# Geometric Models for Linguistics

Matilde Marcolli and Doris Tsao

Ma191b Winter 2017
Geometry of Neuroscience

References for this lecture:


1. Alexander Port, Iulia Gheorghita, Daniel Guth, John M.Clark, Crystal Liang, Shival Dasu, Matilde Marcolli, *Persistent Topology of Syntax*, arXiv:1507.05134
2. Karthik Siva, Jim Tao, Matilde Marcolli, *Spin Glass Models of Syntax and Language Evolution*, arXiv:1508.00504
3. 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*, arXiv:1510.06342
4. Kevin Shu, Sharjeel Aziz, Vy-Luan Huynh, David Warrick, Matilde Marcolli, *Syntactic Phylogenetic Trees*, arXiv:1607.02791
5. Kevin Shu, Matilde Marcolli, *Syntactic Structures and Code Parameters*, arXiv:1610.00311
6. Matilde Marcolli, *Syntactic Parameters and a Coding Theory Perspective on Entropy and Complexity of Language Families*, Entropy 2016, 18(4), 110

## Question: Language and Machines

• Natural Language Processing has made enormous progress in problems like automated translation

• but can we use computational (mathematical) techniques to better understand how the human brain processes language?

• some of the main questions:

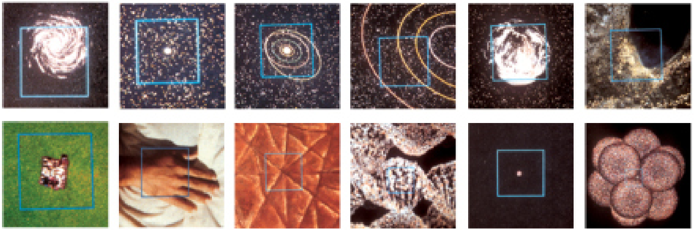
- Language acquisition (poverty of the stimulus): how does the learning brain converge to *one* grammar?
- How is language (in particular syntax) stored in the brain?
- How do languages change and evolve in time? quantitative, predictive modeling?

• Plan: approach these questions from a mathematical perspective, using tools from geometry and theoretical physics

## Language at different scales

- units of sound (phonology)
- words (morphology)
- sentences (syntax)
- global meaning (semantics)

Physics requires different mathematical models at different scales (relativity/cosmology, classical physics, quantum physics, string theory,...)



Expect different mathematical models of Linguistics at different scales

- focus on the "large scale structure" of language: syntax

## Syntax and Syntactic Parameters

• one of the key ideas of modern Generative Linguistics: Principles and Parameters (Chomsky, 1981)

- *principles*: general rules of grammar
- *parameters*: binary variables (on/off switches) that distinguish languages in terms of syntactic structures

• this idea is very appealing for a mathematician: at the level of syntax a language can be described by a set of coordinates given by binary variables

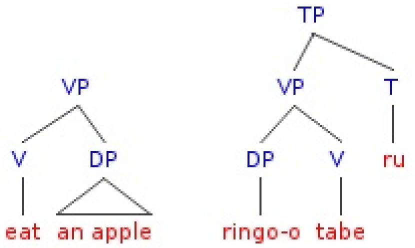
• however, surprisingly no mathematical model of Principles and Parameters formulation of Linguistics has been developed so far

## What are the binary variables?

- Example of parameter: head-directionality

(head-initial versus head-final)

English is head-initial, Japanese is head-final




VP= verb phrase, TP= tense phrase, DP= determiner phrase

- Other examples of parameters:
  - *Subject-side*
  - *Pro-drop*
  - *Null-subject*

## Main Problems

• there is no complete classification of syntactic parameters

• there are hundreds of such binary syntactic variables, but not all of them are "true" syntactic parameters (conflations of deep/surface structure)

• Interdependencies between different syntactic parameters are poorly understood: what is a good independent set of variables, a good set of coordinates?

• syntactic parameters are dynamical: they change historically over the course of language change and evolution

• collecting reliable data is hard! (there are thousands of world languages and analyzing them at the level of syntax is much more difficult for linguists than collecting lexical data; few ancient languages have enough written texts)

## Databases of syntactic structures of world languages

1. Syntactic Structures of World Languages (SSWL) `http://sswl.railsplayground.net/`
2. TerraLing `http://www.terraling.com/`
3. World Atlas of Language Structures (WALS) `http://wals.info/`
4. another set of data from Longobardi–Guardiano, Lingua 119 (2009) 1679-1706
5. more complete set of data announced by Longobardi, not yet available

• First Step: data analysis of syntax of world languages with various mathematical tools (persistent topology, etc.)

• we used the most extensive database currently available: SSWL with 116 "variables" (syntactic "parameters") and 253 world languages (but... some problems with SSWL)

## Problems of SSWL data

- Very non-uniformly mapped across the languages of the database: some are 100% mapped, while for some only very few of the 116 parameters are mapped
- Linguists criticize the choice of binary variable (not all of them should count as "true" parameters)

• the data of Longobardi–Guardiano are more reliable, but only 28 languages (almost all of them Indo-European) and 63 parameters

• linguistic question: can languages that are far away in terms of historical linguistics end up being close in terms of syntactic parameters?

