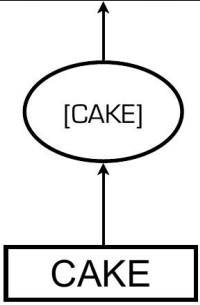
Standard view of (written) word comprehension

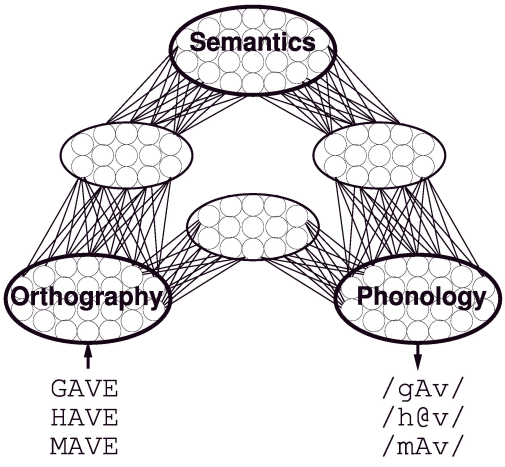


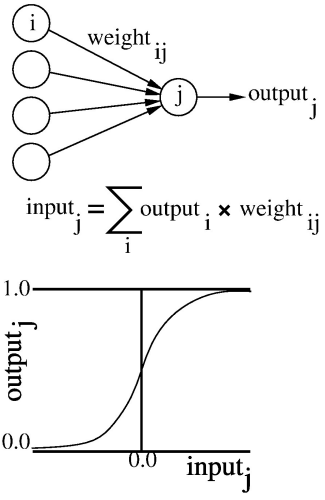
Recognize the word (as a distinct "localist" unit), then activate its meaning

- Logogen Model (Morton, 1969)
- Entry Opening Model (Forster & Davis, 1984)
- Multistage Activation Model (Borowsky & Besner, 1993)
- Dual-Route Cascaded Model (Coltheart et al., 2001)
- Connectionist Dual-Process Model(+) (Perry et al., 2007, 2010)

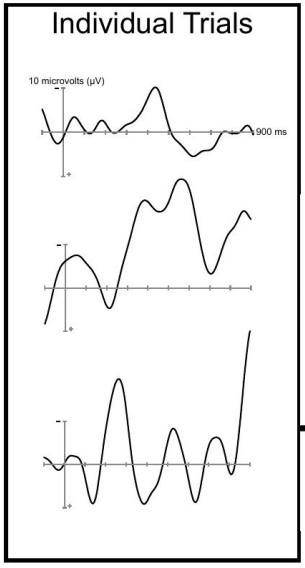
A neural-network framework for word reading

Triangle model (Seidenberg & McClelland, 1989; Plaut et al., 1996; Harm & Seidenberg, 2004)

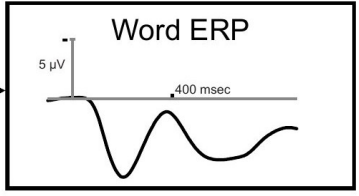




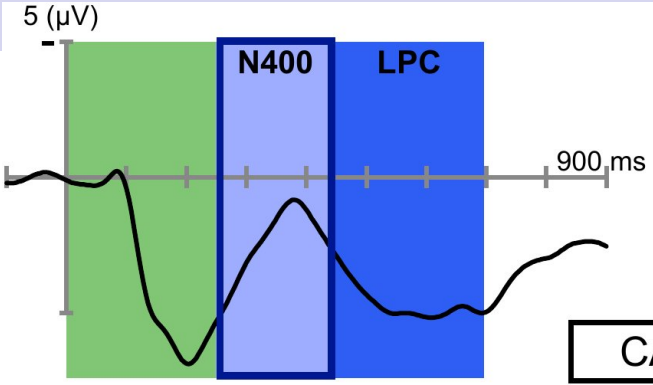
Event-Related Potentials



Signal Averaging

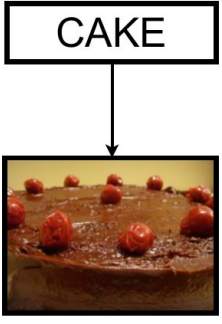


N400



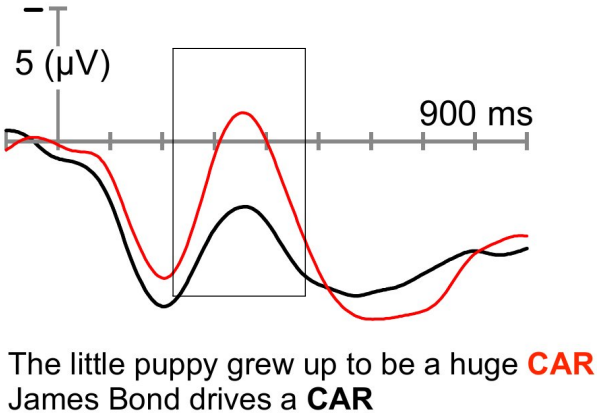
N400: Semantic Access

(Kutas & Federmeier, 2000; Lau, Phillips, & Poeppel, 2008)



