16th International Conference on Austronesian Linguistics
22 June 2024
Manila, Philippines

Emergence of divergent phonotactics in Austronesian: a distributional typological approach

(https://bit.ly/3VAAI2B)

Shelece Easterday
Blaine Billings
Clemens Mayer
Department of Linguistics, University of Hawai'i at Mānoa

Background

Austronesian phonology is often described as unremarkable.

Statements about syllable structure and phonotactics in the family emphasize their relative simplicity.

Background

*"Viewed crosslinguistically, Austronesian languages tend to be **fairly inconspicuous with regard to basic phonological features.** [...] The most common syllable structures are **(C)V** and **(C)V(C)**."*

(Adelaar & Himmelmann 2005: 115)

*"[L]anguages in this subgroup are frequently **phonologically less complex than those of many other linguistic groupings in the world**. Syllable structures tend to approximate a simple **CV** type."*

(Lynch, Ross, & Crowley 2002: 34)

Background

Blust (2013) illustrates outlying phonotactic patterns with a handful of language-specific and subgroup-specific examples.

However, to date, there is no reference work **quantifying the relative frequency of phonotactic patterns** within the family.

(But see Donohue to appear on the segmental phonology of Malayo-Polynesian languages of Southeast Asia)

Research questions

1. What is the range and distribution of phonotactic patterns in Austronesian?
2. How do Austronesian phonotactic patterns compare to global patterns?
3. What is the geographic patterning of divergent patterns within Austronesian?
4. How have diverging patterns emerged?

We take a **distributional typology** approach to these questions (Bickel 2015).

Methodology: language sample

148 Austronesian languages (currently)

Selected for:

- genealogical diversity
- geographical representation
- adequate phonological description in source (usually a reference grammar)

Methodology: data

Languages coded for:

- Maximal onset and coda size
- Obligatoriness of onset
- Biconsonantal onset patterns
- Properties of word-internal codas
- Diphthong/complex nucleus inventories
- Vowel hiatus patterns
- Stress patterns

Methodology: global context

For some of our comparisons, we use a global sample of **178 languages** from the Syllable Structure chapter (Easterday to appear) in the ATLAs database (Inman et al. to appear).

- no family represented by more than one language
- geographically diverse

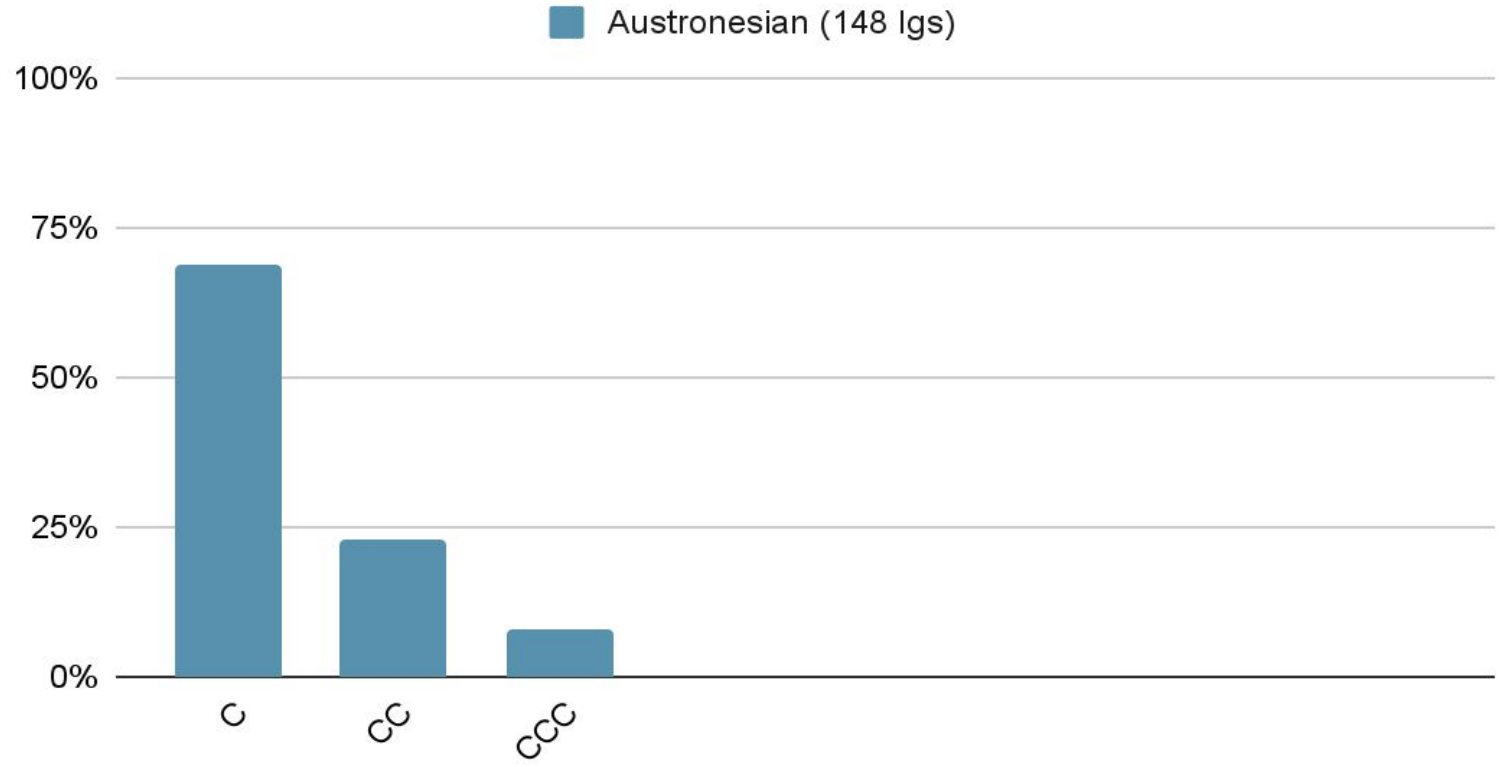
Results: onset patterns

Previous claim:

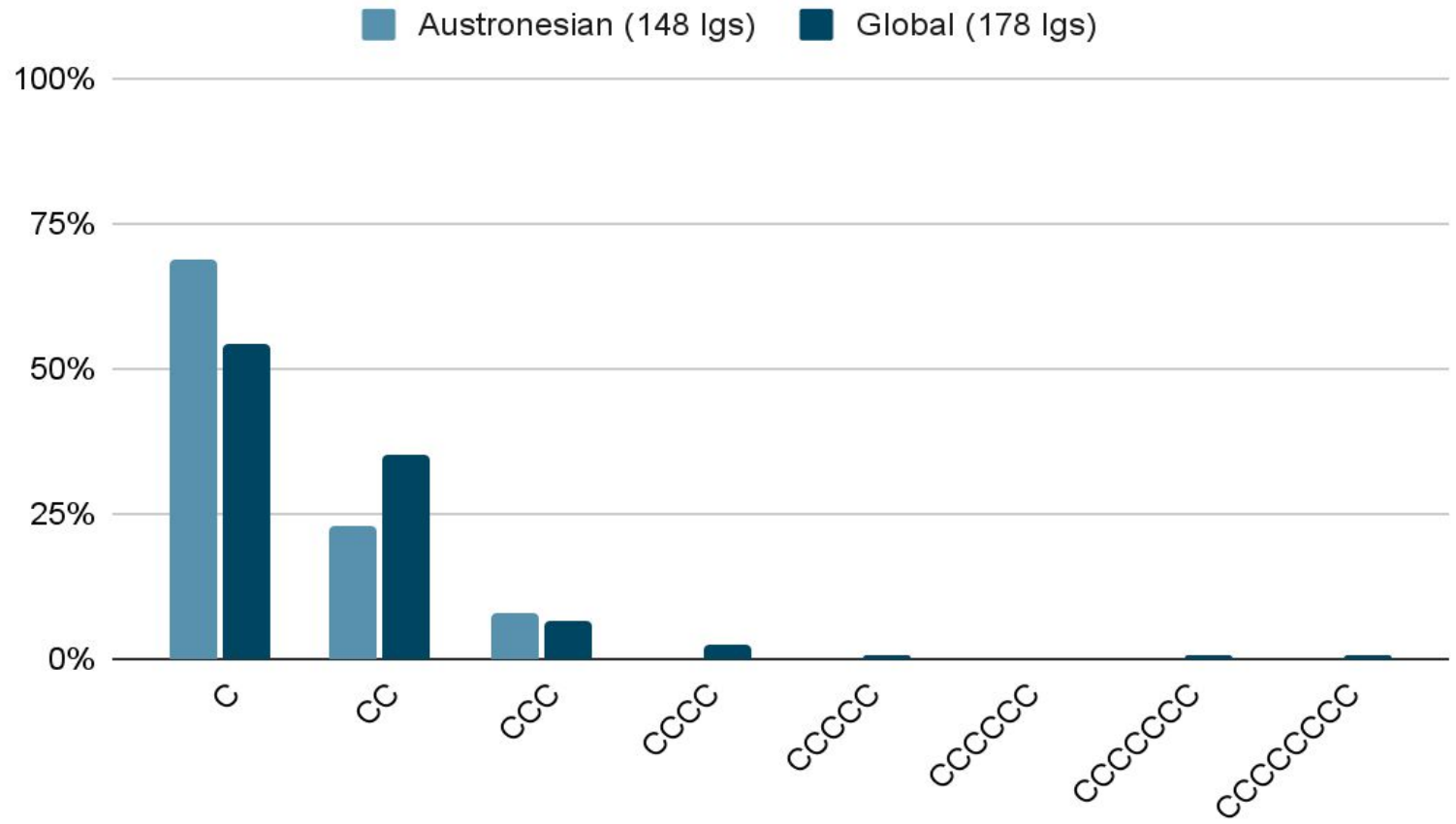
*“The most common syllable structures are **(C)V** and **(C)V(C)**.”*

