Распознавание речи. Произносительные словари

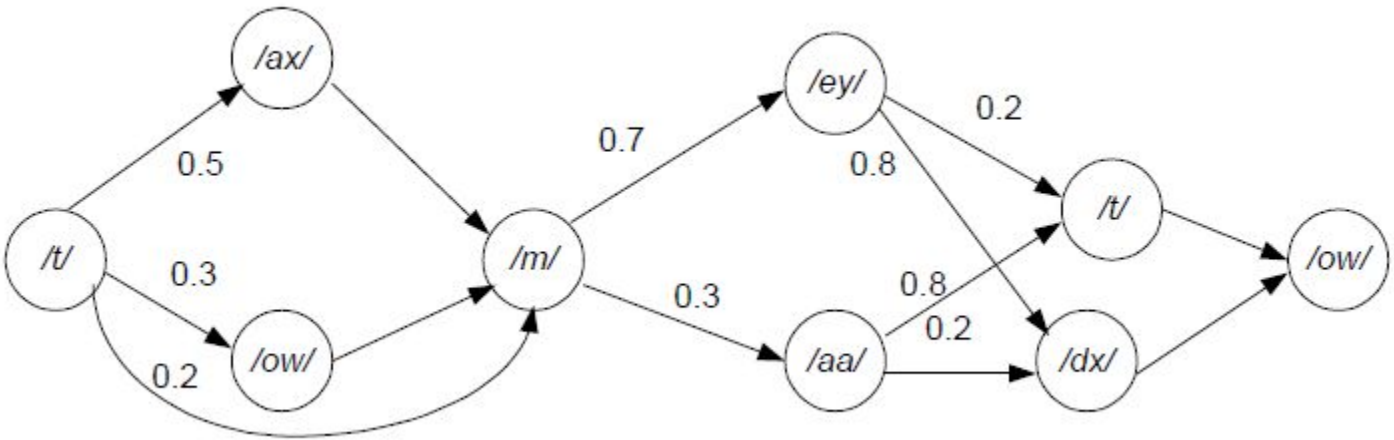
П. А. Холявин

p.kholyavin@spbu.ru

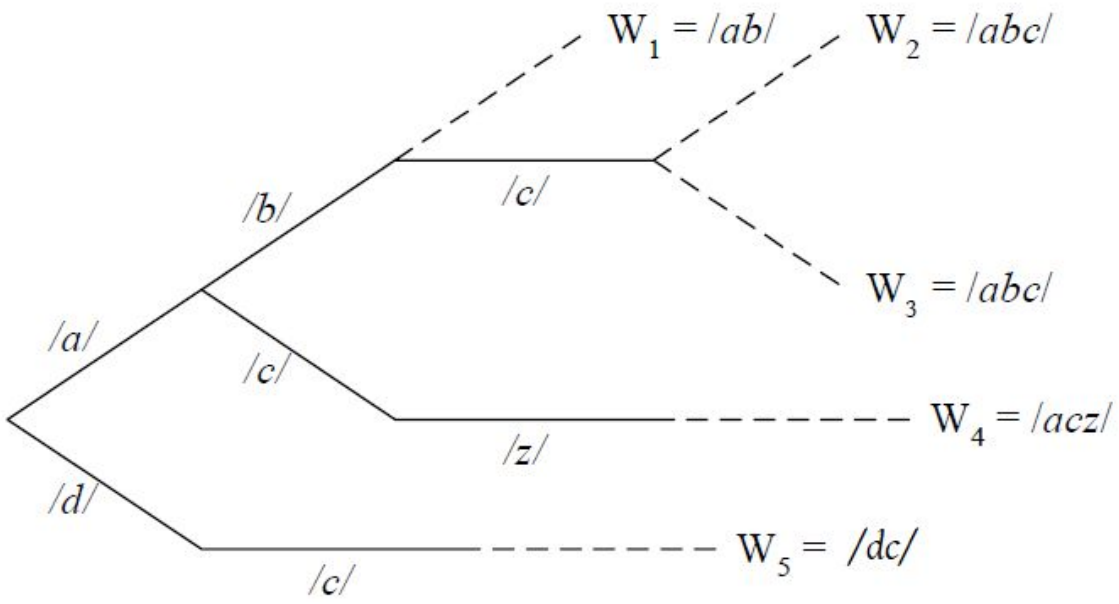
13.03.2024

Произносительные словари

```
нарисо1ван      n a r' i s o0 v a n
нарисо1вано     n a r' i s o0 v a n a
нарисо1ваны     n a r' i s o0 v a n y
нарисова1ть     n a r' i s a v a0 t'
```



