

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',

the latter term felt to avoid possible nationalistic nuances or chauvinism some would attribute to the former one. The term 'Tai', spelled thus to differentiate it from 'Thai', the national language of Thailand—paralleled in the local Thai spelling distinction with and without the equivalent of a final *-y*—may be felt to have covert resonances of this sort.

A further complexity in the Thai/Tai case arises in Thai institutions, and occasionally elsewhere, when academics use 'Tai' and its *y*-less local counterpart in two distinct polysemous applications:

(i.) On some occasions, these terms are used inclusively to designate the entire language family, i.e. all the varieties under the Proto-Tai node, in a way similar to the usage of Fangkuei Li. In this usage, one could indeed say that Standard (Central) Thai is a Tai language.

(ii.) But for others, Tai and its *y*-less local counterpart would denote a language related to Standard Thai but a minority language spoken outside the current political boundaries of Thailand, e.g. White Tai. Following this usage, it would be contradictory to claim that Standard Thai could be a Tai language.

It is well to keep in mind that 'Tai' in an ethnolinguistic sense did not originate with Li or in modern academia. A form of 'Tai' (without *-y*) occurs some two dozen times on the Sukhothai inscriptions and frequently in inscriptions and texts thereafter. The term is especially common in calendrical expressions where specifically Tai forms of reckoning are contrasted with those of other ethnolinguistic groups, especially the Khom or Meng. Whether these latter refer to Khmers and Mons, as is widely held, is not at issue here: the significant point is that 'Tai' is used in a sense that clearly contrasts with these other terms. Part of the contrast directly involves linguistic forms explicitly presented in the calendrical expressions. Somewhat later a related contrast of *Tai-yai* / *Tai-noi*, the 'greater' and 'lesser' Tais, can be traced in traditional sources. Of course the historical usages could hardly inform us exactly how we should apply the term in the current context but at least the term has had a long and interesting indigenous development of its own and is not simply a modern academic construct. (The term 'Kadai', on the other hand, falls exactly into the latter category.)

3. Tai and Tai Studies

A fixed conception of 'Tai' reflecting to some extent the linguists' usage above has been given a boost by the institutional cohesion of Tai Studies or 'Tai-(khadi)-su'ksa' over the past two decades. This has been emerging as an established field of inquiry in a number of Thai institutions and elsewhere. In Thailand this could

be linked to an earlier 'khadi-chao-ban' or local cultural studies movement, which gained momentum in the 1970's. Agricultural practices, textiles and dress, music, art and drama, ritual beliefs and customs, architecture, folktales, historical accounts and all-important interrelationships among Tai social, economic and political organisation are now all finding a place in this new integrative field of study.

Of course, the comparative study, in Thailand, of Tai peoples and cultures is nothing new in itself. Nationalists and traditionalist academics of earlier periods, such as Luang Wichit Wathakan and Phaya Anuman Rajadhon, were important in defining and stimulating such studies in the past, especially within the Thai nationalist framework of the 1930's and 1940's—and Western scholars have had their input too. Partly because of improving foreign relations in the 1980's in the Thai-Chinese, Thai-Vietnamese and other domains, local scholars in Thailand have had the chance to travel and enlarge their 'Tai-su'ksa' data base. Coincidental with the political opening up has been an onslaught of globalization and rapid urban cultural change, with many urban and educated Thai people turning to seek their cultural roots. The institutionalization of Tai-(khadi)-su'ksa speaks to this need.

Two forms of enlargement in the 1980's and 1990's can be distinguished:

- (i) increasing interest in frontier areas within the larger Tai-Kadai grouping and
- (ii) greater inclusion of synthesis of comparative disciplinary perspectives.

But what is the place of the Comparative Method in linguistics in these movements? One answer is given by Merrilee H. Salmon (1996:203), who observes:

Anthropologists, archaeologists, historians sociologists and other social scientists all use some form of the Comparative Method to discover common causes or shared origins of phenomena they want to explain...Linguists enjoy advantages over ...other social scientists...because the latter have no analogue to the regularity of sound change that allows linguists to exploit sound-meaning correspondences.

Methodological interaction, emphasised by Salmon, is one obvious way in which historical linguistics affects other disciplines, but there may be other more hidden ways too. In the Tai-su'ksa case, the linguistic Comparative Method may be implicitly setting up what is taken to be 'Tai' for others to accept as a sort of bounding category. Thai is, it may be that language is assumed all along to lie behind what is taken as 'Tai'. It is this assumption that may make some anthropologists and other researchers uncomfortable.

This situation begins to uncover some dangers to be guarded against in the Tai-su'ksa movement. One danger is a tendency to ignore nearby Southeast Asian and Southern Chinese cultural traditions that would be considered non-Tai on strictly linguistic grounds. The fact that a cultural trait, belief or practice is widely found among Tai speakers need not imply that it is quintessentially, originally or exclusively Tai. A compensatory danger is to ignore what Salmon has called attention to above: the regularity of sound change as constitutive of the Comparative Method in linguistics. This sets linguistic relationships and definitions based on them apart from other patterns of similarity and dissimilarity. This returns us to the premise above that biogenetic and cultural similarities do not necessarily replicate linguistic relationships.

4. Regularity and linguistic reconstruction

What is 'Tai' linguistically raises the broader question of the status of the Comparative Method. A keystone of the Comparative Method is the regularity of sound change as established following from the Neogrammarians of the last century. As noted above, a major milestone in Tai historical reconstruction is the work of Fangkuei Li (1977), which appears to establish Proto-Tai as a very firm hypothesis indeed. From afar, the over 1000 cognate forms and their reconstructions seem to be nearly a clockwork example of the Comparative Method at its best.

On close inspection there are a number of exceptions to regular sound changes evident or implicit in Li's work. These are frequently mentioned in his notes. Tonal irregularity is one particularly common phenomenon encountered in the Handbook. One case is where certain cognate items show High-class (proto voiceless) initials in some languages but Low-class (proto voiced) ones in others, with predictable effects on modern tone. One way that Li suggests to account for this apparent irregularity is to propose voice alternation in the proto language—perhaps a morphophonemic process of some sort.

Another way to account for the same apparent irregularity would be to multiply the inventory of proto initial phonemes, a choice preferred by Gedney ([1979], 1989), who proposed an extra series with differential voicing effects on modern tone. Thus the Standard Thai word for *khā:w* 'rice' shows an alternation that led Gedney to postulate an initial voiced aspirate in the proto language.

These examples go to show that there need to be both craft and a degree of subjective judgment in applying the sometimes conflicting principles of the Comparative Method, especially when exceptions are to be dealt with. Sometimes more than one plausible solution can be proposed. Results then are not obtained through a strictly mechanical clockwork algorithm, appearances in Li's Handbook to

the contrary. Even Li changed his mind on certain reconstructions after the publication of his Handbook (Li 1989).

Recent work of Luo (1997) both extends the inventory of 'Gedney series' items and uncovers many additional cases of what would be irregularity, from the standpoint of Li's reconstructions. It is unlikely that Li would have been totally unaware of the magnitude of these residual items. Perhaps, owing to their irregularity, he chose not to include some of them in his reconstruction process, adding to the superficially textbook-like impression of his work.

5. Alternative historical models

Questions regarding the tree-diagram model of linguistic relationships as well as similar mother-daughter metaphors have been raised by some (e.g. Durie and Ross 1996, and papers therein). Factors seen to pose problems include incomplete or exceptional sound change, areal diffusion and other types of language contact, especially massive borrowing, long-term bilingualism and socially-sensitive bidialectal competence. All of these factors are of great importance in the Tai area, where there has been much language contact and where many speakers are presently bi- or multi-dialectal in competence. Apart from those whose first language is one of the two national languages Thai or Lao, most are bilingual in Vietnamese, Chinese, Burmese, Assamese or other non-Tai languages. In the past as well as the present, contact factors must have been crucial in Tai language history—including factors such as long-term bilingualism with the arising of contact varieties, hybrids, and perhaps creolisation.

Lutz (1992) is among those who propose more radical models. Concentrating on Semitic, Lutz even questions the very notion of a coherent proto language as a unitary linguistic variety spoken by a cohesive social group. Hence by implication he calls into question the family-tree model. He attributes the arising of Semitic similarities to early forms of contact, a process of dynamic diffusion he sees as characterised by polygenesis and entropy. Because of the relatively shallow time depth of Proto-Tai in Li's sense, probably at most two millennia, the Tai family tree could not be easily be implicated in the sort of analysis characterizing the markedly deeper time frame of data considered by Lutz. However if a project were to push back consideration of Tai to earlier stages (Sino-Tai, Austro-Tai) then the issues raised by Lutz could become compelling (see Section 9).

