MORE EVIDENCE FOR THE GENETIC RELATIONSHIP BETWEEN
AUSTRONESIAN AND KAM-TAI[1]

Baoya Chen and **Feng Wang**

Center for Chinese Linguistics, Peking University
Department of Chinese Language and Literature, Peking University

ABSTRACT

This paper provides more evidence for the genetic relationship between Austronesian and Kam-Tai. After an examination of the previous studies, it is preferred to compare modern languages directly at current stage, though some reconstructed Proto-languages are also used with caution. Dehong Dai, as a representative of Kam-Tai, and Indonesian, as a representative of Austronesian, have been compared, and systematic sound correspondences between them are established. According to Rank analysis, there are more Dai-Indonesian related words in High rank than those in Low rank, which indicates genetic relationship. Updated evidence and rank analysis show that both Kam-Tai languages and Austronesian languages are genetically related, respectively. Therefore, according to transitivity of genetic relatedness, Kam-Tai and Austronesian should be genetically related. Moreover, sound correspondences between Proto-Tai and Indonesian have been worked out. And Rank analysis confirms the genetic relationship. Finally, via the similar procedure, it is found that the genetic relationship between Austronesian with either Chinese or Tibeto-Burman are not confirmed because the related words between them in High rank are less than those in Low rank.

SUBJECT KEYWORDS
Genetic relationship, Kam-Tai, Austronesian,Rank analysis

1. INTRODUCTION

It is generally accepted that there exists an Austronesian language phylum (or Malayo-Polynesian language phylum) in Southeastern Asia and the Pacific area, which

consists of more than 1,000 languages. More than 250 million people are speaking these languages. Another language phylum, Austroasiatic, including more than 150 languages, is also distributed in Southeastern Asia. These languages are used by more than 40 million people. William Schmidt (1906) put the two language phyla together and called the new super language group Austric, because the two phyla share some common affixes. Reid (1994) provided more shared verbal configurations to prove the existence of Austric. Reid (2005) examined studies on Austric family in recent decades and concluded that Austronesian and Austroasiatic are genetically related based on cumulative lexical and morphological evidence.

Kam-Tai languages spread broadly in Southwestern China and Southeastern Asia. Kam-Tai languages have little resemblance with Austronesian languages, while they look like Chinese, Tibeto-Burman or Hmong-Mien. Li (1937) united the four language groups into the Sino-Tibetan family, mainly because they are tonal and monosyllabic.

Benedict (1942, 1976, 1991) argued against the genetic relationship between Austronesian and Austroasiatic. He also disagreed with Kam-Tai being genetically related to Chinese. To him, the resemblances in structures like affix, morphology, tone and monosyllable may not be the evidence of genealogy since they are often the results of language contact. Benedict proposed another language family, called Austro-Tai, which consists of Austronesian, Kam-Tai, Kadai and Hmong-Mien. Benedict listed some basic words with sound correspondences between Austronesian, Kam-Tai and Hmong-Mien to support his claim. Meng (1990) and Ni (1988) also added some corresponding words between Austronesian and Kam-Tai, but rigorous correspondences are still needed. We will list more complete correspondences and use rank analysis to further prove genetic relationship between Austronesian and Kam-Tai.

Zhengzhang (1995) proposed a super language family, called Sino-Austric, which includes Austronesian, Austroasiatic, Chinese, Tibeto-Burman, Hmong-Mien and Kam-Tai, based on a few basic words shared by these languages. Pan (1995) supported this idea and provided more evidence from syllabic typology and cognates. Starosta (2005) had a similar proposal and drew a detailed tree for these languages.

Proto-East Asian

Tibeto-Burman-Yangtzean   Austronesian

Tibeto-Burman   Proto-Yangtzean   Formosan   Extra-Formosan

Sino-Bodic   Himalayo-Burman   Austroasiatic

Sinitic   Kamarupan   Qiangic   Munda   Mon-Khmer   Malayo-Polynesian

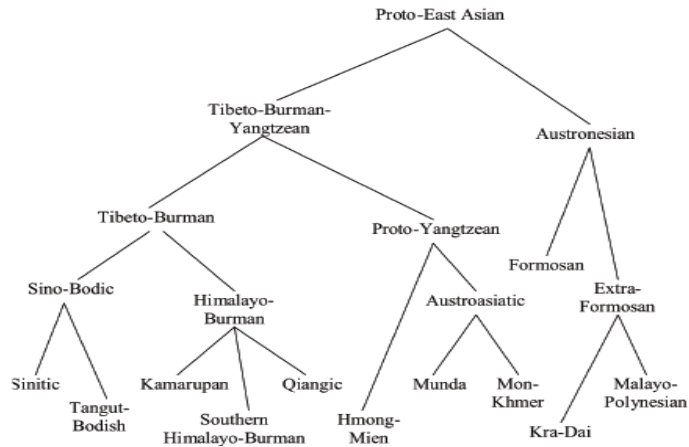Tangut-Bodish   Southern Himalayo-Burman   Hmong-Mien   Kra-Dai

Figure 1 Proto-East Asian

In the above arguments, the genetic relationship between Kam-Tai and Austronesian is a crucial point. The evidence for or against their genetic relationship needs more discussions.

In the studies of Indo-European languages, common morphologies and sound correspondences are two basic criteria in defining the genetic relationship between languages. The arguments for genetic relationship by Schmidt, Reid, Li, Benedict and other scholars can be classified into either one or the other. However, when the contact between languages is heavy, the criterion of common morphologies may meet problems because they may be due to borrowing. Some still tenable shared inflections are often corresponding inflections between languages. Therefore, both criteria can be attributed to correspondence in the end (Chen 1999b:217-218). Now more and more scholars put more weights on sound correspondences. They believe that if the sound correspondence between basic words of different languages could be established, their genetic relatedness is confirmed. From the perspective of language contact, this hypothesis depends on two preconditions: firstly, language contact could not have caused systematic sound correspondences; secondly, basic words are immune to language contact.

Not only is the internal sound change regular, it is found that sound matches in language contact are also regular (Chen 1994), and language contact may result in the systematic correspondences between the borrowing words and the original ones. Moreover, there is no limit in borrowing. The basic words, even kernel words, may be

borrowed. The regularity between the borrowing words and the original ones is as rigorous as that between cognates. Therefore, it is difficult to distinguish whether the sound correspondences in the oldest layer are due to language split or language contact (Chen 1994). In recent years, studies on language contact have shown that sound correspondence is a necessary condition for genetic relatedness, but not by any means a sufficient condition on its own.

Neogrammarians claimed that 'sound laws admit of no exception', which was taken for granted as the working hypothesis for years in historical linguistics. According to our continuous records on homophones in the Dai language in recent decades, the counterevidence is not found. If there is any exception of regular sound change, there must be a reason for it, such as language contact or on-going lexical diffusion (Wang 1969), etc. Up to now, we did not find any exception without reason. It may be safely concluded that the regularity of sound change is supported by empirical data. According to this supposition, sound correspondences between languages will be resulted in after their separation from the ancestor language. In this paper, the sound correspondences between Kam-Tai and Austronesian will be extablished at first, as rigorously as possible, then the relationship between Kam-Tai and Austronesian will be reexamined by means of Rank analysis (Chen 1994).

2. METHODOLOGIES

2.1. The Priority of Modern Languages

In order to explore the relationship between Austronesian and Kam-Tai, the comparative works can start from the proto-languages, Proto-Austronesian and Proto-Kam-Tai, or from modern languages. Different starting points will be associated with diverse advantages and disadvantages.

In the comparison of Proto-languages, the first problem will be the variant distribution of morphemes among branches. Taking Austronesian as an example, if a morpheme has corresponding reflexes in all modern Austronesian languages, it can be reconstructed in Proto-Austronesian almost undoubtedly. The problem lies in that many morphemes are not distributed in all modern languages. What is to determine which branches are more important than the others? Obviously, scholars do not all make the same choices. Even though faced with the same materials, the reconstructed forms for Proto-Austronesian have come out looking far from identical with each other, and even the quantities of proto-forms are not equal. In other words, if different subgrouping

proposals are chosen, the languages will be weighed unequally during the reconstruction process. Various interpretations on sound change can also cause inconsistent reconstructions although based on the same data. Hence, reconstructions by scholars often vary in many aspects. In fact, with the collection of new Austronesian materials and improved interpretations of sound changes, fresh systems of Proto-Austronesian were proposed. In the earlier time, Dempwolff (1934-1938) and Dyen (1963, 1965, 1971) were the representatives. Later there came Tsuchida (1976) and Blust (1970, 1977, 1980, 1983-84, 1988, 1989, 1999). Blust (1980) subgrouped the Austronesian languages as follows:
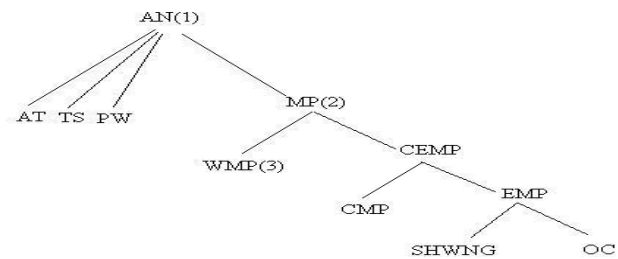
 Figure 2 Blust's classification on Austronesian languages (Adapted from Blust 1980:11)

(AN=Austronesian; AT=Atayalic (Formosa); TS=Tsouic (Formosa); PW = Paiwan (Formosa); MP=Malayo-Polynesian (all AN languages outside Formosa); WMP =Western Malayo-Polynesian. CEMP=Central-Eastern Malayo-Polynesian. CMP=Central Malayo-Polynesian. EMP=the languages of the SHWNG and OC groups. SHWNG=South Halmahera-West New Guinea. OC=Oceanic.)

It should be noted that it is still controversial on how to subgroup the Austronesian languages. Tsuchida (1976) had a quite different classification from Blust (1980). Sagart (2005) put forward a classification, and modified some proto-forms reconstructed by Blust accordingly. Here is Sagart's tree diagram for Austronesian languages.

```
                        PAN
    ┌────────┬──────────┬──────────┬──────────┬──────────┐
 Paiwan   Puyuma   East Coast   Rukai    Western   Northern
                     Linkage    Tsouic
        ┌──────┬──────┬──────┬──────┬──────────┬──────────┐
     Siraya  Amis  Bunun  PMP    AAK   Kavalan        Basai
                                                      Trobiawan
                            etc.
                            many
                          languages
```
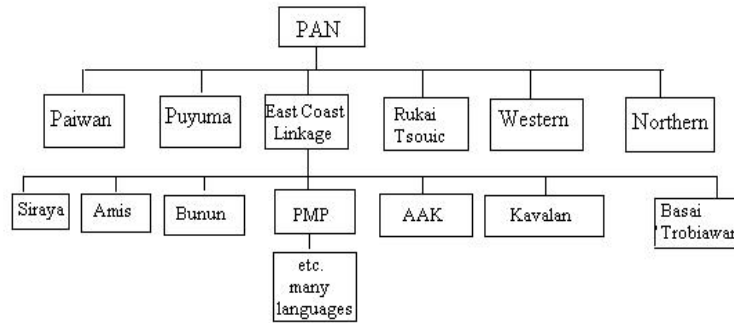
Figure 3 Sagart's classification on Austronesian languages (Modified from Sagart 2005)

        In recent decades, the genetic structure of Austronesian languages was drawn
quite differently. The most complicated is to determine the position of Austronesian
languages in Taiwan (Starosta 1995, Li 1995a, Ho 1995). According to Blust (1999),
more than 17 proposals have been made. For instance, Dyen (1963) used the
lexicostatistic method and concluded that Atayalic and Tsouic should be two independent
branches under the Taiwan Austronesian. Dyen (1965b) proposed that the first split of
Proto-Austronesian is Atayalic (Atayal and Seediq) vs. Eastern Formosan (Amis, Paiwan,
Bunun, and Thao). Ferrell (1969) doubted the existence of Proto-Formosan, which should
be divided into three branches according to him: Atayalic (Atayal and Seediq), Tsouic
(Tsou, Kanakanabu and Sarroa) and Paiwanic (the rest). Tsuchida (1976) accepted
Proto-Formosan, but he put Tsouic and Paiwanic together to contrast with Atayalic.
Harvey (1982) classified Austronesian languages into four groups, Atayalic, Tsouic, the
remaining Formosan languages, and Amis-Malayo-Polynesian. Li (1985) divided
Formosan into three, Northern Formosan, Tsouic and Paiwanic. After careful examination
on the proposals relevant to Taiwan Austronesian, Blust (1999) concluded that current
phonological, lexical or morphosyntactic evidence is not sufficient to prove the existence
of Proto-Formosan.

