Durational and Prosodic Patterning at Discourse Boundaries in Japanese Spontaneous Monologs

Kiyoko Yoneyama[†*] Janice Fon[‡], & Hanae Koiso[†]
†The National Institute for Japanese Language, ∗Daito Bunka University
‡National Taiwan Normal University
Email: yoneyama@kokken.go.jp, jfon@cc.ntnu.edu.tw, koiso@kokken.go.jp

ABSTRACT

The aim of this paper is twofold. First, the paper attempts to investigate durational and prosodic patterning at clausal boundaries in Japanese spontaneous monologues by using the prosodic labels of the X-JToBI system ([1]) in the *Corpus of Spontaneous Japanese* ([2]). The analyses of prosodic cues around clausal boundaries (filled pauses, syllable duration, pause duration and clausal boundary tones) confirmed that these factors contribute to highlighting discourse disjunctures in Japanese spontaneous monologues. Secondly, the paper attempts to evaluate the validity of discourse break indices in describing discourse structure. Results showed that the both the interlabeler reliability (α = .95) and agreement (81.52%) are quite high, suggesting that the DBI labeling system is valid to some extent.

1. INTRODUCTION

The aim of this study is twofold. First, the paper attempts to investigate durational and prosodic patterning at clausal boundaries in Japanese spontaneous monologs. By using the prosodic labels of the X-JToBI system in the *Corpus of Spontaneous Japanese* (hereafter CSJ), prosodic and durational cues that occur around clausal boundaries were analyzed.

Secondly, this paper is to provide data on interlabeler reliability on the discourse-labeling system used. So far, four studies including this have used the discourse boundary indices (hereafter DBI) to label discourse hierarchy and have provided reliable results on relations between prosodic cues and degrees of discourse disjuncture, although no study has yet investigated the validity of this labeling schema. Therefore, investigation regarding interlabeler reliability is conducted in order to investigate this issue.

2. PROCEDURES

Twenty monologues in CSJ were used in this study (10 male and 10 female speakers aged between 20 and 40). The monologues covered a wide range of topics, from relatively neutral matters such as life in cities to personal experiences.

Our discourse labeling followed procedures in [3]. First, a monologue was divided into clauses, which were extracted automatically by referring to the morphological information of the monologues in CSJ, and hand-corrected afterwards. Next, relationships between two adjacent clauses were labeled using the DBI. three levels of DBI were recognized. DBI2 refers to a disjuncture when two clauses belonged to different discourse purposes. DBI1 was labeled when two clauses belonged to different discourse purposes but are strongly related. DBI0 was labeled when two clauses belonged to the same discourse purpose. An example of discourse labeling is shown in Figure 1. A graduate student labeled the DBIs for all twenty monologs.

3. DATA ANALYSES

3.1. Filled pauses

The occurrence of filled pauses at clausal boundaries was analyzed. Table 1 shows the distribution of different types of boundary syllables summed across all 20 monologs with regard to three degrees of discourse disjuncture. Boundary syllables seldom occurred alone (syll). They were often followed by unfilled pauses (syll+UP) or unfilled pauses plus filled pauses (syll+UP+FP).

Boundary Syllable type	DBI labeling DBI0	DBI1	DBI2	Total
syll	205 (8.9)	33 (3.7)	5 (2.1)	243 (7.1)
syll +UP	1549 (67.5)	558 (62.3)	126 (52.9)	2233 (65.1)
syll + UP+FP	541 (23.6)	305 (34.0)	107 (45.0)	953 (27.8)
Total	2295 (100)	896 (100)	238 (100)	3429 (100)

Table 1: Distribution of the number of cases at different discourse levels regarding boundary syllable types.

Figure 2 shows the percentages of different types of boundary syllables with regards to degrees of discourse disjuncture. One can see from the figure that the percentage of syll+UP+FP is higher as the discourse boundary size gets bigger. An ANOVA was conducted on the percentages of syll+UP+FP. The results showed that the percentages of syll+UP+FP at discourse boundaries were significantly different among different DBIs [$F(2, 34269 = 78.110, p < .0001)$]. Post-hoc Benferrioni tests showed that syll+UP+FP was the highest at DBI2 and the lowest at DBI0 ($p < .0001$). In other words, filled pauses occur more frequently as discourse disjunctures become bigger.

