On the Cognitive Complexity of Phonotactic Constraints

James Rogers
Dept. of Computer Science
Earlham College
`jrogers@cs.earlham.edu`

Slide 1

`http://cs.earlham.edu/~jrogers/slides/stonybrook.ho.pdf`

Joint work with Jeff Heinz (UDel), Sean Wibel, Maggie Fero and Dakotah Lambert (EC)

Slide 2

Some simple patterns

(1) Primary stress falls on the final syllable

(2) Primary stress falls on the antepenultimate syllable

(3) In words of five or more syllables primary stress falls on the antepenultimate syllable

Slide 3

Some simple patterns

(4) Primary stress falls on the initial syllable if it is heavy, else the peninitial syllable.

(5) Primary stress falls on the leftmost heavy syllable

(6) Secondary stress falls on every third syllable counting left from the antepenultimate syllable.

Slide 4

Some simple patterns

(7) Final syllable is heavy

(8) All heavy syllables get some stress

(9) There are always an odd number of heavy syllables

Slide 5

Some simple patterns

(10) Primary stress falls on some syllable. (At least one)

(11) Primary stress falls on at most one syllable.

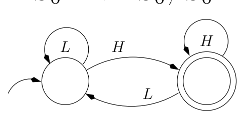
(12) Primary stress falls on exactly one syllable.

Slide 6

Complexity of Simple Patterns

(7) Sequences of 'L's and 'H's which end in 'H':

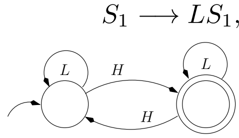
$$S_0 \longrightarrow LS_0,\ S_0 \longrightarrow HS_0,\ S_0 \longrightarrow H$$



$$(L + H)^*H$$

(9) Sequences of 'L's and 'H's which contain an odd number of 'H's:

$$S_0 \longrightarrow LS_0,\ S_0 \longrightarrow HS_1,$$
$$S_1 \longrightarrow LS_1,\ S_1 \longrightarrow HS_0,\ S_1 \longrightarrow \varepsilon$$

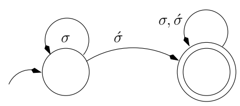


$$(L^*HL^*HL^*)^*L^*HL^*$$

Some More Simple Patterns

(10) Sequences of 'σ's and '$\acute{\sigma}$'s which contain at least one '$\acute{\sigma}$':

$$S_0 \longrightarrow \sigma S_0,\ S_0 \longrightarrow \acute{\sigma} S_1,$$
$$S_1 \longrightarrow \sigma S_1,\ S_1 \longrightarrow \acute{\sigma} S_1,\ S_1 \longrightarrow \varepsilon$$

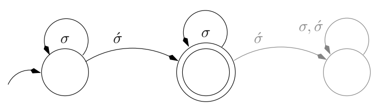


$\sigma^* \acute{\sigma} (\sigma + \acute{\sigma})^*$

Slide 7

(12) Sequences of 'σ's and '$\acute{\sigma}$'s which contain exactly one '$\acute{\sigma}$':

$$S_0 \longrightarrow \sigma S_0,\ S_0 \longrightarrow \acute{\sigma} S_1,$$
$$S_1 \longrightarrow \sigma S_1,\ S_1 \longrightarrow \varepsilon$$



$\sigma^* \acute{\sigma} \sigma^*$

Cognitive Complexity from First Principles

What kinds of distinctions does a cognitive mechanism need to be sensitive to in order to classify an event with respect to a pattern?

Slide 8

Reasoning about patterns

- What objects/entities/things are we reasoning about?
- What relationships between them are we reasoning with?

Some Assumptions about Linguistic Behaviors

Slide 9

- Perceive/process/generate linear sequence of (sub)events
- Can model as strings—linear sequence of abstract symbols
  - Discrete linear order (initial segment of $\mathbb{N}$).
  - Labeled with alphabet of events

    Partitioned into subsets, each the set of positions at which some event occurs.

Word models

Slide 10

$$\langle \mathcal{D}, \triangleleft, <, P_\sigma \rangle_{\sigma\in\Sigma}$$

$(+1)$ $\langle \mathcal{D}, \triangleleft, P_\sigma \rangle_{\sigma\in\Sigma}$ $(<)$ $\langle \mathcal{D}, <, P_\sigma \rangle_{\sigma\in\Sigma}$

$\mathcal{D}$	—	Finite
$<$	—	Linear order on $\mathcal{D}$
$\triangleleft$	—	Successor wrt $<$
P_σ	—	Subset of $\mathcal{D}$ at which σ occurs
		(P_σ partition $\mathcal{D}$)

$$CCVC = \langle\ \{0,1,2,3\},\ \{\langle i, i+1\rangle \mid 0 \leq i < 3\},\ \{0,1,3\}_C,\ \{2\}_V\ \rangle$$

$$\langle\ \ \mathcal{D}\ \ \ \ \triangleleft\ \ \ \ P_C\ \ \ \ P_V\ \ \rangle$$

Slide 11

An Alphabet for Stress Patterns

Syllable Weight				Stress		
L	=	Light		σ	=	Unstressed
H	=	Heavy		$\acute{\sigma}$	=	Primary Stress
S	=	Super Heavy		$\grave{\sigma}$	=	Secondary Stress
σ	=	Arbitrary		$\overset{+}{\sigma}$	=	Some Stress
				$\bar{\sigma}$	=	Not Primary Stress
				$\overset{*}{\sigma}$	=	Arbitrary Stress

eg: $\acute{\mathrm{H}}\mathrm{L}\grave{\mathrm{H}}$

Slide 12

Local Constraints

- Blocks of adjacent symbols
  - k-factors
- End markers: ‘$\rtimes$’, ‘$\ltimes$’

$$\mathrm{F}^2_{(}\rtimes\sigma\sigma\acute{\sigma}\ltimes) = \{\rtimes\sigma,\ \sigma\sigma,\ \sigma\acute{\sigma},\ \acute{\sigma}\ltimes\}$$
$$\mathrm{F}^3_{(}\rtimes\sigma\sigma\acute{\sigma}\ltimes) = \{\rtimes\sigma\sigma,\ \sigma\sigma\acute{\sigma},\ \sigma\acute{\sigma}\ltimes\}$$
$$\mathrm{F}^6_{(}\rtimes\sigma\sigma\acute{\sigma}\ltimes) = \{\rtimes\sigma\sigma\acute{\sigma}\ltimes\}$$

Strictly k-Local Constraints

- Co-occurrence of negative atomic local constraints
  - Conjunctions of negated k-factors

Slide 13

(1) Primary stress falls on the final syllable

$$\neg\bar{\sigma}\ltimes \qquad (\mathrm{SL}_2)$$

(2) Primary stress falls on the antepenultimate syllable

$$\neg\acute{\sigma}\overset{*}{\sigma}\overset{*}{\sigma}\overset{*}{\sigma} \;\wedge\; \neg\acute{\sigma}\overset{*}{\sigma}\ltimes \;\wedge\; \neg\acute{\sigma}\ltimes \qquad (\mathrm{SL}_4)$$

Cambodian

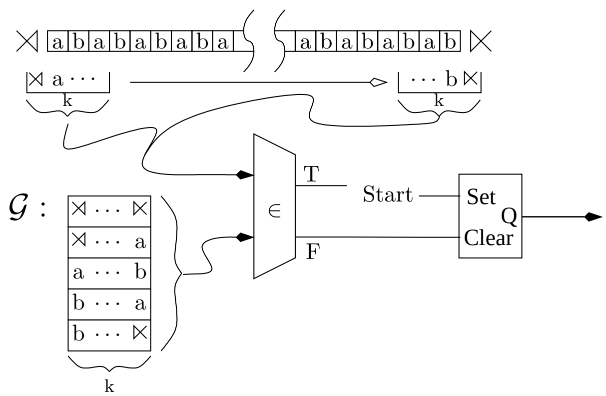
Slide 14

1)	In words of all sizes, primary stress falls on the final syllable.	$\neg\sigma\ltimes \wedge \neg\grave{\sigma}\ltimes$	(SL_2)
1b)	Primary stress does not fall before the final syllable.	$\neg\acute{\sigma}\overset{*}{\sigma}$	(SL_2)
2)	In words of all sizes, secondary stress falls on all heavy syllables.	$\neg\mathrm{H}$	(SL_1)
3)	Light syllables occur only immediately following heavy syllables.	$\neg\rtimes\overset{*}{\mathrm{L}} \wedge \neg\overset{*}{\mathrm{L}}\overset{*}{\mathrm{L}}$	(SL_2)
[4)	Light monosyllables do not occur.	$\neg\rtimes\acute{\mathrm{L}}\ltimes$	(SL_3)]

Cambodian stress is SL_2.

Scanners

Slide 15



Recognizing an SL_k stringset requires only remembering the k most recently encountered symbols.

Character of Strictly k-Local Sets

Slide 16

Theorem (Suffix Substitution Closure):
A stringset L is strictly k-local iff whenever there is a string x of length $k-1$ and strings w, y, v, and z, such that

$$\begin{array}{ccccccc} w & \cdot & \overbrace{x}^{k-1} & \cdot & y & \in L \\ v & \cdot & x & \cdot & z & \in L \end{array}$$

then it will also be the case that

$$w \cdot x \cdot z \in L$$

$\star\, CCC$ is SL_3

$$\begin{array}{ccccc|l} V & \cdot & CC & \cdot & VC & \in \star\, CCC \\ CV & \cdot & CC & \cdot & V & \in \star\, CCC \\ \hline V & \cdot & CC & \cdot & V & \in \star\, CCC \end{array}$$

But $\star\, CCC$ is not SL_2:

$$\begin{array}{ccccc|l} C & \cdot & C & \cdot & VC & \in \star\, CCC \\ V & \cdot & C & \cdot & CV & \in \star\, CCC \\ \hline C & \cdot & C & \cdot & CV & \notin \star\, CCC \end{array}$$

Slide 17

Alawa

- In words of all sizes, primary stress falls on the penultimate syllable.
- [—Except in monosyllables]

$$\begin{array}{rl}\mathcal{G}_{\text{Alawa}} = \{ & \rtimes\sigma\sigma,\ \rtimes\sigma\acute{\sigma},\ \rtimes\acute{\sigma}\sigma, \\ & \sigma\sigma\sigma,\ \sigma\sigma\acute{\sigma},\ \sigma\acute{\sigma}\sigma, \\ & \acute{\sigma}\sigma\ltimes,\ \rtimes\acute{\sigma}\ltimes \qquad \}\end{array}$$

$\rtimes\sigma$	$\acute{\sigma}$	$\sigma\ltimes$
$\rtimes$	$\acute{\sigma}$	$\ltimes$
$\star\ \rtimes\sigma$	$\acute{\sigma}$	$\ltimes$

$\rtimes\sigma$	σ	$\acute{\sigma}\sigma\ltimes$
$\rtimes\acute{\sigma}$	σ	$\ltimes$
$\star\ \rtimes\sigma$	σ	$\ltimes$

Alawa stress is in $\text{SL}_3 - \text{SL}_2$.

Slide 18

SL Hierarchy

Theorem 1 (SL-Hierarchy)

$$SL_1 \subsetneq SL_2 \subsetneq SL_3 \subsetneq \cdots \subsetneq SL_i \subsetneq SL_{i+1} \subsetneq \cdots \subsetneq SL$$

Every Finite stringset is SL_k for some k: Fin $\subseteq$ SL.

There is no k for which SL_k includes all Finite stringsets.

SL_k is learnable in the limit from positive data.

SL is not.

Slide 19

Cognitive interpretation of SL

- Any cognitive mechanism that can distinguish member strings from non-members of a (properly) SL_k stringset must be sensitive, at least, to the length k blocks of consecutive events that occur in the presentation of the string.
- Any cognitive mechanism that is sensitive *only* to the co-occurrence of length k blocks of consecutive events in the presentation of a string will be able to recognize *only* SL_k stringsets.

Sequential: This corresponds to being sensitive, at each point in the string, to the immediately prior sequence of $k-1$ events.

Parallel: This corresponds to being sensitive to the presence of simple contiguous blocks in the string.

Slide 20

Strictly Local Stress Patterns

StressTyp2 Database (2015)—699 languages, 106 formally distinct patterns

9 are SL_2	Abun West, Afrikans, ... Cambodian,... Maranungku
44 are SL_3	Alawa, Arabic (Bani-Hassan),...
23 are SL_4	Dutch,...
3 are SL_5	Asheninca, Bhojpuri, Hindi (Fairbanks)
1 is SL_6	Icua Tupi
26 are not SL	Amele, Bhojpuri (Shukla Tiwari), Arabic (Classical), Hindi (Kelkar), Yidin,...

75% are SL, all $k \leq 6$. 50% are SL_3.

Obligatoriness: Some-$\acute{\sigma}$

Slide 21

$\rtimes\sigma$	$\overbrace{\sigma\cdots\sigma}^{k-1}$	$\grave{\sigma}\ltimes$
$\rtimes\acute{\sigma}$	$\overbrace{\sigma\cdots\sigma}^{k-1}$	$\sigma\ltimes$
$\star\ \sigma$	$\overbrace{\sigma\cdots\sigma}^{k-1}$	$\sigma\ltimes$

Some-$\acute{\sigma} \notin$ SL

How can any stress pattern be SL?

Locally definable stringsets

Slide 22

$$\begin{array}{ccc} f \in F_k(\rtimes\cdot\Sigma^*\cdot\ltimes) & w \models f & \stackrel{\text{def}}{\Longleftrightarrow} & f \in F_k(\rtimes\cdot w\cdot\ltimes) \\ \varphi\wedge\psi & w\models\varphi\wedge\psi & \stackrel{\text{def}}{\Longleftrightarrow} & w\models\varphi \text{ and } w\models\psi \\ \neg\varphi & w\models\neg\varphi & \stackrel{\text{def}}{\Longleftrightarrow} & w\not\models\varphi \end{array}$$

$$\varphi\vee\psi \ \equiv\ \neg(\varphi\wedge\psi)$$

$$L = L(\varphi) \stackrel{\text{def}}{=} \{w\in\Sigma^* \mid w\models\varphi\}$$

$$\text{SL}_k \equiv \bigwedge_{f_i\notin\mathcal{G}} [\neg f_i] \subsetneq \text{LT}_k$$

Some-$\acute{\boldsymbol{\sigma}}$ again

Slide 23

$$\text{Some-}\acute{\sigma} = L(\acute{\sigma})$$

$$\text{Some-}\acute{\sigma} \in \text{LT}_1$$

NKL

Slide 24

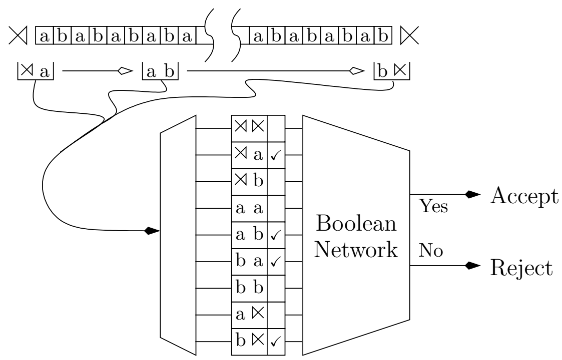
- Primary stress falls on the final syllable if it is Heavy
- Else on the initial syllable if it is Light
- Else on the penultimate syllable

$\varphi_{\text{NKL}} =$

$\acute{\text{H}}\ltimes$	final syllable if it is Heavy
$\vee\ (\neg\acute{\text{H}}\ltimes \wedge \rtimes\acute{\text{L}})$	Else on the initial if it is Light
$\vee\ (\neg\acute{\text{H}}\ltimes \wedge \neg\rtimes\acute{\text{L}} \wedge \acute{\sigma}\overset{*}{\sigma}\ltimes)$	Else on the penultimate syllable

LT Automata

Slide 25



Membership in an LT_k stringset depends only on the set of k-Factors which occur in the string.

Recognizing an LT_k stringset requires only remembering which k-factors occur in the string.

Character of Locally Testable sets

Slide 26

Theorem 2 (k-Test Invariance) *A stringset L is Locally Testable iff*

there is some k such that, for all strings x and y,

if $\rtimes \cdot x \cdot \ltimes$ and $\rtimes \cdot y \cdot \ltimes$ have exactly the same set of k-factors

then either both x and y are members of L or neither is.

Definition 1 (k-Local Equivalence)

$$w \equiv^L_k v \overset{def}{\iff} F_k(\rtimes w \ltimes) = F_k(\rtimes v \ltimes).$$

$$\mathrm{LT}_1 \subsetneq \mathrm{LT}_2 \subsetneq \mathrm{LT}_3 \subsetneq \cdots \subsetneq \mathrm{LT}_i \subsetneq \mathrm{LT}_{i+1} \subsetneq \cdots \subsetneq \mathrm{LT}$$

Slide 27

Cognitive interpretation of LT

- Any cognitive mechanism that can distinguish member strings from non-members of a (properly) LT_k stringset must be sensitive, at least, to the *set* of length k contiguous blocks of events that occur in the presentation of the string—both those that do occur and those that do not.
- Any cognitive mechanism that is sensitive *only* to the occurrence or non-occurrence of length k contiguous blocks of events in the presentation of a string will be able to recognize *only* LT_k stringsets.

Sequential: This corresponds to being sensitive, at each point in the string, to the set of length k blocks of events that occurred at any prior point.

Parallel: This corresponds to being sensitive to the presence of sets of simple contiguous blocks in the string.

Slide 28

Murik

- Primary stress falls on the leftmost heavy syllable
- else the initial syllable

Murik

- Primary stress falls on the leftmost heavy syllable
- else the initial syllable
- No more than one heavy syllable occurs in any word

Slide 29

Murik

- Primary stress falls on the leftmost heavy syllable
- else the initial syllable
- No more than one heavy syllable occurs in any word

Slide 30

$$L_{\text{Murik}} = \neg \text{H} \ \wedge \ (\acute{\text{H}} \vee \rtimes\acute{\sigma}) \ \wedge \cdots$$

Murik

- Primary stress falls on the leftmost heavy syllable
- else the initial syllable
- No more than one heavy syllable occurs in any word

Slide 31

$$\rtimes\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\acute{\mathrm{H}}\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\ltimes \quad \in L_{\text{Murik}}$$

$$\rtimes\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\acute{\mathrm{H}}\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\acute{\mathrm{H}}\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\ltimes \quad \notin L_{\text{Murik}}$$

$$\mathrm{F}^k_(\,\rtimes\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\acute{\mathrm{H}}\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\ltimes) = \mathrm{F}^k_(\,\rtimes\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\acute{\mathrm{H}}\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\acute{\mathrm{H}}\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\ltimes)$$

$$= \{\rtimes\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1},\ \overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\acute{\mathrm{H}},\ \ldots,\ \acute{\mathrm{H}}\overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1},\ \overbrace{\mathrm{L}\cdots\mathrm{L}}^{k-1}\ltimes\}$$

Culmanitivity: (at most) One-$\acute{\sigma}$

$$\rtimes\overbrace{\sigma\cdots\sigma}^{k-1}\acute{\sigma}\overbrace{\sigma\cdots\sigma}^{k-1}\ltimes \quad \in L_{\text{One}-\acute{\sigma}}$$

Slide 32

$$\rtimes\overbrace{\sigma\cdots\sigma}^{k-1}\acute{\sigma}\overbrace{\sigma\cdots\sigma}^{k-1}\acute{\sigma}\overbrace{\sigma\cdots\sigma}^{k-1}\ltimes \quad \notin L_{\text{One}-\acute{\sigma}}$$

One-$\acute{\sigma}$ is not LT (hence not SL)

Slide 33

FO(+1)

Models: $\langle \mathcal{D}, \triangleleft, P_\sigma \rangle_{\sigma\in\Sigma}$

First-order Quantification (over positions in the strings)

$x \triangleleft y$	$w, [x \mapsto i, y \mapsto j] \models x \triangleleft y$	$\stackrel{\text{def}}{\iff}$	$j = i + 1$
$P_\sigma(x)$	$w, [x \mapsto i] \models P_\sigma(x)$	$\stackrel{\text{def}}{\iff}$	$i \in P_\sigma$
$\varphi \wedge \psi$	$\vdots$		
$\neg$	$\vdots$		
$(\exists x)[\varphi(x)]$	$w, s \models (\exists x)[\varphi(x)]$	$\stackrel{\text{def}}{\iff}$	$w, s[x \mapsto i] \models \varphi(x)]$ for some$i \in \mathcal{D}$

FO(+1)-Definable Stringsets: $L(\varphi) \stackrel{\text{def}}{=} \{w \mid w \models \varphi\}$.

$$\varphi_{\text{One-}\acute{\sigma}} = (\exists x)[\acute{\sigma}(x) \wedge (\forall y)[\acute{\sigma}(y) \rightarrow x \approx y]$$

Slide 34

Character of the FO(+1) Definable Stringsets

Definition 2 (Locally Threshold Testable) *A set L is* Locally Threshold Testable *(LTT) iff there is some k and t such that, for all $w, v \in \Sigma^*$:*

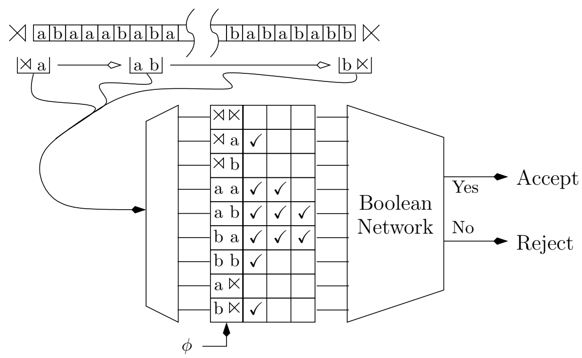
if for all $f \in F_k(\rtimes \cdot w \cdot \ltimes) \cup F_k(\rtimes \cdot v \cdot \ltimes)$
either $|w|_f = |v|_f$ or both $|w|_f \geq t$ and $|v|_f \geq t$,

then $w \in L \iff v \in L$.

Theorem 3 (Thomas) *A set of strings is First-order definable over $\langle \mathcal{D}, \triangleleft, P_\sigma \rangle_{\sigma\in\Sigma}$ iff it is* Locally Threshold Testable.

Slide 35

LTT Automata



Membership in an FO(+1) definable stringset depends only on the multiplicity of the k-factors, up to some fixed finite threshold, which occur in the string.

Slide 36

Cognitive interpretation of FO(+1)

- Any cognitive mechanism that can distinguish member strings from non-members of a (properly) FO(+1) stringset must be sensitive, at least, to the multiplicity of the length k blocks of events, for some fixed k, that occur in the presentation of the string, distinguishing multiplicities only up to some fixed threshold t.
- Any cognitive mechanism that is sensitive *only* to the multiplicity, up to some fixed threshold, (and, in particular, not to the order) of the length k blocks of events in the presentation of a string will be able to recognize *only* FO(+1) stringsets.

Sequential: This corresponds to being able count up to some fixed threshold.

Parallel: This corresponds to being sensitive to the multiplicity of simple contiguous blocks in the string.

No-H-before-H́

- Primary stress falls on the leftmost heavy syllable
- (Murik), Maori, Yidin, Kashmiri, …

Slide 37

$$\star\ \mathrm{H} \ldots \acute{\mathrm{H}}$$

$$\rtimes \overbrace{\grave{\mathrm{L}}\mathrm{L}\cdots\grave{\mathrm{L}}\mathrm{L}}^{2kt}\ \acute{\mathrm{H}}\mathrm{H}\ \overbrace{\grave{\mathrm{L}}\mathrm{L}\cdots\grave{\mathrm{L}}\mathrm{L}}^{2kt}\ \grave{\mathrm{H}}\mathrm{H}\ \overbrace{\grave{\mathrm{L}}\mathrm{L}\cdots\grave{\mathrm{L}}\mathrm{L}}^{2kt} \ltimes$$

$$\equiv^{L}_{k,t}$$

$$\star \rtimes \underbrace{\grave{\mathrm{L}}\mathrm{L}\cdots\grave{\mathrm{L}}\mathrm{L}}_{2kt}\ \grave{\mathrm{H}}\mathrm{H}\ \underbrace{\grave{\mathrm{L}}\mathrm{L}\cdots\grave{\mathrm{L}}\mathrm{L}}_{2kt}\ \acute{\mathrm{H}}\mathrm{H}\ \underbrace{\grave{\mathrm{L}}\mathrm{L}\cdots\grave{\mathrm{L}}\mathrm{L}}_{2kt} \ltimes$$

Precedence—Subsequences

Definition 3 (Subsequences)

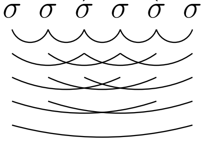
$$v \sqsubseteq w \overset{def}{\Longleftrightarrow} v = \sigma_1 \cdots \sigma_n \ \textit{and}\ w \in \Sigma^* \cdot \sigma_1 \cdot \Sigma^* \cdots \Sigma^* \cdot \sigma_n \cdot \Sigma^*$$

$$P_k(w) \overset{def}{=} \{v \in \Sigma^k \mid v \sqsubseteq w\}$$

$$P_{\leq k}(w) \overset{def}{=} \{v \in \Sigma^{\leq k} \mid v \sqsubseteq w\}$$

Slide 38

$\sigma\ \sigma\ \acute{\sigma}\ \sigma\ \grave{\sigma}\ \sigma$



$\sigma\sigma, \sigma\acute{\sigma}, \acute{\sigma}\sigma, \sigma\grave{\sigma}, \grave{\sigma}\sigma$
$\sigma\acute{\sigma}, \sigma\sigma, \acute{\sigma}\grave{\sigma}$
$\sigma\sigma, \sigma\grave{\sigma}, \acute{\sigma}\sigma$
$\sigma\grave{\sigma}, \sigma\sigma$
$\sigma\sigma$

$$P_2(\sigma\sigma\acute{\sigma}\sigma\grave{\sigma}\sigma) = \{\sigma\sigma, \sigma\acute{\sigma}, \sigma\grave{\sigma}, \acute{\sigma}\sigma, \acute{\sigma}\grave{\sigma}, \grave{\sigma}\sigma\}$$

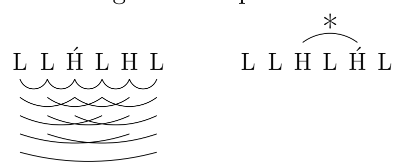
$$P_{\leq 2}(\sigma\sigma\acute{\sigma}\sigma\grave{\sigma}\sigma) = \{\varepsilon, \sigma, \acute{\sigma}, \grave{\sigma}, \sigma\sigma, \sigma\acute{\sigma}, \sigma\grave{\sigma}, \acute{\sigma}\sigma, \acute{\sigma}\grave{\sigma}, \grave{\sigma}\sigma\}$$

Slide 39

Strictly Piecewise Stringsets—SP

Strictly k-Piecewise Definitions

- Co-occurrence of negative atomic piecewise constraints
  - Conjunctions of negated k-sequences



Membership in an SP_k stringset depends only on the individual ($\leq k$)-subsequences which do and do not occur in the string.

Slide 40

Character of the Strictly k-Piecewise Sets

Theorem 4 *A stringset L is Strictly k-Piecewise Testable iff it is closed under subsequence:*

$$w\sigma v \in L \Rightarrow wv \in L$$

Every naturally occurring stress pattern requires Primary Stress

$\Rightarrow$

No naturally occurring stress pattern is SP.

But SP can forbid multiple primary stress: $\neg\acute{\sigma}..\acute{\sigma}$

Cognitive interpretation of SP

Slide 41

- Any cognitive mechanism that can distinguish member strings from non-members of a (properly) SP_k stringset must be sensitive, at least, to the length k (not necessarily consecutive) sequences of events that occur in the presentation of the string.
- Any cognitive mechanism that is sensitive *only* to the length k sequences of events in the presentation of a string will be able to recognize *only* SP_k stringsets.

Sequential: This corresponds to being sensitive, at each point in the string, to up to $k-1$ events distributed arbitrarily among the prior events.

Parallel: This corresponds to being sensitive to the order of inidividual events in the string.

k-Piecewise Testable Stringsets

Slide 42

PT_k-expressions

$$\begin{array}{ccccc} p \in \Sigma^{\leq k} & & w \models p & \stackrel{\text{def}}{\iff} & p \sqsubseteq w \\ \varphi \wedge \psi & & w \models \varphi \wedge \psi & \stackrel{\text{def}}{\iff} & w \models \varphi \text{ and } w \models \psi \\ \neg\varphi & & w \models \neg\varphi & \stackrel{\text{def}}{\iff} & w \not\models \varphi \end{array}$$

k-Piecewise Testable Stringsets (PT_k):

$$L(\varphi) \stackrel{\text{def}}{=} \{w \in \Sigma^* \mid w \models \varphi\}$$

$$\text{One-}\acute{\sigma} = L(\acute{\sigma} \wedge \neg\acute{\sigma}..\acute{\sigma})$$

Membership in an PT_k stringset depends only on the set of $(\leq k)$-subsequences which occur in the string.

SP_k is equivalent to $\bigwedge_{p_i \notin \mathcal{G}} [\neg p_i]$

Slide 43

Character of Piecewise Testable sets

Theorem 5 (k-Subsequence Invariance) *A stringset L is Piecewise Testable iff*

there is some k such that, for all strings x and y,

if x and y have exactly the same set of $(\leq k)$-subsequences

then either both x and y are members of L or neither is.

$$w \equiv_k^P v \overset{\text{def}}{\Longleftrightarrow} P_{\leq k}(w) = P_{\leq k}(v).$$

Slide 44

Cognitive interpretation of PT

- Any cognitive mechanism that can distinguish member strings from non-members of a (properly) PT_k stringset must be sensitive, at least, to the set of length k subsequences of events that occur in the presentation of the string—both those that do occur and those that do not.
- Any cognitive mechanism that is sensitive *only* to the set of length k subsequences of events in the presentation of a string will be able to recognize *only* PT_k stringsets.

Sequential: This corresponds to being sensitive, at each point in the string, to the set of all length k subsequences of the sequence of prior events.

Parallel: This corresponds to being sensitive to the presence of sets of patterns of ordered events in the string.

First-Order(<) definable stringsets

Slide 45

$\langle \mathcal{D}, <, P_\sigma \rangle_{\sigma\in\Sigma}$

First-order Quantification over positions in the strings

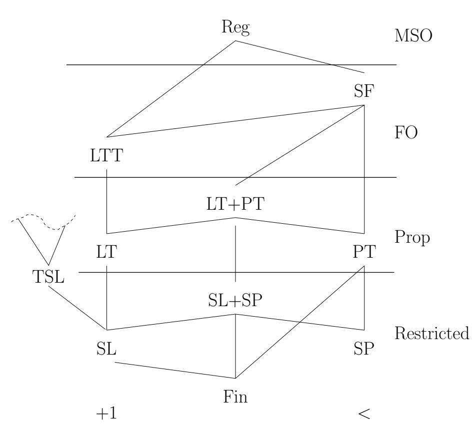
$$
\begin{array}{lrcl}
x < y & w, [x \mapsto i, y \mapsto j] \models x < y & \overset{\text{def}}{\iff} & i < j \\
P_\sigma(x) & w, [x \mapsto i] \models P_\sigma(x) & \overset{\text{def}}{\iff} & i \in P_\sigma \\
\varphi \wedge \psi & \vdots & & \\
\neg\varphi & \vdots & & \\
(\exists x)[\varphi(x)] & w, s \models (\exists x)[\varphi(x)] & \overset{\text{def}}{\iff} & w, s[x \mapsto i] \models \varphi(x)] \\
 & & & \text{for some} i \in \mathcal{D}
\end{array}
$$

$$\varphi_{\text{no-H-before-}\acute{\text{H}}} = \neg(\exists x, y)[x < y \wedge \text{H}(x) \wedge \acute{\text{H}}(y)]$$

Star-Free (SF) stringsets.

Sub-Regular Hierarchies

Slide 46



Slide 47

Yidin

- Primary stress on the leftmost heavy syllable, else the initial syllable
- Secondary stress iteratively on every second syllable in both directions from primary stress
- No light monosyllables

Explicitly:

- Exactly one $\acute{\sigma}$ (One-$\acute{\sigma}$)
- $\acute{\mathrm{L}}$ implies no $\overset{*}{\mathrm{H}}$ (No-$\overset{*}{\mathrm{H}}$-with-$\acute{\mathrm{L}}$)
- σ and $\overset{+}{\sigma}$ alternate (Alt)
- First H gets primary stress (No-$\bar{\mathrm{H}}$-before-$\acute{\mathrm{H}}$)
- $\acute{\mathrm{L}}$ only if initial (Nothing-before-$\acute{\mathrm{L}}$)
- No $\acute{\mathrm{L}}$ monosyllables (No $\rtimes\acute{\mathrm{L}}\ltimes$)

Slide 48

Yidin Constraints wrt Local Hierarchy

- One-$\acute{\sigma}$ — $(\exists! x)[\acute{\sigma}(x)]$ — ($\mathrm{LTT}_{1,2}$)
- No-$\bar{\mathrm{H}}$-before-$\acute{\mathrm{H}}$ — $\neg(\exists x, y)[x < y \wedge \bar{\mathrm{H}}(x) \wedge \acute{\mathrm{H}}(y)]$ — (SF)
- No-$\overset{*}{\mathrm{H}}$-with-$\acute{\mathrm{L}}$ — $\neg(\overset{*}{\mathrm{H}} \wedge \acute{\mathrm{L}})$ — (LT_1)
- Nothing-before-$\acute{\mathrm{L}}$ — $\neg\sigma\acute{\mathrm{L}}$ — (SL_2)
- Alt — $\neg\sigma\sigma \wedge \neg\acute{\sigma}\acute{\sigma} \wedge \neg\acute{\sigma}\grave{\sigma} \wedge \neg\grave{\sigma}\acute{\sigma} \wedge \neg\grave{\sigma}\grave{\sigma}$ — (SL_2)
- No $\rtimes\acute{\mathrm{L}}\ltimes$ — $\neg\rtimes\acute{\mathrm{L}}\ltimes$ — (SL_3)

Yidin is SF

Slide 49

Yidin constraints wrt Piecewise Hierarchy

- One-$\acute\sigma$ — $\acute\sigma \wedge \neg\acute\sigma..\acute\sigma$ — (PT_2)
- No-$\bar{\mathrm{H}}$-before-$\acute{\mathrm{H}}$ — $\neg\bar{\mathrm{H}}..\acute{\mathrm{H}}$ — (SP_2)
- No-$\overset{*}{\mathrm{H}}$-with-$\acute{\mathrm{L}}$ — $\neg\overset{*}{\mathrm{H}}..\acute{\mathrm{L}} \wedge \neg\acute{\mathrm{L}}..\overset{*}{\mathrm{H}}$ — (SP_2)
- Nothing-before-$\acute{\mathrm{L}}$ — $\neg\sigma..\acute{\mathrm{L}}$ — (SP_2)
- Alt — Not PT: $\overbrace{\sigma\grave\sigma\cdots\sigma\grave\sigma}^{2k} \equiv^{P}_{k} \overbrace{\sigma\grave\sigma\cdots\sigma\grave\sigma}^{2k}\grave\sigma$ — (SF)
- No $\rtimes\acute{\mathrm{L}}\ltimes$ — $\acute{\mathrm{L}} \rightarrow (\sigma..\acute{\mathrm{L}} \vee \acute{\mathrm{L}}..\sigma)$ — (PT_2)

Yidin is SF

Slide 50

Yidin wrt Co-occurrence of Local and Piecewise Constraints

One-$\acute\sigma$	$\mathrm{LTT}_{1,2}$	$\mathrm{PT}_{\mathbf{2}}$
Some-$\acute\sigma$	$\boxed{\mathrm{LT}_{\mathbf{1}}}$	$\boxed{\mathrm{PT}_{\mathbf{1}}}$
At-Most-One-$\acute\sigma$	$\mathrm{LTT}_{1,2}$	$\boxed{\mathrm{SP}_{\mathbf{2}}}$
No-$\bar{\mathrm{H}}$-before-$\acute{\mathrm{H}}$	SF	$\boxed{\mathrm{SP}_{\mathbf{2}}}$
No-$\overset{*}{\mathrm{H}}$-with-$\acute{\mathrm{L}}$	LT_1	$\boxed{\mathrm{SP}_{\mathbf{2}}}$
Nothing-before-$\acute{\mathrm{L}}$	$\boxed{\mathrm{SL}_{\mathbf{2}}}$	$\boxed{\mathrm{SP}_{\mathbf{2}}}$
Alt	$\boxed{\mathrm{SL}_{\mathbf{2}}}$	SF
No $\rtimes\acute{\mathrm{L}}\ltimes$	$\boxed{\mathrm{SL}_{\mathbf{3}}}$	PT_2

Yidin is co-occurrence of SL and PT constraints or of LT and SP constraints

Slide 51

co-SL Stringsets

$$L \in \text{co-SL} \quad \Leftrightarrow \overline{L} \in \text{SL}$$

SL **:** Conjunctions of negative literals

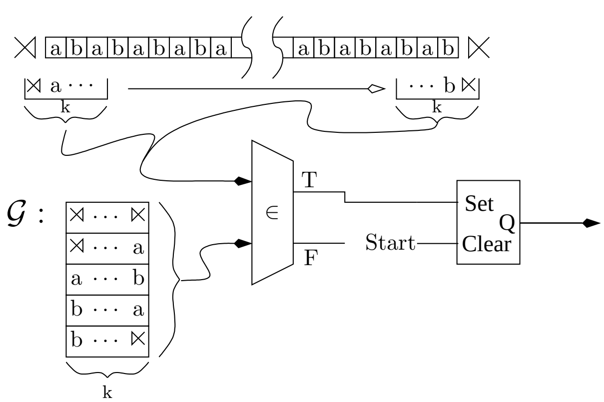
co-SL **:** Disjunctions of positive literals

$$\text{Some-}\acute{\sigma} \text{ is co-SL:} \quad \acute{\sigma}$$

Literal co-SL **:** Single positive literal

Slide 52

co-SL$_k$ Scanners



Recognizing a co-SL$_k$ stringset requires only remembering the k most recently encountered symbols.

Slide 53

Cognitive interpretation of co-SL

- Any cognitive mechanism that can distinguish member strings from non-members of a co-SL_k stringset must be sensitive, at least, to the length k blocks of consecutive events that occur in the presentation of the string.
- Any cognitive mechanism that is sensitive *only* to the *disjunctive* occurrence of length k blocks of consecutive events in the presentation of a string will be able to recognize *only* co-SL_k stringsets.

Sequential: This corresponds to being sensitive, at each point in the string, to the immediately prior sequence of $k-1$ events.

Parallel: This corresponds to being sensitive to the presence of simple contiguous blocks in the string.

Slide 54

Yidin wrt Local, co-Local and Piecewise Constraints

One-$\acute{\sigma}$	$LTT_{1,2}$	$PT_{\mathbf{2}}$
Some-$\acute{\sigma}$	$\boxed{\text{co-}SL_{\mathbf{1}}}$	$PT_{\mathbf{1}}$
At-Most-One-$\acute{\sigma}$	$LTT_{1,2}$	$\boxed{SP_{\mathbf{2}}}$
No-$\bar{\text{H}}$-before-$\acute{\text{H}}$	SF	$\boxed{SP_{\mathbf{2}}}$
No-$\overset{*}{\text{H}}$-with-$\acute{\text{L}}$	LT_1	$\boxed{SP_{\mathbf{2}}}$
Nothing-before-$\acute{\text{L}}$	$\boxed{SL_{\mathbf{2}}}$	$\boxed{SP_{\mathbf{2}}}$
Alt	$\boxed{SL_{\mathbf{2}}}$	SF
No $\rtimes\acute{\text{L}}\ltimes$	$\boxed{SL_{\mathbf{3}}}$	PT_2

Yidin is co-occurrence of SL, literal co-SL and SP constraints.

Slide 55

Stress Patterns wrt Local Constraints

- SL — 80 of 106 patterns
- LT

  None
- LTT

  Bulgarian, Lithuanian, Mam, Murik
- SF

  Amele, Arabic (Classical SPD), Bhojpuri (per Shukla Tiwari), Buriat, Cheremis (East), Cheremis (Meadow), Cheremis (Mountain), Chuvash, Golin, Hindi (per Jones), Kashmiri, Klamath, Komi, Kuuku Yau, Maori, K. Mongolian (Stuart), K. Mongolian (Bosson), Nubian, Sindhi, Yidin
- Reg

  Arabic (Negev Bedouin), Arabic (Cyrenaican Bedouin)

Slide 56

Stress Patterns wrt Piecewise Constraints

- SP

  None
- PT

  59 of 106 patterns, including:

  Abun West, Afrikaans, Agul North, Alawa, Amele, Anguthimri, Anyula, Arabic (Cairene), Arabic (Classical SPD), Arabic Damascene, ...
- SF

  All remaining patterns that are not strictly Regular
- Reg

  Arabic (Negev Bedouin), Arabic (Cyrenaican Bedouin)

Slide 57

Stress Patterns wrt Co-occurrence of Local and Piecewise Constraints

- SL + SP — 80 of 106 patterns
- SL + PT — All remaining patterns that are not strictly Regular
- LT + SP — All patterns that are SL + PT
- SF — None
- Reg
  Arabic (Negev Bedouin), Arabic (Cyrenaican Bedouin)

Slide 58

Stress Patterns wrt Co-occurrence of SL, co-SL and SP Constraints

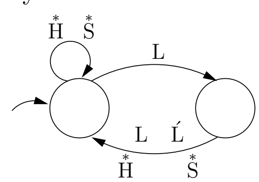
- SL + **co-**SL + SP — 98 of 106 patterns
- LT + SP — 6 patterns (2 abstract types of constraint)
- SL + PT — same 6
- SF — None
- Reg
  Arabic (Negev Bedouin), Arabic (Cyrenaican Bedouin)

Arabic (Negev Bedouin)

- In sequences of light syllables, secondary stress falls on the even numbered syllables, counting from the left edge of the sequence.
- This pattern is used only for the sake of defining main stress. Secondary stress is absent on the surface.

Slide 59 **Without reference to secondary stress**

- Odd number of unstressed light syllables precedes a light syllable with primary stress



Arabic (Negev Bedouin) with explicit secondary stress

Slide 60

$$\varphi_{\mathrm{Lalt}} = \neg\mathrm{LL} \wedge \neg\grave{\mathrm{L}}\grave{\mathrm{L}} \wedge \neg\grave{\mathrm{L}}\acute{\mathrm{L}} \wedge \neg\acute{\mathrm{L}}\grave{\mathrm{L}} \wedge \neg\overset{*}{\mathrm{H}}\mathrm{L} \wedge \neg\overset{*}{\mathrm{S}}\mathrm{L}$$

If secondary stress is explicit, then Arabic (Negev Bedouin) is LT

Slide 61

LT constraints

$$\neg(\mathrm{H} \wedge \acute{\mathrm{H}}\ltimes),\ \neg(\grave{\mathrm{H}} \wedge \acute{\mathrm{H}}\ltimes),\ \neg(\mathrm{S} \wedge \acute{\mathrm{H}}\ltimes)$$

Unifying:

$$\acute{\mathrm{H}}\ltimes \rightarrow \neg X, \quad X \in \{\mathrm{H}, \grave{\mathrm{H}}, \mathrm{S}\}$$

With culminativity, these are also PT:

$$\neg(\mathrm{H} \wedge \acute{\mathrm{H}}\ltimes) \wedge \neg\acute{\sigma}..\acute{\sigma} = (\neg\mathrm{H} \wedge \neg\acute{\sigma}..\acute{\sigma}) \vee (\acute{\mathrm{H}}..\overset{*}{\sigma} \wedge \neg\acute{\sigma}..\acute{\sigma})$$

Slide 62

Some Constraints

- Forbidden syllables (SL_1, SP_1)
  - No heavy syllables
- Required syllables (LT_1, PT_1)
  - Some primary stress
- Forbidden initial/final syllables (SL_2, SF)
  - Cannot start with unstressed light
  - Cannot start with unstressed heavy
  - Cannot end with stressed light
- Forbidden adjacent pairs (SL_2, SF)
  - No adjacent unstressed
  - No adjacent secondary stress
  - No heavy immediately following a stressed light

. . .

Slide 63

Properly Regular Constraints

- Alternation (Reg)
  - Arabic (Negev Bedouin), . . .
  - This class of constraints accounts for all properly regular stress patterns (that are known to us).