

Visibly pushdown automata on trees: universality and u -universality

Véronique Bruyère Marc Ducobu Olivier Gauwin

Abstract

An automaton is universal if it accepts every possible input. We study the notion of u -universality, which asserts that the automaton accepts every input starting with u . Universality and u -universality are both EXPTIME-hard for non-deterministic tree automata. We propose efficient antichain-based techniques to address these problems for visibly pushdown automata operating on trees. One of our approaches yields algorithms for the universality and u -universality of hedge automata.

1 Introduction

The model-checking framework provided many successful tools for decades, starting from the seminal work of Büchi. A lot of them rely on the links between logics used to express properties on words, and automata allowing to check them. Some of these results have been adapted to trees, and more recently to words with a nesting structure.

Visibly pushdown automata (VPAs) have been introduced to process such words with nesting [AM04]. VPAs are similar to pushdown automata, but operate on a partitioned alphabet: a given letter is associated with one action (push or pop), and thus cannot push when firing a transition, and pop when firing another. Such automata were introduced to express and check properties on control flows of programs, where procedure calls push on the stack, and returns pop. They are also suitable to express properties on XML documents [KMV07]. These are usually represented as trees, but are serialized as a sequence of opening and closing tags, also called the *linearization* of this document, or its corresponding *XML stream*.

Processing such streams without building the corresponding tree is permitted by online algorithms. It is often crucial to detect *at the earliest position* of the stream whether it satisfies a given property or not. When the property is given by an automaton, we call this automaton *u -universal* when the stream begins with word u , and u ensures that the whole stream is accepted by the automaton, whatever it contains after u . Indeed, this is a variant of universality of automata: universality is ϵ -universality, and amounts to assert that the property will be true for every possible stream, and thus can be asserted before reading the first letter. While universality of automata is a very strong property, u -universality

arises each time an automaton checks the *presence of a pattern* in trees, and this pattern appears in u .

A delay in the detection of a violation may be exploited to break firewalling systems when they use XML for logs [BJLW08], or to perform a denial of service attack on a remote program. In a less critical sense, it can also be used in XML validators, to assert validation or non-validation of a document before reading it entirely. For program traces, this is usually addressed by online verification algorithms operating on words but without considering the nesting relation between program calls and returns [KV01]. In the XML setting, some streaming algorithms have been proposed. Most of them are not earliest, and require a delay between the position where acceptance/refusal can be decided, and the position where it is claimed.

Indeed, testing u -universality is computationally hard on linearizations of trees. When the property is specified by a deterministic automaton, this can be checked in cubic time. On non-deterministic automata, u -universality becomes EXPTIME-complete [GNT09]. Non-determinism naturally arises when automata are obtained from logic formulas, as for instance XPath expressions with descendant axis [FDL11, GN11].

In this paper we propose new algorithms for deciding universality and u -universality of non-deterministic tree automata on unranked trees accessed through their linearization. Our goal is to obtain algorithms that outperform the usual approach consisting in determinizing the automaton. We want our algorithms for u -universality to be incremental, in that, for a letter a , deciding the ua -universality should reuse as much information from u -universality computation as possible. Indeed we want to find the earliest position allowing to assert acceptance, so we have to test u -universality for every prefix u before that point.

We use *antichains* to get smaller objects to manipulate, and develop other ad-hoc methods. Antichains have been applied recently to decision problems related to non-deterministic automata: universality and inclusion for finite word automata [DWDHR06], and for non-deterministic bottom-up tree automata [BHH⁺08]. Some simulation relations are also known on unranked trees [Srb06] but it is unclear whether they can help for our problems, as they do in other contexts [ACH⁺10, DR10]. Nguyen [Ngu09] proposed an algorithm for testing the universality of VPAs. This algorithm simultaneously performs an on-the-fly determinization and reachability checking by \mathcal{P} -automaton. The notion of \mathcal{P} -automaton introduced in [EHRS00, EKS03] provides a symbolic technique to compute the sets of all reachable configurations of a VPA. This algorithm has been later improved by Nguyen and Ohsaki [NO12] by introducing antichains of over transitions of \mathcal{P} -automaton, in a way to generate reachable configurations as small as possible. Our algorithms for universality are alternative to this one since we do not use the regularity property of the set of reachable configurations. And our techniques for incrementally testing u -universality are totally new wrt this algorithm. A problem similar to u -universality is addressed in [MV09] in the context of query answering. Their algorithm applies to non-deterministic VPAs recognizing a canonical language of a query, but the automata are assumed to only accept prefixes u for which u -universality holds, which is precisely the goal

of our algorithms.

We contribute two algorithms for checking u -universality of VPAs on linearizations of unranked trees. The first algorithm is by reduction to u -universality (and also universality) of hedge automata. *Hedge automata* are the standard automaton model used for unranked trees [BKMW01], and runs in a bottom-up manner. Hedge automata are similar to XML schema models like DTDs or Relax NG. The second algorithm is a direct algorithm on VPAs. Such an algorithm was known in the deterministic case [GNT09], and relied on the incremental computation of safe states. This algorithm cannot be generalized to the non-deterministic case, as sets of safe states do not contain enough information. Instead, we use sets of safe configurations, which may be infinite, but manipulated through finite antichains. We show how SAT solvers can be used to update these antichains.

The paper is structured as follows. In Section 2 we define trees, visibly push-down automata and the problem of u -universality. Section 3 details our first algorithm, relying on a translation to hedge automata. Section 4 contains our second algorithm, namely the incremental computation of sets of safe configurations.

2 Trees, Automata and u -universality

2.1 Unranked Trees

We recall here the standard definition of unranked trees, as provided for instance in [CDG⁺07]. Let Σ be a finite *alphabet*, and Σ^* (resp. Σ^+) be the set of all words (resp. non empty words) over Σ . The empty word is denoted by ϵ . Given two words $v, w \in \Sigma^*$ over Σ , v is a *prefix* (resp. *proper prefix*) of w if there exists a word $v' \in \Sigma^*$ (resp. $v' \in \Sigma^+$) such that $vv' = w$. Let \mathbb{N}_0 be the set of all non-negative integers.

An *unranked tree* t over Σ is a partial function $t : \mathbb{N}_0^* \rightarrow \Sigma$ such that the domain is non-empty, finite and prefix-closed. The domain is denoted by $nodes(t)$ and contains the *nodes* of the tree t , with the root being the empty word ϵ . The function t labels each node p with a letter $t(p)$ of Σ . A node labeled by $a \in \Sigma$ is called an *a-node*. The set of all unranked trees over Σ is denoted by T_Σ .

The *subtree* of t rooted at node p of t is the tree denoted by $t|_p$, which domain is the set of nodes p' such that $pp' \in nodes(t)$ and verifying $t|_p(p') = t(pp')$. For a given node $p \in nodes(t)$, we call *children* of p the nodes $pi \in nodes(t)$ for $i \in \mathbb{N}_0$, and use the usual definitions for parents, ancestors and descendants. The *height* of a tree is the length of its longest branch (with the length being the number of nodes).

Example 1. Let $t_1 : \{\epsilon, 1, 2, 3, 4, 5, 51, 52\} \rightarrow \{a, b, c\}$ such that $t_1(\epsilon) = c$, $t_1(1) = a$, $t_1(2) = a$, $t_1(3) = a$, $t_1(4) = a$, $t_1(5) = b$, $t_1(51) = b$, $t_1(52) = b$. Tree t_1 is an unranked tree with height 3. It can be represented as in Figure 1.

Another example is $t_2 : \{\epsilon, 1, 11, 12, 121, 122, 2, 3, 31, 32, 33, 34\} \rightarrow \{a, b, c\}$ as illustrated in Figure 2.

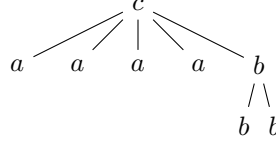


Figure 1: Representation of unranked tree t_1 .

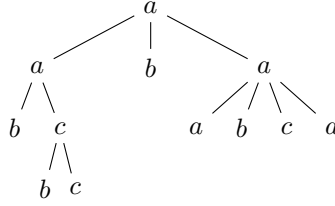


Figure 2: Representation of t_2 .

Linearization Trees can be described by well-balanced words which correspond to a depth-first traversal of the tree. An opening tag is used to notice the arrival on a node and a closing tag to notice the departure of a node. For each $a \in \Sigma$, let a itself represent the opening tag and \bar{a} the related closing tag. The *linearization* $[t]$ of $t \in T_\Sigma$ is the *well-balanced* word over $\Sigma \cup \bar{\Sigma}$, with $\bar{\Sigma} = \{\bar{a} \mid a \in \Sigma\}$, inductively defined by:

$$[t] = a [t_1] \cdots [t_n] \bar{a}$$

with $a = t(\epsilon)$ and the root has n children. We denote by $[T_\Sigma]$ the set of linearizations of all trees in T_Σ . Let $PPref(T_\Sigma)$ denote the set of all proper prefixes of $[T_\Sigma]$: $PPref(T_\Sigma) = \{u \in (\Sigma \cup \bar{\Sigma})^* \mid \exists v \in (\Sigma \cup \bar{\Sigma})^+, uv \in [T_\Sigma]\}$.

Example 2. Let t_1 and t_2 be the trees defined in Example 1, then

$$\begin{aligned} [t_1] &= c a \bar{a} a \bar{a} a \bar{a} a \bar{a} b b \bar{b} b \bar{b} \bar{b} \bar{c} \\ [t_2] &= a a b \bar{b} c b \bar{b} c \bar{c} \bar{c} \bar{a} b \bar{b} a a \bar{a} b \bar{b} c \bar{c} a \bar{a} \bar{a} \end{aligned}$$

2.2 Visibly pushdown automata

Visibly pushdown automata (VPAs, [AM04, AM09]) are pushdown automata operating on a partitioned alphabet where only call symbols can push, return symbols can pop, and internal symbols can do transitions without considering the stack.

In this paper we only consider languages of unranked trees, so we use VPAs as unranked trees acceptors, operating on their linearization (also named *streaming tree automata* [GNR08]). This corresponds to the following restrictions. First, the alphabet is only partitioned into call symbols Σ and return symbols $\bar{\Sigma}$, and does not contain internal symbols. Second, all linearizations recognized by these VPAs are such that all pairs of matched call a and return \bar{b} are such that $a = b$,

corresponding to the label of the tree of the corresponding node. Third, all linearizations are well-matched and single-rooted, so the acceptance condition is that a final state is reached on empty stack.

Definition 3. A visibly pushdown automaton \mathcal{A} over alphabet Σ is a tuple $\mathcal{A} = (Q, \Sigma, \Gamma, Q_i, Q_f, \Delta)$ where Q is a finite set of states containing initial states $Q_i \subseteq Q$ and final states $Q_f \subseteq Q$, a finite set Γ of stack symbols, and a finite set Δ of rules. Each rule in Δ is of the form $q \xrightarrow{a:\gamma} q'$ with $a \in \Sigma \cup \overline{\Sigma}$, $q, q' \in Q$, and $\gamma \in \Gamma$.

The left-hand side of a rule $q \xrightarrow{a:\gamma} p \in \Delta$ is (q, a) if $a \in \Sigma$, and (q, a, γ) if $a \in \overline{\Sigma}$. A VPA is *deterministic* if it has at most one initial state, and it does not have two distinct rules with the same left-hand side.

A *configuration* of a VPA \mathcal{A} is a pair (q, σ) where $q \in Q$ is a state and $\sigma \in \Gamma^*$ a stack content. A configuration is *initial* (resp. *final*) if $q \in Q_i$ (resp. $q \in Q_f$) and $\sigma = \epsilon$. For $a \in \Sigma \cup \overline{\Sigma}$, we write $(q, \sigma) \xrightarrow{a} (q', \sigma')$ if there is a transition $q \xrightarrow{a:\gamma} q'$ in Δ verifying $\sigma' = \gamma \cdot \sigma$ if $a \in \Sigma$, and $\sigma = \gamma \cdot \sigma'$ if $a \in \overline{\Sigma}$. We extend this notation to words, by writing $(q_0, \sigma_0) \xrightarrow{a_1 \cdots a_n} (q_n, \sigma_n)$ whenever there exist configurations (q_i, σ_i) such that $(q_{i-1}, \sigma_{i-1}) \xrightarrow{a_i} (q_i, \sigma_i)$ for all $1 \leq i \leq n$. From $u \in (\Sigma \cup \overline{\Sigma})^*$ and the set of configurations $\mathcal{C} \subseteq Q \times \Gamma^*$, we also define $Post_u(\mathcal{C})$ as the set of configurations (q', σ') for which there exists a configuration $(q, \sigma) \in \mathcal{C}$ such that $(q, \sigma) \xrightarrow{u} (q', \sigma')$.

A *run* of a VPA \mathcal{A} on a linearization $[t] = a_1 \cdots a_n$ of $t \in T_\Sigma$ is a sequence $(q_0, \sigma_0) \cdots (q_n, \sigma_n)$ of configurations (q_i, σ_i) such that (q_0, σ_0) is initial, and for every $1 \leq i \leq n$, $(q_{i-1}, \sigma_{i-1}) \xrightarrow{a_i} (q_i, \sigma_i)$. Such a run is *accepting* if (q_n, σ_n) is final. A tree $t \in T_\Sigma$ is *accepted* by \mathcal{A} if there is an accepting run on its linearization $[t]$. The set of accepted trees is called the *language* of \mathcal{A} and is written $L(\mathcal{A})$.

2.3 Universality and u -universality

We conclude the preliminaries with the notions of universality and u -universality, that we will study in the remainder of the paper.

Definition 4. A tree automaton \mathcal{A} over Σ is said *universal* if \mathcal{A} accepts all trees $t \in T_\Sigma$. Let $u \neq \epsilon$ be a prefix of $[t_0]$ for some tree $t_0 \in T_\Sigma$. The tree automaton \mathcal{A} is said *u -universal* if for all trees $t \in T_\Sigma$, if u is a prefix of $[t]$, then t is accepted by \mathcal{A} .

In other words, u -universality allows to assert that any tree linearization beginning with u is accepted by the automaton. The two previous definitions does not depend on the tree automaton \mathcal{A} but only on the language $L(\mathcal{A})$. Therefore they are independent on the kind of tree automata that are used, as soon as they are equivalent.

Our objective is to propose *incremental* algorithms for u -universality, in the following sense. The linearization $[t_0]$ of a given tree t_0 is read letter by letter, and while \mathcal{A} is not u -universal for the current read prefix u of $[t_0]$, the

next letter of $[t_0]$ is read. For instance Algorithm 1 shows how u -universality is checked incrementally. When processing a new letter, we try to reuse prior computations as much as possible. The automaton can be supposed to be not universal, otherwise it is u -universal for all such words u .

Algorithm 1 Checking u -universality incrementally

```

function INCREMENTAL- $u$ -UNIVERSALITY( $\mathcal{A}, w$ )
   $i \leftarrow 1$ 
  while  $i \leq |w|$  do
    if  $\mathcal{A}$  is  $w_1 \cdots w_i$ -universal then
      return True
    end if
     $i \leftarrow i + 1$ 
  end while
  return False
end function

```

It has been shown in [GNT09] that u -universality is EXPTIME-complete for VPAs, but in PTIME for deterministic VPAs. Determinization is in exponential time for VPAs, and our algorithms aim at avoiding this exponential blowup.

An incremental u -universality check as described in Algorithm 1 is very useful. First, given a tree t_0 , it allows a streaming membership test of t_0 in \mathcal{A} : its linearization $[t_0]$ is read letter by letter, and the algorithm declares as soon as possible whether t_0 is accepted by \mathcal{A} . Second, when a property (of XML documents for instance) is given by a tree automaton, then Algorithm 1 detects at the earliest position of $[t_0]$ whether t_0 satisfies the property.

3 Hedge automata approach

We present algorithms for testing universality and u -universality of a non deterministic visibly pushdown automaton. The approach followed in this section is based on a translation of the VPA into an hedge automaton. Algorithms with several optimizations are then provided for checking universality and u -universality of hedge automata.

3.1 Hedge automata

We present the standard notion of hedge automata [BKMW01, CDG⁺07], the usual automaton model for expressing properties on XML documents. Indeed, a hedge automaton resembles a DTD: a DTD is a set of rules like $a \rightarrow b^+c$ saying that children of an a -node must be a non empty sequence of b -nodes followed by a c -node. Hedge automata are a bit more expressive than DTDs, in that regular languages operate on states instead of labels, enabling for instance to distinguish two kinds of a -nodes.

