

ASYMMETRIES BETWEEN PRODUCTION, PERCEPTION AND COMPREHENSION OF FOCUS TYPES IN JAPANESE

Hyun Kyung Hwang

JSPS fellow, National Institute for Japanese Language and Linguistics, Japan

hwang@g.ninjal.ac.jp

Abstract

Distinct types of focus in Tokyo Japanese are investigated, with special attention given to their acoustic and perceptual characteristics. The results of production, prominence rating, and context retrieval tests demonstrate that prosodic differences between Informational Focus (IF) and Contrastive/Corrective Focus (CF) are subtle, but perceptible to listeners: CF involving a slightly greater F0 fall and longer duration is perceived more prominent than IF. However, it is also revealed that the perceived differences of prominence do not guide listeners to identify the focus types, implying that the prosodic differences between IF and CF in the speech signal alone do not play a functional role in terms of communication.

Index Terms: Prosody, Informational Focus, Contrastive Focus, Prominence Rating, Context Retrieval

1. Introduction

Information conveyed by a single sentence can vary depending on context. Frequently, speakers also *prosodically* mark the semantic differences in order to convey the information to listeners, which is redundant in the presence of contextual cues. Among different kinds of information that context triggers, the primary concern of the current study is prosodic characteristics associated with discourse-givenness and two types of focus.

Discourse-given (Given) material is what has already been made salient in the discourse. Prosodically, Given material is predicted to be nonprominent [1, 2, 3]. On the other hand, a focus can often be understood as the part that corresponds to the answer to the *wh*-part of *wh*-questions [4]. This use of focus is referred to *Informational focus* (IF) in this study. Another use of focus is *Contrastive/Corrective Focus* (CF), which has a limited set of alternatives [5, 6]. It has been widely recognized that focused constituents generally receive prosodic prominence. Although the prosodic realizations of focus have attracted intensive interest, there is no consensus regarding the role of *contrastiveness* associated with focus in the prosodic literature [7, 8, 9, 10, 11, 12]. Some report that contrastive and non-contrastive foci in English exhibit prosodically distinct realizations [7, 9, 10, 11]. More specifically, Breen et al. [9] observe that CF in an object position is produced with a higher maximum F0, while no acoustic difference is found between CF and IF in subject and verb positions. Other than English, Avesani and Vayra [13] observe prosodic differences between distinct types of focus in Florentine Italian. On the contrary, Bolinger [8] and 't Hart et al. [12] argue that there are no consistent acoustic differences between the subtypes of focus in English. Hanssen et al. [14] report a similar observation in Dutch.

Scant research has looked into issues related to the perception and comprehension of focus types. Krahmer &

Swerts [11] demonstrate that a contrastive adjective tends to be perceptually more prominent than a new adjective. Yet, Breen et al. [9] report that the acoustic cues associated with CF are not systematically interpreted as CF.

These controversial findings highlight the need for further research on phonetic realizations, and more comprehensive approach integrating the perception and comprehension of different types of focus in more varieties of languages. Also, this issue has theoretical implications for the interface between syntax and phonology: there is an extensive debate as to whether semantic accounts on the syntax-phonology interface should assume different subtypes of focus [15], or simply assume a single category of focus [16] or givenness [17].

In Japanese, while the prosodic realizations of focus have extensively explored, most previous studies on the relationship between focus and prosody deal with the characteristics of CF [18, 19]. Despite methodological differences, there is a general agreement that focus expands the pitch range of focused items, followed by post-focal pitch compression. Yet, acoustic characteristics of distinct subtypes of focus have rarely been explicitly discussed. Further, the perception and comprehension of subtypes of focus has not been addressed for this language.

Therefore, the goal of the current paper is to investigate distinct focus types in Japanese, and to provide more comprehensive data by considering various aspects of focus. Specifically, the following unresolved questions regarding distinct types of focus in Japanese are addressed; 1) Are the distinct subtypes of focus prosodically differentiated by speakers? 2) Are acoustic differences between IF and CF perceptually noticeable to listeners? 3) Do listeners retrieve and use acoustic information to distinguish IF vs. CF?

2. Methods

2.1. Production

2.1.1. Materials

The structure of tested material is given below.

- (1) Topic – Obj1 – Obj2 – Verb

The information condition of the preceding object (Obj1) was varied: it was given IF, CF, or discourse-given status. All the other elements except Obj1 were controlled to be Given information. A topic phrase and a second object (Obj2) were the targets of pre-focus and post-focus effect, respectively.

