The Extension Condition and Merge Optimality

Matilde Marcolli

Stony Brook, Linguistics, November 21, 2024 this talk is based on:

- Matilde Marcolli, Richard K. Larson, Riny Huijbregts, *Extension Condition "violations" and Merge optimality constraints*, preprint (...ok not really but almost there)

The Goals:

1. show how linguistics constraints can be theoretically *derived* from the mathematical structure

   (what does it mean to derive a constraint? think Emmy Noether's theorem in physics: symmetry $\Rightarrow$ conservation law)

2. specific case: use mathematical formulation of Minimalism to investigate linguistic phenomena that apparently violate the Extension Condition

no linguists were harmed in the making of this work (...hopefully)

Thinking about constraints

Linguistics: many constraints are empirically observed and then theoretically generalized (argued to be natural from a theoretical perspective): for example

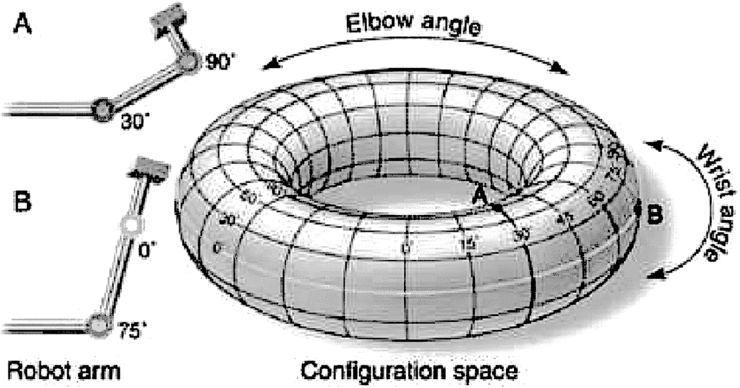
- constraints on movement
- constraints on phrase structure (eg final-over-final condition)
- constraints on agreement

How does one think about constraints in a physical theory?

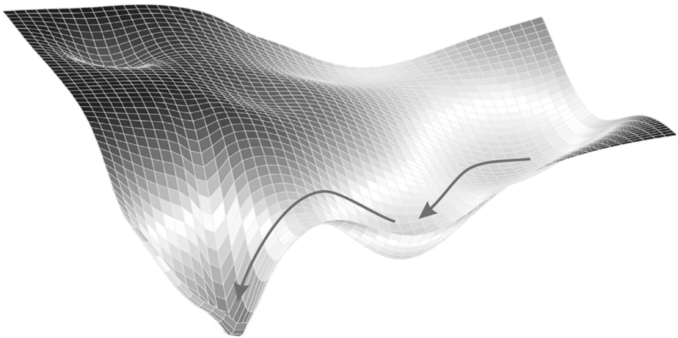
Physics: *hard* and *soft* constraints

- hard constraints: dictated by the "kinematics" (the *geometric* or *algebraic* structure underlying the physical model)
- soft constraints: dictated by *optimization* (of an action functional, energy, cost function) over a geometric space (or algebraic structure) defined by the hard constraints

violation of hard constraints invalidates the model; violation of soft constraints should occur rarely



hard constraints: are dictated by the intrinsic geometric or algebraic structure of the model, solutions of the equation of motion *necessarily* satisfy these constraints



soft constraints: optimization over an energy landscape, equilibrium positions are favored, small oscillations around an equilibrium positions are likely, positions very far from equilibrium are very unlikely (quantifiable in terms of estimating distance from equilibrium positions)

Questions

- which constraints in linguistics are hard constraints and which are soft constraints?
- in physics the intrinsic algebraic and geometric structure of the model dictates the constraints: can one also *derive* linguistics constraints from the algebraic structure rather than abstract them from empirical observation and analysis?
- quantification of violations to soft constraints predicts frequency/likelihood of occurrence of corresponding phenomena: does mathematical prediction meet linguistic observation?

for syntactic phenomena mathematical prediction is possible through the mathematical formulation of Merge and Minimalism

Summary of Mathematical Minimalism

- M. Marcolli, N. Chomsky, R.C. Berwick, *Mathematical Structure of Syntactic Merge. An Algebraic Model for Generative Linguistics*, MIT Press, 2025 (in print)

Main aspects of the model:

- magma of syntactic objects
- Hopf algebra of workspaces
- Merge action on workspaces (Hopf algebra Markov chain)
- optimality constraints on Merge action
- head, complement, phases, labeling
- Externalization (Parameters & Projections)
- syntax–semantics interface (Birkhoff factorization)

syntactic objects and workspaces

- $\mathcal{SO}_0$ finite set of lexical items and syntactic features
- $\mathcal{SO}$ countable set of syntactic objects, with the algebraic structure of free commutative nonassociative magma generated by $\mathcal{SO}_0$

$$\mathcal{SO} = \mathrm{Magma}_{c,na}(\mathcal{SO}_0, \mathfrak{M}) \cong \mathfrak{T}_{\mathcal{SO}_0}$$

  canonically isomorphic set of (non-planar) binary rooted trees with leaves labelled in $\mathcal{SO}_0$
- countable set $\mathfrak{F}_{\mathcal{SO}_0}$ of workspaces: binary rooted forests $F = T_1 \sqcup \cdots \sqcup T_n$ with components T_i in $\mathfrak{T}_{\mathcal{SO}_0}$
- accessible terms $T_v \subset T$, non-root vertex v

Merge and action on workspaces

- Hopf algebra of workspaces: vector space $\mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0})$ with product $F \otimes F' \mapsto F \sqcup F'$ and coproduct $\Delta(F) = \sqcup_i \Delta(T_i)$

$$\Delta(T) = \sum_{\underline{v}} F_{\underline{v}} \otimes T/F_{\underline{v}}$$

extraction of accessible terms + cancellation of deeper copies

- action of Merge on workspaces: $S, S' \in \mathfrak{T}_{\mathcal{SO}_0}$

$$\mathfrak{M}_{S,S'} = \sqcup \circ (\mathfrak{B} \otimes \mathrm{id}) \circ \delta_{S,S'} \circ \Delta$$

all possible Merge transformations $\mathcal{K} = \sqcup \circ (\mathfrak{B} \otimes \mathrm{id}) \circ \Pi_{(2)} \circ \Delta$

