Splicing systems

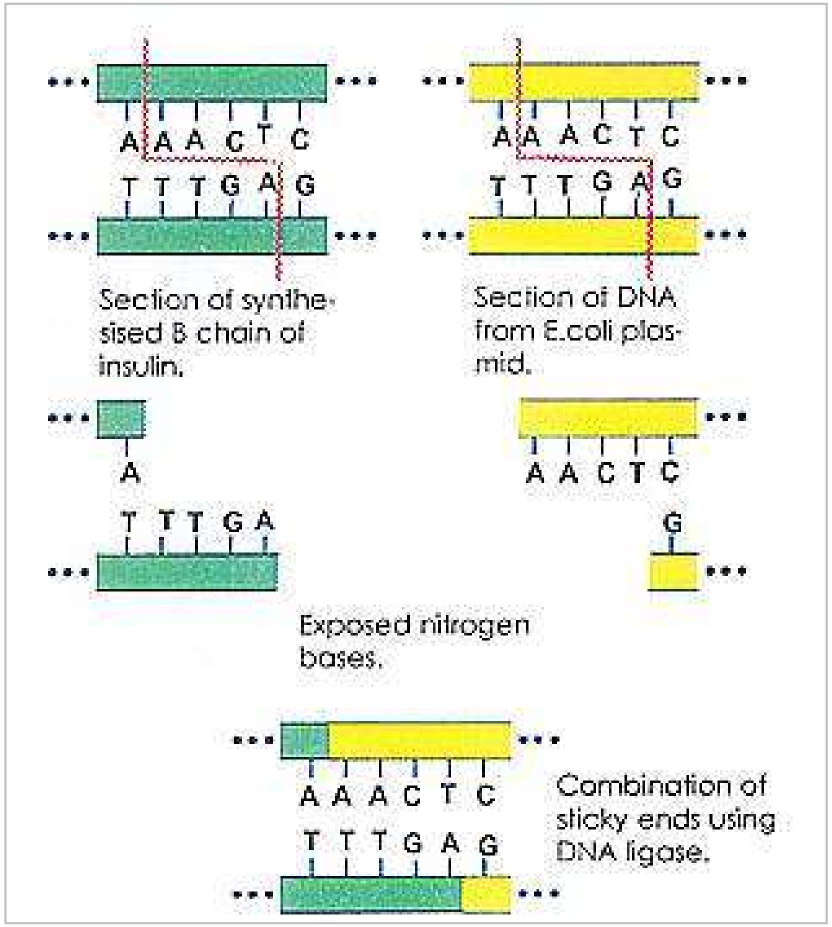
Clelia De Felice

Dipartimento di Informatica e Applicazioni, University of Salerno, ITALY

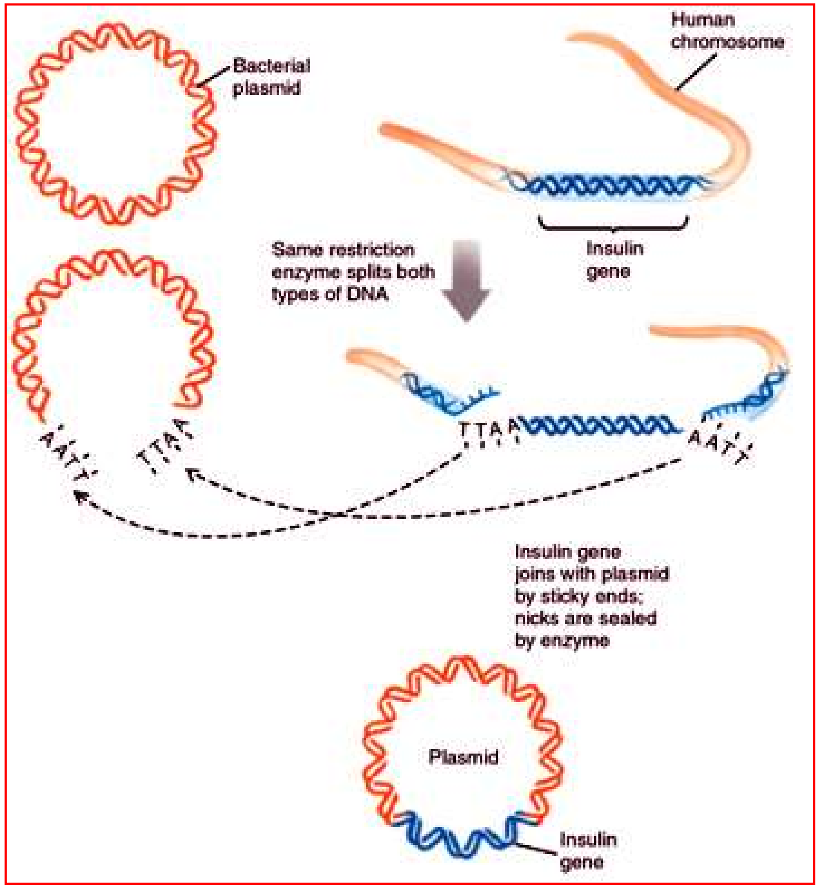
Colloquium "Unconventional Models of Computation"

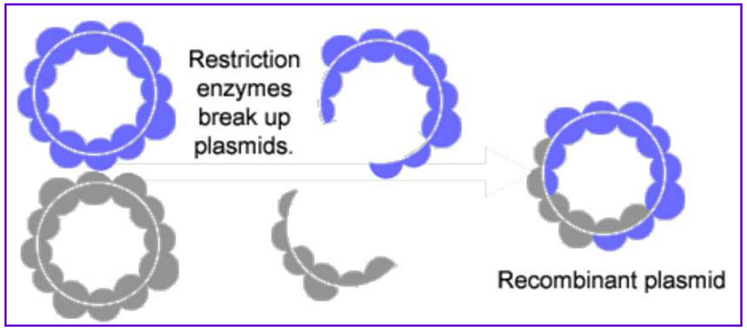
in honour of Giancarlo Mauri – Cremona, September 28, 2009

THE MECHANISM



Source: Watson, J.D., Gilman, M., Witkovski., Zoller, M. Recombinant DNA, pg 78.





HIGHLIGHTS

Linear case

1) Strictly locally testable languages are exactly languages generated by null context splicing systems [Head 1987]

