1

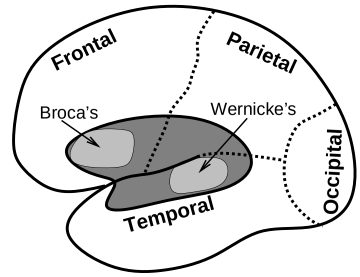
Language

How can we possibly simulate language abilities in neural terms?...

- (We can't... But many aspects of language are not "special") :
- Just another set of input/output paths.
- Levels: phonemes/letters, words, phrases, sentences, paragraphs, and beyond..
- Huge combinatorial power: distributed reps over time!

2

Biological Substrates of Language



- Broca's = speech output, syntax, grammar (surface production):
active maintenance of context to perform syntactic processing

- Wernicke's = semantic comprehension + output (deep):
interconnected overlapping distributed info about semantics

3

Traditional view of language

- Language competence defined by knowledge of **rules and exceptions** (eg. *i before e except after c*)
- Knowledge about words is stored in a central mental **lexicon** (dictionary)
- Each word has a lexical representation that is linked to information about its orthography, phonology, semantics

4

Neural net / Connectionist View of Language

- Language is another set of input-output mappings (eg orthography to phonology, orthography to semantics)
- These mappings are trained up using the same learning algorithms used elsewhere (e.g., vision)
- The same pathways handle both rules and exceptions
- Hard to tell what is "regular" vs "exceptional"
  - regular: *clown, down*
  - exception: blown.. *but blown goes with grown*
- *Distributed lexicon*: Knowledge about words is embodied in reciprocal mappings between phonology, orthography, semantics – **there is no central "word representation"**

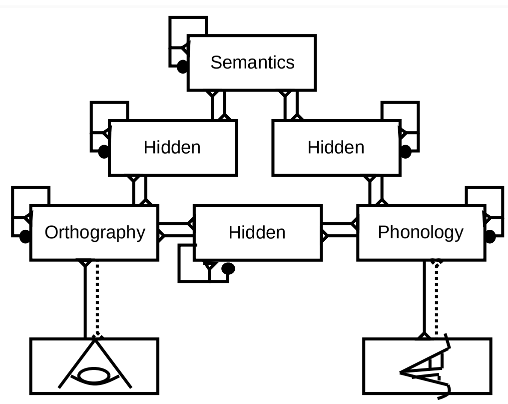
5

Questions

- What general processes are involved in reading, and how do these sometimes fail (e.g., in dyslexia)?
- How are we able to read "cat", "yacht", and "nust"?
- Why do kids say "I goed to school" *after* first saying "I went"?
- How do words come to mean anything?
- How do we go beyond words to sentences?

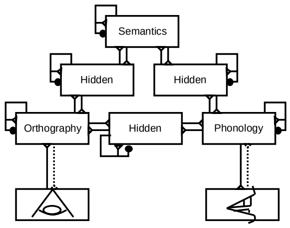
6

Distributed Lexicon Model



7

Distributed Lexicon Model & Reading

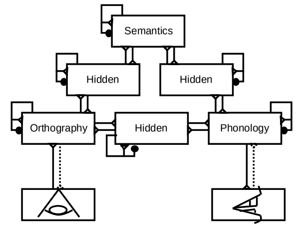


Two Routes

- Direct route: orthography to phonology
- Indirect route: orthography → semantics → phonology

8

Simulating Different Kinds of Dyslexia



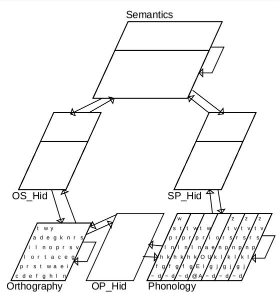
Phonological: nonwords ("nust") impaired.

Deep: phono + semantic errors ("dog" as "cat") + visual errors ("dog" as "dot") + more errors with "truth" (abstract) than "chair" (concrete)

Surface: nonwords OK + semantic access impaired + difficulty reading exception words ("yacht") + visual errors.