Представление лексикона как дерева



Допустим, у нас есть слова *ban, band, banned, bat, beef.* Как будет выглядеть дерево?

Методы создания словарей

1. Экспертный

2. Автоматический (grapheme-to-phoneme, G2P)
а) по правилам
б) с помощью машинного обучения
в) с помощью систем распознавания речи

3. Гибридный (?)

Автоматическая транскрипция

G2P (Grapheme-to-Phoneme)

1. Определение фонемного состава (а какие фонемы?)
2. Определение фонетического качества звуков

+ проблема вариативности: какой вариант выбрать для системы?

Вариативность

Орфография	**CORPRES**	**SibLing**
двойной согласный *военного*	/vajé**n**ava/	/vajé**nn**ava/
выпадение фонемы /j/ *мавзолеем*	/mavzalʲé**j**im/	/mavzalʲéim/
ассимиляция по мягкости *поднялась*	/pa**dn**ʲilás'/	/pa**dʲnʲ**ilás'/

Фонемная транскрипция

1) По словарю

```
PHOENIX  F IY1 N IH0 K S
PHOENIX'S  F IY1 N IH0 K S IH0 Z
PHONE  F OW1 N
PHONE'S  F OW1 N Z
PHONED  F OW1 N D
PHONEMATE  F OW1 N M EY2 T
PHONEME  F OW1 N IY0 M
PHONEMES  F OW1 N IY0 M Z
PHONEMIC  F AH0 N IY1 M IH0 K
PHONES  F OW1 N Z
PHONETIC  F AH0 N EH1 T IH0 K
PHONETICALLY  F AH0 N EH1 T IH0 K L IY0
PHONETICS  F AH0 N EH1 T IH0 K S
PHONEY  F OW1 N IY0
PHONIC  F AA1 N IH0 K
PHONICS  F AA1 N IH0 K S
PHONING  F OW1 N IH0 NG
PHONOGRAPH  F OW1 N AH0 G R AE2 F
PHONOGRAPHS  F OW1 N AH0 G R AE2 F S
PHONOLOGICAL  F OW2 N AH0 L AA1 JH IH0 K AH0 L
PHONOLOGY  F AH0 N AA1 L AH0 JH IY2
```

Фонемная транскрипция

2) По правилам

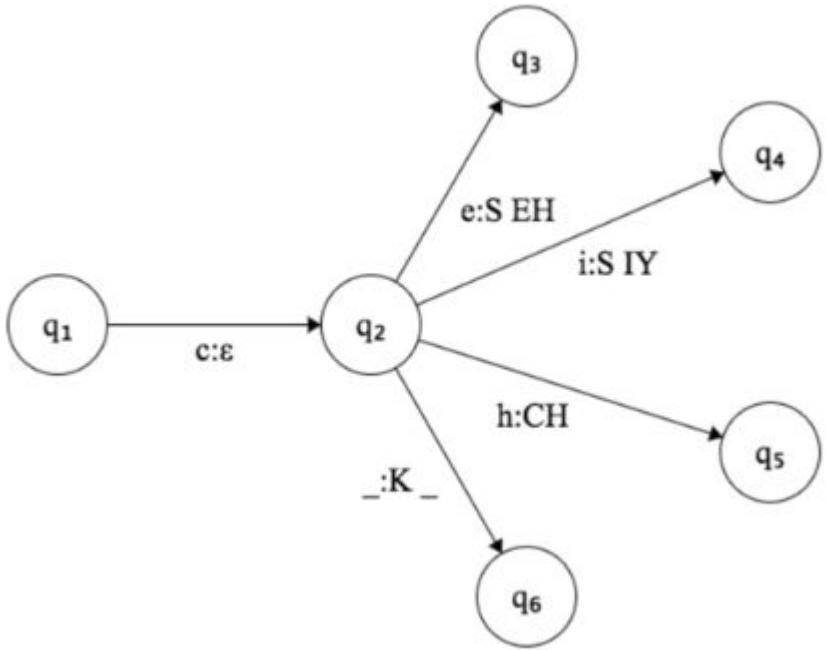
Правила могут кодироваться в конечных автоматах, …

```
def final_devoicing(trans):
    """apply to individual words"""
    for cur in reversed(trans):
        if cur.name not in data.obstruents:
            break
        cur.devoice()

def reflexive_suffix(trans):
    """apply before a_to_i() to individual words"""
    if trans.last.match(["t|t'", "s'", "a"], reverse=True):
        trans[-3].name = "c"
        trans.remove(trans[-2])

def genitive_ending(trans):
    """apply before vowel_reduction() to individual words"""
    if trans.last.match(["o|e", "g", "o"], reverse=True):
        trans[-2].name = "v"
    elif trans.last.match(["o|e", "g", "o", "s'", "a"], reverse=True):
        trans[-4].name = "v"
```