• Guideline: given what is available at present, use SSWL data, but keeping limitations in mind

## Phylogenetic Algebraic Geometry of Languages

• Linguistics has studied in depth how languages change over time (Philology, Historical Linguistics)

• Usually via lexical and morphological analysis

• Goal: understand the historical relatedness of different languages, subdivisions into families and sub-families, phylogenetic trees of language families

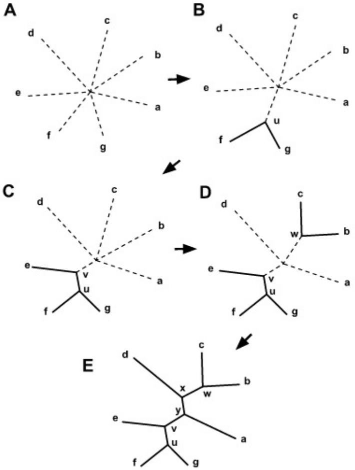
• Historical Linguistics techniques work best for language families where enough ancient languages are known (Indo-European and very few other families)

• Can one reconstruct phylogenetic trees computationally using only information on the modern languages?

• controversial results about the Indo-European tree based on *lexical data*: Swadesh lists of lexical items compared on the existence of cognate words (many problems: synonyms, loan words, false positives)

• Some phylogenetic tree reconstructions using syntactic parameters by Longobardi–Guardiano using their parameter data

• Hamming distance between binary string of parameter values + neighborhood joining method

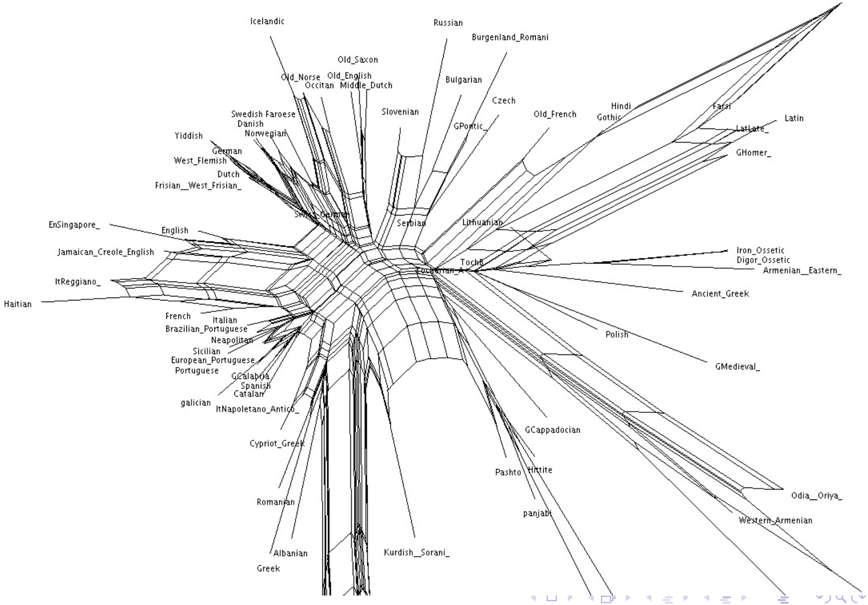


Expect problems: SSWL data and phylogenetic reconstructions

- known problems related to the use of Hamming metric for phylogenetic reconstruction
- SSWL problems mentioned above (especially non-uniform mapping)
- dependence among parameters (not independent random variables)
- syntactic proximity of some unrelated languages

- Phylogeny Programs for trees and networks
  - PHYLIP
  - Splittree 4
  - Network 5

# Checking on the Indo-European tree where good Historical-Linguistics



## Indeed Problems

- misplacement of languages within the correct family subtree
- placement of languages in the wrong subfamily tree
- proximity of languages from unrelated families (all SSWL)
- incorrect position of the ancient languages

• different approach: subdivide into subfamilies (some a priori knowledge from morpholexical linguistic data) and use Phylogenetic Algebraic Geometry (Sturmfels et al.) for statistical inference of phylogenetic reconstruction

## General Idea of Phylogenetic Algebraic Geometry

• Markov process on a binary rooted tree (Jukes-Cantor model)

• probability distribution at the root $(\pi, 1-\pi)$
(frequency of $0/1$ for parameters at root vertex) and transition matrices along edges $M^e$ bistochastic

• observed distribution at the $n$ leaves polynomial function

$$\Phi : \mathbb{C}^{4n-5} \to \mathbb{C}^{2^n}, \quad \Phi(\pi, M^e) = p_{i_1,\ldots,i_n}$$

defines an *algebraic variety*

$$V_T = \overline{\Phi(\mathbb{C}^{4n-5})} \subset \mathbb{C}^{2^n}$$

• (Allman–Rhodes theorem) ideal $\mathcal{I}_T$ defining $V_T$ generated by all $3 \times 3$ minors of all *edge flattenings* of tensor $P = (p_{i_1,\ldots,i_n})$:
$2^r \times 2^{n-r}$-matrix $Flat_{e,T}(P)$

$$Flat_{e,T}(P)(u,v) = P(u_1, \ldots, u_r, v_1, \ldots, v_{n-r})$$

where edge $e$ removal separates boundary distribution into $2^r$ variable and $2^{n-r}$ variables

