Representing and learning non-linear sound patterns in natural language

Adam Jardine, Department of Linguistics & Cognitive Science

ajardine@udel.edu

Big Questions

- Sounds and word order patterns in natural language are not arbitrary
- Why do languages have some kinds of patterns and not others?
- How does this relate to how children acquire language?

Takeaway

- Many sound patterns are describable as illegal substrings (Rogers et al., 2013)
- This provides a strong, computational theory with insights into learning (García et al., 1990; Heinz, 2010)
- However, some sound patterns, especially regarding tone (pitch), require non-linear representations (Goldsmith, 1976)
- We can extend the 'illegal substructure' idea to these representations using *subgraphs*

A Tone Pattern

Digo (Kenya & Tanzania; Kisseberth, 1984):

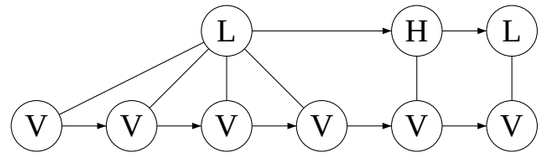
	'I am...'	'He/she is...'
'praising'	ni-na-togor-a	a-na-togór-a
'waking up'	ni-na-ramuk-a	a-na-ramúk-a
'pressing'	ni-na-gandamiz-a	a-na-gandamíz-a
'adding to'	ni-na-onjerez-a	a-na-onjeréz-a

Analysis: A high tone (H) 'shifts' to the end of the word

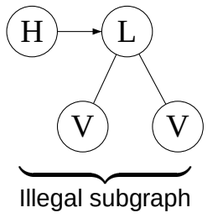
H		H	
a-na-gandamiza	→	a-na-gandamiz-a	'he is pressing'

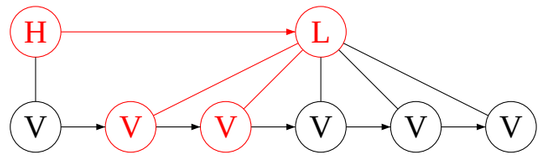
Representing Tone as Graphs

- We can view this non-linear relationship between tones and vowels as a labeled graph
- To the right is graph representation of *a-na-gandamíz-a* 'he/she is pressing'



- Valid graphs in the Digo pattern can be said to exclude this subgraph
- Thus, **á-na-gandamiz-a*, where the tone matches up with its prefix vowel, is excluded





Learning Graph Patterns

- 'Banned substructure' patterns allow a simple method for learning from positive data: remember substructures of a certain size for each data point (Heinz, 2010)
- Subgraphs up to *size* 3 of *diameter* 1 of *a-na-gandamíz-a* graph:

size: number of edges (lines)
diameter: longest minimum path between nodes (circles)

Acknowledgements & References

Acknowledgements

This work is indebted to Prof. Jeffrey Heinz and the students in the computational linguistics and phonology study groups at the UD Department of Linguistics & Cognitive Science.

References

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.

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

Kisseberth, C. W. (1984). Digo tonology. In Clements, G. and Goldsmith, J. A., editors, *Autosegmental Studies in Bantu Tone*, pages 105–182. Foris Publications.

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.