Lecture 8: Externalization
Ma 191c: Mathematical Models of Generative Linguistics

Matilde Marcolli

Caltech, Spring 2024

Summary so far of the "Strong Minimalist Theorem"

- *syntactic objects* are the free nonassociative commutative magma on the set $\mathcal{SO}_0$ of lexical items and syntactic features
- *workspaces* are the commutative Hopf algebra of binary forests
- *Merge* is a Hopf algebra Markov chain
- *minimal search* is a grading of the coproduct that selects External/Internal Merge as leading order terms; the Sideward Merge terms are needed for the Markov chain property but are subdominant for structure formation
- *minimal yield* can be derived from the Hopf algebra grading (as a no-complexity-loss principle)
- *countercyclic movement* and Late Merge are the insertion Lie algebra of the dual Hopf algebra
- *head and complement, phases, labeling* are additional structures on a subdomain of syntactic objects defined in terms of a combinatorial abstract head function

Additional remaining topics (before discussing Externalization)

1. Theta Theory: thematic roles
   - related to the structure of head and complement
   - needed for semantic parsing, in addition to head and labeling discussed earlier
   - mathematically related to operad and colored operad structures
2. obligatory control: restriction to diagonals mentioned earlier: FormCopy

Linguistic preliminaries: more on head, complement, modifiers, specifiers

Head: A lexical or phrasal element that is essential in forming a phrase.

Complement: A phrasal element that a head must combine with or a head select. These include direct object, indirect object, predicative complement, and oblique complement.

Modifier: A phrasal element not selected by the verb functions as a modifier to the head phrase.

complements are mandatory, modifiers are not

complements:

 John placed Kim behind the garage.
 John kept him behind the garage.
*John stayed Kim behind the garage.

*John placed him busy.
 John kept him busy.
*John stayed him busy.

modifiers:

John deposited some money in the bank.
John deposited some money in the bank on Friday.

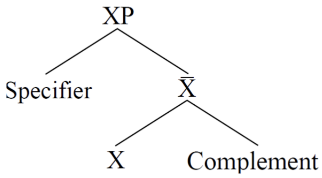
additional notion: specifier

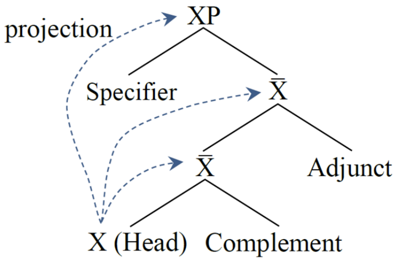
examples:

a little dog, *the* little dogs (indefinite or definite article)
this little dog, *those* little dogs (demonstrative)
my little dogs, *their* little dog (possessive adjective)
every little dog, *each* little dog, *some* little dog, *either* dog, *no* dog (quantifying)
my friend's little dog, *the Queen of England's* little dog (possessive phrase)

Note: language-dependent rules, for example on whether articles are required or not (or even exist)

More general: Spec position Spec–Head–Complement





using the older "X-bar theory" terminology and notation
(as in Government & Binding)

thematic roles

- predicate may be verb, adjective, preposition, noun
- argument is (usually) *required* by a predicate to complete its meaning
- predicates can take one or more arguments (valence of predicate)
- adjuncts can accompany predicates but not required
- theta-roles (thematic roles) are *syntactic* notions about number, type, placement of obligatory arguments

Examples

arguments:

Jill likes **Jack**.

Sam fried **the meat**.

The old man helped **the young man**.

Sam put the pen **on the chair**.

Larry does not put up **with that**.

Bill is getting **on my case**.

adjuncts:

Jill **really** likes Jack.

Jill likes Jack **most of the time**.

Jill likes Jack **when the sun shines**.

Jill likes Jack **because he's friendly**.

theta-roles: the verb *to put* assigns 3 thematic roles (valence 3): someone puts something somewhere; the verb *to give* also: someone gives something to someone, different thematic role

(reflected in cases in morphology of languages with noun declension)

Mathematical preliminary: operads

- operad (in category of Sets) is a collection $\mathcal{O} = \{\mathcal{O}(n)\}_{n\geq 1}$ of sets of n-ary operations (with n inputs and one output), with composition laws

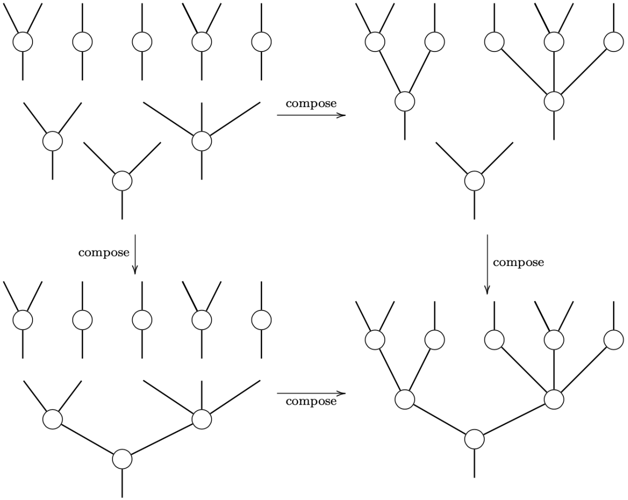
$$\gamma : \mathcal{O}(n) \times \mathcal{O}(k_1) \times \cdots \times \mathcal{O}(k_n) \to \mathcal{O}(k_1 + \cdots + k_n)$$

  plugging output of operations in $\mathcal{O}(k_i)$ into the i-th input of operations in $\mathcal{O}(n)$

- associativity of operad composition

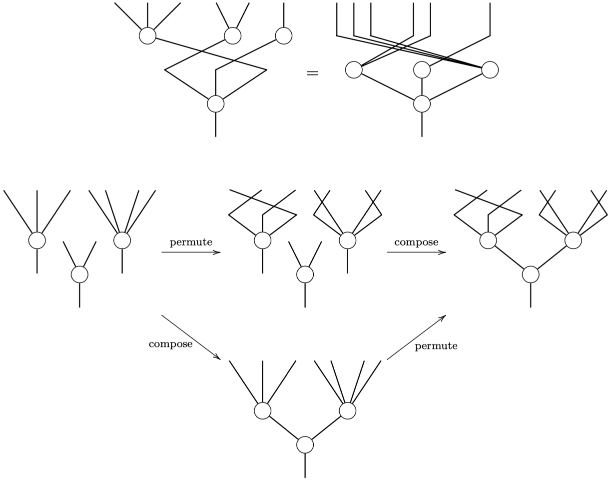
$$\gamma(\gamma(T, T_1, \ldots, T_m), T_{1,1}, \ldots, T_{1,n_1}, \ldots, T_{m,1}, \ldots, T_{m,n_m}) =$$
$$\gamma(T, \gamma(T_1, T_{1,1}, \ldots, T_{1,n_1}), \ldots, \gamma(T_m, T_{m,1}, \ldots, T_{m,n_m}))$$

- unit $1 \in \mathcal{O}(1)$ for composition



associativity of operad composition

case of symmetric operad: two equivariance conditions with respect to actions of symmetric groups



algebras over operads

- algebra $\mathcal{A}$ over an operad $\mathcal{O}$ (in Sets) is a set $\mathcal{A}$ on which the operations of $\mathcal{O}$ act
- can use elements of $\mathcal{A}$ as inputs for operations in $\mathcal{O}(n)$

$$\gamma_{\mathcal{A}} : \mathcal{O}(n) \times \mathcal{A}^n \to \mathcal{A}$$

- compositionality: for $T \in \mathcal{O}(m)$, $T_i \in \mathcal{O}(n_i)$, $\{a_{i,j}\}_{j=1}^{n_i} \subset \mathcal{A}$

$$\gamma_{\mathcal{A}}(\gamma_{\mathcal{O}}(T, T_1, \ldots, T_m), a_{1,1}, \ldots, a_{1,n_1}, \ldots, a_{m,1}, \ldots, a_{m,n_m}) =$$

$$\gamma_{\mathcal{A}}(T, \gamma_{\mathcal{A}}(T_1, a_{1,1}, \ldots, a_{1,n_1}), \ldots, \gamma_{\mathcal{A}}(T_m, a_{m,1}, \ldots, a_{m,n_m})) .$$

with $\gamma_{\mathcal{O}}$ composition in operad and $\gamma_{\mathcal{A}}$ operad action