        Since the subgrouping of Austronesian languages is so diverse, the reliability of
reconstruction of Proto-Austronesian is weakened a lot. For this reason, Ho (1999:77)
adopted a very rigorous criterion to define cognates of Austronesian: "Only when their
corresponding reflexes can be found in OC, Hesperonesian and Formosan, they are
qualified to be Austronesian cognates. Moreover, among Formosan, the reflexes should

be seen in two of the three groups, AT, TS and PW, at least." This requirement on the distribution of corresponding reflexes maybe strengthens the basis of reconstruction. However, it should be realized that the degree of such strengthening depends on whether the distribution requirement matches the genetic structure of Austronesian.

Due to the problems in Proto-Austronesian, it is reasonable to compare modern languages of Austronesian and Kam-Tai directly. Such comparison may avoid the subjectivity in reconstruction. In this paper, the modern Dai in Dehong will be compared with Indonesian. According to the transitivity of genetic relatedness, we may extend the result of comparison to identify the relationship between Kam-Tai and Austronesian. If Indonesian is genetically related to Dai, the languages genetically related to Indonesian or Dai may share the same ancestor languages with them. The Dehong Dai is selected because our studies on it are relatively thorough. Indonesian is the Austronesian language used by the most population. And it is well studied.

Some forms are perfectly corresponding between Dai and Indonesian, but not so between Proto-Tai and Indonesian, for example:

	Proto-Tai	Indonesian	Dai	Longzhou Zhuang
door	*tu1	pin**tu**	(la3) tu6	
fart	*tlot7	qen**tut**	tot9	
eye	*tra1	ma**ta**	ta6	ha:1

This case may be due to too few examples being sampled. It is also possibly due to incorrect reconstruction of Proto-Tai. In order to explain the origin of "h-" of Longzhou Zhuang, Li (1977) proposed the distinction of *t- and *tr- in Proto-Tai. The *tr- of Proto-Tai changed into h- of Longzhou zhuang. However, another alternative explanation is also available. That is, the *t- changed into h- before -a. This proposal can also put "eye" into the regular sound correspondence again.

Of course, there are some limitations when using modern language to do comparative works. The major disadvantage is that the earlier distinction may have been lost. For example (A corresponding to B will be written as A:B):

Onset	Lexical items	Dai	Proto-Tai	Indonesian
hm:b	pig	mu1	*hmu1	(ba)bi
hm:b	bear	mi1	*hmi1	bi(ruaŋ)

hm:b	fruit	ma:k9	*hmak9	buah
hm:b	new	matɯ5	*hmotɯ5	baru
m:b	sweet potato	man2	*mən2	(u)bi
m:b	reamer	mit8	*mit10	(sa)bit [2]
m:m	ant	mot8	*mot8	(se)mut
m:m	come, arrive	ma2	*ma2	ma(ri)
m:m	you	matɯ2	*mɯŋ2	(ka)mu

If compare Proto-Tai with Indonesian, three sets of onset correspondences could be found:

	Onset of Proto-Tai	Onset of Indonesian
1	hm	b
2	m	b
3	m	m

If compare Dehong Dai, only two sets of correspondences could be found:

	Onset of Dai	Onset of Indonesian
1	m	b
2	m	m

The reason is simply that *hm- and *-m of Proto-Tai merge into m- in Dai. Considering such factor, the reconstruction of Proto-language may be referred to when necessary. Since Proto-Tai reconstructed in Li (1977) has been generally accepted, it will be also used to compare with Indonesian.

2.2 Rigorous Match and Complete Correspondence

Sound correspondences are the basis of comparative studies. Semantically, the meanings of corresponding items should match each other rigorously. For instance, in the comparison of English and Chinese, 'hand' in English should be used to match 'shou 手 [hand]' in Chinese, not 'jian 肩[shoulder]'. If the meanings are not exactly the same, the discrepancy should be explained. The loose requirement in semantic equivalence may result in lots of chance correspondence. Therefore, it is necessary to set rigorous semantic requirements. The semantic correspondences must be proved by literature, archaeological

evidence or universal patterns of semantic changes (Chen 1994, 1999a, Wang 2006:5-7).

In another side, if phonetic correspondences are well controlled, the demerits caused by loose semantic match may be cancelled partially. The most rigorous sound correspondences should be complete correspondences, which require that each phonetic element of a morpheme should fall into a set of correspondences between languages (Chen 1999a, 2004). This requirement can help to cancel fraud correspondences due to loose semantic match. The chance resemblance may be also excluded. Moreover, later layers due to borrowing can be excluded partially. Here are examples of m-<>h- between Dai and English (Chen 1994:211):

Dai	ma4	mi2	mɯ2	ma:ŋ2	man2	mai3
English	horse	have	hand	half	he	hot

Such correspondences are chance resemblances, because the match of 'a' in Dai and 'orse' in English represented in morpheme 'horse' cannot be supported by more examples. That is to say, 'ma4' vs. 'horse' is not corresponding completely. The correspondence of m-<>h- between Dai and English is an illusion due to chance.

An extreme example of complete sound correspondence due to chance is given below (Quoted from Ting 2000):

Chinese	yan51 燕	yan51 嚥
English	swallow	swallow

The two morphemes are homophonous in Chinese and English, respectively. They are complete correspondences according to our above definition. However, such correspondences cannot get further support because it is difficult for us to find more examples for the match between Chinese 'y-' and English 'sw-', or any other match between the two. Therefore, the systematic complete correspondences may be the necessary and sufficient condition to exclude chance resemblance. In another side, if we loosen semantic requirement and do not insist on complete correspondences, lots of chance resemblances may swarm in and interfere with the analysis on genetic relatedness.

2.3 Complete Correspondence

Sound correspondences between two languages should be based on the whole

lexicon, not only on a part of it. We made rank analysis of Sino-Tai kernel morphemes (Chen 1994), which does not mean that the sound correspondences only depended on the matches of the 200 kernel morphemes. In our following analysis, the first step is to find all sound correspondences between Chinese and Tai, then those which fall into kernel morphemes will be observed.

If the separation of two languages is too long, the retained cognates may be not sufficient to establish the systematic complete sound correspondences even after searching in the whole lexicon. In the Sino-Tai comparison, we tried to divide a syllable into three parts, initial, final and tone. If the sound correspondence of any part cannot be established, it would be treated as incomplete sound correspondence. Such analysis based on constituents of syllable may be suitable if the number of corresponding morphemes is sufficient. However, if only a few related morphemes left due to a long-time separation, the above analysis may meet problems. In that situation, many initials or finals may cover only one example. We cannot talk about correspondence at all. According to previous comparative studies of Kam-Tai and Austronesian, their separation should be quite early. The cognates between them would not be many. To deal with such case, it may be helpful to divide syllable into even smaller element. Then, each small element may cover several examples. Considering the lexical item 'moon' in Dai [lən6] and Indonesian [bulan], the second syllable of Indonesian is comparable to Dai as below:

	Onset	Nucleus	Ending
Dai	l	ə	n
Indonesian	l	a	n

The complete correspondence requires that onset, nucleus and ending of a syllable are corresponding between two languages, respectively. That is, there are at least two examples to support each kind of match. For example, the onset correspondences between Dai and Indonesian: (The zero position will be marked as "0".)

Onset	Lexical items	Dai	Indonesian	Nucleus	Ending
l:l	moon	lən6	bulan	ə:a	n:n
l:l	tongue	lin4	lidah	i:i	n:0
l:l	deep	lək8	djeluk	ə:u)	k:k
l:l	forget	lɯɯm2	lupe		

The first column indicates that four examples support the onset sound correspondence (OSC) "l:l". It should be noted that the strengths of the four are not equivalent. The relatedness of "moon" morpheme between Dai and Indonesian is supported not only from OSC, but also from nucleus sound correspondence (NSC) "ə:a" and ending sound correspondence (ESC) "n:n"(see evidence below). The morpheme "tongue" is similar. However, the relatedness of "deep" between Dai and Indonesian could not be supported by NSC, because there is only one example of "ə:u" among all the candidates of cognates. This kind of match will be marked by an additional ")". The relatedness of morpheme 'forget' can be only supported by OSC.

Here are examples to support nucleus of "moon" between Dai and Indonesian.

NSC	Lexical items	Dai	Indonesian	ESC	OSC
ə:a	moon	lən6	bulan	n:n	l:l
ə:a	top, above	(pa3)lə1	atas		l:t

Examples for ending of "moon" between Dai and Indonesian are listed below.

ESC	Lexical items	Dai	Indonesian	NSC	OSC
n:n	moon	lən6	bulan	ə:a	l:l
n:n	eat	kin6	makan	i:a)	k:k

The forms of "moon" of the two languages are corresponding in onset, nucleus and ending. Such case would satisfy the criteria of a complete sound correspondence. In above tables, when one of OSC, NSC and ESC is discussed the other two are also listed as reference. We will continue this format in the following paragraphs since such format shows the status of sound correspondence straightforwardly.

3. DAI-INDONESIAN SOUND CORRESPONDENCES
3.1 Dai-Indonesian Complete Sound Correspondences in Broad Style
3.1.1 Onset Sound Correspondence between Dai and Indonesian
        (Note, in the column of "R=Rank", "1" indicates the first 100 kernel words (Swadesh 100 basic words), while "2" indicates the second 100 kernel words (Swadesh 1952, Chen 1994). "In-root" = "Indonesian root".)

OSC	Lexical items	Dai	Indonesian	In-root	NSC	ESC	R
k:g	pair	ku6	gu	gu	u:u	0:0	
k:g	bite,gnaw	kat7	gitgit	-git	a:i	t:t	1
k:g	salt	kə6	garam	ga-	ə:a	0:0	2
k:k	eat	kin6	makan	-kan	i:a)	n:n	1
k:k	I	kau6	aku	-ku	a:u	u:0)	1
l:d	nose	(hu2) laŋ6	hiduŋ	-duŋ	a:u	ŋ:ŋ)	1
l:d	gallbladder	li6	empedu	-du	i:u)	0:0	
l:l	moon	lən6	bulan	-lan	ə:a	n:n	1
l:l	tongue	lin4	lidah	li-	i:i	n:0	1
l:l	forget	lɯɯm2	lupa	lu-			
l:l	deep(water)	lək8	djeluk	-luk	ə:u)	k:k	
l:n	farmfield	la2	tanah	-nah	a:a	0:h	
l:n	bird	lok8	manuk	-nuk	o:u	k:k	1
l:n	child, son	luk8 (tsa:i2)	anak	-nak	u:a	k:k	2
l:n	young, tender	lum5	anom	-nom	u:o)	m:m	
l:n	this	lai4	ini	-ni	a:i	i:0	1
l:t	top, above	(pa3)lə1	atas	-tas	ə:a		
l:t	mice	lu1	tikus	ti-	u:i		
l:t	thick	la1	tebal	te-	a:e)	0:0	2
l:t	black	lam6	hitam	-tam	a:a	m:m	1
m:b	pig	mu1	babi	-bi	u:i	0:0	
m:b	sweet photato	man2	ubi	-bi	a:i	n:0	
m:b	fruit	ma:k9	buah	buah	a:ua)	k:h	2
m:b	shoulder	(ho1ma5); ma5	bahu	ba-	a:a	0:0	
m:b	new	maɯ5	baru	ba-			1
m:b	reamer	mit8	sabit	-bit	i:i	t:t	
m:m	ant	mot8	semut	-mut	o:u	t:t	
m:m	come, arrive	ma2	mari	ma-	a:a	0:0	1
m:m	you	maɯ2	kamu	-mu	a:u	ɯ:0)	1
p:p	fish	pa6	patin	pa-	a:a	0:0	1
p:p	circumrotate	pan5; (taɔn5)	putar	pu-			

s:s	wash	suk8	basuh	-suh	u:u	k:h	2
s:s	you(pl.)	su1	saudara	sau-	u:a	0:u)	
t:t	eye	ta6	mata	-ta	a:a	0:0	1
t:t	door	la3tu6	pintu	-tu	u:u	0:0	
t:t	die	ta:i6	mati	-ti	aa:i)	i:0	1
t:t	fart	tot9	qentut	-tut	o:u	t:t	
t:t	fall	tok9; tok9 (ha:i1) [3]	jatuh	-tuh	o:u	k:h	2
x:k	rightside	xa1	kanan	ka-	a:a	0:0	2
x:k	laugh, smile	xo1	dekah	-kah	o:a)	0:h	2

In the above table, some OSCs are associated with NSC and ESC, which would be complete sound correspondences. Whereas, some rows, the ")" in NSC column or in ESC column suggests incomplete sound correspondences.

