Trees with decidable theories

Damian Niwiński

University of Warsaw

joint work with Teodor Knapik, Paweł Urzyczyn, and Igor Walukiewicz

ASL North American Annual Meeting, Boulder, CO, May 2014

Decidable vs. undecidable

Turing, Church (1936). Arithmetic of natural numbers is undecidable.

All “interesting” mathematical theories are undecidable.

But

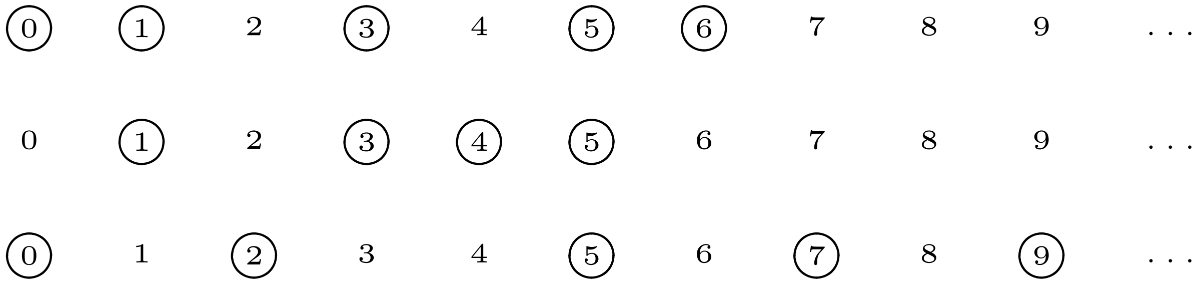
- Decidability of mathematical theories is crucial in automatic verification.
- Delimitating decidable fragments of an undecidable theory (e.g., arithmetics) reveals a fine structure of the theory.

Büchi (1960). Monadic second order theory (MSO) of $\langle \omega, succ \rangle$ is decidable.

This subsumes, among others,

Presburger (1929). First order theory of $\langle \omega, + \rangle$ is decidable.

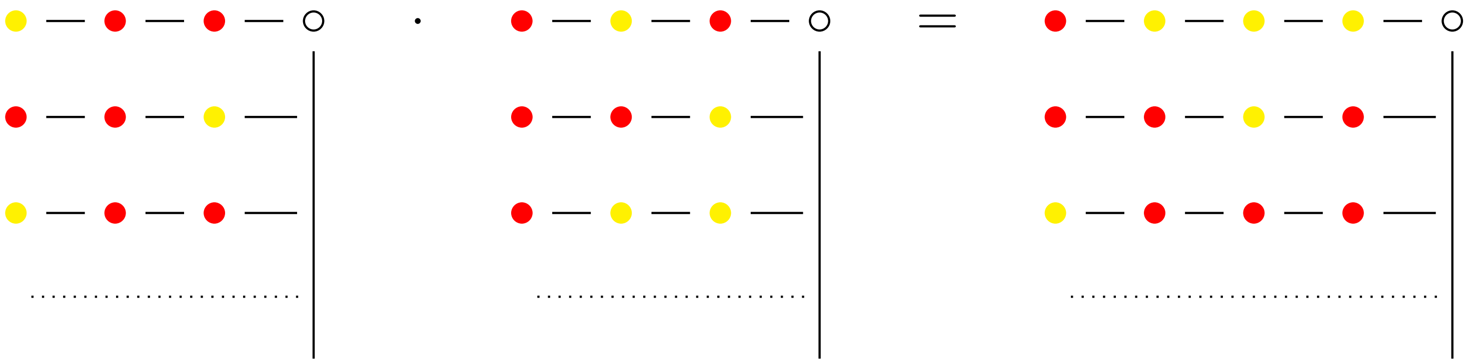
$$\begin{array}{cccccccc}
1 & 1 & 0 & 1 & 0 & 1 & 1 & \\
0 & 1 & 0 & 1 & 1 & 1 & & + \\
\hline
1 & 0 & 1 & 0 & 0 & 1 & 0 & 1
\end{array}$$



Rabin (1969). MSO theory of $\mathbb{T}_2 = \langle 2^*, succ_0, succ_1 \rangle$ is decidable.

This subsumes, among others,

Skolem (1930). First order theory of $\langle \omega, \cdot \rangle$ is decidable.



$$2^{3} \cdot 3^{6} \cdot 5^{3} \quad \cdot \quad 2^{5} \cdot 3^{6} \cdot 5^{4} \quad = \quad 2^{8} \cdot 3^{12} \cdot 5^{7}$$

A great number of decidability results follows from Rabin's theorem.

An equivalent formalism of tree automata is used for better complexity bounds.

An interpretation of a structure $\mathcal{A} \hookrightarrow \mathbb{T}_2$ yields decidability of $Th(\mathcal{A})$.

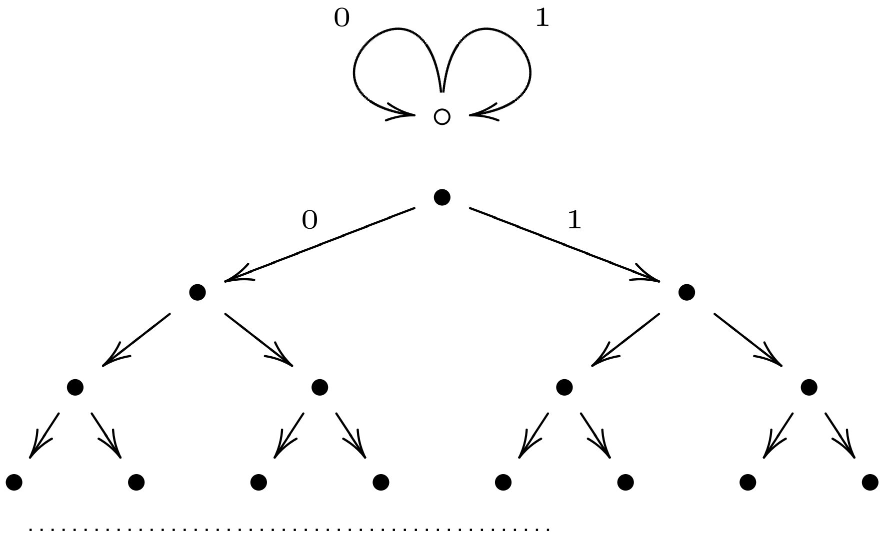
Another construction interprets **all*** **models** of a formula.

$$\varphi \quad \mapsto \quad \mathbf{\Phi}(X)$$
$$\mathcal{A} \models \varphi \quad \Leftrightarrow \quad \mathbb{T}_2 \models \mathbf{\Phi}[\mathcal{A}]$$

This yields decidability of the **satisfiability** problem for numerous logics with the **tree model property**.

Grädel & Walukiewicz (1999). Guarded first-order logic with fixed points is decidable.

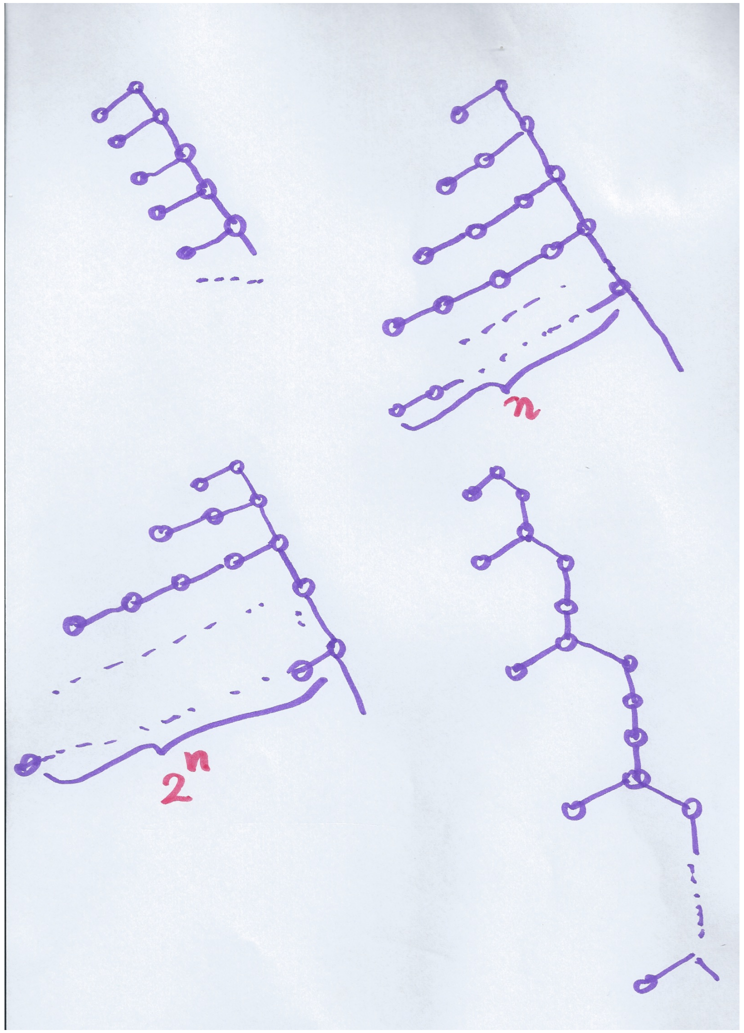
Generalizations of Rabin's Theorem



Courcelle & Walukiewicz (1997). The MSO theory of the unfolding of a graph reduces to the MSO theory of the original graph.

