The Quantifier Alternation Hierarchy of Synchronous Relations

Diego Figueira

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR5800, F-33400 Talence, France
CNRS, ReLaX, UMI2000, Siruseri, India
diego.figueira@labri.fr

Varun Ramanathan

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR5800, F-33400 Talence, France
CNRS, ReLaX, UMI2000, Siruseri, India
varun.ramanathan@labri.fr

Pascal Weil

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR5800, F-33400 Talence, France
CNRS, ReLaX, UMI2000, Siruseri, India
pascal.weil@labri.fr

Abstract

The class of synchronous relations, also known as automatic or regular, is one of the most studied subclasses of rational relations. It enjoys many desirable closure properties and is known to be logically characterized: the synchronous relations are exactly those that are defined by a first-order formula on the structure of all finite words, with the prefix, equal-length and last-letter predicates. Here, we study the quantifier alternation hierarchy of this logic. We show that it collapses at level Σ_3 and that all levels below admit decidable characterizations. Our results reveal the connections between this hierarchy and the well-known hierarchy of first-order defined languages of finite words.

2012 ACM Subject Classification Theory of computation → Logic; Theory of computation → Transducers

Keywords and phrases synchronous relations, automatic relations, first-order logic, characterization, quantifier alternation

Digital Object Identifier 10.4230/LIPIcs.MFCS.2019.29

Funding Work supported by ANR DeLTA, grant ANR-16-CE40-0007.

1 Introduction

We study classes of relations on finite words, within the class of rational relations. Synchronous relations [8] – also studied as *regular relations* [3] and *automatic relations* [5] – form a subclass of rational relations which is well-behaved from many standpoints. Contrary to rational relations, they enjoy crucial effective properties such as closure under intersection and complement. As a consequence, most paradigmatic problems are decidable for synchronous relations, in the same way as they are for regular languages. Further, they admit clean characterizations both in terms of automata and logic, providing yet more evidence of the connections between logic, formal languages and automata. Due to this good behavior, this class finds various applications in verification [6, 1], automatic structures [5], the theory of transducers and database theory [2].

Synchronous relations contain natural relations such as equality, prefix, or equal-length. In fact, any letter-to-letter transduction, alphabetic morphism or length-preserving rational relation lies within synchronous relations [4].

Synchronous relations are those that are accepted by multi-tape finite automata. A k-tape automaton over an alphabet $\mathbb{A}$ can be naturally seen as an NFA over the alphabet of

k-tuples $\hat{\mathbb{A}} = (\mathbb{A} \cup \{\perp\})^k$ that reads k input words $w_1, \dots, w_k \in \mathbb{A}^*$ simultaneously, from left to right, the i-th transition reading the tuple from $\hat{\mathbb{A}}$ composed of the i-th letters of each word w_j (or $\perp$ if $i > |w_j|$). Synchronous relations can also be described as finite unions of the componentwise concatenation of a length-preserving rational relation with a recognizable relation – two other well-studied classes of relations [4].

On the other hand, relations can be defined by logical formulæ interpreted on words in $\mathbb{A}^*$: a formula φ with free variables $z_1, \dots, z_k$ defines the k-ary relation of all tuples $(w_1, \dots, w_k)$ such that φ holds with the interpretation $z_i \mapsto w_i$ $(1 \leq i \leq k)$. Eilenberg, Elgot and Shepherdson [9] showed that a relation is synchronous if, and only if, it can be defined in this way by a first-order formula using the prefix, equal-length and last letter predicates.

This characterization opens the possibility of exploring classes of synchronous relations specified by fragments of first-order logic. In the present work, we study the quantifier alternation hierarchy in this logic, that is, the classes of relations defined by formulæ with a bounded number of alternations of existential and universal quantifier blocks. This is a natural way of providing small, well-behaved classes (closed under boolean combinations) of synchronous relations. We show that the hierarchy collapses at level Σ_3 and we give clean combinatorial characterizations for its different layers, namely Σ_1, its boolean closure $\mathcal{B}\Sigma_1$, Σ_2 and $\mathcal{B}\Sigma_2$. These characterizations reveal strong links with the classical Σ_1- and Σ_2- fragments of the first-order theory on finite words with the order $<$ relation and letter predicates. Interestingly, the notion of *subwords*, which plays a central role in the characterization of $\Sigma_1[<]$ and $\mathcal{B}\Sigma_1[<]$, must be replaced here by the more subtle notion of *synchronized subwords*.

We also show that these characterizations are decidable: given a synchronous relation, one can decide whether it is defined by a formula in Σ_1 (resp. $\mathcal{B}\Sigma$, Σ_2, $\mathcal{B}\Sigma_2$). Our results provide therefore a complete decision procedure for the alternation hierarchy of synchronous relations.

Section 2 introduces technical preliminaries. Our main results are all stated in Section 3, and their proofs are given in the ensuing sections: Section 4 for the collapse of the hierarchy, Section 5 for what concerns the Σ_1- and $\mathcal{B}\Sigma_1$-fragments and Section 6 for the Σ_2- and $\mathcal{B}\Sigma_2$-fragments.

2 Preliminaries

For any set A and $\bar{a} \in A^k$, we denote by $\bar{a}(i)$ its i-th component, an element of A. If $w \in A^*$ is a word, we denote by $|w|$ its length and, for any $1 \leq i \leq j \leq |w|$, by $w[i]$ the letter of w in i-th position, and by $w[i..j]$ the factor $w[i] \cdots w[j]$ of w between positions i and j. To simplify notation, we let $w[i..j] = \varepsilon$ (the empty word) whenever $1 \leq i \leq j \leq |w|$ does not hold. If u, v are words, we let $u \sqcap v$ be the longest common prefix of u and v.

We will consider relations of a fixed arity $k \geq 2$, over a fixed alphabet $\mathbb{A}$ with at least two letters. Let $\perp$ be a symbol not in $\mathbb{A}$, and let $\mathbb{A}_\perp = \mathbb{A} \cup \{\perp\}$. We will often work with the alphabet $\mathbb{A}_\perp^k$, the direct product of k copies of $\mathbb{A}_\perp$.

Synchronous relations

Given $w_1, \dots, w_k \in \mathbb{A}^*$, we define the *synchronized word* $\bar{w}$ of the tuple $(w_1, \dots, w_k)$, written $\bar{w} = w_1 \otimes \cdots \otimes w_k$, to be the word in $(\mathbb{A}_\perp^k)^*$ such that:

- $|\bar{w}| = \max(|w_1|, \dots, |w_k|)$; and
- for every $i \in \{1, \dots, |\bar{w}|\}$ and $j \in \{1, \dots, k\}$, we have $\bar{w}[i](j) = w_j[i]$ if $i \leq |w_j|$, and $\bar{w}[i](j) = \perp$ otherwise.

For example, $abba \otimes c \otimes de = (a,c,d)(b,\bot,e)(b,\bot,\bot)(a,\bot,\bot)$. We let SW_k be the set of all k-synchronized words, that is, $SW_k = \{w_1 \otimes \cdots \otimes w_k : w_1,\ldots,w_k \in \mathbb{A}^*\}$. For $S = \{s_1,\ldots,s_n\} \subseteq \{1,\ldots,k\}$ such that $s_1 < \cdots < s_n$, we define the *projection* $\pi_S \colon SW_k \to (\mathbb{A}^n_\bot)^*$ as $\pi_S(w_1 \otimes \cdots \otimes w_k) = w_{s_1} \otimes \cdots \otimes w_{s_n}$. In the case of a singleton $S = \{i\}$, note that $\pi_S : SW_k \to \mathbb{A}^*$, and we simply write π_i. If $R \subseteq (\mathbb{A}^*)^k$ is a k-ary relation, the *synchronized language of* R, denoted by L_R, is the language $\{w_1 \otimes \cdots \otimes w_k \colon (w_1,\ldots,w_k) \in R\} \subseteq (\mathbb{A}^k_\bot)^*$. The relation R is said to be *synchronous* if L_R is regular. The set of synchronous relations, of arbitrary arity, is denoted by **Sync**.

MSO over finite words

In the classical setting introduced by Büchi (see [21]), languages over an alphabet $\mathbb{A}$ are described by formulæ interpreted over the set of positions of a finite word, using the binary word ordering predicate $<$ and the unary letter predicates a ($a \in \mathbb{A}$) – where $a(i)$ holds if the word carries letter a in position i. Büchi's Theorem [7, 21] states that a language is regular if and only if it is definable by a closed monadic second order formula in this logic, written $\mathbf{MSO}[<, \{a\}_{a\in\mathbb{A}}]$, or $\mathbf{MSO}[<]$ if $\mathbb{A}$ is understood. If φ is a closed formula in $\mathbf{MSO}[<]$, we let $|\varphi|$ be the language in $\mathbb{A}^*$ it defines, and if $\mathcal{F}$ is a set of formulæ, $|\mathcal{F}|$ denotes the class $\{|\varphi| : \varphi \in \mathcal{F}\}$.

First-order formulæ in Büchi's logic define a strict subclass of regular languages, that of *star-free* languages (see [19, 12, 21, 22]). The quantifier alternation hierarchy within $\mathbf{FO}[<]$ forms a strict infinite hierarchy, and it has been the object of intense study (see [17, 18] for an overview). In the sequel, we will only use results regarding the $\mathbf{\Sigma}_1$, $\mathcal{B}\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$ and $\mathcal{B}\mathbf{\Sigma}_2$ fragments of $\mathbf{FO}[<]$, possibly enriched with the constant predicate **max**, which stands for the last position in a word (see below in this section and Section 6). Recall that $\mathcal{B}$ designates the boolean closure; that $\mathbf{\Sigma}_1$ is the set of existential formulæ, of the form $\exists z_1 \cdots \exists z_n\ \varphi$ with φ quantifier-free; and that $\mathbf{\Sigma}_2$ consists in the formulæ of the form $\exists z_1 \cdots \exists z_n\ \varphi$ with φ in $\mathcal{B}\mathbf{\Sigma}_1$. The $\mathbf{\Pi}_i$ fragment consists of the negations of the formulæ in $\mathbf{\Sigma}_i$ (e.g., $\mathbf{\Pi}_1$ are all formulæ of the form $\forall z_1 \cdots \forall z_n\ \varphi$ with φ quantifier-free). We will sometimes write $\mathbf{FO}[<](\mathbb{A}^*)$ (or fragments thereof) when we want to make explicit the alphabet $\mathbb{A}$ we work with.

