

A CROSS-LINGUISTIC STUDY ON DISCOURSE AND SYNTACTIC BOUNDARY CUES IN SPONTANEOUS SPEECH: USING DURATION AS AN EXAMPLE

Janice FON

Department of English,
National Taiwan Normal University, Taipei
jfon@ling.ohio-state.edu

ABSTRACT

This study focuses on the acoustic-phonetic cues at discourse and syntactic boundaries in English, Guoyu, Putonghua, and Japanese. Speech was elicited by having talkers describe the events in *The Pear Story* film. Recorded data were transcribed and segmented into discourse and syntactic units. Acoustic-phonetic measurements of syllable duration and syllable onset intervals (SOIs) were taken on the digitized data. A comparison of discourse/syntax and acoustics was made in order to examine boundary cues in speech. Results showed that final lengthening of boundary syllables and SOIs is the most universal cue for signaling structural boundaries, and the degree of final SOI lengthening is reflective of disjuncture hierarchy. However, there are also language-specific cues. In English, in addition to final lengthening, initial pitch-accented syllables and SOIs are also lengthened. In Guoyu and Putonghua, the scope of final syllable lengthening is widened to include the penultimate syllable. English and Putonghua are similar to each other in that there is no reflection of hierarchy in the degree of final syllable lengthening, while Guoyu and Japanese are more alike in this regard since both languages show some reflection of hierarchy through the degree of boundary syllable lengthening. Bigger structural boundaries are signaled by a smaller degree of lengthening.

1. INTRODUCTION

Syllable duration is influenced by many factors, including segmental effects such as phone types, suprasegmental effects such as stress, tone, and pitch accent, structural effects such as final lengthening at word, phrasal, clausal, and discourse boundaries, and pragmatic effects such as genre, rate, and style [1, 3, 5, 8, 9, 10, 11, 12, 14, 15, 17, 18, 20, 21, 22, 23, 24]. On the other hand, silent pause duration and thus syllable onset intervals (SOIs) are mainly affected by structural and pragmatic parameters only [3, 4, 7, 9, 13, 16, 19, 22, 25, 26]. This study discusses how syllable duration and SOIs reflect discourse hierarchy in English, Mandarin and Japanese.

2. METHODS

2.1 Subjects

Subjects from three languages—English, Japanese, and Mandarin, and two dialects of Mandarin—Guoyu and Putonghua, were recruited. All speakers were recruited from Columbus, Ohio. In order to attain homogeneity, only native speakers of

Central Ohio English from Columbus and neighboring counties, Tokyo Japanese from Tokyo and three neighboring prefectures (i.e., Chiba, Saitama, and Kanagawa), Taipei Guoyu from Taipei and Taipei County, and Beijing Putonghua from Beijing area were included. Subjects were either born and raised in the language area or moved there before age three, and had no exposure to other languages before that. Four females and four males were recruited for each language/dialect group. There are in total 4 (subjects) \times 2 (genders) \times 4 (languages/dialects) = 32 subjects.

2.2 Stimuli and Equipments

The Pear Story [2], a short film without spoken language, was played to the subjects in order to elicit speech. Recordings were done with a SHURE SM10A head-mounted microphone connected to a SONY DAT DTC-790 recorder through a Symetrix SX202 Dual Mic Preamp preamplifier, and using Maxell R-64DA DAT tapes. A D-to-D transfer was done and the sampling rate was set at 44100 Hz and was later downsampled to 22050 Hz for further analyses.

2.3 Procedure

Recordings were made in the Phonetics Laboratory in the Department of Linguistics, The Ohio State University. Subjects were tested individually in a quiet room. To avoid any accommodation effect that subjects might have, the author asked an experimenter who is also a native or a near-native speaker of the dialect of one of the three languages to conduct the experiment for each language group. Subjects were first shown the short film and were then asked to describe it afterwards as if talking to a friend. The experiment took about 20–30 min, and each subject was paid \$5 for his participation.

2.4 Measurements and Labeling

Syllable and SOI durations were measured for each language and discourse labeling was determined using discourse segment purpose (DSP) as a unit [5]. Three levels of disjuncture were recognized. DSP0 refers to the situation when two clauses are linked by the same DSP while DSP1 refers to the situation when two clauses are linked by a different but related DSP. Finally, DSP2 refers to the situation when two clauses are linked by different and unrelated DSPs.

