Some Logical Characterizations of the Dot-Depth Hierarchy and Applications

By: Francine Blanchet-Sadri

F. Blanchet-Sadri, "Some Logical Characterizations of the Dot-Depth Hierarchy and Applications." Journal of Computer and System Sciences, Vol. 51, No. 2, 1995, pp 324-337.

Abstract:
A logical characterization of natural subhierarchies of the dot-depth hierarchy refining a theorem of Thomas and a congruence characterization related to a version of the Ehrenfeucht—Fraïssé game generalizing a theorem of Simon are given. For a sequence $\bar{m} = (m_1, \ldots, m_k)$ of positive integers, subclasses $\mathcal{L}(m_1, \ldots, m_k)$ of languages of level k are defined. $\mathcal{L}(m_1, \ldots, m_k)$ are shown to be decidable. Some properties of the characterizing congruences are studied, among them, a condition which insures $\mathcal{L}(m_1, m_k)$ to be included in $\mathcal{L}(m'_1, \ldots, m'_k)$. A conjecture of Pin concerning tree hierarchies of monoids (the dot-depth being a particular case) is shown to be false.

Article:

I. INTRODUCTION

Traditionally, algebraic automata theory uses monoids as models for finite state machines. One looks at a finite state machine as processing sequences of symbols drawn from a finite input alphabet. Denoting the input alphabet by A, the universe of possible inputs is the free monoid A^* and a finite state machine can be thought of as a quotient of A^* by a finite index congruence $\sim$. $A^* / \sim$ being a finite monoid, one is then led to investigate relationships between the structure of this algebraic system and the combinatorial processing of input sequences. The theory of varieties of Eilenberg constitutes an elegant framework for discussing these relationships between combinatorial descriptions of languages and algebraic properties of their recognizers. The interplay between the two points of view leads to interesting classifications of languages and finite monoids.

Let A be a given alphabet. The regular, or recognizable, languages over A are those subsets of A^* constructed from the finite languages over A by the boolean operations as well as the concatenation product and the star. The star-free languages consist of those regular languages which can be obtained from the finite languages by boolean operations and the concatenation product only. According to a fundamental theorem of Schiitzenberger [25], $L \subseteq A^*$ is star-free if and only if its syntactic monoid $M(L)$ is finite and aperiodic, that is, $M(L)$ contains only trivial subgroups. For example, $(ab)^*$ is star-free since $(ab)^* = ((aA^* \cap A^*b) \setminus (A^*aaA^* \cup A^*bbA^*)) \cup \{1\}$, where 1 is the empty word. But $(aa)^*$ is not star-free, a consequence of the theorem of Schfizenberger. General references on the star-free languages are McNaughton and Papert [19], Eilenberg [11], or Pin [21].

Natural classifications of the star-free languages are obtained based on the alternating use of the boolean operations and the concatenation product. Let $A^+ = A^* / \{1\}$. Define $A^+ \mathcal{B}_0 = \{ L \subseteq A^+ \mid L$ is finite or cofinite$\}$, $A^+ \mathcal{B}_{k+1} = \{ L \subseteq A^+ \mid L$ is a boolean combination of languages of the form $L_1 \ldots L_n$ $(n \geq 1)$ with $L_1 \ldots, L_n \in A^+ \mathcal{B}_k\}$. For technical reasons, only nonempty words over A are considered to define this hierarchy; in particular, the complement operation is applied with respect to A^+ . The language classes $A^+ \mathcal{B}_0$, $A^+ \mathcal{B}_1$, ... form the so-called dot-depth hierarchy introduced by Cohen and Brzozowski [9]. The union of the classes $A^+ \mathcal{B}_0$, $A^+ \mathcal{B}_1$... is the class of star-free languages.

Most of our attention will be directed toward a closely related hierarchy, this one in A^*. It was introduced by Straubing [28]. Let $A^*\mathcal{L}_0 = \{\emptyset, A^*\}$, $A^*\mathcal{L}_{k+1} = \{ L \subseteq A^* \mid L$ is a boolean combination of languages of the form $L_0a_1 L_1a_2 \dots a_nL_n$ $(n \geq 0)$ with $L_0, \dots, L_n \in A^* \mathcal{L}_k$ and $a_1, \dots, a_n \in A\}$. Let $A^*\mathcal{L} = \bigcup_{k \geq 0} A^*\mathcal{L}_k$ is star-free if and only if $L \in A^*\mathcal{L}_k$ for some $k \geq 0$. The dot-depth of L is the smallest such k. The Straubing hierarchy appears to be the more fundamental of the two for reasons explained in [29]. For details concerning the Straubing hierarchy and its relation to the dot-depth hierarchy, see Pin [21 or 22].

In the framework of semigroup theory, Brzozowski and Knast [6] showed that the dot-depth hierarchy is infinite, in fact, that $A^*\mathcal{B}_{k+1} \supsetneq A^+ \mathcal{B}_k$ for $k \geq 0$. Thomas [31] gave a new proof of this result, which shows also that the Straubing hierarchy is infinite, based on a logical characterization of the dot-depth hierarchy that he obtained in [30]. His proof does not rely on semigroup theory; instead, an intuitively appealing model-theoretic technique was applied: the Ehrenfeucht-Fraisse game.

It was the work of Büchi [8] and Elgot [12] that first showed how to use certain formulas of mathematical logic in order to describe properties of regular languages. These formulas (known as monadic second-order formulas) are built up from variables $x, y, \dots$, set variables $X, Y, \dots$, a 2-place predicate symbol < and a set $\{ Q_a \mid a \in A\}$ of 1-place predicate symbols in one-to-one correspondence with the alphabet A. Starting with atomic formulas of the form $x < y$, Q_ax , Xx, and $x = y$, formulas are built up in the usual way by means of the connectives $\neg$, $\vee$, $\wedge$ and the quantifiers $\exists$ and $\forall$ binding up both types of variables. A word w on A satisfies a sentence φ if φ is true when variables are interpreted as integers, set variables as sets of integers, the predicate < as the usual relation on integers and the formula Q_ax as the letter in position x in w is an a.

Ladner [16] and McNaughton [18] were the first to consider the case where the set of formulas is restricted to first-order, that is, when set variables are ignored. They proved that the languages defined in this way are precisely the star-free languages.

Thomas [30] showed that the dot-depth hierarchy corresponds in a very natural way with a classical hierarchy of first-order logic based on the alternation of existential and universal quantifiers. Perrin and Pin [20] gave a substantially different proof of the result of Thomas for the Straubing hierarchy.

For each $k \geq 0$, there is a variety V_k of finite monoids, or M-variety, such that for $L \subseteq A^*$, $L \in A^*\mathcal{V}_k$" if and only if $M(L) \in V_k$. An outstanding open problem is whether one can decide if a star-free language has dot-depth k; this is equivalent to the question "is V_k decidable?," i.e., does there exist an algorithm which enables us to test if a finite monoid is or is not in V_k? The variety V_0 consists of the trivial monoid alone. The variety V_1 consists of all finite $\mathcal{J}$-trivial monoids [26]. Straubing [29] conjectured an effective criterion, based on the syntactic monoid of the language, for the case $k = 2$. His condition is shown to be necessary, in general, and sufficient in an important special case, i.e., for an alphabet of two letters. The condition is formulated in terms of a novel use of categories in semigroup theory, recently developed by Tilson [32].

This paper is concerned with applications of some logical characterizations of the Straubing hierarchy. The aim of Section 2 is to give those logical characterizations of the star-free languages. They are useful in attacking the decidability question. A logical characterization of natural subhierarchies of the Straubing hierarchy refining the logical characterizations of Thomas is given. As an application we can get upper bounds on the dot-depth of star-free languages by considering their descriptions in the first-order logical language. We state the version of the EhrenfeuchtFraisse game which was used in [31] to prove that the Straubing hierarchy is infinite. Then we give a characterization of the star-free languages in terms of congruences defined in that paper generalizing a result of Simon. A characterization of the varieties of monoids related to the Straubing hierarchy through Eilenberg's correspondence is stated. For a sequence $\overline{m} =(m_1,\dots,m_k)$ of positive integers, subclasses $\mathcal{L}(m_1, \dots, m_k)$ of languages of level k are defined.

In Section 3, we study some properties of the characterizing congruences. This section establishes an induction lemma and a condition which ensures $\mathcal{L}(m_1, ..., m_k)$ to be included in $\mathcal{L}(m'_1, ..., m'_{k'})$

Section 4 deals with a first application of the above logical characterizations. We show that a conjecture of Pin concerning tree hierarchies of monoids (the Straubing hierarchy being a particular case) is false. Decidability and inclusion problems are discussed. $\mathcal{L}(m_1, ..., m_k)$ are shown to be decidable. Other applications of the above logical characterizations are subjects of [l-5]. The study of properties of the characterizing congruences and equation systems for the varieties of monoids corresponding to the levels of the Straubing hierarchy are closely related.

In the following, φ will be called a Σ_k-formula if $\varphi = (Q\bar{x})\ \psi$, where ψ is quantifier-free and where $(Q\bar{x})$ is a string of k alternating blocks of quantifiers such that the first block contains only existential ones. Similarly, if $(Q\bar{x})$ consists of k blocks beginning with a block of universal quantifiers, $(Q\bar{x})\ \psi$ is a Π_k-formula. A $B(\Sigma_k)$-formula will denote a boolean combination of Σ_k-formulas. If ~ is a congruence on A^*, the set of all ~-classes will be denoted by $A^*/\sim$. If $L \subseteq A^*$ is a union of ~-classes, we will say that L is a ~-language. All the semigroups considered in this paper are finite (except for free semigroups and free monoids). We refer the reader to the books by Eilenberg [11], Lallement [17], Pin [21], and Enderton [13] for all the other algebraic and logical terms not defined in this paper.

2. SOME LOGICAL CHARACTERIZATIONS OF THE STRAUBING HIERARCHY

2.1. *A Quantifier Complexity Characterization*

Let us first state the logical characterization of the Straubing hierarchy mentioned by Thomas. One identifies any word $w \in A^*$, say of length $|w|$, with a word model $w = <\{1, ..., w\}, <^w, (Q_a^w)_{a \in A}>$, where the universe $\{1, ..., |w|\}$ represents the set of positions of letters in the word w, $<^w$ denotes the < -relation in w, and Q_a^w are unary relations over $\{1, ..., |w|\}$ containing the positions with letter a for each $a \in A$. Sometimes it is convenient to assume that the position sets of two words u, v are disjoint; then one takes any two nonoverlapping segments of the integers as the position sets of u and v. Let $\mathcal{L}$ be the first-order language with equality and nonlogical symbols <, Q_a, $a \in A$. Then the satisfaction of $\mathcal{L}$-sentence φ in a word w, written $w \vDash \varphi$, is defined in a natural way, and we say that $L \subseteq A^*$ is defined by the $\mathcal{L}$-sentence φ if $L= L(\varphi)$— $\{w \in A^* \mid w \vDash \varphi\}$. We also consider the formulas 0 (false) and 1 (true). Observe that $L(0) = \emptyset$ and $L(1)= A^*$.

THEOREM 2.1 (Thomas [30]). *A language $L \subseteq A^*$ belongs to $A^*\mathcal{V}_k$ if and only if L is defined by a $B(\Sigma_k)$-sentence of $\mathcal{L}$.*

COROLLARY 2.1 (Ladner [16] and McNaughton [18]). *A language L is star-free if and only if there exists a first-order $\mathcal{L}$-sentence φ such that $L = L(\varphi)$.*

For $k \geq 1$, let us define subhierarchies of $A^*\mathcal{V}$ as follows: for all $m \geq 1$, let $A^*\mathcal{V}_{k,m} = \{ L \subseteq A^* \mid L$ is a boolean combination of languages of the form $L_0a_1L_1a_2 \dots a_nL_n$ $(0 \leq n \leq m)$ with $L_0, \dots, L_n \in A^*\mathcal{V}_{k-1}$ and $a_1, \dots, a_n \in A\}$. We have $A^*\mathcal{V}_k = \bigcup_{m \geq 1} A^*\mathcal{V}_{k,m}$. Easily, $A^*\mathcal{V}_{k,m} \subseteq A^*\mathcal{V}_{k+1,m}$, $A^*\mathcal{V}_{k,m+1}$. Similarly, subhierarchies of $A^+\mathcal{B}_k$ can be defined. One can show that $\mathcal{V}_{k,m}$ is a *-variety of languages. Let the corresponding M-varieties be denoted by $\mathcal{V}_{k,m}$. We have that for $k \geq 1$, $m \geq 1$, $L \in A^*\mathcal{V}_{k,m}$ if and only if $M(L) \in \mathcal{V}_{k,m}$.

In $A^+\mathcal{B}_1$ several hierarchies and classes of languages have been studied; the most prominent examples are the β-hierarchy [7], also called depth-one finite cofinite hierarchy, and the class of locally testable languages. In Thomas [30] it was shown that both are characterized by natural restrictions on the form of Σ_1 -sentences of a certain first-order language extending $\mathcal{L}$.

