Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Bilingual Production of Narrow Subject Focus in Japanese: Spelunking in Prosody

Onae Parker[1], *Christine Shea*[1,2]

[1]University of Iowa, Department of Linguistics
[2]University of Iowa, Department of Spanish and Portuguese
onae-parker@uiowa.edu, christine-shea@uiowa.edu

Abstract

Languages differ in how they mark prosodic focus–whether syntactically/morphologically or phonetically/phonologically or both, which bears important implications for bilingual language acquisition. This study presents data on subject prosodic focus marking by L1 English/L2 Japanese, heritage Japanese speakers and L1 Japanese speakers. English primarily marks subject focus prosodically with stress and pitch fall, while Japanese also uses an obligatory postpositional particle, /-ga/. Fundamental frequency (F0) is used for subject focus in both languages, but intensity is used more consistently in English. Previous studies have examined L1 Japanese/L2 English production of subject focus, but there is little data looking at L1 English/L2 Japanese, and even less examining heritage Japanese speakers. Data was collected using a semi-spontaneous production task. Results show that i) L2 speakers used intensity more than the other groups; ii) HS speakers used a smaller F0 fall than the L1 group overall, but a greater fall when the /-ga/ focus particle was present. These results suggest that L2 learners transfer L1 prosodic cues to their L2. Heritage speakers, however, combine prosodic marking strategies from their dominant and non-dominant languages.

Index Terms: bilingualism, pragmatics, focus, prosody

1. Introduction

Japanese and English provide a case study of two languages that differ in focus marking in multiple ways, particularly in the expression of narrow subject focus. English primarily uses pitch and stress (realized as higher intensity and duration) [1] and rarely uses morphology or syntax to mark focus. Japanese, on the other hand, uses both morphology and prosody to mark subject focus. The obligatory particle /-ga/ [2], [3], and a local rise in F0 operate together to indicate subject focus. Japanese does not consistently rely on duration or intensity cues [4], [5], [6]. Post-focal compression is a cue available in both languages, but it is unclear whether there are fine-grained distinctions in its use. Thus, bilingual speakers of English and Japanese must navigate potentially opposing pressures not only between the focus marking systems of the two languages, but also between their prosodic marking systems. The investigation reported here seeks to test the waters in examining how these systems interact in English-Japanese bilingual speakers' production of Japanese.

Very few, if any, studies have examined the prosodic characteristics of focus in L1 English-L2 Japanese production. However, there have been a few studies in the other direction—[7] and [8] examined L1 Japanese L2 English learners' production of focus in English. [7] found that Japanese speakers tended to produce the highest pitch peak on the left sentence periphery, instead of on the focused word like the English controls did. [8] observed low proficiency Japanese learners producing flatter pitch contours in the focus condition than English controls, although this improved with proficiency. Duration and intensity were not examined in these studies. [9], however, in investigating L1 Korean L2 English focus production, observed that Korean speakers did not increase duration or intensity (or pitch, in lower proficiency speakers), cues which are not available in Korean.

Fewer if any studies look at focus prosody in heritage speakers (HS) of Japanese, although studies on HSs of other languages have suggested that they pattern between monolingual speakers and L2 speakers in their prosody, employing some strategies characteristic of monolinguals, and some characteristic of L2 speakers [10], [11]. [11] found that heritage speakers of Spanish patterned in-between the native Spanish controls and L2 speakers—namely, while they used clefting and some other syntactic marking that the controls used and the L2 speakers did not, they also used the prosodic cues of pitch expansion and post-focal compression, that the L2 speakers relied on, and the controls did not. In [10] as well, heritage Mandarin speakers patterned between the controls and L2 speakers, producing pitch expansion more closely to the controls than the L2 group, but not patterning as closely with their post-focal compression. Results such as these tend to suggest that heritage speakers pattern more closely to the "baseline" than L2 speakers, or produce more target-like prosody, due to their exposure at a young age, more naturalistic input, or a combination of both. It is of particular interest, as HSs have been said to have an "advantage" in acquiring phonology over their L2 counterparts (see [12] for a more comprehensive discussion of this issue), but this phonological advantage may flatten at the higher level of prosody, and the interaction between prosody and discourse. Comparing L2 speakers and HSs can illuminate what variables predict the acquisition of (focus) prosody, and what experiential factors impact how unbalanced bilinguals navigate the multiple dimensions of pragmatics, prosody, and even syntax in producing focus.

Due to limitations in our data that will be expanded upon later, this study presents rough acoustic explorations into inter-group differences in the usage of F0 and intensity between L1 speakers of Japanese, L1 English L2 speakers of Japanese, and English-dominant speakers of Heritage Japanese. The analyses seek to answer the following questions:

10.21437/SpeechProsody.2024-104

1. Do L1 speakers, L2 speakers, and Heritage Speakers of Japanese differ in the degree of F0 fall after the focused noun?

2. Do these groups differ in the degree of intensity fall between the focused noun and the post-focus constituents?

3. Does the presence or absence of the focus particle /-ga/ affect the use of F0?

2. Methods

2.1. Participants

Participants were L1 Japanese (L1), Japanese Heritage Speakers (HL) (dominant language English) and L2 Japanese learners (L1 English). As the task took approximately 15 minutes to complete, participants were not compensated.

Table 1: *Participant information*

	# Participants	Mean Age (SD)	Mean Proficiency
L1 Speakers	5 (3 Female)	49.6 (15.64)	6/6
HS Speakers	4 (3 Female)	26.5 (4.51)	5.5/6
L2 Speakers	8 (2 Female)	33.57 (12.66)	3.75/6

All participants were recruited by word of mouth, and their bilingual status (HS or L2) was determined ultimately via a language background questionnaire. Participants self-rated their proficiency on a scale from 0 (cannot speak at all) to 6 (fluent). Participants were identified as HSs if they indicated any of the following: they had learned both English and Japanese at an age younger than five years old; and/or spoke Japanese with their family; and/or were born in Japan and moved to America at a young age (<10 years old); and had parent(s) from Japan. These questions were intended to catch HSs from a broad range of experiences—those who acquired Japanese and English simultaneously at the beginning, those born in Japan and who had moved to America at a young age, those who may not know when they were first exposed to Japanese but report using it with their family and having Japanese parent(s).

2.2. Procedures and Stimuli

This study was conducted during the COVID pandemic in 2020, and therefore was distributed online due to restrictions at the time. The basic design of the task was to have participants answer pre-recorded questions designed to elicit the target focus type. The critical condition elicited only narrow subject focus. Distractors elicited broad focus and object focus.

Table 2: *Example Question Stimulus*

Question	Target Answer
Onaka suita-no? Nani tabetai? Pan tabetai?	*(Iya),* ***ramen(-ga)*** *tabetai*
Are you hungry? What do you want to eat? Do you want to eat bread?	(No), I want to eat **ramen**.

In total, there were 5*4=20 critical stimuli and 5*4=20 distractor stimuli, but this pool was divided into two lists of 5*2=10 critical stimuli and 5*2=10 distractor stimuli, such that each participant encountered only one list of 20 total stimuli. This was done to avoid boredom and encourage consistent attention across the entire experiment.

The experimental task was constructed on PsychoPy v2021.2.3 and conducted on Pavlovia.com [13]. The procedure was as follows (Figure 1). Participants first answered a language background survey on Qualtrics, after which they were directed to Pavlovia.com, where the experimental task would be completed. Participants had three practice sessions before beginning the study. Their task was to imagine that they were having a conversation with a friend or close acquaintance. A participant would first see a screen containing an image depicting some scenario (for example, a hungry person). At the same time they would hear a question in Japanese, like that in Table 2. Upon clicking the screen, they would see an image of the answer to that question. In this scenario, they might see an image of noodles. They were instructed to answer the question in full sentences, using the information presented to them. This process repeated until the end of the experiment. This design attempted to elicit the most spontaneous responses from participants, avoiding influencing their syntax as much as possible. Participants across groups consistently responded to these images using the target word (e.g., "ramen" in our example)—any responses that used a non-target word (e.g. if they said "noodles") were eliminated from the analysis.

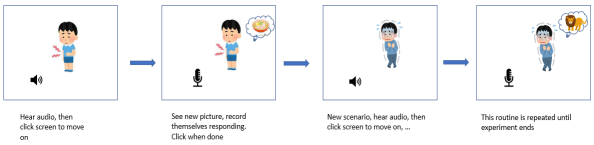


Figure 1: *Illustration of experiment flow*

All speakers were prompted to speak in the casual register. The questions were recorded in casual Japanese, and the instructions, by asking participants to imagine speaking with a close friend or family member, presented a context in which the casual register is used. This decision was made in light of the fact that HSs are most often exposed to informal spoken language as a result of the environment in which they learn their HL—the home and other less formal environments [14]. Therefore, it is more likely to elicit more naturalistic data by allowing HSs to speak in the register with which they are more familiar.

2.3. Acoustic and Statistical Analyses

Prior to a discussion of the acoustic analysis, a word about the limitations of this data is necessary. This study was originally designed to investigate morphosyntax—whether the bilinguals would omit the focused particle /-ga/, or use it similarly to the control group. We did find high rates of omission across all groups, raising the question that led to the present analysis: if a participant was not marking focus morphosyntactically, were they marking focus prosodically, and how? However, this means that while the nouns and verbs in the stimuli were controlled for difficulty level, they were not controlled for with acoustic analysis in mind. The stimuli differ in length (2-4 morae), and some contained voiceless obstruents. Furthermore, the critical stimuli always contained focus at the beginning of the sentence by nature of the task, which is a serious pitfall for reaching a balanced acoustic analysis of pitch activity. And finally, there was a mix of pitch accent patterns (Japanese is a pitch accent language, in which the presence and location of a high F0 peak is lexically contrastive), a further severe

limitation. Thus, this study presents a limited investigation into participants' prosodic behavior when marking subject focus.

The voice recordings collected via Pavlovia were initially stored as .webm files. These files were converted to .wav files on WavePad software, and saved as 44.1 kHz, 16 bit files. Any files with background noise were noise reduced on Audacity v 3.2.3 before being analyzed on Praat [15]. Because this study was conducted online, participants used personal devices to record their responses. This resulted in some participants' recordings being noisier than others'. Although all recordings underwent noise reduction as mentioned just previously, any recordings that were still of such poor quality that the response was unrecognizable, or pitch curves could not be reliably extracted, were removed from the analysis. Thus, ultimately any recordings that a) were of poor quality as described above, or involved b) long pauses/hesitations, laughing mid-speech, or non-target words or c) one-word responses, were removed from the analysis. This resulted in 144 recordings to be analyzed.

Because our variable of interest was the pitch fall between the high tone of the focused noun, and the pitch of the syllables following the focused word, F0 measurements were taken on the vowel of the syllable with the H tone on the focused word, and the vowels of the predicate. The predicates alternated between having two syllables (as in *kowai*), or three (as in *tabetai*). Finally, although the prompt was given in casual Japanese as described earlier, some participants still used honorifics—namely, they added the suffix *-desu* to the end of some predicates. The pitch on these final suffixes was not measured, but only the vowels that were consistent across participants. Measurements of pitch on the particle /-ga/ were taken, but not included in the analysis, due to frequent omissions and the small sample size, precluding a meaningful analysis. However, presence or absence of the particle was added as an interaction in the statistical analysis for pitch.

The maximum F0 of each target vowel for each word was measured, and the mean of the F0 measurements across the post-focal vowels was calculated, to account for the different number of post-focal syllables in the conditions. The raw F0 measurement (in Hertz) of the focus pitch, and the mean F0 of the post-focal vowels, were normalized to Z-scores in R [16] to account for sex differences. The difference in the Z-scores were then taken to form the variable for F0 fall. For these same vowels, the mean intensity was taken and the values were also normalized via Z-scores. One could have normalized each participants' F0 based on their own F0 range, but the approach adopted here was more appropriate for the small sample size, resulting in a slightly more conservative analysis.

Multi-level models were run in R, using the lme4 package [17], to investigate the following questions:

1) Do L1 speakers, L2 speakers, and Heritage Speakers of Japanese differ in the degree of F0 fall between the focused noun, and the post-focus constituents?
    a. Does the presence or absence of the focus particle /-ga/ affect the use of F0?
2) Do these groups differ in a fall in intensity between the focused noun and the post-focus constituents?

Two models were run to investigate these questions. Model 1 (Table 3) investigated the effect of the interaction between the fixed effects of Group (Helmert coded) and the presence/absence of the particle /-ga/ on F0 fall (normalized to Z-scores), calculated as the difference between the pitch peak of the focused word and the average F0 of the post-focal predicate. Subject was included as a random effect. Model 2 (Table 4), investigated the fixed effect of Group on intensity fall (the difference between the mean intensity of the focused word and post-focus predicate). Due to the small sample size, adding a fixed effect of the presence of /-ga/ and person as a random intercept resulted in singularity, and thus both were dropped.

3. Results and Analysis

Model 1 indicates that overall the HS group produced a smaller F0 fall after the focused word than the L1 group did (β = -0.51, SE = 0.193, t = -2.658, p = 0.015). Furthermore, there was a significant effect for the interaction between the HS group and the presence of /-ga/ (β=0.54, SE = 0.13825, t = 3.919, p = 0.001), indicating that HSs use a greater F0 fall when they use the focus particle than when they omit the focus particle. To compare mean F0 fall between all 3 groups, multiple comparisons were performed using the emmeans package [18] (Figure 2). No other significant differences were found between HS and L2 speakers, or L2 speakers and L1 speakers.

As seen in Table 4, the model for intensity identified a significant difference between the L2 group and the HS and L1 groups (β = 0.97, SE = 0.289, t = 3.346, p = 0.0026), suggesting that the L2 speakers produce a larger fall in intensity after the focused word than the L1 speakers and the HS speakers. Multiple comparisons were done for this model as well (Figure 3) but found no other notable differences.

4. Discussion

1. Do L1 speakers, L2 speakers, and Heritage Speakers of Japanese differ in the degree of pitch fall between the focused noun, and the post-focus constituents?

The analysis revealed a difference in F0 fall only between HSs and L1 speakers, and this difference seemed to lie mostly in an interaction between F0 fall and particle usage in the HS group. These results suggest either that there are no differences in post-focal F0 reduction between Japanese and English, or that these bilingual groups did not transfer F0 cue usage from English to Japanese. However, this analysis lacks any degree of subtlety, and there is no way to differentiate between these possibilities. Furthermore, if one were to perform a more gradient analysis of the pitch contour, they might find evidence of more nuanced differences between the groups. This is a direction to take for future work, as the literature leaves unanswered the exact nature of post-focal suppression/compression in English vs. Japanese.

Table 3: *Multi-Level Model of Post-Focus Pitch Fall by Group and Focus Particle*

Predictors	Estimates	CI	p
(Intercept)	-0.08	-0.39 – 0.22	0.568
Gapresent	0.04	-0.18 – 0.26	0.703
Heritage Group	-0.51	-0.92 – -0.11	**0.015**
L2 Group	0.13	-0.07 – 0.32	0.188
Gapresent × HS	0.54	0.25 – 0.83	**0.001**
Gapresent × L2	-0.01	-0.15 – 0.12	0.841
Random Effects			
σ^2	0.05		
$\tau_{00\ Number}$	0.21		
ICC	0.80		
$N_{\ Number}$	17		
Observations	28		
Marginal R^2 / Conditional R^2	0.283 / 0.858		

Table 4: *Fixed Effects Model of Post-focus Intensity Fall by Group*

Predictors	Estimates	CI	p
(Intercept)	-0.15	-0.45 – 0.16	0.339
HS Group	0.03	-0.38 – 0.44	0.886
L2 Group	0.32	0.12 – 0.52	**0.003**
Random Effects			
σ^2	0.58		
$\tau_{00\ Number}$	0.00		
$N_{\ Number}$	17		
Observations	28		
Marginal R^2 / Conditional R^2	0.293 / NA		

2. Do these groups differ in intensity fall between the focused noun and the post-focus constituents?

L2 speakers produced a greater fall in intensity after the focused word, than did L1 speakers and HSs. This result seems to point towards transfer of stress from the L1 (English), but the analysis lacks the power to lend substance to this difference. Future work with better methodology, more controlled stimuli, and better acoustic measurement should look at both intensity and duration to paint a more accurate and complete picture of how L2 (and potentially HS) speakers employ stress.

3. Does the presence or absence of the particle /-ga/ affect the use of pitch?

There was a significant interaction between /-ga/ and F0 fall in the HS group. The nature of this interaction is mysterious. A possible interpretation could be that when HSs do use the particle, they also use F0 to further accentuate the focused nature of the word. In other words, the saliency of the presence of the particle demands the additional cue of F0. But the limited nature of the data prevents much speculation. In addition, there was an interesting (though non-significant) trend in the L1 group in the opposite direction, such that when /-ga/ was omitted, the pitch fall was greater. If this trend were to be replicated in with a larger sample size and methodology more appropriate for prosodic analysis, this would suggest that when L1 speakers are not using the morphosyntactic cue for focus, they then rely more upon the prosodic cue to accomplish this pragmatic emphasis. Conversely, it could mean that when the morphosyntactic cue is present, prosodic marking is dampened. This would be an intriguing thread to follow for the interaction between the syntax-pragmatics interface and prosody.

Figure 2. *Between-group Comparisons in Mean Pitch Fall for Model 1*

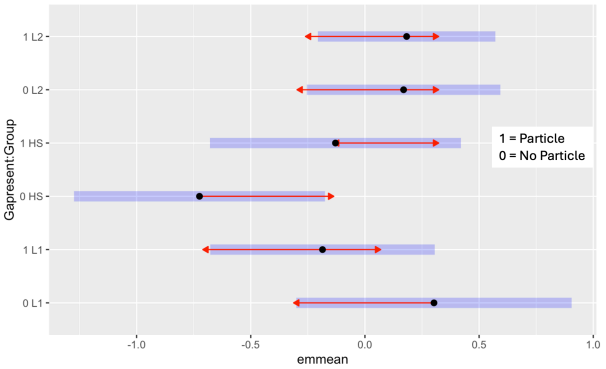
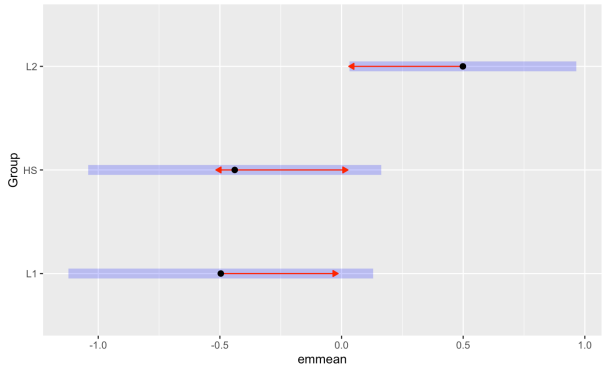


Figure 3. *Between-group Comparisons in Mean Intensity Fall for Model 2*



5. Conclusion

In summary, this exploratory analysis identified a) a possible interaction between usage of the particle and F0 in marking focus in the L1 group and HS group, and b) possible transfer from English in the usage of intensity to mark focus in the L2 group. Not insignificantly, this investigation observed differences in the prosodic behavior between the bilingual groups especially—the HSs indeed seemed to pattern between the L1 group and the L2 group, supporting the prediction that their different acquisitional trajectories (such as age of acquisition) affects their choice of prosodic strategies in marking focus in Japanese, and points to a multifaceted conference of prosodic and morphosyntactic cues that is worth investigating further with better stimuli, better methodology and analyses, and a bigger sample size.

6. References

[1] J. J. Venditti, K. Maekawa, and M. E. Beckman, *Prominence Marking in the Japanese Intonation System*. Oxford University Press, 2008.

[2] N. Nakagawa, *Information structure in spoken Japanese*. in Topics at the Grammar-Discourse Interface 8. Berlin: Language Science Press, 2020.
[3] M. Shimojo, "Properties of particle 'omission' revisited," *Toronto Working Papers in Linguistics*, 26, 2006.
[4] Y. C. Lee *et al.*, "A crosslinguistic study of prosodic focus," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Institute of Electrical and Electronics Engineers Inc., Aug. 2015, pp. 4754–4758.
[5] Y. Sugiyama, C. T. J. Hui, and T. Arai, "The effect of fo fall, downstep, and secondary cues in perceiving Japanese lexical accent," *The Journal of the Acoustical Society of America*, vol. 150, p. 2865, 2021.
[6] M. E. Beckman and J. B. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology Yearbook*, vol. 3, pp. 255–309, 1986.
[7] A. Fujimori, N. Yoshimura, and N. Yamane, "Japanese Learners' Acquisition of English L2 Prosody: L1 Transfer and Effects of Classroom Instruction," *Ars Linguistica*, *22*, p. 105-118, 2016.
[8] A. Fujimori, N. Yamane, N. Yoshimura, M. Nakayama, B. Teaman, and K. Yoneyama, "Development of L2 Prosody," in *Generative SLA in the age of minimalism: features, interfaces, and beyond : selected proceedings of the 15th Generative approaches to second language acquisition conference*, T. Leal, E. Shimanskaya, and C. A. Isabellli, Eds., John Benjamins, 2022, pp. 137–156.
[9] J. Liu, Y. Xu, and Y. Lee, "Post-focus compression is not automatically transferred from Korean to L2 English*," *Phonetics Speech Sci.*, vol. 11, no. 2, pp. 15–21, Jun. 2019.
[10] Y. Chen, Y. Xu, and S. Guion-Anderson, "Prosodic Realization of Focus in Bilingual Production of Southern Min and Mandarin," *Phonetica*, vol. 71, no. 4, pp. 249–270, Jul. 2015
[11] J.Y. Kim, "Heritage speaker's use of prosodic strategies in focus marking in Spanish." *International Journal of Bilingualism*, vol. *23*, no. 5, pp. 986-1004, 2019.
[12] C. Chang. "Phonetics and phonology" in *The Cambridge handbook of heritage languages and linguistics*, S. Montrul and M. Polinsky, Eds., Cambridge: Cambridge University Press, 2021, pp. 581-612.
[13] J. Brooks, "Peirce, J., & MacAskill, M. (Eds.). Building Experiments in PsychoPy," *Perception*, vol. 48, no. 2, pp. 189–190, Jan. 2019
[14] G. Scontras, Z. Fuchs, and M. Polinsky, "Heritage language and linguistic theory," *Frontiers in Psychology*, 2015.
[15] P. Boersma and V. van Heuven, "Speak and unSpeak with PRAAT," vol. 5, no. 9, 2001.
[16] R Core Team. (2020). "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org
[17] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *J. Stat. Soft.*, vol. 67, no. 1, 2015.
[18] Lenth R (2023). "emmeans: Estimated Marginal Means, aka Least-Squares Means." R package version 1.8.4-1, https://CRAN.R-project.org/package=emmeans