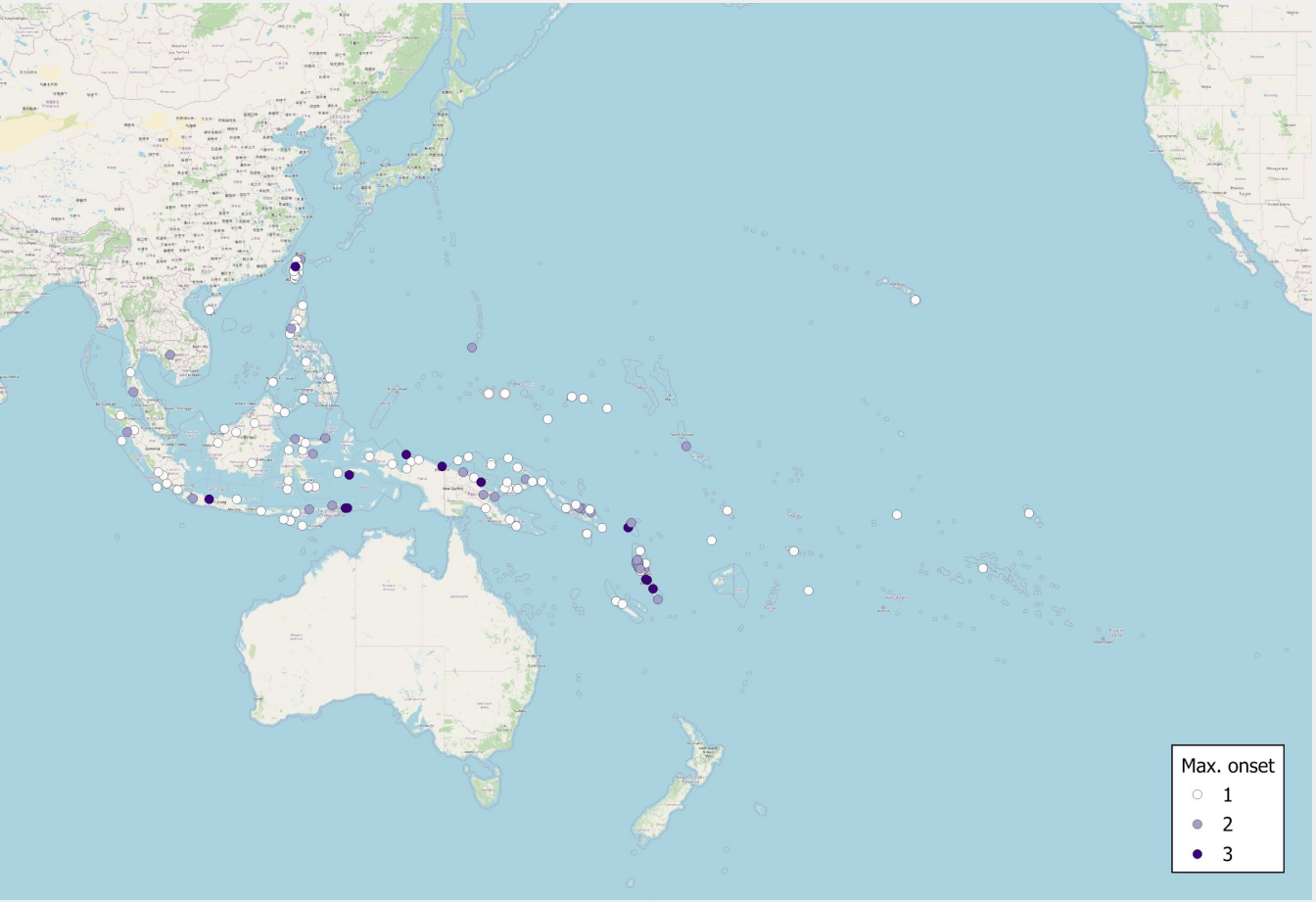
(Adelaar & Himmelmann 2005: 115)

Maximal onset size



Maximal onset size

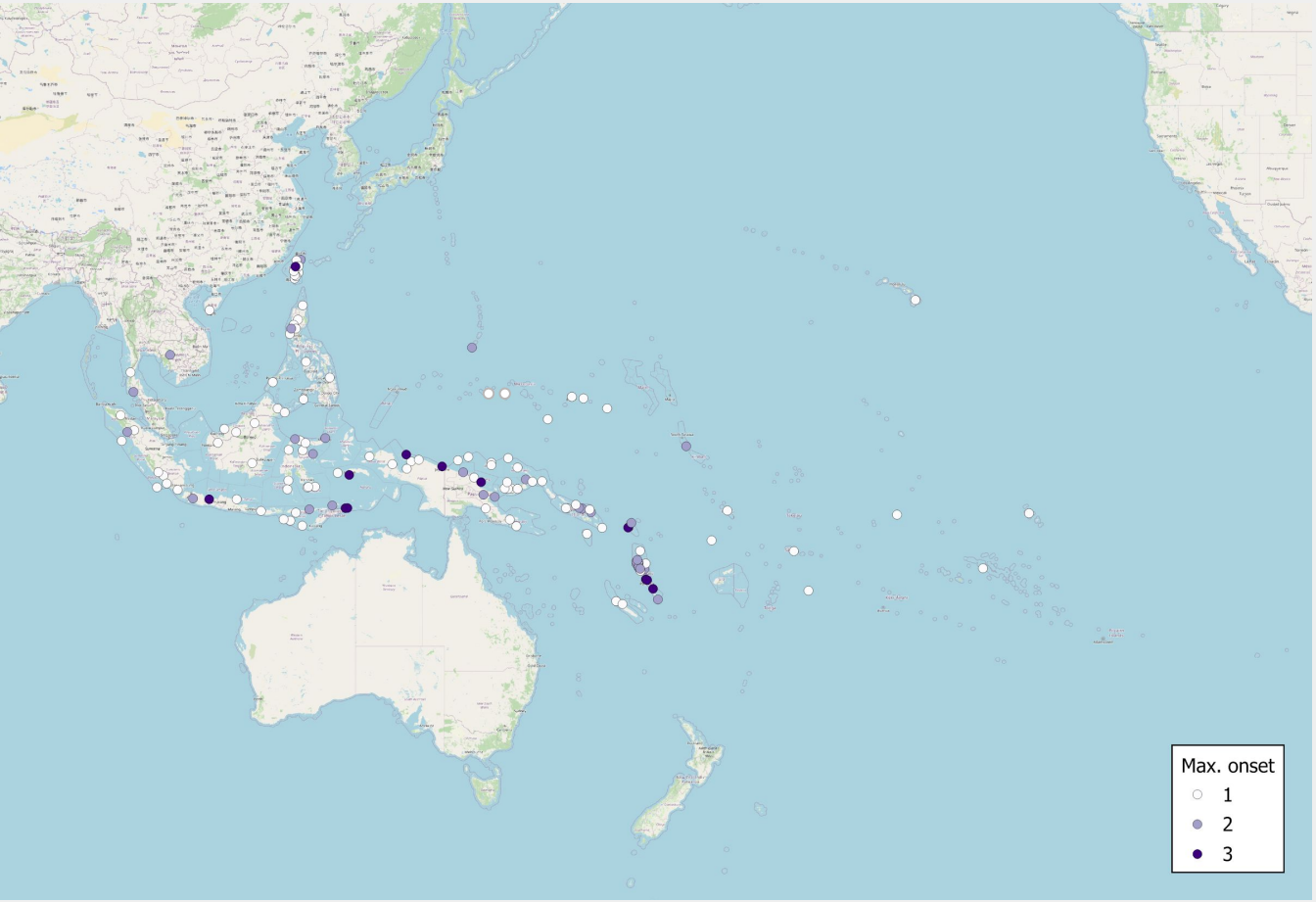




Map made using QGIS 3.36.3; basemap OpenStreetMap

Areas where onsets are almost exclusively simple:

- Philippines
- Borneo
- New Britain
- Polynesia



Map made using QGIS 3.36.3; basemap OpenStreetMap

Languages with **CCC** onsets (12/148) are concentrated in Vanuatu and scattered elsewhere.

e.g. Sie

/**ntru**/ ‘loya cane’

(Crowley 1998: 20)

e.g. Luang

/**tnjamni**/ ‘grave’

(Taber & Taber 2015: 17)

Results: onset patterns

Previous claim:

*“A fair number of languages, including [...] many Philippine languages [...] have **mandatory onsets**.”*

(Adelaar & Himmelmann 2005: 117)

