

Компьютерная лингвистика и интеллектуальные технологии:
по материалам международной конференции «Диалог 2016»

Москва, 1–4 июня 2016

ОЗВУЧИВАНИЕ АНГЛОЯЗЫЧНЫХ СЛОВОУПОТРЕБЛЕНИЙ В СИСТЕМЕ РУССКОЯЗЫЧНОГО СИНТЕЗА «ТЕКСТ-РЕЧЬ» С ПОМОЩЬЮ ПРАКТИЧЕСКОЙ ТРАНСКРИПЦИИ

Черепанова О. Д. (cherepanova.od@gmail.com)

МГУ им. М. В. Ломоносова, Москва, Россия

Ключевые слова: синтез речи по тексту, нормализация текста, практическая транскрипция, оценка, разборчивость речи, естественность речи, русский язык, английский язык

PHONETIZATION OF ENGLISH WORDS IN RUSSIAN TEXT-TO-SPEECH SYNTHESIS BY MEANS OF ORTHOGRAPHIC TRANSCRIPTION

Cherepanova O. D. (cherepanova.od@gmail.com)

Moscow State University, Moscow, Russia

This paper aims to demonstrate how the orthographic transcription method used in Russian Text-to-Speech systems may help to phonetize English words and word combinations occurring in Russian texts. These are mostly names of companies, mass media, different brands and computer programs. As English words are written in the Latin alphabet they must be rendered into a graphic or phonetic system that the Russian synthesizer can interpret. In our work we propose to use the English-Russian orthographic transcription in order to transform English words into the Cyrillic script. Alternatively to transliteration orthographic transcription has the advantage that it relies

on word pronunciation. It also allows to phonetize English words using only the Russian sound set without extending the TTS-system with English phonemes or extra acoustic databases. After English words are presented as a sequence of Cyrillic symbols they can be treated on all following synthesis stages in the same way as normalized Cyrillic words are. So the method proposed is universal and can potentially be used by any Russian TTS-system. In this article we also describe a perceptual experiment that was carried out in order to assess the quality of English words synthesized in this way. The proposed transcription method allowed to achieve a 87% intelligibility rate of English words phonetized by a Russian TTS-program and to highlight several arguable points in the English-Russian orthographic transcription rules.

Key words: text-to-speech-synthesis, text normalization, orthographic transcription, evaluation, intelligibility, naturalness, Russian, English

1. Введение¹. Метод практической транскрипции

Сегодня практически в любом русскоязычном тексте, будь то научная или публицистическая статья, новостной сайт в сети Интернет или инструкция по эксплуатации технического прибора, встречается большое количество вставок на латинице, чаще всего — на английском языке. В первую очередь это институциональные имена собственные: названия компаний и организаций, издательств, средств массовой информации, брендов, программного обеспечения. Чтобы в системе русскоязычного синтеза речи по тексту озвучивать англоязычные словоупотребления, которые представляют собой последовательность знаков латинского алфавита, необходимо перевести их в ту же графическую (или фонетическую) систему, которая используется для озвучивания обычного русскоязычного текста, написанного на кириллице. Для решения данной задачи нами предлагается **метод практической транскрипции**. В отличие от транслитерации, которая широко используется в ситуациях, когда через другую графическую систему требуется передать *написание* последовательности знаков определенного языка на кириллице, практическая транскрипция позволяет передать ее *звучание*, и именно это важно при озвучивании английских словоупотреблений русскоязычной системой синтеза речи по тексту. Действительно, при практической транскрипции мы переводим последовательность символов графической системы *языка-источника* $L1$ (здесь — латиницы) в последовательность символов графической системы *языка-восприемника* $L2$ (кириллицы), произношение которой в языке-восприемнике ближе всего к произношению исходной последовательности в языке-источнике [Успенский 2002]. Имея в распоряжении фонетическую транскрипцию английского слова, по предложенным в настоящей работе правилам англо-русской практической транскрипции можно получить следующие примеры обработки англоязычных вставок:

¹ Статья написана по материалам дипломной работы автора, выполненной на кафедре ТИПЛ филологического ф-та МГУ под руководством в. н. с., д. ф. н. О. Ф. Кривновой.

