Normalization of Sequential Top-Down Tree-to-Word Transducers

Gregoire Laurence
joint work with
Aurelien Lemay, Joachim Niehren, Slawek Staworko, and Marc Tommasi

Mostrare Project
INRIA Lille - Nord Europe
University of Lille

May 27, 2011

Why STWs?

XSLT

```
<xsl:output method="xml"/>
```

```
<xsl:output method="text"/>
```

Why STWs?

XSLT

```
<xsl:output method="xml"/>
```

```
<xsl:output method="text"/>
```

- Tree-to-word transformations

Why STWs?

XSLT

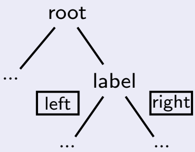
```
<xsl:output method="xml"/>
```

```
<xsl:output method="text"/>
```

- Tree-to-word transformations

XSLT framework

```
<xsl:template match="label">
 left
<xsl:apply-templates/>
 right
</xsl:template>
```

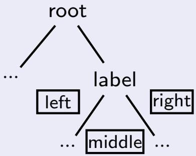


- concatenation on the output
- XSLT is TURING COMPLET

Why STWs?

XSLT inspired framework

```
<xsl:template mode="q" match="label">
  <!-- binary symbol -->
 left
<apply-child rank=1 mode="p" />
 middle
<apply-child rank=2 mode="p" />
 right
</xsl:template>
```



- ranked input
- finite states machine (template = state)
- deterministic (equivalence undecidable for non det string transducers [Griffiths 68])

Related work

Top-down tree transducers [Maneth et Al. PODS 2010]

- Tree-to-tree transducer
- Learning results
- Normalization algorithm
- Allow copy

Visibly pushdown transducers [Raskin et Al. ALP 2008]

- Tree-to-words transducer
- Works on stream of events

Why the normalization problem?

Learning

Grammatical inference

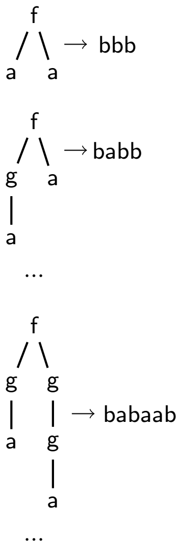
Classic problem

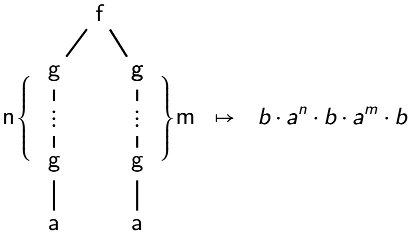
- String transducers [Choffrut, 1978]
- Top-down transducers [Engelfriet et Al. JCSS 2009]
- Bottom-up transducers [Friese et Al. DLT 2010]
- Macro tree transducers (open problem)

Outline

Outline

Example of Transformation

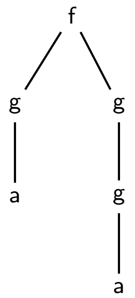




$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

Sequential Top-Down Tree-to-Word Transducers (STWs)

Machine definition

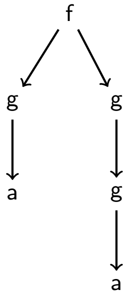


$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

Sequential Top-Down Tree-to-Word Transducers (STWs)

Machine definition

- Top Down tree traversal
- Deterministic
- Each node visited exactly once

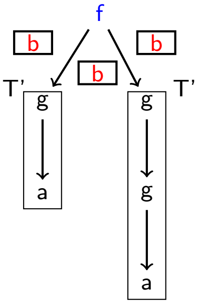


$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

Sequential Top-Down Tree-to-Word Transducers (STWs)

Machine definition

- Top Down tree traversal
- Deterministic
- Each node visited exactly once
- Output concatenation
- $T(f(t_1, t_2)) = b \cdot T'(t_1) \cdot b \cdot T'(t_2) \cdot b$