Dixon (1997), considering mainly Australian data, explores the 'punctuated equilibrium' adapted from evolutionary biology as a model to account for language change, but with an important ecological/geographical input. The family-tree model often taken for granted in traditional discussions of language history is decentred,

but not entirely eradicated, by Dixon. It becomes but one relatively quick phase in the linguistic evolutionary picture. Characteristic of periods of social upheaval and migration, the branching-tree development is held to alternate with longer, slower stable eras characterised by gradual areal contact and diffusion, more like what Lutz has in mind for the desert cultures characteristic of Semitic languages. While Dixon proposes this interplay of fast branching and slow diffusion for language groups generally, for Australian languages it seems especially apt because of Australia's ecological situation, with few impenetrable obstacles and with few documented social upheavals.

Where the lay of the land and political/military factors are different, such as in the Tai area, particular ecologies along with political and military configurations will interact to affect the working-out of local developmental stages. Thus we might expect one pattern of language change for upland Tai mountain valley communities relatively remote from neighbours, such as Northern Zhuang and Buyi (Bouyei), with another developmental pattern characteristic of larger river basins with wide flood plains and easier communications. Historical factors, such as evidence bearing on Tai (Zhuang/Nung) population turmoil at the time of Nong Zhigao's rebellion against the Song (c. 1050), or more recently the impact of Standard Thai on Lao and on other Tai varieties, might also be seen as supporting a form of the model proposed by Dixon.

Hand in hand with new more theoretical directions of the sort sketched above go new discoveries in the field, also set to contribute perspectives to comparative Tai studies. In the time since Li's Handbook, new data have become available of immense value to comparative Tai studies. To name just a few sources: the copious field notes of William J. Gedney edited by his students; Zhuang studies of Pranee Kullavanijaya and colleagues; and a range of Tai and Kadai publications written or edited by Jerold A. Edmondson, David B. Solnit and colleagues. An important field discovery made by Theraphan L. Thongkum and Peter A. Ross is the retention of voicing in etymological 'Low' series stops in Tai varieties of the Vietnam-China border area. This would be directly predicted by Li's reconstructions, but not firmly attested at the time he was working. The discovery thus serves to validate the Comparative Method as deployed by Li, very much as the inscriptional discovery of Hittite laryngeals in the 1920's was predicted decades earlier by the careful comparative Indo-European reconstructions of de Saussure.

6. Positioning proto-Tai

Proto-Tai could be interpreted either, methodologically, as an abstract set of linguistic relationships organized by the Comparative Method or else, substantively, as a concrete ancestral sort of Tai speech. The reconstruction of Proto-Tai is a

project directly in line with assumptions and techniques of Western linguistics as developed first in the context of the Indo-European languages. In general, the methods were simply applied to the Tai languages with few modifications. Jones (1965, 1966) however provides some critical discussion on this point, calling attention to literary evidence that has sometimes influenced reconstructions in a manner not entirely in line with strict practice.

Part of the method requires the selection of related languages from several branches for purposes of comparison. The major three-way division of the Tai languages by Li (1977) in reconstructing Proto-Tai has been mentioned above, but other schemes are possible. It is interesting that vocabulary distribution was responsible for Li's original three-way classification, although several important phonological isoglosses also follow the division.

A degree of consensus now follows the usage of Li (1977), who groups the Tai family in to the three branches mentioned above: Northern, Central and Southwestern. Luo (1997) has raised the possibility of a separate Northwestern branch. (It is crucial to distinguish 'Central Tai', e.g. Nung and southern Zhuang varieties, from 'Central Thai' of Thailand.) Exactly how to further subgroup what Li calls the Southwestern branch has not been completely resolved, as criteria have differed. Brown (1965), Hartmann (1980), Chamberlain (1972), Luo (1997) and others have considered various proposals in detail.

The terms Diac, Kam-Tai (Kam-Thai), Tai-Kadai (Thai-Kadai) and (in Chinese research) Zhuang-Dong have been used for a very inclusive family grouping, including varieties more remotely connected. Several of the latter languages have been subgrouped as Dong-Shui or as Kam-Sui; others as Kadai, although increasingly Kadai has been used to refer to the larger group as a whole. As of the time of writing, the larger inclusive group would represent well over 80 million speakers.

Not all scholars have followed this usage. As noted above, A.-G. Haudricourt (1948) has suggested restricting the term Tai even more—essentially to only those languages whose speakers typically refer to themselves as Tai, viz, to the Southwestern branch of Li's Tai family along with some Central languages. Even nearer to Thailand, Lao speakers do not normally refer to themselves Tai.

On the other hand, in terms of sheer numbers, it is unquestionably the case that the majority of speakers of Tai languages in Li's sense would in fact refer to themselves as Tai, Thai or with a cognate term; perhaps this argument from numbers is sufficient justification to impose the name on the rest? Problems of this sort—politically sensitive (or insensitive) namings of language families—are hardly avoidable.

Early serious work on comparison of Tai dialects can be seen in some 19th-century wordlists. In the pioneering work of Grierson (1903) and Maspero (1911) we see the first extensive comparative study of a number of Tai dialects along the lines of the Comparative Method. Important steps were taken by Haudricourt and Li, during the 1930's and 1940's, that laid the foundation for the reconstruction of Proto-Tai, effectively establishing Comparative-historical Tai Studies as a subdiscipline. In 1956 Vichin Panupong (Chantavibulya) wrote the first rigorous work in Thai on a Tai dialect (that of Songkhla). This was followed by critical developments of the following decade, including work of Haas (1958), Henderson (1959), Egerod (1961), Gedney (1964), Jones (1965) and Brown (1965). More recently many others have contributed both to the Tai data base and to the theory and practice of how to compare Tai dialects and how to reconstruct Proto-Tai.

Li's (1977) work incorporates much progress made in these earlier studies, and as well he relies heavily on his own valuable field data on Tai varieties in China, where he did the pioneering work. His version of Proto-Tai is rich in initials, including many complex clusters. Also proposed are rich vowel contrasts, but vowel length is not considered conservative; see also Sarawit (1973) for another proposal and Strecker (1983) for comparison. Most have followed the four Proto-Tai tones reconstructed by Li, which relate in a rough way to the four orthographic categories of the Thai-Lao writing system: unmarked open words (Li's tone A), words marked *máy-è:k* (B), *máy-tho:* (C), and unmarked closed (D). In this diachronic scheme, physiological mechanisms associated with initial consonant articulation (mainly voicing) produced splitting of the four earlier Proto-Tai tones; initials then merged and (new) tonal systems shifted in different ways, giving rise to varied phonological systems and to patterns of regular correspondence among modern daughter languages.

Exactly when these developments occurred is impossible to specify through the Comparative Method alone, but the incorporation of loanwords into Thai, especially from Malay, can provide useful clues. Pairs like *durian* (Malay)/ *thurian* (Thai) or *sago/sa:khu:* would appear to date the Thai devoicing change to a time after Malay contact. (Indic loans are less satisfactory for this purpose as they may have entered Thai from intermediate languages in which they were affected by prior sound changes, with spelling remaining conservative.)

Another type of chronological evidence comes from spelling variation on Thai inscriptions. Voiceless nasals provide a good example. Vocabulary items with sound nasal *m-* in modern Thai are from at least two sources: (i) from an etymological voiceless nasal **hm-* (High tone class items) and (ii) from a plain nasal **m-* (Low tone class items). Modern Thai spelling is generally conservative, so the orthography retains direct representations of **hn* and **m*, while the formerly different consonant sounds have now merged as plain /m/, but now with proper tonal

differentiation. Thus *mù*: ‘group’ is spelled with cluster *hm*, the silent *h* (*ho*’:-*nam*) now interpreted as indicating tone class. The Wat Sorasak Inscription dated 1417 spells this item with a plain nasal *m*-. This suggests that at that time the merger was in progress, along with an orthographic interpretive shift. The orthographic cluster *hm*- was not yet fully recognized as a marker of tonal class rather than of consonantal quality (Fine Arts Department 2526/1983:128).

