

О.Н. Карпов, О.И. Лучинкина

Днепропетровский национальный университет им. О. Гончара

ПРИНЯТИЕ РЕШЕНИЯ НЕОДНОЗНАЧНОЙ СМЫСЛОВОЙ ТРАКТОВКИ РЕЧЕВОГО ВЫСКАЗЫВАНИЯ НА ГРАФЕ

Розглядається обробка речевого висловлювання на морфемному рівні. На основі досліджень ведеться створення системи синтезу мови з фонем.

Введение. Процесс обработки текста на естественном языке можно разбить на несколько уровней: разбор, анализ и синтез. На первых двух уровнях происходит обработка исходных данных и вывод их в формализованном виде. В задаче распознавания слитной речи на первых двух уровнях мы получаем некоторое фонемное представление звукового сигнала.

Во многих языках, в том числе и в русском, существуют достаточно регулярные правила чтения – правила соответствия между буквами и фонемами. Первоначально человек воспринимает звуки, не осмысливая их, на физическом уровне. На уровне синтеза происходит обработка полученных на первых двух уровнях данных и формирование ответа на естественном языке. Таким образом, идет процесс осмысления данных на основе синтаксиса, грамматики, семантики и т. д.

Основной проблемой, с которой сталкиваются разработчики ПО синтеза речи, – неоднозначность трактовки речевого высказывания. Приведем несколько примеров.

Пример 1. Абонемент. В украинской речи может быть распознано и как «або не мент» и как «абонемент».

Пример 2. Ананас. Может быть распознано как «а на нас», а может как «ананас».

Пример 3. Сезоне. Может быть распознано как «сизо не», а может как «сезоне».

То есть, на основе одного и того же набора фонем, мы можем построить несколько гипотез, каждая из которых может оказаться верной.

© О.Н. Карпов, О.И. Лучинкина, 2008

Обзор существующих решений. В данной статье рассматривается обработка речевого высказывания на морфемном уровне. На

сегодняшний день – данная задача очень актуальна для систем распознавания речи. Это объясняется растущими требованиями к дикторонезависимости разрабатываемых систем и увеличению размера словаря. Морфемный уровень встречается в системе Sirius [3]. Также программный продукт с морфемным уровнем обработки слитной речи, Vocative Russian ASR Engine, предлагает компания Vocative.

Использование морфемного уровня и пофонемного распознавания может привести к незначительной потере качества распознавания, которое может быть компенсировано на других уровнях обработки сигнала, однако, способствует повышению скорости распознавания.

Постановка задачи. Пусть задана входная последовательность фонем $\{x_k\}$, состоящая из N последовательно идущих фонем, не разделенных никакими знаками препинания и пробелами, представляющая собой фонетическую транскрипцию некоторого речевого высказывания.

Необходимо преобразовать данную последовательность $\{x_k\}$ в фразу $\{W_j\}$, состоящую из K слов, отвечающую данному фонетическому набору.

Ввиду того, что в результате анализа слитной речи мы не всегда можем определить ударные гласные, будем считать, что все гласные безударные.

Решение задачи. Текст, на естественном языке, представляет собой упорядоченный поток символов. Символы обрабатываются последовательно, один за другим, в порядке их местоположения в тексте.

На первоначальных уровнях анализа перед нами стоит задача разбиения высказывания на элементы первичного анализа. В качестве элемента первичного анализа будем брать фонемы. В качестве 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, наш алфавит:

- Гласные: а а! е е! о! и и! у у! ы ы! э!.
- Согласные: б, б' в в' г г' д д' ж 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 ш.

На уровне синтеза речевого высказывания происходит морфологический синтез последовательности фонем в последовательность отдельных слов.

Таким образом, происходит формирование внутреннего формализованного представления текста путем сопоставления последовательности фонем.

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

Реализация уровня синтеза текста требует наличия некоторой формальной модели. Рассмотрим модель слов русского языка (рис.1.).

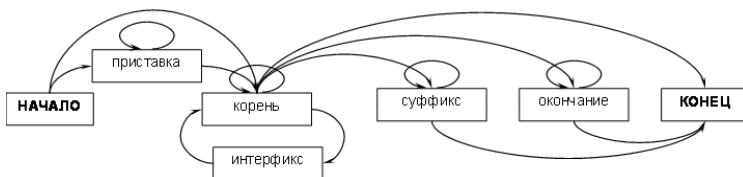


Рис.1. Правила построения слова из морфем

Сформулируем вышеизложенную схему на языке предикатов.

Введем переменные:

- i – номер текущей морфемы, $i = 1 \dots N$;
- j – номер текущего слова, $j = 1 \dots K$;
- M_{i-1} – предыдущая морфема;
- M_i – текущая морфема;
- W_j – текущее слово.

Введем следующие предикаты:

- $other(M_i)$ – морфема M_i является целым словом (например, предлог);
- $prefix(M_i)$ – морфема M_i является приставкой;
- $root(M_i)$ – морфема M_i является корнем;
- $interfix(M_i)$ – морфема M_i является интерфиксом;
- $suffix(M_i)$ – морфема M_i является суффиксом;
- $end(M_i)$ – морфема M_i является окончанием;
- $empty(M_i)$ – морфема M_i отсутствует;
- $combine(M_i, W_j)$ – добавить морфему M_i в конце слова W_j ;
- $new(M_i, W_j)$ – создать слово W_j и началом его будет морфема M_i ;
- $add(W_j)$ – добавить слово W_j в предложение.

На основе правил словообразования построим логическую систему:

$$\forall M_{i-1}, M_i \left(prefix(M_{i-1}) \wedge (prefix(M_i) \vee root(M_i)) \right) \rightarrow \\ \rightarrow combine(M_i, W_j)$$

$$\begin{aligned} & \forall M_{i-1}, M_i (\text{root}(M_{i-1}) \wedge (\text{root}(M_i) \vee \text{int erfix}(M_i) \vee \\ & \vee \text{suffix}(M_i) \vee \text{end}(M_i)) \rightarrow \rightarrow \text{combine}(M_i, W_j)) \\ & \forall M_{i-1}, M_i (\text{int erfix}(M_{i-1}) \wedge \text{root}(M_i) \rightarrow \text{combine}(M_i, W_j)) \\ & \forall M_{i-1}, M_i \{\text{suffix}(M_{i-1}) \wedge (\text{suffix}(M_i) \vee \text{end}(M_i)) \rightarrow \\ & \rightarrow \text{combine}(M_i, W_j)\} \end{aligned}$$

$$\begin{aligned} & \forall M_{i-1}, M_i (\text{end}(M_{i-1}) \wedge \text{end}(M_i) \rightarrow \text{combine}(M_i, W_j)) \\ & \forall M_{i-1}, M_i (\text{empty}(M_{i-1}) \wedge (\text{other}(M_i) \vee \text{prefix}(M_i) \vee \\ & \vee \text{root}(M_i)) \rightarrow \text{add}(W_{j+1}) \wedge \text{new}(M_i, W_{j+1})) \\ & \forall M_{i-1}, M_i (\text{other}(M_{i-1}) \wedge (\text{other}(M_i) \vee \text{prefix}(M_i) \vee \\ & \vee \text{root}(M_i)) \rightarrow \text{add}(W_{j+1}) \wedge \text{new}(M_i, W_{j+1})) \\ & \forall M_{i-1}, M_i (\text{root}(M_{i-1}) \wedge (\text{other}(M_i) \vee \text{prefix}(M_i) \vee \\ & \vee \text{root}(M_i)) \rightarrow \text{add}(W_{j+1}) \wedge \text{new}(M_i, W_{j+1})) \\ & \forall M_{i-1}, M_i (\text{suffix}(M_{i-1}) \wedge (\text{other}(M_i) \vee \text{prefix}(M_i) \vee \\ & \vee \text{root}(M_{i-1})) \rightarrow \text{add}(W_{j+1}) \wedge \text{new}(M_i, W_{j+1})) \\ & \forall M_{i-1}, M_i (\text{end}(M_{i-1}) \wedge (\text{other}(M_i) \vee \text{prefix}(M_i) \vee \\ & \vee \text{root}(M_{i-1})) \rightarrow \text{add}(W_{j+1}) \wedge \text{new}(M_i, W_{j+1})) \end{aligned}$$

Объединяя полученные условия, получим:

$$\begin{aligned} & \forall M_{i-1}, M_i ((\text{prefix}(M_{i-1}) \wedge (\text{prefix}(M_i) \vee \text{root}(M_i))) \vee \\ & \vee (\text{root}(M_{i-1}) \wedge (\text{root}(M_i) \vee \text{int erfix}(M_i) \vee \text{suffix}(M_i) \vee \\ & \vee \text{end}(M_i))) \vee (\text{int erfix}(M_{i-1}) \wedge \text{root}(M_i)) \vee \\ & \vee (\text{suffix}(M_{i-1}) \wedge (\text{suffix}(M_i) \vee \text{end}(M_i))) \vee \\ & \vee (\text{end}(M_{i-1}) \wedge \text{end}(M_i)) \rightarrow \text{combine}(M_i, W_j)) \vee \\ & \forall M_{i-1}, M_i (\text{empty}(M_{i-1}) \vee \text{root}(M_{i-1}) \vee \\ & \vee \text{other}(M_{i-1}) \vee \text{root}(M_{i-1}) \vee \text{suffix}(M_{i-1}) \vee \text{end}(M_{i-1})) \wedge \\ & \wedge (\text{other}(M_i) \vee \text{prefix}(M_i) \vee \text{root}(M_i)) \rightarrow \text{add}(W_{j+1}) \wedge \text{new}(M_i, W_{j+1})) \end{aligned}$$

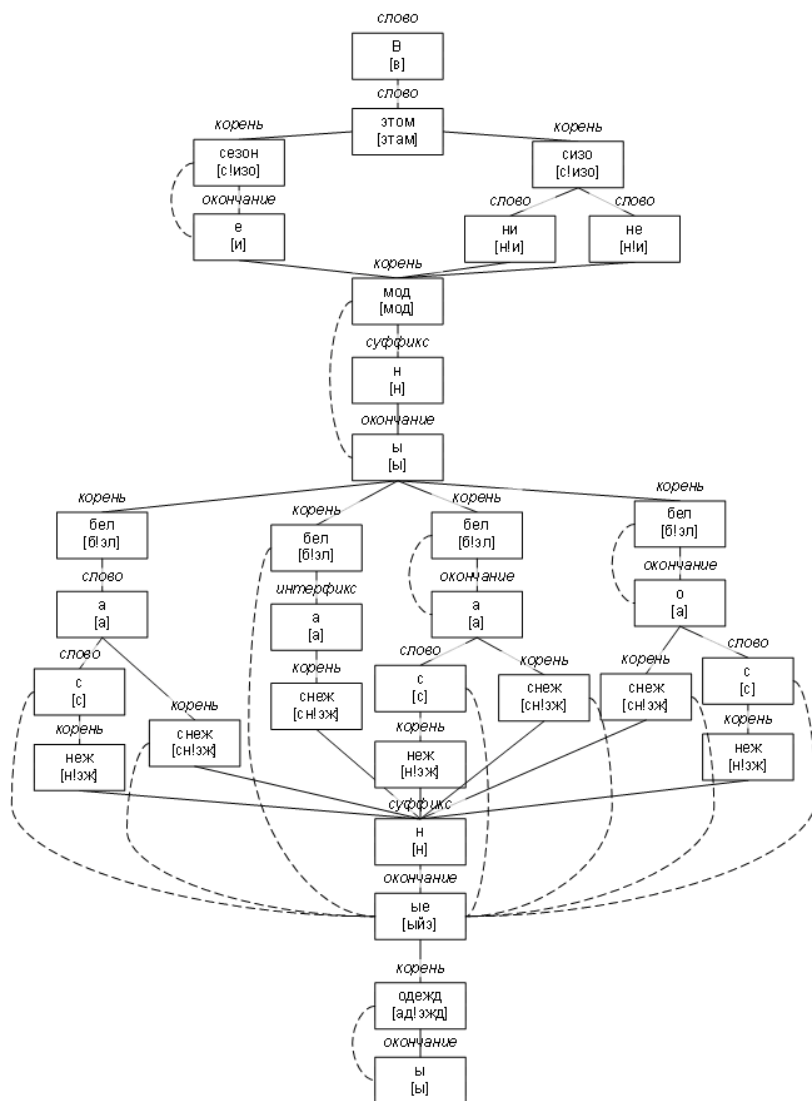


Рис.2. Разбор предложения «В этом сезоне модны белоснежные одежды»

Для русского языка задача морфологического разбора решается при помощи некоторых правил словообразования и словоизменения. В ходе синтеза речевого высказывания особую сложность представляет выделение отдельных слов из текста, так как в слитном тексте мы не можем с точностью говорить о символе «пробел». Результатом отсутствия границ слов является синтез множества морфем, отвечающих одному и тому же набору фонем, но разным границам слов. На основе гипотез, полученных в результате морфологического анализа последовательности фонем, можно построить некоторый конечный граф. Каждой его вершине соответствует очередная распознанная морфема (то есть последовательность фонем).

Рассмотрим следующий пример. Возьмем фразу «В этом сизо не модны белоснежные одежды». Результатом разбора последовательности морфем, отвечающей данной фразе, мы получим 21 возможную гипотезу (рис.2.).

В ходе исследования графа становится очевидной неоднозначность трактовки некоторых отдельных частей речевого высказывания. Мы не можем с точностью сказать, какое из высказываний является верным, а какое нет. Мы можем только строить некоторые предположения относительно того, что та или иная часть графа наиболее отвечает нашему речевому высказыванию.

В ходе построения гипотез также возможен вариант, когда мы правильно определяем границу слова, но возникает неоднозначность в его трактовке при переводе транскрипции морфем в естественную речь. Так на рис.2 это можно увидеть в словах «бело» и «бела». Это обусловлено тем, что транскрипция для разных морфем может быть одинаковой.

Одним из ключевых факторов в построении системы принятия решения о выборе гипотезы является статистика встречаемости слов и следования морфем в словах. Как и в естественной речи, зная тему разговора, часто достаточно услышать что-то весьма неопределённое, чтобы догадаться, о чём идёт речь. И наоборот: верно услышанное слово может быть принято за услышанное неверно, если оно не вяжется со смыслом или темой разговора. Таким образом, знание темы разговора позволяет нам принять верное решение, потому что определенная предметная область подразумевает специфический словарь. Значит, имея данные относительно встречаемости слов в контексте предметной области, мы можем с некоторой определенностью принять решение о верности той или иной фразы.

В ходе изучения данной проблемы ведется разработка программного продукта, осуществляющего синтез конечной конфигурации исходных символов. На основе фонем при помощи их базы строят некоторые гипотезы относительно исходного речевого высказывания. Как видно из приведенного выше примера при анализе фразы «В этом сезоне модны белоснежные одежды» результатом работы программы стало 21 грамматически и морфологически верная гипотеза. Причем были получены как положительные, так и отрицательные (например, «В этом сизо не модны белоснежные одежды») гипотезы.

Принадлежность слова предметной области определяется встречаемостью слова и его форм в контексте предметной области. Таким образом, мы приходим к необходимости обучения системы, то есть в сборе статистики встречаемости слов и следования морфем в них. Сбор статистических данных осуществляется на основе обработки текстового материала предметной области. Входной материал преобразуется в текстовую транскрипцию, которая, в свою очередь, преобразуется в последовательность морфем. Минимально необходимый словарь должен создаваться на базе наиболее употребляемых фраз и выражений.

На основе полученных в результате обучения данных для каждого расхождения в графе мы можем найти некоторую вероятность появления. Тогда верность цепочки – это число, находящееся в интервале от 0 до 1. Максимальное значение из этого интервала дает нам максимальную уверенность в верности цепочки, а минимальное значение – в ее ошибочности.

Анализ различных тематических предложений показал, что в полученных вариантах существуют некоторые слова, которые не несут большой смысловой нагрузки. К ним, например, относятся предлоги «в», «а», «с». При обработке цепочки с такими словами их стоит брать с некоторым коэффициентом или не учитывать вовсе.

Таким образом, очевидно, что недостаточно просто создать систему, которая бы работала на некоторых правилах построения слова из морфем. Также очень важной особенностью подобной системы является сбор статистических данных относительно встречаемости морфем в контексте предметной области.

Рассмотренный выше пример соответствует предметной области «Мода». На основе текстового материала данной области был создан некоторый небольшой словарь из 200 слов. Анализировался текст на 600 знаков, отвечающий данной предметной области.

Результаты анализа для данного предложения отображены в таблице 1.

Таблица.1

Статистика появления морфем для фразы «В этом сезоне модны белоснежные одежды»

Мор-фема	Тип	Встре-чаемость	Вероятность перехода				
			При-ставка	Корень	Интер-фикс	Суффикс	Окон-чание
а	окончание	0,01600	–	–	–	–	0
о	окончание	0,00900	–	–	–	–	0
е	окончание	0,02600	–	–	–	–	0
о	интерфикс	0,00800	–	0,2500	–	–	–
а	слово	0,00200	0	0			
не	слово	0,01800	0	0	–	–	–
ни	слово	0,00200	0	0	–	–	–
с	слово	0,01800	0	0	–	–	–
бел	корень	0,00400	0	0	0,5650	0	0,4350
нежн	корень	0,00000	0	0	0	0	0
сезон	корень	0,00800	0	0	0	0	1,0000
сизо	корень	0,00000	0	0	0	0	0
снеж	корень	0,00400	0	0	0	1,0000	0
н	суффикс	0,05400	–	–	–	0	1,0000

Таким образом, видим, что слово «сизо» не присутствует в контексте данной предметной области, в результате чего мы можем отбросить гипотезу, в которой встречается слово «сизо». В обучающем тексте также не встречалось слово «бел». Однако вероятности следования появления интерфикса и окончания незначительны. Следовательно, не стоит исключать слова «бело» и «бела», хотя наиболее вероятным является вариант со словом «белоснежный».

Выводы. Задача синтеза речи – сгенерировать конечную конфигурацию слов по некоторой последовательности символов. При решении данной задачи мы сталкиваемся с проблемой неоднозначной трактовки исходного высказывания, то есть мы получаем несколько конкурирующих гипотез. Конечно, мы не можем точно определить, какая из полученных гипотез верна, а какая нет. Однако мы можем разработать некоторый аппарат,

который бы на основе статистики вхождения слов в контекст предметной области, смог бы выбрать нужную гипотезу с наибольшей точностью.

Для решения данной задачи разрабатывается некоторая система синтеза речи. Задачей ее является построение конечного графа, из которого мы четко можем определить места, в которых возникла неоднозначность выбора. Тогда задача поиска наиболее верного решения сводится не к проверке всех цепочек полностью, а к оценке некоторых цепочек графа. Для этого предложен метод вероятностного морфемного представления речи и предложен алгоритм, позволяющий на основе статистических данных построить фразу из последовательности фонем.

Предложенная схема в состоянии выполнять морфологический разбор и синтез. Однако, при выборе решения мы не можем быть полностью уверенными, что оно верное.

Библиографические ссылки

1. **Ронжин А.Л.** Система автоматического распознавания русской речи SIRIUS. / А.Л. Ронжин, А.А. Карпов, И.В. Ли, // Научно-теоретический журнал «Искусственный интеллект», №3 – 2005, С. 590-601.
2. **Богданов Д.С.** Технология формирования речевых баз данных / Д.С. Богданов, А.В. Брухтий, О.Ф. Кривнова, А.Я. Подрабинович, Г.С. Строкин // Организационное управление и искусственный интеллект. Института системного анализа РАН. – 2003.
3. **Люггер Джордж Ф.** Искусственный интеллект: стратегии и методы решения сложных проблем. Четвертое издание. – М, 2005 – 864 с.
4. **О.Н. Карпов.** Исследование вариативности трактовки фонетической структуры речевых высказываний. / О.Н. Карпов, О.И. Лучинкина, // Научно-теоретический сборник «Проблемы математического моделирования» – Днепропетровск, 2008.

Надійшла до редколегії 26.06.08