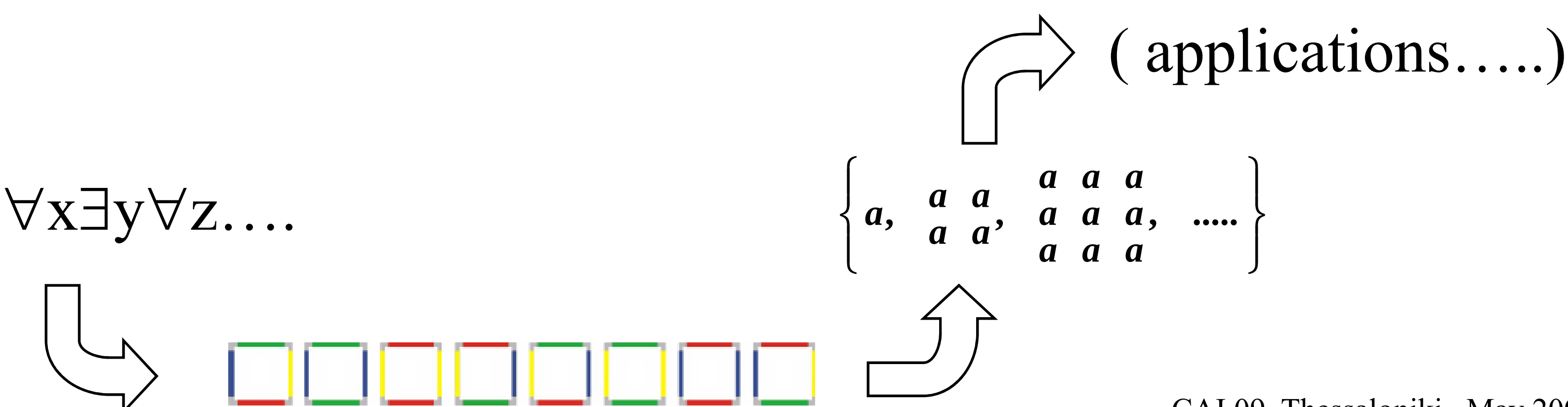
Picture languages: from Wang tiles to Grammars

Alessandra Cherubini
Politecnico di Milano, Dept.of Math.“F.Brioschi”, Italy
alessandra.cherubini@polimi.it



Contents

- Generalities on pictures
- Generalities on Wang tiles
- Recognizable 2D languages
- Grammar models based on tiles
- Final remarks

Goal

Generalizing formal language theory from 1D to 2D.

Finite alphabet Σ $=\{a,b\}$

String over Σ → Picture over Σ

w=abaaba

$$p = \begin{matrix} a & b & a & b \\ b & a & b & a \\ a & b & a & b \end{matrix}$$

Size of p $=|p|= (p_{row},p_{col}) = (3,4)$

$\Sigma^{*,*}= \Sigma^{+,+}\cup\{\lambda\}$ set of all pictures over Σ

$L \subseteq \Sigma^{*,*}$ 2D (or picture) language

- $p_{row}=1$: new definitions reduce to definitions for 1D case
- 2D languages inherit as many as possible properties of 1D languages

Basic notions on pictures and 2D languages (1)

$p\ ,\ q \in \Sigma^{*,*}$

- Row concatenation $\ominus$: $p \ominus q = \begin{array}{|c|}\hline p \\ \hline q \\ \hline\end{array}$
- Column concatenation $\obar$: $p \obar q = \begin{array}{|c|c|}\hline p & q \\ \hline\end{array}$
- Row and column stars: $p^{*\ominus}$, $p^{*\obar}$

} Partial operations

Above operations extend to languages

$$L^{0\ominus} = L^{0\obar} = \{\lambda\}$$

$$L^{i\ominus} = L \ominus L^{(i-1)\ominus},\ L^{i\obar} = L \obar L^{(i-1)\obar}$$

$$L^{*\ominus} = \cup_{i\geq 0} L^{i\ominus},\ L^{*\obar} = \cup_{i\geq 0} L^{i\obar}$$

Basic notions on pictures and 2D languages (2)

- Clockwise rotation : $p = \begin{matrix} p(1,1) \dots p(1,k) \\ \vdots \quad \ddots \quad \vdots \\ p(h,1) \dots p(h,k) \end{matrix}$; $\quad rot(p) = \begin{matrix} p(h,1) \dots p(1,1) \\ \vdots \quad \ddots \quad \vdots \\ p(h,k) \dots p(1,k) \end{matrix}$

$p \in \Sigma^{*,*}$, $q \in \Delta^{*,*}$

- Cartesian product $\otimes$: $r = p \otimes q \in (\Sigma \times \Delta)^{*,*}$, $r(i,j)=(p(i,j),q(i,j))$

Above operations extend to languages

- Row column combination $\oplus$ of $L_1, L_2 \subseteq \Sigma^{*,*}$:

$p \in L_1 \oplus L_2$ iff

all rows and columns of p are in L_1 and L_2 respectively.

Basic notions on pictures and 2D languages (3)

$$p = \begin{matrix} p(1,1) \dots p(1,k) \\ \vdots \quad \ddots \quad \vdots \\ p(h,1) \dots p(h,k) \end{matrix} \in \Sigma^{*,*}$$

- domain of p: dom (p) $=\{1,2,\dots,h\}\times\{1,2,\dots,k\}$
- subdomain of p: $d=\{x, x+1,\dots, x'\}\times\{y, y+1,\dots, y'\}=[x,y;x',y']$
  $1\leq x\leq x'\leq h,\quad 1\leq y\leq y'\leq k$
- subpicture of p associated to d, q=spic (p,d):
  $q(i,j)=p(i+x-1, j+y-1)$ for all $1\leq i\leq x'-x+1,\ 1\leq j \leq y'-y+1$
- subpicture (block) of p: spic(p,d) for some d subdomain of d
- Simplot's operator **: $p \in L^{**}$ iff
  $\exists\, \Pi$ partition of dom(p) s.t. spic (p,d) $\in L$ for all $d \in \Pi$

Basic notions on pictures and 2D languages (4)

Δ, Σ finite alphabets

- projection of $q \in \Delta^{*,*}$, $\pi: \Delta \to \Sigma$: $p = \pi(q) \in \Sigma^{*,*}$
  s.t. $|p|=|q|$ and $\forall (i,j) \in dom(q)$ $p(i,j) = \pi(q(i,j))$
- replacement of q by q' in p, $p[q'/q]$: $p \in \Delta^{*,*}$, $d=[x,y;x',y']$ subdomain of dom (p), $q=spic(p,d)$, $q' \in \Sigma^{*,*}$ s.t. $|q|=|q'|$

$$p[q'/q](i,j) = \begin{cases} q'(i-x+1, j-y+1) & \text{for all } x \le i \le x', y \le j \le y' \\ p(i,j) & \text{otherwise} \end{cases}$$

$\Delta = \{a,b\}$, $\Sigma = \{c,d,e\}$

$$p = \begin{array}{cccc} a & a & a & b \\ b & b & b & b \\ b & b & b & b \end{array} \qquad q = \begin{array}{cc} a & a \\ b & b \end{array} \qquad q' = \begin{array}{cc} c & c \\ d & e \end{array} \qquad p[q'/q] = \begin{array}{cccc} a & c & c & b \\ b & d & e & b \\ b & b & b & b \end{array}$$

Basic notions on pictures and 2D languages (5)

- homogeneous partition of p: $\Pi_p = \{d_1, \ldots, d_r\}$, partition of dom(p), s.t. $\forall\ 1 \leq i \leq r\ \exists a_i \in \Delta \quad spic(p, d_i) \in \{a_i\}^{*,*}$

$$p = \begin{array}{|cc|c|c|} \hline a & a & a & b \\ \hline b & b & b & b \\ \cline{3-4} b & b & b & b \\ \hline \end{array}$$