- $\mathfrak{B}$ grafting at the root
- include $S' = 1$ (unit, formal empty object) so $\mathfrak{M}_{\beta, S/\beta} \circ \mathfrak{M}_{\beta,1}$ for Internal Merge
- combined $\mathcal{K}$ is a Hopf algebra Markov chain (more later)

Optimality constraints on Merge

- cost function weighting terms of the coproduct (cost of extraction, cost of cancellation): Minimal Search $\delta_{S,S'}$ in $\mathfrak{M}_{S,S'}$ picks zero-cost terms (as leading order)
- effect of the Merge $\mathfrak{M}_{S,S'}$ on different size measures of workspace: Resource Restriction, select optimal terms

Note: here also clear difference between structural constraints (dictated by algebraic structure) and soft constraints (dictated by optimality); first type admit no violations within the model, second type admits possible violations, but *increasingly rare with the amount of violation* (how far from equilibrium)

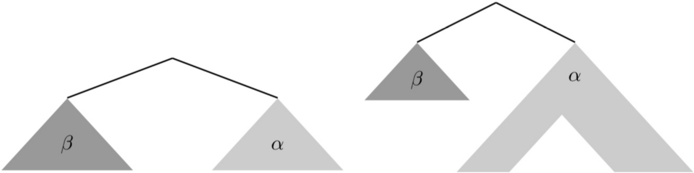
different proposed forms of Merge

- Internal/External Merge satisfy optimality
- Sideward Merge: violate optimality
- Countercyclic movement: violate structural constraints (EC violations)

Extension Condition

- structure formation (Merge) only grows structures at the root

Why?



External Merge and Internal Merge satisfy Extension Condition

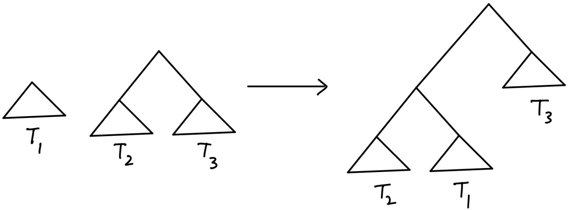
What does the principle exclude?

Two types of mathematical operations of tree do *not* grow at root:

1. insertion at edges (Lie algebra)
2. grafting at leaves (operad)

insertion Lie algebra

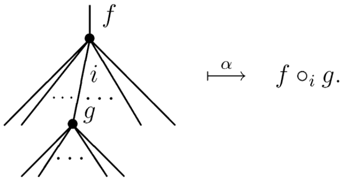
What linguists call countercyclic movement



structure growth that violates the Extension Condition

algebraically this operation defines a pre-Lie structure and an associated Lie algebra

Operad composition: output/input grafting at leaves



another structure growth that violates Extension Condition

trees as operations with several inputs and one output: composition by plugging in output to a new input

linguistic use? interfacing syntax with morphology

Extension Condition in Mathematical Minimalism

it is a structural constraint (hard constraints in the physics sense)

- growth at the root from use of grafting $\mathfrak{B}$ in Merge operation
- combination of structures by merging at an internal vertex/edge does not have the same algebraic properties

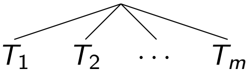
What is the algebraic property that characterizes the structure growth operation $\mathfrak{B}$ (grafting at a new root) that is *not* satisfied by EC-violating growth?

Note: crucial part of the structure building operation of Merge is the Hopf algebra structure of workspaces, so *expect* a characterization of EC in terms of Hopf algebra properties

What does physics suggest?

Note: the same grafting operator $\mathfrak{B}$ plays role in physics

$$\mathcal{B}(T_1 \sqcup \cdots \sqcup T_m) =$$



(binary in the case of Merge, arbitrary arity in physics)

Main algebraic property of $\mathfrak{B}$: cocycle condition

$$\Delta \circ \mathcal{B} = \mathcal{B} \otimes 1 + (\mathrm{id} \otimes \mathcal{B}) \circ \Delta$$

Hochschild cocycle in the Hopf algebra cohomology
Two main roles of this identity

1. recursive Dyson-Schwinger equations $X = \mathfrak{B}(P(X))$ equation encodes generative process of the combinatorial objects
2. cocycle $\mathfrak{B}$ gives a *universal property* for the Hopf algebra, with respect to mapping to other bialgebras/Hopf.

Result: no growth that is EC-violating can satisfy cocycle condition

Extension Condition *violations* in linguistics?

Four categories of proposed EC violations:

1. head-to-head movement,
2. head-to-phrase movement (including phrasal affixes and syntactic cliticization),
3. phrase-to-head movement (like verb-particle alternation),
4. phrase-to-phrase movement (operator-variable phenomena).

There is a clear sense, on linguistics grounds, that these four are not on the same level and that there is a hierarchy between these phenomena, in order or increasing "difficulty"

Apparent EC violations

all these four phenomena appear to present EC-violation, because they can all fit a kind of derivation procedure:

$$X, \; [\ldots Y \ldots] \tag{1}$$

$$[X \; [\ldots Y \ldots]] \;\; \mathrm{EM} \tag{2}$$

$$[X - Y \; [\ldots \not{Y} \ldots]] \quad \text{“EC violation”} \tag{3}$$

with $X - Y$ representing adjunction of Y to X
this fits the “insertion at internal edges” EC-violating growth

Head-to-Head Movement

as above with X and Y syntactic atoms (heads):

$$X, \; [_{YP} \dots Y \dots]$$

$$[X \; [_{YP} \dots Y \dots]] \quad \mathrm{EM}$$

$$[X - Y \; [_{YP} \dots \not{Y} \dots]] \quad \text{“EC violation”}$$

Examples: English “subject auxiliary inversion” (a), French V-to-T movement (b), Germanic “verb second” (c), verb-initial word order in VSO languages like Welsh (d), incorporations in Malayalam (e)

a. [$_{CP}$ C−will [$_{TP}$ Mary ~~will~~ leave]]

b. Pierre [T−mange [$_{VP}$ souvent ~~mange~~ des pommes]]
Pierre eats often of the apples
'Pierre often eats apples'

c. Gestern [$_{CP}$ C−hat [Hans gelacht ~~hat~~]]
Yesterday has Hans laughed has
'Yesterday, Hans laughed.'

