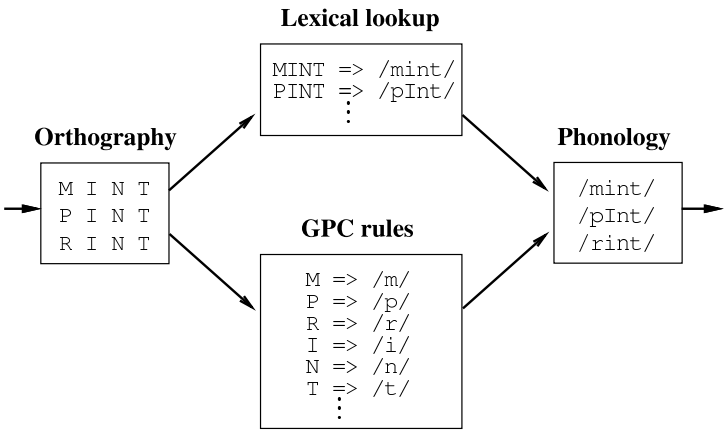
Dual-route theory of word reading

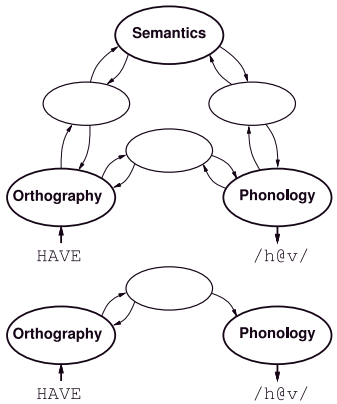
- Systematic spelling-sound knowledge takes the form of grapheme-phoneme correspondence (GPC) rules (e.g., G ⇒ /g/, A_E ⇒ /A/)
- Applying GPC rules produces correct pronunciations for *regular* words (GAVE) and *nonwords* (MAVE), but incorrect pronunciations for *exception* words (HAVE)
- Exception words therefore require a separate lexical look-up procedure



Seidenberg and McClelland (1989, *Psych. Rev.*)

Method

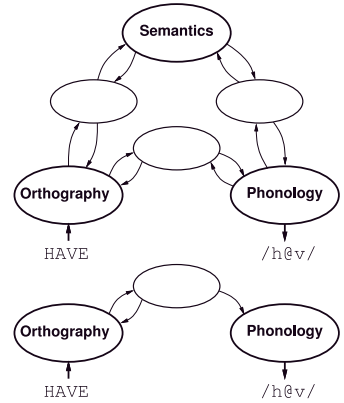
- Feedforward network trained with back-propagation to pronounce 2897 monosyllabic words, sampled proportional to logarithm of actual word frequencies.
- Representations of orthography and phonology based on context-sensitive triples of letters ($_{M}A_{K}$, "Wickelgraphs") or phonemic features ($_{stop}long_{fricative}$, "Wickelfeatures").



Seidenberg and McClelland (1989, *Psych. Rev.*)

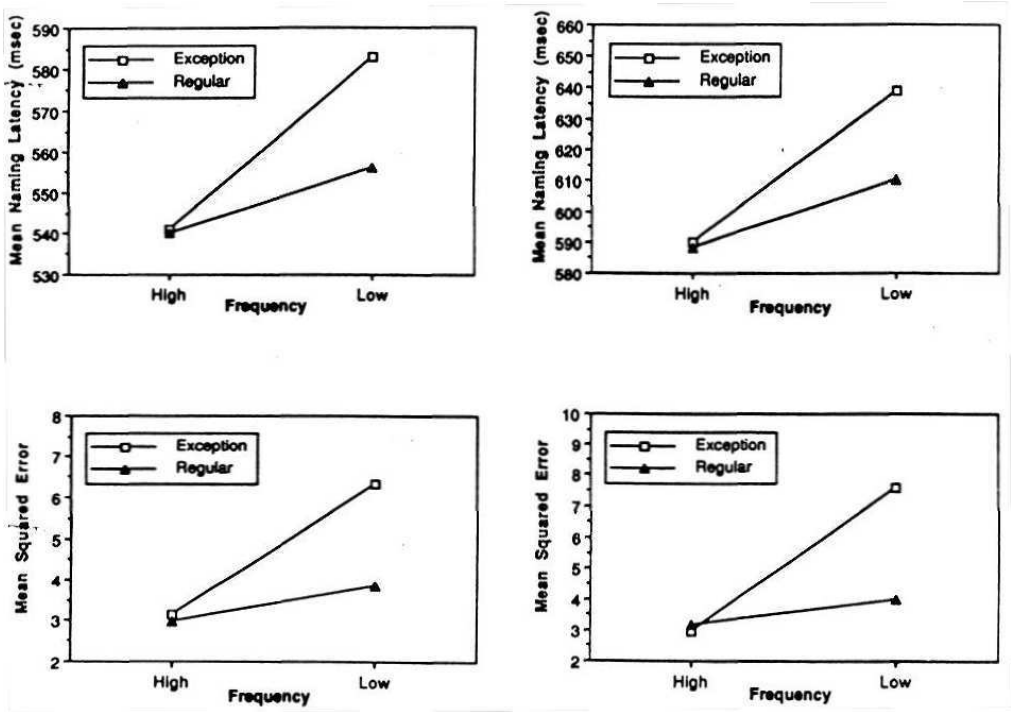
Method

- Feedforward network trained with back-propagation to pronounce 2897 monosyllabic words, sampled proportional to logarithm of actual word frequencies.
- Representations of orthography and phonology based on context-sensitive triples of letters ($_{M}A_{K}$, "Wickelgraphs") or phonemic features ($_{stop}long_{fricative}$, "Wickelfeatures").



Results

- After 250 training epochs, network correctly pronounces 97.3% of words, including most exception words.
- Error pattern accounts for many empirical effects of frequency and consistency on naming latencies.



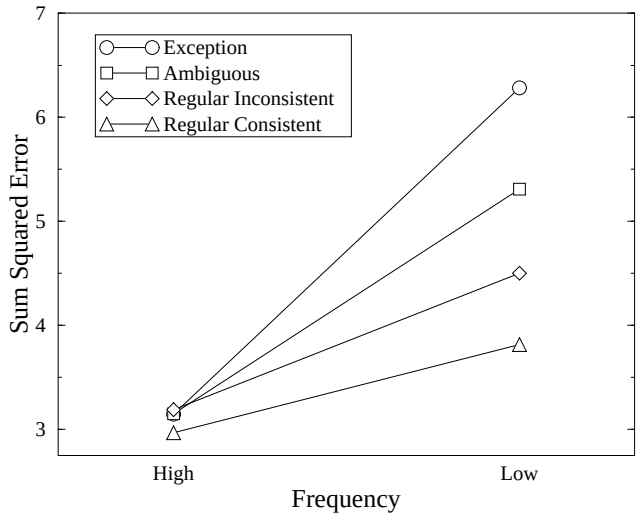
Frequency-by-consistency interaction (SM89)

Regular Consistent: Many *friends* and no *enemies* [LATE, DUST]

Regular Inconsistent: Many friends but at least one enemy [GAVE (cf.HAVE), MINT (cf.PINT)]

Ambiguous: About equal numbers of friends and enemies [DOWN, KNOWN; POUR, SOUR]

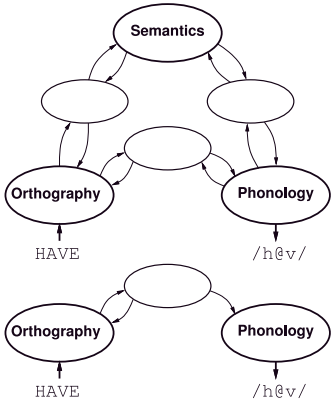
Exception: Few if any friends and many enemies [HAVE (cf. GAVE), PINT (cf. MINT)]



Seidenberg and McClelland (1989, *Psych. Rev.*)

Method

- Feedforward network trained with back-propagation to pronounce 2897 monosyllabic words, sampled proportional to logarithm of actual word frequencies.
- Representations of orthography and phonology based on context-sensitive triples of letters ($_{M}A_{K}$, "Wickelgraphs") or phonemic features ($_{stop}long_{fricative}$, "Wickelfeatures").



