
МЕТОДИЧЕСКИЕ И АЛГОРИТМИЧЕСКИЕ ОСНОВЫ ОБРАБОТКИ И АНАЛИЗА РЕЧЕВЫХ И ЗВУКОВЫХ СИГНАЛОВ

УДК 811.161.1

Н. С. СМЕРНОВА, М. В. ХИТРОВ

ФОНЕТИЧЕСКИ ПРЕДСТАВИТЕЛЬНЫЙ ТЕКСТ ДЛЯ ФУНДАМЕНТАЛЬНЫХ И ПРИКЛАДНЫХ ИССЛЕДОВАНИЙ РУССКОЙ РЕЧИ

Приведен фонетически представительный текст, разработанный с применением новейших достижений в области лингвистических технологий. Полнота покрытия текстом фонетических единиц русской речи позволяет использовать его при формировании речевых корпусов для разработки и оценки экспертных и автоматических речевых систем различного назначения.

Ключевые слова: фонетически представительный текст, фонетически сбалансированный текст, статистические характеристики русской речи, частотность и дистрибуция фонетических единиц.

Если обратиться к прикладным областям речевых исследований, то можно заметить, что сегодня использование небольших фонетически представительных текстов при создании автоматических систем синтеза и распознавания речи уже не столь актуально, и приоритет отдается машинным методам статистического моделирования с использованием обучающих массивов текстовых и речевых данных очень больших объемов. Такие массивы данных называют также базами данных или корпусами (текстовыми или речевыми). Часто под корпусом понимают преимущественно те массивы данных, которые переведены в электронную форму и специальным образом обработаны, структурированы и аннотированы для целей разработки речевых приложений [1]. В настоящей статье опорным текстовым корпусом будет называться большой по объему массив текстов различного жанра, использованный нами для получения опорного статистического распределения фонетических единиц русской речи.

Однако наряду с разработкой речевых систем не менее актуальной остается задача выработки объективных критериев оценки их качества, и в этом случае тестовым материалом для оценки и сравнения систем автоматического синтеза и распознавания речи могут стать небольшие фонетически представительные тексты (ФПТ), позволяющие оценить полноту покрытия системой фонетических единиц целевого языка и выявить возможные недостатки ее работы. Кроме того, на таких текстах удобно проводить быструю подстройку системы под нового диктора.

В общем случае под фонетически представительным (репрезентативным) понимается такой текстовый материал, в котором частотное распределение фонетических единиц (фонем, аллофонов, слогов) соответствует общеязыковому распределению, получаемому из статистического анализа опорного текстового корпуса. В задачах, предполагающих исследование

региональной вариативности речевых характеристик, в качестве дополнительного критерия фонетической представительности текста должно рассматриваться наличие фонетических позиций и контекстов, способствующих проявлению региональной речевой специфики говорящего.

Фонетическая представительность, подобно фонетической сбалансированности, естественным образом предполагает присутствие в тексте всех фонем целевого языка в их основной дистрибуции. Фонетически сбалансированные и фонетически представительные тексты традиционно используются в качестве материала для изучения фонетических характеристик звучащей речи. Преимущество использования фонетически представительных текстов состоит, прежде всего, в их компактности наряду с информационной насыщенностью. С одной стороны, такие тексты обычно невелики по объему, а с другой — отражают фонетическое многообразие языковой системы не хуже произвольно взятых текстовых массивов значительного объема. Это достигается путем кропотливой работы по конструированию текста — наполнением его словами, содержащими требуемые фонетические единицы, а также сокращением его объема путем удаления элементов с низкой информативностью. В результате получается удобный для прочтения материал (обычно не более 600 слов), позволяющий исследовать характер реализации и варьирования в речи носителей определенного языка значимых фонетических характеристик и сформировать полноценный речевой портрет говорящего.

В русистике известно несколько фонетически представительных текстов, составленных на основе списков наиболее частотных слогов, приведенных в работах В. М. Елкиной и Л. С. Юдиной [2, 3]. Один из них — „Был тихий серый вечер“ [4] — лег в основу материала для Фонетического фонда русского языка.

