

Система синтеза белорусской речи по тексту

Ю.С. Гецевич,
аспірант

Б.М. Лобанов,
доктор технических наук



Рассмотрены особенности фонетических систем белорусского и русского языков и структура белорусскоязычного синтезатора речи по тексту. Описаны структура и состав белорусского электронного словаря, а также программная реализация синтезатора белорусской речи.

Abstract

Peculiarities of the phonetic systems of the Belarusian and Russian languages and a structure of the Belarusian text-to-speech synthesizer are considered. A structure and composition of the Belarusian electronic dictionary, as well as the programme realization of a Belarusian speech synthesizer of are described.

Основные методы верификации речевых данных

Для некоторых славянских языков, таких как русский, чешский, польский, украинский, уже существуют практически используемые или экспериментальные образцы синтезаторов речи по тексту (СРТ) [1]. В литературе нет, однако, никаких сведений о создании синтезаторов речи по тексту для белорусского языка. Работа является продолжением проводимых ранее исследований, базирующихся на разработке русскоязычного синтезатора речи [2]. Результатом этой работы стало создание двуязычной системы синтеза речи по тексту. Синтез речи по тексту на двух славянских языках — белорусском и русском — предполагает создание фонетико-акустической базы данных, построенной на единых принципах, отражающих внутри- и межъязыковую специфику фонетических систем и позиционно-комбинаторных явлений, порождающих аллофонию речи.

1. Особенности фонетических систем белорусского и русского языков

Фонетические системы языков, относящихся к группе славянских, имеют между собой значительное сходство, однако каждый из них обладает также специфическими

особенностями, иногда значительными. Исследуемые фонетические системы белорусского и русского языков относительно близки. В белорусском языке насчитывается 41 фонема, из них 6 гласных и 35 согласных, а в русском — 42 фонемы: гласных — 6 и согласных — 36. В табл. 1 представлена обобщённая информация о фонемном составе двух языков и об их различии по способу и месту образования. В каждой ячейке таблицы представлены имена фонем, характеризующихся определённым способом и местом образования, для белорусского и русского языков порядке «сверху — вниз». Для обозначения фонем используются традиционные для каждого языка буквы алфавита.

Таблица 1

Фонетические системы белорусского и русского языков

Способ образования Место образования		Согласные											Гласные	Передняя	Высокая	Осужденная
		Глухие			Звонкие			Сонорные								
		Взрывные	Аффрикаты	Щелевые	Взрывные	Аффрикаты	Щелевые	Држащие	Носовые	Боковые	Плавные	Гласные				
Заднеязычные	Мягкие	к'		х'	~	~	ж'				й	у	0	1	1	
	Твёрдые	к		х	~	~	ж					о	0	0	1	
Среднеязычные	Мягкие		ч'	ш'				р'				а	0	0	0	
	Твёрдые		ч	ш		дж	жс	р				э	1	0	0	
Переднеязычные	Мягкие	~	ц'	с'	~	дз'	з'		н'	л'		ы	0	1	0	
	Твёрдые	т	ц	с	д		з		н	л		и	1	1	0	
Губные	Мягкие	п'		ф'	б'		в'		м'							
	Твёрдые	п		ф	б		в		м		ў					

русского языка порядке «сверху — вниз». Для обозначения фонем используются традиционные для каждого языка буквы алфавита.

В табл. 1 затемнены ячейки, фонетическое качество звуков которых имеет практически полное сходство для каждого из языков. Как видно из таблицы, количество таких ячеек в процентном отношении ко всем используемым ячейкам довольно значительно — 71%. Отличительные особенности фонетических систем белорусского и русского языков заключаются в следующем.

