

Adjective Placement in Korean Numeral–Classifier Constructions: A Corpus Study of Sortal Classifiers and Measure Expressions

Kitaek Kim¹, Kum-Jeong Joo^{2†}, Min-Chang Sung^{3†}, & Yunchuan Chen⁴

¹Seoul National University, South Korea, ²Induk University, South Korea,

³Gyeongin National University of Education, South Korea,

⁴Duke University, United States

ABSTRACT

This study examines the placement of adjectives in Korean numeral–classifier constructions, using Mandarin Chinese as a baseline for comparison. Drawing on the morphologically annotated NIKL POS-tagged corpus as the primary dataset and koTenTen concordances as a qualitative follow-up source, we examine where adjectives occur relative to numerals and classifiers and whether corpus evidence allows adjectives to intervene between the numeral and the classifier. The Korean data show that postnominal numeral–classifier constructions are substantially more frequent than prenominal constructions, and that adjective-bearing numeral–classifier strings are rare overall. In contrast, data from the Mandarin CCL corpus show that adjectives can occur between the numeral and the classifier, particularly when the classifier has a more noun-like or measurement interpretation. In the Korean corpora examined, no cases were found in which an adjective appears between a numeral and a highly grammaticalized sortal classifier; the only apparent exceptions involve a single noun-like measure expression, *swul* ‘spoon’. These findings suggest that Korean strongly preserves numeral–classifier adjacency, and that apparent exceptions arise primarily from elements that retain nominal properties rather than from canonical sortal classifiers.

Keywords: numeral–classifier constructions; sortal classifiers; measure expressions; adjective placement; numeral–classifier adjacency; Korean; Mandarin Chinese; corpus linguistics

1. Introduction

Numeral–classifier constructions have received particular attention in discussions

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2023S1A5A8080501).

† Corresponding author: jookj@induk.ac.kr, mcsung@ginue.ac.kr



of how nominal meaning is mapped onto morphosyntax (Borer, 2005; Rothstein, 2017). This is because they bear directly on the grammatical encoding of counting and individuation and, more broadly, on the mass-count distinction across languages. In number-marking languages such as English, the contrast is visible across multiple diagnostics: count nouns typically pluralize and combine directly with numerals (e.g., *two books*), whereas mass nouns resist plural marking in their default interpretations and require measure expressions (e.g., *two glasses of water*). In numeral-classifier languages such as Mandarin Chinese, by contrast, numerals ordinarily cannot modify nouns directly; instead, a unit¹⁾ expression must intervene. Chierchia (1998) proposes that in some classifier languages, bare nouns are argumental or kind-denoting and number-neutral, and that counting is grammatically mediated by unit expressions.

Korean is widely described as a (generalized) classifier language, permitting bare nouns with number-neutral interpretations and exhibiting complex interactions among classifiers, plural marking, and nominal structure (Kim & Melchin, 2018; Park, 2022). Korean numeral-unit expressions appear in multiple surface orders, including a postnominal N-Num-Unit construction and a prenominal Num-Unit_{GEN}-N construction.

(1) a. N-Num-Unit

chayk sey kwen
book three Cl
'three books'

b. Num-Unit_{GEN}-N

sey kwen-uy chayk
three Cl-GEN book
'three books'

1) The element occurring after the numeral in numeral-classifier constructions has been described using various terms in the literature (e.g., classifier, massifier, measure word). Following work that adopts *unit word(s)* as a cover term for this position (e.g., Börjars et al., 2018), we use Unit (also *unit word* / *unit expression*) to refer collectively to both (i) highly grammaticalized sortal classifiers and (ii) more lexical, noun-like measure expressions. This choice is motivated by our empirical aim—tracking the distribution of adjectival material relative to the numeral and the post-numeral Unit position across languages—while remaining agnostic about the precise syntactic category status of individual items. In the title and abstract, however, we retain the term classifier in a broader sense, following common usage in the typological literature on numeral-classifier languages.

Beyond word order, the Korean syntax literature often posits an especially tight dependency between numerals and classifiers: the numeral and classifier form a close complex, and it is commonly claimed that no material can intervene between them (Joo, 2017; K.-S. Kim, 2010). One manifestation of this closeness is that adjectival material between the numeral and the unit word results in degraded acceptability, as illustrated in (2).

- (2) *maykcwu sey ppalkan pyeng
 beer three red bottle
 'three red bottles of beer'

(K.-S. Kim, 2010, p. 357, (20))

At the same time, many speakers judge certain numeral–unit strings with a size adjective to be relatively natural when the unit expression is a measure expression (3a), while parallel strings with prototypical sortal classifiers are completely unacceptable (3b).

- (3) a. chamkilum twu khun swul
 sesame oil two big spoon
 'two big spoons of sesame oil'
 b. *sakwa twu khun kay
 apple two big Cl²⁾
 'two big apples'

Importantly, the intervention case in (3a) can be analyzed as involving a conventionalized measure expression such as *khun-swul* or *cakun-swul* 'big spoon/small spoon', where the adjective may plausibly be part of a lexicalized unit noun rather than a productive modifier intervening inside a numeral–classifier complex. This possibility highlights a broader point emphasized in previous work on Korean classifiers: unit expressions vary along a continuum between nounhood and classifierhood, and noun-like measure expressions may behave differently from highly grammaticalized classifiers (Choi, 2011; Kim & Park, 2017). For example, Choi (2011) distinguishes “genuine” classifiers from ordinary nouns that may be

2) Abbreviations: Adj=adjective; N=noun; Num=numeral; Cl=classifier; Cl_{GEN}=classifier with a genitive marker; Unit=unit word/unit expression; Unit_{GEN}= unit word with a genitive marker; GEN=genitive; TOP=topic; PAST=past; DECL=declarative.

used as classifiers (p. 525). Genuine classifiers such as *kay*, *myeng*, and *kwen* lack lexical descriptive content and behave as clitic-like elements attaching to numerals, whereas ordinary nouns such as *pyeng* ‘bottle’ or *can* ‘glass’ retain lexical properties and therefore behave differently from highly grammaticalized genuine classifiers.

Mandarin provides a well-studied comparison point. In addition to the canonical Num–Unit–Adj–N order, Mandarin allows a pattern in which the adjective appears between the numeral and the unit word (Num–Adj–Unit–N). Notably, this marked pattern is typically associated with size adjectives and noun-like measure expressions (often termed *massifiers*, *mensural classifiers*, or *measure words*, as distinct from sortal or count classifiers; in this study, they will henceforth be referred to as measure expressions) (Cheng & Sybesma, 1998, 1999; Jiang et al., 2022). See (4a–b) for examples.

(4) a. Num–Adj–[measure]–N

yi	da	ping	jiu
one	big	Cl-bottle	liquor

b. *Num–Adj–[sortal]–N

yi	da	ge	ren
one	big	Cl-individual	person

(Li et al., 2008, p. 254, (3a–b))

As illustrated in (4a–b), when a noun-like measure expression is used with the uncountable noun *jiu* ‘liquor,’ an adjective may intervene between the numeral and the classifier/unit word, yielding the Num–Adj–Unit–N sequence. However, this order is far more restricted when a prototypical sortal classifier is used, as shown in (4b).