## Procedure

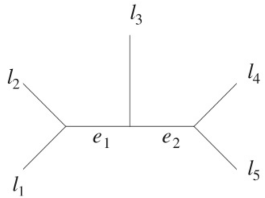
- set of languages $\mathcal{L} = \{\ell_1, \ldots, \ell_n\}$ (selected subfamily)
- set of SSWL syntactic parameters mapped for all: $\pi_i$, $i = 1, \ldots, N$
- gives vectors $\pi_i = (\pi_i(\ell_j)) \in \mathbb{F}_2^n$
- compute frequencies

$$P = \{p_{i_1,\ldots,i_n} = \frac{N_{i_1,\ldots,i_n}}{N}\}$$

  with $N_{i_1,\ldots,i_n}$ = number of occurrences of binary string $(i_1, \ldots, i_n) \in \mathbb{F}_2^n$ among the $\{\pi_i\}_{i=1}^N$
- Given a *candidate tree* $T$, compute all $3 \times 3$ minors of each flattening matrix $Flat_{e,T}(P)$, for each edge
- evaluate $\phi_T(P)$ minimum absolute value of these minors
- smallest $\phi_T(P)$ selects best among candidate trees

## Simple examples

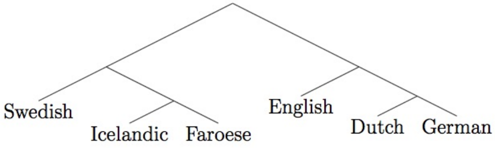
• PHYLIP and Splittree 4 misplace the position of Portuguese among the Latin languages, but phylogenetic invariants identify the correct tree ($\ell_1$ = French, $\ell_2$ = Italian, $\ell_3$ = Latin, $\ell_4$ = Spanish, $\ell_5$ = Portuguese)




$$\mathrm{Flat}_{e_2}(P) = \begin{pmatrix} p_{00000} & p_{00001} & p_{00010} & p_{00011} \\ p_{00100} & p_{00101} & p_{00110} & p_{00111} \\ p_{01000} & p_{01001} & p_{01010} & p_{01011} \\ p_{01100} & p_{01101} & p_{01110} & p_{01111} \\ p_{10000} & p_{10001} & p_{10010} & p_{10011} \\ p_{10100} & p_{10101} & p_{10110} & p_{10111} \\ p_{11000} & p_{11001} & p_{11010} & p_{11011} \\ p_{11100} & p_{11101} & p_{11110} & p_{11111} \end{pmatrix}$$

$$\mathrm{Flat}_{e_1}(P) = \begin{pmatrix} p_{00000} & p_{00001} & p_{00010} & p_{00011} & p_{00100} & p_{00101} & p_{00110} & p_{00111} \\ p_{01000} & p_{01001} & p_{01010} & p_{01011} & p_{01100} & p_{01101} & p_{01110} & p_{01111} \\ p_{10000} & p_{10001} & p_{10010} & p_{10011} & p_{10100} & p_{10101} & p_{10110} & p_{10111} \\ p_{11000} & p_{11001} & p_{11010} & p_{11011} & p_{11100} & p_{11101} & p_{11110} & p_{11111} \end{pmatrix}$$

• PHYLIP and Splittree 4 misplace the relative position of sub-branches of the Germanic languages, but phylogenetic invariants identify the correct tree (similar computation)



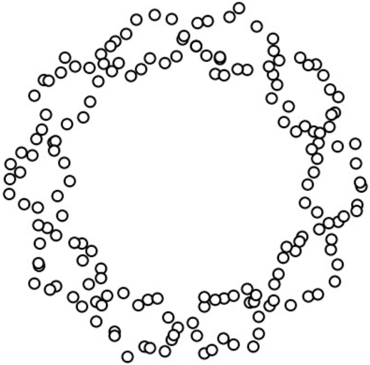

with correct subdivision into North Germanic and West Germanic sub-branches

Conclusion: work with smaller subfamilies, then paste together subtrees; use PHYLIP to generate candidate subtrees and phylogenetic algebraic geometry to select among them

## Persistent Topology of Syntax

• Alexander Port, Iulia Gheorghita, Daniel Guth, John M.Clark, Crystal Liang, Shival Dasu, Matilde Marcolli, *Persistent Topology of Syntax*, arXiv:1507.05134