9

The Model



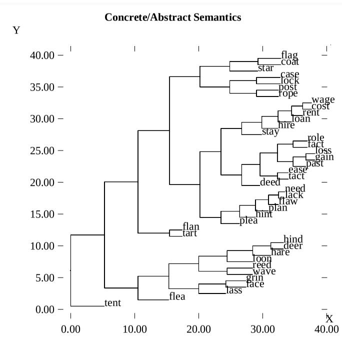
Trained on all pathways (ortho ⇔ phono etc)
for 40 4-letter monosyllabic words (eg flag, star)

Concrete & abstract words use different pools of semantic units

Abstract words activate fewer semantic units than concrete words

10

Corpus and Semantics



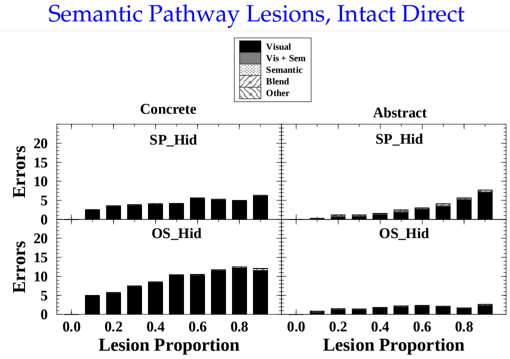
semantic reps made up of distributed features
(e.g. concrete = *living*; abstract = *has-duration*)

11

[dyslex.proj]

12

Semantic Pathway Lesions, Intact Direct



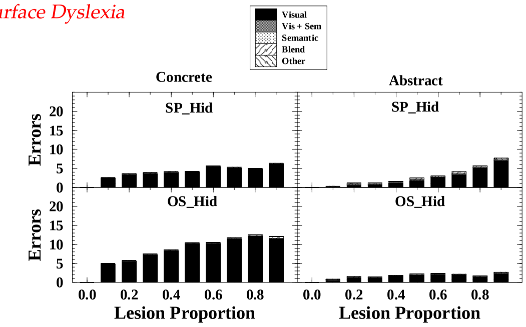
- visual errors with semantic pathway lesions; no semantic errs!

- (more for concrete than abstract..)

13

Semantic Pathway Lesions, Intact Direct

Surface Dyslexia

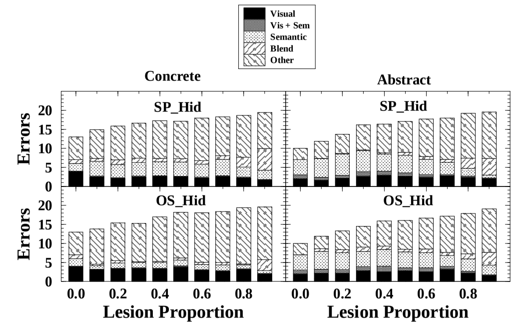


- visual errors with semantic pathway lesions; no semantic errs!

- (more for concrete than abstract..)

