

# МЕТОДЫ ОБНАРУЖЕНИЯ ЭМПИРИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ (Вычислительные системы)

2001 год

Выпуск 167

УДК 621.3:534.4:003.035

## ТЕКСТ МИНИМИЗИРОВАННОЙ ДЛИНЫ ДЛЯ ОБУЧЕНИЯ СИСТЕМ РАСПОЗНАВАНИЯ СЛИТНО ПРОИЗНОСИМЫХ ЧИСЕЛ<sup>1</sup>

А.В.Кельманов, С.А.Хамидуллин

### В в е д е н и е

Несмотря на заметный прогресс в области создания интеллектуальных средств устного диалога "человек-компьютер", проблема распознавания речевых сигналов полностью еще не решена. Известные распознающие системы по-прежнему не способны с высокой надежностью воспринимать устную слитную речь. Отсутствие приемлемого решения проблемы является следствием неадекватности имеющихся моделей речевых сигналов.

Напомним, что фонема — это минимальная смысловозначительная звуковая единица языка, а триграмма — это сочетание (комбинация) трех элементов из алфавита фонем.

Существующие модели слабо отражают объективное взаимовлияние произносимых звуков друг на друга, выражающееся в фонетической изменчивости. Эти модели опираются на хорошо известные лингвистические данные о звуковом строе языка, которые в своей совокупности не дают полного представления об

---

<sup>1</sup>Работа выполнена при финансовой поддержке РФНФ, проект № 00-04-00217а.

имеющем место триграммном взаимодействии фонем в устной речи. Под триграммным взаимодействием (изменчивостью) фонем понимается зависимость произносимой фонемы от предыдущей (левой) и последующей (правой) фонемы.

Неполная картина о триграммных связях в языке, а также отсутствие адекватных моделей триграммного взаимовлияния фонем не позволяют приемлемым алгоритмическим образом (без потери надежности распознавания) выделить (на этапе обучения компьютера), а затем избирательно подавить (на этапе распознавания) мешающий эффект фонетической изменчивости.

Для уточнения эффектов триграммного взаимовлияния фонем в устной речи требуется обработка орфографических и транскрибированных репрезентативных текстов с целью получения статистических данных о частотах встречаемости триграмм, а также выявления триграмм, не допустимых в языке. Без этих данных построение математической модели генерации устной речи, учитывающей триграммное взаимодействие фонем, представляется невозможным. С другой стороны, для построения системы распознавания устной речи требуется универсальный обучающий текст, удовлетворяющий специфическим требованиям, о которых будет сказано ниже. Здесь отметим лишь, что одно из главных и очевидных требований к этому тексту состоит в том, что его транскрипция должна полностью отражать все допустимые в языке эффекты фонетической изменчивости, в том числе и эффекты триграммного взаимовлияния фонем.

Проблема формирования универсальных обучающих текстов рассматривалась в ряде работ. В [1] был представлен английский текст, транскрипция которого содержит все биграммы (комбинации из двух фонем) языка. Чтение этого текста занимает около 20 минут. Аналогичная задача для русского языка решена в [2]. Биграммные модели звукового строя языка позволили создать более совершенные системы распознавания устной речи, в основе которых лежат скрытые марковские последовательности. К сожалению, модели биграммного взаимодействия не позволили полностью закрыть проблему распознавания слитной речи.

В последние годы внимание исследователей было сосредоточено на изучении триграммного состава языка. По-видимому,

первые результаты в этом направлении получены в [3], где описан синтезатор русской речи, основанный на триграммной модели русского языка. Чуть позже аналогичные результаты были получены в [4-5]. По оценкам экспертов модель триграммного фонетического взаимодействия, по сравнению с биграммной моделью, позволила существенно улучшить качество (разборчивость) синтезированной речи и заметно приблизила синтезированную речь к натуральной. Улучшение качества и натуральности синтезированной речи позволяет надеяться на повышение надежности тех систем распознавания устной речи, в которые будет заложена модель триграммного взаимодействия фонем.

Системы распознавания устной речи, в которых задействованы триграммные модели, на настоящий момент еще не построены. Имеющихся данных пока недостаточно, а обучающие тексты сформированы лишь частично, так как процесс формирования этих текстов весьма нетривиален. Так, сформированный к настоящему времени обучающий текст (лингвистический ресурс) [6-7] содержит около 1500 слов, а его транскрипция покрывает только около 5000 наиболее частых триграмм русского языка (множество всевозможных триграмм русского языка содержит  $53^3$  элементов, из них допустимых в языке — около  $10^5$ ).

Напомним, что универсальный обучающий текст — это такой буквенно-цифровой текст, который необходимо устно ввести в компьютер с целью его обучения распознаванию произвольной устной слитной речи. При решении более простых, но, тем не менее, весьма важных задач устного диалога человек-компьютер на "усеченных" языках достаточно проанализировать триграммный состав множества всевозможных фраз из этого "усеченного" языка и сформировать "усеченный" (не универсальный) обучающий текст. В частности, в задаче распознавания слитно произнесенных чисел от 0 до 1000 не требуется распознавание любых фраз и слов. Поэтому универсальный обучающий текст излишен. Так как в настоящее время русскоязычные обучающие тексты с указанными свойствами и ориентированные на распознавание чисел от 0 до 1000 отсутствуют, создание подобных текстов представляется актуальной задачей.

