


ARTICLE

# Sensitivity to syntactic dependency formation in child second language processing: a study of numeral quantifiers in Korean

Hyunwoo Kim<sup>1</sup>, Kitaek Kim<sup>2</sup>, Kyuhee Jo<sup>3</sup> and Haerim Hwang<sup>4</sup> 

<sup>1</sup>Department of English Language and Literature, Yonsei University, Seoul, South Korea; <sup>2</sup>Department of English Language Education, Seoul National University, Seoul, South Korea; <sup>3</sup>Department of English Language Education, Gyeongin National University of Education, Incheon, South Korea; <sup>4</sup>Department of English, The Chinese University of Hong Kong, Hong Kong, China

**Corresponding author:** Haerim Hwang; Email: [haerimhwang@cuhk.edu.hk](mailto:haerimhwang@cuhk.edu.hk)

(Received 13 January 2023; Revised 17 March 2023; Accepted 25 April 2023)

## Abstract

This study investigates whether child second language (L2) learners can use syntactic information during the processing of sentences involving unbounded dependencies and how their processing patterns compare to those of child monolinguals and adult L2 learners. Through a self-paced reading experiment involving the numeral quantifier (NQ) construction in Korean, we tested participants' sensitivity to agreement violations between a noun phrase (NP) and an NQ in local and nonlocal conditions. The results showed that a subset of child L2 learners who demonstrated target-like knowledge of NP-NQ agreement in an offline task spent a longer processing time in the NP-NQ mismatch than in the NP-NQ match condition, in both local and nonlocal contexts. These child L2 learners' processing patterns were comparable to those observed in child monolinguals and adult L2 learners. These findings suggest that child and adult L2 learners rely on the same system of syntactic representations and processing mechanisms that guide first language processing.

**Keywords:** Child L2 processing; numeral quantifier; syntactic dependency; self-paced reading

## 1. Introduction

One critical question raised in the field of second language (L2) research is to what extent nonnative speakers can achieve native-like processing. Researchers have provided different explanations as to how L2 processing compares to first language (L1) processing, particularly in using (morpho)syntactic information (see Juffs & Rodriguez, 2014, for a review). Regarding this issue, the role of maturational constraints on L2 acquisition and processing has been a topic of considerable

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.



theoretical debate (Slabakova, 2016). One view keeps with the critical period hypothesis for L2 acquisition (Birdsong, 2014a), which posits that there exists a qualitative difference in linguistic competence between L1 speakers and adult L2 learners. Conversely, other researchers contend that L1 and L2 competence and processing mechanisms are not qualitatively distinct from each other. Processing models that support this view include the Full Transfer/Full Access/Full Parse model (Dekydtspotter et al., 2006) and Fundamental Similarity Hypothesis (e.g., Hopp, 2007). These models place greater emphasis on language use experience and memory capacity, rather than on maturational constraints, in guiding L2 processing, predicting that adult L2 learners can attain native-like processing with increased levels of proficiency and/or language experience.

Although the issue of fundamental difference or similarity between L1 and L2 processing remains unresolved, Schwartz (2004, p. 64) proposed that the “L1 child–L2 child–L2 adult comparison” helps identify the “age-dependent difference in L2 ultimate attainment,” by comparing proficiency-matched child and adult L2 learners with the same L1. Following this idea, several studies have compared child L2 learners with adult L2 learners and child L1 speakers, and yet they have primarily focused on offline comprehension (Kim & Schwartz, 2022; Unsworth, 2005). Although some studies have presented evidence of comparable processing capabilities between L1 and L2 children (e.g., Marinis, 2007; van Dijk et al., 2022), few studies have specifically investigated whether child L2 learners are fundamentally different from adult L2 learners and/or child L1 speakers in the domain of syntactic processing.

To address this gap, the current study investigates whether L2-Korean learners aged 10–12, who acquired their L2 before puberty, can draw on structural cues to compute local and nonlocal dependencies in Korean sentences. Specifically, the study compares processing behaviors between child L1 speakers and child L2 learners, as well as between child L2 and adult L2 learners, in their computation of agreement in Korean numeral quantifier (NQ) constructions. Given the inconclusive evidence regarding the availability of grammatical information in child L2 processing, comparing child L2 learners with child L1 speakers and adult L2 learners in terms of their processing of Korean NQs provides an interesting test case for the issue of whether syntactic processing in L2 children is guided by the same systems operative in L1 and/or adult L2 processing.

## 2. Literature review

### 2.1. Models of L2 sentence processing

The literature has produced mixed findings regarding the ability of L2 learners to utilize structural cues during online sentence comprehension in a manner comparable to native speakers. Previous research has diverged on whether L2 learners can construct detailed syntactic structures as efficiently as native speakers during processing.

Some researchers propose fundamental differences between L1 and L2 processing, such as Clahsen and Felser’s (2006, 2018) Shallow Structure Hypothesis (SSH). According to the SSH, adult L2 learners are more limited than L1 speakers in their ability to use fully specified structural information during online processing and tend to rely more heavily on semantic and pragmatic cues (Clahsen & Felser, 2006; Marinis et al., 2005). This hypothesis makes different predictions depending on the onset age

of L2 acquisition, attributing adult L2 learners' shallow parsing to neurocognitive factors, such as age of acquisition. However, the SSH also posits that structure-based, native-like processing is possible if L2 acquisition takes place before substantial maturational changes (Birdsong, 2014b; Birdsong & Vanhove, 2016).

Recent studies have explored the onset age effect exclusively in the domain of morphological processing. For example, Bosch et al. (2019) found that L2 onset age significantly affected the speed of L2 speakers' morphological processing in a target language. In their study, adult Russian-German bilinguals were asked to make lexical decisions on German target verbs that were visually represented on a computer screen immediately after being exposed to German prime verbs. The primes and targets formed morphologically related pairs, one with an unmarked stem (the infinitive form, e.g., *sterben* 'to die') and the other with a marked stem (the third-person singular present-tense form, e.g., *stirbt*). The results showed that the facilitation of the priming effect gradually diminished as L2 onset age increased from 0 to 11, beyond which the effect remained stable. Similar findings have been reported by other studies, which have proposed critical periods for acquiring L2 morphosyntactic information (e.g., Hartshorne et al., 2018; but see Stepanov et al., 2020, which indicated that child L2 learners only exhibit native-like processing for the target construction that is similar to their L1 counterpart). It should be noted, however, that these studies have mainly focused on morphological processing. As Clahsen and Felser (2006) noted, the SSH has the potential to apply to various processing domains, which highlights the need for systematic investigations of the onset age effect in areas beyond morphological processing.

Some researchers propose that L2 learners can use structural cues as efficiently as L1 speakers in real-time processing, attributing L2 learners' processing difficulties to their limited memory resources, rather than neurocognitive factors (e.g., Hopp, 2014, 2017; Omaki & Schulz, 2011; Witzel et al., 2012). Proponents of this "fundamental similarity" position assume that L1 and L2 processing operate on the same structural architecture, and divergent L2 processing stems from processing-related challenges. For example, some studies indicate that L2 learners with high working-memory capacity are reported to approximate native processing (e.g., Dussias & Piñar, 2009; Frenck-Mestre, 2002; Hopp, 2014). Other research suggests that high proficiency in an L2 enables learners to converge on native-like processing of various types of structural information (e.g., Fernandez et al., 2018; Frenck-Mestre, 2002; Witzel et al., 2012).

In summary, some approaches argue for fundamental differences between L1 and L2 processing that arise from maturational factors, such as onset age for L2 acquisition, whereas others adopt the fundamental identity position and maintain that divergent L2 processing patterns can be attributed to processing-related variables, such as working memory and proficiency. However, despite their potential to account for both adult and child L2 processing, these theoretical positions have not been extensively examined in the context of child L2 processing. To address this problem, the current study investigates child L2 learners' use of structural cues during online processing and compares their processing behavior with that of child L1 speakers and adult L2 learners.

## 2.2. Numeral quantifier constructions in Korean

In classifier languages like Korean and Japanese, numeral information in a noun phrase (NP) is signaled via a combination of a number and a quantifier (or classifier),

for example, *haksayng sey-myeng* (student three-NQ, ‘three students’). One crucial grammatical condition for the formulation of the NQ construction is agreement between an NQ and its host NP: The NQ must be matched with the associated NP in terms of inherent semantic properties of the NP (e.g., shape, function, animacy; Lee, 2000). Similar to lexical gender assignment in Spanish and Italian, the type of NQ is determined by the NP’s lexical properties. In the Korean sentence (1a), for example, the human NP *haksayng* ‘student’ is modified by the quantifier *myeng*, which denotes a human entity; the co-occurrence of *haksayng* ‘student’ with *kay*, a quantifier denoting a non-human, non-animate entity, would yield an agreement violation. In contrast, the non-human entity *sakwa* ‘apple’ in (1b) can only be modified by *kay*; it cannot co-occur with *myeng*.

- (1) a. Na-nun haksayng-ul sey-myeng/\*kay pwa-ss-ta.<sup>1</sup>  
 I-TOP student-ACC three-NQ<sub>human</sub>/NQ<sub>object</sub> see-PST-DECL  
 ‘I saw three students.’  
 b. Na-nun sakwa-lul sey-kay/\*myeng sa-ss-ta.  
 I-TOP apple-ACC three-NQ<sub>object</sub>/NQ<sub>human</sub> buy-PST-DECL  
 ‘I bought three apples.’

While local NP-NQ agreement is determined by concordance between NQ types and individual nouns’ membership, the relationship is also captured by a syntactic constraint; that is, the two elements must be sisters dominated by the same node in the base-generated position (Miyagawa, 1989). One of the most compelling testaments to the operation of this constraint is the floating NQ phenomenon, whereby an NQ is stranded as a result of an NP undergoing syntactic movement, for example via scrambling. The Korean sentences in (2) illustrate this point.

- (2) a. Haksayng-i sicang-eyse [sakwa-lul sey-kay/\*myeng]  
 student-NOM market-LOC apple-ACC three-NQ<sub>object</sub>/NQ<sub>human</sub>  
 sa-ss-ta.  
 buy-PST-DECL  
 ‘Students bought three apples at the market.’  
 b. Sakwa-lul<sub>i</sub> haksayng-i sicang-eyse [t<sub>i</sub> sey-kay/\*myeng]  
 apple-ACC student-NOM market-LOC three-NQ<sub>object</sub>/NQ<sub>human</sub>  
 sa-ss-ta.  
 buy-PST-DECL  
 ‘Students bought three apples at the market.’

From a generative perspective, the relationship between a moved NP and its NQ is explained as an unbounded syntactic dependency (Ko, 2007; Lee, 2000). In (2a), for example, the NQ phrase *sey-kay* ‘three-NQ<sub>object</sub>’ modifies the object NP *sakwa* ‘apple’ in the post-nominal position, forming a local syntactic constituent. In (2b), however, the NQ is stranded in the base-generated position while the object NP undergoes movement to the sentence-initial position. In this case, the fronted NP must fulfill the

<sup>1</sup>Abbreviations: ACC = accusative marker; DECL = declarative marker; NQ = numeral quantifier; LOC = locative marker; NOM = nominative marker; PST = past tense marker; Q = question marker; TOP = topic marker

locality requirement by leaving a trace within the NQ phrase so as to create a long-distance dependency (Miyagawa, 1989; Sportiche, 1988). Consequently, the NQ is associated exclusively with the fronted object NP *sakwa* ‘apple’ although it is linearly closer to the subject NP *haksayng* ‘student’. This floating NQ phenomenon is well-attested in several languages with a rich case-marking system, such as Korean and Japanese (Miyagawa, 1989; Park & Sohn, 1993; Sohn, 2001), but is precluded in languages without case-marking, such as Chinese, and does not occur in some languages that do have case-marking, such as Russian (Bondarenko & Davis, 2021; Fitzpatrick, 2006; Madariaga, 2005; Pesetsky, 1982; Stepanov & Stateva, 2018).

### 2.3. Previous studies on L2 acquisition and processing of numeral quantifier constructions

The floating NQ phenomenon in Korean and Japanese has attracted considerable attention in L1 and L2 acquisition research, with a particular focus on testing adult speakers’ knowledge of semantic and structural constraints imposed on floating NQs. Because the literature on native speakers’ acquisition and use of the NQ construction is extensive and beyond the scope of this paper, we focus on studies that have investigated nonnative speakers’ acquisition of floating NQs (see Fukuda, 2017, for an overview of L1 acquisition studies). Relevant to the present study’s focus, we mention one L1 study that examined early acquisition of NQ floating in Japanese. Using a picture-selection task, Suzuki and Yoshinaga (2013) found that five-year-old children successfully interpreted floating NQs in both SOV and OSV structures, whereas four-year-old children showed some variation in their interpretations. Based on this finding, we expect Korean L1 children in our study, aged between 10 and 12 years, to have little difficulty understanding NQ floating in SOV and OSV sentences.

In the L2 literature, only a few studies have investigated the nonnative acquisition of floating NQs, with most focusing on adult learner performance in judgment tasks. One of the early studies was conducted by Sorace and Shomura (2001), who tested the knowledge of the structural constraint underlying Japanese floating NQs among adult English speakers. The study employed two types of intransitive sentences, unergatives and unaccusatives, as illustrated in (3) and (4). The distinct structural properties of the two constructions, namely that the subject of unergatives is base-generated in the external argument position while the subject of unaccusatives is initially projected in the internal argument position (Burzio, 1986), allowed the researchers to test whether L2 learners would show sensitivity to floating NQ violations in these sentences. Because the Japanese unergative verb (e.g., *oyoi-da* ‘swam’) selects a subject NP (e.g., *shoonen* ‘boy’) as an external argument, the NQ can only license the subject NP in the adjacent position (3a), while the displacement of the two elements gives rise to a locality violation (3b). In contrast, the subject NP of an unaccusative verb (e.g., *sat-ta* ‘left’) is an internal argument, and the NQ can license the subject NP in both the nonlocal (4b) and local (4a) domains.

#### (3) Unergative

- a. *Shoonen-ga san-nin umi-de oyoi-da.*  
 boy-NOM 3-NQ sea-in swim-PST  
 ‘Three boys swam in the sea.’

- b. \*Shoonen-ga umi-de san-nin oyoi-da.  
 boy-NOM sea-in 3-NQ swim-PST  
 ‘Three boys swam in the sea.’ (Sorace & Shomura, 2001, p. 263 (20a–b))

(4) Unaccusative

- a. Kyaku-ga futari kaijyou-kara sat-ta.  
 guest-NOM 2NQ event hall-from leave-PST  
 ‘Two guests left the event hall.’
- b. Kyaku-ga kaijyou-kara futari sat-ta.  
 guest-NOM event hall-from 2NQ leave-PST  
 ‘Two guests left the event hall.’ (Sorace & Shomura, 2001, p. 263 (21a–b))

In a timed acceptability judgment task, Sorace and Shomura (2001) found that beginner-level learners with classroom instruction failed to show sensitivity to the violation of NQ floating in unergatives (e.g., (3b)), whereas intermediate-level learners who had resided in Japan for nine months showed judgment patterns similar to, but not the same as, a native speaker control group. These results led the researchers to the conclusion that the intermediate-level learners, but not the beginner-level learners, had some knowledge of the interaction of verbs’ lexical-semantic features and the syntactic operations of Japanese NQs. Sorace and Shomura pointed out that their findings alone could not show whether L2 learners can fully acquire the necessary knowledge because the intermediate-level learners showed weaker sensitivity with some verb types, and they suggested that future work should involve advanced and near-native learners.

Part of this issue was addressed by Lee (2011), who investigated advanced heritage Korean speakers’ judgment of floating NQs in Korean. Using the unergative-unaccusative contrast in Korean, which is similar to the Japanese counterpart, she tested two groups of adult bilinguals: early bilinguals who were born and grew up in the USA, and late bilinguals who arrived in the USA between the ages of 4 and 14. In a written acceptability judgment task, the early bilinguals did not distinguish between unergatives and unaccusatives in their judgments of floating NQs in Korean, unlike the late bilinguals and a control group of Korean speakers. Furthermore, even the late bilinguals were more reluctant than the native speakers to accept NQ floating in unaccusatives. Lee concluded that even advanced heritage speakers can have incomplete representations of NQ floating in Korean.

