

Взаимно однозначный англо-русский транслитератор

Инструкция по переводу

1 Введение

Перед вами инструкция по взаимно однозначной транслитерации слов с русского языка на английский и обратно. Она похожа на повседневно используемый «транслит» с тем главным отличием, что каждому русскому слову сопоставляется единственное английское и наоборот. В частности, это значит, что по записи на другом языке можно однозначно восстанавливать исходное слово.

Словом мы, конечно, будем называть любую последовательность букв алфавита, а не только те, что есть в словаре.

2 Транслитерация слова

Начнем с того, что нам дано русское либо английское слово из строчных букв и нужно получить его транслитерацию на другом языке.

Интуиция подсказывает переводить его побуквенно, то есть для каждой буквы взять запись на другом языке и из них составить ответ. Мы будем делать почти так. Ниже приведены правила перевода букв.

а \longleftrightarrow a	и \longleftrightarrow i	с \longleftrightarrow s	ъ \longleftrightarrow oa
б \longleftrightarrow b	й \longleftrightarrow y	т \longleftrightarrow t	ы \longleftrightarrow ea
в \longleftrightarrow v	к \longleftrightarrow k	у \longleftrightarrow ou	ь \longleftrightarrow ie
г \longleftrightarrow g	л \longleftrightarrow l	ф \longleftrightarrow f	э \longleftrightarrow oe
д \longleftrightarrow d	м \longleftrightarrow m	х \longleftrightarrow h	ю \longleftrightarrow u
е \longleftrightarrow e	н \longleftrightarrow n	ц \longleftrightarrow c	я \longleftrightarrow ya
ё \longleftrightarrow yo	о \longleftrightarrow o	ч \longleftrightarrow ch	кь \longleftrightarrow q
ж \longleftrightarrow j	п \longleftrightarrow p	ш \longleftrightarrow sh	кс \longleftrightarrow x
з \longleftrightarrow z	р \longleftrightarrow r	щ \longleftrightarrow shch	ув \longleftrightarrow w
тф \longleftrightarrow th	пф \longleftrightarrow ph		
тх \longleftrightarrow tf	пх \longleftrightarrow pf		

Правила буква-буква понятны, но есть и такие, где слева либо справа стоит последовательность из нескольких букв. Такие правила приоритетнее. Для перевода слова его нужно разбить на части, стоящие в

правилах, и применить соответствующие правила, соблюдая притом следующий закон.

Закон (основной). При переводе слово нужно разбивать на как можно более крупные части, стоящие в правилах.

Пример 1. $\mathcal{ENG}(\text{дом}) = \text{dom}$, $\mathcal{ENG}(\text{улица}) = \text{oulica}$, $\mathcal{RUS}(\text{wall}) = \text{увалл}$, $\mathcal{RUS}(\text{table}) = \text{табле}$ (разбиение на буквы, применение правил для букв).

Пример 2. $\mathcal{RUS}(\text{our}) = \text{ур}$: слово our можно разбить на части двумя способами: o|u|r либо ou|r . Первые две части в первом способе сливаются в одну во втором, значит, выбирается второй; затем части переводятся по правилам и результаты записываются друг за другом в одно слово. Аналогично $\mathcal{RUS}(\text{road}) = \text{рьд}$, $\mathcal{RUS}(\text{this}) = \text{тфис}$, $\mathcal{ENG}(\text{кувшин}) = \text{kwshin}$.

Это не все правила, которые понадобятся. Сейчас, например, $\mathcal{ENG}(\text{ё}) = \mathcal{ENG}(\text{йо}) = \text{yo}$, что нас, естественно, не устраивает. То же самое происходит с любым правилом, содержащим с одной стороны одну букву, а с другой – более одной. Чтобы это исправить, придется добавить бесконечные цепочки правил. В каждой из них между буквами исходного правила «вклинивается» дополнительный символ-разделитель. Ниже каждая цепочка записана под исходным правилом, из-за которого ее пришлось добавить.