d. [$_{CP}$ C−gwelodd [$_{TP}$ Siôn ~~gwelodd~~ y defaid]]
saw Siôn the sheep
'Siôn saw the sheep'

e. Kuṭṭikkə [$_{VP}$ uraŋŋ-aṇam [PRO ~~uraŋŋ~~]]
child.dat sleep-want
'The child wants to sleep'

Head-to-Phrase Movement

as above but X is a phrase and Y is an atom:

$$XP, \; [_{YP} \ldots Y \ldots]$$

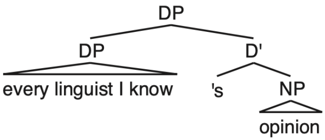
$$[XP \; [_{YP} \ldots Y \ldots]] \quad \mathrm{EM}$$

$$[XP - Y \; [_{YP} \ldots \not{Y} \ldots]] \quad \text{“EC violation”}$$

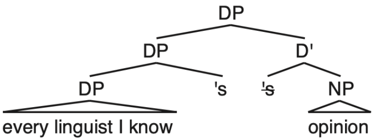
Examples: phrasal affixes or syntactic clitics (English genitive morpheme –’s)

a. [Fred] **’s** opinion about the English genitive is different from mine.
b. [The man on the Clapham omnibus] **’s** opinion about the English genitive is poorly thought out.
c. [Every linguist I know] **’s** opinion about the English genitive involves functional categories.
d. [That young hotshot who was recently hired at Princeton that I was just telling you about] **’s** opinion about the English genitive is simply wrong.
e. Even [that colleague who shares an office with you] **’s** opinion about the English genitive is not to be trusted.

a.



b.



Phrase-to-Head Movement

as above but now X is a head and Y is phrase:

$$X, \ [_{ZP} \ldots YP \ldots]$$

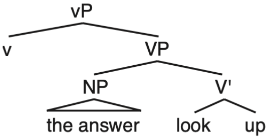
$$[X \ [_{ZP} \ldots Y \ldots]] \quad \text{EM}$$

$$[X - YP \ [_{ZP} \ldots \cancel{YP} \ldots]] \quad \text{“EC violation”}$$

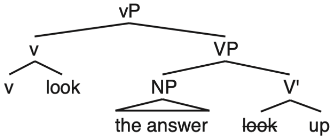
Examples: verb-particle alternation

a. Leslie **looked up** the answer.
b. Leslie **looked** the answer **up**.

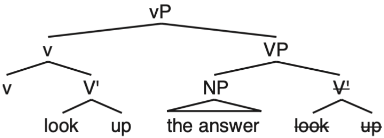
a.


vP
v
VP
NP
V'
the answer
look
up

b.


vP
v
VP
v
look
NP
V'
the answer
~~look~~
up

c.


vP
v
VP
v
V'
NP
~~V'~~
look
up
the answer
~~look~~
~~up~~

Phrase-to-Phrase Movement

as above but X and Y are both phrases:

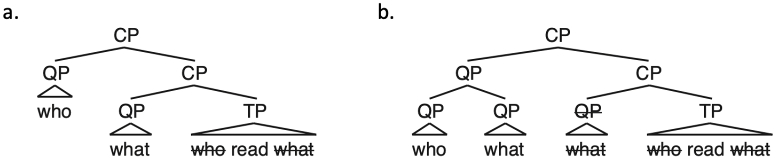
$$XP, \ [_{ZP} \dots YP \dots]$$

$$[XP \ [_{ZP} \dots Y \dots]] \quad \mathrm{EM}$$

$$[XP - YP \ [_{ZP} \dots \cancel{YP} \dots]] \quad \text{“EC violation”}$$

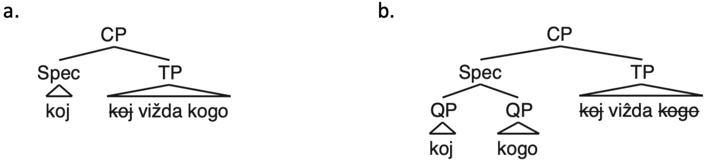
Examples: operator-variable phenomena

quantifier absorption for multiple *wh*-questions



c. [WH x: person(x)] [WH y: thing(y)] [x read y] $\Longrightarrow$
[WH x, WH y: person(x) & thing(y)] [x read y]

English: Who read what?



Bulgarian: "Koj vižda kogo?" (who saw what?)

But are these *really* EC-violations? ... math says no!

why no? because EC is an *algebraic property* of structure formation: do not expect to see violations

What replaces EC-violating derivation then? Possible answer: Sideward Merge

$X, \; [\ldots Y \ldots]$		$X, \; [\ldots Y \ldots]$	
$X - Y \; [\ldots \not{Y} \ldots]]$	**SM**	$[X \; [\ldots Y \ldots]]$	EM
$[X - Y \; [\ldots \not{Y} \ldots]]$	EM	$[X - Y \; [\ldots \not{Y} \ldots]]$	"EC violation"

Now the left column has no EC-violations

Forms of Sideward Merge

action of Merge on workspaces: $S, S' \in \mathfrak{T}_{\mathcal{SO}_0}$

$$\mathfrak{M}_{S,S'} = \sqcup \circ (\mathfrak{B} \otimes \mathrm{id}) \circ \delta_{S,S'} \circ \Delta$$

start with workspace $F = T \sqcup T'$ with two syntactic objects:

1. $S = T_v \subset T$ an accessible term of the syntactic object T, and $S' = T'$ other syntactic object, with resulting new workspace

$$\mathfrak{M}_{S,S'}(F) = \mathfrak{M}(T_v, T') \sqcup T/T_v \, .$$

2. $S = T_v \subset T$ as above and $S" = T'_w \subset T'$ an accessible term of the syntactic object T', with resulting new workspace

$$\mathfrak{M}_{S,S'}(F) = \mathfrak{M}(T_v, T'_w) \sqcup T/T_v \sqcup T'/T'_w \, .$$

3. $S = T_v \subset T$ and $S' = T_w \subset T$ two disjoint accessible terms of the same syntactic object T, with resulting new workspace

$$\mathfrak{M}_{S,S'}(F) = \mathfrak{M}(T_v, T_w) \sqcup T/(T_v \sqcup T_w) \sqcup T' \, .$$

