

Lecture 11: Syntax–Semantics Interface

Ma 191c: Mathematical Models of Generative Linguistics

Matilde Marcolli

Caltech, Spring 2024

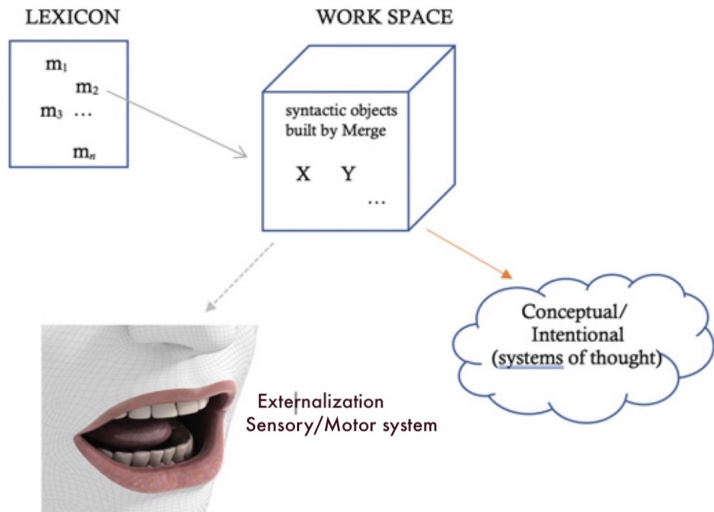
this part based on

- Matilde Marcolli, Robert C. Berwick, Noam Chomsky, *Syntax-semantics interface: an algebraic model*, arXiv:2311.06189

also included in the book:

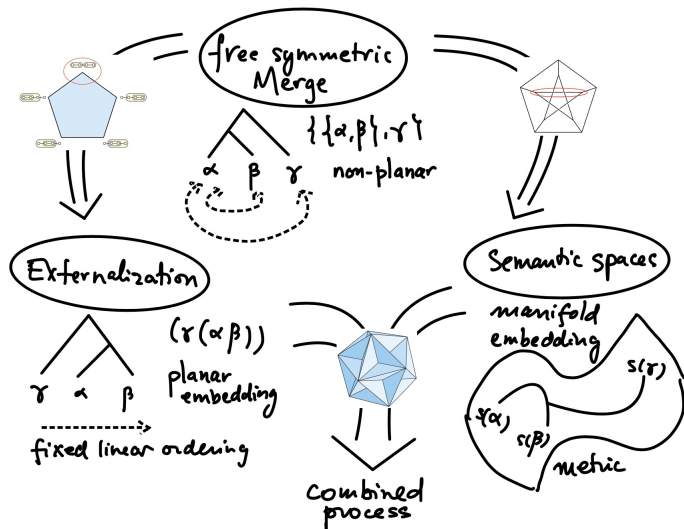
Matilde Marcolli, Noam Chomsky, Robert C. Berwick, “Mathematical structure of syntactic Merge”, MIT Press.

Syntax-Semantics interface (Conceptual-Intentional System)



(from Chomsky et al. "Merge & the Strong Minimalist Thesis")

Preview of what this same picture looks like in our setting:



Syntax-Semantics interface: conceptual requirements

(not all would agree, but we take these as our background assumptions)

- 1 Autonomy of syntax
 - 2 Syntax supports semantic interpretation
 - 3 Semantic interpretation is, to a large extent, independent of externalization
 - 4 Compositionality
- autonomy of syntax: Merge computational generative process of syntax independent of semantics
 - syntax-first view: syntax-semantic interface proceeds *from* syntax *to* semantics
 - two channels: from core Merge mechanism to Conceptual-Intentional system (syntax-semantics interface) and to Sensory-Motor system (externalization)
 - compositionality: consistency across syntactic sub-structures

Basic properties

our view of distinction between the roles of syntax and semantics:

- 1 Syntax is a computational process.
 - 2 Semantics is *not* a computational process and is in essence grounded on a notion of topological proximity.
- first statement is clear in the context of generative linguistics,
 - second claim means possible additional structures on the side of semantics (metric, linear, semiring, etc) instantiate or quantify proximity relations, don't play a computational generative role like syntax

key idea (“Birkhoff factorization”)

- start with an assignment of semantic values to lexical items $s : \mathcal{SO}_0 \rightarrow \mathcal{S}$
- want to extend this to an assignment for syntactic objects that is *consistent over substructures*
- *key idea*: incorporate consistency checking in an inductively defined function
- a simple mapping $\phi : \mathcal{SO} \rightarrow \mathcal{S}$ (which will include inconsistent cases) modified in a recursive way
- this recursive construction of consistency checking is known in physics as “Bogolyubov preparation”

(conceptually similar to e.g. truth-values assignments using trees in context free setting, but recursive checking is directly built into the value function, and context-free hypothesis is not needed)

- recursive construction of consistency checking needs two main ingredients:
 - ① a way of extracting substructures T_v of a given structure T :
checking separately T_v and T/T_v
 - ② a way of separating out, in the target space \mathcal{S} , agreement and disagreement (or to filter by levels of agreement and disagreement)
- Note: the first is on the side of syntax, the second on the side of semantics
- the first is provided by the *coproduct* (Hopf algebra structure); the second is some form of projection (known as Rota–Baxter structure)

Note: the Hopf algebra coproduct at the heart of Merge is both

- ① *structure builder*
- ② *parser*

consistency over substructures (an example)

- consider the sentences “France is a republic”, “France is hexagonal” and “France is a hexagonal republic”
- first two have clear unproblematic semantic parsing, the third seems awkward
- syntactic object of the form $T = \{a, \{b, \{c, d\}\}\}$, with a, b, c, d the lexical items France, is, hexagonal, republic
- Hopf algebra coproduct produces terms of the form $T_v \otimes T/T_v$ including

$$c \otimes \{a, \{b, d\}\} \quad \text{and} \quad d \otimes \{a, \{b, c\}\}$$

where quotient structures $\{a, \{b, d\}\}$ and $\{a, \{b, c\}\}$ are the two sentences “France is a republic” and “France is hexagonal” (uncontroversial semantic parsing)