e.g. Ilocano

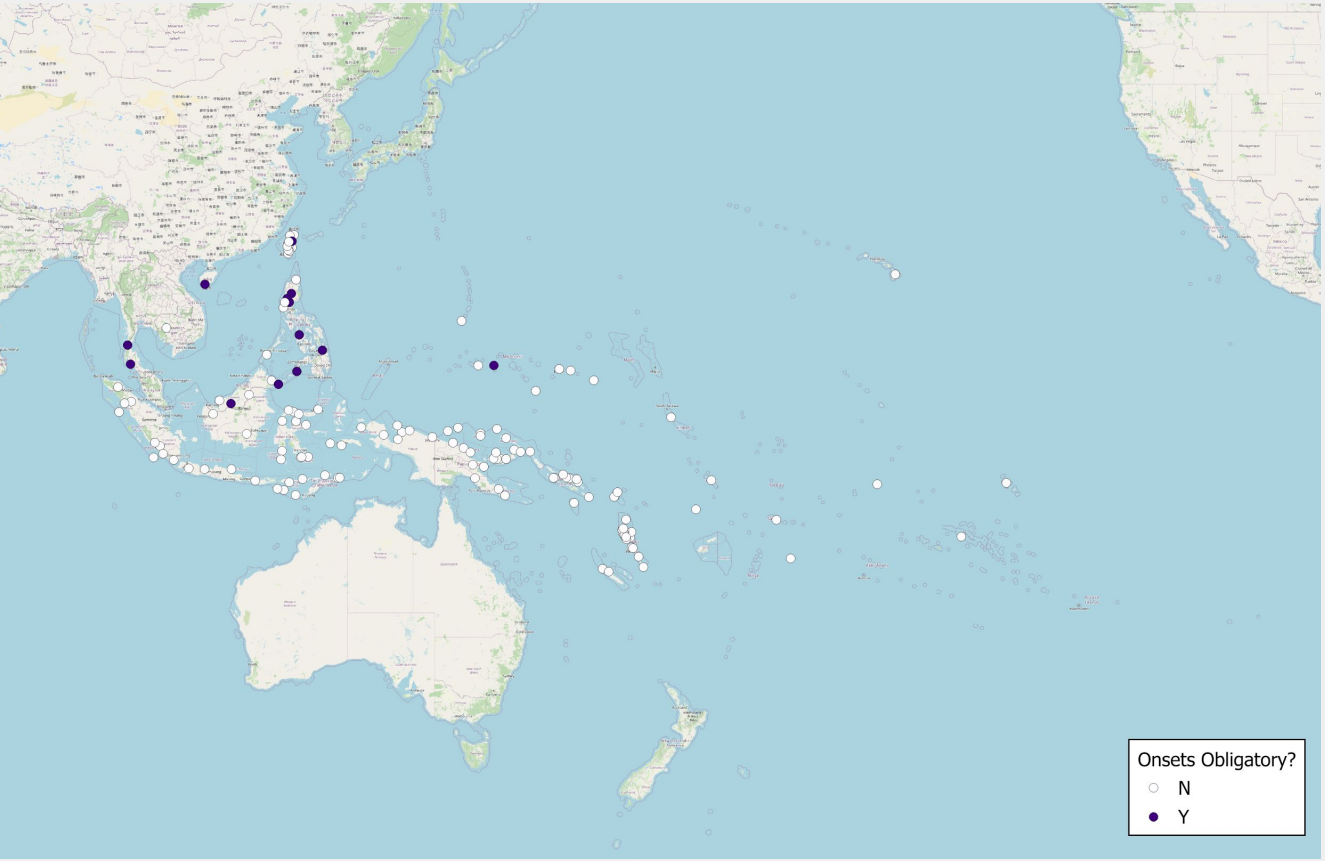
“Every syllable in Ilocano is composed of a consonantal onset and vowel, with an optional consonantal coda.”

/ʔa.rak/ ‘wine’

/ʔag.sa.ŋit/ ‘to cry’

/na.sam.ʔit/ ‘sweet’

(Rubino 1997: 28)



Map made using QGIS 3.36.3; basemap OpenStreetMap

Obligatory onsets are a minority pattern: only **14/148 lgs** show this feature.

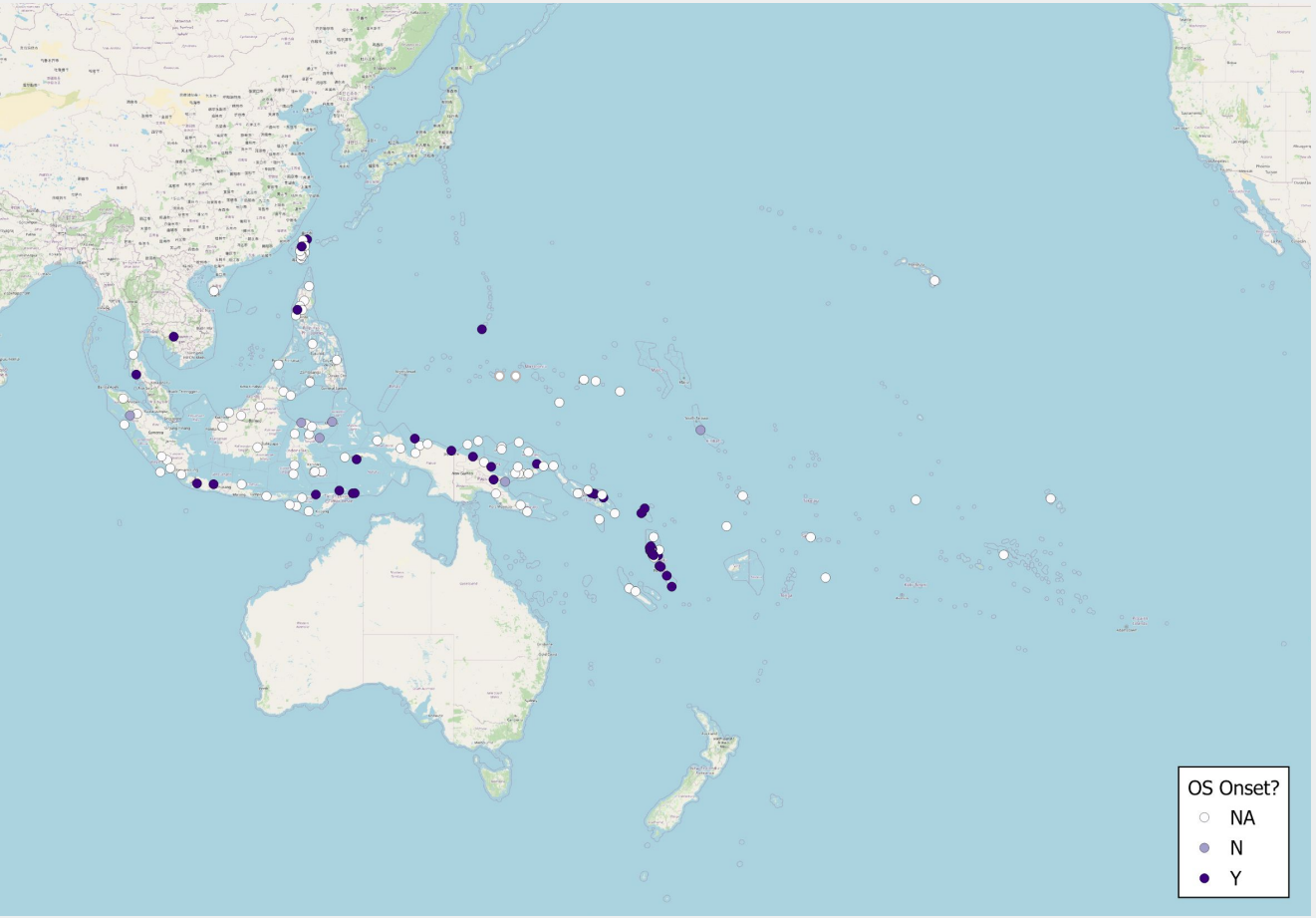
*(Languages with complex onsets are much more frequent at **46/148**.)*

Results: onset patterns

Previous claim:

*"Syllable-internal consonant clusters are typically restricted to onset position and usually consist of **nasal plus obstruent** or **obstruent plus glide or liquid**."*

(Adelaar & Himmelmann 2005: 115)



Map made using QGIS 3.36.3; basemap OpenStreetMap

40/46 lgs with complex onsets have the shape **OS** (obstruent-sonorant).

e.g. CHamoru

/**sjenti**/ ‘feel’

(Chung 2020: 654)

e.g. Nese

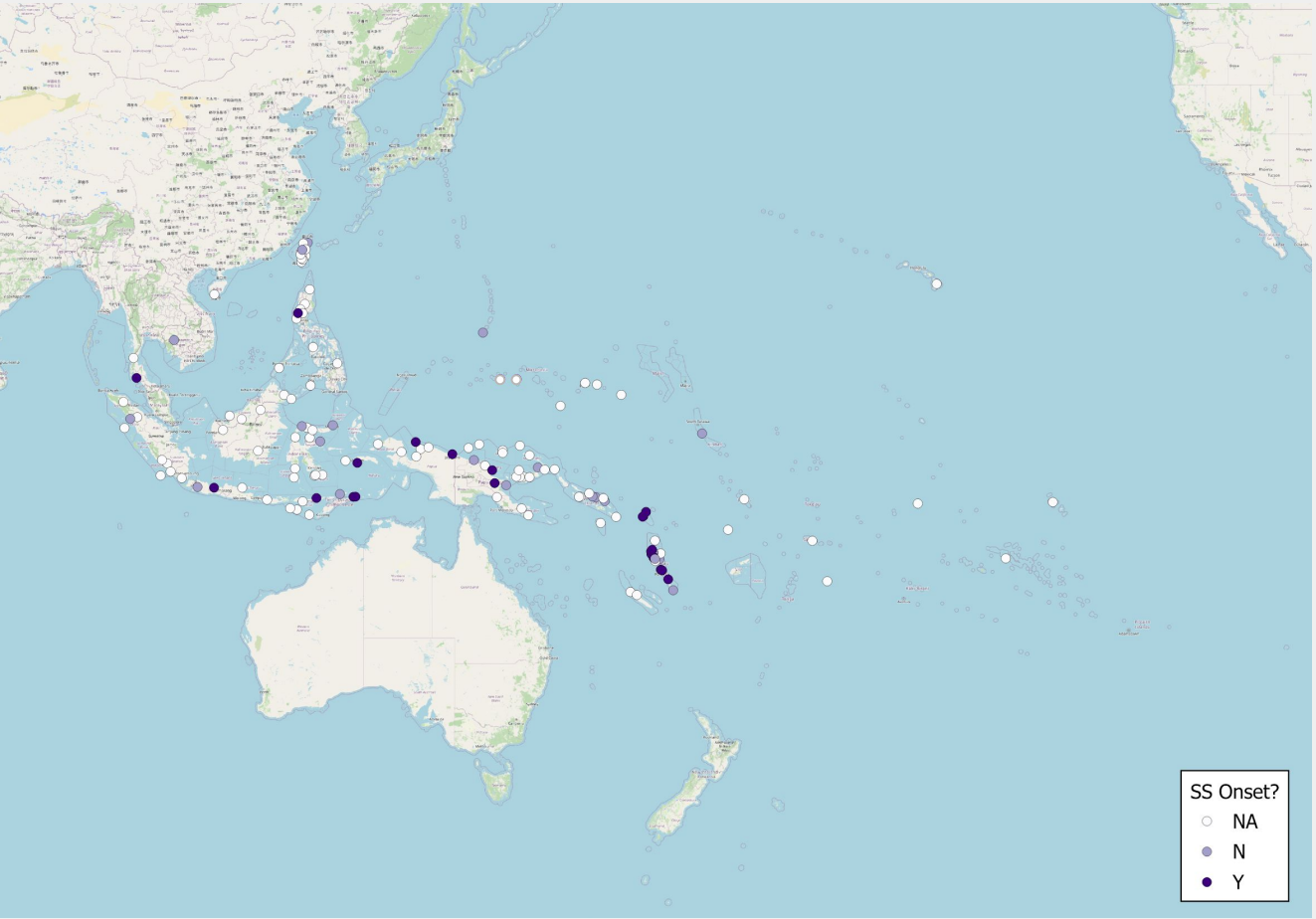
/**tro**/ ‘stand’