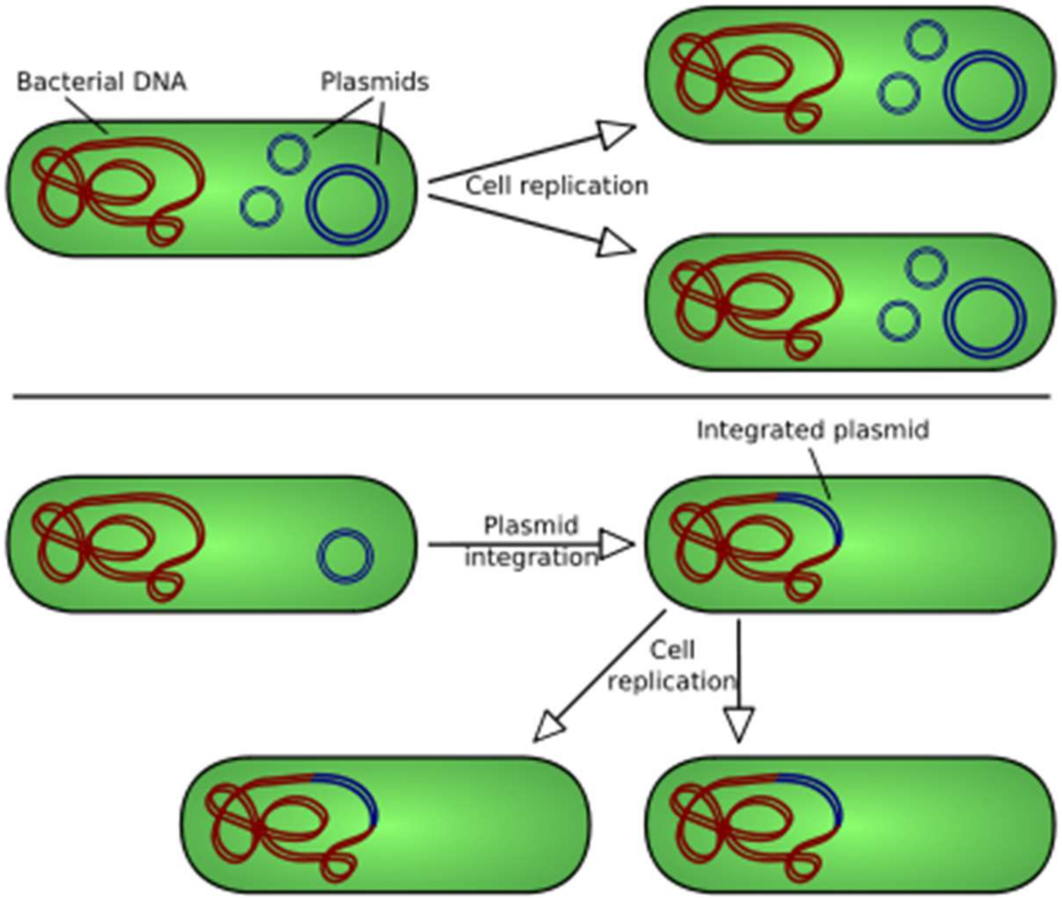
2) reflexive languages are constructed by constants

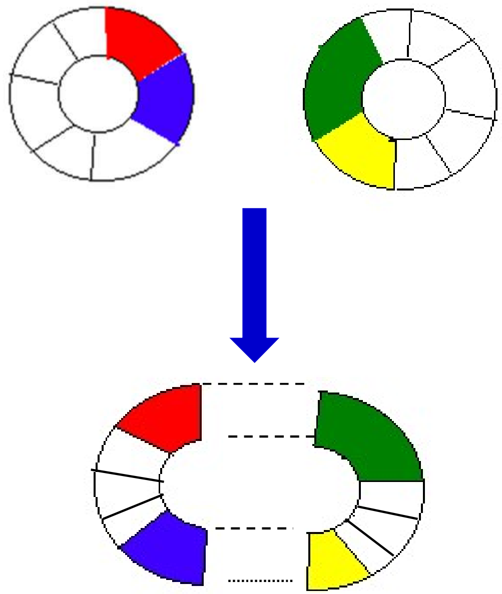
Circular case: the relation between

1) complete systems and pure unitary languages (in particular, free monoids generated by group codes)

2) marked systems and P4-free graphs.

CIRCULAR SPLICING





THE MODEL

$^{\sim}aaab$ = circular word

- $^{\sim}w$ = equivalence class of w w.r.t. the <u>*conjugacy relation*</u>
- ***Circular language L***: set of circular words

$w, w' \in A^*$, $w \sim w'$ $\Leftrightarrow$ $w=xy$, $w' = yx$

(linear) $\mathbf{X} \xrightarrow{\varphi} \mathbf{{}^{\sim}X}$ (circular)

Example: $X = \{abb, bba, bb\}$, $^{\sim}X = \{^{\sim}abb, ^{\sim}bb\}$

(circular)

Lin(L) = Full linearization of L $\rightleftarrows$ **L** (linear)