Merge operad and action on syntactic objects

- operad $\mathcal{M}$ freely generated by a single commutative binary operation $\mathfrak{M}$
- have $\mathcal{M}(1) = \{\mathrm{id}\}$, $\mathcal{M}(2) = \{\mathfrak{M}\}$, $\mathcal{M}(3) = \{\mathfrak{M} \circ (\mathrm{id} \times \mathfrak{M}), \mathfrak{M} \circ (\mathfrak{M} \times \mathrm{id})\}$, etc.
- action $\gamma_{\mathcal{SO}} : \mathcal{M}(n) \times \mathcal{SO}^n \to \mathcal{SO}$ with $\gamma_{\mathcal{SO}}(T, T_1, \dots, T_n)$ with $T \in \mathcal{M}(n)$ and $T_i \in \mathcal{SO}$ for $i = 1, \dots, n$ abstract binary rooted tree in $\mathfrak{T}_{\mathcal{SO}_0} = \mathcal{SO}$ obtained by grafting root of the syntactic object T_i to the i-th leaf of $T \in \mathcal{M}(n)$
- if syntactic objects T_i have n_i leaves, then syntactic object $\gamma(T, T_1, \dots, T_n)$ obtained in this way has $n_1 + \cdots + n_k$ leaves

a subtlety about abstract trees

- no choice of ordering of leaves, so better way of writing action
- $\mathfrak{T}_{\mathcal{SO}_0,k}$, $\mathfrak{T}_k$ abtract binary trees with k leaves (with/without leaves lexical items)
- for $T \in \mathfrak{T}_n$ and $T_\ell \in \mathfrak{T}_{\mathcal{SO}_0,k_\ell}$

$$\gamma(T, \{T_\ell\}_{\ell \in L(T)})$$

root of the tree T_ℓ is grafted to the leaf $\ell \in L(T)$

- associativity of compositions

$$\gamma(\gamma(T, \{T_\ell\}_{\ell \in L(T)}), \{T'_{\ell'}\}_{\ell' \in L(\gamma(T,\{T_\ell\}_{\ell \in L(T)}))})$$

$$= \gamma(T, \{\gamma(T_\ell, \{T'_{\ell'}\}_{\ell' \in L(T_\ell)})\}_{\ell \in L(T)})$$

action of permutation groups

- first equivariance condition

$$\gamma(T \circ \tau, \{T_{\tau(\ell)}\}) = \gamma(T, \{T_\ell\}) \circ \tau'$$

$\tau \in \mathrm{Sym}(L(T)) \simeq S_n$, with $n \in \# L(T)$, and S_n the symmetric group, and with $\tau' \in \mathrm{Sym}(L(\gamma(T, \{T_\ell\})) \simeq S_{\sum_\ell k_\ell}$ that permutes the n blocks of k_ℓ leaves, leaving each block unchanged

- second equivariance condition

$$\gamma(T, \{T_\ell \circ \sigma_\ell\}_{\ell \in L(T)}) = \gamma(T, \{T_\ell\}_{\ell \in L(T)}) \circ \underline{\sigma}$$

$\sigma_\ell \in \mathrm{Sym}(L(T_\ell)) \simeq S_{k_\ell}$ with
$\underline{\sigma} \in \mathrm{Sym}(L(\gamma(T, \{T_\ell\})) \simeq S_{\sum_\ell k_\ell}$ that permutes the leaves within each block of k_ℓ leaves, leaving the position of the blocks unchanged

- for simplicity of notation use $\gamma(T, T_1, \ldots, T_n)$ instead of $\gamma(T, \{T_\ell\}_{\ell \in L(T)})$

compatibility with head function

- $\mathcal{M}_h(n)$ set of pairs (T, h_T) an abstract binary rooted tree $T \in \mathfrak{T}_n$ (with no labeling at the n leaves) and a *head function* $h_T : V^{int}(T) \to L(T)$
- composition in $\mathcal{M}$ induces operad structure on $\mathcal{M}_h = \{\mathcal{M}_h(n)\}$
- operad compositiom

$$\gamma_{\mathcal{M}_h} : \mathcal{M}_h(n) \times \mathcal{M}_h(k_1) \times \cdots \times \mathcal{M}_h(k_n) \to \mathcal{M}_h(k_1 + \cdots + k_n)$$

- data $h_T, h_{T_1}, \ldots, h_{T_n}$ combine to define a *head function* on $T' = \gamma_{\mathcal{M}}(T, T_1, \ldots, T_n)$:
  - all vertices of T' that are vertices of one of the trees T_i: set $h_{T'}(v) = h_{T_i}(v)$
  - vertices of T' that are non-leaf vertices of T, we define $h_{T'}(v)$ by: *head function* on T determines a leaf $h_T(v) \in L(T)$; $T_{i(\ell)}$ denote the tree that is grafted to the leaf $\ell \in L(T)$; take $h_{T'}(v) := h_{T_{i(h_T(v))}}$

Theta theory and operads

- so far only used *head function*: more refined information theta-theory: identify and remove the ill-formed sentences by structure of dependants (complements) of the *head*, in particular assignment of θ-roles (thematic roles)
- theta-theory models thematic relations between predicates and their arguments
- predicates assign *θ-roles* to their arguments
- θ-roles of arguments of predicates: "theme", "agent", "experiencer", "locative", "instrument", "possessor", etc
- matching condition: there should be a one-to-one correspondence between theta roles and arguments these are assigned to

Dichotomy in semantics:

1. External Merge (EM) sole responsible for assignment of theta-roles: argument structure, propositional domain
2. Internal Merge (IM) does no theta-structure: clausal domain, information-related, non-argument structure, displacement

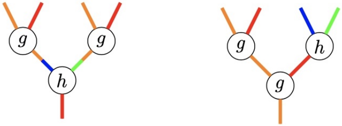
- this dichotomy is called "duality" in semantics in the linguistics literature
- in our setting theta-theory related to operad structure of syntactic objects (which only uses EM)

colored operads

- notion of colored operads similar to passing from group to groupoid etc: make operad compostions defined under some matching conditions (matching "colors" in a set Θ)
- collection $\mathcal{O} = \{\mathcal{O}(c, c_1, \ldots, c_n)\}$ of sets, with $c, c_i \in \Theta$ for $i = 1, \ldots, n$, c_i are color labels of inputs and c color label of output
- composition laws have to match colors

$$\begin{aligned} \gamma : \quad & \mathcal{O}(c, c_1, \ldots, c_n) \times \mathcal{O}(c_1, c_{1,1}, \ldots, c_{1,k_1}) \times \cdots \times \mathcal{O}(c_n, c_{n,1}, \ldots, c_{n,k_n}) \\ & \to \mathcal{O}(c, c_{1,1}, \ldots, c_{1,k_1}, \ldots, c_{n,1}, \ldots, c_{n,k_n}) \end{aligned}$$

- similar associativity, unity (one unit 1_c per color), and symmetric properties main idea: use colors for different θ-roles and matching rules of colored operad composition ensure correct consistent assignment of θ-roles



Example of compositions with mismatched and with matched color assignments (assignment of θ-roles)

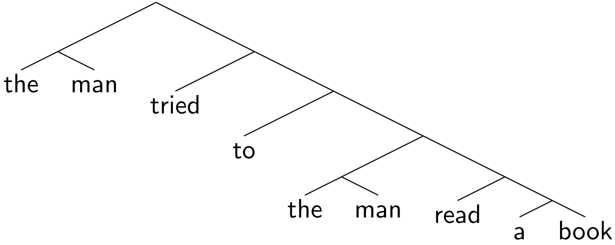
Theta theory and colored operads

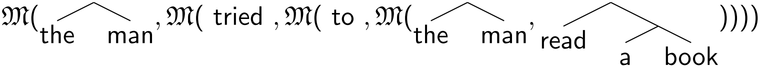
- set Θ of θ-roles and θ relations: labels "predicate" or "argument" and for arguments θ-roles labels "theme", "agent", "experiencer", "locative", "instrument", "possessor", etc.
- $\mathrm{Dom}_\Theta(h) \subset \mathrm{Dom}(h)$ set of syntactic objects $T \in \mathcal{SO}$ in domain of *head function* h admitting assignment of labels in Θ to edges of T compatible, on each substructure accessible term T_v with *head* and complement determined by *head function* h
- set $\mathrm{Dom}_\Theta(h) \subset \mathcal{SO}$ determines a colored operad $\mathcal{M}_{h,\Theta} = \{\mathcal{M}_{h,\Theta}(\theta, \theta_1, \ldots, \theta_n)\}$
- consequence: all n-ary theta-structures (elements $(T, h_T, \theta_T) \in \mathcal{M}_{h,\Theta}(\theta, \theta_1, \ldots, \theta_n)$) are composition of binary theta-structures through repeated application of binary External Merge building elements of $\mathcal{M}$

Obligatory control: an example ("Merge and SMT" §5.3)

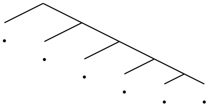
- sentence "the man tried to read a book"

$$\{\{\text{the}, \text{man}\}, \{\text{tried}, \{\text{to}, \{\{\text{the}, \text{man}\}, \{\text{read}, \{\text{a}, \text{book}\}\}\}\}\} =$$

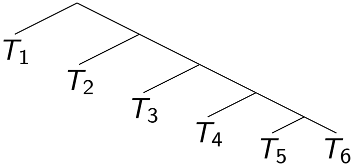




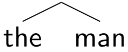
- in terms of operad action: element $T \in \mathcal{M}(6)$ of the form

$T =$ 

- inputs $T_1, \ldots, T_6$ in $\mathcal{SO}^6$, output syntactic object



- here take

$T_1 = T_4 =$ 

$T_2 =$ tried $T_3 =$ to

$T_5 =$ read $T_6 =$ 

Restriction to diagonals (FormCopy)

- for subset $\mathcal{I} \subset \{1, \dots, n\}$ diagonal

$$\mathrm{Diag}_{\mathcal{I}} = \{(T_1, \dots, T_n) \in \mathcal{SO}^n \mid T_i = \hat{T} \in \mathcal{SO}, \forall i \in \mathcal{I}\}$$

- in this example $\mathrm{Diag}_{1,4} \subset \mathcal{SO}^6$ (as inputs to operad action)
- FormCopy in the linguistics literature is restriction to a diagonal
- distinction between case of *repetitions* and *copies*: repetitions are *isomorphic* syntactic objects but not *identical*, copies are the same syntactic object
- effect: if repetitions, usual coproduct Δ in action of Merge
- if copies identified so can only extract all at once (or same accessible term from all) or none

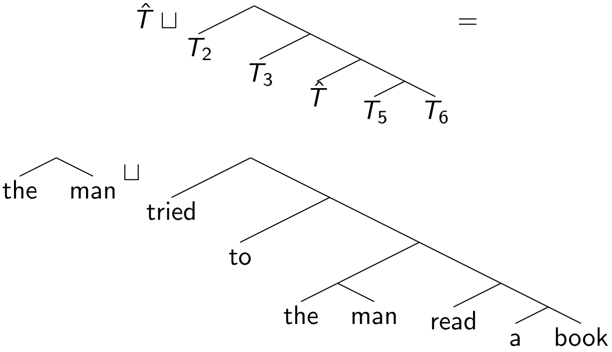
- accessible term extraction with FormCopy and obligatory control:

$$\Delta_{\mathcal{I}}(T) = \sum F_{\underline{v}} \otimes T/F_{\underline{v}} + \sum (F_{\underline{v}} \sqcup \hat{T}_v) \otimes (T/F_{\underline{v}})//\hat{T}_v$$

  sums are over subforests $F_{\underline{v}} \subset T$ such that $\hat{T} \cap F_{\underline{v}} = \emptyset$, and where we write $T//\hat{T}_v$ to denote the quotient with respect to all occurrences of $\hat{T}_v$ in T as accessible terms of all the identified copies in $\mathrm{Diag}_{\mathcal{I}}$

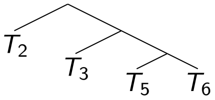
- still coassociative coproduct

- in the example considered start with a workspace forest

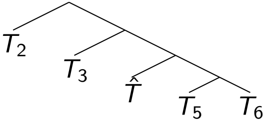


- a term in the coproduct $\Delta_{\mathcal{I}}$

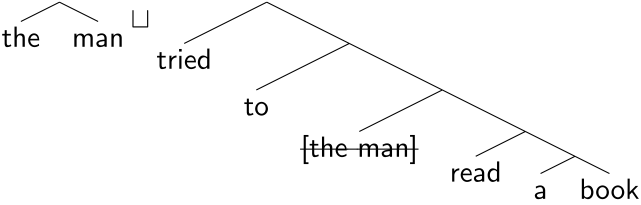
$$\hat{T} \sqcup T//\hat{T} \;\otimes 1 , \quad \text{with} \quad T//\hat{T} =$$



$$T =$$

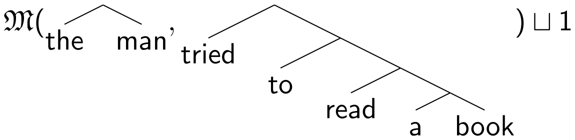


- this means coproduct term of the form

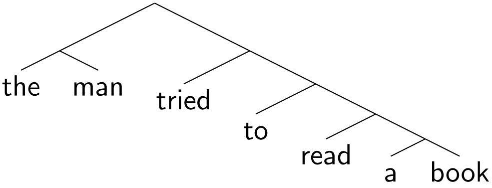


$\otimes 1$

- coproduct term that is targeted by the External Merge producing

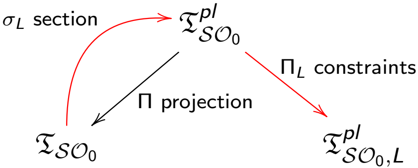


=



- if repetitions instead of copies (example “many people like many people”) then the two accessible terms would be extracted independently by Δ not simultaneously by $\Delta_{\mathcal{I}}$ (so would not have cancellation of deeper copy)

Externalization as a two-step procedure



- language-dependent assignment of planar embedding (consistent with word-order constraints of specific language)
- language-dependent quotient that eliminates non-viable constructions that are either non-parsable or violate other language-specific constraints (syntactic parameters)

planarity and morphism of magmas

- free commutative non-associative magma of syntactic objects

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

- free non-commutative non-associative magma (planar binary rooted trees)

$$\mathcal{SO}^{nc} = \mathrm{Magma}_{na,nc}(\mathcal{SO}_0, \mathfrak{M}^{nc}) = \mathfrak{T}^{pl}_{\mathcal{SO}_0}$$

- it generates the planar binary rooted trees with leaves labelled by $\mathcal{SO}_0$

$$\mathcal{SO}^{nc} \simeq \mathfrak{T}^{pl}_{\mathcal{SO}_0}$$

  write these as T^π (with T for abstract tree, π for choice of planar embedding)

- *asymmetric* Merge

$$\mathfrak{M}^{nc}(T_1^\pi, T_2^\pi) = \overset{\wedge}{T_1^\pi \quad T_2^\pi} \neq \overset{\wedge}{T_2^\pi \quad T_1^\pi} = \mathfrak{M}^{nc}(T_2^\pi, T_1^\pi)$$

- $\exists$ projection $\Pi : T^\pi \mapsto T$ (forgetting planar structure):

$$\Pi : \mathfrak{T}^{pl}_{\mathcal{SO}_0} = \mathcal{SO}^{nc} \twoheadrightarrow \mathcal{SO} = \mathfrak{T}_{\mathcal{SO}_0}$$

- Π is a morphism of magmas and canonical (independent of choices)

- Problem: the map $\Pi : \mathcal{SO}^{nc} \to \mathcal{SO}$ runs in the *opposite direction* to Externalization
- and... there is no morphism of magmas going the other way from $\mathcal{SO}$ to $\mathcal{SO}^{nc}$
- because since $(\mathcal{SO}, \mathfrak{M})$ is commutative it should map to a commutative sub-magma of $(\mathcal{SO}^{nc}, \mathfrak{M}^{nc})$
- but $(\mathcal{SO}^{nc}, \mathfrak{M}^{nc})$ does not have nontrivial commutative sub-magmas: if a nonempty planar tree T^π is in a commutative sub-magmas then $\mathfrak{M}^{nc}(T^\pi, T^\pi)$ also is but this contradicts commutativity since

$$\mathfrak{M}^{nc}(T^\pi, \mathfrak{M}^{nc}(T^\pi, T^\pi)) \neq \mathfrak{M}^{nc}(\mathfrak{M}^{nc}(T^\pi, T^\pi), T^\pi)$$

so what is to be done?

Externalization first step: section σ_L of the projection Π

- can construct (non-canonically: dependent on choices) a non-unique *section* $\Pi \circ \sigma_L = \mathrm{id}$

$$\mathfrak{T}^{pl}_{\mathcal{SO}_0} \underset{\Pi\,\mathrm{proj}}{\overset{\sigma_L}{\rightleftarrows}} \mathfrak{T}_{\mathcal{SO}_0}$$

- a *choice* of a point in each fiber $\Pi^{-1}(T)$ of the projection
- taking the one-way street Π in the opposite direction comes at a cost (loss of some good properties of the map):
  1. $\sigma_L : \mathfrak{T}_{\mathcal{SO}_0} \to \mathfrak{T}^{pl}_{\mathcal{SO}_0}$ is not a morphism of magmas
  2. $\sigma_L : \mathfrak{T}_{\mathcal{SO}_0} \to \mathfrak{T}^{pl}_{\mathcal{SO}_0}$ is not unique and depends on choices
- linguistic consequences:
  1. Merge can act either *before* Externalization (New Minimalism SMT) or *after* (on planar trees as in Old Minimalism) but *not both ways consistently*
  2. Externalization is *necessarily* language-dependent and not uniquely defined

first step $\sigma_L : \mathfrak{T}_{\mathcal{SO}_0} \to \mathfrak{T}^{pl}_{\mathcal{SO}_0}$

- planarization σ_L via a language-dependent non-unique section of the projection
- only requirement on σ_L is compatibility with word-order parameters of given language L
- obtain in this way a planar tree $T^{\pi_L} = \sigma_L(T)$ for every syntactic object $T \in \mathcal{SO}$ no further restriction

Externalization second step: other constraints

- need further elimination of those objects $T^{\pi_L} \in \mathcal{SO}^{nc}$ that violate linguistic constraints (more syntactic parameters) of a particular language L (not word order related)
- other language dependent conditions: *theta-theory*, obligatory control, etc (eliminate trees that fail these)

quotient map $\Pi_L : \mathfrak{T}^{pl}_{\mathcal{SO}_0} \to \mathfrak{T}^{pl,L}_{\mathcal{SO}_0}$ projection that eliminates what does not satisfy these further constraints