14

Semantic Pathway Lesions, Lesioned Direct

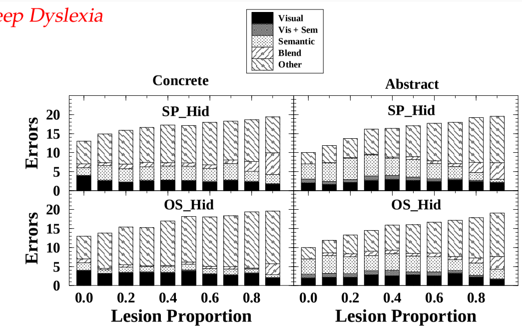


- multiple errors types
- more abstract semantic errs than concrete

15

Semantic Pathway Lesions, Lesioned Direct

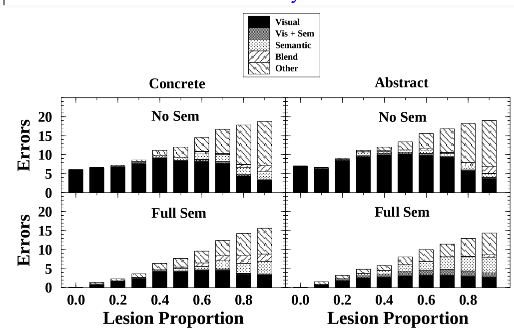
Deep Dyslexia



- multiple errors types
- more abstract semantic errs than concrete

16

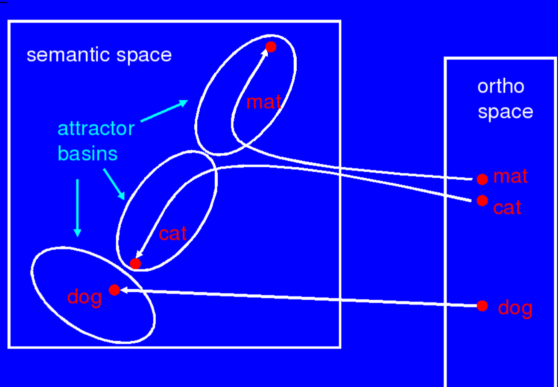
Direct Pathway Lesion



- minor direct damage: just vis errors
- more damage: semantic errs ⇒ deep dyslexia even with Full Sem

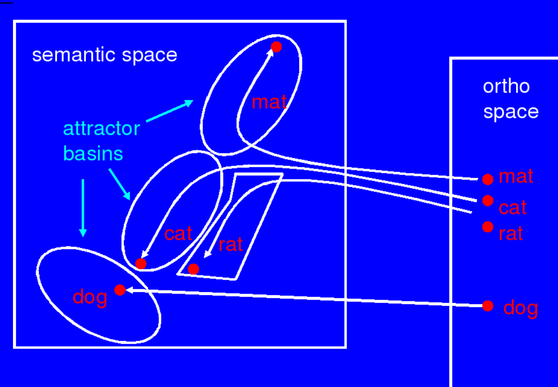
17

Explaining Deep Dyslexia

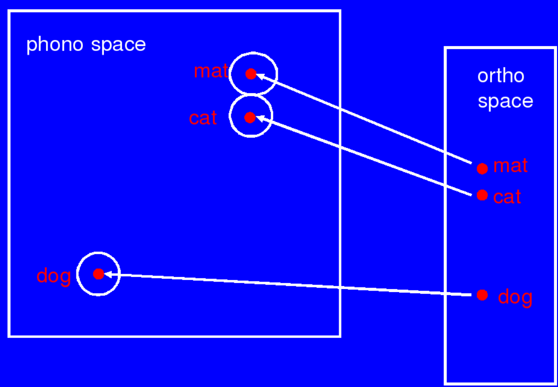


18

Explaining Deep Dyslexia



19 Explaining Surface Dyslexia

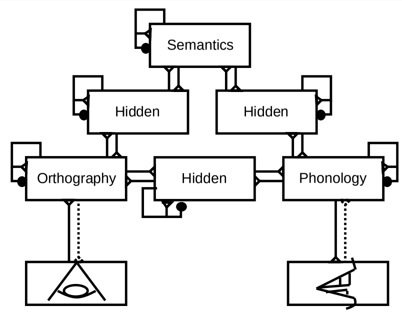


20 Abstract vs Concrete: Summary

- Semantic pathway lesions hurt concrete words more than abstract words
- Concrete words are more strongly represented (more units active) than abstract words in the semantic pathway
- Learning is a function of activation, so the semantic pathway *learns* more about concrete words
- The more semantic pathway learns about concrete words the less direct pathway learns
- The less the direct pathway learns, the less it is able to support performance on its own

- With full direct pathway lesions, the model makes more semantic errors for abstract than concrete
- Abstract words have less distinctive semantic reps than concrete words
- The model is more likely to fall into wrong semantic attractor for abstract words

21 Reading: Distributed Lexicon Model

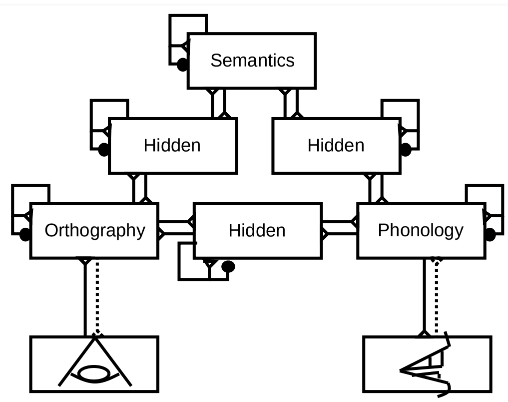


- Distributed reps (not localized to one region).
- Interactive (not modules), leads to interesting divisions of labor.