- (1) *Bloomberg* ['blu:mɒɹg] — /блү+мберг/
 (2) *Microsoft* ['maɪkrəʊsɒft] — /ма+йкрософт/
 (3) *British Airways* ['brɪtɪʃ'eəweɪz] — /бри+тиш э+рвэйс/².

Система правил англо-русской практической транскрипции была разработана нами на основе работ [Ермолович 2009], [Гиляревский, Старостин 1985]. В качестве иллюстрации в таблице 1 представлен пример правил практической транскрипции для фонемы [ɑ:].

Таблица 1. Правила англо-русской практической транскрипции (на примере фонемы [ɑ:])³

Англ. фонема	Практическая транскрипция	Примечание	Примеры
ɑ:	а	основной вариант	<i>castle</i> [kɑ:sl] — /ка+сл/
	ал	если обозначается сочетанием с буквой l	<i>palm</i> [pɑ:m] — /па+лм/
	ар	если обозначается сочетанием с буквой r	<i>partners</i> [pɑ:tɹnəz] — /па+ртнэрс/

Нужно отметить, что изначально правила создавались указанными авторами с целью графической передачи английских имен собственных в русских текстах. В целях адаптации оригинальных правил Ермоловича и Гиляревского к возможностям и требованиям русскоязычного синтеза речи мы провели несколько перцептивных экспериментов на разборчивость англоязычных словоупотреблений, озвученных русскоязычным синтезатором с использованием предложенной ими системы практической транскрипции. С учетом анализа речи телеведущих и дикторов российских новостных каналов (см. ниже раздел 2) и результатов проведенных экспериментов (см. ниже раздел 3) исходные правила были скорректированы под задачу синтеза речи и протестированы дополнительно в озвученном синтезированном варианте.

К важным преимуществам предлагаемого нами способа обработки слов на латинице относится среди прочего его *оптимальность*: вместо составления обширных словарных списков или написания дополнительного языкового модуля мы используем сравнительно небольшой набор правил англо-русской

² Здесь и далее практическая транскрипция выделяется с двух сторон косой чертой («/»), а знаком «+» обозначается ударение. По умолчанию ударение ставится на тот же слог, что и в фонетической транскрипции английского слова.

³ [Ермолович 2009: 52]

практической транскрипции для перевода английских слов на кириллицу. Кроме того, предлагаемый метод *универсален*, то есть может быть использован в любой русскоязычной системе синтеза речи: после того, как вставка на латинице переведена в кириллическую графическую систему, она может обрабатываться на всех дальнейших этапах синтеза речи по тем же правилам, что и обычные (нормализованные) русские слова на кириллице. Однако, поскольку озвучивание англоязычных вставок на основе правил практической транскрипции осуществляется с помощью звуков русского языка (в популярных конкатенативных синтезаторах аллофонного типа), принципиально важной становится оценка качества и приемлемости такого способа озвучивания с точки зрения русскоязычных пользователей данных систем. Этой задаче посвящена основная часть настоящей работы.

2. Основные критерии оценки качества синтезированной речи

Существует большое количество различных подходов к оценке систем синтеза речи по тексту (например, см. [Campbell 2007]). Чтобы оценить качество озвучивания английских слов и словосочетаний в русскоязычной системе синтеза, вероятно, стоит обратить внимание на две основные качественные характеристики синтезированной речи: **разборчивость** (количество правильно распознанных единиц) и **естественность** (оценка близости синтезированной речи к естественной).

2.1. О разборчивости озвученных англоязычных вставок

Обычно оценка качества синтезированной речи начинается с параметра разборчивости. В этой связи нужно отметить, что около 90% англоязычных вставок в русских текстах — это институциональные имена собственные и любые другие названия, обычно в той или иной степени знакомые пользователю (*The New York Times, Forbes, Apple, Bloomberg*). Соответственно, под разборчивостью (или распознаваемостью) синтезированного словоупотребления в проведенном нами исследовании понимается способность слушателя определить, о какой сущности идет речь, или же (в ситуациях, когда название или слово слушателю незнакомо) письменно воспроизвести примерное написание озвученного фрагмента. Для оценки разборчивости англоязычных вставок, озвученных с помощью практической транскрипции, нами были проведено несколько целевых перцептивных экспериментов. Систематические ошибки испытуемых при восприятии определенных словоупотреблений позволили выделить ряд спорных вопросов в оригинальных правилах практической транскрипции, на которых имеет смысл остановиться подробнее, что будет сделано в разделе 3.

