Languages recognized by finite supersoluble groups *

Olivier Carton† Jean-Eric Pin† Xaro Soler-Escrivà‡

April 2009

Abstract

In this paper, we give two descriptions of the languages recognized by finite supersoluble groups. We first show that such a language belongs to the Boolean algebra generated by the modular products of elementary commutative languages. An elementary commutative language is defined by a condition specifying the number of occurrences of each letter in its words, modulo some fixed integer. Our second characterization makes use of counting functions computed by transducers in strict triangular form.

Eilenberg's variety theorem [7] is a powerful tool for classifying regular languages. It states that, given a variety of finite monoids $\mathbf{V}$, the class of languages $\mathcal{V}$ whose syntactic monoid belongs to $\mathbf{V}$ is a *variety of languages*, that is, a class of regular languages closed under finite union, complement, left and right quotients and inverse of morphisms. Further, the correspondence $\mathbf{V} \to \mathcal{V}$ between varieties of finite monoids and varieties of languages is one-to-one and onto.

Eilenberg's theorem can be used in both ways: given a variety of languages, one can look for the corresponding variety of monoids, or, given a variety of monoids, one can seek for a combinatorial description of the corresponding variety of languages. Examples abound in the literature: for instance, aperiodic monoids correspond to star-free languages, $\mathcal{J}$-trivial languages to piecewise testable languages, etc. We refer the reader to [10] for a survey.

It is therefore natural to ask for a nice characterization of the variety of languages corresponding to the variety of groups. The answer to this frequently asked question is unfortunately negative: there is no known satisfactory answer to this question. The reason is hidden in the complexity of finite groups since a solution would probably require a description of the languages recognized by each finite simple group . . .

However, solutions are known for some important subvarieties: abelian groups [7], p-groups [7, 17, 18], nilpotent groups [7, 16] and soluble groups [14, 18]. The

*The authors acknowledge support from the AutoMathA programme of the European Science Foundation. The second author was supported by the Grant AINV07/093 from the Conselleria d'Empresa, Universitat i Ciència de la Generalitat Valenciana and the third author was supported by the Grant PR2007-0164 from MEC of Spain.

†LIAFA, Université Paris VII and CNRS, Case 7014, 75205 Paris Cedex 13, France.

‡Dpt. de Matemàtica Aplicada, Universitat d'Alacant, Sant Vicent del Raspeig, Ap. Correus 99, E – 03080 Alacant.

aim of this paper is to complete these results by giving a description of the languages corresponding to the variety of supersoluble groups.

We first proceed in Section 1 to an algebraic study of the variety of supersoluble groups. Most of the results of this section were actually known before, but we try to present them in a selfcontained way that suits our needs for the next sections. We show in particular that the variety of supersoluble groups is generated by the Borel groups $B_n(\mathbb{F}_p)$ for all $n > 0$ and all primes p.

In Section 2, we first state, in a slightly improved version, the description of the languages recognized by abelian groups. In this new version, these languages are described as a disjoint union of "elementary languages", which have a simple combinatorial description. We also define the modular concatenation product, an operation on languages first introduced by Straubing [14].

Our main result (Corollary 2.8) states that the languages recognized by supersoluble groups can be obtained in two steps: first take the modular products of elementary languages and then take the Boolean algebra generated by these languages.

In the last part of the paper, we give another characterization of the languages recognized by supersoluble groups, which relies on the following idea. Given a function τ from words to numbers and an integer r, consider the language of all words u such that $\tau(u) = r$. One can say that this language is defined by *counting modulo* τ. This leads to the idea of describing regular languages by suitable *counting functions*. It turns out that this idea is very successful for describing group languages: for instance, languages of abelian groups can be described by counting letters and languages of p-groups can be described by counting subwords. Our second description of the languages recognized by supersoluble groups (Corollary 2.14) makes use of counting functions computed by transducers in *strict triangular form* (the precise definition can be found in Section 2.4). It would be nice to have a simple combinatorial description of these transducers, but there is unfortunately no evidence that such a description exists.

1 The variety of supersoluble groups

Throughout this paper, the term *variety* will be used to mean a class of finite groups (or monoids) closed under finite direct products, subgroups (submonoids) and morphic images.

Given two groups G and H, we use the standard notation $H \leqslant G$ (respectively $H \cong G$) to mean that H is a subgroup of G (respectively H is isomorphic to G). Finally, for any element $g \in G$ and any subgroup H of G, the g-conjugate of H is the set $H^g = g^{-1}Hg$.