3.1.2 Nucleus Sound Correspondence between Dai and Indonesian

NSC	Lexical items	Dai	Indonesian	In-root	OSC	ESC	R
a:a	farmfield	la2	tanah	-nah	l:n	0:h	
a:a	rightside	xa1	kanan	ka-	x:k	0:0	2
a:a	fish	pa6	patin	pa-	p:p	0:0	1
a:a	sesame	ŋa2(lo5)	ləŋa	-ŋa	ŋ:ŋ	0:0	
a:a	eye	ta6	mata	-ta	t:t	0:0	1
a:a	shoulder	(ho1ma5); ma5	bahu	ba-	m:b	0:0	
a:a	leg	xa1	paha	-ha	x:h)	0:0	2
a:a	tiny chaff	ham2	sekam	-kam	h:k)	m:m	
a:a	ash	tau6	pirau	-rau		u:u	1
a:a	come, arrive	ma2	mari	ma-	m:m	0:0	1
a:a	black	lam6	hitam	-tam	l:t	m:m	1
a:a	night	(ka:ŋ6)xam6	semalam	-malam	x:l)	m:m	1
a:a	branches	xa6; ŋa6	tjagak	-gak			
a:i	fire	fai2	api	-pi		i:0	1
a:i	sweet photato	man2	ubi	-bi	m:b	n:0	

40

a:i	seed	fan2	bibit	bi-	f:b)	n:0	1
a:i	dream	fan1	mimpi	-pi	f:p)	n:0	
a:i	this	lai4	ini	-ni	l:n	i:0	1
a:i	bite, gnaw	kat7	gitgit	-git	k:g	t:t	1
a:u	flea	mat7	kutu	-tu		t:t	
a:u	nose	(hu2) laŋ6	hiduŋ	-duŋ	l:d	ŋ:ŋ)	1
a:u	blow	pau5	hembus	-bus	p:b)		2
a:u	i	kau6	aku	-ku	k:k	u:0)	1
a:u	you	matɯ2	kamu	-mu	m:m	ɯ:0)	1
e:a	green	xeu1	hijau	-jau		u:u	1
e:a	frog	xet9	katak	-tak	x:t	t:t	
ə:a	moon	lən6	bulan	-lan	l:l	n:n	1
ə:a	top, above	(pa3)lə1	atas	-tas	l:t		
ə:a	salt	kə6	garam	ga-	k:g	0:0	2
i:i	tongue	lin4	lidah	li-	l:l	n:0	1
i:i	reamer	mit8	sabit	-bit	m:b	t:t	
o:u	bird	lok8	manuk	-nuk	l:n	k:k	1
o:u	ant	mot8	semut	-mut	m:m	t:t	
o:u	fart	tot9	qentut	-tut	t:t	t:t	
o:u	fall	tok9; tok9 (ha:i1)	jatuh	-tuh	t:t	k:h	2
o:u	overlay	lop9	kup	kup	l:k)	p:p)	
u:a	child, son	luk8 (tsa:i2)	anak	-nak	l:n	k:k	2
u:a	you(pl.)	su1	saudara	sau-	s:s	0:u)	
u:i	pig	mu1	babi	-bi	m:b	0:0	
u:i	mice	lu1	tikus	ti-	l:t		
u:u	door	(la3)tu6	pintu	-tu	t:t	0:0	
u:u	wash	suk8	basuh	-suh	s:s	k:h	2
u:u	understand	hu4	tahu	-hu	h:h)	0:0	1
u:u	pair	ku6	gu	gu	k:g	0:0	

3.1.3 Ending Sound Correspondence between Dai and Indonesian

41

ESC	Lexical items	Dai	Indonesian	In-root	OSC	NSC	R
0:0	rightside	xa1	kanan	ka-	x:k	a:a	2
0:0	pig	mu1	babi	-bi	m:b	u:i	
0:0	fish	pa6	patin	pa-	p:p	a:a	1
0:0	sesame	ŋa2lo5	ləŋa	-ŋa	ŋ:ŋ)	a:a	
0:0	eye	ta6	mata	-ta	t:t	a:a	1
0:0	shoulder	(ho1)ma5; ma5	bahu	ba-	m:b	a:a	
0:0	leg	xa1	paha	-ha	x:h)	a:a	2
0:0	gallbladder	li6	empedu	-du	l:d	i:u)	
0:0	door	(la3) tu6	pintu	-tu	t:t	u:u	
0:0	salt	kə6	garam	ga-	k:g	ə:a	2
0:0	understand	hu4	tahu	-hu	h:h)	u:u	1
0:0	come	ma2	mari	ma-	m:m	a:a	1
0:0	thick	la1	tebal	te-	l:t	a:e)	2
0:0	pair	ku6	gu	gu	k:g	u:u	
0:h	farmfield	la2	tanah	-nah	l:n	a:a	
0:h	laugh, smile	xo1	dekah	-kah	x:k	o:a)	2
i:0	fire	fai2	api	-pi		a:i	1
i:0	die	ta:i6	mati	-ti	t:t	aa:i)	1
i:0	this	lai4	ini	-ni	l:n	a:i	1
k:h	fruit	ma:k9	buah	buah	m:b	a:ua)	2
k:h	wash	suk8	basuh	-suh	s:s	u:u	2
k:h	fall	tok9; (tok9 ha:i1)	jatuh	-tuh	t:t	o:u	2
k:k	bird	lok8	manuk	-nuk	l:n	o:u	1
k:k	child, son	luk8 (tsa:i2)	anak	-nak	l:n	u:a	2
k:k	deep(water)	lək8	djeluk	-luk	l:l	ə:u)	
m:m	tiny chaff	ham2	sekam	-kam	h:k)	a:a	
m:m	black	lam6	hitam	-tam	l:t	a:a	1
m:m	young, tender	lum5	anom	-nom	l:n	u:o)	
m:m	night	(ka: ŋ6)	semalam	-malam	x:l)	a:a	1

		xam6					
n:0	sweet photato	man2	ubi	-bi	m:b	a:i	
n:0	seed	fan2	bibit	bi-	f:b)	a:i	1
n:0	tongue	lin4	lidah	li-	l:l	i:i	1
n:0	dream	fan1	mimpi	-pi	f:p)	a:i	
n:n	moon	lən6	bulan	-lan	l:l	ə:a	1
n:n	eat	kin6	makan	-kan	k:k	i:a)	1
t:t	ant	mot8	semut	-mut	m:m	o:u	
t:t	flea	mat7	kutu	-tu		a:u	
t:t	fart	tot9	qentut	-tut	t:t	o:u	
t:t	reamer	mit8	sabit	-bit	m:b	i:i	
t:t	frog	xet9	katak	kat-	x:k)	e:a	
t:t	bite, gnaw	kat7	gitgit	-git	k:g	a:i	1
u:u	ash	tau6	pirau	-rau		a:a	1
u:u	green	xeu1	hijau	-jau		e:a	1

3.2 Problems

Four important questions may be addressed on sound correspondences listed above. Firstly, many Indonesian words are polysyllabic, but only one syllable of a polysyllabic Indonesian word is selected to compare with the monosyllabic word of Dai. The question is how to explain the rest of the polysyllabic Indonesian words. Where did they go? Secondly, what element of Indonesian is corresponding to tone of Dai? In this sense, even OSC, NSC and ESC all support the relatedness, it cannot be counted as a complete correspondence. Thirdly, there are some one-to-many matches. The cause is not yet known. Fourthly, some of the morphemes which are used to support the complete correspondence do not belong to complete correspondences.

These problems suggest that layers of sound correspondences in different time depths are perhaps not stratified, and some details of sound change in early time depths need further studies. Therefore, tables in chapter 3.1 just list the sound correspondences in broad sense. However, such regular sound correspondences are unlikely to be chance resemblances. The question turns out to be whether language contact or language split result in them?

4. RANK ANALYSIS ON DAI-INDONESIAN RELATED MORPHEMES

We have tried to divide Swadesh 200 basic words into two ranks according to their stability in language evolution. High rank consists of Swadesh 100 basic words (Swadesh 1955), while Low rank consists of the remaining 100 basic words after some modifications (Chen 1994). Ting (2000) asked the reason of the division. Our answer had be given before Ting's questioning: the division of High / Low rank is based on two observations. Firstly, in those languages with genetic relationship, the ratio of cognates in High rank is higher than that in low rank. Germanic languages and Northern Chinese dialects support such distribution of cognates, so do Dai dialects. On the contrary, in those languages in contact, for instance, Dai dialects and Southwestern Mandarin, borrowings in high rank are less than those in low rank. Secondly, during the contact between Dai and Southwestern Mandarin, the more words a morpheme in Dai can construct, and the more widely it distributes among Tai languages, the more difficult it is to be replaced by Southwestern Mandarin words (Chen 1994:94-124). The morphemes in the High rank are generally more active than those in the Low rank when they are used to construct words. Considering the related morphemes, those in the High rank are more widely distributed than those in the Low rank. For instance, counting the related morphemes reflected in all the three groups of Tai, there are 61 percentages in High rank, while 53 percentage in Low rank.

We would like to emphasize once again that the spirit of rank analysis is not to represent Kam-Tai and Chinese via the 200 kernel morphemes. The point is that the high rank and low rank as two groups can be samples to reflect the contrast between distribution of cognates and borrowing words.

In this spirit, the Sino-Tai related words in the early time were examined. It is found that the related words in High rank are far less than those in Low rank. Therefore, these related words are not sufficient to prove the genetic relatedness between Chinese and Kam-Tai.

The rank analysis has been extended to basic morpheme. The Sino-Tai related morphemes are discussed in terms of relative rank (Chen 1999a; 2004). Relative rank analysis assumes that the stability of a morpheme may be different in different language groups. The distribution of a morpheme among a language group would be an index of its stability. According to this index, the relative ranks of morphemes of a certain language group could be defined. After this kind of analysis, the Sino-Tai related morphemes in the relative high rank are less than those in the relative low rank. It may be concluded that the related morphemes in basic words could not prove the genetic relationship of Sino-Tai,

either.

The absolute rank analysis within 200 kernel morphemes and relative rank analysis of basic words work only in their own scopes. The absolute rank analysis ignores the relativity of languages and assumes that the stability of any particular morpheme is the same in world languages. If "hand" belongs to high rank in Indo-European languages, it would be so in the other languages. In fact, high rank groups and low rank groups in different languages vary to some degree. The variation is ignored in the absolute rank analysis. Though relative rank analysis can deal with this discrepancy, yet it may encounter another problem. When two languages get into contact after their separation from a common ancestor, the shared cognates may be quite different from borrowing words due to later contact in terms of sound correspondence if the separation is long enough. However, if the interval is not long enough, it would be hard to separate the early related words from the later related words. For instance, Sino-Tai related words from the Old Chinese period and those from the Middle Chinese period are difficult to be differentiated. In this situation, the distribution of the early related words among languages could not be specified. In relative rank analysis, the larger the sample size is, the more later related words are mixed into the early related words.

Relative rank analysis requires that the genetic tree of languages is drawn at first. Depending on the distribution of corresponding morphemes in the tree, the relative high/low rank will then be defined. This paper mainly starts from the living modern languages, and the genetic tree and reconstructed materials will only serve as references. The relative rank analysis will not be implemented, but some results based on relative rank analysis in previous studies may be quoted.

4.1 Rank Analysis on Dai-Indonesian Complete Correspondences

A collection of complete correspondences with OSC, NSC and ESC is generated as below after having excluded the rows without OSC, NSC or ESC. (The order is OSC+NSC+ESC.)

Lexical items	Dai	Indonesian	In-root	OSC	NSC	ESC	R
pair	ku6	gu	gu	k:g	u:u	0:0	
bite, gnaw	kat7	gitgit	-git	k:g	a:i	t:t	1
salt	kə6	garam	ga-	k:g	ə:a	0:0	2

moon	lən6	bulan	-lan	l:l	ə:a	n:n	1
tongue	lin4	lidah	li-	l:l	i:i	n:0	1
farmfield	la2	tanah	-nah	l:n	a:a	0:h	
bird	lok8	manuk	-nuk	l:n	o:u	k:k	1
child, son	luk8(tsa:i2)	anak	-nak	l:n	u:a	k:k	2
this	lai4	ini	-ni	l:n	a:i	i:0	1
black	lam6	hitam	-tam	l:t	a:a	m:m	1
pig	mu1	babi	-bi	m:b	u:i	0:0	
sweet photato	man2	ubi	-bi	m:b	a:i	n:0	
shouder	(ho1)ma5; ma5	bahu	ba-	m:b	a:a	0:0	
reamer	mit8	sabit	-bit	m:b	i:i	t:t	
ant	mot8	semut	-mut	m:m	o:u	t:t	
come, arrive	ma2	mari	ma-	m:m	a:a	0:0	1
fish	pa6	patin	pa-	p:p	a:a	0:0	1
wash	suk8	basuh	-suh	s:s	u:u	k:h	2
eye	ta6	mata	-ta	t:t	a:a	0:0	1
door	(la3)tu6	pintu	-tu	t:t	u:u	0:0	
fart	tot9	qentut	-tut	t:t	o:u	t:t	
fall	tok9; tok9(ha:i1)	jatuh	-tuh	t:t	o:u	k:h	2
rightside	xa1	kanan	ka-	x:k	a:a	0:0	2

Summary of the related words in ranks as follows:

	Related words
High rank	9
Low rank	5

The 9 percentage in the high rank is larger that 5 percentage in the low rank. Such distinction is consistent with the distribution of related words due to language split. Therefore, it is suggested that Dai and Indonesian is genetically related.

