Computational Modeling and Human Language

Sandiway Fong
University of Arizona

LING 581 and 506 guest lecture. April 19th and 21st 2022.

Two Lectures

- Part 1: Background (*Today*)
- Part 2: Combinatorics (*Thursday*)

Why Computation?

- Recent work by Chomsky has highlighted the role of *computational efficiency* in **UG**, a theory of **I-Language**.
- References:
  - 2021: ***Minimalism***: *Where Are We Now, and Where Can We Hope to Go***.** *Gengo Kenkyu*)**160**: 1–41.
    - https://www.jstage.jst.go.jp/article/gengo/160/0/160_1/_article/-char/en
  - 2022: ***Genuine Explanation*** *and the* ***Strong Minimalist Thesis***. MIT Linguistics Colloquium video.
    - https://www.dropbox.com/s/npoxkqezudbeh9i/Chomsky_Lecture_20220401.mp4

Why Computation?

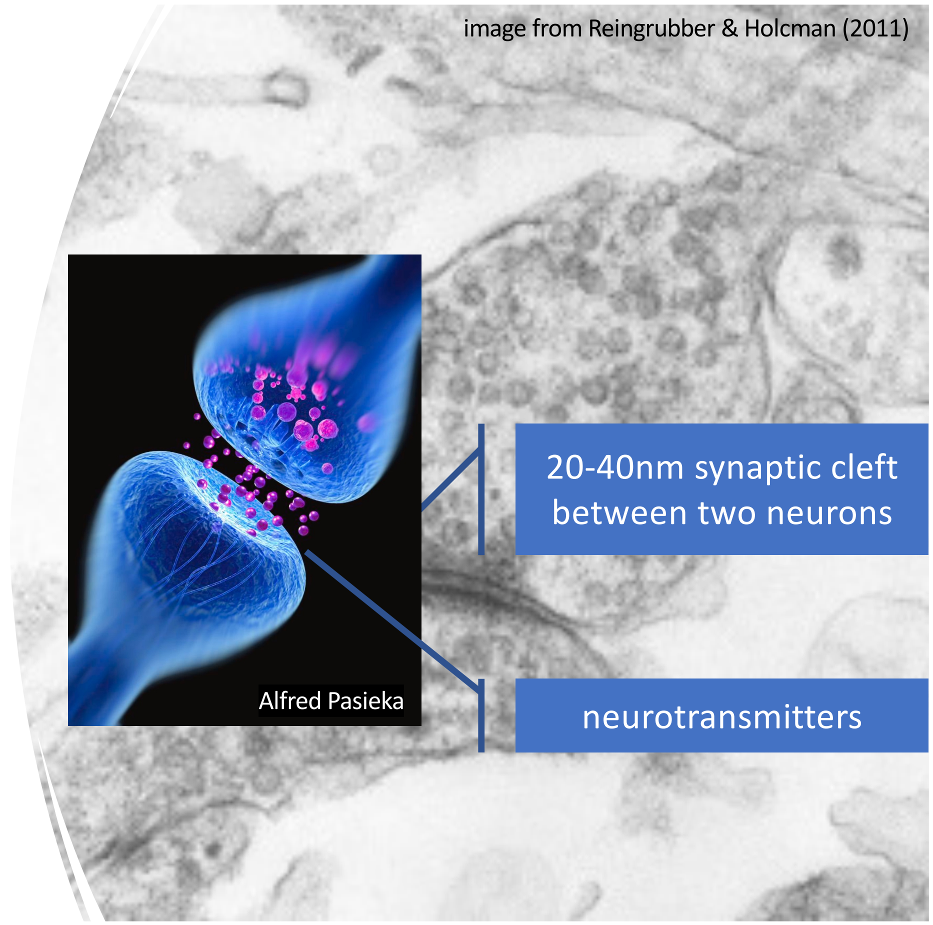
- Recent work by Chomsky has highlighted the role of *computational efficiency* in **UG**, a theory of **I-Language**.
- References:
  - 2021: ***Minimalism***: *Where Are We Now, and Where Can We Hope to Go*. *Gengo Kenkyu*)**160**: 1–41.
    - https://www.jstage.jst.go.jp/article/gengo/160/0/160_1/_article/-char/en
  - 2022: ***Genuine Explanation*** *and the* ***Strong Minimalist Thesis***. MIT Linguistics Colloquium video.
    - https://www.dropbox.com/s/npoxkqezudbeh9i/Chomsky_Lecture_20220401.mp4

