Logical Methods in Computer Science
Vol. 8 (3:11) 2012, pp. 1–30
www.lmcs-online.org

Submitted Feb. 12, 2011
Published Aug. 13, 2012

ON LOGICAL HIERARCHIES WITHIN FO^2-DEFINABLE LANGUAGES *

MANFRED KUFLEITNER [a] AND PASCAL WEIL [b]

[a] University of Stuttgart, Germany
e-mail address: kufleitner@fmi.uni-stuttgart.de

[b] Univ. Bordeaux, LaBRI, UMR 5800, F-33400 Talence, France
CNRS, LaBRI, UMR 5800, F-33400 Talence, France
e-mail address: pascal.weil@labri.fr

Abstract. We consider the class of languages defined in the 2-variable fragment of the first-order logic of the linear order. Many interesting characterizations of this class are known, as well as the fact that restricting the number of quantifier alternations yields an infinite hierarchy whose levels are varieties of languages (and hence admit an algebraic characterization). Using this algebraic approach, we show that the quantifier alternation hierarchy inside $\mathsf{FO}^2[<]$ is decidable within one unit. For this purpose, we relate each level of the hierarchy with decidable varieties of languages, which can be defined in terms of iterated deterministic and co-deterministic products. A crucial notion in this process is that of condensed rankers, a refinement of the rankers of Weis and Immerman and the turtle languages of Schwentick, Thérien and Vollmer.

Many important properties of systems are modeled by finite automata. Frequently, the formal languages induced by these systems are definable in first-order logic. Our understanding of its expressive power is of direct relevance for a number of application fields, such as verification.

The first-order logic we are interested in, in this paper, is the first-order logic of the linear order, written $\mathsf{FO}[<]$, interpreted on finite words. It is well-known that the languages that are definable in this logic are exactly the star-free languages, or equivalently the regular languages whose syntactic monoid is aperiodic (that is: satisfies an identity of the form $x^{n+1} = x^n$ for some integer n) [24, 18] (see also [5, 20, 27, 29]); and that deciding whether a finite automaton accepts such a language is PSPACE-complete [3].

Fragments of first-order logic defined by the limitation of certain resources have been studied in detail. For instance, the quantifier alternation hierarchy, with its close relation with the dot-depth hierarchy of star-free languages, offers one of the oldest open problems in formal language theory: we know that the hierarchy is infinite and that its levels are

1998 ACM Subject Classification: F.4.3, F.4.1.
Key words and phrases: alternation hierarchy, two-variable fragment of first-order logic, rankers.
* An extended abstract of this paper has been published in the proceedings of MFCS 2009.
[a] The first author was supported by the German Research Foundation (DFG) under grant DI 435/5-1.
[b] The second author was supported by the grant ANR 2010 BLAN 0202 01 FREC.

LOGICAL METHODS IN COMPUTER SCIENCE
DOI:10.2168/LMCS-8 (3:11) 2012

characterized algebraically (by a property of the syntactic monoids), but we do not know whether these levels (besides levels 0 and 1) are decidable. In contrast, it is known that the quantifier alternation hierarchy for the first-order logic of the successor, $\mathsf{FO}[S]$, collapses at level 2 [34, 21].

Another natural limitation considers the number of variables in a formula. This limitation has attracted a good deal of attention, as the trade-off between formula size and number of variables is known to be related with the trade-off between parallel time and number of processes, see [11, 1, 9].

It is well-known that every first-order formula is equivalent to one using at most three variables. On the other hand, the first-order formulas using at most two variables, written $\mathsf{FO}^2[<]$, are strictly less expressive. The class of languages defined by such formulas admits many remarkable characterizations [31]. To begin with, a language is $\mathsf{FO}^2[<]$-definable if and only if it is recognized by a monoid in the pseudovariety **DA** [33] (a precise definition will be given in Section 2). As with the characterization of $\mathsf{FO}[<]$-definability by aperiodic monoids, this characterization implies decidability. The $\mathsf{FO}^2[<]$-definable languages are also characterized in terms of unambiguous products of languages (see Section 6.3) and in terms of the unary fragment of propositional temporal logic [7] (see Section 1.3). For a survey of these properties, the reader is referred to [31, 6].

In this paper, we consider the quantifier alternation hierarchy within the two-variable fragment of first-order logic. We denote by $\mathsf{FO}^2_m[<]$ the fragment of $\mathsf{FO}^2[<]$ consisting of formulas using at most 2 variables and at most m alternating blocks of quantifiers. In the sequel, we omit specifying the predicate $<$ and we write simply FO, FO^2 or FO^2_m.

Schwentick, Thérien and Vollmer introduced the so-called *turtle programs* to characterize the expressive power of FO^2 [26]. These programs are sequences of directional instructions of the form *go to the next a to the right, go to the next b to the left*. More details can be found in Section 1.2 below. Turtle programs were then used, under the name of *rankers*, by Weis and Immerman [37] (first published in [36]) to characterize FO^2_m in terms of rankers with m alternations of directions (right *vs.* left). Their subtle characterization, Theorem 1.8 below, does not yield a decidability result. It forms however the basis of our results.

Rankers are actually better suited to the study of a natural alternation hierarchy within the unary fragment of propositional temporal logic (Sections 1.3 and 4), than to the study of the quantifier alternation within $\mathsf{FO}^2[<]$. For the latter, we define the notion of *condensed* rankers, which introduce a notion of efficiency in the path they describe in a word, see Section 3.

Recent results of Kufleitner and Lauser [16] and Straubing [28] show that $\mathcal{FO}^2_m$ (the set of FO^2_m-definable languages) forms a variety of languages. We show that the classes of languages defined by condensed rankers with at most m changes of directions also form varieties of languages, written $\mathcal{R}_m$ and $\mathcal{L}_m$ depending on whether the initial move is towards the right or towards the left (Section 3.3). The meaning of these results is that membership of a language L in these classes depends only on the syntactic monoid of L. This justifies using algebraic methods to approach the decidability problem for $\mathcal{FO}^2_m$ — a technique that has proved very useful in a number of situations (see for instance [20, 31, 30, 6]).

In fact, we use this algebraic approach to show that the classes $\mathcal{R}_m$ and $\mathcal{L}_m$ are decidable (Section 3.5), and that they admit a neat characterization in terms of closure under alternated deterministic and co-deterministic products (Section 3.4). Moreover, we show

(Theorem 5.1) that

$$\mathcal{R}_m \cup \mathcal{L}_m \subseteq \mathcal{FO}^2_m \subseteq \mathcal{R}_{m+1} \cap \mathcal{L}_{m+1}.$$

This shows that one can effectively compute, given a language L in $\mathcal{FO}^2$, an integer m such that L is in $\mathcal{FO}^2_{m+1}$, possibly in $\mathcal{FO}^2_m$, but not in $\mathcal{FO}^2_{m-1}$. That is, we can compute the quantifier alternation depth of L within one unit. As indicated above, this is much more precise than the current level of knowledge on the general quantifier alternation hierarchy in $\mathsf{FO}[<]$.

We conjecture that $\mathcal{FO}^2_m$ is actually equal to the intersection of $\mathcal{R}_{m+1}$ and $\mathcal{L}_{m+1}$. This would prove that each $\mathcal{FO}^2_m$ is decidable.

Many of these results were announced in [14], with a few differences. In particular, the definition of the sets $\underline{R}^{\mathsf{X}}_{m,n}$ (Section 1.2) in [14] introduced a mistake which is corrected here. The proof of [14, Theorem 2] contained a gap: we do not have a proof that the classes defined by the alternation hierarchy within unary temporal logic are varieties. And the proof of [14, Proposition 2.9] also contained a gap: the correct statement is Theorem 4.3 below.

1. Rankers and logical hierarchies

Let A be a finite alphabet. We denote by A^* the set of all words over A (that is, of sequences of elements of A), and by A^+ the set of non-empty words. If u is a length n ($n > 0$) word over A, we say that an integer $1 \leq i \leq n$ is an *a-position* of u if the i-th letter of u, written $u[i]$, is an a. If $1 \leq i \leq j \leq n$, we let $u[i;j]$ be the factor $u[i] \cdots u[j]$ of u.

FO denotes the set of first-order formulas using the unary predicates $\mathbf{a}$ ($a \in A$) and the binary predicate $<$, and FO^2 denotes the fragment of FO consisting of formulas which use at most two variable symbols.

If u is a length n ($n > 0$) word over A, we identify the word u with the logical structure $(\{1, \ldots, n\}, (\mathbf{a})_{a \in A})$, where $\mathbf{a}$ denotes the set of a-positions in u. Formulas from FO are naturally interpreted over this structure, and we denote by $L(\varphi)$ the language *defined by* the formula $\varphi \in \mathsf{FO}$, that is, the set of all words which satisfy φ.

1.1. Quantifier-alternation within FO^2.

We now concentrate on FO^2-formulas and we define two important parameters concerning such formulas. To simplify matters, we consider only formulas where negation is used only on atomic formulas so that, in particular, no quantifier is negated. This is naturally possible up to logical equivalence. Now, with each formula $\varphi \in \mathsf{FO}^2$, we associate in the natural way a parsing tree: each occurence of a quantification, $\exists x$ or $\forall x$, yields a unary node, each occurrence of $\vee$ or $\wedge$ yields a binary node, and the leaves are labeled with atomic or negated atomic formulas. The *quantifier depth* of φ is the maximum number of quantifiers along a path in its parsing tree.

With each path from root to leaf in this parsing tree, we also associate its quantifier label, which is the sequence of quantifier node labels ($\exists$ or $\forall$) encountered along this path. A *block* in this quantifier label is a maximal factor consisting only of $\exists$ or only of $\forall$, and we define the *number of blocks* of φ to be the maximum number of blocks in the quantifier label of a path in its parsing tree. Naturally, the quantifier depth of φ is at least equal to its number of blocks.

We let $\mathsf{FO}^2_{m,n}$ ($n \geq m$) denote the set of first-order formulas with quantifier depth at most n and with at most m blocks and let FO^2_m denote the union of the $\mathsf{FO}^2_{m,n}$ for all n. We also denote by $\mathcal{FO}^2$ ($\mathcal{FO}^2_m$, $\mathcal{FO}^2_{m,n}$) the class of FO^2 (FO^2_m, $\mathsf{FO}^2_{m,n}$)-definable languages.

Remark 1.1. Recall that a language is *piecewise testable* if it is a Boolean combination of languages of the form $A^*a_1A^* \cdots a_kA^*$ ($a_i \in A$). It is an elementary observation that the piecewise testable languages coincide with $\mathcal{FO}^2_1$. It is well-known that this class of languages is decidable (see Section 2 below).

1.2. **Rankers.** A *ranker* is a non-empty word on the alphabet $\{\mathsf{X}_a, \mathsf{Y}_a \mid a \in A\}$. Rankers define positions in words: given a word $u \in A^+$ and a letter $a \in A$, we denote by $\mathsf{X}_a(u)$ (resp. $\mathsf{Y}_a(u)$) the least (resp. greatest) integer $1 \leq i \leq |u|$ such that $u[i] = a$. If a does not occur in u, we say that $\mathsf{Y}_a(u)$ and $\mathsf{X}_a(u)$ are not defined. If in addition q is an integer such that $0 \leq q \leq |u|$, we let

$$\mathsf{X}_a(u, q) = q + \mathsf{X}_a(u[q+1; |u|])$$
$$\mathsf{Y}_a(u, q) = \mathsf{Y}_a(u[1; q-1]).$$

These definitions are extended to all rankers: if r' is a ranker, $\mathsf{Z} \in \{\mathsf{X}_a, \mathsf{Y}_a \mid a \in A\}$ and $r = r'\mathsf{Z}$, we let

$$r(u, q) = \mathsf{Z}(u, r'(u, q))$$

if $r'(u, q)$ and $\mathsf{Z}(u, r'(u, q))$ are defined, and we say that $r(u, q)$ is undefined otherwise. In particular: rankers are processed from left to right.

Finally, if r starts with an X- (resp. Y-) letter, we say that r defines the position $r(u) = r(u, 0)$ (resp. $r(u) = r(u, |u| + 1)$), or that it is undefined on u if this position does not exist.

Remark 1.2. Rankers were first introduced, under the name of *turtle programs*, by Schwentick, Thérien and Vollmer [26], as sequences of instructions: go to the next a to the right, go to the next b to the left, etc. These authors write $(\rightarrow, a)$ and $(\leftarrow, a)$ instead of X_a and Y_a. Weis and Immerman [37] write $\triangleright_a$ and $\triangleleft_a$ instead, and they introduced the term *ranker*. We rather follow the notation in [6, 13, 4], where X and Y refer to the future and past operators of PTL.

Example 1.3. The ranker $\mathsf{X}_a\mathsf{Y}_b\mathsf{X}_c$ (go to the first a starting from the left, thence to the first b towards the left, thence to the first c towards the right) is defined on bac and bca, but not on abc or cba.

By $L(r)$ we denote the language of all words on which the ranker r is defined. We say that the words u and v *agree on a class* R of rankers if exactly the same rankers from R are defined on u and v. And we say that two rankers r and s *coincide* on a word u if they are both defined on u and $r(u) = s(u)$.

Example 1.4. If $r = \mathsf{X}_{a_1} \cdots \mathsf{X}_{a_k}$ (resp. $r = \mathsf{Y}_{a_k} \cdots \mathsf{Y}_{a_1}$), then $L(r)$ is the set of words that contain $a_1 \cdots a_k$ as a subword, $L(r) = A^*a_1A^* \cdots a_kA^*$.

The *depth* of a ranker r is defined to be its length as a word. A *block* in r is a maximal factor in $\{\mathsf{X}_a \mid a \in A\}^+$ (an X-block) or in $\{\mathsf{Y}_a \mid a \in A\}^+$ (a Y-block). If $n \geq m$, we denote by $R^{\mathsf{X}}_{m,n}$ (resp. $R^{\mathsf{Y}}_{m,n}$) the set of m-block, depth n rankers, starting with an X- (resp. Y-)

block, and we let $R_{m,n} = R^{\mathsf{X}}_{m,n} \cup R^{\mathsf{Y}}_{m,n}$ and $\underline{R}^{\mathsf{X}}_{m,n} = \bigcup_{m' \leq m, n' \leq n} R^{\mathsf{X}}_{m',n'} \cup \bigcup_{m' < m, n' < n} R^{\mathsf{Y}}_{m',n'}$. We define $\underline{R}^{\mathsf{Y}}_{m,n}$ dually and we let $\underline{R}^{\mathsf{X}}_{m} = \bigcup_{n \geq m} \underline{R}^{\mathsf{X}}_{m,n}$, $\underline{R}^{\mathsf{Y}}_{m} = \bigcup_{n \geq m} \underline{R}^{\mathsf{Y}}_{m,n}$ and $\underline{R}_m = \underline{R}^{\mathsf{X}}_m \cup \underline{R}^{\mathsf{Y}}_m$.

Remark 1.5. Readers familiar with [37] will notice differences between our $\underline{R}^{\mathsf{X}}_{m,n}$ and their analogous $R^{\star}_{m\triangleright,n}$; introduced for technical reasons, it creates no difference between our $\underline{R}_{m,n}$ and their $R^{\star}_{m,n}$, the classes which intervene in crucial Theorem 1.8 below.

1.3. **Rankers and unary temporal logic.** Let us depart for a moment from the consideration of FO^2-formulas, to observe that rankers are naturally suited to describe the different levels of a natural class of temporal logic. The symbols X_a and Y_a $(a \in A)$ can be seen as modal (temporal) operators, with the *future* and *past* semantics respectively. We denote the resulting temporal logic (known as *unary temporal logic*) by TL: its only atomic formula is $\top$, the other formulas are built using Boolean connectives and modal operators. Let $u \in A^+$ and let $0 \leq i \leq |u| + 1$. We say that $\top$ holds at every position i, $(u, i) \models \top$; Boolean connectives are interpreted as usual; and $(u, i) \models \mathsf{X}_a\varphi$ (resp. $\mathsf{Y}_a\varphi$) if and only if $(u, \mathsf{X}_a(u, i)) \models \varphi$ (resp. $(u, \mathsf{Y}_a(u, i)) \models \varphi$). We also say that $u \models \mathsf{X}_a\varphi$ (resp. $\mathsf{Y}_a\varphi$) if $(u, 0) \models \mathsf{X}_a\varphi$ (resp. $(u, 1 + |u|) \models \mathsf{Y}_a\varphi$).

TL is a fragment of *propositional temporal logic* PTL; the latter is expressively equivalent to FO and TL is expressively equivalent to FO^2 [13].

As in the case of FO^2-formulas, one may consider the parsing tree of a TL-formula and define inductively its depth and number of alternations (between past and future operators). If $n \geq m$, the fragment $\mathsf{TL}^{\mathsf{X}}_{m,n}$ (resp. $\mathsf{TL}^{\mathsf{Y}}_{m,n}$) consists of the TL-formulas with depth n and with m alternated blocks, in which every branch (of the parsing tree) with exactly m alternations starts with future (resp. past) operators. Branches with less alternations may start with past (resp. future) operators. The fragments $\mathsf{TL}_{m,n}$, $\underline{\mathsf{TL}}^{\mathsf{X}}_{m,n}$, $\underline{\mathsf{TL}}^{\mathsf{Y}}_{m,n}$, $\underline{\mathsf{TL}}^{\mathsf{X}}_{m}$, $\underline{\mathsf{TL}}^{\mathsf{Y}}_{m}$ and $\underline{\mathsf{TL}}_{m}$ are defined according to the same pattern as in the definition of $R_{m,n}$, $\underline{R}^{\mathsf{X}}_{m,n}$, $\underline{R}^{\mathsf{Y}}_{m,n}$, $\underline{R}^{\mathsf{X}}_{m}$, $\underline{R}^{\mathsf{Y}}_{m}$ and $\underline{R}_m$. We also denote by $\mathcal{TL}^{\mathsf{X}}_{m,n}$ ($\mathcal{TL}^{\mathsf{X}}_{m}$, $\underline{\mathcal{TL}}_{m}$, etc.) the class of $\mathsf{TL}^{\mathsf{X}}_{m,n}$ ($\mathsf{TL}^{\mathsf{X}}_{m}$, TL_m, etc.) -definable languages.

Proposition 1.6. *Let $1 \leq m \leq n$. Two words satisfy the same $\underline{\mathsf{TL}}^{\mathsf{X}}_{m,n}$ formulas if and only if they agree on rankers from $\underline{R}^{\mathsf{X}}_{m,n}$. A language is in $\underline{\mathcal{TL}}^{\mathsf{X}}_{m,n}$ if and only if it is a Boolean combination of languages of the form $L(r)$, $r \in \underline{R}^{\mathsf{X}}_{m,n}$.*

Similar statements hold for $\underline{\mathsf{TL}}^{\mathsf{Y}}_{m,n}$, $\underline{\mathsf{TL}}^{\mathsf{X}}_{m}$, $\underline{\mathsf{TL}}^{\mathsf{Y}}_{m}$ and $\underline{\mathsf{TL}}_{m}$, relative to the corresponding classes of rankers.

