Benjamin Macaulay*

Issues in systematizing the elicitation and analysis of syllable structure

Abstract: This reply explores issues with documenting prosodic structures given their unique relationship with metalinguistic knowledge. New evidence and perspectives are incorporated into a deeper investigation of Budai Rukai and the analytical decisions that underlie its reanalyzed prosodic system. This case study is then taken as a basis for discussion of a possible standardized protocol for the documentation of prosodic structures, in which diverse types of phonological evidence are integrated and contextualized.

Keywords: syllable structure; prosody; language documentation; Austronesian; metalinguistic awareness

1 Introduction

In this volume's target article, I argue that speaker intuitions are overused as evidence for syllable structure in descriptive linguistics, and that these intuitions are more likely to be influenced by metalinguistic factors than other types of evidence for syllable structure. The paper presented a case study of Budai Rukai, a Formosan language for which speaker intuitions converge on a model of the syllable that differs from the one indicated by its intonational phonology, likely due to the use of Chinese in education and fieldwork interviews. The responses to this paper have given a broad range of perspectives, not only on how best to model Budai Rukai's prosodic structure but on how to contextualize diverse types of phonological evidence more generally, as well as how new strategies for investigating syllable structure may be implemented in the field. I extend my deepest gratitude to the authors of the commentaries for their thought-provoking discussion, which I hope will serve as the starting point for a wider discussion on how syllables and other prosodic structures are elicited and reported.

In responding to these papers, I separate the points of discussion into two main themes: issues involving the model of Budai Rukai proposed in the target paper

---

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

(Section 2), and more general points regarding the elicitation process and analysis (Section 3). Section 4 will conclude.

2 Analytical decisions and modeling Budai Rukai prosody

The target paper uses a case study of Budai Rukai prosody in order to support its theoretical claims. This section discusses topics that either specifically involve the reanalysis of Budai Rukai prosody, or concern general analytical decisions that are best exemplified through the Budai Rukai case study. Section 2.1 provides details about the phonetics of stress in Budai Rukai; Section 2.2 provides additional arguments for the generalization of the updated (C)V syllable template to unaccented material; Section 2.3 comments on the feasibility of removing glides from the phoneme inventory; and Section 2.4 discusses the possibility of an analysis of Budai Rukai that does not assume a universal prosodic hierarchy.

2.1 Stress and its realization in Budai Rukai

The target article's reanalysis of Budai Rukai syllable structure takes as a primary piece of evidence the alignment of the pitch accent H* (or H*L in interrogatives). This pitch accent is described as anchoring to the stressed syllable in the word, which requires an analysis of the location of stress in the language. Two authors responding to the target paper bring up alternatives for the analysis of stress in Budai Rukai: Himmelmann (Section 4.1) notes recent evidence from Western Austronesian languages finding that intonational elements are more reliably anchored to prosodic domain edges than any identifiable prominent/"stressed" syllable; Smith (Section 3) questions whether pitch accent may be separate from another identifiable location of stress.

To begin, I will expand on the existing model of Budai Rukai stress. Despite the reanalysis of syllable structure from the CVX syllables of Chen (2006) or Liu (2011) to CV.V, I analyze stress as remaining within a sequence that previous authors have identified as a stressed syllable. That is to say, where previous authors describe a stressed CÝX syllable, I have analyzed either CÝ.V or CV.Ý (not CV.V), and where previous authors describe an unstressed CVX syllable, I have analyzed this as CV.V (not CÝ.V or CV.Ý).

As noted by Smith (Section 3), stress is traditionally analyzed as the confluence of several acoustic features. Hayes (1995: 6) lists four possible features of a stressed syllable: a) increased intensity, b) increased duration, c) anchoring of a pitch movement, and d) identification via speaker judgments. While the inclusion of speaker judgments in this list could be examined in as much detail as this volume

has done for the analysis of syllable structure, I will focus on the three acoustic features here.

Firstly, intensity does not reliably peak in any particular position (for the speaker I consulted, and in the speech styles used in our interviews). As a small example, I have measured the average intensity during each vocoid in several tokens of *laimai* "clothes" and *laimai=li* "my clothes" from the current study, with the results shown in Table 1.

As can be seen here, there is little change in intensity over the course of the word, with few generalizable patterns aside from perhaps a slight lowering in utterance-final position. The vocoid that bears the f0 peak (i_1 in *laimai* and a_2 in *laimai=li*) does not also bear an intensity maximum. Perhaps it might in other speech registers or with other speakers, but while analyzing the three hours of this speaker's data (as well as the archival recordings from Huang and Lai 2012 when they were available), I noticed no reliable patterns in the intensity contour at the word level (either in productions of words in isolation or within larger utterances).

Sмiтн (Section 3) notes possible lengthening in Budai Rukai words on vocoids that do not bear the pitch accent's f0 peak and suggests the possibility that lengthening, rather than accent position, is the true marker of prominence in Budai Rukai, as it is argued to be in many Philippine languages (Kaufman and Himmelmann 2024). I agree with Sмiтн here that there is evidence of the durational and pitch cues of stress having split (from a presumed initial state where they marked the same position). However, I believe that the anchor of the pitch accent melody is a better candidate for the prominent syllable both in Budai Rukai and the other modern Formosan languages I have examined.

I have noted penultimate lengthening in some of the Formosan languages for which the penult is not the accented syllable. For example, during my study of Pazeh (Macaulay 2020), which aligns by default a pitch accent to the final syllable, I noticed

Table 1: Average intensity (in dB) during vocoids from productions in isolation of *laimai*(=*li*) "(my) clothes" in Budai Rukai.

Instance	l	a_1	i_1	m	a_2	i_2	l	i_3
laimai token 1		86	85		84	78		
laimai token 2		80	80		80	76		
laimai token 3		80	80		80	77		
Average		**82**	**82**		**81**	**77**		
laimai=li token 1		87	86		86	87		83
laimai=li token 2		86	85		83	84		83
laimai=li token 3		82	81		82	80		80
Average		**85**	**84**		**84**	**84**		**82**

that some tokens would have an (unaccented) penult longer than other nearby vowels. Examples are shown in Figures 1 and 2, which show annotated pitch tracks of *dalum* [ˀda.dum] "water" and *mulanguy* [mu.ɾa.ŋui] "AF-swim," each of which has a vowel [a] in the penult that is longer than other nearby (including accented) vowels. Interestingly, these lengthened syllables are distinct from syllables with underlying long vowels, which receive an additional pitch accent when present (ibid.).

