

Conceptual and Linguistic Factors Affect Entity Processing and Labeling

Sarah Hye-yeon Lee (sarahhl@sas.upenn.edu)
Department of Linguistics, University of Pennsylvania
Philadelphia, PA 19104-6228 USA

Lea Mangifesta (leamangi@sas.upenn.edu)
Department of Linguistics, University of Pennsylvania
Philadelphia, PA 19104-6228 USA

Anna Papafragou (anna4@sas.upenn.edu)
Department of Linguistics, University of Pennsylvania
Philadelphia, PA 19104-6228 USA

Abstract

How do learners acquire the meanings of nouns? Given the complex linguistic and non-linguistic input present in the learning environment, how do learners identify the concepts denoted by nouns? In other words, how does a learner map the input language to non-linguistic concepts? In the current study, we focus on the case of mass-count language and physical entities (e.g., objects and substances). We conduct novel word extension experiments to investigate whether conceptual and linguistic factors universally affect label extension to restructured entities in languages that do and do not have a grammatical mass-count distinction (Experiment 1: English; Experiment 2: Korean, respectively). We find that objecthood and linguistic (count/mass) context both modulate how speakers extend labels to restructured novel entities.

Keywords: object; substance; ontological category; mass-count; word learning; label extension; cross-linguistic analysis

Introduction

The ontological distinction between objects and substances is one of the most fundamental distinctions that the human mind makes of the physical world (e.g., Prasada, Ferenz, & Haskell, 2002). Objects refer to individuals: they can be counted and individuated (e.g., vase, table). Substances, on the other hand, are “stuff” that cannot be counted in the same way (e.g., clay, wood). The nature of the link between count-mass language and nonlinguistic ontological categories has been debated within linguistics, philosophy, and psychology. Earlier researchers have proposed that learning mass-count syntax provides a foundation for how learners perceive objects and substances. Quine (1960) suggested that the differentiation between mass and count categories could facilitate children in acquiring conceptual-semantic knowledge related to physical objects, including quantification and individuation. According to Quine, count nouns, but not mass nouns, “possess built in modes, however arbitrary, of dividing their reference” (p. 91). Related to this proposal, developmental studies have demonstrated that children are likely to associate count nouns with objects, and mass nouns with substances (e.g., Brown, 1952; Dickinson, 1988; Imai, Gentner, & Uchida, 1994; Landau, Smith, & Jones, 1988; Samuelson & Smith, 1999; Soja, 1992). For

example, in Brown (1952), 3- to 5-years old children were presented with a scene that depicted “a pair of hands performing a kneading sort of motion, with a mass of red confetti like material ... piled into a blue-and-white striped container.” The scene was described in a count noun frame (“Do you know what a sib is? In this picture you can see a sib”), a mass noun frame (Do you know what sib is? In this picture you can see some sib.”), or a verb frame (“Do you know what it means to sib? In this picture you can see sipping”). When children were asked to map the new word onto either an object-match, a substance-match, or an action-match scene, they strongly preferred the object-match when the word was first presented to them as a count noun, the substance-match when presented with the mass noun, and the action-match when the word was presented in a verb frame.

However, many developmental psychologists have challenged the Quinian idea, and suggested instead that children’s acquisition of count and mass nouns is scaffolded by a preverbal understanding of the distinction between objects and substances. MacNamara (1972, 1982) proposed that the acquisition of mass and count categories is facilitated by conceptual and semantic knowledge obtained through preverbal categories like “object” and “substance” (see also Barner & Snedeker, 2005). Indeed, children are sensitive to the object-substance distinction from very early on, before acquiring language (e.g., Spelke, 1995; Hespos et al., 2009; vanMarle & Scholl, 2003; Imai & Gentner, 1997).