Given a family of varieties of groups $(\mathbf{H}_i)_{i \in I}$, their *join* consists of all groups which are quotients of subgroups of direct products $H_1 \times \cdots \times H_n$ with $H_k \in \mathbf{H}_{i_k}$, for some $i_k \in I$. It is the smallest variety of groups which contains $\mathbf{H}_i$ for all $i \in I$. Given two varieties of groups $\mathbf{U}$ and $\mathbf{V}$, the *product variety* $\mathbf{U} * \mathbf{V}$ consists of all groups G having a normal subgroup $U \in \mathbf{U}$ such that $G/U \in \mathbf{V}$.

For a prime p, $\mathbf{G}_p$ denotes the variety of all p-groups. For any positive integer d, $\mathbf{Ab}(d)$ denotes the variety of all abelian groups of exponent dividing d. A group G is *supersoluble* if it has a normal series with cyclic factors. In particular, all nilpotent groups are supersoluble and any supersoluble group is soluble. The class of all finite supersoluble groups form a variety of groups. A description of supersoluble groups in terms of p-groups and abelian groups was given in [3].

Proposition 1.1 *The variety of all finite supersoluble groups is the join of the varieties* $\mathbf{G}_p * \mathbf{Ab}(p-1)$*, where* p *runs over all primes.*

Proposition 1.1 is actually a particular case of a general result from the theory of formations of groups (see [6, IV] for more details on this topic). The class $\mathbf{U}$ of supersoluble groups is a saturated formation and, as such, can be defined by a local formation function f which associates with each prime p a (possibly empty) formation of groups $f(p)$. The class $\mathbf{U}$ is locally defined by $f(p) = \mathbf{Ab}(p-1)$, for all primes p [5], [6, IV, (3.4)]. In general, a saturated formation $\mathbf{F}$ possesses many local definitions but it has a unique full and integrated local formation function F [6, IV, (3.7)]. In the supersoluble case, it is given by $F(p) = \mathbf{G}_p * \mathbf{Ab}(p-1)$, for all primes p, [6, IV, (3.8)]. Now, given a saturated formation $\mathbf{F}$ and its full and integrated local definition F, it is easy to see that $\mathbf{F}$ is just the formation generated by $F(p)$, when p runs over all primes. Since $F(p)$ is included in $\mathbf{F}$ for all primes p, so is their join $\mathbf{H}$. Conversely, assume that $\mathbf{F} \setminus \mathbf{H}$ is not empty and consider $G \in \mathbf{F}$ of minimal order such that $G \notin \mathbf{H}$. By [6, II, (2.5)], G has a unique minimal normal subgroup N. If p is a prime dividing the order of N then $G \in F(p)$, [6, IV, (3.2)]. This contradiction shows that $\mathbf{F} = \mathbf{H}$.

There is another interesting characterization of this variety. Let p be a prime number and $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ the field with p elements. Let us denote by $GL_n(\mathbb{F}_p)$ the group of all invertible $n \times n$ matrices with entries in $\mathbb{F}_p$ and by $B_n(\mathbb{F}_p)$ the group of all invertible upper triangular matrices of $GL_n(\mathbb{F}_p)$. The group $B_n(\mathbb{F}_p)$ is known as the *standard Borel subgroup* of $GL_n(\mathbb{F}_p)$. Finally, we denote by $U_n(\mathbb{F}_p)$ the group of unitriangular matrices of $B_n(\mathbb{F}_p)$ (upper triangular matrices with ones on the diagonal) and by $D_n(\mathbb{F}_p)$ the group of all diagonal matrices of $B_n(\mathbb{F}_p)$. It is well known that $U_n(\mathbb{F}_p)$ is a Sylow p-subgroup of $GL_n(\mathbb{F}_p)$ which is normal in $B_n(\mathbb{F}_p)$ and that $B_n(\mathbb{F}_p) = U_n(\mathbb{F}_p)D_n(\mathbb{F}_p)$, see [2, pages 50, 64]. Since $D_n(\mathbb{F}_p)$ is isomorphic with $(\mathbb{F}_p^*)^n$, it follows that $B_n(\mathbb{F}_p)$ belongs to the variety $\mathbf{G}_p * \mathbf{Ab}(p-1)$.

The next theorem is also well known.

