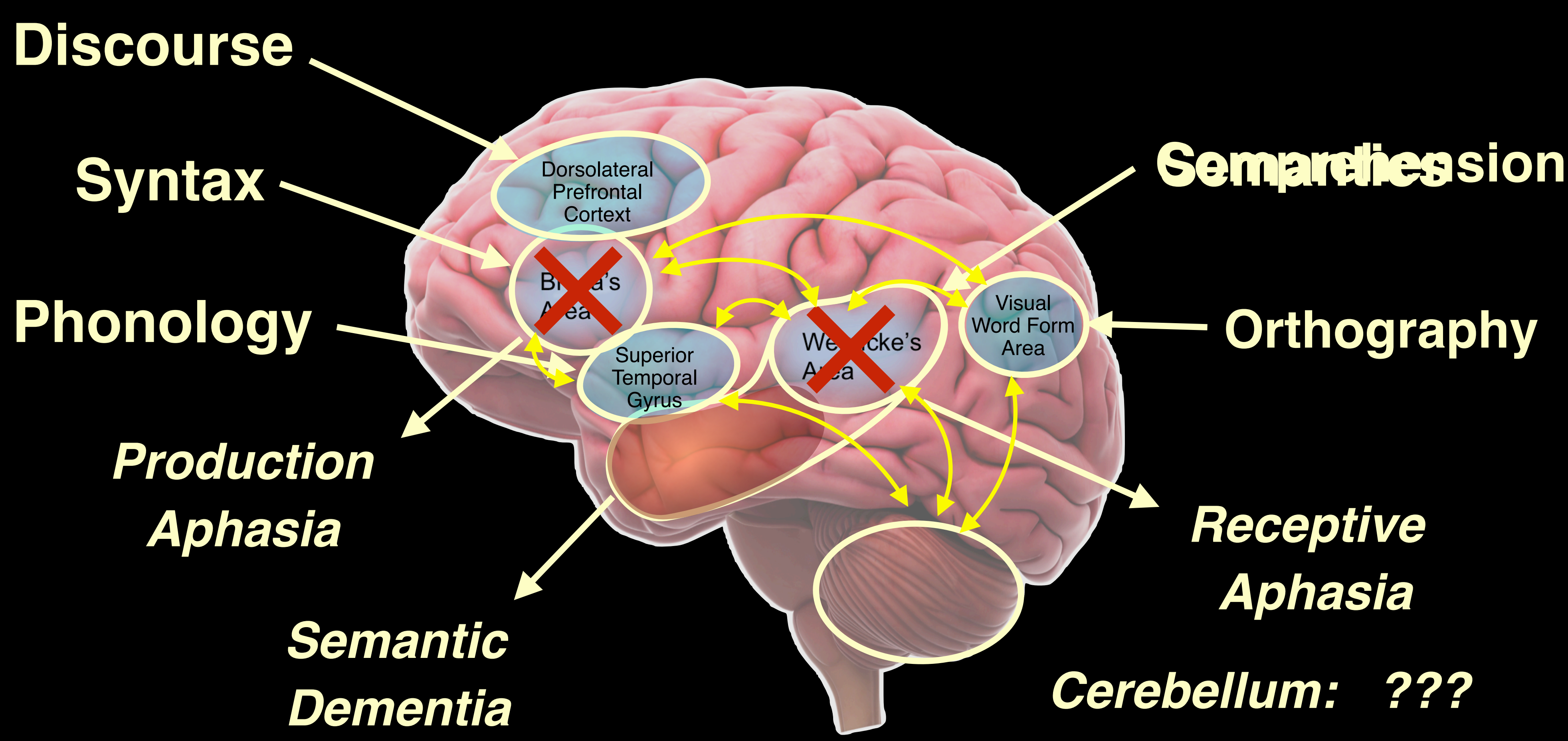
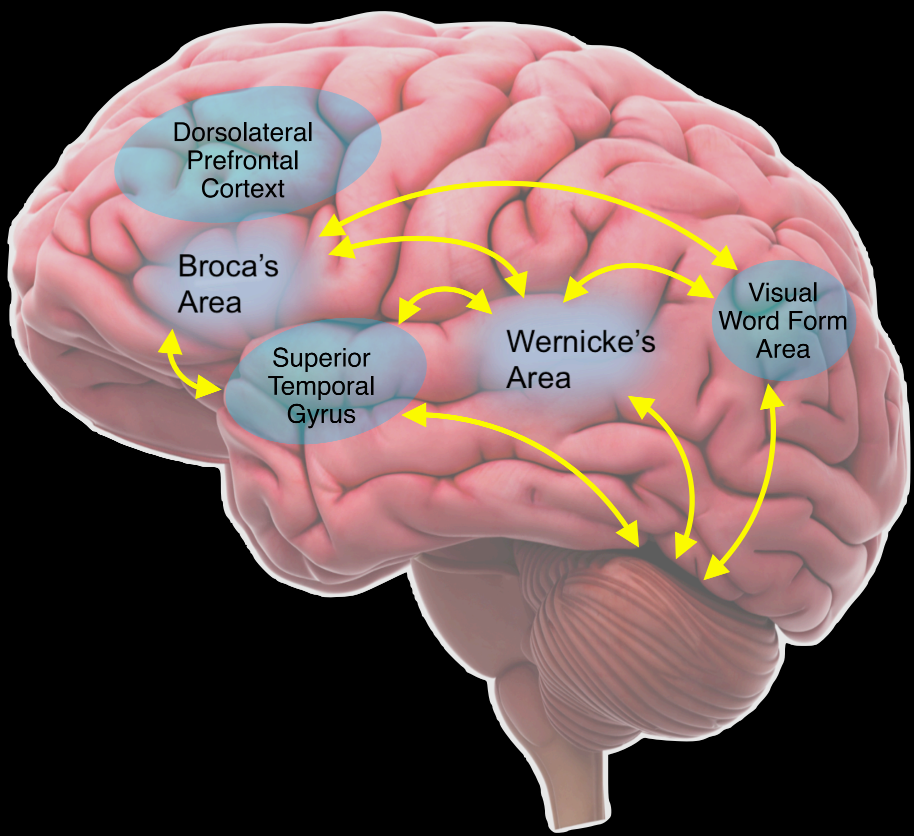
Language Processing

