Learning Phonological Mappings by Learning Strictly Local Functions

Jane Chandlee and Adam Jardine

University of Delaware

Phonology 2013 – UMass

Main objectives

- Identify locality as a near-universal, defining property of many phonological input-output mappings.
- Present the SLFLA, a learning algorithm that uses locality as an inductive principle to generalize these phonological mappings from finite data.
- Discuss simulations and theoretical results.

Phonological mappings

- Final devoicing: (ba:d, ba:t)
- -son $\Rightarrow$ -voice / __#
- *[-son, +voice]# >> IDENT(voice)

Phonological mappings

- (CAD, CBD)
- A $\Rightarrow$ B / C __ D
- *CAD >> FAITH(A$\Rightarrow$B) (Baković 2013)
- Locality as a property of the *mapping*.
- Tesar (to appear): phonological maps are output-driven.

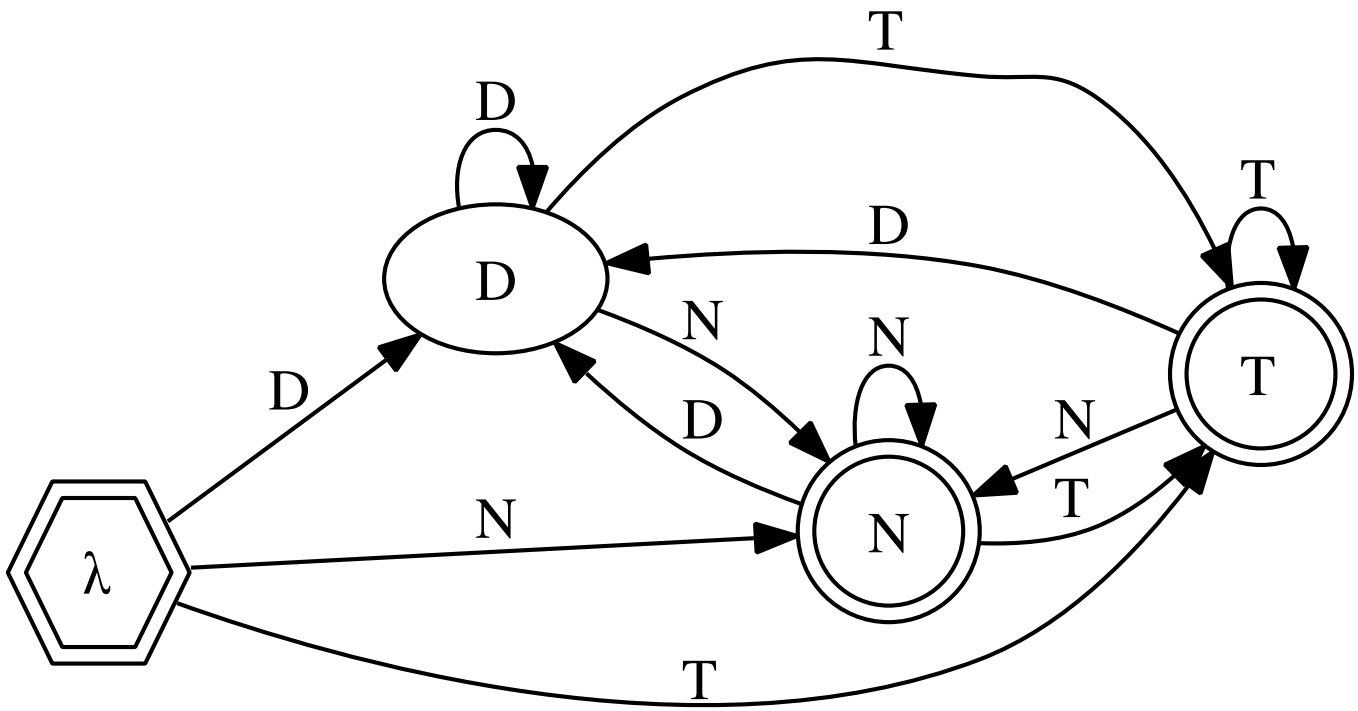
Strictly Local Languages

- Defined with grammars of k-length substrings (i.e., k-factors)

(1) Words can't end with a voiced obstruent: SL-2 grammar that includes {N#, T#} but excludes D#.

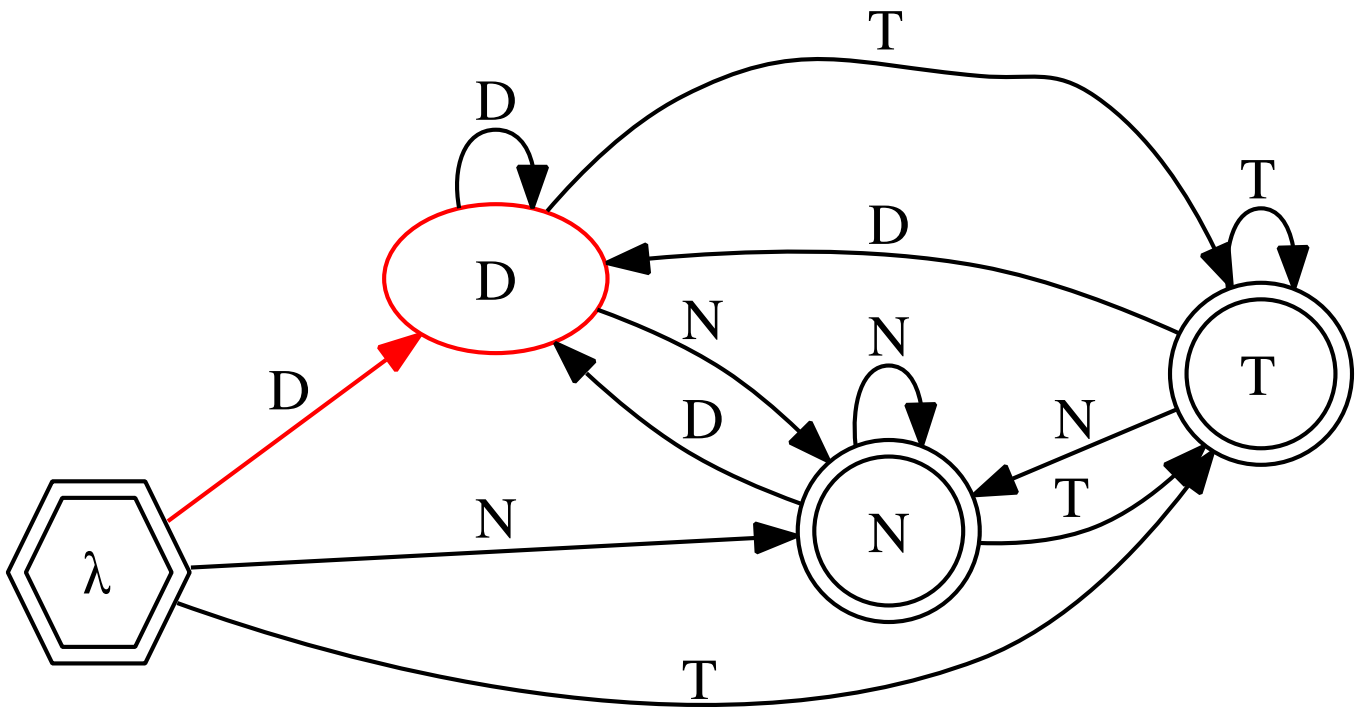
McNaughton & Papert (1971), Rogers & Pullum (2011), Rogers et al. (2012)

