

SYLLABLE DURATION AND DISCOURSE ORGANIZATION AT INTONATIONAL PHRASE BOUNDARIES IN TAIWAN SOUTHERN MIN

Sheng-Fu Wang, Janice Fon

Graduate Institute of Linguistics, National Taiwan University, Taiwan
sftwang0416@gmail.com, jfon@ntu.edu.tw

ABSTRACT

This study examined syllable duration at discourse boundaries in spontaneous speech in Taiwan Southern Min with an eight-hour monologue corpus contributed by sixteen speakers, evenly split by gender and age. In order to control for prosodic boundary type, we only examined discourse boundaries coincided with an intonational phrase boundary followed by a pause. In addition, to tease apart the possible effect of final particles, we subset the data according to whether a boundary was marked with a final particle. Results showed a disyllabic domain for pre-boundary lengthening in the dataset without final particles and a trisyllabic domain in the dataset with final particles. In addition, overall, syllable duration at the pre-boundary position showed an inverse relationship with discourse boundary strength, especially in young speakers' speech.

Keywords: spontaneous speech, Southern Min, intonational phrase, final lengthening

1. INTRODUCTION

Syllable duration has been shown to encode the presence of syntactic and discourse boundaries. The most universal usage of syllable duration is final-lengthening, the prolongation of the pre-boundary syllable or segment [5, 12]. There are also some language-specific cues such as penultimate lengthening, which is found in Mandarin Chinese [4, 5].

As for indicating boundary strength, syllable duration seems to be a less robust cue. Yet, one particular pattern was observed in the variety of Mandarin spoken in Taiwan, where the pre-boundary syllable duration signals boundary strength inversely: shorter duration was observed at stronger boundaries [4, 5]. Since the same pattern was not found in Mandarin spoken in China, what has been found in Taiwan Mandarin may be referred to as a dialectal difference.

Taiwan Southern Min (TSM), more commonly called "Taiwanese", refers to the variety of the Min language spoken in Taiwan, where it is the mother

tongue of approximately 70% of the population [7]. TSM is special in having close contact with Taiwan Mandarin. Since TSM was rooted in Taiwan hundreds of years before Mandarin became dominant for political reasons, TSM might be seen as a substrate language that influences certain characteristics of Taiwan Mandarin. Thus, research on TSM may help explain relevant findings in Taiwan Mandarin, especially when speakers at different ages were included since they would also differ in their relative proficiency between Mandarin and TSM. Any age differences in the results may reveal the sources of the observed durational patterns. For the present study, we focus on how syllable duration reflect discourse organization, namely the presence and strength of discourse boundaries in spontaneous TSM. We wanted to answer two research questions: First, whether TSM also has the inverse relationship between pre-boundary syllable duration and discourse boundary strength. Second, whether TSM also has penultimate lengthening in addition to final lengthening.

2. METHOD

2.1. Corpus

The full dataset contained eight hours of spontaneous monologue-like speech elicited in the form of an interview. There were sixteen speakers, evenly split by gender and age. To avoid the confounding factor of dialectal differences, all of the speakers were from the same region (Taichung). The recordings were transcribed, romanized, and aligned at the syllable level using Praat [2] with the aid of the EasyAlign plug-in [6].

2.2. Discourse Labeling

Discourse annotation was conducted by adopting the model proposed in [8], which annotates basic "purpose units" and the hierarchical relationship between these units in discourse. In the current study, basic discourse units were defined as units containing one main verb, following the definition of a sim-

ple clause in [10]. Next, the relationship, or rather, the strength, of the boundaries between these units was specified with a four-level scale called “Discourse Boundary Index”.

DBI0, the lowest level, was used to label two adjacent clauses describing the same event or entity, as in the following cases: between a matrix and a subordinate clause, between two clauses showing a parallel syntactic pattern, between a tag question and its preceding clause, and between clauses sharing the same anaphora. DBI1 was often used to label the following cases: anaphoric change or update, change of aspect or the introduction of a new time reference, comment, etc.

DBI2 was used in situations where the boundary clearly differentiates two themes or episodes, yet these themes and episodes are still within a larger general topic. DBI3 was an additional label for handling radical shifts of themes which may be considered as boundaries of totally different monologues or interviews. This label was often used when the interviewer directed the interviewee to another totally unrelated topic.