$\sigma(a) = \{c\}^{*,*}$, $\sigma(b) = \{d\}^{*,*} \cup \{d\ e\}$

$$\left\{ \begin{array}{cccc} c\,c\,c\,d & c\,c\,c\,d & c\,c\,c\,d & c\,c\,c\,d \\ d\,d\,d\,e\ , & d\,d\,d\,d\ , & d\,d\,d\,d\ , & d\,d\,d\,e \\ d\,d\,d\,d & d\,d\,d\,e & d\,d\,d\,d & d\,d\,d\,e \end{array} \right\} = \sigma(p, \Pi_p)$$

$\sigma: \Delta \rightarrow 2^{\Sigma^{*,*}}$,

- substitution of p induced by the homogeneous partition Π_p:

$p \in \Delta^{*,*}$, $q_i = spic(p, d_i)$,

$$\sigma(p, \Pi_p) = \{p[q'_1/q_1][q'_2/q_2] \ldots [q'_r/q_r] \mid q'_i \in \sigma(a_i)\}$$

$L \subseteq \Delta^{*,*}$, homogeneous partition set $\Pi = \{(p, \Pi_p) \mid p \in L\}$ with Π_p homogeneous partition of p,

- substitution of L induced by Π: $\sigma(L, \Pi) = \cup_{p \in L}\, \sigma(p, \Pi_p)$

Hierarchy of picture languages (?)

Chomsky's hierarchy for

1D languages:

- Regular
- Context free
- Context sensitive

2D languages

- REC robust theory
- several approaches
- few attempts

mainly based on tiles

Wang tiles (1)

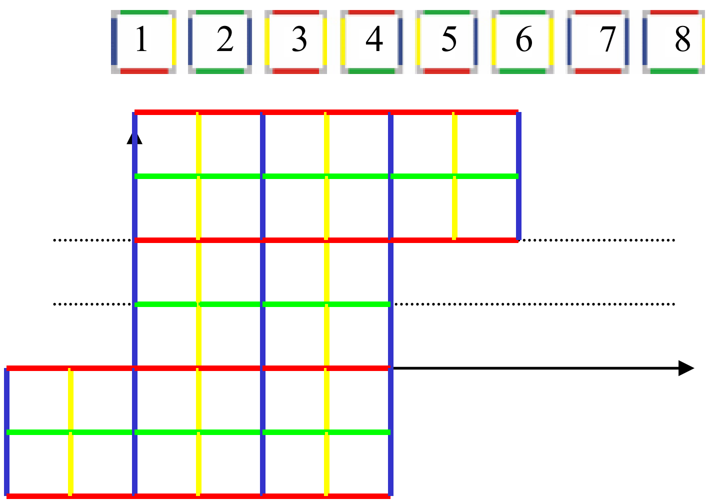
Wang tile: unit square t with colored edges N(t),E(t),S(t),W(t)

T finite set of Wang tiles.

T admits a valid tiling if there is $\tau: Z^2 \rightarrow T$ such that $\forall i,j \in Z$

$$N(\tau(i,j))=S(\tau(i,j+1)),\ S(\tau(i,j))=N(\tau(i,j-1)),$$
$$W(\tau(i,j))=E(\tau(i-1,j)),\ E(\tau(i,j))=W(\tau(i+1,j))$$

Plane tiling problem: Determine whether a given T admits a valid tiling. (Wang,1961)



admits the valid tiling

$\tau(2h,2k)=8,$
$\tau(2h+1,2k)=4,$
$\tau(2h,2k+1)=1,$
$\tau(2h+1,2k+1)=5$

with double periodicity!
but the set admits also non periodic tiling.

Wang tiles (2)

- The plane tiling problem is undecidable. (Berger 66)

- there are finite sets of tiles which admit valid tilings but no periodic valid tiling (aperiodic sets)
  - First aperiodic set: 20426 Wang tiles (Berger 66)
  - Smaller (known) aperiodic set: 13 Wang tiles (Culik 96)
- It is undecidable whether a given finite set of Wang tiles can tile the plane periodically. (Gurevich and Koriakov 1972)
- Each finite set of Wang tiles admitting a valid tiling admits a valid tiling s. t. for each pattern appearing in the tiling there is an integer n such that the pattern appears in all squares of tiling of size n (quasi periodic tiling). (Durand 99)

Quasiperiodicity function, f:x→n, n minimal size of squares in which appear all square patterns of size x of the tiling.

Wang tiles (3)

- A quasiperiodic tiling is periodic iff its quasiperiodicity function is bounded by $x \to x+c$ for some constant c
- A tile set forming a quasiperiodic tiling which is not periodic, can form an uncountable number of different tilings

Tiling finite regions of the plane: a finite set T of Wang tile admits a valid tiling of the region $d=\{1,2,\ldots,n\}\times\{1,2,\ldots,m\}$ with colored boundary iff there is a map $\tau: d \to T$ so that common edges of adjacent tiles have the same color and border tiles agree with boundary color (NP-complete problem, van Emde Boas, 83)

Valid tilings via set of local rules

- T: finite set of Wang tiles
- $L \subseteq T^4$ set of local rules
- T admits a valid tiling of the plane (a finite region) iff each 2×2 pattern appearing in the tiling is in L

Local picture languages over $\Gamma \cup \{\#\}$

$p \in \Gamma^{*,*}$, bordered picture

$$\hat{p} = \begin{array}{ccc} \# & \# \ldots\ldots\ldots\ldots & \# \\ \# & p & \# \\ \# & \# \ldots\ldots\ldots\ldots & \# \end{array} , \quad \# \notin \Gamma$$

θ finite set of square pictures of size 2 (tiles) over $\Gamma \cup \{\#\}$,
$p \in Loc(\theta)$ iff 2×2 subpictures of $\hat{p}$ are in θ.

Recognizable picture languages

Equivalent definitions

REC

Wang Systems (WS): (Q,Σ,T) with
$Q=C\cup\{b\}$, Σ finite sets,
$T\subseteq Q^4 \times \Sigma$ (set of labelled Wang tiles)

$p\in \Sigma^{*,*}$ is recognized by (Q,Σ,T) iff dom(p), with the boundary colored by b, admits a valid tiling τ s.t.

- $\tau((i,j))= (q_1,q_2,q_3,q_4,p(i,j))$
- no inner edge is labelled by b. (De Prophetis, Varricchio, 97)

Tiling systems (TS): $(\Sigma,\Gamma,\theta,\pi)$ with
Σ, Γ finite sets, θ finite set of tiles over $\Gamma\cup\{\#\}$, $\pi: \Gamma\rightarrow \Sigma$.

$p\in \Sigma^{*,*}$ is recognized by $(\Sigma,\Gamma,\theta,\pi)$ iff there exists $q\in\Gamma^{*,*}$ s.t.

- $\hat{q} \in Loc(\theta)$ and
- $\pi(q) = p$ (Giammarresi, Restivo , 92-96)

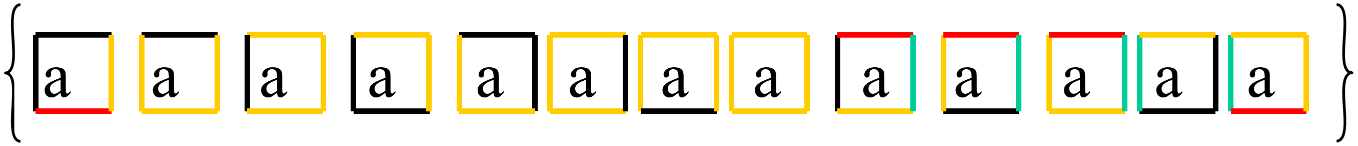
hv-languages $\longrightarrow$ domino systems (DS)

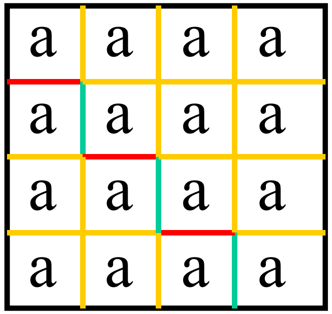
An (usual) example in a less usual way

Square languages via WS:

C={red, yellow, green}, Σ={a},

T=





Other "local" languages

hv-languages: tiles of sizes (1,2) and (2,1)

Locally testable 2D languages:LT

$\sim_{r,s} \subseteq \Sigma^{*,*} \times \Sigma^{*,*}$: $p \sim_{r,s} q$ iff

they have the same sets of subpictures of size (r,s)

$L \in LT$ iff L is union of $\sim_{r,s}$ -classes for some r,s

Locally threshold testable 2D languages :LTT

$\sim_{r,s}$+ counting occurrences of subpictures till a threshold

- hv-languages $\subset$ LOC $\subset$ LT $\subset$ LTT $\subset$ REC

Piecewise testable languages: PT

~r,s based on "scattered" subpictures

Closure properties of REC

REC is closed under

- Boolean union and intersection
- concatenations, and their closures
- rotations
- projections
- Cartesian product
- Simplot's operator

REC substitution: $\sigma: \Delta \rightarrow 2^{\Sigma^{*,*}}$ s.t. $\forall a \in \Delta$ $\sigma(a)$ is in REC

REC is closed under

- REC substitution induced by homogeneous partition sets
- Universal REC substitutions (i.e. REC substitutions induced by the set of all homogeneous partitions of each picture). (C., Crespi Reghizzi, Pradella, San Pietro, 06)

Negative results

REC is not closed under complementation

REC has NP-complete parsing problem

1) for all picture $p \in \Sigma^{*,*}$ compute $q \in \Gamma^{*,*}$ s.t. $\pi(q)=p$

2) verify that $q \in Loc(\theta)$

(but the parsing is succesfully tackled encoding the problem into SAT (Crespi Reghizzi, Pradella, 07))

Emptiness problem is undecidable for REC

Universe problem is undecidable for REC

Determinism and unambiguity (1)

REC definition is implicitely non deterministic

REC is not closed under complementation

→ smaller families to eliminate non determinism

UREC: L is in UREC iff there is a tiling system $(\Sigma,\Gamma,\theta,\pi)$ recognizing L s.t. $\forall p \in L \ \exists! \ q \in L(\Gamma)$ s.t. $p = \pi(q)$ $(\Sigma,\Gamma,\theta,\pi)$ is unambiguos for L

(Anselmo, Giammarresi, Madonia, Restivo, 06)

- UREC$\subset$REC
- UREC closed under disjoint union, intersection, projection and rotation
- UREC is not closed under row, column concatenations and their closures

Determinism and unambiguity (2)

Col-UREC: L is in Col-UREC iff there is a tiling system $(\Sigma,\Gamma,\theta,\pi)$ recognizing L s.t. $\forall p \in L$ once computed the local symbols of a column the local symbols of the "next" (r2l/l2r) one are uniquely determined

Similar definition for Row-UREC, Diag-UREC

- Col-UREC$\cap$Row-UREC$\subset$Col-UREC$\cup$Row-UREC $\subset$UREC

$(\Sigma,\Gamma,\theta,\pi)$ is tl2br deterministic iff $\forall \gamma_1,\gamma_2, \gamma_3 \in \Gamma$, $\forall a \in \Sigma$ there is at most one tile $\begin{matrix}\gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_4\end{matrix}$ in θ s.t. $\pi(\gamma_4)=a$

Similar definition for other "corner 2 corner" directions

Determinism and unambiguity (3)

DREC languages recognized by a deterministic TS

Up to now determinism along diagonals. Other strategies?

- $(\Sigma,\Gamma,\theta,\pi)$ is snake deterministic iff $\Gamma = \Gamma_1 \cup \Gamma_2$, $\theta = \theta_1 \cup \theta_2$ (Γ_1,Γ_2; θ_1,θ_2 disjoint sets) s.t. all tiles in $(\Sigma,\Gamma_1,\theta_1,\pi)$ is tl2br deterministic, $(\Sigma,\Gamma_2,\theta_2,\pi)$ is tr2bl deterministic, tiles in θ_i have the first row in $\Gamma_{3-i} \cup \{\#\}$ and the second in $\Gamma_i \cup \{\#\}$, no tile in θ_2 has # # as first row

L(snake-DTS)=L(t2b-UTS) (Lonati, Pradella, 09)

Snake-DREC: closure under rotation of L(snake-DTS)

Snake-DREC= Col-UREC$\cup$Row-UREC

- Build the preimages in $\Gamma^{*,*}$ by a finite sequence of unique choices (Reinhardt, 98)

Determinism and unambiguity (4)

Is decidable whether a tiling system is deterministic

Is undecidable whether a tiling system for L is unambiguous

Is decidable whether a tiling system for L is col-unambiguous

DREC is closed under complementation.
What about UREC?

REC and automata: 4 way finite automaton

In 1D local language+projection give a finite automaton

In 2D? A scanning procedure is missed

4NFA, $\mathcal{A} = (\Sigma, Q, D, q_0, q_a, q_r, \delta)$ (Blum and Hewitt, 59-67) with

- Σ,Q finite sets, D={t,b,l,r},
- $q_0, q_a, q_r \in Q$ initial, accepting, rejecting states
- $\delta(q,a) \ni (q', d)$, $d \in D$.

$\mathcal{A}$ reads picture from tl corner, accepts reaching q_a, is border-sensitive, does not read all pixels of p.

4DFA, deterministic version of 4NFA

$L(4DFA) \subset L(4NFA) \subset REC$;

$L(4DFA) \subset UREC$, L(4DFA) incomparable with DREC

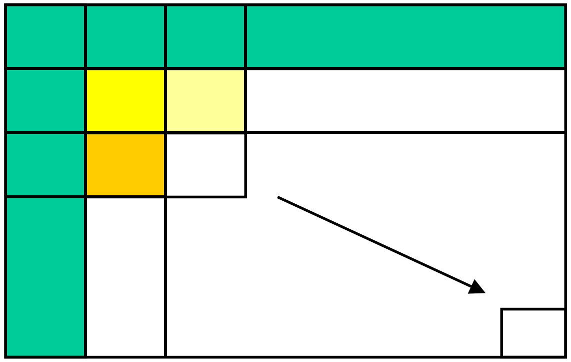
L(4DFA), L(4NFA) not closed under concatenations and closure operators.

REC and automata: 2D on line tasselation automaton

2OTA $\mathcal{A} = (\Sigma, Q, I, F, \delta)$ (Inoue and Nakamura,77) with

- Σ, Q finite sets; $I, F \subseteq Q$
- $\delta: Q \times Q \times \Sigma \rightarrow 2^Q$

Restricted type of 2 dimensional cellular automata computing by diagonal



2DOTA deterministic version

$L(2DOTA) \subset L(2UOTA) \subset L(2OTA)$

$L(2OTA) = REC$

$L(2UOTA) = UREC$

$L(2DOTA) \subset DREC$

L(2DOTA) incomparable with L(4DFA)

REC and automata: tiling automaton (1)

TA : tiling system+ scanning procedure+ data structure

Several scanning strategies:
by rows, by columns, snake like…

Good procedure?

computable next position function

all the picture is filled

contiguity property

a corner as start point

main direction

Data structure:

able to
- remember some local already scanned symbols needed to compute transition
- insert new state

linear size in p_{row} or p_{col}

depends on scanning procedure

REC and automata: tiling automaton (2)

(tl2br)-TA: (T,S,D_0,δ) (Anselmo,Giammaresi, Madonia 07) where:

- $T=(\Sigma,\Gamma,\theta,\pi)$ is a tiling system
- S is a tl2br directed scanning stratedy
- D_0 is the initial content of the data structure
- $\delta: (\Gamma\cup\{\#\})^3\times(\Sigma\cup\{\#\})\rightarrow 2^{\Gamma\cup\{\#\}}$: $\gamma_4\in\delta(\gamma_1,\gamma_2,\gamma_3,a)$ iff

$$\begin{matrix}\gamma_1 & \gamma_2\\ \gamma_3 & \gamma_4\end{matrix} \in \theta \quad \text{and} \quad \pi(\gamma_4)=a$$

Analogous definitions for other directions

DTA deterministic version of TA

$L(DTA)\subset L(TA)=REC$

$L(2DOTA) \subset L(DTA)=DREC$

REC and automata: quadripolic automaton (1)

Free doubly ranked monoid over Σ:

think to pictures on Σ with row and column concatenations
pict (Σ)

Quadripolic relations over Q:

$4\text{-}Rel_{m,n}(Q)$ set of $R \subseteq Q^m \times Q^n \times Q^m \times Q^n$,

$4\text{-}Rel(Q)=(4\text{-}Rel_{m,n}(Q))$

Θ vertical multiplication: $R \in 4\text{-}Rel_{m,n}(Q)$, $S \in 4\text{-}Rel_{p,n}(Q)$,

$R \,\Theta\, S=\{(w_1,w_2,w_3,w_4)\}$ s.t.