Building on this Mandarin–Korean contrast, the present study uses Mandarin as a baseline diagnostic and centers the empirical analysis on Korean. For Mandarin, we conduct a targeted calibration search in the Peking University Center for Chinese Linguistics (CCL) corpus (Zhan et al., 2003, 2019) to confirm that both Num–Unit–Adj–N and Num–Adj–Unit–N are attested, in line with prior analyses, and to establish how the two orders are distributed across sortal classifiers and noun-like measure expressions. The Mandarin CCL corpus search was designed as a targeted calibration exercise to confirm the availability of Num–Adj–Unit–N alongside Num–Unit–Adj–N, in line with prior analyses (Cheng & Sybesma, 1998, 1999; Luo et al., 2017), rather than as a comprehensive cross-linguistic comparison. Our primary

quantitative analysis examines Korean numeral-unit expressions using the morphologically annotated NIKL POS-tagged corpus, with the billion-word-scale koTenTen web corpus used as a qualitative verification tool for low-frequency candidates. Our goal is to assess, using naturally occurring corpus data, how robust numeral-unit adjacency is in Korean and whether any apparent exceptions are restricted to noun-like measure expressions.

2. Literature Review

2.1. Adjective placement in Chinese classifier structures

Cheng and Sybesma (1998, 1999) propose that the classifier position in Mandarin hosts both (i) sortal classifiers and (ii) noun-like measure expressions, which they term *massifiers*. Sortal classifiers are typically associated with counting individuals and are drawn from a relatively closed inventory, while noun-like measure expressions are more open-class and supply contextually salient units (container, portion, quantity), yielding measure or portion readings. The distinction is independently supported by interpretation and acquisition data: sortal classifiers select individuals, whereas noun-like measure expressions allow non-individual or portion-based interpretations (P. Li et al., 2008). One influential diagnostic for this contrast is adjectival intervention. The canonical order in adjective-containing classifier structures is Num-Unit-Adj-N, but Cheng and Sybesma observe that the marked Num-Adj-Unit-N order is available with noun-like measure expressions and not (or much less freely) with sortal classifiers. This asymmetry follows if measure expressions retain enough nominal content to be accessible to adjectival modification, whereas sortal classifiers behave like functional heads and therefore resist it (see (4a-b)).

Subsequent work has shown that pre-classifier modification is more narrowly constrained than this diagnostic would predict: it is largely confined to size adjectives, and some sortal classifiers permit it when the classifier-noun combination provides a dimension along which size can be measured (Luo et al., 2017). We therefore take the Mandarin facts to show that pre-unit adjectives are restricted, without assuming that nounhood alone determines their distribution.

2.2. Adjective placement in Korean classifier structures

A substantial body of research has examined Korean numeral-unit constructions from syntactic, semantic, and discourse-functional perspectives (e.g., Barrie et al., 2022; Choi, 2011; Hong, 2019; C. Kim, 2005; J.-B. Kim, 2011, 2013; R. Kim, 2025; Park, 2009, 2022), as well as from psycholinguistic processing and acquisition perspectives (e.g., Joo, 2017; Joo & Kim, 2016, 2023; H. Kim et al., 2023; O.-S. Lee, 2015; K. Lee & Lee, 2009). To our knowledge, however, no previous study has linked adjective placement in Korean numeral-unit constructions to the lexical status of the unit expression (highly grammaticalized classifiers vs more noun-like measure expressions) or to a Mandarin-style pre-classifier adjective diagnostic. The closest work in this domain is Joo (2017), which investigates the acquisition of adjective-containing Korean classifier constructions. When an adjective is included in the two Korean classifier constructions (i.e., N-Num-Unit and Num-Unit_{GEN}-N), Joo (2017, p. 4) argues that, given the two constraints—(a) adjectives must occur in the prenominal position and (b) the numeral-classifier unit cannot be separated—only three adjective-containing classifier constructions are permitted, as shown in (5a–c).

- (5) a. Num-Unit_{GEN}-Adj-N
 sey-kwen-uy khun chayk
 three-Cl-GEN big book
- b. Adj-Num-Unit_{GEN}-N
 khun sey-kwen-uy chayk
 big three-Cl-GEN book
- c. Adj-N-Num-Unit
 khun chayk sey-kwen
 big book three-Cl

Joo reports that native Korean children ($n=30$, age range=5;2–6;8) demonstrate adult-like knowledge, accepting possible orderings (Adj-N-Num-Unit, Adj-Num-Unit_{GEN}-N, Num-Unit_{GEN}-Adj-N) but rejecting impossible orderings in which an adjective intervenes between the numeral and the classifier (e.g., N-Num-Adj-Unit, Num-Adj-Unit_{GEN}-N) in acceptability judgment experiments. However, Joo's experimental items involve only one sortal classifier (*kay*) with countable nouns, which raises the question of whether children would allow the intervention of an adjective between a numeral and a noun-like measure expression. In other words,

it remains unclear whether a structure with a noun-like measure expression such as (6a) is acceptable, or at least more acceptable than (6b), which uses a sortal classifier. Adopting Choi's (2011) terminology, we hypothesize that the contrast in (6) may correlate with the distinction between genuine classifiers and ordinary nouns used as unit expressions. This hypothesis is evaluated against the corpus data below.

- (6) a. ?*twu khun pyeng-uy mwul*
 two big bottle-GEN water
 'two big bottles of water'
- b. **twu khun kay-uy wusan*
 two big Cl-GEN umbrella
 'two big umbrellas'

2.3. Previous corpus studies on Korean classifier structures

In Korean, corpus research has primarily focused on establishing the baseline distribution of numeral-classifier constructions and their word-order variation. O.-S. Lee (2015), drawing on the spoken portion of the Sejong Corpus (approximately 2.05 million words), reports that postnominal N-Num-Unit constructions are substantially more frequent in production than prenominal Num-Unit_{GEN}-N constructions. However, the corpus component in O.-S. Lee's study is relatively limited in scope, as it serves primarily to contextualize results from a psycholinguistic self-paced reading experiment rather than to provide a systematic large-scale distributional analysis.

Kim and Park (2017), drawing on corpus data from the National Institute of Korean Language, demonstrate that distributional patterns of Korean classifier constructions vary systematically according to lexical origin (native vs. Sino-Korean/loan nouns) and degree of classifier typicality. Their frequency analysis across N-Num-Unit, Num-Unit_{GEN}-N, and Num-Unit configurations shows that items retaining stronger lexical properties tend to display a higher proportion of Num-Unit uses (e.g., *twu pyeng* 'two bottles'), whereas more grammaticalized, dependent-noun-like classifiers more consistently favor N-Num-Unit order. These results suggest that word-order asymmetries in Korean classifier constructions are shaped not only by structural factors but also by lexical status along a continuum between nounhood and classifierhood. While Kim and Park identify such an important distributional pattern, they do not report in detail the overall size of the

corpora examined, a crucial piece of information for evaluating frequency-based findings in corpus research.

