

Japanese Learners' Acquisition of English L2 Prosody: L1 Transfer and Effects of Classroom Instruction

Atsushi Fujimori, Noriko Yoshimura, and Noriko Yamane

Shizuoka University, University of Shizuoka, and University of British Columbia

1. Introduction

Second language (L2) learners experience difficulty at interfaces between the modules of Grammar (Nakayama and Yoshimura, 2015). According to the Interface Hypothesis, an external interface between syntax and discourse is more vulnerable than an internal interface between syntax and semantics (Belletti et al., 2007; Sorace and Filiaci, 2006; Tsimpli and Sorace, 2006). Pragmatically salient elements of an utterance such as focus are often identified with the help of prosodic properties of speech (Healey, 2003). It is anticipated that L2 learners face problems in producing prosodic focus marking when their first language (L1) is different from L2 in implementing focus (Baker, 2010; Gut and Pillai, 2014). This study is concerned with prosodic focus marking in English L2 acquisition at the syntax-discourse-phonology interface. We particularly examine how Japanese-speaking learners of English as a foreign language (EFL) differ from native speakers of English in producing focus prosodically, and whether explicit classroom instruction helps L2 learners acquire the prosodic focus marking in English.

The paper is organized as follows: Section 2 discusses basic linguistic facts pertinent to the manifestation of information focus in English and Japanese and presents a brief review of previous acquisition studies on L2 prosody. We identify possible problems for Japanese EFL learners to overcome during the acquisition of English L2 prosodic focus marking. Section 3 considers results of our comprehension and production tasks in Experiment 1. Section 4 investigates some significant effects of our classroom experiment on the learning of English L2 prosody. Section 5 concludes this discussion by analyzing the two sets of empirical results from the educational as well as acquisition perspectives.

2. Background

2.1. Focus marking

Information focus is a semantically non-presupposed part, and the most important and prominent constituent of a sentence, which is typically realized as an answer to the *wh* constituent in a *wh*-question (Zubizarreta, 1998).¹ How information focus is encoded differs across languages. In English, information focus is marked phonologically, receiving a highest pitch on the stressed syllable (Brown, 1983; Ito et al., 2004; Pierrehumbert and Hirschberg, 1990). The placement of accent is flexible and context-dependent. In (1b), for example, the object noun *cake* is an answer to the *wh* constituent *what* in (1a), and is assigned the prosodic prominence in the sentence. Likewise, in (2b), the subject noun *John* is focused as a response to the *wh* constituent *who* in (2a), thereby receiving the prosodic prominence in the sentence.

- (1) a. What did John eat?
b. He ate the [cake]_F.

- (2) a. Who ate the cake?
b. [John]_F ate it.

In Japanese, information focus can be encoded morphologically with the Case marker *ga* (Kuno, 1973; Heycock, 2008). In (3b), for example, *Taroo* is a non-presupposed part of the sentence as a response to *dare* ‘who’ in the question in (3a). The focused noun is marked with *ga*. That is, *ga* functions to morphologically identify focus in the sentence.

- (3) a. (Kyodai-no nakade) dare-ga dokushin desu ka
(Among your brothers), ‘who is single?’
b. [Taroo]_F-ga dokushin desu.
‘Taro is single.’

¹ The commonly held view is that there are two types of focus, information (thetic) focus and contrastive (categorical/identificational) focus (see Kiss (1998) and Zubizarreta (1998) for detailed discussion of their syntactic and semantic distinctions). We will only deal with information focus in the present paper.

In addition, in Japanese, the highest pitch tends to be placed on the sentence-initial word by default, and the pitch goes down towards the end of a sentence (Pierrehumbert and Beckman, 1988). Note that this “downstepping” (H*L) contour overlaps with the *ga*-marked sentence-initial word in focus in the case of (3b).

2.2. Previous L2 studies

A few studies have been conducted on the acquisition of English prosodic focus marking by L2 learners. Nava (2008) investigated ten L1 Spanish L2 English participants’ (five at advanced and the other five at intermediate proficiency level) oral production in a question and answer experiment. In Spanish, information focus appears with prosodic prominence at the right edge of the sentence, as shown in (4b).