When learning new words, children rely on the ontological distinction between objects and substances. For example, 2-year olds learning English are sensitive to the object-substance status of a referent during word learning, prior to acquiring mass-count syntax: When novel nouns labeled novel objects and non-solid substances in neutral syntax (e.g., “This is my blicket”), children extended object labels to other entities of the same shape and extended substance labels to other entities of the same material (Soja, Carey, & Spelke, 1991). Similar findings have also been reported for children who are learning languages that lack mass-count syntax. Like English-learning children, Japanese-learning children and Mandarin-learning children distinguish solid objects from nonsolid substances when extending new nouns (Imai & Gentner, 1997; Li, Dunham, & Carey, 2009). Thus, it seems

clear that the development of ontological categories such as “object” and “substance” precede the acquisition of the mass-count distinction.

However, the ontological categories of “object” and “substance” do not exhaust the interpretation of mass-count semantics. Instead, children’s knowledge of noun phrase semantics is rooted in individuation, rather than in the object-substance distinction. Indeed, evidence suggests that children have an abstract notion of “individual” from very early in language acquisition, as expressed in the content of their early nouns, and in their ability to count abstract individuals such as sounds, actions, and holes (Starkey, Spelke, & Gelman, 1990; Wynn, 1990; Giralt & Bloom, 2000). Studies on language development provide results along the same lines—children are sensitive to syntactic information when learning nouns that refer to sounds (Bloom, 1994), puddles (Soja, 1992) and collections of things (Bloom & Keleman, 1995). Thus, children appear to have a rich understanding of individuation across both physical and abstract domains. These abilities suggest that children converge on the adult interpretation of count noun semantics almost from the beginning, and use count nouns to quantify over individuals (i.e., count noun → individual) (Bloom, 1999). Thus, both objecthood and the semantics of count nouns seem to be scaffolded by an abstract notion of individuation. Understanding the precise nature of individuals is therefore crucial in investigations of the noun learning question.

Conceptual Signatures of Individuals and Commitment to Structure

What exactly are the conceptual differences between individuals like objects and non-individuals like substances? It has been proposed that conceptual objecthood can be characterized in terms of the manner in which we think about the entity’s structure (Prasada, Ferenz, & Haskell, 2002). Prasada, Ferenz, & Haskell (2002) proposed that objects, but not substances, possess non-arbitrary structure; the cues that indicate non-arbitrariness of structure being regularity of structure, repetition of structure, and the existence of structure-dependent functions. These three properties led participants to describe novel entities using count syntax (*There is a blicket*) as opposed to mass syntax (*There is blicket*), and therefore pointed to objecthood (as opposed to substancehood). However, it is unclear whether viewers are sensitive to an entity’s conceptual structure independent of language.

In recent work, Lee, Ji, and Papafragou (Submitted) propose that individuated entities including spatial objects are characterized by having a well-defined internal structure. In other words, objects have spatial parts that are arranged in a designated spatial configuration. A table, for example, has parts like the tabletop and legs that are arranged in a certain way. A principle that follows from this proposal is that objects, but not substances, will resist changes to their structure. That is, a table leg cannot be placed above the tabletop – the resulting entity may not count as a table anymore. However, sand can be played around with and

would still count as sand. Using a series of non-linguistic tasks, Lee, Ji, and Papafragou show that viewers are indeed more sensitive to structural disruptions and restructurings to objects than to substances, supporting their proposal that objects possess a well-defined internal structure and that this abstract principle is available to viewers independent of language. Thus, it seems that adult viewers are more strongly committed to the structure of individuals than that of non-individuals.

How do these conceptual factors interact and combine with linguistic factors during word learning and affect a learner’s commitment to entity structure? In the present study, we are interested in how conceptual and linguistic factors affect the referential scope of nouns in terms of entity structure. We examine the role of entity type (object vs. substance) and linguistic context (count vs. mass) in the labeling and processing of restructured entities. We test this with English speakers (Experiment 1) and Korean speakers (Experiment 2). We predict that entity type (object vs. substance) would modulate how speakers extend labels to restructured entities: if objects indeed possess a designated internal structure, speakers would be less willing to extend labels to restructured objects than to restructured substances. We also predict that linguistic (count/mass) context would modulate how speakers extend labels to restructured entities: if count syntax supports individuated entity construal, speakers would be less willing to extend count labels to restructured entities, and the opposite would hold for mass labels.

