Breadth-first signature of trees and rational languages

Victor Marsault,
joint work with Jacques Sakarovitch

CNRS / Telecom-ParisTech, Paris, France

Developments in Language Theory 2014, Ekateringburg,
2014–08–30

Breadth-first serialisation of languages and numeration systems:

The rational case

Victor Marsault,
joint work with Jacques Sakarovitch

CNRS / Telecom-ParisTech, Paris, France

Developments in Language Theory 2014, Ekateringburg,
2014–08–30

Outline

We call *tree* a...

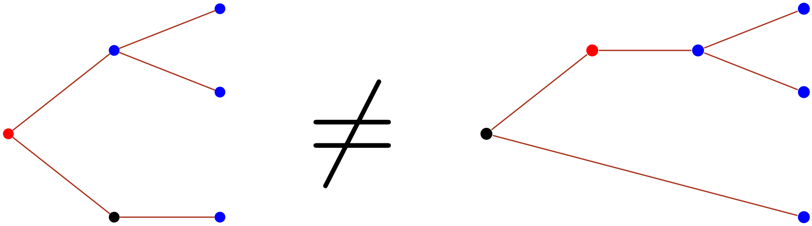
Directed graph which is

- **Rooted:** a node is called *the root* (leftmost in the figures)
- **Directed outward from the root:** there is a unique path from the root to every other node.
- **Ordered:** the children of every node are ordered
  (In the figures, lower children are smaller.)

We call *tree* a...

Directed graph which is

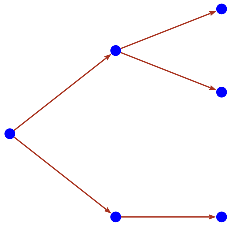
- **Rooted:** a node is called *the root* (leftmost in the figures)
- **Directed outward from the root:** there is a unique path from the root to every other node.
- **Ordered:** the children of every node are ordered
  (In the figures, lower children are smaller.)



We call *tree* a...

Directed graph which is

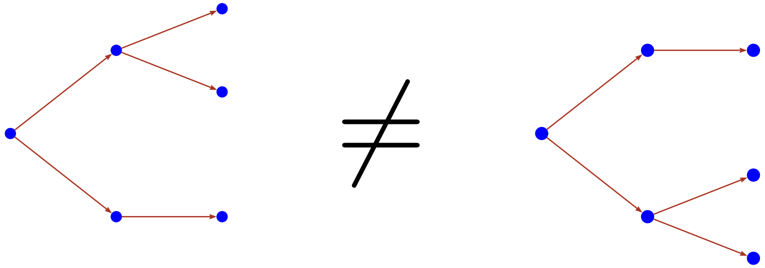
- **Rooted:** a node is called *the root* (leftmost in the figures)
- **Directed outward from the root:** there is a unique path from the root to every other node.
- **Ordered:** the children of every node are ordered
  (In the figures, lower children are smaller.)



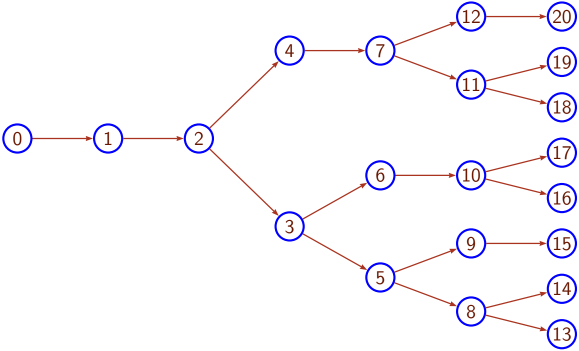
We call *tree* a...

Directed graph which is

- **Rooted:** a node is called *the root* (leftmost in the figures)
- **Directed outward from the root:** there is a unique path from the root to every other node.
- **Ordered:** the children of every node are ordered
  (In the figures, lower children are smaller.)

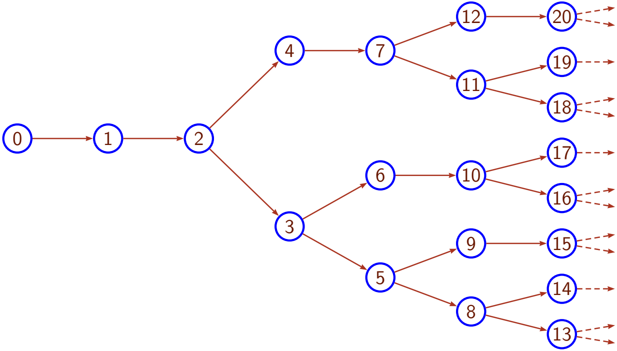


Every tree has a canonical breadth-first traversal



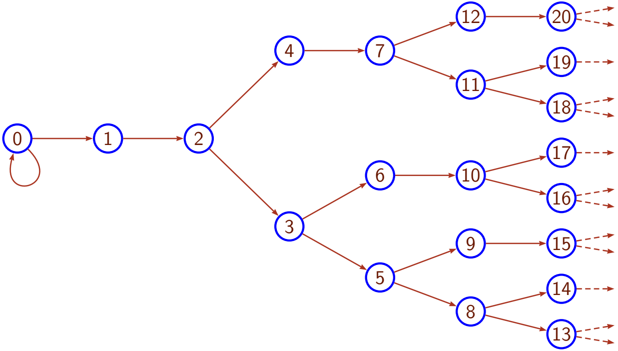
Two more features

- We consider infinite trees only.



Two more features

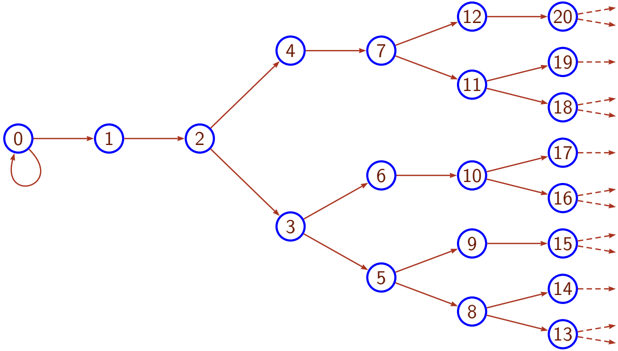
- We consider infinite trees only.
- For convenience, there is loop on the root.



Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

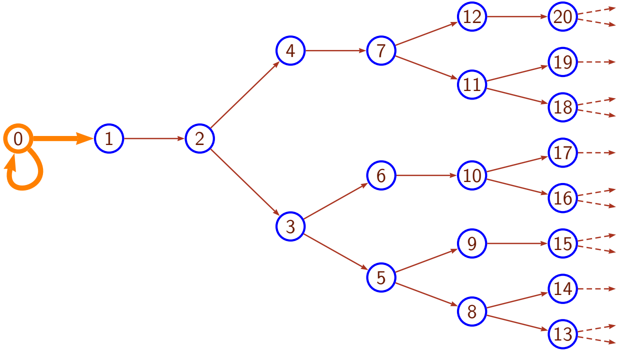


$\mathbf{s} =$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

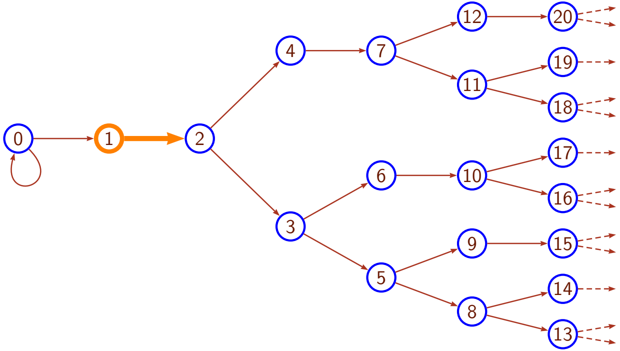


$$\mathbf{s} = 2$$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

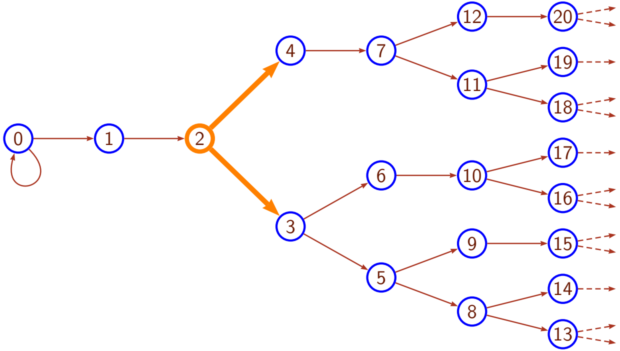


$$\mathbf{s} = 2\ 1$$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

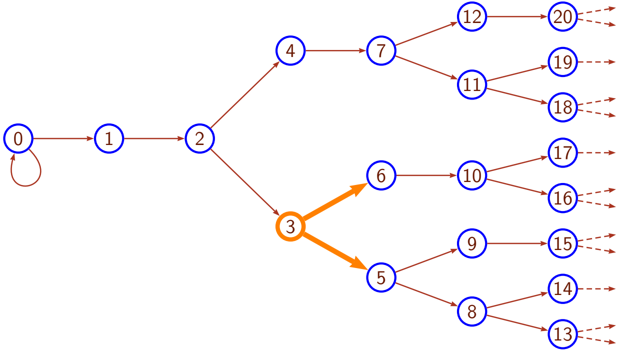


$\mathbf{s} = 2\ 1\ 2$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

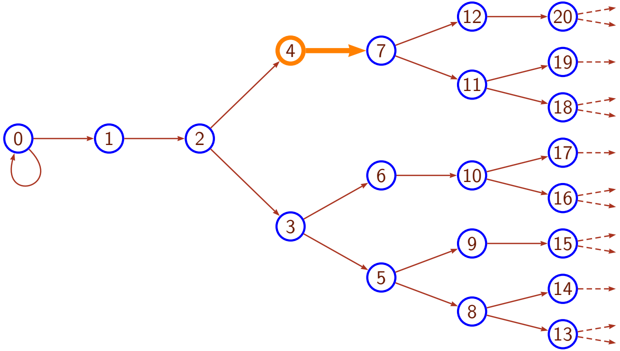


$\mathbf{s} = 2\ 1\ 2\ 2$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

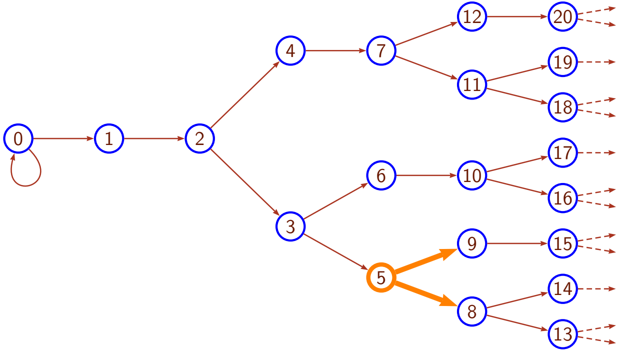


$\mathbf{s} = 2\ 1\ 2\ 2\ 1$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

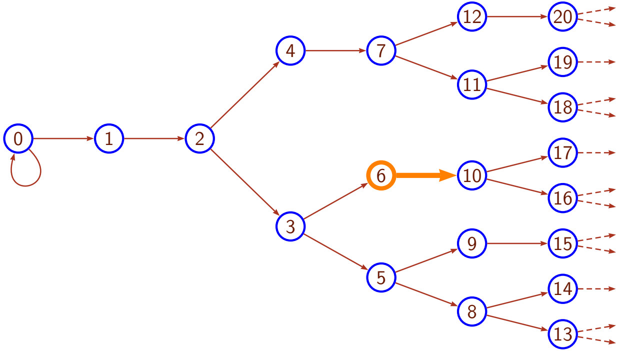


$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

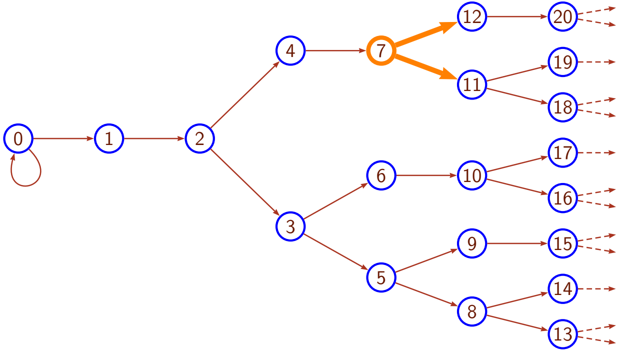


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1$$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

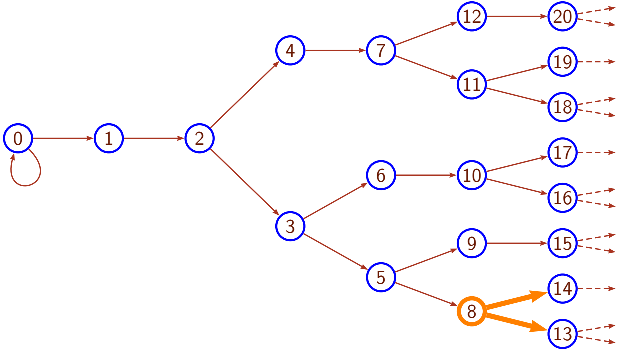


$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

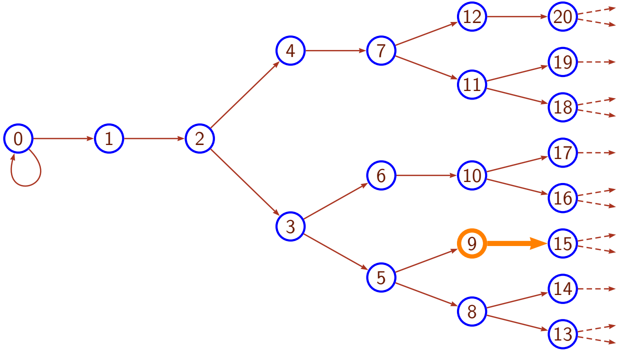


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2$$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

