

巴賽語 (Basay) の音節目録と音韻類型論的位置づけ

—辞書データに基づく計量的分析—

蔡永桂 (Tsai Yung-kuei)

2026年6月

basay.tw/research/2026-06-basay-syllable/

License: CC BY 4.0

巴賽語 (Basay) の音節目録と音韻類型論的位置づけ

—辞書データに基づく計量的分析—

要旨

本稿は、台湾北部に話されたオーストロネシア系言語、巴賽語 (Basay) の音節目録を辞書データ (3,324エントリ) から計量的に抽出し、その音韻構造を記述するとともに、世界言語の音節規模との類型論的比較を行う。PAN (Proto-Austronesian) 再建形ソースを除く 2,364 エントリを対象に、頻度2以上の音節 486 種を確定した。巴賽語は、そり舌側面音 (l̥) ・有声歯茎側面摩擦音 (ɬ) ・軟口蓋鼻音 (ŋ) ・硬口蓋歯茎摩擦音 (ʃ) などの特徴的な音素を有し、複雑な頭子音連続 (クラスター) を許容する。その結果として 486 種という音節目録は、ハワイ語 (約 60 種) や日本語 (約 100 種) を大幅に上回り、北京語 (約 400 種・声調なし) を超える規模であり、台湾原住民語の中でも音韻的に複雑な言語に位置づけられる。

キーワード: 巴賽語・平埔族・音節目録・音韻類型論・オーストロネシア語族・消滅危機言語

1. はじめに

巴賽語 (Basay、バサイ語) は、台湾北部の宜蘭平野から台北盆地北東部にかけて居住した平埔族 (台湾平地原住民) の一集団、巴賽族が話したオーストロネシア系言語である。17世紀のオランダ統治期および清朝期の記録に断片的な語彙が残されているが、日本統治期を経て 20 世紀前半までに母語話者がほぼ途絶したとされる (李壬癸 1996; 2000)。現在は消滅危機言語として分類されており、台湾中央研究院語言學研究所を中心とした記録保存・復元活動が続けられている。

本稿の目的は二つある。第一に、現存する巴賽語辞書データを用いて音節目録を計量的に抽出し、正書法とIPA（国際音声記号）の対照を明示しながら、巴賽語の音節体系を記述することである。第二に、抽出された音節目録の規模を世界言語と比較することで、巴賽語の音韻類型論的位置を明らかにすることである。

なお、本稿では辞書エントリから機械的に音節を抽出する計量的手法を採用している。この手法は文法記述に基づく手作業の音節分析を代替するものではなく、あくまで語彙データから観察される分布の記述を目的とする。

2. データと方法

2.1 使用データ

分析に使用したのは、巴賽語辞書データベース（`basay_dict.jsonl`、3,324エントリ）である。各エントリは巴賽語形式（正書法）、中国語・日本語・英語訳、出典ソースのフィールドを含む。出典ソースには以下の区分がある：**B**（巴賽語固有語彙）、**T**（台湾語借用語）、**M**（閩南語借用語）、**S**、**V**、**PAN**（Proto-Austronesian 再建形）。

本稿では、PAN 再建形（`source: PAN`、960エントリ）を除外した 2,364 エントリを分析対象とした。PAN 再建形は現代巴賽語の記録語彙ではなく、比較言語学的な再建形であるため、実証的な音節目録の記述には適さないと判断した。

2.2 正書法体系

巴賽語の正書法は、以下の対照表に従ってラテン文字により記される（表1参照）。

表1 巴賽語正書法・IPA 対照表

正書法	IPA	説明
<code>n'</code>	ŋ	軟口蓋鼻音
<code>s'</code>	ʃ	硬口蓋歯茎摩擦音
<code>l'</code>	ʎ	そり舌側面音
<code>z'</code>	ʙ	有声歯茎側面摩擦音
<code>o'</code>	ə	中舌中段母音（シュワー）
<code>'</code> （韻尾）	ʔ	声門閉鎖音（前音節韻尾）
<code>q</code>	q	口蓋垂／咽頭音（推定）
<code>ts</code>	ts	歯茎破擦音

