Identifying the Tai Substratum in Cantonese[1]

Robert S. Bauer
Institute of Language and Culture for Rural Development
Mahidol University at Salaya, Thailand

1.0 Introduction

A number of common Cantonese words seem to lack any etymological relationship with semantically-equivalent words in Mandarin. How can these lexical differences be accounted for? Given that the same phenomenon occurs in other Chinese or Han dialects, the question holds much interest for anyone studying the historical development of the Han dialects. It thus seems quite natural that the question should also be an issue in Sino-Tibetan historical comparative linguistics. Indeed, twenty years ago Hashimoto (1976) believed it was too early to talk about a Sino-Tibetan language family on a par with Indo-European until Sinologists had adequately answered the fundamental question, what accounts for the apparent diversity of basic vocabulary within the Sinitic half of the family? He (p. 2) drew attention to the fact that the comparison of basic words in the Han dialect families of Mandarin (northern group), Wu, Gan, Xiang (central group), and Min, Yue, and Kejia (southern group) indicates a fair portion of it derives from diverse sources. Table 1 below lists forms from the seven major Han dialect families for four lexemes for which we observe several patterns of relationship: for 'to cover' there appears to be a clear division between the northern and central groups with their open syllables, on the one hand, and the southern group with syllables closed by the bilabial nasal, on the other; in the case of 'frog' all dialect families except Kejia and Wu may share one root morpheme which had a velar initial and a bilabial stop final consonant; for the third person singular pronoun Mandarin and Xiang share the same form, but forms in the other dialect families clearly derive from different etyma; the morpheme for 'this' is shared by Xiang and Gan and possibly Wu but takes quite diverse phonetic shapes in the other dialects.

Table 1. Basic vocabulary in 7 Han dialect families (from HFCH).

English Gloss	Mand.-Beijing	Wu-Suzhou	Xiang-Changsha	Gan-Nanchang	Yue-Guangzhou	Kejia-Meixian	Min-Xiamen
cover, to	kai^{5}	kE^{5}	kai^{5}	kai^{5}	khɐm^{3}	khɛm^{2}	khɐm^{5}
frog	xa^{1} ma^{0}	ts^{h}in^{1} o^{1}	ka^{7} mau^{0}	ha^{2} ma^{0}	kɐp^{7a} na:^{3}	ts^{h}iaŋ^{1} kuai^{3} ɛ^{0}	ts^{h}an^{2} kap^{7} a
he, she	t^{h}a^{1}	li^{1}	t^{h}a^{1}	tɕ^{h}iɛ^{2}	khɵy^{4}	ki^{2}	i^{1}
this	tʂei^{5}	kE^{1}	ko^{7}	kɔ^{3}	ni:^{1}	e^{3}	tsit^{7}

[1] This is a much revised version of the paper presented at the *4th International Symposium on Languages and Linguistics, Pan-Asiatic Linguistics* in Bangkok on January 9, 1996.

This bit of comparative lexical material is enough to demonstrate that the so-called Han "dialects" are not as homogeneous as the term *dialect* might suggest. However, what is more important and relevant to the purpose of this paper is the question, what has created this etymological diversity among the various Han dialect families? As will become clear in the following discussion, some scholars believe that historical contact among Han dialects and non-Han languages must account for some of it. In a wide-ranging paper on this very topic Wang (1995:18) has noted that "the deeper we probe into these questions of contact and differentiation, the more we are likely to realize the importance of the non-Han languages in the formation of the Han dialects, and that China is indeed a complex Sprachbund . . ." Determining the contribution of the non-Han languages to the historical development of the Han dialects is a subject of great significance; yet its magnitude and complexity may help explain why we still do not have an etymological dictionary of Han dialectal vocabulary (as opposed to the Chinese characters which are not necessarily equivalent to words). Working with the reading pronunciations of the Chinese characters in the different dialects rather than vocabulary items in speech, Sinologists have generally focused their attention on the changes that have taken place in the pronunciations of the standard Chinese characters; as a result, they have tended to shy away from tackling the more difficult questions associated with the emergence and evolution of the Han dialects.

Over the last few years my own interest in this subject has been mainly limited to the southern group of Han dialect families of Yue, Min, and Kejia. In my approach I have focused on a particular kind of basic vocabulary, viz., characterless morphemes -- words which are not etymologically associated with standard Chinese characters. I have begun with Yue, the dialect family with which I am most familiar, and then have systematically expanded the scope of my research by tracing lexical connections to the neighboring Min and Kejia dialects, and then to the non-Han languages of Southeast Asia. This paper attempts to identify the origins of 29 Cantonese words which appear to lack any etymological relationship with standard Chinese characters.

