LING/C SC 581:
Advanced Computational Linguistics

Lecture 28

Prof. Sandiway Fong

Continuing on from last time

- So far:
  - **Einstein's Miracle Creed:**
    - there's a simple basis (*although things looks complicated*)
  - **Biology:**
    - from the outside, we ain't that special (*brain size etc*.)
  - **Disruptive event**:
    - small mutation unlocks Merge.
  - **I-Language**:
    - doesn't use pre-existing linear system, operations rely on structure
  - **Not evolved for communication**
    - Nature prioritized thought (*not well-designed for communication*)

Aside: Behaviorism and Corpus Linguistics

2 min clip

Aside: Behaviorism and Corpus Linguistics

- **Einstein's Miracle Creed:**
  - there's a simple basis (*e.g. for physical laws*)

  can't be done by (ingesting) corpora
  - LLM-style

Three Factors in Language Design

(Chomsky 2005)

- "**I-language** develops in the individual through the interaction of three factors: **(1)** biological endowment, **(2)** external data, **(3)** natural law, including general principles of computational efficiency, of crucial significance for a computational system such as language."

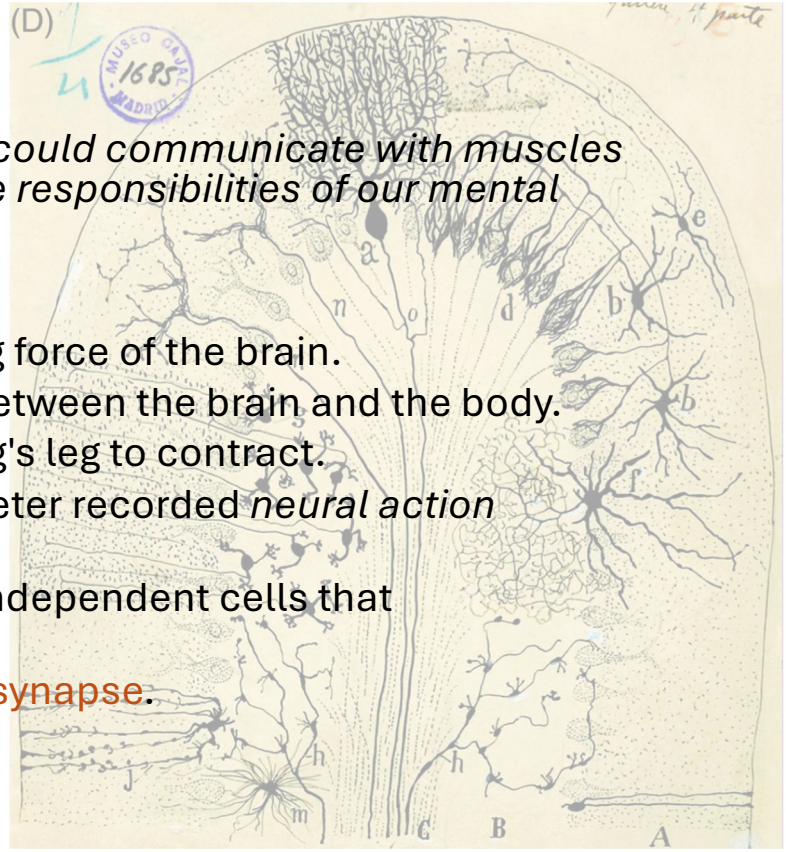
- "… accessing principles of efficient computation, understood as "third factor" principles of natural law."

Slow Brain Bottleneck

(Sotelo 2020)

> "*The first diffculty was to imagine how the brain could communicate with muscles and organs and, at the same time, take on all the responsibilities of our mental life.*"

- A timeline:
  - Aristotle: *pneuma*, or divine breath, the driving force of the brain.
  - Descartes: "*animal spirits*" flowed in nerves between the brain and the body.
  - Galvani (1791): electrical charge caused a frog's leg to contract.
  - Du Bois-Reymond (1848): sensitive galvanometer recorded *neural action potential* (voltage change).
  - Cajal (1890): nerve structures are formed by independent cells that communicate with each other.
  - Waldeyer (1891): neuron. Sherrington (1897): synapse.



