

## On a complete set of generators for dot-depth two\*

F. Blanchet-Sadri\*\*

*Department of Mathematics, University of North Carolina, Greensboro, NC 27412, USA*

Received 27 August 1991

---

### Abstract

A complete set of generators for Straubing's dot-depth-two monoids has been characterized as a set of quotients of the form  $A^*/\sim_{(n,m)}$ , where  $n$  and  $m$  denote positive integers,  $A^*$  denotes the free monoid generated by a finite alphabet  $A$ , and  $\sim_{(n,m)}$  denote congruences related to a version of the Ehrenfeucht–Fraïssé game. This paper studies combinatorial properties of the  $\sim_{(n,m)}$ 's and in particular the inclusion relations between them. Several decidability and inclusion consequences are discussed.

---

### 1. Introduction

This paper deals with the problem of the decidability of the different levels of the dot-depth hierarchy and in particular with its second level. The problem is a central one in language theory. Its study is justified by its recognized connections with the theory of automata, formal logic and circuit complexity. The method used relies on a game-theoretical approach that was introduced by Thomas [19] and used in [1–5].

In Section 1.1, we recall the basic definitions and well-known results concerning the dot-depth hierarchy. Section 1.2 includes the definition of the game that is used in the proofs of our results and Section 1.3 presents a summary of our main results in this paper.

#### 1.1. The dot-depth hierarchy

First, notation and basic concepts are introduced in order to define the decidability problem of the dot-depth hierarchy.

---

\* This material is based upon work supported by the National Science Foundation under Grant No. CCR-9101800. The Government has certain rights in this material.

\*\* E-mail: blanchet@steffi.acc.uncg.edu.

Let  $A$  be a finite set of letters. The regular languages over  $A$  are those subsets of  $A^*$ , the free monoid generated by  $A$ , constructed from the finite languages over  $A$  by the boolean operations, the concatenation product and the star. The star-free languages are those regular languages which can be obtained from the finite languages by the boolean operations and the concatenation product only. According to Schützenberger [14],  $L \subseteq A^*$  is star-free if and only if its syntactic monoid  $M(L)$  is finite and aperiodic (or  $M(L)$  contains no nontrivial subgroups). General references on the star-free languages are [9, 10, 12].

Natural classifications of the star-free languages are obtained based on the alternating use of the boolean operations and the concatenation product. Classes of languages  $A^*\mathcal{V}_0, A^*\mathcal{V}_1, \dots$  introduced by Straubing in [16] form a hierarchy which is closely related to the so-called *dot-depth hierarchy* introduced by Cohen and Brzozowski in [7]:  $A^*\mathcal{V}_0$  consists of the empty set and  $A^*$ , and  $A^*\mathcal{V}_{k+1}$  denotes the class of languages over  $A$  which are boolean combinations of languages of the form  $L_0a_1L_1a_2 \dots a_mL_m$  ( $m \geq 0$ ) with  $L_0, \dots, L_m \in A^*\mathcal{V}_k$  and  $a_1, \dots, a_m \in A$ . Let  $A^*\mathcal{V} = \bigcup_{k \geq 0} A^*\mathcal{V}_k$ .  $L \subseteq A^*$  is star-free if and only if  $L \in A^*\mathcal{V}_k$  for some  $k \geq 0$ . The *dot-depth* of  $L$  is defined as the smallest such  $k$ .

The fact that the dot-depth hierarchy is infinite has long been known [6], i.e.  $A^*\mathcal{V}_{k+1} \not\subseteq A^*\mathcal{V}_k$  for every  $k$  (a proof, using games, is given in [20]). The question of effectively determining the dot-depth of a given star-free language remains open. Simon [15] has shown that one can decide if a given language has dot-depth one, and Straubing [17] gave a decision procedure for  $k = 2$  but which works only for an  $A$  with two letters. Straubing conjectured that his algorithm works for an arbitrary  $A$ . Results relative to the characterization of dot-depth-two languages are the subject of this paper.

For  $k \geq 1$ , let us define subhierarchies of  $A^*\mathcal{V}_k$  as follows: for all  $n \geq 1$ , let  $A^*\mathcal{V}_{k,n}$  denote the class of boolean combinations of languages of the form  $L_0a_1L_1a_2 \dots a_mL_m$  ( $0 \leq m \leq n$ ), with  $L_0, \dots, L_m \in A^*\mathcal{V}_{k-1}$  and  $a_1, \dots, a_m \in A$ . We have  $A^*\mathcal{V}_k = \bigcup_{n \geq 1} A^*\mathcal{V}_{k,n}$ . Easily,  $A^*\mathcal{V}_{k,n} \subseteq A^*\mathcal{V}_{k+1,n}$  and  $A^*\mathcal{V}_{k,n} \subseteq A^*\mathcal{V}_{k,n+1}$ .

The Straubing hierarchy gives examples of  $*$ -varieties of languages. One can show that  $\mathcal{V}, \mathcal{V}_k$  and  $\mathcal{V}_{k,n}$  are  $*$ -varieties of languages. According to Eilenberg, there exist monoid varieties  $V, V_k$  and  $V_{k,n}$  corresponding to  $\mathcal{V}, \mathcal{V}_k$  and  $\mathcal{V}_{k,n}$ , respectively.  $V$  is the variety of aperiodic monoids. We have that, for  $L \subseteq A^*$ ,  $L \in A^*\mathcal{V}$  if and only if  $M(L) \in V$ , for each  $k \geq 0$ ,  $L \in A^*\mathcal{V}_k$  if and only if  $M(L) \in V_k$  and, for  $k \geq 1, n \geq 1$ ,  $L \in A^*\mathcal{V}_{k,n}$  if and only if  $M(L) \in V_{k,n}$ . The dot-depth-two problem reduces to characterizing effectively the monoids which are in  $V_2$  but not in  $V_1$ .

### 1.2. The Ehrenfeucht–Fraïssé game technique

Our contributions related to the decidability problem of the dot-depth hierarchy were obtained by carefully exploiting the Ehrenfeucht–Fraïssé game technique. This method was used by Thomas [19] to give a new proof of the infinity of the dot-depth hierarchy.

First, one regards a word  $w \in A^*$  of length  $|w|$  as a word model  $w = \langle \{1, \dots, w\}, <^w, (Q_a^w)_{a \in A} \rangle$ , where the universe  $\{1, \dots, |w|\}$  represents the set of positions of letters in  $w$ ,  $<^w$  denotes the  $<$ -relation in  $w$ , and  $Q_a^w$  are unary relations over  $\{1, \dots, |w|\}$

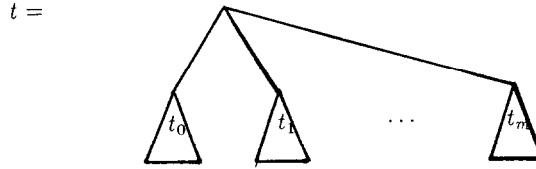
containing the positions with letter  $a$ , for each  $a \in A$ . To any  $k$ -tuple  $\bar{m} = (m_1, \dots, m_k)$  of positive integers, where  $k \geq 0$ , and any words  $u$  and  $v$  from  $A^*$  corresponds a game  $\mathcal{G}_{\bar{m}}(u, v)$  which is played between two players I and II on the word models  $u$  and  $v$ . A play of the game consists of  $k$  moves. In the  $i$ th move, player I chooses, in  $u$  or in  $v$ , a sequence of  $m_i$  positions of letters; then player II chooses, in the remaining word, also a sequence of  $m_i$  positions of letters. After  $k$  moves, by concatenating the position sequences chosen from  $u$  and  $v$ , two sequences of positions  $p_1 \dots p_n$  ( $p_1 \leq \dots \leq p_n$ ) from  $u$  and  $q_1 \dots q_n$  ( $q_1 \leq \dots \leq q_n$ ) from  $v$  have been formed where  $n = m_1 + \dots + m_k$ . Player II has won the play if the two subwords (a word  $a_1 \dots a_n$  is a subword of a word  $w$  if there exist words  $w_0, \dots, w_n$  such that  $w = w_0 a_1 w_1 a_2 \dots a_n w_n$ ) given by the position sequences  $p_1 \dots p_n$  and  $q_1 \dots q_n$  coincide. One writes that  $u \sim_{\bar{m}} v$  if player II has a winning strategy to win each play of the game  $\mathcal{G}_{\bar{m}}(u, v)$ .  $\sim_{\bar{m}}$  naturally defines a congruence on  $A^*$ . The importance of  $\sim_{\bar{m}}$  lies in the fact that  $V_k$  can be characterized in terms of the monoids  $A^*/\sim_{(m_1, \dots, m_k)}$ . Thomas [18, 19] and Perrin and Pin [11] infer that, for  $k \geq 1$ ,  $M \in V_k$  if and only if there exists  $\bar{m} = (m_1, \dots, m_k)$  such that  $M$  divides  $A^*/\sim_{\bar{m}}$ , or, more precisely, for  $k \geq 1$ ,  $n \geq 1$ ,  $M \in V_{k,n}$  if and only if there exists  $\bar{m} = (n, m_1, \dots, m_{k-1})$  such that  $M$  divides  $A^*/\sim_{\bar{m}}$  ( $V_k = \{A^*/\sim \mid \sim \supseteq \sim_{(m_1, \dots, m_k)}\}$  for some  $(m_1, \dots, m_k)$  and  $V_{k,n} = \{A^*/\sim \mid \sim \supseteq \sim_{(n, m_1, \dots, m_{k-1})}\}$  for some  $m_1, \dots, m_{k-1}$ ). Hence the monoids  $A^*/\sim_{\bar{m}}$  form a class of monoids that generate  $V$  in the sense that every finite aperiodic monoid is a morphic image of a submonoid of a monoid of the form  $A^*/\sim_{\bar{m}}$ .

### 1.3. Decidability and inclusion results

Reference [5] characterizes the dot-depth-two monoids of the form  $A^*/\sim_{\bar{m}}$ :  $A^*/\sim_{(m_1, \dots, m_k)}$  is of dot-depth two if and only if  $k = 2$ , or  $k = 3$  and  $m_2 = 1$ . This paper studies dot-depth-two monoids, and in particular the monoids  $A^*/\sim_{(n, m)}$  and  $A^*/\sim_{(n, 1, m)}$ . More specifically, in Sections 2, 3 and 4 we study combinatorial properties of the  $\sim_{(n, m)}$ 's and  $\sim_{(n, 1, m)}$ 's and the inclusion relations between them. Studying properties of the  $\sim_{(n, m)}$ 's and  $\sim_{(n, 1, m)}$ 's sheds some light on the dot-depth-two syntactic monoids. Several decidability and inclusion consequences are discussed in Section 5.

Reference [3] shows that, for  $|A| = 2$ ,  $A^*\mathcal{V}_{2,1} \subsetneq A^*\mathcal{V}_{2,2} = A^*\mathcal{V}_{2,3} \subsetneq A^*\mathcal{V}_{2,4} = A^*\mathcal{V}_{2,5} \subsetneq \dots$  and, for  $|A| \geq 3$ ,  $A^*\mathcal{V}_{2,1} \subsetneq A^*\mathcal{V}_{2,2} \subsetneq A^*\mathcal{V}_{2,3} \subsetneq \dots$  (here  $|A|$  denotes the cardinality of  $A$ ). Let  $A^*/\sim_{\bar{m}}$  be of dot-depth two. The 2-dot-depth (abbreviated *2dd*) of  $A^*/\sim_{\bar{m}}$  is defined as the smallest  $n$  for which  $A^*/\sim_{\bar{m}} \in V_{2,n}$ . In Section 5.1, we show that the 2-dot-depth of all the generators  $A^*/\sim_{(n, m)}$  can be determined from the values of  $|A|$ ,  $n$  and  $m$ , and the 2-dot-depth of all the monoids of the form  $A^*/\sim_{(n, 1, m)}$  where  $|A| = 2$  can be determined from the value of  $n$ . An upper bound for the 2-dot-depth of all the monoids of the form  $A^*/\sim_{(n, 1, m)}$  where  $|A| \geq 3$  is given.

Section 5.2 deals with a conjecture of Pin. A special case of one of the results in this paper implies that a conjecture of Pin concerning tree hierarchies of monoids (the dot-depth and the Straubing hierarchies being particular cases) is false. More precisely,  $\{\emptyset, A^*\}$  is associated with the tree reduced to a point. Then to the tree



is associated the boolean algebra  $\mathcal{V}_t$  which is generated by all the languages of the form  $L_{i_0}a_1L_{i_1}a_2 \dots a_rL_{i_r}$ , with  $0 \leq i_0 < \dots < i_r \leq m$ , where, for  $0 \leq j \leq r$ ,  $L_{i_j} \in \mathcal{V}_{t_{i_j}}$ . Pin [13] conjectured that  $\mathcal{V}_t \subseteq \mathcal{V}_{t'}$  if and only if  $t$  is extracted from  $t'$ .

Section 5.3 contains some results on equations. Equations are used to define monoid varieties. Abstract arguments show that every monoid variety can be ultimately defined by a sequence of equations. We give equations satisfied in the monoid varieties  $V_{2,n}$ .

In Section 5.4, generalizations of the inclusion results of Section 4 to arbitrary  $\sim_{(m_1, \dots, m_k)}$  are discussed.

The reader is referred to [8, 12] for all the algebraic and logical terms not defined in this paper. In the following sections, we assume  $|A| \geq 2$  (unless otherwise stated).  $A^+$  will denote  $A^* \setminus \{1\}$ , where 1 denotes the empty word;  $|w|_a$  ( $w\alpha$ ) the number of occurrences of the letter  $a$  in a word  $w$  (the set of letters in a word  $w$ );  $m_1, \dots, m_k$ ,  $n_1, \dots, n_{k'}$ ,  $m, m', n$  and  $n'$  positive integers,  $\lfloor x \rfloor$  the largest integer smaller than or equal to  $x$  and  $\lceil x \rceil$  the smallest integer larger than or equal to  $x$ .

## 2. Some basic properties of the congruences $\sim_{\bar{m}}$

### 2.1. An induction lemma

This section is concerned with an induction lemma for the  $\sim_{\bar{m}}$ 's.

In what follows, if  $w = a_1 \dots a_n$  is a word and  $1 \leq i \leq j \leq n$ ,  $w[i, j]$ ,  $w(i, j)$ ,  $w[i, j]$  and  $w[i, j]$  will denote the segments  $a_i \dots a_j$ ,  $a_{i+1} \dots a_{j-1}$ ,  $a_{i+1} \dots a_j$  and  $a_i \dots a_{j-1}$ , respectively.

