

ORIGINAL ARTICLE

Processing focus in native and non-native speakers of English: an eye-tracking study in the visual world paradigm

Haoyan Ge^{1*}, Iris Mulders², Xin Kang³, Aoju Chen² and Virginia Yip⁴

¹School of Education and Languages, The Open University of Hong Kong, Hong Kong SAR, China, ²Department of Languages, Literature and Communication & Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, The Netherlands, ³Department of Linguistics and Modern Languages & Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong SAR, China and ⁴Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong SAR, China
*Corresponding author. Email: hge@ouhk.edu.hk

(Received 27 May 2020; revised 25 April 2021; accepted 12 May 2021; first published online 07 June 2021)

Abstract

This “visual-world” eye-tracking study investigated the processing of focus in English sentences with preverbal *only* by L2 learners whose L1 was either Cantonese or Dutch, compared to native speakers of English. Participants heard *only*-sentences with prosodic prominence either on the object or on the verb and viewed pictures containing an object-focus alternative and a verb-focus alternative. We found that both L2 groups showed delayed eye movements to the alternative of focus, which was different from the native speakers of English. Moreover, Dutch learners of English were even slower than Cantonese learners of English in directing fixations to the alternative of focus. We interpreted the delayed fixation patterns in both L2 groups as evidence of difficulties in integrating multiple interfaces in real time. Furthermore, the similarity between English and Dutch in the use of prosody to mark focus hindered Dutch learners’ L2 processing of focus, whereas the difference between English and Cantonese in the realization of focus facilitated Cantonese learners’ processing of focus in English.

Keywords: L2 processing; focus; visual world eye tracking; Cantonese learners of English; Dutch learners of English

Introduction

The question of whether sentence processing in the second language (L2) is fundamentally similar or different from sentence processing in the native language (L1) has been widely debated (e.g., Clahsen & Felser, 2006; Sorace, 2011). Within the domain of information structure, whether L2 learners can comprehend focus in the same way as L1 speakers has gained considerable attention in L2 acquisition (e.g., Akker & Cutler, 2003; Liu & Lee, 2021; Ortega-Llebaria & Colantoni, 2014; Reichle & Birdsong, 2014; Slabakova et al., 2012; Zubizarreta & Nava, 2011).

© The Author(s), 2021. Published by Cambridge University Press.

Information structure, also known as information packaging, refers to the interface between the structure and meaning of utterances, and is constrained by the context and the interlocutors' mental representations of information (Chafe, 1976; Lambrecht, 1994). Focus is a key concept of informational structure. It commonly refers to new or contrastive information about a topic in a sentence. The interpretation of focus involves multiple levels of knowledge including syntax, semantics, prosody, and pragmatics (Lambrecht, 1994). In English, the focus is typically realized by assigning a nuclear pitch accent (indicated by capital letters) to the focal element(s) (indicated by subscript F), as in (1b). Nuclear pitch accents are manifested in expanded pitch range, increased intensity, and longer duration (Gussenhoven, 1983; Selkirk, 1995). The focus-to-prosody mapping plays an important role in determining the felicity of the utterance in a discourse. Accentuation on the object *APPLE* (1b) is felicitous to the question, whereas accentuation on the verb *ATE* (1c) is not.

- (1) a. Question: Did John eat the pear?
 b. Answer: No, he ate the [APPLE]_F.
 c. Answer: No, he ATE the [apple]_F.

However, the presence of a nuclear pitch accent does not contribute to the meaning of focus directly, as it does not change the truth conditions of the sentence. In sentences with the focus particle *only*, prosodic information is not only relevant for pragmatic felicity of focus, but directly contributes to the meaning of utterances. In English preverbal *only*-sentences, different positions of prosodic prominence trigger different interpretations of focus and affect the truth conditions of the sentences (Jackendoff, 1972; Rooth, 1992). For example, to correctly comprehend a sentence like “*John only ate the apple*”, listeners need to identify the scope of the focus particle *only*, which can associate with any focused element found in the following phrase, and integrate prosodic information for successful semantic parsing. If *apple* carries prosodic prominence, listeners will understand that John ate nothing else but the apple. If *ate* receives prosodic prominence, listeners will know that John did nothing else to the apple other than eating it.

The investigation of focus in *only*-sentences is interesting in the field of L2 processing for two reasons. First, the focus processing in *only*-sentences involves multiple levels of linguistic knowledge, including syntax, prosody, semantics, and pragmatics, as stated above. Over the past two decades, there has been considerable investigation into interface structures, with a particular emphasis on whether L2 learners experience difficulties in integrating different levels of linguistic knowledge (Sorace, 2011; Sorace & Filiaci, 2006; Sorace & Serratrice, 2009; White, 2011). According to the Interface Hypothesis (Sorace, 2011), internal interfaces involving components of the language system (e.g., syntax–semantics) are less likely to be problematic, whereas the external interfaces involving a cognitive system not specific to language (e.g., syntax–pragmatics) are the prime locus of protracted delays and difficulty in L2 acquisition, regardless of the L1–L2 differences (Hopp, 2009; Sorace & Serratrice, 2009; Sorace, 2011). However, it is unclear how the Interface Hypothesis could account for the L2 processing of multiple interfaces that are sensitive to both internal and external interfaces, such as the processing of focus.

Second, the realization of focus is language-specific. Different languages use different linguistic devices (e.g., morphosyntax and prosody) and to different extents to mark focus (Lambrecht, 1994). It is not clear how the Interface Hypothesis could account for the differences between L1 and L2. In the field of L2 prosody, the Prosodic-Learning Interference Hypothesis was recently proposed by Tremblay et al. (2016) to explain the similarities and differences between L1 and L2 prosodic systems. They proposed that the L2 learning of prosodic cues (e.g., signaling word boundaries) is more difficult when the L1 and L2 use similar prosodic cues than when the L1 and L2 differ in the use of prosodic cues. They compared Korean learners' and English learners' use of fundamental frequency (F0) rise as a cue to word-final boundaries in French. Korean is similar to French in the use of F0 to mark the edge of a phrase, whereas English is different from French and does not mark the phrasal boundary with prosodic prominence. Their results showed that Korean learners of French had greater difficulty using F0 rise as a cue to segment speech in French than English learners of French.

Against this background, the present eye-tracking study in the visual world paradigm investigated the processing of focus in sentences with preverbal *only* by advanced Cantonese learners of English and Dutch learners of English. It aimed to obtain a clear understanding of the factors underlying similarities and/or differences in the processing of focus in *only*-sentences between L1 and L2.

The paper is organized as follows. We first outline the properties of focus in *only*-sentences in English, Dutch, and Cantonese (see Section "Focus in sentences with 'only' in English, Dutch and Cantonese"), and review previous research on the processing of focus by L1 and L2 speakers (see Sections "L2 processing of focus" and "L1 processing of focus in *only*-sentences in visual world paradigm"). Then we present the research questions and hypotheses (Section "The current study"). We describe the eye-tracking experiments in Section "Method", and report the results from L1 and L2 speakers in Section "Results". Finally, we discuss how our findings can shed light on L2 processing of focus (Section "Discussion"), and offer a brief conclusion in Section "Conclusion".

Focus in sentences with "only" in English, Dutch, and Cantonese

Both English and Dutch use *only* (or the Dutch equivalent *alleen*) to construct the focus meaning. Semantically, *only* presupposes the existence of an alternative set in the discourse, and asserts that the entity in focus has some characteristics that other alternatives lack (Jackendoff, 1972; Rooth, 1992; Zimmermann & Onea, 2011). Prosodically, different positions of accentuation generate different alternative interpretations and affect the truth conditions of the sentences with *only* (Jackendoff, 1972; Mulders & Szendrői, 2016; Rooth, 1992). As illustrated in (2), when accentuation is placed on *BUCKET*, it not only triggers an object-focus reading¹ that John is carrying nothing else but the bucket, but also a set of objects that are not carried by John. When *CARRYING* is accented, a verb-focus reading and a set of other actions are triggered obligatorily: John is doing nothing to the bucket but carrying it. In a situation where John is carrying a bucket and a suitcase, (2a) is false, while (2b) is true. In another situation where John is carrying and washing the bucket, (2a) is true, but (2b) is false.

- (2) a. John is **only** carrying the [BUCKET]_F. (object focus)
 Alternative set = {x: John is carrying x}
 b. John is **only** [CARRYING]_F the bucket. (verb focus)
 Alternative set = {x: John is x the bucket}

English and Dutch are very similar with respect to *only*-sentences. In both languages, *only* (or the Dutch equivalent *alleen*) can be adjoined to maximal projections (XPs), and is associated with the focused element in the following phrase. There are independent differences between English and Dutch syntax though. In English, verbs are preferably placed adjacent to the direct object, leading to the more frequent placement of *only* before the verb (Bouma *et al.*, 2007). Unlike English, Dutch has OV word order with the verb moving to second position in main clauses, and it has a more flexible word order in the middle of sentences. Dutch has a general preference for *alleen* “only” to directly precede the focused element (see example (3a)), as reported in corpus-based research (Foolen *et al.*, 2009). However, in (3b), the V2 movement of the focused verb linearly removes the focused element away from *alleen* in the surface order. In that sense, although both English and Dutch use prosodic prominence to realize focus, they are not identical in how they use prosodic cues to mark focus in *only*-sentences in terms of the focus position.

- (3) a. John draagt **alleen** De [EMMER]_F
 John carry only The bucket
 “John is only carrying the BUCKET.”
 b. John [DRAAGT]_F de Emmer **alleen**
 John carry the Bucket only
 “John is only CARRYING the bucket.”

While Dutch is very similar to English in the use of prosody to mark focus, Cantonese differs from English substantially in this respect. The use of prosody in Cantonese is highly constrained. Specifically, there is no clear evidence for on-focus pitch expansion in Cantonese (Man, 2002; Wu & Xu, 2010). Instead, longer duration and higher intensity are manifested in Cantonese-focused elements (Gu & Lee, 2007; Wu & Xu, 2010). Apart from this prosodic device, Cantonese uses different focus particles, including *zing6hai6* “only”², *zaa3* “only”, *ze1* “only”, *sin3* “only then”, *zau6* “only”, and *dak1* “only”, in different sentence positions to convey the focus meaning of *only* (Fung, 2000; Lee, 2019; Matthews & Yip, 2011). A full discussion of all the focus particles in Cantonese is beyond the scope of this study. In what follows, we briefly discuss three focus particles in Cantonese, *zing6hai6*, *zaa3*, and *dak1*, which have drawn the most theoretical attention. Semantically, these focus particles have similar functions to English *only*: specifying the focused element, introducing an alternative set, and contributing to the truth conditions of the sentence.