4.2 Rank Analysis on Incomplete Sound Correspondences

The complete sound correspondence is a rigorous requirement, which can exclude the chance resemblance. However, incomplete sound correspondences are still usable, especially if two of three kinds of sound correspondences, OSC, NSC and ESC, are established for a morpheme. There are two possibilities for incomplete sound correspondences. Firstly, they are due to chance resemblance. Secondly, they are due to language contact or language split, but there are too few examples to establish sound correspondences. In the first situation, the distribution of "related morphemes" among the rank should be level since it is caused by chance without any bias. In the second situation, either language contact or language split will cause the distinction between two ranks.

Now we are going to analyze the distribution of "related morpheme" supported by more than one kind of sound correspondences.

Lexical items	Dai	Indonesian	In-root	OSC	NSC	ESC	R
fire	fai2	api	-pi		a:i	i:0	1
flea	mat7	kutu	-tu		a:u	t:t	
ash	tau6	pirau	-rau		a:a	u:u	1
green	xeu1	hijau	-jau		e:a	u:u	1
branches	xa6;ŋa6	tjagak	-gak		a:a		
seed	fan2	bibit	bi-	f:b)	a:i	n:0	1
dream	fan1	mimpi	-pi	f:p)	a:i	n:0	
understand	hu4	tahu	-hu	h:h)	u:u	0:0	1
tiny chaff	ham2	sekam	-kam	h:k)	a:a	m:m	
pair	ku6	gu	gu	k:g	u:u	0:0	
bite;gnaw	kat7	gitgit	-git	k:g	a:i	t:t	1
eat	kin6	makan	-kan	k:k	i:a)	n:n	1
I	kau6	aku	-ku	k:k	a:u	u:0)	1
nose	(hu2) laŋ6	hiduŋ	-duŋ	l:d	a:u	ŋ:ŋ)	1
gallbladder	li6	empedu	-du	l:d	i:u)	0:0	
moon	lən6	bulan	-lan	l:l	ə:a	n:n	1
tongue	lin4	lidah	li-	l:l	i:i	n:0	1
deep(water)	lək8	djeluk	-luk	l:l	ə:u)	k:k	
farmfield	la2	tanah	-nah	l:n	a:a	0:h	
bird	lok8	manuk	-nuk	l:n	o:u	k:k	1

child; son	luk8 (tsa:i2)	anak	-nak	l:n	u:a	k:k	2
young; tender	lum5	anom	-nom	l:n	u:o)	m:m	
this	lai4	ini	-ni	l:n	a:i	i:0	1
top, above	(pa3)lə1	atas	-tas	l:t	ə:a		
mice	lu1	tikus	ti-	l:t	u:i		
thick	la1	tebal	te-	l:t	a:e)	0:0	2
black	lam6	hitam	-tam	l:t	a:a	m:m	1
pig	mu1	babi	-bi	m:b	u:i	0:0	
sweet photato	man2	ubi	-bi	m:b	a:i	n:0	
fruit	ma:k9	buah	buah	m:b	a:ua)	k:h	2
shoulder	(ho1)ma5; ma5	bahu	ba-	m:b	a:a	0:0	
reamer	mit8	sabit	-bit	m:b	i:i	t:t	
ant	mot8	semut	-mut	m:m	o:u	t:t	
come, arrive	ma2	mari	ma-	m:m	a:a	0:0	1
you	matɯ2	kamu	-mu	m:m	a:u	ɯ:0)	1
sesame	ŋa2lo5	ləŋa	-ŋa	ŋ:ŋ)	a:a	0:0	
blow	pau5	hembus	-bus	p:b)	a:u		2
fish	pa6	patin	pa-	p:p	a:a	0:0	1
wash	suk8	basuh	-suh	s:s	u:u	k:h	2
you(pl.)	su1	saudara	sau-	s:s	u:a	0:u)	
eye	ta6	mata	-ta	t:t	a:a	0:0	1
door	(la3)tu6	pintu	-tu	t:t	u:u	0:0	
die	ta:i6	mati	-ti	t:t	aa:i)	i:0	1
fart	tot9	qentut	-tut	t:t	o:u	t:t	
fall	tok9; tok9 (ha:i1)	jatuh	-tuh	t:t	o:u	k:h	2
leg	xa1	paha	-ha	x:h)	a:a	0:0	2
frog	xet9	katak	-kat	x:k)	e:a	t:t	
laugh, smile	xo1	dekah	-kah	x:k	o:a)	0:h	2
night	(ka: ŋ6	semalam	-malam	x:l)	a:a	m:m	1

	)xam6					

The result is summarized in the following table：

	Related morpheme
High rank	20
Low rank	8

The percentage of related morpheme in high rank, 20%, is more than that in low rank, 8%. Therefore, two remarks may be made: 1. The distinction of distribution of related morphemes between high rank and low rank is quite obvious. Some incomplete examples should be related morphemes indeed. However, there is no method available to detect them. 2. The calculation confirms the result based on complete sound correspondences. The genetic relatedness of Dai and Indonesian is further confirmed.

4.3 Interpretation of Rank Analysis

Since the Dai-Indonesian complete sound correspondences are too few, it is difficult to explain the sound change occurred, and stratify layers of those correspondences. Based on the distribution of related morphemes between two ranks, the ratio in high rank is higher than that in low rank. Such distribution could not be only due to contact. Is it possible that first language contact and later language split result in this distribution? Sino-Tai related morphemes in the layer of bronze and domestic animals were borrowed into Kam-Tai from Chinese due to their contact at the Bronze time. After that, they are separate. (Chen 2004) These related morphemes distributed more in low rank than high rank. Such distribution is different from the behavior of Dai-Indonesian related morphemes. Therefore, there is no enough evidence to argue that the distribution of Dai-Indonesian related morphemes was first caused by language contact and later by split.

How about split first plus contact later? Since language contact always increases the ratio of related morphemes in low rank, this hypothesis may bring two possible outcomes. Firstly, the later contact changes the earlier distribution of related morphemes due to language split, higher ratio in high rank. The later distribution may be that ratio in high rank is equal to that in low rank, or even lower than that in low rank. This does not fit the actual distribution of Dai-Indonesian related morphemes. Another

possibility is that, the later contact is not heavy enough to interfere with the earlier distribution due to language split. Therefore, it still keeps the higher ratio in high rank, which is consistent with distribution of Dai-Indonesian related morphemes.

If there is no contact after two languages split from a common ancestor, the related morphemes in high rank are surely more than those in low rank. Such hypothesis satisfies the distribution of Dai-Indonesian related morphemes.

In summary, both of the two possible hypotheses indicate that Dai and Indonesian are genetically related.

5. TRANSITIVITY OF GENETIC RELATEDNESS AND THE AUSTRO-YUE FAMILY

According to the analysis in chapter 4, it may be concluded that Dai and Indonesian are genetically related. The genetic relatedness is of transitivity (Chen 1994). That is to say, if language A and language B are genetically related, meanwhile, language B and language C are genetically related, then it could be deduced that language A and language C are genetically related, too. In this spirit, by rank analysis, we will first confirm the genetic relatedness among Austronesian languages one of which is Indonesian, and then genetic relatedness among Kam-Tai languages one of which is Dai, therefore, the genetic relatedness between Proto-Kam-Tai and Proto-Austronesian will be supported.

5.1 The Genetic Relationship of Austronesian Languages

Early reconstructions of Proto-Austronesian were trying to prove the genetic relatedness of Austronesian languages based on sound correspondences, which has been argued to be necessary but not sufficient condition for genetic relatedness in several occasions. Now, we are applying rank analysis to Austronesian languages.

Taking into consideration materials from Ho (1999), the kernel morphemes of Austronesian languages are listed in the following table: (PAN=Proto-Austronesian，PA=Proto-Tayic，PP=Proto-Paiwanic，PT=Proto-Tsou, henceforth.)

Lexical items	Index	PAN	PA	PP	PT	R
ear	7	caliŋaɦ	caŋiraʔ	caljŋa	caĺiŋaɦa	1
two	32	ɖusa	dusaʔ	ɖusa	řúSa	1

give	78	bəgay	bəgay	pa-vai		1
bone	39	cuqəlai		cuqəlai	cuʔúlałə	1
fire	10	śapuy	hapuy	sapuy	apúžu	1
horn	50	təquŋ		təquŋ	suʔ₁úŋu	1
name	55	ŋadan		ŋadan	ŋázánə	1
you	28	su	ʔisuʔ	su-	su	1
stone	91	batuɦ	batu-nux		vátuɦu	1
water	73	jalum		zalum	čahlúmu	1
head louse	18	kucuɦ	kucuʔ	kucu	kúcúɦu	1
I	15	(a)ku	akuʔ	ku-	́aku	1
new	56	vaquɦ		vaqu-an	vaʔ₂əruɦu	1
blood	4	daga[]	dagaʔ	daq	cará ʔ₁ə	1
eye	45	maca		maca	macá	1
one	58	́ta		ita	cáni	1
moon	88	bulal	bural		vuláhlə	1
this	93	(i)niɦ	ni		iniɦi	1
eat	8	kan	kan	k-əm-an	k₁-um-ánə	1
liver	52	qacay		qacay	ʔ₁₄acayi	1
ash	2	qabu	qabu-liq	qavu	ʔ2avuʔ4u	1
see	90	kita	kitaʔ		kíta	1
road	24	dalan	daran	ɗalan	čalánə	1
skin	38	kulic	luliC(bark)	kulic	kul ́ici	1
hot	49	jaŋjaŋ		zaŋzaŋ	čaŋəčaŋə	1
person	59	caw		cawcaw	cáw	1
breats	5	zuzuh	nunuh	tutu	ɵuɵu	1
kill	42	macay		pa-pacay	pacáyi	1
tree	31	kaśuy	kahuy	hasiw	káiwu	1
stomach	69	bicuka		vicuka	civúka	1
we(inclusive)	95	ita	ʔita		(-ita)	1
bite	3	kagac	k-um-agac	k-ə-ac	k₁-um-áraca	1
swim	27	laŋuy	l-um-aŋuy	l-əm-aqis	laŋúžu	1
rain	60	qudal		qudal	ʔ₂účahjə	1

father	47	amaɦ		k-ama	ámaɦa	2
dig	6	kaliɦ	kariʔ	k-əm-ali	ˊkaliɦi	2
five	11	lima	rimaʔ	lima	lˊima	2
flow	12	qańud	qaluic	səqalud	-ʔ₂añúču	2
four	13	se-pat	səpat	səpa	Sǝ́pátə	2
back	36	likuj		likuz	(liku[cřč])	2
left side	17	wiri[]	ʔiril	ka-viri	wˊiríɦi	2
living	35	qujip		pa-quzip	-ʔ₂učípi	2
mother	54	-inaɦ		k-ina	inaɦa	2
string	63	ṭalis		calis	talíSi	2
sew	26	ṭaqiś	c-cum-aqis	c-əm-aqis	t-um-áʔ₃iɵi	2
thin	92	liśipis	hlipis		hlˊipisi	2
three	30	tǝlu	təruʔ	ˊtǝlu	túlu	2
vomit	83	muˊtaq	mutaq	muˊtaq		2
wash	72	sinaw		s-əm-ənaw	sináwu	2
child	40	alak		alak	-ahlákə	2

The result is summarized into the following table:

	Related morphemes
High rank	34
Low rank	16

The ratio of related morphemes in high rank, 34%, is significantly higher than that in low rank, 16%, which suggests the genetic relatedness of Austronesian languages.