Jones (1965), Brown (1965), Chamberlain (1972), Hartmann (1980) and others have proposed sub-branching structures for the Southwestern Tai languages. Criteria, and therefore results, have varied somewhat. Some have taken the current status of the Low orthographic series of consonants as criterial. If so, Southern and Central Thai and Lao dialects would group together (with Low-series stops aspirated and voiceless), with the remaining varieties grouping together in a different sub-branch (with Low-series stops unaspirated voiceless).

For modern varieties, particularly in Thailand, an added difficulty both for theory and for practice is the strong influence of Standard Central Thai on regional varieties and the widespread bidialectal language competence of many upcountry speakers. Within Thailand there is substantial diglossic dialect mixing; thus Lao dialects in Central Thailand have been influenced by Central Tai tones (Wilaiwan Khanittanan 1973) while a stable form that could be called urban hybrid Southern Thai is constituted by a mixture of Standard (Central) Thai lexical items and segmental phonemes which are assigned — through regular rule—local (Southern) tones (Diller 1987).

These issues can affect field investigations crucially. M.R. Kalaya Tingsabadh (1988) reports that in some local dialects in central Thailand speakers do not show the ability to produce independent citation forms in their own dialect; rather, they revert to forms essentially of the standard language. This produces an observer’s paradox if a linguist relies on isolation-form evidence only in field investigations.

7. Widespread Tai features

Although the Comparative Method, based as it is on regular sound change, is responsible for determining a tree structure for the Tai family, other linguistic features have their patterns as well. Specific vocabulary items and other linguistic features involving syntax, etc., may, or may not, be distributed in a manner similar to arrangements that the process of phonological comparison and reconstruction produces.

It is important to emphasise that many of the features noted in this section have an areal distribution or are affected by forms of contact, and so are more like

diffusible cultural artefacts. In tracing Tai linguistic relationships, it is useful to *distinguish rather sharply the criterial regularity of sound change*, the dynamo of the Comparative Method, from other sporadic and even haphazard distributions of other linguistic features, such as classifier phenomena (Matisoff 1992).

Typologically, all Tai languages found so far are tonal. Systems of three to eight contrasting tones have been reported. For the related Kam-Sui languages, nine contrastive tones are characteristic of some varieties of Kam (Dong). The tone systems are remarkably regular in terms of systematic correspondences, as are initial consonants and clusters. Final consonants are especially stable. As noted above, it is this phonological regularity more than anything else that has made the enterprise of Tai historical linguistics a viable one. None of the non-Tai language families of Southeast Asia would show the particular totality of sound changes characterising the Tai languages, although some diachronic phenomena, such as devoicing with effect on pitch or tone, are found widely throughout the area.

Other similarities give cohesion to the language family, although many of these structural features are shared with a number of nearby Southeast and East Asian languages and so are not definitively Tai. Tai languages are mainly monosyllabic, Standard Tai and Lao being partial exceptions because of borrowed vocabulary, and all Tai languages are completely lacking in inflectional morphology, as opposed to derivational morphology, which could be taken as represented by compounding. Modifiers typically follow modified items in the Tai languages, but there are a few exceptions mainly due to contact.

Virtually all Tai varieties, and other most Mainland Southeast languages east of Burmese as well, could be characterized as at least loosely subject - verb - object (SVO). The Tai language Khamti of India is at least partially SOV and represents an exception due to contact (Wilaiwan Khanittanan 2526/1983). Other orders are used for pragmatic marking or discourse-sensitive purposes.

Numeral classifiers are used in all Tai languages, but classifier semantics and syntax vary quite markedly and most neighboring non-Tai languages show similar systems. Standard Thai has proliferated an original much simpler classifier system for stylistic purposes and for sociolinguistic marking (Juntanamalaga 1989). The order NOUN + NUMBER + CLASSIFIER is usual for the Southwestern members of the group, whereas NUMBER + CLASSIFIER + NOUN occurs to the north. In some in-between languages, such as varieties of Black Tai in Vietnam, both orders are used with pragmatic difference. (I am indebted to Cam Trong for this information). Even Central Thai admits certain special set phrases which appear to follow the northern order.

Tai languages share a number of widespread Southeast and East Asian grammaticalization patterns, such as the use of verbs meaning ‘give’, ‘go’, ‘come’, ‘get’, ‘ascend’, etc. in auxiliary-like or other functional usages (Clark 1978; Matisoff 1991; Diller 1993; Enfield, this volume).

Lexically, Li’s work shows that some thousand words are shared at least by two out of three of the branches for Tai he proposes. Strangely, certain lexical categories that one might regard as basic do not show good pan-Tai cognates. These include question words (like ‘who’, ‘what’, ‘why’, ‘how’), which vary considerably, as do many conjunctions, prepositions, preverbal auxiliaries and common verbs like ‘speak’, ‘look at’, ‘return’ and even ‘go’ and ‘be’. In some cases the variation is partially explained through differential grammaticalization as main verbs are demoted to preposition-like or auxiliary-like words, etc.

Pronoun, kinship and deictic systems show considerable variation, but with a few core cognates usually shared. Even body-part terms, including words for ‘hand’, considered basic vocabulary by many linguists, vary somewhat—some perhaps for taboo-associated reasons. In terms of lexicon, modern Standard Thai is a rather poor representative of the Tai language family: it has demoted many original Tai words to the category of vulgar or taboo, or has replaced them with Mon-Khmer, Indic, Taeciw (or even English) substitutes.

Some dozen Southwestern languages have traditional alphabetic-syllabic scripts of Indic provenance, and this may affect their status as distinct languages in the eyes of their speakers. Some groups in Vietnam and China formerly used specially-adapted Chinese characters for local writing; recently Romanizations have been introduced in these areas. (For more on the history of Tai writing systems, see Nantana Danvivathana 1987; Hartmann 1986).

8. The homeland problem

Most linguists working in the field would probably accept a time depth for Li’s level of Proto-Tai of less than two millennia. A locus for Proto-Tai at this level—a Tai homeland—remains controversial and cannot directly be established by invoking the Comparative Method as narrowly confined to mapping sound change. On the other hand, evidence is good for a cohesive homeland area, because the rather high degree of shared lexicon and regular correspondences in the Tai languages taken as a whole would be difficult to explain without it. If there had been no rather cohesive area to which Tai varieties (in Li’s sense) were confined, say, some 1500 years ago, then the degree of similarity among the modern Tai languages would be very unusual, in terms of what is known about other language families of the world. Tai would then represent a peculiar and inexplicable exception. There

might be a similar case for a somewhat later secondary homeland for the Southwestern branch of the family, as these languages show even stronger cohesion among themselves.

Chamberlain, Cholthira Satyawadhna, Edmondson, Hartmann, Placzek and others (this volume) have made useful contributions to the homeland problem and the reader is referred to these studies for more detail. Philological evidence such as how flora, fauna and agricultural practices are referred to would not rule out the region encompassing parts of the watersheds of the Xi River system in Guangxi extending to northern tributaries of the Red River in Vietnam.

One important piece of evidence in line with this hypothesis is the relatively higher degree of linguistic variation among dialects in the Northern group—especially the case if the contiguous Dong-Shui, Li, Lati and other cousin-like Tai-related languages are included. Everything being equal, one expects to find greater variation over a smaller area where a language family has been represented for a longer time—compare for example British Isles with Australia in terms of degree of language variation—although geography, forced migration, contact and other factors surely complicate this. Clearly, rugged terrain and isolated valleys of the upper Xi system would promote and preserve dialect differentiation more effectively than the great lowland Chao Phraya plain.

In terms of folklore, the present Zhuang and Buyi speakers in Guangxi-Guizhou—who number many millions—appear to lack oral traditions of having migrated from elsewhere, but this is true for other Tai groups as well, and can hardly be taken as convincing evidence. For further discussion of more historical evidence, see Kanchani La'ongsri (2528/1985) and Terwiel (1991). Also, Taylor (1983), Pulleyblank (1983) and Barlow (1987) discuss Chinese sources with insight and their observations generally confirm the Xi-River/Red-River view suggested above, although other possibilities remain open.

Chinese written evidence relating to early ethnography requires extreme caution. Some have made rash leaps of faith—e.g. relying on modern Mandarin pronunciations of characters written many centuries ago which may have undergone semantic change—leading to quandaries or folk-etymologizing. A similar process gave rise to speculative imagining that Tais migrated from the Altai Mountains—a supposition still to be encountered in certain Thai circles.