Strictly Local Languages



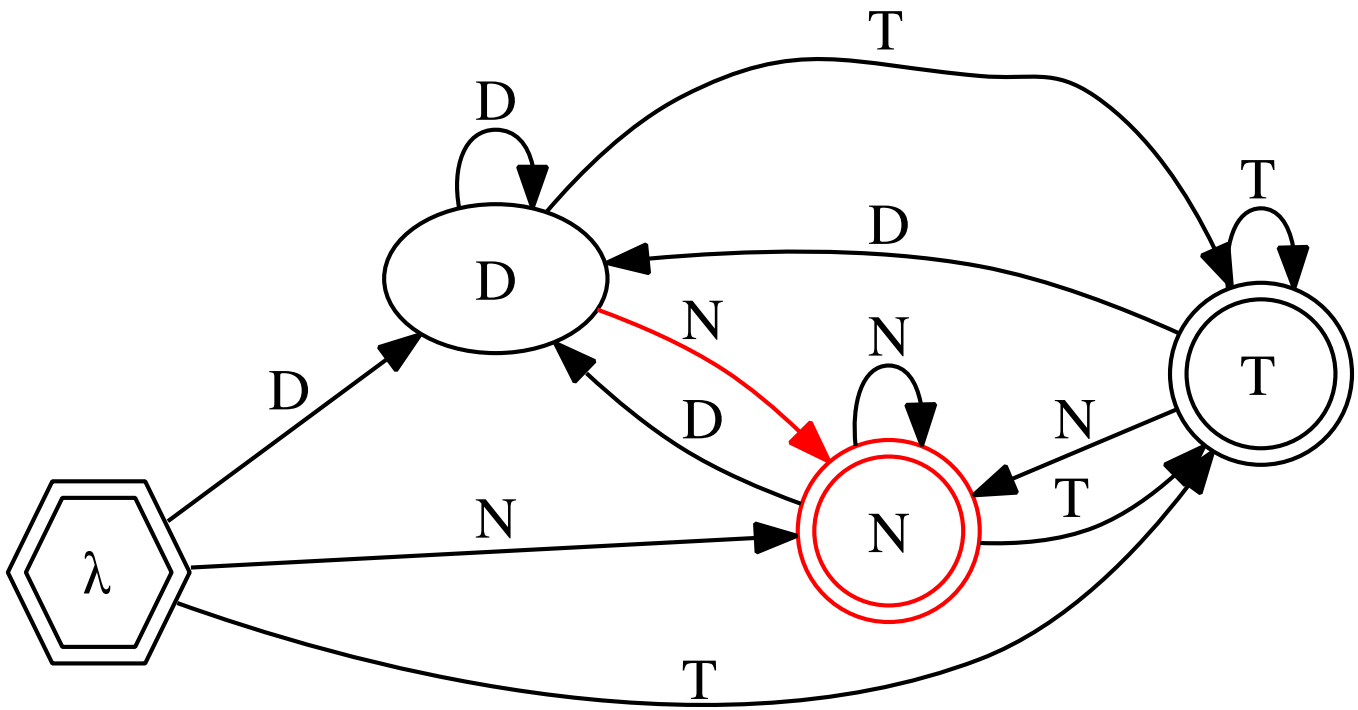
Strictly Local Languages

DNT



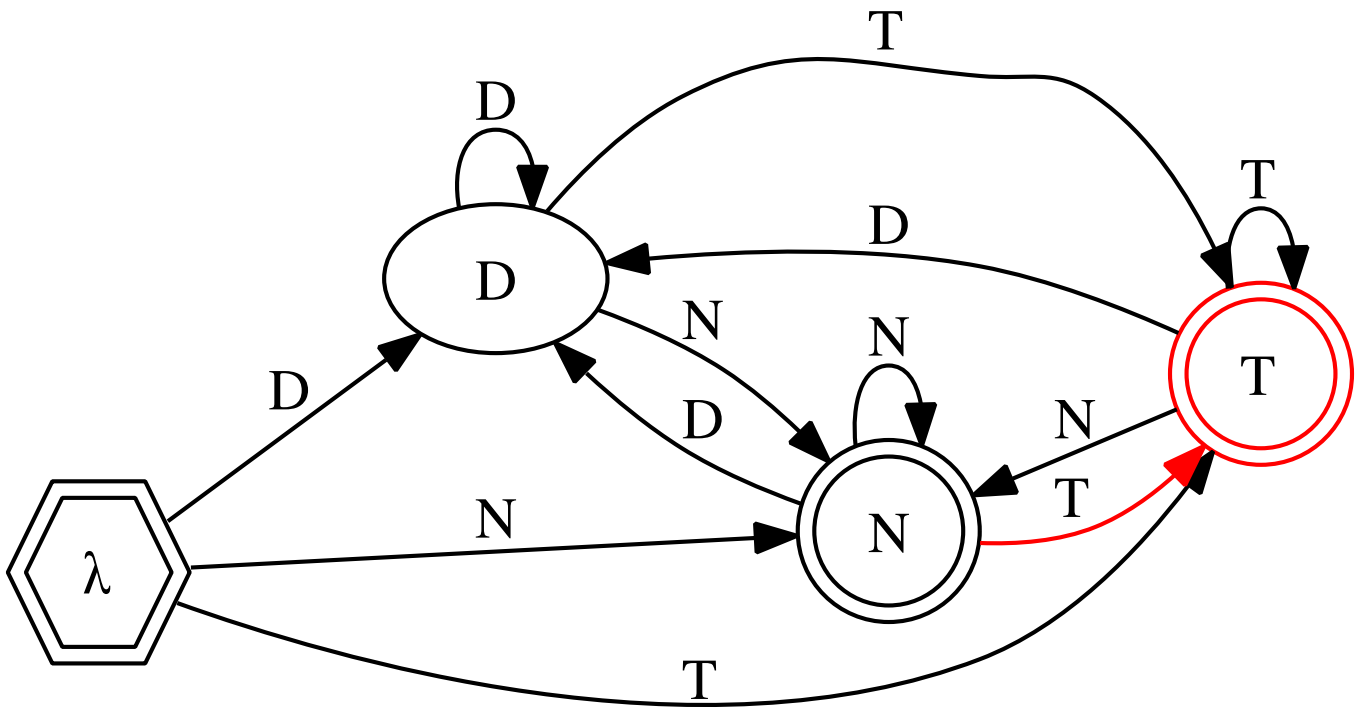
Strictly Local Languages

DNT



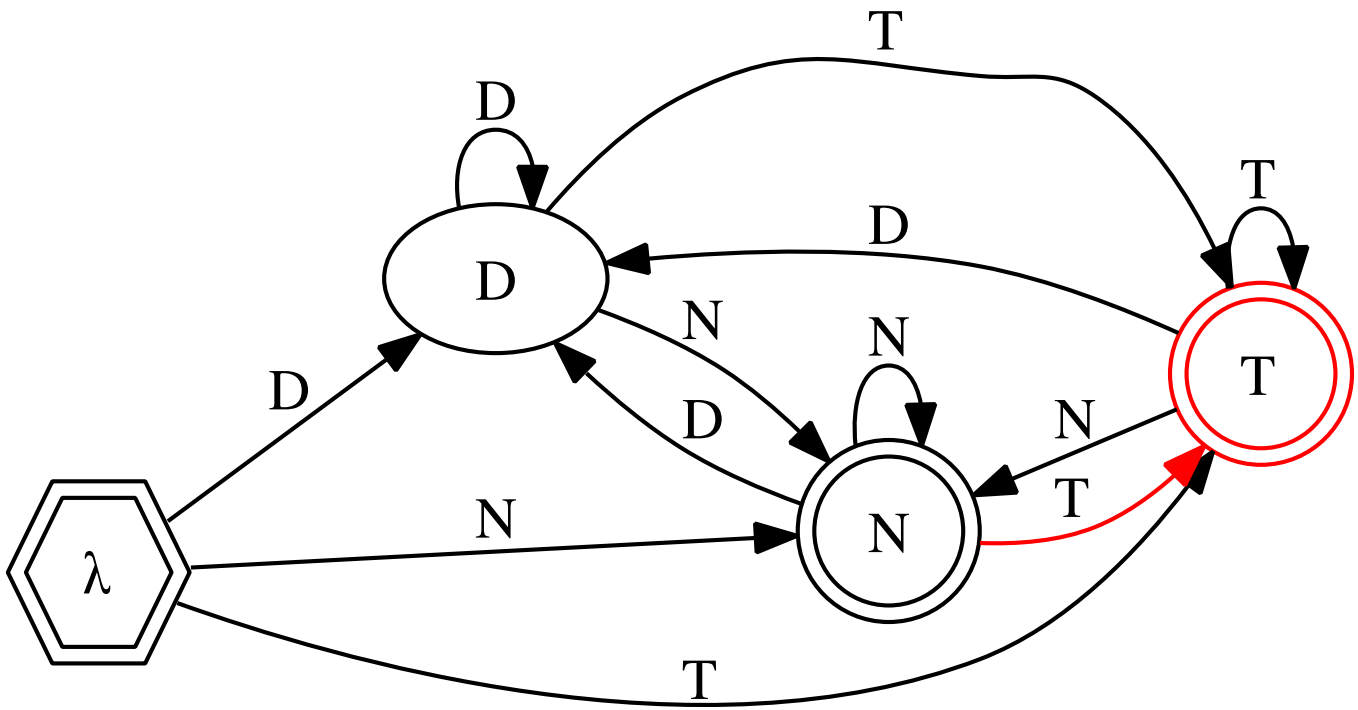
Strictly Local Languages