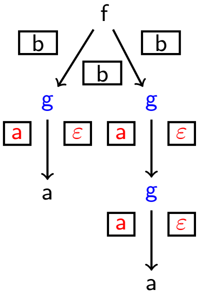


$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

Sequential Top-Down Tree-to-Word Transducers (STWs)

Machine definition

- Top Down tree traversal
- Deterministic
- Each node visited exactly once
- Output concatenation
- $T(f(t_1, t_2)) = b \cdot T'(t_1) \cdot b \cdot T'(t_2) \cdot b$
- $T'(g(t_1)) = a \cdot T'(t_1)$

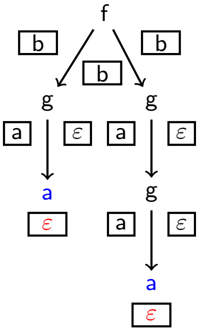


$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

Sequential Top-Down Tree-to-Word Transducers (STWs)

Machine definition

- Top Down tree traversal
- Deterministic
- Each node visited exactly once
- Output concatenation
- $T(f(t_1, t_2)) = b \cdot T'(t_1) \cdot b \cdot T'(t_2) \cdot b$
- $T'(g(t_1)) = a \cdot T'(t_1)$
- $T'(a) = \varepsilon$

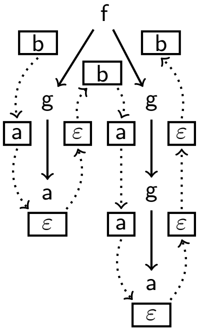


$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

Sequential Top-Down Tree-to-Word Transducers (STWs)

Machine definition

- Top Down tree traversal
- Deterministic
- Each node visited exactly once
- Output concatenation
- $T(f(t_1, t_2)) = b \cdot T'(t_1) \cdot b \cdot T'(t_2) \cdot b$
- $T'(g(t_1)) = a \cdot T'(t_1)$
- $T'(a) = \varepsilon$

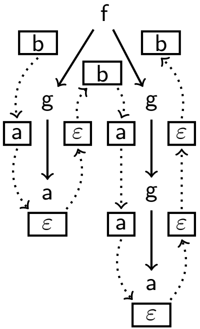


$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

Sequential Top-Down Tree-to-Word Transducers (STWs)

Machine definition

- Top Down tree traversal
- Deterministic
- Each node visited exactly once
- Output concatenation
- $q(f(x_1, x_2)) = b \cdot q'(x_1) \cdot b \cdot q'(x_2) \cdot b$
- $q'(g(x_1)) = a \cdot q'(x_1)$
- $q'(a) = \varepsilon$



$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

Expressivity

Input domain

- Regular tree language
- Ranked
- Path-closed (recognizable by deterministic top-down automata)

Output

- CFL

Transformation

- no reordering
- no copy

Problem

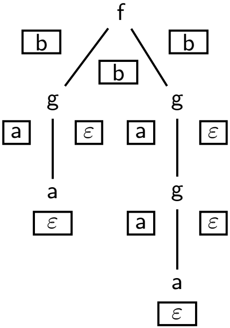
What's the most adapted normal form for such machine?
Does every STWs can be normalized in such form?
What's the cost of this operation?

Outline

On the road to an earliest form

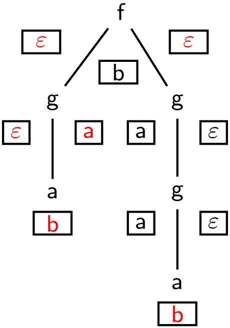
$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

earliest



$$q_0(f(x_1, x_2)) \to b \cdot q_1(x_1) \cdot b \cdot q_1(x_2) \cdot b$$
$$q_1(g(x_1)) \to a \cdot q_1(x_1)$$
$$q_1(a) \to \varepsilon$$

another definition

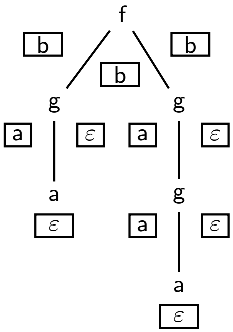


$$q_0(f(x_1, x_2)) \to q_2(x_1) \cdot b \cdot q_1(x_2)$$
$$q_1(g(x_1)) \to a \cdot q_1(x_1)$$
$$q_2(g(x_1)) \to q_2(x_1) \cdot a$$
$$q_1(a) \to b$$
$$q_2(a) \to b$$