- coproduct also also contains terms like

$$\{c, d\} \otimes \{a, b\}$$

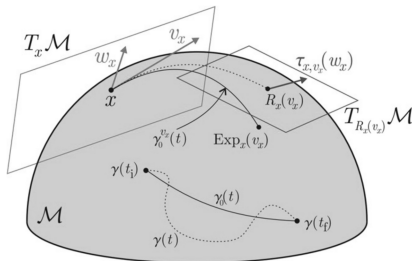
that track the precise location (the extracted term $\{c, d\}$ “hexagonal republic”) where the assignment of semantic values runs into problems

Basic model of semantic parsing

- main constructions:
 - 1 **semantic space** \mathcal{S} (topological/metric, interpolation/convexity): geodesically convex Riemannian manifold; mapping of syntactic objects (using head/phases)
 - 2 **evaluations** (of distances or agreement/disagreement): algebra or semiring \mathcal{R} with some “threshold operation” R (Rota-Baxter structure)
 - 3 **semantic probes**: extract evaluations (comparative) from image of syntactic objects in semantic space

Semantic spaces setup

- **distance/proximity** on \mathcal{S} and **convexity** (both ok in vector spaces)
- semantic space \mathcal{S} geodesically convex region in Riemannian manifold (M, g) (min length geodesic arcs between points)



- $\{\lambda s + (1 - \lambda)s' \mid \lambda \in [0, 1]\}$ segment between s, s' in \mathcal{S}
- suppose also have some **comparison functions**

$$\mathbb{P} : \mathcal{S} \times \mathcal{S} \rightarrow [0, 1] \quad \text{or} \quad \mathfrak{C} : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$$

similarity/relatedness \mathbb{P} or agreement/disagreement \mathfrak{C}

- e.g. *cosine similarity* in vector spaces $\mathfrak{C}(s, s') = \frac{\langle s, s' \rangle}{\|s\| \|s'\|}$

Semantic spaces

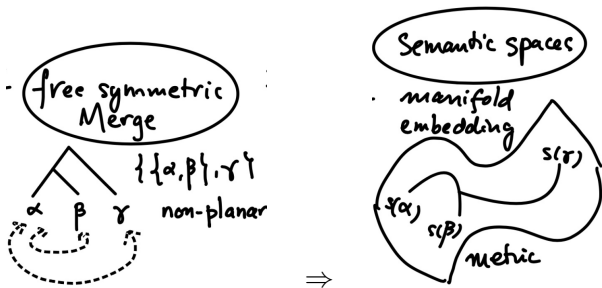
- in this setting assume that conceptual/semantic spaces exist independently of their interface with the language faculty: for instance for parsing of images, spatial data, other embodied perceptions
- the language calculator that is responsible for syntax interfaces with other structures responsible for detecting similarity/contrast and proximity relations
- semantics is an “open condition”: can vary assignment in semantic spaces depending on context etc, but can compare semantic values in an open neighborhood (proximity/interpolation)

Similar ideas on semantic/conceptual spaces

- Yu.I. Manin, M. Marcolli, *Semantic Spaces*, Math. Comput. Sci. 10 (2016), no. 4, 459–477.
- P. Gärdefors, *Conceptual spaces: the geometry of thought*, MIT Press, 2000.
- P. Gärdefors, *The geometry of meaning: semantics based on conceptual spaces*, MIT Press, 2014.
- S.Y. Chung, D.D. Lee, H. Sompolinsky, *Classification and geometry of general perceptual manifolds*, Phys. Rev. X 8 (2008), 031003.
- Andrea E. Martin, *Compositional neural architecture for language*, Journal of Cognitive Neuroscience 32 (2020) 8, 1407–1427

mapping syntactic objects to semantic space

- $s : \mathcal{SO}_0 \rightarrow \mathcal{S}$ (semantic values for lexical items)
- structures $T = \mathfrak{M}(\alpha, \beta)$: get $s(T)$ by
 - geodesic arc γ in \mathcal{S} between $s(\alpha)$ and $s(\beta)$
 - $T \in \text{Dom}(h)$ (head function: parsable) root vertex of T placed on arc γ close to $h(T)$ displaced according to value of $\mathbb{P}(s(\alpha), s(\beta))$ or $\mathcal{C}(s(\alpha), s(\beta))$



evaluations and thresholds (Rota–Baxter structures)

- **Rota–Baxter algebra** (\mathcal{R}, R) of weight -1 : commutative associative algebra \mathcal{R} , linear operator $R : \mathcal{R} \rightarrow \mathcal{R}$ with

$$R(a)R(b) = R(aR(b)) + R(R(a)b) - R(ab)$$

- effect: R and $1 - R$ spit \mathcal{R} into *subalgebras* \mathcal{R}_{\mp}
- **Rota–Baxter semiring** (\mathcal{R}, R)

$$R(a) \odot R(b) = R(a \odot R(b)) \boxminus R(R(a) \odot b) \boxminus R(a \odot b) \quad (\text{weight } +1)$$

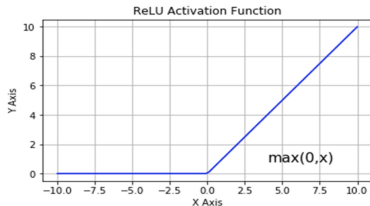
$$R(a) \odot R(b) \boxminus R(a \odot b) = R(a \odot R(b)) \boxminus R(R(a) \odot b) \quad (\text{weight } -1)$$

Meaning of this Rota–Baxter identity? it's the behavior of a “threshold operator”

Example: Max-plus (tropical) semiring and ReLU

- max-plus semiring (tropical semiring)

$$\mathcal{R} = (\mathbb{R} \cup \{-\infty\}, \max, +)$$



- ReLU operator $R : x \mapsto x^+ = \max\{x, 0\}$ is a Rota–Baxter operator of weight $+1$ on \mathcal{R}