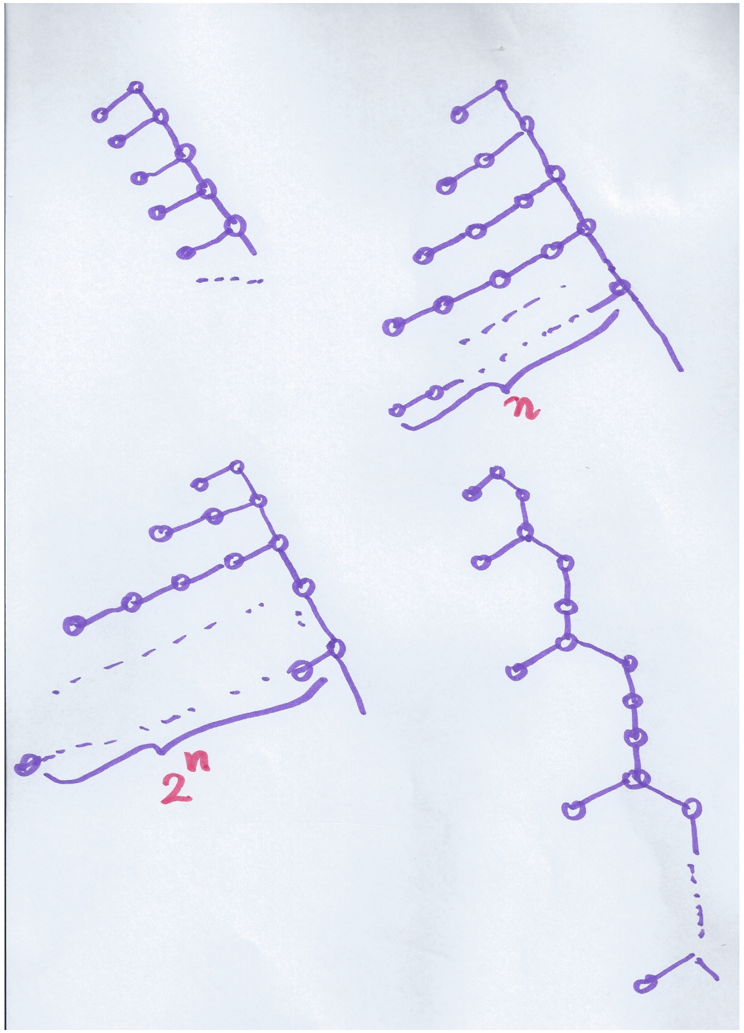
Muchnik (unpublished, ca. 1990), **Walukiewicz 1996.** The MSO theory of a tree-like structure M^* over an arbitrary structure M reduces to the MSO theory of M.

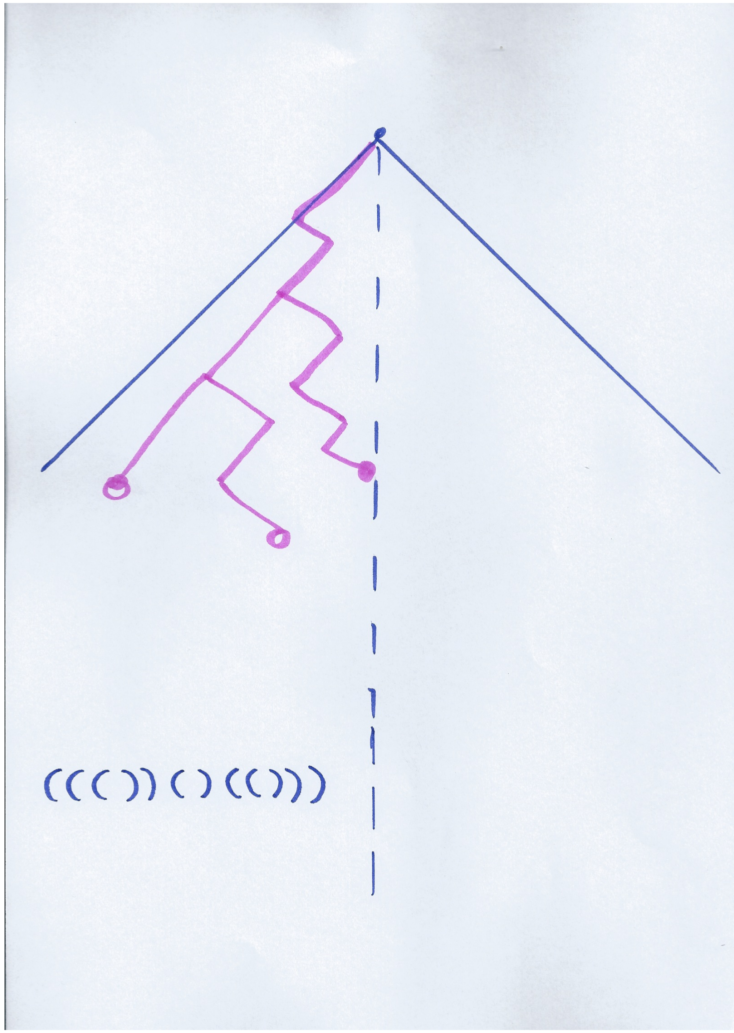
What about different shapes of trees ?



MSO theory of a recursive tree can be $\mathbf{\Pi}^1_1$-hard (cf. **Thomas 2010**).

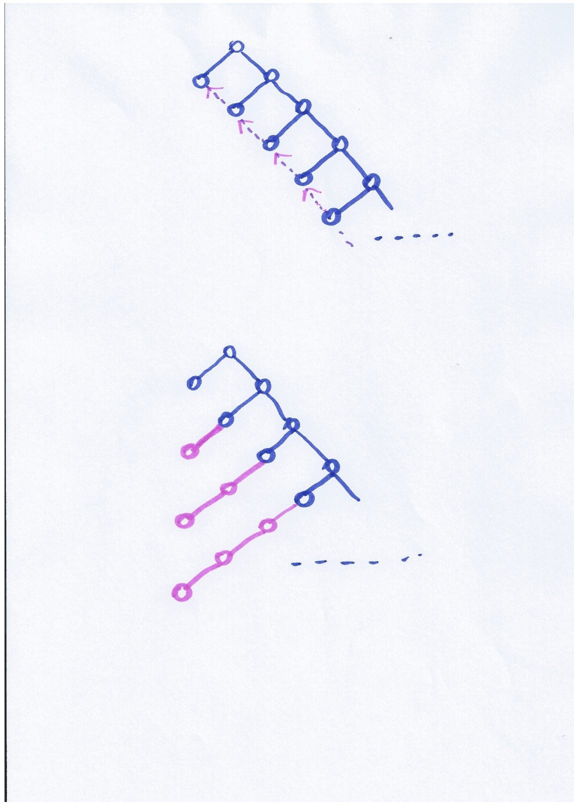
On positive side





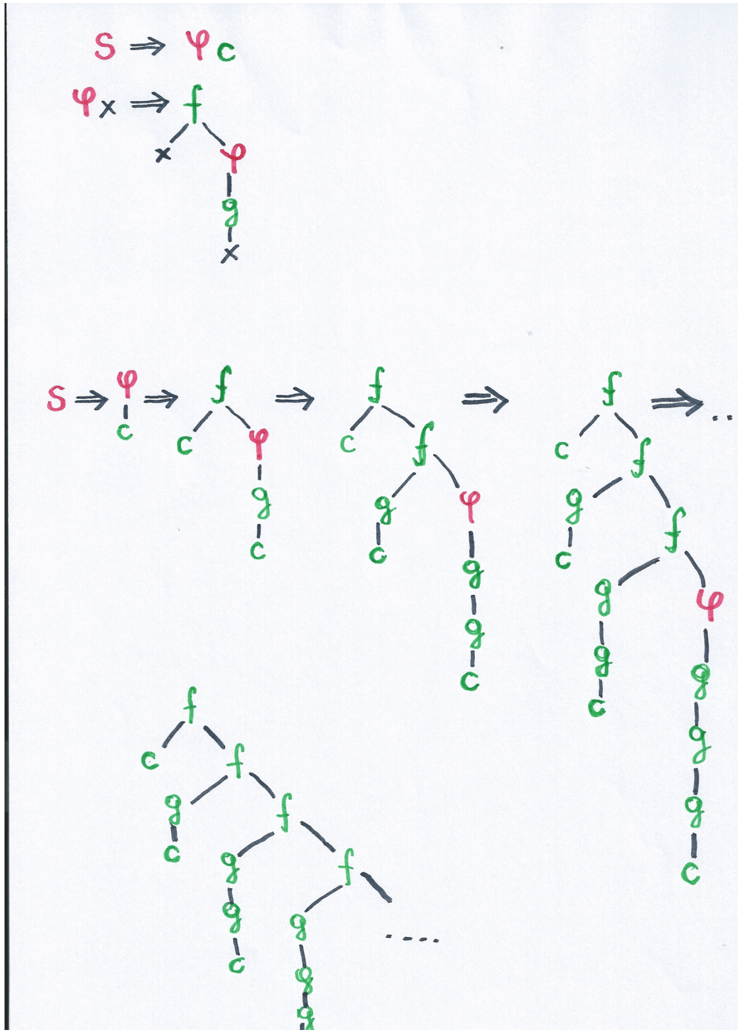
MSO theories of *algebraic* trees are decidable (cf. **Courcelle** 1995).