2.5 Data Selection

Boundary syllables can be followed by various combinations of pauses. For English, Guoyu, and Putonghua, the most common boundary syllable types are either syllables by themselves or syllables that are followed by unfilled pauses. In Japanese, boundary syllables seldom occur alone. They are often followed by unfilled pauses or unfilled pauses plus filled pauses. In order to facilitate cross-linguistic comparisons, for all subsequent analyses, only cases containing boundary syllables that are followed by unfilled pauses are included. This is to maintain a more homogeneous set of data and facilitate cross-linguistic comparisons, despite the spontaneous nature of the study. All of the results reported below are statistically significant unless noted otherwise.

3. RESULTS

3.1 English

Figure 1 shows the patterning of syllable and pause duration at various discourse boundaries in English. The *x*-axis indicates position. The bars in the upper part of the graph (corresponding to the left *y*-axis) indicate syllable duration while those in the lower part (corresponding to the right *y*-axis) indicate pause duration. In other words, the total length of the bars (combining the upper and the lower parts) indicates SOI.

From the figure, one sees that although syllable duration is lengthened at Position 0 (hereafter P_0) as compared to that at Positions -1 and -2 (hereafter P_{-1} and P_{-2} , respectively), there is also lengthening at Position 1 (hereafter P_1), especially at Levels DSP1 and DSP2. In fact, at the DSP2 level, the degree of lengthening between P_0 and P_1 is about the same. In addition, discourse hierarchy is surprisingly reflected by the initial rather than the final syllable lengthening effect. On the other hand, SOI patterning follows what was predicted in that the highest degree of lengthening occurs at P_0 . However, there is still a slight lengthening effect at P_1 , especially at Level DSP2. Both initial and final SOI lengthening effects reflect boundary strength.

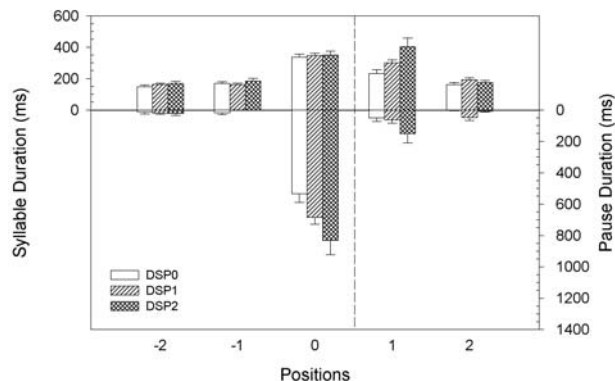


Figure 1: A bar graph illustrating the patterns of syllable and pause duration at structural boundaries in English. The *x*-axis shows syllable positions relative to the boundary. ‘0’ indicates boundary syllables located right before the relevant boundaries, ‘-1’ and ‘-2’ indicate syllables that are one and two syllables

before boundary syllables, respectively, and ‘1’ and ‘2’ indicate syllables that are one and two syllables after the relevant boundaries, respectively. In other words, the boundary is between ‘0’ and ‘1’, as indicated by the dashed line. The left *y*-axis indicates syllable duration and the right *y*-axis indicates pause duration. The error bars indicate standard error.

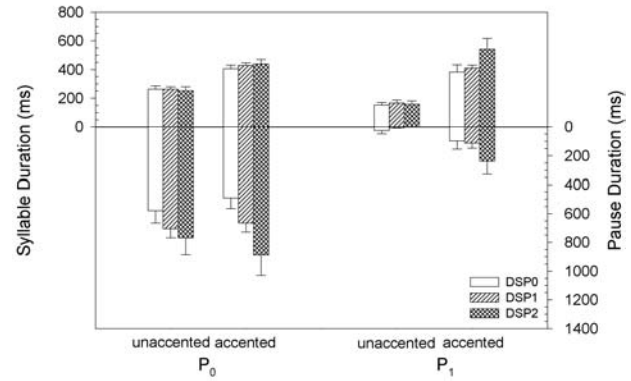


Figure 2: A bar graph illustrating the duration patterns of accented and unaccented syllables and SOIs at P_0 and P_1 . The *x*-axis indicates positions, and the *y*-axes indicate syllable (left) and pause duration (right). The error bars indicate standard error.