Optimality constraints on Merge

- External Merge, Internal Merge, Sideward Merge are all represented in the action of Merge on workspaces

$$\mathcal{K} = \sum_{S,S'} \mathfrak{M}_{S,S'} = \sqcup \circ (\mathfrak{B} \otimes \mathrm{id}) \circ \Pi_{(2)} \circ \Delta$$

- Chomsky in recent Merge & SMT formulation suggested that SM is eliminated by optimality constraints that only EM and IM satisfy
  1. Minimal Search
  2. Resource Restriction (Minimal Yield)
- both can be quantified in the mathematical formulation
- soft constraints: optimality with respect to a cost function
- deviation from optimality can be accurately measured

Optimality violations of Sideward Merge (Minimal Search)

will focus more on Resource Restriction but quick idea about Minimal Search

- assign a cost to extraction of accessible term and to deletion of deeper copy (coproduct operation) and to grafting
- cost is proportional to either depth of location of accessible term or size of accessible term; cost of deletion of deeper copy is inversely proportional to cost of extraction; cost of merging combines the two costs
- EM and IM result in zero-cost operations
- SM has nonzero cost

Counting costs for Minimal Search

- two main possible ways depending on viewing the structures top-down or bottom up:
  - top down: cost of accessing lower terms higher the deeper they are (d_v dist T_v from root)
  - bottom up: first locate accessible terms that are atomic, then of increasingly large size ($\ell(T_v)/\ell(T)$ relative size in # leaves)
- cost of extraction and of cancellation of deeper copy compensate each other
- cost of merging depending on cost of things merged and of operation
- bottom-up case: $\mathfrak{c}(T_v) = \ell(T_v)/\ell(T)$, $\mathfrak{c}(T/T_v) = 1 - \ell(T_v)/\ell(T)$

$$\mathfrak{c}(\mathfrak{M}(A, B)) := \mathfrak{b}(A, B) - \mathfrak{c}(A) - \mathfrak{c}(B)$$

- $\mathfrak{b}(A, B)$ counts number of components from which A, B is taken: $\mathfrak{b}(A, B) = 1$ if A and B from the same component and $\mathfrak{b}(A, B) = 2$ if A and B from different components

Minimal Search costs for forms of Merge

- External and Internal Merge are zero-cost

$$\begin{array}{ll} \mathrm{EM}: & \mathfrak{c}(\mathfrak{M}(T, T')) = 2 - \mathfrak{c}(T) - \mathfrak{c}(T') = 0 \\ \mathrm{IM}: & \mathfrak{c}(\mathfrak{M}(T_v, T/T_v)) = 1 - \mathfrak{c}(T_v) - \mathfrak{c}(T/T_v) = 0 . \end{array}$$

- Sideward Merge is *not* zero-cost

$$\begin{array}{ll} \mathrm{SM}(1): & \mathfrak{c}(\mathfrak{M}(T_v, T')) = 2 - \mathfrak{c}(T_v) - 1 = 1 - \mathfrak{c}(T_v) > 0 \\ \mathrm{SM}(2): & \mathfrak{c}(\mathfrak{M}(T_v, T'_w)) = 2 - \mathfrak{c}(T_v) - \mathfrak{c}(T_w) > 0 \\ \mathrm{SM}(3): & \mathfrak{c}(\mathfrak{M}(T_v, T_w)) = 1 - \mathfrak{c}(T_v) - \mathfrak{c}(T_w) > 0 , \end{array}$$

- Minimal Search gives high cost to SM involving extraction of atomic elements (*both* for bottom-up and top-down search): favors short derivations that move large pieces over long derivations moving small pieces
- Resource Restriction pulls in the opposite direction: favors longer derivations with small extractions (also favored by combined costs!)

in dynamics operations weight $t^{\mathfrak{c}}$ for $t > 0$ weight parameter: $t \to 0$ leading terms

Optimality violations of Sideward Merge (Resource Restriction)

$\Phi : \mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0}) \to \mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0})$ transformation of workspaces (in particular $\Phi = \mathfrak{M}_{S,S'}$) satisfies

- *no divergence* if $b_0(\Phi(F)) \leq b_0(F)$, the number of components is non-increasing, a condition that ensures that derivations consisting of iterations of such transformations do not diverge;
- *no information loss* if $\alpha(\Phi(F)) \geq \alpha(F)$, the number of accessible terms is non-decreasing, namely no amount of syntactic information is lost in the process.
- *Minimal Yield*: for combined size $\sigma(F) = \alpha(F) + b_0(F)$ minimal change is $\sigma(\Phi(F)) = \sigma(F) + 1$

Type of Merge	*Coproduct*	b_0	α	σ
External	Δ^c and Δ^d	-1	$+2$	$+1$
Internal	Δ^c	0	$+1$	$+1$
Internal	Δ^d	0	0	0

Δ^d and Δ^c two different forms of the coproduct (cancellation of the deeper copies at the two interfaces: trace at CI and no trace in externalization)

Merge	*Coproduct*	b_0	α	σ
SM(1)	Δ^c	0	$+1$	$+1$
SM(1)	Δ^d	0	0	0
SM(2)	Δ^c	$+1$	0	$+1$
SM(2)	Δ^d	$+1$	-2	-1
SM(3)	Δ^c	$+1$	0	$+1$
SM(3)	Δ^d	$+1$	-2	-1

first form of SM is indistinguishable from IM in terms of size counting

second look at the Extension Condition

1. syntactic composition always *grows* the structure resulting in more complex components
2. this *growth only happens at the root* of the tree(not by insertions at any lower vertices/edges),

second property already discussed (in terms of grafting $\mathfrak{B}$); first property distinct soft constraint: no complexity loss

- start with workspace $F = \sqcup_a T_a$ with component T_a
- root vertex v_a of the component T_a becomes, track where it goes under Φ
- it becomes a vertex of some component $T'_{a'}$ of new workspace
- no complexity loss: new component more complex (Hopf algebra degree as proxy for tree complexity measure)

$$\deg(T'_{a'}) = \#L(T_{a'}) \geq \#L(T_a) = \deg(T_a)$$

- satisfied by EM and IM: fails for *all* forms of SM

Question: relation between these constraints

- different optimization constraints: Minimal Search and Resource Restriction (incl No Complexity Loss)
- are they all needed?
- are they independent?
- is any of them more fundamental?

