

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...

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 2

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?”

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 3

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.

4/18/17 COSC 494/594 CCN (slide based on O'Reilly) 4

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 lumps 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 reposes but these colourless green ideas sleep furiously.”
- “‘Twas brillig, and the slithy toves...”
 - But each word has *some* overlap with real words..

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 5

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, ...

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 6

Distributed Representations of Words

I cnduo't bvliece taht I culod auclacty uesdtanrd waht I was rdnaieg. Unisg the icndebliire pweor of the hmuan mnid, aocdcnrig to rseecrah at Cmabrigde Uinervtisy, it dseno't mtttaer in waht oderr the lterets in a wrod are, the olny irpoamtnt tilng is taht the frsit and lsat ltteer be in the rhgit pclae. The rset can be a taotl mses and you can sitll raed it whoutit a pboerlm. Tihis is bucseae the huamn mnid deos not raed ervey ltteer by istlef, but the wrod as a wlohe. Aaznmig, huh? Yaeh and I awlyas tghhuot slelimg was ipmorant! See if yuor fdreins can raed tihis too.

4/18/17

COSC 494/594 CCN
(slide < O'Reilly)

7

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, semantics

4/18/17

COSC 494/594 CCN
(slide < Frank)

8

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”**

4/18/17

COSC 494/594 CCN
(slide < Frank)

9

Language and Thought

- Socrates (d. 399 BCE): “that which we know we must surely be able to tell” (Plato, *Laches* 190c)
- Socrates on definition:
 - ... what is that common quality, which is the same in all these cases, and which is called courage? (*Laches* 191e)
 - Well then, show me what, precisely, this ideal is, so that, with my eye on it, and using it as a standard, I can say that any action done by you or anybody else is holy if it resembles this ideal, or, if it does not, can deny that it is holy. (*Euthyphro* 6e)
 - And so of the virtues, however many and different they maybe, they all have a common nature which makes them virtues. (*Meno* 72)

4/18/17

COSC 494/594 CCN

10

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 *M* are *P*
 - S* is *M*
 - ∴ *S* is *P*
- More reliable, because more mechanical

COSC 312 — Turing Machines

11

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

4/18/17

COSC 494/594 CCN

12

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)

4/18/17

COSC 494/594 CCN

13

Thought as Calculation


- “By *rationation* 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”

4/18/17

COSC 494/594 CCN

14

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
- Turing (1912–1954): neural nets, artificial intelligence, “Turing test”

4/18/17

COSC 494/594 CCN

15

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 represented 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”?

4/18/17

COSC 494/594 CCN

16

Connectionist AI & Cognitive Science

- Knowledge is represented at a *subsymbolic* 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

4/18/17

COSC 494/594 CCN

17

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?

4/18/17

COSC 494/594 CCN

18

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

4/18/17 COSC 494/594 CCN 19

A. Biology of Language

4/18/17 COSC 494/594 CCN 20

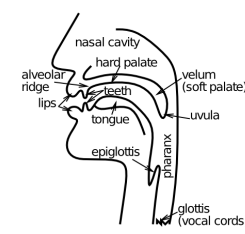
Language is Instinctual

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

4/18/17 COSC 494/594 CCN 21

Speech Output



VOWELS

	Front	Near front	Central	Near back	Back
Close	i	y	ɨ	ɯ	u
Near close	ɪ	ʏ	ɨ	ɯ	ʊ
Close mid	e	ø	ɘ	ɵ	o
Mid	ə				
Open mid	ɛ	œ	ɜ	ɞ	ɔ
Near open	æ				
Open	a	ɶ	ɚ	ɛ	ɔ

Vowels at right & left of bullets are rounded & unrounded.

VOWELS

	Front	Near front	Central	Near back	Back
Close	ɪ	ʏ			u
Near close	e	ø			o
Close mid	ɛ	œ			ɔ
Mid	ə				
Open mid	ɛ	œ			ɔ
Near open	æ				
Open	a				ɔ

Vowels at right & left of bullets are rounded & unrounded.