If it is required that the reflexes of a proto-form distribute in all subgroups, the kernel morphemes of Proto-Austronesian would be recounted in the following table:

Lexical items	Index	PAN	PA	PP	PT	R
ear	7	caliŋaɦ	caɲiraʔ	caljŋa	calˊiŋaɦa	1
two	32	ḍusa	dusaʔ	ḍusa	řúSa	1
fire	10	śapuy	hapuy	sapuy	apúžu	1
you	28	su	ʔisuʔ	su-	su	1
head louse	18	kucuɦ	kucuʔ	kucu	kúcúɦu	1

I	15	(a)ku	aku?	ku-	ʿaku	1
blood	4	daga[]	daga?	daq	cará?₁ə	1
eat	8	kan	kan	k-əm-an	k₁-um-ánə	1
ash	2	qabu	qabu-liq	qavu	?2avu?4u	1
see	90	kita	kita?		kíta	1
path, road	24	dalan	daran	ɗalan	čálánə	1
skin	38	kulic	luliC(bark)	kulic	kulíci	1
breasts	5	zuzuh	nunuh	tutu	өuөu	1
tree	31	kaśuy	kahuy	hasiw	káiwu	1
bite	3	kagac	k-um-agac	k-ə-ac	k₁-um-áraca	1
swim	27	laŋuy	l-um-aŋuy	l-əm-aqis	laŋúžu	1
dig	6	kalifi	kari?	k-əm-ali	ʿkalifi	2
five	11	lima	rima?	lima	líma	2
flow	12	qańud	qaluic	səqalud	-?₂añúču	2
four	13	se-pat	səpat	səpa	Sə́pátə	2
left-side	17	wiri[]	?iril	ka-viri	wírífi	2
sew	26	ʈaqiś	c-cum-aqis	c-əm-aqis	t-um-á?₃iөi	2
three	30	təlu	təru?	ʈəlu	túlu	2

The statistic result would be in the following table:

	Related morphemes
High rank	15
Low rank	7

The ratio of related morphemes in high rank, 15%, is significantly higher than that in low rank, 7%, which suggests the genetic relatedness of Austronesian languages also.

5.2 The Genetic Relationship of Kam-Tai Languages

It is generally accepted that Kam-Tai languages are from a common ancestor. The rank analysis has confirmed this claim (Chen 1994:232-4; 2004). To make more certain, let's do further rank analysis on the related morphemes between Kam-Tai given recently by Ostapirat (2005:110):

Lexical items	Tai	Kamsui	Hlai	Gelao	KDai tone	R
moon	dɯɯan	njaan	ŋaːn	daan(LH)	a	1
water	naam	nam	nom	(əɯ C)	c	1
fire	fai	wii	pei	pai	a	1
fish	plaa	paa(K)	da	lau	a	1
head louse	hau	tuu	tshou	ta	a	1
horn	khau	qaau	hau(Bd)	qa	a	1
eye	taa	daa	tsha	tau	a	1
nose	daŋ	ʔnaŋ	doŋ	daŋ(LH)	a	1
hand	mɯɯɯ	mjaa	meɯ	mpau	a	1
this	nii	naai	nei	ni	b/c	1
you	mɯɯŋ	maa(Lk)	meɯ	maa(By)	a	1
I	kuu	(ju)	hou(BD)	kuu(By)	a	1
tooth	fan	wjan	phen	pan	a	1
full	(tem)	tik	thiːʔ	tei	d	1
dog	maa	hmaa	ma	mpau	a	1
blood	lɯɯat	phjaat	daːt	plɒ	d	1
bone	duuk	laak	rɒːʔ	taŋ	d	1
tail	(haaŋ)	hət	tshut	tshan	d	1
ear	huu	qhaa	(zai)	zau	a	1
stone	hin	tin	tshiːn	(pɣaa)	a	1
cloud	faa	faa	fa(BD)	phaa	c	1
bird	nok	nok	(taƫ)	ntau	d	1
seed	fan(WT)	wan	phen	pa(QS)	a	1
head	klau	ku	rau	(klɒ B)	c	1
knee	klau	quu	(rou)	qo(LZ)	b	1
live	tap	tap	(ŋaːn)	tæ(LZ)	d	1
oil	man	man	man(B)	mal(LH)	a	1
road	hon	khun	kuːn	qan	a	1
long	rii	ʔɣaai	loi(B)	ðii C(BY)	a	1

black	dam	ʔnam	dom C	ʔdam(BY)	a	1
dry	khatɯ	khu C	khetɯ	xau	b	1
smoke	khwan	kwan	hwo:n	qɒ	a	1
green	khiau	cju	khi:u	(ten)	a	1
walk	pai	paai	pei	pai	a	1
kill	khaa	haa	hau	(ven)	c	1
eat	kin	tsjaan	khan (BC)	kaan(BY)	a	1
come, arrive	maa	hmaa	petɯ(BD)	mu	a	1
rain	fon	fən	pun	(jal)(LH)	a	1
leaf	bai	waa(LK)	betɯ	(vu)	a	1
leg	khaa	qaa	ha	qau	a	2
child	luuk	laak	dɯ:ʔ	lei	d	2
living	dip	ʔdjup	ri:p	te	d	2
bear	mii	ʔmii	mui	mi(LZ)	a	
sesame	ŋaa	ʔŋaa	ketɯ(BD)	ŋklau	a	
shoulder	baa	wie(LK)	va	baa(LH)	b	
nail	lep	ljap	li:p	kle	d	
navel	dɯɯ	ʔdwaa	retɯ	zo(QS)	a	
excrement	khii	qee	hai	qɒ	c	
grandmother	jaa B	jaa C	tsatɯ (BD)	zjɒ C	b/c	
fart	tot	tət	thu:t	tæ(LZ)	d	

The rank analysis can be summarized in the following table:

	Related morphemes
High rank	39
Low rank	3

The ratio of related morphemes in the high rank, 39%, is significantly higher than that in the low rank, 3%, which suggests the genetic relatedness of Kam-Tai languages.

In summary, independent studies point out the same conclusion that Kam-Tai languages and Austronesian languages are genetically related internally, respectively. This may serve as a good basis to use transtivity of genetic relatedness. Dai as a Kam-Tai language and Indonesian as an Austronesian language can bring together Kam-Tai and Austronesian. Since Dai and Indonesian are genetically related, Kam-Tai and Austronesian should be gentically related, too.

5.3 The Austro-Yue Family (Austronesian and Kam-Tai)

Now according to transitivity of genetic relatedness and rank analysis, we are to assure you further of the genetic relationship between Austronesian and Kam-Tai.

Benedict's Austro-Tai sometimes may be misunderstood because Hmong-Mien is often included in the "Tai" group. We would introduce another name for our result. In ancient China, people speaking Kam-Tai languages were called Bai-Yue. So the genetic group of Austronesian and Kam-Tai can be called Austro-Yue. Here the "Yue" instead of "Tai" covers Kam-Tai without Hmong-Mien.

6. EVIDENCE FROM COMPARISONS OF PROTO-LANGUAGES

6.1 Proto-Tai-Indonesian Sound Correspondences

According to the above comparison of modern languages, Dai and Indonesian, their genetic relationship is defined. Via the transitivity of genetic relatedness, the genetic relationship of Kam-Tai and Austronesian is further confirmed in previous paragraphs. The reason we do not use the reconstructed materials is discussed before. However, if the reconstruction of a proto-language is generally accepted, it can be also used in comparison and rank analysis. Proto-Tai (PT) reconstructed in Li (1977) has been widely used without major modification in these years. It may be compared with Indonesian.

6.1.1 Onset Sound Correspondence between Proto-Tai and Indonesian

OSC	Lexical items	Dai	PT	Indonesian	In-root	NSC	ESC	R
	pond		*dəm2	kolam		ə:a	m:m	
	fire	fai2	*vɛi2	api	-pi	ɛ:i	i:0	1
	right side	xa1	*khwa1	kanan	ka-	a:a	0:0	2
	branch	xa6	*ɣa6	tjagak		a:a		

	salt	kə6	*klɯe1	garam	ga-	ɯe:a	0:0	2
	tiny chaff	ham2	*rəm2	sekam	-kam	ə:a	m:m	
	ash	tau6	*vləu6	pirau	-rau	ə:a	u:u	1
	blow	pau5	*pou5	hembus	-bus	o:u		2
	laugh	xo1	*xrue1	dekah	-kah	ue:a	0:h	2
	green	xeu1	*khiau1	hijau	-jau	a:a	u:u	1
	night	(ka:ŋ6)xam6	*ɣəm6	semalam	-malam	ə:a	m:m	1
ʔb:b)	shoulder	ho1ma5; ma5	*ʔba5	bahu	ba-	a:a	0:0	
ʔbl:l)	moon	lən6	*ʔbl/rɯen1	bulan	-lan	ɯe:a	n:n	1
ʔd:d)	nose	(hu2)laŋ6	*ʔdəŋ1	hiduŋ	–duŋ	ə:u	ŋ:ŋ)	1
ʔdl:t)	black	lam6	*ʔdl/rəm1	hitam	-tam	ə:a	m:m	1
f:p)	dream	fan1	*fən1	mimpi	-pi	ə:i	n:0	
g:g)	pair	ku6	*gu6	gu	gu	u:u	0:0	
hm:b	pig	mu1	*hmu1	babi	-bi	u:i	0:0	
hm:b	fruit	ma:k9	*hmak9	buah*	buah	a:a	k:h	2
hm:b	new	maɯ5	*hmoɯ5	baru	ba-			1
hm:k)	flea	mat7	*hmət7	kutu	-tu	ə:u	t:t	
hn:n	young; tender	lum5	*hnum5	anom	-nom	u:o)	m:m	
hn:n	that		*hna3	sana		a:a	0:0	1
hn:t	top, above	pa3lə1	*hnɯe1	atas	-tas	ɯe:a		
hn:t	mice	lu1	*hnu1	tikus	ti-	u:i		
hn:t	thick	la1	*hna1	tebal	te-	a:e	0:0	2
hn:t	tight	xeŋ6	*hneet9	ketat	-tat		t:t	
k:g)	bite;gnaw	kat7	*kət7	gitgit	-git	ə:i	t:t	1
k:k	eat	kin6	*kin1	makan	-kan	i:a)	n:n	1
k:k	I	kau6	*ku1	aku	-ku	u:u	0:0	1
kh:h)	leg	xa1	*kha1	paha	-ha	a:a	0:0	2
kl:k)	overlay	lop9	*klop7	kup	kup	o:u	p:p	
l:l	tongue	lin4	*lin4	lidah	li-	i:i	n:0	1

57

l:l	forget	lɯm2	*lɯɯm2	lupa	lu-			
l:l	deep (water)	lək8	*lɯɯk9	djeluk	-luk	ɯ:u	k:k	
l:n)	child; son	luk8(ts a:i2)	*luk10	anak	-nak	u:a)	k:k	2
m:b	sweet photato	man2	*mən2	ubi	-bi	ə:i	n:0	
m:b	reamer	mit8	*mit10	sabit	-bit	i:i	t:t	
m:m	ant	mot8	*mot8	semut	-mut	o:u	t:t	
m:m	come,arr ive	ma2	*ma2	mari	ma-	a:a	0:0	1
m:m	you	maɯ2	*mɯɯŋ2	kamu	-mu	ɯ:u		1
n:n	farmfield	la2	*na2	tanah	-nah	a:a	0:h	
n:n	this	lai4	*nɛi4	ini	-ni	ɛ:i	i:0	1
ŋl:ŋ)	sesame	ŋa2 (lo5)	*ŋl/ra2	ləŋa	‾ŋa	a:a	0:0	
nr:n)	bird	lok8	*nl/rok8	manuk	-nuk	o:u	k:k	1
p:p)	circumro tate	pan5	*pən5	putar	pu-			
pl:p)	fish	pa6	*pla1	patin	pa-	a:a	0:0	1
r:h)	understa nd	hu4	*ru4	tahu	-hu	u:u	0:0	1
t:t	door	la3tu6	*tu1	pintu	-tu	u:u	0:0	
t:t	fall	tok9; tok9(h a:i1)	*tok7	jatuh	-tuh	o:u	k:h	2
tl:t)	fart	tot9	*tlot7	qentut	-tut	o:u	t:t	
tr:t	eye	ta6	*tra1	mata	-ta	a:a	0:0	1
tr:t	die	ta:i6	*trai1	mati	-ti	a:i)	i:0	1
v:b)	seed	fan2	*vən2	bibit	bi-	ə:i	n:0	1
z:s)	wash	suk8	*zuk8	basuh	-suh	u:u	k:k	2

