PROSODIC REALIZATION OF SYNTACTIC PHRASE AND CLAUSE BOUNDARIES IN TOKYO JAPANESE

Shinichiro Ishihara

Centre for Languages and Literature, Lund University
shinichiro.ishihara@ostas.lu.se

ABSTRACT

This paper presents the results of a production experiment on the syntax–prosody mapping in Tokyo Japanese, whereby the phonetic realization of syntactic phrase and clause boundaries were systematically compared. The data consist of 672 utterances of read speech from fourteen subjects. The results corroborate earlier findings on the mapping of syntactic phrases to phonological phrases, but do not support the claim that clauses are mapped to intonational phrases. Data from a few subjects suggest instead that there may be a mapping of discourse-related domains (such as topic and comment) to intonational phrases.

Keywords: syntax–prosody mapping, Japanese, syntactic phrase, clause, intonational phrase

1. INTRODUCTION

Recently proposed theories of intonation [1, 2, 3, 4, 5] put forward a strong version of the syntax–prosody mapping hypothesis, which states i) that there are, language universally, three prosodic categories—phonological word (ω), phonological phrase (φ) and intonational phrase (ι)—above the foot, and ii) that these three prosodic categories show correspondence to three syntactic categories, words, phrases, and clauses, respectively. Match Theory [5], for example, explicitly hypothesizes that syntactic phrases are mapped to φs and clauses to ιs.

In the case of (Tokyo) Japanese, however, it is still unclear whether there is such a clear-cut distinction between the phrase-level and clause-level syntax–prosody mapping. Previous studies have shown that left edges of syntactic constituents—be it phrasal or clausal—are realized with high f_0-peaks. For example, the left edge of a relative clause has shown to be marked by a high f_0-peak [6, 7]. [8] showed that left edges of syntactic phrases (XPs) are realized with an f_0-boosting phenomenon called *metrical boost* and that metrical boost is larger when there are two syntactic boundaries at one position than when there is only one syntactic boundary. In these studies, however, no distinction was made between phrase and clause boundaries, or between φ and ι.

According to [9], five phonetic cues in (1) distinguish ιs from φs.

(1) a. ι-final f_0-lowering
b. creaky vowels at the end of ι
c. obligatory pauses at the end of ι
d. stronger pitch resets at the onset of ι
e. larger pitch rises at the onset of ι

Using these cues, [9] showed that syntactic clauses are mapped as ιs in Japanese, as clause boundaries are prosodically marked differently from phrase boundaries. However, since the coordinated clauses tested in this study were all main clauses, it still remains to be examined whether embedded clauses, such as clausal objects, exhibit phonetic cues in (1). This point is particularly important because it has already been acknowledged [4] that some languages do not map embedded clauses to ι despite the proposed mapping hypothesis.

Opposite results have also been reported (though with a caveat). In the study of the prosody of embedded clauses [10] and relative clauses [11] containing a single content word, it was found that the embedded/relative clause undergoes downstep[1] relative to the preceding material, indicating that they are contained in a single φ, which entails that these clauses are not mapped to ιs. However, since Japanese is known to be subject to the so-called *rhythmic (a.k.a. binarity) effect* [8, 12], which triggers grouping of two phonological words into a single φ, the one-word embedded/relative clauses may have been grouped together with the preceding word, forming a φ, due to this effect. It remains to be seen if downstep can be observed in longer embedded clauses.

2. EXPERIMENT

2.1. Methodology

A phonetic production experiment was conducted with fourteen native speakers of Japanese (nine females, five males). They were undergraduate or graduate students at a university located in Tokyo

and were from Tokyo or surrounding areas.

Four conditions (named 0xp, 1xp, 2xp, and cp) were tested with 4 lexical items each. The stimuli (target sentences + 192 fillers) were presented on a computer screen one sentence at a time. Each subject read all of the stimuli three times, in three different pseudo-randomized orders. This procedure yielded a total of 672 samples of the target sentences (14 speakers×4 conditions×4 items×3 repetitions).

