



# Why does the Tai-Kadai word for 'ten' end in -t ?

Laurent Sagart

## ► To cite this version:

Laurent Sagart. Why does the Tai-Kadai word for 'ten' end in -t ?. 23èmes Journées de Linguistique Asie Orientale, Jul 2010, Paris, France. halshs-03941330

**HAL Id: halshs-03941330**

**<https://shs.hal.science/halshs-03941330>**

Submitted on 16 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Why does the Tai-Kadai word for 'ten' end in -t ?<sup>1</sup>

Laurent Sagart

CRLAO

(Ecole des Hautes Etudes en Sciences Sociales and CNRS),  
Paris, France

## I. The problem.

The Tai-Kadai (TK; also Kra-Dai ; 侗台; 侗泰) languages are spoken over a large area extending from the region of Guangdong and Guangxi in S. China to SE Asia and even NE India. The following map shows five components:

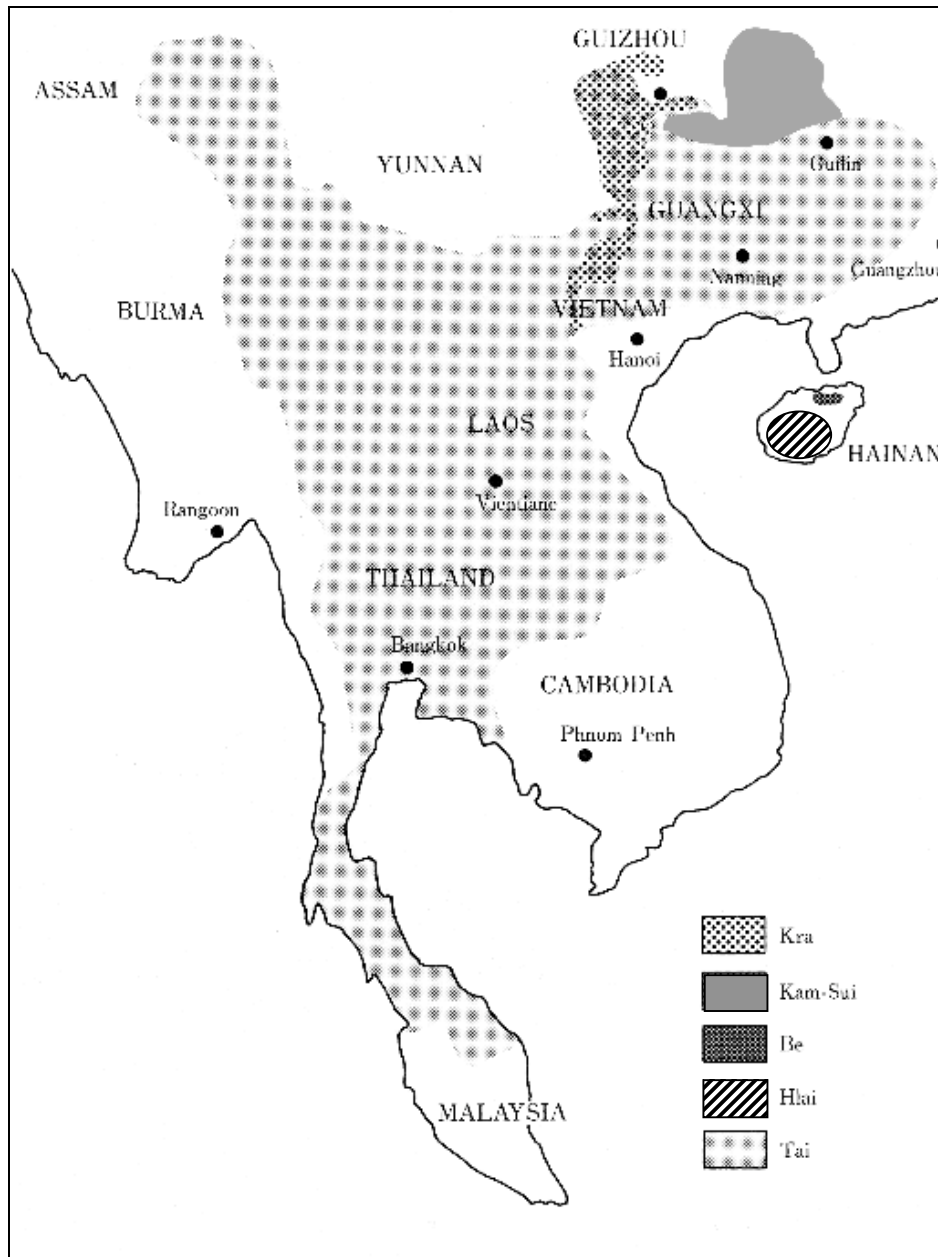


Figure 1: the Tai-Kadai languages (after Ostapirat 2005a, slightly edited)