FO over the structure of all finite words

We now turn to the signature introduced by Eilenberg *et al.* [9] to discuss synchronous relations over $\mathbb{A}^*$, namely $\sigma = [\preceq, \mathsf{eq}, (\ell_a)_{a\in\mathbb{A}}]$. These predicates are interpreted as follows:

- $(w_1,w_2) \models x \preceq y$ if and only if w_1 is a *prefix* of w_2;
- $(w_1,w_2) \models \mathsf{eq}(x,y)$ if and only if w_1 and w_2 have *equal length*;
- $w \models \ell_a(x)$ if and only if the *last letter* of w is a.

Every formula φ with free variables $z_1,\ldots,z_k$ defines a k-ary relation written $\|\varphi\|$, namely:

$$\|\varphi\| = \{(w_1,\ldots,w_k) \in (\mathbb{A}^*)^k : (w_1,\ldots,w_k) \models \varphi\}.$$

Let $\mathbf{FO}[\sigma]$ denote the set of first order formulæ with signature σ, and for any $\mathcal{F} \subseteq \mathbf{FO}[\sigma]$ let $\|\mathcal{F}\|$ denote the set of relations definable by formulæ in $\mathcal{F}$. For convenience, we write $x \prec y$ for $(x \preceq y) \wedge \neg(y \preceq x)$, and L_φ for $L_{\|\varphi\|}$. For example, $\varphi(x_1,x_2,x_3) = (x_3 \preceq x_1) \wedge (x_3 \preceq x_2) \wedge \forall z\ (x_3 \prec z \to \neg(z \preceq x_1 \wedge z \preceq x_2))$ defines the set $\|\varphi\|$ of all triples (w_1,w_2,w_3) such that $w_3 = w_1 \sqcap w_2$.

Types and type sequences

For a letter $\bar{a} = (a_1, \dots, a_k) \in \mathbb{A}_\perp^k$, the *type of* $\bar{a}$ is the subset of $\{1, \dots, k\}^2$ $\mathsf{type}(\bar{a}) = \{(i,j) : a_i = a_j \neq \perp\}$. The *type* of a synchronized word $\bar{w} = \bar{a}_1 \cdots \bar{a}_n$ is given by $\mathsf{type}(\bar{w}) = \bigcap_{1 \le i \le n} \mathsf{type}(\bar{a}_i)$. For example, $\mathsf{type}((a, \perp, a, b)) = \{(1,3), (3,1), (1,1), (3,3), (4,4)\}$ and $\mathsf{type}((a, \perp, a, b)(\perp, \perp, b, b)) = \{(3,3), (4,4)\}$.

In particular, if $\bar{w} \in SW_k$, the successive values $T_1 \supsetneq T_2 \cdots \supsetneq T_n$ taken by the types of the prefixes of $\bar{w}$ form the *type sequence* of $\bar{w}$, written $\mathsf{type\text{-}seq}(\bar{w})$. In such a sequence, we say that T_i is an *end type* if either $i = n$, or $(j,j) \in T_i \setminus T_{i+1}$ for some $j \le k$ – that is, if $\bar{w} = w_1 \otimes \cdots \otimes w_k$, T_i is an end type in $\mathsf{type\text{-}seq}(\bar{w})$ if the length of the longest prefix of $\bar{w}$ of type T_i is equal to $|w_j|$ for some j. If T is a type, we let $\mathbb{A}_T$ be the set of T-*compatible letters*, $\mathbb{A}_T = \{\bar{a} \in \mathbb{A}_\perp^k : T \subseteq \mathsf{type}(\bar{a}) \subseteq T^\star\}$, where $T^\star = \{(i,j) : (i,i), (j,j) \in T\}$; and let $\mathbb{A}_{-,T} = \{\bar{a} \in \mathbb{A}_\perp^k : T = \mathsf{type}(\bar{a})\}$. If T' is a type such that $T \subsetneq T'$, we also let $\mathbb{A}_{T',T} = \{\bar{a} \in \mathbb{A}_\perp^k : T = T' \cap \mathsf{type}(\bar{a})\}$. Hence, if $\bar{w}\bar{a} \in SW_k$ and $\bar{w}$ has type T (resp. T'), then $\mathsf{type}(\bar{w}\bar{a}) = T$ if and only if $\bar{a} \in \mathbb{A}_T$ (resp. $T \subsetneq T'$ and $\bar{a} \in \mathbb{A}_{T',T}$).

It follows that, if $\bar{T} = (T_1, \dots, T_n)$ is a type sequence and $K(\bar{T})$ is the set of synchronized words $\bar{w}$ such that $\mathsf{type\text{-}seq}(\bar{w}) = \bar{T}$, then

$$K(\bar{T}) = \mathbb{A}_{-,T_1} \mathbb{A}_{T_1}^* \mathbb{A}_{T_1,T_2} \mathbb{A}_{T_2}^* \cdots \mathbb{A}_{T_{n-1},T_n} \mathbb{A}_{T_n}^*. \tag{1}$$

Note that this product of languages is *deterministic*, that is, given $\bar{w}$, we can determine $\mathsf{type\text{-}seq}(\bar{w})$ and its unique factorization in the product (1) by reading $\bar{w}$ from left to right: the first letter determines T_1, the next factor is the longest written in $\mathbb{A}_{T_1}$, the first letter not in $\mathbb{A}_{T_1}$ (together with T_1) determines T_2, etc.

If $\bar{w} = \bar{w}_1 \cdots \bar{w}_n$ is this factorization, with $\bar{w}_i \in \mathbb{A}_{T_{i-1},T_i} \mathbb{A}_{T_i}^*$ for each i ($\mathbb{A}_{T_-,T_i} \mathbb{A}_{T_i}^*$ for $i = 1$), we say that $\bar{w}_i$ is the *i-th type factor* of $\bar{w}$, written $\mathsf{type\text{-}factor}_i(\bar{w})$.

Synchronized subwords

We denote by $\sqsubseteq$ the (*scattered*) *subword relation* on $\mathbb{A}^*$ (sometimes called *subsequence*): if $u, v \in \mathbb{A}^*$, we have $u \sqsubseteq v$ if there exists a strictly increasing function $p\colon \{1, \dots, |u|\} \to \{1, \dots, |v|\}$, called the *witness function*, such that, for all $i \in \{1, \dots, |u|\}$, $u[i] = v[p(i)]$.

Given $\bar{w} = w_1 \otimes \cdots \otimes w_k$ and $\bar{w}' = w'_1 \otimes \cdots \otimes w'_k$, we say that $\bar{w}$ is a *synchronized subword* of $\bar{w}'$, denoted by $\bar{w} \sqsubseteq_{\mathsf{s}} \bar{w}'$ if and only if $\bar{w} \sqsubseteq \bar{w}'$, with a witness function p which is

- *type preserving*: $\mathsf{type}(\bar{w}[1..i]) = \mathsf{type}(\bar{w}'[1..p(i)])$ for all $1 \le i \le |\bar{w}|$; and
- *end preserving*: $p(|w_j|) = |w'_j|$ for all $j \in \{1, \dots, k\}$.

▶ **Lemma 1.** *For $\bar{u}, \bar{u}' \in SW_k$ with type sequences $\bar{T}$ and $\bar{T}'$, we have $\bar{u} \sqsubseteq_{\mathsf{s}} \bar{u}'$ if and only if $\bar{T}$ is a subsequence of $\bar{T}'$ with a witness function $t\colon \{1, \dots, |\bar{T}|\} \to \{1, \dots, |\bar{T}'|\}$ such that, for every i, the i-th type factor of $\bar{u}$ is a subword of the $t(i)$-th type factor of $\bar{u}'$, and they further have the same last letter if $\bar{T}_i$ is an end type of $\bar{T}$.*

Proof. Suppose first that $\bar{u} \sqsubseteq_{\mathsf{s}} \bar{u}'$ and let $p\colon \{1, \dots, |\bar{u}|\} \to \{1, \dots, |\bar{u}'|\}$ be a witness function. Let $\bar{T}$ and $\bar{T}'$ be the type sequences of $\bar{u}$ and $\bar{u}'$ and let $\bar{u} = \bar{u}_1 \cdots \bar{u}_n$, $\bar{u}' = \bar{u}'_1 \cdots \bar{u}'_m$ be the type factorizations of $\bar{u}$ and $\bar{u}'$. For each $1 \le i \le n$, if $\bar{v}_i = \bar{u}_1 \cdots \bar{u}_i$, then $T_i = \mathsf{type}(\bar{v}_i) = \mathsf{type}(\bar{u}'[1..p(|\bar{v}_i|)])$, so $\bar{T}$ is a subsequence of $\bar{T}'$. Let $t(i)$ be such that $T'_{t(i)} = T_i$. Since p is type-preserving, the factor $\bar{u}_i$ is a subword of $\bar{u}'_{t(i)}$. Both have the same last letter if T_i is an end type for $\bar{u}$, since p is end-preserving.

Conversely, suppose that $T_i = T'_{t(i)}$ and that $\bar{u}_i \sqsubseteq \bar{u}'_{t(i)}$, with witness function p_i (with domain $\{1, \dots, |\bar{u}_i|\}$). Let p be the function on $\{1, \dots, |\bar{u}|\}$ obtained by "concatenating" the p_i: $p(|\bar{u}_1 \cdots \bar{u}_{i-1}| + h) = |\bar{u}'_1 \cdots \bar{u}'_{t(i)-1}| + p_i(h)$. It is directly verified that p witnesses $\bar{u} \sqsubseteq_{\mathsf{s}} \bar{u}'$. ◀

Given a quasi-order $\preceq$ over a domain X, the $\preceq$-*upward closure* of an element $x \in X$ is the set $\uparrow_{\preceq} x = \{x' \in X : x \preceq x'\}$. If $S \subseteq X$, we also let $\uparrow_{\preceq} S = \bigcup_{x \in S} \uparrow_{\preceq} x$. Finally, S is $\preceq$-*upward closed* if $S = \uparrow_{\preceq} S$. Henceforward, we write $\uparrow w$ and $\uparrow S$ as short for $\uparrow_{\sqsubseteq} w$ and $\uparrow_{\sqsubseteq} S$; and we write $\uparrow_{\mathsf{s}} \bar{w}$ and $\uparrow_{\mathsf{s}} S$ as short for $\uparrow_{\sqsubseteq_{\mathsf{s}}} w$ and $\uparrow_{\sqsubseteq_{\mathsf{s}}} S$.

