Abstract
Vietnamese has been demonstrated to be a Mon-Khmer Austroasiatic language (Haudricourt 1954, Shorto 2006), albeit one which differs substantially from the typical Austroasiatic phonological template (Alves 2001). Some of that linguistic transformation was most likely due in part to language contact with Chinese, primarily through the massive lexical borrowing that took place over the past two millennia. However, the question of the sociolinguistic conditions under which this borrowing occurred over this large period of time has nevertheless been little described. The main purpose of this paper is to consider the borrowing of grammatical vocabulary in particular from Chinese into Vietnamese to exemplify the long-term Sino-Vietnamese language contact. This requires an exploration of the socio-historical context in which the elements of Chinese came into Vietnamese and a sorting out of the spoken versus literary means of transmission of linguistic borrowing. This case study in the borrowing of grammatical vocabulary sheds light on the issues of language contact and linguistic borrowing when a prestigious written language is accessible to a linguistic community.

Overview
A database being amassed by this author indicates that well over 400 Vietnamese words, considered native vocabulary today, were most likely borrowed via a spoken means of transmission around the time of the Han Dynasty (though some possibly as late as the beginning of the Tang Dynasty, which began in the 7th century CE). This large number of early loanwords at least in part the result of the immigration of some twenty thousand Chinese soldier-settlers who were sent to Vietnamese and brought with them many of the cultural customs and material trappings of Chinese civilization (Taylor 1983:49). The Han Dynasty was, however, the only period in which such a large quantity of spoken Chinese was directly borrowed without the powerful influence of written Chinese. It is the assertion here that the early foundation of Chinese culture in the Han Dynasty coupled with the second major spread of Chinese culture during the powerful Tang Dynasty (7th to 10th

1 The database is based on work by numerous scholars, including primarily Haudricourt 1954, Wang 1958, Mei 1970, Pulleyblank 1981 and 1984, and Nguyễn T. C. 1995. Admittedly, complete certainty of loanword status of the words in the database is impossible to achieve. Instead, the author has evaluated a range of high to low certainty based on the overall phonetic and semantic patterns, coupled with historically documented details about the kinds of social contact at that time. Of the over 500 words, nearly 300 have been evaluated as highly likely Old Chinese loanwords, about 150 are at medium certainty, and about 40 are at low certainty.
centuries CE) in Vietnam served as a socially prestigious platform from which Vietnamese literate in Chinese could spread Chinese vocabulary, including grammatical vocabulary, into Vietnamese regardless of the number of actual bilingual Chinese speakers in Vietnam.2

Material borrowing, in contrast with borrowing of syntactic and phonological patterns, may occur from languages of high status even without a bilingual population (Sakel 2007). This appears to be the case in Vietnam, in which the initial era of Chinese political domination was marked by a substantive and influential population of Chinese settlers. Subsequently, the direct influence of the Chinese population was diminished as they were nativized (Taylor Ibid.:52). There have been numerous instances throughout history when groups of Chinese maintained small but financially influential communities in Vietnam, and written Chinese has constantly been an important part of the upper levels of Vietnamese society, but there has never been an era in which Chinese was spoken throughout Vietnam. Thus, it must be concluded that, over the past thousand years since the time of Vietnamese political independence from China, the time during which the bulk of Chinese vocabulary was borrowed into Vietnamese, written texts have been the primary source of this borrowing.

The focus of this paper, transfer of grammatical vocabulary, is particularly telling of the increased borrowing via literary texts. While Vietnamese syntactic structure has largely been unaffected by Chinese and maintains a primarily Southeast Asian template (Alves 2001), the amount of grammatical vocabulary in Vietnamese of Chinese origins is significant. They constitute several major categories, including connective words, passive voice markers, classifiers and general measure words, among others (Lê 2002, Alves 2005 and 2007).

The earliest well-known linguistic description of Vietnamese appears in the 1651 Vietnamese-Portuguese-Latin dictionary of Alexandre de Rhodes, the “Dictionarium Annamicum, Lusitanum, et Latinum”.3 The introduction to the text contains a grammar section, and grammatical words and examples of their usage are provided throughout the dictionary’s 9,000 entries. Exploration of the data (referring to a 1991 translation into Vietnamese of the original Latin text) shows that, structurally, Vietnamese syntax has changed little since the 1600s. While the dictionary was influenced to a good deal by Central Vietnamese, with some lexical and phonological characteristics specific to that region, the text can still be considered representative of general Vietnamese grammar. Overall, the data in the dictionary clearly show that Vietnamese at that time was a topic-comment language with other typological characteristics similar to Vietnamese today.

