

凱達格蘭 · 巴賽語

Ketagalan · Basay — 從記憶到再生

[首頁](#) [文法](#) [教育推進](#) [研究成果](#) [研究筆記](#) [語音合成](#) [辭典](#)

[Home](#) / [Research](#) / [This article](#)

[中文](#) [日本語](#) [English](#)

Reanalysis of the Basay Syllable Inventory

Correcting a Mixed-Source Analysis and Evidence for Language Contact

Author: Tsai, Yung-kuei (蔡永桂)

Date: June 23, 2026

Type: Methodological correction (phonological analysis / language contact)

License: CC BY 4.0 **Citation ID:** basay.tw/research/2026-06-basay-syllable-revised/

Abstract

This paper corrects a prior analysis of the Basay (巴賽語) syllable inventory that pooled native vocabulary (source=B), Yilan dialect data (source=T+M), and other source types, reporting 486 syllable types. Source-separated reanalysis yields 266 types across 22 onsets for native vocabulary (B) and 315 types across 38 onsets for the Yilan dialect (T+M). The mixed analysis produced three specific errors: (1) CV was reported as the dominant structure, whereas source=B shows CVC dominance (54%); (2) phonemes /q/, /z/, /ʃ/, /ʒ/ were attributed to Basay, whereas they are absent from source=B and exclusive to T+M; (3) the inventory was overstated at 486 types. The T+M-

exclusive phonemes correspond to those of Kavalan (噶瑪蘭語), spoken in the geographically adjacent Yilan Plain, and are interpreted as evidence of phonological contact borrowing. This paper documents the errors, their sources, and the corrected results, and argues that source separation is a non-negotiable methodological step in phonological description from multilayered lexical databases.

Keywords: Basay, syllable inventory, source separation, methodological correction, language contact, Kavalan, Formosan languages

Cite this article

APA:

Tsai, Y.-k. (2026). Reanalysis of the Basay syllable inventory: Correcting a mixed-source analysis and evidence for language contact. basay.tw.
<https://basay.tw/research/2026-06-basay-syllable-revised/en/>

BibTeX:

```
@misc{tsai2026syllableRevised_en,  
  author = {Tsai, Yung-kuei},  
  title = {Reanalysis of the {Basay} Syllable Inventory},  
  year = {2026},  
  month = {6},  
  url = {https://basay.tw/research/2026-06-basay-syllable-revised/en/},  
  note = {Correcting a mixed-source analysis and evidence for Kavalan  
language contact}  
}
```

1. Introduction: The Mixed Analysis and Its Problems

A preceding study pooled all non-PAN entries (2,364) from the Basay lexical database and reported 486 syllable types (frequency ≥ 2). That analysis attributed phonemes including /q/, /z/, /l/, and /ɓ/ to Basay and described the inventory as CV-dominant.

The database contains entries assigned to six source codes (Table 1). Pooling B (native vocabulary), T and M (Yilan area dialects), and S (a source with suspected heavy Kavalan admixture, excluded here) without separation treats distinct phonological layers as a single system. This is a methodological error that this paper corrects.

Source	Entries	Layer
B	1,117	Native Basay vocabulary
T	588	Trobiawan dialect, Yilan area
M	541	Trobiawan (vocabulary-only collection)
S	113	Suspected Kavalan admixture — excluded
V	5	Unidentified — excluded
PAN	960	Proto-Austronesian reconstructions — excluded

2. Corrected Results

2.1 Scale

	Mixed (prior)	B only	T+M only
Entries	2,364	1,117	1,129
Syllable types	486	266	315
Onset categories	48	22	38
Shared ($B \cap T+M$)	—	128	128

2.2 Syllable Structure

Structure	B	T+M	Mixed (prior)
V	4	4	4
VC	1	2	(misclassified as CVC)
VV	2	2	(misclassified as CVV)
VVC	1	0	(misclassified)
CV	75	63	54
CVC	134	159	30
CVV	36	27	15
CVVC	7	26	—
other	6	22	—