Interpretation + unfolding + interpretation + unfolding ...



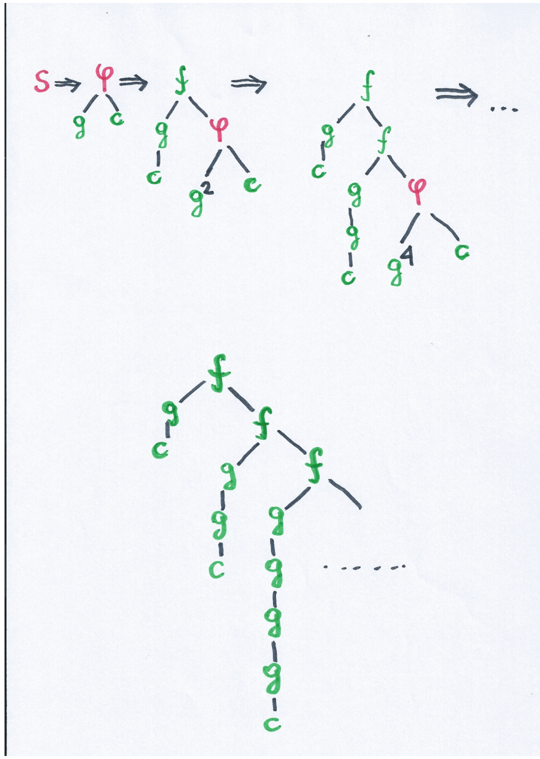
Caucal observed (in 1990s) that alternating interpretation and unfolding gives rise to a rich family of trees. This resulted in **Caucal's hierarchy (2002)**.

Generating trees by 1st order grammars (algebraic)



Generating trees by 2nd order grammars

$$S \Rightarrow \phi g c$$
$$\phi \xi x \Rightarrow f(\xi x)\,(\phi(Copy \xi) x)$$
$$Copy \xi z \Rightarrow \xi(\xi z)$$



Higher-order tree grammars — definitions

Types $\mathcal{T}$ $\quad \tau ::= \mathbf{0} \,|\, \tau \to \tau$

Nonterminals $N = \{N_\tau\}_{\tau \in \mathcal{T}}$

Variables $\mathcal{X} = \{\mathcal{X}_\tau\}_{\tau \in \mathcal{T}}$

Signature constants $f, g, c, \ldots : \mathbf{0}^k \to \mathbf{0}$

Grammar $\mathcal{G} = (\Sigma, V, S, E)$

with Σ a *signature*, $V \subseteq \bigcup_{\tau \in \mathcal{T}} N_\tau$, $\quad V \ni S : \mathbf{0}$,

and E a finite set of *productions* of the form

$$\mathcal{F} z_1 \ldots z_m \quad \Rightarrow \quad w$$

with $V \ni \mathcal{F} : \tau_1 \to \tau_2 \cdots \to \tau_m \to \mathbf{0}$, $\quad z_i \in \mathcal{X}_{\tau_i}$,

and w an applicative term over $\Sigma \cup V \cup \{z_1 \ldots z_m\}$ of type $\mathbf{0}$.

Derivations

We assume that a grammar $\mathcal{G}$ is **deterministic**,

i.e., one production per nonterminal.

Hence there is a unique outermost derivation

$$S = t_0 \to_{\mathcal{G}} t_1 \to_{\mathcal{G}} t_2 \to_{\mathcal{G}} \dots$$

producing the tree $[\![\mathcal{G}]\!]$ generated by $\mathcal{G}$.

Levels

$\ell(\mathbf{0}) = 0,\qquad \ell(\tau_1 \to \tau_2) = \max(1 + \ell(\tau_1), \ell(\tau_2))$

The model checking problem

Given a grammar $\mathcal{G}$ and a formula φ, decide if $[\![\mathcal{G}]\!] \models \varphi$.

Here, a tree $t : \{1, 2, \ldots, M\}^* \supseteq dom\, t \to \{f, g, c, \ldots\}$ is considered as a logical structure

$$\mathbf{t} = \left\langle dom\, t, f^{\mathbf{t}}, g^{\mathbf{t}}, c^{\mathbf{t}}, \ldots,\ succ_1^{\mathbf{t}}, \ldots, succ_M^{\mathbf{t}} \right\rangle$$

where $f^{\mathbf{t}}(w) \Leftrightarrow t(w) = f$, and
$succ_i^{\mathbf{t}}(w, wi)$, whenever $wi \in dom\, t$.

Reduction of a grammar $\mathcal{G}$ of level n to $\mathcal{G}^{\alpha}$ of level $n-1$

For types, $\tau \mapsto \tau^{\alpha}$,

- $\alpha : \mathbf{0} \mapsto \mathbf{0}$,
- $\alpha : (\mathbf{0}^k \to \mathbf{0}) \mapsto \mathbf{0}$,
- $\alpha : (\tau_1 \to \cdots \to \tau_n) \mapsto (\tau_1^{\alpha} \to \cdots \to \tau_n^{\alpha})$

For terms, $t : \tau \mapsto t^{\alpha} : \tau^{\alpha}$,

- $\alpha : \mathcal{F} \mapsto \mathcal{F}^{\alpha}$,
- $\alpha : z \mapsto z$, for any parameter z,
- $\alpha : (ts) \mapsto (t^{\alpha} s^{\alpha})$, whenever $s : \tau$ with $\ell(\tau) \geq 1$,
- $\alpha : (ts) \mapsto ((@t^{\alpha})s^{\alpha})$, whenever $s : \mathbf{0}$ (hence $t^{\alpha}, s^{\alpha} : \mathbf{0}$).

Reduction of grammars cont'd

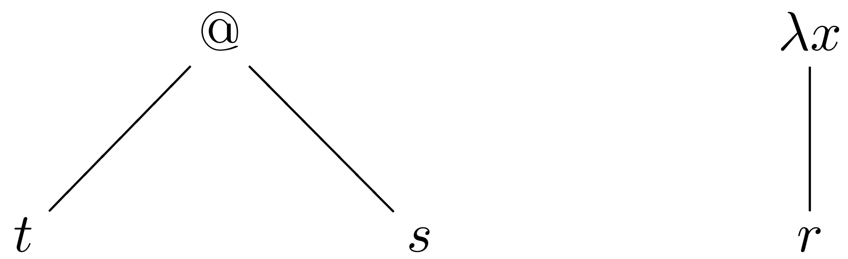
$$\mathcal{G} = (\Sigma, V, S, E) \quad \mapsto \quad \mathcal{G}^{\alpha} = (\Sigma^{\alpha}, V^{\alpha}, S^{\alpha}, E^{\alpha})$$

where

$E : \mathcal{F}\phi_1 \ldots \phi_m y_1 \ldots y_n \Rightarrow r,$ with $y_1 \ldots y_n : \mathbf{0}$ then

$E^{\alpha} : \mathcal{F}^{\alpha}\phi_1 \ldots \phi_m \Rightarrow \lambda y_1 \ldots \lambda y_n . r^{\alpha}.$