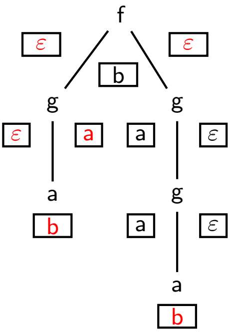
On the road to an earliest form

$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

<u>earliest</u>



<u>another definition</u>



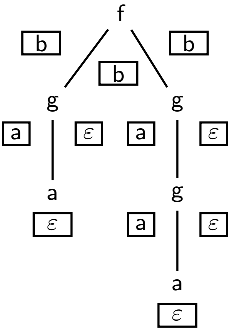
- as soon as possible
  - top-down
  - left-to-right
- smallest definition

- no clear semantic
- bigger definition

On the road to an earliest form

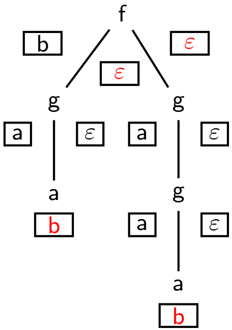
$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

earliest STW (wrt input)



$$q_0(f(x_1, x_2)) \to b \cdot q_1(x_1) \cdot b \cdot q_1(x_2) \cdot b$$
$$q_1(g(x_1)) \to a \cdot q_1(x_1)$$
$$q_1(a) \to \varepsilon$$

earliest wrt output (VPT)

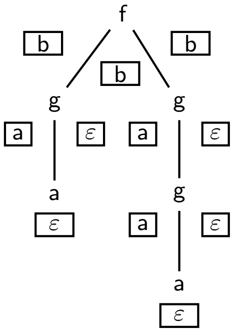


$$q_0(f(x_1, x_2)) \to b \cdot q_1(x_1) \cdot q_1(x_2)$$
$$q_1(g(x_1)) \to a \cdot q_1(x_1)$$
$$q_1(a) \to b$$

On the road to an earliest form

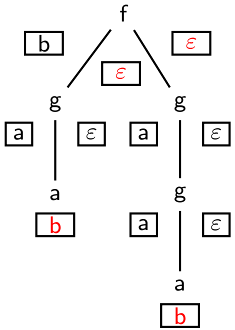
$$\boxed{f(g^n(a), g^m(a)) \mapsto b \cdot a^n \cdot b \cdot a^m \cdot b}$$

earliest STW (wrt input)



- as soon as possible
  - top-down
  - left-to-right
- same size

earliest wrt output (VPT)



- as soon as possible
  - depth first
- same size

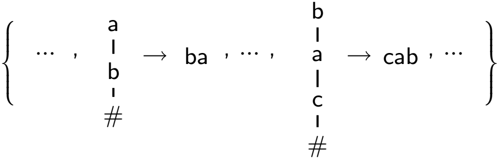
Bad news for earliest wrt output

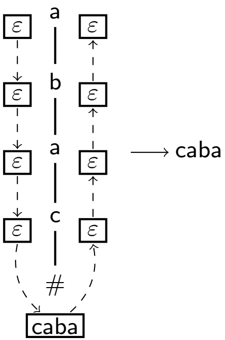
Reverse Transformation

$$\left\{ \quad \ldots \;, \quad \begin{array}{c} \mathsf{a} \\ | \\ \mathsf{b} \\ | \\ \# \end{array} \rightarrow \mathsf{ba} \;, \ldots, \quad \begin{array}{c} \mathsf{b} \\ | \\ \mathsf{a} \\ | \\ \mathsf{c} \\ | \\ \# \end{array} \rightarrow \mathsf{cab} \;, \ldots \quad \right\}$$