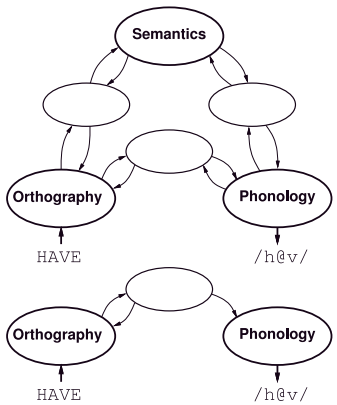
Results

- After 250 training epochs, network correctly pronounces 97.3% of words, including most exception words.
- Error pattern accounts for many empirical effects of frequency and consistency on naming latencies.
- Fails to pronounce nonwords as well as skilled readers.

Seidenberg and McClelland (1989, *Psych. Rev.*)

Method

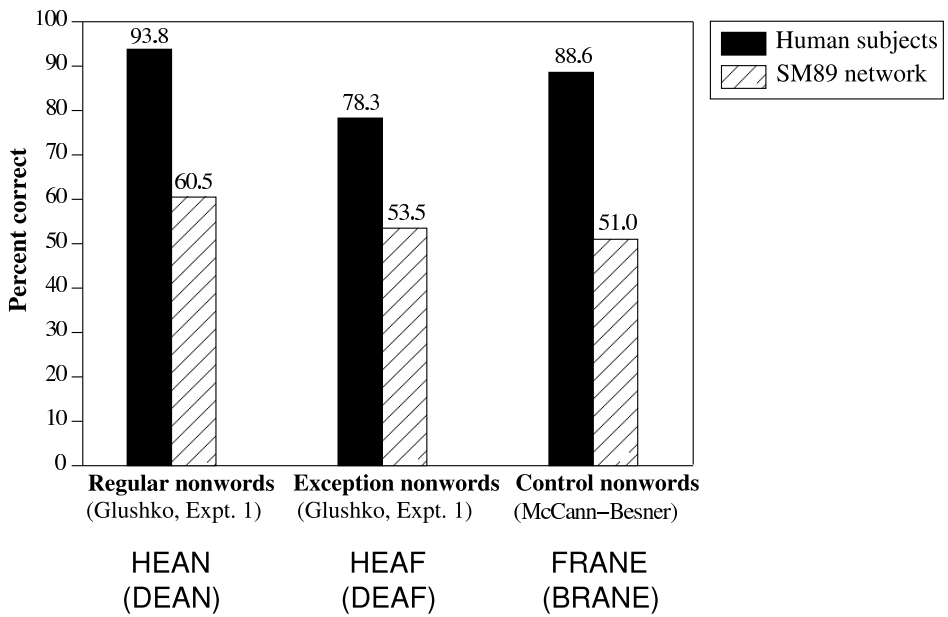
- Feedforward network trained with back-propagation to pronounce 2897 monosyllabic words, sampled proportional to logarithm of actual word frequencies.
- Representations of orthography and phonology based on context-sensitive triples of letters ($_{M}A_{K}$, "Wickelgraphs") or phonemic features ($_{stop}long_{fricative}$, "Wickelfeatures").



Results

- After 250 training epochs, network correctly pronounces 97.3% of words, including most exception words.
- Error pattern accounts for many empirical effects of frequency and consistency on naming latencies.

Nonword reading (SM89)



Representation and generalization: Condensing regularities

The "dispersion" problem

LOG			$_{*}L_{O}$	$_{L}O_{G}$	$_{O}G_{*}$
GLAD		$_{*}G_{L}$	$_{G}L_{A}$	$_{L}A_{D}$	$_{A}D_{*}$
SPLIT	$_{*}S_{P}$	$_{S}P_{L}$	$_{P}L_{I}$	$_{L}I_{T}$	$_{I}T_{*}$

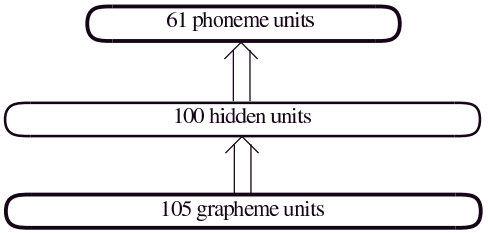
Capturing orthographic and phonological structure

- *Phonotactic constraints*: Possible phoneme sequences are strongly constrained by the structure of the articulatory system and by language-specific learning.
- *Alphabetic principle*: Parts of written words (graphemes) correspond to parts of pronunciations (phonemes).

Ordering of graphemes and phonemes is (virtually) unambiguous within clusters of consonants/vowels.

	Onset				Vowel			Coda		
	S	P	G	L	O	A	I	G	D	T
LOG				L	O			G		
GLAD			G	L		A			D	
SPLIT	S	P		L			I			T

Simulation: Feedforward network



Training

- Trained with back-propagation on 2998 monosyllabic words (SM89 corpus plus additional 101 words) using log-frequencies to scale weight changes
- Error measured by *cross-entropy* between states and targets:

$$C = -\sum_{j} t_j \log(a_j) + (1 - t_j) \log(1 - a_j)$$

- Adaptation of connection-specific learning rates (delta-bar-delta; Jacobs, 1988).
- After 300 training epochs, network pronounces the *entire* training corpus correctly (100% correct)