2.2. Естественность озвученных англоязычных вставок в русскоязычном контексте

Второй параметр качества синтеза английских словоупотреблений — **естественность** — является с нашей точки зрения более неоднозначным и субъективным. С одной стороны, метод практической транскрипции в русскоязычном синтезе, при всех его достоинствах, «обречен» на малую естественность озвученных англоязычных вставок относительно норм произносительного стандарта английского языка. С другой стороны, если понимать под естественностью «соответствие норме, эталону», то каким именно нормам должны соответствовать озвученные английские словоупотребления? Нормам произношения английского или русского языка? Можно ли вообще четко определить, как произносительные нормы русского языка распространяются на английские слова в русскоязычном контексте?

Для того чтобы изучить, как русскоговорящие дикторы произносят английские названия в русском контексте, нами был проанализирован аудио- и видеоматериал новостных сайтов и технические обзоры Интернет-магазинов. Выбор данных ресурсов объясняется тем, что произношение телеведущих и дикторов в официальных источниках информации зачастую рассматривается как некий произносительный стандарт, который влияет на распознаваемость акцентно окрашенных английских слов носителями русского языка и который, следовательно, надо учитывать при озвучивании англоязычных вставок русскоговорящим синтезатором.

Анализ материала показал, что русскоязычные дикторы имеют привычку фонетически модифицировать английские слова, заимствованные в русскую речь. Для английских звуков, отсутствующих в русском языке, подыскиваются ближайшие по звучанию звуки-замены: вместо *Twitter* [ˌtwɪtə] и *Forbs* [fɔːps] произносится «русифицированное» [тв'йтър], [форпс]. В окончании *-ing* на месте носового [ɪŋ] в британском английском произносится русское [ьнк] («консалтинг», «драйвинг» — [кансáлт'ьнк], [дрáйв'ьнк]), а в союзе *and* не происходит редукции, поэтому в названии компании *Ernst & Young* мы слышим отчетливое [энт] вместо английского [ən].

В результате анализа нами было выделено несколько факторов, которые влияют на степень выраженности русского произносительного акцента при озвучивании англоязычных фрагментов. Безусловно, большое значение имеет **уровень владения английским языком** говорящего. Но, как показал анализ аудиоматериала, говорящий лишь в редких случаях полностью переключается на английский язык вне зависимости от его уровня владения языком. Не менее важным фактором является **распространенность** англоязычной вставки в повседневной жизни носителей русского языка. Чем выше частотность слова, тем выше вероятность того, что оно будет озвучено диктором «по-русски». Так, в проанализированных материалах социальная сеть *Twitter* действительно произносится всегда русифицировано. То же самое можно сказать о *Bluetooth*, *Word*, *Amazon*, *Microsoft*, *Facebook* и многих других англоязычных вставках. Однако даже менее распространенные слова (*WhatsApp*,

Verge, *Silencer* и др.) в проанализированном материале часто произносятся говорящими «по-русски». Еще один существенный фактор — **длина** англоязычной вставки. Целые фразы на английском языке и просто длинные словосочетания редко употребляются в русскоязычных текстах, но во всех прослушанных записях они озвучиваются диктором с английским произношением (насколько это позволяет уровень владения английским языком говорящего): *we are hunted*, *Glass Explorer Edition*; *News of the world*. Тем не менее, длинные англоязычные фрагменты иногда озвучиваются русифицировано, особенно если данные словосочетания достаточно известны: *Work & Travel*, *Windows Phone 7*, *Amazon Kindle Paperwhite*.

Таким образом, если говорить о естественности синтеза речи по тексту, можно сделать вывод, что при озвучивании отдельных английских слов и не очень длинных словосочетаний вполне допустимо (а иногда даже желательно) использовать звуковой материал русского языка. С этой точки зрения практическая транскрипция как способ передачи англоязычных вставок при синтезе речи кажется вполне оправданной.