Language in the Brain



Language and the Brain



Interactive Activation Model

- First connectionist model…

Interactive Activation Model

McClelland & Rumelhart, 1981

- **Word superiority effect** *(Reicher, 1969)***:**
  - **Faster to recognize a letter in the context of a word than in a pseudoword or alone (even controlling for frequency and "legality"):**

    WORK ⇒ forced choice: K or R?

    faster than

    SORK ⇒ forced choice: K or R?

    faster than

    K ⇒ forced choice: K or R?

  - **Works for pronounceable pseudowords** *(e.g., recognizing A in MAVE)***:**

    ⇒ *encoding of orthographic "rules"?*

    No! *Constraint satisfaction through interactive activation*

Interactive Activation Model of Letter Perception

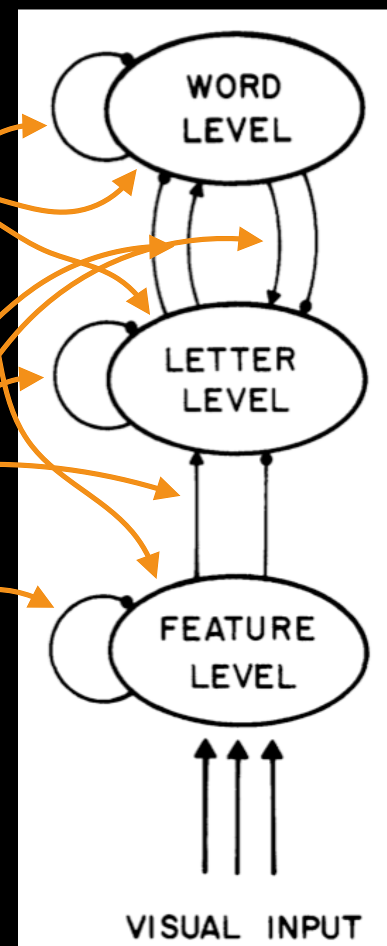
McClelland & Rumelhart, 1981

- **Assumptions:**
  - **multilayered / hierarchical sets of representations**
  - **parallel, *"interactive"* processing at all levels (simultaneous top down + bottom up influences):**

    bottom up:
    features (letters) excite objects (words) with which they are consistent

    top down:
    objects (words) excite their features (letters)