Example: $L=\{\sim abb, \sim bb\}$, $Lin(L) = \{abb, bba, bab, bb\}$

Definition

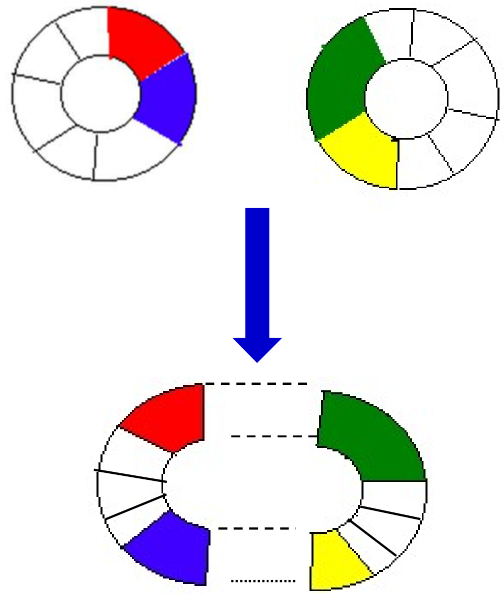
A circular language L is *regular* (resp. *context-free*) if Lin(L) is *regular* (resp. *context-free*).

$\Rightarrow$ languages closed under the conjugacy relation

PAUN CIRCULAR SPLICING SYSTEMS

$S = (A,I,R)$, A = finite alphabet, I = initial circular language $\subseteq {}^{\sim}A^*$,
R = set of the rules $r = u_1 \# u_2 \$ u_3 \# u_4$, $u_i \in A^*$, $\#, \$ \notin A$.

$({}^{\sim}w', {}^{\sim}w'') \vdash_r {}^{\sim}w$ if $w' = u_2 x u_1$, $w'' = u_4 y u_3$, $w = u_2 x u_1 u_4 y u_3$



Definition [Head, Paun, Pixton, Handbook of Formal Languages, Vol. 2, 1996]

$L(S)$ = circular splicing language ***generated*** by $S = (A,I,R)$ is the smallest language which contains I and is invariant under iterated splicing by rules in R.

Problem 1

The computational power of circular splicing systems

Problem 2

Finding a characterization of the class of (regular) circular languages generated by circular splicing systems

Problem 3

Given S =(A,I,R), is it decidable whether L(S) is regular?

Problem 4

Given a regular circular language L, is it decidable whether there exists S =(A,I,R) such that L=L(S)?

STATE OF THE ART

✓ ***Circular splicing systems***

Results [Paun, Handbook of Formal Languages], [Pixton, TCS, 2000]: S=(A,I,R)

- I regular circular, **additional** conditions on R, **self-splicing** $\Rightarrow$ L(S) regular circular
- Lin(I) regular $\Rightarrow$ Lin(I) in a full AFL closed under conjugacy relation

✓ ***Finite circular splicing systems*** S =(A,I,R) with I, R **finite** sets

Example: $\tilde{}\{w \in A^* \mid |w|=2n, n\geq 0\}$ is generated by a finite splicing system

$\tilde{}\{w \in \{a,b\}^* \mid |w|_a=2n, |w|_b = 2m, n,m \geq 0\}$ is generated by a finite splicing system

$\tilde{}((aa)^*b)$ is regular and cannot be generated by a finite splicing system

$\tilde{}\{ a^n b^n \mid n,m \geq 0\}$ is context-free and is generated by a finite splicing system

Results: L(S) may be regular, context-free, context-sensitive [Fagnot, JM04]

Case A ={a} [Bonizzoni, DeF, Mauri, Zizza, Rairo 2004, DAM 2005] *:*The class of regular circular languages generated has been characterized. It is decidable whether L(S) is regular

Special families of circular languages [B, DF, M, Z, Rairo 2004] :

X* closed under conjugacy relation, X regular, X* cycle closed (for each simple cycle c in m.DFA of X*, c in X*) $\Rightarrow$ ~X* generated by finite systems.

Example group codes.

✓ *CSSH systems*

Rules: a # 1 $ b # 1 or 1 # a $ 1 # b or 1# a $ b # 1 or a # 1 $ 1 # b, with a, b letters

Example S = (A,I,R), A = {a,b,c }, I = ~{aac, bcba }, R = { c # 1 $ b # 1, 1 # a $ b # 1}

Result: L(S) context-free [Fagnot, JM04]

(1,3)-CSSH systems: S=(A,I,R), R = {a #1 $ b # 1| a,b $\in$ A}

TWO PARTICULAR CASES
of (1,3)-CSSH systems

Complete systems

Marked systems

✓ *Complete systems*: S =(A,I,R), alph(I) = A, I finite

$$R = \{a \# 1 \$ b \# 1 \mid a, b \in A\} = A \times A$$

Example S=({a,b,c}, ~{ac, bb}, {a#1$a#1, a#1$b#1, a#1$c#1, b#1$b#1, c#1$c#1, b#1$c # 1})

Remark Splicing = concatenation + closure under conjugation

$$(\sim xa, \sim yb) \vdash r \sim xayb$$

AN OLD PROBLEM...

CONDITIONS UNDER WHICH A CONTEXT-FREE LANGUAGE L IS REGULAR: PROPERTIES OF A CONTEXT-FREE GRAMMAR G WITH L(G)=L

- properties of L [Autebert, J. Beauquier, Boisson, Latteaux, Nivat, Ehrenfeucht, Haussler, Rozenberg,… 80's]

- properties of a context-free grammar G with L(G)=L [Ehrenfeucht, Haussler, Rozenberg, 80's]

✓ ***Example*** **(Dyck languages)**

$G = (\{X\}, A, P, X)$, $A = \{a,b\}$,

$P = \{ X \to 1,\ X \to XaXbX,\ X \to XbXaX \}$

$$L(G) = \{w \text{ in } A^* \mid |w|_a = |w|_b \}$$

✓ ***Example*** **(generalized Dyck languages)**

$Y = \{aba, c\}$

$P = \{ X \to 1,\ X \to XcX,\ X \to XaXbXaX\}$

(In the previous example take $Y = \{ab, ba\}$)

Definition $w = a_{i_1} \ldots a_{i_h} \in Y$, $a_{i_j} \in A$, $p_w = X \to X\, a_{i_1}\, X\, a_{i_2} \ldots X\, a_{i_h}\, X$

$\mathbf{G_Y = (\{X\}, A, P, X)}$,

$P = \{ X \to 1 \} \cup \{p_w \mid w \in Y \}$

STRUCTURE OF L(G)?

INSERTION [Haussler, Inf. Sci. 1983]

$$Z \leftarrow Y = \{w_1\, y\, w_2 \mid w_1 w_2 \in Z,\ y \in Y\}$$

Example

$Y = \{aba, c\}$, $Y^{\leftarrow 2} = Y \leftarrow Y = \{caba, acba, abca, abac,$
$cc, aababa, ababaa, abaaba\}$

ITERATED INSERTION [Haussler, Inf. Sci. 1983]

$$Y^{\leftarrow 0} = 1\ ,\quad Y^{\leftarrow n} = Y^{\leftarrow n-1} \leftarrow Y,\quad Y^{\leftarrow *} = \cup_{n \geq 0}\ Y^{\leftarrow n}$$

Example

$Y = \{aba, c\}$, $Y^{\leftarrow *} = \{1, aba, c, caba, acba, abca, abac,$
$cc, aababa, ababaa, abaaba,$
$ccaba, cabaaba\ldots\}$

✓ ***Pure unitary languages*** **(GENERALIZED DYCK LANGUAGES)**

Definition [Ehrenfeucht, Haussler, Rozenberg, TCS 1983]

L is a *pure unitary language* $\Leftrightarrow \exists\, Y \subseteq A^*$, Y finite: $L = L(G_Y)$.

- **Theorem 1** **[Ehrenfeucht, Haussler, Rozenberg, TCS 1983]**
  L is a pure unitary language, i.e., $\exists$ Y finite: L =L(G_Y) $\Leftrightarrow$ $\exists$ Y finite: L = $Y^{\leftarrow *}$

- **Theorem 2** **[Ehrenfeucht, Haussler, Rozenberg, TCS 1983]**

A pure unitary language L = $Y^{\leftarrow *}$ is regular $\Leftrightarrow$ Y is *subword unavoidable* in alph(Y)*
(i.e. $\exists$ k s.t. each word w $\in$ alph(Y)*, with |w| $\geq$ k has a word of Y as a factor)

  - (well quasi-orders)

- **Proposition** **[Ehrenfeucht, Haussler, Rozenberg, TCS 1983]**

For any regular set R $\subseteq$ A*, it is decidable whether or not R is subword unavoidable

COMPLETE SYSTEMS AND PURE UNITARY LANGUAGES

✓ *REMIND. Complete systems*: $S = (A, I, R)$, $alph(I) = A$, I finite

$$R = \{a \# 1 \$ b \# 1 \mid a, b \in A\} = A \times A$$

Theorem **[DCM'09]** $L = L(S)$ with S *complete* system $\Leftrightarrow$
$\exists$ Y finite and closed under the conjugacy relation : $Lin(L) = L(G_Y) \setminus 1$ $\Leftrightarrow$
$\exists$ Y finite and closed under the conjugacy relation : $Lin(L) = Y^{\leftarrow *} \setminus 1$

Corollary **[DCM'09]** $S = (A, I, R)$ *complete* system.

- $L(S)$ is context-free
- $L(S)$ is a regular circular language $\Leftrightarrow$ $Lin(I)$ is subword unavoidable.
- It is decidable whether $L(S)$ is a regular circular language.

- ✓ *(1,3)-simple systems:* S=(A,I,R), R = {a #1 \$ a # 1| a $\in$ B $\subseteq$ A}
  - a # 1 \$ a # 1 = (a,a)

- ✓ *(1,3)-simple systems with 1 rule:* R = {(a,a)}

Complete systems $\Leftrightarrow$ (1,3)-*simple systems* with 1 rule

by means of

φ: S → S' , S (1,3)-simple system with 1 rule, S' complete

Example S=(A,I,R), I =~{baca}, R = {(a,a)} →
S'=(A,I',R'), I' = ~{de}, R'= {d#1\$d#1, e#1\$e#1, d#1\$e#1}

Theorem [Siromoney, Subramanian, Dare, ICPIA , LNCS 654, 1992]
(1,3)-simple system S $\Rightarrow$ L(S) regular

FALSE

Corollary [DCM'09]

- ➢ The class of regular circular languages generated by *(1,3)-simple systems S with 1 rule* has been characterized.

- ➢ *(1,3)-simple systems S with 1 rule :* it is decidable whether L(S) is regular

✓ ***Marked systems*** **[DeF, Fici, Zizza, FCT 2007]**

(1,3)-CSSH systems S=(A,I,R), R $\subseteq$ A x A, **I = A = alph(R)**

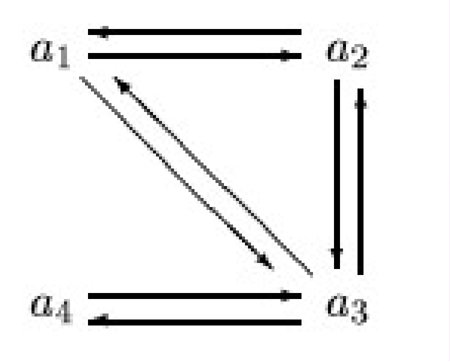
Example S = (A, I,R), A = I = {c, b}, R = {(c, b)}

- ✓ The class of regular circular languages generated by S has been characterized.
  - ✓ It is decidable whether L(S) is regular

[Bonizzoni, DeF, Fici, Zizza, to appear in Nat. Comp., 2009]

- Extended to marked systems with self-splicing
- Reviewed in a graph theoretical setting

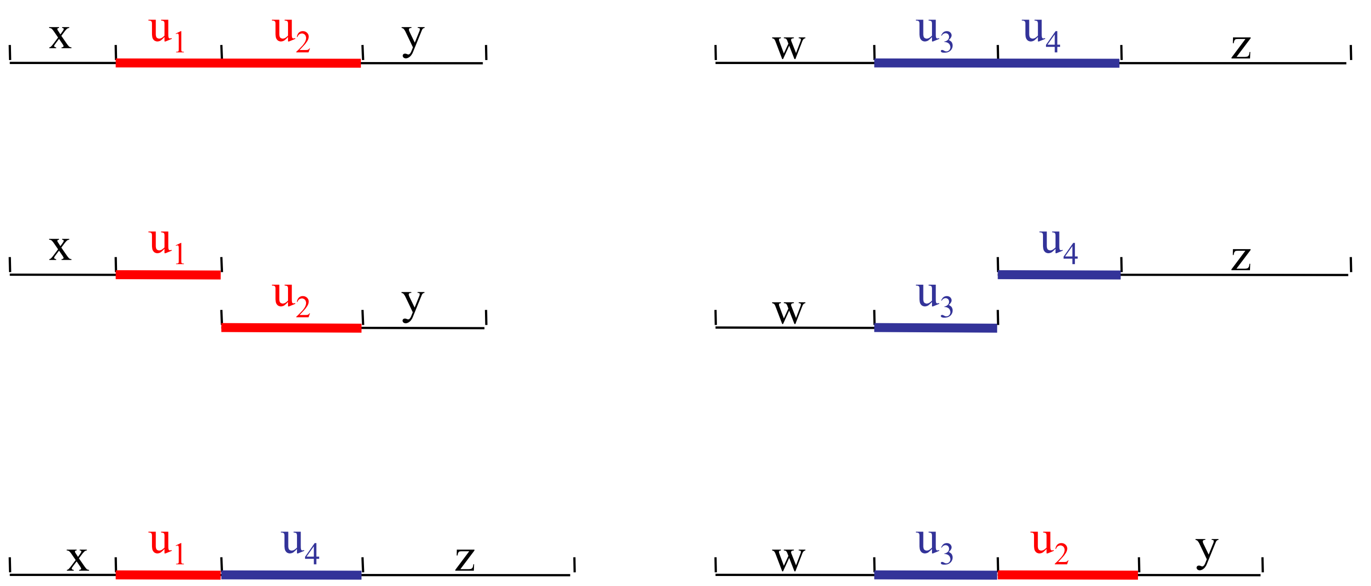
A marked system generates a regular circular language
if and only
if its graph is P_4-free



PAUN'S LINEAR SPLICING (1996)

$r = u_1 \mid u_2 \;\$\; u_3 \mid u_4$ rule

σ: $(x\, u_1 u_2\, y,\;\; w u_3 u_4\, z)$ $\quad$ $(x\, u_1\, u_4\, z\,,\;\; w u_3\, u_2\, y)$



Pattern recognition

cut

paste

✓ ***Paun's linear splicing system*** **S_{PA} = (A, I, R)**

A=finite alphabet; $I \subseteq A^*$ **initial language;** $R \subseteq A^*|A^*\$A^*|A^*$ set of rules;

$L(S_{PA}) = I \cup \sigma(I) \cup \sigma^2(I) \cup \ldots = \cup_{n \geq 0} \sigma^n(I)$ **splicing language**

✓ ***Finite linear splicing system*** S =(A,I,R) with I, R **finite** sets

(aa)* is regular and cannot be generated by a finite splicing system

Problem 1

Characterize regular languages
generated by finite linear Paun splicing systems

Problem 2

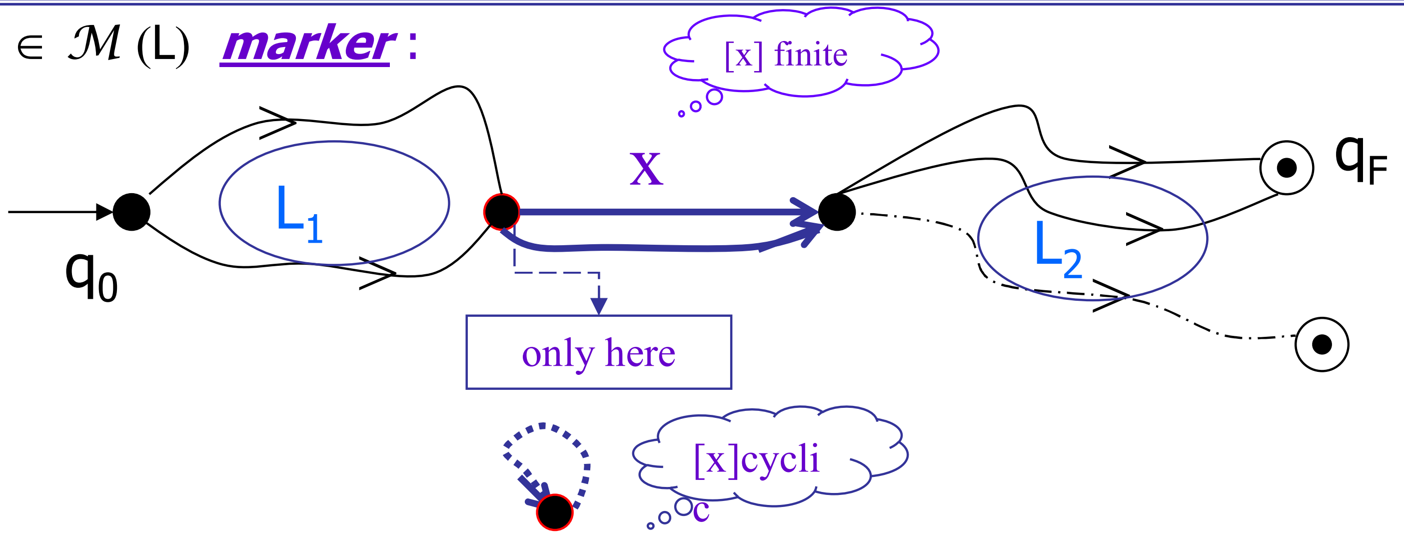
Given L regular, can we decide whether L is generated by a finite linear Paun splicing system?

PARTIAL RESULTS

✓ ***Marker Languages*** **[Bonizzoni, De Felice, Mauri, Zizza, DAM 2005]**

Let L be a regular language, $\mathcal{A} = (A, Q, \delta, q_0, F)$ minimal for L. Consider the Syntactic Congruence (w.r.t. L) $\equiv_L$

- $[x] \in \mathcal{M}(L)$ ***marker*** :



- Denote $[x]_1 = [x]$ if $[x]$ is finite, otherwise $[x]_1 = [x] \cup 1$

Theorem $L([x]) = L_1 [x]_1 L_2$ is a finite splicing language

Marker Language associated with [x]

✓ ***Reflexive Languages*** **[Bonizzoni, D.F., Mauri, Zizza, DLT03; Bonizzoni, D.F., Zizza, TCS 2005]**

The characterization of *reflexive Paun* splicing languages

Generated by finite splicing systems which are "reflexive", i.e.,

$u_1 \# u_2 \ \$ \ u_3 \# u_4 \in R \Rightarrow u_1 \# u_2 \ \$ \ u_1 \# u_2 \ , \ u_3 \# u_4 \ \$ \ u_3 \# u_4 \in R$

by means of

- finite set of (Schutzenberger) constants C
- finite set of factorizations of these constants into 2 words

✓ **As a consequence we have:**

The characterization of Head splicing languages, since

Finite Head splicing system → Finite Paun splicing system, reflexive and symmetric

✓ **It is decidable** whether a regular language is generated by a reflexive splicing system **[Goode,Pixton 2007], [Bonizzoni, Mauri 2005], [Bonizzoni 2009]**