Theorem 1.2 *A group belongs to the variety* $\mathbf{G}_p * \mathbf{Ab}(p-1)$ *if and only if it is isomorphic to a subgroup of* $B_n(\mathbb{F}_p)$ *for some* $n > 0$*.*

Proof. Let G be a group belonging to $\mathbf{G}_p * \mathbf{Ab}(p-1)$. There exists a normal subgroup N of G such that N is a p-group and $G/N \in \mathbf{Ab}(p-1)$. By [6, A, (11.3)], there exists a subgroup C of G such that $G = NC$ and $C \cong G/N$. If V is a simple $\mathbb{F}_pG$-module then N centralizes V, by [6, B, (3.12)]. In particular, this implies that V is a simple $\mathbb{F}_pG$-module if and only if V is a simple $\mathbb{F}_pC$-module. Now, since $C \in \mathbf{Ab}(p-1)$, any simple $\mathbb{F}_pC$-module has dimension 1 over $\mathbb{F}_p$ [6, B, (9.8)]. Thus, any simple $\mathbb{F}_pG$-module has dimension 1 over $\mathbb{F}_p$.

Let R be the regular $\mathbb{F}_pG$-module and let $\{0\} = R_0 < R_1 < \cdots < R_n = R$ be a composition series of R as $\mathbb{F}_pG$-module. We use this composition series in order to chose a basis of R over $\mathbb{F}_p$. Any factor R_{i+1}/R_i is a simple $\mathbb{F}_pG$-module and then $\dim_{\mathbb{F}_p}(R_{i+1}/R_i) = 1$ for $i = 0, \dots, n-1$. We choose $r_1 \in R_1$ a basis of R_1 over $\mathbb{F}_p$. Now, let $r_2 + R_1$ a basis of R_2/R_1. Notice that $\{r_1, r_2\}$

is a linearly independent set. Now, assume we have chosen $\{r_1, \ldots, r_i\}$ in that way and take $r_{i+1} + R_i$ a basis of R_{i+1}/R_i. Then $\{r_1, \ldots, r_{i+1}\}$ is a linearly independent set. Let $\mathcal{B} = \{r_1, \ldots, r_n\}$ be a basis of R over $\mathbb{F}_p$ constructed in this way. Let $\varphi : G \to GL_n(\mathbb{F}_p)$ the matrix representation of G afforded by $\mathcal{B}$. Since G acts faithfully over R, it follows that G is isomorphic to a subgroup of $GL_n(\mathbb{F}_p)$. Moreover, the choice of $\mathcal{B}$ assures that the matrix associated to any element g of G is a triangular matrix. Thus, G is isomorphic with a subgroup of $B_n(\mathbb{F}_p)$.

Corollary 1.3 *The variety of supersoluble groups is generated by the Borel groups $B_n(\mathbb{F}_p)$ for all $n > 0$ and all primes p.*

2 Languages

We shall denote by $\mathcal{U}_p$ the variety of languages associated with the variety of groups $\mathbf{G}_p * \mathbf{Ab}(p-1)$ and by $\mathcal{U}$ the variety of languages associated with the variety of supersoluble groups. Proposition 1.1 shows that $\mathcal{U}$ is the join of the varieties of languages $\mathcal{U}_p$, for any prime p. These varieties of languages will be described in Section 2.3. Before that, we need a precise description of the varieties of languages corresponding to $\mathbf{Ab}(n)$ and to $\mathbf{G}_p$.

2.1 Languages recognized by Abelian groups

Let us call *n-commutative* a language recognized by a group in $\mathbf{Ab}(n)$. The set of n-commutative languages of A^* is denoted by $\mathcal{A}b(n)(A^*)$. A description of these languages was given in [7]. It relies on the fact that this variety is generated by the cyclic groups of order n.

Proposition 2.1 *For each alphabet A, the n-commutative languages of A^* form the Boolean algebra generated by the languages of the form*

$$F(a, k, n) = \{u \in A^* \mid |u|_a \equiv k \bmod n\} = ((B^*a)^n)^*(B^*a)^k B^*,$$

where $a \in A$, $B = A \setminus \{a\}$ and $0 \leqslant k < n$.