DNT



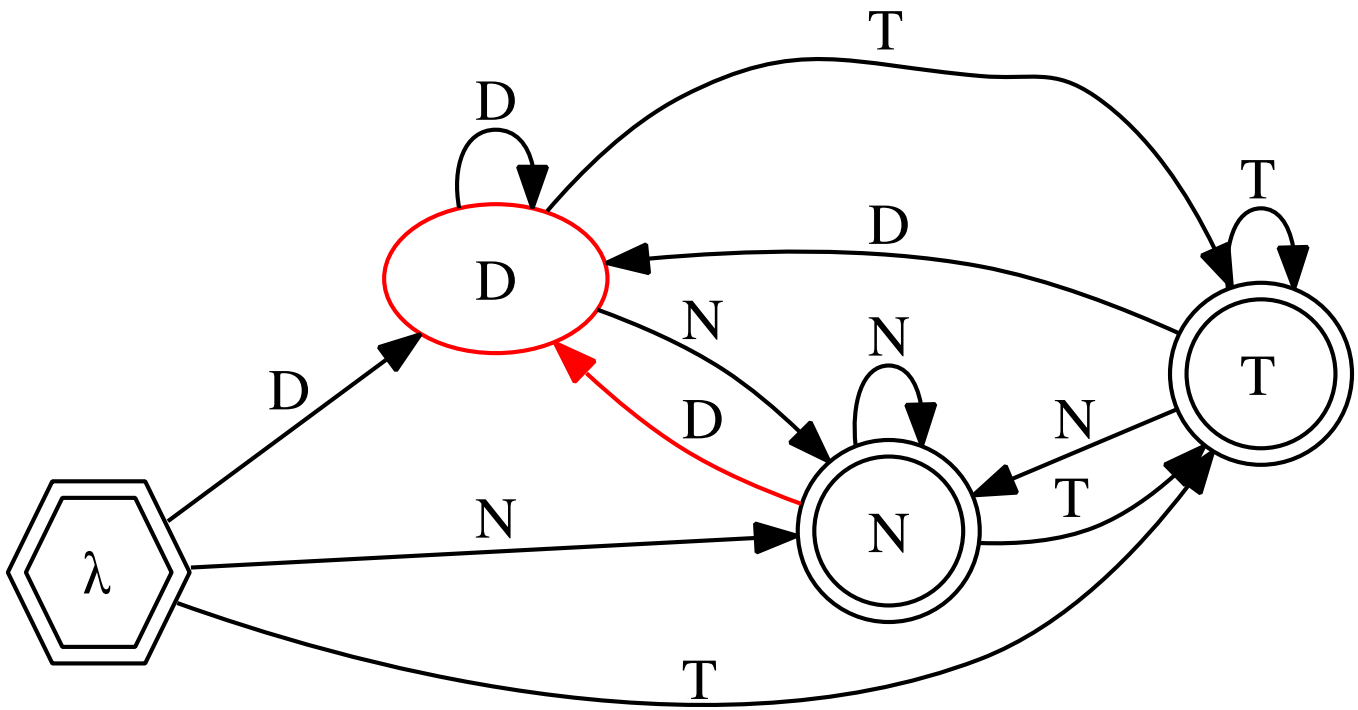
Strictly Local Languages

✓ DNT



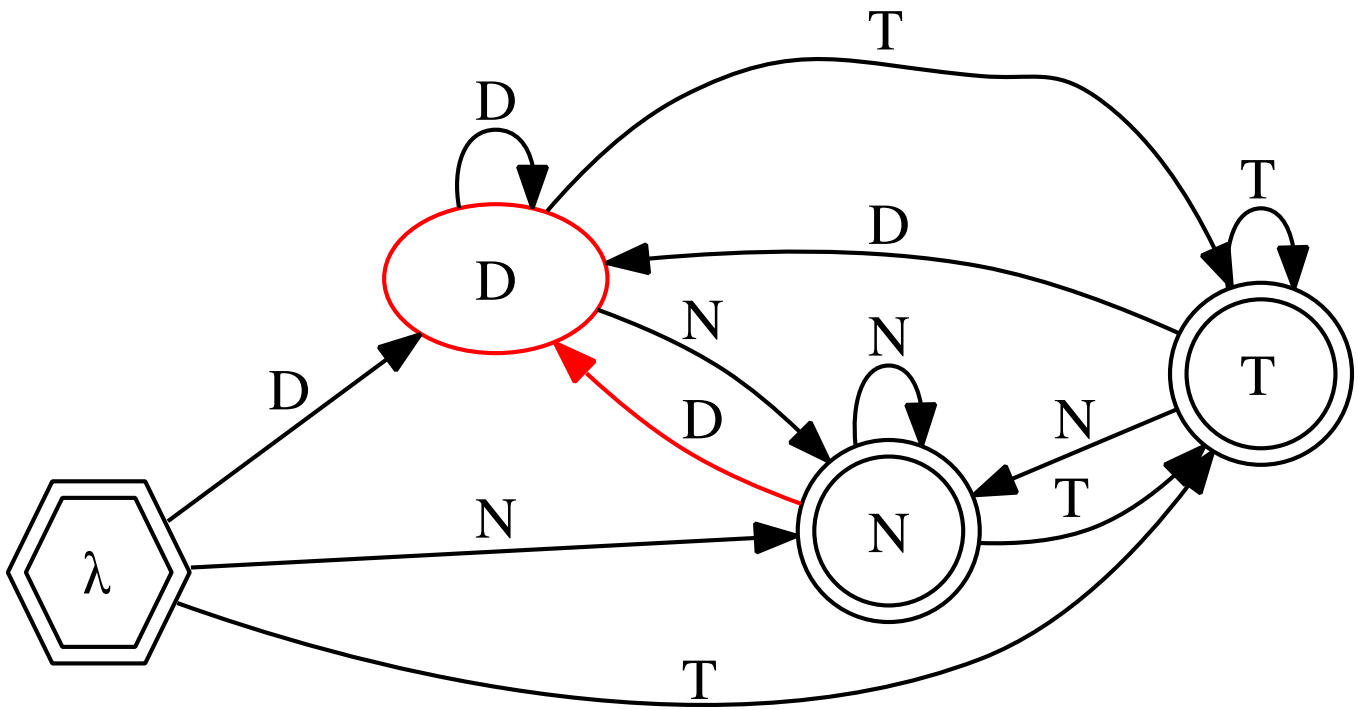
Strictly Local Languages

DND



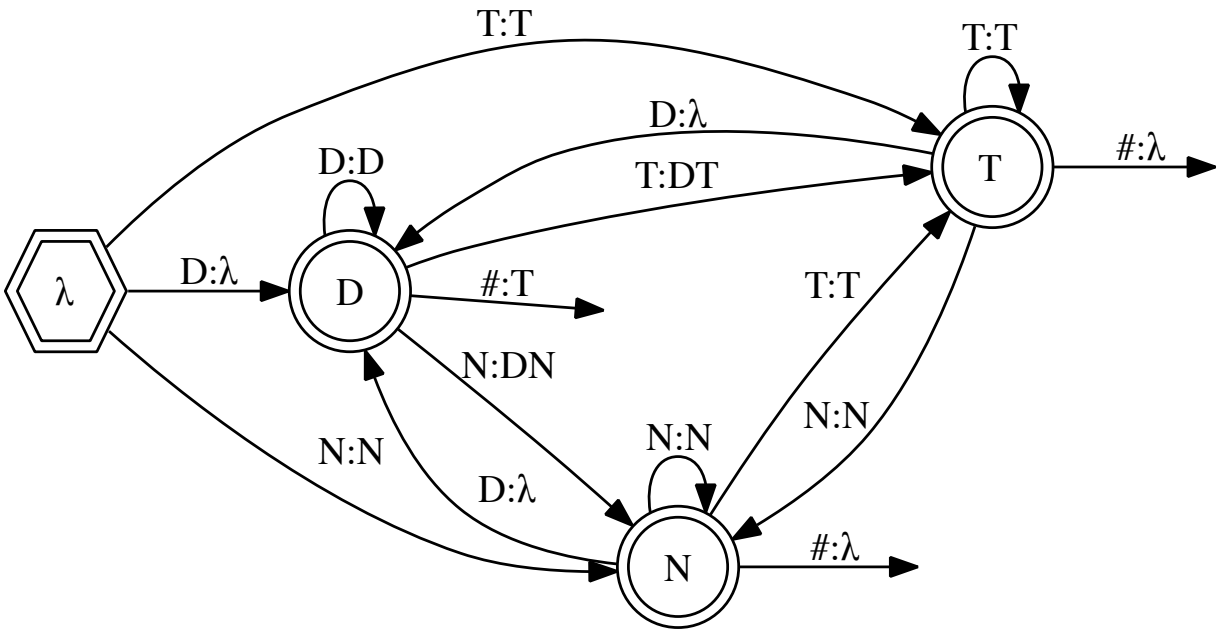
Strictly Local Languages

*DND