<sup>1</sup> I thank Lawrence A. Reid, John U. Wolff and R. D. Zorc for providing valuable information on Philippine languages.

Three branches are often recognized: I. Tai: Tai+Kam-Sui+Be; II. Hlai; III. Kra (also called Kadai, Geyang etc.). From the point of view of numerals, there are two kinds of Tai-Kadai languages: those that have numerals related to Chinese, like those of the Tai and Kam-Sui subgroups (exemplified by Buyi and Dong in Table 1) and those that have numerals related to Austronesian, like those of the Hlai and Kra subgroups (exemplified by Hlai and Buyang in Table 2).

	OC	MC	Buyi	Dong
1	一 *ŋi[t]	'jit	it 7	ət 7
2	二 *ni[j]-s	nyijH	n'i 8	n'i 6
3	三 *[s.r][ə]m	sam	sa:m 1	sa:m 1'
4	四 *s.li[j]-s	sijH	si 5	si 5'
5	五 *ɕ.ŋ'aʔ	nguX	ɣa 3	ŋo 4
6	六 *k.ruk	ljuwk	zoʔ 7	ljok 8
7	七 *s-ŋi[t]	tshit	tsat	sət 7'
8	八 *p <sup>s</sup> ret	peat	piat 7	pet 9
9	九 *(tə.)[k](w)uʔ	kjuwX	ku3	t'u 3
10	十 *[g][i]p or *t.ɕ[ə]p	dzyip	tsip 8	ɕəp 8

Table 1: TK languages with numerals related to Chinese. Old Chinese numerals in the Baxter-Sagart system.

In these languages, the phonetic shape of the numerals is closer to Middle Chinese (mid-first millennium CE) than to Old Chinese (early to mid- first millennium BCE): OC consonant clusters have been reduced in '3', '4', '7' and '8'; preinitial elements have been lost in '9' and '10'; suffixes have disappeared in '2' and '4'; tones have appeared. The velar initial in '10' has been replaced by a palatal affricate. The TK and MC tones have been related by systematic sound correspondences (Wulff 1934). It is particularly noteworthy that these TK language share with Middle Chinese (and with most living Chinese dialects) the irregular change of the vowel in '3'. This was a central vowel in Old Chinese: it changed to Middle Chinese /a/ in an irregular (and mostly unexplained) change. Long a: in Dong and Buyi corresponds to MC /a/, not to OC schwa. Unless Buyi and Dong languages are Chinese dialects (they certainly are not) it would appear that these numerals were borrowed from Chinese after OC. Borrowing also explains the lack of aspiration in '7'. Proto-Kam-Tai lacked aspirated stops and used its own unaspirated voiceless stops for Chinese aspirated stop initials.

	PAn	PMP	Buyang	Hlai Baoding	Hlai Tongshi	P-Hlai
1	isa	isa	am 24	tsuu 2, tsheu 3	uu 3	ci C
2	duSa	duha	ɛa 54	ɬau 3	ɬau 3	alau C
3	telu	telu	tu 54	fu 3	tshu 3	utu C
4	Sepat	e(m)pat	pa 54	tshau 3	tsho 3	atəu C
5	RaCep	lima	ma 312	pa 4	pa 4	ama A
6	?	enem	nam 54	tom 4	tom 1	(ə)num A
7	RaCep-i-tuSa *	pitu	tu 312	thou 1	thou 1	?tu A
8	RaCep-a-telu *	walu	ma 0 ðu 312	gou 1	gou 1	aRu A
9	RaCep-i-Sepat *	siwa	va 11	fau 3	fau 3	upal C
10	iCit	puluq	put 54	fu:t	fu:t	apu:c

Table 2: Tai-Kadai languages with numerals related to Austronesian. Sources: Li Jinfang 1999 (Buyang); Ouyang Jueya and Zheng Yiqing 1980 (Li); Ostapirat 2004 (proto-Hlai). The reconstruction of PAn numerals follows Sagart (2004). PAn forms followed by an asterisk are additive expressions based on '5': they evolved into the PMP forms in the columns next to them. The PAn words for 'eight' and 'nine' had multiplicative and subtractive competitors (for details see Sagart 2004).