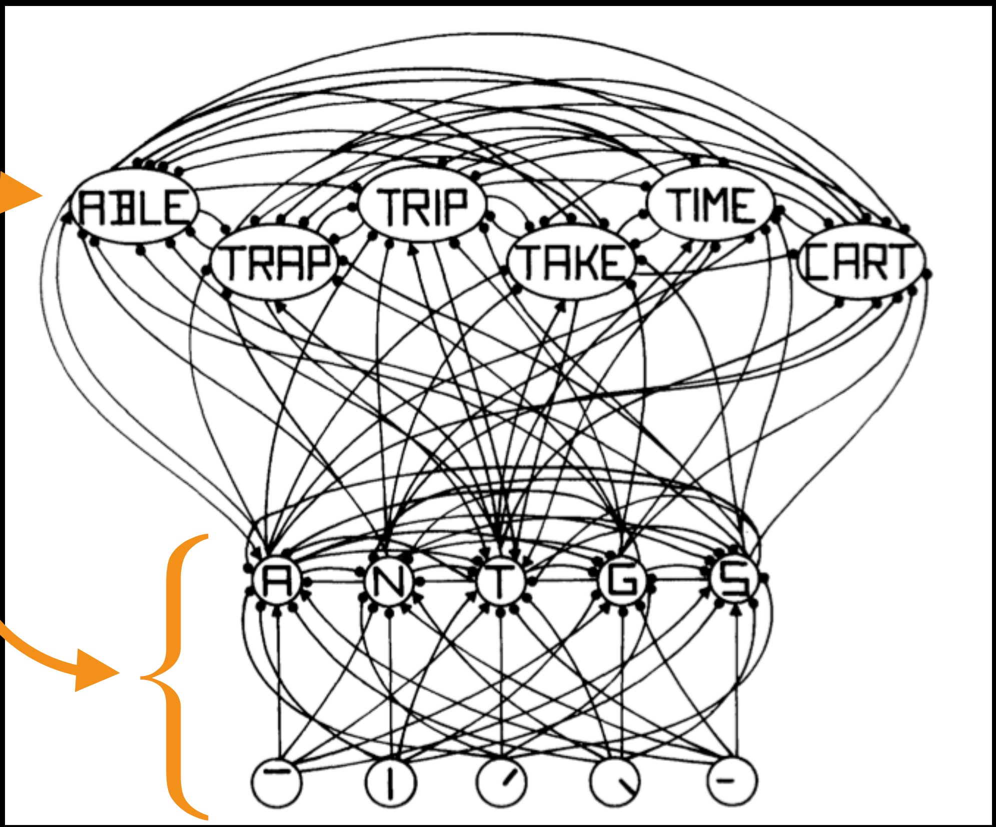
    competition:
    alternatives within a layer inhibit one another
  - **letter perception: constraint satisfaction through settling process over the network**