Experiment 1

We conducted a word extension task to investigate how objecthood and count-mass context affect how English speakers generalize labels to structurally disrupted entities.

Participants

We recruited 40 adult native English speakers from Prolific.


Stimuli

Visual Stimuli We used 16 pairs of images, each depicting a familiar object (e.g., toilet paper roll) and substance (e.g., some toilet paper). In 10 pairs, the object was the artifact made from the substance counterpart (e.g., vase-clay), and in 2 pairs, the object was a natural kind and the substance was an artifact made from the object counterpart (e.g., onion-chopped onion). In the remaining 4 pairs, both the object and the substance were artifacts (e.g., toilet paper roll-pile of toilet paper). We created spatially restructured versions of each entity by switching the positions of the second and third vertical strip of the image (see Table 1).

The original stimuli came from a pool of images that were normed in a manner similar to Li, Dunham, and Carey’s (2009) Experiment 3, where participants were asked to rate the entities in their original (not restructured) form on a scale of 1-7, with 1 being a good object and 7 being a good substance. The stimuli were rated by fifteen naïve native English speakers that did not participate in any of the other experiments reported in this study. Items categorized as

objects had a mean rating of 2.62 (SD=2.25), and items categorized as substances had a mean rating of 4.81 (SD=2.31), with people reasonably rating substances higher than objects on our response scale ($t(478)=-10.55, p<.001$). Following the rating scales in Li et al. (2009), we additionally tested the stimuli on several features that have been associated with objecthood or lack thereof (see section 1.1): the complexity of their overall shape and outline (1=not at all complex; 7=extremely complex), the degree to which their function depended on their overall shape and outline (1=not at all dependent; 7=extremely dependent), as well as their cohesiveness/solidity (1=not at all cohesive/solid; 7=extremely cohesive/solid). Items categorized as objects had a higher complexity rating ($M=4.12, SD=1.95$) than substances ($M=3.34, SD=1.90$) ($t(478)=4.45, p<.001$). Moreover, object stimuli ($M=5.22, SD=1.84$) were rated higher than substance stimuli ($M=3.68, SD=2.02$) in terms of shape-dependent function. Similarly, object stimuli ($M=5.04, SD=1.86$) were rated higher than substance stimuli ($M=3.81, SD=2.03$) in terms of cohesiveness ($t(474)=6.90, p<.001$). These results are consistent with Li et al.’s findings and confirm our choice of items.

Table 1. Sample entity images (Experiments 1-2)

	original	restructured
object		
substance		

Linguistic Stimuli In addition to manipulating the entity type that participants are shown, we manipulated the linguistic context that accompanied each item. We had two conditions: count syntax and mass syntax. We created 16 nonce English nouns, all of which were one or two syllables/characters. On each trial, we introduced each novel label (“alien word”) in a full sentence like “You will see { a_{count} gorp/ \emptyset_{mass} gorp}.” before showing them the first image. After displaying the second image, we used the novel word again in a sentence with the same syntax (“Was that also a_{count} gorp/ \emptyset_{mass} gorp?”). In the count syntax condition, the label was preceded by *a/an* and in the mass syntax condition, the label was unmarked.

Procedure

The experiment was hosted online on PennController IBEX (Zehr & Schwarz, 2018), and participants completed them remotely via the internet. They were told that they will be learning some alien words from an alien language and that the alien is very picky and strict about language use. At the beginning of each trial, a fixation cross was displayed. After the fixation cross, participants saw nonsense labels (“alien words”) in either count or mass syntax (“You will see {a gorp/gorp}”) for 5000ms. This was followed by a brief

(100ms) exposure to the original entity. After a 2000ms mask, they were shown the restructured entity for 100ms and were asked to decide if it could also be labeled with the same nonsense noun (“Was that also {a gorp/gorp}?”).

Results

The proportion of the participants’ Yes responses (likelihood of word extension) was analyzed using Generalized Linear Mixed Effects models (*glmer*). Results are plotted in Figure 1. Participants were more likely to extend the novel noun to the structurally disrupted entity in the Substance condition than in the Object condition ($z=-3.485, p<.001$). Linguistic Context Moreover, the likelihood of extending the novel noun to the disrupted entity was indeed modulated by linguistic context: participants were more likely to label the structurally disrupted entity with the novel noun when the noun was introduced with mass syntax than with count syntax ($z=-2.36, p<.05$).

