BOUNDARY STRENGTH AND PREDICTABILITY EFFECTS ON DURATIONAL CUES AT TONE SANDHI GROUP BOUNDARIES IN TAIWAN SOUTHERN MIN

Sheng-Fu Wang

Institute of Linguistics, Academia Sinica, Taiwan
sftwang@gate.sinica.edu.tw

ABSTRACT

This study examined syllable duration at the right edge of Tone Sandhi Groups (TSGs) in an 8-hour spontaneous speech corpus in Taiwan Southern Min. TSG boundaries under three prosodic conditions were compared: matching with an intonational phrase boundary (TSG+IP), an intermediate phrase boundary (TSG+ip), and neither (TSG-only). We also examined whether the correlation between syllable duration and predictability (lexical frequency, surprisal, and informativity) was affected differently at these three boundary conditions. Results showed significant penultimate and final lengthening at TSG boundaries in all conditions, with longer pre-boundary syllables in the TSG+IP condition, followed by the TSG+ip and TSG-only conditions. Analyses also revealed that the positive correlation between syllable duration and surprisal/informativity was weaker towards all three types of boundaries. To summarize, TSG boundaries exhibited durational patterns previously reported for IP boundaries, with an incremental strengthening of final lengthening when overlapping with larger prosodic breaks.

Keywords: predictability, tone sandhi, Taiwan Southern Min, duration, boundary strength

1. INTRODUCTION

This study aimed to investigate the interaction between pre-boundary lengthening and probabilistic reduction based on predictability at the boundaries of Tone Sandhi Groups (henceforth TSGs) in Taiwan Southern Min. There are two specific research questions: First, do the pre-boundary durational patterns incrementally change as TSG boundaries match up with higher prosodic breaks? Second, does the interaction between predictability effects and pre-boundary durational variability also differ for TSG boundaries in different conditions?

Tone sandhi in Taiwan Southern Min refers to a phonological process where all syllables except for the rightmost one in a TSG switch their tones. The formation of TSGs has long been tied to syntactic configurations. Even though there are different accounts [4, 12], they share the assumption that right edges to certain syntactic constituents in some conditions trigger a TSG break.

TSG patterns in two example sentences adapted from [12, 4] are shown in (1), where TSG breaks are marked by the # symbol. Example (1a) shows how the right edges of the noun phrase ti^{44}-baʔ^{2} 'pork' and the quantifier phrase tsit^{4} kun^{44} 'one catty' are marked with TSG breaks, and the tone of the first syllable switches in the surface forms ti^{22}-baʔ^{2} and tsit^{2} kun^{44}. Interestingly, as shown in (1b), it is possible to have a whole sentence consisting of only one TSG, given the right configuration, i.e., the exceptionality of pronouns and phrases that are potentially TSGs being governed [12].

(1) a. ti$^{44\rightarrow22}$-baʔ^{2}# tsit$^{4\rightarrow2}$ kun^{44}# sã$^{44\rightarrow22}$ kʰo^{44}#
pork one catty three dollars
'Pork is three dollars a catty (= 0.6 kg)'

b. i$^{44\rightarrow22}$ kjoŋ$^{24\rightarrow22}$-kjoŋ$^{24\rightarrow22}$ kjo$^{21\rightarrow53}$ gwa$^{53\rightarrow44}$ ke$^{44\rightarrow22}$ kʰwã$^{21\rightarrow53}$ pwã$^{21\rightarrow53}$ tjam$^{53\rightarrow44}$-tsiŋ$^{44\rightarrow22}$ ku$^{53\rightarrow44}$ tsʰeʔ^{2}#
he by-force ask I more read half hour long book
'He insisted that I read for another half an hour.'

For the purpose of this discussion, it is important to note that TSGs can be identified by the tonal changes within a sentence, and their formation is likely governed by syntactic principles. Thus, any changes in TSG formation may result in alterations in meaning that reflect differences in syntactic structure. Examples of this can be seen in (2) from [6, 12]: removing the TSG break by pronouncing the second syllable in mwa^{24}-a^{53} in the base tone results in 'sesame seed' becoming a modifier of the adjective twa^{22}, leading to a change in the phrase's meaning. In summary, TSGs have a distinct phonological definition that likely reflects the syntactic structure.