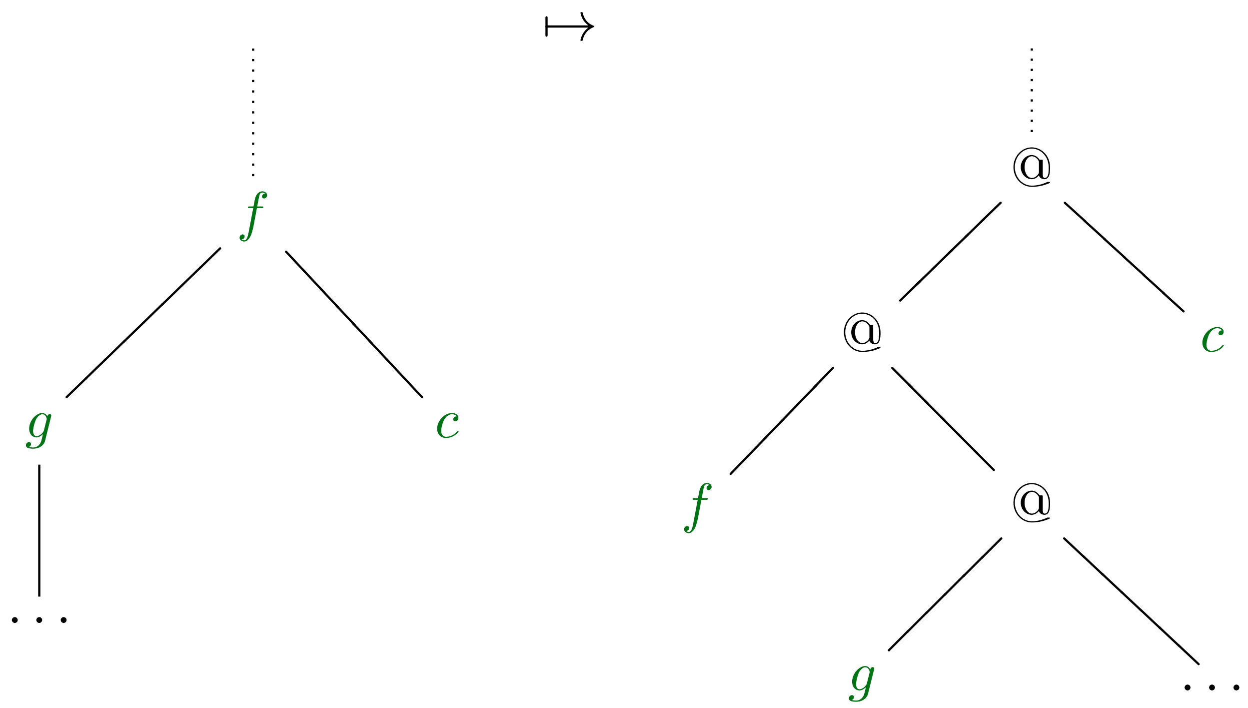
Here the λy_i's and @ are new constants with $\lambda y_i : \mathbf{0} \to \mathbf{0}$ and $@ : \mathbf{0}^2 \to \mathbf{0}$.



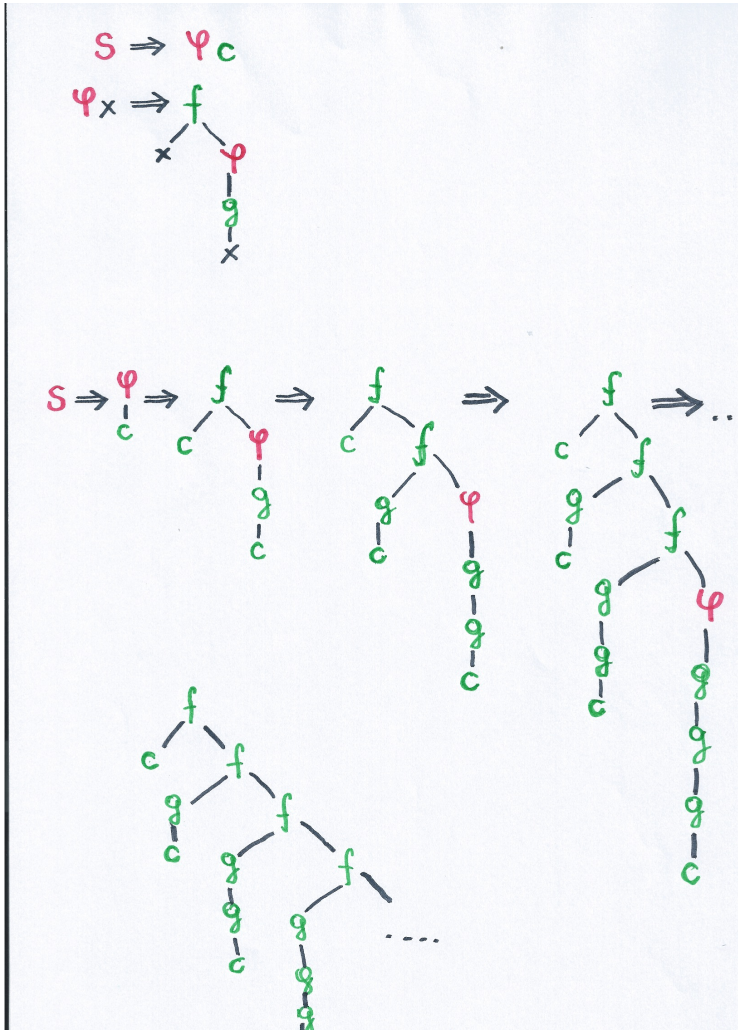
The tree is a $[\![\mathcal{G}^{\alpha}]\!]$ is a λ-definition of $[\![\mathcal{G}]\!]$.

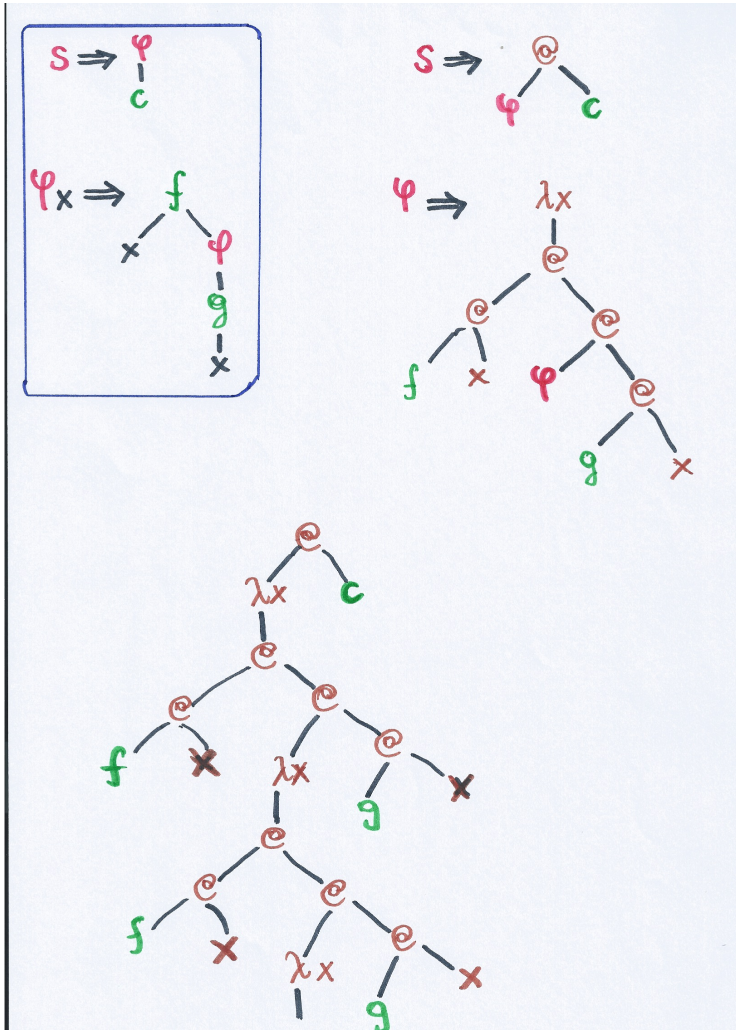
Goal: to interpret

$[\![\mathcal{G}]\!]$ in $[\![\mathcal{G}^{\color{red}\alpha}]\!]$



Reduction level 1 to level 0 – example





Reduction level 2 to level 1 – example

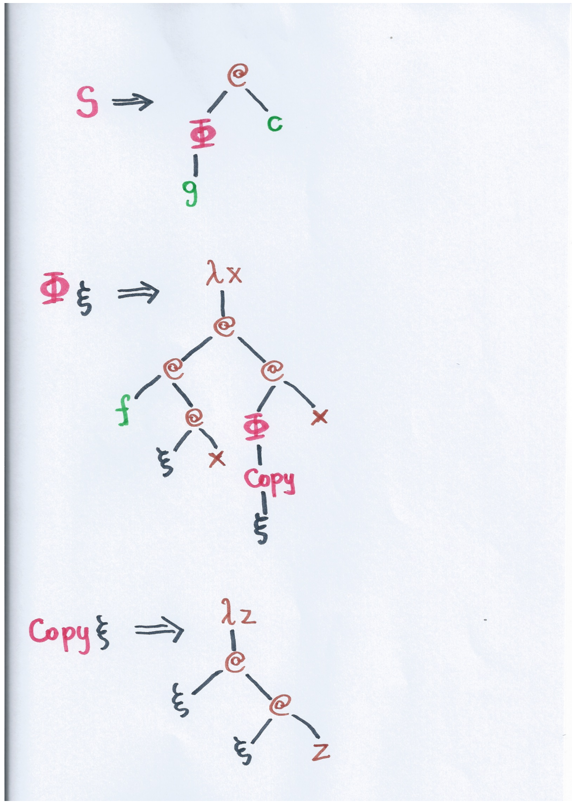
$$
\begin{array}{lcl}
S & \Rightarrow & \phi g c \\
\phi \xi x & \Rightarrow & f\,(\xi x)\,(\phi(Copy\xi)x) \\
Copy\xi z & \Rightarrow & \xi(\xi z)
\end{array}
$$