Slow Brain Bottleneck

The War of the Soups and the Sparks

(Valenstein, 2005)

"*How can the electragonist explain the 0.2–0.4 millisecond delay that everyone agrees occurs at synapses?*"
(Cannon)

chemical neurotransmitters

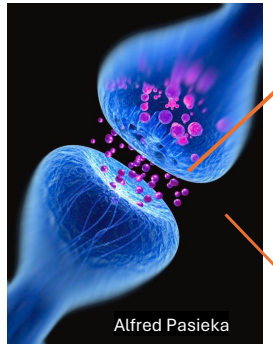
Even after the Nobel Prize was awarded to Otto Loewi and Henry Dale in 1936, most neurophysiologists did not accept neurohumoral transmission of the nerve impulse as a general principle, although many were willing to concede that chemical transmission might be adequate for the sluggish response of visceral organs. Neurophysiologists in general were convinced that only electrical transmission is fast enough to activate skeletal muscles, and for them the possibility that nerve impulses at brain synapses might be transmitted chemically was not worth thinking about. The neurophysiologist John Eccles wrote in 1936 that the "presumed chemical nature of the synaptic transmitter in the central nervous system . . . is almost entirely based on an extrapolation from the ACh [acetylcholine] hypothesis for sympathetic ganglia" and that the evidence that this could be applicable to the central nervous system was "almost negligible."[1]

Slow Brain Bottleneck

- 3rd Factor (*computational efficiency*)
- Sensor/brain mismatch
  - slow brain limits what sensory inputs can be analyzed
  - mid-20th C: chemical neural communication, not electrical
    - *The War of Soups and Sparks* (Valenstein, 2005)
- Example: can we "see" a single photon?
  - *Adaptation: **neural filters** only allow a signal to pass to the brain to trigger a conscious response when at least 5-9 arrive within < 100 ms.* (Gibbs 1996)
  - **Yes** (Tinsley *et al*. 2016) vs. **No** ~5-7 (Hecht, Schlaer & Pirenne 1942)
  - *Single photon priming effect* (peak ~3.5 secs)

neurotransmitters

predates evolution of neurons



20-40nm synaptic cleft between two neurons

Human Eye

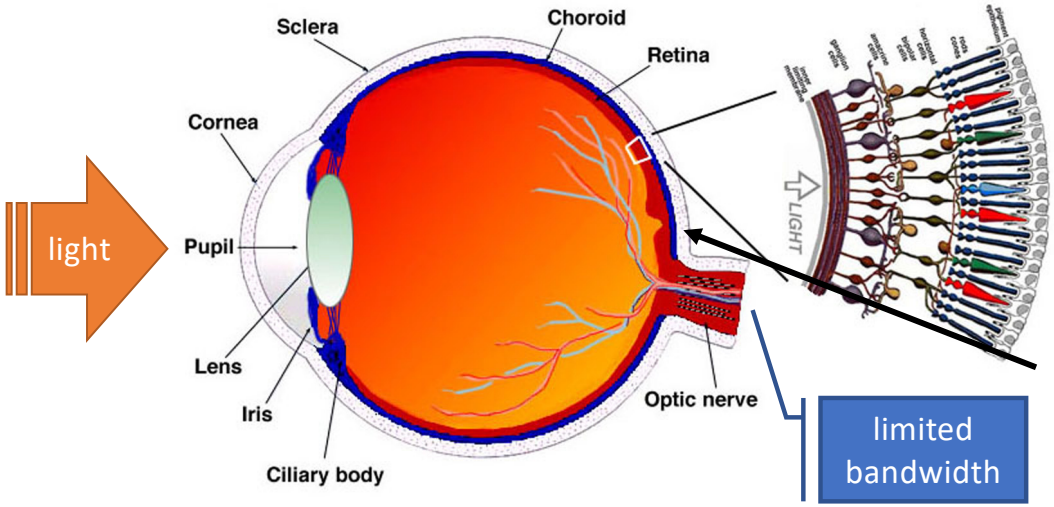


Fig. 1.1. A drawing of a section through the human eye with a schematic enlargement of the retina.

