Using Locality to Learn Long-distance Phonological Processes

Jane Chandlee and Jeffrey Heinz

University of Delaware

LSA 2015 | Portland, OR

Main objectives

- Define *Output Strict Locality* (OSL), a restrictive and defining computational property of phonological UR $\mapsto$ SR maps.
- Analyze various long-distance processes as the composition of OSL maps.
- Demonstrate how the OSL property can be used to learn phonological processes.

Processes as Maps

- Phonological processes are *maps* from an input (UR) to an output (SR) (cf. Tesar 2014).
- We want to identify the best (i.e., most restrictive) characterization of the class of maps that is needed for phonology.
  - Why? Identifying this class helps us understand 1) what is a 'possible' phonological process and 2) how processes are learned.

Learnability

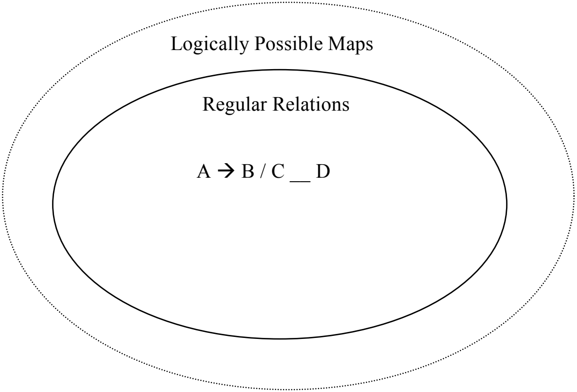
For a given class of maps, we want to know

1. Is there an algorithm that we can prove will learn any of the maps in the class?
2. Can it do so *efficiently* (i.e., with a reasonable amount of time and resources)?
3. What is the nature of the data it needs to learn the maps?

Methodology

1. Identify the defining property of the class of maps.
2. Prove that an algorithm that uses this property as an inductive principle will learn *all* and *only* the maps in that class.
3. Implication: no other maps can be learned and will therefore not be attested (i.e., the learner *is* the hypothesis space!).

Phonological maps are Regular Relations



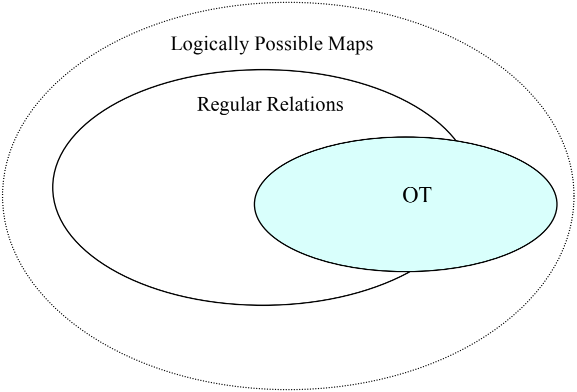
Johnson (1972); Kaplan and Kay (1994); Koskenniemi (1983)

Phonological maps are Regular Relations

- A → B / C __ D
- *$\mathrm{CAD} \gg \mathrm{FAITH}$(A↦B)

Baković (2013)

Phonological maps are Regular Relations

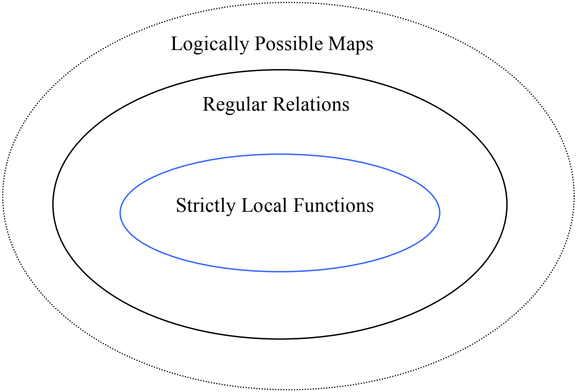


Riggle (2004); Gerdemann & Hulden (2012); Buccola 2013

Phonological maps are Regular Relations

- Not the best characterization.
  - Regular relations can also describe rules that are unattested and phonologically 'odd'.
- Doesn't help with learning.
  - The class of regular relations is not learnable from positive data.