Фонемная транскрипция

3) Статистические методы и машинное обучение:

- Марковские цепи
- FST (конечные автоматы)
- Нейронные сети: LSTM, трансформеры, …



Стыки слов

Кот бежит /kod bʲiʒɨt/

Отец дома /atʲe[dz] dóma/

Раз в жизни /raʒ (v) ʒɨzʲnʲi/

Фонетическая транскрипция

1. Отражение коартикуляции звуков, влияние ударения

2. Стили произношения и типы произнесения

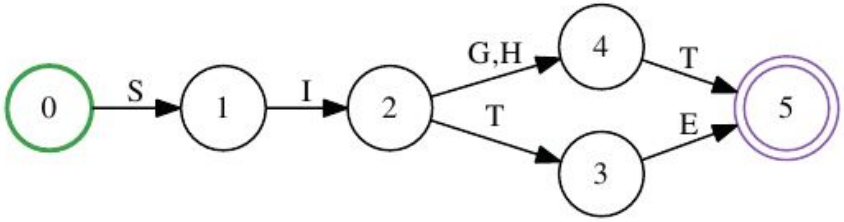
Типы произнесения: полный и неполный (невозможно восстановить фонемный состав)

[gəvɐˈrʲit]
[gəˈrʲit]
[grʲit]

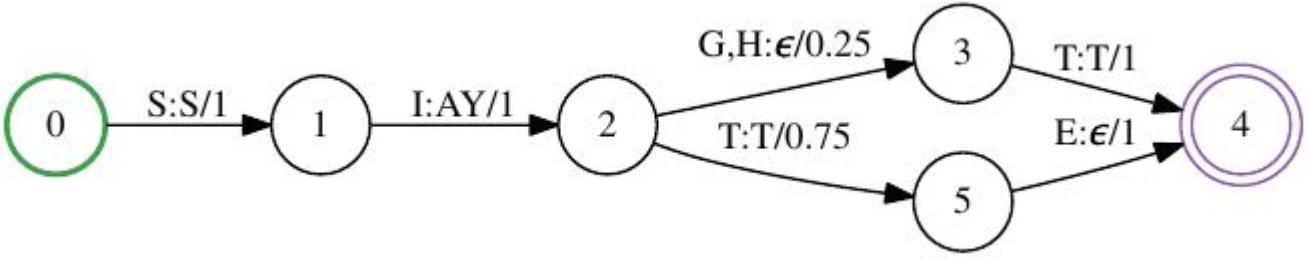
3. Влияние других просодических явлений

Взвешенные конечные преобразователи (WFST)

1. Конечный автомат (finite-state acceptor)



2. Конечный преобразователь (finite-state transducer)



