COGNITIVE SCIENCE

Filling gaps in early word learning

Years of research has shown that children do not learn words at random, but in distinct patterns. Why do we observe the patterns that we do? By using network science and investigating the words that children don't learn, researchers have potentially uncovered a general property of word learning as a process of gap forming and filling.

Thomas T. Hills and Cynthia S. Q. Siew

Much human endeavour is devoted to filling some kind of 'gap': businesses provide products and services that are missing in the world and scientists carry out research that fills gaps 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 Sizemore et al. in a recent paper in *Nature Human Behaviour* investigating how children fill knowledge gaps in their semantic lexicon as they learn new words[1].

Knowledge gaps in word learning refer to cavities or holes in the semantic structure of words that children know. These gaps represent sparse areas of the semantic space that can be filled in as new words are learned.

To study the development and closure of these gaps in children's early language acquisition, Sizemore and colleagues[1] used an innovative combination of methods from graph theory and topological analysis to depict words that children know as a network: nodes represent words with connections (or edges) between words that have similar features. For instance, 'banana' and 'cheese' are two connected nodes in the semantic network because they are both yellow.

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 were able to detect when knowledge gaps first emerged in the network and when these gaps later 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. 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.

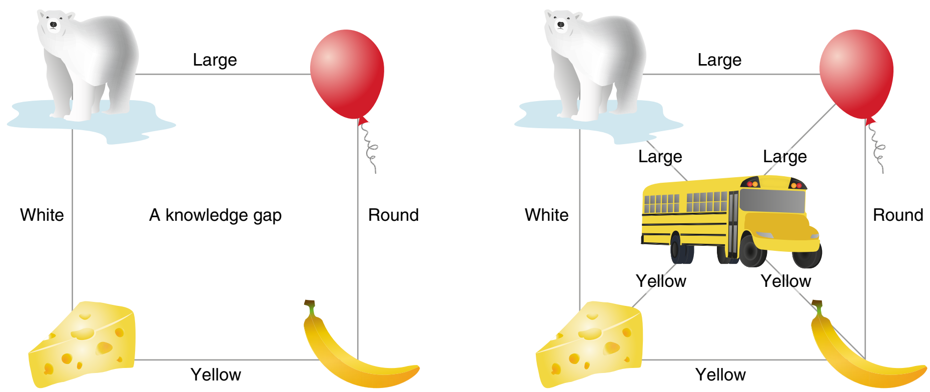

Fig. 1 | An example of a knowledge gap in a child's semantic network. The gap is filled when the word 'bus' is acquired.

These results reveal that language acquisition is remarkably robust to the variability of language input as well as to what the child already knows. Put another way, it shows that it is the structure of the learning environment, not previous knowledge, that is important. 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[2]) and prioritize the acquisition of distinctive words[3] appear to better describe children's learning. These models lead to the pattern of gap formation and filling observed in children.

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? The evidence suggests 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[4].

This may represent a universal property of learning in complex environments. That is, initially learning general, coarse-grained 'landmarks' in the learning environment creates gaps that are later filled by more detailed, fine-grained information. For example, in many domains including language, we learn typical and mid-level members of categories first and only later learn atypical and subordinate category members, such as birds that do not fly or vehicles with one wheel[5]. Any system of learning that starts with learning mid-level categories may therefore show gap-forming and filling behaviour. The research by Sizemore et al.[1] identifies this potentially universal process of gap forming and gap filling, where information is learned in initially sparse representations that are later filled in with more detail. This may allow learners to efficiently acquire the complex structure of real-world environments in a form that is optimally efficient however much information the learner currently has[6].

As a final point, it is important to realize that a gap in any given feature space may not be a gap in a different feature space. The network edges used by Sizemore et al.[1] were constructed of shared features (for example,

Table 1 | Combination of growth models and network attributes reveals gaps in our understanding of children's word learning

Growth models	Edges			
	Phonology	Semantics	Features	Multiplex
Distinctiveness	?	?	+ (ref. 3) + (ref. 1)	?
Preferential acquisition	?	+ (ref. 2)	+ (ref. 1) – (ref. 2)	+ (ref. 10)
Lure of associates	+ (ref. 9)	+ (ref. 7)	– (ref. 2)	?
Preferential attachment	?	– (ref. 2)	– (ref. 2)	?

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.

'made of wood' and 'is white'). But edges can be based on a variety of relationships such as co-occurrences in child-directed language[7], phonology[8,9], semantic relationships available from free association data[2] ("say the first word that comes to mind when I say 'cat'"), or combinations of the above[10]. Each of these approaches has established that different kinds of 'connective tissue' in language matter in relation to what words children learn — they are each predictive of early word learning, although in different ways. But do these different ways of connecting words to one another reveal the same kinds of learning properties? For example, will gaps form and fill in the phonological space occupied by words and reveal the same patterns of preferential acquisition? This brings us back to gap filling as a human endeavour, one that drives scientific inquiry in various academic fields. Alongside the new research reported here, the combination of models and network connectivity reveals gaps in what we know about how children learn new words, and therefore areas where future research can investigate new models on a variety of network structures (see Table 1). ❐

Thomas T. Hills* and Cynthia S. Q. Siew
Department of Psychology, University of Warwick, Coventry, UK.
**e-mail: t.t.hills@warwick.ac.uk*

Published online: 7 September 2018
https://doi.org/10.1038/s41562-018-0428-y

References

1. Sizemore, A. E., Karuza, E. A., Giusti, C. & Bassett, D. A. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-018-0422-4 (2018).
2. Hills, T. T., Maouene, M., Maouene, J., Sheya, A. & Smith, L. *Psychol. Sci.* **20**, 729–739 (2009).
3. Engelthaler, T. & Hills, T. T. *Cogn. Sci.* **41**, 120–140 (2017).
4. Hart, B. & Risley, T. R. *Meaningful Differences in the Everyday Experience of Young American Children* (Paul H. Brookes, Baltimore, MD, 1995).
5. Rosch, E. & Lloyd, B. B. *Cognition and Categorization* (Erlbaum Associates, Hillsdale, NJ, 1978).
6. Kemp, C. & Tenenbaum, J. B. *Proc. Natl Acad. Sci. USA* **105**, 10687–10692 (2008).
7. Hills, T. T., Maouene, J., Riordan, B. & Smith, L. B. *J. Mem. Lang.* **63**, 259–273 (2010).
8. Siew, C. S. Q. *Front. Psychol.* **4**, 553 (2013).
9. Storkel, H. L. *J. Child Lang.* **29**, 251–274 (2002).
10. Stella, M., Beckage, N. M. & Brede, M. *Sci. Rep.* **7**, 46730 (2017).

Competing interests
The authors declare no competing interests.