(Takau 2016: 65)

e.g. Lamaholot

/**blaha**/ ‘long’

(Kroon 2016: 264)



Map made using QGIS 3.36.3; basemap OpenStreetMap

23/46 lgs with complex onsets have the shape **SS** (sonorant-sonorant).

e.g. Nanggu

/**nj**ɔ/ ‘my (CL.V)”

(Vaa 2013: 112)

e.g. Tobati

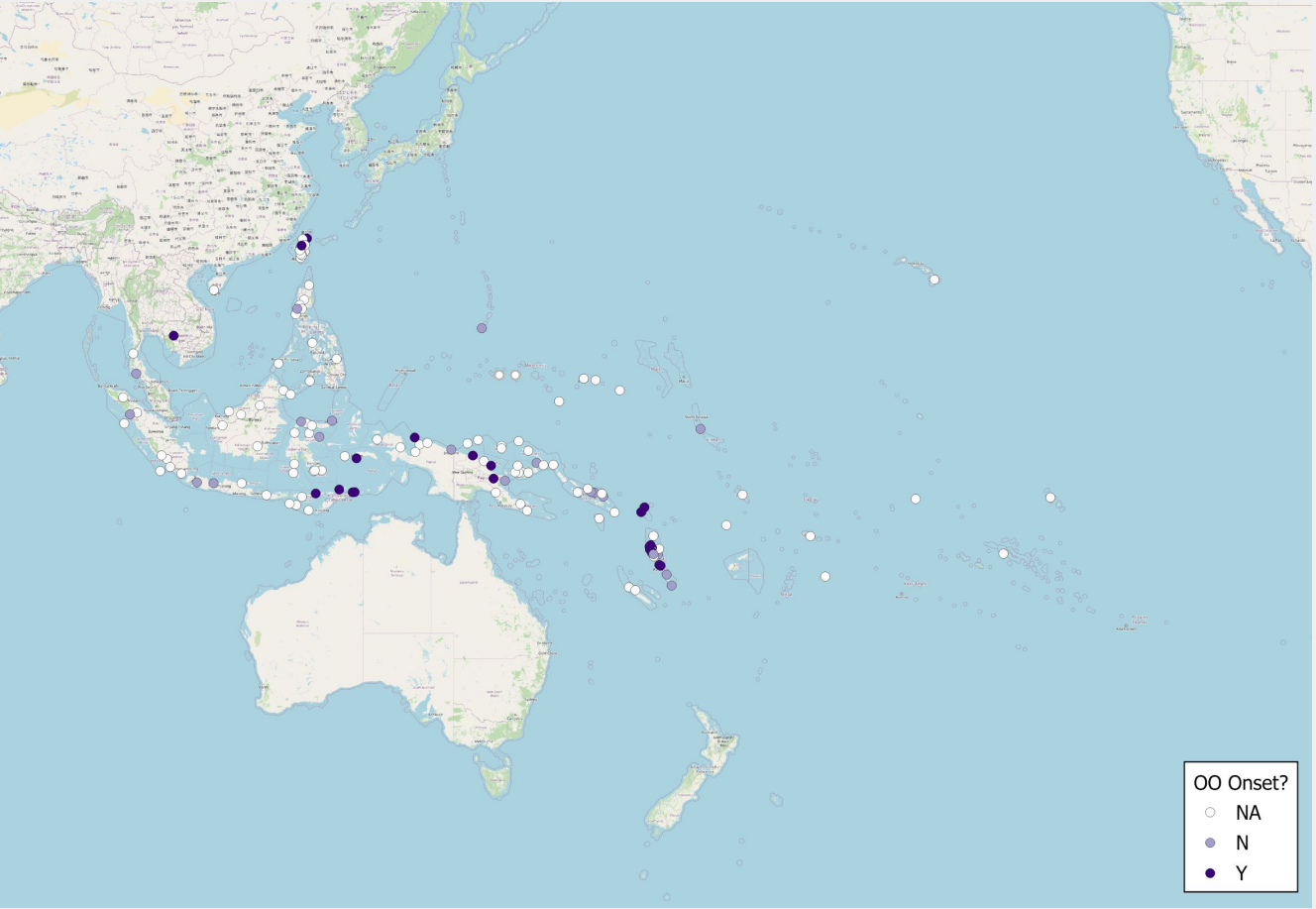
/**rw**ador/ ‘six’

(Donohue 2002: 189)

e.g. Urak Lawoi’

/**ml**upatʃ/ ‘jump’

(Saengmani 1979: 41)



Map made using QGIS 3.36.3; basemap OpenStreetMap

22/46 lgs with complex onsets have the shape **OO** (obstruent-obstruent).

e.g. Leti

/**pt**una/ ‘star’

(Van Engelenhoven 2004: 67)

e.g. Thao

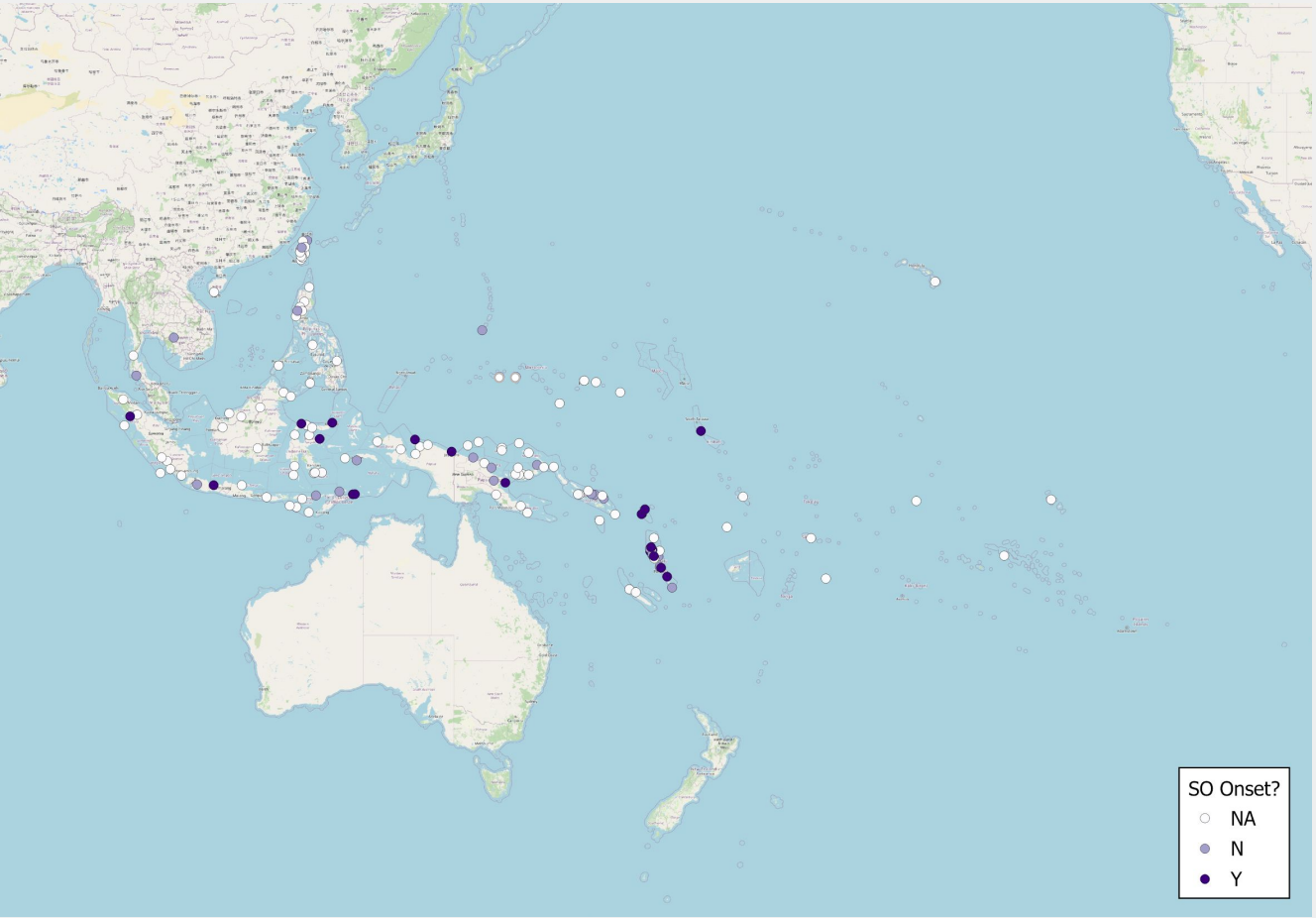
/**qt**iła/ ‘salt’

(Blust 2003: 20)

e.g. Lelepa

/**sk**ei/ ‘INDEF’

(Lacrampe 2014: 42)



Map made using QGIS 3.36.3; basemap OpenStreetMap

19/46 lgs with complex onsets have the shape **SO** (sonorant-obstruent).

e.g. Biak

/**mk**un/ ‘little’