Wyatt (1984:6-11) has suggested on historical grounds that for the later Tang period (c. 800 A.D.) ancestors of the Southwestern Tai may have already been somewhat to the south of the Red River. This would have included the region around Mu'ang Thaeng (Dien Bien Phu; see also Chamberlain 1972). Certain

pervasive folk traditions (as reported to the writer by Cam Trong) especially among the Black and White Tais, may be seen as providing some support for this possibility.

The weight of this evidence will surely erode traditional support for the Nan Chao theory that Tai speakers were dominant in that Yunnanese Nan Chao kingdom, a theory with no dependable supporting linguistic evidence.

9. More ultimate Tai relationships

Most linguists working in the field are quick to distinguish the Comparative Method as applied quite strictly by Sarawit (1973), Li (1977), Gedney (1989), Luo (1997) and others from methodologically looser historical comparisons pointing further and further back to greater time depths than Proto-Tai in Li's sense. In general, and especially the case for relationships considered here, the deeper the time depth, the fewer the regular cognate sets and the greater the difficulty of applying the Comparative Method in its strict sense. One can only conclude: the more problematic the long-range genetic relationships postulated on the basis of such attenuated evidence. Yet, as Matisoff (1990) has convincingly documented, the Tai languages have been prone to megalocomparisons of this sort.

9.1 Problems with typological comparison

At the time of writing there is certainly no general consensus among comparative-historical linguists as to whether (or how) the Tai languages are to be related externally to other language families or stocks. The main candidates for relationship have been Sino-Tibetan (Chinese, Burmese, Lolo, etc.), Austronesian (Malay, Tagalog, etc.) and there is need to consider Austroasiatic (Mon, Khmer, Lawa, etc.). The fact is that no external relationship has yet been conclusively demonstrated, using the normal rigorous methods of historical linguistics. Nor should one automatically discount the possibility that some level of Tai/Tai-Kadai represents a separate family.

Most of the problem involves criteria and methods used to establish genetic relationship. Are we to rely mainly on typological structure, on shared vocabulary or on the establishment of regular sound laws? Are we to compare modern languages and note sporadic similarities or are we to reconstruct proto-languages and compare them systematically?

Typological characterizations of languages are important, but as noted in Section 7 since typological features often have areal distributions and are subject to shifts and changes, this sort of analysis does not necessarily establish genetic connection in a conclusive way. Take the case of (i.) [NOUN] + [ADJECTIVAL

QUALIFIER] word order, and its opposite order: (ii.) [ADJECTIVAL QUALIFIER] + [NOUN]. We need go no further than Spanish and English to see these respective orders. Since Spanish and English are related languages of the Indo-European family, clearly the order of adjectival qualifiers is not a shared feature of all the related languages in the family. Even a single language may employ both patterns, e.g. Russian or French.

In the case of the Tai languages, as well as most Austronesian and Austroasiatic languages, the first order above is predominant for adjectival qualifiers, but in Chinese, the second. However, as the Spanish-English case shows, this fact alone cannot establish anything conclusively about genetic grouping.

Other syntactic features, such as subject-verb-object or subject-object-verb basic word order are likewise unsuitable as absolute genetic criteria. The Indo-European languages show both of these orders and in some branches, such as Italo-Romance, basic typological shifts have occurred.

Morphologically, the Austronesian and Austroasiatic languages show strong tendencies toward bisyllabic word structure, prefixing, infixing and, especially for Austronesian, verbal morphology (e.g. to mark transitivity). These features are not characteristic of Chinese or Tai, which lack verbal morphology and tend, at least apparently in earlier stages, to be monosyllabic. Are we to accept these facts as important genetic indications, or not? Again, evidence from other language families casts doubt on features like these as criterial in any absolute sense.

Nor is simply being a tonal language or not proof of genetic connection. We know that tone systems can develop in languages which were previously not tonal. This has occurred (and is presently happening) in the Austroasiatic family, and even the Austronesian family has a few tone languages. In the Austroasiatic family some languages are non-tonal, some partially so or with a register distinction (such as Khmer) while some, like Vietnamese, have become completely tonal. The Sino-Tibetan family similarly varies. A number of languages and dialects close to Tibetan are non-tonal, or only beginning to develop tones. In the Southeast Asian area, tones appear to arise in association with changes in initial or final consonant features; in general, as consonant sounds fall together as indistinguishable, tonal distinctions become more important. Also, it appears that tones can be lost, as has apparently happened in Korean. So tone cannot be taken as an indicator of genetic relationship in any absolute way.

Shared vocabulary and regular sound correspondences are typically involved in establishing genetic relationship, but it is often difficult in practice to determine whether lexical items should be considered inherited words or borrowed ones. Regular sound correspondences can be the product of either borrowing or

inheritance, to follow the traditional distinction. Of course careful comparative work can often clarify these relationships, but not always. How much vocabulary needs to be shared, and how regular do phonological relationships have to be to establish a genetic relationship conclusively? Answers to these questions have varied greatly and there is no real present consensus among linguists.

9.2 The classification of de Lacouperie (1886)

The latter part of the 19th century saw several conflicting attempts to classify the languages in East and Southeast Asia into macro-families or stocks. In some cases little-known languages were classified on the basis of very scanty evidence—perhaps a traveler’s hurriedly-written list of ten or twenty words. Even for the better known languages, comparisons were generally a matter of appealing to vague typological features or exhibiting lists of similar-looking words.

A representative early attempt at macro-classification was that of Rev. J. Edkins (1871), a missionary with many years’ experience in China, who proposed that Chinese was related to Semitic, or rather that the languages of Babylon and Chinese belonged to one and the same family. He also claimed on the basis of the organic metaphor that the Chinese thus belonged to the Hamitic race. This was not widely accepted, although a Babylonian/Sumerian - Chinese link of a different sort was promulgated by A. Terrien de Lacouperie in 1886 (de Lacouperie 1886:115). The latter used Edkin’s data and basic line of argument to reach different conclusions.

De Lacouperie is responsible for the first comprehensive attempt to put the East and Southeast Asian languages into an all-encompassing genetic scheme. He also seems to be responsible for originating the Nan Chao theory mentioned above. This was picked up by Dodd and others, later popularized in Thai by Luang Wichit Wattakan, came to be taught in Thai schools as part of a nationalistic curriculum and thus even now is difficult to expunge. Although neither de Lacouperie’s language classification nor the Nan Chao theory would receive much support today from serious scholars of linguistics or history, his ideas are still worth considering in terms of method and categories of analysis used. In particular, it is instructive to consider the presuppositions behind de Lacouperie’s use of words like ‘mixed’ and ‘pure’, even though this analysis may recall unfortunate Nazi preoccupations of a related sort that followed his work.

The main features of de Lacouperie’s approach are (i.) attention to details of syntactic typology; (ii.) extensive appeal to language mixing, both of lexicon and of syntactic typologies; the approach included arranging languages on a hybridization scale from pure or unmixed, through mixed, hybridized, and final to fully hybrid; (iii.)

consequent unwillingness to construct a unique branching taxonomic structure to account for linguistic diversity in the area under consideration.

Although many of the particular claims as to patterns of mixing or hybridization among Southeast Asian languages advanced by de Lacouperie—especially when these are linked to race—would be objectionable in the light of more recent evidence, his general interest in what we would now call areal features and language contact is again surprisingly modern in essential tone.

In fact, de Lacouperie's approach prefigures contemporary critiques of the Comparative Method and he should be given credit for raising over a century ago, in the Southeast Asian context, many of the issues of current interest to Dury & Ross (1996), Dixon (1997) and Lutz (1993) and other researchers as noted above. Attention to what we would call today syntactic typology involved a remarkably rigorous scheme quite similar to widely-accepted methods in modern typological studies. In particular, de Lacouperie's work prefigures that of Greenberg and of more recent typologists such as Hawkins and Vennemann. Three of the most salient features used by Greenberg in his typological analysis are present in de Lacouperie's scheme: characteristic SVO, SOV, etc. predicate orders; relative order of genitive and possessed noun; relative order of adjective and qualified noun. Interestingly, De Lacouperie refers to these parameters in numerical designations which provide typological profiles of different languages (somewhat opaquely referred to as a language's 'ideological indices'). Thai, by this scheme, is "2-4-6-8": i.e. noun + genitive; noun + qualifier; SVO order (actually VO and SV).