We shall need an improved version of this result, which avoids using complementation. Let $A = \{a_1, \ldots, a_s\}$ be an alphabet. Let us call *n-elementary commutative* a language of the form

$$F(r_1, \ldots, r_s, n) = \{u \in A^* \mid |u|_{a_1} \equiv r_1, \ \ldots, \ |u|_{a_s} \equiv r_s \bmod n\}$$

where $r_1, \ldots, r_s \in \{0, \ldots, n-1\}$. Thus, with the notation of Proposition 2.1,

$$F(r_1, \ldots, r_s, n) = F(a_1, r_1, n) \cap \ldots \cap F(a_s, r_s, n)$$

Proposition 2.2 *A language is n-commutative if and only if it is a disjoint union of n-elementary commutative languages.*

Proof. Let $A = \{a_1, \ldots, a_s\}$, let G be a group in $\mathbf{Ab}(n)$ and let $\varphi : A^* \to G$ be a morphism. If L is recognized by φ, then $L = \varphi^{-1}(P)$ for some subset P of

G. Put $\varphi(a_1) = g_1, \ldots, \varphi(a_s) = g_s$. Let $u \in A^*$ and, for $1 \leqslant i \leqslant s$, let $|u|_{a_i} \equiv r_i \bmod n$. Adopting an additive notation for G, we get

$$\varphi(u) = \sum_{1 \leqslant i \leqslant s} |u|_{a_i} g_i = \sum_{1 \leqslant i \leqslant s} r_i g_i$$

Therefore $u \in L$ if and only if $\sum_{1 \leqslant i \leqslant s} r_i g_i \in P$ and hence

$$L = \bigcup_{(r_1, \ldots, r_s) \in E} F(r_1, \ldots, r_s, n)$$

where $E = \{(r_1, \ldots, r_s) \mid \sum_{1 \leqslant i \leqslant s} r_i g_i \in P\}$. This concludes the proof, since the languages $F(r_1, \ldots, r_s, n)$ are clearly pairwise disjoint.

2.2 Languages recognized by p-groups

A few auxiliary definitions are required to describe the variety of languages $\mathcal{G}_p$ associated with $\mathbf{G}_p$, for a given prime p.

A word $u = a_1 a_2 \cdots a_n$ (where $a_1, \ldots, a_n$ are letters) is a *subword* of a word v if v can be factored as $v = v_0 a_1 v_1 \cdots a_n v_n$. For instance, ab is a subword of $cacbc$. Given two words u and v, we denote by $\binom{v}{u}$ the number of distinct ways to write u as a subword of v.

More formally, if $u = a_1 a_2 \cdots a_n$, then

$$\binom{v}{u} = \operatorname{Card}\{(v_0, v_1, \ldots, v_n) \mid v_0 a_1 v_1 \cdots a_n v_n = v\}$$

Observe that if u is a letter a, then $\binom{v}{a}$ is simply the number of occurrences of the letter a in v, also denoted by $|v|_a$.

The following result is credited to Eilenberg and Schützenberger in [7].

Proposition 2.3 *A language of A^* is recognized by a p-group if and only if it is a Boolean combination of the languages*

$$S(u, r, p) = \{v \in A^* \mid \binom{v}{u} \equiv r \bmod p\},$$

for $0 \leqslant r < p$ and $u \in A^*$.

Another characterization, given in [17, 18], relies on a variation of the concatenation product, called the *modular concatenation product* and first introduced in [14]. Let $L_0, \ldots, L_k$ be languages of A^*, let $a_1, \ldots, a_k$ be letters of A and let r and p be integers such that $0 \leqslant r < p$. We define $(L_0 a_1 L_1 \cdots a_k L_k)_{r,p}$ as the set of all words u in A^* such that the number of factorizations of u in the form $u = u_0 a_1 u_1 \cdots a_k u_k$, with $u_i \in L_i$ for $0 \leqslant i \leqslant k$, is congruent to r modulo p.

Proposition 2.4 *A language of A^* is recognized by a p-group if and only if it is a Boolean combination of languages of the form $(A^* a_1 A^* \cdots a_k A^*)_{r,p}$, where $0 \leqslant r < p$, $k \geqslant 0$ and $a_1, \ldots, a_k \in A$.*

Contrary to the concatenation product, the modular concatenation product does not distribute over union. For instance, if $A = \{a, b\}$,

$$(\{b\}a\{1, ba\})_{1,2} = \{ba, baba\}, \quad (\{bab\}a\{1, ba\})_{1,2} = \{baba, bababa\}$$
$$\text{but } (\{b, bab\}a\{1, ba\})_{1,2} = \{ba, bababa\}$$

since $baba = (b)a(ba) = (bab)a(1)$. However, a weaker property holds.