Similarly to English *only*, the preverbal *zing6hai6* may associate with the verb (4a), object (4b), or entire VP (4c), based on the contextual and prosodic information (*i.e.*, primarily variation in duration and intensity).

- (4) Keoi5 **zing6hai6** ling1zyu6 go3 tung2
 She only carrying CL³ bucket
 a. “She is only CARRYING the bucket.” (verb focus)
 b. “She is only carrying the BUCKET.” (object focus)
 c. “She is only CARRYING THE BUCKET.” (VP focus)

Unlike *zing6hai6*, the sentence-final particle *zaa3* can associate with any leftward constituent, the object (5a), verb (5b), or even the subject (5c). *Zaa3* also contributes a sense of exclusion to sentences⁴ (Fung, 2000; Lee, 2019). Moreover, *zing6hai6* and *zaa3* could be used together in one sentence to encode the meaning of *only* (6), giving rise to an object-focus (6a) or verb-focus (6b) reading.

- (5) Keoi5 ling1zyu6 go3 tung2 **zaa3**
 She carrying CL bucket only
 a. “She is only carrying the BUCKET.” (object focus)
 b. “She is only CARRYING the bucket.” (verb focus)
 c. “Only SHE is carrying the bucket.” (subject focus)
- (6) Keoi5 **zing6hai6** ling1zyu6 go3 tung2 **zaa3**
 She only carrying CL bucket only
 a. “She is only carrying the BUCKET.” (object focus)
 b. “She is only CARRYING the bucket.” (verb focus)

Apart from *zing6hai6* and *zaa3*, the focus meaning of *only* could also be conveyed by *dak1* in Cantonese⁵ (Tang, 2002). *Dak1* can be associated with the subject (7a), or appear in a postverbal position and associate with the object (7b), implying that “she ate only three apples (and no more than three)”.

- (7) a. **Dak1** [keoi5]_F ling1zyu6 go3 tung2
 Only she carrying CL bucket
 “Only SHE is carrying the bucket.”
- b. Keoi5 sik6 **dak1** [saam1 go3 ping4gwo2]_F
 She eat only three CL apple
 “She ate only THREE APPLES.”

In addition, Chinese speakers, including Cantonese speakers, prefer to have an overt (or contextually implied) negation conjunct, as in (8), to realize focus meaning (Shyu, 2010).

- (8) Keoi5 zing6hai6 ling1zyu6 go3 [tung2]_F m4hai6 ling1zyu6 go3 doi2
 She only carrying CL bucket not carrying CL bag
 “She is only carrying the bucket, not carrying the bag.”

The above examples demonstrate a rich repertoire of focus particles in Cantonese, which makes the use of prosody optional to encode focus meaning (Lee, 2019). In contrast to English that rests on prosody to mark focus, Cantonese relies heavily on focus particles for the same purpose and demonstrates a strong feature of the

syntax–discourse interface in focus structuring (Lee, 2019; Xu, 2004). Thus, English and Cantonese differ in both the linguistic devices they use to realize focus and the extent to which the same linguistic device is used.

L2 processing of focus

Previous studies mainly investigated L2 processing of focus in sentences without *only* (e.g., Akker & Cutler, 2003; Ortega-Llebaria & Colantoni, 2014; Reichle & Birdsong, 2014; Slabakova *et al.*, 2012). The L2 processing of focus in *only*-sentences has not been systematically examined.

In an event-related potentials (ERPs) study, Reichle and Birdsong (2014) asked English learners with high- and low-proficiency of French to read *wh*-questions, followed by responses instantiating focus marked by *c'est . . . que* “it is . . . that” cleft construction either in appropriate or inappropriate contexts. They found that high-proficiency L2 learners showed similar patterns compared to L1 speakers of French, suggesting the possibility of native-like processing of syntactically encoded focus in L2.

Akker and Cutler (2003) examined how Dutch learners of English use prosodic information to comprehend focus. In their experiments, participants first heard a question that elicited focus on different elements (e.g., *Which bones were found by the archaeologist?* OR *Which archaeologist found the bones?*), then heard an answer involving the target phoneme (e.g., [d] in the bearing word *dinosaur*), which was either with or without prosodic prominence (e.g., *The bones of the DINOSAUR were found by the Cuban archaeologist* OR *The bones of the dinosaur were found by the CUBAN archaeologist*). Participants were asked to detect the target phoneme as quickly as possible. Native speakers of English were faster at detecting the target phoneme when the word bearing the target phoneme carried prosodic prominence or was focused and the effect of prosody and focus interacted (i.e., the effect of prosodic prominence was smaller for the focused words than for the non-focused words). The interaction between prosody and focus was, however, absent in Dutch learners of English. Akker and Cutler attributed Dutch learners’ non-native-like performance to reduced efficiency in integrating prosody into focus in L2.

However, Ortega-Llebaria and Colantoni (2014) suggested that native-like L2 processing of prosodic focus was possible when L2 learners’ L1 used similar strategies to L2 to encode focus. They compared L2 learners of English whose L1 was either Spanish or Mandarin. While Spanish primarily uses word order to express focus, Mandarin uses prosody to encode focus by expanding the pitch range and duration of the word (Liu, 2009; Wang & Xu, 2011), which is more similar to English at the acoustic level (Gussenhoven, 2006; Ladd, 2008; Xu & Xu, 2005). In their comprehension task, participants were required to select one of the three possible answers with prosodic prominence in different positions (e.g., (a) *TOBY fell out of the tree*; (b) *Toby FELL OUT of the tree*; (c) *Toby fell out of the TREE*.) to best answer the question (e.g., *Did Bobby fall out of the tree?*). Mandarin learners were observed to pattern with native controls, whereas

Spanish learners were significantly less accurate, suggesting the role of L1 in the comprehension of focus. Taken together, these previous studies mainly looked at whether syntactic or prosodic information would facilitate the processing of focus in contexts, where the use of syntactic or prosodic cues would not affect the propositional meaning of sentences.

Complementing these studies, Ge et al. (2021) investigated how L2 learners of English whose L1 was either Cantonese or Dutch-processed English sentences with *only* with a focus on either the verb or the object (e.g., *The fox is only LICKING the honey* vs. *The fox is only licking the HONEY*) in a “make-sense” task (i.e., judging whether a sentence was a sensible response in a certain context). Their results indicated that placement of accentuation affected how quickly and how accurately L1 English speakers and Dutch learners of English comprehended the *only*-sentences, whereas it hardly played a role in Cantonese learners’ comprehension. However, their study only looked at the focus-to-prosody mapping without investigating the use of prosody to disambiguate focus. Moreover, their findings are based on measurements that tap into the end stage of L2 comprehension process. It remains to be investigated whether L2 learners can process focus in *only*-sentences in a native-like way in real time.

L1 processing of focus in *only*-sentences in visual world paradigm

To the best of our knowledge, two eye-tracking studies in the visual world paradigm have investigated the L1 online processing of focus in *only*-sentences. In these two studies, the prosodic information directly contributed to the semantics of the utterances: one cannot simply compute the full meaning of the utterances without knowing the position of prosodic focus.

Gennari et al. (2004) measured overall fixations to an entity during the course of an entire utterance, using a visual setup involving three characters, for example, a boy, a man, and a woman. In the picture, the boy had a glass of milk, the man had a cup of coffee and a glass of milk, and the woman was holding a tray. Participants heard utterances in two conditions, either with or without accentuation on the direct object, like *milk* in (9). They were asked to judge whether the utterance was a true description of the picture or not.

- (9) a. The mother only gave some milk to the boy. (Expected response: FALSE)
 b. The mother only gave some MILK to the boy. (Expected response: TRUE)
 (Gennari et al., 2004: 246)

Gennari et al. hypothesized that L1 processing of focus would be fast in *only*-sentences and that L1 English speakers could immediately use prosodic cues to decide which object carried the focus and, therefore, which set of alternatives should be invoked for the interpretation. Gennari et al. found that the “boy’s milk” drew a significantly higher proportion of looks when “milk” was accented (9b), relative to when it was not (9a). However, the overall proportion of looks to a particular entity was not time-locked to the particular time window in which accentuation appeared. Therefore, Gennari et al.’s results cannot be interpreted as evidence for the fast processing of prosodic focus.

Circumventing the methodological limitation in Gennari *et al.* (2004), Mulders & Szendrői (2016) examined the effects of accentuation on the time course of fixation patterns in the processing of Dutch sentences with the focus particle *alleen* “only”. The experimental auditory stimuli varied in the position of accentuation, either on the direct object or the indirect object (e.g., *Ik heb alleen SELDERIJ aan de brandweerman gegeven* “I only gave CELERY to the fireman” vs. *Ik heb alleen selderij aan de BRANDWEERMAN gegeven* “I only gave celery to the FIREMAN”). The participants also viewed co-presented visual stimuli containing the alternatives of the direct object focus or indirect object focus. They found that Dutch speakers’ fixations started to diverge across the conditions upon hearing the indirect object *brandweerman* “fireman”. They also found evidence for anticipatory eye movements slightly before the indirect object. Their findings provided evidence for the fast integration of prosodic and semantic information on online processing of sentences with *alleen* “only” in L1 Dutch speakers.

In sum, the visual world paradigm can serve as an effective method to investigate when prosodic information is integrated with the processing of *only*-sentences. However, it is still far from clear how L2 learners process focus in sentences with *only*, and whether there is any difference between L1 and L2 processing in this respect.

The current study

The current study aimed to attain a clear understanding of the factors underlying similarities and/or differences in the processing of focus in *only*-sentences between L1 and L2. To this end, we investigated the processing of focus in English sentences with preverbal *only* (e.g., *The dinosaur is only carrying the BUCKET, not carrying the suitcase* vs. *The dinosaur is only CARRYING the bucket, not throwing the bucket*) by L1 English speakers and L2 learners whose L1 was either Cantonese or Dutch. We raise the following two research questions:

- I. How do L1 English speakers use prosodic information to comprehend focus in *only*-sentences in real time?
- II. Do Cantonese learners of English and Dutch learners of English process focus in *only*-sentences in the same way as L1 English speakers?

Regarding the first research question, we hypothesized that L1 English speakers would use prosodic information immediately to interpret focus in *only*-sentences, based on the previous findings on L1 processing of focus in native Dutch speakers (Mulders & Szendrői, 2016). To be more specific, L1 English speakers would show increased fixations to the visual display containing the focus alternative when hearing words with prosodic prominence.

With respect to the second research question, we formulated two opposing hypotheses for the L2 learners, drawing reference to theoretical perspectives of the Interface Hypothesis (Sorace, 2011) and the Prosodic-Learning Interference Hypothesis (Tremblay *et al.*, 2016). According to the Interface Hypothesis, both groups of L2 learners would perform differently from L1 English speakers, regardless of the L1–L2 pairs, as the processing of focus in sentences with *only* involves

multiple levels of linguistic knowledge. Specifically, both groups of L2 learners would show delayed fixations or not fixate on the focus alternative when hearing accented words, compared to L1 English speakers.