The recorded target sentences were annotated using Praat [13]. They were first forced aligned for segment and word boundaries using Julius ver. 4.5 [14], Julius Segmentation Kit ver. 4.3.1, and pyjuliusalign ver. 2.0 [15]. Word boundaries were checked and corrected manually. Then, for each word, f_o-maximum and f_o-minima before and after the f_o-maximum were measured, with the help of Praat scripts. f_o-maxima were chosen either from a point that is judged as corresponding to the peak of a lexical pitch accent (H*+L in X-JToBI notation, [16]) or on the BPM at the end of each word. In addition, the presence of a detectable pause after Word1 and Word2 was checked by listening to the data and by visually examining pitch contours.

2.2. Stimuli

The four conditions share the same first four nouns (Word1–Word4) and the verb, which are all lexically accented. The cp condition additionally contained a complementizer (*to*) and a verb. (2)–(5), which are one of the four sets of lexical items used in the study, differ in terms of the number and the type of syntactic boundaries in front of the target word, Word2 (*Naoya*).

(2) No XP boundary (0xp)
[$_{NP}$ *Yuuta-to Naoya-wa*] [$_{VP}$ *imooto-o*
Y.-and N.-TOP sister-ACC
paatii-ni maneita]
party-to invited
'Yuta and Naoya invited their sisters to the party.'

(3) One XP boundary (1xp)
Yuuta-wa [$_{VP}$ *Naoya-o* [$_{NP}$ *imooto-no*
Y.-TOP N.-ACC sister-GEN
paatii-ni] *maneita*]
party-to invited
'Yuta invited Naoya to his sister's party.'

(4) Two XP boundaries (2xp)
Yuuta-wa [$_{VP}$ [$_{NP}$ *Naoya-no imooto-o*]
Y.-TOP N.-GEN sister-ACC
paatii-ni maneita]
party-to invited
'Yuta invited Naoya's sister to the party.'

(5) Clause boundary (cp)
Yuuta-wa [$_{CP}$ *Naoya-ga imooto-o paatii-ni*
Y.-TOP N.-NOM sister-ACC party-to
maneita to] *omotteita*
invited that was.thinking
'Yuta believed that Naoya invited his sister to the party.'

In 0xp (2), there is no XP-boundary between Word1 and Word2 as they form an NP together. In 1xp (3), there is a VP-boundary between Word1 and Word2. In 2xp (4), in addition to the VP-boundary, there is a boundary of the NP containing Word2 and Word3. Lastly, in cp (5), there is an embedded clause boundary. If the number of syntactic boundaries (0xp, 1xp, 2xp) and/or the type of boundary (phrase or clause) affects prosody, the realization of the target word is expected to reflect these effects.

2.3. Predictions

Based on the the previous findings summarized in §1, two predictions can be made as to the realization of the target word (Word2) in the four conditions (0xp, 1xp, 2xp, cp). The first prediction concerns the mapping of syntactic phrases to φs. When comparing 0xp, 1xp, and 2xp, the f_o-peak of Word2 in 0xp is predicted to be lower than that of 1xp and 2xp, because Word2 of 0xp is subject to downstep due to the lack of syntactic boundary to its left [8, 17, 18]. Furthermore, when 1xp and 2xp are compared, it is predicted that the f_o-peak of Word2 in 2xp is realized higher than that in 1xp, because *metrical boost* is stronger when more than one syntactic boundary co-occurs at one position [8].

(6) Predictions 1: 0xp < 1xp < 2xp

The second prediction concerns the mapping of clauses to ιs. According to (1), an ι-boundary shows a higher pitch rises than a φ-boundary. If the embedded clause in cp is mapped to an ι, it is expected that the f_o-peak of Word2 is higher in cp than in 1xp and 2xp.