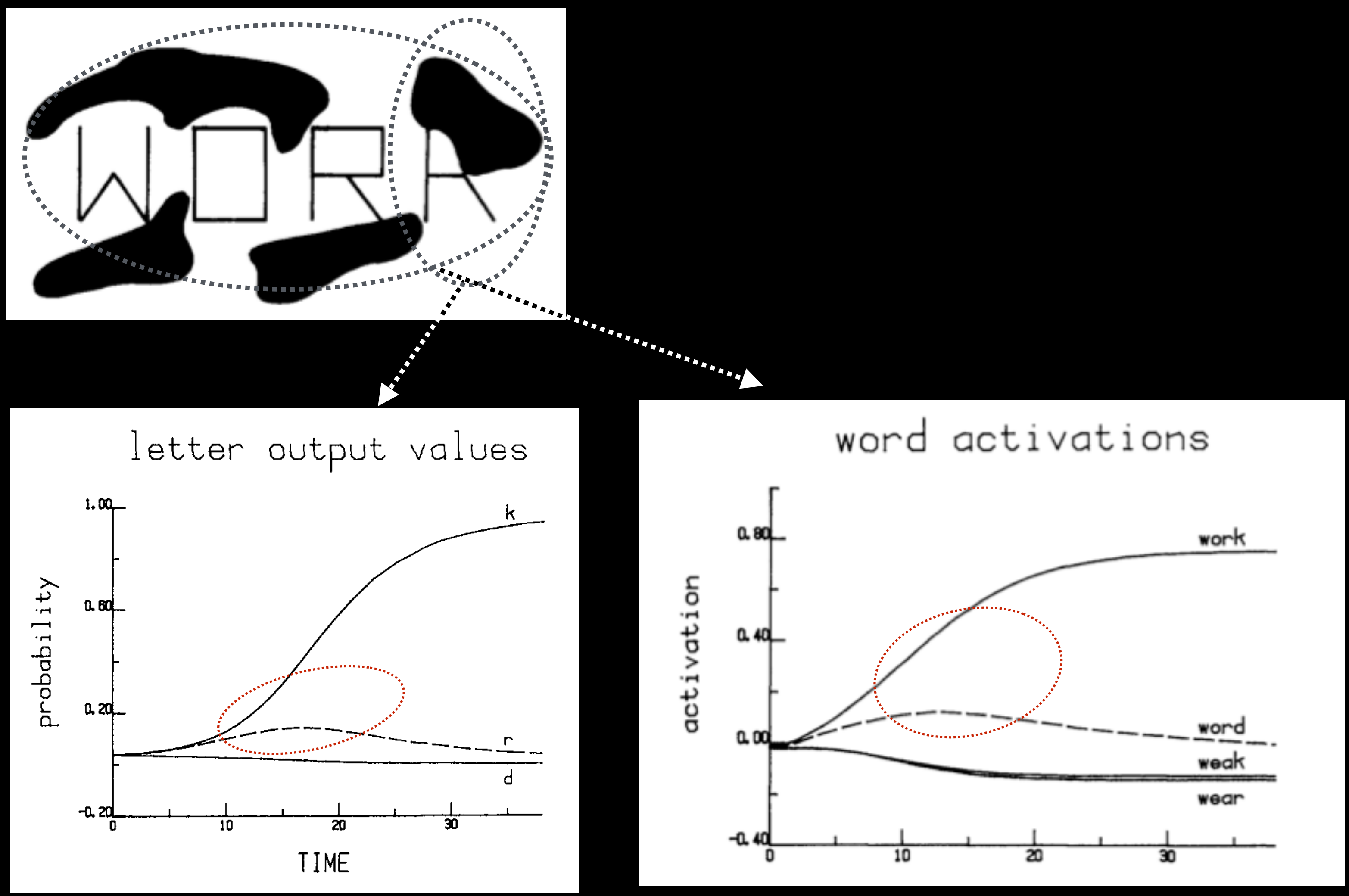
Interactive Activation Model of Letter

McClelland & Rumelhart, 1981

- **Model:**
  - **4 sets of position-specific feature and letter units**
  - **Units for all 4 letter words** *(above freq of 4 in Kucera &Francis, 1967; cf Brysbaert & New, 2009)*
  - **Base activities (biases) monotonically related to frequency of occurrence**



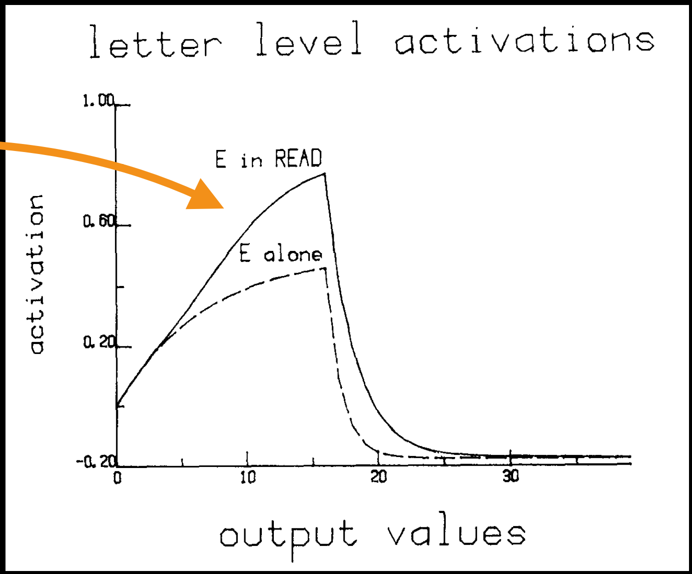
Dynamics



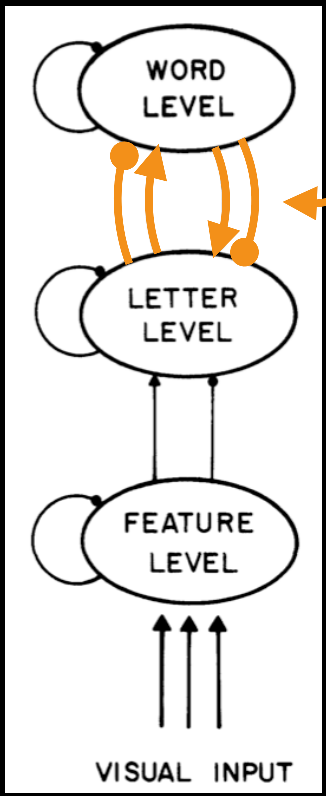
letter output values
probability
TIME
k
r
d
word activations
activation
work
word
weak
wear

