Logical Characterizations of Local and Long-distance Phonological Agreement

Jane Chandlee

Department of Computer Science
Haverford College

May 7, 2016

Characterizing phonological maps

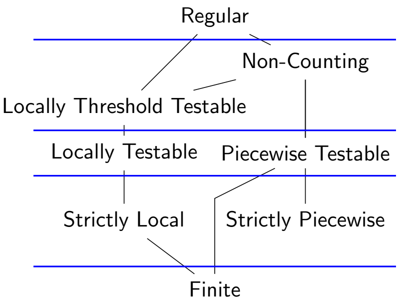
- **Goal**: identify the most restrictive computational complexity class that is sufficiently expressive for the set of possible phonological maps.
- Known computational complexity classes have multiple, converging characterizations.
  - finite state automata
  - formal language theory
  - algebraic
  - logical

Characterizing phonological maps

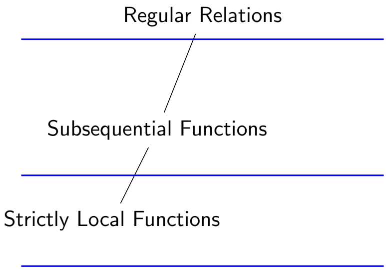
- Phonological maps are **regular relations** (Johnson 1972, Kaplan & Kay 1994)
- **Subregular** classes of functions appear to better characterize phonological maps (Mohri 1997; Chandlee et al. 2012; Gainor et al. 2012; Chandlee and Heinz 2012; Heinz and Lai 2013; Luo 2013; Payne 2013; Jardine, to appear)...
  - ...but not all classes are currently well understood.

Subregular hierarchies

Formal Languages



String-to-string maps



(Mohri 1997, Rogers & Pullum 2011, Rogers et al. 2013, Chandlee 2014)

Phonotactics

- Local phonotactics are **SL languages** (Heinz 2007, Heinz 2009)

(1) In a language with final devoicing:

a. Allowable substrings: $\{\rtimes$V, VV, V$\ltimes$, $\rtimes$D, $\rtimes$T, DV, DD, TV, ... $\}$

b. Prohibited substrings: $\{$ D$\ltimes$ $\}$

- Long-distance phonotactics are **SP** (Heinz 2010) or **TSL languages** (Heinz et al. 2011)

(2) In a language with sibilant harmony:

a. Allowable subsequences: { s...s, ʃ...ʃ, ...}

b. Prohibited subsequences: { s...ʃ, ʃ...s }

Maps

- Local phonological maps are **SL functions** (Chandlee 2014, Chandlee et al. 2015)

(3) D$\ltimes$ $\mapsto$ T$\ltimes$

  - FST characterization
  - Language-theoretic characterization

- **Hypothesis**: Long-distance maps are **SP functions**

(4) s...ʃ $\mapsto$ s...s

Why use logic?

- Goal is to build on what is known of SL functions to identify the class of SP functions.
- Logic provides a natural extension from local to long-distance.

Examples of agreement

- Adjacent

(5) Korean (Lee and Pater 2008)
/pa**p**mul/ ↦ [pa**m**mul] 'rice water'

- Non-adjacent but bounded

(6) Ndonga (Viljoen 1973, Rose and Walker 2004)

a. /ku**n**i**l**a/ ↦ [ku**n**i**n**a] 'sow for'
b. /**n**iki**l**a/ ↦ [**n**iki**l**a] 'season for'

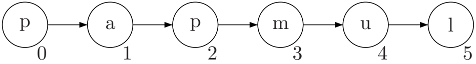
- Unbounded

(7) Kikongo (Meinhof 1932, Odden 1994, Rose and Walker 2004)
/kudu**m**ukisi**l**a/ ↦ [kudu**m**ukisi**n**a] 'to cause to jump for'

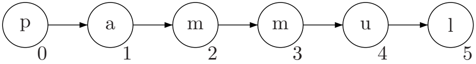
String graphs

pa**pm**ul $\mapsto$ pa**mm**ul

g

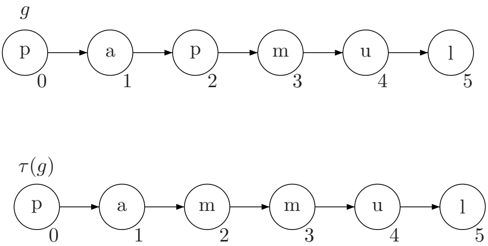


$\tau(g)$



String graphs

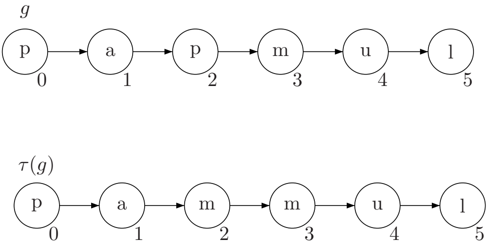
papmul $\mapsto$ pammul



- Set of nodes: $\{0, 1, 2, 3, 4, 5\}$
- Successor relation: $S(0) = 1$, $S(1) = 2$, etc. (also written $1 \lhd 2$)

