

# Automata, algebras and logic for languages of labeled birooted trees

David Janin

GT Graphe et logique - 26th of march



# Plan of the talk

- 1 Labeled birooted trees
- 2 Automata :  $\text{MSO}^\uparrow$
- 3 Algebra :  $\text{BOOL}(\text{MSO}^\uparrow)$
- 4 Expressiveness :  $\text{BOOL}(\text{MSO}^\uparrow) \subset \text{MSO}$
- 5 Conclusion

# 1. Labeled birooted trees

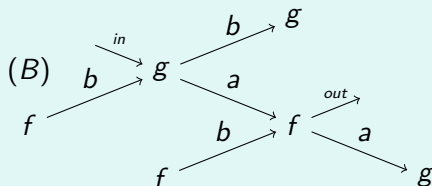
Within Free inverse monoids

## Labeled Munn's trees (or birooted $F$ -trees)

Let  $F$  be a finite vertex alphabet, let  $A$  be a finite edge alphabet  $A$ .

### Birooted $F$ -tree

A finite directed tree-shaped graph with  $F$ -labeled vertices and  $A$ -labeled edges, both forward and backward deterministic,



with distinguished input root and output output root.

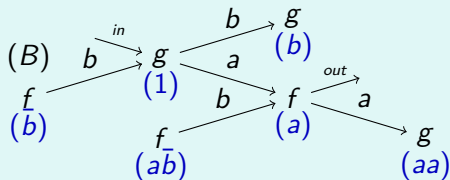
## Labeled Munn's trees (or birooted $F$ -trees)

The **free group**  $FG(A)$  generated by  $A$  ordered by the (reduced) **prefix order**, e.g.  $1 = a\bar{a} \leq_p a = a\bar{b}b \leq_p a\bar{b} \leq_p a\bar{b}a$ .

### Birooted $F$ -tree

A pair  $B = \langle t, u \rangle$  with :

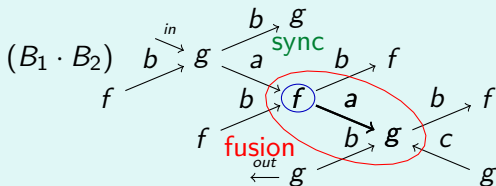
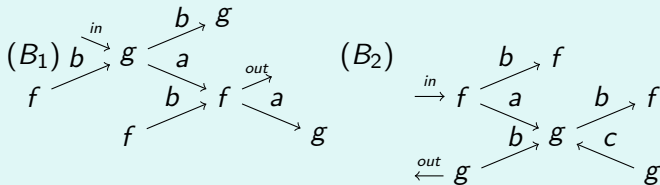
- $t : FG(A) \rightarrow F$  a finite labeled tree with prefix closed  $dom(t)$ ,
- input root  $1 \in dom(u)$  and output root  $u \in dom(t)$ .



Here:  $B = \langle t, u \rangle$  with  $dom(t) = \{1, b, \bar{b}, a, \bar{a}, aa\}$  with  $u = a$ .

## Birooted $F$ -trees product

Product = sync + fusion



or the *undefined birooted tree*  $0$  when fusion fails.

## Birooted $F$ -trees product

Let  $B_1 = \langle r, u \rangle$  and  $B_2 = \langle s, v \rangle$  be two birooted  $F$ -trees.

Let  $r_u$ , the  $u$ -translation of  $r$ , be defined by  $\text{dom}(r_u) = \bar{v} \cdot \text{dom}(r)$  and, for every  $x \in \text{dom}(r_u)$  by  $r_u(x) = r(ux)$ .