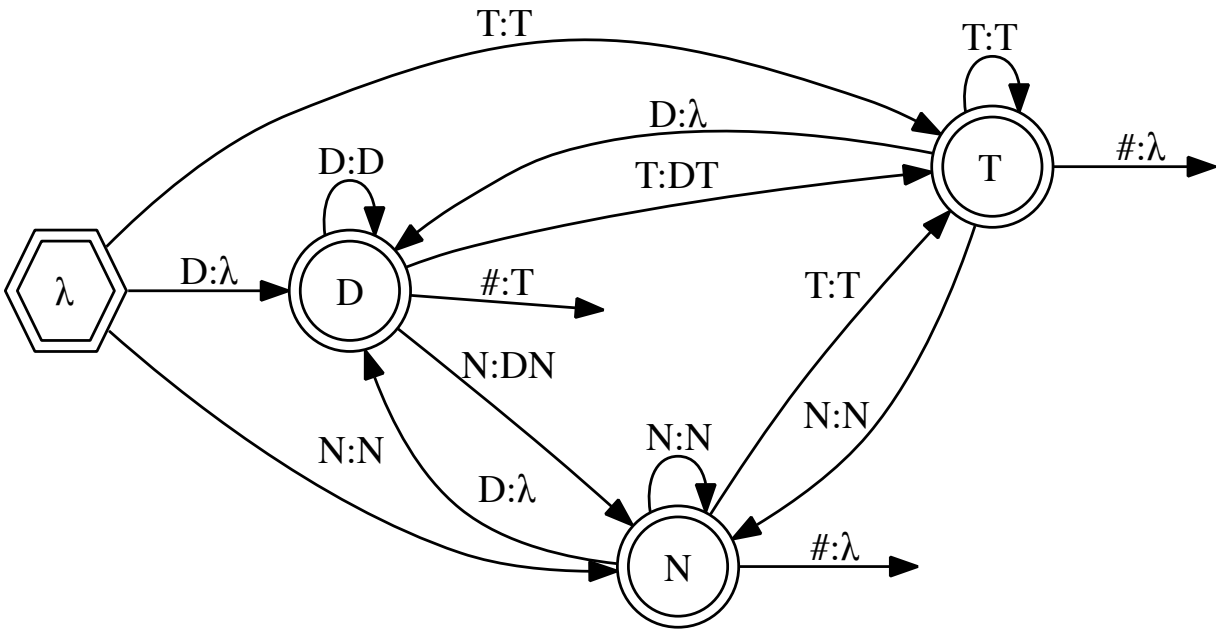


Strictly Local Functions

(2) $D \Rightarrow T \; / \; \_ \; \#$



Strictly Local Functions

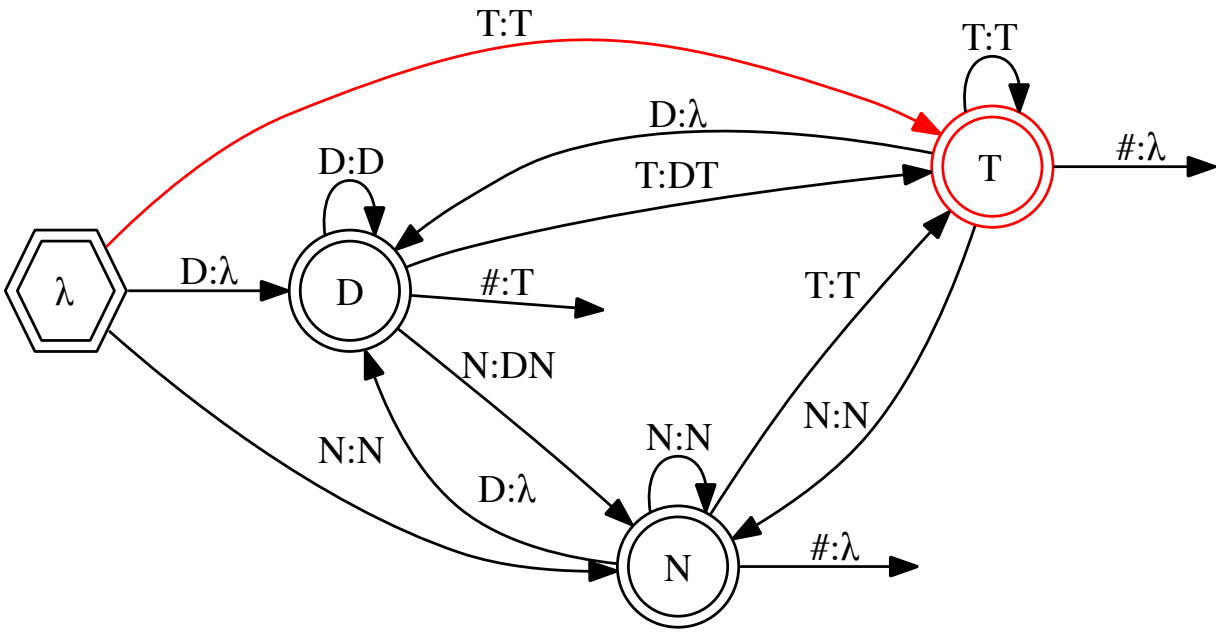


Input: T N D

State: λ

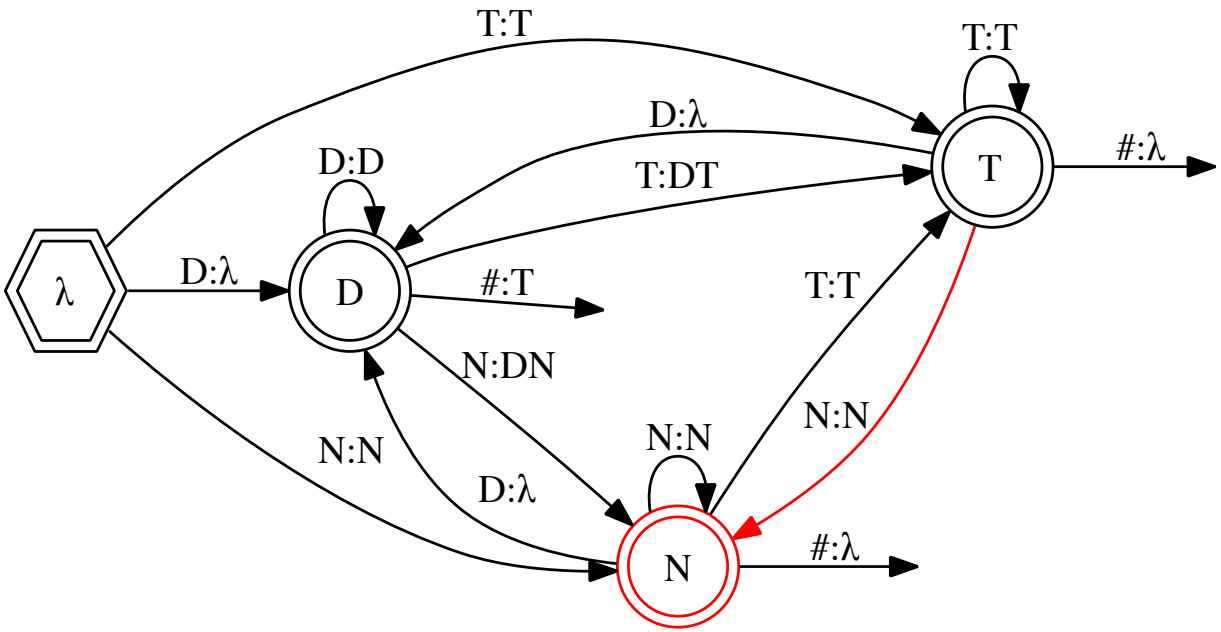
Output:

Strictly Local Functions



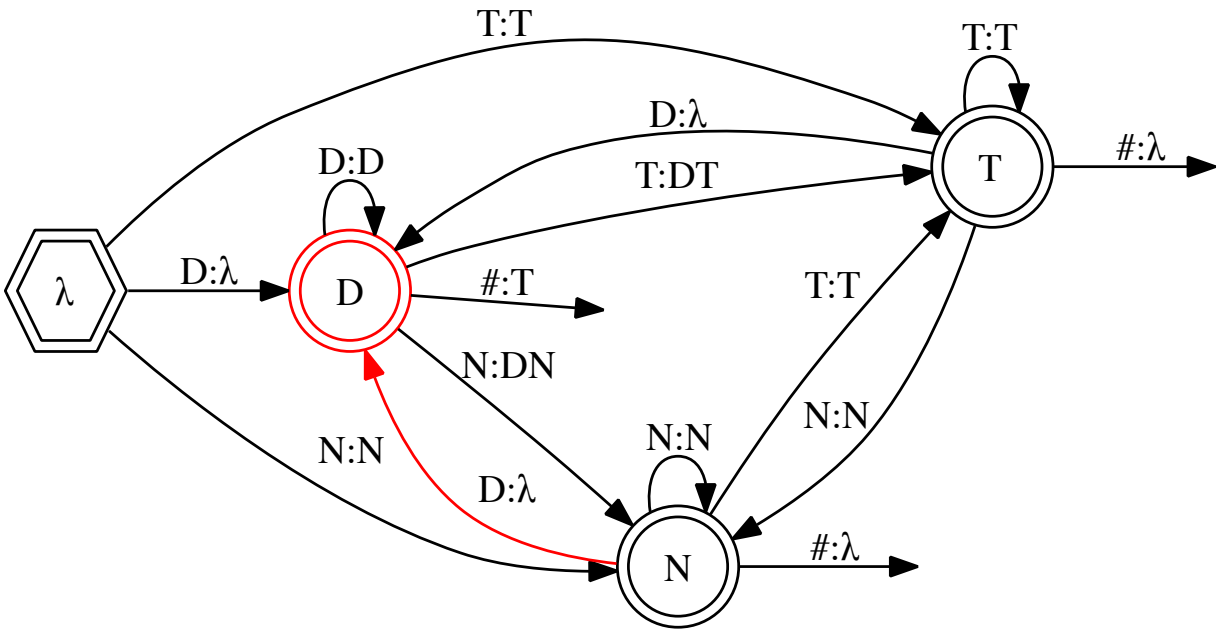
Input: T N D
State: $\lambda \Rightarrow$ T
Output: T

Strictly Local Functions



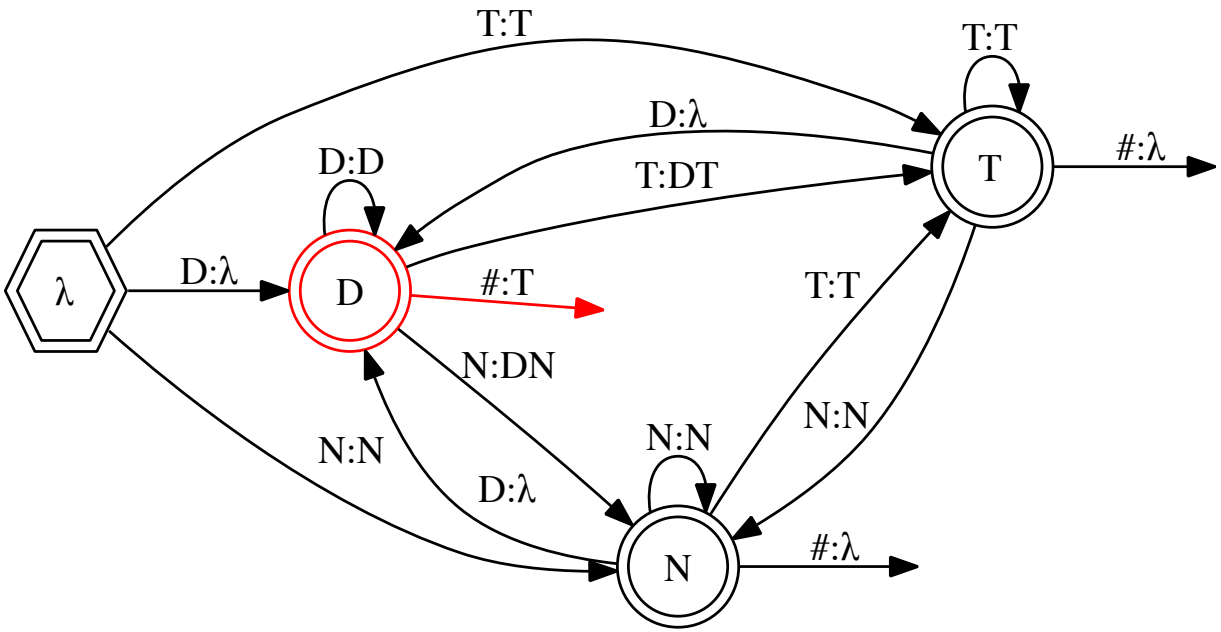
Input:		T	N	D
State:	λ	⇒ T	⇒ N	
Output:		T	N	

Strictly Local Functions



Input:		T	N	D
State:	λ	$\Rightarrow$ T	$\Rightarrow$ N	$\Rightarrow$ D
Output:		T	N	λ

Strictly Local Functions



Input:		T	N	D	
State:	λ	$\Rightarrow$ T	$\Rightarrow$ N	$\Rightarrow$ D	
Output:		T	N	λ	T

Empirical Coverage

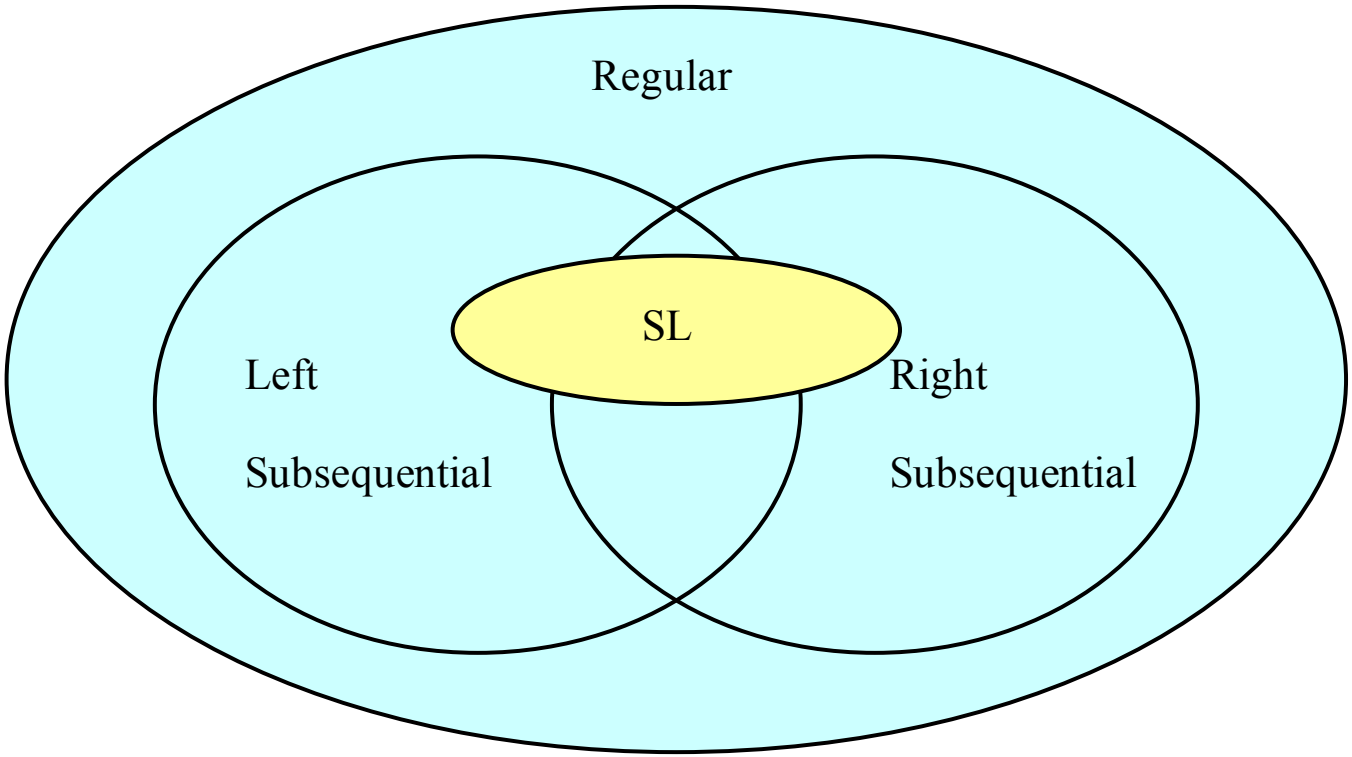
(3) $x_i \Rightarrow y_i \ / \ \mathrm{U} \_\_ \mathrm{V}$

