How can we possibly simulate language abilities in neural terms?

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

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.

Language
How can we possibly simulate language abilities in neural terms? Beyond
levels: phonemes/letters, words, phrases, sentences, paragraphs, and
just another set of input/output paths.

(We can’t... But many aspects of language are not “special”)

How can we possibly simulate language abilities in neural terms?

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!
Biological Substrates of Language

Broca's

Wernicke's

Occipital

Frontal

Parietal

Temporal
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
Traditional view of language

• Language competence defined by knowledge of rules and exceptions (e.g., *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, and semantics

Language competence defined by knowledge of rules and exceptions
Neural net / Connectionist View of Language

• Language is another set of input-output mappings (e.g., 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...

Language is another set of input-output mappings (e.g., orthography to semantics)
Neural net / Connectionist View of Language

- Language is another set of input-output mappings (e.g., 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"
  - But blown goes with grown

- “But blown” is an exception.
  - “ blown: blown”
  - “down, down”

These mappings are trained up using the same learning algorithms (e.g., orthography to semantics, phonology)

Languages is another set of input-output mappings (e.g., orthography to semantics)

But blown goes with grown

Neural net / Connectionist View of Language
Neural net / Connectionist View of Language

- Language is another set of input-output mappings (e.g., orthography to phonology, orthography to semantics)
- These mappings are trained 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"
  - Distributed lexicon: Knowledge about words is embodied in reciprocal mappings between phonology, orthography, and semantics - there is no central "word representation"

Distributed lexicon: Knowledge about words is embodied in reciprocal mappings between phonology, orthography, and semantics - there is no central "word representation"

- Implausible in the case of "blown"
  - exception: blown
  - regular: down, grown

The same pathways handle both rules and exceptions used elsewhere (e.g., vision)

These mappings are trained using the same learning algorithms (e.g., orthography to semantics)

Languages is another set of input-output mappings (e.g., orthography to semantics)

Neural net / Connectionist View of Language
• What general processes are involved in reading, and how do these sometimes fail (e.g. in dyslexia)?
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 "must"?
Questions

• Why do kids say “I goed to school” after first saying “I went”?

• How are we able to read “cat”, “yacht”, and “must”?

• Sometimes fail (e.g. in dyslexia)?

• What general processes are involved in reading, and how do these
Questions

• How do words come to mean anything?

• Why do kids say “I goed to school” after first saying “I went”?

• How are we able to read “cat”, “yacht”, and “must”?

• Sometimes fail (e.g., in dyslexia)

• What general processes are involved in reading, and how do these...
Questions

• How do we go beyond words to sentences?

• How do words come to mean anything?

• Why do kids say “I goed to school” after first saying “I went”?

• How are we able to read “cat”, “yacht”, and “must”?

• Sometimes fail (e.g., in dyslexia)

• What general processes are involved in reading, and how do these
Distributed Lexicon Model & Reading
Two Routes

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

Distributed Lexicon Model of Reading
Simulating Different Kinds of Dyslexia
Simulating Different Kinds of Dyslexia

Phonological: nonwords ("nust") impaired.
Simulating Different Kinds of Dyslexia

**Phonological:**
- more errors with “truth” (abstract) than “chair” (concrete)
- nonwords (“nust”) impaired.

**Deep:**
- phon + semantic errors (“dog” as “cat”)
- visual errors (“dog” as “dot”)
- more errors with “truth” (abstract) than “chair” (concrete)
- nonwords (“must”) impaired.
Simulating Different Kinds of Dyslexia

**Phonological:**
- nonwords OK + no semantic errors (but semantic access impaired).

**Surface:**
- difficulty reading exception words (e.g., "yacht")
- more errors with abstract (e.g., "truth") than concrete (e.g., "chair")
- visual errors (e.g., "dog" as "cat") + semantic errors (e.g., "dog" as "cat")

**Deep:**
- phonological: nonwords "must" impaired
  - impaired

Simulations of Different Kinds of Dyslexia

Diagram showing the relationships between phonology, orthography, semantics, and hidden layers.
Trained on all pathways (ortho ⇔ phono etc) for 40 4-letter monosyllabic words (e.g. flag, star)
Concrete & abstract words use different pools of semantic units.

Abstract words activate fewer semantic units than concrete words.

For 40 4-letter monosyllabic words (e.g., flag, star), abstract words activate fewer semantic units than concrete words.