## Persistent Topology of Data Sets

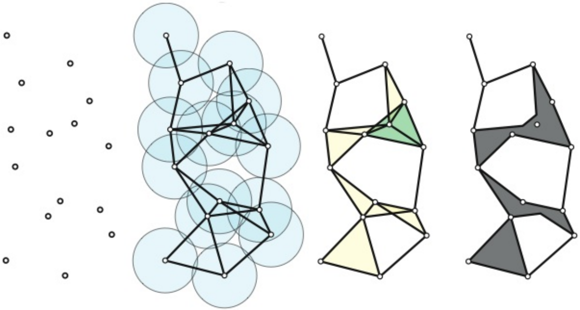


how data cluster around topological shapes at different scales

## Vietoris–Rips complexes

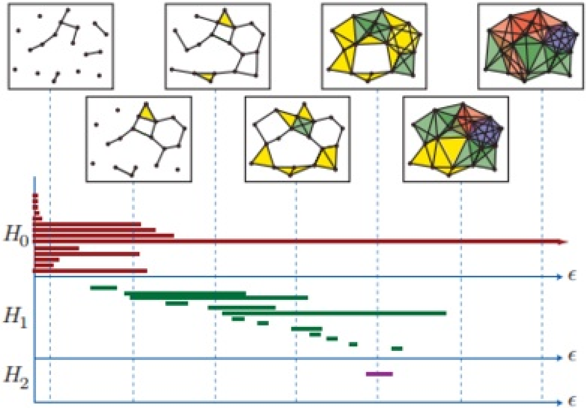
- set $X = \{x_\alpha\}$ of points in Euclidean space $\mathbb{E}^N$, distance $d(x,y) = \|x-y\| = (\sum_{j=1}^N (x_j - y_j)^2)^{1/2}$
- Vietoris-Rips complex $R(X,\epsilon)$ of scale $\epsilon$ over field $\mathbb{K}$:

$R_n(X,\epsilon)$ is $\mathbb{K}$-vector space spanned by all unordered $(n+1)$-tuples of points $\{x_{\alpha_0}, x_{\alpha_1}, \ldots, x_{\alpha_n}\}$ in $X$ where all pairs have distances $d(x_{\alpha_i}, x_{\alpha_j}) \le \epsilon$



(image by Jeff Erickson)

• inclusion maps $R(X, \epsilon_1) \hookrightarrow R(X, \epsilon_2)$ for $\epsilon_1 < \epsilon_2$ induce maps in homology by functoriality $H_n(X, \epsilon_1) \to H_n(X, \epsilon_2)$




(image by `forty.to`)

barcode diagrams: births and deaths of persistent generators

## Persistent Topology of Syntactic Parameters

• Data: 253 languages from SSWL with 116 parameters

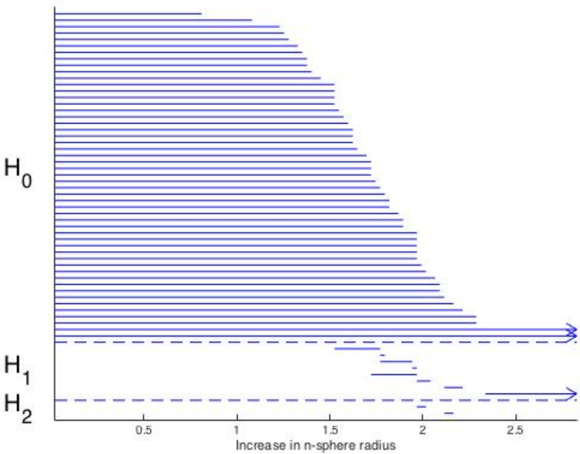
• if consider all world languages together too much noise in the persistent topology: subdivide by language families

• Principal Component Analysis: reduce dimensionality of data

• Compute Vietoris–Rips complex and barcode diagrams

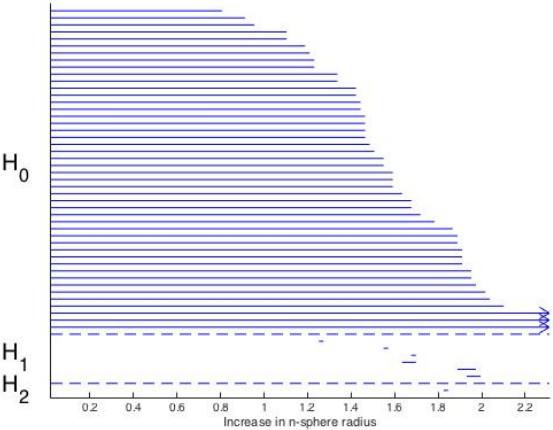
- Persistent $H_0$: clustering of data in components – language subfamilies
- Persistent $H_1$: clustering of data along closed curves (circles) – linguistic meaning?

# Persistent Topology of Indo-European Languages




- Two persistent generators of $H_0$ (Indo-Iranian, European)
- One persistent generator of $H_1$

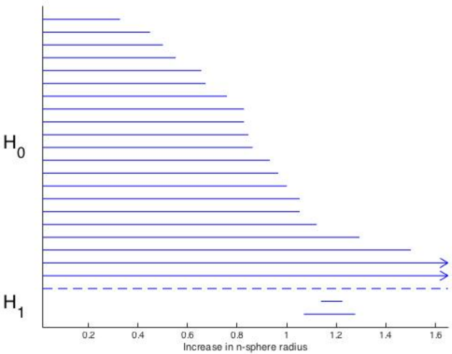
# Persistent Topology of Niger–Congo Languages




- Three persistent components of $H_0$
(Mande, Atlantic-Congo, Kordofanian)
- No persistent $H_1$

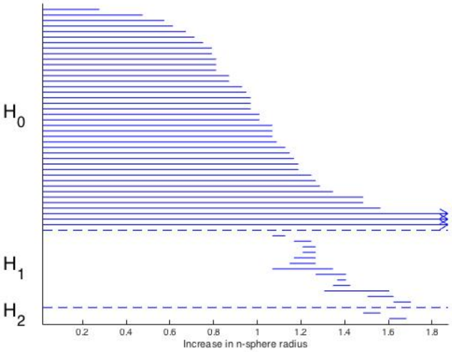
## What is the Indo-European $H_1$?