The Vietnamese grammatical vocabulary inventory, on the other hand, has changed noticeably over the past three and a half centuries. In a comparison of the grammatical vocabulary of the 1600s (both de Rhodes’ work and a dictionary of archaic Vietnamese by Vuong 2002) and that of today, in some cases, there are preservations or minimal semantic and phonetic changes of some grammatical words. In other cases, some words have changed more substantially in their semantico-syntactic functions and are in the ongoing process of grammaticalization. Finally, there are grammatical words in the pre-modern era which do not exist today or which have very limited usage in modern Vietnamese, and a

---

2 Consider the Latin loanword “via,” which is considered a formal register word in English, in contrast with the more neutral English word “through”.  
3 Other works that precede de Rhode’s work are discussed in Jacque 2002.
noticeable number of those words are not of Chinese origins. It is this last category of words that are of particular interest in this study.

In the following sections, the eras of socio-historical Sino-Vietnamese contact are described, and then linguistic data are provided to demonstrate how Sino-Vietnamese grammatical vocabulary were borrowed over the past few centuries through biliteracy rather than spoken bilingualism.

**Historical Sociolinguistic Background**

The eras of Sino-Vietnamese contact are here divided into four general categories based on the nature of the sociolinguistic contact: (a) the Han Dynasty era (1st century BCE to 2nd century CE), (b) the Tang Dynasty era (7th to 10th centuries CE), (c) the era of Vietnamese independence (11th century to the modern era), and (d) the modern era (20th century to the present). Besides the first era, all the other eras are marked by situations in which Chinese is largely transmitted via writing rather than an influential Chinese vernacular community.

Documented Sino-Vietnamese language contact begins early in the Han Dynasty. The ancestors of the modern Vietnamese resided, at that time, primarily in modern day northern and north-central Vietnam, with a cultural center in the Red River Delta. In the Eastern Han dynasty at the beginning of the Christian era, the Chinese administration mandated the adoption of Chinese cultural customs throughout Vietnam, including Chinese family and household customs and accoutrements (e.g., Taylor Ibid.:33-34). The tools of administration left lexical imprints (e.g., *giấy* “paper” (Sino-Vietnamese *chữ*: Chinese ㄓ), *họ* “family name” (Sino-Vietnamese *hộ*, Chinese ㄏ “household”), etc.), though these etyma were nativized and later reborrowed with standardized, Sino-Vietnamese literary readings (the second readings in the previous examples).

Another crucial burst of language contact occurred during the time when large groups of Chinese soldier-settlers and the establishment of an elite Sino-Vietnamese class, who, despite their eventual “Vietnamization,” maintained some sense of Chinese identity for centuries. As noted, there are perhaps hundreds of these words which belong to a core of Vietnamese culture, and thus this contact was indeed significant. During this period, there occurred the borrowing of at least a few hundred Chinese words, mostly nouns and some verbs, but almost no grammatical words.

Chinese power wavered after the Han dynasty. To what extent sociolinguistic contact led to borrowing before and into the early Tang dynasty several centuries later is less clear, though there were certainly Chinese leaders, armies, and traders in Vietnam in this era, and this was the period during which Chinese-style Buddhism began to flourish (Taylor Ibid.:80-84). At the very least, it can be said that Vietnam’s continuing status as part of China coincides with a progression into further sinicization of culture and language, a process which was complete among the modern varieties of Chinese spoken in Southern China, where there had been numerous non-Chinese groups prior to the Han Dynasty.

China regained its political strength in the Tang Dynasty, and a large-scale spread of Chinese writing ensued throughout East Asia—including Japan, Korea, and Vietnam—via the Chinese rhyme dictionaries. These texts, containing tens of thousands of Chinese characters, provided access to the entirety of the Chinese lexicon. From this era on, the vast majority of Chinese loanwords have maintained their official, standardized, literary readings, in contrast with the vernacular pronunciations of the Han Dynasty loans.