Trained on all pathways (ortho ↔ phono etc.)
Corpus and Semantics

Concrete/Abstract Semantics

Example: concrete = living, abstract = has-duration

Semantic reps made up of distributed features (e.g., concrete = living; abstract = has-duration)
Semantic Pathway Lesions, Intact Direct

- Visual errors with semantic pathway lesions; no semantic errors!

Bars represent error distribution with various lesion proportions.
- Visual errors with semantic pathway lesions, no semantic errors!

- More for concrete than abstract.

Semantic Pathway Lesions, Intact Direct

Abstract

Concrete
Semantic Pathway Lesions, Intact Direct Surface Dyslexia

- Visual errors with semantic pathway lesions; no semantic errors!
- More for concrete than abstract.

Graphs showing lesion proportion and errors for concrete and abstract words with different lesion types.
more abstract semantic errs than concrete
- multiple errors types

Semantic Pathway Lesions, Lesioned Direct

- abstract errors vs concrete errors
- bar charts showing error distribution across lesion proportions for different conditions (Vis, Sem, Blend, Other, Vis + Sem)
Semantic Pathway Lesions, Lesioned Direct

- more abstract semantic errors than concrete
- multiple errors types

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

Direct Pathway Lesion

Concrete

Abstract

Lesion Proportion

Lesion Proportion

0.0 0.2 0.4 0.6 0.8

0 5 10 15 20

Errors

Errors

Full Sem

No Sem

Full Sem

No Sem

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

Direct Pathway Lesion
Explaining Deep Dyslexia
Explaining Surface Dyslexia

dog

cat

mat

ortho

space

phon space
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.
- Learning is a function of activation, so the semantic pathway learns more about concrete words the less direct pathway learns.
- The more semantic pathway learns about concrete words, the less direct pathway learns.
- The less direct pathway learns, the less it is able to support performance on its own.
- Abstract words in the semantic pathway are more strongly represented (more units active) than concrete words, but concrete words hurt concrete words more than abstract words.
With full direct pathway lesions, the model makes more semantic errors for abstract than concrete words.

Abstract words have less distinctive semantic reps than concrete words.

The model is more likely to fall into wrong semantic attractor for abstract words.

Errors for abstract than concrete

With full direct pathway lesions, the model makes more semantic
Reading: Distributed Lexicon Model

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

![Diagram of Distributed Lexicon Model]
Questions

• How do we go beyond words to sentences?

• How do words come to mean anything?

• Why do kids say “I goed to school” after first saying “I went”?

• How are we able to read “cat”, “yacht”, “nust”?

Distributed Lexicon (ortho, phono, sem)

Sometimes fail (e.g. in dyslexia)

What general processes are involved in reading, and how do these...

Questions
Distributed Lexicon Model
Regularities in pronunciation are often partial, context dependent.
Regularities in pronunciation are often partial, context dependent:

Regularities & Exceptions: A Continuum
Regularities & Exceptions: A Continuum

Regularities in pronunciation are often partial, context dependent:

\[ \text{bint} \quad \text{in mint, hint, stint} \quad \text{(regular)} \]

\[ \text{pint} \quad \text{vs pint} \quad \text{(exception)} \]

Regularities in pronunciation are often partial, context dependent:

Regulatilities & Exceptions: A Continuum
Regularities & Exceptions: A Continuum

Regularities in pronunciation are often partial, context dependent:

- bint, mint, hint, stint, ... (regular)
- vs pint (exception)

but also:

- mind, find, hind, ... (regular)
- mine, fine, dine, ... (regular)
- in mint, hint, stint, ... (regular)

Regularties in pronunciation are often partial, context dependent:
Regularities & Exceptions: A Continuum

Pronunciation depends on context:

\[ \text{mine, fine, dine, \ldots (regular)} \]
\[ \text{but also: mind, kind, Hind, \ldots (regular)} \]
\[ \text{vs pint (exception)} \]
\[ \text{in mint, hint, shint, \ldots (regular)} \]

Regularities in pronunciation are often partial, context dependent:

Regularities & Exceptions: A Continuum
Regularities & Exceptions: A Continuum

Regularities in pronunciation are often partial, context dependent:

*bint* vs *pint* (exception)

\[\text{bint, hint, stint, ... (regular)}\]

\[\text{mind, find, hind, ... (regular)}\]

\[\text{mine, fine, dine, ... (regular)}\]

\[\text{but also: mind, hind, Hind, ... (regular)}\]

\[\text{vs pint (exception)}\]

\[\text{in mint, hint, stint, ... (regular)}\]

\[\text{pint}\]

Exceptions are extreme of context dependent.