A *well-quasi-order* (*wqo*) is a quasi-order $(X, \preceq)$ such that for every infinite sequence $(x_i)_{i \in \mathbb{N}}$ of elements of X, there exist $i < j$ such that $x_i \preceq x_j$. A crucial observation is that, if $\preceq$ is a wqo, then any set has a finite number of $\preceq$-minimal elements.

It is a classical result (Higman's lemma [10], see also [11, chap. 6]) that the subword order on $\mathbb{A}^*$ is a wqo. Unsurprisingly, the same holds for the synchronized subword order.

▶ **Proposition 2.** *For every k, $(SW_k, \sqsubseteq_{\mathsf{s}})$ is a well-quasi-order.*

Proof. If $(\bar{w}_n)_n$ is an infinite sequence of elements of SW_k, we can extract an infinite subsequence of elements with the same type sequence $\bar{T} = (T_1, \ldots, T_n)$ (since there are only finitely many type sequences). Similarly, we can further extract an infinite subsequence where, for each end type T_i, all the i-th type factors end with the same letter.

On this subsequence, $\sqsubseteq_{\mathsf{s}}$ coincides with the intersection of the subword order applied to each of the n type factors. The result follows since the subword order is a wqo and wqo's are closed under intersection. ◀

Bounded subword and synchronized subword classes

It is well-known [21] that $\|\mathbf{\Sigma}_1[<]\|$ is the set of languages of the form $\uparrow\mathcal{S}$, where $\mathcal{S}$ is a set of words (which can be assumed to be finite by the wqo property). Similarly, $\|\mathcal{B}\mathbf{\Sigma}_1[<]\|$ is the set of finite unions of languages of the form $\uparrow\mathcal{S} \setminus \uparrow\mathcal{S}'$, where $\mathcal{S}, \mathcal{S}'$ are finite sets of words. For $h \in \mathbb{N}$, let $\sim_h$ be the equivalence relation on $\mathbb{A}^*$ defined by $w_1 \sim_h w_2$ if and only if w_1 and w_2 have the same subwords of length at most h. Then we also know [13, 21] that $\|\mathcal{B}\mathbf{\Sigma}_1[<]\|$ is the set of finite unions of $\sim_h$-classes, also known as the set of *piecewise testable* languages.

We introduce analogous definitions for synchronized subwords. If $h \in \mathbb{N}$, we let $\approx_h$ be the equivalence relation on synchronized words defined by $\bar{w}_1 \approx_h \bar{w}_2$ if $\bar{w}_1$ and $\bar{w}_2$ have the same set of synchronized subwords of length less than or equal to h. We let $\mathcal{V}_h$ be the set of equivalence classes of $\approx_h$ and $\bar{\mathcal{V}}_h$ be its Boolean closure. Finally, we let $\bar{\mathcal{V}} = \bigcup_{h \in \mathbb{N}} \bar{\mathcal{V}}_h$.

3 Summary of results

We start with an overview of our main results. Their proofs are discussed in the next sections. Theorem 3 refines the already mentioned 1969 result of Eilenberg *et al.* [9], which states that the relations definable in $\|\mathbf{FO}[\sigma]\|$ are exactly the synchronous relations.

▶ **Theorem 3.** *For any alphabet having at least two letters,*

$$\|\mathbf{\Sigma}_1[\sigma]\| \subsetneq \|\mathcal{B}\mathbf{\Sigma}_1[\sigma]\| \subsetneq \|\mathbf{\Sigma}_2[\sigma]\| \subsetneq \|\mathcal{B}\mathbf{\Sigma}_2[\sigma]\| \subsetneq \|\mathbf{\Sigma}_3[\sigma]\| = \|\mathbf{FO}[\sigma]\| = \mathbf{Sync}.$$

The characterizations of the $\mathbf{\Sigma}_1[\sigma]$- and the $\mathcal{B}\mathbf{\Sigma}_1[\sigma]$-fragments are in terms of synchronized subwords, rather than ordinary subwords as in the case of word languages.

▶ **Theorem 4.** *For any relation R, $R \in \|\mathbf{\Sigma}_1[\sigma]\|$ if and only if $L_R = \uparrow_{\mathsf{s}} L_R$.*

▶ **Theorem 5.** *For any relation R, $R \in \|\mathcal{B}\mathbf{\Sigma}_1[\sigma]\|$ if and only if $L_R \in \bar{\mathcal{V}}$.*

We note the following corollary of Theorem 4, which follows from the wqo property of the synchronized subword order (Proposition 2).

▶ **Corollary 6.** *For every formula $\varphi \in \mathbf{\Sigma}_1[\sigma]$ with k free variables, there exists a finite set $S \subseteq SW_k$ such that $L_\varphi = \uparrow_s S$.*

In contrast, the characterizations of the $\mathbf{\Sigma}_2[\sigma]$- and the $\mathcal{B}\mathbf{\Sigma}_2[\sigma]$-fragments reduce to the corresponding logical fragments for word languages.

▶ **Theorem 7.** *For any relation R, $R \in \|\mathbf{\Sigma}_2[\sigma]\|$ if and only if $L_R \in |\mathbf{\Sigma}_2[<]|$.*

▶ **Corollary 8.** *For any relation R, $R \in \|\mathcal{B}\mathbf{\Sigma}_2[\sigma]\|$ if and only if $L_R \in |\mathcal{B}\mathbf{\Sigma}_2[<]|$.*

These characterizations can then be used to prove the decidability of the membership problems for the different fragments of $\mathbf{FO}[\sigma]$.

▶ **Theorem 9.** *Given a fragment $\mathcal{F} \in \{\mathbf{\Sigma}_1[\sigma], \mathcal{B}\mathbf{\Sigma}_1[\sigma], \mathbf{\Sigma}_2[\sigma], \mathcal{B}\mathbf{\Sigma}_2[\sigma]\}$ and a synchronous relation R (say, an automaton accepting L_R), it is decidable whether $R \in \|\mathcal{F}\|$.*

▶ Remark 10. The decidability of membership in $\|\mathbf{\Sigma}_2[\sigma]\|$ and $\|\mathcal{B}\mathbf{\Sigma}_2[\sigma]\|$ follows directly from the decidability of $|\mathbf{\Sigma}_2[<]|$ [15] and $|\mathcal{B}\mathbf{\Sigma}_2[<]|$ [16], see also [17].

4 Collapse of the alternation hierarchy

Here we prove Theorem 3. The equality $\|\mathbf{FO}[\sigma]\| = \mathbf{Sync}$ was proved in [9]. The collapse at level Σ_3 is established by a folklore argument, in which the runs of a classical (1-tape) automaton are first-order encoded using additional tapes.

Specifically, let R be a k-ary synchronous relation and let $\mathcal{A}$ be a DFA accepting the language L_R, with state set Q. We fix $0, 1$, two distinct letters of $\mathbb{A}$, which will be used to encode the runs of $\mathcal{A}$. If $\bar{w} \in SW_k$ and if $q \in Q$, we let u_q be the word of length $|\bar{w}|$ which carries the letter 1 at each position i such that $\mathcal{A}$ is in state q after reading the i first letters of $\bar{w}$, and carries letter 0 everywhere else.

A $\mathbf{\Sigma}_3[\sigma]$-formula φ with free variables $z_1, \ldots, z_k$ defining R is obtained as follows. A tuple of variables $Y = (y_q)_{q \in Q}$ is quantified existentially, and the rest of the formula, in $\mathcal{B}\mathbf{\Sigma}_2[\sigma]$, verifies that the words u_q $(q \in Q)$ assigned to these variables encode the run of $\mathcal{A}$ as above. More precisely, each of these words must have length $|\bar{w}|$ (verified in $\mathbf{\Pi}_2[\sigma]$) and at every position, exactly one of them carries a 1 (verified in $\mathbf{\Pi}_1[\sigma]$). Moreover, the first letter of each u_q must be 1 exactly when q is the state reached from the initial state when reading the tuple of first letters of the words assigned to $z_1, \ldots, z_k$; the notion of first letter, or rather of length 1 prefix is $\mathbf{\Pi}_2[\sigma]$-definable. Similarly, the last state must be one of the final states of $\mathcal{A}$, which is readily verified using (without any quantifier) the last letter predicates ℓ_a. Finally, the compatibility of the u_q's with the transitions of the run of $\mathcal{A}$ on $\bar{w}$ can be encoded with a $\mathbf{\Pi}_2[\sigma]$-formula. Thus $\mathbf{Sync}$ is contained in $\|\mathbf{\Sigma}_3[\sigma]\|$

We now turn to the proof of the strictness of containments in Theorem 3. We observe that a unary relation on $\mathbb{A}$ is nothing but a language in $\mathbb{A}^*$. In that case, the notion of type is trivial, and $w \sqsubseteq_s w'$ if and only if $w \sqsubseteq w'$ and they have the same last letter. In view of Theorems 4 and 5, it follows that a^* is in $\|\mathcal{B}\mathbf{\Sigma}_1[\sigma]\|$ but not in $\|\mathbf{\Sigma}_1[\sigma]\|$. The other containments are strict because they are in the classical framework. This completes the proof of Theorem 3.

5 $\Sigma_1[\sigma]$ and its boolean closure

We prove Theorems 4 and 5, and we exhibit a decision procedure for membership in $(\mathcal{B})\mathbf{\Sigma}_1[\sigma]$.

5.1 Characterization of $\Sigma_1[\sigma]$

The proof of Theorem 4 is a consequence of the following three properties:

- if φ is a formula of $\mathbf{\Sigma}_1[\sigma]$, then L_φ is $\sqsubseteq_{\mathsf{s}}$-upward closed (Lemma 12 below);
- the relation defined by the $\sqsubseteq_{\mathsf{s}}$-upward closure of any synchronized word is $\mathbf{\Sigma}_1[\sigma]$-definable (Lemma 13 below);
- $\sqsubseteq_{\mathsf{s}}$ is a well-quasi-order on SW_k (Proposition 2 above).

We first show how these three properties imply Theorem 4.

