

*О.И. Лучинкина, О.Н. Карпов, доктор технических наук
(Днепропетровский государственный университет имени Олеся Гончара)*

МОДЕЛИРОВАНИЕ ЛИНГВИСТИЧЕСКИХ УРОВНЕЙ СИСТЕМЫ РАСПОЗНАВАНИЯ СЛИТНОЙ РЕЧИ

В тезисах предложена модель построения системы распознавания слитной речи, как системы искусственного интеллекта. Приведены принципы построения фреймов, описывающих отдельные этапы системы.

На сегодняшний день существует множество систем распознавания речи. Большинство из них направлены на распознавание изолированных слов. Согласно опубликованным на сегодняшний день данным, надежность распознавания речи для систем, которые работают с однословными командами, достигает 99.5, командами, которые состоят из двух слов, - 97.5, из трех слов – 92.5, из четырех слов – 91.9. Но при проектировании системы распознавания слитной речи перед разработчиками появляется проблема распознавания не только языкового сигнала, но и лингвистического содержания.

Системы распознавания речи можно классифицировать по таким признакам [2,3]:

- тип речи (слитная или изолированная);
- тип элементов словаря (фонемы, слоги, слова, и др.);
- зависимость от диктора;
- степень детализации;
- размер словаря.

Каждая такая система характеризуется своими методами и алгоритмами.

Как правило, работа системы распознавания речи делится на два вида:

- распознавание голосовых меток;
- распознавание лексических элементов

Первый подход допускает распознавание фрагментов языка по заранее записанному образцу. Этот подход широко используется в относительно простых системах, предназначенных для выполнения заранее записанных команд, например, системы голосового управления транспортом.

Второй подход сложнее. При его реализации из потока речи выделяются отдельные лексические элементы - фонемы и алофоны, которые потом объединяются в составы и морфемы. Строго говоря, именно этот подход и используется в "настоящих" системах распознавания речи. При создании системы распознавания слитной речи важно не просто разработать механизм, позволяющий математически описать звуковой сигнал. Разработка подобной системы включает лингвистический аппарат, позволяющий анализировать распознанные данные и формировать выводы на основе языковых правил.

Рассмотрим один из возможных подходов к разработке системы распознавания слитной речи, учитывающей особенности языка и произношения.

На начальных уровнях анализа перед нами стоит задача разбиения высказывания на элементы первичного анализа. В качестве элемента первичного анализа будем брать фонемы. Как фонетического алфавита будем использовать набор из 48 фонем: 12 - для гласных звуков (учитывая, что каждая гласная может быть ударной и безударной) и 36 - для согласных (учитывая, что каждая согласная может быть твердой и мягкой).

Таким образом, получаем алфавит фонем:

- Гласные: а а! е е! о! и и! у у! ы ы! э!.
- Согласные: б, б' в в' г г' д д' ж з з' й к к' л л' м м' н н' п п' р р' с с' т т' ф ф' х х' ц ч ш щ.

Одним из важнейших узлов системы является транскриптор.

Фонетическая транскрипция [1,4] - это особый вид записи речи, который используется для фиксации на листе особенностей ее звучания. Она предназначена для описания произношения слов. Текст на естественном языке представляет собой упорядоченный поток символов. Символы обрабатываются последовательно, друг за другом, в порядке их расположения в тексте. Обратный транскриптор – это механизм, позволяющий преобразовывать произносимое слово в его запись.

При создании фонетического транскриптора необходимо принимать во внимание логику транскрипции. Проблема заключается в том, что невозможно поставить в соответствие каждой букве ее конкретный звук. Так, например, транскрипция слова "еж" - [jош], а слова "кожа" - [кожа].

Звук буквы меняет свое значение в зависимости от расположения в слове, ударению, порядку следования, свойств букв. Таким образом, транскриптор строится на основе правил, определенных конкретным языком.

Так как перед нами стоит задача анализа слитной речи, а не просто анализа отдельных слов, возникает очередная проблема. Зачастую, при быстром произнесении фразы, некоторые фонемы на границах слов «теряются». Таким образом, кроме правил транскрипции изолированных слов необходимо учитывать и правила, описывающие транскрибирование на границах слов.

Писать программный продукт, который будет проводить десятки однотипных проверок для каждой буквы при транскрибировании, не является целесообразным. Поэтому проанализировав современные методы решения подобных проблем можно прийти к заключению, что целесообразно строить такую систему как систему искусственного интеллекта, анализирующую набор правил на основе некоторых заранее заданных методов.