Целью данной работы является создание такого русскоязычного текста минимизированной длины, который обеспечивает обучение систем распознавания слитно произнесенных чисел с учетом эффекта триграммного взаимодействия фонем.

### 1. Сущность и особенности проблемы

Проблема формирования текста для обучения систем распознавания устной речи возникает всякий раз, когда появляются новые данные о звуковом строе языка или когда исследуются новые модели образования речи. В этом смысле триграммная модель звукового строя языка не является исключением. К обучающему тексту, отражающему объективно существующие эффекты триграммного взаимодействия фонем, предъявляются следующие естественные требования: 1) в состав текста должны входить осмысленные слова, словосочетания или предложения; 2) транскрипция обучающего текста должна содержать все триграммы из исходного текста, т.е. из того текста, который будет устно вводиться в компьютер с целью распознавания; 3) обучающий текст должен иметь минимальную длину; 4) текст должен формироваться в предположении, что словосочетания и предложения будут произноситься слитно, без пауз.

Первое требование осмысленности позволяет диктору произносить читаемый текст естественно (без напряжений). Второе требование обеспечивает полноту учета в обучающем тексте тех эффектов взаимовлияния фонем, которые будут иметь место при произнесении распознаваемых устных команд. Здесь следует иметь в виду, что обучающий текст в общем случае отличен от того текста, который будет наговариваться при распознавании.

Отдельного пояснения требует третье условие минимальности длины обучающего текста. Дело в том, что процесс обучения компьютера — утомительный и однообразный. Он состоит в многократном чтении одного и того же обучающего текста группой из нескольких десятков дикторов и записи речевых сигналов на жесткие носители. На этапе обучения можно было бы предложить каждому диктору произнести все числа от 0 до 1000, т.е. начитать весь исходный текст. Однако в этом случае (без минимизации длины обучающих текстов) объемы памяти, тре-

буемые для сохранения сигналов, окажутся непомерно завышенными, а сам процесс обучения будет слишком длительным. По нашим предварительным оценкам, минимизация длины обучающего текста позволит сократить объем памяти, требующийся для хранения сигналов, примерно на порядок. Наконец, четвертое требование слитности произнесения обеспечивает необходимую степень редукции фонем, в особенности на стыках слов во фразах.

Таким образом, сущность проблемы создания обучающих текстов состоит в таком формировании этих текстов, при котором перечисленные выше требования удовлетворяются одновременно. Главная особенность обучающих текстов состоит в том, что при минимальной длине их фонетическая транскрипция должна отражать все многообразие фонетических (звуковых) явлений языка (или "усеченного" языка).

## 2. Возможные подходы к решению проблемы

Основная трудность формирования обучающего текста состоит в одновременном соблюдении четырех упомянутых выше требований. Рассмотрим три возможных подхода к решению проблемы.

Первый подход может базироваться исключительно на алгоритмах дискретной оптимизации. Этот (полностью компьютерный) подход предполагает применение алгоритмов решения задачи о вычленении из (в общем случае бесконечной) символьной последовательности (транскрипции орфографического текста) таких подпоследовательностей, конкатенация которых удовлетворяют определенным ограничениям (эти ограничения фактически указаны в сформулированных выше требованиях). Насколько известно авторам, точный полиномиальный алгоритм решения указанной оптимизационной задачи на сегодняшний день еще не найден. Следует также отметить, что признаком получения точного решения задачи формирования обучающего текста минимальной длины является выполнение двух очевидных условий: 1) в транскрипции алгоритмически сформированного текста частота встречаемости каждой из допустимых триграмм равна 1, 2) цепочки слов согласованы по смыслу. Как показывает опыт,

чисто математические методы решения подобных задач очень редко приводят к успеху, так как в сформированных с помощью компьютера текстах почти всегда содержатся цепочки несогласованных по смыслу слов. Алгоритмы же согласования слов по смыслу на настоящее время еще не построены.

В качестве второго подхода к решению проблемы можно предложить комбинирование алгоритмов, упомянутых в первом подходе, с итеративной корректировкой текста (для согласования слов по смыслу) и его пополнением экспертом-фонетистом. Наконец, третий подход может быть основан на принципе первоначального и предпочтительного включения в формируемый текст тех слов или словосочетаний, фонетическая транскрипция которых содержит наиболее редкочастотные триграммы, встретившиеся в исходном тексте. Причем на первом этапе обучающий текст можно сформировать с использованием только исходного текста, не привлекая другие текстовые базы данных.

Второй и третий подходы предполагают участие эксперта-фонетиста для смыслового согласования слов в цепочках. При этом для второго подхода в качестве начального приближения могут использоваться результаты, полученные с применением первого. С другой стороны, тексты, полученные в рамках второго и третьего подходов, могут служить приближением для точных оптимизационных методов из первого подхода.

Ниже описаны результаты формирования обучающего текста, полученные в рамках третьего подхода.

### 3. Формирование текста

Источником данных для формирования искомого обучающего текста, содержащего все трехфонемные комбинации, возникающие при слитном произнесении чисел от 0 до 1000, является легко получаемая цифровая запись списка всех этих чисел. Орфографический эквивалент этого списка (список фраз) и его фонетическая транскрипция были получены с помощью программного обеспечения, реализующего ранее созданный алгоритм фонетического транскрибирования [4], которое встроено в синтезатор устной речи "Текстофон" [5].