Note: Negative ("N") is up

Semantic incongruity



- Larger N400 to a semantically incongruous sentence completion (Kutas & Hillyard, 1980)

N400: Summary of 40+ years of research

The N400 IS sensitive to:

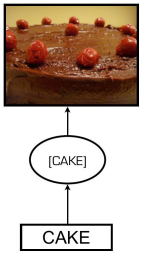
- Sentence congruity
- Discourse congruity
- Semantic priming
- Lexical association
- Concreteness
- Ambiguity

The N400 IS NOT sensitive to:

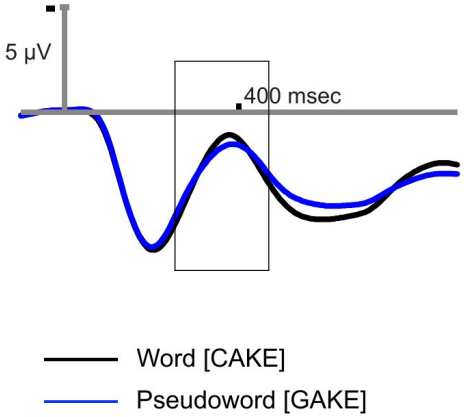
- Visual characteristics (e.g., font, case)
- Syntactic constraints/violations (e.g., number, gender, verb argument structure)

- Tends to be localized to classic "semantic" areas (middle/anterior temporal lobes)
- Effects are reduced in aphasic patients
- Not restricted to language but reflects semantic processing more generally

Lexical models and the N400



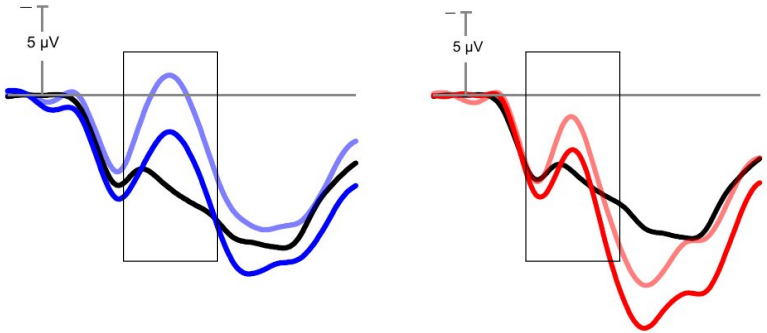
- IF an input must be recognized as a word prior to semantic access, and the N400 represents semantic access...
- THEN there **should be no N400 for pseudowords**



Not the case

(Rugg & Nagy, 1987; Deacon, Dynowska, Ritter, & Grose-Fifer, 2004; Laszlo & Federmeier, 2007; **Laszlo & Federmeier, 2011**)

Orthographic effects in sentence contexts



The genie was ready to grant his third and final:
WISH CLAM DISH TKNT WRSH

- Both words and illegal strings show effects of orthographic similarity to an expected sentence completion (Laszlo & Federmeier, 2009)

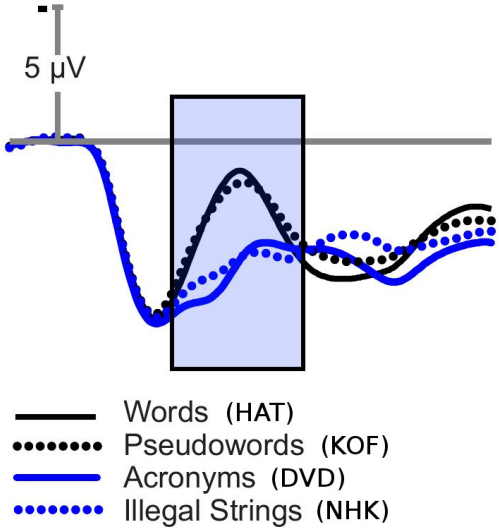
Meaningfulness vs. orthographic regularity: Acronyms

Item-level ERP data for letter strings presented in isolation (Laszlo & Federmeier, 2011):

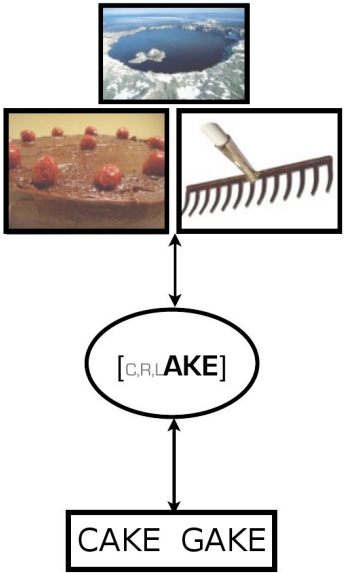
- words (HAT)
- pseudowords (KOF)
- acronyms (DVD)
- illegal strings (NHK)

  - N400 magnitudes reflect **orthographic regularity** rather than meaningfulness

  - **Problem**: How can this be reconciled with findings that the N400 reflects **semantic** access/integration?



