Benjamin Macaulay*

# Speaker judgments alone cannot diagnose syllable structure

**Abstract:** The ease and consistency with which speakers of many languages provide direct judgments about syllable structure has been taken by scholars as evidence that these judgments are accurate and sufficient argumentation for an analysis of syllable structure in descriptive works. This paper questions whether the results of direct elicitation tasks reliably indicate a prosodic domain that is meaningful in the language's phonology and provides alternative forms of evidence for syllable structure from intonation. This is done through a case study of Budai Rukai, a Formosan language whose contact relationship with Sinitic languages affects how speakers respond to syllable judgment tasks.

**Keywords:** syllable structure; intonation; language documentation; Austronesian; metalinguistic awareness

## 1 Introduction

Grammars, phonology sketches, and other descriptive linguistic works nearly always include some analysis of the target language's syllable structure. This may take the form of a dedicated section of the work or simply be implied by the existence of syllable boundaries in phonological transcriptions or other indirect mentions to syllable breaks, onsets, codas, and the like. What is often missing from these works is the argumentation behind how the author(s) decided on the model of syllable structure they present: how was syllable structure elicited, what other aspects of the phonology corroborate the model, and what competing hypotheses were ruled out? This type of information is generally provided for other phonological structures, for example, phoneme contrasts are routinely argued for in descriptive works with the use of minimal pairs.

    Often the argumentation behind a work's model of syllable structure is not made explicit because speakers are able to give direct judgments about syllable structure in their language. As these judgments converge across speakers, the judgments are assumed to be reliable and are presented "as is" without corroborating evidence or argumentation. However, the syllable is not only active in these types of parsing

**\*Corresponding author: Benjamin Macaulay**, Lund University, Box 201, 221 00 Lund, Sweden, E-mail: benjamin.macaulay@englund.lu.se

tasks. There are also countless phonological rules that reference syllable structure across the world's languages. These phonological rules may also serve as diagnostics for syllable structure and crucially, may point to a different definition of the syllable than is found via direct elicitation. I argue that in the case of such a clash, it is the syllable as defined by the language's phonological rules that is the "true" syllable, and that language-external factors shape the divergent results of direct elicitation tasks. Nevertheless, it is these divergent results that are reported in the literature, due to their presumed reliability. These mischaracterized descriptions of syllable structure are then passed on to typologists, historical linguists, and other scholars in theoretical linguistics without being questioned.

I will illustrate these claims via a case study of Budai Rukai, a Formosan language (i.e., an Austronesian language of Taiwan). Previous scholars of the language have mostly converged on a model of the prosodic structure of words in the language. This includes description of two levels of representation in particular: the mora, which takes the structure (C)V, and the syllable, which is maximally either CVV (Chen 2006) or CVVV (Liu 2011), where V represents a vocoid (either a vowel or a glide). I have replicated their model of the "syllable" in a direct elicitation task but also find evidence from work on Budai Rukai's intonational system that it is a smaller (C)V domain (i.e., the previously reported "mora") that is referenced by the language's syllable-level phonology. I argue that the larger CVV(V) "syllables" that are given in direct elicitation tasks are an artifact of field methodology, language contact, and education, rather than a domain relevant to the language's phonological processes.

The extralinguistic factors that affect how speakers respond to syllable judgment tasks are by no means limited to Budai Rukai. In fact, the interference from contact languages and education leave endangered and Indigenous languages particularly vulnerable to having their syllable structures misreported due to reliance on direct elicitation tasks. As these are precisely the languages for which accurate language documentation is most crucial, it is critical that direct elicitation be supplanted (or at least supported) by more reliable diagnostics for syllable structure in the fieldwork process. In the case of Budai Rukai, it is intonational phonology that provides the best insight into the language's true syllable structure. Given how many languages lack documentation of their intonation entirely, it is likely that many intonational patterns that challenge other languages' reported syllable structures are yet to be discovered.

The paper is structured as follows: Section 2 will describe the role of direct elicitation in linguistic fieldwork generally, and how it affects the content of published descriptive works; Section 3 will provide an introduction to the two prosodic domains under discussion, the mora and the syllable, as well as the kinds of metalinguistic knowledge that speakers may have about them; Section 4 will give an overview of Budai Rukai phonology, the existing descriptions of the language's syllable structure, and a reanalysis based on novel data primarily from intonational

phonology; Section 5 will discuss the direct elicitation of syllable structure specifically in Budai Rukai, as well as evidence of interference from contact languages; and Section 6 will comment on the scope of the problem generally, with reference to a typological study of the Formosan languages that concluded in many other similar cases of reanalyzed syllable structure. Section 7 will conclude.

# 2 The role of native speaker intuitions in phonological study

Many structures in phonology are not accessible to speakers. One well-known type of example is the "near merger," a situation in which speakers reliably produce a measurable phonetic distinction between two words but claim that the words sound identical, such as the pair *source* and *sauce* in New York City English as described by Labov et al. (1972).

On the other hand, there are phonological structures that speakers provide direct intuitions about. When there is speaker agreement (however, small the sample may be, including $n = 1$), these intuitions are seen as evidence for the structures they describe. When used as evidence, speaker intuitions can either appear alongside other types of evidence or stand alone as the sole argumentation for an analysis. In the phonology literature, the use of speaker intuitions as evidence is more likely in those topics for which speakers more easily give intuitions. Within these topics, some are more likely than others to rely on intuitions as the sole form of evidence. This section will first discuss the appearance of speaker intuitions alongside other evidence (Section 2.1), followed by a discussion of syllabicity judgments as a type of speaker intuition often presented alongside no corroborating evidence (Section 2.2).

## 2.1 Speaker intuitions as support for a larger analysis

Some topics that make use of speaker intuitions commonly provide them as a final "confirmation" for an analysis based on other types of evidence. Analyses of certain articulatory combinations as complex segments /$\widehat{AB}$/ versus segment sequences /AB/ often contain mention that the final analysis (based primarily on other evidence from distribution, alternations, symmetry in the inventory, tonal alignment, etc.) confirms the intuitions of native speakers.[1] To give some examples, Vance's (2008)

---

**1** It is likely that speaker judgments are used as evidence for monosegmental versus bisegmental representations as native speakers can be asked directly if a stimulus (or part thereof) contains "one or two sounds," which is then interpreted as one versus two segments by the linguist. Given that

analysis of Japanese affricates as single segments, Casali's (1995) analysis of Moghamo prenasalized consonants as single segments, and Zonneveld and Trommelen's (1980) analysis of monosegmental versus bisegmental diphthongs in Dutch all mention that their results reflect the intuitions of native speakers.

These intuitions may be from authors who are native speakers of the language under study, as in the case of Zonneveld and Trommelen (1980), who note the split between monosegmental and bisegmental diphthongs in Dutch as "a fact which we all intuitively accept" (p. 279), seemingly including the two authors themselves as well as unnamed other native speakers in agreement on the intuition.

In other cases, nonspeaker (or non-native speaker) authors cite intuitions from native speakers consulted for linguistic study. These intuitions are presented in a variety of ways and play different roles depending on the study. For example, many mentions of native speaker intuitions are short and presented without detail on the context in which the intuition was provided. Vance (2008: 82–7) bases his analysis of Japanese affricates mainly on distributional evidence but adds that "[a]s far as the intuitions of Japanese native speakers are concerned, there's no doubt that the affricate analysis is right in both cases [palatalized and nonpalatalized voiceless coronals]" (p. 83), and later "[a]s with [the voiceless coronals], the intuitions of Japanese native speakers leave no doubt that [the voiced coronals] should be analyzed as single affricate phonemes" (p. 85). On the other end of the spectrum, Casali (1995: 164) devotes a full section of the paper to native speaker intuitions, presenting information about the speaker, the methodology by which data was obtained, and data sets that resulted from the elicitation.

This is not to say that short or indirect mentions of native speaker intuitions are bad or should in all cases be replaced by more in-depth discussion and data presentation. In all three of the papers mentioned above, native speaker intuitions were presented alongside various other types of evidence. In each case, the other evidence presented would have just as well argued the paper's claims without the addition of the speaker intuitions.

## 2.2 Speaker intuitions as standalone evidence

There are, however, areas of phonology in which native speaker intuitions often appear as the sole evidence for a proposed structure. One topic that is commonly discussed in the phonology literature using speaker intuitions as the sole form of evidence is syllable structure.

---

many such mentions of speaker intuitions do not provide their methodology, this is at best an assumption in many cases.

In addition to being able to parse speech into segments (as in the cases mentioned in Section 2.1), native speakers are often able to provide intuitions on the number of syllables in a word or utterance and to the location of syllable boundaries (Blevins 1995: §1.4, and references therein). These intuitions are often presented under the assumption that they faithfully represent the organization of segments into syllables in the speaker's phonology. As an example, the native speaker intuitions presented by Casali (1995) in his analysis of Moghamo prenasalized stops included judgments on the location of syllable boundaries: the speaker is asked to parse sequences like [sambe] "seven" into syllables and gives the judgment [sa.mbe] rather than *[sam.be], which has a more typical sonority profile. This is taken as support for Casali's analysis of [mb] as a prenasalized stop, rather than a sequence of two segments. While Casali provides various other types of evidence pointing to the presence of prenasalized stops in the language, the fact that the speaker intuitions are the only source of syllable boundaries for [sa.mbe] (and the other data presented in the same section) is a tacit assumption that the intuitions are reliable indicators of syllable structure.

This assumption is possibly bolstered by survivorship bias in the literature: works that present a model of the syllable based on both speaker intuitions and other phonological evidence are likely to be those cases in which there is no clash between the two sources of evidence. If native speaker intuitions point to one model of the syllable while other facts about the phonology point to another, then an author may choose to foreground one over the other rather than present both.

The assumption that speakers can reliably identify syllable structure is possibly the root of an asymmetry between how syllable structure and other phonological structures are presented in descriptive works such as grammars and phonology sketches. These works generally assume some model of the syllable in the language they describe, either in a section dedicated to syllable structure or transcribed in phonological/phonetic data throughout the work. However, despite how common it is for these works to have a concrete analysis of syllable structure in the languages they describe, this analysis is often accompanied by little to no information about what evidence it is based on and what alternatives (if any) have been ruled out.

The lack of argumentation surrounding syllable structure stands in stark contrast to how many other phonological structures are treated in descriptive works. Take, as one example of myriad, Pan's (2012) grammar of Saaroa, an Austronesian language of Taiwan. Pan gives a detailed account of the syllable (pp. 32–5), including the minimal and maximal syllable, the conditions under which codas can occur, and the vowel sequences that are allowed and disallowed tautosyllabically. This model feeds into following discussions of stress (pp. 35–8) and vowel deletion (pp. 38–41). The model of Saaroa syllable structure Pan presents, however, is given no

explanation for how it came into being. Statements like "unlike a long vowel that form[s] just one syllable, vowels in a sequence constitute two syllables" (p. 33) are given without context: were speakers given stimuli with long vowels and vowel sequences and asked to syllabify them? (And was there agreement across speakers?) Is the distinction something that the author found while analyzing the language's phonetics, but chose not to discuss in the final publication?[2] It is unlikely that such a discussion was cut for brevity (the work is already over 400 pages), suggesting that it is more likely that the syllable boundaries under question were elicited directly from the speaker.[3]

On the other hand, Pan (2012) provides detailed evidence for other structures of Saaroa phonology that he presents. Minimal pairs are provided to evidence contrasts between phoneme pairs like /k/ versus /m/, /ts/ versus /v/, and /s/ versus /ŋ/, all cases in which the modern linguist would assume contrast without extraordinary evidence to the contrary. The grammar also documents an incredible ten marginal phonemes in the language's loan phonology, and the differential structures of hyper- versus hypoarticulated speech in the underlying and surface representations. This is all to say that Pan's grammar is not lacking in evidence and argumentation for its model of phonology in general. Rather, it has become normalized in the field to present models of syllable structure, likely directly elicited, as is without mention of methodology or consideration of possible alternatives.

Pan's grammar of Saaroa is by no means unique in its presentation of syllable structure. Looking at other grammars of Austronesian languages in Taiwan, Chang's (2006) grammar of Paiwan also contains a dedicated section on syllable structure, which does not provide evidence, argumentation, or possible alternatives (pp. 31–4). The same goes for Teng's (2008) grammar of Puyuma, Rau's (1992) grammar of Atayal, and the list goes on. Nor are these types of descriptions at all limited to studies of languages in Taiwan: the same direct style of presentation can be found in Doornenbal's (2009) grammar of Bantawa (Sino-Tibetan, Nepal), Klamer's (2010) grammar of Teiwa (Papuan, Indonesia), Heath's (1998) grammar of Koyra Chiini (Nilo-Saharan,

---

**2** I have reached out to Chia-jung Pan to ask about the analytical decisions discussed here and made a similar inquiry to Chun-Mei Chen, the author of the largest existing work on Budai Rukai phonology (Chen 2006), to be discussed later in the paper. As I did not receive a reply from either, I base the discussion in this paper on only that information available to me (and other readers of these works).
**3** A third, more cynical, possibility exists: there may be some (non-native speaker) field linguists who document syllable structure as they themselves perceive it. As the goal of this paper is to argue for external evidence as the primary basis of syllable structure analyses over intuitions in general, this includes intuitions of linguists that have made it into descriptive works. Thus, there is no harm giving existing descriptive works the benefit of the doubt, that the syllable structures they present are sourced from native speaker intuitions rather than their own.