Word Superiority Effect

- Letter in *context* of word get *top-down support* from word unit



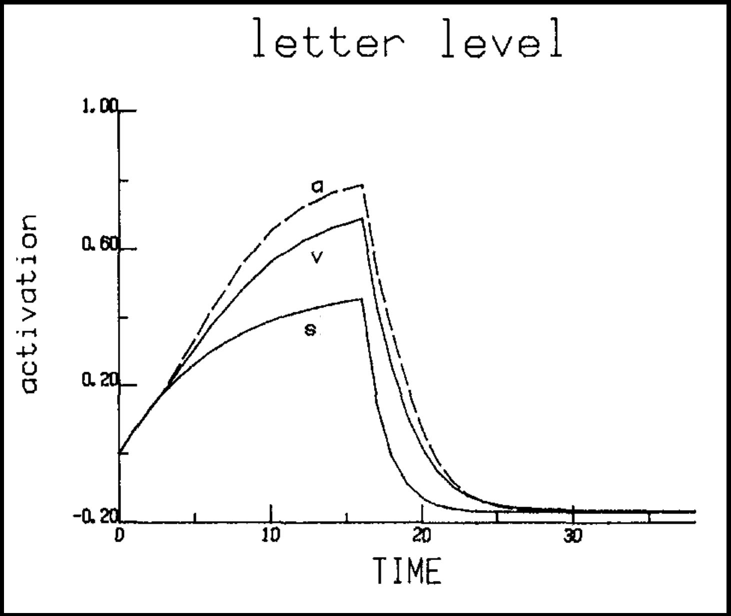
INTERACTIVE ACTIVATION



Pseudoword Superiority Effect

- Letters in MAVE benefit by top down effect of ***partial matches*** to real words (SAVE, HAVE, HATE, etc.):

NEIGHBORHOOD EFFECT
(words that have 3 letters in common)



Critical Dynamic Effects

- **Rich get richer** *("early bird gets the worm"):*
  - **Words that have higher *initial bias* (frequency) get *more active* by *suppressing competitors* early on**

HYSTERESIS

- **Gang effect:**
  - **Freq (initial bias): move = male > save**
  - **But move has *fewer neighbors*, so save ≈ male > move**

QUASI-REGULARITIES

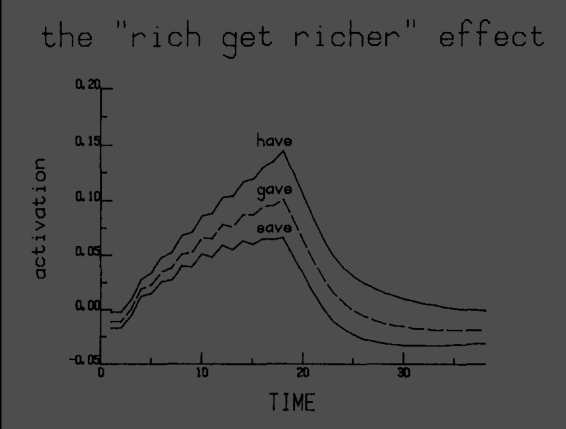
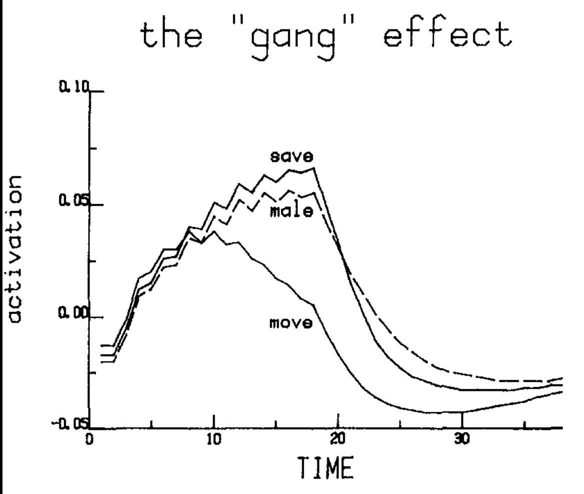


Figure 11. The rich-get-richer effect. (Activation functions for the nodes for *have*, *gave*, and *save* under presentation of *MAVE*.)



