



L1 Effects on L2 comprehension of focus-to-prosody mapping: A comparison between Cantonese and Dutch learners of English

Haoyan Ge¹, Aoju Chen², Virginia Yip¹

¹The Chinese University of Hong Kong, Hong Kong

²Utrecht University, The Netherlands

hyge@cuhk.edu.hk, aoju.chen@uu.nl, vcymatthews@cuhk.edu.hk

Abstract

This study investigates how L2 learners use prosody in the comprehension of focus in sentences with the focus particle *only* and how L1 modulates L2 comprehension, through a cross-linguistic comparison between Cantonese- and Dutch-learners of English. The realization of focus is language specific: prosody is the primary device to encode focus in Dutch and English whereas it is less important than other linguistic devices in Cantonese to mark focus. In a comprehension experiment, participants were presented with question-answer dialogues and were asked to judge whether the answer made sense for the question in a certain context. The results revealed significant differences between the L2 groups: Cantonese learners showed similar percentage of 'yes' judgments and reaction times (RTs) across the conditions regardless of prosody and focus position, whereas Dutch learners, like native English controls, showed a significantly lower percentage of 'yes' judgments and longer RTs for answers with inappropriate prosody than those with appropriate prosody. Our findings reveal L1 effects on L2 comprehension of the mapping between focus and prosody.

Index Terms: prosody, focus, L2 comprehension, Cantonese, Dutch

1. Introduction

1.1. Prosody and focus

Languages differ to a great extent in how they use linguistic devices to realize and interpret focus. In West Germanic languages like English and Dutch, the focus of a sentence is typically realized by assigning an accent to the focal word [1]. It is generally agreed that there is a correlation between prosodic prominence and focus in a sentence. According to [2], prosody is a necessary condition of focus and the focused element must contain the most prosodic prominence in a sentence. By contrast, the use of prosody to realize focus in Cantonese, a tonal language with six lexical tones, is confined to duration-related cues [3][4]. Cantonese also uses a large number of particles in different sentence positions and varies word order to achieve the same purpose [5][6].

Previous studies have shown that native speakers of English and Dutch rely on prosody to interpret focus in ambiguous sentences with the focus particle *only* [7][8][9]. Different prosodic prominence triggers different interpretations of sentences with the focus particle *only* and affects the truth-value of the sentences [2][10]. In [7], native speakers of English first heard a story and then a dative sentence with prosodic prominence either on the indirect object or direct object (e.g., *Bill only gave the book to SUE* vs. *Bill only gave the BOOK to*

Sue). They were asked to judge whether the *only*-sentence was a true description of the story. It was found that native speakers of English could successfully use prosodic cues alone to resolve ambiguities involving *only*. They associated prosodic prominence on *SUE* with the interpretation that Bill gave the book to Sue but nobody else whereas prosodic prominence on *BOOK* with the interpretation that Bill gave the book but nothing else to Sue. Unlike speakers of English, native speakers of Taiwan Mandarin have difficulty with the use of prosodic information in comprehending focus involving *zhi* 'only' [11], despite the fact that Taiwan Mandarin does utilize prosody to express focus [12]. More specifically, using a similar task and test sentences to [7], [11] examined how Taiwan Mandarin-speaking adults made use of prosodic cues to resolve ambiguity in dative constructions with *zhi* 'only' in Taiwan Mandarin, as in (1). It was found that the participants consistently associated *zhi* 'only' with direct object in their interpretation, regardless of the position of prosodic prominence. The findings suggest that Taiwan Mandarin adults were insensitive to prosody in resolving ambiguity in sentences with *zhi* 'only'.

(1) a. *Laoshi zhi song SHU gei Yuehan*

b. *Laoshi zhi song shu gei YUEHAN*

Teacher only gave the book to John

1.2. L1 effect on L2 comprehension

The role of L1 in L2 comprehension has been the subject of some debate. On the one hand, some researchers ascribe an influential role to L1 in L2 comprehension, suggesting that everything from phonology to syntax to lexicon is filtered through the L1 parser in some way [13][14]. On the other hand, other researchers minimize the role of L1 in real-time processing. They argue that L2 comprehension is not influenced by L1 but instead relies more on lexical and pragmatic strategies that may be universal in nature [15][16]. A recent study by [17] suggested that L1 experience moulded the perception of L2 sentence prosody and meaning. There is also the possibility that L2 learners manage to acquire native-like competence in certain aspects of L2 comprehension.