String graphs

pa**pm**ul $\mapsto$ pa**mm**ul



- Labeling function: $\ell(0) = \mathsf{p}$, $\ell(1) = \mathsf{a}$, etc.

Natural class predicates

- $\mathsf{V}(x) = \mathsf{TRUE}$ iff $\ell(x) = \mathsf{a} \vee \ell(x) = \mathsf{u} \vee \ldots$
- $\mathsf{N}(x) = \mathsf{TRUE}$ iff $\ell(x) = \mathsf{n} \vee \ell(x) = \mathsf{m} \vee \ldots$
- $\mathsf{C}(x) = \mathsf{TRUE}$ iff $\neg\, \mathsf{N}(x) \wedge \neg\, \mathsf{V}(x)$

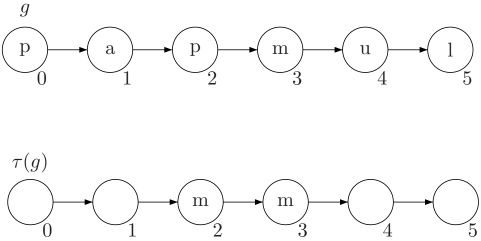
Substring predicates

- $\mathsf{CN}(x) = (\exists y)\ [\mathsf{C}(x) \wedge \mathsf{N}(y) \wedge x \lhd y]$

Graph transduction

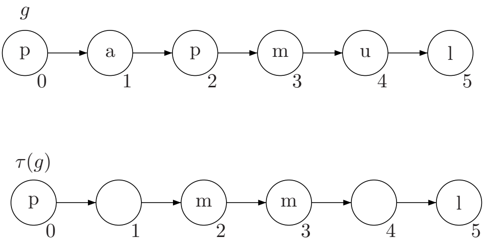
- Predicates define the nodes and labels of the output graph in terms of the input graph (Engelfriet and Hoogeboom 2001)

$$\varphi^0_N(x) = \mathsf{N}(x) \vee \mathsf{CN}(x)$$



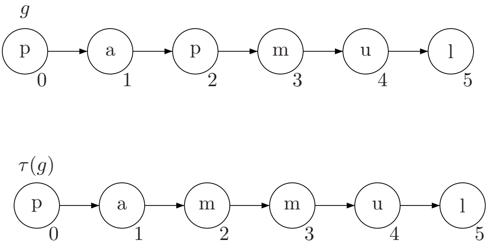
Graph transduction

$$\varphi_C^0(x) = \mathsf{C}(x) \wedge \neg\ \mathsf{CN}(x)$$



Graph transduction

$$\varphi_V^0(x) = \mathsf{V}(x)$$



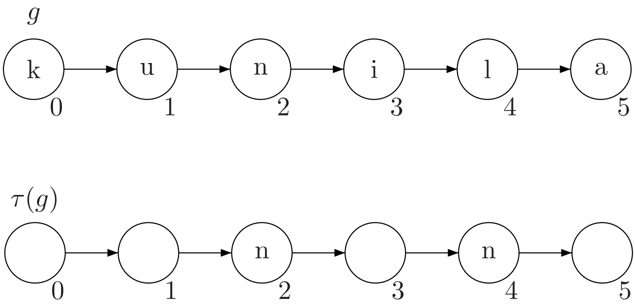
Non-adjacent but bounded

ku**n**i**l**a ↦ ku**n**i**n**a

$$\mathsf{NVL}(x) = (\exists y, z)\ [\mathsf{N}(y) \wedge \mathsf{V}(z) \wedge \mathsf{L}(x) \wedge y \triangleleft z \wedge z \triangleleft x]$$