For him, both (Mandarin) Chinese and modern Thai are mixed languages, while Mon (Peguan) is one of the few unmixed original ones. Interestingly, in the setting up of language stocks, the Tai languages, or at least Thai, are classified in effect as essentially *Austroasiatic* (although that term was not used until later), i.e. along with Mon-Khmer: the family name given is Mon-Taïc. The Tai family proper is referred to as Tai-Shan, and is a member of the Indo-Pacific Stock, which includes Mon-Khmer (Austroasiatic).

The Tai-Shan linguistic formation has taken place in historical times in Pre-China. It has evolved from the intermingling of southern languages belonging chiefly, though not exclusively, to the Mon type, with Chinese and other languages of the Kuenlunic family...The Tai-Shan vocabulary is thoroughly mixed, to such an extent that one-third of its words are common with ancient words of the Mandarin Chinese (1886:111).

The Kuenlunic Family mentioned above is de Lacouperie's designation for what we would now call Sino-Tibetan, however Chinese was also seen as a mixed language, or as he says: "The modern Chinese are a hybrid race, and their speech is a

hybrid speech” (ibid). This is claimed due to the mingling of the Bak invaders’ tongue and the original Indo-Pacific languages. The Bak, from Persia, are said to have imported “many words of the Akkado-Sumerian and Babylonian civilization.”

Returning to Tai, de Lacouperie reflects on the fact that Tai languages (unlike most Austroasiatic languages) are tonal. He proposes Tai “developed tones originally as a compensation by natural equilibrium to phonetic losses undergone in the everlasting process of intermingling” (1886:67). We see here the beginning of an explanation for the origin of tone through loss of consonantal distinctions prefiguring that proposed and developed by Haudricourt, Jones, Brown, Matisoff, Li and others.

It is of interest that de Lacouperie in effect presents two classification schemes—one applying to the state of East and Southeast Asia before the Chinese (Bak) invasion, and one subsequent to that, in which almost all varieties are mixed or hybridized. Note that this in effect denies the applicability of a single family-tree genetic representation for these languages. Interestingly, it raises the sorts of issue discussed by Dixon (1997) in developing his model.

We are confident today that virtually all of the Mon-Khmer cognates exhibited by de Lacouperie as evidence for an Austroasiatic-Tai connection (e.g. items like *thale*: ‘sea, lake’) are fairly recent loans. We can deduce this both from their lack in Proto-Tai and from their wide presence as cognates in other Austroasiatic languages. Data of this sort were perhaps not available to de Lacouperie.

In spite of this, there do seem to be intriguing Tai-Austroasiatic lexical connections which were not noticed by de Lacouperie. The following Thai forms were represented in Proto-Tai (Li 1977)—i.e. they could not merely be a matter of later borrowing into Southwestern Tai languages. These forms also show plausible pan-Austroasiatic cognates as well, and/or occur in Khmer or Mon inscriptions predating Tai contact, as one supposes along normal historical lines.

sa:n	‘to weave’	ráp	‘receive’
khlo:ng	‘canal’	tha:ng	‘path’
kát	‘bite’	mày	‘new’
khaeng	‘hard’	lo:y	‘float’
phà:	‘split’	kho’:	‘neck’
khi:	‘ride’	ta:	‘old man’

(Analysis of forms such as these is planned for future work. For convenience, I use a straight-forward adaptation of the Royal Institute and Haas transcriptions.)

Vocabulary of such earlier strata — which varies somewhat in terms of how convincing cognate forms are—needs to be distinguished from Mon or Khmer borrowings into Thai varieties in Thailand at a much later date. This latter vocabulary (words like *thale:*) appears in (Central, etc.) Thai but is lacking in elsewhere in Tai varieties and cannot be assumed for the stage of Proto-Tai. For the Austroasiatic evidence, see Huffman (1977); Jenner (1981); Shorto (1971). Until more analysis is done—especially until a convincing version of Proto-Austroasiatic becomes available—it is premature to attempt to set up regular correspondences for vocabulary of this sort. 79?

A complication is that certain forms, including *khlo':ng* 'canal', *nga:* 'ivory', according to Norman and Mei (1976) were borrowed from Austroasiatic into Archaic Chinese as well. In this Norman and Mei counter traditional Chinese chauvinistic attitudes which deny that 'barbarian' languages could ever loan any important vocabulary into Chinese. Other plausible Chinese-Austroasiatic-Tai convergences involve the items *phà:* 'split', *khi:* 'ride', and *càp* 'catch'. Note the link here with de Lacouperie's century-old proposals.

Furthermore, many words like cognates of *khàe:k* '(stranger), guest', *phae:* 'raft' etc. occur in Tai, Chinese and Vietnamese. Vocabulary of this sort once led to a hypothesis connecting Tai directly with Vietnamese. This was advanced by H. Maspero, who considered Vietnamese a Tai-Chinese 'mixture', a position accepted by Schmidt (1926) and by others for a time. However Haudricourt (1966) has shown quite conclusively that Vietnamese is instead an Austroasiatic language with no direct relationship to Tai, but with many Chinese loans, some shared by Tai as well. There thus appear to be different strata of vocabulary sharing that need to be distinguished.

9.3 Schlegel (1902), Benedict and the origin of the Austro-Tai hypothesis

In 1902 Gustave Schlegel in his *Siamese Studies* proposed a connection between the Tai languages and Austronesian (Malayo-Polynesian). For years this hypothesis lay dormant until revived almost simultaneously by K. Wulff (1942) and by Paul K. Benedict (1942, 1966, 1975; see also Thurgood 1985). Benedict's work is similar to de Lacouperie's in its interest in what Matisoff (1990) has called megalocomparison: the building up of ever-encompassing linguistic stocks of great time depth. Benedict generally rejects massive mixing of stocks as a general explanation and is more guided by the family-tree metaphor, with some use of the notion of substrata. Over the years through Benedict's various publications, his constructed Austro-Tai macro-family has differed greatly. It has variously included, and then excluded, Austroasiatic. In the latest version there is said to be an Austro-Tai substratum in Austroasiatic, accounting for certain lexical and typological

similarities. In the latest version (1990), Japanese has been added to the family as well (see also Thurgood 1985).

On close reading, a genetic connection between Austronesian and Austroasiatic is actually to be found in the work of de Lacouperie (1886:130). He defined an Indo-Pacific Stock, two branches of which were somewhat opaquely called Indo-Chinese (Austroasiatic plus Tai, but excluding Chinese) and Inter-Oceanic (essentially Austronesian). The Austronesian-Austroasiatic connection was taken up and developed by W. Schmidt (1906, 1926), who referred to the family as Austric. Tai was excluded from Schmidt's Austric and grouped with Chinese.

Schlegel and Benedict base their arguments on lexical similarities and call attention to some lexical items which do indeed appear common to Tai and Austronesian. These include a few plausible comparisons such as the following, in which total developments in Thai have not been pursued.

Thai		Proto-Tai (after Li 1977)	Proto-Austronesian (after Dyen 1971)
ta:y	'die'	tra:y	matay
ta:	'eye'	tra:	mata
fay	'fire'	vay	apuy
ku:	'I'	ku:	aku
du'an	'month'	'blu'en	bulan ('moon')
dò:k	'flower'	'blo'k	burak

Unfortunately for the Austro-Tai case, many additional far less convincing relationships are presented by Benedict (1975, 1990), who not infrequently resorts to loose resemblances, semantic leaps and to a practice known as 'proto-form stuffing'—the making up of maximal earlier forms to account for all desired modern cognate relationships. An important part of this hypothesis is the assumption of a fairly hefty borrowing of Austro-Tai vocabulary into Archaic Chinese, including rather basic cultural vocabulary. This resembles the Norman - Mei hypothesis for Austroasiatic as above.

Although items like those above suggest some etymological relationship between Proto-Tai and an Austronesian language, Benedict fails to convince most linguists who have specialized in Tai historical linguistics that the relationship should be treated as a general genetic one, rather than as some form of early contact, borrowing or the like. Leaders in the field such as Li have remained unconvinced, and Gedney (1976) has criticised the Austro-Tai hypothesis on a number of counts, including complete failure to account for Tai tones; see also the critique by Hartmann (1986). Dyen (1971), from the Austronesian side, showed that using

Benedict's loose resemblance techniques Austronesian could be just as well related to Indo-European. Finally, Diffloth (1977), from the Austroasiatic side, has expressed doubts about Benedict's claim that Austro-Tai has formed a significant substratum in Austroasiatic.