A *hedge* h over a finite alphabet Σ is a sequence (empty or not) of unranked trees over Σ . The set of all hedges over Σ is denoted by H_Σ . For instance, given the trees t_1 and t_2 from Example 1, the sequence $t_1 t_2 t_1$ is a hedge.

Definition 5. A hedge automaton over Σ is a tuple $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ where Q is a finite set of states, $Q_f \subseteq Q$ is the set of final states, and Δ is a finite set of transition rules of the following type:

$$(a, L, q)$$

where $a \in \Sigma$, $q \in Q$, and $L \subseteq Q^*$ is a regular language over Q , called a horizontal language.

We denote by $\mathcal{H}_\mathcal{A}$ the set of all horizontal languages of \mathcal{A} . Note that for every $a \in \Sigma$ and $q \in Q$, we can assume that there is only one L such that $(a, L, q) \in \Delta$. Indeed, we can replace all rules (a, L', q) by one rule (a, L, q) where L is the union of all such L' . A hedge automaton is *deterministic* if for all pairs of rules (a, L_1, q_1) and (a, L_2, q_2) we have $L_1 \cap L_2 = \emptyset$ or $q_1 = q_2$.

A *run* of \mathcal{A} on a tree $t \in T_\Sigma$ is a tree $r \in T_Q$ with the same domain as t such that for each node $p \in \text{nodes}(r)$ and its n children $p1, p2, \dots, pn$, if $a = t(p)$ and $q = r(p)$, then there is a rule $(a, L, q) \in \Delta$ with $r(p1)r(p2) \dots r(pn) \in L$. In particular, to apply the rule (a, L, q) at a leaf, the empty word ϵ has to belong to L . Intuitively, a hedge automaton \mathcal{A} operates in a bottom-up manner on a tree t : with a run r , it assigns a state to each leaf, and then to each internal node, according to the states assigned to its children. We use notation $t \xrightarrow[\mathcal{A}]{} q$ to indicate the existence of a run r on t that labels the root of t by the state q . Such a run r is *accepting* if q is final, i.e. $r(\epsilon) \in Q_f$. An unranked tree t is *accepted* by \mathcal{A} if there exists an accepting run on it. The *language* $L(\mathcal{A})$ of \mathcal{A} is the set of all unranked trees accepted by \mathcal{A} .

Example 6. Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton over $\Sigma = \{a, b, c\}$ with $Q = \{q_a, q_b, q_c, q_f\}$, $Q_f = \{q_f\}$, and $\Delta = \{(a, L_1, q_a), (b, L_1, q_b), (c, L_1, q_c), (a, L_2, q_f), (a, L_3, q_f), (b, L_3, q_f), (c, L_3, q_f)\}$ where $L_1 = Q^*$, $L_2 = q_b q_c$ and $L_3 = Q^* q_f Q^*$.

Let t_1 and t_2 the trees from Example 1. Figure 3 represents a run r_1 of \mathcal{A} on t_1 and two runs, r_2 and r_3 , of \mathcal{A} on t_2 . The runs r_1 and r_2 are not accepting, whereas r_3 is accepting. The tree t_1 is not accepted by \mathcal{A} , whereas t_2 is accepted by \mathcal{A} . The language of \mathcal{A} is the set of all trees having a subtree s which root is an a -node and has two children with $s(1) = b$ and $s(2) = c$.

3.2 From VPAs to hedge automata

In this section, we describe a translation of VPAs into hedge automata, with the aim to transfer universality and u -universality testing of a VPA to a hedge automaton.

Theorem 7. Let \mathcal{A} be a VPA. Then one can construct a hedge automaton \mathcal{A}_H such that for all $t \in T_\Sigma$, $[t] \in L(\mathcal{A})$ iff $t \in L(\mathcal{A}_H)$.

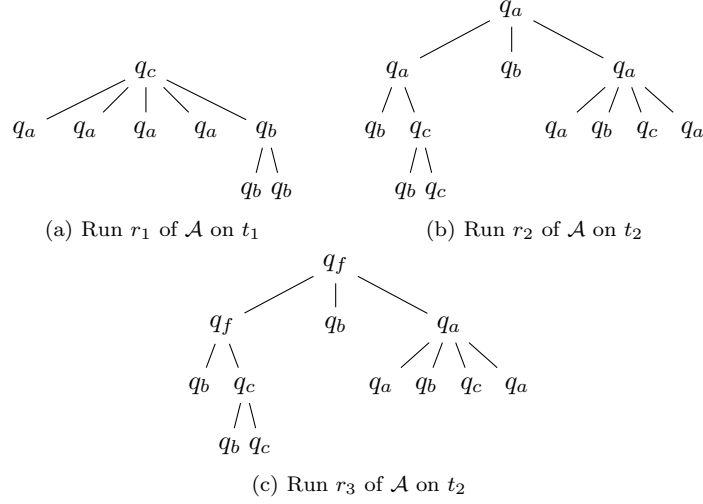


Figure 3: Examples of runs

Proof. Let $\mathcal{A} = (Q, \Sigma, \Gamma, Q_i, Q_f, \Delta)$ be a VPA. We define the hedge automaton $\mathcal{A}_H = (Q', \Sigma, Q'_f, \Delta')$ such that

- $Q' = Q \times Q$
- $Q'_f = Q_i \times Q_f$
- $\Delta' = \{(a, L_{s,s'}, (q, q')) \mid \exists \gamma \in \Gamma, q \xrightarrow{a:\gamma} s \in \Delta \text{ and } s' \xrightarrow{\bar{a}:\gamma} q' \in \Delta\}$ where $L_{s,s'} = \{(s, q_1) \cdot (q_1, q_2) \cdots (q_{n-1}, q_n) \cdot (q_n, s') \mid n \geq 0, s, q_1, \dots, q_n, s' \in Q\} \cup K_{s,s'}$, and $K_{s,s'} = \emptyset$ if $s \neq s'$, $K_{s,s'} = \{\epsilon\}$ otherwise.

Notice that each language $L_{s,s'}$ is regular. Let us prove for all $t \in T_\Sigma$ and $q, q' \in Q$ that:

$$(q, \epsilon) \xrightarrow{[t]} (q', \epsilon) \iff t \xrightarrow{\mathcal{A}_H} (q, q')$$

As a consequence, we will have $[t] \in L(\mathcal{A})$ iff $t \in L(\mathcal{A}_H)$.

We proceed by induction on the height of t . We begin with the basic case $\text{height}(t) = 1$, i.e. t be a a -leaf for some $a \in \Sigma$. Then $t \xrightarrow{\mathcal{A}_H} (q, q')$ iff $\exists s \in Q, \gamma \in \Gamma$ such that $q \xrightarrow{a:\gamma} s \in \Delta$ and $s \xrightarrow{\bar{a}:\gamma} q' \in \Delta$ (recall that $\epsilon \in L_{s,s}$). This is equivalent to $(q, \epsilon) \xrightarrow{[t]} (q', \epsilon)$.

Let $i > 1$ and suppose that the property holds for all trees of height less than i . Let t be a tree of height i such that $a = t(\epsilon)$ and the root has n children.

Let r be a run of \mathcal{A}_H on t such that $(q, q') = r(\epsilon)$. Then, by definition of \mathcal{A}_H , there exist $q_1, \dots, q_{n+1} \in Q$ and $\gamma \in \Gamma$ such that $r(1) = (q_1, q_2)$, $r(2) = (q_2, q_3), \dots, r(n) = (q_n, q_{n+1})$, $q \xrightarrow{a:\gamma} q_0 \in \Delta$ and $q_{n+1} \xrightarrow{\bar{a}:\gamma} q' \in \Delta$. We know by

induction hypothesis that $(q_i, \epsilon) \xrightarrow{[t_i]} (q_{i+1}, \epsilon)$ for all $1 \leq i \leq n$. It follows that $(q_0, \epsilon) \xrightarrow{[h]} (q_{n+1}, \epsilon)$ where $h = t_1 t_2 \dots t_n$. We have also $(q_0, \gamma) \xrightarrow{[h]} (q_{n+1}, \gamma)$ since h is an edge, and thus $(q, \epsilon) \xrightarrow{[t]=a[h]\bar{a}} (q', \epsilon)$.

Suppose now that $(q, \epsilon) \xrightarrow{[t]} (q', \epsilon)$. So there exist $q_1, \dots, q_{n+1} \in Q$ and $\gamma \in \Gamma$ such that $q \xrightarrow{a:\gamma} q_1 \in \Delta$, $q_{n+1} \xrightarrow{\bar{a}:\gamma} q' \in \Delta$, and $(q_i, \gamma) \xrightarrow{[t_i]} (q_{i+1}, \gamma)$ for all i . By induction hypothesis, $t_i \xrightarrow{\mathcal{A}_H} (q_i, q_{i+1})$ for all i , and thus $t \xrightarrow{\mathcal{A}_H} (q, q')$. \square

As a consequence of Theorem 7, universality and u -universality testing of a VPA \mathcal{A} is transferred to the hedge automaton \mathcal{A}_H .

3.3 Checking universality

A standard method to check universality of a hedge automaton is to determinize it, complement it, and check for emptiness. As determinization is in exponential time [CDG⁺07], we propose in this section an antichain-based algorithm for checking universality without explicit determinization.

Such an algorithm has been proposed in [BHH⁺08] for finite (ranked) tree automata. In the context of hedge automata, additional difficulties have to be solved due to the fact that the accepted trees are unranked.

In our approach, the main idea is to find as fast as possible one tree rejected by the hedge automaton (if it exists) by performing a kind of bottom-up implicit determinization. Antichains will limit the computations.

3.3.1 Macrostates and *Post* operator

To test universality of a hedge automaton \mathcal{A} , we have to check that all the trees of T_Σ belong to $L(\mathcal{A})$. Instead of working with trees we work with sets of states, which are called *macrostates*. A macrostate is associated with each tree t : it is the set of all the states q labeling the root of a run of \mathcal{A} on t , i.e. such that $t \xrightarrow{\mathcal{A}} q$. To compute the macrostates, we make bottom-up computations by applying a *Post* operator defined as follows.

Definition 8. Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton. A macrostate is a set of states $P \subseteq Q$. A macrostate word $\pi = P_1 P_2 \dots P_n$, $n \geq 0$, is a word over the alphabet 2^Q . We denote by $\bar{\pi}$ the set $\{p_1 p_2 \dots p_n \mid p_i \in P_i, \forall i, 1 \leq i \leq n\}$. Given $a \in \Sigma$ and π a macrostate word, let

$$Post_a(\pi) = \{q \in Q \mid \exists (a, L, q) \in \Delta : L \cap \bar{\pi} \neq \emptyset\}$$

For $\mathcal{P} \subseteq 2^Q$ a set of macrostates, let

$$Post(\mathcal{P}) = \{Post_a(\pi) \mid a \in \Sigma, \pi \in \mathcal{P}^*\} \cup \mathcal{P}$$

and $Post^*(\mathcal{P}) = \bigcup_{i \geq 0} Post^i(\mathcal{P})$ such that $Post^0(\mathcal{P}) = \mathcal{P}$, and for all $i > 0$, $Post^i(\mathcal{P}) = Post(Post^{i-1}(\mathcal{P}))$.

When $\pi = \epsilon$, $Post_a(\epsilon)$ is the set of all states that can be assigned to an a -leaf of a tree, with $a \in \Sigma$. If an a -node has n children to which the macrostates P_1, \dots, P_n have been assigned, then $Post_a(P_1 \cdots P_n)$ is the set of all states that can be assigned to this node. The next lemma is immediate.

Lemma 9. *Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton and $t \in T_\Sigma$ be such that its root is an a -node with n children. Let $P_i = \{q \in Q \mid t_{|i} \xrightarrow{\mathcal{A}} q\}$ for $1 \leq i \leq n$. Then*

$$Post_a(P_1 \cdots P_n) = \{q \in Q \mid t \xrightarrow{\mathcal{A}} q\}.$$

Given \mathcal{P} a set of macrostates, $Post(\mathcal{P})$ is the set of all macrostates that belong to \mathcal{P} or can be obtained via $Post_a(\pi)$ with any letter $a \in \Sigma$, and any macrostate word $\pi = P_1 P_2 \cdots P_n$ with $P_i \in \mathcal{P}, \forall i$. More precisely, we have:

Lemma 10. *Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton and $i \geq 1$. A macrostate P belongs to $Post^i(\emptyset)$ iff there exists a tree $t \in T_\Sigma$ with $height(t) \leq i$ such that $P = \{q \in Q \mid t \xrightarrow{\mathcal{A}} q\}$.*

Proof. We proceed by induction on i .

The basic case, $i = 1$, directly follows from $Post^1(\emptyset) = \{Post_a(\epsilon) \mid a \in \Sigma\}$ and $Post_a(\epsilon) = \{q \mid t \xrightarrow{\mathcal{A}} q\}$ with t being an a -leaf.

Let $i > 1$ and suppose that the property holds for all $j, 1 \leq j < i$.
 (\Rightarrow) Let $P \in Post^i(\emptyset)$. If $P \in Post^{i-1}(\emptyset)$, then the property holds by induction hypothesis. Otherwise there exist $n \geq 0$, $P_1, \dots, P_n \in Post^{i-1}(\emptyset)$, and $a \in \Sigma$, such that $P = Post_a(P_1 \cdots P_n)$. By induction hypothesis, $\forall k, 1 \leq k \leq n$, $\exists t_k \in T_\Sigma$ such that $height(t_k) < i$ and $P_k = \{q \mid t_k \xrightarrow{\mathcal{A}} q\}$. Let t be the tree with the a -root and the n subtrees t_1, \dots, t_n . Then $height(t) \leq i$ and $P = \{q \mid t \xrightarrow{\mathcal{A}} q\}$ by Lemma 9.

(\Leftarrow) Let $t \in T_\Sigma$ with $height(t) \leq i$ and $P = \{q \in Q \mid t \xrightarrow{\mathcal{A}} q\}$. If $height(t) < i$, then by induction hypothesis $P \in Post^{i-1}(\emptyset) \subseteq Post^i(\emptyset)$. Otherwise let a be the label of the root of t and $t_{|1}, \dots, t_{|n}$ its n subtrees. Let $P_k = \{q \in Q \mid t_{|k} \xrightarrow{\mathcal{A}} q\}$, $1 \leq k \leq n$. As $height(t_{|k}) < i$, we have by induction hypothesis that $P_k \in Post^{i-1}(\emptyset)$. By Lemma 9, $P = Post_a(P_1 \cdots P_n)$, and thus $P \in Post^i(\emptyset)$. \square

Given a tree $t \in T_\Sigma$ we define P_t as the macrostate $P_t = \{q \in Q \mid t \xrightarrow{\mathcal{A}} q\}$. More generally, given a hedge $h = t_1 t_2 \cdots t_n \in H_\Sigma$ we denote by π_h the macrostate word $\pi_h = P_{t_1} P_{t_2} \cdots P_{t_n}$. The previous lemmas indicate that $Post^*(\emptyset) = \{P_t \mid t \in T_\Sigma\}$, and more generally that $(Post^*(\emptyset))^* = \{\pi_h \mid h \in H_\Sigma\}$.

The next proposition is an immediate consequence of Lemmas 9 and 10.

Proposition 11. *Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton. Then \mathcal{A} is universal iff $\forall P \in Post^*(\emptyset), P \cap Q_f \neq \emptyset$.*

3.3.2 Relations and universality algorithm

Our method for checking universality of a hedge automaton is to compute $Post^*(\emptyset)$ by iteratively applying the $Post$ operator. However to get $Post(\mathcal{P})$, we have to compute \mathcal{P}^* which is an infinite set of macrostate words. To circumvent this problem, we represent a macrostate word by a relation as described below, with the advantage that the set of relations is now finite.

We first introduce some notation. Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton and $\mathcal{H}_{\mathcal{A}}$ be the set of horizontal languages appearing in its transition rules. We recall that these languages are regular. Let $L \in \mathcal{H}_{\mathcal{A}}$ and \mathcal{B}_L be a (word) automaton over the alphabet Q that accepts L . Let S_L be its set of states, I_L its set of initial states, and F_L its set of final states. We denote by $\mathcal{B}_{\mathcal{A}}$ the automaton which is the disjoint union of all the automata \mathcal{B}_L with $L \in \mathcal{H}_{\mathcal{A}}$. Its set of states is denoted by $S_{\mathcal{A}} = \bigcup_{L \in \mathcal{H}_{\mathcal{A}}} S_L$. A run in $\mathcal{B}_{\mathcal{A}}$ from state $s \in S_{\mathcal{A}}$

to state $s' \in S_{\mathcal{A}}$ labeled by word $w \in Q^*$ is denoted by $s \xrightarrow{w} s'$.

