Algebraic analyses of phonology

For the workshop on algebraic models of generative linguistics

Dakotah Lambert

09 Mar 2024

The question

- Phonotactics $\subseteq$ regular
  - Necessary?

The question

- Phonotactics $\subseteq$ regular
  - Necessary? (seems not)
  - FO[$<$]?

The question

- Phonotactics $\subseteq$ regular
  - Necessary? (seems not)
  - FO[$<$]? (better)
  - Lower?
- Only forbidding substrings/subsequences/etc?

The question

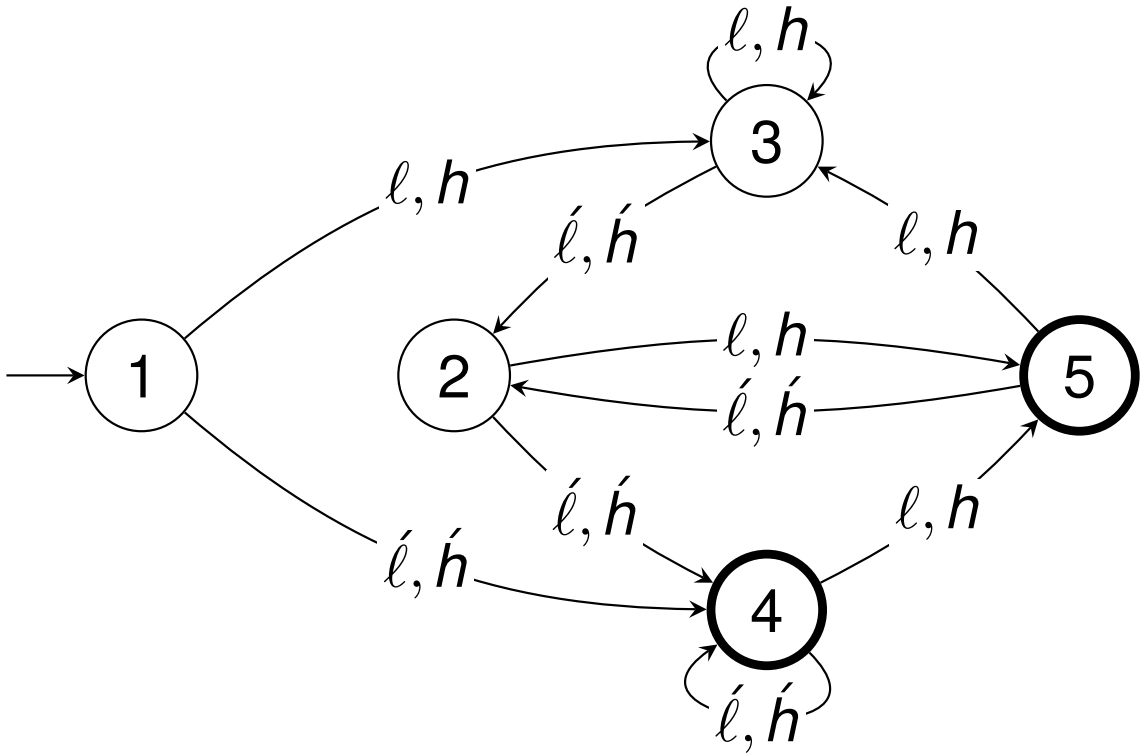
- Phonotactics $\subseteq$ regular
  - Necessary? (seems not)
  - FO[$<$]? (better)
  - Lower?
- Only forbidding substrings/subsequences/etc?
- No: Uyghur backness harmony, Karanga Shona tone, ...
- Propositional logic? Maybe!

Outline

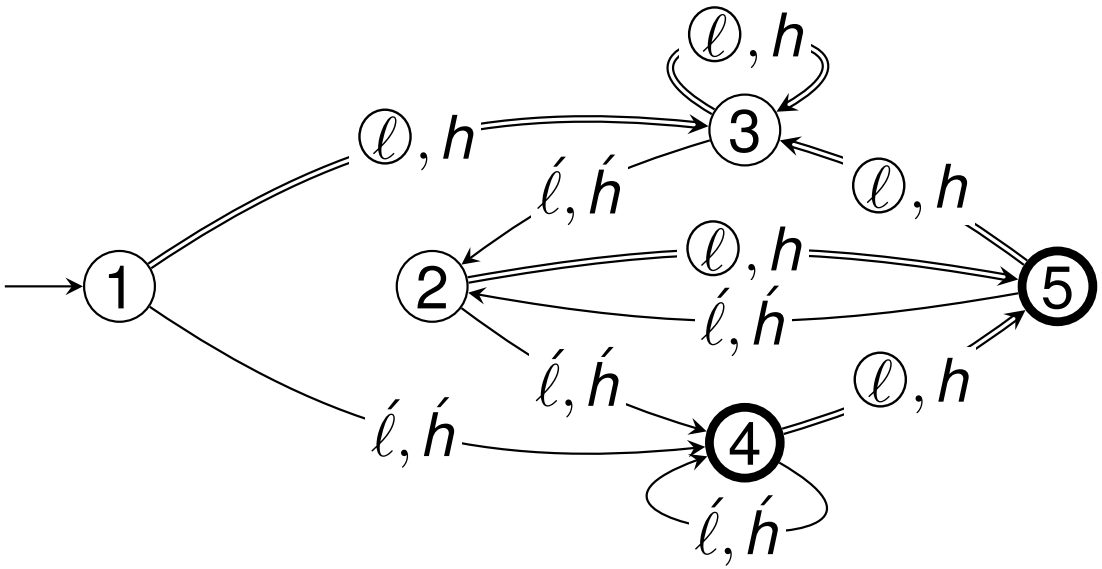
- Bounded stress: simple! Only reverse/definite?
- Varieties
- Long-distance constraints: Culminativity
- Unbounded stress
- Harmony
  - Symmetric
  - Asymmetric
  - Uyghur

