

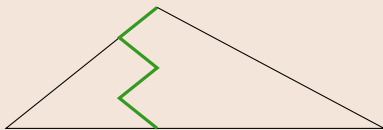
# MOTIVATION

- Understanding tree formalisms
- To construct reasonable fragments of XPATH with data
- To simplify formulas in verification.

## POTTHOF'S EXAMPLE

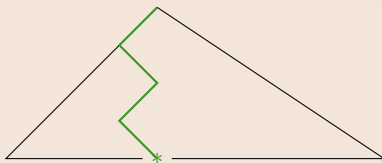
### EXISTS A MAXIMAL PATH OF ODD LENGTH

- “There is a maximal path of odd length” is definable in  $\text{FO}[\leq, \textit{left}, \textit{right}]$  over ordered binary trees.



## POTTHOF'S EXAMPLE: CONT.

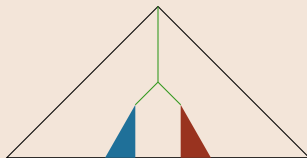
WHEN THE PARITY OF DEPTHS OF ALL LEAVES IS THE SAME



$(\text{left} \cdot \text{right})^* \text{left path}$ .

WHEN THIS IS NOT THE CASE

Find a maximal depth node where the property fails.

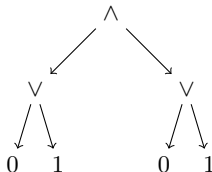


The formula expressing the property is not of the form “there is a path”.

# BOOLEAN EXPRESSIONS

## POSITIVE BOOLEAN EXPRESSIONS

Let  $A = \{\vee, \wedge, 0, 1\}$ , and consider the set of ordered binary trees that evaluate to 1 (and are well-formed).



## THEOREM (HEUTER, POTTHOFF)

*The above language is not definable in chain logic. (This language is aperiodic.)*

## WORDS AND THEIR STRUCTURE

- Concatenation is a privileged operation on  $A^*$ .
- The set of words  $A^*$  with concatenation forms a monoid.

**Recognizability:**  $\alpha : A^* \rightarrow S$  and  $F \subseteq S$  such that  $L = \alpha^{-1}(F)$ .

**Syntactic monoid of  $L$ :** Smallest monoid recognizing  $L$ .

### APERIODICITY

A semigroup  $\langle S, \cdot \rangle$  is **aperiodic** iff there is  $n$  such that  $s^n = s^{n+1}$  for all  $s \in S$ .

### THEOREM (SCHÜTZENBERGER, MCNAUGHTON & PAPERT)

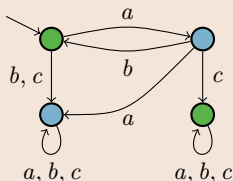
*A regular language is FO definable iff its syntactic monoid is aperiodic.*

# AUTOMATA VS. ALGEBRA

## A FOL DEFINABLE LANGUAGE

Let  $L = (ab)^*c\Sigma^*$ . It is first-order definable.

## THE MINIMAL AUTOMATON FOR $L$



## REMARKS

- The aperiodicity property is not that visible from the structure of the automaton.
- Every property of syntactic algebra is also a property of the minimal automaton. **The converse is not true.**
- **This is good**, as long as we know that properties of interest are properties of syntactic algebras.

## WHAT ALGEBRA GIVES US

- The structure gives new notions: idempotent, Green relations, . . .
- (pseudo) varieties. Limit the search space.
- Krohn-Rhodes decomposition theorem: A group free semigroup is a divisor of an iterated wreath product of  $U_2$ .
  - This implies LTL is equivalent to FO over words.
  - Decidability of Until-hierarchy for LTL [Therien, Wilke].

# WHAT STRUCTURE FOR TREES

- $A$ -algebras
- Pre-clones
- Forest algebras
- Seminearrings.



# A-ALGEBRAS

## RANKED ALPHABET

$A_0, A_1, A_2, \dots$

## FREE STRUCTURE: RANKED TREES

A letter in a node determines the number of children. These children are numbered.  
Free  $A$ -algebra: trees with operations of adding a letter on a top.