(van den Heuvel 2006: 38)

e.g. Sakao

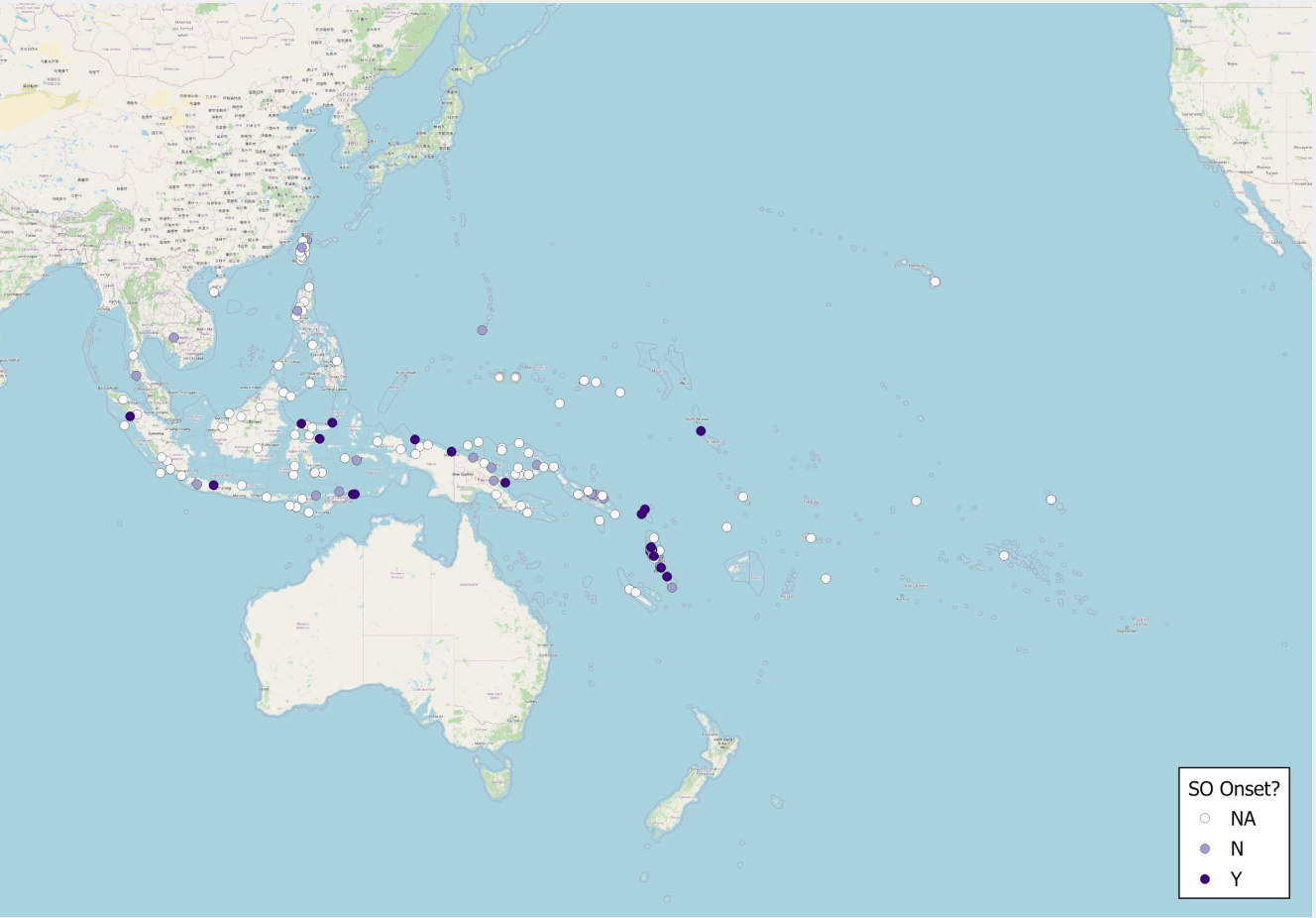
/**rt**ateʁ/ ‘my sisters’

(Touati 2014: 73)

e.g. Gilbertese

/**ŋk**e/ ‘when (PAST)’

(Groves et al. 1985: 18)



Map made using QGIS 3.36.3; basemap OpenStreetMap

6 languages in the sample were reported to have **only SO shapes**:

- Balantak
- Batak Karo
- Gilbertese
- Tondano
- Totoli
- Yabem

CC onset shape	**Austronesian** (46 lgs)	**Global** (78 lgs)
OS	40 lgs *(87%)*	74 lgs *(95%)*
SS	23 lgs *(50%)*	45 lgs *(58%)*
OO	22 lgs ***(48%)***	31 lgs ***(40%)***
SO	19 lgs ***(41%)***	21 lgs ***(27%)***

In comparison to the global sample, Austronesian languages are (somewhat) **more likely to have obstruent-final CC onsets,** and (somewhat) **less likely to have sonorant-final CC onsets.**

Globally, **obstruent-final CC onsets** are more likely to be found in languages with maximal onsets of **3 Cs or more**. In Austronesian, these are usually found in languages with maximal onsets of **2 Cs**.

There are **9 languages** with all shapes (OS, OO, SO, and SS):

Vanuatu and Santa Cruz Islands

- Araki
- Axamb
- Nafsan
- Nalögo
- Vaeakau-Taumako
- Wanohe

Maluku

- Leti
- Luang

New Guinea

- Biak

In Vanuatu, the deletion of unstressed interconsonantal vowels, often ***high vowels in pretonic position***, has led to the historical emergence of diverse onset cluster types:

e.g. Nanggu

POc ‘eye’	pre-PRSC	PRSC	Nanggu
*mata	*mala	*na mnɒ	mnɔ

(Vaa 2013: 105; Ross & Næss 2007: 467)

e.g. Merei /ˈtlui/ ~ Tiale /tuˈlui/ ‘pull’

Merei /ˈlmana/ ~ Tiale /liˈmana/ ‘his/her hand’

(Chung 2005: 8)

Similar ***optional processes*** are reported to operate synchronically in Nanggu, Araki, Mavea, and Lelepa.

Results: coda patterns

Previous claims:

*"The most common syllable structures are **(C)V** and **(C)V(C)**."*

(Adelaar & Himmelmann 2005: 115)

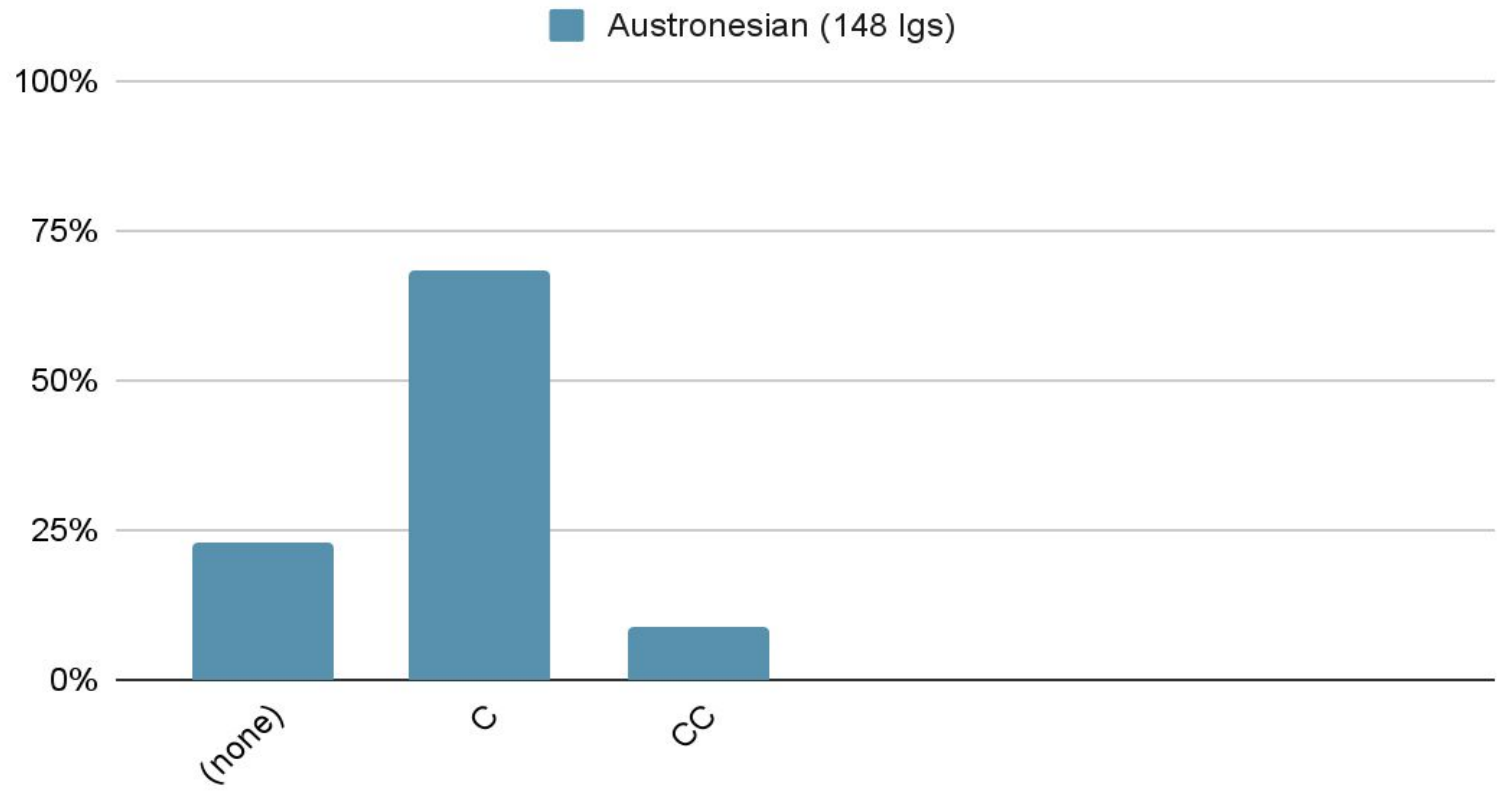
*"Syllable structures tend to approximate a simple **CV** type."*

(Lynch, Ross, & Crowley 2002: 34)