3. Разборчивость синтезированных англоязычных вставок (экспериментальные данные)

3.1. Материал и методика эксперимента

Чтобы оценить качество синтеза англоязычных вставок по предложенным правилам практической транскрипции и проверить текущие корректировки исходных правил, нами был проведен перцептивный эксперимент с участием 16 русскоязычных испытуемых (с неоконченным или окончанным высшим образованием, в возрасте от 18 до 54 лет). Два участника эксперимента практически не владели английским языком, остальные же знали английский на уровне не ниже начинающего. Из русских новостных текстов нами было отобрано 27 англоязычных слов и словосочетаний на латинице. Практическая транскрипция, написанная нами вручную для всех 27 словосочетаний, была озвучена с помощью русского синтезатора «Агафья»⁴. Каждому испытуемому необходимо было прослушать 27 записей с озвученными англоязычными словоупотреблениями, отвечая после каждой записи на соответствующее контрольное задание онлайн-опросника. В заданиях испытуемый должен был либо выбрать правильный ответ из нескольких вариантов, либо самостоятельно написать услышанное слово/словосочетание (данный тип задания выбирался нами для тех англоязычных вставок, которые достаточно часто употребляются носителями русского языка в повседневной жизни).

⁴ Разработан речевой группой ОТиПЛ МГУ [Кривнова 1998], [Бабкин, Захаров 1999].

3.2. Результаты эксперимента

Для каждого из 27 словосочетаний был подсчитан процент его распознаваемости испытуемыми. Если определять разборчивость озвученных англоязычных вставок как относительное количество испытуемых, успешно их распознавших, то с помощью правил англо-русской практической транскрипции, использованных в настоящем исследовании, достигается средний уровень разборчивости 86,6%.

Отметим еще раз, что значительная часть озвученных словосочетаний — в той или иной степени знакомые носителям русского языка имена собственные. Как и следовало ожидать, стимулы, успешно распознанных всеми без исключения испытуемыми, включают названия известных компаний и организаций (*Dream Works, Adidas Global*), компьютерных программ и сервисов сети Интернет (*Microsoft Word, WebMoney*), а также название футбольной команды *Manchester United*. Тем не менее, среди идентифицированных в 100% случаев словосочетаний оказались и явно менее распространенные *Market Watch* и *Drag and Drop*, в то время как популярная социальная сеть *Twitter* была распознана лишь каждым третьим испытуемым. Как видно из примеров, приведенных выше, известность названия напрямую на его распознаваемость не влияет. Систематические ошибки испытуемых позволили выделить следующие спорные вопросы в использованных правилах англо-русской практической транскрипции:

1. Место словесного ударения. Целый ряд английских слов был заимствован в русский язык со смещением ударения на последний (или предпоследний) слог. Речь идет о таких словах как *email, Facebook, Android, YouTube, Manchester* и многих других. Оказалось, что озвучивание подобных вставок с ударением «по-английски» достаточно сильно понижает их распознаваемость⁵:

(4) *Email* [ˈi:meɪl] /и+мейл/ — распознаваемость 45%
 /имэ+йл/ — 96%

(5) *Facebook* [ˈfeɪsbʊk] /фе+йсбук/ — 67%.
 /фэйсбу+к/ — 78%.

2. Фонема [w]. Гиляревский и Старостин предлагают передавать фонему [w] через /в/ перед буквой «у» и через /у/ во всех остальных случаях (*woods* — /вудс/, *windows* — /уи+ндоус/). В связи с систематическими ошибками испытуемых в таких словах как *Twitter* («тулитер», «tuitер»), *Windows* («уинданс»), *Paperwhite* («paperlight») данную фонему мы предлагаем передавать только через /в/. Откорректированные правила транскрипции позволили улучшить разборчивость синтезированных слов, содержащих звук [w]:

⁵ Стоит отметить, что распознаваемость примеров (4) и (5), скорее всего, повысилась также из-за более удачной транскрипции дифтонга [eɪ] как /эй/.

Microsoft Word /ма+йкрософт ворд/, *Market Watch* /ма+ркет вотч/ и *Web Money* /вэб ма+ни/ были успешно распознаны в 100% случаев, *Black and White* /блэк энд ва+йт/ — в 97%, *swift* /свифт/ — почти в 85%.