$$\Downarrow$$

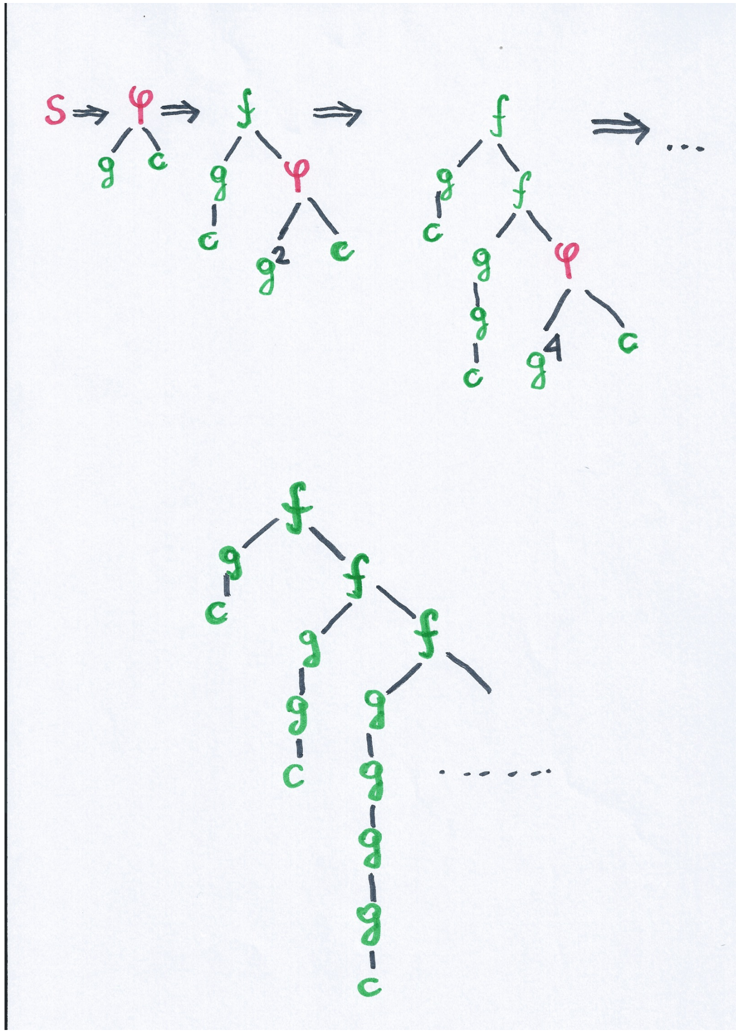
$$
\begin{array}{lcl}
S & \Rightarrow & @(\phi g)c \\
\phi \xi & \Rightarrow & \lambda x\,@\,(@f(@\xi x))\;(@\phi(Copy\xi)x) \\
Copy\xi & \Rightarrow & \lambda z\,@\xi\,(@\xi z)
\end{array}
$$

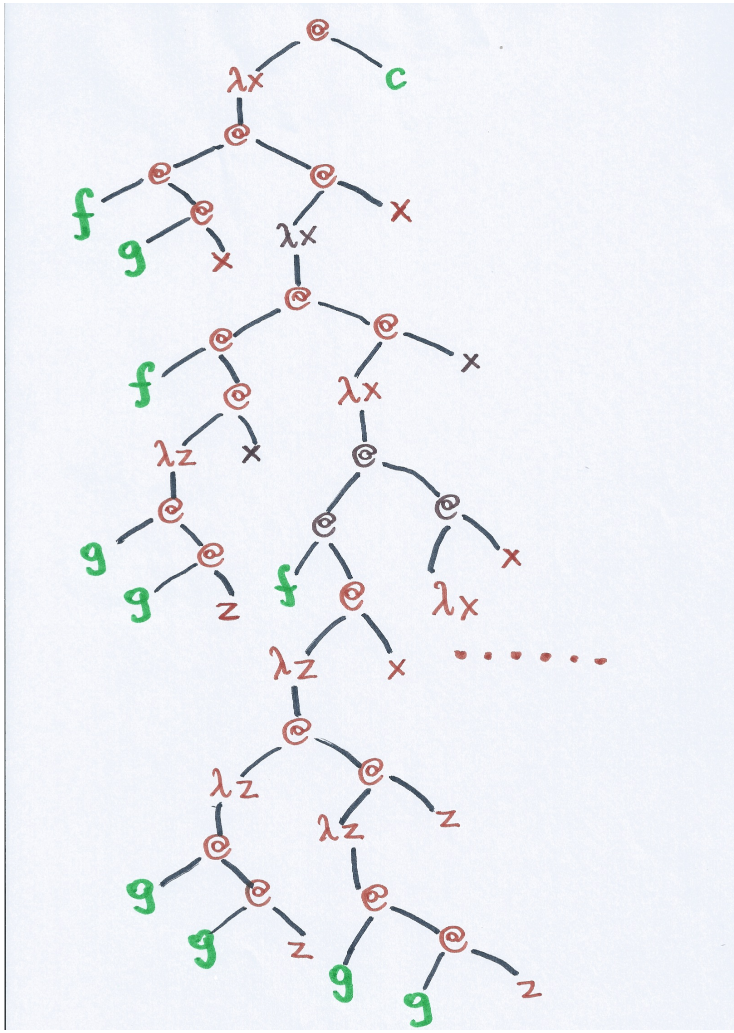
$$\Downarrow$$

S	$\Rightarrow$	$@(\phi g)c$
$\phi\xi$	$\Rightarrow$	$\lambda x\, @\, (@f(@\xi x))\ (@\phi(Copy\xi)x)$
$Copy\xi$	$\Rightarrow$	$\lambda z\, @\xi\, (@\xi z)$