Proof of Theorem 4. If R is $\mathbf{\Sigma}_1[\sigma]$-definable, then $L_R = \uparrow_{\mathsf{s}} L_R$ by Lemma 12. Conversely, suppose that $L_R = \uparrow_{\mathsf{s}} L_R$ and let S be a $\sqsubseteq_{\mathsf{s}}$-minimal subset of L_R, such that $\uparrow_{\mathsf{s}} S = \uparrow_{\mathsf{s}} L_R$. Since $\sqsubseteq_{\mathsf{s}}$ is a wqo by Proposition 2, S is finite, say, $S = \{\bar{w}_1, \ldots, \bar{w}_m\}$. By Lemma 13, for every $1 \le i \le m$, there exists a formula $\varphi_i \in \mathbf{\Sigma}_1[\sigma]$ such that $L_{\varphi_i} = \uparrow_{\mathsf{s}} \bar{w}_i$. Letting $\varphi = \bigvee_{1\le i\le m} \varphi_i$, we see that $L_\varphi = \bigcup_{1\le i\le m} \uparrow_{\mathsf{s}} \bar{w}_i = \uparrow_{\mathsf{s}} S = \uparrow_{\mathsf{s}} L_R = L_R$, and hence $R = \|\varphi\| \in \|\mathbf{\Sigma}_1[\sigma]\|$. ◀

Before we prove Lemma 12, we establish the following technical lemma.

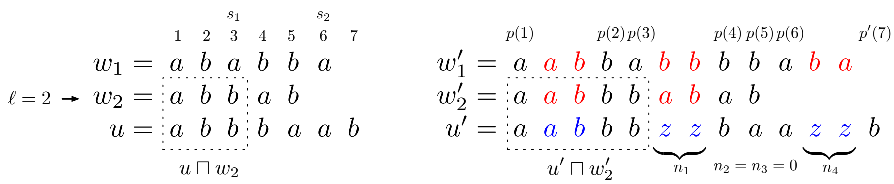
▶ **Lemma 11.** *Let $\bar{w} = w_1 \otimes \cdots \otimes w_k$, and $\bar{w}' = w_1' \otimes \cdots \otimes w_k'$ such that $\bar{w} \sqsubseteq_{\mathsf{s}} \bar{w}'$. For all $u \in \mathbb{A}^*$, there exists $u' \in \mathbb{A}^*$ such that $\bar{w} \otimes u \sqsubseteq_{\mathsf{s}} \bar{w}' \otimes u'$.*

Proof. Let $p\colon \{1, \ldots, |\bar{w}|\} \to \{1, \ldots, |\bar{w}'|\}$ be the witness function for $\bar{w} \sqsubseteq_{\mathsf{s}} \bar{w}'$. Let $s_1 = \max_{i\le k} |u \sqcap w_i|$ and $\ell \le k$ be such that $|u \sqcap w_\ell| = s_1$. Finally, let $s_2 = \min(|\bar{w}|, |u|)$, and $u[s_1{+}1..s_2] = a_1 \cdots a_m$. That is, $u = (u {\sqcap} w_\ell)\ (a_1 \cdots a_m)\ u[s_2{+}1..|u|]$, with the understanding that $u[s_2 + 1..|u|] = \varepsilon$ if $|u| \le |\bar{w}|$.

For $1 \le i \le m$, let $n_i = p(s_1 + i) - p(s_1) - 1$, and $n_{m+1} = |\bar{w}'| - p(s_2)$. Let also z be an arbitrary letter of $\mathbb{A}$. We then define

$$u' = w_\ell'[1..p(s_1)]\ z^{n_1}\, a_1 \cdots z^{n_m}\, a_m\, z^{n_{m+1}}\ u[s_2 + 1..|u|].$$

Example. Let $\bar{w} = w_1 \otimes w_2$, $\bar{w}' = w_1' \otimes w_2'$ and u be defined as follows.



Now let p' be the function defined on $\{1, \ldots, \max(|\bar{w}|, |u|)\}$, which extends p by letting $p'(|\bar{w}| + j) = |\bar{w}'| + j$ for $1 \le j \le |u| - |\bar{w}|$. We show that p' is a witness for $\bar{w} \otimes u \sqsubseteq_{\mathsf{s}} \bar{w}' \otimes u'$.

By construction, p' is increasing and $(\bar{w} {\otimes} u)[i] = (\bar{w}' {\otimes} u')[p'(i)]$ for every $i \le \max(|\bar{w}|, |u|) = |\bar{w} {\otimes} u|$. We must now show that, for each such i, $\mathsf{type}((\bar{w} {\otimes} u)[1..i]) = \mathsf{type}((\bar{w}' {\otimes} u)[1..p'(i)])$, and that $p'(|u|) = |u'|$. For convenience, we write $\bar{w}_u$ for $\bar{w} \otimes u$ and $\bar{w}'_{u'}$ for $\bar{w}' \otimes u'$.

- If $1 \le i \le s_1$, then $p'(i) = p(i)$, the u-component of each letter of $\bar{w}_u[1..i]$ coincides with its w_ℓ-component, and the u'-component of each letter of $\bar{w}'_{u'}[1..p(i)]$ coincides with its w_ℓ'-component. It follows that $\mathsf{type}(\bar{w}_u[1..i])$ is the symmetric transitive closure of $\mathsf{type}(\bar{w}[1..i]) \cup \{(\ell, k+1)\}$. Similarly, $\mathsf{type}(\bar{w}'_{u'}[1..p(i)])$ is the symmetric transitive closure of $\mathsf{type}(\bar{w}'[1..p(i)]) \cup \{(\ell, k+1)\}$. Since $\mathsf{type}(\bar{w}[1..i]) = \mathsf{type}(\bar{w}'[1..p(i)])$, we have $\mathsf{type}(\bar{w}_u[1..i]) = \mathsf{type}(\bar{w}'_{u'}[1..p'(i)])$. In particular, we have $u[i] = w_\ell[i] = w_\ell'[p'(i)] = u'[p'(i)]$. If $i = |u|$, then $s_1 = i$ and, by definition, $u' = w_\ell'[1..p(i)]$. It follows that $|u'| = p(i) = p'(i)$.

- If $s_1 < i \leq s_2$, again we have $p'(i) = p(i)$. Moreover, $\mathsf{type}(\bar{w}_u[1..i]) = \mathsf{type}(\bar{w}[1..i]) \cup \{(k+1, k+1)\}$ since the u-component differs from any other component on at least one position less than or equal to i. For the same reason, $\mathsf{type}(\bar{w}'_{u'}[1..p'(i)]) = \mathsf{type}(\bar{w}'[1..p(i)]) \cup \{(k+1, k+1)\} = \mathsf{type}(\bar{w}_u[1..i])$. Also, by definition of the n_j, we get $u[i] = u'[p(i)]$ and, as above, if $i = |u|$, we find that $p(i) = p(|u'|)$.
- If $s_2 < i \leq |u|$, then $p'(i) - |\bar{w}'| = i - |\bar{w}| = |u[s_2+1..i]|$. In particular, $p'(|u|) = |\bar{w}'| + |u[s_2+1..i]| = |u'|$. Moreover, $\mathsf{type}(\bar{w}_u[1..i]) = \{(k+1, k+1)\} = \mathsf{type}(\bar{w}'_{u'}[1..p'(i)])$. ◀

▶ **Lemma 12.** *If φ is a formula in $\mathbf{\Sigma}_1[\sigma]$, then L_φ is $\sqsubseteq_\mathsf{s}$-upward closed.*

Proof. First observe that if the synchronized words $\bar{w} = w_1 \otimes \cdots \otimes w_k$ and $\bar{w}' = w'_1 \otimes \cdots \otimes w'_k$ satisfy $\bar{w} \sqsubseteq_\mathsf{s} \bar{w}'$, then, for all $i, j \in \{1, \ldots, k\}$, we have:

- $w_i \preceq w_j$ if and only if $w'_i \preceq w'_j$;
- $|w_i| = |w_j|$ if and only if $|w'_i| = |w'_j|$;
- if $|w_i| = |w'_i| > 0$, then w_i and w'_i have the same last letter.

We now proceed by induction on the number of quantified variables of φ. If φ is quantifier-free, these three properties show that L_φ is $\sqsubseteq_\mathsf{s}$-upward closed.

If φ is not quantifier-free, we have $\varphi(y_1, \ldots, y_k) = \exists x\ \psi(y_1, \ldots, y_k, x)$ for some $\psi \in \mathbf{\Sigma}_1[\sigma]$. Let $\bar{w}, \bar{w}' \in SW_k$ such that $\bar{w} \sqsubseteq_\mathsf{s} \bar{w}'$ and $\bar{w} \models \varphi$. Then there is $u \in \mathbb{A}^*$ such that $\bar{w} \otimes u \models \psi$. By Lemma 11, there also exists $u' \in \mathbb{A}^*$ such that $\bar{w} \otimes u \sqsubseteq_\mathsf{s} \bar{w}' \otimes u'$. Since $\|\psi\|$ is $\sqsubseteq_\mathsf{s}$-upward closed by induction, $\bar{w}' \otimes u' \models \psi$, and hence $\bar{w}' \models \varphi$. This completes the proof. ◀

▶ **Lemma 13.** *If $\bar{w}$ is a synchronized word, then the relation defined by $\uparrow_\mathsf{s}\bar{w}$ is $\mathbf{\Sigma}_1[\sigma]$-definable.*

Proof. Let $\bar{w} = w_1 \otimes \cdots \otimes w_k \in SW_k$. We define a formula $\varphi(z_1, \ldots, z_k)$ (dependent on $\bar{w}$) whose synchronized language is $\uparrow_\mathsf{s}\bar{w}$, using existential quantification on a set consisting of one variable for each w_i and one for each position within w_i. Formally, let $X = \{x_{i,j} : 1 \leq i \leq k, 1 \leq j \leq |w_i|\}$. Then $\varphi(z_1, \ldots, z_k) = \exists X. \psi(z_1, \ldots, z_k, X)$, where ψ is the conjunction of the following formulæ for every $i \in \{1, \ldots, k\}$:

(1) $z_i = x_{i,|w_i|}$;
(2) for every $1 \leq j < |w_i|$: $x_{i,j} \prec x_{i,j+1}$;
(3) for every $1 \leq j \leq |w_i|$: $\ell_{w_i[j]}(x_{i,j})$;
(4) for every $1 \leq i' \leq k$ and $j \leq \min\{|w_i|, |w_{i'}|\}$: $\mathsf{eq}(x_{i,j}, x_{i',j})$;
(5) for every $1 \leq i' \leq k$ and $j \leq \min\{|w_i|, |w_{i'}|\}$ such that $w_i[1..j] = w_{i'}[1..j]$: $x_{i,j} = x_{i',j}$.

