9. Language

Language Involves all of Cognition
- Perception: hearing & reading words
- Attention: picking out words, speakers from many
- Motor: speech, writing, etc.
- Memory: semantics, specific content – how do you encode plot of a book?
- Executive Function: maintaining context, planning speech, syntax structure...

Language is Special
- Symbols — thought is reduced to symbols to reconstruct thoughts
- Syntax
- Temporally-extended sequences
- Cultural transmission
- Embedded levels of structure (generativity):
  - “The horse racing past the barn fell”
  - “Isn’t it true that example-sentences that people that you know produce are more likely to be accepted?”

Language Controversies
- How special is language: just co-opting existing neural mechanisms vs. innate language modules?
  - Poverty of Stimulus (POS)
- Rules vs. regularities: Is there anything special about rule-like behavior in language?
  - Spelling to sound: Exceptions also have sub-rules
  - Over-regularization (add “-ed” = “goed”) – evidence of rule system coming online?
  - Truly variable-like behavior? Generative, abstract.

What is Truly Novel?
- Pure syntax: “Adj Adj Noun Verb Adverb”
- “Colorless green ideas sleep furiously”
  - But: “ Newly formed bland ideas are inexpressible in an infuriating way.”
  - “It can only be the thought of verdure to come, which prompts us in the autumn to buy these dormant white lamps of vegetable matter covered by a brown papery skin, and lovingly to plant them and care for them. It is a marvel to me that under this cover they are labouring unseen at such a rate within to give us the sudden awesome beauty of spring flowering bulbs. While winter reigns the earth repose but these colorless green ideas sleep furiously.”
- “‘Twas brillig, and the slithy toves…”
  - But each word has some overlap with real words.

Time and Language
- Distributed representation broken down and sent over a sequential channel:
  The summer is a fun time for going to the beach, dancing, …
Distributed Representations of Words
I couldn't believe that I couldn't use the word "waltz" I was writing. Using the incorrect spelling of the human mind, a sequence of letters at Cambridge University, it seems that "waltz" in waltz is still pronounced differently than in the other claus. The net can be a taotl model and you can still read it without it being aliased. This is because the human mind does not read every letter by itself, but the word as a whole.

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... 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"

Calculus
- In Latin, calculus means pebble
- In ancient times calculi were used for calculating (as on an abacus), voting, and many other purposes
- Now, a calculus is:
  - an mechanical method of solving problems
  - by manipulating discrete tokens
  - according to formal rules
- Examples: algebraic manipulation, integral & differential calculi, logical calculi

Formal Logic
- Originally developed by Aristotle (384–322 BCE)
- A syllogism:
  - All men are mortal
  - Socrates is a man
  - ∴ Socrates is mortal
- Formal logic: the correctness of the steps depend only on their form (syntax), not their meaning (semantics):
  - All men are P
  - S is P
  - ∴ S is P
- More reliable, because more mechanical
9. Language

Common Features of Calculi

- Information (data) representation is:
  - formal (info. represented by arrangements)
  - finite (finite arrangements of atomic tokens)
  - definite (can determine symbols & syntax)

- Information processing (rule following) is:
  - formal (depends on arrangement, not meaning)
  - finite (finite number of rules & processing time)
  - definite (know which rules are applicable)

Thought as Calculation

- “By ratiocination I mean computation.”
  — Thomas Hobbes (1588–1679)

- “Then, in case of a difference of opinion, no discussion … will be any longer necessary … It will rather be enough for them to take pen in hand, set themselves to the abacus, and … say to one another, “Let us calculate!” — Leibniz (1646–1716)

- Boole (1815–64): his goal was “to investigate the fundamental laws of those operations of mind by which reasoning is performed; to give expression to them in the symbolical language of a Calculus”

Early Investigations in Mechanized Thought

- Leibniz (1646–1716): mechanical calculation & formal inference
- Boole (1815–1864): “laws of thought”
- Jevons (1835–1882): logical abacus & logical piano
- von Neumann (1903–1957): computation & the brain