- (4) a. ¿De qué te ríes?
 at what you laugh-PRS-PROG
 ‘What are you laughing at?’
- b. ¡Un pingüino está [bailando]_F!
 a penguin be-PRS-3SG dance-PROG
 ‘A penguin is dancing.’ (Nava, 2008: 158)

If an L1 transfer effect occurs, it is predicted that the participants would incorrectly put prosodic focus on the final word in the L2 utterance. The results showed that Spanish learners of both high and low proficiency incorrectly preferred placing prosodic prominence sentence-finally in English, as shown in (5b).

- (5) a. Why are you looking out the window?
 b. Madonna just walked [by]_F! (L1 Spanish-L2 English)
 c. ([Madonna]_F just walked by! (L1 English)
- (Nava, 2008: (16))

Grosser (1997) also observed the erroneous placement of prosodic prominence by German learners of English. For example, they placed the prosodic prominence on the sentence-final non-focused word *absent* in “nobody was absent” in answering the question “Is anybody absent today?” He attributed this error to the difference in

prominence patterns between German and English. Similarly, Gut (2009) found in his corpus study of 25 English-learning speakers from various L1 backgrounds that the learners had a general tendency to place a main pitch accent on the final word of an utterance, regardless of whether the word in question was new or given information (see also Ramirez Verdugo (2002)).

As for instructional effects on L2 prosody, Hardison (2004) conducted an experiment with a pretest-posttest design to examine the effectiveness of computer-assisted prosody learning. Sixteen English-speaking low-intermediate learners of French participated in 13 practice sessions where they were individually asked to read a set of French sentences aloud at a conversational rate. The students received auditory (hearing their utterance) and visual (seeing their pitch contour on a screen) feedback in real time. The pitch contour of a model's speech was also displayed on the same screen, so that the participants could compare the model pitch contours with their own. The pretest-posttest comparison indicated that their French prosody significantly improved after the training with audio-visual feedback, and their utterances sounded intelligible to native speakers of French (see also Levis and Pickering (2004)).

These previous findings point to two significant issues in L2 acquisition of prosodic focus marking, namely, L1 transfer and effects of explicit instruction. As such, our first question concerns whether Japanese EFL learners show a “down-stepping” phenomenon in producing English L2 prosody. Our second question is to examine whether explicit classroom instruction can help Japanese EFL learners produce appropriate English L2 prosody. To the best of our knowledge, no research has been conducted on the acquisition of English prosodic focus marking by Japanese EFL learners. If L1 Japanese affects L2 English in prosodic focus marking, it is predicted that the prosodic prominence will appear on the sentence-initial position. From the educational perspective, it is important to evaluate the effectiveness of classroom prosodic instruction on the remedy of L1 prosodic transfer in English L2 acquisition. To address these questions, we conducted two experiments on Japanese monolingual college students studying English in Japan.

3. Experiment 1

3.1. Participants

Experiment 1 was administered to examine the first research question, i.e., whether

Japanese EFL learners showed L1 transfer by placing a prosodic prominence sentence-initially. Sixteen Japanese college EFL learners were divided into two groups: Advanced Group ($n=6$) and Novice Group ($n=10$). The AG learners were third-year students majoring in English. Their TOEIC scores were all over 700, and their English proficiency was at CEFR B1 or B2 level. The NG learners were first-year students whose major was science. Their average TOEIC score was 418.5, and their English proficiency was at CEFR A2 level. Ten native speakers of North American English also participated in the experiment as a control group.

3.2. Tasks

The participants performed on a production task where they were asked to read aloud three written dialogs in pairs at a conversational rate. The dialogs took the form of questions and answers, as shown in (6) to (8). They were borrowed from a junior high school textbook *Sunshine English Course 2* (Kairyudo), thereby confirming that the vocabulary and the sentence structures of test tokens were comprehensible for the participants. The participants were given several minutes to practice, and at the end of the session, their utterances were recorded in Audacity. The intrinsic frequency (F0) of each vowel was measured in Praat.

(6) Token 1

Q: Where did you go last Sunday?

A: I went [fishing]_F with my friend [in the river]_F.

(7) Token 2

Q: Did you catch any fish?

A: Yes. I caught [three]_F fish.

(8) Token 3

(A: I saw your sister in the park this morning.)

Q: Oh, really? What was she doing?

A: She was [running]_F with her friend.

Before the production task, the participants also took a comprehension task in order to examine whether they could correctly identify focus in the information structure. The

participants were asked to mark a single word of each sentence in the dialogs (6)-(8) that they thought the most prominent with a nucleus stress in order for communication to be achieved. For each conversation, the focus was identified as part of the statement that answered the corresponding question in the context.

