

Minimalism and Computation: Where Are We Heading?

Sandiway Fong

University of Arizona

sandiway@arizona.edu

Caltech Workshop, June 8–9th 2024

Quite excited about recent theory developments and opportunities for Computational Modeling ...

- Chomsky publications mentioned:
 - LSLT: *Logical Structure of Linguistic Theory* (1955; 1975)
 - MI: *Minimalist Inquiries* (1998; 2000)
 - 3Factors: *Three Factors in Language Design* (2005)
 - GK: *Minimalism: Where Are We Now, and Where Can We Hope to Go* (2021)
 - MC: *The Miracle Creed and SMT* (to appear)

Chomsky LSLT

- **Knowledge of language** (*grammaticalness*) vs. **performance** (*acceptability*).
 - from intro to LSLT (*20 years later*)

It should be clear, then, that a generative grammar does not take the point of view of the speaker or hearer. Rather, it is a description, in the most neutral possible terms, of knowledge possessed by the speaker-hearer and put to use in normal discourse. A generative grammar is a theory of competence.

The related distinction between “acceptable” and “grammatical” utterances is also drawn in *LSLT* (cf., e.g., §100.2). We may take “acceptability” to be a concept of the theory of performance, and “grammaticalness” a related notion of the theory of competence.

LSLT

- Particular theory was
 - Transformational Generative Grammar
 - theory of competence only
 - *no theory of when we can violate principles/rules*
 - transformations were (sub-)construction-specific and overlapping
 - *"A well-known example is passive forms of ECM constructions, which seemed at one time to be generated both by raising and passive transformations."*
 - *John was believed (by Mary) to be telling the truth*
- replaced by a general theory of movement
 - Move-NP, Move-wh, ..., Move- α , ..., Merge

The Third Factor

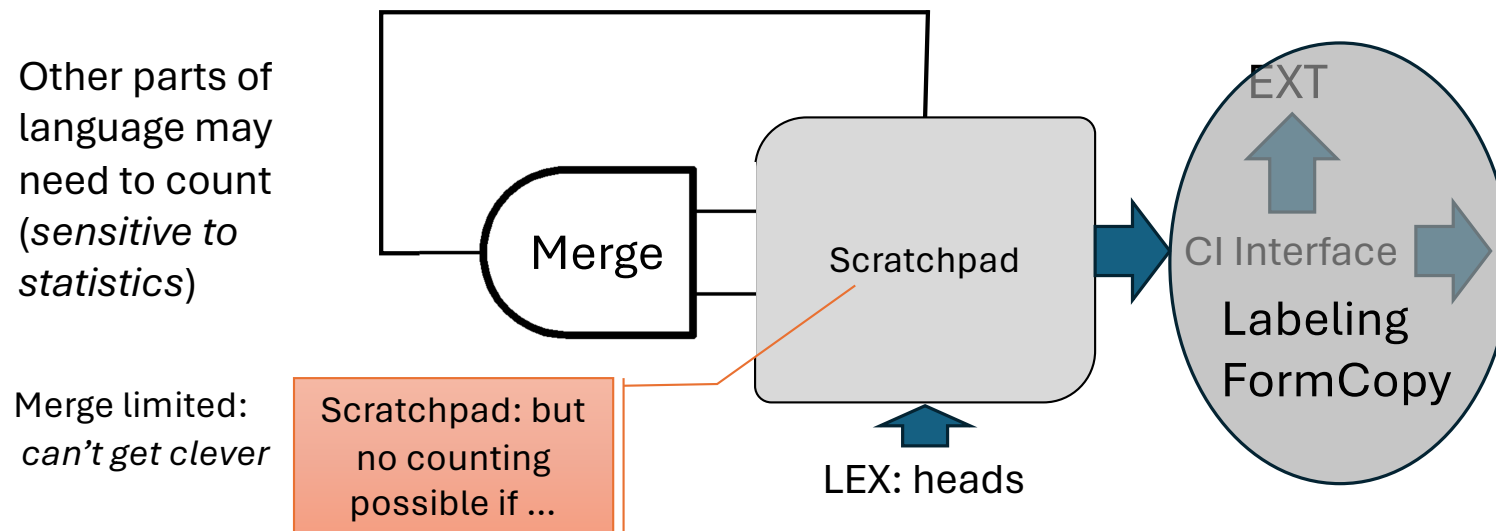
- Chomsky 3Factors:
 - Near uniform genetic endowment (*no known variants, no time*)
 - Narrow variation through experience
 - Third Factor:
 - organic (*slow chemical brain*)/**any** system limitations
 - Resource Restrictions (RR)
 - combinatorics (also Merge *scratchpad* size)
 - *scratchpad*: Workspace (WS)
- Maximal simplicity
 - Evolvability (simplest possible Merge)
 - Explanatory adequacy (*Chomsky MC*)
 - WS size: minimize access, minimize additions? (Minimal Yield)



impose a ceiling:
bound from above

Merge is *recursive*

- i.e. output of Merge (can) feed Merge
- *Idea that there could be a single, simple operation that recursively constructs hierarchical objects is very attractive:*



Free Merge

- *one sense of free:*
 - *anything* in the WS can Merge with *anything else*
 - *a non-starter (both empirically and computationally)*
 - *have to define what's 'anything' is in WS (accessibility)*
- generate-and-test
 - maybe filtered at the interfaces (CI/EXT)
 - violates Third Factor
 - (perhaps worked for GB Theory computational modeling)