(7) Predictions 2: 1xp, 2xp < cp

In addition to the expected differences in f_o-peak height, pauses are expected consistently before word2 in cp, as per (1c), if the embedded clause is mapped to ι and preceded by another ι.

3. RESULTS

The pooled data of all the fourteen subjects confirmed Prediction 1 while they did not confirm Prediction 2. Figure 1 shows the normalized means of the f_o-maximum on the target word (Word2) in each condition. 0xp, in which no metrical boost but downstep is expected, shows the lowest f_o-maximum. 1xp and 2xp show higher f_o-peaks compared to 0xp. Furthermore, 2xp is significantly higher than 1xp (2-sided t-test: $t = -3.9389, df = 333.1, p = 9.974\text{e}-05$). In contrast, there is no significant difference between 2xp and cp (2-sided t-test: $t = 0.11212, df = 327.79, p = 0.9108$).

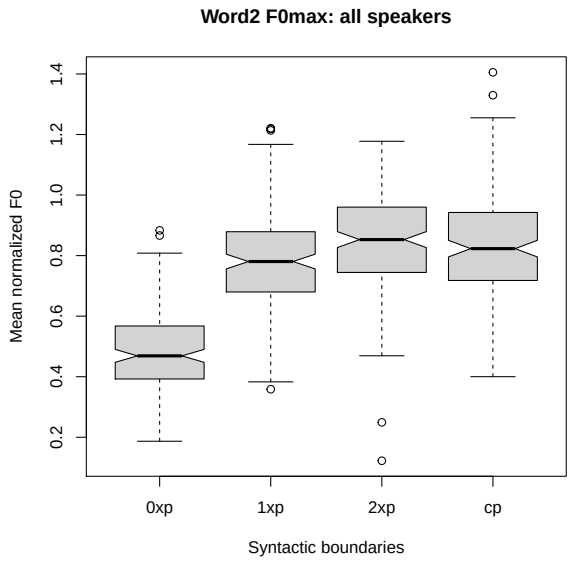

Figure 1: Boxplot of normalized means of f_o-maximum on Word2 for all subjects

Table 1 shows the frequency of detectable pauses after Word1 and Word2. The frequency of pauses after Word1 in cp (41.7%) is comparable to those in 1xp and 2xp (36.3% and 38.7%, respectively), which suggests that there is no sign of ι-boundary at this position. Another point to note is that there are much fewer pauses after Word1 in 0xp (5.4%) and after Word2 in 2xp (17.3%) than other places. This is presumably because these positions are between two coordinated nouns (*Yuuta-to Naoya-wa* 'Yuta-and Naoya-TOP' in (2); *Naoya-no imooto-o* 'Naoya-GEN sister-ACC' in (4)), that is, where there is no syntactic XP boundaries. At these positions, downstep was also observed (as in 0xp in Figure 1).

There were inter-speaker variations that need to be noted. In what follows, speakers are numbered (01–14) and indexed with f/m to indicate the gender.

	Word1		Word2	
0xp	5.4%	(9)	72.6%	(122)
1xp	36.3%	(61)	53.0%	(89)
2xp	38.7%	(65)	17.3%	(29)
cp	41.7%	(70)	33.9%	(57)

Table 1: Percentages (and counts) of pauses detected after Word 1 and Word2 in each condition. The number of total data points in each position and condition is 168.

As for Prediction 1 in (6), while all but one speaker (05m) showed a significant contrast in 0xp < 1xp, a half of the speakers did not show the expected contrast 1xp < 2xp.

(8) a. Significant difference (1xp < 2xp):
Speakers 01f, 03f, 04m, 06f, 07m, 09m, 10f
b. No significant difference (1xp $\not<$ 2xp):
Speakers 02f, 05m, 08m, 11f, 12f, 13f, 14f

Regarding Prediction 2 in (7), i.e, the comparison of 2xp and cp, the subjects can be divided into three groups, as in (9). The first group (9a) did not show any significant difference between 2xp and cp, similarly to the pooled data in Figure 1. The second group (9b) showed a significantly lower f_o-max in cp than in 2xp (2xp > cp), i.e., in the opposite direction to the prediction. Only the last group (9c) showed the expected difference, 2xp < cp. This means that the first two groups, (9a) and (9b), do not confirm the prediction of the clause-level mapping, while group (9c) is in line with it.