6.1.2 Nucleus Sound Correspondence between Proto-Tai and Indonesian

NSC	Lexical items	Dai	PT	Indonesian	In-root	OSC	ESC	R

	new	maɯ5	*hmoɯ5	baru	ba-	hm:b		1
	tight	xeŋ6	*hneet9	ketat	-tat	hn:t	t:t	
	forget	lʊm2	*lɯʊm2	lupa	lu-	l:l		
a:a	right side	xa1	*khwa1	kanan	ka-		0:0	2
a:a	branch	xa6	*ɣa6	tjagak				
a:a	green	xeu1	*khiau1	hijau	-jau		u:u	1
a:a	shoulder	ho1ma5; ma5	*ʔba5	bahu	ba-	ʔb:b)	0:0	
a:a	fruit	ma:k9	*hmak9	buah	buah	hm:b	k:h	2
a:a	that		*hna3	sana		hn:n	0:0	1
a:a	leg	xa1	*kha1	paha	-ha	kh:h)	0:0	2
a:a	come, arrive	ma2	*ma2	mari	ma-	m:m	0:0	1
a:a	farmfield	la2	*na2	tanah	-nah	n:n	0:h	
a:a	sesame	ŋa2lo5	*ŋl/ra2	ləŋa	-ŋa	ŋl:ŋ)	0:0	
a:a	fish	pa6	*pla1	patin	pa-	pl:p)	0:0	1
a:a	eye	ta6	*tra1	mata	-ta	tr:t	0:0	1
a:e	thick	la1	*hna1	tebal	te-	hn:t	0:0	2
a:i)	die	ta:i6	*trai1	mati	-ti	tr:t	i:0	1
ɛ:i	fire	fai2	*vɛi2	api	-pi		i:0	1
ɛ:i	this	lai4	*nɛi4	ini	-ni	n:n	i:0	1
ə:a	pond		*dəm2	kolam			m:m	
ə:a	tiny chaff	ham2	*rəm2	sekam	-kam		m:m	
ə:a	ash	tau6	*vləu6	pirau	-rau		u:u	1
ə:a	night	ka:ŋ6xam6	*ɣəm6	semalam	-malam		m:m	1
ə:a	black	lam6	*ʔdl/rəm1	hitam	-tam	ʔdl:t)	m:m	1
ə:i	dream	fan1	*fən1	mimpi	-pi	f:p)	n:0	
ə:i	bite; gnaw	kat7	*kət7	gitgit	-git	k:g)	t:t	1
ə:i	sweet photato	man2	*mən2	ubi	-bi	m:b	n:0	
ə:i	seed	fan2	*vən2	bibit	bi-	v:b)	n:0	1
ə:u	nose	hu2laŋ6	*ʔdəŋ1	hiduŋ	-duŋ	ʔd:d)	ŋ:ŋ)	1

ə:u	flea	mat7	*hmət7	kutu	-tu	hm:k)	t:t	
i:a)	eat	kin6	*kin1	makan	-kan	k:k	n:n	1
i:i	tongue	lin4	*lin4	lidah	li-	l:l	n:0	1
i:i	reamer	mit8	*mit10	sabit	-bit	m:b	t:t	
o:u	blow	pau5	*pou5	hembus	-bus			2
o:u	overlay	lop9	*klop7	kup	kup	kl:k)	p:p	
o:u	ant	mot8	*mot8	semut	-mut	m:m	t:t	
o:u	bird	lok8	*nl/rok8	manuk	-nuk	nr:n)	k:k	1
o:u	fall	tok9; tok9(h a:i1)	*tok7	jatuh	-tuh	t:t	k:h	2
o:u	fart	tot9	*tlot7	qentut	-tut	tl:t)	t:t	
u:a)	child; son	luk8(ts a:i2)	*luk10	anak	-nak	l:n)	k:k	2
u:i	pig	mu1	*hmu1	babi	-bi	hm:b	0:0	
u:i	mice	lu1	*hnu1	tikus	ti-	hn:t		
u:o)	young;te nder	lum5	*hnum5	anom	-nom	hn:n	m:m	
u:u	pair	ku6	*gu6	gu	gu	g:g)	0:0	
u:u	I	kau6	*ku1	aku	-ku	k:k	0:0	1
u:u	understa nd	hu4	*ru4	tahu	-hu	r:h)	0:0	1
u:u	door	la3tu6	*tu1	pintu	-tu	t:t	0:0	
u:u	wash	suk8	*zuk8	basuh	-suh	z:s)	k:k	2
ɯ:u	deep (water)	lək8	*lɯk9	djeluk	-luk	l:l	k:k	
ɯ:u	you	matɯ2	*mɯŋ2	kamu	-mu	m:m		1
ɯe:a	salt	kə6	*klɯe1	garam	ga-		0:0	2
ɯe:a	moon	lən6	*ʔbl/rɯe n1	bulan	-lan	ʔbl:l)	n:n	1
ɯe:a	top, above	pa3lə1	*hnɯe1	atas	-tas	hn:t		
ue:a)	laugh	xo1	*xrue1	dekah	-kah		0:h	2

6.1.3 Ending Sound Correspondence between Proto-Tai and Indonesian

ESC	Lexical items	Dai	PT	Indonesian	In-root	OSC	NSC	R
	new	maɯ5	*hmoɯ5	baru	ba-	hm:b		1
	forget	ltɯm2	*ltɯtɯm2	lupa	lu-	l:l		
	blow	pau5	*pou5	hembus	-bus		o:u	2
	mice	lu1	*hnu1	tikus	ti-	hn:t	u:i	
	you	maɯ2	*mɯɯŋ2	kamu	-mu	m:m	ɯ:u	1
	top, above	pa3lə1	*hntɯe1	atas	-tas	hn:t	ɯe:a	
0:0	right side	xa1	*khwa1	kanan	ka-		a:a	2
0:0	shoulder	ho1ma5; ma5	*ʔba5	bahu	ba-	ʔb:b)	a:a	
0:0	that		*hna3	sana		hn:n	a:a	1
0:0	leg	xa1	*kha1	paha	-ha	kh:h)	a:a	2
0:0	come, arrive	ma2	*ma2	mari	ma-	m:m	a:a	1
0:0	sesame	ŋa2lo5	*ŋl/ra2	ləŋa	-ŋa	ŋl: ŋ)	a:a	
0:0	fish	pa6	*pla1	patin	pa-	pl:p	a:a	1
0:0	eye	ta6	*tra1	mata	-ta	tr:t	a:a	1
0:0	thick	la1	*hna1	tebal	te-	hn:t	a:e	2
0:0	pig	mu1	*hmu1	babi	-bi	hm:b	u:i	
0:0	pair	ku6	*gu6	gu	gu	g:g)	u:u	
0:0	I	kau6	*ku1	aku	-ku	k:k	u:u	1
0:0	understand	hu4	*ru4	tahu	-hu	r:h)	u:u	1
0:0	door	la3tu6	*tu1	pintu	-tu	t:t	u:u	
0:0	salt	kə6	*klɯe1	garam	ga-		ɯe:a	2
0:h	farmland	la2	*na2	tanah	-nah	n:n	a:a	
0:h	laugh	xo1	*xrue1	dekah	-kah		ue:a)	2
i:0	die	ta:i6	*trai1	mati	-ti	tr:t	a:i)	1
i:0	fire	fai2	*vɛi2	api	-pi		ɛ:i	1
i:0	this	lai4	*nɛi4	ini	-ni	n:n	ɛ:i	1
k:h	fruit	ma:k9	*hmak9	buah	buah	hm:b	a:a	2
k:h	fall	tok9	*tok7	jatuh	-tuh	t:t	o:u	2
k:k	bird	lok8	*nl/rok8	manuk	-nuk	nr:n)	o:u	1
k:k	child;	luk8	*luk10	anak	-nak	l:n)	u:a)	2

	son	(tsa:i2)						
k:k	wash	suk8	*zuk8	basuh	-suh	z:s)	u:u	2
k:k	deep (water)	lək8	*lʊɪk9	djeluk	-luk	l:l	ɯ:u	
m:m	pond		*dəm2	kolam			ə:a	
m:m	tiny chaff	ham2	*rəm2	sekam	-kam		ə:a	
m:m	night	(ka:ŋ6)xam6	*ɣəm6	semalam	-malam		ə:a	1
m:m	black	lam6	*ʔdl/rəm1	hitam	-tam	ʔdl:t)	ə:a	1
m:m	young;tender	lum5	*hnum5	anom	-nom	hn:n	u:o)	
n:0	dream	fan1	*fən1	mimpi	-pi	f:p)	ə:i	
n:0	sweet photato	man2	*mən2	ubi	-bi	m:b	ə:i	
n:0	seed	fan2	*vən2	bibit	bi-	v:b)	ə:i	1
n:0	tongue	lin4	*lin4	lidah	li-	l:l	i:i	1
n:n	eat	kin6	*kin1	makan	-kan	k:k	i:a)	1
n:n	moon	lən6	*ʔbl/rɯen1	bulan	-lan	ʔbl:l)	ɯe:a	1
ŋ:ŋ)	nose	hu2laŋ6	*ʔdəŋ1	hiduŋ	−duŋ	ʔd:d)	ə:u	1
p:p	overlay	lop9	*klop7	kup	kup	kl:k)	o:u	
t:t	tight	xeŋ6	*hneet9	ketat	-tat	hn:t		
t:t	bite; gnaw	kat7	*kət7	gitgit	-git	k:g)	ə:i	1
t:t	flea	mat7	*hmət7	kutu	-tu	hm:k)	ə:u	
t:t	reamer	mit8	*mit10	sabit	-bit	m:b	i:i	
t:t	ant	mot8	*mot8	semut	-mut	m:m	o:u	
t:t	fart	tot9	*tlot7	qentut	-tut	tl:t)	o:u	
u:u	green	xeu1	*khiau1	hijau	-jau		a:a	1
u:u	ash	tau6	*vləu6	pirau	-rau		ə:a	1

6.1.4 Rank Analysis of Proto-Tai-Indonesian Complete Correspondences

Only taking into consideration complete correspondences, their distribution

among ranks are listed in the following table.

OSC	Lexical items	Dai	PT	Indonesian	In-root	NSC	ESC	R
hm:b	pig	mu1	*hmu1	babi	-bi	u:i	0:0	
hm:b	fruit	ma:k9	*hmak9	buah	buah	a:a	k:h	2
hn:t	thick	la1	*hna1	tebal	te-	a:e	0:0	2
k:k	I	kau6	*ku1	aku	-ku	u:u	0:0	1
l:l	tongue	lin4	*lin4	lidah	li-	i:i	n:0	1
l:l	deep (water)	lək8	*lɯk9	djeluk	-luk	ɯ:u	k:k	
m:b	sweet photato	man2	*mən2	ubi	-bi	ə:i	n:0	
m:b	reamer	mit8	*mit10	sabit	-bit	i:i	t:t	
m:m	ant	mot8	*mot8	semut	-mut	o:u	t:t	
m:m	come, arrive	ma2	*ma2	mari	ma-	a:a	0:0	1
m:m	you	maɯ2	*mɯɯŋ2	kamu	-mu	ɯ:u		1
n:n	farmfield	la2	*na2	tanah	-nah	a:a	0:h	
n:n	this	lai4	*nɛi4	ini	-ni	ɛ:i	i:0	1
t:t	door	la3tu6	*tu1	pintu	-tu	u:u	0:0	
t:t	fall	tok9; tok9(ha:i1)	*tok7	jatuh	-tuh	o:u	k:h	2
tr:t	eye	ta6	*tra1	mata	-ta	a:a	0:0	1
tr:t	die	ta:i6	*trai1	mati	-ti	a:i)	i:0	1

From the above table, 7 examples of complete sound correspondences belong to high rank, while 3 examples belong to low rank. Again, this distribution supports the genetic relationship between Proto-Tai and Indonesian.

