Large-Scale Modeling of Lexical Processes

Christopher T. Kello (ckello@gmu.edu)
Department of Psychology, George Mason University
Fairfax, VA 22015 USA

Connectionist models of lexical processing have come a long way over the past 25 years. The literature contains small-scale models that have been used in support of specific arguments (e.g., Kello & Plaut, 2003), as well as larger-scale models that have been used in support of general theories (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996). All of these models have been invaluable in providing some theoretical grounding for debates about lexical acquisition and processing, both normal and impaired.

However, all extant models have at least one limitation in common: they only simulate performance with monosyllabic words. This limitation is serious because it calls into question whether the current computational principles and findings will hold up when the full extent of a lexicon is considered. Perhaps the biggest challenge of multisyllabic words is their variability in length.

In current models, orthographic and phonological representations are *slot-based*: each representational unit stands for one or more letters or phonemes in a given position (e.g., one unit might represent the letter P at the beginning of a word). Such slot-based codes raise problems of alignment. How should positions be coded in order to capture the relevant similarities among words that differ in length and hierarchical structure? To illustrate, how can slot-based representations approximate the orthographic and phonological similarities among words like WITH, WHITER, HITHER, WHITENER, HEREWITH, and WITHHOLD? Better yet, how can such representations be learned?

Recently, Kello and his colleagues introduced a connectionist architecture, termed the *autosequencer*, that learns fixed-length representations of variable-length sequences (Kello, Sibley, & Colombi, 2004). The architecture embodies a general theory in which *junction* representations are learned to serve the dual-purpose of 1) integrating sequences of perceptual inputs relevant to a word, as well as 2) generating sequences of motor outputs relevant to word. The inputs and outputs might be speech sounds or letters, for instance. Kello et al. demonstrated that the architecture is capable of learning orthographic and phonological representations of words, even when those words are multisyllabic and vary from 1 to 20 elements in length.

I will present a large-scale model of lexical processing that is based on the theory of junction representations. The model includes representations for over 45,000 words of English, and is able to process just as many nonwords. The model uses autosequencers to learn orthographic and phonological representations of words, and it uses text co-occurrences to learn semantic representations (the latter are proxies for what are planned to be junction representations in future models). I will present preliminary tests of the model against naming and lexical decision response times and error rates for over 30,000 words in the Elexicon database (Balota et al., 2002).

The model architecture was originally inspired by empirical results from the tempo-naming task (Kello, 2004), and modeling work on phonological development (Plaut & Kello, 1999). One can think of the architecture as an implementation and extension of the analogy model originally proposed by Glushko (1979). Localist nodes are used to coordinate interactions among orthographic, phonological, and semantic representations, which are all coded in a distributed fashion. Both words and nonwords invoke graded patterns of activation across the localist nodes, which implement an analogical process of sorts.

References

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., et al. (2002). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. http://elexicon.wustl.edu.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception & Performance, 5*, 674-691.

Kello, C. T. (2004). Control over the time course of cognition in the tempo-naming task. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 942-955.

Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory & Language, 48*(1), 207-232.

Kello, C. T., Sibley, D. E., & Colombi, A. (2004). Using simple recurrent networks to learn fixed-length representations of variable-length strings. In *Proceedings of the AAAI Symposium on Compositional Connectionism.* Washington, DC.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The emergence of language* (pp. 381-415). Mahweh, NJ: Erlbaum.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115.