Proof. Since every ranker can be viewed as a TL-formula, it is easily verified that if u and v satisfy the same $\underline{\mathsf{TL}}^{\mathsf{X}}_{m,n}$-formulas, then they agree on rankers from $\underline{R}^{\mathsf{X}}_{m,n}$. To prove the converse, it suffices to show that a $\underline{\mathsf{TL}}^{\mathsf{X}}_{m,n}$-formula is equivalent to a Boolean combination of formulas that are expressed by a single ranker. That is: we only need to show that modalities can be brought outside the formula. This follows from the following elementary logical equivalences:

$$\begin{aligned} \mathsf{X}_a(\varphi \wedge \psi) &\equiv \mathsf{X}_a\varphi \wedge \mathsf{X}_a\psi, \\ \mathsf{X}_a(\varphi \vee \psi) &\equiv \mathsf{X}_a\varphi \vee \mathsf{X}_a\psi, \\ \mathsf{X}_a(\neg\varphi) &\equiv \mathsf{X}_a\top \wedge \neg\mathsf{X}_a\varphi. \end{aligned}$$

□

Remark 1.7. Together with Example 1.4, this proposition confirms the elementary observation that a language is $\underline{\mathsf{TL}}_1$ (resp. $\underline{\mathsf{TL}}_1^{\mathsf{X}}$, $\underline{\mathsf{TL}}_1^{\mathsf{Y}}$) definable if and only if it is piecewise testable (see Remark 1.1). It follows that $\underline{\mathsf{TL}}_1$-definability is decidable.

1.4. **Rankers and FO^2.** The connection established by Weis and Immerman [37, Theorem 4.5] between rankers and formulas in $\mathsf{FO}^2_{m,n}$, Theorem 1.8 below, is much deeper. If x, y are integers, we let $\mathrm{ord}(x, y)$, the *order type* of x and y, be one of the symbols $<$, $>$ or $=$, depending on whether $x < y$, $x > y$ or $x = y$.

Theorem 1.8 (Weis and Immerman [37]). *Let $u, v \in A^*$ and let $1 \leq m \leq n$. Then u and v satisfy the same formulas in $\mathsf{FO}^2_{m,n}$ if and only if*

- **(WI *1)*** *u and v agree on rankers from $\underline{R}_{m,n}$,*
- **(WI *2)*** *if the rankers $r \in \underline{R}_{m,n}$ and $r' \in \underline{R}_{m-1,n-1}$ are defined on u and v, then* $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$.
- **(WI *3)*** *if $r \in \underline{R}_{m,n}$ and $r' \in \underline{R}_{m,n-1}$ are defined on u and v and end with different direction letters, then* $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$.

Corollary 1.9. *For each $n \geq m \geq 1$, $\underline{\mathcal{TL}}_{m,n} \subseteq \mathcal{FO}^2_{m,n}$ and $\underline{\mathcal{TL}}_m \subseteq \mathcal{FO}^2_m$.*

Proof. Let L be a $\underline{\mathsf{TL}}_{m,n}$-definable language. For each $u \in L$, let φ_u be the conjunction of the $\mathsf{FO}^2_{m,n}$-sentences satisfied by u and let φ be the disjunction of the formulas φ_u ($u \in L$). Since $\mathsf{FO}^2_{m,n}$ is finite (up to logical equivalence), the conjunctions and disjunctions in the definition of φ are all finite. We show that $L = L(\varphi)$.

A word v satisfies φ if and only if it satisfies φ_u for some word $u \in L$. Then v satisfies the same $\mathsf{FO}^2_{m,n}$-sentences as u and, by comparing the statements in Proposition 1.6 and Theorem 1.8, we see that u and v satisfy the same $\underline{\mathsf{TL}}_{m,n}$-formulas. Since L is defined by such a formula, it follows that $v \in L$. Conversely, every word $v \in L$ satisfies φ since it satisfies φ_v, which is logically equivalent to a term in the disjunction defining φ. This concludes the proof. □

2. On varieties and pseudovarieties

Recent results show that the FO^2_m-definability of a language L can be characterized algebraically, that is, in terms depending only on the syntactic monoid of L. This justifies exploring the algebraic path to tackle the decidability of this definability problem. Eilenberg's theory of varieties provides the mathematical framework. In this section, we summarize the information on monoid and variety theory that will be relevant for our purpose. For more detailed information and proofs, we refer the reader to [20, 2, 31, 32, 30], among other sources.

A *semigroup* is a set equipped with a binary associative operation. A *monoid* is a semigroup which contains a unit element. The set A^* of all words on alphabet A, equipped with the concatenation product, is the *free monoid* on A: it has the specific property that, if $\varphi\colon A \to M$ is a map into a monoid, then there exists a unique monoid morphism $\psi\colon A^* \to M$ which extends φ. Apart from free monoids, the semigroups and monoids which we will consider in this paper are finite.

If A is a finite alphabet and M is a finite monoid, we say that a language $L \subseteq A^*$ is *recognized* by M if there exists a morphism $\varphi\colon A^* \to M$ such that $L = \varphi^{-1}(\varphi(L))$.

Example 2.1. If $u \in A^*$ and $B \subseteq A$, let

$$\mathsf{alph}(u) = \{a \in A \mid u = vaw \text{ for some } v, w \in A^*\},$$
$$[B] = \{u \in A^* \mid \mathsf{alph}(u) = B\}$$

Let φ be the following morphism from A^* into the direct product of $|A|$ copies of the 2-element monoid $\{1, 0\}$ (multiplicative): for each letter $a \in A$, $\varphi(a)$ is the A-tuple in which every component is 1, except for the a-component. It is elementary to show that $[B] = \varphi^{-1}(\varphi([B]))$ and hence, $[B]$ is accepted by a monoid that is *idempotent* (every element is equal to its own square) and commutative. Conversely, one can show that every language recognized by an idempotent and commutative monoid is a Boolean combination of languages of the form $[B]$ $(B \subseteq A)$.

A *pseudovariety* of monoids is a class of finite monoids which is closed under taking direct products, homomorphic images and submonoids. A class of languages $\mathcal{V}$ is a collection $\mathcal{V} = (\mathcal{V}(A))_A$, indexed by all finite alphabets A, such that $\mathcal{V}(A)$ is a set of languages in A^*. If $\mathbf{V}$ is a pseudovariety of monoids, we let $\mathcal{V}(A)$ be the set of all languages of A^* which are recognized by a monoid in $\mathbf{V}$. The class $\mathcal{V}$ has important closure properties: each $\mathcal{V}(A)$ is closed under Boolean operations and under taking residuals (if $L \in \mathcal{V}(A)$ and $u \in A^*$, then Lu^{-1} and $u^{-1}L$ are in $\mathcal{V}(A)$); and if $\varphi\colon A^* \to B^*$ is a morphism and $L \in \mathcal{V}(B)$, then $\varphi^{-1}(L) \in \mathcal{V}(A)$. Classes of recognizable languages with these properties are called *varieties* of languages, and Eilenberg's theorem (see [20]) states that the correspondence $\mathbf{V} \mapsto \mathcal{V}$, from pseudovarieties of monoids to varieties of languages, is one-to-one and onto. Moreover, the decidability of membership in the pseudovariety $\mathbf{V}$, implies the decidability of the variety $\mathcal{V}$: indeed, a language is in $\mathcal{V}$ if and only if its (effectively computable) syntactic monoid is in $\mathbf{V}$.

For every finite semigroup S, there exists an integer, usually denoted ω, such that every element of the form s^ω in S is idempotent. The *Green relations* are another important concept to describe semigroups and monoids: if S is a semigroup and $s, t \in S$, we say that $s \leq_{\mathcal{J}} t$ (resp. $s \leq_{\mathcal{R}} t$, $s \leq_{\mathcal{L}} t$) if $s = utv$ (resp. $s = tv$, $s = ut$) for some $u, v \in S \cup \{1\}$. We also say that $s \mathrel{\mathcal{J}} t$ is $s \leq_{\mathcal{J}} t$ and $t \leq_{\mathcal{J}} s$. The relations $\mathcal{R}$ and $\mathcal{L}$ are defined similarly.

Pseudovarieties that will be important in this paper are the following.

- $\mathbf{J_1}$, the pseudovariety of idempotent and commutative monoids; as discussed in Example 2.1, the corresponding variety of languages consists of the Boolean combinations of languages of the form $[B]$.

- $\mathbf{R}$, $\mathbf{L}$ and $\mathbf{J}$, the pseudovarieties of $\mathcal{R}$-, $\mathcal{L}$- and $\mathcal{J}$-trivial monoids; a monoid is, say, $\mathbf{R}$-trivial if each of its $\mathcal{R}$-classes is a singleton. The variety of languages corresponding to $\mathbf{J}$ was described by Simon (see [20]): it is exactly the class of piecewise testable languages, i.e., the class of FO^2_1-definable languages, see Remarks 1.1 and 1.7.

- $\mathbf{A}$, the variety of *aperiodic* monoids, i.e., monoids in which $x^\omega = x^{\omega+1}$ holds for each x. Celebrated theorems of Schützenberger, McNaughton and Papert and Kamp show that the corresponding variety of languages consists of the star-free languages, the languages that are definable in FO, and the languages definable in propositional temporal logic, see for instance [20, 32, 5, 29, 30].

- $\mathbf{DA}$ is the pseudovariety of all monoids in which $(xy)^\omega x (xy)^\omega = (xy)^\omega$ for all x, y. This pseudovariety has many characterizations in combinatorial, algebraic and logical terms. Of particular interest to us is the fact that the corresponding variety of languages consists of

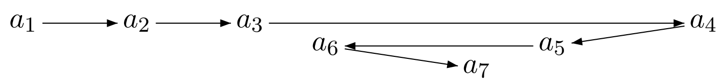

Figure 1: The positions defined by r in u, when $r = \mathsf{X}_{a_1}\mathsf{X}_{a_2}\mathsf{X}_{a_3}\mathsf{X}_{a_4}\mathsf{Y}_{a_5}\mathsf{Y}_{a_6}\mathsf{X}_{a_7}$ is condensed on u

the languages that are definable in FO^2, and equivalently, of the languages that are defined in unary temporal logic, see [31, 32, 6, 13, 35] among others.

- Straubing showed that, for each $m \geq 1$, $\mathcal{FO}^2_m$ is a variety of languages, and he described the corresponding pseudovariety of monoids, which we write $\mathbf{FO}^2_m$, in terms of iterated block products [28]. We will not need to discuss the definition of the block product here, retaining only that this characterization does not imply decidability, and that Straubing gave identities (using products and ω-powers like the identities given above for $\mathbf{A}$ and $\mathbf{DA}$) which he conjectures define each $\mathbf{FO}^2_m$. Establishing this conjecture would prove the decidability of FO^2_m-definability.

- Kufleitner and Lauser also showed that, for each $n \geq m \geq 1$, $\mathcal{FO}^2_{m,n}$ and $\mathcal{FO}^2_m$ form varieties of languages, using a general result on *logical fragments* [16, Cor. 3.4]. Their result also does not imply a decidability statement.
- On a given monoid M, we define the congruences $\sim_\mathbf{K}$ and $\sim_\mathbf{D}$ as follows.
  - $u \sim_\mathbf{K} v$ if and only if, for each idempotent e in M, we have either $eu, ev <_\mathcal{J} e$ or $eu = ev$,
  - $u \sim_\mathbf{D} v$ if and only if, for each idempotent e in M, we have either have $ue, ve <_\mathcal{J} e$ or $ue = ve$.

If $\mathbf{V}$ is a pseudovariety of monoids, we say that the monoid $M \in \mathbf{K} ⓜ \mathbf{V}$ if $M/{\sim_\mathbf{K}} \in \mathbf{V}$, and $M \in \mathbf{D} ⓜ \mathbf{V}$ if $M/{\sim_\mathbf{D}} \in \mathbf{V}$. The classes $\mathbf{K} ⓜ \mathbf{V}$ and $\mathbf{D} ⓜ \mathbf{V}$ are pseudovarieties as well, which are usually defined in terms of Mal'cev products with the pseudovarieties $\mathbf{K}$ and $\mathbf{D}$, see [23, Thm 4.6.50] or [12, 10].

The following equalities are well-known [20]:

$$\mathbf{K} ⓜ \mathbf{J_1} = \mathbf{K} ⓜ \mathbf{J} = \mathbf{R}, \;\; \mathbf{D} ⓜ \mathbf{J_1} = \mathbf{D} ⓜ \mathbf{J} = \mathbf{L}.$$

3. Condensed rankers

Our main tool to approach the decidability of FO^2_m-definability lies in the notion of condensed rankers, a variant of rankers which was introduced implicitly by Weis and Immerman to prove Theorem 6.4 below (see [37, Theorem 4.7]). Recall that a ranker can be seen as a sequence of directional instructions (see Example 1.3). We say that a ranker r is *condensed on u* if it is defined on u, and if the sequence of positions visited *zooms in* on $r(u)$, never crossing over a position already visited, see Figure 1. Formally, $r = \mathsf{Z}_1 \cdots \mathsf{Z}_n$ is condensed on u if there exists a chain of open intervals

$$(0; |u| + 1) = (i_0; j_0) \supset (i_1; j_1) \supset \cdots \supset (i_{n-1}; j_{n-1}) \ni r(u)$$

such that for all $1 \leq \ell \leq n-1$ the following properties are satisfied:

- If $\mathsf{Z}_\ell \mathsf{Z}_{\ell+1} = \mathsf{X}_a \mathsf{X}_b$ then $(i_\ell, j_\ell) = (\mathsf{X}_a(u, i_{\ell-1}), j_{\ell-1})$.

- If $\mathsf{Z}_\ell\mathsf{Z}_{\ell+1} = \mathsf{Y}_a\mathsf{Y}_b$ then $(i_\ell, j_\ell) = (i_{\ell-1}, \mathsf{Y}_a(u, j_{\ell-1})$.
- If $\mathsf{Z}_\ell\mathsf{Z}_{\ell+1} = \mathsf{X}_a\mathsf{Y}_b$ then $(i_\ell, j_\ell) = (i_{\ell-1}, \mathsf{X}_a(u, i_{\ell-1}))$.
- If $\mathsf{Z}_\ell\mathsf{Z}_{\ell+1} = \mathsf{Y}_a\mathsf{X}_b$ then $(i_\ell, j_\ell) = (\mathsf{Y}_a(u, j_{\ell-1}), j_{\ell-1})$.

Remark 3.1. The i_ℓ and j_ℓ are either 0 or $1 + |u|$, or positions of the form $r'(u)$ for some prefix of r' of r. More precisely, if r_ℓ is the depth ℓ prefix of r ($\ell < n$), then $r_\ell(u) = i_\ell$ if $\mathsf{Z}_{\ell+1}$ is of the form X_a, and $r_\ell(u) = j_\ell$ if $\mathsf{Z}_{\ell+1}$ is of the form Y_a.

Remark 3.2. If $r = r_1r_2$ is condensed on u, then $r(u) > r_1(u)$ if r_2 starts with an X-letter, and $r(u) < r_1(u)$ if r_2 starts with a Y-letter.

Example 3.3. The ranker $\mathsf{X}_a\mathsf{Y}_b\mathsf{X}_c$ is defined on the words bac and bca, but it is condensed only on bca.

Rankers in $\underline{R}_{1,n}$ and rankers of the form $\mathsf{X}_a\mathsf{Y}_{b_1}\cdots\mathsf{Y}_{b_k}$ or $\mathsf{Y}_a\mathsf{X}_{b_1}\cdots\mathsf{X}_{b_k}$ are condensed on all words on which they are defined.

Condensed rankers form a natural notion, which is equally well-suited to the task of describing FO^2_m-definability (see Theorem 3.11 below). With respect to TL, for which Proposition 1.6 shows a perfect match with the notion of rankers, they can be interpreted as adding a strong notion of unambiguity, see Section 6.3 below and the work of Lodaya, Pandya and Shah [17].

Let us say that two words u and v *agree on condensed rankers from a set* R of rankers, if the same rankers in R are condensed on u and v. We write $u \rhd_{m,n} v$ (resp. $u \lhd_{m,n} v$) if u and v agree on condensed rankers in $\underline{R}^{\mathsf{X}}_{m,n}$ (resp. $\underline{R}^{\mathsf{Y}}_{m,n}$).

If r is a ranker, let $L_c(r)$ be the language of all words on which r is condensed. We define $\mathcal{R}_m$ (resp. $\mathcal{L}_m$) to be the Boolean algebra generated by the languages of the form $L_c(r)$, $r \in \underline{R}^{\mathsf{X}}_{m,n}$ (resp. $\underline{R}^{\mathsf{Y}}_{m,n}$), $n \geq m$.

3.1. **Technical properties of condensed rankers.** A factorization $u = u_-au_+$ of a word $u \in A^*$ is called the *a-left factorization* of u if $a \notin \mathsf{alph}(u_-)$. Symmetrically, $u = u_-au_+$ is the *a-right factorization* of u if $a \notin \mathsf{alph}(u_+)$. Thus, the a-left (resp. a-right) factorization of a identifies the first occurrence of a when reading u from the left (resp. the right).

Lemmas 3.4 and 3.5 admit an elementary verification.

Lemma 3.4. *Let s be a ranker, $a \in A$ and $r = \mathsf{X}_a s$. Let also $u \in A^+$ and let $u = u_-au_+$ be its a-left factorization. Then r is condensed on u if and only if*

- *s is condensed on u_+ if s starts with an X-block;*
- *s is condensed on u_- if s starts with a Y-block.*

A dual statement holds if r is of the form $r = \mathsf{Y}_a s$, with respect to the a-right factorization of u.

Lemma 3.5. *Let r be a ranker and $a \in A$. Let also $u \in A^+$ and let $u = u_-au_+$ be its a-left factorization.*

If r starts with an X-letter, then

- *r is defined on u_- if and only if r is defined on u, r does not contain X_a or Y_a and, for every prefix p of r ending with an X-letter, $p\mathsf{Y}_a$ is not defined on u.*
- *r is condensed on u_-*
  - *if and only if r is defined on u_- and condensed on u,*
  - *if and only if r is condensed on u, r does not contain X_a or Y_a and, if p is the maximal prefix of r consisting only of X-letters, then $p\mathsf{Y}_a$ is not defined on u,*

  – *if and only if* r *is condensed on* u, r *does not contain* X_a *or* Y_a *and, if* $p = \mathsf{X}_{b_1} \cdots \mathsf{X}_{b_k}$ $(k \geq 1)$ *is the initial* X*-block of* r, *then* $\mathsf{X}_a \mathsf{Y}_{b_k} \cdots \mathsf{Y}_{b_1}$ *is defined on* u.
- r *is defined on* u_+ *if and only if* $\mathsf{X}_a r$ *is defined on* u *and, for every prefix* p *of* r *ending with a* Y*-letter,* $\mathsf{X}_a p \mathsf{Y}_a$ *is defined on* u.
- r *is condensed on* u_+ *if and only if* $\mathsf{X}_a r$ *is condensed on* u.