Слог традиционно считается минимальной произносительной единицей, и потому оценка встречаемости слогов может лечь в основу формирования текстового материала. Однако на частоту встречаемости и состав выделяемых слогов оказывает влияние ряд факторов, в частности, характер опорного текстового материала (на основе которого получены показатели частотности слогов), используемая система транскрипции текста и степень ее подробности, а также принятая стратегия слогаделения. В отношении частотного распределения слогов, приведенного в работах [2, 3] и использованного впоследствии при составлении текста „Был тихий серый вечер“, следует отметить, что оно было получено на текстах радиотехнической тематики и с применением довольно спорной теории деления на открытые слоги, предложенной Л. В. Бондарко [5]. Кроме того, в классификации [2, 3] не различаются предупредные и заударные слоги (а для некоторых гласных фонем — также ударные и безударные варианты), что приводит к серьезным упрощениям в оценках частотности и сочетаемости аллофонов русских фонем.

Исследовав методологию и инструментарию, ранее использовавшиеся при составлении фонетически представительных текстов, авторы разработали несколько иной подход, предполагающий, в частности, применение более подробной транскрипции текстового материала (с учетом предупредной/заударной позиции гласного), преимущественно стилистически нейтрального текстового материала для получения опорной статистики, а также увеличение объема опорного текстового материала. Кроме того, поскольку существующие теории слогаделения допускают вариативность межслоговых границ для сочетаний ГС и СС (Г — гласный, С — согласный) и, как следствие, по-разному представляют состав и количество слогов русского языка (ср., например, принципы, предложенные М. В. Ломоносовым, Р. И. Аванесовым, Л. В. Щербой, Л. В. Бондарко [6]), было решено в качестве базовых единиц при составлении текста использовать последовательности СГ, поскольку при любом подходе они относятся к одному слогу. При этом в последовательностях типа ГГ в качестве самостоятельных элементов выделялись гласные, а на конце слова допускались закрытые слоги типа СГС. В качестве дополнительных критериев учитывалась встречаемость двухфонемных и трехфо-

немных сочетаний (так называемых дифонов и трифонов). Кроме наиболее частотных фонетических единиц в текст были введены звуковые последовательности и позиции, диагностически важные для выявления региональной речевой специфики.

Данный принцип построения текста был предпочтен „слоговому“ как более адекватный и экономичный для получения фонетической представительности. Если следовать слоговому принципу построения текста, то для полноценного выявления региональных особенностей говорящих потребовалось бы дополнительно включить в текст целый ряд низкочастотных звуко сочетаний и позиций, что в комбинации с обеспечением высокой слоговой представительности неизбежно привело бы к увеличению объема текстового материала. Так, в составе частотных слогов отсутствует целый ряд элементов, чрезвычайно важных для исследования вариативности русской речи — в частности, конечный мягкий <-вь> и другие мягкие губные. Например, первый в списке по теории Аванесова [7] слог с конечным „ф“ [к а_{з/уд} ф'] имеет ранг 989, слог [б О_{уд} ф'] — 1352, а первый по частотности слог с конечным мягким „п“ (слово „степь“) — лишь 3993. Подобная ситуация наблюдается и в отношении ряда других важных в диагностическом плане звуковых элементов. Кроме того, известно, что на качественные характеристики гласных в русском языке преимущественное влияние оказывает левый контекст, и неслучайно при различных подходах к слогаделению именно последовательность СГ неизменно относится к одному слогу.

Материал для получения опорной статистики был скомпонован из текстов классической и современной литературы, а также современной публицистики (отекстованные интервью, репортажи, дискуссии). Он включает в себя более 460 тыс. словоформ, более 1 млн слогов (по сравнению с более 100 тыс. в работах [2, 3]), более 2,5 млн фонемоупотреблений. Была оценена встречаемость фонем (монофонов), звуко сочетаний (двух- и трехфонемных) и слогов (по трем различным сценариям слогаделения). Кроме того, для ряда фонем был составлен список фонетических позиций и контекстов, потенциально значимых для выявления региональной вариативности русской речи (например, мягкие губные в конечной позиции, определенные сочетания согласных). Опорный текстовый корпус и его статистические характеристики приведены в работах [8, 9].

