УДК 62.506.2

Ю. П. ШАБАНОВ-КУШНАРЕНКО, д-р техн. наук,
Е. А. СОЛОВЬЕВА

БИОНИЧЕСКИЕ АСПЕКТЫ МОДЕЛИРОВАНИЯ РЕЧЕВОГО ПОВЕДЕНИЯ ЧЕЛОВЕКА

Математические модели различных способностей человека ши-
ю применяются в технике, которая, развиваясь, в свою очередь
вит перед бионикой новые задачи. В связи с созданием вычисли-
ьных машин, способных частично заменять человека в сфере
еллектуальной деятельности, появился ряд требующих разре-
ния проблем, в частности проблемы общения человека с машиной
ашин между собой.

Несмотря на совершенствование входных и выходных устро ЭЦВМ, а также методов автоматизации программирования, про общения до сих пор недостаточно эффективен. Человеку удоб выражать свои мысли в языковой форме, поэтому его непос ственное общение с машиной на естественном человеческом яз или на языке, близком к человеческому, было бы наиболее це сообразным. Таким образом, возникает необходимость созда транслятора с естественного языка на машинный, т. е. дейст щей модели человеческого языка. В сущности, не имеет значе происходит ли общение человека с машиной в виде устной или п менной речи. Важно другое — умение машины владеть язы понимать его смысловую и грамматическую структуру. В дан работе не затрагивается вопрос о том, насколько такой язык об ния будет отличаться от человеческого (видимо, он станет п ближаться к естественному языку по мере совершенствован

Известно, что современные искусственные языки плохо п способлены для описания творческих функций человека, поэт при использовании машин для автоматизации интеллектуал деятельности также следует обучать их естественному язы Способность машины перерабатывать словесную информацию о бенно необходима при решении таких важных прикладных за как машинный перевод, автоматическое реферирование и анно рование текстов, обработка данных в системах материально-тех ческого снабжения и др. Модель естественного языка необход в деле разработки и внедрения автоматизированных систем уп ления.

Итак, практика выдвинула задачу создания действующей м дели человеческого языка, которая по мере совершенствова будет все более ему соответствовать. Пока полная модель не созда большое значение для теории и практики представляют част математические модели, описывающие отдельные компоненты ре вого общения.

Не вызывает сомнения, что «язык есть не нечто, посторон по отношению к человеку, что можно изучать лишь как некото «памятник» эпохи, направления или художественного творче отдельных людей, а часть самого человека в такой мере, в какой частью человека является его способность ход на двух ногах в вертикальном положении, создавать ору труда, мыслить понятиями и пр.» [1, с. 19]. Поэтому для модели вания языка (в частности, русского) требуется прежде всего и чение его как «части самого человека», что является одной из п воочередных задач бионики.

Речевое поведение можно изучать с помощью методов биоф зики, нейрохирургии и других наук, исследуя процессы, проис дящие в мозгу человека при речевом общении. Подобные иссле вания важны, однако проще изучать и моделировать язык с п мощью макроподхода — кибернетического метода «черного ящик Информационно-функциональный характер кибернетическ

моделирования позволяет без анализа структуры человеческого мозга достичь хороших результатов при исследовании способности человека перерабатывать словесную информацию.

В данной работе на материале русского языка рассматриваются результаты моделирования словесного поведения человека, которое является неотъемлемой частью его речевого поведения. Под словесным поведением понимаем не обычное речевое общение, а переработку человеком (как преобразователем информации, которому применимы понятия входных и выходных сигналов) отдельных словоформ. Одним из этапов моделирования словесного поведения является создание алгоритмов автоматического морфологического анализа, которые позволяют получать морфологическую информацию о каждой словоформе.

Психологические эксперименты показывают, что результаты решения каким-либо человеком даже простейших морфологических или субморфологических (связанных с частями слов, например [2]) задач могут несколько отличаться от результатов других людей, что объясняется прежде всего различием словарного запаса и неодинаковым знанием грамматики испытуемыми. В принципе можно создать столько моделей какой-нибудь способности человека обрабатывать словесную информацию, а также моделей речи человека, сколько существует их прототипов — людей, говорящих на русском языке. Математическое описание речи отдельного индивидуума представляет известный интерес, однако особую практическую ценность имеет изучение и моделирование поведения идеально грамотного человека (точно руководствующегося правилами грамматики [3]), т. е. создание моделей языка.