The purpose of this subsection is to give a logical characterization, which follows from an analysis of the proof of Theorem 2.1, of the subhierarchies of $A^*\mathcal{V}$ refining the theorem of Thomas. It will be useful to extend $\mathcal{L}$ by adding constant symbols s, for every natural number s. For a word model w, the interpretation s^w of s will be the

sth element of w. Let $\varphi(x_1, \ldots, x_m)$ be a formula in which $x_1, \ldots, x_m$ are the unique free variables. Let $s_1, \ldots, s_m$ be positive integers. The meaning and usage of $\varphi(s_1, \ldots, s_m)$ should be quite clear in what follows. $\varphi(s_1, \ldots, s_m)$ is obtained from $\varphi(x_1, \ldots, x_m)$ by replacing simultaneously all free occurrences of x_1 in φ by the constant $s_1, \ldots, x_m$ by s_m. The interpretation of the formula $\varphi(\bar{x}) = \varphi(x_1, \ldots, x_m)$ in a word model w with universe $\{1, \ldots, |w|\}$ and elements $s_1, \ldots, s_m \in \{1, \ldots, |w|\}$ is defined in the natural way; we write $w \vDash \varphi(s_1, \ldots, sm)$ if φ is satisfied in w it when interpreting x_i by s_i for $1 \leq i \leq m$.

A logical characterization of the subhierarchies of $A^*\mathcal{V}$ is based on the following two lemmas. In what follows, if $w = a_1 \ldots a_n$ is a word and $1 \leq s \leq s' \leq n$, $w[s, s']$, $w(s, s')$, $w(s, s']$, and $w[s, s')$ will denote respectively the segments $a_s \ldots a_{s'}$, $a_{s+1} \ldots a_{s'}$, $_1$,$a_{s+1} \ldots a_{s'}$ *and as* $\ldots a_{s'}$ $_1$.

LEMMA 2.1 (Perrin and Pin [20]). *For* $k \geq 0$ *and for each* $B(\Sigma_k)$*-sentence* φ, *there exist* $B(\Sigma_k)$*-formulas* $\varphi_1(x)$, $\varphi_r(x)$, $\varphi_m(x, y)$ *in which x (x, y) is (are) the unique free variable(s) and such that for every n and for every word w of length n* we *have*

1. $w \in L(\varphi_l(s))$ *if and only if* $w[1, s) \in L(\varphi)$, *and*

2. $w \in L(\varphi_r(s))$ *if and only if* $w(s, n] \in L(\varphi)$ *for every integer s such that* $1 \leq s \leq n$, *and*

3. $w \in L(\varphi_m(s, s'))$ *if and only* $w(s, s') \in L(\varphi)$ *for every integers s, s' such that* $1 \leq s < s' \leq n$.

Proof. We define φ_m for every formula φ. φ_m is constructed by induction as follows (the constructions are similar for φ_1 and φ_r): if φ is quantifier-free, then $\varphi_m = \varphi$. Otherwise, we set $(\exists z\varphi)_m = \exists z((x < z < y \wedge \varphi_m)$, $(\forall z\varphi)_m = \forall z((x < z < y \rightarrow \varphi_m)$, $(\varphi \vee \psi)_m = \varphi_m \vee \psi_m$, $(\neg\varphi)_m = \neg\varphi_m$, $(\varphi \wedge \psi)_m = \varphi_m \wedge \psi_m$. Then one can verify by induction on $k \geq 0$ the following properties:

- *if* φ and ψ are equivalent formulas, then φ_m and ψ_m are equivalent;
- if φ is $B(\Sigma_k)$, then φ_m is equivalent to a $B(\Sigma_k)$-formula;
- let φ be a sentence. If $|w| = n$ and if $1 \leq s < s' \leq n$, w satisfies $\varphi_m(s, s')$ if and only if $w(s, s')$ satisfies φ.

LEMMA 2.2. *Given a* $B(\Sigma_k)$*-formula* $\varphi(x_1, \ldots, x_n)(n \geq 1)$, *there is a system* $\langle \bar{L}^j \rangle_{j<p}$ *of sequences* $\bar{L}^j = \langle L_0^j, \ldots, L_n^j \rangle$ *of languages* $L_i^j \in A^*\mathcal{V}_k$ *and* $\langle \bar{a}^j \rangle_{j<p}$ *of sequences* $\bar{a}^j = \langle a_1^j, \ldots, a_n^j \rangle$, $a_i^j \in A$ *such that for any w and* $s_1 < \ldots < s_n$ *in* $\{1, \ldots, |w|\}$, $w \vDash \varphi(s_1, \ldots, s_n)$ *if and only if there is* $j < p$ *such that*

1. $w[1, s1) \in L_0^j$ *and* $Q_{a_1^j}^w\ s_1$,

2. $w(s_i, s_{i+1}) \in L_i^j$ *and* $Q_{a_{1-1}^j}^w\ s_{i+1}$, $1 \leq i < n$,*and*

3. $w(s_n, |w|\,] \in L_n^j$.

Proof: By induction on k (see the proof of Theorem 2.1 [30] . If $n = 0$, this is just Theorem 2.1).

Let φ be an $\mathcal{L}$-sentence. If φ is a boolean combination of the Σ_k -sentences $\varphi_1, \ldots, \varphi_n$, define the *quantifier rank* $q_r(\varphi)$ to be the maximum number of quantifiers occurring in the leading block of one of the formulas $\varphi_1, \ldots, \varphi_n$. Let us now prove a refinement of Thomas' theorem.

THEOREM 2.2. *Let $k \geq 1$, $m \geq 1$. A language $L \subseteq A^*$ is defined by a $B(\Sigma_k)$-sentence* of $\mathcal{L}$, φ, *where* qr($(\varphi) \leq m$ *if and only if L belongs to* $A^*\mathcal{V}_{k,m}$.

Proof. The case $k = 1$ is the following. Let $m \geq 1$. Let L be a language of the form $A^*a_1A^*a_2 \ldots a_mA^*$, where $a_i \in A$, $i = 1,\ldots, m$. We have to find a boolean combination of Σ_1 -sentences defining L such that qr$(\phi) \leq m$. The assertion it $w \in L$ can be expressed by a Σ_1-sentence as follows: $\exists x_1 \exists x_2 \ldots \exists x_m (x_1 < x_2 < \ldots < x_m \wedge Q_{a_1}x_1 \wedge \ldots \wedge Q_{a_m}x_m)$. Hence L is defined by a sentence of the required form.

Conversely, we show that a given Σ_1-sentence $\exists x_1 \ldots \exists x_m\varphi(\bar{x})$ • • • ax,,,g9(,t) defines a language in $A^*\mathcal{V}_{1,m}$,where $\varphi(\bar{x})$ is equivalent to a conjunction of atomic formulas of the form Q_ax, $x < y$ or $x = y$ (for x, y variables and $a \in A$) or their negation. Let ord$_1(\bar{x})$, …, ord$_r(\bar{x})$ be the conjunctions saying $x_{i_1} \leq \ldots \leq x_{i_m}$, where $\{i_1 \ldots i_m\} = \{1, \ldots, m\}$. Then $\exists\bar{x}\varphi(\bar{x})$ is equivalent to $\bigvee_{1 \leq i \leq r} \exists\bar{x}(\text{ord}_i(\bar{x}) \wedge \varphi(\bar{x}))$. Let us consider a typical member of this disjunction, say $\exists\bar{x}(x_1 < \ldots < \text{xm} \wedge \varphi(\bar{x}))$ (identify variables if equalities occur between the x_i's). It suffices to show that the language L defined by $\psi = \exists\bar{x}(x_1 < \ldots < x_m \wedge \varphi(\bar{x}))$ is in $A^*\mathcal{V}_{1,m}$. But ψ defines either Ø or is equivalent to a disjunction of formulas of the form $\exists\bar{x}(x_1 < \ldots < x_m \wedge \varphi^1(\bar{x}))$, where $\varphi^1(\bar{x})$ is a conjunction of atomic formulas of the form Q_ax, $\neg Q_ax$ for x a variable and $a \in A$. In either case, L is seen to belong to $A^*\mathcal{V}_{1,m}$. For example, $L(\exists xQ_ax) = A^*aA^*$, $L(\exists x\neg Q_ax) = \bigcup_{b \in A, b \neq a} A^*bA^*$, $L(\exists y \exists z\, (y < z \wedge Q_ay \wedge Q_bz)) = A^*aA^*bA^*$ and $L(\exists y\exists z(\neg(y < z) \wedge Q_ay \wedge \neg Q_bz)) = L(\exists y(Q_ay \wedge \neg Q_by)) \cup L(\exists y\exists z(z < y \wedge Q_ay \wedge \neg Q_bz))$.

Now let us assume that $k >1$, *in* 1. Let L be a language of the form $L_0a_1L_1a_2 \ldots a_mL_m$, where $a_i \in A^*\mathcal{V}_{k-1}$, $i=0,\ldots,m$. We have to find a boolean combination φ of Σ_k -sentences defining L such that qr$(\varphi) \leq m$. By Theorem 2.1, let $\varphi^0, \varphi^1, \ldots, \varphi^m$ be $B(\Sigma_{k-1})$-sentences defining $L_0, L_1, \ldots, L_m$, respectively. We can find $B(\Sigma_{k-1})$-formulas $\varphi_1^0(x)$, $\varphi_m^1(x, y)$, $\varphi_m^2(x, y)$, …, $\varphi_r^m(x)$ satisfying Lemma 2.1. Hence the assertion $w \in L$ can be expressed by the following sentence: $\exists x_1 \exists x_2 \ldots \exists x_m (x_1 < x_2 < \ldots < x_m \wedge Q_{a_1}x_1 \wedge Q_{a_2}x_2 \wedge \ldots \wedge Q_{a_m}x_m \wedge \varphi_1^0(x_1, x_2) \wedge \varphi_m^2(x_2, x_3) \wedge \ldots \wedge Q_r^m(x_m))$, which is equivalent to a $B(\Sigma_k)$-sentence of the required form since $(x_1 < \ldots < x_m \wedge Q_{a_1}x_1 \wedge \varphi_r^m(x_m))$ is equivalent to a $B(\Sigma_{k\,1})$-formula or a Π_{k-1}-formula.

Conversely, consider a Σ_k-sentence $\exists x_1 \ldots \exists \text{xm}\varphi(\bar{x})$, where $\varphi(\bar{x})$ is a $B(\Sigma_{k-1})$-formula. As in the proof of the case $k = 1$, $m \geq 1$, it suffices to consider a Σ_k -sentence of the form $\psi = \exists x_1 \ldots \exists x_m (x_1 < \ldots < x_m \wedge \varphi(\bar{x}))$. Then, by Lemma 2.2, there is a system $\langle \bar{L}^j \rangle_{j<p}$ of sequences $\bar{L}^j = \langle L_0^j, \ldots, L_m^j \rangle$ of languages $L_i^j \in A^*\mathcal{V}_{k-1}$ and $\langle \bar{a}^j \rangle_{j<p}$ of sequences $\bar{a}^j = \langle a_1^j, \ldots, a_m^j \rangle$, $a_i^j \in A$ such that for any w and $s_1 < \bullet\bullet\bullet < s_m$ in $\{1, \ldots, |w|\}$, $w \models \varphi(s_1 \ldots, s_m)$ if and only if there is $j<p$ such that $w \in L_0^j a_1^j L_1^j a_2^j \ldots a_m^j L_m^j$. But for every $j<p$, $L_0^j a_1^j L_1^j a_2^j 4 \ldots a_m^j L_m^j \in A^*\mathcal{V}_{k,m}$. Hence ψ defines a boolean combination of languages of the required form and the proof is complete.

2.2. A Congruence Characterization Related to a Version of the Ehrenfeucht-Fraissé Game

Thomas [31], in order to show that the dot-depth hierarchy is infinite, defined some congruences which we state after describing the version of the Ehrenfeucht-Fraisse game which was used in his proof. Those congruences will be shown to characterize the star-free languages. The next three paragraphs restate [31].

First we define what we mean by $\bar{m}$-formulas of $\mathcal{L}$. For a sequence $\bar{m} = (m_1, \ldots, m_k)$ of positive integers, where $k \geq 0$, let length$(\bar{m}) = k$ and sum$(\bar{m}) = m_1 + \ldots + m_k$. The set of $\bar{m}$-formulas of $\mathcal{L}$ is defined by induction on length$(\bar{m})$: if length$(\bar{m}) = 0$, it is the set of quantifier-free $\mathcal{L}$-formulas; and for $\bar{m} = (m, m_1, \ldots, m_k)$, an $\bar{m}$-formula is a boolean combination of formulas $\exists x_1 \ldots \exists x_m\varphi$, where φ is an $(m_1, \ldots m_k)$-formula. We write $u \equiv_m v$ if u and v satisfy the same $\bar{m}$-sentences of $\mathcal{L}$. For $\bar{m} = (m_1, \ldots, m_k)$, the $\bar{m}$-formulas of $\mathcal{L}$ are seen to be $B(\Sigma_k)$-formulas φ such that qr$(\varphi) \leq m_1$. Moreover, languages in $A^*\mathcal{V}_{k,m}$ are defined by $(m, m_2, \ldots, m_k)$-formulas for some m_i, $i = 2, \ldots, k$ and m. The following game $\mathcal{G}\bar{m}(u, v)$ is useful for showing $\equiv_m$-equivalence.