Как известно, алфавит символов русского языка включает 33 элемента (буквы). В этот алфавит дополнительно включен символ "\_", обозначающий пробел.

Алфавит фонем включает 53 элемента. В него входит подмножество {А, О, Э, Ы, И, У} имен гласных (ударных) фонем, включающее 6 элементов, и подмножество {^, Ж, Ш, в, о, е, и, ы, у} имен редуцированных гласных фонем, состоящее из 9 элементов. Символом ^ обозначаются аллофоны гласных "а" и "о", соответствующие первой степени редукции, а символом Ш — аллофоны этих же гласных для второй степени редукции. Символ Ж служит для обозначения аллофонов гласных "э", "е" в первой степени редукции, а символ в — для обозначения этих же гласных во второй степени редукции. Символом о обозначена слаборедуцированная гласная "о" в словах "кто", "что", "то" и т.п., а символом е — слаборедуцированная гласная "е" в словах "тем", "кем", "чем" и т.п. Символами и, ы, у обозначены безударные гласные "и", "ы", "у" соответственно.

Подмножество имен согласных {Бб, Вв, Гг, Дд, Жж, Зз, Кк, Лл, Мм, Нн, Пп, Рр, Сс, Тт, Фф, Хх, Ц, ч, Ш, ш, Щ} состоит из 37 элементов. В этом подмножестве строчные символы служат для обозначения мягких согласных. Наконец, в алфавит фонем входит символ -, обозначающий паузу.

Пример. При произнесении числа 795 (семьсот девяносто пять) имеем транскрипцию "\_сЖмСОДдвЖНОСТ\$пАт\_", в которую входит 20 (18 плюс две паузы в начале и в конце фразы) фонем и 18 триграмм.

Для получения затранскрибированного исходного текста все числа от 0 до 1000 были через пробел записаны в один входной файл. Цифровой текст с помощью программного транскриптора [4] был развернут в буквенный эквивалент. Длина исходного (буквенного) текста составляет 4896 символов. Полученный (фонетический) файл с транскрипциями послужил источником для оценивания частот встречаемости триграмм. В ходе формирования текста подсчитывался коэффициент избыточности. Под коэффициентом избыточности понимается отношение общего числа триграмм (фонем) в тексте к числу различных триграмм (фонем) в этом тексте. Для точного решения задачи формирования

обучающего текста минимальной длины коэффициент триграммной избыточности у искомого обучающего текста, очевидно, будет равен 1.

В результате обработки исходного текста было установлено, что в транскрипции этого текста имеется всего 33 (32 плюс пауза) различных фонемы из 53 (52 плюс пауза), допустимых в русском языке. Длина транскрипции исходного текста — 16183 фонемы (без пауз). Коэффициент фонемной избыточности равен  $16183/32 = 505.72$ . Ниже приведена табл. 1, в которой перечислены все возможные фонемы при произнесении чисел от 0 до 1000. Справа от символа фонемы указана ее частота встречаемости в исходном тексте. Фонемы упорядочены по частоте их встречаемости в исходном тексте.

Т а б л и ц а 1

1	-	2002	8	Ж	950	15	В	490	22	Н	292	29	з	100
2	Т	1616	9	Ш	920	16	р	490	23	ч	201	30	К	80
3	с	1201	10	т	830	17	Э	480	24	Ы	201	31	М	20
4	б	1192	11	м	580	18	е	410	25	Р	110	32	Г	20
5	С	1110	12	Ц	580	19	И	390	26	и	110	33	л	1
6	О	991	13	А	551	20	п	300	27	ы	110			
7	д	950	14	Д	506	21	Ш	300	28	^	101			

Транскрипция исходного текста включает всего 16183 триграмм, из них различных — 314. Коэффициент триграммной избыточности равен  $16183/314=51.54$ . Все различные триграммы, встретившиеся в транскрипции исходного текста, сведены в табл. 2 в частотном порядке встречаемости. Как и в табл. 1, справа от текста триграммы указана ее частота встречаемости в исходном тексте.