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 22

Speech Output

	Labial		Dental		Alveolar		Palatal		Retroflex		Palatal		Velar		Uvular		Pharyngeal		Glottal		
Billabial	m	ɱ																			
Dental			f	v	θ	ð	s	z	ʃ	ʒ											
Alveolar			p	b	t	d	ʈ	ɖ	c	ɟ	k	g	q	ɢ							
Palatal							ç	ʝ	x	ɣ	ɕ	ʝ									
Velar											ŋ	ɴ									
Uvular													ʁ								
Pharyngeal																					
Glottal																					ʔ

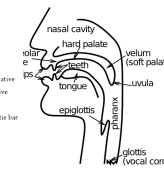
Where symbols appear in pairs, the one to the right represents a normally voiced consonant, except for murmured ɢ. Shaded areas denote articulations judged to be impossible. Light grey boxes are unusual articulations of the voicing.

CONSONANTS (NON-PULMONIC)

Glottal fricative	h
Labiodental fricative	f
Labial fricative	ɸ
Alveolar fricative	s
Alveolar fricative	z
Palatal fricative	ç
Palatal fricative	ʝ
Velar fricative	ʁ
Uvular fricative	ʀ

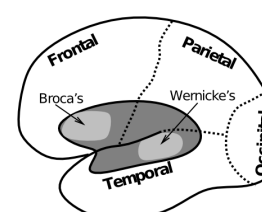
CONSONANTS (CO-ARTICULATED)

M	Voiceless labiodental velar approximant
W	Voiceless labiodental velar approximant
ɥ	Voiceless labiodental palatal approximant
ç	Voiceless palatalized labiodental fricative
ʝ	Voiceless palatalized labiodental fricative
ʁ	Voiceless palatalized labiodental fricative
ʀ	Voiceless palatalized labiodental fricative
ɕ	Simultaneous s and ʃ (sɕ)
ɟʝ	Simultaneous ʒ and ʝ (ʒʝ)



4/18/17 COSC 494/594 CCN (slide < O'Reilly) 23

Biology 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

4/18/17 COSC 494/594 CCN (slide < Frank) 24

Aphasias

- *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

4/18/17 COSC 494/594 CCN 25

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

4/18/17 COSC 494/594 CCN 26

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

4/18/17 COSC 494/594 CCN 27

B. Reading and Dyslexia in the Triangle Model

4/18/17 COSC 494/594 CCN 28

Reading: The "Triangle Model"

Note: There is no single "lexicon," no "word units"

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 29

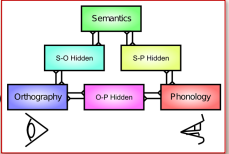
Acquired Dyslexia

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

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 30

Hypotheses

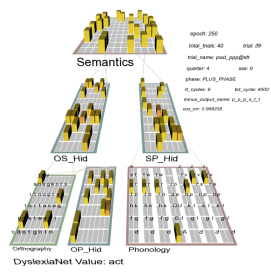
- Phonological dyslexia
 - difficulty reading nonwords (“must” or “mave”)
 - explained by damage to direct orthography → phonology pathway
- Deep dyslexia
 - may make semantic substitutions (“orchestra” → “symphony”)
 - may make visual errors (“dog” → “dot”)
 - significant damage to direct pathway
-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



4/18/17 COSC 494/594 CCN 31

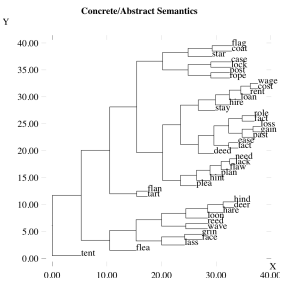
The dyslex Model

- 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



4/18/17 COSC 494/594 CCN 32

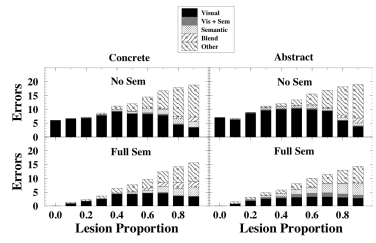
Word Corpus



- 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

4/18/17 COSC 494/594 CCN 33

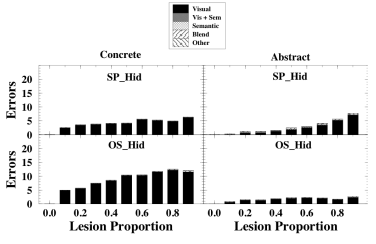
Partial Direct Pathway Lesions (with or without Semantic Pathway)



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

4/18/17 COSC 494/594 CCN (slide < Frank) 34

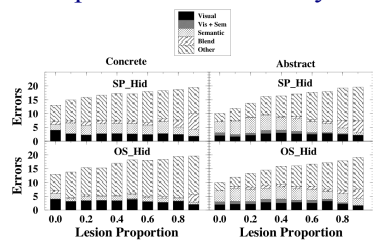
Partial Semantic Pathway Lesions



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

4/18/17 COSC 494/594 CCN (slide < Frank) 35

Partial Semantic Pathway Lesions with Complete Direct Pathway Lesions



- multiple errors types
- more abstract semantic errors than concrete
- cf. deep dyslexia

