Linguistic Research 43(1): 1-30
DOI: 10.17250/khisli.43.1.202603.001

Definite descriptions in L2 acquisition: An approximate replication of Ionin et al. (2012)*

Vatcharit Chantajinda[a] · Hyun Bae[b] · Joseph Moran[a] · Jacee Cho[a]**
(University of Wisconsin–Madison[a] · University of Southern California[b])

Chantajinda, Vatcharit, Hyun Bae, Joseph Moran, and Jacee Cho. 2026. Definite descriptions in L2 acquisition: An approximate replication of Ionin et al. (2012). *Linguistic Research* 43(1): 1-30. The present study is an approximate replication of Ionin et al. (2012). In a picture-based comprehension task (similar to an act-out task) used in the original study, L1-English speakers interpreted English definite plurals as referring to (i) a set of previously mentioned objects and (ii) all objects within the context. However, L1-Korean L2-English speakers allowed only the former. This non-targetlike behavior was attributed to L1 transfer from the semantics of Korean demonstratives. Methodologically, the task might reflect participants' preference rather than their representations for definite plurals since participants could express only one possible interpretation in the picture-based comprehension task. We thus modified the task changing it into a picture-based acceptability judgement task wherein participants were shown both interpretations. Participants were thirty-three L1-English and thirty-one L1-Korean L2-English speakers. Ordinal logistic regression models revealed that L1-Korean participants accepted both interpretations. We argue that non-targetlike behavior in the original study is attributable to task effects. We conclude that L1-Korean speakers compute both interpretations and have targetlike representations of English definite plurals. **(University of Wisconsin–Madison · University of Southern California)**

Keywords definiteness, demonstrative, article, transfer, replication

* We would like to thank Tania Ionin for kindly sharing all materials used in the original study with us. Our appreciation also goes to the members of the SLA lab at UW–Madison for their help with participant recruitment and pilot testing, especially Eleanor Sand who contributed to the initial stage of the current study. We are also grateful for constructive comments provided by the audience at EuroSLA32 in Birmingham as well as thoughtful discussions with Keisuke Kume. Support for this research was provided by the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation.
** First and corresponding author

1. Introduction

The English article system has been found to present learning difficulties to second language (L2) learners whose first language (L1) lacks an article system such as Chinese, Korean, Russian, Japanese, and Thai (e.g., Ionin et al. 2004; Trenkic and Pongpairoj 2013; J. Cho 2017; Snape 2018). Additionally, the acquisition of the semantics of English definite plurals (such as *the dogs*) poses further challenges for L2 learners who must not only acquire the article system itself, but also learn to distinguish between two possible interpretations of definite plurals: (i) a uniquely identifiable set of entities in discourse vs. (ii) all of the entities in the context that satisfy the description.

This study considers the comprehension of English definite descriptions (*the dog, the dogs*) and demonstrative descriptions (*that book, those dogs*) by replicating the work of Ionin et al. (2012). The original study investigated the L2 acquisition of English articles by speakers of L1 Korean, a language that lacks articles but has demonstratives. Results from a written production task and a picture-based comprehension task suggest that intermediate proficiency L2 learners are significantly less likely to distinguish between the definite article *the* and demonstrative *that* in descriptions (e.g., *the dogs* = *those dogs*), suggesting that L1-Korean participants transferred the semantics of the Korean demonstrative *ku* 'that' onto their L2. Such transfer was claimed to contribute to non-targetlike performance.

While the original study presents robust findings on L2-English speakers' interpretation patterns and use of the article and demonstrative, questions remain as to whether the observed results reflect the participants' underlying representations for L2 English definite descriptions or their preferred interpretation for English definite plurals. In particular, insofar as the original study's comprehension task required participants to read instructions then complete a series of drawings according to those instructions in a manner similar to an act-out task (Goodluck 1996), their drawings necessarily exhibited only one possible interpretation of the definite description. That is, results from the comprehension task do not rule out the possibility that participants had targetlike comprehension of definite descriptions (i.e., two interpretations), but nevertheless completed the task according to their preferred interpretation of the definite description which was modulated by L1 transfer of Korean demonstrative semantics. Accordingly, the present replication study seeks to evaluate L1-Korean

L2-English speakers' comprehension of definite and demonstrative descriptions by modifying the picture-based comprehension task into a picture-based acceptability judgment task (AJT) in which participants are asked to judge whether the drawings match the instruction lines.

1.1 Definite plurals in English

Definite plurals in English (*the* + a plural noun phrase (NP), e.g., *the dogs*) have two possible interpretations. They can refer to either a set of uniquely identifiable entities in the context, described by the uniqueness approach (Hawkins 1991), or the maximal/total set of the entities in the context which satisfy the description, referred to as the maximality interpretation (Sharvy 1980; Heim 1991) or inclusiveness (Hawkins 1978). This ambiguity is demonstrated in example (1) below:

(1) Context: There are six dogs in a park and two children are watching them.
A: Look, three dogs are white!
B: The dogs are adorable.

In this example, Speaker A mentions that three dogs are white, forming a specific set of all the dogs within the discourse. Speaker B then uses the definite plural *the dogs* which has two possible interpretations here: *the dogs* can refer to either (i) the three white dogs mentioned by Speaker A or (ii) all six dogs in the park. In this paper, we refer to the interpretation in (i) as 'same' interpretation and in (ii) as 'all' interpretation.

The uniqueness exemplified by definite plurals as in (1) is also shared by demonstrative plurals (*those* + plural NP, e.g., *those books*), and the two are interchangeable in certain context—evidence that the definite article and demonstratives share a central semantics of denoting uniquely. However, the two varieties are not universally interchangeable insofar as demonstrative NPs only denote uniquely relative to entities that are immediately perceptible via textual introduction or demonstration such as a pointing gesture, whereas definite NPs can denote entities that are not immediately perceptible as long as uniqueness is established in the context. This is illustrated in (2) and (3) below, taken from the original study (Ionin et al.

2012: 75). The felicity of *the* and *that* in the two examples demonstrates a functional overlap between the two determiners to denote unique referents. In (3), *that* is pragmatically infelicitous because it carries an implication of contrast between two nouns of the same category, e.g., *a woman* and *another woman*.

(2) The curtain rose. A woman came onto the stage. Then the/that woman started singing and dancing.
(3) The curtain rose. A woman and a man came onto the stage. Then the/#that woman started singing and dancing.

1.2 Demonstrative descriptions in Korean

Unlike English, Korean lacks an article system, but it possesses three demonstrative determiners (the proximal form *i* 'this,' the distal form *ce* 'that over there,' and the neutral form *ku* 'that') which behave similarly to English demonstratives insofar as they can be used to pick out an individual that is salient given the context of utterance (see Sohn 1999: 210). While all three forms can be used in deictic contexts, the distal form *ce* cannot be used in anaphoric contexts, and the neutral form *ku* is generally preferred (Lee and Song 2010).