Externalization as correspondence

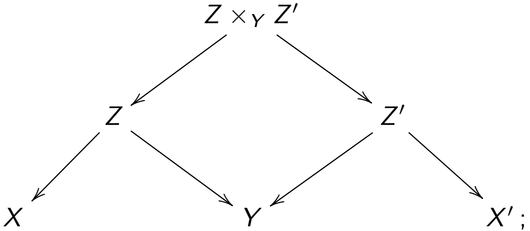
- two-step externalization: section of a projection followed by another projection ... correspondence
- the simplest way of describing transformation is through functions $f : X \to Y$ (single valued $x \mapsto f(x)$)
- but sometimes functions are not the best way of going from X to Y and a better notion is correspondences

$$X \longleftarrow Z \longrightarrow Y$$

climbing one arrow "the wrong way" then going down the other one (includes the case of multivalued functions)

correspondences

- a correspondence Z transfers structures (e.g. vector bundles, spaces of functions, etc.) from X to Y, pulling back to Z and pushing forward to Y
- in a category $\mathcal{C}$ that has pullbacks correspondences as 1-morphisms in a 2-category of *spans* in $\mathcal{C}$.
- 2-category $\mathrm{Spans}(\mathcal{C})$ with
    - objects given by the objects of $\mathcal{C}$;
    - 1-morphisms given by correspondences $\mathcal{C}$-diagrams
    - composition given by the pullback



- 2-morphisms between spans $X \leftarrow Z_1 \rightarrow Y$ and $X \leftarrow Z_2 \rightarrow Y$ are morphisms $Z_1 \rightarrow Z_2$ in $\mathcal{C}$ with commutative

cospans vs spans (co-correspondences)

- diagrams for algebras (if correspondences for spaces)

$$\begin{array}{ccc} & \mathcal{E} & \\ \nearrow & & \nwarrow \\ \mathcal{A} & & \mathcal{A}' \end{array}$$

- typically for algebras morphisms are given by bimodules
- but can also consider co-correspondences as spans

$$\begin{array}{ccc} & \mathcal{E} & \\ \swarrow & & \searrow \\ \mathcal{A} & & \mathcal{A}' \end{array}$$

- composition

$$(\mathcal{A}' \stackrel{g}{\leftarrow} \mathcal{E}' \to \mathcal{A}') \circ (\mathcal{A} \leftarrow \mathcal{E} \stackrel{f}{\to} \mathcal{A}'')$$

pullback: restricted direct sum

$$\mathcal{E} \oplus_{\mathcal{A}'} \mathcal{E}' = \{(e,e') \,|\, f(e) = g(e')\}$$

What happens to action on workspaces in Externalization?

- since all Merge operations happen with symmetric Merge before externalization it seems one cannot see at all this action after externalization (because magma structure not preserved by planarization $\sigma_{\mathcal{L}}$)
- but one can still see part of it
- $\mathcal{A}_{na,c} = (\mathcal{V}(\mathfrak{T}_{\mathcal{SO}_0}), \mathfrak{M})$ non-associative commutative algebra
- representations for a non-associative algebras $\mathcal{A}$ are just *linear maps* (not algebra homomorphisms) $\rho : \mathcal{A} \to \mathrm{End}(\mathcal{V})$, endomorphisms of vector space $\mathcal{V}$
- fix an argument of Merge: $\mathfrak{M}^T(T') := \mathfrak{M}(T, T')$

- then representation (in the above sense) from action on workspaces $F = \sqcup_a T_a$

$$\rho(T)(F) = \sqcup \circ (\mathfrak{M}^T \otimes 1) \circ \Delta(F) = \sqcup_a (\mathfrak{M}(T, T_{a,v}) \sqcup T_a / T_{a,v})$$

suffices to determine full action if known for all T

- image of the $\rho(T)$ recovers image of the $\mathfrak{M}_{S,S'}$

$$\bigcup_T \rho(T)(\mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0})) = \bigcup_{S,S'} \mathfrak{M}_{S,S'}(\mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0})),$$

here union as the common span as vector spaces

- the projection part is compatible with action of asymmetric Merge
- but section $\sigma_{\mathcal{L}}$ is not a magma morphism so only projection in the other direction is compatible with Merge action

$$\begin{array}{ccccc}
 & & \mathcal{A}_{na,nc,\mathcal{L}} \otimes \mathcal{V}(\mathfrak{F}^{pl,\mathcal{L}}_{\mathcal{SO}_0}) & \xrightarrow{\rho^{pl,\mathcal{L}}} & \mathcal{V}(\mathfrak{F}^{pl,\mathcal{L}}_{\mathcal{SO}_0}) \\
 & \nearrow^{\Pi_{\mathcal{L}}\otimes\Pi_{\mathcal{L}}} & & \nearrow_{\Pi_{\mathcal{L}}} & \\
\mathcal{A}_{na,nc} \otimes \mathcal{V}(\mathfrak{F}^{pl}_{\mathcal{SO}_0}) & \xrightarrow{\rho^{pl}} & \mathcal{V}(\mathfrak{F}^{pl}_{\mathcal{SO}_0}) & & \\
\downarrow{\scriptstyle \Pi\otimes\Pi} & & \downarrow{\scriptstyle \Pi} & & \\
\mathcal{A}_{na,c} \otimes \mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0}) & \xrightarrow{\rho} & \mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0}) & & .
\end{array}$$

- but climbing up the projection Π with the section $\sigma_{\mathcal{L}}$ leads to only a partially defined Merge action on the image
- indeed in old Minimalism, where Merge is after planarization, Merge is partially defined with specific conditions on domains

role of syntactic parameters

- syntactic parameters account for syntactic variation across languages
- part of externalization (determine choice of σ_L and Π_L
- assume all syntactic parameters are binary (sometimes ternary as undefined value due to relations)
- set of syntactic parameters of *possible languages* as a subset $\mathcal{P} \subset \mathbb{F}_2^N$ with some large ($N \geq 200$) configuration space
- large number of relations expected
- rough picture: $q : \mathbb{F}_2^N \to \mathbb{F}_2^M$ projection to word-order parameters $\Rightarrow$ responsible for constraints on choice of σ_L
- remaining parameters affect Π_L
- can think of setting parameters as a collection of maps to Grassmannians for $\pi = (\pi_i) \in \mathcal{P}$

$$E_{i,\ell} : \mathcal{P} \to \mathrm{Gr}(d_{\pi_i,\ell}, d_\ell)$$

that selects subspaces of $\mathcal{V}(\mathfrak{T}^{pl}_{\mathcal{SO}_0})$ compatible with value of a given parameter

Data on syntactic parameters

1. Syntactic Structures of World Languages (SSWL) now TerraLing
2. World Atlas of Language Structures (WALS)
3. another set of data from Longobardi–Guardiano, 2009
4. more complete set of data by Giuseppe Longobardi's LanGeLin Collaboration, 2016

• Data Analysis of syntax of world languages with various mathematical tools (dimensional reduction, persistent topology, phylogenetic algebraic geometry, etc.)

SSWL list of parameters (253 languages)

The 116 binary variables recorded in the SSWL database include:

- variables describing word order properties, from *01*–Subject Verb to *22*–Noun Pronomial Possessor
- variables *A01–A04* describing relations of adjectives to nouns and degree words
- variable *AuxSel01* about the selection of auxiliary verbs
- variables *C01–C04* related to word order properties of complementarizer and clause and adverbial subordinator and clause
- variables *N201–N211* on properties of numerals
- variables *Neg01–Neg14* on negation
- variables *OrderN301–OrderN312* on word order properties involving demostratives, adjectives, nouns, and numerals
- variables *Q01–Q15* regarding the structure of questions
- variables *Q16Nega–Q18Nega* and *Q19NegQ–Q22NegQ* on answers to negative questions
- variables *V201-V202* on declarative and interrogative Verb-Second
- variables *w01a–w01c* on indefinite mass nouns in object position
- variables *w02a–w02c* on definite mass nouns in object position

SSWL list of parameters

- variables *w03a–w03d* on indefinite singular count nouns in object position
- variables *w04a–w04c* on definite singular count nouns in object position
- variables *w05a–w05c* on indefinite plural count nouns in object position
- variables *w06a–w06c* on definite plural count nouns in object position
- variables *w06a–w06c* on definite plural count nouns in object position
- variables *w07a–w07d* on nouns with (intrinsically) unique referents in object position
- variables *w08a–w08d* on proper names in object position
- variables *w09a–w09b* on order of article and proper names in object position
- variables *w10a–w10c* on proper names modified by an adjective in object position
- variables *w11a–w11b* on order of proper names and adjectives in object position
- variables *w12a–w12f* on order of definite articles and nouns in object position
- variables *w20a–w20e* on singular count nouns in vocative phrases
- variables *w21a–w21e* on proper nouns in vocative phrases
- variables *w22a–w22e* on plural nouns in vocative phrases.