Two different sets of the same structure were tested. Different orders between direct objects and indirect objects were also recorded, as the word order in Japanese is relatively free. In order to vary the information status of Obj1, each target sentence was uttered as an answer to a different prompt question as shown in Table1.

Table 1: An example set of test sentences (direct object-indirect object order): a. Given, b. IF, c. CF

	Prompt	Target
a	Yúu-wa mémo-o Náo-ni mise-ta? 'Did Yuu show Nao the memo?'	un, Yúu-wa mémo-o Náo-ni mise-ta-yo. 'Yes, Yuu showed Nao the memo.'
b	Yúu-wa náni-o Náo-ni mise-ta? 'What did Yuu show Nao?'	Yúu-wa mémo-o Náo-ni mise-ta-yo. 'Yuu showed Nao the memo.'
c	Yúu-wa ronbun-o Náo-ni mise-ta? 'Did Yuu show Nao the paper?'	uun, Yúu-wa mémo-o Náo-ni mise-ta-yo. 'No, Yuu showed Nao the memo.'

A target sentence was constructed to be a plausible answer for any one of three prompt questions. Note that the target sentences in Table 1 are identical across the three information types, except *un* ‘yes’ in the Given (Table 1a) or *uun* ‘no’ in the CF condition (Table 1c). In recording, participants were asked to make a pause after those words in order to circumvent any effect of those words.

The target sentences are semantically ambiguous. In the Given case, the direct object *memo* is already made salient in the prompt question, therefore making this part of the target discourse-given. On the other hand, the direct object in Table (1b) and (1c) receives focus: in the IF case, the direct object is the part of an answer (the target) that corresponds to the wh-part of a wh-interrogative (the prompt). The same object in Table (1c) differs from that in (1b) in that there is an explicit contrast set for the element, which is *ronbun* ‘the paper’ in the prompt question.

In constructing test material, only accented items were used since the accentedness of a preceding word influences the F0 of following material [19, 20]. Also, the length of target phrases and their accent location were controlled: each phrase was 3-mora long, and bears an accent in the first mora across the tested material.

2.1.2. Recordings and measurements

Recordings were carried out in 3 separate sessions depending on their information condition, and each session was recorded with at least a day interval. In each session, 4 target sentences were interspersed with 4 filler sentences, in a pseudo-randomized order. The order of test material was identical for all participants. Participants were instructed to repeat the list of tested sentences ten times. The materials were recorded in a sound-attenuated booth at the National Institute for Japanese Language and Linguistics, using a Marantz digital recorder (PMD 661) and a SHURE microphone (Beta 58A). Participants were instructed to give natural renditions at a comfortable speed.

Two female and two male speakers of Tokyo Japanese, aged 24-39 participated in the recording. All were born and grew up in Tokyo or surrounding areas, and had no history of speech or hearing impairment.

In total, 480 utterances were obtained (2 object orders x 3 information types x 2 sets x 4 speakers x 10 repetitions). For acoustic measurements, phrase boundaries were first manually marked on each utterance. In measuring fundamental

frequencies, maximum F0 (MaxF0) and Minimum F0 (MinF0) values of each target phrase, and pitch falls from Obj1 to Obj2 (MaxF0/MinF0 of Obj1-MaxF0/MinF0 of Obj2) were extracted using a Praat script [21]. Duration of target phrases was also calculated.

2.2. Perception

A perception experiment was performed to test if listeners can perceive the observed acoustic differences among the distinct information conditions. A subset of the target sentences collected in the production test was presented. Following the methodology of [22], participants were asked to rate how prominent Obj1 was, *i.e.* the target of the on-focus effect, along the scale of 1 (not prominent at all) to 5 (very prominent). 30 native speakers of Tokyo Japanese participated in the test. Participants were students at International Christian University and Tokyo University of Foreign Studies. In total, 360 responses were collected (2 object orders x 3 information types x 2 sets x 30 participants).

