Neural networks for word recognition: Is a hidden layer necessary?

Frédéric Dandurand†, Thomas Hannagan‡, and Jonathan Grainger†

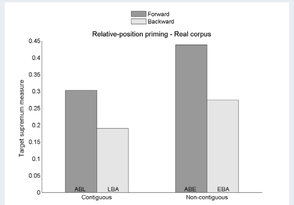
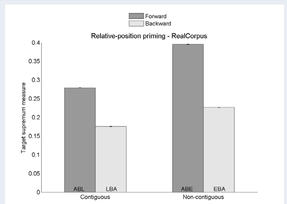
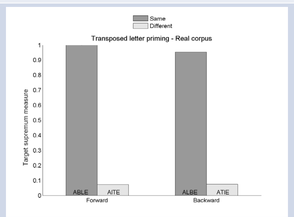
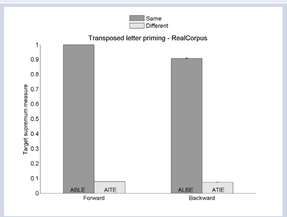
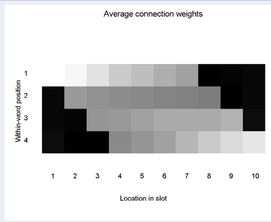
†Laboratoire de psychologie cognitive, CNRS & Université de Provence, 3 Place Victor Hugo, 13331 Marseille Cedex 3, France

‡Laboratoire de Sciences Cognitives et Psycholinguistique, EHESS/CNRS/DEC-ENS, École Normale Supérieure, 29 rue d'Ulm, 75005 Paris

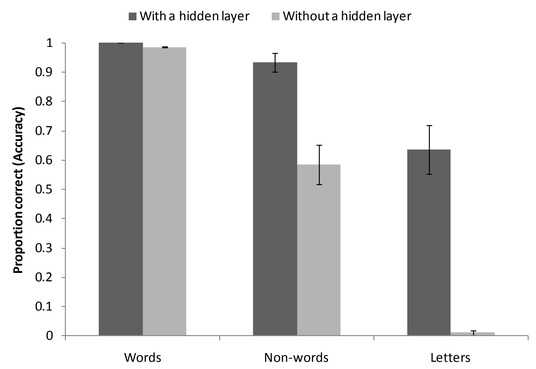
European Research Council

Task: Learn to map letters presented in all 7 positions of an alphabetic array (i.e., location-invariance) to 1179 English words of four letters (e.g., #ABLE##### where # are blank slots)
Coding: Local (sparse) – inputs: 26 binary letter values per slot position – outputs: one unit per word **Model**: feedforward neural network (backpropagation learning with momentum)

Feedforward model comparison

	With a hidden layer	Without a hidden layer
Network topology	1179 outputs 91 hidden units 260 inputs	1179 outputs 260 inputs
Relative-position priming 2 factors manipulated (1) order (2 levels) forward (for ABLE : ABL, BLE, ABE, ALE) backward (for ABLE : LBA, ELB, EBA, ELA) (2) contiguity (2 levels) contiguous (for ABLE : ABL, BLE, LBA, ELB) non-contiguous (for ABLE : ABE, ALE, EBA, ELA)		
Transposed-letter priming 2 factors manipulated (1) origin of central, or inner, letters (2 levels) same word (for ABLE : ABLE, ALBE) different word (for ABLE : AITE, ATIE) (2) order (2 levels) forward (for ABLE : ABLE, AITE) backward(for ABLE : ALBE, AITE)		
Accuracy (all words, anagrams)	100%, 100%	98.6%, 95%
Model size and complexity	91 hidden units 132 219 connection weights	> 0 hidden unit < 307 719 connection weights
Coding & segregating anagrams	Holographic overlap coding Small changes of the inputs result in small differences in hidden layer representations Differences in patterns of hidden layer activations are monotonically related to differences in identity and position of input letters. (Hannagan, Dandurand, & Grainger, in press)	 Significant interaction between Location and Position ($p < 0.001$)

Lexical decision



Condition	Examples	Success criterion
word	#ABLE##### ######WITH ###DOOR###	Output unit of target word (only) activated above threshold (0.9)
Nonword (random strings)	#JKTS##### ######HIQL ###BXGA###	All output units activated below threshold (0.9)
Letters only	#AAAA#### #####HHHH# #LLLL#####	All output units activated below threshold (0.9)

Conclusions

1. Hidden layer necessary for word-nonword discrimination, but not for priming
2. Coding in networks without a hidden layer: essentially based on the presence of letters but where letter contributions are modulated using the interaction between within-word position and within-slot location. This modulation allows networks to factor in some information about letter position, which is sufficient to segregate most (95%) anagrams.

References

Dandurand, F., Grainger, J., & Dufau, S. (2010). Learning location invariant orthographic representations for printed words. *Connection Science, 22*(1), 25-42. doi:10.1080/09540090903085768

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes, 23*(1), 1-35.

Grainger, J., & van Heuven, W. J. B. (2003). Modeling letter position coding in printed word perception. In *The Mental lexicon* (pp. 1-23). New York: Nova Science Publishers.

Hannagan, T., Dandurand, F., & Grainger, J. (in press). Broken symmetries in a location invariant word recognition network, *Neural Computation*.

McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing*. Boston, MA: MIT Press.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227-234.

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*(1), 56-115.

Ratcliff, R., McKoon, G., & Gomez, P. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review, 111*(1), 159-182.

Rayner, K., White, S., Johnson, R., Liversedge, S. (2006). Raeding Wrods With Jubmled Lettres; There Is a Cost. *Psychological Science, 17*(3), 192-193

Contact: frederic.dandurand@gmail.com

This project was supported by the Agence Nationale de la Recherche (grant no. ANR-06-BLAN-0337) and the European Research Council (ERC-230313).

Presented at the annual meeting of the Cognitive Science Society (CogSci 2010), Portland, August 11 -14, 2010.