3. Суффиксы *-er* и *-ed*. Изначально по правилам Ермоловича фонема [ə] передавалась транслитерацией (*galaxy* ['gæləksi] — /гэлакси/, *Amazon* ['æməzən] — /э+мазон/, *Russian* ['rʌʃən] — /ра+шиан/). В таких словах как *partner* ['pa:tənə] /па+ртнер/ гласный [e] смягчал предшествующий согласный [n] и по правилам ассимиляции смягчался и предшествующий зубной [t] ([пáрт'н'ер]). В результате были получены такие ответы, как «партиз», «partes», «parties», «partiers», «partirs» и «parteurs». Вопрос о том, насколько критично при озвучивании англоязычных словоупотреблений смягчение согласных перед гласными переднего ряда, свойственное русскому языку, требует отдельного рассмотрения, поэтому нами был проведен анализ русской речи дикторов (см. выше). Оказалось, что телеведущие при произнесении англоязычных вставок смягчают согласные по большей части так же, как и в русском языке: *Christies* — [кр'йс'т'ьс]; *Acoustic* — [гкúс'т'ьк]. Однако есть ряд суффиксов, перед которыми (или в которых) смягчение обычно не происходит. Сюда относятся суффиксы *-er* (*Warner Brothers, partners*), *-ed* (*associated, limited*), *-ment* (*entertainment*). В связи с этим по откорректированным нами правилам практической транскрипции суффиксы *-er*, *-ed*, *-ment* транскрибируются как /эр/, /эд/, /мэнт/ соответственно.
4. Транскрипция фонем [ð] и [θ]. Передача английского [θ] через /с/ в *BBC Earth* [bi: bi: si: z:θ] /би би си эрс/ привела к низкому уровню различимости вставки (31%): многие участники эксперимента услышали вместо *earth* слово *ears*. Учитывая, что состав согласных звуков в русском и английском языках не позволяет установить однозначного соответствия между ними, подобных трудностей при практической транскрипции англоязычных вставок на кириллицу избежать нельзя. При передаче фонем [ð] и [θ] приходится так или иначе выбирать между русскими согласными [д], [т] (по Гиляревскому), [з] и [с] ([Мещеряков et al. 2011]) или [в], [ф] ([Ротаров 2003]), с помощью которых уже транскрибируются английские [d], [t], [z], [s], [v], [f]. Анализ речи русских дикторов говорит в пользу передачи фонем [ð] и [θ] через /з/ и /с/, об этом же свидетельствует и уже укоренившаяся в русском языке форма «блютуз» для английского *Bluetooth*. В наших откорректированных правилах англо-русской практической транскрипции мы остановились именно на этом варианте передачи фонем [ð] и [θ].

4. Заключение

В настоящей работе рассматривается одна из задач нормализации русскоязычного текста — обработка англоязычных слов и словосочетаний на латинице. Для озвучивания англоязычных словоупотреблений предлагается метод практической транскрипции, позволяющий решать поставленную задачу,

пользуясь только звуковым материалом русского языка. Благодаря корректировкам, внесенным в исходные правила англо-русской практической транскрипции в результате тестирования, удалось достичь 87% распознаваемости английских словоупотреблений, озвучиваемых данным методом с помощью русскоязычного синтезатора «Агафья». В целом, результат можно считать положительным. К основным преимуществам метода практической транскрипции относится его доступность и универсальность. Данный метод позволяет озвучивать английские слова и словосочетания на латинице в русскоязычной системе синтеза без дополнения синтезатора отдельным модулем для английского языка, как это делается, например, в синтезаторах-полиглотах polySVOX. В последнем случае для добавления звуковых единиц английского языка в звуковую библиотеку русскоязычного синтезатора необходимо либо использовать два разных голоса для разных языков, либо искать для записи материала донора-билингва. Универсальность метода практической транскрипции заключается в том, что на выходе блока нормализации формируются «нормализованные» слова на *кириллице*, которые на всех дальнейших этапах синтеза обрабатываются программой по тем же правилам, что и обычные слова русского языка. Благодаря этому одни и те же правила практической транскрипции можно включить в любую русскоязычную систему синтеза речи по тексту, независимо от фонетической транскрипции, которая в ней используется.