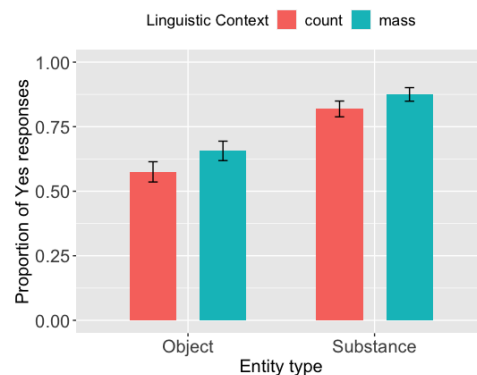


Figure 1: Proportion of Yes responses in Experiment 1

Discussion

We conducted a word extension experiment to investigate how objecthood and linguistic (count-mass) information affect learners in deciding whether to extend novel words to restructured instances of the referred entity. Our results illustrate two main findings. First, learners draw on objecthood information during novel word learning and extension: they are more likely to extend the label to restructured entities when the label is applied to substances than when it was applied to objects. This finding is in line with the proposal that learners have a stronger commitment to structure in learning names for individuated entities (objects). Moreover, we find that linguistic (count-mass) context mediates the likelihood of word extension to the restructured entity. When the noun is introduced in mass syntax, learners are more likely to extend the noun to restructured instances of the entity as opposed to when the noun is introduced in count syntax. That is, mass-count syntax affects how learners decide the scope of the referential domain of the same entity.

What is particularly interesting is what happens in conditions when the linguistic context diverges from the count-mass status of the English noun (Object-Mass or Substance-Count conditions), given that our adult

participants were exposed to familiar entities that they already know English words for. For example, when participants saw a picture of a toilet paper roll, but was introduced to it in mass context (e.g., “You will see gorp.”) or when they saw a picture of some toilet paper, but was introduced to it in count syntax (e.g., “You will see a gorp.”), the given linguistic context diverges from the count-mass status of the English noun. In the Object-Mass condition, participants were more likely to extend the novel label to the restructured entity than in the Object-Count condition, and in the Substance-Count condition, participants were less likely to extend the novel label to the restructured entity than in the Substance-Mass condition. This suggests that even for familiar entities, a novel word learning context encourages participants to capitalize on count-mass information, in a way that count context encourages stronger commitment to entity structure, and vice versa for mass context. This effect of linguistic context also provides us with the confirmation that participants were indeed treating the task as a novel word learning study. If not, the likelihood of label extension may not be as systematically amenable to effects of linguistic contexts as in our data.

Our data also suggests that entity type has a stronger effect on the likelihood of label extension than linguistic context. That is, the effect of linguistic context does not override the effect of entity type – mass context with object stimuli does not bring the likelihood of label extension as high as the Substance conditions, and count context with substance stimuli does not bring the likelihood of label extension as low as the Object conditions. These results suggest that the conceptual distinction between entity types has a stronger effect than linguistic context on word extension to entities with non-canonical structure.

Experiment 2

In Experiment 2, we investigate whether speakers of a classifier language (Korean) also capitalize on entity type and linguistic context information to determine whether or not to extend a novel word to a restructured entity.

Classifier Languages vs. Non-classifier Languages

The core ontological distinction between objects and substances is grammatically encoded in many natural languages. Previous traditional research has drawn a clear distinction between two classes of languages: those that grammatically encode entity type, known as mass-count languages or non-classifier languages, and those that do not, referred to as classifier languages (e.g., Quine, 1969; Lucy, 1992; Chierchia, 1998; Borer, 2005; cf., Doetjes, 1997; Cheng & Sybesma, 1999; Yi, 2010; Strickland, 2015; Kim, 2021; Yi, 2021).