3.3. Results

The measurement results of the production task showed that the majority of the control group (NS) placed the highest pitch on the focused word, as shown in figures 1 - 3.^{2,3} The mean accuracy rates on pitch placement for AG and NG were 22% and 13%, respectively, which were both quite low. An ANOVA revealed that the mean accuracy rates of the three groups were significantly different ($F(2,75)=13.320, p<.000$). Post-hoc tests indicated that AG and NG were significantly different from NS ($ps<.000$). Noteworthy is that most of the learners placed the prosodic prominence on the subject, not on the focused word, as shown in figures 1 and 3. In figure 2, the prominence was also placed on the non-focused word, either the subject pronoun (*I* in (6) and *She* in (8)) or the verb (*caught* in (7)).



Figure 1 Distribution of prominence in token 1

² In token 3, 60% of the English speakers placed the prominence on the verbal part, either *was* or *running*.

³ We noticed some variants in the control group's utterances. This is partially due to simple contexts of the test dialogs. After the recording, some native speakers of English pointed out that they could place the prominence on a word other than the expected, as they elaborated the contexts to which they accommodated the test tokens on their own. We need to control for the contexts in our future research.

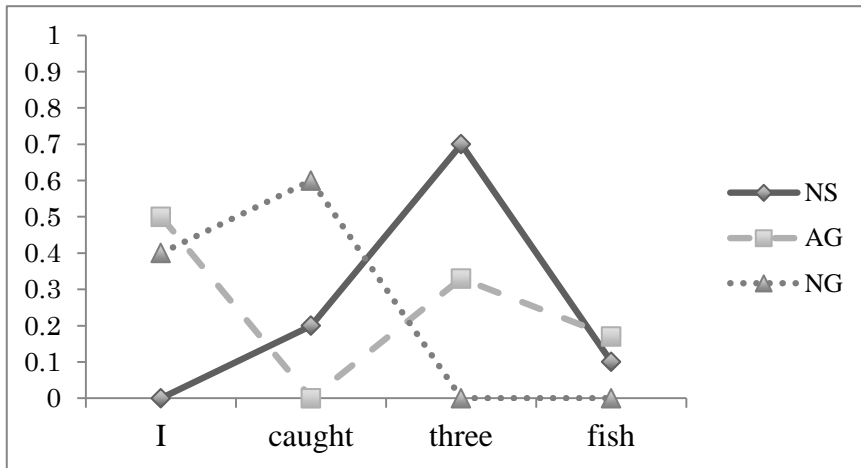


Figure 2 Distribution of prominence in token 2

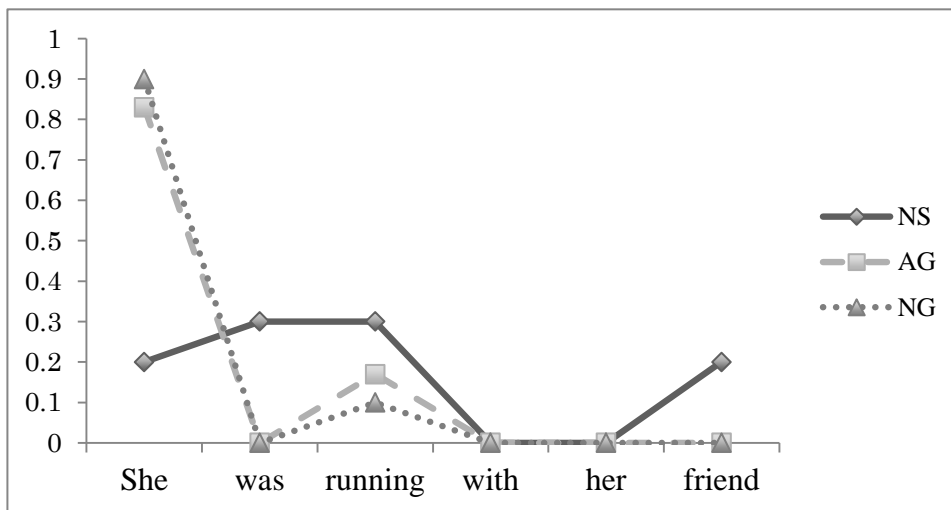


Figure 3 Distribution of prominence in token 3

Regarding the comprehension task, the mean correct percentages of AG and NG were 89% and 80%, respectively. These results indicated that both groups performed quite well on the task. Throughout the tokens, the performance of the AG learners was consistently at a high accuracy rate, as shown in Table 1. The answers of the NG learners were split between *fishing* and *river* in Token 1, although both were relevant to the *wh*-question *where did you go*. This suggests that the NG learners at least noticed that a key word substituting for the *wh*-word needed to be focused in the sentence. Notice also that none of the participants chose the subject noun as the prominent word in the sentence on the comprehension task, unlike in the production task.