$w_1=z_1v_1$, $w_3=z_3v_3$ for some $v_1,v_3 \in Q^m$, $z_1,z_3 \in Q^p$

$(v_1,w_2,v_3,u) \in R$, $(z_1,u,z_3,w_4) \in S$ for some $u \in Q^n$

Horizontal multiplication is defined similarly

→ 4-Rel(Q) doubly ranked monoid

Natural extension of homomorphism notion

REC and automata: quadripolic automaton (2)

QA $\mathcal{A}=(X,Q,\delta,F_W,F_S,F_E,F_N)$ (Bozapilidis, Grammatikoupoulou,05):

- X doubly ranked set
- Q finite set of states,
- $F_W,F_S,F_E,F_N \subseteq Q$ poles of acceptance
- δ: set of maps $\delta_{m,n}: \Sigma^{(m,n)} \rightarrow$ 4-$Rel_{m,n}(Q)$

δ^* morphism from Pict(X) $\rightarrow$ 4-Rel (Q) extending δ

p $\in$ Pict $_{m,n}$(X) is accepted by $\mathcal{A}$ iff

$$\delta^*(p) \cap (F_W{}^m \times F_S{}^n \times F_E{}^m \times F_N{}^n) \neq \varnothing$$

Algebraic point of view

→ weighted version

→ rational power series

REC and logic

- REC coincides with the set of picture languages definable by EMSO formulas on the signature $\{S_1,S_2,\{P_a\}_{a\in\Sigma}\}$ (Giammarresi, Restivo, Seibert, Thomas 94)
- REC coincides with the set of picture languages definable by EMSO formulas on the signature $\{S_1,S_2,\{P_a\}_{a\in\Sigma}\}$ having the form $\exists X\varphi(X)$ with $\varphi(X)$ first-order formula (Matz, 98)
- LTT coincides with the set of picture languages definable by first order formulas on the signature $\{S_1,S_2,\{P_a\}_{a\in\Sigma}\}$

2D-Regular expressions (1)

Regular expressions over Σ: expressions built over Σ and $\varnothing$ using operators from

$$R=\{\cup,\cap,{}^{C}, \ominus, \obar, {}^{*\ominus}, {}^{*\obar}\}$$

L(RE) family of picture languages denoted by regular expressions

L(RE) and REC are different $\Rightarrow$ Restrict R!

$R_1=\{\cup,\cap, \ominus, \obar, {}^{*\ominus}, {}^{*\obar}\} \Rightarrow$ CFRE $\Rightarrow$ L(CFRE)

- hv-languages $\subset$ L(CFRE) $\subset$ REC

R_1+projections $\Rightarrow$ PCFRE $\Rightarrow$ L(PCFRE)

- L(PCFRE)=REC (Giammarresi, Restivo, 96)
- L $\in$ REC over Σ iff there are S_1, S_2 regular 1D languages over Γ and $\pi: \Gamma \rightarrow \Sigma$ s.t. $L= \pi(S_1 \oplus S_2)$.

2D-Regular expressions (2)

$R_2=\{\cup,\cap,^C, \ominus, \varnothing\!\!\!\!\mid\} \Rightarrow$ SFRE $\Rightarrow$ L(SFRE)

L(SFRE) not comparable with REC (Matz, 98)

CORNERS$=\{p\in\{a,b\}^{*,*} \mid \forall i,j,h,k\ p(i,j)=p(i,h)=p(k,j)=b \rightarrow p(h,k)=b\}$

CORNERS$\in$L(SFRE)

a corner is labelled by a

$K=\cup_{x,y,z,t\in\{a,b\},xyzt\in b^*ab^*}(x \varnothing\!\!\!\!\mid \varnothing^C \varnothing\!\!\!\!\mid y)\ominus \varnothing^C\ominus(z \varnothing\!\!\!\!\mid \varnothing^C \varnothing\!\!\!\!\mid t)$

CORNERS$= (\varnothing^C \ominus (\varnothing^C \varnothing\!\!\!\!\mid K \varnothing\!\!\!\!\mid \varnothing^C) \ominus \varnothing^C)^C$

CORNERS$\notin$ REC tricky counting argument

CORNERS $\in$ PT

PT not comparable with SFRE….

2D-Regular expressions (3)

Aim: increase the power of CFRE.

Adding operators: $\ominus a^{+}$, ⦶ a^{+} as individuals objects to be applied to expressions and composed each other with some restrictions

$\Rightarrow$ regular expressions with restricted operators (Matz 97)

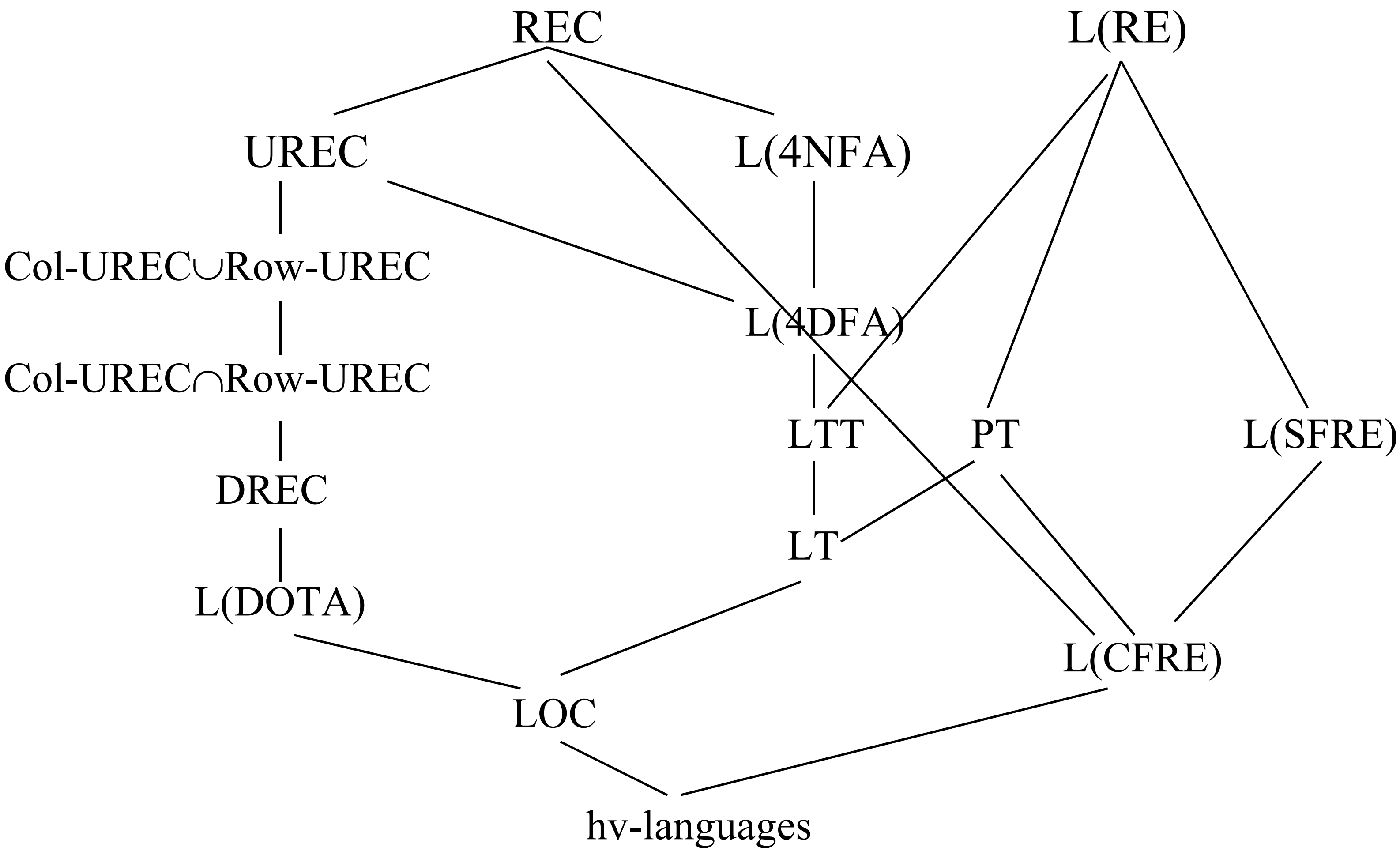
Does it reach REC?