At the end of the Tang Dynasty, Vietnam gained political independence. While there continued to be Sino-Vietnamese contact through trade, politics, religion, and
education and some amount of Chinese immigration (Luong 1988, Châu 2006), with few exceptions, there are no instances of large-scale migrations of Chinese into Vietnam in this era that would have resulted in widespread spoken bilingualism. This combination of factors—little Chinese immigration and ready access to Chinese vocabulary without the need for native speakers—supports the idea of a mainly literary means of transmission. Perhaps somewhat ironically, at the same time that the Vietnamese increasingly sought political independence from China, the Chinese political and educational model grew in influence in Vietnam. This is exemplified by the creation of the Confucian university, the Văn Miếu (文廟 wén miào) “Temple of Literature,” shortly after independence from China, thereby establishing a long-term literary Chinese tradition in Vietnam.

With this simultaneous independence from China but strengthened ability of Chinese writing as a center of education in Vietnam in the Post-Tang Dynasty era, it can be assumed that borrowing from Chinese continued to be primarily through biliteracy of the Vietnamese literary elite. Modeling of the Chinese socio-political and cultural systems continued even into the 1800s (Woodside 1971). This was the case regardless of the size of the Chinese population in Vietnam, which did increase at times, particularly in the late 1800s under French interest in Chinese labor and managerial skills. The influential Chinese merchant class moved easily throughout Vietnam, but also continued to establish permanent family-managed properties and businesses (Ibid. 272). However, over a period of several decades, the immigration of many dozens of thousands of Chinese, many of whom came from neighboring Guangdong and Guangxi provinces, did not result in massive lexical borrowing of spoken Cantonese or indeed any other Southern variety of Chinese. Over the centuries, larger numbers of Chinese loanwords entered daily, spoken Vietnamese, but these were literary, non-dialectal readings. Loanwords from Cantonese are few in number, a few dozen at most, and are limited mainly to the domain of food (e.g., chiến “to pan fry” (Sino-Vietnamese tiên; Chinese 煎 jiān; Cantonese jīn), lap xưởng “Chinese sausage” (Sino-Vietnamese lạp tràng; Chinese 臘腸 là cháng; Cantonese laahp cheungkin)). This situation is in sharp contrast with the hundreds of Han Dynasty era loanwords which span numerous semantic domains and which have remained part of the Vietnamese lexicon for two thousand years. Finally, it is worth noting that these loanwords are clearly recent borrowings based on their close phonetic matches, and none of them appear to be typical Yue or Cantonese dialectal words.

In the modern era, from the early 20th century, it is clear that borrowing from Chinese into Vietnamese occurred mainly as a result of biliteracy among Vietnamese. The large-scale spread of “Sino-neologisms” (i.e., translation by Chinese and Japanese of Western concepts and terms using Chinese morphs, typically combinations of two morphs) led to the borrowing of many thousands of “Chinese” words, but these came from both Japanese and Chinese texts. A number of influential Vietnamese studied in Japan in the early 20th century, helping to stimulate the spread of these words (Sinh 1993). As these loanwords were borrowed primarily from writings, they are consistently pronounced with literary Sino-Vietnamese readings, and notably not with any dialectal pronunciations. At the same time that Sino-neologisms entered Vietnamese, there was massive growth in

---

4 One possibility considered by this author is that the neighboring Pinghua Chinese, distinct from Yue Chinese, spoken in modern day Guangxi province, where Chinese schools existed, could have been a source of the so-called “southern koine” (Hashimoto 1978). Exploration of Pinghua lexical and phonological data in Li 1998 shows no traces.
literacy rates in Vietnam—from 5% to 20% prior to World War II (DeFrancis 1977:218) to 90% today. This increase in literacy also corresponds to the time when the national Vietnamese Quốc Ngữ alphabet became an important aspect of literacy campaigns (Marr 1981:137 and 181). Finally, the intentional standardization of the massive quantities of new vocabulary of largely Chinese origins (the Vietnamese lexicon grew from 40,000 in 1945 to hundreds of thousands within a few decades (Marr 1981:168, Nguyễn et. al. 2002:19) further magnified the impact of these Chinese lexical imports on both the spoken and written Vietnamese lexicon. All in all, the borrowing of Chinese words came into Vietnamese via the written word.

Linguistic Data and Grammatical Vocabulary
The Vietnamese lexicon has been described as being 70% Chinese in origin, with technical vocabulary constituting 80%. However, in a loanword typology project utilizing a list of about fifteen hundred words, only 27% of Vietnamese vocabulary was shown to be from Chinese (Alves 2007a). However, this number must be considered low since that study did not include grammatical vocabulary, names, common vocabulary in the region, among other categories of words to which the Chinese lexicon is the source in Vietnamese. Still, by focusing on a set of more core vocabulary, as the study did, it does suggest that 70%, which includes the entirety of a dictionary, is not a realistic figure either if the goal is to determine the depth of influence on spoken Vietnamese in contrast with specialized vocabulary. No studies thus far have indicated the percentage of vocabulary of Chinese origin based on a set of high-frequency Vietnamese vocabulary. Such a study would logically be expected to show a number somewhere between 27% and 70% and more accurately and realistically portray the role Chinese has played in the Vietnamese lexicon. Regardless, the number of words of Chinese origin must be considered substantive even if little more than a third of core Vietnamese vocabulary is Chinese in origin.