- At least 96% of the approx. 5500 processes in P-Base (v1.95, Mielke 2008) are Strictly Local.
  - Substitution, deletion, epenthesis, general affixation (Chandlee 2014, Chandlee & Heinz (in prep.))
  - Bounded metathesis, local partial reduplication (Chandlee et al. 2012, Chandlee & Heinz 2012)

Non-SL

- Vowel harmony with transparent vowels (Gainor et al. 2012, Heinz & Lai 2013)
- Long-distance consonant harmony (Luo 2013)
- Long-distance dissimilation (Payne 2013)
- Certain tonal patterns (Jardine 2013)
- Unbounded displacement/metathesis (Chandlee & Heinz 2012)
- Nonlocal partial reduplication (Riggle 2003)

Learning SL



OSTIA (Oncina et al. 1993)

- **O**nward **S**ubsequential **T**ransducer **I**nduction **A**lgorithm
- Identifies the class of subsequential functions in the limit from positive data.

Gildea & Jurafsky (1996)

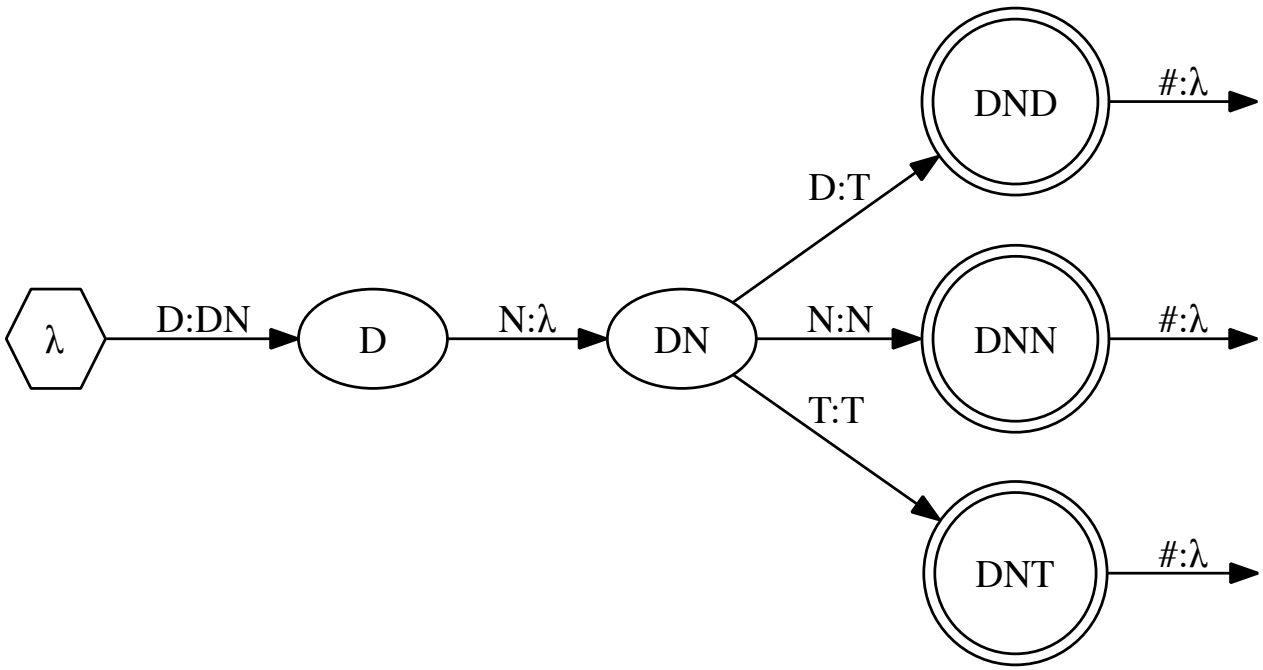
- OSTIA failed to learn the English flapping rule using data from a natural language corpus.
- Modified OSTIA with three learning biases:
  1. Faithfulness: underlying segments realized similarly on the surface
  2. Community: similar segments behave similarly
  3. Context: rules access contextual variables

Gildea & Jurafsky (1996)

- OSTIA failed to learn the English flapping rule using data from a natural language corpus.
- Modified OSTIA with three learning biases:
  1. Faithfulness: underlying segments realized similarly on the surface
  2. Community: similar segments behave similarly
  3. Context: rules access contextual variables

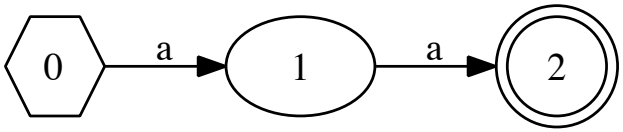
Onward prefix tree transducer

{(DND, DNT), (DNT, DNT), (DNN, DNN)}

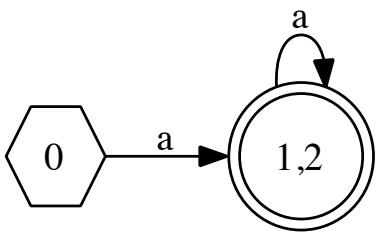


State merging

{ aa }



aa*



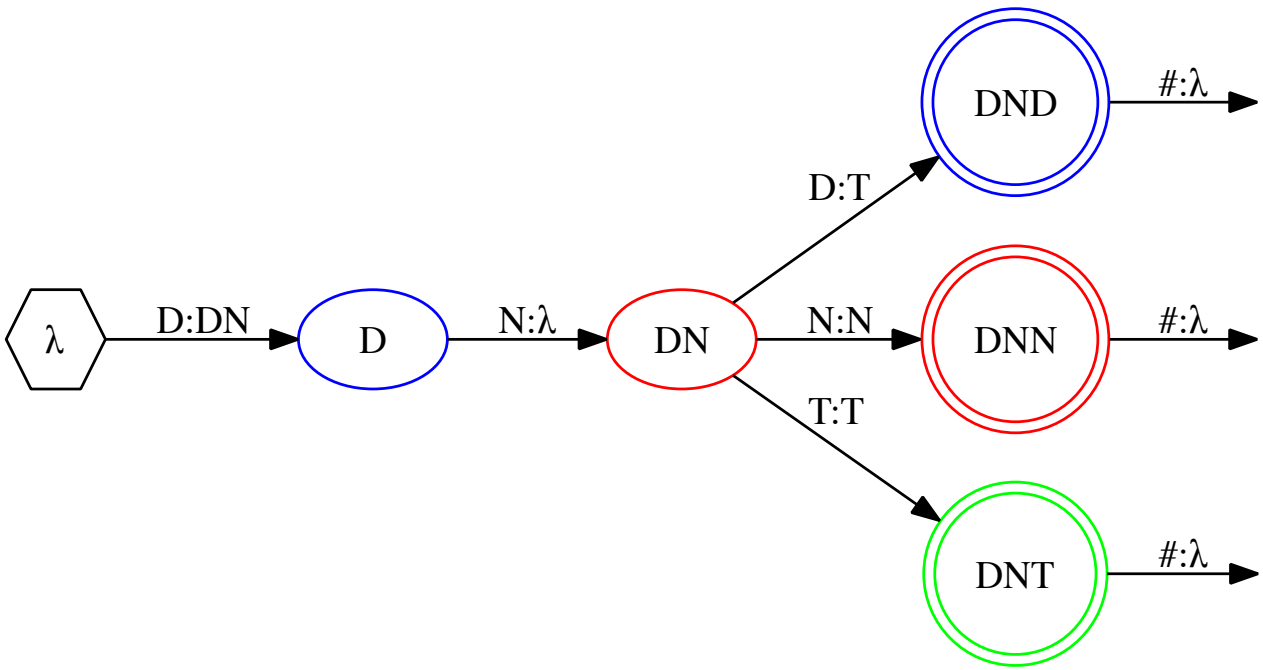
State merging

- OSTIA: merges any two states as long as the result is still a subsequential transducer (allowing for additional merges to remove resulting non-determinism)
- SLFLA: maintains subsequentiality *and* enforces strict locality