Table 1 Answer distribution in comprehension task

Token	AG (<i>n</i> =6)		NG (<i>n</i> =10)	
T1 (6)	[fishing] _F 83%	river 17%	[fishing] _F 50%	river 50%
T2 (7)	[three] _F 83%	fish 17%	[three] _F 90%	caught 10%
T3 (8)	[running] _F 100%		[running] _F 100%	

4. Experiment 2

4.1. Participants

Experiment 2 was administered to examine whether explicit classroom instruction could help Japanese EFL learners improve their production of prosodic focus marking. Ten participants were chosen from the NG group to participate in this experiment because they had great difficulty in producing appropriate prosodic focus marking in Experiment 1.

4.2. Tasks

The participants took a perception task to examine whether they correctly perceived prosodic prominence. In the task, they were asked to identify the phonetically most salient word of each sentence and to mark it on a written answer sheet while they were listening to a recorded conversation by English native speakers. The test tokens were those used in Experiment 1 (see the tokens in (6) to (8) above).

In addition, the learners participated in eight repetitions of once-a-week oral instruction sessions (10 minutes per session). The instructor first confirmed that the participants correctly identified focus in comprehension. Then, the instructor taught them the basic rule, according to which the focused word is phonetically salient in English no matter where it occurs in the sentence. The participants were given opportunities to practice reading the test tokens aloud while listening to the recordings of the model speech. They were not given any “feedback” of their own speech during the sessions, as we tested the effectiveness of “feedforward” in the sense of de Bot (1980). That is, we examined whether the participants’ production could improve, if receiving information of the learned linguistic object (i.e., the basic rule of focus-prominence mapping), without information of the learners’ output (see also

Fujimori et al., 2015).

After the series of the instruction sessions, the participants retook a production task with the same set of the test tokens. This task was considered a posttest, whose results were compared with those of the pretest in Experiment 1.

4.3. Results

The mean accuracy rate in the perception task was 97%, as summarized in Table 2, which indicated that the NG learners correctly perceived prosodic prominence. The results also suggested that they had no hearing impairment.

Table 2 Answer distribution in perception task

Token	Prominent word	NG (<i>n</i> =10)	
T1 (6)	fishing	fishing 100%	
T2 (7)	three	three 100%	
T3 (8)	running	running 90%	friend 10%

In the pretest of the production task, only four out of 30 tokens (13.3%) showed the appropriate placement of the highest pitch on the focused element, as in Table 3. However, in the posttest, the appropriate response rate improved to 73.3%. A one-sample T-test indicated that there was a significant difference between the two tests ($t(59)=23.043$, $p<.000$). Table 3 also shows that although the sentence subject was prominent quite often at the pretest, it did not receive such erroneous prosodic prominence at the posttest: 60% decreased to 10% in (6), 40% to 10% in (7), and 90% to 60% in (8).⁴

⁴ One might wonder why the appropriate response rate was relatively low in Token 3 (8). Namely, 40% of the participants placed the prosodic prominence on the focused word “running”, as seen in Table 3. However, important is the fact that the rate improved from 10% to 40% after the instruction. Their improvement is also seen in the pitch patterns of Figure 6. The relatively high rate of the erroneous prominence production on the subject “she” may be due to the acoustic characteristics of the sibilant [ʃ] and the high vowel [i] contained in the pronoun: Sibilants have a long noise duration and among sibilants, the voiceless palato-alveolar fricative [ʃ] has the greatest noise amplitude (e.g., Jongman et al., 2000). The high-front vowel [i] universally contains a high pitch (Whalen and Levitt, 1995). Bear in mind that native speakers’ production of the same pronoun in the same sentence often gets contracted to the next word “was”: [ʃi wəz] becomes [ʃwəz] and the pronoun does not receive the highest pitch.