Цель проектирования экспертных систем [5] заключается в разработке программных продуктов, которые могут решать задачи, которые являются тяжелыми для решения экспертами в связи с необходимостью обработки большого количества данных, или применения большого количества правил. С помощью экспертных систем мы получаем результаты, которые не уступают по качеству и эффективности решениям, до которых могут прийти эксперты. В большинстве случаев экспертные системы решают задачи, которые трудно формализовать, или задачи, которые не имеют алгоритмического решения.

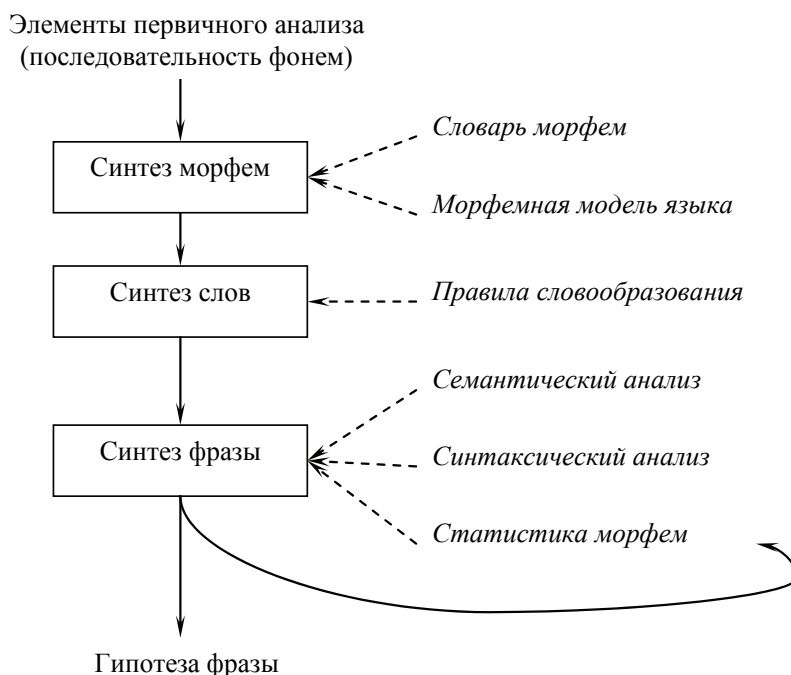


Рис.1. Синтез фразы из речевого сигнала

При разработке транскриптора целесообразно использовать статичную экспертную систему с четко заданными правилами и загодя известными элементами. Для того чтобы решить какие именно правила нужно использовать для транскрибирования нужно

1. Выделить элементы системы - звуки
2. Выделить правила транскрипции.
3. Расставить приоритет для каждого из правил.
4. Записать правила в формализованном виде.

В целом уровень дальнейший синтез предложения можно охарактеризовать следующей схемой, представленной на рис. 1.

Первоочередной задачей при синтезе слов является объединение фонем в последовательность морфем. Будем рассматривать следующие типы морфем: приставка, корень, интерфикс, суффикс, окончание, целое слово.

После распознавания фонем и подбора наиболее вероятных цепочек морфем получившийся набор гипотез далее используется для формирования цепочек слов. На основе каждой поступившей гипотезы фразы, представленной в виде последовательности морфем, формируется еще несколько гипотез, представленных уже последовательностью гипотез слов.

Одной из основных проблем распознавания слитной речи является особенность интерференции звуков на стыках слов. Таким образом, при разработке алгоритмов распознавания слитной речи необходимо учитывать не только правилам транскрибирования слов, но и уделить особенное внимание ряду правил, описывающих межсловные фонетические явления. Эти правила можно классифицировать следующим образом [6,7]:

1. Если в начале слова стоит сочетание фонем /йи/, причем гласная безударная, оно переходит в фонему /ы/ в случае, если первое слово заканчивается на твердую согласную. (город в Якутии /го!рат в йику!т'ии/ → /го!рат в йку!т'ии/).

2. Первая в слове гласная /и/ после всех твердых согласных переходит в фонему /ы/ (лист ивы /л'и!ст ы!вы/).

3. Безударные гласные редуцируются до полного исчезновения, если они находятся:

а) между одинаковыми согласными (мясо сырое /м'а!са сыро!йе/ → /м'а!с сыро!йе/)

б) после одной из парных по глухости-звонкости согласных и перед соответствующей парной согласной (степи большие /с'т'е!пи бал'шы!йе/ → /с'т'е!п' бал'шы!йе/).

4. Фонемы /т/ и /д/, стоящие после /с/ и /з/ соответственно, редуцируются до полного исчезновения (есть порох /йэ!с'т' по!рах/ → /йэ!с' по!рах/).