22 Questions

- What general processes are involved in reading, and how do these sometimes fail (e.g., in dyslexia)?
  Distributed lexicon (ortho, phono, sem)
- How are we able to read "cat", "yacht", and "nust"?
- Why do kids say "I goed to school" *after* first saying "I went"?
- How do words come to mean anything?
- How do we go beyond words to sentences?

23 Distributed Lexicon Model



24 Regularities & Exceptions: A Continuum

Regularities in pronunciation are often partial, *context dependent*: bint

i in *mint, hint, stint*, ... (regular)

vs *pint* (exception)

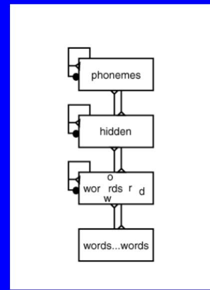
but also: *mind, find, hind*,... (regular)
mine, fine, dine,... (regular)

Pronunciation depends on context.

Exceptions are extreme of context dependent.

Need a range of context dependency for regulars and exceptions.

25 Reading as Object Recognition



- Input to network = 7 letter slots
- Words were trained in multiple positions:

word
 word
  word
   word

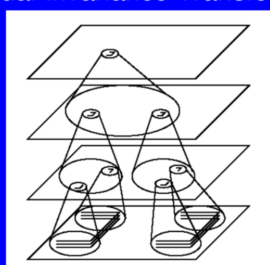
- Hierarchy of processing layers

26 Gradual Invariance Transformations

collapse

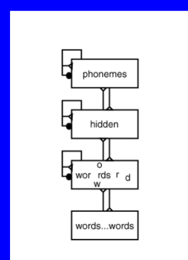
conjoin

conjoin



- Increasing receptive field size gives you two options:
  - Encode location-sensitive conjunctions of features (e.g., a horizontal line below and to the left of a vertical line)
  - Collapse over location or size info

27 Reading as Object Recognition



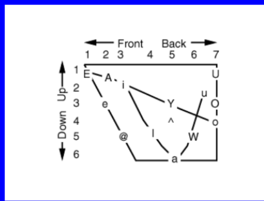
- Second layer ("ortho code") is like V2: It learns a mix of position-invariant individual letter representations, and position-sensitive conjunctions...

28 Phonological Representations

- Same 7 slot vowel-centered representations as before:
  - face = fffAsss
  - grin = grrinnn
  - star = sttarrr
  - post = pppOstt
- Except instead of using a localist rep of each phoneme, we use a distributed rep
- This allows us to represent the fact that phonemes vary in their similarity to one another...

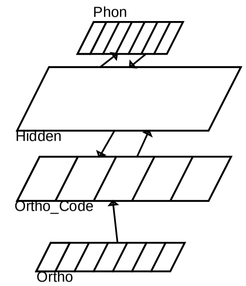
29

Phonology Features: Vowels



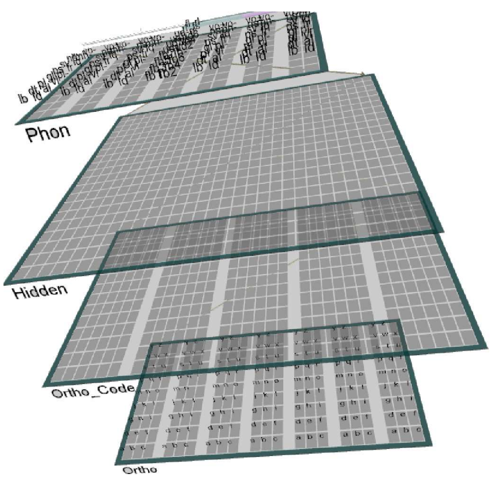
30

Reading Model



detailed model of the "direct" reading pathway (ortho→phono)
- trained to pronounce large set of regular & exception words
- generalization testing: nonwords (eg, nust)