Maybe semantic access occurs *during* rather than after recognition....

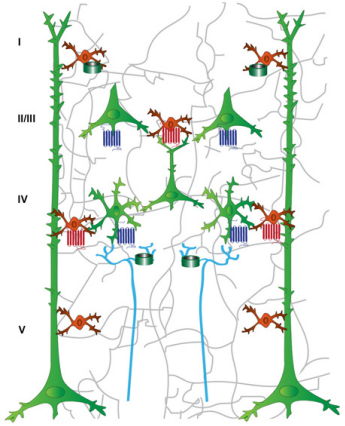


- IF an input must merely be **similar** to known words to initially engage semantics...
- THEN there should be N400s for words and pseudowords but not for illegal strings (NHK) nor for acronyms (DVD) because these have few if any "neighbors"
- Not really a "lexical" model anymore....

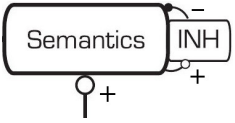
Question: Why would the comprehension system work this way?

Excitation, inhibition, and transient semantic activation

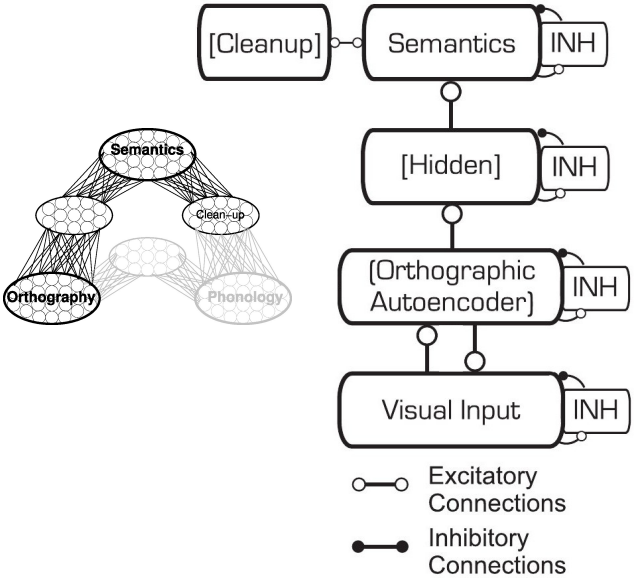
- Information in cortex is represented largely by excitatory neurons (pyramidal cells)
- Local inhibitory interneurons serve to maintain overall activation levels
- Projections between areas, including bottom-up input, are restricted to be excitatory, whereas local interactions are both excitatory and inhibitory
- Leads to cortical activation dynamics in which early bottom-up excitation is followed by suppressive within-layer inhibition (and cooperative interactions)



Proposal: N400 reflects **transient over-activation** of semantics during comprehension

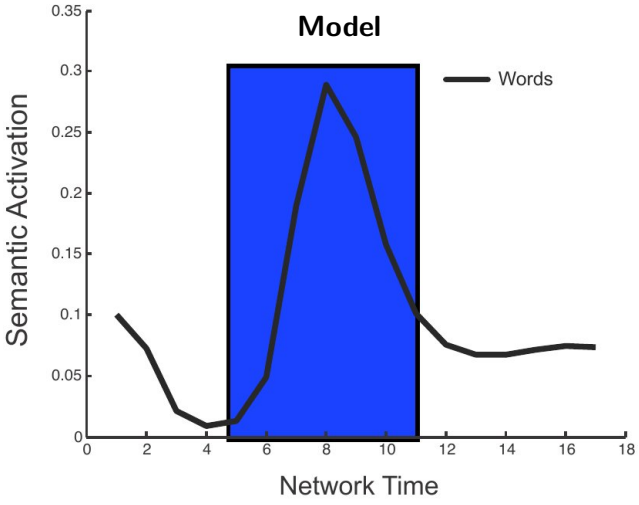


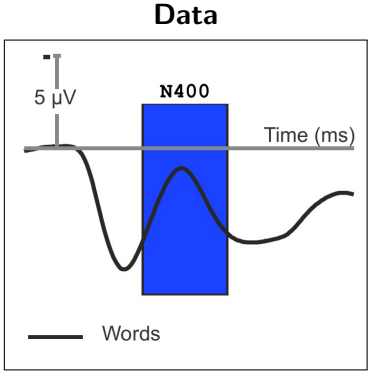
Laszlo and Plaut (2012, *Brain and Language*)

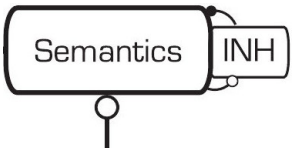


- Separate excitatory and inhibitory unit populations
- Only excitatory connections between layers
- Excitatory units represent information; inhibitory units modulate overall activation
- N400 reflected in **summed semantic activity**
- Trained to regenerate visual input and to activate meaning of 62 words (CVC) and 15 acronyms (*C*)