The game $\mathcal{G}\bar{m}(u, v)$, where $\bar{m} = (m_1, \ldots, m_k)$, is played between two players *I* and *II* on the word models u and v. A play of the game consists of k moves. In the ith move, player *I* chooses, in u or in v, a sequence of m_i positions; then player *II* chooses, in the remaining word (v or u), also a sequence of m_i positions. Before each move, player I has to decide whether to choose his next elements from *u* or from v. After *k* moves, by concatenating the position sequences chosen from *u* and chosen from *v*, two sequences $\bar{p} = p_1 \ldots p_n$ from *u* and $\bar{q} = q_1 \ldots q_n$ from *v* have been formed, where $n = \text{sum}(\bar{m})$. Player *II* has won the play if the map $p_i \to q_i$ respects $<$ and the predicates Q_a, $a \in A$ (i.e., $p_i <^u p_j$ if and only if $q_i <^v q_j$, $Q_a^u p_i$ if and only if $Q_a^v q_i$, $a \in A$ for $1 \leq i, j \leq n$). Equivalently, the two subwords in *u* and *v* given by the position sequences $\bar{p}$ and $\bar{q}$ should coincide. If there is a winning strategy for *II* in the game to win each play we say that player *II* wins $\mathcal{G}\bar{m}(u, v)$ and write $u \sim_{\bar{m}} v$; $\sim_{\bar{m}}$ naturally defines a congruence on A^* which we will denote also by $\sim_{\bar{m}}$

The standard Ehrenfeucht-Fraissé game is the special case of $\mathcal{G}\bar{m}(u, v)$, where $\bar{m} = (1, ..., 1)$. For a detailed discussion see Rosenstein [24] or Fraissé [14]. If length($\bar{m}$) = *k* and $\bar{m}$ = (1, ..., 1) we write $\mathcal{G}k(u, v)$ instead of $\mathcal{G}\bar{m}(u, v)$ and *u v* instead of *u* -,, v. Note that in this case the Wi-formulas are up to equivalence just the formulas of quantifier depth *k* (Remark. One should not confuse *.(§k(u*, v) and
(k)(u, v); a play of the game *k(u, r)* consists of *k* moves but a play of the game *(k)(u, v)* of 1 move). We have the following important.

THEOREM 2.3 (Ehrenfeucht and Fraissé [10]). *For all* $\bar{m} = (m_1, \ldots, m_k)$ *with* $k > 0$ *and* $m_i > 0$ *for* $i = 1, \ldots k$, *we have* $u \equiv_m v$ *if and only i f* $u \sim_{\bar{m}} v$.

Simon [26] calls $\sim_{(\bar{m})}$-languages piecewise testable languages. They constitute level 1 of the Straubing hierarchy. The purpose of this subsection is to characterize similarly the hierarchy, each level of it and also each subhierarchy.

To do so, we use Theorem 2.1 and Theorem 2.2 and follow the technique used in [30]. For a word *w*, we can define, by induction on length($\bar{m}$), a sentence φ_w^m which in a certain sense guarantees the satisfaction of all $\bar{m}$-sentences of $\mathcal{L}$ which are satisfied by *w*. We have the following.

LEMMA 2.3. 1. $W \vDash \varphi_w^m$.

2. φ_w^m *is equivalent to a* $\bar{m}$*-sentence.*

3. *For all w and u, if* $u \vDash \varphi_w^m$ *then every* $\bar{m}$*-sentence satisfied in w is also satisfied in u.*

We can now prove the following.

THEOREM 2.4. *L is star-free if and only if L is* $\sim_m$*-language for some in.*

Proof: If $L = \emptyset$, then *L* is an empty union of classes of some congruence $\sim_m$. If $L = A^*$, *L* can be taken as the union of all classes of some congruence $\sim_m$. Hence consider $L \in A^* \mathcal{V}_k$ for some $k \geq 1$. Then by Theorem 2.1 *L* is defined by a $B(\Sigma_k)$-sentence of $\mathcal{L}$, or a $\bar{m}$-sentence of $\mathcal{L}$, φ, for some $\bar{m} = (m_1, \ldots, m_k)$. Hence $L = L(\varphi) = \{w \in A^* \mid w \vDash \varphi\}$. Let us show that $\sim_m \subseteq \sim_L$ (here, $x \sim_L y$ if and only if for all $u, v \in A^*$, $uxv \in L$, if and only if $uyv \in L$ and $\sim_L$ is the congruence of minimal index with the property that *L* is a ~-language). Let $u, v \in A^*$. Suppose that $u \sim_m v$. Suppose that $xuy \in L$. $\sim_m$ being a congruence, we have that $xuy \sim_m xvy$. We have assumed that $xuy \in L$, which means that $xuy \vDash \varphi$. We want to show that $xvy \vDash \varphi$. But by Theorem 2.3, we get $xuy \equiv_m xvy$, which means that *xuy* and *xvy* satisfy the same $\bar{m}$-sentences of $\mathcal{L}$, φo being a $\bar{m}$-sentence, we get that, since $xuy \vDash \varphi$, $xvy \vDash \varphi$. Hence, $xvy \in L$. Similarly, we show that $xvy \in L$ implies that $xuy \in L$. Hence $u \sim_L v$. Since $\sim_m \subseteq \sim_L$ we have that *L* is a $\sim_m$-language.

Let *L* be a $\sim_m$-language for some $\bar{m}$. Then *L* is a union of classes of the congruence $\sim_m$. $\sim_m$ being a finite index equivalence relation (see Rosenstein [24]), it has only finitely many equivalence classes. Let $w_1, \ldots, w_m$ be a set of representatives. In order to show that *L* is star-free, it suffices to show that $[w_i]_{\sim_m}$ is star-free for $w_i \in L$. $\varphi_{w_i}^m$ denotes the conjunction of all $\bar{m}$-sentences of $\mathcal{L}$ satisfied by w_i. Note that, since there are only finitely many atomic and negated atomic formulas in the language, the conjunction will be of bounded length. We will show that $[w_i]_{\sim_m}$ is defined by . $\varphi^m_{w_i}$, and that . $\varphi^m_{w_i}$ being a first-order sentence, using Corollary 2.1, we will get the result. If $v_{\sim_m}$ w_i, then using Theorem 2.3, we get v $\equiv_m w_i$, implying by Lemma 2.3(1) and (2) that $v \vDash \varphi^m_{w_i}$. Now let $v \vDash \varphi^m_{w_i}$. Let us show that $v_{\sim_m}$ w_i. By Theorem 2.3, we have to show that v and w_i satisfy the same $\bar{m}$-sentences. Let φ be a $\bar{m}$-sentence such that wi $\vDash \varphi$. Since by hypothesis v $\vDash \varphi^m_{w_i}$, using Lemma 2.3(3) we get $v \vDash \varphi$. Now, let φ be a $\bar{m}$-sentence such that $v \vDash \varphi$. Choose the unique j with $w_j \sim_m v$ and suppose that $j \neq i$. By Theorem 2.3, we get wj $\vDash \varphi$. Since wj $\nsim_m w_i$, there are two cases which can happen.

Case 1. There is a $\bar{m}$-sentence ψ such that $w_j \vDash \psi$, $w_i \nvDash \psi$. Since $w_i \sim_m v$ we get $v \vDash \psi$. From $v \vDash \varphi^m_{w_i}$, we get V $\vDash \neg\, \psi$. Contradiction.

u

Case 2. There is a $\bar{m}$-sentence ψ such that $w_i \vDash \psi$, $w_j \nvDash \psi$. From v $\vDash \varphi^m_{w_i}$ and $w_i \vDash \psi$, we get $v \vDash \psi$. From $w_j \vDash \neg\, \psi$ and $w_j \sim_m v$ we get $v \vDash \neg\, \psi$. Contradiction. Hence wi $\vDash \psi$.

In the course of the proof of Theorem 2.4, using Theorem 2.2, we have in fact proved the following corollaries.

COROLLARY 2.2. *$L \in A^*\mathcal{V}_k$ if and only if L is a $\sim_m$-language for some $\bar{m}$ =(m_1,..., m_k).*

COROLLARY 2.3. *$L \in A^*\mathcal{V}_{k,m}$ if and only if L is a $\sim_m$-language for some $\bar{m}$ =(m, m_2, ..., m_k).*

Theorem 2.4 states precisely which are the important congruences related to the study of star-free languages. Section four will be concerned with an application of Theorem 2.4 and its corollaries. In the sequel $\mathcal{L}(m_1, \ldots, m_k)$ will denote the class of $\sim_{(m_1,\ldots,m_k)}$-languages. We end this section with a few notes on Theorem 2.4.

Kleene's theorem [15], stated in terms of congruences, asserts that L is regular if and only if there exists a finite index congruence ~ such that L is a ~-language. Schützenberger's theorem [25] states that L is star-free if and only if there exists a finite index aperiodic congruence ~ such that L is a ~-language. As a consequence of Theorem 2.4 we get a logical proof of the easiest side of Schützenberger's theorem, the $\sim_m$ being finite index aperiodic congruences (see Rosenstein [24] and the results in the next section). Two proofs of the Schützenberger's theorem have been given so far. Schützenberger's proof is done by recurrence on the cardinality of the syntactic monoid and uses Green's relations. The other proof, obtained independently by Cohen and Brzozowski and Meyer, is based on the decompositions as wreath products of semigroups. The last proof appears in Eilenberg's book [11].

Theorem 2.4 implies that the problem of deciding whether a language has dot-depth k is equivalent to the problem of effectively characterizing the monoids $M = A^* / \sim$ with $\sim \supseteq \sim_m$ for some $\bar{m} = (m_1 , \ldots, m_k)$, i.e., $V_k = \{A^*/\sim \mid \sim \supseteq \sim_m$ for some $\bar{m} = (m_1, \ldots, m_k)\}$.

3. SOME PROPERTIES OF THE CHARACTERIZING CONGRUEN CES

3.1. *An Induction Lemma*

The following lemma is a basic result (similar to one in [24] regarding $\sim_k$) which will allow us to resolve games with k +1 moves into games with k moves and thereby allows us to perform induction arguments.

LEMMA 3.1. *Let $\bar{m}$ = (m_1,..., m_k). $u \sim_{(m,m_1,\ldots,m_k)} v$ if and only if*

1. *for every $p_1, \ldots, p_m \in u$ ($p_1 \leq \ldots \leq p_m$) there are $q_1 \ldots, q_m \in v$ ($q_1 \leq \ldots \leq q_m$) such that*

    (a) *$Q^u_a p_i$ if and only if $Q^v_a q_i$, $a \in A$ for $1 \leq I \leq m$,*

    (b) $u[1,p_1) \sim_m v[1, q_1)$,

(c) $u(p_i, p_{i+1}) \sim_m v(q_i, q_{i+1})$ *for* $1 \leq i \leq m - 1$,

(d) $u(p_m, |u|] \sim_m v(q_m, |v|]$, *and*

2. *for every* $q_1, \ldots q_m \in v$ $(q_1 \leq \ldots \leq q_m)$ *there are* $p_1, \ldots p_m \in u$ $(p_1 \leq \ldots \leq p_m)$ *such that* (a)-(d) *hold.*

Proof. Suppose that player *II* has a winning strategy in $\mathcal{G}(m, m_1, \ldots, m_k)(u, v)$ and suppose that $p_1, \ldots p_m \in u$, $p_1 \leq \ldots \leq p_m$. Using the strategy we can find positions $q_1, \ldots, q_m \in v$, $q_1 \leq \ldots \leq q_m$ such that if player *I* chooses $p_1, \ldots, p_m \in u$ at his first move, then player *II* should choose $q_1, \ldots, q_m \in v$. Moreover, $Q_a^u\, p_i$ if and only if $Q_a^v q_i$, $a \in A$ for $1 \leq i \leq m$. There are now k moves left in the game $\mathcal{G}(m, m_1, \ldots, m_k)(u, v)$. Whenever player *I* chooses positions in $u[1, p_1)$ or if $v[1, q_1)$, the strategy, since it produces a win for player *II*, will always choose positions in $v[1, q_1)$ or $u[1, p_i)$. Thus player *II*'s winning strategy for $\mathcal{G}(m, m_1, \ldots, m_k)(u, v)$ includes within it a winning strategy for $\mathcal{G}\overline{m}(u[1, p_1), v[1, q_1))$, and similarly it includes a winning strategy for $\mathcal{G}\overline{m}(u(p_i, p_{i+1}), v(qi, q_{i+1}))$ for $1 \leq i \leq m — 1$, and $\mathcal{G}\overline{m}(u(p_m, |u|], v(q_m, |v|])$. This proves 1. By symmetry, 2 also holds.