	$x \leq 0, y \leq 0$	$x \geq 0, y \leq 0$	$x \leq 0, y \geq 0$	$x \geq 0, y \geq 0$
$x^+ + y^+$	0	x	y	$x + y$
$(x^+ + y)^+$	0	$\begin{cases} x + y & x + y \geq 0 \\ 0 & x + y \leq 0 \end{cases}$	y	$x + y$
$(x + y^+)^+$	0	x	$\begin{cases} x + y & x + y \geq 0 \\ 0 & x + y \leq 0 \end{cases}$	$x + y$
$(x + y)^+$	0	$\begin{cases} x + y & x + y \geq 0 \\ 0 & x + y \leq 0 \end{cases}$	$\begin{cases} x + y & x + y \geq 0 \\ 0 & x + y \leq 0 \end{cases}$	$x + y$
max	0	x	y	$x + y$

$$x^+ + y^+ = \max\{(x^+ + y)^+, (x + y^+)^+, (x + y)^+\}$$

Example: Boolean semiring

- Boolean semiring $\mathcal{B} = (\{0, 1\}, \vee, \wedge) = (\{0, 1\}, \max, \cdot)$
- identity as Rota–Baxter operator of weight -1 (no room for more subtle thresholds in this case)

Example: Viterbi (probabilistic) semiring and threshold

- Viterbi semiring $\mathcal{P} = ([0, 1], \max, \cdot, 0, 1)$
- threshold operators c_λ with $0 \leq \lambda \leq 1$

$$c_\lambda(x) = \begin{cases} x & x < \lambda \\ 1 & x \geq \lambda \end{cases}$$

- Rota–Baxter operators of weight -1

Viterbi semiring: Rota-Baxter identity for threshold

	$x < \lambda, y < \lambda$	$x \geq \lambda, y < \lambda$	$x < \lambda, y \geq \lambda$	$x \geq \lambda, y \geq \lambda$
$c_\lambda(xy)$	xy	xy	xy	$\begin{cases} xy & xy < \lambda \\ 1 & xy \geq \lambda \end{cases}$
$c_\lambda(x)c_\lambda(y)$	xy	y	x	1
$c_\lambda(c_\lambda(x)y)$	xy	y	xy	1
$c_\lambda(xc_\lambda(y))$	xy	xy	x	1

max first two rows:

$$\max\{c_\lambda(xy), c_\lambda(x)c_\lambda(y)\} = c_\lambda(x)c_\lambda(y)$$

max last two rows: $c_\lambda(x)c_\lambda(y)$

so Rota-Baxter structure of weight -1

Note: for RB weight -1 in semiring, additional condition

$$R(x \odot y) \square R(x) \odot R(y) = R(x) \odot R(y)$$

(for Birkhoff factorization to work well): Boolean and Viterbi OK

semantic probes: example

- semantic space \mathcal{S} has *probes* by functions $\Upsilon : \mathcal{S} \rightarrow \mathbb{R}$ checking degree of agreement or disagreement with a particular semantic hypothesis
- eg a chosen vector v_Υ in vector space models

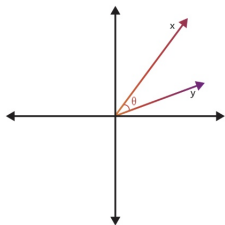
$$\Upsilon(s) = \langle s, v_\Upsilon \rangle$$

- given a semantic space \mathcal{S} , a probe $\Upsilon : \mathcal{S} \rightarrow \mathbb{R}$, a map $s : \text{Dom}(h) \subset \mathcal{SO} \rightarrow \mathcal{S}$ assigning semantic values to (parsable) syntactic objects
- build from these a *test of semantic agreement/disagreement*

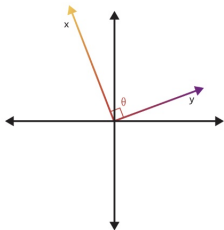
$$\Upsilon_{s,h} : T \mapsto \begin{cases} \Upsilon(s(T)) = \langle s(T), v_\Upsilon \rangle & T \in \text{Dom}(h) \\ -\infty & T \notin \text{Dom}(h). \end{cases}$$

$$\phi_{\Upsilon,s,h} : \mathfrak{F}\mathcal{SO}_0 \rightarrow \mathbb{R} \cup \{-\infty\}$$

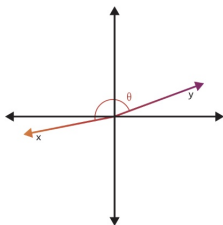
$$\phi_{\Upsilon,s,h}(F) = \sum_a \Upsilon_{s,h}(T_a), \quad \text{for } F = \sqcup_a T_a$$



Angle θ close to 0°
 $\text{Cos}(\theta)$ close to 1 → **Similar vectors**



Angle θ close to 90°
 $\text{Cos}(\theta)$ close to 0 → **Orthogonal vectors**



Angle θ close to 180°
 $\text{Cos}(\theta)$ close to -1 → **Opposite vectors**

inner product $\langle s, v_\gamma \rangle$ with a fixed probe vector v_γ detects agreement/disagreement (can normalize to cosine similarity)

can similarly evaluate agreement/disagreement $\langle s(T), s(T') \rangle$

Note: semantic probes define characters (measuring agreement/disagreement with respect to given probe)

characters

- start with $\phi : \mathcal{H} \rightarrow \mathcal{R}$ where \mathcal{H} Hopf algebra and \mathcal{R} Rota-Baxter algebra ... **but** ϕ only morphism of commutative algebras (multiplicative: independent structures go to independent values)
- these maps are called *characters* of the Hopf algebra
- **Note:** target \mathcal{R} does not have same kind of “computational” structure as source \mathcal{H} (no coproduct operation), but has RB projection R
- **Birkhoff factorization** of ϕ into $\phi_{\pm} : \mathcal{H} \rightarrow \mathcal{R}_{\pm}$ (alg homom)

$$\phi = (\phi_- \circ S) \star \phi_+ \quad \text{or} \quad \phi_+ = \phi_- \star \phi$$

- product \star is determined by coproduct of \mathcal{H}

$$(\phi_1 \star \phi_2)(x) = (\phi_1 \otimes \phi_2) \Delta(x)$$

Birkhoff factorization (using coproduct of \mathcal{H} and RB of \mathcal{R})

- Birkhoff factorizations are constructed inductively

$$\phi_-(x) = -R(\phi(x) + \sum \phi_-(x')\phi(x''))$$

$$\phi_+(x) = (1 - R)(\phi(x) + \sum \phi_-(x')\phi(x''))$$

where $\Delta(x) = 1 \otimes x + x \otimes 1 + \sum x' \otimes x''$ with x', x'' of lower degree