На основе статистик, полученных на опорном материале, с учетом фонемных позиций и комбинаций, способствующих выявлению региональной произносительной специфики, был составлен новый фонетически представительный текст. Он состоит из 533 слов, 1197 слогов (по числу гласных). Всего текст насчитывает 2902 фонемоупотребления. Текст включает в себя как описательную, так и богатую диалоговую часть (все коммуникативные типы); в нем представлены все фонемы русского языка во всех допустимых аллофонах (включая межсловные озвонченные аллофоны непарных глухих русских фонем / х /, / ч /, / ц / и / щ / — соответственно Х_{озв}, Ч_{озв}, Ц_{озв} и Щ_{озв}). В тексте присутствует более 99 % сочетаний типа СГ (из них 98 % — наиболее частотные 258), 92 % возможных в русском языке двухфонемных сочетаний (из них 62 % — 250 наиболее частотных), значительно расширен (по сравнению с существующими текстами) набор сочетаний „согласный + ударный гласный“, возможных в русской речи. Для трех рассмотренных вариантов слогаделения (по [6, 7, 10]) доля покрытия типов слогов не ниже 70 % (что на 4—5 % выше, чем в тексте „Был ... вечер“). Отметим также, что в текст целенаправленно были введены слова со звуко сочетаниями, важными для исследования региональной и индивидуальной произносительной вариативности, в том числе иностранного происхождения, что привело к повышению доли низкочастотных слогов.

В табл. 1 приведены данные о встречаемости в нашем тексте аллофонов русских фонем в сопоставлении с их статистическим распределением в опорном текстовом корпусе. При обозначении аллофонов русских фонем используются следующие конкретизаторы: п/уд — предупредительный, з/уд — заударный, б/уд — безударный, озв — озвонченный, ' — мягкий.

В тексте присутствует 56 типов аллофонов русских фонем (как уже упоминалось выше, отсутствуют лишь редкие безударные аллофоны фонемы / е /).

Таблица 1

Аллофон	Ранг в опорном корпусе	Ранг и встречаемость в тексте	Аллофон	Ранг в опорном корпусе	Ранг и встречаемость в тексте
a ₁ -й п/уд	1	1 (137)	г'	23	29 (39)
a ₃ /уд	2	2 (132)	ф	36	30 (38)
и ₃ /уд	3	3 (131)	Ы _{уд}	40	31 (38)
й	7	4 (124)	г	32	32 (36)
и _п /уд	4	5 (123)	в'	37	33 (34)
О _{уд}	9	6 (119)	ж	38	34 (34)
т	6	8 (115)	м'	39	35 (34)
А _{уд}	5	7 (114)	ч	28	36 (33)
н	8	9 (107)	У _п /уд	31	37 (32)
р	12	10 (101)	б	29	38 (31)
к	14	11 (95)	ш	30	39 (30)
с	10	12 (88)	Ы _п /уд	43	40 (29)
в	11	13 (88)	ц	42	41 (27)
Е _{уд}	13	14 (86)	д'	35	42 (23)
м	17	15 (77)	х	41	43 (20)
д	21	16 (68)	к'	44	44 (17)
п	16	17 (63)	п'	45	45 (16)
л'	18	18 (62)	щ	46	46 (12)
a ₂ -й п/уд	20	19 (60)	б'	47	47 (11)
Ы ₃ /уд	24	20 (54)	з'	49	48 (9)
н'	15	21 (54)	ф'	51	49 (7)
с'	25	22 (53)	О _б /уд	48	50 (5)
л	19	23 (51)	г'	50	51 (5)
У _{уд}	34	24 (50)	х'	53	52 (4)
р'	26	25 (47)	х _{озв}	52	53 (3)
з	27	26 (47)	ч _{озв}	55	54 (1)
У ₃ /уд	33	27 (44)	ц _{озв}	56	55 (1)
И _{уд}	22	28 (42)	щ _{озв}	58	56 (1)

Как видно из табл. 1, распределение частотности аллофонов в разработанном тестовом материале достаточно близко к распределению в опорном корпусе.

Совпадает состав 14 наиболее частотных аллофонов и 13 наиболее редких (разница в ранге — не более 3). В частотах остальных 39 аллофонов наблюдаются более существенные различия в рангах. В среднем разница в рангах составляет 2,89; максимальная разница в ранге наблюдается для ударного „У“ — 10.

Основные статистические характеристики созданного фонетически представительного текста приведены в табл. 2.

