Overcoming Poverty of Stimulus with Structure and Parameters

Dakotah Lambert, Jon Rawski, Jeffrey Heinz

DEPARTMENT OF LINGUISTICS
STONY BROOK

iACS
INSTITUTE FOR ADVANCED COMPUTATIONAL SCIENCE

Talk in a Nutshell

Phonological Induction succeeds despite PovStim (Pearl in press)

- sparse, underdetermined, Zipf-distributed input
- lack of invariance in the signal

Successful learning requires **restricted hypothesis space/class**

- What are structural conditions on this space?
  - Necessary: **Regular** (Kaplan & Kay 1994)
  - Necessary and Sufficient: **Subregular** (Heinz 2018)
- Why does it look that way?
  - It emerges from learning parameters! (Heinz 2010)
  - A typology of mental representations
  - A typology of learning algorithms
  - Learners entertain the **simplest** of these

Today we will show this via phonotactics

Cognitive Complexity from First Principles

What kinds of distinctions does a cognitive mechanism need to be sensitive to in order to classify forms with respect to a pattern?

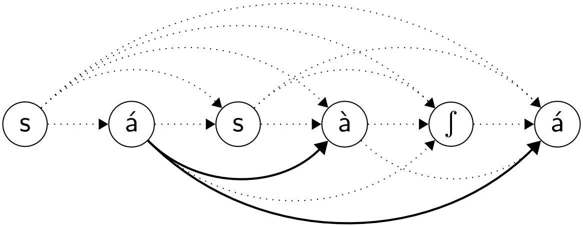
Reasoning about patterns

- What objects/entities/things are we reasoning about?
- What relationships between them are we reasoning with?

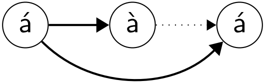
Finite Model Theory

- A ‘model’ of a word is a representation of it.
- A (Relational) Model contains two parts:
  - A **domain**: a finite set of elements.
  - **Relations** over domain elements.
- Every word has a model.
- Different words have different models.
- General: strings, trees, autosegmental graphs, etc
- Models are structures, sub-structures are called **Factors**

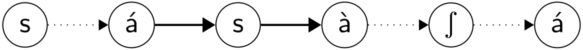
Precedence Model



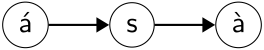
$$\langle \mathcal{D}; <, \mathsf{s}, \int, \text{á}, \text{à} \rangle$$



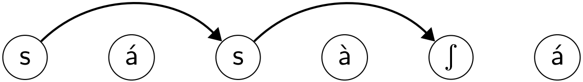
Successor Model



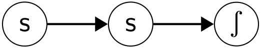
$$\langle \mathcal{D}; \lhd, \mathsf{s}, \textstyle\int, \text{á}, \text{à} \rangle$$



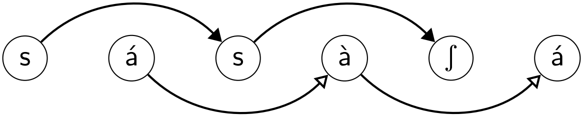
Tier-Successor Model



$$\langle \mathcal{D}; \triangleleft^{\{\text{s},\text{ʃ}\}}, \text{s}, \text{ʃ}, \text{á}, \text{à} \rangle$$



Multi-Tier Successor Model



$$\langle \mathcal{D}; \lhd^{\{\mathsf{s},\int\}}, \lhd^{\{\mathsf{\grave{a}},\mathsf{\acute{a}}\}}, \mathsf{s}, \int, \mathsf{\grave{a}}, \mathsf{\acute{a}}\rangle$$

Online Learning Algorithms

Generalizing String Extension Learning (Heinz 2010), one can infer grammars from data by collecting

I. k-factors, but ignore words longer than k
II. Any k-factors
III. Sets of k-factors
IV. Multisets of k-factors, saturated by a constant t

Example: Learner sees 'aaaab', biased with $<$ model, k and $t = 2$

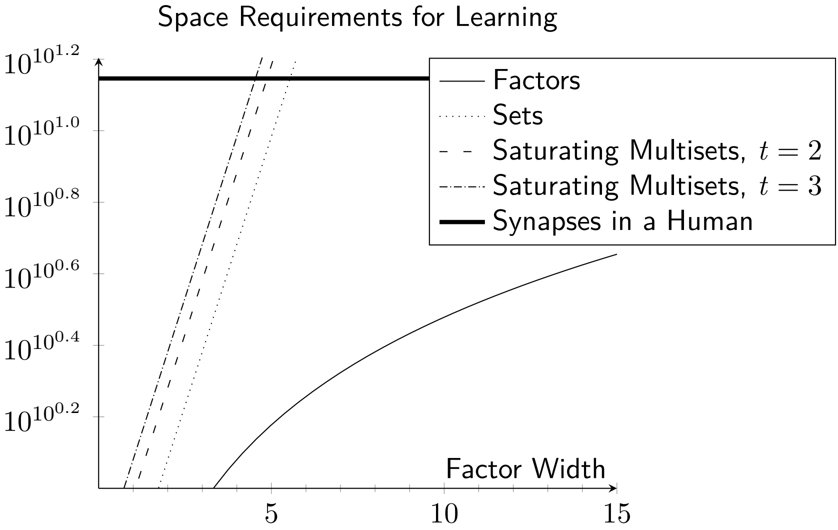
Algorithm	Resulting Grammar
I	$\varnothing$
II	$\{\text{aa}, \text{ab}\}$
III	$\{\{\text{aa}, \text{ab}\}\}$
IV	$\{\{\langle \text{aa}, 2\rangle, \langle \text{ab}, 1\rangle\}\}$

Deriving the Subregular Hierarchy

Regular

Star-Free

(IV) Saturating Multisets of Factors	Piecewise Testable	Locally Threshold Testable	Tier-Based Locally Threshold Testable	Multi-Tier Locally Threshold Testable
(III) Sets of Factors		Locally Testable	Tier-Based Locally Testable	Multi-Tier Locally Testable
(II) All Factors	Strictly Piecewise	Strictly Local	Tier-Based Strictly Local	Multi-Tier Strictly Local
(I) Certain Factors	Complement Strictly Piecewise	Complement Strictly Local	Complement Tier-Based Strictly Local	Complement Multi-Tier Strictly Local
	$<$	$\lhd$	$\lhd^{\mathcal{T}}$	$\{\lhd^{\mathcal{T}}\}$


Space Requirements for Learning
$10^{10^{1.2}}$
$10^{10^{1.0}}$
$10^{10^{0.8}}$
$10^{10^{0.6}}$
$10^{10^{0.4}}$
$10^{10^{0.2}}$
Factor Width
5
10
15
Factors
Sets
Saturating Multisets, $t = 2$
Saturating Multisets, $t = 3$
Synapses in a Human

Subregular + Space

Regular

Star-Free

(IV) Saturating Multisets of Factors	Piecewise Testable	Locally Threshold Testable	Tier-Based Locally Threshold Testable	Multi-Tier Locally Threshold Testable
(III) Sets of Factors		Locally Testable	Tier-Based Locally Testable	Multi-Tier Locally Testable
(II) All Factors	Strictly Piecewise	Strictly Local	Tier-Based Strictly Local	Multi-Tier Strictly Local
(I) Certain Factors	Complement Strictly Piecewise	Complement Strictly Local	Complement Tier-Based Strictly Local	Complement Multi-Tier Strictly Local
	$<$	$\lhd$	$\lhd^{\mathcal{T}}$	$\{\lhd^{\mathcal{T}}\}$

Subregular + Space + Linear Time

Regular

Star-Free

(IV) Saturating Multisets of Factors	Piecewise Testable	Locally Threshold Testable	Tier-Based Locally Threshold Testable	Multi-Tier Locally Threshold Testable
(III) Sets of Factors		Locally Testable	Tier-Based Locally Testable	Multi-Tier Locally Testable
(II) All Factors	Strictly Piecewise	Strictly Local	Tier-Based Strictly Local	Multi-Tier Strictly Local
(I) Certain Factors	Complement Strictly Piecewise	Complement Strictly Local	Complement Tier-Based Strictly Local	Complement Multi-Tier Strictly Local
	$<$	$\lhd$	$\lhd^{\mathcal{T}}$	$\{\lhd^{\mathcal{T}}\}$

Do the Predictions Match the Evidence?

Typological Data:

- many examples of SL, SP, TSL, MTSL
- Other classes: virtually none
- Aksenova & Deshmukh 2018: Multi-Tier Interactions restricted, suggests bias toward the linear-time classes

Laboratory Learning Results:

- Lai 2015, McMullin & Hansson 2019, Finley 2009: Learners biased towards SL, SP, TSL patterns

Conclusion

Learners efficiently overcome impoverished data through biases towards simple representations and learning parameters