An 18th Century English-Chinese Dictionary[*]

(unpublished manuscript)

Xuan Di
di.xuan@yahoo.com

The present paper includes results of my study on a manuscript *Ying-ji-li-Guo Yi-Yu* (<<口英咭口利国译语>>, henceforth YY, approximately translated as *Wordlist of the English*). In 1931, Walter Fuchs reported the discovery of this dictionary where he considered the dictionary to have been compiled by a Chinese who spoke poor English and made frequent mistakes. However, such an assumption faces two problems: (1) is the historical period possible to license the existence of such a bilingual Chinese? (2) YY is a dictionary among a series of dictionaries compiled after the same decree, its entries are one of the standard wordlists. However, the compilation process hypothesized by Fuchs (1931) doesn't confirm to the standards of official collections in general, according to which, compilations are not done by a single person. The nature of the languages recorded and used in the bilingual dictionary cannot be answered by a brief glance through the pages. It is a question that must be answered after scrutiny on the entries themselves. The written forms of the 'English' that are manifested in the manuscripts provide partial information; more cues are provided by the phonological system in the Chinese dialect that is used in transcribing the foreign input.

Besides solving the puzzle of the manuscript, the purpose of this paper is also to use this manuscript as a probe to investigate a method that disentangle strands of a complex linguistic system that contains factors from both historical change and language contact.

The organization of the paper is as the following: firstly, historical facts and other compilation related information are presented in section 2; matrix-based methodology and results in analyzing the Chinese transcription of English and Portuguese entries are explained in section 3. Section 4 presents the conclusion: the

dictionary is compiled by a Portuguese second language learners of English who provide a source "English" (a mixture of English and Portuguese), a Chinese officer who speaks a higher register of the then Mandarin and uses this register of his to transcribe the "English" he heard, and a scribe who doesn't know English at all and made some mechanical mistakes. Section 5 points out further studies that need be done on the manuscript.

2. Philological Background
2.1. Basics about the dictionary

Ying-ji-li-Guo-Yi-Yu (YY) is a bilingual dictionary/glossary literally translated as "English Country Translated Language". In the manuscript, there are 743 word entries. YY is more of a glossary than a dictionary, because its entries don't have illustrating sentences. When there are more than one English words that can translate into a given Chinese word (e.g. *gao*^{55} 'tall, high'), only one is recorded; vice versa, every English word/phrase (whether it is ambiguous or not, e.g. *bank*) maps onto one Chinese word/phrase.

The entries of YY conform to other bilingual glossaries in the 18th century. A typical entry has three sections: one for the written form of the source language, an English word (1a); the second section has a Chinese word of the same meaning (1b); the third part contains transcription of the pronunciation of the source language, recorded in Chinese characters (1c).

a. "English"	(1)	*Sea*	(2)	*Provider*
b. Chinese		海		都
				綱
c. transcription		洗		多未羅播

The first line is the written form of the foreign language that is being recorded. Here in (1) is the 45th word in the glossary/dictionary. In the first part, there is the English word *sea* (1a). In the second part of the entry, the Chinese character 海 corresponds to the sound /hai^{214}/ that means "the expanse of salt water that covers most of the earth's surface and surrounds its landmasses" (1b). The third section of the entry is, at least it is meant to be, the pronunciation of the English *sea*. The Chinese character 洗 is pronounced as /ɕi^{214}/ in Present Day Standard Mandarin[1].

When there are more than one characters in the Chinese word (2b), the convention is to write them vertically, the same rule as writing any Chinese sentence at that time. In the transcription, the characters are written in one line horizontally, and they are read from the right to the left. So in (2c), the four characters are 多未羅

---

[1] When ignoring the tone, a fricative followed by a high front unrounded vowel sounds very close to /si:/ in the ear of a present-day Northern Mandarin speaker. The exact mechanism that induces two different syllables is outside the scope of the present paper: in part, it requires careful laboratory work; what's more, the exact quality of speakers two hundred year's ago is difficult to reconstruct.

播, read from right to left in Modern Mandarin, we get: 播 /po/羅 /luo/ 未/vei/多 /tuo/[2].

In the convention of the compilation of Chinese Foreign dictionaries, the transcriber is required to find the character that has the closest sound in the recording dialect, the dialect that is used to transcribe the foreign sounds. In other bilingual dictionaries of the same series, efforts have been paid to spell CV combinations by *fan-qie*, a common method in notating the sound of low frequency words in ancient China. For instance, to represent a syllable X_{CaVb} , two characters A_{CaVa}, B_{CbVb} are used. The reader shall take the onset of the sound represented by the first character, and link it to the rhyme of the second character. In YY, unlike its sister glossaries, *fan-qie* is not implemented.

To conduct the research, I idealize the transcriber as someone who has done his best in finding the best matching characters in his dialect to record the sound he heard. In §2.2 I will present arguments for assuming such an oral-based compilation process.

2.2 compilation process for bilingual dictionaries in general
2.2.1 classification

YY is not the only glossary that has been compiled in the history of China. It is one of the 66 Chinese Foreign Glossaries (*Hua-Yi-Yi-Yu* <<华夷译语>>) that are edited by the central governments in the Ming and Qing dynasties[3] (Feng 1981).

According to Feng (1981), the Chinese-Foreign Glossaries are classified into four types. Type I, one glossary, edited in 1382 by Huo-yuan-jie (火源洁) and Ma-yi-chi-he (马懿赤黑). Printed in 1389, this is a recording of a Mongol variety. The Mongol-Chinese glossary has Chinese words and the Chinese transliteration of the Mongol, without the written form of Mongol. Type II, 10 glossaries, edited by the Bureau of Translation (Si-yi-Guan, set up in 1407), an institute under the Board of Rites (*Li-bu*) in the Ming dynasty (1368-1644). This type of "glossaries" has an appendix—memorials (*Lai-wen*), the official documents from the states whose languages are recorded. The non-Chinese include Tartar, Burmese, Tibetan, etc. Type III, edited by the Bureau of Interpreters (Tong-wen Guan) in the Ming dynasty. These glossaries cover 13 languages, such as Japanese, Korean, Vietnamese, Mongol, Tartar, etc. The fourth type of glossaries is edited in the Qing dynasty. The 42 Chinese-Foreign glossaries are compiled in 1748 when the Bureau of Translators and the Bureau of Interpreters were merged. Most of this set of glossaries are finished around

---

[2] Whether consonant clusters are represented by words that shares similar rhyme remain to be confirmed by further studies.