31



32

Nonword Performance

Regularity tests (Glushko): bint → /bint/

Pseudo-homophones (McCann & Besner):
phoyce → /fYs/, choyce → /CYs/

Matched regularity/exception cases (Taraban):
High freq: poes → /pOz/, goes → /gOz/, does → /d^z/
Low freq: mose → /pOs/, poes → /pOz/, lose → /lUz/

Nonword Set	Model	PMSP	People
Glushko regulars	95.3	97.7	93.8
Glushko exceptions raw	79.0	72.1	78.3
Glushko exceptions alt OK	97.6	100.0	95.9
McCann & Besner ctrls	85.9	85.0	88.6
McCann & Besner homoph	92.3	n/a	94.3
Taraban & McClelland	97.9	n/a	100.0[1]

33

Reading Summary

- One network can learn both regular pronunciations and exceptions, and it can generalize properly to nonwords
- Network learns a good mix of context-dependent and context-invariant representations on its own

34

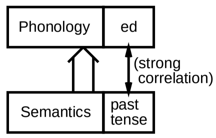
Questions

- What general processes are involved in reading, and how do these sometimes fail (e.g., in dyslexia)? *Distributed lexicon (ortho, phono, sem)*
- How are we able to read "cat", "yacht", and "nust"? *Range of context dependent reps & continuum of regularity-exception*
- Why do kids say "I goed to school" *after* first saying "I went"?
- How do words come to mean anything?
- How do we go beyond words to sentences?

35

Past Tense Simple model

Input: Some units represent word identity, others inflection
Output: Phonology



36

Learning the Past Tense of Verbs
Rumelhart & McClelland (1986)

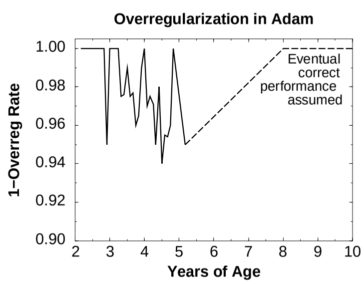
Children exhibit three stages of development:

- Stage 1: Small number of verbs in past tense
  - Very high frequency
  - Majority are irregular
  - *Correct performance*
  - Examples: came, got, gave, looked, needed, took, went
- Stage 2: Larger number of verbs
  - Mostly regular
  - Examples: wiped, pulled
  - *Can generate past tense for invented verbs (rick → ricked)*
  - *Over-regularize words that were correct in stage 1 (goed)*
- Stage 3: Regular and irregular forms coexist
  - Regained use of correct irregular forms

37

Past Tense: U-Shaped Curve

This is the interesting target developmental phenomenon:



38

U-Shaped History

Initially: explained in terms of separate, overzealous rule system.

Then: Rumelhart & McClelland, U-shaped curve based on slow network processing of regularities.

BUT: trained irregulars first, then regs.

(much controversy ensues)

Later: Plunkett et al., etc, manipulate enviro in graded way (continuously add reg verbs to training set instead of all at once).

39

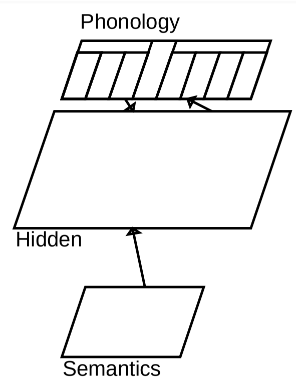
U-Shaped Model in Leabra

Can we get a U-shaped curve without building the explanation into the environment?

- Key: all previous connectionist accounts used feedforward, backprop nets – no attractor dynamics.
- Problem: backprop (pure error-driven learning) leads to steady decrease in error; hard to explain increase in error...
- Interactivity, competition & Hebbian learning produce network that is in dynamic balance between reg & irreg mappings.
- Small tweaks can shift it one way or the other (priming model). String of regular trials will lead to overregularization...