Because the notion of 'one' is a frequent source of grammaticalizations, numerals for 'one' tend to be displaced relatively often in language-internal developments. This appears to have taken place in Buyang and, less certainly, in Hlai. The correspondence between Buyang and Austronesian numerals above 'one' is more straightforward (Sagart 2004). In general, Buyang numerals match the second syllable of the corresponding Austronesian words, except when the initial consonant was a stop and the medial was -l-: in such cases ('3' and '10') a cluster is formed which Buyang reduces to the initial stop. In '8', the initial syllable of PMP appears to be reflected by Buyang ma 0.

In Sagart (2004, 2005), based in part on the correspondence between numerals, I have argued that the Tai-Kadai languages form a sister branch to Malayo-Polynesian within Austronesian. The AN vocabulary in TK is very basic (as shown by Benedict 1942). Much of the non-Austronesian vocabulary in TK is shared with Austroasiatic (Maspero, as cited and expanded in Haudricourt 1954). The most likely scenario at this point is that after the first Austronesian movement out of Taiwan, an Austronesian ancestor of the TK group left the Philippines probably in the second millennium BCE to settle the coast of Guangxi or north Vietnam; and that once on the mainland that language was partly relexified by a local Austroasiatic language; after which it evolved into the most common recent ancestor of all TK languages alive today. This story is different from the scenario in Benedict (1942) where the TK languages were treated as coordinate with Austronesian as a whole (including the languages of Taiwan) within an 'Austro-Tai' macro-family. Benedict's view is still on the whole adhered to by Ostapirat 2005a, though in some respects Ostapirat 2005a oscillates between Benedict's view and mine.

The sound correspondences between Austronesian are not completely known, but Benedict's pioneering work (1942) has recently been furthered in Ostapirat (2005a). One interesting

difference between them relates to Austronesian final -q, a voiceless uvular stop (thought to have been a fricative -h at the time when Benedict wrote). Benedict thought the corresponding TK sound was -t, based on the Austronesian word for '10' \*puluq which has -t in TK (Table 2). Benedict looked for similar examples to bolster his case. He came up with a comparison between the Austronesian word \*daRaQ 'blood' and Li (=Hlai) *dat*, *tlat* 'blood':

"Li *p'uo*t~*fu*o't '10' attest to a pair of shifts, viz. final -h > -t., as in \**darah* 'blood', Li *dat*~*tlat* 'blood', and -u- > -uo- (...). The development here has been of the type *puluh* > *p'luh* > *p'lut* > *p'ut* > *p'uo*t." (Benedict 1942:583).

However Hlai *dat*, *tlat*, or Tai *lüät* 'blood' do not correspond phonologically to Austronesian \*daRaQ 'blood'. By Benedict's own principles Austronesian voiced stops should correspond to implosive stop initials in TK; and Austronesian medial -R- normally falls in TK. Benedict's comparison for 'blood' is in error and the comparison for '10' is alone in showing the correspondence Austronesian -q : TK -t. Progress was accomplished when Ostapirat (2005a:119) showed that Austronesian \*-q corresponds to TK -k in two good examples:

	PAn	Tai	Hlai (Baoding)
Taro	biRaQ	<i>phuak</i>	<i>ge:k</i>
Otter	Sanaq	<i>naak</i>	<i>te:k</i>

In another paper, Ostapirat (2005b:278) cited a TK form for 'vomit' as an example of PTK \*t-:

Be	Siamese	Kam
<i>duak</i> (Low)	<i>raak</i> (Low)	<i>taak</i> (High)

He argues that Be *d*- and Siamese *r*- have undergone intervocalic voicing, but original \*t- is preserved by Kam.

If Ostapirat is right about the initial, this word probably reflects PMP \*u(n)taq 'to vomit', providing us with a third example of the correspondence PAN/PMP -q : TK -k.