Extending these earlier studies’ findings, Fukuda (2017) tested knowledge of Japanese NQ floating with L2 Japanese learners who had taken second- or third-year Japanese classes in the USA, as well as heritage Japanese speakers who were born in the USA or immigrated to the USA at an early age. When asked to rate the acceptability of sentences involving floating and non-floating NQs in unergatives and unaccusatives, the heritage speakers performed at the same level as Japanese native speakers, indicating target-like knowledge of the structural constraint underlying NQ floating. In contrast, the L2 classroom-learners did not show target-like judgment patterns. Fukuda interpreted these results as showing evidence of native-like representations in the heritage speakers but not in the L2 learners.

Fukuda’s (2017) findings for heritage speakers are at odds with Lee’s (2011) findings for early bilinguals, who showed nontarget-like performance. However, as both authors acknowledged, their samples (13 early and 14 late bilinguals in Lee’s

study; 6 heritage and 10 L2 learners in Fukuda's) were too small to yield results that could be easily generalized. Moreover, neither of the studies tested advanced (non-heritage) L2 learners' knowledge of floating NQs, and they employed offline judgments rather than investigating online processing. Another problem with these studies is that information other than the unergative-unaccusative contrast may also affect participants' judgment patterns. For example, Fukuda found that the participants' acceptability judgments of NP-NQ agreement varied depending on the animacy status of subjects in the unaccusative sentences and the telicity of events in the unergative sentences. Such confounding factors render it difficult to draw a firm conclusion regarding L2 learners' use of structural information in their processing of NP-NQ agreement.

Kim (2018) addressed these issues by examining both the knowledge and processing of floating NQs in Korean with a larger sample of advanced late learners of Korean (L1-Mandarin Chinese,  $n = 32$ ) and by employing SOV and OSV sentences instead of unergative and unaccusative structures. In an acceptability judgment task where NP-NQ agreement was presented in local and nonlocal domains, as in (2), the L2 learners displayed native-like judgment patterns, rejecting infelicitous sentences where an object NP (e.g., *apples*) was incorrectly paired with an NQ in terms of an animacy feature (e.g., *apples-ACC students-NOM market-LOC three-kayl\*myeng bought* 'Students bought three apples at the market'). The learners also exhibited sensitivity, albeit delayed compared to native speakers, to the NP-NQ mismatch during real-time self-paced reading, spending longer times on sentences with an NP-NQ agreement violation than on the grammatical counterparts in both local and nonlocal domains. Based on these findings, Kim (2018) concluded that these advanced learners not only possessed native-like knowledge of the floating NQ construction but were able to apply that knowledge during online processing.

Collectively, these previous L2 studies suggest that while the structural constraint underlying floating NQs presents difficulty for some heritage speakers with early exposure to the majority language and L2 learners with intermediate proficiency, it is still acquirable and can even be processed in a native-like way by highly advanced learners. However, given that all of these studies targeted adult learners, a question arises as to whether such native-like attainment and processing can be achieved by child L2 learners. It is conceivable that child L2 learners would fail to converge on native-like processing of floating NQs because agreement computations between non-adjacent items require considerable processing resources (Cunnings, 2017), of which children have less than adults. Alternatively, it is possible that child learners would have little difficulty with the floating NQ construction if they had sufficient L2 proficiency and had acquired the L2 before puberty.

### 3. The present study

This study examines whether child L2 learners can exploit structural cues to detect NP-NQ agreement violations during the online processing of the Korean NQ construction. Two specific research questions (RQs) are posed, as follows:

RQ1. Do child L2 learners show structural sensitivity to Korean NQ and NP agreement to the same extent as child L1 speakers?

RQ2. How do child L2 learners' processing patterns compare to those of adult L2 learners?

To address these questions, we conducted a self-paced reading experiment that involved Korean sentences in canonical and scrambled word orders in which an NQ modifies its associated NP in a local or nonlocal domain. We also administered a picture-based interpretation task to screen L2 children for adequate levels of target knowledge. Only those learners who achieved an accuracy level above the threshold (50% for both local and nonlocal agreement conditions) were included in the self-paced reading task.

## 4. Methods

### 4.1. Participants

The study involved 78 children (47 girls and 31 boys) aged 10–12 ( $M = 10.7$ ,  $SD = 0.7$ ), including 52 L2 learners as an experimental group (cL2 group) and 26 native speakers of Korean as a control group (cL1 group). The cL2 group consisted of participants from immigrant families in South Korea, who had various L1 backgrounds, mostly Mandarin Chinese ( $n = 30$ ) or Russian ( $n = 19$ ), with one L1-speaker each of Mongolian, Thai, and Vietnamese. We later excluded the Thai-speaking child's data because Thai allows floating numeral quantifiers (Jenks, 2013), in order to avoid a potentially confounding factor from L1 influence. Although the other languages have NP-NQ structures, none of them allows nonlocal agreement, and so we included speakers of these languages. As a result, a total of 51 children contributed to the L2 data. According to their school records and responses to a language background survey, the L2 children were born and raised in their home country and exclusively spoke their native language during at least the first four to five years of their life. None had prior exposure to Korean before they came to Korea with their parents at the mean age of 7 (Range: 4–9.7,  $SD = 1.5$ ). We therefore regard them as early L2 learners of Korean.

The children in the cL2 group had resided in Korea for an average of 3.4 years (Range: 1.5–8,  $SD = 1.3$ ). They were fifth-grade students at a public primary school in Korea at the time of data collection. Following the school's system, they took a two-year-long intensive Korean language class upon entering the school, where they studied general courses in Korean two hours a day. After the program, they were assigned to regular classes along with other Korean-speaking peers. The students' teacher indicated that they were fluent in Korean and had little difficulty with listening, speaking, reading, and writing in the language at the time of testing.

We assessed the L2 participants' Korean proficiency using two measures: (a) the Diagnostic Assessment of Korean Language Proficiency (DAKLP, Noh et al., 2020) and (b) self-reports. The DAKLP comprised 25 multiple-choice questions that assessed participants' reading, vocabulary, and grammar skills in Korean. We opted for this task because it was designed for testing children learning Korean as a second language, and it only required receptive skills. Participants' accuracy scores on the DAKLP ranged from 52 to 100%, indicating variability in their proficiency levels. For self-reports, each participant rated their proficiency in four domains (reading, writing, speaking, and listening) using a 10-point scale. The mean rating across the



**Table 1.** Participant information

Group	Age	Years of studying Korean	Years in Korea	Onset age of Korean acquisition	Self-reported proficiency (1–10)
cL1 ( $n = 26$ )	10.0 (0.0)	–	–	–	–
cL2 ( $n = 51$ )	11.0 (0.6)	3.4 (1.2)	3.4 (1.3)	7.1 (1.5)	6.8 (1.6)
aL2 ( $n = 32$ )	25.5 (2.3)	4.5 (2.7)	1.8 (1.4)	21.3 (1.7)	6.9 (0.8)

Note: Values in parentheses indicate standard deviations.

four domains was 7.3 ( $SD = 1.6$ ). Each rating was significantly correlated with the others (all  $r_s > .7$ ), and with the DAKLP scores (all  $r_s > .5$ ).

To compare the results from the cL2 group with those from adult L2 learners, we retrieved data from Kim's (2018) adult L2 group (aL2 group,  $n = 32$ , mean age = 25). Because the current study adopted the same linguistic stimuli and experimental set-up used in Kim's study, it was deemed appropriate to make direct comparisons between the child and adult learners from each study. Due to the different proficiency tests administered to each group (the DAKLP to the cL2 group and a C-test to the aL2 group), we compared the two groups' proficiency in terms of their Korean learning experience and self-ratings. The aL2 group had spent a longer time studying Korean than the cL2 group ( $t(82) = -4.338$ ,  $p < 0.001$ , Cohen's  $d = -0.975$ ), while the cL2 group had spent a longer time residing in Korea ( $t(82) = 2.981$ ,  $0.004$ , Cohen's  $d = 0.670$ ) and had an earlier onset age of Korean acquisition than the aL2 group ( $t(82) = -40.455$ ,  $p < 0.001$ , Cohen's  $d = -9.089$ ). The two groups' self-ratings were not significantly different ( $t(82) = 0.064$ ,  $p = 0.949$ , Cohen's  $d = 0.014$ ).

Participant information is presented in Table 1.

## 4.2. Materials

### 4.2.1. Picture-based interpretation task

This task was conducted as a preliminary step for selecting participants who had sufficient understanding of NP-NQ agreement in local and nonlocal domains.

The experimental items consisted of 16 Korean *wh*-questions in canonical SOV and scrambled OSV word order, as in (5). Each sentence included the *wh*-phrase *myech* 'how many' followed by an NQ, asking participants to identify the number of group members in question. The goal was to test whether the participants could successfully choose the NP associated with the NQ in the *wh*-phrase in the canonical and scrambled word-order conditions. The target NP in the canonical SOV condition is adjacent to the NQ, while the target NP in the scrambled OSV condition is in the sentence-initial position, constituting a long-distance dependency with the NQ. Other words within the sentence were held constant across the conditions. Two types of NQ were used: the animal-noun quantifier *mali* and the human-noun quantifier *myeng*, both associated with animate nouns and the most frequent classifiers in Korean (Kim & Yang, 2006). All words in the sentences, including the NQs, appeared in Korean textbooks used at the participants' school.

## (5) a. Canonical SOV condition

Phaynte-ka tongmwulwen-eyse kolilla-lul myech-mali  
 panda-NOM zoo-LOC gorilla-ACC how many-NQ  
 kentuli-ess-nayo?  
 touch-PST-Q

'How many gorillas do the pandas touch at the zoo?'

## b. Scrambled OSV condition

Kolilla-lul phaynte-ka tongmwulwen-eyse myech-mali  
 Gorilla-ACC panda-NOM zoo-LOC how many-NQ  
 kentuli-ess-nayo?  
 touch-PST-Q

'How many gorillas do the pandas touch at the zoo?'

Each sentence was paired with a picture showing two groups of characters in different numbers, either animals or humans, as illustrated in Fig. 1. The names of the characters were printed in Korean below each image. The position of the target image was counterbalanced across items.

The experimental items were counterbalanced across two lists using a Latin-square design (8 items per condition), and each participant encountered only one condition of a given item. The experimental sentences were intermixed with 16 distractor items. Because the NQ was always associated with the object NP in the experimental sentences, distractor items included sentences with an NQ modifying the subject NP (e.g., *rabbit-NOM wood-in deer-by how many-NQ be chased* 'How many rabbits are chased by the deer in the wood?'). The experimental and distractor items were pseudo-randomized such that items in the same condition or of the same type never appeared in a row.

## 4.2.2. Self-paced reading task

The materials for the self-paced reading task were adopted from Kim's (2018) study. They consisted of 24 sets of Korean sentences with NP-NQ match and mismatch



Figure 1. Sample picture in the picture-based interpretation task.

conditions in canonical and scrambled word orders as in (6). All of the sentences consisted of seven words in seven regions.

- (6) a. match/mismatch conditions in the canonical word order  
 [Haksayng-i]<sub>R1</sub> [sicang-eyse]<sub>R2</sub> [sakwa-lul]<sub>R3</sub> [sey-kay/\*sey-myeng]<sub>R4</sub>  
 student-NOM market-LOC apple-ACC three-NQ<sub>object/NQ<sub>human</sub></sub>  
 [sa-se]<sub>R5</sub> [chinkwu-wa]<sub>R6</sub> [nanw-ess-ta.]<sub>R7</sub>  
 buy-and friend-with share-PST-DECL  
 ‘The students bought three apples at the market and shared them with their friends.’
- b. match/mismatch conditions in the scrambled word order  
 [Sakwa-lul<sub>i</sub>]<sub>R1</sub> [haksayng-i]<sub>R2</sub> [sicang-eyse]<sub>R3</sub> [sey-kay/\*sey-myeng]<sub>R4</sub>  
 apple-ACC student-NOM market-LOC three-NQ<sub>object/NQ<sub>human</sub></sub>  
 [sa-se]<sub>R5</sub> [chinkwu-wa]<sub>R6</sub> [nanw-ess-ta.]<sub>R7</sub>  
 buy-and friend-with share-PST-DECL  
 ‘The students bought three apples at the market and shared them with their friends.’
- Question  
 Haksayng-i sakwa-lul sa-ss-nayo?  
 student-NOM apple-ACC buy-PST-Q  
 ‘Did the students buy apples?’

Two types of NQs were consistently used in the experimental sentences: the non-animate NQ *kay* in the match conditions, and the human NQ *myeng* in the mismatch conditions. The experimental items were distributed in a 2 × 2 Latin-square design (agreement × word order) across four lists, and each participant was randomly assigned to one of the lists so that s/he saw only one of the four versions of an item. The experimental items were interleaved with 30 fillers consisting of sentences with NQ constructions in which the NQ modifies the subject NP and transitive/intransitive sentences without NQs. The fillers with NQ constructions employed one of three types of NQs: the object-denoting NQ *kay*, the human-denoting NQ *myeng*, and the animal-denoting NQ *mali*. The lexical items used in the experimental and filler sentences were selected from the vocabulary lists for beginner to intermediate learners of Korean provided by the International Standard Curriculum of Korean Language (Kim et al., 2011).

### 4.3. Procedure

Both cL1 and cL2 groups first completed a picture-based interpretation task. The cL2 group additionally completed the DAKLP as the measure of their Korean proficiency. Based on their performance in the interpretation task, we only included a subset of participants who met the inclusion threshold for the self-paced reading task (see the results section below).

#### 4.3.1. Picture-based interpretation task

The task was implemented in a pencil-and-paper format. Participants saw pictures on a computer screen while reading questions on a sheet and writing answers. For

example, given the images in Fig. 1 and one of the questions in (5), participants were expected to write “2” because the corresponding picture shows two gorillas. Participants were also given an additional choice of “I don’t know” in case they were unsure of the answer. Prior to the task, participants received written and verbal instructions on the task, followed by two practice items. In pre- and post-test interviews, all participants responded that they had little difficulty understanding the task instructions. The task took approximately 15 minutes for the cL1 group and 25 minutes for the cL2 group.

#### 4.3.2. Self-paced reading task

The self-paced reading task was conducted about two months after the interpretation task. The task was implemented using a noncumulative moving window display (Just et al., 1982) via a web-based platform using the Ibx Farm 0.3.9 software (Drummond, 2013) under the supervision of the third author. During the experiment, participants read a target sentence word by word with each word comprising a region. A series of dashes indicated the positions of the words, and participants pressed the spacebar to reveal each word. After each sentence, a verification question appeared, as in (6), and participants clicked on a “yes” or “no” response. The position of the correct answer was counterbalanced across items. Before the task, participants received oral and written instructions and worked through five practice items. All the participants confirmed that they understood the task procedure. The task took approximately 15 minutes for the cL1 group and 25 minutes for the cL2 group.

## 5. Results

### 5.1. Picture-based interpretation task

We first checked for the selection of the “I don’t know” option, which was chosen by no participant. Fig. 2 presents the mean accuracy scores for the experimental sentences in each group. As the graph indicates, the cL1 group was generally more accurate than the cL2 group in both the canonical SOV and the scrambled OSV conditions. Within each group, participants showed lower accuracy in the OSV than the SOV condition, indicating the greater difficulty of resolving the nonlocal NP-NQ dependency compared to the local NP-NQ dependency.

Given that the primary purpose of the interpretation task was to identify participants who demonstrated a sufficient understanding of NP-NQ agreement, we focused on individuals, rather than groups, who exhibited above-chance performance. Among the 51 participants in the cL2 group, 19 (10 L1-Chinese and 9 L1-Russian speakers) scored more than 4 in both SOV and OSV conditions. This subset group’s mean accuracy scores were 7.1 ( $SD = 0.9$ ) for the SOV condition and 5.9 ( $SD = 1.0$ ) for the OSV condition. In the cL1 group, a majority of participants (23 of 26) scored above 4 in both conditions, with average scores of 7.3 ( $SD = 0.9$ ) for the SOV condition and 6.4 ( $SD = 0.9$ ) for the OSV condition.