正書法	IPA	説明
ts'	tʃ	硬口蓋歯茎破擦音
j	j~dʒ	接近音または破擦音（文脈依存）
v	v	有声唇歯摩擦音

声門閉鎖音（'）は前音節の韻尾音素として表記されるものであり、独立した音節頭子音（onset）ではない。したがって、' で始まる表記（例：'a、'uɪ）は前音節韻尾の転写アーティファクトとして音節目録から除外した。

2.3 音節抽出手順

音節の抽出は以下の手順で行った。

1. 各エントリの巴賽語形式（正書法）から注記・括弧・代替形を除去し、第一形式のみを使用した。
2. 母音字（a, e, i, o, u および特殊母音文字）を核とする音節分割アルゴリズムを適用した。具体的には、(C*)V(V?)(C?) の構造を基本とし、後続する子音が次音節の onset として解釈できる場合（後続が母音で始まる場合）はコーダに含めない方針を採った。
3. 抽出された音節のうち、スペース・句読点・数字・ハイフン起始・' 起始のものをノイズとして除外した。
4. 頻度1の音節をさらに除外した（辞書の転写ゆれ・借用語アーティファクトの排除）。
5. 残る頻度2以上の音節を最終的な目録とした。

3. 音節目録の記述

3.1 全体統計

上記の手順により、486種の音節（頻度2以上）が確定した。頻度の分布は表2のとおりである。

表2 頻度区分別音節数

頻度区分	種数	全体比
高頻度（50回以上、★）	27種	5.6%
中頻度（10～49回、☆）	113種	23.3%
低頻度（2～9回）	346種	71.2%
合計	486種	100%

高頻度音節のほとんどは単純な CV 構造であり (la, ma, sa, ta, se, ka, pa など) 、これはオー
ストロネシア語族に広く見られる音節選好と一致する。

3.2 頭子音 (Onset) 別分布

onset 別の分布を表3に示す。

表3 主要 onset 別音節数 (頻度2以上)

Onset	IPA	音節種数	代表音節 (高頻度順)
∅ (頭子音なし)	—	15	a, i, u, au, o
b	b	21	ba, be, bu, bo, bun
h	h	20	ha, hi, he, ho, hu
j	j~dʒ	6	ja, jan, jen, ju
k	k	31	ka, ku, ki, ke, kə
l	l	40	la, li, lu, lan, lai
l'	l̥	5	la, li, lai, la, le
m	m	39	ma, man, mu, mi, mal
n	n	37	na, nan, nu, ni, nə
n'	ɲ	6	ɲa, ɲo, ɲan, ɲu
p	p	30	pa, pu, pan, pə, pi
q	q	21	qa, qu, qo, qai, qul
r	r	17	ru, ri, ra, re, rit
s	s	49	sa, se, su, si, san
s'	ʃ	3	ʃi, ʃa, ʃe
t	t	37	ta, te, tu, ti, tan
ts	ts	15	tsu, tsa, tse, tsat
ts'	tʃ	2	tʃi, tʃa
v	v	24	va, vu, van, vi
w	w	8	wa, wan, wai, wak
y	j	2	ja, jan (y表記)
z	z	22	za, zu, zo, zan
z'	ʒ	5	ʒa, ʒu, ʒian, ʒaz
クラスター (ml' , mn, kn, tmなど)	各種	計28	ml' a, kna, tma など

onset **l** の 40 種が最多であり、**s** (49種) がそれに続く。これは接頭辞・接中辞の発達したオーストロネシア語族において、流音および歯茎音が語形成上重要な役割を担う傾向と対応している。

3.3 音節構造

確認された主要な音節構造は以下のとおりである。

- ・ V : a, i, u など (母音のみ)
- ・ CV : la, ma, sa, ba など (最多、基本型)
- ・ CVC : lan, man, tan, bun など
- ・ CVV : lau, mai, tiu など (二重母音核)
- ・ CVVC : laan, maan など (長母音+コーダ)
- ・ CCV : kna, tma, ml' a, sja など (クラスター)
- ・ CCVC : knat, mnan, tmat など