Even in Pazeh, where this lengthening occurs much more often than in the Budai Rukai data from the current study, I argue that it is less reliable as an indicator of prominence than the pitch accent. Firstly, while the Pazeh data contained the most lengthened unaccented penults in the Formosan data I have surveyed, these lengthened penults are not the norm: not all tokens of these words have lengthened penults, but all have pitch accents aligned with the final syllable of the prosodic word. An additional Pazeh-specific pattern relevant to this discussion is that syllables with complex nuclei in Pazeh (i.e., CV_1V_1 or CV_1V_2 but not CGV or CVC) surface with a second instance of the pitch accent when not in final position. For this reason, I see the pitch accent as tied to the language's metrical structure more closely than the sporadic lengthening seen on the penult.
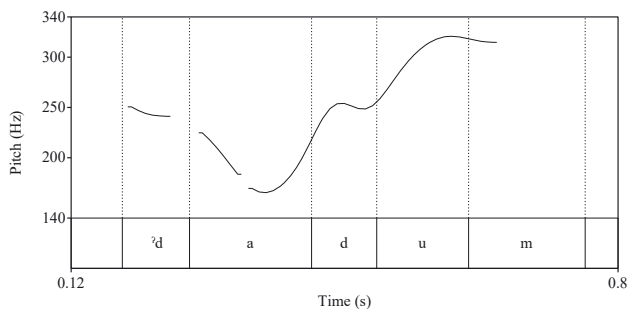
Figure 1: Pitch track of *dalum* "water" produced in isolation in Pazeh.
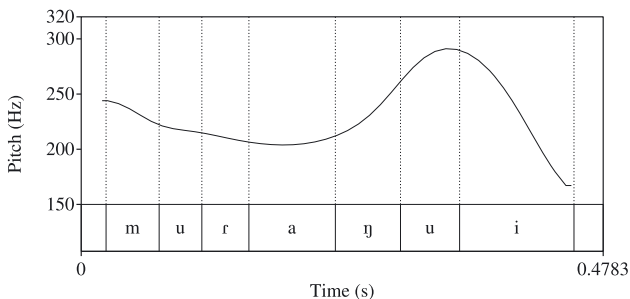
Figure 2: Pitch track of *mulanguy* "AF-swim" produced in isolation in Pazeh.
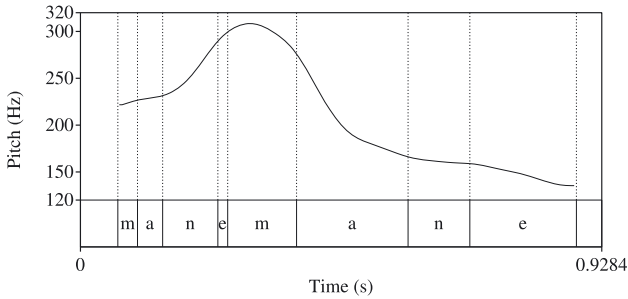
Figure 3: Pitch track of *manemane* "who" produced in isolation in Budai Rukai.

In the Budai Rukai data I collected, I have found only two words produced with lengthened unaccented penults. One is *luange* "cow" from the target article, and the other is *manemane* "who," a pitch track of which is shown in Figure 3.

What both of these have in common is that they are not surfacing with the (L+)H* pitch accent melody used in declarative and sarcastic utterances. Instead, they are surfacing with the (L+)H*L pitch accent associated with interrogatives. In the case of *manemane* "who" this is expected as it is a *wh*-word, and the lengthening disappears when this word is not in IP-final position. The token of *luange* "cow" presented in the target article is actually from a production of the forced-choice utterance shown in (1):

(1) *i-kane=su   ku kuka   luange*
     FUT-eat=2sg.NOM ACC chicken cow
     "Are you going to eat chicken, or beef?"

Forced-choice utterances in Budai Rukai are expressed with an interrogative pitch accent (L+)H*L on each disjunct, which is why *luange* has this pitch accent.[1]

The declarative (L+)H* and interrogative (L+)H*L pitch accent melodies in Budai Rukai reach their f0 peak in roughly the same place, as can be seen in declarative-interrogative minimal pairs (cf. Macaulay 2021a: 150–3). The primary difference between them is that (L+)H*L has a sharp fall in f0 directly after the peak while (L+)H* ends with a steady descent from the peak to the IP-final L% boundary tone. Since all cases of unaccented penultimate lengthening in the Budai Rukai data I elicited are on words that have received the (L+)H*L pitch accent, I suspect that the lengthening is a feature of this pitch accent (or the intonational contour it is associated with) rather than a marker of prominence, pending further investigation on this topic.[2] If so, perhaps an earlier stage of (Budai or Proto-) Rukai underwent the

---

1 Pitch tracks of two tokens of the full utterance can be seen in Macaulay (2021a: 159).

2 As it does not appear on all (L+)H*L-marked words, it would also be an optional feature.
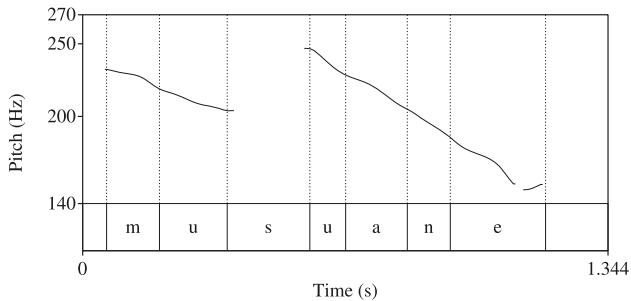
Figure 4: Pitch track of *musuane* "2sg.ᴏʙʟ" produced in isolation in Budai Rukai.

same accent-lengthening split that Pazeh did, and subsequently associated the lengthening with the interrogative pitch accent.

This is not to say that the only evidence for early peaks in sequences like the [ú.a] in *luange* "cow" comes from utterances with interrogative pitch accents. An example of the 2*sg* oblique pronoun *musuane* produced in isolation is shown in Figure 4, including a [ú.a] environment with the declarative pitch accent (L+)H*.

Here, the microprosody of the preceding [s] occludes the exact location of the nuclear H*'s peak; however, it must occur during the second [u], which contains the highest f0 range of any of the vocoids. Moreover, the penultimate [a] does not appear lengthened here (although the boundary at this and other vowel junctures is difficult to precisely identify).

Another reason to see the location of accent as indicative of prominence in Budai Rukai is that its position is lexically determined. The pitch accent anchors to a window that has been described as the antepenult-to-penult (although it may have widened slightly with the target paper's reanalysis). As Hɪᴍᴍᴇʟᴍᴀɴɴ (Section 4.2) notes, after reanalysis there are many antepenultimate-stressed words that end in either an echo vowel or a reanalyzed glide.

