Fachbereich Informatik der Universität Hamburg

Vogt-Kölln-Str. 30 ◊ D-22527 Hamburg / Germany

University of Hamburg - Computer Science Department

Mitteilung Nr. 302/01 • Memo No. 302/01

Undecidability of $\mathcal{ALC}_{\mathcal{RA}}$

Michael Wessel

Arbeitsbereich KOGS

FBI-HH-M-302/01

February 2001

Undecidability of $\mathcal{ALC}_{\mathcal{RA}}$

Michael Wessel

University of Hamburg, Computer Science Department,
Vogt-Kölln-Str. 30, 22527 Hamburg, Germany

Abstract

This paper answers the question whether the logic $\mathcal{ALC}_{\mathcal{RA}}$ might be decidable – which was left open in [4], [3], [2] – to the negative.

1 Introduction and Motivation

This paper answers a question which was left open in our previous work. We are investigating the extension of the standard description logic $\mathcal{ALC}$ with composition-based role inclusion axioms of the form $S \circ T \sqsubseteq R_1 \sqcup \cdots \sqcup R_n$. Please refer to our previous work for a discussion of the considered description logics ([4], [3], and [2]), a discussion of related work, other descriptions logics etc.

Even though we know since [3] that the logic $\mathcal{ALC}_{\mathcal{RA}^{\ominus}}$ is undecidable it was not immediate to adopt the $\mathcal{ALC}_{\mathcal{RA}^{\ominus}}$ undecidability-proof to $\mathcal{ALC}_{\mathcal{RA}}$, since $\mathcal{ALC}_{\mathcal{RA}}$ and $\mathcal{ALC}_{\mathcal{RA}^{\ominus}}$ differ in one fundamental aspect: in contrast to $\mathcal{ALC}_{\mathcal{RA}^{\ominus}}$, $\mathcal{ALC}_{\mathcal{RA}}$ enforces global *role disjointness* (see our previous work for a more thorough discussion). Otherwise the logics are identical.

The structure of this paper is as follows: we first define the syntax and semantics of $\mathcal{ALC}_{\mathcal{RA}}$, then prove its undecidability by giving a reduction from *Post's Correspondence Problem (PCP)*, which is the main contribution of this paper.

2 Syntax and Semantics of $\mathcal{ALC}_{\mathcal{RA}}$

In the following we will define the syntax and semantics of the logic $\mathcal{ALC}_{\mathcal{RA}}$. We start with the set of well-formed concept expressions (concepts for short):

Definition 1 (Concept Expressions) Let $\mathcal{N_C}$ be a set of concept names, and let $\mathcal{N_R}$ be a set of role names (roles for short), such that $\mathcal{N_C} \cap \mathcal{N_R} = \emptyset$. The set of concept expressions (or concepts for short) is defined inductively:

1. Every concept name $C \in \mathcal{N_C}$ is a concept.
2. If C and D are concepts, and $R \in \mathcal{N_R}$ is a role, then the following expressions are concepts as well: $(\neg C)$, $(C \sqcap D)$, $(C \sqcup D)$, $(\exists R.C)$, and $(\forall R.C)$.

The set of concepts is the same as for the language $\mathcal{ALC}$. If a concept starts with "(", we call it a compound concept, otherwise a concept name or atomic concept. Brackets may be omitted for the sake of readability if the concept is still uniquely parsable.

We use the following abbreviations: if $R_1, \ldots, R_n$ are roles, and C is a concept, then we define $(\forall R_1 \sqcup \cdots \sqcup R_n.C) =_{def} (\forall R_1.C) \sqcap \cdots \sqcap (\forall R_n.C)$ and $\exists R_1 \sqcup \cdots \sqcup R_n.C =_{def} (\exists R_1.C) \sqcup \cdots \sqcup (\exists R_n.C)$. Additionally, for some $CN \in \mathcal{N_C}$ we define $\top =_{def} CN \sqcup \neg CN$ and $\bot =_{def} CN \sqcap \neg CN$ (therefore, $\top^\mathcal{I} = \Delta^\mathcal{I}$, $\bot^\mathcal{I} = \emptyset$).

Before we can proceed, we need some auxiliary definitions. The set of *roles* being used within a concept C is defined:

Definition 2 (Used Roles, $\mathsf{roles}(C)$)

$$\mathsf{roles}(C) =_{def} \begin{cases} \emptyset & \text{if } C \in \mathcal{N_C} \\ \mathsf{roles}(D) & \text{if } C = (\neg D) \\ \mathsf{roles}(D) \cup \mathsf{roles}(E) & \text{if } C = (D \sqcap E) \\ & \text{or } C = (D \sqcup E) \\ \{R\} \cup \mathsf{roles}(D) & \text{if } C = (\exists R.D) \\ & \text{or } C = (\forall R.D) \end{cases}$$

For example, $\mathsf{roles}(((\forall R.(\exists S.C)) \sqcap \exists T.D)) = \{R, S, T\}$.

As already noted, we are investigating the satisfiability of $\mathcal{ALC}$ concepts w.r.t. a set of role axioms of the form $S \circ T \sqsubseteq R_1 \sqcup \cdots \sqcup R_n$. More formally, the syntax of these role axioms and of the considered role boxes containing these axioms is as follows:

Definition 3 (Role Axioms, Role Box, Admissible Role Box) If $S, T, R_1, \ldots, R_n \in \mathcal{N_R}$, then the expression $S \circ T \sqsubseteq R_1 \sqcup \cdots \sqcup R_n$, $n \geq 1$, is called a *role axiom*. If $ra = S \circ T \sqsubseteq R_1 \sqcup \cdots \sqcup R_n$, then $\mathsf{pre}(ra) =_{def} (S, T)$ and $\mathsf{con}(ra) =_{def} \{R_1, \ldots, R_n\}$.

If $n = 1$, then ra is called a *deterministic* role axiom. In this case we also write $T = \mathsf{con}(ra)$ instead of $T \in \mathsf{con}(ra)$.

A finite set $\mathfrak{R}$ of role axioms is called a *role box*.

Let $\mathsf{roles}(ra) =_{def} \{S, T, R_1, \ldots, R_n\}$, and $\mathsf{roles}(\mathfrak{R}) =_{def} \bigcup_{ra \in \mathfrak{R}} \mathsf{roles}(ra)$.

The *semantics* of a concept is specified by giving a Tarski-style interpretation $\mathcal{I}$ that has to satisfy the following conditions:

Definition 4 (Interpretation) An *interpretation* $\mathcal{I} =_{def} (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a non-empty set $\Delta^{\mathcal{I}}$, called the domain of $\mathcal{I}$, and an interpretation function $\cdot^{\mathcal{I}}$ that maps every concept name to a subset of $\Delta^{\mathcal{I}}$, and every role name to a subset of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.

Additionally, for all roles $R, S \in \mathcal{N_R}$, $R \neq S$: $R^{\mathcal{I}} \cap S^{\mathcal{I}} = \emptyset$. All roles are interpreted disjointly then.

The following functions on $\mathcal{I}$ will be used: The *universal relation* of $\mathcal{I}$ is defined as $\mathcal{UR}(\mathcal{I}) =_{def} \bigcup_{R \in \mathcal{N_R}} R^{\mathcal{I}}$, and the universal relation w.r.t. a set of role names $\mathcal{R}$ as $\mathcal{UR}(\mathcal{I}, \mathcal{R}) =_{def} \bigcup_{R \in \mathcal{R}} R^{\mathcal{I}}$.

If $\langle i, j \rangle \in \mathcal{UR}(\mathcal{I})$, the edge is called an *incoming edge* for j.

Given an interpretation $\mathcal{I}$, every (possibly compound) concept C can be uniquely interpreted ("evaluated") by using the following definitions (we write $X^{\mathcal{I}}$ instead of $\cdot^{\mathcal{I}}(X)$):

$$
\begin{array}{rcl}
(\neg C)^{\mathcal{I}} & =_{def} & \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}} \\
(C \sqcap D)^{\mathcal{I}} & =_{def} & C^{\mathcal{I}} \cap D^{\mathcal{I}} \\
(C \sqcup D)^{\mathcal{I}} & =_{def} & C^{\mathcal{I}} \cup D^{\mathcal{I}} \\
(\exists R.C)^{\mathcal{I}} & =_{def} & \{\, i \in \Delta^{\mathcal{I}} \mid \exists j \in C^{\mathcal{I}} : \langle i, j \rangle \in R^{\mathcal{I}} \,\} \\
(\forall R.C)^{\mathcal{I}} & =_{def} & \{\, i \in \Delta^{\mathcal{I}} \mid \forall j : \langle i, j \rangle \in R^{\mathcal{I}} \Rightarrow j \in C^{\mathcal{I}} \,\}
\end{array}
$$

It is therefore sufficient to provide the interpretations for the concept *names* and roles, since the extension $C^{\mathcal{I}}$ of every concept C is uniquely determined then.

In the following we specify under which conditions a given interpretation is a *model* of a syntactic entity (we also say an interpretation *satisfies* a syntactic entity):

Definition 5 (The Model Relationship) An interpretation $\mathcal{I}$ is a model of a concept C, written $\mathcal{I} \models C$, iff $C^{\mathcal{I}} \neq \emptyset$.

An interpretation $\mathcal{I}$ is a model of a role axiom $S \circ T \sqsubseteq R_1 \sqcup \cdots \sqcup R_n$, written $\mathcal{I} \models S \circ T \sqsubseteq R_1 \sqcup \cdots \sqcup R_n$, iff $S^{\mathcal{I}} \circ T^{\mathcal{I}} \subseteq R_1^{\mathcal{I}} \cup \cdots \cup R_n^{\mathcal{I}}$.

An interpretation $\mathcal{I}$ is a model of a role box $\mathfrak{R}$, written $\mathcal{I} \models \mathfrak{R}$, iff for all role axioms $ra \in \mathfrak{R}$: $\mathcal{I} \models ra$.

An interpretation $\mathcal{I}$ is a model of $(C, \mathfrak{R})$, written $\mathcal{I} \models (C, \mathfrak{R})$, iff $\mathcal{I} \models C$ and $\mathcal{I} \models \mathfrak{R}$.

Definition 6 (Satisfiability) A syntactic entity (concept, role box, concept with role box, etc.) is called *satisfiable* iff there is an interpretation which satisfies this entity; i.e., the entity has a model.

Then, the *satisfiability problem* is to decide whether a syntactic entity is satisfiable or not.

In order to demonstrate the *consequences of disjointness for roles,* please consider $\mathfrak{R} = \{R \circ S \sqsubseteq A \sqcup B, S \circ T \sqsubseteq X \sqcup Y, A \circ T \sqsubseteq U, B \circ T \sqsubseteq V, R \circ X \sqsubseteq U, R \circ Y \sqsubseteq V\}$. Then, $(\exists R.((\exists S.\exists T.\top) \sqcap \forall Y.\bot) \sqcap \forall A.\bot, \mathfrak{R}(C))$ is unsatisfiable, since $\forall A.\bot$ forces to choose $B \in \mathsf{con}(R, S)$, and $\forall Y.\bot$ forces to choose $X \in \mathsf{con}(S, T)$. Due to $B \circ T \sqsubseteq V$ and $R \circ X \sqsubseteq U$ there must be a non-empty intersection between U and V. The unsatisfiability is caused by a subtle interplay between the role box and the concept.

An important relationship between concepts is the subsumption relationship, which is a partial ordering on concepts w.r.t. their specificity:

Definition 7 (Subsumption Relationship) A concept D *subsumes* a concept C, $C \sqsubseteq D$ (w.r.t. to $\mathfrak{R}$), iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ holds for all interpretations $\mathcal{I}$ (that are also models $\mathfrak{R}$).

Since a full negation operator if provided, the subsumption problem can be reduced to the concept satisfiability problem: $C \sqsubseteq D$ iff $C \sqcap \neg D$ is unsatisfiable.