- Bogolyubov preparation is the expression

$$\tilde{\phi}(x) := \phi(x) + \sum \phi_-(x')\phi(x'')$$

that inductively incorporates the search for possible inconsistencies in substructures

Semirings

- character $\phi : \mathcal{H}^{semi} \rightarrow \mathcal{R}$ morphism from a sub-semiring of \mathcal{H}
- **semiring case**: inductive Birkhoff factorization

$$\phi_{-}(x) = R(\tilde{\phi}(x)) = R(\phi(x) \boxtimes \phi_{-}(x') \odot \phi(x''))$$

$$\begin{aligned}\phi_{+}(x) &= (\phi_{-} \star \phi)(x) = \phi(x) \boxtimes \phi_{-}(x) \boxtimes \phi_{-}(x') \odot \phi(x'') \\ &= \phi_{-} \boxtimes \tilde{\phi}\end{aligned}$$

multiplicative with respect to the semiring product

$$\phi_{\pm}(xy) = \phi_{\pm}(x) \odot \phi_{\pm}(y)$$

ϕ_{-} part is detecting *where right/wrong agreement*: substructures where consistent/inconsistent parsing is located (for a given probe and evaluation)

Boolean example

- Boolean semiring $\mathcal{B} = (\{0, 1\}, \vee, \wedge) = (\{0, 1\}, \max, \cdot)$
- map $\phi : \mathfrak{T}_{\mathcal{SO}_0} \rightarrow \mathcal{B}$ is an assignment of truth values extended to $\phi : \mathfrak{F}_{\mathcal{SO}_0} \rightarrow \mathcal{B}$ by $\phi(F) = \prod_i \phi(T_i)$ for $F = \sqcup_i T_i$
- identity as Rota–Baxter operator of weight -1
- Bogolyubov preparation

$$\tilde{\phi}(T) = \max\{\phi(T), \phi(F_{\underline{v}})\phi(T/F_{\underline{v}}), \dots, \phi(F_{\underline{v}_N})\phi(F_{\underline{v}_{N-1}}/F_{\underline{v}_N}) \cdots \phi(T/F_{\underline{v}_1})\}$$

- again example of “France is a hexagonal republic”

$$\begin{aligned} \tilde{\phi}\left(\begin{array}{c} \diagup \quad \diagdown \\ a \quad \quad b \quad \quad c \quad \quad d \end{array}\right) &= \max\left\{\phi\left(\begin{array}{c} \diagup \quad \diagdown \\ a \quad \quad b \quad \quad c \quad \quad d \end{array}\right), \phi(a)\phi\left(\begin{array}{c} \diagup \quad \diagdown \\ b \quad \quad c \quad \quad d \end{array}\right), \right. \\ &\quad \phi(c)\phi\left(\begin{array}{c} \diagup \quad \diagdown \\ a \quad \quad b \quad \quad d \end{array}\right), \phi(d)\phi\left(\begin{array}{c} \diagup \quad \diagdown \\ a \quad \quad b \quad \quad c \end{array}\right), \\ &\quad \left. \phi\left(\begin{array}{c} \diagup \quad \diagdown \\ b \quad \quad c \quad \quad d \end{array}\right)\phi(a), \phi\left(\begin{array}{c} \diagup \quad \diagdown \\ c \quad \quad d \end{array}\right)\phi\left(\begin{array}{c} \diagup \quad \diagdown \\ a \quad \quad b \end{array}\right), \dots\right\} \end{aligned}$$

where ... stands for terms of the coproduct involving a forest

- $0/1 = \text{consistent/inconsistent}$ identifies the only consistent substructures “France is hexagonal” and “France is a republic”

toy model example in vector space semantics

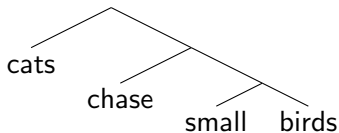
- use probes $\Upsilon(s) = \langle s, v_\Upsilon \rangle$
- character

$$\phi_{\Upsilon, s, h}(T) = \begin{cases} \Upsilon(s(T)) = \langle s(T), v_\Upsilon \rangle & T \in \text{Dom}(h) \\ -\infty & \text{otherwise.} \end{cases}$$

- Birkhoff factorization of $\phi_{\Upsilon, s, h}$ with respect to ReLU Rota–Baxter operator
- $\phi_{\Upsilon, s, h, -}(T)$ identifies maximum value $\phi_{\Upsilon, s, h}(F_{\underline{v}_N}) + \phi_{\Upsilon, s, h}(F_{\underline{v}_{N-1}}) + \dots + \phi_{\Upsilon, s, h}(F_{\underline{v}_1}) + \phi_{\Upsilon, s, h}(T)$ over all nested sequences with all $\phi_{\Upsilon, s, h}(F_{\underline{v}_i}) > 0$
- identifies where are chains of substructures where maximum consistent agreement with the chosen probe is achieved
- **problem:** this only uses the *head* $h(T)$ for location in semantic space (not right: values of lexical items only): see the issue better in some examples

example

- sentence like



- lexical items as vectors in a vector space model of semantics
- probe vector v_γ associated to “predation”
- vectors v for “cats”, “chase”, “birds” positively correlate to v_γ probe (as predator, prey, hunting action)

extracting coproduct terms and building recursive factorization

$$\phi_{\tau,s,h,-}(T) = (\max\{\phi_{\tau,s,h}(T), \phi_{\tau,s,h}(F_{\underline{v}})^+ \phi_{\tau,s,h}(T/F_{\underline{v}}), \\ \dots, \phi_{\tau,s,h}(F_{\underline{v}_N})^+ \phi(F_{\underline{v}_{N-1}}/F_{\underline{v}_N}) \cdots \phi(T/F_{\underline{v}_1})\})^+$$

for example terms like

$$\phi_{\tau,s,h}(\text{chase} \vee \text{small birds})^+ \phi_{\tau,s,h}(\text{cats})$$

$$\phi_{\tau,s,h}(\text{small birds})^+ \phi_{\tau,s,h}(\text{cats} \vee \text{chase})$$