В белорусском языке отсутствуют следующие фонемы:

- мягкие согласные **Т, Д, Ш, Ч, Р;**
- мягкая и твёрдая **Г***

В белорусском языке есть ряд специфических фонем, отсутствующих в русском:

- плавная **Ў;**
- мягкая **Ц** и твёрдая **Ч;**
- мягкая аффриката **Дз** и твёрдая **Дж;**
- мягкая и твёрдая щелевая **Гх.**

* произношение взрывного звука Г возможно в некоторых заимствованных словах (гандаль, ганак, гузік) и в коренных буквосочетаниях ЗГ, ДЗГ, ДЖГ (мазіг, бразгаць, джаць), в остальных случаях произносится звонкий фрикативный звук «Гх».

2. Структура белорусскоязычного синтезатора речи по тексту

Синтез устной речи по тексту осуществляется на основе лексико-грамматического анализа входного текста путём моделирования процессов речеобразования с учётом правил произношения звуков и интонирования, свойственных белорусскому языку. Орфографический текст документа (книги, статьи, веб-страницы и т.п.) поступает на вход синтезатора и далее подвергается последовательной обработке рядом специализированных процессоров в соответствии с общей структурой синтезатора речи по тексту, представленной на *рис. 1*. Синтезатор включает четыре основных модуля: текстовый процессор, просодический процессор, фонетический процессор и акустический процессор. Каждый из этих модулей поддерживается наборами соответствующих БД и правил.

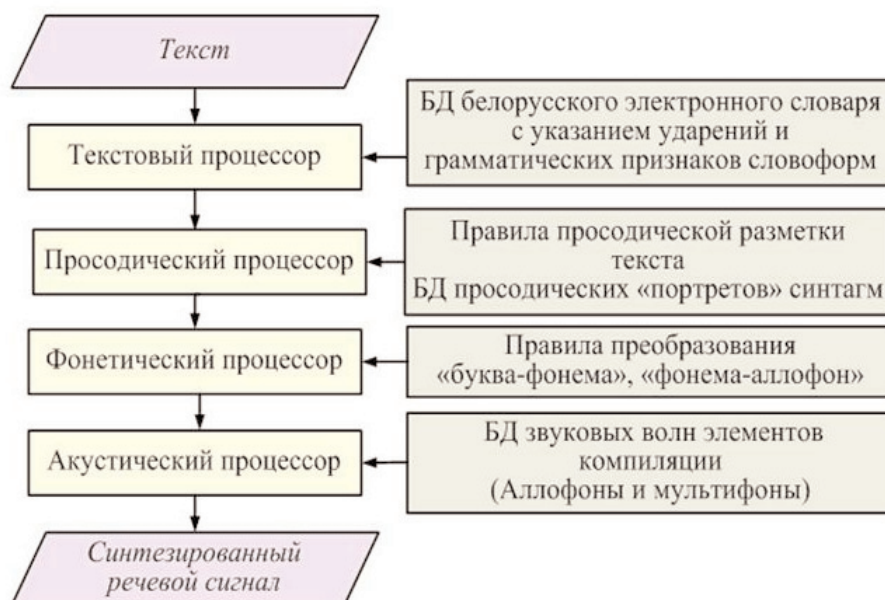


Рис. 1. Общая структура синтезатора речи по тексту

Входной орфографический текст подвергается ряду последовательных обработок в каждом из процессоров. Текстовый процессор обрабатывает входной орфографический текст в следующей последовательности: очистка текста, преобразование символов (аббревиатур, сокращений, чисел и др.), расстановка словесных ударений и грамматических признаков словоформ. Преобразованный текст поступает на входы просодического, а затем фонетического процессора. В результате работы просодического процессора, текст делится на синтагмы, акцентные единицы (АЕ), который далее размечается на элементы акцентных единиц (ЭАЕ): интонационное предъядро, ядро и заядро.

