



**HAL**  
open science

# The Higher Phylogeny of Austronesian: a Response to Winter

Laurent Sagart

► **To cite this version:**

Laurent Sagart. The Higher Phylogeny of Austronesian: a Response to Winter. *Oceanic Linguistics*, University of Hawai'i Press, 2013, 52 (1), pp.249-254. hal-00781136

**HAL Id: hal-00781136**

**<https://hal.archives-ouvertes.fr/hal-00781136>**

Submitted on 25 Jan 2013

**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.

## *Squib*

### **The Higher Phylogeny of Austronesian: a Response to Winter.<sup>1</sup>**

Laurent Sagart

CNRS, PARIS

This paper is a response to criticism by Winter (2010) of Sagart (2004). The author gives examples of compound numerals being affected by several apparently irregular changes outside of Austronesian; argues that the number of changes supposed in his Austronesian model is realistic; presents new evidence of Formosan numerals explainable only on the basis of the PAN forms he posited; explains the order of establishment of disyllabic numerals as depending on two factors: cardinal order and number of competitors; gives Austronesian examples showing that the drive to disyllabism does apply to morphologically complex forms; ascribes the limited similarities between the phylogenies of Blust and Ross to chance. Finally, the author claims that the only realistic explanation to the nesting of six related isoglosses is a sequence of innovations.

**1. INTRODUCTION.** In Sagart (2004) I observed that the six numerals \*lima 5, \*enem 6, \*pitu 7, \*walu 8, \*Siwa 9 and \*puluq 10 are reflected in Formosan languages according to the implicational hierarchy \*puluq < \*Siwa < \*walu < \*enem < \*lima < \*pitu; that these six etyma occupy geographical areas nested like matryoshka dolls; and further, that the coastal edges of these areas, beginning with the largest, follow a path leading counterclockwise from northwest Taiwan to the east coast. I noted that this situation is anomalous if, as is generally assumed, these numerals were part of PAN, but fully expected if the numerals for 5-10 became established successively in post-PAN times. I proposed that new numerals for 5-10 were being created even as the early

---

<sup>1</sup> The author thanks E. Buchi, X. Gouvert, Y. Greub, G. Jacques, F. Kortlandt, A. Manaster Ramer and P. Videsott for assistance with Indo-European numeral systems.

Austronesians were expanding in Taiwan, along the coastal route I identified. While this implicational and geographical evidence suffices to establish the short numerals 5-10 as post-PAN innovations (a point to which I return in my conclusion), I observed that Pazeh, one of the three northwestern languages without reflexes of the six innovative numerals, has for '7', '8' and '9' five-syllable expressions meaning 5+2, 5+3, 5+4, each of which contains the phonetic ingredients for building \*pitu, \*walu and \*Siwa: thus \*pitu can be related to Pazeh *xasebidusa* through a PAN precursor \*RaCep-i-duSa, considering that \*pitu may involve a \*tuSa variant of '2' (with initial \*d- changed to \*t- on the analogy of \*telu '3', as in Amis, Rikavung Puyuma and Thao); \*walu to *xasebatulu* through PAN \*RaCep-a-telu, if the penultimate vowel is allowed to be lost, the resulting cluster is reduced to -l-, and -p- lenited to -w- before a low vowel; and \*Siwa to *xasebisupat* through PAN \*RaCep-i-Sepat, given a vowel change and another labial stop lenition before low vowel. Additionally, derivation of the short forms involves stress-conditioned pruning of initial and final material. Readers are referred to my original paper for a detailed exposition of my proposal, and especially to Table 2 on p. 418, where a sequence of six ordered sound changes (with several possible variants) collectively derives the short forms out of the long ones. Here I respond to Winter's criticisms (Winter 2010), in the order he presented them.

**2. DERIVATION OF THE POST-PAN NUMERALS.** Winter opens on the observation that the sound changes used to derive 7-8-9 "do not have parallels in the Pre-Pituish languages", a major weakness in his view. In my model, the changes affecting 7-8-9 did occur before Pituish, but only after the breakup of PAN, and only in that primary branch of PAN which leads to Pituish. There is therefore no reason why they should affect the pre-Pituish languages: Pazeh, Saisiat and Luilang. He calls the changes themselves "ad hoc". Under the comparative method, a sound change can apply to a very limited number of words, if the context triggering it is uncommon. Part of the context conditioning the changes I proposed is a particularly fast speech rate due to pronouncing 6-

10 on the same tempo as 1-5, all disyllables, in rhythmic counting. When \*RaCep-i-tuSa, \*RaCep-a-telu, \*RaCep-i-Sepat were pronounced on this accelerated tempo, phonetic overcrowding ensued. Local changes applied regularly to sounds within these expressions, but not outside of them. Phonetic normality was restored when the long expressions were reduced to CVCV(C) disyllables.