Pronunciation depends on context:

Regularities & Exceptions: A Continuum
Regularities & Exceptions: A Continuum

Regularities in pronunciation are often partial, context dependent:

- mint, hint, stint, ... (regular)
- pint (exception)

Exceptions are extreme of context dependent:

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

Pronunciation depends on context.

Regularties in pronunciation are often partial, context dependent.
Hierarchy of processing layers

- Word
- Word
- Word positions
- Words were trained in multiple

Input to network = 7 letter slots

Reading as Object Recognition
Gradual Invariance Transformations

- Encode location-sensitive conjunctions of features (e.g.,

Increasing receptive field size gives you two options:

- Collapse over location or size into a horizontal line below and to the left of a vertical line,
- Encode location-sensitive conjunctions of features (e.g.,

...
Reading as Object Recognition

- Position-sensitive conjunctions...
- Individual letter representations, and
- It learns a mix of position-invariant
- Second layer ("ortho code") is like V2;
Phonological Representations

- face = fffAsss
- grin = grrinnn
- star = sttarrr
- post = ppOstt

This allows us to represent the fact that phonemes vary in their similarity to one another.

Except instead of using a localist rep of each phoneme, we use a distributed rep.

- star = sttarrr
- grin = grrinnn
- face = fffAsss

Same 7 slot vowel-centered representations as before.
Phonology Features: Vowels
A detailed model of the "direct" reading pathway (ortho→phon) - trained to pronounce large set of regular & exception words - generalization testing: nonwords (e.g., must).
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 → /dˆz/
Low freq: mose → /pOs/, poes → /pOz/, lose → /lUz/

<table>
<thead>
<tr>
<th></th>
<th>Model</th>
<th>PMSP</th>
<th>People</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taraban × McLeodland</td>
<td>100.0</td>
<td>97.9</td>
<td></td>
</tr>
<tr>
<td>Glushko regulars</td>
<td>94.3</td>
<td>92.3</td>
<td>88.6</td>
</tr>
<tr>
<td>Glushko exceptions alt OK</td>
<td>96.9</td>
<td>97.6</td>
<td>97.0</td>
</tr>
<tr>
<td>Glushko exceptions raw</td>
<td>78.3</td>
<td>72.1</td>
<td>79.0</td>
</tr>
<tr>
<td>Glushko regulars</td>
<td>93.8</td>
<td>9.7</td>
<td>95.3</td>
</tr>
</tbody>
</table>
• One network can learn both regular pronunciations and exceptions.

Network learns a good mix of context-dependent and context-invariant representations on its own.

• One network can learn both regular pronunciations and exceptions.

and it can generalize properly to nonwords.
Questions

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

Input: Some units represent word identity, others inflection
Output: Phonology (strong correlation)

Past tense

Semantics

ed

Phonology
Children exhibit three stages of development: Rumelhart & McClelland (1986) Learning the Past Tense of Verbs

- Stage 1: Small number of verbs in past tense
  - Examples: came, got, gave, looked, needed, took, went
  - Correct performance
  - Majority are irregular
  - Very high frequency

Children exhibit three stages of development: Rumelhart & McClelland (1986) Learning the Past Tense of Verbs
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
  - Can generate past tense for invented verbs (rick → ricked)
  - Mostly regular
  - Examples: wiped, pulled

- **Stage 3:** Large number of verbs
  - Correct performance
  - Very high frequency

Rumelhart & McClelland (1986) Learning the Past Tense of Verbs
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
  - Can generate past tense for invented verbs (rick → ricked)
  - Over-regularize words that were correct in stage 1 (goed)
  - Examples: wiped, pulled

- Stage 3: Regular and irregular forms coexist
  - Regained use of correct irregular forms
  - Over-regularize words that were correct in stage 1 (goed)
  - Correct performance
  - Majority are irregular
  - Very high frequency

Rumelhart & McClelland (1986) Learning the Past Tense of Verbs
This is the interesting target developmental phenomenon:

Past Tense: U-Shaped Curve
U-Shaped History

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

U-Shaped History
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 reg's.

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

(much controversy ensues)

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

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

U-Shaped History
U-Shaped Model in Leabra

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

- In dynamic balance between reg & irreg mappings.

- Interactivity, competition & Hebbian learning produce network that is in dynamic balance between reg & irreg mappings.