[3] *Hua-Yi-Yi-Yu* in its narrow sense refers to the glossary edited in 1382. Here I am following Feng (1981) in adopting its broad sense. The name for this set of glossaries has been translated as "Chinese-Barbarian Glossary" (Wild 1943), or "Chinese-barbarian-translated-language' (Bolton 2003). Because the royal nationality in the Qing dynasty was nomad, the Manchus were sensitive to the word *yi* 'barbarian'. The Bureau of Translators was then renamed from "the Barbarian Language Hall" (literally) into "the Translation Hall" (literally). As the YY is compiled after the decree of a Manchu Emperor, a less offending word 'foreign' is used in translating the name of the series to which YY belongs.

1750 (Zhu et al. 1992). Among the type IV glossaries, there are six European languages (English, French, German, Italian, Latin and Portuguese) and nine Tibeto-Burman languages (including *the Glossary of Bai-Ma*); all of which are relevant to the following discussion[4].

2.2.2 The decree

YY, like the other 41 glossaries, is compiled upon the mandate of Emperor Qian-Long (1735-1796). The decree on 20th June, 1748 (*Qing-Shi-Lu: Gao-Zong-Shi-Lu* vol. 324) serves as a historical evidence for the compilation of the type IV Chinese-Foreign Glossaries.

Text of the decree in the first half of the ninth month of Emperor Qian-Long (translated by Fuchs (1931), with my modification, numbering and high-lighting):

> "Previously edited glossaries should be searched for and corrections should be made. Each of them must, **[1] in accordance with the rules applied to the Tibetan collection, be re-examined, divided into categories and be compiled as part of the series**. … **[2] All European glossaries may be edited in Xian-an-Gong Palace for convenience**. … Furthermore, with reference to **[3] the foreigners beyond the border, as well as the Inner Tibetan** … any existing written forms of their languages shall be collected; then in accordance with the form and arrangement of the Tibetan part, the sound and the meaning of the words shall be taken and explained by Chinese characters placed under the native expression. Then, **[4] all shall be copied**, presented to the Throne and handed over to the (Si-Yi-) Guan for examination. **[5] By this means, We shall manifest Our prosperous administration.**"

Five points could be drawn from the decree.

[1] Since the compilation of the Chinese-Foreign Glossaries has always been governmental behaviour[5], the data collectors who are low ranking officers don't have

---

[4] People who are not familiar with the history may confuse two kinds of glossaries which are very different in nature. The Chinese Foreign Glossaries (*Yi-yu*) and the 'foreign language books' (*fan-hua*). The collection of the Chinese-Foreign glossary is edited by officers in the royal court, serving political use, for the convenience of dealing with foreign affairs, carefully hand-written manuscripts. The editing of *Fan-hua*, on the other hand, is for commercial use, printed by private institutes, often without any foreign written forms. For example, if a *fan-hua* were to have an entry for the *sea*, it would only have (1b) and (1c). The most confusing ones to our YY are several versions of the Red Hair Common Language. According to Zhou (2003), these includes an earliest 'Hong-mao Mai-mai Tong-yong Gui-hua' (Red Hair Buying Selling Commoon Ghost-language), reported by Samuel Wells Williams in 1834 (a version printed in Fo Shan, Guang Dong, unknown date, lost); a 'Hong-mao Hua' (Red-hair Language), reported by Short when cataloguing for a library in the Land Brandenburg in 1840 (lost); a 'Hong-mao Tong-yong Fan-hua' (The Red Hair's Common Foreign Language) kept in the British Library, printing time unknown, approximated by the library as in 1850 (2 vol.); and a 'Hong-mao Fan-hua' (Red Hair Foreign Language), kept in Taiwan, printed in the late Qing Dynasty.

[5] There is a potential confusion between two kinds of glossaries which are different in nature. The Chinese Foreign Glossaries and the "foreign language books" (Fan-hua). The collection of the Chinese-foreign glossaries is edited by the royal court, for the convenience of dealing with foreign affairs. The

the right to decide what should be included in the word list. These linguists went to do their fieldwork with officially predetermined word lists. For these 42 glossaries, they should be based on the Glossary of Lhasa Tibetan, decided by someone in a higher position than the editors.

However, in practise, the wordlists are not exactly the same across Type IV glossaries. Although all type four glossaries are edited after the decree, the numbers of lexical entries (the 'word list') are different. Among the six European languages, five of them (French, German, Italian, Latin and Portuguese) bear the heading "Xi-Yang-Guan" (The Western Countries Institute). These glossaries are, as ordered by the decree, "in accordance with" the Glossary of Lhasa Tibetan, which has the heading "Xi-Fan-Guan-Yi-Yu" (the Western Minority Institute Translated Language). They contain 2103 entries, and they are the largest glossaries among all type four glossaries, sharing the same word list, i.e. same Chinese word entries in the same sequence.

Whereas the English glossary uses a much smaller word list, only 740 words, if it were intact. The same word list is used by the other nine dialects/languages in the Tibetan area in the Qing dynasty (present day Tibet, Sichuan and Qinghai provinces).

Why is English so differently treated from the other European languages?

The answer in part can find some clue from the decree itself.

[2] This sentence suggests that the five European glossaries are edited in Beijing. In mid-Qing, there are missionaries from the then important European countries. They are masters of Latin, as well as their native languages. Therefore, it is natural, as Fuchs proposed, to summon those missionaries in Beijing and have them work with Chinese officers to finish the glossary editing without field-workers' traveling afar. Although Latin is not a native language for any missionaries, it is a dominant language for them to communicate with each other, also to communicate with the Chinese before they learned Chinese (Fang 1987: 957-67)[6].