Bad news for earliest wrt output

Reverse Transformation

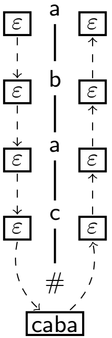




Bad news for earliest wrt output

Reverse Transformation

$$\left\{ \cdots , \begin{array}{c} a \\ | \\ b \\ | \\ \# \end{array} \rightarrow ba , \cdots , \begin{array}{c} b \\ | \\ a \\ | \\ c \\ | \\ \# \end{array} \rightarrow cab , \cdots \right\}$$



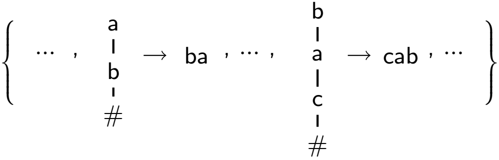
$\longrightarrow$ caba

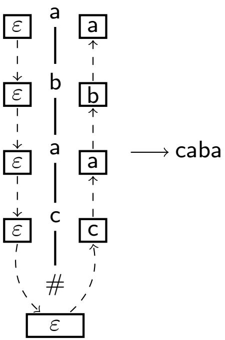
$$\begin{aligned}
q_\varepsilon(a(x_1)) &\rightarrow q_a(x_1) \\
q_a(b(x_1)) &\rightarrow q_{ba}(x_1) \\
&\cdots \\
q_w(a(x_1)) &\rightarrow q_{a\cdot w}(x_1) \\
q_w(b(x_1)) &\rightarrow q_{b\cdot w}(x_1) \\
q_w(c(x_1)) &\rightarrow q_{c\cdot w}(x_1) \\
&\cdots \\
q_w(\#) &\rightarrow w \\
&\cdots
\end{aligned}$$

Infinite number of states

Bad news for earliest wrt output

Reverse
Transformation





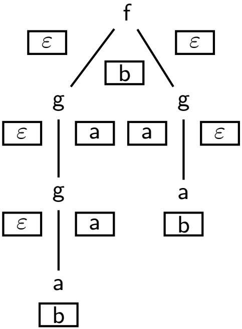
$$
\begin{aligned}
q(a(x_1)) &\to q(x_1)\cdot a\\
q(b(x_1)) &\to q(x_1)\cdot b\\
q(c(x_1)) &\to q(x_1)\cdot c\\
q(\#) &\to \varepsilon
\end{aligned}
$$

Outline

Push up output annotations

$$\boxed{f(g^m(a), g^n(a)) \mapsto b \cdot (a)^m \cdot b \cdot (a)^n \cdot b}$$

$$
\begin{aligned}
q_0(f(x_1, x_2)) &\to q_1(x_1) \cdot b \cdot q_2(x_2) \\
q_1(g(x_1)) &\to q_1(x_1) \cdot a \\
q_1(a) &\to b \\
q_2(g(x_1)) &\to a \cdot q_2(x_1) \\
q_2(a) &\to b
\end{aligned}
$$

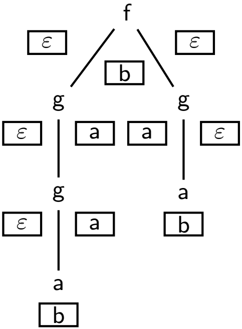


Output as soon as possible following the top-down order

Push up output annotations

$$\boxed{f(g^m(a), g^n(a)) \mapsto b \cdot (a)^m \cdot b \cdot (a)^n \cdot b}$$

$$\begin{aligned} q_0(f(x_1,x_2)) &\to q_1(x_1)\cdot b \cdot q_2(x_2) \\ q_1(g(x_1)) &\to q_1(x_1)\cdot a \\ q_1(a) &\to b \\ q_2(g(x_1)) &\to a \cdot q_2(x_1) \\ q_2(a) &\to b \end{aligned}$$