With respect to adjective placement, Joo (2017) conducted a corpus analysis of five high-frequency unit words (*kay*, *myeng*, *kwen*, *pyeng*, and *khep*) using the Sejong Corpus (approximately 10 million words across written and spoken registers). She reports that adjective-containing numeral-unit constructions are extremely rare overall, with only 23 tokens of Num-Unit_{GEN}-Adj-N, 7 tokens of Adj-N-Num-Unit, and no tokens of Adj-Num-Unit_{GEN}-N. Despite these findings, the corpus size remains relatively limited for the investigation of low-frequency constructions, and the study does not systematically distinguish between highly grammaticalized classifiers and more noun-like measure expressions. As a result, it leaves open the question of whether adjective placement interacts with the lexical status of the unit expression in a way comparable to the Mandarin pre-classifier adjective diagnostic.

Large-scale corpora allow rare constructions—such as adjective-containing numeral-unit expressions—to be examined with broader empirical coverage. Building on analyses of Mandarin pre-classifier adjectival modification (Cheng & Sybesma, 1998, 1999; Luo et al., 2017) and prior corpus studies of Korean numeral-classifier constructions (Joo, 2017; Kim & Park, 2017; O.-S. Lee, 2015), the present study takes Mandarin as a baseline and focuses on Korean corpora to evaluate the distribution of adjectives in numeral-unit expressions and the strength of numeral-unit adjacency. In particular, the corpus search is designed to determine whether clear intervention cases can be verified and, if so, whether they cluster around noun-like measure expressions rather than highly grammaticalized sortal classifiers. In addition, the study replicates earlier frequency findings using different corpus datasets. The research questions addressed in this study are as follows:

RQ1. In Korean corpus usage, which construction is more frequent: prenominal numeral expressions (Num-Unit_{GEN}-N) or postnominal expressions (N-Num-Unit), and how does this vary by register (spoken vs. written)?

RQ2. When adjectives occur in Korean numeral-unit expressions, where do they appear relative to the numeral and the unit expression, and is there corpus evidence for adjective intervention between the two?

RQ3. Against the Mandarin baseline—where Num-Adj-Unit-N is independently known to be available—do the Korean corpus results suggest a stronger distributional preference for numeral-unit adjacency, and, if any apparent Korean exceptions exist, are they concentrated in noun-like measure expressions?

3. Methods

3.1. Corpus data

We used two sets of Korean corpus data to analyze the frequencies and usage patterns of numeral-unit constructions. One is the NIKL POS-tagged corpus, and the other is the koTenTen corpus. The NIKL POS-tagged corpus was built by NIKL (National Institute of Korean Language) in 2020. Its size is 3 million *eojeols*. Here, an *eojeol* in Korean is a spacing unit in writing, which combines a meaningful stem (noun, verb, adjective) with grammatical morphemes. For example, (7) consists of three *eojeols* separated by spaces, and each *eojeol* consists of a stem and grammatical morphemes. For convenience, we refer to *eojeols* as spacing units hereafter.

- (7) na-nun hakkyo-ey ka-ss-ta
I-TOP school-to go-PAST-DECL
'I went to school'

The NIKL POS-tagged corpus contains both a written subcorpus (2 million spacing units) and a spoken subcorpus (1 million spacing units). The written subcorpus is composed of Korean newspaper articles from 2009 to 2017. Collected from three major newspaper publishers, the articles covered 11 topics such as economics, science, culture, and sports. The spoken subcorpus consists of three registers: public monologue (13.6%), public conversation (63.1%), and private conversation (23.3%). The public monologue contains news broadcasts, lectures, and conference presentations. The public conversation includes radio interviews, radio discussions, and podcasts. The private conversation contains daily dialogues. Every word in the corpus was manually annotated using a Korean POS tag list. The list consists of 47 tags over 7 major categories such as substantive, predicate, modifier, and postposition, and the target items in the present study (i.e., unit expressions) are tagged as bound nouns (code: NNB) in the category of substantive-noun.

The koTenTen corpus is a web-crawled corpus with 1.7 billion spacing units. It belongs to the TenTen Corpus Family in Sketch Engine, which contains 10+ billion-word corpora over 50 languages such as English, Hindi, Urdu, and Korean. The data in the koTenTen corpus were collected in August and September 2012 and from December 2017 to April 2018, from a variety of web domains such as *.org*, *.kr*, and *.com*. Every spacing unit in the corpus was POS-tagged by *HanNanum*

using a simplified set of seven tags (i.e., adjective, adverb, interjection, noun, numeral, pronoun, verb). Every spacing unit was also lemmatized for the base form of the stem, which enabled us to collect the usage of a classifier combined with various grammatical morphemes.

To examine adjective placement in numeral-unit constructions in Mandarin, we used the CCL corpus. It comprises both modern and classical Chinese written texts. For the purposes of this study, we restrict our analysis to the modern Chinese subcorpus, which includes literary works, newspaper articles, blog essays, and academic journal publications. This subcorpus contains approximately 581 million characters.

3.2. Corpus analysis

We adopt a corpus-based approach to linguistic analysis (e.g., Kim & Oh, 2025; Sung et al., 2026), focusing on Korean and using Mandarin as a baseline for calibration. For Korean, we implement strict POS-sequence templates over morphologically annotated data and treat intervention as the presence of adjectival material in a strictly ordered local sequence between a numeral and a unit word. We combine quantitative extraction from the NIKL POS-tagged corpus with manual concordance inspection, and we additionally use the billion-word-scale 2018 Korean web corpus as a qualitative verification tool for very low-frequency candidates.

Our primary dataset, i.e., the NIKL POS-tagged corpus, is a morphologically analyzed Korean corpus, in token-per-row CSV format, where each row corresponds to a single morph token annotated with a surface form and POS label, and indexed by document and sentence identifiers (e.g., `doc_id`, `sentence_id`). This corpus was analyzed using Python code in a Google Colab environment. Searches were implemented over morphologically segmented strings and restricted to sentence boundaries; document identifiers were used only for file-level organization and not as the linguistic search boundary. Because the study is corpus-based, non-attestation is interpreted as evidence about distributional rarity in the corpora examined, not as direct proof of categorical ungrammaticality.

To examine the usage patterns of classifier constructions and adjective placement in Korean, we extracted the eight contiguous POS-sequence templates shown in Table 1 and computed their frequencies. The first four constructions contain postnominal classifiers, while the other four contain prenominal classifiers. The first and the fifth constructions represent the base forms of the postnominal and

prenominal classifier constructions, respectively, whereas the remaining constructions include adjectives. Among the adjective-containing constructions, the fourth and the seventh represent the intervention patterns, in which an adjective intervenes between the numeral and the unit word, and these are the focus of our analysis. In addition, the third pattern is also noteworthy, as the adjective modifies the unit rather than the noun. As a baseline calibration, we also queried the CCL corpus for the two Mandarin word orders Num-Unit-Adj-N and Num-Adj-Unit-N and manually inspected concordances to confirm attestation in line with prior literature.