4/18/17 COSC 494/594 CCN (slide < Frank) 36

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

4/18/17

COSC 494/594 CCN
(slide < Frank)

37

emergent Demonstration dyslex.proj

4/18/17

COSC 494/594 CCN

38

C. Spelling to Sound Mapping in Word Reading

Modeling the direct pathway

4/18/17

COSC 494/594 CCN

39

Regularities & Exceptions

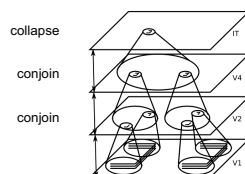
- Regularities in pronunciation are often partial, context dependent: bint
- i in mint, hint, stint,... (regular) vs. pint (exception)
- but also: mind, find, hind,... (regular) mine, fine, dine,... (regular)
- Pronunciation depends on context
- Exceptions are extremes of context dependence
- Need a range of context dependency for regulars and exceptions.

4/18/17

COSC 494/594 CCN
(slide < Frank)

40

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

4/18/17

COSC 494/594 CCN
(slide < Frank)

41

Illustration of Spatial Invariance

I cnduo't bveleee taht I culod aulaclyt uesdtannrd waht I was rdnaieg. Unisg the icndeblire pweor of the hmuam mnid, aocdcnrig to rseeerah at Cmabrigde Uinervtisy, it dseno't mtttaer in waht oderr the lterets in a wrod are, the olny irpoamtnt tihng is taht the frsit and lsat ltteer be in the rhgit pclae. The rset can be a taotl mses and you can sitll raed it whoutit a pboerlm. Tihis is bucseae the huamn mnid deos not raed ervey ltteer by istlef, but the wrod as a wlohe. Aaznmig, huh? Yaeh and I awlyas tghhuot slelimg was ipmorantt! See if yuor fdreins can raed tihis too.

4/18/17

COSC 494/594 CCN

42

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

4/18/17 COSC 494/594 CCN 43

Reading Model (Direct Pathway)

- Detailed model of the “direct” reading pathway (ortho → phono)
- Trained to pronounce large set of regular & exception words
 - nearly 3000 sampled according to frequency
- Generalization testing: nonwords (e.g., nust)

4/18/17 COSC 494/594 CCN (slide based on Frank) 44

Phonological Representations

- Same 7 slot vowel-centered representations as before:
 - face = fffAsss
 - grin = grrinnn
 - star = starr
 - post = pppOstt
- 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

4/18/17 COSC 494/594 CCN (slide < Frank) 45

ss Network Structure

- Ortho**: 7 slots with 27 letter units
 - words shift to right for spatial invariance
- OrthoCode**: receptive fields are 3 Ortho slots
 - learns groups of 1-3 letters
- Hidden**: spatial invariance + increased complexity
 - encodes context-sensitive regularities
- Phon**: 7 slots centered on vowel

4/18/17 COSC 494/594 CCN 46

Test Sets of Nonwords

- Glushko regulars**: nonwords constructed to match strong regularities, for example must, 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 (alt OK in the table).
- McCann & Besner ctrls**: 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 poses (like high frequency words goes or does), and mose, like lower frequency pose or lose.

4/18/17 COSC 494/594 CCN 47

Nonword Performance

Regularity tests (Glushko): bint → /bint/

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

Matched regularity/exception cases (Taraban):
 High freq: poses → /pOz/, goes → /gOz/, does → /dʒz/
 Low freq: mose → /pOs/, poses → /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

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 48

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

4/18/17 COSC 494/594 CCN (slide < Frank) 49

emergent Demonstration: ss.proj

4/18/17 COSC 494/594 CCN 50

D. Latent Semantics in Word Co-occurrence

And the Emergence of Semantic Representations through the Generalized Hebbian Algorithm

4/18/17 COSC 494/594 CCN 51

How Do Words Come to Mean Anything?