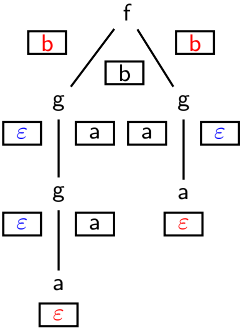


Output as soon as possible following the top-down order

Push up output annotations

$$\boxed{f(g^m(a), g^n(a)) \mapsto b \cdot (a)^m \cdot b \cdot (a)^n \cdot b}$$

$$q_0(f(x_1, x_2)) \to b \cdot q_1(x_1) \cdot b \cdot q_2(x_2) \cdot b$$
$$q_1(g(x_1)) \to q_1(x_1) \cdot a$$
$$q_1(a) \to \varepsilon$$
$$q_2(g(x_1)) \to a \cdot q_2(x_1)$$
$$q_2(a) \to \varepsilon$$

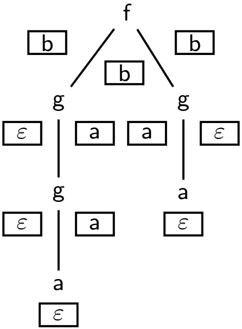


Output as soon as possible following the top-down order

Push left output annotations

$$\boxed{f(g^m(a), g^n(a)) \mapsto b \cdot (a)^m \cdot b \cdot (a)^n \cdot b}$$

$$
\begin{aligned}
q_0(f(x_1, x_2)) &\to b \cdot q_1(x_1) \cdot b \cdot q_2(x_2) \cdot b \\
q_1(g(x_1)) &\to q_1(x_1) \cdot a \\
q_1(a) &\to \varepsilon \\
q_2(g(x_1)) &\to a \cdot q_2(x_1) \\
q_2(a) &\to \varepsilon
\end{aligned}
$$

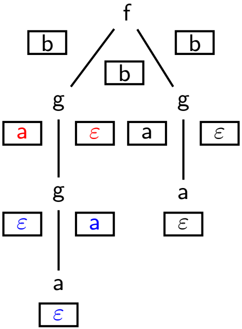


Output as soon as possible following the left-to-right preorder

Push left output annotations

$$\boxed{f(g^m(a), g^n(a)) \mapsto b \cdot (a)^m \cdot b \cdot (a)^n \cdot b}$$

$$
\begin{aligned}
q_0(f(x_1, x_2)) &\to b \cdot q_3(x_1) \cdot b \cdot q_2(x_2) \cdot b \\
q_1(g(x_1)) &\to q_1(x_1) \cdot a \\
q_1(a) &\to \varepsilon \\
q_2(g(x_1)) &\to a \cdot q_2(x_1) \\
q_2(a) &\to \varepsilon \\
q_3(g(x_1)) &\to a \cdot q_2(x_1)
\end{aligned}
$$

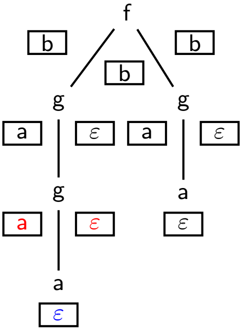


Output as soon as possible following the left-to-right preorder

Push left output annotations

$$\boxed{f(g^m(a), g^n(a)) \mapsto b \cdot (a)^m \cdot b \cdot (a)^n \cdot b}$$

$$
\begin{aligned}
q_0(f(x_1, x_2)) &\to b \cdot q_3(x_1) \cdot b \cdot q_2(x_2) \cdot b \\
q_1(g(x_1)) &\to q_1(x_1) \cdot a \\
q_1(a) &\to \varepsilon \\
q_2(g(x_1)) &\to a \cdot q_2(x_1) \\
q_2(a) &\to \varepsilon \\
q_3(g(x_1)) &\to a \cdot q_3(x_1) \\
q_3(a) &\to \varepsilon
\end{aligned}
$$