Symbolic AI & Cognitive Science

- Western philosophy has generally assumed knowledge and cognition can be described in terms of language-like structures (a language of thought)
- Symbolic AI and CogSci assumed intelligence resides in:
  - the structures of a knowledge representation language
  - deduction-like formal rules for their manipulation
- Knowledge represented at a symbolic level:
  - atomic word-level categories
  - related by sentence-like logical structures
- Semantics is reduced to syntax
  - that is meaning is representational and manipulated formally (calculation)
- The brain doesn’t matter because all computers are equally powerful
  - competence is relevant, but not performance
- Is a language of thought “the only game in town”?

Connectionist AI & Cognitive Science

- Knowledge is represented at a sub-symbolic level:
  - in terms of minute, quantitative features
  - related by low-level, often statistical connections
  - grounded in sensorimotor interactions
- Knowledge is more akin to an image than a sentence
- The brain is very relevant because:
  - it shows us how to do connectionist information processing
  - it places tight constraints on models, because brains have to work in real time: performance is critical (100 step rule etc.)
  - on the other hand, e.g., symbolic approaches try to account for unlimited nesting, connectionist don’t need to

Some 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 children say “I goed to school” when previously they said “I went”?
- How do words come to mean anything?
- How do we go beyond words to sentences?
Outline: Language

A. Biology of Language
B. Reading and Dyslexia in the Triangle Model
C. Spelling to Sound Mappings in Word Reading
D. Latent Semantics in Word Co-occurrence
E. Syntax and Semantics in a Sentence Gestalt

Language is Instinctual
Evidence:
• Skeletal specializations for speech in our earliest hominid ancestors
• Left hemisphere specialization evident before birth
• Ability to recognize any phoneme at birth — basis for learning one’s native languages
• On the other hand, writing is not instinctual
Aphasia

- Aphasia = an acquired language disorder
  - disorder of higher-order processing, not articulation
- Wernicke’s aphasia (receptive aphasia)
  - damage to Brodmann’s area 22 (temporal lobe: semantics)
  - can’t understand language, but can speak correctly
  - but not necessarily meaningfully
  - may eventually talk nonstop nonsense (because no self-monitoring?)
- Broca’s aphasia (expressive aphasia)
  - damage to Brodmann’s areas 44 & 45 (frontal lobe: motor sequencing)
  - comprehension intact, but difficulty speaking correctly: “dog → walk”
  - may eventually become mute
  - deficits in understanding syntactically complex sentences (due to FC role in temporally extended behavior?)
- LH lesions have analogous effects on signing and on reading and writing

Wernicke and Broca in Right Hemisphere?

- LH is dominant in language, so what do the corresponding areas do in RH?
- They are involved in language prosody (rhythm and intonation)
  - musical aspects
  - important in conveying meaning
- Lesion in RH Wernicke ⇒ inability to understand intonation
- Lesion in RH Broca ⇒ speaking in flat tones

Reading and Writing

- Written language is an invention, not innate
- LH involved in written language as in spoken
- LH parietal areas involved in written language
- Damage to LH Brodmann’s areas 39 & 40 (parietal) ⇒ acquired illiteracy
  - i.e., an acquired inability to understand written language

B. Reading and Dyslexia in the Triangle Model

Reading: The “Triangle Model”

Note: There is no single “lexicon,” no “word units”

Acquired Dyslexia

- Phonological: nonword (“nust”) errors
- Deep: phonological + semantic errors (“dog” → “cat”) + visual errors (“dog” → “dot”)
- Surface: exception (“yacht”) errors + visual errors + impaired semantic access
Hypotheses