In contrast, according to the Prosodic-Learning Interference Hypothesis, Cantonese learners of English would show increased fixations to the focus alternatives and at a similar rate, relative to L1 English speakers, because Cantonese differs greatly from English in the use of prosody to mark focus. Given the similarity between Dutch and English in the use of prosodic cues to encode focus, we would expect this similarity to hinder Dutch learners' processing of focus. Thus, Dutch learners would show delayed fixations or no increased fixations to the visual displays that involved focus alternatives, compared to L1 English speakers.

Method

Participants

Forty native English speakers, 40 Cantonese learners of English, and 35 Dutch learners of English participated in this study. None of them reported deficits in vision or hearing. The participants were unaware of the purpose of the experiment and were paid 5 Euros or equivalent for their participation. This study was conducted in accordance with research ethical procedures at the universities where the experiments took place with informed consent from all participants. The participants filled out a language background questionnaire. The L1 English speakers were exchange students at a research university in Hong Kong. They had very limited or no proficiency in Cantonese, Mandarin, or other varieties of Chinese at the time of testing. The Cantonese learners of English were recruited from undergraduate students at the same university. They were not fluent in Mandarin according to their self-reports⁶. The Dutch learners of English were recruited from undergraduate students from a research university in the Netherlands. The background information of the three groups of participants is summarized in Table 1.

To examine whether there was any difference between the two L2 groups in terms of the English proficiency, two-sample *t* tests were conducted on the scores obtained from the language questionnaires. The Cantonese learners started learning English at a significantly younger age and had learned English for a substantially longer time than the Dutch learners, whereas the Dutch learners rated themselves higher than the Cantonese learners on overall English proficiency (speaking and listening). Crucially, there was no difference between the two L2 groups regarding their IELTS scores or equivalents, $t(72) = 1.45$, $p = .15$. Thus, the two L2 groups were matched for proficiency level in English.

Task and materials

This eye-tracking study adopted the "look and listen" task in which participants heard auditory stimuli and looked at co-present pictures. The participants were not asked to give any behavioral responses such as pressing keys or clicking a mouse. The "look and listen" version of the visual world paradigm has been widely used in previous studies (e.g., Altmann & Kamide, 1999, 2007; Kang et al., 2020;

Table 1. Language background of participants (*SD* in parentheses)

	L1 English	Cantonese learners	Dutch learners
Number	40	40	35
Male: Female	19:21	20:20	11:24
Age range	19–38	18–28	17–29
Mean age	21.8 (4.72)	20.5 (2.33)	21.3 (2.39)
Mean age starting English	N/A	3.15 (1.68)	7.58 (3.51)
Years of learning English	N/A	17.6 (2.84)	12.6 (4.52)
Mean IELTS score or equivalent ^a	N/A	7.78 (0.34)	7.65 (0.42)
Self-evaluation of English proficiency ^b	Native	Advanced	Advanced
Overall	6	4.17 (0.60)	4.55 (0.49)
Speaking	6	4.25 (0.44)	4.76 (0.43)
Listening	6	4.37 (0.52)	4.97 (0.17)

^aThe HKDSE English Language Examination scores for the Cantonese learners and the VWO scores for the Dutch learners were converted to IELTS scores, based on the standards between the IELTS and HKDSE English Language Examination conducted by Hong Kong Examinations and Assessment Authority (HKEAA) (<http://www.hkeaa.edu.hk/en/recognition/benchmarking/hkdse/ielts/>) and comparison between VWO (English) and CERF (<https://www.cambridgeenglish.org/scale-score-converter/>).

^bOn a 1–6 scale: 1 = almost no knowledge/fluency/understanding; 2 = limited knowledge/fluency/understanding; 3 = some knowledge/fluency/understanding; 4 = good knowledge/fluency/understanding; 5 = excellent knowledge/fluency/understanding; 6 = native.

Salverda *et al.*, 2011). In a task-oriented visual world paradigm, participants are instructed to move or point at the target entity. Their increased fixations on the target visual display may be due to the task demands instead of the natural process of language comprehension (e.g., Altmann and Kamide 2004; Brown-Schmidt & Tanenhaus, 2008; Eberhard *et al.*, 1995; Salverda *et al.*, 2011). The “look and listen” version can avoid this kind of interference from task demands. Moreover, participants in the “look and listen” paradigm do not need to use conscious response strategies. Thus, the “look and listen” paradigm examines online language processing without interruption and can provide an implicit record of cognitive processes as auditory stimuli unfold over time.

Forty auditory stimuli were constructed in two conditions (i.e., object-focus vs. verb-focus), as in (10). In each stimulus, the story established a context that included the alternatives of object-focus and verb-focus, and thus elicited the use of the focus particle *only* in the target sentence (10a and 10b).

- (10) Story: *The dinosaur has a bucket and a suitcase. He was going to carry them and throw them. Then he changed his mind.*
- a. The dinosaur is only carrying the **BUCKET**, not carrying the suitcase. (Object-focus; alternatives: carrying other things, in this context, carrying the suitcase).
 - b. The dinosaur is only **CARRYING** the bucket, not throwing the bucket. (Verb-focus; alternatives: performing other actions on the bucket, in this context, throwing the bucket).

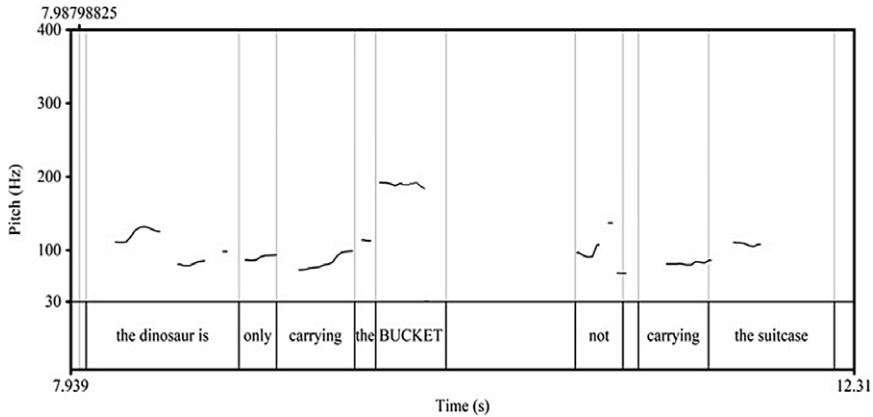


Figure 1. Pitch Track for Example of the Object-Focus Condition Utterance.

To add variation to the stimuli, 48 fillers were constructed. To match the experimental items, each filler involved a similar story and a *not*-fragment. Unlike the experimental items, the fillers did not include the focus particle *only*, as in (11).

- (11) Introduction: *The dog has a broom, a paper, and a tomato. He was going to wash the tomato. Then he changed his mind.*
- The dog is squeezing the tomato, not washing the tomato.
 - The dog is getting the tomato, not washing the tomato.

The experimental trials and fillers were recorded by a male native speaker of British English at 44.1 k Hz sampling frequency with 16-bit resolution in a sound-proof booth. He was instructed to produce the auditory stimuli with the appropriate prosody. Each stimulus was scaled to 70 dB SPL in mean intensity using Praat (Version 6.0.39; Boersma & Weenink, 2018). To ensure that prosody was placed either on the object or the verb, acoustic measurements were conducted in Praat. Examples of representative contours in the two experimental conditions are shown in Figures 1 and 2.

The verb was significantly longer in the verb-focus condition ($Mean = 463$ ms, $SD = 38.73$) than in the object-focus condition ($Mean = 401$ ms, $SD = 38.66$) ($t(38) = -5.07$, $p < .001$). The object was significantly longer in the object-focus condition ($Mean = 453$ ms, $SD = 71.75$) than in the verb-focus condition ($Mean = 388$ ms, $SD = 72.06$) ($t(38) = 2.89$, $p = .0063$).

Apart from the experimental trials and fillers, a practice session was constructed, including two items in two experimental conditions and two items in the filler conditions. A counterbalanced experimental design was used to distribute 40 experimental sentences, 48 fillers, and 8 practice items, resulting in 2 lists. In total, each participant was presented with 48 items (4 practice items + 2 experiment conditions \times 10 experiment items + 24 fillers). All the stimuli were cross-checked by two native speakers of English (one American English-speaking female and one

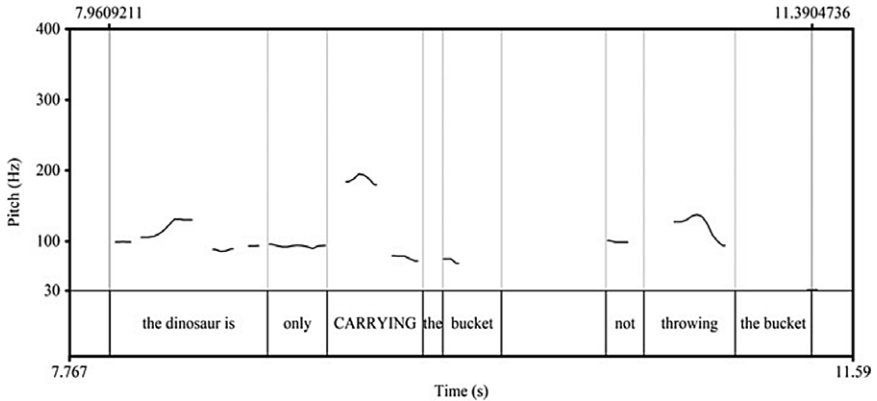


Figure 2. Pitch Track for Example of the Verb-Focus Condition Utterance.

British English-speaking male) to make sure the experimental sentences were natural. Pseudo-randomized orders were created for each list, in which two experimental items never directly followed each other in any list. That is, an experimental item from the verb-focus condition never followed an experimental item from the object-focus condition or vice versa. No more than two items from the same condition occurred successively.

The visual scenes were designed using a 2×2 grid design. Each image contained four scenes, depicting the same cartoon character performing different actions – one scene as the target, two as competitors, and one as the distractor. To control for the potential confounds caused by the spatial locations of four scenes, we varied their positions across the trials.

The experimental auditory stimuli in (10) correspond to the visual display in Figure 3. The target scene was about the cartoon character performing the target action (e.g., the dinosaur carrying the bucket). The distractor depicted the cartoon performing an irrelevant action (e.g., the dinosaur throwing the suitcase). One competitor picture involved the object-focus alternative (e.g., the dinosaur carrying the suitcase) and the other competitor involved the verb-focus alternative (e.g., the dinosaur throwing the bucket).

For each stimulus, the visual display and auditory stimuli were presented to the participants simultaneously, followed by a silence for 1,000 ms. There was no preview time because the story was long enough for the participants to explore the visual scene. The silence period in the end allowed the analysis of eye movements even after the completion of the auditory stimuli.

Procedure

The participants were tested individually in an eye-tracking laboratory at two testing sites: the L1 English speakers and Cantonese learners of English in Hong Kong and the Dutch learners of English in the Netherlands. One experimenter monitored each participant's eye movements from a screen outside the eye-tracking cabin throughout the task to make sure the participants were not looking away. All the

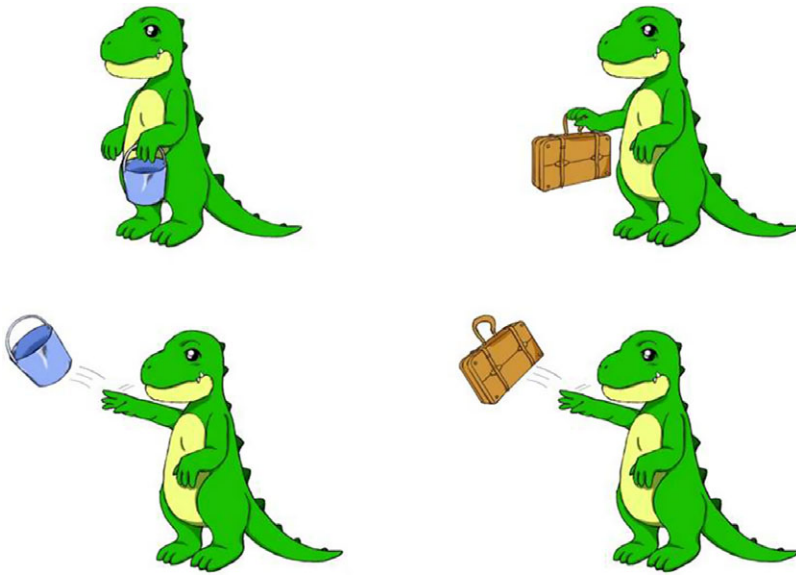


Figure 3. Example of Visual Stimulus for the Experiment.

participants achieved above 90% of gaze samples (calculated by dividing the number of eye-tracking samples that were correctly identified by the number of attempts) with a mean percentage at 95.26%, indicating that they were consistently looking at the visual stimuli during the experiment.

Native English speakers and Cantonese learners of English

The participants were positioned comfortably without a chinrest, with their eyes at a distance of 60–65 cm from a 23 inch (1,920 × 1,080) display monitor. Their eye movements were recorded with a Tobii TX300 eye tracker in remote mode, at a 300 Hz sampling rate. Freedom of movement was 37 × 17 cm at a 65 cm distance and gaze accuracy was 0.47°. Tobii Studio was used to display the stimuli and collect the data. The height of the table where the eye tracker was placed could be adjusted to get an optimal image of the eyes. Only valid data with at least one eye being successfully tracked were analyzed.

The experiment began with a 9-point automatic calibration procedure using a red dot on a white background. When the calibration was successful, the experiment started. The participants first saw instructions displayed on the screen and then began with a practice session of four trials to familiarize themselves with the task. The practice session was followed by a small break during which the participants could ask questions about the experiment. The testing session started when the participants were ready. The participants heard the auditory stimuli and saw the corresponding pictures displayed on the screen at the same time. Between each trial, a cartoon character, that is, a pink pig, unrelated to the animal characters used in the study, appeared in the center of the screen for 1,000 ms, which allowed the

participants to redirect attention to the center of the screen or to adjust their sitting position. The eye tracker recorded the eye movements of the participants throughout the experiment. No feedback was given during the experiment. Although the eye tracker did not require head stabilization, the participants were asked to sit still and to avoid body movements as much as possible. Each testing session lasted for about 20 min.

Dutch learners of English

The procedure of the experiment conducted in the Netherlands was very similar to that in Hong Kong. The experiment in the Netherlands was conducted with an EyeLink 1000 eye tracker and programmed in ZEP, a system for implementing and running psycholinguistic experiments (<https://www.beexy.nl/zep1/wiki/doku.php>). The participants' right eye movements were recorded in remote mode using a target sticker to track head movements, at a 500 Hz sampling rate. Participants were seated at a distance of 60–65 cm from the screen where the visual image was presented. The height of the participants' chair was adjusted to get an optimal image of the eye.

After the experimenter had ensured a clear image of the pupil, corneal reflection, and target sticker, the experimenter left the testing booth and a 13-point calibration and validation procedure was initiated from the control room. These were repeated until the calibration and validation were successful. Each trial was preceded by a fixation target in the middle of a blank screen. An automatic drift check was applied as the participant fixated this fixation target and a recalibration initiated if the drift check indicated a drift of more than 20 pixels. The length of the experiment for each participant was similar to that conducted in Hong Kong.

Predictions

The presence of *only* prepared the participants for the upcoming focus, and prompted them to generate a focus alternative and search for the picture depicting the alternative of focus. Once the participants proceeded to verify and disambiguate the meaning of focus based on the prosodic information, their looks were expected to diverge across the two conditions. For an object-focus experimental trial like “*The dinosaur is only carrying the BUCKET, not carrying the suitcase*”, “the suitcase carried by the dinosaur” is the alternative to the focus meaning. The participants would proceed with referential looking and fixate on the picture involving the focus (the bucket carried by the dinosaur) when hearing *BUCKET*. At the same time, the presence of *only* and prosodic prominence on *BUCKET* would trigger participants' interpretation that the dinosaur was not carrying something else. Therefore, the participants would also look at the picture indicating the object-focus alternative (the suitcase carried by the dinosaur), assuming that they looked at the pictures that reflected interpretations under consideration. Moreover, upon hearing *not*, the participants would look more at the picture that involved the object-focus alternative (the suitcase carried by the dinosaur).

In contrast, for a verb-focus trial like “*The dinosaur is only CARRYING the bucket, not throwing the bucket*”, “*throwing the bucket*” is the alternative of

Table 2. Predictions of fixations during three critical time windows across the two conditions

Object-focus condition		Verb-focus condition	
Critical time windows	Predictions	Critical time windows	Predictions
carrying	No increased fixations to the alternative of object-focus (the <i>suitcase</i> carried by the dinosaur)	CARRYING	No increased fixations to the alternative of verb-focus (the dinosaur <i>throwing</i> the bucket)
BUCKET	Increased fixations to the alternative of object-focus (the <i>suitcase</i> carried by the dinosaur)	bucket	Increased fixations to the alternative of verb-focus (the dinosaur <i>throwing</i> the bucket)
not	Increased fixations to the alternative of object-focus (the <i>suitcase</i> carried by the dinosaur)	not	Increased fixations to the alternative of verb-focus (the dinosaur <i>throwing</i> the bucket)

verb-focus. However, when participants heard the word *CARRYING*, they did not know what the upcoming object would be, which could be either the bucket or the suitcase. The possible corresponding alternative of focus during the time window of “*CARRYING*” was the dinosaur throwing either the bucket or the suitcase. Therefore, the participants would not fixate more on the picture of “the dinosaur throwing the bucket” (verb-focus alternative) until they heard the unaccented object (e.g., *bucket*). Moreover, looks targeting the display of the verb-focus alternative (the dinosaur throwing the bucket) were expected to increase during the time window of *not*. Table 2 presents the summary of the predictions during the three critical time windows (the first verb, the first object, and not) across the two conditions.

We expected that L1 speakers of English would use prosodic information to resolve the ambiguity of focus rapidly. Therefore, during the time windows of the first object and *not*, they were expected to perform more fixations on the picture of the object-focus alternative (e.g., the dinosaur carrying the suitcase) in the object-focus condition than in the verb-focus condition. They were also expected to look more at the picture of the verb-focus alternative (e.g., the dinosaur throwing the bucket) in the verb-focus condition than in the object-focus condition when hearing the first object and *not*.

If L2 learners could process focus on *only*-sentences in a native-like way, we predicted that they would display increased fixations to focus alternatives with a similar speed to L1 English speakers. Otherwise, they would either show no more looking to focus alternatives, or exhibit delayed looking patterns compared to L1 speakers of English.

Preprocessing, coding, and analysis

Experimental trials in which eye movements could not reliably be tracked were excluded from the analyses. This resulted in the exclusion of 14.2% of all trials (3.6% for L1 English speakers, 5.8% for Cantonese learners, and 4.9% for Dutch

learners). For each participant, we exported the raw eye gaze data (timestamp and gaze tracking data) using the default algorithm of Tobii Studio software for the L1 English speakers and Cantonese learners of English, and the default algorithm of EyeLink software for the Dutch learners of English. Then, we used an R script to conduct the data preprocessing and converted the raw eye-tracking data as a binomial outcome (depending on whether a participant looked at the area of interest (AOI): 0 = Not looked; 1 = Looked), which was suitable for the analysis of fixation proportion for each AOI and each time window. For example, in the time window of “verb1” (e.g., from the onset to the offset of “carrying”), each raw sample was coded as either 1 or 0 for each AOI. Then, across all the samples in the entire time window “verb1”, we calculated the proportion of fixation for each AOI, first by trial, then by condition, and finally by the participant.

Four AOIs were identified for each visual display: “Target”, “Verb-focus Alternative”, “Object-focus Alternative”, and “Distractor”. Fixations were assigned to the AOI they occurred on. For ease of reference, we take the visual display in Figure 3 as an example. The AOI “Target” referred to the picture that involved the focus meaning of the sentence, that is, the one that depicted the dinosaur carrying the bucket. The AOI “Verb-focus Alternative” referred to the picture involving the verb-focus alternative, that is, the one in which the dinosaur was throwing the bucket. The AOI “Object-focus Alternative” referred to the picture that involved the object-focus alternative in which the dinosaur was carrying something other than the bucket, that is, the suitcase. When talking about the AOI “Distractor”, it referred to the picture in which the dinosaur was throwing the suitcase.

Each experimental auditory stimulus was divided into 11 time windows for analyzing eye movements over time, as illustrated in (12). The onset and offset of each window were determined using Praat. The “gap” referred to the interval between the offset of “object1” and the onset of “not”. The final time window “offset500” referred to the auditory interval 500 ms after the offset of the sentence. For each time window, we analyzed the fixation samples falling between 200 ms after window onset and 200 ms after the window offset, considering that it takes 200 ms to launch a saccade driven by linguistic input (Altmann & Kamide, 2004).

(12) The animal is | only | verb1 | the1 | object1 | gap | not | verb2 | the2 | object2
| offset500.

For each time window, for example, “verb1” (e.g., from the onset to offset of “carrying/CARRYING”), each sample of fixation was coded as either 1 or 0 for each AOI. Then, across all the samples in each time window, we first calculated the proportions of fixation on each AOI for each trial. Then we averaged the proportion of fixations by condition and participant. This approach of data analysis was taken mainly for two reasons. First, we conducted the experiment with eye trackers in different sampling rates (300Hz for native English speakers and Cantonese learners of English, 500Hz for Dutch learners of English). The number of continuous samples from the two eye trackers was different. Second, and more importantly, each trial involved different verbs and objects, which differed in duration from trial to trial. For example, “verb1” in trial 1 (i.e., “carrying”) has a duration of 500 ms, whereas “verb1” in trial 2 (i.e., “kicking”) has a duration of 450 ms. If the

Growth Curve Analysis was used, we would end up analyzing different parts of the sentences in different trials, not necessarily the accented verbs or objects. Therefore, to minimize the influence of the variation in the duration of the spoken words, we reported proportions of fixations synchronized on a trial-by-trial basis with words in the experimental sentences (cf. Altmann & Kamide, 2007; Mirković & Altmann, 2019; Mulders & Szendrői, 2016).

To examine whether AOI and *Condition* affected L1 and L2 speakers' proportions of fixation, we used linear mixed-effects models in the *lme4* package (Bates et al., 2015) for L1 and L2 speakers separately in the R statistical program (Version 3.6.2; R Core Team, 2020). For each group, we conducted data analysis for nine time windows separately, that is, “verb1”, “the1”, “object1”, “gap”, “not”, “verb2”, “the2”, “object2”, and “offset500”. As we focused on the differences in eye gazes on the pictures of focus alternatives that could be attributed to the prosodic difference between the two conditions, we only included the AOIs “Object-focus Alternative” and “Verb-focus Alternative” in the analyses. In the models, we included fixed effects of AOI (Object-focus Alternative vs. Verb-focus Alternative), *Condition* (object-focus vs. verb-focus), and their interaction. As the data points had been averaged across the subjects, we included random intercepts for *subject* as well as random slopes for AOI, *Condition*, and their interaction. For each time window, we took the backward elimination approach, starting with the most complicated model that included all fixed effects and their interactions, and the random effects and slopes (R code: `lmer(proportion of fixation ~ AOI * Condition + (1 + AOI * Condition|subject))`) (Bates et al., 2015). Then we used the “step” function in the *lmerTest* package (Kuznetsova et al., 2017) to reduce the models by eliminating nonsignificant fixed and random effects or interactions. The analysis was conducted to test whether there was a significant interaction between AOI and *Condition*. In the section below, we report the results separately for the three groups of participants.

Results

L1 English speakers

Figure 4 presents the mean proportion of fixation time for each AOI in the object-focus and verb-focus conditions for L1 English speakers. Table 3 summarizes the output from the final best-fit model for each time window.

Against the prediction, there was no significant two-way interaction between AOI and *Condition* during the time windows of “object1” and “gap”. Specifically, there were no more eye movements toward the alternatives of focus during the above two time windows for the native English speakers. However, as shown in Figure 4, the English speakers' fixations to the AOI “Verb-focus Alternative” started to climb after hearing the accented “verb1” (during the time window of “the1”), and their fixations on the AOI “Object-focus Alternative” began to increase after hearing accented “object1” (during the time window “gap”). These looking patterns can be interpreted as English speakers' attempts to use prosodic information to verify the correspondent alternative of the focus meaning. We discuss the possible reasons for

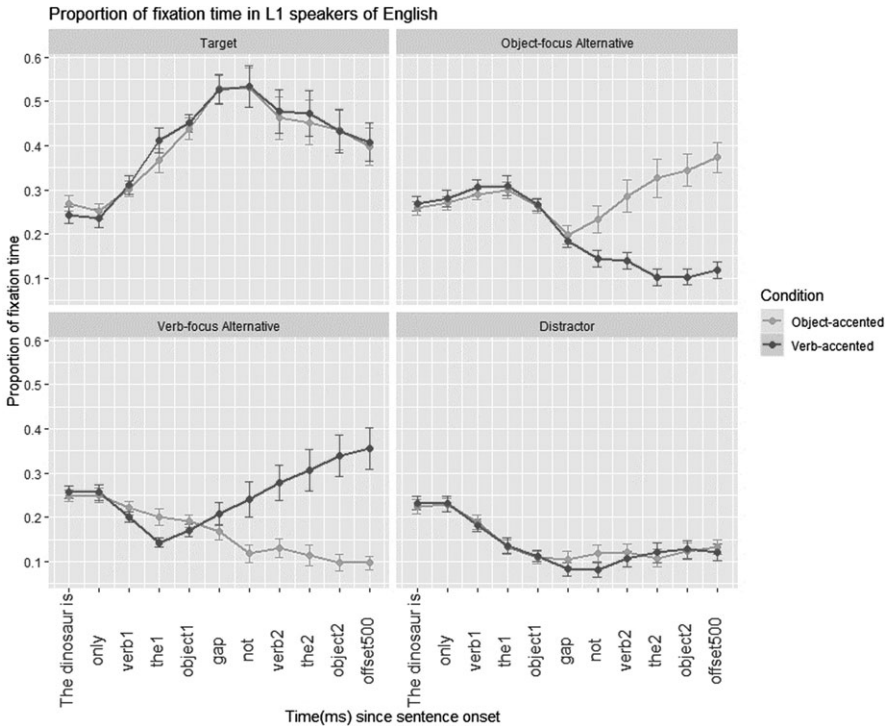


Figure 4. Proportion of Fixation Time on the Four AOIs Over Time in the Object-Focus and Verb-Focus Conditions by the L1 English Speakers. The Error Bars Indicate ± 1 Standard Error (SE)..

the lack of a significant interaction of AOI and *Condition* during these two time windows in the Discussion section.

There was a significant interaction between AOI and *Condition* for the time windows of “not”, “verb2”, “the2”, “object2”, and “offset500”. In other words, the L1 English speakers’ looking patterns began to differ significantly during the “not” time window across the two conditions. There were more fixations on the AOI “Object-focus Alternative” in the object-focus condition than in the verb-focus condition ($\beta = -.089, t = -2.723, p = .0096$). Significantly more fixations were also found on the AOI “Verb-focus Alternative” in the verb-focus condition than in the object-focus condition ($\beta = .123, t = 3.31, p = .002$). The L1 English speakers’ fixation patterns provided evidence that they have successfully computed the meaning of focus in *only*-sentences based on prosodic information as early as the time window “not”.

Cantonese learners of English

Figure 5 presents the mean proportion of fixation on each AOI in each time window across the conditions in the Cantonese learners of English. The model parameters for each time window obtained from the best-fit model are presented in Table 4.

Table 3. Model parameters for each time window in the L1 English speakers (** $p < .001$, ** $p < .01$)

Time window	Estimate	SE	<i>t</i>	<i>p</i>
verb1:				
(Intercept)	0.298	0.009	30.160	
AOI	-0.087	0.014	-6.221	<.001***
the1:				
(Intercept)	0.299	0.018	16.507	
AOI	-0.099	0.026	-3.853	<.001***
Condition	0.009	0.026	0.360	.719
AOI:Condition	-0.067	0.036	-1.842	.067
object1:				
(Intercept)	0.265	0.010	25.460	
AOI	-0.084	0.015	-5.700	<.001***
gap:				
(Intercept)	0.223	0.008	27.690	
not:				
(Intercept)	0.232	0.029	8.143	
AOI	-0.115	0.034	-3.401	<.001***
Condition	-0.089	0.034	-2.626	.0098**
AOI:Condition	0.211	0.048	4.428	<.001***
verb2				
(Intercept)	0.286	0.031	9.243	
AOI	-0.155	0.037	-4.180	<.001***
Condition	-0.147	0.037	-3.944	<.001***
AOI:Condition	0.294	0.053	5.595	<.001***
the2:				
(Intercept)	0.327	0.036	9.261	
AOI	-0.213	0.045	-4.776	<.001***
Condition	-0.224	0.045	-5.026	<.001***
AOI:Condition	0.416	0.063	6.596	<.001***
object2:				
(Intercept)	0.345	0.032	10.618	
AOI	-0.247	0.042	-5.893	<.001***
Condition	-0.242	0.042	-5.785	<.001***
AOI:Condition	0.482	0.059	8.156	<.001***

(Continued)

Table 3. (Continued)

Time window	Estimate	SE	<i>t</i>	<i>p</i>
offset500:				
(Intercept)	0.373	0.032	11.772	
AOI	-0.277	0.042	-6.530	<.001***
Condition	-0.256	0.042	-6.031	<.001***
AOI:Condition	0.513	0.059	8.572	<.001***

Bold values refer to the *p* value of the interaction between AOI and Condition.

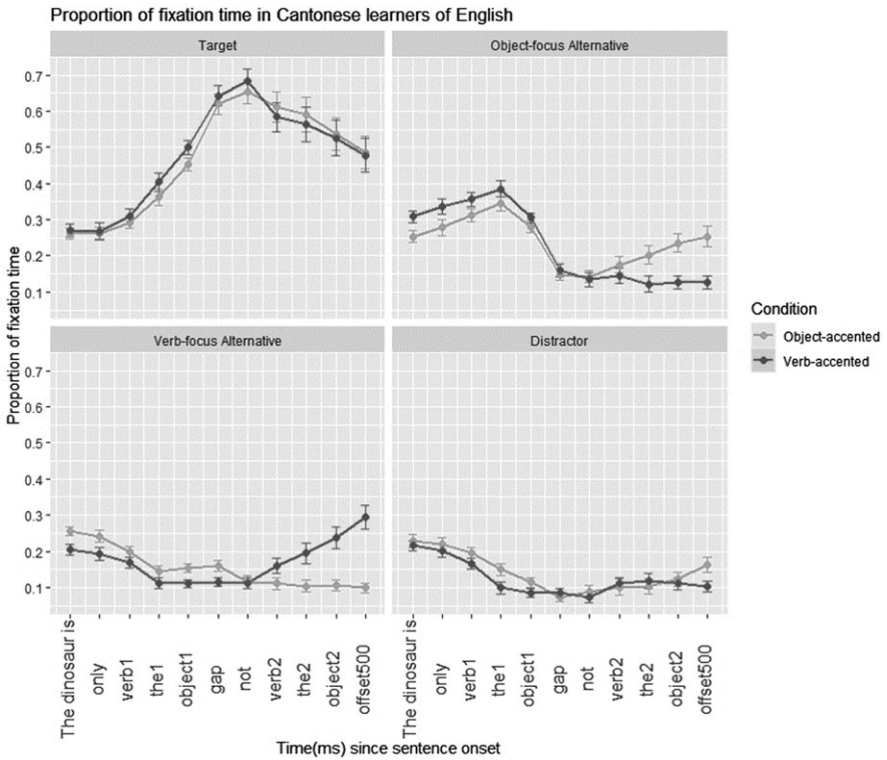


Figure 5. Proportion of Fixation Time on the Four AOIs Over Time in the Object-Focus and Verb-Focus Conditions by the Cantonese Learners of English. The Error Bars Indicate ± 1 SE.

There was a significant interaction effect of AOI and *Condition* during the time windows “verb1”, “the1”, “object1”, and “gap”. We predicted that the AOIs representing the alternatives of focus would not be more fixated until the time window “object1” if participants were able to use prosodic information to resolve the ambiguity of focus. Thus, the most relevant findings were the interaction between AOI and *Condition* during the time windows “object1” and “gap”. A closer look at the interaction effect for each time window revealed that the Cantonese learners fixated more on the AOI “Verb-focus Alternative” in the object-focus condition than in the

Table 4. Model parameters for each time window in the Cantonese learners of English (** $p < .01$, * $p < .05$, *** $p < .001$)

Time window	Estimate	SE	<i>t</i>	<i>p</i>
verb1:				
(Intercept)	0.540	0.063	8.599	
AOI	-0.113	0.025	-4.609	<.001***
Condition	0.189	0.089	2.140	.034*
AOI:Condition	-0.073	0.035	-2.109	.037*
the1:				
(Intercept)	0.745	0.068	10.950	
AOI	-0.201	0.027	-7.514	<.001***
Condition	0.187	0.096	1.946	.053
AOI:Condition	-0.073	0.038	-1.936	.055
object1:				
(Intercept)	0.531	0.049	10.702	
AOI	-0.126	0.019	-6.458	<.001***
Condition	0.164	0.070	2.337	.021*
AOI:Condition	-0.069	0.028	-2.512	.013*
gap:				
(Intercept)	0.122	0.049	2.459	
AOI	0.013	0.019	0.671	.504
Condition	0.123	0.069	1.776	.078
AOI:Condition	-0.056	0.027	-2.066	.041*
not:				
(Intercept)	0.188	0.038	4.964	
AOI	-0.025	0.014	-1.741	.084
verb2:				
(Intercept)	0.303	0.063	4.821	
AOI	-0.064	0.024	-2.642	.009**
Condition	-0.189	0.087	-2.175	.032*
AOI:Condition	0.079	0.034	2.324	.022*
the2:				
(Intercept)	0.401	0.071	5.639	
AOI	-0.099	0.027	-3.621	<.001***
Condition	-0.425	0.099	-4.315	<.001***
AOI:Condition	0.172	0.039	4.454	<.001***

(Continued)

Table 4. (Continued)

Time window	Estimate	SE	<i>t</i>	<i>p</i>
object2:				
(Intercept)	0.490	0.064	7.664	
AOI	-0.128	0.024	-5.236	<.001***
Condition	-0.589	0.088	-6.692	<.001***
AOI:Condition	0.239	0.035	6.948	<.001***
offset500:				
(Intercept)	0.562	0.074	7.555	
AOI	-0.154	0.029	-5.380	<.001***
Condition	-0.768	0.104	-7.420	<.001***
AOI:Condition	0.321	0.041	7.903	<.001***

Bold values refer to the *p* value of the interaction between AOI and Condition.

verb-focus condition during both time windows “object1” ($\beta = -.043$, $t = -2.548$, $p = .013$) and “gap” ($\beta = -.045$, $t = -2.718$, $p = .0097$). This pattern was absent in the L1 speakers of English during the same time windows.

No significant interaction of AOI and *Condition* was observed during the time window “not”. Post hoc analysis showed that the Cantonese learners did not fixate more on the AOI of focus alternative across conditions. The interaction effect became evident from the time window “verb2” and onwards (Table 4), indicating that the Cantonese learners showed more fixations to the AOIs that represented the alternatives of focus. It seems that the Cantonese learners were computing the focus meaning of *only*-sentences based on prosodic cues similarly to the L1 English speakers, but in a slower way.

Dutch learners of English

Figure 6 shows the mean proportion of fixation on the four AOIs in the Dutch learners of English. The model parameters for each time window are presented in Table 5.

The Dutch learners of English demonstrated different looking patterns, compared to the L1 English speakers. There was no significant interaction effect of AOI and *Condition* during the time window “object1” (Table 5). During the time window “gap”, there was significant AOI \times *Condition* interaction. However, post hoc comparison showed significantly more looks on the AOI “Object-focus Alternative” in the verb-focus condition than in the object-focus condition ($\beta = .091$, $t = 2.797$, $p = .0069$), and on the AOI “Verb-focus Alternative” in the object-focus condition than in the verb-focus condition ($\beta = -.054$, $t = -2.291$, $p = .0289$). The “looks going the wrong way” were also observed in the Cantonese learners of English during “object1” and “gap” time windows, but only for the AOI “Verb-focus Alternative”.

At the time windows of “not” and “verb2”, there was a significant interaction effect between AOI and *Condition*. However, the looking pattern in the time

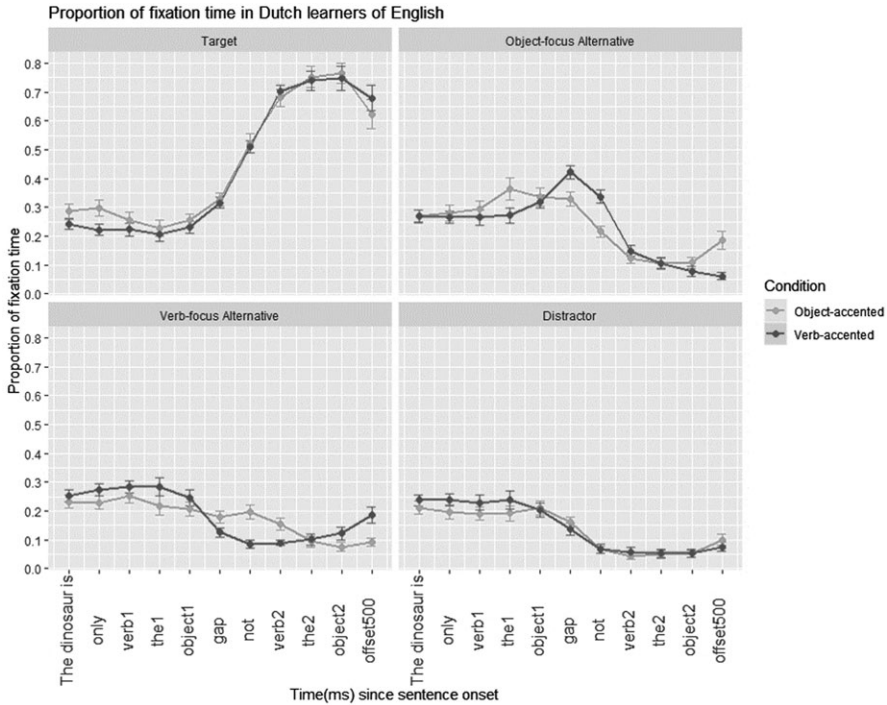


Figure 6. Proportion of Fixation Time on the Four AOIs Over Time in the Object-Focus and Verb-Focus Conditions by the Dutch Learners of English. The Error Bars Indicate ± 1 SE.

window “not” was similar to that of the time window “gap”: significantly more looks to the AOI “Object-focus Alternative” in the verb-focus condition than in the object-focus condition ($\beta = .121, t = 4.797, p < .001$), and significantly more fixations to the AOI “Verb-focus Alternative” in the object-focus condition than in the verb-focus condition ($\beta = -.112, t = -4.303, p < .001$).

During the time window “verb2”, post hoc comparison revealed significantly more looks on the AOI “Verb-focus Alternative” in the object-focus condition than in the verb-focus condition ($\beta = -.064, t = -2.794, p = .007$), and no significant difference of fixation on the AOI “Object-focus Alternative” between the two conditions ($\beta = .025, t = 1.016, p = .317$). This looking pattern for the Dutch learners was different from that of the English speakers and the Cantonese learners. From the time window “object2” and onwards, a significant interaction effect was observed. The Dutch learners began to fixate more on the AOIs of focus alternatives across the conditions.

It seems that the Dutch learners of English were computing the corresponding meaning of focus in *only*-sentences based on the prosodic cues in a different way from the L1 English speakers and the Cantonese learners of English. In the next section, we discuss how the results relate to the research questions.

Table 5. Model parameters for each time window in the Dutch learners of English (** $p < .001$, ** $p < .01$, * $p < .05$)

Time window	Estimate	SE	<i>t</i>	<i>p</i>
verb1:				
(Intercept)	0.274	0.013	21.410	
the1:				
(Intercept)	0.651	0.119	5.467	
AOI	-0.144	0.047	-3.085	.003**
Condition	-0.402	0.168	-2.386	.019*
AOI:Condition	0.155	0.066	2.351	.020*
object1:				
(Intercept)	0.613	0.036	17.220	
AOI	-0.160	0.014	-11.470	<.001***
gap:				
(Intercept)	0.631	0.076	8.302	
AOI	-0.151	0.029	-5.051	<.001***
Condition	0.383	0.108	3.561	<.001***
AOI:Condition	-0.146	0.042	-3.456	<.001***
not:				
(Intercept)	0.253	0.071	3.536	
AOI	-0.019	0.028	-0.664	.508
Condition	0.589	0.101	5.833	<.001***
AOI:Condition	-0.234	0.039	-5.903	<.001***
verb2:				
(Intercept)	0.063	0.064	0.990	
AOI	0.030	0.025	1.211	.229
Condition	0.204	0.090	2.257	.026*
AOI:Condition	-0.089	0.035	-2.525	.013*
the2:				
(Intercept)	0.101	0.012	8.528	
object2:				
(Intercept)	0.170	0.059	2.847	
AOI	-0.031	0.023	-1.365	.176
Condition	-0.182	0.084	2.171	.033*
AOI:Condition	0.076	0.033	2.318	.023*

(Continued)

Table 5. (Continued)

Time window	Estimate	SE	<i>t</i>	<i>p</i>
offset500:				
(Intercept)	0.374	0.070	5.331	
AOI	-0.094	0.027	-3.474	<.001***
Condition	-0.563	0.098	-5.768	<.001***
AOI:Condition	0.219	0.038	5.720	<.001***

Bold values refer to the *p* value of the interaction between AOI and Condition.

Discussion

This “visual world” eye-tracking study aimed to reach a better understanding of the factors underlying L1–L2 similarities and/or differences in the processing of focus. We investigated the use of prosodic information in resolving the ambiguity of focus meaning in sentences with *only* by L1 and L2 speakers of English in real time. Our work was designed to address two research questions. First, we wanted to examine the real-time processing of focus in L1 English speakers. Second, we wanted to ask whether the L2 processing of focus in *only*-sentences was similar to or different from L1 processing of the same structure. In what follows, we will first discuss the real-time processing of focus in *only*-sentences in L1 speakers of English (Section “L1 processing of focus in *only*-sentences”), then the L2 processing by the two groups of L2 learners of English (Section “L2 processing of focus in *only*-sentences”), and finally how to account for the L1–L2 difference within the frameworks of the Interface Hypothesis and the Prosodic-Learning Interference Hypothesis.

L1 processing of focus in *only*-sentences

Our results indicate that L1 English speakers processed focus in a way that was largely consistent with our predictions. They looked at the different focus alternatives to compute the meanings associated with different prosodic information when hearing “not”. These looking patterns showed that L1 English speakers used the prosodic information to comprehend focus in *only*-sentences.

One might wonder whether L1 English speakers’ looking patterns during the “not” time window might also appear in sentences without *only* if the sentences contained contrastive prosody⁷. To assess this possibility, we analyzed L1 English speakers’ proportion of fixation during the critical time window of “not” in the fillers which did not involve *only*. Our results showed no effect of AOI × Condition interaction during the time window of “not” in the fillers, indicating that L1 English speakers did not search for the focus alternative when *only* was absent. Therefore, the fixation patterns observed in L1 English speakers provide clear evidence that they have successfully computed the meaning of focus in *only*-sentences.

Although L1 English speakers’ fixations to the alternative of focus began to increase immediately after hearing the accented verb (during time window “the1”) or object (during time window “gap”) (Figure 4), this effect did not reach significance, as reflected in the time windows of “object1” and “gap” (Table 3). This finding contrasts

with Mulders and Szendrői's (2016) in which L1 Dutch speakers processed prosodic focus in sentences with *alleen* "only" immediately after hearing accented words.

We interpret this difference between our study and Mulders and Szendrői's (2016) in terms of the nature of the task and the composition of the stimuli. In our study, L1 English speakers just needed to "look and listen" and were not required to give any explicit response, whereas L1 Dutch speakers in Mulders and Szendrői's study were required to judge whether the visual display was a true description of the auditory stimulus. The fixation proportions on the focus alternative in their study may have increased because the participants needed to verify the picture and give an explicit response by the end of each trial. The "look and listen" task in our study revealed an implicit record of focus processing without the interference of task demands. Further research is needed to compare the processing of focus in eye-tracking experiments with and without explicit tasks.

Moreover, the composition of our stimuli may also have led to the lack of evidence showing immediate integration of prosody and semantics in earlier time windows. Recall that our experimental trials and fillers involved the "not" fragment (e.g., . . . , *not carrying the suitcase*). Our participants might have become accustomed to the fact that the interpretation of focus would be revealed in the "not" fragment. This in turn might have discouraged them from actively directing looks to the visual display representing the focus alternative before hearing the "not" fragment.

L2 processing of focus in only-sentences

Compared to L1 English speakers, Cantonese learners of English showed increased fixations to the focus alternative, but delayed fixation divergence, relative to the native English speakers. Dutch learners of English did not show significantly more looks toward the focus alternatives across conditions until a much later stage, relative to L1 English speakers and Cantonese learners. Comparing the two L2 groups, it seems that Cantonese learners of English showed faster processing speed than Dutch learners of English in utilizing the prosodic information to interpret focus associated with *only*.

A further remark concerns the "looks going the wrong way" in the two L2 groups during the time windows of "object1" and "gap". Recall that Cantonese learners began to show more fixations to the AOI "Verb-focus Alternative" in the object-focus condition than in the verb-focus condition in the "object1" time windows, whereas Dutch learners showed similar but delayed fixation patterns in the "gap" time window for both AOIs that involve focus alternatives. We speculate that the "looks going the wrong way" could be explained by L2 learners' delayed use of prosodic information. When they heard an accented object, for example, *BUCKET*, they directed more looks to the AOI involving the object, for example, picture in which the dinosaur is throwing the bucket (cf. Eberhard *et al.*, 1995). If this line of speculation is correct, Dutch learners were slower than Cantonese learners in making use of the prosodic information. The effect of delayed use of prosodic information in L2 learners could be tested in further studies.

One might wonder whether L2 learners' delayed fixation divergence was due to their reliance on the lexical meaning of the "verb2". In other words, their fixations to the pictures of the focus alternative could be interpreted as referential looking rather

than processing of the meaning of focus associated with different prosodic cues. If the fixation patterns observed in the time window “verb2” and onwards could be explained by referential looking, Dutch learners were even slower than Cantonese learners in lexical activation of “verb2”⁸. We think that this explanation is not plausible for two reasons. First, the words of “verb2” (e.g., *drinking* and *carrying*) were high-frequency verbs and the two L2 groups were matched in their English proficiency. It seems that language proficiency and word frequency would not lead to different rates of lexical activation in the two groups of L2 learners. Second, previous research on L2 lexical activation found cognate facilitation: L2 learners were faster in processing cognates than non-cognates (Dijkstra et al., 1999; Van Assche et al., 2013) and cognate translations produced a larger priming effect than non-cognate translations (Davis et al., 2010; Voga & Grainger, 2007). Among the 24 English verbs used as “verb2” stimuli in the current study, there were eight Dutch cognates (e.g., *drinking* – *drinken*, *baking* – *bakken*), but no Cantonese cognate or cognate translations. If this was referential looking, we would expect Dutch learners to be even faster than Cantonese learners in activating the “verb2”, but not the other way around. However, note that in the current study, Cantonese learners were faster than Dutch learners in showing increased fixation to the alternatives of focus. Thus, we think it is unlikely that the referential looking could explain the faster processing observed in Cantonese learners but not in Dutch learners. To further investigate this issue, future research could include both *only*-sentences with and without the “not”-fragments and see when the fixations start to diverge in L2 learners.

Accounting for the L1–L2 difference

Previous studies have suggested two possible accounts for explaining the L1–L2 differences. According to the Interface Hypothesis, both Cantonese learners and Dutch learners are predicted to have some difficulty in processing focus in *only*-sentences in real time, showing different eye movements or delayed eye gaze patterns. In contrast, the Prosodic-Learning Interference Hypothesis predicts that the processing of focus in English *only*-sentences will be more difficult for Dutch learners compared to Cantonese learners due to the similarity between Dutch and English. On the other hand, the difference between Cantonese and English makes L2 processing of focus easier for Cantonese learners of English.

We have observed delayed eye movement patterns in both L2 groups. Our results were largely in line with the Interface Hypothesis. It seems that both L2 groups in our study, compared to L1 English speakers, did have some difficulty in integrating multiple information sources online to resolve ambiguous focus, regardless of their L1s. Although Dutch learners of English were able to reach native-like performance in comprehending the mapping between focus and prosody without disambiguating focus interpretations (Ge et al., 2021), they might still find the real-time processing of focus more demanding. To achieve the correct interpretation of focus in *only*-sentences in real time, L2 learners not only need the knowledge of focus-to-prosody mapping, but also need to update the prosodic, semantic, and syntactic information dynamically, both based on the discourse and as the *only*-sentences unfold over time. The integration of different levels of linguistic knowledge may create more

computational demands and thus delay the L2 processing of focus in sentences with *only*.

Nonetheless, we also found that Cantonese learners, compared to Dutch learners, were faster in making use of prosodic cues to disambiguate focus meaning in *only*-sentences. This observation is not compatible with the Interface Hypothesis. We think that the difference between Cantonese learners and Dutch learners could be accounted for by the Prosodic-Learning Interference Hypothesis. As discussed in Section “Focus in sentences with ‘only’ in English, Dutch and Cantonese”, Cantonese heavily relies on a rich inventory of focus particles to encode focus, whereas Dutch uses prosody as the primary strategy to mark focus. From the traditional perspective of L1 effect on L2 processing where the similarity in L1 and L2 are assumed to facilitate L2 processing, Cantonese learners should have more difficulty in using prosodic cues to interpret focus meaning in *only*-sentences in English than Dutch learners, whereas Dutch learners should find the use of prosodic cues easier than Cantonese learners for the same purpose. Yet, it was Cantonese learners who were able to integrate prosodic cues to comprehend focus meaning in a way more similar to L1 English speakers than Dutch learners. As discussed in Section “Focus in sentences with ‘only’ in English, Dutch and Cantonese”, English and Dutch are similar but not identical in their marking of focus in *only*-sentences. The differences between Dutch and English have consequences for the preferred position of *only* relative to the position of the verb, which may contribute to Dutch learners’ difficulty in processing focus in English. When Dutch learners of English heard the English preverbal *only*-sentences, they might have treated the verb following *only* as the focused element and needed to make more effort to integrate prosodic information with the meaning of *only* to reach a focus interpretation. Since our L2 groups were matched in their English proficiency, the observed difference between Cantonese learners and Dutch learners suggests that the prosodic similarity between English and Dutch may pose more processing challenges for Dutch learners than Cantonese learners, consistent with the Prosodic-Learning Interference Hypothesis.

Another possibility for the difference between the two L2 groups could be related to L2 learners’ understanding of English prosodic focus marking⁹. Although the two L2 groups did not differ significantly in their English proficiency, reflected in their IELTS scores, Cantonese learners had a much longer experience of learning English than Dutch learners (Table 1). Cantonese learners might have a better knowledge of English prosodic focus marking due to their longer exposure to English than Dutch learners since this knowledge would not be necessarily reflected in IELTS or other English proficiency assessments. Further work is desired to address this issue.

Our findings complement and extend the previous research in a number of ways. First, our study contributes to a better understanding of online processing of focus in both L1 and L2. Our findings also provide new empirical evidence for L2 processing of information structure, from two typologically divergent and genetically unrelated L1s (Cantonese/Dutch). Our study in general suggests a delayed processing of focus in sentences with *only* in both L2 groups in real time, regardless of their L1s.

Second, our study advances our understanding of interface structures in L2 acquisition by investigating the multiple interfaces structure that involves both

internal and external interfaces. Our findings indicate that multiple interfaces do pose difficulty to the L2 processing of focus in *only*-sentences, as manifested in L2 learners' delayed fixation divergence. Furthermore, our results demonstrate that similarity between the L1 and L2 in the use of prosody to mark focus can hinder L2 processing in the domain of focus, whereas the categorical difference between L1 and L2 in terms of the use of prosody for focus marking can facilitate L2 processing in the same domain. We have shown that Dutch learners of English, unlike Cantonese learners of English, have greater difficulty in using prosodic cues to disambiguate focus in *only*-sentences. Our findings provide supporting evidence for the Prosodic-Learning Interference Hypothesis, from a perspective other than speech segmentation. Further research is needed to examine L2 learners' use of prosody in different structures with a variety of L1–L2 combinations.

Finally, the different results obtained from our study and previous eye-tracking research on L1 processing of focus (Mulders & Szendrői, 2016) highlight the importance of the nature of tasks involved in the visual world paradigm. In task-based visual search, the patterns of fixations and saccades are strongly influenced by the task performed. Further eye-tracking studies shall consider the task-based effects and include both natural visual search (e.g., “look and listen”) and tasks, which require an explicit response, such as key pressing and mouse clicking.

Despite the contributions, the current study is not without limitation. First, this study investigated how L1 and L2 speakers of English used prosodic information, which was carried by a whole word to resolve the ambiguity of focus in online sentence processing. We calculated the distribution of fixations launched in the time span between the onset and offset of each critical word. While this approach of data analysis is sufficient to address our research questions, we acknowledge that through our approach we cannot distinguish a fixation launched early in one word from a fixation launched later in the same word. Furthermore, the L1–L2 differences observed in the current study were accounted for under the theoretical framework of the Interface Hypothesis. However, we cannot exclude the possibility that both L2 groups might just show delayed L2 processing in general, regardless of interface structures. Further studies shall compare the L2 processing of non-interface structures and interface structures, and examine whether L2 learners are slower in both structures or only slower in interface structures. In addition, the current study investigated the L2 processing of focus by only two groups of English learners. Further research is needed to explore L2 learners of a variety of language pairs in order to have an in-depth understanding of L2 processing of focus in general.

Conclusion

This eye-tracking study in the visual world paradigm investigated real-time L1 and L2 processing of focus in sentences with “*only*”. In the “look and listen” task without requiring any explicit response from the participants, L1 speakers of English were able to integrate prosodic information with focus meaning. The L2 data revealed delayed fixation patterns in both Cantonese learners of English and Dutch learners of English. Moreover, Dutch learners of English were even slower than Cantonese learners of English in directing fixations to the alternative focus. We interpreted the

delayed eye movements in both L2 groups as evidence of difficulty in integrating multiple levels of linguistic information to resolve ambiguities of focus in L2 online processing, consistent with the Interface Hypothesis. The difference between the two L2 groups was in line with the Prosodic-Learning Interference Hypothesis: the similarity between English and Dutch in the use of prosody to realize focus may hinder Dutch learners' L2 processing of focus, whereas the difference between English and Cantonese in the realization of focus may facilitate Cantonese learners' processing of focus in English. Our findings have theoretical implications on L2 processing of focus in real time. The challenge for future research will be to investigate L2 processing of focus in a variety of language pairs with different methodologies.

Acknowledgments. The authors would like to thank James Britton, Alex Brouwer, and Rachida Ganga for their assistance with the study. We thank our participants and gratefully acknowledge the support from the Utrecht Institute of Linguistics – OTS, the Childhood Bilingualism Research Centre, the Research Institute for Bilingual Learning and Teaching, the University of Cambridge – CUHK Joint Laboratory for Bilingualism, and the CUHK – Peking University – University System of Taiwan Joint Research Centre for Language and Human Complexity. We have benefited from the discussion with Stephen Matthews, Roumyana Slabakova, Hannah Lam, Ziyin Mai, and Jiangling Zhou. We thank the anonymous reviewers for their constructive comments and helpful suggestions on earlier versions of this manuscript.

References

- Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition*, 6(2), 81–96.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.
- Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson, & F. Ferreira (Eds.), *The Interface of language, vision, and action: Eye movements and the visual world* (pp. 347–368). Psychology Press.
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57, 502–518.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. <http://arxiv.org/abs/1506.04967>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer software]. Version 6.0.39. <http://www.praat.org/>
- Bouma, G., Hendriks, P., & Hoeksema, J. (2007). Focus particles inside prepositional phrases: A comparison of Dutch, English, and German. *Journal of Comparative German Linguistics*, 10, 1–24.
- Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, 32(4), 643–684.
- Chafe, W. (1976) Givenness, contrastiveness, definiteness, subjects and topics. In C. N. Li (Ed.), *Subject and topic* (pp. 27–55). Academic Press.
- Claahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences*, 10(12), 564–570.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd edn. Lawrence Erlbaum Associates.
- Davis, C., Sánchez-Casas, R., García-Albea, J. E., Guasch, M., Molero, M., & Ferré, P. (2010). Masked translation priming: Varying language experience and word type with Spanish–English bilinguals. *Bilingualism: Language and Cognition*, 13(2), 137–155.

- Dijkstra, T., Grainger, J., & Van Heuven, W. J. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*(4), 496–518.
- Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye-movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.
- Foolen, A., Van Gerrevink, R., Hogeweg, L., & Prawiro-Atmodjo, P. (2009). The placement of focus particles in Dutch. *Linguistics in the Netherlands*, *26*(April 2016), 51–63.
- Fung, R. S. Y. (2000). Final particles in standard Cantonese: Semantic extension and pragmatic inference. Doctoral dissertation, The Ohio State University.
- Ge, H., Chen, A., & Yip, V. (2021). Comprehension of focus-to-accentuation mapping in sentences with only by advanced Cantonese learners and Dutch learners of English. *Studies in Second Language Acquisition*, *43*(1), 25–49.
- Gennari, S., Meroni, P. L., & Crain, S. (2004). Rapid relief of stress in dealing with ambiguity. In J. Trueswell, & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 245–259). MIT Press.
- Gu, W. & Lee, T. (2007). Effects of tonal context and focus on Cantonese F0. In *Proceedings of 16th International Conference Phonetic Science*, Saarbrücken, 1033–1036.
- Gussenhoven, C. (1983). *On the grammar and semantics of sentence accents*. Foris.
- Gussenhoven, C. (2006). Types of focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation* (pp. 83–100). Kluwer.
- Hopp, H. (2009). The syntax-discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, *12*, 463–483.
- Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. MIT Press.
- Kang, X., Joergensen, G. H., & Altmann, G. T. M. (2020). The activation of object-state representations during online language comprehension. *Acta Psychologica*, *210*, 103162.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.
- Ladd, R. (2008). *Intonational phonology*. Cambridge University Press.
- Lambrecht, K. (1994). *Information structure and sentence form: Topics, focus, and the representations of discourse referents*. Cambridge University Press.
- Lee, P. P. L. (2019). *Focus manifestation in Mandarin Chinese and Cantonese: A comparative perspective*. Routledge.
- Linguistic Society of Hong Kong. (2002). *Yueyu pinyin zibiao [Guide to LSHK Cantonese Romanization of Chinese Characters]*, 2nd edn. Linguistic Society of Hong Kong.
- Liu, F. (2009). *Intonation systems of Mandarin and English: A functional approach*. Doctoral dissertation, University of Chicago.
- Liu, J., & Lee, Y. C. (2021). Focus prosody by Korean learners of English. *Linguistic Approaches to Bilingualism*.
- Man, V. C. (2002). Focus effects on Cantonese tones: An acoustic study. In *Proceedings of 1st International Conference on Speech Prosody*, Aix-en-Provence, France (pp. 467–470).
- Matthews, S., & Yip, V. (2011). *Cantonese: A comprehensive grammar*. Routledge.
- Mirković, J., & Altmann, G. T. (2019). Unfolding meaning in context: The dynamics of conceptual similarity. *Cognition*, *183*, 19–43.
- Mulders, I., & Szendrői, K. (2016). Early association of prosodic focus with alien “only”: Evidence from eye movements in the visual-world paradigm. *Frontiers in Psychology*, *7*, 1–19.
- Ortega-Llebaria, M., & Colantoni, L. (2014). L2 English intonation: Relations between form-meaning associations, access to meaning, and L1 transfer. *Studies in Second Language Acquisition*, *36*(2), 331–353.
- R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Reichle, R. V., & Birdsong, D. (2014). Processing focus structure in L1 and L2 French: L2 proficiency effects on ERPs. *Studies in Second Language Acquisition*, *36*(3), 535–564.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75–116.

- Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica*, *137*(2), 172–180.
- Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 550–569). Blackwell.
- Shyu, S.-I. (2010). Focus interpretation of Zhi “only” associated arguments in Mandarin triadic constructions. *Linguistics*, *48*(3), 671–716.
- Slabakova, R., Kempchinsky, P., & Rothman, J. (2012). Clitic-doubled left dislocation and focus fronting in L2 Spanish: A case of successful acquisition at the syntax–discourse interface. *Second Language Research*, *28*(3), 319–343.
- Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, *1*, 1–33.
- Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, *22*, 339–368.
- Sorace, A., & Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, *13*(2), 195–210.
- Tang, S. W. (2002). Focus and dak in Cantonese. *Journal of Chinese Linguistics*, *30*(2), 266–309.
- Tremblay, A., Broersma, M., Coughlin, C. E., & Choi, J. (2016). Effects of the native language on the learning of fundamental frequency in second-language speech segmentation. *Frontiers in Psychology*, *7*, 985.
- Trommelem, M., & Zonneveld, W. (1999). Word stress in Western Germanic languages. In H. van der Hulst (Ed.), *Word prosodic systems in the languages of Europe* (pp. 478–515). Mouton de Gruyter.
- Van Assche, E., Duyck, W., & Brysbaert, M. (2013). Verb processing by bilinguals in sentence contexts: The effect of cognate status and verb tense. *Studies in Second Language Acquisition*, *35*(2), 237–259.
- Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition*, *35*(5), 938–952.
- White, L. (2011). Second language acquisition at the interfaces. *Lingua*, *121*, 577–590.
- Wang, B., & Xu, Y. (2011). Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics*, *39*, 595–611.
- Wu, W., & Xu, Y. (2010). Prosodic focus in Hong Kong Cantonese without post-focus compression. In *Proceedings of the 5th International Conference Speech Prosody*, Chicago (pp. 1–4).
- Xu, L. (2004). Manifestation of informational focus. *Lingua*, *114*(3), 277–299.
- Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197.
- Zimmermann, M., & Onea, E. (2011). Focus marking and focus interpretation. *Lingua*, *121*(11), 1651–1670.
- Zubizarreta, M. L., & Nava, E. (2011). Encoding discourse-based meaning: Prosody vs. syntax. Implications for second language acquisition. *Lingua*, *121*(4), 652–669.

Cite this article: Ge, H., Mulders, I., Kang, X., Chen, A., and Yip, V. (2021). Processing focus in native and non-native speakers of English: an eye-tracking study in the visual world paradigm. *Applied Psycholinguistics* *42*, 1057–1088. <https://doi.org/10.1017/S0142716421000230>