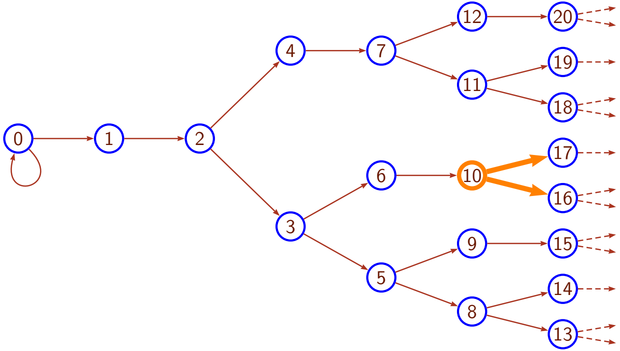


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 1$$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

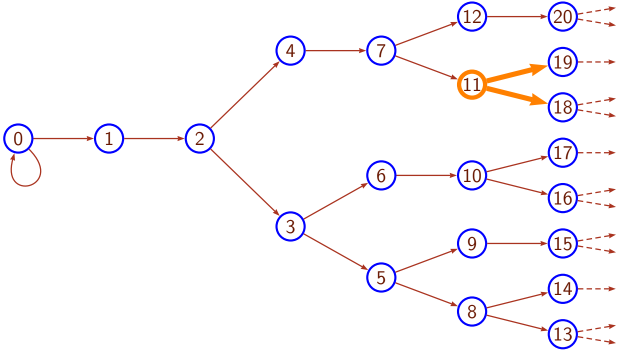


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 1\ 2$$

Signature of a tree

Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.

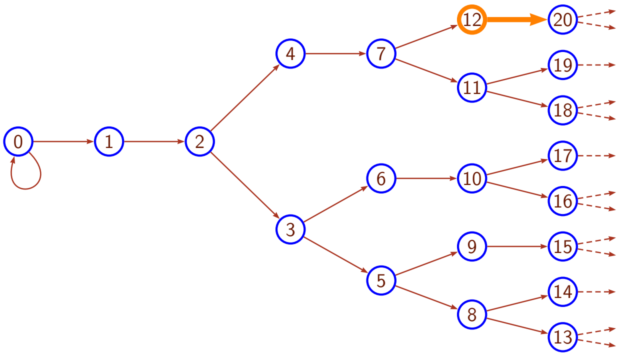


$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 1\ 2\ 2$$

Signature of a tree

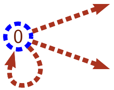
Definition

The **signature** of a tree is the sequence of the degrees of the nodes taken in breadth-first order.



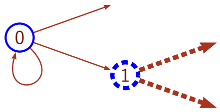
$$\mathbf{s} = 2\ 1\ 2\ 2\ 1\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 1\ \cdots$$

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



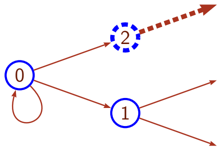
The signature is characteristic of a tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



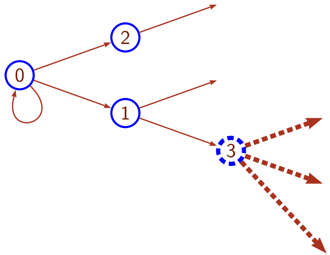
The signature is characteristic of a tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



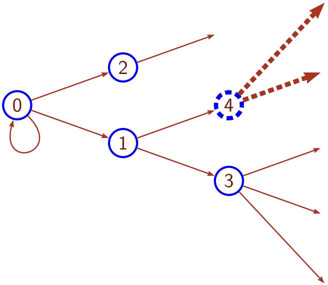
The signature is characteristic of a tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



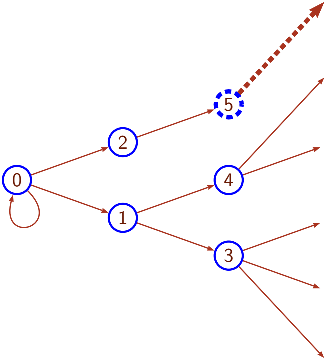
The signature is characteristic of a tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^\omega$$



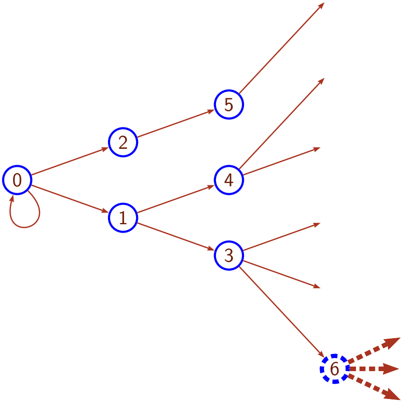
The signature is characteristic of a tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



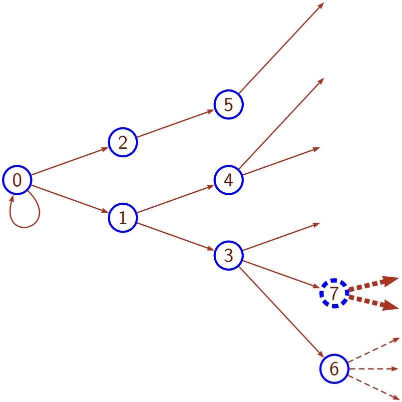
The signature is characteristic of a tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



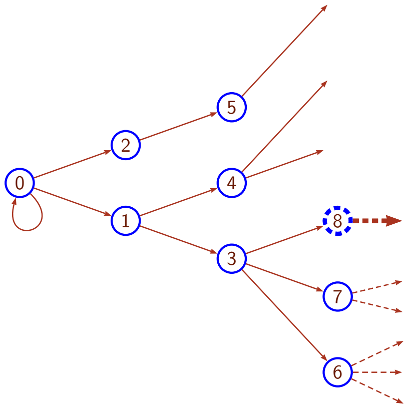
The signature is characteristic of a tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



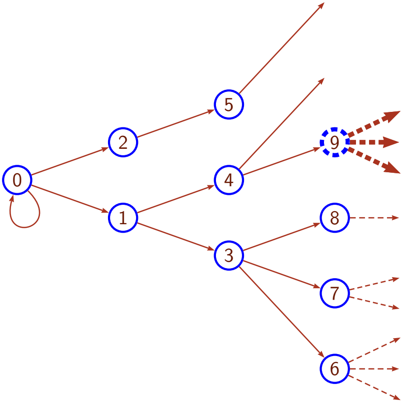
The signature is characteristic of a tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



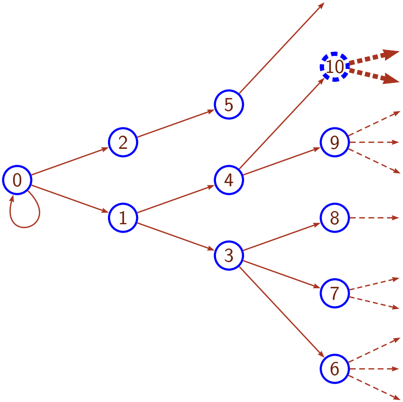
The signature is characteristic of a tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



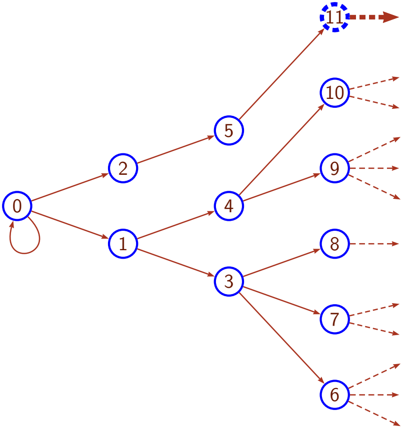
The signature is characteristic of a tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



The signature is characteristic of a tree

$$\mathbf{s} = (\ 3\ \ 2\ \ 1\ )^{\omega}$$



Alphabets are ordered hence

prefix-closed languages $=$ labelled trees.

