Child language input does not reflect world frequency: Typical and atypical feature description across development

Anonymous CogSci submission

Abstract

Language provides children a powerful source of information about the world. From language alone, simple distributional learning models can recover enough information to perform comparably to non-native college applicants on the TOEFL (Landauer & Dumais, 1997). Blind children learn the same kinds of relationships among perceptual categories as sighted children, without any of the relevant visual input (Landau & Gleitman, 1985). However, language does not perfectly reflect the world: the most typical features of natural kinds may often go unremarked. For instance, adults rarely describe the color of an orange carrot, as world knowledge makes this description redundant. Given children's nascent world knowledge, does parents' speech to children follow this pattern? From longitudinal corpus data of parent-child communication (Goldin-Meadow et al., 2014) between 14–58 months, we extracted usage data for 684 high-frequency concrete nouns and co-occurring adjectives. Independent raters coded the typicality of over 2,000 unique adjective–noun pairs on a 7-point Likert scale (interrater reliability: r = 0.8 in a subset of the data). If language statistics reflect world statistics, description should be dominated by the typical (strong negative skew); however, across all ages, we see descriptors concentrated in the atypical range (positive skewness = 0.38). Parents were reliably more likely to use typical descriptors when talking to younger rather than older children. Overall, child language input reflects notable more than typical features, but increased description of typical features early in development may provide a foothold for young learners.

Keywords: referential pacts; parent-child communication

Children learn a tremendous amount about the structure of the world around in just a few short years. From the rules that govern the movement of physical objects to the hierarchical structure of natural categories and even relational structures among social and cultural groups (Baillargeon, 1994; Legare & Harris, 2016; Rogers & McClelland, 2004). Where does the information for this rapid acquisition come from? Undoubtedly, a sizeable comoponent in at least some domains comes from direct experience observing and interacting with the world (Sloutsky & Fisher, 2004; Stahl & Feigenson, 2015). But, another important source of information arises from cultural learning: the information in the language of people around them (Landauer & Dumais, 1997; Rhodes, Leslie, & Tworek, 2012).

How similar is the information available from children's direct experience to the information available in the language children hear? Several lines of work suggest that they may be surprisingly similar. One compelling area of work has been the comparison of semantic structures learned by congentinally blind children to those of their sighted peers. In serveral domains that would at first blush rely heavily on perceptual information–like color or in verbs of perception (e.g. *look*, *see*), blind children's semantic similarity judgments are quite similar to those of sighted children (Landau, Gleitman, & Landau, 2009). Further, blind adults' judgments of perceptual verbs are sensitive to highly detailed information like variation in (e.g. blaze vs. glow), just like sighted adults (Bedny, Koster-Hale, Elli, Yazzolino, & Saxe, 2019). Another piece of evidence in favor of the redundance of vision and language is the broad success of statistical models trained on the statistics of language in approximating human judgments across a variety of domains (Landauer & Dumais, 1997; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013, @devlin2018).

Experiment 1: Judgements

Method

Participants

Stimuli

Procedure

Design

Results

Acknowledgements

Place acknowledgments (including funding information) in a section at the end of the paper.

References

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science*, *3*(5), 133–140.

Bedny, M., Koster-Hale, J., Elli, G., Yazzolino, L., & Saxe, R. (2019). There's more to "sparkle" than meets the eye: Knowledge of vision and light verbs among congenitally blind and sighted individuals. *Cognition*, *189*, 105–115.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv Preprint arXiv:1810.04805*.

Landau, B., Gleitman, L. R., & Landau, B. (2009). *Language and experience: Evidence from the blind child* (Vol. 8). Harvard University Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211.

Legare, C. H., & Harris, P. L. (2016). The ontogeny of cultural learning. *Child Development*, *87*(3), 633–642.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, *109*(34), 13526–13531.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, *133*(2), 166.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94.