In deictic use, *ku* behaves more similarly to the demonstrative *that* than the article *the* as in (4). In anaphoric contexts, *ku* can be used to refer to a unique antecedent as in (5). In non-anaphoric contexts, *ku* can be used in situations where the uniqueness is satisfied by the common ground between interlocutors, as in (6) but the referent is neither immediately available nor mentioned explicitly in the discourse. In this way, *ku* has been argued to match the semantics of the definite descriptor *the* in certain contexts (H. Cho 1999). In sum, the demonstrative *ku* has adopted partial functionality of a definite determiner while maintaining the core semantics of demonstratives as demonstrated by the examples (4) – (6) below from H. Cho (1999, as cited in Ionin et al. 2012: 76–77).[1]

1 We edited H. Cho's (1999) glossing in these examples to make the glosses more accurate and to follow the Leipzig Glossing Rules.

(4) Ku chayk caymi iss-ni? (pointing to a book)
DEM book interesting be-Q
'Is that book interesting?'

(5) Tom-i nay-key chayk-ul sa-cwu-ess-ta.
Tom-NOM me-DAT book-ACC buy-BEN-PST-DECL
Ku chayk-un caymi iss-ess-ta.
DEM book-TOP interesting be-PST-DECL
'Tom bought a book for me. The book was interesting.'

(6) Ku coffee shop-eyse manna-ca.
DEM coffee shop-at meet-IMP
'See you at the coffee shop.'

2. The original study

Ionin et al. (2012) investigated the acquisition of definite and demonstrative descriptions in L2 English by L1-Korean learners, focusing on how they interpret definite and demonstrative descriptions. Due to the semantics of the Korean demonstrative *ku* 'that', which overlaps with the semantics of *the* in anaphoric contexts and with the semantics of *that* in non-anaphoric contexts, Ionin et al. (2012) hypothesized that Korean learners would exhibit L1 transfer effects by mapping the semantics of *ku* onto both *the* and *that* in English. Ionin et al. predicted that Korean learners would treat both *the* and *that* as equivalents of *ku*, as a result of which they would not be able to distinguish between definite and demonstrative descriptions. In contexts where L1-English speakers prefer *that*, L1-Korean L2-English learners were expected to allow both *the* and *that*, suggesting a significant influence of L1 transfer.

To test these predictions, Ionin et al. (2012) employed a combination of a forced-choice production task, a proficiency test, and a picture-based comprehension task. Participants included 21 L1-English speakers and 48 L1-Korean-speaking learners of L2 English, tested in both the U.S. and South Korea. The L2 learners were divided into intermediate and advanced groups based on their proficiency test scores. In the production task, each test item consisted of a short story with a blank space which participants were asked to fill in by choosing from *a*, *one*, *the*, or *that*.

In the comprehension task, participants were presented with a three-line

instruction set for each item with twelve objects: six of one kind and six of another. With a sentence introducing the objects, the three-line command asked participants to draw geometric shapes on the objects. The three test conditions were demonstrative plurals (e.g., *those balloons*), definite plurals (e.g., *the books*), and indefinite plurals (e.g., *some cars*). In the demonstrative condition, participants were expected to act on a maximal set of the object mentioned in the previous instruction line since only salient referents should be acted upon. In the definite condition, due to the two possible interpretations of English definite plurals, participants could act on a set of the object mentioned in the previous line (Figure 1) or on all the members of the object (Figure 2), indicating the 'same' and 'all' interpretations, respectively. Likewise, we modified the two examples from test materials used in the original study.[2] In the indefinite condition, added as a control condition, participants should not act upon the previously mentioned objects because indefinites are not used with unique antecedents.

Here are six cups and six houses.

1. Please draw circles around two houses.
2. Now, please draw lines below two cups.
3. Now, please draw arrows above the houses.

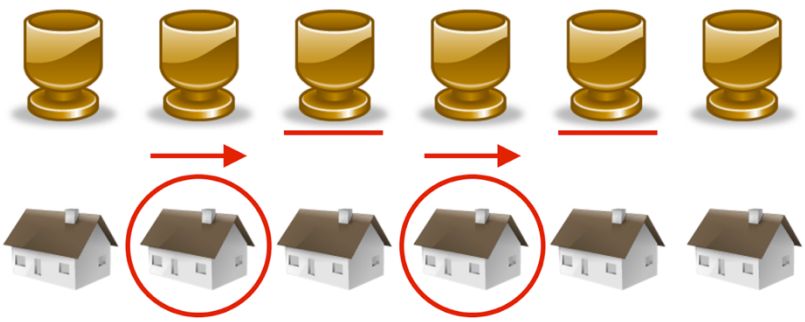

Figure 1. Example item with a 'same' interpretation response

2 We thank Dr. Tania Ionin for sharing the test materials with us.

Here are six cups and six houses.

1. Please draw circles around two houses.
2. Now, please draw lines below two cups.
3. Now, please draw arrows above the houses.

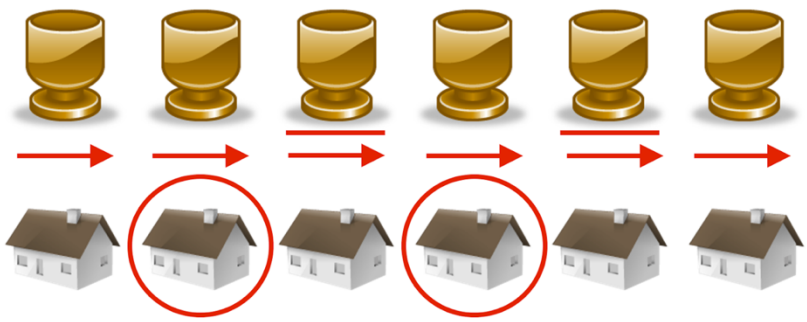

Figure 2. Example item with an 'all' interpretation response

The results confirmed that while L1-Korean L2-English learners could distinguish between definite and demonstrative descriptions, this distinction was not as robust as in L1-English native speakers. Furthermore, while target-like performance was observed in the production task for the advanced learners (but not the intermediate), in the picture-based comprehension both advanced and intermediate L2 learners demonstrated target-divergent performance. Table 1 summarizes the results from the comprehension task which is relevant to our replication. In the definite condition, L1-English speakers allow both 'all' and 'same' interpretations, 41% and 58%, respectively. L1-Korean L2-English speakers at both proficiency levels, however, showed an overwhelming preference for the 'same' interpretation over the 'all' interpretation in both the demonstrative and definite conditions indicating transfer of the semantics of Korean demonstratives. These findings supported the hypothesis of L1 transfer affecting the L1-Korean L2-English learners' use and interpretation of English articles.

Table 1. Main findings of the original study's comprehension task

Condition	L1-English	Advanced L2-English	Intermediate L2-English
Demonstrative plurals	96% of the 'same' interpretation	100% of the 'same' interpretation	81% of the 'same' interpretation (5% 'all' and 8% 'different objects')
Definite plurals	58% of the 'same' interpretation and 41% of the 'all' interpretation	97% of the 'same' interpretation	68% of the 'same' interpretation (5% 'all', 22% 'different objects')
Indefinite plurals	100% of different object responses	97% of different object responses	92% of different object responses (5% 'same')

3. The current replication study

3.1 Motivation for replication

In recent years, replication research has gained prominence in the social sciences—including second language acquisition—as a means of evaluating the generalizability of influential findings (Porte and McManus 2019; McManus 2022). Ionin et al.'s (2012) study merits replication for both theoretical and methodological reasons. Numerous studies on L2 article acquisition repeatedly showed that speakers whose L1 lacks articles have difficulty in acquiring articles due to L1 influence. To explain the mechanisms of L1 effects in L2 article acquisition, three different approaches have been proposed: lexical transfer, feature reassembly (Lardiere 2008, 2009), and the Morphological Congruency Hypothesis (Jiang et al. 2011). While there is a good number of studies on L2 article acquisition from the feature reassembly account (J. Cho 2017; Tuniyan 2018; Feng 2019; Zhang 2021) and the Morphological Congruency Hypothesis (Ionin et al. 2021; Snape et al. 2023), there have been relatively few experimental studies that investigated the role of L1 lexical transfer in L2 article acquisition.[3] Ionin et

3 While the Feature Reassembly Hypothesis and the Morphological Congruency Hypothesis provide broader theoretical accounts of L2 acquisition of articles, both the original study and the present replication were not designed to directly test the hypotheses. Accordingly, we limit our discussion to the patterns observed for definite plurals in anaphoric contexts in L1 vs L2.

al. (2012) is one of the few studies that systematically investigated whether article-less L1 speakers' non-targetlike use of English articles stems from lexical transfer of demonstrative semantics in the L1. Thus, replicating Ionin et al.'s study is important for evaluating the generalizability of their findings as well as the theoretical tenability of the lexical transfer approach to L2 article acquisition. Replicating Ionin et al. is also important on the methodological level. In recent years, there has been much discussion on methodological quality in applied linguistics research. One issue raised in Plonsky (2024) is the validity of instruments used for measuring linguistic behaviors or relevant constructs (e.g., L2 anxiety, motivation).

Ionin et al's study used a picture-based comprehension task in which participants drew pictures based on written instructions containing definite descriptions. In anaphoric contexts, this task allows for both a maximal and a same interpretation of definite plurals. However, because participants generated their own drawings, their responses may reflect a preference—possibly shaped by L1 transfer of the Korean *ku*—rather than a clear indication of their comprehension. That is, participants might understand both interpretations but consistently choose one, making it difficult to determine whether the results reflect their interlanguage grammar or task-induced bias.

The present study is an approximate replication[4] of Ionin et al. (2012), with three key modifications designed to address the concerns outlined above:

1. Task redesign: We replaced the picture-based comprehension task with a picture-based acceptability judgment task (AJT). Rather than producing drawings, participants judged the acceptability of pre-made images representing both 'all' and 'same' interpretations of definite plurals. This design isolates comprehension from production and allows for a more direct assessment of interpretation. An example of the AJT format is shown in Figure 4.
2. Expanded conditions: We included non-anaphoric contexts to test whether participants can derive the 'all' interpretation of definite plurals without prior mention.
3. Updated data analysis: Given the ordinal nature of our outcome variable,

4 An approximate replication is a replication study that modifies two or more variables of the original study (Porte and McManus 2019: 78).

we employed cumulative link mixed models, following the recommendations of Taylor et al. (2023).

These changes aim to clarify whether L2 learners' responses reflect genuine comprehension or task-driven preferences. Table 2 compares the original study and the current replication.

Table 2. Comparison between the original study and the current replication

	Ionin et al. (2012)	The current replication
Task	Forced-choice production task Proficiency test Picture-based comprehension task	Forced-choice production task Proficiency test Picture-based 4-point AJT
Participant	21 L1-English 48 L1-Korean L2-English (24 advanced and 24 intermediate)	33 L1-English 31 L1-Korean L2-English (proficiency scores: 26–36)
Statistical analysis	One-way ANOVA	Cumulative link mixed model

3.2 Methodology

Participants were 33 L1-English (male = 8, female = 25) and 31 L1-Korean L2-English (male = 11, female = 20) speakers. The L1-English speakers were recruited from a large public university in the United States. Recruited in the same manner and by word of mouth, the Korean participants had not moved to any English-speaking country before the age of twelve. They were thus assumed to be comparable to the Korean speakers in the original study who received a large amount of EFL input in South Korea. The two participant groups in the current study differed in English proficiency ($F(1, 62) = 68.11$, $p < 0.001$, 95% CI [−5.28, −3.22], $d = 2.1$) with L1-English speakers performing better than L1-Korean L2-English speakers in the proficiency test. Participants' demographic information is reported in Table 3. Unlike the original study, L1-Korean L2-English participants were not divided into different levels of proficiency. In our replication, proficiency is treated as a continuous variable in the regression model in order to capture subtle effects of graduate increases in proficiency on L2 article comprehension (see Leal 2018 for theoretical and methodological rationales for treating proficiency as a continuous variable).

Table 3. Participants' information

	L1-English (n = 33)		L1-Korean L2-English (n = 31)	
	Mean (SD)	Range	Mean (SD)	Range
Age at Testing (AAT)	19(1.8)	18–26	30.3(8.1)	19–55
Age of Acquisition (AOA)	NA	NA	10.5(4.3)	4–26
Proficiency (max = 39)	36.6(1.3)	33–39	32.4(2.6)	26–36

Test materials for the forced-choice production task and the proficiency test were identical to those used by Ionin et al., described in the previous section.[5] Modifications were made only to the picture-based comprehension task. The original comprehension task was modified into a picture-based AJT on a 4-point Likert scale and an additional option of "I don't know." Participants completed the tasks in the same order as in the original study: the picture-based AJT, the proficiency test, and the forced-choice production task. In the picture-based AJT, participants saw figures identical to those in the original study, but with manipulated geometric drawings. They were asked to judge if the drawings are acceptable according to the instruction lines. The judgment task consisted of 30 test items with six items in five conditions. The conditions were manipulated via two main factors: context (anaphoric vs. non-anaphoric) and interpretation ('same' vs. 'all') plus acceptability (acceptable vs. unacceptable) as shown in Table 4. The target items were acceptable items while unacceptable items were added to counterbalance the design. Based on these modifications, it is possible to determine if L1-Korean L2-English participants' performance reflects the acquisition or L1-based preference of the 'same' and 'all' interpretations in English definite plurals as participants responded to both interpretations. The three tasks were administered via Qualtrics.

5 The proficiency test consists of 40 items. However, due to a technical issue of the test implementation (a typo in one test item), one item was removed from the analysis resulting in 39 items in total.

Table 4. Test conditions

Interpretation	Context	Acceptability	Example
same	anaphoric	acceptable	Figure 3
same	anaphoric	unacceptable	Figure 4
all	anaphoric	acceptable	Figure 5
all	non-anaphoric	acceptable	Figure 6
all	non-anaphoric	unacceptable	Figure 7

Each item comprised a four-line instruction. The first line introduced two objects in the format of 'Here are six X (object 1) and six Y (object 2).' The second, third, and fourth lines provided instructions on how drawings should be made onto the two objects, analogous to the instructions in the original study. In anaphoric contexts, one object was mentioned twice with *the* in the second mention. The other object was thus referred to only once without *the*. In non-anaphoric contexts, one object was mentioned twice without *the* in both mentions. *The* was thus used with the other object in its first and only mention. The first two conditions, i.e., *same/anaphoric/acceptable* and *same/anaphoric/unacceptable*, in Table 4 examined the 'same' interpretation in which the use of *the* in the second mention should refer to the same set of the object in the first mention. In the *same/anaphoric/acceptable* condition (Figure 3), since the second mention marked with *the* refers to the same entity in the first mention, the drawings are acceptable. In the *same/anaphoric/unacceptable* condition (Figure 4), the second mention mismatches the previously mentioned referent, conflicting with the 'same' interpretation and leading to unacceptability. The other three conditions, i.e., *all/anaphoric/acceptable*, *all/non-anaphoric/acceptable*, and *all/non-anaphoric/unacceptable*, tested the 'all' interpretation in which the use of *the* in the second mention can refer maximally. In the *all/anaphoric/acceptable* condition (Figure 5) which is an anaphoric context, drawings in the second mention were made onto the maximal set of one object yielding the 'all' interpretation and acceptability. Central to our study is this condition where Ionin et al. (2012) found that L1-Korean learners of English rarely showed the 'all' interpretation while L1-English speakers provided both interpretations. In the *all/non-anaphoric/acceptable* condition (Figure 6), *the* in a non-anaphoric context referred to all members of an object also leading to the 'all' interpretation and acceptability. In the *all/non-anaphoric/unacceptable* condition (Figure 7), drawings

were not made to the maximal set of the object with *the* in a non-anaphoric context. The incongruence between anaphoricity and interpretation resulted in unacceptability. We acknowledge that the two non-anaphoric conditions can be considered anaphoric due to the presence of the first introductory line setting the context of each item. For example, in Figure 6, where *the apples* were not mentioned in any numbered line, participants could resort to the introductory line where *six apples* were introduced. However, we decided to maintain the non-anaphoric conditions as is to maximize the similarities between the original and present studies.

The experiment was created in a 2×2×2 design with *Context*, *Interpretation*, and *Acceptability* as independent variables. The context factor has two levels (anaphoric vs. nonanaphoric), the interpretation factor has two levels (same vs. all) and the acceptability factor has two levels (acceptable vs. unacceptable). In a fully crossed factorial design, our experiment would have yielded eight conditions. But our experiment employed a partially crossed design which included only five test conditions shown in Table 4. Our rationale for omitting three conditions is the following.

Items in the *all/anaphoric/unacceptable* condition, which were not included, would be similar to items in the *same/anaphoric/unacceptable* condition in Figure 4. That is, the geometric drawings in the second mention would be on only some objects that do not match those in the first mention. The inclusion of both conditions would outnumber items in other conditions. That is, there would be 36 items in total with 12 identical items. Furthermore, it is not possible to create a test condition with the 'same' interpretation in the non-anaphoric context. This explains why two more conditions, namely, *same/non-anaphoric/acceptable* and *same/non-anaphoric/unacceptable*, were absent from the experimental design.

Here are six cups and six balls
1. Draw triangles around two balls.
2. Draw a circle around one cup.
3. Draw arrows below the balls.

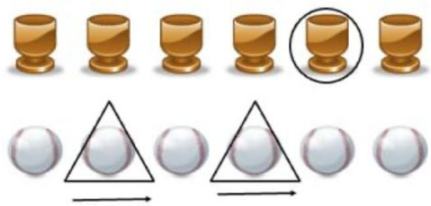

Figure 3. Example of the same/anaphoric/acceptable condition

(8)
Here are six cars and six books.
1. Draw lines above two cars.
2. Draw arrows above four books.
3. Draw circles around the cars.

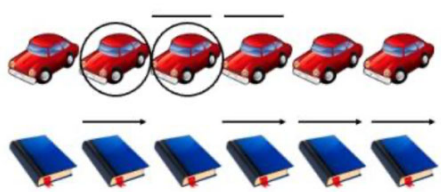

Figure 4. Example of the same/anaphoric/unacceptable condition

Here are six pencils and six bananas.
1. Draw arrows above two pencils.
2. Draw squares around four bananas.
3. Draw circles around the pencils.

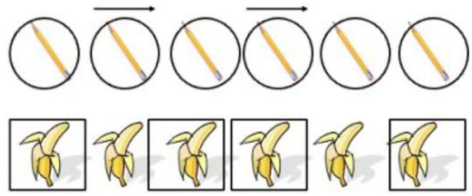

Figure 5. Example of the all/anaphoric/acceptable condition

Here are six books and six apples.
1. Draw circles around two books.
2. Draw arrows above the apples.
3. Draw stars on two books.

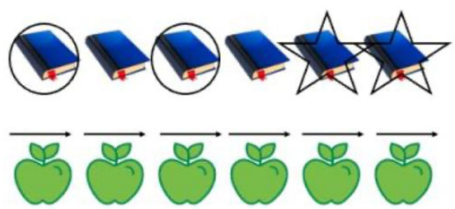

Figure 6. Example of the all/non-anaphoric/acceptable condition

Here are six pens and six chairs.
1. Draw arrows below two pens.
2. Draw a square around a pen.
3. Draw circles around the chairs.

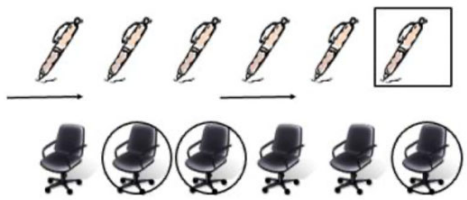

Figure 7. Example of the all/non-anaphoric/unacceptable condition

4. Results

4.1 Descriptive analysis

Firstly, one "I don't know" response, which accounted for 0.05% of the data set, was identified and excluded from further analysis.[6] Descriptive data from the forced-choice production task seemed to indicate that L1-Korean participants were able to provide *the*, *that*, and *a* correctly. However, since our replication's focus is on the picture-based AJT, only the results from this task are reported below.

Table 5 reports an overview of mean (M) and standard deviation (SD) ratings by *condition* and *language group* which is also visualized as Figure 8. Among L1-English participants, ratings were markedly higher for *all/anaphoric* items than for *same/anaphoric* items, suggesting a clear preference for the 'all' interpretation over the 'same' interpretation for definite plurals in anaphoric contexts. In contrast, L1-Korean participants showed minimal difference between the two anaphoric conditions with the 'same' interpretation slightly higher than the 'all' interpretation.

6 An anonymous reviewer raised a concern about the issue of indeterminacy in L2 knowledge. As discussed in Ellis (2004: 260), "... indeterminacy arises when learners lack implicit knowledge of a feature". Indeterminacy is generally indicated by ratings hovering around the middle point of the rating scale. Following recent experimental L2 studies, we included the "I don't know" option which is considered as evidence of indeterminacy in L2 knowledge (Yuan and Zhao 2010).

Ratings for *all/anaphoric* and *all/non-anaphoric* items in L1-English participants were almost identical. However, for L1-Korean participants, *all/non-anaphoric* items received higher acceptability. Both of the unacceptable conditions received lower ratings than all three acceptable conditions in both L1-English and L2-English speakers.

Table 5. Means and standard deviations of ratings by condition and language group

Condition	L1-English M (SD)	L1-Korean L2-English M (SD)
same/anaphoric/acceptable	2.70 (0.71)	3.33 (0.61)
same/anaphoric/unacceptable	2.04(0.76)	2.41(0.95)
all/anaphoric/acceptable	3.95 (0.11)	3.26 (0.77)
all/non-anaphoric/acceptable	3.96 (0.09)	3.70 (0.51)
all/non-anaphoric/unacceptable	2.15(0.77)	2.57(0.91)

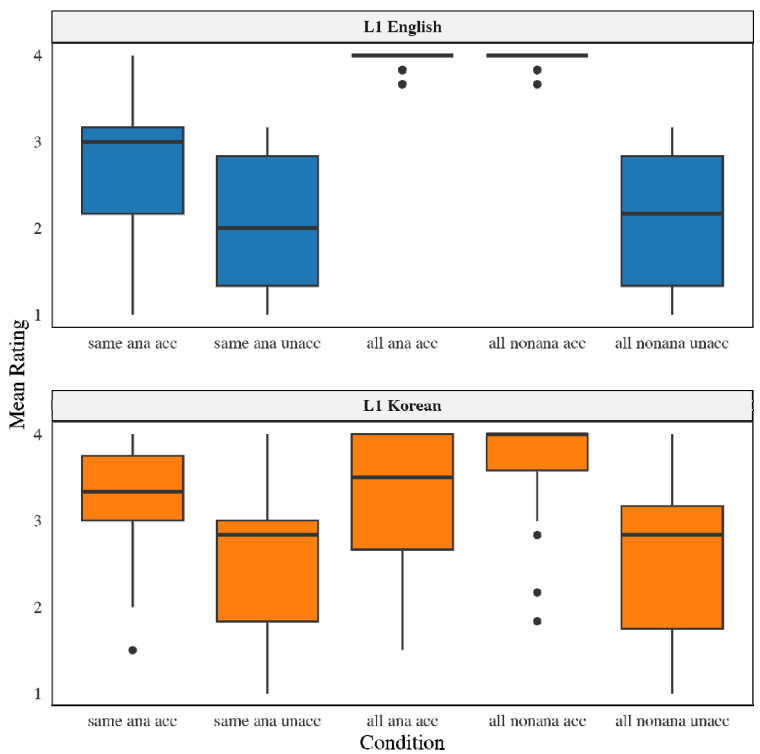

Figure 8. Box plot of the ratings by condition and language group

4.2 Inferential analysis

In order to better understand the effects of *anaphoricity* (anaphoric vs. non-anaphoric contexts), *interpretation* (same vs. all), *language group* (L1 vs. L2), and *acceptability* (acceptable vs. unacceptable target sentences) on participants' ratings, we conducted a series of statistical analyses using cumulative link mixed models (CLMMs). These models are preferred over ANOVA, which was used in the original study, and linear regression models with a *z*-transformation of rating scores since the outcome variable is considered ordinal and thus not continuous, as suggested by Taylor et al. (2023). All models were fit using the *clmm* function in the *ordinal* package (Christensen 2023) in R Studio (R Core Team 2023), treating ratings as the response variable and random intercepts for *participant* and *item*. For each model, we report statistically significant main effects and interactions where relevant, as well as 95% confidence intervals (CIs) and effect sizes via Cohen's D (*d*), following Plonsky and Oswald (2014), who provide a guide on how effect sizes should be interpreted in L2 research.[7]

First, we built a CLMM to compare between acceptable and unacceptable items to examine whether both groups rated them differently. The model demonstrated an effect of *acceptability* but not *language group*. We divided the data set by *language groups* to avoid unnecessary comparisons which could emerge from pairwise comparisons of an interaction between *acceptability* and *language group* (e.g., acceptable items in L1-English speakers vs. unacceptable items in L1-Korean L2-English speakers). The same pattern was observed in both L1-English and L2-English models. That is, unacceptable items were significantly less acceptable than acceptable items (L1-English: $\beta = -6.288$, SE = 1.08, $z = -5.85$, $p < .001$, 95% CI [−8.4 ,−4.18], $d = 2.25$; L2-English: $\beta = -2.37$, SE = 0.28, z −8.61 3.52, $p < .001$, 95% CI [−2.91, −1.83], d =1.2). The effect sizes were large in both groups. Since unacceptable stimuli were added as control items to counterbalance the design, they were later removed from further analyses. Thus, effects of acceptability are not presented or discussed further.

Reported in Table 6, our main model examined the effects of *anaphoricity*

7 Regression tables of all models to be reported below can be found in Supplementary Materials on https://osf.io/gu8yn/overview?view_only=32ab94a8d9504b9c9987a99e258ccca7.

(anaphoric vs. non-anaphoric) and *interpretation* ('all' vs. 'same') across both language groups (L1 vs. L2), including two-way interactions for the *language group* predictor (i.e., *anaphoricity* × *language group* and *interpretation* × *language group*). The reference level is the 'all' interpretation for *interpretation*, L1-English speakers for *language group*, and the non-anaphoric context for *anaphoricity*. Table 6 reports the results from the model. First, *proficiency* was also included as a fixed factor with no interaction with the other fixed factors, but its effect did not emerge. The model revealed significant effects of *interpretation* and *language group*. However, the two interactions were found to be statistically significant, overshadowing the effects of the two fixed factors. We thus conducted further analyses to disentangle the interactions.

Table 6. Regression analysis of the main model

Predictor	β	SE	z	p	95% CI
Interpretation (all vs. same)	−4.827	0.426	−11.34	< 0.001	−5.66, −3.99
Language group (L1 vs. L2 English)	−3.089	0.552	−5.594	< 0.001	−4.17, −2.01
Anaphoricity (anaphoric vs. non-anaphoric)	0.43	0.568	0.757	0.449	−0.68, 1.54
Proficiency	0.082	0.07	1.116	0.243	−0.06, 0.22
Interpretation × Language group	4.919	0.454	10.833	< 0.001	4.03, 5.81
Anaphoricity × Language group	1.248	0.606	2.06	0.039	0.06, 2.43

To further explore how these effects manifested within each group, the data was split into several subsets to isolate the effects of the different conditions for the two language groups. We began by dividing the data according to *language group* and conducted analyses of ratings in anaphoric contexts. The first model on *interpretation* tested how L1-English participants interpreted anaphoric contexts by comparing their ratings for the 'all' vs. 'same' interpretations. The model showed a significant main effect for *interpretation* (β = −6.41, SE = 0.63, z = −10.13, p < .001, 95% CI [−7.65, −5.17], d = −2.7). These results indicate that L1-English speakers preferred the 'all' interpretation over the 'same' interpretation in anaphoric contexts with a large effect size. Similarly, the second *interpretation* model tested L2-English participants' ratings for the two interpretations in anaphoric contexts. The model revealed no effect of

interpretation (β = 0.05, SE = 0.23, z = 0.23, p > .05, 95% CI [−0.39, 0.5], d = 0.04), meaning that L2-English speakers did not differentiate between the two interpretations in anaphoric contexts with a negligible effect size.