Phonological maps are Strictly Local Functions



Chandlee (2014)

Strictly Local Functions

- Input Strictly Local Functions (ISL)
- Output Strictly Local Functions (OSL)

Input Strictly Local Functions

- An ISL function is actually ISL-k for some integer k.
- The output for a given input can be determined by only paying attention to k contiguous segments at a time (i.e., bounded memory).
- For a phonological map, k = the length of the target + the triggering context.

Input Strictly Local Functions

(1) Nasal place assimilation

a. $f(\text{iN+possible}) = \text{impossible}$

$k = |\text{Np}| = 2$

Input Strictly Local Functions

i N p o s s i b l e

Input Strictly Local Functions

#	i	N	p	o	s	s	i	b	l	e	#
#	i										

Input Strictly Local Functions

#	i	N	p	o	s	s	i	b	l	e	#
#	i										

Input Strictly Local Functions

#	i	N	p	o	s	s	i	b	l	e	#
#	i		mp								

Input Strictly Local Functions

#	i	N	p	o	s	s	i	b	l	e	#
#	i		mp	o							

Input Strictly Local Functions

#	i	N	p	o	s	s	i	b	l	e	#
#	i		mp	o	s						

ISL: Typological Prediction

- A significant range of phonological processes (local substitution, deletion, epenthesis, synchronic metathesis) can be modeled with ISL functions.
- Approximately 95% of the processes in P-Base (v.1.95, Mielke (2008)) meet this criterion (i.e., are ISL).
- Chandlee, Jardine, Heinz (2014) also show that the opaque maps discussed in Baković (2007) are ISL.

ISL: Learning Result

The class of ISL functions is efficiently identifiable in the limit from positive data (Chandlee, Eyraud, Heinz 2014) in the sense of de la Higuera (1997).

Non-ISL processes

Because ISL functions only pay attention to the input, the trigger of the process must be present underlyingly.

Nasal spreading

(2) Johore Malay (Onn 1980)

a. [+nasal][-cons] ↦ [+nasal][+nasal]

b. pəŋawasan ↦ pəŋãw̃ãsan, 'supervision'

Output Strictly Local Functions

#	p	ə	ŋ	a	w	a	s	a	n	#
#	p									

Output Strictly Local Functions

#	p	ə	ŋ	a	w	a	s	a	n	#
#	p	ə								

Output Strictly Local Functions

#	p	ə	ŋ	a	w	a	s	a	n	#
#	p	ə	ŋ							

Output Strictly Local Functions

#	p	ə	ŋ	a	w	a	s	a	n	#
#	p	ə	ŋ	ã						

Output Strictly Local Functions

#	p	ə	ŋ	a	w	a	s	a	n	#
#	p	ə	ŋ	ã	w̃					

Output Strictly Local Functions

#	p	ə	ŋ	a	w	a	s	a	n	#
#	p	ə	ŋ	ã	w̃	ã				

Spreading versus Harmony

- Though a target of the assimilation can be arbitrarily far from the trigger, the iterative nature of the spreading makes it a local process.
- This contrasts with harmony processes in which intervening 'potential' targets are skipped (Hansson 2001, Rose & Walker 2004).

Transparent Vowels

(3) Finnish front-back harmony (Goldsmith 1995, van der Hulst & van de Weijer 1995)

a.	værttinæ	‘spinning wheel’
b.	værttinæ-llæ-ni-hæn	‘with spinning wheel, as you know’
c.	tuoli	‘chair’
d.	tuoli-llɑ	‘on the chair’

Transparent Vowels

værttinæ-llA-nI-hAn

↧ ✓ OSL

værttinæ-llæ-nɯ-hæn

↧ ✓ OSL

værttinæ-llæ-ni-hæn

(Clements 1977; Vago 1980)