(9) a. No significant difference between 2xp and cp ($2\text{xp} \not< \text{cp}, 2\text{xp} \not> \text{cp}$):
Speakers 02f, 03f, 08m, 09m, 13f
b. cp significantly lower than 2xp (2xp > cp):
Speakers 01f, 04m, 06f, 07m, 10f
c. cp higher than 2xp (cp > 2xp)
Speakers 05m, 11f, 12f, 14f

There is a noteworthy correlation between the grouping in (8) (1xp vs. 2xp) and that in (9) (2xp vs. cp): Group (9b) speakers all belong to Group (8a), while Group (9c) speakers belong to (8b).

Sample pitch tracks of cp taken from these two groups illustrate this difference. Pitch tracks from Group (9b) tend to show a lower Word2 f_o-peak in relation to the preceding word (Word1) in cp (Figure 2). This group also showed a similar pattern in 1xp. The Word2 f_o-peak values in 1xp and cp were therefore comparable to each other. Figure 2 shows downstep from Word1 to Word2 followed by the pitch resetting on Word3, which means that Word1 and Word2 are phrased together within a single φ.

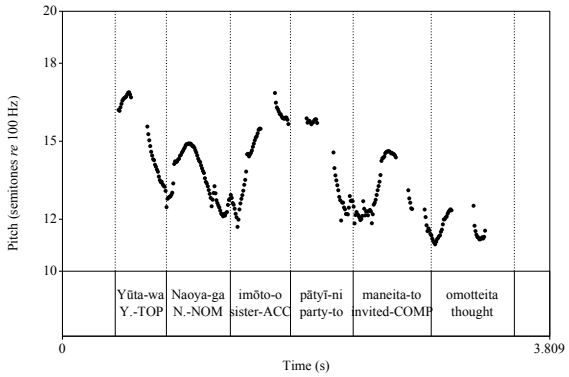

Figure 2: Sample pitch track of cp from a speaker in Group (9b). Word2 is downstepped relative to Word1.

This phrasing is presumably due to the rhythmic effect [8, 12], which can be explained as avoidance of φ containing a single prosodic word. Since both Word1 and Word2 are single-word NPs in 1xp and cp, they are phrased together due to this effect. This phrasing pattern excludes the possibility to have an ι boundary between Word1 and Word2, both in 1xp and cp conditions.

In contrast to Group (9b), Group (9c) speakers showed a high f_0-peak on Word2, as exemplified in Figure 3. Furthermore, these speakers showed a tendency to insert a pause between Word1 and Word2 in cp. Given that pauses are one of the phonetic cues of ι boundary (see (1c)), it can be concluded that there is an ι boundary at this position.

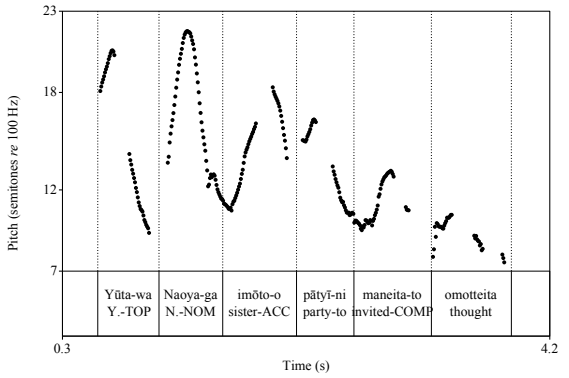

Figure 3: Sample pitch track of cp from a speaker in Group (9c). Word2 shows a high f_0-peak, preceded by a pause.