LanGeLin list of parameters
(64 languages including microvariations)

FGP	gramm. person	GSI	grammaticalised inalienability
FGM	gramm. Case	ALP	alienable possession
FPC	gramm. perception	GST	grammaticalised Genitive
FGT	gramm. temporality	GEI	Genitive inversion
FGN	gramm. number	GNR	non-referential head marking
GCO	gramm. collective number	STC	structured cardinals
PLS	plurality spreading	GPC	gender polarity cardinals
FND	number in D	PMN	personal marking on numerals
FSN	feature spread to N	CQU	cardinal quantifiers
FNN	number on N	PCA	number spread through cardinal adjectives
SGE	semantic gender	PSC	number spread from cardinal quantifiers
FGG	gramm. gender	RHM	Head-marking on Rel
CGB	unbounded sg N	FRC	verbal relative clauses
DGR	gramm. amount	NRC	nominalised relative clause
DGP	gramm. text anaphora	NOR	NP over verbal relative clauses/ adpositional genitives
CGR	strong amount		
NSD	strong person	AER	relative extrap.
FVP	variable person	ARR	free reduced rel
DGD	gramm. distality	DOR	def on relatives
DPQ	free null partitive Q	NOD	NP over D
DCN	article-checking N	NOP	NP over non-genitive arguments

LanGeLin list of parameters

DNN	null-N-licensing art	PNP	P over complement
DIN	D-controlled infl. on N	NPP	N-raising with obl. pied-piping
FGC	gramm. classifier	NGO	N over GenO
DBC	strong classifier	NOA	N over As
XCN	conjugated nouns	NM2	N over M2 As
GSC	c-selection	NM1	N over M1 As
NOE	N over ext. arg.	EAF	fronted high As
HMP	NP-heading modifier	NON	N over numerals
AST	structured APs	FPO	feature spread to genitive postpositions
FFS	feature spread to struct. APs	ACM	class MOD
ADI	D-controlled infl. on A	DOA	def on all +N
DMP	def matching pron. poss.	NEX	gramm. expletive article
DMG	def matching genitives	NCL	clitic poss.
GCN	Poss°-checking N	PDC	article-checking poss.
GFN	Gen-feature spread to Poss°	ACL	enclitic poss. on As
GAL	Dependent Case in NP	APO	adjectival poss.
GUN	uniform Gen	WAP	wackernagel adjectival poss.
EZ1	generalized linker	AGE	adjectival Gen
EZ2	non-clausal linker	OPK	obligatory possessive with kinship nouns
EZ3	non-genitive linker	TSP	split deictic demonstratives
GAD	adpositional Gen	TSD	split demonstratives
GFO	GenO	TAD	adjectival demonstratives
PGO	partial GenO	TDC	article-checking demonstratives
GFS	GenS	TLC	Loc-checking demonstratives
GIT	Genitive-licensing iterator	TNL	NP over Loc

Most extensive treatment of syntactic parameters:

- Ian Roberts, *Parameter Hierarchies and Universal Grammar*, Oxford University Press, 2019
  - formulated within the Minimalism framework
  - extensive empirical evidence on syntactic variation
  - parameters organized into *hierarchies*
  - parameters as "emergent properties"
  - extensive description, but still not fully incorporated as a theoretical/mathematical model of externalization

Evidence for relations between parameters

- some relations explicitly known for linguistic reasons (Longobardi et al.)
- some visible through data analysis:
  - deviation from Markovian behavior (evolution as Markov model on a tree – phylogenetic trees of languages): issues with hypothesis of identically distributed independent random variables
  - coding theory perspective: collection of languages $\mathcal{L}$ comparative view of their parameters (binary code): if random code with independent variables would be around the Gilbert-Varshamov curve in the space of code parameters but many outliers high above
  - dimensional analysis finds actual dimension much lower ($d \sim 30$ among $N = 116$ for SSWL and $d \sim 15$ for $N = 83$ of LanGeLin)
  - higher recoverability of some parameters in sparse distributed memory (Kanerva network) models

some references

- Sitanshu Gakkhar, Matilde Marcolli, *Syntactic Structures and the General Markov Models*. Math. Comput. Sci. 18 (2024), no. 1, Paper No. 4.
- Kevin Shu, Matilde Marcolli, *Syntactic Structures and Code Parameters*, Math. Comput. Sci. 11 (2017) N.1, 79-90
- Alexander Port, Taelin Karidi, Matilde Marcolli, *Topological Analysis of Syntactic Structures*, Math. Comput. Sci. 16 (2022), no. 1, Paper No. 2, 68 pp.
- Jeong Joon Park, Ronnel Boettcher, Andrew Zhao, Alex Mun, Kevin Yuh, Vibhor Kumar, Matilde Marcolli, *Prevalence and recoverability of syntactic parameters in sparse distributed memories*, in "Geometric Structures of Information 2017", Lecture Notes in Computer Science, Vol. 10589 (2017) 1–8

Example: Parametric comparison as codes

• Kevin Shu, Matilde Marcolli, *Syntactic Structures and Code Parameters*, Math. Comput. Sci. 11 (2017), no. 1, 79–90.

• Matilde Marcolli, *Syntactic Parameters and a Coding Theory Perspective on Entropy and Complexity of Language Families*, Entropy 2016, 18(4), 110

- select a group of languages $\mathcal{L} = \{\ell_1, \ldots, \ell_N\}$
- with the binary strings of n syntactic parameters form a code $\mathcal{C}(\mathcal{L}) \subset \mathbb{F}_2^n$
- compute code parameters $(R(\mathcal{C}), \delta(\mathcal{C}))$ code rate and relative minimum distance
- analyze position of (R, δ) in space of code parameters
- get information about "syntactic complexity" of $\mathcal{L}$

code parameters $\mathcal{C} \subset \mathbb{F}_2^n$

- transmission rate (encoding)

$$R(\mathcal{C}) = \frac{k}{n}, \quad k = \log_2(\#\mathcal{C}) = \log_2(N)$$

for q-ary codes in $\mathbb{F}_q^n$ take $k = \log_q(N)$

- relative minimum distance (decoding)

$$\delta(\mathcal{C}) = \frac{d}{n}, \quad d = \min_{\ell_1 \neq \ell_2} d_H(\ell_1, \ell_2)$$

Hamming distance of binary strings of ℓ_1 and ℓ_2

- error correcting codes: optimize for maximal R and δ but constraints that make them inversely correlated
- bounds in the space of code parameters (R, δ)

Bounds on code parameters

• Gilbert-Varshamov curve (q-ary codes)

$$R = 1 - H_q(\delta), \quad H_q(\delta) = \delta \log_q(q-1) - \delta \log_q \delta - (1-\delta) \log_q(1-\delta)$$

q-ary Shannon entropy: asymptotic behavior of volumes of Hamming balls for large n

• The Gilbert-Varshamov curve represents the typical behavior of large random codes (Shannon Random Code Ensemble)

• Plotkin curve $R = 1 - \delta/q$: asymptotically codes below Plotkin curve $R \leq 1 - \delta/q$

• more significant asymptotic bound (Manin '82) between Gilbert-Varshamov and Plotkin curve

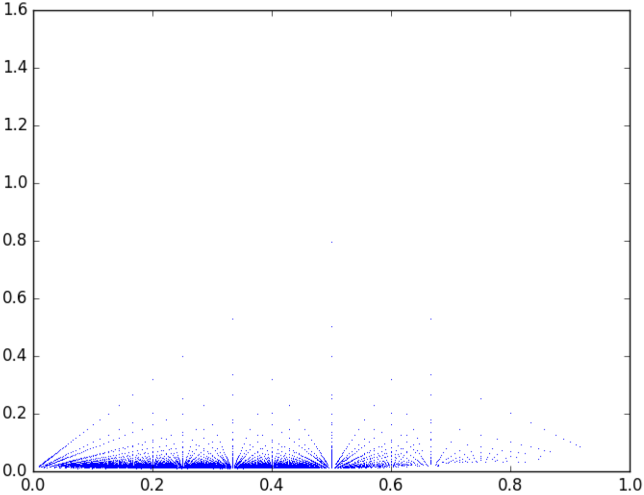
$$1 - H_q(\delta) \leq \alpha_q(\delta) \leq 1 - \delta/q$$

separates a region with dense code points with infinite multipliciites (below) and one with isolated code points with finite multiplicity (good codes above): difficult to find examples

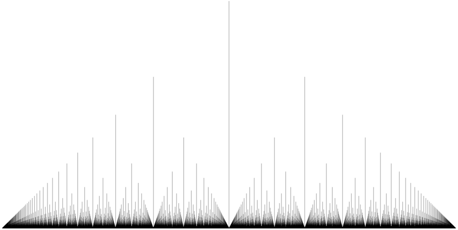
• asymptotic bound not explicitly computable (related to Kolmogorov complexity of codes, Manin–Marcolli)

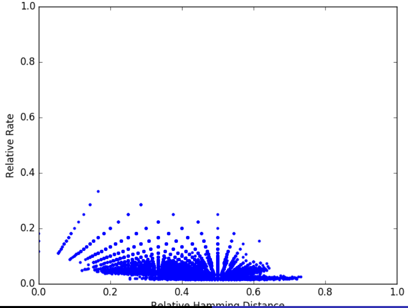
• difficult to construct codes above the asymptotic bound: examples from algebro-geometric codes from curves (but only for $q \geq 49$ otherwise entirely below the GV curve)