Ostapirat (2005) takes the -t ending in the TK word for '10' to be another regular correspondent of Austronesian -q. He supposes the conditioning factor is the preceding vowel. If it is /u/ then one has TK -t; otherwise, TK -k. This is not an illogical interpretation given the small number of available examples. However it would be phonetically strange for a -k to be fronted to a -t following a *back* vowel. I would like to propose another interpretation: the only regular correspondent of Austronesian -q is TK -k; the correspondence of -q to -t is unique to 'ten'. It is due to a special development in the history of the MP word for '10', not to phonological conditioning.

## II. Austronesian numerals and subgrouping.

It is generally agreed that the Austronesian homeland was in Taiwan, c. 4000-3000 BCE. The languages of Taiwan represent the first diversification of the family, at a time when Austronesian was confined to Taiwan. As to the 1200 or so Austronesian languages outside of Taiwan, it is generally agreed that they share a common ancestor, generally called proto-Malayo-Polynesian (PMP). The position of PMP vis-a-vis the languages of Taiwan is a matter of controversy: Blust



(1999) thinks that PAn split into ten branches, of which 9 are spoken only in Taiwan and one — PMP — only outside of Taiwan. But a number of scholars think that there are less than nine primary branches on Taiwan and that PMP is not a primary branch, but part of a SE coast subgroup which also includes Amis. My own phylogeny (Sagart 2004) is of the latter type. It is based on innovations in the basic vocabulary, especially the numerals from 5 to 10.

Of the PMP numerals : PMP \*isa 'one', \*duha 'two', \*telu 'three', \*e(m)pat 'four', \*lima 'five', \*enem 'six', \*pitu 'seven', \*walu 'eight', \*siwa 'nine' and \*puluq 'ten', only those for '1' to '4' were part of PAn. The numeral for '5' was \*RaCep; \*lima existed but basically meant 'hand' although it may have been used as '5' in competition with \*RaCep. The numeral for 'ten' was \*iCit. There were no fixed numerals between 'five' and 'ten'. 'Six' was either 'two times three', or 'five plus one'; 'seven' was 'five plus two'; 'eight' was either 'five plus three' or 'two times four'; 'nine' was either 'five plus four' or 'ten minus one'. Some of these analytic expressions are preserved in NW Taiwan languages like Pazeh:

six	xasebuza < *RaCep-esa ('five-one')
seven	xasebidusa < *RaCep-i-duSa ('five-i-two')
eight	xasebaturu < *RaCep-a-telu ('five-a-three')
nine	xasebisupat < *RaCep-i-Sepat (('five-i-four'))

Pazeh represents an early stage in the development of AN, in which the PMP numerals for 6 to 9 did not exist. They became established one after the other between PAn and PMP. The origin of the PMP word for 'six': \*enem, is unknown. The word for 'seven' \*pitu is a reduction from \*RaCep-i-duSa ('five-i-two'), the word for 'eight', \*walu is a reduction of \*RaCep-a-telu ('five-a-three'); the word for nine a reduction of \*RaCep-i-Sepat (('five-i-four')). The word for 'ten': \*puluq is probably related to \*puluŋ 'all' (the fingers). The order in which these innovations became fixed can be recovered by reference to the languages that reflect them. This gives the following tree:

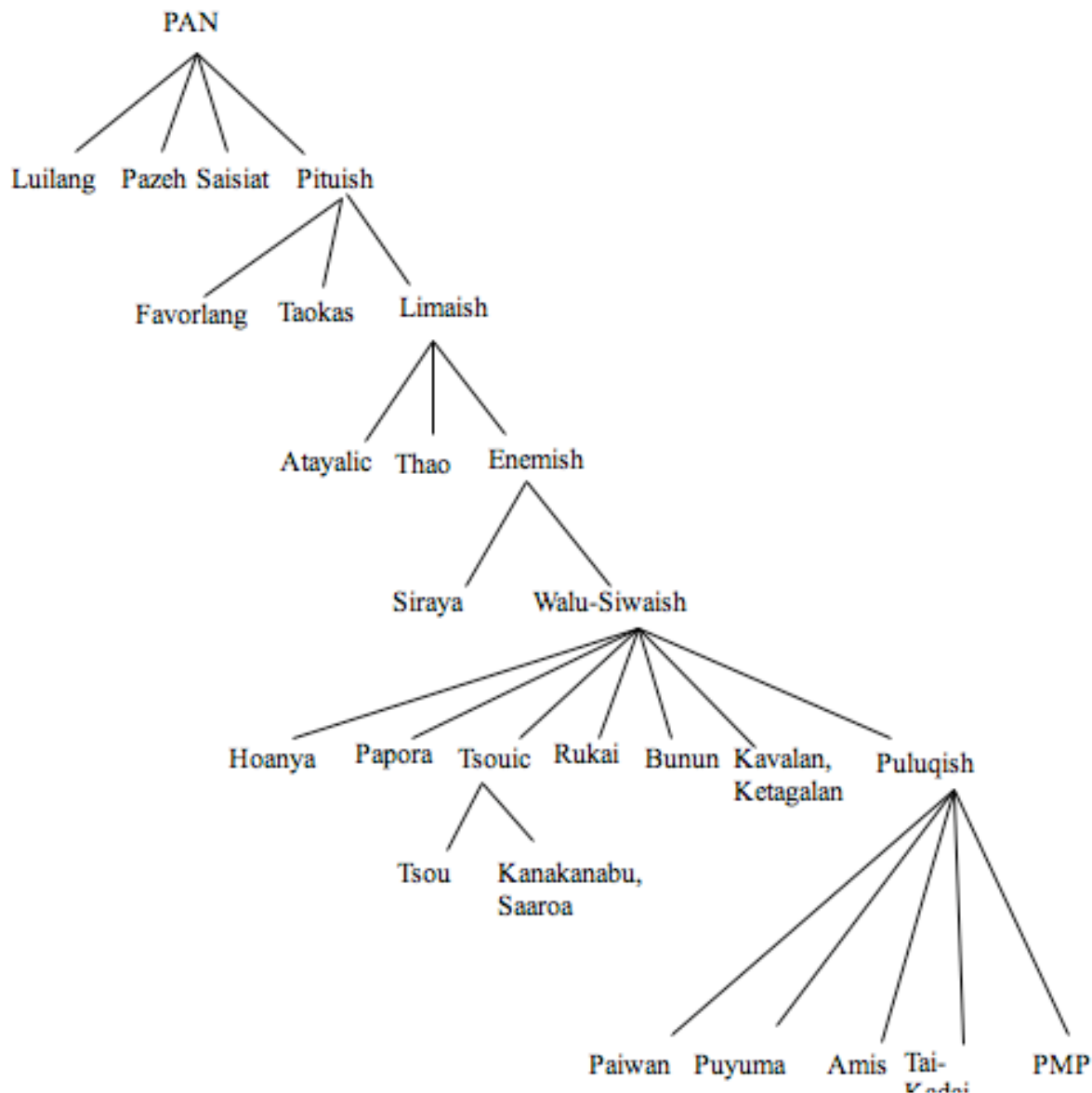


Figure 2. Higher Austronesian phylogeny after Sagart 2010.

As can be seen from this tree, \*puluq was innovated shortly before PMP. It is reflected only in three languages of SE Taiwan: Paiwan, Puyuma and Amis; outside of Taiwan, in PMP and in those TK languages that did not replace their numerals with Chinese loanwords. But as we have seen, there is this curious irregularity in Tai-Kadai, where the word ends in -t instead of expected -k.

The first consequence of this is that TK is not a sister language of PAN, as Benedict thought: because it shares some post-PAN innovations with PMP, it must be a sister language of PMP.

I now come to the question of why the word for 'ten' ends in -t in Tai-Kadai.

### III. A proposal to explain why the TK word for '10' ends in -t.

I have argued (Sagart 2004, 2005) that PTK and PMP form a subgroup. Since that subgroup has no name yet, I will call it 'Extra-Formosan' (EF), a term earlier introduced by Reid to designate all AN languages outside of Taiwan. In Reid's usage, the term 'Extra-Formosan' was synonymous with 'Malayo-Polynesian' in Blust's usage. In the present usage, it is more encompassing. I take the

phylogeny in Figure 3 to be broadly correct outside of Formosan, once a node is added that dominates Malayo-Polynesian and a Tai-Kadai branch is added under that node.