- **Phonological dyslexia** — difficulty reading nonwords (“nust” or “mave”) — explained by damage to direct orthography → phonology pathway
- **Deep dyslexia** — may make semantic substitutions (“orchestra” → “symphony”) — may make visual errors (“dog” → “dot”) — significant damage to direct pathway
- **Surface dyslexia** — intact nonword reading, but access to semantics impaired — impaired pronunciation of exception words (“yacht”) — explained by damage to indirect (semantic) pathway

The dyslex Model

- 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

**Word Corpus**

- Cluster plot of semantic similarity for words used in model
- Words that are semantically close are sometimes confused for each other in simulated deep dyslexia

**Partial Direct Pathway Lesions** (with or without Semantic Pathway)

- Minor direct damage: just visual errors
- More damage: semantic errors ⇒ deep dyslexia even with full semantics
- cf. phonological and deep dyslexia

**Partial Semantic Pathway Lesions**

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

**Partial Semantic Pathway Lesions with Complete Direct Pathway Lesions**

- Multiple errors types
- More abstract semantic errors than concrete
- cf. deep dyslexia
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

C. Spelling to Sound Mapping in Word Reading

Modeling the direct pathway

Regularities & Exceptions

- 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 extremes of context dependence
- Need a range of context dependency for regulars and exceptions.

Gradual Invariance Transformation

Increasing receptive field size gives 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 information

Illustration of Spatial Invariance

I cndu't bveliee taht I culod aulaclty usdtdnm whah I was rdnueg. Unisg the icndeblire pwer of the huma mnid, aoeclrnig to seecraht @Cmbrigde Univertisy, it dseno't mtaer in watdh oderr the itterts in a wrod are, the oly iproauntnt thng is tah the frsit and lats titeer be in the rghit pelae. The rset can be a tao1 mses and you can stll raed it wouht a pboerlm. Tlhs is bcesae the huamn mnid does not raed eveyt titeer by istelf, but the wrod as a wlohe. Aaznmig, huh? Yaeh and I awlyas tghhuot skelimp was ipmorannt! See if your frdreins can raed tlhs too.
9. Language

Word Reading as Object Recognition

- Word reading modeled as spatially invariant object recognition
- V1-like: Words show up in different locations in input
- V2/V4-like: Extracts more complex combinations of letters, also developing more invariant representations
  - Integrates individual letters or multi-letter features over multiple different locations
- IT-like: Fully spatially invariant representation of the word

Reading Model (Direct Pathway)

- Detailed model of the "direct" reading pathway (ortho → phon)
- Trained to pronounce large set of regular & exception words
  - Nearly 3000 sampled according to frequency
- Generalization testing: nonwords (e.g., must)

Phonological Representations

- Same 7 slot vowel-centered representations as before:
  - Face = /fAas/
  - Grin = /gOz/
  - Star = /sArt/
  - Post = /pOz/
- Except instead of using a localist representation of each phoneme, we use a distributed representation
- This allows us to represent the fact that phonemes vary in their similarity to one another

Test Sets of Nonwords

- Glushko regulars: Nonwords constructed to match strong regularities, for example nust, which is completely regular (e.g., must, bust, trust, etc.).
- Glushko exceptions: Nonwords that have similar English exceptions and conflicting regularities, such as bint (could be like mint, but also could be like pint). We score these items either according to the predominant regularity, or also including close exceptional cases (all OK in the table).
- McCann & Besner controls: These are pseudo-homophones and matched controls, that sound like actual words, but are spelled in a novel way. For example choyce (pronounced like choice), and the matched control is phoyce.
- Taraban & McClelland: Has frequency matched regular and exception nonwords, for example poes (like high frequency words goes or does), and mose, like lower frequency pose or lose.

Nonword Performance

Regularity tests (Glushko): bint → /bint/.

Pseudo-homophones (McCann & Besner):

- Phonoyce → /pOz/, goes → /pOz/.

Matched regularity/exception cases (Taraban):

- High freq: poes → /pOz/, goes → /pOz/, does → /d’z/.
- Low freq: mose → /pOz/, poes → /pOz/, lose → /lUz/.