1.3. The present study

Given the cross-linguistic differences in the role of prosody in production and comprehension of focus, it is unclear whether advanced L2 learners whose L1 does not use prosody to encode focus to the same extent as L2 can successfully comprehend the mapping between focus and prosody in L2 and whether there is any influence from L1.

The present study compared two groups of L2 learners of English whose L1s are Cantonese and Dutch respectively. The comparison between these two L2 groups makes it possible to

examine the role of L1 in L2 comprehension and tease apart the potential L1 effects from the general L2 processing effects. If different performance is found between Cantonese- and Dutch-learners of English with Dutch learners performing like native English controls, it would provide evidence for the influence of L1 in the form of both negative and positive transfer. If similar performance is observed in the two L2 groups but different from native controls, it would suggest little L1 effect but general L2 processing effects. The research questions are as follows:

- I. To what extent is the relationship between focus and prosody reflected in L2 learners' comprehension of focus in sentences with *only* in English?
- II. How do the differences and similarities between L1 and L2 modulate L2 comprehension?

Answers to these two questions will contribute to the ongoing debate on the role of L1 influence in L2 processing and have implications for L2 comprehension in general.

2. Method

2.1. Participants

We recruited for the experiment 40 Cantonese learners of English (21 males) between the ages of 18 and 28 (Mean age=20), 35 Dutch learners of English (11 males) between the ages of 17 and 29 (Mean age=21), and 40 native speakers of English (19 males) ranged in age from 19 to 38 (Mean age=21). Among the native controls, 19 of them spoke American English and 21 of them British English. All the L2 participants were advanced learners of English based on their scores of IELTS or equivalent proficiency tests. On average, the Cantonese learners started learning English at a younger age (Mean age=3) than the Dutch learners (Mean age=11). None of the participants had self-reported deficits in vision or hearing.

2.2. The 'make-sense' experiment and predictions

A 'make-sense' experiment was conducted to address the research questions. On each trial, participants listened to a short story and were presented with a question-answer dialogue about the short story. They were then asked to judge whether the answer (sentence with *only*) made sense as a response to the question. The answer sentences were systematically varied in prosody such that either the verb or object is accented, leading to either contextually appropriate or inappropriate prosodic patterns, as illustrated in (2). Participants' judgments and RTs were measured. Similar paradigms had been used in previous studies of comprehension of the focus-to-prosody mapping in sentences without the focus particle *only* [18][19].

If L2 learners of English exhibit native-like comprehension of the focus-to-prosody mapping in sentences with *only* in English, we would expect them to judge answers with appropriate prosody more frequently as making sense than answers with inappropriate prosody and to respond faster in the cases of appropriate prosody than in the cases of inappropriate prosody.

2.3. Design and Materials

The short stories were composed to provide background information of the dialogues, introducing the agents, the actions and objects involved in the dialogues. For the experimental dialogues, there were two versions of each question and two versions of each answer. Following the Nuclear Stress Rule

[20], verb-focus is more marked relative to object-focus in SVO sentences because the most deeply embedded node in the syntactic structure receives prosodic prominence. Therefore, one version of the question set up an object-focus in the answer whereas the other version set up a verb-focus in the answer. Two variables were embedded in the answer sentences: prosody (appropriate vs. inappropriate) and focus position (object vs. verb). This gave rise to four experimental conditions, as illustrated in (2), where the target words carrying prosodic prominence are in bold letters and the subject noun always receives an accent to achieve metrical well-formedness at the prosodic phrasal level [21].

(2) *Story: The fox has some honey and an ice cream. She was going to lick and freeze both of them. Then she changed her mind.*

- a. *object-focus with appropriate prosody*

Question: I wonder what the fox is licking.

*Answer: The fox is only licking the **HONEY**.*

- b. *verb-focus with appropriate prosody*

Question: I wonder what the fox is doing with the honey.

*Answer: The fox is only **LICKING** the honey.*

- c. *object-focus with inappropriate prosody*

Question: I wonder what the fox is licking.

*Answer: The fox is only **LICKING** the honey.*

- d. *verb-focus with inappropriate prosody*

Question: I wonder what the fox is doing with the honey.