40

The Past Tense Model



41

Phonological Representations

- Same 7 slow vowel-centered reps as before:
  - face = fffAsss
  - grin = grrinnn
  - star = sttarrr
  - post = pppOstt
- Small addition: 8th slot to represent extra inflection
  - started - sstarttD
  - starting = sstarttG

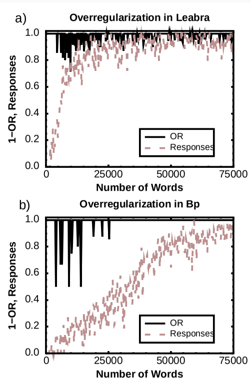
42

The Past Tense Model

Inflection	Reg sfx	Regular/Irregular examples
Base	–	I walk to the store daily. I go to the store daily.
Past	-ed	I walked to the store yesterday. I went to the store yesterday.
3rd pers sing	-s	She walks to the store daily. She goes to the store daily.
Progressive	-ing	I am walking to the store now. I am going to the store now.
Past participle	-en	I have walked to the store before. I have gone to the store now.

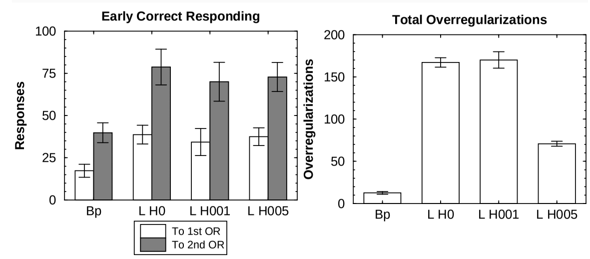
43

Past Tense Results



44

Past Tense Results



45

Past Tense: Summary

- Leabra past tense model shows that you can get U-shaped pattern from a model without manipulating the training environment
- Achieves a substantial level of correct responding prior to onset of overregularization
- Overregularization continues at a low, sporadic rate over an extended period of time. (does this eventually go away?)

46

Past Tense: Summary

- The controversy continues!
- Rules account predicts that past tense acquisition will be sudden and that it will be insensitive to semantic factors...
- Counterexample from Ramscar (2002):
  - “frink” in the context of “drink” = frank
  - “frink” in the context of “blink” = frinked

Table 2
Examples of the "semantic context" passages used in Experiment 2

Passage 1—irregular context—primes *drink*

In a traditional spring rite at Moscow University Hospital, the **terminally ill patients all** *frink* **in the onset of good weather**, consuming vast quantities of vodka and pickled fish. In 1996, his favorite vodka glass in hand, cancer **patient Ivan Borovich** ________ **around** 35 vodka shots and 50 pickled sprats; it is not recorded whether this helped in his treatment.

Passage 2—regular context—primes *blink* and *wink*

In a classical symptom of Howson's syndrome, **patients all** *frink* **in their right eye** if they are left handed or left eye if right handed, their eyelids opening and closing rapidly and uncontrollably. **In 1996**, in extreme discomfort due to his bad eye, Howson's **patient Ivan Borovich** ________ **around** 35 **times per minute** for two days, causing severe damage to the muscles in his left eyelid.

Ramscar (2002)

47 Past Tense: Summary

- The controversy continues!
- Rules account predicts that past tense acquisition will be sudden and that it will be insensitive to semantic factors...
- Counterexample from Ramscar (2002):
  - "frink" in the context of "drink" = frank
  - "frink" in the context of "blink" = frinked
- Neuroimaging irreg & reg past tenses show overlapping neural activation (Joanisse & Seidenberg, 2005) – not two systems.
- See also McClelland vs Pinker smackdown in *Trends in Cognitive Sciences (2002)*

48 Questions

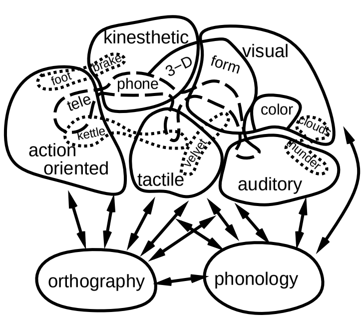
- What general processes are involved in reading, and how do these sometimes fail (e.g., in dyslexia)?

  Distributed lexicon (ortho, phono, sem)
- How are we able to read "cat", "yacht", and "nust"? *Range of context dependent reps & continuum of regularity-exception*
- Why do kids say "I goed to school" *after* first saying "I went"? *Dynamic balance between regular & exception mapping*
- How do words come to mean anything?
- How do we go beyond words to sentences?

49 How Do Words Come to Mean Anything?

- What Gives Words Their Meaning?
- Where Does this Meaning Come From?

50 What Gives Words Their Meaning?: Distributed Semantics



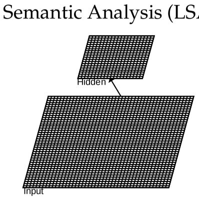
Semantics is distributed across specialized processing areas.

51 Where Does this Meaning Come From?: Correlational Semantics

Hebbian learning encodes structure of word co-occurrence.

Same idea as:

- V1 receptive field learning: learn the strong correlations.
- Similar to Latent Semantic Analysis (LSA)



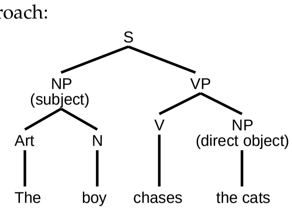
52 [sem.proj]

53 Multiple-Choice Quiz

0.	neural activation function	5.	attention
A	spiking rate code membrane potential pt	A	competition inhibition selection binding
B	interactive bidirectional feedforward	B	gradual feature conjunction spatial invariance
C	language generalization nonwords	C	spiking rate code membrane potential point
1.	transformation	6.	weight based priming
A	emphasizing distinctions collapsing diffs	A	long term changes learning
B	error driven hebbian task model based	B	active maintenance short term residual
C	spiking rate code membrane potential pt	C	fast arbitrary details conjunctive
2.	bidirectional connectivity	7.	hippocampus learning
A	amplification pattern completion	A	fast arbitrary details conjunctive
B	competition inhibition selection binding	B	slow integration general structure
C	language generalization nonwords	C	error driven hebbian task model based
3.	cortex learning	8.	dyslexia
A	error driven task based hebbian model	A	surface deep phonological reading problem
B	error driven task based	B	speech output hearing language nonwords
C	gradual feature conjunction spatial invar	C	competition inhibition selection binding
4.	object recognition	9.	past tense
A	gradual feature conjunction spatial invar	A	overregularization shaped curve
B	error driven task based hebbian model	B	speech output hearing language nonwords
C	amplification pattern completion	C	fast arbitrary details conjunctive

54 Sentences: Beyond just semantics

Traditional approach:



Alternative approach:

Distributed reps of sentence *meaning*: The sentence Gestalt!

(gestalt = unified configuration of elements that can't be described merely as a sum of parts)

55 Sentence Comprehension

- We want to build an internal model of the situation
- e.g., "The teacher drank Pepsi in the classroom"
  - Who/what is the agent? teacher
  - What is the patient (object)? Pepsi
  - What did the agent do? drink
  - Where? classroom (and so on)...
- Goal: Teach a model to understand sentences
- Present one word at a time
- Want the model to be able to answer questions, e.g., Who is the agent? (has to be able to do this even if agent not currently in input)

56 Toy World

People: *busdriver* (adult male), *teacher* (adult female), *schoolgirl*, and *pitcher* (boy).

Actions: *eat, drink, stir, spread, kiss, give, hit, throw, drive, rise*.

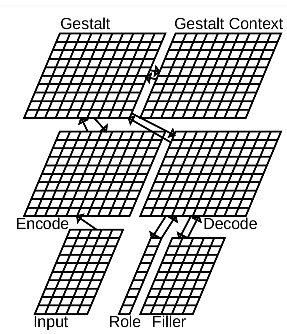
Objects: *spot* (the dog), *steak, soup, ice cream, crackers, jelly, iced tea, kool aid, spoon, knife, finger, rose, bat* (animal), *bat* (baseball), *ball, ball* (party), *bus, pitcher,* and *fur*

Locations: *kitchen, living room, shed,* and *park*.