etc, are above ReLU threshold because of positive correlation: see good agreement with probe over substructures

$$\phi_{\tau,s,h,-}(\text{cats} \begin{array}{c} \diagup \quad \diagdown \\ \text{chase} \quad \begin{array}{c} \diagup \quad \diagdown \\ \text{small} \quad \text{birds} \end{array} \end{array}) =$$

$$(\max\{\phi_{\tau,s,h}(\text{cats} \begin{array}{c} \diagup \quad \diagdown \\ \text{chase} \quad \begin{array}{c} \diagup \quad \diagdown \\ \text{small} \quad \text{birds} \end{array} \end{array}), \phi_{\tau,s,h}(\text{cats})^+ \phi_{\tau,s,h}(\text{chase} \begin{array}{c} \diagup \quad \diagdown \\ \text{small} \quad \text{birds} \end{array})\},$$

$$\phi_{\tau,s,h}(\text{small})^+ \phi_{\tau,s,h}(\text{cats} \begin{array}{c} \diagup \quad \diagdown \\ \text{chase} \quad \text{birds} \end{array}), \phi_{\tau,s,h}(\text{birds})^+ \phi_{\tau,s,h}(\text{cats} \begin{array}{c} \diagup \quad \diagdown \\ \text{chase} \quad \text{small} \end{array})\},$$

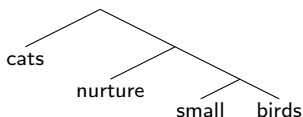
$$\phi_{\tau,s,h}(\text{chase} \begin{array}{c} \diagup \quad \diagdown \\ \text{small} \quad \text{birds} \end{array})^+ \phi_{\tau,s,h}(\text{cats}),$$

$$\phi_{\tau,s,h}(\text{small} \begin{array}{c} \diagup \quad \diagdown \\ \text{birds} \end{array})^+ \phi_{\tau,s,h}(\text{cats} \begin{array}{c} \diagup \quad \diagdown \\ \text{chase} \end{array}), \dots\}^+$$

where ... stands for terms of the coproduct involving a forest

here *head* is either verb “chase” or noun “bird” on the subtrees and if R is ReLU (or some threshold) get agreement with probe on substructures

- take same sentence but replace “chase” with “nurture”



- with the same probe v_r semantic vector of “predation”
- now some substructures are filtered out by ReLU because disagreement with probe

$$\phi_{r,s,h}(\text{nurture } \text{small } \text{birds})^+ = 0$$

- also final application of $R = \text{ReLU}$ to max in $\phi_{r,s,h,-} = R(\tilde{\phi}_{r,s,h})$ also cuts off all structures with *head* the verb “nurture” because of disagreement with probe

$$\phi_{r,s,h}(\text{cats } \text{nurture})^+ = 0, \quad \phi_{r,s,h}(\text{cats } \text{nurture } \text{birds})^+ = 0, \quad \text{etc}$$

but ... **another example**

- consider again example of “France is a hexagonal republic”
- in a vector space model of semantics all the lexical items “France”, “hexagon(al)”, “republic” have corresponding vectors, say v_{Fr} , v_{hex} , $v_{rep} \in \mathcal{S}$
- choose a probe v_{Γ} , for example the vector associated to “government”
- would have $\langle v_{\Gamma}, v_{Fr} \rangle > 0$, $\langle v_{\Gamma}, v_{rep} \rangle > 0$, but $\langle v_{\Gamma}, v_{hex} \rangle \leq 0$
- same tree structure... *but* now see that just looking at the *head* leaf of the trees does *not* help detect where problem
- **what is needed:** one can see in this example that what one wants to have is a point $s(T) \in \mathcal{S}$ that displaces the position of $s(h(T))$ in the direction of the *complement* of the head, just slightly if large agreement, a lot if disagreement so that $\phi_{\Gamma, s, h}(T)$ can differ significantly from $\phi_{\Gamma, s, h}(h(T))$ if there is disagreement inside the structure T

first improvement: use topology, metric, convexity, interpolation on the semantic space

As discussed before assume a semantic space with:

- use some notion of **distance/proximity** on \mathcal{S} and **convexity** (both ok in vector spaces)
- semantic space \mathcal{S} geodesically convex region in Riemannian manifold (M, g) (min length geodesic arcs between points)
- $\{\lambda s + (1 - \lambda)s' \mid \lambda \in [0, 1]\}$ segment between s, s' in \mathcal{S}
- also suppose have some **comparison functions**

$$\mathbb{P} : \mathcal{S} \times \mathcal{S} \rightarrow [0, 1]$$

$$\mathfrak{C} : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$$

likelihood of relatedness \mathbb{P} or degree of agreement/disagreement \mathfrak{C} (also ok in vector spaces for instance $\mathfrak{C}(s, s') = \frac{\langle s, s' \rangle}{\|s\| \|s'\|}$)

- discuss the case with \mathbb{P} (case of \mathfrak{C} similar)

- Viterbi semiring $\mathcal{P} = ([0, 1], \max, \cdot, 0, 1)$
- threshold operators c_λ with $0 \leq \lambda \leq 1$

$$c_\lambda(x) = \begin{cases} x & x < \lambda \\ 1 & x \geq \lambda \end{cases}$$

- Rota–Baxter operators of weight -1
- extend $s : \mathcal{SO}_0 \rightarrow \mathcal{S}$ to a map $s : \text{Dom}(h) \rightarrow \mathcal{S}$ inductively
 - for $\alpha, \beta \in \mathcal{SO}_0$

$$T = \mathfrak{M}(\alpha, \beta) = \alpha \frown \beta \Rightarrow s(T) = p s(\alpha) + (1 - p) s(\beta)$$

$$p = \begin{cases} p_{\alpha, \beta} & \alpha = h(T) \\ 1 - p_{\alpha, \beta} & \beta = h(T) \end{cases} \quad \text{with } p_{\alpha, \beta} := \mathbb{P}(s(\alpha), s(\beta))$$