Background

- **FL**: Faculty of Language
  - Biological organ: part of the brain, part of our innate endowment, enables acquisition and use.
  - Inner workings inaccessible to consciousness.
- **I-Language**
  - Internal language of thought, optionally externalized.
- **UG**: Universal Grammar
  - A generative theory of I-Language
- **Genuine Explanation**:
  - UG must be (A) descriptively adequate, (B) explain how language is learnable (given **POS**), and (C) plausibly evolved.
- **SMT**: Strong Minimalist Thesis
  - Nature came up with an optimal solution (to I-Language)
- Computational efficiency:
  - the human brain is slow
  - UG makes efficient use of limited resources

Brain is slow

- **Computational efficiency** (and **bandwidth**) are important considerations for all organic systems:
  - our sensory apparatus can generate vast amounts of data
  - a slow (*chemical*) brain limits what can be analyzed
  - we (selectively) throw out/ignore almost all of the signal



source: Reuters

Human Eye

- 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 it for scene analysis
- Other sensors:
  - **olfactory thresholds** at parts per billion (ppb) (Wackermannová et al., 2016)
  - **eardrums** can detect tiny vibrations: hydrogen atom size (Fletcher & Munson, 1933)
  - brain doesn't need this either

90% of the area is peripheral vision: 50% of the nerve fibers

fovea pit (0.2mm diameter) 20/20 vision, color: 50%

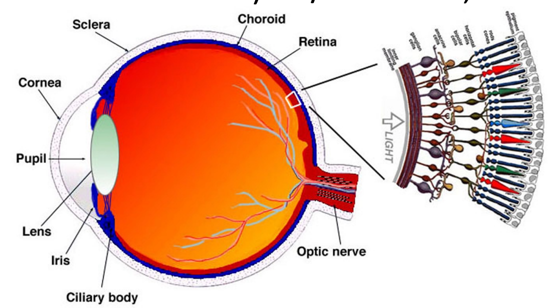
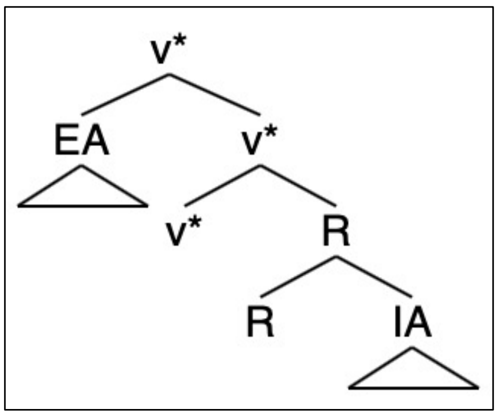


Fig. 1.1. A drawing of a section through the human eye with a schematic enlargement of the retina.

A Formal Model of I-Language

- Phrases in UG are represented as sets.
- Sets are the simplest formal objects for representing **unordered** hierarchical phrase structure.
  - a set can contain sets as members.



Set representation:

`{EA, {v∗, {R, IA}}}`

- v* and R, heads, come from LEX

Minimalist representation:

- **Inclusiveness**
- no labels (*and no room for labels*)
- no room for syntactic features (*too strong*?)

Use of Set Theory

- Limited use of set theory for representation:
  - no {}
  - no infinite sets
  - no union, intersection, difference, powerset operations
  - no equality
  - not a set, perhaps a multi-set
  - naïve set theory paradoxes seem not to apply
    - $\{x: P(x)\}$, P some property
    - Russell's Paradox: suppose $P(x)$ is $x \notin x$
    - *Barber Paradox*

Use of Set Theory

UG: sub-phrases are **unordered**

- However, set theory allows for the construction of **ordered** elements too
- Define $x < y$ `iff` $x \in y$
- **Example**:
  - `<a, b> = {a, {a, b}}`
  - `<a, b, c> = <a, <b, c>> = {a, {a, {b, {b, c}}}}`
- **Example**:
  - ordinals `{}, {{}}, {{}, {{}}},` *and so on* …
  - more readably as $\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}$ ($\emptyset$ = empty set)
  - generally, $\mathit{successor}(x) = x \cup \{x\}$

Use of Set Theory

- **Question:**
  - does anything block the formation of ordered elements in this model?
- **Note 1**:
  - ordinal operation $x \cup \{x\}$ not formulable under **Simplest Merge**
  - e.g. {Ø, {Ø}, {Ø, {Ø}}, {Ø, {Ø}, {Ø, {Ø}}}} cannot be constructed
  - the above expression maps to 4 in $\mathbb{N}$.