So-called Kadai languages including Li/Hlai varieties spoken on Hainan Island and Laha (in Vietnam) have a particular importance in Benedict's construction, since they show more transparently a good deal of Austronesian vocabulary (such as number terms) not present in other Tai languages. These languages, for Benedict, are in a sense pivotal between Tai and Austronesian: in his hypothesis they would seem to represent a much more conservative (lexically more Austronesian) form of Tai.

Of course, even if it were clear that the so-called Kadai languages were not particularly conservative, this would not deny the Austro-Tai hypothesis, but merely might cast doubt on certain alleged evidence for it. More positively, Blust (1988) has recently argued for the possibility of a set of monosyllabic Austronesian roots that code semantic notions widely throughout the family, but usually appear with different prefixes, etc., as words in the different languages. There are some striking similarities between a few of Blust's monosyllabic roots and Thai (or Proto-Tai) forms, such as the following:

Thai		Proto-Tai (after Li 1977)	Proto-Austronesian root (Blust 1988)
ka:ng	'spread'	kang	kang
tàt	'cut'	tat	tas
pìt	'close'	-	pet
sâe:k	'insert'	-	sek

More roots of this sort would go far towards clarifying an Austronesian-Tai connection, although some of these roots appear to be represented in Sino-Tibetan as well.

9.4 *Conrady (1896), Grierson (1903) and Sino-Tai*

Conrady (1898), redefining de Lacouperie's terminology, set up an Indo-Chinese family with Chinese, Tai and Tibeto-Burman branches. Grierson followed and extended this classification, unequivocally placing Chinese and Tai together in what he called the Siamese-Chinese language family. He was also among the first comparative Tai scholars to discuss regular sound correspondences among the Tai languages, rather than merely assembling lists of similar-seeming words. Grierson made no attempt to demonstrate the Tai-Chinese relationship, but rather assumed it, probably on the basis of typology and shared vocabulary.

Schmidt (who in 1906 had proposed the Austric hypothesis) also accepted the Tai-Chinese connection in his encyclopaedic and influential 1926 language survey, and Wulff (1934) and others further developed the relationship. More recently Li (1976) published a list of correspondences between Archaic Chinese and Proto-Tai in support of his contention that the two were genetically related. E.H.S. Simmonds, writing in the 1969 edition of the authoritative *Encyclopaedia Britannica*, supported the Sino-Tai hypothesis, calling Tai “an important branch of Sino-Tibetan”. However in later editions of the *Encyclopaedia*, articles by Egerod and Haas have expressed skepticism; even Li himself, writing in the 1974 edition, described Tai as “traditionally related to Sino-Tibetan” but admitted that the relationship had “never definitely been established”. Li’s assessment remains the case today, although Prapin Manomaivibool (1975, 1976), Luo (1997) and others have made progress in clarifying important issues.

Karlgren noticed that some Tai items such as cognates for *khra:m* ‘indigo’ had Ancient Chinese cognates with simple initials (e.g. **lâm*, Karlgren 1964 item 609k), indicating a quite early etymological relationship—i.e. to allow Archaic Chinese or all varieties of Chinese to simplify subsequently. More recently Prapin Manomaivibool (1975, 1976) has supported a richer form of the hypothesis, in which various layers or strata of Chinese-related vocabulary are recognized and analyzed for Tai. Thus the forms related to *plù’ak* ‘peel, bark’ and *phiw* ‘skin’ are both seen as cognate to the archaic Chinese reconstructed form **pljag*—the former representing (she suggests) a genetic connection with the latter taken for a later Chinese-to-Tai borrowing from a time when the Chinese form had undergone sound change and was pronounced more like **pju*. This work (1975) also presents an impressive list of over 600 Tai-Chinese cognate forms, although only about half of these conform rigorously to rules of regular correspondence set up by the author (perhaps indicating again different strata). Luo (1997) expands this list and establishes more regular relationships.

9.5 Genetic relationships or complex contact?

As suggested above, some lexical items are shared by two three or even four of the main language families in the Southeast Asia area, giving support to a notion of areal vocabulary for these and similar items. For example, is Archaic Chinese **nuo*: ‘crossbow’ to be linked to Old Mon *nga:*, to Tai *nâ:*, and to Malay *panah*?

Note that the crossbow is just the type of artefact that one might expect to diffuse widely—along with its name—over early Southeast Asian language communities, especially among those which survived by mastering its technology. It is reasonable to suppose that shifting patterns of military subjugation, forced relocation and uneasy alliance, as well as other more friendly types of inter-ethnic

association, must have characterized the Southeast Asian mainland in the distant past, as well as more recently.

The Laha of Ban Ung Kam in Vietnam, and their overlord-neighbours, the Black Tai, provide an excellent example of the sort of economic relationships which must once have been very common in the Southeast Asian past. According to Condominas (1980:281), prior to the present Vietnamese regime, the Laha were living in a stable economically-defined relationship of servitude to nearby Black Tai overlords, probably for many years (generations?). Laha and Black Tai villages are separate, but rice growing and control of property are regulated by strict inter-ethnic norms. For twenty days per month Laha workers (called *pua-pai*) were forced to provide labour for the Black Tai overlords (in five separate categories of dominance), including tending of buffaloes, repair of houses, collection of bamboo, cultivation of gardens, etc. Surplus produce was regularly presented by the Laha to their Black Tai masters, along with other ritual services. Being in the lower social position, the Laha had to become bilingual to communicate with the Black Tai, and increasingly they came to use Tai even for ritual purposes, folksongs, etc. Condominas reports that the Laha use Black Tai in all public discourse, with Laha restricted to the home. Unquestionably they would have mixed Tai into their speech, perhaps accounting for the etymologically mixed Austronesian-Tai wordlists collected by French colonial administrators and inspected by Benedict. Fortunately Condominas has carefully recorded particulars of the Laha contact situation, but we can only suppose that tiered labour arrangements of this sort have been the norm, rather than the exception, for much of mainland Southeast Asia into the distant past. Also, when labour demands became too exploitative, or when overlords failed to provide protection from outsiders, communal escape and migration often became options. Cases of this are frequently mentioned in local chronicle sources. Recalling Dixon's equilibrium stage, we can expect that over a long period of time, perhaps even several thousands of years, shifting arrangements of this sort would seriously affect the discernment, if not the viability, of simple family tree relationships.

10. Tracing the Tai linguistic past

Sections above have suggested the value of a critique of the Comparative Method as applied to the Tai languages. A preliminary problem encountered lies in the name 'Tai' itself, a term not without some problems. It is important to stress that names of this sort are not directly provided by the Comparative Method but must be assigned by professional consensus.

A distinction was drawn between two types of linguistic data of use in analysing the Tai languages. One type of data, regular sound correspondences, can be considered the essential evidence needed for the Comparative Method to operate.

This produces hypotheses of historical sound change and resulting branching structures: in effect, the Tai linguistic family tree. Hypotheses remain open to testing and revision as new data or new insights regarding conditioning factors, etc., may come to light. A strict application of the Comparative Method in theory may be essentially positivistic in ignoring historical records, social context, contact and other ecological factors. In practice, though, as Sections 4-6 have established, scholars applying the Comparative Method in the Tai case have shown that professional craft and judgment are required and that non-phonological evidence may be of value in motivating certain aspects of linguistic reconstruction.

Ecological factors become centrally important in analysing data of the second type, including syntactic typology and areal distribution of many lexicosemantic features. It is these linguistic features that are more like pottery types, textile designs or food preparation methods, in that areal distribution throughout Southeast Asia is the frequent pattern and it would be dangerous to claim, without extensive research, that data of these types could be exclusively Tai.

The Comparative Method as strictly applied has worked well for the level of Proto-Tai, and perhaps can be employed for Proto-Tai-Kadai, but for time depths deeper than these nodes the method becomes more problematic as sets of regularly corresponding items become more and more numerous, but with fewer and fewer relevant items in each set. We enter into the sort of conditions that Dixon (1997) indicates are the limits of strict comparative reconstruction.