Во многих работах, посвященных моделированию языка, в явном или неявном виде применяются морфологические анализ и синтез, которые обычно осуществляются при помощи словаря основ или словоформ, а также вручную (частично или полностью) перед введением информации в машину. Множество входных слов невелико и часто ограничивается какой-либо узкой областью, например математическими текстами. Все это объясняется не только трудностями формализации языка, но и спецификой задачи машинного перевода, на решение которой в основном ориентируются исследователи. Общепринятый путь анализа, при котором в начале осуществляются словарный поиск и извлечение информации основы, иногда стараются объединить с подходом к языку как к цепи морфем [4], использующим максимум информации флексий (и некоторых других элементов) до обращения к словарю основ.

Предлагаемые нами алгоритмы отличаются прежде всего тем, что множество входных слов достаточно велико (за основу берется словарь русского языка [5] объемом в 104 000 слов с учетом всевозможных синтетических форм этих слов), а используемые словари ничтожно малы по сравнению с входным множеством. При получении искомых грамматических признаков исследуется минимально

необходимое для этой цели количество букв слова (обычно пос них), независимо от его членения на морфемы.

В единичных случаях такой информации о словоформе оказ ется недостаточно для точного решения. Если же оно необход то для анализа слов, неточно обрабатываемых алгорит можно использовать предлагаемые ниже методы (отдельно в комбинации друг с другом).

1. *Метод словарей.* В качестве формального признака исп зуется все слово. При этом появится словарь исключений, об которого должен быть намного меньше объема входного жества.

2. *Метод обращения.* В алгоритме (программе) должно б предусмотрено обращение к человеку или к алгоритму (если т вой имеется) в целях получения признака, необходимого точного решения.

3. *Метод дополнения.* К входному слову приписываются о или несколько дополнительных признаков, позволяющих ре задачу точно.

4. *Метод ограничения.* Входное множество слов огранич ется таким образом, чтобы модель функционировала безошибо

Хотя метод 1 является самым трудоемким, его использова наиболее целесообразно, если объем словаря исключений неве В противном случае наиболее приемлемым окажется метод 2. М 3, снижающий автоматизированность модели, и метод 4 желате не применять, так как они несколько уменьшают ценность а ритмов.

В предлагаемых здесь моделях получение высокой вероятно их достоверных предсказаний основано прежде всего на прим нии метода 1.

При построении моделей используем принцип, заключающи в том, что словоформа, изолированная от контекста, обла всеми грамматическими значениями некоторой грамматичес категории, которыми эта словоформа может характеризовать в различных контекстах. Это означает, что допускаются неод значные ответы машины, которые считаются точными, если соот ствуют предложенному выше принципу. Например, для глаго ной формы *вели* ответ «изъявительное наклонение и повелитель наклонение» будет верным, так как *вели* может использовать в контекстах в грамматических значениях как изъявительно так и повелительного наклонения. Подавляющее большинс словоформ характеризуется одним грамматическим значени любой грамматической категории за исключением слов-омон фов, для которых выходной сигнал будет состоять из конъюнк признаков. Предложенный принцип подтверждается многочисл ными психологическими экспериментами.

Плодотворной является идея об использовании взаимосв и взаимозависимости между различными грамматическими ка гориями и значениями. Исследование таких зависимостей полез

например, при составлении алгоритмов анализа многих грамматических категорий для минимизации таких алгоритмов.

В целях использования глубинных структур языка мы неограниченно расширяем входное множество всевозможными словами, псевдословами и их формами, образованными на основании правил грамматики русского языка. При этом подразумевается, что существует распознающая процедура, способная отличать слова и псевдослова от произвольных цепочек букв и позволяющая таким образом формировать входное множество. Алгоритмы, оперирующие словами из неограниченного множества, не содержат исключений, довольно просты и дают достоверные предсказания с высокой вероятностью. Для получения точного решения при неограниченном множестве входных слов можно использовать методы 2—4.

