Unavoidable Sets and Regularity of Languages Generated by (1,3)-Circular Splicing Systems

Clelia De Felice, Rocco Zaccagnino, Rosalba Zizza

Department of Computer Science

University of Salerno

TPNC 2014

Our aim

Focus on the still unknown relations between regular languages and CSSH systems.

Prove of necessary conditions for (1,3)-CSSH systems generating regular languages

Hybrid systems: relation between the regularity of the languages generated and the notion of unavoidability introduced by Ehrenfeucht, Haussler and Rozenberg in 1983.

Circular Splicing Systems and Open questions

Formal model of a generative mechanism of circular words inspired by a recombinant behaviour of circular DNA (Head 1992).

Defined by a finite alphabet A, an initial set I of circular words and a set R of rules.

Open questions:

- Decide whether a splicing language is regular
- Decide whether a regular circular language is a splicing language
- Characterize the structure of the regular splicing circular languages

Our Plan

- Background
- Outline of results
- Regular Sets and Unavoidable Sets
- Our approach: Hybrid Systems
- Future perspectives

The model: circular words and languages

Circular word

$^{\sim}w$ = equivalence class of w under *coniugacy relation* $\sim$

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

Example: $^{\sim}aaab = \{aaab, aaba, abaa, baaa\}$

Circular language

Circular language L = set of circular words.

$$\text{(linear) } X \rightarrow {}^{\sim}X = \{{}^{\sim}w \mid w \in X\} \text{ (circular)}$$

$$\text{(circular) } L \rightarrow Lin(L) = \{w \in A^* \mid {}^{\sim}w \in L\} \text{ (linear)}$$

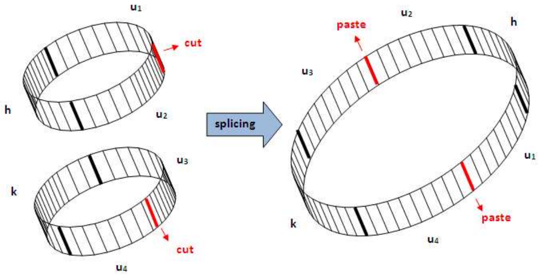
A circular language L is regular, i.e., $L \in Reg^{\sim}$, if and only if $Lin(L)$ is *regular*. (Similarly, $CF^{\sim}$, $CS^{\sim}$).

The model: circular splicing operation

Păun circular splicing operation

Let $r = u_1\#u_2\$u_3\#u_4 \in R$.

$(\sim w', \sim w'') \vdash_r \sim w$ se $w' = u_2 h u_1, w'' = u_4 k u_3, w = u_2 h u_1 u_4 k u_3$



Example:

$r = a\#b\$c\#d$ *and* $w' = bdda, w'' = daac, w = bddadaac.$

Then, $(\sim w', \sim w'') \vdash_r \sim w.$

The model: circular splicing systems

Păun Circular splicing system (1996)

Triple $S = (A, I, R)$:

- A finite alphabet
- I *initial* circular language ($I \subseteq \sim A^*$)
- R set of *rules* ($R \subseteq A^* \# A^* \$ A^* \# A^*$ con $\#, \$ \notin A$)
    - symmetric: $\forall u_1 \# u_2 \$ u_3 \# u_4 \in R$, $u_3 \# u_4 \$ u_1 \# u_2 \in R$.

Circular language generated by $S = (A, I, R)$

$L(S) =$ the smallest circular language that contains I and closed under iterated applications of the rules in R .

Finite Circular Splicing languages

C(Fin,Fin)= class of circular languages generated by finite circular splicing systems (I, R finite sets).

Results:

- Unlike the linear case, $C(Fin, Fin)$ is not comparable with $Reg^{\sim}$.
  Examples:
  $^{\sim}(aa)^* \in Reg^{\sim}$ *is a finite splicing circular language.*
  $^{\sim}a^n b^n \in CF^{\sim}$ *is a finite splicing circular language.*
- $C(Fin, Fin) \subset CS^{\sim}$ [Berstel, Boasson, Fagnot (2012)].
- Particular cases: *alphabetic, semi-simple (CSSH), marked, complete monotone* [see next slides].

Alphabetic systems and CSSH

Alphabetic systems [Berstel, Boasson, Fagnot (2012)]

$S = (A, I, R)$ is alphabetic if for each rule $u_1 \# u_2 \$ u_3 \# u_4 \in R$:

$$|u_i| \leq 1, 1 \leq i \leq 4$$

Semi-simple splicing systems (CSSH) [Ceterchi, Martín-Vide, Subramanian (2004)]

$S = (A, I, R)$ finite is a CSSH if for each $u_1 \# u_2 \$ u_3 \# u_4 \in R$:

$$u_1 u_2, u_3 u_4 \in A$$

If $u_1 u_2 = u_3 u_4 \in A$ then S is *simple*.

Known results

Alphabetic systems [Berstel, Boasson, Fagnot (2012)]

If I is context-free, then $L(S)$ is context-free.

Marked systems [De Felice, Fici, Zizza (2009)]

CSSH systems, $I = A = alph(R)$ and rules $a\#1\$b\#1$ ($(1,3)$ type).

- characterization of the structure of $L(S)$
- decidable to estabilish whether $L(S)$ is regular

Complete monotone systems [Bonizzoni, De Felice, Zizza (2010)]

CSSH systems, $alph(I) = A$, $R = \{a\#1\$b\#1 | a, b \in A\} = A \times A$.

- characterization of the systems that generate regular splicing circular languages
- decidable to estabilish whether $L(S)$ is regular

Regular Languages and Unavoidable Sets

Unavoidable Set [Ehrenfeucht, Haussler, Rozenberg (1983)]

Let $X, Y \subseteq A^*$. Y is *unavoidable* in X if $\exists k_0 \in N$ such that $\forall x \in X$, with $|x| > k_0$, $\exists y \in Y$ such that $x = x_1 y x_2$.

Examples:
$Y = \{aa, b\}$ is unavoidable in A^*.
$Y = \{aa, ab\}$ is avoidable in A^*. In fact, Y is avoidable in b^*.

- **A complete $(1,3)$-CSSH system generates a regular circular language if and only if $Lin(I)$ is unavoidable [Bonizzoni, De Felice, Zizza (2010)].**

Generalization of a result in [Ehrenfeucht, Haussler, Rozenberg (1983)]

Let $S = (A, I, R)$ be a $(1,3)$-CSSH system. If $L(S)$ is regular then $Lin(I)$ is unavoidable in $Pref(Lin(L(S)))$.

Hybrid Systems

Definition (Hybrid System)

(1, 3)-CSSH system $S = (A, I, R)$ such that:

- $A = alph(R) = alph(I)$
- For any $a_i, a_j \in A$ there are $b_1, \ldots, b_k$, with $k \geq 2$ such that: $b_1 = a_i$, $b_k = a_j$ and for all $h \in \{1, \ldots, k-1\}$, $(b_h, b_{h+1}) \in R$.

Examples:

- *The* $(1, 3)$*-CSSH system* $S = (A, I, R)$*, with* $A = \{a, b\}$*,* $I = {\sim}\{ab, a, b\}$ *and* $R = \{(a, b)\}$*, is hybrid.*
- *The* $(1, 3)$*-CSSH system* $S' = (A, I', R')$*, with* $I = {\sim}\{ab, a\}$ *and* $R' = \{(a, a)\}$*, is not hybrid.*

Hybrid Systems

Proposition

If $S = (A, I, R)$ is a hybrid system, then $Pref(Lin(S)) = A^*$.

Corollary (1)

Let $S = (A, I, R)$ be a hybrid system, let $Y = Lin(I)$. If $L(S)$ is a regular circular language, then Y is unavoidable.

Corollary (2)

Let $S = (A, I, R)$ be a hybrid system. If $L(S)$ is a regular circular language, then for any $a \in A$ there exists $n \in N$ such that $a^n \in I$.

A necessary condition for Regularity

Proposition

Let $S = (A, I, R)$ be a $(1, 3)$-CSSH, let $a \in A$. If $(a, b) \in R$ for all $b \in A \setminus \{a\}$ and $Lin(I) \cap a^* = \emptyset$, then $L(S)$ is not regular.

Future Perspectives

Conjecture

Let $S = (A, I, R)$ be a hybrid system, let $Y = Lin(I)$. $L(S)$ is regular if and only if $d(R) < 4$ and Y is unavoidable, where $d(R)$ is the diameter of R.

Example 1:
The $S = (A, I, R)$ *with* $A = \{a, b\}$, $I = {}^{\sim}\{ab, a, b\}$ *and* $R = \{(a, b)\}$.

- $d(R) < 4$
- $Lin(I)$ *is unavoidable.*

$$Lin(L(S)) = \{a, b\}^+ \setminus (a^+a \cup b^+b) \in Reg$$

Example 2:
The $S = (A, I, R)$ *with* $A = \{a, b\}$, $I = {}^{\sim}\{ab\}$ *and* $R = \{(a, b)\}$.
Observe that $d(R) < 4$ *but* $Lin(I)$ *is avoidable.*

$$Lin(L(S)) \cap a^* b^* = \{a^n b^n | n > 0\} \notin Reg$$