England, however, hadn't had an established power to initiate any informal or official trade with the Qing Dynasty; neither had it sent any (English-speaking) missionaries before Morrison (1782-1834) (Coblin 2003). Presumably, the Qing Dynasty didn't want to reach out to a peripheral island for linguistic data collecting, especially when the motivation for collecting the data was politically driven: to serve the need in foreign affairs. Collecting English for the Qing Dynasty officers was

---

compilation of the latter is for the purpose of commercial use, printed (in general) by private institutes. The most confusing ones to our YY are several version of the Red Hair Common language (Zhou 2003), which includes an earliest unofficial glossary 'Hong-mao Mai-mai Tong-yong Gui-hua' (Red-Hair Buying-Selling Common Ghost Language), reported by Samuel Wells Williams in 1834 (a version printed in Fo Shan, Guang Dong, unknown date, lost); a "Hong-mao Hua" (Red-Hair Language), reported by Short when cataloguing for a library in the Land Brandenburg in 1840, now lost; a "Hong-mao Tong-yong Fan-hua" (The Red Hair's Common Foreign Language) kept in the British Library, date unknown, approximated by the library as in 1850, (2 volumes); and a "Hong-Mao Fan-hua" (Red Hair Foreign Language), kept in Taiwan, printed in the late Qing dynasty (19th century).

[6] The 'European Hall' (Xi-Yang Guan) in the Bureau of Translators set up in 1792, teaching Latin and Russian to Chinese students by Russian missionaries; and teaching Chinese to Russian students. In 1732, most of the students in the Latin class could use Latin satisfactorily (Fang 1987).

probably something that will not be useful in a foreseeable future (Earl George Macartney was not going to arrive in China until more than forty years later (Eames 1914: 48); thus if there are any mistakes, nobody would ever have figured it out. Therefore, English was treated as some 'remote' minority language and grouped with the non-Lhasa Tibetan glossaries.

If a foreign missionary claimed that he could speak and write a language called English[7], the linguist officers might have been willing to add a language to meet the order [3] in the decree.

[3] "Foreigners beyond the border" might be realized as English, since all other languages in type four glossaries are mentioned in the decree.

[4] The YY we now see is not the original version(s) that are used in the editing. Because all the manuscripts will be copied again by scribes.

[5] The glossaries are initially to serve political purposes. Where the international relation is not established, quality of the book doesn't affect any practical value. This makes it understandable that the scribe English glossary paid less attention to the content than to the format of the glossary (c.f. §2.3).

2.2.3 Historic cues on the compilation procedure

As far as I am aware of, no historical record has been found on the compiling procedure. A general assumption regarding to the process is[8]: there is a data collector, might be an officer in the Bureau of Translators (or from the Bureau of Interpreters) who is literate in Chinese[9]. And the Chinese transcriptions of the target foreign languages/dialects are based on what the officer *heard* from his informants.

This assumption is based on the translation tradition of the Buddhist cannon. In the 7th century, when there was a need for high quality translation of Buddhist cannon, Xuanzang (or Hiuen Tsiang, 596-664) recorded an "Eight Steps" translation process (Wang 1997). The first three steps are: (1) one person chants the Sanskrit (or Pali), (2) a different person interprets it, (3) a third expert records the interpretation on of the second; plus another five steps in checking and securing the quality of the translation.

---

[7] Some years' later, my personal experience, when talking with many non-English speaking people from Europe, seems to support this hypothesis at a rather trivial level. The undergraduates in Holland and Norway often consider themselves as 'capable of speaking English', while their 'English' is often the English pronunciation of a Dutch/Norwegian word: such as using 'house-work' when meaning to say 'homework'.

[8] This is based on the discussion in 2003 with Professor Nie Hongyin and Dr. Sun Bojun (CASS).

[9] The Chinese have a tradition of editing monolingual dictionary. Whoever is an officer in the government should in principle be a bachelor, which implies that they know how to use a monolingual dictionary and familiar with the method in spelling-out an unfamiliar word by *fan-qie*, or by a homophone, or by using another word of the same syllable but of different tones (for instance, if this method were used in English, a dictionary might denote the noun *permit*, they may say 'read as the verb *permit*, stress on the second syllable').

A simplified version of this method is still used in the early Qing Dynasty, when translating the western technological works (Wang 2000). Since translation and glossary editing are both bilingual data processing, one might borrow the 'read-out' convention from translation and apply it to the other.

2.2.4 Evidence from another glossary for multiple editors

Because YY shares the same wordlist with the nine Tibetan glossaries, to seek for cues in its compilation, we may gather some hints from these glossaries. We can take an example from the Glossary of Bai-Ma (Nishida and Sun 1990), a sister language to Tibetan, daughter of Himalayish, Tibeto-Burman. There is an ideal case for the oral-oriented compilation process.

The table below illustrates the 383rd and 387th word entries in the Glossary of Bai-Ma. For the 383rd entry, the Chinese written form is 旨, whose reconstructed pronunciation is */ tʂi / (Li and Zhou 1998); it means 'the decree' in English. The Chinese transcription for the pronunciation in Bai-Ma is 少稍, the reconstructed sound of which is */ ʂiau ʂau /. The Roman transliteration for the Tibetan Devanageri of *decree* is *Rgyal-hu-sag-mkhan-reg*. The phonetic representation [ʑi^{314} ndɛ^{53}] is based on the fieldwork by Nishida and Sun.

The two Chinese entries are homophones, although their Bai-ma are not homophones and are written differently. The transliteration is the same two characters 少稍. An explanation that evokes minimal assumptions is that this is a mistake.

(3)

Chinese word Entry	English translation of the word	Chinese transcription of BM	Latin trans-literation of the BM written form	Corresponding form in modern oral speech
a. (383) 旨 */ tʂi/	the decree	少稍 */ ʂiau ʂau /	Rgyal-hu-sag-mkhan-reg	[ʑi^{314} ndɛ^{53}]
b. (387) 纸 */ tʂi/	Paper	少稍 */ ʂiau ʂau /	Shor-shob-reg	[ʃɔ^{13} ʃuɛ^{53}]

Such a mistake could be made in the process of an oral interpretation process. One possibility is that the informant of Bai-ma didn't read the Chinese characters, for had he read it, he would have been able to know the meaning of the entry from the written form. Recalling the Buddhist cannon style of translation, if the Chinese entry is being read to the Bai-ma informant, when the context is not clear, he might fail to disambiguate the meanings of homophone and produce the same word to be transcribed. However, it is also possible that the oral informant didn't make a mistake. Here in this case, both 383rd word and the 387th word are under the religion section, where 'decree' could refer to the order/testimony from the God, whereas 'paper' could be associated with 'paper that is used in recording the sutra'. In this case, they may use the same word to represent two semantically close words. Whether the oral informant made a mistake or not, it is clear that the written form informant is not the

same person as the oral informant, because the written form informant distinguished the two words.