Полученные алгоритмы позволяют сделать заключение о больших возможностях формального описания морфологии русского языка.

Математические модели способности идеального грамотного человека решать задачи морфологической классификации получены в виде алгоритмов, реализованных на ЭЦВМ. Решена задача морфологической классификации глаголов русского языка на разряды (инфинитив, личные формы, причастия, деепричастия), т. е. анализа категории репрезентации. Получены модели определения наклонения, времени, лица, числа, рода, спряжения, анализа суженной парадигмы, повелительного наклонения (собственно глаголов), проанализированы атрибутивные формы (причастия и деепричастия). Составлены алгоритмы определения глагольных признаков (времени и залога) причастий, а также разделения причастий на полные и краткие. Рассматриваются задачи об определении признаков вида и залога глагольных форм, задачи синтеза глаголов и субморфологической классификации.

Приведем записанные на алгоритмическом языке АЛГОЛ (для транслятора АЛГОЛ—ЦЭМИ на ЭЦВМ «Урал-14») отлаженные программы алгоритмов анализа категорий времени и числа глаголов русского языка в личной форме. Входным сигналом этих алгоритмов может быть произвольный глагол или псевдоглагол русского языка; входной алфавит — русский, расширенный знаком «—», который может применяться в глаголах повелительного наклонения (на *-ка*) и для выражения редубликации (или удвоения слов). Выходными сигналами для первого алгоритма служат признаки времени: непрошедшее и прошедшее, а также сигнал отсутствия времени (для глаголов повелительного наклонения). Множество выходных сигналов второго алгоритма объединяет два грамматических значения категории числа — единственное и множественное.

Количество входных слов N для каждого просчета на машине может быть любым, для изменения этого количества в программе достаточно заменить одну перфокарту. В данном случае $N = 100$ (для определения вероятностей достоверного предсказания

моделей выбрано N-1000). Максимальная длина слова в предлага мых алгоритмах равна 80 символам.

Эталонный язык в трансляторе АЛГОЛ — ЦЭМИ расшир введением текстовых величин и действий над ними. Значения текстовых величин являются последовательности литер (пуст строка, буква, цифра, знак операции и т. п.). Мы пользовали операциями соединения текстовых значений (ТЗ) и выделения час значения текстового выражения, осуществляемого с помощ выделителя (**from** < индексное выражение > **thru** < индекс выражение >), следующего за текстовым выражением, котор заключено в круглые скобки. Операция соединения ТЗ предста ляется знаком | (вертикальная черта); результат ее состоит из л тер первого ТЗ, за которыми следуют литеры второго ТЗ. Дейст выделителя состоит в выделении подпоследовательности литер значения предшествующего ему текстового выражения, начин с литеры, номер которой равен значению первого индексного вы жения, и кончая литерой, номер которой равен значению втор индексного выражения включительно.

Программа определения времени глагола имеет вид

```
Begin integer N;
    N: = 100;
    begin text array А[1:N], Б[1:80];
        integer i, j, L;
        text C;
        for i: = 1 step 1 until N do
            А[i]: = intext (ПК, 80);
        for i: = 1 step 1 until N do
        begin for j: = 1 step 1 until 80 do
            begin Б[j]: = (А[i]) from j thru j;
                if Б[j] = ( ) then
                begin L: = j — 1; go to М1; end;
            end;
М1:         if (Б[L — 1] | Б[L] = 'ся') V (Б[L — 1] | Б[L] = 'сь')
                then L: = L — 2;
            go to if (Б[L] = 'у') V (Б[L] = 'ю') V
                V (Б[L] = 'м') V (Б[L] = 'т')
                then НВ else if (Б[L] = 'и') V
                V (Б[L — 2] | Б[L — 1] | Б[L] = '-ка')
                then Н;
            go to if (Б[L] = 'и') V (Б[L] = 'о') then
                (if Б[L — 1] = 'л') then ПВ else Н);
            go to if Б[L] = 'ь' then
                (if Б[L — 1] = 'ш' then НВ else Н);
            go to if Б[L] = 'е' then
                (if (Б[L — 2] | Б[L — 1] = 'ет') V
                V (Б[L — 2] | Б[L — 1] = 'ит')
                then НВ else Н) else ПВ;
```

```
НВ:         outtext (АЦПУ, А[i] | 'непрошедшее время');
                go to М2;
ЛВ:         outtext (АЦПУ, А[i] | 'прошедшее время');
                go to М2;
Н:          outtext (АЦПУ, А[i] | 'нет времени');
М2:         newline (АЦПУ);
        end;
    end;
end.
```