*Answer: The fox is only licking the **HONEY**.*

To add variation to the stimuli, two types of fillers were included. The answers in the fillers were incorrect in half of the time either because of semantic errors (e.g., referring to *licking* as *drinking* or *fox* as *bear*) or pronunciation errors (e.g., mispronouncing *fox* as *fax*). The prosody was appropriate in half of the fillers with error-free answers and inappropriate in the other half of the fillers with error-free answers. The same held for answers of the fillers that contain either semantic or pronunciation errors.

In total, there were 160 experimental dialogues and 160 fillers, which were distributed over the four conditions via a Latin Square design. Four lists of dialogues were created such that each dialogue appeared in every experimental condition but not in the same list. The four lists were then pseudo-randomized. Each participant was presented with only one list, which included 88 dialogues (4 experimental conditions × 10 experimental dialogues + 20 fillers with errors + 20 fillers without error + 8 practice dialogues).

The stimuli were recorded by a male native speaker of British English at 44.1kHz sampling frequency with 16 bits resolution in a recording booth. He was asked to produce the stimuli as naturally as possible with the appropriate prosody. To create stimuli with contextually inappropriate prosody, the answers to the questions on the object were combined with the questions on the verb and the answers to the questions on the verb were combined with the questions on the object. The intensity of the stimuli was normalized to 70dB.

2.3.1. Prosodic analysis

To ensure that the focus-marking prosodic prominence was placed in the right position, the answer sentences of the experimental dialogues were subjected to a phonetic analysis

on peak height, mean pitch and duration of the object nouns and verbs using Praat. Because the answers to the “inappropriate” conditions were identical to those of the “appropriate” conditions, only experimental stimuli in the “appropriate” conditions were measured.

A two-sample t-test revealed no significant difference between the object-focus and verb-focus conditions in terms of the overall duration of the answer sentence ($t(78)=0.692$, $p=0.49$). Significant differences were found between the two conditions regarding peak height and mean pitch. The peak height of the verb was significantly higher in the verb-focus condition than in the object-focus condition ($t(78)=-23.02$, $p<0.001$). The peak height of the object in the object-focus condition was significantly higher than that in the verb-focus condition ($t(78)=19.37$, $p<0.001$). The mean pitch of the verb was significantly lower in the object-focus condition than that in the verb-focus condition ($t(78)=-16.62$, $p<0.001$). The mean pitch of the object was much lower in the verb-focus condition than that in the object-focus condition ($t(78)=18.67$, $p<0.001$). Although no significant difference was found between the two conditions with respect to the mean duration of the verb ($t(78)=-0.31$, $p=0.76$), the object duration in the object-focus condition was remarkably longer than that in the verb-focus condition ($t(78)=2.84$, $p=0.006$).

2.4. Procedure

This study was carried out in accordance with research ethical laws at the Chinese University of Hong Kong (CUHK) and Utrecht University (UU) with informed consent from all participants. The participants filled in a language background questionnaire before proceeding to the experiment.

Each trial was set up in the E-Prime as follows: first, a cross appeared in the center of the screen for 1000ms. Then, a short story was presented, followed by a question. An answer was played 2500ms after the end of the question. The two options of ‘make-sense’ judgments “YES or NO” were displayed on the screen at the end of the answer. The participants were instructed to press the buttons on an RT box connected with E-Prime 2.0 to indicate their response. If the answer made sense to the question, they were supposed to press the “YES” button, otherwise the “NO” button. They were further instructed to press the buttons as quickly as possible, but not before the end of the answer sentence. The RTs were recorded at the end of each answer sentence until a button was pressed and the judgments were automatically recorded by E-prime 2.0.

Each testing session lasted about 20 minutes, starting with eight practice trials, aiming to familiarize the participants with the task. The participants were unaware of the purpose of the experiment and were paid HK\$30 or €5 for their participation.

3. Analysis & Results

Separate analyses were conducted on the judgments data and the RT data.