Many authors have proposed that in mass-count languages, count nouns denote objects and mass nouns denote substances (e.g., Bloom, 1999; Gordon, 1985). In English, for example, objects are usually referred to with count nouns (e.g., *a table*) and substances are typically described by mass nouns (e.g., *wood*). However, the count-mass linguistic

distinction does not correlate perfectly with the conceptual object-substance distinction; for example, both the count noun *cows* and the mass noun *cattle* can be used to refer to the same groups of objects. It appears, therefore, that count nouns denote individuals but mass nouns are unspecified for individuation (and could, under certain circumstances, individuate). Indeed, studies of quantity judgments in 4-year-olds and adults demonstrate that some mass nouns (*furniture*) do denote individuals (Barner & Snedeker, 2005).

However, not all languages distinguish between count nouns and mass nouns. Classifier languages such as Korean, Japanese, or Mandarin use numeral classifiers to numerically quantify any noun. In these languages, classifiers are used regardless of whether the noun refers to an object or a substance, and they often provide specifics regarding the entity’s shape, functionality, or animacy (Strickland, 2016). According to the traditional view, classifier languages treat all nouns alike irrespective of the ontological status of what is denoted by the noun. In other words, classifier languages do not distinguish between mass and count nouns at all. Some researchers even suggest that all nouns in classifier languages are mass nouns (Chierchia, 1998; Lucy, 1992; Quine, 1969; Borer, 2005; Kulkarni et al., 2013).

Given these cross-linguistic differences, a long-standing question has been whether speakers of non-classifier and classifier languages differ in the ontological distinction and construal of physical entities. Many studies have investigated the potential effects of cross-linguistic differences on conceptualizations of entity type (e.g., Whorf, 1956; Quine, 1969; Imai & Gentner, 1997; Imai & Mazuka, 2007; Barner et al., 2009; Lucy, 1992; Papafragou, 2005).

While speakers of classifier and non-classifier languages perform differently in object classification tasks, this does not mean that it is due to differences in perception. Li, Dunham, and Carey (2009) asked Japanese, Mandarin, and English speakers to rate a series of novel entities on a scale from 1 to 7 regarding whether they were objects or substances. They found that the ratings did not differ across the three speaker groups: all three languages classified the novel entities in the same way. It seems that once the mass-count distinction is taken out of the equation and the task is purely activating the conceptual object-substance condition, participants perform the same regardless of language. These results challenge Whorfian ideas and suggest a universalist approach that the object-substance distinction is universal in cognition.

To sum up, past research varies in the extent to which they attribute effects of language to entity construal and conceptualization. In the current work, we revisit this issue and also investigate how the conceptual system interacts with mass-count syntactic information cross-linguistically.

Cross-linguistic differences in the encoding of the mass-count distinction provide a unique opportunity to explore entity conceptualizations and related cognitive processes. While researchers have previously investigated questions of cross-linguistic differences and the conceptual underpinnings of entities, they have not fully addressed the current question of how speakers of classifier and non-classifier languages

integrate linguistic and non-linguistic information to construe and label non-canonically structured items.

Participants

We recruited 30 adult native Korean speakers from Prolific. Five participants were excluded from data analysis due to poor performance on the Korean knowledge questionnaire. The exclusion criteria on the Korean knowledge questionnaire were pre-specified prior to data analysis.

Stimuli

Entity Type We used the same set of visual stimuli from Experiment 1.

Linguistic Context While classifiers are not always mandatory in Korean, they are required in some linguistic contexts. One such context is when counting non-countable entities (substances) (Kim, 2021). For example, the Korean numeral *han* means ‘one.’ *Han* by itself is syntactically incomplete and must be combined with a classifier such as *kay* (the general classifier) to be grammatical. However, there is another form of the numeral ‘one,’ *hana*, which is a cardinal pronominal numeration. This means that it is syntactically complete and thus it can replace both the numeral and the classifier *han kay* ‘one CL.’ However, cardinal numerations can only substitute when the noun is a countable noun. In cases with non-countable mass nouns, *han* and a classifier are obligatory and *hana* is not permitted. This can be seen in the following examples:

- (1) a. sakwa han-kay
apple one-CL
b. sakwa hana
apple one-count
- (2) a. ssal han-doe
rice one-CL
b. *ssal hana
rice one-count

