

Shape Bias in Korean Noun Learning: Evidence from Property Ratings and Corpus Analysis in a Classifier Language

Jongmin Jung¹ & Eon-Suk Ko^{1,2}

¹Center for Data Science in Humanities, Chosun University

²Department of English Language and Literature, Chosun University

jongminjung169@gmail.com; consukko@chosun.ac.kr

Children acquire words with remarkable efficiency, and researchers have proposed various perceptual and cognitive mechanisms underlying this process (e.g., Markman, 1990; Smith et al., 2002). In number-marking languages such as English, the *shape bias*—a robust tendency to generalize novel labels based on shape similarity—has been proposed to be related to accelerated vocabulary growth (Gershkoff-Stowe & Smith, 2004; Samuelson & Smith, 1999), supported by obligatory count–mass morphosyntax (Imai & Gentner, 1997; Soja et al., 1991). However, in classifier languages including Korean, number marking is not obligatory; instead, classifiers (e.g., *gae*) encode individuation in numeral constructions. This raises a central theoretical question: does the shape bias emerge from cognitive and perceptual mechanisms independent of obligatory count–mass morphology, or does its development require language-specific grammatical support? The present study is the first to examine cognitive, perceptual, and linguistic properties of children's early expressive vocabulary in Korean. We test the theoretical prediction that shape-centered multi-feature words structure early lexical development even in the absence of obligatory count–mass morphology, thereby providing evidence for a universal cognitive foundation for the *shape bias* across typologically distinct languages.

Twenty-nine native Korean-speaking adults (22 female; aged 18–43+ years) rated 260 common nouns from the Korean MacArthur–Bates Communicative Development Inventory (K-MBCDI:WS; Pae & Lee, 2011) on three dimensions: (1) countability, (2) perceptual bias, and (3) solidity. Items were classified into a category when adult agreement exceeded 70%. Shape-centered multi-feature words were defined following Perry and Samuelson (2011) as items meeting one or more of: [count + shape + solid], [shape + count], [shape + solid], or [count + solid]. These adult ratings were applied to vocabulary data from 1,589 Korean children aged 18–36 months (788 girls; $M = 26.79$ months, $SD = 5.42$) drawn from a university child language laboratory corpus and the Wordbank database (<https://wordbank.stanford.edu>). Age of acquisition (AoA) was defined as the earliest age at which $\geq 50\%$ of children produced a given word. Vocabulary scores were analyzed using beta regression to examine the relationship between the proportion of shape-centered nouns and children's K-MBCDI percentile scores.

The results revealed that 86.92% of K-MBCDI nouns were classified as countable and 90.00% as shape-biased by native Korean-speaking adults. In contrast, only 57.31% of items were classified as solid. Accordingly, 225 nouns (86.54%) were classified as shape-centered multi-feature words. Spearman correlations confirmed positive but partially dissociable associations among countability, shape bias, and solidity ($\rho_s = .15-.58$). Logistic growth modeling revealed that Korean children acquire shape-centered words more rapidly and earlier than non-shape-centered words (Figure 1). Beta regression analyses further showed that the proportion of shape-centered nouns in children's vocabulary was significantly associated with K-MBCDI percentile scores across ages 18–30 months ($\beta = 1.63$, $p < .001$). This advantage reversed in the 31–36 month period ($\beta = -21.38$, $p < .001$), consistent with a

developmental shift toward more diverse lexical acquisition strategies as children's vocabularies expand beyond shape-centered nouns (Figure 2).

The prominence of countable, shape-biased nouns in Korean children's early vocabulary emerged without the morphosyntactic scaffolding available in number-marking languages, suggesting that cognitive–perceptual mechanisms—which precede and are only partially reshaped by language-specific grammar—are sufficient to support shape-based word generalization. Critically, convergence of multiple features (countability, shape bias, solidity) into shape-centered multi-feature words—rather than any single property alone—predicted early vocabulary advantage, pointing to a multi-dimensional attractor structure as the mechanism underlying the shape bias across languages. Taken together, these results support a universal cognitive foundation for early noun learning, with linguistic systems playing a modulatory rather than determinative role.

References

- Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, 75(4), 1098–1114.
- Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, 62(2), 169–200. doi: 10.1016/s0010-0277(96)00784-6
- Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, 14(1), 57–77.
- Pae, S., & Lee, Y. (2011). *Korean MacArthur-Bates Communicative Development Inventories: User's guide and technical manual*. Seoul: Mindpress.
- Perry, L. K., Kucker, S. C., Horst, J. S., & Samuelson, L. K. (2022). Late bloomer or language disorder? Differences in toddler vocabulary composition associated with long-term language outcomes. *Developmental Science*, e13342.
- Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, 73(1), 1–33.
- Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, 13(1), 13–19.
- Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning. *Cognition*, 38(2), 179–211.

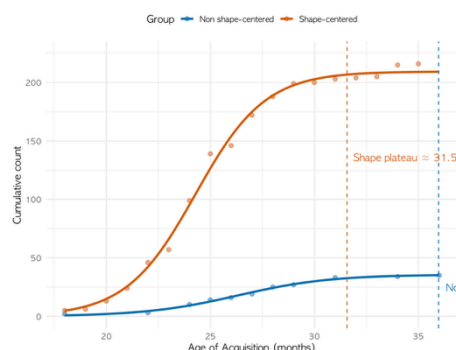


Figure 1. Predicted Developmental Growth of Shape-centered Nouns and Non Shape-centered Nouns

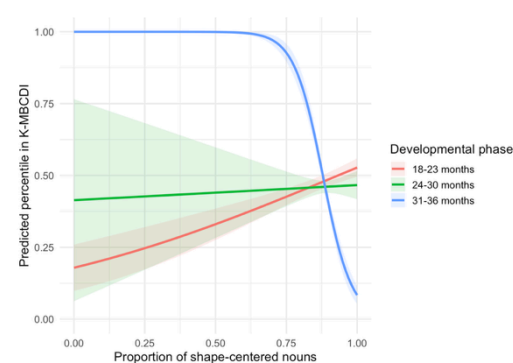


Figure 2. Developmental Patterns in the Associations Between the Proportion of Shape-Centered Nouns and Children's Vocabulary Scores