**Lemma 2.1.**  $u \sim_{(n, \bar{m})} v$  if and only if

• for every  $p_1, \dots, p_n \in u$  ( $p_1 \leq \dots \leq p_n$ ), there exist  $q_1, \dots, q_n \in v$  ( $q_1 \leq \dots \leq q_n$ ) such that

- (1)  $Q_a^u p_i$  if and only if  $Q_a^v q_i$ ,  $a \in A$  for  $1 \leq i \leq n$ ,
- (2)  $u[1, p_1] \sim_{\bar{m}} v[1, q_1]$ ,
- (3)  $u(p_i, p_{i+1}) \sim_{\bar{m}} v(q_i, q_{i+1})$  for  $1 \leq i \leq n-1$ ,
- (4)  $u(p_n, |u|] \sim_{\bar{m}} v(q_n, |v|]$ , and

• for every  $q_1, \dots, q_n \in v$  ( $q_1 \leq \dots \leq q_n$ ), there exist  $p_1, \dots, p_n \in u$  ( $p_1 \leq \dots \leq p_n$ ) such that (1)–(4) hold.

### 2.2. An inclusion lemma

This section is concerned with an inclusion lemma which gives conditions which insure  $\sim_{(n_1, \dots, n_{k'})}$  to be included in  $\sim_{(m_1, \dots, m_k)}$ . A trivial condition is the following:  $k \leq k'$  and there exist  $1 \leq i_1 < \dots < i_k \leq k'$  such that  $m_1 \leq n_{i_1}, \dots, m_k \leq n_{i_k}$ .

Define  $\mathcal{N}(m_1, \dots, m_k) = (m_1 + 1) \cdots (m_k + 1) - 1$ . We can show that  $x^N \sim_{(m_1, \dots, m_k)} x^{N+1}$  ( $N = \mathcal{N}(m_1, \dots, m_k)$ ) and that  $N$  is the smallest  $n$  such that  $x^n \sim_{(m_1, \dots, m_k)} x^{n+1}$  for  $|x| = 1$ . It follows that if  $u, v \in A^*$  and  $u \sim_{(m_1, \dots, m_k)} v$ , then  $|u|_a = |v|_a < \mathcal{N}(m_1, \dots, m_k)$  or  $|u|_a, |v|_a \geq \mathcal{N}(m_1, \dots, m_k)$ . The following lemma follows easily from Lemma 2.1 and the above remarks.

**Lemma 2.2.**  $\sim_{(m_1, \dots, m_k)} \subseteq \sim_{(\mathcal{N}(m_1, \dots, m_k))}$  and  $\sim_{(m_1, \dots, m_k)} \not\subseteq \sim_{(\mathcal{N}(m_1, \dots, m_k) + 1)}$ . Consequently, a necessary condition for  $\sim_{(n_1, \dots, n_k)}$  to be included in  $\sim_{(m_1, \dots, m_k)}$  is  $\mathcal{N}(m_1, \dots, m_k) \leq \mathcal{N}(n_1, \dots, n_k)$ . Moreover, if  $k \leq k'$  and there exist  $0 = j_0 < \dots < j_{k-1} < j_k = k'$  such that  $m_i \leq \mathcal{N}(n_{j_{i-1}+1}, \dots, n_{j_i})$  for  $1 \leq i \leq k$ , then  $\sim_{(n_1, \dots, n_{k'})} \subseteq \sim_{(m_1, \dots, m_k)}$ .

### 3. Some positions of a word $w$

This section is concerned with some positions of a word  $w$ , i.e.  $(m)$  positions,  $(m, m')$  positions where  $m > m'$ , and  $(m)_r$  positions where  $|A| = r$ .

#### 3.1. $(m)$ positions

We give induction lemmas for the  $\sim_{(n, m)}$ 's after defining the positions which spell the first occurrences of every subword of length  $\leq m$  of a word  $w$  (or the  $(m)$  first positions in  $w$ ).

Let  $w \in A^+$  and let  $w_1$  denote the smallest prefix of  $w$  such that  $w_1\alpha = w\alpha$  (call the last position of  $w_1$ ,  $p_1$ ); let  $w_2$  denote the smallest prefix of  $w(p_1, |w|]$  such that  $w_2\alpha = (w(p_1, |w|])\alpha$  (call the last position of  $w_2$ ,  $p_2$ ); ... let  $w_m$  denote the smallest prefix of  $w(p_{m-1}, |w|]$  such that  $w_m\alpha = (w(p_{m-1}, |w|])\alpha$  (call the last position of  $w_m$ ,  $p_m$ ).

If  $|w\alpha| = 1$ ,  $p_1, \dots, p_m$  are the  $(m)$  first positions in  $w$  and the procedure terminates. If  $|w\alpha| > 1$ ,  $p_1, \dots, p_m$  are among the  $(m)$  first positions in  $w$ . To find the others, we repeat the process to find the  $(m)$  first positions in  $w[1, p_1]$  and the  $(m - i + 1)$  first positions in  $w(p_{i-1}, p_i)$  for  $2 \leq i \leq m$ .

We can define similarly the positions which spell the last occurrences of every subword of length  $\leq m$  of  $w$  (or the  $(m)$  last positions in  $w$ ). The  $(m)$  first and the  $(m)$  last positions in  $w$  are called the  $(m)$  positions in  $w$ . The number of  $(m)$  positions in  $w$  is bounded above by  $2\mathcal{J}(m)$ , where  $\mathcal{J}(m)$  denotes the index of  $\sim_{(m)}$ .

Consider the following example: let  $A = \{a, b, c\}$  and

$$w = \overline{abbbbbb} \overline{aa} \overline{abbb} \overline{ba} \overline{aaab} \overline{cb} \overline{cc} \overline{cb} \overline{bbbbb} \overline{cb} \overline{bbcb} \overline{ab} \overline{ab} \overline{aa} \overline{abbb} \overline{cc} \overline{cc} \overline{cb} \overline{cb} \overline{bbb} \overline{ab} \overline{cb} \overline{bb} \overline{bb} \overline{cb} \overline{cb} \overline{ab} \overline{bb}.$$

The overlined positions of  $w$  are the  $(5)$  first positions in  $w$ .

The following facts hold.

**Fact 1:** If  $p$  is among the  $(m)$  first positions of  $w$ , then  $w[1, p]$  can be divided (using the above process) into at most  $m$  segments  $w_1, \dots, w_m$  whose last positions, say  $p_1, \dots, p_{m-1}, p$  ( $p_1 \leq \dots \leq p_{m-1} \leq p$ ), spell the first occurrence of a subword of length

$\leq m$  of  $w$  (also,  $p_1, \dots, p_{m-1}$  are among the positions which spell such occurrences). It is clear that  $w[1, p] \not\sim_{(m)} w[1, p]$ , and if  $p_1 < \dots < p_{m-1} < p$ , then  $w[1, p] \sim_{(m)} w[1, p]v$  if and only if  $v\alpha \subseteq w_m\alpha$ , where  $v \in A^*$ .

**Fact 2:** If  $p$  is among the  $(m)$  first positions of  $w$ , then consider the decomposition of  $w[1, p]$  as in fact 1. If  $p_1 < \dots < p_{m-1} < p$ , then  $p$  is the first occurrence of its letter in  $w(p_{m-1}, |w|]$ . If, in addition,  $q \in w(p_{m-1}, p)$  and  $q$  is also among the  $(m)$  first positions of  $w$ , then  $q$  is the first occurrence of its letter in  $w(p_{m-1}, |w|]$ .

**Fact 3:** If  $p$  and  $q$  ( $p < q$ ) are among the  $(m)$  first positions of  $w$ , and there is no such position between  $p$  and  $q$ , then there exist  $p_1, \dots, p_{m-1} \in w$  ( $p_1 < \dots < p_{m-1} < p$ ) such that  $p_1, \dots, p_{m-1}, p$  and  $p_1, \dots, p_{m-1}, q$  spell the first occurrences of subwords of length  $\leq m$  of  $w$ , or  $w(p, q) = 1$  and there exist  $p_1, \dots, p_{m-1} \in w$  ( $p_1 \leq \dots \leq p_{m-1} = p$ ) such that  $p_1, \dots, p_{m-1}, q$  spell the first occurrence of a subword of length  $\leq m$  of  $w$  (also, in both situations,  $p_1, \dots, p_{m-1}$  are among the positions which spell such occurrences). To see this, consider the decomposition of  $w[1, q]$  as in fact 1. Fact 2 implies the result.

Similar facts hold for the  $(m)$  last positions of  $w$ .

**Lemma 3.1.** Let  $u, v \in A^+$  be such that  $u \sim_{(n, n')} v$  and let  $p_1, \dots, p_t \in u$  ( $p_1 < \dots < p_t$ ) ( $q_1, \dots, q_r \in v$  ( $q_1 < \dots < q_r$ )) be the  $(m)$  positions in  $u$  ( $v$ ), where  $m \leq \lfloor (\mathcal{N}(n, n') - 1)/2 \rfloor$ . Then

- $t = t'$ .
  - $Q_a^u p_i$  if and only if  $Q_a^v q_i$ ,  $a \in A$  for  $1 \leq i \leq t$ .
- Here  $p_1 = q_1 = 1$ ,  $p_t = |u|$  and  $q_r = |v|$ .

**Proof.** We show the result for  $m = \lfloor (\mathcal{N}(n, n') - 1)/2 \rfloor$  (the proof is similar for the other values of  $m$ ). For each  $1 \leq i \leq t$ , there exist  $1 \leq i_1 \leq \dots \leq i_{m-1} \leq i$  (or  $i \leq i_{m-1} \leq \dots \leq i_1 \leq t$ ) such that  $p_{i_1}, \dots, p_{i_{m-1}}, p_i$  (or  $p_i, p_{i_{m-1}}, \dots, p_{i_1}$ ) spell the first (or the last) occurrence of a subword of length  $\leq m$  in  $u$ . The result follows from Lemma 2.1 by considering different plays of the game  $\mathcal{G} = \mathcal{G}_{(n, n')}(u, v)$ . In a first round of games, one for each  $1 \leq i \leq t$ , player I, in the first move, among  $p_{i_1}, \dots, p_{i_{m-1}}, p_i$ , chooses  $p_{i_1}, p_{i_{2n}}, \dots$ , and  $p_i$  for a total of at most  $n$  positions since  $m \leq nn'$ . In a second round of games, one for each pair  $(i, j)$  with  $1 \leq i, j \leq t$ , for the first move, player I chooses among  $p_{i_1}, \dots, p_{i_{m-1}}, p_i, p_{j_1}, \dots, p_{j_{m-1}}, p_j$ , put in linear order, the  $(n' + 1)$ th from the left, the  $(n' + 1)$ th from the right, the  $2(n' + 1)$ th from the left, the  $2(n' + 1)$ th from the right, ... for a total of no more than  $n$  positions since  $\mathcal{N}(1, m) \leq \mathcal{N}(n, n')$ . More details follow.

If  $n = 1$ , consider the plays of the game  $\mathcal{G}$  where player I, in the first move, chooses  $p_i$  for some  $1 \leq i \leq t$ . If  $n > 1$ , let  $f_1, \dots, f_r \in u$  ( $f_1 < \dots < f_r$ ) ( $f'_1, \dots, f'_{r'} \in v$  ( $f'_1 < \dots < f'_{r'}$ )) be the  $(m)$  first positions in  $u$  ( $v$ ). Then  $r = r'$ , and  $Q_a^u f_i$  if and only if  $Q_a^v f'_i$ ,  $a \in A$  for  $1 \leq i \leq r$ . To see this, for each  $1 \leq i \leq r$ , by fact 1, there exist  $1 \leq i_1 \leq \dots \leq i_{m-1} \leq i$  such that  $f_{i_1}, \dots, f_{i_{m-1}}, f_i$  spell the first occurrence of a subword of length  $\leq m$  in  $u$ . In a round of games, one for each  $1 \leq i \leq r$ , player I, in the first move, among  $f_{i_1}, \dots, f_{i_{m-1}}, f_i$ , chooses  $f_{i_1}, f_{i_{2n}}, \dots$ , and  $f_i$  for a total of no more than  $n$  positions. A similar statement is valid for the  $(m)$  last positions  $l_1, \dots, l_s \in u$  ( $l'_1, \dots, l'_{s'} \in v$  ( $l'_1 < \dots < l'_{s'}$ )) in  $u$  ( $v$ ).

The proof is complete if we show the following:

- (1)  $f_i = l_j$  if and only if  $f'_i = l'_j$ , and

(2)  $f_i < l_j$  if and only if  $f'_i < l'_j$ , for  $1 \leq i \leq r$ ,  $1 \leq j \leq s$ .

We show (2) (the proof for (1) is similar). Assume that, for some  $1 \leq i \leq r$ ,  $1 \leq j \leq s$ ,  $f_i < l_j$  but  $f'_i \geq l'_j$ . Consider the play of the game  $\mathcal{G}$  where player I, in the first move among  $f_{i_1}, \dots, f_{i_{m-1}}, f_i, l_j, l_{j_{m-1}}, \dots, l_{j_1}$ , chooses the  $(n' + 1)$ th from the left, the  $(n' + 1)$ th from the right, the  $2(n' + 1)$ th from the left, the  $2(n' + 1)$ th from the right, ... for a total of at most  $n$  positions. Call them  $r_1, \dots, r_n$  ( $r_1 \leq \dots \leq r_n$ ). There exist  $s_1, \dots, s_n$  ( $s_1 \leq \dots \leq s_n$ ) satisfying Lemma 2.1. We have that if  $r_k = f_{i_{k(n'+1)}}$ , then  $s_k \geq f'_{i_{k(n'+1)}}$ ; if  $r_{n+1-k} = l_{j_{k(n'+1)}}$ , then  $s_{n+1-k} \leq l'_{j_{k(n'+1)}}$ ; if  $r_k = f_i$ , then  $s_k \geq f'_i$ ; and if  $r_k = l_j$ , then  $s_k \leq l'_j$ .

If  $n$  is odd,  $l_j = r_{(n+1)/2}$ . Put  $r = r_{((n+1)/2)-1}$ ,  $r' = r_{((n+1)/2)+1}$ ,  $s = s_{((n+1)/2)-1}$  and  $s' = s_{((n+1)/2)+1}$ . We have that  $u(r, l_j) \sim_{(n')} v(s, s_{(n+1)/2})$ . Let the positions (among the  $< \mathcal{N}(1, m)$  positions considered) which are in  $u(r, l_j)$  be denoted by  $p'_1, \dots, p'_{n'}$  ( $p'_1 \leq \dots \leq p'_{n'}$ ) ( $p'_{n'} = f_i$ ). However, the word (of length  $\leq n'$ ) spelled by  $p'_1, \dots, p'_{n'}$  is in  $u(r, l_j)$  but not in  $v(s, s_{(n+1)/2})$  since by assumption  $f'_i \geq l'_j$ .