this is not obvious in Merge & SMT

but can be compared precisely in math formulation of Merge:

- Minimal Search and Resource Restriction are not redundant constraints and play different roles
- different SM types do not simultaneously minimize both: MS favors shorter derivation, RR favors smaller extraction (possibly causing longer derivations)
- but in asymptotic behavior of the dynamics, both achieve the same effect (more on this later)

Ranking forms of Sideward Merge by distance from optimality

- observed that SM(1) behaves "like IM" in RR (b_0, α, σ)
- but violates NCL: by how much?
- $F = T \sqcup T' \mapsto \mathfrak{M}(T_v, T') \sqcup T/T_v$, root of T' goes in $\mathfrak{M}(T_v, T')$ with $\deg(\mathfrak{M}(T_v, T')) > \deg(T')$ no NCL violation, but root of T goes to T/T_v with $\deg(T/T_v) = \deg(T) - \deg(T_v) < \deg(T)$ (NCL violation)
- smallest violation for $\deg(T_v) = 1$ (atomic element)
- how far are SM(2) & SM(3)? both have larger NCL violations by $\deg(T_v) + \deg(T_w) \geq 2$ (and violations of optimality in the counting b_0, α, σ)

Ranking linguistic phenomena by optimality violation

- all the four classes of phenomena discussed have a proposed SM-based derivation that involves a composition EM∘SM that replaces an IM-type movement operation
- just seen that RR & NCL select SM(1) with an atomic $T_v = \alpha$ as closest to optimal: resulting composition EM∘SM

$$\mathfrak{M}(\mathfrak{M}(T', \alpha), T/\alpha)$$

- then evaluate how far from IM is this EM∘SM
- how to evaluate if "IM-like"?
- IM takes a T to a $\mathfrak{M}(T_v, T/T_v)$ with $\deg(\mathfrak{M}(T_v, T/T_v)) = \deg(T)$
- how far from being degree preserving?
- $\deg(\mathfrak{M}(\mathfrak{M}(T', \alpha), T/\alpha)) - \deg(T) = \deg(T')$ so closest to IM when $T' = \beta$ is also atomic

$$\mathfrak{M}(\mathfrak{M}(\beta, \alpha), T/\alpha)$$

resulting ranking by optimality violations (by RR, NCL, and deg)

- *smallest overall violations:* Head-to-Head movement

$$\mathfrak{M}(\mathfrak{M}(\beta,\alpha), T/\alpha)$$

- *smallest RR/NCL violations, but larger* deg *violation*: Head-to-Phrase Movement

$$\mathfrak{M}(\mathfrak{M}(T',\alpha), T/\alpha)$$

- *smallest* deg *violation, larger RR/NCL violations*: Phrase-to-Head Movement

$$\mathfrak{M}(\mathfrak{M}(\beta, T_v), T/T_v)$$

- *both larger* deg *and RR/NCL violations*: Phrase-to-Phrase Movement

$$\mathfrak{M}(\mathfrak{M}(T', T_v), T/T_v)$$

same as what would be expected from linguistic considerations

is Sideward Merge needed for structural reasons?

are (small) optimality violations structurally needed?
YES ... for Markovian property of Merge

- in Merge & SMT it is assumed that "Merge is Markovian" just qualitatively as "memoryless" process (action only depends on current workspace)
- because of the underlying Hopf algebra structure, Markovian also has a more precise meaning
- Hopf algebra Markov chains: introduced in
  - Persi Diaconis, C. Y. Amy Pang, and Arun Ram, "Hopf Algebras and Markov Chains: Two Examples and a Theory." *Journal of Algebraic Combinatorics*, 39, no. 3 (2014): 527–585.
- shown in MCB that Merge action on workspaces is a Hopf algebra Markov chain in this sense
- but... proof requires full $\mathfrak{M}_{S,S'}$ that includes SM in addition to EM and IM

How does that work? (what role for optimality constraints?)

What is a Hopf algebra Markov chain?

- Hopf algebra $\mathcal{H}$ with a *linear* map $\mathcal{K} = \sqcup \circ \mathcal{Q} \circ \Delta$ (uses coproduct and product to decompose and recompose and in between a transformation $\mathcal{Q}$ acts on the pieces of the decomposition)
- $\mathcal{H}$ has a preferred linear basis X: matrix representation $\mathcal{K} = (K_X(x,y))_{x,y\in X}$ in that basis
- $K_X(x,y) \geq 0$ and, for all $x \in X$ there is at least one $y \in X$ such that $K_X(x,y) > 0$

- K_X has Perron–Frobenius eigenfunction $\eta = \sum_x \eta(x)\, x$ with eigenvalue $\lambda > 0$:

$$\sum_y K_X(x,y)\eta(y) = \lambda\, \eta(x) \text{ with } \eta(x) > 0,\ \forall x \in X$$

- $x \mapsto x/\eta(x)$ rescales basis to get *stochastic matrix*:

$$\hat{K}_X(x,y) = \frac{1}{\lambda}\frac{\eta(y)}{\eta(x)} K_X(x,y)$$

$$\sum_y \hat{K}_X(x,y) = \frac{1}{\lambda}\frac{1}{\eta(x)} \sum_y K_X(x,y)\eta(y) = 1$$

to check: Hopf algebra Markov chain need to check these properties, tricky bit is Perron–Frobenius

Perron–Frobenius theorem (non-negative matrices)

- $N \times N$ real matrix $A = (A_{ij})$ with $A_{ij} \geq 0$
- irreducible: $\forall i, j\ \exists m \geq 1$ with $(A^m)_{ij} > 0$
- equivalent property: associate to A a graph G_A with N vertices and oriented edge from i to j iff $A_{ij} > 0$, then A irreducible means that G_A is strongly connected: given any two vertices there is an oriented path from the first to the second
- period h_A of A is greatest common divisor of the m such that $(A^m)_{ii} > 0$ (independent of i if A irreducible)

then...