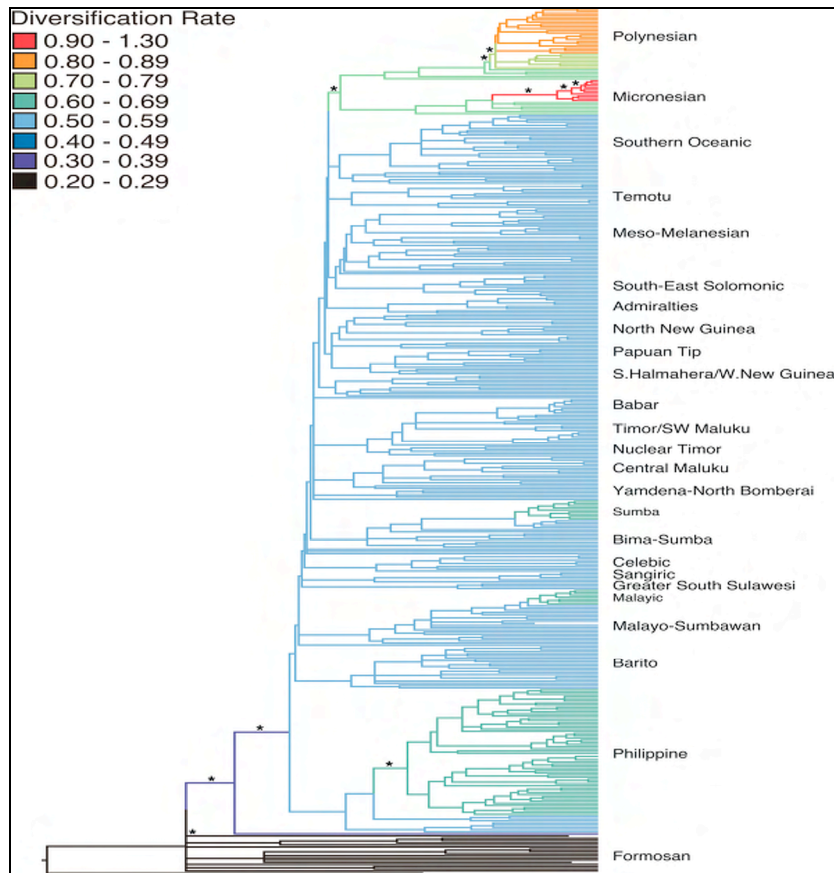


Figure 3: Austronesian phylogeny after Gray, Drummond and Greenhill 2009.

In Kalinga, a language of the northern Philippines (Tryon et al. 1995, Vol. 4, 43-45), the word for 'ten' is *simpulu* (where *sim-* includes \*sa 'one' and a nasal linker, and *pulu* is the regular reflex of PMP \*puluq 'ten', with regular loss of \*-q); but the words for 'eleven' and 'twelve' are *simpu:lut osa* (analysed as *sin-pulu=t osa*, where *osa* < \*esa 'one' and the clitic =*t* is 'and') and *simpu:lut duwa* (analysed as *sin-pulu=t duwa*, where *duwa* < \*duSa 'two' and the same clitic =*t* as in 'eleven').

In Tagalog, '10' is *sampu*, a contraction of \*sa- 'one', a nasal linker and \*puluq 'ten'. Numbers above twenty include the morpheme '10' followed by another numeral,<sup>2</sup> as in:

*dalawampú't dalawá* twenty-two (< \*daduha 'two' + nasal linker + \*puluq 'ten' + \*at 'and' + \*daduha 'two')

*dalawampú't isá* twenty-one (< \*daduha 'two' + nasal linker + \*puluq 'ten' + \*at 'and' + \*esa 'one')

The -*t* ending in -*pú't* is generally believed to be the result of incorporating the conjunction *at* 'and'. This morpheme appears to be a borrowing from the Kapampangan conjunction *at* (L. Reid, R.D. Zorc, p.c. April 2007), where *a* < PAN \*e. Thus Kalinga *ot* and Kapampangan *at* reflect an earlier \*et 'and, and then'. Zorc reconstructs proto-Northern Philippine \*et on the basis of forms from Kapampangan, Kalinga, Inibaloi, Kankanay, Sambal, Bolinao and Botolan Sambal. The meanings vary from 'and' to 'yet' and 'still'.

<sup>2</sup> This does not happen in numbers between 10 and 20 because Tagalog uses another strategy to form those numerals.