webvision.med.utah.edu/

Non-optimal design: inverted eye (vertebrates)

- cf. cuttlefish (sensors in front, no blind spot)
- retina (0.5mm thick): 90% of the area is peripheral vision; 50% of the nerve fibers
- saccadic movement

- fovea pit (0.2mm diameter)
- tightly-packed cones (20/20 vision; color)
- 50% of the nerve fibers

Sensor/Brain Mismatch

- Sensor performance is incredibly good:
  - **sensitivity** down to the single photon level (Tinsley *et al.*, 2016)
  - **resolution** (77 cycles/degree) (Curcio et al. 1990)
    - *within a factor of 3 of an eagle's*
  - we don't need/use it for scene analysis
- Other sensors:
  - **olfactory thresholds** at parts per billion (ppb) (Wackermannová et al., 2016)
  - **eardrums** can detect tiny vibrations: down to the size of a hydrogen atom (Fletcher & Munson, 1933)
  - brain doesn't need/use this either

source: Reuters

Avians: same architecture but *souped-up*:

- one/two foveae (*also denser*)
- higher flicker fusion rates
- more types of cones

Sensor/Brain Mismatch

- Language is an organic system
  - The brain is highly pre-wired in parts (*evolution*)
  - A slow brain (*mostly chemical*) constrains what is possible
  - **Computational efficiency** and **bandwidth** are important considerations for all organic subsystems
  - We (selectively) throw out almost all of the signal

Contrasting Models of Language

- **Chomsky**: part out
  - **hierarchy** (core syntax) using simplest conceivable operations
    - e.g. memoryless Binary Merge : BM(*A*, *B*) replaces *A* and *B* with {*A*, *B*} *and nothing more*
    - ideally unparameterized (*universal*)
  - **linear order** relegated to externalization to sensorimotor systems
    - a locus of language variation
    - automatically acquired by the language learner
- **General Purpose Deep Neural Networks:**
  - inspired by biology: artificial neurons with connections to other neurons
  - **deep** means stacked in multiple layers
  - **machine learning**: data intensive and compute intensive (*made recently practical*)
  - e.g. GPT-2 48 layer Transformer Neural Net Model with **1.5 billion parameters**.

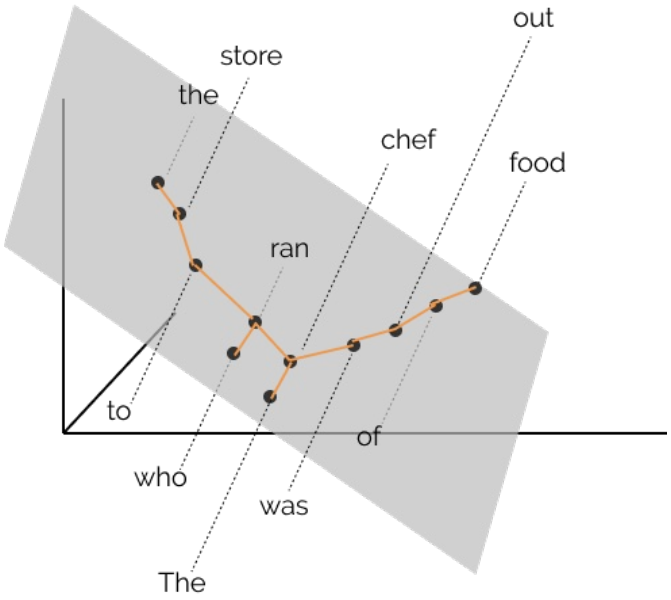
Syntax Distance Hypothesis

Words are contextual vectors (ELMo; BERT):