Realization that a fast speech rate affects compound numerals particularly, producing shortenings and other sound changes not seen in the rest of the vocabulary, goes back at least to Schmid (1964: 232) who specifically links them to the act of counting:

"In der Tat geht es ja beim Zählen oft so eilig zu, dass die umständlichen längeren Zahlwörter (...) nur undeutlich und manchmal in verstümmelter Form hergesagt werden. (...) Übermäßige Verkürzung von Zahlwörtern ist also an sich nicht Erstaunliches."

As examples one may cite the informal Rumanian numerals *paişpe* '14', *cinşpe* '15', *şaişpe* '16', *opşpe* '18' which are reduced from the corresponding formal numerals *patrusprezece*, *cincisprezece*, *şasesprezece*, *optsprezece*, the inherited forms (where *spre* < Lat. *super* 'over', *zece* < Lat. *decem* 'ten'). In Danish, informal *tres* '60' is reduced from the (extremely) formal *tre-sinds-tyve* (< 'three times twenty') and *halvtreds* 'fifty' is for *halvtredje-sinds-tyve* (< 'half third times twenty' [=two times twenty plus a half of a third twenty]). Modern Greek *trianta* '30', *saranta* '40' are reduced from Classical Greek *triakonta*, *tessarakonta*. In Hongkong Cantonese rapid counting and price-giving, trisyllabic numerals in which the second syllable is *sahp* [səp L] '10' have informal variants with *ah* [ə L] for [səp L]; additional changes reduce the first syllable's final consonant: *saam-sahp-sei* '34' > *saa-ah-sei*, *chat-sahp-sei* '74' > *cha'-ah-sei* (tone marks omitted). None of the changes just described affects any word outside of the compound numerals. Most are not analogically motivated. Note the importance of prunings in Rumanian, Danish and Greek. Note also that in Rumanian and Danish (as in e.g. Javanese *rolas* '12' < *loro* 'two' + *walas* 'count back'.) the short numerals are made up of elements from separate morphemes. The absence of clear evidence for lexical conditioning above is remarkable in all the changes above: in all cases phonological conditioning produces the correct output. This fits with my claim that they are regular

changes conditioned by a fast speech rate induced by rhythmic counting.

Winter next states that the drive to disyllabism in Austronesian languages does not apply to morphologically complex forms. Yet Iban (Acehnese, Maloh, etc.) *lapan* ‘eight’ originates in \*dua ‘two’ + *alap* ‘fetch’ + -an ‘two taken away’. Adelaar (1992: 117; 412) gives the origin of Malay *satu* ‘one’ as \*(ə)sa- ‘one’ + \*batu ‘stone’. Tagalog *sampu* ‘ten’ is from \*sa ‘one’ + nasal linker + \*puluq ‘ten’. Javanese *rolas* ‘twelve’ is from *loro* ‘two’ + *wəlas* ‘count back’.

Winter finds it strange that the word for 7 was innovated before 5 and 6. A distinction must be made between the moment when an innovation appears and the moment when it becomes established, having driven its competitor(s) out of existence. What counts for the trees in my paper is their order of establishment, which in turn results from the interaction of two factors: cardinal order, according to which lower numerals are established before higher ones; and the number of synonymous competitors a numeral expression had to overcome before it became established. My claim is that \*pitu ‘7’ was established before \*enem ‘6’ because there were no competitors that \*pitu needed to overcome, while \*enem had two (one additive, another multiplicative). The case of 5 is different. The displacement of PAN \*Racəp ‘5’ by \*lima, originally ‘hand’, was not motivated by the drive to disyllabism: the reason why \*lima became established as ‘5’ later than \*pitu may simply be that \*lima did not acquire the meaning ‘5’ until after \*pitu became established as ‘7’.

