Locality and learning over autosegmental representations

Adam Jardine and Jeffrey Heinz

Dept. of Linguistics & Cognitive Science
University of Delaware

LSA 2016
January 7, 2016

Introduction

- How do we learn over non-linear phonological structures?

Introduction

- Autosegmental phonological representations (APRs) still offer important insights in phonology (McCarthy, 2010; de Lacy, 2011; Walker, 2014, etc.), especially in tone (Hyman, 2011, 2014)
- There is cross-linguistic variation in what constitutes a well-formed APR
- How do we learn these language-specific well-formedness constraints?

Introduction

- Grammatical inference (de la Higuera, 2010) gives us ways to learn grammars organized around **local substructures** (García et al. 1990; Heinz 2010a; Jardine and Heinz, to appear)
- "Learning" = efficiently inducing the target pattern exactly given finite data (de la Higuera, 1997)

Introduction

- This talk extends this idea of learning local substructures to APRs
- By invoking grammars organized around **subgraphs**, we get a learning model which
  - **can** learn attested language-specific well-formedness constraints on APRs
  - **can** learn long-distance generalizations for which no proven learner exists
  - **cannot** learn complex, unattested patterns which require counting over the representation

Computational locality (strings)

- *blick* vs. **bnick*
- Constraint: *$\#$*bn*
- This is a **forbidden substructure constraint**

$$*\,\boxed{\#}\boxed{b}\boxed{n}$$

$$\boxed{\#}\boxed{b}\boxed{l}\boxed{ɪ}\boxed{k}\boxed{\#} \qquad \boxed{\boldsymbol{\#}\,\mathbf{b}\,\mathbf{n}}\boxed{ɪ}\boxed{k}\boxed{\#}$$

- $\#$*bn* is a **3-factor** of *$\#$*bnɪk*$\#$ (but not $\#$*blɪk*$\#$)
- Formal languages described with forbidden *k*-factors are the **Strictly *k*-Local (SL) languages** (McNaughton and Papert, 1971)

Computational locality (strings)

Learning forbidden substructure constraints (strings)

- Fix k

$$\texttt{fact}_k(w) = \{u \mid u \text{ is a } k\text{-factor of } w\}$$

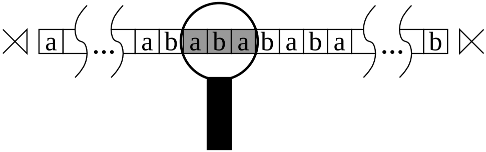
$$\texttt{fact}_k(D) = \bigcup_{w \in D} \texttt{fact}_k(w)$$

- $\texttt{fact}_3(\#bl\imath k\#) = \{\#bl, bl\imath, l\imath k, \imath k\#\}$

Computational locality (strings)

Learning forbidden substructure constraints (strings)

- `fact`$_k$ scans each input string with a window of size k



Rogers and Pullum (2011); Rogers et al. (2013)

Computational locality (strings)

Learning forbidden substructure constraints (strings)

- Let F_k be the set of all possible k-factors
- Learner based on $\texttt{fact}_k$:

<u>Time</u>	<u>Datum</u>	<u>Grammar</u>
(start)		F_k
0	w_0	$F_k - \texttt{fact}_k(\{w_0\})$
1	w_1	$F_k - \texttt{fact}_k(\{w_0, w_1\})$
2	w_2	$F_k - \texttt{fact}_k(\{w_0, w_1, w_2\})$
	$\ldots$	
n	w_n	$F_k - \texttt{fact}_k(\{w_0, w_1, w_2, ..., w_n\})$

- For any target grammar G of forbidden k-factors, there is some data point w_i for which the learner converges to G (García et al., 1990; Heinz, 2010b)
- $\texttt{fact}_k(D)$ calculated in time linear in size of D

Computational locality (strings)

- SL languages are are provably, efficently learnable
- **Tier-based Strictly *k*-Local (TSL) languages** are also so learnable, even without prior knowledge of the tier (Jardine and Heinz, to appear)

Tone and non-linear phonology

- Tone seems to need some representation beyond strings (Hyman, 2014)
- Some tone patterns (e.g., Hirosaki Japanese) are beyond power of even TSL grammars (Jardine, forthcoming)
- How do we learn over non-linear representations?
- First step: **local** theory of constraints for non-linear representations

