

## What’s the smallest part of spinach? A new experimental approach to the count/mass distinction

Sea Hee Choi & Tania Ionin\*

**Abstract.** This paper reports on a study that uses a novel methodology, the *minimal parts identification task*, in order to probe the relationship between morphosyntax and interpretation. English, Korean and Mandarin Chinese differ from one another with regard to the count/mass distinction. Building on prior research, this study examines whether speakers of these three languages also differ in how they interpret count vs. mass nouns. The findings, while uncovering some language-specific effects of morphosyntax, point to the importance of universality, and suggest that interpretation drives morphosyntax rather than the other way around.

**Keywords.** atomicity; count/mass distinction; English; Korean; Mandarin Chinese

**1. Introduction.** The object/substance distinction is cognitive, while the count/mass distinction is linguistic; in plural-marking languages like English, there are a number of diagnostics for whether a noun is count or mass, see Table 1. According to Chierchia (1998a, b, 2010, 2015), the relevant semantic distinction underlying the count/mass morphosyntax is *atomicity*: a noun is atomic iff there exists a minimal unit that has the property denoted by the noun. Thus, the minimal unit of *chair* is a chair, but there is no minimal unit of *mustard*. In languages like English, which have a fully grammaticized count/mass distinction, the relationship between atomicity and morphosyntax is not direct: e.g., *furniture* is atomic yet mass, while *chocolate(s)* can be either mass or count (see Table 1). There is also cross-linguistic variation with regard to which nouns are count vs. mass: e.g., *spinach* is mass in English but count in French; *beans* is count in English but mass in Russian. Such nouns have been labeled *flexible* nouns in the literature: note that nouns can be flexible both across languages (as in the case of *beans* and *spinach*) and within a language (as in the case of *chocolate(s)* or *stone(s)* in English).

diagnostic	count nouns	mass nouns
indefinite article <i>a</i> (count)	√ <i>a chair / chocolate / bean</i>	* <i>a furniture/mustard/spinach</i>
plural marking (count)	√ <i>chairs / chocolates / beans</i>	* <i>furnitures/mustards/spinaches</i>
ability to occur in bare (determiner-less) form (mass)	* <i>I bought chair / bean.</i>	√ <i>I bought furniture / mustard / spinach / chocolate.</i>
<i>many</i> (count) vs. <i>much</i> (mass)	<i>many chairs / chocolates / beans</i>	<i>much furniture / mustard / spinach / chocolate</i>

Table 1: diagnostics for the count/mass distinction in English

Unlike in plural-marking languages, in generalized classifier (GC) languages, where plural marking is optional, the relationship between atomicity and morphosyntax is direct. For example, in Korean, only atomic nouns can combine with the plural marker *-tul* (Kim 2005). In Mandarin, the atomicity distinction is directly encoded in the classifier system (Cheng and Sybesma 1998).

---

\* Thanks to the members of the Illinois Experimental Linguistics group and to the ELM audience for their feedback. This project was funded by an NSF dissertation research improvement grant, BCS-1823762. Authors: Sea Hee Choi, University of Illinois at Urbana-Champaign ([schoi76@illinois.edu](mailto:schoi76@illinois.edu)) & Tania Ionin, University of Illinois at Urbana-Champaign ([tionin@illinois.edu](mailto:tionin@illinois.edu)).

A number of experimental studies, beginning with Barner and Snedeker (2005), have investigated how speakers of both plural-marking languages (English, French) and GC languages (Japanese, Mandarin, Korean) interpret different types of nouns. The present study follows in this tradition, but uses a novel methodology in order to directly examine which nouns are interpreted as atomic or non-atomic across different languages. We ask the following research question: *Does the morphosyntax of the count/mass distinction in a given language affect speakers’ interpretation of nouns as atomic vs. non-atomic? or, Is interpretation driven by semantic universals, independently of language-specific morphosyntax?*

The first objective of our study is methodological: to develop a task that directly asks participants about interpretation, targeting the concept of atomicity, and does not provide any morphosyntactic cues to interpretation. The second objective is theoretical: to examine whether there are differences in the interpretation of flexible nouns between English and GC languages, in the absence of morphosyntactic cues.

**2. Background.** In this section, we first discuss how interpretation of count and mass nouns has been experimentally studied in prior literature, and then move into the specifics of the three languages under investigation in our study.

2.1. PRIOR STUDIES ON THE INTERPRETATION OF COUNT AND MASS NOUNS. Prior studies have used two types of tasks to address the relationship between the object/substance distinction and count/mass morphosyntax: the object/substance rating task and the quantity judgment task.

Barner, Inagaki and Li (2009) used the object/substance rating task with native speakers of English and of Japanese, asking them to rate 100 common nouns with regard to whether they denote objects, substances, both, or neither. All nouns were presented in bare singular forms (no plural marking or determiners). Despite clear morphosyntactic differences between English and Japanese, the two participant groups performed very similarly: nouns that are count in English (e.g., *ball*) were classified as denoting objects in both languages; nouns that are mass in English (e.g., *water*) were classified as substance-denoting in both languages; and many food-denoting nouns (e.g., *pizza*, *banana*) received a mix of ratings, in both languages.