Definition 12. Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton and π a macrostate word. Then $\text{rel}(\pi) \subseteq S_{\mathcal{A}} \times S_{\mathcal{A}}$ is the relation

$$\text{rel}(\pi) = \{(s, s') \mid s \xrightarrow{w} s' \text{ with } w \in \pi\}.$$

In other words, if $\pi = P_1 \cdots P_n$ with $P_i \subseteq Q$ for all i , then (s, s') belongs to $\text{rel}(\pi)$ iff there is a path in $\mathcal{B}_{\mathcal{A}}$ from s to s' that is labeled by a word $p_1 \cdots p_n \in \pi$. The notation rel is naturally extended to sets \mathcal{W} of macrostate words as $\text{rel}(\mathcal{W}) = \{\text{rel}(\pi) \mid \pi \in \mathcal{W}\}$.

Notice there are finitely many relations $r \subseteq S_{\mathcal{A}} \times S_{\mathcal{A}}$, since $S_{\mathcal{A}}$ is a finite set. If \mathcal{R} is a set of relations $r \subseteq S_{\mathcal{A}} \times S_{\mathcal{A}}$, then \mathcal{R}^* denotes the set of all relations obtained by composing relations in \mathcal{R} : $\mathcal{R}^* = \{r_1 \circ r_2 \circ \cdots \circ r_n \mid n \geq 0 \text{ and } r_i \in \mathcal{R} \text{ for all } 1 \leq i \leq n\}$. In particular \mathcal{R}^* contains the identity relation $\text{id}_{S_{\mathcal{A}}}$ over $S_{\mathcal{A}}$, obtained when $n = 0$.

Lemma 13. Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton. If \mathcal{P} a set of macrostates and \mathcal{R} a set of relations such that $\text{rel}(\mathcal{P}) = \mathcal{R}$, then $\text{rel}(\mathcal{P}^*) = \mathcal{R}^*$.

Proof. Let us prove that for any macrostate word $\pi = P_1 \cdots P_n$, $\text{rel}(\pi) = \text{rel}(P_1) \circ \cdots \circ \text{rel}(P_n)$; the lemma is an immediate consequence.

Let $(s, s') \in \text{rel}(P_1 \cdots P_n)$, that is, $\exists w = p_1 \cdots p_n \in \pi : s \xrightarrow{w} s'$. Let $s = s_1, s_2, \dots, s_n, s_{n+1} = s' \in S_{\mathcal{A}}$ be such that $s_i \xrightarrow{p_i} s_{i+1}$ for all i . As $p_i \in P_i$ and $(s_i, s_{i+1}) \in \text{rel}(P_i)$, it follows that $(s, s') \in \text{rel}(P_1) \circ \cdots \circ \text{rel}(P_n)$.

Conversely, let $(s, s') \in \text{rel}(P_1) \circ \cdots \circ \text{rel}(P_n)$. Let $s = s_1, s_2, \dots, s_n, s_{n+1} = s' \in S_{\mathcal{A}}$ be such that $(s_i, s_{i+1}) \in \text{rel}(P_i)$ for all i . By definition, for all i , there exists $p_i \in P_i$ such that $s_i \xrightarrow{p_i} s_{i+1}$. So for $w = p_1 \cdots p_n$, we have $s_1 \xrightarrow{w} s_{n+1}$ showing that $(s, s') \in \text{rel}(P_1 \cdots P_n)$. \square

The $Post$ operator is adapted to relations in the following way.

Definition 14. Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton, $r \subseteq S_{\mathcal{A}} \times S_{\mathcal{A}}$ a relation, and $a \in \Sigma$ a letter. Then

$$Post_a(r) = \{q \in Q \mid \exists(a, L, q) \in \Delta, \exists(s, s') \in r : s \in I_L \text{ and } s' \in F_L\}.$$

Lemma 15. Let $a \in \Sigma$ and π be a macrostate word, then $Post_a(\pi) = Post_a(\text{rel}(\pi))$.

Proof. For $a \in \Sigma$ and π a macrostate word, we have

$$\begin{aligned} Post_a(\pi) &= \{q \in Q \mid \exists(a, L, q) \in \Delta : L \cap \bar{\pi} \neq \emptyset\} \\ &= \{q \in Q \mid \exists(a, L, q) \in \Delta, \exists s, s' \in S_{\mathcal{A}}, \exists w \in \bar{\pi} : s \xrightarrow{w} s', s \in I_L \text{ and } s' \in F_L\} \\ &= \{q \in Q \mid \exists(a, L, q) \in \Delta, \exists(s, s') \in \text{rel}(\pi) : s \in I_L \text{ and } s' \in F_L\} \\ &= Post_a(\text{rel}(\pi)). \end{aligned}$$

□

Lemma 16. Let \mathcal{P} be a set of macrostates, then $Post(\mathcal{P}) = \{Post_a(r) \mid a \in \Sigma, r \in \text{rel}(\mathcal{P})^*\} \cup \mathcal{P}$.

Proof. By definition, $Post(\mathcal{P}) = \{Post_a(\pi) \mid a \in \Sigma, \pi \in \mathcal{P}^*\}$. By Lemma 15, this set is equal to $\{Post_a(\text{rel}(\pi)) \mid a \in \Sigma, \pi \in \mathcal{P}^*\}$ which is equal to $\{Post_a(r) \mid a \in \Sigma, r \in \text{rel}(\mathcal{P})^*\}$ by Lemma 13. □

We are now able to propose an algorithm to check universality of hedge automata. With Algorithm 2, the set $Post^*(\emptyset)$ is computed incrementally and the universality test is performed thanks to Proposition 11. More precisely, at step i , variable \mathcal{P} is used for $Post^i(\emptyset)$ and variable \mathcal{R}^* is used for $\text{rel}(\mathcal{P})^*$. We compute \mathcal{R}^* with Function COMPOSITIONCLOSURE, and then possible new macrostates with $\{Post_a(r) \mid a \in \Sigma, r \in \mathcal{R}^*\}$. The algorithm stops when no new macrostate is found or the hedge automaton is declared not universal.

Let us detail Function COMPOSITIONCLOSURE($\mathcal{R}^*, \mathcal{R}'$) which computes the set $(\mathcal{R}^* \cup \mathcal{R}')^*$. In Algorithm 3, we show how to compute $(\mathcal{R}^* \cup \mathcal{R}')^*$ given the inputs \mathcal{R}^* and \mathcal{R}' , without recomputing \mathcal{R}^* from \mathcal{R} . Initially, *Relations* is equal to $\mathcal{R}^* \cup \mathcal{R}'$ and will be equal to $(\mathcal{R}^* \cup \mathcal{R}')^*$ at the end of the computation. *ToProcess* contains the relations that can produce new relations by composition with an element of *Relations*.

Proposition 17. Given \mathcal{R}^* and \mathcal{R}' , Algorithm 3 computes $(\mathcal{R}^* \cup \mathcal{R}')^*$.

Proof. Let *Relations* be the set computed by Algorithm 3. Clearly, $\text{Relations} \subseteq (\mathcal{R}^* \cup \mathcal{R}')^*$. Assume by contradiction there exists r that belongs to $(\mathcal{R}^* \cup \mathcal{R}')^* \setminus \text{Relations}$. Then $r \notin \mathcal{R}^* \cup \mathcal{R}'$ and we can suppose wlog that $r = r'_2 \circ r'_1$ with $r'_1, r'_2 \in \text{Relations}$. Notice that at least one element among r'_1, r'_2 has been added to *ToProcess* during the execution of Algorithm 3, since otherwise $r'_1, r'_2 \in \mathcal{R}^*$ and then $r \in \mathcal{R}^*$. If r'_1 is the last one (among r'_1, r'_2) to be popped from *ToProcess*, then the relation $r'_2 \circ r'_1$ is added to *NewRelations*, which leads to a contradiction. The conclusion is similar if r'_2 is the last one to be popped. □

Algorithm 2 Checking universality

```
function UNIVERSALITY( $\mathcal{A}$ )
   $\mathcal{P} \leftarrow \emptyset$ 
   $\mathcal{R}^* \leftarrow \{id_{S_A}\}$ 
  repeat
     $\mathcal{P}_{new} \leftarrow \{Post_a(r) \mid a \in \Sigma, r \in \mathcal{R}^*\}$ 
    if  $\exists P \in \mathcal{P}_{new} : P \cap F = \emptyset$  then
      return False // Not universal
    end if
     $\mathcal{R}' \leftarrow \text{rel}(\mathcal{P}_{new} \setminus \mathcal{P}) \setminus \mathcal{R}^*$ 
    if  $\mathcal{R}' \neq \emptyset$  then
       $\mathcal{P} \leftarrow \mathcal{P} \cup \mathcal{P}_{new}$ 
       $\mathcal{R}^* \leftarrow \text{COMPOSITIONCLOSURE}(\mathcal{R}^*, \mathcal{R}')$ 
    end if
  until  $\mathcal{R}' = \emptyset$ 
  return True // Universal
end function
```

Algorithm 3 Computing $(\mathcal{R}^* \cup \mathcal{R}')^*$

```
function COMPOSITIONCLOSURE( $\mathcal{R}^*, \mathcal{R}'$ )
   $Relations \leftarrow \mathcal{R}^* \cup \mathcal{R}'$ 
   $ToProcess \leftarrow \mathcal{R}'$ 
  while  $ToProcess \neq \emptyset$  do
     $rel \leftarrow \text{POP}(ToProcess)$ 
     $NewRelations \leftarrow \emptyset$ 
    for  $r \in Relations$  do
       $NewRelations \leftarrow NewRelations \cup \{r \circ rel, rel \circ r\}$ 
    end for
     $ToProcess \leftarrow ToProcess \cup (NewRelations \setminus Relations)$ 
     $Relations \leftarrow Relations \cup NewRelations$ 
  end while
  return  $Relations$ 
end function
```

3.3.3 Antichain-based optimization

In this section we explain how to use the concept of antichain for saving computations. We show that it is sufficient to only compute the \subseteq -minimal elements of $Post^*(\emptyset)$ for checking universality.

Consider the set 2^Q of all macrostates, with the \subseteq operator. An *antichain* \mathcal{P} of macrostates is a set of pairwise incomparable macrostates with respect to \subseteq . Given a set \mathcal{P} of macrostates, we denote by $\lfloor \mathcal{P} \rfloor$ the \subseteq -*minimal* elements of \mathcal{P} , similarly we denote by $\lceil \mathcal{P} \rceil$ the \subseteq -*maximal* elements of \mathcal{P} . A set \mathcal{P} of macrostates is \subseteq -*upward closed* (resp. \subseteq -*downward closed*) if for all $P \in \mathcal{P}$ and $P \subseteq P'$ (resp. $P' \subseteq P$), we have $P' \in \mathcal{P}$. The same notions can be defined for a set of relations (instead of macrostates).

Definition 18. Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton. Let $\mathcal{P} \subseteq 2^Q$ be a set of macrostates, let

$$Post_{\sqcup}(\mathcal{P}) = \lfloor Post(\mathcal{P}) \rfloor$$

and $Post_{\sqcup}^*(\mathcal{P}) = \cup_{i \geq 0} Post_{\sqcup}^i(\mathcal{P})$ such that $Post_{\sqcup}^0(\mathcal{P}) = \lfloor \mathcal{P} \rfloor$, and for all $i > 0$, $Post_{\sqcup}^i(\mathcal{P}) = Post_{\sqcup}(Post_{\sqcup}^{i-1}(\mathcal{P}))$.

Lemma 19. Given \mathcal{P} a set of macrostates, for all $P \in Post^*(\mathcal{P})$, there exists $P' \in Post_{\sqcup}^*(\mathcal{P})$ such that $P' \subset P$.

Proof. The proof is done by induction on i such that $Post^*(\mathcal{P}) = \cup_{i \geq 0} Post^i(\mathcal{P})$, and on the next two observations:

- Given $a \in \Sigma$, and r, r' two relations over $S_{\mathcal{A}}$, if $r \subseteq r'$ then $Post_a(r) \subseteq Post_a(r')$.
- Let $r_1, \dots, r_n, r'_1, \dots, r'_n$ be relations over $S_{\mathcal{A}}$, if $r_i \subseteq r'_i, \forall 1 \leq i \leq n$, then $r_1 \circ \dots \circ r_n \subseteq r'_1 \circ \dots \circ r'_n$.

□

Notice that thanks to Lemma 16, given an antichain of macrostates \mathcal{P} , we can compute $Post_{\sqcup}(\mathcal{P})$ as $\lfloor \{Post_a(r) \mid a \in \Sigma, r \in \lfloor rel(\mathcal{P})^* \rfloor\} \cup \mathcal{P} \rfloor$. We have the next counterpart of Proposition 11.

Proposition 20. Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton. \mathcal{A} is universal if and only if $\forall P \in Post_{\sqcup}^*(\emptyset), P \cap Q_f \neq \emptyset$.

Proof. The proof is based on Proposition 11.

(\Rightarrow) As $Post_{\sqcup}^*(\emptyset) \subseteq Post^*(\emptyset)$, the proof is immediate.

(\Leftarrow) Suppose that $\forall P \in Post_{\sqcup}^*(\emptyset), P \cap Q_f \neq \emptyset$. Let $P' \in Post^*(\emptyset)$. By Lemma 19, $\exists P \in Post_{\sqcup}^*(\emptyset) : P \subseteq P'$. It follows that $P' \cap Q_f \neq \emptyset$. □

Algorithm 4 checks whether a given hedge automaton is universal by computing incrementally $Post_{\sqcup}^*(\emptyset)$. It is an adaptation of Algorithm 2.

Algorithm 4 Checking universality

```
function UNIVERSALITY( $\mathcal{A}$ )  
   $\mathcal{P} \leftarrow \emptyset$   
   $\mathcal{R}_{min}^* \leftarrow \{id_{S_{\mathcal{A}}}\}$   
  repeat  
     $\mathcal{P}_{new} \leftarrow \lfloor \{Post_a(r) \mid a \in \Sigma, r \in \mathcal{R}_{min}^*\} \rfloor$   
    if  $\exists P \in \mathcal{P}_{new} : P \cap F = \emptyset$  then  
      return False // Not universal  
    end if  
     $\mathcal{R}' \leftarrow \text{rel}(\mathcal{P}_{new} \setminus \mathcal{P}) \setminus \mathcal{R}_{min}^*$   
    if  $\mathcal{R}' \neq \emptyset$  then  
       $\mathcal{P} \leftarrow \lfloor \mathcal{P} \cup \mathcal{P}_{new} \rfloor$   
       $\mathcal{R}_{min}^* \leftarrow \lfloor \text{COMPOSITIONCLOSURE}(\mathcal{R}_{min}^*, \mathcal{R}') \rfloor$   
    end if  
  until  $\mathcal{R}' = \emptyset$   
  return True // Universal  
end function
```

Notice that in Algorithm 4, to compute $\lfloor \text{rel}(\mathcal{P})^* \rfloor$, we first make a call to Function COMPOSITIONCLOSURE and then we only keep the \subseteq -minimal elements of the result. An optimisation could be, at each step of the COMPOSITIONCLOSURE computation, to only consider the minimal elements.

3.4 Checking u -universality

In this section, given \mathcal{A} a hedge automaton and $u \neq \epsilon$ a word in $PPref(T_{\Sigma})$, we propose a method to check whether \mathcal{A} is u -universal. This method is incremental, as explained in Section 2.3. As in the previous section, we first propose our approach, then transform it into an algorithm (thanks to relations), and finally propose some optimizations.

We need the following notation. Let u be the current read proper prefix of $[t_0]$ for a given tree t_0 . If $u = a_1[h_1]a_2[h_2] \cdots a_n[h_n]$ with $a_i \in \Sigma, h_i \in H_{\Sigma}$, for $1 \leq i \leq n$, then $\text{open}(u) = a_1a_2 \cdots a_n$. In other words, a_1, a_2, \dots, a_n are the read open tags which closing tags have not been read yet. The partial reading of t_0 according to u indicates a current list of ancestors respectively labeled by a_1, a_2, \dots, a_n as depicted in Figure 4.