Representations

Phonology

Onset

s S C z Z j f v T D p b t d k g m n h l r w y

Vowel

a @ e i o u A E I O U W Y ^

Coda

r l m n N b g d ps ks ts s z f v p k t S Z T D C j

Orthography

Onset

Y S P T K Q C B D G F V J Z L M N R W H CH GH GN PH PS RH SH TH TS WH

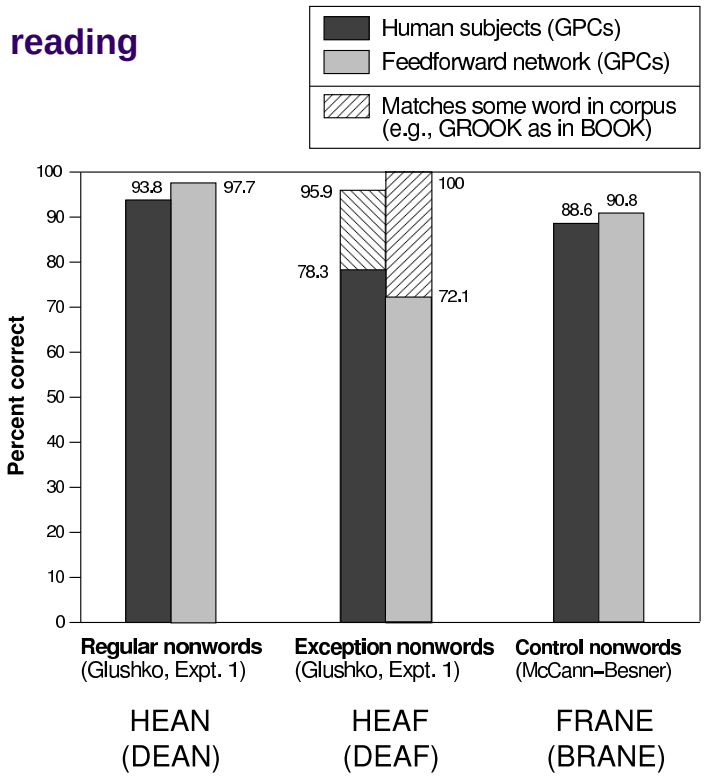
Vowel

E I O U A Y AI AU AW AY EA EE EI EU EW EY IE OA OE OI OO OU OW OY UE UI UY

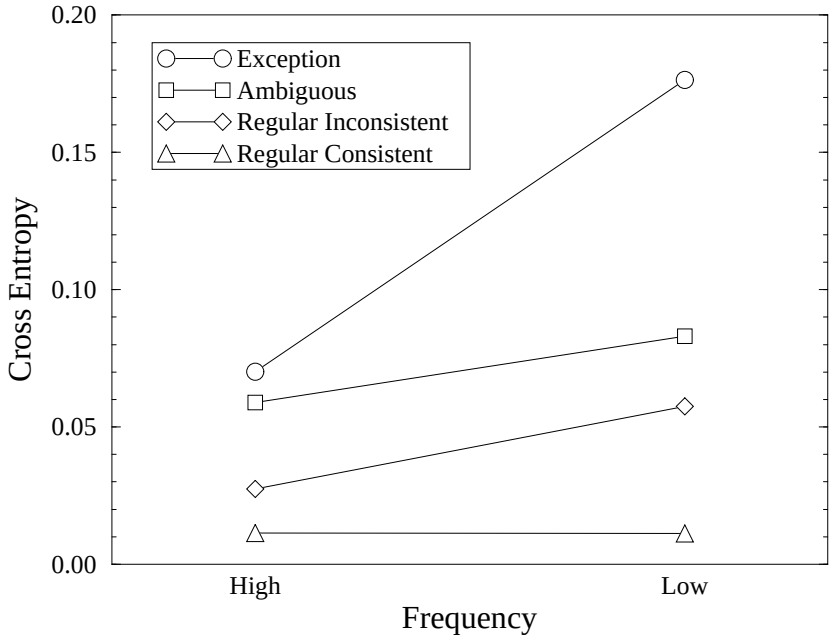
Coda

H R L M N B D G C X F V J S Z P T K Q BB CH CK DD DG FF GG GH GN KS LL NG NN PH PP PS RR SH SL SS TCH TH TS TT ZZ U E ES ED

Nonword reading

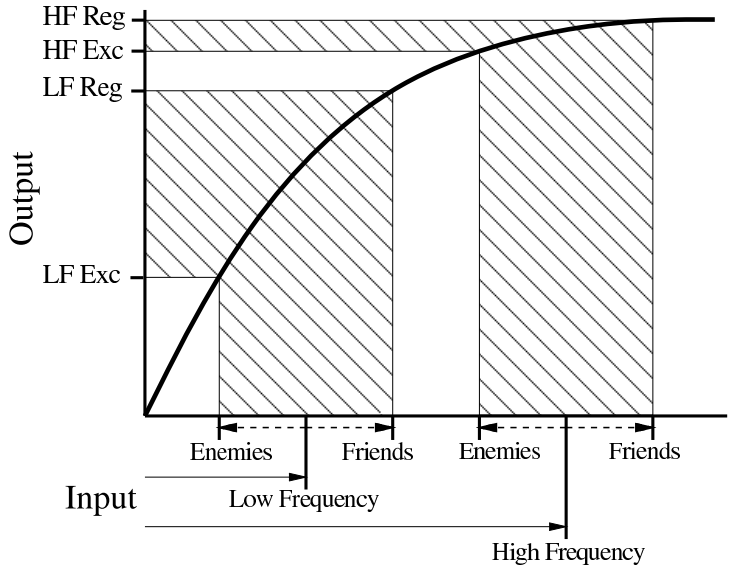


Frequency-by-consistency interaction



Frequency-by-consistency interaction

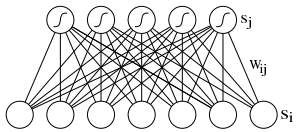
The *additive* combination of frequency and consistency has an nonlinear (asymptotic) effect on output activations



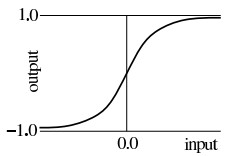
Analytic account of frequency/consistency effects

Train with Hebbian learning on set of patterns indexed by p:

$$\triangle w_{ij} = a_i a_j \qquad w_{ij} = \sum_p a_i(p)\, a_j(p)\, \mathrm{freq}(p)$$



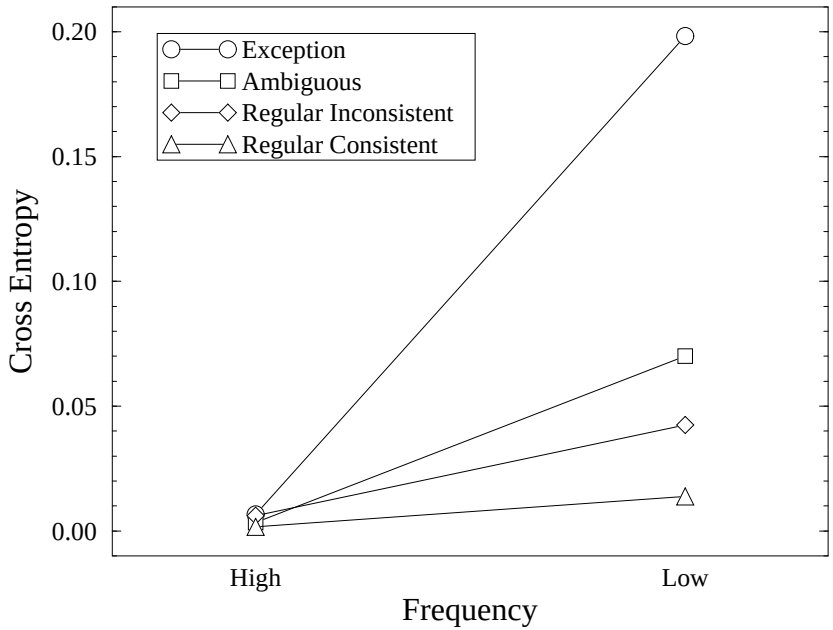
Response of output unit j to test pattern t:



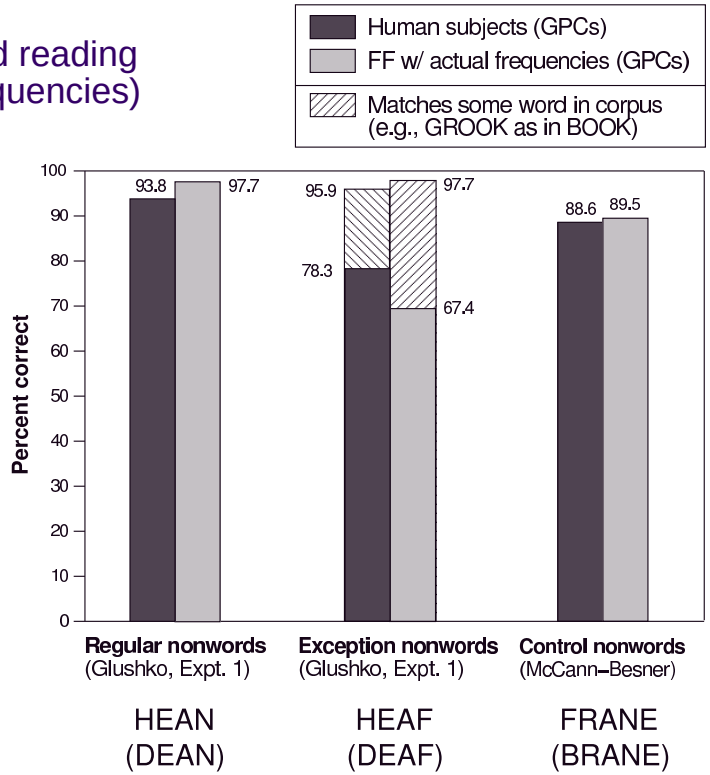
$$\begin{aligned} a_j(t) &= \sigma\left(\sum_i a_i(t)\, w_{ij}\right) \\ &= \sigma\left(\sum_p a_j(p)\, \mathrm{freq}(p) \sum_i a_i(t)\, a_i(p)\right) \\ &= \sigma\left(\sum_p a_j(p)\, \mathrm{freq}(p)\, \mathrm{sim}(t,p)\right) \\ &= \sigma\left(\mathrm{freq}(t) + \sum_f \mathrm{freq}(f)\, \mathrm{sim}(f,t) - \sum_e \mathrm{freq}(e)\, \mathrm{sim}(e,t)\right) \end{aligned}$$

where f indexes *friends* of pattern t and e indexes *enemies*.

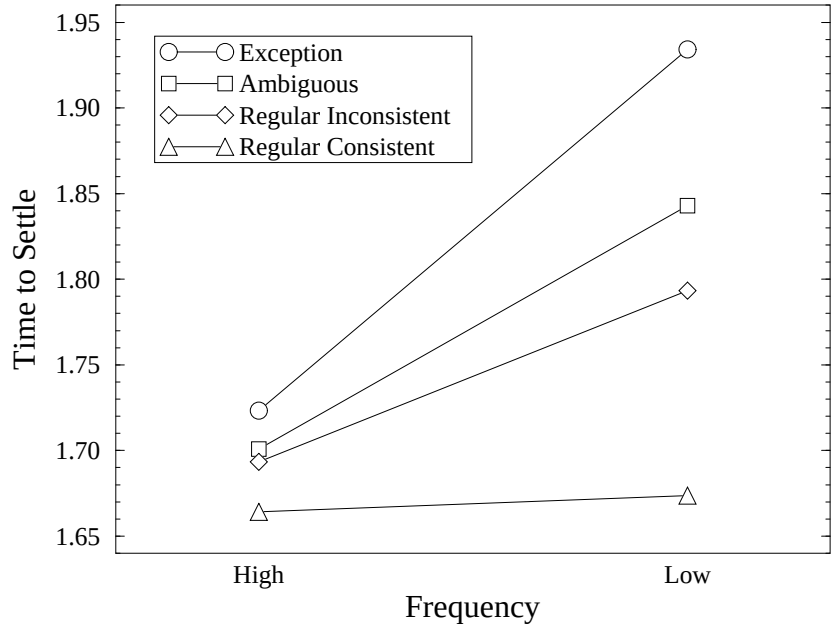
Frequency-by-consistency interaction (raw frequencies)



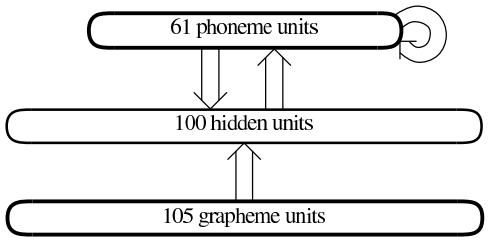
Nonword reading (raw frequencies)



Frequency-by-consistency interaction



Simulation: Attractor network

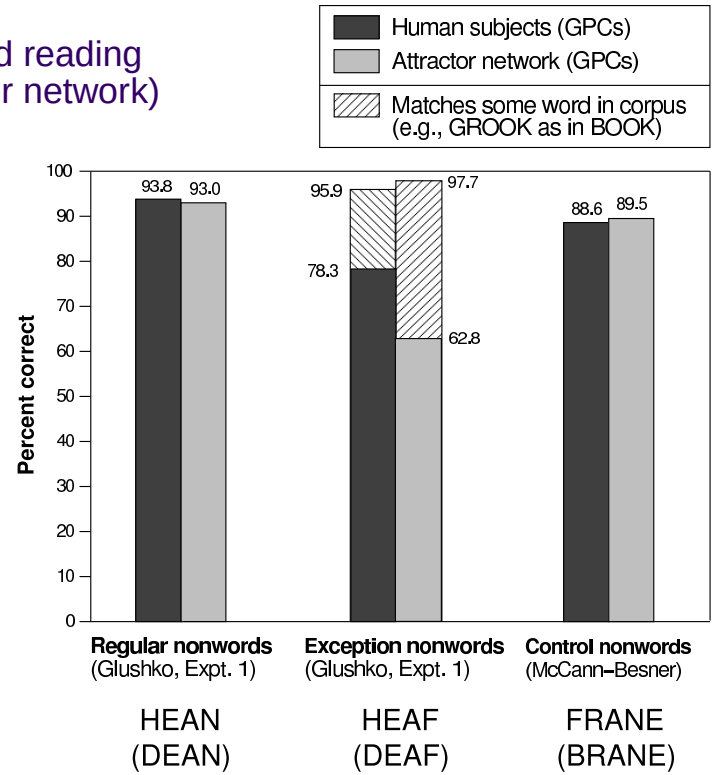


$$\eta_j^{[t]} = \tau \sum_i a_i^{[t-\tau]} w_{ij} \;+\; (1-\tau)\,\eta_j^{[t-\tau]} \qquad a_j^{[t]} = \frac{1}{1+\exp\left(-\eta_j^{[t]}\right)}$$

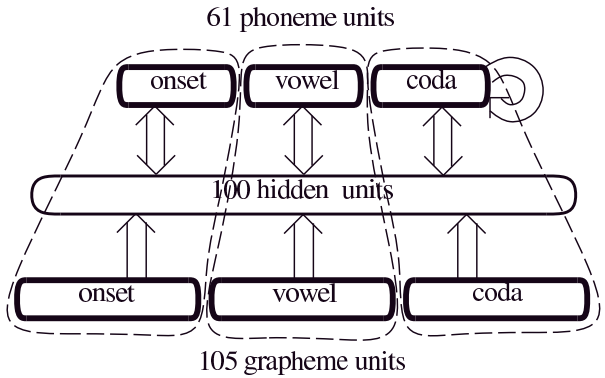
Training

- Trained with a continuous version of back-propagation through time (Pearlmutter, 1989), using actual word frequencies, cross-entropy, and delta-bar-delta
- Run for 2.0 units of time, receiving no error before time 1.0; discretization $\tau = 0.2$ reduced to 0.01 at end of training
- During testing, network responds when phoneme states stop changing
- After 1900 training epochs, the network pronounces all but 25 words correctly (99.2% correct)

