

ФОНЕТИЧЕСКОЕ ПРЕОБРАЗОВАНИЕ ЗАИМСТВОВАННЫХ СЛОВ ДЛЯ СИНТЕЗАТОРА РЕЧИ НА ЯКУТСКОМ ЯЗЫКЕ

Леонтьев Ньургун Анатольевич

К.т.н., доцент, Северо-Восточный федеральный университет им. М. К. Аммосова (г. Якутск)
leonza@mail.ru

PHONETIC CONVERSION OF BORROWED WORDS FOR A SPEECH SYNTHESIZER IN THE YAKUT LANGUAGE

N. Leontiev

Summary. This article describes the problem of pronunciation of borrowed words in a speech synthesizer in the Yakut language. The official spelling of the borrowed words is given, the national corpus of the Yakut language is analyzed. A variant of the pronunciation of borrowed words by replacing syllables with adapted versions is proposed.

Keywords: speech synthesis, language corpora, speech processing, borrowed words, Yakut language.

Аннотация. В данной статье описывается проблема произношения заимствованных слов в синтезаторе речи на якутском языке. Приводятся официальное правописание заимствованных слов, анализируется национальный корпус якутского языка. Предлагается вариант произношения заимствованных слов с помощью замены слогов на адаптированные варианты.

Ключевые слова: синтез речи, языковой корпус, обработка речи, заимствованные слова, якутский язык.

Введение

Современный язык не является жесткой конструкцией, он меняется во соответствии с временем, в него входят новые понятия, новые сущности, а также происходит заимствование слов, чаще всего из языков имеющих сильное влияние на жизнь и культуру народа. Произношение словозаимствования изменяется под влиянием согласно фонетическим правилам и привычкам родного языка. Фонетические привычки языка ведут к искажению исходного слова, что вносит определенный акцент, но не мешает смысловому пониманию слова.

«Вариативность, будучи объективным свойством языковой системы, является одним из самых ярких проявлений непрерывного эволюционного развития языка. Любое иноязычное слово, входя в язык-реципиент, проходит различные этапы адаптации, первичной из которых является фонетическая адаптация. Важным моментом в процессе фонетической адаптации является многообразие форм передачи фонетического состава иноязычных слов в новом языке в период их вхождения — наличие фоновариантов.

Важным моментом в процессе фонетической адаптации является многообразие форм передачи фонетического состава иноязычных слов в новом языке в период их вхождения — наличие фоновариантов. Это значит, что одно и то же слово может иметь не одну звуковую оболочку или акцентную структуру, которая не обладает смыслоразличительным характером и которая образовалась заменой фонем или изменением позиции ударения в слове. При возникновении вариативности необходимым условием фонетической адаптации слова является прежде всего стабилизация его формы, отбор из нескольких вариантов того варианта, который закреплен в литературном употреблении в качестве нормативного.

На вариативность заимствованных слов может влиять такой факт, как применение транслитерации или транскрипции при заимствовании» [1].

«Табасаранский язык имеет более трех тысяч заимствованных слов их русского языка. Анализ персидских заимствований в табасаранском языке позволил ему выявить около 400 иранизмов. Многие арабизмы, усвоенные несколько веков тому назад, и поныне занимают

прочное место в табасаранской лексике, в которой их количество достигает 800 единиц. Из азербайджанского языка заимствовано около 1300 лексических единиц» [2].

“Следует отметить, что со времен проникновения русских слов в якутский язык, всегда идет тенденция к фонетизированию (якутизации) заимствованных слов. Это объясняется различием звуковых систем русского и якутского языков, а также стихийной фонетизацией заимствованных слов в дореволюционное время. Еще в то время в якутском языке было заимствовано устным путем свыше трех тысяч слов и это обстоятельство создало устойчивую традицию” [3].

Многие языки имеют заимствованные слова из русского и других языков, например в казахском [4], бурятском, монгольском, даже в языке аборигенов Папуа-Новая Гвинея.

ЯКУТСКИЙ ЯЗЫК

Согласно Постановлению Правительства Республики Саха (Якутия) «О Правилах орфографии и пунктуации языка саха» от 22 декабря 2015 г. № 501:

«IV. Правописание заимствованных из русского языка слов 1. Правописание основы слов

При правописании заимствованных из русского языка слов установлены следующие правила: 1) Заимствованные русские слова, фонетически и грамматически давно освоенные якутским языком, типа: *остуол, куорат, дьыала, бирикээс, оскуола, сокуон, дьааһыла, ачыкы*, а также слова, получившие широкое употребление в разговорной практике, легко приспособившиеся к фонетическим нормам якутского языка, типа *автобус — оптуобус, журналист — суруналыыс, конкурс — куонкурус, рынок — ырыынак, студент — устудьуон, тема — тизмэ*, пишутся в фонетизированной форме.

2) Слова, относящиеся к области общественно-политической, научно-технической лексики, не поддающиеся фонетической обработке, такие, как: *гроссмейстер, неолит, полиграфия, фортепьяно*, пишутся в исходной форме. П р и м е ч а н и е. Такие слова в случае широкого распространения в разговорной практике их якутизированного варианта впоследствии могут быть преобразованы в письме в пользу последнего.

