

9. Methods and models of Методы и алгоритмы распознавания слов казахского языка: монография // Астана: ЕНУ им. Л.Н.Гумилева, 2010. 132с.

10. A.Sharipbayev, G.Bekmanova, B.Yergesh, A.Buribayeva, and M.K.Karabalayeva. Intellectual morphological analyzer based on semantic networks. Processings of the OSTIS-2012, 2012, pp.397-400.

УДК 004.5:811.512.145

## **РЕЧЕВОЙ ЧЕЛОВЕКО-МАШИННЫЙ ИНТЕРФЕЙС НА ТАТАРСКОМ ЯЗЫКЕ**

*А.Ф. Хусаинов, Институт прикладной семиотики Академии наук Республики Татарстан Казанский (Приволжский) федеральный университет, Казань, Россия  
khusainov.aidar@gmail.com*

### **Аннотация**

В работе приводятся результаты исследований по созданию речевого человеко-машинного интерфейса на татарском языке. Он включает в себя два основных элемента: систему автоматического распознавания речи и систему синтеза речи. Планируется использовать возможности речевого ввода и вывода информации на татарском языке при создании множества диалоговых систем: машинного переводчика, интеллектуального помощника и т.д.

**Ключевые слова:** распознавание речи, синтез речи, речевой интерфейс, татарский язык

## **SPEECH HUMAN-MACHINE INTERFACE FOR THE TATAR LANGUAGE**

*A.F. Khusainov, Institute of Applied Semiotics of the Tatarstan Academy of Sciences Kazan (Volga region) federal university, Kazan, Russia, khusainov.aidar@gmail.com*

In this paper, we describe our recent work of creation speech human-machine interface for the Tatar language. Our work consists of two main elements: speech recognition system and speech synthesizer. These systems will be used in mobile and desktop applications, for instance, machine translation system, smart assistant.

**Key words:** speech recognition, speech synthesis, speech interface, the Tatar language

### **1 Введение**

Увеличение числа информационных систем и условий их работы диктует развитие новых средств взаимодействия с ними. Одним из актуальных способов взаимодействия является использование произнесённых команд и сообщений на естественном языке. Речевой интерфейс получает всё большее распространение благодаря нескольким факторам: естественности использования речи для человека, удобству использования в определенных условиях, а также развитию технологий анализа, синтеза речи и понимания текста.

Данная работа описывает последние результаты, полученные при создании систем автоматического распознавания и синтеза речи в контексте использования татарского языка. Для решения обеих задач используется корпусный подход: система распознавания обучается на базе создаваемого многодикторного корпуса слитной речи, синтезатор речи – на основе студийных записей профессиональных дикторов.

### **2 Система распознавания слитной татарской речи**

Система распознавания речи состоит из 4 основных элементов:

Акустические модели: модели произношения акустических единиц языка (фонем, дифононов и т.д.);

Модели фонетических транскрипций слов (лексический уровень): словарь используемых слов, фонетические транскрипции слов;

Уровень языковой модели: описываются правила употребления слов языка;

Декодер – производит анализ входного речевого сигнала на основе моделей.

### 2.1. Акустические модели

Акустические модели создаются для условий относительно качественных записей: 16 бит в секунду, 16 кГц. Их можно будет использовать для распознавания речи, например, в офисных помещениях, перед компьютером, для анализа выступлений в не очень шумных помещениях. В дальнейшем планируется создавать отдельные корпуса для речи, передаваемой через телефонный канал связи, а также эфирной речи (телевидение и радио).

Крупнейшие разработчики систем распознавания речи используют для обучения речевые корпуса общей продолжительностью в тысячи часов речи. Однако экспериментально установлено, что результаты, устойчивые к смене пола, возраста и особенностей произношения диктора, можно достичь, начиная с 50 часов записей.

Общие характеристики речевого корпуса для распознавания речи, созданные в рамках исследования, приведены в Табл. 1. Все записи речевого корпуса имеют метаописания: сохраняется информация о дикторах (пол, возраст, родной язык, наличие диалекта), условиях записи (звуковое оборудование, шумовые условия).

**Таблица 1. Характеристики татарского речевого корпуса**

Параметр	Значение
Количество дикторов корпуса читаемой речи	341
Общая продолжительность корпуса читаемой речи	44:59:51
Средняя продолжительность записи	7:55
Продолжительность спонтанной речи	5:19:33
Количество дикторов корпуса спонтанной речи	168

### 2.2. Модель фонетических транскрипций слов

На основе созданных акустических моделей можно построить систему распознавания фонем татарского языка. Для перехода с уровня отдельных фонем на уровень слов необходимо создание алгоритма построения фонетических транскрипций слов. Данный алгоритм работает на основе правил графем-фонемных преобразований.