• look at the distribution of code parameters for small sets of languages (pairs or triples) and SSWL data



• in lower region of code parameter space a superposition of two Thomae functions ($f(x) = 1/q$ for $x = p/q$ coprime, zero on irrationals)





Thomae function

- defined by

$$f(x) = \begin{cases} 1 & x = 0 \\ q^{-1} & x \in Q,\ x = p/q,\ q > 0 \\ 0 & \text{otherwise.} \end{cases}$$

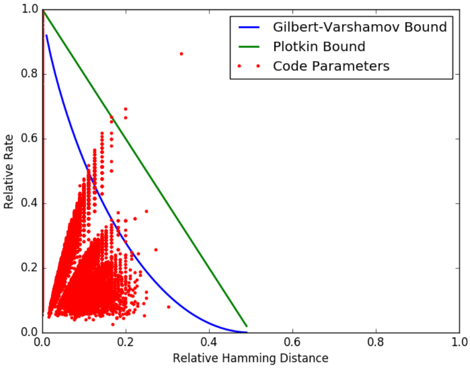
- note that fixing absolute rate of code $k = \log_2(\#\mathcal{L})$

$$(\delta(C), R(C)) = (\frac{d}{n}, \frac{k}{n})$$

- for $k = 1$ looking at graph of $d/n \mapsto 1/n$ where $n =$ parameters mapped for entire set $\mathcal{L}$
- so expect to see overlapping graphs of several Thomae functions
- the interesting part is where the points accumulate (depending on d/n values)
- lower regions of (δ, R)-space: random codes at most at GV bound

• more interesting what happens in the upper regions of the code parameter space

• take larger sets of randomly selected languages and syntactic parameters in the SSWL database

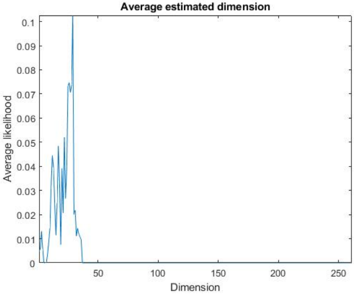


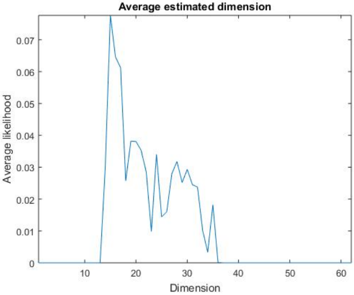
codes better than algebro-geometric above GV, asymptotic, and Plotkin

very far from random identically distributed variables behavior

Example: estimated dimension of syntactic parameters

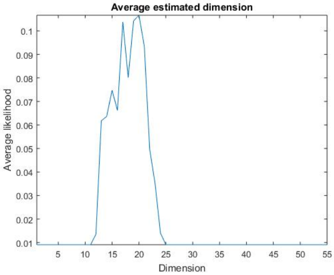
* Alexander Port, Taelin Karidi, Matilde Marcolli, *Topological Analysis of Syntactic Structures*, Math. Comput. Sci. 16 (2022), no. 1, Paper No. 2, 68 pp.

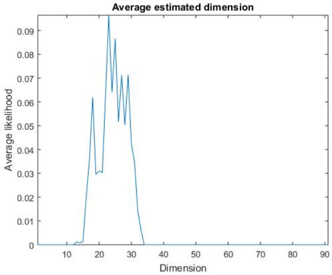




- Dimension of SSWL syntactic variables peak $d \sim 30$ (116 dim ambient space)
- Dimension of LanGeLin syntactic variables peak $d \sim 15$ (83 dim ambient space)

Family specific relations: dimension drop from $d \sim 30$ of SSWL





- Niger-Congo languages (SSWL data) $d \sim 20$
- Indo-European languages (SSWL data) $d \sim 23$

also see more data analysis in

* Sitanshu Gakkhar, Matilde Marcolli, *Syntactic Structures and the General Markov Models*. Math. Comput. Sci. 18 (2024), no. 1, Paper No. 4.

Heat Kernel Method

- Andrew Ortegaray, Robert C. Berwick, Matilde Marcolli, *Heat Kernel analysis of Syntactic Structures*, Math. Comput. Sci. 15 (2021), no. 4, 643–660.

General questions:

- What is the structure of relations between syntactic parameters?
- Which parameters cluster together?
- Do syntactic parameters span a manifold?
- What is the geometry/topology of this manifold?
- Are some parameters more dependent/independent from others?

Geometric methods of dimensional reduction:

- M. Belkin, P. Niyogi, *Laplacian eigenmaps for dimensionality reduction and data representation*, Neural Comput. 15 (6) (2003) 1373–1396

• *Problem*: low dimensional representations of data sampled from a probability distribution on a manifold

• *Main Idea*: build a graph with neighborhood information, use Laplacian of graph, want low dimensional representation that maintains local neighborhood information

• *Key Result*: graph Laplacian for a set of data point sampled from a uniform distribution on a manifold converges to Laplace–Beltrami operator on the manifold for large sets (using heat kernel and relation to Laplacian)

• Use to construct optimal (preserving information on manifold geometry) mapping of data sets to low dimensional spaces via eigenfunctions of Laplacian

Laplace–Beltrami operator and heat kernel

• on $\mathbb{R}^N$

$$\Delta f(x) = \sum_i \frac{\partial^2}{\partial x_i^2} f(x)$$

heat kernel equation

$$\frac{\partial}{\partial t} u(x,t) = \Delta u(x,t)$$

solutions with initial heat distribution $f(x)$

$$H^t f(x) = \int_{\mathbb{R}^N} f(y) H^t(x,y) dy$$

convolution with heat kernel

$$H^t(x,y) = (4\pi t)^{-k/2} \exp(-\frac{\|x-y\|^2}{4t})$$

Heat kernel and approximating the Laplacian

- Laplacian and heat kernel:

$$-\Delta f(x) = \frac{\partial}{\partial t} H^t f(x)|_{t=0}$$

$$= \lim_{t\to 0} \frac{(4\pi t)^{-k/2}}{t} \int_{\mathbb{R}^N} e^{-\frac{\|x-y\|^2}{4t}} f(y)dy - \frac{(4\pi t)^{-k/2}}{t} f(x) \int_{\mathbb{R}^N} e^{-\frac{\|x-y\|^2}{4t}} dy$$

- approximation: (uniform sampling of y)

$$\frac{(4\pi t)^{-k/2}}{t\, n} (f(x) \sum_{i=1}^{n} e^{-\frac{\|y_i - x\|^2}{4t}} - \sum_{i=1}^{n} e^{-\frac{\|y_i - x\|^2}{4t}} f(y_i))$$

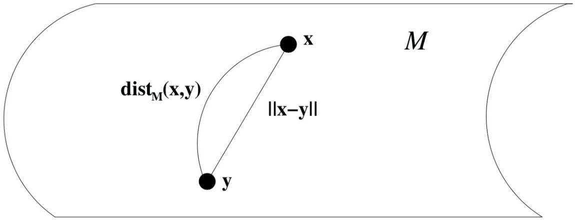
$$= C \, \frac{(4\pi t)^{-(k+2)/2}}{n} L^{t,n} f$$

- how to extend this idea from flat $\mathbb{R}^N$ to curved manifolds?

Laplacian approximation on manifolds

- geodesic distance and ambient Euclidean distance $\mathrm{dist}_{\mathcal{M}}(x,y) \geq \|x-y\|$
- exponential map $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ takes lines through origin to geodesics
- on compact manifolds chord distance approximates geodesic distance

$$\mathrm{dist}_{\mathcal{M}}(x,y) = \|x-y\| + O(\|x-y\|)$$



Step 1: replace integral on $\mathcal{M}$ with integral on small open set $\mathcal{U}$ around a point $x \in \mathcal{M}$

• can do this because for $\mathcal{U} \subset \mathcal{M}$ open and $d^2 = \inf_{y \notin \mathcal{U}} \|x-y\|^2$

$$\left| \int_{\mathcal{U}} e^{-\frac{\|x-y\|^2}{4t}} f(y) d\mu_y - \int_{\mathcal{M}} e^{-\frac{\|x-y\|^2}{4t}} f(y) d\mu_y \right| \leq M \|f\|_\infty e^{-d^2/4t}$$

• then can use exponential map $v \mapsto \exp_x(v)$ to parameterize neighborhood $\mathcal{U}$ of $x \in \mathcal{M}$

• at point x where exp map centered

$$\Delta_{\mathcal{M}} f(x) = \Delta_{\mathbb{R}^k} \tilde{f}(0), \quad \tilde{f}(v) = f(\exp_x(v))$$

- S. Rosenberg, *The Laplacian on a Riemannian manifold*, Cambridge University Press, 1997.