If r *starts with a* Y*-letter, then*

- r *is defined on* u_- *if and only if* $\mathsf{X}_a r$ *is defined on* u, r *does not contain* X_a *or* Y_a *and, for every prefix* p *of* r *ending with an* X*-letter,* $\mathsf{X}_a p \mathsf{Y}_a$ *is not defined on* u.
- r *is condensed on* u_- *if and only if* $\mathsf{X}_a r$ *is condensed on* u.
- r *is defined on* u_+ *if and only if* r *is defined on* u *and, for every prefix* p *of* r *ending with a* Y*-letter,* $p\mathsf{Y}_a$ *is defined on* u.
- r *is condensed on* u_+
  – *if and only if* r *is defined on* u_+ *and condensed on* u,
  – *if and only if* r *is condensed on* u *and, if* $p = \mathsf{Y}_{b_1} \cdots \mathsf{Y}_{b_k}$ $(k \geq 1)$ *is the initial* Y*-block of* r, *then* $p\mathsf{Y}_a$ *is defined on* u.

We also note the following, very useful characterization of the relations $\triangleright_{m,n}$ and $\triangleleft_{m,n}$.

Proposition 3.6. *The families of relations* $\triangleright_{m,n}$ *and* $\triangleleft_{m,n}$ $(n \geq m \geq 1)$ *are uniquely determined by the following properties.*

(1) $u \triangleright_{1,n} v$ *if and only if* $u \triangleleft_{1,n} v$, *if and only if* u *and* v *have the same subwords of length at most* n.
(2) *If* $m \geq 2$, *then* $u \triangleright_{m,n} v$ *if and only if* $\mathsf{alph}(u) = \mathsf{alph}(v)$, $u \triangleleft_{m-1,n-1} v$ *and for each letter* $a \in \mathsf{alph}(u)$, *the* a*-left factorizations* $u = u_- a u_+$ *and* $v = v_- a v_+$ *satisfy* $u_- \triangleleft_{m-1,n-1} v_-$ *and* $u_+ \triangleright_{m,n-1} v_+$ $(u_+ \triangleright_{m-1,n-1} v_+$ *if* $n = m)$.
(3) *If* $m \geq 2$, *then* $u \triangleleft_{m,n} v$ *if and only if* $\mathsf{alph}(u) = \mathsf{alph}(v)$, $u \triangleright_{m-1,n-1} v$ *and for each letter* $a \in \mathsf{alph}(u)$, *the* a*-right factorizations* $u = u_- a u_+$ *and* $v = v_- a v_+$ *satisfy* $u_+ \triangleright_{m-1,n-1} v_+$ *and* $u_- \triangleleft_{m,n-1} v_-$ $(u_- \triangleleft_{m-1,n-1} v_-$ *if* $n = m)$.

Proof. Statement (1) follows directly from Examples 1.4 and 3.3. Let us now assume that $m \geq 2$.

Suppose that $\mathsf{alph}(u) = \mathsf{alph}(v)$, $u \triangleleft_{m-1,n-1} v$ and for each $a \in \mathsf{alph}(u)$, the a-left factorizations $u = u_- a u_+$ and $v = v_- a v_+$ satisfy $u_- \triangleleft_{m-1,n-1} v_-$ and $u_+ \triangleright_{m,n-1} v_+$ if $n > m$ ($u_+ \triangleright_{m-1,n-1} v_+$ if $n = m$). Let $r \in \underline{R}^{\mathsf{X}}_{m,n}$ be condensed on u. If r starts with a Y-letter, then $r \in \underline{R}^{\mathsf{Y}}_{m-1,n-1}$, and hence r is condensed on v since $u \triangleleft_{m-1,n-1} v$. If instead r starts with an X-letter, say $r = \mathsf{X}_a s$, we consider the a-left factorizations of u and v. If s starts with a Y-letter, then $s \in \underline{R}^{\mathsf{Y}}_{m-1,n-1}$, s is condensed on u_- (Lemma 3.4) and hence s is condensed on v_- since $u_- \triangleleft_{m-1,n-1} v_-$, from which it follows again that r is condensed on v. Finally, if s starts with an X-letter, then s is condensed on u_+ by Lemma 3.4. Moreover, $s \in \underline{R}^{\mathsf{X}}_{m,n-1}$ if $n > m$. If $n = m$, we have in fact $r \in \underline{R}^{\mathsf{X}}_{m-1,n}$ (since r starts with two X-letters) and hence $s \in \underline{R}^{\mathsf{X}}_{m-1,n-1}$. Since $u_+ \triangleright_{m,n-1} v_+$ if $n > m$ and $u_+ \triangleright_{m-1,n-1} v_+$ if $n = m$, it follows that s is condensed on v_+, and hence r is condensed on v.

Conversely, let us assume that $u \triangleright_{m,n} v$, that is, u and v agree on condensed rankers in $\underline{R}^{\mathsf{X}}_{m,n}$. Considering rankers in $R^{\mathsf{X}}_{1,1} \subseteq \underline{R}^{\mathsf{X}}_{m,n}$ shows that $\mathsf{alph}(u) = \mathsf{alph}(v)$. Similarly, considering rankers in $\underline{R}^{\mathsf{Y}}_{m-1,n-1} \subseteq \underline{R}^{\mathsf{X}}_{m,n}$ shows that $u \triangleleft_{m-1,n-1} v$. Finally, let $a \in \mathsf{alph}(u)$ and let $u = u_- a u_+$ and $v = v_- a v_+$ be a-left factorizations.

Let $s \in \underline{R}^{\mathsf{Y}}_{m-1,n-1}$ be condensed on u_-. Note that s contains neither X_a nor Y_a, since $a \notin \mathsf{alph}(u_-)$. If s starts with a Y-letter, then $r = \mathsf{X}_a s$ is condensed on u (Lemma 3.4) and since $r \in \underline{R}^{\mathsf{X}}_{m,n}$, r is condensed on v as well, which implies that s is condensed on v_-. If instead s starts with an X-letter, then s is condensed on u and hence on v. Moreover, if $p = \mathsf{X}_{b_1} \cdots \mathsf{X}_{b_k}$ is the maximal prefix of s consisting only of X-letters, then $\mathsf{X}_a \mathsf{Y}_{b_k} \cdots \mathsf{Y}_{b_1} \in \underline{R}^{\mathsf{X}}_{2,n}$ is condensed on u (Lemma 3.5). Since $\underline{R}^{\mathsf{X}}_{2,n} \subseteq \underline{R}^{\mathsf{X}}_{m,n}$, it is condensed on v as well and hence, s is condensed on v_-.

Finally, assume that $s \in \underline{R}^{\mathsf{X}}_{m,n-1}$ ($\underline{R}^{\mathsf{X}}_{m-1,n-1}$ if $n = m$) is condensed on u_+. The reasoning is similar: if s starts with an X-letter, then $\mathsf{X}_a s \in \underline{R}^{\mathsf{X}}_{m,n}$ is condensed on u. Therefore $\mathsf{X}_a s$ is condensed on v and s is condensed on v_+. If instead s starts with a Y-letter, then s is condensed on u and $s \in \underline{R}^{\mathsf{Y}}_{m-1,n-2}$ ($\underline{R}^{\mathsf{X}}_{m-2,n-2}$ if $n = m$). In particular $s \in \underline{R}^{\mathsf{X}}_{m,n}$ and hence, s is condensed on v as well. Moreover, if p is the initial Y-block of s, then $p\mathsf{Y}_a$ is condensed on u. Note that $p\mathsf{Y}_a \in \underline{R}^{\mathsf{Y}}_{1,n-1} \subseteq \underline{R}^{\mathsf{X}}_{m,n}$, so $p\mathsf{Y}_a$ is condensed on v and s is condensed on v_+. □

Lemma 3.7. *Let $n \geq m \geq 2$, $u, v \in A^*$, $a \in A$ and let $u = u_- a u_+$ and $v = v_- a v_+$ be a-left factorizations. If $u \rhd_{m,n} v$, then $u_- \rhd_{m,n-1} v_-$ ($u_- \rhd_{m-1,n-1} v_-$ if $n = m$). And if $u \lhd_{m,n} v$, then $u_+ \lhd_{m,n-1} v_+$ ($u_+ \lhd_{m-1,n-1} v_+$ if $n = m$). Dual statements hold for the factors of the a-right factorizations of u and v if $u \lhd_{m,n} v$ or $u \rhd_{m,n} v$.*

Proof. We give the proof if $n > m$; it is easily adapted to the case where $n = m$.

Assume that $u \rhd_{m,n} v$ and $r \in \underline{R}^{\mathsf{X}}_{m,n-1}$ is condensed on u_-. By Lemma 3.5, we have:

- If r starts with an X-letter, then r is condensed on u, r does not contain occurrences of X_a or Y_a, and if $p = \mathsf{X}_{b_1} \cdots \mathsf{X}_{b_k}$ is the initial X-block of r, then $q = \mathsf{X}_a \mathsf{Y}_{b_k} \cdots \mathsf{Y}_{b_1}$ is condensed on u. Since $q \in R^{\mathsf{X}}_{2,k+1}$ and $k < n$, we have $q \in \underline{R}^{\mathsf{X}}_{m,n}$ and hence r and q are condensed on v. Therefore r is condensed on v_-.

- If r starts with a Y-letter, then $\mathsf{X}_a r$ is condensed on u. But $r \in \underline{R}^{\mathsf{Y}}_{m-1,n-2}$, so $\mathsf{X}_a r \in \underline{R}^{\mathsf{X}}_{m,n}$ and hence $\mathsf{X}_a r$ is condensed on v. It follows that r is condensed on v_-.

Assume now that $u \lhd_{m,n} v$ and $r \in \underline{R}^{\mathsf{Y}}_{m,n-1}$ is condensed on u_+. Then

- If r starts with an X-letter (which is possible only if $m \geq 2$), then $r \in \underline{R}^{\mathsf{X}}_{m-1,n-2}$ and $\mathsf{X}_a r$ is condensed on u. But $\mathsf{X}_a r \in \underline{R}^{\mathsf{X}}_{m-1,n-1} \subseteq \underline{R}^{\mathsf{Y}}_{m,n}$, so $\mathsf{X}_a r$ is condensed on v and r is condensed on v_+.

- If instead r starts with a Y-letter, then r is condensed on u and if p is the initial Y-block of r, then $p\mathsf{Y}_a$ is condensed on u. But $r, p\mathsf{Y}_a \in \underline{R}^{\mathsf{Y}}_{m,n}$, so r and $p\mathsf{Y}_a$ are condensed on v, and r is condensed on v_+. □

3.2. **Condensed rankers, rankers and FO^2.** We now show that, in the characterization of $\mathcal{FO}^2_{m,n}$ in Theorem 1.8, condensed rankers can be used just as well. This is done in Theorem 3.11. The first step is to relate agreement on rankers and agreement on condensed rankers. We start with a technical lemma.

Lemma 3.8. *If a ranker $r \in \underline{R}^{\mathsf{Z}}_{m,n}$ ($\mathsf{Z} \in \{\mathsf{X}, \mathsf{Y}\}$) is defined but not condensed on u, and if s is the maximal prefix of r which is condensed on u, then one of the following holds, for some $\ell \geq 1$:*

- *$r = s\mathsf{X}_b t$, $s = s_0 \mathsf{Y}_a \mathsf{X}_{b_1} \cdots \mathsf{X}_{b_{\ell-1}}$ and $s_0\mathsf{Y}_a(u) \leq s(u) < s_0(u) \leq s\mathsf{X}_b(u)$;*
- *$r = s\mathsf{Y}_b t$, $s = s_0 \mathsf{X}_a \mathsf{Y}_{b_1} \cdots \mathsf{Y}_{b_{\ell-1}}$ and $s_0\mathsf{X}_a(u) \geq s(u) > s_0(u) > s\mathsf{Y}_b(u) = s_0\mathsf{Y}_b(u)$.*

Moreover s_0 is not empty, $s_0 \in \underline{R}^{\mathsf{Z}}_{m-1,n-\ell}$; $s\mathsf{X}_b(u) = s_0(u)$ (resp. $s\mathsf{Y}_b(u) = s_0(u)$) if the last letter of s_0 is in $\{\mathsf{X}_b, \mathsf{Y}_b\}$; and $s\mathsf{X}_b(u) = s_0\mathsf{X}_b(u)$ (resp. $s\mathsf{Y}_b(u) = s_0\mathsf{Y}_b(u)$) otherwise.

Proof. Rankers in $R_{1,n}$ are condensed on each word on which they are defined (Example 3.3). Therefore we have $m \geq 2$.

By hypothesis, $s \neq r$. We consider the case where the first letter after s is an X-letter, the other case is dual. Then r is of the form $r = s\mathsf{X}_b t$, where t may be empty. In view of Example 3.3, $s = s_0 \mathsf{Y}_a \mathsf{X}_{b_1} \cdots \mathsf{X}_{b_{\ell-1}}$ for some non-empty s_0 and $\ell \geq 1$. Since s is condensed on u but $s\mathsf{X}_b$ is not, we have the following (see Remark 3.2):

$$s_0\mathsf{Y}_a(u) < s_0\mathsf{Y}_a\mathsf{X}_{b_1}(u) \cdots < s_0\mathsf{Y}_a\mathsf{X}_{b_1}\cdots\mathsf{X}_{b_{\ell-1}}(u) = s(u) < s_0(u),$$

and $s\mathsf{X}_b(u) \geq s_0(u)$. More precisely, $s\mathsf{X}_b(u)$ is the first b-position to the right of $s(u)$, so $s\mathsf{X}_b(u) = s_0(u)$ if $s_0(u)$ is a b-position (*i.e.*, if s_0 ends with X_b or Y_b), and $s\mathsf{X}_b(u) = s_0\mathsf{X}_b(u)$ otherwise. ☐

Proposition 3.9. *Let $n \geq m \geq 1$, $u, v \in A^+$ and $\mathsf{Z} \in \{\mathsf{X}, \mathsf{Y}\}$. If u and v agree on condensed rankers in $\underline{R}^{\mathsf{Z}}_{m,n}$ and if $r \in \underline{R}^{\mathsf{Z}}_{m,n}$ is defined on both u and v, then there exists $r' \in \underline{R}^{\mathsf{Z}}_{m,n}$ which is condensed on u and v and coincides with r on both words.*

Proof. The result is trivial if $m = 1$, since rankers in $R_{1,n}$ are condensed on each word on which they are defined (Example 3.3). We now assume that $m \geq 2$.

Let p and q be positions in u and v and let $r \in \underline{R}^{\mathsf{Z}}_{m,n}$ such that $r(u) = p$ and $r(v) = q$. If r is not condensed on u, then r is not condensed on v (since the two words agree on condensed rankers). With the notation of Lemma 3.8, r coincides on both u and v with $r' = s_0 t$, $s_0\mathsf{X}_b t$ or $s_0\mathsf{Y}_b t$ (depending on the last letter of s_0 and on the letter following s in r), which starts with the same letter as r. If r' is not condensed on u and v, we repeat the reasoning. This process must terminate since each iteration reduces the depth of r'. ☐

Proposition 3.10. *Let $n \geq m \geq 1$, $u, v \in A^+$ and $\mathsf{Z} \in \{\mathsf{X}, \mathsf{Y}\}$. If u and v agree on condensed rankers in $\underline{R}^{\mathsf{Z}}_{m,n}$, then they agree on rankers from the same class.*

Proof. If u and v do not agree on rankers from $\underline{R}^{\mathsf{Z}}_{m,n}$, let $r \in \underline{R}^{\mathsf{Z}}_{m,n}$ be a minimum depth ranker on which u and v disagree. Without loss of generality, we may assume that $u \in L(r)$ and $v \notin L(r)$. In particular, r is not condensed on u.

Let s, s_0 and t be as in Lemma 3.8. Without loss of generality again, we may assume that the letter following s in r is X_b. Since s is condensed on u and $s\mathsf{X}_b$ is not, the ranker s is condensed on v and $s\mathsf{X}_b$ is not. Moreover, $s\mathsf{X}_b$ coincides on u with $s' = s_0$, or $s_0\mathsf{X}_b$, depending on the last letter of s_0. Observe that s' is shorter than r, so s' is defined on v. In particular, there exists a b-position in v to the right of s, which is not to the left of s_0 (since $s\mathsf{X}_b$ is not condensed on v). It follows that $s\mathsf{X}_b(v) = s'(v)$. Let now $r' = s't$: then r' is shorter than r, it coincides with r on u, and it is not defined on v since s' coincides with s on that word. This contradicts the minimality of r. ☐

We can now prove the following variant of Theorem 1.8.

Theorem 3.11. *Let $u, v \in A^*$ and let $1 \leq m \leq n$. Then u and v satisfy the same formulas in $\mathsf{FO}^2_{m,n}$ if and only if*

(WI 1c) *u and v agree on condensed rankers from $\underline{R}_{m,n}$,*

(WI 2c) *if the rankers $r \in \underline{R}_{m,n}$ and $r' \in \underline{R}_{m-1,n-1}$ are condensed on u and v, then* $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$.

(WI 3c) *if $r \in \underline{R}_{m,n}$ and $r' \in \underline{R}_{m,n-1}$ are condensed on u and v and end with different direction letters, then* $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$.

Proof. We need to prove that together, Properties (**WI 1**), (**WI 2**) and (**WI 3**) are equivalent to Properties (**WI 1c**), (**WI 2c**) and (**WI 3c**).

Let us first assume that (**WI 1**), (**WI 2**) and (**WI 3**) hold. It is immediate that (**WI 2c**) and (**WI 3c**) hold. If (**WI 1c**) does not hold, let r be a ranker in $\underline{R}_{m,n}$ which is condensed on v and not on u. Since (**WI 1**) holds, r is defined on u. Let s_0, s and t be as in Lemma 3.8 and let us assume, without loss of generality, that the letter following s in r is X_b. Then s_0 and $s\mathsf{X}_b$ are defined on both u and v, with $s_0 \in \underline{R}_{m-1,n-1}$ and $s\mathsf{X}_b \in \underline{R}_{m,n}$. Since r is condensed on v, we have $s\mathsf{X}_b(v) < s_0(v)$, and since $s\mathsf{X}_b$ is not condensed on u, we have $s_0(v) \leq s\mathsf{X}_b(v)$, contradicting Property (**WI 2**). Thus (**WI 1c**) holds.