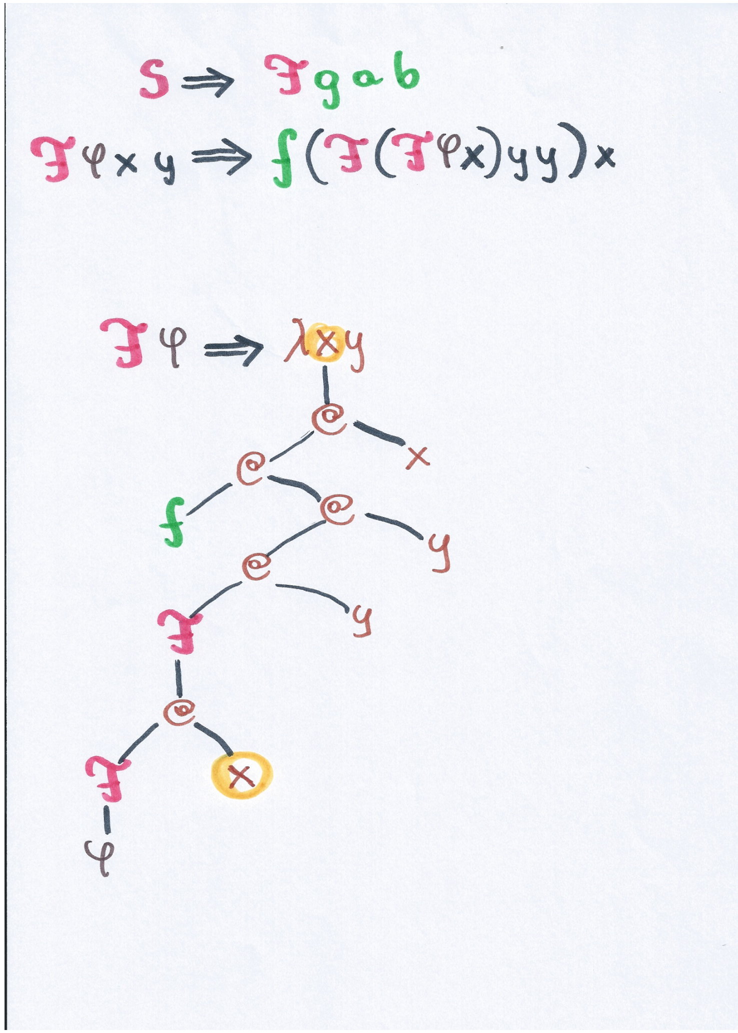


Reduction level 2 to level 1 – example cont'd

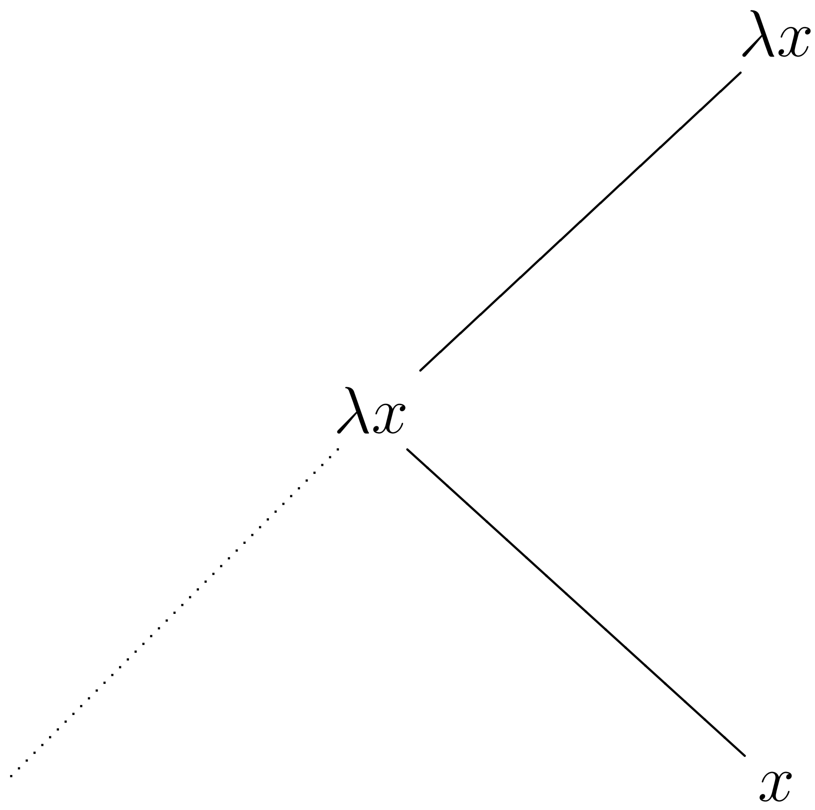


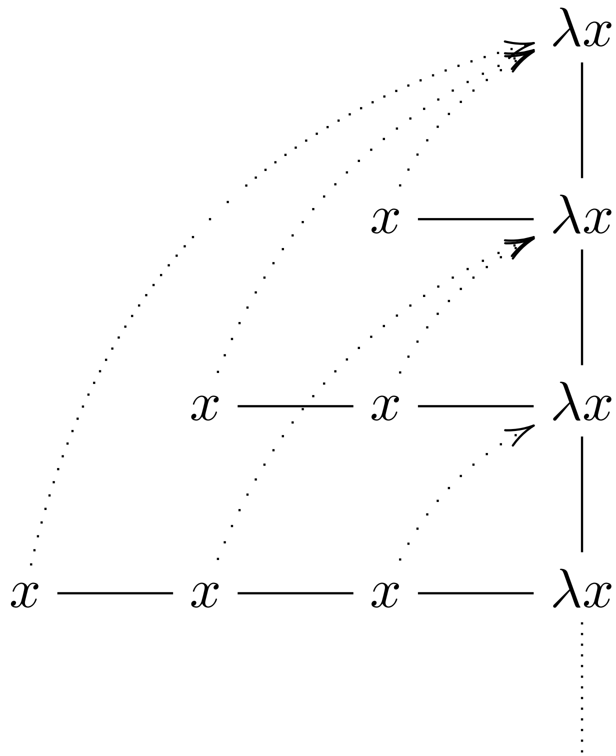


A problem may arise with a conflict of binding.



Ambiguity.





Explicit definition of binding leads to **undecidability**.

A term of level $k > 0$ is *unsafe* if it contains an occurrence of a parameter of level strictly less than k.

An *occurrence* of an unsafe term t is *unsafe*, unless it is in the context $\ldots (ts) \ldots$

$$\mathcal{F} \varphi x y \quad \Rightarrow \quad f \, (\mathcal{F}(\mathcal{F} \varphi \; \textcircled{x} \;) y y) \, x$$

A grammar without such occurrences is **safe**.

Note. If a grammar $\mathcal{G}$ is safe, so is $\mathcal{G}^{\alpha}$.

Lemma. If $\mathcal{G}$ is safe then the MSO theory of the tree $[\![\mathcal{G}]\!]$ is recursively reducible to the MSO theory of the tree $[\![\mathcal{G}^{\alpha}]\!]$.

Note. A grammar $\mathcal{G}$ of level ≤ 1 is always safe and $[\![\mathcal{G}]\!]$ has decidable MSO theory.

Theorem (KNU 2002). The MSO theory of the tree generated by a safe grammar of any level is decidable.

Theorem (Caucal 2002). The hierarchy of trees generated by safe grammars of level n coincides with the hierarchy obtained by interpretation $+$ unfolding
($\rightarrow$ **Caucal's hierarchy**).

But safety is not the frontier of decidability.

Theorem (Ong 2006). The MSO theory of the tree generated by any grammar is decidable.

Preceded by Aehlig, de Miranda and Ong 2005 for level **2**, and independently KNUW 2005, *via* **panic automata** (of level 2).

Further development

Hague, Murawski, Ong and Serre 2008: another proof *via* **collapsible automata** of any level.

Kobayashi & Ong 2009: another proof *via* a type system.

Salvati & Walukiewicz 2012: another proof *via* **Krivine machine**.

Language-theoretic characterization of trees

By the complexity of sets of words $\{w \in dom\, t : t(w) = f\}$.

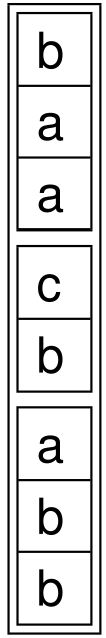
Let $t = [\![\mathcal{G}]\!]$.

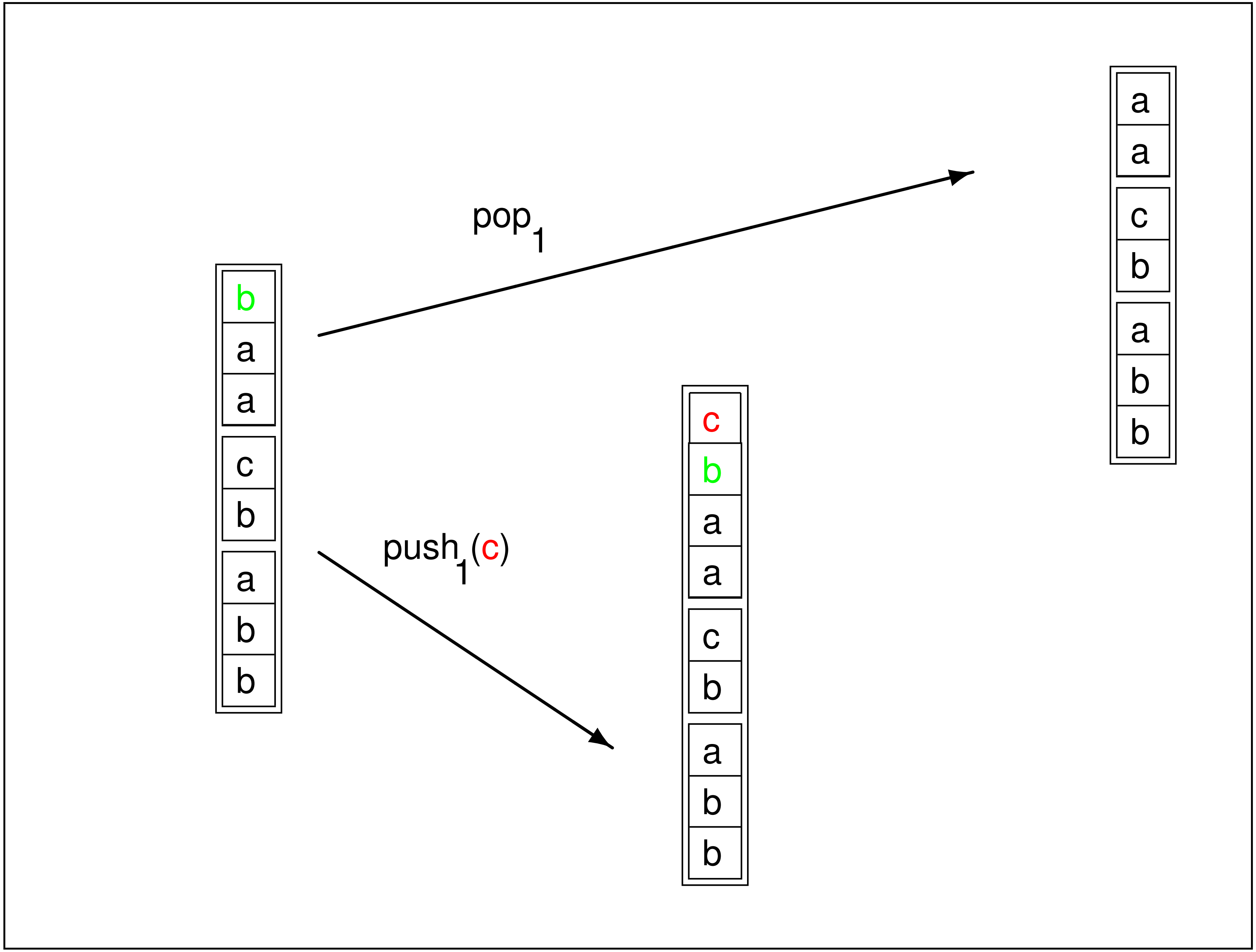
level 0	regular	
level 1	deterministic pushdown	Courcelle
safe level n	deterministic pushdown of level n	KNU 2002
level 2	panic automata	KNUW 2005
level n	collapsible automata of level n	HMOS 2008