Free Merge

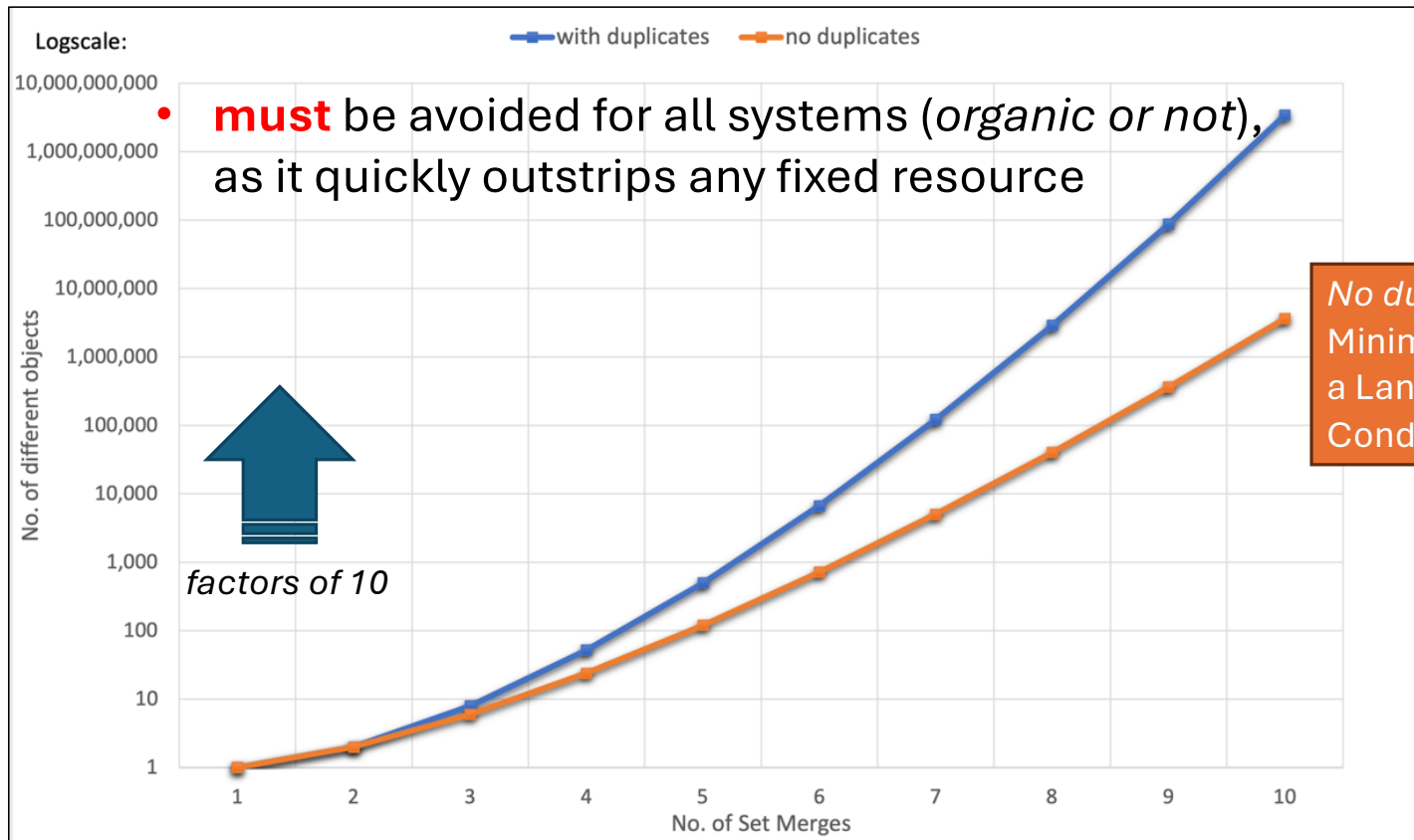
- Nasty combinatorics
 - not computationally efficient (*for another sense, see later slide*)
- Perhaps can be made practical with performance limits?
 - *e.g. whatever comes back first within a timeout*
- **Example** (without LSCs):

- Initial WS: a b
- 1 Merge (EM):
 - {a, b}
- 2 Merges (IM):
 - {a, {a, b}} {b, {a, b}}
- 3 Merges (IM):
 - {{a,b},{a,{a,b}}} {b,{a,{a,b}}} {a,{a,{a,b}}}
 - {{a,b},{b,{a,b}}} {b,{b,{a,b}}} {a,{b,{a,b}}}