- naive guess: is it the Anglo-Norman bridge?
(but... lexical not syntactic!)
- No, definitely not the Anglo-Norman bridge!




Persistent topology of the Germanic+Latin languages

Answer: It's all because of Ancient Greek!




Persistent topology with Hellenic (and Indo-Iranic) branch removed

• it is related to influences (at the syntactic level) of the Hellenic branch on some Slavic languages (consistent with independent observations in new data by Longobardi, not analyzed yet topologically)

## So, what does topology tell us?

- it captures known historical-linguistics phenomena (clustering of syntactic structures by language families and sub-families)
- it is sensitive to more subtle phenomena, which are not seen in "phylogenetic trees" of languages: influences across different language sub-families ($H_1$ persistent generators)
- it can provide additional useful information on understanding how language (at the syntactic level) evolves

## Syntactic Parameters in Kanerva Networks

• 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*, arXiv:1510.06342

– Address two issues: relative prevalence of different syntactic parameters and "degree of recoverability" (as sign of underlying relations between parameters)
– If corrupt information about one parameter in data of group of languages can recover it from the data of the other parameters?
– Answer: different parameters have different degrees of recoverability
– Used 21 parameters and 165 languages from SSWL database

• Towards a possible model of how syntax is stored in the brain (Kanerva networks as models of associative memory)

Kanerva networks (sparse distributed memories)

- P. Kanerva, *Sparse Distributed Memory*, MIT Press, 1988.
- field $\mathbb{F}_2 = \{0, 1\}$, vector space $\mathbb{F}_2^N$ large $N$
- uniform random sample of $2^k$ hard locations with $2^k << 2^N$
- median Hamming distance between hard locations
- Hamming spheres of radius slightly larger than median value (access sphere)
- *writing to network*: storing datum $X \in \mathbb{F}_2^N$, each hard location in access sphere of $X$ gets $i$-th coordinate (initialized at zero) incremented depending on $i$-th entry ot $X$
- *reading at a location*: $i$-th entry determined by majority rule of $i$-th entries of all stored data in hard locations within access sphere

Kanerva networks are good at reconstructing corrupted data

## Procedure

• 165 data points (languages) stored in a Kanerva Network in $\mathbb{F}_2^{21}$ (choice of 21 parameters)

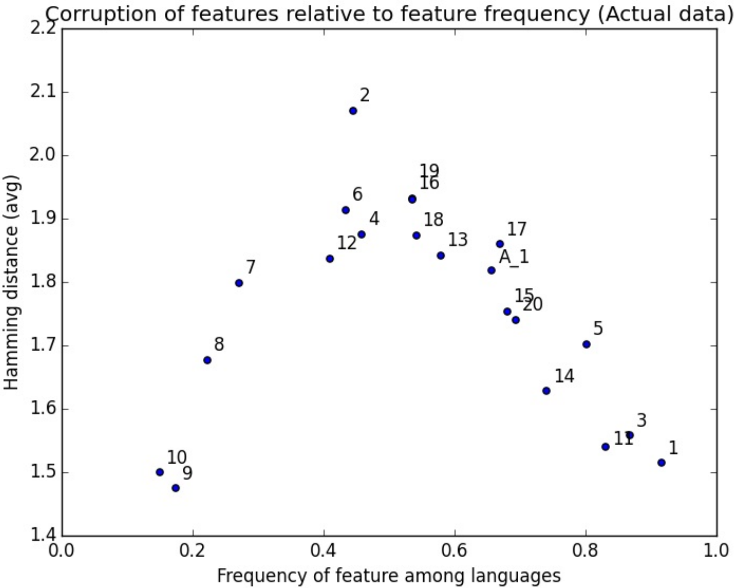
• corrupting one parameter at a time: analyze recoverability

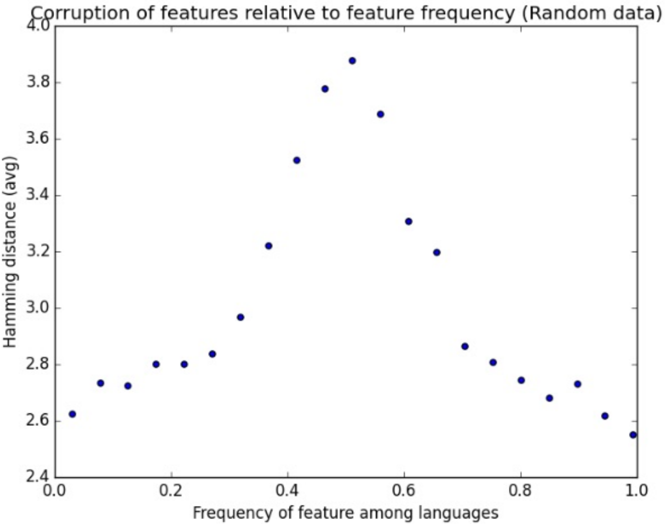
• language bit-string with a single corrupted bit used as read location and resulting bit string compared to original bit-string (Hamming distance)

• resulting average Hamming distance used as score of recoverability (lowest = most easily recoverable parameter)