Next, we took a different subset of the data to compare the L1 and L2 groups to test the effects of *anaphoricity* (anaphoric vs. nonanaphoric contexts) for the 'all' interpretation. Similar to the previous analyses, two models were built: one for L1-English speakers and the other for L2-English speakers. The L1 model showed no effect of *anaphoricity* (β = 0.41, SE = 0.6, z = 0.69, p > 0.05, 95% CI [−0.76, 1.58], d = −0.1). That is, L1-English participants rated items with the 'all' interpretation similarly regardless of anaphoricity with a negligible effect size. The L2 model, on the other hand, revealed that L2-English participants' interpretation of 'all' was affected by *anaphoricity* (β = 1.99, SE = 0.3, z = 6.67, p < .001, 95% CI [1.41, 2.58], d = −0.65). That is, for L2-English speakers, the 'all' interpretation was more acceptable in non-anaphoric contexts than anaphoric ones with a small effect size. English proficiency was also added to all L2 models. However, it did not reach statistical significance in any model. The models reported did not include proficiency since its inclusion did not significantly improve the models.

In order to determine how the three primary conditions of interest (i.e., *same/anaphoric, all/anaphoric*, and *all/non-anaphoric*) differed between the two language groups, we further divided the data into three subsets for each of these conditions in order to test how *language group* affected the ratings for each condition. The first model compared L1 and L2 speakers' ratings for the 'same' interpretation in anaphoric contexts (*same/anaphoric*) and showed a main effect of *language group* (β = 2.29, SE = 0.62, z = 3.7, p < .001, 95% CI [1.08, 3.5], d = −0.92). This indicates that L1-Korean L2-English participants rated the *same/anaphoric* condition as more acceptable than L1-English participants with a relatively large effect size. Likewise, the second model compared L1 and L2 speakers' ratings for the 'all' interpretation in anaphoric contexts (*all/anaphoric*). In this model, *language group* emerged as a significant factor (β = −4.58, SE = 0.92, z = −4.99, p < .001, 95% CI [−6.39, −2.78], d = 1.28). To explain, L1-English speakers rated the *all/anaphoric* condition higher than L1-Korean L2-English speakers did with a large effect size. Finally, the last model compared L1 and L2 speakers' ratings for the 'all' interpretation in non-anaphoric. The model revealed a main effect of *language group* (β = −2.02, SE = 0.89, z = −2.26, p < .05, 95% CI [−3.77, −0.27], d = 0.65). This means that L1-English speakers provided

higher ratings than L1-Korean L2-English speakers in the *all/non-anaphoric* condition with a medium effect size.

5. Discussion and conclusion

Our replication set out to examine whether the 'all' interpretation of English definite plurals is available to L1-Korean L2-English speakers. In the original study by Ionin et al. (2012), L1-Korean speakers demonstrated a strong preference for the 'same' interpretation (97% of the time) while L1-English speakers allowed both 'all' and 'same' interpretations (41% and 58%, respectively). The authors attributed such non-targetlike behavior to transfer from the semantics of demonstratives in L1 Korean. However, the design of the original study did not allow us to see if both interpretations are available to L1-Korean speakers : the large number of 'same' responses in L1-Korean L2-English speakers might reflect their preference but not their comprehension of English definite plurals. We thus modified the task by using a picture-based AJT. With this design, participants were able to show whether their grammars contain both 'same' and 'all' interpretations for definite plurals.

Overall, the results of our picture-based AJT point to a partial confirmation of the original study. Compared to L1-English speakers, L1-Korean L2-English speakers are more likely to allow the 'same' interpretation for definite plurals in anaphoric contexts (M = 2.70 for the L1-English group vs. M = 3.31 for the L2-English group). This finding supports Ionin et al.'s claim that L1-Korean L2-English speakers differ from L1-English speakers in their interpretation for definite plurals. However, there are two findings of this replication study that differ from the original study. First, unlike the original study, our study found that L1-English speakers rated the 'all' interpretation significantly higher than the 'same' interpretation. The original study showed that L1-English speakers did not have a strong preference for one interpretation over the other for definite plurals (58% for the 'same' interpretation and 41% for the 'all' interpretation). However, in our replication study, L1-English speakers preferred the 'all' interpretation over the 'same' interpretation. The difference in their ratings for 'all' versus 'same' is statistically significant (M = 3.95 vs. M = 2.7). So, why did L1-English speakers in the replication study give a relatively low rating (M = 2.7 out of 4) to the 'same' interpretation? A possible explanation could

be that L1-English speakers might have come up with sentences that are more natural than the target sentences presented in the picture-based AJT. For example, participants were asked to judge the target sentence "Draw arrows below the balls" in the context of "Draw triangles around two balls. Draw a circle around one cup" as in Figure 3 (repeated here as Figure 9).

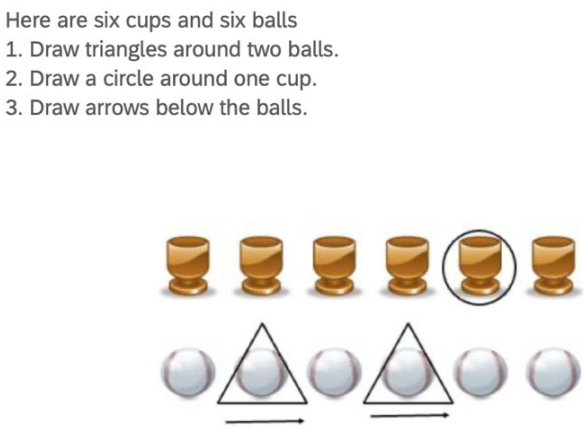

Figure 9. Example of the same/anaphoric/acceptable condition

The preference for demonstrative NPs among L1-English speakers may reflect a broader pragmatic sensitivity to contextual appropriateness. It is possible that L1-English speakers gave low ratings to target sentences like "Draw arrows below the balls" which are ambiguous between two interpretations ('same' and 'all') because unambiguous sentences containing demonstrative NPs like "Draw arrows below these/those two balls" would be considered contextually more appropriate. In natural discourse, speakers often choose expressions that minimize ambiguity and maximize referential clarity (Grice 1975). Demonstrative NPs like "these two balls" explicitly anchor the referent in the visual or discourse context, whereas "the balls" leaves room for interpretive ambiguity. Thus, the lower ratings for the 'same' interpretation may not indicate the intended meaning was not available, but rather a judgment that the sentence was suboptimal given the visual context. However, in the picture comprehension task used in the original study, participants did not have to evaluate the acceptability/appropriateness of test sentences or pictures. They were asked to show

by drawing how they interpreted the test sentences, similar to what is asked of participants in an ask-out task (Goodluck 1996). We believe that this methodological difference yielded different results from L1-English speaker controls between the original study and the replication study.