2.3 Onset Distribution

Onset	IPA	B	T+M	Mixed attribution
h	h	✓	—	Incorrectly pooled
s'	ʃ	✓	—	Incorrectly pooled
ts'	tʃ	✓	—	Incorrectly pooled
q	q	—	✓	Incorrectly attributed to Basay
z	z	—	✓	Incorrectly attributed to Basay
z'	ʒ	—	✓	Incorrectly attributed to Basay
ʎ'	ʎ	—	✓	Incorrectly attributed to Basay
mʎ'	mʎ	—	✓	Incorrectly attributed to Basay
vʎ'	vʎ	—	✓	Incorrectly attributed to Basay

3. The Kavalan Contact Hypothesis

Six onset types exclusive to T+M — /q/, /z/, /ʒ/, /ʎ/, /mʎ/, /vʎ/ — are documented phonemes of Kavalan (Li 2000), the Formosan language indigenous to the Yilan Plain adjacent to the Basay Yilan dialect territory. Their combined frequency in T+M reaches approximately 176 tokens across 34 syllable types, and they are entirely absent from source=B.

Under Thomason & Kaufman's (1988) contact typology, systemic integration of borrowed phonemes at this scale indicates intensive contact, not incidental lexical borrowing. The parallel increase of CVVC structures in T+M (26 types vs. 7 in B) — particularly involving the Kavalan-derived z and z' onsets — extends the contact evidence to the level of syllable structure.

The absence of *h*, */ʃ/*, and */tʃ/* from T+M, all present in source=B, may reflect convergence toward the Kavalan phonological type, which lacks these segments. This constitutes possible contact-induced attrition in the T+M variety.

4. Methodological Lessons

The mixed analysis produced four compounding errors, each traceable to source pooling:

1. **Inflated inventory size:** 486 types vs. B=266, T+M=315 (overlap 128)
2. **Distorted structure profile:** CV apparent dominance masks CVC dominance in native vocabulary (134 types, 50%)
3. **Misattributed phonemes:** */q/*, */z/*, */l/*, */ɟ/* described as native Basay phonemes
4. **Obscured contact signal:** T+M's Kavalan-derived phonemes merged with native vocabulary, preventing detection of the contact pattern

For documentary databases covering multiple dialect strata, source separation before phonological analysis is not optional. The source labels in such databases encode linguistically meaningful distinctions; ignoring them is equivalent to pooling data from different languages.

5. Revised Claims

Prior claim	Corrected claim
Basay has 486 syllable types	Native vocabulary: 266 types; Yilan dialect: 315 types

Prior claim	Corrected claim
CV is dominant (54 types)	In native vocabulary, CVC is dominant (134 types, 54%)
/q/, /z/, /l/, /ɣ/ are Basay phonemes	These are absent from native vocabulary; T+M only
48 onset categories	Native vocabulary: 22; Yilan dialect: 38

6. Conclusion

Source-separated reanalysis corrects the mixed-analysis account of Basay phonology in three respects: inventory size, syllable structure dominance, and phoneme attribution. The T+M-exclusive phonemes are reinterpreted as Kavalan contact borrowings, a hypothesis supported by geographic adjacency, phoneme identity, high token frequency, and structural parallelism. The corrected files (source=B and T+M syllable inventories) are publicly available at basay.tw; the prior mixed analysis is preserved in the archive with a correction notice linking to this paper.

References

- Blust, R. (1999). Subgrouping, circularity and extinction. In E. Zeitoun & P. J.-K. Li (Eds.), *Selected papers from the Eighth International Conference on Austronesian Linguistics* (pp. 31–94). Academia Sinica.
- Blevins, J. (1995). The syllable in phonological theory. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206–244). Blackwell.
- Li, Paul Jen-kuei. (1996). *The Formosan Tribes and Languages in I-Lan*. Yilan: Yilan County Government.

- Li, Paul Jen-kuei. (2000). *The Phonological System of Taiwan Austronesian Languages*. Taipei: Crane Publishing.
- Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.
- Institute of Linguistics, Academia Sinica (Ed.). *Basay Lexical Database* (basay_dict.jsonl). Taipei: Academia Sinica.

[← Back to Research](#)

"Correction is the foundation of scientific integrity."

© 2026 basay.tw · CC BY 4.0 | [Return to Home](#)