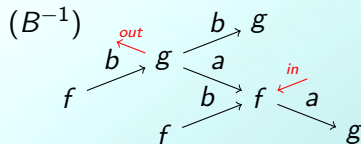
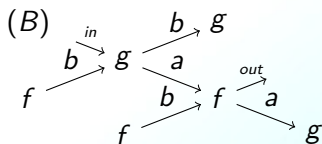
### Product $B_1 \cdot B_2$

- If  $r_u$  and  $s$  agree on  $\text{dom}(r_u) \cap \text{dom}(s)$  then  $B_1 \cdot B_2 = \langle t, w \rangle$  defined by  $w = uv$ ,  $\text{dom}(t) = \text{dom}(r) \cup u \cdot \text{dom}(s)$  with:
  - for every  $x \in \text{dom}(r)$ ,  $t(x) = r(x)$
  - for every  $y \in \text{dom}(s)$ ,  $t_u(y) = s(y)$ .
- Otherwise,  $B_1 \cdot B_2 = 0$  the undefined birooted  $F$ -tree.

# An inverse semigroup

## Semigroup/Monoid

The set  $\mathcal{B}(F)$  of birooted  $F$ -tree is a semigroup.  
 When  $F$  is a singleton, it is isomorphic to  $FIM(A)$ .  
 Otherwise, it is extended with a unit 1.



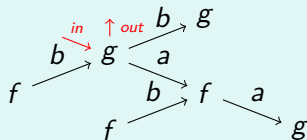
## Lemma (Inverses)

For every  $B \in \mathcal{B}(F)$  there exists a unique  $B^{-1} \in \mathcal{B}(F)$  such that  $B \cdot B^{-1} \cdot B = B$  and  $B^{-1} \cdot B \cdot B^{-1} = B^{-1}$ .

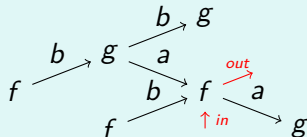
## Left and right projections

Elements  $B \cdot B^{-1}$  and  $B^{-1} \cdot B$  are idempotents

With  $B = \langle t, u \rangle$  we have  $B^R = B \cdot B^{-1} = \langle t, 1 \rangle$



and we have  $B^L = B^{-1} \cdot B = \langle t_u, 1 \rangle$



with  $B^R \cdot B = B \cdot B^L = B$ .

## The natural order

### Definition (Natural Order)

For every  $B_1$  and  $B_2$ , let  $B_1 \leq B_2$  when  $B_1 = B_1 B_1^{-1} B_2$  (equivalently  $B_1 = B_2 B_1^{-1} B_1$ ).

### Lemma

For every non trivial  $B_1 = \langle t_1, u_1 \rangle$  and  $B_2 = \langle t_2, u_2 \rangle$ , we have

$$B_1 \leq B_2 \text{ if and only if } \text{dom}(t_1) \supseteq \text{dom}(t_2), u_1 = u_2 \text{ and } t_1(v) = t_2(v) \text{ for every } v \in \text{dom}(t_2)$$

i.e.  $B_1 \leq B_2$  when  $B_2$  is a sub birooted  $F$ -tree of  $B_1$  with same roots.

### Lemma (Idempotents)

The idempotents are the subunits  $U(\mathcal{B}(F)) = \{B \in \mathcal{B}(F) : B \leq 1\}$ .

# Languages of birooted $F$ -trees

## Operations on languages

Given  $M$  and  $N \subseteq \mathcal{B}(F) - 0$ :

- **sum** :  $M + N = M \cup N$ ,
- **product** :  $M \cdot N = \{B \cdot C : B \in M, C \in N, B \cdot C \neq 0\}$ ,
- **star** :  $M^* = \bigcup_{n \in \mathbb{N}} M^n$  with  $M^0 = \{1\}$  and  $M^{n+1} = M \cdot M^n$ ,
- **inverse** :  $M^{-1} = \{B^{-1} : B \in M\}$ ,
- **context projection** :  $M^E = \{B : B \in M, B \cdot B = B\}$ ,
- **left and right projection** :  $M^L = \{B^{-1} \cdot B : B \in M\}$  and  $M^R = \{B \cdot B^{-1} : B \in M\}$ ,
- **others** ...

with upward or downward closed variants.

