Cognitive Science 33 (2009) 1183–1186

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01062.x

Learning Representations of Wordforms With Recurrent Networks: Comment on Sibley, Kello, Plaut, & Elman (2008)

Jeffrey S. Bowers,[a] Colin J. Davis[b]

[a]*Department of Experimental Psychology, University of Bristol*
[b]*Department of Psychology, Royal Holloway, University of London*

Received 12 February 2008; received in revised form 27 June 2008; accepted 6 August 2008

Abstract

Sibley et al. (2008) report a recurrent neural network model designed to learn wordform representations suitable for written and spoken word identification. The authors claim that their *sequence encoder* network overcomes a key limitation associated with models that code letters by position (e.g., CAT might be coded as C-in-position-1, A-in-position-2, T-in-position-3). The problem with coding letters by position (slot-coding) is that it is difficult to generalize knowledge across positions; for example, the overlap between CAT and TOMCAT is lost. Although we agree this is a critical problem with many slot-coding schemes, we question whether the *sequence encoder* model addresses this limitation, and we highlight another deficiency of the model. We conclude that alternative theories are more promising.

Keywords: Slot-coding; Connectionism; Alignment problem; Word identification; Symbols; Position-invariance

In a recent contribution to this journal, Sibley, Kellow, Plaut, and Elman (2008) report a recurrent PDP model designed to learn wordform representations suitable for written and spoken word identification. Their *sequence encoder* network is claimed to address a limitation associated with the input coding schemes commonly employed in models of reading, that is, models that include some version of slot-coding. In these models, each letter or phoneme is associated with a given position in a word (e.g., CAT might be coded as C-in-position-1, A-in-position-2, and T-in-position-3). These coding schemes fail to explain a variety of empirical findings (e.g., Bowers, Davis, & Hanley, 2005; Davis, 1999; Davis 2006; Davis & Bowers, 2004, 2006; Grainger, Granier, Farioli, Van Assche, & van Heuven,

Correspondence should be sent to Jeffrey S. Bowers, Department of Experimental Psychology, University of Bristol, 12a Priory Road, BS8-1TU, Bristol, UK. E-mail: j.bowers@bris.ac.uk

2006; Perea & Lupker, 2003, 2004), but the critical limitation noted by the authors is that slot-coding makes it difficult to generalize knowledge across positions. For example, the words CAT and TOMCAT are composed of entirely different input codes, given that the CAT in TOMCAT is coded by C-in-position-4, A-in-position-5, and T-in-position-6. Davis (1999) termed this the ''alignment problem''.

In an attempt to address this problem, Sibley et al. developed a recurrent connectionist model that takes a series of letters or stress-marked phonemes as input and reproduces them in the correct sequence at an output layer. This sequence encoder learned to reproduce over 70,000 written and spoken words. Furthermore, its performance was sensitive to orthotactic and phonotactic regularities of the training regime: It was much better at reproducing the letters and phonemes of pseudowords compared to illegal sequences. It should be noted, however, that existing PDP and localist models already exhibit sensitivity to orthotactic structure. Indeed, such effects can be observed in the Interactive Activation model of word identification and its variants (e.g., Grainger & Jacobs, 1996), as well as the SOLAR model, which learns localist word representations (Davis, 1999).

According to the authors, a critical feature of the model is that it learns a unique distributed pattern of activation for each letter in a given position of a given word length; for example, the C in CAT is coded as C-in-position-1-for-3-letter-words. The representations of letters overlap to the extent that they are encoded in similar positions, and they come from words of a similar length. This is said to solve the alignment problem: ''Sequence encoder representations do not engender the alignment problem because conjunction patterns are created for length-specific positions (e.g., the last position in a 5-letter word)'' (p. 7). Nothing else is said, nor any evidence presented, to support this conclusion. However, if the model develops conjunctive codes for letter identities at a given position for a given word length, the C in CAT and TOMCAT should be coded very differently, as they occur in different places, and in different word lengths. Thus, it is hard to see how generalization is improved.

