

УС в разных текстах. Для дословного перевода УС идеален машинный переводчик Google. Переводчик Prompt дает разные способы перевода: эквивалентные, дословные и даже ошибочные (*plain as day – равнина как день*), поэтому не рекомендуется использовать его для перевода УС.

Таблица 3. – Функции прослушивания и транскрипции в онлайн-словарях

Функции онлайн-словарей	google.translate	translate.ru (Prompt)	multitran.ru	lingvo.ru
озвучивание	+	+-	+-	+- (только пословно)
транскрипция	-	+-	+-	+- (отдельных слов)

Функция прослушивания фраз (в том числе целого УС) доступна только в машинном переводчике Google (таблица 3). Остальные словари дают возможность прослушивания отдельного слова. Но переводчик Google не предоставляет транскрипцию слов, в отличие от других рассмотренных словарей.

Таким образом, на начальном этапе освоения английского языка и английских УС для их прослушивания и упрощенного перевода можно использовать переводчик Google. На продвинутом этапе изучения английских УС, а также для качественного перевода УС с английского на русский и с русского на английский рекомендуем пользоваться онлайн-словарем Multitran.

Зяноўка Я.С.

Аб'яднаны інстытут праблем інфарматыкі НАН Беларусі, Мінск

РАСПРАЦОЎКА ЛІНГВІСТЫЧНЫХ РЭСУРСАЎ ДЛЯ АЎТАМАТЫЧНАЙ ТРАНСКРЫПЦЫ БЕЛАРУСКАМОЎНЫХ ТЭКСТАЎ НА АНГЛІЙСКОЙ МОВЕ

Англійская мова з’яўляецца інтэрнацыянальнай мовай. Гэта родная мова для англічан, аўстралійцаў, новазеландцаў, большай часткі насельніцтва ЗША, адна з дзвюх афіцыйных моў Ірландыі, Канады, Мальты і многіх іншых краін. Гэта трэцяя мова па колькасці гаворачых пасля кітайскай і іспанскай моў. Больш за 1 мld чалавек выкарыстоўвае яе (уключаючы носьбітаў і тых, для каго англійская мова не з’яўляецца роднай). Шмат людзей лічаць, што ў будучыні яна стане лінгва франка – універсальнай мовай, на якой зможа размаўляць увесь свет. Але, нягледзячы на такую распаўсюджанасць, застаецца вялікая колькасць людзей, якія не ведаюць англійскую мову.

Каб спрасціць працэс камунікацыі на дадзенай мове, аўтар у суполцы з камандай лабараторыі распазнавання і сінтэзу маўлення АІП НАН Беларусі распрацавала спіс адпаведнасцей “англійскі алафон – беларускі алафон” для атрымання магчымасці беларускамоўным карыстальнікам пазнаёміцца з фанетычнай будовай англійскай мовы, выкарыстоўваць транскрыпцыю англійскага гучання беларускімі сімваламі, праслухваць англійскія словы на беларускі лад.

За аснову распрацоўкі ведаў быў узяты электронны слоўнік “Carnegie Mellon University Pronouncing Dictionary (CMU)” [1], распрацаваны прыватным універсітэтам Карнэгі – Меллан (Піцбург, штат Пенсільванія). CMU – слоўнік у адкрытым доступе, які змяшчае арфаэпічны код (код у алафонным выглядзе), чытабельны машынай. Слоўнік распрацаваны на базе паўночна-амерыканскага вырыянта англійскай мовы і налічвае больш за 134000 слоў з указаннем правільнага вымаўлення. Слоўнік асабліва карысны для распазнавання і сінтэзу маўлення, так як мае адпаведнасці “слова – алафонны выгляд” па наборах фанем ARPAbet – стандарта для англійскага вымаўлення. Бягучы набор фанем змяшчае 39 фанем, уключаючы націскныя, часткова націскныя і ненаціскныя галосныя і зычныя (малюнак 1).

ABSOLUTE	AE1 B S AH0 L UW2 T
ABSOLUTELY	AE2 B S AH0 L UW1 T L IY0
ABSOLUTENESS	AE1 B S AH0 L UW2 T N AH0 S
ABSOLUTES	AE1 B S AH0 L UW2 T S
ABSOLUTION	AE2 B S AH0 L UW1 SH AH0 N
ABSOLUTISM	AE1 B S AH0 L UW2 T IH2 Z AH0 M
ABSOLUTIST	AE0 B S IH0 L UW1 T IH0 S T
ABSOLVE	AH0 B Z AA1 L V
ABSOLVE(1)	AE0 B Z AA1 L V
ABSOLVED	AH0 B Z AA1 L V D

Малюнак 1. – Фрагмент змешчаных слоў у алафонным выглядзе CMU

Па спісе алафонаў англійскай мовы было выбрана ў сярэднім па 3-7 слоў у літарным і алафонным выглядзе на кожную літару алфавіта для пакрыцця ўсіх алафонаў. Такім чынам, атрымалася 69 слоў (столькі ж колькі і алафонаў у англійскай мове). Асноўнай задачай у дадзеным выпадку стаў не пошук мінімальнай колькасці слоў для пакрыцця ўсіх алафонаў, а наадварот, як мага больш слоў для

далейшага параўнання алафонаў у розных словах і вылучэння адпаведных алафонаў для беларускай мовы (малюнак 2).