Tone and non-linear phonology

Mende word tone (Leben, 1973; Goldsmith, 1976; Leben, 1978)

a. kɔ́	H	‘war’	b. pɛ́lɛ́	HH	‘house’	c. háwámá	HHH	‘waist’
d. kpà	L	‘debt’	e. bɛ̀lɛ̀	LL	‘pants’	f. kpàkàlì	LLL	‘stool’
g. mbû	F	‘owl’	h. ngílà	HL	‘dog’	i. félàmà	HLL	‘junction’
j. mbǎ	R	‘rice’	k. nìká	LH	‘cow’	l. ndàvúlá	LHH	‘sling’
m. mbǎ̂	R-F	‘comp.’	n. nyàhâ	LF	‘woman’	o. nìkílì	LHL	‘nut’

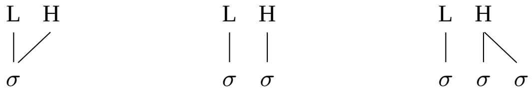
*LLH, *HHL, *RH, *RHH, etc.

Tone and non-linear phonology

Mende word tone (Leben, 1973; Goldsmith, 1976; Leben, 1978)

H	a. kɔ́	H	‘war’	b. pɛ́lɛ́	HH	‘house’	c. háwámá	HHH	‘waist’
L	d. kpà	L	‘debt’	e. bɛ̀lɛ̀	LL	‘pants’	f. kpàkàlì	LLL	‘stool’
HL	g. mbû	F	‘owl’	h. ngílà	HL	‘dog’	i. félàmà	HLL	‘junction’
LH	j. mbǎ	R	‘rice’	k. nìká	LH	‘cow’	l. ndàvúlá	LHH	‘sling’
LHL	m. mbǎ̂	R-F	‘comp.’	n. nyàhâ	LF	‘woman’	o. nìkílì	LHL	‘nut’

*LLH, *HHL, *RH, *RHH, etc.



Tone and non-linear phonology

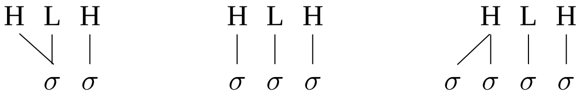
Hausa tone-integrating suffixes (Newman, 1986, 2000)

a. jáa	H	‘pull’	b. jíráa	HH	‘wait for’	c. béebíyáa	HHH	‘deaf mute’
c. wàa	L	‘who?’	d. màcè	LL	‘woman’	e. zàmfàrà	LLL	‘Zamfara’
f. jàakíi	LH	‘donkey’	g. jìmìnúu	HHL	‘ostriches’	h. bàbbàbbàkú	LLLH	‘roasted’
i. fáadì	HL	‘fall’	j. hántúnàa	LLH	‘noses’	k. búhúnhúnàa	HHHL	‘sacks’
l. mântá	FH	‘forget’	m. káràntá	HLH	‘read’	n. kákkáràntá	HHLH	‘reread’

Tone and non-linear phonology

Hausa tone-integrating suffixes (Newman, 1986, 2000)

a. jáa	H	‘pull’	b. jíráa	HH	‘wait for’	c. béebíyáa	HHH	‘deaf mute’
c. wàa	L	‘who?’	d. màcè	LL	‘woman’	e. zàmfàrà	LLL	‘Zamfara’
f. jàakíi	LH	‘donkey’	g. jìmìnúu	HHL	‘ostriches’	h. bàbbàbbàkú	LLLH	‘roasted’
i. fáadì	HL	‘fall’	j. hántúnàa	LLH	‘noses’	k. búhúnhúnàa	HHHL	‘sacks’
l. mântá	FH	‘forget’	m. káràntá	HLH	‘read’	n. kákkáràntá	HHLH	‘reread’



Tone and non-linear phonology

Kukuya word tone (Hyman, 1987; Zoll, 2003)

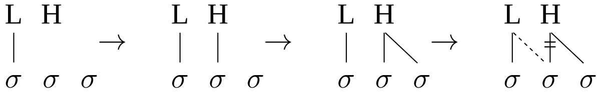
a. kâ ‘to pick’	F	b. sámà ‘conversation’	HL	c. káràgà ‘entangled’	HLL
d. sǎ ‘knot’	R	e. kàrá ‘paralytic’	LH	f. m^{w}àrə̀gí ‘brother’	LLH
g. bá ‘palms’	H	h. bágá ‘show knives’	HH	i. bálágá ‘fence’	HHH
j. bvǐ̀ falls’	R-F	k. pàlî ‘goes out’	LF	l. kàlə́gì ‘turns’	LHL

Tone and non-linear phonology

Previous approaches:

- Derivational frameworks parameterized directionality (e.g., Archangeli and Pulleyblank, 1994) and employed language specific rules for quality-dependent spreading (Hyman, 1987)

Kukuya:



- In OT, ALIGN for directionality and *CLASH for Kukuya (Zoll, 2003)

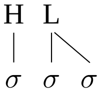
Tone and non-linear phonology

Previous approaches:

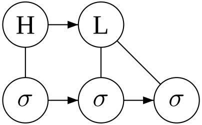
- Unclear how language-specific rules are learned
- ALIGN can generate complex patterns (Eisner, 1997) and cannot capture ‘edge-in’ patterns like in N. Karanga Shona (Hewitt and Prince, 1989; Zoll, 2003)

Computational locality (APRs)

- All of these patterns are **local**
- What is a substructure in an APR?

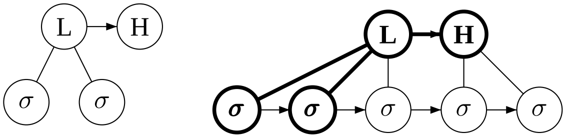


- APRs are **graphs** (Goldsmith, 1976; Coleman and Local, 1991)



Computational locality (APRs)

- Let a **subgraph** be some finite, connected piece of a graph



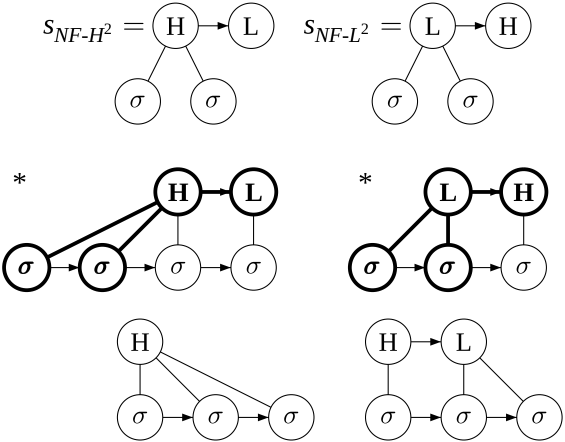
- Let k be number of nodes

Computational locality (APRs)

- A local grammar specifies sets of **forbidden subgraphs** $\{s_1, s_2, ..., s_n\}$, where each s_i is a subgraph
- This is a set of **inviolable** and **language-specific** constraints

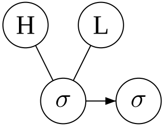
Case study: Mende

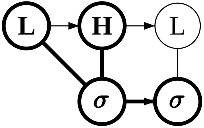
Plateaus in Mende

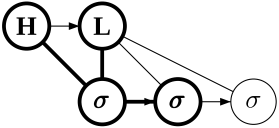


Case study: Mende

Contours in Mende

$s_{NF\text{-}Cont} =$ 

* 

* 

- c.f. Zhang (2000)

Case study: Mende

Melody constraint in Mende

$$s_{HLH} = \text{H} \rightarrow \text{L} \rightarrow \text{H}$$

$$\begin{tikzcd} {*}\ \text{L} \arrow[r] \arrow[d, dash] & \mathbf{H} \arrow[r] \arrow[d, dash] & \mathbf{L} \arrow[r] \arrow[d, dash] & \mathbf{H} \arrow[d, dash] \\ \sigma \arrow[r] & \sigma \arrow[r] & \sigma \arrow[r] & \sigma \end{tikzcd}$$

Local language-specific well-formedness

- Mende grammar: $\{s_{HLH}, s_{NF\text{-}Cont}, s_{NF\text{-}H^2}, s_{NF\text{-}L^2}\}$
- $k = 4$
- Language-specific constraints in Hausa, Kukuya, N. Karanga all local in this way (Jardine and Heinz, AMP)
- Long-distance patterns, like Hirosaki Japanese, are local over APRs (Jardine, forthcoming)
- Because they are local, do not overgenerate like ALIGN

Learning local APR constraints

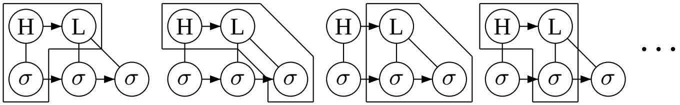
- Fix k

$$\mathtt{subg}_k(g) = \{s \mid s \text{ is a } k\text{-subgraph of } g\}$$

$$\mathtt{subg}_k(D) = \bigcup_{g \in D} \mathtt{subg}_k(g)$$

Learning local APR constraints

- Cognitive interpretation of $\texttt{subg}_k$ is the same as $\texttt{fact}_k$: scan input structures, remembering substructures of size k