Таблица 2

Типы единиц	Типы единиц в ФПТ относительно опорного корпуса, %	Общее покрытие единиц опорного корпуса в ФПТ, %
Фонемы	96,6	99,9
Последовательности СГ	72,3	99,9
Дифоны	46,2	91,6
Трифоны	6,6	42,3
Слоги (по Аванесову)	6,2	74,1
Слоги (по Щербе)	5,9	72,6
„Открытые“ слог	5,6	74,1

Из табл. 2 видно, что наш текст обеспечивает практически стопроцентное покрытие фонемного состава опорного текстового корпуса. Столь же высокий процент покрытия обеспечивают и присутствующие в сформированном тексте сочетания СГ (отсутствующие 118 типов таких последовательностей составляют менее 1 % опорного корпуса). Из числа возможных типов слога в разработанном тексте присутствует лишь 5—6 %, однако они покрывают 73—75 % всех слогов, встречающихся в опорном корпусе. В тексте встречается чуть менее половины (46 %) возможных в русском языке дифонов, однако при этом общая степень покрытия реализаций дифонов опорного корпуса достигает 92 %. Состав трифонов опорного корпуса наиболее обширен и насчитывает более 35 тыс. типов. Созданный текст включает более 2 тыс. типов трифонов (7 %), что покрывает 43 % всех реализаций трифонов опорного корпуса.

Всего в тексте присутствует 1197 гласных и 1705 согласных, консонантный коэффициент 1,42, что несколько выше, чем в опорной статистике (1,35 в опорном корпусе; 1,38—1,39 — по литературным источникам [11]). К более высокому значению консонантного коэффициента привело введение в текст слов с диагностическими консонантными последовательностями и позициями, а также дополнение текста словами с низкочастотными звуками и звукосочетаниями (в основном консонантными). Таким образом, фонемный состав текста был сбалансирован для получения более надежных результатов исследований.

Приведем фонетически представительный текст.

Дом, в котором я живу, расположен на окраине маленького городка, у самой подошвы горы. Здесь мягкий климат и редко идут дожди. Ночью небосвод бывает так густо усеян звездами, что кажется, будто все миллиарды их из нашей галактики разбросаны вверх над моей головой. Летним утром, как только я открываю окно, моя большая комната наполняется запахом цветов. Ветки черешен смотрят мне в окна, и легкий теплый ветер усыплет мой письменный стол белыми лепестками.

Я слушаю щебет птиц. Вот с искрометным задором пропел зяблик. Где-то дятел устраивает дупло. А это черные дрозды — поют не хуже соловьев. Прямо передо мной внизу — пестрый узор из крыш городских домов, а вдалеке, на краю горизонта, тянется серебряная цепь снеговых вершин... Весело жить в такой земле! Отрадное чувство разливается в жилах: вокруг величественные горы, воздух чистый и свежий, солнце яркое, небо синее — чего еще желать?.. Бьют настенные часы над камином: пять, шесть, семь, восемь, девять... Нужно торопиться в бюро. Минуты две-три ищущу в шкафу электрическую схему, привезенную французским коллегой. Наконец заглядываю в портфель, нахожу ее внутри и вкладываю в книгу. После этого плотно закрываю жалюзи, однако сквозь щелки все равно пробивается солнечный свет. Выхожу на крыльцо и запираю ключом дверь.

Спустившись ниже к центру города, иду бульваром. Часть дороги проходит по пешеходному мосту через реку. Гибкие стальные тросы держат невысокий мост. Они привязаны к специальным тяжелым якорям, врытым в землю. Останавливаюсь в начале мостика у ограды, чтобы полюбоваться рельефными склонами горных хребтов, всматриваюсь в речную рябь. Под мостом с шумом плещутся мелкие рыбешки, возмущая водную гладь. Откуда-то доносится музыка: ноктюрн Шопена — позывные местной радиостанции.

Вдруг позади себя я слышу: „Сережа, неужели это ты? Вот так встреча!“ Я узнаю этот низкий голос. Обобщаю — так и есть: Андрей Сафонов! Очень радостно видеть его вновь. С Андреем мы знакомы с конца восьмидесятых — служили вместе в армии. Его, энергичного и общительного, всегда на помощь готового прийти, любили все. Меня покорила его честность и недюжинная сила. Мы были дружны, но потом, мало-помалу, связь наша оборвалась.