Как показывают результаты анализа речи телеведущих, носители русского языка привыкли озвучивать и воспринимать на слух английские слова в их русифицированной форме. Таким образом, озвучивание английских слов с помощью звукового материала русского языка, подобранного по правилам англо-русской практической транскрипции, может повысить не только разборчивость, но и естественность синтезируемой русской речи.

Литература

1. *Бабкин А. В., Захаров Л. М.* (1999) Оценка качества системы синтеза речи, разработанного в МГУ // Труды Межд. семинара по компьютерной лингвистике и ее приложениям «Диалог99». М..
2. *Гиляревский Р. С., Старостин Б. А.* (1985) Иностранные имена и названия в русском тексте: Справочник // 3-е изд., М.
3. *Ермолович Д. И.* (2009) Методика межъязыковой передачи имен собственных // М.
4. *Кривнова О. Ф.* (1998) Автоматический синтез русской речи по произвольному тексту (вторая версия с женским голосом) // Труды Международного семинара по компьютерной лингвистике и ее приложениям «Диалог'98». М.
5. *Мещеряков Р. В., Тиунов С. Д., Лирмак Ю. М., Шевкунова А. Е.* (2011) Речевые технологии в задаче обучения студентов-носителей русского языка произношению на иностранном языке. // «Анализ разговорной русской речи» (АРЗ-2011): Труды пятого междисциплинарного семинара, СПб.: ГУАП, С. 78–82.
6. *Успенский В. А.* (2002) Труды по нематематике // М., ОГИ, С. 390–412.

References

1. *Babkyn A. V., Zakharov L. M.* (1999) Quality assessment of the MSU Text-to-Speech system [Otsenka kachestva sistemy sinteza rechi, razrabotannogo v MGU], Proc. Int. seminar in computational linguistics and its applications “Dialogue’99” [Trudy mezhdunarodnogo seminarara po komp’yuternoy lingvistike i yeyo prilozheniyam “Dialog 1999”], Moscow.
2. *Gilyarevskiy R. S., Starostin B. A.* (1985) Foreign names in Russian texts: Handbook [Inostrannye imena i nazvaniya v russkom tekste: Spravochnik], 3rd edition, Moscow.
3. *Ermolovich D. I.* (2009) A method of interlingual translation of proper nouns [Metodika mezhjazykovoy peredachi imen sobstvennykh], Moscow.
4. *Krivnova O. F.* (1998) Automatic Text-to-Speech synthesis (second version with a women voice) [Avtomaticheskiy sintez russkoy rechi po proizvol’nomu tekstu (vtoraya versiya s zhenskim golosom)], Proc. Int. seminar in computational linguistics and its applications “Dialogue 1998” [Trudy mezhdunarodnogo seminarara po komp’yuternoy lingvistike i yeyo prilozheniyam “Dialog 1998”], Moscow.
5. *Meshcheryakov R. V., Tiunov S. D., Lirmak Ju. M., Shevkunova A. E.* (2011) Speech technologies in the task of foreign language pronunciation education of Russian-speaking students [Rechevye tekhnologii v zadache obucheniya studentov-nositeley russkogo jazyka proiznosheniyu na inostrannom jazyke], “Russian spoken language analysis”: Proc. 5th interdisciplinary seminar [Analiz razgovornoy russkoy rechi (ARZ-2011): Trudy pyatogo mezhdistsiplinarnogo seminarara)], GUAP, St. Petersburg, pp. 78–82.
6. *Uspenskiy V. A.* (2002) Writings on non-mathematics [Trudy po nematematike], OGI, Moscow, pp. 390–412.
7. *Campbell N.* (2007) Chapter 2. Evaluation of speech synthesis. From Reading Machines to Talking Machines. // L. Dybkjoer et al. Evaluation of Text and Speech Systems, Springer Verlag, pp. 29–64.
8. *Potapov V.* (2003) The American English Interference in Russian on the Segmental Level, Proc. of the intern. workshop “Speech and computer” (SPECOM’2003), Moscow, pp. 283–290.