Conversely, let us assume that (**WI 1c**), (**WI 2c**) and (**WI 3c**) hold. Then (**WI 1**) holds by Proposition 3.10. Let us verify Property (**WI 2**): suppose that $r \in \underline{R}_{m,n}$ and $r' \in \underline{R}_{m-1,n-1}$ are defined on u and v. In view of (**WI 1c**), Proposition 3.9 shows that there exist rankers $s \in \underline{R}_{m,n}$ and $s' \in \underline{R}_{m-1,n-1}$ which are condensed on u and v, and which coincide with r and r', respectively, on both words. By (**WI 2c**), we have $\mathrm{ord}(s(u), s'(u)) = \mathrm{ord}(s(v), s'(v))$, and hence $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$. Thus Property (**WI 2**) holds. The verification of (**WI 3**) is identical. □

These results imply the following statement, which refines Corollary 1.9 and can be proved like that Corollary, using Propositions 1.6 and 3.10, and Theorem 3.11.

Corollary 3.12. *For each $m \geq 1$, we have $\underline{\mathcal{TL}}^{\mathsf{X}}_m \subseteq \mathcal{R}_m \subseteq \mathcal{FO}^2_m$ and $\underline{\mathcal{TL}}^{\mathsf{Y}}_m \subseteq \mathcal{L}_m \subseteq \mathcal{FO}^2_m$.*

3.3. **Condensed rankers determine a hierarchy of varieties.**

We now examine the algebraic properties of the relations $\triangleright_{m,n}$ and $\triangleleft_{m,n}$.

Lemma 3.13. *The relations $\triangleright_{m,n}$ and $\triangleleft_{m,n}$ are finite-index congruences.*

Proof. The relations $\triangleright_{m,n}$ and $\triangleleft_{m,n}$ are clearly equivalence relations, of finite index since $\underline{R}_{m,n}$ is finite. We now verify that if $b \in A$ and if u and v are $\triangleright_{m,n}$-equivalent, then so are ub and vb (resp. bu and bv).

The proof is by induction on $m+n$. The property of having the same subwords of length n is easily seen to be a congruence (and the proof of this fact can be found in [20] as it is related to Simon's theorem on piecewise testable languages). In view of Proposition 3.6 (1), this shows that $\triangleright_{1,n}$ and $\triangleleft_{1,n}$ are congruences.

Let us now assume that $n \geq m \geq 2$ and $u \triangleright_{m,n} v$. By Proposition 3.6 (2), we have $\mathsf{alph}(u) = \mathsf{alph}(v)$ and $u \triangleleft_{m-1,n-1} v$. It follows that $\mathsf{alph}(ub) = \mathsf{alph}(bu) = \mathsf{alph}(vb) = \mathsf{alph}(bv)$, and that $ub \triangleleft_{m-1,n-1} vb$ and $bu \triangleleft_{m-1,n-1} bv$ by induction.

Let now $a \in \mathsf{alph}(u) \cup \{b\}$. If $a \in \mathsf{alph}(u)$ and if $u = u_-au_+$ and $v = v_-av_+$ are a-left factorizations, then the a-left factorizations of ub and vb are $u_-\ a\ (u_+b)$ and $v_-\ a\ (v_+b)$. And the a-left factorizations of bu and bv are $(bu_-)\ a\ u_+$ and $(bv_-)\ a\ v_+$ — unless $a = b$, in which case these factorizations are εbu and εbv. By Proposition 3.6 (2) we have $u_- \triangleleft_{m-1,n-1} v_-$ and $u_+ \triangleright_{m,n-1} v_+$ ($u_+ \triangleright_{m-1,n-1} v_+$ if $n = m$). By induction, we have $u \triangleright_{m,n-1} v$, $bu_- \triangleleft_{m-1,n-1} bv_-$ and $u_+b \triangleright_{m,n-1} v_+b$ ($u_+b \triangleright_{m-1,n-1} v_+b$ if $n = m$).

If $a \notin \mathsf{alph}(u)$, and hence $a = b$, the a-left factorizations of ub and vb (resp. bu and bv) are $u\,b\,\varepsilon$ and $v\,b\,\varepsilon$ (resp. $\varepsilon\,b\,u$ and $\varepsilon\,b\,v$), and we do have $u \triangleleft_{m-1,n-1} v$ and $u \triangleright_{m,n-1} v$ ($u \triangleright_{m-1,n-1} v$ if $m = n$).

Thus all the conditions in Proposition 3.6 (2) are satisfied, whether a occurs in u and v or not, and we have established that $ub \triangleright_{m,n} vb$ and $bu \triangleright_{m,n} bv$. The proof regarding $\triangleleft_{m,n}$ is symmetric. □

Lemma 3.14. *If $\varphi\colon A^* \to B^*$ is a morphism and if $u, v \in A^*$ are $\triangleright_{m,n}$-equivalent (resp. $\triangleleft_{m,n}$-equivalent), then so are $\varphi(u)$ and $\varphi(v)$.*

Proof. We carry out the proof for the congruence $\triangleright_{m,n}$ by induction on $m+n$. The proof for $\triangleleft_{m,n}$ is symmetrical.

For $m = 1$, we show that if a ranker $r \in \underline{R}^{\mathsf{X}}_{1,n}$ is condensed on $\varphi(u)$, then it is condensed on $\varphi(v)$. If $u = a_1 \cdots a_\ell$, the word $\varphi(u)$ has a natural factorization in blocks, namely the $\varphi(a_i)$ and the sequence of positions in $\varphi(u)$ defined by the prefixes of r visits (some of) the $\varphi(a_i)$-blocks. This yields a factorization of r, $r = r_1 r_2 \cdots r_k$, where all the positions in $\varphi(u)$ visited while running r_1 are in the same block, say, $\varphi(a_{j(1)})$; then all the positions visited by the prefixes of r between r_1 (excluded) and $r_1 r_2$ (included) are in the block $\varphi(a_{j(2)})$ with $j(2) > j(1)$; and so on. In particular, the ranker $\mathsf{X}_{a_{j(1)}} \cdots \mathsf{X}_{a_{j(k)}}$ is defined on u, and hence on v. Therefore $v = v_0 a_{j(1)} v_1 \cdots a_{j(k)} v_{k+1}$. By construction, each r_i is defined on $\varphi(a_{j(i)})$, so r is defined on $\varphi(v)$, and condensed on that word (Example 3.3).

We now let $n \geq m \geq 2$ and $u \triangleright_{m,n} v$. It is immediate that $\mathsf{alph}(\varphi(u)) = \mathsf{alph}(\varphi(v))$ since u and v have the same alphabet. By Proposition 3.6 (2), we have $u \triangleleft_{m-1,n-1} v$, and by induction it follows that $\varphi(u) \triangleleft_{m-1,n-1} \varphi(v)$. Let now $b \in \mathsf{alph}(\varphi(u))$ and let $\varphi(u) = x_- b x_+$ and $\varphi(v) = y_- b y_+$ be b-left factorizations. The occurrence of b thus singled out in $\varphi(u)$ sits in some $\varphi(a)$, $a \in A$, and the corresponding occurrence of a in u is the leftmost one: we have an a-left factorization $u = u_- a u_+$ and a b-left factorization $\varphi(a) = x' b x''$ such that $x_- = \varphi(u_-) x'$ and $x_+ = x'' \varphi(u_+)$. Similarly, the leftmost occurrence of b in $\varphi(v)$ sits in some $\varphi(a')$, $a' \in A$: a' is the leftmost letter in v such that b occurs in $\varphi(a')$. If $a' \neq a$, the consideration of the rankers $\mathsf{X}_a \mathsf{Y}_{a'}$ and $\mathsf{X}_{a'} \mathsf{Y}_a$, which are simultaneously defined or not defined on u and v, yields a contradiction. Therefore $a' = a$ and if $v = v_- a v_+$ is the a-left factorization, then $y_- = \varphi(v_-) x'$ and $y_+ = x'' \varphi(v_+)$. By Proposition 3.6 (2) again, we have $u_- \triangleleft_{m-1,n-1} v_-$ and $u_+ \triangleright_{m,n-1} v_+$ ($u_+ \triangleright_{m-1,n-1} v_+$ if $n = m$). By induction, it follows that the same relations hold between the φ-images of u_-, v_-, u_+ and v_+, and we have $x_- \triangleleft_{m-1,n-1} y_-$ and $x_+ \triangleright_{m,n-1} y_+$ ($x_+ \triangleright_{m-1,n-1} y_+$ if $n = m$) by Lemma 3.13. Therefore $\varphi(u) \triangleright_{m,n} \varphi(v)$ by Proposition 3.6 (2). □

For each $n \geq m \geq 1$, let $\mathbf{R}_{m,n}$ (resp. $\mathbf{L}_{m,n}$) be the pseudovariety of monoids generated respectively by the monoids of the form $A^*/\triangleright_{m,n}$ (resp. $A^*/\triangleleft_{m,n}$). Since $\triangleright_{m,n'}$ refines $\triangleright_{m,n}$ when $n' \geq n$, the sequence $(\mathbf{R}_{m,n})_n$ is increasing and we let $\mathbf{R}_m$ be its union (a pseudovariety as well). The pseudovariety $\mathbf{L}_m$ is defined similarly, as the union of the $\mathbf{L}_{m,n}$.

Corollary 3.15. *If $\gamma\colon A^* \to M$ is a morphism into a monoid in $\mathbf{R}_{m,n}$, then there exists a morphism $\beta\colon A^*/\triangleright_{m,n} \to M$ such that $\gamma = \beta \circ \pi_A$, where $\pi_A\colon A^* \to A^*/\triangleright_{m,n}$ is the projection morphism. The same result holds for $\mathbf{L}_{m,n}$ and the quotient $A^*/\triangleleft_{m,n}$.*

Proof. By definition, there exists an onto morphism $\delta\colon N \to M$, and an injective morphism $\imath\colon N \hookrightarrow A_1^*/\triangleright_{m,n} \times \cdots \times A_k^*/\triangleright_{m,n}$. Let B be the disjoint union of the A_i, and for each i, let π_i be the morphism from B^* to A_i^* which erases all the letters not in A_i. By Lemma 3.14, $\triangleright_{m,n}$-equivalent elements have $\triangleright_{m,n}$-equivalent images, so we have a morphism $\pi\colon B^*/\triangleright_{m,n} \to \prod_i A_i^*/\triangleright_{m,n}$ as in Figure 2.

For each letter $a \in A$, we then pick a word $\varphi(a)$ in $\pi_B^{-1}\pi^{-1}\delta^{-1}\gamma(a) \subseteq B^*$: this defines a morphism $\varphi\colon A^* \to B^*$ such that $\delta \circ \pi \circ \pi_B \circ \varphi = \gamma$. By Lemma 3.14 again, there exists a morphism $\psi\colon A^*/\triangleright_{m,n} \to B^*/\triangleright_{m,n}$ such that $\psi \circ \pi_A = \pi_B \circ \varphi$. It follows that if $u \triangleright_{m,n} v$, then $\pi_B\varphi(u) = \pi_B\varphi(v)$, and hence $\gamma(u) = \gamma(v)$. This concludes the proof. □

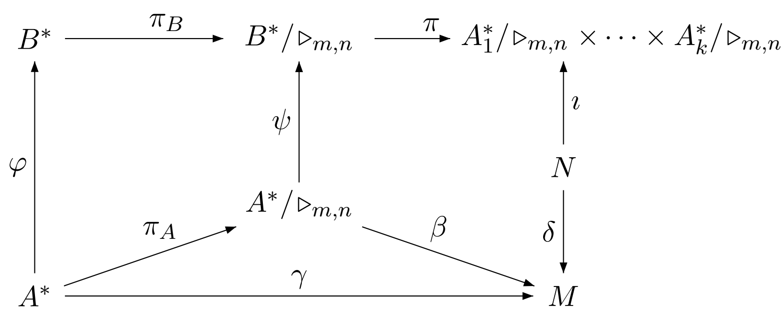

Figure 2: A commutative diagram

Corollary 3.16. *For each $m \geq 1$, $\mathcal{R}_m$ and $\mathcal{L}_m$ are varieties of languages and the corresponding pseudovarieties of monoids are $\mathbf{R}_m$ and $\mathbf{L}_m$.*

Proof. Every $L_c(r)$ ($r \in \underline{R}^{\mathsf{X}}_{m,n}$) is a union of $\triangleright_{m,n}$-classes, and hence it is recognized by $A^*/\triangleright_{m,n}$. Therefore every language in $\mathcal{R}_m$ is recognized by a monoid in $\mathbf{R}_m$ (and indeed, by $\pi_A\colon A^* \to A^*/\triangleright_{m,n}$ for n large enough).

Conversely, suppose that $L \subseteq A^*$ is recognized by a morphism $\gamma\colon A^* \to M$, into a monoid $M \in \mathbf{R}_m$. Then $M \in \mathbf{R}_{m,n}$ for some $n \geq m$ and by Corollary 3.15, there exists a morphism $\beta\colon A^*/\triangleright_{m,n} \to M$ such that $\gamma = \beta \circ \pi_A$. It follows that L is also accepted by π_A, L is a union of $\triangleright_{m,n}$-classes, and hence $L \in \mathcal{R}_m$. ☐

Example 3.17. It follows from Proposition 3.6 (i) that $\mathcal{R}_1 = \mathcal{L}_1$ is the variety of piecewise testable languages, and $\mathbf{R}_1 = \mathbf{L}_1 = \mathbf{J}$, the pseudovariety of $\mathcal{J}$-trivial monoids.

Remark 3.18. The proof of Corollary 3.16 also establishes that for each $n \geq m \geq 1$, the Boolean algebra $\mathcal{R}_{m,n}$ generated by the languages of the form $L_c(r)$, $r \in \underline{R}^{\mathsf{X}}_{m,n}$, defines a variety of languages, for which the corresponding pseudovariety of monoids is $\mathbf{R}_{m,n}$. In variety-theoretic terms, Corollary 3.15 states that $A^*/\triangleright_{m,n}$ is the free object of $\mathbf{R}_{m,n}$ over the alphabet A. The symmetrical statement also holds for $\mathcal{L}_{m,n}$ and the monoids $A^*/\triangleleft_{m,n}$.

We note the following containments.

Corollary 3.19. *For each $m \geq 1$, $\mathbf{R}_m$ and $\mathbf{L}_m$ are contained in $\mathbf{DA}$, and also in $\mathbf{R}_{m+1} \cap \mathbf{L}_{m+1}$.*

Proof. Since $\mathcal{FO}^2$ is the variety of languages corresponding to $\mathbf{DA}$, Corollary 3.12 yields the containment of $\mathbf{R}_m$ and $\mathbf{L}_m$ in $\mathbf{DA}$. Similarly, $\mathcal{R}_m$ and $\mathcal{L}_m$ are contained in both $\mathcal{R}_{m+1}$ and $\mathcal{L}_{m+1}$ by definition of these classes of languages – and this in turn implies the containment of the corresponding pseudovarieties. ☐

3.4. **Condensed rankers and deterministic products.** Recall that a product of languages $L = L_0a_1L_1 \cdots a_kL_k$ ($k \geq 1$, $a_i \in A$, $L_i \subseteq A^*$) is *deterministic* if, for $1 \leq i \leq k$, each word $u \in L$ has a unique prefix in $L_0a_1L_1 \cdots L_{i-1}a_i$. If for each i, the letter a_i does not occur in L_{i-1}, the product $L_0a_1L_1 \cdots a_kL_k$ is called *visibly deterministic*: this is obviously a particular case of a deterministic product.

The definition of a *co-deterministic* or *visibly co-deterministic* product is dual, in terms of suffixes instead of prefixes. If $\mathcal{V}$ is a class of languages and A is a finite alphabet, let $\mathcal{V}^{det}(A)$ (resp. $\mathcal{V}^{vdet}(A)$, $\mathcal{V}^{codet}(A)$, $\mathcal{V}^{vcodet}(A)$) be the set of all Boolean combinations of

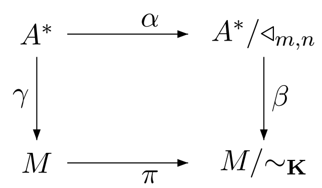

Figure 3: $M \in \mathbf{K} ⓜ \mathbf{L}_m$

languages of $\mathcal{V}(A)$ and of deterministic (resp. visibly deterministic, co-deterministic, visibly co-deterministic) products of languages of $\mathcal{V}(A)$.

Pin gave algebraic characterizations of the operations $\mathcal{V} \longmapsto \mathcal{V}^{det}$ and $\mathcal{V} \longmapsto \mathcal{V}^{codet}$, see [19, 22].

Proposition 3.20. *If $\mathcal{V}$ is a variety of languages and if $\mathbf{V}$ is the corresponding pseudovariety of monoids, then $\mathcal{V}^{det}$ and $\mathcal{V}^{codet}$ are varieties of languages and the corresponding pseudovarieties are, respectively, $\mathbf{K} ⓜ \mathbf{V}$ and $\mathbf{D} ⓜ \mathbf{V}$.*

This leads to the following statement.

Theorem 3.21. *For each $m \geq 1$, we have $\mathcal{R}_{m+1} = \mathcal{L}_m^{vdet} = \mathcal{L}_m^{det}$, $\mathcal{L}_{m+1} = \mathcal{R}_m^{vcodet} = \mathcal{R}_m^{codet}$, $\mathbf{R}_{m+1} = \mathbf{K} ⓜ \mathbf{L}_m$ and $\mathbf{L}_{m+1} = \mathbf{D} ⓜ \mathbf{R}_m$. In particular, $\mathbf{R}_2 = \mathbf{R}$ and $\mathbf{L}_2 = \mathbf{L}$.*

The proof uses the following technical property of monoids in $\mathbf{DA}$, whose proof can be found for instance in [6, Lemma 4.2].

Fact 3.22. Let $\sigma\colon A^* \to S$ be a morphism into a monoid $S \in \mathbf{DA}$. If $u, v \in A^*$, $a \in \mathsf{alph}(v)$ and $\sigma(u) \mathrel{\mathcal{R}} \sigma(uv)$, then $\sigma(uva) \mathrel{\mathcal{R}} \sigma(u)$.

Proof of Theorem 3.21. It is immediate from the definition that $\mathcal{L}_m^{vdet} \subseteq \mathcal{L}_m^{det}$.

Let $u \in A^*$ and let $B = \mathsf{alph}(u)$. For each $a \in B$, let $u = u_-^{(a)} a u_+^{(a)}$ be the a-left factorization of u. Let $[B]$ be the language of all strings with alphabet B, $[B] = \{u \in A^* \mid \mathsf{alph}(u) = B\}$. Observe that

$$[B] = \bigcap_{a \in B} L_c(\mathsf{X}_a) \setminus \bigcup_{a \notin B} L_c(\mathsf{X}_a) = \bigcap_{a \in B} L_c(\mathsf{Y}_a) \setminus \bigcup_{a \notin B} L_c(\mathsf{Y}_a).$$

This shows that $[B] \in \mathcal{R}_1 = \mathcal{L}_1$. (It is also well-known that $[B]$ is piecewise testable, and hence $[B] \in \mathcal{R}_1 = \mathcal{L}_1$.)