Proposition 2.5 *Let $L_0, \ldots, L_k$ be languages of A^* and let $i \in \{0, \ldots, k\}$. Suppose that L_i is the disjoint union of the languages $L_{i,1}, \ldots, L_{i,\ell}$. Then each modular product $(L_0a_1L_1 \cdots a_kL_k)_{r,p}$ is a union of intersections of languages of the form $(L_0a_1L_1 \cdots L_{i-1}a_iL_{i,j}a_{i+1}L_{i+1} \cdots a_kL_k)_{s,p}$, with $1 \leqslant j \leqslant \ell$ and $0 \leqslant s < p$.*

Proof. We claim that $(L_0a_1L_1 \cdots a_kL_k)_{r,p}$ is equal to

$$\bigcup_{\substack{r_1+\ldots+r_\ell \equiv r \bmod p \\ 0 \leqslant r_1,\ldots,r_\ell < p}} \ \bigcap_{1 \leqslant j \leqslant l} (L_0a_1L_1 \cdots L_{i-1}a_iL_{i,j}a_{i+1}L_{i+1} \cdots a_kL_k)_{r_j,p}$$

For a given word u, consider the set $F(u)$ of all k-uples $(u_0, u_1, \ldots, u_k)$ such that $u = u_0a_1u_1 \cdots a_ku_k$, with $u_0 \in L_0, \ldots, u_k \in L_k$. The set $F(u)$ is the disjoint union of the sets $F_j(u)$ defined by

$$F_j(u) = \{(u_0, u_1, \ldots, u_k) \in F(u) \mid u_i \in L_{i,j}\}$$

It follows that $|F(u)| = \sum_{1 \leqslant j \leqslant \ell} |F_j(u)|$ and hence $|F(u)| \equiv r \mod p$ if and only if there exist $r_1, \ldots, r_\ell$ such that $r_1 + \ldots + r_\ell \equiv r$ and $|F_1(u)| \equiv r_1 \mod p$, $\ldots$, $|F_\ell(u)| \equiv r_\ell \mod p$. This proves the claim and the proposition.
Coming back to the previous example, one has

$$(\{b\}a\{1, ba\})_{0,2} = A^* \setminus \{ba, baba\}, \quad (\{bab\}a\{1, ba\})_{0,2} = A^* \setminus \{baba, bababa\}$$

Therefore

$$(\{b\}a\{1, ba\})_{0,2} \cap (\{bab\}a\{1, ba\})_{1,2} = \{bababa\}$$
$$(\{b\}a\{1, ba\})_{1,2} \cap (\{bab\}a\{1, ba\})_{0,2} = \{ba\}$$

and the union of these two languages is exactly $(\{b, bab\}a\{1, ba\})_{1,2}$.

2.3 Languages recognized by supersoluble groups

The aim of this section is to prove our main result, which describes the languages of $\mathcal{U}_p$.

Theorem 2.6 *Let L be a language of A^*. The following conditions are equivalent:*

(1) *L is recognized by a group in $\mathbf{G}_p * \mathbf{Ab}(p-1)$,*

(2) *L is a Boolean combination of languages of the form $(L_0a_1L_1 \cdots a_kL_k)_{r,p}$, where each L_i is a $(p-1)$-elementary commutative language,*

(3) *L is a Boolean combination of languages of the form $(L_0a_1L_1 \cdots a_kL_k)_{r,p}$, where each L_i is a Boolean combination of $(p-1)$-elementary commutative languages.*

We can also formulate our result in terms of varieties of languages.

Corollary 2.7 *For every alphabet A, $\mathcal{U}_p(A^*)$ is the Boolean algebra generated by the languages of the form $(L_0a_1L_1\cdots a_kL_k)_{r,p}$, where each L_i is a $(p-1)$-elementary commutative language of A^*.*

Corollary 2.8 *For every alphabet A, $\mathcal{U}(A^*)$ is the Boolean algebra generated by the languages of the form $(L_0a_1L_1\cdots a_kL_k)_{r,p}$, where each L_i is a $(p-1)$-elementary commutative language of A^*, for any prime p.*

We shall give three different proofs. The first one relies on the wreath product principle, the second one on the representation theory of finite semigroups and the third makes use of matrix representations of transducers.

2.3.1 Proof using the wreath product principle

We shall need two auxiliary tools to characterize the languages of $\mathcal{U}_p$. The first one is an operation on groups introduced in [17, 18] to study the modular concatenation product.