As we can see above in (1), *sakwa* ‘apple’ may combine with either *han kay* or *hana* because it refers to an individual entity that can be counted. In contrast, we see in (2) that *ssal* ‘rice’ is not permitted to combine with *hana* because it is a non-countable noun, but rather it may only appear with the classifier phrase *han doe* ‘one CL.’ This grammatical distinction drawn between countable and non-countable items seems to resemble the count-mass distinction we have seen in count-mass languages, and it maps pretty well to the conceptual object-substance distinction. We made use of this property in Korean to design our linguistic stimuli.

As in Experiment 1, in addition to manipulating the entity type that participants are shown, we manipulated the linguistic context that accompanied each item. We had two conditions: count context and mass context. We created 16 nonce Korean nouns, all of which were two or three syllables/characters. On each trial, we introduced each novel

alien word in a full sentence like (3) before showing them the first image. After displaying the second image, we used the novel alien word again in a sentence with the same syntax as in (4). In the count syntax condition, the noun was followed by *hana* ‘one-count’, and in the mass syntax condition, it was followed by *com* ‘some.’

- (3) po-a po-a, thalamwun-i
see-IMP see-IMP thalamwun-NOM
{hana_{count}/com_{mass}} iss-e.
one-count/some exist-DECL
‘Look! Here’s a thalamwun.’ (thalamwun=nonce word)
- (4) pangkum po-n sacin-ey-to
just.before see-ADJ image-DAT-also
thalamwun-i {hana_{count}/com_{mass}}
thalamwun-NOM one-count/some
iss-ess-ni?
exist-PST-Q
‘Was there also {a/some} thalamwun in the image that you just saw?’

Procedure

All recruitment information and instructions were presented in Korean. The experiment began with a language background questionnaire, followed by a Korean knowledge questionnaire that we designed. Afterwards, participants completed the main experiment. The trial structure of Experiment 2 was identical to that of Experiment 1, with the exception of the sentences and questions being presented in Korean.

Results

Results are plotted in Figure 2. As in Experiment 1, participants were more likely to label the structurally disrupted entity with the novel noun in the Substance condition than in the Object condition ($z=-4.87$, $p<0.001$). Again, as in Experiment 1, the likelihood of labelling the disrupted entity with the novel noun was modulated by linguistic context: participants were more likely to label the structurally disrupted entity with the novel noun when the noun was introduced with mass syntax than with count syntax ($z=-3.28$, $p=0.001$).

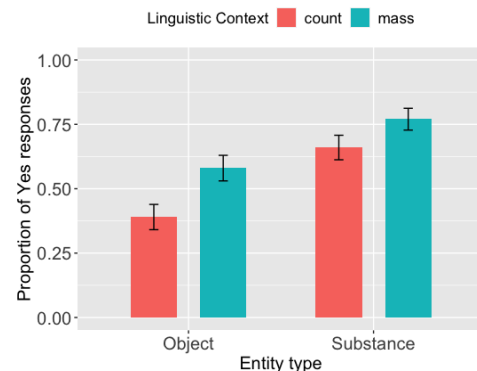


Figure 2: Proportion of Yes responses in Experiment 2

Discussion

In Experiment 2, we investigated whether speakers of a classifier language (Korean) also capitalize on entity type and linguistic context information to determine whether or not to extend a novel word to a restructured entity. The results suggest that speakers of classifier languages are also sensitive to entity type and mass-count context during novel word learning and extension. Like speakers of non-classifier languages (Experiment 1), they are more likely to extend the label to restructured entities when the label is applied to substances than when it was applied to objects, and also when the noun is introduced in mass context than in count context.