(2) a. mwa$^{24\rightarrow22}$-a^{53}# twa^{22}# e$^{24\rightarrow22}$
sesames seed big 's
sjo$^{44\rightarrow22}$-piã^{53}#
bun
'buns with big sesame seeds'

b. mwa$^{24\rightarrow22}$-a$^{53\rightarrow44}$ twa^{22}# e$^{24\rightarrow22}$
sesames seed big 's
sjo$^{44\rightarrow22}$-piã^{53}#
bun
'buns as big as sesame seeds' (i.e., tiny buns)

There has been disagreement over whether TSGs are a part of the prosodic hierarchy in Taiwan Southern Min, along with other levels such as syllable, word, intermediate phrase (ip), and intonational phrase (IP). While some studies, such as [16, 9, 10], treat TSGs as a level of prosodic unit below the IP and above the word, other studies and corpora, such as [15, 14, 23], exclude TSGs from these levels of prosodic annotation and treat ip as the level between IP and the word. The separation of TSGs from IP and ip mirrors how they are defined in different terms: As mentioned earlier, TSGs have a clear phonological definition that likely reflects morphosyntactic patterning. On the other hand, IP and ip are mostly defined by potentiality categorical perception of acoustic cues.

It is worth noting that TSG boundaries were shown to be more likely to align with larger prosodic boundaries (i.e., IP and ip, as compared with word and syllable) in a study that directly investigated the relationship between the distribution of TSGs and prosodic units [15]. The authors suggest that this finding indicates that a purely morpho-syntactic approach cannot fully explain the distribution of TSGs in natural language.

To complement the ongoing research on TSG and the prosodic structure in Taiwan Southern Min, we aimed to answer two questions in this study. Firstly, we investigated whether TSGs exhibit pre-boundary lengthening effects similar to those seen in intonational phrases. We went beyond the binary distinction between citation and sandhi tones reported in previous studies [8, 22] by examining syllable duration in final, penultimate, and ante-penultimate positions before a TSG boundary.

Secondly, we investigated whether overlapping with IP and ip boundaries influenced pre-boundary lengthening and its interaction with probabilistic reduction at TSG boundaries. Previous studies have reported that linguistic units with higher lexical frequency or contextual probability tend to be reduced in speech production. One issue related to this correlation is whether the relationship between acoustic cues and predictability is direct or mediated by the prosodic structure [2, 21, 13], and directly examining whether predictability effect covaries with the strength of the prosodic boundary has been used to answer this question [1]. We also explored whether the predictability effects are modulated as a function of a syllable's distance from a TSG boundary, which has been reported as a characteristic of IP boundaries in Taiwan Southern Min [22].

In summary, this study aimed to provide a more detailed description of durational patterning at TSG boundaries. Crucially, we sought to examine whether TSGs themselves exhibit patterns that are observed for higher prosodic units such as the IP, and how overlaps with different levels of prosodic units may change those durational patterns.

2. METHOD

2.1. Speech corpus

Speech data came from an 8-hour spontaneous speech corpus with recordings from 16 speakers. Each speaker contributed approximately 30 minutes of monologue-like speech from sociolinguistic interviews. The corpus included annotations of word segmentation, TSG breaks, and prosodic breaks. A 30-minute sample of the corpus was prosodically annotated by two annotators, who had a high agreement rate of 93.2%. For analysis, we focused on final, penultimate, and ante-penultimate positions in TSGs longer than three syllables. Our dataset included 10,934 TSGs consisting of 46,414 syllables (35,545 words).

2.2. Written corpus and language models

In this study, we estimated informativity and surprisal using trigram language models trained on a written corpus consisting of approximately 4.7 million words (from 6 million syllables). The corpus contained materials from various sources such as fiction, prose, poetry, conversation transcripts, and academic writings, among others [7]. Trigram language models were trained in both directions using the SRILM toolkit [18, 19] with modified Kesner-Ney smoothing [5] to obtain smoothed bigram and unigram probabilities.

2.3. Variables in analyses

The statistical analyses were performed using mixed effects regression models with the LME4 package [3] in R. Three types of predictability variables were examined: **unigram surprisal** ($-\log P(x)$), which measured lexical frequencies; bigram surprisal ($-\log P(x|context)$), which measured local contextual predictability, and **bigram informativity** ($-\sum_{\text{context}} P(\text{context}|x) \log P(x|\text{context})$), the average of a word's surprisal across all contexts. Bigram surprisal and informativity were measured both in the forward and backward directions [17]. All these variables were log-transformed (base 10) and normalized.