What is clear from this mistake is that the Chinese transcription is not based on the written form. The officer based on speech when transcribing the non-Chinese languages. Later in §2.3.2, we can see similar type of mistakes occurring in YY.

2.2.5 Sum up

The compilation is a complex process in editing these Type IV Foreign-Chinese dictionaries. Each of the glossaries is not edited by a single person, or by one informant with one data-collecting officer. There are at least a written form informant, an oral informant, and a Chinese officer. The Chinese transcription is based on oral input, not on the written form of the foreign language. When the glossary has been collected, a scribe makes a copy, which is the manuscript that we found in the Forbidden City.

2.3 the compilation of YY
2.3.1 the contents of YY

YY consists of two volumes. In the remaining manuscript, there are 734 words, subcategorized into twenty categories (astronomy, geography, weather, persons, parts of the body, architecture, implements, food and drink, clothing, colours, religion, humanity, location, plants, animals, jewels, fragrance and herb, numerals, human activities and miscellany).

On the first page of each category there is the name of the category as well as two entries. In the body of each category there are four entries on each page. The first page(s) of Volume One is (are) missing; on the first page of Volume Two, there is the title *Ying-ji-li-Guo Yi-yu*. However, unlike the other five European bilingual glossaries, there isn't the name of the editing institutes. Fuchs (1931) and Feng (1981) suggest that the Type IV glossaries started editing in 1748. Zhu J. et al. (1992) states that the compilation has finished around 1750.

Apart from the missing first page(s), the rest of YY consists the same wordlist as the Tibetan dialects. Therefore, I estimate that there is only one page (two entries) missing. The compilation originally contains 740 words.

2.3.2 Support for pronunciation oriented compilation from another glossary

In §2.2.4, I present data that supports the oral-oriented practices in providing transcription. The transcription is based on an oral-informant who didn't even read the written forms in the entry.

The fact that more than three dozen entries have mismatched written English and transcription suggests that this assumption is ready. A typical instance is like the 575th word entry (4). This word is in the 'category of human activities' (*Ren-shi Men*), where entries are all about the behaviour and manner of a human being. The Chinese word is 迟, which means '(to do something) slowly' in the context. The entry is next to an entry meaning '(to do something) quickly'. However, this Chinese character can also represent the meaning 'to be late'. In (4) we can see that the English

reconstructed from the transcription is *too late*, the written form is *Very Slowly*. Both are acceptable for translating 迟, and they are not different in terms of registers.

(4)

Chinese word Entry	English written form in the entry	Chinese transcrip-tion of English	Reconstructed English from the Chinese transcription
(#575) 迟 late	Very Slowly	都 列 du lie	*too late

In other instances of such mismatch, the two English phrases/words are not always completely different. Sometimes, the transcription doesn't have the article that showed up in the written form; sometimes, the transcriptions contain more words than that in the written form.

This suggests that the transcriptions are not based on the written forms. The same speech-oriented practice in other bilingual dictionaries is also applied in transcribing English in YY.

2.3.3 Cues for the final copy process by a scribe

Although the wordlists are of the same sets of words as those in the Tibetan dialects, the order of the entries inside each category has been rearranged in YY. Other glossaries are ordered purely according to semantic relations among the entries, whereas in YY, the entries inside each category are arranged according to the morpheme number of the Chinese word.

This rearrangement is probably not done by the editors, rather, it is altered due to the likings of the scribe. Take the 40th and the 80th word entries[10] for example:

(5) #40   (6) #80

a. "English" *Good Time*   *Good Time*
b. Chinese   水 water   世
            界 the world
c. transcription 達挖   墨低多個
pinyin of the transcriptions wa da ge duo di mo

According to the Tibetan glossaries, the 40th entry is *the world*, and the entry for *water* follows that for *the world*; it is the 45th entry. However, in YY, the first time where there is a need to rearrange the entries to meet the 'new order', the scribe made a mistake. What he should have done is to order the entries according to the number of morphemes of each Chinese word; so he should take the 45th entry *water* and insert it before the entry for *world*. However, he copied the English written form for *the world* (which for reason unknown to me is *Good Time*), and then he seemed to realize that he needed to make a new order, so he copied on the Chinese character for *water* down, as well as the Chinese transcription for *water* (5). Then he went on with his

---

[10] Numbering of entries in YY all include the missing two entries, to make the matching among glossaries easier.

copying of the other monomorphemic entries and then started his bimorphemic entries for the *Geography Area* with the 'right' entry for *the world*.

2.3.4 Sum up

The compiling process for YY is as complex as other bilingual dictionaries/glossaries. The Chinese transcription is based on something uttered by a speech-informant; it is not a transliteration of the English written form. Some mistakes found in YY are made by the scribe, not during data-collecting.

2.4 Assumptions on the compilation

Both historical and philological facts suggest that the Chinese transcription is based on oral speech. In YY, to figure out the Chinese dialect that is used in transcribing the 'English' must base the reconstruction on the phonological mapping. It is rational to be confident in believing that the mapping between the phonological system of the transcribing dialect and the system of the oral informant's phonological system can be figured out by a scientific method.

In §3, I describe the methods used in figuring out the nature of the phonological system of the transcribing Chinese dialect and the nature of the phonological system of the speech-informant.

3. Methodology

The basic method is to using historical cues facilitate the solution of a six-row matrix, which consists of six rows of related phonological systems (§3.3).

3.1 the unknowns and the assumptions

At the bird-eye view of YY, the system recorded is an 'English' that is provided by a Portuguese (§3.2). In phonology-based study, schematically speaking, what we can get from the manuscript is over 700 words from Language X, being transcribed into a language Y by some three thousand symbols. It is not transparent from the manuscript what the phonological system of X is. At the outset, it seems that we have two variables but only one equation.

(7) $X_{\{x1, x2, x3,...xN\}} \longleftrightarrow Y_{\{y1, y2, y3,...yN\}}$

However, with five assumptions it is not impossible to solve the puzzle. Assumptions:

(a) Every language is a system; its phonology is systematic;

(b) Each register of the idiolect of a person has a systematic phonology.