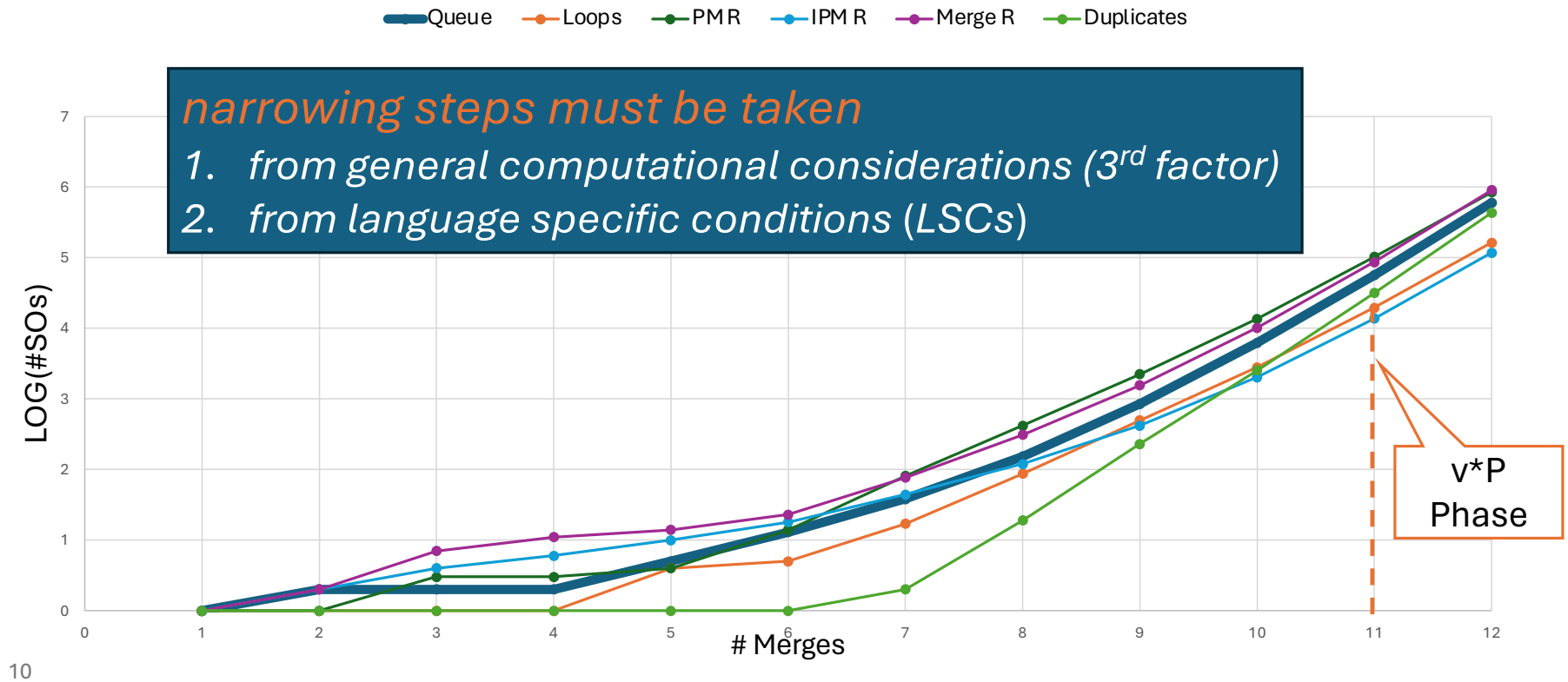
Duplicate should be listed twice!
same structure derived in 2 ways:
{a₁, {a₂, b}}

1. {a₁, {a₁, {a₂, b}}}
2. {a₂, {a₁, {a₂, b}}}

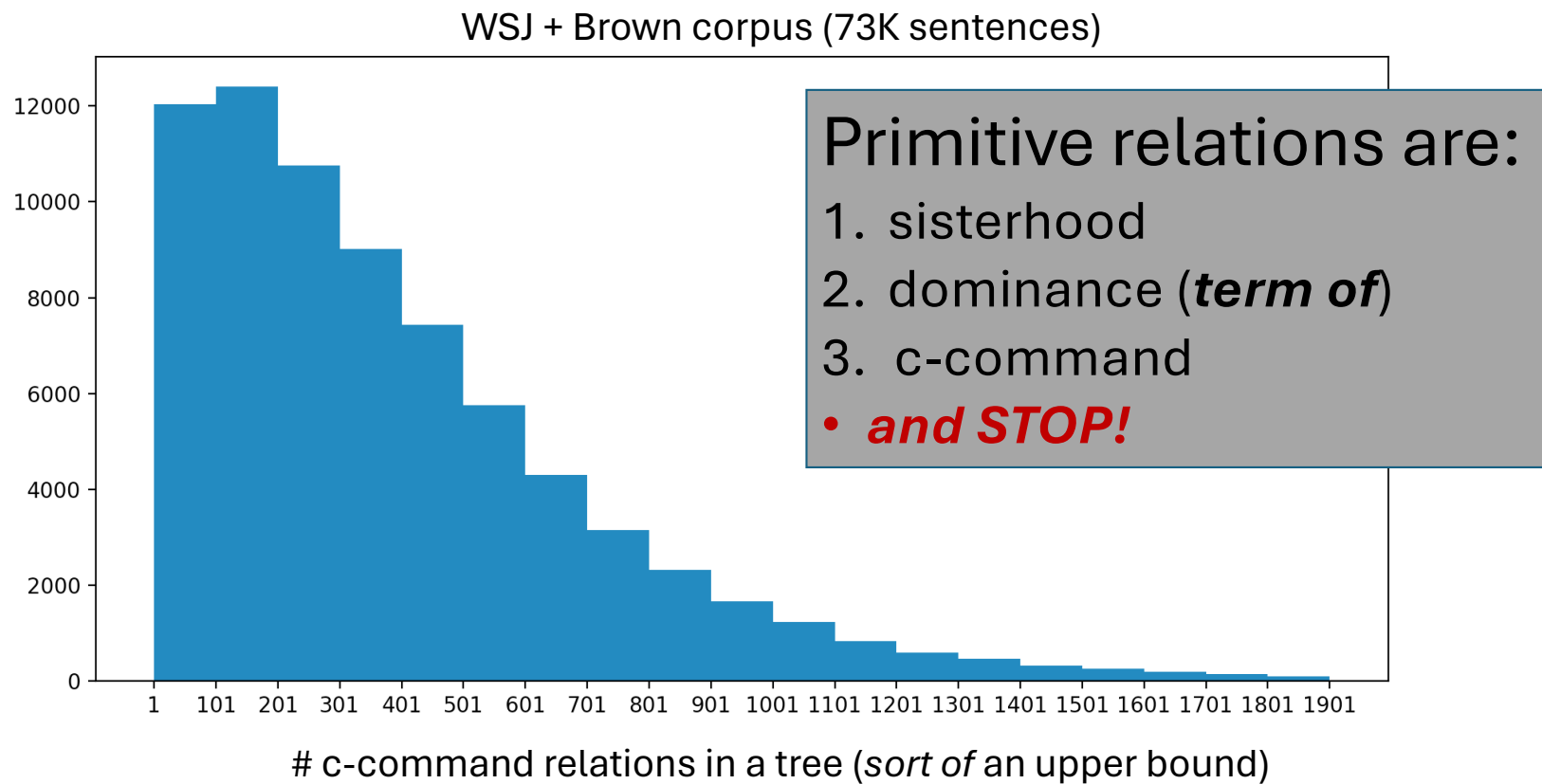
A Note on Combinatorics



A Note on Combinatorics



C-Command Combinatorics



Evolution: possible relations

Organize the primitive relations on structure as:

- 1. term of
- 2. sister of
- 3. 1+2 = c-command
- 4. there is no 4.