Conversely, assuming that 1 and 2 hold, we describe a winning strategy for player *II* in $\mathcal{G}(m, m_1, \ldots, m_k)(u, v)$. If player *I* chooses positions $p_1, \ldots, p_m \in u$ $(p_1 \leq \ldots \leq p_m)$ on his first move, then player *II* uses 1 to find positions $q_1, \ldots, q_m \in v$ $(q_1 \leq \ldots \leq q_m)$. Thereafter, whenever player *I* chooses positions of $u[1, p_1)$ or $v[1, q_1)$, player *II* uses his winning strategy in $\mathcal{G}\overline{m}(u[1, p_1), v[1, q_1))$ to respond; and similarly, whenever player *I* chooses positions of $u(p_i, p_{i+1})$ or $v(q_i, q_{i+1})$ $(u(p_m, |u|]$ or $v(qm, |v|])$, player *II* uses his winning strategy in $\mathcal{G}\overline{m}(u(p_i, p_{i+1}), v(q_i, q_{i+1}))$ $(\mathcal{G}\overline{m}(u(p_m, |u|], v(qm, |v|]))$ to reply. Since there are only k subsequent moves in the game and $\sim_{(m_1,\ldots,m_k)}$ implies $\sim(m'_1, \ldots, m'_k)$ for all $m'_i \leq m_i$, player *I* can choose no more than k times from $u[1, p_1)$ or $v[1, q_1)$, $(u(p_i, p_{i+1})$ or $v(q_i\, q_{i+1}))$ $(u(p_m, |u|]$ or $v(q_m, |v|])$ and no more than m_i positions each time. Hence player II's winning strategies in $\mathcal{G}\overline{m}(u[1, p_1), v[1, q_1))$, $(\mathcal{G}\overline{m}(u(p_i, p_{i+1}), v(q_i, q_{i+1})))$ $(\mathcal{G}\overline{m}(u(p_m, |u|], v(q_m, |v|]))$ provides him with moves in all contingencies. If, on the other hand, player *I* chooses positions $q_1, \ldots, q_m \in v$, then player *II* uses 2 to find his correct first move and then proceeds analogously to the above. Thus player *II* has a winning strategy in $\mathcal{G}(m, m_1, \ldots, m_k)(u, v)$.

3.2. A Condition for Inclusion

Let us find a condition which ensures $\mathcal{L}(m_1, \ldots, m_k) \subseteq \mathcal{L}(m'_1, \ldots, m'_{k'})$. A trivial condition is the following: $k \leq k'$ and there exist $1 \leq i_1 < \ldots < i_k \leq k'$ such that $m_1 \leq m'_{i_1}, \ldots, m_k \leq m'_{i_k}$.

Define $\mathcal{N}(m_1, \ldots, m_k) = (m_1 + 1) \ldots (m_k + 1) — 1$.

PROPOSITION 3.1. *For* $N = \mathcal{N}(M1, \ldots, m_k) \geq 2$, $xyz^{N-2}zx \sim_{(m_1,\ldots,m_k)} xyz^{N-1}zx$.

Proof. The proof is similar to the one of a property of $\sim_k$ in [31]. Consider the natural decompositions of $u = xyx^{N-2}zx$ and $v = xyx^{N-1}zx$ into x- (y- or z-) segments. Before each move we have in u and v certain segments in which positions have been chosen, and others where no positions have been chosen. Call a maximal segment of succeeding x- (y- or z-) segments without chosen positions a gap. (a gap may be empty). Before each move there is a natural correspondence between the gaps in u and v (given by their order). *II* should play to what we call the *$(m_i, \ldots, m_k)$-strategy*, namely guarantee the following condition before each move: when $m_i + \ldots \bullet + m_k$ elements are still to be chosen by both players, two corresponding gaps should both consist of any number $\geq \mathcal{N}(m_i, \ldots, m_k)$ of x-(y- or z-) segments, or else should both consist of the same number $< \mathcal{N}(m_i, \ldots, m_k)$ of x- (y- or z-) segments. By induction on $k — i$ it is easy to see that *II* always can choose his segments in this manner; of course, inside his segments, *II* should pick exactly those positions which match the positions chosen by *I* in the corresponding segments.

Note that $\mathcal{N}(1, \ldots, 1) = 2^k — 1$. By putting $y = z = 1$ in the above proposition, we get as a corollary that if $m, m' \geq 2^k — 1$, then $(w)^m \sim_k (w)^{m'}$. $y = z = 1$ imply $x^N \sim_{(m_1,\ldots,m_k)} x^{N+1}$ $(N = \mathcal{N}(m_1, \ldots, m_k))$ and N is seen to be the smallest n such that $x^n \sim_{(m_1,\ldots,m_k)} x^{n+1}$ $|x| = 1$. Moreover, we see that if $u, v \in A^*$ and $u \sim_{(m_1,\ldots,m_k)} v$, then $|u|_a = |v|_a < \mathcal{N}(m_1, \ldots, m_k)$ or $|u|_a, |v|_a \geq \mathcal{N}(m_1, \ldots, m_k)$ (here, $|w|_a$ denotes the number of occurrences of the letter a in a word w). Also, similarly to the above proof, one can show that if $u \sim_{(m_1,\ldots,m_k)} v$ and $k \geq 2$, then either $u = v$ or u and v have a common prefix and suffix of length $\geq m_1 \ldots m_k$.

PROPOSITION 3.2. 1. $\sim_{(m_1,\ldots,m_k)} \subseteq \sim_{(\mathcal{N}(m_1,\ldots,m_k))}$ *and* 2. $\sim_{(m_1,\ldots,m_k)} \nsubseteq \sim_{(\mathcal{N}(m_1,\ldots,m_k)+1)}$.

Proof. By the preceding proposition, choosing $|x| = 1$, we have

$$U = x^{\mathcal{N}(m_1,\ldots,m_k)} \sim_{(m_1,\ldots,m_k)} = x^{\mathcal{N}(m_1,\ldots,m_k)+1} = v.$$

$= x^{\mathcal{N}(m_1,\ldots,m_k)+1}$ is a subword of length $\mathcal{N}(m_1, \ldots, m_k) + 1$ of v but not of u. This gives 2. 1 follows easily from Lemma 3.1.

Another condition for $\mathcal{L}(m_1, \ldots, m_k)$ to be included in $\mathcal{L}(m_1', \ldots, m_{k'}')$ is stated in the following.

PROPOSITION 3.3. *If* $k \leq k'$ *and there exist* $0 = j_0 < \ldots < j_{k-1} < j_k = k'$ *such that* $mi \leq \mathcal{N}(m_{j_{i-1}+1}', \ldots, m_{j_i}')$ for $1 \leq I \leq k$, *then* $\sim_{(m_1',\ldots,m_{k'}')} \subseteq \sim_{(m_1,\ldots,m_k)}$.

Proof. The result comes from the following observation: for $1 \leq i < j \leq k'$, we have $\sim_{(m_1',\ldots,m_i',\ldots,m_j',\ldots,m_{k'}')} \subseteq \sim_{(m_1',\ldots,m_{j-1}',}$ $\mathcal{N}(m_i', \ldots, m_j'), m_{j+1}', \ldots, m_{k'}'$, which is a consequence of Proposition 3.2, part 1.

Proposition 3.3 implies that if $n \geq \text{sum}(\overline{m})$ and $u \sim_n v$, then $u \sim_{n1} v$.

If $\sim_{\left(m_1',\ldots,m_{k'}'\right)} \subseteq \sim_{(m_1,\ldots,m_k)}$, then $\sim_{(m_1',\ldots,m_{k'}')} \subseteq \sim_{(\mathcal{N}(m_1,\ldots,m_k))}$.

Hence by Proposition 3.2, $\mathcal{N}(m_1, \ldots, m_k) \leq \mathcal{N}(m_1', \ldots, m_{k'}')$. Does the condition ($k < k'$ or ($k = k'$ and $m_1' \geq m_1$)) and $\mathcal{N}(m_1, \ldots, m_k) \leq \mathcal{N}(m_1', \ldots, m_{k'}'$ imply that $\sim_{(m_k',\ldots,m_{k'}')} \subseteq \sim_{(m_1,\ldots,m_k)}$? For $k = 1$, it is true. Section 4 includes partial results in this direction. $\mathcal{N}(m_1, \ldots, m_k)$ will appear several times in the sequel.

4. AN ANSWER TO A CONJECTURE OF PIN

First we introduce some terminology. The study of the concatenation product leads to the definition of the Schützenberger product of finite monoids. The reader is referred to [27] for the important properties of this construction. Let $M_1, \ldots, M_n$ be finite monoids. The *Schützenberger product* of $M_1, \ldots, M_n$, denoted by $\Diamond_n(M_1, \ldots, M_n)$ is the submonoid of upper triangular $n \times n$ matrices with the usual product of matrices of the form $p = (p_{ij})$, $1 \leq i, j \leq n$, in which the (i, j)-entry is a subset of $M_1 \times \ldots \times M_n$ and all of whose diagonal entries are singletons, i.e.,

1. $P_{ij} = \emptyset$ if $i > j$,
2. $p_{ii} = \{(1, \ldots, 1, m_i, 1, \ldots, 1)\}$ for some $m_i \in M_i$, (here, m_i is the ith component in the tuple),
3. $p_{ij} \subseteq \{(m_1, \ldots, m_n) \in M_1 \times \ldots \times M_n \mid m_1 = \ldots = m_{i-1} = 1 = m_{j+1} = \ldots m_n\}$.

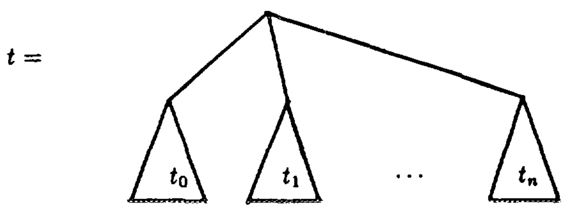
Condition 2 allows us to identify the coefficient p_{ii} with an element of M_i and condition 3 p_{ij} with a subset of $M_i \times \ldots \times M_j$. If $\mu = (m_i, \ldots, m_j) \in M_i \times \ldots \times M_j$ and $\mu' = (m_j', \ldots, m_k') \in M_j \times \ldots \times M_k$, then we define $\mu\mu' = (m_i, \ldots, m_{j-1}, m_j m_j', m_{j+1}', \ldots, m_k')$. This product is extended to sets in the usual fashion; addition is given by set union.

Straubing [27] has demonstrated that if the languages $Li \subseteq A^*$ $(0 \leq i \leq n)$ are recognized by the monoids M_i, then the language $L_0 a_1 L_1 a_2 \ldots a_n L_n$, where the a_i are letters, is recognized by the monoid $\Diamond_{n+1}(M_0, \ldots, M_n)$. It is easy to verify that if $0 \leq i_0 < \ldots < i_r \leq n$, then $\Diamond_{r+1}(M_{i_0}, \ldots, M_{i_r})$ is a submonoid of $\Diamond n+1(M_0, \ldots, M_n)$. This implies that the monoid $\Diamond_{n+1}(M_0, \ldots, M_n)$ recognizes all languages of the form $L_{i_0}a_1L_{i_1}a_2 \ldots a_rL_{i_r}$, where L_{i_k} is recognized by M_{i_k}. A partial converse has been established. The case $n = 1$ has been treated by Reutenauer [23] and the general case by Pin [22]. We have that if a language $L \subseteq A^*$ is recognized by $\Diamond_{n+1}(M_0, \ldots, M_n)$ then L is in the boolean algebra generated by the languages of the form $L_{i_0}a_1L_{i_1}a_2 \ldots a_rL_{i_r}$ where $0 \leq i_0 < \ldots < i_r \leq n$, where for $0 \leq k \leq r$, $a_k \epsilon A$, and L_{i_k} is a language recognized by M_{i_k}.

Let W be a M-variety. We define $\Diamond W$ to be the variety of all finite monoids that divide some Schützenberger product $\Diamond_n(M_1, \ldots, M_n)$ for some n, where $M_i \in W$ for $i = 1, \ldots, n$. From the above discussion, we have that for $k \geq 0$, $V_{k+1} = \Diamond V_k$. In particular, $V_1 = J = \Diamond 1$ and $V_2 = \Diamond J$, where I denotes the variety consisting of the trivial monoid alone and J of all finite $\mathcal{J}$-trivial monoids.

4.1. Decidability and Inclusion Problems

Pin [22] demonstrated that the Straubing hierarchy is a particular case of a more general construction obtained in associating varieties of languages not to integers but to trees under the following fashion. A variety of languages is associated by definition to the tree reduced to a point. Then to the tree



is associated the boolean algebra generated by the languages of the form $L_{i_0}a_1L_{i_1}a_2 \ldots a_rL_{i_r}$ with $0 \leq i_0 < \ldots < i_r \leq n$, where for $0 \leq j \leq r$, L_{i_j} is member of the variety of languages associated to the tree t_{i_j}. Since the Schützenberger product is perfectly adapted to the operation $(L_0, \ldots, L_n) \rightarrow L_0a_1L_1a_2 \ldots a_nL_n$, it permits us to construct, without reference to languages, hierarchies of varieties of monoids corresponding, via Eilenberg's theorem, to the hierarchies of languages precedently constructed; i.e., starting with a variety of monoids W, we associate with each tree t, respectively with each set of trees T, a variety of monoids $\Diamond_t(W)$ ($\Diamond_T(W)$). Descriptions of the hierarchies of monoids are given after a few definitions.

We will denote by $\mathcal{T}$ the set of trees on the alphabet $\{a, \bar{a}\}$. Formally, $\mathcal{T}$ is the set of words in $\{a, \bar{a}\}^*$ congruent to 1 in the congruence generated by the relation $a\bar{a} = 1$. Intuitively, the words in $\mathcal{T}$ are obtained as follows: we draw a tree and starting from the root we code a for going down and $\bar{a}$ for going up. For example,

is coded by $aa\bar{a}aa\bar{a}a\bar{a}a\bar{a}\bar{a}\bar{a}a\bar{a}$. The number of leaves of a word t in $\{a, \bar{a}\}^*$, denoted by $l(t)$ is by definition the number of occurrences of the factor $a\bar{a}$ in t. Each tree t factors uniquely into $t = at_1\bar{a}at_2\bar{a} \ldots at_n\bar{a}$, where $n \geq 0$ and where the t_i's are trees. We have then $l(t) = \Sigma_{1 \leq i \leq n}\, l(t_i)$. Let t be a tree and let $t = t_1at_2\bar{a}t_3$ be a factorization of t. We say that the occurrences of a and $\bar{a}$ defined by this factorization are related if t_2 is a tree. Let t and t' be two trees. We say that t is extracted from t' if t is obtained from t' by removing in t' a certain number of related occurrences of a and $\bar{a}$. We now state the algebraic interpretation of the above stated hierarchy construction using the Schützenberger product.

To each tree t and to each sequence $W_1, \ldots, W_{l(t)}$ of varieties of monoids, we associate a variety of monoids $\Diamond_t(W_1, \ldots, W_{l(t)})$ defined recursively by:

1. $\Diamond_1(W) = W$ for every M-variety W,

2. if $t = at_1\bar{a}at_2\bar{a} \dots at_n\bar{a}$ with $n \geq 0$ and $t_1, \dots, t_n \in \mathcal{T}$, $\Diamond_t(W_1, \dots, W_{l(t)})$ is the variety of monoids M such that M divides some $\Diamond_n(M_1, \dots, M_n)$ with $M1 \in \Diamond_{t_1}(W1, \dots, W_{l(t_1)})$, …, $Mn \in \Diamond_{t_n}(W_{l(t_1)+\dots+l(t_{n-1})+1}, \dots, W_{l(t_1)+\dots+l(t_n)})$.

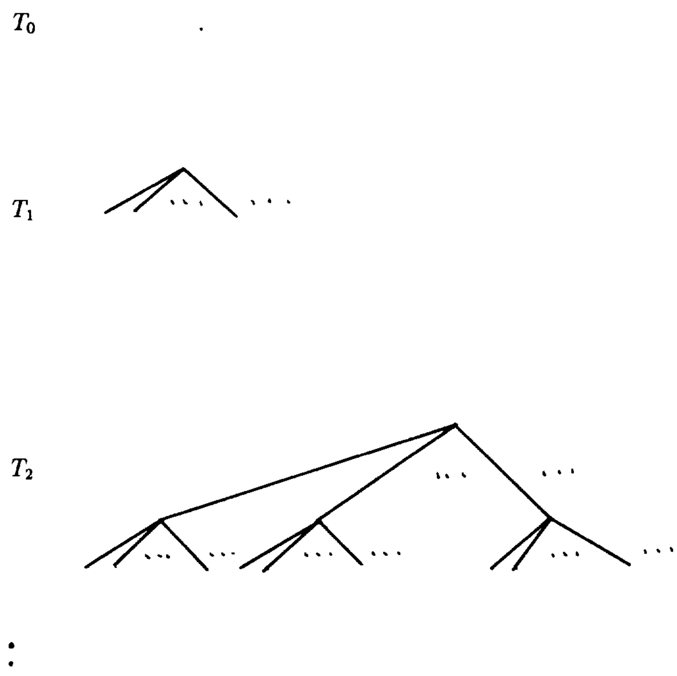
When $W_1 = \bullet\bullet\bullet = W_{l(t)} = W$, we denote simply $\Diamond_t(W)$ the variety $\Diamond_t(W_1, \dots, W_{l(t)})$. More generally, if T is a language contained in , we denote $\Diamond_T(W)$ the smallest variety containing the varieties $\Diamond_t(W)$ with $t \in T$.

The following proposition allows us, by recurrence, to describe the languages associated to the varieties $\Diamond_t(W_1, \dots, W_{l(t)})$ for each tree t.

PROPOSITION 4.1 (Pin [22]). *Let n be a positive integer and let $W_0, \dots, W_n$ be M-varieties. We denote respectively by $\mathcal{W}_j$ and $\mathcal{W}$ the *-varieties of languages corresponding to W_j $(0 \leq j \leq n)$ and to $\Diamond_{(aa)^{n+1}}(W_0, \dots, W_n)$. Then for each alphabet A, $A^*\mathcal{W}$ is the boolean algebra generated by the languages of the form $L_{i_0}a_1L_{i_1}a_2 \dots a_rL_{i_r}$, where $0 \leq i_0 < \dots < i_r \leq n$, where for $0 \leq j \leq r$, $a_j \in A$, and $L_{i_j} \in A^*\mathcal{W}_{i_j}$.*

The above proposition implies that if $t = at_1\bar{a}at_2\bar{a} \dots at_n\bar{a}$ with $t_1, \dots, t_n \in T$, we have $\Diamond_t(W) = \Diamond_{(a\bar{a})^n}(\Diamond_{t_1}(W), \dots, \Diamond_{t_n}(W))$.

The Straubing hierarchy V_k can be described in the following fashion. Let T_k be the sequence of languages defined by $T_0=\{1\}$ and $T_{k+1}=(aT_k\bar{a})^*$. Intuitively, we can represent the languages by trees infinite in width:



PROPOSITION 4.2. *For $k \geq 0$, $V_k = \Diamond_{T_k}(I)$. In particular, $\Diamond_{T_0}(I) = I$, $\Diamond_{T_1}(I) = J$, $\Diamond_{T_2}(I) = \Diamond J$.*

Proof. It is an immediate consequence of Proposition 4.1.

More precisely, we have the following.

PROPOSITION 4.3. *For $k \geq 1$, $m \geq 1$, $V_{k,m} = \Diamond_{(aT_{k-1}a)^{m+1}}(I)$.*

Proof Let $\mathcal{W}_{k,m}$ be the *-variety of languages corresponding to

$$\Diamond_{(aT_{k-1}a)^{m+1}}(I) = \Diamond_{(aa)^{m+1}}(\Diamond_{T_{k-1}}(I)).$$

We have to establish the equality $\mathcal{W}_{k,m} = \mathcal{V}_{k,m}$. Proposition 4.1 and $V_k = \lozenge_{T_k}(I)$ of the preceding proposition show that for each alphabet A, $A^* \mathcal{W}_{k,m}$ is the boolean algebra generated by the languages of the form $L_0a_1L_1a_2 \ldots a_nL_n$, where $0 \leq n \leq m$, $L_0, \ldots, L_n \in A^* \mathcal{V}_{k-1}$, and $a_1, \ldots, a_n \in A$. The result clearly follows.

Let $\overline{m} = (m_1, \ldots, m_k)$. By induction on k, we define a tree $t\overline{m}$ as follows: if length($\overline{m}$) = 1, then $t\overline{m} = (a\bar{a})^{m_1+1}$, for $\overline{m} = (m, m_1, \ldots, m_k)$, $t\overline{m} = (at(\overline{m}_1 \ldots, m_k)\,\bar{a})^{m+1}$. One can also observe that $l(t(m_1, \ldots, m_k))$ *is* $\mathcal{N}(m_1, \ldots, m_k) + 1$.

Let t be a tree and let $\mathcal{V}_t$ be the *-variety of languages associated with $\lozenge_t(I)$. We have the following.

PROPOSITION 4.4. $\mathcal{V}_{t(m_1,\ldots,m_k)} = \mathcal{L}(m_1, \ldots, m_k)$. *(Here, it is understood that for each alphabet A, A** $\mathcal{V}_{t(m_1,\ldots,m_k)}$ *is the class of* $\sim_{(m_1,\ldots,m_k)}$*-languages in A*. Let us denote it by A** $\mathcal{L}(m_1, \ldots, m_k)$).

Proof. The proof is by induction on k. If $k = 1$, then $\lozenge_{t(m_1)}(I) = V1, m_1$ by Proposition 4.3. The result then follows from Corollary 2.3. Suppose it is true for k, i.e., letting $\overline{m} = (m_1, \ldots, m_k)$, $\mathcal{V}_{tm} = \mathcal{L}\overline{m}$. Let us show that $\mathcal{V}_{t(m,m_1,\ldots,m_k)} = \mathcal{L}(m, m_1, \ldots, m_k)$. From $\lozenge_{t(m,m_1,\ldots,m_k)}(I) = \lozenge_{(atma)^{m+1}}(I) = \lozenge_{(aa)^{m+1}}(\lozenge_{tm}(I))$, using the induction hypothesis and Proposition 4.1, we can conclude that for each alphabet A, $A^*\mathcal{V}_{t(m,m_1,\ldots,m_k)}$ is the boolean algebra generated by the languages of the form $L_0a_1L_1a_2 \cdots a_mL_m$, where for $0 \leq j \leq m$, $a_j \in A^*$ and $L_j \in A^* \mathcal{L}(m_1, \ldots, m_k)$. The result follows since each $\sim(m, m_1,\ldots,m_k)$-*class* is a boolean combination of sets of the form $L_0a_1L_1a_2 \cdots a_mL_m$, where each L_j is a $\sim_{(m_1,\ldots,m_k)}$-class.

The following result perhaps constitutes a first step towards the general solution of the decidability problem.

PROPOSITION 4.5 (Pin [22]). *For each tree t, the variety* $\lozenge_t(I)$ *is decidable.*

Using Proposition 4.4 and Proposition 4.5, we get the following.

PROPOSITION 4.6. *For, fixed* $(m_1, \ldots, m_k)$, *the M-variety* $\lozenge_{t(m_1,\ldots,m_k)}(I)$ *is decidable, or the *-variety of languages* $\mathcal{L}(m_1, \ldots, m_k)$ *is decidable.*

Among the many problems concerning these tree hierarchies, is the comparison between the varieties inside a hierarchy. More precisely, the problem consists in comparing the different varieties $\lozenge_t(W)$ (or even $\lozenge_T(W)$). A partial result and a conjecture on this problem was given in Pin [22]. It was shown that for every variety W, if t is extracted from t', then $\lozenge_t(W) \subseteq \lozenge_{t'}(W)$, and it was conjectured that if $t, t' \in T'$, $\lozenge_t(I)$ is contained in $\lozenge_{t'}(I)$ if and only if t is extracted from t'. Here, T' denotes the set of trees in which each node is of arity different from 1.

THEOREM 4.1. *The above conjecture is false.*

To see this, $\mathcal{L}_{(1,2)} \subseteq \mathcal{L}_{(2,1)}$ by Lemma 4.7 of the next section. Hence $\lozenge_{t(1,2)}(I) \subseteq \lozenge_{t(2,1)}(I)$ by Proposition 4.4. But it is easy to verify that the tree $t(1, 2)$ is not extracted from the tree $t(2, 1)$. The main step of the proof of Theorem 4.1 is given in the next section.

4.2. The Conjecture is False

This section is devoted to the proof of Theorem 4.1 of the preceding section. The proof goes through seven lemmas, Lemmas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7. When is $\sim_{(2,m_2')} \subseteq \sim_{(1,m_2)}$? Of course, if $m_2' \geq m_2$, it is true. We will be considering the case when $m_2' < m_2$, or, $m_2' + 1 \leq m_2$. Assume that $u \sim_{(2,1)} v$ and $|u|_a, |v|_a > 0$. Let $u = u_0au_1 \cdots au_n$, $v = v_0av_1 \cdots av_m$, where $n = |u|_a$, $m = |v|_a$. If $Q_a^u p_i$, $q_a^v q_j$ for $i = 1, \ldots, n$, $j = 1, \ldots, m$, then $u_i = u(p_i, p_{i+1})$, $i = 1, \ldots, n - 1$, $v_j = v(q_j, q_{j+1})$, $j = 1, \ldots, m - 1$. $u_0 = u[1, p_1)$, $v_0 = v[1, q_1)$, $u_n = u(p_n, |u|]$, $v_m = v(q_m, |v|]$.

LEMMA 4.1. 1. $u_0 \sim_1 v_0$, $u_1 \sim_1 v1$, $u_{n-1} \sim_1 v_{m-1}$, $u_n \sim_1 v_m$,
2. $u_2au_3 \ldots au_{n-2} \sim_1 v_2av_3 \ldots av_{m-2}$.

Proof 1. Player *I*, in the first move chooses two consecutive *a*'s among the first or the last two ones (of *u* or *v*). Since u $\sim_{(2,1)} v$, player *II* chooses two consecutive *a*'s, the same occurrences among the first or the last two ones (of *v* or *u*). The result follows from Lemma 3.1.

2. Let *w* be a subword of length ≤ 1 of $u_2au_3 \cdots au_{n-2}$ (or of $v_2av_3 \cdots av_{m-2}$). Hence *w* is a subword of $v_2av_3 \cdots av_{m-2}$ (or of $u_2au_3 \cdots au_{n-2}$) because *aawaa* is a subword of length $\leq \mathcal{N}(2, 1) = 5$ of *u* (or of *v*) ($\sim_{(2,1)} \subseteq \sim_{(\mathcal{N}(2,1))}$) by Proposition 3.2(I).

LEMMA 4.2. 1. $u_1au_2 \cdots au_n \sim_{(2)} v_1av_2 \cdots av_m$,

$u_2au_3 \cdots au_n \sim_{(2)} v_2av_3 \cdots av_m$,

$u_3au_4 \cdots au_n \sim_{(2)} v_3av_4 \cdots av_m$;

2. $u_0au_1 \cdots au_{n-1} \sim_{(2)} v_0av_1 \cdots av_{m-1}$,

$u_0au_1 \cdots au_{n-2} \sim_{(2)} v_0av_1 \cdots av_{m-2}$,

$u_0au_1 \cdots au_{n-3} \sim_{(2)} v_0av_1 \cdots av_{m-3}$.

Proof: 1. Let $1 \leq i \leq 3$. Let *w* be a subword of length ≤ 2 in $u_iau_{i+1} \cdots au_n$. Consider $w' = a^iw$ of length $\leq i + 2 \leq \mathcal{N}(2, 1)$. $u \sim_{(\mathcal{N}(2,1))} v$ (Proposition 3.2(1)) and the fact that w' is a subword of *u* of length $\leq \mathcal{N}(2, 1)$ imply that w' is also a subword of *v* and, hence, *w* a subword in $v_iav_{i+1} \cdots av_m$. Similarly, for subwords of $v_iav_{i+1} \cdots av_m$. For 2, we consider wa^i.

LEMMA 4.3. 1. $u_0 \sim_{(2)} v_0$,
2. $u_n \sim_{(2)} v_m$.

Proof 1. Let $w = w_1 \ldots w_{|w|}$ be a subword of length ≤ 2 in u_0. Let $p, p' \in u$ be such that $p \leq p' < p_1$ and $Q^u_{w_1}p$, $Q^u_{w_{|w}}p'$. Consider the following play of the game $\mathcal{G}(2, 1)(u, v)$. In the first move, player *I* chooses *p* and p_1. Using Lemma 3.1, there is $q \in v$, $q < q_1$, $Q^v_{w_1}$ q, and $u(p, p_1) \sim_1 v(q, q_1)$. Since $w_{|w|}$ is a subword of length ≤ 1 in $u(p, p_1)$ and $u(p, p_1) \sim_1 v(q, q_1)$, $w_{|w|}$ is a subword of length ≤ 1 in $v(q, q_1)$. Hence *w* is also a subword in v_0. Similarly, for subwords of v_0. For 2, let $w = w_1 \ldots w_{|w|}$ be a subword of length ≤ 2 in u_n. Let $p, p' \in u$ be such that $p_n < p' \leq p$ and $Q^u_{w_{|u}}p$, $Q^u_{w_1}$ p'. In the first move, player *I* chooses p_n and *p*. The result follows similarly as 1.

LEMMA 4.4. 1. $u_0au_1 \sim_{(2)} v_0av_1$,
2. $u_{n-1}au_n \sim (2)\ v_{m-1}av_m$.

Proof. 1. We will show that $u_0au_1 \sim_{(2)} v_0av_1$. The proof is similar for 2. Let $w = w_1 \ldots w_{|w|}$ be a subword of length ≤ 2 in u_0au_1 (similar if starting with v_0av_1). We want to show that *w* is a subword of v_0av_1. If *w* is a subword of u_0, *w* is also a subword of v_0 by Lemma 4.3(1). If not, let *j*, $1 \leq j \leq |w|$, be the first index such that $w_1 \ldots w_j$ is not a subword of u_0 but $w_1 \ldots w_{j-1}$ is a subword of u_0. We have that w1 $\ldots w_{j-1}$ is a subword of v_0 by Lemma 4.3(1) but we do not have that $w_1 \ldots w_j$ is a subword of v_0 (if we had, $w_1 \ldots w_j$ would be in u_0 for the same reason). If $w_j = a$, $w_1 \ldots w_j$ is a subword of u_0a and v_0a, and since $u_1 \sim_1 v_1$ by Lemma 4.1(1) and $1 \leq j \leq |w|$, *w* is a subword of v_0av_1. If $w_j \neq a$, let *p* be the first position in *u* after p_1 such that $Q^u_{w_j}$ p. Now, since $u_1 \sim_1 v_1$ by Lemma 4.1(1), w_j occurs between q_1 and q_2. Let *q* be the first position in *v* after q_1 such that $Q^v_{w_j}q$. If $|w_j \ldots w_{|w|}| \leq 1$, the proof is complete. If not, i.e., $|w_j \ldots w_{|w|}| > 1$ then $j = 1$, $|w| = 2$. Consider the following play of the game $\mathcal{G}(2, 1)(u, v)$. Player *I* in the first move, chooses positions *p* and p_2 in *u*. Player *II* should choose *q* in *v*. If not, *II* would choose a position q' in *v* such that $q' > q$ because he needs at least one *a* before q', and *q* is the first position in *v* after q_1 such that $Q^u_{w_1}$q. But then, player *I*, in the second move could choose an occurrence of w_1

from $v[1, q')$ (not possible for *II* in $u[1, p)$ from the choice of j and the fact that $w_j \neq a$). Player *II* cannot choose a position q'' such that $Q^v_a q''$ before q_2 because he needs at least one a before q. Since there is no a between p and p_2, there should not be any between q and q''. Hence player *II* should choose q and q_2. Hence $u(p, p_2) \sim_1 v(q, q_2)$ and 1 follows.

LEMMA 4.5. *Let $p'_1, \ldots, p'_s$ in u ($p'_1 < \cdots < p'_s$) ($q'_1, \ldots, q'_{s'}$ in v ($q'_1 < \cdots < q'_{s'}$)) be the positions which spell the first and the last occurrences of every letter in u (v). Then*

1. $s = s'$,
2. *$Q^u_b p'_i$ if and only if $Q^v_b q'_i$, $b \in A$ for $1 \leqslant i \leqslant s$,*
3. *$u[1, p'_i) \sim_{(2)} v[1, q'_i)$ and $u(p'_i, |u|] \sim_{(2)} v(q'_i, |v|]$ for $1 \leqslant i \leqslant s$,*
4. *$u(p'_i, p'_{i+1}) \sim_1 v(q'_i, q'_{i+1})$ for $1 \leqslant i \leqslant s-1$,*
5. *for $1 \leqslant i \leqslant s-1$ and for every $p' \in u(p'_i, p'_{i+1})$, there exists $q' \in v(q'_i, q'_{i+1})$ such that*
   a. *$Q^u_b p'$ if and only if $Q^v_b q'$, $b \in A$,*
   b. *$u(p'_i, p') \sim_1 v(q'_i, q')$.*

Also, there exists $q' \in v(q'_i, q'_{i+1})$ (which may be different from the one which satisfies a, b) *such that* a *and*

   c. *$u(p', p'_{i+1}) \sim_1 v(q', q'_{i+1})$.*

Similarly, for every $q' \in v(q'_i, q'_{i+1})$, there exists $p' \in u(p'_i, p'_{i+1})$ such that a, b *hold (also* a, c *hold) and*

6. *for $1 \leqslant i \leqslant s-1$ and for every $p''_1, p''_2 \in u(p'_i, p'_{i+1})$ ($p''_1 < p''_2$), there exist $q''_1, q''_2 \in v(q'_i, q'_{i+1})$ ($q''_1 < q''_2$) such that*
   d. *$Q^u_b p''_j$ if and only if $Q^v_b q''_j$, $b \in A$ for $1 \leqslant j \leqslant 2$,*
   e. *$u(p''_1, p''_2) \sim_1 v(q''_1, q''_2)$.*

Similarly, for every q''_1, $q''_2 \in v(q'_i, q'_{i+1})$ ($q''_1 < q''_2$), there exist p''_1, $p''_2 \in u(p'_i, p'_{i+1})$ ($p''_1 < p''_2$) such that d and e hold.

Proof: 1 holds since $u \sim_{(2.1)} v$, by Section 3, implies that $|u|_b = |v|_b < \mathcal{N}(2, 1) = 5$ or $|u|_b, |v|_b \geq \mathcal{N}(2, 1)$ for every $b \in A$.

2 holds, since $\sim_{(2,1)} \subseteq \sim_{(1,1)}$ and we may consider the plays of the game $\mathcal{G}(1, 1)(u, v)$, where player *I* in the first move chooses p'_i for some i, $1 \leq i \leq s$.

3 follows from the arguments in the proofs of Lemma 4.2 and Lemma 4.3, since p'_i (q'_i) is either the first or the last occurrence of a letter in u(v) (in Lemma 4.2 and Lemma 4.3 we were considering p_1 (q_1) which are the first occurrences of the letter a in u (v) and p_n (q_m) which are the last occurrences of that letter in u (v)).

4, 5, and 6 follow by considering different plays of the game $\mathcal{G}(2, 1)(u, v)$. First, from the choice of the p'_r's and the q'_r's and Lemma 3.1, if p'_i (q'_i) is among the positions chosen in u (v) by player *I* in the first move, then q'_i (p'_i) should be among the ones chosen in v (u) by player *II* in the first move. Second, if the positions chosen by player *I* in the first move are in $u(p'_i, p'_{i+1})$($v(q'_i, q'_{i+1})$), then the positions chosen by player *II* in the first move should be in $v(q'_i, q'_{i+1})$ ($u(p'_i, p'_{i+1})$) for the same reasons. For 4, consider the play of the game $\mathcal{G}(2, 1)(u, v)$, where player *I*, in the first move, chooses p'_i and p'_{i+1}; for 5, *I* chooses p'_i and p', or p' and p'_{i+1}; for 6, he chooses p''_1 and p''_2.

LEMMA 4.6. *Let $p'_1, \ldots, p'_s$ in u ($p'_1 < \ldots < p'_s$)($q'_1, \ldots, q'_s$ in v($q'_1 < \ldots < p'_s$)) be the positions which spell the first and last occurrences of every letter in u (v) so (satisfying) 2, 3, 4, 5, and 6 of Lemma 4.5. For i fixed* *between 1 and s -1, let* p_1^n, …, p''_{s_i} *in* $u(p'_i, p'_{i+1})$ (p''_1 < … < p''_{s_i}) (q''_1, …, $q''_{s'_i}$ *in* $v(q'_i, q'_{i+1})$ (q''_1< … < $q''_{s'_i}$)) *be the positions which spell the first and the last occurrences of every letter in* $u(p'_i , p'_{i+1})$ ($v(q'_i, q'_{i+1})$). *Then*

1. $s_i = s'_i$,

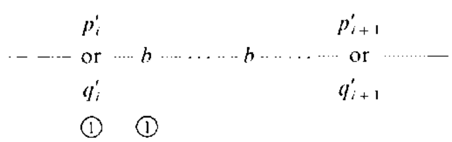
2. $Q_b^u p''_j$ *if and only if* $Q_b^v p''_j$, $b \in A$ *for* $1 \leq j \leq s_i$ *and*

3. $u[1, p''_j) \sim_{(2)} v[1, q''_j)$ *and* $u(p''_j , |u|] \sim_{(2)} v(q''_j , |v|]$ *for* $1 \leq j \leq s_i$.

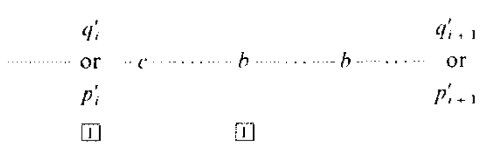
Proof. By 4 of Lemma 4.5 we have $u(p'_i , p'_{i+1}) \sim_1 v(q'_i, q'_{i+1})$. Now, if in one of these segments, either $u(p'_i, p'_{i+1})$ or $v(q'_i , q'_{i+1})$, there is only one occurrence of some letter and in the other segment there are two or more occurrences of that same letter, then player *I* in the first move could choose two of these occurrences (not possible for *II* in the remaining segment, contradicting 6 of the preceding lemma). Hence 1 holds.

For 2, consider any two letters, say $b \neq c$, in $u(p'_i\ p'_{i+1})$ (and, hence, in $v(q'_i\ q'_{i+1})$ by Lemma 4.5(4)) and consider their first and last occurrences in $u(p'_i, p'_{i+1})$ and $v(q'_i , q'_{i+1})$ (by 1, the numbers of these occurrences agree). We claim that we have the same pattern: there are six possibilities, namely, pattern 1, *bbcc*; or pattern 2, *bcbc*; or pattern 3, *bccb*; or pattern 4, *cbbc*; or pattern 5, *cbcb*; or pattern 6, *ccbb*. Expressed differently, the subwords formed by these occurrences are the same (the proof is similar if only one occurrence of a letter instead of a first and a last: the patterns would be shorter words). Let us separate different patterns by considering plays of the game $\mathcal{G}(2, 1)(u, v)$. We will illustrate the plays by diagrams. The first move of *I* will be indicated by [circle with 1 in middle] and the first move of *II* by [square with 1 in middle]. In each diagram, the segment between the positions chosen by *I* in move 1 $\nsim_1$, the segment between the positions chosen by *II* in move 1, in contradiction with Lemma 4.5(5) or (6). We show how to separate patterns 1-2-3 from patterns 4-5-6, pattern 1 from patterns 2 and 3, and pattern 2 from pattern 3. The separation of the patterns 4, 5, and 6 is similar to the separation of 1, 2, and 3. To separate patterns 1-2-3 from patterns 4-5-6:

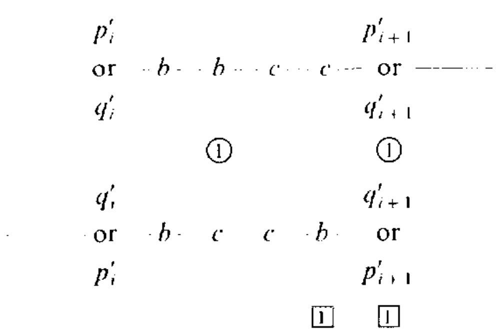
pattern 1, 2, or 3.



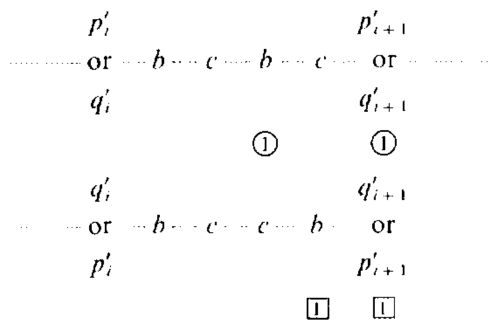
pattern 4, 5, or 6.



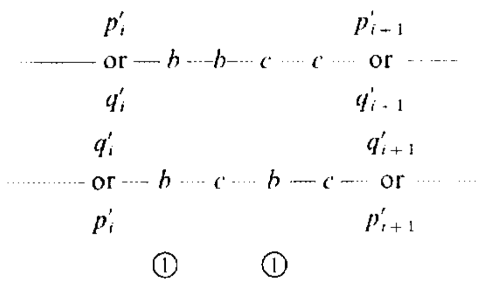
The above diagram is in contradiction with Lemma 4.5(5) (*II* has to choose the first occurrence of *b* but there is an occurrence of *c* between the positions that he chooses which is not the case for *I*). To separate patterns 1 and 3,



To separate patterns 2 and 3,



To separate patterns 1 and 2,



Here, player *II* cannot choose two b's separated by a c (in contradiction with Lemma 4.5(6)).

The diagrams above show that any two letters obey the same pattern. $Q_b^u p''_1$ if and only if $Q_b^v q''_1$ is clear. Now, by induction on j, assume $Q_b^u p''_k$, if and only if $Q_b^v q''_k$ for $1 \le k \le j$. Suppose, say $Q_b^u p''_{j+1}$ + and $Q_b^v q''_{j+1}$ with $b \neq c$. But b and c have the same pattern in $u(p'_i, p''_j]$ and in $v(q'_i, p''_j]$ by the induction hypothesis and the result follows.

We now prove 3. Let $1 \le j \le s_i$. We will show that $u[\,1, p''_j\,) \sim_{(2)} v[1, q''_j\,)$ (the proof is similar for $u(p''_j, |u|] \sim_{(2)} v(q''_j, |v|])$. Let $w = w_1 \ldots w_{|w|}$ be a subword of length ≤ 2 in $u[\,1, p''_j\,)$ (it is similar if in $v[1, q''_j\,)$). We want to show that w is a subword of $v[\,1, q''_j\,)$. If $|w| = 1$, then there is an occurrence of w_1 in $u[\,1, p'_i]$ (and, hence, in $v[1, q'_i]$) from the choice of the p'_r's and the q'_r's and Lemma 4.5(1, 2) and the proof is complete. If $|w| = 2$, and w is in $u[\,1, p'_i)$, then w is in $v[1, q'_i)$ by Lemma 4.5(3). If there is an occurrence of w_1 in $u[\,1, p'_i)$ (and, hence, in $v[1, q'_i)$ by Lemma 4.5(3)) and $Q_{w_2}^u p'_i$ (and hence $Q_{w_2}^v q'_i$ by Lemma 4.5(2)) the proof is complete. Otherwise, there is an occurrence of w_1 in $u[\,1, p'_i]$ (and, hence, in $v[\,1, q'_i]$) from the choice of the p'_r's and q'_r's and Lemma 4.5(1, 2) and also an occurrence of w_2 in $u(p'_i, p''_j\,)$. From the choice of the p''_r's, there exists k, $k < j$, such that $Q_{w_2}^u p''_k$. Hence, from the choice of the q''_r's and (1, 2), $Q_{w_2}^v p''_k$.

LEMMA 4.7. $\sim_{(2,1)} \subseteq \sim_{(1,2)}$.

Proof. Suppose that $u \sim_{(2,1)} v$. Then there is a winning strategy for player *II* in the game $\mathcal{G}(2, 1\,)(u, v)$ to win each play. Let us describe a winning strategy for player *II* in the game $\mathcal{G}(1, 2\,)(u, v)$ to win each play. Let p be a position in u chosen by player *I* in the first move. Suppose $Q_a^u p$ for some $a \in A$.

Case 1. $|u|_a = |v|_a < 5 = \mathcal{N}(1, 2) = \mathcal{N}(2, 1)$. If p is the ith occurrence of a in u chosen by player *I* in the first move, then player *II* chooses the same occurrence of a in v, say position q. The fact that $u[1, p) \sim_{(2)} v[1,q)$ and $u(p, |u|] \sim_{(2)} v(q, |v|]$ follows from Lemmas 4.2, 4.3, and 4.4.

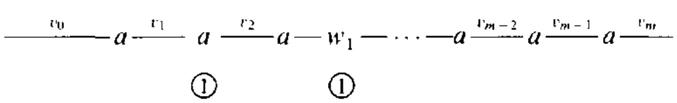
Case 2. $|u|_a = |v|_a = 5$. Same as case 1.

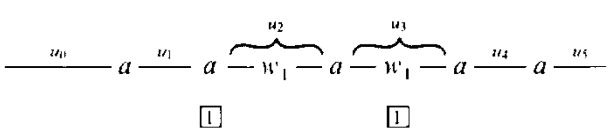
Case 3. $|u|_a = 5$, $|v|_a > 5$. We include this case because the strategy here for player *II* is very easy but the arguments in Case 4 are enough to prove the lemma. If p is the ith occurrence of a in u ($1 \le i \le 2$) chosen by player *I* in the first move, then player *II* chooses the same occurrence of a in v, say position q. If p is the (6 - i)th occurrence of a in u ($1 \le i \le 2$), player *II* chooses the (m — $i + 1$)th occurrence of a in v. The fact that $u[1, p)$ $\sim_{(2)}$ $v[1,q)$ and $u(p, |u|]$ $\sim_{(2)}$ $v(q, |v|]$ follows from Lemmas 4.2, 4.3, and 4.4. If $p = p_3$, then player *II* chooses q, an a, among the middle ones in v, i.e., among $q_3, \ldots, q_{m-2}$. Lemma 4.2 implies that $u_3au_4au_5 \sim_{(2)} v_3av_4 \ldots av_m$ and $u_0au_1au_2 \sim_{(2)} v_0av_1 \ldots av_{m-3}$. Observe that if we show $u_0au_1au_2 \sim_{(2)} v_0av_1av_2$ and $u_3au_4au_5 \sim_{(2)} v_{m-2}av_{m-1}av_m$ the proof is complete, since we will have $u_0au_1au_2 \sim_{(2)} v[1,q)$ and $u_3au_4au_5 \sim_{(2)}$ $v(q, |v|]$ for any position q among $q_3, \ldots, q_{m-2}$. If player *I* had chosen p among the middle positions in v, then player *II* would choose p_3 in u. So let us show that $u_0au_1au_2 \sim_{(2)} v_0av_1av_2$. The proof of u3au4au5$\sim_{(2)}$ $v_{m-2}av_{m-1}av_m$ is similar.

First, let w be a subword of length ≤ 2 in $v_0av_1av_2$. Then w is a subword of length ≤ 2 in $v_0av_1 \ldots av_{m-3}$. But since $u_0au_1au_2 \sim_{(2)} v_0av_1 \ldots av_{m-3}$, w is a subword of $u_0au_1au_2$.

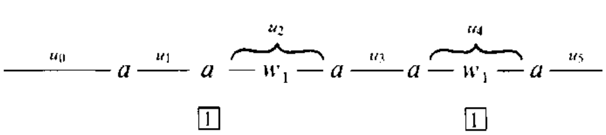
Now, let $w = w_1 \ldots w_{|w|}$ be a subword of length ≤ 2 in $u_0au_1au_2$. We want to show that w is a subword of $v_0av_1av_2$. If w is a subword of u_0au_1, w is a subword of v_0av_1 by Lemma 4.4(1). If not, let j be the first index such that $w_1 \ldots w_j$ is not a subword of u_0au_1 but that $w_1 \ldots w_{j-1}$ is a subword of u_0au_1. We have to consider the case where $j = 1$ and the case where $j = 2$. In each case, $u_0au_1au_2 \sim_{(2)} v_0av_1av_2$ will follow by considering different plays of the game $\mathcal{G}(2, 1)(u, v)$. We will illustrate the plays by diagrams. The first move of *I* will be indicated by [circle with 1 in middle] and the first move of *II* by [square with 1 in middle].

$j = 1$. We have that w_1 is not a subword of v_0av_1 ; $w_1 \neq a$ since otherwise w_1 would be in u_0au_1, contradicting the choice of j. So let p' be the first position in u after p_2 such that $Q^u_{w_1}p'$. Now, since $u_0au_1au_2 \sim_{(2)} v_0av_1 \ldots av_{m-3}$ and w_1 is not in v_0av_1, w_1 occurs between q_2 and q_{m-2}. Let q' be the first position in v after q_2 such that $Q^u_{w_1}q'$; q' is not between q_2 and q_3 in v because then we would have w_1aaaa in v but not in u. Hence q' is between q_3 and q_{m-2}. Consider the following play of the game $\mathcal{G}(2, 1)(u, v)$ (illustrated in the diagram below). Player *I* in the first move chooses q_2 and q'. Player *II* should choose an occurrence of a before the first occurrence of w_1 in u (which is in u_2) because in v_0av_1 there is no occurrence of w_1 and, since he needs at least one a before the occurrence of a that he chooses, he has to choose p_2. *II* also needs at least one a between and after the positions that he chooses. Player *II* cannot win this play of the game, a contradiction on the fact that $u \sim_{(2,1)} v$ (*II* cannot win, since there is no occurrence of w_1 between the positions chosen by player *I* in the first move, but there is an occurrence of w_1 between the positions chosen by player *II* in the first move). Hence $j = 1$ is eliminated. (Remark. That $j = 1$ is eliminated can also be seen by considering the play of the game $\mathcal{G}(2, 1)(u, v)$, where player *I* in the first move chooses q_1 and q_3. There is no occurrence of w_1 between q_1 and q_3 but there is one between p_1 and p_3 or p_1 and p_4.)



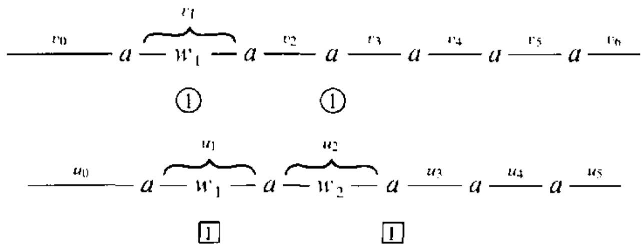


or



$j = 2$. We have that w_1 is a subword of v_0av_1 , but we do not have that w_1w_2 is a subword of v_0av_1. If $w_2 = a$, w_1w_2 is a subword of v_0av_1a and, hence, of $v_0av_1av_2$. So, assume that $w_2 \neq a$ and let p' be the first position in u after p_2 such that $Q^u_{w_2}p'$. Now, since $u_0au_1au_2 \sim_{(2)} v_0av_1 \ldots av_{m-3}$, w_2 occurs between q_2 and q_{m-2}. Let q' be the first position in v after q_2 such that $Q^v_{w_2}q'$. Suppose that q' is not between q_2 and q_3 in v. If the first occurrence of

w_1 in v is in v_1 (and hence in u_1 by Lemma 4.1(1)), consider the following play of the game $\mathcal{G}(2, 1)(u, v)$ (illustrated in the diagram below). Player *I* in the first move chooses the first occurrence of w_1 in v and q_3 in v. Player *II* cannot win this play of the game, a contradiction on the fact that $u \sim_{(2,1)} v$ (*II* cannot win, since there is no w_2 between the positions chosen by player *I* in the first move, but there is an occurrence of w_2 between the positions chosen by player *II* in the first move):

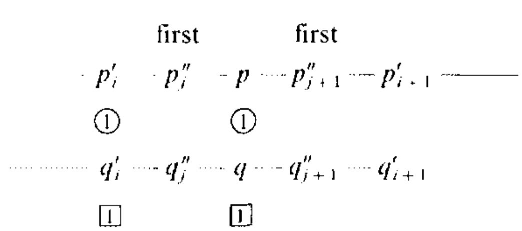


or

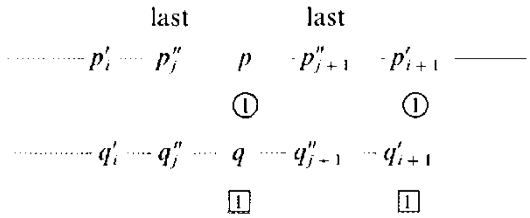
If the first occurrence of w_1 in v is in v_0a, player *I* in the first move chooses q_1 and q_3 in v. Player *II* cannot win this play of the game, for the same reason as above. Hence q' should be between q_2 and q_3.

Case 4. $|u|_a > 5$, $|v|_a > 5$. Let $p'_1, \ldots, p'_s$ in u ($p'_1 < \ldots < p'_s$) ($q'_1, \ldots, q'_s$ in v ($q'_1 < \ldots < q'_s$)) be the positions which spell the first and the last occurrences of every letter in u (v) satisfying (2, 3, 4, 5, 6) of Lemma 4.5. Now if p is any middle position in u (among $p_3, \ldots, p_{n-2}$) chosen by player *I* in the first move, then $p \in u(p'_i, p'_{i+1})$ for some i, $1 \le i \le s — 1$. Then player *II* chooses a middle position q in v (among $q_3, \ldots, q_{m-2}$) as follows. Let $p''_1, \ldots, p''_{s_i}$ in $u(p'_i, p'_{i+1})$ ($p''_1 < \ldots < p''_{s_i}$) ($q''_1, \ldots, q''_{s_i}$ in $v(q'_i, q'_{i+1})$ ($q''_1 < \ldots < q''_{s_i}$)) be the positions which spell the first and the last occurrences of every letter in $u(p'_i, p'_{i+1})$ ($v(q'_i, q'_{i+1})$) satisfying (2, 3) of Lemma 4.6. First, if $p = p''_j$ for some j, $1 \le j \le s_i$, then let $q = q''_j$; $u[1, p) \sim_{(2)} v[1, q)$ and $u(p, |u|] \sim_{(2)} v(q, |v|]$ follow from Lemma 4.6(3). Second, if $p \in u(p''_j, p''_{j+1})$ for some j, $1 \le j \le s_i — 1$, then q will be chosen according to the following rules, rules 1 to 4, which describe different plays of the game $\mathcal{G}(2, 1)(u, v)$. Rules 1 to 4 depend on p''_j and p''_{j+1} being first or last occurrences of letters in $u(p'_i, p'_{i+1})$ (Remark. It can happen that, for example, p''_j is both a first and a last occurrence of a letter; in such a case, q will be chosen according to any of the rules that apply). We will illustrate the plays by diagrams. The first move of *I* will be indicated as before by [circle with 1 in the middle] and the first move of *II* by {square with 1 in the middle].

Rule 1. Rule 1 is an application of Lemma 4.5(5). If p''_j and p''_{j+1} are first occurrences of letters in $u(p'_i, p'_{i+1})$, then consider the play of the game $\mathcal{G}(2, 1)(u, v)$, where in move 1, player *I* chooses p'_i and p. Player *II* should choose q'_i and a position q in $v(q'_i, q'_{i+1})$ such that $Q^v_a q$ and $u(p'_i, p) \sim_1 v(q'_i, q)$. Since p''_j and p''_{j+1} (and hence (q''_j and q''_{j+1}) are first occurrences of letters in $u(p'_i, p'_{i+1})$ ($v(q'_i, q'_{i+1})$), q must be in $v(q''_j, q''_{j+1})$ (otherwise there would be contradiction with $u(p'_i, p) \sim_1 v(q'_i, q)$). More precisely, q is not in $v(q'_i, q''_j)$ and $q \ne q''_j$ since otherwise there would be an occurrence of the letter of p''_j in $u(p'_i, p)$ but not in $v(q'_i, q)$; q is not in $v(q''_{j+1}, q'_{i+1})$, since otherwise there would be an occurrence of the letter of q''_{j+1} in $v(q'_i, q)$ but not in $u(p'_i, p)$; $q \ne q''_{j+1}$ since otherwise $Q^v_a q''_{j+1}$ and, hence, $Q^u_a p''_{j+1}$, contradicting the fact that p''_{j+1} is the first occurrence of a letter in $u(p'_i, p'_{i+1})$ ($Q^u_a p$ and $p < p''_{j+1}$):

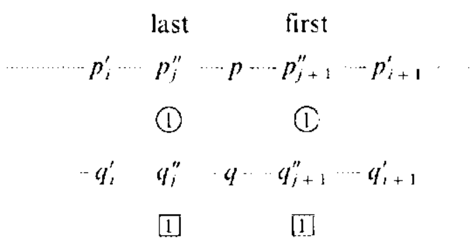


Rule 2. Rule 2 is an application of Lemma 4.5(5). If p''_j and p''_{j+1} are last occurrences of letters in $u(p'_i, p'_{i+1})$, then player *I*, in the first move chooses p and p'_{i+1}. Player *II* should choose q'_{i+1} and a position q in $v(q'_i, q'_{i+1})$ such that $Q^v_a q$ and $u(p, p'_{i+1}) \sim_1 v(q, q'_{i+1})$. Similarly as in Case 1, q must be in $v(q''_j, q''_{j+1})$:

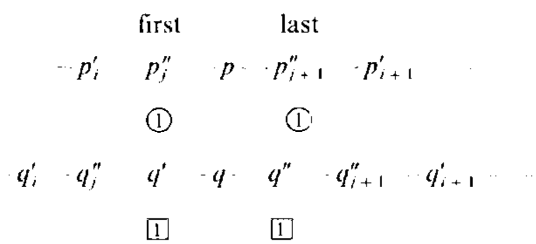


Rules 3 and 4 are applications of Lemma 4.5(6).

Rule 3. If p''_j is the last occurrence of a letter in $u(p'_i, p'_{i+1})$ and p''_{j+1} is the first occurrence of a letter in $u(p'_i, p'_{i+1})$, then player *I*, in the first move chooses p''_j and p''_{j+1}. Hence there exist q' and q'' in $v(q'_i, q'_{i+1})(q' < q'')$ such that $Q^v_b q'$ if and only if $Q^u_b p''_j$ if and only if $Q^v_b q''_j$, $Q^v_b q''$ if and only if $Q^u_b p''_{j+1}$ if and only if $Q^v_b q''_{j+1}$, $b \in A$ and $u(p''_j , p''_{j+1}) \sim_1 v(q' , q'')$. $q' \leq q''_j$ (since p''_j is the last occurrence of the letter of q' and p''_j in $v(q'_i, q'_{i+1})$) and $q''_{j+1} \leq q''$ (since q''_{j+1} c is the first occurrence of the letter of q'' and q''_{j+1} in $v(q'_i, q'_{i+1})$); $q' < q''_j$ or $q''_{j+1} < q''$ would contradict $u(p''_j, p''_{j+1}) \sim_1 v(q', q'')$. More precisely, $q' < q''_j$ ($q''_{j+1} < q''$) would imply an occurrence of the letter of q''_j (q''_{j+1}) in $v(q', q'')$ but there is no such occurrence in $u(p''_j, p''_{j+1})$. Hence $q' = q''_j$ and $q'' = q''_{j+1}$. Since $u(p''_j, p_j^{+1''}) \sim_1 v(q''_j , q''_{j+1})$, there exists q in $v(q''_j , q''_{j+1})$ such that $Q^v_a q$:



Rule 4. If p''_j is the first occurrence of a letter in $u(p'_i, p'_{i+1})$ and p'_{j+1} is the last occurrence of a letter in $u(p'_i, p'_{i+1})$, then player *I*, in the first move chooses p''_j and p''_{j+1}. Hence there exist q' and q'' such that $q''_j \leq q' < q'' \leq q''_{j+1}$ and satisfying $Q^v_b q'$ if and only if $Q^u_b p''_j$ if and only if $Q^v_b q''_j$, $Q^v_b q''$ if and only if $Q^u_b p''_{j+1}$ if and only if $Q^v_b q''_{j+1}$, $b \in A$, and $u(p''_j , p''_{j+1}) \sim_1 v(q', q'')$. Since $u(p''_j, p''_{j+1}) \sim_1 v(q', q'')$, there exists q in $v(q', q'')$ such that $Q^v_a q$:



In Rules 1 to 4, the facts that $u[1, p) \sim_{(2)} v[1,q)$ and $u(p, |u|] \sim_{(2)} v(q, |v|]$ will follow similarly as Lemma 4.6(3). We show $u(p, |u|] \sim_{(2)} v(q, |v|]$ for Rule 4. Let $w = w_1 \ldots w_{|w|}$ be a subword of length ≤ 2 in $v(q, |v|]$ (it is similar if in $u(p, |u|]$). We want to show that w is a sub-word of $u(p, |u|]$. If $|w| = 1$, then there is an occurrence of w_1 in $v[q'_{i+1} \; |v|]$ (and hence in $u[p'_{i+1}, |u|]$) from the choice of the p'_r's and the q'_r's and Lemma 4.5(1, 2) and the proof is complete. If $|w| = 2$, and w is in $v(q'_{i+1}, |v|]$, then w is in $u(p'_{i+1}, |u|]$ by Lemma 4.5(3). If there is an occurrence of w_2 in $v(q'_{i+1}, |v|]$ (and, hence, in $u(p'_{i+1}, |u|]$ by Lemma 4.5(3)) and $Q^v_{w_1} q'_{i+1}$ (and hence $Q^u_{w_1} p'_{i+1}$ by Lemma 4.5(2)) the proof is complete. Otherwise, there is an occurrence of w_2 in $v[q'_{i+1}, |v|]$ (and, hence, in $u[p'_{i+1}, |u|]$) from the choice of the p'_r's and the q'_r's and Lemma 4.5(1, 2) and there is also an occurrence of w_1 in $v(q, q'_{i+1})$. From the choice of the q''_r's, there exists k, $k \geq j+1$, such that $Q^v_{w_1} q''_k$. Hence, from the choice of the p''_r 's and Lemma 4.6(1, 2), $Q^u_{w_1} p''_k$. The result follows.

REFERENCES

1. F. Blanchet-Sadri, "Some Logical Characterizations of the Dot-Depth Hierarchy and Applications," Ph.D. thesis, McGill University. May 1989.
2. F. Blanchet-Sadri, Games, equations and the dot-depth hierarchy, Comput. Math. App!. 18 (1989), 809-822.
3. F. Blanchet-Sadri, On dot-depth two, RAIRO Inform. Theor. Appl. 24 (1990), 521-529.
4. F. Blanchet-Sadri, Games, equations and dot-depth two monoids, Discrete App!. Math. 39 (1992), 99 III.
5. F. Blanchet-Sadri, Equations and dot-depth one, Semigroup Forum 47 (1993), 305-317.
6. J. A. Brzozowski and R. Knast, The dot-depth hierarchy of star-free languages is infinite, J. Comput. System Sci. 16 (1978), 37-55.
7. J. A. Brzozowski and I. Simon, Characterizations of locally testable events, Discrete Math. 4 (1973), 243 271.
8. J. R. Bilchi, Weak second-order arithmetic and finite automata, Z. Math. Logik Grundlag. Math. 6 (1960), 66 92.
9. R. S. Cohen and J. A. Brzozowski, Dot-depth of star-free events, J. Comput. System Sci. 5 (1971), 1-15.
10. A. Ehrenfeucht, An application of games to the completeness problem for formalized theories, Fund. Math. 49 (1961), 129-141.
11. S. Eilenberg, "Automata, Languages and Machines," Vol. B, Academic Press, New York, 1976.
12. C. C. Elgot, Decision problems of finite automata design and related arithmetics, Trans. Amer. Math. Soc. 98 (1961), 21-52.
13. H. B. Enderton, "A Mathematical Introduction to Logic," Academic Press, New York, 1972.
14. R. Fraisse, "fours de Logique Mathematique," Tome 2, Gauthiers Villars, Paris, 1972.
15. S. C. Kleene, Representation of events in nerve nets and finite automata, in "Automata Studies, pp. 3-42," Princeton Univ. Press, Princeton, 1956.
16. R. E. Ladner, Application of model theoretic games to discrete linear orders and finite automata, Inform. Control 33 (1977), 281-303.
17. G. Lallement, "Semigroups and Combinatorial Applications," Wiley, New York, 1979.
18. R. McNaughton, Algebraic decision procedures for local testability, J. Math. Systems Theory 8 (1974), 64-76.
19. R. McNaughton and S. Papert, "Counter-free Automata," MIT Press, Cambridge, MA, 1971.
20. D. Perrin and J.-E. Pin, First-order logic and star-free sets, J. Comput. System Sci. 32 (1986), 393-406.
21. J.-E. Pin, "Varietes de Langages Formels," Masson, Paris, 1984. ("Varieties of Formal Languages," Plenum, London, 1986.)
22. J.-E. Pin, Hierarchies de concatenation, RAIRO Infirm. Theor. 18 (1984), 23-46.
23. C. Reutenauer, Sur les varietes de langages et de monoides, in "Lect. Notes in Comput. Sci., Vol. 67, pp. 260 265," Springer-Verlag, Berlin, 1979.
24. .1. G. Rosenstein, "Linear Orderings," Academic Press, New York, 1982.
25. M. P. Schützenberger, On finite monoids having only trivial subgroups, Inform. Control 8 (1965), 190-194.
26. I. Simon, Piecewise testable events, in "Proceedings, Second Cl Conf.," Lect. Notes in Comp. Sci. 33, Springer-Verlag, Berlin, 1975,214-222.
27. H. Straubing, A generalization of the Schützenberger product of finite monoids, Theoret. Comput. Sci. 13 (1981), 137 150.
28. H. Straubing, Finite semigroup varieties of the form V* D, J. Pure Appl. Algebra 36 (1985), 53-94.
29. H. Straubing, Semigroups and languages of dot-depth two, in "Proceedings, Thirteenth ICALP," Lect. Notes in Comput. Sci., Vol. 226, pp. 416-423, Springer-Verlag, New York, 1986.
30. W. Thomas, Classifying regular events in symbolic logic, J. Comput. System Sci. 25 (1982), 360-376.
31. W. Thomas, An application of the Ehrenfeucht Fraisse game in formal language theory, Bull. Soc. Math. France 16 (1984), 11-21.
32. B. Tilson, Categories as algebra, J. Pure App!. Algebra 48 (1987), 83-198.