As exploratory analyses, we investigated whether the L1 and L2 participants who were retained after the screening task exhibited a comparable level of understanding of the target structure. We thus compared the accuracy of the subset of native speakers (cL1<sub>sub</sub>) and the subset of nonnative speakers (cL2<sub>sub</sub>), using mixed-effects logistic regression (Baayen, 2008). The likelihood of a correct response was modeled as a function of two fixed effects, group (cL1<sub>sub</sub>, cL2<sub>sub</sub>) and condition (SOV, OSV), all coded using deviation coding ( $-.5$  assigned to cL1<sub>sub</sub> and OSV, and  $.5$  assigned to

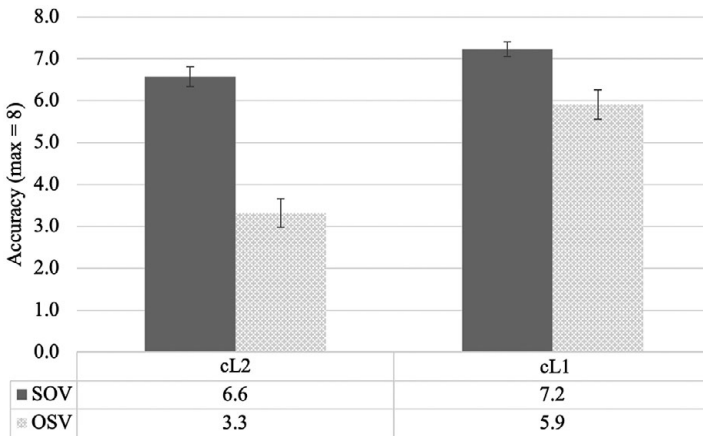


Figure 2. Mean accuracy of the picture-based interpretation task. Error bars indicate standard errors.

cL2<sub>sub</sub> and SOV) centered around the mean. We had initially included by-participant and by-item random slopes for all the fixed effects (Barr et al., 2013). However, the model with the maximal random-effects structure failed to converge, and we ended up omitting the by-participant random slope for group and the by-item random slope for condition (e.g., Barr et al., 2013); and the resulting model formula was Accuracy  $\sim$  Group \* Condition + (1 + Condition | Participant) + (1 + Group | Item). The modeling was conducted using the *glmer* function in the *lme4* package (Bates et al., 2015) in R version 3.6.3 (R Core Team, 2019). All dataset and R codes that were used in this study are available at the Open Science Framework: <https://osf.io/38jm5/>. The model only showed a robust effect of condition ( $\beta = -1.07$ ,  $SE = 0.26$ ,  $p < 0.001$ ), with higher scores for the SOV than the OSV condition, indicating that both groups performed comparably on the task.

We also investigated how this subset group of L2 children, who showed target-like performance in the interpretation task, differed from the remaining L2 participants in terms of Korean learning experience and proficiency. Independent sample *t*-tests revealed that the cL2<sub>sub</sub> group had spent a longer time in Korea ( $t(50) = 2.286$ ,  $p = 0.027$ , Cohen's  $d = 0.666$ ), had higher average self-ratings ( $t(50) = 2.164$ ,  $p = 0.035$ , Cohen's  $d = 0.631$ ), and had higher scores on the DAKLP ( $t(50) = 3.114$ ,  $p = 0.003$ , Cohen's  $d = 0.908$ ) than the remaining L2 participants. These results suggest that the high performance of the cL2<sub>sub</sub> group comparable to the cL1 group can be attributed to their increased experience with Korean and their higher proficiency in the target language.

Based on the results of the interpretation task, we included these subsets of participants (19 cL2, 23 cL1) in the following self-paced reading experiment. When we compared the subset of child L2 participants with the adult L2 learners from Kim (2018), the two groups did not differ in their years of studying Korean ( $t(49) = 0.313$ ,  $p = 0.756$ ), but the cL2 group had spent a longer time in Korea ( $t(49) = 5.536$ ,  $p < 0.001$ , Cohen's  $d = 1.603$ ), had an earlier onset age of Korean acquisition ( $t(49) = -30.338$ ,  $p < 0.001$ , Cohen's  $d = -8.786$ ), and had higher self-ratings ( $t(49) = 4.042$ ,  $p < 0.001$ , Cohen's  $d = 1.171$ ) than the aL2 group. Information of the two groups is summarized in Table 2.

**Table 2.** Information of child and adult L2 learners

Group	Age	Years of studying Korean	Years in Korea	Onset age of Korean acquisition	Self-reported proficiency (1–10)
cL2 ( <i>n</i> = 19)	11.9 (0.6)	4.2 (1.7)	4.2 (1.7)	7.5 (1.4)	8.1 (1.5)
aL2 ( <i>n</i> = 32)	25.5 (2.3)	4.5 (2.7)	1.8 (1.4)	21.3 (1.7)	6.9 (0.8)

Note: Values in parentheses indicate standard deviations.

### 5.2. Self-paced reading task

Three L1-Korean speakers were excluded from further analyses because they did not complete the self-paced reading task, leaving 20 in the cL1 group. Also removed were two participants in the aL2 group who scored less than 60% on accuracy on the verification questions. The mean accuracy rates for the verification questions were 84.7% ( $SD = 8.2$ ) in the cL1, 79.5% ( $SD = 13.1$ ) in the cL2, and 91.1% in the aL2 group. The aL2 group's accuracy scores were significantly higher than those of the cL1 group ( $p = 0.027$ ) and the cL2 group ( $p < 0.001$ ). There was no significant difference between the cL1 and cL2 groups' scores ( $p = 0.142$ ). Items for which participants provided incorrect responses on the verification questions were removed, affecting 10.6% of the data in the cL1 group, 12.1% in the cL2 group, and 6.1% in the aL2 group.

Following Kim (2018), the reading time (RT) data were trimmed by eliminating outlying values exceeding 5,000 ms (2.6% in the cL1 group, 1.1% in the cL2 group, and 0.5% in the aL2 group) and those beyond 2 standard deviations from each participant's mean RT (1.9% in the cL1 group, 2.4% in the cL2 group, and 6.4% in the aL2 group). Figs. 3–5 show RT profiles of the three groups (mean RTs and standard deviations for each region are provided in the Supplementary Material). We focused on the NQ phrase as the critical region (R4, e.g., *sey-kay/\*myeng* 'three-NQ<sub>object</sub>/NQ<sub>human</sub>' in (6)) because it is the earliest point where agreement checking potentially takes place. The subsequent two regions, R5 (e.g., *sa-se* 'buy-and' in (6)) and R6 (e.g., *chinkwu-wa* 'friend-with' in (6)), were also included in data analyses as spillover regions.

To compare the three groups' RT patterns in detail, we ran linear mixed-effects regression. For statistical analyses, the RTs were log-transformed for normal distribution (Ratcliff, 1993). We then residualized them to adjust for variability in individuals' reading speed and the length of words across items (Trueswell et al., 1994). The residual RTs were derived by subtracting the estimated RTs, obtained from a linear model that included the number of characters as a fixed effect and participant as a random effect (formula: log-transformed RT ~ number of characters + (1 | Participant)), from the actual log-transformed RTs. Three regression models were fit to the residual RTs in the critical and spillover regions, each including fixed effects of group (cL1, cL2, aL2), agreement (match, mismatch), and word order (SOV, OSV), along with random effects of participant and item. The group factor was coded using Helmert contrasts, with the first contrast being between L1 and L2 groups (cL1 versus cL2 and aL2) and the second contrast being between the two L2 groups (cL2 versus aL2). The agreement and word order factors were centered around the mean and coded using deviation coding, with  $-.5$  assigned to match and OSV, and  $.5$  to mismatch and SOV conditions. We initially constructed the model with the maximal random-effects structure, and then simplified the model by removing the

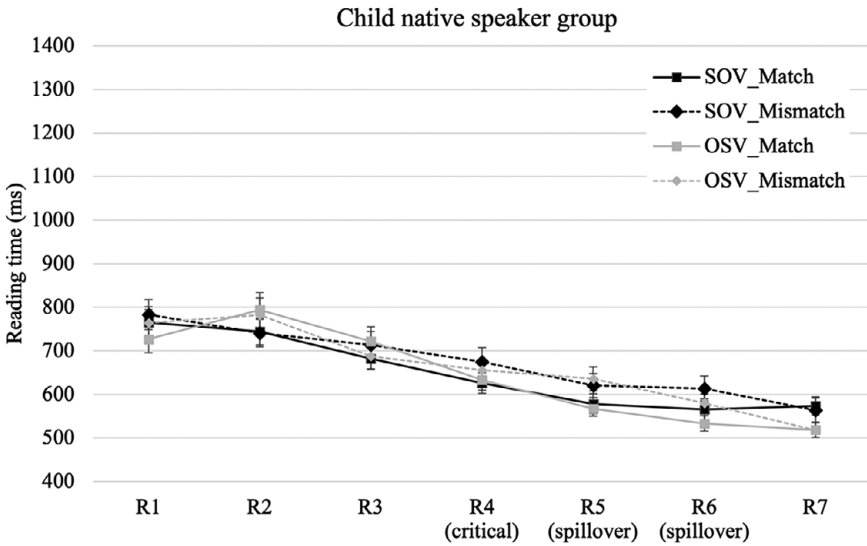


Figure 3. Child L1 speakers' reading time profile. Error bars indicate standard errors.

