Prevalence and Recoverability of Syntactic Parameters in Sparse Distributed Memories

Matilde Marcolli

GSI 2017: Geometric Structures of Information
Paris, November 5–9, 2017

Talk based on

- 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 (in this conference volume)

General 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
- 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

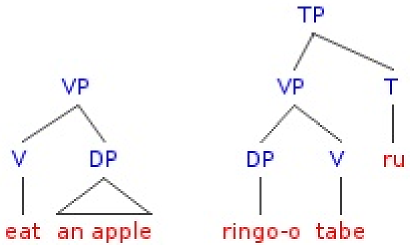
• however, surprisingly no mathematical model of Principles and Parameters formulation of Linguistics has been developed so far

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 by Giuseppe Longobardi, 2016

• Data Analysis of syntax of world languages with various mathematical tools (persistent topology, etc.)

• Goal: quantitively detect dependence relations between syntactic parameters

Expression frequencies of parameters among languages

- Example: Word Order: SOV, SVO, VSO, VOS, OVS, OSV

Word Orders	Percentage		
SOV	41.03%	Subject-initial	Specifier-Head
SVO	35.44%		
VSO	6.90%	Subject-medial	Head-Specifier
VOS	1.82%	Subject-final	
OVS	0.79%		
OSV	0.29%	Subject-medial	Specifier-Head

Very unevenly distributed across world languages

• Word order distribution: a neuroscience explanation?

- D. Kemmerer, *The cross-linguistic prevalence of SOV and SVO word orders reflects the sequential and hierarchical representation of action in Broca's area*, Language and Linguistics Compass, 6 (2012) N.1, 50–66.

• Internal reasons for diachronic switch?

- F.Antinucci, A.Duranti, L.Gebert, *Relative clause structure, relative clause perception, and the change from SOV to SVO*, Cognition, Vol.7 (1979) N.2 145–176.

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

Memory items in SDM: most items unrelated but most pairs linked by few intermediaries

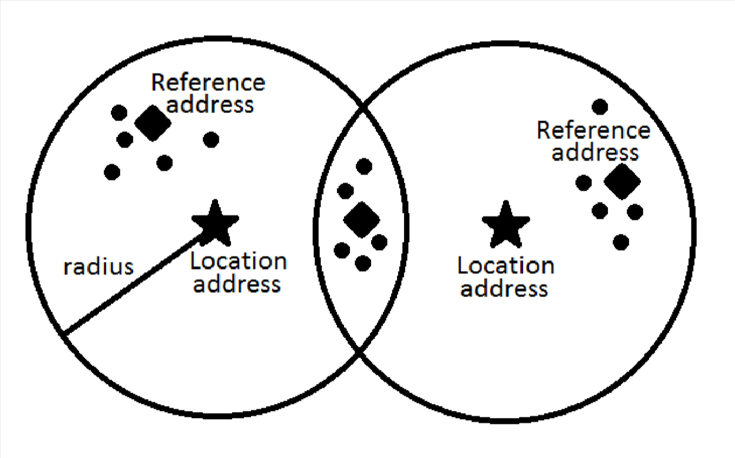


illustration from: Ján Kvak, *Creating and Recognizing Visual Words Using Sparse Distributed Memory* writing at ξ stores a copy at each hard location within the access sphere of ξ; reading at x retrieves content of all hard locations in the access sphere of x and averages them by majority rule

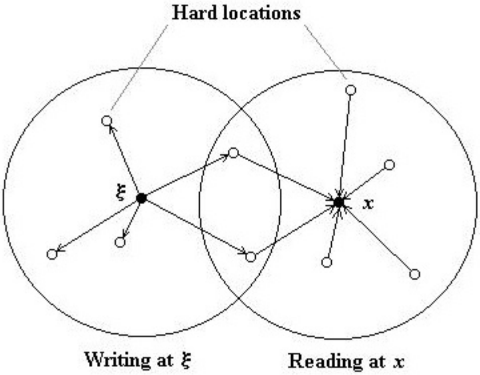


illustration from Jim Marshall's lecture notes on SDM proposed as a realistic computational model of how information is stored and retrieved in human memory

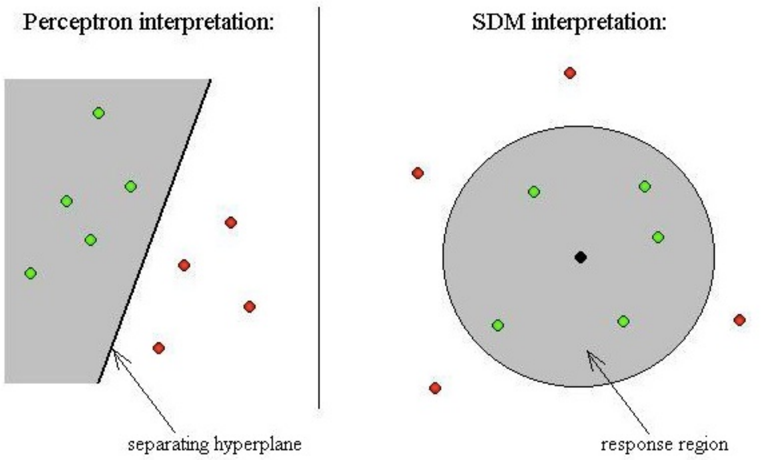


illustration from Jim Marshall's lecture notes on SDM

Procedure

• 165 data points (languages) stored in a Kanerva Network in $\mathbb{F}_2^{21}$ (choice of 21 parameters that are accurately mapped for all of these languages and for which one knows there are relations)

• corrupting one parameter at a time: analyze recoverability

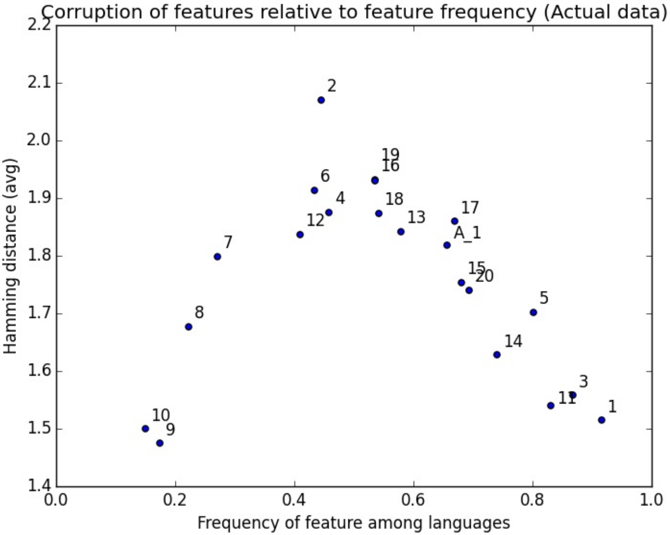
• language bit-string with a single corrupted bit used as read location and resulting bit string compared to original bit-string (Hamming distance)

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

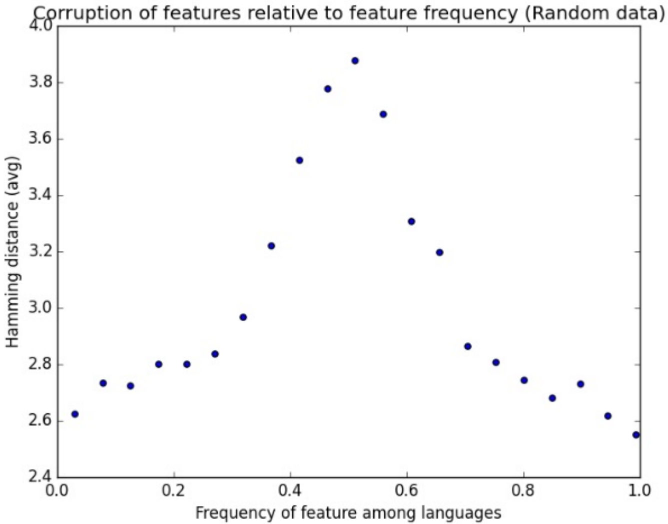
Parameters and frequencies (as classified in SSWL)

- 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)

Recoverability in Kanerva Networks

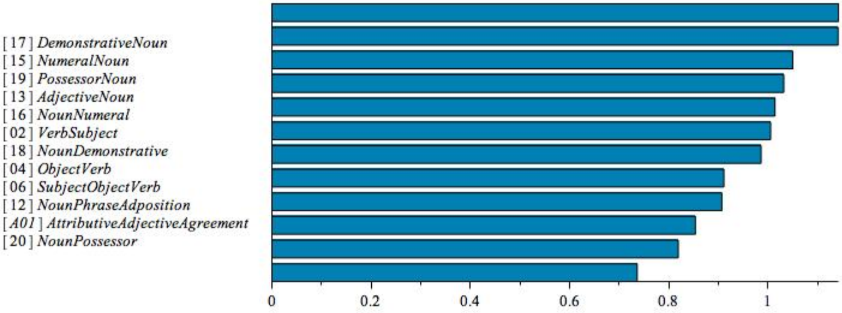


Specific effects due to individual parameters



Overall effect related to relative prevalence of a parameter

More refined effect after normalizing for prelavence (extracting effect of syntactic dependencies)



- Overall effect relating recoverability in a Kanerva Network to prevalence of a certain parameter among languages (depends only on frequencies: see in random data with assigned frequencies)

- Additional effects (that deviate from random case) which detect possible dependencies among syntactic parameters: increased recoverability beyond what effect based on frequency

- Possible neuroscience implications? Kanerva Networks as models of human memory (parameter prevalence linked to neuroscience models)

- More refined data if divided by language families?

. . . WORK IN PROGRESS