If  $n$  is even, put  $r = r_{n/2}$ ,  $r' = r_{(n/2)+1}$ ,  $s = s_{n/2}$  and  $s' = s_{(n/2)+1}$ . If  $n' \leq 2$ ,  $r = f_i$  and  $r' = l_j$ .  $s \geq f'_i$  and  $s' \leq l'_j$  together with  $f'_i \geq l'_j$  lead to a contradiction. If  $n' > 2$ , we have that  $u(r, r') \sim_{(n')} v(s, s')$ . Let the positions (among the  $< \mathcal{N}(1, m)$  positions considered) which are in  $u(r, r')$  be denoted by  $p'_1, \dots, p'_{n'-1}$  ( $p'_1 \leq \dots \leq p'_{n'-1}$ ). However, the word (of length  $< n'$ ) spelled by  $p'_1, \dots, p'_{n'-1}$  is in  $u(r, r')$  but not in  $v(s, s')$  since by assumption  $f'_i \geq l'_j$ .  $\square$

The following lemmas are from [5] and give necessary and sufficient conditions for  $\sim_{(n, m)}$ -equivalence.

**Lemma 3.2.** Let  $u, v \in A^+$  and let  $p_1, \dots, p_t \in u$  ( $p_1 < \dots < p_t$ ) ( $q_1, \dots, q_{t'} \in v$  ( $q_1 < \dots < q_{t'}$ )) be the  $(m)$  positions in  $u$  ( $v$ ).  $u \sim_{(1, m)} v$  if and only if

- $t = t'$ ,
- $Q_a^u p_i$  if and only if  $Q_a^v q_i$ ,  $a \in A$  for  $1 \leq i \leq t$ , and
- $u(p_i, p_{i+1}) \sim_{(1)} v(q_i, q_{i+1})$  for  $1 \leq i \leq t - 1$ .

**Lemma 3.3.** Let  $n > 1$ . Let  $u, v \in A^+$  and let  $p_1, \dots, p_t \in u$  ( $p_1 < \dots < p_t$ ) ( $q_1, \dots, q_{t'} \in v$  ( $q_1 < \dots < q_{t'}$ )) be the  $(m)$  positions in  $u$  ( $v$ ).  $u \sim_{(n, m)} v$  if and only if:

- $t = t'$ .
- $Q_a^u p_i$  if and only if  $Q_a^v q_i$ ,  $a \in A$  for  $1 \leq i \leq t$ .
- $u(p_i, p_{i+1}) \sim_{(n-2, m)} v(q_i, q_{i+1})$  for  $1 \leq i \leq t - 1$ .
- for  $1 \leq i \leq t - 1$  and for every  $r_1, \dots, r_{n-1} \in u(p_i, p_{i+1})$  ( $r_1 < \dots < r_{n-1}$ ), there exist  $s_1, \dots, s_{n-1} \in v(q_i, q_{i+1})$  ( $s_1 < \dots < s_{n-1}$ ) such that
  - (1)  $Q_a^u r_j$  if and only if  $Q_a^v s_j$ ,  $a \in A$  for  $1 \leq j \leq n - 1$ ,
  - (2)  $u(r_j, r_{j+1}) \sim_{(m)} v(s_j, s_{j+1})$  for  $1 \leq j \leq n - 2$ ,
  - (3)  $u(p_i, r_1) \sim_{(m)} v(q_i, s_1)$ .

Also, there exist  $s_1, \dots, s_{n-1} \in v(q_i, q_{i+1})$  (which may be different from the positions which satisfy (1), (2) and (3)) ( $s_1 < \dots < s_{n-1}$ ) such that (1), (2) and

- (4)  $u(r_{n-1}, p_{i+1}) \sim_{(m)} v(s_{n-1}, q_{i+1})$

hold. Similarly, for every  $s_1, \dots, s_{n-1} \in v(q_i, q_{i+1})$  ( $s_1 < \dots < s_{n-1}$ ), there exist  $r_1, \dots, r_{n-1} \in u(p_i, p_{i+1})$  ( $r_1 < \dots < r_{n-1}$ ) such that (1), (2) and (3) hold (also (1), (2) and (4) hold).

Similarly, for every  $s_1, \dots, s_n \in v(q_i, q_{i+1})$  ( $s_1 < \dots < s_n$ ), there exist  $r_1, \dots, r_n \in u(p_i, p_{i+1})$  ( $r_1 < \dots < r_n$ ) such that (5) and (6) hold.

Let  $w \in A^+$  and  $|w\alpha| = r \geq 2$ . In the proof of Lemma 4.11, we will talk about the  $(m)_r$  first positions of  $w$  (they form a subset of the  $(m)$  first positions of  $w$ ). They are defined

recursively as follows: let  $w_1$  denote the smallest prefix of  $w$  such that  $w_1\alpha = w\alpha$  (call the last position of  $w_1$ ,  $p_1$ ); let  $w_2$  denote the smallest prefix of  $w(p_1, |w|]$  such that  $w_2\alpha = (w(p_1, |w|])\alpha$  (call the last position of  $w_2$ ,  $p_2$ ); ... let  $w_m$  denote the smallest prefix of  $w(p_{m-1}, |w|]$  such that  $w_m\alpha = (w(p_{m-1}, |w|])\alpha$  (call the last position of  $w_m$ ,  $p_m$ ).  $\max\{p_1, \dots, p_m\}$  is an  $(m)_r$  first position in  $w$ . If  $r = 2$ , the procedure terminates. If  $r > 2$ , we repeat the process to find the  $(m)_{r-1}$  first positions in  $w[1, p_1)$  and the  $(m-i+1)_{r-1}$  first positions in  $w(p_{i-1}, p_i)$  for  $2 \leq i \leq m$ .

We can define similarly  $(m)_r$  last positions of  $w$ . The  $(m)_r$  first and the  $(m)_r$  last positions in  $w$  are called the  $(m)_r$  positions in  $w$ . The number of  $(m)_r$  positions in  $w$  is bounded above by  $2 \sum_{i=0}^{r-2} m^i$ .

Consider the example in Section 3.1:

$$w = abbbbbbbaaaabbbbaaaab\bar{c}bccccbbbbbcb\bar{c}bbcababaaaaa\bar{b}bbcccccb\bar{c}bcbabab- \\ \bar{c}bcbcbcb\bar{a}abb.$$

The overlined positions of  $w$  are the  $(5)_3$  first positions in  $w$ .

The following facts hold.

**Fact 5:** Let  $p_1, \dots, p_t \in u$  ( $p_1 < \dots < p_t$ ) ( $q_1, \dots, q_t \in v$  ( $q_1 < \dots < q_t$ )) be the  $(m)_r$  first positions in  $u$  ( $v$ ). Let  $p'_1, \dots, p'_s \in u$  ( $p'_1 < \dots < p'_s$ ) ( $q'_1, \dots, q'_s \in v$  ( $q'_1 < \dots < q'_s$ )) be the  $(m)$  first positions in  $u$  ( $v$ ) (here  $p_t = p'_s$  and  $q_t = q'_s$ ). Assume that  $t = t'$ ,  $Q_a^u p_i$  if and only if  $Q_a^v q_i$ ,  $a \in A$  for  $1 \leq i \leq t$ ,  $u[1, p_1) \sim_{(2m+1)} v[1, q_1)$ , and  $u(p_i, p_{i+1}) \sim_{(2m+1)} v(q_i, q_{i+1})$  for  $1 \leq i \leq t-1$ . We can conclude that  $s = s'$ ,  $Q_a^u p'_i$  if and only if  $Q_a^v q'_i$ ,  $a \in A$  for  $1 \leq i \leq s$ , and  $u(p'_i, p'_{i+1}) \sim_{(1)} v(q'_i, q'_{i+1})$  for  $1 \leq i \leq s-1$ . To see this, we consider the gaps in  $u$  formed by the  $p_i$ 's and the gaps in  $v$  formed by the  $q_i$ 's. Let  $u'$  be such a gap in  $u$  and  $v'$  be its corresponding gap in  $v$ . One of the following is true:

- $u'\alpha = v'\alpha$  and  $u', v'$  do not contain all the words of length  $\leq m$  over their alphabet of size 2;
- $u' = u''au'''$ ,  $v' = v''av'''$  where  $u''\alpha = v''\alpha$ ,  $\{a\} \not\subseteq u''\alpha$ ,  $u'''\alpha = v'''\alpha$ ,  $u''$ ,  $v''$  are free from  $(m)$  first positions, and  $u'''$ ,  $v'''$  do not contain all the words of length  $< m$  over their alphabet of size 2;
- $u'\alpha = v'\alpha$  and  $u', v'$  are free from  $(m)$  first positions.

The fact follows. A similar fact holds for the  $(m)_r$  last positions in  $u$  ( $v$ ).

**Fact 6:** Lemma 3.2 together with fact 5 imply the following. Let  $p_1, \dots, p_t \in u$  ( $p_1 < \dots < p_t$ ) ( $q_1, \dots, q_t \in v$  ( $q_1 < \dots < q_t$ )) be the  $(m)_r$  positions in  $u$  ( $v$ ). If  $t = t'$ ,  $Q_a^u p_i$  if and only if  $Q_a^v q_i$ ,  $a \in A$  for  $1 \leq i \leq t$ ,  $u[1, p_1) \sim_{(2m+1)} v[1, q_1)$ ,  $u(p_i, p_{i+1}) \sim_{(2m+1)} v(q_i, q_{i+1})$  for  $1 \leq i \leq t-1$ , and  $u(p_t, |u|] \sim_{(2m+1)} v(q_t, |v|]$ , then  $u \sim_{(1,m)} v$ .

## 4. Inclusion relations

### 4.1. Between the congruences $\sim_{(n,m)}$

The purpose of this section is to find necessary and sufficient conditions for  $\sim_{(n',m')}$  to be included in  $\sim_{(n,m)}$ . The proofs provide different winning strategies for either player I or player II. Lemma 2.2 implies that a necessary condition is  $\mathcal{N}(n, m) \leq \mathcal{N}(n', m')$ . Applications are discussed in the next section.

**Lemma 4.1.** (1) If  $\mathcal{N}(n, n') \geq \mathcal{N}(1, m)$  and if either  $|A| = 2$ , or  $n \neq 2$ , or  $n' \leq 2$ , then

$$\sim_{(n, n')} \subseteq \sim_{(1, m)}.$$

$$(2) \sim_{(2, m)} \subseteq \sim_{(1, m+1)}.$$

**Proof.** First, we show (1). It is sufficient to show the result for  $m = \lfloor (\mathcal{N}(n, n') - 1)/2 \rfloor$ . The result is obvious for  $n = 1$ . Assume  $n > 1$  and let  $u, v \in A^+$  be such that  $u \sim_{(n, n')} v$ . There is then a winning strategy for player II to win each play of the game  $\mathcal{G} = \mathcal{G}_{(n, n')}(u, v)$ . Let us show that  $u \sim_{(1, m)} v$  by using Lemma 3.2. Let  $p_1, \dots, p_t$  ( $p_1 < \dots < p_t$ ) ( $q_1, \dots, q_t$  ( $q_1 < \dots < q_t$ )) denote the  $(m)$  positions in  $u$  ( $v$ ). The first two conditions of Lemma 3.2 hold by Lemma 3.1. To show that the third condition of Lemma 3.2 also holds, let  $1 \leq i \leq t - 1$  and let  $p \in u(p_i, p_{i+1})$  (the proof is similar if starting in  $v(q_i, q_{i+1})$ ). Assume  $Q_a^u p$ ,  $a \in A$ . We are looking for  $q \in v(q_i, q_{i+1})$  satisfying  $Q_a^v q$ . We consider the following four cases (the hypotheses  $|A| = 2$ , or  $n \neq 2$ , or  $n' \leq 2$  will be used in case 4 only).

*Case 1:*  $p_i$  and  $p_{i+1}$  are among the  $(m)$  first positions in  $u$ . First, assume that there exist  $1 \leq i_1 < \dots < i_{m-1} < i$  such that  $p_{i_1}, \dots, p_{i_{m-1}}, p_i$  and  $p_{i_1}, \dots, p_{i_{m-1}}, p_{i+1}$  spell the first occurrence of subwords of length  $\leq m$  in  $u$ . Consider the play of the game  $\mathcal{G}$ , where player I, in the first move, chooses among the positions  $p_{i_1}, \dots, p_{i_{m-1}}, p$  the  $(n')$ th,  $(2n')$ th,  $\dots$ , and  $p$  for a total of at most  $n$  positions since  $m \leq nn'$ . Call them  $r_1, \dots, r_{n-1}$ ,  $p$  ( $r_1 \leq \dots \leq r_{n-1} < p$ ). Hence there exist  $s_1, \dots, s_{n-1}$ ,  $q$  ( $s_1 \leq \dots \leq s_{n-1} < q$ ) satisfying Lemma 2.1 (in particular,  $Q_a^v q$ ). We have that  $r_j = p_{i_{j n'}}$  if and only if  $s_j = q_{i_{j n'}}$ . Let the positions (among the positions  $p_{i_1}, \dots, p_{i_{m-1}}$ ) which are in  $u(r_{n-1}, p)$  be denoted by  $p'_1, \dots, p'_{n'-1}$  ( $p'_1 \leq \dots \leq p'_{n'-1}$ ).  $u(r_{n-1}, p) \sim_{(n')} v(s_{n-1}, q)$  and  $p'_1, \dots, p'_{n'-1}, p_i \in u(r_{n-1}, p)$  imply that  $q \in v(q_i, q_{i+1})$ . More precisely,  $q \notin v(s_{n-1}, q_i]$  since otherwise there would be an occurrence of the word (of length  $\leq n'$ ) spelled by  $p'_1, \dots, p'_{n'-1}, p_i$  in  $u(r_{n-1}, p)$  but not in  $v(s_{n-1}, q)$ ;  $q \notin v(q_{i+1}, |v|]$  since otherwise there would be an occurrence of the word (of length  $\leq n'$ ) spelled by  $p'_1, \dots, p'_{n'-1}, p_{i+1}$  in  $v(s_{n-1}, q)$  but not in  $u(r_{n-1}, p)$  (the letter of  $p_{i+1}$  differs from the letters in  $u(p_{i_{m-1}}, p)$  since otherwise there would be contradiction with the fact that  $p_{i_1}, \dots, p_{i_{m-1}}, p_{i+1}$  spell the first occurrence of a subword of length  $\leq m$  in  $u$ );  $q \neq q_{i+1}$  since otherwise  $Q_a^v q_{i+1}$  and hence  $Q_a^u p_{i+1}$ , contradicting the fact that  $p_{i_1}, \dots, p_{i_{m-1}}, p_{i+1}$  spell the first occurrence of a subword of length  $\leq m$  in  $u$  ( $Q_a^u p$  and  $p_{i_{m-1}} < p < p_{i+1}$ ).