Since discourse structure was an independent variable in this study, discourse annotation was conducted solely on the transcription, independent of the recordings.

2.3. Prosodic Labeling and Syllable Position

Since acoustic cues might be mediated by the prosodic structure (i.e., more pronounced cues were caused by the placement of bigger prosodic breaks), it was important to control for the level of prosodic boundaries that corresponded to the discourse boundaries investigated. A ToBI-style Break Index labelling system [1, 13] was used to identify the major (i.e., intonational phrase) and minor prosodic boundaries, as well as the boundaries with segmental truncations and hesitations. In addition, in order to further restrict the type of prosodic break associated with discourse boundaries, we only included boundaries that are followed by a pause, thereby ensuring that we are only looking at boundaries that are of the same prosodic category.

Syllable position was annotated with the nearby discourse boundary as reference. Syllable duration at four positions before and after a boundary was the target of this study. The pre-boundary positions, from the one immediately preceding the boundary, were referred to as “p-1”, “p-2”, “p-3”, and “p-4”. The post-boundary positions were referred to as “p+1”, “p+2”, “p+3”, and “p+4”.

Since the presence of final particles, which signals the finality of a sentence as well as providing

some pragmatic functions, might also alter the overall prosodic pattern at discourse boundaries, we further divided the data into two subsets: one with final particles and one without final particles.

3. RESULTS

For each subset of data, a four-way mixed-design analysis of variance was run with DBI, age, and gender as between-subject factors and position as a within-subject factor. F and df values were corrected in case of violation of sphericity.

3.1. Dataset with final particles

The results showed main effects of position [$F(6.46, 7961.91) = 17.61, p < .001, \eta^2_G = .011$] and gender [$F(1, 1233) = 7.15, p < .01, \eta^2_G = .008$; female > male, 201.84 ms vs. 186.96 ms], as well as an interaction between DBI and position [$F(12.91, 7961.91) = 2.30, p < .01, \eta^2_G = .003$].

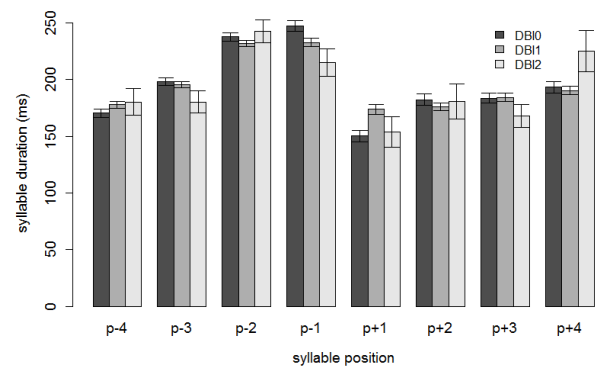


Figure 1: Patterning of syllable duration as a function of DBI and position at discourse boundaries in the dataset with final particles (Error bars indicate standard error.)

The interaction between DBI and position is shown in Figure 1. *Post hoc* analyses with Bonferroni adjustments showed significant final and penultimate lengthening across three DBI levels, although the difference between final and penultimate syllable duration only showed up at DBI0 (DBI2: p-1 > all but p-2, p+2, p+4, $p < .05$; p-2 > all but p-1, $p < .05$; no difference between p-1 and p-2; DBI1: p-1 & p-2 > other positions, $p < .01$; DBI0: p-1 > p-2, $p < .05$, p-1 & p-2 > other positions, $p < .001$). A significant ante-penultimate lengthening effect was also found at DBI1 and DBI0 (p-3 > p-4, $p < .01$ at both DBI1 and DBI0). In addition, an initial shortening effect was found cross three DBI levels (p+1 < pre-boundary positions, $p < .05$ for all comparisons). As for the effect of DBI at different syllable

ble positions, results showed that DBI had a negative effect on syllable duration at the p-1 position. Multiple comparisons with Bonferroni adjustments showed that DBI0 syllables were longer than DBI1 syllables at this position ($p < .05$). The DBI effect surfaced in a positive way at the p+1 position, where DBI1 syllables were longer than DBI0 syllables ($p < .01$).

3.2. Dataset without final particles

The results showed main effects of position [$F(6.60, 19308.57) = 51.10, p < .001, \eta^2_G = .017$], age [$F(1, 2322) = 13.43, p < .001, \eta^2_G = .001$, old speakers having longer syllable duration, 191.85 ms vs. 182.61 ms], and gender [$F(1, 2322) = 8.62, p < .001, \eta^2_G = .001$, female speakers having longer syllable duration, 193.50 ms vs. 181.83 ms], as well as an interaction between DBI and position [$F(13.19, 19308.57) = 2.56, p < .01, \eta^2_G = .002$].

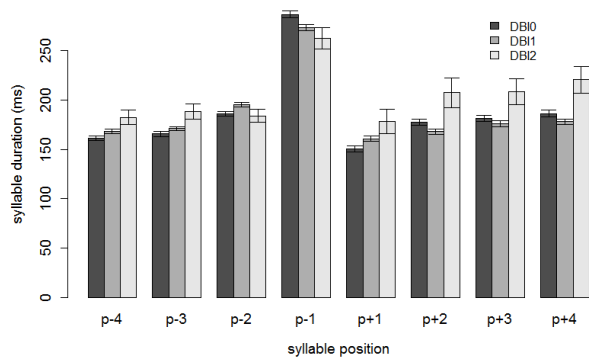


Figure 2: Patterning of syllable duration as a function of DBI and position at discourse boundaries in the dataset without final particles (Error bars indicate standard error.)

The interaction between DBI and position is plotted in Figure 2. *Post hoc* analyses with Bonferroni adjustments showed significant final lengthening across three DBI levels (DBI2: p-1 > all but p+4, $p < .05$; DBI1: p-1 > other positions, $p < .001$; DBI0: p-1 > other positions, $p < .001$). Penultimate lengthening was found for DBI1 (p-2 > other positions except for p-1, $p < .001$) and DBI0 (p-2 > p-3, p-4, $p < .05$). In addition, initial shortening was found at three DBI levels (for all three levels, p+1 < other positions, $p < .001$). As for the effect of DBI at different positions, at the p-1 position, DBI1 was shown to be significantly shorter than DBI0 ($p < .001$), showing an inverse relationship between syllable duration and the hierarchical organization of discourse. At other positions, the observed DBI

effects were all positive (p-4: DBI2 > DBI0; p-3: DBI2 > DBI0; p-2: DBI1 > DBI0, ; p+1, DBI1 > DBI0; p+2: DBI2 > DBI1 & DBI0; p+3: DBI2 > DBI1 & DBI0; p+4: DBI2 > DBI1 & DBI0. $p < .05$ for all comparisons).

4. DISCUSSION

Our first research question concerns the relationship between discourse boundary strength and pre-boundary syllable duration. The present study shows that pre-boundary syllable duration negatively correlated with discourse boundary strength. This kind of inverse relationship has also been found in Taiwan Mandarin, but not in the Mandarin spoken in China [4, 5]. The presence of this “negative hierarchy effect” in Taiwan Southern Min suggests a possibility that Taiwan Mandarin has taken this prosodic feature from its contact with Min in Taiwan, although the claim might have to be mitigated, since age did not interact significantly with the observed negative hierarchy effect in the present study, showing that different generations of speakers do not differ in the employment of this feature.

The presence of this negative discourse hierarchy effect in Taiwan Mandarin and its absence thereof in the Mandarin spoken in China was speculated as a result of rhythmic differences between these two varieties in [4]: since silent pause is a very important cue to discourse hierarchy [5], the negative correlation between syllable duration and discourse boundary strength may be a mechanism to highlight pause duration. Since Mandarin speakers in China prefer a faster speech rate [14], they may be more sensitive to long pauses following a boundary so that they can more effectively use pauses as a cue. In contrast, Taiwan Mandarin speakers may relatively be more tolerant to long pauses due to slower speech rates so that they have to inversely use syllable duration to further highlight pauses as a cue to discourse strength. Further studies on the rhythmic properties of Mandarin and TSM are needed to solidify this claim.

Our second research question concerns penultimate lengthening and can be rephrased as “whether TSM has a disyllabic domain for pre-boundary lengthening”. The answer is affirmative, as the results for both datasets in this study showed a significant final and penultimate lengthening effect, as has been found in the Mandarin dialects spoken in Taiwan and China [4]. Neither English, a stress language, nor Japanese, a pitch-accent language, exhibits a disyllabic range of lengthening before discourse or syntactic boundaries [5]. This finding

has implications concerning linguistic typology and prosodic realization: the existence of penultimate lengthening in Taiwan Southern Min and Mandarin seems to suggest that the disyllabic domain of pre-boundary strengthening may be a feature for tone languages.

Tone languages, especially those like Mandarin and Min, may indeed need a larger domain of pre-boundary strengthening. The autosegmental view on intonational phonology [9] assumes that a syllable has to be stressed to bear a tonal target. Following this point of view, tone languages must have a very dense stress assignment because most, if not all, syllables bear lexical tones. This would imply a lack of durational flexibility due to the lack of the frequent stressed-unstressed contrasts and the durational differences that go with such contrasts. The lack of flexibility in syllable duration may thus be compensated with a larger domain of lengthening when there is a need to signal the presence of a discourse boundary.

Since the dataset with final particles also shows an ante-penultimate lengthening effect, that is, a trisyllabic domain for pre-boundary lengthening, another possible explanation concerns the word template in TSM. The fact that Mandarin and Taiwan Southern Min employs a wider domain of pre-boundary lengthening may be attributed to the word template preference. There is empirical evidence that the disyllabic word type are the most frequent one in Taiwan Mandarin [15] and Taiwan Southern Min [11]. The same tendency has been noted for Standard Chinese [3], which refers to the Mandarin dialect spoken in China. Thus, it is certainly possible that this similarity in word type frequency contributed to the range of pre-boundary lengthening found in these two languages, which may explain why the pre-boundary lengthening domain became trisyllabic once a final particle is attached at the boundary.

5. CONCLUSION

Syllable duration at discourse boundaries that coincided with an intonational phrase boundary in TSM patterns similarly with what has been found in Taiwan Mandarin. For the inverse relationship between pre-boundary syllable duration and discourse boundary strength, more studies on speech rhythm may help answer why TSM and the Taiwan variety of Mandarin have this cue, but not the variety spoken in China. As for the domain of pre-boundary lengthening, TSM, just like Mandarin, has a disyllabic domain, showing that it could be a feature for

tone languages, although word template might also play a role.

6. REFERENCES

- [1] Beckman, M. E., Hirschberg, J., Shattuck-Hufnagel, S. 2005. The original ToBI system and the evolution of the ToBI framework. In: Jun, S.-A., (ed), *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press 9–54.
- [2] Boersma, P., Weenink, D. 2010. Praat: doing phonetics by computer [computer program], version 5.1. 44.
- [3] Duanmu, S. 1999. Stress and the development of disyllabic words in Chinese. *Diachronica* 16(1), 1–35.
- [4] Fon, J., Johnson, K., Chen, S. 2011. Durational Patterning at Syntactic and Discourse Boundaries in Mandarin Spontaneous Speech. *Language and Speech* 54(1), 5–32.
- [5] Fon, Y.-J. J. 2002. *A cross-linguistic study on syntactic and discourse boundary cues in spontaneous speech*. PhD thesis The Ohio State University.
- [6] Goldman, J. 2011. EasyAlign: an automatic phonetic alignment tool under Praat. *Proceedings of Interspeech 2011: 12th Annual Conference of the International Speech Communication Association, Florence, Italy, August 27-31* 3233–3236.
- [7] Government Information Office, 2012. *The Republic of China Yearbook 2012*. Taipei: The Executive Yuan.
- [8] Grosz, B., Sidner, C. 1986. Attention, intentions, and the structure of discourse. *Computational linguistics* 12(3), 175–204.
- [9] Ladd, D. 2008. *Intonational phonology*. Cambridge: Cambridge University Press.
- [10] Li, C. N., Thompson, S. A. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- [11] Ministry of Education, 2010. Term Frequency Statistics of Taiwan Southern Min. <http://203.64.42.97/bang-cham/thau-iah.php>.
- [12] Oller, D. 1973. The effect of position in utterance on speech segment duration in English. *The journal of the Acoustical Society of America* 54, 1235–1247.
- [13] Peng, S.-H., Beckman, M. E. 2003. Annotation conventions and corpus design in the investigation of spontaneous speech prosody in Taiwanese. *Proceedings of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition, Tokyo, Japan, April 13-16* 17–22.
- [14] Scollon, R. 1985. The machine stops: Silence in the metaphor of malfunction. *Perspectives on silence* 21–30.
- [15] Tseng, S.-C. 2004. Processing Mandarin spoken corpora. *Traitement Automatique des Langues. Special issue: Spoken corpus processing* 45(2), 89–108.