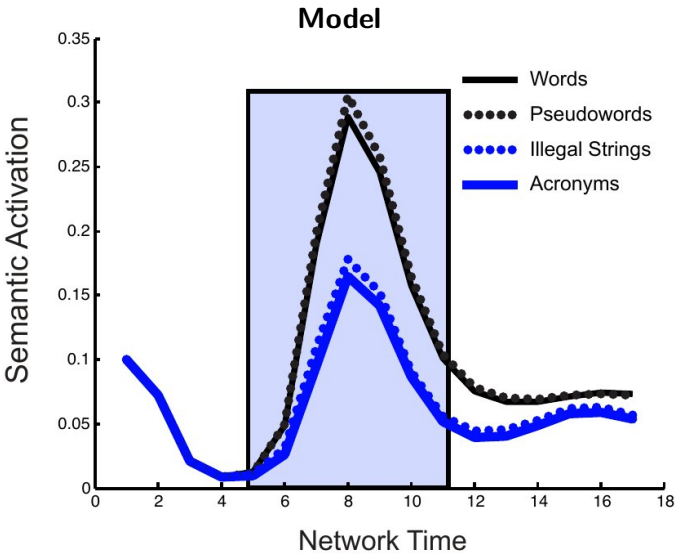
Dynamics of semantic activation in the model

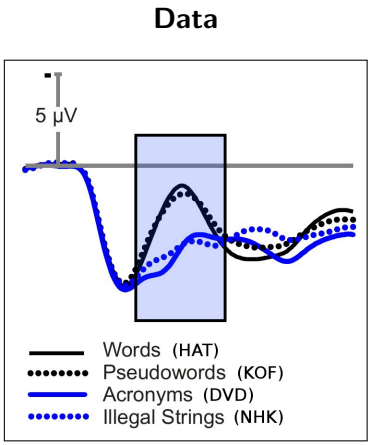




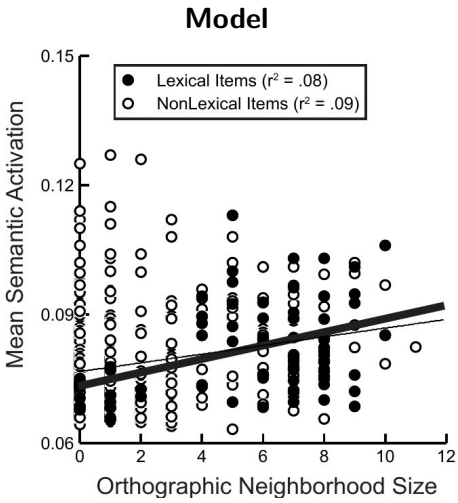


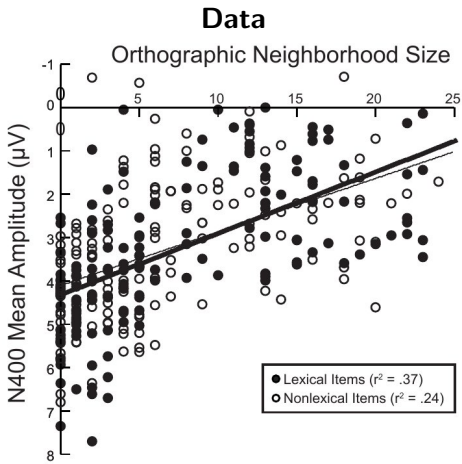
Orthographic and semantic influences on "N400"