6.1.5 Analysis of Proto-Tai-Indonesian Incomplete Correspondences

Now taking in consideration the incomplete correspondences between Proto-Tai and Indonesian in the same principle applied in chapter 4.2:

Lexical	Dai	PT	Indonesian	In-root	OSC	NSC	ESC	R

items								
pond		*dəm2	kolam			ə:a	m:m	
fire	fai2	*vɛi2	api	-pi		ɛ:i	i:0	1
right side	xa1	*khwa1	kanan	ka-		a:a	0:0	2
salt	kə6	*klɯe1	garam	ga-		ɯe:a	0:0	2
tiny chaff	ham2	*rəm2	sekam	-kam		ə:a	m:m	
ash	tau6	*vləu6	pirau	-rau		ə:a	u:u	1
green	xeu1	*khiau1	hijau	-jau		a:a	u:u	1
night	(ka:ŋ6)xam6	*ɣəm6	semalam	-malam		ə:a	m:m	1
shoulder	ho1ma5; ma5	*ʔba5	bahu	ba-	ʔb:b)	a:a	0:0	
moon	lən6	*ʔbl/rɯen1	bulan	-lan	ʔbl:l)	ɯe:a	n:n	1
black	lam6	*ʔdl/rəm1	hitam	-tam	ʔdl:t)	ə:a	m:m	1
dream	fan1	*fən1	mimpi	-pi	f:p)	ə:i	n:0	
pair	ku6	*gu6	gu	gu	g:g)	u:u	0:0	
pig	mu1	*hmu1	babi	-bi	hm:b	u:i	0:0	
fruit	ma:k9	*hmak9	buah	buah	hm:b	a:a	k:h	2
new	maɯ5	*hmoɯ5	baru	ba-	hm:b			1
flea	mat7	*hmət7	kutu	-tu	hm:k)	ə:u	t:t	
young;tender	lum5	*hnum5	anom	-nom	hn:n	u:o)	m:m	
that		*hna3	sana		hn:n	a:a	0:0	1
top, above	pa3lə1	*hnɯe1	atas	-tas	hn:t	ɯe:a		
mice	lu1	*hnu1	tikus	ti-	hn:t	u:i		
thick	la1	*hna1	tebal	te-	hn:t	a:e	0:0	2
tight	xeŋ6	*hneet9	ketat	-tat	hn:t		t:t	
bite; gnaw	kat7	*kət7	gitgit	-git	k:g)	ə:i	t:t	1
eat	kin6	*kin1	makan	-kan	k:k	i:a)	n:n	1
I	kau6	*ku1	aku	-ku	k:k	u:u	0:0	1
leg	xa1	*kha1	paha	-ha	kh:h)	a:a	0:0	2
overlay	lop9	*klop7	kup	kup	kl:k)	o:u	p:p	

tongue	lin4	*lin4	lidah	li-	l:l	i:i	n:0	1
deep (water)	lək8	*lɯk9	djeluk	-luk	l:l	ɯ:u	k:k	
sweet photato	man2	*mən2	ubi	-bi	m:b	ə:i	n:0	
reamer	mit8	*mit10	sabit	-bit	m:b	i:i	t:t	
ant	mot8	*mot8	semut	-mut	m:m	o:u	t:t	
come, arrive	ma2	*ma2	mari	ma-	m:m	a:a	0:0	1
you	maɯ2	*mɯŋ2	kamu	-mu	m:m	ɯ:u		1
farmfield	la2	*na2	tanah	-nah	n:n	a:a	0:h	
this	lai4	*nɛi4	ini	-ni	n:n	ɛi:i	i:0	1
sesame	ŋa2lo5	*ŋl/ra2	ləŋa	-ŋa	ŋl:ŋ)	a:a	0:0	
bird	lok8	*nl/rok8	manuk	-nuk	nr:n)	o:u	k:k	1
fish	pa6	*pla1	patin	pa-	pl:p)	a:a	0:0	1
understa nd	hu4	*ru4	tahu	-hu	r:h)	u:u	0:0	1
door	la3tu6	*tu1	pintu	-tu	t:t	u:u	0:0	
fall	tok9; tok9(h a:i1)	*tok7	jatuh	-tuh	t:t	o:u	k:h	2
fart	tot9	*tlot7	qentut	-tut	tl:t)	o:u	t:t	
eye	ta6	*tra1	mata	-ta	tr:t	a:a	0:0	1
die	ta:i6	*trai1	mati	-ti	tr:t	a:i)	i:0	1
seed	fan2	*vən2	bibit	bi-	v:b)	ə:i	n:0	1
wash	suk8	*zuk8	basuh	-suh	z:s)	u:u	k:k	2

There are 21 examples in the high rank, while only 7 in the low rank, and such distribution suggests the genetic relatedness between Indonesian and Proto-Tai.

6.2 Comparison of Proto-Tai and Proto-Austronesian

Ostapirat (2005) compared Kam-Tai and Proto-Austronesian reconstructed by Blust. The related morphemes identified in Ostapirat (2005) can be used in rank analysis as below.

Lexical items	Pan	Tai	Kamsui	Hlai	Gelao	KD	R

						tone	
moon	bulaN	dɯan	njaan	ŋa:n	daan(LH)	a	1
water	daNum	naam	nam	nom	(əɯ C)	c	1
fire	Sapuy	fai	wii	pei	pai	a	1
head louse	kuCu	hau	tuu	tshou	ta	a	1
eye	maCa	taa	daa	tsha	tau	a	1
nose	ijuŋl	daŋ	ʔnaŋ	doŋ	daŋ(LH)	a	1
hand	(qa)lima	mɯɯɯ	mjaa	meɯ	mpau	a	1
this	i-ni	nii	naai	nei	ni	b/c	1
you	kamu	mɯɯŋ	maa(Lk)	meɯ	maa(By)	a	1
I	aku	kuu	(ju)	hou(BD)	kuu(By)	a	1
tooth	nipen	fan	wjan	phen	pan	a	1
bird	manuk (PMP)	nok	nok	(tatr)	ntau	d	1
head	qulu	klau	ku	rau	(klɒ B)	c	1
grease, oil	simaR	man	man	man(B)	mal(LH)	a	1
black	tidem	dam	ʔnam	dom C	ʔdam(BY)	a	1
eat	kaen	kin	tsjaan	khan (BC)	kaan(BY)	a	1
leaf	(ʔbabag)	bai	waa(LK)	beɯ	(vu)	a	1
leg	paqa	khaa	qaa	ha	qau	a	2
child	aNak	luuk	laak	dɯ:ʔ	lei	d	2
living	qudip	dip	ʔdjup	ri:p	te	d	2
bear	Cumay	mii	ʔmii	mui	mi(LZ)	a	
sesame	leŋa	ŋaa	ʔŋaa	keɯ(BD)	ŋklau	a	
shoulder	qabaRa	baa	wie(LK)	va	baa(LH)	b	
excrement	Caqi	khii	qee	hai	qɒ	c	
grandmother	aya	jaa B	jaa C	tsaɯ (BD)	zjɒ C	b/c	
fart	qe(n)tut	tot	tət	thu:t	tæ(LZ)	d	

The results are calculated in the following table:

	Related morphemes
High rank	17
Low rank	3

The ratio of related morphemes between Kam-Tai and Austronesian in the high rank, 17%, is far higher than that in the low rank. The distribution supports the genetic relatedness of Kam-Tai and Austronesian.

7. A LOOK AT MALAY-YUE AND SINO-TIBETAN

The genetic relationship between Kam-Tai and Austronesian has been demonstrated in previous chapters. Now comes the question - are they genetic related to Sino-Tibetan? If either Kam-Tai or Austronesian is proved to be genetically related to Chinese, the affirmative answer to the question may be generated, because the genetic relatedness between Chinese and Tibeto-Burman is generally accepted. The transitivity of genetical relatedness may connect Chinese, Tibeto-Burman, Kam-Tai and Austronesian.

However, either Sino-Tai or Sino-Austronesian comparison does not prove their genetic relatedness. In Sino-Tai studies, it is found that the more basic the morphemes are, the lower ratio of the related morphemes could be found (Chen 1994, 2004). Therefore, in current stage the Sino-Tai related morphemes based on complete sound correspondences should be treated as borrowing.

In Sino-Austronesian studies, Sagart did a lot of comparative studies since 1990 (Sagart 1990, 1993, 1995, 2005). Meanwhile, his proposal has been criticized by several scholars, such as Blust (1995), Li (1995b), Pulleyblank (1995) and Starostin (1995). Numerous problems such as lack of kernel morphemes in the comparison, unacceptable view on Old Chinese, and some misunderstanding on Austronesian morphology and phonology have been addressed. However, Sagart keeps refining the comparative works. Recently, Sagart (2005) updated his proposal and arguments. The new diagram is shown as follows.
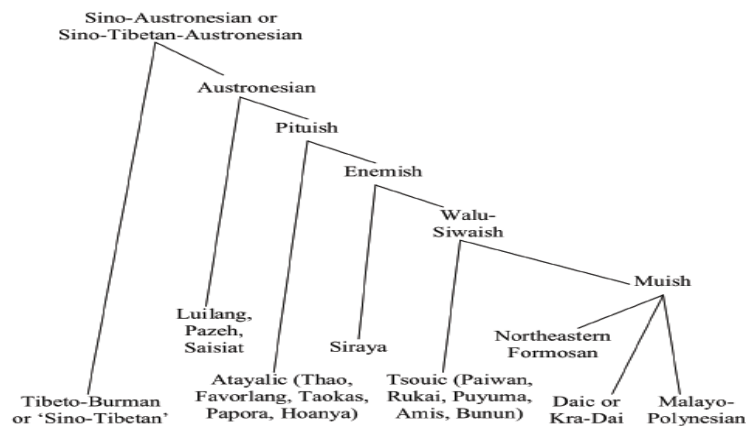
Figure 4 Sagart's new diagram (Adapted from van Driem 2005)

These languages have been lumped together in Zhengzhang (1995) and Pan (1995). Sagart (2005:163-5) discussed the related morphemes in kernel words. He also noted that numerals and pronouns involve many factors and whether forms in these domains are related is hard to judge. Therefore, the caclulation of kernel morphemes should be used with caution. Now let's make a rank analysis on the following materials taken from his latest result of Sino-Austronesian comparison (Sagart 2005).

Lexical items	Chinese	PAN	Old Chinese	Rank
water	水	daNum	濤 [b]t-ɦimʔ	1
egg	蛋	qiCeluR	卵 [a]Cə-lo[r]ʔ	1
bone	骨頭	kukut	骨 [a]kut	1
this	這	di	時 [b]di Iʔ	1
horn, angle	角	(q)uRuŋ	角 [a]k-rok	1
year	歲	kawaS	歲 [b]s-hwat-s S![4]	2
salt	鹽	siRaH1	鹵 [a]raʔ S!	2
head	頭	quluH1	首 [b]hluʔ	1
lie down; sleep	睡	_zem	寢 [b]tshimʔ	1
breast, milk	乳	nunuH1	乳 [b]noʔ	1
hot	熱	qa(i)ŋet	熱 [b]ŋet	1
say	說	kawaS	話 [a]m-kw-rat-s	1

			日 [b]wat	
come	來	duwa	於 [b]wa	1
snake	蛇	bulay	蛇 [b]m-la[r]	2
mother	母親	ina(_q)	女 [b]nraʔ	2
dig; hollow out	挖	_kut	掘 [b]m-kut	2
to flow	流	qaluR	水 [b]hlu[r]ʔ	2
thick	厚	_tul	敦 [a]tu[r,n]	2
far	遠	ma-dawiN	遠 [b]wa[r,n]ʔ V!	2
old; aged	老	_daŋ	丈 [b]draŋʔ	2
to hunt	打獵	qaNup	獵 [b]Cə-lap	2
to wash; rinse	洗	basuq	溲 [b]s(r)uʔ	2
sharp	尖	Cazem	[GSR 660a] [b]tsim	2
worm	蟲	[]ulej	蚓 [b]linʔ F?	2
to spit out; vomit	嘔吐	utaq	吐 [a]tha	2
broad, wide	寬	_baŋ	旁 [a]baŋ	2
think	想(思考)	nemnem	念 [a]nim-s	2

According to the above materials, the rank analysis would be like the follows:

	Related morphemes
High rank	11
Low rank	16

The ratio of related morphemes between Old Chinese and Austronesian in the high rank, 11%, is lower than that in the low rank, 16%. The distribution does not support the genetic relatedness of Chinese and Austronesian. Sino-Austronesian related words may be due to language contact.

Sagart (2005) listed related morphemes between Austronesian and Tibeto-Burman. Here is their distribution among ranks:

Lexical items	Chinese	PAN	Tibeto-Burman	Rank
month; moon	月亮	qiNaS	b.s-la	1
egg	蛋	qiCeluR	b. twiy<t-l-ʔ	1

this	這	di	tib. ndi'this'	1
horn, angle	角	(q)uRuŋ	b.ruŋ=rwaŋ	1
wind	風	bali	b. g-liy	2
salt	鹽	siRaH1	b. la I!	2
head	頭	quluH1	Lushai lu	1
lie down; sleep	睡	_zem	tib. gzim	1
breast, milk	乳	nunuH1	b. nuw	1
say	說	kawaS	tib. s-go	1
foot	腳	kakay	b. kriy	1
come	來	duwa	b. s-wa	1
snake	蛇	bulay	p-loloish lay1/2	2
mother	母親	ina(_q)	b. m-na	2
dig; hollow out	挖	_kut	Kachin kot	2
to flow	流	qaluR	b. twiy<t-l-,lwiy	2
thick	厚	_tul	PS tu:r	2
far	遠	ma-dawiN	b.wiy	2
to hunt	打獵	qaNup	Chepang krup	2
to wash; rinse	洗	basuq	Lushiai shuk	2
to spit out; vomit	嘔吐	utaq	b.(m-)tuk V!	2
broad, wide	寬	_baŋ	boro go2 baŋ1	2
think	想(思考)	nemnem	tib. s-nym-pa	2

The result is summarized in the following table:

	Related morphemes
High rank	10
Low rank	13

The ratio of related morphemes between Tibeto-Burman and Austronesian in the high rank, 10%, is lower than that in the low rank, 13%. The distribution suggests that they are due to language contact between Tibeto-Burman and Austronesian.