In their data, however, high f_0-peaks and pauses were observed in other conditions as well. Group (9c) speakers showed a high f_0-peak in 1xp and 2xp as well as cp. Note that none of the Group (9c) speakers showed the contrast 1xp $<$ 2xp, as in (8b).

In fact, the location of pauses seems to be rather correlated with the location of topic marker *wa*. The topic marker *wa* is attached to Word1 in 1xp, 2xp, and cp, and to Word2 in 0xp. Three of the four speakers in Group (9c) show a clear tendency to place a pause after a noun marked with *-wa*. Table 2 shows that pauses are consistently found in all the locations with *-wa* in these speakers' data.

	Word1		Word2	
0xp	0.0%	(0)	100.0%	(36)
1xp	75.0%	(27)	36.1%	(13)
2xp	80.6%	(29)	25.0%	(9)
cp	100.0%	(36)	11.1%	(4)

Table 2: Percentages (and counts) of pauses detected after Word 1 and Word2 in each condition by speakers 11f, 12f, and 14f.

4. DISCUSSION

The results of this study suggest that there is no mapping of embedded clause to ι in Japanese. Although this conclusion is still in line with Match Theory, as claimed by [5], the results also showed a stronger correlation between the location of the topic marker *wa* and the presence of pauses for some speakers, which seems to suggest that ιs may be more directly correlated with discourse-related notions such as topic and comment.

It has been claimed that ιs correlate with illocutionary forces, or speech acts associated with them [19, 20, 21, 22]. There is also a claim that the introduction of topics is a type of speech act [23]. Taking these claims into consideration, Ishihara [24] proposes that, contrary to the syntax–prosody mapping hypothesis presented in §1, ι is not part of the syntax–prosody mapping, and that it is rather part of the discourse–prosody mapping: Speech acts are mapped as ιs in the prosodic representation. Further investigation is needed to test this claim.

Another point to be noted is the rhythmic effect observed in Group (9b) speakers' data. Downstep on Word2 in cp (Figure 2) suggests that two single-word NPs can be grouped into a single φ even when there is a clause boundary in between. This results explains the otherwise puzzing results in [10, 11].

5. ACKNOWLEDGMENT

This study was supported by Swedish Research Council (2018-01539). I sincerely thank Natsumi Goto, Thea Johansson, and Dillen Luis Smit for their help with data annotation, and anonymous reviewers for useful comments. All errors are my own.

6. REFERENCES

[1] J. Itô and A. Mester, "Prosodic adjunction in Japanese compounds," in *Formal Approaches to Japanese Linguistics 4 (FAJL4)*, 2007, pp. 97–111.

[2] ——, "Recursive prosodic phrasing in Japanese," in *Prosody Matters: Essays in Honor of Elisabeth Selkirk*, T. Browsky, S. Kawahara, T. Shinya, and M. Sugahara, Eds. London: Equinox Publishing, 2012, pp. 280–303.

[3] ——, "Prosodic subcategories in Japanese," *Lingua*, vol. 124, no. 1, pp. 20–40, 2013.

[4] E. Selkirk, "On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure," *Gengo Kenkyu*, vol. 136, pp. 35–74, 2009.

[5] ——, "The syntax-phonology interface," in *The Handbook of Phonological Theory*, 2nd ed., J. Goldsmith, J. Riggle, and A. Yu, Eds. Oxford, UK: Blackwell, 2011, pp. 435–484.

[6] T. Uyeno, H. Hayashibe, and K. Imai, "On pitch contours of declarative, complex sentences in Japanese," *Annual Bulletin of Research Institute of Logopedics and Phoniatrics*, vol. 13, pp. 175–187, 1979.

[7] T. Uyeno, H. Hayashibe, K. Imai, H. Imagawa, and S. Kiritani, "Syntactic structure and prosody in Japanese: A study on pitch contours and the pauses at phrase boundaries," *Annual Bulletin of Research Institute of Logopedics and Phoniatrics*, vol. 15, pp. 91–108, 1981.