Proposition 1 $\mathcal{ALC}_{\mathcal{RA}}$ does not have the *finite model property*, i.e. there are pairs $(C, \mathfrak{R})$ that have no finite models.

Proof 1 As a counter-example to a finite model property assumption in $\mathcal{ALC}_{\mathcal{RA}}$, please consider $(\exists R.\exists R.\top) \sqcap (\forall S.\exists R.\top)$ w.r.t. $\{R \circ R \sqsubseteq S, R \circ S \sqsubseteq S, S \circ R \sqsubseteq S, S \circ S \sqsubseteq S\}$, which has no finite model (see [4],[3] for a proof). $\square$

Intuitively, the disjointness requirement ensures that the given example can only be fulfilled by infinite models: whenever one tries to create a finite model, it must by cyclical. But then, the presence of the cycle invalidates the model, since the role axioms will enforce a non-empty intersection between the roles R and S $(R^{\mathcal{I}} \cap S^{\mathcal{I}} \neq \emptyset)$.

3 Proving Undecidability of $\mathcal{ALC}_{\mathcal{RA}}$

Basically, the proof is by means of a reduction from *Post's Correspondence Problem (PCP)*. It is well-known that checking whether a given PCP of sufficiently large size has a solution is an undecidable problem.

In order to make the proof more transparent and comprehensible (since the reduction is rather technical, see below) we use terminology from formal language theory. The structure of the proof is as follows: Given a PCP K, we define two corresponding context-free grammars $\mathcal{G}_{1,K}$ and $\mathcal{G}_{2,K}$ such that $\mathcal{L}(\mathcal{G}_{1,K}) \cap \mathcal{L}(\mathcal{G}_{2,K}) \neq \emptyset$ iff the PCP K has a solution. The grammars $\mathcal{G}_{1,K}$ and $\mathcal{G}_{2,K}$ are transformed into the grammars $\mathcal{G}'_{1,K}$ and $\mathcal{G}'_{2,K}$, that have the following properties: $w_1 w_2 \ldots w_n \in \mathcal{L}(\mathcal{G}_{1,K})$ iff $w_1\#w_2\#\ldots w_n\# \in \mathcal{L}(\mathcal{G}_{1,K})$, and $w_1 w_2 \ldots w_n \in \mathcal{L}(\mathcal{G}_{2,K})$ iff $\#w_1\#w_2 \ldots \#w_n \in \mathcal{L}(\mathcal{G}'_{2,K})$; i.e., they differ from their original versions with respect to an "odd"- resp. "even"-interleaving of the new symbol "#". By construction of these grammars we obviously have $\mathcal{L}(\mathcal{G}'_{1,K}) \cap \mathcal{L}(\mathcal{G}'_{2,K}) = \emptyset$. It will subsequently become clear why this property is of utmost importance. The reader should bear in mind that this complicated construction is necessary in order to ensure that the disjointness requirement of $\mathcal{ALC}_{\mathcal{RA}}$ does not become violated. Moreover, the original PCP K has a solution iff $\{\#\}\mathcal{L}(\mathcal{G}'_{1,K}) \cap \mathcal{L}(\mathcal{G}'_{2,K})\{\#\} \neq \emptyset$.[1] We will then define a role box $\mathfrak{R}_{\mathfrak{K}}$ and a concept term E such that $(E, \mathfrak{R}_{\mathfrak{K}})$ is satisfiable iff $\mathcal{L}_{\mathcal{K}} = \emptyset$. Basically, $\mathfrak{R}_{\mathfrak{K}}$ is constructed by *reversing* the productions of the grammars $\mathcal{G}'_{1,K}$, $\mathcal{G}'_{2,K}$. The construction of $\mathcal{G}'_{1,K}$ and $\mathcal{G}'_{2,K}$ is rather technical, since a certain "normal form" of the productions must be achieved in order to be able to reverse them into syntactically well-formed role axioms. This normal form is quite similar to the well-known Chomsky Normal Form. Since the emptiness problem for $\mathcal{L}_{\mathcal{K}}$ is undecidable, satisfiability for $(E, \mathfrak{R}_{\mathfrak{K}})$ is as well.

We start with some basic definitions:

Definition 8 (Context-Free Grammar, Language) A context-free grammar $\mathcal{G}$ is a quadruple $(\mathcal{V}, \Sigma, \mathcal{P}, S)$, where $\mathcal{V}$ is a finite set of variables or non-terminal symbols, Σ is finite alphabet of terminal symbols with $\mathcal{V} \cap \Sigma = \emptyset$, and $\mathcal{P} \subseteq \mathcal{V} \times (\mathcal{V} \cup \Sigma)^+$ is a set of productions or grammar rules. $S \in \mathcal{V}$ is the start variable. The *language* generated by a context-free grammar $\mathcal{G}$ is defined as $\mathcal{L}(\mathcal{G}) = \{\, w \mid w \in \Sigma^+, S \xrightarrow{\star} w \,\}$ (see [1]). In the following, we will only consider languages with $\epsilon \notin \mathcal{L}(\mathcal{G})$ – ϵ is the *empty word* – and we can therefore write $\mathcal{L}(\mathcal{G}) = \{\, w \mid w \in \Sigma^+, S \xrightarrow{+} w \,\}$.

[1] The expression $\{\#\}\mathcal{L}(\mathcal{G}'_{1,K}) \cap \mathcal{L}(\mathcal{G}'_{2,K})\{\#\}$ denotes the language $\mathcal{L}_K =_{def} \{\, \#\alpha\# \mid \alpha\# \in \mathcal{L}(\mathcal{G}'_{1,K}), \#\alpha \in \mathcal{L}(\mathcal{G}'_{2,K}) \,\}$

Definition 9 (PCP) A *Post's Correspondence Problem (PCP) K* over an alphabet $\mathcal{A}$ is given by a finite set of pairs $K = \{(x_1, y_1), (x_2, y_2), \dots (x_k, y_k)\}$, where x_i, y_i are (non-empty!) words over a given alphabet $\mathcal{A}$: $x_i, y_i \in \mathcal{A}^+$. A *solution* to a PCP is sequence of indices $(i_1, i_2, \dots, i_n) \in \{1 \dots k\}$ with $n \geq 1$ such that $x_{i_1} x_{i_2} \dots x_{i_n} = y_{i_1} y_{i_2} \dots y_{i_n}$ (see [1]).

For example, the PCP $K = \{(1, 101), (10, 00), (011, 11)\}$ has the solution (1,3,2,3), since $x_1 x_3 x_2 x_3 = \boxed{1}\boxed{011}\boxed{10}\boxed{011} = 101110011 = \boxed{101}\boxed{11}\boxed{00}\boxed{11} = y_1 y_3 y_2 y_3$ (the example is taken from [1]).

Lemma 1 (PCP Undecidable) It is undecidable whether a given PCP with $|\mathcal{A}| \geq 2$ and $k \geq 9$ has a solution (see [1]).

In the following it suffices to consider (sufficiently large) PCPs with $|\mathcal{A}| = 2$. Whatever $\mathcal{A}$ is, we name its elements by a_1 and a_2: $\mathcal{A} = \{a_1, a_2\}$.

Definition 10 (Auxiliary Definitions) Let $x \in \mathcal{A}^+$, $x = a_1 \dots a_n$. We define $|x| =_{def} n$, $\mathsf{first}(x) =_{def} a_1$, and $\mathsf{rest}(x) =_{def} a_2 \dots a_n$. Let $\mathsf{postfixes}(x) =_{def} \{\, w \mid \exists v \in \mathcal{A}^\star : x = vw, w \neq \epsilon \,\}$ (e.g. $\mathsf{postfixes}(1011) = \{1011, 011, 11, 1\}$). Additionally, $\mathsf{even}_\#(a_1 \dots a_n) =_{def} a_1\# \dots a_n\#$, and $\mathsf{odd}_\#(a_1 \dots a_n) =_{def} \#a_1 \dots \#a_n$ (e.g. $\mathsf{even}_\#(abc) = a\#b\#c\#$, $\mathsf{odd}_\#(abc) = \#a\#b\#c$).

Let $K = \{(x_1, y_1), (x_2, y_2), \dots (x_k, y_k)\}$ be a PCP over the alphabet $\mathcal{A}$. It is well-known from formal language theory that the emptiness problem for intersections of context-free languages is undecidable. Given a PCP K, we can define two grammars $\mathcal{G}_{1,K}$ and $\mathcal{G}_{2,K}$ such that K has the solution $(i_1, \dots, i_n)$ iff $i_{i_n} \dots i_{i_2} i_{i_1} x_{i_1} x_{i_2} \dots x_{i_n} \in \mathcal{L}(\mathcal{G}_{1,K}) \cap \mathcal{L}(\mathcal{G}_{2,K})$ iff $i_{i_n} \dots i_{i_2} i_{i_1} y_{i_1} y_{i_2} \dots y_{i_n} \in \mathcal{L}(\mathcal{G}_{1,K}) \cap \mathcal{L}(\mathcal{G}_{2,K})$. Please note that the PCP solution $(i_1, \dots, i_n)$ appears *reversed* in the word.

Let $\mathcal{A}' = \mathcal{A} \cup \{i_1, \dots, i_k\}$. Then, the context-free grammars $\mathcal{G}_{1,K}$ and $\mathcal{G}_{2,K}$ are defined as follows (see also [1]):

- $\mathcal{G}_{1,K} = (\{S_1\}, \mathcal{A}', \mathcal{P}_1, S_1)$, where
  $\mathcal{P}_1 = \{S_1 \to i_1 x_1 \mid \dots \mid i_k x_k\} \cup \{S_1 \to i_1 S_1 x_1 \mid \dots \mid i_k S_1 x_k\}$, and
- $\mathcal{G}_{2,K} = (\{S_2\}, \mathcal{A}', \mathcal{P}_2, S_2)$, where
  $\mathcal{P}_2 = \{S_2 \to i_1 y_1 \mid \dots \mid i_k y_k\} \cup \{S_2 \to i_1 S_2 y_1 \mid \dots \mid i_k S_2 y_k\}$.

It is interesting to note that these grammars are *deterministic* – each word of the generated languages has one unique parse tree. Applied to the example PCP $K = \{(1, 101), (10, 00), (011, 11)\}$ we get the two context-free grammars

- $\mathcal{G}_{1,K} = (\{S_1\}, \{0,1\} \cup \{i_1, i_2, i_3\}, \mathcal{P}_1, S_1)$, where
  $\mathcal{P}_1 = \{S_1 \to i_1 1 \mid i_2 10 \mid i_3 011\} \cup \{S_1 \to i_1 S_1 1 \mid i_2 S_1 10 \mid i_3 S_1 011\}$, and
- $\mathcal{G}_{2,K} = (\{S_2\}, \{0,1\} \cup \{i_1, i_2, i_3\}, \mathcal{P}_2, S_2)$, where
  $\mathcal{P}_2 = \{S_2 \to i_1 101 \mid i_2 00 \mid i_3 11\} \cup \{S_2 \to i_1 S_2 101 \mid i_2 S_2 00 \mid i_3 S_2 11\}$.

It is easy to verify that $i_3 i_2 i_3 i_1 101110011 \in \mathcal{L}(\mathcal{G}_{1,K}) \cap \mathcal{L}(\mathcal{G}_{2,K})$, since $(1,3,2,3)$ is a solution to K.

We already mentioned that we are aiming at a grammar whose productions can be "reversed" in order to get a valid role box. In a second step we therefore transform $\mathcal{G}_{1,K}$ and $\mathcal{G}_{2,K}$ and get the grammars $\mathcal{G}'_{1,K}$ and $\mathcal{G}'_{2,K}$, which have the "odd and even interleaving-property" (see above). These grammars are defined as follows:

- $\mathcal{G}'_{1,K} = (\mathcal{V}'_1, \mathcal{A}' \cup \{\#\}, \mathcal{P}'_1, S_1)$

$$
\begin{array}{rl}
\mathcal{V}'_1 = & \{S_1\} \cup \\
& \{\boxed{a\#} \mid a \in \mathcal{A}'\} \cup \\
& \{\boxed{w\#} \mid x \in \{x_1, \ldots, x_k\}, \\
& \qquad w \in \mathsf{postfixes}(x)\} \cup \\
& \{\boxed{S_1 x\#} \mid x \in \{x_1, \ldots, x_k\}\}
\end{array}
$$

$$
\begin{array}{rl}
\mathcal{P}'_1 = & \{\boxed{a\#} \to a\# \mid a \in \mathcal{A}'\} \cup \\
& \{S_1 \to \boxed{i_1\#}\ \boxed{x_1\#} \mid \cdots \mid \boxed{i_k\#}\ \boxed{x_k\#}\} \cup \\
& \{S_1 \to \boxed{i_1\#}\ \boxed{S_1 x_1\#} \mid \cdots \mid \boxed{i_k\#}\ \boxed{S_1 x_k\#}\} \cup \\
& \{\boxed{S_1 x_1\#} \to S_1\ \boxed{x_1\#}, \ldots, \boxed{S_1 x_k\#} \to S_1\ \boxed{x_k\#}\} \cup \\
& \{\boxed{x\#} \to \boxed{\mathsf{first}(x)\#}\ \boxed{\mathsf{rest}(x)\#} \mid n \in 1 \ldots k, \\
& \qquad x \in \mathsf{postfixes}(x_n), |x| \geq 2\}
\end{array}
$$

- $\mathcal{G}'_{2,K} = (\mathcal{V}'_2, \mathcal{A}' \cup \{\#\}, \mathcal{P}'_2, S_2)$

$$
\begin{array}{rl}
\mathcal{V}'_2 = & \{S_2\} \cup \\
& \{\boxed{\#a} \mid a \in \mathcal{A}'\} \cup \\
& \{\boxed{\#w} \mid y \in \{y_1, \ldots, y_k\}, \\
& \qquad w \in \mathsf{postfixes}(y)\} \cup \\
& \{\boxed{S_2\#y} \mid y \in \{y_1, \ldots, y_k\}\}
\end{array}
$$

$$
\begin{array}{rl}
\mathcal{P}'_2 = & \{\boxed{\#a} \to \#a \mid a \in \mathcal{A}'\} \cup \\
& \{S_2 \to \boxed{\#i_1}\ \boxed{\#y_1} \mid \cdots \mid \boxed{\#i_k}\ \boxed{\#y_k}\} \cup \\
& \{S_2 \to \boxed{\#i_1}\ \boxed{S_2\#y_1} \mid \cdots \mid \boxed{\#i_k}\ \boxed{S_2\#y_k}\} \cup \\
& \{\boxed{S_2\#y_1} \to S_2\ \boxed{\#y_1}, \ldots, \boxed{S_2\#y_k} \to S_2\ \boxed{\#y_k}\} \cup \\
& \{\boxed{\#y} \to \boxed{\#\mathsf{first}(y)}\ \boxed{\#\mathsf{rest}(y)} \mid n \in 1 \ldots k, \\
& \qquad y \in \mathsf{postfixes}(y_n), |y| \geq 2\}
\end{array}
$$

If we write expressions like "$\boxed{\#y} \rightarrow \boxed{\#\mathsf{first}(y)}\,\boxed{\#\mathsf{rest}(y)}$" and for example, $y = 101$, then this construction denotes the production "$\boxed{\#101} \rightarrow \boxed{\#1}\,\boxed{\#01}$", since $\mathsf{first}(y) = 1$ and $\mathsf{rest}(y) = 01$. What happens if for some $i, j \in 1 \ldots k$, $i \neq j$, $x_i = x_j$ (or $y_i = y_j$)? In this case, also $\boxed{x_i\#} = \boxed{x_j\#}$ and $\boxed{S_1 x_i\#} = \boxed{S_1 x_j\#}$. If, for example, $x_1 = 11$ and $x_2 = 11$, $\mathcal{G'}_{1,K}$ would contain the productions $S_1 \rightarrow \boxed{i_1\#}\boxed{11\#}$, $S_1 \rightarrow \boxed{i_2\#}\boxed{11\#}$ as well as $S_1 \rightarrow \boxed{i_1\#}\boxed{S_1 11\#}$ and $S_1 \rightarrow \boxed{i_2\#}\boxed{S_1 11\#}$ (and $\boxed{S_1 11\#} \rightarrow S_1\boxed{11\#}$, $\boxed{11\#} \rightarrow \boxed{1\#}\boxed{1\#}$ etc., of course).

As already noted, due to the construction, we have $\mathcal{G'}_{1,K} \cap \mathcal{G'}_{2,K} = \emptyset$, since words in $\mathcal{G'}_{1,K}$ have the form $i_{i_n}\# \ldots \# i_{i_2} \# i_{i_1} \# x_{i_1} \# x_{i_2} \# \ldots \# x_{i_n}\#$, and words in $\mathcal{G'}_{2,K}$ have the form $\# i_{i_n}\# \ldots \# i_{i_2} \# i_{i_1} \# y_{i_1} \# y_{i_2} \# \ldots \# y_{i_n}$. The relationship with the PCP K is the following:

Corollary 1 A PCP K has the solution $(i_1, \ldots, i_n)$ iff $\# i_{i_n}\# \ldots \# i_{i_2} \# i_{i_1} \# x_{i_1} \# x_{i_2} \# \ldots \# x_{i_n}\# \in (\{\#\}\mathcal{L}(\mathcal{G'}_{1,K})) \cap (\mathcal{L}(\mathcal{G'}_{2,K})\{\#\})$. Consequently, K has no solution iff $(\{\#\}\mathcal{L}(\mathcal{G'}_{1,K})) \cap (\mathcal{L}(\mathcal{G'}_{2,K})\{\#\}) = \emptyset$. Emptiness for this language is therefore undecidable.

Applied to the example

- $\mathcal{G}_{1,K} = (\{S_1\}, \{0, 1, i_1, i_2, i_3\}, \mathcal{P}_1, S_1)$,

  $\mathcal{P}_1 = \{S_1 \rightarrow i_1 1 \mid i_2 10 \mid i_3 011\} \cup \{S_1 \rightarrow i_1 S_1 1 \mid i_2 S_1 10 \mid i_3 S_1 011\}$

  becomes

  $\mathcal{G'}_{1,K} = (\mathcal{V'}_1, \{\#, 0, 1, i_1, i_2, i_3\}, \mathcal{P'}_1, S_1)$ with

  $$\mathcal{V'}_1 = \{\ S_1, \boxed{0\#}, \boxed{1\#}, \boxed{i_1\#}, \boxed{i_2\#}, \boxed{i_3\#}, \boxed{10\#}, \boxed{011\#}, \boxed{11\#}, \boxed{S_1 1\#}, \boxed{S_1 10\#}, \boxed{S_1 011\#}\ \}$$

  $$\begin{aligned}\mathcal{P'}_1 = \ & \{\boxed{0\#} \rightarrow 0\#, \boxed{1\#} \rightarrow 1\#, \boxed{i_1\#} \rightarrow i_1\#, \boxed{i_2\#} \rightarrow i_2\#, \boxed{i_3\#} \rightarrow i_3\#\} \cup \\ & \{S_1 \rightarrow \boxed{i_1\#}\ \boxed{1\#}, S_1 \rightarrow \boxed{i_2\#}\ \boxed{10\#}, S_1 \rightarrow \boxed{i_3\#}\ \boxed{011\#}\} \cup \\ & \{S_1 \rightarrow \boxed{i_1\#}\ \boxed{S_1 1\#}, S_1 \rightarrow \boxed{i_2\#}\ \boxed{S_1 10\#}, S_1 \rightarrow \boxed{i_3\#}\ \boxed{S_1 011\#}\} \cup \\ & \{\boxed{S_1 1\#} \rightarrow S_1\boxed{1\#}, \boxed{S_1 10\#} \rightarrow S_1\boxed{10\#}, \boxed{S_1 011\#} \rightarrow S_1\boxed{011\#}\} \cup \\ & \{\boxed{10\#} \rightarrow \boxed{1\#}\ \boxed{0\#}, \boxed{011\#} \rightarrow \boxed{0\#}\ \boxed{11\#}, \boxed{11\#} \rightarrow \boxed{1\#}\ \boxed{1\#}\}, \text{ and}\end{aligned}$$

- $\mathcal{G}_{2,K} = (\{S_2\}, \{0, 1, i_1, i_2, i_3\}, \mathcal{P}_2, S_2)$,

  $\mathcal{P}_2 = \{S_2 \rightarrow i_1 101 \mid i_2 00 \mid i_3 11\} \cup \{S_2 \rightarrow i_1 S_2 101 \mid i_2 S_2 00 \mid i_3 S_2 11\}$

  becomes

  $\mathcal{G'}_{2,K} = (\mathcal{V'}_2, \{\#, 0, 1, i_1, i_2, i_3\}, \mathcal{P'}_2, S_2)$ with

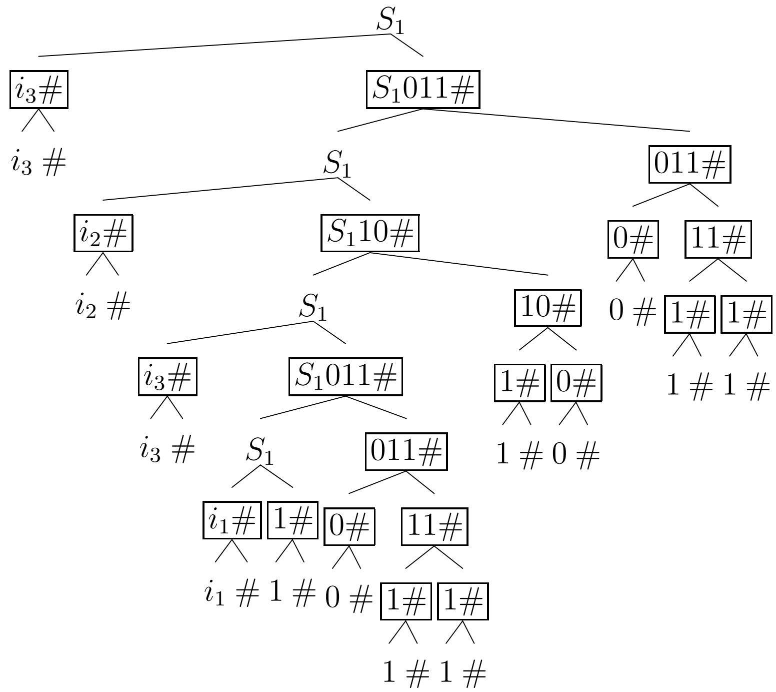

Figure 1: $i_3\#i_2\#i_3\#i_1\#1\#0\#1\#1\#1\#0\#0\#1\#1\# \in \mathcal{L}({\mathcal{G}'}_{1,K})$

$$
\begin{aligned}
{\mathcal{V}'}_2 = \{\ & S_2, \boxed{\#0}, \boxed{\#1}, \\
& \boxed{\#i_1}, \boxed{\#i_2}, \boxed{\#i_3}, \\
& \boxed{\#10}, \boxed{\#011}, \boxed{\#11}, \\
& \boxed{S_2\#101}, \boxed{S_2\#00}, \boxed{S_2\#11}\ \}
\end{aligned}
$$

$$
\begin{aligned}
{\mathcal{P}'}_2 = \ & \{\boxed{\#0} \to \#0, \boxed{\#1} \to \#1, \boxed{\#i_1} \to \#i_1, \boxed{\#i_2} \to \#i_2, \boxed{\#i_3} \to \#i_3\} \cup \\
& \{S_2 \to \boxed{\#i_1}\ \boxed{\#101}, S_2 \to \boxed{\#i_2}\ \boxed{\#00}, S_2 \to \boxed{\#i_3}\ \boxed{\#11}\} \cup \\
& \{S_2 \to \boxed{\#i_1}\ \boxed{S_2\#101}, S_2 \to \boxed{\#i_2}\ \boxed{S_2\#00}, S_2 \to \boxed{\#i_3}\ \boxed{S_2\#11}\} \cup \\
& \{\boxed{S_2\#101} \to S_2\boxed{\#101}, \boxed{S_2\#00} \to S_2\boxed{\#00}, \boxed{S_2\#11} \to S_2\boxed{\#11}\} \cup \\
& \{\boxed{\#101} \to \boxed{\#1}\ \boxed{\#01}, \boxed{\#01} \to \boxed{\#0}\ \boxed{\#1}, \\
& \ \boxed{\#00} \to \boxed{\#0}\ \boxed{\#0}, \boxed{\#11} \to \boxed{\#1}\ \boxed{\#1}\}.
\end{aligned}
$$

We can verify that $i_3\#i_2\#i_3\#i_1\#1\#0\#1\#1\#1\#0\#0\#1\#1\# \in \mathcal{L}({\mathcal{G}'}_{1,K})$ (see Figure 1) and $\#i_3\#i_2\#i_3\#i_1\#1\#0\#1\#1\#1\#0\#0\#1\#1 \in \mathcal{L}({\mathcal{G}'}_{2,K})$ (see Figure 2).

The grammars ${\mathcal{G}'}_{i,K}$, $i \in \{1,2\}$ have the following important property: whenever a word w is derivable by *some* non-terminal $A \in {\mathcal{V}'}_i$ such that $A \xrightarrow{+} w$, then there is no other non-terminal $B \in {\mathcal{V}'}_i$ with $A \neq B$ such that also $B \xrightarrow{+} w$. We can even put "the grammars together", and still this property holds:

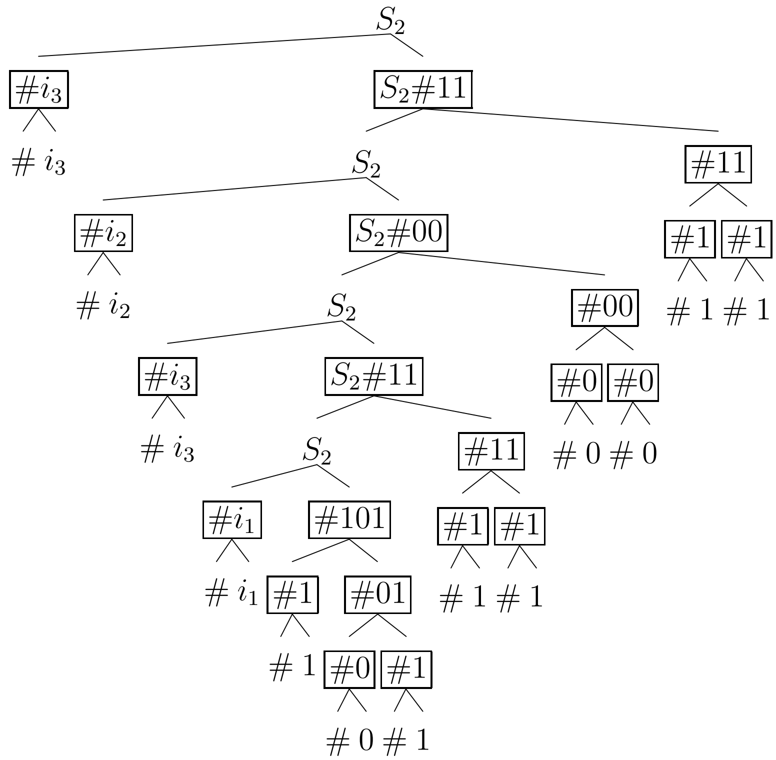

Figure 2: $\#i_3\#i_2\#i_3\#i_1\#1\#0\#1\#1\#1\#0\#0\#1\#1 \in \mathcal{L}(\mathcal{G}'_{2,K})$

Lemma 2 Let $\mathcal{G}_K =_{def} (\mathcal{V}'_1 \cup \mathcal{V}'_2, \{\#, 0, 1, i_1, i_2, i_3\}, \mathcal{P}'_1 \cup \mathcal{P}'_2, \emptyset)$ be the union of the two grammars $\mathcal{G}'_{1,K}$ and $\mathcal{G}'_{2,K}$ as defined above.[2]

Then, for all $w \in \{\#, 0, 1, i_1, i_2, i_3\}^+$, for all $A, B \in \mathcal{V}'_1 \cup \mathcal{V}'_2$, if $A \xrightarrow{+} w$ and $B \xrightarrow{+} w$, then $A = B$.

Proof 2 First of all note that $\mathcal{V}'_1 \cap \mathcal{V}'_2 = \emptyset$. Now assume $A \in \mathcal{V}'_1$, $B \in \mathcal{V}'_2$, and there is a word w such that $A \xrightarrow{+} w$ and $B \xrightarrow{+} w$. Since $\mathcal{V}'_1 \cap \mathcal{V}'_2 = \emptyset$, $A \neq B$ would follow, a contradiction. We must therefore show that there is no such w. It can be easily verified that $\{\, w \mid A \xrightarrow{+} w \,\} \cap \{\, w \mid B \xrightarrow{+} w \,\} = \emptyset$, since the "odd-even-interleaving" of "#" holds not only for $A = S_1$ and $B = S_2$, but for *all* words derivable by *some* pair of non-terminals A, B. But then we must have $A, B \in \mathcal{V}'_i$, for some $i \in \{1, 2\}$ – in the following we will show the lemma for $i = 1$; the case for $i = 2$ is analogous and left out here for the sake of brevity.

The proof is a simple induction on $|w|$. Obviously, for each w with $A \xrightarrow{+} w$ for some $A \in \mathcal{V}'_1$ we have $|w| = 2j$ for some $j \in \mathbb{N} \setminus \{0\}$:

[2] This grammar has no starting symbol, since we do not consider the language of this grammar. Its purpose is just to act as a "container data structure".

- If $j = 1$, $|w| = 2$ and $A \xrightarrow{+} w$, then $w = a\#$ for $a \in \{0, 1, i_1, i_2, i_3\}$. Therefore, $A = \boxed{a\#}$ and $\boxed{a\#} \rightarrow a\# \in \mathcal{P}'_1$, which is the only rule with "$a\#$" on its right hand side. Therefore, $B \not\xrightarrow{+} a\#$, if $A \neq B$.

- If $|w| = 2j$, $j \geq 2$, $j \in \mathbb{N}$ and $A \xrightarrow{+} w$, then there must be a production $P \in \mathcal{P}'_1$ with $P = (A \rightarrow XY)$, $X, Y \in \mathcal{V}'_1$. In the following we can forget about the productions with $A = \boxed{a\#}$, since they derive words of length two. We have $w = w_X w_Y$, and $X \xrightarrow{+} w_X$, $Y \xrightarrow{+} w_Y$. Since $|w_X| < |w|$, $|w_Y| < |w|$ (note that there are no productions of the form $X \rightarrow \epsilon$!), the induction hypothesis holds, and thus there are no $X' \neq X$ with $X' \xrightarrow{+} w_X$ or $Y' \neq Y$ with $Y' \xrightarrow{+} w_Y$. Therefore, X and Y uniquely determine the production $P = (A \rightarrow XY)$. One can easily check that there is no production $P' = (B \rightarrow XY)$ with $B \neq A$.[3]

  However, we also need to argue that there is no other partition of w, with $w = w'_X w'_Y$, $w'_X \neq w_X$, $w'_Y \neq w_Y$, such that $X' \xrightarrow{+} w'_X$, $Y' \xrightarrow{+} w'_Y$ and $P' = (B \rightarrow X'Y')$, $P' \in \mathcal{P}'_1$ with $A \neq B$. If $X' = X$ and $Y' = Y$, we already know that $A = B$, since P' is the only production with X and Y on its right hand side. Otherwise we can make a case distinction, assuming $A \neq B$ and derive a contradiction in every case:

  - $P = (S_1 \rightarrow \boxed{i_n\#}\ \boxed{x_n\#})$, $A = S_1$, $X = \boxed{i_n\#}$, $Y = \boxed{x_n\#}$, for some $n \in 1 \dots k$,
    * $P' = (\boxed{S_1 x_m \#} \rightarrow S_1 \boxed{x_m\#})$, $B = \boxed{S_1 x_m\#}$, $X' = S_1$, $Y' = \boxed{x_m\#}$: note that $w'_X = i_n\# \dots$ (since $w = w_X w_Y = w'_X w'_Y$ and $w_X = i_n\# \dots$, due to $\boxed{i_n\#} \xrightarrow{+} w_X$). Since also $w'_X \in \mathcal{L}(\mathcal{G}_{1,K})$, this shows that $w'_X = i_n\# \dots \mathsf{even}_\#(x_n)$. Since already $w = i_n\#\mathsf{even}_\#(x_n)$ and $w = w'_X w'_Y$ it follows that $w'_X = w$ and therefore $w'_Y = \epsilon$ which contradicts $w'_Y = \mathsf{even}_\#(x_m)$ (note that the PCP K does not contain empty words in its word list).
    * $P' = (\boxed{x\#} \rightarrow \boxed{\mathsf{first}(x)\#}\ \boxed{\mathsf{rest}(x)\#})$, $B = \boxed{x\#}$, $X' = \boxed{\mathsf{first}(x)\#}$, $Y' = \boxed{\mathsf{rest}(x)\#}$, for some $n \in 1, \dots, k$, $x \in \mathsf{postfixes}(x_n)$, $|x| \geq 2$ (where x_n is the nth word in the PCP K): obvious, since $\mathsf{first}(x) \neq i_n$, because $x \in \{a_1, a_2\}^\star$ (recall that $\{a_1, a_2\}$ is the alphabet of the PCP K), but $i_n \notin \{a_1, a_2\}$.
  - $P = (\boxed{S_1 x_n\#} \rightarrow S_1\ \boxed{x_n\#})$, $A = \boxed{S_1 x_n\#}$, $X = S_1$, $Y = \boxed{x_n\#}$, for some $n \in 1 \dots k$

[3] As already noted, if there were some $x_i = x_j$ (in the PCP K of size k) for $i \neq j$, $i, j \in 1 \dots k$, then the productions $\boxed{S_1 x_i\#} \rightarrow S_1\ \boxed{x_i\#}$ and $\boxed{S_1 x_j\#} \rightarrow S_1\ \boxed{x_j\#}$ would coincide, since $\boxed{x_i\#} = \boxed{x_j\#}$ and also $\boxed{S_1 x_i\#} = \boxed{S_1 x_j\#}$.

* $P' = (S_1 \rightarrow \boxed{i_m\#}\ \boxed{x_m\#})$, $B = S_1$, $X' = \boxed{i_m\#}$, $Y' = \boxed{x_m\#}$: w'_X must start with $i_m\#$. This shows that w_X must start with $i_m\#$ and thus has the form $w_X = i_m\# \ldots \mathsf{even}_\#(x_m)$. This leads to the conclusion that $w_Y = \epsilon$ which contradicts $w_Y = \mathsf{even}_\#(x_n)$.
* $P' = (\boxed{x\#} \rightarrow \boxed{\mathsf{first}(x)\#}\ \boxed{\mathsf{rest}(x)\#})$, $B = \boxed{x\#}$, $X' = \boxed{\mathsf{first}(x)\#}$, $Y' = \boxed{\mathsf{rest}(x)\#}$: obvious, see above.

– $P = (\boxed{x\#} \rightarrow \boxed{\mathsf{first}(x)\#}\ \boxed{\mathsf{rest}(x)\#})$, $A = \boxed{x\#}$, $X = \boxed{\mathsf{first}(x)\#}$, $Y = \boxed{\mathsf{rest}(x)\#}$, for some $n \in 1, \ldots, k$, $x \in \mathsf{postfixes}(x_n)$, $|x| \geq 2$ (where x_n is the nth word in the PCP K)

* $P' = (\boxed{x'\#} \rightarrow \boxed{\mathsf{first}(x')\#}\ \boxed{\mathsf{rest}(x')\#})$, $B = \boxed{x\#}$, $X' = \boxed{\mathsf{first}(x')\#}$, $Y' = \boxed{\mathsf{rest}(x')\#}$, for some $n \in 1, \ldots, k$, $x' \in \mathsf{postfixes}(x_n)$, $|x| \geq 2$: obviously, $A = B$ iff $x' = x$. Therefore, $A \neq B$ iff $x' \neq x$. However, then either $\mathsf{first}(x) \neq \mathsf{first}(x')$ or $\mathsf{rest}(x) \neq \mathsf{rest}(x')$. In both cases the contradiction is immediate, since $w = w_X w_Y = \mathsf{even}_\#(x)$ and $w = w'_X w'_Y = \mathsf{even}_\#(x')$.
* $P' = (S_1 \rightarrow \boxed{i_m\#}\ \boxed{x_m\#})$, $B = S_1$, $X' = \boxed{i_m\#}$, $Y' = \boxed{x_m\#}$: obvious, since w'_X starts with $i_m\#$ and $i_m \notin \{a_1, a_2\}$.
* $P' = (\boxed{S_1 x_m\#} \rightarrow S_1\boxed{x_m\#})$, $B = \boxed{S_1 x_m\#}$, $X' = S_1$, $Y' = \boxed{x_m\#}$: obvious, since w'_X starts with $i_n\#$ for some $n \in 1 \ldots k$ and $i_n \notin \{a_1, a_2\}$.

□

The key-observation is now that one can simply *reverse* the productions of the grammar $\mathcal{G}_K$ in order to get a role box $\mathfrak{R}_\mathfrak{K}$. That is, each production of the form $A \rightarrow BC \in \mathcal{P'}_1 \cup \mathcal{P'}_2$ yields a role axiom $B \circ C \sqsubseteq A \in \mathfrak{R}_\mathfrak{K}$. The terminals and non-terminals of $\mathcal{G}_K$ are considered as roles now. If a word can be derived "top down" by the grammar using a derivation tree, then it is possible to "parse" this word in a bottom-up style using the role axioms. The previous lemma ensures that the disjointness-requirement of $\mathcal{ALC}_{\mathcal{RA}}$ cannot be violated during this "bottom-up" parsing process.

The following lemma fixes the relationship between words that are derivable by the grammar (by *some* non-terminal, not necessarily only S_1 or S_2) and the models of the role box corresponding to this grammar:

Lemma 3 Let $\mathcal{G}_K = (\mathcal{V}, \Sigma, \mathcal{P}, S)$ be the grammar constructed in Lemma 2. W.l.o.g. we assume $(\mathcal{V} \cup \Sigma) \subseteq \mathcal{N}_\mathcal{R}$. Let $w \in \Sigma^+$, $w = w_1 \ldots w_n$ be a word with $n \geq 2$, and $\mathcal{I}$ be a model of $(\exists w_1 . \ldots \exists w_n . \top, \mathfrak{R}_\mathfrak{K})$ with $\mathfrak{R}_\mathfrak{K} =_{def} \{ B \circ C \sqsubseteq A \mid A \rightarrow B\ C \in \mathcal{P} \}$.

Let $<x_0, x_1> \in w_1^{\mathcal{I}}, \ldots <x_{n-1}, x_n> \in w_n^{\mathcal{I}}$ be an arbitrary path in the model $\mathcal{I}$ corresponding to w. Note that the individuals x_i must not necessarily be distinct (e.g., there might be i, j such that $x_i = x_j$).

Let $A \in \mathcal{V}$ be an arbitrary non-terminal of $\mathcal{G}_K$. Then, $<x_0, x_n> \in A^{\mathcal{I}}$ holds in *all* models $\mathcal{I}$ of $(\exists w_1. \ldots \exists w_n.\top, \mathfrak{R}_{\mathfrak{K}})$ iff there is a derivation of w having A as the root node: we write $A \xrightarrow{+} w$.

Proof 3 **"$\Leftarrow$"** If $A \xrightarrow{+} w$, then $|w| = 2j$, $j \in \mathbb{N} \setminus \{0\}$. Using induction we show that $<x_0, x_n> \in A^{\mathcal{I}}$ (again we focus on $\mathcal{G'}_{1,K}$):

- If $|w| = 2$, then there must be a production of the form $\boxed{a\#} \rightarrow a\# \in \mathcal{P}$, for $a \in \{a_1, a_2, i_1, \ldots, i_k\}$. This shows that $w = a\#$. If $\mathcal{I}$ is a model of $\mathfrak{R}_{\mathfrak{K}}$ and $<x_0, x_1> \in a^{\mathcal{I}}$, $<x_1, x_2> \in \#^{\mathcal{I}}$, then, due to $a \circ \# \sqsubseteq \boxed{a\#} \in \mathfrak{R}_{\mathfrak{K}}$ we have $<x_0, x_2> \in \boxed{a\#}^{\mathcal{I}}$ in all models.
- If $|w| = 2j$, $j \geq 2$, then there must be a production of the form $A \rightarrow XY \in \mathcal{P}$ such that $w = w_X w_Y$, $w_X = w_1 \ldots w_m$, $w_Y = w_{m+1} \ldots w_{2j}$, $X \xrightarrow{+} w_X$, $Y \xrightarrow{+} w_Y$. Since $|w_X| < |w|$ and $|w_Y| < |w|$ the induction hypothesis holds and therefore, $<x_0, x_m> \in X^{\mathcal{I}}$ and $<x_{m+1}, x_{2j}> \in X^{\mathcal{I}}$ in *all* models. Therefore, due to $X \circ Y \sqsubseteq A \in \mathfrak{R}_{\mathfrak{K}}$ also $<x_0, x_{2j}> \in A^{\mathcal{I}}$ in all models.

"$\Rightarrow$" If $<x_0, x_n> \in A^{\mathcal{I}}$ holds in *all* models $\mathcal{I}$ of $(\exists w_1. \ldots \exists w_n.\top, \mathfrak{R}_{\mathfrak{K}})$, then we may say (with a slight abuse of terminology) that the presence of $<x_0, x_n> \in A^{\mathcal{I}}$ is a *logical consequence* of $(\exists w_1. \ldots \exists w_n.\top, \mathfrak{R}_{\mathfrak{K}})$. Please note that among the models with $<x_0, x_n> \in A^{\mathcal{I}}$ there is also a model $\mathcal{I}$ in which the x_i's are *distinct* individuals such that $\Delta^{\mathcal{I}} = \{x_0, \ldots, x_n\}$, and $<x_0, x_1> \in w_1^{\mathcal{I}}, \ldots <x_{n-1}, x_n> \in w_n^{\mathcal{I}}$ corresponds to a linear path of *direct* edges. An edge $<x, z> \in \mathcal{UR}(\mathcal{I})$ is called *direct* if there is no $y \in \Delta^{\mathcal{I}}$, $y \neq x$, $y \neq z$ such that $<x, y> \in \mathcal{UR}(\mathcal{I})^+$ and $<y, z> \in \mathcal{UR}(\mathcal{I})^+$. Now one can easily construct a derivation tree for w showing that $A \xrightarrow{+} w$ by inspecting the nesting of role compositions leading to $<x_0, x_n> \in A^{\mathcal{I}}$ in this model. More formally this could be shown by using induction as well, and the proof would be very similar to the previous one. □

Given an arbitrary word w, the expression $\exists w.C$ is defined in the obvious way: $\exists w.C =_{def} C$ if $w = \epsilon$, and $\exists w.C =_{def} \exists \mathsf{first}(w).(\exists \mathsf{rest}(w).C)$ otherwise (if $w \neq \epsilon$).

We still need to argue that the role box $\mathfrak{R}_{\mathfrak{K}}$ admits models; i.e., given an arbitrary word w, is it always the case that $(\exists w.\top, \mathfrak{R}_{\mathfrak{K}})$ is satisfiable? Please note that this is *not* granted for arbitrary role boxes in $\mathcal{ALC}_{\mathcal{RA}}$, due to the disjointness-requirement. However, for the role box $\mathfrak{R}_{\mathfrak{K}}$ we know this for sure:

Corollary 2 Let w be a word over some alphabet. Then, $(\exists w.\top, \mathfrak{R}_{\mathfrak{K}})$ is satisfiable (in $\mathcal{ALC}_{\mathcal{RA}}$).

Proof 4 Suppose that $w = w_1 \dots w_n$, $(\exists w_1.\dots\exists w_n.\top, \mathfrak{R}_{\mathfrak{K}})$ is unsatisfiable. Obviously, $(\exists w_1.\dots\exists w_n.\top, \mathfrak{R}_{\mathfrak{K}})$ can only become unsatisfiable if the disjointness requirement cannot be fulfilled, e.g. in any interpretation $\mathcal{I}$ in which $\mathcal{I} \models \exists w_1.\dots\exists w_n.\top$ and $\mathcal{I} \models \mathfrak{R}_{\mathfrak{K}}$ holds there exist at least two roles $S, T \in \mathcal{N_R}$ such that $S^{\mathcal{I}} \cap T^{\mathcal{I}} \neq \emptyset$ is enforced by the role box. Then, *every* model must contain $S^{\mathcal{I}} \cap T^{\mathcal{I}} \neq \emptyset$ (otherwise we could find another model, and $(\exists w.\top, \mathfrak{R}_{\mathfrak{K}})$ would be satisfiable). Of course, for $n = 1$, $(\exists w_1.\top, \mathfrak{R}_{\mathfrak{K}})$ is always satisfiable. According to Lemma 3, for $n \geq 2$ we have $\langle x_0, x_n \rangle \in S^{\mathcal{I}}$ in *every* model iff $S \xrightarrow{+} w$ and $\langle x_0, x_n \rangle \in T^{\mathcal{I}}$ in *every* model iff $T \xrightarrow{+} w$. However, this is a contradiction to Lemma 2. □