Now let $n > m \geq 1$. It follows from Proposition 3.6 that the $\triangleright_{m+1,n}$-class of u is the intersection of $[B]$, the $\triangleleft_{m,n-1}$-class of u and the products KaL ($a \in B$) where K is the $\triangleleft_{m,n-1}$-class of $u_-^{(a)}$ and L is the $\triangleright_{m+1,n-1}$-class of $u_+^{(a)}$ if $n > m+1$, the $\triangleright_{m,n-1}$-class of $u_+^{(a)}$ if $n = m+1$.

By definition of an a-left factorization, each of these products is visibly deterministic and, since every $\triangleleft_{m,n-1}$-class is a language in $\mathcal{L}_m$, we have shown that the $\triangleright_{m+1,n}$-class of u is in $\mathcal{L}_m^{vdet}$. Thus $\mathcal{R}_{m+1} \subseteq \mathcal{L}_m^{vdet}$.

To establish the last inclusion, namely $\mathcal{L}_m^{det} \subseteq \mathcal{R}_{m+1}$, we rather show $\mathbf{K} ⓜ \mathbf{L}_m \subseteq \mathbf{R}_{m+1}$.

Let $\gamma\colon A^* \to M$ be a surjective morphism, onto a monoid $M \in \mathbf{K} ⓜ \mathbf{L}_m$: we want to show that there exists a morphism from $A^*/\triangleright_{m+1,n}$ onto M for some $n > m$. Since $M \in \mathbf{K} ⓜ \mathbf{L}_m$, the monoid $M/{\sim_{\mathbf{K}}} \in \mathbf{L}_m$ and by Corollary 3.15, there exists an integer n and a morphism $\beta\colon A^*/\triangleleft_{m,n} \to M/{\sim_{\mathbf{K}}}$ such that $\beta \circ \alpha = \pi \circ \gamma$, where α is the projection from A^* onto $A^*/\triangleleft_{m,n}$ and π is the projection from M onto $M/{\sim_{\mathbf{K}}}$, see Figure 3.

Let ℓ be the maximal length of a strict $\mathcal{R}$-chain in M, that is: if $x_k <_{\mathcal{R}} \ldots <_{\mathcal{R}} x_1$ in M, then $k \leq \ell$. We show that, for any $u, v \in A^*$,

$$u \triangleright_{m+1,\ell|A|+n+1} v \Longrightarrow \gamma(u) = \gamma(v). \tag{3.1}$$

If $n' = \ell|A| + n + 1$, this implies the existence of a morphism from $A^*/\triangleright_{m+1,n'}$ onto M, as announced.

To prove implication (3.1), it suffices to show that we have

$$u \triangleright_{m+1,\ell|\mathsf{alph}(u)|+n+1} v \Longrightarrow \gamma(u) = \gamma(v), \tag{3.2}$$

which we prove by induction on $|\mathsf{alph}(u)|$. If $|\mathsf{alph}(u)| = 0$, then $u = \varepsilon$, $\mathsf{alph}(v) = \emptyset$ and $v = \varepsilon$ as well, so that $\gamma(u) = \gamma(v)$.

Now suppose that $u \neq \varepsilon$ and assume that $u \triangleright_{m+1,\ell|\mathsf{alph}(u)|+n+1} v$. Let $u = u_1 a_1 \cdots a_k u_{k+1}$ be the factorization of u such that each u_i is a word, each a_i is a letter and

$$1 \mathrel{\mathcal{R}} \gamma(u_1) >_{\mathcal{R}} \gamma(u_1 a_1) \cdots >_{\mathcal{R}} \gamma(u_1 a_1 \cdots u_k a_k) \mathrel{\mathcal{R}} \gamma(u_1 a_1 \cdots a_k u_{k+1}).$$

Then $k + 1 \leq \ell$, so $k < \ell$. Moreover, by Fact 3.22 (and Corollary 3.19), for each $1 \leq i \leq k$, $a_i \notin u_i$, so that each product $u_i a_i (u_{i+1} \cdots a_k u_{k+1})$ is an a_i-left factorization $(1 \leq i \leq k)$.

An easy induction on k, using Lemma 3.7, shows that v can then be factored as

$$v = v_1 a_1 v_2 \cdots a_k v_{k+1},$$

where $u_i \triangleright_{m+1,\ell|\mathsf{alph}(u)|+n-i+1} v_i$ for each $1 \leq i \leq k+1$. Moreover, for $1 \leq i \leq k$, $|\mathsf{alph}(u_i)| < |\mathsf{alph}(u)|$. Since $i \leq k < \ell$, we have $\ell|\mathsf{alph}(u)| + n - i \geq \ell|\mathsf{alph}(u_i)| + n + 1$, and by induction, we have $\gamma(u_i) = \gamma(v_i)$. However, it is possible that $\mathsf{alph}(u_{k+1}) = \mathsf{alph}(u)$, so we cannot conclude that $\gamma(u_{k+1}) = \gamma(v_{k+1})$.

But we do have the following:

$$u_{k+1} \triangleright_{m+1,\ell|\mathsf{alph}(u)|+n-k} v_{k+1} \text{ and } \gamma(u') = \gamma(v'),$$

where $u' = u_1 a_1 \cdots u_k a_k$ and $v' = v_1 a_1 \cdots v_k a_k$. The first relation implies that u_{k+1} and v_{k+1} are $\triangleleft_{m,\ell|\mathsf{alph}(u)|+n-k-1}$-equivalent. Since $k < \ell$, we have $\ell|\mathsf{alph}(u)| + n - k - 1 \geq n$, so $u_{k+1} \triangleleft_{m,n} v_{k+1}$ and hence, $\pi\gamma(u_{k+1}) = \pi\gamma(v_{k+1})$, that is, $\gamma(u_{k+1}) \sim_{\mathbf{K}} \gamma(v_{k+1})$.

Moreover, there exists a string $x \in A^*$ such that $\gamma(u') = \gamma(u' u_{k+1} x)$. Let ω be an integer such that every ω-power is idempotent in M: then $\gamma(u') = \gamma(u')\gamma(u_{k+1}x)^\omega$.

Now observe that $\gamma(u_{k+1}x)^\omega \mathrel{\mathcal{J}} \gamma(u_{k+1}x)^\omega \gamma(u_{k+1})$, since $\gamma(u_{k+1}x)^\omega = \gamma(u_{k+1}x)^{2\omega}$. It follows from $\gamma(u_{k+1}) \sim_{\mathbf{K}} \gamma(v_{k+1})$ that $\gamma(u_{k+1}x)^\omega \gamma(u_{k+1}) = \gamma(u_{k+1}x)^\omega \gamma(v_{k+1})$. Therefore we have

$$\gamma(u')\gamma(u_{k+1}) = \gamma(u')\gamma(v_{k+1}) \quad \text{and hence}$$
$$\gamma(u) = \gamma(u')\gamma(u_{k+1}) = \gamma(u')\gamma(v_{k+1}) = \gamma(v')\gamma(v_{k+1}) = \gamma(v).$$

This concludes the proof of Formula (3.2), and therefore of Theorem 3.21. □

3.5. **Structure of the $\mathbf{R}_m$ and $\mathbf{L}_m$ hierarchies.** It turns out that the hierarchies of pseudovarieties given by the $\mathbf{R}_m$ and the $\mathbf{L}_m$ were studied in the semigroup-theoretic literature (Trotter and Weil [35], Kufleitner and Weil [15]). In [15], they are defined as the hierarchies of pseudovarieties obtained from $\mathbf{J}$ by alternated applications of the operations

$\mathbf{X} \mapsto \mathbf{K} \circledm \mathbf{X}$ and $\mathbf{X} \mapsto \mathbf{D} \circledm \mathbf{X}$. Theorem 3.21 shows that these are the same hierarchies as those considered in this paper[1]. The following results are proved in [15, Section 4].

Proposition 3.23. *The hierarchies* $(\mathbf{R}_m)_m$ *and* $(\mathbf{L}_m)_m$ *are infinite chains of decidable pseudovarieties, and their unions are equal to* $\mathbf{DA}$*. Moreover, every* m*-generated monoid in* $\mathbf{DA}$ *lies in* $\mathbf{R}_{m+1} \cap \mathbf{L}_{m+1}$.

The results in [35, 15] go actually further, and give defining pseudoidentities for the pseudovarieties $\mathbf{R}_m$ and $\mathbf{L}_m$.

Remark 3.24. The way up in the $\mathbf{R}_m$-$\mathbf{L}_m$ hierarchy, by means of Mal'cev products with $\mathbf{K}$ and $\mathbf{D}$, is strongly reminiscent of the structure of the lattice of band varieties [8]. This observation is no coincidence, and forms the basis of the results in [15] which are used here.

4. The $\mathbf{R}_m$ hierarchy and unary temporal logic

We have seen in Corollary 3.12 that $\underline{\mathcal{TL}}_m^{\mathsf{X}} \subseteq \mathcal{R}_m$ and $\underline{\mathcal{TL}}_m^{\mathsf{Y}} \subseteq \mathcal{L}_m$. In Theorem 4.3 below, we prove a weak converse. Let us however make the following observation.

Proposition 4.1. *We have*

$$\underline{\mathcal{TL}}_1^{\mathsf{X}} = \underline{\mathcal{TL}}_1^{\mathsf{Y}} = \mathcal{R}_1 = \mathcal{L}_1,$$
$$\underline{\mathcal{TL}}_2^{\mathsf{X}} = \mathcal{R}_2, \quad \underline{\mathcal{TL}}_2^{\mathsf{Y}} = \mathcal{L}_2.$$

Proof. The statement concerning $\underline{\mathcal{TL}}_1$ was already proved in Remark 1.7. Let us now establish that $\mathcal{R}_2 \subseteq \underline{\mathcal{TL}}_2^{\mathsf{X}}$. We show, by induction on $n \geq 2$, that if u and v agree on rankers in $\underline{R}_{2,2n}^{\mathsf{X}}$, then they agree on condensed rankers in $\underline{R}_{2,n}^{\mathsf{X}}$: $u \triangleright_{2,n} v$. We use the characterization of $\triangleright_{2,n}$ in Proposition 3.6.

The consideration of 1-letter rankers shows that $\mathsf{alph}(u) = \mathsf{alph}(v)$. Moreover, since $\underline{R}_{1,n-1}^{\mathsf{Y}}$ is contained in $\underline{R}_{2,2n}^{\mathsf{X}}$, and since these rankers are condensed where they are defined, we find that $u \triangleleft_{1,n-1} v$. Similarly, let $u = u_- a u_+$ and $v = v_- a v_+$ be a-left factorizations, and let $s \in \underline{R}_{1,n-1}^{\mathsf{Y}}$. Then s is condensed on u_- if and only if s is defined on u_-, if and only if $\mathsf{X}_a s$ is defined on u (Lemma 3.5). Since $\mathsf{X}_a s \in \underline{R}_{2,2n}^{\mathsf{X}}$ and u and v agree on such rankers, it follows that $\mathsf{X}_a s$ is defined on v, and s is condensed on v_-. Thus $u_- \triangleleft_{1,n-1} v_-$.

Now we need to show that $u_+ \triangleright_{2,n-1} v_+$ if $n \geq 3$, $u_+ \triangleright_{1,1} v_+$ if $n = 2$. Suppose first that $n = 2$ and consider $s \in \underline{R}_{1,1}^{\mathsf{X}}$, condensed on u_+. Then $s = \mathsf{X}_b$ for some $b \in A$ and the consideration of $r = \mathsf{X}_a \mathsf{X}_b$ (in $\underline{R}_{2,2}^{\mathsf{X}}$) shows that s is condensed on v_+ as well. This settles the case $n = 2$.

Let us now assume that $n \geq 3$ and let us show that $u_+ \triangleright_{2,n-1} v_+$. By induction, it suffices to show that u_+ and v_+ agree on rankers in $\underline{R}_{2,2n-2}^{\mathsf{X}}$. So let $s \in \underline{R}_{2,2n-2}^{\mathsf{X}}$ be defined on u_+. Then for every prefix p of s ending with a Y-letter, $\mathsf{X}_a p \mathsf{Y}_a$ is defined on u (Lemma 3.5). Since $\mathsf{X}_a p \mathsf{Y}_a \in \underline{R}_{2,2n}^{\mathsf{X}}$, it follows that $\mathsf{X}_a p \mathsf{Y}_a$ is defined on v, and hence s is defined on v_+. This concludes the proof. ☐

[1]More precisely, the pseudovarieties $\mathbf{R}_m$ and $\mathbf{L}_m$ in [15] are pseudovarieties of semigroups, and the $\mathbf{R}_m$ and $\mathbf{L}_m$ considered in this paper are the classes of monoids in these pseudovarieties.

Example 4.2 below shows that the statement of Proposition 4.1 cannot be extended to the higher levels of the hierarchy.

Example 4.2. We show in this example that $\underline{\mathcal{TL}}^{\mathsf{X}}_3$ is properly contained in $\mathcal{R}_3$. More precisely, let $r_0 = \mathsf{X}_a\mathsf{Y}_b\mathsf{X}_c \in R^{\mathsf{X}}_{3,3}$. We show that $L_c(r_0)$, a language in $\mathcal{R}_3$, is not $\underline{\mathsf{TL}}^{\mathsf{X}}_3$-definable.

Let $u_n = (bc)^n(a(bc)^n)^n$ and $v_n = (bc)^n b(a(bc)^n)^n$ $(n \geq 1)$. It is easily verified that r_0 is condensed on u_n, and that it is defined and not condensed on v_n: that is, for each n, $u_n \in L_c(r_0)$ and $v_n \notin L_c(r_0)$.

We now show that u_n and v_n agree on all rankers in $\underline{R}^{\mathsf{X}}_{3,n}$, so that any $\underline{\mathsf{TL}}^{\mathsf{X}}_3$-definable language contains either both u_n and v_n, or neither – and hence $L_c(r_0)$ is not $\underline{\mathsf{TL}}^{\mathsf{X}}_3$-definable.

Let $r \in \underline{R}^{\mathsf{X}}_{3,n}$. If r starts with a Y-letter, then any two words ending with $(a(bc)^n)^n$ agree on r. In particular, u_n and v_n agree on r. Similarly, if r starts with an X-letter and does not contain the letters X_a or Y_a, then any two words starting with $(bc)^n$ agree on r, so u_n and v_n agree on r.

Finally, assume that r starts with an X-letter and that $r = s_0\mathsf{Z}^{(1)}_a s_1 \cdots \mathsf{Z}^{(k)}_a s_k$ with $k > 0$, each $\mathsf{Z}^{(i)} \in \{\mathsf{X}, \mathsf{Y}\}$ and each s_i a (possibly empty) ranker avoiding the letters X_a and Y_a. We denote by p_i the prefix $p_i = s_0\mathsf{Z}^{(1)}_a s_1 \cdots \mathsf{Z}^{(i)}_a$.

Suppose first that $r \in R^{\mathsf{X}}_{1,n}$. Then p_i coincides with X^i_a on u_n as well as on v_n. Therefore r is defined and coincides with $\mathsf{X}^k_a s_k$ on both words.

Suppose now that $r \in R^{\mathsf{X}}_{2,n}$, say $r = r'r''$ with r' a non-empty string of X-letters and r'' a non-empty string of Y-letters. If r' is shorter than p_1, then $\mathsf{Z}^{(1)} = \mathsf{Y}$ and r is not defined on either u_n or v_n. If $\mathsf{Z}^{(1)} = \mathsf{X}$, let i be maximal such that p_i is a prefix of r', say $r' = p_i s'_i$. Then $i > 0$ and p_i coincides with X^i_a on u_n, as well as on v_n.

If $i = k$, then r is defined on u_n and v_n, and it coincides with $\mathsf{X}^k_a s_k$ on both words.

If $1 \leq i < k$, s'_i is non-empty and s_i is defined on $(bc)^n$, then p_{i+1} coincides with X^i_a on u_n and v_n. Thus r is defined on u_n (resp. v_n) if and only $i \geq k - i$, and in that case, it coincides with $\mathsf{X}^{k-2i}_a s_k$.

If $1 < i < k$, s'_i is non-empty and s_i is not defined on $(bc)^n$, or if s'_i is empty, then p_{i+1} coincides with X^{i-1}_a on u_n and v_n. Thus r is defined on u_n (resp. v_n) if and only $i > k - i$, and in that case, it coincides with $\mathsf{X}^{k-2i-1}_a s_k$.

Finally, if $1 = i < k$, s'_i is non-empty and s_i is not defined on $(bc)^n$, or if s'_i is empty, then r (and even p_{i+1}) is not defined on either u_n or v_n.

Finally, let us assume that $r \in R^{\mathsf{X}}_{3,n}$, say $r = r'r''r'''$ with r' and r''' non-empty strings of X-letters and r'' a non-empty string of Y-letters. Again, let i be maximal such that p_i is a prefix of r' ($i = 0$ if p_1 is not a prefix of r') and let j be maximal such that p_j is a prefix of $r'r''$. Then $r'r'' = p_j s'_j$ for some prefix s'_j of s_j. By the previous analysis, if $i \leq 1 < j$, then $r'r''$ is not defined on u_n nor on v_n, and hence neither is r. In all other cases, $r'r''$ is defined on both words and coincides with $\mathsf{X}^{2i-j}_a s'_j$ or $\mathsf{X}^{2i-j-1}_a s'_j$. Since $(k-j)+(2i-j) \leq k$, r is defined on u_n and v_n, and coincides on these words with $\mathsf{X}^{k-j+2i-j}_a s_k$ or $\mathsf{X}^{k-j+2i-j-1}_a s_k$.

To conclude this example, note that u_n and v_n disagree on rankers in $\underline{R}^{\mathsf{X}}_4$. More precisely, the ranker $\mathsf{X}_a\mathsf{Y}_c\mathsf{X}_b\mathsf{Y}_a$ is defined on u_n but not on v_n. Further getting ahead of ourselves, we note that this example also shows (in view of Theorem 5.1) that $\underline{\mathcal{TL}}_3$ is properly contained in $\mathcal{FO}^2_3$.

Finally we prove a result on the containment of the $\mathcal{R}_m$ and $\mathcal{L}_m$ hierarchies in the $\underline{\mathcal{TL}}_m$ hierarchy.

Theorem 4.3. *Let $m \geq 1$. Then $\mathcal{R}_m \subseteq \underline{\mathcal{TL}}^{\mathsf{X}}_{2m-1}$ and $\mathcal{L}_m \subseteq \underline{\mathcal{TL}}^{\mathsf{Y}}_{2m-1}$.*

More precisely, for all $n \geq m$, $\mathsf{Z} \in \{\mathsf{X}, \mathsf{Y}\}$ and $u, v \in A^*$, if u and v agree on rankers in $\underline{R}^{\mathsf{Z}}_{2m-1,2n-1}$, then they agree on condensed rankers in $\underline{R}^{\mathsf{Z}}_{m,n}$.