The role of scalar curvature

- exp map locally invertible: $\mathcal{B} \subset \mathcal{U}$ with inverse, change coords

$$\int_{\mathcal{B}} e^{-\frac{\|x-y\|^2}{4t}} f(y) d\mu_y = \int_{\exp_x^{-1}(\mathcal{B})} e^{-\frac{\phi(v)}{4t}} \tilde{f}(v) \det(d\exp_x(v)) dv$$

with $\phi(v) = \|v\|^2 + O(\|v\|^4)$ (chord and geodesic dist)

- asymptotics of exp map

$$|\Delta_{\mathbb{R}^k} \det(d\exp_x(v))| = \frac{\kappa(x)}{3} + O(\|v\|)$$

κ scalar curvature

$$\Delta_{\mathbb{R}^k}(\tilde{f} \det(d\exp_x(v))(0) = \Delta_{\mathbb{R}^k} \tilde{f}(0) + k\frac{\kappa(x)}{3} f(x)$$

Cancellation of curvature terms

- then obtain

$$\frac{\partial}{\partial t}((4\pi t)^{-k/2}\int_{\mathcal{B}} e^{-\frac{\|x-y\|^2}{4t}} f(y) d\mu_y)|_{t=0} = \Delta_{\mathcal{M}} f(x) + \frac{k}{3}\kappa(x) f(x) + C f(x)$$

using previous and relation of $\Delta_{\mathcal{M}} f(x)$ and $\Delta_{\mathbb{R}^k} \tilde{f}(0)$

- then obtain

$$\lim_{t\to 0}(4\pi t)^{-k/2}(\int_{\mathcal{M}} e^{-\frac{\|x-y\|^2}{4t}} f(x) d\mu_y - \int_{\mathcal{M}} e^{-\frac{\|x-y\|^2}{4t}} f(y) d\mu_y) = \Delta_{\mathcal{M}} f(x)$$

Belkin–Niyogi method main idea:

- then show that using a sampling approximation for $\mathbb{R}^k$ this gives

$$\lim_{n\to\infty}(4\pi t_n)^{-(k+2)/2} L^{t_n,n} f(x) = \frac{\Delta_{\mathcal{M}} f(x)}{\mathrm{Vol}(\mathcal{M})}$$

where $L^{t_n,n}$ is a graph-Laplacian approximation of the heat kernel

Main idea of Belkin–Niyogi heat kernel method

- k-dimensional compact smooth manifold $\mathcal{M}$ isometrically embedded in some $\mathbb{R}^N$
- data $\mathcal{S} = \{x_1, \ldots, x_n\}$ sampled from a uniform distribution in the induced measure on $\mathcal{M}$
- associated graph Laplacian $L = L^{t,n} = D^{t,n} - W^{t,n}$

$$L^{t,n} f(x) = f(x) \sum_j \exp\left(-\frac{\|x - x_j\|^2}{4t}\right) - \sum_j f(x_j) \exp\left(-\frac{\|x - x_j\|^2}{4t}\right)$$

- diagonal $D_{i,i} = D^{t,n}_{i,i} = \sum_j W^{t,n}_{i,j}$

Main Result: for sampled data $\mathcal{S} = \{x_1, \ldots, x_n\}$ from uniform distribution on $\mathcal{M}$ take $t_n = n^{-(k+2+\alpha)^{-1}}$ with $\alpha > 0$: for some $C > 0$

$$\lim_{n\to\infty} C \frac{(4\pi t_n)^{-\frac{k+2}{2}}}{n} L^{t_n,n} f(x) = \Delta_{\mathcal{M}} f(x)$$

for $f \in \mathcal{C}^\infty(\mathcal{M})$ with $\Delta_{\mathcal{M}}$ = Laplace-Beltrami operator on $\mathcal{M}$

• this shows the graph Laplacian of a point cloud data set converges to the Laplace–Beltrami operator on the underlying manifold

Why useful for low dimensional embeddings?

- given map $f : \mathcal{M} \to \mathbb{R}$, points near x will map to points near $f(x)$ if gradient ∇f is sufficiently small
- minimizing square gradient reduces to finding eigenfunctions of the Laplace–Beltrami operator: Stokes theorem

$$\int_{\mathcal{M}} \|\nabla f\|^2 = \int_{\mathcal{M}} f \Delta_{\mathcal{M}} f$$

normalized local extrema are eigenfunctions

$$\lambda_n = \inf_{X_n} \frac{\int_{\mathcal{M}} \|\nabla f\|^2}{\int_{\mathcal{M}} f^2}$$

X_n complement of span of previous eigenfunctions

• Use to construct optimal mapping of data sets to low dimensional spaces via eigenfunctions of Laplacian

Low dimensional embeddings algorithm

- setting: data points $x_1, \ldots, x_k \in \mathcal{M} \subset \mathbb{R}^\ell$ on a manifold; find points $y_1, \ldots, y_k$ in a low dimensional $\mathbb{R}^m$ ($m << \ell$) that *represent* the data points x_i

- Step 1 (a): adjacency graph (ϵ-neighborhood): an edge e_{ij} between x_i and x_j if $\|x_i - x_j\|_{\mathbb{R}^\ell} < \epsilon$

- Step 1 (b): adjacency graph (n nearest neighborhood): egde e_{ij} between x_i and x_j if x_i is among the n nearest neighbors of x_j or viceversa

- Step 2: weights on edges: heat kernel

$$W_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{t} \right)$$

if edge e_{ij} and $W_{ij} = 0$ otherwise; heat kernel parameter $t > 0$

• Step 3: Eigenfunctions for connected graph (or on each component)

$$L\psi = \lambda D\psi$$

diagonal matrix of weights $D_{ii} = \sum_j W_{ji}$; Laplacian $L = D - W$ with $W = (W_{ij})$; eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{k-1}$ and ψ_j eigenfuctions

$$\psi_i : \{1, \ldots, k\} \to \mathbb{R}$$

defined on set of vertices of graph

• Step 4: Mapping by Laplace eigenfunctions

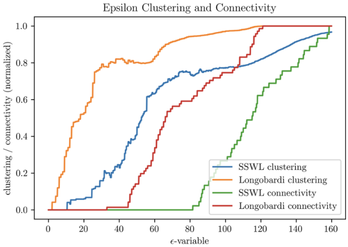
$$\mathbb{R}^\ell \supset \mathcal{M} \ni x_i \mapsto (\psi_1(i), \ldots, \psi_m(i)) \in \mathbb{R}^m$$

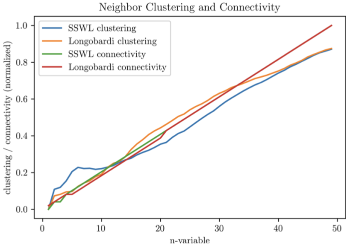
map by first m eigenfunctions

• Belkin–Niyogi: *optimality* of embedding by Laplace eigenfunctions

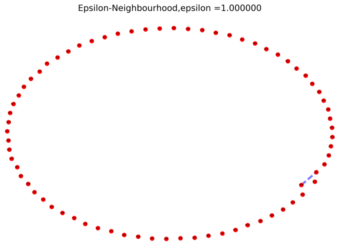
Heat Kernel analysis of Syntactic Parameters

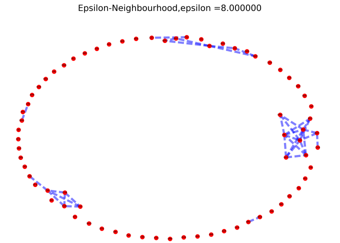
• Connectivity-clustering properties in ϵ-neighborhood and nearest-neighbor (SSWL data and LanGeLin data)



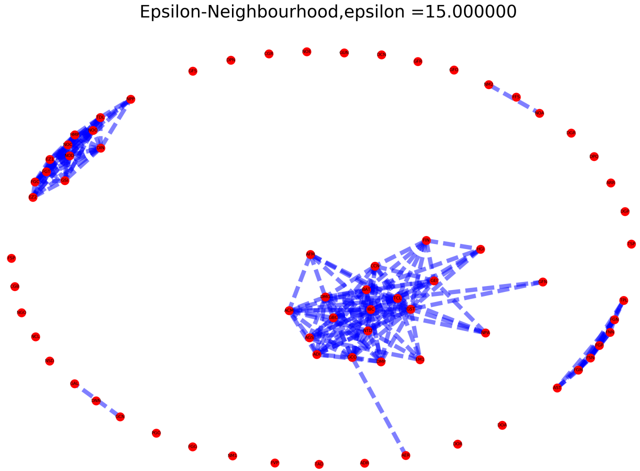


Graphs with ϵ-neighborhood Longobardi data

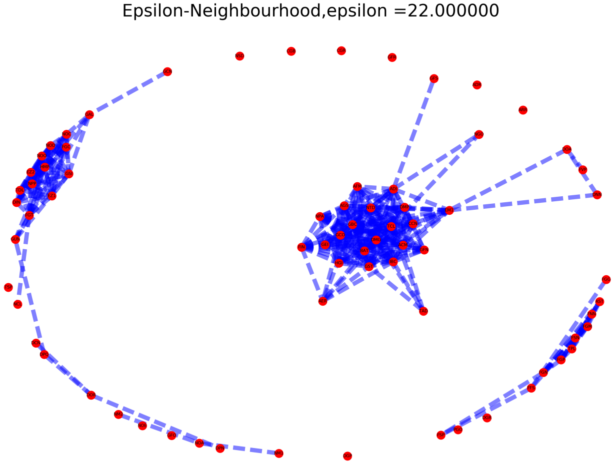




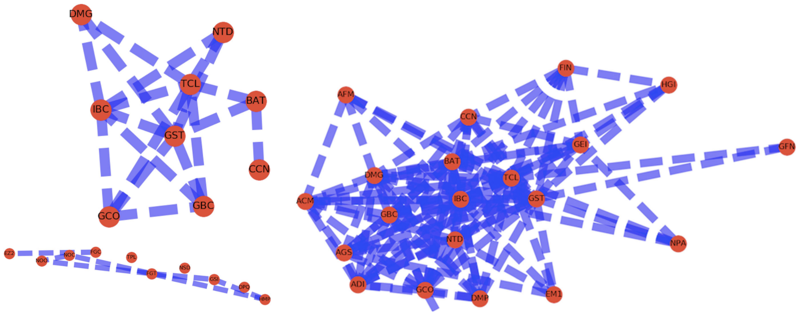
Graphs with ϵ-neighborhood Longobardi data