Effect of orthographic neighborhood size

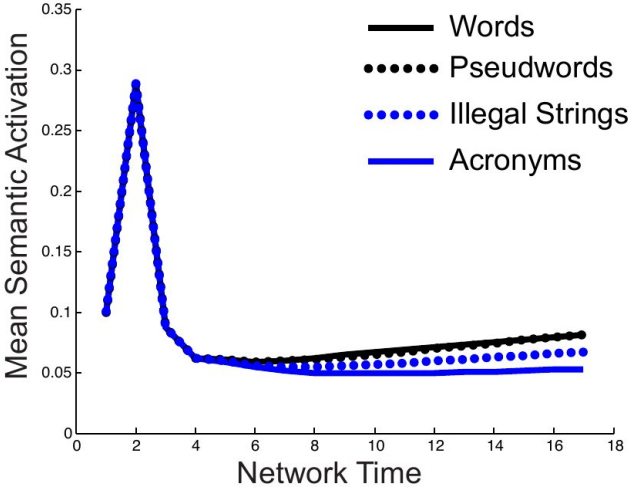




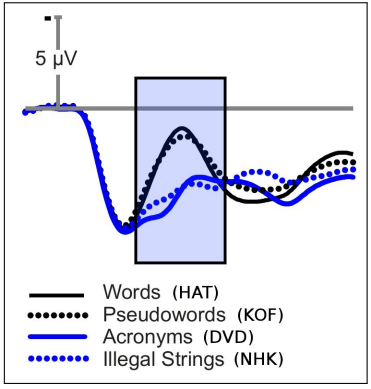
Unconstrained model

Same connectivity, no constraints on excitatory/inhibitory weight values

Model

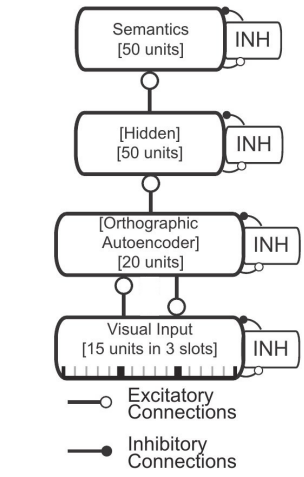


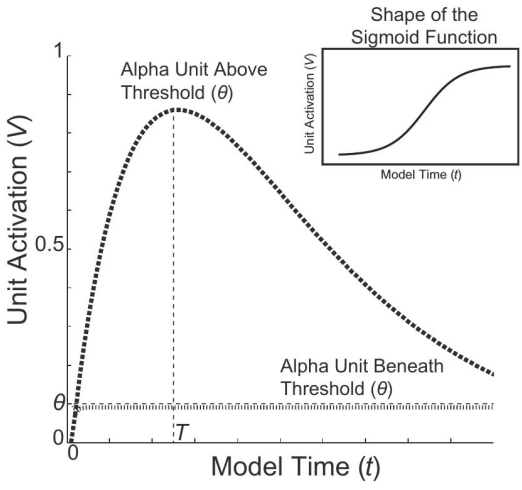
Data



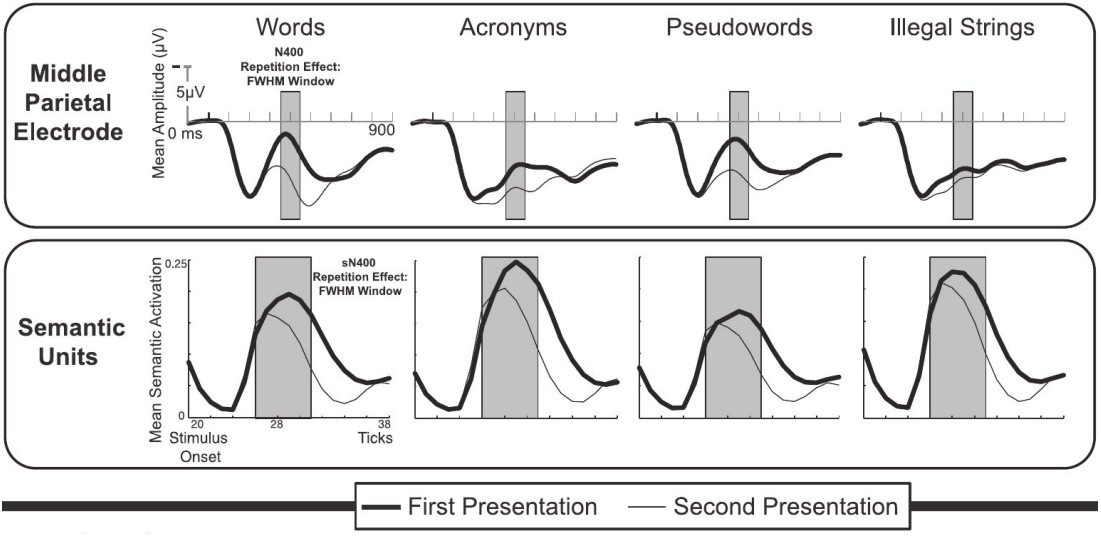
Laszlo and Armstrong (2014, *Brain and Language*)

Extended the model to address effects of **repetition** by adding activation-based **decay**





Laszlo and Armstrong (2014) results



Laszlo/Armstrong/Plaut account

- Distribution of excitation and inhibition between and within cortical layers leads to specific activation dynamics: bottom-up **over-activation** followed by reduction due to competitive (and cooperative) interactions
- The N400 ERP component can be understood as reflecting this temporary over-activation within lexical semantics
- Degree of over-activation in semantics depends on **co-activation** due to input similarity structure (e.g., orthographic neighborhood), as well as consistency with **pre-activated information** (e.g., repetition, sentence context [not implemented]).