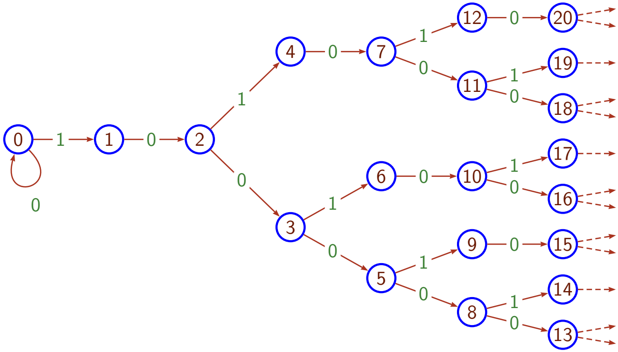
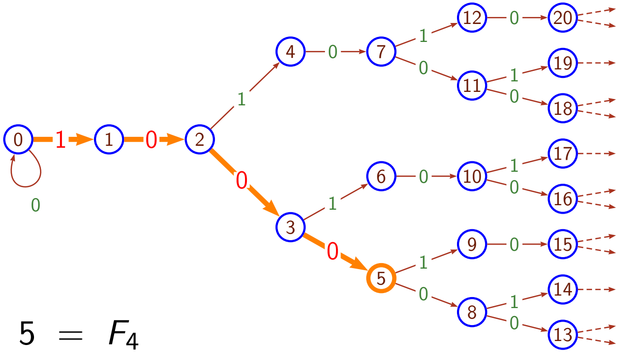


Figure : Integer representations in the Fibonacci numeration system.

Prefix-closed languages and labelled trees

Alphabets are ordered hence

prefix-closed languages = labelled trees.



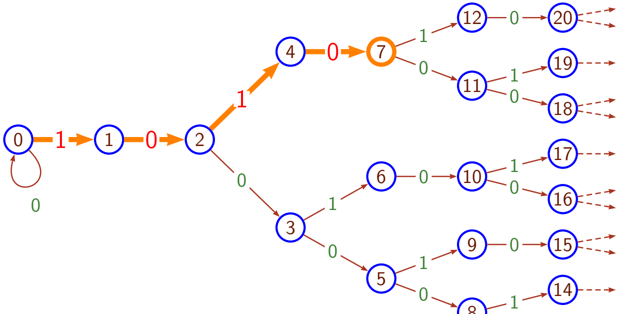
$5 = F_4$

Figure : Integer representations in the Fibonacci numeration system.

Prefix-closed languages and labelled trees

Alphabets are ordered hence

prefix-closed languages $=$ labelled trees.

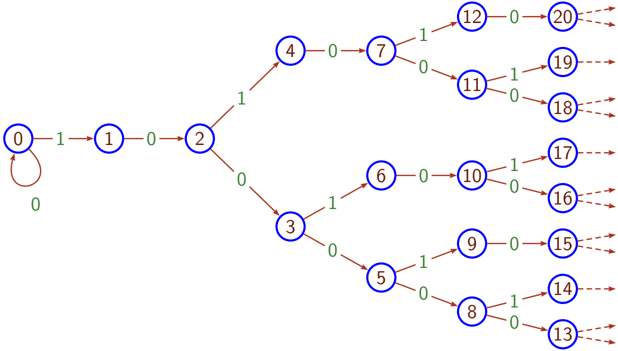


$$7 = 5 + 2 = F_4 + F_2$$

Figure : Integer representations in the Fibonacci numeration system.

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



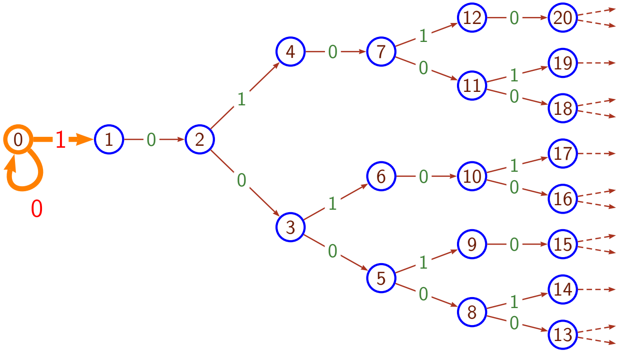
$\mathbf{s} =$

$\lambda =$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



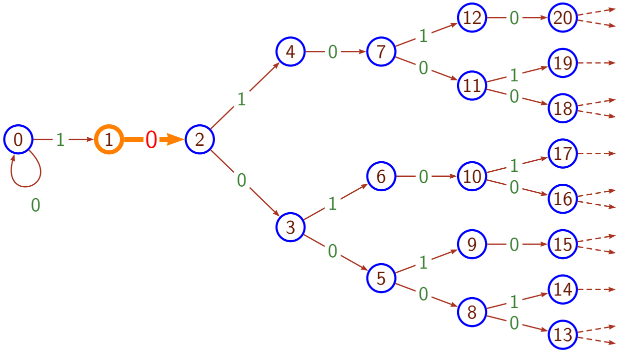
$\mathbf{s} = 2$

$\lambda = 01$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



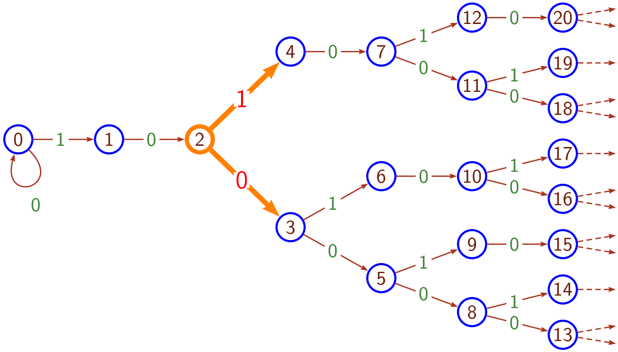
$\mathbf{s} = 2\ 1$

$\lambda = 01\ 0$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.