(c) Diachronic phonological change and synchronic phonological differences are rule based (i.e. systematic).

(d) Language transfer is rule based.

(e) One idealization: the transcribing officer made perfect matches between sounds, namely, his transference of his Chinese phonological system to the 'English' he perceived is rule based.

When actually working on the manuscript, the historical cues also help in limiting the range of languages/dialects under the study.

3.2 Portuguese speakers speaking English

Although it is titled as *English* glossary, almost a quarter of the entries are not English. Rather, they are words from Portuguese. For these Portuguese written forms, the transcription from the oral informant usually can be reconstructed into the corresponding Portuguese words/phrases.

It is not likely that the informants are English native speakers for two reasons. Firstly, as mentioned in §2, there isn't any historical record about any English person in the Qing Court before 1750. Secondly, if the informants were native English speakers, the motivation of including a considerately amount of Portuguese words is puzzling.

My BA dissertation has suggested that this could be a dictionary of Proto-Pidgin, a language that is more standard than the Pidgin English around the Chinese coast a century later. Because it is a ship/trade language, it carries some Portuguese words, as they were dominating the trade between Asia and Europe before the British came to power. As I mentioned in my thesis, this hypothesis is not unproblematic. Firstly, there are many low-frequency words about religion and political issues (domestic and foreign affairs); a tradesman's language is less likely to contain words on these subject matters. Secondly, YY is edited upon the order of the Qing Court, where a businessman can gain money but not much social recognition. Granting a tradesman the power to be an informant is not very likely. Besides, a hypothesis that this is Proto-Pidgin doesn't answer the question about the phonological system of the oral informant. The person could be a native Chinese speaker, whatever his dialect is, speaking a language that contains English and Portuguese lexicon; alternatively, it could be a Portuguese speaking the 'English' he had in mind. Even more wildly, one can guess that it could be someone with yet another native phonological background, being the informant.

Comparatively speaking, it is more difficult for history to create an environment that is ready for such a Chinese speaker rather than for a Portuguese speaker. The crucial factor that prevents a wild guess is the phonological systems that we have found underlying the transcription. The minimal assumption on the 'English' would be containing two linguistic elements: English and Portuguese, since it is more likely for a Portuguese to produce a hybridized 'English', I take the non-linguistic facts to suggest that the phonological system is a Portuguese based English.
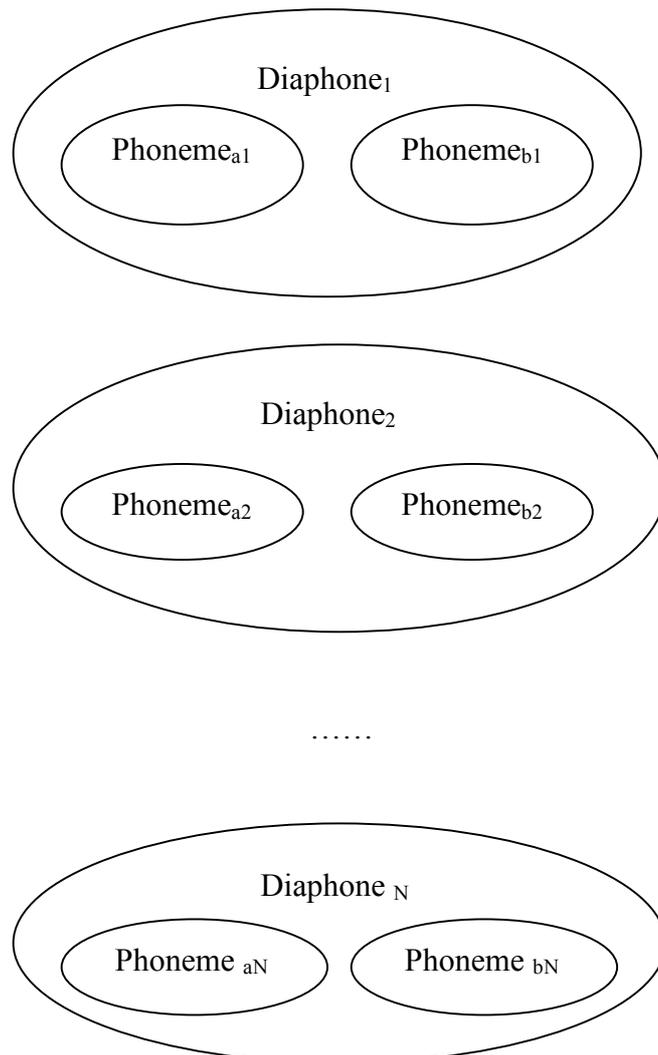
3.3 Diaphone and a phonological matrix

In this section, I introduce the notion of *diaphone*, which is important when talking about the system of transcription, a system that contains two dimensions. Then I show how the model is used in deducing the phonological systems that is being transcribed and is used in transcribing.

3.3.1 Diaphone

"Diaphone" was introduced by Johns (1934) when comparing the speech of one person with that of another. A diaphone refers to 'a sound used by one group of speakers together with the other sounds which replace it consistently in the pronunciation of other speakers'. The sound that is used by the oral informant is being consistently (by idealization) replaced by sounds in the transcriber's dialect. Johns also takes different phonemes in different registers spoken by a single person as diaphone. The second extension will also be a useful notion in the present work.

In the present paper, I modify Johns' definition as: *given the same phonological environment*, a phoneme$_a$ in one idiolect$_A$ is consistently used in transcribing another phoneme$_b$ in another idiolect $_B$. If the phoneme$_{a1}$ is the same as the phoneme$_{b1}$, we can say that the two members of a diaphone are identical; if phoneme$_{a1}$ is different from phoneme$_{b1}$, we can say that there are two different members of a diaphone$_1$. In other words, diaphones are used to refer to two sets of phonemes, each set of which is used by a speaker of a language.

(8)

The diagram (8) illustrates the relation between a diaphone and a pair of phonemes. We can also formulate the set of diaphones by a diaphonic matrix. In the matrix, each row is a phonological system, one of the oral-informant, the other of the transcriber. The two vectors are of one dimension if the two phonological systems are underlyingly identical. The two vectors are two-dimension if the two systems are not the same.

3.3.2 A phonological matrix

The Chinese officer who conducted the transcription is most likely a monolingual, otherwise, had he known English he would have been able to detect the difference between the written form and his transcription.