These facts provide the explanation for the unexpected -t in the Tai-Kadai word for 'ten'. I propose that the innovation of \*et 'and, yet, still' occurred at that (EF) node, and that in EF \*et was used (perhaps in variation with other conjunctions) between \*puluq 'ten' and a following numeral.<sup>3</sup> In the TK branch, the sequence \*puluq-et was reduced to \*pulu't, according to the clitic incorporation pattern we have seen in Kalinga and Tagalog<sup>4</sup> (but independently from them). At a later stage the form ending in -t became generalized in all contexts, displacing whatever form existed previously (\*puluq or \*puluk, depending on chronology). The replacement was completed in PTK. On the MP side the conjunction is preserved in northern Philippines, occasioning the developments we have described in Kalinga and Tagalog. Outside of the Philippines, it became lost in the common ancestor of non-Philippine MP languages. Thus a single innovation and a single extinction can explain why reflexes of EF forms \*et and \*puluq-et are found only in Tai-Kadai and in the Philippines.

This scenario allows for a phonetically less curious account of the PTK - PMP sound correspondences; it ascribes apparently independent and irregular developments in TK and Philippine languages to a single nonphonological cause; third, the innovations which underlie the appearance of a -t at the end of reflexes of \*puluq are compatible with the tree in Figure 2.

## References

- Benedict, P.K. (1942) Thai, Kadai and Indonesian: a new alignment in Southeastern Asia. *American Anthropologist*, n.s., 44, 576-601.
- Blust, R. (1999) Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In: Elizabeth Zeitoun and Paul Jen-kuei Li (eds.) *Selected Papers from the Eighth International Conference on Austronesian linguistics*, 31-94. Taipei: Institute of Linguistics (preparatory office).
- Haudricourt, A.-G. (1954) De l'origine des tons du vietnamien. *Journal Asiatique* 1954, 242: 69-82.
- Li Jinfang (1999) *Buyang Yu Yanjiu*. Beijing: Zhongyang Minzu Daxue.
- Ostapirat, W. (2004) Proto-Hlai sound system and lexicons. in: Ying-ching Lin, Fang-min Hsu, Chun-chih Lee, Jackson T.S. Sun, Hsiu-fang Yang and Dah-an Ho (eds), *Studies in Sino-Tibetan languages, papers in honor of Professor Hwang-cherng Gong on his seventieth birthday*, 121-175. Nankang: Institute of Linguistics, Academia Sinica.
- Ostapirat, W. (2005a) Tai-Kadai and Austronesian: notes on phonological correspondences and vocabulary distribution. In L. Sagart, R. Blench and A. Sanchez-Mazas (eds) *The peopling of East Asia: Putting together Archaeology, Linguistics and Genetics* 109-133. London: RoutledgeCurzon
- Ostapirat, W. (2005b) Be's obstruent consonants in Kradai perspective, in Ho Dah-an and Ovid Tzeng (eds) *POLA forever*, 275-291. Nankang: Academia Sinica.
- Ouyang Jueya and Zheng Yiqing (1980) *Li Yu Jian zhi*. Peking, Chinese Academy of Social Sciences.
- Sagart, L. (2004) The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43,2: 411-444.
- Sagart, L. (2005) Tai-Kadai as a subgroup of Austronesian. In L. Sagart, R. Blench and A. Sanchez-Mazas (eds) *The peopling of East Asia: Putting together Archaeology, Linguistics and Genetics* 177-181. London: RoutledgeCurzon.
- Sagart, L. (2010) Is Puyuma a primary branch of Austronesian ? *Oceanic Linguistics* 49, 1 (June 2010)
- Tryon, D. T. (1995) (ed.). *Comparative Austronesian Dictionary. An Introduction to Austronesian Studies*. Trends in Linguistics, Documentation 10. Mouton de Gruyter, Berlin and New-York, 1995. 5 volumes.
- Wulff, K. (1934) *Chinesisch und Tai*. Sprachvergleichende Untersuchungen. Copenhagen.

<sup>3</sup> Other innovations can be assigned to the EF node: the shift of PAn \*Su '2nd person pronoun singular' to 2nd-person plural; the rise of a new form \*bujak for 'flower'.

<sup>4</sup> A difference is that while in the Philippine languages =(ə)t was added to a stem which had already lost final -q as a result of sound change, in the TK branch the EF form \*puluq-et probably evolved via final cluster simplification: \*puluq-t > pulut or \*puluk-t > pulut.