Алфавит татарского языка состоит из 39 символов: к буквам русского алфавита в нём добавлены татарские буквы Ә-ә, Ө-ө, Ү-ү, Ж-ж, Ң-ң, Һ-һ. Вопрос о составе фонемного алфавита татарского языка не имеет однозначного ответа, поэтому при его формировании учитывались как имеющиеся на данный момент результаты фонетических исследований татарского языка, так и особенности работы программных средств распознавания речи, заключающиеся в необходимости выделения основных звуков языка, которые влияют на смысл произнесенного, группируя при этом схожие по звучанию звуки. В результате проведенного исследования был сформирован алфавит из 57 фонем татарского языка. На основе определенного алфавита фонем были сформулированы акустические закономерности в татарском языке. В конечном итоге для случая татарского языка было выделено 37 правил фонетической транскрипции [1].

### 2.3. Языковая модель татарского языка

Задача создания языковых моделей возникает при решении множества задач, от проверки орфографии и до систем машинного перевода. Во всех случаях языковая модель призвана описывать существующие в языке закономерности и на их основе уметь оценивать вероятности произнесения конкретных последовательностей слов.

Татарский язык относится к группе агглютинативных языков и имеет богатую морфологию. При построении стандартных статистических языковых моделей для таких языков возникает проблема с большим числом словоформ, которые необходимо включать в словарь. Большое количество различных аффиксальных цепочек, которые могут следовать за основой слова, делает невозможным построение словаря адекватных размеров с небольшим уровнем OOV (out of vocabulary) слов. Решением этих проблем является уменьшение базовой моделируемой единицы до элемента, меньшего чем слово. В качестве базовых подходов в текущем исследовании были выбраны следующие: слова, морфемы, основы плюс аффиксальная цепочки, статистически выделенные морфы, слоги, буквы.

Исходной информацией для обучения языковой модели выступил текстовый корпус татарского языка [2]. Полученный для работы фрагмент после процедуры фильтрации и разделения на обучающую и тестовую части имеет следующие характеристики, Табл. 2.

**Таблица 2. Характеристики текстового корпуса**

Параметр	Значение
Количество файлов	217 294
Количество слов	69 810 033
Количество слогов	186 014 478 (2,66 / слово)
Количество морфем	110 280 448 (1,58 / слово)
Количество морфов	93 458 542 (1,34 / слово)
Количество основ и афф. цепочек	97 461 218 (1,4 / слово)
Количество букв	434 636 548 (6,23 / слово)
Размер	901 МБ

С учётом отсутствия до настоящего момента публикаций на тему построения и сравнения различных видов статистических языковых моделей для татарского языка, схема эксперимента была составлена таким образом, чтобы собрать максимально полную оценку влияния факторов на качество итоговой языковой модели. Так, были построены и проанализированы отдельные статистические модели для всех комбинаций из следующих категорий:

1. Тип элемента – 6 типов: слово, слог, морфема, морф, основа и аффиксальная цепочка, буква;

1. Размерность n-грамм: биграммы, триграммы, 4-граммы (5-граммы для модели на основе букв);

2. Алгоритм сглаживания модели – 5 типов: абсолютное сглаживание, Good-Turing, Kneser-Ney, Witten-Bell, модифицированный алгоритм Kneser-Ney.

Качество построенной модели оценивалось по таким показателям, как логарифм вероятности для тестового подкорпуса, perplexity (степень уверенности модели при анализе тестовых данных), OOV (количество элементов тестовой выборки, не вошедших в словарь) и размеру модели (по числу используемых n-грамм).

По результатам построения моделей был сделан вывод о том, что с точки зрения алгоритма сглаживания наилучшие результаты показали основной и модифицированный алгоритмы Kneser-Ney. Среди 95 построенных моделей наилучшее качество показала пословная модель, далее следуют модели на основе морфем и основ с аффиксальной цепочкой, морфов, слогов и букв, Табл. 3.

**Таблица 3. Сравнение языковых моделей**

Базовый элемент	Log вероятности, тыс.
Слово (4-грамм)	-12 209,0
Основа+цепочка (4-грамм)	-12 386,7
Морфема (4-грамм)	-12 638,7
Морф (4-грамм)	-12 772,4
Слог (4-грамм)	-14 282
Буква (5-грамм)	-20 741,5

В заключительном эксперименте была построена биграммная модель на основе классов слов для словаря в 20 тысяч элементов. Для выделения классов слов использовался алгоритм Брауна [3]. Построенная модель отличается нулевым значением внесловарных слов, небольшим размером, однако уступает стандартным пословным моделям в качестве описания тестового подкорпуса. Интерес представляет результат разбиения словаря из 20 тысяч слов на классы: автоматически выделенные классы объединяют слова со схожими значениями. Например, в отдельные классы выделены названия населённых пунктов, чисел, годов, фамилий, названий стран, профессий и т.д.

Для использования при распознавании была использована пословная 3-граммная модель со словарём в 100 тысяч наиболее частотных слов татарского языка.

#### **2.4. Распознавание слитной татарской речи**

В качестве декодера был использован инструмент Julius [4]. Дикторонезависимая система распознавания слитной татарской речи была построена на основе пословных моделей. Система реализована в двух вариантах: в виде консольного приложения, предоставляющего также и служебную информацию, и в виде оконного приложения, демонстрирующего распознанный текст. Для удобства пользователей был реализован алгоритм автоматического определения границ фраз.

Разработанная система распознавания речи будет адаптирована для использования в конкретных приложениях (например, словарях, машинном переводчике, интеллектуальном помощнике), в течение 2016 года станет доступна для использования на сайте [5].

### **3 Система автоматического синтеза татарской речи**

Задача синтеза речи состоит в формировании аудиосигнала на основе фразы, представленной в текстовом виде. Большая часть подходов к синтезу речи основывается на конкатенативном подходе (дифонный синтез [6], Unit selection [7]). Исходной информацией в данных подходах служат выделенные из речевых записей акустические единицы. Начиная с 2002 года, набирает популярность параметрический подход, в котором базовыми элементами являются статистические модели звуков языка.

При разработке синтезатора татарской речи используется параметрический подход на основе скрытых Марковских моделей (HMM-based speech synthesis, HTS [8]). Для обучения скрытых Марковских моделей необходимо наличие аннотированного речевого корпуса.

Особенностью создания корпуса для синтеза речи являются требования к качеству используемой аппаратуры и условиям записи. Для создания корпуса татарского языка был задействован профессиональный диктор, записи были сделаны в звукозаписывающей студии, в звуконепроницаемом помещении с использованием профессионального оборудования.

В записанных аудиофайлах экспертами были вручную размечены все интонационные группы, отмечены заимствованные и акцентные слова, после чего для всего прочитанного текста была построена фонетическая транскрипция.

Далее, была реализована модель разметки корпуса, основные элементы которой можно представить следующим образом:

1. Уровень фонем: текущая фонема, две предшествующие, две последующие фонемы.

2. Уровень слогов: тип слога (V, VC, CV, CVC, VCC, CVCC); позиция фонемы в слоге; количество фонем в предыдущем, текущем, последующем слоге; номер текущего слога в слове; гласная в текущем слоге.

3. Уровень слов: часть речи, количество слогов для предыдущего, текущего, следующего слова; количество предшествующих и последующих слов во фразе.

4. Уровень фразы: количество слов/слогов в предыдущей, текущей, последующей фразе.

#### 4. Заключение

Разработанные системы автоматического распознавания слитной речи и её синтеза позволяют начать работы по внедрению речевого человеко-машинного интерфейса на татарском языке.

Дальнейшее развитие речевых технологий предусматривает совместное использование результатов исследований в области семантического анализа текста на татарском языке, что позволит создавать интеллектуальные системы. Планируется разработка мобильных приложений, предоставляющих возможности машинного перевода, работы со словарями, помощи слабовидящим, диктовки и т.д.

#### Список литературы

1. Хусаинов, А.Ф. Система автоматического распознавания фонем татарского языка // Компьютерная обработка тюркских языков: труды первой международной конференции. (Астана, 3-4 октября, 2013). Астана: ЕНУ им. Л. Н. Гумилева, 2013. С. 211–217.

2. Dz. Suleymanov, O.A. Nevzorova, and B. Khakimov. National Corpus of the Tatar Language “Tugan Tel”: Structure and Features of Grammatical Annotation. Proc. International Conference Georgian Language and modern Technology. (Tbilisi, 2013). P. 107-108.

3. P. F. Brown, V. J. Della Pietra, P. V. deSouza, J. C. Lai and R. L. Mercer. Class-Based n-gram Models of Natural Language, Computational Linguistics 18(4), 1992. P. 467-479.

4. Open-Source Large Vocabulary Continuous Speech Recognition Engine [Электронный ресурс]. URL: <https://github.com/julius-speech/julius>.

5. Программные продукты, локализованные на татарский язык [Электронный ресурс]. URL: <http://tatsoft.tatar>.

6. Moulines, E., Charpentier, F. “Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones”. In Speech Communication, 9 (5/6), 1990, pp. 453–467.

7. Sagisaka, Y. ATR v-talk speech synthesis system. In Proc. ICSLP-92, 1992, Banff, Canada.

8. Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., Kitamura, T. “Simultaneous modeling of spectrum, pitch and duration in HMM-based speech synthesis”. In Proc. Eurospeech, 1999, pp. 2347–2350.

УДК 161.112:811.512.1-027.271

#### SIMILARITIES AND DIFFERENCES OF TURKIC LANGUAGES

*Adali, Eşref, PhD (Engineering), Full Professor, Istanbul Technical University, Turkey  
e-mail : [adali@itu.edu.tr](mailto:adali@itu.edu.tr)*

Although the origin of the Turkic languages area the same, by the time they have been changed. In this paper we will show the similarities and differences of Turkic languages. We take six Turkic languages in our study which are Turkish, Azeri, Turkmen, Uzbek, Uighur, Kazakh and Tatar languages. In this context, the alphabets, phonology, morphology and syntax of Turkic language will be shown.

**Keywords:** Turkic Languages, alphabets, phonology, morphology, syntax