Otherwise, using fact 3 of Section 3,  $u(p_i, p_{i+1}) = 1$  and there exist  $1 \leq i_1 \leq \dots \leq i_{m-1} = i$  such that  $p_{i_1}, \dots, p_{i_{m-1}}, p_{i+1}$  spell the first occurrence of a subword of length  $\leq m$  in  $u$ . In such a situation, we show that  $v(q_i, q_{i+1}) = 1$ . To see this, consider the play of the game  $\mathcal{G}$ , where player I, in the first move, chooses among  $p_{i_1}, \dots, p_{i_{m-1}}, p_{i+1}$  the  $(n')$ th,  $(2n')$ th,  $\dots$ , and  $p_{i+1}$  for a total of at most  $n$  positions since  $m \leq nn'$ . Call them  $r_1, \dots, r_{n-1}$ ,  $p_{i+1}$  ( $r_1 \leq \dots \leq r_{n-1} \leq p_i$ ). Hence there exist  $s_1, \dots, s_{n-1}$ ,  $q_{i+1}$  ( $s_1 \leq \dots \leq s_{n-1} \leq q_i$ ) satisfying Lemma 2.1. We have, as before,  $r_j = p_{i_{j n'}}$  if and only if  $s_j = q_{i_{j n'}}$ . Let the positions (among  $p_{i_1}, \dots, p_{i_{m-1}}$ ) which are in  $u(r_{n-1}, p_{i+1})$  be denoted by  $p'_1, \dots, p'_{n'-1}$  ( $p'_1 \leq \dots \leq p'_{n'-1}$ ). If  $v(q_i, q_{i+1}) \neq 1$ , then let  $q \in v(q_i, q_{i+1})$ . The word spelled by  $p'_1, \dots, p'_{n'-1}, q$  is then in  $v(s_{n-1}, q_{i+1})$  but not in  $u(r_{n-1}, p_{i+1})$ , contradicting the fact that  $u(r_{n-1}, p_{i+1}) \sim_{(n')} v(s_{n-1}, q_{i+1})$ .

*Case 2:*  $p_i$  and  $p_{i+1}$  are among the  $(m)$  last positions in  $u$ . Similar to case 1.

*Case 3:*  $p_i(p_{i+1})$  is among the  $(m)$  first ( $(m)$  last) positions in  $u$ . Assume case 1 or case 2 do not apply. Then there exist  $1 \leq i_1 \leq \dots \leq i_{m-1} \leq i$ ,  $i+1 \leq j_{m-1} \leq \dots \leq j_1 \leq t$ ,

such that  $p_{i_1}, \dots, p_{i_{m-1}}, p_i (p_{i+1}, p_{j_{m-1}}, \dots, p_{j_1})$  spell the first (last) occurrence of a subword of length  $\leq m$  in  $u$ . Consider the play of the game  $\mathcal{G}$  where player I, in the first move, chooses among the positions  $p_{i_1}, \dots, p_{i_{m-1}}, p_i, p, p_{i+1}, p_{j_{m-1}}, \dots, p_{j_1}$  (at most  $\mathcal{N}(1, m)$ ) the  $(n' + 1)$ th from the left, the  $(n' + 1)$ th from the right, the  $2(n' + 1)$ th from the left, the  $2(n' + 1)$ th from the right, ..., for a total of at most  $n$  positions since  $\mathcal{N}(1, m) \leq \mathcal{N}(n, n')$ , and call them  $r_1, \dots, r_n$  ( $r_1 \leq \dots \leq r_n$ ). There exist  $s_1, \dots, s_n$  ( $s_1 \leq \dots \leq s_n$ ) satisfying Lemma 2.1. We have if  $r_k = p_{i_{k(n'+1)}}$ , then  $s_k \geq q_{i_{k(n'+1)}}$ ; if  $r_{n+1-k} = p_{j_{k(n'+1)}}$ , then  $s_{n+1-k} \leq q_{j_{k(n'+1)}}$ ; if  $r_k = p_i$ , then  $s_k \geq q_i$ ; and if  $r_k = p_{i+1}$ , then  $s_k \leq q_{i+1}$ .

If  $n$  is odd,  $p = r_{(n+1)/2}$  and let  $q = s_{(n+1)/2}$ . We have that  $Q_a^v q$ . Put  $r = r_{((n+1)/2)-1}$ ,  $r' = r_{((n+1)/2)+1}$ ,  $s = s_{((n+1)/2)-1}$ , and  $s' = s_{((n+1)/2)+1}$ . We have that  $u(r, p) \sim_{(n')} v(s, q)$  and  $u(p, r') \sim_{(n')} v(q, s')$ . Let the positions (among the  $\leq \mathcal{N}(1, m)$  positions considered above) which are in  $u(r, p)$  be denoted by  $p'_1, \dots, p'_{n'}$  ( $p'_1 \leq \dots \leq p'_{n'}$ ) and those in  $u(p, r')$  by  $r'_1, \dots, r'_{n'}$  ( $r'_1 \leq \dots \leq r'_{n'}$ ) ( $p'_{n'} = p_i$  and  $r'_1 = p_{i+1}$ ). Since the words spelled by  $p'_1, \dots, p'_{n'}$  ( $r'_1, \dots, r'_{n'}$ ) must be in  $v(s, q)$  ( $v(q, s')$ ), it follows that  $q \in v(q_i, q_{i+1})$ .

If  $n$  is even, put  $r = r_{n/2}$ ,  $r' = r_{(n/2)+1}$ ,  $s = s_{n/2}$  and  $s' = s_{(n/2)+1}$ . We have that  $u(r, r') \sim_{(n')} v(s, s')$ . Let the positions (among the  $\leq \mathcal{N}(1, m)$  positions considered above) which are in  $u(r, r')$  be denoted by  $p'_1, \dots, p'_{n'}$  ( $p'_1 \leq \dots \leq p'_{n'}$ ) ( $p = p'_{(n'+1)/2}$ ,  $p_i = p'_{[(n'+1)/2]-1}$ ,  $p_{i+1} = p'_{[(n'+1)/2]+1}$ ). The word (of length  $\leq n'$ ) spelled by these positions is in  $v(s, s')$ . Let  $q'_1, \dots, q'_{n'}$  ( $q'_1 \leq \dots \leq q'_{n'}$ ) spell that same word in  $v(s, s')$ .  $q = q'_{[(n'+1)/2]}$  is such that  $Q_a^v q$  and  $q \in v(q_i, q_{i+1})$ .

Case 4:  $p_i (p_{i+1})$  is among the  $(m)$  last  $((m)$  first) positions in  $u$ . Assume cases 1–3 do not apply. There exist  $1 \leq i_1 \leq \dots \leq i_{m-1} \leq i$ ,  $i+1 \leq j_{m-1} \leq \dots \leq j_1 \leq t$ , such that  $p_{i_1}, \dots, p_{i_{m-1}}, p_{i+1} (p_i, p_{j_{m-1}}, \dots, p_{j_1})$  spell the first (last) occurrence of a subword of length  $\leq m$  in  $u$ . First of all, the letter of  $p$  is not the letter of  $p_i$  nor the letter of  $p_{i+1}$ . Hence  $q = q_i$  or  $q = q_{i+1}$  are eliminated. Also, the letters of  $p_i$  and  $p_{i+1}$  differ. Hence, if  $|A| = 2$  the proof is complete. So for the rest of the proof, we assume  $|A| > 2$ .

If  $n$  is odd, or  $n$  is even and  $n' \leq 2$ , we consider the play of the game  $\mathcal{G}$  where player I, in the first move, chooses as in case 3. Using the same notations as in case 3, we show the existence of  $q$  in  $v(q_i, q_{i+1})$  such that  $Q_a^v q$ . We have that if  $r_k = p_{i_{k(n'+1)}}$ , then  $s_k \geq q_{i_{k(n'+1)}}$ ; if  $r_{n+1-k} = p_{j_{k(n'+1)}}$ , then  $s_{n+1-k} \leq q_{j_{k(n'+1)}}$ ; if  $r_k = p_i$ , then  $s_k \leq q_i$ ; and if  $r_k = p_{i+1}$ , then  $s_k \geq q_{i+1}$ .

For  $n$  odd,  $q < q_i$  would imply an occurrence of the word (of length  $\leq n'$ ) spelled by  $p_i, r'_2, \dots, r'_{n'}$  in  $v(q, s')$  but not in  $u(p, r')$ , and  $q > q_{i+1}$  would imply an occurrence of the word (of length  $\leq n'$ ) spelled by  $p'_1, \dots, p'_{n'-1}, p_{i+1}$  in  $v(s, q)$  but not in  $u(r, p)$  (here we use the fact that the letter of  $p_i$  does not occur in  $u(p_i, p_{j_{m-1}})$  and the letter of  $p_{i+1}$  does not occur in  $u(p_{i_{m-1}}, p_{i+1})$ ).

For  $n$  even and  $n' \leq 2$ ,  $r = p_i$  and  $r' = p_{i+1}$ . We have that  $s \leq q_i$  and  $s' \geq q_{i+1}$ .  $s < q_i$  would imply an occurrence of the word (of length  $\leq n'$ ) spelled by  $p_i$  in  $v(s, s')$  but not in  $u(r, r')$ , and  $q_{i+1} < s'$  would imply an occurrence of the word (of length  $\leq n'$ ) spelled by  $p_{i+1}$  in  $v(s, s')$  but not in  $u(r, r')$ . Hence  $s = q_i$  and  $s' = q_{i+1}$ .  $u(r, r') \sim_{(n')} v(s, s')$  implies the existence of  $q \in v(q_i, q_{i+1})$  such that  $Q_a^v q$ .

If  $n$  is even,  $n \geq 4$  and  $n' > 2$ , we consider the play of the game  $\mathcal{G}$  where player I, in the first move, chooses among the positions  $p_{i_1}, \dots, p_{i_{m-1}}, p_i, p_{i+1}$  the  $(n')$ th, the  $(2n')$ th, ...,  $p_i$  and  $p_{i+1}$  for a total of at most  $n$  positions since  $1 + \lceil m/n' \rceil \leq n$ . Call them  $r_1, \dots, r_n$  ( $r_1 \leq \dots \leq r_{n-2} < r_{n-1} < r_n$ ). Hence there exist  $s_1, \dots, s_n$

$(s_1 \leq \dots \leq s_{n-2} < s_{n-1} < s_n)$  satisfying Lemma 2.1. We have that  $r_j = p_{i_{j'n'}}$  if and only if  $s_j = q_{i_{j'n'}}$ ,  $r_{n-1} = p_i$ ,  $s_{n-1} = q_i$ ,  $r_n = p_{i+1}$  and  $s_n = q_{i+1}$ .  $u(p_i, p_{i+1}) \sim_{(n')} v(q_i, q_{i+1})$  implies the existence of  $q$  in  $v(q_i, q_{i+1})$  such that  $Q_a^v q$ .

Now, we show (2). The proof is similar to the proof of (1) except for case 4. Let  $u, v \in A^+$  be such that  $u \sim_{(2,m)} v$ . If  $p_i (p_{i+1})$  is among the  $(m+1)$  last  $((m+1)$  first) positions in  $u$ , there exist  $1 \leq i_1 \leq \dots \leq i_m \leq i$ ,  $i+1 \leq j_m \leq \dots \leq j_1 \leq t$ , such that  $p_{i_1}, \dots, p_{i_m}, p_{i+1} (p_i, p_{j_m}, \dots, p_{j_1})$  spell the first (last) occurrence of a subword of length  $\leq m+1$  in  $u$ . Consider the play of the game  $\mathcal{G}_{(2,m)}(u, v)$  where player I, in the first move, chooses  $p_i$  and  $p_{i+1}$ . Similarly to case 4 ( $n$  even and  $n' \leq 2$ ), we can conclude that player II has to choose  $q_i$  and  $q_{i+1}$ . The existence of  $q \in v(q_i, q_{i+1})$  such that  $Q_a^v q$  follows.  $\square$

**Lemma 4.2.** *If  $|A| \geq 3$ , then  $\sim_{(2,m)} \not\sim_{(1,m+3)}$ .*

**Proof.** The result is obviously true if  $m < 5$  since  $\mathcal{N}(1, m+3) > \mathcal{N}(2, m)$ . So assume  $m \geq 5$  and let  $A$  contain at least the three letters  $a, b$  and  $c$ . Define

$$w_m = \dots (uv)(uav)(uv)(uav)\underline{uv}(uv)(uav)(uv)(uav) \dots,$$

and

$$w'_m = \dots (uv)(uav)(uv)(uav)\underline{uav}(uv)(uav)(uv)(uav) \dots,$$

where  $u = (ab)^{\mathcal{N}(2,m)}$  and  $v = (ca)^{\mathcal{N}(2,m)}$ , and where the total number of  $u$ - and  $v$ -segments preceding and following the underlined segments is exactly  $m+2$ . For instance, if  $m = 5$ ,

$$w_5 = v(uav)(uv)(uav)\underline{uv}(uv)(uav)(uv)u,$$

and

$$w'_5 = v(uav)(uv)(uav)\underline{uav}(uv)(uav)(uv)u,$$

where  $u = (ab)^{17}$  and  $v = (ca)^{17}$ .  $w_m$  and  $w'_m$  are not  $\sim_{(1,m+3)}$ -equivalent. To see this, we illustrate a winning strategy for player I. Player I, in the first move, chooses the middle  $a$  of the underlined segment  $w'_m$ . Player II cannot win this play of the game  $\mathcal{G}_{(1,m+3)}(w_m, w'_m)$  since the last  $b$  of the  $u$ -segment of the underlined segments is an  $(m+3)$  last position, and the first  $c$  of the  $v$ -segment of the underlined segments is an  $(m+3)$  first position.

We now show that  $w_m$  and  $w'_m$  are  $\sim_{(2,m)}$ -equivalent:

$$w_m \sim_{(2,m)} \dots (uv)(uav)(uv)\overline{ab}(uav)\underline{uv}(uv)\overline{ca}(uav)(uv)(uav) \dots,$$

and

$$w'_m \sim_{(2,m)} \dots (uv)(uav)(uv)\overline{ab}(uav)\underline{uav}(uv)\overline{ca}(uav)(uv)(uav) \dots,$$

where the total number of  $u$ - and  $v$ -segments preceding and following the underlined segments is as before. The above equivalences are true since  $(ab)^{\mathcal{N}(2,m)} \sim_{(2,m)} (ab)^{\mathcal{N}(2,m)+1}$  and  $(ca)^{\mathcal{N}(2,m)} \sim_{(2,m)} (ca)^{\mathcal{N}(2,m)+1}$ . Notice that the segment up to and including the overlined  $ab$ -segment contains all the words of length  $\leq m$  over  $\{a, b, c\}$ .

The same is true for the segment starting with the overlined  $ca$ -segment. Call the word which is  $\sim_{(2,m)}$ -equivalent to  $w_m$  by  $w'_m$  and the word  $\sim_{(2,m)}$ -equivalent to  $w'_m$  by  $w''_m$ . To see that  $w_m$  and  $w'_m$  are  $\sim_{(2,m)}$ -equivalent, it is sufficient to show that  $w'_m$  and  $w''_m$  are  $\sim_{(2,m)}$ -equivalent. We distinguish the two cases where player I first picks 2 positions from  $w'_m$  or player I first picks 2 positions from  $w''_m$ . Note that  $(ab)^{\mathcal{N}(2,m)} \sim_{(m)} (ab)^{\mathcal{N}(2,m)} a$  and  $(ca)^{\mathcal{N}(2,m)} \sim_{(m)} a(ca)^{\mathcal{N}(2,m)}$ .