И наконец, последняя функция просодического процессора — установка в соответствии с БД просодических «портретов» синтагм значений амплитуды (А), длительности фонем (Т) и частоты основного тона (F0) для каждого ЭАЕ. Задача фонетического процессора заключается в преобразовании орфографического текста в

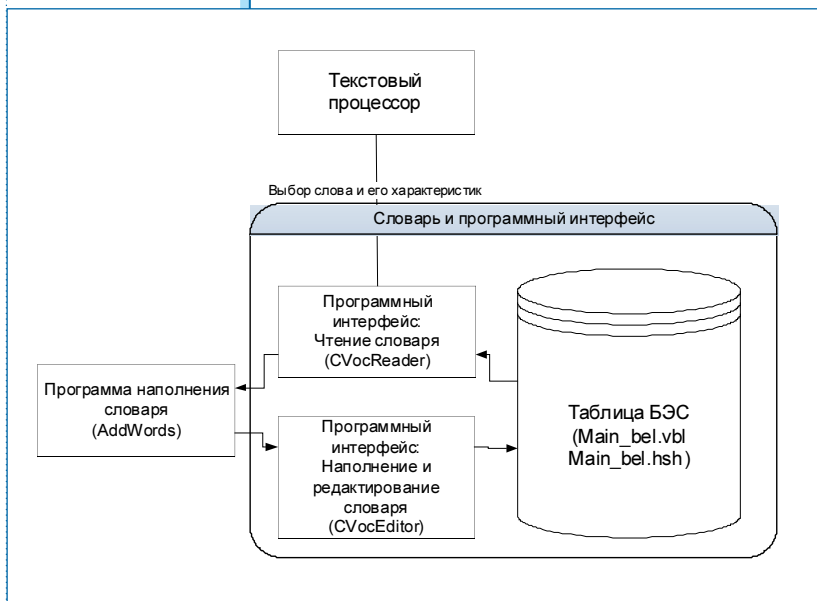


Рис.2. Структурная схема белорусского электронного словаря

Таблица 2

Сравнительные характеристики словарей белорусского и русского языков

Часть речи	Количество словоформ (белорусский)	Количество словоформ (русский)
Существительное	591 036	465 230
Глагол	624 501	989 064
Прилагательное	874 997	626 842
Наречие	6 541	1 353
Числительное	571	410
Предлог	121	87
Междометие	96	180
Частица	58	86
Союз	46	68
Всего	2 097 967	2 083 320

3. Белорусский электронный словарь

Полная словарная база содержит 2 097 967 словоформ белорусской речи. В табл. 2 приведены сравнительные характеристики состава используемых словарей для белорусского и русского языков. Словарь белорусского языка построен на базе словаря М.В. Бирылы [3], русского языка — А.А. Зализняка [4].

фонемный, а также в генерации позиционных и комбинаторных аллофонов. Акустический процессор на основе информации о том, какие аллофоны необходимо синтезировать, а также какие просодические характеристики должны быть приписаны каждому аллофону, генерирует речевой сигнал путём компиляции отрезков естественных звуковых волн соответствующих аллофонов и мультифонов (диаллофоны и аллослоги).

Текстовый, просодический, фонетический и акустический процессоры синтезатора речи в своей основе языконезависимые, а языковая специфика (в нашем случае специфика белорусского языка) задаётся соответствующим набором баз данных и знаний (правил). На рис. 1 языконезависимые блоки представлены прямоугольниками в левой части, а языкозависимые — в правой части. Три из четырёх блоков синтезатора (см. рис. 1) — просодический, фонетический и акустический процессоры — в наименьшей степени требуют коррекции языковых баз данных и знаний при переходе от синтеза русской речи к белорусской. Коренные изменения связаны, в основном, с созданием электронного белорусского словаря.

Архитектура белорусского электронного словаря (БЭС) представляет собой базу данных с конкретными функциями для её чтения и наполнения (рис. 2).

Таблица базы данных белорусского электронного словаря наполняется записями с тремя полями: орфографическое слово S1, позиции ударений в слове S2, теги слова S3 (табл. 3). На диске компьютера таблица БЭС сохраняется в файлах main_bel.vbl (данные записей) и main_bel.hsh (данные проиндексированных записей).

Для чтения БЭС используется программный интерфейс класса CVocReader. Его архитектура приводится в табл. 4.