Question to Noam:

- Could 3 could be a later stage than 1 and 2 alone? In the current stage of evolution, 3. is available to all of I-Language, but we'll probably never know if there was a proto-I-Language in which 3 was not yet "connected".
- But that would mean Minimal Search (MS) not yet possible.
- MS relies on 3 being available plus RR (a given).
- **Interesting. Will have to think about it.**

***Free* Merge**

- *another* sense of **free**
 - e.g. IM can apply or not
 - output is legible in either case
 - not (formal or interpretable) feature-driven
 - can be replaced with a different theory, see later.

Another notion of Computational Efficiency?

- Chomsky GK:
 - [The problem of evolvability] will be overcome to the extent that the structures of I-language are generated by the **simplest operations**. The *Strong Minimalist Thesis* SMT sets this outcome as a prime goal of the theory of language.
 - Nature then took the usual course of seeking the simplest such operation (reconstruction), relying on natural law (in this case, **principles of computational efficiency**), and on a particular concept of simplicity. The result is a system satisfying the **Basic Property**. Its deepest and most surprising property, structure-dependence, relies on elementary properties of computational efficiency: adopting the simplest computational operation and the no-tampering condition NTC, which basically says "don't add bells and whistles but keep to pure elegance." **That's the simplest imaginable condition of computational efficiency.**

Simplest computational operation: Merge

- Merge
 - two items (*simpler 1 item Merge is excluded*)
 - two stage item pick:
 - stage 1: pick item α from the Workspace (WS)
 - stage 2: pick β , a subterm of α (Internal Merge), or
 - stage 2: pick **another** item β from WS (External Merge)
 - *simpler to pick the same item twice, no?*
 - no modifying α or β (NTC)
 - put $\{\alpha, \beta\}$ into WS
 - no Labeling, nothing else
 - maximally simple (*maximally efficient?*)

Simplest computational operation: Merge

- Markovian Assumption
 - Merge is memoryless
 - *No record/history of Merge operations is kept*
 - *Merge doesn't get access to/can't use its own history*
 - *If it could, it could do Sideways/Parallel Merge and Long Distance Extraction in one go*
 - **conceptually solves the Delete Problem**
 - put $\{\alpha, \beta\}$ into WS, **delete** inputs α and β from WS
(*would be an extra operation otherwise*)
 - *have to assume unused WS items propagate*
 - black box: other operations can't watch Merge either (*could record*)
- Merge is *silo-ed*:
 - e.g. can't access Labels (*computed at INT*)
 - e.g. can't see/use results of FormCopy

Simplest computational operation: Merge

- Maximal simplicity doesn't necessarily mean "free" or be unconstrained
 - **Simple** could mean limiting options
- EM/IM must be defined not to explode search space
 - *computational efficiency*
 - **Minimal Yield (MY)** is a **design principle**: *not active during Merge*
 - *no operation can increase the number of **selectable** items by more than the bare minimum, one*
 - example of a Resource Restriction (RR)
 - MY excludes Merge variants (*by design*) ...

Merge Variants

Simplest Merge excludes head movement:

- e.g. extended verbal projection : R to v to T (to C) (*with tuck-in*)
- heads don't raise: *operation eliminated from I-language*
- "*Head-movement is a familiar operation, but it is illegitimate. Head-movement is not formulable.*" (Chomsky GK)
- must be re-constituted at EXT (**Amalgamation**)

Merge Variants

MY excludes (useful) Merge variants ...

- e.g. **Sideways Merge** (Nunes 2001)
 - WS: $\{X, Y\} Z \dots \Rightarrow_{\text{SideM}} \{X, Y\} \{X, Z\} \dots$ (out: +2)
 - Example (Chomsky 1993):
 - *which claim that John made was he willing to discuss?*
 - $\{..\{.., X\}\} = [\text{was he willing to discuss } \textit{which claim}]$
 - $Z = [\text{OP that John made } \emptyset_p]$
 - *useful also for parasitic gaps etc.*
- another conception:
 - **sideways merge requires selecting 3 times*
- yet another conception:
 - blocked by MS (see later slide)

Merge Variants

MY excludes (useful) Merge variants ... contd.

- e.g. **Parallel Merge** (Citko 2005)
 - WS: X Y Z ... $\Rightarrow_{\text{ParaM}}$ {X, Y} {X, Z} ... (out: +3)
 - (embedded) ATB *wh*-questions (Williams 1978)
 - *I wonder what Gretel recommended and Hansel read*
 - *actually, blocked if Sideways Merge is blocked* (so other conceptions apply too)
 - i.e. Parallel = EM followed by Sideways
 - better handled by FormSet (see later slides)

We also aren't allowed to game MY

- e.g. **Sideways Merge + Delete:**
 - WS: {A, B} C D ... $\Rightarrow_{\text{SideM+D}}$ {A, B} {A, C} ... (still out: +1)