Таблица 2

1	<i>СОТ</i>	375	36	<i>ШмС</i>	110	71	<i>ди</i>	100	106	<i>ТШ\$</i>	50
2	<i>СТ\$</i>	300	37	<i>\$т.</i>	110	72	<i>вт.</i>	100	107	<i>ТСО</i>	50
3	<i>ТрИ</i>	290	38	<i>е%Н</i>	110	73	<i>Эмд</i>	100	108	<i>Тпс</i>	50
4	<i>ЦЦ\$</i>	290	39	<i>ч%</i>	102	74	<i>эмд</i>	100	109	<i>Гд</i>	50
5	<i>Ц\$т</i>	250	40	<i>л%</i>	101	75	<i>мСС</i>	100	110	<i>ОТС</i>	50
6	<i>дос</i>	210	41	<i>В\$</i>	101	76	<i>сти</i>	100	111	<i>ТШЭ</i>	45
7	<i>е%Х</i>	210	42	<i>Де</i>	101	77	<i>эд%</i>	100	112	<i>ТпА</i>	45
8	<i>ч%Т</i>	200	43	<i>с%</i>	101	78	<i>ДДВ</i>	95	113	<i>сАД</i>	40
9	<i>всв</i>	200	44	<i>Шм</i>	101	79	<i>ТсЭ</i>	95	114	<i>ссД</i>	40
10	<i>дос</i>	200	45	<i>дд%</i>	100	80	<i>ТВО</i>	95	115	<i>Ц\$д</i>	40
11	<i>%ТЫ</i>	200	46	<i>е%т</i>	100	81	<i>ШЭс</i>	90	116	<i>Т\$д</i>	36
12	<i>мдв</i>	200	47	<i>НОС</i>	100	82	<i>Эвс</i>	90	117	<i>Т\$В</i>	34
13	<i>%сА</i>	200	48	<i>вСТ</i>	100	83	<i>дЭв</i>	90	118	<i>Т\$п</i>	34
14	<i>д%с</i>	200	49	<i>ДИН</i>	100	84	<i>рв.</i>	90	119	<i>Т\$с</i>	34
15	<i>%мС</i>	200	50	<i>СОР</i>	100	85	<i>вст</i>	90	120	<i>Т\$Ш</i>	34
16	<i>%тС</i>	200	51	<i>псд</i>	100	86	<i>рИ.</i>	90	121	<i>\$Тр</i>	34
17	<i>с%м</i>	200	52	<i>ОР\$</i>	100	87	<i>Эм.</i>	90	122	<i>Т\$Д</i>	34
18	<i>мСО</i>	200	53	<i>\$эд</i>	100	88	<i>ИН.</i>	90	123	<i>Т\$Т</i>	34
19	<i>тСО</i>	200	54	<i>рИЦ</i>	100	89	<i>Ат.</i>	90	124	<i>\$ВО</i>	32
20	<i>Осв</i>	190	55	<i>вдд</i>	100	90	<i>ВА.</i>	90	125	<i>\$сЭ</i>	32
21	<i>Эст</i>	190	56	<i>рвС</i>	100	91	<i>эм.</i>	90	126	<i>\$ДВ</i>	32
22	<i>сэм</i>	190	57	<i>ОСТ</i>	100	92	<i>пАт</i>	90	127	<i>\$дс</i>	22
23	<i>Тыр</i>	190	58	<i>ДвЭ</i>	100	93	<i>ст.</i>	90	128	<i>сТВ</i>	20
24	<i>сЭм</i>	190	59	<i>Ш\$з</i>	100	94	<i>Р\$К</i>	80	129	<i>ТпА</i>	20
25	<i>ДВА</i>	190	60	<i>СТ</i>	100	95	<i>НАЦ</i>	70	130	<i>тч%</i>	20
26	<i>Ырв</i>	190	61	<i>рИС</i>	100	96	<i>Одд</i>	65	131	<i>АТТ</i>	20
27	<i>ВОс</i>	190	62	<i>ТТр</i>	100	97	<i>ОТТ</i>	60	132	<i>с%М</i>	20
28	<i>АЦЦ</i>	170	63	<i>п%т</i>	100	98	<i>ОТп</i>	60	133	<i>ддЭ</i>	20
29	<i>свТ</i>	160	64	<i>вЭс</i>	100	99	<i>ОДД</i>	60	134	<i>\$Ш\$</i>	20
30	<i>сАТ</i>	160	65	<i>СТО</i>	100	100	<i>ОТВ</i>	60	135	<i>АДД</i>	20
31	<i>СОЦ</i>	125	66	<i>ВАЦ</i>	100	101	<i>ОТШ</i>	60	136	<i>дДВ</i>	20
32	<i>\$с%</i>	112	67	<i>%НО</i>	100	102	<i>ОТс</i>	60	137	<i>т^д</i>	20
33	<i>Тр</i>	112	68	<i>ИЦЦ</i>	100	103	<i>Ддс</i>	55	138	<i>АТ.</i>	20
34	<i>дс</i>	111	69	<i>ССО</i>	100	104	<i>ДдЭ</i>	50	139	<i>тпА</i>	20
35	<i>В\$с</i>	110	70	<i>ИСТ</i>	100	105	<i>Тч%</i>	50	140	<i>АТ</i>	20