Таблица 3

Архитектура базы данных БЭС	
Название поля	Тип
Орфографическое слово	Строка (<=255 символов)
Позиция ударения	Вектор целых чисел
Тег	Строка (<=255 символов)

Таблица 4

Архитектура программного класса CVocReader для чтения словаря

Функция класса CVocReader	Комментарий
Открытие/Закрытие	Словарь переводится в режим чтения/общего доступа
Получить количество слов в словаре	Реальное количество слов
Поиск слова	Выбирается запись (3 поля) из таблицы БД
Поиск следующего слова	Выбираются остальные записи. (В словаре могут быть слова омографы, поэтому эта функция покажет все хранимые варианты слов)
Получить последнюю ошибку	Во время чтения словаря эта функция постоянно должна проверяться для следующей корректной работы всего процесса синтеза речи

Для редактирования БЭС используется программный интерфейс класса CVocEditor. Его архитектура приводится в таблице 5.

Таблица 5

Архитектура программного класса CVocEditor для редактирования словаря

Функция класса CVocEditor	Комментарий
Открытие/Закрытие	Словарь переводится в режим редактирования/общего доступа
Добавить слово	Добавляется запись с ЛГК (3 поля) в таблицу БД



Удалить слово	Удаляется запись с ЛГК (3 поля) из таблицы БД
Поиск слова	По орфографической записи слова ищется соответствующая запись с ЛГК в БЭС
Поиск слова по маске	По условной записи (через регулярные записи) слова ищется соответствующая запись с ЛГК в БЭС
Получить последнюю ошибку	Во время чтения словаря эта функция постоянно должна проверяться для следующей корректной работы всего процесса синтеза речи
Обновить/Перестроить индекс БЭС	Синхронизация индексов для поиска слов в БЭС. (Должна исполняться после любых редактирований таблицы)

В табл. 6 представлен фрагмент используемой словарной базы.

Таблица 6

Список словарной базы с указанием ударения (`) и грамматических категорий (Тэг). (фрагмент, всего — 1 260 794)

Слова_Тэг (существительные)	Слова_Тэг (глаголы)
...	...
зака`зчык_ NNAMO	падвыша`ць_ VIC
зака`зчыка_ NNAMG	падвыша`ю_ VIIR1
зака`зчыку_ NNAMD	падвыша`еш_ VIIR2
зака`зчыка_ NNAMA	падвыша`е_ VIIR3
зака`зчыкам_ NNAMI	падвыша`ем_ VIIR1P
зака`зчыку_ NNAMR	падвыша`еце_ VIIR2P
зака`зчыкі_ NNAMPO	падвыша`юць_ VIIR3P
зака`зчыкаў_ NNAMPG	падвыша`й_ VIM2
зака`зчыкам_ NNAMPD	падвыша`йце_ VIM2P
зака`зчыкаў_ NNAMPA	падвыша`ў_ VIIPM
зака`зчыкамі_ NNAMPI	падвыша`ла_ VIIPF
зака`зчыках_ NNAMPR	падвыша`ла_ VIIPN

Отдельная группа алгоритмов разработана для преобразования чисел в порядковые и количественные числительные.

В табл. 7–9 приведены правила преобразования чисел в числительные с соответствующими базами данных, а в таблице 10 — соответствующие текстовые примеры.

Таблица 7

Список соответствий — TIs «окончание =>склонение порядкового числительного» (фрагмент, всего — 480).

Число	Окончание	Окончание
...
12	-га	дванаццатага
12	-е	дванаццатае
12	-ы	дванаццаты
12	-м	дванаццатым
12	-мі	дванаццатымі
12	-му	дванаццатаму
12	-ае	дванаццатае
12	-ай	дванаццатай
12	-аю	дванаццатаю
12	-ую	дванаццатую
12	-х	дванаццатых
12	-ыя	дванаццатыя
12	-ым	дванаццатым
...

Таблица 8

Список TIs — «число, склонение =>количественное числительное» (фрагмент, всего 388).