Summary

- Accounts for empirical findings regarding letter perception:
  - word superiority effects
  - orthographically legal pseudoword superiority effect (MAVE)
  - frequency effects
  - masking and stimulus quality effects
  - timing of context
  - Made new predictions
    (e.g., superiority effect should extend to orthographically *illegal* strings that are neighbors of orthographically legal words *(such as SPCT given SPAT, SPIT and SPOT)*
- Exhibits principles of constraint satisfaction:
  - *Effects of quasi-regularities* (neighborhood effects): rule-like behavior without explicit representation of rules
  - *Dynamics* (e.g., hysteresis)

Actual and Simulated Results (Probability Correct) From Johnston (1978) Experiments

	Constraint	
Result class	High	Low
Actual data		
Forced choice	.77	.79
Free report	.54	.54
Simulation		
Forced choice	.77	.76
Free report	.56	.54

Actual and Simulated Results of the McClelland & Johnston (1977) Experiments (Proportion of Correct Forced Choice)

	Target type		
Result class	Word	Pseudoword	Single letter
Actual data			
High BF	.81	.79	.67
Low BF	.78	.77	.64
Average	.80	.78	.66
Simulation			
High BF	.81	.79	.67
Low BF	.79	.77	.67
Average	.80	.78	.67

Note. BF = bigram frequency.

Levels of Analysis

- **Classical:**
  - **Phonology / Orthography**
    sounds/letters
  - **Morphology**
    meaningful subcomponents of words
  - **Lexical**
    meanings of words
  - **Syntax**
    word order

- **But also:**
  - **Prosody**
    tone of voice, stress, cadence
  - **Affect**
    emotional meaning
  - **Discourse**
    topics

Rules and Exceptions

- **Traditional view of language**
  - **knowledge of:**
    rules
    exceptions
  - **spelling** *(orthography):*
    "i" before "e" (achieve, die, relieve, alien…)
    except after "c" (receive, conceive, ceiling…)
    except except "re-", "eign" (reinforce, reify, reign, deign…)
  - **pronunciation** *(phonology):*
    "k" is spoken (kite, kick, klutz…),
    except before "n" (knife, knee, knave…)
    except except knish, knuddel…
  - **conjugation:**
    past = present + ed (train-trained, love-loved, jump-jumped)
    except for __ee__ (sleep-slept, creep-crept); or __ow (grow-grew, blow-blew)
    except except steep-steeped, or row-rowed… or speak-spoke, swim-swam, see-saw, go-went…

Rules and Exceptions

- OK, so there are lots of exceptions

- **But, without rules, how could we explain:**
  (Fodor & Pylyshyn, 1988)

  - generativity:

    we can say meaningful things we've never said before,
    that someone else can understand who has never heard them before

  - compositionality:

    we seem to be able to recombine parts (words and sentences)
    in new ways that are nevertheless meaningful / understandable

  - but how could we do these things if there was not some systematicity
    to how we do the combining and recombining…

  - that is, if there were no rules

Classical Linguistics

- Generativity of language must be rule-based:
  - "Colorless green ideas sleep furiously" *vs.*
  - the same rule can be used to add 2+2, 3079+63 or any other pair of numbers

– But following the rules doesn't seem to be necessary or sufficient

  - you know what this means even though it violates the rules:
    - *"Me cookies like!"*

  - you don't really know what this means, even though you know it obeys the rules:
    - *"Colorless green ideas sleep furiously"*

- Competence vs. performance…

Competence vs. Performance

- **Competence:**
  - **knowledge of the rules of language**
  - **focus of traditional *linguistics***
  - **must be studied by "grammaticality judgements"**
    - *can't be studied directly, since it involves implicit knowledge — the rules are "latent variables"*
    - *can't be studied by examining performance directly, since performance is not perfect*

- **Performance:**
  - **use of language** *(Wittgenstein: "meaning is use")*
  - **traditional linguistics view** *(Chomsky: poor measure of <u>competence</u>)***:**
    - *noisy execution of competence, and therefore really just a reflection of <u>in</u>competence*
  - **the focus of psycholinguistics and statistical / computational linguistics:**
    - *dysfluencies / deviations from rules ⇒ statistical features and processing constraints*
      ⇏ impaired competence
  - **we'll come back to that later; for the moment, more about the traditional view…**

Phrase Structure Grammar

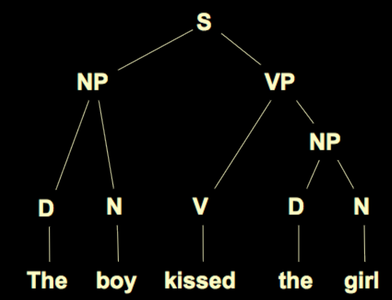
- **Grammar vs. syntax:**
  - – grammar = rules
  - – syntax = rules for word order

- **Simple sequential model** *(left-right grammar)***:**
  - – production:
    - *construction of utterances based on rules for combining elements (constituents) — "rewrite rules"*
    - *example:*

"the boy kissed the girl" ⟶

S → NP + VP
NP → D (+A) + N
VP → V + NP
N → boy, girl
V → kissed
D → the

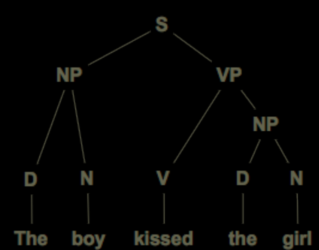
⟶ **Parse tree:**

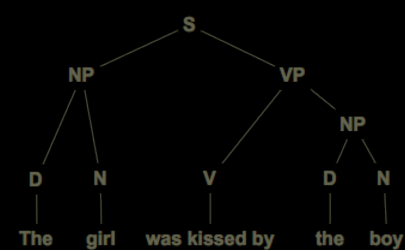


  - **– comprehension:**
    - – *parsing (decomposing) utterances using knowledge of the rules of composition*

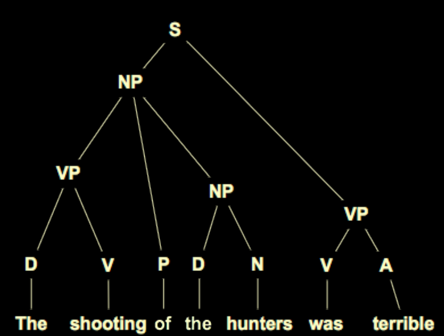
Limitations to Phrase Grammar

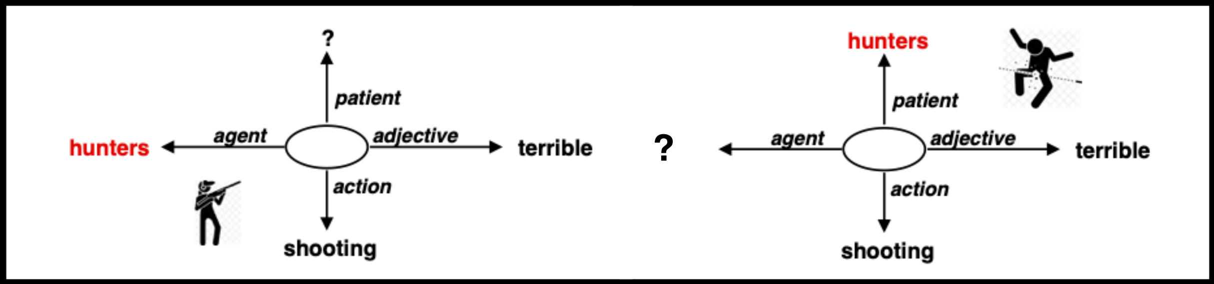
- **This can't be the basis for *semantic* *(or "propositional")* knowledge:**
  - **two sentences that have *different parses* may have the *same meaning*:** *(i.e., represent the same proposition)*





  - **a sentence with only a single parse may be ambiguous:** *(i.e., represent two different propositions)*





Transformational Grammar

- This was Chomsky's big insight:
  - *phrase* grammar governs the *deep structure* of language:

    *the **meaning** of an utterance*

  - *transformational* grammar governs the *surface structure* of language:

    the *form* of an utterance

  - *Transformation rules* translate deep structure ⟶ surface structure and back
- Examples…