Winter then outlines a kind of statistical argument aiming to show that the similarity between \*pitu, \*walu and \*Siwa and complex forms in Pazeḥ is due to chance. The argument, not quantified, is not sufficiently explicit for discussion. A chance explanation is in any case strongly counter-indicated by forms like Makatao *sipat* ‘9’, whose medial stop and final -t are explained by my \*(RaCep-i-)Supat, but not by \*Siwa. For ‘8’, I have cited Amis *falū* < \*balu, whose initial stop makes better sense coming from \*(RaCe)p-a-telu (with intervocalic voicing of -p-, as in Pazeḥ *xasebatulu*) than from \*walu. A west-coast precursor of Amis *falū* can now be cited: Papora *bahalu* ‘8’ (Ogawa and Li 2006). I take this to be re-vocalized from an earlier \*bahlu < \*batlu with the -tl-

cluster (resulting from the syncope of \*e) hypothesized at stage 6 of my derivation.

Winter finds the number of changes —six— needed in my model to collectively derive three short forms out of three long ones too high. At least seven are needed to derive the Rumanian short forms for ‘14’, ‘15’ and ‘16’ (above) from the corresponding long ones: 1) loss of *-ze-*,<sup>2</sup> 2) pruning of *-ce*, 3) palatalization<sup>3</sup> of *-s-* in *spre*, 4) loss of *-r-* in *şpre*, 5) loss of *-tr-* in ‘14’, 6) loss of *-ci-* in ‘15’, 7) loss of *-s-* in *-ase-* in ‘16’.<sup>4</sup>

Winter states that disyllabism-maintaining or disyllabism-reestablishing changes are usually single changes. Malay *satu* ‘one’ < \*sa-batu requires at least lenition of *-b-* to *-w-* preceding low vowel, loss of the first *-a-* (*suatu* is attested) and loss of *-w-*. Tagalog *sampu* ‘ten’ < \*sa-n(g)a-puluq requires loss of unstressed *-a-* in the linker, labial assimilation and loss of *-l-*. In Atayal, Mayrinax *mayalpuy* ‘ten’ (< *mayal-lpuy* ‘ten’+‘count’) is seen reduced to *malpu* in Atayal point #1a of Ogawa and Li (2006), this further to *mappu* and finally to *mapu* in different varieties of Squliq.

**3. BORROWABILITY OF NUMERALS.** Winter offers the opinion that "the cultural ties between many of the Formosan language communities are very close", this forming an "optimal basis for borrowing relationships". He suggests, without being more specific, that borrowing of innovative numerals, better than a retention from PAN, explains the quinary systems of Saisiat, Luilang and Pazeh. If this is the explanation, it can only account for a small part of the data, since the relevant numerals are in large part not cognate. He cites Oceanic parallels, but accounts of the origins of those systems by either Blust (2008) or Lynch (2009) do not particularly imply the spread of loanwords. Winter argues that the numerals 5-10 can be borrowed (this is not in dispute). The question is to what extent. The borrowed scores in Tadmor and Haspelmath (n.d.) place them between basic and cultural categories, closer to the former. Sagart added that the risk from borrowing was further reduced by the circumstance that numerals are most often borrowed from a

---

2 *Unsprece* ‘11’, *doisprece* ‘12’ etc. are attested.

3 In compensation for the loss of /c/ (Schmid 1964:231, fn. 222). Schmid presents an alternative, assimilatory account of the palatalization of /s/: *Xsprezeče* > *Xspreče* > *Xspče* > *Xšpče* > *Xšpe*.

4 Vowel *-i-* in *paişpe*, *şaişpe* spread analogically from ‘12’ and ‘13’ and is discounted.

state language in the context of a monetary economy, both elements lacking in neolithic Taiwan. Winter's portrayal of this as a simplistic theory that statehood is necessary for numeral borrowing misrepresents my view. He does not deny that borrowing from non-state languages is uncommon, anyway.

He suggests that sound changes "might provide a more stable and borrowing-resistant indicator of subgrouping than numerals" (286). He notes (fn. 3) my argument that sound changes routinely spread across language boundaries, objecting that "borrowing often creates irregular sound correspondences that can be detected in subgrouping hypotheses based on the comparative method". Unless I misunderstand, we are talking at cross purposes: by 'spreading of sound changes' I mean the spreading across boundaries of regular changes *per se*. Thus, palatalization of velar stops preceding -j-, originating in Manchuria around 300 years ago, is spreading south across Sinitic language boundaries, as an entirely regular process. Such changes do not produce any irregular sound correspondences which could distinguish them from internally-motivated changes. In section 4.3 of my paper, I have argued that several consonant mergers in Formosan history, some used in Blust's subgrouping, are of this type.