Table 1. Queried constructions for the NIKL POS-tagged corpus

No.	Construction	Example
1	N-Num-Unit	<i>sakwa twu kay</i> ‘two apples’
2	Adj-N-Num-Unit	<i>khun sakwa twu kay</i> ‘two big apples’
3	N-Adj-Num-Unit	<i>sakwa khun twu kay</i> ‘two big apples’
4	<i>N-Num-Adj-Unit</i>	<i>sakwa twu khun kay</i> ‘two big apples’
5	Num-Unit _{GEN} -N	<i>twu kay-uy sakwa</i> ‘two apples’
6	Adj-Num-Unit _{GEN} -N	<i>khun twu kay-uy sakwa</i> ‘two big apples’
7	<i>Num-Adj-Unit_{GEN}-N</i>	<i>twu khun kay-uy sakwa</i> ‘two big apples’
8	Num-Unit _{GEN} -Adj-N	<i>twu kay-uy khun sakwa</i> ‘two big apples’

Note. The intervention patterns are highlighted in bold italics.

The Python-based automatic search of the eight classifier constructions in the NIKL POS-tagged corpus found 3,418 cases, and a follow-up manual inspection excluded 2,079 non-target cases³⁾. Among the final set of 1,339 target cases, adjective-containing numeral-unit expressions were extremely low-frequency, echoing Joo’s (2017) report for the Sejong Corpus.

Because the adjective-containing numeral-unit expressions were rare, we supplemented the primary quantitative analysis with a qualitative search in the

3) The majority of the excluded tokens involved sequences that matched the POS pattern but did not form a single numeral-unit expression. For example, the following POS templates were used to extract N-Num-Unit patterns: NNG-NR-NNB, NNG-SN-NNB, and NNG-MMN-NNB. However, these templates also returned strings such as *ohu 2-si* ‘2 p.m.’, *pal-wol i-il* ‘August 2’, and *kawunte 3-myeng* ‘three among them’, which do not constitute numeral-unit noun phrases and were therefore excluded from the analysis.

billion-word-scale Korean web corpus (koTenTen) hosted on Sketch Engine. We used POS-based queries to retrieve candidate strings containing numerals, unit words, and adjectival material. However, given the limitations of the Sketch Engine output, we refrained from making any claims beyond verifying whether the Num-Adj-Unit pattern is attested in Korean corpora. Because the underlying HanNanum tagger collapses the full complexity of Korean morphosyntax into only seven categories (i.e., adjective, adverb, interjection, noun, numeral, pronoun, verb), tagging errors are inevitable. For example, items automatically tagged as adjectives frequently turned out to be adverbs or other constructions upon manual inspection. For this reason, manual verification of each candidate token was an essential step in the analysis, and the counts reported here reflect only those tokens that were confirmed as adjective-bearing constructions through hand-coding. To investigate adjective placement in Korean numeral-unit constructions, we conducted a corpus search targeting the eight constructions in Table 1. Because retrieving all possible adjective-unit combinations in large web corpora proved practically infeasible, we limited the search to a set of representative unit expressions that are among the most frequent in Korean numeral-unit constructions. This set included both sortal classifiers (*kay*, *kwen*, *myeng*; the three “genuine” classifiers identified by Choi (2011)) and noun-like measure expressions (*pyeng* ‘bottle’, *khep* ‘cup’, *can* ‘glass’, *pongci* ‘bag’, *kulus* ‘bowl’, *thong* ‘container’, *swul* ‘spoon’, and *sangca/paksu* ‘box’). All instances in which these unit expressions occurred together with either an adjective or a noun were retrieved from the corpus and subsequently categorized according to their word-order patterns.

The initial extraction yielded more than 40,000 tokens in total, including 20,020 tokens for Adj-N-Num-Unit, 12,412 tokens for Adj-Num-Unit_{GEN}-N, and 10,016 tokens for Num-Unit_{GEN}-Adj-N. When the classifier *kay* alone was considered, the corresponding numbers were 16,357, 8,052, and 7,133 tokens, respectively. Given the large number of tokens, full manual verification was impractical. We therefore reduced the dataset by restricting it to tokens containing size adjectives (e.g., *khun* ‘big’, *cahun* ‘small’), given that the baseline language Mandarin allows the marked Num-Adj-Unit-N order with size adjectives. After applying this size-adjective restriction for the comparative dataset, the remaining tokens were manually inspected. During this process, we excluded cases in which the noun and the numeral-unit sequence were separated by intervening particles or case markers (e.g., *khun sakwa-ka twu kay* ‘two big apples’, where the nominative marker *-ka* intervenes between the noun and the numeral-unit sequence). We also excluded cases in which

the adjective formed part of a relative clause modifying the noun rather than directly modifying the unit expression. The remaining tokens were then used to identify genuine instances of adjective placement relative to the numeral-unit boundary.

4. Results

4.1. Corpus analysis: Chinese CCL corpus

Consistent with prior literature on pre-classifier adjectival modification, a targeted baseline search in the CCL corpus returns instances of both the canonical Num-Unit-Adj-N order and the marked Num-Adj-Unit-N. Following Luo et al. (2017), who show that pre-classifier modification is particularly associated with size-denoting adjectives, we restrict the Chinese corpus investigation to *da* ‘big’ and *xiao* ‘small’. For methodological consistency, we use the numeral *yi* ‘one’ in combination with either the general sortal classifier *ge* or the common measure expression *ping* ‘bottle’ when retrieving Num-Unit-Adj-N and Num-Adj-Unit-N sequences from the corpus.

The corpus data show that with the sortal classifier *ge*, the dominant word order is overwhelmingly Num-Unit-Adj-N. The corpus returned 73,562 tokens containing *da* ‘big’ and 95,602 tokens containing *xiao* ‘small’, whereas only a very small number of tokens exhibited the Num-Adj-Unit-N order (75 tokens with *da* and 78 tokens with *xiao*). In total, 153 tokens (75 + 78) out of 169,317 tokens exhibit the Num-Adj-Unit-N order, corresponding to merely 0.1% of the dataset, suggesting that this pattern is extremely rare with the sortal classifier *ge*. See (8) for an example of Num-Unit-Adj-N.⁴⁾

- (8) yi ge da zhengke
 one Cl big politician
 ‘one big politician’

Notably, with the measure expression *ping*, the pattern is reversed. The dominant

4) An example of Num-Adj-Unit-N, which is very rare in the corpus, is as follows.