2.3. Comprehension

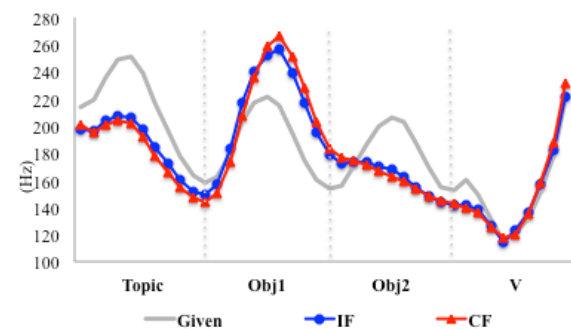
A prompt retrieval test using the prosodic cues of a given target sentence was conducted in order to investigate whether listeners indeed use prosodic cues from the speech signal, and retrieve the information that speakers intended to convey by distinguishing the two types of focus. The same participants of the perception test took part in a forced-choice question-answer matching task for which the identical stimuli of the perception experiment were utilized. Participants were instructed to listen to each utterance, and to choose the question that they thought the speaker was answering. Three choices, the prompt questions for Given/IF/CF information type in the production test, were provided.

3. Results and discussion

3.1. Acoustic differences

Time-normalized F0 curves in Hz, averaged all renditions by 4 speakers are demonstrated in Figure 1. Overall, IF and CF exhibit extremely similar curves: compared to the Given case, both types of focus involve on-focus F0 expansion (Obj1) and post-focus F0 compression (Obj2).

Figure 1: Time-normalized mean F0 contours: Vertical lines indicate phrase boundaries – Topic, Obj1, Obj2, and Verb.



Surprisingly, it is also observed that F0 peaks of the pre-focus phrase (Topic) is remarkably lower in IF and CF than in the Given case. For the two types of focus, the curves in Figure 1 do not seem to show distinct prosodic correlates of *contrastiveness* with respect to F0. The statistical significance of these observations was tested using JMP 9. Table 2 below summarizes the results of one-way ANOVAs. All reported effects were significant at the $p < 0.05$ level.

Table 2: Results of statistical analyses.

df (2, 477)	MaxF0	MinF0	Duration
Topic	F=32.9736 P<.0001*	F=10.5731 P<.0001*	F=0.1423 P=.8674
Obj1	F=23.4691 P<.0001*	F=3.5000 P=.0310*	F=7.6145 P=.0006*
Obj2	F=15.9853 P<.0001*	F=1.2833 P=.2781	F=28.9798 P<.0001*
Obj1-Obj2	F=297.6041 P<.0001*	F=5.2637 P<.0055*	

In the pre-focus region (*i.e.* Topic), there is a significant main effect of information status in F0 values, whereas no significant difference was found in duration. IF and CF pattern together with respect to MaxF0 & Min F0: post-hoc comparisons reveal that the MaxF0 of Topic in the Given condition is significantly higher than that in IF and CF (Given > IF = CF), indicating *pre-focus pitch compression* for both types of focus.

In the on-focus region (*i.e.* Obj1), IF and CF yield significantly higher MaxF0s compared to the Given case (CF = IF > Given). The same tendency that IF patterns together with CF is also found in the comparison of duration; compared to Given, both types of focus exhibit longer duration of Obj1 (CF = IF > Given). With respect to MinF0, only Given and CF are significantly different (Given > IF > CF).

Turning to the post-focus region (*i.e.* Obj2), the two types of focus exhibit significantly lower Max F0s (Given > IF=CF), confirming post-focus compression. However, no significant effect of information condition was observed in MinF0s (CF = IF = Given). Duration, on the other hand, differentiates all the information types (Given > IF > CF).

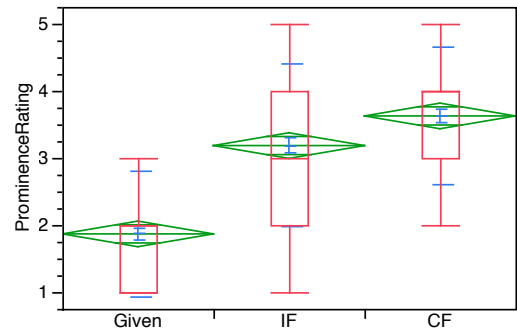
Finally, the MaxF0 fall from Obj1 to Obj2 differentiates all the information conditions (CF > IF > Given). For MinF0s of Obj1-Obj2, on the other hand, it is indicated that IF patterns together with Given (Given = IF > CF).

In comparing IF and CF, while the difference is subtle, the result demonstrates that CF involves slightly greater F0 fall and shorter duration of post-focus material than IF. Similar results on the acoustic differences between IF and CF were reported in English [9], and Seoul Korean [23]. In both studies, CF appears to be prosodically stronger.