Let $G_1,\ldots,G_r$ be groups. Denote by $K = \mathbb{F}_p[G_1\times\cdots\times G_r]$ the group algebra of $G_1\times\cdots\times G_r$ over $\mathbb{F}_p$. The *Schützenberger product over $\mathbb{F}_p$* of the groups $G_1,\ldots,G_r$, denoted by $\mathbb{F}_p\diamondsuit(G_1,\ldots,G_r)$, is the subgroup of $GL_r(K)$ made up of matrices $m=(m_{i,j})$ such that

(1) $m_{i,j}=0$, for $i>j$,

(2) $m_{i,i}=(1,\ldots,1,g_i,1,\ldots,1)$ for some $g_i\in G_i$,

(3) $m_{i,j}\in\mathbb{F}_p[1\times\cdots\times 1\times G_i\times\cdots\times G_j\times 1\times\cdots\times 1]$, for $i<j$.

The following result was first proved in [17, 18].

Proposition 2.9 *Let, for $0\leqslant i\leqslant k$, L_i be a language of A^* recognized by a group G_i. Then the language $(L_0a_1L_1\cdots a_kL_k)_{r,p}$ is recognized by the group $\mathbb{F}_p\diamondsuit(G_0,\ldots,G_k)$.*

Our second tool, the sequential transducer of a morphism, is required to characterize the languages recognized by the wreath product of two monoids.

Let G be a group and let $\varphi\colon A^*\to G$ be a monoid morphism. Set $B_G = G\times A$. The *sequential function associated with φ* is the function $\sigma_\varphi\colon A^*\to B_G^*$ defined by

$$\sigma_\varphi(a_1a_2\cdots a_n) = (1,a_1)(\varphi(a_1),a_2)\cdots(\varphi(a_1\cdots a_{n-1}),a_n)$$

Straubing's wreath product principle [14, 15, 13] leads immediately to the following result.

Proposition 2.10 *For every alphabet A, $\mathcal{U}_p(A^*)$ is the smallest Boolean algebra containing $\mathcal{A}b(p-1)(A^*)$ and the languages of the form $\sigma_\varphi^{-1}(V)$, where σ_φ is the sequential function associated with a morphism $\varphi\colon A^*\to G$, with $G\in\mathbf{Ab}(p-1)$, and V is a language of B_G^* recognized by a p-group.*

We are now ready to prove our main theorem.

Proof. (2) implies (3) is trivial.

(3) implies (1). By Proposition 2.1, any Boolean combination of $(p-1)$-elementary commutative languages is recognized by a group in $\mathbf{Ab}(p-1)$. Further, Proposition 2.9 shows that, if each language L_i is recognized by a group G_i, then the language $(L_0a_1L_1\cdots a_kL_k)_{r,p}$ is recognized by the group $G = \mathbb{F}_p\diamondsuit(G_0,\ldots,G_k)$. Consequently, it just remains to show that if the groups G_i are all in $\mathbf{Ab}(p-1)$, then G is an element of $\mathbf{G}_p*\mathbf{Ab}(p-1)$. Let $\pi : G \to G_0\times\cdots\times G_k$ be the surjective morphism which maps each matrix onto the product of its diagonal elements. Thus if $m \in G$, $\pi(m) = m_{0,0}\cdots m_{k,k}$. We claim that $\mathrm{Ker}(\pi)$ is a p-group. Indeed, if m belongs to $\mathrm{Ker}(\pi)$, then $m_{i,j} = 0$ if $i > j$, $m_{i,i} = (1,\ldots,1)$ for $i = 0,\ldots,k$ and $m_{i,j} \in \mathbb{F}_p[1\times\cdots\times1\times G_i\times\cdots\times G_j\times1\times\cdots\times1]$, for $i < j$. Notice that, for $i < j$, the (i,j)-th entry of m can be written as $\sum_{h\in G_i\times\cdots\times G_j}\alpha_h h$ for some $\alpha_h \in \mathbb{F}_p$. Since there are exactly $p^{|G_i|\cdots|G_j|}$ elements of this form, the order of $\mathrm{Ker}(\pi)$ is a power of p (more precisely, $\prod_{i<j}p^{|G_i|\cdots|G_j|}$) and $\mathrm{Ker}(\pi)$ is a p-group. Therefore, $G \in \mathbf{G}_p * \mathbf{Ab}(p-1)$.

(1) implies (2). With the notation of Proposition 2.10, it suffices to show that the languages of $\mathcal{A}b(p-1)(A^*)$ and the languages $\sigma_\varphi^{-1}(V)$ are of the form described in (2). For the languages of $\mathcal{A}b(p-1)(A^*)$, this follows directly from Proposition 2.1. Consider now a language $\sigma_\varphi^{-1}(V)$, where σ_φ is the sequential function associated with a morphism $\varphi\colon A^* \to G$, with $G \in \mathbf{Ab}(p-1)$, and V is a language of B_G^* recognized by a p-group. Since σ_φ^{-1} commutes with Boolean operations, we may assume by Proposition 2.3, that $V = S(u,r,p)$ with $0 \leqslant r < p$ and $u \in B_G^*$. Since $B_G = G \times A$, u is a word of the form $(g_1,c_1)\cdots(g_k,c_k)$, where $g_1,\ldots,g_k \in G$ and $c_1,\ldots,c_k \in A$. Thus V is the set of words $v \in B_G^*$ such that

$$\mathrm{Card}\,\{(v_0,v_1,\ldots,v_k) \mid v_0(g_1,c_1)v_1\cdots v_{k-1}(g_k,c_k)v_k = v\} \equiv r \bmod p$$

Let us now compute $\sigma_\varphi^{-1}(V)$. If $u = a_1\cdots a_n$, then

$$\sigma_\varphi(a_1\cdots a_n) = (1,a_1)(\varphi(a_1),a_2)\cdots(\varphi(a_1\cdots a_{n-1}),a_n)$$

Therefore u belongs to $\sigma_\varphi^{-1}(V)$ if and only if it belongs to

$$\left(\varphi^{-1}(h_1)c_1\varphi^{-1}(h_2)c_2\ \cdots\varphi^{-1}(h_k)c_kA^*\right)_{r,p}$$

where $h_1 = g_1$, $h_2 = (g_1\varphi(c_1))^{-1}g_2$, $\ldots$, $h_k = (g_{k-1}\varphi(c_{k-1}))^{-1}g_k$. Since G is in $\mathbf{Ab}(p-1)$, the languages $\varphi^{-1}(h_1),\ldots\varphi^{-1}(h_k)$ are, by Proposition 2.2, a disjoint union of $(p-1)$-elementary commutative languages. To conclude the proof, it remains to use Proposition 2.5 to "distribute" the modular product $(L_0a_1L_1\cdots a_kL_k)_{r,p}$ over this disjoint union.

2.3.2 Proof using representation theory

The second proof relies on a result [1] describing the variety of languages corresponding to a Mal'cev product of the form $\mathbf{LG}_p \circledM \mathbf{V}$. We actually don't need the full version of this theorem and we shall only state it when $\mathbf{V}$ is a variety of finite groups. In this case, $\mathbf{LG}_p \circledM \mathbf{V} = \mathbf{G}_p * \mathbf{V}$ and the result can be formulated as follows.

Theorem 2.11 (See [1, Corollary 6.3]) *Let $\mathbf{V}$ be a variety of finite groups and $\mathbf{W} = \mathbf{G}_p * \mathbf{V}$. Let $\mathcal{V}$ and $\mathcal{W}$ be the varieties of languages corresponding* *to* $\mathbf{V}$ *and* $\mathbf{W}$*, respectively. Then for each alphabet* A*,* $\mathcal{W}(A^*)$ *is the Boolean algebra generated by the languages of the form* $(L_0a_1L_1\cdots a_kL_k)_{r,p}$*, where each* L_i *belongs to* $\mathcal{V}(A^*)$*.*

When $\mathbf{V} = \mathbf{Ab}(p-1)$, we get immediately:

Corollary 2.12 *For every alphabet* A*,* $\mathcal{U}_p(A^*)$ *is the Boolean algebra generated by the languages of the form* $(L_0a_1L_1\cdots a_kL_k)_{r,p}$ *where each* L_i *is a* $(p-1)$*-commutative language.*

In order to obtain the stronger version stated in Theorem 2.6 and Corollary 2.7, one needs again to use Proposition 2.5 as we did in the first proof.

2.3.3 Proof using matrix representation of transducers

A general method to study operations on regular languages using transducers was given by Pin and Sakarovitch [11, 12]. This method can be used, for instance, to show that the marked product of n languages is recognized by the Schützenberger product of these languages. It can easily be adapted (see next Section) to prove that a language of the form $(L_0a_1L_1\cdots a_kL_k)_{r,p}$, where each L_i is a $(p-1)$-elementary commutative language of A^*, is recognized by an upper triangular group of matrices over $\mathbb{F}_p$.