Proof. Without loss of generality, we may assume $\mathsf{Z} = \mathsf{X}$. The proof is by induction on m. The result is trivial if $m = 1$, since $2m-1 = 1$ and $2n-1 \geq n$. We now assume that $m \geq 2$ and u, v agree on rankers in $\underline{R}^{\mathsf{X}}_{2m-1,2n-1}$.

We use the characterization of $\triangleright_{m,n}$ in Proposition 3.6: the consideration of length 1 rankers shows that $\mathsf{alph}(u) = \mathsf{alph}(v)$. Since $\underline{R}^{\mathsf{Y}}_{2m-3,2n-3}$ is contained in $\underline{R}^{\mathsf{X}}_{2m-1,2n-1}$, we have $u \triangleleft_{m-1,n-1} v$ by induction. Now, for each letter $a \in \mathsf{alph}(u)$, let $u = u_- a u_+$ and $v = v_- a v_+$ be the a-left factorizations. We want to show that $u_- \triangleleft_{m-1,n-1} v_-$ and $u_+ \triangleright_{m,n-1} v_+$ ($u_+ \triangleright_{m-1,n-1} v_+$ if $m = n$). By induction, it suffices to show that u_- and v_- agree on rankers in $\underline{R}^{\mathsf{Y}}_{2m-3,2n-3}$, and u_+ and v_+ agree on rankers in $\underline{R}^{\mathsf{X}}_{2m-1,2n-3}$ ($\underline{R}^{\mathsf{X}}_{2m-3,2n-3}$ if $m = n$). In the rest of the proof we silently rely on the results of Lemma 3.5.

Let $s \in \underline{R}^{\mathsf{Y}}_{2m-3,2n-3}$ be defined on u_-. If s starts with a Y-block, then $\mathsf{X}_a s \in \underline{R}^{\mathsf{X}}_{2m-2,2n-2}$ and $\mathsf{X}_a s$ is defined on u. Moreover, if p is any prefix of s, then $\mathsf{X}_a p \mathsf{Y}_a \in \underline{R}^{\mathsf{X}}_{2m-1,2n-1}$ is not defined on u. It follows that s is defined on v_-.

If instead s starts with an X-block, then $s \in \underline{R}^{\mathsf{X}}_{2m-4,2n-4}$ and s is defined on u. If p is any prefix of s, then $p\mathsf{Y}_a \in \underline{R}^{\mathsf{X}}_{2m-3,2n-3}$ and $p\mathsf{Y}_a$ is not defined on u. As all these rankers are in $\underline{R}^{\mathsf{X}}_{2m-1,2n-1}$, the same holds on v and s is defined on v_-.

Let now $s \in \underline{R}^{\mathsf{X}}_{2m-1,2n-3}$ ($s \in \underline{R}^{\mathsf{X}}_{2m-3,2n-3}$ if $n = m$) be defined on u_+. If s starts with an X-block, then $\mathsf{X}_a s \in \underline{R}^{\mathsf{X}}_{2m-1,2n-2}$ ($\mathsf{X}_a s \in \underline{R}^{\mathsf{X}}_{2m-3,2n-2}$ if $n = m$) and $\mathsf{X}_a s$ is defined on u. Moreover, for each prefix p of s ending with a Y-letter, $\mathsf{X}_a p \mathsf{Y}_a \in \underline{R}^{\mathsf{X}}_{2m-1,2n-1}$ ($\mathsf{X}_a p \mathsf{Y}_a \in \underline{R}^{\mathsf{X}}_{2m-3,2n-1}$ if $n = m$) and $\mathsf{X}_a p \mathsf{Y}_a$ is defined on u. As all these rankers are in $\underline{R}^{\mathsf{X}}_{2m-1,2n-1}$, the same holds on v and s is defined on v_+.

If instead s starts with a Y-block, then $s \in \underline{R}^{\mathsf{Y}}_{2m-2,2n-4}$ ($s \in \underline{R}^{\mathsf{Y}}_{2m-4,2n-4}$ if $n = m$) and s is defined on u. Moreover, if p is any prefix of s ending with a Y-letter, $p\mathsf{Y}_a \in \underline{R}^{\mathsf{Y}}_{2m-2,2n-3}$ ($p\mathsf{Y}_a \in \underline{R}^{\mathsf{Y}}_{2m-4,2n-3}$ if $n = m$) and $p\mathsf{Y}_a$ is defined on u. As all these rankers are in $\underline{R}^{\mathsf{X}}_{2m-1,2n-1}$, the same holds on v and s is defined on v_+. □

The containment of $\mathcal{R}_m$ and $\mathcal{L}_m$ into $\underline{\mathcal{TL}}^{\mathsf{X}}_{2m-1}$ and $\underline{\mathcal{TL}}^{\mathsf{Y}}_{2m-1}$, respectively, is not very precise, unfortunately, especially in view of Theorem 6.4 below.

5. The $\mathbf{R}_m$ hierarchy and FO^2_m

The objective of this section is to prove the following theorem.

Theorem 5.1. *Let $m \geq 1$. Every language in $\mathcal{R}_m$ or $\mathcal{L}_m$ is FO^2_m-definable, and every FO^2_m-definable language is in $\mathcal{R}_{m+1} \cap \mathcal{L}_{m+1}$. Equivalently, we have*

$$\mathbf{R}_m \vee \mathbf{L}_m \subseteq \mathbf{FO}^2_m \subseteq \mathbf{R}_{m+1} \cap \mathbf{L}_{m+1},$$

where $\mathbf{V} \vee \mathbf{W}$ denotes the least pseudovariety containing $\mathbf{V}$ and $\mathbf{W}$.

5.1. **Are the containments in Theorem 5.1 strict?** In the particular case where $m = 1$, we know that $\mathbf{R}_2 \cap \mathbf{L}_2 = \mathbf{R} \cap \mathbf{L} = \mathbf{J} = \mathbf{R}_1 \vee \mathbf{L}_1$: this reflects the fact that $\mathcal{FO}^2_1$ is the class the piecewise testable languages. However, we conjecture that this equality does not hold for larger values of m.

Conjecture 5.2. For $m \geq 2$, $\mathbf{R}_m \vee \mathbf{L}_m$ is properly contained in $\mathbf{R}_{m+1} \cap \mathbf{L}_{m+1}$.

The following example proves the conjecture for $m = 2$.

Example 5.3. $L = \{b, c\}^* ca\{a, b\}^*$ is FO^2_2-definable, by the following formula:

$$\begin{aligned} & \exists i \quad (\mathbf{c}(i) \wedge (\forall j \ (j < i \rightarrow \neg \mathbf{a}(j))) \wedge (\forall j \ (j > i \rightarrow \neg \mathbf{c}(j)))) \\ \wedge\ & \exists i \quad (\mathbf{a}(i) \wedge (\forall j \ (j < i \rightarrow \neg \mathbf{a}(j))) \wedge (\forall j \ (j > i \rightarrow \neg \mathbf{c}(j)))) \\ \wedge\ & \forall i \quad (\mathbf{b}(i) \rightarrow (\exists j \ (j < i \wedge \mathbf{a}(j)) \vee (\exists j \ (j > i \wedge \mathbf{c}(j)))). \end{aligned}$$

The words $u_n = (bc)^n(ab)^n$ are in L, while the words $v_n = (bc)^n b(ca)^n$ are not. Almeida and Azevedo showed that $\mathbf{R}_2 \vee \mathbf{L}_2$ is defined by the pseudo-identity $(bc)^\omega(ab)^\omega = (bc)^\omega b(ab)^\omega$ [2, Theorem 9.2.13 and Exercise 9.2.15]). In particular, for each language K recognized by a monoid in $\mathbf{R}_2 \vee \mathbf{L}_2$, the words u_n and v_n (for n large enough) are all in K, or all in the complement of K. Therefore L is not recognized by such a monoid, which proves that $\mathbf{R}_2 \vee \mathbf{L}_2$ is strictly contained in $\mathbf{FO}^2_2$, and hence also in $\mathbf{R}_3 \cap \mathbf{L}_3$. It also shows that $\underline{\mathcal{TL}}_2$ is properly contained in $\mathcal{FO}^2_2$.

Finally, we formulate the following conjecture.

Conjecture 5.4. For each $m \geq 1$, $\mathbf{FO}^2_m = \mathbf{R}_{m+1} \cap \mathbf{L}_{m+1}$.

5.2. **Proof of Theorem 5.1.** Corollary 3.12 already established that every language in $\mathcal{R}_m$ or $\mathcal{L}_m$ is FO^2_m-definable[2].

In view of Theorem 3.11, to establish that $\mathcal{FO}^2_m$ is contained in $\mathcal{R}_{m+1} \cap \mathcal{L}_{m+1}$, it suffices to prove the following result.

> *For each $n \geq m \geq 1$, if $u \triangleright_{m+1,2n} v$ or $u \triangleleft_{m+1,2n} v$, then Properties (**WI 1c**), (**WI 2c**) and (**WI 3c**) hold for m, n.*

The result is trivial if $m = 1$, since in that case, only Property (**WI 1c**) is non-vacuous.

So we now assume that $m \geq 2$, and $u \triangleright_{m+1,2n} v$ or $u \triangleleft_{m+1,2n} v$. Property (**WI 1c**) holds trivially, by definition of the $\triangleright_{m+1,2n}$ and $\triangleleft_{m+1,2n}$ relations. We now concentrate on proving that Properties (**WI 2c**) and (**WI 3c**) also hold for m, n, a task that will be completed in Section 5.2.3.

5.2.1. *The case where r and r' start with opposite directions.*

Proposition 5.5. *Let $n \geq m \geq 1$, $r = \mathsf{Y}_{a_1} s \in \underline{R}^{\mathsf{Y}}_{m,n}$ and $r' = \mathsf{X}_c$. If $u, v \in A^*$, r is condensed on u and v and $u \triangleright_{m,n+1} v$ or $u \triangleleft_{m+1,n+1} v$, then* $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$. *The dual statement (involving $r = \mathsf{X}_{a_1} s \in \underline{R}^{\mathsf{X}}_{m,n}$ and $r' = \mathsf{Y}_c$) holds as well.*

[2]Of course, the same fact can be proved by the direct construction of an FO^2_m-formula for each $\triangleright_{m,n}$-class (by induction on m and using Proposition 3.6).

Proof. First suppose that $u \lhd_{m+1,n+1} v$, that is, u and v agree on condensed rankers in $\underline{R}^{\mathsf{Y}}_{m+1,n+1}$. We are in exactly one of the following three situations:

- $r\mathsf{Y}_c$ is defined on u, in which case $r'(u) < r(u)$;
- $r\mathsf{Y}_c$ is undefined on u and c is the last letter to occur in r, in which case $r'(u) = r(u)$;
- $r\mathsf{Y}_c$ is undefined on u and c is not the last letter to occur in r, in which case $r(u) < r'(u)$.

The same trichotomy holds for v. Since $r\mathsf{Y}_c \in \underline{R}^{\mathsf{Y}}_{m+1,n+1}$, u and v agree on $r\mathsf{Y}_c$ (Proposition 3.10), and hence $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$.

Let us now assume that $u \rhd_{m,n+1} v$, so that u and v agree on condensed rankers in $\underline{R}^{\mathsf{X}}_{m,n+1}$. If $m = 1$ then r is of the form $r = \mathsf{Y}_{a_1} \cdots \mathsf{Y}_{a_k}$ and we observe again that

- either $\mathsf{X}_c\mathsf{X}_{a_k} \cdots \mathsf{X}_{a_1} \in R^{\mathsf{X}}_{1,n+1}$ is defined on u, and we have $r'(u) < r(u)$;
- or $\mathsf{X}_c\mathsf{X}_{a_k} \cdots \mathsf{X}_{a_1}$ is undefined on u and $c = a_k$, and we have $r'(u) = r(u)$;
- or $\mathsf{X}_c\mathsf{X}_{a_k} \cdots \mathsf{X}_{a_1}$ is undefined on u and $c \neq a_k$, and we have $r'(u) > r(u)$.

The same holds for v since $\mathsf{X}_c\mathsf{X}_{a_k} \cdots \mathsf{X}_{a_1} \in \underline{R}^{\mathsf{X}}_{1,n+1}$ and such rankers are condensed where they are defined. Therefore we have $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$.

We now assume that $m \geq 2$. Let $u = u_-cu_+$ and $v = v_-cv_+$ be c-left factorizations. We distinguish two cases depending on the direction of the second letter of r.

First suppose that $r = \mathsf{Y}_{a_1}\mathsf{Y}_{a_2}s'$. If $a_1 \notin \mathsf{alph}(u_+)$, then $r(u) < r'(u)$ (because r is condensed on u). Since $u_+ \rhd_{m,n} v_+$, we have $\mathsf{alph}(u_+) = \mathsf{alph}(v_+)$, so $r(v) < r'(v)$ as well. If instead $a_1 \in \mathsf{alph}(u_+) = \mathsf{alph}(v_+)$, let $u_+ = u_0a_1u_1$ and $v_+ = v_0a_1v_1$ be the a_1-right factorizations. Then

$$\begin{aligned}\mathrm{ord}(r(u), r'(u)) &= \mathrm{ord}(\mathsf{Y}_{a_2}s'(u_-cu_0), r'(u_-cu_0)) \text{ and}\\ \mathrm{ord}(r(v), r'(v)) &= \mathrm{ord}(\mathsf{Y}_{a_2}s'(v_-cv_0), r'(v_-cv_0)).\end{aligned}$$

Since $(u_-cu_0)a_1u_1$ and $(v_-cv_0)a_1u_1$ are a_1-right factorizations as well, we deduce from Lemma 3.7 that $u_-cu_0 \rhd_{m,n} v_-cv_0$ and it follows by induction on the length of r that

$$\mathrm{ord}(\mathsf{Y}_{a_2}s'(u_-cu_0), r'(u_-cu_0)) = \mathrm{ord}(\mathsf{Y}_{a_2}s'(v_-cv_0), r'(v_-cv_0)).$$

The other case is $r = \mathsf{Y}_{a_1}\mathsf{X}_{b_1}s'$. If $a_1 \in \mathsf{alph}(cu_+) = \mathsf{alph}(cv_+)$ then $r'(u) < r(u)$ and $r'(v) < r(v)$. If instead $a_1 \notin \mathsf{alph}(cu_+) = \mathsf{alph}(cv_+)$, we first consider the case where r has a single alternation, i.e., $r = \mathsf{Y}_{a_1}\mathsf{X}_{b_1} \cdots \mathsf{X}_{b_k}$. We have $r(u) < r'(u)$ if and only if r is defined on u_-, and hence condensed (Example 3.3). Since $u_- \rhd_{m,n} v_-$ (Lemma 3.7), this is the case if and only if r is defined on v_-. Hence, if r is defined on u_-, we have $r(u) < r'(u)$ and $r(v) < r'(v)$. If r is not defined on u_-, but $\mathsf{Y}_{a_1}\mathsf{X}_{b_1} \cdots \mathsf{X}_{b_{k-1}}$ is defined on u_- and $b_k = c$, then the same holds for v and we have $r(u) = r'(u)$ and $r(v) = r'(v)$. Otherwise, we have $r(u) > r'(u)$ and $r(v) > r'(v)$.

The last situation arises if r is of the form $r = \mathsf{Y}_{a_1}\mathsf{X}_{b_1} \cdots \mathsf{X}_{b_k}\mathsf{Y}_d s''$. In particular, $m \geq 3$. If $\mathsf{Y}_{a_1}\mathsf{X}_{b_1} \cdots \mathsf{X}_{b_k}$ is defined on u_-c, then it is defined on v_-c as well (by the same reasoning as in the previous paragraph) and we have $r(u) < r'(u)$ and $r(v) < r'(v)$.

Similarly, if $\mathsf{Y}_{a_1}\mathsf{X}_{b_1} \cdots \mathsf{X}_{b_{k-1}}$ is not defined on u_- and v_-, then we have $r'(u) < r(u)$ and $r'(v) < r(v)$.

Finally, let us assume that $\mathsf{Y}_{a_1}\mathsf{X}_{b_1} \cdots \mathsf{X}_{b_k}$ is not defined on u_-c or v_-c, but $\mathsf{Y}_{a_1}\mathsf{X}_{b_1} \cdots \mathsf{X}_{b_{k-1}}$ is defined on u_- and v_-. Let $u_+ = u_0b_ku_1$ and $v_+ = v_0b_kv_1$ be b_k-left factorizations. Then

$$\begin{aligned}\mathrm{ord}(r(u), r'(u)) &= \mathrm{ord}(\mathsf{Y}_d s''(u_-cu_0), r'(u_-cu_0)) \text{ and}\\ \mathrm{ord}(r(v), r'(v)) &= \mathrm{ord}(\mathsf{Y}_d s''(v_-cv_0), r'(v_-cv_0)).\end{aligned}$$

Since $u \triangleright_{m,n+1} v$, we have $u_+ \triangleright_{m,n} v_+$, and by Lemma 3.7 , $u_- \triangleright_{m,n} v_-$ and $u_0 \triangleright_{m,n-1} v_0$. Therefore $u_-cu_0 \triangleright_{m,n-1} v_-cv_0$. Since $\mathsf{Y}_d s'' \in \underline{R}^{\mathsf{Y}}_{m-2,n-2}$ is condensed on both u_-cu_0 and v_-cv_0, we conclude by induction on the length of r that $\mathrm{ord}(\mathsf{Y}_d s''(u_-cu_0), r'(u_-cu_0)) = \mathrm{ord}(\mathsf{Y}_d s''(v_-cv_0), r'(v_-cv_0))$ and hence $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$.

This concludes the proof. □

Proposition 5.6. *Let $n > m \geq 1$, let $r = \mathsf{X}_a s \in \underline{R}^{\mathsf{X}}_m$ and $r' = \mathsf{Y}_b s' \in \underline{R}^{\mathsf{Y}}_m$ such that $|r| + |r'| \leq n$, and let $u, v \in A^*$ such that r and r' are condensed on u and v. If $u \triangleright_{m+1,n} v$ or $u \triangleleft_{m+1,n} v$, then* $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$.

Proof. Without loss of generality, we assume that $u \triangleright_{m+1,n} v$. We proceed by induction, first on m. If $m = 1$, then $r = \mathsf{X}_{a_1} \cdots \mathsf{X}_{a_k}$ and $r' = \mathsf{Y}_{b_1} \cdots \mathsf{Y}_{b_\ell}$ with $k + l \leq n$. We observe that if $p = r\mathsf{X}_{b_\ell} \cdots \mathsf{X}_{b_1}$ is defined on u, then $r(u) < r'(u)$; if p is not defined on u, but $a_k = b_\ell$ and $r\mathsf{X}_{b_{\ell-1}} \cdots \mathsf{X}_{b_1}$ is defined on u, then $r(u) = r'(u)$; and in all other cases, $r(u) > r'(u)$. The same holds for v, and this completes the proof in case $m = 1$.