Perron-Frobenius:

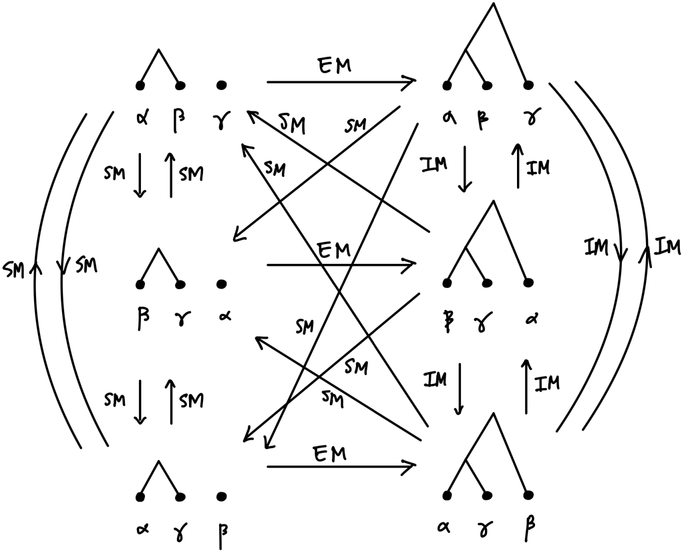
- $\exists$ PF-eigenvalue $\lambda_A > 0$ simple ($=$ spectral radius $\rho(A)$)
- $\exists$ left/right PF-eigenvector $\underline{w}_A, \underline{v}_A$ with positive entries
- these are the only eigenvectors that are positive (up to scalar multiples)
- there are h_A complex eigenvalues λ with $|\lambda| = \lambda_A$, each $\lambda = \lambda_A \zeta$ (ζ root of 1) and simple

So this says the key to this property is a geometric condition: graph G_A is strongly connected

to show Merge is a Hopf algebra Markov chain: need to show it has this property

for action on workspaces (starting with a fixed set of lexical items/features in $\mathcal{SO}_0$; also Merge always builds structure so only look at action on workspaces with nontrivial set of edges: can never return to zero edges through Merge)... SM is needed for strong connectedness

simplest example Merge action on deg = 3 workspaces



$$K_X = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Perron-Frobenius eigenvalue $\lambda' = 2 + \sqrt{2}$ and Perron-Frobenius eigenvector $\eta' = (\sqrt{2}, \sqrt{2}, \sqrt{2}, 1, 1, 1)^\tau$ (column)

$$\hat{K}_X = \frac{1}{\lambda} \frac{\eta(y)}{\eta(x)} K_X(x,y) = \begin{pmatrix} 0 & \frac{1}{2+\sqrt{2}} & \frac{1}{2+\sqrt{2}} & 0 & \frac{1}{2+2\sqrt{2}} & \frac{1}{2+2\sqrt{2}} \\ \frac{1}{2+\sqrt{2}} & 0 & \frac{1}{2+\sqrt{2}} & \frac{1}{2+2\sqrt{2}} & 0 & \frac{1}{2+2\sqrt{2}} \\ \frac{1}{2+\sqrt{2}} & \frac{1}{2+\sqrt{2}} & 0 & \frac{1}{2+2\sqrt{2}} & \frac{1}{2+2\sqrt{2}} & 0 \\ \frac{\sqrt{2}}{2+\sqrt{2}} & 0 & 0 & 0 & \frac{1}{2+\sqrt{2}} & \frac{1}{2+\sqrt{2}} \\ 0 & \frac{\sqrt{2}}{2+\sqrt{2}} & 0 & \frac{1}{2+\sqrt{2}} & 0 & \frac{1}{2+\sqrt{2}} \\ 0 & 0 & \frac{\sqrt{2}}{2+\sqrt{2}} & \frac{1}{2+\sqrt{2}} & \frac{1}{2+\sqrt{2}} & 0 \end{pmatrix}$$

bistochastic so uniform distribution $\xi(x) = 1$ stationary distribution of the Markov chain $\xi \hat{K}_X = \xi$ ... not so interesting... but not accounting for optimality constraints !

Optimality constraints in the Hopf algebra Markov chain

cost functions: Minimal Search, Minimal Yield, No Complexity Loss

$$K^{c_{MS}}_{X,t} = \begin{pmatrix} 0 & 1 & 1 & 0 & t^{1/3} & t^{1/3} \\ 1 & 0 & 1 & t^{1/3} & 0 & t^{1/3} \\ 1 & 1 & 0 & t^{1/3} & t^{1/3} & 0 \\ 1 & 0 & 0 & 0 & t^{1/2} & t^{1/2} \\ 0 & 1 & 0 & t^{1/2} & 0 & t^{1/2} \\ 0 & 0 & 1 & t^{1/2} & t^{1/2} & 0 \end{pmatrix} \quad \text{with Minimal Search costs}$$

$$K^{c_{MY}}_{X,t} = \begin{pmatrix} 0 & 1 & 1 & 0 & t^{-1} & t^{-1} \\ 1 & 0 & 1 & t^{-1} & 0 & t^{-1} \\ 1 & 1 & 0 & t^{-1} & t^{-1} & 0 \\ t & 0 & 0 & 0 & 1 & 1 \\ 0 & t & 0 & 1 & 0 & 1 \\ 0 & 0 & t & 1 & 1 & 0 \end{pmatrix} \quad \text{with Minimal Yleld costs}$$

$$K^{c_{CL}}_{X,t} = \begin{pmatrix} 0 & 1 & 1 & 0 & t^2 & t^2 \\ 1 & 0 & 1 & t^2 & 0 & t^2 \\ 1 & 1 & 0 & t^2 & t^2 & 0 \\ 1 & 0 & 0 & 0 & t & t \\ 0 & 1 & 0 & t & 0 & t \\ 0 & 0 & 1 & t & t & 0 \end{pmatrix} \quad \text{with Complexity Loss costs}$$

$$K^{c}_{X,t} = \begin{pmatrix} 0 & 1 & 1 & 0 & t^{4/3} & t^{4/3} \\ 1 & 0 & 1 & t^{4/3} & 0 & t^{4/3} \\ 1 & 1 & 0 & t^{4/3} & t^{4/3} & 0 \\ t & 0 & 0 & 0 & t^{3/2} & t^{3/2} \\ 0 & t & 0 & t^{3/2} & 0 & t^{3/2} \\ 0 & 0 & t & t^{3/2} & t^{3/2} & 0 \end{pmatrix} \quad \text{with all costs combined}$$

same basic calculation for all cases

$$K_{a,b,c}(t) := \begin{pmatrix} 0 & 1 & 1 & 0 & t^a & t^a \\ 1 & 0 & 1 & t^a & 0 & t^a \\ 1 & 1 & 0 & t^a & t^a & 0 \\ t^c & 0 & 0 & 0 & t^b & t^b \\ 0 & t^c & 0 & t^b & 0 & t^b \\ 0 & 0 & t^c & t^b & t^b & 0 \end{pmatrix}$$