The initial lengthening effect on both syllable and SOI was unexpected. One possible explanation is that the effect might be related to pitch accents. Higher-level discourse segments are more likely to have accented full referring NPs as subjects, and accented syllables are more likely to be lengthened. Therefore, lengthening in position P_1 might be a reflection of the greater likelihood of having a subject with initial accent in that position. Figure 2 shows that pitch accents lengthen syllable and pause duration in general, regardless of positions. However, it is more effective in lengthening syllables than SOIs, and at P_1 than P_0 . Thus, the initial strengthening effect observed above seems to be due solely to the existence of pitch accents. Unaccented syllables are lengthened only at P_0 , DSP-final positions.

3.2 Guoyu

Figure 3 shows the syllable and pause duration patterning at various discourse boundaries in Guoyu. From the figure, one notices that, unlike in English, the final lengthening effect at P_0 in Guoyu only stretches syllable duration slightly, although SOIs are still lengthened to a comparable extent. Also, unlike English, lengthening starts early at P_{-1} . Another observation is that there is no duration increase of syllable that corresponds to the magnitude of discourse disjuncture. Instead, the degree of syllable lengthening tends to *decline* as discourse disjuncture becomes bigger, which is somewhat counterintuitive. However, boundary SOIs mirror discourse boundary strength in a positive manner as in English.

3.3 Putonghua

Figure 4 shows the syllable and pause duration patterning at various discourse boundaries in Putonghua. From the figure, one sees that there is very little lengthening at P_0 for syllable duration

as compared to the syllable before it, which is similar to what was observed in Guoyu. Also, like Guoyu, the lengthening effect starts early, at P_{-1} . In terms of SOI, the effect is more localized on the boundary position. Discourse hierarchy is reflected mainly by SOIs. No differentiation in the degree of syllable lengthening regarding boundary strength was found in Putonghua.

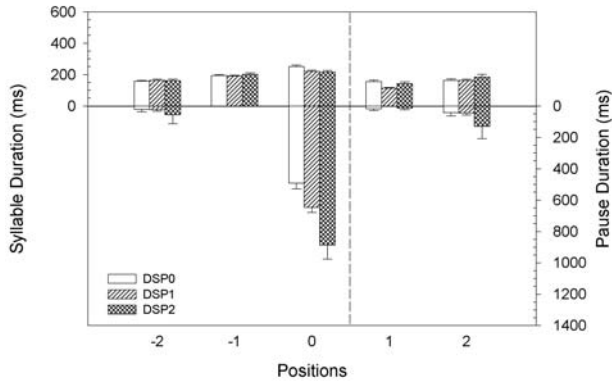


Figure 3: A bar graph illustrating the patterns of syllable and pause duration at structural boundaries in Guoyu. The x - and the y -axes are the same as those in Figure 1. The dashed line depicts the location of the boundary. The error bars indicate standard error.

3.4 Japanese

Figure 5 shows the syllable and pause duration patterning at various discourse boundaries in Japanese. One sees that unlike English and Mandarin, the syllable and SOI lengthening effects are localized at P_0 . As in Guoyu, discourse hierarchy is reflected by a decrease in syllable duration and an increase in SOIs at the boundary position. In other words, bigger disjunctures will result in shorter boundary syllables and longer boundary SOIs.

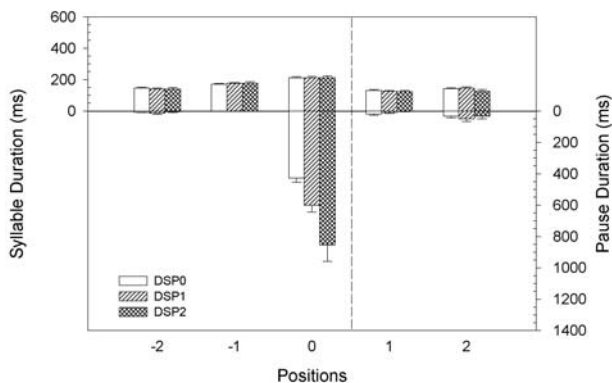


Figure 4: A bar graph illustrating the patterns of syllable and pause duration at structural boundaries in Putonghua. The x - and the y -axes are the same as those in Figure 1. The dashed line indicates the boundary location. The error bars indicate standard error.