As for grammatical vocabulary, no studies have been found to quantify the percentage of function words of Chinese origin, though the percentage must indeed be substantial. Based on collections of such words in Lê 2002 and Alves 2005 and 2007a, connective words are largely of Chinese origin, dozens of measure words and some major classifiers are from Chinese, a number of preverbal grammatical morphs are Chinese, and a majority of the words in the complex system of pronoun reference come from Chinese. However, Chinese grammatical vocabulary, which entered Vietnamese at different times, did so for the most part in the Post-Tang era since most are pronounced with their literary readings from this era. In fact, it may be the case that a majority of Chinese grammatical vocabulary entered Vietnamese after the 1600s and even as late as the turn of the 20th century.

Nguyễn Đình Hòa (1991) identified archaic lexical items in de Rhodes’ dictionary which are not part of modern Vietnamese. Exploration in the dictionary of Vương (2002), which is based on numerous ancient Vietnamese writings from the past several centuries, also reveals additional archaic vocabulary in Vietnamese which has been replaced over the centuries. Unfortunately, more detailed statistical studies of the timing of the historical changes in written records, which would serve to clarify and verify the ideas in this study, are non-existent. Still, the position that some of the words in Vietnamese in the 1600s are not in modern, standard usage is feasible and can be confirmed by native speaker intuitions and simple corpus queries. Both lexical loss and replacement did take place over the past few centuries, and the realm of grammatical vocabulary also shows this kind of change.
Such is the case for the archaic Vietnamese words bèn “but” and ãu là “or”, which have the modern Sino-Vietnamese counterparts nhùng “but” (仍 rēng) and hoặc “or” (或 huò).

Of particular note in this study are the numerous examples in de Rhodes’ dictionary of non-Chinese grammatical vocabulary which were subsequently replaced by synonymous Chinese vocabulary. The grammatical functions are wide ranging, including a number of important grammatical lexical categories. Examples are shown in Table 1, which contains grammatical words found in de Rhodes’ dictionary which are not Chinese in origin and their mainstream, modern counterparts of Sino-Vietnamese origin. Forms in the 17th century column marked with an asterisk are extremely literary and/or have very limited usage in modern Vietnamese.

Table 1: 17th Century Grammatical Words in Vietnamese and their Modern Sino-Vietnamese Replacements

<table>
<thead>
<tr>
<th>Grammatical Functions</th>
<th>17th Century</th>
<th>Modern Era</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantity</td>
<td>đối* “every”</td>
<td>mọi “every” (每 méi)</td>
</tr>
<tr>
<td></td>
<td>phô “the various (higher social status)”</td>
<td>các “the various” (各 gē)</td>
</tr>
<tr>
<td>Sentence connecting</td>
<td>bèn “but”</td>
<td>nhùng “but” (仍 rēng)</td>
</tr>
<tr>
<td></td>
<td>chưng* “because”</td>
<td>vì⁵ (non-literary reading) “because” (為 wēi)</td>
</tr>
<tr>
<td></td>
<td>âu là* “or”</td>
<td>hoặc “or” (或 huò)</td>
</tr>
<tr>
<td>Negation</td>
<td>chảng⁶ “no/not”</td>
<td>không “no/not” (空 kōng)</td>
</tr>
<tr>
<td>Location</td>
<td>ca “at”</td>
<td>tại “at” (在 zài)</td>
</tr>
<tr>
<td>Comparison and intensification</td>
<td>ngât “equal to”</td>
<td>bàng (Old-Sino-Vietnamese) (平 ping)</td>
</tr>
<tr>
<td>Time terms</td>
<td>chưng (progressive marker)</td>
<td>đặng (progressive marker) (當 dāng)</td>
</tr>
<tr>
<td></td>
<td>đặc “time/instance”</td>
<td>lân “time/instance” (輪 lún)</td>
</tr>
<tr>
<td>Grammatical adverbs</td>
<td>bui “only”</td>
<td>chi “only” (只 zhī)</td>
</tr>
<tr>
<td></td>
<td>chìn “truly”</td>
<td>thật “truly” (實 shí)</td>
</tr>
<tr>
<td></td>
<td>nghĩ “by oneself”</td>
<td>tự “by oneself” (自 zì)</td>
</tr>
</tbody>
</table>