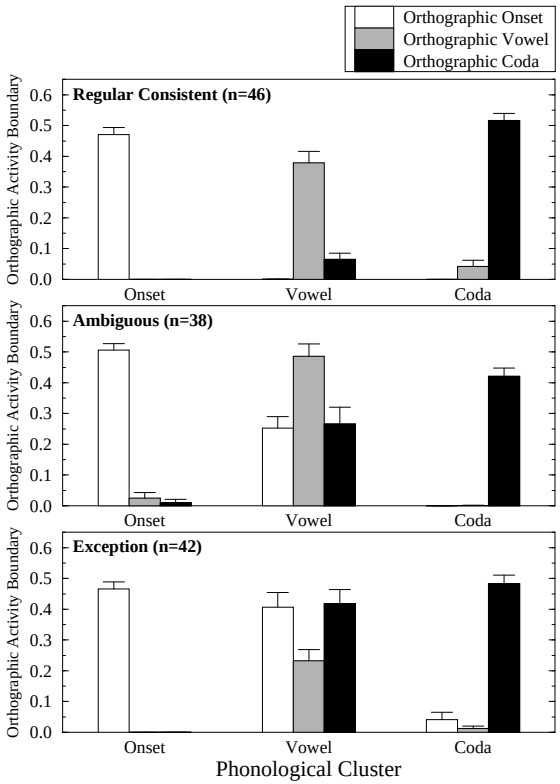
Nonword reading (attractor network)



Generalization with componential attractors

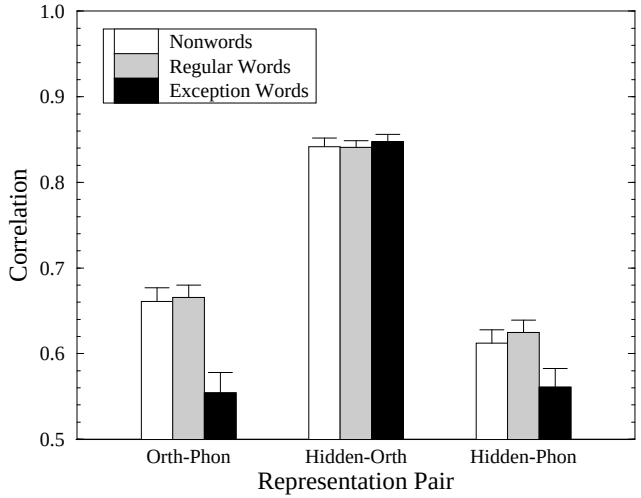


- Very strong orthography-phonology systematicity *within* consonant clusters (less for vowels); relative independence *between* clusters (except for vowels in exception words).
- Connectionist learning is sensitive to which parts of the input reliably predict (e.g., are correlated with) each part of the output.
- Network develops *componential* attractors for words that can recombine to support nonword reading.



Similarity structure among representations

- Two sets of representations are structured similarily if their pairwise similarities are correlated
- Tested with 48 body-matched triples of nonwords, regular words, and exception words (MAVE, GAVE, HAVE)



Impaired reading in “surface” dyslexia

- Brain damage to left temporal lobe (stroke, head injury, or degenerative disease) in premorbidly literate adult
- Severe impairment to semantics, or to mapping from semantics to phonology
- Word reading accuracy influenced by frequency and consistency:

	Correct Performance					
Patient	HFR	LFR	HFE	LFE	%Reg’s	NW
MP	95	98	93	73	90	95.5
KT	100	89	47	26	85	100

- Exception words produce *regularization* errors:

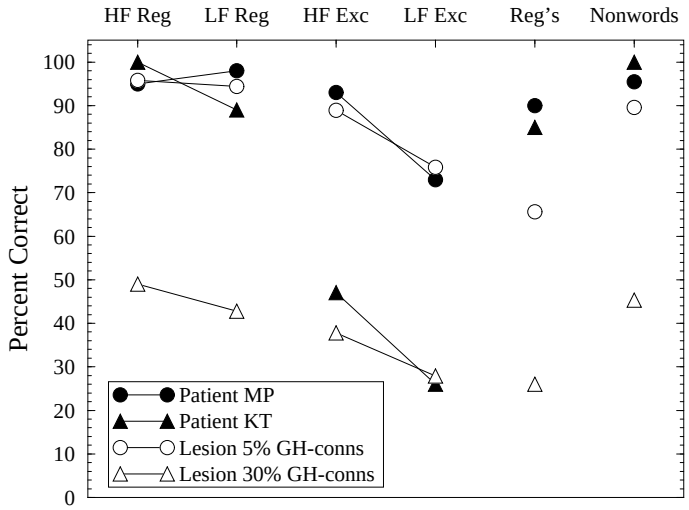
DEAF	⇒ “deef”	FLOOD	⇒ “flude”
SAID	⇒ “sayed”	GONE	⇒ “goan”
BROAD	⇒ “brode”	STEAK	⇒ “steek”
SHOE	⇒ “show”	SEW	⇒ “sue”
ONE	⇒ “own”	SOOT	⇒ “suit”

- Nonword reading accuracy is normal
- Word and nonword naming latencies are normal

Surface dyslexia: Damage to phonological pathway?

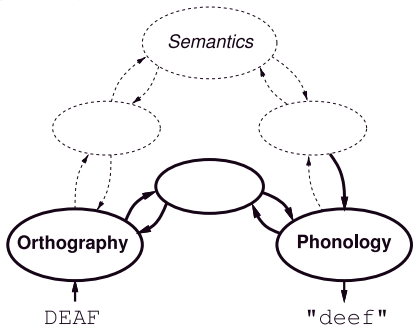
Attractor network: Lesioning procedure

- Remove specified proportion of connections between two groups of units
- Results averaged over 50 different instances of lesion at given severity/location



Division of labor between pathways

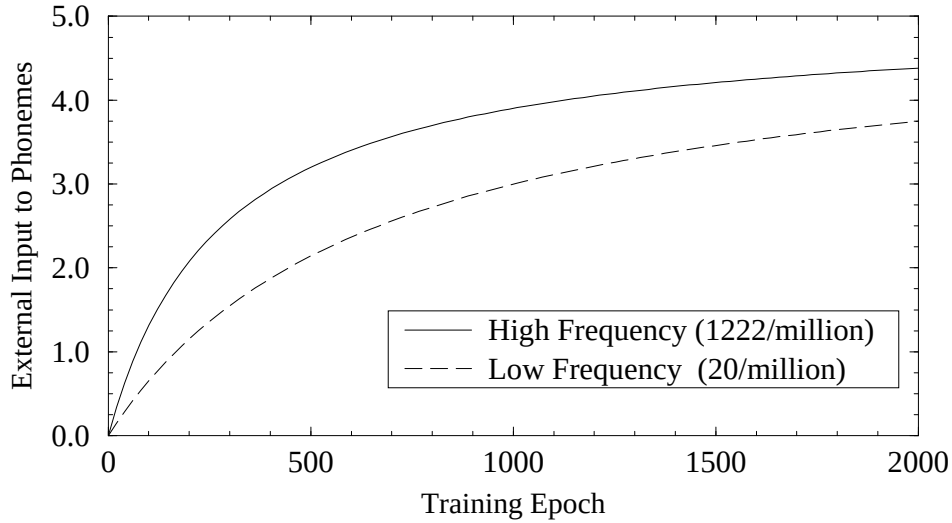
- Phonological and semantic pathways combine to support oral reading.
- As semantic pathway develops, demands on phonological pathway diminish.
- Removal of semantic pathway by brain damage reveals latent inadequacies of intact phonological pathway.