$\text{ч} \longleftrightarrow \text{ch}$	$\text{у} \longleftrightarrow \text{ou}$	$\text{ъ} \longleftrightarrow \text{oa}$	$\text{къ} \longleftrightarrow \text{q}$
$\text{цк}^*\text{x} \longleftrightarrow \text{ck}^+\text{h}$	$\text{ой}^*\text{ю} \longleftrightarrow \text{oy}^+\text{u}$	$\text{ох}^*\text{a} \longleftrightarrow \text{oh}^+\text{a}$	$\text{кг}^+\text{ъ} \longleftrightarrow \text{kg}^*\text{oa}$
$\text{ш} \longleftrightarrow \text{sh}$	$\text{э} \longleftrightarrow \text{oe}$	$\text{ы} \longleftrightarrow \text{ea}$	$\text{кс} \longleftrightarrow \text{x}$
$\text{ск}^*\text{x} \longleftrightarrow \text{sk}^+\text{h}$	$\text{ой}^*\text{e} \longleftrightarrow \text{oy}^+\text{e}$	$\text{ex}^*\text{a} \longleftrightarrow \text{eh}^+\text{a}$	$\text{кг}^+\text{c} \longleftrightarrow \text{kg}^*\text{s}$
$\text{щ} \longleftrightarrow \text{shch}$	$\text{ё} \longleftrightarrow \text{yo}$	$\text{ь} \longleftrightarrow \text{ie}$	$\text{ув} \longleftrightarrow \text{w}$
$\text{шт}^*\text{ч} \longleftrightarrow \text{sht}^+\text{ch}$	$\text{йи}^*\text{o} \longleftrightarrow \text{yi}^+\text{o}$	$\text{ий}^*\text{e} \longleftrightarrow \text{iy}^+\text{e}$	$\text{уф}^+\text{в} \longleftrightarrow \text{ouf}^*\text{v}$
	$\text{я} \longleftrightarrow \text{ya}$		
	$\text{йи}^*\text{a} \longleftrightarrow \text{yi}^+\text{a}$		

Когда с одной стороны правила после буквы написана *, а с другой +, это означает, что той буквы, после которой *, может быть любое на единицу меньшее количество, чем той, после которой +. Таким образом, без использования * и + пришлось бы выписывать бесконечные последовательности правил, например, $\text{ш} \longleftrightarrow \text{sh}$, $\text{cx} \longleftrightarrow \text{skh}$, $\text{скx} \longleftrightarrow \text{skkh}$, ... Это и есть цепочка.

Каждая цепочка начинается с правила, в котором с одной стороны стоит одна буква, а с другой – слово из нескольких букв. Ту самую одну букву будем называть головой цепочки. Например, только что мы выписали цепочку с головой ш, или, иначе говоря, это цепочка правил буквы ш.

Пример 3. Рассмотрим добавленную цепочку правил для буквы ч.
 $\mathcal{ENG}(ч) = ch$, $\mathcal{ENG}(цх) = ch$, $\mathcal{ENG}(цкх) = ckch$, $\mathcal{ENG}(цккх) = ckckh$ и так далее. Действительно, в левой части правила цепочки стоит *, в правой – +; символ-разделитель здесь к/k – в русских словах его на один меньше, чем в соответствующих им английских. Очевидно, $RUS(ch) = ч$, $RUS(ckh) = цх$, $RUS(ckkh) = цкх$ и так далее.

Пример 4. 1. $\mathcal{ENG}(щ) = shch$, $\mathcal{ENG}(штч) = shtch$, $\mathcal{ENG}(шттч) = shtttch$, $\mathcal{ENG}(штттч) = shttttch$ и так далее.
 2. $\mathcal{ENG}(кѣ) = q$, $\mathcal{ENG}(кѣѣ) = koa$, $\mathcal{ENG}(кѣѣѣ) = kgoa$, $\mathcal{ENG}(кѣѣѣѣ) = kggoa$ и так далее.
 3. $RUS(w) = ув$, $RUS(ouv) = уфв$, $RUS(oufv) = уффв$, $RUS(ouffv) = уфффв$ и так далее.
 4. $RUS(ea) = ы$, $RUS(eha) = ea$, $RUS(ehha) = еха$, $RUS(ehhha) = ехха$ и так далее.