Learning local APR constraints

- Let S_k be the set of all possible k-subgraphs
- Learner based on $\texttt{subg}_k$:

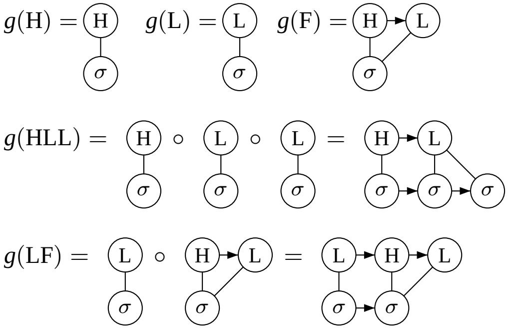
<u>Time</u>	<u>Datum</u>	<u>Grammar</u>
(start)		S_k
0	g_0	$S_k - \texttt{subg}_k(\{g_0\})$
1	g_1	$S_k - \texttt{subg}_k(\{g_0, g_1\})$
2	g_2	$S_k - \texttt{subg}_k(\{g_0, g_1, g_2\})$
	$\dots$	
n	g_n	$S_k - \texttt{subg}_k(\{g_0, g_1, g_2, ..., g_n\})$

- For any target grammar G of forbidden k-subgraphs, there is some data point g_i for which the learner converges to G
- Work in graph theory suggests $\texttt{subg}_k(D)$ is efficient (Ferreira, 2013)

Learning local APR constraints

A note on the input data

- Given correspondence of string symbols to graph primitives, input can be strings (Jardine and Heinz, 2015)



Discussion/conclusions

- This method learns autosegmental constraints
- This is thanks to a **theory** of constraints which captures the typology of tone patterns **locally**
- The result is a sufficient, yet restrictive model of tone-pattern learning
- It is able to capture long-distance patterns

Discussion/conclusions

- We can extend ideas to Optimal Domains Theory (Cassimjee and Kisseberth, 2001), correspondence theory (Rose and Walker, 2004; Shih and Inkelas, 2014), and others
- We can extend idea of locality to learning APR transformations (a la Chandlee, 2014, et seq.)

Acknowledgements

Many thanks to James Rogers, Rémi Eyraud, the Berkeley Phonology Phorum, audiences at AMP 2015 and NECPhon 2017, the students of Jeffrey Heinz's spring 2014 Doing Computational Phonology seminar, and the members of the UD Phonology & Phonetics and Computational Linguistics groups for their comments, questions, and inspiration.

References I

Archangeli, D. and Pulleyblank, D. (1994). *Grounded Phonology*. Cambridge: MIT Press.

Cassimjee, F. and Kisseberth, C. (2001). Zulu tonology and its relationship to other Nguni languages. In Kaji, S., editor, *Cross-linguistic studies of tonal phenomena: tonogenesis, Japanese accentology, and other topics.*, pages 327–359. Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA).

Chandlee, J. (2014). *Strictly Local Phonological Processes*. PhD thesis, University of Delaware.

Coleman, J. and Local, J. (1991). The “No Crossing Constraint” in autosegmental phonology. *Linguistics and Philosophy*, 14:295–338.

de la Higuera, C. (1997). Characteristic sets for polynomial grammatical inference. *Machine Learning*, 27(2):125–138.

de la Higuera, C. (2010). *Grammatical Inference: Learning Automata Grammars*. Cambridge University Press.

References II

de Lacy, P. (2011). Markedness and faithfulness constraints. In Oostendorp, M. V., Ewen, C. J., Hume, E., and Rice, K., editors, *The Blackwell Companion to Phonology*. Blackwell.

Eisner, J. (1997). What constraints should OT allow? Talk handout, Linguistic Society of America, Chicago. ROA#204-0797. Available at http://roa.rutgers.edu/.

Ferreira, R. (2013). *Efficiently Listing Combinatorial Patterns in Graphs*. PhD thesis, Università degli Studi di Pisa.

García, P., Vidal, E., and Oncina, J. (1990). Learning locally testable languages in the strict sense. In *Proceedings of the Workshop on Algorithmic Learning Theory*, pages 325–338.