5. Фонемы /т/ и /д/, стоящие после /с/ и /з/ соответственно, редуцируются до полного исчезновения (хвост коровы /хво!ст каро!вы/ → /хво!с каро!вы/).

6. Согласная /й/ в конце слова редуцируется до полного исчезновения, если ей предшествует безударная гласная, а следующее слово начинается с любой фонемы, кроме ударной гласной (красный шар /кра!сный ша!р/ → /кра!сны ша!р/).

7. На стыке двух знаменательных слов глухие согласные /п/, /п'/, /т/, /т'/, /к/, /к'/, /ф/, /ф'/, /с/, /с'/, /ш/, /ш'/, /ц/, /ч/ озвончаются перед фонемами /б/, /д/, /г/, /з/ или /ж/. На стыке служебного и знаменательного слова внутрисловные правила ассимиляции по глухости-звонкости сохраняются, т.е. в положении перед глухими шумными согласными звонкие шумные согласные оглушаются, и на их месте выступают глухие шумные, в положении перед звонкими шумными согласными, кроме /в/, /в'/, глухие шумные озвончаются, и на их месте выступают звонкие шумные (с дороги /здара!г'и/, в лесу /вл'эсу!/).

8. Сочетание фонем /с'т'/ в конце слова переходит в фонему /щ/, если следующее слово начинается с /ч/ (есть чему /йэ!с'т' чэму!/ → /йэ!щ чэму!/).

9. Если на стыке двух слов находятся одинаковые согласные, то согласная первого слова редуцируется (лес сосновый /л'э!с сасно!вый/ → /л'э! сасно!вый/).

Следует обратить внимание на тот факт, что фонема связана со словом. Она вычисляется в словоформе не в морфеме, а в словоформе. Поэтому определение стыков слов является важнейшим фактором, обуславливающим сегментацию речевого потока на фонемы.

Интересно проследить эволюцию русского алфавита, из которого в конечном итоге бы-

ли удалены некоторые неиспользуемые буквы. Например буквы \square (кси), \square (пси) обозначали по два звука, кс и пс.

Изучение старославянского алфавита с его сочетаниями звуков (например, пси и кси) подтолкнуло к созданию дополнительного уровня обработки фонем – добавление правил, соответствующих некоторым исключенным из алфавита букв.

Последним этапом обработки является синтез фразы.

В ходе построения подобной системы на этапе генерации фразы мы можем встретить несколько типов ошибок, наиболее распространенные из которых:

1. Правильно определена граница слова, но возникает неоднозначность в его трактовке при переводе транскрипции морфем на естественный язык.

2. Границы слова определены неверно.

Таким образом построение лингвистического аппарата системы распознавания речи сводится к построению фреймов, представляющих собой последовательность уровней обработки сигнала. Каждый из них характеризуется исключительно своими правилами, и законами из разных отраслей наук - акустика, лингвистика, физики, математика и других. При построении системы распознавания речи как фреймовой модели, каждый из этих уровней будет представлять собой отдельный фрейм, решающий свои задачи независимо от других.

Кроме того, система распознавания слитной речи требует большого словаря и текстовой базы для обучения. Поэтому одним из важных уровней работы системы является ее обучение. Отдельным этапом обучения системы является прямой транскриптор.

Выводы

Система распознавания слитной речи является многоуровневой сложной системой. Ее построение требует тщательного анализа правил и методов, которые используются на каждом из ее уровней. Каждый из уровней использует свои методы, алгоритмы, входные данные. Сложность системы, большое количество данных, которые необходимо обработать, и необходимость принятия экспертного решения обуславливает выбор метода ее построения как экспертной системы, каждый узел которой является фреймом, который характеризуется своими правилами, методами, алгоритмами и данными.

Список литературы

1. Акишина А. А, Барановская. С.А. Русская фонетика на фоне общей. – М., 2007 – 104 с.
2. Карпов О.Н. Технология построения устройств распознавания речи. Монография. – Д., 2001. – 184 с.
3. Лучинкіна О.І., Карпов О.М., Аксьоненко П.Ю. Транскриптор як один з вузлів загальної експертної системи розпізнавання мови. Актуальні проблеми автоматизації та інформаційних технологій. Т.14. – Д., 2010 – с.68-78.
4. Попова Т.В., Губанова Т.В. Практическая фонетика русского языка. Учеб. пособие. – Тамбов, 2001.
5. Попов Е.В. Экспертные системы: Решение неформализованных задач в диалоге с ЭВМ. – М.: Наука. 1987.
6. Аванесов Р. И. Русская литературная и диалектная фонетика. – М.: Просвещение. – 1974.
7. Аванесов Р. И. Русское литературное произношение. – М.: Просвещение. – 1984.