At the greater time depths, an important question regarding the scope of application of the Comparative Method lies in how we are to recognize, in the Southeast Asian and southern Chinese context, genetic inheritance in the linguistic sense and distinguish it from massive contact with borrowing? —Or to what extent this traditional two-way distinction remains tenable at great time depths. Although the Proto-Tai grouping is now firmly established through technical principles of the Comparative Method, what is the status of earlier borrowing from non-Tai or pre-Tai languages into (Proto-)Tai?

In the Proto-Tai lexical inventory established by Li (1977), Luo (1997) and others, only a limited number of cognate items, such as *pi*: ‘year’, are clearly represented in Proto-Tai, occurring in virtually all daughter varieties, but appear to lack plausible cognates in other language families. Are we forced to admit that this rather small pool of such *pi*: items is what we will consider as candidates for an ultimate Tai (or Tai-Kadai) vocabulary? Perhaps the group’s identification or label is less important than the hypothesis it suggests, which might be called the ‘*pi*:-people scenario’.

In this speculation, a set of speakers—call them the **pi:* people—would have come into intense contact, in one way or another, over a protracted period, with Austronesian-speaking speakers (the **'blu'an* ‘moon’ people), Austroasiatic speakers (the **glo':ng* ‘canal’ people) and—massively—Sino-Tibetan speakers, the *plù'ak* ‘bark’ people) who contributed to the linguistic amalgam basic vocabulary like number terms and hundreds of further lexical items. Tiered labour situations with bilingual children, similar to what was reported for the Laha above, at times may have been the context for stages of the lexical integration. Multiple entries and contacts at different periods cannot be ruled out, to witness the paired *plù'ak/phiw* ‘skin’ items proposed by Manomaivibool (1976). Also, the various component carriers may have made multiple contacts among themselves before the *pi:* people entered the picture. Proto-Tai and—depending on definition—earlier stages of Tai-Kadai, would then need to be considered a synthesis of at least these elements, which could hardly be accounted for by the Comparative Method and which perhaps can never be completely untangled to everyone’s satisfaction.

Probably as the result of increasingly cohesive technological and socio-political practices, treated by Hartmann and other contributors to this volume, by about two millennia ago a Proto-Tai synthesis in the sense of Li (1977) became firm—or perhaps a somewhat earlier synthesis, to the extent that successful reconstructions of Proto-Kam-Tai or Proto Tai-Kadai can be established. At such a point the Comparative Method takes hold as an effective tool and with later centuries a clear family-tree structure begins to be discernible. But this scenario is but one speculation, posed here in the hope that it may stimulate reactions, further research and lead to clearer visions of the past.

REFERENCES

- Anuman Rajadhon, Phraya. 2499. *Niruktisat. [Etymology]*. Bangkok: Royal Institute.
- Ballard, W.L. 1985. "Aspects of the linguistic history of South China." *AP* 24 (2):163-185.
- Barlow, Jeffrey G. 1987. "The Zhuang minority peoples of the Sino-Vietnamese frontier in the Song period." *JSAS* 18(2):250-269.
- Benedict, Paul K. 1942. "Thai, Kadai and Indonesian: A new alignment in Southeastern Asia." *AA* 44:576-601.
- Benedict, Paul K. 1966. "Austro-Thai studies." *BSN* 3(1):2.
- Benedict, Paul K. 1975. *Austro-Thai: Language and Culture*. New Haven, Conn.: HRAF Press.
- Benedict, Paul K. 1990. *Japanese - Austro-Tai*. Ann Arbor: Karoma.
- Blust, Robert A. 1988. "Beyond the morpheme: Austronesian root theory and related matters." In Richard McGinn ed., *Studies in Austronesian Linguistics*. Ohio University Monographs in International Studies. Southeast Asia Series, No. 76:3-90.
- Brown, J. Marvin. 1965. *From Ancient Thai to Modern Dialects*. Bangkok: Social Science Association Press of Thailand.
- Chamberlain, James R. 1972. "The origin of the Southwestern Tai." *BARL* 7-8.
- Chamberlain, James R. 1973. "An introduction to Proto-Tai zoology." Ph.D. dissertation, University of Michigan.
- Clark, Marybeth. 1978. *Coverbs and Case in Vietnamese*. Canberra: Pacific Linguistics B-48.
- Condominas, Georges. 1980. *L'espace social à propos de l'Asie du Sud-Est*. Paris: Flammarion.
- Conrady, August. 1896. *Eine indochinesische Causativ-Denominativ-Bildung und ihr Zusammenhang mit den Tonaccenten*. Leipzig: Otto Harrassowitz.
- de Lacouperie, Terrien. 1886. "The languages of China before the Chinese." *TPS* (Reprinted 1969.)
- Diffloth, Gérard. 1976. "An appraisal of Benedict's views on Austroasiatic and Austro-Thai relations." *South-East Asian Studies*, Discussion Paper No. 82, Kyoto University.
- Diller, Anthony. 1987. "Reflections on Tai diglossic mixing." *Orbis* 32:1-2.
- Diller, Anthony. 2533/1990. "Phasa Trakul Thai: Buang Lang Naew Khwam-khit Baep Upalak [The Tai family of languages: behind the metaphors.]." In Cholthira Satyawadhna, A. Diller and B.J. Terwiel eds., *Khon Thai Doem Mai Dai Yu Thi Ni*. [The Proto Tais did not live here.]. Bangkok: Muang Boran.
- Diller, Anthony. 1993. "Grammaticalisation in Tai." Paper presented at *The Third International Conference on Thai Studies*, London.
- Dixon, R.M.W. 1997. *The Rise and Fall of Languages*. Cambridge University Press.

- Durie, Mark and Malcolm Ross, eds. 1996. *The Comparative Method Reviewed*. Oxford University Press.
- Dyen, Isidore. 1972. "The Austronesian languages and Proto-Austronesian." *CTL* 8:1-55. The Hague: Mouton.
- Edkins, Joseph. 1871. *China's place in Philology*. London. Np.
- Edmondson, Jerold A. and David B. Solnit., eds. 1997. *Comparative Kadai: The Tai Branch*. Summer Institute of Linguistics and the University of Texas at Arlington.
- Edmondson, Jerold A. and David B. Solnit. (eds.) 1988. *Comparative Kadai: Linguistic Studies Beyond Tai*. Summer Institute of Linguistics and the University of Texas at Arlington.
- Egerod, Søren. 1961. "Studies in Thai dialectology." *AO* 26:43-91.
- Fine Arts Department. 2526. *Charuk Sukhothai*. Bangkok: Fine Arts Department.
- Gedney, William J. 1964. *A Comparative Sketch of White, Black and Red Tai*. Bangkok: Social Sciences Review Special Publication 1.
- Gedney, William J. 1976. "On the Thai evidence for Austro-Thai." *CAAAL* 3:39-48. Also: *Ajia Afuriku go no Keisu Keisan* 6:65-86.
- Gedney, William J. [1979], 1989. "Evidence for another series of voiced initials in Proto-Tai." In Robert J. Bickner et al. eds., *Selected Papers on Comparative Tai Studies*. Michigan Papers on South and Southeast Asia. Center for South and Southeast Asian Studies, University of Michigan. No. 29:229-270.
- Gedney, William J. 1985. "Confronting the unknown: tonal splits and the genealogy of Tai-Kadai." In Graham Thurgood et al. eds., *LSTA*. Pacific Linguistics C-87:116-124.
- Grierson, 1903. *Linguistic Survey of India*. Calcutta: Office of the Superintendent of Government Printing.
- Haas, Mary R. 1958. "The tones of four Tai dialects." *Bulletin of the Institute of History and Philology, Academia Sinica (Taipei)* 29.
- Haas, Mary R. 1969a. *The Prehistory of Languages*. The Hague: Mouton.
- Haas, Mary R. 1969b. "Thai." *EB* 21. (14th Edition)
- Haas, Mary R. 1978. *Language, Culture and History* (Essays by Mary R. Haas). Stanford University Press.
- Hartmann, John F. 1980. "A model for the alignment of dialects in Southwestern Tai." *JSS* 68(1):72-86.
- Hartmann, John F. 1986. "Style, scope and rigor in comparative Tai research." In Robert J. Bickner et al. eds., *Papers from a Conference on Thai Studies in Honor of William J. Gedney*. Michigan Papers on South and Southeast Asia, Center for South and Southeast Asian Studies, University of Michigan, 25:171-184.
- Haudricourt, A.-G. 1948. "Les phonèmes et le vocabulaire du thai commun." *JA* 236:197-238.
- Haudricourt, A.-G. 1966. "The limits and connections of Austroasiatic in the northeast." In Norman H. Zide ed., *SCAL* 5:44-56.