Mali), and so on.[4] The models of syllable structure in these descriptive works are then adopted by other scholars in journal articles and the like.

This is not to say that descriptive works are entirely lacking in argumentation for the syllable structures they present, rather that models of syllable structure that lack argumentation can and do appear in published descriptive works without raising suspicion. Some counterexamples within descriptive works of Austronesian languages in Taiwan include Zeitoun's (2007) grammar of Mantauran Rukai, Blust's (1999) phonology of Pazeh, and Tung's (1964) grammar of Tsou, all of which provide arguments for their presented models of syllable structure based on evidence like glide formation, sites of epenthesis, and patterns in reduplication.

In summary, many native speakers are able to provide intuitions about specific phonological structures without linguistic training. The strength of their intuitions and the agreement between speakers has been taken by many linguists as a sign of their reliability in identifying a domain that is relevant to the rest of their language's phonology. The ease of eliciting data on syllable structure directly from speakers, and the presumed reliability of that data, has given syllable structure a unique place in descriptive works on phonology: many of these works neither provide information about how the data they present were elicited or analyzed nor explore any alternatives to structures indicated by directly elicited data. There is little room left to question whether what speakers have accessed is truly the syllable (as used elsewhere in the phonology), nor to examine possible faults in the methodology of how syllable structure is elicited in field or laboratory settings. In turn, these unconfirmed models of syllable structure color both the rest of the phonological analysis within the work, as well as the work of other scholars who adopt the model.

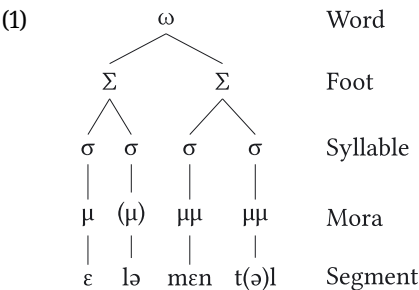## 3 The mora and the syllable in phonology

The two phonological domains relevant to this paper are the syllable and the mora. Section 3.1 will define these terms and describe them in the context of the prosodic

---

**4** It is not my intention to single out any of the works cited here or undermine their incredible importance as descriptive records of these languages. Instead, I mean to demonstrate that: i) "as is" presentation of directly elicited syllable structure is common in descriptive linguistic works and ii) this style of presentation is not limited to any one language group or subgroup of scholars. Please note also that these works were chosen not because they are unique among descriptive work in phonology, quite the opposite: they are representative of trends in descriptive work as a whole. The grammars of languages on Taiwan are familiar to me from my other research projects, and the other examples were chosen at random. It only took three attempts to find three grammars with syllable structure presented "as is," which I take as confirmation that this is a real practice in the field.

hierarchy, and Section 3.2 will discuss speaker intuitions of these domains. Section 3.3 will summarize these discussions' findings and their relevance to the elicitation process.

## 3.1 The prosodic hierarchy

There are three levels of the prosodic hierarchy that lie between the segment and the prosodic word: the foot, the syllable, and the mora. An example of these domains can be seen in (1), which shows the prosodic structure of the English word *elemental*.

(1)

| | | |
|---|---|---|
| ω | | Word |
| Σ Σ | | Foot |
| σ σ σ σ | | Syllable |
| μ (μ) μμ μμ | | Mora |
| ɛ lə mɛn t(ə)l | | Segment |

The largest of these subword domains is the foot, a grouping of two syllables (or morae, depending on the language and analysis) with one "strong" and one "weak" position, often in a rhythmic alternation, as is the case with *elemental*. Models of foot structure in descriptive work are directly affected by how the syllable and mora are analyzed. However, as the foot is less commonly the target of direct elicitation in field and laboratory studies, this paper will focus primarily on the syllable and mora.

The syllable can be defined in a number of ways. In terms of phonetics, the syllable may represent oscillations in jaw movements (Ohala 2008), or in intensity or other phonetic correlates of sonority (Parker 2002, and references therein). This paper is primarily concerned with the syllable as defined by phonology, in which it is a domain referenced by numerous operations: phonological rules may target consonants either at the beginning or end of a syllable, the syllable may be the target of reduplication, the syllable may be the host of stress or in some languages the tone-bearing unit, and so on (Blevins 1995). These definitions stand alongside the unit that speakers intuitively break words into, as was discussed in Section 2.2. The syllable has been argued to be a universal in spoken languages by authors including Greenberg (1962), with its definition extended to encompass the hold-movement oscillations in signed languages (Liddell 1984), while other scholars have questioned whether all speech can be discretely parsed into syllables, and whether all languages' phonologies reference the syllable as a domain (Hyman 1983, 2011).

The mora is a unit of phonological weight, which groups syllables of certain structures together based on shared behaviors. Some of the phonological behaviors that have been analyzed through mora structure include certain syllable types being

available word-finally (cf. Hammond 1997 on vowel quality in English), attracting/repelling stress (Hyman 1985), and being able to host contour tones (Duanmu 1990, among others). Some models posit that all syllables have at least one mora (Hayes 1989; Hyman 1985), while others allow for the possibility of zero-mora syllables (Piggott 1995). In many languages, onset consonants do not contribute weight to the syllable, only the rime (i.e., the nucleus and coda), although there are a number of languages for which onsets can also bear morae, such as Pirahã (Everett and Everett 1984).

As an example of mora structure, the initial syllable /ɛ/ in (1) has been assigned only one mora (μ), as the vowel /ɛ/ cannot appear word-finally in English (*ɛ#). On the other hand, the penultimate syllable /mɛn/ is bimoraic, as the coda consonant /n/ also bears a mora. As it is bimoraic, it is a licit word-final syllable in English: *men* /mɛn/; *amen* /eɪmɛn/.

It should be noted, however, that the calculation of syllable weight is highly language- and analysis-dependent, and the aforementioned diagnostics for mora structure do not necessarily lead to identical results. For example, Cantonese has constraints on which vowel-qualities may be syllable-final, as well as which rimes may bear contour tones. Both diagnostics have been used to argue for mora structure in the language: like Hammond's (1997) model of English mora structure cited above, Cheung (1986) ascribes monomoraicity to short vowels like /ɐ/ which cannot appear syllable-finally (*Cɐ#, but Cɐj#, Cɐn#, Cɐt# all licit) and bimoraicity to long vowels like /aː/ which can (Caː#, Caːj#, Caːn#, Caːt# all licit). On the other hand, Iacoponi (2014, inter alia) takes rime-tone compatibility as the basis of his model of mora structure in Cantonese: CVT syllables are monomoraic as they can, in nonderived contexts, only bear level tones (Caːt$^H$, Caːt$^M$, Caːt$^L$, but *Caːt$^{LM}$, *Caːt$^{MH}$, *Caːt$^{ML}$), while CV and CVR syllables, regardless of vowel length, can bear contour tones (e.g., Cɐn$^{LM}$, Cɐj$^{MH}$, Caː$^{ML}$ all licit). This leaves syllables like Caːt in the awkward position of being bimoraic in models of mora structure based on vowel quality, and monomoraic in models of mora structure based on contour tone availability. As this example shows, the mora is in many cases less easily and intuitively parsed (by linguist and speaker alike) than the syllable.

## 3.2 Can speakers access the mora?

Hayes (1995: 28–9) notes the ability of native speakers to access various levels of the prosodic hierarchy by being asked to "tap" a certain number of times during the production of a word.[5] Depending on the word provided and the number of taps requested, the taps will generally be assigned by the speaker to syllables that bear a

---

5 Hayes notes that this exercise was used with students of "beginning phonetics," specifically as a method for helping those students who "cannot hear stress" to access prosodic structure in their

certain level of prominence: for example, when asked to assign six taps to the word *reconciliation*, the speaker will assign one tap to each of the six syllables in the word. If asked to assign three taps, the speaker will tap on the first, third, and fifth syllable: the prominent positions in each of the word's three metrical feet. By asking speakers to intuit taps in various numbers, Hayes is able to replicate a metrical grid representation of the word's prosodic structure. His results are reproduced here, with the metrical grid above (with each "x" marking a level of prominence above each syllable) and number of "taps" below (from p. 29):

(5)

| | | | | | x | |
|---|---|---|---|---|---|---|
| | x | | | | x | |
| | x | | x | | x | |
| | x | x | x | x | x | x |
| | re | con | ci | li | a | tion |
| 1 tap: | | | | | τ | |
| 2 taps: | τ | | | | τ | |
| 3 taps: | τ | | τ | | τ | |
| 4 taps: | | | ??? | | | |
| 5 taps: | | | ??? | | | |
| 6 taps: | τ | τ | τ | τ | τ | τ |

The success of this exercise raises the question: if the mora represents an identifiable prosodic domain in languages like English, can it be accessed by the same method as the syllable and foot were accessed by Hayes' method?

Using the vowel quality and presence of coda as diagnostics, the word *reconciliation* has ten morae: one each in the syllables /ɹɛ/ and /sɪ/, each an open syllable with a short vowel (note that neither is a licit syllable word-finally in English), and two each for /kən/ and /ʃən/ which receive one mora for the vowel and one for the coda, as well as two morae each for /li/ and /eɪ/ which contain bimoraic long vowels. If Hayes' "tapping" methodology can be extended to the mora level with English speakers, then the result should be as shown in (2):

(2)

| | σ | σ | σ | σ | σ | σ | Syllable |
|---|---|---|---|---|---|---|---|
| | μ | μμ | μ | μμ | μμ | μμ | Mora |
| | ɹɛ | kən | sɪ | li | eɪ | ʃən | Segment |
| 10 taps?: | τ | ττ | τ | ττ | ττ | ττ | |

There are two major issues with trying to access the English mora this way. The first is that while syllables like [[kə]ᵤ[n]ᵤ]σ and [[ʃə]ᵤ[n]ᵤ]σ have each mora associated with

---

native language. Whether these students can be considered so-called "naive speakers," a term and categorization otherwise avoided in this paper, or otherwise, is left to the reader.

a discrete substring (i.e., hypothetically, a speaker could produce a two-tap structure like [kə]$_τ$[n]$_τ$ or [ʃə]$_τ$[n]$_τ$), the syllables /li/ and /eɪ/ are not as easily broken up.[6]

The other issue with "tapping" as a means to elicit the mora in English is that speakers do not have the same type of intuitions at this level as they do for the syllable. A speaker could be coached to tap once on syllables that would be illicit word-finally and twice on others: experimental studies like Westby (1984) find that English speakers have knowledge of the constraint against word-final short/lax vowels that can be accessed in a laboratory setting. That said, this type of coaching defeats the purpose of Hayes' tapping methodology, in which the number of taps that correspond to the syllable, foot, secondary stress, and word stress of the word provided do not require any additional coaching.

It seems that in the case of English, the nonlinguist speaker needs as much training as the linguist to parse their speech into morae. On the other hand, there are languages for which the mora is more readily accessible, such as Japanese.

In (Tokyo) Japanese, the mora is a domain more often referred to by phonological operations than any larger "syllable" proposed for the language. The mora in Japanese can take the forms V, C(j)V, the syllabic nasal N, or what is known in the literature as Q, the first timing unit of a geminate consonant (Vance 2008), while proposals of a larger syllable in the language take roughly the maximal form CjVX. It is the mora, rather than the syllable, in Japanese that is referenced by long-standing poetic traditions such as haiku and tanka, as well as more recent innovations such as the "moraic assonance" of modern Japanese hip hop (Tsujimura and Davis 2008). The mora is the bearer of pitch accent and the tone-bearing unit (Pierrehumbert and Beckman 1988: 119–21), arguably a better predictor for duration than other proposed prosodic domains (Port et al. 1987),[7] and a common target of speech errors (Kubozono 1989).

The metalinguistic knowledge Japanese speakers have about mora structure is mentioned in numerous works. Tsujimura and Davis (2008: p. 159) note that "Japanese speakers naturally divide words into moras rather than syllables," and that Japanese speakers "hear the place name *London* as consisting of four moras while English speakers hear it as comprising two syllables" (emphasis theirs). Japanese speakers' instinctual parsing of stimuli into morae has seen some confirmation in experimental work: McQueen et al. (2001) find that listeners are better able to recognize sequences within larger stimuli when the edges of those sequences align with mora boundaries.

Taking a step back, this evidence demonstrates two things: i) that there is a prosodic domain into which Japanese speakers easily (and perhaps automatically)

---

**6** While possible, parses like [li]$_τ$[i]$_τ$ and [e]$_τ$[ɪ]$_τ$ are wholly unintuitive to me as a native speaker, and I suspect that other English speakers feel similarly.

**7** However, see Warner and Arai (2001) for counterarguments.

parse speech and ii) that this domain can take the shapes V, C(j)V, syllabic N, and Q (the first half of a geminate consonant). The association of this domain with the "mora" level of the modern prosodic hierarchy is an artifact of linguistic theory: while both the English and Japanese "mora" may be a single consonant or short vowel, it is clear that there are also major differences in the role they play in how speakers organize speech sounds. In terms of being the domain speakers more easily parse speech into, the Japanese "mora" is more similar to the English "syllable" than the English "mora." This raises the question: could what has been described as the "mora" in Japanese be isomorphic to a Japanese "syllable"? Such an account would allow for the definition of the "syllable" to encompass its relative availability to speakers in both languages.