3.2. Syllable duration and pause duration

The syllable and pause durations around clausal boundary were analyzed. In this study, "syll+UP" shown in Table 1 were analyzed in this study. Syllable and pause durations around clause boundaries were calculated.

ISBN 1-876346-48-5

<<トレッキングツアーの構成メンバーについての説明>>
. . . . . ((省略)) . . . . .
<<トレッキングツアーの一日の説明>>
DBI2 でトレッキングの一日を簡単に御説明いたしますと
行程１朝食 DBI1 まず朝シェルパが持ってきてくれるモーニングティーで目を覚まします
DBI0 このモーニングティーってのは紅茶かコーヒーかっていうのを選べるんですけれども
DBI0 それで目を覚ましまして
DBI0 簡単に身支度を整えて
DBI0 荷物のパッキングだけをして
DBI0 朝食に向かいます
行程２準備 DBI1 で 私達が朝食を食べてる間に荷造りしたものをゾッキョに結び付けて
DBI0 ポーター達は出発の準備をしております
行程３出発 DBI1 で食事が終わりますと
DBI0 大体一日約六時間の行程でトレッキングを行ないました
. . . . . ((省略)) . . . . .

Figure 1: Example of discourse labeling

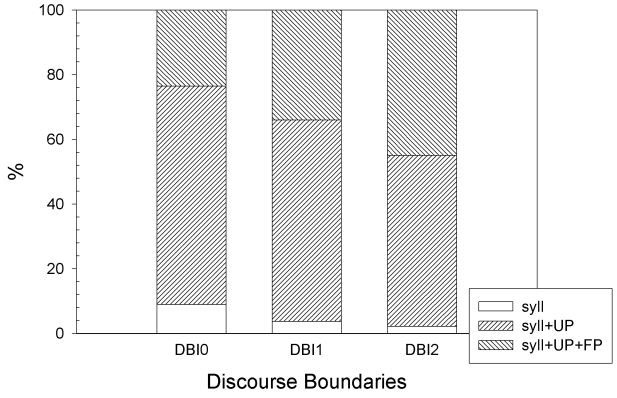

Figure 2: The percentages of boundary syllable types with regards to degrees of discourse disjuncture.

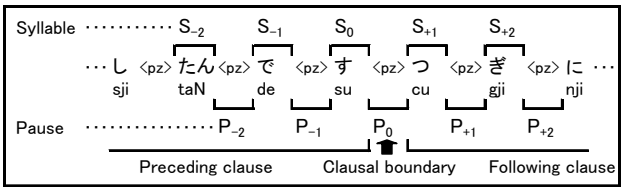

Figure 3: Positions of syllable and pause durations.

Figure 3 shows the positions of syllable and pause durations in this analysis. For syllable durations, the last syllable before a clausal boundary is S0. S-2 and S-1 precede S0 in the same clause whereas S+1 and S+2 follow it in the following clause. For pause durations, P0 occurs at a clausal boundary whereas P-2 and P-1 occur in the preceding clause and P+1 and P+2 occur in the following clause.

Figure 4 shows the patterns of syllable and pause duration at clausal boundaries. Position numbers are used for both syllable and pause. For example, Position 0 refers to both S0 and P0. The *x*-axis shows positions relative to the boundary. The left *y*-axis indicates syllable duration and the right *y*-axis indicates pause duration. The error bars indicate standard error.

For syllable duration, a two-way mixed design ANOVA (DBI × Position) was performed. Main effects of DBI and position were observed. Syllable durations at

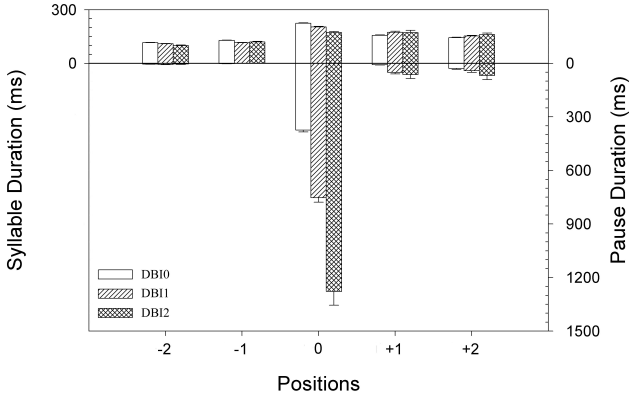

Figure 4: Patterning of syllable and pause duration at clausal boundaries.