The	chef	who	ran	to	the	store	was	out	of	food
.4 -.2 .3	.1 .9 -.2	.3 -.4 .2	.7 -.4 0	.4 0 -.5	.1 -.6 .2	.3 .1 -.6	.1 .9 -.8	.3 .1 .8	-.8 .3 -.6	0 .7 -.9

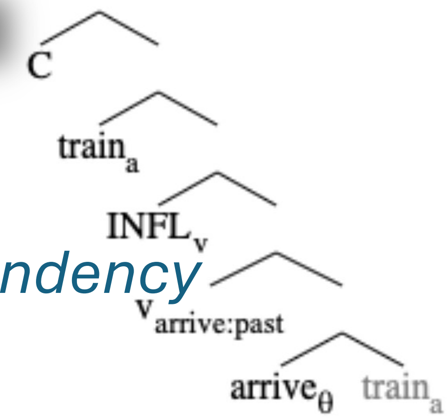
Hewitt and Manning (2019):

- **The syntax distance hypothesis**: There exists a linear transformation B of the word representation space under which vector distance encodes parse trees.



https://nlp.stanford.edu/~johnhew/structural-probe.html

INT/EXT: $\{C, \{train_a, \{INFL_v, \{v_{arrive:past}, \{arrive_\theta, train{:}a_a\}\}\}\}\}$



(Simplest) Merge

- **Basic Property** of I-Language: *structural dependency*
  - *not linear order!* (surprising: *even simpler*)
- **SMT**:
  - what is the simplest mechanism that results in structure?
- (Mathematical) **Merge**: (Marcoli et al. 2025)
  - two items X and Y (*binary*)
  - create {X, Y} (set: *no order*)
  - **recurse**: *Merge can feed Merge* (**WS**)
  - no embellishments!

non-recursive precursors?
- Core-Merge (Fujita 2014)
- Conjoin (Progovac 2015)

(Simplest) Merge

- Nature makes **Merge** available for cognition
- Natural Numbers ($\mathbb{N}$)
  - lexicon (**LEX**): h (*lexical item*: **LI**, a head h)
  - Workspace (**WS**): {h} (*suppose WS items: sets*)
  - **Merge**: {h} h
    - select WS item *X* and *Y*, a (sub-)**term of** *X*
    - create {X, Y} {{h}, h}

Internal Merge (IM)
"*simplest case*"

  - WS': {{h}, h} (*WS': state of WS after 1 Merge*)
  - WS'': {{{h}, h}, h} (h *a term of* {{h}, h} in WS')
    or {{{h}, h}, {h}} ({h} *a term of* {{h}, h} in WS')
    or {{{h_1}, h_2}, h_1} (h *a term of* {{h}, h} in WS')

Symbolic Model: Merge

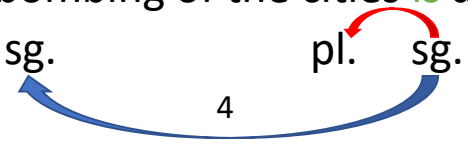
- Simplest I-language-specific primitive
  - simpler linear concatenation primitive not available
- Recursive operation (Merge *feeds* Merge)
- Markovian assumption
  - *no access to history by* Merge *or external systems*
  - compute without knowledge of the steps before
- 3rd factor: **keep object building to a minimum**
- Two conceptually necessary cases:
  - External Merge (**EM**): needed to encode predicate-argument structure
  - Internal Merge (**IM**): needed for discourse/information-related functions