- (i) yi da ge hongshu
 one big Cl sweet potato
 ‘one big sweet potato’

order is Num-Adj-Unit-N, yielding 640 tokens with *da* and 496 tokens with *xiao*. By contrast, the Num-Unit-Adj-N order occurs much less frequently, with 76 tokens containing *da* and 81 tokens containing *xiao*. In total, 1,136 tokens (640 + 496) out of 1,293 exhibit the Num-Adj-Unit-N order, corresponding to 87.9% of the dataset, suggesting that this pattern is dominant with the measure expression *ping*. See (9) for an example of Num-Adj-Unit-N.

- (9) yi da ping shui
 one big bottle water
 ‘one big bottle of water’

This confirms that, in Mandarin, adjectival material can occur between the numeral and the unit word in naturalistic data. The pattern is attested with both unit types, but the relative frequencies differ sharply: it accounts for 87.9% of the retrieved *ping* tokens and only 0.1% of the retrieved *ge* tokens. This asymmetry is consistent with Cheng and Sybesma’s (1998, 1999) observation that noun-like measure expressions are more accessible to adjectival modification than sortal classifiers. It does not, however, adjudicate between a nounhood-based account and the measurement-based account of Luo et al. (2017), on which pre-classifier modification is licensed by the dimension being measured rather than by the lexical category of the unit word.

4.2. Corpus analysis: Korean NIKL POS-tagged corpus

Table 2 shows the frequency distributions of two major Korean numeral-unit constructions in the written and spoken subcorpora.

Table 2. Distribution of Korean numeral-unit constructions by register (written vs. spoken) in the NIKL POS-tagged corpus

Type	Written	Spoken	Sum
N-Num-Unit	511 (67.8%)	427 (76.7%)	938 (71.5%)
Num-Unit _{GEN} -N	243 (32.2%)	130 (23.3%)	373 (28.5%)
Sum	754 (100.0%)	557 (100.0%)	1,311 (100.0%)

Table 2 reveals a clear preference for the N-Num-Unit construction over the Num-Unit_{GEN}-N, accounting for 71.5% of numeral-unit constructions. The pattern N-Num-Unit occurs 938 times in total, with 511 tokens in the written subcorpus and 427 tokens in the spoken subcorpus. By contrast, the Num-Unit_{GEN}-N appears substantially less frequently, with 373 tokens overall (243 in the written subcorpus and 130 in the spoken subcorpus). This distribution indicates that the N-Num-Unit construction is the dominant pattern in both registers. This robust preference for postnominal placement replicates the general tendency reported in previous corpus and processing studies (e.g., O.-S. Lee, 2015; Joo & Kim, 2016; Kim & Park, 2017), suggesting that the dominance of N-Num-Unit is stable across corpora and annotation schemes. At the same time, the Num-Unit_{GEN}-N is consistently attested in both written and spoken language, suggesting that it functions as a productive but secondary alternative to the canonical N-Num-Unit structure.

When normalized by corpus size, the same asymmetry remains: the postnominal pattern occurs at a substantially higher rate than the prenominal pattern in both subcorpora, with the contrast especially pronounced in the spoken subcorpus. More specifically, N-Num-Unit corresponds to approximately 255.5 tokens per million spacing units in the written subcorpus and 427 per million in the spoken subcorpus, whereas Num-Unit_{GEN}-N corresponds to about 121.5 and 130 per million, respectively.

Table 3. Attested word orders in Korean numeral-unit phrases

No.	Type	Example	Token
1	N-Num-Unit	<i>sakwa twu kay</i> ‘two apples’	938 (70.1%)
2	Adj-N-Num-Unit	<i>khun sakwa twu kay</i> ‘two big apples’	16 (1.2%)
3	N-Adj-Num-Unit	<i>sakwa khun twu kay</i> ‘two big apples’	0 (0%)
4	<i>N-Num-Adj-Unit</i>	<i>sakwa twu khun kay</i> ‘two big apples’	7 (0.5%)
5	Num-Unit _{GEN} -N	<i>twu kay-uy sakwa</i> ‘two apples’	373 (27.9%)
6	Adj-Num-Unit _{GEN} -N	<i>khun twu kay-uy sakwa</i> ‘two big apples’	0 (0%)
7	<i>Num-Adj-Unit_{GEN}-N</i>	<i>twu khun kay-uy sakwa</i> ‘two big apples’	0 (0%)
8	Num-Unit _{GEN} -Adj-N	<i>twu kay-uy khun sakwa</i> ‘two big apples’	5 (0.4%)
Sum			1,339 (100%)

Note. The intervention patterns are highlighted in bold italics.

Table 3 shows that numeral-unit constructions containing adjectives are extremely rare in the corpus, with only 28 tokens (2.1%=28/1,339) in total. This distribution is consistent with Joo’s (2017) findings based on the Sejong Corpus, which show that adjective-containing classifier phrases constitute only a very small proportion of numeral expressions. Among these, 16 tokens occur in the Adj-N-Num-Unit pattern and 5 tokens in the Num-Unit_{GEN}-Adj-N pattern. Notably, 7 tokens (0.5%=7/1,339) occur in the N-Num-Adj-Unit pattern, in which an adjective appears between the numeral and the unit word. All of these cases involve the measure expression *swul* ‘spoon’.⁵⁾

- (10)
- a.

kancang

han

khun

swul

soy sauce

one

big

spoon

‘a big spoon of soy sauce’
- b.

socwu

twu

khun

swul

soju

two

big

spoon

‘two big spoons of soju (alcohol)’

Because the NIKL POS-tagged corpus returned only 28 tokens containing adjectives, we examined additional web data in the koTenTen corpus using Sketch Engine and manually inspected concordance lines for candidate adjective-containing numeral-unit constructions.

4.3. Corpus analysis: Korean koTenTen corpus via Sketch Engine

Table 4 shows the observed frequencies of adjective placement patterns for 11 major Korean unit words in the Sketch Engine corpus.

5) It should be noted that in another construction, namely the N-Num-Unit pattern, forms such as *khunswul* and *cakunswul*, which are written without a space between *khun* or *cakun* and *swul*, were POS-coded as nouns. These forms are compound nouns consisting of the adjectives *khun* ‘big’ and *cakun* ‘small’ combined with *swul* ‘spoon’. Following the original coding of the NIKL POS-tagged corpus, we did not treat these cases (15 tokens in total) as adjective-added constructions.

Table 4. Distribution of adjective placement in Korean numeral-unit constructions in the Sketch Engine corpus (limited to size adjectives)

Unit word	Adj-N- Num-Unit	N-Adj- Num-Unit	<i>N-Num- Adj-Unit</i>	Adj-Num- Unit _{GEN} -N	<i>Num-Adj- Unit_{GEN}-N</i>	Num- Unit _{GEN} -Adj-N
<i>Kay</i>	2,493	0	0	650	0	3,714
<i>Myeng</i>	5	0	0	7	0	324
<i>Kwen</i>	17	0	0	49	0	58
<i>Pyeng</i>	23	0	0	0	0	1
<i>Khep</i>	0	1	0	0	0	0
<i>Can</i>	10	1	0	2	0	5
<i>Pongci</i>	4	0	0	2	0	0
<i>Kulus</i>	3	0	0	0	0	1
<i>Thong</i>	27	0	0	3	0	0
<i>Swul</i>	1	0	196	0	2	0
<i>Sangca/Paksu</i>	5	0	0	0	0	0
<i>Sum</i>	2,588	2	196	713	2	4,103

Note. The intervention patterns are highlighted in bold italics. The full set of corpus examples is available at: https://osf.io/hu5wk/overview?view_only=1211b6abd2ee442baf6c9501c9062c72

Table 4 should be interpreted primarily as a diagnostic overview rather than as a direct frequency comparison, because the Sketch Engine queries were restricted to selected unit expressions and the retained set was further narrowed through manual filtering. In addition, low or zero attestation in the datasets examined is interpreted as evidence of distributional rarity in usage rather than as direct proof of categorical ungrammaticality. Even so, the distribution is revealing: adjectives rarely occur between a numeral and a unit word across the surveyed expressions, with the exception of *swul* ‘spoon’. For the more grammaticalized sortal classifiers *kay*, *myeng*, and *kwen*, the attested adjective-bearing tokens overwhelmingly occur in orders that keep the numeral and the unit adjacent. In more noun-like measure expressions such as *pyeng*, *can*, *pongci*, and *thong*, adjective placement between the numeral and the unit remains highly restricted. However, for *swul* ‘spoon’, we find cases in which an adjective appears between the numeral and the unit. See (11) for an example of Num-Adj-Unit_{GEN}-N and (12) for examples of N-Num-Adj-Unit. Example (12a) illustrates a typical use of *swul* with a food ingredient, whereas

(12b) shows a metaphorical extension referring to a smile.

- (11) Num-Adj-Unit_{GEN}-N
- | | | | |
|---|-------|-----------|-------|
| 4 | cakun | swul-uy | sokum |
| 4 | small | spoon-GEN | salt |
- ‘four small spoons of salt’

- (12) N-Num-Adj-Unit
- a. milkalwu han khun swul
- flour one big spoon
- ‘one big spoon of flour’
- b. wusum han khun swul
- smile one big spoon
- ‘one big spoon of smile’

When we analyzed the NIKL POS-tagged corpus, we found both *khunswul* and *cakunswul*, single compound words with no space between the elements, as well as cases where an adjective and a unit word appear with a space between them (e.g., *khun swul*, *cakun swul*). We therefore further examined all such forms (i.e., *khun swul*, *cakun swul*, *khunswul*, *cakunswul*) in Sketch Engine. The search yielded a total of 1,762 tokens without a space and 391 tokens with a space, suggesting that these expressions may occur both as lexicalized compounds⁶⁾ and as adjective-unit phrases.

Notably, this ambiguity is not confined to *swul*. Measure expressions such as *pyeng* ‘bottle’ or *can* ‘glass’ also retain lexical nounhood and can be used as ordinary count nouns denoting containers, not only as classifier-like unit expressions, as seen in (13a-b).

- (13) a. sey pyeng-uy cakun pyeng
- three bottle-GEN small bottle
- ‘three small bottles’
- b. cakun can han can
- small glass one glass
- ‘one small glass’

6) We acknowledge that spacing alone does not establish a categorical distinction between lexicalized compounds and productive adjective-unit phrases. What can be established, however, is that both *khunswul* and *cakunswul* are attested as lexical items. Both *khunswul* and *cakunswul* are listed as bound nouns in the Standard Korean Language Dictionary (National Institute of Korean Language).

In (13a), the first occurrence of *pyeng* forms part of the prenominal numeral-unit expression, whereas the second denotes the lexical container noun modified by *cakun* ‘small’. Similarly, in (13b), the final *can* functions as the unit expression associated with the preceding lexical expression *cakun can* ‘small glass’. These examples illustrate the dual availability of container nouns as lexical nouns and as unit expressions.

It is also important to note two cases of N-Adj-Num-Unit with *can* ‘glass’ and *khep* ‘cup’, as seen in (14a–b).

- | | | | | |
|---------|------------------------------|------------|-----|-------|
| (14) a. | Maysilcwu | cokumanhan | han | can |
| | plum wine | small | one | glass |
| | ‘a small glass of plum wine’ | | | |
| b. | yokwuluthu | cakun | han | khep |
| | yogurt | small | one | cup |
| | ‘a small cup of yogurt’ | | | |

Note that in N-Adj-Num-Unit, the adjective modifies the measure expression: *cokumanhan* ‘small’ modifies *can* ‘glass’ in (14a), and *cakun* ‘small’ modifies *khep* ‘cup’ in (14b), which suggests the nounhood of *can* and *khep*.

5. Discussion and Conclusion

This study uses Mandarin as a baseline calibration and focuses primarily on Korean corpora to investigate adjective placement in numeral-unit expressions, in particular whether adjectival material can occur between a numeral and a unit word. Prior work establishes that Mandarin allows a restricted pre-classifier adjective order (Num-Adj-Unit-N) alongside the canonical Num-Unit-Adj-N (Cheng & Sybesma, 1998, 1999; Luo et al., 2017; Jiang et al., 2022). In contrast, the Korean literature has argued for an especially tight numeral-classifier complex to which “nothing can intervene” (K.-S. Kim, 2010; Joo, 2017). Using corpus evidence, we reassess how robust numeral-unit adjacency is in Korean in naturalistic usage and ask whether any apparent exceptions are restricted to noun-like measure expressions.

RQ1 asked about the baseline distribution of the two major Korean numeral-unit constructions. In the NIKL POS-tagged corpus, the postnominal N-Num-Unit pattern overwhelmingly outnumbers the prenominal Num-Unit_{GEN}-N pattern in both

written and spoken subsets (Table 2), replicating the strong postnominal preference reported in earlier studies (O.-S. Lee, 2015; Joo & Kim, 2016; Kim & Park, 2017).

Several, not mutually exclusive, explanations can account for the distributional asymmetry between N-Num-Unit and Num-Unit_{GEN}-N, observed in the present study as well as in previous work. First, the results can be interpreted from the perspective of information packaging. Research on Korean numeral-unit constructions and typologically comparable Japanese numeral-quantifier constructions has shown that the choice among structurally available configurations may reflect systematic differences in how referents and quantities are packaged in discourse, rather than mere stylistic variation (Amazaki, 2006; Downing, 1996; C. Kim, 2005; J.-B. Kim, 2011, 2013). For example, J.-B. Kim (2011, 2013) argues that floated numeral-classifier sequences introduce new quantitative information and mark the onset of the rheme in the thematic structure of the clause. His account is developed for the floating type rather than for the juxtaposed N-Num-Unit pattern examined here; we extend the underlying discourse principle to the latter, on the assumption that both orders place the quantity expression after the lexical noun. On this view, the N-Num-Unit order aligns naturally with the given-before-new and theme-before-rheme organization characteristic of Korean discourse: the lexical noun is introduced first as a referential or thematic anchor, while the following numeral-unit sequence supplies more specific quantitative information. By contrast, the prenominal Num-Unit_{GEN}-N order may package quantity and nominal reference together within a single complex nominal expression, a pattern that may be better suited to planned or edited registers. This interpretation is consistent with the comparatively larger share of Num-Unit_{GEN}-N in the written subcorpus (32.2%) than in the spoken subcorpus (23.3%) observed in Table 2. A further, complementary explanation comes from the processing-based account proposed by Joo and Kim (2016), which builds on Hawkins's (2004) domain-minimization account. Under this view, the classifier, or unit expression in the terminology adopted here, is the head of the unit phrase. N-Num-Unit is predicted to be preferred over Num-Unit_{GEN}-N insofar as it minimizes the processing domain calculated from the unit head to the element with which the phrase is syntactically integrated. Although these accounts are compatible with the corpus pattern, the present study does not directly annotate information structure, specificity, givenness, or processing difficulty. A full account of the discourse-functional and processing motivations underlying the observed asymmetry is therefore beyond its scope. Future research combining corpus-based discourse annotation with processing measures could further substantiate these explanations.

RQ2 asked where adjectives occur when they appear in Korean numeral expressions and whether any naturally occurring cases show an adjective intervening between the numeral and the unit word. In both the NIKL POS-tagged corpus and the supplementary koTenTen corpus, adjective intervention between the numeral and the unit word was extremely rare. All candidate intervention tokens involved *swul* ‘spoon’, though some caution is warranted because it is unclear whether *khun swul* ‘big spoon’ should be analyzed as a lexicalized unit noun or as a measure expression modified by an adjective. However, as discussed above, such potential adjective modification appears to be limited to measure expressions that retain noun-like properties rather than to grammatical sortal classifiers. Taken together, the results suggest that apparent intervention in Korean is highly restricted; any candidates are expected to occur only with expressions that retain substantial lexical nounhood.

Importantly, the apparent intervention cases involving *khun swul* ‘big spoon’, together with the dual uses of *can* ‘glass’ and *pyeng* ‘bottle’ in (13) and the adjectivally modified *can* ‘glass’ and *khep* ‘cup’ expressions in (14), all involve noun-like measure expressions; no comparable pattern is observed with highly grammaticalized sortal classifiers such as *kay*, *myeng*, or *kwen*. This distribution is compatible with Choi’s (2011) distinction between genuine classifiers such as *kay*, *myeng*, and *kwen* and ordinary nouns that may function as classifier-like unit expressions, such as *can* ‘glass’ and *pyeng* ‘bottle’. Choi analyzes genuine classifiers as clitic-like elements that enter into a particularly close dependency with numerals, whereas ordinary nouns retain greater lexical independence. The present corpus findings extend this distinction to adjective placement: the apparent intervention candidates occur with noun-like measure expressions, not with the genuine classifiers examined here.

In this respect, the Korean pattern shows only a partial parallel with Mandarin. In the targeted Mandarin comparison, pre-classifier adjectival modification is much more frequent with the noun-like measure expression *ping* ‘bottle’ than with the highly grammaticalized sortal classifier *ge*. However, the broader Mandarin pattern cannot be reduced to lexical nounhood alone, because some sortal classifiers also permit pre-classifier size adjectives when the classifier–noun combination provides a dimension along which size can be measured (Luo et al., 2017). The Korean data are more restricted: the few attested Num–Adj–Unit sequences involve *swul*-based expressions that either have lexicalized counterparts or independently permit adjectival modification, rather than clear cases of productive violation of numeral–classifier adjacency. Thus, lexical nounhood appears to play an important role in the

Korean cases, whereas Mandarin adjective placement reflects both the lexical properties of the unit expression and the semantic conditions governing measurement.

RQ3 asked whether the Korean corpus patterns, viewed against the Mandarin baseline, suggest a stronger numeral–unit adjacency constraint in Korean. The contrast is clear at the level of distributional evidence. Mandarin is independently known to allow Num–Adj–Unit–N in restricted environments, and our CCL baseline check confirms its attestation alongside Num–Unit–Adj–N. Korean, by contrast, shows a markedly stronger preference for numeral–unit adjacency in the corpora examined. Thus, although both languages are often grouped as classifier languages, the present corpus evidence suggests that the Korean corpora examined show stronger distributional evidence for numeral–unit adjacency than the Mandarin baseline queried here.

Taken together, the corpus data reveal a clear cross-linguistic difference in the distribution of adjectives within numeral–unit constructions. In Mandarin, Num–Adj–Unit–N is independently known to occur under restricted morphosyntactic and semantic conditions, and the targeted CCL comparison shows that it is substantially more frequent with *ping* ‘bottle’ than with *ge*. Korean, by contrast, shows a much stronger distributional preference for numeral–unit adjacency: among the unit expressions examined, apparent intervention was attested only with *swul* ‘spoon’, an expression that also occurs in lexicalized forms such as *khunswul* and *cakunswul*. The two languages thus differ not only in frequency but also in the range of conditions under which pre-unit adjectival material is licensed.

Several limitations follow from the corpus-based design. First, the absence of a construction in a corpus does not entail ungrammaticality; low-frequency but acceptable patterns may simply be unattested in the sampled data. Second, orthographic variation and lexicalization can obscure category boundaries in web corpora, increasing the need for manual verification. The present findings should therefore be interpreted as corpus-based evidence about usage preferences rather than as direct evidence of grammatical possibility. This caution also applies to broader syntactic analyses of numeral–unit constituency. Within Little et al.’s (2022) typology, Korean has been analyzed both as a classifier-for-numeral language, with a [[Num Cl] N] structure (Barrie et al., 2022), and as a classifier-for-noun language, with a [Num [Cl N]] structure (R. Kim, 2025). Because surface distributional patterns alone do not distinguish between these constituency analyses, the present corpus results do not adjudicate between them; resolving the issue will require independent syntactic and experimental evidence.

Despite these caveats, the present study makes two contributions. Empirically, it provides corpus-based baselines for adjective placement in numeral-unit expressions, confirming the availability of adjectival intervention in Mandarin and documenting the rarity of such intervention in Korean naturalistic data. Analytically, it sharpens the cross-linguistic contrast by showing that numeral-unit adjacency appears more robust in Korean than in Mandarin in actual usage, while also identifying a plausible locus for apparent Korean exceptions in noun-like measure expressions.