To provide some context on echo vowels in Southwest Taiwan (i.e., in Tsou, Saaroa, Kanakanavu, and Rukai), in these languages the echo vowel is not optional, and always surfaces unless affected by another rule (/NV/ sequences optionally surface as [N:] in Budai Rukai, Saaroa, and Kanakanavu, especially word-finally). The vowels are not shorter in duration than underlying vowels, but often longer due to final lengthening. Nor do they have differences in vowel quality compared to underlying vowels, with the exception of /a/ copying to [ə] in Rukai.[3] While I have not asked speakers for judgments on the availability of echo vowels following final glides

---

3 In Tsou, Saaroa, and Kanakanavu, the cognate non-low central vowel is /ɨ/ rather than /ə/. In these languages, final /aC/ surfaces as [aCɨ] where this [ɨ] is identical in quality to underlying /ɨ/. In any case, there is no centralization of echo vowels aside from the a > ə merger in Rukai.

(like *[la.i.ma.jə] for "clothes"), such structures are not attested in any of the data I have collected on Budai Rukai, nor in previous descriptions like Li (1973), Chen (2006), or Liu (2011).[4] In Ross's (1992: 49–50) discussion of the development of echo vowels in Budai Rukai, they are absent on glide-final words and present elsewhere. If there are glides in this position, they do not trigger the insertion of echo vowels.

With this in mind, the wealth of final echo vowels and high vocoids in the target paper's antepenultimate-stressed lexical items does not mean that an analysis of the Budai Rukai lexicon is possible in which all are penultimate-stressed. Given the number of Proto-Austronesian roots that are consonant-final, echo vowels are present in a large amount of the modern Rukai lexicon, including in penultimate-stressed words. Examples of Budai Rukai trisyllables with both $CVCVCV_{echo}$ and $CVCVCV_{echo}$ structures are listed by Ross (1992: 50). If stress assignment happens before echo vowel insertion, as HIMMELMANN suggests, then one might be able to remove "antepenultimate" from the list of stress patterns.[5] In doing so, however, a category of lexical items with final stress emerges to encode words of the type /CVCVC/ > [CVCVCV_{echo}]. While much remains to be tested in terms of the Budai Rukai lexicon and morphophonology, and the secondary effects of the target paper's reanalysis of CXX to CV.V, I suspect that noise has been added to the system rather than reduced.

In sum, Budai Rukai words have a lexically specified prominent syllable to which a pitch accent melody anchors whenever a pitch accent is available (i.e., in IP-final position). While nonaccented syllables may show lengthening, they tend to do so only in certain utterance types, and for this reason, I see lengthening as a less compelling indicator of prominence than pitch accent alignment for Budai Rukai.

2.2 Generalizing the CV template to nonaccented syllables

Both HIMMELMANN (Section 4.2) and SMITH (Section 3) pose an important question about the current study's reanalysis of sequences like [ai] to [a.i]: if the basis of this reanalysis is the alignment of pitch accents in the language, then why should the reanalysis affect vocoids that are not near an accent? For example, why should the word for "clothes" be reanalyzed from [láj.maj] to [la.í.ma.i] rather than [la.í.maj]?

---

4 Regarding the other languages with echo vowels (Tsou, Saaroa, Kanakanavu), descriptions preceding my own study (Macaulay 2021a) generally do not include final glides.

5 There is evidence that echo vowels are excluded from the domain of stress assignment in Mantauran Rukai (Macaulay 2021b). Mantauran has undergone a restructuring of its prosodic system, including a shift from right-aligned to left-aligned stress, and for this reason I refrain from drawing conclusions about Budai based on Mantauran. However, the pattern that HIMMELMANN suggests is attested within the Rukai language cluster.

My initial reasoning for extending the analysis of one syllable per vocoid to all sequences was a general desire for efficiency in phonological models. Segmental phonologists (at least generativists and others who use a similar phoneme/allophone distinction) take as a primary goal of analysis that the phoneme inventory should be as small as possible. For example, the American English segments [t] [d] [ɾ] can be analyzed into a three-phoneme system /t/ /d/ /ɾ/ or, due to [ɾ] having an identifiable conditioning environment and alternations with [t] [d], a two-phoneme system /t/ /d/. Since the two-phoneme system is smaller, this is the more efficient model, and the one taken by many scholars (cf. Chomsky and Halle 1968: 191 for one example of many).

Syllable structures can also be seen as existing in finite, language-specific inventories. This is often how they are presented in descriptive works: Chen's (2006: 212) phonology sketch of Budai Rukai gives the inventory of attested syllable types as V, VV, CV, and CVV as one example. If one considers a language's syllable structure inventory to be a closed set (as phoneme inventories are treated), then an analysis is more desirable if it is able to use a smaller inventory to encode the language's phonological structures. As the small (C)V syllables are generally the least controversial, maximally efficient syllable type inventories will trend toward (C)V-only, which as Szigatvári notes, dovetails with the work of CV phonology scholars. Of course, it is unlikely that all scholars see syllable structures as inventories in the same way as phonemes to begin with. After all, the minimal phoneme inventory is not only most efficient in an abstract way, but it also seeks to reflect those categories used to contrast lexical items, a function rarely served by syllable structure (with some exceptions like the English names *Ida* [ai.da] vs. *Aida* [a.i.da]; Blevins 1995).

While acknowledging these differences between phoneme and syllable structure inventories, I make the assumption that an analysis that includes fewer syllable structures is more efficient. The consequence of this is that the inclusion of each additional syllable type into the model requires justification. In the absence of definitive evidence for max CV versus max CVX models, max CV is thus more desirable under the above theoretical assumptions. For this reason, I took the stance in the target article that the max CV syllable should hold across the language's phonology, rather than only for accented sequences where it is best evidenced. Regarding the unaccented [maI] sequence in *laimai* "clothes," the burden of proof should thus rest on claims that this sequence is tautosyllabic (which introduces an additional syllable type to the inventory), rather than the default assumption that it conforms to the max CV syllable structure evidenced in accented sequences.

This is not to say that the target article has exhausted avenues for investigating the structure of the unaccented [ma(.)I] sequence in the Budai Rukai word for "clothes," and I will explore one more here with the data currently available to me. If [la.í.maj] is the structure of the word for "clothes," then there exist one [i] and one [j] in the word, and phonetic differences between the two should mirror general findings in the acoustic properties of glides versus high vowels. Phonetic studies have found that glides [j] [w] have a narrower constriction between the tongue and palate

than their high-vowel counterparts [i] [u] (Maddieson and Emmorey 1985; Padgett 2008). While I have not directly measured the constriction between the tongue and palate in Budai Rukai speakers, the size of this constriction affects the first formant (F1): a tighter constriction has a lower F1 value (measured in Hertz), while a wider opening in the oral cavity has a higher F1 value. Although F1 is not a perfect measure of oral aperture, it is the phonetic cue measured in both the Maddieson and Emmorey (1985) and Padgett (2008) studies.

With this in mind, I revisited the three tokens each of *laimai* "clothes" and *laimai=li* "my clothes" produced in isolation as part of the current study. Using Praat's formant tracking feature (Boersma and Weenink 2024), I measured the maximum F1 during each instance of [a] and minimum F1 during each instance of [i]/[j]. The results are shown in Table 2:

Table 2: Maximal F1 values of [a] and minimal F1 values of [i]/[j] (in Hz) of productions in isolation of *laimai*(=*li*) "(my) clothes" in Budai Rukai.

Instance	*l*	a_1	i_1	*m*	a_2	i_2	*l*	i_3
laimai token 1		909	447		1,053	530		
laimai token 2		891	434		817	407		
laimai token 3		822	539		792	577		
Average		**874**	**473**		**887**	**505**		
laimai=li token 1		893	490		764	522		370
laimai=li token 2		798	531		856	563		494
laimai=li token 3		850	555		850	566		436
Average		**847**	**525**		**823**	**550**		**433**

Using the notation in Table 2, it is i_2 that is suspected by HIMMELMANN and SMITH to be a glide [j], while i_1 has generally been accepted by the authors responding to the target article as a vowel [i]. Since it is not adjacent to another vowel, i_3 is also unlikely to be controversial if described as a vowel [i]. This means that the expected relative F1 values of these vowels are $i_2 < i_1 = i_3$ if i_2 is a glide, or $i_2 = i_1 = i_3$ if i_2 is a vowel. As seen in Table 2, however, the F1 of i_2 is actually higher than that of i_1 in five of the six tokens examined. This is not an effect of being in utterance-final position, as the effect is found in all three tokens of *laimai=li*, in which i_3 has an even lower F1 (perhaps due to a lack of coarticulation from a preceding low vowel [a]).

While this is too small a sample size to show a definitive difference in the F1 minima of i_1 versus i_2, it is notable that the expected difference is not present in five of six of the tokens, and that the measured differences trend in the opposite direction. As for i_2's heightened F1, I leave it to future studies to confirm this effect

and speculate on its source. In the meantime, I see the above data as in support of the a_2i_2 sequence having the same heterosyllabic structure as a_1i_1,[6] and more generally of the max CV syllable extending to nonaccented material.

Another possibility for investigating vocoid sequences phonetically is brought up by Himmelmann (Section 3): since work in Articulatory Phonology (Browman and Goldstein 1986, inter alia) has found differences in timing relationships beween onset-nucleus and nucleus-coda junctures, the status of a segment as a nucleus (=vowel) versus onset/coda (=consonant) should be able to be determined via articulatory study. While I see the value of this approach to determining syllable structure, and am a proponent of incorporating Articulatory Phonology's findings into general works in phonology, I also foresee some difficulties with applying this method to vocoid sequences. In disambiguating CVG ~ CV.V sequences (such as those in *laimai*), the timing relationships between the VG and V.V sequences would be identical, as both a coda consonant and a following syllable's nucleus are in antiphase with the preceding nucleus. Timing relationships might have more to say in disambiguating CGV ~ CV.V sequences; however, even then, as both G and V are using the dorsal gesture, it may be difficult to determine the amount of temporal overlap between the gestures.

2.3 The issue of glide-consonant alternations

The biggest hurdle in analyzing Budai Rukai without phonemic glides, and the area most sorely in need of re-elicitation, is the fortition of /j/ /w/ but not /i/ /u/ to [ð] [v] in onset position (Liu 2011). Smith (Section 4.1) notes that this is not only a synchronic alternation but is also reflected in diachronic changes to word-internal *y *w and suggests that a model with phonemic /j/ /w/ best suits modern Budai Rukai's synchronic phonology as well.

Of course, it must be the case that glides were present in the inventory of some previous stage of the language, as they were directly inherited from Proto-Austronesian. However, even in earlier stages of Rukai and its neighbors who developed echo vowels (Tsou, Saaroa, Kanakanavu), final glides did not pattern with other final consonants. While all other consonant-final words inherited from PAN received echo vowels, glide-final words did not in Proto-Rukai or Tsou, and echo

---

6 A third possibility in which a_2i_2 represents a tautosyllabic sequence [ai] with two full vowels in the nucleus is excluded here as it predicts a later f0 peak in *laimái=li* "my clothes" than is observed. For discussion of pitch accent alignment in syllables with two full vowels within the Formosan context, see Macaulay's (2021a) discussion of Kavalan (pp. 401–4) and Bunun (pp. 488–93).

vowels appear in Saaroa and Kanakanavu in reflexes of *uy but not for most other final *y *w, as shown in (2) (with echo vowels **bolded**):[7]

(2)   Reflexes of PAn final glides in languages with echo vowels:

Language	PAn *y#	PAn *w#	PAn other C#
Proto-Rukai	*bábuy > *báboy 'wild pig'	*láŋaw > *a [La]Láŋaw 'big fly'	*púluq > *póLok-**o** 'ten'
Tsou	*Cumay > *cmoi* 'black bear'	*Ciqaw > *czou* 'river fish'	*Nusuŋ > *luhŋ**u** 'mortar'
Saaroa	*aCay > *pa-pa-paci* 'kill' (but *Naŋuy > *maka-lhangol**o** 'swim')	*babaw > *i-vavu* 'above, up'	*qabaŋ > *'abang**e** 'boat'
Kanakanavu	*kulay > *kulái* 'insect' (but *Sapuy > *apúl**u* 'fire')	*Ciqaw > *ci'áu* 'river fish'	*daNum > *canúm**u* 'water'

These reflexes of PAn final *y *w show at least some vowel-like behavior as early as when the echo vowels developed, although when exactly that was, and whether it was a single development is not set in stone (Li 1977: 25–6). The vocoids in question have been analyzed as full vowels in modern Tsou, Saaroa, Kanakanavu, and Mantauran Rukai (Zeitoun 2007), leaving Budai Rukai as an odd one out among this language group. Of course, the dearth of phonetic studies on other varieties of Rukai is a current hurdle for this line of inquiry; however, I think it is worth investigating an earlier vocalization of PAn final *y *w.

As for the alternation between [ð] [v] and the high vocoids in modern Budai Rukai, there may be possible ways to encode them without positing glide phonemes /j/ /w/. Sᴍɪᴛʜ (Section 4.1) notes that /ð/ /v/ elsewhere in Budai Rukai have their source in PAn *y *w, just as the alternating segments do. With this in mind, it may be possible to encode the alternating segment as one of the fricative phonemes /ð/ /v/, which vocalizes word-finally or before a suffix starting with a non-low vowel or consonant. Further investigation could exclude this analysis, for example, by identifying /ð/- or /v/-final roots that trigger echo vowel insertion; however, none seem to appear in Chen (2006), Liu (2011), or the data from the current study.

---

7 Data in this set from Proto-Rukai are from Ross (1992), and the rest are from the Austronesian Comparative Dictionary (Blust and Trussel 2020; Blust et al. 2023) with some adaptations to orthography for consistency with current usage.

2.4 Building a prosodic hierarchy for Budai Rukai

The target article, in its reanalysis of Budai Rukai prosody, draws parallels to (Tokyo) Japanese, a language in which small (C)V units are well-evidenced in the language's phonology while the status of larger (C)VX units is controversial and difficult to evidence. In Budai Rukai, I argued in the target article as well as the previous two sections of this reply that heavy syllables are unlikely to exist in the language despite being common responses in direct elicitation tasks with Budai Rukai speakers (in both my own study and Chen's 2006 study, according to her commentary in this volume).

With a reanalyzed syllable tier, the mora tier becomes redundant, since as BLEVINS (Section 2) notes, languages with a maximum syllable of $(C_0)V$ can only have a mora tier isomorphic to the syllable tier. In Section 4.5 of the target paper, I conclude that both levels are present and are isomorphic. This analytical decision was an agnostic one, aimed to appeal to universalist approaches to the prosodic hierarchy in what I believed was a lack of evidence to disambiguate 1) whether one versus both tiers were projected and 2) if only one was projected, whether it was the syllable or mora tier.

BLEVINS (Section 2–3) describes a number of languages that resist an analysis in which the mora, syllable, foot, and prosodic word are all projected and contain a fixed amount of material. With these accounts in mind, it seems that perhaps the question of whether each level of the prosodic hierarchy is projected (and is projected only once) should be part of the analytical process for field and documentation phonologists.

The target article stops short of this step; however, commentary by LABRUNE (Section 4.2) about her own process in analyzing the prosody of Tokyo Japanese has provided new insights that I will incorporate into the updated model of Budai Rukai. Labrune's (2012, inter alia) analysis of Japanese reached the same critical juncture as the target article's analysis of Budai Rukai: a small unit that is maximally CjV (or a moraic nasal or geminate consonant timing slot) is well-evidenced, while a larger CjVX unit that had been described elsewhere as the "syllable" was poorly evidenced. LABRUNE's eventual conclusion is to project a single tier containing the smaller unit, which she labels the mora tier rather than the syllable tier.

LABRUNE (Section 4.2) lists several reasons for choosing to label the tier containing the CjV units as the mora tier rather than the syllable tier. Among them are two that I will repeat here: 1) these units do not maximize their onsets, which I assume refers to cases like the moraic nasal /N/ (e.g. [sa.N.e.N] rather than *[sa.Ne.N] for 三円 *san-en* "three yen"); and 2) these units are "perceived as isochronous," which leans into the use of the mora in many phonological analyses to represent a timing slot. Insofar as

these two properties represent "mora-like" behavior, the presence of the opposite behavior in a domain of the same size can be interpreted as "syllable-like."

Budai Rukai does not have an analog to Japanese's moraic nasal /N/; however, it does have a similar structure that is derived through an optional rule. This rule takes inputs of the form /NV/ and replaces the vowel with an elongation of the nasal: [N:]. The rule is found not only in Budai Rukai but also in two other languages in the same linguistic area, Saaroa and Kanakanavu.[8] In my work on Budai Rukai, most instances of this rule's application have involved /ŋV/ rather than /mV/ or /nV/ (while all commonly undergo the rule in Saaroa and Kanakanavu); however, this may represent a higher frequency of this sequence in the elicited Budai Rukai data. Examples of this rule can be seen in Figure 1 of the target article, which contains pitch tracks of four tokens of *taúpungu* "dog," all of which have a long final [ŋ:] rather than the expected [ŋu]. While the lengthened nasals [N:] in these languages are similar to Japanese /N/ in having a nasal segment take up a timing slot similar to a CV syllable,[9] their distribution differs in one key way. In Japanese, the moraic nasal /N/ can appear before a vowel without resyllabifying into its onset as noted above. However, in Budai Rukai (and Saaroa and Kanakanavu), all tokens of [N:] I have found directly precede either a consonant or the utterance edge. If *[N:V] is indeed unavailable in these languages, then perhaps this is to avoid syllabification into a following onset.

The behavior of the Japanese mora as a timing unit also differs from Budai Rukai's (C)V unit. As noted by Sᴍɪᴛʜ and discussed in Section 2.1 above, this unit can lengthen, sporadically in conjunction with specific intonational contours, as well as before an IP edge. If the (C)V unit is a mora, then this lengthening should introduce an additional mora. However, this added length does not affect the location of pitch accents, while an enclitic like *=li* "1sg.ᴘᴏss" (which would also be one mora in this view) does. These behaviors depart from the Japanese mora, which is more consistent as a timing unit.

With these two pieces of evidence in mind, if Budai Rukai does not project both a mora and syllable tier, it is more likely that it is the syllable tier that houses the (C)V units, rather than the mora tier. The prosodic hierarchy as projected in Budai Rukai is thus as shown in (3) (pending investigation into the projection of the foot):

---

8 This rule has been described for Saaroa and Kanakanavu both in previous works (cf. Pan 2016 for Saaroa and Tsuchida 1976 for Kanakanavu), as well as in the dissertation study from which the target article presents data (Macaulay 2021a).

9 Given that these languages were in a contact situation with Japanese during the Japanese occupation of Taiwan, similarities between Formosan [N:] and Japanese /N/ may reflect change under contact rather than coincidence. However, there is not to my knowledge good evidence of when [N:] developed in Formosan.

(3)

```
              ω
            ╱   ╲
           σ     σ
          ╱ ╲    │
         t  a    i     tái 'sweet potato'
```
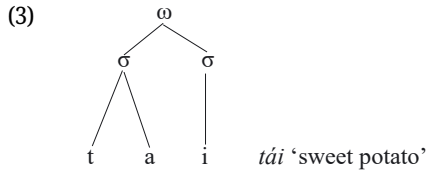
3 The elicitation and analysis of prosodic structure

The target paper raises issues surrounding what has become standard methodology for the elicitation and reporting of syllable structures. As Labrune (Section 2) notes, students of field linguistics are often given end goals for the description of prosodic structures without specific processes by which to reach them. What types of elicitation tasks and analytical decisions would most rigorously lead to a model of a language's prosodic phonology is also controversial, and many possibilities have been explored in the papers in this volume.

A full manual for the prosodic field linguist is outside the scope of this paper. However, I think the commenting authors would agree on two main points: 1) that field linguists should be given some protocol on working with prosodic structures that matches in rigor what they are taught about segmental phonology; 2) that by giving field linguists such a protocol, we could expect future documentation work to be more transparent in how its model of the syllable and other prosodic structures is developed.

In this section, I pick up some of the issues touched on by the commentaries that might shape a future standard protocol in the elicitation, analysis, and reporting of prosodic structures. Section 3.1 defines "direct speaker judgments" in the context of elicitation tasks, while Section 3.2 explores ways of systematizing fieldwork on prosody.

3.1 What is and isn't a direct speaker judgment?

Two types of evidence for prosodic structures are contrasted in the target paper. The type of evidence that the paper argues for, and that which forms the basis of the reanalyzed model of Budai Rukai prosody, is evidence from intonation, more specifically pitch accent alignment. The target paper presents this as "indirect" evidence for syllable structure, as the elicitation task did not prompt the interviewee to draw on their (implicit or explicit) knowledge of syllable structure any more than is done automatically during speech. This is contrasted with the evidence that had formed the basis of the previous model of Budai Rukai (Chen 2006; Liu 2011). Without (at the time) documentation of how the previous model came to be, the target paper

assumes that "direct speaker judgments" played a central role, on the basis of their use in general linguistic works, as well as a replication of the previous model using such methodology.

As SMITH (Section 1) notes, the assumption that previous authors working on Budai Rukai used a specific methodology (in absence of explicit documentation) is a large one. Luckily, CHEN (Section 3) has given us a vivid account of the fieldwork that led to the model of Budai Rukai prosody in her 2006 dissertation. CHEN describes the use of direct speaker judgments on syllabicity, as well as many of the sociolinguistic factors that affected participant responses, and how these shaped the model presented in the dissertation. She also describes the use of some instrumental and intonational evidence in building the model. With this account, a fuller picture of Budai Rukai structure and usage emerges, one in which diverse types of evidence can be accounted for.

With this in mind, HIMMELMANN (Section 2) questions the term "direct speaker judgment": can elicitation tasks and their responses be meaningfully divided into "direct" versus "indirect," and if so, what defines this divide? While the comments on the target paper have broadened my view on the use of direct speaker judgments, I still believe that they are a meaningful subset of elicitation tasks, and that the evidence from them should be considered in a certain light (to be expanded upon in the following section).

One of the end goals of an analysis of syllable structure is to locate the syllable boundaries within words. I would argue that any elicitation task in which the expected output is a word separated into syllables is one that yields direct speaker judgments. The most straightforward version of this is to indicate a specific word to the interviewee and ask them to divide it into syllables, which is possible for many speakers who have learned a term for "syllable" through education or culture. This is the task described in Section 5.1 of the target article, and the one most prone to interference (in the case of Budai Rukai, most likely due to the use of Mandarin in education and the elicitation task). LABRUNE (Section 3.1) also describes the results of this task on French speakers, again with possible interference from education and a prescriptive hyphenation convention.

An interviewer does not have to ask a participant specifically to syllabify for the task to yield such an output, however. BLEVINS (Section 3) notes several instances of speakers being given a task where they are asked to break a word into smaller pieces generally. In this case, the output is argued to be a prosodic constituent, but not necessarily the syllable: Kachok speakers will break material into sesquisyllables (possibly a type of foot) (Olsen 2018), and Nhanda speakers will break a word into two pieces, applying lengthening so that each adheres to the language's minimal word constraint. While the output of these tasks may not be a direct judgment of syllabicity in that speakers may answer with a unit other than the syllable, the "directness" is

still present in that speakers know they are being asked to identify the location of boundaries within a word.

The above tasks may also be performed in reverse, in which an interviewer presents a word that has been separated into smaller units, and the participant is asked whether the boundaries are in the correct position. This task was described in Section 5.1 of the target article, as well as by Chen (Section 3). These tasks are important as they are one of the only possible sources of negative evidence. They also involve "direct" speaker judgments as the above two tasks: even if the interviewer does not explicitly use the term "syllable" to describe the units in which the material is separated, many participants may make the connection independently, and in doing so activate the same metalinguistic knowledge that affects the other "direct" tasks. How likely it is for them to do so depends on other aspects of the task: I expect that participants may react differently to audio versus text stimuli, as well as stimuli with boundaries closer versus further from a "canonical" syllable template, among other factors.

An edge case in between "direct" and "indirect" speaker judgment tasks is one in which speakers are asked to repeat a word more slowly or clearly. The presentation of this task does not explicitly invoke the concept of syllables (or any prosodic domain), but there are multiple ways in which a participant may respond to the task. Some participants may lengthen each syllable nucleus, but not indicate syllable boundaries in any way. Others may separate the word into syllables or other prosodic units with pauses in-between. Speakers may also provide slow- or carefulspeech forms broken into subunits outside of elicitation tasks, as Hanson (2010, as cited by Blevins) describes for speakers of Yine. These forms, spontaneously or elicited, may or may not have been produced by the speaker with prescriptive norms and other metalinguistic knowledge in mind.

On the other end of the spectrum, some types of data that can be elicited and analyzed for markers of syllable structure do not involve the explicit separation of words into syllables. The most extreme example is corpus or archival data, in which the elicitation task (if one was present) may be simply to speak freely. In this data, speakers are "using" syllables in the application of their language's phonological rules, but likely not drawing on metalinguistic knowledge to draw attention to or make claims about the location of syllable boundaries. Some types of evidence cannot be obtained by analyzing this type of data alone, most notably negative evidence and the effects of speech rate (cf. Chen for comments on the effect of speech rate on syllable structure). However, as I argue in the target paper, some types of evidence for syllabic structure, such as pitch accent alignment, are evident in data which was elicited from speakers who were not prompted to identify or manipulate prosodic boundaries. For this reason, I see the use of this kind of evidence as

"indirect" in a study of syllable structure, versus the "direct" elicitation task mentioned above.

Of course, due to the need for diverse types of evidence (and dearth of materials available for many languages), studies in phonology generally do not rely solely on corpora and archival recordings.[10] In the target paper's study, data were elicited primarily via translation tasks of stimuli presented in Mandarin, which consisted of either words in isolation, sentences/single utterances, or short dialogues (the direct speaker judgments described in Section 5.1 were elicited after these data). CHEN (Section 3) describes a similar procedure in which target words are placed in a carrier sentence. This type of data elicitation is not as separated from metalinguistic knowledge as the type of archival data described above, as participants will generally have some knowledge that the interviewer is interested in pronunciation or linguistic structure, depending on how the interviewer presents their project. Additionally, there are other types of interference from metalinguistic knowledge that affect the standard field interview task: interviewees may adjust their register when they know they are taking part in a study, up to and including avoiding the use of stigmatized speech sounds (cf. Everett 2004 for examples from Pirahã).

To summarize, the distinction between "direct" and "indirect" evidence for syllable structure has been drawn here primarily as the two clusters of tasks that elicit them interact differently with speakers' metalinguistic knowledge. Since many speakers have knowledge about the existence of syllables and prescriptive models of the syllable from education or culture, a task which invokes or activates this knowledge is one whose output includes "direct" syllable judgments. While some tasks (like the "speak slower" task) may be interpreted differently by participants, and responded to with or without direct speaker judgments on syllabicity, there is a clear link between tasks that either mention syllables or the separation of words into subunits and the use of metalinguistic knowledge to inform responses. For this reason, descriptive studies of prosody should incorporate "indirect" evidence in models of syllable structure in order to mitigate the interference of metalinguistic knowledge.

3.2 The incorporation of speaker judgments into syllable structure analysis

The target paper takes a firm stance on the relative value of "direct" versus "indirect" evidence for syllable structure. Since tasks that elicit "direct" speaker judgments are

---

10 On the other hand, in the case of dormant languages, it is *only* archival recordings that are available. For an example of a prosodic analysis based solely on archival recordings, see Bowern et al.'s (2013) reanalysis of Yidiny.

more susceptible to interference from metalinguistic knowledge, the target paper suggests that the use of this type of evidence (which is currently standard in field linguistics) be replaced with tasks that yield "indirect" types of evidence.

The commentaries have raised two main objections to this line of reasoning. The first, discussed by Myers (Section 1), is that the output of both types of tasks are systematic reactions to stimuli, and that the output of direct judgment tasks should be analyzed rather than discarded. Yang furthermore points out that any one type of evidence is insufficient on its own (including indirect evidence), and that a multivariate system of evidence better informs descriptive works. The second objection is that there are a number of situations in which direct speaker judgments are the best or only evidence. Blevins (Section 3) notes cases like Hanson's (2010) study of Yine in which speakers would spontaneously break longer words into units like [ko.wt͡ʃo.-ha.ta.tna.kna] "they are fishing again," in which onset clusters like /wt͡ʃ/ are rare cross-linguistically and a maximal violation of the Sonority Sequencing Principle (Clements 1990). By standing in contrast to expected structures, these judgments become compelling evidence for syllable structure. As an additional point in favor of direct speaker judgments, Chen's (Section 3) and the target paper's direct judgment task both included negative evidence, i.e., participants being able to reject specific parses. Without "direct" tasks, negative evidence can only be approximated by statistical analysis of large-scale corpora, which is unavailable for many languages.

In light of these points, I will revisit the target paper's theoretical claims. It seems that many of the commenting authors and I are in agreement on two things: 1) that some elicitation practices are susceptible to interference from metalinguistic knowledge and 2) that standard practices in fieldwork training and publication have not encouraged the use of a broad spectrum of evidence or a transparent analytical process in order to mitigate this issue. With this in mind, we can ask: what should the field linguist do with direct speaker judgments on syllabicity (or other prosodic structures)?

I propose four modifications to the fieldwork process that seek to add consistency and transparency to documentation work in prosody, with commentary below:
– Order tasks within-participant from "indirect" to "direct"
– Formalize metalinguistic knowledge
– Determine the dominant prosodic unit
– Attribute sources of evidence in publication

As discussed in the previous section, elicitation tasks differ in how likely they are to activate certain types of metalinguistic knowledge in the participant. Once this knowledge is activated, there is no guarantee that responses to other tasks later in the elicitation session will not also be affected. One way to mitigate this effect is to order tasks within the elicitation sessions so that "direct" judgment tasks come later

than free response or translation tasks. Of course, a task that involves a direct judgment for one structure may provide indirect evidence for another, and perhaps a translation task will prime the speaker in a way that biases the collection of some other type of data. With this in mind, the order of tasks in an elicitation session will depend on its specific goals.

In the target paper's case study of Budai Rukai, the source of interference in the elicitation task is easily identified: all or nearly all Budai Rukai speakers are bilingual in 1–2 Sinitic languages and have undergone formal education in Mandarin, which has a strict definition of the "syllable." As BLEVINS (Section 3) notes, different speech communities have different circumstances with regard to metalinguistic knowledge, and the direct judgments of syllabicity by Yine speakers in Hanson's (2010) study have little in common with the syllable structure of the dominant language (Spanish). The case of Yine syllabification is especially compelling as it deviates from certain baseline expectations, namely structures in the dominant language and typological norms regarding sonority and syllable structure. For this reason, it would benefit the prosodic fieldworker to go into a study with a set of structures that are expected given contact languages (and the language in which interviews are conducted) as a baseline against which to compare the results of elicitation tasks. In some cases, the relevant metalinguistic knowledge can be investigated directly, for example, for speakers who were familiar with a term for "syllable," they can be asked about the context in which they learned the term, and how it was defined.

BLEVINS (Section 3) describes a number of field studies of languages in which speakers readily divided words into units other than the syllable, including the mora, foot, and minimal word template.[11] For this reason, it cannot be assumed that any template into which speakers readily parse words is the syllable (despite a suggestion to that effect in the target paper). As LABRUNE (Section 4.4) discusses, this has implications for the fieldwork process and can best be approached by identifying this unit (which she gives the label "prosodeme"), before identifying its relationship to the rest of the language's prosodic hierarchy. This initial identification could take a form as BLEVINS describes for her (2001) study of Nhanda and Olsen's (2018) study of Kachok, in which participants are asked (perhaps iteratively) to break words into smaller pieces in a way that feels natural, and the results compared across participants when possible.

---

11 There also exists the possibility that a speech community does not converge on a specific level of the prosodic hierarchy, or in the case of Budai Rukai, that tasks used to identify a "dominant" prosodic unit converge on some template which is not part of the language's phonology. Whether tasks that minimize the influence of Chinese may find the reanalyzed max CV syllable as dominant in Budai Rukai remains to be seen.

Finally, there needs to be some change in how syllable structure is presented in descriptive works. Sмɪтн (Section 1) notes that in my critique of existing models of Budai Rukai, I made several large assumptions about the methodologies used by Chen (2006) and Liu (2011). This is true; however, I would argue that the standard practice of presenting models of syllable structure without the evidence or analytical process that led to it forces the reader to make similar assumptions (at the very least that the study conformed to some set of standard practices, whatever they may be). While it cannot be expected that each author of a grammar or phonology sketch provide as detailed a description as Cнɛɴ has for this volume, I believe that a few small additions may go a long way to allowing the reader to contextualize and evaluate descriptions of syllable structure: 1) whether direct and/or indirect evidence was used; 2) whether any parses were specifically excluded on the basis of negative evidence (rather than simply being unattested); and 3) whether there were any clashes between different types of evidence (as was found in the target paper).

The above proposals are not meant solely as suggestions for the field linguist, but as considerations for the linguistic community as a whole: peer reviewers, conference attendees, and those who teach field methods to new linguists all affect the field's norms and practices. Nor do I want to ignore that any additional demands placed on field linguists are demands on time and resources. As Hɪммɛʟмɑɴɴ (Section 2) notes, there are limits to how much can be investigated in a field setting. One way to mitigate this is to identify types of evidence that can be investigated in data elicited for other purposes. For example, Liu and Huang (2024) have modeled the prosodic structure of Bunun by analyzing pitch patterns in corpus data with methods similar to the target article. With this approach, a broader spectrum of evidence can be obtained without an excessive number of additional elicitation tasks.

At the same time, I believe it is worth the extra effort to ensure that linguistic documentation is consistent and accurate. As Cнɛɴ (Section 3) points out, the structures that linguists include in our works can wind up taking on an "official" status, such as in the Formosan language materials put out by Taiwan's Council of Indigenous Peoples. One such material was Huang and Lai's (2012) dictionary, which presented Budai Rukai with Chinese-like syllable structures as discussed in the target article. What emerges here is a cycle: structures from a dominant language enter descriptive works through biases in the elicitation process, and then are cemented in materials aimed at native language education, and possibly the emerging grammars of young students. As it is highly unlikely that anyone involved in such documentation projects is doing so to promote assimilation to dominant languages, we owe it to ourselves to foresee issues like this in our future works.

To summarize, I have proposed in this section ways in which fieldwork in prosody might be better systematized while hopefully not significantly increasing

the burden on the researcher. I leave a fuller exploration of revising the prosodic fieldwork protocol to future work.

4 Conclusion

While the target paper takes a hard line against the use of direct speaker judgments as evidence for syllable structure, the diverse contributions of the commenting authors have painted a more nuanced picture in which speaker judgments can be contextualized and incorporated alongside indirect evidence into a fuller account of a language's prosodic structure. In this way, interference from metalinguistic knowledge can be acknowledged and mitigated without discarding valuable information provided by speakers.

Not all scholars agree on which analytical decisions will lead to the most efficient or accurate model of a language's prosodic structure. However, it is also clear that there is a need for more standardization in fieldwork practices surrounding prosody. The commentaries, as well as this reply, have given some suggestions as to what better practices might look like. Hopefully, this volume serves as the first step to the linguistic community developing a standard protocol for prosodic documentation.

In addition, the discussion in this volume's commentaries has led to further refinement of the model of Budai Rukai's prosodic structures, and in doing so also unveiled new questions about the language that invite future study. The reanalysis of Budai Rukai throughout this volume highlights the importance of revisiting existing models of prosody with instrumental data and diverse types of evidence and underscores the need to consider language contact, culture, and metalinguistic awareness in documentation work.

References

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

Blevins, Juliette. 2001. *Nhanda: An aboriginal language of Western Australia*. Honolulu: University of Hawai'i Press.

Blust, Robert & Stephen Trussel. 2020. *Austronesian comparative dictionary*. Revised 6/21/2020. Available at: http://www.trussel2.com/ACD.

Blust, Robert, Stephen Trussel & Alexander D. Smith. 2023. *CLDF dataset derived from Blust's "Austronesian Comparative Dictionary" (v1.2)*. Available at: https://doi.org/10.5281/zenodo.7741197.

Boersma, Paul & David Weenink. 2024. Praat: Doing phonetics by computer [Computer program]. Version 6.4.17. Available at: http://www.praat.org/.

Bowern, Claire, Barry Alpher & Erich Round. 2013. Yidiny stress, length, and truncation reconsidered. Poster presentation at *NELS* 44.

Browman, Catherine P. & Louis M. Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3. 219–252.

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

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. Cambridge, MA: MIT Press.

Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In Johan Kingston & Mary Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, 282–333. Cambridge: Cambridge University Press.

Everett, Daniel L. 2004. Coherent fieldwork. In Piet van Sterkenberg (ed.), *Linguistics today: Facing a greater challenge*, 141–162. Amsterdam: John Benjamins.

Hanson, Rebecca. 2010. *A grammar of Yine (Piro)*. Melbourne: Research Centre for Linguistic Typology, La Trobe University PhD thesis.

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

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.

Kaufman, Daniel & Nikolaus P. Himmelmann. 2024. Suprasegmental phonology. In Alexander Adelaar & Antoinette Schapper (eds.), *The Oxford guide to the Western Austronesian languages*. Oxford: Oxford University Press.

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. 1977. The internal relationships of Rukai. *Bulletin of the Institute of History and Philology* 48(1). 1–92.

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

Liu, Roger Cheng-yen & Hui-chuan J. Huang. 2024. Intonational elements and alignment in Bunun dialects. Paper presented at *AFLA* 31.

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

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

Macaulay, Benjamin K. 2021b. Interactions between prosody and morphology in Mantauran Rukai. In Nabil Hathout, Fabio Montermini & Juliette Thuilier (eds.), *Proceedings of the International Symposium of Morphology (ISMo 2021)*, 84–87. Toulouse: Université de Toulouse Jean Jaurès.

Maddieson, Ian & Karen Emmorey. 1985. Relationship between semivowels and vowels: Cross-linguistic investigations of acoustic difference and coarticulation. *Phonetica* 42. 163–174.

Olsen, Emily Long. 2018. *The sound patterns of Kachok in the context of Bahnaric and North-Bahnaric studies*. New York: The Graduate Center, City University of New York PhD thesis.

Padgett, Jaye. 2008. Glides, vowels, and features. *Lingua* 118. 1937–1955.

Pan, Chia-jung. 2016. 拉阿魯哇語語法概論 [Outline of Saaroa grammar]. New Taipei City, Taiwan: Council of Indigenous Peoples.

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

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

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