Because he doesn't know English, knowing the meaning of the 'English' words from the Chinese entries doesn't help him in guessing the sound of the word. Therefore, we can assume that the phonemes abstracted from his transcription represent the actual phonemes he perceived. Namely, if a sound is not being perceived, he could not 'filling the blank' because he has a mental representation of the sound himself. A typical instance of diaphone is where a phoneme$_{ax1}$ in one idiolect$_{AX}$ is consistently used in transcribing another phoneme$_{by1}$ in another idiolect $_{BY}$. Collecting all the corresponding phonemes in the manuscript, we get a two-dimension matrix with the phonemes perceived [Phoneme $_{ax1}$, Phoneme $_{ax2}$, … , Phoneme $_{axN}$] and phonemes produced [Phoneme $_{by1}$, Phoneme $_{by2}$, … , Phoneme $_{byN}$] (9).

(9)

AX	Phoneme $_{ax1}$	Phoneme $_{ax2}$	……	Phoneme $_{axN}$
BY	Phoneme $_{by1}$	Phoneme $_{by2}$	……	Phoneme $_{byN}$

Learners of a second language often transfer their native phonological system onto the second language. For instance, if there isn't any [±voiced] distinction in someone's native dialect, but only the [±aspiration] distinction; given a language that distinguish both features, such as Sanskrit, the person might reproduce [p] for [b, p] and [p^h] for [b^h, p^h]. Therefore, the direction of transfer is from Chinese to 'English', i.e. making phonotactic adaptation of the 'English' according to the phonology of the Chinese dialect (10).

(10)

A	Phoneme $_{a1}$	Phoneme $_{a2}$	……	Phoneme $_{aN}$
	By transfer rules ↓	By transfer rules ↓	By transfer rules ↓	By transfer rules ↓
AX	Phoneme $_{ax1}$	Phoneme $_{ax2}$	……	Phoneme $_{axN}$
	By transfer rules ↓	By transfer rules ↓	By transfer rules ↓	By transfer rules ↓
BY	Phoneme $_{by1}$	Phoneme $_{by2}$	……	Phoneme $_{byN}$

The phonology of the transcribing Chinese linguist and that of Modern Beijing dialect are historically related. Historical changes are systematic[11], i.e. given a group of kinship dialects (languages/varieties/…), when the condition of a set of features is fixed, the features may change to another set of features[12]. In an idealized situation, the changes are rule governed and every exception to the rule has its own rule. Therefore, the relationship between a daughter phoneme and its mother phoneme can be represented by a finite set of transformations.

The Phonological system of the transcriber's (BY) must conform to the generalizations of Chinese dialects, which all can be transformed into Modern Mandarin, Beijing dialect, because Chinese characters each represent a syllable, they encode the conditions for phonological change. Therefore, by finite steps of transformation, i.e. by rules that result from synchronic and diachronic studies on Chinese dialects, we can find the mother language of the two dialects (11). Therefore, by finite steps of transformation, we can get from the system of Present-Day Beijing Dialect to the unknown transcribing dialects.

(11)

the transcribing   Beijing Dialect
dialect
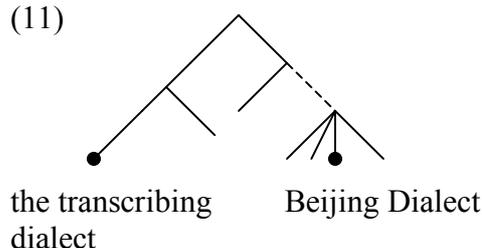
Linguists working on Chinese dialects, present-day dialects and historic dialects, based on studies of *Yun-shu* (dictionaries classifying characters according to their rhymes, usually containing *fan-qie*), rhymed prose and poems, and transliteration of Buddhist cannons, have figured out under what condition (Cn) will a phoneme in Dialect $_\alpha$ corresponds to a phoneme in Dialect $_a$ (12).

(12)   Phoneme $_{\alpha 1}$ → Phoneme $_{a1}$ /_____[Cn]

From the 7th century to present, scholars in different historical periods classify most of the Chinese characters into groups (either by onsets or by rhymes); since the 16th century, there have been descriptions of the diachronic changes and synchronic variations among the groups. Since Bernhard Karlgren, the reconstruction of these systems using IPA is also conducted outside and inside China.

---

[11] Admittedly, language changes are more complicated than the 'family tree' representation. Language contact also affects a language. However, since comparative phonology deduces the reconstructed forms by an idealization that nullifies contact, and I depend on the reconstructed phonological representations, I will also ignore the language contact factor in discussing historical changes.
[12] If consider phonological features are structured features (Moren 2003), the change can be changes of the hierarchy.

With these cumulative knowledge, we can check the pronunciation of each characters, which is a monosyllabic morpheme containing conditions for phonological changes.

Now we can set up a four-dimension matrix[13], with extra steps of phonological rules that mapping Beijing Dialect to the transcribing dialect. Therefore, (10) can be modified into (13).

(13)

Present-day Beijing dialect	Phoneme $_{\alpha 1}$	Phoneme $_{\alpha 2}$	……	Phoneme $_{\alpha N}$
	via M steps of transformations: historical comparison: with possible merge of several phonemes, or split one phoneme into multiple daughters, or sound shifts (moving from one column to another, such as Grimm's law).			
A $_{\text{dialect of the transcriber}}$	Phoneme $_{a1}$	Phoneme $_{a2}$	……	Phoneme $_{aN}$
	By transfer rules ↓	By transfer rules ↓	By transfer rules ↓	By transfer rules ↓
AX $_{\text{phonemes perceived}}$	Phoneme $_{ax1}$	Phoneme $_{ax2}$	……	Phoneme $_{axN}$
	By transfer rules ↓	By transfer rules ↓	By transfer rules ↓	By transfer rules ↓
BY $_{\text{phonemes produced}}$	Phoneme $_{by1}$	Phoneme $_{by2}$	……	Phoneme $_{byN}$

On the 'English' side, similar operation can be conducted, namely, we can link the English dialect to present-day English. Besides, there is another set of rules that are not the same as family-internal dialects. Namely, the transfer from one's native phonological system (a background) to a second language is based on different sets of rules from that in historical comparison. Historical changes are conditioned on the position of the segments, second language transfers are operations that apply across the board, they are not conditioned on the position of the phonological segment. Here, is the transfer from Portuguese to English (illustrated in (14)).