Cheyette and Plaut (2017, *Cognition*)

Extended Laszlo/Armstrong/Plaut model to address much broader range of effects

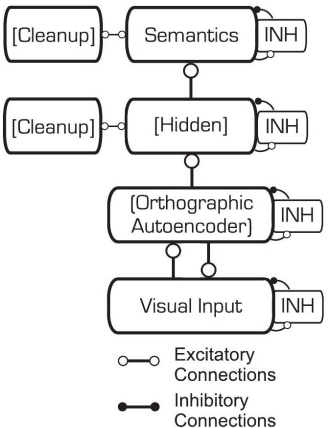
- semantic priming, richness, frequency, repetition, orthograhic neighborhood (and interactions)

	N400	Behavior
Semantic priming	related $<$ unrel.	related $>$ unrel.
Semantic richness	high $>$ low	high $>$ low
Word frequency	high $<$ low	high $>$ low
Repetition	yes $<$ no ($\Downarrow$)	yes $>$ no ($\Uparrow$)
Richness x Rep.	$\Downarrow$ high $>$ $\Downarrow$ low	$\Uparrow$ high $>$ $\Uparrow$ low
Frequency x Rep.	$\Downarrow$ high $<$ $\Downarrow$ low	$\Uparrow$ high $<$ $\Uparrow$ low
Orth. neighborhood	high $>$ low	variable

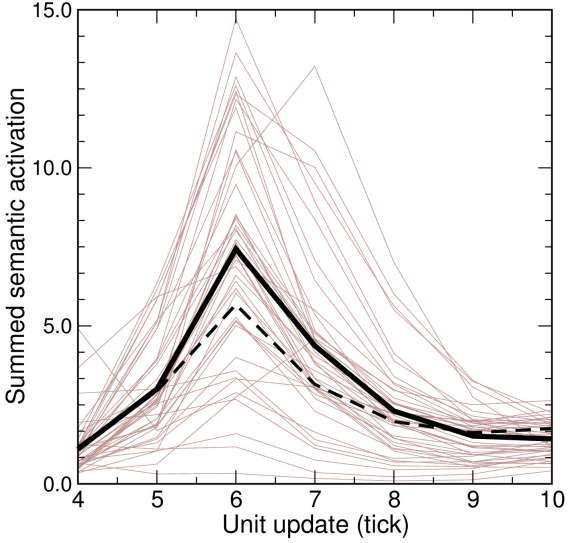
Note: Some effects are in opposite direction for N400 vs. behavior

Cheyette and Plaut (2017, *Cognition*)

- Larger vocabulary (176 words) varying in word frequency and semantic richness (number of semantic features)
- Added excitatory "clean-up" layers
- Simplified decay function (cf. Gotts & Plaut, 2002)
- Trained on pairs of words, where each word has an *associated* word that follows it much more often than chance (Plaut, 1995)
- Tested each word as *target* preceded by each word as *prime* (including itself, for *repetition*)
- N400 measure is peak amplitude of **summed semantic activity** while processing target



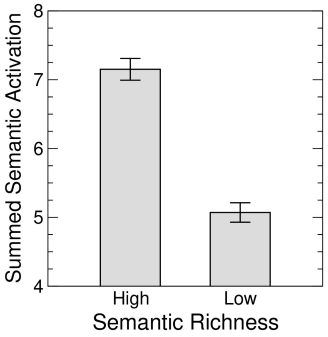
Network dynamics (individual items and means)

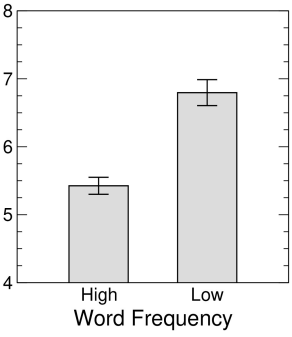


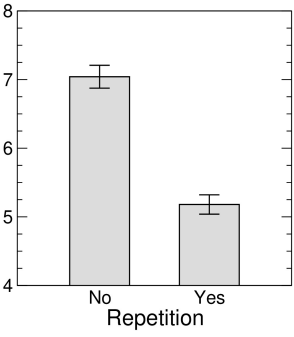
Solid = low-frequency, low-richness, repeated; Dashed = high-frequency

N400: Semantic richness, word frequency, repetition (main effects)

	N400
Semantic richness	high $>$ low
Word frequency	high $<$ low
Repetition	yes $<$ no ($\Downarrow$)

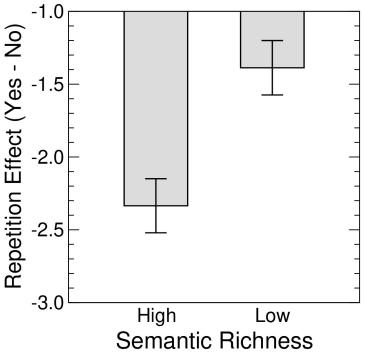


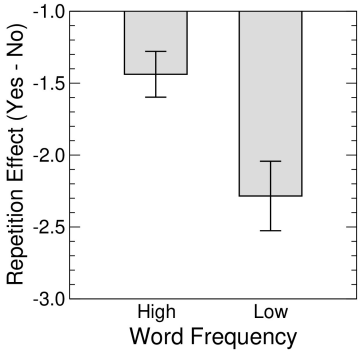




N400: Semantic richness, word frequency (interactions with repetition)

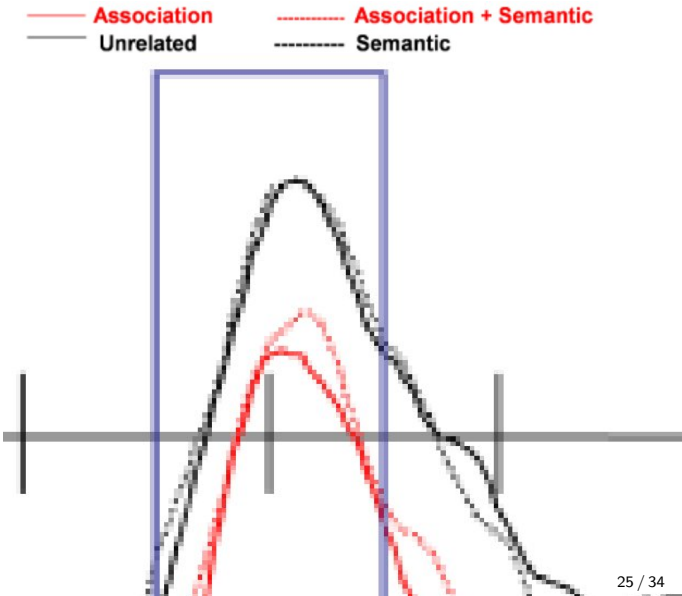
	N400
Richness x Repetition	$\Downarrow$ high $>$ $\Downarrow$ low
Frequency x Repetition	$\Downarrow$ high $<$ $\Downarrow$ low





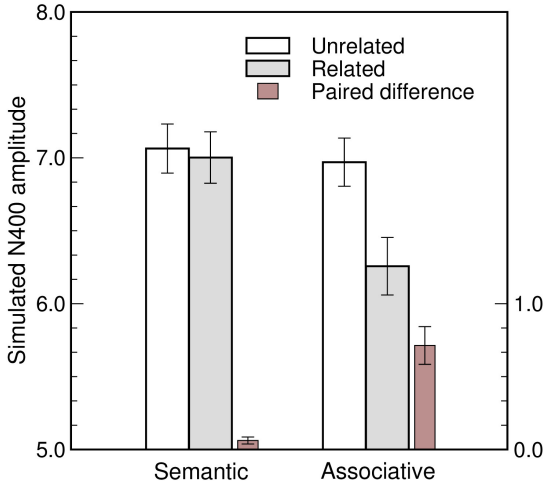
Distinguishing semantic and associative priming on N400

- "Semantic" relatedness (shared features) is often conflated with associative relatedness in empirical studies.
- Rhodes and Donaldson (2014, *Psychophys.*) found an effect of associative relatedness but **no effect of pure semantic relatedness** on N400 amplitude (see also Koivisto & Revonsuo, 2001, *Cog. Br. Res.*)



N400: Semantic and associative priming

	N400
Semantic priming	related $<$ unrel. (???)



N400 and behavioral effects

	N400	Behavior
Semantic priming	related $<$ unrel.	related $>$ unrel.
Semantic richness	high $>$ low	high $>$ low
Word frequency	high $<$ low	high $>$ low
Repetition	yes $<$ no ($\Downarrow$)	yes $>$ no ($\Uparrow$)
Richness x Rep.	$\Downarrow$ high $>$ $\Downarrow$ low	$\Uparrow$ high $>$ $\Uparrow$ low
Frequency x Rep.	$\Downarrow$ high $<$ $\Downarrow$ low	$\Uparrow$ high $<$ $\Uparrow$ low
Orth. neighborhood	high $>$ low	variable

- How should behavior (in lexical decision) be modeled?
  - Problem: Repetition suppresses neural activity but facilitates behavior

Repetition suppression and neural synchrony

Possible mechanisms for behavioral facilitation under repetition suppression (Gotts, Chow & Martin, 2012, *Cog. Neuropsych.*)

- **Representational "sharpening":** Differential reduction of poorly tuned neural responses
  - But careful measurements of short-term repetition suppression indicate proportional scaling rather than sharpening (McMahon & Olson, 2007; Weiner et al., 2010)
  - fMRI-adaptation study found broadening rather than sharpening (Gotts et al., 2014)
- **Neural synchrony:** Increase in spike synchrony under suppression amplifies downstream impact
  - General support from single-cell recordings (Brunet et al., 2014), MEG (Ghuman et al., 2008) and intracranial EEG (Engell & McCarthy, 2014)

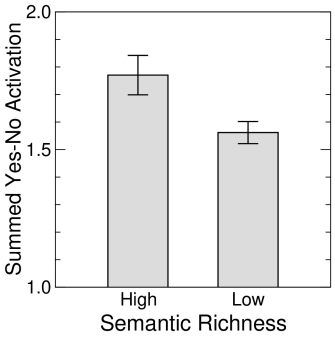
Current approach: Cannot use neural synchrony directly but can make its underlying cause (degree of suppression) available to a **response system**

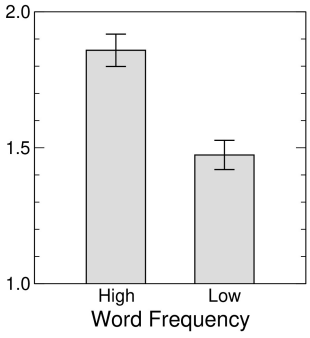
Trained response system

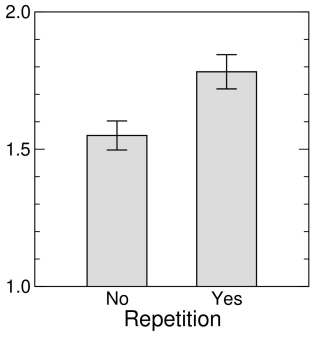
- Same computational/architectural principles as in comprehension network (separate excitation/inhibition; activation-based decay)
- Created 176 *pseudowords* with same orthographic structure as words (CVC) but not in comprehension training set
- Inputs: both *instantaneous* semantic activations and *integrated* semantic activations (which determine activation-based decay)
- 25 hidden units (with 5 inhibitory units, 10 "clean-up" units); connections from integrated semantics are unconstrained
- Outputs: "Yes" and "No" response units (1 inhibitory unit)
- Trained on pairs of words and pseudowords (fully crossed) with Yes/No targets only for second stimulus (target)
- **Behavioral measure:** sum of Yes−No activation during target

Behavior: Semantic richness, word frequency, repetition (main effects)

	Behavior
Semantic richness	high > low
Word frequency	high > low
Repetition	yes > no ($\Uparrow$)

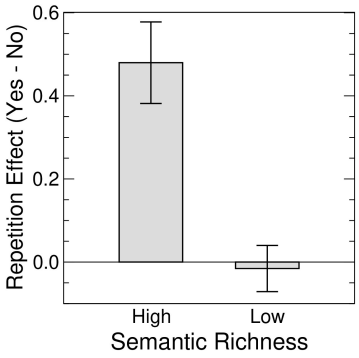


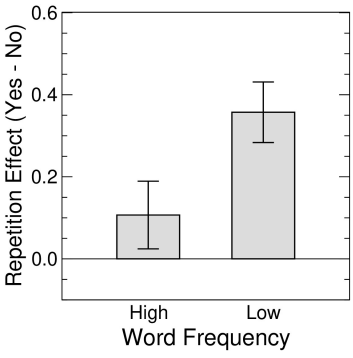




Behavior: Semantic richness, word frequency (interactions with repetition)

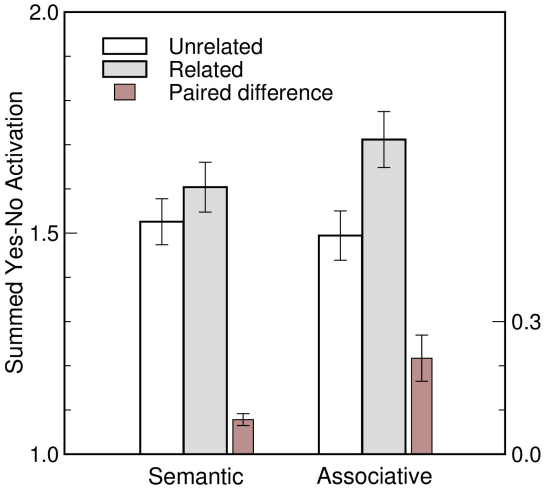
	Behavior
Richness x Repetition	⇑ high > ⇑ low
Frequency x Repetition	⇑ high < ⇑ low





Behavior: Semantic and associative priming

	Behavior
Semantic priming	related > unrel.



N400 and behavioral effects

	✔ N400	✔ Behavior
Semantic priming	related $<$ unrel.	related $>$ unrel.
Semantic richness	high $>$ low	high $>$ low
Word frequency	high $<$ low	high $>$ low
Repetition	yes $<$ no ($\Downarrow$)	yes $>$ no ($\Uparrow$)
Richness x Rep.	$\Downarrow$ high $>$ $\Downarrow$ low	$\Uparrow$ high $>$ $\Uparrow$ low
Frequency x Rep.	$\Downarrow$ high $<$ $\Downarrow$ low	$\Uparrow$ high $<$ $\Uparrow$ low
Orth. neighborhood	high $>$ low	variable

Conclusions

- The N400 can be understood as reflecting transient over-activation of semantics (reflecting pre-activated and co-activated features) due to distribution of excitation and inhibition in cortex
- A distributed connectionist network exhibits analogous dynamics, and is influenced by lexical and context variables in similar ways, but only when it separates excitation and inhibition appropriately
- The dynamics of word comprehension differ fundamentally from that in the standard "lexical" theory in which recognition precedes semantic access
- Behavioral performance can be modeled by a trained response system that is sensitive to degree of repetition suppression (perhaps approximating neural synchrony)