Table 3 Distribution of prominence in pretest and posttest

Token	Prominent word	Pretest			Posttest	
T1 (6)	fishing	I 60%	fishing 30%	went 10%	fishing 90%	I 10%
T2 (7)	three	caught 60%	I 40%		three 90%	I 10%
T3 (8)	running	She 90%	running 10%		She 60%	running 40%

Moreover, the individual analysis of the results revealed that the eight participants who did not perform well in the pretest could appropriately produce the correct prominence after the instruction, as summarized in Table 4.

Table 4 Number of individual performances by the NG group in pretest and posttest

Appropriate placement of prosodic prominence	Pretest (n=10)	Posttest (n=10)
Three (out of three tokens)	0	4
Two	0	4
One	4	2
None	6	0

Furthermore, the pitch measurement showed that in the posttest, the focused word was given higher pitch than that in the pretest, as shown in Figures 4 to 6. As a consequence, the mean pitch range between the focused word and the lowest F0 was greatly expanded in the posttest, as summarized in Table 5: 80.08 increased to 101.82 in (6), 12.76 to 62.37 in (7), and 42.01 to 59.89 in (8).

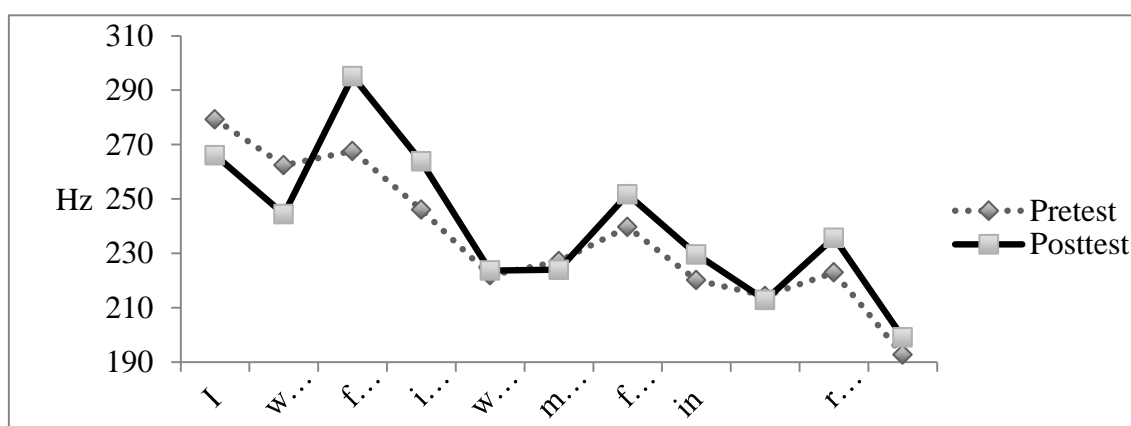


Figure 4 Mean pitch patterns of Token 1 in pretest and posttest

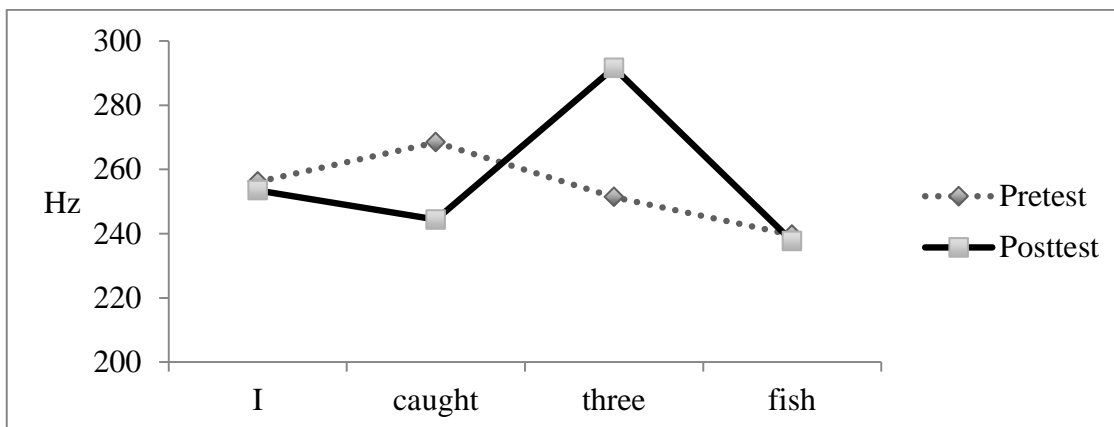


Figure 5 Mean pitch patterns of Token 2 in pretest and posttest

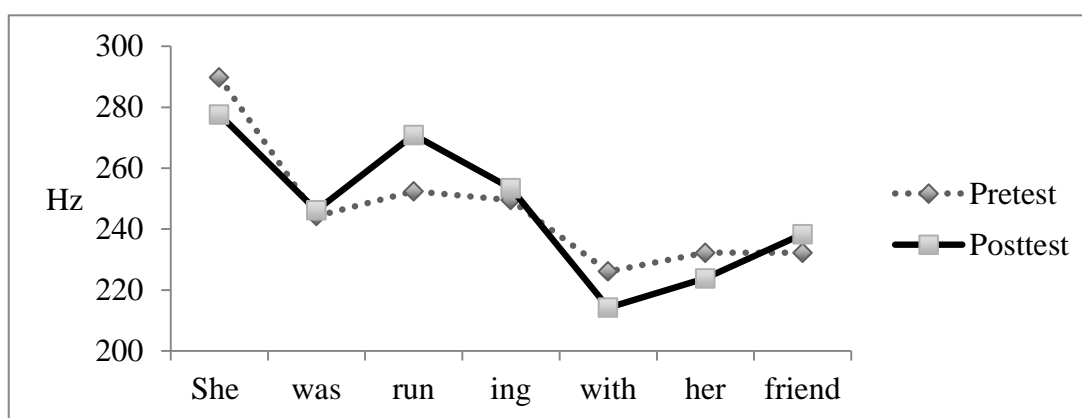


Figure 6 Mean pitch patterns of Token 3 in pretest and posttest

Table 5 Mean pitch ranges between the focus and the lowest F0 (Hz)

	Pretest	Posttest
Token 1 (6)	80.08	101.82
Token 2 (7)	12.76	62.37
Token 3 (8)	42.01	59.89

Notice also that in the posttest, the sentential subject lowered in F0 in each token.

5. Discussion

In the present study we investigated the following two research questions.

- (i) Do Japanese EFL learners show a “down-stepping” phenomenon in producing English L2 prosody?

- (ii) Can explicit classroom instruction help Japanese EFL learners produce appropriate English L2 prosody?

The results of our first experiment showed that Japanese EFL learners tended to place a prosodic prominence on the sentence-initial word rigidly, whereas English native speakers shifted prosodic marking depending on which word is focused. That is, our answer to the question (i) is positive. Given the comprehension result that the learners could correctly understand the focused element in the sentence, we suppose that the learners had difficulty in mapping their understanding of focus onto their prosodic production of focus, regardless of their English proficiency (see also Yamane et al. (2015) for similar results). From our discussion in Section 2, we maintain that the downstep contour incorrectly occurred at the left-edge of the sentence due to an effect of their L1 Japanese prosodic property.

Our answer to the second question is also positive. The L2 learners participated in a series of oral instruction sessions, and were offered several opportunities to practice the dialogs while listening to the relevant recordings, without any feedback of their own speech from instructors. Nevertheless, their production of prosodic focus marking improved to a large extent. Note that this large improvement was made only through oral instruction, without any visual representation of the physical speech. This seems to suggest that explicit oral instruction can be effective in helping Japanese EFL learners map their focus interpretation onto their prosodic production in English L2 performance.

To conclude, the present study confirmed the effectiveness of classroom instruction on the acquisition of English L2 prosody. A question remains with respect to whether such instruction effects will remain available for L2 learners after a long interval. Another question concerns whether L2 learners can apply their learned knowledge to similar but novel utterances in real-life communication. These questions require further investigation.

Acknowledgements

This paper is an extended version of our talk presented at the 41st annual meeting of the Japan Society of English Language Education (JASELE). Portions of the materials in this discussion were presented in Fujimori et al. (2014) and Yoshimura et al. (2015).

We thank the participants in the experiments and the audience at the conference for their generous assistance and encouraging questions. We appreciate Valerie Wilkinson proofreading our draft. We would also like to thank *Ars Linguistica* reviewers for very helpful comments. This work has partially been supported by the Grants-in-Aid for Scientific Research (B) 26284077 and (C) 26370700 from the Japan Society for the Promotion of Science.