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

4/18/17 COSC 494/594 CCN (slide < Frank) 52

Distributed Semantics

Semantics is distributed across specialized processing areas

4/18/17 COSC 494/594 CCN (slide < Frank) 53

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

4/18/17 COSC 494/594 CCN 54

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)

4/18/17 COSC 494/594 CCN (slide < Frank) 55

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 and Σ is diagonal: $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$, where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$ are the *singular values*
- Compute rank k approximation of X by keeping the largest k singular values:

$$X_k = U_k \Sigma_k V_k^T$$
 (that is, factor X through a lower-dimensional semantic space: a bottleneck to extract meaningful dependencies)
- The rows \mathbf{t}_i^T of U_k are k -dimensional representations of the terms (and the columns of V_k^T are k -dimensional representations of the documents)
- Terms p and q can be compared by the inner product of $\Sigma_k \mathbf{t}_p$ and $\Sigma_k \mathbf{t}_q$ (cosine similarity)

4/18/17 COSC 494/594 CCN 56

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 $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$
- We can extract these with a series of neural nets $y_i = \mathbf{w}_i^T \mathbf{x}_i$
- Let $\mathbf{x}_1 = \mathbf{x}$ and $\Delta \mathbf{w}_1 = \epsilon [y_1 \mathbf{x}_1 - y_1 \mathbf{w}_1 y_1]$ (Hebbian + normalization)
- We find weight vector $\mathbf{w}_1 \rightarrow U_{\cdot 1}$ (1st eigenvector of $E\{\mathbf{x}\mathbf{x}^T\}$) and output variance $\langle y_1^2 \rangle \rightarrow \sigma_1^2$ (1st eigenvalue)
- This procedure can be repeated to extract the other principal components by a variant of the Gram-Schmidt algorithm

4/18/17 COSC 494/594 CCN 57

Computation by GHA (continued)

- To compute the 2nd PC, subtract the first PC from input and pass it to 2nd neural net
- Let $\mathbf{x}_2 = \mathbf{x} - \sigma_1^2 U_{\cdot 1} \approx \mathbf{x}_1 - \mathbf{w}_1 y_1$ (modified input to train output 2)
- Therefore,

$$\begin{aligned} \Delta \mathbf{w}_2 &= \epsilon [y_2 \mathbf{x}_2 - y_2 \mathbf{w}_2 y_2] \\ &= \epsilon [y_2 (\mathbf{x}_1 - \mathbf{w}_1 y_1) - y_2 \mathbf{w}_2 y_2] \\ &= \epsilon [y_2 \mathbf{x} - y_2 (\mathbf{w}_1 y_1 + \mathbf{w}_2 y_2)] \end{aligned}$$
- In general, let $\mathbf{x}_i = \mathbf{x}_{i-1} - \mathbf{w}_{i-1} y_{i-1}$ and

$$\Delta \mathbf{w}_i = \epsilon [y_i \mathbf{x}_i - y_i \mathbf{w}_i y_i] = \epsilon \left[y_i \mathbf{x} - y_i \sum_{k=1}^{i-1} \mathbf{w}_k y_k \right]$$
- In this way, weight vectors converge to the eigenvectors and the rows of U yield semantic representations (i.e., incremental LSA)

4/18/17 COSC 494/594 CCN 58

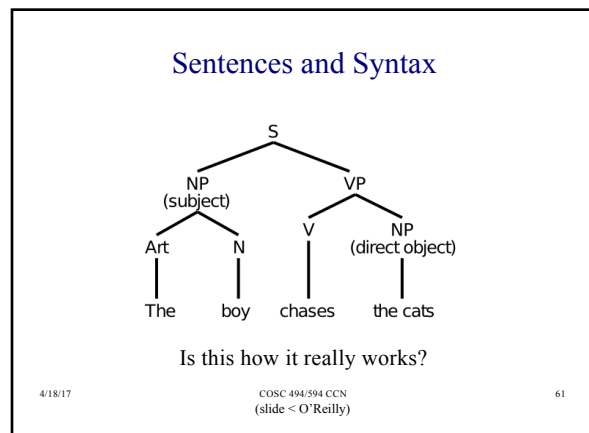
emergent Demonstration: sem.proj

4/18/17 COSC 494/594 CCN 59

E. Syntax and Semantics in a Sentence Gestalt

How to understand language without parsing

4/18/17 COSC 494/594 CCN 60



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)

4/18/17 COSC 494/594 CCN 62

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

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 63

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?