Differential L1-speakers' judgment patterns in the original study and in our study as well as variability in L1-speakers' ratings in our study offer insights into the issue of linguistic indeterminacy (variable interpretation or judgments). In L2 research, indeterminacy is typically considered an L2-specific phenomenon. However, as shown in Ayoun (2022), the grammars of L1 speakers can also display indeterminacy (see also Dalrymple and Kaplan 2000). Considering the complexity of the semantics of plural definites whose interpretations depend on multiple factors including pragmatic and contextual factors, indeterminacy is an inherent property of plural definites which is necessarily reflected in L1 grammars as shown in our data from L1 speakers.[8]

The second and the most important finding of our replication study that is different from the original study is L1-Korean L2-English speakers' interpretation of definite plurals in the anaphoric condition. While the findings of the original study indicated that L1-Korean L2-English speakers predominantly allowed only the 'same' interpretation (97%), our replication data showed that L1-Korean L2-English speakers allowed both 'all' and 'same' interpretations at the same degree (i.e., no statistically significantly different preference). This indicates that L2 speakers know that definite plurals in anaphoric contexts can have two meanings. Unlike the original study where only one interpretation could be selected, our modification presented both interpretations and allowed us to measure the difference in ratings between these two conditions. The lack of statistically significant differences between 'all' and 'same' interpretations in anaphoric contexts suggests that L1-Korean speakers possess both interpretations in their L2 grammars. However, because the 'acceptable' items in anaphoric contexts featured drawings where both interpretations were plausible, it is not immediately clear whether our results reflect genuine comprehension of the semantics of definite plurals or merely the degree to which participants found the pictures contextually permissible.[9] The presence of 'unacceptable' control items designed to counterbalance the 'acceptable' target items may have led L1-Korean

8 We thank an anonymous reviewer for pointing out this important issue.

9 We thank an anonymous reviewer for pointing out this important issue

participants to rate these items similarly across both interpretations simply by virtue of not being clearly infelicitous. We believe that our target items provided enough information to avoid infelicity without contextually biasing participants towards one interpretation or another. Given the results for L1-English participants, results for L1-Korean participants may likewise reflect pragmatic sensitivity rather than interpretive availability, further motivating the need for future work to explore the way contextual appropriateness affects the availability of different interpretations, a key requirement for linking hypotheses in experimental semantics (Harris 2021). This experimental change could explain the difference in findings between the original and replication studies.

These findings suggest that L2 speakers have the targetlike representations for definite plurals (i.e., definite plurals are ambiguous between the 'same' and 'all' interpretations in anaphoric contexts). However, while L1 speakers tend to prefer the 'all' interpretation over the 'same' interpretation for definite plurals in anaphoric contexts, L2 speakers show no preference. We speculate that L1-English speakers' response patterns might be due to the fact that in evaluating some target sentences, they might have come up with better expressions for the 'same' interpretation than definite plurals (e.g., these/those two balls vs. the balls). While L1 speakers were able to do so while performing the picture-based AJT, L2 speakers were unable to (or did not) conjure up alternative expressions for the target sentences since L2 processing is arguably more effortful than L1 processing (Hopp 2022). We acknowledge that via our offline judgment task, it is not possible to determine whether the differences between L1 and L2 speakers' judgment patterns are at the representational level or at the processing level.

A promising direction for future research would be the incorporation of online processing measures to complement the findings from our offline AJT. While our AJT design allowed us to isolate interpretation from production, it does not capture the real-time cognitive processes involved in interpreting definite plurals. Online methods such as eye-tracking, self-paced reading, or event-related potentials could reveal whether L2 learners momentarily entertain both 'same' and 'all' interpretations before settling on one, or whether they experience processing difficulty when encountering ambiguous definite descriptions. This distinction is crucial, as prior research suggests that L2 speakers show differences between online processing and offline judgment patterns (see Orfitelli and Polinsky 2017 for an overview of studies

showing the discrepancy between online and offline performance in L2). Future studies could therefore further elucidate the relationship between representation and integration of ambiguous definite descriptions by incorporating online processing techniques to more fully characterize the interplay between representation, preference, and processing in L2 learners' interpretation of definite plurals.

As mentioned earlier, the non-anaphoric context in our study might not be truly non-anaphoric. We intended to minimize changes to the original study by keeping the first line of the instructions which introduced objects such as "Here are six books and six apples" in Figure 6 (repeated below as Figure 10). Even though definite plurals (e.g., "Draw arrows above the apples") appeared without prior mention except for the introductory line, participants could interpret "the apples" as referring to "six apples" mentioned in the introduction due to the presence of "Here" which could introduce anaphoric support for definite plurals. Future replication could consider removing this line to test the 'all' interpretation in non-anaphoric contexts. The introductory line might be unnecessary since all of the objects are common and hence need no label. In addition, even though the introductory line is not included, it is likely that the visual presentation of objects forms anaphoricity in all cases. That is, with pictures (e.g., six books and six apples in Figure 10), all items can be interpreted as anaphoric. Including this line also established the first mention and made the 'all' interpretation plausible even in the 'same' interpretation conditions. In the 'same' interpretation conditions, there are two competing prior mentions, i.e., one in the introductory line and the other before the definite plurals. Even though the 'same' interpretation can be derived due to the proximity between the target line and the line that mentioned the same object, participants could still refer back to the introductory line to derive the 'all' interpretation. This interpretation was thus available in all conditions.

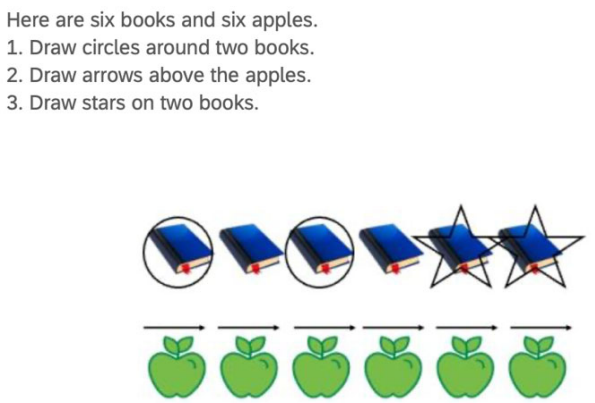

Figure 10. Example of the all/non-anaphoric/acceptable condition

In conclusion, this present approximate replication of Ionin et al. (2012) with modifications to the methodology aimed to investigate whether L1-Korean L2-English speakers are able to compute both 'all' and 'same' interpretations for definite plurals in anaphoric contexts. Differential response patterns between L1-English speakers and L1-Korean L2-English speakers in the original study were attributed to transfer from the semantics of demonstratives in Korean, as Ionin et al. argued. We do not dispute the claim about L1 transfer. Our replication aimed to further investigate whether L1-Korean L2-English speakers' strong preference for the 'same' interpretation in the original study is a matter of preference or inability to compute the 'all' interpretation for definite plurals. With a modified picture-based task, this replication showed that both interpretations are available to L1-Korean L2-English speakers. Specifically, while L2 speakers allowed both 'all' and 'same' interpretations at the same degree, L1 speakers gave lower ratings to the 'same' interpretation in the same/anaphoric condition. We argue that this observed difference between L1 and L2 speakers indicates that while L2 speakers may have acquired the semantics of definite plurals, they have not yet fully developed native-like pragmatic sensitivity.[10]

10 We thank an anonymous reviewer for their comment on this issue.

References

Ayoun, Dalila. 2022. Indeterminacy in L1 French grammars: The case of gender and number agreement. *Journal of French Language Studies* 32(3): 327–359.