Non-adjacent but bounded

$$\varphi_N^0(x) = N(x) \vee NVL(x)$$



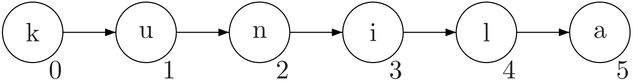
Non-adjacent but bounded

$$\varphi^0_L(x) = \mathsf{L}(x) \wedge \neg\, \mathsf{NVL}(x)$$

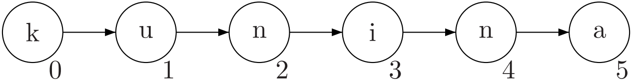
$$\varphi^0_V(x) = \mathsf{V}(x)$$

$$\varphi^0_D(x) = \mathsf{D}(x)$$

g



$\tau(g)$



Unbounded

tuku**nid**i $\mapsto$ tuku**nin**i

$$\mathrm{NVL}(x) = (\exists y, z)\ [\mathrm{N}(y) \wedge \mathrm{V}(z) \wedge \mathrm{L}(x) \wedge y \triangleleft z \wedge z \triangleleft x]$$

Unbounded

tu**n**iki**d**i ↦ tu**n**iki**n**i

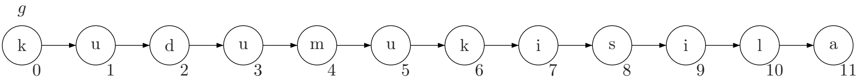
$$\mathrm{N}{-}{-}{-}\mathrm{L}(x) = (\exists v, w, y, z)\ [\mathrm{N}(y) \wedge \mathrm{L}(x) \wedge y \triangleleft z \wedge z \triangleleft w \wedge w \triangleleft v \wedge v \triangleleft x]$$

Unbounded

kudu**m**ukisi**l**a ↦ kudu**m**ukisi**n**a

$$\mathrm{N}{-}{-}{-}{-}{-}{-}\mathrm{L}(x) = (\exists s, t, u, v, w, y, z)\ [\mathrm{N}(y) \wedge \mathrm{L}(x) \wedge y \triangleleft z \wedge z \triangleleft w \wedge w \triangleleft v \wedge v \triangleleft u \wedge u \triangleleft t \wedge t \triangleleft s \wedge s \triangleleft x]$$

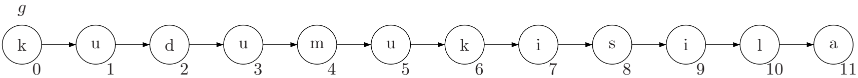
Instead of successor, **precedence**



- Successor: $x \lhd y$
  - S(0) = 1, S(1) = 2, ...
- Precedence: $x < y$
  - P(1) = {0}, P(2) = {0, 1}, ..., P(10) = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9 }

Instead of substring, **subsequence**

kudu**m**ukisi**l**a ↦ kudu**m**ukisi**n**a



$$\mathsf{NL}(x) = (\exists y)\ [\mathsf{N}(y) \wedge \mathsf{L}(x) \wedge y < x]$$

Categories

<table>
<tr><td>Local</td><td colspan="2">Long-distance</td></tr>
<tr><td>Adjacent</td><td colspan="2">Non-adjacent</td></tr>
<tr><td colspan="2">Bounded</td><td>Unbounded</td></tr>
<tr><td colspan="2">SL</td><td>SP</td></tr>
</table>

More examples of bounded LDA

(8) **Transvocalic**: Anywa, Basaa, Bemba, Bukusu, Dime, Gaagudju, Herero, Ila, Kipare, Koyra, Kwanyama, Lamba, Luba, Mayali, Mwiini, Nyangumarta, Pangwa, Pare, Pende, Punu, Ruund, Sawai, Shambaa, Sudanese, Suku, Tonga, Yabem, Zayse, ...

(9) **Co-occurrence restrictions within roots or morphemes**: Adhola, Alur, Anywa, Basque, Chaha, Chol, Ganda, Hausa, Ineseno Chumash, Izere, Kalasha, Karaim, Komi-Permyak, Kukuya, Luo, Malto, Ndebele, Ngizim, Pengo, Pohnpeian, Shilluk, Siwi, Teke-Gabon, Tiene, Yucatec Mayan, Zulu, ...

