

Система автоматической транскрипции имен собственных на китайский язык¹

Володина А. А.
kliaksa-n@yandex.ru

Ёлкин С.В.
elkin_serg@mail.ru

Клышинский Э.С.
klyshinsky@mail.ru

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. Задача осложняется тем, что 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”, то есть перевод значащей части слова (если таковая есть). Второй 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 литературный 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. При этом иногда “близость” эта довольно условна и перевод определяется скорее традицией. Например, звук “р” всегда передается звуком “л”.

При создании правил были вычленены соответствия, 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 проведены соответствия только для русских слогов соответствующим китайским слогам, без указания конкретного иероглифа, употребляющегося при записи. В дальнейшем было сделано сопоставление каждого слога языка-посредника, разработанного нашей группой для системы автоматической транскрипции «Трансскриба» [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

¹ Работа поддержана грантом РФФИ 03-01-00353

случае отсутствия для некоторых слогов специального иероглифа, использующегося для заимствований и прочих фонетических соответствий, представляется целесообразным использовать сходные по прочтению иероглифы.

Для перевода русской фамилии на китайский язык, первым делом необходимо разбить ее на слоги. Разбиение осуществляется следующим образом:

слогом считается

(1) CV,

(2) V,

(3) C,

(4) (C)Vн,

(5) (C)Vнь,

(6) (C)Vй.

(где C – согласный, V – гласный)

При этом 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 необходимо переводить фиксированным образом. При этом оставшееся слово необходимо разбивать на слоги.

- в (иванов, петров, кожухов) – 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. Для того 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 эти фамилии с традиционным переводом.

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

Литература

1. Бондаренко А.В. и др. Автоматизация процесса транскрипции для задачи многоязыковой передачи имен собственных / Известия Академии Наук. Теория и системы управления. Том 43, N. 6, 2004. сс. 892-898
2. Бондаренко А.В. и др. Метод фонетической транскрипции с использованием единого промежуточного фонетического представления / Препринт Института Прикладной Математики им Келдыша М.В. РАН, №90, 2003 г. 28 с.

Таблица 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 | | | |
| Г | ge | ge/gao | ge | ge | ji | guo/g e | gu | ge | ge | jiu | jia | - |
| Д | de | da | jie | jie | ji | duo | du | de | de | jiu | jia | ji |
| Ж | re | zha | re | re | ri | ruo | ru | ri | re | - | - | ri |
| З | zi | za | ze | ze | zi | zuo | zu | zi | ze | jiu | jia | ji |
| Й | yi | ya | ye | - | yi | you | yu | - | 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 | | 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 | | | 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 | feiyu | 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 | - | - |
| 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 | - | shi |
| Щ | shi | sha | xie | xie | shi | shuo | shu | shi | she | shu | sha | shi |

Продолжение таблицы

| | АНЬ | АН | ЕНЬ/ ЭНЬ | ЕН/ ЭН/ ЫН | ЁН | ИН/ИН Ь | УН/ОН | ЮН | ЯН | АЙ/ ОЙ | АУ/АУ | ЕЙ/ ЭЙ | ОУ | ЫЙ | УЙ | ЮЙ |
|----|------|------|-------------|------------------|-------|------------|-------|------|------|-----------|-------|-------------|------|------|-----|------|
| {} | an | an | yin | yin | | yin | yong | yong | yang | ai | ao | ei | ou | ei | | yu |
| Б | ban | ban | ben | ben | | bing | beng | | | bai | bao | bei | | bei | | |
| В | wan | wan | wen | wen | | wen | weng | | | wai | | wei | | wei | | |
| Г | gan | gan | gen | geng | | gen | gong | | | gai | gao | gei | gou | gei | | |
| Д | dan | dan | den | deng | | din | dong | | | dai | dao | dei | dou | dei | dui | |
| Ж | ran | ran | ren | ren | | - | rong | | | - | rao | - | rou | - | rui | |
| З | zan | zan | zen | zen | | - | zong | | | zai | zao | zei | zou | zei | zui | |
| Й | yang | yang | yan | yan | | yin | yong | | | - | yao | - | you | - | | |
| К | kan | kan | ken | ken | | - | kong | | | kai | kao | - | | - | kui | |
| Л | lan | lan | leng | leng | | lin | long | | | lai | lao | lie | lou | lei | | lü |
| М | man | man | man | man | | ming | - | | | mai | mao | Mei/ mei | mou | mei | | |
| Н | nan | nan | nian | nian | | nin | nong | | | nai | nao | nei | - | nei | | nü |
| П | ban | ban | ben | beng | | bing | beng | | | bai | bao | bei | pou | bei | | |
| Р | lan | lan | leng | leng | | lin | long | | | lai | lao | lei | lou | lei | | lü |
| С | sang | sang | shen | sheng | xiong | xin | song | xun | | sai | sao | - | sou | - | sui | xu |
| Т | tan | tan | den | deng | qiong | jun | tong | | | tai | dao | dei | dou | dei | tui | |
| Ф | fan | fan | fen | fen | | - | feng | | | fei | | fei | fou | fei | | |
| Х | han | han | hen | heng | | xing | hong | | | hai | hao | hei | hou | hei | hui | |
| Ц | can | can | cen | ceng | | qin | cong | | | cai | cao | - | cou | - | cui | |
| Ч | chan | chan | chen | chen | | - | chun | | | chai | chao | - | | - | | |
| Ш | shan | shan | shen | sheng | | - | sheng | | | shai | shao | shei | shou | shei | | shui |
| Щ | shan | shan | shen | sheng | | - | sheng | | | shai | shao | shei | shou | shei | | shui |