A syllable's position within a TSG (initial/medial, ante-penultimate, penultimate, final) and the boundary's prosodic strength (TSG+IP, TSG+ip, TSG-only) were two other crucial variables. Their interaction was also included in the analysis.

Some controlling variables were included in the analysis: Speech rate (syllable count per second in each TSG), syllable count of the word a syllable occurs, the surface tonal category of a syllable, and each syllable's baseline duration. Following a similar method reported in [20], baseline duration was calculated from the predictions of a linear regression model trained to predict the duration of a syllable based on its component phonemes.

3. RESULTS

3.1. Full model statistics

In the full model, all variables except for unigram surprisal (β = 0.015, p = 0.23) and bigram forward informativity (β = 0.011, p = 0.4) significantly improved the model fit. The other three predictability variables had a positive estimate, indicating that higher surprisal/informativity was associated with longer syllable duration (backward informativity: β = 0.063, $p < .001$; forward surprisal: β = 0.018, $p < .0001$; backward surprisal: β = 0.015, $p < .001$).

There was also a significant interaction between boundary strength and prosodic position. Figure 1 shows the estimated syllable duration as a function of prosodic position and boundary type. Across different boundary types, there was significant final and penultimate lengthening (p < .0001 for all pairs of comparisons). The main difference among boundary types was the duration of the pre-boundary syllable: TSG+IP boundaries had the longest final syllable, followed by TSG+ip and TSG-only boundaries (p < .0001 in all pairs). The penultimate syllables at TSG+IP boundaries were also longer than the TSG-only counterparts (p < .001). In other words, having higher levels of prosodic breaks coinciding with TSG breaks did result in a significant difference in boundary cues.

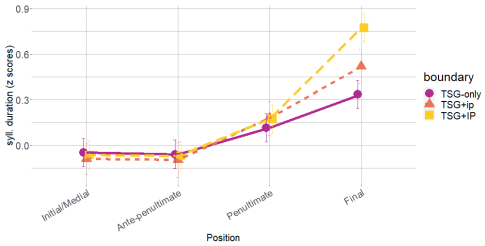

Figure 1: Syllable duration as a function of prosodic position (x axis) and boundary strength (color/shape). Error bars indicate standard errors.

As for the controlling variables, syllable duration was shorter with faster speech rates (β = -0.19, $p < .0001$), in longer words (β = -0.19, $p < .0001$), and with more phonological neighbors (β = -0.029, $p < .05$).

3.2. Predictability effects at different conditions

Next, we fitted a series of regression models that only contained one of the predictability variables so that their effect sizes had an interpretation free of issues such as suppression [24]. In these models, we also included the three-way interaction between the predictability variable, prosody position, and boundary strength. Post-hoc analyses were conducted using the lstrends() function in the lsmeans [11] package in R.

Figure 2 illustrates the interaction between informativity, prosodic position, and boundary strength. The effect of forward informativity (upper panel) was significant in all conditions except the final position for all three types of boundaries (p < .0001). **Backward informativity** (lower panel) was significant under the same conditions (TSG+ip at the ante-penultimate position: $p < .001$; all other non-final conditions: $p < .0001$).

The effect of **Unigram surprisal** is shown in the top panel of Figure 3. Similar to informativity, it was a significant factor for all but the final position across three boundary types (TSG-only: $p < .0001$; TSG+ip: $p < .05$ at initial/medial, $p < .01$ at ante-

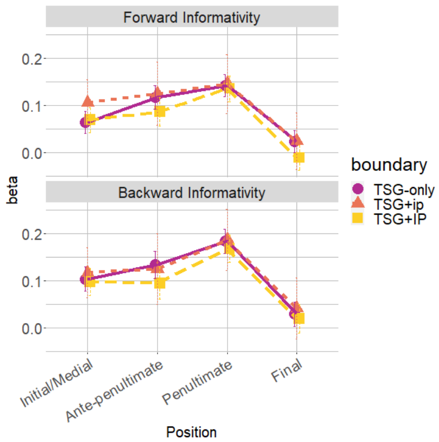

Figure 2: The effect size of **informativity** as a function of direction (panel), prosodic position (x axis) and boundary strength (color/shape). Error bars indicate standard errors.

penultimate, $p < .0001$ at penultimate; TSG+IP: $p < .0001$ at initial/medial and final, $p < .001$ at ante-penultimate).