— Здравствуй, Андрюша! Как ты тут оказался?

— Командировка в архив: предлагают снять сюжет про судоверфь. Вчера приехали — сегодня уезжаем. Вернее, улетаем — к четверем в аэропорт.

— Так скоро? И куда?

— Следующим пунктом Уфа. Прямой рейс. А ты, значит, теперь здесь живешь? Давно?

— Два года будет в феврале. Обменял свою городскую квартиру на бревенчатую избу.

— Серьезно? Не жалеешь?

В глазах моего приятеля мелькнул веселый огонек.

— Нисколько. Отдыхаю от километровых пробок, сутолоки и пыли.

— По-прежнему плывешь против течения? Счастливый ты человек, Сергей.

— Ладно, расскажи лучше о себе. Мы тысячу лет не виделись. Как жизнь? Как семья?

— Все у нас хорошо, все здоровы. Мы с женой работаем, дочь гимназию заканчивает.

— Ну а Федор как? Учится?

- Он в этом году поступил в медицинский.
 — Какой молодец! Поздравляю вас!
 — Спасибо.
 — Кстати, ты позавтракал?
 — Немного кофе выпил в гостинице. С удовольствием бы съел что-нибудь.
 — Недалеко отсюда есть кафе. Мы привыкли там есть. Пойдем, провожу тебя. Вполне приличный сервис, разнообразное меню. Одно из их „фирменных“ блюд — рыба по-бенгальски. Рекомендую: вкус необыкновенный — для настоящих гурманов.
 — Звучит слишком изысканно. Попроще ничего нет?
 — Как насчет яичницы с грибами?
 — В самый раз. А ты торопишься? Может быть, составишь мне компанию?
 За разговорами и воспоминаниями незаметно пролетел завтрак. Приближалось время сказать „до свидания“. Мы расстались в твердом намерении больше не терять друг друга из виду.

СПИСОК ЛИТЕРАТУРЫ

1. *Кривнова О. Ф.* Фонетическое обеспечение для построения речевого корпуса // Акустика речи. Медицинская и биологическая акустика. Сб. тр. XIII сессии Российского акустического общества. Т. 3. М.: ГЕОС, 2003. С. 118—122.
2. *Елкина В. М., Юдина Л. С.* Статистика слогов русской речи // Вычислительные системы. Новосибирск, 1964. Вып. 10. С. 58—78.
3. *Елкина В. М., Юдина Л. С.* Статистика открытых слогов русской речи // Там же. Вып. 14. С. 55—91.
4. *Степанова С. Б.* Фонетические свойства русской речи: реализация и транскрипция: Дис. ...канд. филол. наук. Л., 1988.
5. *Бондарко Л. В.* Фонетика современного русского языка. Л., 1998. С. 196—211
6. *Бондарко Л. В.* Структура слога и характеристики фонем // Вопросы языкознания. 1967. № 1. С. 34—46.
7. *Аванесов Р. И.* О слогоразделе и строении слога в русском языке // Там же. 1954. № 6. С. 88.
8. *Смирнова Н. С., Чистиков П. Г.* Программа анализа фонетических статистик в текстах на русском языке и ее использование для решения прикладных задач в области речевых технологий // Матер. XXVII Междунар. конф. „Диалог“. М., 2011. С. 632—644.
9. *Smirnova N., Chistikov P.* Statistics of Russian Monophones and Diphones // Proc. of Speccom-2011. Kazan, Russia, 2011. P. 218—223.
10. *Щерба Л.В.* Теория русского письма. Л., 1983. С. 29—33.
11. Фонетика спонтанной речи / Под ред. *Н. Д. Светозаровой.* Л., 1988. С. 210.

Сведения об авторах

- Наталья Сергеевна Смирнова** — канд. филол. наук; ООО „ЦРТ“, Санкт-Петербург; руководитель группы лингвистов; E-mail: nsmirnova@speechpro.com
- Михаил Васильевич Хитров** — канд. техн. наук; ООО „ЦРТ“, Санкт-Петербург; генеральный директор; Санкт-Петербургский национальный исследовательский университет информационных технологий, кафедра речевых информационных систем; зав. кафедрой; E-mail: khitrov@speechpro.com

Рекомендована кафедрой
речевых информационных систем

Поступила в редакцию
22.10.12 г.