No Free Merge

- Can't tell if structure is built by IM or EM:
 - $\{XP, \{v^*, \{R, XP\}\}\}$
 - more efficiently built by IM XP but not available (Duality)
 - *no indices*, $\{XP_i, \{v^*, \{R, XP_i\}\}\}$ violate NTC,
 - *cf. Move- α : can tell if something is a trace*
- Duality of Semantics (Chomsky GK)
 - LSC: *specific to the language organ*
 - External Merge (EM) build θ -configurations **ONLY**
 - *theta roles not as features, but configurationally identified*
 - Internal Merge (IM) discourse/information/everything else
 - Box theory: A/A'-segregation blocks **Improper Movement**
 - **What was asked to read? / *Who is possible John will see?*

Simplest computational operation: Merge

- EM:
 - perhaps I-Language has a only small number of possible θ -configurations per clause
 - *what about serial verbs?*
- IM:
 - Chomsky MC: IM no longer **recursive** (Box theory)
 - i.e. *no longer feeds Merge*
 - *output is inaccessible (to Merge)*
 - local only (Phase-limited)

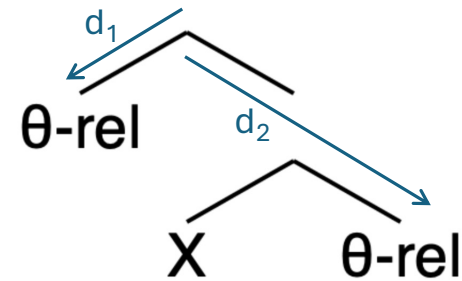
θ -aware Merge

- Chomsky (p.c.):
 - Well, there are no marking for IM vs. EM.
 - INT reads the computed structure and determines how to interpret identical inscriptions.
 - **That's true, but it doesn't mean that IM can't observe theta theory (and duality ...), crashing and hence cancelling the preferred derivation.**
 - **Theta positions are detectable everywhere.**
- [T] All relations and structure-building operations (SBO) are thought-related, with semantic properties interpreted at CI. (Chomsky MC)
- Merge is θ -driven:
 - EM builds θ -configurations efficiently
(as quickly and simply as possible)

Simplest computational operation: Merge

- External Merge (EM): select $X, Y \in WS$
 - assumption: selected X and Y are distinct WS elements
 - *arguably less work to select X twice (more efficient)*
 - but we don't seem to need $\{X, X\}$ in language (*same X*)
 - don't see $Agree(X, X)$ either
- Chomsky (p.c.):
 - **one possibility might be Moro's analysis of copula, which derives "I am I/me" from $\{be, \{I, I\}\}$.**

Minimal Search



- What are we searching for?
 - theta-relevant objects
 - Simplest idea: see first theta-relevant object, done
 - *a lower copy is not even seen/detected*
 - *in fact, Merge has no notion of "copy"*
 - *there is no findall, compute distances, then $d_1 < d_2$*
 - *i.e. not a sort operation*
 - *also, distance d_i not reported/used (e.g. Path)*
 - no feature propagation/inheritance for Merge objects
- External Merge (EM) (Chomsky p.c.):
 - **We assume that Merge like other operations observes it.**
 - **That's why only members of WS, not their terms, are eligible for EM.**

Simplest computational operation: Merge

- What about Pair Merge?
 - is asymmetric Merge still needed?
- What about FormSet?
 - needed for **unbounded, unstructured sequences (UUS)** (Chomsky GK)
 - set members observe parallelism (*different from Merge*)
 - *John, Bill, my friends, the actor who won the Oscar, ... ran, danced, took a vacation* (GK: 31)
 - $S_1 = \{\text{John, Bill, my friends, the actor who won the Oscar}\}$
 - $S_2 = \{\text{ran, danced, took a vacation}\}$
 - S_1 and S_2 can have distinct cardinality (Chomsky GK fn. 47)

UUS: Relative clause stacking

Example:

9) *the student who lives here who studies English whom I know*

FormSet applies to:

10) (a) {student, {who {~~student~~, {lives here}}}}

(b) {student, {who, {~~student~~, {studies English}}}}

(c) {student, {who, {I, {know, ~~student~~}}}}

- relative CPs need not be identical (Williams, 1978)

Optionally spelling out (conjunctive form) as:

11) *the student who lives here, who studies English **and** whom I know*

UUS: Relative clause stacking

[GIF animation of compute steps not visible in PDF version]