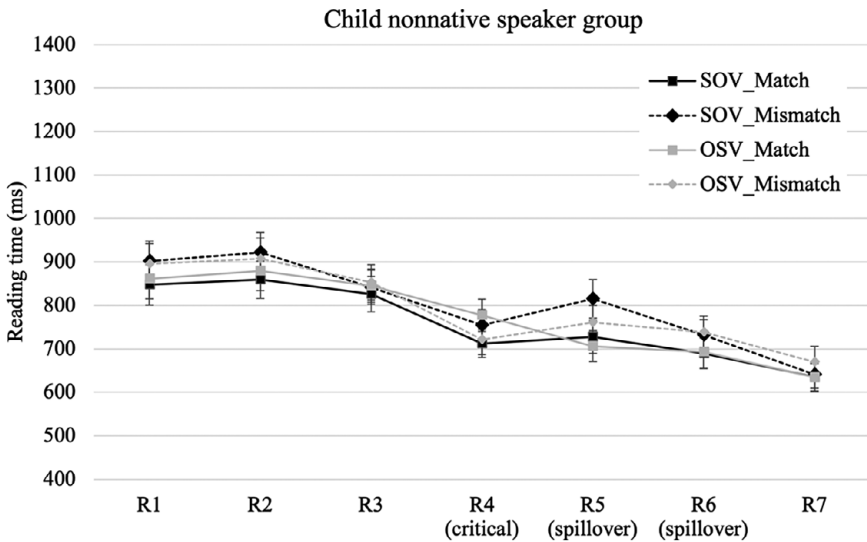


Figure 4. Child L2 speakers' reading time profile. Error bars indicate standard errors.

by-participant random slope for group and the by-item random slopes for agreement and for word order (e.g., Barr et al., 2013). The model outputs are presented in Table 3.

The only effects that reached significance in the model for the critical region (R4) were the main effect of group between L1 and L2 groups and the main effect of

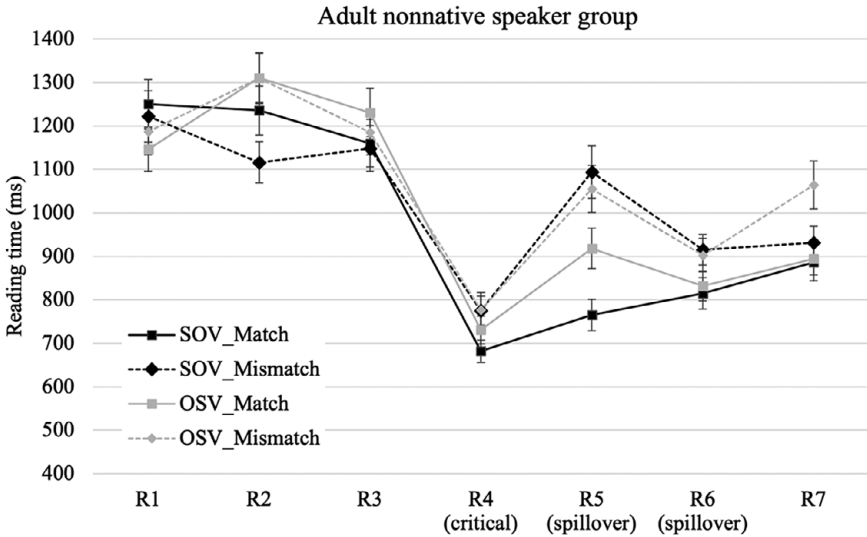


Figure 5. Adult L2 speakers' reading time profile. Error bars indicate standard errors.

group between cL2 and aL2 groups. There was no main effect associated with agreement or word order, and no interaction between factors. These results indicate that all three groups did not show sensitivity to the NP-NQ agreement violation at this region.

In the model for the first spillover region (R5), there was a main effect of agreement, with longer RTs in the mismatch than the match condition. We also found a two-way interaction between group (cL2 versus aL2) and agreement and a three-way interaction among group (cL2 versus aL2), agreement, and word order. These interactions suggest that the cL2 and aL2 groups showed different processing patterns depending on the agreement and word-order conditions. To inspect these interactions in detail, we conducted by-group analyses using linear mixed-effects regression, with agreement and word order as fixed effects (centered), and participant and item as random effects. The fixed factors were coded using deviation coding (match and OSV:  $-.5$ , mismatch and SOV:  $.5$ ). Although we initially generated the maximal random-effects structure permitted by the design, we simplified the structure by only including the by-participant random slope for agreement when the models failed to achieve convergence. Model outcomes for each group at R5 are presented in Table 4.

The model for the cL1 group revealed a main effect of agreement, with longer reading times for the mismatch than for the match condition. There was no main effect of word order and no interaction between agreement and word order. Likewise, the model for the cL2 group only showed a main effect of agreement, with no main effect of word order and no interaction of word order and agreement. These findings provide evidence that both cL1 and cL2 groups showed sensitivity to the agreement violation in the first spillover region. For the aL2 group, we also found a main effect of agreement, and yet this effect interacted with word order. To examine this interaction in detail, we conducted post-hoc comparisons using the *emmeans* package with Tukey's HSD (Lenth, 2019) in R. Results showed that the effect of agreement was



**Table 3.** Model outcomes for the critical and spillover regions

		$\beta$	SE	<i>p</i>
Region 4 (critical)	Intercept	−0.035	0.015	0.017*
	Group (L1 vs. L2)	0.071	0.031	0.023*
	Group (cL2 vs. aL2)	0.206	0.035	<0.001***
	Agreement	0.013	0.018	0.465
	Word order	−0.015	0.019	0.429
	Group (L1 vs. L2) × Agreement	−0.020	0.040	0.622
	Group (cL2 vs. aL2) × Agreement	0.021	0.043	0.631
	Group (L1 vs. L2) × Word order	−0.015	0.041	0.714
	Group (cL2 vs. aL2) × Word order	0.005	0.043	0.909
	Agreement × Word order	0.062	0.038	0.106
	Group (L1 vs. L2) × Agreement × Word order	0.086	0.084	0.309
Region 5 (spillover)	Group (cL2 vs. aL2) × Agreement × Word order	−0.074	0.089	0.408
	Intercept	−0.036	0.016	0.030*
	Group (L1 vs. L2)	0.010	0.030	0.747
	Group (cL2 vs. aL2)	−0.034	0.037	0.371
	Agreement	0.102	0.020	<0.001***
	Word order	−0.018	0.019	0.356
	Group (L1 vs. L2) × Agreement	−0.040	0.043	0.358
	Group (cL2 vs. aL2) × Agreement	−0.118	0.046	0.012*
	Group (L1 vs. L2) × Word order	0.063	0.043	0.141
	Group (cL2 vs. aL2) × Word order	0.058	0.045	0.198
	Agreement × Word order	0.058	0.038	0.127
Region 6 (spillover)	Group (L1 vs. L2) × Agreement × Word order	−0.014	0.084	0.869
	Group (cL2 vs. aL2) × Agreement × Word order	−0.191	0.089	0.033*
	Intercept	−0.096	0.017	<0.001***
	Group (L1 vs. L2)	0.035	0.028	0.261
	Group (cL2 vs. aL2)	0.006	0.039	0.875
	Agreement	0.062	0.018	0.001**
	Word order	0.001	0.016	0.955
	Group (L1 vs. L2) × Agreement	−0.022	0.041	0.588
	Group (cL2 vs. aL2) × Agreement	−0.014	0.043	0.750
	Group (L1 vs. L2) × Word order	−0.026	0.036	0.475
	Group (cL2 vs. aL2) × Word order	0.028	0.039	0.475
Agreement × Word order	0.022	0.033	0.515	
Group (L1 vs. L2) × Agreement × Word order	−0.027	0.074	0.714	
Group (cL2 vs. aL2) × Agreement × Word order	−0.064	0.078	0.416	

Note: Formula: Residual RT ~ Group \* Agreement \* Word order + (1 + Agreement \* Word order | Participant) + (1 + Group | Item).

\*\*\**p* < .001;

\*\**p* < .01;

\**p* < .05

significant only in the SOV condition ( $p < 0.001$ ), but not in the OSV condition ( $p = 0.203$ ). These results indicate that only the child L1 and L2 learners, but not the adult L2 learners, showed sensitivity to agreement violations in both the local and the nonlocal conditions in the first spillover region.

In the second spillover region (R6), we only found a main effect of agreement, induced by longer RTs in the mismatch than in the match condition (see Table 3). The single effect of agreement without an interaction with group indicates that all three groups showed sensitivity to the agreement violation. This result was confirmed by separate by-group analyses, which were conducted using mixed-effects regression in the same manner as in the analysis of the first spillover region. Model outcomes for each group at R6 are presented in Table 5.

**Table 4.** Outcomes of by-group analyses in the first spillover region

Group	Factor	$\beta$	SE	<i>p</i>
cL1	Intercept	-0.056	0.016	0.002**
	Agreement	0.056	0.027	0.037*
	Word order	-0.010	0.027	0.715
cL2	Agreement × Word order	-0.033	0.054	0.545
	Intercept	-0.029	0.018	0.131
	Agreement	0.075	0.029	0.009**
aL2	Word order	0.025	0.032	0.444
	Agreement × Word order	0.049	0.057	0.394
	Intercept	-0.022	0.036	0.538
	Agreement	0.173	0.034	<0.001***
	Word order	-0.068	0.035	0.050
	Agreement × Word order	0.158	0.067	0.019*

Note: Formula: Residual RT ~ Agreement \* Word order + (1 + Agreement | Participant) + (1 | Item).

\*\*\**p* < .001;

\*\**p* < .01;

\**p* < .05

**Table 5.** Outcomes of by-group analyses in the second spillover region

Group	Factor	$\beta$	SE	<i>p</i>
cL1	Intercept	-0.105	0.017	<0.001***
	Agreement	0.062	0.027	0.021*
	Word order	0.024	0.027	0.373
	Agreement × Word order	-0.003	0.053	0.959
cL2	Intercept	-0.073	0.023	0.005**
	Agreement	0.048	0.027	0.070
	Word order	-0.017	0.039	0.662
	Agreement × Word order	0.004	0.053	0.933
aL2	Intercept	-0.111	0.038	0.007**
	Agreement	0.077	0.027	0.005**
	Word order	-0.004	0.027	0.877
	Agreement × Word order	0.063	0.054	0.245

Note: Formula: Residual RT ~ Agreement \* Word order + (1 + Agreement | Participant) + (1 | Item).

\*\*\**p* < .001;

\*\**p* < .01;

\**p* < .05

The models showed a single effect of agreement, but no interaction of agreement with word order, for the cL1 and for the aL2 groups, and the model for the cL2 group showed a marginal effect of agreement. These results indicate that all three groups spent longer times for the sentences in the mismatch condition than for those in the match condition in the second spillover region.

In summary, the analyses of the three groups' processing patterns confirmed that all groups showed online sensitivity to the NP-NQ agreement violations; yet in the nonlocal condition, the effect emerged earlier for the cL1 and cL2 groups (at the first spillover region) than for the aL2 group (at the second spillover region).

## 6. Discussion

The primary goal of this study was to test the extent to which young L2 learners use syntactic information to detect agreement violations during the online processing of

the NQ construction in Korean. To this end, we conducted a picture-based interpretation task and a self-paced reading task. In this section, we discuss how our findings address the RQs of this study.

### 6.1. Processing patterns of child L1 and L2 speakers

Our first research question concerned whether child L2 learners exhibit similar processing patterns as child L1 speakers. The age-matched L1 and L2 children in this study, who possessed knowledge of NP-NQ agreement, patterned alike in their detection of the agreement violations in both local and nonlocal conditions. This finding is in line with previous studies demonstrating comparable processing abilities between L1 and L2 children (e.g., Marinis, 2007). These results suggest qualitatively identical structure-building routines in child L1 and L2 processing.

The target-like processing of the L2 children in this study may be attributed to at least two factors: the children's early onset of L2 acquisition and their extensive experience with Korean. All the L2 children in this study acquired the L2 before 10, an age range that lies within a period showing superior neural plasticity. However, it is difficult to convincingly argue for the decisive role of the maturational effect associated with the onset age of L2 acquisition to explain our L2 children's processing performance, because early L2 acquisition is often confounded with more L2 exposure. Indeed, our L2 children had substantial experience with the L2 through the intensive Korean program at their school as well as linguistic immersion in diverse social contexts. Crucially, despite the early onset age of L2 acquisition in all of the child participants, only the subset of children who had more extensive experience with Korean displayed a sufficient understanding of NP-NQ agreement in the picture-based interpretation task.

More compelling evidence for the stronger role of language experience would be obtained by comparing the child L2 learners' processing patterns with those of the adult learners. It is reasonable to assume that comparable processing patterns between the child and adult L2 learners would counter the idea that the onset age of L2 acquisition is determinant of native-like L2 attainment and processing. To address this issue, we now turn to our second research question, discussing the comparison of processing patterns between the child and adult L2 learners.

### 6.2. Processing patterns of child L2 and adult L2 learners

The comparison of the L2 children's results with the adult learner data revealed the two groups' largely comparable processing patterns. Although the NP-NQ agreement effect emerged earlier for the child than the adult learners, delayed processing does not necessarily indicate a reduced ability to use target information (Jackson & Dussias, 2009). Rather, the group difference lies in the quantitative domain (i.e., the timing of the computation of syntactic structures).

Note that the child L2 group had earlier exposure to Korean and had spent a longer time in Korea than the adult group, while both groups had a similar amount of time studying Korean and were currently immersed in a Korean-speaking environment. Their quantitatively distinct processing behaviors are unlikely to be due to differences in their cognitive abilities. We assume the adult, college-educated learners to be cognitively more mature than the L2 children in the current study. Nevertheless, in

the adult group, the expected agreement effect emerged only in the local condition in the first spillover region, and the effect in the nonlocal condition was delayed to the second spillover region. This delayed effect may be associated with the complexity of the NQ construction, which may have placed considerable demands on the learners' cognitive resources (for similar findings in the context of unbounded dependencies, see Dekydtspotter et al., 2006). In contrast, the child L2 learners showed faster detection of the agreement violations, despite their presumably limited working-memory capacity and greater difficulty managing their processing resources compared to adults (e.g., Kharitonova et al., 2015). It appears that the challenges of maintaining and retrieving relevant information for the children were offset by their early and extensive exposure to Korean, allowing them to employ the target syntactic information in a rapid and efficient manner.

An alternative account for the differences between the child and adult L2 groups may be related to different degrees of knowledge regarding the target structure. Note that for the self-paced reading task in the current study, we selected the subset of child L2 learners who showed sufficient knowledge in the picture-based interpretation task, whereas no such screening had been implemented for the adult comparison group, the participants in Kim's (2018) study. Although the adult learner group in the earlier study showed acceptability judgment patterns comparable to those from a native control group, this group performance does not reflect individuals' knowledge of the target structure. As a reviewer pointed out, it is therefore possible that some of the learners from the adult L2 group did not have native-like knowledge of NP-NQ agreement, leading to their deferred integration of the information. This possibility calls for a future study that administers the same screening task to child and adult participants for a precise comparison of their processing behaviors.

As noted earlier, however, the timing difference between the child and adult learners is a characteristic of a quantitative difference, and we found little evidence of qualitative differences between the two learner groups. Both groups showed a target-like sensitivity to the NP-NQ violations, which indicates the same structure-building routines underlying child L2 and adult L2 processing. Considering that both L2 groups had extensive L2 learning experience in an immersive environment, our findings suggest that L2 experience is a crucial factor that leads to native-like processing, consistent with previous studies highlighting naturalistic exposure as a necessary condition for learners' ability to process syntactic dependencies (Pliatsikas & Marinis, 2013).

## 7. Conclusions

The current study investigated the extent to which child L2 processing differs from child L1 and adult L2 processing. As far as qualitative differences are concerned, the child L2 and adult L2 groups showed fundamentally similar processing patterns. These findings suggest that child and adult L2 learners employ the same system of syntactic representations and processing mechanisms operative in L1 processing, at least when the learners have achieved high levels of L2 proficiency through extensive L2 experience. These results support the position that argues for fundamental similarity between L1 processing and L2 processing (e.g., Cunnings, 2017; Fernandez et al., 2018; Hopp, 2014, 2017). Future research should investigate whether the effects

found in this study can be generalized to other learner populations and linguistic phenomena by testing learners with diverse L1 backgrounds and L2 learning experiences on their processing of a broader set of syntactic dependencies.

**Supplementary material.** The supplementary material for this article can be found at <http://doi.org/10.1017/langcog.2023.16>.

**Data availability statement.** All dataset and R codes that were used in this study are available at the Open Science Framework: <https://osf.io/38jm5/>.

## References

- Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge University Press.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Birdsong, D. (2014a). The critical period hypothesis for second language acquisition: Tailoring the coat of many colors. In M. Pawlak & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism. Studies in honor of David Singleton* (pp. 43–50). Springer.
- Birdsong, D. (2014b). Dominance and age in bilingualism. *Applied Linguistics*, 35, 374–392.
- Birdsong, D., & Vanhove, J. (2016). Age of second language acquisition: Critical periods and social concerns. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 163–182). De Gruyter.
- Bondarenko, T., & Davis, C. (2021). Sub-extraction asymmetries and linearization in Russian. *Journal of Slavic Linguistics*, 29, 1–13.
- Bosch, S., Veríssimo, J., & Clahsen, H. (2019). Inflectional morphology in bilingual language processing: An age-of-acquisition study. *Language Acquisition*, 26, 339–360.
- Burzio, L. (1986). *Italian syntax: A government and binding approach*. Reidel.
- Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, 27, 107–126.
- Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, 40, 693–706.
- Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, 20, 659–678.
- Dekydtpotter, L., Schwartz, B. D., & Sproule, R. A. (2006). The comparative fallacy in L2 processing research. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 33–40). Cascadilla Proceedings Project.
- Drummond, A. (2013). *Ibex Farm*. Retrieved from <http://spellout.net/ibexfarm/>
- Dussias, P. E., & Piñar, P. (2009). Sentence parsing in L2 learners: Linguistic and experience-based factors. In W. C. Ritchie & T. K. Bathia (Eds.), *The new handbook of second language acquisition* (pp. 295–318). Emerald.
- Fernandez, L., Höhle, B., Brock, J., & Nickels, L. (2018). Investigating auditory processing of syntactic gaps with L2 speakers using pupillometry. *Second Language Research*, 34, 201–227.
- Fitzpatrick, J. M. (2006). The syntactic and semantic roots of floating quantification (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Boston, MA.
- French-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 217–236). Elsevier.
- Fukuda, S. (2017). Floating numeral quantifiers as an unaccusative diagnostic in native, heritage, and L2 Japanese speakers. *Language Acquisition*, 24, 169–208.
- Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 177, 263–277.

- Hopp, H. (2007). *Ultimate attainment at the interfaces in second language acquisition: Grammar and processing*. Grodil Press.
- Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition*, 21, 250–278.
- Hopp, H. (2017). The processing of English which-questions in adult L2 learners: Effects of L1 transfer and proficiency. *Zeitschrift für Sprachwissenschaft*, 36, 107–134.
- Jackson, C. N., & Dussias, P. E. (2009). Cross-linguistic differences and their impact on L2 sentence processing. *Bilingualism: Language and Cognition*, 12, 65–82.
- Jenks, P. (2013). Quantifier Float, Focus, and Scope in Thai. Proceedings of BLS39. Retrieved from <http://ling.auf.net/lingbuzz/001960>.
- Juffs, A., & Rodriguez, G. A. (2014). *Second language sentence processing*. Routledge.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228–238.
- Kharitonova, M., Winter, W., & Sheridan, M. A. (2015). As working memory grows: A developmental account of neural bases of working memory capacity in 5-to 8-year old children and adults. *Journal of Cognitive Neuroscience*, 27, 1775–1788.
- Kim, H. (2018). Second language processing of Korean floating numeral quantifiers. *Journal of Psycholinguistic Research*, 47, 1101–1119.
- Kim, J. B., & Yang, J. (2006). Processing Korean numeral classifier constructions in a typed feature structure grammar. In *International Conference on Text, Speech and Dialogue* (pp. 103–110). Springer.
- Kim, J. S., Kim, C. S., Kang, H. H., Kim, J. W., Kim, H., & Lee, J. H. (2011). *Korean educational model for international standards (written in Korean)*. National Institute of Korean Language.
- Kim, K., & Schwartz, B. D. (2022). Learnability in the acquisition of the English tough construction by L1-Korean adult and child L2 learners. *Second Language Research*, 38, 259–287.
- Ko, H. (2007). Asymmetries in scrambling and cyclic linearization. *Linguistic Inquiry*, 38, 49–83.
- Lee, C. (2000). Numeral classifiers, (in-)definites and incremental theme in Korean. In C. Lee & J. Whitman (Eds.), *Korean syntax and semantics: LSA institute workshop* (p. 91). Thaeaksa.
- Lee, T. (2011). Grammatical knowledge of Korean heritage speakers: Early vs. late bilinguals. *Linguistic Approaches to Bilingualism*, 1, 149–174.
- Lenth, R. (2019). Emmeans: Estimated marginal means, aka least-squares means. <https://CRAN.R-project.org/package=emmeans>
- Madariaga, N. (2005). Russian patterns of floating quantification: (Non-)agreeing quantifiers. In P. Kosta, G. Hassler, L. Schurcks, & N. Thielemann (Eds.), *Linguistic investigations into formal description of Slavic languages, Potsdam Linguistic Investigation*. Frankfurt-am-Main.
- Marinis, T. (2007). On-line processing of passives in L1 and L2 children. In A. Belikova, Meroni, & M. Umeda (Eds.), *Proceedings of the 2nd conference on generative approaches to language acquisition North America (GALANA)* (pp. 265–276). Cascadilla Proceedings Project.
- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, 27, 53–78.
- Miyagawa, S. (1989). *Structure and case marking in Japanese*. Academic Press.
- Noh, J., Jang, M., & Park, J. (2020). Trends and challenges in research on the diagnostic assessment of Korean language proficiency and remedial instruction for KSL. *Eomunhak*, 148, 33–63.
- Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, 33, 563–588.
- Park, M., & Sohn, K. (1993). Floating quantifiers, scrambling, and the ECP. In S. Choi (Ed.), *Japanese/Korean linguistics 3* (pp. 187–203). CSLI Publications.
- Pesetsky, D. (1982). Complementizer-trace phenomena and the nominative island condition. *The Linguistic Review*, 1, 297–344.
- Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, 16, 167–182.
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510–532.
- Schwartz, B. D. (2004). Why child L2 acquisition? In J. van Kampen & S. Baauw (Eds.), *Proceedings of GALA 2003* (Vol. 1, pp. 47–66). Netherlands Graduate School of Linguistics (LOT).

- Slabakova, R. (2016). *Second language acquisition*. Oxford University Press.
- Sohn, H. (2001). *The Korean language*. Cambridge University Press.
- Sorace, A., & Shomura, Y. (2001). Lexical constraints on the acquisition of split intransitivity: Evidence from L2 Japanese. *Studies in Second Language Acquisition*, 23, 247–278.
- Sportiche, D. (1988). A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry*, 19, 425–449.
- Stepanov, A., Andreetta, S., Stateva, P., Zawiszewski, A., & Laka, I. (2020). Anomaly detection in the processing of complex syntax by early L2 learners. *Second Language Research*, 36, 371–397.
- Stepanov, A., & Stateva, P. (2018). Countability, agreement and the loss of the dual in Russian. *Journal of Linguistics*, 54, 779–821.
- Suzuki, T., & Yoshinaga, N. (2013). Children's knowledge of hierarchical phrase structure: Quantifier floating in Japanese. *Journal of Child Language*, 40, 628–655.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285–318.
- Unsworth, S. (2005). Child L2, adult L2, child L1: differences and similarities. A study on the acquisition of direct object scrambling in Dutch. (Unpublished doctoral dissertation). Utrecht University, Utrecht: LOT.
- van Dijk, C., Dijkstra, T., & Unsworth, S. (2022). Cross-linguistic influence during online sentence processing in bilingual children. *Bilingualism: Language and Cognition*, 25, 691–704.
- Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics*, 33, 419–456.

---

**Cite this article:** Kim, H., Kim, K., Jo, K. & Hwang, H. (2023). Sensitivity to syntactic dependency formation in child second language processing: a study of numeral quantifiers in Korean *Language and Cognition* 15: 763–785. <https://doi.org/10.1017/langcog.2023.16>