- then inductively if $T = \mathfrak{M}(\alpha, \beta) \in \text{Dom}(h) \subset \mathcal{SO}$ take

$$s(T) = p s(T_1) + (1 - p) s(T_2) \quad \text{and} \quad p_{s(T_1), s(T_2)} = \mathbb{P}(s(T_1), s(T_2))$$

$$p = \begin{cases} p_{s(T_1), s(T_2)} & h(T) = h(T_1) \\ 1 - p_{s(T_1), s(T_2)} & h(T) = h(T_2) \end{cases}$$

- Viterbi valued characters $\phi_{s,\mathbb{P},h} : \mathfrak{TSO}_0 \rightarrow \mathcal{P}$
(used for probabilistic assignments)

$$\phi_{s,\mathbb{P},h}(T) = p_{s(T_1),s(T_2)}$$

and $\phi_{s,\mathbb{P},h}(T) = 0$ outside $\text{Dom}(h)$

- Birkhoff factorization of $\phi_{s,\mathbb{P},h}$ with respect to threshold Rota–Baxter operators identifies substructures with large semantic agreement between constituent parts

$$c_\lambda(\max\{\phi_{s,\mathbb{P},h}(T), c_\lambda(\phi_{s,\mathbb{P},h}(T_v)) \cdot \phi_{s,\mathbb{P},h}(T/T_v)\}) =$$

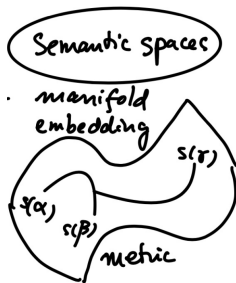
$$c_\lambda(\max\{p_{s(T_1),s(T_2)}, c_\lambda(p_{s(T_{v,1})s(T_{v,2}))} \cdot p_{s(T_1),s(T_2)}\}) = c_\lambda(p_{s(T_1),s(T_2)})$$

max value when all $p_{s(T_{v,1})s(T_{v,2})} \geq \lambda$ and $p_{s(T_1),s(T_2)} \geq \lambda$

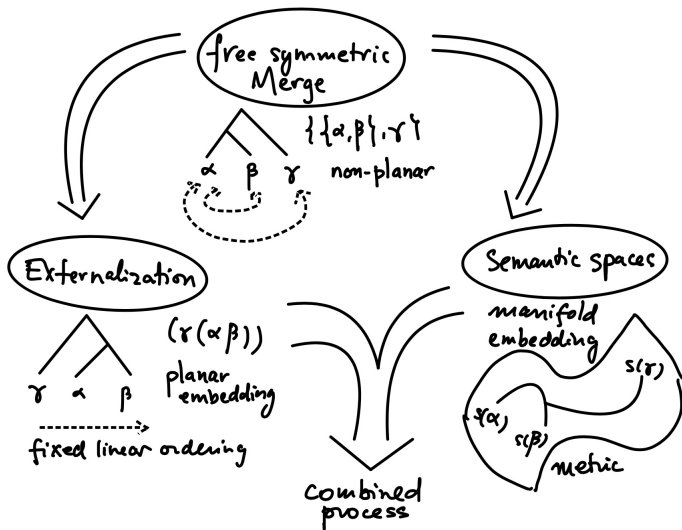
- maximizers are accessible terms that carry large semantic agreement between their constituent parts
- similar example with \mathfrak{C} and ReLU Rota–Baxter operator

image of syntax inside semantics

- semantic space \mathcal{S} geodesically convex Riemannian manifold
- if for $s \neq s'$ one has $\mathbb{P}(s, s') \in (0, 1)$
- obtain embeddings of trees $T \in \text{Dom}(h) \subset \mathcal{T}_{SO_0}$ inside \mathcal{S}
- edges geodesic arcs, vertices points
 $s(T) = p s(T_1) + (1 - p) s(T_2)$ on geodesics arcs
- the image $\mathcal{I}(T_v)$ of a subtree T_v as the union of the images $\mathcal{I}(T_{v,1})$ and $\mathcal{I}(T_{v,2})$, where $T_v = \mathfrak{M}(T_{v,1}, T_{v,2})$, and the geodesic arc between $s(T_{v,1})$ and $s(T_{v,2})$ with root vertex at $s(T_v)$



Externalization and syntax-semantics interface



goal: giving a geometric model for the Conceptual/Intentional and Sensory/Motor channels and their interaction

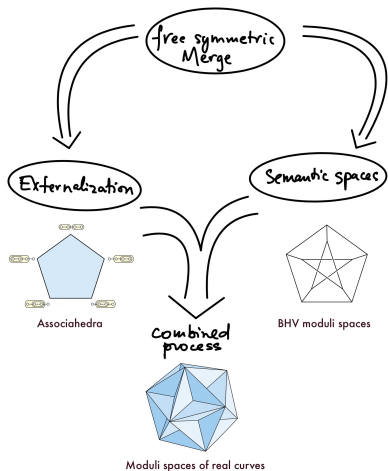
main ideas

- planarization of Externalization is language dependent, we need a space where different planarization can be compared
⇒ **associahedron**.
- hierarchical structures produced by free symmetric Merge acquire a metric structure through mapping to semantic spaces
- metric structure keeps track of semantic relatedness across substructures
- assignment of metric data on (non-planar) binary rooted trees described by **BHV moduli space**
- so previous picture becomes geometric relation between associahedra and BHV moduli spaces of metric trees (relation realized by **moduli space of real curves** of genus zero with marked points)

some references

- S.L. Devadoss, J. Morava, *Navigation in tree spaces*, Adv. in Appl. Math. 67 (2015), 75–95
- L. Billera, S. Holmes, K. Vogtmann, *Geometry of the space of phylogenetic trees*, Advances in Applied Mathematics 27 (2001) 733–767.
- F. Apéry, M. Yoshida, *Pentagonal structure of the configuration space of five points in the real projective line*, Kyushu J. Math. 52 (1998) N. 1, 1–14.
- S.L. Devadoss, *Tessellations of moduli spaces and the mosaic operad*, Homotopy Invariant Algebraic Structures, Contemporary Mathematics 239 (1999) 91–114.

