Spin Glass models of Syntactic Parameters

Matilde Marcolli
MAT1509HS: Mathematical and Computational Linguistics

University of Toronto, Winter 2019, T 4-6 and W 4, BA6180 this lecture based on:

- Karthik Siva, Jim Tao, Matilde Marcolli, *Syntactic Parameters and Spin Glass Models of Language Change*, Linguistic Analysis, Vol. 41 (2017) N. 3-4, 559–608.

Spin Glass viewpoint

• historical examples syntactic parameters can flip by effect of interaction between languages: Sanskrit is believed to have flipped some syntactic parameters by influence of Dravidian languages...

• physicist viewpoint: binary variables (up/down spins) that flip by effect of interactions: Spin Glass Model

– focus on linguistic change caused by language interactions

– think of syntactic parameters as spin variables

– spin interaction tends to align (ferromagnet)

– strength of interaction proportional to bilingualism (MediaLab)

– role of temperature parameter: probabilistic interpretation of parameters & amount of code-switching in bilingual populations

– not all parameters are independent: entailment relations

– Metropolis–Hastings algorithm: simulate evolution

The Ising Model of spin systems on a graph G

• graph: vertices = languages, edges = language interaction (strength proportional to bilingual population); over each vertex a set of spin variables (syntactic parameters)

• configurations of spins $s : V(G) \to \{\pm 1\}$

• magnetic field B and correlation strength J: Hamiltonian

$$H(s) = -J \sum_{e \in E(G) : \partial(e) = \{v, v'\}} s_v s_{v'} - B \sum_{v \in V(G)} s_v$$

• first term measures degree of alignment of nearby spins

• second term measures alignment of spins with direction of magnetic field

Equilibrium Probability Distribution

- Partition Function $Z_G(\beta)$

$$Z_G(\beta) = \sum_{s:V(G)\to\{\pm 1\}} \exp(-\beta H(s))$$

- Probability distribution on the configuration space: Gibbs measure

$$\mathbb{P}_{G,\beta}(s) = \frac{e^{-\beta H(s)}}{Z_G(\beta)}$$

- low energy states weight most
- at low temperature (large β): ground state dominates; at higher temperature (β small) higher energy states also contribute

Average Spin Magnetization

$$M_G(\beta) = \frac{1}{\#V(G)} \sum_{s:V(G)\to\{\pm 1\}} \sum_{v\in V(G)} s_v \, \mathbb{P}(s)$$

- Free energy $F_G(\beta, B) = \log Z_G(\beta, B)$

$$M_G(\beta) = \frac{1}{\#V(G)} \frac{1}{\beta} \left(\frac{\partial F_G(\beta, B)}{\partial B} \right) |_{B=0}$$

- if all syntactic parameters were independent: just have several uncoupled Ising models (low temperature: converge to more prevalent up/down state in initial configuration; high temperature fluctuations around zero magnetization state)

Syntactic Parameters and Ising/Potts Models

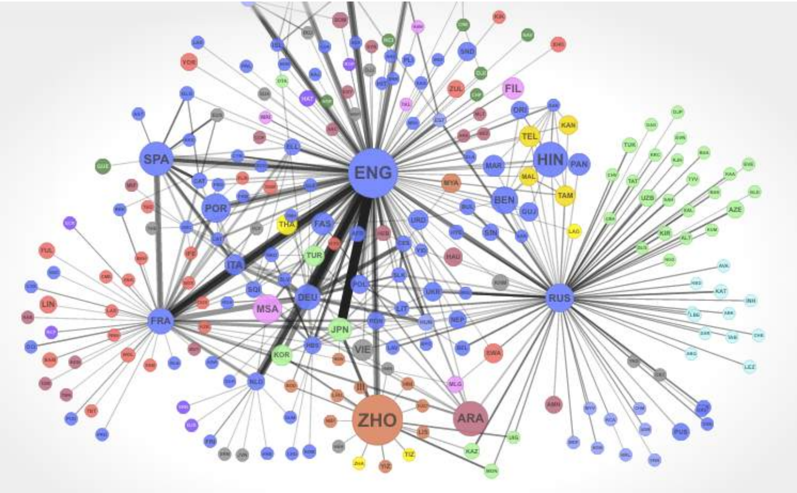
• characterize set of $n = 2^N$ languages $\mathcal{L}_i$ by binary strings of N syntactic parameters (Ising model)

• or by ternary strings (Potts model) if take values ± 1 for parameters that are set and 0 for parameters that are not defined in a certain language

• a system of n interacting languages = graph G with $n = \#V(G)$

• languages $\mathcal{L}_i$ = vertices of the graph (e.g. language that occupies a certain geographic area)

• languages that have interaction with each other = edges $E(G)$ (geographical proximity, or high volume of exchange for other reasons)



graph of language interaction (detail) from Global Language Network of MIT MediaLab, with interaction strengths J_e on edges based on number of book translations (or Wikipedia edits)

• if only one syntactic parameter, would have an Ising model on the graph G: configurations $s : V(G) \to \{\pm 1\}$ set the parameter at all the locations on the graph

• variable interaction energies along edges (some pairs of languages interact more than others)

• magnetic field B and correlation strength J: Hamiltonian

$$H(s) = - \sum_{e \in E(G): \partial(e) = \{v, v'\}} \sum_{i=1}^{N} J_e \, s_{v,i} \, s_{v',i}$$

• if N parameters, configurations

$$\underline{s} = (s_1, \ldots, s_N) : V(G) \to \{\pm 1\}^N$$

• if all N parameters are independent, then it would be like having N non-interacting copies of a Ising model on the same graph G (or N independent choices of an initial state in an Ising model on G)

Metropolis–Hastings

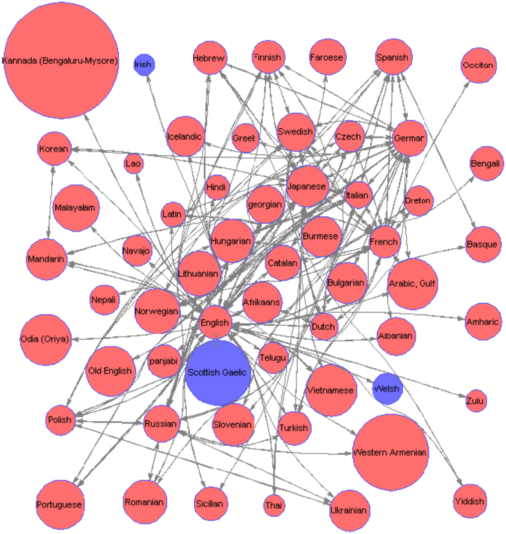
- detailed balance condition $\mathbb{P}(s)\mathbb{P}(s \to s') = \mathbb{P}(s')\mathbb{P}(s' \to s)$ for probabilities of transitioning between states (Markov process)
- transition probabilities $\mathbb{P}(s \to s') = \pi_A(s \to s') \cdot \pi(s \to s')$ with $\pi(s \to s')$ conditional probability of proposing state s' given state s and $\pi_A(s \to s')$ conditional probability of accepting it
- Metropolis–Hastings choice of acceptance distribution (Gibbs)

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

satisfying detailed balance

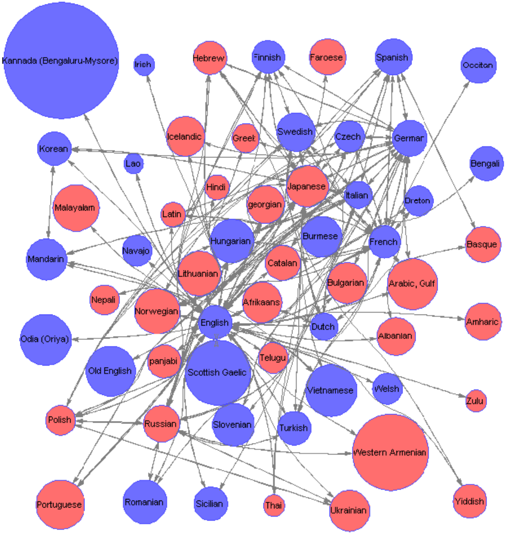
- selection probabilities $\pi(s \to s')$ single-spin-flip dynamics
- ergodicity of Markov process $\Rightarrow$ unique stationary distribution

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

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

	p_1	p_2
ℓ_1	$+1$	$+1$
ℓ_2	-1	0
ℓ_3	$+1$	$+1$
ℓ_4	$+1$	-1

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

Strong Deixis $+1$: governs possible positions of demonstratives in the nominal domain
Strong Anaphoricity $+1$: obligatory dependence on an antecedent in a local and asymmetric relation to anaphor

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

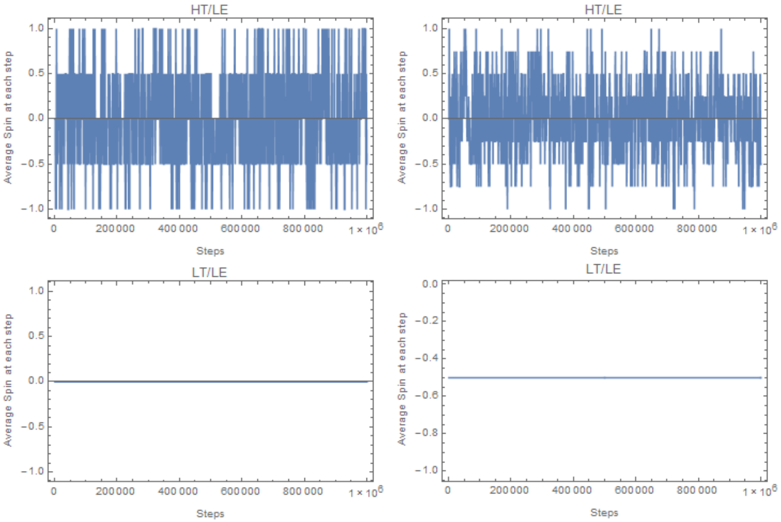
$$\pi_A(s \to s \pm 1 \pmod 3) = \begin{cases} 1 & \text{if } \Delta_H \leq 0 \\ \exp(-\beta\Delta_H) & \text{if } \Delta_H > 0. \end{cases}$$

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

Equilibrium configuration

(p_1, p_2)	HT/HE	HT/LE	LT/HE	LT/LE
ℓ_1	$(+1, 0)$	$(+1, -1)$	$(+1, +1)$	$(+1, -1)$
ℓ_2	$(+1, -1)$	$(-1, -1)$	$(+1, +1)$	$(+1, -1)$
ℓ_3	$(-1, 0)$	$(-1, +1)$	$(+1, +1)$	$(-1, 0)$
ℓ_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 (at least the 28 languages and 63 parameters of Longobardi–Guardiano with all their 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

• to be discussed later: a coding theory perspective on code parameters; induced dynamics on the space of codes shows more easily long term behavior of the system

How to improve this dynamical model?

• language change is related to mechanisms of language acquisition

• dynamical systems models of language acquisition were proposed by Berwick and Niyogi based on a Markov model on a space of possible grammats (in the formal languages sense)

• would like to couple the spin glass dynamics capturing language interaction through code-switching and bilingualism to a dynamical model of language acquisition

Next to be discussed: how to detect relations between syntactic parameters? what is the manifold of syntax?