How many words is a picture (or a definition) worth? A distributional perspective on learning new word meanings

Matthew Z. Borman[*1], Gary Lupyan[1]

[*]Corresponding Author: borman@wisc.edu
[1]Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA

As adults, we continue to learn new word meanings. We can learn new words through ostensive labeling events where a word denotes a clear referent in context, or by having the word explicitly defined for us (Hahn & Gershkoff-Stowe, 2010). However, people also learn word meanings through exposure to how words are used in text (Nagy et al., 1985; Saragi et al., 1978). Here, we examine the relative effectiveness of different ways of learning new word meanings, finding that more ostensive experiences are not necessarily more effective than indirect learning via merely observing how a word is used.

Both research and intuition suggest that explicit/direct experiences with new words (often times via definitions or ostensive referents) are efficient and effective ways of learning new word meanings (Gruhn, et al., 2020; Watts, 1995). In comparison, the knowledge we gain from experience with words in natural text may seem somewhat fuzzy, imprecise, and variable from one instance to another. This variability, however, provides rich distributional information, helping link the new word to already known words.

One crucial aspect of word knowledge requires learners to generalize to new situations or different modalities. Though efficient, do these more explicit, direct experiences also yield generalizable word knowledge? Conversely, have we underestimated the richness that naturalistic text imparts during learning? In Experiment 1, we ask whether richer but less precise contexts (sentences), or more explicit/direct contexts (images and definitions) best yield generalization to other modalities or types of text. Experiment 2 builds on this finding by demonstrating that surprisingly little exposure is required for the distributional patterns of naturalistic text to efficiently impart word meaning.

Experiment 1

To test the how well different word learning experiences impart generalizable word knowledge, participants (N=58) were exposed to 12 novel word meanings (e.g. "the empty space at the top of a container") and pseudowords (Keuleers & Brysbaert, 2010) in one of three conditions where they either: read a definition, viewed four images depicting the new word's meaning, or read five sentences generated using ChatGPT (OpenAI, 2023) that used the word in context without defining it. To test how well participants learned the word meanings, we showed them new unlabeled images, definitions, and (cloze) sentences for each trained meaning and asked them to match it to one of the words presented. Because we were interested in generalization, our analysis only included responses for a given word if the participant answered correctly when tested in its exposure condition. A mixed effects logistic regression model was used to analyze the relationship between exposure condition and generalization X2 (1, N=58) = 4.79, p=.028 and participants who learned from sentences (M=.33, SE=.05) were more accurate in generalizing to other test conditions compared to participants who learned via images (M=.19, SE=.04) or definitions (M=.21, SE=.04). In sum, learning from passive exposure to text better supported generalization to situations that involved other types of word knowledge and visual knowledge.

Experiment 2

What do these results, then, say about human cognition? If we have underestimated the richness that linguistic experience affords during word learning, we may have also underestimated one of the processes believed to underlie word learning – distributional learning. To assess the role of distributional learning with minimal exposure, participants (*N*=86) learned three rare words (Brysbaert, et al., 2019) by reading ten sentence contexts sampled from COCA (Davies, 2008-). After exposure, participants provided definitions for the newly learned target words. A separate set of participants (N=30) defined and reported their familiarity with these words (without receiving any exposure). Sentence embeddings (Reimers & Gurevych, 2019) were then computed for the definitions collected from the experimental, high, and low familiarity groups, and evaluated for similarity to dictionary definitions. Bootstrapped means of these embedding similarities showed that participants with just ten exposures, *M*=.29, 95% CI [.28, .29], moved away from definitions of people who reported not knowing the word, *M*=.20, 95% CI [.19, .21], and towards definitions of those who reported knowing the word *M*=.37, 95% CI [.36, .38].

Our findings show how learners leverage the richness of natural language to gain generalizable, expert-like word meaning knowledge from surprisingly little exposure. Ongoing work is exploring how even more minimal text exposure and controlling for RTs may provide a window into relative efficiencies.

References

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*(2), 467–479. https://doi.org/10.3758/s13428-018-1077-9

Davies, Mark. (2008-) The Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/.

Gruhn, S., Segers, E., & Verhoeven, L. (2020). Moderating role of reading comprehension in children's word learning with context versus pictures. *Journal of Computer Assisted Learning*, *36*(1), 29–45. https://doi.org/10.1111/jcal.12387

Hahn, E. R., & Gershkoff-Stowe, L. (2010). Children and Adults Learn Actions for Objects More Readily Than Labels. *Language Learning and Development*, *6*(4), 283–308. https://doi.org/10.1080/15475441003635315

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633. https://doi.org/10.3758/BRM.42.3.627

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. *Reading Research Quarterly*, *20*(2), 233–253. https://www.jstor.org/stable/747758.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 3982–3992. https://doi.org/10.18653/v1/d19-1410

Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, *6*(2), 72–78. https://doi.org/10.1016/0346-251X(78)90027-1

Watts, S. M. (1995). Vocabulary instruction during reading lessons in six classrooms. *Journal of Literacy Research*, *27*(3), 399–424. https://doi.org/10.1080/10862969509547889