Returning to our example PCP K, the following role box will be constructed:

$$
\begin{array}{rl}
\mathfrak{R}_{\mathfrak{K}} = & \{\ 0 \circ \# \sqsubseteq \boxed{0\#}, 1 \circ \# \sqsubseteq \boxed{1\#}, \\
& \quad i_1 \circ \# \sqsubseteq \boxed{i_1\#}, i_2 \circ \# \sqsubseteq \boxed{i_2\#}, i_3 \circ \# \sqsubseteq \boxed{i_3\#}, \\
& \quad \boxed{i_1\#} \circ \boxed{1\#} \sqsubseteq S_1, \boxed{i_2\#} \circ \boxed{10\#} \sqsubseteq S_1, \boxed{i_3\#} \circ \boxed{011\#} \sqsubseteq S_1, \\
& \quad \boxed{i_1\#} \circ \boxed{S_1 1\#} \sqsubseteq S_1, \boxed{i_2\#} \circ \boxed{S_1 10\#} \sqsubseteq S_1, \boxed{i_3\#} \circ \boxed{S_1 011\#} \sqsubseteq S_1, \\
& \quad S_1 \circ \boxed{1\#} \sqsubseteq \boxed{S_1 1\#}, S_1 \circ \boxed{10\#} \sqsubseteq \boxed{S_1 10\#}, S_1 \circ \boxed{011\#} \sqsubseteq \boxed{S_1 011\#}, \\
& \quad \boxed{1\#} \circ \boxed{0\#} \sqsubseteq \boxed{10\#}, \boxed{0\#} \circ \boxed{11\#} \sqsubseteq \boxed{011\#}, \boxed{1\#} \circ \boxed{1\#} \sqsubseteq \boxed{11\#}\ \} \ \cup \\
& \{\ \# \circ 0 \sqsubseteq \boxed{\#0}, \# \circ 1 \sqsubseteq \boxed{\#1}, \\
& \quad \# \circ i_1 \sqsubseteq \boxed{\#i_1}, \# \circ i_2 \sqsubseteq \boxed{\#i_2}, \# \circ i_3 \sqsubseteq \boxed{\#i_3}, \\
& \quad \boxed{\#i_1} \circ \boxed{\#101} \sqsubseteq S_2, \boxed{\#i_2} \circ \boxed{\#00} \sqsubseteq S_2, \boxed{\#i_3} \circ \boxed{\#11} \sqsubseteq S_2, \\
& \quad \boxed{\#i_1} \circ \boxed{S_2\#101} \sqsubseteq S_2, \boxed{\#i_2} \circ \boxed{S_2\#00} \sqsubseteq S_2, \boxed{\#i_3} \circ \boxed{S_2\#11} \sqsubseteq S_2, \\
& \quad S_2 \circ \boxed{\#101} \sqsubseteq \boxed{S_2\#101}, S_2 \circ \boxed{\#00} \sqsubseteq \boxed{S_2\#00}, S_2 \circ \boxed{\#11} \sqsubseteq \boxed{S_2\#11}, \\
& \quad \boxed{\#1} \circ \boxed{\#01} \sqsubseteq \boxed{\#101}, \boxed{\#0} \circ \boxed{\#1} \sqsubseteq \boxed{\#01}, \\
& \quad \boxed{\#0} \circ \boxed{\#0} \sqsubseteq \boxed{\#00}, \boxed{\#1} \circ \boxed{\#1} \sqsubseteq \boxed{\#11}\ \}.
\end{array}
$$

The "first part" of this role box corresponds to $\mathcal{P}'_1$, and the "second part" to $\mathcal{P}'_2$.

One can now use this role box to solve the membership problem of $\mathcal{L_K}$. For example, consider $w \in \mathcal{L_K}$, with $w = \#i_3\#i_2\#i_3\#i_1\#1\#0\#1\#1\#1\#0\#0\#1\#1\#$. The following concept term is unsatisfiable w.r.t. $\mathfrak{R}_{\mathfrak{K}}$, since $w \in \mathcal{L_K}$. Recall that $w \in \mathcal{L_K}$ iff $w = \#\alpha\#$, $\alpha\# \in \mathcal{L}(\mathcal{G}'_{1,K})$ and $\#\alpha \in \mathcal{L}(\mathcal{G}'_{2,K})$. Consider

$$
\begin{array}{l}
(((\forall \#.\forall S_1.C) \sqcap (\forall S_2.\forall \#.D) \sqcap \\
\qquad \exists \#i_3\#i_2\#i_3\#i_1\#1\#0\#1\#1\#1\#0\#0\#1\#1\#.\neg(C \sqcap D)), \mathfrak{R}_{\mathfrak{K}})
\end{array}
$$

Any model of this example would also be a model of $(\exists \#i_3\#i_2\#i_3\#i_1\#1\#0\#1\#1\#1\#0\#0\#1\#1\#.\top, \mathfrak{R}_{\mathfrak{K}})$. Let $\langle x_0, x_1 \rangle \in \#^{\mathcal{I}}$, $\langle x_1, x_2 \rangle \in i_3^{\mathcal{I}}$, $\langle x_3, x_4 \rangle \in \#^{\mathcal{I}}$, $\langle x_4, x_5 \rangle \in i_2^{\mathcal{I}}$, $\dots$, $\langle x_{26}, x_{27} \rangle \in \#^{\mathcal{I}}$ (see also

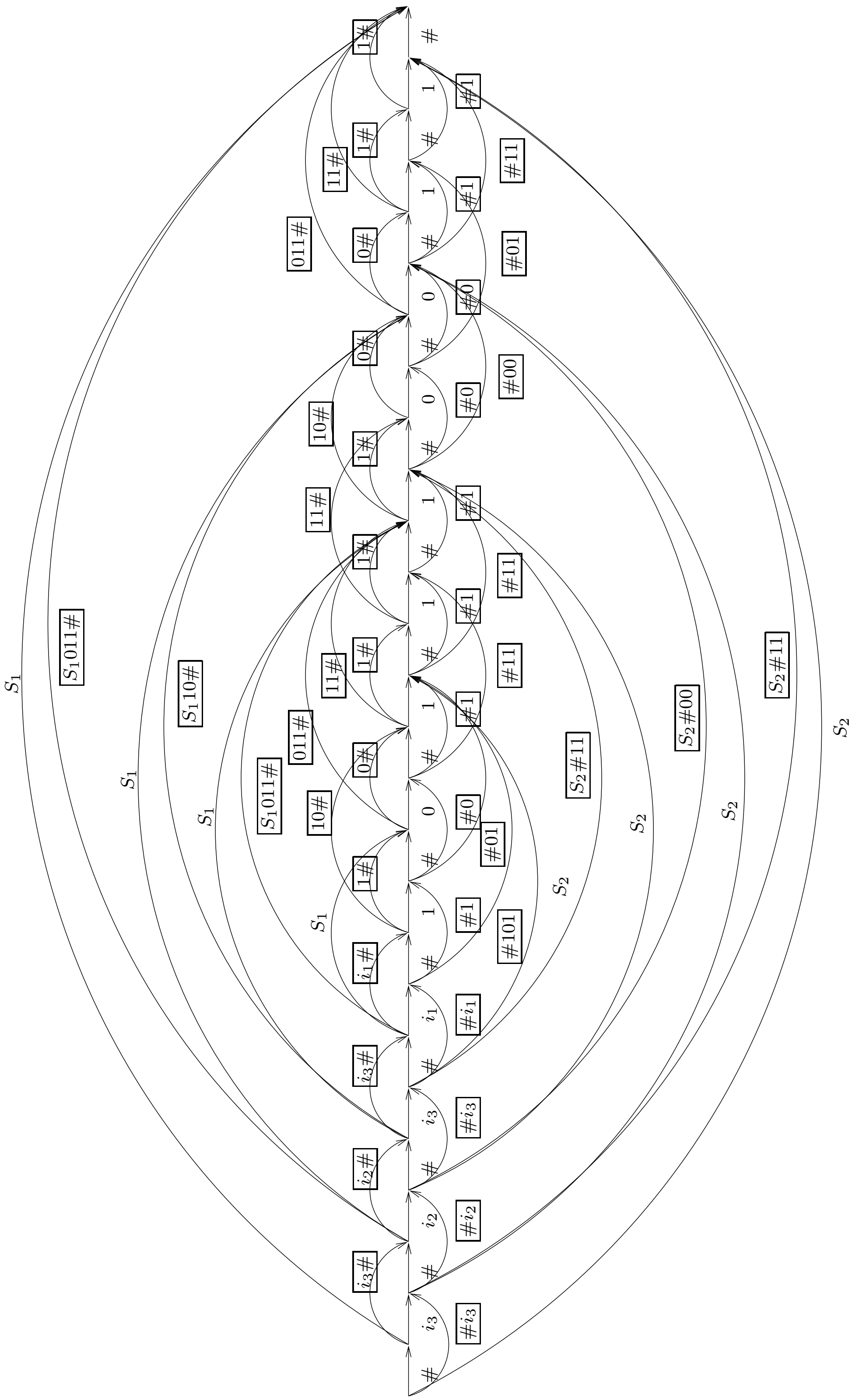

Figure 3: “Bottom up parsing” of $\#i_3\#i_2\#i_3\#i_1\#1\#0\#1\#1\#1\#0\#0\#1\#1\#$

Figure 3). Due to Lemma 3, since $i_3\#i_2\#i_3\#i_1\#1\#0\#1\#1\#1\#0\#0\#1\#1\# \in \mathcal{L}({\mathcal{G}'}_{1,K})$, we have $<x_1, x_{27}> \in S_1^{\mathcal{I}}$. But then, due to $x_0 \in (\forall\#.\forall S_1.C)^{\mathcal{I}}$, also $x_{27} \in C^{\mathcal{I}}$. Since $\#i_3\#i_2\#i_3\#i_1\#1\#0\#1\#1\#1\#0\#0\#1\#1 \in \mathcal{L}({\mathcal{G}'}_{2,K})$, we have $<x_0, x_{26}> \in S_2^{\mathcal{I}}$. But then, due to $x_0 \in (\forall S_2.\forall\#.C)^{\mathcal{I}}$, also $x_{27} \in D^{\mathcal{I}}$. However, this contradicts $x_{27} \in (\neg(C \sqcap D))^{\mathcal{I}}$. The example is therefore unsatisfiable. Considering Figure 3, it can be seen that the role box performs a "bottom up parsing" of the word $\#i_3\#i_2\#i_3\#i_1\#1\#0\#1\#1\#1\#0\#0\#1\#1\#$. The two derivation trees shown in Figure 1 and 2 can be immediately discovered in Figure 3. One can also clearly see that all roles are interpreted disjointly.

With the auxiliary machinery at hand, we can now show the main result of the paper:

Theorem 1 The satisfiability problem of $\mathcal{ALC}_{\mathcal{RA}}$ is undecidable.

Proof 5 We give an example for a pair $(E, \mathfrak{R}_{\mathfrak{K}})$ for which no algorithm exists that is capable of checking its satisfiability.

Let $\mathcal{G}_K = (\mathcal{V}, \Sigma, \mathcal{P}, S)$ be the grammar of Lemma 2. Let

$$\mathfrak{R}_{\mathfrak{K}} =_{def} \{ B \circ C \sqsubseteq A \mid A \to B\ C \in \mathcal{P} \}.$$

Let $R_? \notin \mathsf{roles}(\mathfrak{R}_{\mathfrak{K}})$, and let

$$\mathfrak{R}'_{\mathfrak{K}} =_{def} \mathfrak{R}_{\mathfrak{K}} \cup \{ R \circ S \sqsubseteq \sqcup_{T \in (\mathsf{roles}(\mathfrak{R}_{\mathfrak{K}}) \cup R_?)} T \mid R, S \in (\{R_?\} \cup \mathsf{roles}(\mathfrak{R}_{\mathfrak{K}})), \neg\exists ra \in \mathfrak{R}_{\mathfrak{K}} : \mathsf{pre}(ra) = (R, S) \}$$

be the *completion* of $\mathfrak{R}_{\mathfrak{K}}$. In the following, the so-called "don't care role" $R_?$ plays a special role.

Then, $(E, \mathfrak{R}'_{\mathfrak{K}})$ is satisfiable iff $\mathcal{L}_{\mathcal{K}} = \emptyset$, where $\mathcal{L}_{\mathcal{K}} =_{def} (\{\#\}\mathcal{L}({\mathcal{G}'}_{1,K})) \cap (\mathcal{L}({\mathcal{G}'}_{2,K})\{\#\})$ (due to Corollary 1, K has no solution then). The concept E is defined as

$$E =_{def} X \sqcap \neg(C \sqcap D) \sqcap Y \sqcap (\forall\#.\forall S_1.C) \sqcap (\forall S_2.\forall\#.D), \text{ with}$$
$$X =_{def} \sqcap_{a \in \Sigma} \exists a.\top \text{ and}$$
$$Y =_{def} \sqcap_{R \in \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}})} \forall R.(X \sqcap \neg(C \sqcap D)).$$

We have to show that $(E, \mathfrak{R}'_{\mathfrak{K}})$ is satisfiable iff $\mathcal{L}_{\mathcal{K}} = \emptyset$:

"$\Rightarrow$" We prove the contra-positive: if $\mathcal{L}_{\mathcal{K}} \neq \emptyset$, then $(E, \mathfrak{R}'_{\mathfrak{K}})$ is *un*satisfiable. Assume to the contrary that $\mathcal{L}_{\mathcal{K}} \neq \emptyset$, but $(E, \mathfrak{R}'_{\mathfrak{K}})$ is satisfiable. Let $\mathcal{I}$ be a model of $(E, \mathfrak{R}'_{\mathfrak{K}})$. Because $\mathcal{I}$ satisfies $\mathfrak{R}'_{\mathfrak{K}}$, it holds that $<x_0, x_n> \in (\bigcup_{R \in \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}})} R^{\mathcal{I}})^+$ implies $<x_0, x_n> \in \mathcal{UR}(\mathcal{I})$. This is ensured by the fact

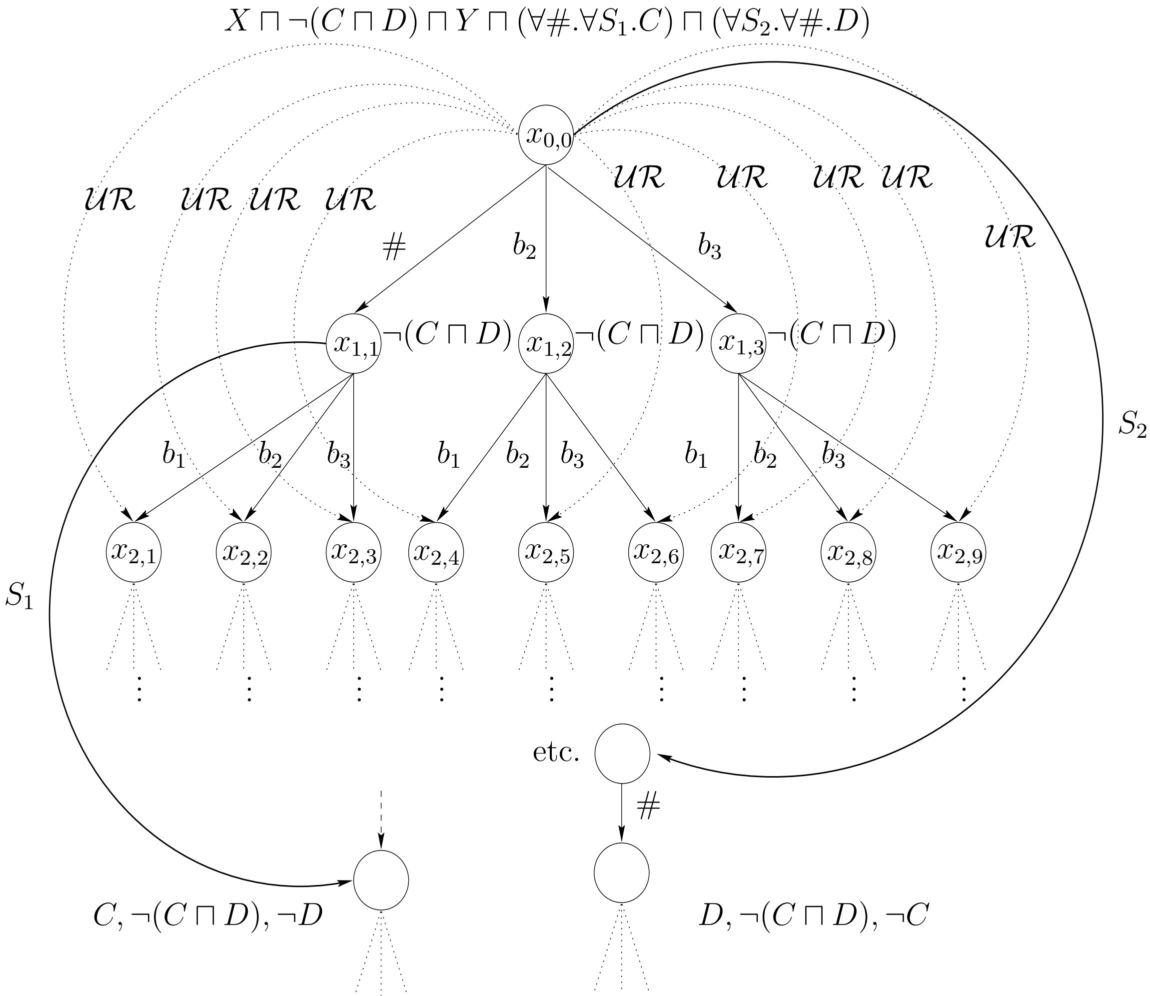

Figure 4: Illustration of the constructed model for $(E, \mathfrak{R}'_{\mathfrak{K}})$. Not all edges are shown; please note that each edge is also a member of $\mathcal{UR}$, even if not labeled with $\mathcal{UR}$.

that the composition of two arbitrary roles from $\mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}})$ is always defined in $\mathfrak{R}'_{\mathfrak{K}}$, due to the completion process. Since $\mathcal{I}$ is a model of E, there is some $x_0 \in E^{\mathcal{I}}$. Due to $x_0 \in (X \sqcap Y)^{\mathcal{I}}$ it holds that $x_0 \in ((\sqcap_{a \in \Sigma} \exists a.\top) \sqcap (\sqcap_{R \in \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}})} \forall R.(\sqcap_{a \in \Sigma} \exists a.\top)))^{\mathcal{I}}$. The model $\mathcal{I}$ therefore represents *all* possible words $w \in \Sigma^+$. Let $w \in \mathcal{L}_{\mathcal{K}}$, with $w = w_1 \ldots w_{n-1} w_n$. Please note that $w = \#\alpha\#$, where $\alpha \in \Sigma^+$, since $\mathcal{L}_{\mathcal{K}} =_{def} (\{\#\}\mathcal{L}({\mathcal{G}'}_{1,K}) \cap (\mathcal{L}({\mathcal{G}'}_{2,K})\{\#\})$, and therefore $\alpha\# \in \mathcal{L}({\mathcal{G}'}_{1,K})$ and $\#\alpha \in \mathcal{L}({\mathcal{G}'}_{2,K})$. Obviously, $\mathcal{I}$ is also a model of $\exists\#\alpha\#.\top$, with $x_0 \in (\exists\#\alpha\#.\top)^{\mathcal{I}}$. Let $x_0, x_1, \ldots, x_n$ be the path in the model corresponding to the word $\#\alpha\#$, i.e. $<x_0, x_1> \in \#^{\mathcal{I}}, \ldots, <x_{n-1}, x_n> \in \#^{\mathcal{I}}$. Due to $\mathfrak{R}_{\mathfrak{K}} \subseteq \mathfrak{R}'_{\mathfrak{K}}$, $\mathcal{I}$ is also a model of $\mathfrak{R}_{\mathfrak{K}}$, and therefore Lemma 3 is applicable. Since $S_1 \xrightarrow{+} \alpha\#$ we have $<x_1, x_n> \in S_1^{\mathcal{I}}$. Since

$S_2 \xrightarrow{+} \#\alpha$ we have $<x_0, x_{n-1}> \in S_2^{\mathcal{I}}$. However, in every model of E we also have $x_0 \in ((\forall\#.\forall S_1.C) \sqcap (\forall S_2.\forall\#.D))^{\mathcal{I}}$, and therefore $x_n \in (C \sqcap D)^{\mathcal{I}}$. But this contradicts $x_n \in \neg(C \sqcap D)^{\mathcal{I}}$ caused by $<x_0, x_n> \in \mathcal{UR}(\mathcal{I})$ and $x_0 \in (\sqcap_{R \in \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}})} \forall R.\neg(C \sqcap D))^{\mathcal{I}}$, and $(E, \mathfrak{R}'_{\mathfrak{K}})$ is therefore unsatisfiable.

"$\Leftarrow$" If $\mathcal{L}_K = \emptyset$, then we show that $(E, \mathfrak{R}'_{\mathfrak{K}})$ is satisfiable by constructing an infinite model. The model $\mathcal{I}$ is constructed incrementally, e.g. $\mathcal{I}_0 \subset \mathcal{I}_1 \subset \mathcal{I}_2 \subset \cdots \subset \mathcal{I}_\omega$, $\mathcal{I} = \mathcal{I}_\omega$. In the following construction, we refer to the set $\mathcal{UR}(\mathcal{I}, \Sigma)$ (not to be confused with $\mathcal{UR}(\mathcal{I})$!) as the *skeleton* of the model $\mathcal{I}$. The skeleton has the form of an infinite tree. An illustration of $\mathcal{I}$ is given in Figure 4. Each node in the model $\mathcal{I}$ has $|\Sigma|$ different *direct successors* in the skeleton; the skeleton of $\mathcal{I}$ is a tree with branching factor $|\Sigma|$.

For each $n \in \mathbb{N} \cup \{0\}$, the skeleton of the interpretation $\mathcal{I}_n$ is a tree of depth n, encoding *all* words w with $|w| \leq n$, i.e. $w \in \bigcup_{i \in \{0,\ldots,n\}} \Sigma^i$. Each word w of length $i = |w|$, $i \leq n$, corresponds to a path from the root node $x_{0,0}$ to some node $x_{i,m}$ at depth i, in all skeletons of the models $\mathcal{I}_n$. Therefore, the skeleton of $\mathcal{I}$ represents *all* words from Σ^+.

Intuitively, the terminal symbols of the words to be parsed by the role box are represented as *direct* edges in the skeleton of the model, whereas the *indirect* edges in this model are inserted to mimic the "bottom-up parsing process" of these words, which is performed by the role box. The model $\mathcal{I}$ is constructed as follows:

- $\mathcal{I}_0 = (\Delta_0^{\mathcal{I}}, \cdot_0^{\mathcal{I}})$, $\Delta_0^{\mathcal{I}} := \{x_{0,0}\}$, $\cdot_0^{\mathcal{I}} := \{\}$
- For $n \in 0, 1, \ldots$,
  $\mathcal{I}_{n+1} = (\Delta_{n+1}^{\mathcal{I}}, \cdot_{n+1}^{\mathcal{I}})$ is constructed from $\mathcal{I}_n = (\Delta_n^{\mathcal{I}}, \cdot_n^{\mathcal{I}})$ as follows:
  1. $\Delta_{n+1}^{\mathcal{I}} := \Delta_n^{\mathcal{I}} \cup \{\, x_{n+1,j} \mid j \in \{1, \ldots, |\Sigma|^{n+1}\} \,\}$,
     $\cdot_{n+1}^{\mathcal{I}} := \cdot_n^{\mathcal{I}}$
  2. $\Sigma = \{b_1, \ldots, b_k\}$, $\forall b_r \in \{b_1, \ldots, b_k\}$:
     $$b_r^{\mathcal{I}_{n+1}} := b_r^{\mathcal{I}_{n+1}} \cup \{\, <x_{n,j}, x_{n+1,k(j-1)+r}> \mid x_{n,j} \in \Delta_n^{\mathcal{I}}, x_{n+1,k(j-1)+r} \in \Delta_{n+1}^{\mathcal{I}} \,\}$$
  3. `while` $\mathcal{I}_{n+1} \not\models \mathfrak{R}_{\mathfrak{K}}$ `do`
     `for each` $R \circ S \sqsubseteq T \in \mathfrak{R}_{\mathfrak{K}}$ `do`
     $T^{\mathcal{I}_{n+1}} := T^{\mathcal{I}_{n+1}} \cup R^{\mathcal{I}_{n+1}} \circ S^{\mathcal{I}_{n+1}}$
     `od`
     `od`
  4. $R_?^{\mathcal{I}_{n+1}} := \{\, <x_{i,j}, x_{n+1,k}> \mid i < n+1, x_{i,j}, x_{n+1,k} \in \Delta_{n+1}^{\mathcal{I}}, <x_{i,j}, x_{n+1,k}> \notin \mathcal{UR}(\mathcal{I}_{n+1}, \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}})) \,\}$

5. $C^{\mathcal{I}_{n+1}} := C^{\mathcal{I}_{n+1}} \cup \{\, x_{n+1,j} \mid \langle x_{0,0}, x_{n+1,j}\rangle \in \#^{\mathcal{I}_{n+1}} \circ S_1^{\mathcal{I}_{n+1}} \,\}$
6. $D^{\mathcal{I}_{n+1}} := D^{\mathcal{I}_{n+1}} \cup \{\, x_{n+1,j} \mid \langle x_{0,0}, x_{n+1,j}\rangle \in S_2^{\mathcal{I}_{n+1}} \circ \#^{\mathcal{I}_{n+1}} \,\}$

We show that $\mathcal{I}$ is a model.

First we show $\mathcal{I} \models \mathfrak{R}'_{\mathfrak{K}}$ and that all roles are disjointly interpreted. We will use induction over n, where $n \in \mathbb{N} \cup \{0\}$:

The base case for $n = 0$ is immediate.

So suppose that $\mathcal{I}_n \models \mathfrak{R}'_{\mathfrak{K}}$, and all roles are interpreted disjointly. Then, after step 3 in the construction we have obviously $\mathcal{I}_{n+1} \models \mathfrak{R}_{\mathfrak{K}}$. Please note that $\mathfrak{R}_{\mathfrak{K}}$ contains only *deterministic* role axioms, so the result of step 3 is well-defined. After step 4 we will have $\mathcal{I}_{n+1} \models \mathfrak{R}'_{\mathfrak{K}}$: note that $\mathfrak{R}_{\mathfrak{d}} = \mathfrak{R}'_{\mathfrak{K}} \backslash \mathfrak{R}_{\mathfrak{K}}$ is the *completed* part of the original role box $\mathfrak{R}_{\mathfrak{K}}$. The axioms $ra \in \mathfrak{R}_{\mathfrak{d}}$ have the form $R \circ S \sqsubseteq \sqcup_{T \in (\mathsf{roles}(\mathfrak{R}_{\mathfrak{K}}) \cup R_?)} T$, where $R, S \in (\{R_?\} \cup \mathsf{roles}(\mathfrak{R}_{\mathfrak{K}}))$, and there exists no other role axiom(s) $ra' \in \mathfrak{R}_{\mathfrak{K}}$ such that $\mathsf{pre}(ra) = (R, S)$. If $\langle x_{i,j}, x_{n+1,k}\rangle$ with $i < n+1$, $x_{i,j}, x_{n+1,k} \in \Delta^{\mathcal{I}}_{n+1}$ has not been added to $\mathcal{UR}(\mathcal{I}_{n+1}, \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}}))$ in step 3 and we therefore have $\langle x_{i,j}, x_{n+1,k}\rangle \notin \mathcal{UR}(\mathcal{I}_{n+1}, \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}}))$, then $\langle x_{i,j}, x_{n+1,k}\rangle \in R_?^{\mathcal{I}_{n+1}}$ has been added by step 4. Because $R_?$ occurs on the right hand side of *every* role axiom $ra' \in \mathfrak{R}_{\mathfrak{d}}$ we have $\mathcal{I}_n \models \mathfrak{R}'_{\mathfrak{K}}$ after step 4.

However, we also need to argue that all roles are still interpreted disjointly after step 4. Since $\mathcal{L}_K = \emptyset$ and due to Lemma 2 and Lemma 3 we know that all roles are interpreted disjointly after step 3. To see that this still holds after step 4, observe that $\mathfrak{R}_{\delta}$ is tailored in such a way that it is always *safe* to add $\langle x_{i,j}, x_{n+1,k}\rangle \in R_?^{\mathcal{I}_{n+1}}$ in step 4. The $R_?$-edges act as "don't care" edges and simply cannot violate the disjointness requirement: suppose we added $\langle x_{i',j'}, x_{n+1,k}\rangle \in R_?^{\mathcal{I}_{n+1}}$, but $\mathcal{I}_{n+1}$ already contained $\langle x_{i,j}, x_{i',j'}\rangle \in R^{\mathcal{I}_{n+1}}$, and $\langle x_{i,j}, x_{n+1,k}\rangle \in S^{\mathcal{I}_{n+1}}$ (added by step 3). Since there is no role axiom $ra \in \mathfrak{R}_{\mathfrak{K}}$ with $\mathsf{pre}(ra) = (R, R_?)$, but instead $R \circ R_? \sqsubseteq \sqcup_{T \in (\mathsf{roles}(\mathfrak{R}_{\mathfrak{K}}) \cup R_?)} \in \mathfrak{R}_{\mathfrak{d}}$, the insertion of $R_?$ did not "invalidate" the model, since the role S appears on the right hand side of this role axioms.

We now prove that $x_{0,0} \in E^{\mathcal{I}}$, i.e. $x_{0,0} \in ((\sqcap_{a \in \Sigma} \exists a.\top) \sqcap (\sqcap_{R \in \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}})} \forall R.(\sqcap_{a \in \Sigma} \exists a.\top)) \sqcap (\forall \#.\forall S_1.C) \sqcap (\forall S_2.\forall \#.D))^{\mathcal{I}}$. For each node $x_{i,j} \in \Delta^{\mathcal{I}}$ with $i \neq 0$, $j \neq 0$ we have $\langle x_{0,0}, x_{i,j}\rangle \in \mathcal{UR}(\mathcal{I})$ (note that $\{R_?^{\mathcal{I}} \subseteq \mathcal{UR}(\mathcal{I})$, and $R_? \in \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}}))$, and each node has the required $k = |\Sigma|$ direct successors, $b_1, \ldots, b_k$. Since this holds for $x_{0,0}$ and for any (arbitrarily chosen) $x_{i,j}$ as well, we have $x_{0,0} \in X^{\mathcal{I}}$, $x_{i,j} \in X^{\mathcal{I}}$, and finally $x_{0,0} \in (\sqcap_{R \in \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}})} \forall R.X)^{\mathcal{I}}$. However, also $x_{0,0} \in (\sqcap_{R \in \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}})} \forall R.\neg(C \sqcap D))^{\mathcal{I}}$ holds: assume the contrary. Let n be the smallest level in the tree corresponding to the *skeleton of* $\mathcal{I}$ for which there is some node $x_{n,i_n} \in \Delta^{\mathcal{I}}$ with $x_{n,i_n} \in C^{\mathcal{I}}$, $x_{n,i_n} \in D^{\mathcal{I}}$. Since this node lies at depth

n in the skeleton, we already have $x_{n,i_n} \in C^{\mathcal{I}_n}$, $x_{n,i_n} \in D^{\mathcal{I}_n}$. Due to the construction, $x_{n,i_n} \in C^{\mathcal{I}_n}$ iff $<x_{0,0}, x_{n,i_n}> \in \#^{\mathcal{I}_n} \circ S_1^{\mathcal{I}_n}$, and $x_{n,i_n} \in D^{\mathcal{I}_n}$ iff $<x_{0,0}, x_{n,i_n}> \in S_2^{\mathcal{I}_n} \circ \#^{\mathcal{I}_n}$. Let w be the corresponding path of (maximal) length n in the skeleton, with $w = w_1 \ldots w_n$, $<x_{0,0}, x_{1,i_1}> \in w_1^{\mathcal{I}_n}$, $\ldots <x_{n-1,i_{n-1}}, x_{n,i_n}> \in w_n^{\mathcal{I}_n}$, with $w_i \in \{a_1, a_2, \#, i_1, \ldots, i_k\}$, leading from $x_{0,0}$ to x_{n,i_n}. By construction of $\mathcal{I}$ we know that #-edges can only occur as part of the skeleton, and therefore we must have $w_1 = \#$ and $w_n = \#$. But this means that w has the form $w = \#\alpha\#$ – we have $<x_{0,0}, x_{1,i_1}> \in \#^{\mathcal{I}_n}$, $<x_{1,i_1}, x_{n,i_n}> \in S_1^{\mathcal{I}_n}$ and $\#\alpha \in \{\#\}\mathcal{L}({\mathcal{G}'}_{1,K})$, and also $<x_{0,0}, x_{n-1,i_{n-1}}> \in S_2^{\mathcal{I}_n}$, $<x_{n-1,i_{n-1}}, x_{n,i_n}> \in \#^{\mathcal{I}_n}$, therefore $\alpha\# \in \mathcal{L}({\mathcal{G}'}_{2,K})\{\#\}$. This shows that $w \in \mathcal{L}_{\mathcal{K}}$ and $\mathcal{L}_{\mathcal{K}} \neq \emptyset$, contradicting the assumption. Therefore it holds that $x_{0,0} \in (\sqcap_{R \in \mathsf{roles}(\mathfrak{R}'_{\mathfrak{K}})} \forall R.\neg(C \sqcap D))^{\mathcal{I}}$. Hence, it is shown that $\mathcal{I} \models (E, \mathfrak{R}'_{\mathfrak{K}})$. □

4 Discussion & Conclusion

The decidability status of $\mathcal{ALC}_{\mathcal{RA}}$ was an open question for quite a time now. $\mathcal{ALC}_{\mathcal{RA}}$ was first defined in [4], where we even conjectured that it might be decidable.

It should be noted that, even though full $\mathcal{ALC}_{\mathcal{RA}}$ is undecidable, there might be certain classes of *admissible* role boxes that might be useful for spatial reasoning applications with description logics. E.g., it is still unsolved whether $\mathcal{ALC}_{\mathcal{RA}}$ instantiated with a role box corresponding to the RCC8 composition table might be decidable. Perhaps special-purpose (and therefore decidable) reasoning calculi can be invented to turn special instantiations of $\mathcal{ALC}_{\mathcal{RA}}$ into suitable and computable frameworks for spatial reasoning with description logics. Please note that we have identified a decidable fragment of $\mathcal{ALC}_{\mathcal{RA}}$, called $\mathcal{ALC}_{\mathcal{RASG}}$ which offers a special class of admissible role boxes (see [2]). However, we must admit that $\mathcal{ALC}_{\mathcal{RASG}}$ in its current form (with its very strong admissibility criterion) is not very useful for spatial reasoning with description logics. However, perhaps the insights gained from $\mathcal{ALC}_{\mathcal{RASG}}$ can be further exploited in order to design a less restrictive admissibility criterion for role boxes. But this is future work.

5 Acknowledgments

I would like to thank Volker Haarslev, Bernd Neumann and Ralf Möller for valuable discussions on the topics covered in this paper.

References

[1] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979.

[2] M. Wessel. Decidable and undecidable extensions of $\mathcal{ALC}$ with composition-based role inclusion axioms. Technical Report FBI–HH–M–301/01, University of Hamburg, Computer Science Department, December 2000. Available at `http://kogs-www.informatik.uni-hamburg.de/~mwessel/report5.{ps.gz | pdf}`.

[3] M. Wessel. Obstacles on the way to spatial reasoning with description logics – undecidability of $\mathcal{ALC}_{\mathcal{RA}^{\ominus}}$. Technical Report FBI–HH–M–297/00, University of Hamburg, Computer Science Department, October 2000. Available at `http://kogs-www.informatik.uni-hamburg.de/~mwessel/report4.{ps.gz | pdf}`.

[4] M. Wessel, V. Haarslev, and R. Möller. $\mathcal{ALC}_{\mathcal{RA}}$ – $\mathcal{ALC}$ with Role Axioms. In F. Baader and U. Sattler, editors, *Proceedings of the International Workshop in Description Logics 2000 (DL2000)*, number 33 in CEUR-WS, pages 21–30, Aachen, Germany, August 2000. RWTH Aachen. Proceedings online available from `http://SunSITE.Informatik.RWTH-Aachen.DE/Publications/CEUR-WS/Vol-33/`.