Phonetisaurus

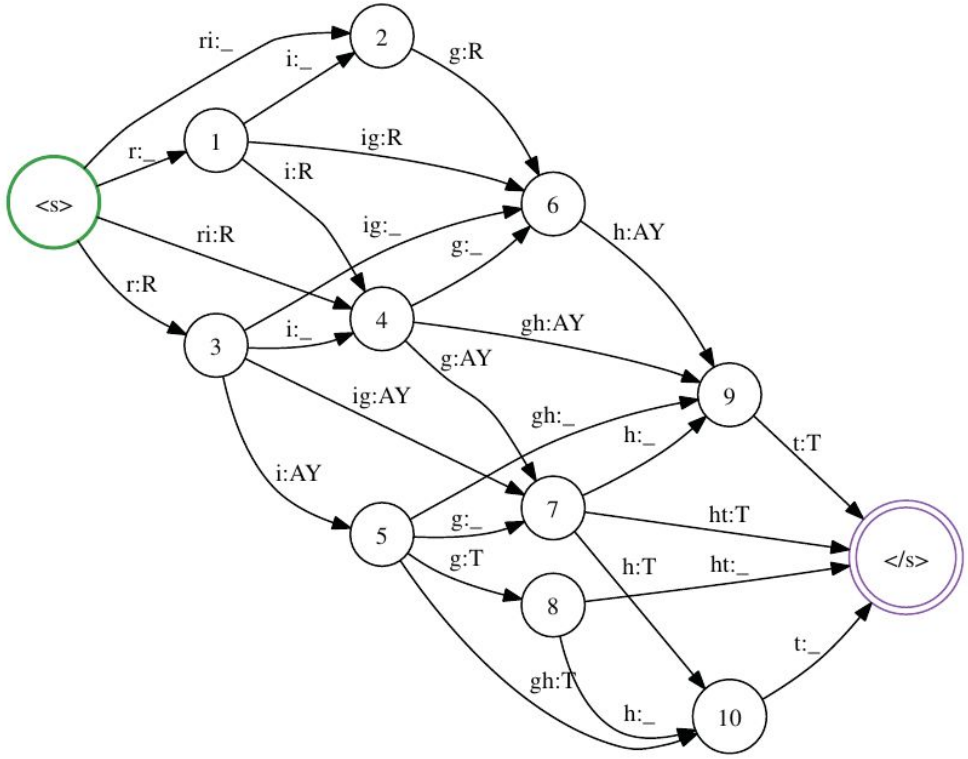
1. Выравнивание обучающего материала (alignment)

T	E	S	T		T	A	S	T	E
|	|	|	|		|	|	|	|	|
T	EH	S	T		T	EY	S	T	ϵ

T	E	X	T	B	O	O	K		T	E	X	T	B	O,O	K
|	|	|	|	|	|	|	|		|	|	|	|	|	|	|
T	EH	K	S	T	B	UH	K		T	EH	K,S	T	B	UH	K

Phonetisaurus

1. Выравнивание обучающего материала



Phonetisaurus

1. Выравнивание обучающего материала

Algorithm 1: EM-driven M2One/One2M

Input: *sequence pairs*, *seq1_max*, *seq2_max*, *seq1_del*, *seq2_del*
Output: γ, AlignedLattices

```
foreach sequence pair (seq1, seq2) do
    lattice ←Seq2FST(seq1, seq2, seq1_max, seq2_max, seq1_del, seq2_del)
foreach lattice do
    Expectation(lattice, γ)
Maximization(γ, total);
```

Algorithm 2: Expectation step

Input: *AlignedLattices*
Output: γ, *total*

```
foreach FSA alignment lattice F do
    α ← ShortestDistance(F)
    β ← ShortestDistance(F^R)
    foreach state q ∈ Q[F] do
        foreach arc e ∈ E[q] do
            v ← ((α[q] ⊗ w[e]) ⊗ β[n[e]]) ⊘ β[0];
            γ[i[e]] ← γ[i[e]] ⊕ v;
            total ← total ⊕ v;
```