Программа определения категории числа глаголов получена в виде

```
Begin integer N;
    N:= 100;
    begin text array А[1:N], Б[1:80];
        integer i, j, L;
        text C;
        for i:= 1 step 1 until N do
            А[i]:= intext (ПК, 80);
        for i:= 1 step 1 until N do
        begin for j:= 1 step 1 until 80 do
            begin Б[j]:= (А[i]) from j thru j;
                if Б[j] = ' ' then
                begin L:= j — 1; go to М1; end;
М1:         if Б[L — 2] | Б[L — 1] | Б[L] = '-ка' then L:= L — 3;
            if (Б[L — 1] | Б[L] = 'ся') ∨
                ∨ (Б[L — 1] | Б[L] = 'сь') then L:= L — 2;
            C:= Б[L — 1] | Б[L];
            if (C = 'ат') ∨ (C = 'ят') ∨ (C = 'ут') ∨ (C = 'ют') ∨
                ∨ (C = 'ем') ∨ (C = 'им') ∨ (C = 'те') then go to М;
            if Б[L] = 'и' then L:= L — 1 else go to Е;
            if Б[L] = 'л' then go to М;
Е:          outtext (АЦПУ, А[i] | 'единственное число');
                go to М2;
М:          outtext (АЦПУ, А[i] | 'множественное число');
М2:         newline (АЦПУ);
        end;
    end;
end.
```

Структура программ определения лица, рода и некоторых других грамматических категорий аналогична.

Ставя своей целью автоматическое получение грамматической информации о слове, в ряде случаев можно автоматически получать частичную смысловую информацию. Это обусловлено тем, что, определяя несинтаксические грамматические значения слов (например, значения времени), «отражающие различные смысловые абстракции» [3, с. 303], мы тем самым частично формализуем смысл. Решение задач морфологического анализа и синтеза позволит

определить для каждого слова его парадигму, т. е. класс сло с одним и тем же лексическим значением.

Полученные алгоритмы используются при дальнейших иссле дованиях (анализ пар слов, фраз и т. д.), но представляют и само стоятельный интерес как математические описания психически функций человека, а также могут найти применение, например в различных системах автоматического анализа текстов.

Процессы переработки человеком словесной информации тре буется изучать не только при решении морфологических задач Грамматика русского языка (представляющая собой обобщенны опыт владения языком многими поколениями людей и являюща яся эффективным средством при обучении человека языку, закреп лении языковых норм и т. д.) предназначена для человека, и ее пра вила в том виде, в каком они записаны, обычно не пригодны дл обучения машины. Бывает, что правила грамматики оказываютс недостаточными для формализации языка. В таких случаях при ходится проводить психологические эксперименты и изучать язы ковые структуры для выявления закономерностей, еще не опи санных в языке. Исследование речевого поведения человека по зволяет глубже проникать в структуру естественного языка создавать «машинные» правила и модели.

ЛИТЕРАТУРА

1. Звегинцев В. А. Теоретическая и прикладная лингвистика. М. «Просвещение», 1968. 335 с.
2. Соловьева Е. А. Математическое описание способности человек анализировать правильность переноса слов.— В сб.: Проблемы бионики Вып. 8. Харьков, 1972, с. 61—67.
3. Грамматика современного русского литературного языка. Отв. ред Н. Ю. Шведова, М., «Наука», 1970. 767 с.
4. Севбо И. П. Пивоварова Е. П. Об алгоритме независимог флективного анализа русского текста.— В кн.: Прикладная лингвистик и машинный перевод. [Сб. статей]. Киев, 1962, с. 66—71.
5. Орфографический словарь русского языка. Изд. 11-е. Под ред. С. Г. Бар хударова и др. М., «Сов. энциклопедия», 1971. 520 с.