Cho, Hyesun. 1999. Interpretation and function of the Korean demonstrative *ku*. *Studies in Modern Grammar* 18: 71–90.

Cho, Jacee. 2017. The acquisition of different types of definite noun phrases in L2-English. *International Journal of Bilingualism* 21(3): 367–382.

Christensen, Rune H. B. 2023. *Ordinal: Regression models for ordinal data*. R package version 2023.12-4.

Dalrymple, Mary and Ronald M. Kaplan. 2000. Feature indeterminacy and feature resolution. *Language* 76(4): 759–798.

Ellis, Rod. 2004. The definition and measurement of L2 explicit knowledge. *Language Learning* 54(2): 227–275.

Feng, Shuo. 2019. The acquisition of English definite noun phrases by Mandarin Chinese speakers. *Studies in Second Language Acquisition* 41(4): 881–896.

Goodluck, Helen. 1996. The act-out task. In Dana McDaniel, Helen S. Cairns, and Cecile McKee (eds.), *Methods for assessing children's syntax*, 147–162. Cambridge, MA: The MIT Press.

Grice, H. P. 1975. Logic and conversation. In Peter Cole and Jerry L. Morgan (eds.), *Syntax and semantics: Vol. 3. Speech acts*, 41–58. New York, NY: Academic Press.

Harris, Jesse. 2021. Acceptability judgments at the syntax-semantics interface. In Grant Goodall (ed.), *The Cambridge handbook of experimental syntax*, 341–370. Cambridge, UK: Cambridge University Press.

Hawkins, John. A. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Routledge.

Hawkins, John. A. 1991. On (in)definite articles: Implicatures and (un)grammaticality prediction. *Journal of Linguistics* 27(2): 405–442.

Heim, Irene. 1991. Artikel und definitheit. In Arnim von Stechow and Dieter Wunderlich (eds.), *Semantik: Ein internationales Handbuch der Zeitgenössischen forschung*. Berlin: De Gruyter.

Hopp, Holger. 2022. Second language sentence processing. *Annual Review of Linguistics* 8: 235–256.

Ionin, Tania, Heejeong Ko, and Kenneth Wexler. 2004. Article semantics in L2 acquisition: The role of specificity. *Language Acquisition* 12(1): 3–69.

Ionin, Tania, Sea Hee Choi, and Qiufen Liu. 2021. Knowledge of indefinite articles in L2-English: Online vs. offline performance. *Second Language Research* 37(1): 121–160.

Ionin, Tania, Soondo Baek, Eunah Kim, Heejeong Ko, and Kenneth Wexler. 2012. That's not so different from the: Definite and demonstrative descriptions in second language acquisition. *Second Language Research* 28(1): 69–101.

Jiang, Nan, Eugenia Novokshanova, Kyoko Masuda, and Xin Wang. 2011. Morphological congruency and the acquisition of L2 morphemes. *Language Learning* 61(3): 940–967.

Lardiere, Donna. 2008. Feature assembly in second language acquisition. In Juana Liceras, Helmut Zobl, and Helen Goodluck (eds.), *The role of formal features in second language acquisition*, 107–140. New York, NY: Lawrence Erlbaum.

Lardiere, Donna. 2009. Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research* 25(2): 173–227.

Leal, Tania. 2018. Data analysis and sampling: Methodological issues concerning proficiency in SLA research. In Aarnes Gudmestad and Amanda Edmonds (eds.), *Critical reflections on data in second language acquisition*, 63–88. Amsterdam: John Benjamins Publishing Company.

Lee, Sun-Hee and Jae-young Song. 2010. Annotating Korean demonstratives. In *Proceedings of the Fourth Linguistic Annotation Workshop*, 162–165. Uppsala, Association for Computational Linguistics.

McManus, Kevin. 2022. Replication research in instructed SLA. In Laura Gurzynski-Weiss and Youjin Kim (eds.). *Research methods in instructed second language acquisition*. Amsterdam: John Benjamins Publishing Company.

Ortifelli, Robyn and Maria Polinsky. 2017. When performance masquerades as comprehension: Grammaticality judgments in experiments with non-native speakers. In Mikhali Kopotev, Olga Lyashevskaya, and Arto Mustajoki (eds.), *Quantitative approaches to the Russian language*, 197–214. London, UK: Routledge.

Plonsky, Luke. 2024. Study quality as an intellectual and ethical imperative: A proposed framework. *Annual Review of Applied Linguistics* 44: 1–15.

Plonsky, Luke and Frederick L. Oswald. 2014. How big is "big"? Interpreting effect sizes in L2 research. *Language Learning* 64(4): 878–912.

Porte, Graeme and Kevin McManus. 2019. *Doing replication research in applied linguistics*. New York, NY: Routledge.

R Core Team. 2023. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Sharvy, Richard. 1980. A more general theory of definite descriptions. *The Philosophical Review* 89(4): 607–624.

Snape, Neal. 2018. Definite generic vs. definite unique in L2 acquisition. *Journal of the European Second Language Association* 2(1): 83–95.

Snape, Neal, Hironobu Hosoi, and Mari Umeda. 2023. Second language processing of English definite noun phrases by Spanish speakers and Japanese speakers. *Journal of Second Language Studies* 6(2): 319–347.

Sohn, Ho-min. 1999. *The Korean language*. Cambridge, UK: Cambridge University Press.

Taylor, Jack E., Guillaume A. Rousselet, Christoph Scheepers, and Sara C. Sereno. 2023. Rating norms should be calculated from cumulative link mixed effects models. *Behavior Research*

Methods 55(5): 2175–2196.
Trenkic, Danijela and Nattama Pongpairoj. 2013. Referent salience affects second language article use. *Bilingualism: Language and Cognition* 16(1): 152–166.
Tuniyan, Elina. 2018. *Second language acquisition of definiteness: A feature-based contrastive approach to second language learnability*. PhD Dissertation. University of Southampton.
Yuan, Boping and Yang Zhao. 2011. Asymmetric syntactic and thematic reconfigurations in English speakers' L2 Chinese resultative compound constructions. *International Journal of Bilingualism* 15(1): 38–55.
Zhang, Lulu. 2021. The influence of first language at the semantics–pragmatic interface: Evidence from definite and demonstrative determiners in L2 English. *Linguistic Approaches to Bilingualism* 11(3): 368–388.

Vatcharit Chantajinda
Graduate Student
University of Wisconsin–Madison
600 N. Park Street
Madison, WI 53705 USA
Email: chantajinda@wisc.edu

Hyun Bae
Lecturer
University of Southern California
3415 South Figueroa Street
Los Angeles, CA 90089 USA
Email: hyunbae@usc.edu

Joseph Moran
Graduate Student
University of Wisconsin–Madison
600 N. Park Street
Madison, WI 53705 USA
Email: jmoran8@wisc.edu

Jacee Cho
Associate Professor
University of Wisconsin–Madison
600 N. Park Street
Madison, WI 53705 USA
Email: jacee.cho@wisc.edu

Received: 2026. 01. 12.
Revised: 2026. 02. 24.
Accepted: 2026. 02. 26.