---

[13] Strictly speaking, since the transformations may not be linear, (12) is not likely to be three dimensions.

(14)

Present-day Beijing dialect	Phoneme $_{\alpha 1}$	Phoneme $_{\alpha 2}$	……	Phoneme $_{\alpha N}$
↓	via M steps of transformations: historical comparison rules: with possible merge of several phonemes, or split one phoneme into multiple daughters, or sound shifts (moving from one column to another, such as Grimm's law).			
A $_{\text{dialect of the transcriber}}$	Phoneme $_{a1}$	Phoneme $_{a2}$	……	Phoneme $_{aN}$
	By transfer rules ↓	By transfer rules ↓	By transfer rules ↓	By transfer rules ↓
AX $_{\text{phonemes perceived}}$	Phoneme $_{ax1}$	Phoneme $_{ax2}$	……	Phoneme $_{axN}$
	↑ transfer (adapt) to ax1	↑ transfer (adapt) to ax1	↑ transfer (adapt) to ax1	↑ transfer (adapt) to ax1
BY $_{\text{phonemes produced}}$	Phoneme $_{by1}$	Phoneme $_{by2}$	……	Phoneme $_{byN}$
	↑ transfer from Portuguese	↑ transfer Portuguese	↑ transfer Portuguese	↑ transfer Portuguese
English dialect	Phoneme $_{be1}$	Phoneme $_{be2}$	……	Phoneme $_{beN}$
↑	via E steps of transformations: historical comparison rules: with possible merge of several phonemes, or split one phoneme into multiple daughters, or sound shifts (moving from one column to another, such as Grimm's law).			
Present-day Standard English	Phoneme $_{\beta 1}$	Phoneme $_{\beta 2}$	……	Phoneme $_{\beta N}$

Although the six-raw matrix in (14) seems complicated; however, now for each entry, we know the phonemes of the transcribing character in Beijing dialect, we also know the sound of the present-day standard English reconstructed from the transcription. Following the transformation rules, we can arrive at one set of 18th century Chinese dialects $_{a1}$, and another set of 18th century English varieties $_{by}$. Then after the transcribing officer makes phonotactic adaptation on the English varieties he perceives, he records it in the transcription. This transcription containing its underlying phonological system, must match (ideally be identical to) one of the Chinese dialects $_{a1}$. The very Chinese dialect $_{a1}$ that is used by the officer who transcribes is the link in solving the problem.

3.5 Sum up

With the notion of diaphone and the matrix that were set up above, a seemingly unsolvable problem is now argued that it is in principle possible to be

solved. The important help we received in making a solution possible is based on the assumption that language change and transfer are systematic and rule governed.

4. Application and the results
4.1 Diaphone between known phonemes (A first approximation)

In actually dealing with the data, step-wise reduction and 'substraction' is applied. First, I picked out all the entries that have well-written English and whose Chinese transcriptions read-out in modern Beijing dialect matches closely to the English written forms. Then I take all the entries with clean Portuguese, set standard Portuguese as the target form and use Beijing dialect to read the transcription.

The process of gathering diaphones is as the following: for the entry *colour*, we have a Portuguese word *cor* /kɔɾ/ in the written form, there are two characters used in the transcription, in present-day Beijing dialect, they are read as /tɕɪʌ.ɚ/. The onset /k/ is transcribed by /tɕ/, the nuclear vowel /o/ is transcribed by /ɪʌ/ and the coda /ɾ/ is transcribed with /ɚ/. Thereby, we can gather three diaphones (15).

(15)   /k/ ⇔ /tɕ/
        /o/ ⇔ /ɪʌ/
        /ɾ/ ⇔ /ɚ/

For all onsets, we arrive at two diaphone charts as shown in (16). One of them is English-Chinese based (16a), the other is Portuguese-Chinese based (16b). In both charts, I only list the diaphones whose phonemes are distinctly at variance. Where the phonemes in a diaphone are similar, they are left out. The phonological environment in Beijing dialect when the transcription is used is listed in the third column. "-" refers to cross-the-board application.

(16) Diaphone charts for onsets
a. When the target language is set as English (RP)

Target phonemes (in English)	Phonemes used in Beijing transcription	Conditions in BJ for the use of the transcribing phonemes
[± aspirate,-continuant]	[-aspirate]	-
[+voiced, ,-continuant] /b/ /d/, /g/	[+nasal] /m/ /n/	-
/h/	/ɕ/	___V$_{[-low]}$
/f/	/h/	___V$_{[+high, +back]}$
/k/	/tɕ/	___V$_{[+ATR]}$

b. When the target language is set as Standard Portuguese

Target phonemes (in English)	Phonemes used in Beijing transcription	Conditions in BJ for the use of the transcribing phonemes
[+voiced, ,-continuant] /b/ /d/, /g/	[+nasal] /m/ /n/	-
/h/	/ɕ/	___V$_{[-low]}$
/f/	/h/	___V$_{[+high, +back]}$
/k/	/tɕ/	___V$_{[+ATR]}$

The [±aspirate] difference cannot be distinguished in the transcription; it is also not possible to find a dialect that differs from Beijing dialect in all and only all of the remaining five instances.

4.2 Interpreting the diaphone charts

The problem remains in attributing which factors to Chinese dialectical differences, which ones to English varieties. If the diaphone is conditioned only on the phonotactic constraints of the Chinese syllables, such as the lower three rows in (16a, 16b), I attribute the mapping operation to Chinese dialectical variation. When the transformation between two phonemes of a diaphone is not conditioned on their phonological positions in Chinese, I attribute the factors to second language transfer.

This is because the phonological changes from Early Modern English to present-day English are not very significant, and the differences among English dialects are mainly on their vowel systems; especially, for the resulting diaphone is based on onsets, at this stage of the study, I idealize the segmental features of the onsets to be the same as present-day RP English. If the study on the vowel system could provide us with arguments against such idealization, it is also possible to amend this assumption. For the onset diaphones, there isn't argument to evoke more complex assumptions so far.

Similarly, because Standard Portuguese is enough in explaining the data dug out so far, there isn't a need to assume the informant 250 years ago spoke a dialect that has an onset system significantly different from comtemporary Standard Portuguese.