Внимательный читатель заметит, что транслитерация все еще не однозначна. К примеру, как перевести *you*? Слово можно разбить на части, стоящие в правилах, следующими способами: $y|o|u$, $yo|u$, $y|ou$. Что выбрать, $yo|u$ или $y|ou$, и каков результат перевода, ёу или йу? Та же проблема возникает при попытке перевода слова *yioyu*: выбрать $yio|y|u$ или $y|i|oyu$?

Подобную ситуацию неоднозначности будем называть конфликтом. Конфликт возникает, когда части правил «накладываются» одна на другую одним или несколькими символами. Выше был продемонстрирован конфликт между правилами из цепочек ё и у.

Всего конфликтов девять: ё-у, ё-ѣ, ё-э, э-ы, ѣ-ы, ё-э-ы, х-ш («икс»-ш), х-щ («икс»-щ), йу-w. Первые восемь конфликтов являются конфликтами цепочек и обозначены головными буквами этих цепочек. Выше мы рассмотрели примеры слов из конфликта ё-у.

Ниже приведены три закона перевода, разрешающие эти конфликты, и показано на примерах их применение.

Закон 1. Чтобы выполнить перевод при конфликте ё-у, ё-ѣ, ё-э, э-ы, ѣ-ы либо ё-э-ы, нужно применить обычные правила из конфликтующих цепочек с тем лишь отличием, что непосредственно перед использованием правила с головной буквой символ-разделитель не требуется.

Закон 2. Чтобы выполнить перевод при конфликте х-ш либо х-щ, нужно провести замену на одно слово вперед по цепочке с головой на исходном языке, затем перевод побуквенно, и, наконец, замену на одно слово назад по цепочке с головой на целевом языке.

Закон 3. При конфликте йу-w перевод происходит без учета закона 1 для последовательности букв йу и ей подобных.

Пример 5. Рассмотрим конфликт ё-у.

1. $\mathcal{EN}\mathcal{G}(\text{йу}) = \text{уоу}$, $\mathcal{EN}\mathcal{G}(\text{йиу}) = \text{уіоу}$, $\mathcal{EN}\mathcal{G}(\text{йиийу}) = \text{уіііоу}$: использовано правило с головой у, перед ним разделитель не требуется, а именно, букв и столько же, сколько і.
2. $\mathcal{EN}\mathcal{G}(\text{ёю}) = \text{уоуу}$, $\mathcal{EN}\mathcal{G}(\text{ёйю}) = \text{уоууу}$, $\mathcal{EN}\mathcal{G}(\text{ёйййю}) = \text{уоууууу}$: использовано правило с головой ё, но после разделитель требуется, а именно, букв у на одну больше, чем й.
3. $\mathcal{EN}\mathcal{G}(\text{йюю}) = \text{уіоуу}$, $\mathcal{EN}\mathcal{G}(\text{йиоййю}) = \text{уііоуууу}$ – правил с головным символом не используется, разделители требуются.

Конфликты ё-ъ, ё-э, э-ы, ь-ы разрешаются абсолютно аналогично.

Пример 6. Рассмотрим конфликт ё-э-ы. Он и вовсе тройной (и такой единственный).

1. $\mathcal{EN}\mathcal{G}(\text{ёы}) = \text{уоеа}$, $\mathcal{EN}\mathcal{G}(\text{ёйы}) = \text{уоуеа}$: одновременно использованы правила с первой (ё) и третьей (ы) головами, разделитель й/у перед третьей не требуется.
2. $\mathcal{EN}\mathcal{G}(\text{йоы}) = \text{уіоеа}$, $\mathcal{EN}\mathcal{G}(\text{йиоы}) = \text{уііоеа}$, $\mathcal{EN}\mathcal{G}(\text{йойы}) = \text{уіоуеа}$, $\mathcal{EN}\mathcal{G}(\text{йиойы}) = \text{уііоуеа}$: использовано правило с головой ы, второй разделитель (й/у, непосредственно перед) не нужен, но первый (и/і) нужен.
3. $\mathcal{EN}\mathcal{G}(\text{йэа}) = \text{уоеаа}$, $\mathcal{EN}\mathcal{G}(\text{йиэа}) = \text{уіоеаа}$, $\mathcal{EN}\mathcal{G}(\text{йэха}) = \text{уоеааа}$, $\mathcal{EN}\mathcal{G}(\text{йиэха}) = \text{уіоеаааа}$: использовано правило с головой э, первый разделитель не нужен, третий нужен.
4. $\mathcal{EN}\mathcal{G}(\text{ёеа}) = \text{уоуеаа}$, $\mathcal{EN}\mathcal{G}(\text{ёйеа}) = \text{уоууеаа}$, $\mathcal{EN}\mathcal{G}(\text{ёеаа}) = \text{уоуеаааа}$, $\mathcal{EN}\mathcal{G}(\text{ёйеаа}) = \text{уоууеаааа}$: использовано правило с головой ё, второй и третий разделители нужны.
5. $\mathcal{EN}\mathcal{G}(\text{йоеа}) = \text{уіоуеаа}$, $\mathcal{EN}\mathcal{G}(\text{йиойеаа}) = \text{уііоууеаааа}$: правила с головами не встречаются, все разделители нужны.

Пример 7. Рассмотрим конфликт х-ш.

1. $\mathcal{EN}\mathcal{G}(\text{кш}) = \text{хш}$: $\text{кш} \rightarrow (\text{замена вперед по цепочке для ш}) \text{ксх} \rightarrow \text{кш} \rightarrow (\text{замена назад по цепочке для х}) \text{хш}$; обратный перевод: $\text{хш} \rightarrow (\text{замена вперед по цепочке для х}) \text{кш} \rightarrow \text{ксх} \rightarrow (\text{замена назад по цепочке для ш}) \text{кш}$.

2. $\mathcal{ENG}(\text{кгш}) = \text{ksh}$: $\text{кгш} \rightarrow \text{кгсх} \rightarrow \text{kgsh} \rightarrow \text{ksh}$; обратно: $\text{ksh} \rightarrow \text{kgsh} \rightarrow \text{кгсх} \rightarrow \text{кгш}$.
3. $\mathcal{ENG}(\text{кcx}) = \text{xkh}$: $\text{кcx} \rightarrow \text{кскх} \rightarrow \text{kskh} \rightarrow \text{xkh}$; обратно: $\text{xkh} \rightarrow \text{kskh} \rightarrow \text{кскх} \rightarrow \text{кcx}$.
4. $\mathcal{ENG}(\text{кхсх}) = \text{kskh}$: $\text{кхсх} \rightarrow \text{кхскх} \rightarrow \text{khskh} \rightarrow \text{kskh}$; обратно: $\text{kskh} \rightarrow \text{khskh} \rightarrow \text{кхскх} \rightarrow \text{кхсх}$.

Пример 8. Рассмотрим конфликт х-щ.

1. $\mathcal{ENG}(\text{кщ}) = \text{xhch}$: $\text{кщ} \rightarrow$ (замена вперед по цепочке для щ) $\text{кшч} \rightarrow \text{kshch} \rightarrow$ (замена назад по цепочке для х) xhch ; обратный перевод: $\text{xhch} \rightarrow$ (замена вперед по цепочке для х) $\text{kshch} \rightarrow \text{кшч} \rightarrow$ (замена назад по цепочке для щ) кщ .
2. $\mathcal{ENG}(\text{кгшч}) = \text{kshch}$: $\text{кгшч} \rightarrow \text{кгшч} \rightarrow \text{kgshch} \rightarrow \text{kshch}$; обратно: $\text{kshch} \rightarrow \text{kgshch} \rightarrow \text{кгшч} \rightarrow \text{кгшч}$.
3. $\mathcal{ENG}(\text{кштч}) = \text{xhtch}$: $\text{кштч} \rightarrow \text{кштч} \rightarrow \text{kshtch} \rightarrow \text{xhtch}$; обратно: $\text{xhtch} \rightarrow \text{kshtch} \rightarrow \text{кштч} \rightarrow \text{кштч}$.
4. $\mathcal{ENG}(\text{кгштч}) = \text{kshtch}$: $\text{кгштч} \rightarrow \text{кгштч} \rightarrow \text{kgshtch} \rightarrow \text{kshtch}$; обратно: $\text{kshtch} \rightarrow \text{kgshtch} \rightarrow \text{кгштч} \rightarrow \text{кгштч}$.

Пример 9. Конфликт йу-w отличается от всех других. Здесь конфликтуют правила из цепочки w и закон 1, предписывающий особым образом переводить слова йу, йиу, йииу и так далее. В этом случае перевод выполняется без учета закона 1, лишь согласно цепочке w.

1. $\mathcal{ENG}(\text{йуw}) = \text{uw}$, $\mathcal{ENG}(\text{йиуw}) = \text{yiw}$, $\mathcal{ENG}(\text{йииуw}) = \text{yiiiw}$.
2. $\mathcal{ENG}(\text{йуфw}) = \text{youv}$, $\mathcal{ENG}(\text{йиуфффw}) = \text{yiiouffv}$.

После введения указанных трех законов цель, наконец, достигается: транслитерация однозначна! Итак, мы можем однозначно переводить любые слова из строчных букв с русского алфавита на английский и обратно.

3 Транслитерация текста

Пусть теперь дан текст на русском либо на английском языке и нужно получить его транслитерацию. Под текстом подразумевается последовательность символов, в которой встречаются участки сплошь из букв алфавита (то есть слова), а также некоторые посторонние символы (к примеру, знаки препинания). Все буквы в слове строчные, за исключением, быть может, первой, которая может быть заглавной.

Пример 10. В описанном выше смысле текст "Погода просто ЗаМеЧа-ТеЛьНаЯ" делится на следующие слова: Погода, просто, За, Ме, Ча, Те, Ль, На, Я.

Чтобы перевести текст, нужно выполнить следующие шаги:

- 1) поделить текст на слова в указанном выше смысле;
- 2) выполнить перевод каждого слова, как описано в предыдущем разделе; если в исходном слове первая буква заглавная, в слове-результате ее тоже нужно сделать заглавной;
- 3) записать подряд результаты транслитерации всех слов, сохранив при этом все стоявшие между ними посторонние символы, и получить таким образом транслитерацию текста.

4 Примеры перевода

Широка страна моя родная,
Много в ней лесов, полей и рек!
Я другой такой страны не знаю,
Где так вольно дышит человек.
В.И. Лебедев-Кумач

Shiroka strana moya rodnaya,
Mnogo v ney lesov, poley i rek!
Ya drougoy takoy stranea ne znau,
Gde tak volieno deashit chelovek.
V.I. Lebedev-Koumach

She walks in beauty, like the night
Of cloudless climes and starry skies;
And all that's best of dark and bright
Meet in her aspect and her eyes;
Thus mellowed to that tender light
Which heaven to gaudy day denies.
G.G. Byron

Ше увалкнс ин быютй, лике тфе нигхт
Оф цлудлесс цлимес анд старрй скъс;
Анд алл тфат'с бест оф дарк анд бригхт
Меет ин хер аспект анд хер ейес;
Тфюс меллоувед то тфат тендер лигхт
Увхич хывен то гаюдй дай деньс.
Г.Г. Байрон