- Perron-Frobenius eigenvector $\eta = (u, u, u, 1, 1, 1)^\tau$

$$u = t^{-c}(1 - t^b + ((1 - t^b)^2 + 2t^a t^b)^{1/2})$$

- Perron-Frobenius eigenvalue

$$\lambda = 1 + t^b + ((1 - t^b)^2 + 2t^a t^b)^{1/2}$$

$$\hat{K}_{abc}(t) = \begin{pmatrix} 0 & \lambda^{-1} & \lambda^{-1} & 0 & \lambda^{-1}u^{-1}t^a & \lambda^{-1}u^{-1}t^a \\ \lambda^{-1} & 0 & \lambda^{-1} & \lambda^{-1}u^{-1}t^a & 0 & \lambda^{-1}u^{-1}t^a \\ \lambda^{-1} & \lambda^{-1} & 0 & \lambda^{-1}u^{-1}t^a & \lambda^{-1}u^{-1}t^a & 0 \\ \lambda^{-1}ut^c & 0 & 0 & 0 & \lambda^{-1}t^b & \lambda^{-1}t^b \\ 0 & \lambda^{-1}ut^c & 0 & \lambda^{-1}t^b & 0 & \lambda^{-1}t^b \\ 0 & 0 & \lambda^{-1}ut^c & \lambda^{-1}t^b & \lambda^{-1}t^b & 0 \end{pmatrix}$$

stationary distribution $\xi \hat{K}_{abc} = \xi$ with $\xi = Z^{-1}(v, v, v, 1, 1, 1)$ with $Z = 3v + 3$

resulting behavior for $t \to 0$ (dominant terms)

- Minimal Search case:

$$\frac{v_{\rm MS}}{Z_{\rm MS}} \sim \frac{1}{3} - \frac{t^{5/6}}{6} - \frac{t^{4/3}}{3} + \frac{t^{5/3}}{4} + \cdots \qquad \frac{1}{Z_{\rm MS}} \sim \frac{t^{5/6}}{6} + \frac{t^{4/3}}{3} - \frac{t^{5/3}}{4} + \cdots$$

- RR, MY: $v_{\rm MY} = 1$ (bistochastic) weights t and t^{-1} of EM and SM(3) cancel
- NCL case:

$$\frac{v_{\rm CL}}{Z_{\rm CL}} \sim \frac{1}{3} - \frac{1}{6} t^3 - \frac{1}{3} t^4 + \cdots \qquad \frac{1}{Z_{\rm CL}} \sim \frac{1}{6} t^3 + \frac{1}{3} t^4 + \cdots$$

- combined costs:

$$\frac{v_{\text{total cost}}}{Z_{\text{total cost}}} \sim \frac{1}{3} - \frac{t^{17/6}}{6} - \frac{t^{13/3}}{3} + \cdots \qquad \frac{1}{Z_{\text{total cost}}} \sim \frac{t^{17/6}}{6} + \frac{t^{13/3}}{3} + \cdots$$

resulting behavior for $t \to 0$ (dominant terms)

- MS, RR/MY, and CL cost functions are different independent soft constraints, weight different forms of SM differently, not same SM preferable
- weighting action of Merge by RR/MY has not effect on the dynamics: convergence to uniform distribution
- other cost functions MS, CL or combined give same limiting distribution: uniform on the *connected* structures (completed EM structure formation) with only remaining IM movement as dynamics
- so effect of SM becomes rare in long range dynamics (asymptotic distribution)
- additional result: for arbitrary size, the *strong connectedness* (hence Perron-Frobenius) property holds also if only use first two smallest violations

$$T \sqcup \beta \mapsto \mathfrak{M}(\beta, \alpha) \sqcup T/\alpha \quad \text{and} \quad T \sqcup T' \mapsto \mathfrak{M}(\beta, \alpha) \sqcup T/\alpha \sqcup T'/\beta$$

Alternatives to Optimality Violations? (no Sideward Merge)

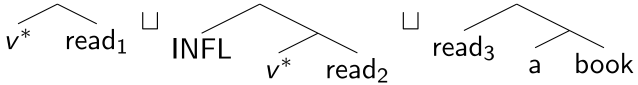
- so we have a coherent picture of SM as avoiding EC-violations, with different optimality violation costs, and necessary for structural reasons
- but... what if there are other possible derivations of the same phenomena that do *not* require SM and use only IM and EM ?
- this may be possible (Riny Huijbregts has some possible solutions)... but

these require additional algebraic structure

discuss an example: Amalgam in I-language or in Externalization?

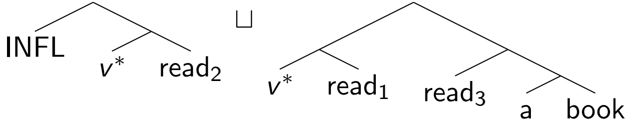
Amalgam in I-language proposal

$F_1 = \text{John} \sqcup$



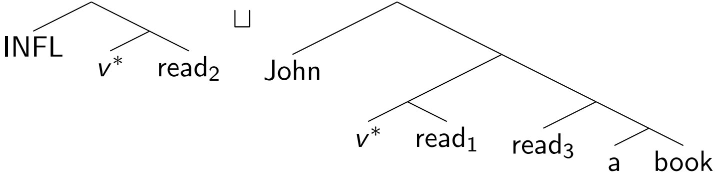
$\mathfrak{M}_{\{v^*,\text{read}_1\},\{\text{read}_3,\text{IA}\}}(F_1) = F_2$ External Merge

$= \text{John} \sqcup$



$\mathfrak{M}_{\text{EA},v^*P}(F_2) = F_3$ External Merge

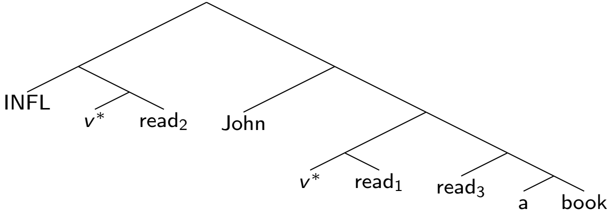
$F_3 =$



external argument (EA), internal argument (IA)

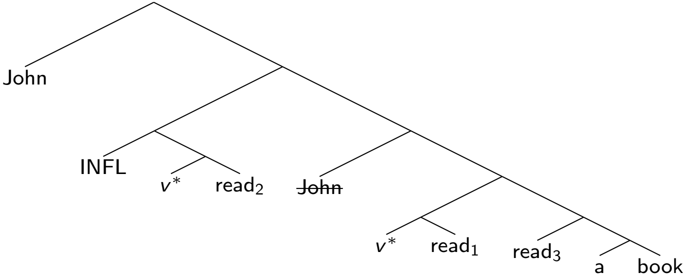
$\mathfrak{M}_{\{\mathrm{INFL},\{v^*,\mathsf{read}\}\},\{\mathrm{EA},\{\{v^*,\mathsf{read}\},\{\mathsf{read},\mathrm{IA}\}\}\}}(F_3) = F_4$ External

$F_4 =$



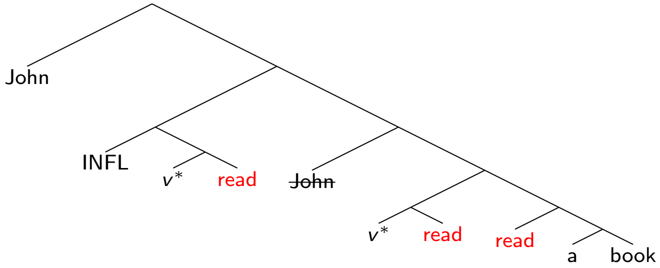
$\mathfrak{M}_{\mathrm{EA},\mathsf{INFL\text{-}P}}(F_4) = F_5$ Internal Merge

$F_5 =$



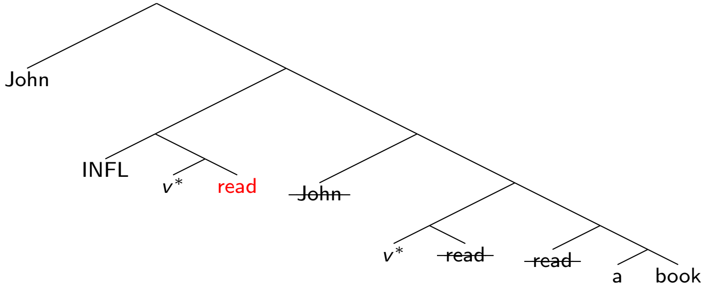
then FormCopy

$F_5 =$



cancellation of deeper copies (this is the tricky bit)

$F_5 =$

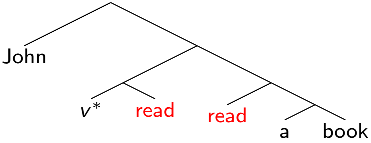


externalized as *John [[read] -s] a book*

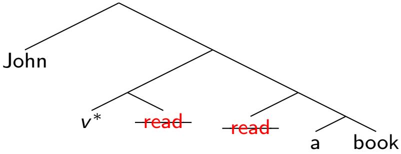
What about implementing this in the mathematical Minimalism?

- restriction to diagonals (FormCopy) is OK and coproduct only extracts once the identified terms
- but cancellation of deeper copies is only realized by the coproduct, so it can only come together with an extraction: so derivation cannot be performed in that order
- if introduce a separate operation of cancellation this is NOT zero-cost (higher cost than SM)... also issue of algebraic properties of this additional operation
- if pairing extraction and cancellation then it is an SM
- the algebraic nature of the model puts very strong structural constraints

pairing of extraction and cancellation (after restriction to diagonals) so it can be performed by a coproduct operation

$F = \text{INFL} \sqcup v^* \sqcup$ 

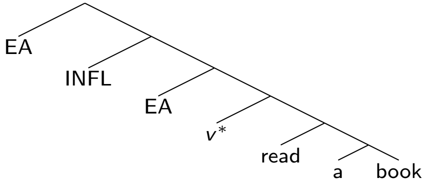
coproduct extraction and cancellation (relevant coproduct term)

$v^* \sqcup \text{read} \otimes \text{INFL} \sqcup$ 

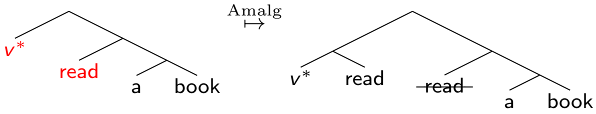
but now this is Sideward Merge

Amalgamation in Externalization proposal

- syntactic object generated by Merge



- head α locally c-commands head β if α c-commands β and no head γ ($\gamma \neq \alpha, \beta$) c-commanded by α and c-commands β
- $\mathrm{Amalg}(X, \{\alpha, \beta\})$ substitutes lexical morphology $\{\alpha, \beta\}$ for the syntactic head α of $T \in \mathcal{SO}$ with lexical morphology $\{\alpha, \beta\}$



- avoids the cancellation/extraction issue ... but what is this operation? Operad insertion at leaves + a cancellation?

Morphology-syntax interface as operad structure

- morphology also creates binary rooted tree structures
- items at the leaves of syntactic trees can be thought of as "modifiable by morphology"
- operation that happens after syntactic structure formation
- acting by insertion (of morphological structure) at the leaves
- insertion at leaves has its own algebraic structure (operad)
- interactions with syntax: word order and morphology (eg apparent free word order) etc
- but still... issue with cancellation operation that is NOT zero cost (at least as costly as SM)

Conclusion

as in physics, just let the algebraic structure do all the work for you: it is so highly constrained that it provides a way to navigate and quantitatively rank different linguistic hypotheses