  Why?
  - Ø = 0
  - {Ø} = 1
  - {Ø, {Ø}} = 2
  - {Ø, {Ø}, {Ø, {Ø}}} = 3

Use of Set Theory

- **Question:**
  - does anything block the formation of ordered elements in this model?
- **Note 2**:
  - the sequence <a, b, c> is represented by the set {a, {a, {b, {b, c}}}}

1) this set can be formed by iterated External Merge (EM) from WS:

   a a b b c

   b c $\Rightarrow_{EM}$ {b, c} $\Rightarrow_{EM}$ {b, {b, c}} $\Rightarrow_{EM}$ {a, {b, {b, c}}} $\Rightarrow_{EM}$ {a, {a, {b, {b, c}}}}

or

2) this set can also be formed from a minimal WS:

   a b c

   b c $\Rightarrow_{EM}$ {b, c} $\Rightarrow_{IM}$ {b, {b, c}} $\Rightarrow_{EM}$ {a, {b, {b, c}}} $\Rightarrow_{IM}$ {a, {a, {b, {b, c}}}}

  - do these violate SMT?
  - I-language builds thoughts only (*theta configurations + other projections*)

Use of Set Theory

wrt. phrases as sets,

- we use $\in$, immediate dominance
- we also employ a special constructor, **Binary Merge**, takes two sets *x* and *y* from a **Workspace** (WS) and creates {*x*, *y*}
  - reminiscent of the ZF Axiom of the Unordered Pair: given sets x, y, form {x, y}
- also have another operation, **Minimal Search** (MS), on sets
- Binary Merge is recursive, i.e. Merge feeds Merge.

Binary Merge gives us the **Basic Property** (BP) of language

- *capable of forming an infinite number of structured thought objects* (*from LEX*)

Research Question:

- what else will we need?

Use of Set Theory

- **Workspace (WS)**: a collection of sets
  - we use ∈, member of the WS
  - each set represents a phrase or item (LI) from LEX
  - Note: WS itself is set-like, like a phrase, but not a phrase
  - WS can contain repetitions of identical inscriptions
  - formally, a LI *x* can be treated as singleton set {*x*}, or the collection could be expanded to include ur-elements/atoms LI

Workspace Binary Merge

Binary Merge (**BM**) has two selection steps:

- 1st step: select any **WS** item *x* that's eligible
- 2nd step: select *y*, a proper sub-term of *x* (**Internal Merge**), or
- select a **WS** item *y*, distinct from *x* (**External Merge**)
- add {x, y} to the WS

Some assumptions:

1. distinctness requirement for *x* and *y*
2. non-selected WS items carry over (**WS Preservation**)
3. selected WS items vanish (**Markovian assumption** perhaps)
4. involves only the current WS (**Markovian assumption**)
5. and nothing more is permitted (**Simplest Merge**) *and no other operations*

step 1

WS: a `{b, c}` d

step 2 step 2

WS': a `{b, {b, c}}` d

or

WS': a `{d, {b, c}}`

Workspace Binary Merge

- **Markovian assumptio**n
  - no access to history of the WS
- Consequences
  - can't tell (*or make use of*) what kind of Merge (IM or EM) has applied
  - **SMT Enabling Function** (M-gaps, see Chomsky ref.)
  - *can't compare previous with current state*

Operation	Previous WS	Current WS
Internal Merge (IM)	{a, b}	{a, {a, b}}
External Merge (EM)	a {a, b}	{a, {a, b}}