Наступным крокам стаў пошук адпаведнасцей “англійскі алафон – беларускі алафон”. У працэсе падбора адпаведнасцей былі вылучаныя як галосныя манафтонгі (AAO – A) і дыфтонгі (AWO – AY), так і зычныя манафтонгі (CH – Ч) і дыфтонгі (JH – ДЖ) (малюнак 2).

Word	Phoneme transcription	Словы па беларуску
ACCESSIBLE	AE0 K S EH1 S AH0 B AH0 L	Э К С Э+ С А Б А Л
ACCOMMODATION	AH0 K AA2 M AH0 D EY1 SH AH0 N	А К А= М А Д Э+ Ш А Н
ALREADY	AO0 L R EH1 D IY0	О Л Р Э+ Д І
ASK	AE1 S K	Э+ С К
ASSEMBLY	AH0 S EH1 M B L IY0	А С Э+ М Б Л І
ASSOCIATION	AH0 S OW2 S IY0 EY1 SH AH0 N	А С О=У С І Э+ Ш А Н
AUGUSTIN	AW0 G UW0 S T IY1 N	АУ Г У= С Т І+ Н
BACHELOR	B AE1 CH AH0 L ERO	Б Э+ Ч А Л Э
BEHAVIORS	B IH0 NH EY1 V Y ERO Z	Б Ы Х Э+ В Й Э Э
BELONG	B IH0 L AO1 NG	Б Ы Л О+ Н
BIMODAL	B AY0 M OW1 D AH0 L	Б А І М О+У Д А Л

Малюнак 2. – Фрагмент атрыманага спіса канвертаваных слоў

Па першапачатковых выніках стала зразумела, што ў файле існуе шмат недапрацовак і несутыкненняў, якія прадстаўлены ніжэй:

Па-першае, усе галосныя дыфтонгі былі пераўтвораны ў спалучэнне галоснага і зычнага гукаў (AWO – AY – AЎ, AY0 – AI – АЙ, EY0 – ЭИ – ЭЙ, OY0 – ОИ – ОЙ, OW – ОУ – ОЎ) у адпаведнасці з мелодыкай гучання англійскай мовы. Агульная колькасць пераўтвораных дыфтонгаў з 39 фанем – 13.

Па-другое, былі знойдзены наяўныя несутыкненні англійскай транскрыпцыі алафонаў з падобраным адпаведнікам (памылкі лінгвіста-эксперта). Колькасць выпраўленых неадпаведных гучанню алафонаў – 6, прычым усе галосныя (AH0, IH0, IH1, AA0, AA1).

Па-трэцяе, варыятыўнасць гучання аднаго і таго ж алафона ў розным асяроддзі. Напрыклад, алафон AH0 не мае адзінага гучання ва ўсіх 69 тэставых словах. Гэты вынік можна зрабіць з таго, што з 39 слоў, у якім ён ужываецца, толькі ў 31 слове ён рэалізаваны як -э, у 5 як -і, і ў астатніх 3 гучыць як -а. Гук R з 22 ужыванняў у 6 не павінен гучаць зусім. Аб апісаных памылках сведчыць табліца ніжэй.

Па-чацвёртае, недакладнасць прыведзеных у слоўніку CMU алафонаў у розных словах, якія маюць супярэчнасці паміж сабой (напрыклад, AE2: AYIGLAE2SAH0Z - А+ЙГЛЭ=CЭЗ - 'aigla:səs i F IH1SHER0MAE2N - ФІ+ ШЭМЭ=Н - 'fɪjəmən).

Вынікі атрыманага спіса адпаведнасцей “англійскі алафон – беларускі алафон” былі убудаваныя ў сэрвіс “Агучаны электронны

граматычны слоўнік” [2] інтэрнэт-платформы corpus.by [3]. На дадзены момант існуе шмат недахопаў, якія звязаныя з несупадзеннямі гукавога ўжывання аднаго і таго ж алафона ў розным асяроддзі, так як слоўнік CMU складзены на базе паўночна-амерыканскага вырыянта англійскай мовы і таму мае пэўныя асаблівасці. Гэта можна выправіць толькі ў выніку стварэння свайго асабістага слоўніка. Таксама трэба разумець, што падбраныя адпаведнасці не змогуць перадаваць дакладнае гучанне англійскіх слоў па прычыне рознага фанетычнага строя дзвюх моў. Аднак, нягледзячы на гэта, якасць успрымання англійскай транскрыпцыі беларускімі сімваламі дастаткова высокая, і з’яўляецца зразумелай.

ЛІТАРАТУРА

1. Carnegie Mellon University Pronouncing Dictionary [Электронны рэсурс]. – Рэжым доступу: <http://svn.code.sf.net/p/cmuspinyin/code/trunk/cmudict/cmudict.0.7a>. – Дата доступу: 25.03.2016.
2. Агучаны электронны граматычны слоўнік [Электронны рэсурс]. – Рэжым доступу: <http://corpus.by/VoicedElectronicGrammaticalDictionary/> – Дата доступу: 25.03.2016.
3. Hetsevich, Yu. The system of generation of phonetic transcriptions for input electronic texts in belarusian / Yu. Hetsevich, V. Mandzik, V. Rusak, A. Hiuntar, T. Okrut, B. Lobanov, S. Lysy, Dz. Dzenisiuk // Pattern Recognition and Information Processing: Proceedings Of The 12th International Conference (28-30 May, Minsk, Belarus) – Minsk : UIIP NASB, 2014. – С.81-85.