

# **The Tai language family and the comparative method**

**Anthony DILLER\***  
Australian National University

As attested to by other contributions to this book, by now it is common to encounter a language family called ‘Tai’ that would include familiar members such as Thai and Lao. Along with these national languages would be their closely-related regional varieties and the family would include less well-known minority languages and varieties such as Khamti, Aiton, Phake, Dehong, Lue, Khuen, White and Black Tai, Nung, Tay, Northern and Southern Zhuang—and scores of other language names could be added. To a large extent, the Tai languages gain their cohesion and family status from the Comparative Method of historical linguistic reconstruction.

A central purpose of this paper is to distinguish and clarify which types of linguistic evidence have a core relevance to the operation of the Comparative Method in the Tai case and which do not—a distinction which I believe has led to misunderstanding in the past, including proposed questionings of or deconstructions of the Tai language family. Also considered is the relation of time depth to how effectively the Comparative Method can be deployed.

In the sections below we investigate some of the background of the usage of ‘Tai’ as in the ‘Tai language family’ and raise general questions of what constraints and limitations need to be kept in mind when we conceptualize or speak of language families and apply this notion in the Thai/Tai context. Sections 1-8 examine how the Comparative Method has been applied in the case of the Tai languages, with Sections 9-10 briefly turning to wider relationships with other languages and families.

## **1. The comparative method and Tai research**

Before attempting to answer questions such as: “to which language family does Thai belong?” it is useful to probe some certain assumptions that lie behind the question itself. One could ask where the family metaphor, as applied to language, has

---

\*Parts of this paper were developed with support from Australian National Research Council grants A59601467 and A59905294. An earlier Thai version of some sections appeared as Diller (2533/1990). For helpful comments, I am indebted to participants of the International Conference on Tai Studies, Bangkok, July 29-31, 1998, organized by the Institute of Language and Culture for Rural Development, Mahidol University.

come from and what similar metaphorical representations have been used in linguistics to represent language history and interrelationships.

In an important way, an early leading figure in Thai language study, Phaya Anuman Rajadhon, has raised the issue of interest in the present inquiry. In his work *Niruktisat* (2499/1956:1:42) he addresses the problem of using a figure of speech like the Thai word *traku:n* 'lineage' for language groupings and notes the misunderstanding or false presuppositions such a metaphor might cause if taken too literally.

A common set of misunderstandings can occur when the linguistic notions of a language family are uncritically extended or abruptly projected onto other modes of investigation. Is it correct to assume that linguistic relationship necessarily correlates with biogenetic-racial or cultural-ethnic factors? After years of quandary, the answer is now clear that while language and culture typically do show pervasive synchronic interconnections, there is no guarantee that all speakers in what is considered a single language family will share any particular biological or cultural traits or that outsiders will lack them. Clearly also the time-depth of a given language family and its contact history will need to be taken into account.

Just as Swedish and Finnish people share many cultural and biogenetic attributes but speak languages classified in different families (Indo-European and Finno-Ugric), so in Mainland Southeast Asia: speakers of Khmer clearly have much culturally in common with Thai and Lao people, although Khmer is assigned to the Austroasiatic family, not to Tai. In terms of physical appearance, typical Swedes and Finns would seem closer than, say, Swedes and Bengalis, although it is the latter two groups that are distantly related in the same Indo-European language family in a way that Swedish and Finnish neighbors are not linguistically related.

If linguistic, cultural and biogenetic relationships do not necessarily pattern together, should we then simply discard the concept of language family entirely, seeing it as an outmoded and potentially deceptive metaphor? This may depend on disciplinary or political perspectives. Anthropologists confronted with Southeast Asia as an interactive cultural complex of long standing have recently been inclined to entertain this possibility, while linguists more involved with the technical specifics of the Comparative Method of historical linguistic reconstruction may be less disposed to reject the model entirely. Much of the difference may lie in the tendency of some to attribute to linguists' family-tree models more than the linguists would intend. Intentional or inadvertent political chauvinism or hegemonic sociocultural control is not the objective of a linguistically defined family tree but, like unwelcome parasites, they can easily attach themselves to it.

We argue here that a plausible approach is to accept language families as valid reifications only to the extent that their implied claims are based on the Comparative Method of historical linguistics, the method that accounts for their construction. In this technically based view, the Tai Language Family is seen as the handiwork of the linguistic Comparative Method as applied to a particular set of languages—in this case, those we now usually refer to as Tai. The method clarifies and elucidates the linguistic relationships, based on the regularity of sound change, and may suggest features of an earlier common language, but does not imply other claims about specifics of history, culture, race or the like. Whether phenomena of these sorts may correlate to any extent or not with the linguists' family tree would be important empirical questions for practitioners of the various disciplines to investigate.