Workspace Binary Merge

Binary Merge

- `BM(x,y)` *only* add `{x, y}`
  - *nothing more is permitted*
- Consequences
  - can't annotate/label the new set
  - **No Tampering Condition** (NTC)
    - can't annotate or modify constituents x or y
  - IM does not form copies
  - no relation between the two *a*'s in
    - `{a, b}` $\Rightarrow_{IM}$ `{a, {a, b}}`
    - at INT, **FormCopy** optionally establishes a relation `(a, a)` between the two identical inscriptions

cf.
- move-α
- copy theory of movement

Use of Set Theory

Axiomatic Set Theory

- e.g. Zermelo-Fraenkel (ZF) Set Theory
  - axioms expressed in 1^{st} order logic + ∈ (*set membership*)
  - **Axiom of Extension**: same set if same members
    - $x \in y$ iff $x \in z$ implies $y = z$
    - doesn't necessarily hold for language
    - we have "occurrences", possibly repetitions/copies
    - **Example** (Chomsky, p.c.):
      - the man who saw **many people** didn't see **many people**
      - can't be derived using just one occurrence of *many people*
    - See also: identity of inscriptions

Workspace Binary Merge

Binary Merge (BM) has two selection steps:

- 1st step: select any WS item *x* that's eligible
- 2nd step: select *y*, a proper sub-term of *x* (**IM**), or
- select a WS item *y*, distinct from *x* (**EM**)

Note:

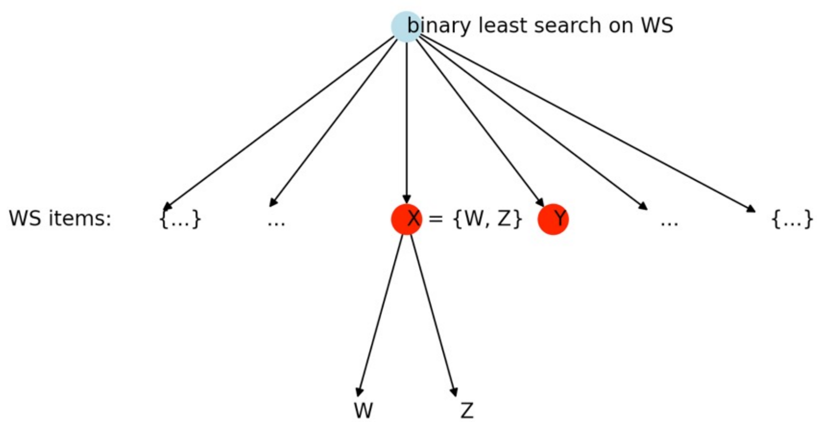
- IM is less computationally complex

Why do we have both then?

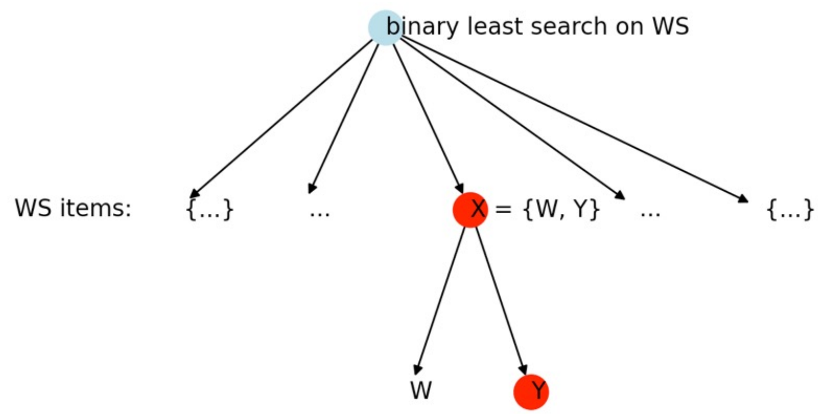
- I-Language generates thought expressions
  - **EM**: introduces theta-position content into structure
  - **IM**: displacement: *wh*-movement, raising, passivization, etc.

Workspace Binary Merge

Σ_B finds a distinct *Y* in WS + LEX



Σ_B **continues** the search within *X.*
Secondary search finds a distinct *Y*.



Workspace Binary Merge

- **Modern humans** emerged in Africa (200,000–300,000 years ago)
  - genetic change: mutation
  - rich evidence of symbolic behavior
- **Simplest Merge** meets evolutionary plausibility
  - *a criterion of Genuine Explanation*
  - simplest mechanism for construction of thought
  - its appearance could coincide with that small evolutionary step
- Question:
  - might IM be an optimization of EM?