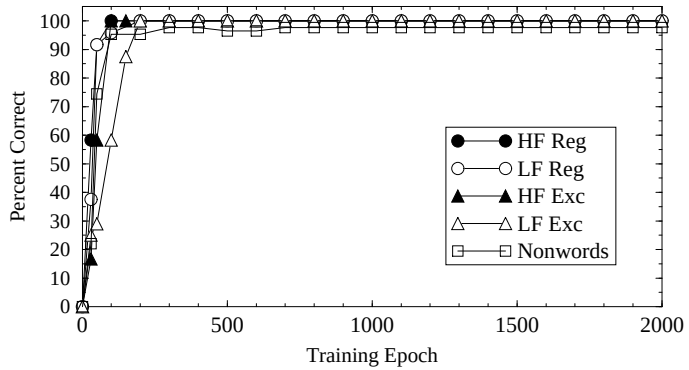


Simulation

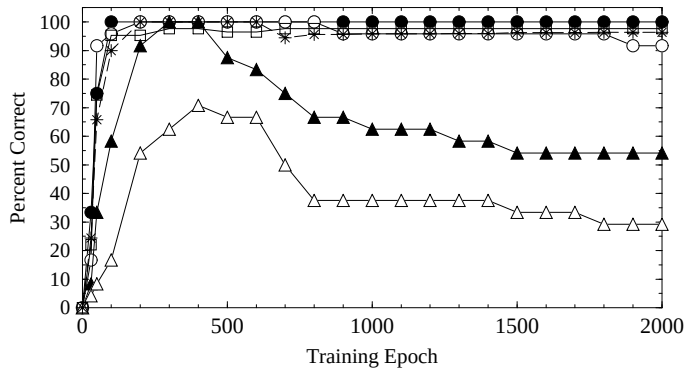
- Feedforward network with weight decay.
- Contribution of semantics approximated by external correct input to phoneme units that increases gradually and is frequency-sensitive.

Contribution of putative semantic pathway

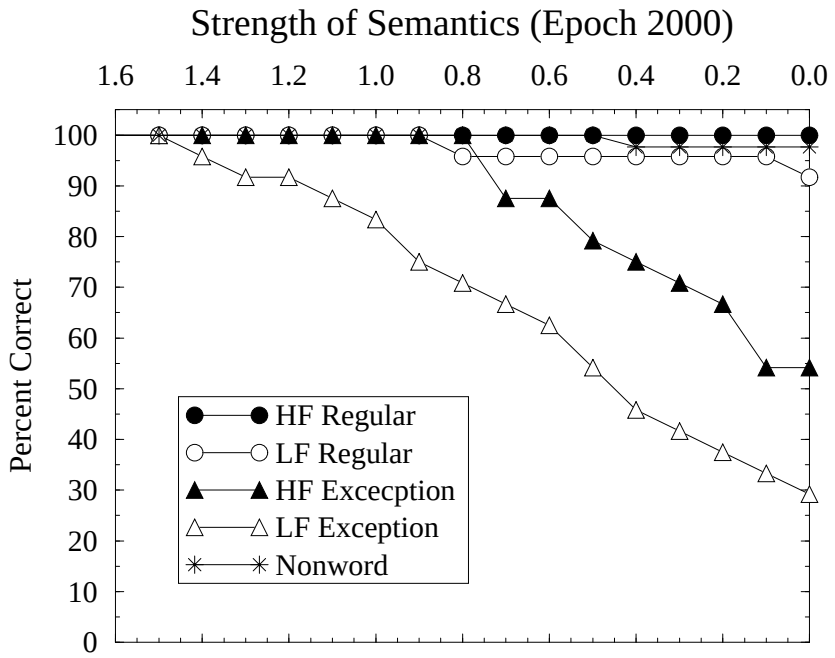




Phonological Pathway in Isolation (Semantics Removed)

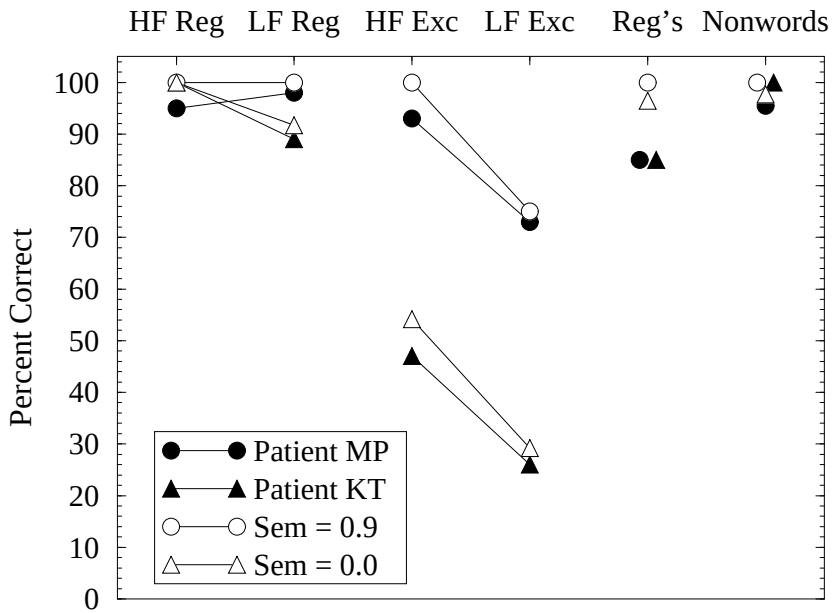


Effect of progressive semantic deterioration



Semantic impairment without surface dyslexia?

Surface dyslexia: Intact but specialized phonological pathway



Semantic impairment without surface dyslexia?

WLP (Schwartz et al., 1979); probable semantic dementia

- PPVT 37%; HF exception reading 98%
- Eventually exhibited surface dyslexia (PPVT 10%; HFE reading 75%)

Semantic impairment without surface dyslexia?

WLP (Schwartz et al., 1979); probable semantic dementia

- PPVT 37%; HF exception reading 98%
- Eventually exhibited surface dyslexia (PPVT 10%; HFE reading 75%)

DRN (Cipolotti & Warrington, 1995); scientist with semantic dementia

- LF exceptions: 95% reading, 29% generating definitions

DC (Lambon Ralph et al., 1995); left school at 14, Alzheimer's type dementia

- LF exceptions: 95% reading, 31% generating definitions

Semantic impairment without surface dyslexia?

WLP (Schwartz et al., 1979); probable semantic dementia

- PPVT 37%; HF exception reading 98%
- Eventually exhibited surface dyslexia (PPVT 10%; HFE reading 75%)

DRN (Cipolotti & Warrington, 1995); scientist with semantic dementia

- LF exceptions: 95% reading, 29% generating definitions

DC (Lambon Ralph et al., 1995); left school at 14, Alzheimer's type dementia

- LF exceptions: 95% reading, 31% generating definitions

EM (Blazely, Coltheart, & Casey, 2005); secretary with semantic dementia

- 98% reading; 34% picture naming

Not due simply to severity of semantic impairment

- Other equally anomic semantic dementia patients exhibit surface dyslexia
  - **GC** (Patterson et al., 1994): 45% picture naming; 38% LFE reading
  - **PC** (Blazely et al., 2005): 29% picture naming; 49% LFE reading

Semantic impairment without surface dyslexia?