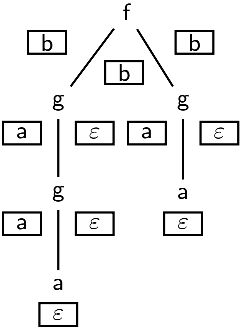


Output as soon as possible following the left-to-right preorder

Push left output annotations

$$\boxed{f(g^m(a), g^n(a)) \mapsto b \cdot (a)^m \cdot b \cdot (a)^n \cdot b}$$

$$\begin{aligned} q_0(f(x_1, x_2)) &\to b \cdot q_2(x_1) \cdot b \cdot q_2(x_2) \cdot b \\ q_2(g(x_1)) &\to a \cdot q_2(x_1) \\ q_2(a) &\to \varepsilon \end{aligned}$$



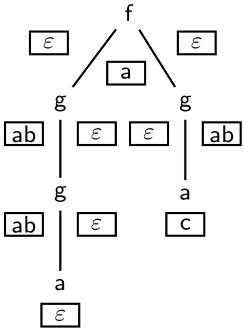
Output as soon as possible following the left-to-right preorder

Combinatorial problems

$$\boxed{f(g^m(a), g^n(a)) \mapsto (ab)^m \cdot ac \cdot (ab)^n}$$

$$\begin{aligned}
q_0(f(x_1, x_2)) &\to q_1(x_1) \cdot a \cdot q_2(x_2) \\
q_1(g(x_1)) &\to ab \cdot q_1(x_1) \\
q_1(a) &\to \varepsilon \\
q_2(g(x_1)) &\to q_2(x_1) \cdot ab \\
q_2(a) &\to c
\end{aligned}$$

Applying an operation can lead to:

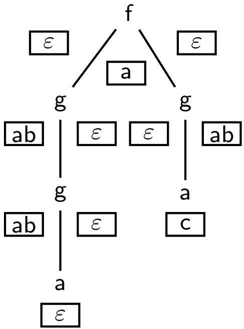


Combinatorial problems

$$\boxed{f(g^m(a), g^n(a)) \mapsto a \cdot (ba)^m \cdot c \cdot (ab)^n}$$

$$\begin{aligned}
q_0(f(x_1, x_2)) &\to q_1(x_1) \cdot a \cdot q_2(x_2)\\
q_1(g(x_1)) &\to ab \cdot q_1(x_1)\\
q_1(a) &\to \varepsilon\\
q_2(g(x_1)) &\to q_2(x_1) \cdot ab\\
q_2(a) &\to c
\end{aligned}$$

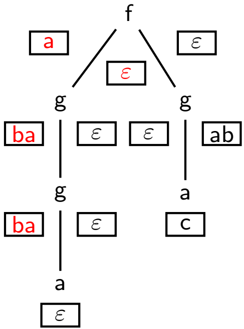
Applying an operation can lead to:



Combinatorial problems

$$\boxed{f(g^m(a), g^n(a)) \mapsto a \cdot (ba)^m \cdot c \cdot (ab)^n}$$

$$
\begin{aligned}
q_0(f(x_1, x_2)) &\to a \cdot q_1(x_1) \cdot q_2(x_2) \\
q_1(g(x_1)) &\to ba \cdot q_1(x_1) \\
q_1(a) &\to \varepsilon \\
q_2(g(x_1)) &\to q_2(x_1) \cdot ab \\
q_2(a) &\to c
\end{aligned}
$$



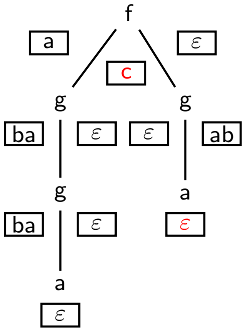
Applying an operation can lead to:

- modify other output annotations

Combinatorial problems

$$\boxed{f(g^m(a), g^n(a)) \mapsto a \cdot (ba)^m \cdot c \cdot (ab)^n}$$

$$q_0(f(x_1, x_2)) \rightarrow a \cdot q_1(x_1) \cdot c \cdot q_2(x_2)$$
$$q_1(g(x_1)) \rightarrow ba \cdot q_1(x_1)$$
$$q_1(a) \rightarrow \varepsilon$$
$$q_2(g(x_1)) \rightarrow q_2(x_1) \cdot ab$$
$$q_2(a) \rightarrow \varepsilon$$



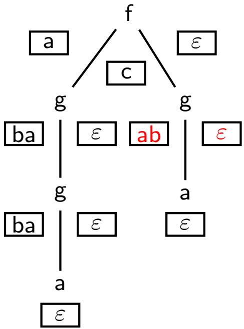
Applying an operation can lead to:

- modify other output annotations

Combinatorial problems

$$\boxed{f(g^m(a), g^n(a)) \mapsto a \cdot (ba)^m \cdot c \cdot (ab)^n}$$

$$\begin{aligned} q_0(f(x_1, x_2)) &\to a \cdot q_1(x_1) \cdot c \cdot q_2(x_2) \\ q_1(g(x_1)) &\to ba \cdot q_1(x_1) \\ q_1(a) &\to \varepsilon \\ q_2(g(x_1)) &\to ab \cdot q_2(x_1) \\ q_2(a) &\to \varepsilon \end{aligned}$$



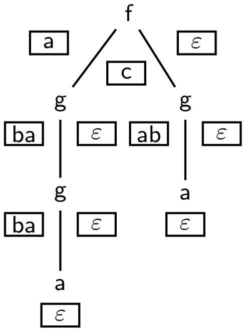
Applying an operation can lead to:

- modify other output annotations
- reapply other operations

Combinatorial problems

$$\boxed{f(g^m(a), g^n(a)) \mapsto a \cdot (ba)^m \cdot c \cdot (ab)^n}$$

$$\begin{aligned}
q_0(f(x_1, x_2)) &\to a \cdot q_1(x_1) \cdot c \cdot q_2(x_2) \\
q_1(g(x_1)) &\to ba \cdot q_1(x_1) \\
q_1(a) &\to \varepsilon \\
q_2(g(x_1)) &\to ab \cdot q_2(x_1) \\
q_2(a) &\to \varepsilon
\end{aligned}$$



Applying an operation can lead to:

- modify other output annotations
- reapply other operations

Normalization results

Theorem

For an STW M, we can construct an equivalent earliest STW M' in time polynomial in the size of M'.

M' size is at most doubly-exponential in the size of M.

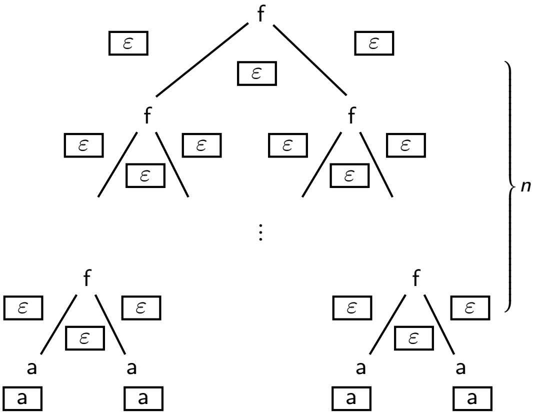
Proof techniques

- Constructive
- Combinatory on words

Outline

Explosion of the size of rules

- **input:** a perfect (well-balanced) binary tree of height n
- **output:** concatenation of all leaves, erasing inner nodes