Theta theory

- **Duality of Semantics**
  - strict separation of tasks between EM and IM
  - Theta positions populated by EM exclusively
  - IM cannot *move* something into a theta position
  - a **Language-specific Condition** (LSC)
- Consequence
  - Theta theory must be built into Merge
  - *Theta positions must be detectable throughout the Merge derivation*
- Question
  - does this violate the Markovian assumption or the **NTC**?

WS: $\texttt{R IA} \Rightarrow_{EM} \{\texttt{R, IA}_\theta\}$

WS: $\texttt{EA } \{\texttt{v*, }\{\texttt{R, IA}_\theta\}\} \Rightarrow_{EM}$
$\{\texttt{EA}_\theta\texttt{, }\{\texttt{v*, }\{\texttt{R, IA}_\theta\}\}\}$

WS: $\{\texttt{v*, }\{\texttt{R, IA}_\theta\}\} \Rightarrow_{IM}$
$\{\texttt{IA}_\theta\texttt{, }\{\texttt{v*, }\{\texttt{R, IA}_\theta\}\}\}$

Non-simplest Merge

SMT Condition

- **Minimal Yield** (MY):
  - no operation can increase the number of selectable items by more than the bare minimum, one.
- **Sideways Merge** variant (Nunes & Hornstein):
  - `{X, Y}, Z, …` ⇒ `{X, Y}, {X, Z}, …` (out: +2)
- **Parallel Merge** variant (Citko):
  - `X, Y, Z` ⇒ `{X, Y}, {X, Z}, …` (out: +2)
- **Merge + Deletion**:
  - `{A, B}, C, D, …` ⇒ `{A, B}, {A, C}, …` (still out: +1)

Non-simplest Merge

- Simplest Binary Merge imposes the least burden on computational resources. Other variants are non-optimal.
- Actually, **Parallel Merge**
  - `X, Y, Z` ⇒ `{X, Y}, {X, Z}, …`

can be simulated using **Sideways Merge** (SiM):

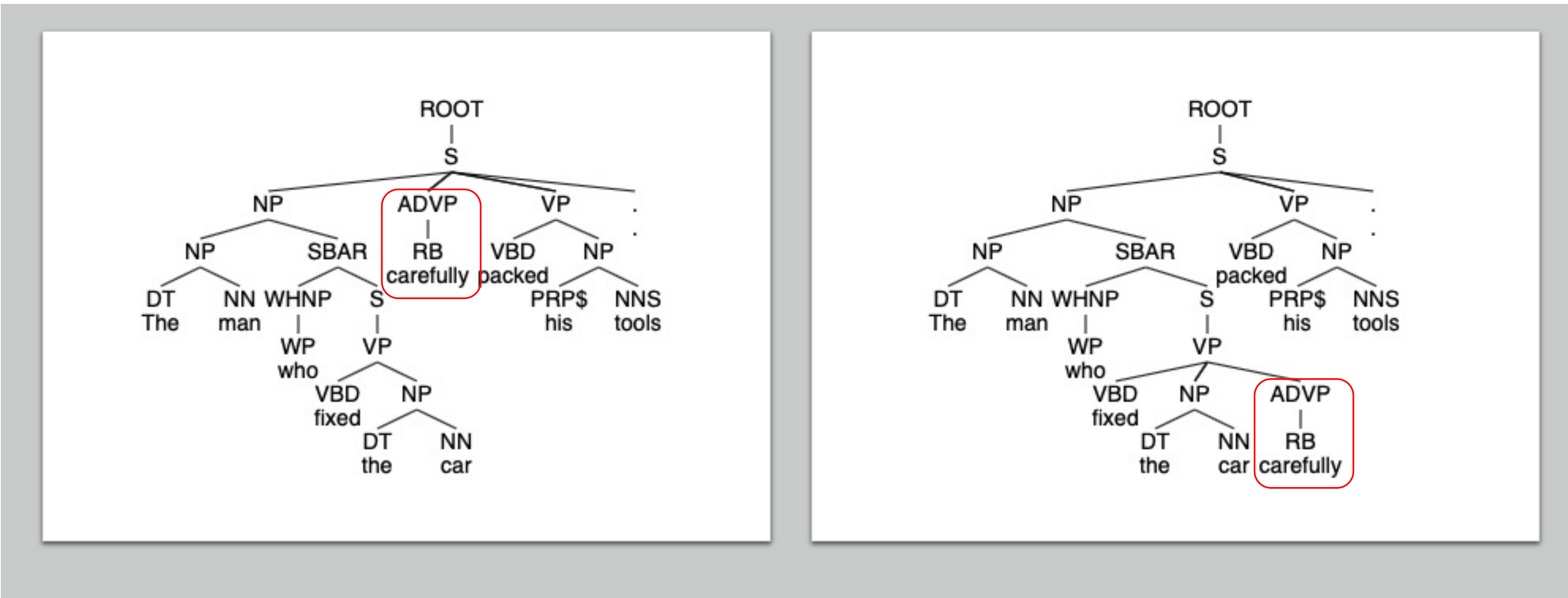
  - WS: `X, Y, Z, …`
  - EM: `{X, Y}, Z, …`
  - SiM: `{X, Y}, {X, Z}, …`