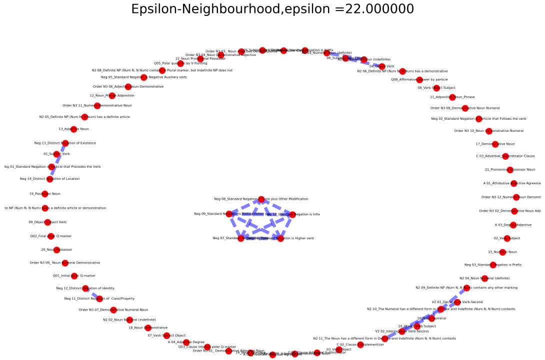
Graphs with ϵ-neighborhood Longobardi data



Structures of parameters relations in the LanGeLin parameters

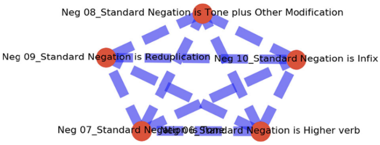


Graphs with ϵ-neighborhood SSWL data



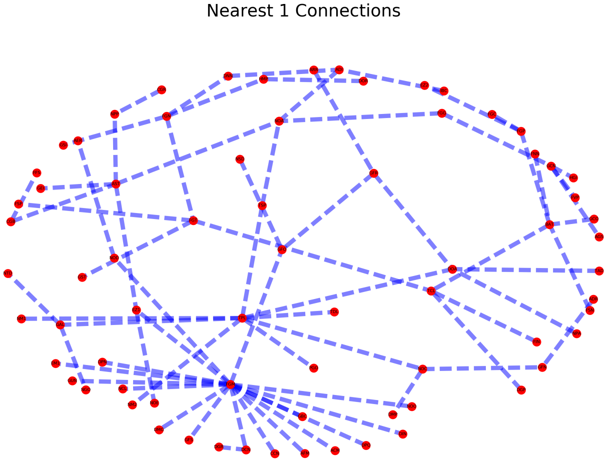
The ϵ-neighborhood construction is better suited to gain connectivity information in the Longobardi data: the SSWL data remain highly disconnected (only small local structures) compatible with SSWL being composed of different blocks of

Structures of parameters in SSWL data

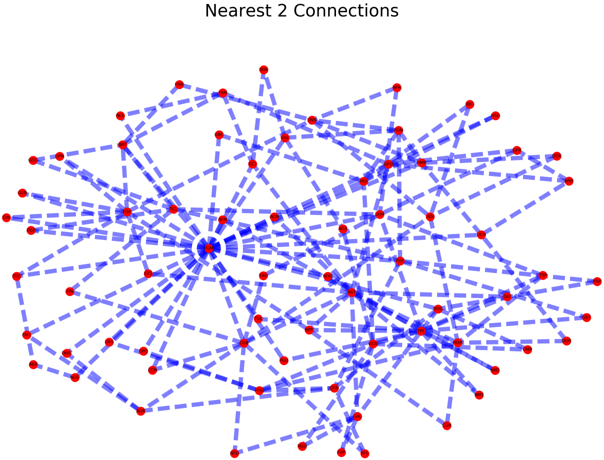


Explanation: properties of Standard Negation, connections Neg 06-10 emerge earlier (some are negation through tone as in some Niger-Congo languages and some Oto-Manguean, some other forms like reduplicated verb, infix); then cluster with other negation forms like Neg 01-05 (position of negation particle with respect to verb); expressed in different language families, relations are not family-specific

Graphs with *n*-neighborhood Longobardi data

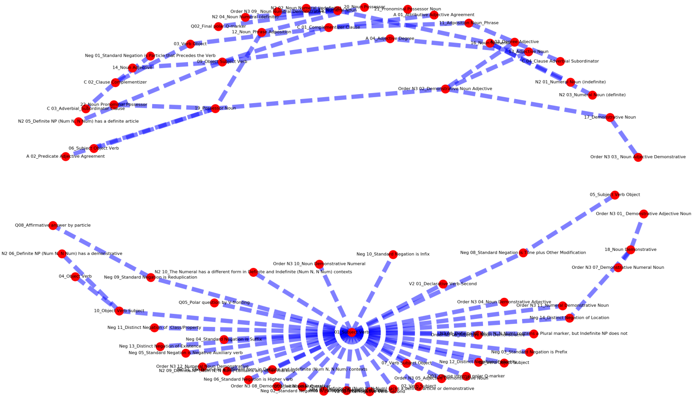


Graphs with *n*-neighborhood Longobardi data

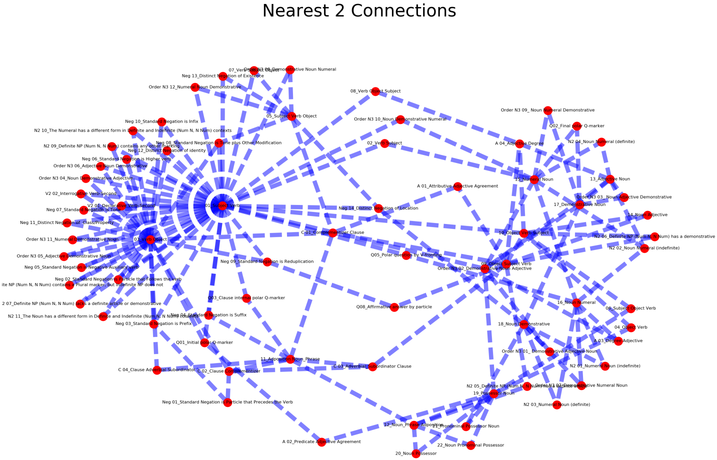


Graphs with *n*-neighborhood SSWL data

Nearest 1 Connections

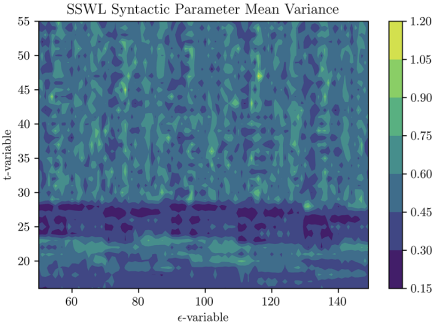


Graphs with *n*-neighborhood SSWL data



Variance

- graphs depend on ϵ-neighborhood and on t-heat kernel variable
- how embeddings depend on parameters: where the obtained coordinates by Laplace eigenfunctions are a set of independent coordinates that captures as much as possible of the data structure
- as in the case of other dimensional reduction methods (like PCA) high variance indicates independent resulting variables capture directions of highest variance in original data



Parameters evolve as dynamical variables (historical linguistics)

- can model evolution as a spin glass system with a set of spins (parameters) for each node (language) and interaction between languages proportional to biligualism; interaction also between parameters (Lagrange multipliers from relations alter the dynamics)
- phylogenetic trees of language families (usually based on morphology) and correlation to syntactic parameters: significant discrepancies with respect to identically distributed independent random variables of phylogenetic Markov models on trees
- topological structures in parameters distributions within given language families (persistent topology structures, sometimes describable in terms of historical linguistics)

general question of the Geometry of Syntax

- region of "possible languages" (A. Moro) among all configurations of syntactic parameters (I.Roberts, G.Longobardi, L.Rizzi,...)
- estimations of dimension and geometric structure of locus of possible languages (topology/geometry)
- comparison of sections σ_L and projection Π_L of externalization (determined by syntactic parameters) for different languages L
- distinguishing parameters that affect word order from parameters detecting other syntactic properties

Question of the geometry of syntactic parameters is a main open problem suitable for mathematical treatment:
parameter setting dynamics

Parameter setting and learnability problem

What is needed:

- large ambient space of parameters, with actual smaller dimensional submanifold (not directly known)
- language L determines parameters $\pi_L = (\pi_{L,i})_{i=1}^N$
- the $\pi_{L,i}$ are instructions for selecting syntactic objects (constraints)
- Note: don't need to learn a generative grammar (as in earlier models): need to learn a set of constraints that filter already formed structures
- significant improvement as there are non-learnability results in formal languages settings and learning grammars (Berwick–Niyogi)
- small set of examples (from specific language) should suffice to determine parameters (relations between parameters help), which in turn then suffice to select larger structures