, которые будут влиять на формообразование элементов множества M_2.

Вводим обозначение грамматических категорий, присущих элементам множества M. Категорию наклонения обозначим как $н(x) = н_1(x) \vee н_2(x)$, где $н_1(x)$, $н_2(x)$ — соответственно категории изъявительного и повелительного наклонений. Категорию времени — $b(x) = b_1(x) \vee b_2(x)$, где $b_1(x)$, $b_2(x)$ — соответственно категории непрошедшего и прошедшего времени. Категорию лица — $л(x) = л_1(x) \vee л_2(x) \vee л_3(x)$, где $л_1(x)$, $л_2(x)$, $л_3(x)$ — соответственно категории первого, второго и третьего лица. Категорию числа — $r(x) = r_1(x) \vee r_2(x)$, где $r_1(x)$, $r_2(x)$ — соответственно категории единственного и множественного чисел. Категорию рода — $p(x) = p_1(x) \vee p_2(x) \vee p_3(x)$, где $p_1(x)$, $p_2(x)$, $p_3(x)$ — соответственно категории мужского, женского и среднего рода.

В процессе развития языка количество личных форм глагола постоянно изменяется в сторону увеличения, а это значит, что рассматриваемое множество M_2 будет представлять собой расширяющуюся динамическую модель. Поэтому множество M_2 можно представить как объединение S и S': $M_2 = S \cup S'$. Здесь S и S' — соответственно множество всех личных форм глагола и множество псевдоглаголов $x \in M_2$. Ставя в соответствие каждому элементу множества S' присущие ему признаки (грамматические категории), можно разбить это множество на элементарные классы, логические формулы которых будут иметь вид: $H_i(x) \wedge b_j(x) \wedge \wedge л_k(x) \wedge r_l(x) \wedge p_m(x)$ $(i = 1, 2;\ j = 1, 2;\ k = 1, 2, 3;\ l = 1, 2;\ m = = 1, 2, 3)$, откуда предполагаемое число элементарных классов $n = 2 \times 2 \times 3 \times 2 \times 3 = 72$ кл.

Неформальные грамматические правила позволяют сделать вывод, что не все признаки, присущие элементам рассматриваемого подмножества S', совместимы. Так, например, категорию рода можно поставить в соответствие только категории прошедшего времени единственного числа, а категория повелительного наклонения сочетается только со вторым лицом непрошедшего времени. Поэтому некоторые классы будут пустыми и действительное их число будет значительно меньше предполагаемого.

---

[1] Шведова Н. Ю. Грамматика современного русского литературного языка. М., «Наука», 1970. 75 с.

В результате исследований были получены следующие непустые элементарные классы.

1. $Ф_1 = \{x: н_1(x) \wedge b_1(x) \wedge r_1(x) \wedge л_1(x)\}$.
2. $Ф_2 = \{x: н_1(x) \wedge b_1(x) \wedge r_1(x) \wedge л_2(x)\}$.
3. $Ф_3 = \{x: н_1(x) \wedge b_1(x) \wedge r_1(x) \wedge л_3(x)\}$.
4. $Ф_4 = \{x: н_1(x) \wedge b_1(x) \wedge r_2(x) \wedge л_1(x)\}$.
5. $Ф_5 = \{x: н_1(x) \wedge b_1(x) \wedge r_2(x) \wedge л_2(x)\}$.
6. $Ф_6 = \{x: н_1(x) \wedge b_1(x) \wedge r_2(x) \wedge л_3(x)\}$.
7. $Ф_7 = \{x: н_1(x) \wedge b_2(x) \wedge r_1(x) \wedge (л_1(x) \vee л_2(x) \vee л_3(x)) \wedge p_1(x)\}$.
8. $Ф_8 = \{x: н_1(x) \wedge b_2(x) \wedge r_1(x) \wedge (л_1(x) \vee л_2(x) \vee л_3(x)) \wedge p_2(x)\}$.
9. $Ф_9 = \{x: н_1(x) \wedge b_2(x) \wedge r_1(x) \wedge (л_1(x) \vee л_2(x) \vee л_3(x)) \wedge p_3(x)\}$.
10. $Ф_{10} = \{x: н_1(x) \wedge b_2(x) \wedge r_2(x) \wedge (л_1(x) \vee л_2(x) \vee л_3(x))\}$.
11. $Ф_{11} = \{x: н_2(x) \wedge b_1(x) \wedge r_1(x) \wedge л_2(x)\}$.
12. $Ф_{12} = \{x: н_2(x) \wedge b_1(x) \wedge r_2(x) \wedge л_2(x)\}$.

Объединение этих классов определит нам все исходное множество M_2. Пустые элементарные классы интереса для исследования не представляют и в данной работе не рассматривались.