- Problem: backprop (pure error-driven learning) leads to steady decrease in error; hard to explain increase in error.

- Key: all previous connectionist accounts used feedforward, backprop nets – no attractor dynamics.

- U-shaped Model in Leadra:
Can we get a U-shaped curve without building the explanation into the environment?

- Small tweaks can shift it one way or the other (priming model). String of regular trials will lead to overregularization.
- Small tweaks can shift it one way or the other (priming model). String of regular trials will lead to overregularization.

...in dynamic balance between reg & irreg mappings:

- Interactivity, competition & Hebbian learning produce network that is...

...decrease in error; hard to explain increase in error:

- Problem: backprop (pure error-driven learning) leads to steady nets – no attractor dynamics.

...no attractor dynamics.

Key: all previous connectionist accounts used feedforward, backprop.
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 = sstarrtt
  – started - started

Same / slow vowel-centered reps as before.
<table>
<thead>
<tr>
<th>Past Tense Model</th>
<th>Regular/Irregular examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past participle</td>
<td>-en</td>
</tr>
<tr>
<td>Progressive</td>
<td>-ing</td>
</tr>
<tr>
<td>3rd pers sing</td>
<td>-s</td>
</tr>
<tr>
<td>Past</td>
<td>-ed</td>
</tr>
<tr>
<td>Base</td>
<td>-</td>
</tr>
</tbody>
</table>

**Examples:**
- I have gone to the store now.
- I have walked to the store before.
- I am going to the store now.
- I am walking to the store now.
- She goes to the store daily.
- She walks to the store daily.
- She has walked to the store before.
- She has gone to the store now.
- I went to the store yesterday.
- I walked to the store yesterday.
- I go to the store daily.
- I walk to the store daily.
Overregularization in BP

Overregularization in Leabra

Past Tense Results
Past Tense Results

Early Correct Responding

Overregularizations

Total Overregularizations

Responses
Past Tense: Summary

- Leabra past tense model shows that you can get U-shaped patterns 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?)

- Leadha past tense model shows that you can get U-shaped patterns from a model without manipulating the training environment.
Past Tense: Summary

- Rules account predicts that past tense acquisition will be sudden and insensitive to semantic factors...
- The controversy continues!
Past Tense: Summary

• The controversy continues!
• Rules account predicts that past tense acquisition will be sudden and insensitive to semantic factors.

Counterexample from Ramscar (2002):

- "frink" in the context of "drink" = franked
- "frink" in the context of "blink" = frinked
Past Tense: Summary

The controversy continues!

• Rules account predicts that past tense acquisition will be sudden and insensitive to semantic factors...

Counterexample from Ramscar (2002):

- "frink" in the context of "drink" = franked
- "frink" in the context of "drink" = frink

The controversy continues!

Past Tense: Summary
Ramscar (2002)

Passage 2—Regular context—Primes think and work

Ramscar (2002) around 35 times per minute for two days, causing severe damage to the muscles in his left eyelid. He lost his ability to close his eye, and his left eyelid was barely controllable. In 1996, he experienced discomfort due to his bad eye. His vision was severely impaired, and he had difficulty opening and closing his eye rapidly and accurately. He was later diagnosed with Howson's syndrome, a rare condition that affects patients' ability to close their eyes quickly.

In a classical symptom of Howson's syndrome, patients all blink in their right eye if they are left handed or left eye if right handed. Their eyelids open and close more rapidly and accurately.

Passage 1—Irregular context—Primes drink and work

Ramscar's (2002) around 35 times per minute for two days, causing severe damage to the muscles in his left eyelid. He lost his ability to close his eye, and his left eyelid was barely controllable. In 1996, he experienced discomfort due to his bad eye. His vision was severely impaired, and he had difficulty opening and closing his eye rapidly and accurately. He was later diagnosed with Howson's syndrome, a rare condition that affects patients' ability to close their eyes quickly.

In a classical symptom of Howson's syndrome, patients all blink in their right eye if they are left handed or left eye if right handed. Their eyelids open and close more rapidly and accurately.

Examples of the "semantically context" Passages used in Experiment 2

Table 2
Past Tense: Summary

• The controversy continues!

• Rules account predicts that past tense acquisition will be sudden and insensitive to semantic factors...

• Counterexample from Ramscar (2002):

  - "frink" in the context of "drink" = frinked
  - "frink" in the context of "blink" = frinked

  Neurorimaging inter alia past tenses show overlapping neural activation (Joanisse & Seidenberg, 2005) – not two systems.