Given u , let $w_i = a_1[h_1] \cdots a_{i-1}[h_{i-1}]$, for $1 \leq i \leq n$, such that $w_1 = \epsilon$. The incremental method is based on the usage of some sets

$$X_{w_i a_i}, \quad 1 \leq i \leq n,$$

such that each $X_{w_i a_i}$ is defined from $X_{w_{i-1} a_{i-1}}$, with the underlying idea that \mathcal{A} is $w_i a_i$ -universal iff $X_{w_i a_i}$ is empty. This permits to check u -universality when u ends with a Σ -symbol. Moreover, we will see that each element of $X_{w_i a_i}$ is a witness of some word v such that the tree t with $[t] = w_i a_i v$ is not accepted

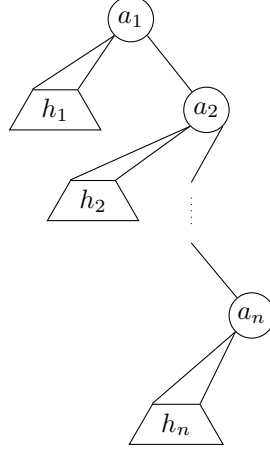


Figure 4: Current reading of a tree t_0 according to the prefix $a_1[h_1]a_2[h_2] \cdots a_n[h_n]$.

by \mathcal{A} . For words u ending with a $\bar{\Sigma}$ -symbol, we will explain at the end of this section how the test of $w_i a_i[h_i]$ -universality can be easily performed using the set $X_{w_i a_i}$.

3.4.1 Incremental approach

Let us give the definition of $X_{w_i a_i}$ for all i . We begin with the basic case $i = 1$, i.e. with set X_a .

We use notation P_{T_Σ} for $Post^*(\emptyset)$ and Π_{H_Σ} for $(Post^*(\emptyset))^*$ as introduced in Section 3.3.1 (recall that $Post^*(\emptyset) = \{P_t \mid t \in T_\Sigma\}$ and $(Post^*(\emptyset))^* = \{\pi_h \mid h \in H_\Sigma\}$ by Lemma 10). Given a set \mathcal{W} of macrostate words, we define $Pref(\mathcal{W})$ as the set $\{\pi \in \Pi_{H_\Sigma} \mid \exists \pi' \in \Pi_{H_\Sigma} : \pi\pi' \in \mathcal{W}\}$.

Basic case We need to define X_a such that $X_a = \emptyset$ iff \mathcal{A} is a -universal, i.e. all trees t such that $[t] = a[h]\bar{a}$ with $h \in H_\Sigma$, are accepted by \mathcal{A} . The test of a -universality is performed in two steps. We first collect all macrostate words $\pi_h \in \Pi_{H_\Sigma}$ (see Lemmas 9 and 10). Then for each of them we compute $Post_a(\pi_h)$ and check whether $Post_a(\pi_h) \cap Q_f \neq \emptyset$ (see Proposition 11). If for some π_h , we have $Post_a(\pi_h) \cap Q_f = \emptyset$, then π_h is a witness of non a -universality of \mathcal{A} , since $a[h]\bar{a}$ is not accepted by \mathcal{A} . More precisely, we have the next definition and proposition.

Definition 21. Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton, and let $a \in \Sigma$ be a letter. We define

$$X_a = \{\pi \in \Pi_{H_\Sigma} \mid Post_a(\pi) \cap Q_f = \emptyset\}.$$

Proposition 22. \mathcal{A} is a -universal iff $X_a = \emptyset$. Moreover, if X_a is not empty, for all $\pi \in X_a$, let $h \in H_\Sigma$ be such that $\pi = \pi_h$. Then $a[h]\bar{a} \in [T_\Sigma \setminus L(\mathcal{A})]$.

Let us now proceed with the general case, that is, the definition of $X_{w_i a_i}$ with $i > 1$. For all proper prefixes $w_j a_j$ of $w_i a_i$, we can suppose that \mathcal{A} is not $w_j a_j$ -universal, otherwise \mathcal{A} would be trivially $w_i a_i$ -universal. We define $X_{w_i a_i}$ and then, explain how to check $w_i a_i$ -universality knowing $X_{w_i a_i}$.

General case Let wa with $a \in \Sigma$. We first define X_{wa} . Let $w = w'a'[h']$ with $a' \in \Sigma$ and $h' \in H_\Sigma$. We suppose that \mathcal{A} is not $w'a'$ -universal, and that $X_{w'a'} \neq \emptyset$. Moreover $X_{w'a'}$ contains a witness of a word v such that the tree t with $[t] = w'a'v$ is not accepted by \mathcal{A} .

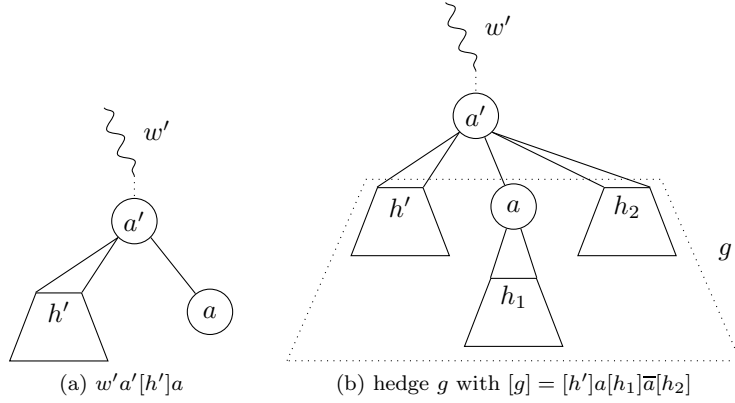


Figure 5: Current reading according to the prefix wa

Let us define the set X_{wa} from the set $X_{w'a'}$. In Figure 5 (a), we indicate the current reading of a tree according to wa : an internal node labeled by a' with a sequence of subtrees equal to h' followed by a child labeled by a . With this figure, we notice that \mathcal{A} is not wa -universal iff there exists $h_1, h_2 \in H_\Sigma$ such that for the hedge g with $[g] = [h']a[h_1]\bar{a}[h_2]$, we have $\pi_g \in X_{w'a'}$ (see Figure 5 (b)). This observation leads to the next definition of X_{wa} .

Definition 23. Let $wa \in P\text{Pref}(T_\Sigma)$ with $w = w'a'[h']$, $a, a' \in \Sigma$ and $h' \in H_\Sigma$. Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton. We define

$$X_{wa} = \{\pi \in \Pi_{H_\Sigma} \mid \pi_{h'} \text{Post}_a(\pi) \in \text{Pref}(X_{w'a'})\}.$$

As for the basic case (see Proposition 22), we have the next proposition.

Proposition 24. \mathcal{A} is wa -universal iff $X_{wa} = \emptyset$. Moreover, if X_{wa} is not empty, then $X_{wa} = \{\pi_h \in \Pi_{H_\Sigma} \mid \exists v : wa[h]\bar{a}v \in [T_\Sigma \setminus L(\mathcal{A})]\}$.

Proof. We proceed by induction on w to prove that $X_{wa} = \{\pi_h \in \Pi_{H_\Sigma} \mid \exists v : wa[h]\bar{a}v \in [T_\Sigma \setminus L(\mathcal{A})]\}$. The basic case, $w = \epsilon$, directly follows from Proposition 22.

Let $w = w'a'[h']$ with $a' \in \Sigma$ and $h' \in H_\Sigma$. Suppose that the property holds for $X_{w'a'}$, i.e. $X_{w'a'} = \{\pi_{h'} \in \Pi_{H_\Sigma} \mid \exists v' : w'a'[h']v' \in [T_\Sigma \setminus L(\mathcal{A})]\}$.

(\subseteq) Let $\pi_h \in X_{wa}$. By definition, $\exists h', h'' \in H_\Sigma : \pi_{h'} Post_a(\pi_h) \pi_{h''} \in X_{w'a'}$. Then, by induction hypothesis, $\exists v' : w'a'[h']a[h]\bar{a}[h'']\bar{a}'v' \in [T_\Sigma \setminus L(\mathcal{A})]$. Let $v = [h'']\bar{a}'v'$, then $wa[h]\bar{a}v \in [T_\Sigma \setminus L(\mathcal{A})]$.
(\supseteq) Let $\pi_h \in \Pi_{H_\Sigma}$ such that $\exists v : wa[h]\bar{a}v \in [T_\Sigma \setminus L(\mathcal{A})]$. So there exists a word v' and hedges h', h'' such that $w'a'[h']a[h]\bar{a}[h'']\bar{a}'v' \in [T_\Sigma \setminus L(\mathcal{A})]$. By induction hypothesis, $\pi_g \in X_{w'a'}$ with $[g] = [h']a[h]\bar{a}[h'']$, and thus $\pi_h \in X_{wa}$. \square

In this section, given a tree t_0 and the current read prefix u of $[t_0]$, we have shown how to test incrementally for u -universality as follows. Suppose that $u = a_1[h_1]a_2[h_2] \cdots a_n[h_n]$ with $a_i \in \Sigma, h_i \in H_\Sigma$, for $1 \leq i \leq n$, and let $w_i = a_1[h_1] \cdots a_{i-1}[h_{i-1}]$, for $1 \leq i \leq n$. We have defined set X_a and then each set $X_{w_i a_i}$, $1 < i \leq n$, from $X_{w_{i-1} a_{i-1}}$, such that \mathcal{A} is $w_i a_i$ -universal iff $X_{w_i a_i}$ is empty.

It should be noted that it is also possible to test whether \mathcal{A} is $w_i a_i[h_i]$ -universal thanks to set $X_{w_i a_i}$. Indeed, by Proposition 24, \mathcal{A} is $w_i a_i[h_i]$ -universal iff $\nexists \pi \in \Pi_{H_\Sigma} : \pi_{h_i} \pi \in X_{w_i a_i}$.

3.4.2 Algorithm for checking u -universality

In this section, we propose an algorithm for u -universality checking. As done before for universality in Section 3.3.2, we need to represent a macrostate word π by the relation $\text{rel}(\pi)$. Definitions 21 and 23 are rephrased as follows. Given a set Y of relations, we define $\text{Pref}(Y)$ as the set $\{r \in \text{rel}(\Pi_{H_\Sigma}) \mid \exists r' \in \text{rel}(\Pi_{H_\Sigma}) : rr' \in Y\}$.

Definition 25. Let $\mathcal{A} = (Q, \Sigma, Q_f, \Delta)$ be a hedge automaton, and let $wa \in \text{PPref}(T_\Sigma)$ with $a \in \Sigma$.

1. If $w = \epsilon$, we define $Y_a = \{r \in \text{rel}(\Pi_{H_\Sigma}) \mid \text{Post}_a(r) \cap Q_f = \emptyset\}$.
2. If $w \neq \epsilon$, given $w = w'a'[h']$ with $a' \in \Sigma$ and $h' \in H_\Sigma$, we define $Y_{wa} = \{r \in \text{rel}(\Pi_{H_\Sigma}) \mid \text{rel}(\pi_{h'})\text{rel}(\text{Post}_a(r)) \in \text{Pref}(Y_{w'a'})\}$.

Lemma 26. $Y_{wa} = \text{rel}(X_{wa})$.

Proof. The proof is done by induction on w .

The basic case, $Y_a = \text{rel}(X_a)$, follows from Lemma 15.

Let $w = w'a'[h']$ with $a' \in \Sigma$ and $h' \in H_\Sigma$. Suppose that $Y_{w'a'} = \text{rel}(X_{w'a'})$ holds. Notice that for $\pi \in X_{wa}$ and $\pi' \in \Pi_{H_\Sigma}$, if $\text{rel}(\pi) = \text{rel}(\pi')$, then $\pi' \in X_{wa}$ (see Lemma 15). We have for $r = \text{rel}(\pi) \in \text{rel}(\Pi_{H_\Sigma})$:

$$\begin{aligned} r \in Y_{wa} &\Leftrightarrow \exists r' \in \text{rel}(\Pi_{H_\Sigma}) : \text{rel}(\pi_{h'})\text{rel}(\text{Post}_a(r))r' \in Y_{w'a'} \\ &\Leftrightarrow \exists \pi' \in \Pi_{H_\Sigma} : \text{rel}(\pi_{h'}\text{Post}_a(\pi)\pi') \in \text{rel}(X_{w'a'}) \\ &\Leftrightarrow \pi \in \text{rel}(X_{wa}) \end{aligned}$$

It follows that $Y_{wa} = \text{rel}(X_{wa})$. \square

The next proposition is the equivalent of Propositions 22 and 24, as a consequence of Lemma 26.

Proposition 27. \mathcal{A} is wa -universal iff Y_{wa} is empty.

By definition of Y_{wa} , it follows that \mathcal{A} is $wa[h]$ -universal, with $h \in H_\Sigma$, iff $\nexists r \in \text{rel}(\Pi_{H_\Sigma}) : \text{rel}(\pi_h)r \in Y_{wa}$.

Let us now describe an algorithm to test whether a hedge automaton \mathcal{A} is u -universal. We recall that t_0 is a given tree and u its current read prefix. This algorithm is incremental and thus has already checked that \mathcal{A} is not wa -universal for all non-empty proper prefixes wa of u thanks to Proposition 27.

More precisely, let $u = a_1[h_1]a_2[h_2] \cdots a_n[h_n]$ and $w_i = a_1[h_1] \cdots a_{i-1}[h_{i-1}]$, for $1 \leq i \leq n$, and suppose that for all i the sets $Y_{w_i a_i}$ have been computed and seen to be non empty. A stack is used to store all triples $(Y_{w_i a_i}, \text{rel}(h_i), a_i)$, $1 \leq i \leq n$, with the triple $(Y_{w_n a_n}, \text{rel}(h_n), a_n)$ at the top of the stack. The stack has a depth equal to the length of $\text{open}(u)$.

In Algorithm 5, four functions are called according to the letter that is currently read in t_0 knowing that u is the last read prefix of t_0 . If it is the first letter a (resp. last letter \bar{a}) of $[t_0]$, then Function $\text{OPENROOT}(a)$ (resp. $\text{CLOSEROOT}(a)$) is called. Otherwise either Function $\text{NEXTOPENTAG}(a)$ or $\text{NEXTCLOSEDTAG}(a)$ is called according to whether a or \bar{a} is the next read letter.

Function $\text{OPENROOT}(a)$ computes the set Y_a as defined in Definition 25. If Y_a is empty, then \mathcal{A} is declared a -universal. Otherwise, the stack is initialized with the triple (Y_a, id, a) .

Function $\text{CLOSEROOT}(a)$ pops the stack to get its unique triple (Y_a, r, a) (since \bar{a} is the last letter of $[t_0]$). It checks whether $t_0 = u\bar{a}$ is accepted by the automaton with the emptiness test of $\text{Post}_a(r) \cap Q_f$.

If $u \neq \epsilon$ and the letter read after u is a with $a \in \Sigma$, then Function $\text{NEXTOPENTAG}(a)$ reads the triple (Y', r', a') at the top of the stack and computes the Y_{ua} from the set Y' (as in Definition 25). If Y_{ua} is empty, then \mathcal{A} is declared ua -universal. Otherwise, the triple (Y_{ua}, id, a) is pushed on the stack. If the letter read after u is \bar{a} with $\bar{a} \in \bar{\Sigma}$, and $u\bar{a} \neq t_0$, then Function $\text{NEXTCLOSEDTAG}(a)$ pops once the stack to get the triple (Y, r, a) (notice that \bar{a} is the closing tag of a in this triple). It then modifies the triple (Y', r', a') at the top of stack, by replacing r' by $r'' = r' \circ \text{rel}(\text{Post}_a(r))$ (see Figure 5 (b)). If there does not exist $s \in \text{rel}(\Pi_{H_\Sigma})$ such that $r''s \in Y'$, then \mathcal{A} is declared to be $u\bar{a}$ -universal.

These four functions return True as soon as they can declare that \mathcal{A} is u -universal for the current read prefix u of $[t_0]$.

3.4.3 Antichain-based optimization

In this section we explain how to use the concept of antichain to avoid some computations when checking for u -universality. In particular we show that it is sufficient to only compute the \subseteq -maximal elements of set Y_{wa} as defined in Definition 25.

Lemma 28. Let $wa \in \text{PPref}(T_\Sigma)$ with $a \in \Sigma$, Y_{wa} is a \subseteq -downward closed set.

Proof. We proceed by induction on w . Notice that for $r, r' \in \text{rel}(\Pi_{H_\Sigma})$ and $a \in \Sigma$, if $r' \subseteq r$, then $\text{Post}_a(r') \subseteq \text{Post}_a(r)$.

Algorithm 5 Functions used for checking u -universality incrementally

```

function OPENROOT( $a$ )
   $Y \leftarrow \emptyset$ 
  for  $r \in \text{rel}(\Pi_{H_\Sigma})$  do
    if  $\text{Post}_a(r) \cap Q_f = \emptyset$  then
       $Y \leftarrow Y \cup \{r\}$ 
    end if
  end for
  if  $Y = \emptyset$  then
    return True //  $u$ -universal with  $u$  the current read prefix
  else
     $\text{Stack} \leftarrow \emptyset$ 
    PUSH( $\text{Stack}, (Y, \text{id}, a)$ )
  end if
end function

function CLOSEROOT( $a$ )
   $(Y, r, a) \leftarrow \text{POP}(\text{Stack})$ 
  if  $\text{Post}_a(r) \cap Q_f = \emptyset$  then
    return False //  $t_0$  is not accepted
  else
    return True //  $t_0$  is accepted
  end if
end function

function NEXTOPENTAG( $a$ )
   $(Y', r', a') \leftarrow \text{TOP}(\text{Stack})$ 
   $Y \leftarrow \emptyset$ 
  for  $r \in \text{rel}(\Pi_{H_\Sigma})$  do
    if  $r' \circ \text{rel}(\text{Post}_a(r)) \in \text{Pref}(Y')$  then
       $Y \leftarrow Y \cup \{r\}$ 
    end if
  end for
  if  $Y = \emptyset$  then
    return True //  $u$ -universal with  $u$  the current read prefix
  else
    PUSH( $\text{Stack}, (Y, \text{id}, a)$ )
  end if
end function

function NEXTCLOSEDTAG( $a$ )
   $(Y, r, a) \leftarrow \text{POP}(\text{Stack})$ 
   $(Y', r', a') \leftarrow \text{POP}(\text{Stack})$ 
   $r' \leftarrow r' \circ \text{rel}(\text{Post}_a(r))$ 
  if  $\nexists s \in \text{rel}(\Pi_{H_\Sigma}) : r' \circ s \in Y'$  then
    return True //  $u$ -universal with  $u$  the current read prefix
  end if
  PUSH( $\text{Stack}, (Y', r', a')$ )
end function

```

Consider the basic case where $w = \epsilon$. By definition $Y_a = \{r \in \text{rel}(\Pi_{H_\Sigma}) \mid \text{Post}_a(r) \cap Q_f = \emptyset\}$. By the previous remark, Y_a is a \subseteq -downward closed set.

Let $w = w'a'[h']$, with $a' \in \Sigma$ and $h' \in H_\Sigma$. Let $r \in Y_{wa}$ and $r' \in \text{rel}(\Pi_{H_\Sigma})$ such that $r' \subseteq r$. Let us show that $r' \in Y_{wa}$. As $r \in Y_{wa}$, $\exists r'' \in \text{rel}(\Pi_{H_\Sigma}) : \text{rel}(\pi_{h'})\text{rel}(\text{Post}_a(r))r'' \in Y_{w'a'}$. As $\text{Post}_a(r') \subseteq \text{Post}_a(r)$ and $Y_{w'a'}$ is \subseteq -downward closed, it follows that $\text{rel}(\pi_{h'})\text{rel}(\text{Post}_a(r'))r'' \in Y_{w'a'}$ and then $r' \in Y_{wa}$. \square

As Y_{wa} is \subseteq -downward closed, it can be described by the antichain $\lceil Y_{wa} \rceil$ of its maximal elements. Let $w = w'a'[h']$ with $a \in \Sigma$ and $h' \in H_\Sigma$, the next lemma shows that it is possible to compute Y_{wa} from $\lceil Y_{w'a'} \rceil$ without knowing the whole set $Y_{w'a'}$.

Lemma 29. *For $r \in \text{rel}(\Pi_{H_\Sigma})$, $r \in Y_{wa}$ iff there exist $r' \in \lfloor \text{rel}(\Pi_{H_\Sigma}) \rfloor$ and $s \in \lceil Y_{w'a'} \rceil$ such that $\text{rel}(\pi_{h'})\text{rel}(\text{Post}_a(r))r' \subseteq s$.*

Proof.

$$\begin{aligned} r \in Y_{wa} &\iff \exists r' \in \text{rel}(\Pi_{H_\Sigma}) : \text{rel}(\pi_{h'})\text{rel}(\text{Post}_a(r))r' \in Y_{w'a'} \quad (\text{Def. 25}) \\ &\iff \exists r' \in \text{rel}(\Pi_{H_\Sigma}), \exists s \in \lceil Y_{w'a'} \rceil : \text{rel}(\pi_{h'})\text{rel}(\text{Post}_a(r))r' \subseteq s \\ &\iff \exists r' \in \lfloor \text{rel}(\Pi_{H_\Sigma}) \rfloor, \exists s \in \lceil Y_{w'a'} \rceil : \text{rel}(\pi_{h'})\text{rel}(\text{Post}_a(r))r' \subseteq s \end{aligned}$$

\square

Based on the previous lemma, Algorithm 6 is an optimized version of Function `NEXTOPENTAG`(u, a) which computes $Y = \lceil Y_{wa} \rceil$ from $Y' = \lceil Y_{w'a'} \rceil$ without computing the entire set Y_{wa} . The idea is to have a set, called *Candidates*, containing all elements that could be potentially in Y . Initially, it is the set $\text{rel}(\Pi_{H_\Sigma})$. Otherwise, suppose that Y has been partially computed, then *Candidates* is the set $\text{rel}(\Pi_{H_\Sigma}) \setminus \{r' \mid \exists r \in Y : r' \subseteq r\}$. Function `MAXIMALELEMENT`(*Candidates*) returns a maximal element of the set *Candidates*.

4 Safe configurations approach

We present an algorithm for testing u -universality of a non-deterministic visibly pushdown automaton \mathcal{A} . This algorithm is a generalization of the algorithm for the deterministic case [GNT09], adding several optimizations to avoid huge computations. As in Section 3.4, the algorithm is incremental in the sense that the linearization $[t_0]$ of a given tree t_0 is read letter by letter, and while \mathcal{A} is not u -universal for the current read prefix u of $[t_0]$, the next letter of $[t_0]$ is read.

4.1 Safe configurations

In the deterministic case [GNT09], the algorithm relies on the incremental computation of the set of safe states. In the non-deterministic case, safe states are not enough to decide u -universality. Indeed In [GNT09], safe states are computed according to the unique run of the deterministic automaton on u . In fact,

Algorithm 6 Optimized Function NEXTOPENTAG

```

function OPTNEXTOPENTAG( $u, a$ )
  ( $Y', r', a'$ )  $\leftarrow$  TOP( $Stack$ )
   $Y \leftarrow \emptyset$ 
   $Candidates \leftarrow \text{rel}(\Pi_{H_\Sigma})$ 
  while  $Candidates \neq \emptyset$  do
     $r \leftarrow \text{MAXIMALELEMENT}(Candidates)$ 
    if  $\exists r'' \in \lfloor \text{rel}(\Pi_{H_\Sigma}) \rfloor, \exists s \in Y' : r' \circ \text{rel}(Post_a(r)) \circ r'' \subseteq s$  then
       $Y \leftarrow Y \cup \{r\}$ 
       $Candidates \leftarrow \{r' \in Candidates \mid r' \not\subseteq r\}$ 
    else
       $Candidates \leftarrow Candidates \setminus \{r\}$ 
    end if
  end while
  if  $Y = \emptyset$  then
    return  $\mathcal{A}$  is  $ua$ -universal
  else
    PUSH( $Stack, (Y, \text{id}, a)$ )
  end if
end function

```

safe configurations (q, σ) are considered, but all these configurations have the same stack σ here, so only states q have to be stored. When the automaton is non-deterministic, we may have several runs on u , and each of them may use a different stack. All these stacks have to be considered for testing u -universality, so we cannot consider only states.

Therefore, we have to consider *safe configurations*, or more precisely sets of safe configurations as described in the next definition. We use notions about VPAs that are defined in Section 2.2, as well sets of configurations that are antichains with respect to \subseteq , or \subseteq -upward (resp. \subseteq -downward) closed sets (see Section 3.3.3).

Definition 30. Let \mathcal{A} be a VPA and $\mathcal{C} \subseteq Q \times \Gamma^*$ be a set of configurations. Let $u \in \text{PPref}(T_\Sigma)$ be a prefix.

- \mathcal{C} is safe for u if for every v such that $uv \in [T_\Sigma]$, there exist $(q, \sigma) \in \mathcal{C}$ and $p \in Q_f$ such that $(q, \sigma) \xrightarrow{v} (p, \epsilon)$ in \mathcal{A} .
- \mathcal{C} is leaf-safe for u if for every $v = \bar{a}v'$ with $\bar{a} \in \bar{\Sigma}$ such that $uv \in [T_\Sigma]$, there exist $(q, \sigma) \in \mathcal{C}$ and $p \in Q_f$ such that $(q, \sigma) \xrightarrow{v} (p, \epsilon)$ in \mathcal{A} .

We write $\text{Safe}(u)$ for $\{\mathcal{C} \mid \mathcal{C} \text{ is safe for } u\}$ and $\text{LSafe}(u)$ for $\{\mathcal{C} \mid \mathcal{C} \text{ is leaf-safe for } u\}$.

Intuitively, as stated in Theorem 32 below, if \mathcal{C} is the set of configurations reached in \mathcal{A} after reading u , then \mathcal{A} is u -universal iff \mathcal{C} is safe for u . Indeed, for every possible v , one can find in \mathcal{C} at least one configuration leading to an

accepting configuration after reading v . We first note that, from the definitions, if a set of configurations \mathcal{C} is safe (resp. leaf-safe) for u , then a larger set \mathcal{C}' is also safe (resp. leaf-safe) for u .

Lemma 31. *Safe(u) and LSafe(u) are \subseteq -upward closed sets.*

Let $Reach(u)$ denote the set of configurations (q, σ) such that $(q_0, \sigma_0) \xrightarrow{u} (q, \sigma)$ for some initial configuration (q_0, σ_0) of \mathcal{A} .

Theorem 32. *\mathcal{A} is u -universal iff $Reach(u) \in Safe(u)$.*

Proof. (\Rightarrow) Assume that \mathcal{A} is u -universal. Consider the set \mathcal{C} of configurations (q, σ) of \mathcal{A} such that there exists $v \in (\Sigma \cup \overline{\Sigma})^*$, $q_i \in Q_i$ and $q_f \in Q_f$ verifying $uv \in [T_\Sigma]$ and $(q_i, \epsilon) \xrightarrow{u} (q, \sigma) \xrightarrow{v} (q_f, \epsilon)$. We have $\mathcal{C} \subseteq Reach(u)$.

Let v be such that $uv \in [T_\Sigma]$. As \mathcal{A} is u -universal, there exists a configuration $(q, \sigma) \in \mathcal{C}$ such that $(q, \sigma) \xrightarrow{v} (q_f, \epsilon)$ with $q_f \in Q_f$. Hence $\mathcal{C} \in Safe(u)$. By Lemma 31, we get $Reach(u) \in Safe(u)$.

(\Leftarrow) Assume now that $Reach(u) \in Safe(u)$, and let v be such that $uv \in [T_\Sigma]$. As $Reach(u) \in Safe(u)$, there exists $(q, \sigma) \in Reach(u)$ and $p \in Q_f$ such that $(q, \sigma) \xrightarrow{v} (p, \epsilon)$. Thus, $uv \in L(\mathcal{A})$, and \mathcal{A} is u -universal. \square

4.2 Incremental definition of safe configurations

In this section, we detail how set $Safe(u)$ of safe configurations can be defined from set $Safe(u')$ with u' a proper prefix of u . In this way, while reading the linearization $[t_0]$ of a given tree t_0 , set $Safe(u)$ with u prefix of $[t_0]$, can be incrementally defined. In the next section, we will turn this approach into an algorithm.

4.2.1 Starting point

The starting point is to begin with $Safe(a)$ for which we recall the definition.

$$Safe(a) = \{\mathcal{C} \mid \forall h \in H_\Sigma, \exists q_f \in Q_f, \exists (q, \sigma) \in \mathcal{C} : (q, \sigma) \xrightarrow{h\bar{a}} (q_f, \epsilon)\}.$$

4.2.2 Reading a letter $\bar{a} \in \overline{\Sigma}$

When reading an $\bar{a} \in \overline{\Sigma}$, we can retrieve safe configurations from prior sets of safe configurations:

$$Safe(u\bar{a}) = Safe(u')$$

where u' is the unique prefix of u such that $u = u'a[h]$. Indeed as shown by Lemma 33 below, we have $Safe(u'a[h]\bar{a}) = Safe(u')$.

Hence, from an algorithmic point of view, we just have to use a stack to store these safe configurations. When opening a , we put $Safe(u')$ on the stack, and when closing \bar{a} , we pop it. As h is a hedge, the stack before reading \bar{a} is exactly the stack after reading a .

Lemma 33. *If $h \in H_\Sigma$, then $Safe(u[h]) = Safe(u)$ and $LSafe(u[h]) = LSafe(u)$.*

Proof. (\supseteq) Assume $\mathcal{C} \in \text{Safe}(u)$, and let v be such that $u[h]v \in [T_\Sigma]$. As h is a hedge, we have $uv \in [T_\Sigma]$. As $\mathcal{C} \in \text{Safe}(u)$, there exists $(q, \sigma) \in \mathcal{C}$ such that $(q, \sigma) \xrightarrow{v} (p, \epsilon)$ with $p \in Q_f$. So $\mathcal{C} \in \text{Safe}(u[h])$.

(\subseteq) Conversely, assume $\mathcal{C} \in \text{Safe}(u[h])$. Let v be such that $uv \in [T_\Sigma]$. We also have $u[h]v \in [T_\Sigma]$, so there exists $(q, \sigma) \in \mathcal{C}$ such that $(q, \sigma) \xrightarrow{v} (p, \epsilon)$ with $p \in Q_f$. Thus $\mathcal{C} \in \text{Safe}(u)$.

The proof is the same for $\text{LSafe}(u[h]) = \text{LSafe}(u)$, except that we only consider v of the form $\bar{a}v'$. \square

In the rest of Section 4, we only treat sets $\text{Safe}(ua)$ since the way of computing sets $\text{Safe}(u\bar{a})$ has been just detailed. The case of sets $\text{Safe}(ua)$ is much more involved.

4.2.3 Reading a letter $a \in \Sigma$

When reading an $a \in \Sigma$, two successive steps are performed, with leaf-safe configurations as intermediate object:

$$\text{Safe}(u) \xrightarrow{\text{Step 1}} \text{LSafe}(ua) \xrightarrow{\text{Step 2}} \text{Safe}(ua)$$

We now detail Step 1 and Step 2, i.e. how $\text{LSafe}(ua)$ can be defined from $\text{Safe}(u)$, and how $\text{Safe}(ua)$ is defined from $\text{LSafe}(ua)$. Proposition 34 gives a first idea of these links. Equivalence (1) states that a set of configurations \mathcal{C} is leaf-safe for ua iff after performing a $\text{Post}_{\bar{a}}(\mathcal{C})$ we get a safe set of configurations for u . Equivalence (2) states that safe configurations for ua are those from which traversing any hedge leads to a leaf-safe set of configurations, i.e. one can safely close the a -node. Proposition 34 thus relates sets $\text{Safe}(u)$, $\text{LSafe}(ua)$, and $\text{Safe}(ua)$, however backwardly. Proposition 38 hereafter will relates them in the right direction.