We now assume that $m \geq 2$ and proceed by induction on n. We first note that if one of r, r' has length 1, then the result was established in Proposition 5.5. We now assume that $|r|, |r'| \geq 2$ (so $|r|, |r'| \leq n - 2$).

Suppose that $n = m + 1$ and let $\beta(r)$ the number of alternating blocks in r: then $\beta(r) \leq |r| \leq n - |r'| \leq n - 2 = m - 1$. The same inequality holds for r' and we conclude by induction on m.

We must now consider the case where $n > m+1 > 2$. In particular, we have $r \in \underline{R}^{\mathsf{X}}_{m,n-2}$ and $r' \in \underline{R}^{\mathsf{X}}_{m+1,n-1}$.

First case: s starts with an X*-block.* Let $u = u_-au_+$ and $v = v_-av_+$ be a-left-factorizations. Then s is condensed on u_+ and v_+ and $u_+ \triangleright_{m+1,n-1} v_+$, so u_+ and v_+ agree on rankers in $\underline{R}^{\mathsf{X}}_{m+1,n-1}$ (Proposition 3.10). In particular, u_+ and v_+ agree on r'. If r' is defined on u_+, then $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(s(u_+), r'(u_+))$. Moreover, r' is defined on v_+ as well and $\mathrm{ord}(r(v), r'(v)) = \mathrm{ord}(s(v_+), r'(v_+))$, so we conclude by induction. If instead r' is not defined on u_+ or v_+, then $r'(u) \leq \mathsf{X}_a(u) < r(u)$ and $r'(v) \leq \mathsf{X}_a(v) < r(v)$.

Second case: s' starts with a Y*-block.* Let $u = u_-bu_+$ and $v = v_-bv_+$ be b-right factorizations. Then $u_- \triangleright_{m+1,n-1} v_-$ by Lemma 3.7 and this case can be handled exactly like the previous one.

Third case: s starts with a Y*-block and s' starts with an* X*-block.* If $\mathsf{X}_a(u) \leq \mathsf{Y}_b(u)$, then $\mathsf{X}_a(v) \leq \mathsf{Y}_b(v)$ (by Proposition 5.5), we have $r(u) < \mathsf{X}_a(u) \leq \mathsf{Y}_b(u) < r'(u)$, and the same inequalities hold for v.

We now assume that $\mathsf{X}_a(u) > \mathsf{Y}_b(u)$ and $\mathsf{X}_a(v) > \mathsf{Y}_b(v)$. In particular, $a \neq b$. Identifying the first a and the last b in u and v, we get factorizations $u = u_-bu_0au_+$ and $v = v_-bv_0av_+$ such that $a \notin \mathsf{alph}(u_-bu_0) \cup \mathsf{alph}(v_-bv_0)$ and $b \notin \mathsf{alph}(u_0au_+) \cup \mathsf{alph}(v_0av_+)$. In particular, $r(u) = s(u_-bu_0)$, $r'(u)$ is the position $s'(u_0au_+)$ in the suffix u_0au_+ of u, and the same holds in v. Moreover, $u = (u_-bu_0)au_+$ is an a-left factorization, $u = u_-b(u_0au_+)$ is a b-right factorization, and the same holds in v. Therefore, and since $u \triangleright_{m+1,n} v$, we have $u_-bu_0 \triangleleft_{m,n-1} v_-bv_0$ by definition and $u_0 \triangleleft_{m,n-2} v_0$ by Lemma 3.7.

Since $s \in \underline{R}^{\mathsf{Y}}_{m-1,n-3}$ and $s' \in \underline{R}^{\mathsf{X}}_{m-1,n-3} \subseteq \underline{R}^{\mathsf{Y}}_{m,n-2}$, Proposition 3.10 shows that, if s is not defined on u_0, then it is not defined on v_0 either, and $r(u) \leq \mathsf{Y}_b(u) < r'(u)$ and similarly, $r(v) < r'(v)$. Symmetrically, if s' is not defined on u_0, then $r(u) < \mathsf{X}_a(u) \leq r'(u)$ and $r(v) < r'(v)$.

Finally, if s and s' are defined on u_0, then

$$\begin{aligned}\operatorname{ord}(r(u), r'(u)) &= \operatorname{ord}(s(u_0), s'(u_0)) \text{ and}\\ \operatorname{ord}(r(v), r'(v)) &= \operatorname{ord}(s(v_0), s'(v_0)),\end{aligned}$$

and we conclude by induction. □

5.2.2. *The case where r and r' start with the same direction.*

Proposition 5.7. *Let $n \geq m \geq 2$, $r \in \underline{R}^{\mathsf{X}}_{m,n}$ starting with an X-letter, and $r' = \mathsf{X}_c$. If $u, v \in A^*$, r is condensed on u and v and $u \triangleright_{m,n+1} v$, then* $\operatorname{ord}(r(u), r'(u)) = \operatorname{ord}(r(v), r'(v))$. *The dual statement (involving $r \in \underline{R}^{\mathsf{Y}}_{m,n}$, $r' = \mathsf{Y}_c$ starting with a Y-letter, and $u \triangleleft_{m,n+1} v$) holds as well.*

Proof. We proceed by induction, first on m. If $m = 2$, then either $r = \mathsf{X}_{a_1} \cdots \mathsf{X}_{a_k}$ or $r = \mathsf{X}_{a_1} \cdots \mathsf{X}_{a_k} \mathsf{Y}_{b_1} \cdots \mathsf{Y}_{b_\ell}$. In the first case, the order type $\operatorname{ord}(r(u), r'(u))$ depends, as in the proof of Proposition 5.6, on whether $\mathsf{X}_c \mathsf{Y}_{a_k} \cdots \mathsf{Y}_{a_1}$ is defined on u, or if it is not defined, whether $a_k = c$ and $\mathsf{X}_c \mathsf{Y}_{a_{k-1}} \cdots \mathsf{Y}_{a_1}$ is defined. Since these rankers are in $\underline{R}^{\mathsf{X}}_{2,n+1}$ and are condensed where they are defined (Example 3.3), we have $\operatorname{ord}(r(u), r'(u)) = \operatorname{ord}(r(v), r'(v))$.

In the second case, where $r = \mathsf{X}_{a_1} \cdots \mathsf{X}_{a_k} \mathsf{Y}_{b_1} \cdots \mathsf{Y}_{b_\ell}$, three cases arise: if $r\mathsf{Y}_c$ is defined on u, then $\mathsf{X}_c(u) < r(u)$; if $r\mathsf{Y}_c$ is not defined and $c = b_\ell$, then $\mathsf{X}_c(u) = r(u)$; in all other cases, $r(u) < \mathsf{X}_c(u)$. Since $u \triangleright_{2,n+1} v$ and $r\mathsf{Y}_c \in \underline{R}^{\mathsf{X}}_{2,n+1}$, Proposition 3.10 shows that $r\mathsf{Y}_c$ is defined on u if and only if it is defined on v, and $\operatorname{ord}(r(u), r'(u)) = \operatorname{ord}(r(v), r'(v))$.

We now assume that $m \geq 3$. If r has less than m alternating blocks, we conclude by induction on m. Let us suppose now that r has m alternating blocks and let us proceed by induction on $|r| \geq m$.

Let $r = \mathsf{X}_a s$. If s starts with a Y-letter (which includes the base case where $|r| = m$), then $s \in \underline{R}^{\mathsf{Y}}_{m-1,n-1}$ is condensed on u_- and v_-. If $c \notin \mathsf{alph}(u_-) = \mathsf{alph}(v_-)$, then $r(u) < r'(u)$ and $r(v) < r'(v)$. In all other cases,

$$\begin{aligned}\operatorname{ord}(r(u), r'(u)) &= \operatorname{ord}(s(u_-), r'(u_-)) \text{ and}\\ \operatorname{ord}(r(v), r'(v)) &= \operatorname{ord}(s(v_-), r'(v_-)).\end{aligned}$$

Since $u_- \triangleright_{m,n} v_-$ by Lemma 3.7, these two order types are equal by Proposition 5.5.

If instead s starts with an X-letter, then $|r| > m$, $s \in \underline{R}^{\mathsf{X}}_{m,n-1}$ is condensed on u_+ and v_+ (Lemma 3.4) and we distinguish two cases. If $c \in \mathsf{alph}(u_- a) = \mathsf{alph}(v_- a)$, then $r'(u) < r(u)$ and $r'(v) < r(v)$. Otherwise

$$\begin{aligned}\operatorname{ord}(r(u), r'(u)) &= \operatorname{ord}(s(u_+), r'(u_+)) \text{ and}\\ \operatorname{ord}(r(v), r'(v)) &= \operatorname{ord}(s(v_+), r'(v_+)).\end{aligned}$$

Since $u_+ \triangleright_{m,n} v_+$, these two order types are equal by induction on n. □

Proposition 5.8. *Let $n \geq m \geq 2$, let $r = \mathsf{X}_a s \in \underline{R}^{\mathsf{X}}_{m}$ and $r' = \mathsf{X}_b s' \in \underline{R}^{\mathsf{X}}_{m-1}$ such that $|r| + |r'| \leq n$, and let $u, v \in A^*$ such that r and r' are condensed on u and v. If $u \triangleright_{m,n} v$, then* $\operatorname{ord}(r(u), r'(u)) = \operatorname{ord}(r(v), r'(v))$. *The dual statement (where r, r' start with Y-blocks and $u \triangleleft_{m,n} v$) holds as well.*

Proof. The proof is by induction on m, and then on n. If one of r and r' has length 1, then the result was established in Proposition 5.7. This takes care of the cases where $n \leq 3$, including the base case $m = n = 2$. We now assume that $|r|, |r'| \geq 2$.

Let us observe that under this assumption, if $n = m$, then the number of alternating blocks in r is less than or equal to $m-2$: indeed it is at most equal to $|r| \leq n-2 = m-2$. The same inequality holds for r', so this situation is handled by induction on m. We can now assume that $n > m$.

Let $u = u_- a u_+ = u'_- b u'_+$ and $v = v_- a v_+ = v'_- b v'_+$ be a-left and b-left factorizations.

First case: $a = b$. If s starts with an X-block and s' starts with a Y-block, then $r'(u) < r(u)$ and $r'(v) < r(v)$. Dually, if s starts with a Y-block and s' starts with an X-block, then $r'(u) > r(u)$ and $r'(v) > r(v)$.

If s and s' both start with a Y-block (which can happen only if $m-1 \geq 2$), then $s \in \underline{R}^{\mathsf{Y}}_{m-1}$ and $s' \in \underline{R}^{\mathsf{Y}}_{m-2}$ are condensed on u_- and v_- and

$$\begin{aligned}\mathrm{ord}(r(u), r'(u)) &= \mathrm{ord}(s(u_-), s'(u_-)) \text{ and}\\ \mathrm{ord}(r(v), r'(v)) &= \mathrm{ord}(s(v_-), s'(v_-)).\end{aligned}$$

Since $u_- \triangleleft_{m-1,n-1} v_-$ and $|s| + |s'| \leq n-2$, we have $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$ by induction on m.

If instead s and s' both start with an X-block, then $s \in \underline{R}^{\mathsf{X}}_{m}$ and $s' \in \underline{R}^{\mathsf{X}}_{m-1}$ are condensed on u_+ and v_+, and we have

$$\begin{aligned}\mathrm{ord}(r(u), r'(u)) &= \mathrm{ord}(s(u_+), s'(u_+)) \text{ and}\\ \mathrm{ord}(r(v), r'(v)) &= \mathrm{ord}(s(v_+), s'(v_+)).\end{aligned}$$

Since $u_+ \triangleright_{m,n-1} v_+$ and $|s| + |s'| \leq n-2$, we have $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$ by induction on n.

Second case: $a \neq b$, s and s' start with X-blocks. Then $s \in \underline{R}^{\mathsf{X}}_{m}$ and $s' \in \underline{R}^{\mathsf{X}}_{m-1}$ are condensed on u_+ and v_+. Without loss of generality, $\mathsf{X}_b(u) < \mathsf{X}_a(u)$, so we have $r(u) = \mathsf{X}_a s(u) = \mathsf{X}_b \mathsf{X}_a s(u) = \mathsf{X}_b r(u)$. In particular, $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(u'_+), s'(u'_+))$. By Proposition 5.7, we also have $\mathsf{X}_b(v) < \mathsf{X}_a(v)$, and hence $\mathrm{ord}(r(v), r'(v)) = \mathrm{ord}(r(v'_+), s'(v'_+))$. Since $u \triangleright_{m,n} v$, we have $u'_+ \triangleright_{m,n-1} v'_+$ and we conclude by induction on n since $|r| + |s'| \leq n-1$.

Third case: $a \neq b$, s and s' start with Y-blocks. This can occur only if $m-1 \geq 2$. Then $s \in \underline{R}^{\mathsf{Y}}_{m-1}$ and $s' \in \underline{R}^{\mathsf{Y}}_{m-2}$ are condensed on u_- and v_-, $r(u) = s(u_-)$ and $r'(u) = s'(u'_-)$, and the same equalities hold for v. Without loss of generality, we may assume that $\mathsf{X}_b(u) < \mathsf{X}_a(u)$, and hence $\mathsf{X}_b(v) < \mathsf{X}_a(v)$ (Proposition 5.7). Let u_0 and v_0 be such that $u = u'_- b u_0 a u_+$ and $v = v'_- b v_0 a v_+$: then u_0 is the left factor in the a-left decomposition of u'_+ and the right factor in the b-left decomposition of u_-. An analogous statement is true for v_0. There are two cases, depending on whether s is defined on bu_0. If this is the case, then $r'(u) < r(u)$. Moreover, we have $u'_+ \triangleright_{m,n-1} v'_+$ and $u_0 \triangleleft_{m-1,n-2} v_0$, so s is defined on bv_0 as well, by Proposition 3.10.

If instead, s is not defined on bu_0 or bv_0, let p be the longest prefix of s which is defined on bu_0 (and hence on bv_0): then p is either empty or a Y-block and $s = p\mathsf{Y}_c t$, where c has no occurrence in $u[\mathsf{X}_b(u); \mathsf{X}_a p(u) - 1]$ (so Y_c is defined on u'_-).

If $\mathsf{Y}_c t$ is defined on u'_-, then $r(u) = s(u_-) = \mathsf{Y}_c t(u'_-)$, so that

$$\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(\mathsf{Y}_c t(u_-), s'(u_-)).$$

Now $u \triangleright_{m,n} v$ implies $u'_- \triangleright_{m,n-1} v'_-$ by Proposition 3.7, so $\mathsf{Y}_c t$ is defined on v'_- and hence we have $\mathrm{ord}(r(v), r'(v)) = \mathrm{ord}(\mathsf{Y}_c t(v'_-), s'(v'_-))$ as well. Since $|\mathsf{Y}_c t| \leq |s| < |r|$, we conclude by induction that $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$.

If $\mathsf{Y}_c t$ is not defined on u'_-, then let $\mathsf{Y}_c q$ be the longest prefix of $\mathsf{Y}_c t$ which is defined on u'_- (and hence on v'_-). Then q is either empty or an X-block and $\mathsf{Y}_c t = \mathsf{Y}_c q \mathsf{X}_d t'$. If $d = b$, then $q\mathsf{X}_d(u'_- b) = \mathsf{X}_b(u)$, so $r(u) = \mathsf{X}_b t'(u)$ and similarly, $r(v) = \mathsf{X}_b t'(v)$. We conclude by induction on m that $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$ since $\mathsf{X}_b t'$ has 2 blocks less than r.

If $d \neq b$, then we have $\mathsf{X}_b(u) < \mathsf{X}_a p \mathsf{Y}_c q \mathsf{X}_d(u)$. If $\mathsf{X}_d t'$ is defined on bu_0, then $r(u)$ lies in u_0 and $r'(u)$ lies in u'_-, so $r(u) > r'(u)$. Similarly $r(v) > r'(v)$, and we are done. If instead $\mathsf{X}_d t'$ is not defined on bu_0, then $\mathsf{X}_a p \mathsf{Y}_c q(u) < \mathsf{X}_b(u)$ and $\mathsf{X}_a p \mathsf{Y}_c q \mathsf{X}_d(u) = \mathsf{X}_b \mathsf{X}_d(u)$, so the condensedness of $r = \mathsf{X}_a p \mathsf{Y}_c q \mathsf{X}_d t'$ on u implies that $\mathsf{X}_b \mathsf{X}_d t'$ is condensed on u as well. The same holds for v, and we have

$$\begin{aligned} \mathrm{ord}(r(u), r'(u)) &= \mathrm{ord}(\mathsf{X}_b \mathsf{X}_d t'(u), r'(u)) \text{ and similarly,} \\ \mathrm{ord}(r(v), r'(v)) &= \mathrm{ord}(\mathsf{X}_b \mathsf{X}_d t'(v), r'(v)). \end{aligned}$$

We conclude by induction on m since $\mathsf{X}_b \mathsf{X}_d t'$ has 2 blocks less than r.

Fourth case: $a \neq b$, s and s' start with different directions. Without loss of generality, we may assume that s starts with an X-block and s' starts with a Y-block. Since r starts with 2 X-letters, the number of alternating blocks of r is less than $|r| - 1 \leq n - 3$. Therefore if $n = m + 1$, $r \in \underline{R}^{\mathsf{X}}_{m-2}$ and $r' \in \underline{R}^{\mathsf{X}}_{m-1}$, a case that can be decided by induction on m. So we now assume that $n \geq m - 2$.

If $\mathsf{X}_b(u) < \mathsf{X}_a(u)$, then the same inequality holds in v (by Proposition 5.7) and we have $r'(u) < r(u)$ and $r'(v) < r(v)$. If instead $\mathsf{X}_a(u) < \mathsf{X}_b(u)$ and $\mathsf{X}_a(v) < \mathsf{X}_b(v)$, then the b-left factorizations of u_+ and v_+ are of the form $u_+ = u_0 b u'_+$ and $v_+ = v_0 b v'_+$.

Several cases arise, according to whether s and s' are defined (and condensed) on u_0 or not. We have $u_+ \triangleright_{m,n-1} v_+$ and $u_0 \triangleright_{m,n-2} v_0$ by Lemma 3.7. It follows as usual that s and s' are defined on v_0 if and only if they are defined on u_0. If s is not defined on u_0 then the order types $\mathrm{ord}(r(u), r'(u))$ and $\mathrm{ord}(r(v), r'(v))$ are both $>$. Therefore, from now on we can assume that s is defined on u_0 and v_0.

If s' is defined on u_0 then we can chop off $u_- a$ from u, $v_- a$ from v, and X_a from r: $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(s(u_+), r'(u_+))$ and $\mathrm{ord}(r(v), r'(v)) = \mathrm{ord}(s(v_+), r'(v_+))$. Since $u_+ \triangleright_{m,n-1} v_+$, $\mathrm{ord}(s(u_+), r'(u_+))$ and $\mathrm{ord}(s(v_+), r'(v_+))$ are equal by induction on n, and hence $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$.

