Model-Theory of the Subregular

James Rogers
Dept. of Computer Science
Earlham College
and
Dept. of Linguistics and Cognitive Science
University of Delaware

`jrogers@cs.earlham.edu`

Slide 1

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

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

Slide 2

Finite and Co-Finite stringsets

Definition 1 (Finite Stringsets (Fin)) *For any alphabet* Σ*:*

- $\emptyset$ *is a finite stringset over* Σ*,*
- *The singleton set* $\{\varepsilon\}$ *is a finite stringset over* Σ*,*
- *For each* $\sigma \in \Sigma$*, the singleton set* $\{\sigma\}$ *is a finite stringset over* Σ*,*
- *If* L_1 *and* L_2 *are finite stringsets over* Σ *then:*
  - $L_1 \cdot L_2$ *is a finite stringset over* Σ*,*
  - $L_1 \cup L_2$ *is a finite stringset over* Σ*.*
- *Nothing else is a finite stringset over* Σ*.*

Definition 2 (Co-Finite (CoFin))

$$L \in \mathit{CoFin} \overset{def}{\Longleftrightarrow} L = \Sigma^* - F,\ F \in \mathit{Fin}.$$

Star-free stringsets

Slide 3

Definition 3 (Star-free Stringsets (SF)) *For any alphabet* Σ*:*

- $Fin \subset SF$,
- *If* $L_1, L_2 \in SF$ *then:*
  - $L_1 \cdot L_2 \in SF$
  - $L_1 \cup L_2 \in SF$
  - $\Sigma^* - L_1 \in SF$
- *Nothing else is a star-free stringset over* Σ.

Word models

Slide 4

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

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

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

$$\begin{array}{rcccccc} CCVC = & \langle & \{0,1,2,3\}, & \{\langle i, i+1\rangle \mid 0 \leq i < 3\}, & \{0,1,3\}_C, & \{2\}_V & \rangle \\ & \langle & \mathcal{D} & \triangleleft & P_C & P_V & \rangle \end{array}$$

Slide 5

Thomas's Word models [Thomas'82]

$$\langle \mathcal{D}, <, \min, \max, S, P, Q_\sigma \rangle_{\sigma \in \Sigma}$$

$$\begin{array}{rcl} \mathcal{D} & \text{—} & \{0, 1, \ldots n\} \\ < & \text{—} & \text{Linear order on } \mathcal{D} \\ \min \in D & \text{—} & \text{Minimum elelment of} D \\ \max \in D & \text{—} & \text{Maximum elelment of} D \\ S(\tau) : D \to D & \text{—} & \text{Successor wrt } <,\ S(\max) = \max \\ P(\tau) : D \to D & \text{—} & \text{Predecessor wrt } <,\ P(\min) = \min \\ Q_\sigma \subseteq D & \text{—} & \text{Subset of } \mathcal{D} \text{ at which } \sigma \text{ occurs} \\ & & (Q_\sigma \text{ partition } \mathcal{D}) \end{array}$$

Slide 6

First-order formulae over word models

Definition 4 ($L^1(\Sigma)$)
$\mathbb{X}_0 = \{x_0, x_1, \ldots\}$, a countably infinite set of position variables.

1. *(**Atomic formulae**)*
   (a) $x \triangleleft y \in L^1$.
   (b) $x \triangleleft^+ y \in L^1$.
   (c) *If $x, y \in \mathbb{X}_0$ then '$x \approx y$' $\in L^1(\Sigma)$.*
2. *(**Truth functional connectives**) If $\varphi, \psi \in L^1(\Sigma)$ then:*
   (a) '$(\varphi \vee \psi)$' $\in L^1(\Sigma)$ *(**disjunction**),*
   (b) '$(\neg\varphi)$' $\in L^1(\Sigma)$ *(**negation**)*
3. *(**First-Order quantifiers**) If $\varphi \in L^1(\Sigma)$ and $x \in \mathbb{X}_0$ then*
   (a) '$(\exists x)[\varphi]$' $\in L^1(\Sigma)$ *(**existential quantification**)*

Defined connectives

Slide 7

1. $(\varphi \wedge \psi) \equiv (\neg((\neg\varphi) \vee (\neg\psi)))$ (**conjunction**),
2. $(\varphi \rightarrow \psi) \equiv ((\neg\varphi) \vee \psi))$ (**implication**),
3. $(\varphi \leftrightarrow \psi) \equiv ((\varphi \wedge \psi) \vee ((\neg\varphi) \wedge (\neg\psi)))$ (**bi-conditional**), ...

Defined quantifiers

1. $(\forall x)[\varphi] \equiv (\neg(\exists x)[\neg\varphi])$ (**universal quantification**).

FO assignment

Slide 8

Definition 5 *An* ***assignment*** *s for a model $\mathcal{W}$ is a partial function from $\mathbb{X}_0$ to the domain of $\mathcal{W}$. The empty assignment is not defined for any variable. If s is an assignment, $x \in \mathbb{X}_0$ and a in the domain of $\mathcal{W}$, then*

$$s[x \mapsto a](y) \stackrel{def}{=} \begin{cases} a & \text{if } y = x, \\ s(y) & \text{otherwise.} \end{cases}$$

FO satisfaction

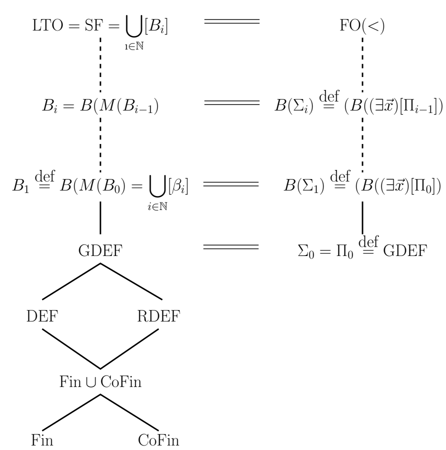
Slide 9

Definition 6 *An assignment s **satisfies** a formula φ in a model $\mathcal{W}$, denoted $\mathcal{W}, s \models \varphi$, iff one of the following holds:*

- $\varphi = x \triangleleft y$, *$s(x)$ and $s(y)$ are both defined and* $\langle s(x), s(y) \rangle \in \triangleleft^{\mathcal{W}}$.
- $\varphi = x \triangleleft^{+} y$, *$s(x)$ and $s(y)$ are both defined and* $\langle s(x), s(y) \rangle \in \triangleleft^{+\mathcal{W}}$.
- $\varphi = P_\sigma(x)$, *$s(x)$ is defined and* $s(x) \in P_\sigma^{\mathcal{W}}$.
- $\varphi =$ '$x \approx y$', *$s(x)$ and $s(y)$ are both defined and* $s(x) = s(y)$,
- $\varphi =$ '$(\psi_1 \vee \psi_2)$' *and either* $\mathcal{W}, s \models \psi_1$ *or* $\mathcal{W}, s \models \psi_2$,
- $\varphi =$ '$(\neg\psi)$' *and* $\mathcal{W}, s \not\models \psi$, $\mathcal{W}, s[x \mapsto a] \models \psi$, *or*
- $\varphi =$ '$(\exists x)[\psi]$' *and, for some a in the domain of $\mathcal{W}$,* $\mathcal{W}, s[x \mapsto a] \models \psi$

Dot Depth Hierarchy [Cohen & Brzozowski'71] Quantifier Alternation Hierarchy [Thomas'82]

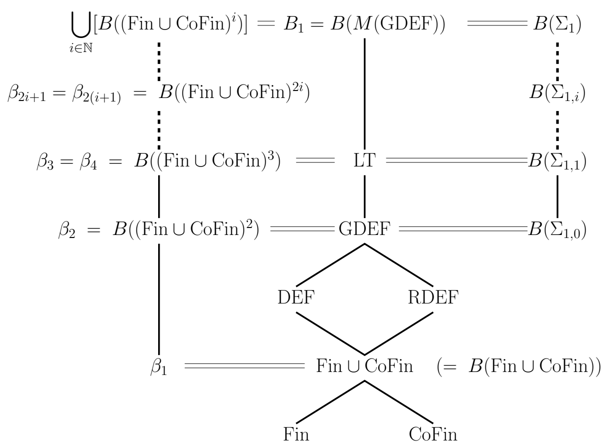
Slide 10



$B_i \stackrel{\mathrm{def}}{=} B(M(B_{i-1}))$ $\quad \Sigma_0 = \Pi_0 \stackrel{\mathrm{def}}{=} \text{q.f.}$, $\Pi_i \stackrel{\mathrm{def}}{=} \neg\Sigma_i$, $\Sigma_i \stackrel{\mathrm{def}}{=} (\exists\vec{x})[\Pi_{i-1}]$

Slide 11

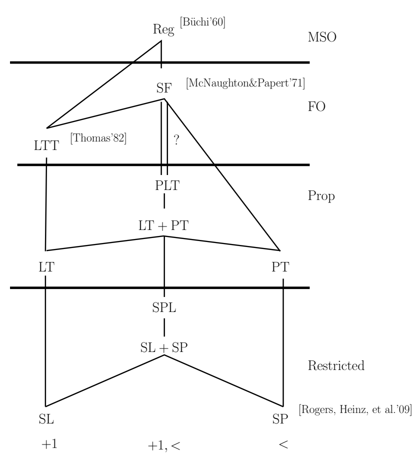
Beta Hierarchy [Brzozowski & Simon'73]
Quantifier-Rank Hierarchy [Thomas'82]



$\beta_i \stackrel{\text{def}}{=} B((\mathrm{Fin}\cup\mathrm{CoFin})^i)$ $\qquad$ $\Sigma_{1,r} \stackrel{\text{def}}{=} \mathrm{FO}(<)$, quantifier rank $= r$

Slide 12

Local Hierarchy [McNaughton & Papert'71]
Piecewise Hierarchy [Simon'75, Rogers, Heinz, et al.'09]



Definition 7 (k-Factors, F_k) *Let $\boldsymbol{F}_k(w)$ denote the set of length k sequences of adjacent symbols that occur in w. If $|w| \leq k$ then $F_k(w)$ is just the (single) sequence of symbols in w.*

$$F_k(L) \stackrel{def}{=} \{F_k(w) \mid w \in L\}.$$

Similarly

Slide 13

$$F_{<k}(w) \stackrel{def}{=} \bigcup_{2<i<k} [F_i(w)]$$

etc.

$$\begin{aligned}
\mathrm{F}_2(\rtimes abab\ltimes) &= \{\rtimes a, ab, ba, b\ltimes\} \\
\mathrm{F}_3(\rtimes abab\ltimes) &= \{\rtimes ab, aba, bab, ab\ltimes\} \\
\mathrm{F}_7(\rtimes abab\ltimes) &= \{\rtimes abab\ltimes\} \\
\mathrm{F}_{\leq 3}(\rtimes abab\ltimes) &= \mathrm{F}_2(\rtimes abab\ltimes) \cup \{\rtimes ab, aba, bab, ab\ltimes\}.
\end{aligned}$$

k-expressions

Definition 8 (k-expressions)

Slide 14

- *If $\varphi = v \in F_{\leq k}(\{\rtimes\} \cdot \Sigma^* \cdot \{\ltimes\})$ then φ is a k-expression over Σ.*
- *If $\psi = (\neg\varphi)$ and φ is a k-expression over Σ then ψ is a k-expression over Σ.*
- *If $\psi = (\varphi_1 \vee \varphi_2)$ and φ_1 and φ_2 are k-expressions over Σ then ψ is a k-expression over Σ.*
- *Nothing else is a k-expression over Σ.*

Slide 15

Definition 9 (Propositional Satisfaction) *For $\mathcal{W}$ a word model and ψ, a k-expression over Σ:*

$$\mathcal{W} \models_\Sigma \psi \overset{def}{\Longleftrightarrow} \begin{cases} \psi \text{ is an atom } v \text{ and } v \in F_{\leq k}(\rtimes \mathcal{W} \ltimes), \\ \psi \text{ is } \neg\varphi \text{ and } \mathcal{W} \not\models_\Sigma \varphi, \\ \psi \text{ is } \varphi_1 \vee \varphi_2 \text{ and either } \mathcal{W} \models_\Sigma \varphi_1 \text{ or } \mathcal{W} \models_\Sigma \varphi_2 \text{ or both.} \end{cases}$$

Slide 16

Strictly Local stringsets

Conjunctions of negative atomic constraints

$$\varphi = \neg f_1 \wedge \neg f_2 \wedge \cdots \wedge \neg f_n = \bigwedge_{f \in F} [\neg f]$$

Definition 10 (SL)

- *L is $SL_k \overset{def}{\Longleftrightarrow} L(\bigwedge_{f \in F}[\neg f])$ for some $F \subseteq F_k(\{\rtimes\} \cdot \Sigma^* \cdot \{\ltimes\})$*
- *$SL = \bigcup_{0<i\in\mathbb{N}}[SL_k]$*

For all $0 < i \in \mathbb{N}$: $\mathrm{SL}_i \subsetneq \mathrm{SL}_{i+1}$.

$\mathrm{Fin} \cup \mathrm{CoFin} \subsetneq \mathrm{SL}$, $\mathrm{DEF} \cup \mathrm{RDEF} \subsetneq \mathrm{SL}$

Slide 17

Abstract characterization of SL_k

Lemma 1 (k-Suffix Substitution Closure (SSC)) *If $L \in SL_k$ then for all strings u_1, v_1, u_2, and v_2 in Σ^* and all $x \in \Sigma^{k-1}$*

$$u_1 \cdot x \cdot v_1 \in L \text{ and } u_2 \cdot x \cdot v_2 \in L \Rightarrow u_1 \cdot x \cdot v_2 \in L.$$

There is no k for which $\mathrm{Fin} \subseteq \mathrm{SL}_k$.

GDEF $\not\subseteq$ SL and SL $\not\subseteq$ GDEF

Slide 18

Closure properties of SL

- SL and SL_k are closed under intersection but not union or complement
- SL and SL_k are not closed under concatenation
- SL_2 is closed under iteration (Kleene-$*$).
- $\mathrm{SL}_{k>2}$ and SL are not closed under iteration
- SL and SL_k are not closed under alphabetic homomorphism.

$\{(ab)^i \mid i \in \mathbb{N}\} \in \mathrm{SL}_2$

$\{(aa)^i \mid i \in \mathbb{N}\} \notin \mathrm{SL}$

Some-$b \stackrel{\text{def}}{=} \{w \in \{ab\}^* \mid |w|_b > 1\} \notin \mathrm{SL}$

Locally Testable stringsets

Definition 11 (LT)

- *A stringset is LT_k iff it is $L(\varphi)$ for some k-expression φ.*
- $LT \stackrel{def}{=} \bigcup_{0<i\in\mathbb{N}}[LT_i]$

For all $0 < i \in \mathbb{N}$: $\mathrm{LT}_i \subsetneq \mathrm{LT}_{i+1}$.

Slide 19 $\mathrm{SL}_k \subseteq \mathrm{LT}_k, \qquad \mathrm{SL}_{k+1} \not\subseteq \mathrm{LT}_k \not\subseteq \mathrm{SL}_{k+1}$

Abstract characterization of $\mathbf{LT}_k$

Lemma 2 (k-Test Invariance) *A language $L \subseteq \Sigma^*$ is LT_k for some $k > 0$ if and only if, for all strings $w, v \in \Sigma^*$:*

$$(F_k(\rtimes w \ltimes) = F_k(\rtimes v \ltimes)) \Rightarrow (w \in L \Leftrightarrow v \in L).$$

Closure properties of LT

- LT and LT_k are closed under all Boolean operations
- LT and LT_k are not closed under concatenation
- LT_k and LT are not closed under iteration
- LT and LT_k are not closed under alphabetic homomorphism.

Slide 20

Some-$b \stackrel{\text{def}}{=} \{w \in \{ab\}^* \mid |w|_b > 1\} \in \mathrm{LT}_1$

One-$b \stackrel{\text{def}}{=} \{w \in \{ab\}^* \mid |w|_b = 1\} \notin \mathrm{LT}_1$

First-order(successor) definable stringsets

Slide 21

Definition 12 (FO(+1)) *A stringset is FO(+1) iff it is $L(\varphi)$ for some first-order sentence in which $\triangleleft^+$ does not occur.*

Abstract characterization of FO(+1)

Definition 13 *A stringset L is (k,t)-**Locally Threshold Testable** ($L \in LTT_{k,t}$) iff whenever $w \equiv_{k,t} v$ then either both w and v are in L or neither are.*

Slide 22

Theorem 1 (Thomas'82) *A stringset is FO(+1) iff there is some k and t such that it is $LTT_{k,t}$.*

FO(+1) is not closed under concatenation, iteration or alphabetic homomporphism.

One-$b \stackrel{\text{def}}{=} \{w \in \{ab\}^* \mid |w|_b = 1\} \in$ FO(+1)

No-c-before-$b \stackrel{\text{def}}{=}$ Some-$b \cdot \{a, b, c\}^* \notin$ FO(+1).

First-order definable stringsets

Slide 23

Definition 14 (FO(<)) *A stringset is FO(<) iff it is $L(\varphi)$ for some first-order sentence (in which $\triangleleft^+$ may occur).*

$\triangleleft$ is FO definable from $\triangleleft^+$.

Abstract characterization of FO(<)

Slide 24

Definition 15 *A stringset is **non-counting** iff there exists some $n > 0$ (depending only on the language) such that for all strings $u, v, w \in \Sigma^*$, where $|v| \geq 1$, and for all $i \geq 1$*

$$uv^n w \in L \Leftrightarrow uv^{n+i} w \in L.$$

FO(<) is closed under concatenation.

FO(<) is not closed under iteration or alphabetic homomorphism.

No-c-before-$b \stackrel{\text{def}}{=}$ Some-$b \cdot \{a, b, c\}^* \in$ FO(<).

$\{(aa)^i \mid i \in \mathbb{N}\} \notin$ FO(<).

Precedence—Subsequences

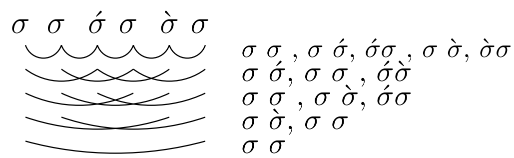
Definition 16 (Subsequences)

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

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

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

Slide 25



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

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

Definition 17 (Piecewise Propositional Satisfaction) *For $\mathcal{W}$ a word model and ψ, a k-expression over Σ:*

Slide 26

$$\mathcal{W} \models^P_\Sigma \psi \overset{def}{\Longleftrightarrow} \begin{cases} \psi \ \textit{is}\ v \in \Sigma^k \ \textit{and}\ v \sqsubseteq \mathcal{W}, \\ \psi \ \textit{is}\ \neg\varphi \ \textit{and}\ \mathcal{W} \not\models^P_\Sigma \varphi, \\ \psi \ \textit{is}\ \varphi_1 \vee \varphi_2 \ \textit{and either}\ \mathcal{W} \models_\Sigma \varphi_1 \ \textit{or}\ \mathcal{W} \models^P_\Sigma \varphi_2 \ \textit{or both.} \end{cases}$$

Strictly Piecewise stringsets

Conjunctions of negative atomic constraints

$$\varphi = \neg f_1 \wedge \neg f_2 \wedge \cdots \wedge \neg f_n = \bigwedge_{f \in F} [\neg f]$$

Slide 27

Definition 18 (SP)

- L *is* $SP_k \overset{def}{\Longleftrightarrow} L(\bigwedge_{f \in F}[\neg f])$ *for some* $F \subseteq \Sigma^{\leq k}$
- $SP = \bigcup_{0<i\in\mathbb{N}}[SP_k]$

For all $0 < i \in \mathbb{N}$: $\mathrm{SP}_i \subsetneq \mathrm{SP}_{i+1}$.

Fin, CoFin $\not\subset$ SP

Character of the Strictly k-Piecewise Sets

Theorem 2 *A stringset L is Strictly k-Piecewise Testable iff it is closed under subsequence:*

Slide 28

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

SP and SP_k, for any $k > 0$, are closed under intersection but not union or complement. SP_k is not closed under concatenation, although SP is.

Piecewise Testable stringsets

Definition 19 (PT)

- *A stringset is PT_k iff it is $L(\varphi)$ for some piecewise k-expression φ.*
- $PT \overset{def}{=} \bigcup_{0<i\in\mathbb{N}}[PT_i]$

Slide 29

For all $0 < i \in \mathbb{N}$: $\mathrm{PT}_i \subsetneq \mathrm{PT}_{i+1}$.

$\mathrm{Fin} \cup \mathrm{CoFin} \subseteq \mathrm{PT}, \quad \mathrm{DEF}, \mathrm{RDEF} \not\subseteq \mathrm{PT}$.

Abstract characterization of $\mathbf{PT}_k$

Lemma 3 (k-Test Invariance) *A language $L \subseteq \Sigma^*$ is PT_k for some $k > 0$ if and only if, for all strings $w, v \in \Sigma^*$:*

$$(P_k(w) = P_k(v)) \Rightarrow (w \in L \Leftrightarrow v \in L).$$

Sub-regular hierarchies

Slide 30