Proposition 34. *Let $ua \in \text{PPref}(T_\Sigma)$ with $a \in \Sigma$.*

$$\mathcal{C} \in \text{LSafe}(ua) \iff \text{Post}_{\bar{a}}(\mathcal{C}) \in \text{Safe}(u) \quad (1)$$

$$\mathcal{C} \in \text{Safe}(ua) \iff \forall h \in H_\Sigma, \text{Post}_{[h]}(\mathcal{C}) \in \text{LSafe}(ua) \quad (2)$$

Proof. (1, \Rightarrow) Let $\mathcal{C} \in \text{LSafe}(ua)$ and $\mathcal{C}' = \text{Post}_{\bar{a}}(\mathcal{C})$. Let us show that $\mathcal{C}' \in \text{Safe}(u)$. By Lemma 33, it is sufficient to prove that $\mathcal{C}' \in \text{Safe}(ua\bar{a})$. Let v such that $ua\bar{a}v \in [T_\Sigma]$. As $\mathcal{C} \in \text{LSafe}(ua)$ and $\bar{a}v$ starts with $\bar{a} \in \bar{\Sigma}$, there exists $(q, \sigma) \in \mathcal{C}$ and (q', σ') such that $(q, \sigma) \xrightarrow{\bar{a}} (q', \sigma') \xrightarrow{v} (p, \epsilon)$ for some $p \in Q_f$. By definition of $\text{Post}_{\bar{a}}(\mathcal{C})$ we have $(q', \sigma') \in \mathcal{C}'$ and thus $\mathcal{C}' \in \text{Safe}(ua\bar{a})$.

(1, \Leftarrow) For the converse, let $\mathcal{C}' = \text{Post}_{\bar{a}}(\mathcal{C}) \in \text{Safe}(u) = \text{Safe}(ua\bar{a})$. Let us show that $\mathcal{C} \in \text{LSafe}(ua)$. Let v be such that $uav \in [T_\Sigma]$ and $v = \bar{b}v'$. We necessarily have $\bar{a} = \bar{b}$. As $\mathcal{C}' \in \text{Safe}(ua\bar{a})$, there exists $(q', \sigma') \in \mathcal{C}'$ such that $(q', \sigma') \xrightarrow{v'} (p, \epsilon)$ with $p \in Q_f$. By definition of $\text{Post}_{\bar{a}}(\mathcal{C})$, there also exists $(q, \sigma) \in \mathcal{C}$ such that $(q, \sigma) \xrightarrow{\bar{a}} (q', \sigma')$ and thus $(q, \sigma) \xrightarrow{v=\bar{a}v'} (p, \epsilon)$ with $p \in Q_f$.

(2, \Rightarrow) Let $\mathcal{C} \in \text{Safe}(ua)$ and $h \in H_\Sigma$. Let us show that $\mathcal{C}' = \text{Post}_{[h]}(\mathcal{C})$ is in $\text{LSafe}(ua)$. Let v such that $uav \in [T_\Sigma]$ and $v = \bar{b}v'$. We must have $\bar{a} = \bar{b}$. We also have $ua[h]\bar{a}v' \in [T_\Sigma]$. As $\mathcal{C} \in \text{Safe}(ua)$, there exists $(q, \sigma) \in \mathcal{C}$ such that $(q, \sigma) \xrightarrow{[h]} (q', \sigma') \xrightarrow{v=\bar{a}v'} (p, \epsilon)$ with $p \in Q_f$. By definition of $\text{Post}_{[h]}(\mathcal{C})$, $(q', \sigma') \in \mathcal{C}'$.

(2, \Leftarrow) Let us assume that for every hedge $h \in H_\Sigma$, $\text{Post}_{[h]}(\mathcal{C}) \in \text{LSafe}(ua)$. Let us show that $\mathcal{C} \in \text{Safe}(ua)$. Let v be such that $uav \in [T_\Sigma]$. Then we have $uav = ua[h]\bar{a}v'$ for some $h \in H_\Sigma$. As $\mathcal{C}' = \text{Post}_{[h]}(\mathcal{C}) \in \text{LSafe}(ua)$, $\bar{a}v'$ starts with $\bar{a} \in \bar{\Sigma}$ and $ua\bar{a}v' \in [T_\Sigma]$, there exists $(q', \sigma') \in \mathcal{C}'$ such that $(q', \sigma') \xrightarrow{\bar{a}v'} (p, \epsilon)$ for some $p \in Q_f$. Hence, by definition of $\text{Post}_{[h]}(\mathcal{C})$, there also exists $(q, \sigma) \in \mathcal{C}$ such that $(q, \sigma) \xrightarrow{[h]} (q', \sigma') \xrightarrow{\bar{a}v'} (p, \epsilon)$ with $p \in Q_f$. \square

We propose now the notion of *predecessor* in a way to get Step 1 and Step 2 in the right direction.

Definition 35. Let $\mathcal{C}, \mathcal{C}'$ be two sets of configurations, $\bar{a} \in \bar{\Sigma}$ and $h \in H_\Sigma$.

- \mathcal{C} is an \bar{a} -predecessor of \mathcal{C}' if $\forall (q', \sigma') \in \mathcal{C}', \exists (q, \sigma) \in \mathcal{C}, (q, \sigma) \xrightarrow{\bar{a}} (q', \sigma')$.
- \mathcal{C} is an h -predecessor of \mathcal{C}' if $\forall (q', \sigma') \in \mathcal{C}', \exists (q, \sigma) \in \mathcal{C}, (q, \sigma) \xrightarrow{[h]} (q', \sigma')$.

Let $\text{Pred}_{\bar{a}}(\mathcal{C}') = \{\mathcal{C} \mid \mathcal{C} \text{ is an } \bar{a}\text{-predecessor of } \mathcal{C}'\}$ and $\text{Pred}_h(\mathcal{C}') = \{\mathcal{C} \mid \mathcal{C} \text{ is an } h\text{-predecessor of } \mathcal{C}'\}$.

From their definitions, the sets of predecessors are \subseteq -upward closed.

Lemma 36. $\text{Pred}_{\bar{a}}(\mathcal{C}')$ and $\text{Pred}_h(\mathcal{C}')$ are \subseteq -upward closed sets.

Predecessors closely relate to the *Post* operator.

Lemma 37. \mathcal{C} is an \bar{a} -predecessor of $\text{Post}_{\bar{a}}(\mathcal{C})$. If \mathcal{C} is an \bar{a} -predecessor of \mathcal{C}' then $\mathcal{C}' \subseteq \text{Post}_{\bar{a}}(\mathcal{C})$. Both properties also hold for $\text{Post}_{[h]}(\mathcal{C})$.

We can now rephrase Proposition 34 in terms of predecessors.

Proposition 38. Let $ua \in \text{PPref}(T_\Sigma)$.

$$\mathcal{C} \in \text{LSafe}(ua) \iff \exists \mathcal{C}' \in \text{Safe}(u), \mathcal{C} \text{ is an } \bar{a}\text{-predecessor of } \mathcal{C}' \quad (3)$$

$$\mathcal{C} \in \text{Safe}(ua) \iff \forall h \in H_\Sigma, \exists \mathcal{C}' \in \text{LSafe}(ua), \mathcal{C} \text{ is a } h\text{-predecessor of } \mathcal{C}' \quad (4)$$

Proof. (3, \Rightarrow) Let $\mathcal{C} \in \text{LSafe}(ua)$. Then by Proposition 34, $\text{Post}_{\bar{a}}(\mathcal{C}) \in \text{Safe}(u)$. Moreover, \mathcal{C} is an \bar{a} -predecessor of $\text{Post}_{\bar{a}}(\mathcal{C})$ by Lemma 37.

(3, \Leftarrow) Let \mathcal{C} be an \bar{a} -predecessor of \mathcal{C}' , with $\mathcal{C}' \in \text{Safe}(u)$. By Lemma 37, $\mathcal{C}' \subseteq \text{Post}_{\bar{a}}(\mathcal{C})$. By Lemma 31, we also have $\text{Post}_{\bar{a}}(\mathcal{C}) \in \text{Safe}(u)$, so $\mathcal{C} \in \text{LSafe}(ua)$ by Proposition 34.

(4) Same proofs, except that \bar{a} has to be replaced by h , for all $h \in H_\Sigma$. \square

Proposition 38 can be used to perform Step 1 and Step 2 of our method. It states that safe sets of configurations are only among predecessors of prior safe sets of configurations. However, the number of hedges to consider in equivalence (4) is infinite. We use relations to overcome this. Also the size of $\text{Safe}(u)$ may be huge and not all configurations of $\text{Safe}(u)$ are crucial for checking u -universality. We use antichains to have a representation of $\text{Safe}(u)$ and to avoid computations of elements which are not crucial. These two concepts are explained in the following in a way to get an algorithm for incrementally checking u -universality.

4.3 An algorithm for u -universality

4.3.1 Antichains

Let $\lfloor \text{Safe}(u) \rfloor$ denote the set of elements of $\text{Safe}(u)$ which are minimal for \subseteq , similarly for $\text{LSafe}(u)$. These antichains are finite objects.

Proposition 39. $\lfloor \text{Safe}(u) \rfloor$ and $\lfloor \text{LSafe}(u) \rfloor$ are finite and only contain finite sets of configurations.

Proof. We begin with the following observation. Let v be such that $[uv] \in T_\Sigma$ and $(q, \sigma) \xrightarrow{v} (p, \epsilon)$ with $p \in Q_f$. Let $u' = \text{open}(u)$ (recall that $\text{open}(u)$ is the word obtained from u by removing all factors that are linearizations of hedges). Let v' be the word obtained from v in the same way. Then $|u'| = |v'|$ and $|u'| = |\sigma|$.

Let $\mathcal{C} \in \text{Safe}(u)$. Then by definition

$$\forall v, uv \in [T_\Sigma] \implies \exists (q, \sigma) \in \mathcal{C}, (q, \sigma) \xrightarrow{v} (p, \epsilon) \text{ with } p \in Q_f.$$

If \mathcal{C} is minimal with respect to \subseteq , then every $(q, \sigma) \in \mathcal{C}$ is used for at least one v in the previous definition. Now by the previous observation, each such (q, σ) belongs to $Q \times \Gamma^{|u'|}$. Hence $\mathcal{C} \subseteq Q \times \Gamma^{|u'|}$, and thus both \mathcal{C} and $\lfloor \text{Safe}(u) \rfloor$ are finite.

The same arguments hold for proving that $\lfloor \text{LSafe}(u) \rfloor$ is finite and contains only finite sets of configurations. \square

We now try to use these antichains in the starting point, and in Steps 1 and 2 of our approach.

4.3.2 Step 1 with antichains: from $\lfloor \text{Safe}(u) \rfloor$ to $\lfloor \text{LSafe}(ua) \rfloor$

For the two steps, the goal is to adapt Proposition 38 so that it uses $\lfloor \text{Safe}(\cdot) \rfloor$ instead of $\text{Safe}(\cdot)$, and $\lfloor \text{LSafe}(\cdot) \rfloor$ instead of $\text{LSafe}(\cdot)$. We begin with Step 1. Implication (\Rightarrow) of equivalence (3) can be directly adapted.

Proposition 40. Let $ua \in \text{PPref}(T_\Sigma)$.

$$\mathcal{C} \in \lfloor \text{LSafe}(ua) \rfloor \implies \exists \mathcal{C}' \in \lfloor \text{Safe}(u) \rfloor, \mathcal{C} \text{ is an } \bar{a}\text{-predecessor of } \mathcal{C}'$$

Proof. Let $\mathcal{C} \in \lfloor L\text{Safe}(ua) \rfloor$ and let $\mathcal{C}' = \text{Post}_{\bar{a}}(\mathcal{C})$. We know from Proposition 34 that $\mathcal{C}' \in \text{Safe}(u)$. Let $\mathcal{C}'_0 \subseteq \mathcal{C}'$ such that $\mathcal{C}'_0 \in \lfloor \text{Safe}(u) \rfloor$. From the definition of \mathcal{C}' we get:

$$\forall c' \in \mathcal{C}', \exists c \in \mathcal{C}, c \xrightarrow{\bar{a}} c'$$

We build \mathcal{C}_0 from these $c \in \mathcal{C}$ but for $c' \in \mathcal{C}'_0$:

$$\mathcal{C}_0 = \{c \in \mathcal{C} \mid \exists c' \in \mathcal{C}'_0, c \xrightarrow{\bar{a}} c'\}$$

Figure 6 illustrates the construction. \mathcal{C}_0 is an \bar{a} -predecessor of \mathcal{C}'_0 , so using

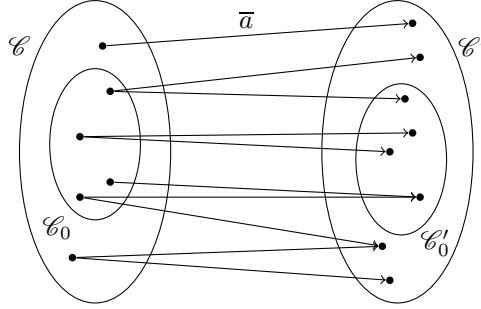


Figure 6: Construction of \mathcal{C}_0

Proposition 38, we get $\mathcal{C}_0 \in L\text{Safe}(ua)$. Furthermore, $\mathcal{C}_0 \subseteq \mathcal{C} \in \lfloor L\text{Safe}(ua) \rfloor$, so $\mathcal{C}_0 = \mathcal{C}$, and \mathcal{C} is obtained as an \bar{a} -predecessor of $\mathcal{C}'_0 \in \lfloor \text{Safe}(u) \rfloor$. \square

Proposition 40 gives us a way to compute $\lfloor L\text{Safe}(ua) \rfloor$ from $\lfloor \text{Safe}(u) \rfloor$: it suffices to take all \bar{a} -predecessors of elements of $\lfloor \text{Safe}(u) \rfloor$ and then limit to those predecessors that are \subseteq -minimal. We can even only consider minimal \bar{a} -predecessors of $\lfloor \text{Safe}(u) \rfloor$ in the following sense: \mathcal{C} is a *minimal \bar{a} -predecessor* of \mathcal{C}' if for all \mathcal{C}'' \bar{a} -predecessor of \mathcal{C}' , $\mathcal{C}'' \subseteq \mathcal{C} \implies \mathcal{C}'' = \mathcal{C}$. We finally obtain:

Corollary 41.

$$\lfloor L\text{Safe}(ua) \rfloor = \lfloor \{\mathcal{C} \mid \mathcal{C} \text{ is a minimal } \bar{a}\text{-predecessor of } \mathcal{C}' \in \lfloor \text{Safe}(u) \rfloor\} \rfloor$$

4.3.3 Step 2 with antichains: from $\lfloor L\text{Safe}(ua) \rfloor$ to $\lfloor \text{Safe}(ua) \rfloor$

The second step for computing $\lfloor \text{Safe}(ua) \rfloor$ from $\lfloor \text{Safe}(u) \rfloor$ relies on the introduction of antichains in equivalence (4) of Proposition 38. Implication (\implies) holds with antichains.

Proposition 42. Let $ua \in P\text{Pref}(T_\Sigma)$.

$$\mathcal{C} \in \lfloor \text{Safe}(ua) \rfloor \implies \forall h \in H_\Sigma, \exists \mathcal{C}' \in \lfloor L\text{Safe}(ua) \rfloor, \mathcal{C} \text{ is a } h\text{-predecessor of } \mathcal{C}'$$

Proof. The proof is in the same vein as for Proposition 40. Let $\mathcal{C} \in \llbracket \text{Safe}(ua) \rrbracket$, and $h \in H_\Sigma$. Let $\mathcal{C}'_h = \text{Post}_{[h]}(\mathcal{C})$. By Proposition 34, $\mathcal{C}'_h \in \text{LSafe}(ua)$. Let $\mathcal{C}''_h \subseteq \mathcal{C}'_h$ such that $\mathcal{C}''_h \in \llbracket \text{LSafe}(ua) \rrbracket$. We know that $\forall c' \in \mathcal{C}'_h, \exists c \in \mathcal{C}$ such that $c \xrightarrow{[h]} c'$. We define $\mathcal{C}_h = \{c \in \mathcal{C} \mid \exists c' \in \mathcal{C}''_h, c \xrightarrow{[h]} c'\}$. For every $h \in H_\Sigma$, \mathcal{C}_h is a h -predecessor of $\mathcal{C}''_h \in \text{LSafe}(ua)$. Consider $\mathcal{C}_\cup = \bigcup_{h \in H_\Sigma} \mathcal{C}_h$, then \mathcal{C}_\cup is also a h -predecessor of \mathcal{C}''_h . Using Proposition 38, we have $\mathcal{C}_\cup \in \text{Safe}(ua)$. As $\mathcal{C}_\cup \subseteq \mathcal{C}$ and $\mathcal{C} \in \llbracket \text{Safe}(ua) \rrbracket$, we also have that $\mathcal{C}_\cup = \mathcal{C}$. Hence \mathcal{C} verifies that $\forall h \in H_\Sigma, \exists \mathcal{C}'' \in \llbracket \text{LSafe}(ua) \rrbracket$ such that \mathcal{C} is a h -predecessor of \mathcal{C}'' . \square

Note that this proof does not use the fact that ua ends with a symbol in Σ , so Proposition 42 also holds when replacing ua by u .

Similarly to Proposition 40, we can restrict h -predecessors to consider to only minimal ones: \mathcal{C} is a *minimal h -predecessor* of \mathcal{C}' if for all \mathcal{C}'' h -predecessor of \mathcal{C}' , $\mathcal{C}'' \subseteq \mathcal{C} \implies \mathcal{C}'' = \mathcal{C}$. We obtain:

Corollary 43.

$$\llbracket \text{Safe}(ua) \rrbracket = \left\{ \mathcal{C} \mid \mathcal{C} = \bigcup_{h \in H_\Sigma} \mathcal{C}_h \text{ with } \mathcal{C}_h \text{ a minimal } h\text{-predecessor of } \mathcal{C}' \in \llbracket \text{LSafe}(ua) \rrbracket \right\}$$

This definition does not provide an algorithm, as it still relies on a quantification over an infinite number of hedges $h \in H_\Sigma$. In fact, only a finite number of such hedges needs to be considered. The reason is that a hedge does not change the original stack during the run of a VPA, so a hedge can be considered as a function mapping each state q to the set of states obtained when traversing h from q . Formally, we have the next definition.

Definition 44. For every $h \in H_\Sigma$, rel_h is the function from Q to 2^Q such that $q' \in \text{rel}_h(q)$ iff $(q, \sigma) \xrightarrow{[h]} (q', \sigma)$ for some $\sigma \in \Gamma^*$.

The number of such functions is finite, and bounded by $|Q| \cdot 2^{|Q|}$. These functions naturally define an equivalence relation of finite index over H_Σ :

$$h \sim h' \iff \text{rel}_h = \text{rel}_{h'}.$$

Let us note H for a subset containing one hedge per \sim -class. We have $|H| \leq |Q| \cdot 2^{|Q|}$. The next lemma indicates that the computation of h -predecessors can be limited to $h \in H$.

Lemma 45. For every $h \in H_\Sigma$, \mathcal{C} is a h -predecessor of \mathcal{C}' iff there exists $h' \in H, h \sim h'$, such that \mathcal{C} is a h' -predecessor of \mathcal{C}' .

Proof. Let us recall the definition of h -predecessor: \mathcal{C} is a h -predecessor of \mathcal{C}' if $\forall (q', \sigma) \in \mathcal{C}', \exists (q, \sigma) \in \mathcal{C}, (q, \sigma) \xrightarrow{h} (q', \sigma)$. Hence if $h \sim h'$ then \mathcal{C} is a h -predecessor of \mathcal{C}' iff \mathcal{C} is a h' -predecessor of \mathcal{C}' . \square

We propose an algorithm for computing such a set H from a VPA \mathcal{A} . Algorithm 7 is based on the definition of hedges, adapted to relations:

- ϵ is the empty hedge, and $\text{rel}_\epsilon(q) = \{q\}$ for every $q \in Q$. We write this function id_Q .
- if h_1, h_2 are two hedges, then $h_1 h_2$ is a hedge, and $\text{rel}_{h_1 h_2} = \text{rel}_{h_2} \circ \text{rel}_{h_1}$.
- if h is a hedge and $a \in \Sigma$, then $ah\bar{a}$ is a hedge, and $\text{rel}_{ah\bar{a}}(q)$ is the set of states q' such that there exists $\gamma \in \Gamma$ verifying:

$$(q, \epsilon) \xrightarrow{a} (p, \gamma) \quad \text{and} \quad (p', \gamma) \xrightarrow{\bar{a}} (q', \epsilon) \quad \text{with } p' \in \text{rel}_h(p).$$

Algorithm 7 uses the variables *ToProcess* and *Functions* with the following meaning. *Functions* contains initially the identity relation id_Q ; at the end of the computation, it contains all functions rel_h , for $h \in H_\Sigma$. *ToProcess* contains all the newly constructed relations, and these relations are used to create other new relations as described in the previous definition by induction.

Algorithm 7 Computing all functions rel_h , for $h \in H_\Sigma$.

```

function HEDGEFUNCTIONS( $\mathcal{A}$ )
  Functions  $\leftarrow \{\text{id}_Q\}$ 
  ToProcess  $\leftarrow \{\text{id}_Q\}$ 
  while ToProcess  $\neq \emptyset$  do
    fct  $\leftarrow \text{POP}(\textit{ToProcess})$ 
    NewFunctions  $\leftarrow \emptyset$ 
    for  $f \in \textit{Functions}$  do
      NewFunctions  $\leftarrow \textit{NewFunctions} \cup \{f \circ \textit{fct}, \textit{fct} \circ f\}$ 
    end for
    for  $a \in \Sigma$  do
       $f \leftarrow f_\emptyset$  //  $f_\emptyset$  maps every  $q \in Q$  to  $\emptyset$ 
      for  $q \xrightarrow{a:\gamma} p \in \Delta$  and  $p' \xrightarrow{\bar{a}:\gamma} q' \in \Delta$  with  $p' \in \textit{fct}(p)$  do
         $f(q) \leftarrow f(q) \cup \{q'\}$ 
      end for
      NewFunctions  $\leftarrow \textit{NewFunctions} \cup \{f\}$ 
    end for
    ToProcess  $\leftarrow \textit{ToProcess} \cup (\textit{NewFunctions} \setminus \textit{Functions})$ 
    Functions  $\leftarrow \textit{Functions} \cup \textit{NewFunctions}$ 
  end while
  return Functions
end function

```

Proposition 46. *Algorithm 7 computes the set $\{\text{rel}_h \mid h \in H_\Sigma\}$.*

Proof. Let *Functions* be the set computed by Algorithm 7. Clearly, *Functions* $\subseteq \{\text{rel}_h \mid h \in H_\Sigma\}$. Assume for contradiction that there exists $r = \text{rel}_h$ with $h \in H_\Sigma$ such that $r \notin \textit{Functions}$. Clearly, $r \neq \text{id}_Q$, and we can suppose wlog that either $r = r'_2 \circ r'_1$ with $r'_1, r'_2 \in \textit{Functions} \setminus \{\text{id}_Q\}$, or there exists $r' \in \textit{Functions}$ such

that for all q , $r(q)$ is the set of q' with $q \xrightarrow{a:\gamma} p \in \Delta$, $p' \xrightarrow{\bar{a}:\gamma} q' \in \Delta$ and $p' \in r'(p)$. Consider the first case. When they have been constructed by Algorithm 7, both r'_1 and r'_2 have been added to *ToProcess* and to *Functions*. After the last element (among r'_1 and r'_2) is popped from *ToProcess*, then $r = r'_2 \circ r'_1$ is built during the loop on $f \in \text{Functions}$, which leads to a contradiction. We also have a contradiction in the second case by considering the loop on $a \in \Sigma$. \square

Consequently we can rephrase our definition of $\lfloor \text{Safe}(ua) \rfloor$ from $\lfloor \text{LSafe}(ua) \rfloor$ given in Corollary 43 by restricting the quantification on h to the finite set H . Therefore we obtain a finite procedure for computing $\lfloor \text{Safe}(ua) \rfloor$ from $\lfloor \text{LSafe}(ua) \rfloor$:

Proposition 47.

$$\lfloor \text{Safe}(ua) \rfloor = \left[\left\{ \mathcal{C} \mid \mathcal{C} = \bigcup_{h \in H} \mathcal{C}_h \text{ with } \mathcal{C}_h \text{ a minimal } h\text{-predecessor of } \mathcal{C}' \in \lfloor \text{LSafe}(ua) \rfloor \right\} \right]$$

4.3.4 Starting point with antichains

It remains to explain how to compute $\text{Safe}(a)$. Clearly, by definition of H , we can compute $\lfloor \text{Safe}(a) \rfloor$ as follows:

Proposition 48.

$$\lfloor \text{Safe}(a) \rfloor = \left[\left\{ \mathcal{C} \mid \forall h \in H, \exists q_f \in Q_f, \exists (q, \sigma) \in \mathcal{C} : (q, \sigma) \xrightarrow{h\bar{a}} (q_f, \epsilon) \right\} \right].$$

4.4 Algorithmic improvements

The previous section resulted in a first algorithm to incrementally compute sets of safe configurations. This algorithm can be improved by limiting hedges to consider, and optimizing operators and predecessors to be computed. The goal here is to avoid the complexity of the on-the-fly determinization procedure.

4.4.1 Minimal hedges

A first improvement is obtained by further restricting hedges to consider. Indeed it suffices to consider *minimal hedges* wrt their function rel_h . Formally, let us write $h \leq h'$ whenever $\text{rel}_h(q) \subseteq \text{rel}_{h'}(q)$ for every $q \in Q$. We denote by $\lfloor H \rfloor$ the \leq -minimal elements of H . Notice that Algorithm 7 that computes the set $\{\text{rel}_h \mid h \in H\}$ can be easily adapted to compute the set of its minimal elements, such that *NewFunctions* and *ToProcess* are restricted to antichains of minimal elements.

From the definition of h -predecessor, for every $\mathcal{C}, \mathcal{C}' \in Q \times \Gamma^*$ we have:

$$\mathcal{C} \text{ } h\text{-predecessor of } \mathcal{C}' \text{ and } h \leq h' \implies \mathcal{C} \text{ } h'\text{-predecessor of } \mathcal{C}' \quad (5)$$

This property can be used to replace $h \in H$ in Proposition 47 by $h \in \lfloor H \rfloor$.

Proposition 49.

$$\lfloor \text{Safe}(ua) \rfloor = \left[\left\{ \mathcal{C} \mid \mathcal{C} = \bigcup_{h \in \lfloor H \rfloor} \mathcal{C}_h \text{ with } \mathcal{C}_h \text{ a minimal } h\text{-predecessor of } \mathcal{C}' \in \lfloor \text{LSafe}(ua) \rfloor \right\} \right]$$

Proof. Let S denote the set

$$\left\{ \mathcal{C} \mid \mathcal{C} = \bigcup_{h \in H} \mathcal{C}_h \text{ with } \mathcal{C}_h \text{ a minimal } h\text{-predecessor of } \mathcal{C}' \in \lfloor \text{LSafe}(ua) \rfloor \right\}$$

Let $\mathcal{C} \in S$. We have: $\mathcal{C} = \underbrace{\mathcal{C}_{h_1} \cup \dots \cup \mathcal{C}_{h_k}}_{h_i \in \lfloor H \rfloor} \cup \underbrace{\mathcal{C}_{h'_1} \cup \dots \cup \mathcal{C}_{h'_n}}_{h'_i \in H \setminus \lfloor H \rfloor}$. Let us show

that $\mathcal{C}_{h_1} \cup \dots \cup \mathcal{C}_{h_k} \cup \mathcal{C}_{h'_1} \cup \dots \cup \mathcal{C}_{h'_{n-1}} \in S$. By induction, this will prove that $\mathcal{C}_{h_1} \cup \dots \cup \mathcal{C}_{h_k} \in S$. We have $h'_n \in H \setminus \lfloor H \rfloor$, so there exists $h_i \in \lfloor H \rfloor$ such that $h_i \leq h'_n$. As \mathcal{C}_{h_i} is a minimal h_i -predecessor of an element \mathcal{C}' in $\lfloor \text{LSafe}(ua) \rfloor$, it follows from (5) that \mathcal{C}_{h_i} is also a minimal h'_n -predecessor of \mathcal{C}' . So $\mathcal{C}_{h_1} \cup \dots \cup \mathcal{C}_{h_k} \cup \mathcal{C}_{h'_1} \cup \dots \cup \mathcal{C}_{h'_{n-1}} \cup \mathcal{C}_{h_i} \in S$. \square

We have also the next proposition.

Proposition 50.

$$\lfloor \text{Safe}(a) \rfloor = \left[\left\{ \mathcal{C} \mid \forall h \in \lfloor H \rfloor, \exists q_f \in Q_f, \exists (q, \sigma) \in \mathcal{C} : (q, \sigma) \xrightarrow{h\bar{a}} (q_f, \epsilon) \right\} \right].$$

4.4.2 An appropriate union operator

Proposition 49 expresses that every set of configurations \mathcal{C} in $\lfloor \text{Safe}(ua) \rfloor$ is the union of \mathcal{C}_h with $h \in \lfloor H \rfloor$. We introduce a new operator to improve the readability and find new properties.

Definition 51. Let S be a finite set, and $A, B \in 2^{2^S \setminus \{\emptyset\}}$. The set $A \sqcup B \in 2^{2^S}$ is defined by:

$$A \sqcup B = \{a \cup b \mid a \in A \text{ and } b \in B\}$$

Operator \sqcup builds sets obtained by taking one set of each of its operands, and performing their union. It is obviously associative and commutative. Notice that the elements of A, B are supposed to be non-empty sets. This will always be the case in the following algorithms using this operator. Proposition 49 can now be rewritten as follows.

Proposition 52.

$$\lfloor \text{Safe}(ua) \rfloor = \left[\bigsqcup_{h \in \lfloor H \rfloor} \{ \mathcal{C}_h \mid \mathcal{C}_h \text{ is a minimal } h\text{-predecessor of } \mathcal{C}' \in \lfloor \text{LSafe}(ua) \rfloor \} \right]$$

When combined with operator $\lfloor \cdot \rfloor$, clauses of the \sqcup operator can be splitted, so that \sqcup is to be computed on smaller sets.

Lemma 53. $\lfloor A \sqcup B \rfloor = \lfloor (A \cap B) \cup (A \setminus B \sqcup B \setminus A) \rfloor$

Proof. (\supseteq) Let $\mathcal{C} \in \lfloor (A \cap B) \cup (A \setminus B \sqcup B \setminus A) \rfloor$. Then $\mathcal{C} \in A \sqcup B$. For contradiction, let us assume that there exists $\mathcal{C}' \subsetneq \mathcal{C}$ such that $\mathcal{C}' \in \lfloor A \sqcup B \rfloor$. If $\mathcal{C}' \in A \cap B$ then $\mathcal{C}' \in (A \cap B) \cup (A \setminus B \sqcup B \setminus A)$, which contradicts \mathcal{C} . So $\mathcal{C}' \notin A \cap B$, and assume wlog that $\mathcal{C}' = a \cup b$ with $a \in A \setminus B$ and $b \in B$. If $b \in A$ then $b \in A \cap B \subseteq A \sqcup B$ and $b \subsetneq \mathcal{C}'$, but this contradicts \mathcal{C}' . If $b \notin A$ then $\mathcal{C}' \in A \setminus B \sqcup B \setminus A$, so $\mathcal{C}' \in A \cap B \cup (A \setminus B \sqcup B \setminus A)$, and $\mathcal{C}' \subsetneq \mathcal{C}$, which contradicts \mathcal{C} .

(\subseteq) Let $\mathcal{C} \in \lfloor A \sqcup B \rfloor$. Let us first show that $\mathcal{C} \in (A \cap B) \cup (A \setminus B \sqcup B \setminus A)$. If $\mathcal{C} \in A \cap B$ this is direct. Otherwise $\mathcal{C} = a \cup b$ with $a \in A \setminus B$ and $b \in B$ (the other case is symmetric). If $b \in A$ then $b \in A \cap B \subseteq A \sqcup B$ and $b \subsetneq \mathcal{C}$, which contradicts the definition of \mathcal{C} . So $b \in B \setminus A$, and $\mathcal{C} \in A \setminus B \sqcup B \setminus A$. Now, assume for contradiction that there exists $\mathcal{C}' \subsetneq \mathcal{C}$ such that $\mathcal{C}' \in \lfloor (A \cap B) \cup (A \setminus B \sqcup B \setminus A) \rfloor$. Then, according to (\supseteq), $\mathcal{C}' \in \lfloor A \sqcup B \rfloor$, which contradicts the definition of \mathcal{C} . \square

Corollary 54. If $A \subseteq B$, then $\lfloor A \sqcup B \rfloor = \lfloor A \rfloor$.

The \sqcup operator also simplifies the definition of $\lfloor \text{Safe}(a) \rfloor$. From this new definition, an algorithm follows.

Proposition 55. $\lfloor \text{Safe}(a) \rfloor = \left\lfloor \bigsqcup_{h \in \lfloor H \rfloor} A_h \right\rfloor$ with

$$A_h = \left\{ \{(q, \sigma)\} \mid q \in Q, \sigma \in \Gamma : \exists q_f \in Q_f : (q, \sigma) \xrightarrow{h\bar{a}} (q_f, \epsilon) \right\}.$$

Proof. 1. Every element of $\bigsqcup_{h \in \lfloor H \rfloor} A_h$ belongs to $\text{Safe}(a)$. Thus $\left\lfloor \bigsqcup_{h \in \lfloor H \rfloor} A_h \right\rfloor \subseteq \lfloor \text{Safe}(a) \rfloor$.

2. Let us show that for each \mathcal{C} in $\text{Safe}(a)$, there exists $\mathcal{C}' \in \bigsqcup_{h \in \lfloor H \rfloor} A_h$ such that $\mathcal{C}' \subseteq \mathcal{C}$. Let $\mathcal{C} \in \text{Safe}(a)$. By definition, for all $h \in \lfloor H \rfloor$ there exists $(q_h, \sigma_h) \in \mathcal{C}$ and $q_f \in Q_f$ such that $(q_h, \sigma_h) \xrightarrow{h\bar{a}} (q_f, \epsilon)$. Let $\mathcal{C}' = \{(q_h, \sigma_h) \mid h \in \lfloor H \rfloor\}$. Then $\mathcal{C}' \subseteq \mathcal{C}$ and $\mathcal{C}' \in \bigsqcup_{h \in \lfloor H \rfloor} A_h$ because $\{(q_h, \sigma_h)\} \in A_h, \forall h$.

3. Assume that there exists $\mathcal{C}_* \in \left\lfloor \bigsqcup_{h \in \lfloor H \rfloor} A_h \right\rfloor \setminus \lfloor \text{Safe}(a) \rfloor$. By 1., there exists \mathcal{C} in $\lfloor \text{Safe}(a) \rfloor$ such that $\mathcal{C} \subsetneq \mathcal{C}_*$; and by 2., there exists $\mathcal{C}' \in \bigsqcup_{h \in \lfloor H \rfloor} A_h$ such that $\mathcal{C}' \subseteq \mathcal{C} \subsetneq \mathcal{C}_*$ in contradiction with the definition of \mathcal{C}_* . Therefore $\left\lfloor \bigsqcup_{h \in \lfloor H \rfloor} A_h \right\rfloor \subseteq \lfloor \text{Safe}(a) \rfloor$.

4. Let $\mathcal{C} \in \lfloor \text{Safe}(a) \rfloor$. By 2., there exists $\mathcal{C}' \in \left\lfloor \bigsqcup_{h \in \lfloor H \rfloor} A_h \right\rfloor$ such that $\mathcal{C}' \subseteq \mathcal{C}$. By 3., it follows that $\mathcal{C} = \mathcal{C}'$ and thus $\lfloor \text{Safe}(a) \rfloor \subseteq \left\lfloor \bigsqcup_{h \in \lfloor H \rfloor} A_h \right\rfloor$. \square

4.4.3 Using SAT solvers to find minimal predecessors

The computation of minimal predecessors is the key operation for Step 1 and Step 2 which respectively compute $\lfloor L\text{Safe}(ua) \rfloor$ from $\lfloor \text{Safe}(u) \rfloor$ and $\lfloor \text{Safe}(ua) \rfloor$ from $\lfloor L\text{Safe}(ua) \rfloor$ using the following formulas (see Corollary 41 and Proposition 52) :

$$\begin{aligned} \lfloor L\text{Safe}(ua) \rfloor &= \lfloor \{ \mathcal{C} \mid \mathcal{C} \text{ is a minimal } \bar{a}\text{-predecessor of } \mathcal{C}' \in \lfloor \text{Safe}(u) \rfloor \} \rfloor \\ \lfloor \text{Safe}(ua) \rfloor &= \left\lfloor \bigsqcup_{h \in \lfloor H \rfloor} \{ \mathcal{C}_h \mid \mathcal{C}_h \text{ is a minimal } h\text{-predecessor of } \mathcal{C}' \in \lfloor L\text{Safe}(ua) \rfloor \} \right\rfloor \end{aligned}$$

We propose a method to compute minimal predecessors by performing multiple calls to a SAT solver. A SAT solver is an algorithm used to efficiently test the satisfiability of a boolean formula φ , that is to check whether there exists a valuation v of the boolean variables of φ that makes φ true. In this case we say that v is a *model* of φ , denoted by $v \models \varphi$.

Most of the SAT solvers require that the boolean formula given as input is a conjunction of clauses (where a clause is a disjunction of literals, and a literal is a variable or its negation). Such formulas are said to be in conjunctive normal form (CNF). In the following all input formulas will be in CNF.

We first detail a method to compute all minimal \bar{a} -predecessor of \mathcal{C}' . It is also valid to compute all minimal h -predecessors of \mathcal{C}' .

Minimal predecessors. We recall that \mathcal{C} is a \bar{a} -predecessor of \mathcal{C}' if for all $(q', \sigma') \in \mathcal{C}'$, there exists $(q, \sigma) \in \mathcal{C}$ such that $(q, \sigma) \xrightarrow{\bar{a}} (q', \sigma')$. Let us write $\varphi_{\bar{a}}(\mathcal{C}')$ for the following boolean formula:

$$\varphi_{\bar{a}}(\mathcal{C}') = \bigwedge_{c' \in \mathcal{C}'} \bigvee_{c \xrightarrow{\bar{a}} c'} x_c,$$

and let $v_{\mathcal{C}}$ be the valuation such that $v_{\mathcal{C}}(x_c) = 1$ iff $c \in \mathcal{C}$. Then we immediately obtain that:

$$v_{\mathcal{C}} \models \varphi_{\bar{a}}(\mathcal{C}') \quad \text{iff} \quad \mathcal{C} \text{ is an } \bar{a}\text{-predecessor of } \mathcal{C}'$$

We define an ordering over valuations as follows, in a way to have a notion of minimal models equivalent to minimal predecessors. Let φ be a CNF boolean formula over the set V of boolean variables, let v and v' be two valuations over V . We define $v' \leq v$ iff for all variables $x \in V$, $v'(x) = 1 \implies v(x) = 1$. We denote $v' < v$ if $v' \leq v$ and $v' \neq v$. We say that a model v of φ is *minimal* if for all model v' of φ , we have $v' \leq v \implies v' = v$. We get the next characterization which also holds for h -predecessors.

Lemma 56. \mathcal{C} is a minimal \bar{a} -predecessor of \mathcal{C}' iff $v_{\mathcal{C}}$ is a minimal model of $\varphi_{\bar{a}}(\mathcal{C}')$.

We can now explain how to compute all the minimal \bar{a} -predecessors of \mathcal{C}' , or equivalently all the minimal models of formula $\varphi_{\bar{a}}(\mathcal{C}')$.

Let φ be a CNF boolean formula over V . First, we explain, knowing a model v of φ , how to compute a model v' of φ such that $v' < v$ (if it exists). Consider the next formula φ' :

$$\varphi' = \varphi \wedge \left(\bigwedge_{x \in V_0} \neg x \right) \wedge \left(\bigvee_{x \in V_1} \neg x \right)$$

where V_0 (respectively V_1) is the set of all variables $x \in V$ such that $v(x) = 0$ (resp. $v(x) = 1$). If φ' has a model v' , it follows from the definition of φ' that v' is a model of φ such that $v' < v$. Otherwise, v is a minimal model of φ . So from a model of φ we can compute a minimal model of φ by repeating the above procedure.

Second, let us explain how to compute all the minimal models of φ . Suppose that we already know some minimal model v of φ , and let V_1 be the set of variables $x \in V$ such that $v(x) = 1$. Consider the formula

$$\varphi' = \varphi \wedge \left(\bigvee_{x \in V_1} \neg x \right).$$

Then a model v' of φ' , if it exists, is a model of φ such that neither $v' < v$ (since v is minimal) nor $v < v'$ (by definition of φ'). With the previous procedure, we thus get a minimal model of φ that is distinct from v . In this way we can compute all minimal models of φ .

This approach has been detailed for minimal \bar{a} -predecessors. It also works for minimal h -predecessors.

Step 1 with SAT solvers. The computation of the set $\lfloor L\text{Safe}(ua) \rfloor$ from $\lfloor \text{Safe}(u) \rfloor$ can also be done using SAT solvers. Indeed, suppose that given $\mathcal{C}'_1 \in \lfloor \text{Safe}(u) \rfloor$, we have computed all the minimal \bar{a} -predecessors of \mathcal{C}'_1 as explained before. Let \mathcal{C}'_2 be another elements of $\lfloor \text{Safe}(u) \rfloor$. As done previously, we can express by boolean formulas, that we want to compute minimal \bar{a} -predecessor of \mathcal{C}'_2 that are either strictly included in some minimal \bar{a} -predecessor of \mathcal{C}'_1 , or incomparable with all minimal \bar{a} -predecessors of \mathcal{C}'_1 .

Step 2 with SAT solvers. The computation of the set $\lfloor \text{Safe}(ua) \rfloor$ from $\lfloor L\text{Safe}(ua) \rfloor$ can be done as in Proposition 52 by using operator \sqcup and exploiting its properties.

Under the hypothesis that $\epsilon \in \lfloor H \rfloor$, an alternative is possible with Proposition 49 stating that $\lfloor \text{Safe}(ua) \rfloor$ is equal to

$$\left\{ \mathcal{C} \mid \mathcal{C} = \bigcup_{h \in \lfloor H \rfloor} \mathcal{C}_h \text{ with } \mathcal{C}_h \text{ a minimal } h\text{-predecessor of } \mathcal{C}' \in \lfloor L\text{Safe}(ua) \rfloor \right\}$$

It is based on the following observations. Fix some \mathcal{C} and \mathcal{C}' in the previous equality. First, if $h = \epsilon$, then \mathcal{C}' is the only minimal h -predecessor of \mathcal{C}' and thus

$\mathcal{C}' \subseteq \mathcal{C}$. Second we know by the proof of Proposition 39 that $\mathcal{C} \subseteq Q \times \Gamma^{|u'|}$ where $u' = \text{open}(u)$. Therefore, instead of computing \mathcal{C} as a union $\bigcup_{h \in [H]} \mathcal{C}_h$, we can compute it starting from \mathcal{C}' and adding elements of $Q \times \Gamma^{|u'|}$ one by one, until we get an element \mathcal{C} of $\text{Safe}(ua)$. By the way it is constructed, $\mathcal{C} \in \lfloor \text{Safe}(ua) \rfloor$. We can check that such an element belongs to $\text{Safe}(ua)$ with Proposition 34 by testing for all $h \in [H]$, whether there exists $\mathcal{C}'' \in \lfloor \text{LSafe}(ua) \rfloor$ such that $\text{Post}_{[h]}(\mathcal{C}) \supseteq \mathcal{C}''$. To get the whole set $\lfloor \text{Safe}(ua) \rfloor$, we need to consider all the possibilities to enlarge \mathcal{C}' with elements of $Q \times \Gamma^{|u'|}$. This task can be done efficiently with the help of SAT solvers (with ideas similar to the ones developed above).

References

- [ACH⁺10] Parosh Aziz Abdulla, Yu-Fang Chen, Lukás Holík, Richard Mayr, and Tomás Vojnar, *When simulation meets antichains*, 16th International Conference on Tools and Algorithms for the Construction and Analysis of Systems, Lecture Notes in Computer Science, vol. 6015, Springer, 2010, pp. 158–174.
- [AM04] Rajeev Alur and P. Madhusudan, *Visibly pushdown languages*, 36th ACM Symposium on Theory of Computing, ACM-Press, 2004, pp. 202–211.
- [AM09] ———, *Adding nesting structure to words*, Journal of the ACM **56** (2009), no. 3, 1–43.
- [BHH⁺08] Ahmed Bouajjani, Peter Habermehl, Lukás Holík, Tayssir Touili, and Tomás Vojnar, *Antichain-based universality and inclusion testing over nondeterministic finite tree automata*, CIAA (Oscar H. Ibarra and Bala Ravikumar, eds.), Lecture Notes in Computer Science, vol. 5148, Springer, 2008, pp. 57–67.
- [BJLW08] Michael Benedikt, Alan Jeffrey, and Ruy Ley-Wild, *Stream Firewalling of XML Constraints*, ACM SIGMOD International Conference on Management of Data, ACM-Press, 2008, pp. 487–498.
- [BKMW01] Anne Brüggemann-Klein, Makoto Murata, and Derick Wood, *Regular tree and regular hedge languages over unranked alphabets: Version 1*, Tech. Report HKTUST-TCSC-2001-05, HKUST Theoretical Computer Science Center Research, 2001.
- [CDG⁺07] H. Comon, M. Dauchet, R. Gilleron, C. Löding, F. Jacquemard, D. Lugiez, S. Tison, and M. Tommasi, *Tree automata techniques and applications*, Available on: <http://www.grappa.univ-lille3.fr/tata>, 2007, release October, 12th 2007.

- [DR10] Laurent Doyen and Jean-François Raskin, *Antichain algorithms for finite automata*, 16th International Conference on Tools and Algorithms for the Construction and Analysis of Systems, Lecture Notes in Computer Science, vol. 6015, Springer, 2010, pp. 2–22.
- [DWDHR06] M. De Wulf, L. Doyen, T. Henzinger, and J. Raskin, *Antichains: A new algorithm for checking universality of finite automata*, Computer Aided Verification (Thomas Ball and Robert Jones, eds.), Lecture Notes in Computer Science, vol. 4144, Springer Berlin / Heidelberg, 2006, pp. 17–30.
- [EHR00] Javier Esparza, David Hansel, Peter Rossmanith, and Stefan Schwoon, *Efficient algorithms for model checking pushdown systems*, CAV (E. Allen Emerson and A. Prasad Sistla, eds.), Lecture Notes in Computer Science, vol. 1855, Springer, 2000, pp. 232–247.
- [EKS03] Javier Esparza, Antonín Kucera, and Stefan Schwoon, *Model checking ltl with regular valuations for pushdown systems*, Inf. Comput. **186** (2003), no. 2, 355–376.
- [FDL11] Nadime Francis, Claire David, and Leonid Libkin, *A Direct Translation from XPath to Nondeterministic Automata*, 5th Alberto Mendelzon International Workshop on Foundations of Data Management, 2011.
- [GN11] Olivier Gauwin and Joachim Niehren, *Streamable fragments of forward XPath*, CIAA (Béatrice B. Markhoff, Pascal Caron, Jean M. Champarnaud, Denis Maurel, Béatrice B. Markhoff, Pascal Caron, Jean M. Champarnaud, and Denis Maurel, eds.), Lecture Notes in Computer Science, vol. 6807, Springer, 2011, pp. 3–15.
- [GNR08] Olivier Gauwin, Joachim Niehren, and Yves Roos, *Streaming tree automata*, Information Processing Letters **109** (2008), no. 1, 13–17.
- [GNT09] Olivier Gauwin, Joachim Niehren, and Sophie Tison, *Earliest query answering for deterministic nested word automata*, 17th International Symposium on Fundamentals of Computer Theory, Lecture Notes in Computer Science, vol. 5699, Springer Verlag, 2009, pp. 121–132.
- [KMV07] Viraj Kumar, P. Madhusudan, and Mahesh Viswanathan, *Visibly pushdown automata for streaming XML*, 16th international conference on World Wide Web, ACM-Press, 2007, pp. 1053–1062.
- [KV01] Orna Kupferman and Moshe Y. Vardi, *Model checking of safety properties*, Formal Methods in System Design **19** (2001), no. 3, 291–314.

- [MV09] P. Madhusudan and Mahesh Viswanathan, *Query automata for nested words*, 34th International Symposium on Mathematical Foundations of Computer Science, Lecture Notes in Computer Science, vol. 5734, Springer Verlag, 2009, pp. 561–573.
- [Ngu09] Tang Van Nguyen, *A tighter bound for the determinization of visibly pushdown automata*, INFINITY, EPTCS, vol. 10, 2009, pp. 62–76.
- [NO12] Tang Van Nguyen and Hitoshi Ohsaki, *On model checking for visibly pushdown automata*, Language and Automata Theory and Applications (Adrian-Horia Dediu and Carlos Martín-Vide, eds.), Lecture Notes in Computer Science, vol. 7183, Springer Berlin / Heidelberg, 2012, pp. 408–419.
- [Srb06] Jiri Srba, *Visibly pushdown automata: From language equivalence to simulation and bisimulation*, Computer Science Logic (Zoltan Sik, ed.), Lecture Notes in Computer Science, vol. 4207, Springer Berlin / Heidelberg, 2006, pp. 89–103.