[8] H. Kubozono, *The Organization of Japanese Prosody*. Tokyo: Kurosio Publishers, 1993.

[9] S. Kawahara and T. Shinya, "The intonation of gapping and coordination in Japanese: Evidence for intonational phrase and utterance," *Phonetica*, vol. 65, no. 1–2, pp. 62–105, 2008.

[10] M. Hirayama and H. K. Hwang, "Not all XPs affect prosody in Japanese," in *NELS 46: Proceedings of the North Eastern Linguistics Society*, C. Hammerly and B. Prickett, Eds., 2016, pp. 95–104.

[11] ——, "Relative clause and downstep in Japanese," in *Supplemental Proceedings of the 2018 Annual Meeting on Phonology*, K. Hout, A. Mai, A. McCollum, S. Rose, and M. Zaslansky, Eds. Washington, DC: Linguistic Society of America, 2019.

[12] T. Shinya, E. Selkirk, and S. Kawahara, "Rhythmic boost and recursive minor phrase in Japanese," in *Proceedings of the Second International Conference on Speech Prosody*, 2004, pp. 345–348.

[13] P. Boersma and D. Weenink, "Praat: doing phonetics by computers [Computer programm]," Version 6.1.53, retrieved 8 September, 2021 from http://www.praat.org, 2021. [Online]. Available: http://www.praat.org

[14] A. Lee and T. Kawahara, "Julius v4.5," 2019.

[15] T. Mahrt, "pyjuliusalign v.2.0," 2019. [Online]. Available: https://github.com/timmahrt/pyJuliusAlign

[16] K. Maekawa, H. Kikuchi, Y. Igarashi, and J. Venditti, "X-JToBI: An extended J_ToBI for spontaneous speech," in *Proceedings of The 7th International Conference on Spoken Language Processing (ICSLP)*, Denver, Colorado, September 16–20 2002, pp. 1545–1548.

[17] E. Selkirk and K. Tateishi, "Syntax and downstep in Japanese," in *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda*, C. Georgopoulos and R. Ishihara, Eds. Dordrecht: Kluwer Academic Publishers, 1991, pp. 519–543.

[18] S. Ishihara, "Japanese downstep revisited," *Natural Language & Linguistic Theory*, vol. 34, no. 4, pp. 1389–1443, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11049-015-9322-8

[19] E. Selkirk, "Comments on intonational phrasing in English," in *Prosodies: With Special Reference to Iberian Languages*, S. Frota, M. Vigário, and M. J. ao Freitas, Eds. Berlin/New York: Mouton de Gruyter, 2005, pp. 11–59.

[20] G. Güneş, "Constraints on syntax-prosody correspondence: The case of clausal and subclausal parentheticals in Turkish," *Lingua*, vol. 150, pp. 278–314, 2014.

[21] ——, "Deriving prosodic structures," Ph.D. dissertation, University of Groningen, 2015.

[22] H. Truckenbrodt, "Intonation phrases and speech acts," in *Parenthesis and Ellipsis: Cross-Linguistic and Theoretical Perspective*, M. Kluck, D. Ott, and M. de Vries, Eds. Berlin/Boston: De Gruyter Mouton, 2015, pp. 301–349.

[23] C. Ebert, *Quantificational Topics: A Scopal Treatment of Exceptional Wide Scope Phenomena*. Dordrecht: Springer, 2009.

[24] S. Ishihara, "On the (lack of) correspondence between syntactic clauses and intonational phrases," in *Prosody and Prosodic Interfaces*, H. Kubozono, J. Ito, and A. Mester, Eds. Oxford, UK: Oxford University Press, 2022, pp. 420–456.

[1] In Japanese, downstep is a pitch range compression triggered by H*+L lexical pitch accents [18] and takes the φ as its domain. That is, f_0-peaks within a φ are cumulatively lowered relative to the previous one, and at the beginning of the following φ, pitch resetting takes places.