- Henderson, Eugénie J.A. 1959. "The tones of the Tai dialect of Songkhla." *Bulletin of the the Institute of History and Philology, Academia Sinica* 30:233-235.
- Huffman, Franklin E. 1977. "An examination of lexical correspondences between Vietnamese and some other Austroasiatic languages." *Lingua* 42(2):171-198.
- Huffman, Franklin E. 1986. *Bibliography and Index of Mainland Southeast Asian Languages*. New Haven and London: Yale University Press.
- Jenner, Philip N. 1981. *A Chrestomathy of Pre-Angkorian Khmer*. Southeast Asia Paper No. 20, Part II. Center for Asian and Pacific Studies, University of Hawaii.
- Jones, Robert B. 1965. "On the reconstruction of proto-Thai." In G.B. Milner et al. eds., *IPLS* 1:194-229.
- Jones, Robert B. 1966. "Comparative Tai studies, a critique." In Ba Shin et al. eds., *Essays Offered to G.H. Luce*. Leiden: E.H. Brill. *Artibus Asiae Supp.* 23 (1):160-3.
- Jones, Robert B. 1970. "Classifier constructions in southeast Asia." *JAOS* 90(1):1-12.
- Kalaya Tingsabadh, M.R. 1988. "Loss of preeminence of citation forms in the study of tones in Thai dialects." Paper presented to *The International Symposium on Language and Linguistics*, Thammasat University, 9-11 August 1988.
- Karlgren, Bernard. *Grammatica Serica Recensa*. Stockholm: Museum of Far Eastern Antiquities.
- Kanchani La'ongsri. 2528. "Thin kamneot khong khon thai. [The original home of the Tai people.]" *Silapawattanatham chabap phiset*.
- Li, Fang Kuei. 1976. "Sino-Tai." *CAAAL* 3:39-48.
- Li, Fang Kuei. 1977. *A Handbook of Comparative Tai*. Oceanic Linguistics Special Publications No. 15. Honolulu: University Press of Hawaii.
- Li, Fang Kuei. 1989. "Proto-Tai *kh and *x." In Jeremy H.C.S. Davidson ed., *Southeast Asian Linguistics: Essays in Honour of Eugenie J.A. Henderson*. School of Oriental and African Studies, University of London.
- Luo, Yongxian. 1997. "The subgroup structure of the Tai languages: A historical-comparative study." *JCL Monograph Series Number* 12.
- Lutz, Edzard. 1992. "Polygenesis and entropy: An alternative evolution applied to semitic linguistics." Ph.D. dissertation, University of California, Berkeley.
- Maspero, Henri. 1911. "Contribution à l'étude système phonétique des langues thai." *BEFEO* 11:153-69.
- Matisoff, James. 1985. "Paul K. Benedict - an appreciation." In Graham Thurgood, et al eds., *LSTA* C-87:16-20.
- Matisoff, James. 1990. "On megalocomparison." *Language* 66(1):106-120.
- Matisoff, James. 1992. "Southeast Asian languages." In William Bright ed., *International Encyclopedia of Linguistics* 4:44-48, Oxford University Press.
- Nantana Danvivathana. 1987. *The Thai Writing System*. Hamburg: Helmut Buske Verlag. *Forum Phoneticum* 39.

- Norman, Jerry and Tsu-lin Mei. 1976. "The Austroasiatics in south China: Some lexical evidence." *MS* 32:274-301.
- Prapin Manomaivibool. 1975. "A study of Sino-Tai lexical correspondences." Ph.D. dissertation, University of Washington.
- Prapin Manomaivibool. 1966. "Layers of Chinese loanwords in Thai." In Thomas W. Gething, et al. eds., *TLHFKL* 179-184.
- Pranee Kullavanijaya. 2526. *Chuang - Chonchat Tai nai Satharanarat Prachachon Chin. [Zhuang - a Thai ethnic group in the Peoples Republic of China]*. Bangkok: Chulalongkorn University.
- Pranee Kullavanijaya, Amara Prasitrathsint and Suvanna Kriengkraipetch. 1996. *Collection of Papers on the Relationship Between the Zhuang and the Thai*. Chulalongkorn University.
- Preecha Juntanamalaga. 1989. "Sociolinguistic issues in Thai classifier usage." *LS* 10:313-330.
- Pulleyblank, E.G. 1983. "The Chinese and their neighbors in prehistoric and early historic times." In David N. Keightley ed., *The Origins of Chinese Civilization*. University of California Press. pp. 429-441.
- Ross, P.A. 1996 "Archaic features in Baan Ha Tay." *JLL* 14(2):53-61.
- Salmon, Merrilee H. 1996. "Causal explanations of linguistic behavior." *Theoretical Linguistics* 22.3:197-205.
- Sarawit, Mary S. 1973. "The proto-Tai vowel system." Ph.D. dissertation, University of Michigan.
- Schlegel, Gustav. 1902. "Siamese studies." *T'oung Pao* n.s.2, supp. 1-128.
- Schmidt, Wilhelm. 1906. *Die Mon-Khmer-Völker, ein Bindeglied zwischen Zentralasiens und Australasiens*. Braunschweig.
- Schmidt, Wilhelm. 1926. *Die Sprachfamilien und Sprachenkreise der Erde*. Heidelberg.
- Shorto, H.L. 1971. *A Dictionary of the Mon Inscriptions from the Sixth to the Sixteenth Centuries*. Oxford University Press.
- Shorto, H.L. 1979. "The linguistic protohistory of mainland South East Asia." In R.B. Smith and W. Watson eds., *Early South East Asia*. Oxford University Press.
- Smalley, William A. 1994. *Linguistic Diversity and National Unity*. University of Chicago Press.
- Southworth, Franklin C. 1964. "Family-tree diagrams." *Language* 40:557-565.
- Strecker, David. 1983. "Proto-Tai vowels revisited: A comparison and critique of the work of Sarawit and Li." *Languages of the Tibeto-Burman Area* 7(2):33-74.
- Taylor, K.W. 1983. *The Birth of Vietnam*. University of California Press.
- Terwiel, B.J. 1978. "The origin of the Tai peoples reconsidered." *OE* 25(2):239-258.
- Thurgood, Graham. 1985. "Benedict's work: past and present." In Graham Thurgood, et al. eds., *LSTA*. Pacific Linguistics C-87:1-15.
- Theraphan L. Thongkum. 1985. "Minority languages of Thailand." *SLP* 5:29-74.

- Theraphan L. Thongkum. 1994. "Implications of the retention of proto voiced plosives and fricatives in the Dai Tho Language of Yunnan for a theory of tonal development and Tai language classification." *Paper in the 27th International Conference on Sino-Tibetan Languages and Linguistics*.
- Tovar, Antonio. 1984. "Linguistic similarity and its significance: comparative procedures." *Proceedings of the 13th International Congress of Linguists* (Tokyo 1982).
- Vichin Panupong. 2499/1956. *Khwam-taek-tang Rawang Phasa Krungtrhep lae Phasa Songkhla*. [The difference between Bangkok Thai and Southern Thai] Chulalongkorn University MA thesis.
- Wichit Wathakan, Luang. 2475/1932. *Khwam-ru Buang-ton Nai Prawatisat*. [Basic historical knowledge.] Bangkok: Duang Pratheep.
- Wilaiwan Khanittanan. 1973. "The Influence of Siamese on Five Lao Dialects." Ph.D. dissertation, University of Michigan.
- Wilaiwan Khanittanan. 1988. "Thai written discourse: change toward a more autonomous style?" Paper presented to *The International Symposium on Language and Linguistics*, Thammasat University, 9-11 August 1988.
- Wulff, K. 1934. *Chinesische und Tai*. Copenhagen: Levin und Munksgaard. (Danske Videnskabernes Selskab, Hist.-filol. Meddelelser 20.3.)
- Wulff, K. 1942. *Über das Verhältnis des Malayo-Polynesischen zum Indochinesischen*. Copenhagen: Levin und Munksgaard. (Danske Videnskabernes Selskab, Hist.-filol. Meddelelser 27.2.)
- Wyatt, David K. 1984. *Thailand: A Short History*. Yale University Press.