Syntax: Active & Passive, phrases.

Some events more probable than others (eg busdrivers eat steak more often than teachers)

57 Network



To answer questions at the end of sentence, net needs to *actively maintain* info about words it has seen... SRN

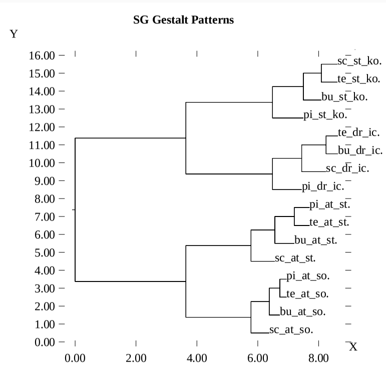
58 Training

- Present words & their roles, one at a time; after each word/role pair, quiz the net on what it has seen up to that point
- The busdriver stirred Kool-Aid
- Present "busdriver" + agent
  - Who is the agent? busdriver
- Present "stirred" + action
  - What is the action? stirred
  - Who is the agent? busdriver
- Present "Kool-Aid" + patient
  - What is the patient? Kool-Aid
  - Who is the agent? busdriver
  - What is the action? stirred

59 Tests

Task	Sentence
Role assignment	
Active semantic	The schoolgirl stirred the kool-aid with a spoon.
Active syntactic	The busdriver gave the rose to the teacher.
Passive semantic	The jelly was spread by the busdriver with the knife.
Passive syntactic	The teacher was kissed by the busdriver.
(control)	The busdriver kissed the teacher.
Word ambiguity	The busdriver threw the ball in the park.
	The teacher threw the ball in the living room.
Concept instantiation	The teacher kissed someone (male).
Role elaboration	The schoolgirl ate crackers (with finger).
	The schoolgirl ate (soup).
Online update	The child ate soup with daintiness.
(control)	The pitcher ate soup with daintiness.
Conflict	The adult drank iced-tea in the kitchen (living-room).

60 Gestalt Representations



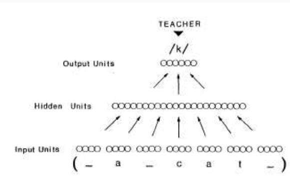
61 Problems with the statistical approach?

- The model makes mistakes for infrequent and/or irregular sentences
- Example: busdriver ate soup; responds with *steak* as patient
- Explanation: Net saw busdriver eating steak 7x more than soup
- Statistical model overrides reality...
- People suffer from similar biases! (How many animals did Moses bring on the ark?)

62 Questions

- What general processes are involved in reading, and how do these sometimes fail (e.g., in dyslexia)?
- How are we able to read "cat", "yacht", and "nust"?
- Why do kids say "I goed to school" *after* first saying "I went"?
- How do words come to mean anything? *Statistics of word co-occurrences.*
- How do we go beyond words to sentences? *Sentence gestalt*

63 Application: NetTalk! (Sejnowksi & Rosenberg, 1986)



- Learns to read & pronounce english text
- Inputs are one of 29 chars (26 letters + space, comma, full stop)
- 7 letter window (provides context). total = 29x7 = 203 inputs.
- Hidden layer of 80 units.
- Output generates one of 60 phonemes, represented by 21 articulation units and 5 units for stress/syllable boundary info.

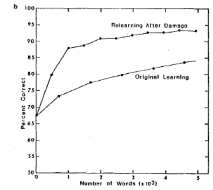
64

NetTalk: Results

- Learns regularities of english speech
- Generalizes to novel words not in training set with 78% accuracy

65

NetTalk: Results



- Knowledge is distributed: relearning after damage much faster than original training
- Distributed (spaced) practice more effective for long term retention than massed practice

66

NetTalk: Impressive, But...

- Solves reading and speaking at once (unlike people)
- Doesn't address specialization of different brain areas in language processing.
- Uses biologically implausible "error backpropagation" method for training weights
- Explicit "teacher" provides correct information on output articulatory units (instead of trial and error learning that we have to do)
- Requires many passes through exact same training set (rather than natural language experiences).