Modeling the Development of Lexicon with DevLex: A Self-Organizing Neural Network Model of Lexical Acquisition

Ping Li & Igor Farkaš (pli, ifarkas@richmond.edu)
Department of Psychology, University of Richmond, Richmond, VA 23173, USA

It is well known in developmental psycholinguistics that young children produce a significant amount of speech errors at a certain stage, which is often associated with U-shaped learning (Bowerman, 1982). Not only do children show errors in morphological development (e.g., the well-known past tense errors), they also display errors in word naming and in spontaneous word production. Various accounts have been offered for these errors, from the child's confusion of related semantic fields, to the child's semantic reorganization, to the child's inability to retrieve the correct items in lexical memory during production. In this study, we investigate the nature and origin of these errors as a function of the organization and reorganization of a developmental lexicon, using DevLex as the basis of our modeling.

DevLex is a self-organizing neural network model that has the following properties (Farkaš & Li, 2002). The model consists of two self-organizing maps interconnected by associative links. The two maps attempt to capture the organization of word meanings (the semantic map) and word pronunciations (the phonological map). The semantic map is a dynamically growing network that learns word representations derived from word cooccurrence matrices in child-directed speech. It organizes word representations incrementally, adding new units to areas with higher lexical density. Semantic map is connected with phonological map that is pre-trained on the PatPho representations (Li & MacWhinney, 2002) of the 550 toddler word-list from the CDI lexical norms (Dale & Fenson, 1996). At every iteration during simulation, the semantic and phonological representations of words are simultaneously presented to both maps. Through self-organization (using Kohonen's algorithm), the network forms an activity on the phonological map in response to the phonological input, and an activity on the semantic map in response to the semantic input. For every word being presented, the model simultaneously learns associative connections between the two maps through Hebbian learning. DevLex is evaluated with respect to its accuracy in semantic representations and in productions (via associative connections from semantics to phonological output).

Our simulation results indicate that DevLex is able to model and provide insights into a range of phenomena in the early lexical acquisition. In particular, the onset of errors (measured as number of words confused in the network) reflects a number of important factors that govern lexical development. (1) The lexical categories of nouns, verbs, adjectives, and closed-class words have different profiles over the course of lexical growth, and their relative proportion, type frequency, and token frequency all affect the kinds of errors made in the network. This finding confirms Bates et al's (1994) argument on the important role that word category composition plays in lexical acquisition. (2) Number of words confused in the semantic map and in production is directly related to word density, measured as the amount of words mapped onto the nearest neighborhood of the target word. In addition, lexical confusion occurs more often for nouns than for other word categories, because of a "noun-bias" in the early vocabulary. This result shows that the source of children's naming errors may be the tight competition among similar neighbors in densely populated regions of the lexicon, consistent with the view that word density can predict the speed and accuracy in children's lexical access (Charles-Luce & Luce, 1990). (3) The rate of vocabulary expansion influences the rate of lexical confusion: the more related words that the network has to learn within a given period, the more likely it will show inaccuracies in the semantic map and in production. This pattern is consistent with the hypothesis that rapid increase in the rate of new words predicts the increase in children's naming errors (Gershkoff-Stowe & Smith, 1997). In sum, DevLex provides a new connectionist model that can simulate a developmental lexicon and relate to realistic language learning with self-organizing principles.

Acknowledgments

This research is supported by NSF (grant #BCS-9975249 awarded to P.L.).

References

Bowerman, M. (1982). Reorganizational processes in lexical and syntactic development. In E. Wanner & L. Gleitman (Eds.), *Language Acquisition: The State of the Art.* Cambridge: Cambridge University Press.

Farkaš, I., & Li, P. (2002). Modeling the development of lexicon with a growing self-organizing map. In H.J. Caulfield et al. (Ed.), *Proceedings of the 6th Joint Conference on Information Sciences* (pp. 553-556), JCIS/Association for Intelligent Machinery, Inc.

Li, P., & McWhinney, B. (2002). PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments and Computers*. (in press)

Dale, P.S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments and Computers, 28*, 125-127.

Bates, E. et al. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21*, 85-123.

Charles-Luce, J., & Luce, P.A. (1990). Similarity neighborhoods of words in young children's lexicons. *Journal of Child Language, 17*, 205-215.

Gerskoff-Stowe, L., & Smith, L.B. (1997). A curvilinear trend in naming errors as a function of early vocabulary growth. *Cognitive Psychology, 34*, 37-71.