3.2. Perceived prominence

The boxplot in Figure 2 demonstrates the rated prominence of Obj1 in the three information conditions. The result exhibited that CF is judged as the most prominent (average 3.63) among the information conditions. Also, listeners judged IF (average 3.18) as more prominent than Given (average 1.87).

Figure 2: Boxplot of rated prominence for 3 information types.



There are significant differences in perceived prominence among the distinct information types ($F(2,357) = 88.7363$, $P < .0001^*$). A post-hoc comparison confirms that CF is perceived most prominent, IF more prominent than Given, suggesting that listeners perceive the prominence differences in the speech signal which were marked by speakers. This result is in accordance with the finding in English [11], implying that a semantic theory should assume distinct subtypes of focus to capture this difference in perception.

3.3. Prosodic effect on context retrieval

In the absence of other contextual information, retrieving a prompt question using only prosodic cues from the target stimuli is a quite difficult task. Overall, listeners identified the matching questions poorly; the average accuracy was 65.8 % for Given, 70.8 % for IF, and 48.3 % for CF. From a functional perspective, the considerable confusion between the two types of focus seems to show that the acoustic differences do not play a critical role in distinguishing focus types. The confusion matrix between the information type of stimuli and of retrieved prompt is given in Table 3.

Table 3: Confusion matrix between information type of stimuli and retrieved prompt questions.

% Responses		Stimuli		
		Given	IF	CF
Responses	Given	65.8	4.17	4.17
	IF	22.5	70.8	47.5
	CF	11.7	25.0	48.3

When stimuli of the Given condition were presented, more than 34 % of them were incorrectly judged as focus. Interestingly, IF responses (22.5 %) were greater than CF ones (11.7 %) in the Given condition. Yet, the two types of focus were often confused with each other. In particular, responses for stimuli of the CF condition were nearly equally divided between IF and CF. It is conceivable that context, which is the primary domain for CF coding, is the dominant cue for the comprehension of CF rather than the prosodic information, which is redundant for listeners.

Although the accuracy rate exhibited by IF stimuli reaches 70.8 %, it is not unreasonable to expect that the ease of access

to IF facilitated the IF responses. That is, the IF case (wh-interrogative & answer pair) is presumably the easiest information condition for listeners to accommodate among the tested information types. It should be noted that 47.5 % of CF data were incorrectly judged as IF, whereas only 25 % of IF were matched with CF prompts. This asymmetry also suggest some bias towards IF, which is not surprising because CF requires a specific context that could be difficult for listeners to come up with.

4. Conclusions

This paper discussed acoustic, perceptual, and functional aspects of two distinct focus types in Japanese through a series of experiments. The comprehensive approach to subtypes of focus in this study attempted to better understand the nature of focus, and to provide empirical data to evaluate theories on the syntax-phonology interface.

Results of the present study demonstrate that both IF and CF consistently yield considerable F0 expansion of focused items and F0 compression of post-focus material. Interestingly, focus in Japanese involves F0 compression not only of post-focus items, but also of pre-focus items.

Yet, a prominence rating test reveals that the two types of focus are prosodically differentiated. Though the prosodic effect of distinct types of focus is quite limited, the statistically significant prosodic differences between two types of focus imply that the subtle difference between IF and CF may play a crucial role in the perception of prominence.

Indeed, the two types of focus are perceptually differentiated. Listeners tend to perceive CF as more prominent than IF. The results of the production and the prominence rating test seem to support semantic accounts where distinct marking of subtypes of focus is assumed [15]. Yet, the debate over distinct types of focus marking is still open as the context retrieval test demonstrates that the perceived prosodic prominence between IF and CF alone does not successfully guide listeners to identify the type of focus.

The results reported in this paper indicate directions for further research. In considering the prosodic realization of focus, further realities of the pre-focus compression in Japanese would be worth pursuing. Also, since the current study only examined two parameters, F0 and duration, more work is necessary to investigate other acoustic correlates of focus. Further, the results of the prominence rating and context retrieval tests do not rule out the possibility that other prosodic cues such as intensity, phonation type, relative prominence between the target and neighboring phrases also contribute to the perception of prominence and identification of focus type. Thus, it is necessary to tease the acoustic cues apart, and to test each parameter separately in order to confirm the finding of the current study. Such a study will allow us to determine more exactly the nature of distinct subtypes of focus.

5. Acknowledgements

I am grateful to John Whitman, Haruo Kubozono, and Timothy Vance for helpful comments and discussions. This work was supported by Grants-in-Aid for JSPS fellows 22 · 00308.

6. References

[1] Ladd, D. R. 1980. *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

[2] Selkirk, E. 2006. "Bengali intonation revisited: An optimality theoretic analysis in which FOCUS stress prominence drives FOCUS phrasing". In C. Lee, M. Gordon and D. Büring (eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation*, 215–244. Dordrecht: Springer.

[3] Féry, C. and V. Samek-Lodovici. 2006. "Focus projection and prosodic prominence in nested foci". *Language* 82.1: 131–150.

[4] Jackendoff, R. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

[5] Chafe, W. 1976. "Givenness, contrastiveness, definiteness, subjects, topics and point of view". In C. N. Li (ed.) *Subject and Topic*, 25–55. New York: Academic Press.

[6] Krifka, M. 2007. "Basic Notions of Information Structure". In C. Féry, G. Fanselow & M. Krifka. (eds.), *Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS) 6*: 13-56. Potsdam: Universitätsverlag Potsdam.

[7] Bartels, C. & Kingston, J. 1994. "Salient pitch cues in the perception of contrastive focus". In P. Boach & R. Van der Sandt (eds.), *Focus & natural language processing [IBM Working Papers on Logic and Linguistics, 6]*, 1-10. Heidelberg, Germany.

[8] Bolinger, D. 1961. "Contrastive accent and contrastive stress". *Language*, 37: 83-96.

[9] Breen, M., Fedorenko, E., Wagner, M., Gibson, E. 2010. "Acoustic correlates of information structure". *Language and Cognitive Processes* 25 (7/8/9): 1044-1098.

[10] Couper-Kuhlen, E. 1984. "A new look at contrastive intonation". In R. Watts & U. Weidman (eds.), *Modes of interpretation: Essays presented to Ernst Leisi*. 137-158. Tübingen, Germany: Gunter Narr Verlag.

[11] Krahmer, E. & Swerts, M. 2001. "On the alleged existence of contrastive accents". *Speech Communication*, 34: 391-405.

[12] 't Hart, J., Collier, R. & Cohen, A. 1990. *A perceptual study of intonation*. Cambridge, UK: Cambridge University Press.

[13] Avesani C and Vayra M. 2003. "Broad, narrow and contrastive focus in Florentine Italian". *Proceedings of The 15th International Congress of Phonetic Sciences*, Barcelona, 1803-1806.

[14] Hanssen J, Peters J and Gussenhoven C. 2008. "Prosodic Effects of Focus in Dutch Declaratives". *Proceedings of Speech Prosody*, 609-612, Campinas, Brazil.

[15] Selkirk, E. 2007. "Contrastive Focus, Givenness and the Unmarked Status of "Discourse-New"". In Féry, C., G. Fanselow, and M. Krifka (eds.): *The Notions of Information Structure*, Interdisciplinary Studies on Information Structure 6: 125–145.

[16] Rooth, M. 1992. "Theory of Focus Interpretation". *Natural Language Semantics* 1: 75–116.

[17] Schwarzschild, R. 1999. "GIVENness, AvoidF and other Constraints on the Placement of Accent". *Natural Language Semantics* 7(2): 141–177.

[18] Poser, W. 1984. *The Phonetics and Phonology of Tone and Intonation in Japanese*. Ph.D. dissertation, MIT.

[19] Pierrehumbert, J., and Beckman, M. 1988. *Japanese Tone Structure*. Cambridge, Mass.: MIT Press.

[20] Kubozono, H. 2007. "Focus and Intonation in Japanese: Does Focus Trigger Pitch Reset?". In Ishihara, S. (eds.), *Interdisciplinary Studies on Information Structure 9: Proceedings of the 2nd Workshop on Prosody, Syntax and Information Structure*.

[21] Xu, Y. 2005-2011. ProsodyPro.praat. Available from: <http://www.phon.ucl.ac.uk/home/yj/ProsodyPro/>

[22] Gussenhoven, C. 1983. "Testing the reality of focus domains". *Language and Speech* 26: 61-80.

[23] Jun, S.-A. 2010. "Focus Domains and Focus Types: Their Realizations in English and Korean". Presentation at the 7th Workshop on Altaic Formal Linguistics.