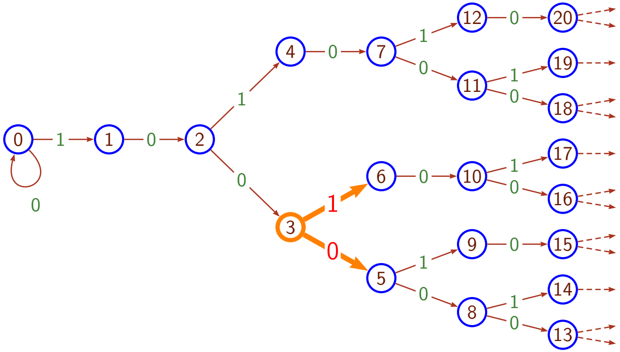


$\mathbf{s} = 2\ 1\ 2$

$\lambda = 01\ 0\ 01$

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.

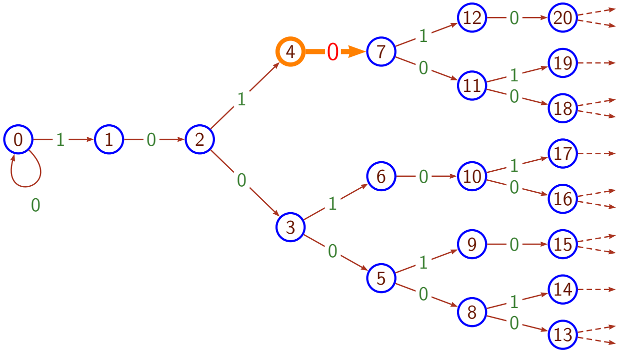


$$\mathbf{s} = 2 \ 1 \ 2 \ \ 2$$

$$\lambda = 01 \ 0 \ 01 \ 01$$

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



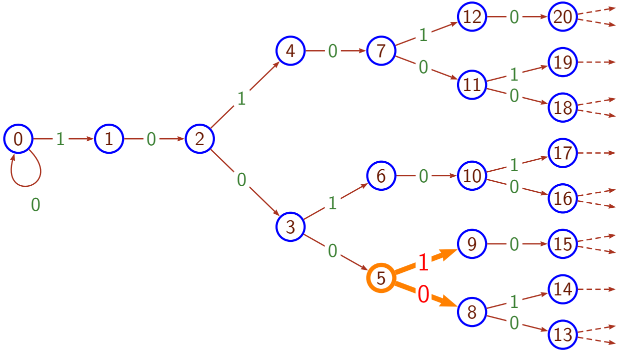
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ \ 1$$

$$\lambda = 01\ 0\ 01\ 01\ 0$$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



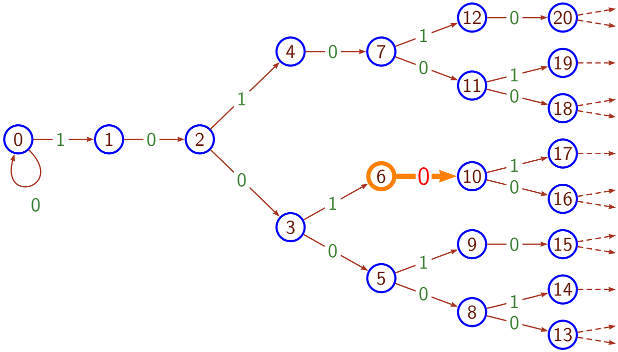
$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2$

$\lambda = 01\ 0\ 01\ 01\ 0\ 01$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



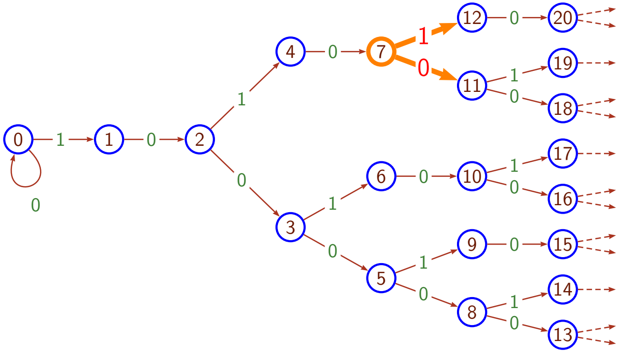
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0$$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



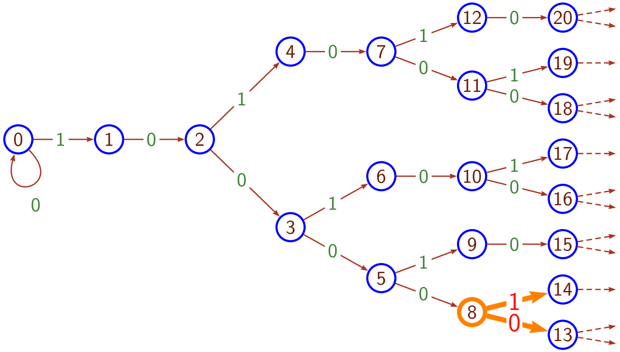
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01$$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



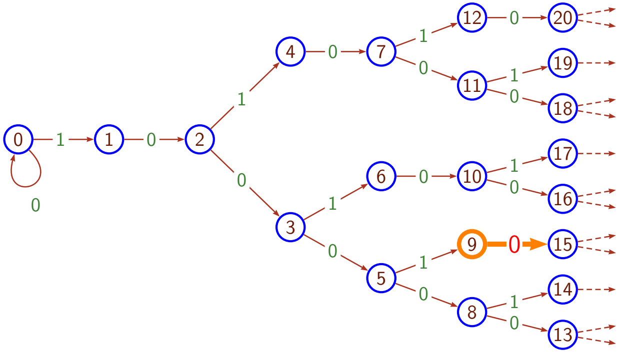
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01\ \ 01$$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



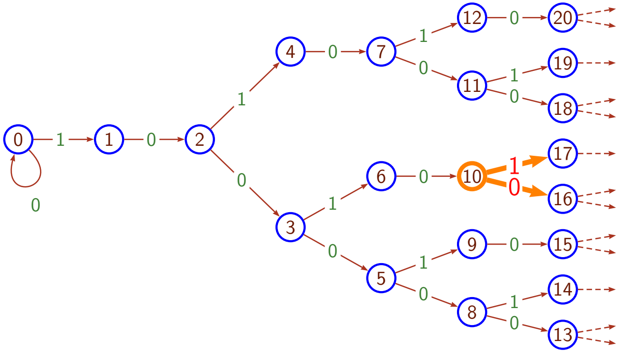
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2\ 1$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01\ 01\ 0$$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.

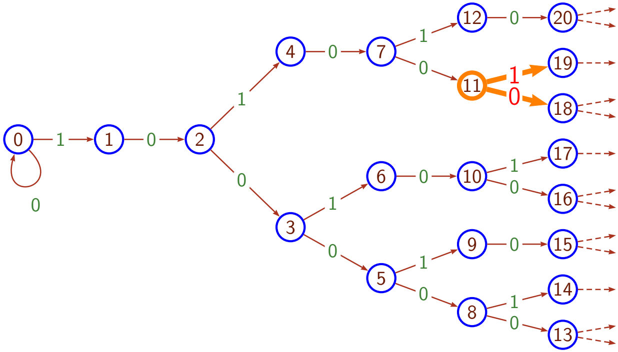


$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2\ 1\ 2$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01\ 01\ 0\ 01$$

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.



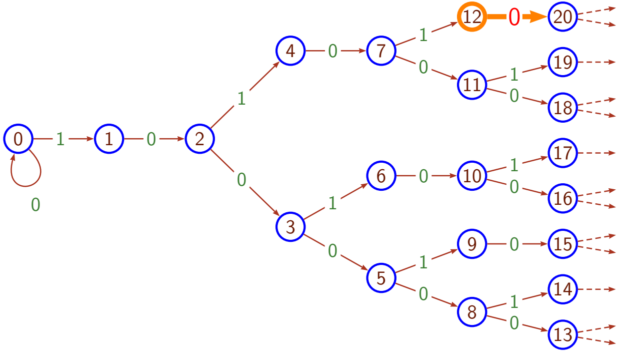
$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2\ 1\ 2\ \ 2$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01\ 01\ 0\ 01\ 01$$

Serialisation of a prefix-closed language

Definition

The **labelling** of a language is the **sequence of arc labels** of its transitions taken in breadth-first order.

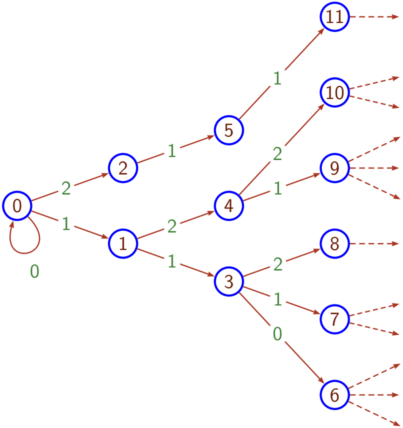


$$\mathbf{s} = 2\ 1\ 2\ \ 2\ 1\ 2\ 1\ 2\ \ 2\ 1\ 2\ \ 2\ 1\ \cdots$$

$$\lambda = 01\ 0\ 01\ 01\ 0\ 01\ 0\ 01\ 01\ 0\ 01\ 01\ 0\ \cdots$$

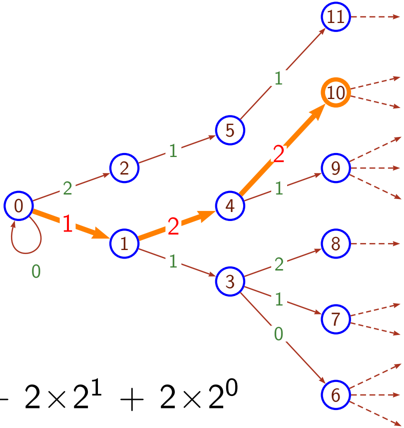
The pair signature/labelling is characteristic

$$\mathbf{s} = (3\ 2\ 1)^\omega$$
$$\lambda = (012\ 12\ 1)^\omega$$



The pair signature/labelling is characteristic

$$\mathbf{s} = (3\ 2\ 1)^\omega$$
$$\lambda = (012\ 12\ 1)^\omega$$



$$10 = 1\times 2^2 + 2\times 2^1 + 2\times 2^0$$

Figure : Non-canonical integer representations in base 2.

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

A word on substitution

A substitution σ is a morphism $A^* \to A^*$.

Running examples

Fibonacci substitution: $\{a, b\} \to \{a, b\}^*$
$a \mapsto ab$
$b \mapsto a$

A word on substitution

A substitution σ is a morphism $A^* \to A^*$.

Running examples

Fibonacci substitution: $\{a, b\} \to \{a, b\}^*$

$a \mapsto ab$

$b \mapsto a$

Periodic substitution: $\{a, b, c\} \to \{a, b, c\}^*$

$a \mapsto abc$

$b \mapsto ab$

$c \mapsto c$

A word on substitution

A substitution σ is a morphism $A^* \to A^*$.

σ is prolongable on a if $\sigma(a)$ starts with the letter a.

Running examples

Fibonacci substitution: $\{a, b\} \to \{a, b\}^*$

$a \mapsto ab$

$b \mapsto a$

Periodic substitution: $\{a, b, c\} \to \{a, b, c\}^*$

$a \mapsto abc$

$b \mapsto ab$

$c \mapsto c$

A word on substitution

A substitution σ is a morphism $A^* \to A^*$.

σ is prolongable on a if $\sigma(a)$ starts with the letter a.
In this case, $\sigma^\omega(a)$ exists and is called a purely substitutive word.

Running examples

Fibonacci substitution: $\{a, b\} \to \{a, b\}^*$
$a \mapsto ab$
$b \mapsto a$

Periodic substitution: $\{a, b, c\} \to \{a, b, c\}^*$
$a \mapsto abc$
$b \mapsto ab$
$c \mapsto c$

Substitutive signature

σ: a substitution $A^* \to A*$ prolongable on a.

f : a letter-to-letter morphism
$f(\sigma^\omega(a))$ is called a subtitutive word.

Substitutive signature

σ: a substitution $A^* \to A*$ prolongable on a.

f : a letter-to-letter morphism
$f(\sigma^\omega(a))$ is called a subtitutive word.

Definitions

let f_σ be the (letter-to-letter) morphism: $A^* \to \mathbb{N}^*$ defined by

- $\forall b,\ f_\sigma(b) = |\sigma(b)|$

We call $f_\sigma(\sigma^\omega(a))$ a **subtitutive signature**.

Substitutive signature

σ: a substitution $A^* \to A*$ prolongable on a.

f : a letter-to-letter morphism
$f(\sigma^\omega(a))$ is called a subtitutive word.

Definitions

let f_σ be the (letter-to-letter) morphism: $A^* \to \mathbb{N}^*$ defined by

- $\forall b,\ f_\sigma(b) = |\sigma(b)|$

We call $f_\sigma(\sigma^\omega(a))$ a **subtitutive signature**.

If g is a morphism such that

- $\forall b,\ |g(b)| = |\sigma(b)|$
- if $g(b) = c_0 c_1 \cdots c_k$ then $c_0 < c_1 < \cdots < c_k$

We call $g(\sigma^\omega(a))$ a **substitutive labelling**.

Example 1 – the Fibonacci signature

$$\sigma(a) = ab \implies f_\sigma(a) = 2$$
$$\sigma(b) = a \implies f_\sigma(b) = 1$$
$$f_\sigma(\sigma^\omega(a)) = 2122121221221212212122\cdots$$

if we choose g:

$g(a) = 01$

$g(b) = 0$

$$g(\sigma^\omega(a)) = 01\,0\,01\,01\,0\,01\,0\,01\,01\,0\,01\,01\,0\,\cdots$$

Example 1 – the Fibonacci signature

$$\sigma(a) = ab \quad \Longrightarrow \quad f_\sigma(a) = 2$$
$$\sigma(b) = a \quad \Longrightarrow \quad f_\sigma(b) = 1$$
$$f_\sigma(\sigma^\omega(a)) = 212212122122121221212 2\cdots$$

if we choose g:

$$g(a) = 01$$
$$g(b) = 0$$
$$g(\sigma^\omega(a)) = 01\,0\,01\,01\,0\,01\,0\,01\,01\,0\,01\,01\,0\,\cdots$$

This pair signature/labelling defines the language of integer representations in the Fibonacci numeration system.

Example 2 – a periodic signature

$\sigma(a) = abc \quad (f_\sigma(a) = 3)$
$\sigma(b) = ab \quad (f_\sigma(b) = 2)$
$\sigma(c) = c \quad (f_\sigma(c) = 1)$

$$\sigma(abc) = abc\,abc \qquad \text{hence } f_\sigma(\sigma^\omega(a)) = (321)^\omega$$

If we choose g:
$g(a) = 012$
$g(b) = 12$
$g(c) = 1$

$$g(\sigma^\omega(a)) = (012\,12\,1)^\omega$$

Example 2 – a periodic signature

$\sigma(a) = abc \quad (f_\sigma(a) = 3)$
$\sigma(b) = ab \quad (f_\sigma(b) = 2)$
$\sigma(c) = c \quad (f_\sigma(c) = 1)$

$$\sigma(abc) = abc\,abc \qquad \text{hence } f_\sigma(\sigma^\omega(a)) = (321)^\omega$$

If we choose g:
$g(a) = 012$
$g(b) = 12$
$g(c) = 1$

$$g(\sigma^\omega(a)) = (012\,12\,1)^\omega$$

This pair signature/labelling defines a *non-canonical* representation of integers in base 2.

Example 3 – the Thue-Morse morphism

$$\sigma(a) = ab \quad (f_\sigma(a) = 2)$$
$$\sigma(b) = ba \quad (f_\sigma(b) = 2)$$
$$f_\sigma(\sigma^\omega(a)) \quad = \quad 2^\omega$$

$\forall$ labelling g, the language is *essentially* $(0+1)^*$.

Forward direction of the theorem

Theorem

L: a prefix-closed language.

Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

Forward direction of the theorem

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

(σ, g): a substitutive signature.
(σ, g) defines a finite automaton $\mathcal{A}_{(\sigma,g)}$.
It is analogous to

- the *prefix graph/automaton* in Dumont–Thomas '89,'91,'93
- or the correspondence used in Maes–Rigo '02.

Forward direction of the theorem

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

(σ, g): a substitutive signature.
(σ, g) defines a finite automaton $\mathcal{A}_{(\sigma,g)}$.
It is analogous to

- the *prefix graph/automaton* in Dumont–Thomas '89,'91,'93
- or the correspondence used in Maes–Rigo '02.

Proposition

The language accepted by $\mathcal{A}_{(\sigma,g)}$ has signature (σ, g).

Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle \mathsf{A}, B, \delta, \{\mathsf{a}\}, \mathsf{A} \rangle$$

$\sigma(\mathsf{a}) = \mathsf{a}\,\mathsf{b}$
$\sigma(\mathsf{b}) = \mathsf{a}$

$g(\mathsf{a}) = 0\,1$
$g(\mathsf{b}) = 0$

Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}, B, \delta\,, \{\mathsf{a}\}\,, \mathsf{A}\,\rangle$$

$$\sigma(\,\mathsf{a}\,) = \mathsf{a}\,\mathsf{b}$$
$$\sigma(\,\mathsf{b}\,) = \mathsf{a}$$

$$g(\,\mathsf{a}\,) = 0\,1$$
$$g(\,\mathsf{b}\,) = 0$$

Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B,\, \delta\,,\, \{\mathsf{a}\},\, \mathsf{A}\,\rangle$$

$\sigma(\mathsf{a}) = \mathsf{a\,b}$
$\sigma(\mathsf{b}) = \mathsf{a}$

$g(\mathsf{a}) = 0\,1$
$g(\mathsf{b}) = 0$

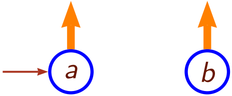
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathrm{A}\,, B,\, \delta\,,\, \{\mathrm{a}\}\,,\, \mathrm{A} \,\rangle$$

$\sigma(\,\mathrm{a}\,) = \mathrm{a\,b}$
$\sigma(\,\mathrm{b}\,) = \mathrm{a}$

$g(\,\mathrm{a}\,) = 0\,1$
$g(\,\mathrm{b}\,) = 0$



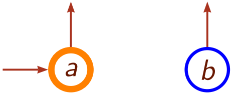
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle \, A \, , B, \delta, \{a\} \, , A \, \rangle$$

$\sigma(a) = ab$
$\sigma(b) = a$

$g(a) = 01$
$g(b) = 0$



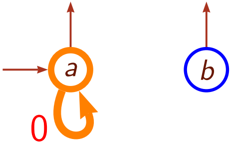
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle \mathsf{A}, B, \delta, \{\mathsf{a}\}, \mathsf{A} \rangle$$

$\sigma(\mathsf{a}) = \mathsf{a}\,\mathsf{b}$
$\sigma(\mathsf{b}) = \mathsf{a}$

$g(\mathsf{a}) = 0\,1$
$g(\mathsf{b}) = 0$



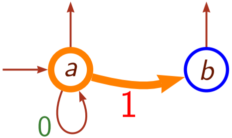
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B, \delta, \{\mathsf{a}\}\,, \mathsf{A}\,\rangle$$

$\sigma(\mathsf{a}) = \mathsf{a}\,\mathsf{b}$
$\sigma(\mathsf{b}) = \mathsf{a}$

$g(\mathsf{a}) = 0\,1$
$g(\mathsf{b}) = 0$



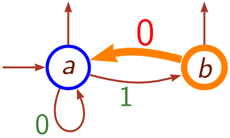
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathrm{A}\,, B, \delta, \{\mathrm{a}\}\,, \mathrm{A}\,\rangle$$

$\sigma(\mathrm{a}) = \mathrm{a\,b}$
$\sigma(\mathrm{b}) = \mathrm{a}$

$g(\mathrm{a}) = 0\,1$
$g(\mathrm{b}) = 0$



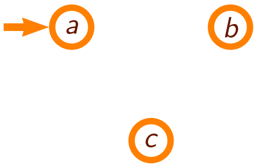
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle \mathsf{A}, B, \delta, \{\mathsf{a}\}, \mathsf{A} \rangle$$

$\sigma(\mathsf{a}) = \mathsf{a\,b\,c}$
$\sigma(\mathsf{b}) = \mathsf{a\,b}$
$\sigma(\mathsf{c}) = \mathsf{c}$

$g(\mathsf{a}) = 0\,1\,2$
$g(\mathsf{b}) = 1\,2$
$g(\mathsf{c}) = 1$



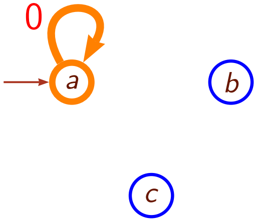
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B, \delta, \{\mathsf{a}\}\,, \mathsf{A}\,\rangle$$

$\sigma(\mathsf{a}) = \mathsf{a}\,\mathsf{b}\,\mathsf{c}$
$\sigma(\mathsf{b}) = \mathsf{a}\,\mathsf{b}$
$\sigma(\mathsf{c}) = \mathsf{c}$

$g(\mathsf{a}) = 0\,1\,2$
$g(\mathsf{b}) = 1\,2$
$g(\mathsf{c}) = 1$



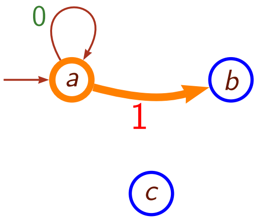
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B, \delta, \{\mathsf{a}\}\,, \mathsf{A} \,\rangle$$

$\sigma(\mathsf{a}) = \mathsf{a\,b\,c}$
$\sigma(\mathsf{b}) = \mathsf{a\,b}$
$\sigma(\mathsf{c}) = \mathsf{c}$

$g(\mathsf{a}) = 0\,1\,2$
$g(\mathsf{b}) = 1\,2$
$g(\mathsf{c}) = 1$



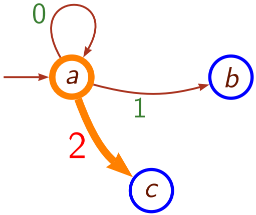
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathsf{A}\,, B, \delta, \{\mathsf{a}\}\,, \mathsf{A}\,\rangle$$

$\sigma(\,\mathsf{a}\,) = \mathsf{a\,b\,c}$
$\sigma(\,\mathsf{b}\,) = \mathsf{a\,b}$
$\sigma(\,\mathsf{c}\,) = \mathsf{c}$

$g(\,\mathsf{a}\,) = 0\,1\,2$
$g(\,\mathsf{b}\,) = 1\,2$
$g(\,\mathsf{c}\,) = 1$



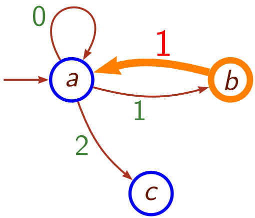
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, \mathrm{A}\,, B, \delta, \{\mathrm{a}\}\,, \mathrm{A}\,\rangle$$

$\sigma(\mathrm{a}) = \mathrm{a\,b\,c}$
$\sigma(\mathrm{b}) = \mathrm{a\,b}$
$\sigma(\mathrm{c}) = \mathrm{c}$

$g(\mathrm{a}) = 0\,1\,2$
$g(\mathrm{b}) = 1\,2$
$g(\mathrm{c}) = 1$



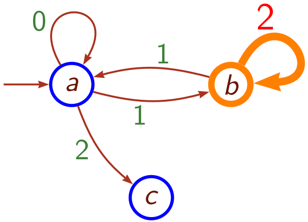
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, A\,, B, \delta, \{a\}\,, A\,\rangle$$

$\sigma(a) = abc$
$\sigma(b) = ab$
$\sigma(c) = c$

$g(a) = 012$
$g(b) = 12$
$g(c) = 1$



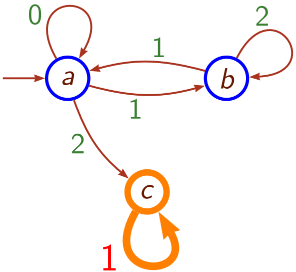
Automaton associated with a subst. signature

$\sigma : A^* \to A^*$ prolongable on a and $g : A^* \to B^*$

$$\mathcal{A}_{(\sigma,g)} = \langle\, A\,, B, \delta, \{a\}\,, A\,\rangle$$

$\sigma(a) = abc$
$\sigma(b) = ab$
$\sigma(c) = c$

$g(a) = 012$
$g(b) = 12$
$g(c) = 1$



Forward direction of the theorem

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

(σ, g): a substitutive signature.
(σ, g) defines a finite automaton $\mathcal{A}_{(\sigma,g)}$.
It is analogous to

- the *prefix graph/automaton* in Dumont Thomas '89,'91,'93
- or the correspondence used in Maes Rigo '02.

Proposition

The language accepted by $\mathcal{A}_{(\sigma,g)}$ has signature (σ, g).

Forward direction of the theorem

Theorem

L: a prefix-closed language.
Signature(L) is substitutive $\Leftrightarrow$ L is accepted by a finite automaton.

(σ, g): a substitutive signature.
(σ, g) defines a finite automaton $\mathcal{A}_{(\sigma,g)}$.
It is analogous to

- the *prefix graph/automaton* in Dumont Thomas '89,'91,'93
- or the correspondence used in Maes Rigo '02.

Proposition

The language accepted by $\mathcal{A}_{(\sigma,g)}$ has signature (σ, g).

Proof: unfold the automaton $\mathcal{A}_{(\sigma,g)}$.

What will be in the augmented version

Abstract Numeration System:
built from an arbitrary regular language.

Dumont-Thomas Numeration system:
built from a substitution

What will be in the augmented version

Abstract Numeration System:
built from an arbitrary regular language.

Dumont-Thomas Numeration system:
built from a substitution

Theorem (augmented version)

Two (prefix-closed) ANS built on language with **same signature** (but **different labelling**) are easily[†] convertible one from the other.

[†] Through a finite, letter-to-letter and pure sequential transducer.

What will be in the augmented version

Abstract Numeration System:
built from an arbitrary regular language.

Dumont-Thomas Numeration system:
built from a substitution

Theorem (augmented version)

Two (prefix-closed) ANS built on language with **same signature** (but **different labelling**) are easily† convertible one from the other.

Theorem (augmented version)

Every DTNS is a prefix-closed ANS.

Every prefix-closed ARNS is easily† convertible to a DTNS.

† Through a finite, letter-to-letter and pure sequential transducer.

Other works: Ultimately periodic signatures

$$\mathbf{s} \;=\; u\, r^{\omega} \qquad \text{with} \qquad r \;=\; r_0\, r_1\, r_2\, \cdots\, r_{q-1}$$

Definition: growth ratio

$$\mathtt{gr}(\mathbf{s}) \quad = \quad \frac{r_0 + r_1 + \cdots + r_{q-1}}{q}$$

Other works: Ultimately periodic signatures

$$\mathbf{s} \;=\; u\, r^{\omega} \qquad \text{with} \qquad r \;=\; r_0\, r_1\, r_2 \cdots r_{q-1}$$

Definition: growth ratio

$$\mathtt{gr}(\mathbf{s}) \;=\; \frac{r_0 + r_1 + \cdots + r_{q-1}}{q}$$

Theorem (MS, to appear)

If $\mathtt{gr}(\mathbf{s}) \in \mathbb{N}$, then **s** generates the language of a finite automaton.
It is linked‡ to the integer base $b = \mathtt{gr}(\mathbf{s})$.

If $\mathtt{gr}(\mathbf{s}) \notin \mathbb{N}$, then **s** generates a non-context-free language.
It is linked‡ to the *rational base* $\frac{p}{q} = \mathtt{gr}(\mathbf{s})$. (cf. Akiyama et al. '08)

‡ It is a non-canonical representation of the integers (using extra digits).

Future works : Directed signatures

Aperiodic signature: $\mathbf{s} = s_0\, s_1\, s_2 \cdots$

$S_n = \frac{1}{n}\Sigma_{k=0}^{n-1} s_k$: partial average of **s**.
α : lim S_n extends the notion of growth ratio.