**4. SPATIAL ISSUES.** Winter first observes that my model does not predict the relative locations of Formosan languages perfectly: Taokas, a Pituic language, should be spoken more to the south. The answer, obviously, is that a migration occurred. His second criticism is the opposite of the first: the spatial pattern predicted by my model is too close to reality. Judging from the number of documented migrations in the past 300 to 400 years, he expects more disruption of the initial pattern. However, the recent period coincides with the high tide of Chinese immigration to the island. Occupation of coastal areas by Chinese settlers evidently set off important population movements (Mabuchi 1954 for Atayal, Blust 2003 for Thao). One should expect lower rates of migration of Formosan peoples before Chinese intrusion.

**5. CONCLUSIONS.** Winter concludes by expressing his preference for the phylogenies of Blust (1999) and Ross (2009), which he calls "very similar" and "converging". However there are no innovation-defined subgroups common to these two proposals. The similarity between them comes entirely from their treating Puyuma and Rukai as primary branches. Ross's two other branches: Tsou and Nuclear Austronesian, clash with Blust's Tsouic. Similarity confined to a subset of the non-innovation-defined branches in two phylogenies is not impressive. I regard it as random. See my criticisms of these proposals in Sagart (2004; 2010) and the rejoinder in Teng and Ross (2010).

Throughout his review, Winter never confronts my observation that the isoglosses for the short numerals 5-10 are nested, as if that pattern was neutral with respect to the question whether they are retentions or innovations. Nesting of isoglosses is common in dialectological maps and is of no consequence when small numbers of unrelated characters are involved. However, nesting of isoglosses for related characters—such as qualitatively different stages in a chain shift, cf. Labov, Ash and Bober 2006:43, 119, and maps 11.4, 11.7; successive stages in the broadening of the context of a sound change, cf. Goossens 1969:51 and map 9; or contiguous members of a paradigm, as in my model—is generally seen as the spatial signature of sequences of innovations. It is indeed difficult to see what mechanism, other than chance, could result in a set of retentions displaying a nesting pattern; and ascribing the nesting of *six* related isoglosses to chance stretches credibility. Those who maintain that the short numerals are retentions have their nesting to explain.

## REFERENCES

- Adelaar, Alexander K. 1992. *Proto-Malayic: the reconstruction of its phonology and parts of its morphology and lexicon*. Pacific Linguistics C-119, Canberra: Research School of Pacific Studies, Dept. of Linguistics, A.N.U.
- Blust, Robert. 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. Taipei: Institute of Linguistics (preparatory office).
- . 2003. *Thao Dictionary*. Nankang: Institute of Linguistics.
- . 2008. Remote Melanesia: One history or two? An addendum to Donohue and Denham. *Oceanic Linguistics* 47:445–59.
- Goossens, Jan. 1969. *Strukturelle Sprachgeographie, eine Einführung in Methodik und Ergebnisse*. Heidelberg: Carl Winter. Universitätsverlag.
- Labov, William, Sharon Ash and Charles Boberg. 2006. *The Atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton/de Gruyter.
- Lynch, John (2009) At sixes and sevens. In *Discovering history through language, papers in honour of Malcolm Ross*, ed. by Bethwyn Evans, 391-411. Pacific Linguistics #605. Canberra: Research school of Pacific and Asian Studies, Australian National University.
- Mabuchi, Toishi. 1954. Migration and distribution of the Formosan aborigines [in Japanese]. *Minzogaku Kenkyu* 18,1-2: 123-154, 18,4:23-72.
- Ogawa, Naoyoshi, and Paul J.-K. Li (2006) *A Comparative Vocabulary of Formosan Languages and Dialects*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Ross, Malcolm D. 2009. Proto Austronesian verbal morphology: a reappraisal. In *Austronesian historical linguistics and culture history. A festschrift for Robert Blust*, ed. by Alexander K. Adelaar and Andrew K. Pawley: 285-316. Canberra: Pacific Linguistics.

- Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43, 2: 411-444.
- . 2010. Is Puyuma a primary branch of Austronesian ? *Oceanic Linguistics* 49, 1: 194-204.
- Schmid, Heinrich. 1964. Zur Entwicklungsgeschichte der romanischen Zahlwörter. *Vox Romanica* 23, 186-238.
- Tadmor, Uri and Martin Haspelmath (n.d.) *World Loanword Database*. Online at <http://wold.livingsources.org/semanticfield>. Accessed June 11, 2011.
- Teng, Stacy F. and Malcolm Ross. 2010. Squib: Is Puyuma a Primary branch of Austronesian? A Reply to Sagart. *Oceanic Linguistics* 49.2:543-558.
- Winter, Bodo. 2010. Squib: A note on the higher phylogeny of Austronesian. *Oceanic Linguistics* 49, 1: 282-287.