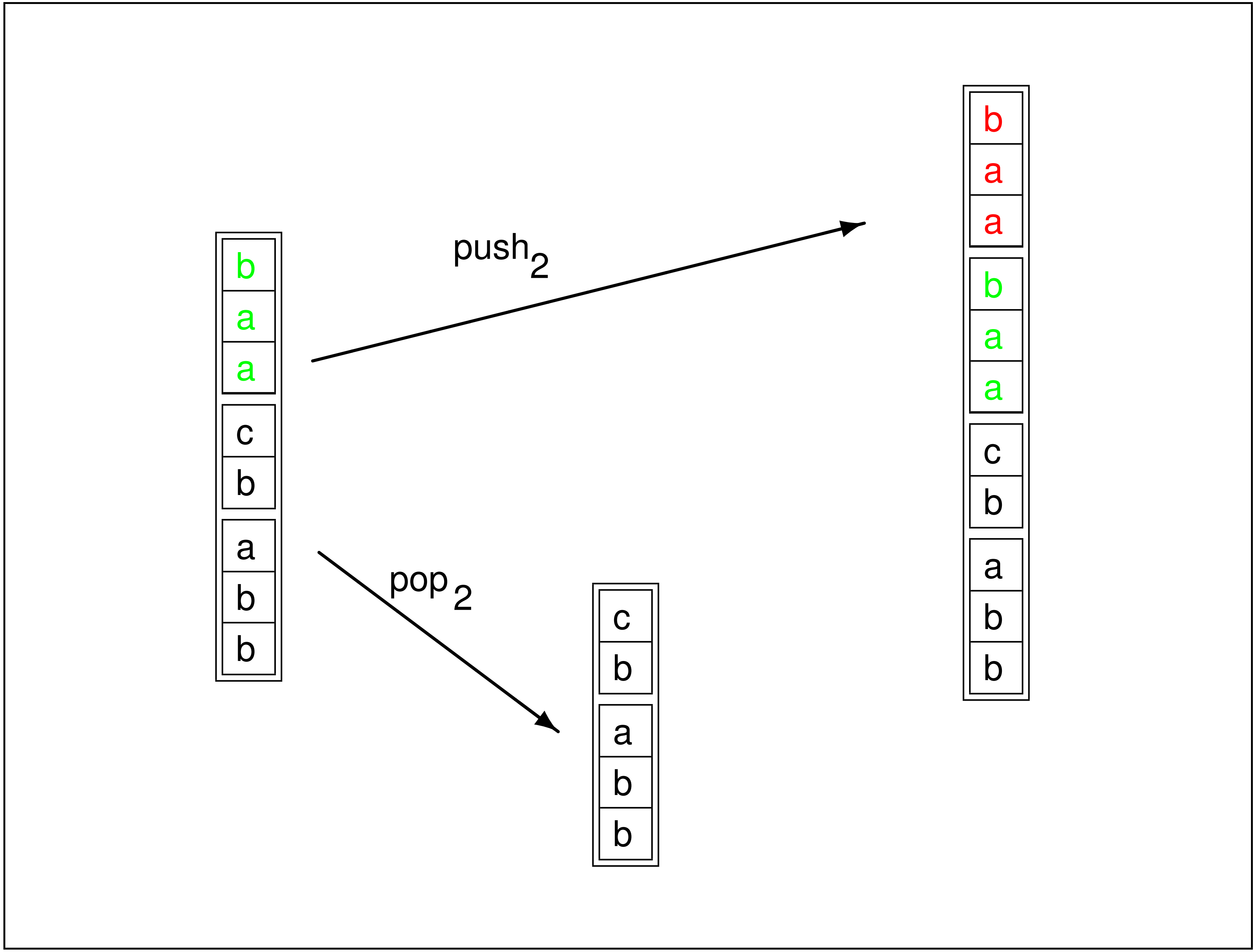
Parys 2012 used these characterizations to separate **safe** from **unsafe** grammars.

Higher order pushdown store

Maslov 1974




pop1
push1(c)
b
a
a
c
b
a
b
b
c
b
a
a
c
b
a
b
b
a
a
c
b
a
b
b



push2
pop 2
b
a
a
c
b
a
b
b

Second-order pushdown stores

A *level* 1 *pushdown store* is a non-empty word $a_1 \ldots a_k$ over Γ.

A *level* 2 *pds* is a non-empty sequence of 1-pds' $[s_1][s_2]\ldots[s_l]$.

Operations :

$$
\begin{aligned}
push_1\langle a\rangle([s_1][s_2]\ldots[s_l][w]) &= [s_1][s_2]\ldots[s_l][wa]\\
pop_1(\alpha[w\xi]) &= \alpha[w]\\
push_2(\alpha[w]) &= \alpha[w][w]\\
pop_2(\alpha[v][w]) &= \alpha[v]
\end{aligned}
$$

[⊥]	$push_1\langle a\rangle$
[⊥ a]	
[⊥ a b]	$push_2$
[⊥ a b] [⊥ a b]	pop_1
[⊥ a b] [⊥ a]	
[⊥ a b] [⊥ a a]	
[⊥ a b] [⊥ a a] [⊥ a a]	
[⊥ a b] [⊥ a a] [⊥ a a] [⊥ a a]	
[⊥ a b] [⊥ a a] [⊥ a a]	pop_2

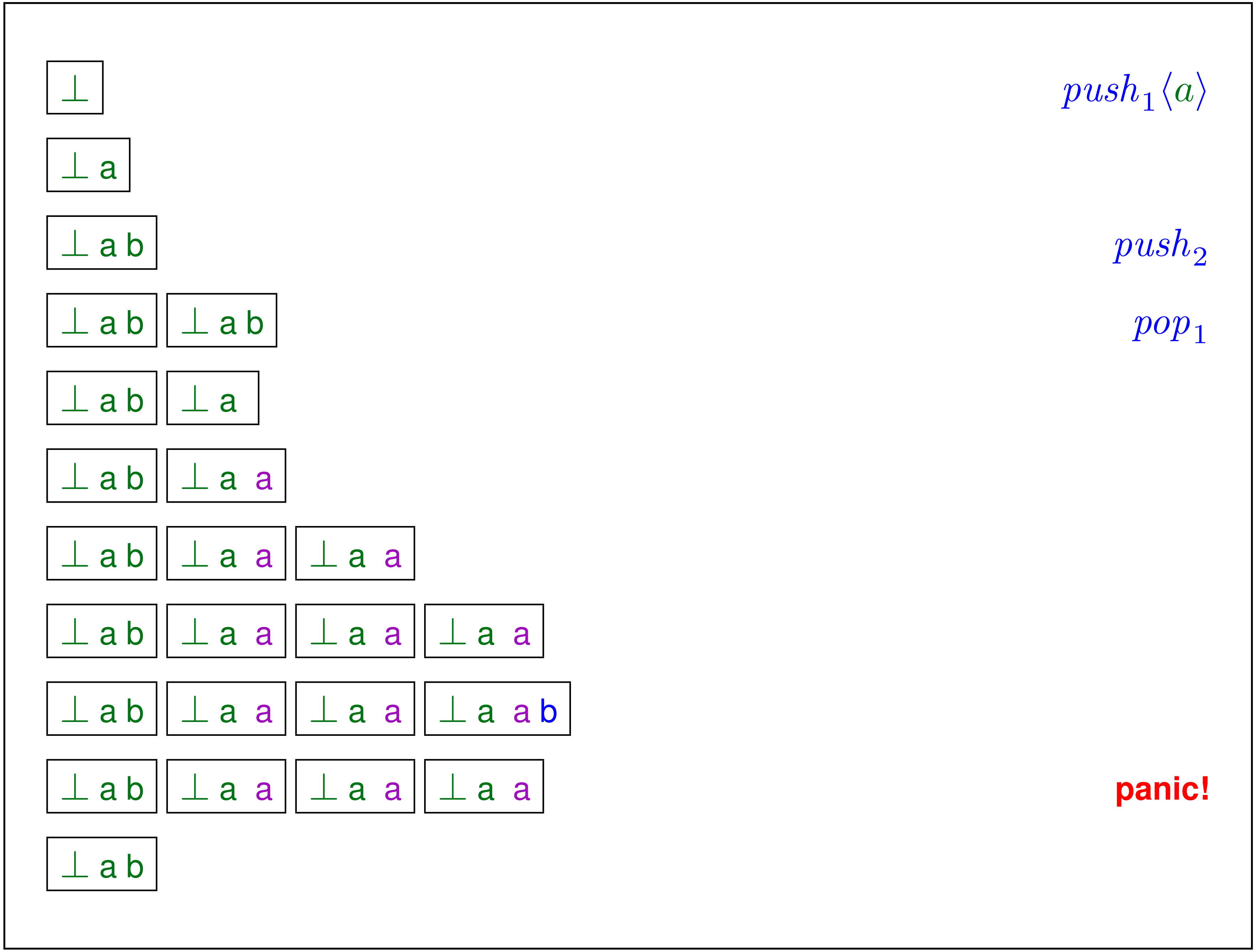
Second-order pushdown stores with time stamps

A *level* 1 *pushdown store* is a non-empty word $a_1 \dots a_k$ over $\Gamma \times \omega$.

A *level* 2 *pds* is a non-empty sequence of 1-pds' $[s_1][s_2] \dots [s_l]$.

Operations (Op_2) :

$$
\begin{aligned}
push_1\langle a\rangle([s_1][s_2]\dots[s_l][w]) &= [s_1][s_2]\dots[s_l][w(a,l)] \\
pop_1(\alpha[w\xi]) &= \alpha[w] \\
push_2(\alpha[w]) &= \alpha[w][w] \\
pop_2(\alpha[v][w]) &= \alpha[v] \\
panic([s_1][s_2]\dots[s_m]\dots[s_l][w(a,m)]) &= [s_1][s_2]\dots[s_m]
\end{aligned}
$$


⊥
⊥ a
⊥ a b
⊥ a b ⊥ a b
⊥ a b ⊥ a
⊥ a b ⊥ a a
⊥ a b ⊥ a a ⊥ a a
⊥ a b ⊥ a a ⊥ a a ⊥ a a
⊥ a b ⊥ a a ⊥ a a ⊥ a a b
⊥ a b ⊥ a a ⊥ a a ⊥ a a
⊥ a b
$push_1\langle a\rangle$
$push_2$
pop_1
panic!

The model checking problem for level 2.

Given a grammar $\mathcal{G}$ and a formula φ, decide if $[\![\mathcal{G}]\!] \models \varphi$.

Reduces to:

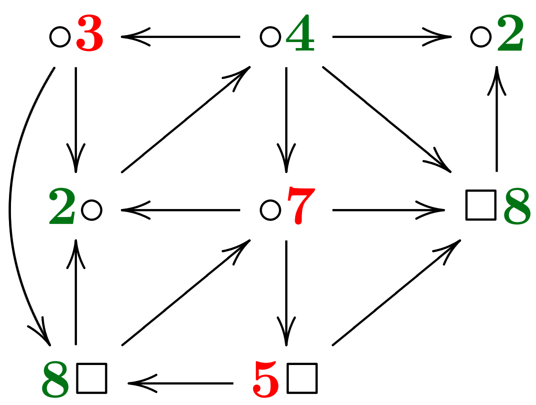
Given a second-order pushdown system with panic $\mathcal{C}$, and a parity tree automaton $\mathcal{A}$, decide if $\mathcal{A}$ accepts the tree $[\![\mathcal{C}]\!]$.

Reduces to:

Given a second-order pushdown systems with panic $\mathcal{C}$, and a parity tree automaton $\mathcal{A}$, decide if Eve wins a certain **parity game** $Game(\mathcal{C} \times \mathcal{A})$.

Parity games

Eve (○) and Adam (□) move a token on a graph.

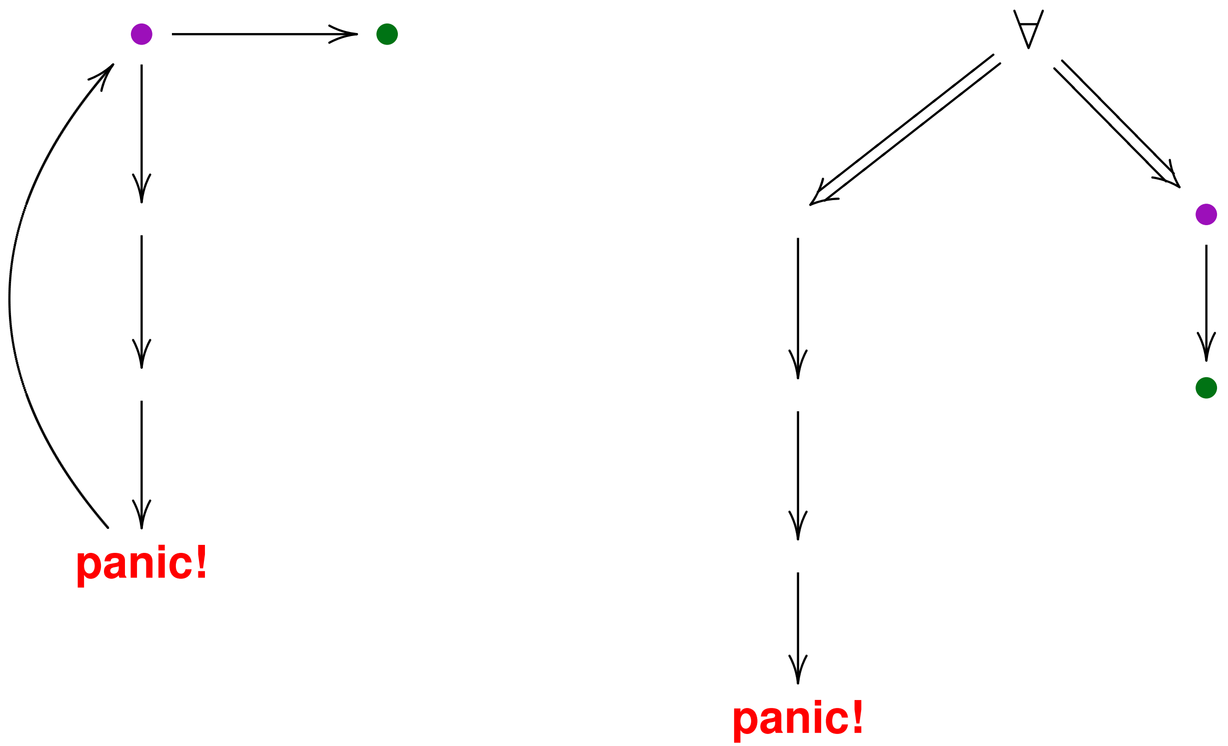


Eve wants to visit **even** priorities infinitely often.

Adam wants to visit **odd** priorities infinitely often.

Maximal priority wins.

Reduction of types is implemented by the structure of the game.



But is **safety** a true restriction ?

Example — panic not needed

Recognize words of the form $w{*}^{n+1}$, where:

– w is a prefix of a correctly parenthesized expression;

– $n = |w|$.

Words like this one: $[\,[\,[\,]\,]\,[\,[\,]\,{*}{*}{*}{*}{*}{*}{*}{*}{*}$

Not a context-free language.

Example (Urzyczyn) — panic seems to be needed

Recognize words of the form $uv *^{n+1}$, where:

– u is a prefix of a correctly parenthesized expression ending with $[$;

– v is a correctly parenthesized expression;

– $n = |u|$.

Words like this one:

$[\ [\ [\ ]\ ]\ [\ [\ ]\ [\ [\ ]\ ]\ \ *\ *\ *\ *\ *\ *\ *$

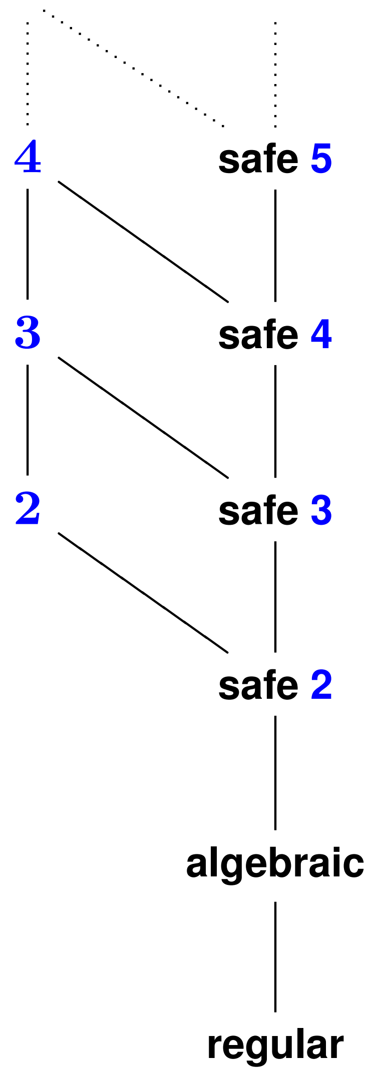
The example is related to the following grammar (**Urzyczyn**).

$$\begin{aligned} S &\Rightarrow D\varphi ab \\ D\varphi xy &\Rightarrow (f\,D(D\varphi x)y\overline{y})\,(f\,(\varphi y)x) \end{aligned}$$

Parys (2011, 2012) proved that the above language U **cannot** be recognized by a deterministic automaton without panic of any level.

The level hierarchy of collapsible pushdown automata is strict **Parys & Kartzow 2012**.

Hierarchy of trees with decidable MSO theories



Questions

Is there a **Caucal–like** hierarchy of unsafe trees ?

Does safety admit some decidable characterization ?

Are there other reasons for decidability (e.g., low entropy) ?