Externalization and syntax-semantics interface



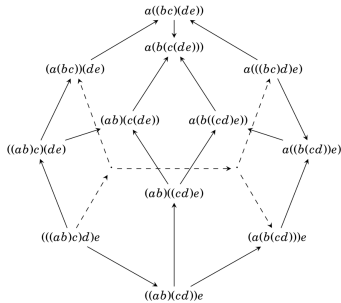
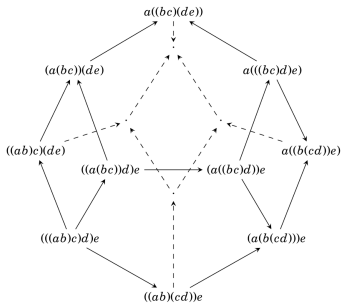
the resulting moduli space parameterizes simultaneous compatible choices of a metric embedding of the non-planar syntactic object and of a planarization

associahedra (Stasheff)

- associahedron K_n convex polytope of dimension $n - 2$, vertices are all the balanced parentheses insertions on an ordered string of n symbols (all planar binary rooted trees on n leaves)
- 1-dimensional associahedron K_3

$$((ab)c) \longleftrightarrow (a(bc))$$

- 2-dimensional K_4 a pentagon
- 3-dimensional K_5 :

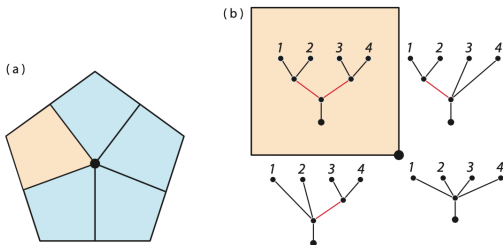


associahedra with metric data

- associahedron K_n decomposed into C_{n-1} cubes of dimension $n - 2$ with Catalan number

$$C_{n-1} = \frac{1}{n} \binom{2n-2}{n-1}$$

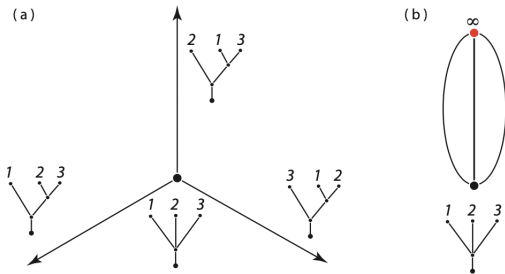
- cubical decomposition of K_4

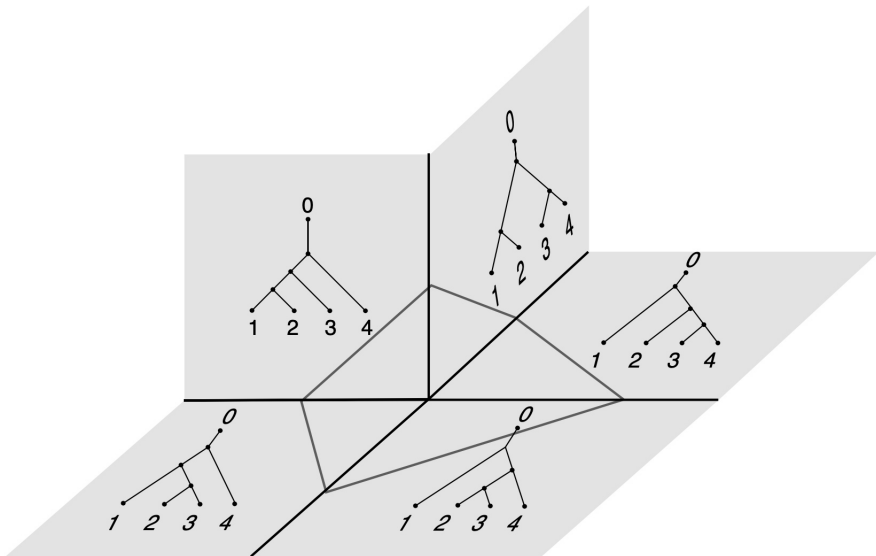


- polytope points as metric structures on planar binary rooted trees: weights in $\mathbb{R}_{\geq 0}$ to the internal edges of the tree

BHV moduli spaces of metric trees (Billera, Holmes, Vogtmann)

- moduli space BHV_n of abstract binary rooted trees with n leaves (with no assigned planar structure) along with weighted internal edges
- one-point compactification BHV^+
- consider all the $(2n - 3)!!$ abstract binary rooted trees with n labeled leaves: they all have $n - 2$ internal edges
- for each tree an orthant $\mathbb{R}_{\geq 0}^{n-2}$: all the possible choices of weight (length) of internal edges
- moduli space BHV_3 and one-point compactification BHV_3^+

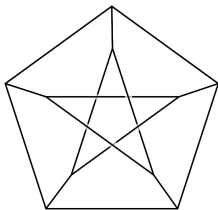




part of moduli space BHV_4

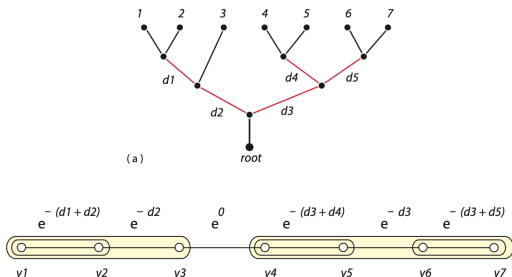
Link structure

- link \mathcal{L}_n of the origin in BHV_n is an $(n - 3)$ -dimensional simplicial complex
- case $n = 3$ just three points
- case $n = 4$ Peterson graph

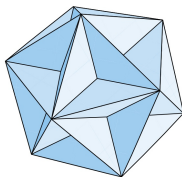
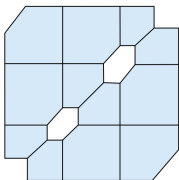


Relation between associahedra and BHVs: moduli spaces of curves

- in relating associahedra and BHV need to account for:
 - different planar structures (in associahedron not in BHV)
 - different permutations of the labels at the leaves (in BHV not in associahedron)
- one geometric object that accounts for both: moduli space of real curves $\bar{M}_{0,n+1}(\mathbb{R})$, orientation double cover $\bar{M}_{0,n+1}^{or}(\mathbb{R})$
- think of ordered leaves of tree as ordered set of points in the real line \mathbb{R}
- take coordinates of points as functions of weights of edges