Продолжение таблицы 2

141	\$ЦЦ	20	176	сГ	20	211	ОсЭ	11	246	\$К	10
142	\$ns	20	177	\$mB	20	212	ДВ	11	247	muC	10
143	\$mT	20	178	\$д	14	213	\$Кс	10	248	ОТч	10
144	\$СО	20	179	Т\$	14	214	ОШ\$	10	249	НН\$	10
145	МНА	20	180	\$дЭ	14	215	Онс	10	250	Эсб	10
146	сТn	20	181	Т\$ч	14	216	с%Т	10	251	КпА	10
147	Р\$Г	20	182	\$ч%	14	217	ИНН	10	252	ТОС	10
148	сТШ	20	183	muд	13	218	руН	10	253	ТЫР	10
149	АТч	20	184	ТОд	13	219	ОГ	10	254	uНА	10
150	Н\$Ц	20	185	mun	12	220	п%Т	10	255	%НА	10
151	mШЭ	20	186	ТОВ	12	221	\$КШ	10	256	\$КТ	10
152	mcЭ	20	187	muД	12	222	Тру	10	257	РН\$	10
153	сДД	20	188	ТОТ	12	223	\$Кп	10	258	До%	10
154	АТn	20	189	ТОД	12	224	лс	10	259	ГДВ	10
155	mBO	20	190	muШ	12	225	uCO	10	260	\$К.	10
156	\$m	20	191	\$пА	12	226	\$ГД	10	261	ГдЭ	10
157	\$mc	20	192	muc	12	227	КТр	10	262	ТВ\$	5
158	\$mШ	20	193	ТОШ	12	228	Ш\$	10	263	ДДо	5
159	сГ.	20	194	Т\$.	12	229	\$Гд	10	264	ОГ.	5
160	%МН	20	195	ТОн	12	230	СНА	10	265	Тп%	5
161	сГч	20	196	muB	12	231	\$Кч	10	266	ТШы	5
162	Т\$С	20	197	ТОс	12	232	CO	10	267	Тс%	5
163	АТВ	20	198	ОТр	12	233	сст	10	268	ТО	2
164	\$дд	20	199	muT	12	234	\$KB	10	269	uч%	2
165	АТШ	20	200	\$ШЭ	12	235	ЫРН	10	270	\$Шы	2
166	сДд	20	201	uTr	12	236	КВО	10	271	\$В\$	2
167	\$дД	20	202	ОДВ	11	237	uns	10	272	ТОч	2
168	\$mп	20	203	удс	11	238	Кч%	10	273	\$До	2
169	АДд	20	204	одс	11	239	дэс	10	274	uд	2
170	АТс	20	205	uДВ	11	240	ОСО	10	275	\$п%	2
171	сТс	20	206	BO	11	241	uCH	10	276	muч	2
172	сТТ	20	207	сЭ	11	242	КШЭ	10	277	дЭ	2
173	%ТН	20	208	uBO	11	243	uШ\$	10	278	д	2
174	mTr	20	209	ОВО	11	244	Кд	10	279	Од	2
175	\$mч	20	210	ucЭ	11	245	КсЭ	10	280	ОдЭ	2

281	<i>идЭ</i>	2	290	<i>Бьс</i>	1	299	<i>цДс</i>	1	308	<i>ис%</i>	1
282	<i>Оч%</i>	2	291	<i>иШы</i>	1	300	<i>АТы</i>	1	309	<i>иВ%</i>	1
283	<i>тй^</i>	2	292	<i>НО</i>	1	301	<i>свч</i>	1	310	<i>ипА</i>	1
284	<i>Ол</i>	1	293	<i>ОпА</i>	1	302	<i>лА</i>	1	311	<i>Ос%</i>	1
285	<i>НОл</i>	1	294	<i>чс</i>	1	303	<i>ШЭ</i>	1	312	<i>Д</i>	1
286	<i>ДН</i>	1	295	<i>НАТ</i>	1	304	<i>ип%</i>	1	313	<i>иШЭ</i>	1
287	<i>счс</i>	1	296	<i>ДНА</i>	1	305	<i>ТО</i>	1	314	<i>тн</i>	1
288	<i>Оп%</i>	1	297	<i>Тьс</i>	1	306	<i>ОШЭ</i>	1			
289	<i>ОШы</i>	1	298	<i>ОДс</i>	1	307	<i>ОВ%</i>	1			

Сущность многошаговой процедуры формирования обучающего текста состояла в следующем. На первом этапе в искомый текст были включены лишь те слова и словосочетания, в транскрипции которых имелись триграммы с частотой встречаемости в затранскрибированном исходном тексте равной 1. Затем в обучающий текст последовательно включались слова и словосочетания, в транскрипции которых имелись наиболее редкочастотные триграммы. На каждом шаге пополнения обучающего текста анализировался его триграммный состав. Очередной проход процесса пополнения останавливался, если на текущем шаге в транскрипции формируемого текста содержались все триграммы из транскрипции исходного текста. Для полученного варианта обучающего текста подсчитывались частоты встречаемости триграмм, и очередной проход процесса формирования с помощью описанной процедуры повторялся снова. При этом полученный вариант обучающего текста рассматривался как исходный текст. Окончательный вариант обучающего текста был получен за три прохода (после выполнения третьего прохода текст, полученный на втором проходе, не изменился).

Сформированный обучающий текст включает 109 слов и фраз. В транскрипции этого текста всего 1225 фонем (32 — различных) и 1225 триграмм (314 — различных). Коэффициенты фонемной и триграммной избыточности для буквенного и фонетического вариантов текста равны соответственно 38.28 и 3.90. В транскрипции сформированного текста 177 из 314 триграмм

встречаются только один раз. Ниже приведен обучающий текст (в виде последовательности чисел, записанных через пробел) и его транскрипция.

*Обучающий текст.* 0 5 6 10 11 13 22 29 33 34 35 36 37 38 41  
42 43 44 45 46 47 48 49 51 52 53 54 55 56 57 58 59 60 72 73 74 75  
76 77 78 79 80 100 101 102 104 105 106 108 110 112 115 116 117 118  
119 131 140 150 160 171 200 201 202 203 205 206 207 208 210 212  
214 215 216 217 218 240 250 260 290 302 303 304 305 306 307 308  
315 316 318 340 350 360 399 412 504 512 591 600 615 616 617 618  
740 750 760 803 901 1000.