In the quantity judgment task, used in Barner and Snedeker (2005) and many subsequent studies, participants are shown two characters, one of whom has two large objects (e.g., two large chairs or two large blobs of mustard), while the other has six small objects (e.g. six small chairs or six small blobs of mustard), and are asked *"Who has more X?"* The choice of two large objects for the answer corresponds to a judgment by volume, while the choice of six small objects corresponds to a judgment by number. This paradigm has been used in English by Barner and Snedeker (2005), Barner, et al. (2009), Inagaki and Barner (2009), and MacDonald and Carroll (2018), and in French by Inagaki and Barner (2009). The paradigm has also been extended to GC languages: it was used in Japanese by Barner, et al. (2009) and Inagaki and Barner (2009); in Mandarin Chinese by Cheung, Li and Barner (2010); and in Korean by MacDonald and Carroll (2018).

Category	Sample noun	Explanation
1. Object-count	chair	Atomic nouns, count-cross-linguistically
2. Flexible-count	bean	Flexible nouns cross-linguistically, count in English
3. Flexible	chocolate	Flexible nouns, both count and mass in English
4. Object-mass	furniture	Superordinate nouns, mass in English
5. Flexible-mass	spinach	Flexible nouns cross-linguistically, mass in English
6. Substance-mass	mustard	Non-atomic nouns, mass cross-linguistically

Table 2: different noun categories, based on cross-linguistic behavior

Each study investigated some of the noun categories listed in Table 2. Note that categories 2 through 5 in Table 2 are labeled based on their behavior in English, but vary with regard to their count/mass morphosyntax cross-linguistically. In contrast, categories 1 and 6 appear to be universal: clearly bounded objects like *chair* are count in any language which has a count/mass distinction, while substance-denoting nouns like *mustard* are always mass. The findings of the studies that have used the quantity judgment task can be summarized as follows. For categories 1, 4 and 6, there is strong evidence of universality: speakers of both plural-marking and GC languages uniformly judge nouns in categories 1 and 4 by number, and those in category 6 - by volume. In contrast, there is much cross-linguistic variability with regard to categories 3 and 5 (category 2 has not, to the best of our knowledge, been tested with the quantity judgment task). In plural-marking languages, judgments are dependent on the morphosyntax: e.g., in English, *chocolates* (count) is judged by number, while *spinach* and *chocolate* (mass) are judged by volume, whereas in French *spinach* is count and is judged by number. In GC languages, the judgments for these noun types fall in-between. Thus, the results of the quantity judgment task suggest that there is both universality (nouns which are always object-denoting, such as *chair/furniture* are always judged by number) and the effects of language-specific morphosyntax.

A potential limitation of the quantity judgment task is that, in plural-marking languages such as English and French, the task confounds morphosyntax with interpretation: count nouns are presented in plural form ("*Who has more chairs / chocolates?*") while mass nouns are presented in singular form ("*Who has more mustard / chocolate / spinach?*"). In GC languages, on the other hand, all nouns are presented in bare form, with no plural marking. Thus, it is possible that the apparent effect of morphosyntax on interpretation (the finding that *spinach* is judged by volume in English but by number in French) is due to the task format, specifically, to whether the noun appeared in singular or plural form in the task. In light of this, we have developed a new task, the *minimal parts identification task (MPIT)*, which avoids this problem by presenting all nouns in bare singular form. Additionally, the MPIT aims to probe directly into speakers' judgments of atomicity, by asking participants about whether a given entity has minimal parts.

**2.2. THE COUNT/MASS DISTINCTION IN ENGLISH, KOREAN AND MANDARIN.** The three languages examined in the present study are English, Korean and Mandarin Chinese. As discussed above, English is a plural-marking language with an obligatory count/mass distinction. As shown in Tables 1 and 2, the count/mass distinction only partially corresponds to the semantic atomicity distinction. English has nouns like *furniture*, which are mass despite denoting atomic entities. English also has nouns like *string(s)*, *stone(s)*, etc., which have both count and mass variants.

Korean and Mandarin are both GC languages, and there is much debate as to they have a grammaticized count/mass distinction (see Chierchia 1998b, 2010; Cheng and Sybesma 1998). Both languages have plural marking, but the Korean plural marker *-tul* has a much wider distribution than the Mandarin plural marker *-men*. According to Kim (2005), Korean *-tul* is directly related to atomicity, being compatible with atomic nouns but not with non-atomic ones. This claim was supported by experimental findings in Choi, Ionin and Zhu (2018), who found that native Korean speakers accepted *-tul* with object-denoting nouns like *chair* as well as those like *furniture* (categories 1 and 4 in Table 2), but not substance-denoting nouns like *oil* (category 6). At the same time, *-tul* is nearly always optional, as discussed by Kim (2005) and Kwon and Zribi-Hertz (2004), among others, so that a bare singular noun is compatible with both singular and plural interpretations (one exception is definite contexts, where *-tul* is obligatory for plural interpretation). In the case of Mandarin, the plural marker *-men* is restricted to [+human] nouns; for more discussion, see Iljic (1994) and Li (1999). At the same time, according to Cheng and

Sybesma (1998, 1999), the atomicity distinction is reflected in the classifier system, with clear syntactic differences in the behavior of count and mass classifiers.

Thus, we have three languages with differences both in the distribution of plural marking, and the correspondence between atomicity and morphosyntax. If interpretation is at least partially influenced by morphosyntax, then we would expect speakers of English, Korean and Mandarin to exhibit differences in their judgments of atomicity. If, on the other hand, interpretation is universal and independent of morphosyntax, then we would expect very similar behavior from all three language groups. Before we can study interpretation, however, we need to establish exactly how the different noun types in Table 2 behave with regard to morphosyntax. While none of the noun types in Table 2 are compatible with *-men* in Mandarin (since they are all [-human]), it is an open question as to which of these noun types are compatible with *-tul* in Korean. Choi, et al. (2018) found that *-tul* was compatible with the nouns in categories 1 and 4, but not category 6; however, they did not test categories 2, 3 and 5. In this study, we administered a grammaticality judgment task (GJT) to native Korean speakers in order to examine the compatibility of *-tul* with all noun types in Table 2; to allow for a cross-linguistic comparison, we administered the GJT in English as well.

**3. Noun selection.** In this study, we tested the six noun types in Table 2 with regard to both morphosyntax (the GJT, section 4) and interpretation (the MPIT, section 5). Before we present those tasks, we discuss how the nouns for the six categories in Table 2 were selected.

For categories 1 and 6, we selected clearly object-denoting and clearly substance-denoting nouns, which are expected to be universally count and mass, respectively. For category 4, we included superordinate object-denoting nouns like *furniture*, *jewelry*, etc., which we had investigated in prior studies on the second language acquisition of the count/mass distinction; see Choi, et al. (2018) and Choi and Ionin (in press). For category 3, we selected nouns which have both singular and plural variants in English (*string(s)*, *chocolate(s)*, etc.). The main challenge was finding nouns for categories 2 and 5, which are invariably count and mass, respectively, in English, yet have different status in other languages. We created a list of nouns which have the potential to be flexible cross-linguistically (primarily names of various fruit and vegetables, foods and materials), and asked linguists from eight obligatory plural marking languages other than English (Spanish, Russian, German, Greek, Brazilian Portuguese, Polish, French, Basque) to categorize the nouns as either ‘count’, ‘mass,’ or ‘flexible’ in their language. Based on the results of this survey, we selected for category 2 those nouns that are count in English but that were categorized as mass or flexible in at least one plural-marking language; for category 5, we selected those nouns that are mass in English but that were categorized as count or flexible in at least one plural-marking language. A total of eight nouns were selected for each noun type in Table 2, 48 nouns total.

**4. Grammaticality judgment task.** The goal of the GJT was to establish the behavior of plural marking in both English and Korean, in order to relate it to interpretation of different noun types in those languages (see section 5). Mandarin was not tested in the GJT, since, as discussed above, the Mandarin plural marker *-men* is incompatible with [-human] nouns.

4.1. PARTICIPANTS. 20 native speakers of English residing in the U.S. (mean age = 21), and 20 native speakers of Korean residing in South Korea (mean age = 22) completed the GJT. About half of the participants in each group had also completed the MPIT about a month earlier.

The English participants were recruited via Amazon's Mechanical Turk, while the Korean participants were recruited via online advertisement. All the participants were tested online via the Survey Gizmo tool.

4.2. MATERIALS. Prior to the actual test, the sentence frames for the GJT were normed to make sure that they were not biased towards object-denoting or substance-denoting interpretations. The norming was done via Mechanical Turk, with 15 native English speakers who did not take part in the main experiment. The participants in the norming task were asked to read sentence frames containing X in the object position (e.g., *We talked about X at school yesterday*) and judge the likelihood of X being an object or a substance on a scale from 1 (definitely a substance) to 5 (definitely an object). The four sentence frames with the most neutral ratings were selected for the GJT. These were: *We talked about X at school yesterday*; *We read about X in the library yesterday*; *We searched for X online yesterday*; and *I thought about X when I was in the kitchen yesterday*.

For the GJT, 48 target sentences were created; each sentence corresponded to one of the four neutral sentence frames, with the target noun in object position, in place of X. There were 48 target nouns, corresponding to the six conditions in Table 2, eight nouns per condition. Two versions of each sentence were created, one with the singular and the other with the plural form of the noun; no determiners were used with the target nouns. Two experimental lists were created, and the singular and plural versions of each sentence were distributed across the two lists using a Latin-square design. Each list contained 30 filler items in addition to the 48 target items; the items were pseudo-randomized for order of presentation. A sample item for one of the categories is given in (1a) for English, and (1b) for Korean. The plural marker is given in parentheses here; in the actual test, a given sentence either did or did not contain the plural marker.

- (1) *sample item for category 2:*
- a. We talked about bean(s) at school yesterday.
  - b. Wuli-nun ecey      hakkyo-eyse khong-(tul)-eytayhay iyaki hay-ss-ta.  
     We-Top yesterday school-Loc bean-Pl-Dat      about talk-Pst-Decl.

The participants were asked to rate the grammaticality of each sentence on a scale from 1 (not acceptable) to 4 (very acceptable). In English, nouns from categories 1 and 2 were expected to be acceptable in plural form and unacceptable in singular form, while the opposite was expected to be the case for nouns in categories 4, 5 and 6; in category 3, both singular and plural forms are grammatical. In Korean, the bare singular form is always grammatical; based on the results in Choi, et al. (2018), the plural *-tul* form was expected to be more acceptable in categories 1 and 4 than in category 6; its acceptability in categories 2, 3 and 5 was an open question.

4.3. DESCRIPTIVE RESULTS. Figure 1 shows the mean ratings of all six noun types in singular and plural forms in both languages. English speakers performed as expected given English morphosyntax, rejecting the singular form in categories 1 and 2, rejecting the plural form in categories 4 through 6, and accepting both forms in category 3 (though the plural was rated higher than the singular in category 3, the singular form still received a mean rating of about 3 on a 1-to-4 scale). Korean speakers always accepted the bare singular forms, as expected. The plural *-tul* forms were fully acceptable in categories 1 and 4, less acceptable in categories 2 and 3, and received the lowest ratings in categories 5 and 6. In sum, the English and Korean speakers had similar judgments of count/mass morphosyntax with flexible-mass and substance-mass nouns, but differed in the other four categories. We now move on to the question of whether speakers of these languages (as well as of Mandarin) also differ on their interpretation of count vs. mass nouns.

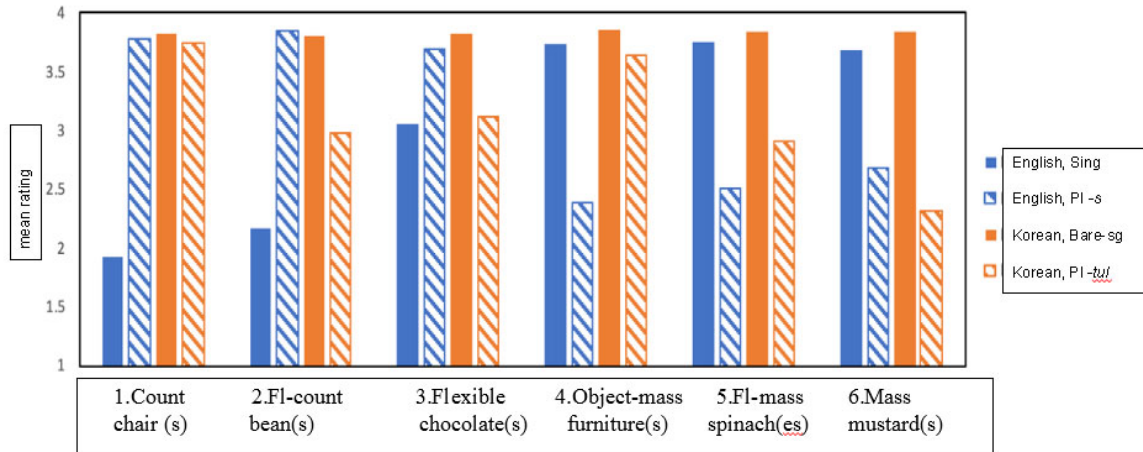


Figure 1: GJT results for English and Korean: mean ratings

**5. Minimal parts identification task.** The MPIT is a new task specially devised to study atomicity. As discussed earlier, the main advantage of the MPIT over the quantity judgment task from Barner and Snedeker (2005) is that the MPIT asks about atomicity (minimal parts) more directly and does not provide morphosyntactic cues, thus addressing interpretation rather than morphosyntax.

5.1. PARTICIPANTS. 20 native speakers of English in the U.S. (mean age = 21), 20 native speakers of Korean in South Korea (mean age = 23) and 20 native speakers of Mandarin Chinese in China (mean age = 20) completed the MPIT. The English speakers were recruited via Amazon’s Mechanical Turk; the Korean speakers and the Mandarin speakers were recruited in online university communities, via advertisement. All participants were tested online via the Survey Gizmo tool.

5.2. MATERIALS. The MPIT materials were first created in English, and then translated into Korean and Mandarin. Participants first read the instructions in (2a); for each test item, they responded first to the question in (2b); if they answered Yes to this question, they were asked to provide the name of the minimal unit, in response to (2c). Only responses to (2b) are reported in this paper; responses to (2c) have not yet been analyzed. The same 48 nouns were used in the MPIT as in the GJT (the six categories in Table 2, eight nouns per category). The 48 items were pseudo-randomized for order of presentation. Each noun appeared in singular bare form, as shown in (2b). There was only one experimental list.

(2) a. Instructions (English version):

When we see something, we can sometimes think of its minimal (smallest) unit. For example, the minimal unit of *table* is *a table*: if you divide *a table* in half, it cannot function as *a table* anymore. However, the minimal (smallest) unit of *water* is vague (it is unclear what the minimal (smallest) unit is): if we divide *water* in half, it will still be *water*. For each item in this task, you will see a word and two questions about the given word. In the first question, please indicate whether you can think of the minimal (smallest) unit for this word, by clicking either ‘yes’ (you can think of the minimal unit, as with *table*), or ‘no’ (you cannot think of the minimal unit, or the minimal unit is vague, as with *water*). In the second question, you will see another question which will ask what is the minimal (smallest) unit of the given object/substance. If you answered “yes” to the first question, then, please type what you think the minimal unit is. For

example, if you see the word “*table*”, you will click ‘yes’ and type ‘*a table*’. If you answered “no” to the first question, please type “N/A” “vague” or “none” in the second question.

- b. Does *chair* have a minimal unit? (Yes)/(No)
- c. If yes, what is the minimal (smallest) unit?

5.3. PREDICTIONS. As discussed in section 2.1, prior studies with the quantity judgment task have found striking uniformity with regard to nouns in categories 1, 4 and 6. We expect the same universality to be manifested in the MPIT: the responses to the question in (2b) should be primarily Yes (there is a minimal unit) for categories 1 and 4 and primarily No (there is no minimal unit) for category 6, in all three languages.

With regard to the other three categories, there are two possibilities. One possibility is that morphosyntax drives interpretation, as was found in studies using the quantity judgment task. In that case, English speakers should give primarily Yes responses to category 2, and primarily No responses to category 5. It is less clear how they would respond to category 3 (flexible nouns like *chocolate(s)*). Since all nouns are presented in bare form in the MPIT, participants see *chocolate* (singular, mass) rather than *chocolates* (plural, count), and, under the influence of morphosyntax, would therefore be more likely to give a No response. Alternatively, English speakers may consider both count and mass interpretations of the noun, and therefore give a mix of Yes and No responses. Moving on to Korean speakers, Figure 1 shows that for categories 2, 3 and 5, they rated the singular form higher than the plural, with no obvious differences among the three categories. Therefore, if morphosyntax drives interpretation, Korean speakers should behave about the same on categories 2, 3 and 5; the same holds for Mandarin speakers, for whom all three categories are equally incompatible with the plural marking *-men*. Exactly what response type might be expected from the Korean and Mandarin speakers on categories 2, 3 and 5 is an open question, but crucially, they should behave differently than English speakers, who should distinguish among the three categories.

Alternatively, it is possible that interpretation is universal and largely independent of morphosyntax, and that the cross-linguistic differences obtained on flexible nouns with the quantity judgment task (Barner et al. 2009, Inagaki and Barner 2009) was due to the specific nature of that task, in which count nouns were presented in plural form but mass nouns in singular form in English, while all nouns were presented in bare singular form in GC languages. In that case, we expect to find no cross-linguistic differences for any category in the MPIT, where all nouns are presented in bare singular form in all three languages.

5.4. RESULTS. Figure 2 presents the MPIT results, as the percentage of Yes (there is a minimal unit) responses to each noun type. Overall, the three groups showed very similar performance. At the same time, compared to English speakers, speakers of Korean and Mandarin showed an overall higher proportion of Yes responses to the minimal parts question in almost every condition, with Mandarin speakers showing the highest proportion of Yes responses among all three groups.

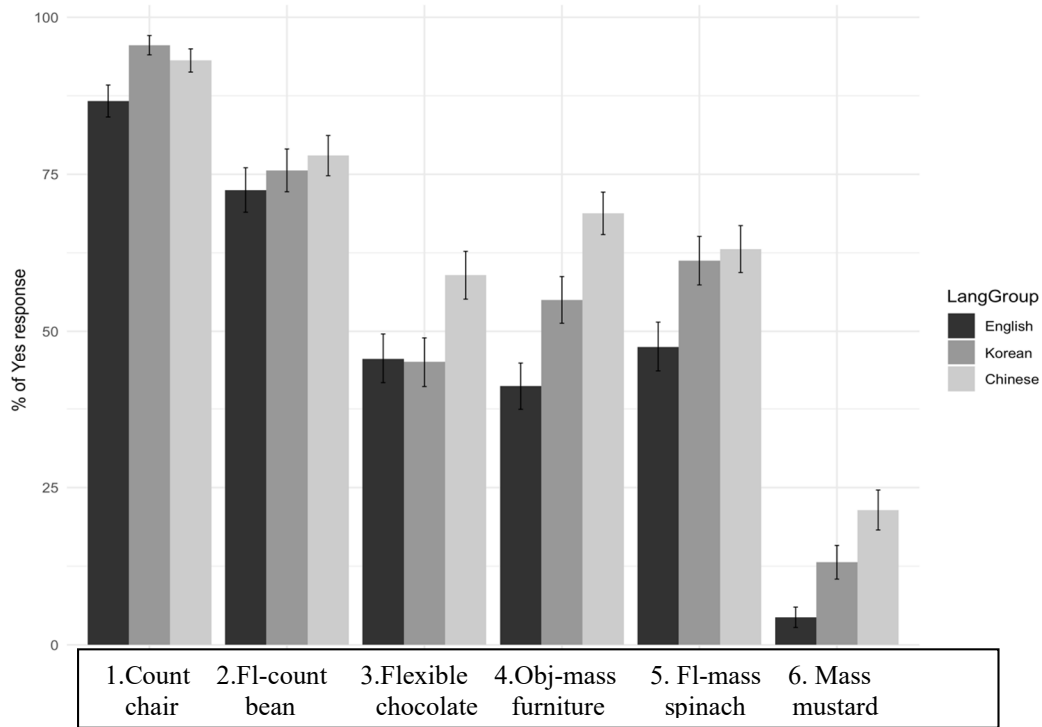


Figure 2: MPIT results (% of Yes, there is a minimal unit responses)

For the statistical analysis, the dependent variable (the response to (2b)) was coded with 1 for Yes and 0 for No. The data were analyzed using a mixed effect logistic regression model in R, see Bates, Mächler, Bolker and Walker (2014). The following fixed effects were included in the model: language group (3 levels) and category (6 levels). Helmert coding was used for the variable of language group: the English group was compared to the full GC group; subsequently, the Korean and Mandarin groups were compared to each other. This allowed us to see a comparison between plural-marking and GC languages, as well as between two different GC languages. The variable of category was coded with contrast coding. An interaction between group and category was also introduced in the model. Subjects and items were included as random effects. Following Barr, Levy, Scheepers and Tily (2013), maximal models were created but since the maximal models did not converge the model was gradually reduced; the final model included by-subject and by-item random intercepts. The model output is given in Table 3. The performance of the English group differed significantly from that of the GC group, while the Korean and Mandarin groups did not differ. There were also significant effects of category for most of the category comparisons. The interactions between group and category were significant for several of the English/GC comparisons, and marginal for one of the Korean/Mandarin comparison.

In order to explore the sources of the interactions, Bonferroni pairwise comparisons were conducted using the *emmeans()* function in R, see Lenth (2020) (the Bonferroni correction for multiple comparisons is automatically implemented in R) as well as a visual examination of the interaction plots using the *plot\_model()* function, see Gelman (2008). With regard to cross-group comparisons, differences were noted only in category 4 (object-mass nouns), where Mandarin speakers gave significantly more Yes responses than English speakers, and marginally more Yes responses than Korean speakers; and in category 5 (flexible-mass nouns), where Mandarin speakers gave marginally more Yes responses than English speakers. There were no group differences on any of the other categories.

Fixed effects	Coefficie	SE	Z	P-value
Intercept	.41	.17	2.37	.01*
Group (English vs. GC)	1.14	.41	2.81	<.01*
Group (Korean vs. Mandarin)	.35	.35	1.01	.31
Category (1 vs. 2-6)	2.48	.16	15.81	<.01*
Category (2 vs. 3-6)	.92	.11	8.53	<.01*
Category (3 vs. 4-6)	-.41	.10	-4.19	<.01*
Category (4 vs. 5-6)	-.13	.10	-1.31	.19
Category (5 vs. 6)	-.05	.10	-.47	.64
Group (English vs. GC) x Category (1 vs. 2-6)	.48	.40	1.20	.23
Group (Korean vs. Mandarin) x Category (1 vs. 2-6)	-.66	.42	-1.58	.11
Group (English vs. GC) x Category (2 vs. 3-6)	-.61	.30	-2.05	.04*
Group (Korean vs. Mandarin) x Category (2 vs. 3-5)	-.10	.27	-.36	.72
Group (English vs. GC) x Category (3 vs. 4-6)	-.69	.27	-2.53	.01*
Group (Korean vs. Mandarin) x Category (3 vs. 4-6)	.35	.24	1.45	.15
Group (English vs. GC) x Category (4 vs. 5-6)	.26	.27	.99	.32
Group (Korean vs. Mandarin) x Category (4 vs. 5-6)	.41	.24	1.75	.08+
Group (English vs. GC) x Category (5 vs. 6)	-.13	.28	-.48	.63
Group (Korean vs. Mandarin) x Category (5 vs. 6)	-.21	.24	-.88	.38
Random Effects	Variance	SD		
Subject	1.07	1.03		
Item	.08	.29		

\* significant at  $p < .05$

+ marginal at  $.05 < p < .10$

Table 3: Model output for MPIT data

Furthermore, the three groups showed fairly similar patterns, but with some variations. All three groups gave significantly more Yes responses to category 1 (object-count nouns) than to any of the other categories, and significantly fewer Yes responses to category 6 (substance-mass nouns) than to any of the other categories. Categories 2 through 5 patterned in between, but with some variability across groups. Out of these four categories, all three groups gave the most Yes responses to category 2 (flexible-count).

**6. Discussion.** The MPIT results show that there are many similarities among the three groups. The highest rates of ‘Yes, minimal unit’ responses obtained for count nouns, and the lowest - for substance-mass nouns, while the other four noun types patterned in between, with slight variations across languages. The English and Korean groups did not differ on any category, despite clear differences in morphosyntax. While the Mandarin group gave more ‘Yes, minimal unit’ responses overall, this reached significance only in the object-mass and flexible-mass categories.

6.1. COMPARISON TO THE QUANTITY JUDGMENT TASK. We now examine how the results of the MPIT compare to the findings of the quantity judgment task in prior studies (see section 2.1). For object-count and substance-mass nouns (categories 1 and 6), the tasks give extremely similar results. Object-count nouns are judged by number in the quantity judgment task, and receive predominantly Yes responses in the MPIT; substance-mass nouns are judged by volume in the quantity judgment task, and receive predominantly No responses in the MPIT. In both tasks, performance is very similar across languages.

On the other hand, differences across tasks emerge with regard to category 4, object-mass nouns like *furniture*. In the quantity judgment task, such nouns were uniformly judged by number, across languages. We expected predominantly Yes responses to this category in the MPIT, but instead got a split between Yes and No responses. We believe that the issue had to do with how the participants interpreted the question about minimal units. The intent was for participants to think along the lines of "When we divide furniture, at some point it stops being furniture: a chair or a table is furniture, but a chair leg or half of a table is no longer furniture. Therefore, furniture has a minimal unit, which is a single piece of furniture". However, it is quite possible that participants interpreted 'minimal unit' as 'minimal unit which is the same across contexts'; while the minimal unit of *chair* is always a chair, the minimal unit of *furniture* could be a chair, or a table, or a couch, etc., and this could be why participants often responded to No. This task confound is likely the reason for why object-mass nouns behaved differently in the MPIT than in the quantity judgment task.

Turning next to the three types of flexible nouns (categories 2, 3 and 5), prior findings with the quantity judgment task found effects of language-specific morphosyntax: English speakers judged *chocolate* and *spinach* by volume, and *chocolates* by number, while speakers of GC languages tended to pattern in between on both *chocolate* and *spinach*. In contrast, in the MPIT, we found that the three language groups performed very similarly on both *chocolate* and *spinach* (despite slightly higher Yes responses in the case of the Mandarin group), as well as on *bean*. This suggests that the cross-linguistic differences regarding flexible nouns found with the quantity judgment task were likely due to the morphosyntactic form in which the nouns were presented in that task. When all nouns are presented in bare form, we see much similarity across languages.

6.2. UNIVERSALITY OR EFFECTS OF MORPHOSYNTAX? The high degree of similarity across the three languages suggests that semantics drives morphosyntax, rather than the other way around. As shown in section 4, English and Korean differ on most categories tested with regard to compatibility with plural marking; and Mandarin does not allow plural marking with any inanimate, [-human] nouns. Yet, with regard to interpretation, the results obtained on the MPIT in section 5 are very similar. Object-denoting nouns are judged as atomic, and substance-denoting ones - as non-atomic. The one exception is object-mass nouns like *furniture*, where, as discussed above, the MPIT likely did not succeed in targeting judgments about minimal units.

Turning to flexible nouns, we found that all three groups gave more Yes responses to category 2 (flexible-count) than to categories 3 (flexible) and 5 (flexible-mass). For English speakers, this is fully consistent with the morphosyntax, since category 2 nouns are obligatorily count, while category 3 and 5 nouns are not. However, morphosyntax cannot explain the performance of Korean and Mandarin speakers on these categories. As shown in Figure 1, category 2 nouns behave almost identically to category 3 and category 5 nouns in Korean, with regard to their compatibility with the plural marker; and in Mandarin, none of these categories allow the plural marker. And yet, with regard to interpretation, the three language groups behaved very similarly.

The one place where we did find cross-linguistic differences was in the overall greater likelihood of Yes responses in the Mandarin group; the higher rates of Yes responses were especially pronounced in categories 3 through 5, but even category 6 (substance-mass nouns) elicited numerically more Yes responses from the Mandarin group than from the other two groups. It is possible that this difference is indeed driven by morphosyntax. As discussed above, while Mandarin has highly restricted use of plural marking, it uses classifiers to make the distinction between atomic (count) and non-atomic (mass) nouns. A detailed investigation of which classifiers in Mandarin are compatible with the nouns tested could shed light onto the relationship between morphosyntax and interpretation in Mandarin.<sup>1</sup>

**7. Conclusion and future directions.** The findings of this study provide evidence that when morphosyntactic cues are absent, noun interpretation is remarkably similar across very different languages. Thus, we can conclude that interpretation drives morphosyntax, rather than the other way around. In languages that grammatically encode the count/mass distinction, atomic, object-denoting nouns are count, whereas non-atomic, substance-denoting nouns are mass. Nouns that can potentially denote either objects or substances, and therefore be perceived as either atomic or not, are flexible within and/or across languages.

Future plans include analyze the qualitative part of the results (responses to question 2c), in order to examine what participants perceive to be minimal units of substances (e.g., when they respond that *mustard* has a minimal unit), and of entities denoted by flexible nouns (e.g., do English speakers consider the minimal unit of *stone* to be a single stone, or a piece of the stone substance?). An analysis of Mandarin classifiers, and how they correspond to interpretation, would also be quite fruitful, as discussed in the previous section.

Given that the MPIT ran into problems with object-mass nouns like *furniture*, another plan for future research would be to modify the MPIT instructions to make it clear that the minimal unit need not always be the same. An additional question could ask about atomicity in a slightly different way: e.g., *If we divide X into 50 parts, will each part still have the property of being X?* For substances like *mustard* or *water*, the answer is Yes. But for *furniture*, the answer is No, not necessarily: we might have 50 pieces of furniture, but we might also have 50 random furniture-parts which are not furniture.

Finally, we acknowledge that the MPIT is a highly metalinguistic task, more so than the quantity judgment task. It would be worthwhile to develop a less metalinguistic version of the MPIT, perhaps one that uses pictures, as the quantity judgment task does. There is still much to be learned about how speakers of different languages perceive entities, and what this means for the count-mass distinction.

## References

- Barner, David, Shunji Inagaki & Peggy Li. 2009. Language, thought, and real nouns. *Cognition* 111(3). 329-344. <https://doi.org/10.1016/j.cognition.2009.02.008>.
- Barner, David & Jesse Snedeker. 2005. Quantity judgments and individuation: evidence that mass nouns count. *Cognition* 97. 41-46. <https://doi.org/10.1016/j.cognition.2004.06.009>.
- Barr, D. J., R. Levy, C. Scheepers & H. Tily. 2013. Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255-278. <https://doi.org/10.1016/j.jml.2012.11.001>.

---

<sup>1</sup> The Mandarin speakers tested on the MPIT in our study also completed a classifier elicitation task (this was administered after the MPIT, and therefore could have no effect on MPIT performance); the results of this task have not yet been analyzed.

- Bates, Douglas M., Martin Mächler, Ben Bolker & Steve Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. *Journal of Statistical Software* 67(1). 47pp. .
- Cheng, Lisa L.-S. & Rint Sybesma. 1998. yi-wan tang, yi-ge tang: classifiers and massifiers. *The Tsing Hua Journal of Chinese Studies* 28(3). 385-412. .
- Cheng, Lisa L.-S. & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30. 509-542. <https://doi.org/10.1162/002438999554192>.
- Cheung, Pierina, Peggy Li & David Barner. 2010. Individuation and quantification: Do bare nouns in Mandarin Chinese individuate? In L. E. Clemens and C.-M. Liu (eds.), *Proceedings of the 22nd North American conference on Chinese Linguistics & the 18th International Conference on Chinese Linguistics*, 395-412. Cambridge, MA: Harvard University.
- Chierchia, Gennaro. 1998a. Plurality of mass nouns and the notion of "semantic parameter". In S. Rothstein (eds.), *Events and Grammar*, 53-103. Dordrecht: Kluwer.
- Chierchia, Gennaro. 1998b. Reference to kinds across languages. *Natural Language Semantics* 6. 339-405. <https://doi.org/10.1023/A:1008324218506>.
- Chierchia, Gennaro. 2010. Mass nouns, vagueness and semantic variation. *Synthese* 174. 99-149. <https://doi.org/10.1007/s11229-009-9686-6>.
- Chierchia, Gennaro. 2015. How universal is the mass/count distinction? Three grammars of counting. In A. Li, A. Simpson and W.-T. D. Tsai (eds.), *Chinese syntax in a cross-linguistic perspective*, 147-175. Oxford: Oxford University Press.
- Choi, Sea Hee & Tania Ionin. in press. Plural marking in the second language: Atomicity, definiteness, and transfer. *Applied Psycholinguistics*. <https://doi.org/10.1017/S0142716420000569>.
- Choi, Sea Hee, Tania Ionin & Yeqiu Zhu. 2018. L1 Korean and L1 Mandarin L2 English learners' acquisition of the count/mass distinction in English. *Second Language Research* 34(2). 147-177. <https://doi.org/10.1177/0267658317717581>.
- Gelman, A. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* 27. 2865-2873. .
- Iljic, Robert. 1994. Quantification in Mandarin Chinese: two markers of plurality. *Linguistics* 32. 91-116. <https://doi.org/10.1515/ling.1994.32.1.91>.
- Inagaki, Shunji & David Barner. 2009. Countability in absence of count syntax: Evidence from Japanese quantity judgments. In S. Inagaki, M. Hirakawa, Y. Hirakawa, H. Sirai, S. Arita, H. Morikawa, M. Nakayama and J. Tsubakita (eds.), *Papers from the Eighth Annual Conference of the Japanese Society for Language Sciences*. Studies in Language Sciences, 111-125. Tokyo, Japan: Kurosio Publishers.
- Kim, C. (2005). The Korean plural marker *tul* and its implications. Ph.D. thesis: University of Delaware.
- Kwon, Song-Nim & Anne Zribi-Hertz. 2004. Number from a syntactic perspective: why plural marking looks 'truer' in French than in Korean. In O. Bonami and P. Cabredo Hofherr (eds.), *Empirical issues in formal syntax and semantics* 5, 133-158.
- Lenth, R. 2020. emmeans: Estimated marginal means, aka least-squares means, aka least-squares means. *R Package* version 1.4.4. .
- Li, Yen-Hui Audrey. 1999. Plurality in a classifier language. *Journal of East Asian Linguistics* 8. 75-99. <https://doi.org/10.1023/A:1008306431442>.
- MacDonald, Danica & Susanne Carroll. 2018. Second-language processing of English mass-count nouns by native-speakers of Korean. *Glossa: a journal of general linguistics* 3(1). 46, 41-27. <http://doi.org/10.5334/gjgl.363>.