# Modeling Lexical-Semantic Networks of School Age Children through a Repeated Word Association Task

Throughout the school years, children's vocabularies undergo massive expansion as they encounter new words in varied academic and non-academic contexts (Verhoeven, van Leeuwee, & Vermeer, 2011; Song et al., 2015). Recent work in network science has explored how toddlers organize their early vocabularies (Beckage, Smith, & Hills, 2011; Peters & Borovsky, 2019; Wojcik, 2018). Our work extends network modeling to vocabulary development in school-age children, focusing on word association as a critical aspect of lexical-semantic organization (De Deyne, Navarro, & Storms, 2013). Study 1 involved reanalysis of a corpus of word associations of children with typical language development (*N* = 22; 12 girls, 10 boys; *M* age = 6y;7m, range 5;3–8;7), taken from a larger dataset that included children with developmental language disorder. In the repeated word association task, participants were asked to generate the first word that came to mind in response to a series of cue words (e.g., *bridge, read, snake, pillow*), with the 48 cues repeated three times each. Cue were pre-recorded and presented through external speakers. Study 2 used the same task to generate a new corpus of word associations from children with typical development (*N* = 21; 14 girls, 7 boys; *M* age = 9y; 6m, range 7;1–11;4) and undergraduate students (*N* = 21; 13 women, 8 men; *M* age = 20y 3m, range 18;0–26;0).

*Semantic Relatedness Estimation.* The semantic relatedness of each cue-response pair was estimated using two large-scale corpora of the English language: Latent Semantic Analysis (LSA: Landauer & Dumais, 1997, http://lsa.colorado.edu/) and Continuous Bag of Words (CBOW: Mandera, Keuleers, & Brysbaert., 2017, http://meshugga.ugent.be/snaut-english/). In Study 1, we used a median split to divide children into younger and older age groups (younger: *M* age = 5y; 9m; older: *M* age = 7y; 5m). We ran linear mixed-effects models, with random effects of participants and cues, to examine effects of age group, list repetition, and their interaction on the LSA and CBOW estimates of semantic relatedness. Both analyses (LSA, CBOW) revealed a significant main effect of list repetition and a significant interaction of age group × repetition, but no main effect of age group. Children's first responses were more closely related in meaning to the cues than later responses. In Repetition 1, responses of older children were more closely related to the cues than those of younger children; the group difference was not significant in Repetition 2 or 3. In Study 2, comparing children with adults, both analyses (LSA, CBOW) revealed significant main effects of age group and list repetition. The interaction of age group × repetition was not significant. As in Study 1, semantic relatedness of responses decreased over list repetitions. Adults' responses were more related to the cues than children's responses, with the difference maintained over repetitions.

*Network Models.* Shared associations (i.e., responses produced by more than one participant) were subjected to network modeling with the cues serving as the nodes in the resulting network. We distinguished three types of shared associations: Local Only (the response occurred multiple times and always with the same cue, e.g., "applause" produced twice in response to *clap*), Global Only (the response occurred multiple times but each time with a different cue, e.g., "fog" produced once in response to *frog* and once in response to *dog*), Global with Local (the response occurred multiple times across the same and different cues, e.g., "green" produced twice in response to *tree* and twice in response to *turtle*). Across studies, Chi-Square analyses indicated higher proportions of Local Only shared associations and lower proportions of Global Only shared associations in the older age group. That is, the responses tended to be more cue-specific and conventional in older children (Study 1) and adults (Study 2). In the models, the larger number of cue-specific responses resulted in slightly lower global clustering coefficients in the lexical networks of older participants. To extract the community

structure of each network, we applied the hierarchical algorithm in Mathematica (Peay, 1974; Wolfram Research Inc., 2014). Here, the increased number of cue-specific responses in older participants resulted in more differentiated community structure (Study 1: 11 communities for younger children vs. 13 for older children; Study 2: 12 communities for children vs. 13 for adults). In the older groups, the cues in each community reflected more coherent sets of thematic and taxonomic relations (e.g., animals, foot-related, school-related, sleep-related) and more of the communities consisted of a single cue. These results suggest that, as children develop, increased numbers of conventional, cue-specific associations may serve to distinguish words meanings, adding structure to the developing lexicon and contributing to more efficient lexical search in circumscribed portions of the network.

References


Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS one*, *6*(5), e19348. https://doi.org/10.1371/journal.pone.0019348

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods, 45*(2), 480-498. https://doi.org/10.3758/s13428-012-0260-7

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), 211-240. https://doi.org/10.1037/0033-295X.104.2.211

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57-78. https://doi.org/10.1016/j.jml.2016.04.001

Peay, E.R. (1974). Hierarchical clique structures. *Sociometry*, *37(1),* 54-65. https://www.jstor.org/stable/2786466

Peters, R., & Borovsky, A. (2019). Modeling early lexico-semantic network development: Perceptual features matter most. *Journal of Experimental Psychology: General*, *148*(4), 763-782. https://doi.org/10.1037/xge0000596

Song, S., Su, M., Kang, C., Liu, H., Zhang, Y., McBride-Chang, C., ... & Shu, H. (2015). Tracing children's vocabulary development from preschool through the school-age years: An 8-year longitudinal study. *Developmental Science*, *18*(1), 119-131. https://doi.org/10.1111/desc.12190

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, *15*(1), 8-25. https://doi.org/10.1080/10888438.2011.536125

Wojcik, E. H. (2018). The development of lexical–semantic networks in infants and toddlers. *Child Development Perspectives*, *12*(1), 34-38. https://doi.org/10.1111/cdep.12252

Wolfram Research Inc. (2014). Mathematica, Version 10.0, Champaign, IL.