In general, according to our rank analysis, the results from Sagart's latest comparison do not suggest the genetical relationship between Austronesian with either Chinese or Tibeto-Burman.

8. CONCLUSION

We would like to restate two crucial steps in our studies. The first is to search morphemes with complete sound correspondences and stratify the layers to define the earliest layer. The second is the rank analysis. Without the first step, the second step cannot get started. If only the first step is done, the factor causing the related words is still under clarification. Rank analysis supposes that there are related morphemes at the earliest time depth as objects. If the separation of languages has been too long to find enough examples for establishing the complete sound correspondences, the rank analysis should not be applied. In such case, some new methods to recognize genetic relationship are needed.

This study confirms the genetic relationship between Austronesian and Kam-Tai. However, its genetic relationship with Sino-Tibetan (Sinitic, Tibeto-Burman and Hmong-Mien) or Austroasiatic (including Mon-Khmer, etc.) is still not justified because the related morphemes between them based on complete sound correspondences are not yet established and the rank analysis therefore can not be applied.

NOTES

2. Reaping hook.
3. To lose something.
4. I! V! F! T! S! irregular Initial, Vowel, Final, Tone, Syllable type according to Sagart.

REFERENCES

BENEDICT, P. K. 1942, Tai, Kadai, and Indonesian: a new alignment in Southeastern Asia. *American Anthropolog*.44:576-601.

_____.1976, Sino-Tibetan: another look. *Journal of the American Oriental Society* 96(2):167-97.

_____.1991, Austric, an 'Extinct' proto-Language. In J. Davidson (eds.) *Austroasiatic languages, Essays in honour of H. L. Shorto*. School of Oriental and African Studies.

BLUST, Robert. 1970. Proto-Austronesian addenda. *Oceanic Linguistics* 9:104-165.

_____. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working Papers in Linguistics* 9.2:1-15.Honolulu, University of Hawaii.

_____. 1980. Austronesian etymologies (I). *Oceanic Linguistics* 19.1:1-181.

_____.1983-84. Austronesian etymologies (II). *Oceanic Linguistics* 22-23:29-149.

_____. 1988. Austronesian etymologies (III). *Oceanic Linguistics* 25:1-123.

_____. 1989. Austronesian etymologies (IV). *Oceanic Linguistics* 28:111-180.

_____.1995. An Austronesianist looks at Sino-Austronesian. In Wang (1995) *ACL*: 283-298.

_____. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Elizabeth Zeitou and Paul Jen-kuei Li (eds.) *Selected papers from the eighth international conference on Austronesian linguistics*: 31-94. Taipei: Academia Sinica.

CHEN, Baoya (陳保亞). 1994. *Lun yuyan jiechu yu yuyan lianmeng 論語言接觸與語言聯盟*. PhD Thesis: Peking University. Published by Yuwen Press, 1996.

_____. 1999a. Hantai guanxici youxu duiying guizebiao 漢台關係詞聲母有序對應規則表.*Yuyanxue luncong 語言學論叢*. 22. Beijing 北京: the Commercial Press 商務印書館.

_____.1999b. *20 shiji zhongguo yuyanxue fangfalun 20 世紀中國語言學方法論*. Ji'nan 濟南：Shandong jiaoyu chubanshe 山東教育出版社.

Chen, Baoya and He, Fang (陳保亞、何方). 2004. Hantai hexin yizhi duiying yusu youjie fenxi 漢台核心一致對應語素有階分析. In Ting, Pang-hsin (丁邦新) and Sun, Hongkai (孫宏開) (eds.): *Hanzangyu tongyuanci yanjiu (III)漢藏語同源詞研究（三）*. Nanning 南寧: Guangxi minzu chubanshe 廣西民族出版社.

DEMPWOLFF, Otto. 1934, 1937, 1938. Vergleichende Lautlehre des austronesischen Wortschatzes.3 vol. *Beiheft zur Zeitschrift für Eingeborenen-Sprachen* 15, 17, 19.

VAN DRIEM, George. 2005. Sino-Austronesian vs. Sino-Caucasian, Sino-Bodic vs. Sino-Tibetan, and Tibeto-Burman as default theory. In Yogendra Prasada Yadava, Govinda Bhattarai, Ram Raj Lohani, Balaram Prasain and Krishna Parajuli (eds.) *Contemporary Issues in Nepalese Linguistics*: 285-338. Kathmandu: Linguistic Society of Nepal.

DYEN, Isidore. 1963. The position of the Malayopolynesian languages of Formosa. *Asian Perspectives* 7.1, 2: 261-271.

_____.1965. Formosan evidence for some new Proto-Austronesian phonemes. *Lingua* 14:285-305.

_____. 1971. The Austronesian languages of Formosa. In Thomas A. Sebeok (eds.) *Current Trends in Linguistics* 8:168-199. The Hague, Mouton.

FERRELL, Raleigh. 1969. *Taiwan aboriginal groups: problems in cultural and linguistic classification*. Institute of ethnology, Academia Sinica, Monograph No.17. Taipei: Academia Sinica.

HARVEY, Mark. 1982. Subgroups in Austronesian. In Halim, Carrington and Wurm (eds.) *Papers from the third international conference on Austronesian linguistics*. Pacific linguistics C-75. Canberra: Department of linguistics, Research School of Pacific and Asian studies, The Australian National University.

HO, Dah-an (何大安). 1995. Lun paiwanqun yuyan de fenqun 論排灣群語言的分群. *Taiwan yanjiu tongxun 臺灣研究通訊* 5-6:19-34.

_____. 1999. Lun yuanshi nandaoyu tongyuanci 論原始南島語同源詞. In Shi, Feng(石鋒) and Pan, Wuyun (潘悟雲) (eds.) *Zhongguo yuyanxue de xintuozhan – qingzhuo Wang Shiyuan jiaoshou 65 huadan 中國語言學的新拓展──慶祝王士元教授六十五華誕*. Hong Kong: Press of City University of Hong Kong:75-84.

LI, Fang-kuei. 1937. Languages and dialects. *The Chinese Year Book*, Shanghai. Also appear in *Journal of Chinese Linguistics* 1973.1

_____. 1977. *A Handbook of Comparative Tai*. The University Press of Hawaii

LI, Paul Jen-kuei (李壬癸) 1985. The position of Atayal in the Austronesian family. In Andrew Pawley and Lois Carrington (eds.) *Austronesian linguistics at the 15th Pacific science congress*: 257-280. Pacific Linguistics C-88. Canberra: Department of linguistics, Research school of Pacific and Asian studies, the Austronesian national university.

　　　　.1995a. Formosan vs. non-Formosan features in some Austronesian languages in Taiwan. In Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho, and Chiu-yu Tseng (eds.) *Austronesian Studies Relating to Taiwan*: 651-682. Symposium Series of the Institute of History and Philology Academia Sinica No.3. Taipei, Academia Sinica.

　　　　.1995b. Is Chinese genetically related to Austronesian? In Wang (1995) *ACL*: 93-112.

MENG, Simu (蒙斯牧). 1990. Yinniyu he dongtaiyu de guanxici 印尼語和侗泰語的關係詞. *Minzu yuwen 民族語文*. 6.

NI, Dabai (倪大白). 1988. Zhongguo de zhuangdongyu yu nandaoyu 中國的壯侗語與南島語. *Zhongyang minzu xueyuan xuebao 中央民族學院學報*. 3.

OSTAPIRAT, Weera. 2005. Kra-dai and Austronesian: notes on phonological correspondences and vocabulary distribution. In Sagart et al (2005) *PEA*: 107-131.

PAN, Wuyun (潘悟雲). 1995. Dui huaao yuxi jiaoshuo de ruogan zhichi cailiao 對華澳語系假說的若干支持材料. In Wang (1995) *ACL*: 113-144.

PULLEYBLANK, E. G.. 1995. Comments from Pulleyblank. In Wang (1995) *ACL*: 325-335.

REED, Lawrence A. 1994. Morphological evidence for Austric. *Oceanic linguistics* 33.2:323-44.

　　　　. 2005. The Current Status of Austric: a review and evaluation of the lexical and morphosyntactic evidence. In Sagart et al (2005) *PEA*: 161-176.

SAGART, Laurent. 1990. Chinese and Austronesian are genetically related. 23rd international conference on Sino-Tibetan languages and linguistics, University of Texas at Atlington. Oct.3-7.

　　　　. 1993. Chinese and Austronesian: evidence for a genetic relationship. *Journal of Chinese linguistics* 21.1:1-62.

　　　　.1995. Some remarks on the ancestry of Chinese. In Wang (1995) *ACL*: 195-223.

　　　　. 2005. Sino-Tibetan-Austronesian: an updated and improved argument. In Sagart et al (2005) *PEA*: 161-176.

SAGART, Laurent, Roger BLENCH and Alicia SANCHEZ-MAZAS (eds.) 2005. *The Peopling of East Asia: Putting together archaeology, linguistics and genetics*(=PEA). London: RoutledgeCurzon.

SCHMIDT, W. 1906. Die Monkhmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. *Archiv der Anthropologie*, n.s. 5: 59-109.

STAROSTA, Stanley. 1995. A grammatical subgrouping of Formosan languages. In Paul Jen-kuei Li，Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho, and Chiu-yu Tseng (eds.) *Austronesian Studies Relating to Taiwan*: 683-726. Symposium Series of the Institute of History and Philology Academia Sinica No.3.Taipei: Academia Sinica.

_____. 2005. Proto-East-Asian and the origin and dispersal of languages of East and Southeast Asia and the Pacific. In Sagart et al (2005) *PEA*: 161-176.

STAROSTIN, Sergai. (1995). Comments from Starostin. In Wang (1995) *ACL*: 393-404.

SWADESH. M. 1952. Lexico-statistic dating of prehistoric ethnic contacts, with special reference to North American Indians and Eskimos, *Proceedings of the American philosophical society* 96.4.

_____.1955. Time depths of American linguistic groupings. *American Anthropologist* 56.

TING, Pang-hsin (丁邦新). 2000. Hanzang yuyan yanjiufa de jiantao 漢藏語言研究法的檢討. *Zhongguo yuwen 中國語文*. 6.

TSUCHIDA, Shigeru. 1976. *Reconstruction of Proto-Tsouic Phonology*. Study of Languages & Cultures of Asia & Africa Monograph Series No.5. Tokyo, Gaikokugo Daigaku.

WANG, Feng (汪鋒). 2006. *Comparison of languages in contact: the distillation method and the case of Bai*. Language and Linguistics Monograph Series B: Frontiers in Linguistics III. Taipei: Institute of Linguistics, Academia Sinica.

WANG, William S-Y(王士元). 1969. Competing changes as a cause of residue. *Language* 45:9-25.

_____. (eds.) 1995. *The ancestry of the Chinese Language* (=ACL). Journal of Chinese linguistics. Monograph 8. Chinese translation by Li Baojia (李葆嘉) et al (2005). Hanyu de zuxian 漢語的祖先. Beijing 北京: Zhonghua Shuju 中華書局.

ZHENGZHANG, Shangfang (鄭張尚芳). 1995. Hanyu yu qinshuyu tongyuan genci ji fuzhui chengfen bijiao de zedui wenti 漢語與親屬語同源根詞及附綴成分比較的擇對問題. In Wang (1995) *ACL*: 269-282.

支持侗台語和南島語親緣關係的更多證據
陳保亞 汪鋒
北京大學漢語語言學中心
北京大學中文系

提要

本文為南島語和侗台語的親緣關係提供了更多證據。在仔細檢查前人的相關研究之後，我們認為目前情況下，最好直接比較現代語言，儘管一些原始語言的重構也用來作為本文的佐證。在比較了作為侗台語代表的德宏傣語和作為南島語代表的印尼語後，二者之間系統的對應建立起來了。在此基礎上的詞階分析顯示，有更多的關係語素落在高階中，也就是支持二者之間的親緣關係。根據最新的材料和詞階分析也進一步確認了侗台諸語言之間和南島諸語言之間的親緣關係。遵照親緣關係的可傳遞性，侗台語應該與南島語同源。我們還進一步比較了原始台語和印尼語，二者之間的關係語素也同樣支持同源關係。應用同樣的比較步驟來分析南島語同漢語或者同藏緬語的關係，關係語素都顯示在高階少而低階多，也就是說，南島語同漢語或者同藏緬語的親緣關係還不能確認。

關鍵詞
親緣關係，侗台語，南島語，詞階分析