Goldsmith, J. (1976). *Autosegmental Phonology*. PhD thesis, Massachussets Institute of Technology.

Haraguchi, S. (1977). *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology*. Tokyo: Kaitakusha.

Heinz, J. (2010a). Learning long-distance phonotactics. *Linguistic Inquiry*, 41:623–661.

References III

Heinz, J. (2010b). String extension learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 897–906. Association for Computational Linguistics.

Hewitt, M. and Prince, A. (1989). Ocp, locality, and linking: the N. Karanga verb. In Fee, E. J. and Hunt, K., editors, *WCCFL 8*, pages 176–191. Stanford: CSLI Publications.

Hyman, L. (1987). Prosodic domains in Kukuya. *Natural Language & Linguistic Theory*, 5:311–333.

Hyman, L. (2011). Tone: Is it different? In Goldsmith, J. A., Riggle, J., and Yu, A. C. L., editors, *The Blackwell Handbook of Phonological Theory*, pages 197–238. Wiley-Blackwell.

Hyman, L. (2014). How autosegmental is phonology? *The Linguistic Review*, 31:363–400.

Jardine, A. and Heinz, J. (2015). A concatenation operation to derive autosegmental graphs. In *Proceedings of the 14th Meeting on the Mathematics of Language (MoL 2015)*, pages 139–151, Chicago, USA. Association for Computational Linguistics.

References IV

Leben, W. R. (1973). *Suprasegmental phonology*. PhD thesis, Massachussets Institute of Technology.

Leben, W. R. (1978). The representation of tone. In Fromkin, V., editor, *Tone—A Linguistic Survey*, pages 177–219. Academic Press.

McCarthy, J. (2010). Autosegmental spreading in Optimality Theory. In Goldsmith, J. A., Hume, E., and Wezels, W. L., editors, *Tones and Features: Phonetic and Phonological Perspectives*, pages 195–222. De Gruyer Mouton.

McNaughton, R. and Papert, S. (1971). *Counter-Free Automata*. MIT Press.

Newman, P. (1986). Tone and affixation in Hausa. *Studies in African Linguistics*, 17(3).

Newman, P. (2000). *The Hausa Language: An encyclopedic reference grammar*. New Haven: Yale University Press.

Rogers, J., Heinz, J., Fero, M., Hurst, J., Lambert, D., and Wibel, S. (2013). Cognitive and sub-regular complexity. In *Formal Grammar*, volume 8036 of *Lecture Notes in Computer Science*, pages 90–108. Springer.

References V

Rogers, J. and Pullum, G. (2011). Aural pattern recognition experiments and the subregular hierarchy. *Journal of Logic, Language and Information*, 20:329–342.

Rose, S. and Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, 80:475–531.

Shih, S. and Inkelas, S. (2014). A subsegmental correspondence approach to contour tone (dis)harmony patterns. In Kingston, J., Moore-Cantwell, C., Pater, J., and Staubs, R., editors, *Proceedings of the 2013 Meeting on Phonology (UMass Amherst)*, Proceedings of the Annual Meetings on Phonology. LSA.

Walker, R. (2014). Nonlocal trigger-target relations. *Linguistic Inquiry*, 45:501–523.

Zhang, J. (2000). The phonetic basis for tonal melody mapping. In Billerey and Lillehaugen, editors, *WCCFL 19 Proceedings*, pages 603–616. Also available as ROA 451.

Zoll, C. (2003). Optimal tone mapping. *LI*, 34(2):225–268.

Hirosaki Japanese (Haraguchi, 1977)

- Exactly one H or F-toned mora in every word

In L_{HJ}	Not in L_{HJ}
H, F, LH, LF, HL,	L, LL, HH, HF,
LLH, LLF, LHL, HLL,	LLL, FLL, LFL, HLF,
LLLH, LLLF, LLHL, LHLL,	LLLL, LLFL, LFLL,
HLLL, ...	FLLL, HLLF, ...

- F can only be word final
- No k can exclude a word L^{k+1}
- One-H generalization is TSL, but final F is SL

Hirosaki Japanese

- Over APRs, HJ describable with

$$\{\, s_{HLH},\ s_{NF\text{-}Cont}, s_{\#\mathrm{L}\#},\ s_{H^2} \}$$

- $s_{\#\mathrm{L}\#} = \# \rightarrow \mathrm{L} \rightarrow \#$ $\quad s_{H^2} =$