Let $\bar{w}' = w'_1 \otimes \cdots \otimes w'_k$. First assume that $\bar{w}' \in \uparrow_\mathsf{s}\bar{w}$: we want to show that $\bar{w}' \in L_\varphi$. Let p be a witness function for $\bar{w} \sqsubseteq_\mathsf{s} \bar{w}'$. For each $1 \leq i \leq k$ and $1 \leq j \leq |w_i|$, let the word $w'_i[1..p(j)]$ be assigned to variable $x_{i,j}$. It is readily verified that $(w'_1, \ldots, w'_k)$ satisfies $\varphi(z_1, \ldots, z_k)$ and hence $\bar{w} \in \|\varphi\|$.

Conversely, suppose that $\bar{w}' \in L_\varphi$. There exists an assignment α for the variables in φ such that $(\bar{w}, \alpha) \vDash \varphi$. We define a function $p\colon \{1, \ldots, |\bar{w}|\} \to \{1, \ldots, |\bar{w}'|\}$ as follows.

If $1 \leq j \leq |\bar{w}|$, there exists $1 \leq i \leq k$ such that $j \leq |w_i|$, and we let $p(j) = |\alpha(x_{i,j})|$. Condition (4) in the definition of φ shows that p is well-defined. Condition (2) shows that it is increasing and Condition (3) shows that $\bar{w}[j] = \bar{w}'[p(j)]$, so that p is a witness function for $\bar{w} \sqsubseteq \bar{w}'$. Finally, Conditions (5) and (1) show that p is type- and end-preserving, thus establishing that it is a witness for $\bar{w} \sqsubseteq_s \bar{w}'$. ◀

5.2 Characterization of $\mathcal{B}\Sigma_1[\sigma]$

The following lemma establishes one of the implications of Theorem 5.

▶ **Lemma 14.** *For every $\varphi \in \mathcal{B}\mathbf{\Sigma}_1[\sigma]$, $L_\varphi \in \bar{\mathcal{V}}$.*

Proof. By a standard transformation, φ is logically equivalent to a formula $\bigvee_{i=1}^{n} \psi_i \wedge \psi'_i$ in disjunctive normal form, where $\psi_i \in \mathbf{\Sigma}_1[\sigma]$ and $\psi'_i \in \mathbf{\Pi}_1[\sigma]$ for every i. By Corollary 6, there exist finite sets $\mathcal{S}_i$ and $\mathcal{S}'_i$ $(1 \le i \le n)$ of synchronized words such that $L_{\psi_i} = \uparrow_{\mathsf{s}}\mathcal{S}_i$ and $L_{\neg\psi'_i} = \uparrow_{\mathsf{s}}\mathcal{S}'_i$. Let $\mathcal{S} = \bigcup_{i=1}^{n} \mathcal{S}_i \cup \mathcal{S}'_i$ and let $h = \max\{|\bar{u}| : \bar{u} \in \mathcal{S}\}$.

If $\bar{w}, \bar{w}'$ are synchronized words such that $\bar{w} \approx_h \bar{w}'$, then $\bar{w}$ and $\bar{w}'$ have the same synchronized subwords of length at most h, and hence the same synchronized subwords in each $\mathcal{S}_i$ and each $\mathcal{S}'_i$ (since these sets contain only words of length at most h). If $\bar{w} \in L_\varphi$, then for some i, $\bar{w}$ contains a synchronized subword in $\mathcal{S}_i$ and none in $\mathcal{S}'_i$. The same holds therefore for $\bar{w}'$, and $\bar{w}' \in L_{\varphi_i \wedge \psi'_i}$. Thus $\bar{w}' \in L_\varphi$, which completes the proof. ◀

To establish the converse implication, we consider $h \in \mathbb{N}$ and $L \in \bar{\mathcal{V}}_h$, such that L is a finite union of $\approx_h$-classes $[\bar{w}_1]_h, \ldots, [\bar{w}_n]_h$, and we show that $L = L_\varphi$ for some $\varphi \in \mathcal{B}\mathbf{\Sigma}_1[\sigma]$.

For each $1 \le i \le n$, let $\mathcal{S}_i$ be the set of synchronized subwords of $\bar{w}_i$ of length at most h, and $\mathcal{S}'_i$ be the complement of $\mathcal{S}_i$ within the set of synchronized words of length at most h. Both are finite and, by Lemma 13, there exist $\mathbf{\Sigma}_1[\sigma]$-formulæ ψ_i and ψ'_i such that $L_{\psi_i} = \uparrow_{\mathsf{s}}\mathcal{S}_i$ and $L_{\psi'_i} = \uparrow_{\mathsf{s}}\mathcal{S}'_i$. Then, for each $1 \le i \le n$, $[\bar{w}_i]_h = \uparrow_{\mathsf{s}}\mathcal{S}_i \setminus \uparrow_{\mathsf{s}}\mathcal{S}'_i = L_{\psi_i \wedge \neg\psi'_i}$ and hence, $L = L_\varphi$ with $\varphi = \bigvee_i \psi_i \wedge \neg\psi'_i$. This completes the proof of Theorem 5, since $\varphi \in \mathcal{B}\mathbf{\Sigma}_1[\sigma]$.

5.3 Deciding membership in $\|\mathbf{\Sigma}_1[\sigma]\|$

In view of Theorem 4 and of the properties of regular languages (namely the decidability of equality), membership decidability for $\|\mathbf{\Sigma}_1[\sigma]\|$ reduces to proving the following proposition.

▶ **Proposition 15.** *Given a regular language $L \subseteq SW_k$, its upward-closure $\uparrow_{\mathsf{s}}L$ is regular and computable.*

We begin with some preliminary definitions, which will also be used in the next section. For $\mathcal{S} \subseteq SW_k$, let $\uparrow^{\ell} \mathcal{S} = \{\bar{w} \in SW_k : \exists \bar{u} \in \mathcal{S}\ \bar{u} \sqsubseteq \bar{w}$ and $\bar{u}, \bar{w}$ have the same last letter$\}$. Let $\mathcal{A}$ be a deterministic automaton accepting L, with state set Q and initial state q_0. For $p, r \in Q$, we let $\mathcal{A}(p, r)$ be the same as $\mathcal{A}$, with p as initial state and $\{r\}$ as final states, and denote by $\mathsf{Lang}(\mathcal{A}(p, r))$ the language accepted by $\mathcal{A}(p, r)$.

We say that a state sequence $\bar{q} = (q_1, \ldots, q_n) \in Q^n$ is $\bar{T}$*-compatible* in $\mathcal{A}$ if q_1 is reachable from q_0 by reading a word in $\mathbb{A}_{-,T_1}\mathbb{A}^*_{T_1}$, q_2 is reachable from q_1 by reading a word in $\mathbb{A}_{T_1,T_2}\mathbb{A}^*_{T_2}$, etc. In addition, we require q_n to be a final state of $\mathcal{A}$. Observe that, given $\bar{T}$, the set of $\bar{T}$-compatible state sequences is finite and computable.

If $\bar{q}$ is $\bar{T}$-compatible, we let $L(\bar{T}, \bar{q}, i)$ be the intersection of the language accepted by $\mathcal{A}(q_{i-1}, q_i)$ with $\mathbb{A}_{T_{i-1},T_i}\mathbb{A}^*_{T_i}$ ($\mathbb{A}_{-,T_1}\mathbb{A}^*_{T_1}$ if $i = 1$). In particular, if $\bar{w} \in SW_k$ and $\mathsf{type\text{-}seq}(\bar{w}) = \bar{T}$, then $\bar{w} \in L$ if and only if there exists a $\bar{T}$-compatible sequence $\bar{q}$ such that $\bar{w} \in L(\bar{T}, \bar{q}, 1) \cdots L(\bar{T}, \bar{q}, n)$ (uniquely determined, due to determinism). Note that the n factors of $\bar{w}$ thus determined are its type factors. In particular, $L = \bigcup L(\bar{T}, \bar{q}, 1) \cdots L(\bar{T}, \bar{q}, n)$, where the union runs over all type sequences $\bar{T}$ and all $\bar{T}$-compatible state sequences $\bar{q}$ of $\mathcal{A}$. This is a finite union, all of whose terms are explicitly computable.

Proof of Proposition 15. For L, $\mathcal{A}$, $\bar{T}$, $\bar{q}$, i as above, let $\widehat{L}(\bar{T}, \bar{q}, i)$ be $\widehat{L}(\bar{T}, \bar{q}, i) = \uparrow L(\bar{T}, \bar{q}, i) \cap \mathbb{A}^*_{T_i}$ if T_i is not an end type or $\widehat{L}(\bar{T}, \bar{q}, i) = \uparrow^{\ell} L(\bar{T}, \bar{q}, i) \cap \mathbb{A}^*_{T_i}$ otherwise. We now show that

$$\uparrow_{\mathsf{s}}L = \bigcup \widehat{L}(\bar{T}, \bar{q}, 1) \cdots \widehat{L}(\bar{T}, \bar{q}, n).$$

Note that the closure $\uparrow K$ of a regular language K is regular and computable (by adding self loops to the states of an automaton accepting K), and the operation $L(\bar{T}, \bar{q}, i) \mapsto \widehat{L}(\bar{T}, \bar{q}, i)$ is therefore computable, implying Proposition 15.

The proof is essentially an application of Lemma 1. Suppose first that $\bar{w} \in \uparrow_{\mathsf{s}} L$, that is, there exists $\bar{u} \in L$ such that $\bar{u} \sqsubseteq_{\mathsf{s}} \bar{w}$. Let $\bar{T} = \mathsf{type\text{-}seq}(\bar{u}) = (T_1, \ldots, T_n)$ and let $\bar{q}$ be the $\bar{T}$-compatible state sequence determined by reading $\bar{u}$ in $\mathcal{A}$. By Lemma 1, $\bar{T} \sqsubseteq \mathsf{type\text{-}seq}(\bar{w})$, with a witness function t such that, for each $1 \leq i \leq n$, the i-th type factor $\bar{u}_i$ of $\bar{u}$ is a subword of $\bar{w}_{t(i)}$, the $t(i)$-th type factor of $\bar{w}$ (with an additional last letter condition if T_i is an end type). Therefore $\bar{u}_i$ is also a subword of $\bar{w}_{t(i-1)+1} \cdots \bar{w}_{t(i)}$, with the same last letter condition in the case of end types. Since $\bar{u}_i \in L(\bar{T}, \bar{q}, i)$, this means that $\bar{w}_{t(i-1)+1} \cdots \bar{w}_{t(i)} \in \widehat{L}(\bar{T}, \bar{q}, i)$ and hence $\bar{w} \in \widehat{L}(\bar{T}, \bar{q}, 1) \cdots \widehat{L}(\bar{T}, \bar{q}, n)$.