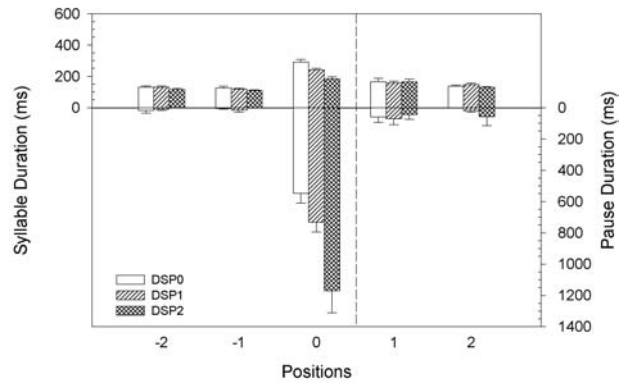


Figure 5: A bar graph illustrating the patterns of syllable and pause duration at structural boundaries in Japanese. The x - and the y -axes are the same as those in Figure 1. The dashed line indicates the boundary location. The error bars indicate standard error.

3.5 Cross-linguistic Comparisons

Comparing across different languages, one sees that not only do the exact location and degree of localization of the final lengthening effect vary, the absolute value of lengthening and its relationship to the discourse hierarchy also differ from language to language. One notices that the magnitude of syllable lengthening is on average larger in English and Japanese and smaller in the two Mandarin varieties, especially at the DSP0 level. At the DSP1 and DSP2 levels, English has the longest boundary syllables while the other three languages do not differ much from one another. Similarly, SOI lengthening is larger in English and Japanese, and smaller in Guoyu and Putonghua, although the difference is not as extreme.

4. DISCUSSION

The final syllable lengthening effect is less universal than it is usually assumed. In English and Japanese, the effect is fairly localized at the boundary syllable P_0 , while in Putonghua and Guoyu, the final lengthening effect starts earlier, at around P_{-1} and peaks at P_0 . On the other hand, SOI final lengthening is more universal. For all four language examined here, all of them display lengthening at the boundary position only. However, the absolute value of lengthening, be it on syllable or SOI, is still language-dependent. English and Japanese tend to have a higher degree of lengthening than Guoyu and Putonghua.

The interaction between syllable duration and pitch accents in English is also an intriguing phenomenon. Although accented syllables tend to be longer in general, they are especially lengthened at DSP-initial positions. SOIs, on the other hand, are not affected by pitch accents as much. In other words, the lengthening effect of pitch accents is more localized on syllable rather than on pause, and more on P_1 than any other positions examined.

The lengthening effect of syllable duration is sometimes also indicative of boundary strengths in most of the languages studied here, although the manifestation of such differs from language to language. In English, there is a positive correlation between

DSP-initial syllable duration and the degree of discourse disjuncture, provided that the syllables are pitch-accented. There is no such relation at the DSP-final position, however. In contrast, in Guoyu and Japanese, the correlation is negative at the DSP-final position in that the amount of lengthening declines when discourse disjuncture increases. On the other hand, Putonghua does not show any reflection of hierarchy through syllable duration.

The relationship between SOI lengthening and discourse hierarchy is more universal. In all four languages examined in this study, all showed a positive correlation between degree of lengthening and structural boundary strengths at the DSP-final position. This might be due to a more universal perception and manipulation of pause and thus SOI intervals than syllables. In fact, the unusual negative correlation between syllable lengthening and discourse disjuncture in Guoyu and Japanese can be viewed as a strategy to lengthen the relative duration of the following pause. Of course, this does not mean languages have to lengthen pause duration at the expense of syllable duration. English and Putonghua are two counterexamples. This would simply imply that languages do have different preferences as to how to indicate discourse hierarchy, and it just so happens that Guoyu and Japanese emphasize more on pause duration than English and Putonghua.

5. ACKNOWLEDGMENTS

This study is part of the author's dissertation project at The Ohio State University and was supported by NIDCD grant no. R01DC04421, granted to Dr. Keith Johnson, the author's advisor. The author wishes to thank Dr. Keith Johnson, Dr. Mary E. Beckman, and Dr. Shari Speer, for their generous support and comments.