recall example:
mechanic
carefully fix - pack

Agreement: linear model inadequate

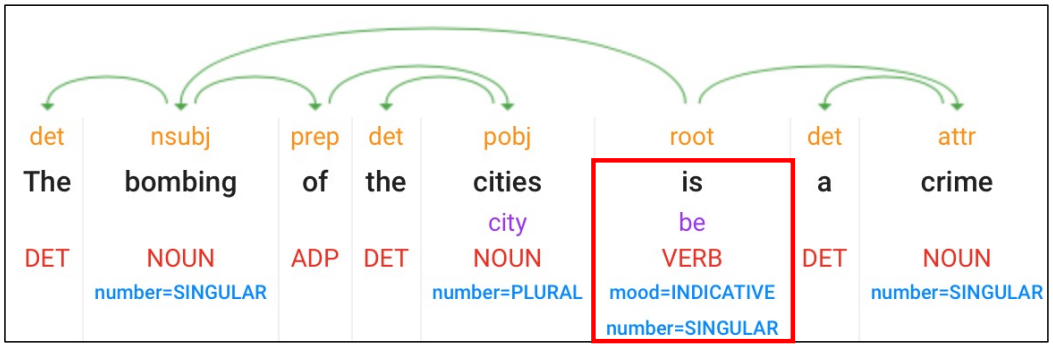
- Examples (Chomsky, *OLLI lecture*, Oct 2021):

  The bombing of the cities is a crime

  *The bombing of the cities are a crime

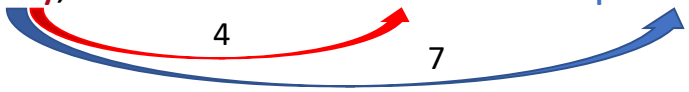
- Google Natural Language parse:



Agreement: linear model inadequate

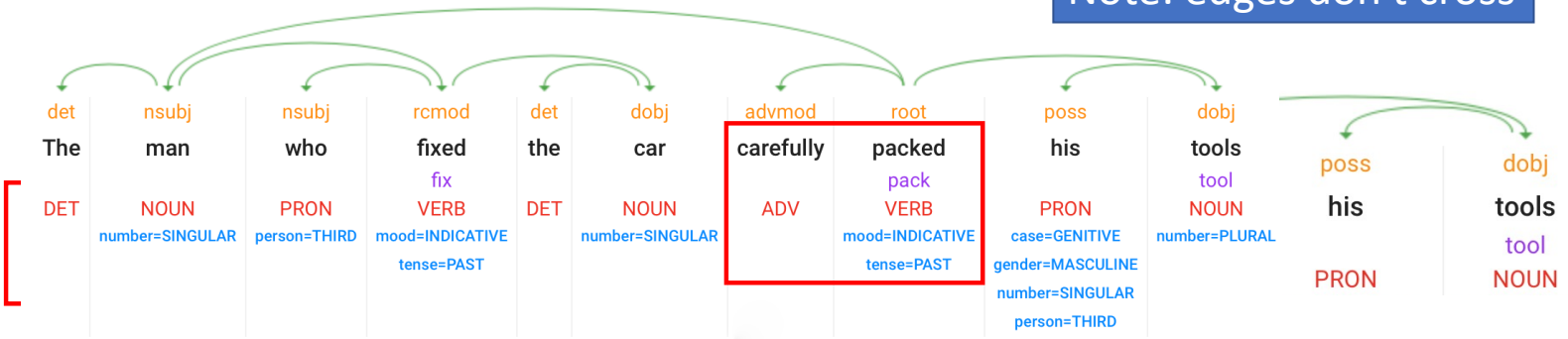
- Examples (Chomsky, *OLLI lecture*, Oct 2021):
  - The man who fixed the car carefully packed his tools