Although classifier languages do not encode entity type explicitly in their nouns, they still share the conceptual principle regarding structure, and use this information to determine the referential scope of nouns. We conclude that speakers of both English and Korean (and more broadly, speakers of both classifier and non-classifier languages) conceptualize individuals (objects) as having well-defined structure, while non-individuals (substances) do not. Moreover, we find that speakers of classifier languages also readily utilize mass-count expressions during word learning tasks. This is consistent with past research that has found that syntactic cues influence how objects are perceived and categorized (e.g., Barner et al. 2009). Finally, as in Experiment 1, we find that entity type has a stronger effect on the likelihood of label extension than linguistic context. The effect of linguistic context does not override the effect of entity. These results suggest that the conceptual distinction between entity types has a stronger effect than linguistic context on word extension to entities with non-canonical structure.

General Discussion

In this work, we aimed to investigate the role of objecthood and linguistic context in entity processing and labeling. We conducted novel word extension experiments to investigate how these factors modulate how speakers extend labels to previously unseen restructured entities. We tested this with speakers of non-classifier and classifier languages (Experiment 1: English and Experiment 2: Korean, respectively) to assess whether these are universal factors that affect both speakers of classifier and non-classifier languages in labeling novel entities. The interaction between entity perception and language is a valuable area of research to gain insight into how we conceptualize, categorize, and cognitively represent the physical world. Our results have important implications for cross-linguistic entity construal and label acquisition for spatial entities.

In both experiments, we found that participants were more likely to extend the label to the restructured substance with the novel noun than the restructured object. That is, regardless of whether one's native language grammatically encodes the distinction between objects and substances, objects and substances differ in terms of their conceptual representational structure: adults have a stronger structural commitment to individuals like objects than to non-

individuals like substances. Learners capitalize on such non-linguistic information when processing and labeling novel entities.

Furthermore, we found that the likelihood of label extension is also modulated by linguistic context: participants were more likely to extend the novel noun to the structurally disrupted entity when the noun was introduced with mass syntax than with count syntax, in both English and Korean. These results support the idea that count syntax supports individuated entity construal, and that the opposite holds in mass syntax. Crucially, Korean speakers also made use of such linguistic information, although the mass-count expressions were not grammatically encoded in the language, but rather expressed by different means.

In sum, learners capitalize on conceptual knowledge and integrate it with linguistic (count/mass) context to determine the referential scope of nouns. Crucially, these conceptual and linguistic factors hold up cross-linguistically even in languages that do not have grammatically encoded mass-count distinction. These findings contribute to the growing literature on the representational structures of objects and substances, as well as shed light on how learners integrate linguistic and non-linguistic information when learning novel nouns.

Language and Thought

What is the nature of the relationship between these conceptual and linguistic factors? Overall, we observe that the effect of linguistic context never overrides the effect of objecthood. Unlike several studies that found that when the syntactic context contradicts one's intuitions about an entity type, people are more likely to favor the syntactic context over their own perceptual concepts (e.g., Gordon, 1985, 1988), we did not find this to be the case.

We consider this data in conjunction with findings showing that infants acquire the object-substance concept before acquiring language, and now that speakers of classifier and non-classifier languages both adhere to the same principles when construing new entities. When we take this all into consideration, a possible explanation may be that the conceptual object-substance distinction precedes the count-mass distinction. We consider the possibility that the human mind first becomes sensitive to the distinction between objects and substances in accordance with universal conceptual criteria, then later integrates this knowledge with linguistic information.

Accordingly, our data challenge the Whorfian idea that speaking a language that has the mass-count distinction, or learning count syntax, shifts the perceptual boundary of what speakers consider to be discrete individuals. We propose that language is not what determines how entities are conceptualized, but rather it is a tool that may check pre-determined categories and disambiguate vague stimuli. This would be consistent with a growing body of literature in the field of language and cognition that believes language to be a tool that aids cognition when available, rather than determining thought.