2.1 Language Families of the Southeast Asia Sprachbund

From their early beginnings down to the present day the language groups of Southeast Asia, i.e., Chinese (or Sinitic), Tibeto-Burman, Austronesian, Tai, Mon-Khmer, and Miao-Yao, have been in contact with each other and have been influencing each other's development. This long history of mutual influence has resulted in areal convergence which has made working with these languages a particularly complicated matter. Gedney (1976:66), a specialist in the description of Tai dialects, has stated the difficulty of sorting out relationships among Southeast Asia's languages as follows:

"Linguists who study Southeast Asia, one of the most complex areas in the world, are constantly struck by instances of apparent convergence, where languages or language groups which are genetically unrelated (or if related, the relationship is so far back in time as to [be] irrelevant), e.g. Thai and Mon-Khmer, or Thai and Tibeto-Burman, or Thai and Vietnamese, or any Southeast Asian language and Chinese, show similarities in grammatical structure or the organization of the semantics of the lexicon. These convergences are clearly the result of contact; how much former bilingualism must be assumed in order to explan these similarities is not yet clear."

A topic of particular interest to me is how the various language groups of Southeast Asia have influenced the development of the southern Han dialect families of Yue, Min, and Kejia. At the same time, I am also deeply concerned with their historical origin and how they have been influencing the development of each other. I have not been alone in pursuing this line of research. With respect to the development of Yue, Li J-z. (1990:28) has delimited the problem as follows:

"*In a nutshell, the Yue dialect area encompasses Min, Zhuang, Yao, and Kejia* languages. In that case, under the circumstances in which numerous languages are enclosed within the same area, how exactly has the Yue language emerged and developed?"

2.2 The Sino-Tai Relationship

Benedict (1975:32, 35, 135), a pioneer in the historical comparative linguistics of Southeast Asian languages, has proposed that Tai, Austronesian, and Miao-Yao belong to a larger family which he has named Austro-Tai, while Tibeto-Burman, Karen, and Chinese constitute the Sino-Tibetan family; Mon-Khmer, a sub-branch of Austroasiatic, is a separate family from these other two. Linguists in China have generally recognized that Sino-Tibetan includes Chinese, Tai, Tibeto-Burman, Karen, and Miao-Yao. However, they have not defined the Sino-Tai genetic relationship according to the traditional sense of this term. Chinese and Tai have been described as not originally sharing a genetic relationship but as having developed such a relationship as a result of their convergence through areal contact. Based on a detailed comparison of basic vocabulary from Chinese, Tibeto-Burman, and Tai, the eminent Chinese Tibeto-Burmanist, Dai Qingxia (1991:63), concluded that Chinese and Tai share a "'genetic' link" which is of a different type than the one which joins Chinese with Tibeto-Burman. Dai's comparison of basic vocabulary clearly showed that Chinese and Tibeto-Burman share cognates with each other but not with Tai. On the other hand, Chinese and Tai share some non-basic vocabulary with each other but not with Tibeto-Burman. He concluded that Sino-Tibetan languages are related through two kinds of genetic relationship:

"one is formed by direct descent from the proto-language (like that of Chinese and T[ibeto]-B[urman], and the other is forged under language contact (like that of Tai and Chinese). This is then one of the major points on which the formation of Sino-Tibetan is different from other families, such as the Indo-European. . . while Chinese and T-B are outwardly dissimilar but have the same origin, Tai and Chinese are similar in appearance but have different origins. . . We can . . . treat the relationship between Chinese and Tai as a genetic one, since the latter has experienced a qualitative change under language contact and has organically assimilated into the family of the influencing language" (Dai 1991:63).

The Tai and Chinese historical linguist, Li Fang Kuei (1976:231-237), compared a number of items of basic vocabulary from modern Thai, modern Chinese, and his reconstructed Old Chinese which he claimed showed regular phonetic correspondences. He made clear his belief that these shared words must have been inherited from the same source language and not borrowed. Prapin (1976:12-25), a student of Li, presented an even larger set of Sino-Tai lexical comparisons for which she claimed the phonetic correspondences established that Old Chinese and Proto-Tai were related not only through their descent from a common ancestor but also contact borrowing. She stated that this borrowing occurred in both directions but did not specify for which words Tai or Chinese was the donor.

In the eyes of most Western linguists the convergence between Chinese and Tai in monosyllabicity, tones, vocabulary, grammatical categories, and word order has resulted from their areal contact. Egerod (1976:52), the prominent Danish sinologist who worked for decades on both Tai and Han dialects, described the Sino-Tai relationship as follows (names of language families have been adjusted to correspond with those used in this paper):

"The similarities between [Tai] and Chinese are non-genetic and just one symptom of the tremendous consequences of the rise, intermingling, and fall of three major language families in present-day China: [Sino-Tibetan], Mon-Khmer, and [Austronesian]. And it is toward [Austronesian] and not toward [Sino-Tibetan] that we have to look for the genetic relationship with [Tai]."

2.3 Sino-Tai Lexical Borrowing

Regardless of whether Tai and Chinese have separate origins, there is no question that their close contact has resulted in each language group influencing the development of the other. The traditional, sinocentric view (much revised by Benedict 1976:60) holds that the Han civilization historically dominated East Asia for more than two millennia, and that the Chinese language exerted great influence on its neighbors, e.g., by exporting vocabulary, tones, and monosyllabicism to Tai and Miao-Yao (Benedict 1976:62) and Vietnamese.