Stress-penult

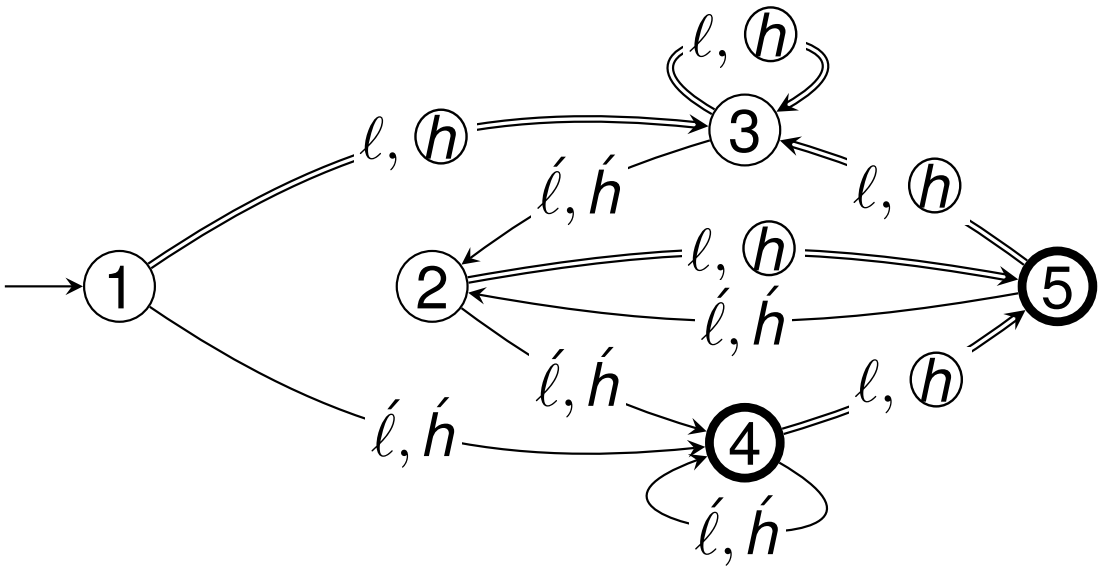
ℓℓ́h, ĺh, or ℓℓ́ĺ
but not ℓ or ĺhh



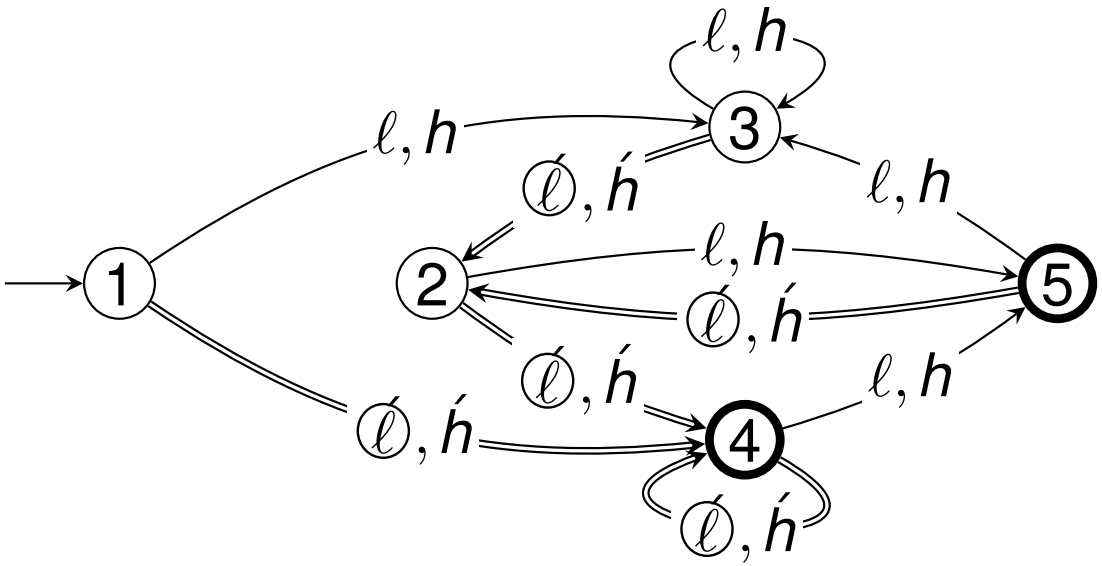
Syntactic semigroup: Actions



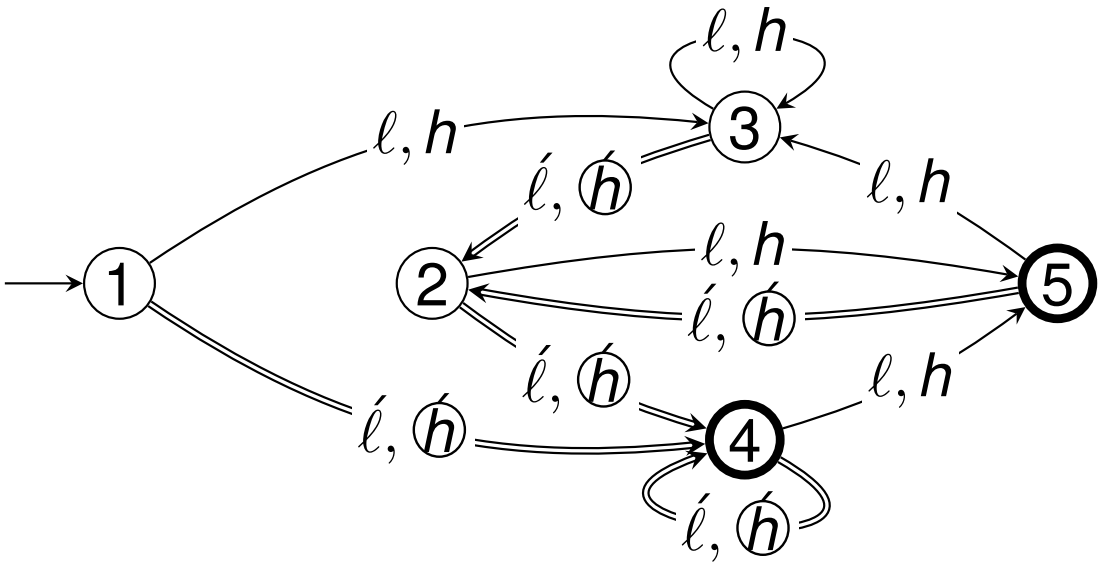
Syntactic semigroup: Actions



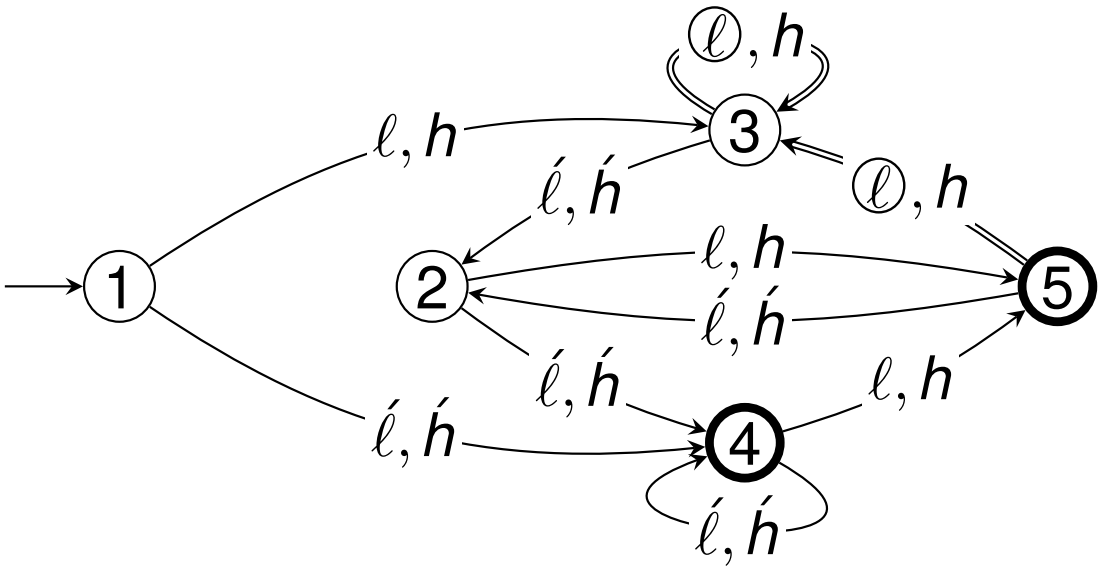
Syntactic semigroup: Actions



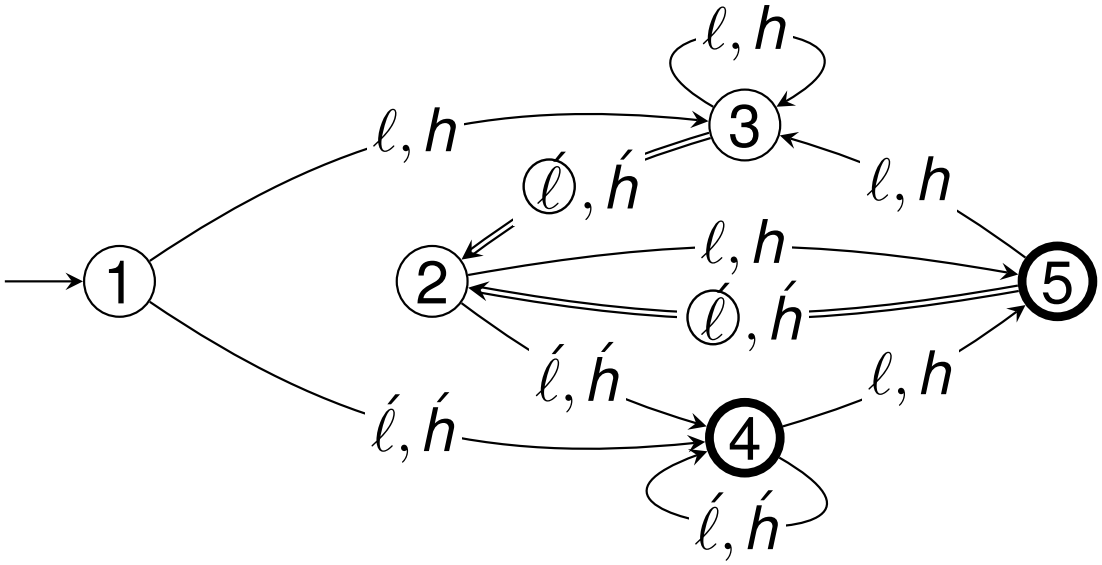
Syntactic semigroup: Actions

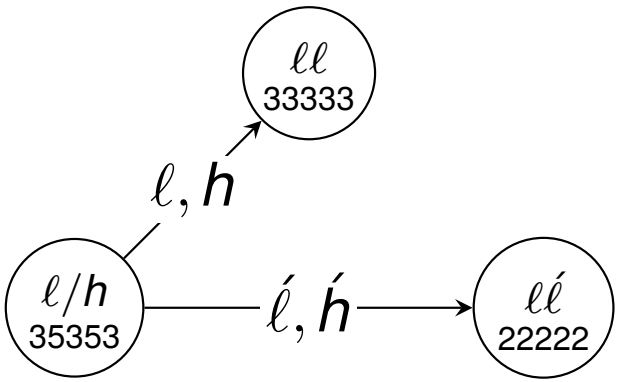


Syntactic semigroup: Actions

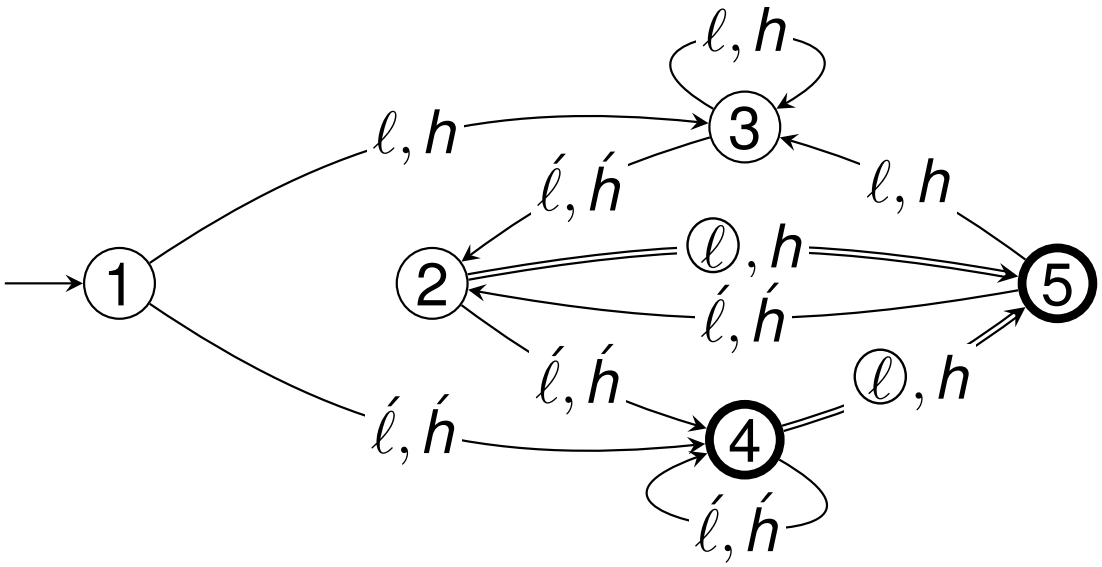