So, if we rule out SiM, we rule out Parallel Merge too.

WS Computation

- Example (*Chomsky refs.*):
  an adverb has to find a VP to modify
  - *the man who fixed the car carefully packed his tools*

Parses #3 and #1 from kbest standalone Stanford Parser



WS Computation

- Example (*Chomsky refs.*):
  - *the man who fixed the car carefully packed his tools*
- WS computation:
  - WS: {v*, {fix, the car}} {v*, {packed, his tools}} carefully
  - WS: {{v*, {fix, the car}}, carefully} WS:{{v*, {packed, his tools}}, carefully}
  - {carefully, {… {v*, {packed his tools}}, carefully}…}}

the man who carefully fixed the car … the man who fixed the car carefully packed …

the man who fixed the car carefully … the man who fixed the car packed his tools carefully

Carefully, the man who fixed the car packed …

WS Computation

Example:

- *He saw him*
- WS: `he, see, v*, he, INFL, C`

he   see: *ϕ-f*   *v*∗: *TNS*   he   INFL: *ϕ-features*   C: *force*

{see, he}

{*v*∗, {see, he}}

{he, {*v*∗, {see, he}}}

theta configuration

{INFL, {he, {*v*∗, {see, he}}}}

Agreement:
c-command

{he, {INFL, {he, {*v*∗, {see, he}}}}}

{C, {he, {INFL, {he, {*v*∗, {see, he}}}}}}

WS Computation

Example:

- *He saw him*
- WS: `he, see, v*, he, INFL, C`

- Form Copy (FC) applies @ INT (Phase level):
  - via c-command (Minimal Search)
  - **{**he, {*v*∗, {see, he}}**}**
  - {C, {he, {INFL, {he, {*v*∗, {see, he}}}}}**}**
  - construct relation (he, he)
  - Ruled out by Theta Theory
  - would be pronounced as **he saw*
- Form Copy (FC) *doesn't* apply:
  - no relation between the two *he*'s
  - pronounced as *he saw him*

Appendix

Russell's Paradox

Axiom

- there exists a set y, for all x, $x \in y$ iff P(x)

Paradox

- suppose P(x) is $x \notin x$
- then there's w st. $w \in w$ iff $w \notin w$, a contradiction

Why?

- there exists a set y, for all x, $x \in y$ iff $x \notin x$
- by existential instantiation, let that set be w
- for all x, $x \in w$ iff $x \notin x$
- by universal instantiation, $w \in w$ iff $w \notin w$
- *contradiction*

Avoid Russell's Paradox
assume a set A
$\{ x \in A \text{ st. } P(x)\}$

Barber Paradox

- **First**, some background:
  - assume all men must shave, and
  - barber is a man
- **Define**
  - barber = one who shaves all and only those who do not shave themselves
- **Derivation**
  - barber can’t shave himself as he only shaves those who don’t shave themselves.
  - but if he doesn’t shave himself, he must be shaved by the barber.
- **Saved if**
  - barber does not need to shave (e.g. recently mentioned medical condition *alopecia*)
  - no such barber exists

Axiomatic Set Theory

- Zermelo-Fraenkel (ZF) Axioms:
  - Axiom of Extensionality: same if same members – not valid for language
    - identity of inscription is the relevant notion
  - Axiom of the Unordered Pair: reminiscent of **Binary Merge** (*see later slides*)
  - Axiom of Separation: *avoids paradox*
  - Axiom of Union
  - Axiom of Powerset
  - Axiom of Infinity
  - Axiom of Replacement
  - …