There remain important issues to explore with this terminology and its linguistic bases. Methodological problems increase when we recede back in time to hypothetical stages like 'Sino-Tai' or 'Austro-Tai'. Below in Section 9 we will consider both the utility and the limitations of the linguistic Comparative Method in establishing terminology and categories at this distant level of macrocomparison. The sections preceding will sketch some important issues in applying the Comparative Method to the more limited set of languages noted above now widely called Tai.

## 2. Nodes, not names

That the Comparative Method provides nodes, rather than names for nodes, may be more apparent to linguist-practitioners than to others outside the discipline. A hierarchically organized array of nodes on a branching tree-like structure would be the hoped-for result of applying the Comparative Method systematically, but the method itself provides names neither for the top node nor for lower nodes.

A name like 'Tai' then, should be taken as arbitrary as far as the linguistic techniques of reconstruction go and just where on the tree to apply a term such as this is similarly arbitrary. Fangkuei Li in his *Handbook of Comparative Tai* (1977) was instrumental in establishing 'Tai' in one particular sense, with his application of the Comparative Method referring nm Proto-Tai reconstructions to a particular node on a family tree. At levels more inclusive—and therefore presumably referring to more ancient nodes—names like Kam-Tai, Tai-Kadai or Kadai are now used more and more, as other contributions to this book will show.

Criteria for recognizing discrete Tai languages and dialects have varied but tonal systems are a good place to start. Nearly 200 distinct tonal systems have been reported, and at least 50 varieties have distinctive ethnonyms of one sort or another. About 60 tonal systems have been reported within Thailand alone; they seem to fall

loosely into four regional groupings, although this has not been very rigorously established. The existence of a large shared vocabulary, a high degree of regularity of correspondence relationships across virtually all tonal systems, along with other highly regular segmental correspondences, is responsible for the success of the Comparative Method as applied to the Tai languages and accounts for why most linguists are confident of the utility of a Tai grouping structurally, even if some may prefer a different name.

There is by now a growing professional among Western-based scholars to apply 'Tai' to a particular node in line with Li's (1977) usage, although colleagues in China typically use 'Zhuang-Dai' to label the same node. Chinese-based scholars might object extending the use of 'Tai' (or Dai in Pinyin transcription) to subsume speakers of Zhuang and Buyi (Bouyei) varieties who do not typically refer to themselves as 'Tai' - a point also considered significant by A.-G. Haudricourt (1948) in his research on the language family. Thus problems both with recognizing the term Tai for the language family in Li's sense and also with using the term Zhuang-Dai following Chinese authorities have more to do with political sensibilities than with substantive linguistic issues.

Thus one problem concerns what the Li's Northern-branch members of the family call themselves. The speakers in question include most of populations the Chinese refer to as 'Zhuang' in the Guangxi-Zhuang Autonomous Region and as 'Buyi' (Bouyei) in neighboring Guizhou; these terms refer to officially-recognized Chinese nationalities (i.e. minorities). The difference between Zhuang and Buyi would appear to depend on arbitrarily on Chinese administrative divisions, not on any particular linguistic facts.

Most of these speakers, when speaking their own languages, simply refer to themselves with the toponym referring to their locality or with some other distinctive qualifier, preceded by a form of *pu-* 'person'. Thus most of these speakers would only call themselves Zhuang or Buyi when speaking Chinese or when referring to themselves specifically the way Chinese authorities would. Nor do these people regularly call themselves (nor wish to call themselves) 'Tai' at all. (Luo Yongxian, p.c., who is a Zhuang speaker from Feng-Shan, adds that there is no regular cognate item meaning 'free' similar to *tai* in his language, nor any other regular plausible cognate form; nor is there in Zhuang dictionaries.) Note the form 'Dai', i.e. 'Tai' in Pinyin Romanization, has a special administrative meaning in Chinese writing. For a small Zhuang-related group not living in Guangxi whose speakers may call themselves *phu-tai*, see Pranee Kullavanijaya (2529/1986:1, 4).

Of course, naming practices could hardly be arbitrary in the wider context of political ideology. The language family to which many languages of Europe and northern India belong has been called both 'Indo-Germanic' and 'Indo-European',