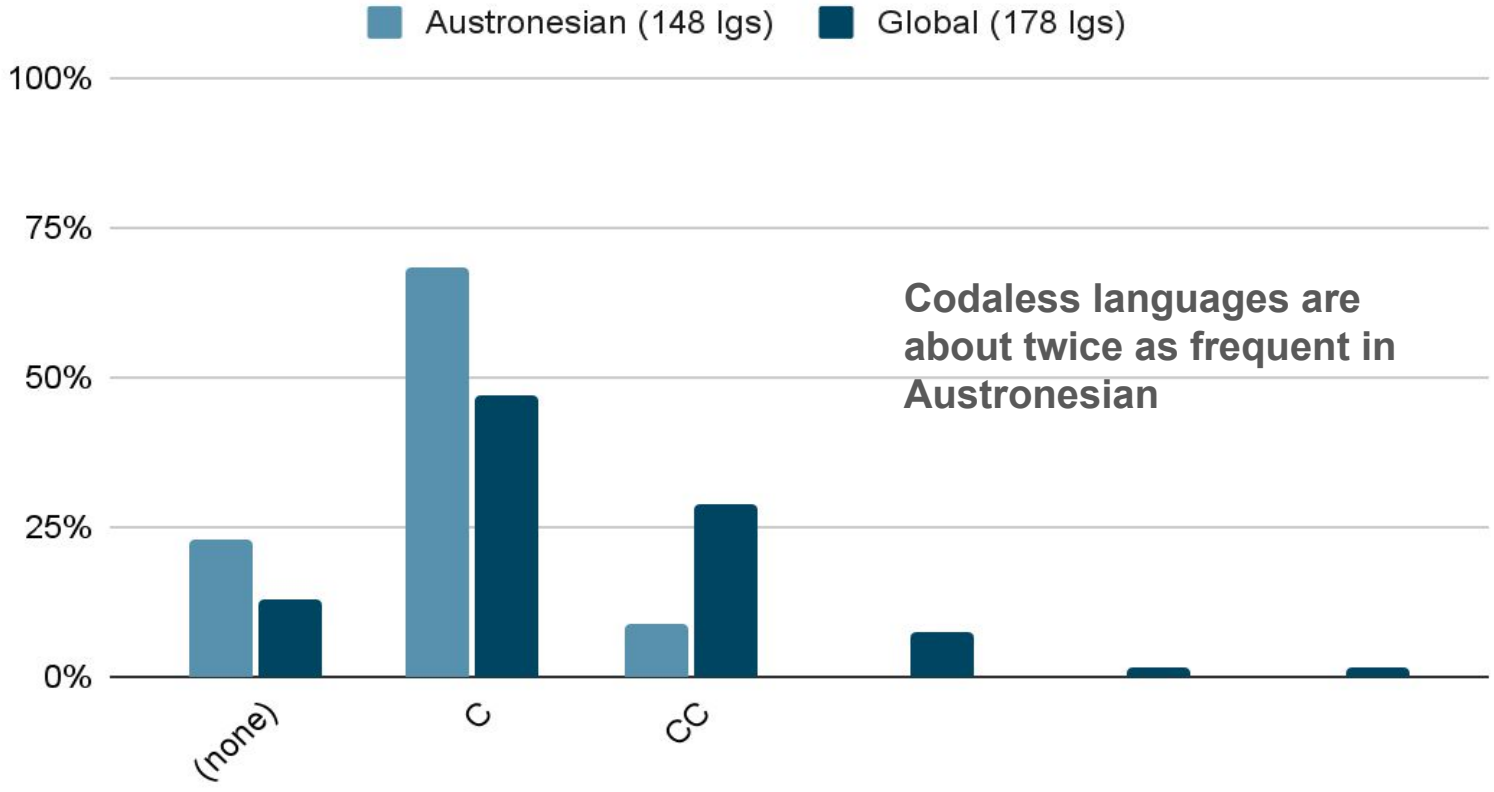
"Syllable-internal consonant clusters are typically restricted to onset position..."

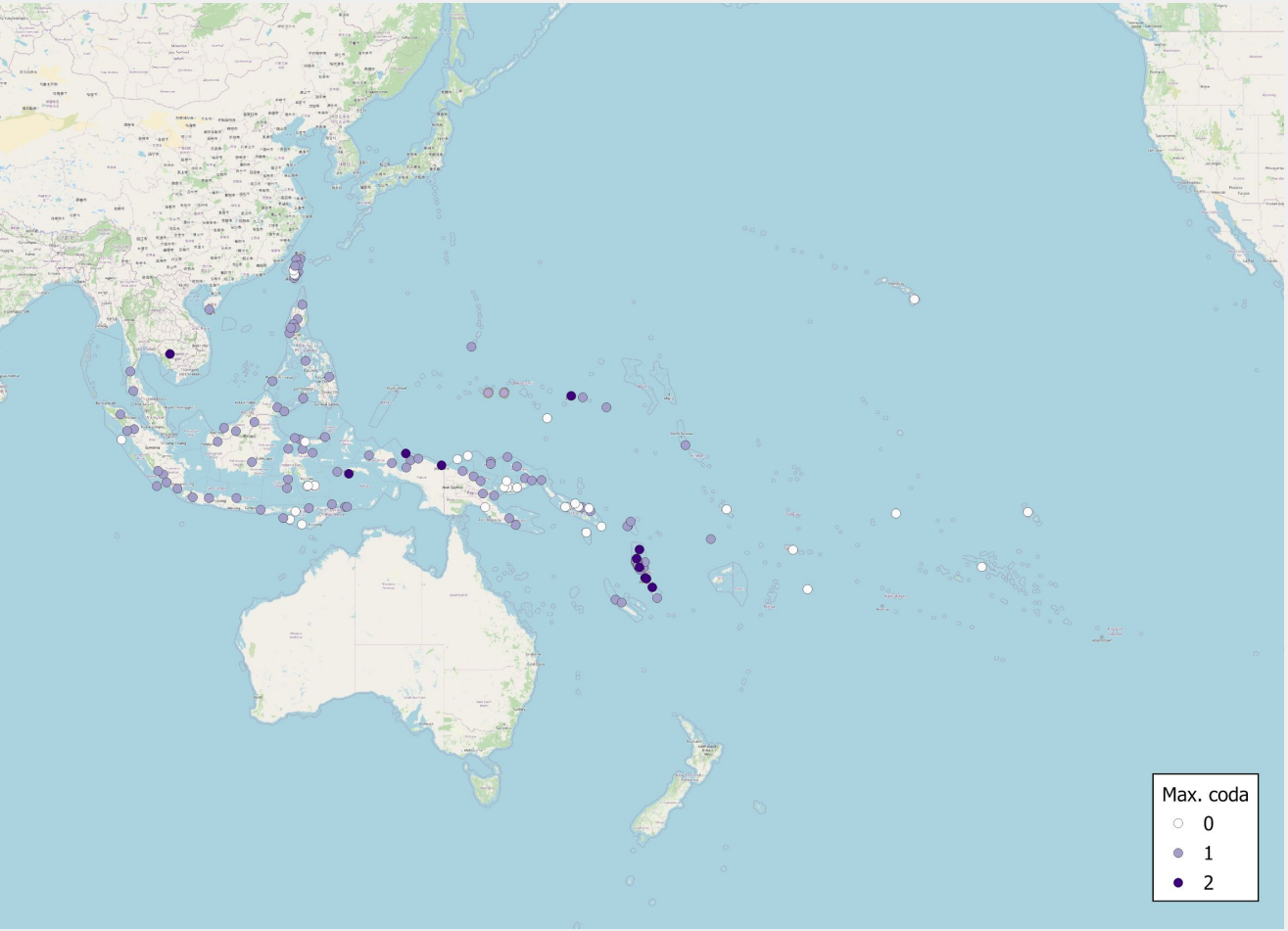
(Adelaar & Himmelmann 2005: 115)

Maximal coda size



Maximal coda size





Map made using QGIS 3.36.3; basemap OpenStreetMap

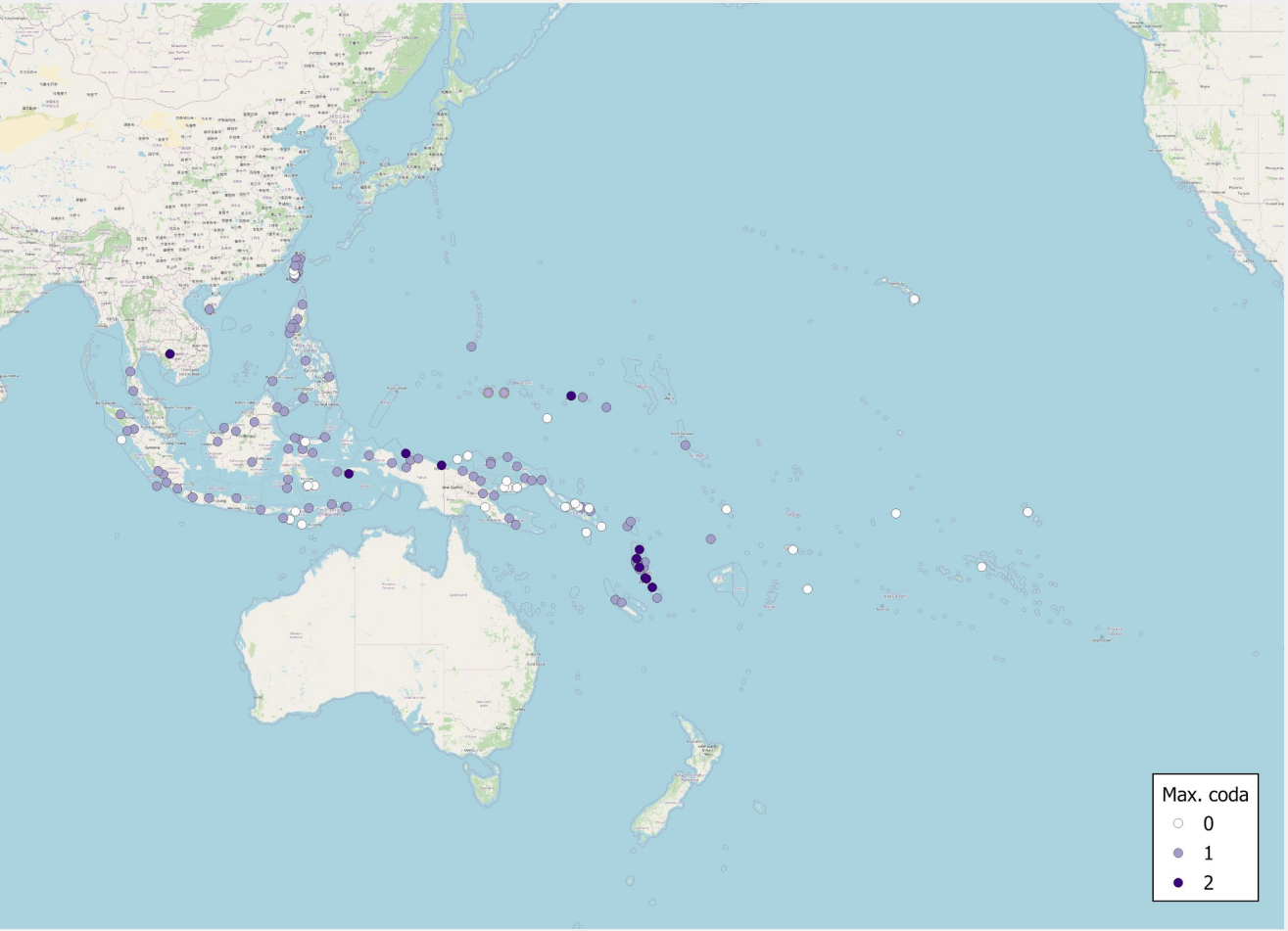
Languages with codas predominate (**114/148**, or **77%** of languages).

Solomons, New Britain, and Polynesia are codaless hotspots.

Complex codas are heavily concentrated in Vanuatu. They tend to be substantially more restricted than complex onsets.

e.g. Nafsan complex coda inventory /**lf rk**/

(Thieberger 2004: 63)



Map made using QGIS 3.36.3; basemap OpenStreetMap

In the sample, we observe codas emerging from word-final unstressed vowel reduction and deletion, often of ***high vowels***.

e.g. Nanggu

/**u**/ is particularly subject to weakening after oral and nasal stops:

/**dɔn(u)**/ ‘here’

Older speakers:

[ˈ**dɔnu̯**]

Younger speakers:

[ˈ**dɔn**]

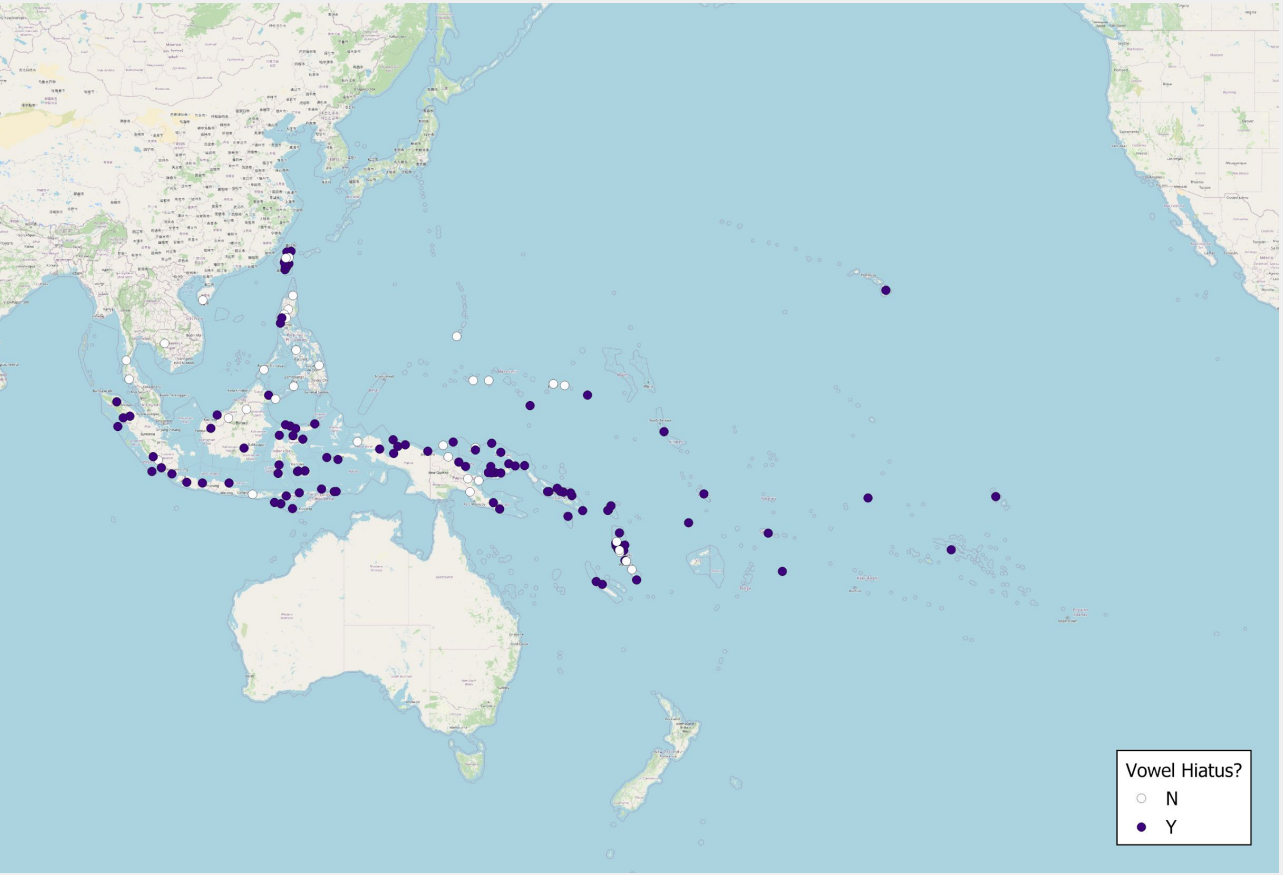
(Vaa 2013: 119)

Results: vowel hiatus

Vowel hiatus is significantly more likely to occur in languages with canonical **(C)V** structure, globally:

- $p < .001$ in 100 language sample stratified for syllable complexity (Easterday 2019)

Since canonical **(C)V** structure is more common in Austronesian than it is globally, we'd expect vowel hiatus to be widespread in the family.

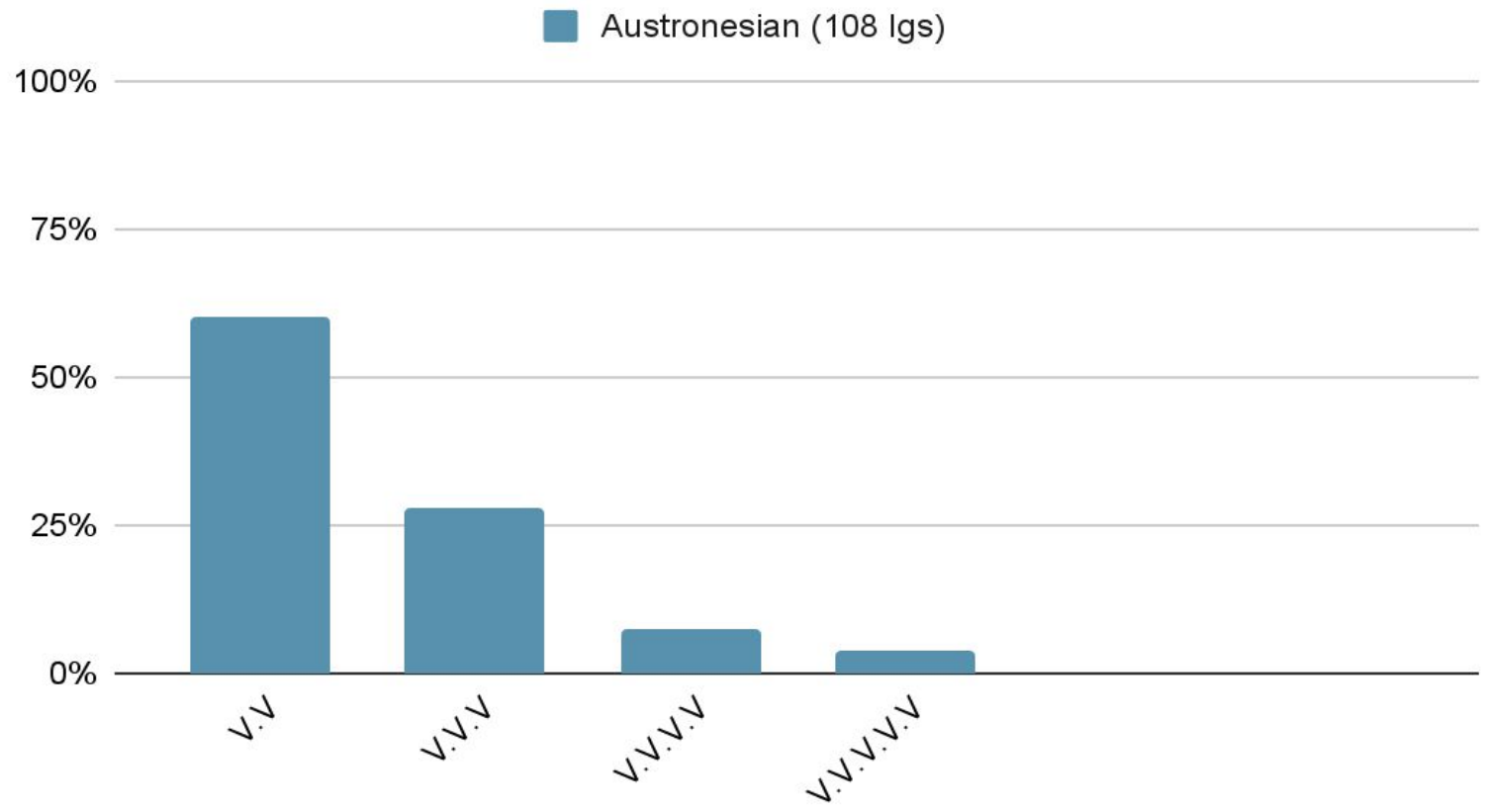


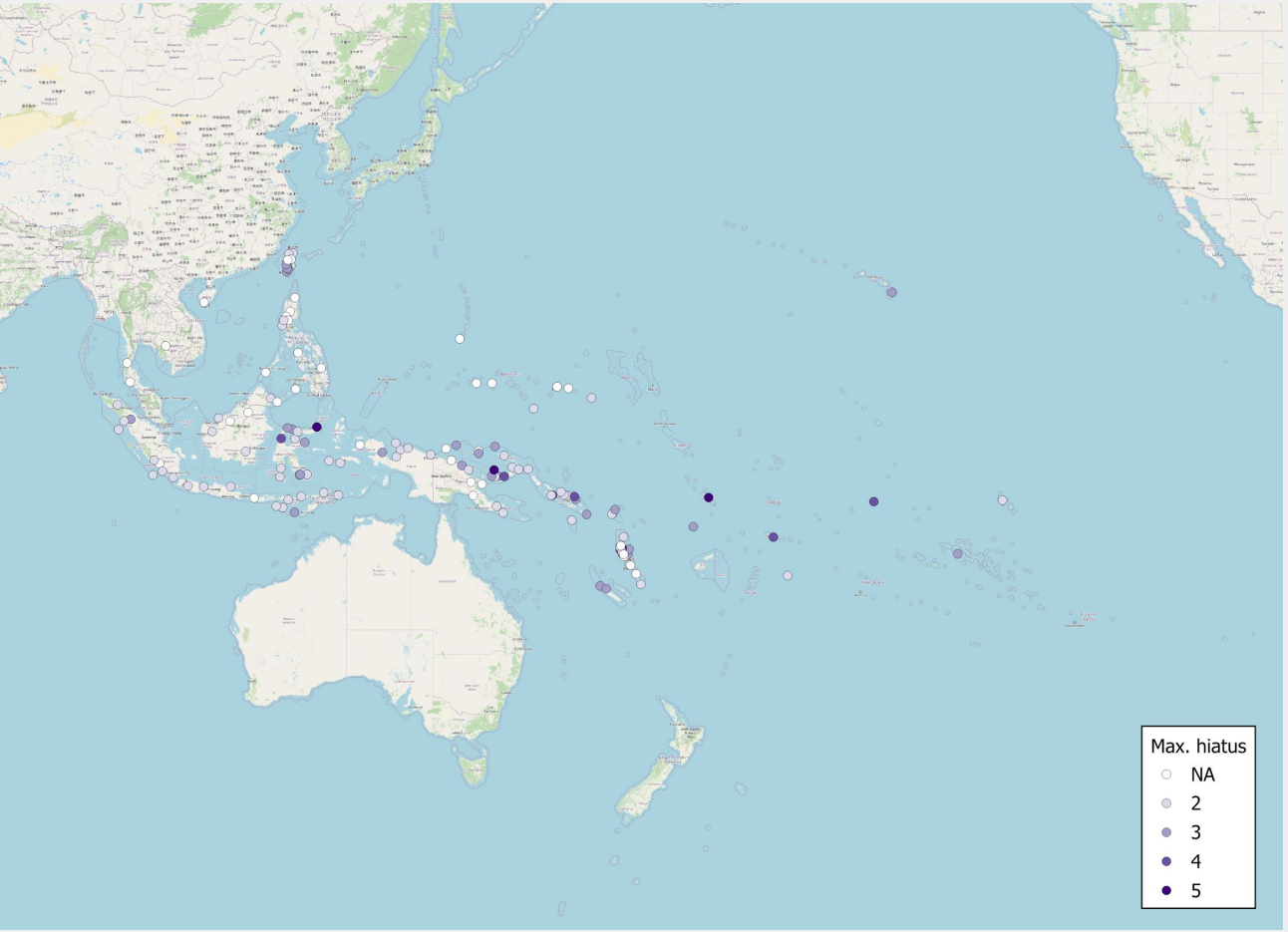
Map made using QGIS 3.36.3; basemap OpenStreetMap

Vowel hiatus within words is widespread outside Philippines, where onsets are usually obligatory.

This pattern is ubiquitous in Polynesia, Solomons, and New Britain, which are all also notable codaless regions, as well as Sulawesi.

Maximal sequence of vowels in hiatus





Map made using QGIS 3.36.3; basemap OpenStreetMap

Extreme manifestations of hiatus (4 or 5 vowels) are reported for languages in Polynesia, New Britain, Vanuatu, and Sulawesi.

e.g. Nakanai

'Whole word [and initial] clusters of two to four vowels may occur.'

/**eiau**/ 'I'

(Johnston 1980: 254)

e.g. Tondano

'[O]ne sequence of five vowels [has] been recorded.'

/**maoaoas**/

'is continually washing'

(Sneddon 1975: 26)

Discussion

Austronesian phonotactics are not uniform.

- Within-family variation tends to cluster in geographical hotspots according to the feature, but not exclusively, and most patterns show some scatter.

Austronesian phonotactics are not inconspicuous from a crosslinguistic perspective.

- Higher prevalence of sonority reversals and plateaus in biconsonantal onsets.
- Higher rates of simple syllable structure and vowel hiatus.

Mahalo nui loa!

References

Adelaar, Alexander and Nikolaus P. Himmelmann (eds.). 2005. The Austronesian languages of Asia and Madagascar. Routledge.

Bickel, Balthasar 2015. Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In Bernd Heine & Heiko Narrog (eds.), The Oxford Handbook of Linguistic Analysis, 2nd edition, Oxford: Oxford University Press.

Blust, Robert. 2003. Thao Dictionary. (Language and Linguistics Monograph Series, A5.) Taipei: Academia Sinica, Institute of Ethnology.

Blust, Robert. 2013. The Austronesian languages. Asia-Pacific Linguistics, School of Culture, History and Language, College of Asia and the Pacific, The Australian National University.

Chung, Sandra. 2020. Chamorro grammar. Santa Cruz: University of California.

Crowley, Terry. 1998. An Erromangan (Sye) Grammar. (Oceanic Linguistics Special Publication, 27.) Honolulu: University of Hawaii Press.

Donohue, Mark. To appear. Malayo-Polynesian languages of Southeast Asia: segmental phonologies. In Alexander Adelaar and Antoinette Schapper (eds.), The Oxford Guide to the Malayo-Polynesian Languages of Southeast Asia.

Donohue, Mark. 2002. Tobati. In John Lynch and Malcolm Ross and Terry Crowley (eds.), The Oceanic Languages, 186-203. Richmond: Curzon.

Easterday, Shelece. 2019. Highly complex syllable structure: A typological and diachronic study. Studies in Laboratory Phonology 9. Berlin: Language Science Press.

Easterday, Shelece. To appear. Syllable structure. In David Inman, Natalia Chousou-Polydouri, Marine Vuillermet, Kellen Parker van Dam, Shelece Easterday, Françoise Rose, Alena Witzlack-Makarevich, Kevin M. Bätscher, Oscar Cocaud-Degrève, Anna Graff, Selma Hardegger, Tai Hong, Thomas C. Huber, Diana Krasovskaya, Raphaël Luffroy, Nora Muheim, André Müller, Alexandra Nogina, David Timothy Perrot, & Balthasar Bickel (eds.), The ATLAs database: Areal Typology of the Languages of the Americas.

Engelenhoven, Aone van. 2004. Leti: A Language of Southwest Maluku. (Verhandelingen van het Koninklijk instituut voor taal-, land- en volkenkunde, 211.) Leiden: KITLV Press.

Groves, Terab'ata R., Gordon W. Groves & Roderick Jacobs. 1985. Kiribatese: An Outline Description. (Pacific Linguistics: Series D, 64.) In Terab'ata R. Groves and Gordon W. Groves and Roderick Jacobs (eds.) Canberra: Research School of Pacific and Asian Studies, Australian National University.

Guérin, Valérie M. P. R. 2011. A grammar of Mavea: An Oceanic language of Vanuatu. (Oceanic Linguistics Special Publication, 39.) Oceanic Linguistics. Honolulu: University of Hawai'i Press.

Mahalo nui loa!

References

Inman, David, Natalia Chousou-Polydouri, Marine Vuillermet, Kellen Parker van Dam, Shelece Easterday, Françoise Rose, Alena Witzlack-Makarevich, Kevin M. Bätscher, Oscar Cocaud-Degrève, Anna Graff, Selma Hardegger, Tai Hong, Thomas C. Huber, Diana Krasovskaya, Raphaël Luffroy, Nora Muheim, André Müller, Alexandra Nogina, David Timothy Perrot, & Balthasar Bickel (eds.). To appear. The ATLAs database: Areal Typology of the Languages of the Americas.

Heuvel, Wilco van den. 2006. Biak: Description of an Austronesian language of Papua. 138. Utrecht: Vrije Universiteit Amsterdam dissertation.

Johnston, Raymond L. 1980. Nakani of New Britain: The grammar of an oceanic language. (Pacific Linguistics B, 70.) Canberra: Australian National University.

Kroon, Yosep Bisara. 2016. A grammar of Solor-Lamaholot: A Language of Flores, Eastern Indonesia. Adelaide: University of Adelaide dissertation.

Lacrampe, Sébastien. 2014. Lelepa: Topics in the grammar of a Vanuatu language. Australian National University dissertation.

Lynch, John, Malcolm Ross, and Terry Crowley (eds.). 2002. The Oceanic languages. Curzon.

Ross, Malcolm, & Åshild Næss. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands. *Oceanic Linguistics* 46(2): 456-48.

Rubino, Carl Ralph Galvez. 1997. A Reference Grammar of Ilocano. Ann Arbor: University of California at Santa Barbara dissertation.

Saengmani, Amon. 1979. Phonology of the Urak Lawoi' language: Adang island. Bangkok: Mahidol University MA thesis.

Sneddon, James N. 1975. Tondano Phonology and Grammar. (Pacific Linguistics: Series B, 38.) Canberra: Research School of Pacific and Asian Studies, Australian National University.

Taber, Kathleen B. & Mark H. Taber. 2015. Luang Grammar and Phonology Sketch. (SIL e-Books, 63.) [Dallas]: SIL International.

Takau, Lana Grelyn. 2016. A grammar of Nese. Australia: University of Newcastle dissertation.

Touati, Benjamin. 2014. Description du sakao, langue océanienne du nord-est Santo (Vanuatu). Université de la Sorbonne (Paris IV) dissertation.

Vaa, Anders. 2013. A Grammar of Engdewu: An Oceanic language of Solomon Islands. University of Oslo dissertation.