3.1. YES-NO judgments

The percentage of YES responses by the three groups of participants was shown in Figure 1. The Cantonese learners of English showed similar numbers of “YES” judgments across the four conditions, regardless of prosody and focus position (object-focus with appropriate prosody: $Mean=86.58\%$, $SD=0.34$; verb-focus with appropriate prosody: $Mean=85.86\%$, $SD=0.35$; object-focus with inappropriate prosody:

$Mean=85.22\%$, $SD=0.36$; verb-focus with inappropriate prosody: $Mean=84.15\%$, $SD=0.37$).

Unlike the Cantonese learners of English, the Dutch learners and the native controls showed a significantly higher frequency of acceptance for answers with appropriate prosody (Dutch: object-focus, $Mean=90.99\%$, $SD=0.29$; verb-focus, $Mean=90.19\%$, $SD=0.30$; English: object-focus, $Mean=80.95\%$, $SD=0.39$; verb-focus, $Mean=86.21\%$, $SD=0.35$) than answers with inappropriate prosody (Dutch: object-focus, $Mean=37.27\%$, $SD=0.48$; verb-focus, $Mean=40\%$, $SD=0.49$; English: object-focus, $Mean=59.32\%$, $SD=0.49$; verb-focus, $Mean=58.47\%$, $SD=0.49$).

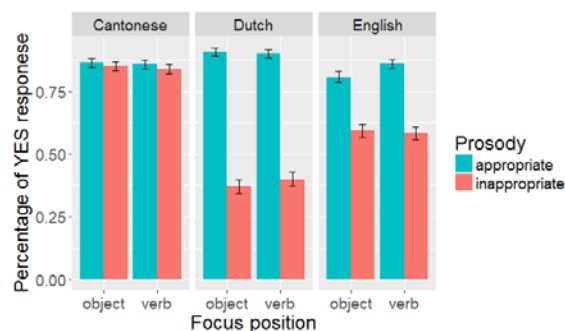


Figure 1: Mean percentage of YES response in the ‘make-sense’ experiment by the Cantonese and Dutch learners of English and the native speakers of English. The error bars indicate $\pm SE$.

To assess the above-mentioned observations, the participants’ “YES-NO” judgments were analyzed using a mixed-effects logistic regression in R with Participant as a random factor. The predictor variables included Group (L1 Cantonese vs. L1 Dutch vs. native English), Prosody (appropriate vs. inappropriate), Focus Position (object-focus vs. verb-focus) and their interaction. Results revealed a main effect of Group ($\chi^2(2)=488.69$, $p<0.001$), a main effect of Accent Placement ($\chi^2(1)=17.7$, $p<0.001$) and a significant Group \times Prosody interaction ($\chi^2(2)=234.4$, $p<0.001$). Post-hoc analyses showed that the group differences lay crucially in the participants’ response to inappropriate prosody. Specifically, the Dutch learners of English gave significantly fewer “YES” judgments to answers with inappropriate prosody than the Cantonese learners of English ($\beta=-4.05$, $SE=0.29$, $z=-13.77$, $p<0.001$). Similarly, the native speakers of English gave significantly fewer “YES” judgments to answers with inappropriate prosody than the Cantonese learners of English ($\beta=-1.51$, $SE=0.22$, $z=-6.91$, $p<0.001$).

3.2. Reaction times

Only RTs in the experimental trials where the answers were judged to make sense were included for further analysis, following [18] and [19]. RTs smaller than 100ms or above 4000ms were excluded. The remaining RTs were then log-transformed to reduce the non-normality of residuals. The mean log-transformed RTs of the groups of participants were shown in Figure 2.

The Cantonese learners of English showed similar RTs across the four conditions. That is, they were not slower in judging answers with inappropriate prosody (object-focus: $Mean=2.7$, $SD=0.3$; verb-focus: $Mean=2.67$, $SD=0.32$) than those with appropriate prosody (object-focus: $Mean=2.7$, $SD=0.31$; verb-focus: $Mean=2.69$, $SD=0.35$).

By contrast, the Dutch learners of English showed longer RTs for answers with inappropriate prosody (object-focus: $Mean=2.7$, $SD=0.34$; verb-focus: $Mean=2.7$, $SD=0.32$) than those with appropriate prosody regardless of the focus position (object-focus: $Mean=2.61$, $SD=0.31$; verb-focus: $Mean=2.64$, $SD=0.3$). Similar patterns were also observed in the native controls: they were slower in responding to answers with inappropriate prosody (object-focus: $Mean=2.7$, $SD=0.34$; verb-focus: $Mean=2.7$, $SD=0.32$) than those with appropriate prosody (object-focus: $Mean=2.61$, $SD=0.31$; verb-focus: $Mean=2.64$, $SD=0.3$).