Conversely, suppose that $\bar{w} \in \widehat{L}(\bar{T}, \bar{q}, 1) \cdots \widehat{L}(\bar{T}, \bar{q}, n)$ for some type sequence $\bar{T}$ and $\bar{T}$-compatible state sequence $\bar{q}$. For each $1 \leq i \leq n$, let $\bar{u}_i \in L(\bar{T}, \bar{q}, i)$ be such that $\bar{u}_i \sqsubseteq \bar{w}_i$, with witness function p_i (and such that $p_i(|u_i|) = |w_i|$ if T_i is an end type). By construction of the $L(\bar{T}, \bar{q}, i)$'s, the word $\bar{u} = \bar{u}_1 \cdots \bar{u}_n$ is in L, with type factors $\bar{u}_1, \ldots, \bar{u}_n$. Moreover $\bar{u} \sqsubseteq_{\mathsf{s}} \bar{w}$ for the witness function obtained by "concatenating" the functions p_i: $p(i) = p_1(i)$ for $i \leq |\bar{u}_1|$, and $p(|\bar{u}_1 \cdots \bar{u}_{i-1}| + h) = |\bar{w}_1 \cdots \bar{w}_{i-1}| + p_i(h)$ for every $1 < i \leq n$ and $1 \leq h \leq |\bar{u}_i|$. ◀

5.4 Deciding membership in $\|\mathcal{B}\boldsymbol{\Sigma}_1[\sigma]\|$

As a first step, we note the following.

▶ **Lemma 16.** *If $\bar{T}$ is a type sequence, then $K(\bar{T})$ is $\mathcal{B}\boldsymbol{\Sigma}_1[\sigma]$-definable.*

Proof. Let $S_{\bar{T}}$ be the set of $\sqsubseteq_s$-minimal elements of $K(\bar{T})$, a finite set by Proposition 2. Then $K(\bar{T}) \subseteq \uparrow_{\mathsf{s}} S_{\bar{T}}$. Moreover, if $\bar{u} \in \uparrow_{\mathsf{s}} S_{\bar{T}}$, then $\bar{T} \sqsubseteq \mathsf{type\text{-}seq}(\bar{u})$ by Lemma 1. It follows that $K(\bar{T}) = \uparrow_{\mathsf{s}} S_{\bar{T}} \setminus \bigcup\{\uparrow_{\mathsf{s}} S_{\bar{T}'} : \bar{T} \sqsubseteq \bar{T}', \bar{T} \neq \bar{T}'\}$. The statement then follows from Theorem 4. ◀

Since there are finitely many type sequences $\bar{T}$ and each $K(\bar{T})$ is computable, membership of a language L in $\|\mathcal{B}\boldsymbol{\Sigma}_1[\sigma]\|$ is equivalent to the membership of each $L \cap K(\bar{T})$ in $\|\mathcal{B}\boldsymbol{\Sigma}_1[\sigma]\|$.

We now fix a type sequence $\bar{T} = (T_1, \ldots, T_n)$. Our next step is a technical characterization of $\|\mathcal{B}\boldsymbol{\Sigma}_1[\sigma]\|$ for languages within $K(\bar{T})$. For each $1 \leq i \leq n$, let $\mathcal{F}_i = \mathcal{B}\boldsymbol{\Sigma}_1[<, \mathbf{max}](\mathbb{A}^*_{T_i})$ if T_i is an end type in $\bar{T}$, and $\mathcal{F}_i = \mathcal{B}\boldsymbol{\Sigma}_1[<](\mathbb{A}^*_{T_i})$ otherwise. Let also $\mathcal{G}_1 = \{\mathbb{A}_{-,T_1}\mathbb{A}^*_{T_1} \cap H : H \in |\mathcal{F}_1|\}$ and, for $i \geq 2$, $\mathcal{G}_i = \{\mathbb{A}_{T_{i-1},T_i}\mathbb{A}^*_{T_i} \cap H : H \in |\mathcal{F}_i|\}$.

▶ **Lemma 17.** *A regular language $L \subseteq K(\bar{T})$ is in $\|\mathcal{B}\boldsymbol{\Sigma}_1[\sigma]\|$ if and only if, for each $\bar{T}$-compatible state sequence $\bar{q}$ and $1 \leq i \leq n$, $L(\bar{T}, \bar{q}, i) \in \mathcal{G}_i$.*

Proof. For convenience, we write $L(\bar{q}, i)$ for $L(\bar{T}, \bar{q}, i)$. First assume that every $L(\bar{q}, i) \in \mathcal{G}_i$, that is, there exists a $\mathcal{B}\boldsymbol{\Sigma}_1[<]$-definable language $H(\bar{q}, i) \subseteq \mathbb{A}^*_{T_i}$ such that $L(\bar{q}, i) = \mathbb{A}_{T_{i-1},T_i}\mathbb{A}^*_{T_i} \cap H(\bar{q}, i)$ (or, if $i = 1$, $L(\bar{q}, 1) = \mathbb{A}_{-,T_1}\mathbb{A}^*_{T_1} \cap H(\bar{q}, 1)$). Then $H(\bar{q}, i)$ is the finite union of languages of the form $H(\bar{q}, i, j) = \uparrow S(\bar{q}, i, j) \setminus \uparrow S'(\bar{q}, i, j)$ (or $\uparrow^\ell S(\bar{q}, i, j) \setminus \uparrow^\ell S'(\bar{q}, i, j)$ if T_i is an end type), for $1 \leq j \leq n_{\bar{q},i}$, where the $S(\bar{q}, i, j)$'s and $S'(\bar{q}, i, j)$'s are finite sets. Then $L(\bar{q}, i)$ is the union of the $L(\bar{q}, i, j) = \mathbb{A}_{T_{i-1},T_i}\mathbb{A}^*_{T_i} \cap H(\bar{q}, i, j)$ (or, if $i = 1$, $L(\bar{q}, 1, j) = \mathbb{A}_{-,T_1}\mathbb{A}^*_{T_1} \cap H(\bar{q}, 1, j)$). If $\bar{\jmath} = (j_1, \ldots, j_n)$ is such that $1 \leq j_i \leq n_{\bar{q},i}$ for each $1 \leq i \leq n$, let $L(\bar{q}, \bar{\jmath}) = L(\bar{q}, 1, j_1) \cdots L(\bar{q}, n, j_n)$. Then L is the (finite) union of the $L(\bar{q}, \bar{\jmath})$. Now let

- $\mathcal{S}(\bar{q}, \bar{\jmath}) = \{\bar{w} \in K(\bar{T}) : \text{ for all } i \in \{1, \ldots, n\}, \mathsf{type\text{-}factor}_i(\bar{w}) \in S(\bar{q}, i, j_i)\}$ and
- $\mathcal{S}'(\bar{q}, \bar{\jmath}) = \{\bar{w} \in K(\bar{T}) : \text{ for some } i \in \{1, \ldots, n\}, \mathsf{type\text{-}factor}_i(\bar{w}) \in S'(\bar{q}, i, j_i)\}$.

Then, $L(\bar{q}, \bar{\jmath}) = K(\bar{T}) \cap (\uparrow\mathcal{S}(\bar{q}, \bar{\jmath}) \setminus \uparrow\mathcal{S}'(\bar{q}, \bar{\jmath})) \in \|\mathcal{B}\boldsymbol{\Sigma}_1[\sigma]\|$. It follows that $L \in \|\mathcal{B}\boldsymbol{\Sigma}_1[\sigma]\|$.

Conversely, suppose that for some $\bar{q}$ and some i, $L(\bar{q}, i) \notin \mathcal{G}_i$. In view of Theorem 5, we want to show that L is not a union of $\approx_r$-classes for any $r \geq 1$. Let r be now fixed. We only need to exhibit words $\bar{w} \in L$ and $\bar{w}' \notin L$ such that $\bar{w} \approx_r \bar{w}'$.

Let $\sim_r^i$ be the relation on $\mathbb{A}_{T_i}^*$ given by $\bar{u} \sim_r^i \bar{v}$ if $\bar{u} \sim_r \bar{v}$ and, either the first letters of $\bar{u}$ and $\bar{v}$ are both in $\mathbb{A}_{T_{i-1}}$ or neither is. By means of contradiction, suppose $L(\bar{q}, i)$ is the finite union of the $\sim_r^i$ classes $[\bar{u}_1], \ldots, [\bar{u}_m]$. Then each $\bar{u}_j \in \mathbb{A}_{T_{i-1},T_i}\mathbb{A}_{T_i}^*$ and, by definition of $\sim_r^i$, $[\bar{u}_j] = \mathbb{A}_{T_{i-1},T_i}\mathbb{A}_{T_i}^* \cap [\![\bar{u}_j]\!]$, where $[\![\bar{u}_j]\!]$ denotes the $\sim_r$-class of $\bar{u}_j$. Therefore $L(\bar{q}, i) = \mathbb{A}_{T_{i-1},T_i}\mathbb{A}_{T_i}^* \cap M$, where M is the union of the $[\![\bar{u}_j]\!]$'s. Since $M \in |\mathcal{B}\mathbf{\Sigma}_1[<](\mathbb{A}_{T_i})|$, this shows that $L(\bar{q}, i) \in \mathcal{G}_i$, a contradiction. (In the case where T_i is an end type, we need to reason with the intersection of $\sim_r^i$ with the same-last-letter equivalence.)