There has been a long-running debate as to whether a "syllable" can meaningfully be identified in Tokyo Japanese. Kubozono (2003) notes a number of benefits to projecting a larger CjVX syllable in the language. The processes he notes, including a language game played by jazz musicians and certain phenomena in loanword phonology and motherese, are not core aspects of Japanese phonology or used by all speakers. A counter-proposal is provided by Labrune (2012), who finds no need for the "heavy syllables" referenced by the processes described by Kubozono. She takes a similar approach to Japanese as Hyman (1983) did with Gokana, i.e., that a mora but not a syllable is projected in the language's phonology. Kawahara (2016) provides additional arguments for the syllable based on claimed phonetic differences between onset and coda consonants.

This debate may or may not have reached a definitive conclusion. Some of the arguments put forth by Kawahara (2016) may also find alternatives without reference to a heavy CjVX syllable. To take one example, /N/ triggers nasalization in a preceding vowel, but not in a following vowel, which Kawahara takes as evidence for /N/ and preceding vowels to form a tautosyllabic sequence: [ãN.a]. However, if the diachronic source of vowel nasalization in Japanese is a mis-timing between the oral closure and velar opening gestures of nasal consonants like /N/, as is common cross-linguistically (Blevins 2004; Watson 1999), then this asymmetry is expected regardless of how /N/ is syllabified.

It is beyond the scope of this paper to go through the claims in Kawahara (2016) point by point and find alternatives without using reference to a syllable larger than the mora. This exercise brings up a number of theoretical questions that descriptive linguists need an answer to regardless of whether one can dig deep enough into Tokyo Japanese to find a syllable larger than one mora. The most important such question, in my view, is as follows: if there is evidence for the mora but not for a larger syllable, then what happens to the syllable tier in the language? Hyman (1983) and Labrune (2012) take the approach that the syllable tier is not projected, and that the mora is the sole tier between the segment and prosodic word. However, there is

an alternative worth exploring: namely, that the mora and syllable are both pro-jected and are identical.[8]

Assuming a hypothetical language Japanese′ that follows Labrune's (2012) model of Tokyo Japanese and has none of the evidence for heavy syllables described by Kubozono (2003) or Kawahara (2016), the mora-as-syllable approach would generate structures as in (3a) (based on the word *hontō* "truth"), while Labrune's (2012) no-syllable approach would generate (3b), and including heavy syllables (as done by Kawahara 2016; Kubozono 2003) would generate the structure in (3c):[9]

(3)   a.
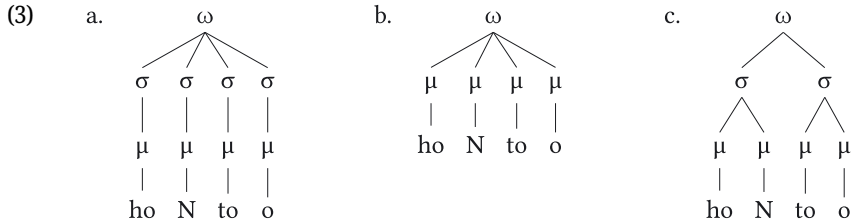


The choice between the isomorphic mora/syllable model in (3a) and the no-syllable model in (3b) depends on the goals of the analysis. Is it desirable to group the intuitively parsed domain of both the English and Japanese(′) speakers on the "syllable" tier? Or is it more desirable to avoid syllable types such as lone /N/ [ɴ] and /Q/ [p̚] [t̚] [k̚] [s̩]? If the latter is the largest hindrance to a mora-as-syllable account of Japanese′, then it should be noted that these syllable types have been described for other languages such as Coptic (Depuydt 1993) and Imdlawn Tashlhiyt Berber (Dell and Elmedlaoui 1985). Even within Japonic, syllabic /s/ has developed in Ōgami, a Ryukyuan language spoken in Okinawa prefecture (Pellard 2009). Even if these syllabic consonants are uncommon cross-linguistically, they must remain an option for those languages (like Japanese′) for which they are the best-evidenced model of the syllable.

## 3.3 What should we expect when documenting syllables and morae?

The terms "syllable" and "mora" appear in myriad descriptive works in phonology but resist a universal definition in terms of their structure, relationship to

---

**8** This would also allow for an analysis in which it is the mora that is not projected, and the previously described mora is relabeled the syllable. Whether the mora is still projected when redundant would neither be easily evidenced nor affect this paper's thesis in any meaningful way.
**9** The prosodic foot, while present in many models of Japanese phonology, is omitted from this example.

speakers' metalinguistic knowledge, and best practices for elicitation (when documented at all).

While speakers of many languages are able to break up speech into consistently sized subdomains when asked (and perhaps automatically during speech perception), it is unclear whether this task can be argued to target one particular level of the prosodic hierarchy across speakers of different languages. Scholars of English have defined both a syllable and a mora in the language's phonology, but it is only the syllable that can reasonably be accessed by non-linguistically trained speakers in a direct elicitation task. Japanese speakers, on the other hand, instinctively parse speech into a domain traditionally labeled the mora, and there are difficulties both in trying to fit the syllable isometrically to the mora and with attempts to evidence a larger syllable in the language.

The field linguist, when working with a speaker who can parse words into smaller prosodic domains, thus cannot be sure without additional evidence i) which level of the prosodic hierarchy is being accessed by the speaker and ii) whether a syllable-mora distinction is active in the language's phonology generally.

There exists also a third type of complication that can arise in the interpretation of directly elicited syllable structures: languages in which speakers are able to parse their speech into a domain that is not active in the language's phonology at all. The following sections of this paper will provide an example of such a situation, namely in Budai Rukai, an Austronesian language of Taiwan. While the English syllable and Japanese mora are both active generally in English and Japanese phonology, respectively, linguists working on Budai Rukai (including myself) have described speaker intuitions of a "syllable" that I argue is not active in the language's phonology but rather is an artifact of a flaw in the elicitation process.

# 4 Syllable structure in Budai Rukai

This section will present an overview of the Budai Rukai syllable, beginning with an introduction to the language and the linguistic profile of its speakers (Section 4.1), then an overview of the language's nonintonational phonology (Section 4.2), followed by how the language's syllable structure has been described in the literature so far (Section 4.3), a description of Budai Rukai's intonational phonology (Section 4.4), and a reanalysis of the language's syllable structure incorporating findings from the intonation study (Section 4.5). Section 4.6 will provide a summary and discussion of this new model's implications for Austronesian studies.

The novel data presented here are from a comparative study of prosody and intonation in Formosan languages, including newly elicited data from thirteen languages/varieties in addition to Budai Rukai (Mantauran Rukai, Saaroa,

**Table 1:** Consonant inventory of Budai Rukai.

|  | Bilabial | Labiodental | Interdental | Alveolar | Retroflex | Palatal | Velar | Laryngeal |
|---|---|---|---|---|---|---|---|---|
| Vl. stop | p |  |  | t |  |  | k | [ʔ] <'> |
| Voiced stop | b |  |  | d | ɖ <dr> |  | g |  |
| Vl. fricative |  |  | θ <th> | s |  |  |  | (h) |
| V. fricative |  | v | ð <dh> |  |  |  |  |  |
| Vl. affricate |  |  |  | ts <c> |  |  |  |  |
| Nasal | m |  |  | n |  |  | ŋ <ng> |  |
| Trill |  |  |  | r |  |  |  |  |
| Lat. appr. |  |  |  | l |  |  |  |  |
| Flap |  |  |  |  | ɺ <lr> |  |  |  |

V.: voiced; Vl.: voiceless; Lat.: lateral; appr.: approximant.

Kanakanavu, Pnguu Tsou, Tjaylaking/Piuma Paiwan, Southern Amis, PatRungan Kavalan, Isbukun Bunun, Tgdaya/Toda/Truku Seediq, and Pazeh).[10] The Budai Rukai data were elicited during 3h of interviews with a female speaker, birth year 1968. The speaker is from Vedai, the village that gives Budai Rukai its name, and the largest village in the Budai-speaking area. The speaker uses Budai Rukai in her daily life (including in its written form, using the orthography shown in Tables 1 and 2), in addition to Taiwanese Southern Min and Mandarin. The elicitation sessions were conducted in Mandarin. At the time of the interviews, the speaker was a graduate student in a field other than linguistics (and had not otherwise received linguistic training or consulted for linguistic studies prior).

**Table 2:** Vowel inventory of Budai Rukai.

|  | Front | Central | Back |
|---|---|---|---|
| High | i, ii |  | u~o <u>, uu~oo <uu> |
| Mid |  | ə <e>, əə <ee> |  |
| Low |  | a, aa |  |

---

**10** The data from this study appear in the author's unpublished PhD dissertation (Macaulay 2021), and a descriptive overview of intonational phonology in the languages covered appears in an edited volume (Macaulay 2023). The discussion in this paper of the elicitation process, the reanalysis of syllable structure in Budai Rukai, and the context of Chinese-language education have not appeared in published works.

## 4.1 Budai Rukai and its speakers

Budai Rukai is one of six varieties of the "Rukai" language cluster. These languages are Formosan, i.e., they are Austronesian languages spoken in Taiwan, and Rukai in particular is spoken in a mountainous area spanning parts of Kaohsiung, Pingtung, and Taitung Counties. Vedai lies in Wutai Township, Pingtung County.

The groups of people who speak "Rukai" languages are culturally and linguistically diverse, and not all members of these groups identify with the term "Rukai." There are also considerable differences between Rukai varieties in terms of phonology, syntax, and lexicon, and not all pairs of varieties are fully mutually intelligible (Tu 1994). Nevertheless, the label "Rukai" conveniently groups the linguistic descendants of Proto-Rukai, be they dialects or separate languages.

The Formosan languages are endangered, Budai Rukai included. Taiwan has had over a century of compulsory education in Japanese (1895–1945) and Mandarin (1946–present), as well as waves of Southern Min-speaking settlers to Taiwan well before these policies, all of which have demanded assimilation (linguistic and otherwise) of the indigenous populations. Because of this, bi- and multilingualism is the norm in Formosan language speech communities. There are likely no remaining monolingual speakers of Formosan languages (Li 2008).[11]

If monolingual Formosan language speakers do exist somewhere, they are not those routinely consulted in linguistic documentation projects, which are typically conducted via translation tasks in one of the contact languages (Southern Min, Mandarin, or Japanese with older speakers). As one example, Blust (2003: 8) describes the elicitation process for his dictionary of Thao, which involved translation tasks from Thao to Southern Min, and then some additional translation of the responses from Southern Min into Mandarin (which is more accessible to many non-Taiwanese scholars).

As noted earlier, the speaker consulted for this study speaks her tribal language Budai Rukai, plus Southern Min and Mandarin. Her linguistic profile is thus typical both among the Budai Rukai speech community, as well as among linguistic consultants who inform Formosan language documentation projects. The final sections of this paper argue that the speaker's experience with education and culture in the dominant language(s) has affected how she has responded to Mandarin-language elicitation tasks. This experience is common to the Budai Rukai speech community, and I argue that the interference it causes during the standard linguistic interview is a factor also in previous and future documentation work on the language with other speakers. Furthermore, if

---

**11** One anonymous reviewer on a previous version of this paper notes that the Wikipedia entry "Rukai language" mentions the existence of monolingual Rukai speakers. The claim is uncited, and I have found no evidence in available surveys that there are monolingual Rukai speakers.

bilingualism is the cause of the interference, then the same type of interference may be expected in work with any speech community for which bilingualism is the norm. This includes virtually all endangered languages,[12] as well as nonendangered languages like Swedish and Dutch, which have very few monolingual speakers.[13]

## 4.2 An overview of Budai Rukai phonology

This study finds an inventory of eighteen consonants in Budai Rukai (excluding those exclusive to loanwords), shown in Table 1 alongside their orthographic equivalents in angle brackets < >. This inventory is equivalent to that presented by Chen (2006), with three exceptions: i) the data elicited include /h/ in loanwords like *hungu* (from Japanese 本 *hon* [hoɴ]); ii) a noncontrastive glottal stop was found in certain positions at slow speech rates, to be discussed in Section 4.5; and iii) glide phonemes /j/ /w/ are not present in the inventory, to be discussed throughout this section.

Like many Formosan languages, Budai Rukai has four contrastive vowel qualities, as shown in Table 2: /a/, /i/, /u/ (which varies between [u] and [o]; see also Chen 2006: 238), and /ə/, written <e> in Rukai orthography. All four vowel qualities have what has been described in the literature as a short-long pair, although an alternative structure for the "long vowels" will be explored in Section 4.5. Note that /ə/ in Rukai (including Budai and other dialects) lacks some of the restrictions placed on it in other languages, namely existing in geminate form, and being able to bear stress.

In addition to consonants and vowels, Budai Rukai has unpredictable stress, falling on the penult or antepenult. Unpredictable stress is uncommon in Formosan (including among other Rukai varieties), found elsewhere only in Kanakanavu (Tsuchida 1976: 30–1), and some analyses of Saaroa (Macaulay 2021). Some examples of penultimate and antepenultimate stress in Budai Rukai are shown in (4)–(5), where stress is marked with the acute accent "´":[14]

---

**12** See Wurm (1991) for discussion on causes of language endangerment and the role of language contact.
**13** In a study by Eurostat (2022), only 4.1 % of Swedish and 7 % of Dutch participants responded that they spoke no "foreign languages."
**14** The sorting of Budai Rukai words into "penultimate" and "antepenultimate" stress categories is dependent on the model of syllable structure one assumes. The examples in (4)–(5) have been chosen as they conform to these categories both in existing models of the Budai Rukai syllable as well as in this paper's reanalysis. Note that these terms are being used to refer to the position of the stressed syllable in surface structure. The interactions between stress assignment and epenthesis, reduplication, and suffixation/encliticization are complex and have become more so with the addition of this paper's reanalysis of syllable structure. It should be mentioned that while some words with surface antepenultimate stress can possibly be analyzed as having an epenthetic "echo vowel" and

(4)       Penultimate stress:
           a.   *karádha* /karáða/ "pangolin"
           b.   *calrínga* /tsaɭíŋa/ "ear"
           c.   *cakéna* /tsakə́na/ "the ground"

(5)       Antepenultimate stress:
           a.   *manémane* /manə́manə/ "what"
           b.   *draúsulu* /ɖaúsul(u)/ "storehouse"
           c.   *kadaéngane* /kadaə́ŋanə/ "land"

The position of stress in Budai Rukai lexical items has been of interest to Austronesianists, notably as the crux of Ross's (1992) proposal that the stress system is related to that in the Philippine languages. However, it has also been argued that the two systems are unrelated (Blust 1997).

## 4.3 The existing model of Budai Rukai syllable structure

The most extensive account of the syllable in Budai Rukai is found in Chen (2006: 211–8), a comparative study of the phonologies of Budai Rukai and Paiwan (including also a morphological sketch of Budai Rukai). Chen finds a maximal syllable of CVV, where V can be a vowel or glide. Notably, even larger syllables than CVV are described by Liu (2011: 42–56), who analyzes words like *kwaw* "eagle" as monosyllabic CVVV (vs. disyllabic /ku.au/ in Chen 2006:214). In both accounts, the minimal syllable is V, as onsets are not obligatory.

     Glides are the only codas permitted in Chen's model and can occupy either the first or second timing slot of heavy CVV syllables. Although codas are not permitted in any part of the word, some words have a final consonant underlyingly, resolved by the epenthesis of an "echo vowel" that copies the preceding vowel's quality (or surfaces as [ə] if the preceding vowel is /a/). These "echo vowels" are common to the other varieties of Rukai as well as in the neighboring languages Saaroa, Kanakanavu, and Tsou (Li 1973).

     Not all sequences of two vocoids are parsed tautosyllabically in Chen's model. For example, *dae* "earth" is transcribed as [da.ə] rather than monosyllabic *[daə]. Other sequences like /ia/, /ua/, /au/, and /iu/, as well as what she transcribes as long vowels, are tautosyllabic and form heavy syllables.

---

underlying penultimate stress, this would not absolve Budai Rukai stress of its unpredictability: both penultimate and antepenultimate stress surface throughout the language's morphological paradigm, and this paper reanalyzes many supposed penultimate-stressed forms as antepenultimate-stressed.

It should also be noted that the use of glides to transcribe Budai Rukai words is common to other works on the language such as Huang and Lai's (2012) dictionary of the language. Even without providing an autosegmental model of the Budai Rukai syllable (as Chen and Liu both do), the use of <w> and <y> over <u> and <i> for vocoids do provide an assumed syllable structure. For example, by listing the word for "cow" as <lwange>, the use of <w> suggests nonsyllabicity of this vocoid, and thus *[lu.a.ŋə] is ruled out as a parse, leaving [lwa.ŋə].

## 4.4 Intonational phonology in Budai Rukai

As part of a larger study of the intonational phonologies of Formosan languages, Macaulay (2021, 2023) presents a model of Budai Rukai intonation in Autosegmental-Metrical (AM) phonology (Pierrehumbert 1980).[15] AM phonology seeks to model the aspects of intonation that involve changes in fundamental frequency (f0), by analyzing the f0 contour of utterances into a sequence of tonal targets, including high tones (H) that represent f0 maxima and low tones (L) that represent f0 minima. These tonal targets form melodies that align either with prosodic domain boundaries ("boundary tones") or with prominent syllables ("pitch accents").

Tonal alignment in Budai Rukai happens at a level described by Macaulay (2021, 2023) as the IP (intonational phrase).[16] This phrase, in all utterance types studied, ends with a final low boundary tone L% (where % marks an IP boundary). While many utterances in Budai Rukai occur within a single IP, an IP boundary appears mid-utterance between disjuncts in forced-choice utterances, as well as between items in an inexhaustive list (in which case H% appears in place of L%). The beginning of the IP (in all utterance types) shows variability between high and low tonal targets, which is an areal feature shared with Mantauran Rukai and nearby Tsou and Saaroa (Macaulay 2021: 690–4). While this may indicate the presence of an optional initial %H boundary tone, there remain questions as to how this tonal target is assigned (ibid., pp. 103–6 for discussion in the context of Mantauran Rukai).

---

**15** An anonymous reviewer has noted alternatives to AM models in intonation. The use of AM over other frameworks by Macaulay (2021, 2023) and here serves to maximize comparability with other works in intonational typology such as Jun (2005, 2014), as well as with work on lexical tonal phonology. For additional discussion of frameworks in intonation, as well as discussion of patterns in Formosan intonation for which the explanatory power of frameworks differs, see Macaulay (2021).
**16** This label is used in these works for the largest prosodic domain at which there is evidence for tonal alignment, pending further investigation into prosodic grouping in Formosan languages. As one possible alternative, the "IP" in Budai Rukai has a number of similarities with what Beckman and Pierrehumbert (1986) label the AP (accentual phrase) in Tokyo Japanese.

Each IP in Budai Rukai surfaces with one pitch accent, regardless of the number of accented syllables within the IP. This constraint is found both in other nearby Formosan languages Tsou and Saaroa (in all utterance types), Kavalan and Puyuma (in interrogatives only), as well as non-Formosan languages, including several varieties of Japanese at the level of the AP (accentual phrase) (Kubozono 2011). In Budai Rukai, every word (excluding particles) has a prominent syllable, which is either the penult or antepenult of a domain including the stem and any suffixes and enclitics. The location of the prominent syllable is lexically specified: for example, *bakále* "knife" and suffixed forms will have a prominent syllable on the penult, whereas *bécenge* "millet" and suffixed forms will have a prominent syllable on the antepenult. It is to the final prominent syllable in the IP that the pitch accent attaches.[17]

The pitch accent melody in declarative utterances (and words in isolation) in Budai Rukai is (L+)H\*, in which the nuclear H\* surfaces on the prominent syllable.[18] The prenuclear L, when present, often shifts to the utterance-initial position, which is a pattern common to the Formosan languages Kavalan (interrogatives only) and Seediq. In polar- and *wh*-interrogatives and forced-choice utterances, a different pitch accent melody (L+)H\*L surfaces, in which there is a sharp fall in f0 (HL) during the prominent syllable rather than a steady descent between the peak (H\*) and final L%.

To summarize, the intonation of those utterance types attested in Budai Rukai is as shown in (6):[19]

(6)      *Multi-word utterances (single IP)*

|     | [(L/H) | Ø | Ø | H\* | L%]$_{IP}$ | Declarative |
|-----|--------|------|------|-----|------------|-------------|
| a.  | [(L/H) | Ø | Ø | H\* | L%]$_{IP}$ | Declarative |
| b.  | [(L/H) | Ø | Ø | H\*L | L%]$_{IP}$ | Polar interrogative |
| c.  | [(L/H) | H\* | Ø | Ø | L%]$_{IP}$ | Sarcasm |
|     |        | PW1 | PW2 | PW3 |            |             |

---

**17** This holds for declarative and interrogative utterances. In sarcastic utterances, it is the initial prominent syllable in the IP, rather than the final, to which the pitch accent anchors (Macaulay 2021: 162–3).

**18** Please note that the asterisk \* in Autosegmental-Metrical phonology denotes a tone as "nuclear," i.e., that tone in a melody which anchors to a prominent syllable. This is not the same usage as the asterisk in syntax (which denotes ungrammaticality) or in historical linguistics (which marks unattested reconstructed forms).

**19** Pitch tracks of example utterances and discussion of how they were analyzed can be found in Macaulay (2021: 149–63). Multi-word utterances are shown with three prosodic words (PW), i.e., three units that contain an accented syllable. The null set symbol Ø marks a position where no pitch accent is aligned despite an underlying accented syllable.

*Multi-word utterances (multiple IP's)*

d.                 [H\*L L%]$_{IP}$     [H\*L L%]$_{IP}$    Forced choice
                   Disjunct1       Disjunct2

e.   [L H\* H%]$_{IP}$   [L H\* H%]$_{IP}$   [H\* L%]$_{IP}$     Item list
    Item1          Item2         ItemFin

*Single-word utterances*

f.   [(L/H) H\* L%]$_{IP}$    Word in isolation

The use of intonation to diagnose syllable structure in the sections to follow makes use of the H\* common to all pitch accents in Budai Rukai, which is aligned in all cases to an accented syllable. This H\* appears one of two ways on the surface: as an f0 maximum (if H\* follows an initial L) or as an inflection point preceding a final fall (if H\* follows an initial H). The primary data for the following section come from instances of words in isolation; however, the wider dataset presented in Section 4.6 makes use also of tokens of words in the IP-final position of declaratives (in which they also receive a pitch accent containing H\*).

## 4.5 The Budai Rukai syllable reanalyzed with evidence from intonation

Instead of the heavy syllables proposed by Chen (2006) and Liu (2011) like CVV and CVVV, I explore an analysis here in which (C)V is the only syllable type. The largest difference between this CV model and previous CVV(V) models of the Budai Rukai syllable is that a maximal CV requires that the nuclear vocoid be a full vowel, rather than a glide. The CVV, CGV, CVG, and CGVG syllables (with V for full vowel, and G for glide) from previous analyses thus become CV.V and CV.V.V with all full vowels.

This shift from transcribing glides to vowels in the language is necessary to account for the distribution of pitch accents in the language. As noted in Section 4.4, all pitch accent types in Budai Rukai contain a nuclear H\* tone whose acoustic output is an f0 maximum. This f0 maximum occurs during the vowel of the stressed syllable to which it is aligned.

What is unexpected, given syllable types like CGV or CVG that contain one vowel and one glide, is that the pitch accent is aligned to any segment other than the full vowel. While nonvowel segments may be the host of tone in other languages (such as /m/ and /ŋ/ in Cantonese; Cheung 1986), these consonants are generally described as "syllabic" consonants, analyzed as syllable nuclei rather than being in the onset or coda. To have a tone-bearing glide would be at odds with the common use of [±SYLLABIC] (and equivalent structures in frameworks like Autosegmental

Phonology and Articulatory Phonology) to differentiate vowels from glides, as it would have to be simultaneously [–SYLLABIC] and occupy the nucleus of the syllable.[20]

Nevertheless, many of the segments in Budai Rukai that have previously been described either as instances of phonemic glides /j/ /w/ (or phonetic glides [j] [w] as the output of glide formation rules) consistently surface with pitch accents aligned to these supposed glides. As an example, the word for "dog" is transcribed by Huang and Lai (2012) as *tawpungu*. From this description, it can be assumed that the word is trisyllabic, and although this work does not transcribe the location of stress, it can be assumed that a pitch accent will be found on the nuclear vowel of either the ante-penultimate or penultimate syllable: [táw.pu.ŋu] or [taw.pú.ŋu]. However, as can be seen in Figure 1, it is the high back vocoid preceding [p] that consistently hosts the pitch accent's f0 peak. If this segment is a glide [w] as claimed by Huang and Lai, this word has the unexpected structure [taẃ.pu.ŋu].[21]

One possible reason for the f0 peaks surfacing on [w] in this word is that the pitch accent may be actually anchored to the preceding vowel ([táw.pu.ŋu]), and that f0
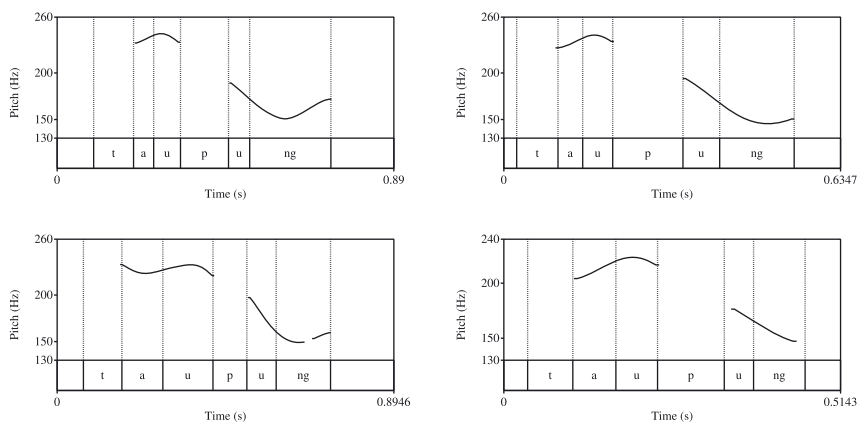


**Figure 1:** Pitch tracks of four tokens of *taúpungu* "dog" in Budai Rukai, produced by the same speaker.

---

**20** In some languages which contrast glides with vowels, the two categories have measurable differences in their phonetics (Chitoran 2002, inter alia), thus the bundle of acoustic features that form "glides" according to these studies could theoretically form a [+SYLLABIC] tone-bearing unit. However, neither has tone-bearing glides entered descriptions of the world's languages nor is there any evidence to deviate from the assumption that tone-bearing vocoids are vowels in the case of Budai Rukai.

**21** Note that utterance-final vowels following nasals are often deleted or devoiced in many (but not all) tokens. This is why the final /u/ is not labeled in the pitch tracks in Figure 1.

peaks in Budai Rukai may be systematically realized late. After all, this exact pattern of "peak delay" is found in other Formosan languages such as Paiwan and certain varieties of Seediq (Macaulay 2021) and is documented in non-Formosan languages like Viennese German (Mücke et al. 2012).

This explanation is unable to account for another fact about pitch accent alignment in Budai Rukai. Vocoid sequences VV in Budai Rukai fall into two categories with regards to their surfacing f0 contour: early peak V́V and late peak VV́. It is the latter with which the [aú] sequences in Figure 1 pattern. The difference between V́V and VV́ sequences can be seen in Figures 2 and 3, which show pitch tracks of a word transcribed by Huang and Lai (2012) as *laymay* "clothes," with and without the 1sg possessive enclitic =*li*.

To illustrate, Figure 2 shows a production of *laymay* in which the f0 peak occurs during the high front vocoid preceding [m], previously transcribed as a glide [j]. Like with *tawpungu* "dog," this leads to an unexpected tone-bearing glide: [laj.maj]. Interestingly, when the possessive enclitic =*li* is adjoined to this word, the f0 appears later, on the following [a]: [laj.máj.li], as shown in Figure 3. While the f0 peak appears during an [aj] sequence in both Figure 2's [laj́.maj] and Figure 3's [laj.máj.li], the former has a late VV́ peak while the latter has an early V́V peak.
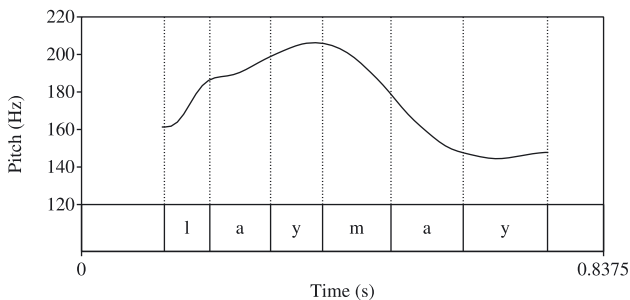


**Figure 2:** Pitch track of a production of *laímai* "clothes" in Budai Rukai.
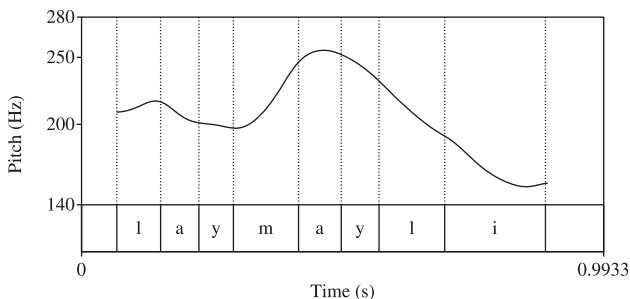


**Figure 3:** Pitch track of a production of *laimái=li* "my clothes" in Budai Rukai.

The simplest explanation for both the tone-bearing glides and the divergent early versus late peak f0 contours is that the "glides" like the [w] in *tawpungu* "dog" and the two [j]'s in *laymay* "clothes" are full vowels that form their own syllables, rather than being tautosyllabic with the preceding [a] in all cases. If [laj.maj] is instead [la.i.ma.i], then the stress alternation between the forms in Figures 2 and 3 is simply stated: the enclitic =*li* is included in the domain of stress assignment, and both forms surface with antepenultimate stress ([la.í.ma.i] vs. [la.i.má.i.li]).

While *tawpungu* "dog" and *laymay* "clothes" show examples of pitch accents on the glides of previously described CVG syllables, the CGV syllables that these works describe often also have pitch accents aligned to the "glide." An example of this can be seen in Figure 4, showing a pitch accent of a word transcribed as *lwange* "cow" by Huang and Lai (2012). In this word, the pitch accent is aligned to the first vocoid, previously transcribed as a [w]. Assuming this is indeed a glide, this word has the unexpected structure [lẃa.ŋə].

Here too, reanalyzing the glide [w] as a full vowel [u] ([lú.a.ŋə]) resolves the issue of tone-bearing glides in the language, as well as the additional problem of why *lwange* "cow" should have an early peak V́V f0 pattern, rather than the late peak VV́ pattern seen in *tawpungu* "dog" and *laymay* "clothes."

The early peak V́V versus late peak VV́ dichotomy extends also to what has been described in the literature as "long vowels." Compare, for example, the [aa] environment in Figures 5 and 6. This sequence is considered a tautosyllabic long vowel [a:] in both Chen's (2006) and Liu's (2011) model of Budai Rukai syllable structure. However, the two surface with different f0 contours: *daane* "house" in Figure 5 has an early peak (V́V) while *kaadraw* "big" in Figure 6 has a late peak. The existing tautosyllabic analysis of these [aa] sequences ([dá:.nə] and [ká:.ɖaw]) fails to allow for these divergent structures. As with the "vowel-glide" sequences in Figures 1–4, the "long vowels" benefit from reanalysis as a heterosyllabic sequence: [dá.a.nə] and [ka.á.ɖa.u] correctly encode the difference in f0 peak alignment.
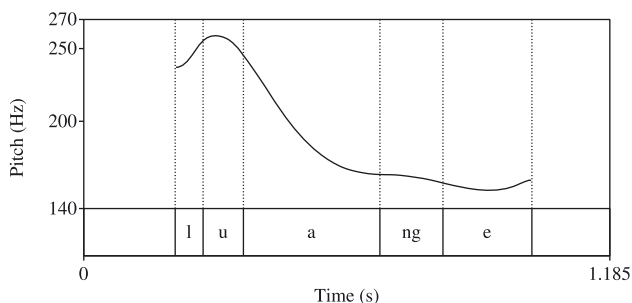


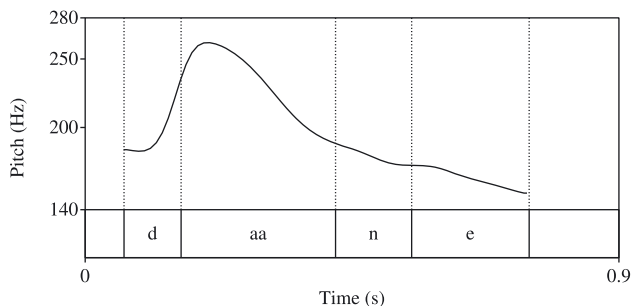**Figure 4:** Pitch track of a production of *lúange* "cow" in Budai Rukai.

**Figure 5:** Pitch track of a production of *dáane* "house" in Budai Rukai.
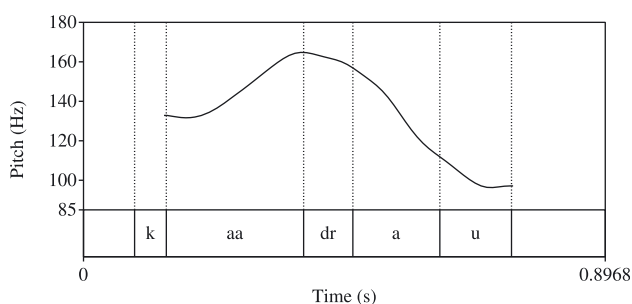


**Figure 6:** Pitch track of a production of *kaádrau* "big" in Budai Rukai. Audio from Huang and Lai (2012).

At this point in the analysis, there is one possibility that remains to be ruled out: that Budai Rukai intonation happens at the mora level, but that multiple morae may be tautosyllabic and form heavy syllables. Under this analysis, all vocoids are full vowels (rather than glides), but more than one may form a syllable similarly to the models proposed by Chen (2006) and Liu (2011). More specifically, while the structure in (7a) has been ruled out on the basis that it requires tone-bearing glides, the structure in (7b) must be ruled out before adopting the reanalysis of syllable structure that this paper presents, shown in (7c).

(7)   a.   *   ω        b.   ω        c.   ω

                  σ             σ           σ   σ

                 μ  μ          μ  μ        μ  μ
                 |  |          |  |        |  |
             C   a  j      C   a  i    C   a  i

The main issue with adopting a model like (7b) for CVV sequences in Budai Rukai is that by differentiating the syllable and mora, there is the implication that both levels can be evidenced in the language's phonology. The data presented in this section so far have shown that the mora is i) the tone-bearing unit of the language, ii) the domain of lexical items that is accented, and iii) the unit by which stress "shifts" under suffixation. So what is left to evidence the syllable in Budai Rukai, if it is larger than CV?

Many of the other structures that might help evidence a syllable larger than CV in other languages are not present in Budai Rukai. For example, there cannot be differences in the phonetics or behavior of onset versus coda consonants, because there are no coda consonants. Stress is not attracted (or repelled) by specific patterns in segmental content (unlike in Pazeh, another Formosan language; Macaulay 2020), so there is no argument to be made that stress-attracting segment sequences form heavy syllables. The shape of Budai Rukai syllables also does not pose any issue after reanalysis. While many scholars are hesitant to analyze Japanese morae like [ʦ] as syllables, the only mora types in Budai Rukai are CV and V, which are also among the most common syllable types in the world's languages (if not *the* most common).

Phonotactics also offers little in terms of evidence for the Budai Rukai syllable. Chen (2006: 236) does note some gaps in $V_1V_2$ vowel sequences such as */uə/; however, this says little about restrictions within a "syllabic" domain as she also analyzes certain sequences such as /a.ə/ as inherently heterosyllabic. Such gaps may also be inherited; Proto-Austronesian vocoid sequences generally occurred in the form of vowel-glide rimes like *-ay, *-aw, *-uy, and *-iw (Blust 1998), which limits the distribution of /a/ and /ə/ in these glide positions even now that the glides have vocalized in Budai Rukai.

One phonological interaction that *does* challenge a model of Budai Rukai with no glides is described by Chen (2006: 262–4), who finds alternations between glides and fricatives like /j/~/ð/ and /w/~/v/. These alternations are limited to stem-final position before an /a/-initial suffix: the [j] in /baj/ [baj] "gift" alternates with the [ð] in /sa-baj-anə/ [sabaðanə] "wedding gift," but the [j] in /apuj/ "fire" remains [j] in /apuj-ini/ [apujini] "his fire" (p. 263). Liu (2011: 27) describes a difference in behavior between /aj-a/, which triggers the alternation and surfaces as [aða], versus /ai-a/, which does not trigger the alternation, surfacing as [aia]. If this is truly the case, then /j/ and /i/ contrast (as with /w/ and /v/) and glides are required in underlying forms in the language.

While I have not yet done a full investigation of proposed /aja/ and /aia/ sequences in Budai Rukai, there are some possible workarounds that may account for these data without the use of a glide-vowel contrast. One is to note an additional difference in the data presented by Liu (2011) and Chen (2006) between /aj-a/ and /ai-a/ environments: the examples given by Liu for /ai-a/ and /au-a/ environments,

i.e., those which do not trigger the glide-fricative alternation, are comparatively short. Namely, /vai-a/ "dawn.IMP" and /bau-a/ "contract_project.IMP," both two-mora stems. The rest of the data that Chen and Liu present show the alternation and have longer stems. The only exception is /baj/ [baj] "give," which surfaces with [ð] in /sa-baj-anə/ "wedding gift" (see above). Even here, the /a/-initial suffix is being added to a base that is larger than two morae. Thus, it could be the case that the alternation is unavailable in the first (bimoraic) foot of the word, blocking the appearance of fricatives [ð] and [v] in /vai-a/ and /bau-a/. Another explanation for the lack of alternation with the stem /bau/ "contract project" is that perhaps this morpheme is a loanword: the translation for this word given by both Liu (2011) and Huang and Lai (2012) (who transcribe this word as *baw*) is 包工程, in which the verb 包 "contract" is pronounced *bāo* [páu] in Mandarin and *baúh* [báuʔ] in Taiwanese Southern Min (Maryknoll Taiwan 2013).

There is, however, one more piece of evidence for the heterosyllabicity of sequences like [a.i]. During the elicitation sessions from which this paper's novel data are presented, the speaker was asked to produce certain words in isolation at different speech rates. At lower, more hyperarticulated speech rates, some sequences like [a.i] were produced with a glottal stop separating the two vocoids: [a.ʔi]. An example can be seen in Figure 7, which shows spectrograms of two careful pronunciations of the word for "taro," a word transcribed by Huang and Lai (2012) as *tay*. In Figure 7a, a glottal stop (labeled ' ) can be seen on the spectrogram as the sharp dip in intensity (i.e., the lightness of the spectrogram compared to during phonation), and the creaky voice at the end of the preceding vowel (seen as the "choppiness" of the spectrogram). These features are absent in Figure 7b, which shows a spectrogram of a token produced without a glottal stop.

While there are languages like Danish that have tautosyllabic [aʔj] sequences in which a glottal stop interrupts the rime (Basbøll 1985), the glottal stop in Budai Rukai does not behave like the one found in Danish. In Danish, a rime like [aʔj] is considered tautosyllabic as it may alternate with a noninterrupted [aj] rime (e.g., in word compounding), and Danish has external evidence for heavy syllables like Caj
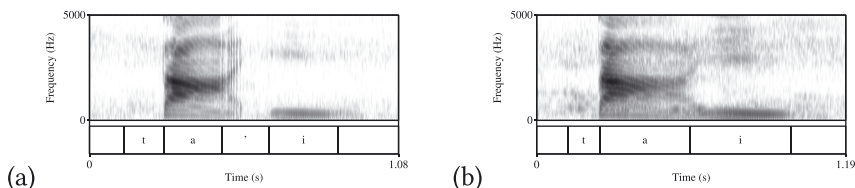


**Figure 7:** Spectrograms of two tokens of *tái* "sweet potato" in Budai Rukai, produced by the same speaker.

generally. In Budai Rukai, the glottal stop was only present (at least in the data I elicited) in hyperarticulated words with vowel-vowel junctures or between a filler word 嗯 [ɤː] and a Budai Rukai vowel-initial utterance. I thus analyze sequences like the [aʔi] in Figure 7a as evidence of the vowels being in separate syllables: [a.ʔi].

## 4.6 Conclusions and implications of the reanalyzed Budai Rukai syllable

Sections 4.4 and 4.5 provide a number of observations from the intonational phonology of Budai Rukai that suggest the need for a reanalysis of the structure of the syllable and its relationship to the mora.

Existing models of prosodic structure in Budai Rukai posit distinct structures for the syllable and mora. What was previously described as the "mora" in Budai Rukai has a number of properties that make it a better candidate for the "syllable": it is the accented unit in lexical items, it is the tone-bearing unit, and it is the domain by which stress "shifts" when a monosyllabic suffix is added. Additionally, the glide-vowel distinction that underlies existing models of the Budai Rukai syllable (by allowing for syllable types such as CGV, CVG, and CGVG) is challenged by the fact that many "glides" described for the languages anchor pitch accents in the language's intonational phonology, a property generally reserved to [+SYLLABIC] segments. Finally, many other features that distinguish light versus heavy syllables in other languages, for example, weight-sensitive stress and unique phonetic features of coda consonants, are missing from Budai Rukai's phonology.

These observations all lead to the conclusion that the previously described "mora" in Budai Rukai is in fact its syllable. This gives the language a maximal syllable of (C)V, rather than the CVV of Chen (2006) or the CVVV of Liu (2011).

One open question that remains is when glides were vocalized in the language: while Proto-Austronesian had glides *y and *w (surviving today as /j/ /w/ in many living Formosan languages), the lack of evidence for glides in Budai Rukai suggests that glides were reanalyzed as vowels at some point in the language's development. It should be noted that a lack of glide-vowel distinction is not unique to Budai Rukai: it is common to Mantauran Rukai, as well as the nearby languages Saaroa, Kanaka-navu, and Tsou.[22] Because of their distribution geographically, the lack of glides and short syllable types that result may be an areal feature that spread through contact (Macaulay 2021: 690–4).

---

**22** Most dialects of Tsou have one glide, realized as [j] intervocalically or as a non-high glide [ɰ̥] elsewhere; see Macaulay (2021: 172–6) for discussion on its phonetic properties. This glide is a recent development from Proto-Tsou(ic) *r and remains /r/ in the Luhtu dialect (Li 1979).

Importantly, the reanalysis of Budai Rukai syllable structure has a direct effect on the categorization of lexical items into those with antepenultimate versus penultimate stress. This categorization was proposed to be cognate to lexical stress in Philippine languages by Ross (1992), and the main counterargument from Blust (1997) was that the categorization in Budai Rukai does not line up closely enough with that in the Philippine languages. But what happens when a word like *dáane* "house," is reanalyzed from a penultimate-stressed word [dá:.nə] to an antepenultimate stressed word [dá.a.nə]? There are a fair number of lexical items that change from penultimate to antepenultimate stress under reanalysis of Budai Rukai's syllable structure. Those that came up over the course of this (relatively small-scale) study are shown in (8):[23]

(8)    Lexical items reanalyzed from penultimate to antepenultimate stress in current study:

|     | Huang&Lai | Original structure | Reanalyzed | Gloss |
| --- | --- | --- | --- | --- |
| (a) | *acilay* | [a.tɕí.laj] | [a.tɕí.la.i] | "water" |
| (b) | *adradraw* | [a.ɖá.ɖaw] | [a.ɖá.ɖa.u] | "wait" |
| (c) | *arakay* | [a.rá.kaj] | [a.rá.ka.i] | "use" |
| (d) | *daane* | [dá:.nə] | [dá.a.nə] | "house" |
| (e) | *galawgaw* | [ga.láw.gaw] | [ga.la.ú.ga.u] | "finger" |
| (f) | *kaadraw* | [ká:.ɖaw] | [ka.á.ɖa.u] | "big" |
| (g) | *kadray* | [ká.ɖaj] | [ká.ɖa.i] | "type of bag" |
| (h) | *kisaalru* | [ki.sá:.ɭu] | [ki.sá.a.ɭu] | "borrow" |
| (i) | *lapanay* | [la.pá.naj] | [la.pá.na.i] | "corn" |
| (j) | *laymay* | [láj.maj] | [la.í.ma.i] | "clothing" |
| (k) | *lwange* | [lwá.ŋə] | [lú.a.ŋə] | "cow" |
| (l) | *muswane* | [mu.swá.nə] | [mu.sú.a.nə] | "2sg.OBL" |
| (m) | *nguduy* | [ŋú.duj] | [ŋú.du.i] | "mouth" |
| (n) | *sawvalay* | [saw.vá.laj] | [sa.u.vá.la.i] | "man" |
| (o) | *tawthu* | [táw.θu] | [tá.u.θu] | "tail" |
| (p) | *vasaw* | [vá.saw] | [vá.sa.u] | "leaf" |

Whether this recategorization of Budai Rukai lexical items affects the strength of arguments for the cognacy of the Budai Rukai and Proto-Philippine stress systems, or if it is merely a recent innovation that stems from the vocalization of glides, remains to be investigated. Hopefully, revisiting the Budai Rukai lexicon with this new model of syllable structure will provide some answers to these and other questions.

---

**23** The column labeled "Huang & Lai" shows how the item appears in Huang and Lai (2012).

# 5 Direct elicitation of syllable structure in Budai Rukai

In the previous section, evidence other than direct elicitation was used to reanalyze the Budai Rukai syllable. This section will discuss what happens when speakers of Budai Rukai are asked to parse words into syllables (Section 5.1), followed by a discussion of contact languages as possible interference in this elicitation task (Section 5.2).

## 5.1 Speaker intuitions of the Budai Rukai syllable in the current study

Despite the differences between this paper's model of the Budai Rukai syllable and the conclusions reached by previous authors such as Chen (2006) and Liu (2011), I was able to replicate the structures of the previous analyses in a direct elicitation task.

For example, the speaker I consulted allowed for a monosyllabic parse of [tái] "taro" but not [dá*(.)ə] "earth," an asymmetry described by Chen (2006: 211–4). Taking this on its face, it would seem that the two words had different prosodic structures, and that certain sequences like [tái] are able to form heavy syllables.

However, this was not the whole story. Without any prompting further than being asked how many syllables were in the word *tái* "taro," the speaker offered multiple parses. To paraphrase her response: "the word can be pronounced as one syllable [tái] [or [táj]], or as two syllables, [tá.ʔi] or [tá.i]." The two-syllable parses were provided at a slower speech rate; spectrograms of these tokens were shown in Figure 7. The speaker then explained that the number of syllables in *tái* was dependent on speech rate.

The variability of parsing *tái* "taro" was not extended to *dáe* "earth"; the speaker maintained that the /aə/ sequence could not be parsed tautosyllabically.

Importantly, the monosyllabic parse of *tái* "taro" and disyllabic parse of *dáe* "earth" are in line with what has been reported in the literature. It is only the alternative disyllabic parses of *tái* that clash with earlier descriptions of Budai Rukai.

## 5.2 Contact language interference in the direct elicitation of syllables

If Budai Rukai phonology points to a maximal CV syllable, then why should speakers be able to parse larger CVV(V) "syllables" in a direct elicitation task? The most likely cause of this discrepancy lies in the languages in contact with Budai Rukai, namely Mandarin and Taiwanese Southern Min, the two largest Sinitic languages spoken in Taiwan. The speaker consulted for the current study, like most (if not all) speakers of

Budai Rukai, speaks both of these languages alongside her tribal language.[24] The speaker was also, at the time of the interviews, a graduate student (in a field other than linguistics) and had undergone many years of education using Mandarin as the medium of instruction. Like much work on Formosan linguistics, the interviews in which data were elicited were conducted in Mandarin.[25]

The first commonality between the directly elicited Budai Rukai syllable structures and the Sinitic contact languages is that both Mandarin and Southern Min allow for large syllables containing up to three vocoids in a row. Examples of the possible CVV and CVVV (where V can be a vowel or glide) syllables in these languages can be seen in (9–12):[26]

(9)       Two-vocoid sequences in surface structure                    (Mandarin)

       a.    (C)aI: 來 *lái* [lǎi] "come"
       b.    (C)eI: 內 *nèi* [nêi] "inside"
       c.    (C)aU: 高 *gāo* [káu] "tall"
       d.    (C)oU: 肉 *ròu* [ʐôu] "meat"
       e.    (C)Ia(ŋ): 下 *xià* [ɕiâ] "down"
       f.    (C)Iɛ(n): 鞋 *xié* [ɕiě] "shoes"
       g.    (C)Ioŋ: 用 *yòng* [jôŋ] "use"[27]
       h.    (C)Ua(N): 掛 *guà* [kuâ] "hang"

---

**24** There is also language contact and bilingualism between Formosan languages: Chen (2011) notes the transfer of certain features from Paiwan to Budai Rukai in a village where both are spoken. As I don't believe that contact with other Formosan languages has affected the syllable structure intuitions discussed in this paper, I will not discuss it in detail here.

**25** Scholars from Taiwan are likely to be proficient in Taiwanese Southern Min and use that language rather than Mandarin during interviews. For many non-Taiwanese scholars like myself, Mandarin is a more accessible language to learn and use in interviews.

**26** Uppercase I, U, and Y used here to refer to high front, high back, and high front rounded vocoids, respectively (be they full vowels or glides). Superscripted "ⁿ" in Southern Min refers to nasalization of the preceding vowel(s) within the syllable. Transcriptions in italics are Pinyin for Mandarin and Pèh-ōe-jī for Southern Min; note the difference between these and IPA for tonal markings. Items here are taken from dictionaries: Harbaugh (1998) for Mandarin and Maryknoll Taiwan (2013) for Southern Min. Transcriptions in square brackets [ ] are meant to roughly represent surface structure rather than underlying structure (which can be rather abstract and vary across scholars for both of these languages). These phonetic structures are roughly based on Duanmu (2007) for Mandarin and Lin (2015) for Southern Min, with nonsyllable-initial vocoids transcribed as full vowels to remain agnostic to the diverse vowel/glide analyses in the literature. As an additional note, while an anonymous reviewer on a previous version of this paper notes a variety of models of the Mandarin syllable in the literature, to my knowledge none challenge the tautosyllabicity of the sequences in (9)–(10).

**27** Other descriptions of this vocalic sequence include [iʊ] and [yə]; see Duanmu (2007: 55) and sources therein for discussion.

     i.    (C)Uən/(*C)Uəŋ: 盾 *dùn* [tûən] "shield"; 甕 *wèng* [wə̂ŋ] "urn"

     j.    (C)Uo: 甕 *duō* [tuó] "many"

     k.    (C)Ɣɛ(n): 雪 *xuě* [ɕɣɛ̀] "snow"

(10)    Three-vocoid sequences in surface structure
     a.    IaI: 崖 *yái* [jǎi] "cliff"[28]
     b.    (C)IaU: 小 *xiǎo* [ɕiàu] "small"
     c.    (C)IoU: 六 *liù* [liôu] "six"
     d.    (C)UaI: 帅 *shuài* [ʂuâi] "handsome"
     e.    (C)UeI: 水 *shuǐ* [ʂuèi] "water"

(11)    Two-vocoid sequences in surface structure (Southern Min)
     a.    (C)aI(ⁿ): 利 *lāi* [lāi] "sharp"
     b.    (C)aU: 老 *lāu* [lāu] "old"
     c.    (C)Ia(C/ⁿ): 名 *mîa* [miǎ] "name"
     d.    (C)Iɛ*(n/t): 烈 *liảt* [liɛ́t] "intense"
     e.    (C)Io(C): 燒 *sio* [sió] "warm"
     f.    (C)IU(ⁿ): 球 *kiû* [kiǔ] "ball"
     g.    (C)Ua(C/ⁿ): 誇 *khoa* [kʰuá] "boast"
     h.    (C)Ue(ʔ): 灰 *hoe* [hué] "ashes"
     i.    (C)UI: 貴 *kùi* [kuì] "expensive"

(12)    Three-vocoid sequences in surface structure
     a.    (C)IaU: 跳 *thiàu* [tʰiàu] "jump"
     b.    (C)UaI(ⁿ): 乖 *koai* [kuái] "well-behaved"

As shown in these examples, CVV and CVVV syllables in Mandarin and Southern Min are not marignal syllable types: they represent both a large portion of the phonotactically licit syllables in these languages and the syllable shapes of many basic vocabulary items and most frequently used words.

    When syllable structure in Budai Rukai was discussed with the speaker consulted for this study, the word used for "syllable" in the interviews was the Mandarin word 音節 *yīnjié*. The speaker was already familiar with this word and likely associated it with a set of structures in Mandarin (and Southern Min) that include those listed in (9)–(12). In Sinitic languages, the syllable is important in phonology: morphemes are most often exactly one syllable, the Chinese characters used to write the languages generally correspond to one syllable, the syllable is the domain of lexical tone, and the target of many phonological operations that depend on tone. The

---

**28** Pronunciation in Taiwan Mandarin. Mainland China uses the pronunciation *yá* [já] (Harbaugh 1998).

syllable as defined by Sinitic languages is one that speakers of Budai Rukai can and do access regularly, as bi- and trilingual (or more) speakers of these languages.

Another commonality between speaker intuitions on Budai Rukai syllable structure and the phonology of Sinitic languages is the pattern shown in (13), namely that sequences like /ai/ allow for tautosyllabic or heterosyllabic parses, while /aə/ sequences must be heterosyllabic:

(13)          Monosyllabic    Bisyllabic
    /ai/     [tai]           [ta.i]
    /aə/     not permitted   [da.ə]

Sinitic languages have structures that roughly correspond to both tautosyllabic [ai] and heterosyllabic [a.i] sequences. For example, Mandarin 愛 *ài* [aî] "love" contains a tautosyllabic [ai] sequence while 阿姨 *āyí* [á.(j)ǐ] "maternal aunt" has a similar sequence that is necessarily disyllabic. The same goes for [au] sequences, which also receive variable parses in the existing Budai Rukai literature: Mandarin has both monosyllabic 小 *xiǎo* [ɕiàu] "small" and disyllabic 下午 *xiàwǔ* [ɕiâ.(w)ù] "afternoon."

What Mandarin and Southern Min do not have are tautosyllabic sequences that resemble Budai Rukai /aə/. There are some vocalic sequences that have been transcribed with [ə], notably the [uə] in Mandarin words like 盾 *dùn* [tûən] "shield," in which it is a short transitional vowel between the nuclear [u] and coda [n]. The closest standalone vowel to Budai Rukai /ə/ by phonetic quality is /ɤ/, used by Mandarin in words like 熱 *rè* [ʐɤ̂] "hot," and in certain dialects of Southern Min in words like 好 *hó* [hɤ̂] "good." This vowel notably does not form diphthongs with the other vowel qualities; sequences like /aɤ/ are only found across syllable boundaries in words like Mandarin 差額 *chāé* [tʂʰá.ɤ̌] "difference in amount." If speakers of Budai Rukai like the one consulted for this study liken /ə/ to Mandarin and Southern Min /ɤ/ and are providing intuitions on syllable structure influenced by these languages, then it is no surprise that this speaker (and those consulted for other studies) would rate poorly any tautosyllabic sequence of /ə/ and another vowel.

In sum, the presence of heavy CVV(V) syllables, the variability in parsing sequences like [ai] tauto- versus heterosyllabically, and the ability for /a/, /i/, /u/, but not /ə/ to form diphthongs, have previously been described as features of Budai Rukai when in fact they better reflect the phonotactics of Sinitic contact languages. The reason these features appear when speakers of Budai Rukai are asked to parse words into syllables is that these speakers have learned what a "syllable" is while attending school in Mandarin and thus associate the concept of "syllable" in general with the set of segmental sequences that form licit syllables in Mandarin (and Southern Min by extension).

# 6 How widespread is the problem?

This section will comment on the breadth of misreported syllable structures, first through other findings within the Formosan languages (Section 6.1), and then with a discussion on the conditions under which syllable structure may be misreported (Section 6.2).

## 6.1 Syllable structure reanalyses beyond Budai Rukai

Budai Rukai was not the only language examined in the current study, and it is not the only one for which patterns in intonational phonology have indicated syllable structures that differ from those presented in the literature (presumably based on direct elicitation tasks). Of the fourteen Formosan varieties studied (Budai/Mantauran Rukai, Pnguu Tsou, Kanakanavu, Saaroa, Tjaylaking/Piuma Paiwan, PatRungan Kavalan, Southern Amis, Isbukun Bunun, Tgdaya/Toda/Truku Seediq, and Pazeh),[29] ten were found to have syllable structures that contradicted existing models in the literature. These ten varieties are shown in Table 3, alongside the syllable structures that were reanalyzed based on intonational phonology in Macaulay (2021).[30]

In addition to Budai Rukai, a number of other nearby Formosan languages also have been described with heavy syllables, only to reveal upon reexamination that (C)V is both the tone-bearing unit and the unit referred to by the stress assignment algorithm. This was the case for: Mantauran Rukai, which has been described with CVV syllables with long vowels (Zeitoun 1997: 163), and with CVG syllables with coda glides (Zeitoun 2007: 23); Kanakanavu, which has been described with CGV and CVG syllables in surface forms (Sung 2016: 16); and Saaroa, which has been described with CVV syllables with a tautosyllabic long vowel (Pan 2012: 30–1). In all cases, these CVV, CGV, and CVG syllables have been reanalyzed as CV.V sequences, with CV the maximal syllable in each language. Additionally, in Tsou, what Chang and Pan (2016)

---

**29** Note that there are several Formosan languages for which novel data were not elicited for this study, including Atayal, Saisiyat, Thao, Sakizaya, and Puyuma. Additionally, only 1–2 dialects of each language were studied, while most have numerous other dialects with their own phonological systems.

**30** Since I cannot reproduce all of the data necessary to argue for these reanalyses as thoroughly as I have done for Budai Rukai in this paper, I refer the interested reader to see the discussion in Macaulay (2021), which has a dedicated section for syllable structure within each language's chapter, which is a "from-scratch" analysis based on novel elicited data using similar evidence and reasoning as presented in this paper for Budai Rukai. In this table, the optionality of onset consonants is left unmarked for brevity.

**Table 3:** Reanalyses of syllable structure in Formosan languages by Macaulay (2021), based on novel data from intonational phonology.

| Language | Previous studies | | Reanalysis |
|---|---|---|---|
| | Syll. type | Source | |
| Budai Rukai | CVV | Chen (2006) | CV.V |
| | CVVV | Liu (2011) | CV.V.V |
| Mantauran Rukai | CVV | Zeitoun (1997) | CV.V |
| | CVG | Zeitoun (2007) | CV.V |
| Tsou | CVV | Chang and Pan (2016) | CGV, CVG |
| Kanakanavu | CGV, CVG | Sung (2016) | CV(.)V |
| Saaroa | CVV | Pan (2012) | CV.V |
| Piuma Paiwan | CVVC | Chen (2006) | CV.VC |
| Kavalan | CV.VC | Hsieh (2016) | CV.VC, CVVC |
| | CV[ʔ]# | Li and Tsuchida (2006) | CV[ʔ]#, CV# |
| Amis | CV[ʔ]# | Chen (2017) | CV[ʔ]#, CV# |
| Bunun | CGV, CVG | Huang and Shih (2016) | CV(.)V |
| Pazeh | CVG | Blust (1999) | CVV |
| | CVV | | CGV |
| | CV | | CV, CVV |

describe as heavy CVV syllables contain sequences of a vowel plus the mid glide [e̞], which cannot bear the pitch accent, and is skipped over by the stress assignment algorithm.[31] These syllables are thus not heavy CVV syllables, but CGV and CVG syllables, leaving Tsou with no syllable type including two full vowels or tone-bearing units (much like its neighboring languages Kanakanavu, Saaroa, and Budai/Mantauran Rukai).

There were two cases in which languages were found to allow heavy syllables, but a subset of the heavy syllables described in the literature was reanalyzed as disyllabic sequences. These were Piuma Paiwan, which was described with CVVC syllables (Chen 2006), reanalyzed to CV.VC, and Bunun, in which many of the CGV and CVG syllables described by Huang and Shih (2016) are bimoraic and must be either CVV or CV.V sequences.

In two other cases, heavy syllables were found where the literature had not described them. One was in Kavalan, which has ultimate stress, but also words like *qayzúan* "inhabit" (with the pitch accent aligned to [u]), for which the only possible parse maintaining ultimate stress is [qay.zúan] with a CVVC syllable. This goes

---

31 For discussion on [e̞]'s phonetic properties, see Macaulay (2021: 172–6).

against Hsieh's (2016: 18–9) analysis, in which all sequences of two vocoids must be either heterosyllabic (V.V) or undergo glide formation (GV or VG).

The other case where reanalysis resulted in new heavy syllables is Pazeh, which Blust (1999) analyzes as having CV, CVG, and $CV_1V_2$ syllables. Macaulay (2020) finds that there are glides in the onset but not the coda, and that there are tautosyllabic long vowels (which Blust excludes). This analysis is based on the presence of secondary pitch accents, which are assigned to nonfinal syllables with two full vowels in the nucleus (long $CV_1V_1$ and $CV_1V_2$ from Blust's CVG). The nonfinal syllables that do not attract pitch accents are those which have a single nuclear vowel: short CV, CVC, and CGV ($CV_1V_2$ in Blust's analysis). In sum, Blust's CVG syllables are reanalyzed as $CV_1V_2$, his $CV_1V_2$ syllables as CGV, and $CV_1V_1$ is added to the inventory of syllable types in the language.

Another type of reanalysis concerned the final syllable in Amis and Kavalan. Both of these languages have been described with a glottal stop epenthesis rule that applies at the end of the word or perhaps a larger prosodic domain: Chen (2017: 54–7) for Amis and Li and Tsuchida (2006: 2) for Kavalan. This rule, as previously described, prohibits final open CV syllables, replacing them with CVC syllables. Upon further investigation, it was found that glottal stop epenthesis only occurs when certain intonational contours are present (and by extension, only occurs in certain utterance types). Because of this, open CV syllabes are found utterance-finally in interrogatives and other related utterance types in both languages, despite being excluded by previous analyses.

## 6.2 What is behind these discrepancies?

In all cases, the original analyses of these languages' syllable structures did not contradict any evidence that had been documented at the time and likely faithfully represented the intuitions provided by the speakers consulted for these documentation projects. The problem is thus not one of academic rigor: these existing studies were conducted by experts from a variety of backgrounds within linguistics, using methodology that is standard in the field. Nor is there some unique property of Formosan languages that would predispose them to documentation errors. Their focus in this paper is merely due to the fact that I have studied them, rather than them having been specifically targeted for this study based on some property of the languages or the literature surrounding them.

Instead, the problem lies in the standard methodology itself, which does not take into account discrepancies between the phonologically active syllable and the "syllable" that results from direct elicitation tasks. Nor does this methodology take into account the extralinguistic factors that lead to these discrepancies. In the case of the Formosan languages, speakers generally have in common the Chinese-language

education that was argued in Section 5.2 to have interfered with the direct elicitation task.[32] The fact that the most common reanalysis in the current study is from heavy CVV(V) syllables (which are licit in the Sinitic contact languages) to shorter, maximum CV syllables, suggests that this particular bias has affected the description of syllables in other Formosan languages as well.

If education in a language other than the target language is a factor that affects the results of elicitation tasks involving syllable structure (and perhaps other speech patterns), then it is endangered and Indigenous languages that are at the highest risk of being improperly documented. In these communities, education is often undergone in the language which is dominant in the region or which is used by the government. If speakers of Indigenous languages gain metalinguistic knowledge (such as a definition of the "syllable") in school, this knowledge will be biased toward the language of instruction and may not reflect structures in the language they speak at home. Sadly, it is precisely these languages that are most in need of descriptive work.

This does not mean, however, that speech communities whose languages are used for education are safe from extralinguistic factors affecting the results of elicitation tasks. Language patterns (including syllable structure) change naturally across generations and see variation across individual speakers. A fixed model of the "syllable" that is taught to students in school cannot account for language change or variation, and furthermore is liable to import errors or controversial facts from descriptive linguistic works.

So, how widespread is the problem? Given that it was the documentation of intonational phonology that revealed the issue with the direct elicitation of syllable structure, the issue of misreported (or at least, underargumented) models of syllable structure likely persists in those language groups whose intonational phonologies are un- or underdocumented. This unfortunately includes most of the world's languages, as modeling intonation has not become standard practice in general descriptive works like grammars.

# 7 Discussion

As speakers without linguistic training can often provide intuitions as to how their language's words are broken up into syllables, it has been taken for granted by many linguists that these intuitions can be directly reported in descriptive works. But many

---

**32** While a couple of the participants were old enough to have undergone some years of education in Japanese during the Japanese occupation of Taiwan, none that I am aware of underwent no Chinese-language education.

speakers consulted for these works are not as lacking in linguistic training as field linguists have assumed: the parsing of words into syllables can be a topic covered in education or find its way into common metalinguistic knowledge from other cultural practices.

So what happens when speakers are bilingual and participate in the cultural practices of languages with divergent definitions of the syllable? This is the case with speakers of languages like Budai Rukai: because they live in Taiwan where Sinitic languages like Mandarin and Southern Min have come to dominate education and other aspects of everyday life, the "syllable" becomes a concept defined in the context of these languages. When a speaker of Budai Rukai is asked to parse words in their language into "syllables" they are happy to do so: but the results are likely to be Budai Rukai words passed through a Sinitic syllable template.

This is a risk faced in all fieldwork in phonology. Languages spoken in an area where Sinitic languages like Mandarin are dominant may be especially susceptible to this type of interference due to their strongly defined and widely known definition of the syllable. However, there is nothing preventing definitions of syllable structure from other regionally dominant languages like English, French, Spanish, and so on, from being accessed by speakers of local languages in direct elicitation tasks.