References

- Barner, D., Inagaki, S., & Li, P. (2009). Language, thought, and real nouns. *Cognition*, 111, 329–344.
- Barner, D., Li, P., & Snedeker, J. (2010). Words as Windows to Thought: The Case of Object Representation. *Current Directions in Psychological Science*, 19(3), 195–200.
- Barner, D., & Snedeker, J. (2005). Quantity judgment and individuation: Evidence that mass nouns count. *Cognition*, 97, 41–66.
- Barner, D., & Snedeker, J. (2006). Children's early understanding of mass-count syntax: Individuations, lexical content, and the number asymmetry hypothesis. *Language Learning and Development*, 2, 163–194.
- Barner, D., Wagner, L., & Snedeker, J. (2008). Events and the ontology of individuals: Verbs as a source of individuating mass and count nouns. *Cognition*, 106, 805–832.
- Bloom, P. (1999). The role of semantics in solving the bootstrapping problem. In R. Jackendoff, P. Bloom, & K. Wynn (Eds.), *Language, logic, and concepts: Essays in memory of John Macnamara* (pp. 285–310). Cambridge, MA: MIT Press.
- Borer, H. (2005). *In Name Only: Structuring Sense: Volume 1*. Oxford: Oxford University Press.
- Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition*, 80, 179–213.
- Cheng, L., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30: 509–42.
- Chierchia, G. (1998a). Reference to kinds across languages. *Natural Language Semantics* 6: 339–405.
- Chierchia, G. (1998b). Plurality of mass nouns and the notion of 'semantic parameter'. In: Rothstein, K. (ed) *Events in Grammar*. New York: Springer, 53–103.
- Chierchia, G. (2010). Mass nouns, vagueness and semantic variation. *Synthese* 174: 99–149.
- Doetjes, J. (1997). *Quantifiers and Selection: On the Distribution of Quantifying Expressions in French, Dutch and English*. Hague, HAG: Leiden University Dissertation.
- Doetjes, J. S. (2012). Count/mass distinctions across languages. In: C. Maienborn, K. von Stechow, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning*, part III. Berlin: De Gruyter, 2559–2580.
- Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science*, 6, 568–584.
- Gordon, P. (1985). Evaluating the semantic categories hypothesis: The case of the count/mass distinction. *Cognition* 20: 209–42.
- Gordon, P. (1988). Count/mass category acquisition: Distributional distinctions in children's speech. *Journal of Child Language* 15:109–128.
- Hespos, S. J., Ferry, A. L., & Rips, L. J. (2009). Five-month-old infants have different expectations for solids and liquids. *Psychological Science*, 20(5), 603–611.
- Imai, M., & Gentner, D. (1997). A crosslinguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, 62, 169–200.
- Kim, K. S. (2021). A continuum-based approach to the count-mass distinction in Korean. *Numeral Classifiers and Classifier Languages*, 98–120.
- Kim, Y. W. (2021). Descriptive function of numeral classifiers. *Numeral Classifiers and Classifier Languages*, 231–257.
- Kodera, M. (2011). A cross-linguistic study of mass-count distinctions. *Hannan Journal of Humanities and Natural Science*, 46(2), 41–50.
- Kulkarni, R., Rothstein, S., & Treves, A. (2013). A statistical investigation into the cross-linguistic distribution of mass and count nouns: morphosyntactic and semantic perspectives. *Biolinguistics*, 7, 132–168.
- Lee, S. H., Ji, Y., & Papafragou, A. (Submitted). Signatures of individuation across objects and events.
- Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on construal. *Cognitive Psychology*, 58, 487–524.
- Lucy, J. (1992). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.
- Papafragou, A. (2005). Relations between language and thought: Individuation and the count/mass distinction. *Handbook of Categorization in Cognitive Science*, 255–275.
- Prasada, S., Ferenz, K., & Haskell, T. (2002). Conceiving of entities as objects and as stuff. *Cognition*, 83(2), 141–165.
- Quine, W.V.O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- Quine, W. V. O. (1969). *Ontological Relativity and Other Essays*. New York, NY: Columbia University Press.
- Soja, N. N., Carey, S., & Spelke, E. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, 38, 179–211.
- Strickland, B. (2016). Language reflects "core" cognition: A new theory about the origin of cross-linguistic regularities. *Cognitive Science*, 41(1), 70–101.
- vanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science*, 14(5), 498–504.
- Whorf, B. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.
- Yi, B. U. (2010). Numeral classifiers and the mass/count distinction. University of Toronto.
- Yi, B. U. (2021). Numeral classifiers and diversity of classifier systems. *Numeral Classifiers and Classifier Languages*, 6–39.
- Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX).