Formal Issues in the Design and Interpretation of Artificial Grammar Learning Experiments

Slide 1

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

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

This work completed, in part, while in residence at the
Radcliffe Institute for Advanced Study

Slide 2

Joint work with:

- Geoffrey K. Pullum
  School of Philosophy, Psychology and Language Sciences
  University of Edinburgh
- Marc D. Hauser
  Depts. of Psychology, Organismic & Evolutionary Biology and Biological Anthropology
  Harvard University
- Jeffery Heinz
  Dept. of Linguistics and Cognitive Science
  University of Delaware
- Gil Bailey, Matt Edlefsen, Molly Visscher, David Wellcome and Sean Wibel
  Dept. of Computer Science
  Earlham College

$(AB)^n$ ***v.s.*** A^nB^n in English

Slide 3

$\{$**(ding dong)**$^n\}$

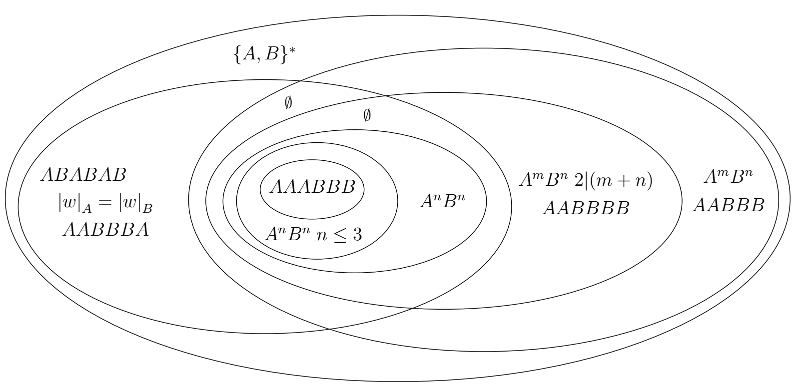
$\{$**people**^{n} **left**$^n\}$

$\{$**people (who were left by people)**^{n} **left**$\}$

$\{$**people (who were left by people)**^{2n} **left**$\}$

Stringset inference experiments

Slide 4



Formal Issues for AGL Experiments

Slide 5

Design

- Identifying relevant classes of patterns
- Finding minimal pairs of stringsets
- Finding sets of stimuli that distinguish those stringsets

Interpretation

- Identifying the class of patterns subject has generalized to
- Inferring the properties of the cognitive mechanism involved
  - properties common to all mechanisms capable of identifying that class of patterns

Assumptions

Slide 6

- Inferred set is not arbitrary
- Principle determining membership is structural
- Inference exhibits some sort of minimality

Slide 7

Dual characterizations of complexity classes

Computational classes

- Characterized by abstract computational mechanisms
- Equivalence between mechanisms
- Means to determine structural properties of stringsets

Descriptive classes

- Characterized by the nature of information about the properties of strings that determine membership
- Independent of mechanisms for recognition
- Support inference about properties of cognitive mechanisms
- Subsume wide range of types of patterns

Slide 8

Sub-regular hierarchies

- Classes of logical descriptions of string models
- Resolve finite-state into dual hierarchy of hierarchies
- Correspondence to cognitive mechanisms
- Relevant to any faculty that deals with structure of sequences of events
- Automata/Grammar-theoretic characterizations

Adjacency—Substrings

Definition 1 (k-Factor)
v is a factor *of w if $w = uvx$ for some $u, v \in \Sigma^*$.*

*v is a k-*factor *of w if it is a factor of w and $|v| = k$.*

Slide 9

$$F_k(w) \stackrel{def}{=} \begin{cases} \{v \in \Sigma^k \mid (\exists u, x \in \Sigma^*)[w = uvx]\} & \textit{if } |w| \geq k, \\ \{w\} & \textit{otherwise.} \end{cases}$$

$$ABABAB$$

$$F_2(ABABAB) = \{AB, BA\}$$
$$F_7(ABABAB) = \{ABABAB\}$$

Strictly Local Stringsets—SL

Strictly k-Local Definitions

Slide 10

$$\mathcal{G} \subseteq F_k(\{\rtimes\} \cdot \Sigma^* \cdot \{\ltimes\})$$
$$w \models \mathcal{G} \stackrel{\text{def}}{\iff} F_k(\rtimes \cdot w \cdot \ltimes) \subseteq \mathcal{G}$$
$$L(\mathcal{G}) \stackrel{\text{def}}{=} \{w \mid w \models \mathcal{G}\}$$

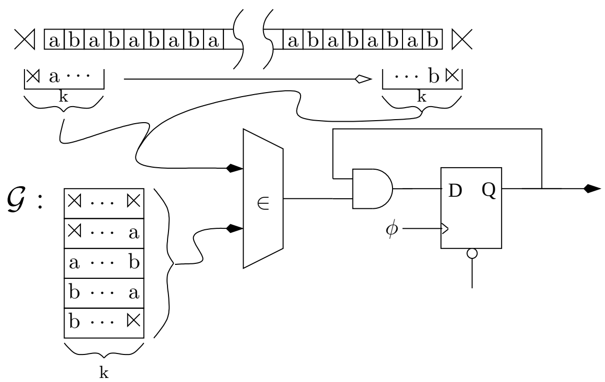
$$\mathcal{G}_{(AB)^n} = \{\rtimes A, AB, BA, B\ltimes\}$$

$$\rtimes ABABAB \ltimes \quad \rtimes A\overset{*}{BB}AB\ltimes$$

Membership in an SL_k stringset depends only on the individual k-factors which occur in the string.

Scanners

Slide 11



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

Character of Strictly k-Local Sets

Slide 12

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

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

Example:

The dog	·	likes	·	the biscuit	$\in L$
Alice	·	likes	·	Bob	$\in L$
The dog	·	likes	·	Bob	$\in L$

SL Hierarchy

Slide 13

Definition 2 (*SL*)
A stringset is Strictly k-Local *if it is definable with an* SL_k *definition.*

A stringset is Strictly Local *(in SL) if it is* SL_k *for some* k*.*

Theorem 1 (SL-Hierarchy)

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

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

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

Cognitive interpretation of SL

Slide 14

- Any cognitive mechanism that can distinguish member strings from non-members of an SL_k stringset must be sensitive, at least, to the length k blocks of events that occur in the presentation of the string.
- If the strings are presented as sequences of events in time, then this corresponds to being sensitive, at each point in the string, to the immediately prior sequence of $k - 1$ events.
- Any cognitive mechanism that is sensitive *only* to the length k blocks of events in the presentation of a string will be able to recognize *only* SL_k stringsets.

Probing the SL boundary

Slide 15

$$(AB)^n = L(\{\rtimes A, AB, BA, B\ltimes\}) \in \text{SL}_2$$

$$\text{Some-B} \stackrel{\text{def}}{=} \{w \in \{A,B\}^* \mid |w|_B \geq 1\} \quad \notin \text{SL}$$

$$\begin{array}{ccccccc} A \ldots A & \cdot & \underbrace{A \ldots A}_{k-1} & \cdot & BA \ldots A & \in \text{Some-}B \\ A \ldots AB & \cdot & \underbrace{A \ldots A}_{k-1} & \cdot & A \ldots A & \in \text{Some-}B \\ \hline A \ldots A & \cdot & \underbrace{A \ldots A}_{k-1} & \cdot & A \ldots A & \notin \text{Some-}B \end{array}$$

		In	Out
SL	$(AB)^n$	$(AB)^{i+j+1}$	$(AB)^i AA (AB)^j$
	$A^m B^n$	$A^{i+k} B^{j+l}$	$A^i B^j A^k B^l$
non-SL	Some-B	$A^i B A^j$	A^{i+j+1}

Locally k-Testable Stringsets

Slide 16

k-Expressions

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

Locally k-Testable Languages (LT_k):

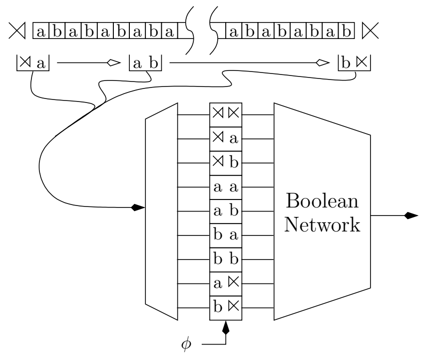
$$L(\varphi) \stackrel{\text{def}}{=} \{w \in \Sigma^* \mid w \models \varphi\}$$

$$\text{Some-}B = L(\rtimes B \vee AB) \qquad (= L(\neg(\neg \rtimes B \wedge \neg AB)))$$

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

LT Automata

Slide 17



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

Character of Locally Testable sets

Slide 18

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.*

$$w \equiv_k^L v \overset{\text{def}}{\Longleftrightarrow} F_k(\rtimes w \ltimes) = F_k(\rtimes v \ltimes).$$

$$\text{Some-}B = \bigcup \{[w]_2^L \mid w \in \{A, B\}^*,\ |w|_B \geq 1 \text{ and } |w| \leq 6\}.$$

LT Hierarchy

Slide 19

Definition 3 (LT)
*A stringset is k-*Locally Testable *if it is definable with an LT_k-expression.*

A stringset is Locally Testable *(in LT) if it is LT_k for some k.*

Theorem 3 (LT-Hierarchy)

$$LT_2 \subsetneq LT_3 \subsetneq \cdots \subsetneq LT_i \subsetneq LT_{i+1} \subsetneq \cdots \subsetneq LT$$

Cognitive interpretation of LT

Slide 20

- Any cognitive mechanism that can distinguish member strings from non-members of an LT_k stringset must be sensitive, at least, to the set of length k blocks of events that occur in the presentation of the string—both those that do occur and those that do not.
- If the strings are presented as sequences of events in time, then this corresponds to being sensitive, at each point in the string, to the length k blocks of events that occur at any prior point.
- Any cognitive mechanism that is sensitive *only* to the set of length k blocks of events in the presentation of a string will be able to recognize *only* LT_k stringsets.

Probing the LT boundary

Slide 21

$$\text{Some-B} = L(\rtimes B \vee AB) \quad \in \text{LT}_2$$

$$\text{One-B} \stackrel{\text{def}}{=} \{w \in \{A,B\}^* \mid |w|_B = 1\} \quad \notin \text{LT}$$

$$A^kBA^k \in \text{One-B} \qquad A^kBA^kBA^k \notin \text{One-B}$$

$$F_k(\rtimes A^kBA^k\ltimes) = F_k(\rtimes A^kBA^kBA^k\ltimes)$$

		In	Out
LT	Some-B	A^iBA^j	A^{i+j+1}
non-LT	One-B	A^iBA^{j+k+1}	$A^iBA^jBA^k$

FO(+1) (Strings)

Slide 22

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

$$AABA = \Big\langle \{0,1,2,3\},\ \{\langle i, i+1\rangle \mid 0 \le i < 3\},\ \{0,1,3\}_A,\ \{2\}_B \Big\rangle$$

First-order Quantification (over positions in the strings)

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

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

$$\text{One-}B = L((\exists x)[B(x) \wedge (\forall y)[B(y) \rightarrow x \approx y]\,])$$

Slide 23

Character of the FO(+1) Definable Stringsets

Definition 4 (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 4 (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*.*

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 24

Cognitive interpretation of FO(+1)

- Any cognitive mechanism that can distinguish member strings from non-members of an 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.
- If the strings are presented as sequences of events in time, then this corresponds to being able count up to some fixed threshold.
- 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.

Probing the FO(+1) boundary

$$\text{One-}B = L((\exists x)[B(x) \wedge (\forall y)[B(y) \rightarrow x \approx y]]) \ \in \text{LTT}$$

$$\text{No-}B\text{-after-}C \stackrel{\text{def}}{=} \{w \in \{A, B, C\}^* \mid \text{no B follows any C}\} \quad \notin \text{LTT}$$

$A^kBA^kCA^k$ and $A^kCA^kBA^k$ have exactly the same number of occurrences of every k-factor.

Slide 25

		In	Out
FO(+1)	One-B	A^iBA^{j+k+1}	$A^iBA^jBA^k$
non-FO(+1)	No-B-after-C	$A^iBA^jCA^k$	$A^iCA^jBA^k$
		$A^iBA^jBA^k$	
		$A^iCA^jCA^k$	

Precedence—Subsequences

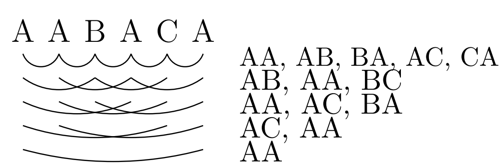
Definition 5 (Subsequences)

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

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

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

Slide 26



$$P_2(AABACA) = \{AA, AB, AC, BA, BC, CA\}$$

$$P_{\leq 2}(AABACA) = \{\varepsilon, A, B, C, AA, AB, AC, BA, BC, CA\}$$

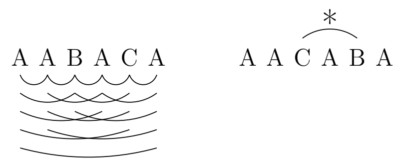
Strictly Piecewise Stringsets—SP

Strictly k-Piecewise Definitions

$$\mathcal{G} \subseteq \Sigma^{\leq k}$$
$$w \models \mathcal{G} \stackrel{\text{def}}{\Longleftrightarrow} P_{\leq k}(w) \subseteq P_{\leq k}(\mathcal{G})$$
$$L(\mathcal{G}) \stackrel{\text{def}}{=} \{w \in \Sigma^* \mid w \models \mathcal{G}\}$$

Slide 27

$$\mathcal{G}_{\text{No-B-after-C}} = \{AA, AB, AC, BA, BB, BC, CA, CC\}$$



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

Character of the Strictly k-Piecewise Sets

Theorem 5 *A stringset L is Strictly k-Piecewise Testable iff, for all $w \in \Sigma^*$,*

$$P_{\leq k}(w) \subseteq P_{\leq k}(L) \Rightarrow w \in L$$

Slide 28

Consequences:

Prefix & Suffix Closure:	$wv \in L \Rightarrow w, v \in L$
Subsequence Closure:	$w\sigma v \in L \Rightarrow wv \in L$
Unit Strings:	$P_1(L) \subseteq L$
Empty String:	$L \neq \emptyset \Rightarrow \varepsilon \in L$

SP Hierarchy

Definition 6 (SP)
A stringset is Strictly k-Piecewise *if it is definable with an SP_k definition.*

A stringset is Strictly Piecewise *(in SP) if it is SP_k for some k.*

Theorem 6 (SP-Hierarchy)

Slide 29

$$SP_2 \subsetneq SP_3 \subsetneq \cdots \subsetneq SP_i \subsetneq SP_{i+1} \subsetneq \cdots \subsetneq SP$$

SP is incomparable (wrt subset) with the Local Hierarchy

$\mathrm{SP}_2 \not\subseteq \mathrm{FO}(+1)$	No-B-after-$C \in \mathrm{SP}_2 - \mathrm{FO}(+1)$
$\mathrm{SL}_2 \not\subseteq \mathrm{SP}$	$(AB)^n \in \mathrm{SL}_2 - \mathrm{SP}$
$\mathrm{SP}_2 \cap \mathrm{SL}_2 \neq \emptyset$	$A^m B^n \in \mathrm{SP}_2 \cap \mathrm{SL}_2$
$\mathrm{Fin} \not\subseteq \mathrm{SP}$	$\{A\} \in \mathrm{Fin} - \mathrm{SP}$

Cognitive interpretation of SP

Slide 30

- Any cognitive mechanism that can distinguish member strings from non-members of an 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.
- If the strings are presented as sequences of events in time, then this corresponds to being sensitive, at each point in the string, to up to $k-1$ events distributed arbitrarily among the prior events.
- 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.

Probing the SP boundary

$$\text{No-}B\text{-after-}C \in \text{SP}_2$$

$$B\text{-before-}C \stackrel{\text{def}}{=} \{w \in \Sigma^* \mid \text{Some B occurs prior to any C}\} \quad \notin \text{SP}$$

$$AABACA \in \text{B-before-C}, \quad AACA \sqsubseteq AABACA, \quad AACA \notin \text{B-before-C}$$

Slide 31

		In	Out
SP	No-B-after-C	$A^iBA^jCA^k$	$A^iCA^jBA^k$
		$A^iBA^jBA^k$	
		$A^iCA^jCA^k$	
	A^mB^n	$A^{i+k}B^{j+l}$	$A^iB^jA^kB^l$
non-SP	B-before-C	$A^iBA^jCA^k$	$A^iCA^jBA^k$
			$A^iCA^jCA^k$
	$(AB)^n$	$(AB)^{i+j+1}$	$(AB)^iAA(AB)^j$

k-Piecewise Testable Stringsets

PT$_k$-expressions

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

Slide 32

k-Piecewise Testable Languages (PT$_k$):

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

$$B\text{-before-}C = L(\neg C \vee BC) \qquad (= L(C \rightarrow BC))$$

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

Character of Piecewise Testable sets

Slide 33

Theorem 7 (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.

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

$$B\text{-before-}C = \bigcup\{[w]_2^P \mid w \in \{A, B\}^*,\ w \models (C \rightarrow BC) \text{ and } |w| \leq 6\}.$$

PT Hierarchy

Slide 34

Definition 7 (SP)
*A stringset is k-*Piecewise Testable *if it is definable with an PT_k definition.*

A stringset is Piecewise Testable *(in PT) if it is PT_k for some k.*

Theorem 8 (PT-Hierarchy)

$$PT_2 \subsetneq PT_3 \subsetneq \cdots \subsetneq PT_i \subsetneq PT_{i+1} \subsetneq \cdots \subsetneq PT$$

PT, SP and the Local Hierarchy

Slide 35

$$
\begin{array}{rl}
\mathrm{SP}_k \subsetneq \mathrm{PT}_k & \\
\mathrm{SP}_{k+1} \not\subseteq \mathrm{PT}_k & \\
\mathrm{PT}_2 \not\subseteq \mathrm{SP} & B\text{-before-}C, \text{One-}B \in \mathrm{PT}_2 - \mathrm{SP} \\
\mathrm{PT}_2 \not\subseteq \mathrm{FO}(+1) & \text{No-}B\text{-after-}C \in \mathrm{PT}_2 - \mathrm{FO}(+1) \\
\mathrm{SL}_2 \not\subseteq \mathrm{PT} & (AB)^n \in \mathrm{SL}_2 - \mathrm{PT} \\
\mathrm{PT}_2 \cap \mathrm{SL}_2 \neq \emptyset & A^m B^n \in \mathrm{PT}_2 \cap \mathrm{SL}_2
\end{array}
$$

Fin $\subseteq$ SP :

$$\Sigma^* = L(\varepsilon), \quad \emptyset = L(\neg\varepsilon), \quad \{\varepsilon\} = L(\bigwedge_{\sigma\in\Sigma}[\neg\sigma]),$$

$$\{w\} = L(w \wedge \bigwedge_{p\in\Sigma^{|w|+1}} [\neg p])$$

$$\{w_1, \ldots, w_n\} = L(\bigvee_{1\leq i\leq n} [w_i \wedge \bigwedge_{p\in\Sigma^{|w_i|+1}} [\neg p]])$$

Cognitive interpretation of PT

Slide 36

- Any cognitive mechanism that can distinguish member strings from non-members of an 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.
- If the strings are presented as sequences of events in time, then 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.
- 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.

Probing the PT boundary

Slide 37

$$B\text{-before-}C, \text{One-}B \in \mathrm{PT}_2$$

$$(AB)^n \notin \mathrm{PT}$$

$$(AB)^k \in (AB)^n \qquad\qquad (AB)^k A \notin (AB)^n$$

$$P_k((AB)^k A) = P_k((AB)^k)$$

		In	Out
PT	B-before-C	$A^iBA^jCA^k$	$A^iCA^jBA^k$
			$A^iCA^jCA^k$
	One-B	A^iBA^{j+k+1}	$A^iBA^jBA^k$
non-PT	$(AB)^n$	$(AB)^{i+j+1}$	$(AB)^iAA(AB)^j$

First-Order(<) definable stringsets

Slide 38

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

First-order Quantification over positions in the strings

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

PT, FO(+1) and FO(<)

Slide 39

Theorem 9 *$PT \subsetneq FO(<)$.*

$$\sigma_1 \cdots \sigma_n \sqsubseteq w \Leftrightarrow (\exists x_1, \ldots, x_n)[\bigwedge_{1 \leq i < j \leq n} [x_i \triangleleft^+ x_j] \wedge \bigwedge_{1 \leq i \leq n} [P_{\sigma_i}(x_i)]$$

Theorem 10 *$FO(+1) \subsetneq FO(<)$.*

+1 is FO definable from <:

$$x \triangleleft y \equiv x \triangleleft^+ y \wedge \neg(\exists z)[x \triangleleft^+ z \wedge z \triangleleft^+ y]$$

$$\text{No-}B\text{-after-}C \subseteq \text{FO}(<) - \text{FO}(+1)$$

$$(AB)^n \subseteq \text{FO}(<) - \text{PT}$$

Star-Free stringsets

Slide 40

Definition 8 (Star-Free Set) *The class of* Star-Free Sets *(SF) is the smallest class of languages satisfying:*

- *$Fin \subseteq SF$.*
- *If $L_1, L_2 \in SF$ then:* $L_1 \cdot L_2 \in SF$, $L_1 \cup L_2 \in SF$, $\overline{L_1} \in SF$.

Theorem 11 (McNauthton and Papert) *A set of strings is First-order definable over $\langle \mathcal{D}, \triangleleft^+, P_\sigma \rangle_{\sigma \in \Sigma}$ iff it is* Star-Free.

PT and LT with Order

$$\varphi \bullet \psi \qquad w \models \varphi \bullet \psi \overset{\text{def}}{\iff} w = w_1 \cdot w_2, \quad w_1 \models \varphi \text{ and } w_2 \models \psi.$$

LTO_k is LT_k plus $\varphi \bullet \psi$

$$\text{No-}B\text{-after-}C = L((\neg C) \bullet (\neg B)) \in \mathrm{LTO}$$

Slide 41

PTO_k is PT_k plus $\varphi \bullet \psi$
Let:

$$\varphi_{A^{=i}} = A^i \wedge \bigwedge_{p \in \Sigma^{i+1}} [\neg p], \qquad \varphi_{\Sigma^*} = \varepsilon$$

Then:

$$\begin{aligned}(AB)^n \quad = \quad & L(\neg(\varphi_{B^{=1}} \bullet \varphi_{\Sigma^*}) \wedge \neg(\varphi_{\Sigma^*} \bullet \varphi_{A^{=1}}) \wedge \\ & \neg(\varphi_{\Sigma^*} \bullet \varphi_{A^{=2}} \bullet \varphi_{\Sigma^*}) \wedge \neg(\varphi_{\Sigma^*} \bullet \varphi_{B^{=2}} \bullet \varphi_{\Sigma^*})) \quad \in \mathrm{PTO}\end{aligned}$$

PTO, LTO and SF

Theorem 12

$$PTO = SF = LTO$$

Slide 42

SF $\subseteq$ PTO, SF $\subseteq$ LTO

Fin $\subseteq$ PTO, Fin $\subseteq$ LTO and both are closed under concatenation, union and complement.

LTO $\subseteq$ PTO $\subseteq$ SF

Concatenation is FO($<$) definable.

Character of FO(<) definable sets

Theorem 13 (McNaughton and Papert) *A stringset L is definable by a set of First-Order formulae over strings iff it is recognized by a finite-state automaton that is* non-counting *(that has an* aperiodic *syntactic monoid), that is, iff:*

Slide 43

there exists some $n > 0$ such that

for all strings u, v, w over Σ

if $uv^n w$ occurs in L

then $uv^{n+i}w$, for all $i \geq 1$, occurs in L as well.

E.g.

$$\frac{\textbf{people who were left (by people who were left)}^{n}\ \textbf{left} \quad \in L}{\textbf{people who were left (by people who were left)}^{n+1}\ \textbf{left} \quad \in L}$$

Cognitive interpretation of FO(<)

Slide 44

- Any cognitive mechanism that can distinguish member strings from non-members of an FO(<) stringset must be sensitive, at least, to the sets of length k blocks of events, for some fixed k, that occur in the presentation of the string when it is factored into segments, up to some fixed number, on the basis of those sets with distinct criteria applying to each segment.
- If the strings are presented as sequences of events in time, then this corresponds to being able to count up to some fixed threshold with the counters being reset some fixed number of times based on those counts.
- Any cognitive mechanism that is sensitive *only* to the sets of length k blocks of events in the presentation of a string once it has been factored in this way will be able to recognize *only* FO(<) stringsets.

Probing the FO(<) boundary

$$BB\text{-before-}C \in \mathrm{FO}(<)$$

$$\text{Even-}B \stackrel{\text{def}}{=} \{w \in \{A,B\}^* \mid |w|_B = 2i,\ 0 \le i\} \notin \mathrm{FO}(<)$$

Slide 45

$$A^iB^nB^n \in \text{Even-}B \quad \text{but} \quad A^iB^{n+1}B^n \notin \text{Even-}B$$

		In	Out
FO(<)	BB-before-C	$A^iBBA^{j+k}CA^l$	$A^iCA^{j+k}BBA^l$
			$A^iBA^jBA^kCA^l$
non-FO(<)	Even-B	B^{2i}	B^{2i+1}

MSO definable stringsets

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

Slide 46

First-order Quantification (positions)

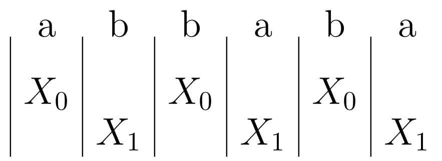
Monadic Second-order Quantification (sets of positions)

$\triangleleft^+$ is MSO-definable from $\triangleleft$.

MSO example

Slide 47

$$
\begin{aligned}
(\exists X_0, X_1)[\quad & (\forall x)[(\exists y)[y \triangleleft x] \vee X_0(x)] \wedge \\
& (\forall x, y)[\neg(X_0(x) \wedge X_1(x))] \wedge \\
& (\forall x, y)[x \triangleleft y \rightarrow (X_0(x) \leftrightarrow X_1(y)] \wedge \\
& (\forall x)[(\exists y)[x \triangleleft y] \vee X_1(x)] \qquad\qquad]
\end{aligned}
$$



Slide 48

Theorem 14 (Chomsky Schützenberger) *A set of strings is Regular iff it is a homomorphic image of a Strictly 2-Local set.*

Definition 9 (Nerode Equivalence) *Two strings w and v are* Nerode Equivalent *with respect to a stringset L over Σ (denoted $w \equiv_L v$) iff for all strings u over Σ, $wu \in L \Leftrightarrow vu \in L$.*

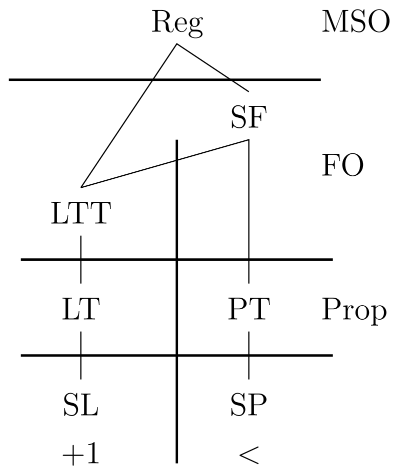
Theorem 15 (Myhill-Nerode) *A stringset L is recognizable by a FSA (over strings) iff $\equiv_L$ partitions the set of all strings over Σ into finitely many equivalence classes.*

Theorem 16 (Medvedev, Büchi, Elgot) *A set of strings is MSO-definable over $\langle \mathcal{D}, \triangleleft, \triangleleft^+, P_\sigma \rangle_{\sigma \in \Sigma}$ iff it is regular.*

Theorem 17 *MSO = ∃MSO over strings.*

Local and Piecewise Hierarchies

Slide 49



Cognitive interpretation of Finite-state

Slide 50

- Any cognitive mechanism that can distinguish member strings from non-members of a finite-state stringset must be capable of classifying the events in the input into a finite set of abstract categories and are sensitive to the sequence of those categories.
- Subsumes *any* recognition mechanism in which the amount of information inferred or retained is limited by a fixed finite bound.
- Any cognitive mechanism that has a fixed finite bound on the amount of information inferred or retained in processing sequences of events will be able to recognize *only* finite-state stringsets.

Slide 51

Probing the FS boundary

$$\text{Even-B} \stackrel{\text{def}}{=} \{w \in \{A,B\}^* \mid |w|_B = 2i,\, 0 \le i\} \in \text{FS}$$

$$\{A^n B^n \mid n > 0\} \notin \text{FS}$$

$$w \equiv_{A^nB^n} v \quad \Leftrightarrow \quad w, v \notin \{A^i B^j \mid i, j \ge 0\} \text{ or } |w|_A - |w|_B = |v|_A - |v|_B\,.$$

		In	Out
FS	Even-B	B^{2i}	B^{2i+1}
non-FS	A^nB^n	A^nB^n	$A^{n-1}B^{n+1}$

Slide 52

Non-FS classes

Additional structure — not finitely bounded

A^nB^n

$D_1 = |w|_A = |w|_B$, properly nested

$D_2 = |w|_A = |w|_B$ and $|w|_C = |w|_D$, properly nested.

Subregular Hierarchy over Trees

$$CFG = SL_2 \ < \ LT \ < \ FO(+1) \ < \ FO(<) \ < \ MSO = \text{FSTA}$$

Slide 53

FLT support for AGL experiments

Model-theoretic characterizations

- very general methods for describing patterns
- provide clues to nature of cognitive mechanisms
- independent of *a priori* assumptions

Grammar- and Automata-theoretic characterizations

- provide information about nature of stringsets
- minimal pairs

Sub-regular hierarchies

- broad range of capabilities weaker than human capabilities
- characterizations in terms of plausible cognitive attributes
- relevant as long as generalizations are based on structure of strings

References

Beauquier, D., and Jean-Eric Pin. 1991. Languages and scanners. *Theoretical Computer Science* 84:3–21.

Büchi, J. Richard. 1960. Weak second-order arithmetic and finite automata. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 6:66–92.

Chomsky, Noam, and M. P. Schützenberger. 1963. The algebraic theory of context-free languages. In *Computer programming and formal systems*, ed. P. Braffort and D. Hirschberg, Studies in Logic and the Foundations of Mathematics, 118–161. Amsterdam: North-Holland, 2nd (1967) edition.

Elgot, Calvin C. 1961. Decision problems of finite automata and related arithmetics. *Transactions of the American Mathematical Society* 98:21–51.

García, Pedro, and José Ruiz. 1996. Learning k-piecewise testable languages from positive data. In *Grammatical Interference: Learning Syntax from Sentences*, ed. Laurent Miclet and Colin de la Higuera, volume 1147 of *Lecture Notes in Computer Science*, 203–210. Springer.

Heinz, Jeffrey. 2007. The inductive learning of phonotactic patterns. Doctoral Dissertation, University of California, Los Angeles.

Heinz, Jeffrey. 2008. Learning long distance phonotactics. Submitted manuscipt.

Heinz, Jeffrey. to appear. On the role of locality in learning stress patterns. *Phonology* .

Kontorovich, Leonid, Corinna Cortes, and Mehryar Mohri. 2006. Learning linearly separable languages. In *The 17th International Conference on Algorithmic Learning Theory (ALT 2006)*, volume 4264 of Lecture Notes in Computer Science, 288–303. Springer, Heidelberg, Germany.

Lothaire, M., ed. 2005. *Applied combinatorics on words*. Cambridge University Press, 2nd edition.

McNaughton, R., and S. Papert. 1971. *Counter-free automata*. MIT Press.

Perrin, Dominique, and Jean-Eric Pin. 1986. First-Order logic and Star-Free sets. *Journal of Computer and System Sciences* 32:393–406.

Rogers, James. 2003. wMSO theories as grammar formalisms. *Theoretical Computer Science* 293:291–320.

Rogers, James, Jeffery Heinz, Gil Bailey, Matt Edlefsen, Molly Visscher, David Wellcome, and Sean Wibel. 2009. On languages piecewise testable in the strict sense. In *Preproceedings of 11th Meeting on Mathematics of Language*. Bielefeld, Germany. To Appear.

Rogers, James, and Geoffrey Pullum. 2007. Aural pattern recognition experiments and the subregular hierarchy. In *Proceedings of 10th Mathematics of Language Conference*, ed. Marcus Kracht, 1–7. University of California, Los Angeles.

Simon, Imre. 1975. Piecewise testable events. In *Automata Theory and Formal Languages: 2nd Grammatical Inference conference*, 214–222. Berlin ; New York: Springer-Verlag.

Thomas, Wolfgang. 1982. Classifying regular events in symbolic logic. *Journal of Computer and Systems Sciences* 25:360–376.