Now, since $L(\bar{q}, i)$ is not a union of $\sim_r^i$-classes, there exist words $\bar{u}_i, \bar{u}'_i \in \mathbb{A}_{T_{i-1},T_i}\mathbb{A}_{T_i}^*$ such that $\bar{u}_i \sim_r \bar{u}'_i$ (and if T_i is an end type they have the same last letter) and exactly one of them is in $L(\bar{q}, i)$. Say $\bar{u}_i \in L(\bar{q}, i)$, such that $\bar{u}_i \in \mathsf{Lang}(\mathcal{A}(q_{i-1}, q_i))$ and $\bar{u}'_i \in \mathsf{Lang}(\mathcal{A}(q_{i-1}, q'_i))$ for some $q_i \neq q'_i$. Assuming wlog that $\mathcal{A}$ is minimal for L, there exist a word $\bar{y}$ and states p, p' of which exactly one is accepting, such that $\bar{y} \in \mathsf{Lang}(\mathcal{A}(q_i, p)) \cap \mathsf{Lang}(\mathcal{A}(q'_i, p'))$. Let $\bar{x} \in L(\bar{q}, 1) \cdots L(\bar{q}, i-1)$, $\bar{w} = \bar{x}\bar{u}_i\bar{y}$ and $\bar{w}' = \bar{x}\bar{u}'_i\bar{y}$. Then exactly one of $\bar{w}$, $\bar{w}'$ is in L. Since $L \subseteq K(\bar{T})$, this implies that $\bar{y} \in \mathbb{A}_{T_i}^*\mathbb{A}_{T_i,T_{i+1}}\mathbb{A}_{T_{i+1}}^* \cdots \mathbb{A}_{T_{n-1},T_n}\mathbb{A}_{T_n}^*$ ($\mathbb{A}_{T_n}^*$ if $i = n$). Consequently, $\bar{w}$ and $\bar{w}'$ have the same type sequence $\bar{T}$, with the same type factors except for the i-th one. Moreover, $\mathsf{type\text{-}factor}_i(\bar{w}) = \bar{u}_i\bar{u}'$ and $\mathsf{type\text{-}factor}_i(\bar{w}') = \bar{u}'_i\bar{u}'$, where $\bar{u}'$ is the longest $\mathbb{A}_{T_i}^*$ prefix of $\bar{u}$. Since $\bar{u}_i \sim_r \bar{u}'_i$, $\bar{u}_i\bar{u}' \sim_r \bar{u}'_i\bar{u}'$. Then $\bar{w} \approx_r \bar{w}'$ is a consequence of Lemma 1. ◀

In view of Lemma 17 and since each $L(\bar{T}, \bar{q}, i)$ is computable (Section 5.3), the decidability of $\|\mathcal{B}\mathbf{\Sigma}_1[\sigma]\|$ will be established if we show that membership in each $\mathcal{G}_i$ is decidable, which is the object of the following lemma.

▶ **Lemma 18.** *Let A be an alphabet and $B \subseteq A$. Then, it is decidable whether a regular language is in $\mathcal{W}_B = \{BA^* \cap L \colon L \in \mathcal{B}\mathbf{\Sigma}_1[<](A)\}$.*

Proof. We will prove the following characterization of $\mathcal{W}_B$: a regular language $K \in \mathcal{W}_B$ if and only if $K \subseteq BA^*$ and for every $b \in B$, $b^{-1}K = \{u \in A^* \colon bu \in K\} \in |\mathcal{B}\mathbf{\Sigma}_1[<](A)|$. Since $\mathcal{B}\mathbf{\Sigma}_1[<]$ membership is decidable [20], the result follows directly.

If $K \in \mathcal{W}_B$, then $K = BA^* \cap L$ for some $L \in |\mathcal{B}\mathbf{\Sigma}_1[<](A)|$. Therefore, $K \subseteq BA^*$ and, for every $b \in B$, $b^{-1}K = b^{-1}L \in |\mathcal{B}\mathbf{\Sigma}_1[<]|$ (since $\mathcal{B}\mathbf{\Sigma}_1[<]$ is closed under left quotients).

Conversely, suppose that each $b^{-1}K$ ($b \in B$) is a $\mathcal{B}\mathbf{\Sigma}_1[<]$-language. Then there exists r such that each of these languages is a union of $\sim_r$-classes. Say that $u \sim_{r+1}^B v$ if $u \sim_{r+1} v$ and, either u, v have the same first letter in B or both their first letters are in $A \setminus B$. Suppose there exist words u, v such that $u \sim_{r+1}^B v$ and $u \in K$. Then $u \in BA^*$ and v has same first letter as u, say b, so that $u = bu'$ and $v = bv'$. In particular, $u' \sim_r v'$. Since $u' \in b^{-1}K$ and $b^{-1}K$ is a union of $\sim_r$-classes, it follows that $v' \in b^{-1}K$ and hence $v \in K$. Therefore, K is the union of the $\sim_{r+1}^B$-classes of a finite set of words $u_1, \ldots, u_n$ and if L is the union of the $\sim_{r+1}$-classes of the same words, then $L \in |\mathcal{B}\mathbf{\Sigma}_1[<]|$ and $L \cap BA^* = K$. That is, $K \in \mathcal{W}_B$. ◀

6 $\Sigma_2[\sigma]$ and its boolean closure

For any alphabet A, an *A-monomial* is a language of the form $A_1^*a_1A_2^*a_2 \cdots A_n^*a_nA_{n+1}^*$, where $A_1, A_2, \ldots, A_{n+1} \subseteq A$ and $a_1, a_2, \ldots, a_n \in A$. An *$A$-polynomial* is a finite union of A-monomials.

▶ Remark 19. It is known [14] that $\Sigma_2[<](A)$ sentences define exactly the set of A-polynomials. A non-trivial consequence is that the set of A-polynomials is closed under intersection.

Not every $\mathbb{A}_\perp^k$-polynomial respects the structural properties (on the positions of $\perp$) of synchronized words. For example $(a,\perp)^*(b,b)(a,a)^*$ is a polynomial over $\mathbb{A}_\perp^k$ for $\mathbb{A} = \{a,b\}$ and $k = 2$ but it does not define a relation. In order to characterize subsets of SW_k which are polynomials, we introduce the notion of $\perp$-*consistency*.

If $\bar{a} = a_1 \otimes \cdots \otimes a_k$ is a synchronized letter in $\mathbb{A}_\perp^k$, we denote by $\tau(\bar{a})$ the set $\{i \in \{1,\ldots,k\} \colon a_i = \perp\}$. A non-empty subset $\bar{A} \subseteq \mathbb{A}_\perp^k$ is said to be $\perp$-*consistent* if all the $\tau(\bar{a})$ $(\bar{a} \in \bar{A})$ take the same value. If that is the case, we let $\tau(\bar{A}) = \tau(\bar{a})$ (for any $\bar{a} \in \bar{A}$). Finally we say that a monomial $\bar{A}_0^*\bar{a}_1\bar{A}_1^* \ldots \bar{a}_n\bar{A}_n^*$ (over $\mathbb{A}_\perp^k$) is $\perp$-*consistent* if and only if every non-empty $\bar{A}_i$ is $\perp$-consistent and the sequence $\tau(\bar{A}_0), \tau(\bar{a}_1), \tau(\bar{A}_1), \ldots, \tau(\bar{a}_n), \tau(\bar{A}_n)$ is $\subseteq$-increasing (where the term $\tau(\bar{A}_i)$ is skipped if $\bar{A}_i = \emptyset$).

We denote by $\bar{\mathcal{P}}$ the set of all $\perp$-*consistent polynomials*, that is, of finite unions of $\perp$-consistent monomials. The following statement follows directly from this definition.

▶ **Lemma 20.** *Let $L = \bar{A}_0^*\bar{a}_1\bar{A}_1^* \ldots \bar{a}_n\bar{A}_n^*$ be an $\mathbb{A}_\perp^k$-monomial. Then $L \subseteq SW_k$ if and only if $L \in \bar{\mathcal{P}}$. Moreover, if L is $\perp$-consistent and $S \subseteq \{1,\ldots,k\}$, then $\pi_S(L)$ is a $\perp$-consistent monomial as well.*

We can now proceed with the proof of Theorem 7. We first show that a $\mathbf{\Sigma}_2[\sigma]$-definable k-ary relation R satisfies $L_R \in \bar{\mathcal{P}}$. Indeed, without loss of generality, R is defined by a $\mathbf{\Sigma}_2[\sigma]$-formula φ with free variables $S = \{z_1,\ldots,z_k\}$, of the form $\varphi = \exists x_1 \ldots \exists x_n\ \psi(x_1,\ldots,x_n,z_1,\ldots,z_k)$, with $\psi \in \mathbf{\Pi}_1[\sigma]$. In particular, $\|\psi\|$ is a $(n+k)$-ary relation and $R = \pi_S(\|\psi\|)$. Lemmas 20 and 21 therefore establish that $L_R \in \bar{\mathcal{P}}$.

▶ **Lemma 21.** *If $R \in \|\mathbf{\Pi}_1[\sigma]\|$ then $L_R \in \bar{\mathcal{P}}$.*

Proof. Since $R \in \|\mathbf{\Pi}_1[\sigma]\|$, it is the complement of a $\mathbf{\Sigma}_1[\sigma]$-definable relation. By Corollary 6, we have $L_R = SW_k \setminus {\uparrow_{\mathsf{s}}}\mathcal{S}$ for some finite set $\mathcal{S}$. Since $\bar{\mathcal{P}}$ is closed under intersection (see Remark 19), we only need to show that $SW_k \setminus {\uparrow_{\mathsf{s}}}\bar{w} \in \bar{\mathcal{P}}$ for a single synchronized word $\bar{w}$.

Let $\bar{T} = (T_1,\ldots,T_n) = \mathsf{type\text{-}seq}(\bar{w})$. By Lemma 1, we see that $\bar{u} \in L_R$ if and only either (1) $\bar{T} \not\sqsubseteq \mathsf{type\text{-}seq}(\bar{u})$, or (2) $\bar{T} \sqsubseteq \mathsf{type\text{-}seq}(\bar{u})$, with witness function t and $\bar{w}_i \not\sqsubseteq \bar{u}_{t(i)}$ for some $1 \le i \le n$ or, (3) again $\bar{T} \sqsubseteq \mathsf{type\text{-}seq}(\bar{u})$ with witness t, where $\bar{w}_i$ and $\bar{u}_{t(i)}$ do not have the same last letter for some i such that T_i is an end-type for $\bar{w}$.