For Chinese, the officers in the Qing dynasty must be able to speak Mandarin book-reading register, i.e. *wen du*, 'the cultivated pronunciation' (Coblin 2003), the official language/dialect that is used for the ease of communication in a vast country manifesting undisguised diversity.

If the Chinese officer is from South-west Mandarin speaking area and speaking *wen du* as the higher register of the diglossia, then this is a system whose phonological system can match all of the three and only the three diaphones in the lower three rows in (16).

If the informant is a Portuguese speaker, who speaks English with his native accent, i.e. transfer the Portuguese system to his English variety; the difference between English-based diaphone (16a) and Portuguese-based diaphone (16b), is easy

to be explained. The transcription cannot record distinctions that are not in the input, while the Portuguese speaker couldn't implement the [±aspirate], a feature that is not phonemic in his own dialect.

Except for the nasalization, the rest of the diaphones in (16b) are all subject to the environment of the Chinese phonemes. This is a difficult point to be interpreted. A more nature explanation maybe achieved after further study on the phonological system. Nevertheless, at this stage, it is still possible to provide an explanation with a little stretch of the imagination. In Portuguese, coda nasal consonants have a strong tendency to spread their feature [+nasal] across the whole syllable, affecting the onset (17). If the "stops changing to nasals" only exists in words that contain a nasal coda, it would be that the Portuguese speaker pronounced the stops in a way so nasalized that it is being perceived by the transcriber as nasals. This hypothesis remains to be thoroughly tested.

4.3 Draw-backs in the first approximation

The study of YY is still at its initial stage. For onset diaphones, I should list the condition for both English and Portuguese dialect, and see if there is any further generalization that could be made. Same data-gathering process on the codes and nuclear vowel diaphone would also help us testifying or falsifying the present hypothesis.

5. Conclusion

I have argued for the solvability of the problem in disentangling and revealing the phonological systems behind the "English"-Chinese glossary, a complex system that consists of historical, dialectical and language transfer factors. Based on the first approximation, the "English" informant is a Portuguese learner of English, a Portuguese who speaks English with a strong Portuguese accent. The Chinese officer who is transcribing, records the oral input in *wen du*, a higher Mandarin register in the mid-18th century, with his own Southwest Mandarin dialectical influence.

The manuscript of YY is a complex system that resembles a *sprachbund* system, where both historical changes and language contact have left traces in the phonological system in question. The method that is used in studying YY is also of importance in study the later systems.

References:
Academia Sinica: *Qing-Shi-Lu: Gao-Zong-Shi-Lu*. Vol 324. 352-3
Bolton, K. 2003 *Chinese Englishes : a sociolinguistic history*, Cambridge : Cambridge University Press, 2003.
Coblin. W. S. 2003. "Robert Morrison and the Phonology of Mid-Qing Mandarin", JRAS, Series 3, 13, 3, pp 339-335.
Eames, J. B. 1914. *The English in China: being an account of the intercourse and relations between England and China from the year 1600 to the year 1843 and a summary of later development.* London: Curzon, New York: Barnes & Noble.

Fang, H. 1987. Zhong-Xi-Jiao-Tong-Shi (A history of Sino-Western Relations). Changsha: Qiulu Shushe.

Fuchs, W. 1931. "Remarks on a new 'Hua-Yi-I-Yu' *Bulletin of the Catholic University*, Peking, No. 8 : 91-97.

Johns, D. 1960. An outline of English Phonetics (9th edition). Cambridge: CUP.

Johns, D. 1950. The Phoneme: Its nature and use. Cambridge: W. Heffer & Sons Itd. 193-200.

Li, Z.H. and C.J. Zhou. 1998. *Hanzi Gujin Yinbiao*. Beijing: Zhonghua Shuju.

Luo. C. 1932. "Zhong-Yuan-Yin-Yun Sheng-Lei-Kao", *Bulletin of the Institute of History and Philology*. Vol 2. Part 4. Peking: Institute of History and Philology.

Moren, Bruce. 2003. The parallel structures model of feature geometry. *Working papers of the Cornell Phonetics Laboratory*.

Ning, J. 1985. *Zhongyuan-Yinyun-Biao-Gao*. Changchun: Jinling Wenshi Chubanshe.

Nishida, T. and H.K. Sun 1990. *Baima Yiyu Yanjiu* [白馬譯語の研究 : 白馬語の構造と系統]

Wang. B. 1997. "Xuanzangs Leben und Werk: Teil I, II, IV", *Tang Yanjiu*, *Tang Research Fundation* vol 3.

Zhou, Z.H. 2003. Wanqing de Yangjingpang Yingyu—Du Daying Tushuguan Suocang *Hongmaofanhua*.

Zhu. J. 1992. *Liangchao Yulan Tushu* (Royal Collection in the Ming and Qing Dynasty) Beijing: the Forbidden City Press. P149.

Wang. Y. 2000. "*Liulun Ming-Qing Zhiji Xi-xue-dong-jian de Tedian yu Zhongxi Xueke Hudong*", [*On the Sino-western Intellectual Communication in the 16th century*.] . Guoji-hanxue [International Sinology]. Vol.6. Zhengzhou: Daxiang Press. P327.

Wild 1943

(The original file written in December 2005 is now lost.
Retyped and revised in November 2009)

English to Chinese dictionary with Mandarin pinyin - learn Chinese faster with MDBG! Â reference book consisting of material quoted from many sources, arranged by category (about 600 of which were compiled between the 3rd and 18th centuries in China). é¡žæ›¸. æµ…è‰. QiÇŽn cÇŽo. *æµ…è‰* | æµ…è‰* | *æµ…è‰. Asakusa, district of Tokyo with an atmosphere of old Japan, famous for the 7th century Buddhist temple, SensÅ-ji. æ·ºè‰. English-German translation for: 18th century. Ã¤Ã¶Ã¼ÃŸ De <> en de â€œ> en en â€"> de ----- de <> en ----- de <> bg de <> bs de <> CS de <> da de <> el de <> eo de <> es de <> fi de <> fr de <> HR de.Â English-German online dictionary developed to help you share your knowledge with others. More information! Contains translations by TU Chemnitz and Mr Honey's Business Dictionary (German-English). Thanks on that account! Links to this dictionary or to single translations are very welcome! Questions and Answers.