6. REFERENCES

- [1] Campbell, W. N. and Isard, S. D. "Segment durations in a syllable frame," *Journal of Phonetics*, 19(1):37-47, 1991
- [2] Chafe, W. L. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production. Advances in discourse processes*, 3. Norwood, NJ: Ablex, 1980.
- [3] Cooper, W. E., Paccia, J. M. and Lapointe, S. G. "Hierarchical coding in speech timing," *Cognitive Psychology*, 10:154-177, 1978.
- [4] Fon, J. and Johnson, K. "Speech Timing Patterning as an Indicator of Discourse and Syntactic Boundaries," *Proceeding of 6th International Congress of Spoken Language Processing*, Beijing, China, 2000, pp. 555-558.
- [5] Fowler, C. A., Levy, E. T. and Brown, J. M. "Reductions of spoken words in certain discourse contexts," *Journal of Memory and Language*, 37(1):24-40, 1998.
- [6] Grosz, B. and Hirschberg, J. "Some Intonational Characteristics of Discourse Structure," *Proceedings of the 2nd International Congress on Spoken Language Processing*, Alberta, Canada, 1992, pp. 429-432.
- [7] Hirschberg, J. and Nakatani, C. H. "A Prosodic Analysis of Discourse Segments in Direction-giving Monologues," *Proceeding of the 4th International Congress of Spoken Language Processing*, Philadelphia, USA, 1996, pp. 286-293.
- [8] Huggins, A. W. F. "An effect of syntax on syllable timing," *Quarterly Progress Report, MIT*, 114:179-185, 1974.
- [9] Katz, W. F., Beach, C. M., Jenouri, K. and Verma, S. "Duration and fundamental frequency correlates of phrase boundaries in productions by children and adults," *The Journal of the Acoustical Society of America*, 99(5):3179-3191, 1996.
- [10] Klatt, D. H. "Vowel lengthening is syntactically determined in a connected discourse," *Journal of Phonetics*, 3:129-140, 1975.
- [11] Koopmans-van Beinum, F. J. and van Donzel, M. E. "Relationship between Discourse Structure and Dynamic Speech Rate," *Proceeding of the 4th International Congress of Spoken Language Processing*, Philadelphia, USA, 1996, pp. 1724-1727.
- [12] Kreiman, J. "Perception of sentence and paragraph boundaries in natural conversation," *Journal of Phonetics*, 10(2):163-175, 1982.
- [13] Ladd, D. R. "Declination 'reset' and the hierarchical organization of utterances," *The Journal of the Acoustical Society of America*, 84(2):530-544, 1988.
- [14] Lehiste, I. "Phonetic disambiguation of syntactic ambiguity," *Glossa*, 7(2):107-121, 1973.
- [15] Lehiste, I., Olive, J. P and Streeter, L. A. "Role of duration in disambiguating syntactically ambiguous sentences," *Journal of the Acoustical Society of America*, 60(5):1199-1202, 1976.
- [16] Maclay, H. and Osgood, C. E. "Hesitation phenomena in spontaneous English speech," *Word*, 15:19-44, 1959.
- [17] Nakatani, L. H., O'Connor, K. D. and Aston, C. H. "Prosodic aspects of American English speech rhythm," *Phonetica*, 38:84-106, 1981.
- [18] Oller, D. K. "The duration of speech segments: The effect of position in utterance and word length," *Journal of the Acoustical Society of America*, 54(5):1235-1247, 1973.
- [19] O'Malley, M. H., Kloker, O. R. and Dara-Abrams, D. "Recovering parentheses from spoken algebraic expressions," *I.E.E.E. Transactions on Audio and Electroacoustics*, AU-21(3):217-220, 1973.
- [20] Quené, H. "Durational cues for word segmentation in Dutch," *Journal of Phonetics*, 20(3):331-350, 1992.
- [21] Quené, H. "Segment duration and accent as cues to word segmentation in Dutch," *The Journal of the Acoustical Society of America*, 94(4):2027-2035, 1993.
- [22] Scott, D. R. "Duration as a cue to the perception of a phrase boundary," *Journal of the Acoustical Society of America*, 71(4):996-1007, 1982.
- [23] Shen, X. S. "A pilot study on the relation between the temporal and syntactic structures in Mandarin," *Journal of the International Phonetic Association*, 22(1-2):35-43, 1992.
- [24] Streeter, L. A. "Acoustic determinants of phrase boundary perception," *Journal of the Acoustical Society of America*, 64(6):1582-1592, 1978.
- [25] Swerts, M. "Filled pauses as markers of discourse structure," *Journal of Pragmatics*, 30(4):485-496, 1998.
- [26] Swerts, M. and Geluykens, R. "Prosody as a marker of information flow in spoken discourse," *Language and Speech*, 37(1):21-43, 1994.