Figure 2: Mean log-transformed RTs of the ‘make-sense’ experiment by the Cantonese and Dutch learners of English and the native speakers of English. The error bars indicate $\pm SE$.

A generalized linear mixed-effect model in the R package *lme4* was conducted on log-transformed RTs to access the above-mentioned observations. The fixed factors included Group (L1 Cantonese vs. L1 Dutch vs. native English), Prosody (appropriate vs. inappropriate), Focus Position (object vs. verb) and their interactions. The factor Participant was included in the model as a random factor. There was a main effect of Group ($\chi^2(2)=6.49$, $p=0.039$), a main effect of Prosody ($\chi^2(1)=16.16$, $p<0.001$) and a significant interaction effect between Group and Prosody ($\chi^2(2)=14.65$, $p<0.001$). Follow-up analyses revealed that the group differences were more evident in answers with inappropriate prosody. Specifically, the Dutch learners of English ($\beta=0.079$, $SE=0.02$, $t=3.23$, $p=0.0012$) and the native speakers of English ($\beta=0.075$, $SE=0.02$, $t=3.19$, $p=0.0014$) were significantly slower than the Cantonese learners of English in judging whether the answers with inappropriate prosody made sense for the questions.

4. Discussion

The results show that appropriate prosody triggered more “YES” judgements than inappropriate prosody in both Dutch learners of English and native speakers of English. Cantonese learners of English, however, were insensitive to inappropriate prosody, treated both appropriate and inappropriate prosody as indistinguishable and judging both similarly, regardless of focus position. Regarding the speed in deciding whether an answer made sense for the question, Dutch learners and native controls were faster in appropriate-prosody condition than in inappropriate-prosody condition, independent of focus position. The effect of prosody was, however, absent in Cantonese learners of English.

Taken together, our results indicate that prosody can affect how accurate and how fast Dutch learners of English and native

controls comprehended English sentences with *only* in different focus conditions independent of focus position, whereas it plays little role in Cantonese learners’ L2 comprehension. The differences between Cantonese and Dutch learners of English imply that L1 plays an important role in L2 learners’ comprehension of the focus-to-prosody interface in English. Given the similarities between Dutch and English in prosodic focus-marking, it is likely that the native-like comprehension of Dutch learners of English results from positive transfer from L1 to L2. In contrast, there is evidence that the non-native-like comprehension of Cantonese learners of English results from negative transfer from L1 to L2. Moreover, L2 proficiency is an unlikely explanation for the differences across groups. Cantonese learners started learning English at an earlier age and had longer time of exposure time. However, it was the Dutch group that matched controls more closely. This can be interpreted as further evidence that L1 effects are behind the differences between the two L2 groups.

5. Conclusion & Future Research

Our cross-linguistic study has examined the L2 comprehension of focus-to-prosody mapping in Cantonese and Dutch learners of English. The findings reveal native-like performance in Dutch learners of English but not in Cantonese learners of English. The differences between the two L2 groups provide evidence for L1 effects on L2 comprehension. These L1 effects are manifested both as positive transfer from Dutch to English, evidenced in the mastery of the focus-to-prosody interface in English, and negative transfer from Cantonese to English, resulting in difficulty in mapping prosody and focus in English.

So far most of the studies on L2 comprehension have investigated aspects of syntax, morphology or pragmatics. The present study has looked at the prosody-focus interface as a new domain of investigation and taken our understanding of how L1 influences L2 comprehension a step forward, by examining the complex interplay between prosody and focus.

Much work still remains to be done in this line of research. First, Dutch learners’ nativelike performance is based on measurements tapping into the end stage of a comprehension process. It is unknown if their underlying processes are native-like. L2 learners may reach the same accuracy in comprehension and at a similar speed but may have undergone a different processing path than native speakers. Further research is needed to reveal how listeners process the focus-to-prosody mapping as a sentence unfolds itself, using eye tracking or ERP methodologies. Second, one might wonder whether Cantonese learners’ non-native-like performance is due to their difficulties in correctly perceiving prosodic prominence in English in the first place. This is not unlikely considering that prosodic prominence is achieved via duration in Cantonese but via both pitch and duration in English. Perception studies on Cantonese learners’ perception of accentuation in English can shed light on this issue.