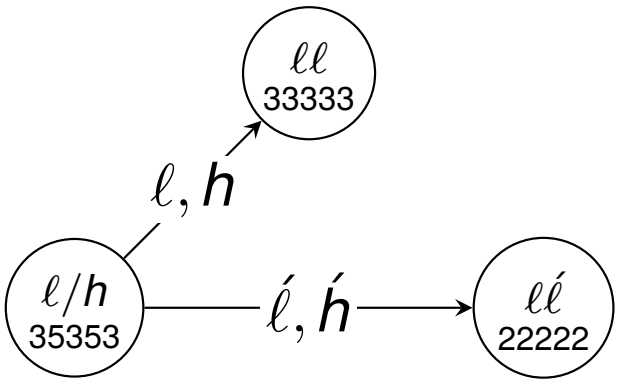
Syntactic semigroup: Actions

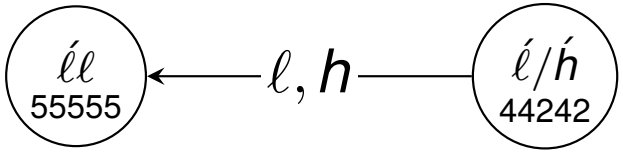




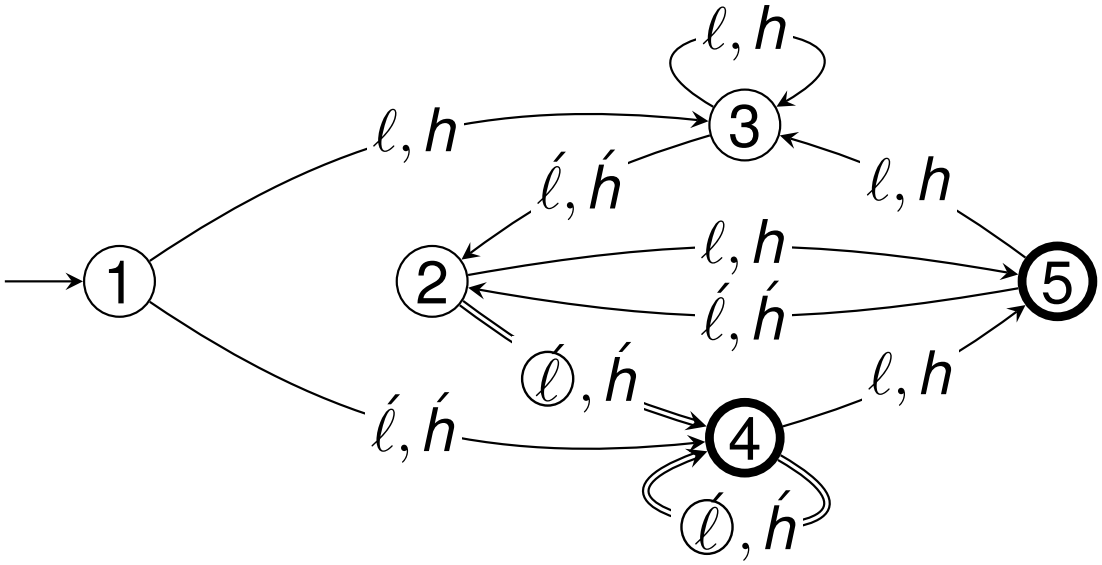
Syntactic semigroup: Actions

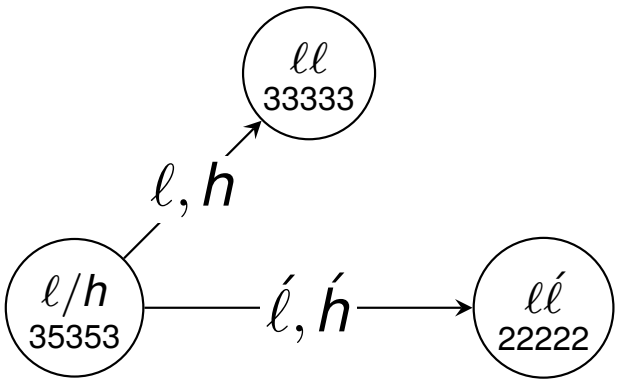


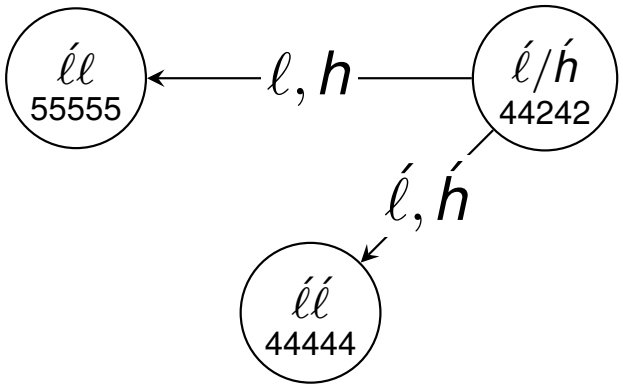




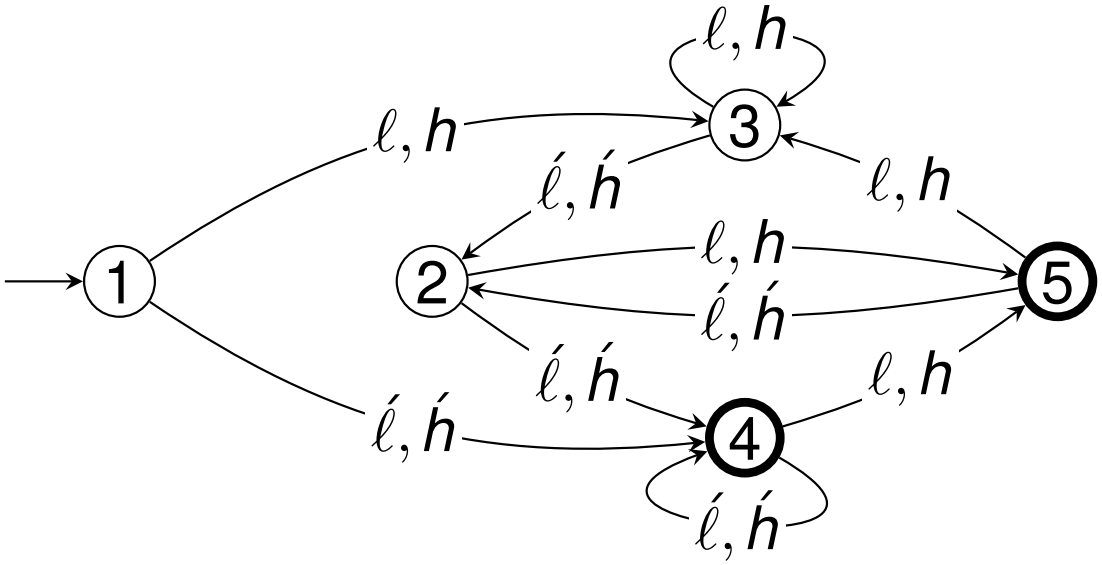
Syntactic semigroup: Actions

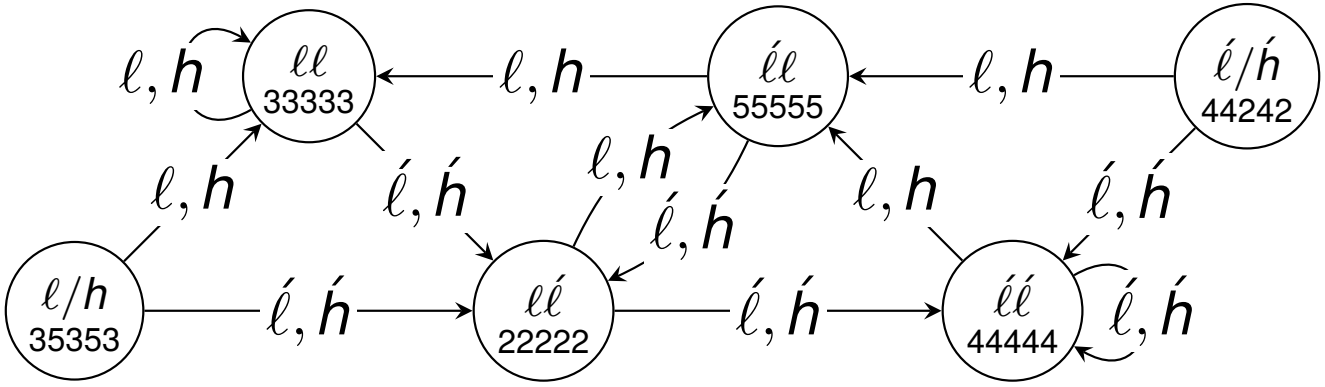




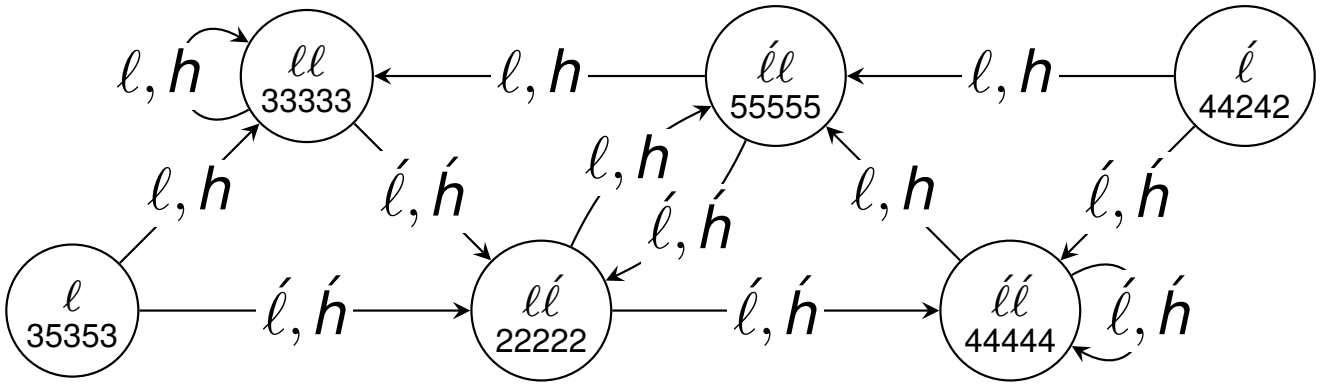


Syntactic semigroup: Actions





Syntactic semigroup: Multiplication table

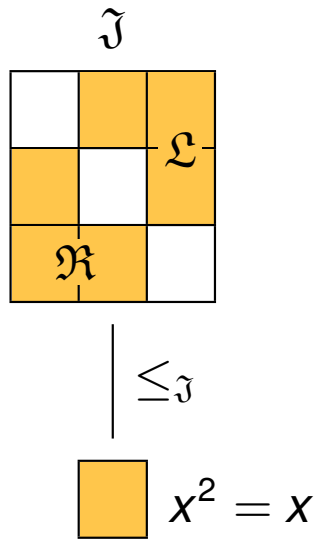


	ℓ	$\acute{\ell}$	$\ell\ell$	$\ell\acute{\ell}$	$\acute{\ell}\ell$	$\acute{\ell}\acute{\ell}$
ℓ	$\ell\ell$	$\ell\acute{\ell}$	$\ell\ell$	$\ell\acute{\ell}$	$\acute{\ell}\ell$	$\acute{\ell}\acute{\ell}$
$\acute{\ell}$	$\acute{\ell}\ell$	$\acute{\ell}\acute{\ell}$	$\ell\ell$	$\ell\acute{\ell}$	$\acute{\ell}\ell$	$\acute{\ell}\acute{\ell}$
$\ell\ell$	$\ell\ell$	$\ell\acute{\ell}$	$\ell\ell$	$\ell\acute{\ell}$	$\acute{\ell}\ell$	$\acute{\ell}\acute{\ell}$
$\ell\acute{\ell}$	$\acute{\ell}\ell$	$\acute{\ell}\acute{\ell}$	$\ell\ell$	$\ell\acute{\ell}$	$\acute{\ell}\ell$	$\acute{\ell}\acute{\ell}$
$\acute{\ell}\ell$	$\ell\ell$	$\ell\acute{\ell}$	$\ell\ell$	$\ell\acute{\ell}$	$\acute{\ell}\ell$	$\acute{\ell}\acute{\ell}$
$\acute{\ell}\acute{\ell}$	$\acute{\ell}\ell$	$\acute{\ell}\acute{\ell}$	$\ell\ell$	$\ell\acute{\ell}$	$\acute{\ell}\ell$	$\acute{\ell}\acute{\ell}$

$x^2 = x$

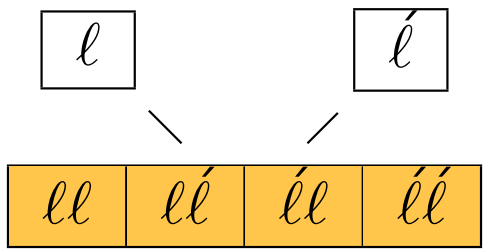
Syntactic semigroup: Green's relations & eggboxes

- $a \leq_{\mathfrak{L}} b$: $a = b$ or $a = sb$ for some s
  a is in the b-column
- $a \leq_{\mathfrak{R}} b$: $a = b$ or $a = bs$ for some s
  a is in the b-row
- $a \leq_{\mathfrak{J}} b$: $a \leq_{\mathfrak{L}} c$ and $c \leq_{\mathfrak{R}} b$ for some c
- $a \mathfrak{L} b$: $a \leq_{\mathfrak{L}} b$ and $b \leq_{\mathfrak{L}} a$
- Similar for $a \mathfrak{R} b$ and $a \mathfrak{J} b$

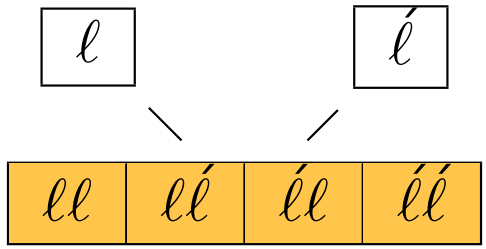


Syntactic semigroup: Stress-penult

	ℓ	ℓ́	ℓℓ	ℓℓ́	ℓ́ℓ	ℓ́ℓ́
ℓ	ℓℓ	ℓℓ́	ℓℓ	ℓℓ́	ℓ́ℓ	ℓ́ℓ́
ℓ́	ℓ́ℓ	ℓ́ℓ́	ℓℓ	ℓℓ́	ℓ́ℓ	ℓ́ℓ́
ℓℓ	ℓℓ	ℓℓ́	ℓℓ	ℓℓ́	ℓ́ℓ	ℓ́ℓ́
ℓℓ́	ℓ́ℓ	ℓ́ℓ́	ℓℓ	ℓℓ́	ℓ́ℓ	ℓ́ℓ́
ℓ́ℓ	ℓℓ	ℓℓ́	ℓℓ	ℓℓ́	ℓ́ℓ	ℓ́ℓ́
ℓ́ℓ́	ℓ́ℓ	ℓ́ℓ́	ℓℓ	ℓℓ́	ℓ́ℓ	ℓ́ℓ́



Definite



Definite: Only care about the last few symbols

- $a \mathfrak{L} b \iff a = b$ ($\mathfrak{L}$-trivial)
- All (and only) idempotents at bottom

Equivalently: $[\![sx^\omega = x^\omega]\!]$ (Brzozowski, 1976; Almeida, 1995)

Reversal

$$ab \mapsto ba$$

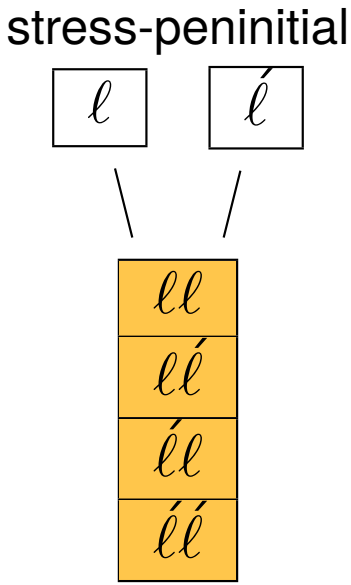
$\mathfrak{L}$ swaps with $\mathfrak{R}$

Reverse definite:

Only care about the first few symbols

- $a \mathfrak{R} b \iff a = b$ ($\mathfrak{R}$-trivial)
- All (and only) idempotents at bottom

Equivalently: $[\![x^\omega s = x^\omega]\!]$ (Brzozowski, 1976; Almeida, 1995)



Bounded stress

Stress anchored to within some fixed distance of a word-edge:
reverse/definite (kind of)

Varieties

Varieties closed under:

- Boolean combinations
- Quotients
- Inverse-homomorphisms

$+$-varieties :: inverse **nonerasing** homomorphisms

Reverse/definite are only $+$-varieties

(Eilenberg, 1976)

Culminativity

Stress does not appear twice.

Long distance: $\acute{\ell}h \ldots h\acute{\ell}$ violates

Local after projection:

$$\begin{array}{ccc} \acute{\ell}h \ldots h\acute{\ell} \\ | \qquad\qquad | \\ \acute{\ell} \qquad\quad \acute{\ell} \end{array}$$

Projection wants to be either empty or single-letter

Tiers

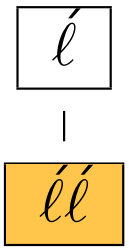
Alphabets: Σ total, Γ salient

Projection: $\pi(\sigma) = \sigma$ if $\sigma \in \Gamma$ else 1
$\pi(xy) = \pi(x)\pi(y)$

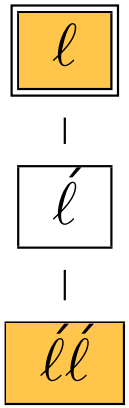
π is a homomorphism

Culminativity: $\pi^{-1}(\{1, \acute{\ell}, \acute{h}\})$ for $\Gamma = \{\acute{\ell}, \acute{h}\}$

Culminativity



$\{1, \acute{\ell}, \acute{h}\}$ on $\Gamma = \{\acute{\ell}, \acute{h}\}$

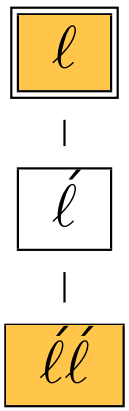


$\pi^{-1}(\{1, \acute{\ell}, \acute{h}\})$

Both definite and reverse definite on projection.
Unprojection is neither!
Idempotent $\ell \equiv 1$ not at bottom.

Tier-based reverse/definite (Heinz et al., 2011; Lambert, 2023)

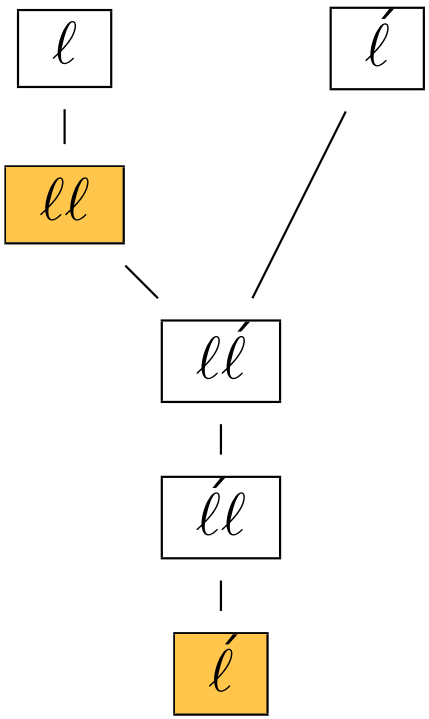
Piecewise-testablitiy



- Piecewise testable $=$ propositional with subsequences
- $\mathfrak{J}$-trivial (Simon, 1975)

Culminativity:
tier-based reverse/definite **and** piecewise testable

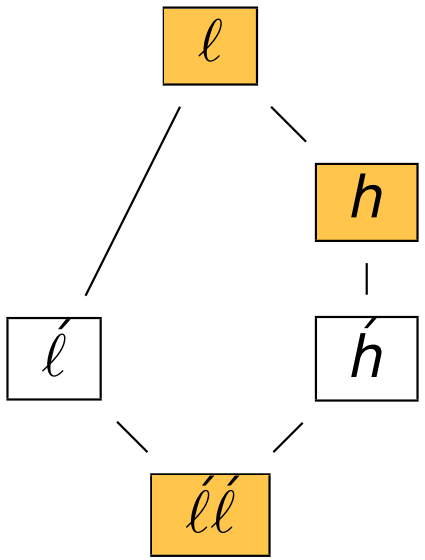
The real stress-penult and stress-peninitial



- $\mathfrak{J}$-trivial: piecewise testable
- Not reverse/definite
- Not tier-based reverse/definite
  - Needs **two** tiers:
  - Culminativity on stress-tier
  - Stress-anchoring on full word
- Multitier reverse/definite
- Bounded stress is long-distance!

Unbounded stress

{Rightmost,Leftmost} heavy else {rightmost,leftmost}
(with culminativity)



All piecewise testable ($\mathfrak{J}$-trivial)

Unbounded stress: Default-to-same

- Rightmost heavy else rightmost: multitier definite

$$[\rtimes \acute{\sigma} \ltimes]_{\{\acute{\sigma}\}} \wedge ([\acute{h} \ltimes]_{\{h,\acute{h}\}} \vee ([\rtimes \ltimes]_{\{h,\acute{h}\}} \wedge \acute{\ell} \ltimes))$$

- Leftmost heavy else leftmost: multitier reverse definite

$$[\rtimes \acute{\sigma} \ltimes]_{\{\acute{\sigma}\}} \wedge ([\rtimes \acute{h}]_{\{h,\acute{h}\}} \vee ([\rtimes \ltimes]_{\{h,\acute{h}\}} \wedge \rtimes \acute{\ell}))$$

Three tiers: stress, heavy, and word

Unbounded stress: Default-to-opposite

- Rightmost heavy else lefttmost: multitier gen. definite

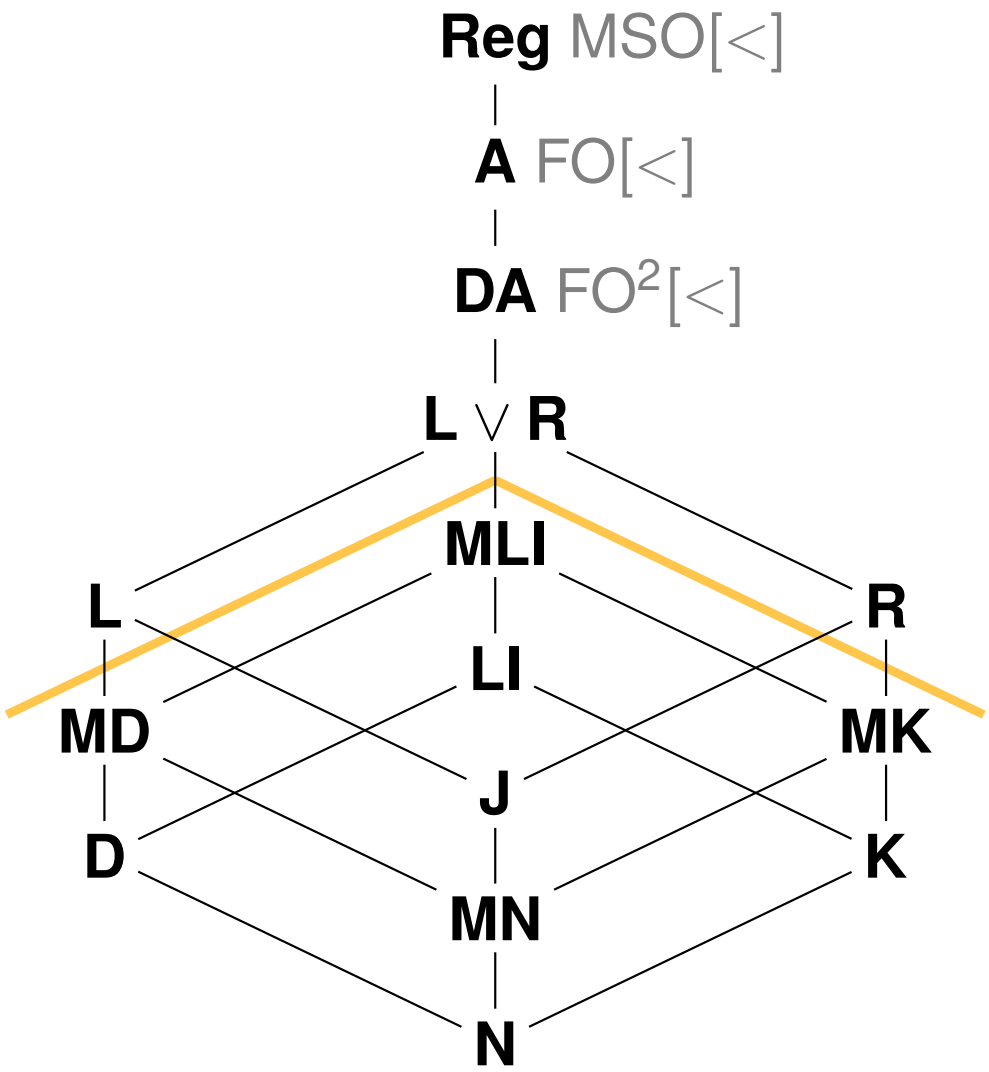
$$[\rtimes \acute{\sigma} \ltimes]_{\{\acute{\sigma}\}} \wedge ([\acute{h} \ltimes]_{\{h,\acute{h}\}} \vee ([\rtimes \ltimes]_{\{h,\acute{h}\}} \wedge \rtimes \acute{\ell}))$$

- Leftmost heavy else rightmost: multitier gen. definite

$$[\rtimes \acute{\sigma} \ltimes]_{\{\acute{\sigma}\}} \wedge ([\rtimes \acute{h}]_{\{h,\acute{h}\}} \vee ([\rtimes \ltimes]_{\{h,\acute{h}\}} \wedge \acute{\ell} \ltimes))$$

Three tiers: stress, heavy, and word

A few varieties



D definite

K rev. def.

LI gen. def.

N co/finite

M*V* multitier *V*

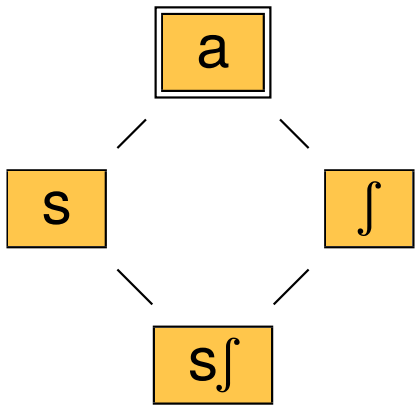
J $\mathfrak{J}$-trivial

L $\mathfrak{L}$-trivial

R $\mathfrak{R}$-trivial

Harmony: Symmetric

Words cannot have both $\int$ and s

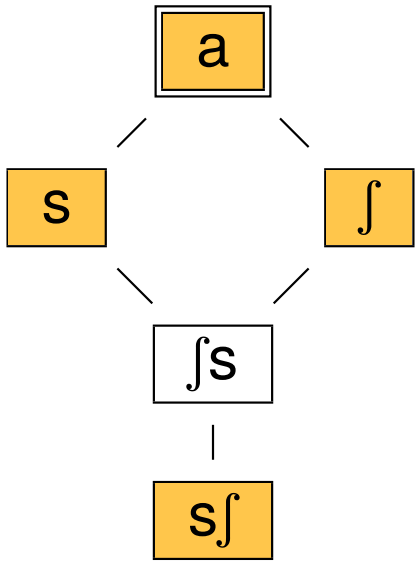


$\mathfrak{J}$-trivial and multitier co/finite

$$[\rtimes \ltimes]_{\{s\}} \vee [\rtimes \ltimes]_{\{\int\}}$$

Harmony: Asymmetric

Words can have ∫...s but cannot have s...∫

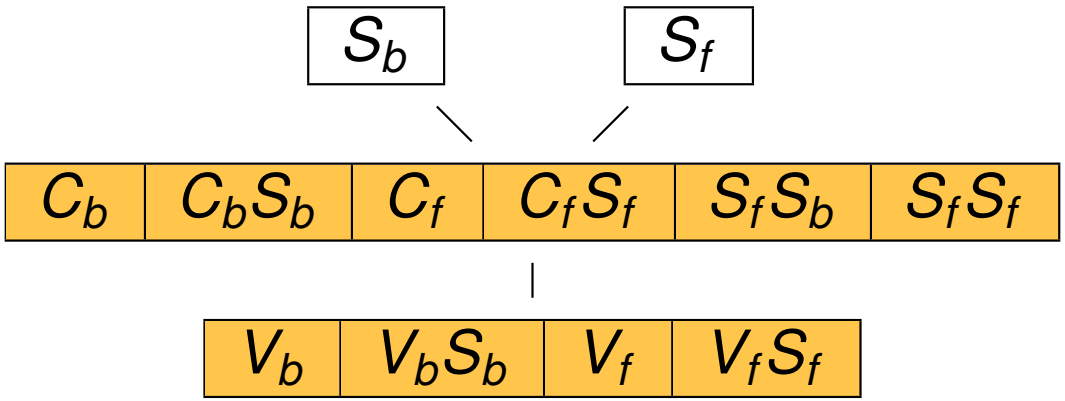


$\mathfrak{J}$-trivial but not **MLI**

$$\int^k \int s s^k \neq \int^k s \int s^k$$

despite having same k-prefix and k-suffix on all tiers

Uyghur backness harmony



NOT $\mathfrak{J}$-trivial
but is multitier definite:

$$
\begin{aligned}
&([V_f \ltimes]_{V_f \cup V_b} \wedge S_f \ltimes) \\
\vee&([V_b \ltimes]_{V_f \cup V_b} \wedge S_b \ltimes) \\
\vee&([\rtimes \ltimes]_{V_f \cup V_b} \wedge [C_f]_{C_f \cup C_b} \wedge C_fS_f \ltimes) \\
\vee&([\rtimes \ltimes]_{V_f \cup V_b} \wedge [C_b]_{C_f \cup C_b} \wedge C_bS_b \ltimes)
\end{aligned}
$$

Software

The Language Toolkit does the classification

`https://hackage.haskell.org/package/language-toolkit`

How? Create semigroup, check membership via equations

Example — “is *expr* multitier definite”:

`:isMTDef` *expr*

`:isVarietyS [xyx*=yx*]` *expr*

Conclusions

- Phonotactics: propositional
  - Substrings (adjacency)
  - Subsequences (precedence)
  - Multitier (adjacency on one or more projections)
- Transducers $\mapsto$ semigroups too
- Same classes exist for functions
- Language Toolkit can classify functions too
- Future work
  - Collect and classify functions
  - Multirelation propositional stuff

References

Jorge Almeida. 1995. Finite Semigroups and Universal Algebra, volume 3 of Series in Algebra. World Scientific, Singapore.

Janusz Antoni Brzozowski. 1976. A generalization of finiteness. Semigroup Forum, 13:239–251.

Samuel Eilenberg. 1976. Automata, Languages, and Machines, volume B. Academic Press, New York, New York.

Jeffrey Heinz, Chetan Rawal, and Herbert G. Tanner. 2011. Tier-based strictly local constraints for phonology. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Short Papers, volume 2, pages 58–64, Portland, Oregon. Association for Computational Linguistics.

Dakotah Lambert. 2023. Relativized adjacency. Journal of Logic, Language and Information, 32(4):707–731.

Imre Simon. 1975. Piecewise testable events. In Helmut Brakhage, editor, Automata Theory and Formal Languages, volume 33 of Lecture Notes in Computer Science, pages 214–222. Springer-Verlag, Berlin.