Transparent Vowels

- Assuming harmony is local (Gafos 1996, Ní Chiosáin & Padgett 1997), then transparent vowels create opaque maps:
  - Underapplication: high vowels don't harmonize
  - Overapplication: non-high vowels that follow non-harmonizing high vowels *do* harmonize

Transparent Vowels

Targeted Constraints analysis (Baković & Wilson 2000)

(4) Wolof ATR harmony (Archangeli & Pulleyblank 1994)

a.	reer-oon	‘was lost’
b.	rɛɛr-ɔɔn	‘had dinner’
c.	tɛɛr-uw-ɔɔn	‘welcomed’

Transparent Vowels

Targeted Constraints analysis (Baković & Wilson 2000)

tɛɛr-Uw-OOn ↦ tɛɛr-uw-ɔɔn

Agree(ATR) prefers the full assimilation candidate tɛɛr-ʊw-ɔɔn, but the targeted constraint →No(+HI,-ATR) selects the candidate the differs minimally from full assimilation (i.e., the transparency candidate): tɛɛr-uw-ɔɔn.

Long-distance Assimilation

(5) Kikongo (Rose and Walker 2004)

a. tu-kun-idi $\mapsto$ tu-kun-ini 'we planted'
b. tu-nik-idi $\mapsto$ tu-nik-ini 'we ground'

Long-distance Assimilation

tu-nik-idi

↧ ✓ OSL

tu-nĩk̃-ĩnĩ

↧ ✓ OSL

tu-nik-ini

Interim Summary

$$\begin{array}{ll} \text{harmony} & \leftarrow \text{OSL} \circ \text{OSL} \\ \downarrow & \\ \text{spreading} & \leftarrow \text{OSL} \\ \downarrow & \\ \text{adjacency} & \leftarrow \text{ISL} \end{array}$$

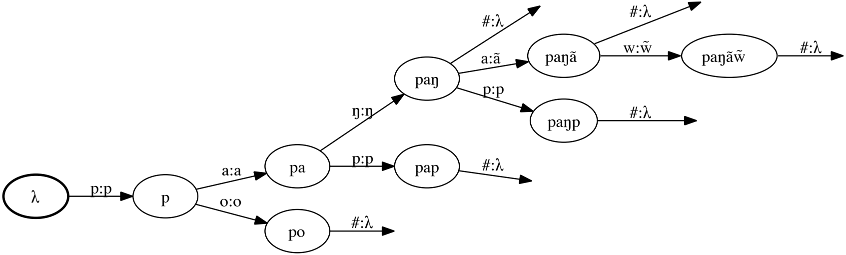
Learning OSL

- Modification to the ISLFL Algorithm already proven to learn the class of ISL functions (Chandlee, Eyraud, Heinz 2014).

Learning OSL

- Starts with a finite state representation of the observed data.
- {(paŋ,paŋ), (paŋa,paŋã), (paŋaw, paŋãw̃), (paŋs, paŋs), (pap, pap), (po, po) ... }