# Parameters and frequencies

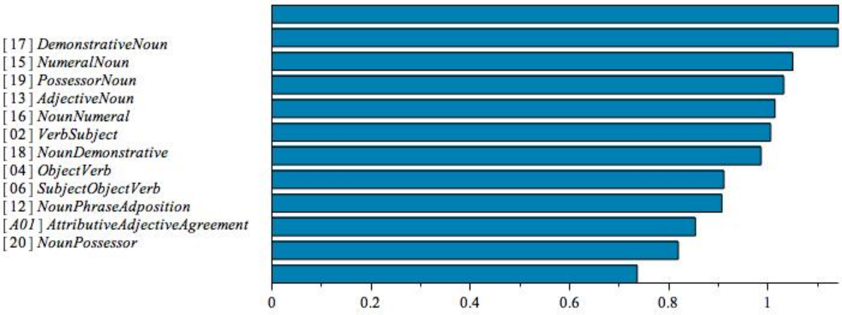
01 Subject-Verb (0.64957267)
02 Verb-Subject (0.31623933)
03 Verb-Object (0.61538464)
04 Object-Verb (0.32478634)
05 Subject-Verb-Object (0.56837606)
06 Subject-Object-Verb (0.30769232)
07 Verb-Subject-Object (0.1923077)
08 Verb-Object-Subject (0.15811966)
09 Object-Subject-Verb (0.12393162)
10 Object-Verb-Subject (0.10683761)
11 Adposition-Noun-Phrase (0.58974361)
12 Noun-Phrase-Adposition (0.2905983)
13 Adjective-Noun (0.41025642)
14 Noun-Adjective (0.52564102)
15 Numeral-Noun (0.48290598)
16 Noun-Numeral (0.38034189)
17 Demonstrative-Noun (0.47435898)
18 Noun-Demonstrative (0.38461539)
19 Possessor-Noun (0.38034189)
20 Noun-Possessor (0.49145299)
A01 Attributive-Adjective-Agreement (0.46581197)


Corruption of features relative to feature frequency (Actual data)
Hamming distance (avg)
Frequency of feature among languages
2.2
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
0.0
0.2
0.4
0.6
0.8
1.0
2
19
16
6
4
18
12
13
17
A_1
7
15
20
5
8
14
3
11
1
10
9




Overall effect related to relative prevalence of a parameter

More refined effect after normalizing for prelavence (syntactic dependencies)

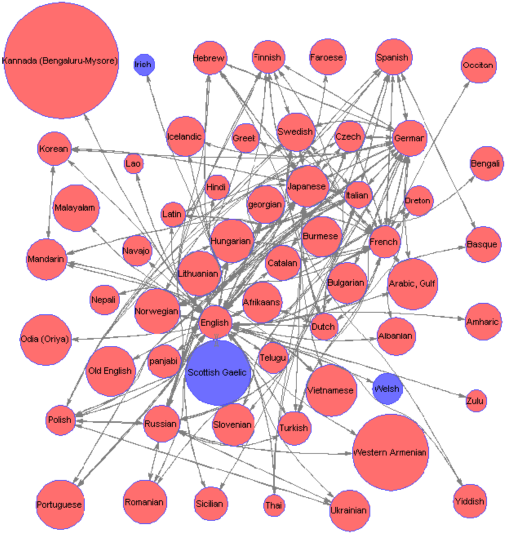



What does this tell us? some SSWL syntactic variables have a much higher degree of recoverability than others: consider them dependent variables; does this reflect how syntax is in fact stored in the human brain?

## Spin Glass Models of Language Evolution

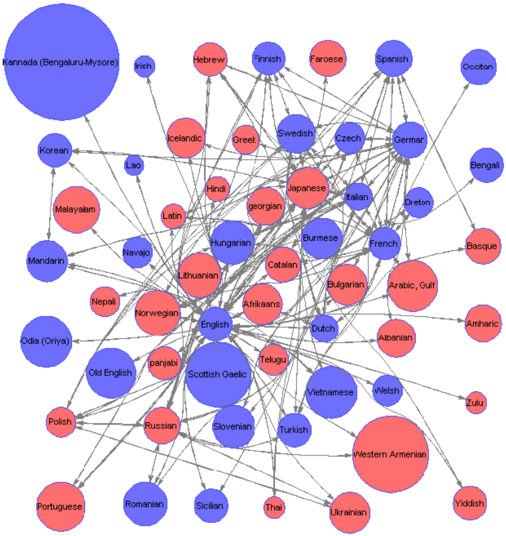
- Karthik Siva, Jim Tao, Matilde Marcolli, *Spin Glass Models of Syntax and Language Evolution*, arXiv:1508.00504
- syntactic parameters are dynamical: change over time, effects of interaction between languages (Ancient Greek switched SOV to SVO from Homeric to Classical; Sanskrit also switched by influence of Dravidian languages; also Old English to Middle English)
- physicist viewpoint: binary variables (up/down spins) that flip by effect of interactions: Spin Glass Model
- graph: vertices = languages, edges = language interaction (strength proportional to bilingual population)
- over each vertex a set of spin variables (syntactic parameters)
- if all syntactic parameters independent: uncoupled Ising models (low temperature: converge to more prevalent up/down state in initial configuration; high temperature fluctuations around zero magnetization state)