State merging

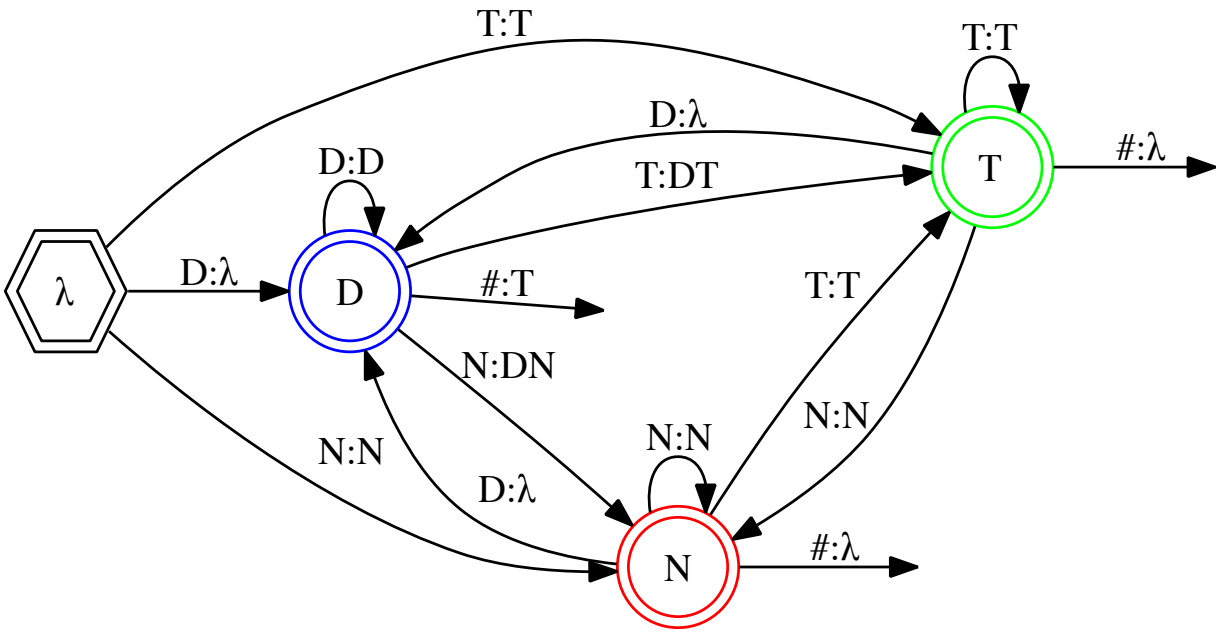
- Merge states in the prefix tree that were reached with the same k-1 input symbols.
- This strategy was previously applied to the learning of SL languages (Heinz 2007) and SL functions (Chandlee & Koirala 2014).

State merging



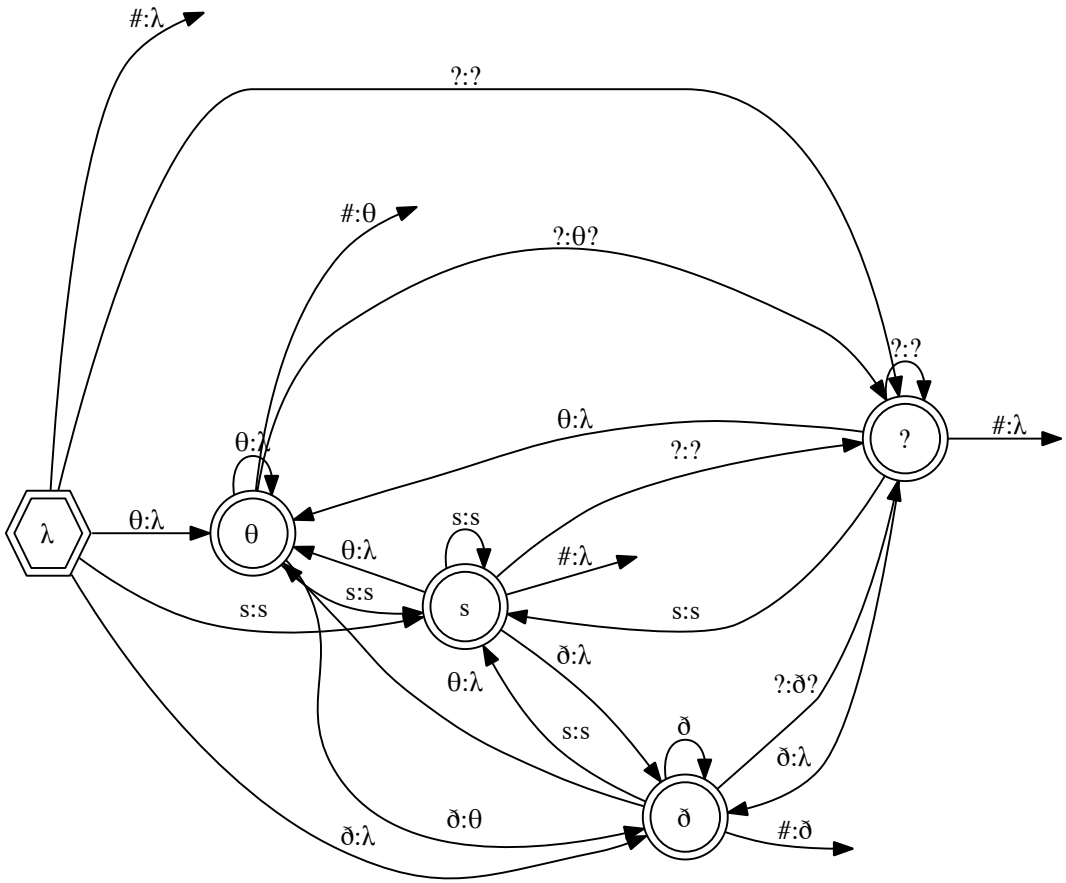
Simultations: Final devoicing

- D $\Rightarrow$ T / $\_\_$ # ($k = 2$)
- Data = all possible $w \in \Sigma^{\leq 5}$ paired with correct surface form according to the rule (363 pairs)



