

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 003.324.1.035

А. А. Володина, С. В. Ёлкин, Э. С. Клышинский

Система автоматической транскрипции имен собственных на китайский язык

Рассматривается проблема автоматической транскрипции имен собственных на китайский язык. Из первичного языкового материала извлечены правила соответствия отдельных букв, буквосочетаний и слогов русского языка китайским слогам и составлена таблица. Намечены перспективы развития системы автоматической транскрипции имен собственных.*

Укрепление культурных связей между государствами среди прочего ставит проблему межъязыкового взаимодействия. Одной из задач в этом случае становится корректная передача имен собственных из одного языка в другой.

Когда речь идет о переводе на китайский язык, ситуация осложняется тем, что почти все слоги китайского языка — значимые слова. Есть некоторое количество слогов, которые используются только для того, чтобы передавать фонетику в именах и заимствованиях, но подобных слогов очень мало. Кроме того, в китайском языке существуют иероглифы с различным написанием, но одинаковым звучанием. Следовательно, даже идя по пути фонетического перевода, приходится решать, какой из иероглифов с таким звучанием выбрать — а это уже, как правило, бывает обусловлено их значением.

Теоретически есть несколько способов перевода имен собственных. Первый (и самый распространенный) — “фонетический перевод”. Слово переводится наиболее близким фонетическим аналогом, возможным в языке, на который делается перевод. Второй — “смысловой перевод”, т. е. перевод значащей части слова (если таковая есть). Этот способ наиболее распространен при переводе топонимов (*река Черная, гора Белая*). Также он может быть оправдан при переводе литературных произведений, где “говорящие” фамилии и названия используются как художественный прием.

При передаче имен собственных на китайский язык используется комбинированный подход. С одной стороны, имена передаются с помощью поиска фонетически близкого аналога. С другой стороны, при наличии нескольких равнозначных вариантов транскрипции, выбор производится по эмпирическим правилам, зачастую по личному отношению к носителю имени. Подобная окраска может приводить к качественному изменению отношения остальных носителей языка к данному человеку, что является, с нашей точки зрения, неприемлемым.

Таким образом, перевод имен собственных на китайский язык является творческим процессом. В связи с этим создание автоматической системы транскрипции на китайский язык оказало бы помощь как при оформлении официальных документов, так и при личных коммуникациях.

Отправной точкой для данной работы являлось построение правил для программы, которая переводила бы русские имена и фамилии на китайский язык. В качестве первичного материала были использованы фамилии, встретившиеся в книгах, переведенных с русского языка на китайский, из собрания ВГБИЛ им. Рудомино. В дальнейшем использовались фамилии и имена, взятые из баз, расположенных в сети Интернет.

Анализ материала показал, что для передачи русских фамилий на китайский язык используется перевод каждого слога на фонетически близкий китайский слог. Однако иногда “близость” эта довольно условна и перевод определяется скорее традицией. Например, звук “р” всегда передается звуком “л”.

При создании правил было выявлено, какими иероглифами передаются те или иные русские слоги. При этом (что вполне объяснимо, учитывая уже описанную специфику перевода на китайский язык) каждый слог может переводиться несколькими разными способами. На начальном этапе были выведены только соответствия русских слогов, китайским, без указания конкретного иероглифа, употребляющегося при записи. В дальнейшем было сделано сопоставление каждого слога языка-посредника, разработанного нашей группой для системы автоматической транскрипции “Транскриба” [1, 2], записи китайского слога латиницей. В качестве развития системы предусматривается сопоставление слогов китайского языка иероглифам, значение которых несет нейтральную или положительную окраску. При этом в случае отсутствия для некоторых слогов специального иероглифа, использующегося для заимствований и прочих фонетических

*Работа поддержана грантом РФФИ 03-01-00353.

соответствий, представляется целесообразным использовать сходные по прочтению иероглифы.

Для перевода русской фамилии на китайский язык прежде всего необходимо разбить ее на слоги. Разбиение осуществляется следующим образом:

слогом считается

- (1) CV,
- (2) V,
- (3) C,
- (4) (C)V_н,
- (5) (C)V_{нь},
- (6) (C)V_й,

где C — согласный, V — гласный.

Слоги (3), (4), (5), (6) выделяются только в том случае, если следующая буква — согласный. При этом согласный в (3) не может быть *н*, *нь* или *й*, если ему предшествует такой гласный, который может присоединить эту букву так, чтобы получился существующий в китайской фонетике слог.

Несколько соседствующих гласных могут передаваться одним дифтонгом. В противном случае первая гласная будет образовывать слог с предшествующим согласным, а остальные передаются самостоятельно.

Если фамилия заканчивается на указанные ниже буквосочетания, их следует переводить фиксированным образом. При этом оставшееся слово необходимо разбивать на слоги:

- в (иванов, петров, кожных) — fu,
- ский, ской (достоевский, слободской) — siqi,
- кий (збанацкий) — qi,
- ный (дыховничный) — nei.

Очевидна возникающая при таком разбиении проблема: если на вход поступает фамилия, не разбитая на слоги, в которой содержится звук “н”, то

может оказаться, что существует несколько способов ее разделения на слоги. Данная проблема не является значительной при переводе с русского языка на китайский, но при переводе с китайского языка на русский слово может быть искажено до неузнаваемости. Чтобы избежать такого искажения, на вход собственно этапа транскрипции имя передается уже разбитым на слоги. Но это достигается на других этапах работы с программой, в данной работе не рассматриваемых. В целях унификации транскрипции программа переводит слоги, идя по слову слева направо, и всякий раз выбирает самый длинный слог из возможных.

Иногда встречаются разночтения. Например, в некоторых случаях две одинаковые буквы подряд могут рассматриваться как одна или как две. Такие разночтения в нашей системе унифицируются — например, мы рассматриваем их как две разные буквы. Такой же унификации подвергаются все подобного рода спорные моменты.

Очевидно, что при таком подходе создаваемая программа не всегда будет переводить известные фамилии так же, как это уже принято, поэтому к программе может прилагаться список исключений, в который будут включены фамилии с традиционным переводом.

Ниже приведены таблицы соответствия слогов русского языка китайским. Данные таблицы были составлены А. А. Володиной на основе эмпирического материала в процессе работы над правилами транскрипции имен собственных с русского языка на китайский.

Таблицы соответствий для русского языка

По вертикали — инициаль (начало слога), по горизонтали — финаль (окончание слога)

	{ }	А	Е	Ё	И	О	У	Ы	Э	Ю	Я	Ь
{ }	{ }	a	ye	you	yi	wo	wu	e	e	yu	ya	—
Б	bo/bu	ba	bie	bie	bi	bo	bu	bo	ba	biyu	biya	bi
В	fu	wa	wei	wei	wei	fu	wu	wo	wei	wu	wa	fu
Г	ge	ge/gao	ge	ge	ji	guo/ge	gu	ge	ge	ju	jia	ji
Д	de	da	jie	jie	ji	duo	du	de	de	jiu	jia	ji
Ж	re	zha	re	re	ri	ruo	ru	ri	re	ru	zha	ri
З	zi	za	ze	ze	zi	zuo	zu	zi	ze	jiu	jia	ji
Й	yi	ya	ye	you	yi	yuo	yu	yi	ye	yu	ya	—
К	ke	ka	ke	ke	ji	ke	ku	ke	ke	qiu	qiya	qi
Л	er	la	lie	lie	li	luo	lu	le	le	liu	lia	li/er
М	mu	ma	mie	mie	mi	mo	mu	mo	mei	miu	miya	mi
Н	ne	na	nie	nie	ni	nuo	nu	ne	na	niu	niya	ni
П	pu	pa	pie	pie	pi	bo	bu	bo	bei	bu	piao	bi
Р	er	la	lie	lie	li	luo	lu	le	le	liu	lia	er
С	si	sa	xie	xie	xi	suo	su	si	se	xiu	xia	si
Т	te	ta	tie	tie	ti	tuo	tu	te	te	jiu	jia	qi
Ф	fu	fa	fei	fei	fei	fu	fu	fei	fei	feiya	feiya	fei
Х	he	ha	he	he	xi	huo	hu	he	he	xiu	xia	he
Ц	ci	ca	ce	ce	ci	cuo	cu	ci	ce	cu	ca	—
Ч	qi	cha	qie	qie	qi/chi	chao	chu	chi	che	chu	cha	qi
Ш	shi	sha	she	she	shi	shuo	shu	shi	she	shu	sha	shi
Щ	shi	sha	xie	xie	shi	shuo	shu	shi	she	shu	sha	shi

	АНЬ	АН	ЕНЬ/ ЭНЬ	ЕН/ЭН/ ЫН	ЁН	ИН/ИНЬ	УН/ОН	ЮН	ЯН	АЙ/ОЙ	АУ/АС	ЕЙ/ЭЙ	ОУ	ЫЙ	УЙ	ЮЙ
{}	an	an	yin	yin	youne	yin	yong	yong	yang	ai	ao	ei	ou	ei	yu	yu
Б	ban	ban	ben	ben	ben	bing	beng	biyun	bian	bai	bao	bei	bowu	bei	buyi	biy u
В	wan	wan	wen	wen	wen	wen	weng	wune	wan	wai	fuao	wei	wow u	wei	wuyi	wu yu
Г	gan	gan	gen	geng	geng	gen	gong	jun	jian	gai	gao	gei	gou	gei	guyi	ju
Д	dan	dan	den	deng	deng	din	dong	jun	jian	dai	dao	dei	dou	dei	dui	jiu
Ж	ran	ran	ren	ren	reng	Rine/ri ni	rong	run	zhan	zhai	rao	reyi	rou	riyi	rui	rui
З	zan	zan	zen	zen	zen	Zine/zi ni	zong	jun	jian	zai	zao	zei	zou	zei	zui	zui
Й	yang	yang	yan	yan	youne	yin	yong	yun	yan	Yiai	yao	yeyi	you	yiyi	yu	yu
К	kan	kan	ken	ken	ken	Jine/jini	kong	qun	qian	kai	kao	keyi	kou	keyi	kui	kui
Л	lan	lan	leng	leng	liene	lin	long	lun	lian	lai	lao	lie	lou	lei	luyi	lü
М	man	man	man	man	man	ming	Mune/ mone	miune	mian	mai	mao	Mei/ mei	mou	mei	muyi	miy u
Н	nan	nan	nian	nian	nian	nin	nong	niune	nian	nai	nao	nei	nou	nei	nuyi	nü
П	ban	ban	ben	beng	beng	bing	beng	pune	pian	bai	bao	bei	pou	bei	buyi	buy i
Р	lan	lan	leng	leng	liene	lin	long	lun	lian	lai	lao	lei	lou	lei	luyi	lü
С	sang	sang	shen	sheng	xiong	xin	song	xun	xian	sai	sao	seyi	sou	seyi	sui	xu
Т	tan	tan	den	deng	qiong	jun	tong	qun	qian	tai	dao	dei	dou	dei	tui	tui
Ф	fan	fan	fen	fen	fen	Fuying/ fuyin	feng	fen	fan	fei	Fawu /fawo	fei	fou	fei	fuyi	fuyi
Х	han	han	hen	heng	heng	xing	hong	xun	xian	hai	hao	hei	hou	hei	hui	xue
Ц	can	can	cen	ceng	ceng	qin	cong	cun	can	cai	cao	ceyi	cou	ceyi	cui	cui
Ч	chan	chan	chen	chen	chen	qin	chun	chun	chan	chai	chao	cheyi	chou	cheyi	chui	chu i
Ш	shan	shan	shen	sheng	sheng	Shine/ shini	sheng	shun	shan	shai	shao	shei	shou	shei	shui	shu i
Щ	shan	shan	shen	sheng	sheng	Shine/ shini	sheng	shun	shan	shai	shao	shei	shou	shei	shui	shu i

СПИСОК ЛИТЕРАТУРЫ

1. Бондаренко А. В. и др. Автоматизация процесса транскрипции для задачи многоязыковой передачи имен собственных // Известия Академии наук. Теория и системы управления. — 2004. — Т. 43. — № 6. — С. 892–898.

2. Бондаренко А. В. и др. Метод фонетической транскрипции с использованием единого промежуточного фонетического представления: Препринт / Ин-т прикладной математики им. М. В. Келдыша РАН. — М., 2003, № 90. — 28 с.

Материал поступил в редакцию 20.05.05