The first condition means that $\bar{u} \in \bigcup K(\bar{T}')$, where the union runs over type sequences $\bar{T}'$ such that $\bar{T} \not\sqsubseteq \bar{T}'$. We saw in Section 2 that this union is in $\bar{\mathcal{P}}$. The second condition places $\bar{u}$ in $C_i = \mathbb{A}_{-,T_1}\mathbb{A}_{T_1}^* \cdots \mathbb{A}_{T_{i-2},T_{i-1}}\mathbb{A}_{T_{i-1}}^*\ L_i\ \mathbb{A}_{T_i,T_{i+1}}\mathbb{A}_{T_{i+1}}^* \cdots \mathbb{A}_{T_{n-1},T_n}\mathbb{A}_{T_n}^*$, where L_i is the set of words in $\mathbb{A}_{T_{i-1}T_i}\mathbb{A}_{T_i}^*$ that do not have $\bar{w}_i$ as a subword. Then L_i is $\mathbf{\Pi}_1[<]$-, and hence $\mathbf{\Sigma}_2[<]$-definable. As a consequence, $L_i \in \bar{\mathcal{P}}$ and, by associativity, $C_i \in \bar{\mathcal{P}}$. Finally, the third condition places $\bar{u}$ in $C_i' = \mathbb{A}_{-,T_1}\mathbb{A}_{T_1}^* \cdots \mathbb{A}_{T_{i-2},T_{i-1}}\mathbb{A}_{T_{i-1}}^*\ L_i'\ \mathbb{A}_{T_i,T_{i+1}}\mathbb{A}_{T_{i+1}}^* \cdots \mathbb{A}_{T_{n-1},T_n}\mathbb{A}_{T_n}^*$, where $L_i' = (\mathbb{A}_{T_i} \setminus \mathbb{A}_{T_{i-1}})\mathbb{A}_{T_i}^* \cap \mathbb{A}_{T_i}^* B_i$, with B_i the set of letters of $\mathbb{A}_{T_i}$ different from the last letter of $\bar{w}_i$. Here too, $L_i' \in \bar{\mathcal{P}}$ and hence $C_i' \in \bar{\mathcal{P}}$. ◀

The following lemma then concludes the proof of Theorem 7.

▶ **Lemma 22.** *If R is a relation such that L_R is a $\perp$-consistent polynomial, then $R \in \|\mathbf{\Sigma}_2[\sigma]\|$.*

Proof. By definition of $\bar{\mathcal{P}}$, the proof reduces to the case where L_R is a $\perp$-consistent monomial, say $L_R = \bar{A}_0^*\bar{a}_1\bar{A}_1^* \cdots \bar{a}_n\bar{A}_n^*$. We now construct a $\mathbf{\Sigma}_2[\sigma]$-formula φ, with set of free variables $Z = \{z_1,\ldots,z_k\}$, which defines R.

Let $w_1,\ldots,w_k \in \mathbb{A}^*$ be such that $\bar{w} = w_1 \otimes \cdots \otimes w_k = \bar{a}_1 \cdots \bar{a}_n$ (they exist due to $\perp$-consistency). Let $X = \{x_{i,j} : 1 \le i \le k, 1 \le j \le |w_i|\}$ and $Y = \{y_1,\ldots,y_k\}$ be sets of variables. We first let $\psi_1(X,Z)$ be the conjunction of the following formulæ for $1 \le i \le k$:

(1) for every $1 \le j < |w_i|$: $x_{i,j} \prec x_{i,j+1}$; (2) $x_{i,|w_i|} \preceq z_i$; (3) for every $1 \le j \le |w_i|$: $\ell_{w_i[j]}(x_{i,j})$; (4) for every $1 \le i' \le k$ and $j \le \min\{|w_i|, |w_{i'}|\}$: $\mathsf{eq}(x_{i,j}, x_{i',j})$. Notice that $\bar{u} \in SW_k$ satisfies $\exists X\ \psi_1(X, Z)$ if and only if $\bar{w}$ is a subword of $\bar{x}$ with witness function given by $p(j) = |x_{h,j}|$ for any $1 \le h \le k$.

The variables in Y are meant to represent the k components of a prefix of $\bar{u}$, which is expressed by $\psi_2(X, Y)$, the disjunction over all subsets H of $\{1, \ldots, k\}$ (H represents the components of $\bar{u}$ which are shorter than that prefix) of the formulæ

$$\bigwedge_{h \in H} (y_h \preceq z_h \wedge \mathsf{eq}(y_h, z_h)) \wedge \bigwedge_{h,i \notin H} (y_h \prec z_h) \wedge \mathsf{eq}(y_h, y_i)) \wedge \bigwedge_{h \in H, i \notin H} \exists r (r \preceq y_i \wedge \mathsf{eq}(r, y_h)).$$

Next, for $\bar{a} \in \mathbb{A}_\perp^k$ and $\bar{A} \subseteq \mathbb{A}_\perp^k$, and recalling that $\tau(\bar{a}) = \{h : \pi_h(\bar{a}) = \perp\}$, we define $\psi_{\bar{a}}(Y) = \bigwedge_{h \notin \tau(\bar{a})} \ell_{\pi_h(\bar{a})}(y_h)$ and $\psi_{\bar{A}}(Y) = \bigvee_{\bar{a} \in \bar{A}} \psi_{\bar{a}}(Y)$. Once $\bar{y}$ is a prefix of $\bar{u}$ and $\bar{w}$ is a subword of $\bar{u}$ with witness function p, if for some $1 \le j \le n$, we have $|\bar{y}| = p(j)$, then $\bar{y}$ satisfies $\psi_{\bar{a}_j}$. We now only need to verify that if $|\bar{y}|$ sits between $p(j)$ and $p(j+1)$ (for some $0 \le j \le n$), then $\bar{y}$ satisfies $\psi_{\bar{A}_j}$. This is done by the formula $\psi_3(X, Y) = \bigwedge_{j=0}^n \chi_j$, where $\chi_0(X, Y) = \left(\bigwedge_{h \notin \tau(\bar{A}_0)} y_h \prec x_{h,1}\right) \to \psi_{\bar{A}_0}(Y)$, $\chi_n(X, Y) = \left(\bigwedge_{h \notin \tau(\bar{A}_n)} x_{h,n} \prec y_h\right) \to \psi_{\bar{A}_n}(Y)$, and for every $0 < j < n$,

$$\chi_j(X, Y) = \left(\bigwedge_{h \notin \tau(\bar{A}_j)} (x_{h,j} \prec y_h) \wedge (y_h \prec x_{h,j+1})\right) \to \psi_{\bar{A}_j}(Y).$$

Finally, R is defined by the $\mathbf{\Sigma}_2[\sigma]$ formula $\varphi(Z) = \exists X \psi_1(X, Z) \wedge \forall Y\ (\psi_2(X, Y) \wedge \psi_3(X, Y))$. ◀

References

1 Parosh Aziz Abdulla, Bengt Jonnson, Marcus Nilsson, and Mayank Saksena. A survey of regular model checking. In *International Conference on Concurrency Theory (CONCUR)*, pages 35–48, 2003.

2 Pablo Barceló, Leonid Libkin, Anthony Widjaja Lin, and Peter T. Wood. Expressive Languages for Path Queries over Graph-Structured Data. *ACM Transactions on Database Systems (TODS)*, 37(4):31, 2012. doi:10.1145/2389241.2389250.

3 Michael Benedikt, Leonid Libkin, Thomas Schwentick, and Luc Segoufin. Definable relations and first-order query languages over strings. *Journal of the ACM*, 50(5):694–751, 2003. doi:10.1145/876638.876642.

4 Jean Berstel. *Transductions and Context-Free Languages*. B. G. Teubner, 1979.

5 Achim Blumensath and Erich Grädel. Automatic Structures. In *Annual IEEE Symposium on Logic in Computer Science (LICS)*, pages 51–62. IEEE Computer Society Press, 2000. doi:10.1109/LICS.2000.855755.

6 Ahmed Bouajjani, Bengt Jonsson, Marcus Nilsson, and Tayssir Touili. Regular Model Checking. In *International Conference on Computer Aided Verification (CAV)*, pages 403–418. Springer, 2000.

7 J. Richard Büchi. On a Decision Method in Restricted Second-Order Arithmetic. In *Proc. Int. Congr. for Logic, Methodology, and Philosophy of Science*, pages 1–11. Stanford Univ. Press, 1962.

8 Christian Choffrut. Relations over Words and Logic: A Chronology. *Bulletin of the EATCS*, 89:159–163, 2006.

9 Samuel Eilenberg, Calvin C. Elgot, and John C. Shepherdson. Sets recognized by n-tape automata. *Journal of Algebra*, 13(4):447–464, 1969. doi:10.1016/0021-8693(69)90107-0.

10 Graham Higman. Ordering by divisibility in abstract algebras. *Proceedings of the London Mathematical Society (3)*, 2(7):326–336, 1952. `doi:10.1112/plms/s3-2.1.326`.

11 M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, Reading, MA, 1983. Reprinted by *Cambridge University Press*, 1997.

12 Robert McNaughton and Seymour Papert. *Counter-Free Automata*. The MIT Press, Cambridge, Mass., 1971.

13 Jean-Éric Pin. *Varieties of Formal Languages*. North Oxford Academic, London, 1986.

14 Jean-Éric Pin and Howard Straubing. Monoids of upper triangular matrices. In *Colloquia Mathematica Societatis Janos Bolyai*, pages 259–272, 1981.

15 Jean-Éric Pin and Pascal Weil. Polynomial Closure and Unambiguous Product. *Theory of Computing Systems*, 30(4):383–422, 1997.

16 Thomas Place and Marc Zeitoun. Going Higher in the First-Order Quantifier Alternation Hierarchy on Words. In *International Colloquium on Automata, Languages and Programming (ICALP)*, pages 342–353, 2014. `doi:10.1007/978-3-662-43951-7_29`.

17 Thomas Place and Marc Zeitoun. The tale of the quantifier alternation hierarchy of first-order logic over words. *SIGLOG News*, 2(3):4–17, 2015. `doi:10.1145/2815493.2815495`.

18 Thomas Place and Marc Zeitoun. Concatenation Hierarchies: New Bottle, Old Wine. In *International Computer Science Symposium in Russia (CSR)*, pages 25–37, 2017. `doi:10.1007/978-3-319-58747-9_5`.

19 Marcel Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.

20 Imre Simon. Piecewise testable events. In H. Barkhage, editor, *Automata Theory and Formal Languages, 2nd GI Conference, Kaiserslautern, May 22–23, 1975*, volume 33 of *LNCS*, pages 214–222. Springer, 1975.

21 Howard Straubing. *Finite Automata, Formal Logic, and Circuit Complexity*. Birkhäuser, Boston, Basel and Berlin, 1994.

22 Howard Straubing and Pascal Weil. An introduction to automata theory. In Deepak D'Souza and Priti Shankar, editors, *Modern applications of automata theory*, volume 2 of *I.I.Sc. Monographs*, pages 3–43. World Scientific, 2012.