Descriptive works are incredibly important, especially given the severity of language endangerment in the world today, and it is crucial that these works are reporting accurate information about the languages they describe. In order to ensure that syllable structure is being reported accurately, two things must happen. First, models of syllable structure should not be appearing in descriptive works without mention of how the data were elicited and analyzed, and what kinds of alternatives were ruled out in the process. This type of presentation has been normalized for other structures in phonology, such as phonemic contrasts, which now appear alongside lists of minimal pairs in nearly all published grammars.

The second change that needs to happen is that the use of indirect evidence from phonology must be centered over direct elicitation in the analysis of syllable structure. One promising source for indirect evidence is intonational phonology, a topic often sorely overlooked itself in descriptive works. This paper has shown some ways in which patterns in intonation can be used to argue for or against specific types of syllable structure, for example, by identifying the location of pitch accents and domain of accent, as well as the effect of affixation on where pitch accents are anchored. This also goes to show the drawbacks of trying to model segmental phonology and prosodic structure in a vacuum, without investigating the workings of a language's intonational phonology.

By recognizing the flaws in direct elicitation as a method of identifying syllable boundaries, and of the importance of incorporating data from intonation into general models of phonology, the quality of descriptive works will improve, and this

effect will trickle down to future reanalyses and typological works that draw on descriptive works as a data source.

# List of symbols

| | |
|---|---|
| ω | phonological word |
| ∑ | metrical foot |
| σ | syllable |
| μ | mora |
| τ | tap |
| C | consonant |
| V | vowel |
| G | glide |
| R | sonorant |
| T | obstruent |
| H | high tone |
| M | mid tone |
| L | low tone |

# References

Basbøll, Hans. 1985. *Stød* in modern Danish. *Folia Linguistica* 19. 1–50.

Beckman, Mary E. & Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.

Blevins, Juliette. 1995. The syllable in phonological theory. In John Goldsmith (ed.), *Handbook of phonological theory*, 206–244. London: Basil Blackwell.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blust, Robert. 1997. Rukai stress revisited. *Oceanic Linguistics* 36. 398–403.

Blust, Robert. 1998. In defense of Dempwolff: Austronesian diphthongs once again. *Oceanic Linguistics* 37. 354–362.

Blust, Robert. 1999. Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38. 321–365.

Blust, Robert. 2003. *Thao dictionary*. Taipei: Academia Sinica.

Casali, Roderic F. 1995. NCs in Moghamo: Prenasalized onsets or heterosyllabic clusters? *Studies in African Linguistics* 24. 151–165.

Chang, Anna Hsiou-Chuan. 2006. *A reference grammar of Paiwan*. Canberra: The Australian National University PhD thesis.

Chang, Henry Y. & Chia-Jung Pan. 2016. 鄒語語法概論 *[Outline of Tsou Grammar]*. New Taipei City, Taiwan: Council of Indigenous Peoples.

Chen, Chun-Mei. 2006. *A comparative study on Formosan phonology: Paiwan and Budai Rukai*. Austin, TX: University of Texas at Austin PhD thesis.

Chen, Chun-Mei. 2011. Phonetic evidence for the contact-induced prosody in Budai Rukai. *Concentric: Studies in Linguistics* 37. 123–154.

Chen, Wei-Shan. 2017. *Phonological status of epiglottal stops, glottal stops and glides in the northern dialect of Amis*. Hsinchu, Taiwan: National Tsing Hua University Master's thesis.

Cheung, Kwan-Hin. 1986. *The phonology of present-day Cantonese*. London: University College London PhD thesis.

Chitoran, Ioana. 2002. A perception-production study of Romanian diphthongs and glide-vowel sequences. *Journal of the International Phonetic Association* 32. 203–222.

Dell, François & Mohamed Elmedlaoui. 1985. Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7. 105–130.

Depuydt, Leo. 1993. On coptic sounds. *Orientalia* 62. 338–375.

Doornenbal, Marius. 2009. *A grammar of Bantawa. Grammar, paradigm tables, glossary and texts of a Rai language of Eastern Nepal*. Utrecht: LOT.

Duanmu, San. 1990. *A formal study of syllable, tone, stress and domain in Chinese languages*. Cambridge, MA: Massachusetts Institute of Technology PhD thesis.

Duanmu, San. 2007. *The phonology of standard Chinese*, 2nd edn. Oxford: Oxford University Press.

Eurostat. 2022. Number of foreign languages known (self-reported) by sex. Data set. Available at: https://ec.europa.eu/eurostat/databrowser/view/edat_aes_l21/.

Everett, Dan & Keren Everett. 1984. On the relevance of syllable onsets to stress placement. *Linguistic Inquiry* 15. 705–711.

Greenberg, Joseph H. 1962. Is the vowel–consonant dichotomy universal? *Word* 18. 73–81.

Hammond, Michael. 1997. Vowel quantity and syllabification in English. *Language* 73. 1–17.

Harbaugh, Rick. 1998. *Chinese characters: A genealogy and dictionary*. New Haven: Yale Far Eastern Publications. Available at: zhongwen.com.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago, IL: The University of Chicago Press.

Heath, Jeffrey. 1998. *A grammar of Koyra Chiini. The Songhay of Timbuktu*. Berlin: Mouton de Gruyter.

Hsieh, Fuhui. 2016. 噶瑪蘭語語法概論 *[Outline of Kavalan Grammar]*. New Taipei City, Taiwan: Council of Indigenous Peoples.

Huang, Tung-Chiou & A-Zhong Lai. 2012. 魯凱語霧台方言詞典 [A dictionary of Budai Rukai]. Available at: http://e-dictionary.apc.gov.tw/dru/search.htm.

Huang, Hui-Chuan & Chao-Kai Shih. 2016. 布農語語法概論 *[Outline of Bunun Grammar]*. New Taipei City, Taiwan: Council of Indigenous Peoples.

Hyman, Larry M. 1983. Are there syllables in Gokana? In Jonathan Kaye, Hilda Koopman, Dominique Sportiche & André Dugas (eds.), *Current approaches to African linguistics*, vol. 2, 171–179. Dordrecht: Foris.

Hyman, John. 1985. *A theory of phonological weight*. Dordrecht: Foris.

Hyman, Larry M. 2011. Does Gokana really have no syllables? Or: what's so great about being universal? *Phonology* 28. 55–85.

Iacoponi, Luca. 2014. Tone-TBU disparity in three phonologies of Cantonese. In John Kingston, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.), *Proceedings of the 2013 annual meeting on phonology*, 1–10. Washington DC: Linguistic Society of America.

Jun, Sun-Ah (ed.). 2005. *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Jun, Sun-Ah (ed.). 2014. *Prosodic typology II: The phonology of intonation and phrasing*. Oxford: Oxford University Press.

Kawahara, Shigeto. 2016. Japanese has syllables: A reply to Labrune. *Phonology* 33. 159–194.

Klamer, Marian. 2010. *A grammar of Teiwa*. Berlin: Mouton de Gruyter.

Kubozono, Haruo. 1989. The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech* 32. 249–278.

Kubozono, Haruo. 2003. The syllable as a unit of prosodic organization in Japanese. In Caroline Féry & Ruben van de Vijver (eds.), *The syllable in optimality theory*, 99–122. Cambridge: Cambridge University Press.

Kubozono, Haruo. 2011. Japanese pitch accent. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 2879–2907. Chichester, UK: Wiley-Blackwell.

Labov, William, Malcah Yaeger & Richard Steiner. 1972. *A quantitative study of sound change in progress*. Philadelphia, PA: U.S. Regional Survey.

Labrune, Laurence. 2012. Questioning the universality of the syllable: Evidence from Japanese. *Phonology* 29. 113–152.

Li, Paul Jen-Kuei. 1973. *Rukai structure*. Taipei: Academia Sinica.

Li, Paul Jen-Kuei. 1979. Variations in the Tsou dialects. *Bulletin of the Institute of History and Philology* 47. 273–300.

Li, Paul Jen-Kuei. 2008. The endangered languages in Taiwan. Keynote speech given at EATS2008, Charles University and SOAS Centre of Taiwan Studies, Prague, April 18, 2008.

Li, Paul Jen-Kuei & Shigeru Tsuchida. 2006. *Kavalan dictionary*. Taipei: Academia Sinica.

Liddell, Scott K. 1984. THINK and BELIEVE: Sequentiality in American Sign Language. *Language* 60. 372–399.

Lin, Philip T. 2015. *Taiwanese grammar: A concise reference*. US: Greenhorn Media.

Liu, Chong-Yu Harry. 2011. *Echo vowels in Budai Rukai*. Hsinchu: National Tsing Hua University Master's thesis.

Macaulay, Benjamin K. 2020. The prosodic structure of Pazeh. *AFLA* 26, 175–191.

Macaulay, Benjamin K. 2021. *Prosody and intonation in Formosan languages*. New York: The Graduate Center, City University of New York PhD thesis.

Macaulay, Benjamin K. 2023. Intonation in Formosan languages. In Paul Jen-Kuei Li, Elizabeth Zeitoun & Rik De Busser (eds.), *Handbook of Formosan languages: The indigenous languages of Taiwan*, 334–370. Leiden: Brill.

Maryknoll Taiwan. 2013. *Taiwanese-English dictionary & English-Taiwanese dictionary*. Taichung City, Taiwan: Maryknoll Language Service Center. Available at: mkdict.net.

McQueen, James M., Takashi Otake & Anne Cutler. 2001. Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language* 45. 103–132.

Mücke, Doris, Hosung Nam, Anne Hermes & Louis Goldstein. 2012. Coupling of tone and constriction gestures in pitch accents. In Phil Hoole (ed.), *Consonant clusters and structural complexity*, 205–230. Berlin: Mouton de Gruyter.

Ohala, John J. 2008. The emergent syllable. In Barbara L. Davis & Krisztina Zajdó (eds.), *The syllable in speech production*, 215–222. New York: Lawrence Erlbaum.

Pan, Chia-Jung. 2012. *A grammar of Lha'alua, an Austronesian language of Taiwan*. Townsville, Australia: James Cook University PhD thesis.

Parker, Stephen G. 2002. *Quantifying the sonority hierarchy*. Amherst, MA: University of Massachusetts at Amherst PhD thesis.

Pellard, Thomas. 2009. *Ōgami: Éléments de description d'un parler du Sud des Ryūkyū*. Paris: École des hautes études en sciences sociales PhD thesis.

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. Cambridge, MA: Massachusetts Institute of Technology PhD thesis.

Pierrehumbert, Janet B. & Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.

Piggott, Glyne L. 1995. Epenthesis and syllable weight. *Natural Language and Linguistic Theory* 13. 283–326.

Port, Robert F., Jonathan Dalby & Michael O'Dell. 1987. Evidence for mora timing in Japanese. *The Journal of the Acoustical Society of America* 81. 1574–1585.

Rau, Victoria Der-Hwa. 1992. *A grammar of Atayal*. Taipei: Crane Publishing.

Ross, Malcolm D. 1992. The sound of Proto-Austronesian: An outsider's view of the Formosan evidence. *Oceanic Linguistics* 31. 23–64.

Sung, Li-May. 2016. 卡那卡那富語語法概論 *[Outline of Kanakanavu Grammar]*. New Taipei City, Taiwan: Council of Indigenous Peoples.

Teng, Stacy Fang-Ching. 2008. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Canberra: ANU Press.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology*. Tokyo: Tokyo Institute for the Study of Languages and Cultures of Africa and Asia.

Tsujimura, Natsuko & Stuart Davis. 2008. Rhyme and the reinterpretation of Hip Hop in Japan. In H. Samy Alim, Awad Ibrahim & Alastair Pennycook (eds.), *Global linguistic flows: Hip Hop cultures, youth identities, and the politics of language*, 157–171. London: Routledge.

Tu, Wen-Chiu. 1994. *A synchronic classification of Rukai dialects in Taiwan: A quantitative study of mutual intelligibility*. Champaign, IL: University of Illinois at Urbana-Champaign PhD thesis.

Tung, T'ung-ho. 1964. *A descriptive study of the Tsou language, Formosa*. Taipei: Academia Sinica.

Vance, Timothy J. 2008. *The sounds of Japanese*. New York: Cambridge University Press.

Warner, Natasha & Takayuki Arai. 2001. Japanese mora-timing: A review. *Phonetica* 58. 1–25.

Watson, Ian. 1999. Phonetics, phonologization, and French nasal vowels. *Oxford University Working Papers in Linguistics, Philology & Phonetics* 4.

Westby, Deanna. 1984. Variable productivity: Evidence from the English lax vowel constraint. *Calgary Working Papers in Linguistics* 10. 56–96.

Wurm, Stephen A. 1991. Language death and disappearance: Causes and circumstances. *Diogenes* 39(153). 1–18.

Zeitoun, Elizabeth. 1997. 萬山方言 [The Mantauran Dialect]. In Paul Jen-Kuei Li (ed.), 高雄縣南島語言 *[The Formosan Languages of Kaohsiung County]*, 159–225. Kaohsiung, Taiwan: 高雄縣政府 [Kaohsiung County Government].

Zeitoun, Elizabeth. 2007. *A grammar of Mantauran (Rukai)*. Taipei: Academia Sinica.

Zonneveld, Wim & Mieke Trommelen. 1980. Egg, onion, ouch! on the representation of Dutch diphthongs. In Wim Zonneveld, Frans van Coetsem & Orrin W. Robinson (eds.), *Studies in Dutch phonology*, 265–292. Den Haag: Martinus Nijhoff.