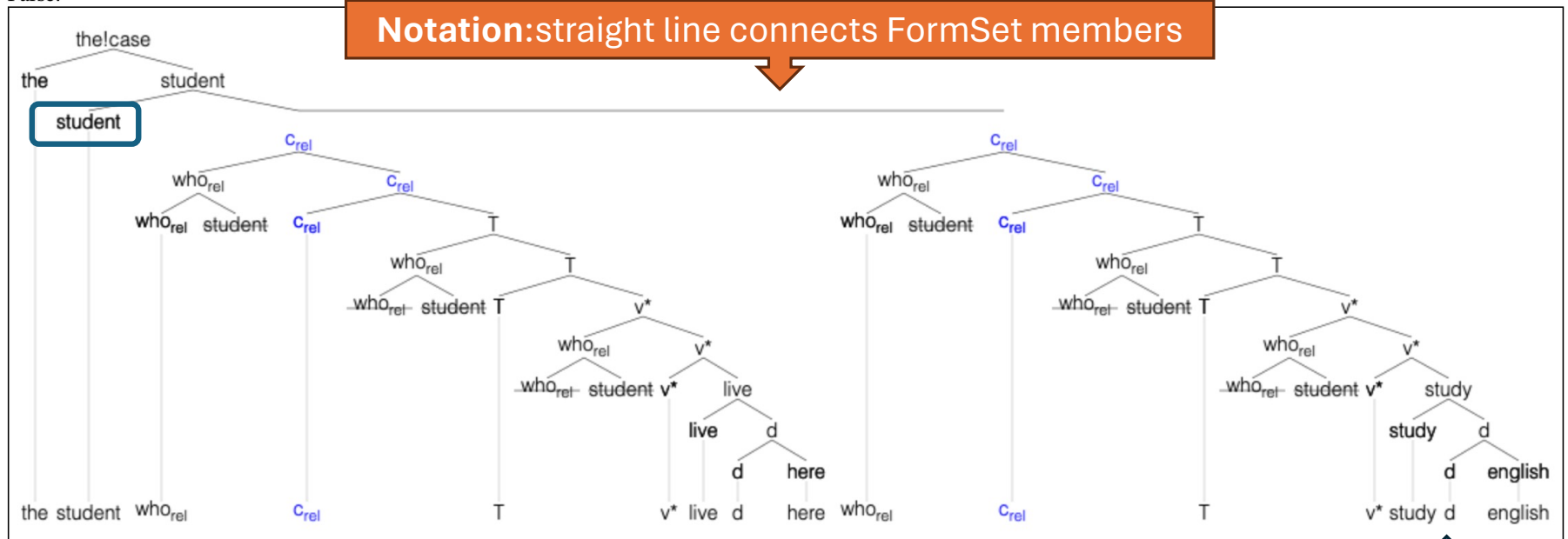
here!D

Initial SO is head of stream :**Operation**
[[d!case!N],[live],[v*!phi],[[student!D],[who_rel!case!N]],[T!phi],[c_rel!rel!T!phi]] :**Stream**

UUS: Relative clause stacking

- note 62: Fong & Ginsburg. Open Linguistics, vol. 9, no. 1, 2023.

Parse:



Spell-out:

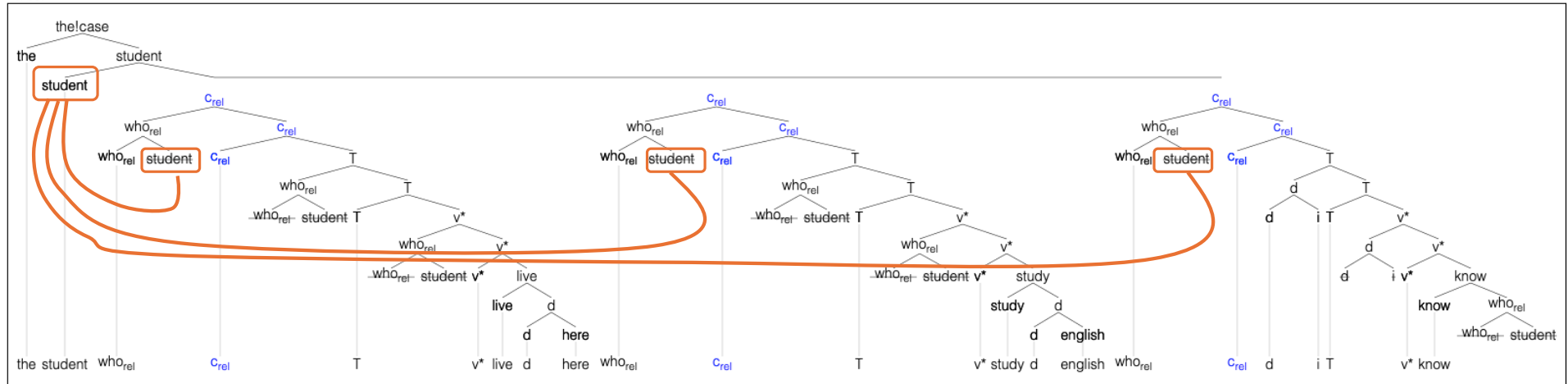
- the student who -s live -acc here who -s study -acc English (after morpheme realization)
- the student who live -s here -acc who study -s English -acc (after affix-hop)
- the student who live -s here -acc who study -s English -acc (after morpheme realization, stage 2)
- the student who lives here who studies English

old implementation
DP

UUS: Relative clause stacking

- *the student who lives here who studies English whom I know*

Parse:



Spell-out:
 the student who -s live -acc here who -s study -acc English whom I -m know (after morpheme realization)
 the student who live -s here -acc who study -s English -acc whom I know -m (after affix-hop)
 the student who live -s here -acc who study -s English -acc whom I know -m (after morpheme realization, stage 2)
 the student who lives here who studies English whom I know

One derivation found.

Not relevant: implementation details, could be Box theory
Important: Parallelism (*members of a set treated uniformly by Merge*)

UUS: Relative clause stacking

- Relative clause stacking parallel to PP stacking (Chomsky GK):
 - 12) (a) *John lived on a farm with his family*
(b) *John lived on a farm **and** with his family*
- IM in unison, targets subject/object:
 - 13) (a) *Which book did John buy and read?*
(b) *which book did John buy ~~which book~~ and read ~~which book~~*
(c) $\{\{\text{John}, \{v^*, \{\text{buy}, \text{which book}\}\}\}, \{\text{John}, \{v^*, \{\text{read}, \text{which book}\}\}\}\}$
- *rationale for Parallel Merge is gone*
 - 14) (a) *John arrived **and** met Bill*
(b) $\{\{v, \{\text{arrive}, \text{John}\}, \{\text{John}, \{v^*, \{\text{meet}, \text{Bill}\}\}\}\}$

Output of FormSet and the target of IM

- Identical inscription target requirement (Williams 1978):
 - 15) (a) *Who *and* when did John see and ignore him?
 - (b) {{John, {v*, {see, who}}}, {{John, {v*, {ignore, him}}}, when}}
- FormSet :
 - 16) (a) *When and where did you see her?*
 - (b) {C_Q, {you, {INFL, {{you, {v*, {see, her}}}, {when, where}}}}

Adjectival and Predicative Noun Phrases

Example:

17) (a) *the politician is greedy and a charlatan*

(b) {politician, {be, {{greedy, politician}, {charlatan, politician}}}}

Similarly:

18) (a) {hallway, {{long, hallway}, {narrow, hallway}, {dark, hallway}}}

(b) *the hallway is long, narrow and dark*

(c) *the long, dark and narrow hallway*

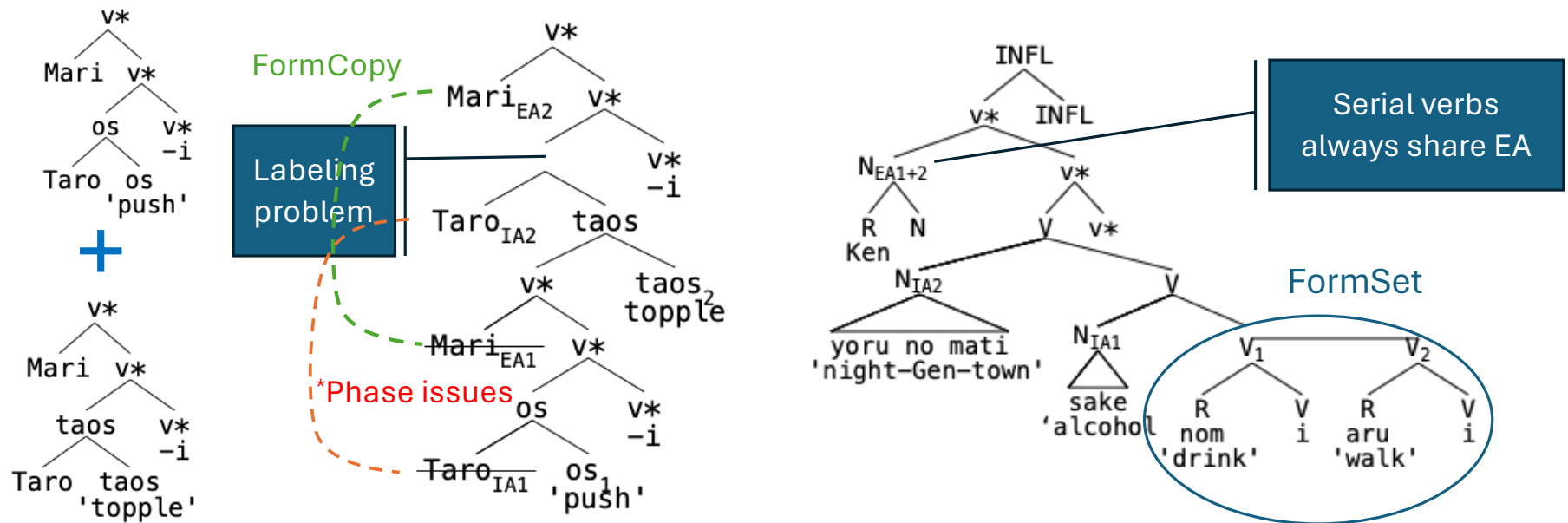
- (Di Scullo 2022) complex cardinals

19) (a) *two hundred and two* (additive complex)

(b) {two hundred, two}

Merge vs. FormSet

- Serial verbs (monoclausal): appears in many languages
- Mari-ga Taro-o osi-taosi-ta / 'Mari toppled Taro by pushing him' (from Nishiyama 1998:185)
- Ken-ga yoru-no mati-o sake-o nomi-arui-ta / 'Ken went bar-hopping at night' (cf. Kageyama 1993)



Joint work with Jason Ginsburg, Hiroshi Terada & Masumi Matsumoto (ELSJ Spring Forum 2024)

How do we first populate the Workspace?

- Not specified (outside I-Language?)
 - Need to select the right heads (*for convergence*)
 - *but how do we know what the right initial state (Lexical Array) is?*
 - Merge stops when?
 - Pre-condition: WS is a single object
 - Termination not in terms of accessibility
- Also important for Parsing
 - suppose Merge is the only **structure-builder** (*no time to evolve anything else*)
 - map sensory input to heads (covert and visible), then Merge fires
 - check generated object against input

Conceptual-Lexical Apparatus

- GK:
 - The rich and intricate concepts that are the atoms of LEX would have no function outside of a system of generation of thought that yields complex expressions that can enter into reasoning, reflection, and other mental acts. If they had emerged before language, they would have been useless waste and would have been discarded. **It is reasonable to suppose that when the capacity emerged for recursive generation of linguistic expressions, it appropriated a lexicon of elementary items available to proto-humans**; and only later, as the combinatorial possibilities of generating language and thought became available, concepts of the distinctive human character appeared – how, we can only guess, but at least it might be possible to explore the question if it falls within the history of humans.
- Mind-internal objects:
 - events, propositions

Phases

- Chomsky MI justification of Phases:
 - **Operative complexity** in some natural sense is reduced, with each stage of the derivation accessing only part of LA.
 - *Third Factor*
- Box theory:
 - c-commanding identical inscriptions (FormCopy)
 - {EA, {v*, {R, IA}}}
 - EA doesn't c-command IA in the same Phase {...}
 - **John saw*
 - (*John, John*) with independent θ -roles but a copy relation
 - *John saw John*
 - repetitions, no copy relation, must be spelled out separately

Workspace Partitioning

Phases impose locality

Observe workspace must be partitioned anyway:

- assume Language needs {XP, YP}
- XP and YP must be constructed separately
 - e.g. can't have heads for XP enter YP construction (*and vice-versa*)
- {XP, YP}
 - *supports divide-and-conquer*
 - can be done simultaneously
 - reduces **operative complexity** (like Phases)

A problem of vacuous recursion eliminated?

- Internal Merge (IM):
 - $\{\alpha, \beta\}$
 - $\{\{\alpha, \beta\}, \beta\}$
 - $\{\{\{\alpha, \beta\}, \beta\}, \beta\}$
 - ...
 - *not blocked by Labeling theory*
- IM free? IM can apply at all stages
- Not limited by Phase

Vacuous recursion eliminable?