## Example: Single parameter dynamics *Subject-Verb* parameter




Initial configuration: most languages in SSWL have +1 for *Subject-Verb*; use interaction energies from MediaLab data

Equilibrium: low temperature all aligned to $+1$; high temperature:




Temperature: fluctuations in bilingual users between different structures ("code-switching" in Linguistics)

## Entailment relations among parameters

• relations recorded in the Longobardi-Guardiano data: cases where one state of a parameter can make another parameter undefined

• Example: $\{p_1, p_2\} = \{\text{Strong Deixis}, \text{Strong Anaphoricity}\}$

| | $p_1$ | $p_2$ |
|---|---|---|
| $\ell_1$ | $+1$ | $+1$ |
| $\ell_2$ | $-1$ | $0$ |
| $\ell_3$ | $+1$ | $+1$ |
| $\ell_4$ | $+1$ | $-1$ |

$\{\ell_1, \ell_2, \ell_3, \ell_4\} = \{\text{English}, \text{Welsh}, \text{Russian}, \text{Bulgarian}\}$

## Modeling Entailment

- variables: $S_{\ell,p_1} = \exp(\pi i X_{\ell,p_1}) \in \{\pm 1\}$, $S_{\ell,p_2} \in \{\pm 1, 0\}$ and $Y_{\ell,p_2} = |S_{\ell,p_2}| \in \{0,1\}$
- Hamiltonian $H = H_E + H_V$

$$H_E = H_{p_1} + H_{p_2} = - \sum_{\ell,\ell' \in \text{languages}} J_{\ell\ell'} \left( \delta_{S_{\ell,p_1}, S_{\ell',p_1}} + \delta_{S_{\ell,p_2}, S_{\ell',p_2}} \right)$$

$$H_V = \sum_{\ell} H_{V,\ell} = \sum_{\ell} J_\ell \, \delta_{X_{\ell,p_1}, Y_{\ell,p_2}}$$

$J_\ell > 0$ anti-ferromagnetic

- two parameters: *temperature* as before and coupling *energy of entailment*
- if freeze $p_1$ and evolution for $p_2$: Potts model with external magnetic field

Acceptance probabilities Metropolis–Hastings dynamics (some binary some ternary variables)

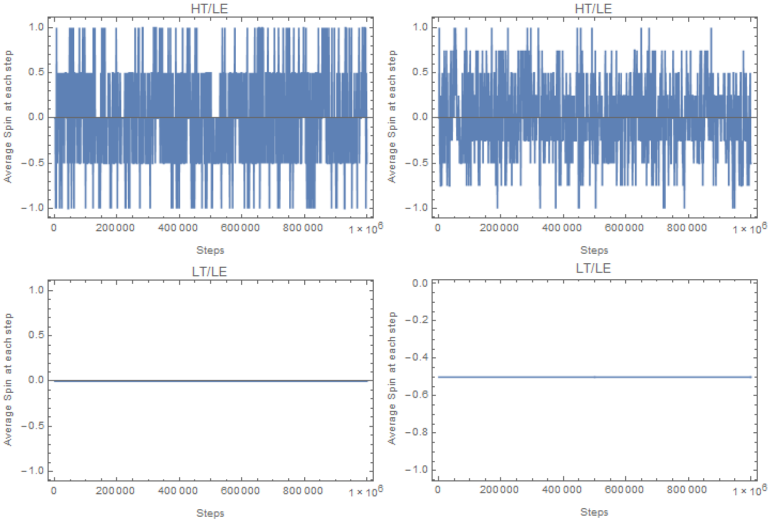
$$\pi_A(s \to s \pm 1 \pmod 3) = \begin{cases} 1 & \text{if } \Delta_H \leq 0 \\ \exp(-\beta \Delta_H) & \text{if } \Delta_H > 0. \end{cases}$$

$$\Delta_H := \min\{H(s + 1 \pmod 3), H(s - 1 \pmod 3)\} - H(s)$$

Equilibrium configuration

| $(p_1, p_2)$ | HT/HE | HT/LE | LT/HE | LT/LE |
|---|---|---|---|---|
| $\ell_1$ | $(+1, 0)$ | $(+1, -1)$ | $(+1, +1)$ | $(+1, -1)$ |
| $\ell_2$ | $(+1, -1)$ | $(-1, -1)$ | $(+1, +1)$ | $(+1, -1)$ |
| $\ell_3$ | $(-1, 0)$ | $(-1, +1)$ | $(+1, +1)$ | $(-1, 0)$ |
| $\ell_4$ | $(+1, +1)$ | $(-1, -1)$ | $(+1, +1)$ | $(-1, 0)$ |

# Average value of spin




$p_1$ left and $p_2$ right in low entailment energy case

• when consider more realistic models (28 languages and 63 parameters of Longobardi–Guardiano with all the entailment relations) very slow convergence of the Metropolis–Hastings dynamics even for low temperature

• how to get better information on the dynamics? consider set of languages as codes and an induced dynamics in the space of code parameters

Coding Theory to study how syntactic structures differ across the landscape of human languages

• Kevin Shu, Matilde Marcolli, *Syntactic Structures and Code Parameters*, arXiv:1610.00311

• 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 $\mathcal{L} = \{\ell_1, \dots, \ell_N\}$
- 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 $(R, \delta)$ 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 $\ell_1$ and $\ell_2$

• error correcting codes: optimize for maximal $R$ and $\delta$ but constraints that make them inversely correlated

• bounds in the space of code parameters $(R, \delta)$

## 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) between Gilbert-Varshamov and Plotkin curve

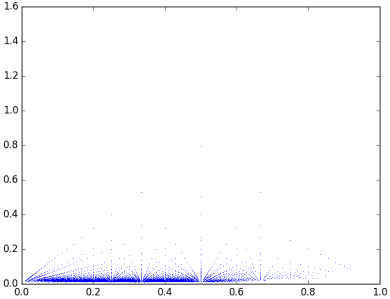
$$1 - H_q(\delta) \leq \alpha_q(\delta) \leq 1 - \delta/q$$

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

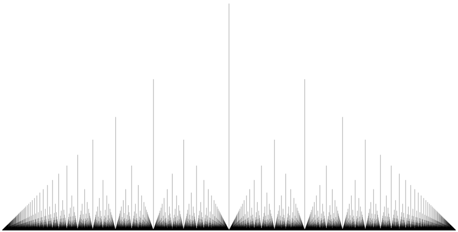
• asymptotic bound not explicitly computable (related to Kolmogorov complexity of codes, Manin–Marcolli)

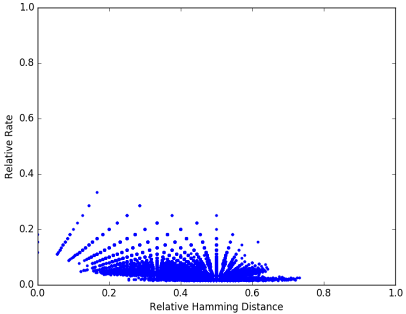
• difficult to construct codes above the asymptotic bound: examples from algebro-geometric codes from curves (but only for $q \geq 49$ otherwise entirely below the GV curve)

• 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)

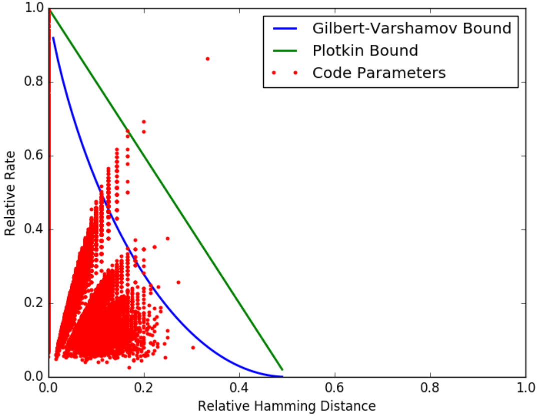





and behaves like the case of random codes with fixed $k = \log_2(N)$

• 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



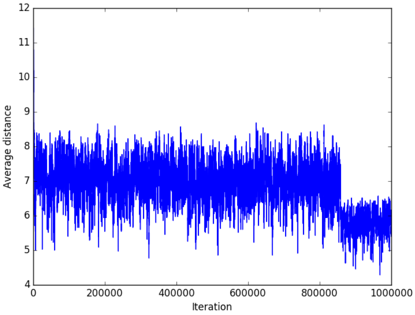

codes better than algebro-geometric above GV, asymptotic, and Plotkin

• Spin Glass Model dynamics for a set of languages $\mathcal{L}$ induces dynamics on codes $\mathcal{C}(\mathcal{L})$ and on code parameters $(R, \delta)$

• without entailment (independent parameters) fixed $R$ and $\delta$ flows towards zero (spoiling code)

• with entailment parameters dynamics can improve code making $\delta$ larger ($R$ fixed)

• in some cases can see better the dynamics on code parameter than with average magnetization of spin glass model



## Further Related Work

- Algebro-Geometric Models of Computational Semantics
  - Yuri Manin, Matilde Marcolli, *Semantic Spaces*, arXiv:1605.04238, to appear in Mathematics in Computer Science

- Generative Grammars and Renormalization
  - Matilde Marcolli, Alexander Port, *Graph Grammars, Insertion Lie Algebras, and Quantum Field Theory*, arXiv:1502.07796, Math. Comput. Sci. 9 (2015), no. 4, 391–408
  - Colleen Delaney, Matilde Marcolli, *Dyson-Schwinger equations in the theory of computation*, arXiv:1302.5040, in "Feynman amplitudes, periods and motives", pp.79–107, Contemp. Math., 648, Amer. Math. Soc., 2015
  - Matilde Marcolli, *Linguistic Merge and Dyson–Schwinger equations in Renormalization*, in preparation

## Conclusions (for now)

• import a set of different mathematical techniques (phylogenetic algebraic geometry, persistent topology, coding theory, statistical mechanics, geometric models of associative memory) in order to *study natural languages as dynamical objects*

• longer term goals: create mathematical and computational models of

1. how languages are acquired?
2. how languages are stored in the brain?
3. how languages change and evolve dynamically in time?

*for human languages viewed at the level of their syntactic structures*