The effect of **forward surprisal** (middle panel in Figure 3) was only significant at the penultimate position for TSG+ip ($p < .01$) and TSG+IP ($p < .0001$) breaks, but for all but the final position for TSG-only breaks (penultimate & ante-penultimate: $p < .0001$, initial/medial: $p < .001$). Finally, the **backward surprisal** effect (bottom panel) showed a three-way difference between prosodic conditions: significantly positive for all but the final position in the TSG-only condition (penultimate & ante-penultimate: $p < .0001$, initial/medial: $p < .01$), not significant at all in the TSG-ip condition, and significantly positive for all but negative at the final position in the TSG-IP condition (initial/medial: $p < .01$, penultimate/final: $p < .0001$). The negative effect suggests that syllables from words predicted to be frequent at the sentence/utterance-final position were actually lengthened at the IP boundary.

4. DISCUSSION & CONCLUSION

Results show that TSG boundaries exhibit lengthening of the penultimate and final syllables even without the boundary overlapping with an IP or ip boundary. In the overlapping scenarios, the duration of the pre-boundary syllable was significantly longer, especially when overlapping with an IP boundary. While this finding does not directly lead to a conclusion the TSG should be incorporated as a level of prosodic units alongside IP and ip, it does show that a linguistic domain defined purely on phonological terms and likely syntactically motivated may still exhibit similar acoustic cues at least in terms of durational patterns.

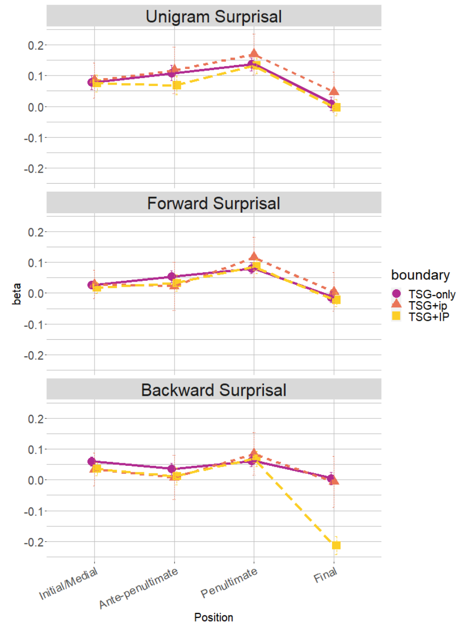

Figure 3: The effect size of **surprisal** as a function of surprisal type (panel), prosodic position (x axis) and boundary strength (color/shape). Error bars indicate standard errors.

TSG-only boundaries were also similar to TSG+IP and TSG+ip boundaries (as well as IP boundaries in general, as shown in [22]) in how the positive correlation between syllable duration and surprisal/informativity usually became neutralized towards a boundary, where final lengthening occurs. As discussed in previous studies, such neutralization is consistent with a view that prosodic modulates the relationship between predictability and acoustic cues, so that when boundary cues are already present, predictability effects become neutralized [21, 13]. Again, the current study shows that this type of interaction can also occur in a linguistic domain other than the IP. As for boundary strength, other than the reversal of backward surprisal's effect at TSG+IP boundaries, overlapping with larger prosodic boundaries did not affect the interaction between predictability and final lengthening at TSG boundaries.

To conclude, this study presents a detailed description of pre-boundary durational variability at TSG breaks in Taiwan Southern Min. We show how TSGs, defined entirely differently from prosodic units such as the IP and ip, still exhibit very similar boundary cues. Future studies with a more comprehensive survey on boundary cues across different types of syntactic units may further shed light on the issue of syntax-prosody interface and the status of TSGs as part of Taiwan Southern Min's prosodic system.

5. REFERENCES

[1] B. Andreeva, B. Möbius, and J. Whang, "Effects of surprisal and boundary strength on phrase-final lengthening," in *Proc. 10th International Conference on Speech Prosody 2020*, 2020, pp. 146–150.

[2] M. Aylett and A. Turk, "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," *Language and speech*, vol. 47, no. 1, pp. 31–56, 2004.