Simplot's operator: to reach REC projection is needed

Diagonal concatenation and its closure
Advanced stars
} One letter alphabet

(Anselmo, Giammarresi, Madonia, 05)

A taxonomy of (some) 2D languages



To state inclusions…

Closure properties, complexity of parsing, the 1D case….

+ some necessary conditions

Vertical pumping lemma: If $L \in REC$, there is $\varphi: N \rightarrow N$ so that $\forall p \in L$ with $p_{row} > \varphi(p_{col})$ there are x, y,q with $|(x \ominus q)_{row}| \leq \varphi(p_{col})$ s.t. $p = x \ominus q \ominus y$ and $\forall i \geq 0$, $p = x \ominus q^{i\ominus} \ominus y \in L$

Lemma (Matz, 98) Let $L \in REC$ over Σ . Let $(M_m \subseteq \Sigma^{m,*} \times \Sigma^{m,*})$ be a sequence s.t. $\forall (p,q) \in M_m \; p \oplus q \in L$,

$\forall (p,q),(p',q') \in M_m \; \{p \oplus q', p' \oplus q\} \not\subseteq L$

Then $|M_m| = 2^{O(m)}$

→ no more than exponentially much space to pass information from a side to the other of a picture in a REC language.

One letter alphabet

Why to consider REC_1: REC on one letter alphabet ?

- Simpler to hand
- Give the shapes of pictures
- Like string languages on three letters alphabets by encoding each picture p as $\varphi(p) \in \{a,h,v\}^*$:

$$\varphi(p) = \begin{cases} a^{p_{row}} h a^{p_{col}-p_{row}-1} & \text{if } p_{row}<p_{col} \\ a^{p_{row}} & \text{if } p_{row}=p_{col} \\ a^{p_{col}} v a^{p_{row}-p_{col}-1} & \text{if } p_{col}<p_{row} \end{cases}$$

➢L is in REC_1 iff $\varphi(L)$ can be recognized by 1-tape nondeterministic Turing machine working , for any input $x \in \{a,h,v\}^*$, within $|x|$ with at most ${}_a|x|$ head reversals (${}_a|x|$ length of the longest prefix of x in a^+) (Bertoni,Goldwurm,Lonati, 07)

What about grammars?

→ Tomorrow

Tile grammars (1)

Tile Grammar (Crespi, Pradella, 05): $G=(\Sigma,N,S,R)$ where

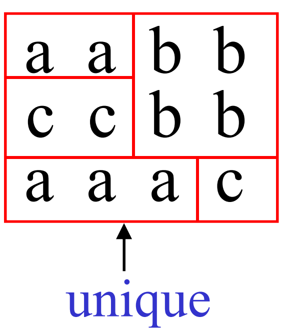
Σ, N are (terminal and non terminal) alphabets,

$S \in N$, starting symbol

R:

$$\begin{cases} A \to t \in \Sigma & \text{fixed size rule} \\ A \to \theta,\ \theta \subseteq (N \cup \{\#\})^{(i,j)} \quad 1 \le i, j \le 2 & \text{variable size rule} \end{cases}$$

Strong homogeneous partition of $p \in (\Sigma \cup N)^{*,*}$: homogeneous partition of dom(p) s.t. adjacent subdomains have distinct labels.



b b b
c c b
a a a

has no strong homogeneous partition

Tile grammars (2)

$p, q \in (\Sigma \cup N)^{*,*}$, $|p|=|q|$, Π_p homogeneous partition of p,

- $(p, \Pi_p) \Rightarrow (q, \Pi_q)$ iff there exist :
  a rule $r: A \rightarrow \rho$, a domain $d \in \Pi_p$: $spic(p,d) \in \{A\}^{*,*}$ s.t. $q=p[s/spic(p,d)]$ with $s=\rho$ if $\rho \in \Sigma$, $s \in Loc(\theta)$ and has a strong homogeneous partition otherwise,
  $\Pi_q = (\Pi_p \setminus \{d\}) \cup$(partition of s translated in d).
- $(p,\Pi) \Rightarrow^n (q,\Pi')$ iff there exists (r,Π'') s.t. $(p,\Pi) \Rightarrow (r,\Pi'')$ and $(r,\Pi'') \Rightarrow^{n-1} (q,\Pi')$
- $\Rightarrow^*$ derivation with finitely many steps
- language generated by G:
  $L(G)=\{p \in \Sigma^{*,*} \mid (S^{|p|}, \Omega_p) \Rightarrow^* (p,\iota_p)\}$,
  Ω_p, ι_p partitions induced by the universal and identical relations on dom(p)

An example (1)

Pictures over {a,b} with one row and one column (not at the border) filled by b and the remainder filled by a

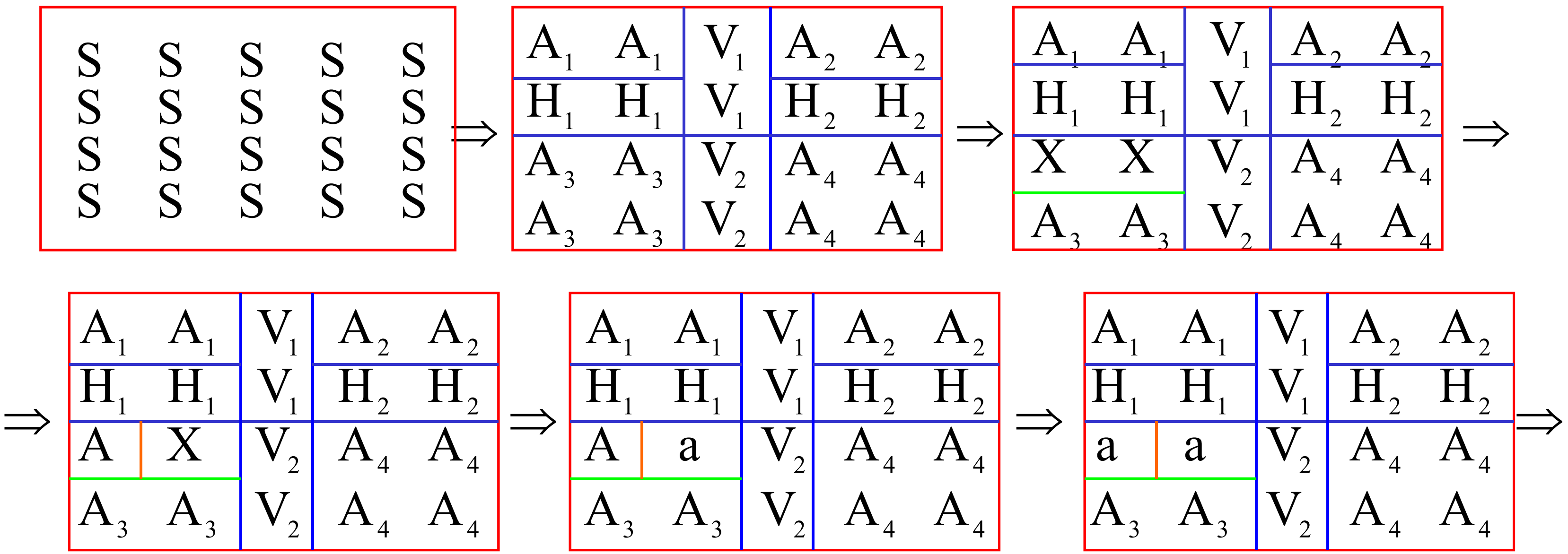
$$S \rightarrow \left\| \begin{matrix} \# & \# & \# & \# & \# & \# & \# \\ \# & A_1 & A_1 & V_1 & A_2 & A_2 & \# \\ \# & A_1 & A_1 & V_1 & A_2 & A_2 & \# \\ \# & H_1 & H_1 & V_1 & H_2 & H_2 & \# \\ \# & A_3 & A_3 & V_2 & A_4 & A_4 & \# \\ \# & A_3 & A_3 & V_2 & A_4 & A_4 & \# \\ \# & \# & \# & \# & \# & \# & \# \end{matrix} \right\|, \qquad A_i \rightarrow \left\| \begin{matrix} \# & \# & \# & \# \\ \# & X & X & \# \\ \# & A_i & A_i & \# \\ \# & A_i & A_i & \# \\ \# & \# & \# & \# \end{matrix} \right\| \,|\, \left\| \begin{matrix} \# & \# & \# & \# \\ \# & X & X & \# \\ \# & \# & \# & \# \end{matrix} \right\|,$$

$$X \rightarrow \left\| \begin{matrix} \# & \# & \# & \# & \# \\ \# & A & X & X & \# \\ \# & \# & \# & \# & \# \end{matrix} \right\|, H_j \rightarrow \left\| \begin{matrix} \# & \# & \# & \# & \# \\ \# & B & H_j & H_j & \# \\ \# & \# & \# & \# & \# \end{matrix} \right\| \,|\, b$$

$$V_j \rightarrow \left\| \begin{matrix} \# & \# & \# \\ \# & B & \# \\ \# & V_j & \# \\ \# & V_j & \# \\ \# & \# & \# \end{matrix} \right\| \,|\, b \;, \; A \rightarrow a \,, \; B \rightarrow b$$

$1 \le i \le 4$, $1 \le j \le 2$, $\|q\|$ means set of tiles in q

An example (2)



$\Rightarrow^+$

a	a	b	a	a
b	b	b	b	b
a	a	b	a	a
a	a	b	a	a

Main properties of TG (1)

- In 1D TGs are context free grammars with a local regular expression in right parts of the rules
- REC $\subseteq$ L(TG)

Hint: on language of squares

$$\theta = \left\| \begin{matrix} \# & \# & \# & \# & \# & \# \\ \# & 1 & 0 & 0 & 0 & \# \\ \# & 0 & 1 & 0 & 0 & \# \\ \# & 0 & 0 & 1 & 0 & \# \\ \# & 0 & 0 & 0 & 1 & \# \\ \# & \# & \# & \# & \# & \# \end{matrix} \right\| \quad S \rightarrow \left\| \begin{matrix} \# & \# & \# & \# & \# & \# \\ \# & 1_w & 0_b & 0_w & 0_b & \# \\ \# & 0_b & 1_w & 0_b & 0_w & \# \\ \# & 0_w & 0_b & 1_w & 0_b & \# \\ \# & 0_b & 0_w & 0_b & 1_w & \# \\ \# & \# & \# & \# & \# & \# \end{matrix} \right\| \cup \left\| \begin{matrix} \# & \# & \# & \# & \# & \# \\ \# & 1_b & 0_w & 0_b & 0_w & \# \\ \# & 0_w & 1_b & 0_w & 0_b & \# \\ \# & 0_b & 0_w & 1_b & 0_w & \# \\ \# & 0_w & 0_b & 0_w & 1_b & \# \\ \# & \# & \# & \# & \# & \# \end{matrix} \right\|$$

Main properties of TG (2)

- REC $\subset$ L(TG)

The language of pictures on {a,b} whose rows are palindromes is in L(TG) but not in REC

$$S \to \left\|\begin{matrix} \# & \# & \# & \# \\ \# & R & R & \# \\ \# & S & S & \# \\ \# & S & S & \# \\ \# & \# & \# & \# \end{matrix}\right\| \mid \left\|\begin{matrix} \# & \# & \# & \# \\ \# & R & R & \# \\ \# & \# & \# & \# \end{matrix}\right\|;$$

$$R \to \left\|\begin{matrix} \# & \# & \# & \# & \# & \# \\ \# & A & R & R & A' & \# \\ \# & \# & \# & \# & \# & \# \end{matrix}\right\| \mid \left\|\begin{matrix} \# & \# & \# & \# & \# & \# \\ \# & B & R & R & B' & \# \\ \# & \# & \# & \# & \# & \# \end{matrix}\right\|$$

$$R \to \left\|\begin{matrix} \# & \# & \# & \# \\ \# & A & A' & \# \\ \# & \# & \# & \# \end{matrix}\right\| \mid \left\|\begin{matrix} \# & \# & \# & \# \\ \# & B & B' & \# \\ \# & \# & \# & \# \end{matrix}\right\|; A \to a, B \to b, A' \to a, B' \to b$$

Main properties of TG (3)

- L(TG) has NP-complete parsing problem
  (Bad but expected news!)
- L(TG) is closed under
  - union,
  - column/row concatenation,
  - column/row closure operators,
  - rotation,
  - projection
- L(TG) is not closed under
  - intersection
  - complementation

TG vs TS

$A \in N$ is non recursive iff there is no derivation of the form $(A,\Pi) \Rightarrow^* (q,\Pi')$ with $spic(p,d) \in \{A\}^{+,+}$ for some $d \in \Pi'$

$A_1, A_2 \in N$ are mutully recursive iff for each i=1,2 there are derivations $(A_i,\Pi_i) \Rightarrow^* (p_i, \Pi'_i)$ with $spic(p_i,d_i) \in \{A_{3-i}\}^{+,+}$ for some $d_i \in \Pi'_i$

Non recursive TG grammar, RFTG: $G=(\Sigma,N,S,R)$ with all non recursive terminals

Corner grammar,CG:N partitionned in N_i $(1 \leq i \leq 4)$, M where M set of non recursive terminals, for each $i \neq j$ $A \in N_i$, $B \in N_j$ are not mutually recursive, the pictures derived by $A \in N_i$ have a rectangle in a corner labelled by N and the remainder labelled by $\Sigma \cup (N \setminus N_i)$

- L(RFTG)=L(CG)=REC

Checking whether a TG grammar is non recursive or CG is undecidable (C.,Crespi Reghizzi,Pradella,San Pietro, 06)

Regional grammars (1)

Regional partition: homogeneous partition where distinct subdomains have distinct labels

Regional picture: picture admitting a regional partition

Regional language: set of regional pictures.

a	a	b	b	b	b
a	a	b	b	b	b
a	a	d	e	e	e
c	c	d	e	e	e

Regional grammar, RTG: TG s.t for all variable size rules $A \to \theta$ the language $Loc(\theta)$ is a local regional language (C.,Crespi Reghizzi, Pradella,08)

Regional grammars (2)

Examples:

- Pictures over {a,b} with one row and one column (not at the border) filled by b and the remainder filled by a
- Pictures with palindromic rules

$$L(RTG) \not\subset REC$$

Loc(θ) (over the alphabet $N \cup \{\#\}$) is regional iff $\theta \cap N^2$ is regional and the incidence graph of $\mathcal{A}_\theta \cap N^2$ is acyclic, where $(A,B) \in \mathcal{A}_\theta$ iff A and B are distinct elements on the same row/column of a tile.

Regional grammars (3)

- The parsing problem for RTG has polynomial time complexity.

  The algorithm is a generalization of CKY algorithm for strings.

  $O(Km^4n^4)$ where K depends on the grammar

$L(RTG) \subset L(TG)$

L(RTG) incomparable with REC

Context-free Průša grammars (1)

(Context free) Průša grammars, PG: $G=(\Sigma,N,R,S)$ with

- Σ,N disjoint finite alphabets
- $S \in N$
- R set of rules : $A \to (\Sigma \cup N)^{*,*}$, $A \in N$ (Průša , 01)

$L(G)=L(G,S)$

$L(G,A)$, $A \in N$, recursively defined, $\forall\ (A \to w) \in R$

– if $w \in \Sigma^{*,*}$ then $w \in L(G,A)$,

– if $w \in (\Sigma \cup N)^{m,n}$, let $P_k = \oplus_{1 \le i \le n}\, p_{k,i}$ with $p_{k,i} = w(k,i)$ if $w(k,i) \in \Sigma$, $p_{k,i} \in L(G,w(k,i))$ if $w(k,i) \in N$, if $\forall (r,s) \in dom(w)$ $(p_{r,s})_{col} = (p_{r+1,s})_{col}$ then $P = \ominus_{1 \le k \le m} P_k \in L(G,A)$.

The right part of each picture is seen as a grid, where non terminals can be replaced maintaing the grid-like stucture

Context free Průša grammars (2)

Example: The grammar generating the language on {a,b}with one row and one column not at the border labelled by b and the remainder labelled by a

$$S \to \begin{array}{ccc} A & V & A \\ H & b & H \\ A & V & A \end{array}, \quad A \to A\,M \mid M, \quad M \to \begin{array}{c} a \\ M \end{array} \mid a, \quad V \to b\,V \mid b,$$

$$H \to b\,H \mid b$$

a	b	a	a	a
a	b	a	a	a
b	b	b	b	b
a	b	b	b	b
a	b	b	b	b

belongs to the language because
"green" pictures are in L(G,A)
"pink" pictures are in L(G,V)
"yellows" pictures are in L(G,H)

Context-free Průša grammars (3)

- PG admits a non terminal normal form where all rules are either $A \rightarrow t$ with $t \in \Sigma$, or $A \rightarrow w$ with $w \in N^{+,+}$
- $L(PG) \subset L(RTG)$

$\subseteq$:

$$A \rightarrow \begin{matrix} B_{1,1} & \dots & B_{1,k} \\ \vdots & \ddots & \vdots \\ B_{h,1} & \dots & B_{h,k} \end{matrix} \quad \Rightarrow \quad A \rightarrow \left\| \begin{matrix} B'_{1,1} & B'_{1,1} & \dots & B'_{1,k} & B'_{1,k} \\ B'_{1,1} & B'_{1,1} & & B'_{1,k} & B'_{1,k} \\ \vdots & & \ddots & & \vdots \\ B'_{h,1} & B'_{h,1} & & B'_{h,k} & B'_{h,k} \\ B'_{h,1} & B'_{h,1} & \dots & B'_{h,k} & B'_{h,k} \end{matrix} \right\|$$

$+\dots\dots$

Context-free Průša grammars (4)

Essentially PG are RTG s.t in the right parts of rules no tiles of the forms

A B C C	A C B C	C C A B	C A C B

≠:

The language of the misaligned palindromes: ribbons of two rows, the first row has a palindrome over {a,b} as prefix, the second as suffix, the remainder is filled by c.

Context-free Kolam (Matz) grammars (1)

Sentential forms over Γ, SF(Γ): non empty well-parenthesized expressions using $\ominus$, $\oplus$ and symbols from Γ.

A sentential form ϕ either defines one picture $< \phi>$ or none.

Context-free Kolam grammar , CKKG: $G=(\Sigma,N,R,S)$ with

- Σ,N disjoint finite alphabets
- $S\in N$
- R set of rules : $A\rightarrow \phi \in SF(\Sigma\cup N)$, $A\in N$ (Matz, 97)

$\gamma,\delta\in SF(\Sigma\cup N)$, $\gamma \Rightarrow \delta$ if there is $(A\rightarrow \phi) \in R$ so that δ results from γ replacing an occurrence of A by ϕ.

$\Rightarrow^*$ as usual

$L(G)=\{<\phi> \mid \phi \in SF(\Sigma), S \Rightarrow^* \phi\}$

Context-free Kolam (Matz) grammars (2)

Normal form (if $\lambda \notin L(G)$): Each rule has the form

$A \rightarrow t \mid B \ominus C \mid B \obar C,\ t \in \Sigma,\ B, C \in N.$

Example:

$S \rightarrow V \obar S \mid A_1 \ominus A_2 \mid B_1 \ominus B_1 \mid a \mid b$

$V \rightarrow A_1 \ominus A_2 \mid B_1 \ominus B_2 \mid a \mid b$

$A_2 \rightarrow V \ominus A_1 \mid a\ ,\ B_2 \rightarrow V \ominus B_1 \mid b\ ,\ A_1 \rightarrow a\ ,\ B_1 \rightarrow b$

is the normal form of the grammar generating pictures with palindromic rules

$S \Rightarrow V \obar S \Rightarrow (A_1 \ominus A_2) \obar S \Rightarrow (A_1 \ominus (V \ominus A_2)) \obar S \Rightarrow (A_1 \ominus ((B_1 \ominus B_2) \ominus A_2)) \obar S \Rightarrow (A_1 \ominus ((B_1 \ominus B_2) \ominus A_2)) \obar (A_1 \ominus (V \ominus A_2)) \Rightarrow (A_1 \ominus ((B_1 \ominus B_2) \ominus A_2)) \obar (A_1 \ominus ((A_1 \ominus A_2) \ominus A_2)) \Rightarrow^{+}$

a	a
b	a
b	a
a	a

Context-free Kolam (Matz) grammars (3)

- The generative power is the same of the grammars originally defined by Siromoney and al. ('73)
- All finite 2D languages are included in L(CFKG)
- L(CFKG) is closed under row/column concatenations and their closures
- L(CFKG) $\subset$ L(PG)

  Each rule $A \rightarrow B \ominus C$ is considered as

$$A \rightarrow \left\| \begin{matrix} \# & \# & \# & \# \\ \# & B & B & \# \\ \# & B & B & \# \\ \# & C & C & \# \\ \# & C & C & \# \\ \# & \# & \# & \# \end{matrix} \right\|$$

- Parsing complexity for CFKG in normal form is $O(m^2n^2(m+n))$

Context-free matrix grammars

Context-free matrix grammar, CFMG: $M=(G,G')$ where

- $G=(T,N,P,S)$ is a context-free string grammar with terminal alphabet $T=\{A_1,A_2,\ldots,A_k\}$,
- $G'=\{G_1,G_2,\ldots,G_k\}$ where $\forall i$, $1\leq i\leq k$, $G_i=(\Sigma,T_i,P_i,A_i)$ is a context free string grammar.

$p=\bigoplus_{1\leq j\leq n} c_j$. $p\in L(M)$ iff there is $A_{x_1}A_{x_2}\ldots A_{x_n}\in L(G)$ s.t $\forall i$, $1\leq j\leq n$, $c_j\in L(G_{x_j})$

Example: $M=(G,\{G_1,G_2\})$ with

- $G=(\{A,B\},\{S_2\},\{S\rightarrow ASA|B\},S)$,
- $G_1=(\{a,b\},\{A,A_1\},\{A\rightarrow bA_1|b,\ A_1\rightarrow aA_1|b\},A)$,
- $G_2=(\{b\},\{B\},\{B\rightarrow bB|b\},B)$

generates the pictures with odd number of columns, first and last rows and central column made of b and the remainder filled by a

- $L(CFMG)\subset L(CKKG)$

Two dimensional right linear grammars

Two dimensional right linear grammars,2 RLG: as context free matrix grammars, with G and G_i (for all G_i in G') regular grammars.

L(2RLG)⊂L(CFMG)

L(2RLG) ⊂L(4DFA)

Grid grammars (1)

Production of a grid picture grammar: a nonterminal symbol on the left-hand side and a square grid on the right-hand side evenly spaced into k^2 subsquares, $k \geq 2$, that are either black or white or labelled with a nonterminal.

Applying a production with left-hand symbol A to a square labelled by A: remove A from the square and subdivide it into smaller squares according to the right-hand side of the production

A derivation starts with the initial nonterminal in the unit square.

Productions are applied repeatedly until there is no nonterminal left, finally yielding a generated picture

The set of all so generated pictures constitutes the picture language generated by the grammar. (Drewes,96)

Grid grammars (2)

Different but basically compatible approach:

Sentential forms ψ over Γ, SF(Γ): either ψ =a$\in$V or ψ=[t_{11},…t_{1k},,…,t_{k1},….t_{kk}] with $t_{ij} \in$ SF(V), $\forall 1 \leq i,j \leq k$.

Each ψ defines a set of pictures $\{\psi\}$:

- if ψ=a then $\{\psi\}=\{p|$ p squares of $\{a\}^{+,+}\}$,
- if ψ=[t_{11},…t_{1k},,…,t_{k1},….t_{kk}] then $\{\psi\}=\{p|$ p square grid pictures s.t $p_{ij} \in \{t_{ij}\}$, all p_{ij} have the same size, p_{11}, p_{1k} ,p_{k1} , p_{kk} are at the bl,br,tl,tr corners resp.$\}$

Grid grammars (3)

Example: ψ= [[a,b,[a,b,b,a],c],a,B,[b,a,a,b]]

B	B	B	B	a	a	b	b
B	B	B	B	a	a	b	b
B	B	B	B	b	b	a	a
B	B	B	B	b	b	a	a
b	a	c	c	a	a	a	a
a	b	c	c	a	a	a	a
a	a	b	b	a	a	a	a
a	a	b	b	a	a	a	a

is the smaller picture in $\{\psi\}$

Grid grammars (4)

Grid grammar,GG: $G=(\Sigma,N,R,S)$ with

- Σ,N disjoint finite alphabets
- $S \in N$
- R set of rules : $A \rightarrow \psi \in SF(\Sigma \cup N)$, $A \in N$

In the literature parameter k in the sentential forms at the right parts of the rules is fixed, k=2 gives quadtrees.
This constraint can be relaxed.

$\gamma,\delta \in SF(\Sigma \cup N)$, $\gamma \Rightarrow \delta$ if there is $(A \rightarrow \phi) \in R$ so that δ results from γ replacing an occurrence of A by ϕ.

$\Rightarrow^*$ as usual

$L(G)=\{$smallest picture in $\{\psi\} \mid \delta \in SF(\Sigma \cup N)$ and $S \Rightarrow^* \psi\}$

Grid grammars (5)

- Generate only squares
- Admit a non terminal normal form: only rules of the forms $A \rightarrow t,\ t \in \Sigma$, or $A \rightarrow [B_{11},\ldots B_{1k},\ldots,B_{k1},\ldots,B_{kk}],\ B_{ij} \in N$
- $L(GG) \subset L(CFKG)$

$$A \rightarrow t \Longrightarrow A \rightarrow (A \obar A_V) \ominus (A_H \obar t) \mid t,$$

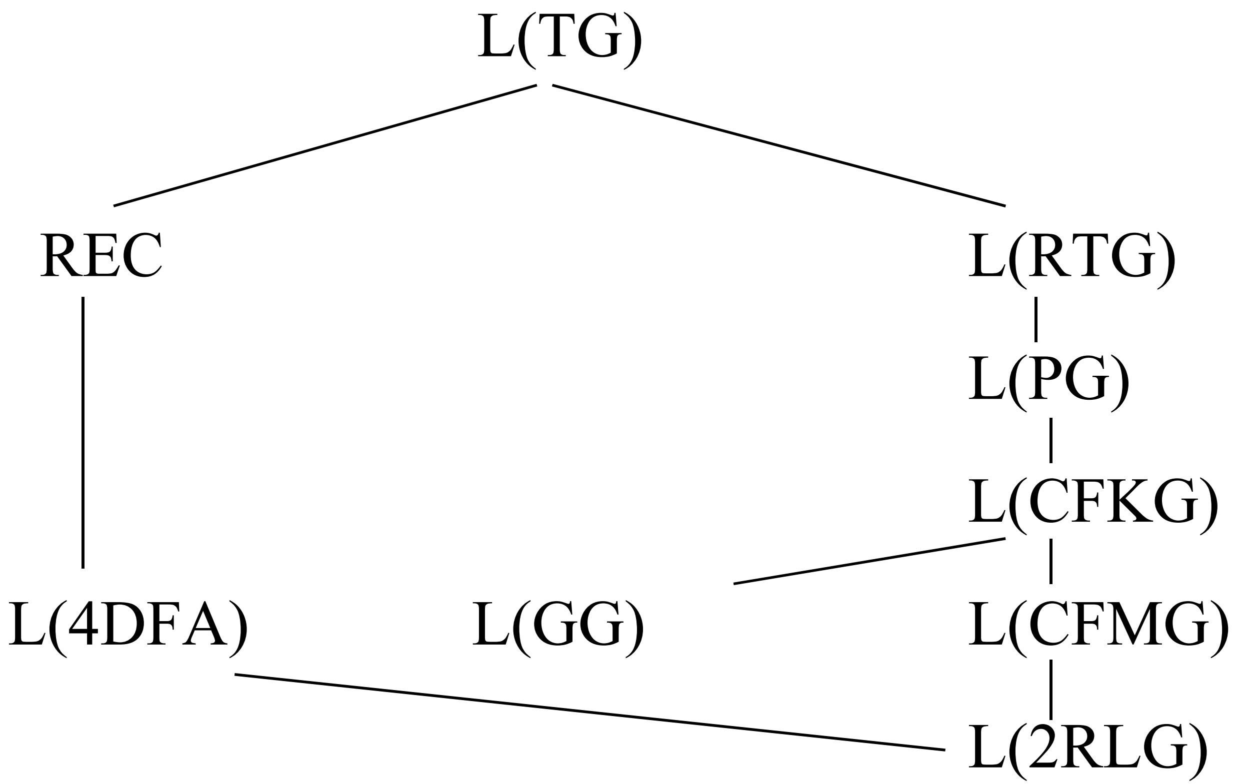
$$A_H \rightarrow A \obar t,\ A_V \rightarrow t \ominus A_V \mid t$$

$$A \rightarrow [B_{11},\ldots B_{1k},\ldots,B_{k1},\ldots,B_{kk}]$$

$$\Longrightarrow A \rightarrow \ominus_{1 \le i \le k}(\obar_{1 \le j \le k} B_{ij})$$

- $L(GG)$ incomparable with $L(CGMG)$

Another taxonomy



Some remarks and open problems

- If REC is the notion corresponding to regular string languages then, to mantain hierarchy, TG grammars is the notion corresponding to context free grammars.
- RTG is a nice model, that includes several well known models usually introduced as a generalization of context free grammars
- If RTG is the right model for generating context free picture languages; what about the right model for regular string languages? Which languages are defined by non recursive TRG grammars?

Some remarks and open problems

How define automata recognizing the same families languages generated by the introduced grammars?

Are there more promising grammatical approaches to "context free" picture languages? Up to now grammars with two set of rules or isometric grammars.

What is the picture language corresponding to 1D Dyck language?