

К КОМПЬЮТЕРНОМУ МОДЕЛИРОВАНИЮ ВАРЬИРОВАНИЯ РУССКОГО КОРНЯ

А.А. Кретов,

О.Г. Артемова,

А.С. Шудрикова

email: kretov@rgph.vsu.ru

Введение

Обработка Текстов на Естественном Языке (Natural Language Processing, NLP) — одна из существенных составляющих искусственного интеллекта (ИИ).

Это совместное направление компьютерных наук и математической лингвистики.

Подходы к формализации языка

- «голливудский» подход - предпочтение отдается *словарю* как декларативной части алгоритма
- «суворовский» подход - предпочтение отдается *грамматике* как его операционной части.

Введение

Описывать варьирование формы корня целесообразно в виде **исчисления**.

Для этого необходимо иметь набор исходных элементов (фонем) и правил их преобразования в наблюдаемый в речи морф в зависимости от позиции, в которой находится корень (как морфема) с составляющими его форму фонемами.

Терминология

Претеза – факультативный звук в **начале** корня

Посттеза – факультативный звук в **конце** корня

Фузия - процесс слияния звуков (фонов), представляющих фонемы соседних морфем, в один звук.

Фузиат - результат *фузии* – неразрывно сплавившиеся морфы соседних морфем.

*леЧь < *леГ=ТЬ*

ЧГТ - бифон

Варьирование корня ЛЕГ-

Корень	Позиция 1	Позиция 2	Позиция 3	Словоформа
<i>лег</i>	л	е	г	ЛЕГла
<i>леж</i>	л	е	ж	поЛЕЖит
<i>лёг</i>	л	ё	г	(он) ЛЁГ
<i>лёж</i>	л	ё	ж	поЛЁЖивал
лог	л	о	г	подЛОГ
лож	л	о	ж	поЛОЖит
лаг	л	а	г	поЛАГать
*леч	л	е	ч	ЛЕЧь

Правила преобразований в морфах

- /лег/ > /лѐг/: если {Е} находится под ударением перед твёрдым согласным, то {Е} => /'ó/ = Ě.
- /лог/ > /лаг/: если {О} перед =á, то /o/ => /a/.
- /г/ > /ж/: если {Г} на суффиксальной стыке находится перед {И/Е/Ь}, то {Г} => /ж/.
- {œ} > [e/o]: если в глаголе, то /e/, если в имени, то /o/.

Преобразования в фузиатах

- Ср. *бы-ли* – *бы-ть*, *да-ли* – *да-ть*, *лег-ли* – *лечь* (= **лег-ть*).
- Следовательно, $Ч = Г+Т$
- Правило: если после К/Г на суффиксальном стыке следует Т, то $КТ/ГТ \Rightarrow Ч^{КТ/ГТ}$.
- *Бифон* – целостное представление {Г} корня и {Т} суффикса инфинитива нельзя сопоставлять с *фонами* – отдельными и независимыми представлениями его частей.

Корень в русском слове имеет две факультативные позиции:

- 1) Претезы: ничего не меняющего звука перед ним (например, *ОСЬМушка* ~ **в***ОСЬМушка*; *КОРа* ~ **с***КОРняк*);
- 2) Посттезы: ничего не меняющего звука после него: например, *страд-∅-ать*, *стра(д>∅)-х* и *стра(д>с)-ть*.

Исследование

Пример позиционного распределения фонем в корневых морфах

№	Корень в словаре	морфы	П1	П2	П3	П4	П5
1	<i>кóпот</i>	кóпот	к	ó	п	о ^б	т
2	<i>(кáпч)</i>	кáпøч	к	á ^о	п	ø ^б	ч ^{тй}
3	<i>копт'</i>	копøт	к	о	п	ø ^б	т
4	<i>(копч)</i>	копøч	к	о	п	ø ^б	ч ^{тй}

Алгоритм анализа базы данных СМОМЯ-86 :

- разбиение корней, упорядоченных по числу алломорфов: от 2 до 8;
- анализ каждого множества отдельно с соблюдением последовательности анализа от коротких морфов к длинным;
- для каждого морфа создаются столбцы. В каждой ячейке каждого столбца соответствующего морфа размещается одна буква, соответствующая одному фону (включая нулевые);
- Для заполнения ячеек используется функция ПСТР;

Алгоритм анализа базы данных СМОМЯ-86 :

- после заполнения всех ячеек с помощью функции ЕСЛИ создаются 8 столбцов сравнений, обозначаемых $s_1, s_2, s_3, \dots, s_8$. При совпадении значений в соответствующих позициях функция ЕСЛИ имеет значение 0, при несовпадении позиций – 1. Для определения суммы значений столбцов в каждой строке создается отдельный столбец СумС;
- сортировка данных по столбцу СумС в порядке возрастания дает информацию о количестве аллофонов в соответствующей позиции;
- столбцы s_1 - s_8 со значением 0 свидетельствуют об отсутствии чередований и в дальнейшем не рассматриваются;

Алгоритм анализа базы данных СМoМЯ-86 :

- анализ каждой позиции осуществляется по следующему алгоритму (пример для 4-х морфных корней):
 - С1 – сравнение позиции X морфемы 1 с позицией X морфемы 2;
 - С2 – сравнение позиции X морфемы 1 с позицией X морфемы 3;
 - С3 – сравнение позиции X морфемы 1 с позицией X морфемы 4;
 - С4 – сравнение позиции X морфемы 2 с позицией X морфемы 3;
 - С5 – сравнение позиции X морфемы 2 с позицией X морфемы 4;
 - С6 – сравнение позиции X морфемы 3 с позицией X морфемы 4;
- Члены чередований упорядочиваются по алфавиту.

Требования по корректному определению позиций фонем в корне

- 1) позиционное совмещение морфов одного корня должно обеспечивать максимальное совпадение фонем (символов) в столбцах;
- 2) при определении состава фонем из рассмотрения исключаются последовательности, содержащие полифоны;
- 3) устранением метатез и сдвигов фонем (символов) в речевой цепи восстанавливаются их дометатезные позиции;
- 4) перед определением позиций в корне буквы Е, Ё, Ю, Я следует заменить на сочетание Й с Е, О, У или А, соответственно, в позиции после букв, обозначающих гласные, и в абсолютном начале слова.

Чередования в русских корнях (V - гласный, C – согласный, S - сонант)

Черед	Раз	Тип	Черед	Раз	Тип	Черед	Раз	Тип
а:о	136	v	а:о:ы	4	v	е:ё:ь:ø	2	v
к:ч	129	c	е:и:о:ь	4	v	е:ё:о	2	v
е:ё	129	v	е:ё:ю	4	v	е:и:ø	2	v
г:ж	93	c	и:о:ь	4	v	ё:ø	2	v
о:ø	78	v	и:ø	4	v	о:е	2	v
е:ø	50	v	у:ø	4	v	б:с	1	c
и:ы	20	v	у:ю	4	v	к:т	1	c
д:ø	19	c	с:т:ø	3	c	к:ц:ч:ш	1	c
в:ø	15	s	в:ы	3	s	к:ч:щ	1	c
ц:ч	12	c	м:ø	3	s	к:ш	1	c
т:ø	11	c	ё:ь	3	v	с:т:ч	1	c
е:ё:ø	10	v	б:ø	2	c	т:ч:ø	1	c
е:о	9	v	г:ø	2	c	р:ø	1	s
е:ь	9	v	г:ж:ø	2	c	а:е:ё:и:о:ø	1	v
о:ы:ø	9	v	г:к	2	c	а:е:ё:и:о	1	v
к:ø	8	c	д:с:ø	2	c	а:е:ё:и:я	1	v
и:о:ø	7	v	д:т:ø	2	c	а:е:и:о:ø	1	v
н:ø	6	s	з:ø	2	c	а:е:о:ø	1	v
а:о:ø	6	v	к:т:ц:ч	2	c	а:е	1	v
е:и:о:ø	6	v	к:ц	2	c	е:ё:и:ø	1	v
к:ч:ø	5	c	к:ц:ч	2	c	е:ё:и:о	1	v
а:е:о	5	v	п:ø	2	c	е:ё:о:ø	1	v
е:о:ø	5	v	с:ø	2	c	е:и:о	1	v
о:ы	5	v	с:т:ч:ø	2	c	е:ы:ø	1	v
г:ж:з:ø	4	c	а:е:и:ø	2	v	е:и	1	v
г:ж:ч	4	c	а:и:о:ø	2	v	и:й:ы	1	v
д:с	4	c	а:и:о:ь	2	v	и:й	1	v
к:т:ø	4	c	а:и:ø	2	v	о:у:ы	1	v
с:т	4	c	а:и:о	2	v	о:у	1	v
а:е:ё:о	4	v	е:ё:и:о:ø	2	v	у:ы	1	v

Заключение

- Компьютерное моделирование варьирования русского корня предполагает как определение множества фонем, так и определение множества позиций, в которых фонема реализуется в том или ином фоне.
- Исчисление и классификация этих позиций составляет ближайшую перспективу нашей исследования.

Спасибо за внимание!