Data in de Rhode’s 1651 Vietnamese-Latin-Portuguese Dictionary reveal the following. First, de Rhodes’ dictionary highlights the diglossic distinction between Chinese vocabulary and Vietnamese, with Chinese morphs having a formal status even higher than

---

5 This particular form is a nativized reading with the huyên tone. The standard literary pronunciation is vi, with the nặng tone. The native huyên tone vs. the literary nặng tone is seen in a number of forms, as discussed in Alves 2005.

6 The form chằng “no/not” in particular has more widespread usage in modern Vietnamese, though statistically, it has significantly lost its status to không “no/not.”
it is today. With literacy in the pre-modern era at a minimum, a small fraction of the entire population, it must be assumed that only biliterate Vietnamese could have been those in control of initiating the spread of such vocabulary, both content words and grammatical vocabulary. Next, de Rhodes’ dictionary shows that Cantonese or other varieties of Yue had contributed extremely little in terms of lexical content by that time, again suggesting that spoken bilingualism in Chinese was relatively unimportant after the first few centuries of Sino-Vietnamese language contact.

While the de Rhodes’ dictionary highlights the highly formal status of Chinese vocabulary in Vietnam in the 17th century, modern era Sino-Vietnamese grammatical vocabulary supports the idea of the literary means of transmission by their notably literary status. In a collection of 156 grammatical Vietnamese words of Chinese origin (Lê 2002: 397-403), many belong to a very formal and/or written register (e.g., nhược (若 ruò) “if,” giả sù (假使 jià shǐ) “in the event that”, and sô dì (所以 suǒ yǐ)). Another characteristic of these words is that some are prefixes in Vietnamese but free morphs in Chinese (e.g., bất “un-” (不 bù), tài “re-” (再 zài), and tôi “-est” (最 zuì)), which suggests that such morphs were not borrowed as part of a grammatical system but rather simply by borrowing the morphs in words, again suggesting borrowing through literacy in Chinese (and in some cases, in Japanese). Some of the borrowed words differ in part of speech from the Chinese forms. For example, the Chinese adverb 果然 (guǒ rán) “as expected” is an adjective in Vietnamese quâ nhiên, and the Chinese adverb 實在 (shí zài) “truly/really” is in Vietnamese thực tại both an adverb “really” and noun “reality”. Finally, it is important to note that some of these grammatical words, which are common in Mandarin, spoken far to the north of Vietnam, are not spoken Cantonese (though they appear in literary Cantonese), for instance, bị (passive marker) (被 bei) and tài “at” (在 zài). This might seem counterintuitive as Mandarin has never been spoken widely in Vietnam, while Cantonese is a virtual lingua franca among Chinese in Vietnam, unless one accepts the idea that the borrowing came via written Chinese texts, which contain essentially Mandarin-style grammar and grammatical vocabulary.

Conclusions

While at some points, some borrowing via spoken transmission did take place, primarily during the initial contact in the early Han Dynasty and a limited scope of borrowing from Cantonese in the modern era, most Sino-Vietnamese borrowing has taken place via a written means of transmission. The limited immigration of Chinese into Vietnam, the substantial adoption of the Chinese written tradition and cultural patterns, and the tendency towards literary status of Chinese vocabulary in Vietnamese all support this position. Among the borrowed grammatical Chinese vocabulary, the vocabulary is higher register, not borrowed from dialectal varieties of Chinese in or near Vietnam, and show unexpected semantico-syntactic shifts from loan source, all highlighting this more literary status and route for borrowing.

These data not only portray a portion of the linguistic history of Vietnam but also serve as a case study of language contact (both spoken and written), of the sociolinguistic

---

7 Spoken Cantonese, like Vietnamese but in contrast with Mandarin Chinese, does not use the free morph 不 bù “no,” which only occurs in bound form in words or in highly literary Cantonese. The other grammatical words noted, however, are free morphs in Cantonese, like Mandarin Chinese but in contrast with Vietnamese.
history of the peoples in East and Southeast Asia, and of broader psycholinguistic issues (i.e., spoken versus written language, semantico-syntactic categories of words).

References