  The controversy continues!
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" = frinked
- "frink" in the context of "blink" = frinked
- "fink" in the context of "drink" = frink

Neuroimaging of 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)
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”? Range of context dependent reps vs. continuum of regularity-exception mapping

• 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?

• Distributed lexicon (ortho, phon, sem) sometimes fail (e.g., in dyslexia)?

• What general processes are involved in reading, and how do these?
How Do Words Come to Mean Anything?

• Where Does This Meaning Come From?

• What Gives Words Their Meaning?

• How Do Words Come to Mean Anything?
What Gives Words Their Meaning?: Distributed Semantics

Semantics is distributed across specialized processing areas.
Where Does this Meaning Come From?

Correlational Semantics

Hebbian learning encodes structure of word co-occurrence.
Where Does this Meaning Come From?

*Correlational Semantics*

Hebbian learning encodes structure of word co-occurrence.

Similar to Latent Semantic Analysis (LSA)

\[ V_1 \text{ receptive field learning: learn the strong correlations.} \]

Same idea as:

Correlational Semantics
Where Does this Meaning Come From?

Correlational Semantics

Hebbian learning encodes structure of word co-occurrence.

Similar to Latent Semantic Analysis (LSA)

• V1 receptive field learning: Learn the strong correlations.

Same idea as:

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

Traditional approach:

Sentences: Beyond just semantics
Beyond just semantics

Traditional approach:

Alternative approach:

Distributed reps of sentence meaning: The sentence Gestalt!
Beyond just semantics

Traditional approach:

S (subject) Art N The boy NP V NP (direct object) chases the cats NP VP

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

The sentence Gestalt: Sentences beyond just semantics
Sentence Comprehension

- We want to build an internal model of the situation.
  - e.g., "The teacher drank Pepsi in the classroom"
    - Who 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

Want the model to be able to answer questions, e.g., Who is the agent?

Present one word at a time

We want to build an internal model of the situation
We want to build an internal model of the situation, e.g., “The teacher drank Pepsi in the classroom.”

- Who 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

- Where? classroom (and so on)...
- What did the agent do? drink
- What is the patient (object)? Pepsi
- Who is the agent? teacher

E.g., “The teacher drank Pepsi in the classroom”

We want to build an internal model of the situation.

Sentence Comprehension
Toy World

People:

- bus driver (adult male)
- teacher (adult female)
- schoolgirl
- pitcher (boy)

Actions:

- eat, drink, stir, spread, kiss, give, hit, throw, drive

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)

Locations:

- kitchen, living room, shed, and park
- bus, pitcher, and fur

Syntax: Active & Passive, phrases.
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, bat (animal), bat (baseball), ball, ball (party), and spoon, knife, finger, rose, bat (animal), bat (baseball), ball, ball (party).

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

Syntax: Active & Passive, phrases, Some events more probable than others (e.g., busdrivers eat steak more often than teachers eat steak).
To answer questions at the end of sentences, net needs to actively maintain info about words it has seen... SRN
• 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.

   • Who is the agent? busdriver
   • What is the patient? Kool-Aid

   • Present "busdriver" + agent
   • Present "Kool-Aid" + patient

   • What is the action? stirred
   • Who is the agent? busdriver
   • What is the action? stirred

   • Present "stirred" + action
   • Present "busdriver" + agent
   • Present "stirred Kool-Aid" + action

   • Who is the agent? busdriver
   • Present "busdriver" + agent

   • The busdriver stirred Kool-Aid.

   Present words & their roles, one at a time; after each word/role pair.

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

Tests
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

- People suffer from similar biases!
  - How many animals did Moses bring on the ark?

- Statistical model overrides reality...

The model makes mistakes for infrequent and/or irregular sentences.
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.

What general processes are involved in reading, and how do these sometimes fail (e.g., in dyslexia)?
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” after first saying “I went”?

• Why do kids say “I goed to school” after saying “I went”?

• How do words come to mean anything? Statistics of word co-occurrence.

• How do we go beyond words to sentences? Sentence gestalt.

• What general processes are involved in reading, and how do these sometimes fail (e.g., in dyslexia)?
Application: NetTalk! (Sejnowski & Rosenberg, 1986)

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

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

![Graph showing learning curves](image)
NetTalk: Impressive, But...
NetTalk: Impressive, But...

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

Natural language experiences.

(\textit{\textbf{NetTalk: Impressive, But...}})