DBIs were significantly different [$F(2, 2230) = 3.41$, $p < .05$]. Syllable durations at the five positions were also significantly different [$F(3.29, 7328.21) = 171.60$, $p < .0001$]. An interaction between DBI and position was also significant [F (6.57, 7328.21) = 12.86, $p < .0001$]. Post-hoc analyses using Bonferroni's adjustments showed that horizontally across different positions, syllable duration was the longest at S0 at the DBI0 and DBI1 ($p < .05$). At the DBI2 level, however, there was not much difference between syllable duration at S0, S+1 and S+2, although syllables at S0 were still longer than those at S-1 and S-2 ($p < .01$). Hierarchically, post-hoc Turkey's-*b* tests showed that syllable duration at S-2, S-1 and S0 corresponded to discourse boundary strength in a negative fashion. At S-2, syllable duration was longer at DBI0 and DBI1 than at DBI2 level ($p < .05$). At S-1, syllables at DBI0 were significantly longer than those at DBI1 ($p < .05$). At S0, it was the longest at DBI0 and the shortest at DBI2 ($p < .05$).

For pause duration, a two-way mixed design ANOVA (DBI × Position) was also performed. Main effects of DBI and position were observed. Pause durations at DBIs were significantly different [$F(2, 2230) = 265.06$, $p < .0001$]. Pause durations at the five positions were also significantly different [$F(1.47, 3281.20) = 1768.69$, $p < .0001$]. An interaction between DBI and position was also significant [$F(2.94, 3281.20) = 215.83$, $p < .0001$]. Post-hoc analyses using Bonferroni's adjustments showed that

horizontally across different positions, pause duration was the longest at P0 at all DBIs ($p < .01$). Hierarchically, post-hoc Turkey's-*b* tests showed that pause duration at P+2, P+1 and P0 corresponded to discourse boundary strength in a positive fashion. At P0, pause duration was the longest at DBI2 and the shortest at DBI0 ($p < .05$). At P+1, it was longer at DBI1 and DBI2 than at DBI0 level ($p < .05$). At P+2, it was significantly longer at DBI2 than at DBI0 ($p < .05$).

In summary, discourse hierarchy is best reflected at Position 0. At this position, syllable duration is reflective of discourse hierarchy in a negative fashion while pause a positive fashion. These patterns were also observed in previous studies [3][4].

3.3. Boundary Tones

The analysis conducted here was based on boundary tones in the X-JToBI system used in CSJ. Boundary tones are classified into two groups: simple (L%) and complex boundary tones (L%H%, L%HL%, L%HLH%, and L%LH%). Table 2 shows the distribution of the number of cases at different discourse levels regarding boundary tones. Due to case constraints, only those that are enclosed in the bold square were used for further analyses.

Boundary Tones	DBI labeling DBI0	DBI1	DBI2	Total
L%	970 (42.3)	510 (56.9)	189 (78.7)	1669 (48.7)
L%H%	569 (24.8)	162 (18.1)	35 (14.6)	766 (22.3)
L%HL%	745 (32.5)	219 (24.4)	16 (6.7)	980 (28.6)
L%HLH%	4 (0.1)	1 (0.1)	0 (0)	5 (0.1)
L%LH%	6 (0.3)	4 (0.5)	0 (0)	10 (0.3)
Total	2294 (100)	896 (100)	240 (100)	3430 (100)

Table 2: Distribution of the number of cases at different discourse levels regarding boundary tones. Numbers in parentheses show %.

Figure 5 shows the percentages of three boundary tones with regards to discourse disjunctures. One can see from the figure that the percentage of simple boundary tones is higher as discourse boundaries become bigger. The percentage of complex boundary tones is lower as discourse disjuncture becomes bigger. An ANOVA was conducted on the percentages of simple boundary tones (L%). The results showed that the percentages of the single boundary tones (L%) at discourse boundaries were significantly different from each other [F (2, 3412) = 73.997, $p < .0001$]. Post-hoc Bonferroni tests showed that L% was used most frequently at DBI2 and least frequently at DBI0 ($p < .001$). On the other hand, the proportions of L%HL% and L%H% were smaller as discourse disjuncture became bigger.

3.4. Discussion

The analyses of this section provided evidence that prosodic cues around clausal boundaries (filled pauses, syllable duration, pause duration and clausal boundary tones) provided evidence that all these factors contribute

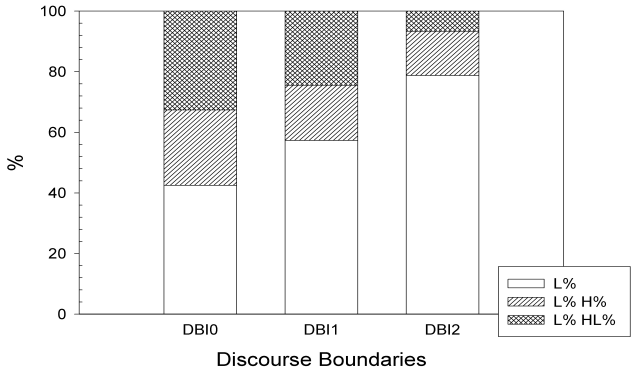

Figure 5: The percentages of boundary tones with regards of discourse junctures.

to highlighting discourse disjunctures in Japanese spontaneous monologues.

Syllable duration and boundary tones seem to reflect the degree of discourse continuation in Japanese spontaneous monologs. The syllable-final lengthening effect is localized on the boundary syllables. The degree of syllable lengthening decreases when discourse boundary strength increases. The distribution of boundary tones reflects degrees of discourse disjuncture as well. L% was used more often at DBI2 than at DBI0 whereas L%HL% was used more often at DBI0 than at DBI2.

Koiso and her colleagues pointed out that turn-taking tends to be observed in Japanese dialogs when the final mora is shorter whereas turn-holding tends to be observed when it is longer. They also found that the complex boundary tone L%HL% tend to be used to indicate that speakers hold their turn whereas the simple boundary tone L% tend to be used to indicate that they have finished and yield their turn [6]. Our results in the monologues seem to be consistent with their findings if we assume that turn-holding/yielding is highly related to discourse continuation.

Filled pause frequency and pause lengthening are also reflective of discourse disjuncture. Although it has been frequently pointed out that pause duration is longer at clausal boundaries than other positions, it is quite interesting that pause duration is also reflective of discourse disjuncture. This might be related to a cognitive process of discourse planning. If discourse disjuncture is bigger, such planning might require more time, resulting in the lengthening of pause duration and filled-pause frequency. One of the interesting findings is that not only was pause duration lengthened at the boundary position, but also at places after the clausal boundaries. This might be related to the fact that the conjunction, *de*, 'and' and fillers that frequently appear with a following pause are often filled in those positions.

4. INTERLABELER RELIABILITY OF DISCOURSE BREAK INDICES

Four studies including this have provided evidence that some prosodic cues highlight degrees of discourse disjuncture in Japanese spontaneous monologs in terms of the DBI labeling system [3][4][5]. Since the validity of

ISBN 1-876346-48-5

the DBI labeling system itself has not yet been tested, the consistency of such was examined.

Three labelers participated in the study. One labeler was the one who labeled DBIs for 20 monologues in this study (Labeler 1) and the other two labelers (Labelers 2 and 3) were newly recruited.

The data were collected from four monologs. About 50 DBI labels were collected from each monolog, resulting in 211 DBI labels in total. The data collecting procedures for Labelers 2 and 3 were as follows. First, the labelers were given the DSP labeling guidelines that were used by Labeler 1. The third author gave a two-hour lecture on the DBI labeling system including a question-answer period. After that, they started labeling the data independently.

There are two ways to evaluate a rating scheme, reliability and agreement. Both were examined in this study. Results showed that the interlabeler reliability was fairly high ($\alpha = .95$), indicating that the labelers could to a large extent distinguish big discourse disjunctures from smaller ones consistently.

The intertranscriber agreement was also high. Of the 211 DBIs labeled, 172 had the same labels from all three labelers, which was 81.52%. The rest of the DBIs had at least two labelers placing the same label.

The level of intertranscriber agreement was different for different DBIs. Labeling was more consistent for DBI0s and DBI2s, as shown in Table 3. Numbers in the 2nd column (All agree) indicate the number of labels that are the same among all three labelers. Numbers in the next three columns (Labler1, Labeler2 and Labeler3) refer to the total number of labels of each labeler. Numbers in parentheses indicate the percentages of labeling agreement.