Lots of principles proposed in the past to limit (vacuous/unwanted) movement, e.g. LastResort, etc.

- IM to Phase boundaries only
 - But what about Object Shift (OS) or raising to surface subject?
 - {R, IA}
 - {IA, {R, IA}} (OS)
 - {EA, {INFL, {EA, v*P}}} (to surface subject)
- Box theory
 - IM no longer iterates (successive cyclically)
 - {IA, {v*, {IA, {R, IA}}}},
 - e.g. possible when IA is a *wh*-phrase
 - *who did John see?*

IM Combinatorics

- Narrowing steps cannot involve *clever devices*:
 - **Example:** Infinite Loop Filter:
 - * $\pi\pi$, where $\pi = (\text{IM } O_1, \dots, \text{IM } O_n)$, $n \geq 1$
 - (O_i = selected subterm; cross- π compare selected O_i)
 - **Example:**
 - $\{\underline{a}, b\}$
 - $\{\underline{a}_1, \{a_2, b\}\}$
 - $\{a_1, \{\underline{a}_1, \{a_2, b\}\}\}$
 - $\{a_1, \{\underline{a}_1, \{a_1, \{a_2, b\}\}\}\}$
 - $\{a_1, \{\underline{a}_1, \{a_1, \{a_1, \{a_2, b\}\}\}\}\}$

and so on ... **without violating MS**

IM Combinatorics

Infinite Loop strategy:

select a and b alternately

- **Example:**

- $\{a, \underline{b}\}$
- $\{b, \{\underline{a}, b\}\}$
- $\{a, \{\underline{b}, \{a, b\}\}\}$
- $\{b, \{\underline{a}, \{b, \{a, b\}\}\}\}$
- $\{a, \{\underline{b}, \{a, \{b, \{a, b\}\}\}\}\}$

and so on ...

IM Combinatorics

- Proposed filter
 - new device (*unless loop-free computation* *stationary toolkit*)
*SMT
 - can't rule all cases anyway (see below)
 - requires access to history of Merge (*violates Markovian assumption*)



- **Example:**

- WS: a b
- {a, b}
- {b, {a, b}}
- {{a, b}, {b, {a, b}}}
- {{b, {a, b}}, {{a, b}, {b, {a, b}}}}
- {{a, b}, {b, {a, b}}}, {{b, {a, b}}, {{a, b}, {b, {a, b}}}}}
- and so on ...

IM selects the WS object from previous round

- compare to ordinals
- $successor(x) = x \cup \{x\}$

Appendix

Head Movement

- Heads don't raise: *head-movement is eliminated from I-language*
 - *Head-movement is a familiar operation, but it is illegitimate. Head-movement is not formulable (fn49: violates **Extension**) ... head-movement typically lacks semantic consequences. (Chomsky 2021)*
 - *Externalization, however, requires an operation that amalgamates the inflectional elements INFL and v along with the roots, presumably cyclically. The operation **AMALGAMATE** forms a complex [INFL, [v, root]]. (Chomsky 2021)*

Object Shift (OS)

- OS visibility:
 - Scandinavian, not English
- OS has semantic effect:
 - *specific*: structurally marked by moving out of the VP (Diesing 1992)
- OS:
 - Icelandic: full NPs and pronouns, mainland Scandinavian: pronouns only
- Holmberg's Generalization:
 - OS is blocked if the main verb which selects the object does not move out of its base position in V°.
 - main verb doesn't move if V2 verb is an auxiliary verb
- Specificity
 - Every woman talked to a child in fifth grade (Enç 1991)
 - discourse linked: *child* (when specific indefinite)
 - specific: structurally marked by moving out of the VP (Diesing 1992)



Object Shift (OS)

(Vikner 2005)

(4) Icelandic: full DPs

- a. Af hverju las_v Pétur aldrei [_{VP} t_v þessa bók]?
why read Peter never this book
- b. Af hverju las_v Pétur þessa bók_i aldrei [_{VP} t_v t_i]?
why read Peter this book never

(5) Icelandic: pronouns

- a. *Af hverju las_v Pétur aldrei [_{VP} t_v hana]?
why read Peter never it
- b. Af hverju las_v Pétur hana_i aldrei [_{VP} t_v t_i]?
why read Peter it never

(6) Danish: full DPs (impossible)

- a. Hvorfor læste_v Peter aldrig [_{VP} t_v den her bog]?
why read Peter never this book
- b. *Hvorfor læste_v Peter den her bog_i aldrig [_{VP} t_v t_i]?
why read Peter this book never

(7) Danish: pronouns

- a. *Hvorfor læste_v Peter aldrig [_{VP} t_v den]?
why read Peter never it
- b. Hvorfor læste_v Peter den_i aldrig [_{VP} t_v t_i]?
why read Peter it never

Rizzi 2004 A/A' source distinction

- (27) a. ?Which problem do you wonder how to solve <which problem>?
b. *How do you wonder which problem to solve <how>?

reduced to effects at the interface. Possibilities that have been investigated (if not in these terms) include constraints that bar PF-vacuous overt movement and others that seek to limit effects on PF (Procrastinate). An LF counterpart is that covert operations are allowed only if they have an effect on interpretation at LF. Another category seeks to reduce “search space” for computation: “Shortest Movement/Attract,” successive-cyclic movement (Relativized Minimality, Subjacency), restriction of search to c-command or minimal domains, and so on. Yet another imposes “local determinability” conditions (barring “look-ahead,” “backtracking,” or comparison of alternatives). I will assume these ideas to be generally on the right track and pursue them further below.²⁵

Chomsky MI