  - Carefully, the man who fixed the car packed his tools

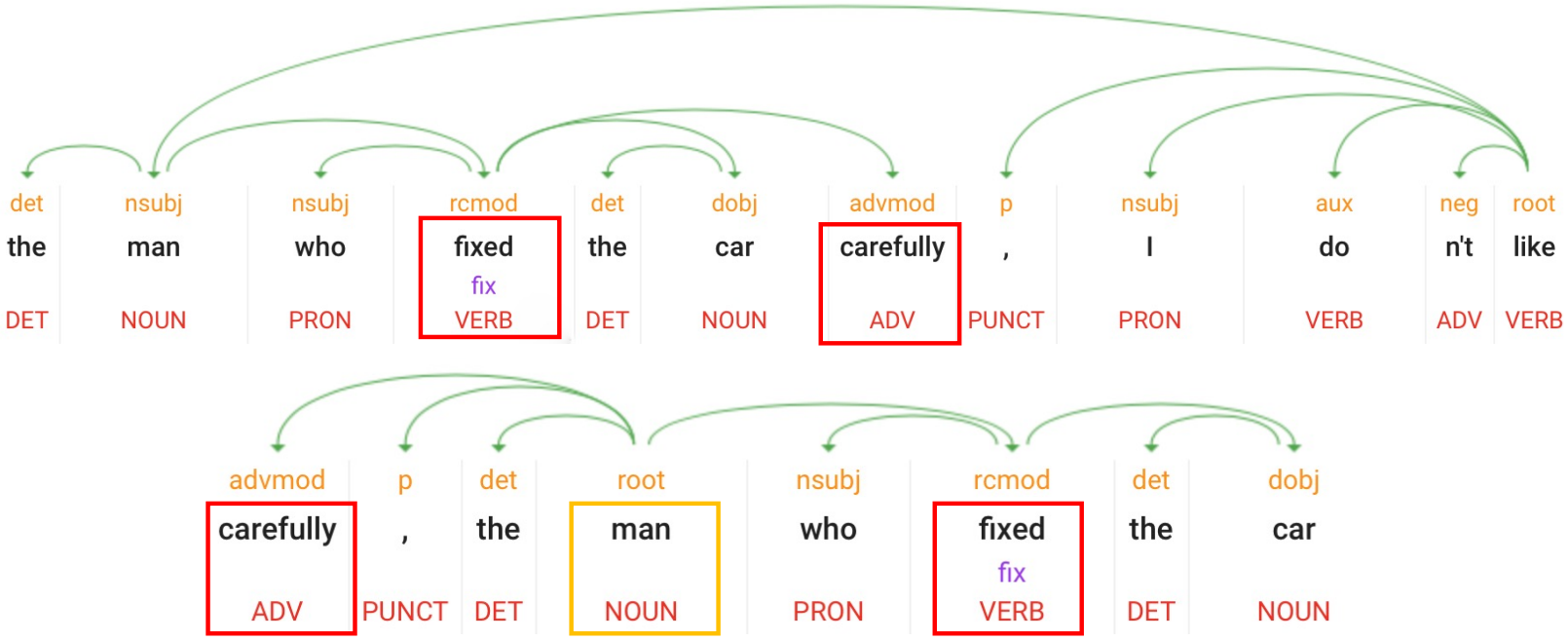


- Google Natural Language parse:

Note: edges don't cross



Agreement: linear model inadequate

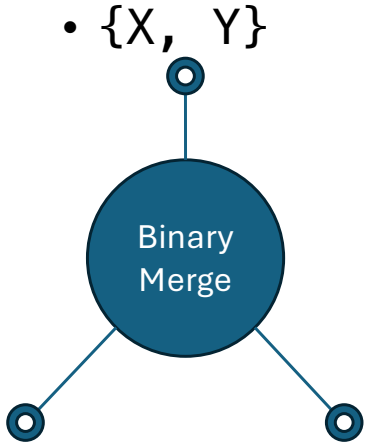


Workspace (WS)

- **EM** is more complex than **IM** (Chomsky)
- *but required by* **FL** *for thought construction*

- For I-Language
  - **LEX** contains (> 1) heads
  - **IM** is basic, *term-of* is the simplest relation
- Need also some θ-configurations:
  - {v, {R, IA_θ}} *unaccusative*
  - {EA_θ, {v, {R, IA_θ}} *transitive*
  - {EA_θ, {v, R}} *unergative*
  - assume v & R are heads (**LEX**)
  - EA & IA could be complex objects
  - *External Merge* (**EM**), brings *sister-of* relation

known to **INT**



Note: *assume* X *&* Y *are distinct, i.e. can't draw the same item twice*