*Case 1:* Assume that player I has chosen 2 positions from  $w''_m$ . Player II will pick exactly corresponding positions in  $w'_m$ , except possibly when player I chooses one position from the  $u$ -segment and one from the  $v$ -segment of the underlined segment of  $w''_m$ . In this situation, player II will pick exactly corresponding positions in the  $uv$ -segment immediately following the underlined segment of  $w''_m$ .

*Case 2:* Assume now that player I has picked his 2 positions from  $w'_m$ . Player II will pick exactly corresponding positions in  $w''_m$ , except possibly when player I picks one position from the  $u$ -segment and one from the  $v$ -segment of the underlined segment of  $w'_m$ , or the middle  $a$  from the underlined segment of  $w'_m$ . In the first situation, player II will pick exactly corresponding positions in the  $uav$ -segment immediately preceding the underlined segment of  $w'_m$ . In the second situation, if the other chosen position is at the left (right) of the middle  $a$  of the underlined segment of  $w'_m$ , then player II will choose his positions in the segment up to and including the last  $a$  of the underlined  $u$ -segment (following and including the first  $a$  of the underlined  $v$ -segment) of  $w'_m$ .  $\square$

**Lemma 4.3.** *Let  $m > m'$  and  $n > 1$ . If  $(2 + (n - 1)M)m \leq n'm'$ , then  $\sim_{(n',m')} \subseteq \sim_{(n,m)}$ , where  $2\lfloor(m - 1)/m'\rfloor M$  is the maximum number of  $(m, m')$  positions in words over  $A$ .*

**Proof.** Let  $u, v \in A^+$  and suppose  $u \sim_{(n',m')} v$ . There is a winning strategy for player II to win each play of the game  $\mathcal{G} = \mathcal{G}_{(n',m')}(u, v)$ . We will show that  $u \sim_{(n,m)} v$  under the stated hypotheses by using Lemma 3.3. Let  $p_1, \dots, p_t \in u$  ( $p_1 < \dots < p_t$ ) ( $q_1, \dots, q_r \in v$  ( $q_1 < \dots < q_r$ )) be the  $(m)$  positions in  $u$  ( $v$ ). The first two conditions of Lemma 3.3 hold by Lemma 3.1. To see that the fifth condition of Lemma 3.3 holds (the third and fourth conditions of Lemma 3.3 will follow similarly), let  $1 \leq i \leq t - 1$  and let  $r_1, \dots, r_n \in u(p_i, p_{i+1})$  ( $r_1 < \dots < r_n$ ) (similar if starting in  $v(q_i, q_{i+1})$ ). We are looking for  $s_1, \dots, s_n \in v(q_i, q_{i+1})$  ( $s_1 < \dots < s_n$ ) satisfying  $Q_a^u r_j$  if and only if  $Q_a^v s_j$ ,  $a \in A$  for  $1 \leq j \leq n$ , and  $u(r_j, r_{j+1}) \sim_{(m)} v(s_j, s_{j+1})$  for  $1 \leq j \leq n - 1$ . We consider the following four cases. Details follow as in Lemma 4.1.

*Case 1:*  $p_i$  and  $p_{i+1}$  are among the  $(m)$  first positions in  $u$ . First, assume that there exist  $1 \leq i_1 < \dots < i_{m-1} < i$  such that  $p_{i_1}, \dots, p_{i_{m-1}}, p_i$  and  $p_{i_1}, \dots, p_{i_{m-1}}, p_{i+1}$  spell the first occurrence of subwords of length  $\leq m$  in  $u$ . Consider the play of the game  $\mathcal{G}$  where player I, in the first move, chooses  $p_{i_m}, p_{i_{2m}}, \dots, r_1$ , the  $(m, m')$  first positions in  $u(r_k, r_{k+1})$  and  $r_{k+1}$  for  $1 \leq k \leq n - 1$ , for a total of at most  $n'$  positions since  $\lceil m/m' \rceil + (n - 1)\lceil m/m' \rceil M \leq n'$ . Obviously,  $q_{i_m}, q_{i_{2m}}, \dots$  should be among the positions chosen in  $v$  by player II in the first move, and there also exist  $s_1, \dots, s_n \in v(q_i, q_{i+1})$  ( $s_1 < \dots < s_n$ ) (corresponding to  $r_1, \dots, r_n$ ) and there exist positions in  $v(s_1, s_2), \dots, v(s_{n-1}, s_n)$  (corresponding to the positions chosen in  $u(r_1, r_2), \dots$ , and  $u(r_{n-1}, r_n)$ ) satisfying Lemma 2.1.  $u(r_j, r_{j+1}) \sim_{(m)} v(s_j, s_{j+1})$  for  $1 \leq j \leq n - 1$  follows by using fact 4 of Section 3.

Otherwise, using fact 3 of Section 3,  $u(p_i, p_{i+1}) = 1$  and there exist  $1 \leq i_1 \leq \dots \leq i_{m-1} = i$  such that  $p_{i_1}, \dots, p_{i_{m-1}}, p_{i+1}$  spell the first occurrence of a subword of length  $\leq m$  in  $u$ . In such a situation, case 1 of Lemma 4.1 shows that  $v(q_i, q_{i+1}) = 1$ .

Case 2:  $p_i$  and  $p_{i+1}$  are among the  $(m)$  last positions in  $u$ . Similar to case 1.

Case 3:  $p_i(p_{i+1})$  is among the  $(m)$  first  $((m)$  last) positions in  $u$ . Assume case 1 or case 2 do not apply. Then there exist  $1 \leq i_1 \leq \dots \leq i_{m-1} \leq i, i+1 \leq j_{m-1} \leq \dots \leq j_1 \leq t$ , such that  $p_{i_1}, \dots, p_{i_{m-1}}, p_i(p_{i+1}, p_{j_{m-1}}, \dots, p_{j_1})$  spell the first (last) occurrence of a subword of length  $\leq m$  in  $u$ . Consider the play of the game  $\mathcal{G}$  where player I, in the first move, chooses  $p_{i_m}, p_{i_{2m}}, \dots, r_1$ , the  $(m, m')$  first positions in  $u(r_k, r_{k+1})$  and  $r_{k+1}$  for  $1 \leq k \leq n-1$ ,  $p_{j_m}, p_{j_{2m}}, \dots$ , for a total of at most  $n'$  positions since  $2\lceil m/m' \rceil + (n-1)\lceil m/m' \rceil M \leq n'$ .

Case 4:  $p_i(p_{i+1})$  is among the  $(m)$  last  $((m)$  first) positions in  $u$ . Assume cases 1–3 do not apply. There exist  $1 \leq i_1 \leq \dots \leq i_{m-1} \leq i$  such that  $p_{i_1}, \dots, p_{i_{m-1}}, p_{i+1}$  spell the first occurrence of a subword of length  $\leq m$  in  $u$ . As in case 4 of Lemma 4.1, the letters of  $p_i, u(p_i, p_{i+1})$  and  $p_{i+1}$  differ. The proof is hence complete if  $|A| = 2$ . Otherwise, we consider the play of the game  $\mathcal{G}$  where player I, in the first move, chooses the positions he chooses in case 1 together with  $p_i$  and  $p_{i+1}$ . The total number of chosen positions is at most  $n'$  since  $\lceil m/m' \rceil + (n-1)\lceil m/m' \rceil M + 2 \leq n'$ .  $\square$

**Lemma 4.4.** If  $|A| = 2$ , then  $\sim_{(2n, \mathcal{N}(1, m))} \subseteq \sim_{(\mathcal{N}(1, n), m)}$ .

**Proof.** Details appear in [3].  $\square$

**Lemma 4.5.**  $\sim_{(2n-1, m)} \not\subseteq \sim_{(2n, 1)}$  and  $\sim_{(2n, m)} \not\subseteq \sim_{(2n+2, 1)}$ .

**Proof.** The result is obvious if  $m = 1$ . So assume  $m > 1$  and let  $a, b \in A$ .  $(ab)^m ab(ab)^m$  is  $\sim_{(1, m)}$ -equivalent to  $(ab)^m ba(ab)^m$ , but they are not  $\sim_{(2, 1)}$ -equivalent. Let  $n$  be fixed and let  $N = \mathcal{N}(2n+1, m)$ .  $w_m = ((ab)^N a(ab)^{2N} b(ab)^N)^n$  is  $\sim_{(2n+1, m)}$ -equivalent to  $w'_m = ((ab)^N b(ab)^{2N} a(ab)^N)^n$ , but they are not  $\sim_{(2n+2, 1)}$ -equivalent. To see that  $w_m$  and  $w'_m$  are not  $\sim_{(2n+2, 1)}$ -equivalent, consider the play of the game  $\mathcal{G}_{(2n+2, 1)}(w_m, w'_m)$  where player I, in the first move, chooses the  $n$  pairs of consecutive  $b$ 's and the last pair of consecutive  $a$ 's in  $w'_m$ . Player II cannot win this play. The  $\sim_{(2n+1, m)}$ -equivalence of  $w_m$  and  $w'_m$  follows the technique in [2].  $\square$

**Lemma 4.6.**  $\sim_{(2n, 2m)} \not\subseteq \sim_{(\mathcal{N}(1, n), m)}$ .

**Proof.** Let  $a, b \in A$ . Let  $u = (ab)^N a^{2m+1} (ab)^N$  and  $v = (ab)^N a^{2m+2} (ab)^N$ , where  $N = \mathcal{N}(2n, 2m)$ . If  $\mathcal{N}(1, n) \equiv 0 \pmod{3}$ , then consider the two words  $(uv)^{(n-1)/3} u(vu)^{(n-1)/3}$  and  $(uv)^{(n-1)/3} v(vu)^{(n-1)/3}$ . If  $\mathcal{N}(1, n) \equiv 1 \pmod{3}$ , consider the words  $u^{n/3} v^{(2n-3)/3} uu$  and  $u^{n/3} v^{(2n-3)/3} u$ . If  $\mathcal{N}(1, n) \equiv 2 \pmod{3}$ , then consider  $(uv)^{(2n-1)/3}$  and  $(vu)^{(2n-1)/3}$ . In each situation, the two given words are  $\sim_{(2n, 2m)}$ -equivalent but are not  $\sim_{(2n+1, m)}$ -equivalent. Let us show the result when  $\mathcal{N}(1, n) \equiv 0 \pmod{3}$  (the other cases are similar). Fix  $n$ . Put  $w_m = (uv)^{(n-1)/3} u(vu)^{(n-1)/3}$  and  $w'_m = (uv)^{(n-1)/3} v(vu)^{(n-1)/3}$ . First,  $w_m$  and  $w'_m$  are not  $\sim_{(2n+1, m)}$ -equivalent. To see this, player I, in the first move, chooses the overlined three  $a$ 's in each of the

$v$ -segments of  $w'_m$  (there are  $(n-1)/3 + (n-1)/3 + 1$  such segments):

$$v = \dots \bar{a}a^m\bar{a}a^m\bar{a} \dots$$

So he chooses a total of  $2n+1$  positions. Player II cannot win this play of the game  $\mathcal{G}_{(2n+1, m)}(w_m, w'_m)$ . The  $\sim_{(2n, 2m)}$ -equivalence of  $w_m$  and  $w'_m$  follows the technique in Lemma 4.2.  $\square$

**Lemma 4.7.** *If  $|A| = 3$ , then  $\sim_{(2, 3)} \subseteq \sim_{(3, 1)}$ .*

**Proof.** Let  $u, v \in A^+$  be such that  $u \sim_{(2, 3)} v$ . If  $u$  and  $v$  contain  $\leq 2$  letters, then the result follows from Lemma 4.4. Otherwise, we want to show that  $u \sim_{(3, 1)} v$ . Let  $p, q$  and  $r$  ( $p < q < r$ ) be positions in  $u$  (the proof is similar if starting in  $v$ ) chosen by player I in the first move (if two of these positions are equal, player II uses his strategy in  $\mathcal{G}_{(2, 3)}(u, v)$ ). The gaps in  $u$  formed by  $p, q$  and  $r$  will be denoted (in order) by  $gap1, gap2, gap3$  and  $gap4$ . We will show that  $p', q'$  and  $r' \in v$  ( $p' < q' < r'$ ) exist such that  $Q_a^u p$  if and only if  $Q_a^v p'$ ,  $Q_a^u q$  if and only if  $Q_a^v q'$ ,  $Q_a^u r$  if and only if  $Q_a^v r'$ ,  $a \in A$ ,  $u[1, p] \sim_{(1)} v[1, p']$ ,  $u(p, q) \sim_{(1)} v(p', q')$ ,  $u(q, r) \sim_{(1)} v(q', r')$  and  $u(r, |u|] \sim_{(1)} v(r', |v|]$ . Since  $u \sim_{(2, 3)} v$ , then the (3) positions  $p_1, \dots, p_t \in u$  ( $p_1 < \dots < p_t$ ) ( $q_1 < \dots < q_t \in v$  ( $q_1 < \dots < q_t$ )) in  $u$  ( $v$ ) satisfy Lemma 3.3. The proof is divided into the following cases. The result follows by considering different plays of the game  $\mathcal{G} = \mathcal{G}_{(2, 3)}(u, v)$ . For each case, we assume that the preceding cases do not apply.  $\mathcal{G}(s, s')$  will abbreviate the play of the game  $\mathcal{G}$  where player I, in the first move, chooses  $s$  and  $s'$  in  $u$ .

*Case 1:*  $q = p_j$  for some  $1 \leq j \leq t$ . Consider  $\mathcal{G}(p, q)$  and then  $\mathcal{G}(q, r)$ .

*Case 2:*  $p = p_j$  for some  $1 \leq j \leq t$  and  $p_j$  is a (2) first position in  $u$  (similar if  $r = p_j$  for some  $1 \leq j \leq t$  and  $p_j$  is a (2) last position in  $u$ ). In order to choose  $q'$  and  $r'$ , consider  $\mathcal{G}(q, r)$ . Let  $p' = q_j$ .

*Case 3:*  $p_j \in gap2$  for some  $1 \leq j \leq t$  and  $p_j$  is a (2) first position in  $u$  (similar if  $p_j \in gap3$  for some  $1 \leq j \leq t$  and  $p_j$  is a (2) last position in  $u$ ). Consider  $\mathcal{G}(q, r)$ , and then  $\mathcal{G}(p, p_j)$ .

*Case 4:*  $gap1$  consists of 1 letter only, say  $gap1$  consists of  $a$ 's only (similar if  $gap4 \sim_{(1)} a$ ). Since, by assumption, the preceding cases do not apply,  $Q_a^u p$  and  $gap2 \sim_{(1)} 1$  or  $gap2 \sim_{(1)} a$ . In order to choose  $q'$  and  $r'$ , consider  $\mathcal{G}(q, r)$ . In such situations,  $u[1, q] \sim_{(3)} v[1, q']$  implies the existence of  $p' \in v[1, q']$  such that  $Q_a^u p, u[1, p] \sim_{(1)} v[1, p']$  and  $u(p, q) \sim_{(1)} v(p', q')$ .

*Case 5:*  $gap1$  consists of 2 letters, say consists of  $a$ 's and  $b$ 's ( $a \neq b$ ) (similar if  $gap4 \sim_{(1)} ab$ ). Let  $c$  be the other letter in  $A$ . We have the following subcases.

*Case 5.1:*  $p_j \in gap3$  for some  $1 \leq j \leq t$ ,  $p_j$  is a (1) first position in  $u$  and  $Q_c^u p_j$ . In order to choose  $r'$ , consider  $\mathcal{G}(p_j, r)$  (player II has to choose  $q_j$ ).

If  $gap2$  and  $u(q, p_j)$  are either 1, or consist of  $a$ 's only or  $b$ 's only, then in order to choose  $p'$ , consider  $\mathcal{G}(p, p_j)$ . In such situations,  $u(p, p_j) \sim_{(3)} v(p', q_j)$  implies the existence of  $q' \in v(p', q_j)$  such that  $Q_a^u q$  if and only if  $Q_a^v q'$  or  $Q_b^u q$  if and only if  $Q_b^v q'$ ,  $u(p, q) \sim_{(1)} v(p', q')$  and  $u(q, p_j) \sim_{(1)} v(q', q_j)$ .

If  $gap2 \sim_{(1)} ab$ , then in order to choose  $q'$ , consider  $\mathcal{G}(q, p_j)$ . In such situations,  $u[1, q] \sim_{(3)} v[1, q']$  implies the existence of  $p' \in v[1, q']$  such that  $Q_a^u p$  if and only if  $Q_a^v p'$  or  $Q_b^u p$  if and only if  $Q_b^v p'$ ,  $u[1, p] \sim_{(1)} v[1, p']$  and  $u(p, q) \sim_{(1)} v(p', q')$ .

If  $gap2$  is either 1, or consists of  $a$ 's only or  $b$ 's only, and if  $u(q, p_j) \sim_{(1)} ab$ , then in order to choose  $p'$ , consider the play where player I chooses  $p$  and the last of the (1) first positions in  $u(q, p_j)$ .

Case 5.2:  $r = p_j$  for some  $1 \leq j \leq t$ ,  $p_j$  is a (1) first position in  $u$  and  $Q_c^u p_j$ . Similar to case 5.1. Here  $r' = q_j$ .

Case 5.3:  $p_j \in gap4$  for some  $1 \leq j \leq t$ ,  $p_j$  is a (1) first position in  $u$  and  $Q_c^u p_j$ . If  $gap2 \sim_{(1)} ab$ , then in order to choose  $q'$  and  $r'$ , consider  $\mathcal{G}(q, r)$ . If  $gap2$  and  $gap3$  are either 1, or consist of  $a$ 's only or  $b$ 's only, then to choose  $p'$  and  $r'$ , consider  $\mathcal{G}(p, r)$ . Otherwise, consider  $\mathcal{G}(p, s)$ , where  $s$  denotes the position following immediately the last of the (1) first positions in  $gap3$  and such that  $Q_a^u r$  if and only if  $Q_a^u s$ , or  $Q_b^u r$  if and only if  $Q_b^u s$ .

Case 6:  $gap1 \sim_{(1)} abc$  and  $gap4 \sim_{(1)} abc$ , where  $A = \{a, b, c\}$ . It is sufficient to consider the following subcases (the others follow similarly).

Case 6.1:  $gap2 \sim_{(1)} abc$ . To choose  $q'$  and  $r'$ , consider  $\mathcal{G}(q, r)$ .

For cases 6.2–6.5, in order to choose  $p'$  and  $r'$ , consider  $\mathcal{G}(p, r)$ .

Case 6.2:  $gap2$  and  $gap3$  are either 1, or consist of  $a$ 's only, or  $b$ 's only or  $c$ 's only.

Case 6.3:  $Q_a^u q$ ,  $a \notin gap2\alpha$  and  $a \notin gap3\alpha$ .

Case 6.4:  $Q_a^u q$ ,  $gap2 \sim_{(1)} a$  and  $gap3 \sim_{(1)} bc$ .

Case 6.5:  $gap2, gap3 \sim_{(1)} ab$ .

In the following cases,  $s$  denotes the first of the (1) last positions in  $gap2$ :  $s'$  denotes the position preceding immediately the first of the (1) last positions in  $gap2$  and satisfying  $Q_a^u p$  if and only if  $Q_a^u s'$ , or  $Q_b^u p$  if and only if  $Q_b^u s'$ , or  $Q_c^u p$  if and only if  $Q_c^u s'$ ;  $s''$  denotes the last of the (1) first positions in  $gap3$ ;  $s'''$  denotes the position following immediately the last of the (1) first positions in  $gap3$  and satisfying  $Q_a^u r$  if and only if  $Q_a^u s'''$ , or  $Q_b^u r$  if and only if  $Q_b^u s'''$ , or  $Q_c^u r$  if and only if  $Q_c^u s'''$ .

Case 6.6:  $gap2 \sim_{(1)} 1$  or  $a$ , and  $gap3 \sim_{(1)} ab$  (similar if  $gap2 \sim_{(1)} a$ ,  $\neg Q_a^u q$ , and  $gap3 \sim_{(1)} bc$ ). If  $Q_c^u r$ , then consider  $\mathcal{G}(p, s'')$ ; otherwise, player I chooses  $p$  and  $s'''$ .

Case 6.7:  $gap2 \sim_{(1)} ab$  and  $gap3 \sim_{(1)} ac$ . If  $\neg Q_c^u p$  and  $\neg Q_b^u r$ , then consider  $\mathcal{G}(s', s''')$ ; if  $\neg Q_c^u p$  and  $Q_b^u r$ , then consider  $\mathcal{G}(s', s'')$ ; if  $Q_c^u p$  and  $\neg Q_b^u r$ , then player I chooses  $s$  and  $s'''$ ; otherwise he chooses  $s$  and  $s'$ .  $\square$

**Lemma 4.8.** If  $|A| \geq 4$ , then  $\sim_{(2, m)} \not\sim_{(3, 1)}$ .

**Proof.** The result is obvious if  $m = 1$  since  $\mathcal{N}(3, 1) > \mathcal{N}(2, m)$ . So assume  $m > 1$  and let  $A$  contain at least the four letters  $a, b, c$  and  $d$ . Let  $N - 1 = \mathcal{N}(2, m)$  and define

$$w_m = (uvd)^N u(duv)^N$$

and

$$w'_m = (uvd)^N v(duv)^N,$$

where  $u = (ab)^N (ca)^N$  and  $v = (ab)^N a(ca)^N$ .  $w_m$  and  $w'_m$  are not  $\sim_{(3, 1)}$ -equivalent. We illustrate a winning strategy for player I. Player I, in the first move, chooses the following (overlined) positions in  $w'_m$ :

$$w'_m = \dots uv\bar{d}(ab)^N \bar{a}(ca)^N \bar{d}uv \dots$$

Player II cannot win this play of the game  $\mathcal{G}_{(3, 1)}(w_m, w'_m)$ .

Now, we show that  $w_m$  and  $w'_m$  are  $\sim_{(2,m)}$ -equivalent. To see this, we distinguish the two cases where player I first picks 2 positions from  $w_m$  or player I first picks 2 positions from  $w'_m$ . Note that  $(ab)^N \sim_{(m)} (ab)^N a$  and  $(ca)^N \sim_{(m)} a(ca)^N$ .

*Case 1:* Assume that player I has chosen 2 positions from  $w_m$ . Player II will pick exactly corresponding positions in  $w'_m$ , except possibly when player I chooses one position from the  $(ab)^N$ -segment of the middle  $u$ -segment of  $w_m$  and one position from the  $(ca)^N$ -segment of the middle  $u$ -segment of  $w_m$ . In this situation, the two positions chosen by player I are in the initial  $(uvd)^{N-1}$   $u$ -segment of  $w_m$ . Player II will pick his positions in the initial  $(uvd)^{N-1}$   $u$ -segment of  $w'_m$  according to his strategy in the game  $\mathcal{G}_{(2,m)}((uvd)^N u, (uvd)^{N-1} u)$  (since  $N-1 = \mathcal{N}(2, m)$ , we have that  $(uvd)^N u \sim_{(2,m)} (uvd)^{N-1} u$ ).

*Case 2:* Assume now that player I has picked his first 2 positions from  $w'_m$ . Player II will pick exactly corresponding positions in  $w_m$ , except possibly when player I picks the middle  $a$  from the middle  $v$ -segment of  $w'_m$ , or one position from the  $(ab)^N$ -segment of the middle  $v$ -segment of  $w'_m$  and one position from the  $(ca)^N$ -segment of the middle  $v$ -segment of  $w'_m$ .

When the positions chosen by player I are in the last  $v(duv)^N$ -segment of  $w'_m$ , player II can pick his positions in the last  $v(duv)^{N-1}$ -segment of  $w_m$  according to his strategy in the game  $\mathcal{G}_{(2,m)}(v(duv)^N, v(duv)^{N-1})$ .

When the positions chosen by player I are in the initial  $(uvd)^N(ab)^N a$ -segment of  $w'_m$ , player II can pick his positions in the initial  $(uvd)^N(ab)^{N-1} a$ -segment of  $w_m$  according to his strategy in the game  $\mathcal{G}_{(2,m)}((uvd)^N(ab)^N a, (uvd)^N(ab)^{N-1} a)$ .  $\square$

**Lemma 4.9.** *If  $|A| \geq 3$ , then  $\sim_{(2,m)} \not\subseteq \sim_{(3,2)}$ .*

**Proof.** The result is obviously true if  $m < 3$  since  $\mathcal{N}(3, 2) > \mathcal{N}(2, m)$ . So assume  $m \geq 3$  and let  $A$  contain at least the three letters  $a$ ,  $b$  and  $c$ . Let  $N-1 = \mathcal{N}(2, m)$  and define

$$w_m = (c(ba)^N ca(ba)^N)^N c(ba)^N c(ca(ba)^N c(ba)^N)^N$$

and

$$w'_m = (c(ba)^N ca(ba)^N)^N ca(ba)^N c(ca(ba)^N c(ba)^N)^N.$$

$w_m$  and  $w'_m$  are not  $\sim_{(3,2)}$ -equivalent. To see this, we illustrate a winning strategy for player I. Player I, in the first move, chooses the three following (overlined) positions in  $w'_m$ :

$$w'_m = \dots \bar{c}\bar{a}(ba)^N c\bar{c} \dots$$

Player II cannot win this play of the game  $\mathcal{G}_{(3,2)}(w_m, w'_m)$ . By Lemma 2.1, player II, in the first move, would need three positions  $p, q, r$  ( $p < q < r$ ) in  $w_m$  satisfying the following conditions (among others):  $Q_c^{w_m} p, Q_a^{w_m} q, Q_c^{w_m} r, 1 \sim_{(2)} w_m(p, q), (ba)^N c \sim_{(2)} w_m(q, r)$ . Assume such positions exist.  $a(ba)^N c \sim_{(2)} w_m(p, r)$  obviously implies that  $w_m(p, r)$  should be a sequence of  $a$ 's and  $b$ 's followed by the letter  $c$ . Hence player II should choose  $p$  and  $r$  as follows ( $p$  is the first overlined position and  $r$  the second one):

$$w_m = \dots \bar{c}(ba)^N c\bar{c} \dots$$

However, there is no position  $q$  between  $p$  and  $r$  satisfying both  $1 \sim_{(2)} w_m(p, q)$  (i.e.  $p$  and  $q$  should be consecutive) and  $Q_a^{w_m} q$ . Similarly to the proof of Lemma 4.8, we can show that  $w_m$  and  $w'_m$  are  $\sim_{(2, m)}$ -equivalent.  $\square$

**Lemma 4.10.** *If  $|A| \geq 3$  and  $n \geq 2$ , then  $\sim_{(2n, m)} \not\sim_{(2n+1, 1)}$ .*

**Proof.** The result is obvious if  $m = 1$  since  $\mathcal{N}(2n+1, 1) > \mathcal{N}(2n, 1)$ . So assume  $m > 1$  and let  $A$  contain at least the three letters  $a, b$  and  $c$ . Let  $n \geq 2$  be fixed,

$$\begin{aligned} x_0 &= x^N u'v(cv'bv)^{n-2} u'x^N, \\ x_1 &= x^N u'bv(cv'bv)^{n-2} u'x^N, \\ x_2 &= x^N u'u(cv'bv)^{n-2} u'x^N, \\ x_{2i+3} &= x^N u'v(cv'bv)^i (cu'bv)(cv'bv)^{n-3-i} u'x^N \quad \text{for } 0 \leq i \leq n-3, \\ x_{2i+4} &= x^N u'v(cv'bv)^i (cv'bu)(cv'bv)^{n-3-i} u'x^N \quad \text{for } 0 \leq i \leq n-3 \end{aligned}$$

and

$$x_{2n-1} = x^N u'v(cv'bv)^{n-2} cu'x^N,$$

where  $N-1 = \mathcal{N}(2n, m)$  and where  $u = (ab)^N(ca)^N$ ,  $v = (ab)^N a(ca)^N$ ,  $u' = (ac)^N(ba)^N$ ,  $v' = (ac)^N a(ba)^N$  and  $x = abc$ . Define

$$w_m = (x_1 \dots x_{2n-1})^N x_n (x_1 \dots x_{2n-1})^N$$

and

$$w'_m = (x_1 \dots x_{2n-1})^N x_0 (x_1 \dots x_{2n-1})^N.$$

We first show that  $w_m$  and  $w'_m$  are not  $\sim_{(2n+1, 1)}$ -equivalent. We illustrate a winning strategy for player I. Player I, in the first move, chooses the following (overlined)  $2n+1$  positions in  $w'_m$ :

$$\begin{aligned} w'_m &= \dots u'v(cv'bv)^{n-2} u' \dots \\ &= \dots (ac)^N (ba)^{N-1} \overline{b\bar{a}\bar{a}b} (ab)^{N-1} \overline{\bar{a}(ca)^N c(ac)^N \bar{a}(ba)^N b(ab)^N \bar{a}(ca)^N} \dots \\ &\quad c(ac)^N \bar{a}(ba)^N b(ab)^N \bar{a}(ca)^{N-1} \overline{c\bar{a}\bar{a}c} (ac)^{N-1} (ba)^N \dots \end{aligned}$$

More precisely, the chosen positions belong to the middle  $u'v(cv'bv)^{n-2} u'$ -segment of the  $x_0$ -segment of  $w'_m$ . They consist of the two middle  $a$ 's of the  $u'v$ -segment, the  $2n-3$  middle  $a$ 's of the  $v$ - and  $v'$ -segments, and the two middle  $a$ 's of the  $vu'$ -segment. Player II cannot win this play of the game  $\mathcal{G}_{(2n+1, 1)}(w_m, w'_m)$ . Player II, in the first move, would need  $2n+1$  positions  $p_1, \dots, p_{2n+1}$  ( $p_1 < \dots < p_{2n+1}$ ) in  $w_m$  satisfying the following conditions (among others):  $Q_a^{w_m} p_i$  for  $1 \leq i \leq 2n+1$ ,  $1 \sim_{(1)} w_m(p_1, p_2)$ ,  $ab \sim_{(1)} w_m(p_{2i}, p_{2i+1})$  for  $1 \leq i \leq n-1$ ,  $ac \sim_{(1)} w_m(p_{2i+1}, p_{2i+2})$  for  $1 \leq i \leq n-1$ , and  $1 \sim_{(1)} w_m(p_{2n}, p_{2n+1})$ . In  $w_m$ , no sequence  $u'v(cv'bv)^{n-2} u'$  exists. The best player II can find is a sequence  $u'bv(cv'bv)^{n-2} u'$ , or a sequence  $u'u(cv'bv)^{n-2} u'$ , or a sequence  $u'v(cv'bv)^i (cu'bv)(cv'bv)^{n-3-i} u'$  for some  $0 \leq i \leq n-3$ , or a sequence  $u'v(cv'bv)^i (cv'bu)(cv'bv)^{n-3-i} u'$  for some  $0 \leq i \leq n-3$ , or a sequence  $u'v(cv'bv)^{n-2} cu'$ . In the first situation, the first  $u'v$ -segment has been replaced by  $u'bv$ . For instance, we

would have

$$w_m = \dots (ac)^N (ba)^N b(ab)^N \bar{a}(ca)^N c(ac)^N \bar{a}(ba)^N b(ab)^N \bar{a}(ca)^N \dots \\ c(ac)^N \bar{a}(ba)^N b(ab)^N \bar{a}(ca)^{N-1} c \bar{a} \bar{a} c (ac)^{N-1} (ba)^N \dots,$$

where the overlined positions are (in order)  $p_3, p_4, p_5, \dots, p_{2n-2}, p_{2n-1}, p_{2n}, p_{2n+1}$ . However, there are no positions  $p_1$  and  $p_2$  before  $p_3$  satisfying  $Q_a^{w_m} p_1$  and  $Q_a^{w_m} p_2$ ,  $1 \sim_{(1)} w_m(p_1, p_2)$ , and  $ab \sim_{(1)} w_m(p_2, p_3)$ . The result similarly follows in the other situations.

We now show that  $w_m$  and  $w'_m$  are  $\sim_{(2n, m)}$ -equivalent. For the proof of  $w_m \sim_{(2n, m)} w'_m$  we distinguish the two cases where player I first picks  $2n$  positions from  $w_m$  or player I first picks  $2n$  positions from  $w'_m$ . Note that  $x_0 \sim_{(m)} x_n$ .

*Case 1:* Assume that player I has chosen  $2n$  positions from  $w_m$ . Player II will pick exactly corresponding positions in  $w'_m$ , except possibly when player I chooses some positions from the middle  $x_n$ -segment of  $w_m$ .

If some of the positions (in the first move) are chosen from the middle  $x_n$ -segment of  $w_m$ , we have

$$w_m = (x_1 \dots x_{2n-1})^N \overbrace{x^N \dots x^N}^{x_n} (x_1 \dots x_{2n-1})^N \\ = (x_1 \dots x_{2n-1})^N x^{M_1} \underline{x^{M_2}} x^{M_3} \dots x^{N_1} \underline{x^{N_2}} x^{N_3} (x_1 \dots x_{2n-1})^N,$$

where  $M_1 + M_2 + M_3 = N$ ,  $N_1 + N_2 + N_3 = N$ ,  $M_2, N_2 \geq m$  and the underlined segments  $x^{M_2}$  and  $x^{N_2}$  are free of chosen positions after the first move.

$$w'_m = (x_1 \dots x_{2n-1})^{N-1} (x_1 \dots x_{2n-1}) x_0 (x_1 \dots x_{2n-1})^N \\ = (x_1 \dots x_{2n-1})^{N-1} (x_1 \dots x_{n-1} x_n x_{n+1} \dots x_{2n-1}) x_0 (x_1 \dots x_{2n-1})^N \\ = (x_1 \dots x_{2n-1})^{N-1} (x_1 \dots x_{n-1} \overbrace{x^{M_1} x^{M_2} x^{M_3} \dots x^{N_1} x^{N_2} x^{N_3}}^{x_n} x_{n+1} \dots x_{2n-1}) \\ x_0 (x_1 \dots x_{2n-1})^N \\ \sim_{(2n, m)} (x_1 \dots x_{2n-1})^N x^{M_1} \underline{x^{N-M_1} u' b v (c v' b v)^{n-2} u' x^N x_2 \dots x_{n-1} x^{M_1} x^{M_2}} \\ x^{M_3} \dots x^{N_1} \underline{x^{N_2} x^{N_3} x_{n+1} \dots x_{2n-1} x^N u' v (c v' b v)^{n-2} u' x^{N_1} x^{N_2} x^{N_3} (x_1 \dots x_{2n-1})^N}.$$

Player II, in the first move, does not choose any position from the above underlined segments. The first underlined segment is  $\sim_{(m)}$ -equivalent to  $x^{M_2}$  and the second one to  $x^{N_2}$ . Player II chooses corresponding positions in the remaining segments.

*Case 2:* Assume now that player I has picked his first  $2n$  positions from  $w'_m$ . Player II will pick exactly corresponding positions in  $w_m$ , except possibly when player I picks

some positions from the middle  $x_0$ -segment of  $w'_m$ . Assume first that  $n \geq 3$ :

$$\begin{aligned} w'_m &= \dots x_0 \dots \\ &= \dots u'v(cv'bv)^{n-2}u' \dots \\ &= \dots u'av''(cv'bv)^{n-3}cv'bv''au' \dots \\ &= \dots u'ayau' \dots \end{aligned}$$

We have the following subcases.

*Case 2.1:* If after the first move, the middle  $y$ -segment of  $x_0$  has some of his  $v$ -,  $v'$ -,  $v''$ - or  $v'''$ -segments free from chosen positions, then the  $i$ th segment of  $y$  (a  $v$ -, a  $v'$ -, a  $v''$ - or a  $v'''$ -segment) is free of chosen positions for some  $1 \leq i \leq 2n - 3$ . Player II can pick the last  $x_{i+1}$ -segment in the initial  $(x_1 \dots x_{2n-1})^N$ -segment of  $w_m$  and play similarly as in case 1.

*Case 2.2:* Otherwise, after the first move, the middle  $y$ -segment of  $x_0$  has none of his  $v$ -,  $v'$ -,  $v''$ - or  $v'''$ -segments free from chosen positions and there are at most three chosen positions not in that middle  $y$ -segment of  $x_0$ .

If none of the chosen positions are at the left (right) of  $y$ , then player II can pick up the last  $x_1$  ( $x_{2n-1}$ )-segment in the initial  $(x_1 \dots x_{2n-1})^N$ -segment of  $w_m$  and play similarly as in case 1.

If two of the chosen positions are at the left of  $y$  and one at the right of  $y$ , then each of the  $v$ -,  $v'$ -,  $v''$ - and  $v'''$ -segments of the middle  $y$ -segment of  $x_0$  contains exactly one chosen position. We may assume that the chosen positions in the  $v$ -,  $v'$ -,  $v''$ - and  $v'''$ -segments are the middle  $a$ 's (otherwise, player II can play as in case 2.1). In this situation, player II can choose the last  $x_{2n-1}$ -segment in the initial  $(x_1 \dots x_{2n-1})^N$ -segment of  $w_m$  and play similarly as in case 1. The situation when one of the chosen positions is at the left of  $y$  and two at the right of  $y$  is similar.

If one of the chosen positions is at the left of  $y$  and one at the right of  $y$ , then either the  $v''$ - or the  $v'''$ -segment of the middle  $y$ -segment of  $x_0$  contains exactly one chosen position which we may assume to be the middle  $a$ . The result follows similarly as in the preceding situation.

Now, if  $n = 2$ ,

$$\begin{aligned} w'_m &= \dots x_0 \dots \\ &= \dots u'ayau' \dots \end{aligned}$$

If after the first move, the middle  $y$ -segment of  $x_0$  is free from chosen positions, then player II can pick the last  $x_2$ -segment in the initial  $(x_1x_2x_3)^N$ -segment of  $w_m$  and play similarly as in case 1. Otherwise, after the first move, the middle  $y$ -segment of  $x_0$  is not free from chosen positions and there are at most three chosen positions not in that middle  $y$ -segment of  $x_0$ . If none of the chosen positions are at the left (right) of  $y$ , then player II can pick up the last  $x_1$  ( $x_3$ )-segment in the initial  $(x_1x_2x_3)^N$ -segment of  $w_m$  and play similarly as in case 1. If two of the chosen positions are at the left of  $y$  and one at the right of  $y$ , then the middle  $y$ -segment of  $x_0$  contains exactly one chosen position which we may assume to be the middle  $a$ . In this situation,

player II can choose the last  $x_3$ -segment in the initial  $(x_1x_2x_3)^N$ -segment of  $w_m$  and play similarly as in case 1. The situation when one of the chosen positions is at the left of  $y$  and two at the right of  $y$  is similar. If one of the chosen positions is at the left of  $y$  and one at the right of  $y$ , then either the initial  $b(ab)^m$ -segment of  $y$  is free of chosen positions, or the middle  $(ab)^ma(ca)^m$ -segment or the last  $(ca)^mc$ -segment. In the first case, player II can pick the last  $x_1$ -segment of the initial  $(x_1x_2x_3)^N$ -segment of  $w_m$  and play as in case 1, in the second case, he can choose the last  $x_2$ -segment, and in the third case, the last  $x_3$ -segment. The result follows.  $\square$

#### 4.2. Between the congruences $\sim_{(n,m)}$ and $\sim_{(n,1,m)}$

The purpose of this section is to give an inclusion relation between  $\sim_{(n',m')}$  and  $\sim_{(n,1,m)}$ . Applications are discussed in the next section.

**Lemma 4.11.** *If  $|A| = r \geq 2$ , then  $\sim_{(2nM, \mathcal{N}(3,m))} \subseteq \sim_{(n,1,m)}$ , where  $2M$  is the maximum number of  $(m)_r$  positions in words over  $A$ .*

**Proof.** Let  $u, v \in A^+$  be such that  $u \sim_{(2nM, 4m+3)} v$ . If  $u$  and  $v$  consist of one letter only, then the result follows by Lemma 2.2 since  $\mathcal{N}(n, 1, m) \leq \mathcal{N}(2nM, 4m+3)$ . Otherwise, we first show the result for  $n = 1$ . We want to show that  $u \sim_{(1,1,m)} v$ . Let  $p$  be a position in  $u$  chosen by player I in the first move (the proof is similar if starting in  $v$ ). Player II chooses a position  $q$  in  $v$  by considering the following play of the game  $\mathcal{G}_{(2M, 4m+3)}(u, v)$ . In the first move, player I chooses the  $(m)_r$  last positions in  $u[1, p]$ , say  $p_1, \dots, p_M$  ( $p_1 \leq \dots \leq p_M$ ), and the  $(m)_r$  first positions in  $u(p, |u|]$ , say  $p_{M+1}, \dots, p_{2M}$  ( $p_{M+1} \leq \dots \leq p_{2M}$ ), for a total of at most  $2M$  positions. There exist  $q_1, \dots, q_{2M}$  in  $v$  ( $q_1 \leq \dots \leq q_{2M}$ ) satisfying Lemma 2.1.  $u(p_M, p_{M+1}) \sim_{(4m+3)} v(q_M, q_{M+1})$  implies the existence of  $q \in v$  such that  $Q_a^u p$  if and only if  $Q_a^v q$ ,  $a \in A$ ,  $u(p_M, p) \sim_{(2m+1)} v(q_M, q)$  and  $u(p, p_{M+1}) \sim_{(2m+1)} v(q, q_{M+1})$ . Since  $u \sim_{(2M, 4m+3)} v$  and  $u[1, p] \sim_{(2m+1)} v[1, q]$ , the following claim concerning the  $(m)_r$  positions in  $u[1, p]$  and  $v[1, q]$  holds. Let  $r_1, \dots, r_s \in u[1, p]$  ( $r_1 < \dots < r_s$ ) ( $r'_1, \dots, r'_s \in v[1, q]$  ( $r'_1 < \dots < r'_s$ )) be the  $(m)_r$  positions in  $u[1, p]$  ( $v[1, q]$ ). We have that  $s = s'$ ,  $Q_a^u r_i$  if and only if  $Q_a^v r'_i$ ,  $a \in A$  for  $1 \leq i \leq s$ ,  $u[1, r_1] \sim_{(2m+1)} v[1, r'_1]$ ,  $u(r_i, r_{i+1}) \sim_{(2m+1)} v(r'_i, r'_{i+1})$  for  $1 \leq i \leq s-1$ , and  $u(r_s, p) \sim_{(2m+1)} v(r'_s, q)$ . A similar claim holds for the  $(m)_r$  positions in  $u(p, |u|]$  and  $v(q, |v|]$ . Using fact 6 (of Section 3), we can conclude that  $u[1, p] \sim_{(1,m)} v[1, q]$  and  $u(p, |u|] \sim_{(1,m)} v(q, |v|]$ .

Now, if  $n > 1$  and player I in the first move chooses  $n$  positions in  $u$  or in  $v$ , a winning strategy for player II in the game  $\mathcal{G}_{(n,1,m)}(u, v)$  to win each play is described as follows. Let  $p'_1, \dots, p'_n$  ( $p'_1 \leq \dots \leq p'_n$ ) be positions in  $u$  chosen by player I in the first move (the proof is similar when starting in  $v$ ). Player II chooses positions  $q'_1, \dots, q'_n$  in  $v$  ( $q'_1 \leq \dots \leq q'_n$ ) by considering the following play of the game  $\mathcal{G}_{(2nM, 4m+3)}(u, v)$ . In the first move, player I chooses the  $(m)_r$  last positions in  $u[1, p'_1]$ , the  $(m)_r$  positions in  $u(p'_1, p'_2), \dots, u(p'_{n-1}, p'_n)$ , and the  $(m)_r$  first positions in  $u(p'_n, |u|]$  for a total of at most  $2nM$  positions. The result follows similarly as above.  $\square$

## 5. Decidability and inclusion results

### 5.1. On 2-dot-depth

This section deals with a first application of the results of the preceding section. Let  $A^*/\sim_{\bar{m}}$  be of dot-depth two, i.e.  $\bar{m}$  is either of the form  $(n, m)$  or  $(n, 1, m)$ . The 2-dot-depth (abbreviated *2dd*) of  $A^*/\sim_{\bar{m}}$  is defined as the smallest  $n$  for which  $A^*/\sim_{\bar{m}} \in V_{2, n}$ . We show that the 2-dot-depth of  $A^*/\sim_{(n, m)}$  is computable for an arbitrary  $A$ , and the 2-dot-depth of  $A^*/\sim_{(n, 1, m)}$  is computable for  $|A| = 2$ . For  $|A| \geq 3$ , an upper bound on the 2-dot-depth of  $A^*/\sim_{(n, 1, m)}$  is given.

**Theorem 5.1.** *The 2-dot-depth of  $A^*/\sim_{(n, m)}$  is computable for an arbitrary  $A$ .*

**Proof.**  $2dd(A^*/\sim_{(1, m)}) = 1$ .  $2dd(A^*/\sim_{(2n, m)}) = 2n$  by Lemma 4.5. If  $|A| = 2$ , then  $2dd(A^*/\sim_{(2n+1, m)}) = 2n$  by Lemmas 4.4 and 4.5. If  $|A| = 3$ , then  $2dd(A^*/\sim_{(3, 1)}) = 2$  by Lemmas 4.5 and 4.7. If  $|A| \geq 4$ , then  $2dd(A^*/\sim_{(3, 1)}) = 3$  by Lemma 4.8. If  $|A| \geq 3$  and  $m > 1$ , then  $2dd(A^*/\sim_{(3, m)}) = 3$  by Lemma 4.9. In the case where  $|A| \geq 3$  and  $n \geq 2$ , we have  $2dd(A^*/\sim_{(2n+1, m)}) = 2n + 1$  by Lemma 4.10.  $\square$

**Theorem 5.2.** *If  $|A| = 2$ , then  $2dd(A^*/\sim_{(n, 1, m)}) = 2n$ .*

**Proof.** First,  $\sim_{(n, 1, m)} \subseteq \sim_{(\mathcal{N}(n, 1), m)}$  by Lemma 2.2, or  $\sim_{(n, 1, m)} \subseteq \sim_{(2n+1, m)}$ . Using the proof of Theorem 5.1,  $2dd(A^*/\sim_{(n, 1, m)})$  is  $d \geq 2n$  since  $2dd(A^*/\sim_{(2n+1, m)}) = 2n$ . Lemma 4.11 shows that  $d \leq 2n$ .  $\square$

**Theorem 5.3.** *If  $|A| = r \geq 3$ , then  $2dd(A^*/\sim_{(n, 1, m)}) \leq 2n \sum_{i=0}^{r-2} m^i$ .*

**Proof.** By Lemma 4.11.  $\square$

### 5.2. On a conjecture of Pin

This section deals with a second application of the results in Section 4.

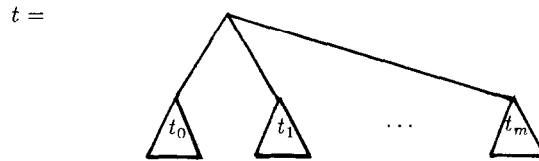
We will denote by  $\mathcal{T}$  the set of trees on the alphabet  $\{a, \bar{a}\}$ . Formally,  $\mathcal{T}$  is the set of words in  $\{a, \bar{a}\}^*$  congruent to 1 in the congruence generated by the relation  $a\bar{a} = 1$ . Intuitively, the words in  $\mathcal{T}$  are obtained as follows: 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  $aa\bar{a}aa\bar{a}\bar{a}\bar{a}\bar{a}\bar{a}\bar{a}\bar{a}\bar{a}$ . 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  $t = at_1\bar{a}at_2\bar{a} \dots at_m\bar{a}$ , where  $m \geq 0$  and where the  $t_i$ 's are trees. Let  $t$  be

a tree and let  $t = t_1 a t_2 \bar{a} t_3$  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  $\bar{a}$ .

To the tree reduced to a point is associated  $\{\emptyset, A^*\}$ . Then to the tree



is associated the boolean algebra  $\mathcal{V}_t$  which is generated by all the languages of the form  $L_{i_0} a_1 L_{i_1} a_2 \dots a_r L_{i_r}$ , with  $0 \leq i_0 < \dots < i_r \leq m$ , where, for  $0 \leq j \leq r$ ,  $L_{i_j} \in \mathcal{V}_{t_{i_j}}$ . In Pin [13] it was shown that if  $t$  is extracted from  $t'$ , then  $\mathcal{V}_t \subseteq \mathcal{V}_{t'}$ , and it was conjectured that if  $t, t' \in \mathcal{T}'$ ,  $\mathcal{V}_t \subseteq \mathcal{V}_{t'}$  if and only if  $t$  is extracted from  $t'$ . Here,  $\mathcal{T}'$  denotes the set of trees in which each node is of arity different from 1.

Let  $\bar{m} = (m_1, \dots, m_k)$ . By induction on  $k$ , we define a tree  $t_{\bar{m}}$  as follows: if  $\text{length}(\bar{m}) = 1$ , then  $t_{\bar{m}} = (a\bar{a})^{m_1+1}$ , for  $\bar{m} = (n, m_1, \dots, m_k)$ ,  $t_{\bar{m}} = (a t_{(m_1, \dots, m_k)} \bar{a})^{n+1}$ . It is easy to see that  $l(t_{(m_1, \dots, m_k)})$  is  $\mathcal{N}(m_1, \dots, m_k) + 1 = (m_1 + 1) \dots (m_k + 1)$ .  $\mathcal{V}_{t_{\bar{m}}} = \mathcal{L}_{\bar{m}}$ , where  $\mathcal{L}_{\bar{m}}$  denotes the  $*$ -variety of languages which are unions of classes of  $\sim_{\bar{m}}$  (details appear in [1]).

**Theorem 5.4.** *The above conjecture is false.*

**Proof.** For any language  $L$  over  $A$ , the syntactic congruence of  $L$  is defined by  $x \sim_L y$  if and only if, for all  $u, v \in A^*$ ,  $uxv \in L$  if and only if  $uyv \in L$ .  $L$  is a union of classes of a congruence  $\sim$  if and only if  $\sim \subseteq \sim_L$ . Now,  $\mathcal{L}_{(1,2)} \subseteq \mathcal{L}_{(2,1)}$  since  $\sim_{(2,1)} \subseteq \sim_{(1,2)}$  (a special case of Lemma 4.1). Hence  $\mathcal{V}_{t_{(1,2)}} \subseteq \mathcal{V}_{t_{(2,1)}}$ . However, it is easy to verify that the tree  $t_{(1,2)}$  is not extracted from the tree  $t_{(2,1)}$ .  $\square$

### 5.3. Equations and the $V_{2,n}$ 's

The problem of finding equations satisfied in the monoid variety  $V_2$  and the monoid varieties  $V_{2,n}$ , a problem related to the decidability of  $V_2$  and the  $V_{2,n}$ 's, is the subject of this section.

Let  $w, w' \in A^*$ . A monoid  $M$  satisfies the equation  $w = w'$  if and only if  $w\varphi = w'\varphi$  for all morphisms  $\varphi: A^* \rightarrow M$ . One can show that the class of monoids  $M$  satisfying the equation  $w = w'$  is a monoid variety, denoted by  $W(w, w')$ . Let  $(w_m, w'_m)_{m > 0}$  be a sequence of pairs of words of  $A^*$ . Consider the following monoid variety:  $W = \bigcup_{n > 0} \bigcap_{m \geq n} W(w_m, w'_m)$ . We say that  $W$  is *ultimately defined* by the equations  $w_m = w'_m$  ( $m > 0$ ): this corresponds to the fact that a monoid  $M$  is in  $W$  if and only if  $M$  satisfies the equations  $w_m = w'_m$  for all  $m$  sufficiently large. The equational approach to varieties is discussed in Eilenberg [9]. Eilenberg showed that every monoid variety

is ultimately defined by a sequence of equations. For example, the monoid variety  $V$  of aperiodic monoids is ultimately defined by the equations  $x^m = x^{m+1}$  ( $m > 0$ ).

**Theorem 5.5.** *Every monoid in  $V_{2,n}$  satisfies  $w_m = w'_m$  for all sufficiently large  $m$ , where  $w_m$  and  $w'_m$  denote the words in Lemmas 4.2, 4.5 and 4.8–4.10 that are shown to be  $\sim_{(n,m)}$ -equivalent, or the words in Lemma 4.6 that are shown to be  $\sim_{(n,2m)}$ -equivalent.*

**Proof.** It is easily seen, using the above lemmas, that monoids in  $V_{2,n}$  satisfy  $w_m = w'_m$  for some  $m$ . This comes from the fact that if  $M \in V_{2,n}$ , then  $M$  divides  $A^*/\sim_{(n,m)}$  for some  $m$ . Since  $A^*/\sim_{(n,m)}$  satisfies  $w_m = w'_m$ ,  $M$  satisfies  $w_m = w'_m$ . Moreover, if  $M$  in  $V_{2,n}$  satisfies  $w_m = w'_m$  for some  $m$ , then it satisfies  $w_{m'} = w'_{m'}$  for all  $m' \geq m$  since  $\sim_{(n,m')} \subseteq \sim_{(n,m)}$  for those  $m'$ .  $\square$

#### 5.4. Generalizations to the congruences $\sim_{\bar{m}}$

This section gives generalizations of some of the results in Section 4.

**Theorem 5.6.** *Let  $k \geq 3$ .  $\sim_{(n_1, \dots, n_k)} \subseteq \sim_{(1,m)}$  if and only if  $\mathcal{N}(n_1, \dots, n_k) \geq \mathcal{N}(1, m)$ .*

**Proof.** The necessity of the condition comes from Lemma 2.2  $\sim_{(n_1, \dots, n_k)} \subseteq \sim_{(\mathcal{N}(n_1, \dots, n_{k-1}), n_k)}$  by Lemma 2.2. Since we have both  $\mathcal{N}(\mathcal{N}(n_1, \dots, n_{k-1}), n_k) = \mathcal{N}(n_1, \dots, n_k) \geq \mathcal{N}(1, m)$  and  $\mathcal{N}(n_1, \dots, n_{k-1}) \neq 2$ , then  $\sim_{(\mathcal{N}(n_1, \dots, n_{k-1}), n_k)} \subseteq \sim_{(1,m)}$  by Lemma 4.1(1). The sufficiency of the condition follows.  $\square$

**Theorem 5.7.** *Let  $k' \geq 2$  and  $n > 1$ . If there exists  $1 \leq k < k'$  such that*

$$m > \mathcal{N}(n_{k+1}, \dots, n_{k'}) \quad \text{and} \quad (2 + (n-1)M)m \leq \mathcal{N}(n_1, \dots, n_k) \mathcal{N}(n_{k+1}, \dots, n_{k'}),$$

*then  $\sim_{(n_1, \dots, n_{k'})} \subseteq \sim_{(n,m)}$  where  $2\lfloor(m-1)/m'\rfloor M$  is the maximum number of  $(m, m')$  positions in words over  $A$ .*

**Proof.** The special case  $k' = 2$  is Lemma 4.3. The general statement follows from that special case and Lemma 2.2 since  $\sim_{(n_1, \dots, n_{k'})} \subseteq \sim_{(n_1, \dots, n_k, n_{k+1}, \dots, n_{k'})} \subseteq \sim_{(\mathcal{N}(n_1, \dots, n_k), \mathcal{N}(n_{k+1}, \dots, n_{k'}))}$ .  $\square$

#### Acknowledgement

The problem of finding conditions for  $\sim_{(n_1, \dots, n_k)}$  to be included in  $\sim_{(m_1, \dots, m_k)}$  was suggested to me by Jean-Eric Pin and Wolfgang Thomas. This research was initiated under an NSERC (Natural Sciences and Engineering Research Council of Canada) Postdoctoral Fellowship. Many thanks to the referee of a preliminary version of this paper for his valuable comments and suggestions.

## References

- [1] F. Blanchet-Sadri, Some logical characterizations of the dot-depth hierarchy and applications, Ph.D. Thesis, McGill University, Montreal, Que. (1989).
- [2] F. Blanchet-Sadri, Games, equations and the dot-depth hierarchy, *Comput. Math. Appl.* 18 (1989) 809–822.
- [3] F. Blanchet-Sadri, Combinatorial properties of some congruences related to dot-depth  $k$ , Tech. Rept. No. 89–27, Department of Mathematics and Statistics of McGill University, Montreal, Que. (1989) 1–29.
- [4] F. Blanchet-Sadri, On dot-depth two, *RAIRO Inform. Théor. Appl.* 24 (1990) 521–529.
- [5] F. Blanchet-Sadri, Games, equations and dot-depth two monoids, *Discrete Appl. Math.*, to appear.
- [6] J.A. Brzozowski and R. Knast, The dot-depth hierarchy of star-free languages is infinite, *Comput. System Sci.* 16 (1978) 37–55.
- [7] R.S. Cohen and J.A. Brzozowski, Dot-depth of star-free events, *Comput. System Sci.* 5 (1971) 1–15.
- [8] H.-D. Ebbinghaus, J. Flum and W. Thomas, *Mathematical Logic* (Springer, New York, 1984).
- [9] S. Eilenberg, *Automata, Languages and Machines*, Vol. B (Academic Press, New York, 1976).
- [10] R. McNaughton and S. Papert, *Counter-Free Automata* (MIT Press, Cambridge, MA, 1971).
- [11] D. Perrin and J.-E. Pin, First-order logic and star-free sets, *Comput. System Sci.* 32 (1986) 393–406.
- [12] J.-E. Pin, *Variétés de Langages Formels* (Masson, Paris, 1984).
- [13] J.-E. Pin, Hiérarchies de concaténation, *RAIRO Inform. Théor.* 18 (1984) 23–46.
- [14] M.P. Schützenberger, On finite monoids having only trivial subgroups, *Inform. and Control* 8 (1965) 190–194.
- [15] I. Simon, Piecewise testable events, in: *Proceedings of the Second GI Conference*, Lecture Notes in Computer Science 33 (Springer, Berlin, 1975) 214–222.
- [16] H. Straubing, Finite semigroup varieties of the form  $V*D$ , *Pure Appl. Algebra* 36 (1985) 53–94.
- [17] H. Straubing, Semigroups and languages of dot-depth two, in: *Proceedings of the Thirteenth ICALP*, Lecture Notes in Computer Science 226 (Springer, New York, 1986) 416–423.
- [18] W. Thomas, Classifying regular events in symbolic logic, *Comput. System Sci.* 25 (1982) 360–376.
- [19] W. Thomas, An application of the Ehrenfeucht–Fraïssé game in formal language theory, *Bull. Soc. Math. France* 16 (1984) 11–21.
- [20] W. Thomas, A concatenation game and the dot-depth hierarchy, in: E. Böger, ed., *Computation Theory and Logic*, Lecture Notes in Computer Science 270 (Springer, New York, 1987) 415–426.