References

- Baker, R. E. (2010). *The Acquisition of English Focus Marking by Non-Native Speakers*. Doctoral dissertation, Northwestern University.
- Belletti A., Bennati, E. & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language and Linguistic Theory*, 25, 657-689.
- Brown, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements*, 67–77. Berlin: Springer Verlag.
- de Bot, C. L. J. (1980). The role of feedback and feedforward in the teaching of pronunciation – an overview. *System*, 8, 35-45.
- Fujimori, A., Yoshimura, N., & Shirahata, T. (2014). Acquisition of prosodic focus marking by Japaense ESL learners. *IECE Technical Report*, TL 2014-20, 49-53.
- Fujimori, A., Yoshimura, N., & Yamane, N. (2015). The development of CALL materials for learning L2 English prosody. *Conference Proceedings of ICT for Language Learning, 8th edition*: 249-254.
- Gass, S., & Selinker, L. (1993). *Language Transfer in Language Learning*. Amsterdam: John Benjamins.
- Grosser, W. (1997). On the acquisition of tonal and accentual features of English by Austrian learners. In A. James & J. Leather (Eds.), *Second language speech: Structure and process*, 211–228. Berlin: De Gruyter.
- Gut, U., & Pillai, S. (2014). Prosodic marking of information structure by Malaysian speakers of English. *SSLA*, 36, 283-302.
- Gut, U. (2009). *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt: Peter Lang.
- Hardison, D. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology* 8: 34-52.
- Heycock, C. (2008). Japanese -wa, -ga, and information structure. In S. Miyagawa & M. Saito (Eds.), *The Oxford Handbook of Japanese Linguistics* (pp. 54-83). New York: Oxford University Press.
- Ito, K., Speer, S. R., & Beckman, M. E. (2004). Informational status and pitch accent distribution in spontaneous dialogues in English. In A. Riehl & T. Savella (Eds.), *Proceedings of the International Conference on Spoken Language Processing*, 279–282. Ithaca, NY: Cornell University.

- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108, 1252-1263.
- Kiss, K. É. 1998. Identificational focus versus information focus, *Language* 74, 245-273.
- Kubozono, H. (1993). *The Organization of Japanese Prosody*. Tokyo: Kurosio Publishers.
- Kuno, S. (1973). *The Structure of the Japanese Language*. Cambridge, MA: MIT Press.
- Levis, J., & Pickering, P. (2004). Teaching intonation in discourse using speech visualization technology. *System* 32: 505-524.
- Nakayama, M., & Yoshimura, N. (2015). The modularity of grammar in L2 acquisition. In M. Nakayama (Ed.), *Handbook of Japanese Psycholinguistics*, 235-270. Berlin: Mouton de Gruyter.
- Pierrehumbert, J. & Beckman, M. (1988). *Japanese Tone Structure*. Linguistic Inquiry Monograph 15. Cambridge: MIT Press.
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication*, 271–311. Cambridge, MA: MIT Press.
- Ramirez Verdugo, D. M. (2002). Non-native interlanguage intonation systems: A study based on a computerized corpus of Spanish learners of English. *ICAME Journal* 26: 115-132.
- Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22, 339-368.
- Tsimpli, I. & Sorace, A. (2006). Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. *Proceeding of the 30th Annual Boston University Conference on Language Development*, 653-664.
- Venditti, J. J., Maekawa K., & Beckman M. E. (2008). Prominence marking in the Japanese intonation system. In Miyagawa, S. & Saito M. (eds.), *The Oxford Handbook of Japanese Linguistics*, 456-512. Oxford: Oxford University Press.
- Vergnaud, J.-R. & Zubizarreta M. L. (2005). The representation of focus and its implications: Towards an alternative account of some intervention effects. In Corver, N. H., Huybregts R., Kleinhenz U. & Koster J. (eds.), *Organizing Grammar. Linguistic Studies in Honor of Henk van Riemsdijk*. Berlin/New York: Mouton de Gruyter.
- Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F 0 of vowels. *Journal of Phonetics*, 23, 349-366.
- Yamane, N., Yoshimura, N., & Fujimori, A. (2015). “Needs of explicit instructions of English prosody for Japanese EFL learners”. Paper presented at Phonology Forum 2015, Osaka University, August 21.
- Yoshimura, N., Fujimori, A., & Shirahata, T. (2015). Focus and prosody in second language acquisition. *Journal of International Relations and Comparative Culture*, 13, 21-36. University of Shizuoka.
- Zubizarreta, M. L. (1998). *Prosody, Focus, and Word Order*. Cambridge, MA: MIT Press.