(see Shaw (1991), Odden (1994), Hansson (2001), Rose & Walker (2004), and references therein)

Domain restrictions

(10) Malto (Mahapatra 1979, Hansson 2001): {t, d} and {ʈ, ɖ} cannot co-occur in tautomorphemic CVC

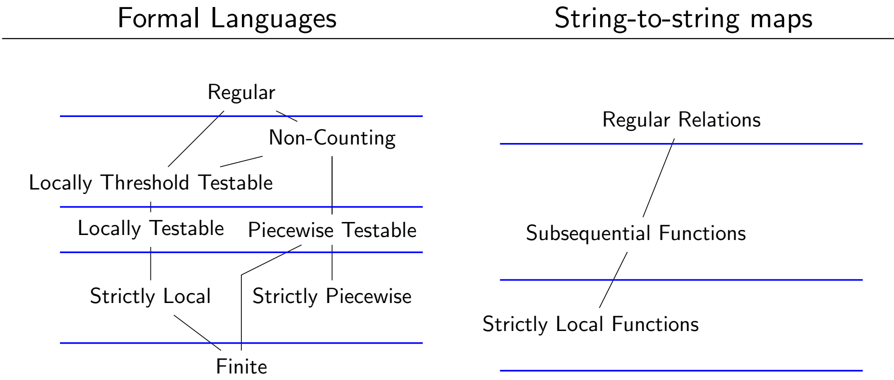
- SL function that targets substrings of length 3?

taʈ ↦ tat	taaʈ ↦ taaʈ
daɖ ↦ dad	danaɖ ↦ danaɖ

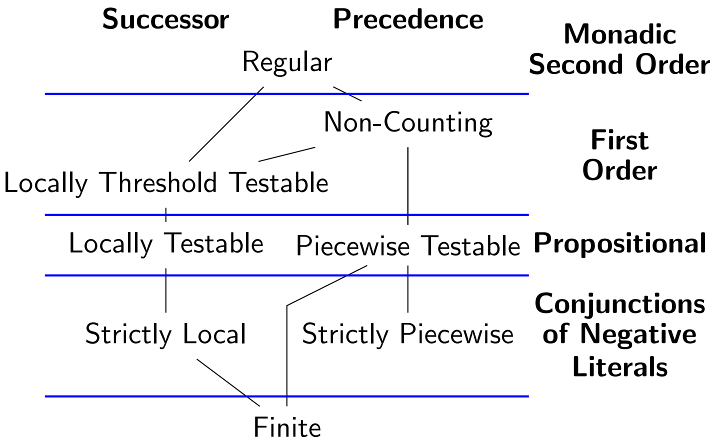
- Or SP function that targets subsequences of length 2 that will never 'see' inputs longer than 3?

taʈ ↦ tat	taat ↦ taat
daɖ ↦ dad	danad ↦ danad

Subregular hierarchies

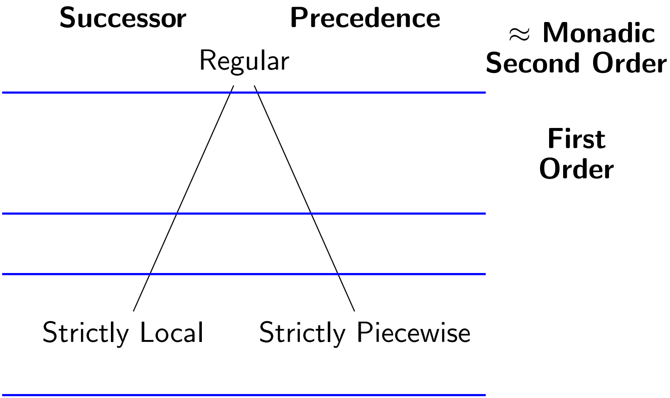


Subregular hierarchy of languages



(Rogers & Pullum 2011, Rogers et al. 2013)

Subregular hierarchy of functions



First Order Logic

- $\forall$, $\exists$
- $\wedge$, $\vee$, $\neg$
- $\rightarrow$, $\leftrightarrow$
- x, y, z, ...
- $P(x)$, $R(x, y)$, ...

'Sub'-First Order Logic

Proposal: SL functions are quantifier-free, unary, monotone graph transductions (Chandlee & Lindell, in prep)

- ~~$\forall, \exists$~~
- $\wedge$, $\vee$, $\neg$
- $\rightarrow$, $\leftrightarrow$
- x, y, z, ...
- $P(x)$, ~~$R(x, y)$~~, ...

Next steps

- Extend to dissimilation maps
- Logical and FST characterizations of SP functions
- Interactions of SL and SP maps?

Conclusions

- Characterizing computational complexity classes from a variety of perspectives leads to a more complete understanding of what the class is.
- Identifying the right class (or combination of classes) for phonological maps likewise leads to insights into what phonology is (and isn't).
- From this perspective, the distinction between local and long-distance phonology is one of representation, not computational complexity.

