

# *Squib*

## **A Note on the Higher Phylogeny of Austronesian**

Bodo Winter

UNIVERSITY OF HAWAII AT MĀNOA

This paper presents a critique of Sagart's (2004) classification of the Formosan languages. Sagart proposes a subgrouping based on a set of innovations in the numeral systems of the Formosan languages. These innovations entail that the higher phylogeny of Austronesian is much more hierarchical than in other subgrouping accounts (e.g., Blust 1999, Ross 2009). According to Sagart, the innovated numeral forms can be derived from the complex numerals of the Formosan language Pazeh. A number of arguments are presented that call this subgrouping account into question. There are reasons to assume that the proposed numeral derivations are highly unlikely, and that similarities between the complex numerals of Pazeh and the innovated forms in other Formosan languages are due to chance. Furthermore, language contact might account for the pattern of numeral innovation. Finally, the migration pattern that is entailed by Sagart's numeral innovations does not take into account possible back and forth migrations of the Formosan tribes.

**1. INTRODUCTION.<sup>1</sup>** Sagart (2004) developed a theory of the higher phylogeny of Austronesian (AN) based on an analysis of the numeral systems of the Formosan languages. It is the most innovative and unconventional approach to AN subgrouping proposed to date. However, this squib advances several arguments that collectively provide grounds for questioning his subgrouping and the patterns of migration that it implies.

Briefly, Sagart has suggested that the numerals \*lima '5', \*enem '6', \*pitu '7', \*walu '8', \*Siwa '9', and \*puluq '10', which are generally assigned to Proto-Austronesian (PAN), are actually post-PAN innovations. According to Sagart, these numerals are innovations that are nested across the Formosan languages in a hierarchical manner: First, \*pitu was innovated, then in a language that already had \*pitu, \*lima was innovated, followed by \*enem, \*walu, \*Siwa, and finally \*puluq. Sagart bases a subgrouping hypothesis on these numerals: all those languages that share the form \*pitu for '7' are thought to derive from a putative ancestor called "Pituish." Those languages that share the form \*walu for '8' and \*Siwa for '9' are thought to derive from a putative ancestor called "Walu-Siwaish" (cf. Sagart 2004:421). All other Austronesian languages are thought to derive from this ancestor. As no forms of \*pitu or any of the other innovated numerals can be found in Pazeh, Luilang, and Saisiyat, these "pre-Pituish" languages are assigned to the highest branch of PAN. This sub-

1. In preparing this paper, I have benefited immensely from comments and suggestions made by Robert Blust, and by Malcolm Ross in his LSA 2009 class in Berkeley. Special thanks also go to Clive Winter for proofreading. I am responsible for any remaining errors.

grouping is much more nested than the subgroupings proposed by Blust (1999) and Ross (2009), which divide the Formosan languages into nine or ten different groups.

As support for this proposal, Sagart (2004:417–19) points to similarities between the numerals 7, 8, and 9, and the forms of the corresponding numerals in Pazehe. Based on known sound correspondences between PAN and Pazehe, he suggests that \*pitu, \*walu, and \*Siwa are reduced forms of complex numerals in PAN, as follows:

(1) PAN	PAZEHE	POST-PAN
*RaCep-i-tuSa	xasebidusa	*pitu ‘7’
*RaCep-a-telu	xasebaturu	*walu ‘8’
*RaCep-i-Sepat	xasebisupat	*Siwa ‘9’

Sagart notes that only six changes are needed to account for these reductions. The beauty of his account lies in the coherent spatial pattern of population movements that can be inferred from the distribution of numerals. The following three sections highlight problems with the proposed subgrouping.

**2. DERIVATION OF THE POST-PAN NUMERALS.** One major weakness of the proposal involves the way in which the allegedly post-PAN numerals are derived: to make the argument work, the suggested derivations assume several sound changes. With the possible exception of one change (pruning of the left pretonic syllable, see Sagart 2004:419), the changes do not have parallels in the pre-Pituih languages. Sagart justifies proposing these ad hoc changes by claiming that they are due to a drive to disyllabism. However, since the drive to disyllabism in Austronesian applies to roots or single content morphemes and not to morphologically complex forms (Blust 2007:1), an explanation based on the drive to disyllabism requires that a reanalysis from morphologically complex forms such as \*RaCep-i-tuSa to simplex forms such as \*RaCepituSa had taken place. This seems highly unlikely, given that speakers of a language that already had a strong preference for disyllabic roots—about 90 percent of PAN lexical roots were disyllabic (Blust 2007)—would surely not have been very much inclined to analyze a form that consists of two disyllables into a five-syllable form.

It also seems strange to assume that the reanalysis of quinary numerals as single unanalyzable wholes started with 7. It is hard to see any motivation for a reanalysis that affects the second in a series of four compound numerals, while leaving the others untouched until future generations of speakers finish the job (Robert Blust, pers. comm.). The numeral system that this order of innovation entails for Sagart’s Pituih (where 6, 8, and 9 are complex numerals but 7 is already a simplex form) appears to be unparalleled in the Austronesian languages (Blust 2009:273).

Furthermore, it is quite likely that the apparent similarity between the PAN numerals 7, 8, and 9 and the complex forms in Pazehe is due to chance. A look at the pre-Pituih languages Luilang, Saisiyat, and Pazehe (cf. Sagart 2004:417) reveals that there are 15 forms that can possibly serve as the basis for deriving the five innovated numerals whose etymologies need to be explained (see table 1, taken from Sagart 2004:414). Many of these 15 forms are complex numerals, and therefore multisyllabic. In total, one has to find only 10 syllables (for five disyllabic numerals) from a 55-syllable set (counting based on table 1, Sagart 2004:414). Within a set of 55 CV syllables, the probability is relatively high that

TABLE 1. NUMERAL FORMS OF THE PRE-PITUISH LANGUAGES

	'5'	'6'	'7'	'8'	'9'	'10'
LUILANG	(na)lup	(na)tsulup	innai	patulunai	patulunai	isit
SAISIYAT	rrasu	saivusa	saivuseaha	makaspat	ra:ha	ranpon
PAZEH	xasep	xasebuza	xasebidusa	xasebaturu, xasebituru	xasebisupat	isit

forms can be found that are similar to the six syllables of \*pitu, \*walu, and \*Siwa (7, 8 and 9). This is especially the case if the criteria for choosing similar forms are rather unconstrained, as when only the place of articulation and the vowel of a given syllable have to match (cf. Sagart 2004:417). Sagart's criteria for choosing similar forms do not even take morpheme boundaries (\*RaCep-i-tuSa > \*pitu) or the contiguity of syllables (\*RaCep-a-telu > \*walu) into account. In addition, if one considers that no similarities can be found between \*enem '6', \*puluq '10', and any of the complex forms of the pre-Pituish languages, the fact that PazeH apparently has forms similar to the numerals \*pitu, \*Siwa, and \*walu seems unremarkable. The similarities pointed out by Sagart are intriguing, but in the absence of any convincing proof that they cannot be due to chance, they should not be used as evidence for any kind of subgrouping.

Finally, Sagart stresses at several points that the derivation of \*pitu, \*walu, and \*Siwa "only" requires six changes. However, considering that these changes apply to just three forms, six changes is not a small number. If one tries to derive three short forms from three long forms, the number of changes is bound to be relatively low, simply because the number of forms that require changes is very small. As mentioned above, these changes are explained by Sagart (2004:418) as an instance of the general trend of the drive to disyllabism in the Austronesian languages. However, the many examples of disyllabism-maintaining or disyllabism-reestablishing changes in Blust (2007) are usually single changes rather than large sets of changes. If the history of the numerals as proposed by Sagart is really due to a drive to disyllabism, it looks strikingly different from the majority of other cases that are due to this preference.

**3. BORROWABILITY OF NUMERALS.** Sagart (2004:412) makes the claim that low numerals belong to a language's core vocabulary, and are therefore relatively more resistant to borrowing than other parts of the lexicon. This is crucial for his overall argument to work, because otherwise the proposed pattern of numeral innovation could be due to spreading via contact (cf. Sagart 2004:421). Sagart argues that low numerals (e.g., 5, 6, 7, and 8) are only likely to be borrowed when the donor language is spoken by a community that forms a state. However, the claim that statehood is necessary for low numerals to be affected by borrowing needs further substantiation before it can serve as a viable basis for Sagart's argument. It seems that "statehood" is too simplistic a criterion to account for the numerous complex sociolinguistic reasons (prestige, trade, exogamy, and so on) that might explain borrowing of low numerals.

In general, cases of numeral borrowing (including the borrowing of low numerals) are very easy to find: in addition to well-known cases like Korean and Japanese, where complete numeral systems have been borrowed from China (cf. Sohn 2001, Shibatani 1990), a Google search of just ten minutes revealed four more cases of low numeral borrowing:

Berber borrowed numerals from Arabic (Souag 2007), Nar-Phu (Tamangic) borrowed numerals from Tibetan (Noonan 2006), Chantyal (Tamangic) replaced its complete numeral system under the influence of Nepali (Noonan 2006), and the Burushaski language (isolate) of Jammu and Kashmir borrowed numerals from Urdu (Munshi 2006). In his study of numeral borrowings in Berber dialects, Souag (2007) points out that 2 and 3 are more often retained than higher numbers. This means that higher numerals—that is, those relevant to Sagart’s proposal—might not fall into the range of those that are particularly resistant to borrowing.

Since “core lexicon” or “basic vocabulary” is still a problematic notion (Haspelmath 2008, Thomason 2001:71–72), it is unclear whether the numerals 5–10 should belong to this category:<sup>2</sup> sound changes therefore might provide a more stable and borrowing-resistant indicator of subgrouping than numerals.<sup>3</sup> In Taiwan, borrowing of numerals seems especially likely because of the long-lasting contact between the different Formosan languages: for example, between Puyuma and Paiwan (Blust 1999:47–51), between Paiwan and Rukai (Tsuchida 1976:10–11), or between Thao and Bunun (Blust 1996). The cultural ties between many of the Formosan language communities are very close, thus providing an optimal basis for borrowing relationships. In this regard, it seems to be more parsimonious to assume that the quinary systems of Luilang, Saisiyat, and Pazeh originate from an innovation that has spread through these communities. This is especially likely since these three languages are situated very close to one other. It is also worth mentioning that quinary systems are likely to arise or spread via contact, as has been the case, for example, in Vanuatu and New Caledonia (Blust 2008:450–52; however, see also Lynch 2009).

#### 4. PROBLEMS WITH THE PROPOSED MOVEMENT PATTERN.

Based on the map employed by Tsuchida (1976:xxx), it appears to be the case that the community which spoke Pazeh was situated south of the Taokas community. However, Taokas has a form *yweto* ‘7’ that is derived from \*pitu, and it is therefore regarded as a Pituish language by Sagart (2004:421). If Taokas is a Pituish language, and Pazeh a pre-Pituish language, then—following the north–south counterclockwise movement pattern Sagart proposes—the Taokas community should have been north of the Pazeh community, and not south of it, as indicated on Tsuchida’s map. Thus there seems to be an inconsistency between the spatial pattern predicted by Sagart and the actual locations of Pazeh and Taokas.

Another problem with the proposed movement pattern and its apparent spatial coherence is the possibility of back and forth migrations. Sagart himself provides evidence for movements of the tribes of Taiwan (2004:436): Atayal and Seediq came from west-central Taiwan to their present-day northern location (Mabuchi 1954, Blust 1996:287); and Thao was previously spoken in the western lowlands, having moved about 300–350 years ago (Blust 1996, 2003). Also, the Amis on the east coast have spread southward in recent times (Malcolm

2. For example, Swadesh (1952) has the numerals ‘one’ through ‘five’ on his 200 wordlist of basic words, but only the numeral ‘two’ on his revised 100-word list (Swadesh 1955), which he considers even more resistant to change than the 200-word list.

3. Sagart (2004:412) mentions that there is “a long list of sound changes having spread across language boundaries” to justify prioritizing innovations in the numeral system over innovations in the sound systems of the Formosan languages. However, this remark misses the fact that borrowing often creates irregular sound correspondences that can be detected in subgrouping hypotheses based on the comparative method (Blust 1999).

Ross, pers. comm.) These examples show that movements of the tribes of Taiwan have happened more than once within the past 300 to 400 years. And since the Neolithic settlement of Taiwan is thought to have happened around 4,000 BC (Blust 1999:54), there must have been many more movements for which we cannot find linguistic evidence. From this perspective, the tight match between Sagart's hierarchy and the coherent spatial pattern that emerges from the hierarchy might actually be problematic: it is unlikely that the languages that matter for Sagart's proposal remained in place throughout the long period since the original Neolithic settlement of Taiwan.

**5. CONCLUSIONS.** It may well be possible to point to weaknesses in some of these arguments against Sagart's proposal. But even if some of the questions raised here turn out to be not as problematic as they initially appear, the sheer number of inconsistencies and hidden assumptions of Sagart's proposal lends credibility to other hypotheses, such as those of Blust (1999) and Ross (2009). Also, the presence of unresolved questions, such as the unexplained etymologies of 6 and 10, or the question as to why \*pitu was the first additive numeral to be reanalyzed as a single morpheme, shed doubt on the overall proposal. The fact that Blust (1999) and Ross (2009) arrive at very similar subgrouping hypotheses while using different methodologies provides converging evidence for a tree of the higher phylogeny of Austronesian that is substantially flatter than that proposed by Sagart (2004).

## REFERENCES

- Blust, Robert. 1996. Some remarks on the linguistic position of Thao. *Oceanic Linguistics* 35:272–94.
- . 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In *Selected papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 31–94. Symposium Series of the Institute of Linguistics, Academia Sinica 1. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- . 2003. *Thao dictionary*. Nankang, Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- . 2007. Disyllabic attractors and anti-antigemination in Austronesian sound change. *Phonology* 24:1–36.
- . 2008. Remote Melanesia: One history or two? An addendum to Donohue and Denham. *Oceanic Linguistics* 47:445–59.
- . 2009. *The Austronesian languages*. Canberra: Pacific Linguistics.
- Haspelmath, Martin. 2008. Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability. In *Aspects of language contact: New theoretical, methodological and empirical findings with special focus on Romancisation processes*, ed. by Thomas Stolz, Dik Bakker, and Rosa Salas Palomo, 43–62. Berlin: Mouton de Gruyter.
- Lynch, John. 2009. At sixes and sevens: The development of numeral systems in Vanuatu and New Caledonia. In *Discovering history through language: Papers in honour of Malcolm Ross*, ed. by Bethwyn Evans, 391–411. Canberra: Pacific Linguistics.

- Mabuchi, Toishi. 1954. Migration and distribution of the Formosan aborigines. *Anthropological Studies* (Tokyo), 18(1–2):123–54; 18(4): 23–72. [In Japanese; cited by Sagart 2004.]
- Munshi, Sadaf. 2006. Jammu and Kashmir Burushaski: Language, language contact, and change. PhD diss., University of Texas at Austin.
- Noonan, Michael. 2006. Contact-induced change in the Himalayas: The case of the Tamangic languages. *International Colloquium on Language Contact and Contact Languages*, University of Hamburg. [Online: <http://www.uwm.edu/~noonan/Contact.paper.pdf>]
- Ross, Malcolm. 2009. Proto Austronesian verbal morphology: A reappraisal. In *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, ed. by Alexander Adelaar and Andrew Pawley, 295–326. Canberra: Pacific Linguistics.
- Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43:411–44.
- Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.
- Sohn, Ho-Min. 2001. *The Korean language*. Cambridge: Cambridge University Press.
- Souag, Lameen. 2007. The typology of number borrowing in Berber. Paper presented at the Fifth Cambridge Postgraduate Conference in Linguistics. [Available online: [http://www.ling.cam.ac.uk/camling/Manuscripts/CamLing2007\\_Souag.pdf](http://www.ling.cam.ac.uk/camling/Manuscripts/CamLing2007_Souag.pdf)]
- Swadesh, Morris. 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):452–63.
- . 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21:121–37.
- Thomason, Sarah. 2001. *Language contact*. Washington, DC: Georgetown University Press.
- Tsuchida, Shigeru. (1976). *Reconstruction of Proto-Tsouic phonology*. Studies of Languages and Cultures of Asia and Africa, Monograph Series 5. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.