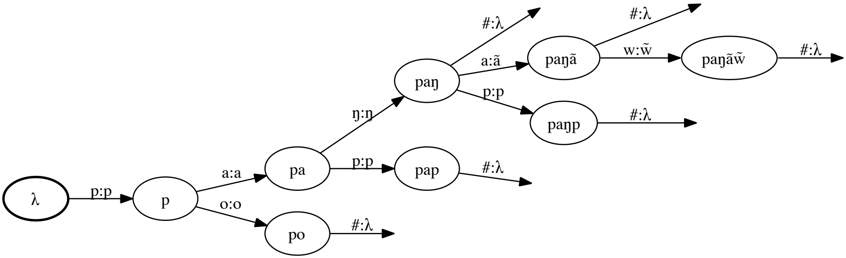
Learning OSL



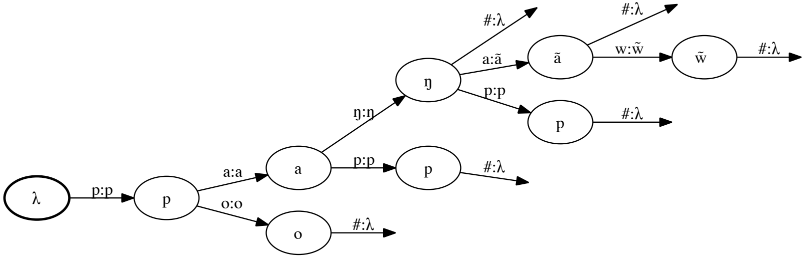
Learning OSL

- Generalize to unobserved inputs by merging states.
- State merging criterion reflects the defining property of the OSL class (i.e., only need to remember the 'most recent' output symbols).

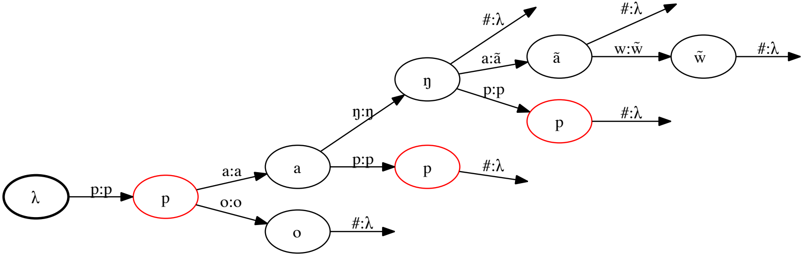
Learning OSL



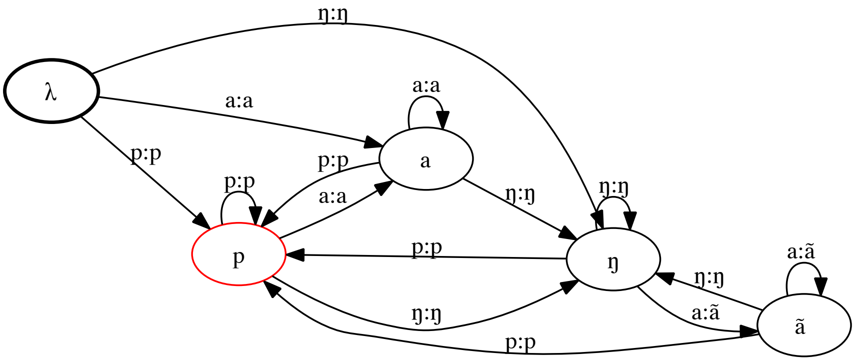
Learning OSL



Learning OSL



Learning OSL



Future Work

- What do the intermediate forms look like?
    - how abstract can they be?
- How is the composition learned?
    - Are the segments that don't surface still available within the 'feature space'?

Future Work

harmony	$\leftarrow$	TSL? or SP?
$\downarrow$		
spreading	$\leftarrow$	OSL
$\downarrow$		
adjacency	$\leftarrow$	ISL

Conclusions

- Proposal for representing and learning long-distance processes as well as an approach to the opacity problem.
- Progress on all goals for a theory of the phonological grammar: sufficiently expressive, maximally restrictive, and learnable.