Географические названия и имена людей пишутся в русском начертании или фонетизированной форме, например: *Япония — Дьоптуон, Алдан — Аллан, Сибирь — Сибиир, Америка — Эмиэрикэ; Иванов, Пе-*

тров, Сидоров, Мария, Фёкла, Пётр, Тумусов; Уйбаныап, Бөтүрүөп, Сиидэрэп, Маарыйа, Сүөкүлэ, Бүөтүр, Тумууһап.

Заимствованные географические названия и мужские фамилии, выраженные полной формой прилагательного, а также названия городов с конечным -ск при орфографировании пишутся с конечными -ай или -эй, например: *Новой Гвинея, Охотской муора, Пекарской, Горькай, Курскай, Минскэй, этиловай испиир, физическэй география*. Примечание. Конечный ударный о не изменяется, например: *Чудской күөл, Чусовской, Донской.*»

Различие в фонетике разных языков начинают резко выявляться при компьютерном синтезе устной речи, на стадии разделения слов на речевые фрагменты. В случае применения не адаптированного варианта возникает проблема произношения фонетически не характерных слогов в якутском языке.

В якутском языке отсутствуют такие звуки как: ц, е, ш, щ, з, ф, в, ж, ю, я. Такие отсутствуют комбинации гласных звуков: **нрав, страх** и т.д. Структура слога якутского языка является следующей: VV, CVC, CVVC, VC, VVC, CVV, CV, VVC, VC и редкое сочетание CVV+рт. C-consonant (согласный звук), V-vowel (гласный звук).

Авторами создан компьютерный синтезатор речи на основе полуслоговых фрагментов устной речи якутского языка. При создании базы данных фрагментов был использован газетный корпус якутского языка [5,6].

Имеются слова с уже установившейся грамматикой, а есть только формирующиеся, которые могут иметь несколько вариантов правописания. Для произношения нового не адаптированного слова предлагается применить триграммы якутского языка. Триграммы якутского языка применяются для идентификации текста языка [7], таким образом наиболее часто используемые триграммы соответствуют частотному свойству языка.

Например слова, с установившимся правописанием предлагается преобразовать в соответствии со словарем: *газета — хаһыат, ложка — ньуоска, свинья сибиинньэ, спирт — испиир, испиирт, ярмарка — дьаарбанка, герой — дьоруой, солдат — саллаат, целковый — солкуобай (применяется в понятии «рубль»), самовар — сылабаар, ведро — биэдэрэ.*

Также широко ходят в якутский язык словозаимствования из иностранных языков, в частности английского языка. Широкое распространение сети Интернет создает прямые словозаимствования с английского языка, раньше такие заимствования проходили, например че-

рез русский язык. Примеры слов на иностранном языке: *Internet* — *Интернет* — *Интэрниэт*, *интиэрнэт*, *whatsapp* — *ватсап* — *батсаан*, *youtube* — *ютьюб* — *үтүүб*, *үтүб*, *iphone* — *айфон* — *айнуон*, *SIM* — *симка* — *сшмкэ*, *smart* — *смарт* — *сымаарт*, *сымаар*, *browser* — *браузер* — *быраһуһэр*, *flash* — *флэш* — *пылээс*, *пылээскэ*, *online* — *онлайн* — *онлайын*, *blogger* — *блоггер* — *булуоггэр*.

Создается набор сегментов речи, например: *нрав* — *нырааб*, *страх* — *ыстыраах*, *пиастр* — *пыастыр*, *фрагмент* — *бырагмиэн*. Произношение заимствованного нового слова в синтезаторе происходит с учетом

данного преобразования, в соответствие структуры слова якутского языка.

Заключение

При ограниченной базе языковых фрагментов речевого синтезатора невозможно произвести корректно заимствованное слово согласно языку оригинала, необходимо фонетически преобразовать его для якутского языка. Для этого можно применять триграммы и подбирать соответственно фрагмент речи. Неустановившиеся фонетика слова позволяет в последующем выбрать один из вариантов произношения.

ЛИТЕРАТУРА

1. Турдуматова Э.Б., «Вариативность фонетической структуры заимствованных слов (на материале англицизмов в русском языке)», Вестник Калмыцкого университета, № 1 (29), 2016, с. 65–70
2. Загиров В.М., Адилов А. Г., Курбанова Р.Г. «Заимствованные служебные слова в табасаранском языке», Филологические науки. Вопросы теории и практики № 1–2 (55), 2016, с. 125–128
3. Васильева Н.М., «Об употреблении заимствованных слов в газете «Саха сирэ» научный альманах № 3(5), 2015, с. 236–238
4. Кусманова К.А. «Употребление заимствованных слов из русского языка в айтысе», Наука и современность № 2–3, 2010, с. 126–131
5. Леонтьев Н.А., Неустроев Н. С. Информационная система «Электронный корпус якутского языка» // Современная наука: Актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2018. № 12. С. 94–97.
6. Leontiev N. The newspaper corpus of the yakut language // в сборнике: Proceedings of the international conference “turkic languages processing” Turklang-2015. Tatarstan Academy of Sciences L. N. Gumilyov Eurasian National University Ministry of Education and Science of the Republic of Kazakhstan Kazan Federal University Institute of Philology and Intercultural Communication. 2015. С. 233–235.
7. Леонтьев Н.А., Слепцов И. А. Идентификация текстового документа с помощью триграмм на материалах якутского языка // Вестник Северо-Восточного федерального университета им. М. К. Аммосова. 2015. № 4 (48). С. 45–50.

© Леонтьев Ньургун Анатольевич (leonza@mail.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»