- $\bar{M}_{0,n+1}(\mathbb{R})$: configurations of $n + 1$ points on $\mathbb{P}^1(\mathbb{R})$ compactified according to collisions (can always fix 0, 1, ∞ by a choice of coordinate in $\mathbb{P}^1(\mathbb{R})$ so $n - 2$ remaining variables)
- associahedra K_n have $\dim n - 2$
- known (Devadoss-Morava): $\bar{M}_{0,n+1}^{or}(\mathbb{R})$ decomposes into a union of $n!$ associahedra K_n glued together ($n!$ permutations of labelled leaves of trees)



Twelve associahedra K_4 assemble into the space $\bar{M}_{0,5}(\mathbb{R})$; orientation double cover gives 24 associahedra assembled into $\bar{M}_{0,5}^{or}(\mathbb{R})$ identified with the great dodecahedron

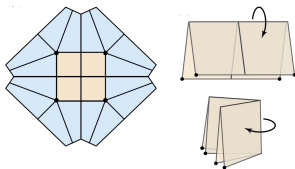
- also known (Devadoss-Morava): projection map

$$\Pi_n : \overline{M}_{0,n+1}^{or}(\mathbb{R}) \rightarrow \text{BHV}_n^+$$

- how these things fit together:

$$n! \cdot C_{n-1} = 2^{n-1} \cdot (2n-3)!!$$

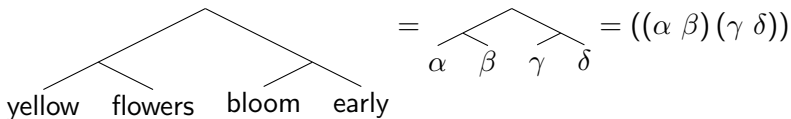
- on left: C_{n-1} cubes of $n!$ associahedra in $\overline{M}_{0,n+1}^{or}(\mathbb{R})$
- on right: $(2n-3)!!$ simplexes dim $(n-3)$ of \mathcal{L}_n
- 2^{n-1} is the (general) multiplicity of the projection map



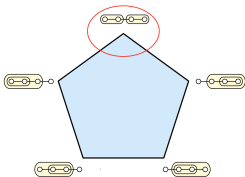
Four squares in adjacent associahedra K_4 are folded together (origami folding) in the projection: in double cover 8 squares identified in $\Pi_4 : \overline{M}_{0,5}^{or} \rightarrow \text{BHV}_4^+$

Example

- a English sentence like



- planar structure selects a vertex of the associahedron K_4

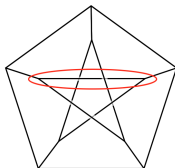


- this associahedron is one of the $4! = 24$ associahedra that correspond to the $4!$ permutations of the leaves labels
- this choice of one vertex on one of the 24 associahedra is the Externalization map

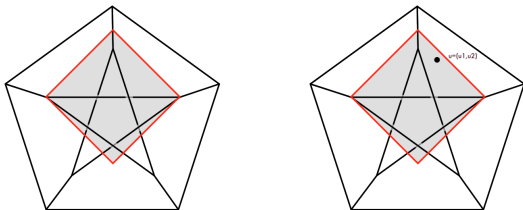
- underlying syntactic object

$$\{\{\alpha, \beta\}, \{\gamma, \delta\}\} = \begin{array}{c} \diagup \quad \diagdown \\ \alpha \quad \beta \quad \gamma \quad \delta \end{array} = \begin{array}{c} \diagup \quad \diagdown \\ \beta \quad \alpha \quad \gamma \quad \delta \end{array} = \dots$$

- this abstract tree is one of the $15 = (2n - 3)!!$, for $n = 4$, possible abstract binary rooted trees on four labeled leaves
- these 15 possibilities are the 15 edges of the link \mathcal{L}_4 of the origin in the moduli space BHV_4 , so the syntactic object selects an edge



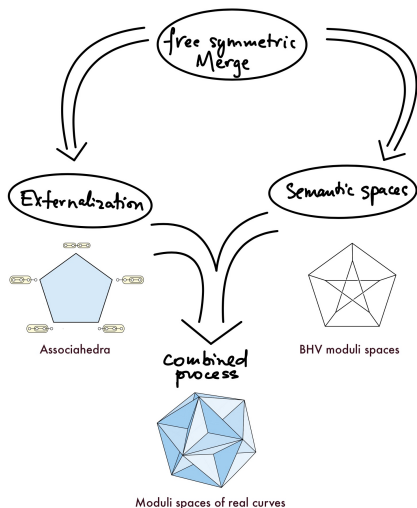
- given a semantic space \mathcal{S} and relatedness measures $u_1 = \mathbb{P}(s(\alpha), s(\beta))$ (relating “yellow” and “flower”) and $u_2 = \mathbb{P}(s(\gamma), s(\delta))$ (relating “blooming” and “early”) in \mathcal{S}
- get two coordinates $(u_1, u_2) \in [0, 1]^2$
- the selected edge of \mathcal{L}_4 is the line $u_1 + u_2 = 1$ (normalized link of origin)
- so obtain a point in (a square tile of) BHV_4^+



- this is the map of the syntax-semantics interface

syntax-semantic interface and Externalization: geometric model

the compatibility of these two maps (Externalization in K_4 and syntax-semantic interface in BHV_4) is through compatibility of associahedra and BHV moduli spaces via $\overline{M}_{0,5}^{or}(\mathbb{R})$, given by the map $\overline{M}_{0,5}^{or}(\mathbb{R}) \rightarrow BHV_4^+$



geometric view of syntactic parameters

- revisit Externalization as language-dependent section of a projection

$$\overline{M}_{0,n+1}^{or}(\mathbb{R}) \begin{array}{c} \xleftarrow{\sigma_{L,n}} \\ \xrightarrow{\Pi_n} \end{array} \text{BHV}_n^+$$

- dependence on syntactic parameters: covering transformations
 $\sigma_{L,n} = \gamma_{L,L',n} \circ \sigma_{L',n}$
- compare with other algorithms like LCA (on domain of *head*)
 $\sigma_{L,n} = \gamma_{L,n} \circ \sigma_{LCA,n}$
- possibility of studying syntactic parameters in terms of properties of symmetric groups (orders etc)