[3] D. Bates, M. Mächler, B. Bolker, and S. Walker, "2015. fitting linear mixed-effects models using lme4," *J. Stat. Softw*, vol. 67, no. 1, pp. 1–48, 2015.

[4] M. Y. Chen, "The syntax of xiamen tone sandhi," *Phonology*, vol. 4, pp. 109–149, 1987.

[5] S. F. Chen and J. Goodman, "An empirical study of smoothing techniques for language modeling," *Computer Speech & Language*, vol. 13, no. 4, pp. 359–394, 1999.

[6] R. L. CHENG, "Tone sandhi in taiwanese," vol. 6, no. 41, pp. 19–42, 1968. [Online]. Available: https://doi.org/10.1515/ling.1968.6.41.19

[7] U. Iunn, "Taiwanese corpus collection and corpus based syllable/word frequency counts for written taiwanese," 2005.

[8] C.-H. Kuo, "Perception and acoustic correlates of the taiwanese tone sandhi group," Ph.D. dissertation, UCLA, 2013.

[9] G. Kuo, "Perceived prosodic boundaries in taiwanese and their acoustic correlates," in *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

[10] ——, "Processing prosodic boundaries in natural and filtered speech," in *Proceedings of the 7th International Conference on Speech Prosody*, 2014, pp. 983–986.

[11] R. V. Lenth, "Least-squares means: The R package lsmeans," *Journal of Statistical Software*, vol. 69, no. 1, pp. 1–33, 2016.

[12] J.-w. Lin, "Lexical government and tone group formation in xiamen chinese," *Phonology*, vol. 11, no. 2, pp. 237–275, 1994.

[13] Z. Malisz, E. Brandt, B. Möbius, Y. M. Oh, and B. Andreeva, "Dimensions of segmental variability: Interaction of prosody and surprisal in six languages," *Frontiers in Communication*, vol. 3, p. 25, 2018.

[14] H.-h. Pan and H.-t. Huang, "Lexical propensity and taiwanese min tone sandhi rules," in *Proceedings of the International Conference on Speech Prosody*, vol. 2020, 2020, pp. 518–522.

[15] H.-h. Pan, H.-t. Huang, and S.-r. Lyu, "The occurrence of Taiwanese Min juncture tones before prosodic boundaries and modification marker," in *Proceedings of the 19th International Congress of Phonetic Sciences*, vol. 2019, 2019, pp. 3423–3427.

[16] S.-h. Peng and M. E. Beckman, "Annotation conventions and corpus design in the investigation of spontaneous speech prosody in taiwanese," in *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, 2003.

[17] S. Seyfarth, "Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation," *Cognition*, vol. 133, no. 1, pp. 140–155, 2014.

[18] A. Stolcke, "Srilm-an extensible language modeling toolkit," in *Seventh international conference on spoken language processing*, 2002.

[19] A. Stolcke, J. Zheng, W. Wang, and V. Abrash, "Srilm at sixteen: Update and outlook," in *Proceedings of IEEE automatic speech recognition and understanding workshop*, vol. 5, 2011.

[20] K. Tang and R. Bennett, "Contextual predictability influences word and morpheme duration in a morphologically complex language (kaqchikel mayan)," *The Journal of the Acoustical Society of America*, vol. 144, no. 2, pp. 997–1017, 2018.

[21] A. Turk, "Does prosodic constituency signal relative predictability? a smooth signal redundancy hypothesis," *Laboratory phonology*, vol. 1, no. 2, pp. 227–262, 2010.

[22] S.-F. Wang, "The interaction between predictability and pre-boundary lengthening on syllable duration in taiwan southern min," *Phonetica*, vol. 79, no. 4, pp. 315–352, 2022. [Online]. Available: https://doi.org/10.1515/phon-2022-0009

[23] S.-F. Wang and J. Fon, "A Taiwan Southern Min spontaneous speech corpus for discourse prosody," in *The Proceedings of Tools and Resources for the Analysis of Speech Prosody, Aix-en-Provence, France*, 2013, pp. 20–23. [Online]. Available: http://www.lpl-aix.fr/~trasp/Proceedings/20427-trasp2013.pdf

[24] L. H. Wurm and S. A. Fisicaro, "What residualizing predictors in regression analyses does (and what it does not do)," *Journal of memory and language*, vol. 72, pp. 37–48, 2014.