## A-ALGEBRA

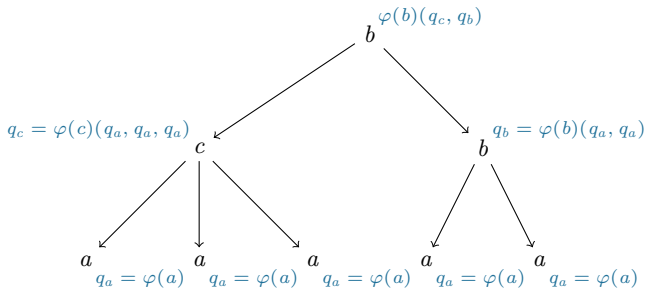
$\langle S, a_0, b_2, c_3 \dots \rangle$

## RECOGNITION

$\varphi : \text{Free}(A) \rightarrow S, S\text{-finite}, L = \varphi^{-1}(F)$

## RECOGNITION

$\varphi : \text{Free}(A) \rightarrow S$ ,  $S$ -finite,  $L = \varphi^{-1}(F)$



Syntactic algebra is the minimal leaves-to-root automaton.

# PRECLONES

## RANKED ALPHABET

$A_0, A_1, A_2, \dots, id \in A_1.$

## FREE STRUCTURE: MULTI-CONTEXTS

Obtained from alphabet by substitution.

- $AM_0$  trees
- $AM_1$  contexts with one variable
- ...

Operation of substitution of multicontexts into multicontext.

## PRE-CLONE

$S = \langle \{S_i\}_{i=0,1,\dots}, \{\cdot\}_J, 1 \rangle$

## RECOGNITION

$\varphi : MA \rightarrow S, S$  finitary,  $F \subseteq S_0, L = \varphi^{-1}(F)$

# FOREST ALGEBRA

Unranked alphabet  $A$ .

## FREE STRUCTURE

- Forests over unranked alphabet
- Forest contexts with one hole.
- Operations:
  - concatenation of forests,
  - putting forest or a context into a hole of a context.

## FOREST ALGEBRA $\langle H, V, act, in_l, in_r \rangle$

- Two monoids:  $H$  and  $V$ . We denote their operations by  $+$  and  $\cdot$ , respectively.
- An action  $act : H \times V \rightarrow H$ . We write  $vh$  for  $act(h, v)$ .
- Two operations  $in_l, in_r : H \rightarrow V$ .
- Axioms:

**ACTION**  $(v \cdot w)h = v(wh)$ ;

**INSERTION**  $in_l(g)h = g + h$  and  $in_r(g)h = h + g$ ;

**FAITHFULNESS** for every two distinct  $v, w \in V$  there is  $h \in H$  with  $vh \neq wh$ ;

## RECOGNIZABILITY

$\varphi : A^\Delta \rightarrow \langle H, V, act, in_l, in_r \rangle$

# SEMINARRINGS

Unranked alphabet:  $A$

## FREE STRUCTURE

- Forest multicontexts
- Operations:
  - Horizontal concatenation
  - Vertical composition: simultaneous substitution.

SEMINARRINGS:  $\langle N, 0, \square, +, \cdot \rangle$

Axioms:

$$(p + q) \cdot r = p \cdot r + q \cdot r$$

$$0 \cdot p = 0$$

$\vdots$

## RECOGNIZABILITY

$\varphi : A^\nu \rightarrow \langle N, 0, \square, +, \cdot \rangle$

# SEMINEARINGS FORGET MORE THAN FOREST ALGEBRAS

## “THERE IS A $b$ ”

- 0: “hopeless” multicontexts: arity 0 without an  $b$ .
- $\square$ : “possible” multicontexts: arity at least 1 without an  $b$ .
- $\infty$ : “good” multicontexts: with  $a$ .

## “SOME NODE HAS LABEL $b$ AND NO ANCESTOR WITH LABEL $c$ ”

- 0: “hopeless” multicontexts.
- $\square$ : “possible” multicontexts: some hole with no  $c$  ancestors.
- $\infty$ : “good” multicontexts: in the language.

Only the first of the two languages is in EF.

## SOLUTION

Instead of properties of a seminearring, one looks at the properties of morphisms  $\varphi : A^\nu \rightarrow \langle N, 0, \square, +, \cdot \rangle$ .

# VARIETIES OF FOREST ALGEBRAS

# VARIETIES OF FOREST ALGEBRAS

## PSEUDO-VARIETY

A **pseudovariety** of finite forest algebras is a collection  $\mathbb{V}$  of finite forest algebras with the following properties:

- $\mathbb{V}$  is closed under **binary product**.  
If  $(H, V), (G, U) \in \mathbb{V}$  then  $(H, V) \times (G, U) \in \mathbb{V}$ .
- $\mathbb{V}$  is closed under faithful quotients of **homomorphic images**.  
If  $(H, V) \in \mathbb{V}$  and  $(H', V')$  is its homomorphic image, then  $\text{faith}(H', V') \in \mathbb{V}$ .
- $\mathbb{V}$  is closed under faithful quotients of **subalgebras**.  
If  $(H, V) \in \mathbb{V}$  and  $(H', V')$  is its subalgebra then  $\text{faith}(H', V') \in \mathbb{V}$ .

## THEOREM

*If  $\mathbb{V}$  is a pseudovariety iff it is ultimately defined by some set of equations.*



# FIRST CORRESPONDENCE THEOREM

## VARIETIES OF FOREST LANGUAGES

Let  $\mathbb{V}$  be a variety of algebras. For every finite alphabet  $A$  define

$$\mathcal{V}(A) = \{L \subseteq H_A : (H^L, V^L) \in \mathbb{V}\}.$$

We call  $\mathcal{V}$  the **variety of forest languages** associated to  $\mathbb{V}$ , and write

$$\mathbb{V} \mapsto \mathcal{V}.$$

## THEOREM

*The mapping  $\mathbb{V} \mapsto \mathcal{V}$  is one-to-one.*

## SECOND CORRESPONDENCE THEOREM

### A PREFIX OPERATOR

If  $L$  is a language of  $A$ -forests and  $p$  is a  $A$ -context then we define

$$p^{-1}L = \{t : pt \in L\}.$$

### THEOREM

Let  $\mathcal{W}$  be an operator assigning to each  $A$  a family  $\mathcal{W}(A)$  of  $A$ -languages.  $\mathcal{W}$  is a variety of languages if and only if the following hold:

- 1 every  $\mathcal{W}(A)$  is closed under boolean operations.
- 2 if  $L \in \mathcal{W}(A)$  and  $p$  is an  $A$ -context then  $p^{-1}L \in \mathcal{W}(A)$ .
- 3 if  $(\alpha, \beta) : A^\Delta \rightarrow B^\Delta$  is a homomorphism, and  $L \in \mathcal{W}(B)$ , then  $\alpha^{-1}(L) \in \mathcal{W}(A)$ .

## CONSEQUENCES

### FACT

$\text{FO}[\leq]$  is a variety of forest languages.

### FACT

Chain logic is a variety of forest languages.

### REMARK

$\text{FO}[\textit{succ}]$  is not a variety of forest languages. Closure under inverse homomorphic images fails.

# WREATH PRODUCT

# WREATH PRODUCT

WREATH PRODUCT:  $(H_1, V_1) \circ (H_2, V_2)$

The product is  $(H_1 \times H_2, V_1 \times V_2^{H_2})$ :

- $(v_1, f)(h_1, h_2) = (v_1 h_1, f(h_1) v_2)$
- $(v, f) \cdot (v', f') = (v v', f'')$  where  $f''(h) = (f(v' h)) \cdot (f'(h))$
- This is a standard definition of wreath product of transformation semigroups (but with actions on the left).
- All axioms are satisfied.

## THEOREM

For every pair

$$\alpha : A^\Delta \rightarrow (G, W) \quad \beta : (A \times G)^\Delta \rightarrow (H, V)$$

there is a morphism into  $(G, W) \circ (H, V)$  that is equal on forest to the sequential composition

$$\alpha \otimes \beta : H_A \rightarrow G \times H$$

Every morphism  $A^\Delta \rightarrow (G, W) \circ (H, V)$  is realized this way.

$TL[\mathcal{A}]$

$TL[\mathcal{A}]$  is the class of languages recognized by iterated product of algebras from  $\mathcal{A}$ .

## THEOREM

We have the following:

- EF is  $TL[\mathcal{U}_1]$
- CLT is  $TL[\mathcal{U}_2]$
- CLT\* is  $TL$ [distributive aperiodic algebras]
- PDL is  $TL$ [distributive algebras]
- FO[ $\leq$ ] is  $TL$ [aperiodic path algebras]
- chain logic is  $TL$ [path algebras]

## CLASSES OF ALGEBRAS

$$\mathcal{U}_1 = (\{0, \infty\}, \{1, 0\})$$

$0 \cdot \infty = 0 \cdot 0 = \infty$ . Syntactic algebra of “some node with  $b$ ”.

$$\mathcal{U}_2 = (\{0, \infty\}, \{1, c_0, c_\infty\})$$

$C_h \cdot h' = h$ . Syntactic monoid of “some node with  $b$  with ancestors only  $c$ ”.

## DISTRIBUTIVE ALGEBRAS

- $H$  commutative
- Action is distributive:  $v(h_1 + h_2) = vh_1 + vh_2$

## PATH ALGEBRAS

- $H$  aperiodic and commutative,
- $vg + vh = v(g + h) + v0$
- $u(g + h) = u(g + uh)$  if  $u = u^2$ .



## INFINITE BASE

$\mathcal{A}$  is a **base** for  $L$  if  $L = TL(\mathcal{A})$

If  $L$  has a finite base then  $L = TL((H, V))$ .

### THEOREM

None of the language classes  $CTL^*$ ,  $FO[\leq]$ , PDL, or graded PDL has a finite base.

### CASE OF APERIODIC $V$

If a logic can define **there is a path labelled**  $(a^n b)^* c$  then it does not have finite base.

### $V$ HAS PERIODS

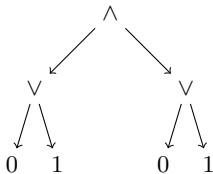
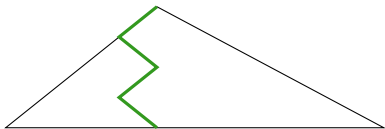
Languages **there is a path**  $a^m b$  with  $m$  a multiple of a fixed prime.

## THEOREM

A forest algebra divides and iterated wreath product of copies of  $\mathcal{U}_1$  iff  $H$  is idempotent and commutative and  $vh + h = vh$ .

This is a characterisation of EF.

# DEFINABILITY PROBLEM



# THE DEFINABILITY PROBLEM

## DEFINITION

Definability problem for a logic  $\mathcal{L}$  is to decide if the language of a given automaton can be defined in  $\mathcal{L}$ .

## WHAT LOGICS ARE KNOWN TO HAVE DECIDABLE DEFINABILITY PROBLEM

- Modal logic.
- EX, EF, EF+EX. [Bojańczyk & W.]
- FO[succ]. [Benedikt & Segoufin]
- EF+F<sup>-1</sup>. [Bojańczyk]
- Boolean closure of the  $\Sigma_1$  fragment of FO[ $\leq_H, \leq_V$ ]  
(an extension of Simon's Theorem). [Bojańczyk & Segoufin & Straubing]
- $\Delta_2$  fragment of FO[ $\leq_H, \leq_V$ ]. [Bojańczyk & Segoufin]

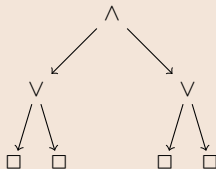
# CONFUSIONS

## HORIZONTAL CONFUSION

$(H, V)$  has *horizontal confusion* if there is a multicontext  $p$  and a set  $G \subseteq H$  with  $|G| > 1$  s.t for every  $g \in G$  and  $x \in \text{holes}(p)$ :

$$G \subseteq p[g/x][G].$$

## EXAMPLE: BOOLEAN EXPRESSIONS



## THEOREM

Chain logic algebras do not have horizontal confusion

## VERTICAL CONFUSION

### VERTICAL CONFUSION

$(H, V)$  has *vertical confusion* if there is a multicontext  $p$  and a set  $\{g_0, \dots, g_{k-1}\} \subseteq H$  with  $k > 1$  s.t. for every  $i = 0, \dots, k - 1$ :

$$p[g_i] = g_j \text{ where } j = (i + 1) \pmod{k}.$$

EXAMPLE: FORESTS WITH A MAXIMAL PATH IN  $(ab)^*$

EXAMPLE: BINARY TREES WITH EVEN PATH LENGTH

### THEOREM

- CTL algebras do not have vertical confusion.
- $\text{FO}[\leq]$  algebras do not have vertical confusion for *uniform* multicontexts.

## COMPARING COUNTING AND NON-COUNTING LOGICS

- CTL\* cannot count the number of successors. Neither can PDL
- FO[ $\leq$ ] is a counting correspondent of CTL\*. Chain logics is that of PDL.

### THEOREM

Let  $(H, V), (H_j, V_j), j = 1, \dots, k$  be such that

- $H$  is idempotent and commutative,
- each  $(H_i, V_i)$  is a path algebra,
- $(H, V)$  divides  $(H_1, V_1) \circ \dots \circ (H_k, V_k)$ .

then

- each  $(H_i, V_i)$  has a **distributive** homomorphic image  $(H'_i, V'_i)$
- $(H, V)$  divides  $(H'_1, V'_1) \circ \dots \circ (H'_k, V'_k)$ .

## REDUCTIONS BETWEEN DIFFERENT DECISION PROBLEMS

UHV: Given a regular language  $L$  of unranked trees,  
decide if  $L$  can be defined in  $FO[\leq_H, \leq_V]$ .



BHV: Given a regular language  $L$  of binary trees,  
decide if  $L$  can be defined in  $FO[\leq_H, \leq_V]$ .



BV: Given a regular language  $L$  of binary trees,  
decide if  $L$  can be defined in  $FO[\leq_V]$ .



UV: Given a regular language  $L$  of unranked trees,  
decide if  $L$  can be defined in  $FO[\leq_V]$ .



CTL\*: Given a regular language  $L$  of unranked trees,  
decide if  $L$  can be defined in CTL\*.

### REMARK

If there is no vertical order then the problem is easy.  
A tree decomposes into a “set of linear orders”.

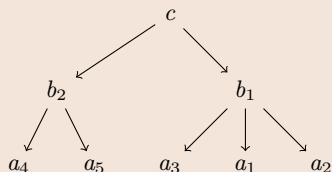
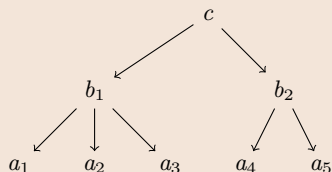


# ORDER INVARIANCE

# ORDER INVARIANCE

## ORDER INVARIANCE

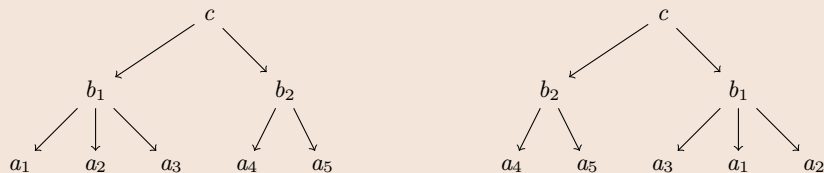
We say that a property of trees is **order invariant** iff it does not depend on the horizontal order.



## ORDER INVARIANCE

### ORDER INVARIANCE

We say that a property of trees is **order invariant** iff it does not depend on the horizontal order.

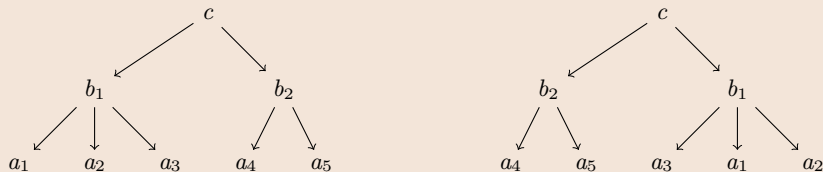


A logic  $\mathcal{L}$  is **order invariant** if every order invariant property definable in  $\mathcal{L}$  is definable in  $\mathcal{L}$  without the horizontal order.

## ORDER INVARIANCE

### ORDER INVARIANCE

We say that a property of trees is **order invariant** iff it does not depend on the horizontal order.



A logic  $\mathcal{L}$  is **order invariant** if every order invariant property definable in  $\mathcal{L}$  is definable in  $\mathcal{L}$  without the horizontal order.

### SOME FACTS

- MSOL is order invariant over ranked trees.
- MSOL is not order invariant over unranked trees.
- MSOL order invariant  $\equiv$  MSOL+counting modulo [Courcelle].
- FOL is not order invariant even over ranked trees.
- FOL[succ] is order invariant over ranked trees [Benedikt & Segoufin].

# OBJECTIVES

- Decidable characterisations of logical fragments
  - Study confusions
  - CTL
- Understanding the power of order invariance
  - What is order invariant  $\text{FO}[\leq_H, \leq_V]$ ?
- What is expressible with finite base
  - Beyond CTL.
- Profinite approach to varieties of trees
  - Delay theorem for trees.