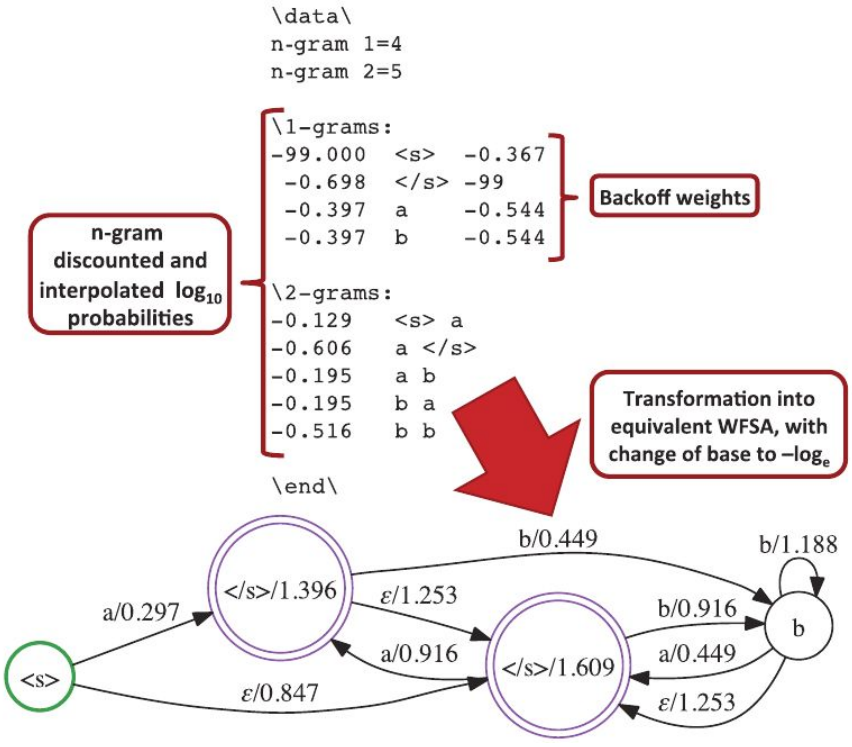
Algorithm 3: Maximization step

Input: γ, *total*
Output: γ_{new}

```
foreach i[e] ∈ γ do
    γ[i[e]]_new ← γ[i[e]] ⊘ total
```