CV が最も普遍的な型であり、CCVC 以上の複雑構造は頻度が低い傾向にある。これは類型論的に予測される音節の有標性の階層 (Blevins 1995; Maddieson 2006) と整合する。

3.4 特徴的な音素

巴賽語の音節目録が他のオーストロネシア語族、特に他の台湾原住民語と比べて際立つ点は以下の4音素の存在である。

そり舌側面音 (l̥、正書法 l') : 台湾原住民語の中でもそり舌音を有する言語は限られる。巴賽語では独立した音素として機能し、**l'a** (la)、**l'i** (li) など14種の音節を構成する。

有声歯茎側面摩擦音 (ɮ、正書法 z') : 世界言語においても稀な音素であり、IPA 一覧において特別な記号が割り当てられている。巴賽語では **z'a** (ɮa)、**z'u** (ɮu) など5種が確認される。

有声唇歯摩擦音 (v) : オーストロネシア語族において **v** を独立音素として持つ言語は少ない。フィリピン系言語などの一部に見られるが、台湾原住民語では巴賽語の特徴とされる (李壬癸 1996)。

複雑な頭子音クラスター (ml', mn, kn, tm など) : **ml'a**、**kna**、**tma**、**mnan** などのクラスターは、単純な CV 型を基本とするポリネシア系言語とは対照的であり、マレー・ポリネシア語群の中でもフィリピン系・台湾系言語の保守的特徴と見られる。

4. 類型論的比較

4.1 音節目録規模の国際比較

音節の「数え方」は言語研究において一様でなく、声調・長短を含むかどうかによって大きく異なる。ここでは声調・長短変種を含まない音節形式の種数を比較基準とした（表4参照）。

表4 主要言語の音節目録規模比較

言語	音節種数（概数）	音節構造の最大型	備考
ハワイ語	約 60 種	CV, V	子音8・母音5
日本語	約 100 種	CV（+撥音・促音）	モーラ単位を基本とする
北京語（普通話）	約 400 種	CVC（韻尾は n/ŋ のみ）	声調込みで約 1,300 種
巴賽語（本稿）	486 種	CCVC	頻度2以上、クラスター含む
ドイツ語	約 2,700 種	CCCCVCCCC	複雑な語末クラスター
英語	約 10,000～15,000 種	最大 CCCVCCCC	語彙規模・借用語の影響大
テルグ語	約 12,000 種	複雑	インド系言語

この比較から、巴賽語の 486 種という数値は、孤立語的・単音節語的傾向を持つ東アジア言語（日本語・北京語）よりも明らかに多く、かつ屈折語的・複雑音節を持つヨーロッパ言語には及ばない、中間的な規模であることが分かる。

4.2 音節複雑性と音素目録の関係

Fenk-Oczlon & Fenk (2021) は、61言語の分析から音素目録の規模と音節あたりの音素数の間に正の相関があることを示した。また Maddieson (2006) は、WALS (World Atlas of Language Structures) のデータを用いて、音節構造が複雑な言語ほど子音目録が大きい傾向があることを指摘している。

巴賽語の場合、**ʃ**、**ʒ**、**ʔ**（韻尾）、**q**、**tʃ**、**v** など類型論的に有標な音素を含む大きな音素目録が、クラスターを許容する複雑な音節構造と相乗して、500 種近い音節目録を生成していると解釈できる。この点で巴賽語は Fenk-Oczlon & Fenk の一般化と整合する事例といえる。

4.3 オーストロネシア語族内での位置

台湾原住民語は「フォルモサ語群」として、オーストロネシア語族の最も基底的な分岐を代表する (Blust 1999)。一般に、ポリネシア系言語（ハワイ語、サモア語など）が CV 型のシンプルな音節を持つのに対し、フィリピン系・台湾原住民系言語は CVC や CCVC を許容する傾向がある。

台湾原住民語の中でも音節構造が複雑な言語として、泰雅語 (Atayal) や排灣語 (Paiwan) が知られている。巴賽語は現存語彙が限られるため厳密な比較は困難だが、有声歯茎側面摩擦音 (ɸ) とそり舌側面音 (l) の併存は、これらの言語にも見られない特異な組み合わせであり、巴賽語の独自性を示す。