WLP (Schwartz et al., 1979); probable semantic dementia

- PPVT 37%; HF exception reading 98%
- Eventually exhibited surface dyslexia (PPVT 10%; HFE reading 75%)

DRN (Cipolotti & Warrington, 1995); scientist with semantic dementia

- LF exceptions: 95% reading, 29% generating definitions

DC (Lambon Ralph et al., 1995); left school at 14, Alzheimer's type dementia

- LF exceptions: 95% reading, 31% generating definitions

EM (Blazely, Coltheart, & Casey, 2005); secretary with semantic dementia

- 98% reading; 34% picture naming

Distribution of surface dyslexia among semantic dementia patients

Woollams, Lambon Ralph, Plaut, and Patterson (2007, *Psych. Rev.*)

Patients

- 100 testing sessions of 51 semantic dementia patients
- Word reading measured on Patterson and Hodges (1992) "surface" list
- Composite semantic score derived from performance on picture naming and spoken word-picture matching

Distribution of surface dyslexia among semantic dementia patients

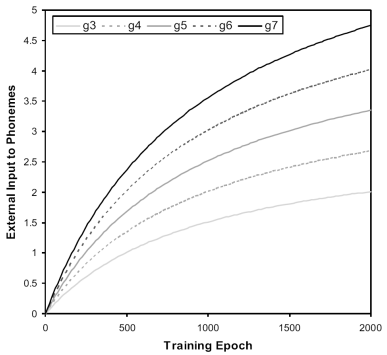
Woollams, Lambon Ralph, Plaut, and Patterson (2007, *Psych. Rev.*)

Patients

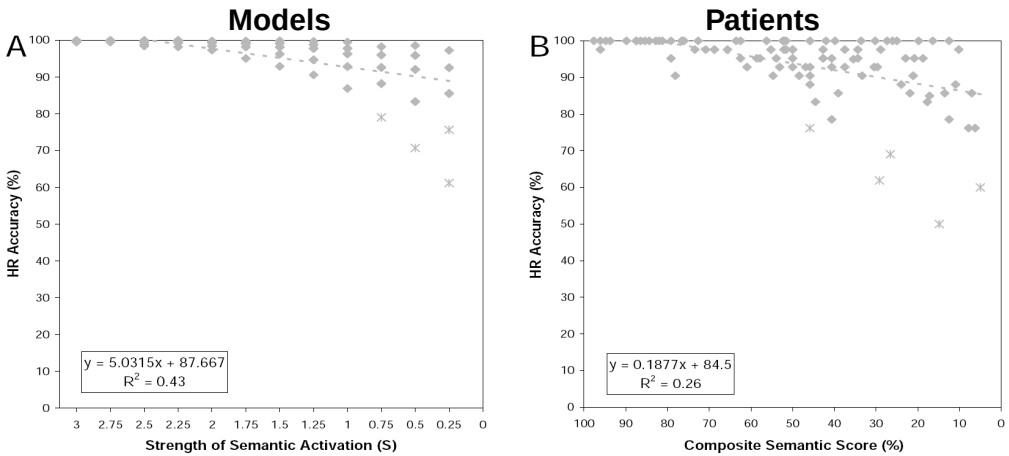
- 100 testing sessions of 51 semantic dementia patients
- Word reading measured on Patterson and Hodges (1992) “surface” list
- Composite semantic score derived from performance on picture naming and spoken word-picture matching

Models

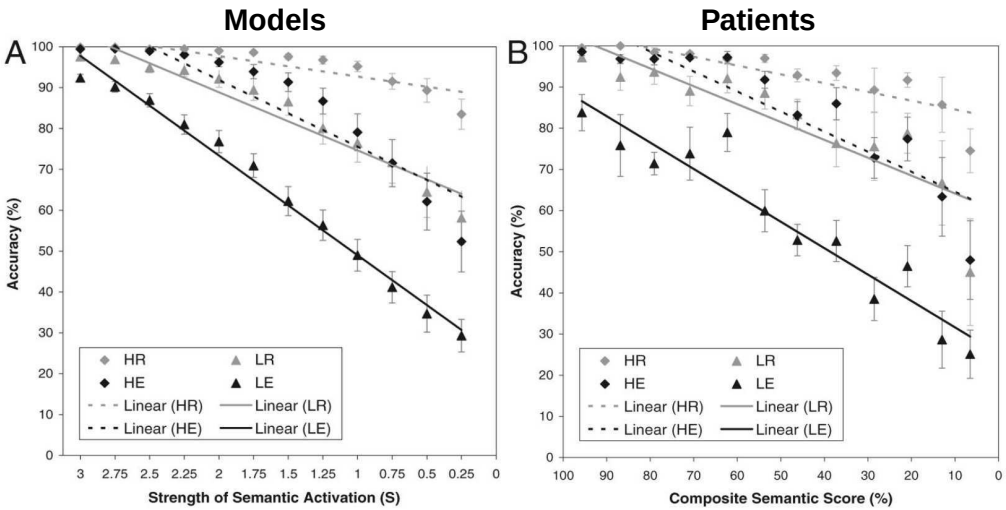
- Trained replications of Plaut (1997) simulations varying only semantic strength ($g = 3 - 7$; original PMSP simulation used $g = 5$)
- Semantic “lesions” make semantic support for phonology *weaker* and *more noisy*



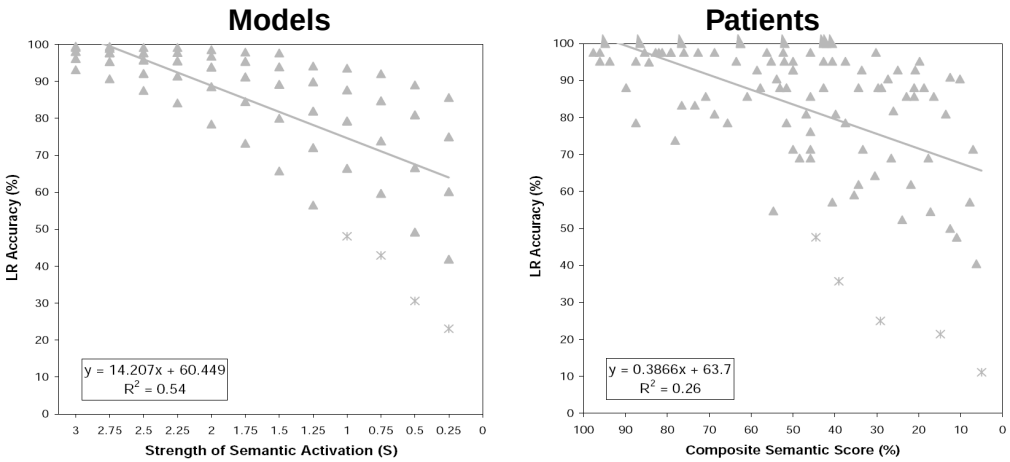
High-frequency regular words



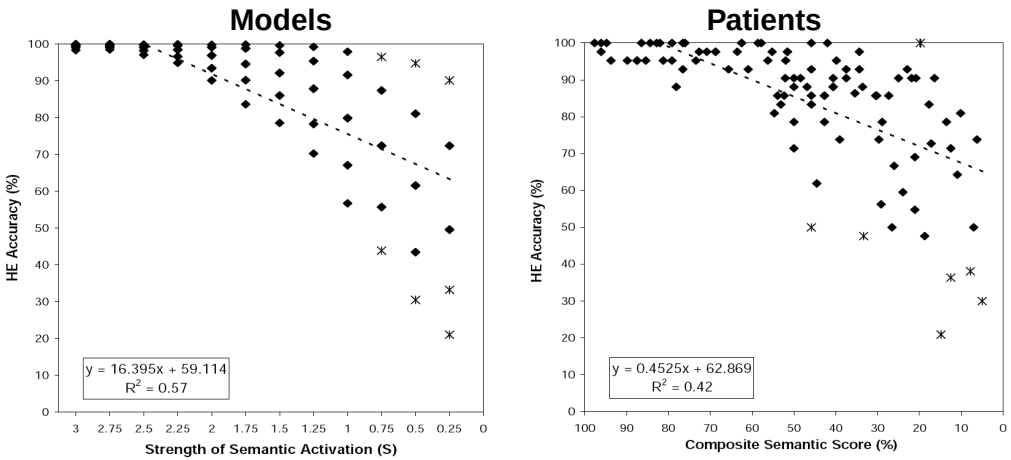
Overall accuracy



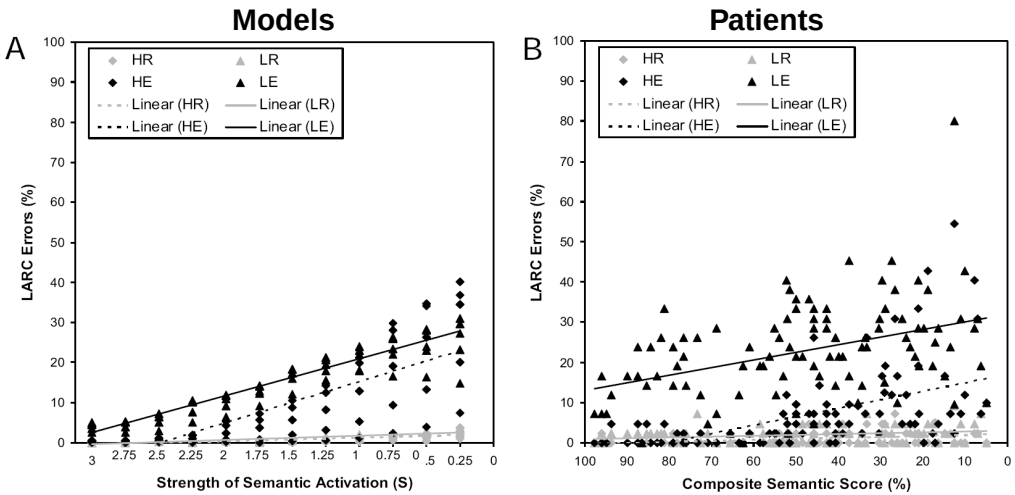
Low-frequency regular words



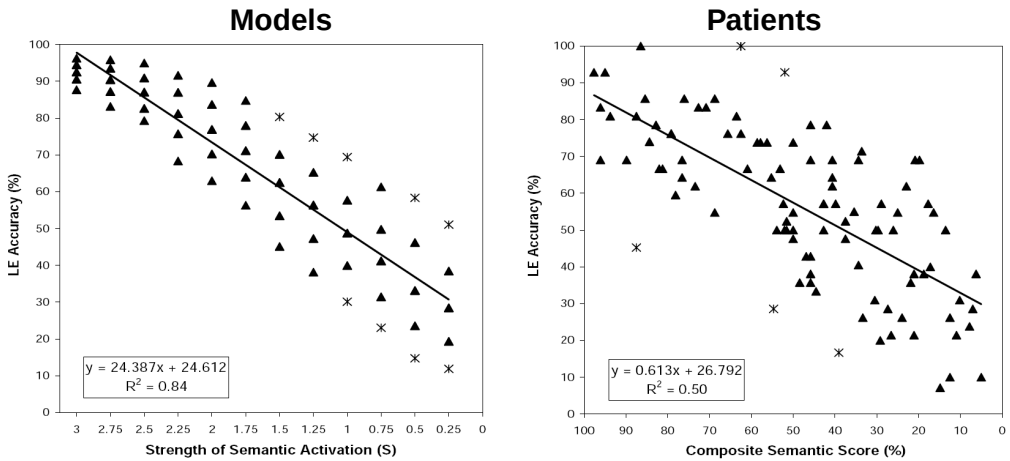
High-frequency exception words



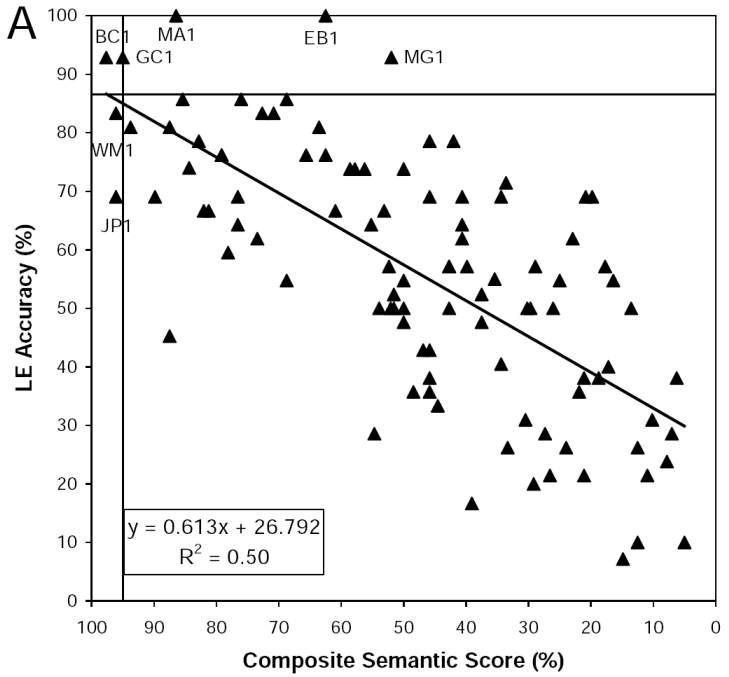
Legitimate Alternative Reading of Component (LARC) errors



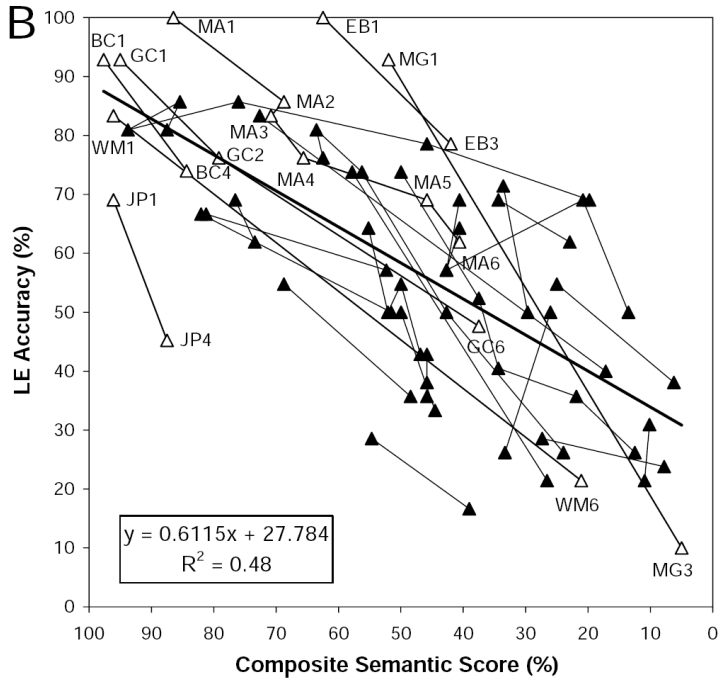
Low-frequency exception words



Outliers (Patients)



Longitudinal observations (Patients)



Concomitant deficits (Patterson et al., 2006, *JCN*)