Indeed, their analysis of the hidden units provides strong evidence against their claim. Sibley et al. carried out a Principle Components Analysis on activation patterns of the hidden units in order to estimate the pattern associated with a given letter in a given position within a word of a given length—what they called the conjunctive pattern for each letter. They then report the average pairwise correlation between conjunctive patterns between same and different letters as a function of the position of the letter and the length of the word. Their Figure 2a presents the similarity of two letters as a function of the number of intervening letters, after partialling out word length. These data show a perfect correlation for same letters in the same position (i.e., the A in ACT and ART are coded by the same distributed pattern), but with three or more intervening letters the similarity is almost entirely eliminated, such that the letters C in CARROT and BOXCAR are not much more similar than the C in CARROT and the Z in FROZEN. Figure 2b of Sibley et al. presents the similarity of two letters embedded in different words as a function of relative word length. The similarity of the same letter was reduced as a function of the difference in word length, such that the same letter embedded in words that differ by three or more letters was smaller than for different letters embedded in words of the same length. For example, the letters C in

CAT and TOMCAT are less similar than the C in CAT and the D in DOG. Given this, it is hard to see how the model solves the alignment problem. In this model, the overlap between the C-A-T in CAT and TOMCAT could not support recognition of embedded words.

By contrast, alternative letter encoding schemes, including spatial coding (Davis, 1999), and various open-bigram models (e.g., Grainger et al., 2006; Whitney, 2001) do solve the alignment problem. For example, in the spatial coding scheme, letters are coded in long-term memory independent of position (the same A unit is involved in coding CAT and TOMCAT), and the input TOMCAT will fully activate the lexical representation for CAT despite the fact that the letters C-A-T occur in different positions and are embedded of words of different lengths.

Another limitation of the model merits brief mention. As noted above, one of the model's key successes is its sensitivity to orthotactic and phonotactic regularity, as illustrated by its poor generalization to illegal letter/phoneme sequences. However, this sensitivity may well be the full extent of the model's knowledge: There is no evidence that the model has acquired the lexical knowledge that would allow it to distinguish words from nonwords. For example, in their Simulation 2, the sequence encoder was trained with ~75,000 written words taken from the *Wall Street Journal* corpus. The model was then tested on the trained items, well-formed pseudowords or illegal letter strings. Performance was similar for the words and pseudowords, and the small differences may have been because of the words having the more common orthotactic constructions. Indeed, when the orthotactics of trained and untrained letter strings were more closely matched in Simulation 1, performance was essentially the same for the trained (100%) and untrained (98%) wordforms. Although the model treats familiar and unfamiliar items similarly, it is straightforward for skilled readers to distinguish words (e.g., ''word'') from pseudowords (e.g., ''werd'') that are matched on sublexical factors such as bigram and trigram frequency. Unless the authors can provide evidence that the model can distinguish between words and nonwords (independent of orthotactic regularity), it seems premature to conclude that ''The sequence encoder is a model of lexical performance in its own right, as demonstrated by its ability to account for data on language user's sensitivity to phonotactics and orthotactics'' (p. 11).

Of course, the work of distinguishing words from nonwords (and reading regular as well as irregular words, etc.) might be accomplished by another component of the lexical system. Indeed, the authors discuss the possibility of using the sequence encoder as the input and output coding schemes for the Seidenberg and McClelland (1989) or Plaut, McClelland, Seidenberg, and Patterson (1996) model of word naming. Thus, lexical knowledge may only be embedded within a scaled-up model that includes the sequence encoder as one part. However, unless the sequence encoder can be modified so that it can solve the alignment problem, it is not clear what advantage this input coding scheme provides.

References

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Automatic semantic activation of embedded words: Is there a 'hat' in 'that'? *Journal of Memory and Language*, *52*, 131–143.

Davis, C. J. (1999). The Self-Organising Lexical Acquisition and Recognition (SOLAR) model of visual word recognition. Unpublished doctoral dissertation, University of New South Wales.

Davis, C. J. (2006). Orthographic input coding: A review of behavioural data and current models. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing p* (pp. 180–206). New York: Psychology Press.

Davis, C. J., & Bowers, J. S. (2004). What do letter migration errors reveal about letter position coding in visual word recognition? *Journal of Experimental Psychology: Human Perception & Performance*, *30*, 923–941.

Davis, C. J., & Bowers, J. S. (2006). Contrasting five theories of letter position coding. *Journal of Experimental Psychology: Human Perception & Performance*, *32*, 535–557.

Grainger, J., Granier, J. P., Farioli, F., Van Assche, E., & van Heuven, W. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 865–884.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518–565.

Perea, M., & Lupker, S. J. (2003). Does jugde activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory & Cognition*, *31*, 829–841.

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, *51*, 231–246.

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.

Seidenberg, M. S., & McClelland, J. L (1989). *A distributed, development model of word recognition and naming Psychological Review*, *96*, 563–568.

Sibley, D. E., Kello, C. T., Plaut, D., & Elman, J. L. (2008). Large-scale modeling of wordform learning and representation. *Cognitive Science*, *32*, 741–754.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review*, *8*, 221–243.