The role of form-to-meaning correspondences in novel word learning and semantic retrieval

Anonymous Author[1]

[1] Anonymous Affiliation

Placeholder for email address

Traditional accounts of language have emphasized that the relationship between wordform and meaning is arbitrary [1]. However, a growing body of work suggests that wordform similarity is positively correlated with semantic similarity [2] and these form-meaning correspondences can benefit word learning and categorization [3]. Even so, the interactions between wordform and meaning in word learning and downstream semantic retrieval tasks remain understudied. This work examined how form-meaning correspondences influence novel word acquisition and how these lexical sources are utilized in the semantic fluency task using a combination of experimental and computational modeling work.

In a language learning experiment, participants learned novel words embedded in sentences (e.g., "I love a X rock"), where the novel word (i.e., X) was either phonologically related (e.g., *zock*) or unrelated (e.g., *dodsy*) to the target noun. In a second experiment, participants learned direct and indirect associations between novel words and target nouns (e.g., "I love a *foobly* Y" and "I love a *foobly mipp*"), where whether the target noun (i.e., Y) shared the starting sound with the indirectly associated novel word (e.g., *melon-mipp* vs. *apple-mipp*) was manipulated. Across both experiments, novel words with form-based overlap facilitated processing of target nouns in a semantic priming task (see top panel of Figure 1). These findings converge with emerging work on how words are learned faster when they have greater phonological overlap with known words (e.g., *dester* is a phonological neighbor of *fester*, *duster*, etc.; see [4]) and individuals automatically use wordform information to make sense of pseudowords (e.g., *tolque* sounds like *torque*; see [5]), and provide novel evidence that wordform and statistical co-occurrence jointly influence the semantic integration of novel words into the lexicon.

To next examine how form-based information influences downstream semantic retrieval, search behavior within the semantic fluency task (e.g., "name all the animals/foods you can in one minute") was analyzed. Blended lexical representations were created by integrating word embeddings learned by two language models trained on speech and text (*speech2vec* and *word2vec*; [6]). These representations were then incorporated into a process model of lexical search and evaluated on fluency responses from children with normal hearing and cochlear implants. Model comparisons indicated that in categories where semantic and phonological information was more strongly correlated (i.e., foods), leveraging both sources to retrieve items in local clusters (e.g., *cauliflower-carrot*, *brownies-cookies*, etc.) was associated with better performance. However, individuals with cochlear implants were unable to adaptively utilize these lexical sources, even though acoustic information was globally amplified in their mental lexicon (see bottom panel of Figure 1). Taken together, these findings shed light on how form-to-meaning mappings and early linguistic experiences not only shape novel word learning and lexical organization, but also how these mappings are actively leveraged in semantic tasks and contribute to efficient lexical processing.