If s' is not defined on u_0, then, as in the third case, we have to split the ranker s' at those points at which it crosses the position $\mathsf{X}_a(u)$. Let $\mathsf{X}_b s' = p_1 q_1 \cdots p_k q_k$ such that all p_i are defined on u_0 and all p_i are starting with an X-letter followed by a (possibly empty) Y-block. The sole exception is p_k which might contain further blocks. Moreover, each p_i is the maximal prefix of $p_i q_i \cdots p_k q_k$ which is defined on u_0. All q_i are defined on $u_- a$ and all q_i are starting with a Y-letter followed by a (possibly empty) X-block. The sole exception is q_k which might be empty or which might contain further blocks. Each q_i is the maximal prefix of $q_i p_{i+1} \cdots p_k q_k$ which is defined on $u_- a$. Since $u_- a \triangleright_{m,n-1} v_- a$ (Lemma 3.7) and $u_0 \triangleleft_{m-1,n-2} v_0$, the same definedness and maximality properties hold on $v_- a$ and v_0.

If q_k is empty, then $k \geq 2$ and p_1 and q_1 are non-empty. We see that $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(s(u_0), p_k(u_0))$ and $\mathrm{ord}(r(v), r'(v)) = \mathrm{ord}(s(v_0), p_k(v_0))$. By induction on n, we have $\mathrm{ord}(s(u_0), p_k(u_0)) = \mathrm{ord}(s(v_0), p_k(v_0))$, and hence $\mathrm{ord}(r(u), r'(u)) = \mathrm{ord}(r(v), r'(v))$.

Finally, if q_k is non-empty, then we have $r(u) > r'(u)$ and $r(v) > r'(v)$. $\square$

5.2.3. *Completing the proof of Theorem 5.1.* Let us (at last!) verify that, if $u \triangleright_{m+1,2n} v$ or $u \triangleright_{m+1,2n} v$, then Properties (**WI 2c**) and (**WI 3c**) hold for m, n. By symmetry, we simply handle the case where $u \triangleright_{m+1,2n} v$.

To verify Property (**WI 2c**), we consider rankers $r \in \underline{R}_{m,n}$ and $r' \in \underline{R}_{m-1,n-1}$ that are condensed on u and v. If both start with X-blocks, Proposition 5.8 shows that $\mathrm{ord}(r(u), r'(u))$ and $\mathrm{ord}(r(v), r'(v))$ coincide. If both start with Y-blocks, the same proposition allows us to conclude, after observing that we have $u \triangleleft_{m,2n-1} v$. And if r and r' start with different direction blocks, we conclude by Proposition 5.6.

To verify Property (**WI 3c**), we consider rankers $r \in \underline{R}_{m,n}$ and $r' \in \underline{R}_{m,n-1}$ that end with different directions, and that are condensed on u and v. If r and r' start with different direction blocks, we again conclude by Proposition 5.6. If both start with X-blocks, then they must have different number of alternations, so we have $r \in R^{\mathsf{X}}_{m_1,n_1}$ and $r' \in R^{\mathsf{X}}_{m_2,n_2}$ for some $n_1 \leq n$, $n_2 \leq n-1$ and for distinct values $m_1, m_2 \leq m$. In particular, one of m_1 and m_2 is less than or equal to $m-1$, and we can apply Proposition 5.8.

We proceed similarly if r and r' both start with Y-blocks, after observing that $u \triangleleft_{m,2n-1} v$. This completes the proof of Theorem 5.1.

6. Consequences

6.1. **Decidability results.** The main consequence we draw of Theorem 5.1 and of the decidability of the pseudovarieties $\mathbf{R}_m$ and $\mathbf{L}_m$ is summarized in the next statement.

Theorem 6.1. *Given an FO^2-definable language L, one can compute an integer m such that L is FO^2_{m+1}-definable, possibly FO^2_m-definable, but not FO^2_{m-1}-definable. That is: we can decide the quantifier alternation level of L within one unit.*

Proof. Let $L \in \mathcal{FO}^2$ and let M be its syntactic monoid. Since each pseudovariety $\mathbf{R}_m \cap \mathbf{L}_m$ is decidable (Proposition 3.23), we can compute the largest m such that $M \notin \mathbf{R}_m \cap \mathbf{L}_m$. By Theorem 5.1, $M \in \mathbf{R}_{m+1} \cap \mathbf{L}_{m+1} \subseteq \mathbf{FO}^2_{m+1}$ and hence L is FO^2_{m+1}-definable. On the other hand, $M \notin \mathbf{FO}^2_{m-1} \subseteq \mathbf{R}_m \cap \mathbf{L}_m$. □

Let us also record the following consequences of Proposition 3.23, Proposition 4.1 and the decidability of $\mathbf{R}_2 \vee \mathbf{L}_2$ (discussed in Example 5.3).

Proposition 6.2. *The classes $\underline{\mathcal{TL}}^{\mathsf{X}}_1 = \underline{\mathcal{TL}}^{\mathsf{Y}}_1 = \underline{\mathcal{TL}}_1 = \mathcal{FO}^2_1$, $\underline{\mathcal{TL}}^{\mathsf{X}}_2$, $\underline{\mathcal{TL}}^{\mathsf{Y}}_2$ and $\underline{\mathcal{TL}}_2$ are decidable.*

6.2. **Infinite and collapsing hierarchies.** The fact that the $\mathbf{R}_m$ and $\mathbf{L}_m$ form strict hierarchies (Proposition 3.23), together with Theorem 5.1, proves that the $\mathcal{FO}^2_m$ hierarchy is infinite. Weis and Immerman had already proved this result by combinatorial means [37, Theorem 4.11], whereas our proof is algebraic. From that result on the $\mathcal{FO}^2_m$ hierarchy, it is also possible to recover the strict hierarchy result on the $\mathbf{R}_m$ and $\mathbf{L}_m$ and the fact that their union is equal to $\mathbf{DA}$.

By the same token, Corollary 3.12 and Theorem 4.3 show that the $\underline{\mathcal{TL}}_m$ (resp. $\underline{\mathbf{TL}}_m$) hierarchy is infinite and that its union is all of $\mathcal{FO}^2$ (resp. $\mathbf{DA}$).

Theorem 6.3. *The hierarchies $\mathcal{FO}^2_m$ and $\underline{\mathcal{TL}}_m$ are infinite, and their union is all of $\mathcal{FO}^2$.*

Similarly, the fact (stated in Proposition 3.23) that an m-generated element of **DA** lies in $\mathbf{R}_{m+1} \cap \mathbf{L}_{m+1}$, shows that an FO^2-definable language in A^* lies in $\mathcal{R}_{|A|+1} \cap \mathcal{L}_{|A|+1}$, and hence in $\mathcal{FO}^2_{|A|+1}$ – a fact that was already established by combinatorial means by Weis and Immerman [37, Theorem 4.7]. It also shows that such a language is in $\underline{\mathcal{TL}}_{2|A|+1}$ by Theorem 4.3.

Theorem 6.4. *A language $L \subseteq A^*$ is FO^2-definable if and only if it is $\mathsf{FO}^2_{|A|+1}$-definable. And it is TL-definable if and only if it is both $\underline{\mathsf{TL}}^{\mathsf{X}}_{2|A|+1}$ and $\underline{\mathsf{TL}}^{\mathsf{Y}}_{2|A|+1}$-definable.*

Even though we arrived at Theorem 6.4 by algebraic means, it is interesting to note that its statement reflects the following combinatorial property (an idea that was already used by Weis and Immerman [37, Theorem 4.7]).

Lemma 6.5. *A ranker that is condensed on a word on alphabet A, has at most $|A|$ alternating blocks.*

Proof. Let u be a word and let r be a ranker that is condensed on u. Without loss of generality, we may assume that $r \in R^{\mathsf{X}}_{m,n}$, say

$$r = \mathsf{X}_{a_1} \cdots \mathsf{X}_{a_{k_1}} \mathsf{Y}_{a_{k_1+1}} \cdots \mathsf{Y}_{a_{k_2}} \cdots \mathsf{Z}_{a_{k_{m-1}+1}} \cdots \mathsf{Z}_{a_{k_m}}$$

with $0 < k_1 < k_2 < \cdots < k_m = n$ and $\mathsf{Z} = \mathsf{X}$ (resp. Y) if m is odd (resp. even). By definition of condensed rankers (and with the notation in that definition, see Section 3), the interval I_{k_h} is of the form $(i_{k_h-1}; \mathsf{X}_{a_{k_h}}(u, i_{k_h-1}))$ if h is odd, of the form $(\mathsf{Y}_{a_{k_h}}(u, j_{k_h-1}); j_{k_h-1})$ if h is even. In either case, $a_{k_{h+1}}$ occurs in u within the interval I_{k_h} but a_{k_h} does not. Since the intervals I_{k_h} are nested, it follows that the letters $a_{k_1}, a_{k_2}, \ldots, a_{k_m}$ are pairwise distinct, and hence $m \leq |A|$. □

6.3. **Infinite hierarchies and unambiguous polynomials.** Finally we note the following refinement of [15, Proposition 4.6]. One of the classical (and one of the earliest) results concerning the languages recognized by monoids in **DA** is the following: they are exactly the disjoint unions of unambiguous products of the form $B_0^* a_1 B_1^* \cdots a_k B_k^*$, where each B_i is a subset of A (Schützenberger [25], see also [31, 32, 6]). Recall that such a product is *unambiguous* if each word $w \in B_0^* a_1 B_1^* \cdots a_k B_k^*$ factors in a unique way as $w = u_0 a_1 u_1 \cdots a_k u_k$ with $u_i \in B_i^*$. Deterministic and co-deterministic products (see Section 3.4) are easily seen to be particular cases of unambiguous products. Propositions 3.21 and 3.23 imply the following statement.

Proposition 6.6. *The least variety of languages containing the languages of the form B^* ($B \subseteq A$) and closed under visibly deterministic and visibly co-deterministic products, is $\mathcal{FO}^2$.*

More precisely, every unambiguous product of the form $B_0^* a_1 B_1^* \cdots a_k B_k^*$, where each B_i is a subset of A, can be expressed in terms of Boolean operations and at most $|A| + 1$ alternated applications of visibly deterministic and visibly co-deterministic products – starting with a visibly deterministic (resp. co-deterministic) product.

The analogous, but weaker statement with the word *visibly* deleted was proved in [15] by algebraic means, and independently by Lodaya, Pandya and Shah using logical and combinatorial arguments [17].

Conclusion. We have related the FO^2_m hierarchy with the $\mathcal{R}_m$-$\mathcal{L}_m$ hierarchy, a hierarchy of varieties of languages which is connected with the alternation of closures under deterministic and co-deterministic products.

The varieties $\mathcal{R}_m$ and $\mathcal{L}_m$ are decidable, but the link we establish with $\mathcal{FO}^2_m$ (Theorem 5.1) is not tight enough to prove decidability of the quantifier alternation hierarchy. We recall the readers of our conjecture (Conjecture 5.4 above), according to which $\mathcal{FO}^2_m$ is equal to the intersection $\mathcal{R}_{m+1} \cap \mathcal{L}_{m+1}$. Establishing this conjecture would prove that each level of the quantifier alternation hierarchy $\mathcal{FO}^2_m$ is decidable.

Finally, we refer the reader to Straubing's result: he showed [28] that the pseudovariety $\mathbf{FO}^2_m$ is the m-th weakly iterated power of the pseudovariety $\mathbf{J}$ of $\mathcal{J}$-trivial monoids (more precisely, $\mathbf{FO}^2_1 = \mathbf{J}$ and $\mathbf{FO}^2_{m+1} = \mathbf{FO}^2_m \Box \mathbf{J}$). This result offers a different avenue to solve the decidability problem for FO^2_m-definability, and our conjecture would show the equality between two algebraic hierarchies which seem completely unrelated.

Acknowledgements

The authors gratefully acknowledge the contribution of the referees, which helped clarify certain points in the paper, suggested simpler proofs for certain technical lemmas and — especially! — pointed out a mistake in one of the proofs.

References

[1] M. Adler and N. Immerman. An $n!$ lower bound on formula size. *ACM Transactions on Computational Logic*, 4:296–314, 2003.
[2] J. Almeida. *Finite Semigroups and Universal Algebra*. World Scientific, Singapore, 1994.
[3] S. Cho and D. T. Huynh. Finite automaton aperiodicity is PSPACE-complete. *Theoretical Computer Science*, 88:96–116, 1991.
[4] V. Diekert and P. Gastin. Pure future local temporal logics are expressively complete for Mazurkiewicz traces. *Information and Computation*, 204:1597–1619, 2006. Conference version in LATIN 2004, LNCS 2976, 170–182, 2004.
[5] V. Diekert and P. Gastin. First-order definable languages. In J. Flum, E. Grädel, and Th. Wilke, editors, *Logic and Automata: History and Perspectives*, Texts in Logic and Games, pages 261–306. Amsterdam University Press, 2008.
[6] V. Diekert, P. Gastin, and M. Kufleitner. A survey on small fragments of first-order logic over finite words. *International Journal of Foundations of Computer Science*, 19:513–548, 2008.
[7] K. Etessami, M. Y. Vardi, and Th. Wilke. First-order logic with two variables and unary temporal logic. *Information and Computation*, 179(2):279–295, 2002.
[8] J.A. Gerhard. The lattice of equational classes of idempotent semigroups. *Journal of Algebra*, 15:195–224, 1970.
[9] M. Grohe and N. Schweikardt. The succinctness of first-order logic on linear orders. *Logical Methods in Computer Science*, 1:1–25, 2005.
[10] T.E. Hall and P. Weil. On radical congruence systems. *Semigroup Forum*, 59:56–73, 1999.
[11] N. Immerman. *Descriptive Complexity*. Springer, 1999.
[12] K. Krohn, J. Rhodes, and B. Tilson. Homomorphisms and semilocal theory. In M. Arbib, editor, *The Algebraic Theory of Machines, Languages and Semigroups*. Academic Press, 1965.
[13] M. Kufleitner. Polynomials, fragments of temporal logic and the variety **DA** over traces. *Theoretical Computer Science*, 376:89–100, 2007. Special issue DLT 2006.
[14] M. Kufleitner and P. Weil. On $\mathbf{FO}^2$ quantifier alternation over words. In *Mathematical Foundations of Computer Science (MFCS 2009)*, number 5734 in Lecture Notes in Computer Science, pages 513–524. Springer-Verlag, 2009.
[15] M. Kufleitner and P. Weil. On the lattice of sub-pseudovarieties of **DA**. *Semigroup Forum*, 81:243–254, 2010.

[16] M. Kufleitner and A. Lauser. Lattices of logical fragments over words. Technical Report Computer Science 2012/03, University of Stuttgart, Faculty of Computer Science, Electrical Engineering, and Information Technology, Germany, March 2012.
[17] K. Lodaya, P. K. Pandya, and S. S. Shah. Marking the chops: an unambiguous temporal logic. In *IFIP TCS 2008*, pages 461–476, 2008.
[18] R. McNaughton and S. Papert. *Counter-Free Automata*. The MIT Press, Cambridge, Mass., 1971.
[19] J.-E. Pin. Propriétés syntactiques du produit non ambigu. In W. Kuich, editor, *Proc. 7th International Colloquium Automata, Languages and Programming (ICALP'80)*, volume 85 of *Lecture Notes in Computer Science*, pages 483–499, Heidelberg, 1980. Springer-Verlag.
[20] J.-E. Pin. *Varieties of Formal Languages*. North Oxford Academic, London, 1986.
[21] J.-E. Pin. Expressive power of existential first-order sentences of Büchi's sequential calculus. *Discrete Mathematics*, 291(1-3):155–174, 2005.
[22] J.-E. Pin, H. Straubing, and D. Thérien. Locally trivial categories and unambiguous concatenation. *Journal of Pure and Applied Algebra*, 52:297–311, 1988.
[23] J. Rhodes and B. Steinberg. *The* q*-theory of finite semigroups*. Springer Monographs in Mathematics. Springer, New York, 2009.
[24] M. P. Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8:190–194, 1965.
[25] M. P. Schützenberger. Sur le produit de concaténation non ambigu. *Semigroup Forum*, 13:47–75, 1976.
[26] Th. Schwentick, D. Thérien, and H. Vollmer. Partially-ordered two-way automata: A new characterization of DA. In W. Kuich, G. Rozenberg, and A. Salomaa, editors, *Proc. of the 5th Int. Conf. on Developments in Language Theory (DLT)*, volume 2295 of *Lecture Notes in Computer Science*, pages 239–250. Springer, 2001.
[27] H. Straubing. *Finite Automata, Formal Logic, and Circuit Complexity*. Birkhäuser, Boston, Basel and Berlin, 1994.
[28] H. Straubing. Algebraic characterization of the alternation hierarchy in $\mathrm{FO}^2[<]$ on finite words. In M. Bezem, editor, *Proc. Computer Science Logic (CSL'11)*, volume 12 of *LIPIcs*, pages 525–537. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2011.
[29] H. Straubing and P. Weil. An introduction to automata theory. In D. D'Souza and P. Shankar, editors, *Modern applications of automata theory*, volume 2 of *I.I.Sc. Monographs*, pages 3–43. World Scientific, 2012.
[30] H. Straubing and P. Weil. Varieties. In J.-E. Pin, editor, *Handbook of Finite Automata*. European Math. Society, to appear.
[31] P. Tesson and D. Thérien. Diamonds are forever: The variety DA. In G. M. Gomes Moreira Da Cunha, P. V. Silva, and J.-E. Pin, editors, *Semigroups, Algorithms, Automata and Languages, Coimbra (Portugal) 2001*, pages 475–500. World Scientific, 2002.
[32] P. Tesson and D. Thérien. Logic meets algebra: The case of regular languages. *Logical Methods in Computer Science*, 3(1):1–37, 2007.
[33] D. Thérien and Th. Wilke. Over words, two variables are as powerful as one quantifier alternation. In *STOC*, pages 234–240, 1998.
[34] W. Thomas. Classifying regular events in symbolic logic. *Journal of Computing Systems and Science*, 25:360–376, 1982.
[35] P. Trotter and P. Weil. The lattice of pseudovarieties of idempotent semigroups and a non-regular analogue. *Algebra Universalis*, 37:491–526, 1997.
[36] Ph. Weis and N. Immerman. Structure theorem and strict alternation hierarchy for FO^2 on words. In J. Duparc and Th. A. Henzinger, editors, *Proc. Computer Science Logic (CSL 2007)*, volume 4646 of *Lecture Notes in Computer Science*, pages 343–357. Springer, 2007.
[37] Ph. Weis and N. Immerman. Structure theorem and strict alternation hierarchy for FO^2 on words. *Logical Methods in Computer Science*, 5:1–23, 2009.