Decidable and Undecidable Fragments of First-Order Concatenation Theory

Lars Kristiansen[1,2] and Juvenal Murwanashyaka[1]

[1] Department of Mathematics, University of Oslo, Norway
[2] Department of Informatics, University of Oslo, Norway

Abstract. We identify a number of decidable and undecidable fragments of first-order concatenation theory. We also give a purely universal axiomatization which is complete for the fragments we identify. Furthermore, we prove some normal-form results.

1 Introduction

1.1 First-order Concatenation theory

First-order concatenation theory can be compared to first-order number theory, e.g., Peano Arithmetic or Robinson Arithmetic. The universe of a standard structure for first-order number theory is the set of natural numbers. The universe of a standard structure for first-order concatenation theory is a set of strings over some alphabet. A first-order language for number theory normally contains two binary functions symbols. In a standard structure these symbols will be interpreted as addition and multiplication. A first-order language for concatenation theory normally contains just one binary function symbol. In a standard structure this symbol will be interpreted as the operator that concatenates two stings. A classical first-order language for concatenation theory contains no other non-logical symbols apart from constant symbols.

In this paper we extend concatenation theory with a binary relation symbol and introduce bounded quantifiers analogous to the bounded quantifiers $(\forall x \leq t)\phi$ and $(\exists x \leq t)\phi$ we know from number theory. Before we go on and state our main results, we will explain some notation and state a few basic definitions.

1.2 Notation and Basic Definitions

We will use $\mathbf{0}$ and $\mathbf{1}$ to denote respectively the bits zero and one, and we use pretty standard notation when we work with bit strings: $\{\mathbf{0}, \mathbf{1}\}^*$ denotes the set of all finite bit strings; $(b)_i$ denotes the i^{th} bit of the bit string b; and $\mathbf{0}\mathbf{1}^3\mathbf{0}^2\mathbf{1}$ denotes the bit string $\mathbf{0111001}$. The set $\{\mathbf{0}, \mathbf{1}\}^*$ contains the empty string which we will denote ε.

Let $\mathcal{L}_{BT}$ denote the first-order language that consist of the constants symbols $e, 0, 1$, the binary function symbol $\circ$ and the binary relation symbol $\sqsubseteq$. We will consider two $\mathcal{L}_{BT}$-structures named $\mathfrak{B}$ and $\mathfrak{D}$.

The universe of $\mathfrak{B}$ is the set $\{\mathbf{0}, \mathbf{1}\}^*$. The constant symbol 0 is interpreted as the string containing nothing but the bit $\mathbf{0}$, and the constant symbol 1 is interpreted as the string containing nothing but the bit $\mathbf{1}$, that is, $0^{\mathfrak{B}} = \mathbf{0}$ and $1^{\mathfrak{B}} = \mathbf{1}$. The constant symbol e is interpreted as the empty string, that is, $e^{\mathfrak{B}} = \varepsilon$. Moreover, $\circ^{\mathfrak{B}}$ is the function that concatenates two strings (e.g. $\mathbf{01} \circ^{\mathfrak{B}} \mathbf{000} = \mathbf{01000}$ and $\varepsilon \circ^{\mathfrak{B}} \varepsilon = \varepsilon$). Finally, $\sqsubseteq^{\mathfrak{B}}$ is the substring relation, that is, $u \sqsubseteq^{\mathfrak{B}} v$ iff there exists bit strings x, y such that $xuy = v$.

The structure $\mathfrak{D}$ is the same structure as $\mathfrak{B}$ with one exception: the relation $u \sqsubseteq^{\mathfrak{D}} v$ holds iff u is a prefix of v, that is, iff there exists a bit string x such that $ux = v$. To improve the readability we will use the symbol $\preceq$ in place of the symbol $\sqsubseteq$ when we are working in the structure $\mathfrak{D}$. Thus, $u \sqsubseteq v$ should be read as "u is a substring of v", whereas $u \preceq v$ should be read as "u is a prefix of v". When we do not have a particular structure in mind, e.g. when we deal with syntactical matters, we will stick to the symbol $\sqsubseteq$.

We introduce the *bounded quantifiers* $(\exists x \sqsubseteq t)\alpha$ and $(\forall x \sqsubseteq t)\alpha$ as syntactical abbreviations for respectively $(\exists x)[x \sqsubseteq t \wedge \alpha]$ and $(\forall x)[x \sqsubseteq t \rightarrow \alpha]$ (x is of course not allowed to occur in the term t), and we define the Σ-formulas inductively by

- α and $\neg\alpha$ are Σ-formulas if α is of the form $s \sqsubseteq t$ or of the form $s = t$ where s and t are terms
- $\alpha \vee \beta$ and $\alpha \wedge \beta$ are Σ-formulas if α and β are Σ-formulas
- $(\exists x \sqsubseteq t)\alpha$ and $(\forall x \sqsubseteq t)\alpha$ and $(\exists x)\alpha$ are Σ-formulas if α is a Σ-formula.

We assume that the reader notes the similarities with first-order number theory. The formulas that correspond to Σ-formulas in number theory are often called Σ_1-formulas or Σ_1^0-formulas. Next we introduce the biterals. The biterals correspond to the numerals of first-order number theory. Let b be a bit string. We define the *biteral* $\overline{b}$ by $\overline{\varepsilon} = e$, $\overline{b\mathbf{0}} = \overline{b} \circ \mathbf{0}$ and $\overline{b\mathbf{1}} = \overline{b} \circ \mathbf{1}$.

A Σ-formula ϕ is called a $\Sigma_{n,m,k}$-formula if it contains n unbounded existential quantifiers, m bounded existential quantifiers and k bounded universal quantifiers. A *sentence* is a formula with no free variables. The *fragment* $\Sigma^{\mathfrak{B}}_{n,m,k}$ ($\Sigma^{\mathfrak{D}}_{n,m,k}$) is the set of $\Sigma_{n,m,k}$-sentences true in $\mathfrak{B}$ (respectively, $\mathfrak{D}$).

To improve the readability we will skip the operator $\circ$ in first-order formulas and simply write st in place of $s \circ t$. Furthermore, we will occasionally contract quantifiers and write, e.g., $\forall w_1, w_2 \sqsubseteq u[\phi]$ in place of $(\forall w_1 \sqsubseteq u)(\forall w_2 \sqsubseteq u)\phi$, and for $\sim \in \{\preceq, \sqsubseteq, =\}$, we write $s \not\sim t$ in place of $\neg s \sim t$.

1.3 Main Results and Related Work

We prove that the fragment $\Sigma^{\mathfrak{B}}_{0,m,k}$ is decidable (for any $m, k \in \mathbb{N}$), and we prove that $\Sigma^{\mathfrak{B}}_{1,2,1}$ and $\Sigma^{\mathfrak{B}}_{1,0,2}$ are undecidable. Furthermore, we prove that the

The Axioms of B

1. $\forall x[\ x = ex \land x = xe\ ]$
2. $\forall xyz[\ (xy)z = x(yz)\ ]$
3. $\forall xy[\ (x \neq y) \rightarrow (\ (x0 \neq y0) \land (x1 \neq y1)\ )\ ]$
4. $\forall xy[\ x0 \neq y1\ ]$
5. $\forall x[\ x \sqsubseteq e \leftrightarrow x = e\ ]$
6. $\forall x[\ x \sqsubseteq 0 \leftrightarrow (x = e \lor x = 0)\ ]$
7. $\forall x[\ x \sqsubseteq 1 \leftrightarrow (x = e \lor x = 1)\ ]$
8. $\forall xy[\ x \sqsubseteq 0y0 \leftrightarrow (x = 0y0 \lor x \sqsubseteq 0y \lor x \sqsubseteq y0)\ ]$
9. $\forall xy[\ x \sqsubseteq 0y1 \leftrightarrow (x = 0y1 \lor x \sqsubseteq 0y \lor x \sqsubseteq y1)\ ]$
10. $\forall xy[\ x \sqsubseteq 1y0 \leftrightarrow (x = 1y0 \lor x \sqsubseteq 1y \lor x \sqsubseteq y0)\ ]$
11. $\forall xy[\ x \sqsubseteq 1y1 \leftrightarrow (x = 1y1 \lor x \sqsubseteq 1y \lor x \sqsubseteq y1)\ ]$

Fig. 1. These are the axioms of the first-order theory B.

fragments $\Sigma^{\mathfrak{D}}_{0,m,k}$ and $\Sigma^{\mathfrak{D}}_{n,m,0}$ are decidable (for any $n, m, k \in \mathbb{N}$), and we prove that $\Sigma^{\mathfrak{D}}_{3,0,2}$ and $\Sigma^{\mathfrak{D}}_{4,1,1}$ are undecidable. Our results on decidable fragments are corollaries of theorems that have an interest in their own right: We prove the existence of several normal forms, and we give a purely universal axiomatization of concatenation theory which is Σ-complete.

Recent related work can be found in Halfon et al. [6], Day et al. [2], Ganesh et al. [3], Karhumäki et al. [8] and several other places, see Section 6 of [3] for further references.

The material in Section 8 of the textbook Leary & Kristiansen [10] is also related to the research presented in this paper. So is a series of papers that starts with with Grzegorczyk [4] and includes Grzegorczyk & Zdanowski [5], Visser [13] and Horihata [7]. These papers deal with the essential undecidability of various first-order theories of concatenation. The relationship between the various axiomatizations of concatenation theory we find in these papers and the axiomatization we give below has not yet been investigated.

The theory of concatenation seems to go back to work of Tarski and Quine, see Visser [13] for a brief account of its history.

2 Σ-complete Axiomatizations

Theorem 1. *Let B and D be the set of axioms given in respectively Figure 1 and Figure 2. For any Σ-sentence ϕ, we have*

$$\mathfrak{B} \models \phi \ \Rightarrow \ B \vdash \phi \quad \textit{and} \quad \mathfrak{D} \models \phi \ \Rightarrow \ D \vdash \phi$$

Proof. We give a brief sketch of the (very long) proof of $\mathfrak{B} \models \phi \Rightarrow B \vdash \phi$. The proof of $\mathfrak{D} \models \phi \Rightarrow D \vdash \phi$ is similar. Full proofs can found in Kristiansen & Murwanashyaka [9].

The Axioms of D

- the first four axioms are the same as the first four axioms of B
5. $\forall x[\ x \preceq e \leftrightarrow x = e\ ]$
6. $\forall xy[\ x \preceq y0 \leftrightarrow (x = y0 \vee x \preceq y)\ ]$
7. $\forall xy[\ x \preceq y1 \leftrightarrow (x = y1 \vee x \preceq y)\ ]$

Fig. 2. These are the axioms of the first-order theory D.

Prove (by induction on the structure of t) that there for any variable-free $\mathcal{L}_{BT}$-term t exists a biteral b such that

$$\mathfrak{B} \models t = b \ \Rightarrow\ B \vdash t = b\ . \tag{1}$$

Prove (by induction on the structure of b_2) that we for any biterals b_1 and b_2 have

$$\mathfrak{B} \models b_1 \neq b_2 \ \Rightarrow\ B \vdash b_1 \neq b_2\ . \tag{2}$$

Use $B \vdash \forall x[x0 \neq e \wedge x1 \neq e]$ when proving (2). Furthermore, prove (by induction on the structure of b_2) that we for any biterals b_1 and b_2 have

$$\mathfrak{B} \models b_1 \sqsubseteq b_2 \ \Rightarrow\ B \vdash b_1 \sqsubseteq b_2 \quad \text{and} \quad \mathfrak{B} \models b_1 \not\sqsubseteq b_2 \ \Rightarrow\ B \vdash b_1 \not\sqsubseteq b_2\ . \tag{3}$$

It follows from (1), (2) and (3) that we have

$$\mathfrak{B} \models \phi \ \Rightarrow\ B \vdash \phi\ . \tag{4}$$

for any ϕ of one of the four forms $t_1 = t_2$, $t_1 \neq t_2$, $t_1 \sqsubseteq t_2$, and $t_1 \not\sqsubseteq t_2$ where t_1 and t_2 are variable-free terms.

Use induction on the structure of b to prove the following claim:

> *If $\phi(x)$ is an $\mathcal{L}_{BT}$-formula such that we have $\mathfrak{B} \models \phi(b) \Rightarrow B \vdash \phi(b)$ for any biteral b, then we also have*
>
> $$\mathfrak{B} \models (\forall x \sqsubseteq b)\phi(x) \Rightarrow B \vdash (\forall x \sqsubseteq b)\phi(x)$$
>
> *for any biteral b.*

Finally, prove (by induction on the structure of ϕ) that we for any Σ-sentence ϕ have $\mathfrak{B} \models \phi \Rightarrow B \vdash \phi$. Use (4) in the base cases, that is, when ϕ is an atomic sentence or a negated atomic sentence. Use the claim and (1) in the case ϕ is of the form $(\forall x \sqsubseteq t)\psi$. The remaining cases are rather straightforward. $\square$

Corollary 2. *The fragments $\Sigma^{\mathfrak{B}}_{0,m,k}$ and $\Sigma^{\mathfrak{D}}_{0,m,k}$ are decidable (for any $m, k \in \mathbb{N}$).*

Proof. We prove that $\Sigma^{\mathfrak{B}}_{0,m,k}$ is decidable. Let ϕ be a $\Sigma_{0,m,k}$-formula. The negation of a $\Sigma_{0,m,k}$-formula is logically equivalent to a $\Sigma_{0,k,m}$-formula (by De Morgan's laws). We can compute a $\Sigma_{0,k,m}$-formula ϕ' which is logically equivalent to $\neg\phi$. By Theorem 1, we have $B \vdash \phi$ if $\mathfrak{B} \models \phi$, and we have $B \vdash \phi'$ if $\mathfrak{B} \models \neg\phi$. The set of formulas derivable from the axioms of B is computably enumerable. Hence it is decidable if ϕ is true in $\mathfrak{B}$. The proof that the fragment $\Sigma^{\mathfrak{D}}_{0,m,k}$ is decidable is similar. □

3 Normal Forms

A proof of the next lemma can be found several places, see e.g. Büchi & Senger [1] or the proof of Theorem 6 in Karhumäki et al. [8]. The lemma is also proved in [9].

Lemma 3. *Let $\mathfrak{A} \in \{\mathfrak{B}, \mathfrak{D}\}$, and let s_1, s_2, t_1, t_2 be $\mathcal{L}_{BT}$-terms. There exist $\mathcal{L}_{BT}$-terms s, t and variables $v_0, ..., v_k$ such that*

(1) $\mathfrak{A} \models (s_1 = t_1 \wedge s_2 = t_2) \;\leftrightarrow\; s_1 0 s_2 s_1 1 s_2 = t_1 0 t_2 t_1 1 t_2$
(2) $\mathfrak{A} \models (s_1 = t_1 \vee s_2 = t_2) \leftrightarrow \exists v_0 \dots v_k [s = t]$
(3) $\mathfrak{A} \models (\neg s_1 = t_1) \leftrightarrow \exists v_0 \dots v_k [s = t]$.

Lemma 4. *Let s_1, t_1 be $\mathcal{L}_{BT}$-terms. There exist $\mathcal{L}_{BT}$-terms s, t and variables $v_1, \dots, v_k$ such that*

(1) $\mathfrak{D} \models s_1 \preceq t_1 \leftrightarrow \exists v_1 [s_1 v_1 = t_1]$ *and (2)* $\mathfrak{D} \models (s_1 \not\preceq t_1) \leftrightarrow \exists v_1 \dots v_k [s = t]$.

Proof. It is obvious that (1) holds. Furthermore, the formula $s_1 \not\preceq t_1$ is equivalent in $\mathfrak{D}$ to the formula

$$(t_1 \preceq s_1 \;\wedge\; t_1 \neq s_1) \;\vee\; \exists xyz[(t_1 = x0y \;\wedge\; s_1 = x1z) \;\vee\; (t_1 = x1y \;\wedge\; s_1 = x0z)]\;.$$

Thus, (2) follows by Lemma 3 and (1). □

Comment: It is not known to us whether the bounded universal quantifier that appears in clause (2) of the next lemma can be eliminated.

Lemma 5. *Let s_1, t_1 be $\mathcal{L}_{BT}$-terms. There exist $\mathcal{L}_{BT}$-terms s, t and variables $v_1, \dots, v_k$ such that (1)* $\mathfrak{B} \models s_1 \sqsubseteq t_1 \leftrightarrow \exists v_1 v_2 [t_1 = v_1 s_1 v_2]$ *and*

(2) $\mathfrak{B} \models s_1 \not\sqsubseteq t_1 \leftrightarrow \forall v_1 \sqsubseteq t_1 \exists v_2 \dots v_k [s = t]$.

Proof. Cause (1) is trivial. Furthermore, observe that $s_1 \not\sqsubseteq t_1$ is equivalent in $\mathfrak{B}$ to the formula $(\forall v \sqsubseteq t_1)\alpha$ where α is

$$\exists x[\; t_1 x = v s_1 \;\wedge\; x \neq e \;] \;\vee\; \exists xyz[\; (t_1 = x0y \;\wedge\; v s_1 = x1z) \;\vee$$
$$(t_1 = x1y \;\wedge\; v s_1 = x0z) \;]\;.$$

If we let $v s_1 \preceq t_1$ abbreviate $\exists x[v s_1 x = t]$, then α can be written as $v s_1 \not\preceq t_1$. Thus, (2) follows by Lemma 3. □

Theorem 6 (Normal Form Theorem I). *Any Σ-formula ϕ is equivalent in $\mathfrak{D}$ to a $\mathcal{L}_{BT}$-formula ϕ' of the form*

$$\phi' \equiv (\mathbf{Q}_1^{t_1} v_1) \dots (\mathbf{Q}_m^{t_m} v_m)(s = t)$$

where $t_1, .., t_m, s, t$ are $\mathcal{L}_{BT}$-terms and $\mathbf{Q}_j^{t_j} v_j \in \{\exists v_j, \exists v_j \preceq t_j, \forall v_j \preceq t_j\}$ for $j = 1, \dots, m$. Moreover, if ϕ does not contain bounded universal quantifiers, then ϕ' does not contain bounded universal quantifiers.

Proof. We proceed by induction on the structure of ϕ (throughout the proof we reason in the structure $\mathfrak{D}$). Suppose ϕ is an atomic formula or the negation of an atomic formula. If ϕ is of the form $s = t$, let ϕ' be $s = t$. Use Lemma 3(3) if ϕ is of the form $s \neq t$. Use Lemma 4 if ϕ is of one of the forms $s \preceq t$ and $s \not\preceq t$.

Suppose ϕ is of the form $\alpha \wedge \beta$. By our induction hypothesis, we have formulas

$$\alpha' \equiv (\mathbf{Q}_1^{t_1} x_1) \dots (\mathbf{Q}_k^{t_k} x_k)(s_1 = t_1) \quad \text{and} \quad \beta' \equiv (\mathbf{Q}_1^{s_1} y_1) \dots (\mathbf{Q}_m^{s_m} y_m)(s_2 = t_2)$$

which are equivalent to respectively α and β. Thus, ϕ is equivalent to a formula of the form $(\mathbf{Q}_1^{t_1} x_1) \dots (\mathbf{Q}_k^{t_k} x_k)(\mathbf{Q}_1^{s_1} y_1) \dots (\mathbf{Q}_m^{s_m} y_m)(s_1 = t_1 \wedge s_2 = t_2)$. By Lemma 3(1), we have a formula ϕ' of the desired form which is equivalent to ϕ. The case when ϕ is of the form $\alpha \vee \beta$ is similar. Use clause (2) of Lemma 3 in place of clause (1).

The theorem follows trivially from the induction hypothesis when ϕ is of one of the forms $(\exists v)\alpha$, $(\forall v \preceq t)\alpha$ and $(\exists v \preceq t)\alpha$. □

Theorem 7 (Normal Form Theorem II). *Any Σ-formula ϕ is equivalent in $\mathfrak{B}$ to a $\mathcal{L}_{BT}$-formula ϕ' of one of the forms*

$$\phi' \equiv (\mathbf{Q}_1^{t_1} v_1) \dots (\mathbf{Q}_m^{t_m} v_m)\,(s = t) \quad \text{or} \quad \phi' \equiv (\exists v)(\mathbf{Q}_1^{t_1} v_1) \dots (\mathbf{Q}_m^{t_m} v_m)\,(s = t)$$

where $t_1, .., t_m, s, t$ are $\mathcal{L}_{BT}$-terms and $\mathbf{Q}_j^{t_j} v_j \in \{\exists v_j \sqsubseteq t_j, \forall v_j \sqsubseteq t_j\}$ for $j = 1, \dots, m$.

Proof. Proceed by induction on the structure of ϕ. This proof is similar to the proof of Theorem 6. A formula of the form $(\forall x \sqsubseteq t)(\exists y)\alpha$ is equivalent (in $\mathfrak{B}$) to one of the form $(\exists z)(\forall x \sqsubseteq t)(\exists y \sqsubseteq z)\alpha$, a formula of the form $(\exists x \sqsubseteq t)(\exists y)\alpha$ is equivalent to one of the form $(\exists y)(\exists x \sqsubseteq t)\alpha$, and a formula of the form $(\exists x)(\exists y)\alpha$ is equivalent to one of the form $(\exists z)(\exists x \sqsubseteq z)(\exists y \sqsubseteq z)\alpha$. Thus, the resulting normal form will contain maximum one unbounded existential quantifier. □

Corollary 8. *The fragment $\Sigma_{n,m,0}^{\mathfrak{D}}$ is decidable (for any $n, m \in \mathbb{N}$).*

Proof. By Theorem 6, any $\Sigma_{n,m,0}$-sentence is equivalent in $\mathfrak{D}$ to a sentence of the normal form $(\exists v_1) \dots (\exists v_k)(s = t)$ (regard the bounded existential quantifiers as unbounded) . The transformation of a $\Sigma_{n,m,0}$-formula into an equivalent formula (in $\mathfrak{D}$) of normal form is constructive. Makanin [11] has proved that it is decidable whether an equation on the form

$$a_n x_n \ldots a_1 x_1 a_0 = b_m y_m \ldots b_1 y_1 b_0$$

where $a_1, ..., a_n, b_1, ..., b_m \in \{\mathbf{0}, \mathbf{1}\}^*$, has a solution in $\{\mathbf{0}, \mathbf{1}\}^*$. It follows that the fragment $\Sigma^{\mathfrak{D}}_{n,m,0}$ is decidable. □

We have not been able to prove that any $\Sigma_{n,m,0}$-sentence is equivalent in $\mathfrak{B}$ to a sentence of the form $(\exists v_1) \ldots (\exists v_k)(s = t)$. See the comment immediately before Lemma 5. Thus, we cannot use Makanin's [11] result to prove that the fragment $\Sigma^{\mathfrak{B}}_{n,m,0}$ is decidable.

Open Problem: Is the fragment $\Sigma^{\mathfrak{B}}_{n,m,0}$ decidable (for any $n, m \in \mathbb{N}$)?

4 Undecidable Fragments

Definition 9. Post's Correspondence Problem, *henceforth* PCP, *is given by*

- *Instance: a list of pairs* $\langle b_1, b'_1 \rangle, \ldots, \langle b_n, b'_n \rangle$ *where* $b_i, b'_i \in \{\mathbf{0}, \mathbf{1}\}^*$
- *Solution: a finite nonempty sequence* $i_1, ..., i_m$ *of indexes such that*

$$b_{i_1} b_{i_2} \ldots b_{i_m} = b'_{i_1} b'_{i_2} \ldots b'_{i_m} \ .$$

We define the map $N : \{\mathbf{0}, \mathbf{1}\}^* \to \{\mathbf{0}, \mathbf{1}\}^*$ *by* $N(\varepsilon) = \varepsilon$, $N(\mathbf{0}) = \mathbf{010}$, $N(\mathbf{1}) = \mathbf{01}^2\mathbf{0}$, $N(b\mathbf{0}) = N(b)N(\mathbf{0})$ *and* $N(b\mathbf{1}) = N(b)N(\mathbf{1})$.

It is proved in Post [12] that PCP is undecidable. The proof of the next lemma is left to the reader.

Lemma 10. *The instance* $\langle b_1, b'_1 \rangle, \ldots, \langle b_n, b'_n \rangle$ *of* PCP *has a solution iff the instance* $\langle N(b_1), N(b'_1) \rangle, \ldots, \langle N(b_n), N(b'_n) \rangle$ *has a solution.*

We will now explain the ideas behind our proofs of the next few theorems. Given the lemma above, it is not very hard to see that an instance $\langle g_1, g'_1 \rangle, \ldots, \langle g_n, g'_n \rangle$ of PCP has a solution iff there exists a bit string of the form

$$\mathbf{01}^5\mathbf{0} N(a_1) \mathbf{01}^4\mathbf{0} N(b_1) \mathbf{01}^5\mathbf{0} \ldots N(a_m) \mathbf{01}^4\mathbf{0} N(b_m) \mathbf{01}^5\mathbf{0} \qquad (*)$$

where

(A) $N(a_m) = N(b_m)$
(B) $N(a_1) = g_j$ and $N(b_1) = g'_j$ for some $1 \leq j \leq n$
(C) $N(a_{k+1}) = N(a_k)N(g_j)$ and $N(b_{k+1}) = N(b_k)N(g'_j)$ for some $1 \leq j \leq n$.

We also see that an instance $\langle g_1, g_1' \rangle, \ldots, \langle g_n, g_n' \rangle$ of PCP has a solution iff there exists a bit string s of the form (*) that satisfies

(a) there is $j \in \{1, \ldots, n\}$ such that $\mathbf{01}^5\mathbf{0}N(g_j)\mathbf{01}^4\mathbf{0}N(g_j')\mathbf{01}^5\mathbf{0}$ is an initial segment of s

(b) if

$$\mathbf{01}^5\mathbf{0}N(a)\mathbf{01}^4\mathbf{0}N(b)\mathbf{01}^5\mathbf{0}$$

is a substring of s, then either $N(a) = N(b)$, or there is $j \in \{1, \ldots, n\}$ such that

$$\mathbf{01}^5\mathbf{0}N(a)N(g_j)\mathbf{01}^4\mathbf{0}N(b)N(g_j')\mathbf{01}^5\mathbf{0}$$

is a substring of s.

In the proof of Theorem 11 we give a formula which is true in $\mathfrak{D}$ iff there exists a string of the form (*) that satisfies (A), (B) and (C). In the proof of Theorem 12 we give formulas which are true in $\mathfrak{B}$ iff there exists a string of the form (*) that satisfies (a) and (b). In order to improve the readability of our formulas, we will write $\#$ in place of the biteral $\overline{\mathbf{01}^5\mathbf{0}}$ and $\mathbf{!}$ in place of the biteral $\overline{\mathbf{01}^4\mathbf{0}}$.

Theorem 11. *The fragment $\Sigma^{\mathfrak{D}}_{3,0,2}$ is undecidable.*

Proof. Let $\psi(x) \equiv (\forall z \preceq x)(z\overline{\mathbf{1}^4} \not\preceq x)$. Observe that ψ contains one bounded universal quantifier. Observe that $\psi(\overline{b})$ is true in $\mathfrak{D}$ iff the bit string b does not contain 4 consecutive ones. Furthermore, let $\phi_n(x_1, ..., x_n, y_1, ..., y_n) \equiv$

$$\begin{aligned}
&(\exists u)\Big(\left(\bigvee_{j=1}^{n} \#x_j\mathbf{!}y_j\# \preceq u \right) \;\wedge \\
&\quad (\forall v \preceq u)\Big[\; v\# \not\preceq u \;\;\vee\;\; v\# = u \;\;\vee\;\; (\exists w_1, w_2)\Big\{\; v\#w_1\mathbf{!}w_2\# \preceq u \wedge \\
&\quad \psi(w_1 w_2) \;\wedge\; \left[\; w_1 = w_2 \;\vee\; \left(\bigvee_{j=1}^{n} v\#w_1\mathbf{!}w_2\#w_1x_j\mathbf{!}w_2y_j\# \preceq u \right) \right] \Big\} \Big] \Big) .
\end{aligned}$$

Let $\langle g_1, g_1' \rangle, \ldots, \langle g_n, g_n' \rangle$ be an instance of PCP. We have

$$\mathfrak{D} \models \phi_n(\overline{N(g_1)}, \ldots, \overline{N(g_n)}, \overline{N(g_1')}, \ldots, \overline{N(g_n')})$$

iff there exists a bit sting of the form (*) that satisfies (A), (B) and (C) iff the instance $\langle g_1, g_1' \rangle, \ldots, \langle g_n, g_n' \rangle$ has a solution. Furthermore ϕ_n is a $\Sigma_{3,0,2}$-formula. It follows that the fragment $\Sigma^{\mathfrak{D}}_{3,0,2}$ is undecidable. $\square$

Theorem 12. *The fragments $\Sigma^{\mathfrak{B}}_{1,2,1}$ and $\Sigma^{\mathfrak{B}}_{1,0,2}$ are undecidable.*

Proof. Let $\vec{x} = x_1, \dots, x_n$, let $\vec{y} = y_1, \dots, y_n$ and let

$$\alpha(\vec{x}, \vec{y}, z) \;\equiv\; \left(\bigvee_{j=1}^{n} \#x_j\mathbf{!}y_j\# \sqsubseteq z \;\wedge\; 0\#x_j\mathbf{!}y_j\# \not\sqsubseteq z \;\wedge\; 1\#x_j\mathbf{!}y_j\# \not\sqsubseteq z \right) .$$

Consider the $\Sigma_{1,2,1}$-formula $\psi_n(\vec{x}, \vec{y}) \equiv$

$$\begin{aligned}
&(\exists u)\Big(\alpha(\vec{x}, \vec{y}, u) \wedge \\
&\quad (\forall v \sqsubseteq u)\Big[\#v\# \not\sqsubseteq u \;\vee\; \overline{\mathbf{1}^5} \sqsubseteq v \;\vee\; (\exists w_1, w_2 \sqsubseteq v)\Big\{ v = w_1\mathbf{!}w_2 \\
&\quad \wedge\; \overline{\mathbf{1}^4} \not\sqsubseteq w_1 \;\wedge\; \overline{\mathbf{1}^4} \not\sqsubseteq w_2 \;\wedge\; \left[w_1 = w_2 \;\vee\; \left(\bigvee_{j=1}^{n} \#w_1x_j\mathbf{!}w_2y_j\# \sqsubseteq u \right) \right] \Big\} \Big] \Big)
\end{aligned}$$

and consider the $\Sigma_1^{1,0,2}$-formula $\gamma_n(\vec{x}, \vec{y}) \equiv$

$$\begin{aligned}
&(\exists u)\Big(\alpha(\vec{x}, \vec{y}, u) \;\wedge\; (\forall w_1, w_2 \sqsubseteq u)\Big\{ \#w_1\mathbf{!}w_2\# \not\sqsubseteq u \;\vee\; \overline{\mathbf{1}^4} \sqsubseteq w_1w_2 \\
&\qquad\qquad \vee\;\; w_1 = w_2 \;\vee\; \left(\bigvee_{j=1}^{n} \#w_1x_j\mathbf{!}w_2y_j\# \sqsubseteq u \right) \Big\} \Big).
\end{aligned}$$

Let $\langle g_1, g_1' \rangle, \dots, \langle g_n, g_n' \rangle$ be an instance of PCP. We have

$$\mathfrak{B} \models \psi_n(\overline{N(g_1)}, \dots, \overline{N(g_n)}, \overline{N(g_1')}, \dots, \overline{N(g_n')})$$

iff

$$\mathfrak{B} \models \gamma_n(\overline{N(g_1)}, \dots, \overline{N(g_n)}, \overline{N(g_1')}, \dots, \overline{N(g_n')})$$

iff there exists a bit sting of the form (*) that satisfies (a) and (b) iff the instance $\langle g_1, g_1' \rangle, \dots, \langle g_n, g_n' \rangle$ has a solution. It follows that the fragments $\Sigma_{1,2,1}^{\mathfrak{B}}$ and $\Sigma_{1,0,2}^{\mathfrak{B}}$ are undecidable. □

The proof of the next theorem is based on the following idea: The instance $\langle g_1, g_1' \rangle, \dots, \langle g_n, g_n' \rangle$ of PCP has a solution iff there exists a bit string of the form

$$\begin{aligned}
&\mathbf{01}^5\mathbf{0}N(a_1)\mathbf{01}^4\mathbf{0}N(b_1)\mathbf{01}^6\mathbf{0}N(a_2)\mathbf{01}^4\mathbf{0}N(b_2)\mathbf{01}^7\mathbf{0}\dots \\
&\qquad\qquad \dots\mathbf{01}^{5+m-1}\mathbf{0}N(a_m)\mathbf{01}^4\mathbf{0}N(b_m)\mathbf{01}^{5+m}\mathbf{0}
\end{aligned}$$

with the properties (A), (B) and (C) given above.

Theorem 13. *The fragment $\Sigma_{4,1,1}^{\mathfrak{D}}$ is undecidable.*

Proof. Let $\mathbf{!}^k \equiv \overline{\mathbf{0}\mathbf{1}^k\mathbf{0}}$. The $\Sigma_{4,1,1}$-formula

$$(\exists u)\Big(\left(\bigvee_{j=1}^{n} \mathbf{!}^5 x_j \mathbf{!}^4 y_j \mathbf{!}^6 \preceq u \right) \ \wedge \ (\forall v \preceq u)\Big[\ v\overline{\mathbf{1}^5\mathbf{0}} \not\preceq u \ \ \vee \ \ v = 0 \ \vee$$

$$(\exists w_1, w_2, y)(\exists z \preceq v)\Big\{ \ v = z0y\overline{\mathbf{1}^5\mathbf{0}}w_1\mathbf{!}^4 w_2 01y \ \wedge \ 1y = y1 \ \wedge$$

$$\left[w_1 = w_2 \ \vee \ \left(\bigvee_{j=1}^{n} v\overline{\mathbf{1}^5\mathbf{0}}w_1 x_j \mathbf{!}^4 w_2 y_j 011y\overline{\mathbf{1}^5\mathbf{0}} \preceq u \right) \right] \Big\} \ \Big] \ \Big)$$

yields the desired statement. Note that y is a solution of the equation $\mathbf{1}y = y\mathbf{1}$ iff $y \in \{\mathbf{1}\}^*$. □

References

1. Büchi, J. R. and Senger, S.: *Coding in the existential theory of concatenation.* Arch. math. Logik **26** (1986/7), 101-106.
2. Day, J., Ganesh, V., He, P., Manea, F. and Nowotka, D.: *The satisfiability of extended word equations: The boundary between decidability and undecidability.* arXiv:1802.00523 (2018).
3. Ganesh, V., Minnes, M., Solar-Lezama, A. and Rinard, M. C.: *Word equations with length constraints: Whats decidable?* In: Biere A., Nahir A., Vos T. (eds) Hardware and Software: Verification and Testing. HVC 2012. Lecture Notes in Computer Science, vol 7857, pp. 209-226. Springer, Berlin, Heidelberg.
4. Grzegorczyk, A.: *Undecidability without arithmetization.* Studia Logica **79** (2005), 163-230.
5. Grzegorczyk, A. and Zdanowski, K.: *Undecidability and concatenation.* pp. 72-91 in "Andrzej Mostowski and Foundational Studies" (eds. by Ehrenfeucht et al.), IOS, Amsterdam, 2008.
6. Halfon, S., Schnoebelen, P. and Zetzsche G: *Decidability, complexity, and expressiveness of first-order logic over the subword ordering.* In *Proc. LICS 2017.* IEEE Computer Society, 1-12.
7. Horihata, Y.: *Weak theories of concatenation and arithmetic.* Notre Dame Journal of Formal Logic, **53** (2012), 203-222.
8. Karhumäki, J., Mignosi, F. and Plandowski, W.: *The expressibility of languages and relations by word equations.* Journal of the ACM **47** (2000), 483-505.
9. Kristiansen, L. and Murwanashyaka, J.: *Notes on fragments of first-order concatenation theory.* arXiv:1804.06367 (2018).
10. Leary, C. and Kristiansen, L.: *A friendly introduction to mathematical logic.* 2nd Edition, Milne Library, SUNY Geneseo, Geneseo, NY, 2015.
11. Makanin, G. S.: *The problem of solvability of equations in a free semigroup.* Mathematics of the USSR-Sbornik **32** (1977), 129-198.
12. Post, E. L.: *A variant of a recursively unsolvable problem.* Bulletin of the American Mathematical Society **52** (1946), 264-268.
13. Visser, A.: *Growing commas. A study of sequentiality and concatenation.* Notre Dame Journal of Formal Logic, **50** (2009), 61-85.