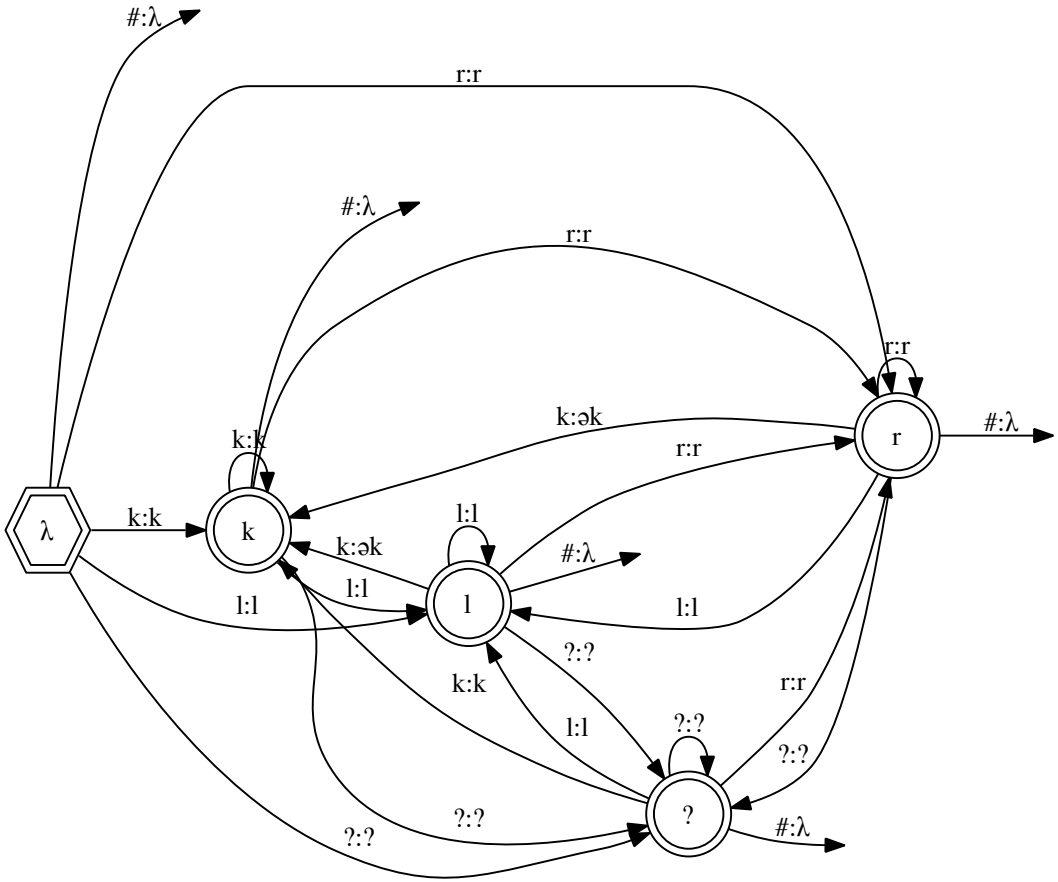
Simulations: Greek fricative deletion

- $\{\theta, \eth\} \Rightarrow \lambda \; / \; \_\_ \; \{s, \theta\}$ $(k = 2)$
- $\Sigma = \{\theta, \eth, s, ?\}$
- 1364 pairs



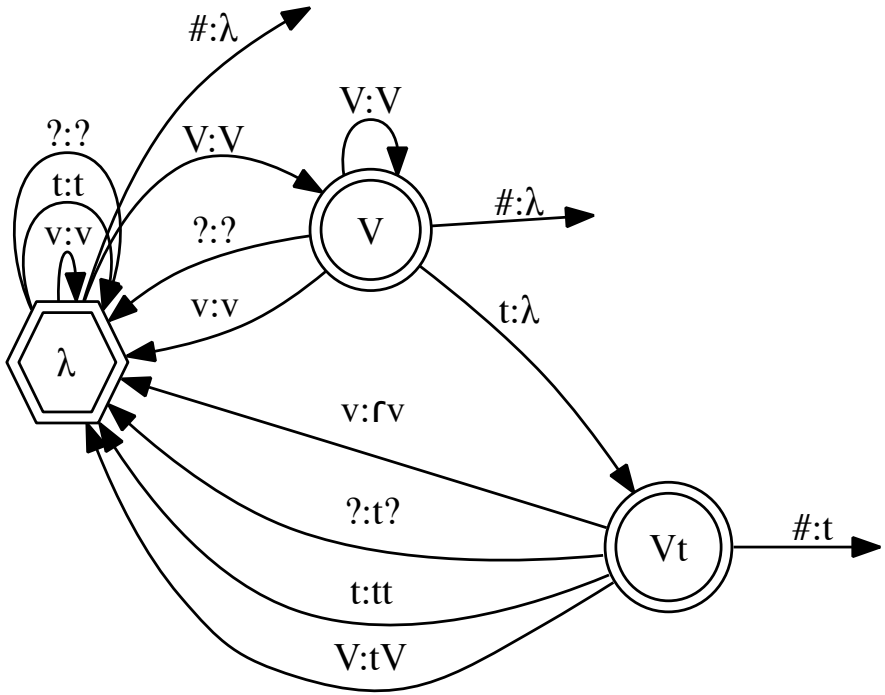
Simulations: Dutch schwa epenthesis

- $\lambda \Rightarrow$ ə / {l, r} __ [-coronal] ($k = 3$)
- $\Sigma = \{$l, r, k, ?$\}$
- 1364 pairs



Simulations: English flapping

- t ⇒ ɾ / V́ __ V ($k = 3$)
- $\Sigma = \{V, v, t, ?\}$
- 1364 pairs



Theoretical result

Theorem

The class of SL functions is learnable from a closed learning sample.

Discussion

- What happens with natural language data?
  - Problem: Data does not provide enough information to generalize a total function.
- Gildea & Jurafsky (1996) address this with the *faithfulness* and *community* biases.
- Relax assumption of total function while maintaining onwardness?

Conclusions

- The property of locality is evident in phonological mappings regardless of the grammatical formalism used to describe them.
- We have shown that an algorithm that makes use of this property can generalize such mappings from positive data.
- If humans also generalize in the way suggested here, this explains why local phonological processes are restricted in this way.

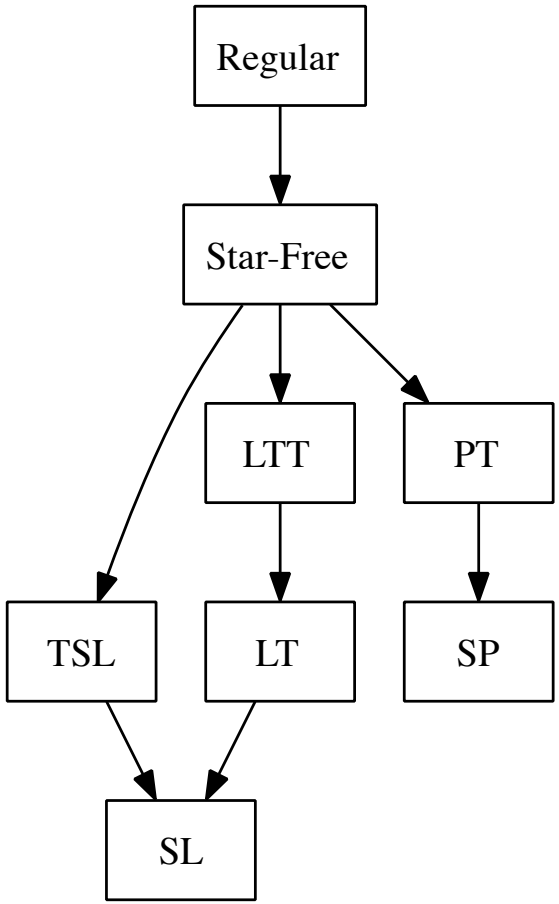
Next steps

- Current implementation will learn mappings for rules that apply *simultaneously*. Modify to learning mappings for *left-to-right* application (merge states with same previous $k-1$ *output* symbols).
- Identify functional counterparts to other subregular languages to account for non-local phenomena.

Acknowledgements

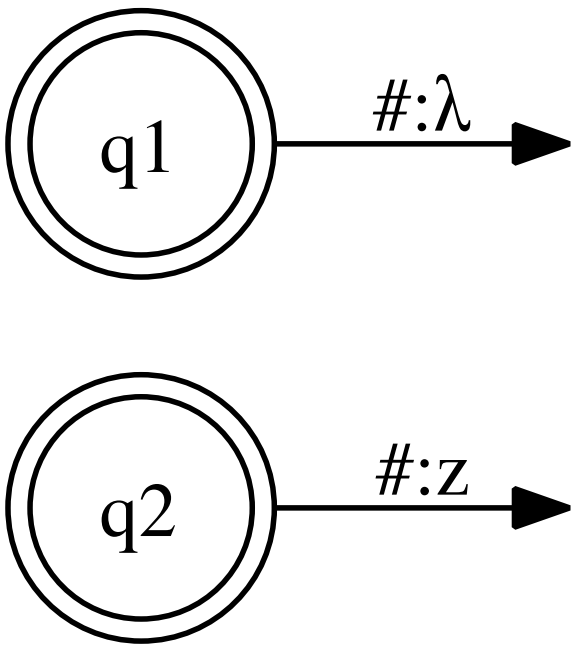
Thanks to Jeffrey Heinz, Rémi Eyraud, and Jim Rogers for valuable feedback and guidance.

Subregular Hierarchy



Closed Learning Sample

- OSTIA and the SLFLA will reject a merge between two states if the following situation results:



Closed Learning Sample

- The learning sample must include data points that prevent this situation: