The decision problem for some logics for finite words on infinite alphabets

Christian Choffrut *
http://www.liafa.jussieu.fr/∼cc
cc@liafa.jussieu.fr

Serge Grigorieff *
http://www.liafa.jussieu.fr/∼seg
seg@liafa.jussieu.fr

February 2008

Contents

Abstract

This paper is a follow-up of a previous paper where the logical characterization of Eilenberg, Elgot and Shepherdson of n-ary synchronous relations was investigated in the case where the alphabet has infinitely

*LIAFA, Université Paris 7 & CNRS, 2, pl. Jussieu 75251 Paris Cedex 05

many letters. Here we show that modifying one of the predicate leads to a completely different picture for infinite alphabets though it does not change the expressive power for finite alphabets. Indeed, roughly speaking, being able to express the fact that two words end with the same symbol leads to an undecidable theory, already for the Σ_2 fragment. Finally, we show that the existential fragment is decidable.

1 Introduction

The purpose of this work is to investigate different theories of the free monoid in the case where it has an infinite, denumerable number of generators.

The study of the theories of the free *finitely generated* monoid, i.e. words on a finite alphabet, dates back to the late sixties and has a wide range of applications in computer science. Recently, infinite alphabets have also been considered in several applications, e.g., in database theory [1, 2] and model checking [6].

It is well-known (Quine, [14] 1946) that adding concatenation leads to an undecidable theory when the free monoid contains at least two generators, so this operation is not considered directly. Variants of elementary predicates are considered which are related to the underlying partial ordering: "u is a prefix of v" (denoted by $u \leq_{\mathtt{pref}} v$), the quasi-ordering: "u has length less than v (denoted by $\mathtt{EqLen}(u, v)$), and the last letter of a word: "u ends with a specific letter a" (denoted by $\mathtt{Last}_a(u)$). This is in essence the structure studied by Eilenberg, Elgot, Shepherdson in [8], 1969, where the authors characterize "à la Büchi" the definable relations in terms of the so-called synchronous automata, and reprove the decidability of the theory (a result first established by Shepherdson, [15] 1959).

Before presenting our contribution let us go through special features of the free monoid generated by a denumerable infinite alphabet to be found in the literature.

1. Vazhenin & Rozenblat, 1981 [18], proved that, for an infinite alphabet, the positive theory of concatenation is decidable. On the opposite, the positive theory of concatenation over a finite alphabet is undecidable (Quine, 1946 [14]). Even the $\forall\exists^3$-positive theory is undecidable (Durnev, 1995 [7]).

2. Let's add to finite automata, registers able to memorize any letter of the alphabet and let's allow comparison between the letter currently read and the contents of the registers. In case of finite alphabets, such

automata are equivalent to the usual ones. However, for an infinite alphabet, this really matters: the universal problem for such non-deterministic register automata is undecidable (Neven & Schwentick & Vianu, 2001 [11, 12]).

3. Answering a question of [8], we proved in [5] that if Σ is infinite then the predicate $\texttt{EqLenEqLast} = \{(ua, va) \mid u, v \in \Sigma^*, |u| = |v|, a \in \Sigma\}$ is not definable with $\texttt{EqLen}$, $\leq_{\texttt{pref}}$ and the $\texttt{Last}_a$'s, $a \in \Sigma$.
Also, if R is definable with $\texttt{EqLen}$, $\leq_{\texttt{pref}}$ and the $\texttt{Last}_a$'s, then
(i) there exists a smallest finite $\Sigma_0 \subset \Sigma$ such that R is definable with $\texttt{EqLen}$, $\leq_{\texttt{pref}}$ and the sole $\texttt{Last}_a$'s where $a \in \Sigma_0$.
(ii) if $\Sigma_0 \neq \emptyset$ then R is definable with $\texttt{EqLen}$, $\leq_{\texttt{pref}}$ and the sole $\texttt{Last}_a$'s where $a \in \Sigma_0$, if and only if, R is invariant under all permutations of Σ which are the identity on Σ_0.
(iii) R is definable with $\texttt{EqLen}$, $\leq_{\texttt{pref}}$ and $|u| \equiv k \mod \ell$ for all $k < \ell \in \mathbb{N}$ (and with no $\texttt{Last}_a$) if and only if R is invariant under all permutations of Σ.
All these results are false for finite alphabets Σ : $\texttt{EqLenEqLast}$ is definable with $\texttt{EqLen}$ and the $\texttt{Last}_a$'s and disproves (iii). Also, $R_{\Sigma_0} = \{xx \mid x \in \Sigma \setminus \Sigma_0\}$ disproves (ii) if $\Sigma \setminus \Sigma_0$ has at least two letters.

The purpose of this paper is to add some new results in that vein. First, let's fix some notations.

Notation 1. Let Σ be an infinite alphabet.

- ε, $\leq_{\texttt{pref}}$ and $\texttt{Pred} : \Sigma^* \to \Sigma^*$ respectively denote the empty word, the prefix ordering on Σ^* and the map such that $\texttt{Pred}(\varepsilon) = \varepsilon$ and $\texttt{Pred}(a_1 \dots a_n) = a_1 \dots a_{n-1}$ (both ε and $\texttt{Pred}$ are definable with $\leq_{\texttt{pref}}$).

- For $k \in \mathbb{Z}$, $\texttt{EqLen}_k$ denotes the binary relation $\{(u, v) \mid |u| = |v| + k\}$ where $|u|$ is the length of u. We also write $\texttt{EqLen}$ in place of $\texttt{EqLen}_0$.

- For $a \in \Sigma$, $\texttt{Last}_a$ denotes the unary relation $\Sigma^* a$ and $\texttt{EqLast}$ denotes the binary relation $\{(ua, va) \mid u, v \in \Sigma^*, a \in \Sigma\}$, i.e. the set of pairs of words which end with the same letter.

In [5], we considered the logic with predicates $\leq_{\texttt{pref}}$, $\texttt{EqLen}$ and the $\texttt{Last}_a$'s, $a \in \Sigma$, and its extension with $\texttt{EqLenEqLast}$ (cf. point 3 above), for words in an infinite alphabet, and introduced the notions of synchronous and "oblivious synchronous" automata which characterize them. As a consequence, we derived the decidability of these logics.

Here we go one step further by considering the predicate $\mathtt{EqLast}$. The picture in that case is completely different since the theory is no longer decidable, a big departure from the case of finite alphabets. More precisely we are able to prove that the existential fragment is decidable while the Σ_2 fragment is undecidable.

Our main results are stated in the next two Theorems. The first result establishes the undecidability of the theory. §2 is devoted to the proof of Point 1. A strong version of Point 2 is given in §5 as Theorem 15.
In contrast, remember that in order to get an undecidable theory in case the alphabet is finite and has at least two letters, one has to consider stronger languages obtained by adding the predicate "u is a suffix of v" (cf. Büchi, [3] 1960), or "u is a factor of v" .

Theorem 1. *Let Σ be an infinite alphabet.*

1. The $\exists\forall\forall$ theory of the structure $\langle \Sigma^*; \leq_{\mathtt{pref}}, \varepsilon, \mathtt{Pred}, \mathtt{EqLast}\rangle$ is undecidable.

2. The Σ_1 theory of $\langle \Sigma^*; \leq_{\mathtt{pref}}, \varepsilon, \mathtt{Pred}, (\mathtt{EqLen}_k)_{k\in\mathbb{Z}}, \mathtt{EqLast}, (\mathtt{Last}_a)_{a\in\Sigma}\rangle$ is decidable.

Let's stress that this undecidability property is specific to infinite alphabets. Indeed, for finite alphabets, $\mathtt{EqLast}$ is definable with the predicates $\mathtt{Last}_a$'s, $a \in \Sigma$ and the theory of $\leq_{\mathtt{pref}}$, $\mathtt{EqLen}$ and the $\mathtt{Last}_a$'s is decidable, cf. [15, 8]. Also, Rabin's celebrated result insures that, for finite or infinite alphabets, the monadic second order theory of $\leq_{\mathtt{pref}}$ and the $\mathtt{Last}_a$'s is decidable.

The second Theorem is concerned with definability or undefinability properties of certain predicates one from another. Its proof is given in §3 (Theorems 5 and 6) and §4 (Theorem 11).

Theorem 2. *Let Σ be an infinite alphabet.*
1. Concatenation is definable in the structure $\langle \Sigma^*; \leq_{\mathtt{pref}}, \mathtt{EqLen}, \mathtt{EqLast}\rangle$.

2. In $\langle \Sigma^*; \leq_{\mathtt{pref}}, \mathtt{EqLast}, (\mathtt{Last}_a)_{a\in\Sigma}\rangle$, neither concatenation nor $\mathtt{EqLen}$ are definable.

3. $\langle \Sigma^*; \leq_{\mathtt{pref}}, \mathtt{EqLen}, \mathtt{EqLast}\rangle$ is interpretable in $\langle \Sigma^*; \leq_{\mathtt{pref}}, \mathtt{EqLast}\rangle$.

Observe that items 1 and 3 in Theorem 2 do no longer hold for finite alphabets: they would violate the undecidability of the theory of the free monoid with concatenation (Quine, 1946 [14]) and that of the monadic second order theory of the free monoid with the successor functions $u \mapsto ua$

and EqLen (Vidal-Naquet, 1972 [19]). As for item 2, non-definability of concatenation also holds for any non-empty finite alphabet and non-definability of EqLen holds whenever Σ has at least two letters.

2 The Σ_2 theory of $\leq_{\mathtt{pref}}, \mathtt{Pred}, \mathtt{EqLast}$ is undecidable

2.1 Coding Post Correspondence Problem with EqLast

Recall the Post Correspondence Problem (PCP) for words in $\{a, b\}^*$:

- An instance $\pi = \{(u_1, v_1), \ldots, (u_k, v_k)\}$ of PCP is a finite subset of pairs of words in $\{a, b\}^*$.
- A non-trivial solution of the PCP for π is a non-empty sequence $(i_1, \ldots, i_r) \in \{1, \ldots, k\}^*$ such that $u_{i_1} \ldots u_{i_r} = v_{i_1} \ldots v_{i_r}$.

As is well known, PCP is undecidable, even if we restrict it to π's for which $k \leq 7$ (Matiyasevich & Senizergues, [9, 10] 1996). It also remains undecidable if we restrict it to families $\pi = \{(u_1, v_1), \ldots, (u_k, v_k)\}$ such that all u_i's and v_i's are non-empty.

To prove the stated undecidability result, we code the Post Correspondence Problem with $\mathtt{EqLast}$ and $\leq_{\mathtt{pref}}$.

Theorem 3. *Let Σ be an infinite alphabet and let $a, b \in \Sigma$. To any instance $\pi = \{(u_1, v_1), \ldots, (u_k, v_k)\}$ of the PCP for pairs of non-empty words in $\{a, b\}^*$, one can recursively associate a closed $\exists\forall\forall$ formula F_π of the language*

$$\mathcal{L} \ = \ \{\leq_{\mathtt{pref}},\ \mathtt{Pred}\ ,\ \varepsilon\ ,\ \mathtt{Last}_a\ ,\ \mathtt{Last}_b\ ,\ \mathtt{EqLast}\}$$

such that π has a non-trivial solution if and only if F_π is true in Σ^*.

Proof. We start with an informal introduction of the encoding of a solution of an instance of PCP. Suppose we are given the instance $(u_1, v_1) = (a, aba)$ and $(u_2, v_2) = (baaab, a)$ which has the solution $w = a\ baaab\ a = aba\ a\ aba$. Consider the coarsest factorization which refines both fatorizations $w = a\ ba\ a\ ab\ a$. A possible tentative consists of introducing an s-th marker γ_s and an s-th marker δ_s, taken in $\Sigma \setminus \{a, b\}$, at the beginning of the s-th occurrence of a factor u_{i_s} and v_{i_s} respectively, together with final markers:

$$\gamma_1\delta_1 a\gamma_2 ba\delta_2 a\delta_3 ab\gamma_3 a\gamma_4\delta_4$$

Since Σ is infinite we can suppose all γ's and all δ's to be distinct. Also, we can express the fact that the γ's and δ's are distinct, but there is no way we can relate γ_i with δ_i for $i = 1, \dots 4$. This is achieved by substituting $\gamma_i a$ for each occurrence of γ_i and $\gamma_i b$ for each occurrence of δ_i, leading to the following encoding:

$$(\gamma_1 a)(\gamma_1 b)a(\gamma_2 a)ba(\gamma_2 b)a(\gamma_3 b)ab(\gamma_3 a)a(\gamma_4 a)(\gamma_4 b)$$

Formally, consider a non-trivial solution $S = u_{i_1} \dots u_{i_r} = v_{i_1} \dots v_{i_r}$ of an instance of the PCP. Let $w_1 \dots w_m$ be the coarsest factorization of S refining the previous two factorizations. For each w_i there are three possible cases:

- either it is a prefix of an occurrence u_{i_s} solely
- or it is a prefix an occurrence of v_{i_ℓ} solely
- or it is a prefix of an occurrence u_{i_s} and of an occurrence of v_{i_ℓ}.

Then set $S' = z_1 w_1 \dots z_m w_m \gamma_{r+1} a \gamma_{r+1} b$ where for $1 \leq i \leq m$ we have $z_i = \gamma_{i_s} a$ in the first case and $z_i = \gamma_{i_\ell} b$ in the second case and $z_i = \gamma_{i_s} a \gamma_{i_\ell} b$ in the last case.

It then suffices to observe that the family of such words S' can be expressed in the theory. Indeed, S' encodes a solution of the PCP for π if and only if it satisfies the following conditions where $\Gamma = \Sigma \setminus \{a, b\}$:

i. Start and end. $i1$. $|S'| \geq 9$
$i2$. There exists $\gamma \in \Gamma$ such that $\gamma a \gamma b$ is a prefix of S'.
$i3$. There exists $\gamma' \in \Gamma$ such that $\gamma' a \gamma' b$ is a suffix of S'.

ii. Markers. Every occurrence in S' of a letter in Γ is immediately followed by an occurrence of a or b.

iii. Markers are distinct. No factor in Γa or Γb occurs twice in S'.

iv. Inductive step of a backward decomposition of S'.
If $\gamma \in \Gamma$ and $x\gamma a$ and $y\gamma b$ are both prefixes of S' then either $x = \varepsilon$ and $y = \gamma a$ or there exists $(u_i, v_i) \in \pi$ and u'_i, v'_i and $\gamma' \in \Gamma$ such that
• $\gamma' a u'_i$ is a suffix of x and $\gamma' b v'_i$ is a suffix of y
• $u'_i \in (\Gamma b \cup \{\varepsilon\}) u_{i,1} \Gamma b u_{i,2} \dots \Gamma b u_{i,m_i}$ where $u_i = u_{i,1} u_{i,2} \dots u_{i,m_i}$ and the $u_{i,j}$'s are $\neq \varepsilon$,
• $v'_i \in v_{i,1} \Gamma a v_{i,2} \dots \Gamma a v_{i,n_i} (\Gamma a \cup \{\varepsilon\})$ where $v_i = v_{i,1} v_{i,2} \dots v_{i,n_i}$ and the $v_{i,j}$'s are $\neq \varepsilon$.

Observe that $m_i \leq |u_i|$ and $n_i \leq |v_i|$ since the $u_{i,j}$'s and $v_{i,j}$'s are $\neq \varepsilon$. Thus, the lengths of u'_i, v'_i are bounded.

We now express conditions *i-iv* in the language $\mathcal{L}$. For $s \geq 2$, let $\texttt{Pred}^{(s)}$

denote the s-th iterate of $\mathtt{Pred}$. First, observe that the predicates $\mathtt{Last}_\Gamma(x)$ (which means $x \in \Sigma^*\Gamma$), $|x| \geq s$ and $|x| = s$ are expressible in $\mathcal{L}$ by the following quantifier-free formulas:

$$\begin{array}{rcl} \mathtt{Last}_\Gamma(x) & \equiv & x \neq \varepsilon \wedge \neg\mathtt{Last}_a(x) \wedge \neg\mathtt{Last}_b(x) \\ |x| \geq s & \equiv & \mathtt{Pred}^{(s-1)}(S') \neq \varepsilon \\ |x| = s & \equiv & \mathtt{Pred}^{(s-1)}(S') \neq \varepsilon \wedge \mathtt{Pred}^{(k)}(S') = \varepsilon \end{array}$$

Condition (i1). See above.

Condition (i2-i3). The predicate $\exists\gamma \in \Gamma\ z \in \Sigma^*\gamma a\gamma b$ is expressible by the formula $\lambda(z)$ such that

$$\lambda(z) \equiv \left\{ \begin{array}{l} |z| \geq 4 \wedge \mathtt{EqLast}(\mathtt{Pred}^{(1)}(z), \mathtt{Pred}^{(3)}(z)) \\ \wedge \mathtt{Last}_b(z) \wedge \mathtt{Last}_a(\mathtt{Pred}^{(2)}(z)) \wedge \mathtt{Last}_\Gamma(\mathtt{Pred}^{(1)}(z)) \end{array} \right.$$

Now, condition (i3) is the quantifier-free formula $\lambda(S')$ and condition (i2) is the formula $\forall x\ ((x \leq_{\mathtt{pref}} S' \wedge |x| = 4)\ \Rightarrow\ \lambda(x))$.

Condition (ii). Consider the formula

$$\forall x \leq_{\mathtt{pref}} S'\ [\mathtt{Last}_\Gamma(\mathtt{Pred}^{(1)}(x)) \Rightarrow (\mathtt{Last}_a(x) \vee \mathtt{Last}_b(x))]$$

Condition (iii). It can be expressed by the $\forall^2$ formula

$$\forall x\, \forall y\, [(\left\{ \begin{array}{l} (x <_{\mathtt{pref}} y \leq_{\mathtt{pref}} S' \wedge \mathtt{EqLast}(x,y) \\ \wedge\ (\mathtt{Last}_a(x) \vee \mathtt{Last}_b(x)) \\ \wedge\ \mathtt{Last}_\Gamma(\mathtt{Pred}(x)) \end{array} \right.) \Rightarrow\ \neg\mathtt{EqLast}(\mathtt{Pred}(x), \mathtt{Pred}(y))]$$

Condition (iv). Let $\omega_0, \ldots, \omega_q, \theta_0, \ldots, \theta_p \in \{a,b\}^*$. The predicates

- $\exists\gamma \in \Gamma\ (x \in \Sigma^*\gamma a\ \wedge\ y \in \Sigma^*\gamma b)$
- $\exists\gamma_1, \ldots, \gamma_p, \gamma'_1, \ldots, \gamma'_q, \gamma \in \Gamma$
  $(x \in \Sigma^*\theta_0\gamma_1 b\theta_1 \ldots \gamma_p b\theta_p\gamma a\ \wedge\ y \in \Sigma^*\omega_0\gamma'_1 a\omega_1 \ldots \gamma'_q a\omega_q\gamma b)$

are expressible in $\mathcal{L}$ by quantifier-free formulas $\mu(x,y)$ and $\nu^{\omega_0,\ldots,\omega_q}_{\theta_0,\ldots,\theta_p}(x,y)$ similar to the above λ.
Condition (iv) can be expressed by the $\forall^2$ formula

$$\begin{array}{l} \forall x\ \forall y\ [(x \leq_{\mathtt{pref}} S' \wedge y \leq_{\mathtt{pref}} S' \wedge \mu(x,y)\ \Rightarrow \\ \qquad (|x| = 2 \wedge |y| = 4 \wedge \lambda(y)) \\ \qquad \vee\ \bigvee_{(u_i,v_i)\in\pi} \bigvee_{(\omega_0,\ldots,\omega_q)\in D(u_i)} \bigvee_{(\theta_0,\ldots,\theta_p)\in D(v_i)} \nu^{\omega_0,\ldots,\omega_q}_{\theta_0,\ldots,\theta_p}(x,y) \end{array}$$

where $D(u)$ is the finite family of tuples $(\alpha_0, \ldots, \alpha_q)$ of non-empty words such that $\alpha_0 \ldots \alpha_q = u$.

Finally, the wanted formula F_π is $\exists S'\ \Phi(S')$ where Φ is the conjunction of all formulas associated to conditions (i) to (iv) above. Since the universal quantifiers of the diverse conjuncts can be factorized, we see that F_π is of the form $\exists\forall^2$. □

2.2 Proof of point 1 of Theorem 1

Let $F_\pi = \exists S'\ \forall x\ \forall y\ \Psi(S', x, y)$ be as in Theorem 3. To eliminate the constants a, b from Ψ, we use condition $i3$ above which insures that b is the last letter of S' and a is the last one of $\texttt{Pred}^{(2)}(S')$ and we express the fact that $a \neq b$ as a property of S'.
Let $G(S', x, y)$ be obtained from $\Psi(S', x, y)$ by replacing every occurrence of the atomic formula $\texttt{Last}_b(z)$ (where z is x or y) by the formula $\texttt{EqLast}(z, S')$ and every occurrence of the atomic formula $\texttt{Last}_a(z)$ by $\texttt{EqLast}(z, \texttt{Pred}^{(2)}(S'))$.
Clearly, F_π is equivalent to the following $\exists\forall\forall$ formula which uses only $\leq_{\texttt{pref}}$, Pred and EqLast together with the sole constant ε :

$$\exists S'\ \forall x\ \forall y\ (\neg\texttt{EqLast}(S', \texttt{Pred}^{(2)}(S'))\ \wedge\ G(S', x, y))$$

3 Definability of concatenation

In this section we show that concatenation of words may be defined with the three predicates $\leq_{\texttt{pref}}$, EqLen, EqLast but that it is independent of the two predicates $\leq_{\texttt{pref}}$ and EqLast.

3.1 Defining concatenation with $\leq_{\texttt{pref}}$, EqLen, EqLast

We first show that the concatenation of two words whose letters are pairwise different can be defined via the two predicates $\leq_{\texttt{pref}}$ and EqLast. This will allow us to show the definability of the concatenation of any two arbitrary words using the three predicates of our structure.

A word is *injective* if it does not contain two occurrences of the same letter. Observe that, in the context of infinite alphabets, there are arbitrarily large injective words.

Proposition 4. *Let Σ be an infinite alphabet. The relation*

$$R = \{(u, v, w) \mid uv \text{ is injective and } w = uv\}$$

is definable by a $\forall\forall$ formula $\rho(u, v, w)$ with $\leq_{\texttt{pref}}$, Pred, ε and EqLast.

Proof. Observe that $(u, v, w) \in R$ if and only if

i. u, v and w are injective words and u, v have no letter in common,

ii. *Trivial case.* If v is empty then $w = u$.

iii. *Initial step of the backward analysis.* If v is not empty then v, w have the same last letter.

iv. *Inductive step of the backward analysis.* If x and y are prefixes of v and w and $|x| \geq 2$ and x, y have the same last letter then $|y| \geq 2$ and $\texttt{Pred}(x), \texttt{Pred}(y)$ also have the same last letter.

v. *Final step of the backward analysis.* If x and y are prefixes of v and w and $|x| = 1$ and x, y have the same last letter then $\texttt{Pred}(y) = u$.

In fact, $(u, v, w) \in R$ clearly implies all these conditions.
Conversely, suppose these conditions hold. Condition ii solves the case where v is empty. So suppose v is non-empty. Conditions iii, iv insure that v is a suffix of w and conditions v shows that the associated prefix of w is u. Thus, $(u, v, w) \in R$.
Finally, it is easy to see that each one of conditions i-v is expressible in the language $(\leq_{\texttt{pref}}, \texttt{Pred}, \varepsilon, \texttt{EqLast})$ with two universal quantifications. □

Theorem 5. *Let Σ be an infinite alphabet. Concatenation is definable in the structure $\langle \Sigma^*; \leq_{\texttt{pref}}, \texttt{EqLen}, \texttt{EqLast} \rangle$.*
Using the extra function Pred *and the constant ε (which are definable with $\leq_{\texttt{pref}}$), concatenation can be defined by a $\exists^3\forall^4$ formula.*

Proof. Observe that $w = uv$ if and only if $v = \varepsilon$ and $w = u$ or $v \neq \varepsilon$ and there exist ξ, η, ζ such that

i. ξ, η are injective words which have no letter in common.

ii. $|\xi| = |u|$, $|\eta| = |v|$, $|\zeta| = |w|$ and $\xi\eta = \zeta$.

iii. u is a prefix of w.

iv. If η', ζ', v', w' are non-empty prefixes of η, ζ, v, w and $|\eta'| = |v'|$ and $|\zeta'| = |w'|$ and η', ζ' have the same last letter then v', w' also have the same last letter.

Using Proposition 4, this is clearly expressible with $\leq_{\texttt{pref}}$, Pred, ε, EqLen and EqLast by a $\exists^3\forall^4$ formula (the existential quantifications being over ξ, η, ζ and the universal ones over η', ζ', v', w'). □

3.2 Non-definability of concatenation with $\leq_{\texttt{pref}}$ and EqLast

Let's denote by $\cdot$ the concatenation operation.
The definition of concatenation obtained in the previous paragraph uses both predicates EqLen and EqLast. Of course, one cannot remove EqLast since the theory of $\leq_{\texttt{pref}}$, EqLen, $(\texttt{Last}_a)_{a\in\Sigma}$ is decidable, cf. [15, 8].
In order to show that we cannot remove EqLen, we use the following simple property: if EqLen or $\cdot$ were definable in Σ^* from $\leq_{\texttt{pref}}$ and the $\texttt{Last}_a$'s, the same would be true in any elementary extension of this structure, in particular in any ultrapower. So, to prove the negative result we are looking for, we construct an ultrapower of the structure

$$\langle \Sigma^*; \leq_{\texttt{pref}}, \texttt{EqLen}, \texttt{EqLast}, \cdot, (\texttt{Last}_a)_{a\in\Sigma}\rangle$$

for which there is a bijection of the domain which does preserve the interpretations of $\leq_{\texttt{pref}}$ and the $\texttt{Last}_a$'s but does not preserve those of $\cdot$ nor EqLen. Let's recall that an ultrafilter $\mathcal{U}$ on $\mathbb{N}$ is a family of subsets of $\mathbb{N}$ closed by intersection and superset and such that, for all $X \subseteq \mathbb{N}$, either X or its complement $\mathbb{N} \setminus X$ is in $\mathcal{U}$. The ultrapower $\mathcal{A}_\mathcal{U}$ of a structure $\mathcal{A}$ is obtained as follows:
- its domain is the set of equivalence classes of the equivalence $\sim_\mathcal{U}$ on $\mathcal{A}^\mathbb{N}$ such that $f \sim_\mathcal{U} g \Leftrightarrow \{i \in \mathbb{N} \mid f(i) = g(i)\} \in \mathcal{U}$,
- the interpretation in $\mathcal{A}_\mathcal{U}$ of function and relation symbols are obtained as follows. First, lift the interpretations in $\mathcal{A}$ to $\mathcal{A}^\mathbb{N}$ in the obvious componentwise way. Second, quotient by $\sim_\mathcal{U}$.
Łoś theorem insures that, for any formula $F(x_1, \ldots, x_n)$, and any $f_1, \ldots, f_n \in \mathcal{A}^\mathbb{N}$, letting $[f_i]$ be the $\sim_\mathcal{U}$ class of f_i, we have

$$\mathcal{A}_\mathcal{U} \models F([f_1], \ldots, [f_n]) \;\Leftrightarrow\; \{i \in \mathbb{N} \mid \mathcal{A} \models F(f_1(i), \ldots, f_n(i))\} \in \mathcal{U}$$

The elementary embedding of $\mathcal{A}$ into $\mathcal{A}_\mathcal{U}$ maps $a \in \mathcal{A}$ to the class of the constant function $\mathbb{N} \to \mathcal{A}$ with value a. For more details on the method of ultrapowers, we refer to standard textbooks, e.g. [4] or [13].

Theorem 6. *Let Σ be a finite or infinite alphabet with at least two letters. Neither* EqLen *nor concatenation is definable in the structure*

$$\mathcal{W} = \langle \Sigma^*; \leq_{\texttt{pref}}, \texttt{EqLast}, (\texttt{Last}_a)_{a\in\Sigma}\rangle$$

Proof. Let $\mathcal{A}$ be the structure $\mathcal{W}$ enriched with the $\cdot$ operation and the EqLen predicate. Consider a non-principal ultrafilter $\mathcal{U}$ on $\mathbb{N}$ and let

$$\mathcal{A}_\mathcal{U} = \langle (\Sigma^*)^\mathbb{N}/\sim_\mathcal{U} \;;\; = \;,\; \leq^\mathcal{U}_{\texttt{pref}} \;,\; \texttt{EqLen}^\mathcal{U} \;,\; \texttt{EqLast}^\mathcal{U} \;,\; (\texttt{Last}^\mathcal{U}_a)_{a\in\Sigma} \;,\; \cdot^\mathcal{U}\rangle$$

be the $\mathcal{U}$-ultrapower of $\mathcal{A}$ with domain $(\Sigma^*)^{\mathbb{N}}/\mathcal{U}$ where $\sim_{\mathcal{U}}$ is the equivalence on $(\Sigma^*)^{\mathbb{N}}$ such that, for $f, g : \mathbb{N} \to \Sigma^*$,

$$f \sim_{\mathcal{U}} g \;\Leftrightarrow\; \{i \mid f(i) = g(i)\} \in \mathcal{U}$$

We denote by $[f]$ the $\sim_{\mathcal{U}}$ equivalence class of $f : \mathbb{N} \to \Sigma^*$. Let $\iota : \mathcal{A} \to \mathcal{A}_{\mathcal{U}}$ be the natural embedding such that $\iota(u)$ is the $\sim_{\mathcal{U}}$ class of the constant function with value u. Since ι is an elementary embedding, in order to prove the theorem, it suffices to construct a bijection of $(\Sigma^*)^{\mathbb{N}}/\mathcal{U}$ into itself which preserves $\leq_{\mathtt{pref}}^{\mathcal{U}}$, the $\mathtt{Last}_a^{\mathcal{U}}$'s and $\mathtt{EqLast}^{\mathcal{U}}$ but does not preserve $\cdot^{\mathcal{U}}$ nor $\mathtt{EqLen}^{\mathcal{U}}$.

For $u \in \Sigma^*$ and $c \in \Sigma$, let's denote by $c^{-1}u$ the word v such that $u = cv$ if u starts with c, else $v = u$. We fix some $c \in \Sigma$ and consider the family $\mathcal{F}$ of $f : \mathbb{N} \to \Sigma^*$ such that $[f]$ admits all $\iota(c^k)$ as prefixes:

$$f \in \mathcal{F} \;\Leftrightarrow\; \forall k \in \mathbb{N}\;\; \iota(c^k) \leq_{\mathtt{pref}}^{\mathcal{U}} [f] \;\Leftrightarrow\; \forall k \in \mathbb{N}\;\; \{n \mid c^k \leq_{\mathtt{pref}} f(n)\} \in \mathcal{U}$$

We define $\Phi : (\Sigma^*)^{\mathbb{N}} \to (\Sigma^*)^{\mathbb{N}}$ as follows: for $f \in (\Sigma^*)^{\mathbb{N}}$,

$$\Phi(f)(i) = \begin{cases} c^{-1}f(i) & \text{if } f \in \mathcal{F} \\ f(i) & \text{if } f \notin \mathcal{F} \end{cases}$$

Intuitively, when there is a non-standard initial block of letters c in $[f]_{\mathcal{U}}$, we remove the first letter of this block, else we do nothing. To prove that Φ induces a bijection on $\mathcal{A}$ which respects $=$, $\leq_{\mathtt{pref}}$, $\mathtt{EqLast}$ and the $\mathtt{Last}_a$'s, we consider $f, g : \mathbb{N} \to \Sigma^*$ and prove that

i. $[f] = [g]$ if and only if $[\Phi(f)] = [\Phi(g)]$,

ii. $[f] \leq_{\mathtt{pref}}^{\mathcal{U}} [g]$ if and only if $[\Phi(f)] \leq_{\mathtt{pref}}^{\mathcal{U}} [\Phi(g)]$,

iii. $\mathtt{EqLast}^{\mathcal{U}}([f], [g])$ if and only if $\mathtt{EqLast}^{\mathcal{U}}([\Phi(f)], [\Phi(g)])$,

iv. $\mathtt{Last}_a^{\mathcal{U}}([f])$ if and only if $\mathtt{Last}_a^{\mathcal{U}}([\Phi(f)])$.

We argue by cases.
Case $f \notin \mathcal{F}$ and $g \notin \mathcal{F}$. Trivial since then $\Phi(f) = f$ and $\Phi(g) = g$.

Case $f \in \mathcal{F}$ and $g \notin \mathcal{F}$. Since $g \notin \mathcal{F}$, there exists k such that $\{n \mid c^k \not\leq_{\mathtt{pref}} g(n)\} \in \mathcal{U}$. Now $\{n \mid c^k \leq_{\mathtt{pref}} f(n)\}$ and $\{n \mid c^{k+1} \leq_{\mathtt{pref}} f(n)\}$ are both in $\mathcal{U}$. Thus, $\{n \mid c^k \leq_{\mathtt{pref}} c^{-1}f(n)\} = \{n \mid c^k \leq_{\mathtt{pref}} \Phi(f)(n)\} \in \mathcal{U}$. In particular, $\{n \mid f(n) \not\leq_{\mathtt{pref}} g(n)\}$ and $\{n \mid \Phi(f)(n) \not\leq_{\mathtt{pref}} g(n)\}$ are both in $\mathcal{U}$. This proves $[f] \not\leq_{\mathtt{pref}}^{\mathcal{U}} [g]$ and $[\Phi(f)] \not\leq_{\mathtt{pref}}^{\mathcal{U}} [\Phi(g)]$. Hence also $[f] \neq [g]$

and $[\Phi(f)] \neq [\Phi(g)]$. Thus, i and ii hold.
Since $\{n \mid |f(n)| \geq 2\} \in \mathcal{U}$ we see that $\{n \mid \texttt{EqLast}(f(n), \Phi(f)(n))\} \in \mathcal{U}$. Thus, $\texttt{EqLast}^{\mathcal{U}}([f], [\Phi(f)])$. Now, EqLast is transitive, hence so is $\texttt{EqLast}^{\mathcal{U}}$. Since $\Phi(g) = g$ we see that

$$\texttt{EqLast}^{\mathcal{U}}([f], [g]) \;\Leftrightarrow\; \texttt{EqLast}^{\mathcal{U}}([\Phi(f)], [g]) \;\Leftrightarrow\; \texttt{EqLast}^{\mathcal{U}}([\Phi(f)], [\Phi(g)])$$

which gives iii.
Since $\mathcal{A} \models \texttt{EqLast}(x, y) \Rightarrow (\texttt{Last}_a(x) \Leftrightarrow \texttt{Last}_a(y))$ we have

$$\mathcal{A}_{\mathcal{U}} \models \texttt{EqLast}_{\mathcal{U}}(x, y) \Rightarrow (\texttt{Last}_a^{\mathcal{U}}(x) \Leftrightarrow \texttt{Last}_a^{\mathcal{U}}(y))$$

which proves assertion iv.

Case $f \notin \mathcal{F}$ and $g \in \mathcal{F}$. By symmetry, equivalences i, iii and iv are similar to the previous case. Let's prove ii. As before, there exists k such that the three sets

$$\{n \mid c^k \not\leq_{\texttt{pref}} f(n)\}\ ,\ \{n \mid c^k \leq_{\texttt{pref}} g(n)\}\ ,\ \{n \mid c^k \leq_{\texttt{pref}} \Phi(g)(n)\}$$

are in $\mathcal{U}$. Let X be their intersection (which is in $\mathcal{U}$). Then,

$$n \in X \;\Rightarrow\; (f(n) \leq_{\texttt{pref}} c^{k-1} \;\Leftrightarrow\; f(n) \leq_{\texttt{pref}} \Phi(g)(n) \;\Leftrightarrow\; f(n) \leq_{\texttt{pref}} g(n))$$

Since $\Phi(f) = f$, this proves that $\{n \mid f(n) \leq_{\texttt{pref}} g(n)\} \in \mathcal{U}$ if and only if $\{n \mid \Phi(f)(n) \leq_{\texttt{pref}} \Phi(g)(n)\} \in \mathcal{U}$, whence ii.

Case $f \in \mathcal{F}$ and $g \in \mathcal{F}$. The set $X = \{n \mid cc \leq_{\texttt{pref}} f(n) \wedge cc \leq_{\texttt{pref}} g(n)\}$ is in $\mathcal{U}$. For $n \in X$, we have $c\Phi(f)(n) = f(n)$ and $c\Phi(g)(n) = g(n)$. From this, we easily deduce conditions i to iv.

We now show that Φ does not preserve EqLen. For instance, let $f(n) = c^{n+1}$ and $g(n) = b^{n+1}$ where b is some letter different from c. Then, for all n, we have $|f(n)| = |g(n)|$ and $|\Phi(f)(n)| = |\Phi(g)(n)| - 1$, so that $\texttt{EqLen}^{\mathcal{U}}([f], [g])$ but $\neg\texttt{EqLen}^{\mathcal{U}}([\Phi(f)], [\Phi(g)])$.

Finally, observe that Φ does not preserve concatenation. For instance, let $f(n) = b$, $g(n) = c^{n+1}$ and $h(n) = bc^{n+1}$ where b is some letter different from c. Then, for all n, we have $h(n) = f(n)g(n)$ and $\Phi(f)(n)\Phi(g)(n) = bc^n \neq bc^{n+1} = \Phi(h)(n)$. Thus, we have $[f] \cdot_{\mathcal{U}} [g] = [h]$ whereas $[\Phi(f)] \cdot_{\mathcal{U}} [\Phi(g)] \neq [\Phi(h)]$. □

4 Interpretability of EqLen with $\leq_{\texttt{pref}}$ and EqLast

The aim of this section is to prove a property which is weaker than that of definability. It relies on the notion of interpretability which we now recall, see, e.g., [16, Paragraph 4.7.].

A structure $\mathcal{A} = \langle A; (S_j)_{j\in J}\rangle$ is *interpretable* in a structure $\mathcal{B}$ if one can define in $\mathcal{B}$ a subset $D \subseteq B$, an equivalence relation $\sim$ on D and relations (or functions) σ_j's which are compatible with $\sim$, in such a way that the quotient of $\langle D; (\sigma_j)_{j\in J}\rangle$ by $\sim$ be isomorphic to $\mathcal{A}$. Though concatenation is not definable with $\leq_{\texttt{pref}}$ and EqLast, it is nevertheless interpretable. A result which is interesting by itself and the proof of which illustrates the definability power of EqLast with $\leq_{\texttt{pref}}$. In Theorem 5 we proved the definability of the concatenation by the predicates EqLen, $\leq_{\texttt{pref}}$ and EqLast. Here we show that EqLen is interpretable with the remaining two predicates $\leq_{\texttt{pref}}$ and EqLast.

First, we have to develop some coding tricks in the vein of that used in the proof of Theorem 1. The idea is the following. If the letters of a word were indexed, then testing that two words are of equal length would reduce to testing whether or not the last values of the indices are equal. This is impossible since there is no total ordering defined in the infinite alphabet, but we can use a weaker property: if we use pairwise different letters as indices, then two words are of the same length if and only if there exists a one-to-one mapping between the two sets of indices. We are thus led to insert an arbitrary letter before each letter of a given word which can be interpreted as indexing that letter, provided all these extra letters are different from one another and from the letters of the word.

4.1 Injective words and the join operator

First, let's introduce some convenient tools.

Definition 7. *1. If $j \in \{1, 2\}$, we denote by $\pi_j(x)$ the map $\Sigma^* \to \Sigma^*$ such that $\pi_j(\varepsilon) = \varepsilon$ (recall that ε denotes the empty word) and, for $x \in \Sigma^*$ and $a \in \Sigma$,*

$$\pi_j(xa) = \begin{cases} \pi_j(x) & \text{if } |xa| \not\equiv j \mod 2 \\ \pi_j(x)a & \text{if } |xa| \equiv j \mod 2 \end{cases}$$

I.e., $\pi_j(x)$ is obtained by keeping only one letter out of two in x, starting with the j-th one. For instance, $\pi_1(abcdefg) = aceg$, $\pi_2(abcdefg) = bdf$.

2. The $\bowtie$ operation on words is a restricted form of the traditional shuffle

and is defined as follows:

$$\rho_1 \dots \rho_p \bowtie \sigma_1 \dots \sigma_q \quad = \quad \rho_1\sigma_1 \dots \rho_s\sigma_s \text{ where } s = \min(p, q)$$

So that $\pi_1(\rho \bowtie \sigma)$ (resp. $\pi_2(\rho \bowtie \sigma)$) is the prefix of ρ (resp. of σ) with length $\min(|\rho|, |\sigma|)$.

3. We let $D = \{\xi \bowtie u \mid \xi$ is injective and no letter of ξ occurs in $u\}$.

Proposition 8. *Let Σ be an infinite alphabet. The following relations are all definable with $\leq_{\texttt{pref}}$ and* `EqLast`*:*

$$x \in D \quad , \quad x \in D \ \wedge \ \xi = \pi_1(x)$$

Proof. First, observe that $x \in D \ \wedge \ \xi = \pi_1(x)$ if and only if
- either both ξ and x are the empty word
- or the following conditions are satisfied

i. ξ is injective and $\xi \neq \varepsilon$ and $|x| \geq 2$,

ii. ξ and $\texttt{Pred}(x)$ have the same last letter,

iii. The last letter of x does not occur in ξ,

iv. If $\xi' \leq_{\texttt{pref}} \xi$ and $x' \leq_{\texttt{pref}} x$ holds and if ξ' and x' have the same last letter then

(a) $|\xi'| = 1$ if and only if $|x'| = 1$

(b) $|\xi'| \geq 2$ if and only if $|x'| \geq 3$

(c) if $|\xi'| \geq 2$ then $\texttt{Pred}(\xi')$ and $\texttt{Pred}^{(2)}(x')$ have the same last letter,

(d) The last letter of $\texttt{Pred}(x')$ does not occur in ξ.

All these conditions are expressible with $\leq_{\texttt{pref}}$ and `EqLast`.

Finally, $x \in D$ if and only if $\exists\xi \ (x \in D \wedge \xi = \pi_1(x))$. □

4.2 Expressing `EqLen` on injective words

Proposition 9. *Let Σ be an infinite alphabet. The predicate*

$$\xi, \eta \text{ are injective words and } |\xi| = |\eta|$$

is definable with $\leq_{\texttt{pref}}$ and `EqLast`*.*

Proof. First, we consider the case where ξ and η have no letter in common, a condition which is expressible with $\leq_{\texttt{pref}}$ and $\texttt{EqLast}$.
In that case, the word $\xi \bowtie \eta$ is also injective and equality $|\xi| = |\eta|$ holds if and only if either both ξ and η are the empty word or there exists an injective x (which is to be $\xi \bowtie \eta$) such that

i. $|x| \geq 2$ and $\texttt{EqLast}(\eta, x)$ and $\texttt{EqLast}(\xi, \texttt{Pred}(x))$,

ii. If $\xi' \leq_{\texttt{pref}} \xi$ and $\eta' \leq_{\texttt{pref}} \eta$ and $x' \leq_{\texttt{pref}} x$ and $|x'| \geq 2$ and $\texttt{EqLast}(\eta', x')$ and $\texttt{EqLast}(\xi', \texttt{Pred}(x'))$ then

   (a) either $|\xi'| = |\eta'| = 1$ and $|x'| = 2$

   (b) or $|\xi'|, |\eta'| \geq 2$ and $|x'| \geq 4$ and $\texttt{EqLast}(\texttt{Pred}(\eta'), \texttt{Pred}^{(2)}(x'))$ and $\texttt{EqLast}(\texttt{Pred}(\xi'), \texttt{Pred}^{(3)}(x'))$.

These conditions are clearly expressible by a formula $\psi(\xi, \eta)$ using $\leq_{\texttt{pref}}$ and $\texttt{EqLast}$.

Let $\omega(x, y)$ be the formula $\forall x' \leq_{\texttt{pref}} x \ \forall y' \leq_{\texttt{pref}} y \ \neg\texttt{EqLast}(x', y')$ which expresses that x, y have disjoint alphabet. In case ξ and η have letters in common, use an auxiliary injective word θ having no letter in common with ξ and η and observe that,

$$|\xi| = |\eta| \quad \Leftrightarrow \quad \exists\theta \ (\psi(\xi, \theta) \ \wedge \psi(\eta, \theta) \ \wedge \omega(\xi, \theta) \ \wedge \omega(\eta, \theta))$$

□

4.3 The basic equivalence on D

Proposition 8 insures that one can get ξ from $\xi \bowtie u$ when $|u| = |\xi|$ using only $\leq_{\texttt{pref}}$ and $\texttt{EqLast}$. However, it is not possible to get u from $\xi \bowtie u$. The following proposition tells the best we can do.

Proposition 10. *Let Σ be an infinite alphabet. The following equivalence relation is definable with $\leq_{\texttt{pref}}$ and $\texttt{EqLast}$:*

$$x \sim y \ \Leftrightarrow \ x, y \in D \ \wedge \ \pi_2(x) = \pi_2(y)$$

Proof. Observe that $x \sim y$ if and only if there exist words ξ, η such that

i. $x \in D \wedge \pi_1(x) = \xi$ and $y \in D \wedge \pi_1(y) = \eta$

ii. ξ, η are injective and $|\xi| = |\eta|$

iii. $\texttt{EqLast}(x, y)$

iv. If $\xi' \leq_{\mathtt{pref}} \xi$, $\eta' \leq_{\mathtt{pref}} \eta$, $x' \leq_{\mathtt{pref}} x$, $y' \leq_{\mathtt{pref}} y$ and $|x'|, |y'| \geq 2$ and $|\xi'| = |\eta'| \geq 1$ and $\mathtt{EqLast}(\xi', x')$ and $\mathtt{EqLast}(\eta', y')$ then

(a) either $|\xi'| = |\eta'| = 1$ and $|x'| = |y'| = 2$

(b) or $|\xi'| = |\eta'| \geq 2$ and $|x'| = |y'| \geq 4$ and $\mathtt{EqLast}(\mathtt{Pred}(\xi'), \mathtt{Pred}^{(2)}(x'))$ and $\mathtt{EqLast}(\mathtt{Pred}(\eta'), \mathtt{Pred}^{(2)}(y'))$.

Finally, we use Proposition 8 and 9 to express the above conditions. □

4.4 The interpretation theorem

Theorem 11. *Let Σ be an infinite alphabet and let D and $\sim$ be as in Definition 7 and Proposition 10. The inverse images in D under π_2 of relations $\leq_{\mathtt{pref}}$, $\mathtt{EqLen}$, $\mathtt{EqLast}$ (resp. $\mathtt{Last}_a$ where $a \in \Sigma$), namely relations*

$$\begin{aligned} \mathtt{PREF} &= \{(x,y) \in D \times D \mid \pi_2(x) \leq_{\mathtt{pref}} \pi_2(y)\} \\ \mathtt{EQL} &= \{(x,y) \in D \times D \mid |\pi_2(x)| = |\pi_2(y)|\} \\ \mathtt{EQLA} &= \{(x,y) \in D \times D \mid \mathtt{EqLast}(\pi_2(x), \pi_2(y))\} \\ \textit{(resp. } \mathtt{LAST}_a &= \{x \in D \mid \mathtt{Last}_a(\pi_2(x))\}\textit{)} \end{aligned}$$

are all definable with $\leq_{\mathtt{pref}}$ and $\mathtt{EqLast}$ (resp. and $\mathtt{Last}_a$).

In particular, the structure

$$\langle \Sigma^*; \leq_{\mathtt{pref}}, \mathtt{EqLen}, \mathtt{EqLast}, (\mathtt{Last}_a)_{a \in \Sigma} \rangle$$

is isomorphic to the quotient under $\sim$ of the structure

$$\langle D; \mathtt{PREF}, \mathtt{EQL}, \mathtt{EQLA}, (\mathtt{LAST}_a)_{a \in \Sigma} \rangle$$

hence is interpretable in

$$\langle \Sigma^*; \leq_{\mathtt{pref}}, \mathtt{EqLast}, (\mathtt{Last}_a)_{a \in \Sigma} \rangle$$

Proof. The definition of the equivalence $\sim$ insures that the relations $\mathtt{PREF}$, $\mathtt{EQL}$, $\mathtt{EQLA}$ and $\mathtt{LAST}_a$ are compatible with $\sim$.
Let $x, y \in D$ and let $\xi = \pi_1(x)$ and $\eta = \pi_1(y)$. Then ξ, η are injective and $x = \xi \bowtie u$ and $y = \eta \bowtie v$ and ξ, u (resp. η, v) have no letter in common. We prove that

$$\begin{aligned} \mathtt{EqLast}(\pi_2(x), \pi_2(y)) &\Leftrightarrow \mathtt{EqLast}(x,y) \\ \mathtt{Last}_a(\pi_2(x)) &\Leftrightarrow \mathtt{Last}_a(x) \\ \mathtt{EqLen}(\pi_2(x), \pi_2(y)) &\Leftrightarrow \exists \xi\ \exists \eta\ (\pi_1(x) = \xi \wedge \pi_1(y) = \eta \wedge |\xi| = |\eta|) \\ \pi_2(x) \leq_{\mathtt{pref}} \pi_2(y) &\Leftrightarrow \exists \widetilde{x}\ \exists \widetilde{y}\ (\widetilde{x} \sim x \wedge \widetilde{y} \sim y \wedge \widetilde{x} \leq_{\mathtt{pref}} \widetilde{y}) \end{aligned}$$

The assertions about $\mathtt{EqLast}$ and $\mathtt{Last}_a$ are obvious since the last letter of u is that of $\xi \bowtie u$. That about $\mathtt{EqLen}$ is easy since $|\xi| = |u|$ and $|\eta| = |v|$. Let's prove the assertion about $\leq_{\mathtt{pref}}$.
$\Leftarrow$. From $\widetilde{x} \leq_{\mathtt{pref}} \widetilde{y}$ we get $\pi_2(\widetilde{x}) \leq_{\mathtt{pref}} \pi_2(\widetilde{x})$. Since $\widetilde{x} \sim x$ and $\widetilde{y} \sim y$, we have $\pi_2(\widetilde{x}) = \pi_2(x)$ and $\pi_2(\widetilde{y}) = \pi_2(y)$. Thus, $\pi_2(x) \leq_{\mathtt{pref}} \pi_2(y)$.
$\Rightarrow$. Assertion $\pi_2(x) \leq_{\mathtt{pref}} \pi_2(y)$ means $u \leq_{\mathtt{pref}} v$. Let θ be any injective word with length equal to $\max(|\xi|, |\eta|)$. It suffices to set $\widetilde{x} = \theta \bowtie u$ and $\widetilde{y} = \theta \bowtie v$. □

Using Theorems 5 and 11, we get the following corollary.

Corollary 12. *Let Σ be an infinite alphabet.*
1. One can interpret the structure $\langle \Sigma^*; =, \cdot \rangle$ in $\langle \Sigma^*; \leq_{\mathtt{pref}}, \mathtt{EqLast} \rangle$.

2. Let A be a finite subalphabet of Σ. One can interpret the structure $\langle A^*; =, \cdot \rangle$ in $\langle \Sigma^*; \leq_{\mathtt{pref}}, \mathtt{EqLast}, (\mathtt{Last}_a)_{a \in A} \rangle$.

5 Decidability of the Σ_1 theory of $\langle \Sigma^*; \leq_{\mathtt{pref}}, \mathtt{EqLast}, (\mathtt{EqLen}_k)_{k \in \mathbb{Z}}, (\mathtt{Last}_a)_{a \in \Sigma} \rangle$

The purpose of this paragraph is to show that the existential fragment of our logic is decidable. This is achieved via the characterization of the predicates in terms of synchronous multi-tape finite automata on infinite alphabets as defined in our paper [5].

Let Σ_0 be some finite subalphabet of Σ. Recall that the ability of a Σ_0-synchronous n-tape finite automata $\mathcal{A}$ on an infinite alphabet Σ is as follows:

- $\mathcal{A}$ distinguishes the sole letters of Σ_0. All the letters in $\Sigma \setminus \Sigma_0$ are treated by $\mathcal{A}$ in the same way except that it can detect whether the scanned letters on two of the n tapes are distinct or not.
- Thus, the k-th transition on input $(u_1, \ldots, u_n)$ depends on the current state of $\mathcal{A}$ and on the truth of the statements

$$u_i[k] = a \quad , \quad u_i[k] = u_j[k] \qquad \text{where } i, j = 1, \ldots, n \text{ and } a \in \Sigma_0$$

  where $u[k]$ is the k-th letter of u in case $|u| \geq k$ and a special marker not belonging to the alphabet Σ otherwise.

The following result characterizes the relations recognized by such automata in terms of logic definability, [5].

Theorem 13. *Let Σ be an infinite alphabet. A relation $R \subseteq (\Sigma^*)^n$ is recognized by some Σ_0-synchronous n-tape finite automaton if and only if it is definable in the structure $\langle \Sigma^*; \leq_{\mathtt{pref}}, \mathtt{EqLen}, \mathtt{EqLenEqLast}, (\mathtt{Last}_a)_{a\in\Sigma_0}\rangle$.*

In order to strengthen our decidability result for existential formulas, we enrich the language as much as possible. A convenient tool is the following straightforward application of Theorem 13.

