

# Analysis of text-reading speech produced by Japanese primary school children (age 6-12): Focus on Topic-Comment structure

Yasuko Nagano-Madsen

Department of Languages and Literatures, University of Gothenburg  
yasuko.nagano-madsen@spraak.gu.se

## Abstract

This paper presents a cross-sectional analysis of text (story)-reading speech produced by Japanese primary school children adopting the CIAIR Children's Voice Speech Corpus, comprising 253 children across six grades. A high percentage of pause insertion after the topic marker *wa* indicates that the Topic-Comment structure is well-comprehended and phonetically marked by Japanese primary school children. The peak relation of fundamental frequency (F0) between the domain of Topic (T) and Comment (C) shows the T>C pattern as dominant with a gradually increasing pitch range difference. The opposite pattern T<C also increases, indicating the upcoming of the focal prosody, which is typically associated with storytelling style.

## Introduction

Age-dependent characteristics of speech are essential for efficient speech communication. This paper presents a cross-sectional analysis of text (story) reading speech produced by Japanese primary school children (age 6-12) focusing on the manifestation of the Topic-Comment structure.

Prosodic correlates of information structure have received much attention in past decades. Kügler & Calhoun (2020) present an overview and typology of prosodic correlates for information structure defined as focus, givenness, and topicality across languages. The role played by prosody is most evident in languages like English

and other Germanic languages lacking grammatical markers to indicate information structure. Conversely, there are languages with rich morphosyntactic means of expressing information structure, which often leads to little or no contribution of prosody. It is unclear where Japanese fits in the typology as it is not included in Kügler & Calhoun's overview.

## Topic-Comment structure in Japanese

The Topic-Comment structure is one of the key concepts with which information structure in Standard Japanese can be encoded (Büring 2007). Japanese has a word order subject-object-predicate, but the sentence structure is dominated by a Topic-Comment relationship. The Topic indicates what is being talked about (often old information), followed by the speaker's comment, i.e., substantially new information. A topic particle *wa* usually follows the Topic.

Other Japanese information structure concepts include the particles *wa* versus *ga*, where some researchers consider *ga* a focus particle (cf. discussion by Heycock 2008). Others consider the key concepts of information structure in Japanese as focus, givenness, and topic/topicality (Féry & Ishihara 2016, Tomioka 2016).

Japanese intonation has been well studied with syntactic structure and focus manifestation (Pierrehumbert & Beckman 1988, Kubozono 1993, Ishihara 2011). However, such studies are experimental with systematically varied

components. The corpus used for the present study consists of short sentences with simple syntactic structure, with no clearly identified focus.

Therefore, this study adopts a Topic-Comment structure as a reference. Two types of prosodic correlates of Topic-Comment structure in Japanese have been reported. The first is a regular pause insertion after the topic marker *wa* (Sugito 1986). The second is intonation in terms of the F0 peak relation between the domain of Topic and Comment. Unless the Topic is contrastive, the F0 tends to be higher in the Comment domain, which carries new information (Nakanishi 2007, Nakagawa 2016).

## Method

### Speech corpus

The study utilized parts of Children Voice Speech Corpus (CIAIRVCV) [24] as a primary material. The corpus contains read-out speech of words in isolation and a story, “The Little Match Girl” by H. C. Andersen in Japanese. The recordings are without annotation and are clipped for each word or sentence(s). The participants are 253 Japanese-speaking Children from the Tokyo district. Table 1 displays the specifications of the children.

Table 1. Specification of speakers.

Grade	Age	Girl	Boy	Total
1	6-7	16	15	31
2	7-8	18	20	38
3	8-9	18	20	38
4	9-10	24	30	54
5	10-11	25	23	48
6	11-12	23	21	44
Total		124	129	253

### Sentences

The story consists of 32 sentences, of which 19 have a topic-comment structure. For analysis, two successive short sentences with a topic-comment

structure were selected. The Topic ends with a topic particle *wa*, and the rest belongs to the Comment. In (1), the Topic is ‘the girl’ and the subject is ‘stomach’; in (2), ‘the small hands’ is both the Topic and the subject.

(1)  
 [Syoojyo-wa]                      [onaka-ga herimasita.]  
 Girl-TOP                              stomach-NOM empty  
 ‘As for the girl, her stomach became empty.’

(The girl became hungry.)’

(2)  
 [Tiisana te wa]                      [imanimo kogoosodesita.]  
 Small hands-TOP                      about to freeze  
 ‘As for her small hands, they are about to freeze.’

### Analysis parameters

The data was analyzed using the following parameters: total reading time and pitch range in semitone for the two successive sentences, pauses, and intonation, which were captured as F0 peak relation in the domain of Topic vs. Comment. All the measurements were done manually using PRAAT, and the obtained data was processed using SPSS for statistical analysis and graphs. SUGI Speech Analyzer produced Figure 5.

## Results

### Reading time

The boxplot in Figure 1 displays the distribution of the total time taken to read two successive sentences for each grade (age). It includes pauses and repetitions. The boxplot shows decreasing medians and standard deviation with increasing ages. A one-way ANOVA revealed significant differences in reading time across the six groups  $F(5, 247)=48.9, p<.001$ . Post-hoc comparisons indicated a significant difference between grades 1 and 2 ( $p < .001$ ) as well as between grades 3 and 4 ( $p < .001$ ). It is characterized that text reading fluency develops in two steps and becomes more stable after grade 4 (9-10 years).

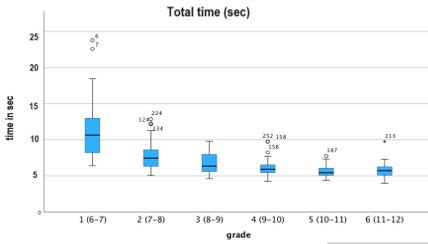


Figure 1. Reading time by grade. Vertical bars denote inter-speaker variability (SD).

### Pitch range

The pitch range (F0 max–F0 min for the entire utterance) distribution is shown in Figure 2 for each grade. The boxplot shows the increasing median and standard deviation from grade 1 to grade 3, and after grade 4, both become more stable, just like the reading time. A one-way ANOVA revealed significant differences in pitch range across the six groups  $F(5, 247)=5.4, p<.001$ . Post-hoc comparisons indicated a significant difference between grade 1 and other grades ( $p < .05$ ). It is concluded that the pitch range measured in semitone does not vary across grades except for grade 1.

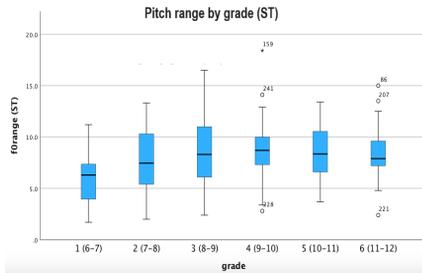


Figure 2. Pitch range. Vertical bars denote inter-speaker variability (SD).

### Pauses

The total number of pauses, including inter- and intra-sentential pauses for the two successive sentences, is shown in Figure 3 for each grade (age). Inter-speaker variability (SD) becomes stable from grade 3.

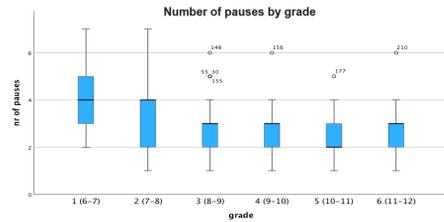


Figure 3. Number of pauses. Vertical bars denote inter-speaker variability (SD).

The location and frequency of pauses are shown graphically for linguistic categories in Figure 4. Pause frequency is calculated in percentages by dividing the number of pauses by their potential numbers.

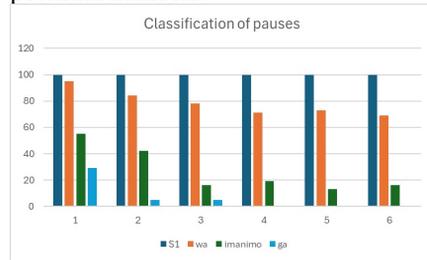


Figure 4. Pause location and frequency in %.

Figure 4 indicates that the location and frequency of pauses become stable from grade 4 (9-10 years). In the text, punctuation marks appear only for the sentences, and all the speakers (100%) inserted a pause after the first sentence. Around 70% of the topic marker *wa* is followed by a pause, while the particle *ga*, considered a focus marker by some researchers, is seldom accompanied by a pause. Another pause is found after the word *imanimo*, ‘almost,’ with a stable but less significant occurrence.

### Intonation

#### F0 peak relation

The analysis shows three main intonation patterns based on the F0 peak relation in Topic (T) and Comment (C). Figure 5 illustrates the three patterns termed Level T=C, Downstep T>C, and Upstep T<C.

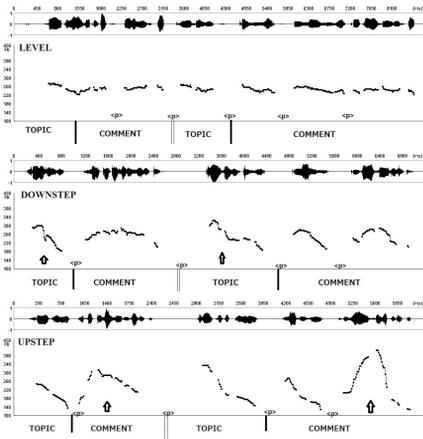


Figure 5. Typical F0 patterns for Level (T=C), Downstep (T>C), and Upstep (T<C). The speakers are all boys at 9 years old.

F0 peak values in the domains of Topic and Comment and the difference between the two F0 peaks are measured in semitones. F0 peak values for Topic (y-axis) and Comment (x-axis) are plotted with a reference line ( $y=x$ ) to examine the difference in pitch range. Tokens below the reference line indicate the intonation pattern T>C, while tokens above the reference line indicate opposite T<C. Figure 6 shows the F0 peak plot for grades 1 and 4 for comparison.

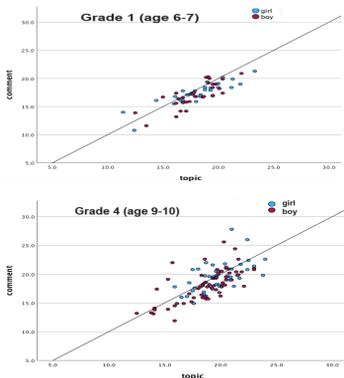


Figure 6. Distribution of F0 peaks with a reference line. Grade 1 (above) and Grade 2 (below).

### F0 peak range

Tables 2 and 3 compare the pitch range, SD, and the number of speakers for the two patterns (T>C and T<C). The mean pitch range is greater in the T<C pattern, indicating that the upstep patterns tend to show focal intonation. For this pattern, the speaker variability (SD) is more variable. The pitch range for the T>C (downstep) pattern is narrower than that of the T<C (upstep) pattern, with more minor standard deviations.

Table 2. F0 peak range for T>C patterns in semitone.

Grade	Mean (ST)	SD	N	%
1	1.22	0.90	51	80
2	1.35	0.79	60	79
3	1.54	1.00	60	78
4	1.49	1.06	77	70
5	1.66	1.06	68	69
6	1.31	0.80	58	70

Table 3. F0 peak range for T<C patterns in semitone.

Grade	Mean (ST)	SD	N	%
1	1.08	0.71	13	20
2	1.83	1.99	16	21
3	1.94	3.05	17	22
4	1.89	1.85	33	30
5	1.62	1.65	30	31
6	1.99	2.23	25	30

### Development of intonation patterns

The results of the developmental change in the intonation patterns are shown graphically in Figure 7. Since most tokens have the T>C pattern with considerable variation in pitch range, the pitch range difference below 0.7 semitones is classified as T=C in this graph to show the changes more clearly. The Mixed pattern indicates that the two sentences have a different pattern, one T>C, the other T<C. The graph shows that the level pattern is gradually reduced while other patterns increase gradually.

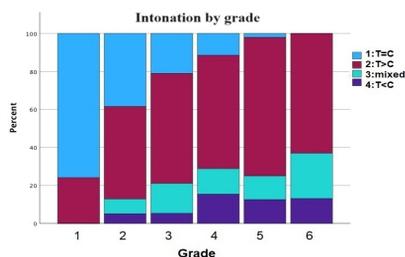


Figure 7. Development of intonation during the primary school ages.

## Discussion

Previous studies have reported that word-level lexical prosodic properties, such as lexical pitch accent in Japanese, are typically acquired before primary school ages (Shirose 2004, Fikkert et al. 2020 for an overview). The present study confirms the claim. All the children analyzed in this study demonstrated in their pitch accent patterns that they are native speakers of Tokyo Standard Japanese.

Several previous studies suggest what follows as the next stage of prosodic development: they include prosodic phrasing, intonation, and focus manifestation (Schwanenflugel et al., 2004; Wells et al., 2004; Frazier et al., 2006, Chen 2011, Chen et al., 2020, Romøren, & Chen 2022). The main developments found in the present study are along these lines.

The results of pause insertion in the present study agree with those reported by Sugito (1986), who analyzed Japanese broadcastings. Sugito found that the topic particle *wa* is usually followed by a pause regardless of punctuation, while pause insertion was rare after the particle *ga*, which some researchers consider a focus marker.

Unlike stable pause insertion, the intonation patterns show more variations and constant developments. The present study also reveals the upcoming focal intonation around grade 4 (9-10 years) in the Comment domain.

The difference in speaking style has a strong influence on the intonation patterns. The T<C (upstep) pattern indicates focal intonation, particularly those with an extensive pitch range. They are found in speech, where children try to read the text in a storytelling style. When the text is read more like a smooth narration, intonation becomes T>C (downstep) instead.

Godde et al. (2021) compared read-aloud material in French for two age groups: children (9–12 years) and adolescents (13–17 years). The results reveal many differences between the two groups, including differences in pausing and prosodic boundaries. The prosodic patterns of Japanese children in reading text are likely to develop further. It should also be noted that reading prosody differs from spontaneous speech (cf, Godde et al. 2020). Therefore, the upcoming focal intonation may be earlier in spontaneous speech in Japanese, which does not require reading competence.

## Conclusion

The Topic-Comment structure in Japanese is well comprehended by primary school children and marked phonetically by inserting a pause in between. The intonational pattern captured as F0 peak relation between Topic and Comment changes throughout the primary school ages. The most significant changes occur between grades 3 (8-9 years) and 4 (9-10 years) regarding reading time, pause insertion, and focal intonation. The focal intonation is associated with the storytelling style.

## Acknowledgment

We gratefully acknowledge using the “CIAIR Children Voice Speech Corpus (CIAIR-VCV)” provided by the Speech Resources Consortium at the National Institute of Informatics, Japan, which was crucial to completing this paper. Chen 2011.

## References

- Büring, D. (2007). Topic and Comment. In P. Hogan (Ed.), *The Cambridge Encyclopedia of the Language Sciences* (p. 866).
- Chen, A. (2011). Tuning information packaging: Intonational realization of Topic and focus in child Dutch. *Journal of Child Language*, 38, 1055–1083.
- Chen, A., Esteve-Gibert, N., Prieto, P., & Redford, M. A. (2020). Development of phrase-level prosody from infancy to late childhood. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody* (pp. 553–562).
- CIAIR, Nagoya University. (2006). *CIAIR Children Voice Speech Corpus (CIAIR-VCV)*. Speech Resources Consortium, National Institute of Informatics. <https://doi.org/10.32130/src.CIAIR-VCV>
- Féry, C., & Ishihara, S. Introduction.(2016). In C. Féry & S. Ishihara (Eds), *The Oxford Handbook of Information Structure* (pp.1-15).
- Fikkert, P., Liqun, L., & Ota, M. (2020). The acquisition of word prosody. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody* (pp. 541-552).
- Frazier, L., Carlson, K., & Clifton, C. Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, 10(6), 244-249.
- Godde, E., Bailly, G., & Bosse, M. (2021). Pausing and breathing while reading aloud: development from 2nd to 7th grade in French-speaking children. *Reading and Writing*, 35, 1 - 27.
- Godde, E., Marie-Line, B., & Gérard, B. (2020). A review of reading prosody acquisition and development. *Reading and Writing*, 33(2), 399-426.
- Heycock, C. (2008). Japanese - WA, - GA, and information structure. In S. Miyagawa (Ed.), *The Oxford Handbook of Japanese Linguistics* (pp. 54-83).
- Ishihara, S. (2011). Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua* 121, 1870-1889.
- Kubozono, H. (1993). *The Organization of Japanese Prosody*. Tokyo: Kuroshio Publishers.
- Kügler, F., & Calhoun, S. (2020). Prosodic Encoding of Information Structure: A typological perspective. In C. Gussenhoven & A. Chen (Eds), *The Oxford Handbook of Language Prosody* (pp.454-467).
- Nakagawa, N. (2016). *Information Structure in Spoken Japanese: Particles, Word Order, and Intonation* (Ph.D. thesis). Kyoto University.
- Nakanishi, K. (2007). Prosody and scope interpretations of the topic marker wa in Japanese. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation* (pp. 177-193). Springer.
- Pierrehumbert, J., & Beckman, M. E. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.
- Romøren, A. S. H., & Chen, A. (2022). The acquisition of prosodic marking of narrow focus in Central Swedish. *Journal of Child Language*, 49(2), 213-238.
- Schwanenflugel, P. J., Hamilton, A. M., Wisenbaker, J. M., Kuhn, M. R., & Stahl, S. A. (2004). Becoming a Fluent Reader: Reading Skill and Prosodic Features in the Oral Reading of Young Readers. *Journal of educational psychology*, 96(1), 119–129.
- Shirose, A. (2004). Acquisition of lexical accent by children and characteristics of accent distribution of Japanese dialects [In Japanese]. *The Japanese Journal of Language in Society*, 7(1), 41-50.
- Tomioka, S. (2016). Information Structure in Japanese. In C. Féry & S. Ishihara (Eds), *The Oxford Handbook of Information Structure* (pp.753-773).
- Wells, B., Peppé, S., & Goulondris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31(4), 749-778.