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 64

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

4/18/17 COSC 494/594 CCN (slide < Frank) 65

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*

4/18/17 COSC 494/594 CCN (slide based on O'Reilly) 66

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

4/18/17 COSC 494/594 CCN (slide based on Frank) 67

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.* vs. *The busdriver kissed the teacher.* (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)

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 68

SG Example/Probe Sentences

- Concept instantiation:** *The teacher kissed someone.* (male). (teacher always kisses a male -- has model picked up on this?)
- Role elaboration:** *The schoolgirl ate crackers.* (with finger); *The schoolgirl ate.* (soup) (these are predominant cases)
- Online update:** *The child ate soup with daintiness.* vs. *The pitcher ate soup with daintiness.* (schoolgirl usually eats soup, so ambiguous *child* is resolved as schoolgirl in first case after seeing soup, but specific input of *pitcher* in second case prevents this updating).
- Conflict:** *The adult drank iced-tea in the kitchen.* (living-room) (iced-tea is always had in the living room).

4/18/17 COSC 494/594 CCN (slide < O'Reilly) 69

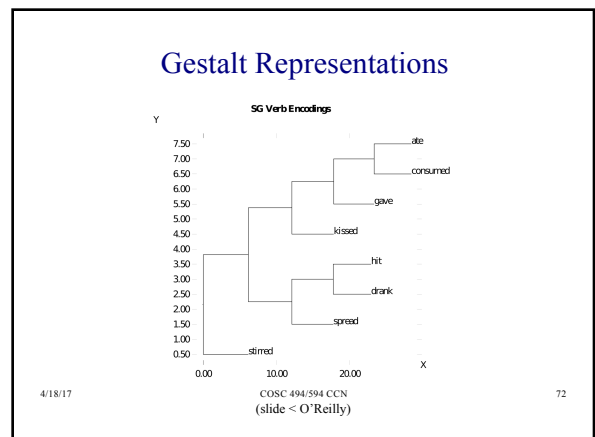
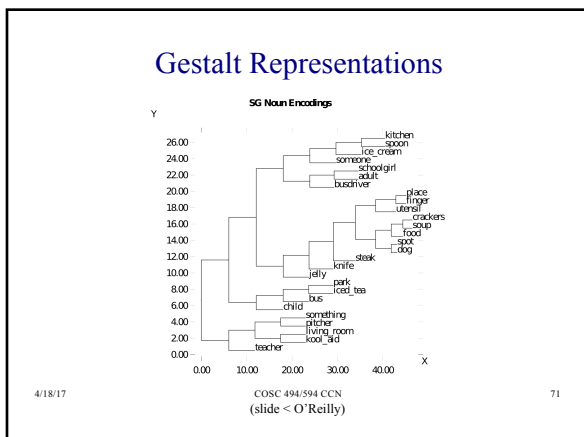
Sentence Gestalt Model

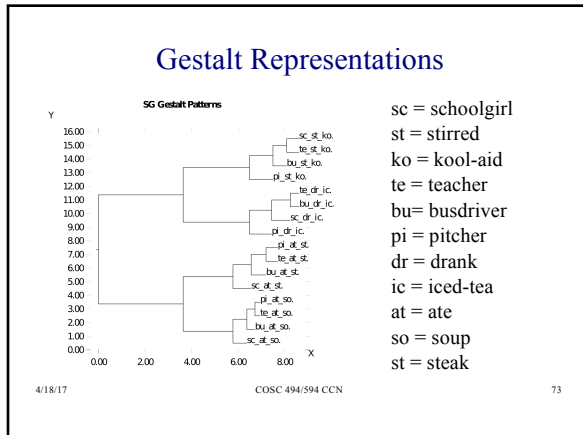
The diagram illustrates the Sentence Gestalt Model architecture. It consists of several layers:

- Input:** The starting point for the model.
- Role Filler:** A layer that processes the input.
- Encode:** A layer that converts the input into a Gestalt representation.
- Decode:** A layer that reconstructs the input from the Gestalt representation.
- Gestalt:** A layer representing the current state of the sentence.
- Gestalt Context:** A layer representing the context of the sentence.

- Single word Input (localist single-unit representations of words)
- Gestalt layer: where distributed representation of sentence meaning develops
- Context: memory for prior words and meaning 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 Filler output layer

4/18/17 COSC 494/594 CCN 70





- ### 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
- 4/18/17 COSC 494/594 CCN (slide < Frank) 74

- ### 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
- 4/18/17 COSC 494/594 CCN 75

- ### 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?
- 4/18/17 COSC 494/594 CCN 76

- ### 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
- 4/18/17 COSC 494/594 CCN 77

emergent Demonstration: sg.proj

4/18/17 COSC 494/594 CCN 78