*Транскрипция обучающего текста.*

Но нАт ШЭст дЭст динн\$ЦЦ\$т ТриНАЦЦ\$т  
ДВАЦЦ\$дДВА ДВАЦЦ\$ддЭвт ТриЦЦ\$тТри  
ТриЦЦ\$тч%Тыр ТриЦЦ\$тнАт ТриЦЦ\$тШЭст  
ТриЦЦ\$тсЭм ТриЦЦ\$тВОсвм СОР\$Кдин СОР\$ГДВА  
СОР\$КТри СОР\$Кч%Тыр СОР\$КнАт СОР\$КШЭст  
СОР\$КсЭм СОР\$КВОсвм СОР\$ГдЭвт пвдд%сАТдин  
пвдд%сАДДВА пвдд%сАТТри пвдд%сАТч%Тыр  
пвдд%сАТнАт пвдд%сАТШЭст пвдд%сАТсЭм  
пвдд%сАТВОсвм пвдд%сАДдЭвт Ш\$зд%сАТ  
сЭмдвсвДДВА сЭмдвсвТТри сЭмдвсвТч%Тыр  
сЭмдвсвТнАт сЭмдвсвТШЭст сЭмдвсвТсЭм  
сЭмдвсвТВОсвм сЭмдвсвДдЭвт ВОсвмдвсвТ СТО  
СТОдин СТОДВА СТОч%Тыр СТОнАт СТОШЭст  
СТОВОсвм СТОдЭвт СТОДе%НАЦЦ\$т  
СТОн%ТНАЦЦ\$т СТОШыСНАЦЦ\$т СТОс%МНАЦЦ\$т  
СТОВс%МНАЦЦ\$т СТОдво%ТНАЦЦ\$т  
СТОТриЦЦ\$тдин СТОСОР\$К СТОпвдд%сАТ  
СТОШ\$зд%сАТ СТОсЭмдвсвТдин ДеЭсти  
ДеЭстидин ДеЭстидВА ДеЭстиТри  
ДеЭстинАт ДеЭстиШЭст ДеЭстисЭм ДеЭстиВОсвм  
ДеЭстидЭвт ДеЭстиДе%НАЦЦ\$т  
ДеЭстич%Тырн\$ЦЦ\$т ДеЭстин%ТНАЦЦ\$т  
ДеЭстиШыСНАЦЦ\$т ДеЭстис%МНАЦЦ\$т  
ДеЭстиВс%МНАЦЦ\$т ДеЭстиСОР\$К  
ДеЭстинвдд%сАТ ДеЭстиШ\$зд%сАТ  
ДеЭстидво%НОСТ\$ ТриСТ\$ДВА ТриСТ\$Три

-ТРИСТ\$%ТЫрв\_ -ТРИСТ\$пАт\_ -ТРИСТ\$ШЭст\_  
 -ТРИСТ\$сЭм\_ -ТРИСТ\$ВОсвм\_ -ТРИСТ\$п%ТНАЦЦ\$т\_  
 -ТРИСТ\$ШыСНАЦЦ\$т\_ -ТРИСТ\$В\$с%МНАЦЦ\$т\_  
 -ТРИСТ\$СОР\$К\_ -ТРИСТ\$пвдд%сАТ\_ -ТРИСТ\$Ш\$зд%сАТ\_  
 -ТРИСТ\$двс%НОСТ\$дЭвст\_ -%ТЫрвСТ\$Дс%НАЦЦ\$т\_  
 -л%тСОТ%ТЫрв\_ -л%тСОДДс%НАЦЦ\$т\_  
 -л%тСОДдвс%НОСТ\$ДИН\_ -ШыССОТ\_  
 -ШыССОТп%ТНАЦЦ\$т\_ -ШыССОТШыСНАЦЦ\$т\_  
 -ШыССОТс%МНАЦЦ\$т\_ -ШыССОТВ\$с%МНАЦЦ\$т\_  
 -с%мСОТСОР\$К\_ -с%мСОТпвдд%сАТ\_ -с%мСОТШ\$зд%сАТ\_  
 -В\$с%мСОТТри\_ -двс%тСОТДИН\_ -ДНАТЫсвч\_

В табл. 3 приведен полный список различных триграмм, входящих в обучающий текст, в порядке частоты их встречаемости. Частота встречаемости, как и ранее, указана справа от триграммы.

Т а б л и ц а 3

1	ЦЦ\$	32	19	сЭм	15	37	лв	9	55	с%Н	7
2	Ц\$т	30	20	рИС	14	38	^ДИ	9	56	%МН	7
3	Три	28	21	ИСТ	14	39	%ТЫ	9	57	рв_	7
4	Эст	27	22	пвд	13	40	ШыС	9	58	ри_	7
5	\$т_	23	23	ОР\$	13	41	_СО	9	59	с%М	7
6	АЦЦ	23	24	дд%	13	42	свТ	8	60	вм_	7
7	НАЦ	21	25	вдд	13	43	ШЭс	8	61	ИЦЦ	7
8	_Тр	21	26	СОР	13	44	вт_	8	62	МНА	7
9	_СТ	19	27	СОТ	11	45	пАт	8	63	ВА_	7
10	_Дс	19	28	Р\$К	11	46	ИН_	8	64	риЦ	7
11	СТО	19	29	мдв	10	47	Осв	8	65	Эм_	6
12	ДсЭ	19	30	двс	10	48	_сЭ	8	66	зд%	5
13	сдс	19	31	свс	10	49	ТЫр	8	67	\$зд	5
14	сти	19	32	Эмд	9	50	Ат_	8	68	ССО	5
15	%сА	18	33	%ЖТ	9	51	Ырв	8	69	%ТН	5
16	д%с	18	34	ДИН	9	52	ВОс	8	70	двс	5
17	СТ\$	18	35	АТ_	9	53	свм	8	71	Ш\$з	5
18	сАТ	16	36	ДВА	9	54	ст_	8	72	_Шы	5

Продолжение таблицы 3

73	мСС	5	108	Тч%	3	143	НН\$	1	178	ОВО	1
74	В\$с	5	109	миШ	3	144	\$КТ	1	179	АТВ	1
75	дэе	5	110	ТОс	2	145	НОа	1	180	\$Кч	1
76	Ээв	5	111	Н\$Ц	2	146	ОТч	1	181	ОТр	1
77	вс%	5	112	Ддэ	2	147	НАТ	1	182	\$Кп	1
78	\$с%	5	113	Тсэ	2	148	Ш\$	1	183	Ос%	1
79	ввт	5	114	сАД	2	149	ИНН	1	184	АТс	1
80	ТНА	5	115	Р\$Г	2	150	Д	1	185	\$КВ	1
81	мСО	4	116	ТОД	2	151	Опс	1	186	Т\$С	1
82	СНА	4	117	ДДВ	2	152	дэ	1	187	НО	1
83	п%Т	4	118	ТШэ	2	153	Шэ	1	188	АТч	1
84	Де%	4	119	Т\$д	2	154	ВО	1	189	лА	1
85	тСО	4	120	ВАЦ	2	155	\$КШ	1	190	АТп	1
86	%мС	4	121	ТОд	2	156	ОДс	1	191	Т\$	1
87	%тС	4	122	\$ЦЦ	2	157	АДд	1	192	ОДД	1
88	%НА	4	123	тцД	2	158	ОТВ	1	193	\$дд	1
89	\$К.	4	124	Т\$В	2	159	ДН	1	194	ОТТ	1
90	мСН	4	125	ТВО	2	160	АТТ	1	195	ОТС	1
91	с%м	4	126	ДВ	2	161	Сд	1	196	АТШ	1
92	дэс	3	127	ОТШ	2	162	д	1	197	ОДд	1
93	ТТр	3	128	СОД	2	163	ОпА	1	198	АДД	1
94	п%т	3	129	ТОВ	2	164	сСТ	1	199	тВО	1
95	свт	3	130	Ц\$д	2	165	ОТс	1	200	\$тп	1
96	Т\$Ш	3	131	СсД	2	166	ОВ\$	1	201	тч%	1
97	ТОШ	3	132	Т\$Д	2	167	ОШ\$	1	202	ОТ.	1
98	Эсв	3	133	тис	2	168	\$Гд	1	203	ОСО	1
99	Тд	3	134	миВ	2	169	Осэ	1	204	тпА	1
100	%НО	3	135	ОТп	2	170	счс	1	205	свч	1
101	ТОп	3	136	тцд	2	171	дс	1	206	ОТ	1
102	Т\$п	3	137	ТпА	2	172	АТ	1	207	миТ	1
103	с%	3	138	ОШэ	1	173	Одс	1	208	тТр	1
104	тип	3	139	Оп%	1	174	Одэ	1	209	ДДе	1
105	НОС	3	140	ОШы	1	175	Ол.	1	210	\$тс	1
106	л%	3	141	ч%	1	176	смд	1	211	ОДВ	1
107	ОСТ	3	142	В\$	1	177	ти.	1	212	Т\$Т	1

213	\$ГД	1	239	ТВ\$	1	265	υВО	1	291	\$ч%	1
214	тсЭ	1	240	сТ.	1	266	Кч%	1	292	чс.	1
215	Оч%	1	241	\$пА	1	267	υШЭ	1	293	\$СО	1
216	сТШ	1	242	\$В\$	1	268	\$ДВ	1	294	ТОС	1
217	\$п%	1	243	Тп%	1	269	ДНА	1	295	КпА	1
218	дДВ	1	244	\$тч	1	270	РН\$	1	296	υДс	1
219	тυ̂	1	245	Ддо	1	271	рсС	1	297	ТШы	1
220	сТВ	1	246	ТО.	1	272	упс	1	298	ТШ\$	1
221	\$К	1	247	\$тТ	1	273	υдс	1	299	руН	1
222	АТЫ	1	248	сДд	1	274	Т\$ч	1	300	υШ\$	1
223	тШЭ	1	249	ддЭ	1	275	КВО	1	301	υНА	1
224	υСО	1	250	сТс	1	276	υШы	1	302	υДВ	1
225	Три	1	251	\$Ш\$	1	277	\$ВО	1	303	Т%с	1
226	КТр	1	252	\$дД	1	278	\$Дс	1	304	υ̂д	1
227	сТч	1	253	ТОч	1	279	\$сЭ	1	305	υч%	1
228	сТп	1	254	ГдЭ	1	280	\$Тр	1	306	уп%	1
229	\$пс	1	255	\$дЭ	1	281	Тпс	1	307	ТСО	1
230	\$Шы	1	256	сТ	1	282	ТЫс	1	308	υдЭ	1
231	\$дс	1	257	υсЭ	1	283	с%Т	1	309	υс%	1
232	сТТ	1	258	ТЫР	1	284	υпА	1	310	ТО	1
233	\$Кс	1	259	\$т̂	1	285	\$д	1	311	υТр	1
234	\$тШ	1	260	ГДВ	1	286	КШЭ	1	312	υВ\$	1
235	К̂д	1	261	\$тВ	1	287	ТОТ	1	313	ЫРН	1
236	сДД	1	262	КсЭ	1	288	Ысс	1	314	Тс%	1
237	тυС	1	263	т̂д	1	289	\$ШЭ	1			
238	тυч	1	264	с%т	1	290	Т\$.	1			

Любая распознаваемая команда (число) может быть представлена в виде последовательности триграмм. Например, число 795 в триграммной транскрипции выглядит следующим образом: с%, с%м, %мс, мсо, сод, одд, ддс, дсс, сс%, с%н, %но, нос, ост, ст\$, т\$п, \$пА, пАт, Ат. При решении задачи обучения именно те 314 триграмм, которые перечислены в табл. 2 и 3, следует вычленять в качестве элементарных эталонов (сигналов). Например, для фонемы /А/ в этих таблицах имеется

7 следующих триграмм: *ВА*-, *ВАЦ*, *НАТ*, *НАЦ*, *пАт*, *сАД*, *сАТ* (с частотами встречаемости 7, 2, 1, 21, 8, 2 и 16 соответственно). Каждая из этих триграмм соответствует различным звуковым оттенкам одной и той же фонемы /А/.

### З а к л ю ч е н и е

В работе найдено приближенное решение задачи формирования русскоязычного текста минимальной длины, в транскрипции которого содержатся всевозможные триграммы, встречающиеся при слитном произнесении чисел от 0 до 1000. Сформированный текст может использоваться при построении более совершенных систем распознавания устной речи. Он позволит исследователям точнее учесть на этапе обучения компьютера и подавить на этапе распознавания устной речи объективно существующие в языке эффекты фонетической изменчивости.

Хотя сформированный текст не имеет минимальную длину, число произносимых фраз, необходимых для обучения компьютера, минимизировано и сокращено по сравнению с исходным текстом почти в 10 раз (с 1001 до 109), при этом число фонем, составляющих текст, уменьшено более чем в 13 раз (с 16183 до 1226), а коэффициент триграммной избыточности снижен с 51.54 до 3.90.

Ввиду математической нетривиальности проблемы, поиск ее точного (с коэффициентом триграммной избыточности равным 1) алгоритмического решения представляется делом будущего. Здесь следует оговориться, что вопрос существования точного решения в настоящий момент остается открытым. Можно лишь гипотетически утверждать, что это решение существует, опираясь на известную избыточность русского языка. Точно также гипотетически можно утверждать, что если точное решение существует, то, скорее всего, оно не единственно в силу той же избыточности языка.

Вместе с этим, резервы поиска лучшего эвристического приближения к точному решению, на наш взгляд, еще не исчерпаны.



Процесс поиска новых приближений видится в такой последовательной замене слов и словосочетаний в уже сформированном тексте, которая шаг за шагом будет уменьшать коэффициент избыточности. При этом, очевидно, потребуется обработка значительных по объему текстовых баз данных.

Наконец заметим, что поиск как точного решения для рассмотренной в работе задачи, так и лучших приближений к решению, важен по той причине, что сформированный текст минимальной длины может служить частью (отрывком) универсального обучающего текста, ориентированного на распознавание произвольной слитной речи.

### Л и т е р а т у р а

1. CUBALA F., et al. Continuous Speech Recognition Results of the BYBLOS System on the DARPA 1000-Word Resource Management Database // IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-88), 1988, Ser.7.8

2. Величко В.М., Саломатина Н.В., Чукалина Н.В., Юдина Л.С. Формирование универсального обучающего текста для систем распознавания речи // Анализ временных рядов и символьных последовательностей. — Новосибирск, 1991.- Вып.141: Вычислительные системы. — С. 93–101.

3. Зиновьева Н.В. и др. Автоматический транскриптор // Всесоюз. конф.: "Автоматическое распознавание слуховых образов" (АРСО-17), Ижевск, 1992 г.: Тез. докл. — Ижевск, 1992. — С. 72–76.

4. Кельманов А.В., Саломатина Н.В., Хайретдинова А.Г., Хамидуллин С.А. Правила и алгоритм преобразования орфографической записи на русском языке в фонетическую транскрипцию // Прикладные системы искусственного интеллекта. — Новосибирск, 1995. — Вып. 150: — Вычислительные системы. — С. 32–92.

5. Кельманов А.В., Кутненко О.А., Саломатина Н.В., Хайретдинова А.Г., Хамидуллин С.А. Макроволновой синтез речи по тексту // Тез. докл. 2 Всерос. с участием стран СНГ конф. "Распознавание образов и анализ изображений: новые информационные технологии" Ульяновск, 1995.- С. 121–123. (Ульяновск. гос. техн. ун-т).

6. Кельманов А.В., Саломатина Н.В. Варьирование лингвистического ресурса при обучении систем распознавания и синтеза слитной речи // Труды IV междунар. конф. "Расознавание образов и анализ изображений: новые информационные технологии" (РОАИ-4-98). — Новосибирск, 1998. — Ч.1. — С. 116-119.

7. Kel'manov A.V., Salomatina N.V. Linguistic Resource for Training in Systems Coherent Speech Recognition and Synthesis // Pattern Recognition and Image Analysis, 1998. — Vol.8, №3. — P. 415-416.

Поступила в редакцию  
2 октября 2000 года