<table>
<thead>
<tr>
<th>Nonword Set</th>
<th>Model</th>
<th>PMSP</th>
<th>People</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glushko regulars</td>
<td>95.3</td>
<td>97.7</td>
<td>95.8</td>
</tr>
<tr>
<td>Glushko exceptions raw</td>
<td>79.0</td>
<td>72.1</td>
<td>78.3</td>
</tr>
<tr>
<td>Glushko exceptions alt OK</td>
<td>85.9</td>
<td>85.0</td>
<td>86.6</td>
</tr>
<tr>
<td>McCann &amp; Besner controls</td>
<td>92.3</td>
<td>n/a</td>
<td>94.3</td>
</tr>
<tr>
<td>Taraban &amp; McClelland</td>
<td>97.9</td>
<td>n/a</td>
<td>100.0</td>
</tr>
</tbody>
</table>

(slides based on Frank)
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.

emergent Demonstration: ss.proj

D. Latent Semantics in Word Co-occurrence

And the Emergence of Semantic Representations through the Generalized Hebbian Algorithm

How Do Words Come to Mean Anything?

- What gives words their meaning?
- Where does this meaning come from?

Distributed Semantics

Semantics is distributed across specialized processing areas.

Distributed Semantics in the Brain

- Semantics:
  - distributed in domain-specific areas
  - fundamentally embodied and grounded
- Anterior pole of temporal lobe
  - perhaps a central hub for coordinating semantics in distributed areas
  - perhaps especially important for abstract words
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)

Latent Semantic Analysis

- Let X be a matrix such that \( x_{ij} \) is the number of occurrences of term \( i \) in document \( j \)
- By singular value decomposition (SVD): \( X = U \Sigma V^T \), where U and V are orthogonal matrices
- We find weight vector \( w \)
- We can extract these with a series of neural nets
- The (left) singular vectors (columns of U) are the eigenvectors of \( X^T X \)
- Extracts the first principal component (PC)
- Recall from supplementary slides to Latent Semantic Analysis (LSA) to Latent Semantic Analysis (LSA)
- Same occurrence.
- Similar to Latent Semantic Analysis (LSA) - Correlational Semantics

Computation by Generalized Hebbian Algorithm

- Recall from supplementary slides to ch. 4 that normalized Hebbian learning extracts the first principal component (Oja’s Rule)
- The (left) singular vectors (columns of U) are the eigenvectors of \( XX^T \) (term co-occurrences) with corresponding eigenvalues \( \sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2 \)
- We can extract these with a series of neural nets \( y_j = w_j^T x \)
- Let \( x_1 = x \) and \( \Delta w_1 = \{ y_1 \} \) (Hebbian + normalization)
- We find weight vector \( w_1 = U \sigma_1^2 \) (first eigenvector of \( XX^T \)) and output variance \( \sigma_1^2 \) (first eigenvalue)
- This procedure can be repeated to extract the other principal components by a variant of the Gram-Schmidt algorithm

Computation by GHA (continued)

- To compute the 2nd PC, subtract the first PC from input and pass it to 2nd neural net
- Let \( x_2 = x - \sigma_1^2 U \) \( y_2 = w_2^T y_1 \) (modified input to train output 2)
- Therefore,
  \[
  \Delta w_2 = \{ y_2 \} (x_1 - y_2 w_2) \approx e_{\{ y_2 \}} (x_1 - y_2 w_2)
  \]
  \[
  \Delta w_2 = \{ y_2 \} (x_1 - y_2 w_2)
  \]
- In general, let \( x_i = x_{i-1} - w_{i-1} y_{i-1} \) and \( \Delta w_i = \{ y_i \} (x_i) \approx e_{\{ y_i \}} (x_i - y_i w_i) \)
- In this way, weight vectors converge to the eigenvectors and the rows of U yield semantic representations (i.e., incremental LSA)

emergent Demonstration: sem.proj

E. Syntax and Semantics in a Sentence Gestalt

How to understand language without parsing
9. Language

Sentences and Syntax

Traditional Language Analysis (used in compilers)