4.4 話者人口と音節規模

Fenk-Oczlon & Fenk (2021) および Lupyan & Dale (2010) は、話者人口の少ない言語ほど音韻体系が複雑になる傾向があると主張している。これは小規模コミュニティでは言語接触・外部からの学習者圧力が弱く、音素目録や音節構造が保守的・複雑なまま維持されやすいためと説明される。巴賽語は話者が消滅した絶滅言語であり、この仮説の極端な事例として、複雑な音節体系が記録段階まで維持されていたことを示唆している。

5. 方法論的注記

本稿の計量的手法にはいくつかの限界がある。

第一に、音節分割アルゴリズムは $(C^*)V(V?)(C?)$ の構造に基づく機械的な処理であり、形態論的境界・語形成上の接辞の切れ目を考慮していない。このため、接辞が含まれる語彙では音節分割が音韻論の実態と乖離する可能性がある。

第二に、辞書データの転写一貫性に依存している。複数の調査者・時代にわたって記録された語彙では、同一音素が異なる表記で記されることがあり、音節の重複計上や過少計上が生じうる。頻度1の音節を除外したのは、この問題を部分的に緩和するための措置である。

第三に、今回確定した 486 種という数値は「辞書に記録された語彙から観察される音節の種類数」であり、「巴賽語の音韻体系が理論的に許容する全音節数」ではない。前者は後者の部分集合であり、語彙の網羅性に依存する。

これらの限界を踏まえた上で、本稿の記述は現存最良のデータに基づく暫定的な音節目録として理解されたい。

6. おわりに

本稿では、巴賽語辞書データを用いた計量的音節抽出により、頻度2以上の音節 486 種を確定し、その音韻的特徴と類型論的位置を論じた。主要な知見は以下のとおりである。

1. 巴賽語の音節目録 (486種) は、ハワイ語 (約60種) ・日本語 (約100種) ・北京語 (約400種) を上回り、世界言語の中では中規模から中上位に位置する。

2. そり舌側面音 (l) 、有声歯茎側面摩擦音 (ɮ) 、有声唇歯摩擦音 (v) 、複雑な頭子音クラスター (ml'、kn、tm など) の存在が、音節目録の規模拡大に寄与している。
3. これらの特徴は、音素目録規模と音節複雑性の正の相関という類型論的一般化と整合し、かつ台湾原住民語群の中でも巴賽語が音韻的に際立った言語であることを示す。
4. 絶滅危機言語・消滅言語の音韻記述において、辞書データを用いた計量的アプローチは、限られたデータから音節体系の概観を得る有効な補完的手法となりうる。

巴賽語の本格的な音韻論的記述のためには、音声コーパスの分析、形態音韻論的な音節境界の検討、および隣接平埔族言語（凱達格蘭語・噶瑪蘭語など）との比較が今後の課題として残る。

参考文献

Blust, R. (1999). Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. 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.

Fenk-Oczlon, G., & Fenk, A. (2021). Linguistic complexity: Relationships between phoneme inventory size, syllable complexity, word and clause length, and population size. *Frontiers in Communication*, 6, 626032.

Li, Paul Jen-kuei (1996). *The Formosan Tribes and Languages in I-Lan*. Yilan: Yilan County Government.)

Li, P. J.-K. [李王癸] (2000). *台灣南島語言的語音符號系統*. 台北：文鶴出版。

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLOS ONE*, 5(1), e8559.

Maddieson, I. (2006). Correlating phonological complexity: Data and validation. *Linguistic Typology*, 10(1), 89–118.

Neergaard, K. D., & Huang, C.-R. (2019). Constructing the Mandarin phonological network: Novel syllable inventory used to identify schematic segmentation. *Complexity*, 2019, 6979830.

中央研究院語言學研究所 (編). 巴賽語辭書データベース (basay_dict.jsonl). 台北：中央研究院.

本研究はオープンデータとして公開されている巴賽語辭書データベースに基づく。データ抽出・分析には Python を使用した。