References

- Archangeli, Diana and Douglas Pulleyblank. 1994. Grounded Phonology. Cambridge: MIT Press.
- Baković, Eric & Colin Wilson. (2000). Transparency, Strict Locality, and Targeted Constraints. In Billerey and Lillehaugen (eds.), *WCCFL 19 Proceedings*, Somerville, MA: Cascadilla Press. 43-56.
- Baković, Eric. (2007). A revised typology of opaque generalisations. *Phonology 24*(2). 217-259.
- Baković, Eric. (2013). *Blocking and Complementarity in Phonological Theory*. Bristol, CT: Equinox.
- Buccola, Brian. (2013). On the expressivity of Optimality Theory versus ordered rewrite rules. In *Proceedings of Formal Grammar 2012 and 2013, volume 8036 of Lecture Notes in Computer Science*, pages 142-158. Berlin: Springer-Verlag.
- de la Higuera, Colin. (1997). Characteristic sets for polynomial grammatical inference. *Machine Learning, 27*(2).125-138.
- Chandlee, Jane. (2014). *Strictly Local Phonological Processes*. Ph. D. thesis, University of Delaware.
- Chandlee, Jane, Rémi Eyraud, and Jeffrey Heinz. (2014a). Learning strictly local subsequential functions. *Transactions of the Association for Computational Linguistics 2*. 491-503.
- Chandlee, Jane, Adam Jardine, and Jeffrey Heinz. (2014b) Learning Repairs for Marked Structures (poster). Phonology 2014. MIT, September 19-21, 2014.
- Clements, G. N. (1977). Neutral vowels in hungarian vowel harmony: an autosegmental interpretation. In *Proceedings of the North Eastern Linguistic Society 7*. 49-64.
- Gafos, Diamandis. (1996). *The Articulatory Basis of Locality in Phonology.* PhD dissertation, Johns Hopkins University.
- Gerdemann, D. and M. Hulden (2012, July). Practical Finite State Optimality Theory. In *Proceedings of the 10th International Workshop on Finite State Methods and Natural Language Processing*, Donostia-San Sebastin, pp. 10-19. Association for Computational Linguistics.

References

- Goldsmith, John. (1985). Vowel harmony in Khalkha Mongolian, Yaka, Finnish and Hungarian. *Phonology Yearbook 2*. 2-23.
- Hansson, Gunnar. (2001). *Theoretical and Typological Issues in Consonant Harmony.* Ph.D. thesis, UC Berkeley.
- Johnson, C. Douglas. (1972). *Formal Aspects of Phonological Description*. The Hague: Mouton. Kaplan, R. M. and M. Kay (1994). Regular models of phonological rule systems. *Computational Linguistics 20*, 371-387.
- Koskenniemi, K. (1983). *Two-Level Morphology: A general computational model for word-form recognition and production*. University of Helsinki, Department of General Linguistics.
- Onn, F. M. (1980). *Aspects of Malay Phonology and Morphology: A Generative Approach*. Kuala Lumpur: Universiti Kebangsaan Malaysia.
- van der Hulst, Harry & J. van de Weijer. (1995). Vowel harmony. In J.A. Goldsmith (ed.), *The Handbook of Phonological Theory*. Oxford: Blackwell. 495-534.
- Mielke, Jeff. (2004). *P-Base 1.95*. http://137.122.133.199/ Jeff/pbase/.
- Ní Chiosáin, Máire and Jaye Padgett. (1997). Markedness, Segment Realization, and Locality in Spreading. Report LRC-97-01, Linguistics Research Center, University of California, Santa Cruz. [ROA-188.]
- Riggle, Jason. (2004). *Generation, Recognition, and Learning in Finite State Optimality Theory*. Ph. D. thesis, UCLA.
- Rose, Sharon and Rachel Walker. (2004). A typology of consonant agreement as correspondence. *Language 80*, 475-531.
- Tesar, Bruce. (2014). *Output-Driven Phonology: Theory and Learning*. Cambridge: Cambridge UP.
- Vago, R. M. (1980). *The sound pattern of Hungarian*. Washington: Georgetown University Press.

Schwa deletion

(6) French (Dell 1973, 1980, Noske 1993)

a. VCəCV ↦ VCCV

(7) tu devenais, ‘you became’

a. ty dəvənɛ ↦ ty dvənɛ

b. ty dəvənɛ ↦ ty dəvnɛ

c. ty dəvənɛ ↦ *ty dvnɛ

Strict Locality

Post-nasal obstruent voicing: Valid surface strings are a SL-2 language.

TVNDV in the language.
NVNTV in the language.

TVNTV *necessarily* in the language.

Optimization

$$\textsc{Ident}, \textsc{Dep} \gg *ab \gg \textsc{Max}$$

Non-regular relation:

$$a^n b^m \mapsto a^n, \text{ if } m < n$$
$$a^n b^m \mapsto b^m, \text{ if } n < m$$