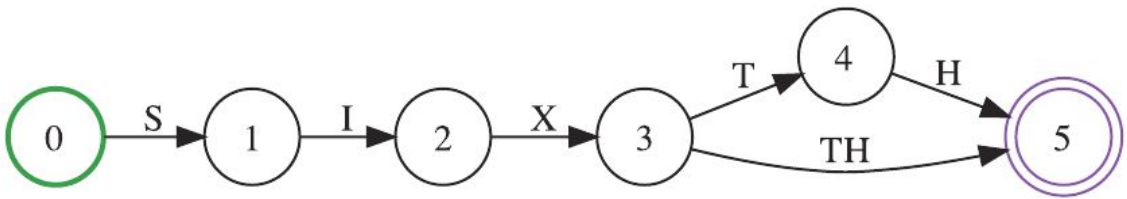
Phonetisaurus

1. Обучение n-граммной модели на графемах



Phonetisaurus

1. Декодирование (генерация транскрипции):
2. Создание конечного акцептора



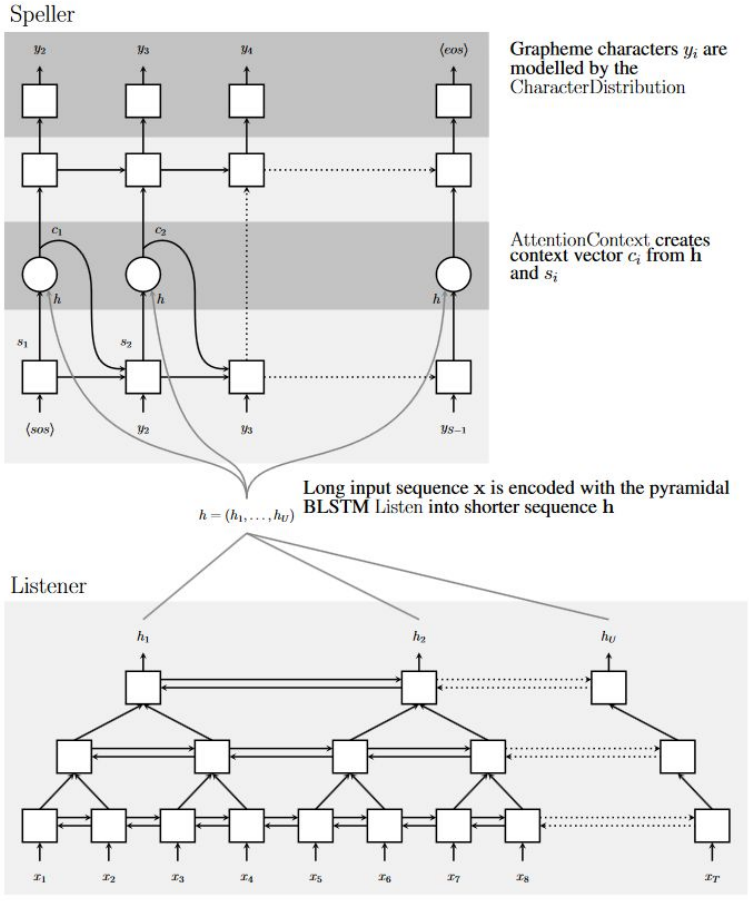
3. Композиция с n-граммной моделью

4. Поиск наилучшего пути

Другие методы

1. Рекуррентные нейронные сети (RNN)
2. Listen, Attend and Spell (BiLSTM + attention)
3. Трансформеры

Listen, Attend and Spell



Генерация транскрипций с помощью ASR

1. Автоматическая разметка на слова
2. Создание матрицы ошибок (confusion matrix)
3. Обучение N-граммной модели для “фонем”
4. Пофонемное распознавание слов/последовательностей слов, для которых нужны транскрипции
5. Удаление транскрипций, которые отличаются от существующих звуками, которые часто путаются системой (п. 2)

Генерация транскрипций с помощью ASR

rank	pronunciations
(1)	? AI N T ER M IE N
(2)	? AI N E2 N T ER M IE N
(3)	N T ER M IE N
(4)	N E2 N T ER M IE N
(5)	? AI N E2 N T ER M IE N
(6)	? E N T ER M IE N

Pronunciation Candidates for "einen Termin"

rank	pronunciations
(1)	N O X AI N T ER M IE N
(2)	N O X ? AI N T ER M IE N
(3)	N O X AI N E2 N T ER M IE N
(4)	N O X E2 N T ER M IE N

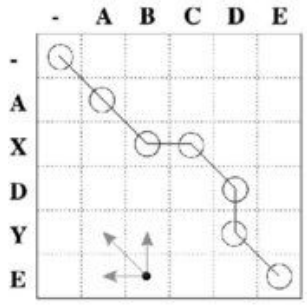
Pronunciation Candidates for "noch einen Termin"

Динамические словари

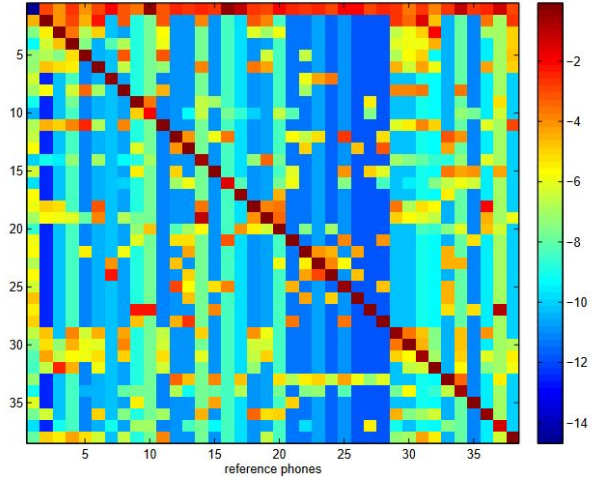
1. Выбор варианта в зависимости от:
2. Темпа
3. Длины слова
4. Фонетического контекста
5. Лексического контекста

Оценка качества транскрипций

1. Word Error Rate
2. Phone Error Rate
3. Phone-based dynamic programming (PDP)



A	A	Match
X	B	Mutate
-	C	Insert
D	D	Match
Y	-	Delete
E	E	Match



Спасибо за внимание!