

# 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](#)

## The Syllable Inventory of Basay: Native Vocabulary (source=B)

*A Source-Separated Analysis of the Phonological System*

**Author:** Tsai, Yung-kuei (蔡永桂)

**Date:** June 23, 2026

**Type:** Original research (linguistic typology / quantitative phonology)

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

### Abstract

This paper presents a syllable inventory of Basay, an extinct Austronesian language of northern Taiwan, based exclusively on native vocabulary entries (source=B, 1,117 entries) drawn from the Basay lexical database. A prior analysis pooling all source types (B, T, M) yielded 486 syllable types; the present source-separated analysis identifies 266 syllable types and 22 onset categories from the native vocabulary alone. CVC is the dominant syllable structure (134 types, 50%), overturning the CV-dominant picture produced by the mixed analysis. The onsets *h*, */ʃ/ (s')*, */tʃ/ (ts')*, and *sj* are distinctive to native vocabulary and absent from the Yilan dialect data (T+M). Conversely,

q, z, /ʒ/ (z'), and /l/ (l'), which appeared prominently in the mixed inventory, are entirely absent from source=B, indicating that these phonemes do not belong to the core Basay phonological system.

**Keywords:** Basay, native vocabulary, syllable inventory, source separation, phonological system, Formosan languages

## Cite this article

### APA:

Tsai, Y.-k. (2026). The syllable inventory of Basay: Native vocabulary (source=B) – A source-separated analysis of the phonological system. basay.tw. <https://basay.tw/research/2026-06-basay-syllable-B/en/>

### BibTeX:

```
@misc{tsai2026syllableB_en,
  author = {Tsai, Yung-kuei},
  title = {The Syllable Inventory of {Basay}: {Native} Vocabulary (source=B)},
  year = {2026},
  month = {6},
  url = {https://basay.tw/research/2026-06-basay-syllable-B/en/}
}
```

## 1. Introduction

Basay is a Formosan Austronesian language formerly spoken by the Basay, a Pingpu (plains indigenous) people of northern Taiwan. Lexical records survive in Dutch colonial sources and Qing dynasty documents; the language ceased to have living speakers by the early twentieth century (Li 1996, 2000). Documentation and revitalization efforts continue at the Institute of Linguistics, Academia Sinica.

A preceding study (the mixed analysis) pooled all non-PAN entries (2,364) from the lexical database and reported 486 syllable types. The database contains entries from multiple distinct source types (Table 1), representing different dialect layers and contact situations. Pooling these without separation conflates distinct phonological systems under a single "Basay" label.

Source	Entries	Content
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)

This paper analyzes source=B exclusively. The T+M Yilan dialect data and the contact-with-Kavalan hypothesis are treated in a companion paper.

## 2. Method

Syllable extraction followed the procedure of the preceding study with two refinements. First, syllable structure classification was corrected to distinguish onset-less syllables: forms without an onset consonant are classified as V-type (V, VC, VV, VVC) rather than conflated with onset-bearing types. Second, complex onset clusters (two or more distinct consonant phonemes) are classified as "other," distinct from single-phoneme digraphs such as ts, n', s', l', z', and ts'.

Table 1. Orthography–IPA correspondence for source=B

Orthography	IPA	Description
n'	ŋ	Velar nasal
s'	ʃ	Palato-alveolar fricative
o'	ə	Mid central vowel (schwa)
' (coda)	ʔ	Glottal stop (syllable-final coda)
ts	ts	Alveolar affricate
ts'	tʃ	Palato-alveolar affricate
sj	sj	Palatal fricative variant
j	j ~ dʒ	Approximant or affricate

Note: source=B contains no occurrences of /l/ (l'), /ʒ/ (z'), /q/, or /z/.

## 3. Results

### 3.1 Overall Statistics

Parameter	Value
Entries analyzed	1,117
Syllable types (frequency $\geq 2$ )	266

Parameter	Value
Onset categories	22
Highest-frequency syllable	la (148 occurrences)
High frequency ( $\geq 50$ )	12 types
Mid frequency (10–49)	53 types
Low frequency (2–9)	201 types

### 3.2 Syllable Structure Distribution

**Note:** **o'** represents the single vowel /ə/; syllables such as **no'** /nə/ and **ko'** /kə/ are classified as CV, not CVC.

Structure	Types	%	Description
V	4	1.5%	Vowel nucleus only (a, i, o, u)
VC	1	0.4%	Vowel + coda (at)
VV	2	0.8%	Diphthong nucleus (ai, au)
VVC	1	0.4%	Diphthong + coda (oat)
CV	75	24.8%	Basic canonical type
CVC	<b>134</b>	<b>53.8%</b>	<b>Most frequent type</b>
CVV	36	13.5%	Diphthong nucleus
CVVC	7	2.6%	Diphthong + coda
other	6	2.3%	Onset clusters etc.

Structure	Types	%	Description
<b>Total</b>	<b>266</b>	<b>100%</b>	

### 3.3 Onset Distribution

Onset	IPA	Types	Tokens	Key syllables
∅	—	8	41	a, i, o, u
b	b	19	207	ba, be, bu
h	h	20	138	ha, hi, he
j	j~dʒ	6	44	ja, jen
k	k	23	228	ka, ke, ku
ɭ	l	32	434	la, li, lu
m	m	21	215	ma, man, mu
n	n	20	211	na, nan, nu
n'	ŋ	3	17	n'a, n'o
p	p	22	243	pa, pu, pi
r	r	10	65	ra, ri, ru
s	s	30	476	se, sa, su
s'	ʃ	3	16	s'i, s'a
sj	sj	3	13	sja, sje

Onset	IPA	Types	Tokens	Key syllables
t	t	26	328	te, ta, ti
ts	ts	8	60	tsa, tse
ts'	tʃ	2	9	ts'i, ts'a
v	v	2	5	va, ve
w	w	5	52	wa, wan

## 4. Discussion

### 4.1 CVC Dominance: A Corrected Picture

The finding that CVC is the dominant structure (54%) represents a significant correction to the mixed analysis. Formosan languages generally retain syllable-final consonants inherited from Proto-Austronesian, in contrast to the coda-less tendencies of Polynesian languages (Blust 1999). Source=B's CVC dominance aligns with this conservative Formosan profile and suggests that the "CV-dominant Basay" picture of the mixed analysis was an artifact of T+M data mixing.

### 4.2 Palatality in Native Vocabulary

The onset series s' /ʃ/, ts' /tʃ/, and sj, which are present in source=B but entirely absent from T+M, indicate a palatal contrast in the native phonological system. A parallel palatal series is attested in Atayal (Li 1996), suggesting a possible areal feature of the northern Taiwan linguistic zone.

### 4.3 Absent Phonemes and Their Implications

The complete absence of /q/, /z/, /ʒ/ (z'), and /l/ (l') from source=B — phonemes that featured prominently in the mixed analysis — indicates that these do not belong to the core Basay phonological inventory. Their appearance in the mixed data is attributable to T+M Yilan dialect entries, where they are present at high frequency and are hypothesized to reflect contact with Kavalan (see companion paper).

---

## 5. Conclusion

---

The source=B-only analysis yields a syllable inventory of 266 types across 22 onset categories, with CVC as the dominant structure (54%). The findings correct three erroneous claims of the mixed analysis: (1) CV dominance, (2) the presence of /q/, /z/, /l/, /ʒ/ as native phonemes, and (3) the 486-type inventory size. Source separation is shown to be a methodological prerequisite for phonological description of multilayered documentary lexical databases.

---

## 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](#)

*"Every syllable counted is a language remembered."*

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