Лік	Склонение	Числительное
...
3	Т	Трыма
30	Т	Трыццацю
300	Т	Трымастамі
4	Т	Чатырма
40	Т	Сарака
400	Т	Чатырмастамі
5	Т	Пяццю
50	Т	Пяццюдзсяцяцю
500	Т	Пяццюстамі
6	Т	Шасцю
60	Т	Шасцюдзсяцяцю
600	Т	Шасцюстамі
8	Т	Васьмю
80	Т	Васьмюдзсяцяцю
800	Т	Васьмюстамі
9	Т	Дзевяццю
90	Т	Дзевяноста
...

Таблица 9

Список Ttl: «число=>название триады»

Число	Название триады
1	Тысяча
2...4	Тысячы
5..20, 30, 40, ... 90	Тысяч
100, 200, ... 900	Тысяч

4. Программная реализация синтезатора белорусской речи

Модульная структура программной реализации синтезатора белорусской речи представлена на рис. 3, где модули, управляющие последовательностью действий других модулей,



называются контроллерами, в то время как модули, реализующие алгоритмы обработки текста или речевого сигнала, называются процессорами.

Главный контроллер системы управляет последовательностью преобразований входных данных, получая на каждом этапе промежуточный результат и передавая его на обработку на следующий этап. Контроллер нормализации текста удаляет из текста символы, не нужные для синтеза речи, убирает случайное дублирование знаков препинания, заменяет похожие символы на один из них, очищает входной текст от недопустимых символов, используя для этого списки букв языка, знаков препинания, а также правила замены символов.

Таблица 10

Примеры преобразования чисел в числительные	
Входной текст	Выходной текст
<i>Количественные числительные</i>	
Настаўнік атрымаў заробак 784921 рубель. Кіраўніцтва зацвердзіла аб'ёмы асноўных відаў прадукцыі для 5 галінаў гаспадаркі.	Настаўнік атрымаў заробак семсот восемдзесят чатыры тысячы дзевяцьсот дваццаць адзін рубель. Кіраўніцтва зацвердзіла аб'ёмы асноўных відаў прадукцыі для пяці галінаў гаспадаркі.
<i>Порядковые числительные</i>	
Завод выпусціў 234000-ы аўтамабіль. У нас няма 123-га байца. У 2010-ым годзе адбудзецца алімпіяда па...	Завод выпусціў двухсоттрыццацічатырохтысячны аўтамабіль. У нас няма сто дваццаць трэцяга байца. У дзве тысячы дзесятым годзе адбудзецца алімпіяда па ...
198000-ая скрынка з цукеркамі выйшла з вытворчага цэха.	Стодзевяноставасьмітысячная скрынка з цукеркамі выйшла з вытворчага цэха.

«Очищенный» таким образом текст подаётся на вход лингвистического контроллера, выходом которого является просодически размеченный текст. Преобразованный текст поступает на вход фонетического процессора, который осуществляет преобразование «буква-фонема», «фонема — аллофон», затем — на вход просодического процессора, который устанавливает текущие значения амплитуды, частоты основного тона и длительности каждого аллофона. Просодически размеченный аллофонный текст поступает затем на вход акустического процессора, который генерирует речевой сигнал с использованием БД звуковых волн аллофонов и мультифонов. Результат преобразований поступает в контроллер формата выходных данных, который осуществляет преобразование в нужный формат: wav или mp3.

На рис. 4 представлен программный интерфейс отладочной версии синтезатора белорусской речи.

В нижней части представлен орфографический текст, подаваемый на вход синтезатора. В верхней части первый столбец отображает порядковый номер просодических синтагм этого текста. Во втором столбце указан интонационный тип каждой синтагмы (Q — разновидности вопросительной интонации, P — повествовательной). В третьем столбце указано количество акцентных единиц в каждой из синтагм. В четвёртом столбце приведён фонемный текст каждой синтагмы, в которой знаком «Ъ» отражен факт объединения служебных



Рис. 3. Программная реализация синтезатора белорусской речи

и знаменательных слов в одно фонетическое слово, знаки «+» и «=» указывают, соответственно, положение в словах полного и частичного ударения, а знак «/» — положение границ акцентных единиц синтагмы. В последнем, пятом, столбце приведён аллофонный текст синтагм.