## Some known facts

### Lemma

*Inverse operator commutes with sum, product, star, context projection and, reversing them, left and right projections.*

### Lemma

$$REC \subset RAT \subset 1-RAT \subseteq \dots \subseteq k-RAT \subseteq \dots \subseteq \bigcup_{k \in \mathbb{N}} k-RAT \subseteq MSO$$

*where  $k$  in  $k-RAT$  is the maximum allowed nested-depth of context projection operators.*

### Question

Strictness of that hierarchy ? What about the downward or upward closed variants ? The extension with (finite) boolean closure ?

## 2. Automata : $\text{MSO}^\uparrow$

Automata for upward closed MSO definable languages

## Birooted $F$ -tree automata

Automaton :  $\mathcal{A} = \langle Q, \delta, \Delta, W \rangle$

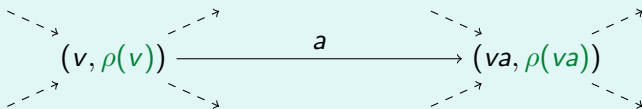
with a (finite) set of states  $Q$ , a **state table**  $\delta : F \rightarrow \mathcal{P}(Q)$ , a **transition table**  $\Delta : A \rightarrow \mathcal{P}(Q \times Q)$  and an **acceptance condition**  $W \subseteq Q \times Q$ .

Run of  $\mathcal{A}$  on  $B = \langle t, u \rangle : \rho : \text{dom}(t) \rightarrow Q$

with, for every  $v \in \text{dom}(t)$ :

- **state coherence** :  $\rho(v) \in \delta(t(v))$ ,

and for every for every  $a \in A$  such that  $va \in \text{dom}(t)$ :

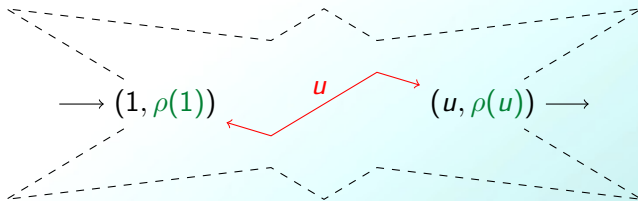


- **transition coherence** :  $(\rho(v), \rho(va)) \in \Delta(a)$ .

## Tile automata languages

### Accepting runs

Run  $\rho : \text{dom}(t) \rightarrow Q$  of  $\mathcal{A}$  on  $B = \langle t, u \rangle$  is **accepting** when  $(\rho(1), \rho(u)) \in W$ .



### Recognized language $L(\mathcal{A})$

The language  $L(\mathcal{A})$  is defined as the set of (non zero) birooted  $F$ -tree  $B \in \mathcal{B}(F)$  such that there is a an accepting run of  $\mathcal{A}$  on  $B$ .

## An example: marking the path between roots

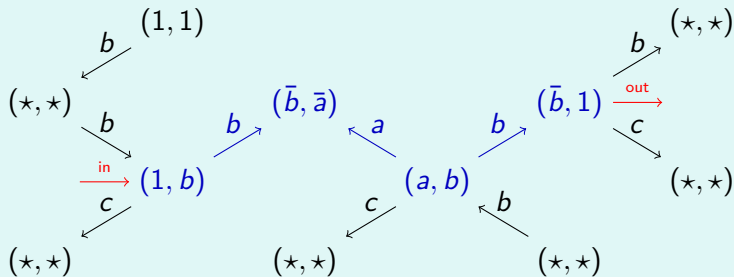
Let  $F = \{\circ\}$  and  $A = \{a, b, c\}$ .

Automaton  $\mathcal{A} = \langle Q, \delta, \Delta, W \rangle$

- $Q = (\star, \star) + \{1, a, \bar{a}, b, \bar{b}\} \times \{1, a, \bar{a}, b, \bar{b}\}$ ,
- $\delta(\circ) = Q$ ,
- $\Delta(x)$  defined by all transitions of one the following form:
  - $(\star, \star) \xrightarrow{x} (y, z)$  when  $y \neq \bar{x}$  and  $z \neq \bar{x}$ ,
  - $(\star, \star) \xleftarrow{x} (y, z)$  when  $y \neq x$  and  $z \neq x$ ,
  - $(y, x) \xrightarrow{x} (\bar{x}, z)$  when  $y \neq x$  and  $z \neq \bar{x}$ ,
  - $(y, \bar{x}) \xleftarrow{x} (x, z)$  when  $y \neq \bar{x}$  and  $z \neq x$ ,
- $W = \left( \{1\} \times \{1, a, b, \bar{a}, \bar{b}\} \right) \times \left( \{1, a, b, \bar{a}, \bar{b}\} \times \{1\} \right)$

## An example: marking the path between roots

### Exemple : a pictured run



### Characteristic property of $L(\mathcal{A})$

The output path  $u$  belongs to  $FG(a + b)$ . Observe that if  $u = x_1 x_2 \cdots x_n$  then  $u$  is **uniquely determined** by the sequence of states  $(1, x_0), (\bar{x}_0, x_1), \cdots, (\bar{x}_{n-1}, x_n), (x_n, 1)$ .

## Expressiveness

### Remark

Every language  $L$  recognized by an automaton is upward closed in the natural order.

### Theorem (MSO<sup>↑</sup>)

*A language of birooted  $F$ -trees  $L \subseteq \mathcal{B}(F)$  is recognized by a finite state tile automaton if and only if  $L$  is definable in MSO and upward closed.*

### Proof.

Slightly technical encoding of automata for mono-rooted tree (with distinguished vertex output) into birooted tree automata. □

### Corollary

*This class of languages is closed under union, intersection, (upward closed) product, (upward closed) star, projection. . .*

## Towards an associated algebra

### Induced premorphism

Let  $\mathcal{A} = \langle Q, \delta, \Delta, W \rangle$  be an automaton and let

$$\varphi : \mathcal{B}(F) \rightarrow \mathcal{P}(Q \times Q)$$

be the mapping defined by  $\varphi(0) = \emptyset$ ,  $\varphi(1) = I_Q$  and, for every non trivial  $B = \langle t, u \rangle$ ,  $\varphi(B)$  is the set of pair  $(\rho(1), \rho(u))$  for runs  $\rho$  of  $\mathcal{A}$  over  $B$ .

### Lemma

*The mapping  $\varphi$  recognizes  $L(\mathcal{A})$  in the sense that  $L(\mathcal{A}) = \varphi^{-1}(\varphi(L(\mathcal{A})))$ .*

### Question

What can we say about  $\varphi^{-1}(X)$  for  $X \subseteq \mathcal{P}(Q \times Q)$  ?

## Towards an associated algebra

### Lemma

*The sub-monoid induced by  $\varphi(\mathcal{B}(F))$  ordered by inclusion is a partially ordered monoid.*

### Lemma

*The mapping  $\varphi$  is a  $\vee$ -premorphisms, i.e. for every  $B$  and  $C \in \mathcal{B}(F)$ :*

- *if  $B \leq C$  then  $\varphi(B) \subseteq \varphi(C)$*
- *$\varphi(B \cdot C) \subseteq \varphi(B) \cdot \varphi(C)$*

### Question

Find a more complete axiomatization of these ordered monoids and premorphisms ?

### 3. Algebra : $\text{BOOL}(\text{MSO}^\uparrow)$

Algebras for the boolean closure of upward closed MSO definable languages

## Algebras for languages of birooted trees

### Definition (Adequately ordered monoid)

A monoid  $S$  such that:

- **the order is stable**, i.e. for every  $x, y$  and  $z \in S$ , if  $x \leq y$  then  $z \cdot x \leq z \cdot y$  and  $x \cdot z \leq x \cdot y$ ,
- **subunits are idempotents**, i.e. for every  $x \in U(S) = \{x \in S : x \leq 1\}$  we have  $xx = x$ ,
- for every  $x \in S$ , both the **left projection**  $x^L = \min\{z \in U(S) : x \cdot z = x\}$  and the **right projection**  $x^R = \min\{z \in U(S) : z \cdot x = x\}$  exists (in  $U(S)$ ).

### Examples

- (1) Every trivially ordered monoid with  $x^L = x^R = 1$ .
- (2) Every naturally ordered inverse monoid with  $x^L = x^{-1}x$  and  $x^R = xx^{-1}$ .
- (3) Every finite partially ordered monoid with idempotent subunits.

## Algebras for languages of birooted trees

### Definition (Adequate premorphism)

A mapping  $\varphi : S \rightarrow T$  with  $S$  and  $T$  two adequately ordered monoids such that:

- $\varphi$  is **monotonic**, **submultiplicative** with  $\varphi(1) = 1$ .
- $\varphi$  preserves **disjoint product**, i.e. for every  $x$  and  $y \in S$  such that  $x \cdot y \neq 0$  and  $x^L \vee y^R = 1$  we have  $\varphi(x \cdot y) = \varphi(x) \cdot \varphi(y)$ ,
- $\varphi$  preserves **left** and **right projections**, i.e. for every  $x \in S$  we have  $\varphi(x^L) = (\varphi(x))^L$  and  $\varphi(x^R) = (\varphi(x))^R$ .

### Example

The canonical mapping  $\varphi : \mathcal{B}(F) \rightarrow \mathcal{P}(Q \times Q)$  induced by an automaton  $\mathcal{A}$ .

# Quasi-recognizable languages

## Definition

A language  $L \subseteq \mathcal{B}(F)$  is **quasi-recognizable** (QREC) when there exists a finite adequately ordered monoid  $S$  and an adequate premorphism  $\varphi : \mathcal{B}(F) \rightarrow S$  such that  $L = \varphi^{-1}(\varphi(L))$ .

## Lemma (Effectiveness)

*Let  $\varphi : \mathcal{B}(F) \rightarrow S$  be an adequate premorphism with finite  $S$ . For every  $B \in \mathcal{B}(F)$ , the image  $\varphi(B)$  of  $B$  is computable in linear time in  $B$ .*

## Proof.

Thanks to some canonical disjoint decomposition property. □

## Expressiveness of QREC

### Theorem

*A language  $L$  is QREC if and only if  $L$  is a finite boolean combination of languages recog. by finite state automata, i.e.  $QREC = \text{BOOL}(\text{MSO}^\uparrow)$ .*

### proof.

(QREC)  $\Rightarrow$  MSO : simulate disjoint decomposition.

(QREC + MSO)  $\Rightarrow$  BOOL(MSO<sup>↑</sup>) : let  $\varphi : \mathcal{B}(F) \rightarrow S$  an adequate premorphism with finite  $S$ . For every  $x \in S$  we have:

$$\varphi^{-1}(x) = \underbrace{\left( \bigcup_{y \in x^\uparrow} \varphi^{-1}(y) \right)}_{\text{MSO}^\uparrow} \cap \underbrace{\left( \bigcup_{y \in x^\downarrow} \varphi^{-1}(y) \right)}_{\text{MSO}^\downarrow}$$

## Expressiveness of QREC

proof (continued).

$\text{BOOL}(\text{MSO}^\uparrow) \Rightarrow \text{QREC}$  : since  $\text{MSO}^\uparrow \subseteq \text{QREC}$  and QREC is closed under complement, it suffices to prove closure under union.

Let  $\varphi_1 : \mathcal{B}(F) \rightarrow S_1$  and  $\varphi_2 : \mathcal{B}(F) \rightarrow S_2$  two adequate premorphisms with finite  $S_1$  and  $S_2$ .

Then  $S_1 \times S_2$  is adequately ordered and

$$\varphi = \langle \varphi_1, \varphi_2 \rangle : \mathcal{B}(F) \rightarrow S_1 \times S_2$$

recognizes the union of any languages  $L_1$  and  $L_2$  resp. recognized by  $\varphi_1$  and  $\varphi_2$ .



## 4. Expressiveness : $\text{BOOL}(\text{MSO}^\uparrow) \subset \text{MSO}$

Where we eventually show that QREC is strictly included in MSO (this result is obtained after the presentation, following Marc Zeitoun remark)

# $\text{BOOL}(\text{MSO}^\uparrow)$ vs $\text{MSO}$

## Theorem

There are  $\text{MSO}$  definable languages of tiles that are not QREC henceforth, equivalently, not  $\text{BOOL}(\text{MSO}^\uparrow)$ .

## Proof.

With  $A = \{a\}$  and  $F = \{\circ\}$ , for every  $k \in \mathbb{N}$ , let  $B_k = \langle t, 1 \rangle$  defined by  $\text{dom}(t) = \{a^n : n \leq k\}$  and let  $L = \{B_{2n}\}_{n \in \mathbb{N}}$ . The language  $L$  is obviously definable in  $\text{MSO}$ .

Assume there is a monotonic mapping  $\varphi : \mathcal{B}(F) \rightarrow S$  that recognizes  $L$ . Let  $k \in \mathbb{N}$ , we have

$\varphi(B_{2k}) \geq \varphi(B_{2k+1}) \geq \varphi(B_{2(k+1)})$ . But since both  $B_{2k}$  and  $B_{2(k+1)}$  belong to  $L$  while  $B_{2k+1}$  does not, this means that  $\varphi(B_{2k}) > \varphi(B_{2k+1}) > \varphi(B_{2(k+1)})$ .

As this holds for arbitrary  $k$ , this means that  $\varphi(B_{2k})$  is an infinite chain in  $S$  hence  $S$  is infinite. □

# Languages of $F$ -terms vs languages of birooted $F$ -trees

Let  $\rho : F \rightarrow \mathcal{P}(A)$  be an arity mapping.

## Definition ( $F$ -terms encodings)

A birooted  $F$ -tree  $B = \langle t, u \rangle$  is an  **$F$ -term** when

- $u = 1$ ,
- $\text{dom}(t) \subseteq A^*$ ,
- for every  $v \in \text{dom}(t)$ , for every  $a \in A$  such that  $va \in \text{dom}(t)$ , we have  $a \in \rho(t(v))$ .

Such an  $F$ -term is **complete** when it is minimal in the natural order, i.e. no vertex can be added without breaking the arity mapping condition.

## Theorem

*Every MSO definable language  $L$  of complete  $F$ -term is QREC.*

## Proof.

Observe that the language  $L$  forms an antichain in  $\mathcal{B}(F)$ .  
Observe that both the upward and downward closure of the language  $L$  are definable in MSO hence definable in QREC.  
Conclude by taking the intersection of these two languages.  $\square$

## 5. Conclusion

Where are we going ? What is all that for ? See  
<http://hal.archives-ouvertes.fr/hal-00784898>  
for all detailed proofs.

## Works in progress

### Algebraic tools

- associated algebraic construction (product, star, refinement) ?
- language classification/decidability ? syntactic monoid ?
- co-recognizability ?
- weakly adequately ordered monoids ?

### Automata theoretic tools

- birooted tree automata vs word universal automata ?
- action refinement with overlaps ?
- walking automata/monotonic cellular automata ?

### Extension

- infinite structures ?
- timed extension ?
- $k$ -ary many rooted trees (and guarded logic) ?