6. Acknowledgements

We are grateful to all the participants from CUHK and UU. We also thank Tracy Au, Julia Fan, Alice Zhu, Rachida Ganga and Alex Brouwer for their help in conducting the experiments. This study was supported by the Global Scholarship Program for Research Excellence from CUHK to the first author.

7. References

- [1] D. R. Ladd. *The structure of intonational meaning: Evidence from English*. Bloomington: Indiana University Press, 1980.
- [2] R. S. Jackendoff. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press, 1972.
- [3] Y. R. Chao. *Cantonese Primer*. Cambridge, MA: Harvard University Press, 1947.
- [4] W. L. Wu and Y. Xu, "Prosodic Focus in Hong Kong Cantonese without Post-focus Compression", *Proc. Speech Prosody 2010*, pp. 1–4, 2010.
- [5] S. Matthews and V. Yip, *Cantonese: A Comprehensive Grammar*. Routledge, 2011.
- [6] K. Cheung, "The phonology of present-day Cantonese", Ph.D. dissertation, University College, London, 1986.
- [7] A. Gualmini, S. Maciukaite, and S. Crain, "Children's Insensitivity to Contrastive Stress in Sentences with ONLY", *Proc. 26th Annu. PCL*, vol. 8, no. 1, pp. 1–14, 2002.
- [8] S. Gennari, L. Meroni, and S. Crain, "Rapid relief of stress in dealing with ambiguity," in *Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions*, J. C. Trueswell and M. K. Tanenhaus, Eds. Cambridge, MA: MIT Press, 2005, pp. 245–259.
- [9] I. Mulders and K. Szendrői, "Early Association of Prosodic Focus with alleen 'only': Evidence from Eye Movements in the Visual-World Paradigm," *Front. Psychol.*, vol. 7, no. March, pp. 1–19, 2016.
- [10] M. Rooth, "A Theory of Focus Interpretation," *Nat. Lang. Semant.*, vol. 1, no. 1, pp. 75–116, 1992.
- [11] S. I. Shyu, "Focus interpretation of zhi 'only' associated arguments in Mandarin triadic constructions," *Linguistics*, vol. 48, no. 3, pp. 671–716, 2010.
- [12] S. Chen, B. Wang, and Y. Xu Y, "Closely related languages, different ways of realizing focus", *Proc. Interspeech*, pp. 1007–1010, 2009
- [13] C. Frenck-Mestre, "An On-Line Look at Sentence Processing in the Second Language," in *Bilingual Sentence Processing*, R. Heredia and J. Altarriba, Eds. 2002, pp. 217–236.
- [14] A. Juffs, "The influence of first language on the processing of wh-movement in English as a second language," *Second Lang. Res.*, vol. 21, no. 2, pp. 121–151, 2005.
- [15] H. Clahsen and C. Felser, "How native-like is non-native language processing?," *Trends Cogn. Sci.*, vol. 10, no. 12, pp. 564–570, 2006.
- [16] T. Marinis, L. Roberts, C. Felser, and H. Clahsen, "Gaps in Second Language Sentence Processing," *Stud. Second Lang. Acquis.*, vol. 27, no. 1, pp. 53–78, 2005.
- [17] M. Ortega-Llebaria and L. Colantoni, "L2 English Intonation," *Stud. Second Lang. Acquis.*, vol. 36, no. 2, pp. 331–353, 2014.
- [18] S. Birch and C. Clifton, "Focus, Accent, and Argument Structure: Effects on Language Comprehension," *Lang. Speech*, vol. 38, no. 4, pp. 365–391, 1995.
- [19] A. Chen, "Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping?," *Lingua*, vol. 120, no. 8, pp. 1926–1939, 2010.
- [20] M. L. Zubizarreta, "Nuclear stress and information structure," in *The Oxford Handbook of Information Structure*, Oxford: Oxford University Press, 2016, pp. 165–184.
- [21] S. Calhoun, "The centrality of metrical structure in signaling information structure: A probabilistic perspective," *Language*, vol. 86, no. 1, pp. 1–42, 2010.