Filling in the gaps: Holes in the semantic network and why they matter

Authors: Thomas T. Hills & Cynthia S. Q. Siew

Cite as: Hills, T. T., & Siew, C. S. (2018). Filling gaps in early word learning. *Nature Human Behaviour*, *2*(9), 622.

[Intro] Why do children learn some words before others? Years of research on how children acquire language has shown that children are not merely randomly learning words--for instance, children tend to learn words like "mama" and "ball" before words like "moose" and "bill". While much research has focussed on the developmental trajectory of words that children learn, a recent paper in Nature Human Behavior investigated the words that children *don't* learn, and how these yet-to-be learned words result in knowledge gaps in children's semantic and lexical spaces that eventually need to be filled in as they learn new words.

Much of human endeavor is devoted to filling some kind of "gap": Businesses provide products and services that are missing in the world and scientists do research that fills a gap in what we currently know. Within the domain of learning, people learn new information to fill gaps in their current understanding. Even young children are not exempt from this enterprise—as demonstrated by a recent paper in Nature Human Behavior investigating how children fill knowledge gaps in their semantic lexicon as they learn new words.

Knowledge gaps in children refer to cavities or holes in the semantic structure of words that children know. These gaps represent sparse areas of the semantic space that need to be filled out via the process of language acquisition.

To study the development and closure of these gaps in children's early language acquisition, researchers from the University of Pennsylvania used an innovative combination of methods from graph theory and topological analysis. The words that children know can be depicted as a network where nodes represent words and connections (or edges) are placed between words that have similar features. For instance, "banana" and "cheese" are two connected nodes in the semantic network because they are both yellow in color.

Gaps are formed when pairs of nodes are connected via multiple paths, leaving a space in the network where a new node could fill the gap (see Fig. 1). By examining how the network developed as children aged, the researchers at the University of Pennsylvania were able to detect when knowledge gaps first emerge in the network and when these gaps are closed.

By simulating different network growth trajectories, the authors were able to investigate possible underlying processes guiding early word learning. These simulations manipulated the probability that a newly added node formed edges. A classic model, preferential attachment, which prioritises new nodes (words) forming edges with nodes that are already well-connected, was a poor fit the observed data. Models where new nodes randomly formed edges were also a poor fit to the observed pattern.

The best fitting models involved assigning each word its own constant affinity for creating edges during development. Of these models, those that prioritized learning more distinctive words first or learned words with many edges to other words in the learning environment (a process called preferential acquisition) showed patterns similar to the observed data.

These results reveal that language acquisition is remarkably robust to the variability of language input as well as what the child already knows. In other words, what matters is the structure of the learning environment, not the way the information is learned. This suggests that growth models typical of other real-world complex systems, such as the Internet and social systems, do not lead to the same patterns of network development that we see in early word learning. Rather, alternative models that emphasize the semantic structure of the learning environment (preferential acquisition; Hills et al., 2009) and prioritize the acquisition of distinctive words (Engelthaler & Hills, 2017) appear to better describe children's learning, which is further corroborated by the formation and filling of knowledge gaps.

Why is the formation and filling in of knowledge gaps seemingly unaffected by the order in which words are learned? Why is preferential attachment not a good fit to the growth observed in semantic networks? We suggest that language is learned in a way that allows a consistent global structure of the lexicon to be acquired despite the inevitable variability in the linguistic input (Hart & Risley, 1995). This may represent a universal property of learning in complex environment, for example, by first learning general (course-grained) features of the learning environment, which create gaps that lead to learning of more detailed (fine-grained and gap filling) features. Thus, Sizemore's new research may represent a new property of learning that has yet to be rigorously quantified.

As a final point, it is important to realize that a gap in feature space may not be a gap in another space. The network edges used in the above research were constructed of shared features (e.g., "made of wood"). But edges can be based on a variety of relationships such as co-occurrence in child-directed language (Hills et al., 2010), phonology (Siew, 2013), semantic relationships available from free association data ("say the first word that comes to mind when I say 'cat'", Hills et al., 2009), or combinations of the above (Stella, Beckage, & Brede, 2017). Each of these approaches has established that different kinds of 'connective tissue' in language matter in relation to what words children learn and are predictive of early word learning, though in different ways. Alongside the new research reported here, these different approaches identify gaps in what we know about how children learn new words, and therefore places where future research can investigate new models on a variety of network structures (see Table 1).

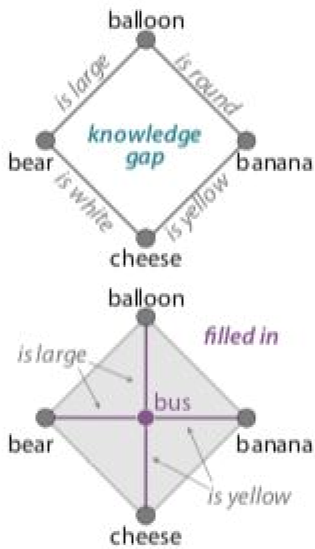

Fig 1. Example of a knowledge gap in a child's semantic network which is filled when the word "bus" is acquired. [This is from Fig. 2 the original Sizemore manuscript, but a reproduction with visual images for the nodes would be very appealing.]

Table 1. Networks, attributes, and models.

		Edges			
		Phonology	Semantics	Features	Multiplex
Growth models	Distinctiveness	?	?	+ Engelthaler & Hills (2017) + Sizemore et al. (2018)	?
	Preferential acquisition	?	+ Hills et al. (2009)	+ Sizemore et al. (2018) - Hills et al. (2009)	+ Stella et al. (2017)
	Lure of associates	+ Storkel (2002)	+ Hills et al. (2010)	- Hills et al. (2009)	?
	Preferential attachment	?	- Hills et al. (2009)	- Hills et al. (2009)	?

Notes: + indicates that the growth model predicts learning on a network with edges of a specific type. - indicates that the model is not predictive. ? indicates unexplored areas for future research.

References

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509-512.

Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports*, *7*, 46730.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition?. *Psychological Science*, *20*(6), 729-739.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259-273.

Siew, C. S. Q. (2013). Community structure in the phonological network. *Frontiers in Psychology*, *4*, 553.

Engelthaler, T., & Hills, T. T. (2017). Feature biases in early word learning: network distinctiveness predicts age of acquisition. *Cognitive Science*, *41*, 120-140.

Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, *29*(2), 251-274.

Sizemore paper (current paper).