Table 4 shows distribution of the five most common clausal endings. Figure 6 shows that the percentages of two-labeler and three-labeler agreement with regard to these endings. The figure shows that sentence final endings elicited the highest level of interlabler agreement while the ending *de* 'and' and *node* 'because' elicited the lowest level of interlabler agreement. In sum, the results showed that DBI labeling consistency was also correlated with clausal endings.

	All agree	Labeler 1	Labeler 2	Labeler 3
DBI0	132 (89.4)	145 (91.0)	144 (91.6)	154 (85.7)
DBI1	24 (56.3)	50 (48.0)	46 (52.2)	35 (68.6)
DBI2	16 (83.0)	16 (100)	21 (76.2)	22 (72.7)

Table 3: Labeling consistency among the three labelers.

Clausal ending	Consistency		Total
	Two agree	Three agree	
te	8 (14.3)	48 (85.7)	56 (100)
sentence final	4 (8.0)	46 (92.0)	50 (100)
node	10 (35.7)	18 (64.3)	28 (100)
ke(re)do(mo)	4 (20.0)	16 (80.0)	20 (100)
de	6 (40.0)	9 (60.0)	15 (100)
Total	32 (100)	137 (81.1)	169 (100)

Table 4: Distribution of five most common clausal endings.

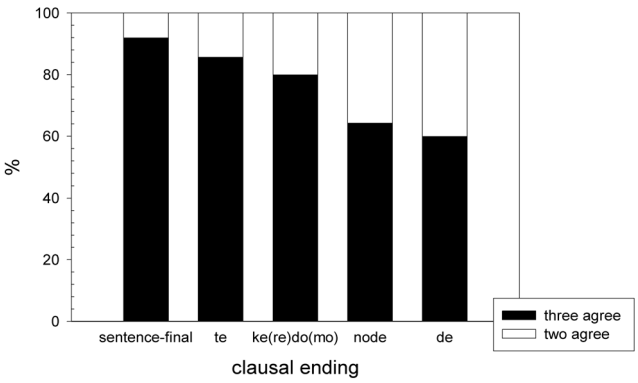

Figure 6: The percentages of two-labeler and three-labeler agreement with regard to five most common endings.

5. CONCLUSION

This study provides a glimpse of how Japanese, and spontaneous speech in general, although "messy", can provide consistent cues to speakers with regards to its discourse structure. Filled pause distribution, boundary tone distribution, and patterning of syllable duration and silent pauses all reflect discourse hierarchy to a certain extent. In addition, this study has also shown that the discourse-labeling scheme used in our series of studies is in general consistent, although labels of DBI0 and DBI2 are more consistent than those of DBI1.

REFERENCES

[1] K. Maekawa, H. Kikuchi, Y. Igarashi, & J.J. Venditti, "X-JToBI: An extended JToBI for spontaneous speech," *Proc. of 7th International Conference on Spoken Language Processing*, pp. 1545-1548, 2002.

[2] K. Maekawa, H. Koiso, S. Furui, & H. Isahara, "Spontaneous speech corpus in Japanese," *Proc. 2nd International Conference on Language Resource and Evaluation*, Athens, Greece, pp.947-952, 2000.

[3] Y-J.J. Fon, *A Cross-linguistic study on syntactic and discourse boundary cues in spontaneous speech*, Unpublished doctoral dissertation, The Ohio State University, Columbus, OH, U.S.A., 2002.

[4] H. Koiso, K. Yoneyama, Y. Maki, & J. Fon, "An analysis of prosodic and discourse structure in the Corpus of Spontaneous Japanese," To appear in JSAI SIG Notes, SIG-SLUD-A203, 2003 (written in Japanese).

[5] K. Yoneyama, H. Koiso, & J. Fon, "A corpus-based analysis on prosody and discourse structure in Japanese spontaneous monologue," To appear in *Proc. of ISCA & IEEE workshop on Spontaneous Speech Processing and Recognition*, 2003.

[6] H. Koiso, Y. Horiuchi, S. Tutiya, A. Ichikawa, & Y. Den, "An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task Dialogs," *Language and Speech*, 41, pp. 295-321, 1998.