- **Lexical analysis:** breaks input stream into lexical units (e.g., words)
- **Syntactic analysis:** parses lexical stream, produces a syntactic structure (parse tree)
- **Semantic analysis:** “decorates” parse tree with semantic information (meanings)
- **Pragmatics:** Interpret semantics (action / effect)

Semantics Affects Syntactic Interpretation

- Time flies like an arrow.
- Fruit flies like a banana.
- The slippers were found by the nosy dog.
- The slippers were found by the sleeping dog.
- Syntax depends on semantics very deeply

The “Gestalt” Alternative

- Just try to get the gist of what the sentence is saying:
- G. W. Bush:
  “Families is where our nation finds hope, where wings take dream.”
- Does this really work? How?

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
  - demonstrate rudimentary understanding
- e.g., Who is the agent? (has to be able to do this even if agent not currently in input)

SG Toy World

- People: busdriver (adult male), teacher, (adult female), schoolgirl, pitcher (boy), adult, child, someone also used
- 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 (sphere), ball (party), bus, pitcher, fur
- Locations: kitchen, living room, shed, park
- Semantic roles: agent, action, patient, instrument, co-agent, co-patient, location, adverb, recipient
Training
- Sentences generated randomly from semantic and syntactic grammars
- 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

SG Example/Probe Sentences
- Active semantic: The schoolgirl stirred the kool-aid with a spoon. (kool-aid can only be the patient, not the agent of this sentence)
- Active syntactic: The busdriver gave the rose to the teacher. (teacher could be either patient or agent — word order syntax determines it).
- Passive semantic: The jelly was spread by the busdriver with the knife. (jelly can’t be agent, so must be patient)
- Passive syntactic: The teacher was kissed by the busdriver. (either teacher or busdriver could be agent, syntax alone determines which it is).
- Word ambiguity: The busdriver threw the ball in the park. The teacher threw the ball in the living room. (ball is ambiguous, but semantically, busdriver throws balls in park, while teacher throws balls in living room)

Sentence Gestalt Model
- Single word input (localist single-unit representations of words)
- Gestalt layer: where distributed representation of sentence meaning develops
- Context: memory for prior words and interpretations of the sentence is encoded
  - copy of Gestalt layer activation state from previous word
  - simple recurrent network (SRN)
- Role input unit is activated, then network is trained to activate appropriate response in Fillens output layer
9. Language

Gestalt Representations

sc = schoolgirl
st = stirred
ko = kool-aid
te = teacher
bu = busdriver
pi = pitcher
dr = drank
ic = iced-tea
at = ate
so = soup
st = steak

Problems with 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 7* more than soup
  Statistical model overrides reality…
  People suffer from similar biases

Rohde (2002) Model

• Uses structured semantic representations in form of slot-filler propositions
• Includes roles: agent, experiencer, goal, instrument, patient, source, theme, beneficiary, companion, location, author, possession, subtype, property, if, because, while, and although
• Departs from notion of unstructured gestalt representation of semantic meaning
• Injects externally more of what model should be developing on own
• Challenge: develop a more naturalistic way of training corresponding semantics

Embodied Semantics

• Most of our semantic representations are grounded and embodied
• The semantic representations used in language understanding are related to those used for:
  — understanding the world
  — controlling our own actions
  — understanding other’s intentions and actions
• I.e., we know what it is to be an agent, to be a patient, to be in a place, to eat, to drink, to hit, etc. etc.
• Our linguistic abilities are very much tied to our “form of life”
• To what extent can (even humanoid) robots understand human language?

An Inverted View

• Pragmatics emerges through statistical regularities in our interactions with the world, including other humans
• Semantics (meaning) emerges through statistical regularities in pragmatic interactions (including communication)
• Syntax emerges through statistical regularities in the communication of semantics

emergent Demonstration: sg.proj