Proposition 14. *Let Σ be an infinite alphabet. Let $\mathtt{Syn}_\Sigma$ be the family of synchronous relations, i.e. of Σ_0-synchronous relations for some finite subalphabet Σ_0 of Σ. Let $\mathtt{SynFun}_\Sigma$ be the family of synchronous functions, i.e. functions $(\Sigma^*)^n \to \Sigma^*$ with graphs in $\mathtt{Syn}_\Sigma$.*
The family $\mathtt{Syn}_\Sigma$ is closed under under Boolean operations, projections and cylindrifications and hence under substitutions of arguments by synchronous functions. The family $\mathtt{SynFun}_\Sigma$ is closed under composition.

We can now state and prove our decidability result.

Theorem 15. *Let Σ be an infinite alphabet. The existential theory of the structure*

$$\langle \Sigma^*; (f)_{f\in\mathtt{SynFun}_\Sigma}, (R)_{R\in\mathtt{Syn}_\Sigma}, \mathtt{EqLast}\rangle$$

is decidable.

In particular, since the function $\mathtt{Pred}$ and all relations $\leq_{\mathtt{pref}}$, $\mathtt{EqLen}_k$ and $\mathtt{Last}_a$ are synchronous, this decidability result applies to the existential theory of Σ^* with this function and these relations.

Proof. As usual, it suffices to decide the truth of formulas of the form $\exists x_1 \ldots \exists x_n\ (\varphi_1 \wedge \ldots \wedge \varphi_p)$ where the φ_i's are atomic formulas or negations of atomic formulas. Proposition 14 allows us to regroup all literals associated to relations in $\mathtt{Syn}_\Sigma$. Thus, we are reduced to decide formulas

$$(*)\ \exists x_1 \ldots \exists x_n\ (R(x_1, \ldots, x_n) \wedge \bigwedge_{(i,j)\in B} \mathtt{EqLast}(x_i, x_j) \wedge \bigwedge_{(i,j)\in C} \neg\mathtt{EqLast}(x_i, x_j))$$

with $B, C \subseteq \{1, \ldots, n\} \times \{1, \ldots, n\}$ and $R \in \mathtt{Syn}_\Sigma^{\Sigma_0}$ for some finite Σ_0.

Claim. *Let $\Gamma = \Sigma_0 \cup \Delta_0 \cup \ldots \cup \Delta_n$ where the Δ_i's are pairwise disjoint subalphabets of $\Sigma \setminus \Sigma_0$, each containing n letters.*
Formula $(*)$ is equivalent to that obtained by restricting $x_1, \ldots, x_n$ to Γ^*.

Proof of Claim. The $\Leftarrow$ direction is trivial. Let's prove the $\Rightarrow$ direction. Suppose $(x_1, \ldots x_n)$ is a solution of $(*)$. Consider the tuple $(y_1, \ldots y_n)$ obtained as follows, where $i, j = 1, \ldots, n$ and $k \in \mathbb{N}$,

- $|y_i| = |x_i|$
- $x_i[k] \in \Sigma_0 \;\Rightarrow\; y_i[k] = x_i[k]$,
- $x_i[k] = x_j[k] \;\Leftrightarrow\; y_i[k] = y_j[k]$
- If $k \neq |x_1|, \ldots, |x_n|$ then the letters $y_1[k], \ldots, y_n[k]$ are in $\Sigma_0 \cup \Delta_0$,
- if $k = |x_i|$ and i is minimal with this property, then the letters $y_1[k], \ldots, y_n[k]$ are in $\Sigma_0 \cup \Delta_i$.

Let now E be an equivalence on $\Delta_1 \cup \ldots \cup \Delta_n$ to be defined below such that $(a, b) \in E$ and $a \in \Delta_i$ and $b \in \Delta_j$ imply $i \neq j$. Let $(z_1, \ldots z_n)$ be obtained from $(y_1, \ldots y_n)$ by identifying pairs of letters in E.
Let $\mathcal{A}$ be a Σ_0-synchronous n-tape finite automaton recognizing R. Since $\mathcal{A}$ only distinguishes letters in Σ_0 and equalities/ disequalities of letters at the same position in the different components, the tuples $(x_1, \ldots, x_n)$, $(y_1, \ldots y_n)$ and $(z_1, \ldots, z_n)$ are simultaneously in or outside R. Let E correspond exactly to the equalities holding between the last letters of $x_1, \ldots, x_n$. Then $(z_1, \ldots, z_n)$ satisfies exactly the same `EqLast` relations than $(x_1, \ldots, x_n)$ does. In particular $(z_1, \ldots, z_n)$ is a solution of $(*)$ which proves the Claim.

To conclude the proof of the theorem, recall that, for a finite alphabet Γ, `EqLast` is recognizable by a Γ-synchronous automaton. Thus, the above Claim insures that $(*)$ reduces to the emptiness problem for such automata on alphabet Γ, which is known to be decidable. □

References

[1] M. Benedikt, L. Libkin, T. Schwentick, and L. Ségoufin. Definable relations and first-order query languages over strings. *Journal of the Association of Computing Machinery*, 50:694–751, 2003.

[2] M. Bojanczyk, A. Muscholl, T. Schwentick, L. Segoufin and C. David. Two-variable logic on words with data. *In Proceedings of LICS'06*, 7–16, 2006.

[3] J.R. Büchi. On a decision method in restricted second order arithmetics. In *Proceedings Int. Cong. in Logic, Methodology and Philosophy of Sc.*, pages 1–11. North-Holland, 1960.

[4] C.C. Chang and J. Keisler. *Model theory.* North-Holland, 1973.

[5] C. Choffrut and S. Grigorieff. Finite n-tape automata on infinite alphabets: extending a Theorem of Eilenberg & al. 2006. Submitted.

[6] S. Demri, R. Lazić and A. Sangnier. Model checking freeze LTL over one-counter automata. *In Proceedings of FoSSaCS'08*, to appear in LNCS, 2008.

[7] V.G. Durnev. Undecidability of the positive $\forall\exists^3$ theory of a free semigroup. *Sibirskii Matematicheskii Zhurnal*, 36(5):1067–1080, Sept. 1995. English translation in *Sibirian Mathematical Journal*, 36(5):917–929, Sept. 1995.

[8] S. Eilenberg, C.C. Elgot, and J.C. Shepherdson. Sets recognized by n-tape automata. *Journal of Algebra*, 3:447–464, 1969.

[9] Y. Matiyasevich and G. Senizergues. Decision problems for semi-thue systems with few rules. In *Proceedings 11th LICS 1996*, pages 523–531. IEEE Computer Soc. Press, 1996.

[10] Y. Matiyasevich and G. Senizergues. Decision problems for semi-thue systems with few rules. *Theoretical Computer Science*, 330(1):145–169, 2005.

[11] F. Neven, T. Schwentick, and V. Vianu. Towards regular languages over infinite alphabets. In *Proceedings 26th MFCS 2001*, LNCS 2136, 560–572 (Extended abstract of [12]).

[12] F. Neven, T. Schwentick, and V. Vianu. Finite state machines for strings over infinite alphabets. In *ACM Transactions on Computational Logic*, 15(3):403–435, 2004.

[13] B. Poizat. *A course in Model theory: an introduction to modern mathematical logic*. Universitext. Springer, 2000.

[14] W. Quine. Concatenation as a basis of arithmetics. *Journal Symbolic Logic*, 11:105–114, 1946.

[15] J.C. Shepherdson. Unpublished, cited in [17].

[16] J.R. Shoenfield. *Mathematical Logic*. Addison-Wesley, 1967.

[17] J.W. Thatcher. Decision problems for multiple successor arithmetics. *Journal Symbolic Logic*, 31:182–190, 1966.

[18] Y.M. Vazhenin & B.V. Rozenblat. Decidability of the positive theory of a free countably generated semigroup. *Mat. Sbornik*, 116:120–127, 1981. English translation in *Math. USSR Sbornik*, 44:109–116, 1983.

[19] G. Vidal-Naquet. Quelques applications des automates d'arbres infinis. In *Automata, Languages and Programming* (Proceedings ICALP 1972), M. Nivat ed., North-Holland 1972, 115–122.