Example 1: Phrase Deletion

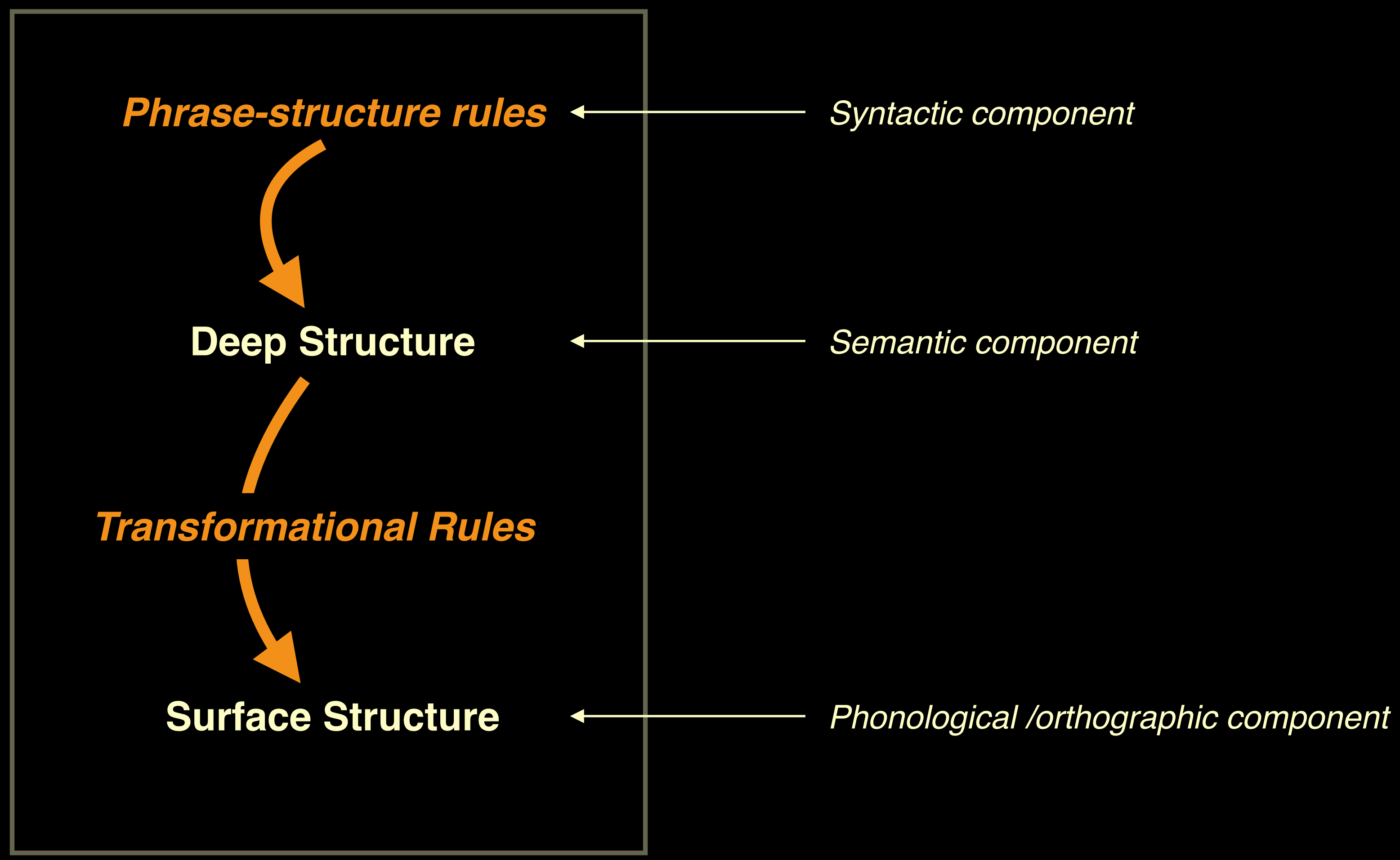
***Surface structure** (utterance):*
***Deep structure** (meaning):*

the shooting of the hunters [by the lunatic] was terrible

the shooting of the hunters [at the deer] was terrible

Transformation rule: **"elision" (deletion)**

Schematic of Chomsky’s Theory



Claims and Problems with Transformational Grammar

- Knowledge of transformational grammar rules is innate and universal
  - we are born with full knowledge of language and just have to discover it
    - very Platonic
- Language acquisition
  - process of hypothesis generation & testing → discovery of which rules are in use
    - do kids really do this?
- *Syntax* is central and pre-ordinate
  - first you ... what to say
    - seems b...
- Transformational grammar is a theory of competence, not performance
  - but many of the defining characteristics of human psychological (neurobiological) function appear to be related to performance, not competence
    - wo... memory ...
- Virtually all of the early work in computational linguistics was spent building parsers based on transformation grammar
  - should be easy if it is based on universal rules, right?

Phrase-structure rules ← *Syntactic component*

Deep Structure ← *Semantic component*

Transformational Rules

Surface Structure ← *Phonological component*

Parallel Distributed Processing

- Language is like everything else we have studied...
- Language is just another set of input-output mappings (e.g., orthography to phonology, orthography to semantics)
- These mapping result from the same simple learning algorithms that extract (complex) statistical regularities from (massive amounts of) experience (i.e., *data*)