Acknowledgments

For ongoing collaboration and inspiration, thanks to:

- Jeffrey Heinz (University of Delaware)
- Adam Jardine (University of Delaware)
- Rémi Eyraud (Aix-Marseilles University)
- Jim Rogers (Earlham College)
- Steven Lindell (Haverford College)

References

- Chandlee, J., A. Athanasopoulou, and J. Heinz (2012). Evidence for classifying metathesis patterns as subsequential. In The Proceedings of the 29th West Coast Conference on Formal Linguistics, Somerville, MA, pp. 303-309. Cascadilla.
- Chandlee, J. and J. Heinz (2012). Bounded copying is subsequential: implications for metathesis and reduplication. In Proceedings of SIGMORPHON 12.
- Chandlee, J. (2014). *Strictly Local Phonological Processes.* Ph.D. thesis, University of Delaware.
- Chandlee, J., R. Eyraud, and J. Heinz. (2015). Output Strictly Local Functions. *Mathematics of Language 2015*.
- Engelfriet, J. and H. J. Hoogeboom. (2001). MSO definable string transductions and two-way finite state transducers. *ACM Transactions on Computational Logic, 1*(4), pg. 1-38.
- Gainor, B., R. Lai, and J. Heinz (2012). Computational characterizations of vowel harmony patterns and pathologies. In The Proceedings of the 29th West Coast Conference on Formal Linguistics, Somerville, MA, pp. 63-71. Cascadilla.
- Hansson, G. (2001). *Theoretical and Typological Issues in Consonant Harmony.* Ph. D. thesis, University of California, Berkeley.
- Hayward, R.J. (1988). Remarks on the Omotic sibilants. *Cushitic-Omotic: Papers from the International Symposium on Cushitic and Omotic languages*, Cologne, January 6-9, 1986, ed. by M. Bechhaus-Gerst and F. Serzisko, pp. 263-299. Hamburg: Helmut Buske.
- Heinz, J. (2007). *The Inductive Learning of Phonotactic Patterns.* Ph. D. thesis, University of California, Los Angeles.
- Heinz, J. (2009). On the role of locality in learning stress patterns. *Phonology* (26), 303-351.
- Heinz, J. (2010). Learning Long-Distance Phonotactics. Linguistic Inquiry, 41(4):623-661.
- Heinz, J., C. Rawal, and H. G. Tanner. (2011). Tier-based Strictly Local Constraints for Phonology. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 58-64, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.
- Heinz, J. and R. Lai (2013). Vowel harmony and subsequentiality. Mathematics of Language.
- Jardine, A. (to appear). Computationally, Tone is different. *Phonology*.

References

- 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.
- Lee, S. and J. Pater. (2008). Phonological inference and word recognition: Evidence from Korean. Ms, Korea University and University of Massachusetts, Amherst.
- Luo, H. (2013). Long-distance consonant harmony and subsequentiality. Unpublished manuscript, University of Delaware.
- Meinhof, C. (1932). *Introduction to the phonology of the Bantu languages*. Trans. by N. J. van Warmelo. Berlin: Dietrich Reimer/Ernst Vohsen.
- Mohri, M. (1997). Finite-state transducers in language and speech processing. Computational Linguistics (23), 269-311.
- Odden, D. (1994). Adjacency parameters in phonology. *Language 70*, pg. 289-330.
- Payne, A. (2013). Restricting phonology: Dissimilation as a subsequential process. NELS 44.
- Rogers, J., J. Heinz, M. Fero, J. Hurst, D. Lambert, and S. Wibel (2012). Cognitive and sub-regular complexity. In 17th Conference on Formal Grammars.
- Rogers, J. and G. Pullum (2011). Aural pattern recognition experiments and the subregular hierarchy. Journal of Logic, Language and Information (20), 329-342.
- Rose, S. and R. Walker. (2004). A typology of consonant agreement as correspondence. *Language 80*(3), pg. 475-531.
- Shaw, P. A. (1991). Consonant harmony systems: the special status of coronal harmony. *The special status of coronals: internal and external evidence*, ed. by C. Paradis and J-F. Prunet, pp. 125-158. San Diego, CA: Academic Press.
- Stirnimann, H. (1983). Praktische Grammatik der Pangwa-Sprache (SW-Tansania). (Studia Ethnographica Friburgensia, 10.) Freiburg: Universitätsverlag Freiburg Schweiz.
- Viljoen, J. J. (1973). *Manual for Ndonga*. Part 1. Pretoria: University of South Africa.

SP FSTs?

(11) Benchnon sibilant harmony (Hayward 1988, Hansson 2001)

a. /sʲap-s/ ↦ [sʲaps] ‘make wet’

b. /ʃir-s/ ↦ [ʃirʃ] ‘bring near’