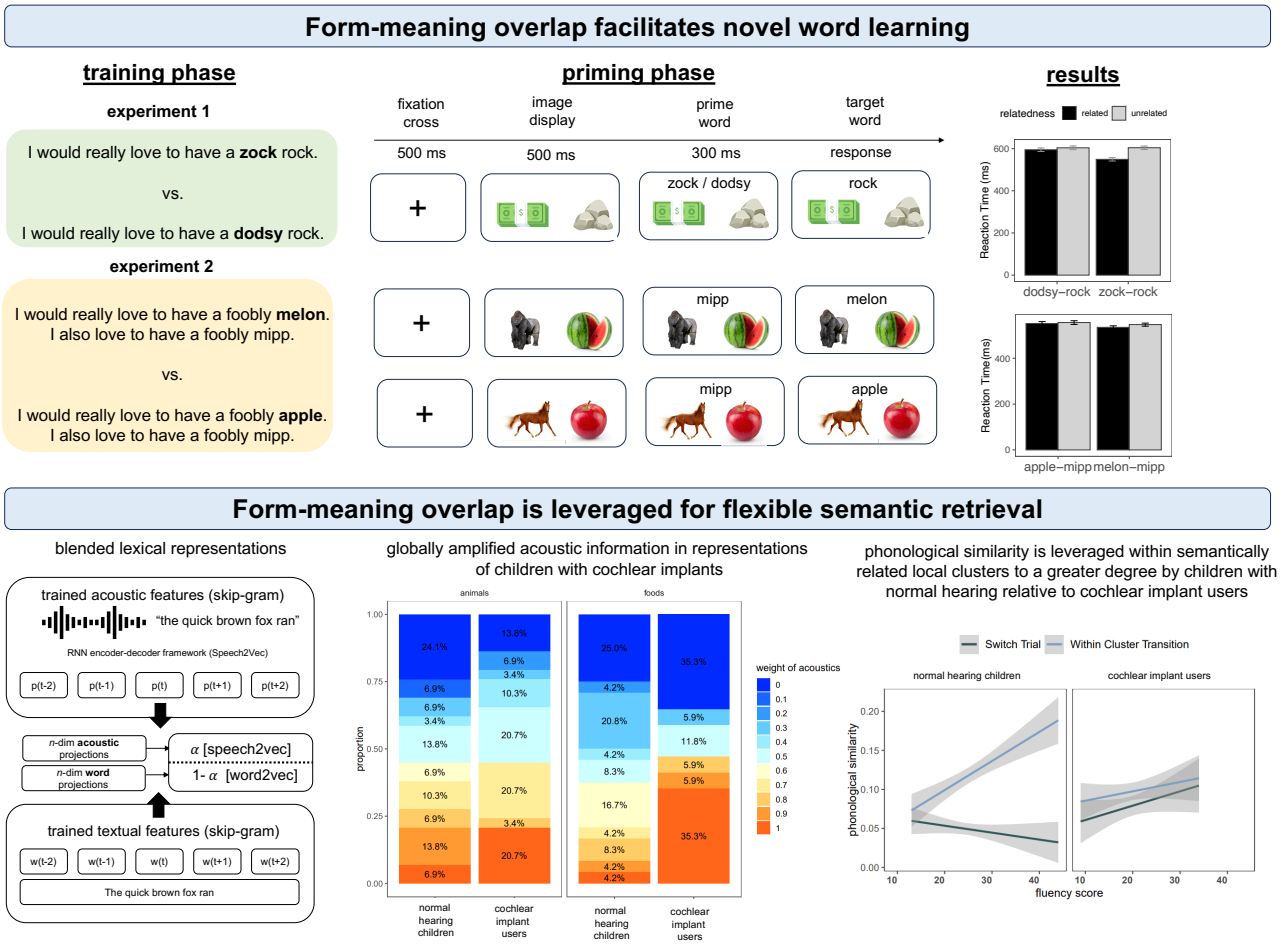

Figure 1: **Top Panel:** Experimental approach and results from two experiments manipulating wordform overlap. Form and statistical co-occurrence jointly influenced semantic integration of novel words. **Bottom panel**: Modeling approach and patterns within the semantic fluency task. Individuals with cochlear implants had higher global activation of acoustic information in their lexicon, but were unable to locally leverage phonological information to retrieve more items.

References

[1] Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97.

[2] Dautriche, I., Mahowald, K., Gibson, E., & Piantadosi, S. T. (2017). Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cognitive Science*, *41*(8), 2149–2169.

[3] Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*(10), 603–615.

[4] James, E., Gaskell, M. G., & Henderson, L. M. (2019). Offline consolidation supersedes prior knowledge benefits in children's (but not adults') word learning. *Developmental Science*, *22*(3), e12776.

[5] Gatti, D., Rodio, F., Rinaldi, L., & Marelli, M. (2024). On humans'(explicit) intuitions about the meaning of novel words. *Cognition*, *251*, 105882.

[6] Chung, Y.-A., & Glass, J. (2018). Speech2vec: A sequence-to-sequence framework for learning word embeddings from speech. *arXiv preprint arXiv:1803.08976*.