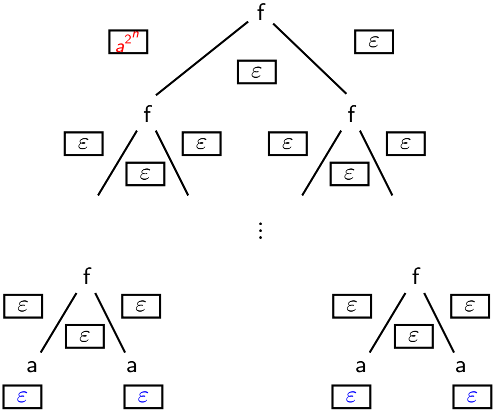
output: a^{2^n}

For $0 \leq i < n$:

$$q_i(f(x_1, x_2)) \rightarrow q_{i+1}(x_1) \cdot q_{i+1}(x_2)$$
$$q_n(a) \rightarrow a$$

Explosion of the size of rules

- **input:** a perfect (well-balanced) binary tree of height n
- **output:** concatenation of all leaves, erasing inner nodes

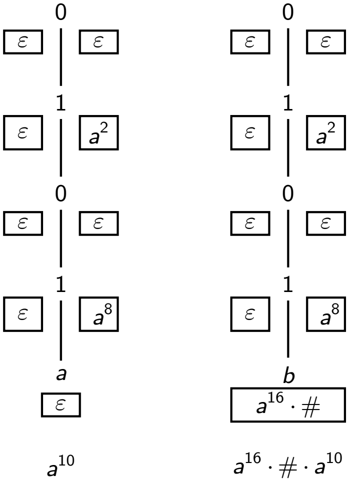


For $0 < i < n$:

$$q_0(f(x_1, x_2)) \to a^{2^n} \cdot q_1(x_1) \cdot q_1(x_2)$$
$$q_i(f(x_1, x_2)) \to q_{i+1}(x_1) \cdot q_{i+1}(x_2)$$
$$q_n(a) \to \varepsilon$$

Size of the first rule is exponential in the size of the initial transducer

Explosion of the number of states



- **input**: $(0|1)^n = x$ binary encoding followed by a or b
- **output**: start always by a^x

For $0 \leq i < n$:

$$q_i(0(x_1)) \to q_{i+1}(x_1)$$

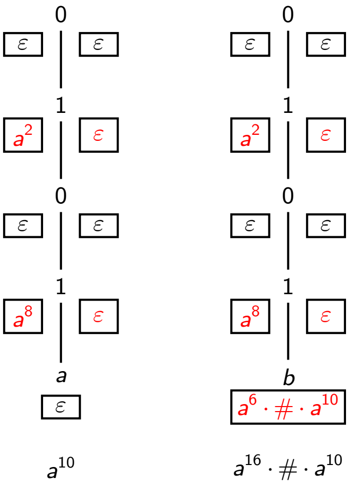
$$q_i(1(x_1)) \to q_{i+1}(x_1) \cdot a^{2^i}$$

$$q_n(a) \to \varepsilon$$

$$q_n(b) \to a^{2^n} \cdot \#$$

n states

Explosion of the number of states



- **input**: $(0|1)^n = x$ binary encoding followed by a or b
- **output**: start always by a^x

For $0 \leq i < n$:

$$q_i^k(0(x_1)) \rightarrow q_{i+1}^k(x_1)$$

$$q_i^k(1(x_1)) \rightarrow a^{2^i} \cdot q_{i+1}^{k+2^i}(x_1)$$

$$q_n^k(a) \rightarrow \varepsilon$$

$$q_n^k(b) \rightarrow a^{2^n-k} \cdot \# \cdot a^k$$

$\sum_{i=0}^{n} 2^i$ states

Obtained number of states is exponential on the number of the initial transducer states

Outline

Minimization

Theorem

Minimization of earliest STWs is in PTIME.

Lemma

Between two earliest STWs, states can be duplicated but keep sames rules.

The equivalence test on earliest STWs is in PTIME [Staworko et Al. FCT 2009]

Lemma

Minimization of arbitrary STWs is NP-complete.

Conclusion

Result

We have defined the class of STW :

- nice expressivity
- with good properties
  - decidable equivalence in PTIME
  - normalization algorithm

Perspectives

- learning algorithm
- allow swapping