References

- Amazaki, O. (2006). *A functional analysis of numeral quantifier constructions in Japanese* (Doctoral dissertation). University at Buffalo, The State University of New York.
- Barrie, M., Chung, H., & An, D.-H. (2022). Numeral classifiers in Korean -ki nominalizations. *Linguistic Research*, 39(3), 499-518.
- Borer, H. (2005). *Structuring sense: Volume I: In name only*. Oxford University Press.
- Börjars, K., Hicks, C., & Payne, J. (2018). *Interdependencies in Chinese noun phrases*. In M. Butt & T. Holloway King (Eds.), *Proceedings of the LFG'18 Conference* (pp. 109-128). CSLI Publications.
- Cheng, L. L.-S., & Sybesma, R. (1998). Yi-wan tang, yi-ge tang: Classifiers and massifiers. *Tsing Hua Journal of Chinese Studies New Series*, 28(3), 385-412.
- Cheng, L. L.-S., & Sybesma, R. (1999). Bare and not so bare nouns and the structure of NP. *Linguistic Inquiry*, 30(4), 509-542.
- Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics*, 6(4), 339-405.
- Choi, K. (2011). On the nature of the dependency between a numeral and a classifier. *Linguistic Research*, 28(3), 517-542.
- Downing, P. A. (1996). *Numeral classifier systems: The case of Japanese*. John Benjamins.
- Hawkins, J. (2004). *Efficiency and complexity in grammars*. Oxford University Press.
- Hong, Y.-T. (2019). The structure of the postnominal numeral classifier construction. *Studies in Generative Grammar*, 29(3), 485-515.
- Jiang, L. J., Jenks, P., & Jin, J. (2022). The syntax of classifiers in Mandarin Chinese. In C.-R. Huang, Y.-H. Lin, I.-H. Chen, & Y.-Y. Hsu (Eds.), *The Cambridge handbook of Chinese linguistics* (pp. 515-549). Cambridge University Press.
- Joo, K.-J. (2017). Acquisition of word order in the adjective-added Korean classifier phrases. *Journal of Cognitive Science*, 18(1), 1-20.
- Joo, K.-J., & Kim, K. (2016). A processing-based account for the preferred ordering of the Korean classifier structures. *Journal of Cognitive Science*, 17(3), 423-440.

- Joo, K.-J., & Kim, K. (2023). L1 transfer in the acquisition of word order in the adjective-added Korean classifier phrases by L1-Chinese and L1-Russian children. *Korean Journal of Applied Linguistics*, 39(1), 3-25.
- Kim, C. (2005). Order and meaning: Numeral classifiers and specificity in Korean. In J. Alderete, C.-H. Han, & A. Kochetov (Eds.), *Proceedings of the 24th West Coast Conference on Formal Linguistics* (pp. 218-226). Cascadia Proceedings Project.
- 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(4), 763-785.
- Kim, H., & Park, M. (2017). Corpus-based analysis of the grammatical category of Korean classifiers. *Journal of Ewha Korean Language and Literature*, 43, 93-117.
- Kim, J.-B. (2011). Floating numeral classifiers in Korean: A thematic-structure perspective. In S. Müller (Ed.), *Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar, University of Washington* (pp. 302-313). CSLI Publications.
- Kim, J.-B. (2013). Floated numeral classifiers in Korean: A non-derivational, functional account. *Lingua*, 133, 189-212.
- Kim, K., & Melchin, P. B. (2018). Modifying plurals, classifiers, and co-occurrence: The case of Korean. *Glossa: A Journal of General Linguistics*, 3(1), Article 25.
- Kim, K.-S. (2010). On the syntax and semantics of Korean classifier constructions. *Studies in Generative Grammar*, 20(2), 349-368.
- Kim, R. (2025). On the constituency of numeral and classifier in Korean. *Studies in Generative Grammar*, 35(4), 485-511.
- Kim, S., & Oh, S.-Y. (2025). First- and second-person pronoun use by East Asian EFL learners: Corpus-based study of argumentative essays. *Language Research*, 61(1), 35-62.
- Lee, K.-O., & Lee, S.-Y. (2009). The acquisition of Korean numeral classifiers. In C. Lee, G. Simpson, & Y. Kim (Eds.), *The handbook of East Asian psycholinguistics, Volume III: Korean* (pp. 137-149). Cambridge University Press.
- Lee, O.-S. (2015). A test of frequency-based processing: Evidence from Korean numeral classifier constructions. *Language Research*, 51(3), 629-652.
- Li, P., Barner, D., & Huang, B. H. (2008). Classifiers as count syntax: Individuation and measurement in the acquisition of Mandarin Chinese. *Language Learning and Development*, 4(4), 249-290.
- Little, C. R., Moroney, M., & Royer, J. (2022). Classifiers can be for numerals or nouns: Two strategies for numeral modification. *Glossa: A Journal of General Linguistics*, 7(1), 1-35.
- Luo, Q., Hsieh, M.-L., & Shi, D. (2017). Pre-classifier adjectival modification in Mandarin Chinese: A measurement-based analysis. *Journal of East Asian Linguistics*, 26(1), 1-36.
- Park, S.-Y. (2009). The syntax of numeral classifiers: A small clause inside a DP. *Language Research*, 45(2), 203-230.
- Park, S.-Y. (2022). Two types of plurals and numeral classifiers in classifier languages: The case of Korean. *Journal of East Asian Linguistics*, 31(2), 139-177.

- Peking University Center for Chinese Linguistics. (n.d.). *CCL Corpus*. Retrieved from http://ccl.pku.edu.cn:8080/ccl_corpus
- Rothstein, S. (2017). *Semantics for counting and measuring*. Cambridge University Press.
- Sketch Engine. (n.d.). *koTenTen - Korean corpus from the web*. Retrieved from <https://www.sketchengine.eu/kotenten-korean-corpus>
- Sung, M.-C., Kim, K., & Nam, B. (2026). Influence of topic-prominent L1s on the use of L2 English copula *be*: A corpus-based study. *International Review of Applied Linguistics in Language Teaching*, 64(1), 499-521.
- Zhan, W., Guo, R., & Chen, Y. (2003). *The CCL Corpus of Chinese Texts*. Center for Chinese Linguistics, Peking University.
- Zhan, W., Guo, R., Chang, B., Chen, Y., & Chen, L. (2019). The building of the CCL corpus: Its design and implementation. *Corpus Linguistics*, 6(1), 71-86.

Kitaek Kim

Professor

Department of English Language Education

Seoul National University

1 Gwanak-ro, Gwanak-gu, Seoul 08826, South Korea

E-mail: kitaek@snu.ac.kr

Kum-Jeong Joo

Assistant Professor

Department of Business English

Induk University

12 Choansan-ro, Nowon-gu, Seoul 01878, South Korea

E-mail: jookj@induk.ac.kr

Min-Chang Sung

Associate Professor

Department of English Education

Gyeongin National University of Education

155 Sammak-ro, Anyang-si, Gyeonggi-do 13910, South Korea

E-mail: mcsung@ginue.ac.kr

Yunchuan Chen

Associate Professor of the Practice

Asian & Middle Eastern Studies, Trinity College of Arts & Sciences

Duke University

Rm 225, 2204 Erwin Road, Durham, NC 27708, United States

E-mail: yunchuan.chen@duke.edu

Received: March 9, 2026

Revised version received: April 7, 2026

Accepted: July 25, 2026