Заключение

Специальное подразделение ЮНЕСКО, занимающееся сохранением языков мира как живым представителем культурного наследия планеты, отнесло белорусский язык к вымирающим — «потенциально угрожаемым» языкам [6].

Если современная тенденция кардинально не изменится, в ближайшие десятилетия мы можем потерять еще один живой язык Земли — славянский белорусский язык. В свете этого факта значимость создания белорусскоязычного синтезатора речи трудно переоценить. В нём программным образом аккумулированы и сохранены для будущих поколений лексика белорусского языка — арфография, акцентуация, словоизменение —

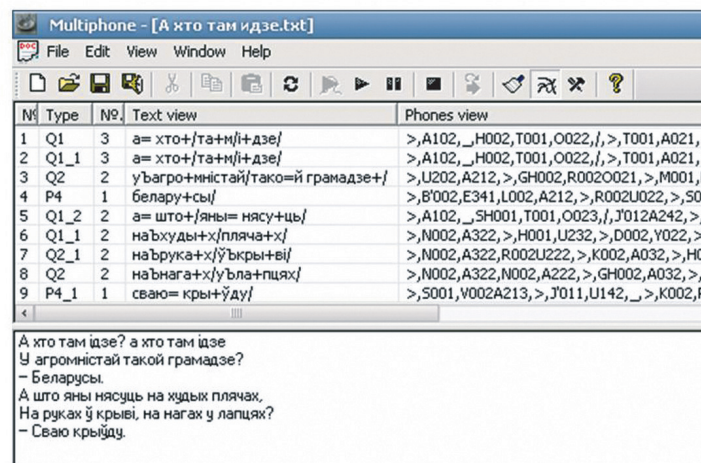


Рис. 4. Программный интерфейс синтезатора белорусской речи



фонетика и правила чтения, интонация и звуки современной белорусской речи. В то же время современным пользователям компьютеров и Интернета этот программный продукт сможет помочь в овладении и совершенствовании знания и навыков белорусской речи, а также использовать его в различных прикладных системах.

В заключение авторы выражают глубокую благодарность заведующему лабораторией интеллектуальных систем БГУ профессору И.В. Совпелю и его сотрудникам за передачу базовой версии белорусского электронного словаря и помощь в его адаптации для решения задач синтеза белорусской речи.

Работа выполнена при поддержке гранта № Ф10Р-006 БРФФИ.

Литература

1. <http://www.speech.cs.cmu.edu/comp.speech/>
2. Лобанов Б.М. Компьютерный синтез и клонирование речи // Б.М. Лобанов, Л.И. Цирульник / Минск: Белорусская Наука, 2008.
3. Зализняк А.А. Грамматический словарь русского языка: Словоизменение. Ок 100 000 слов. 2-е изд., стереотип. М.: Рус. Яз., 1980.
4. Слоўнік беларускай мовы: Арфаграфія. Арфаэпія. Акцэнтацыя. Словазмяненне / Ін-т мовазнаўства імя Я. Коласа АН БССР; Пад рэд. М.В. Бірылы. Мн.: БелСЭ, 1987..
5. <http://ru.wikipedia.org/wiki> — Белорусский язык.

Гецевич Ю.С.,

окончил факультет прикладной математики и информатики Белорусского государственного университета, факультет математики и информатики университета в Манхейме (Германия). Аспирант, младший научный сотрудник лаборатории распознавания и синтеза речи Объединённого института проблем информатики Национальной академии наук Беларуси.

Область научных интересов — методы синтеза белорусской и русской речи по тексту, человеко — машинные системы речевого общения, речевые компьютерные технологии.

E-mail: mix1122@gmail.com