A result of the second author [8] shows that if a language L is recognized by the Schützenberger product of the monoids $M_0, \ldots, M_n$, then L belongs to the Boolean closure of the set of languages of the form $L_{i_0}a_1L_{i_1}\cdots a_rL_{i_r}$ ($0 \leqslant i_0 < i_1 < \ldots < i_r \leqslant n$) where the a_k are letters and the L_{i_k} are recognized by M_{i_k} ($0 \leqslant k \leqslant r$). This result has been extended [9, 18] to upper triangular matrices over $\mathbb{Z}/n\mathbb{Z}$, with the products of the form $(L_0a_1L_1\cdots a_kL_k)_{r,p}$ in place of the marked product. It is not difficult to adapt these arguments to show that a language is recognized by an upper triangular group of matrices over $\mathbb{F}_p$, then it is a Boolean combination of languages of the form $(L_0a_1L_1\cdots a_kL_k)_{r,p}$, where each L_i is a $(p-1)$-elementary commutative language. Then it remains to use the fact that the variety of supersoluble groups is generated by upper triangular group of matrices over $\mathbb{F}_p$, for some p.

2.4 Transducers and languages recognized by supersoluble groups

We now give another description of the variety of languages associated with the variety of supersoluble groups.

A *transducer* with output in $\mathbb{F}_p$ is a 5-tuple $\mathcal{T} = (Q, A, I, F, E)$ where Q is a finite *set of states*, A is the *input alphabet*, $I \subseteq Q$ is the set of *initial* and $F \subseteq Q$ the set of *final* states. The set of transitions E is a finite subset of $Q \times A \times \mathbb{F}_p^* \times Q$. Intuitively, a transition (p, a, r, q) is interpreted as follows: if a is an input letter, the automaton moves from state p to state q and produces the output r.

It is convenient to represent a transition (p, a, r, q) as an edge $p \xrightarrow{a|r} q$. Initial (resp. final) outputs are represented by incoming (resp. outgoing) arrows. A successful path is a sequence of consecutive transitions:

$$q_0 \xrightarrow{a_1|r_1} q_1 \xrightarrow{a_2|r_2} q_2 \quad \cdots \quad q_{n-1} \xrightarrow{a_n|r_n} q_n$$

starting in some initial state and ending in some final state. The *label* of the path is the word $a_1a_2\cdots a_n$. Its *output* is the product $r_1r_2\cdots r_n$. The *function realized by* $\mathcal{T}$ maps each word u of A^* onto the sum of the outputs of all successful paths of label u.

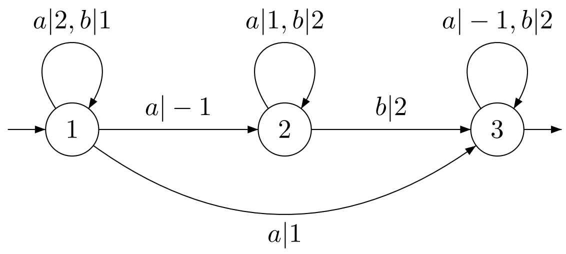

Figure 2.1: A transducer with output in $\mathbb{F}_5$.

For instance, if τ is the transduction realised by the transducer of Figure 2.1, there are five successful paths of input label *abab*.

(1) $1 \xrightarrow{a|2} 1 \xrightarrow{b|1} 1 \xrightarrow{a|-1} 2 \xrightarrow{b|2} 3$ (2) $1 \xrightarrow{a|2} 1 \xrightarrow{b|1} 1 \xrightarrow{a|1} 3 \xrightarrow{b|2} 3$

(3) $1 \xrightarrow{a|-1} 2 \xrightarrow{b|2} 2 \xrightarrow{a|1} 2 \xrightarrow{b|2} 3$ (4) $1 \xrightarrow{a|-1} 2 \xrightarrow{b|2} 3 \xrightarrow{a|-1} 3 \xrightarrow{b|2} 3$

(5) $1 \xrightarrow{a|1} 3 \xrightarrow{b|2} 3 \xrightarrow{a|-1} 3 \xrightarrow{b|2} 3$

The output of the first path is $2 \times 1 \times (-1) \times 2 = 1 \bmod 5$, the output of the other paths are respectively -1, 1, -1 and 1. It follows that $\tau(abab) = 1 - 1 + 1 - 1 + 1 = 1$.

A transducer is in *strict triangular form* if $Q = \{1, \ldots, n\}$, 1 is the unique initial state, n is the unique final state, and its transitions satisfy the three following conditions:

(1) there is no transition from p to q such that $p > q$,
(2) for $p < q$ and for each letter $a \in A$, there is at most one transition from p to q with label a,
(3) for each letter $a \in A$ and every state $q \in Q$ there is exactly one transition of the form $q \xrightarrow{a|r} q$, for some $r \in \mathbb{F}_p^*$.

For instance, the transducer in Figure 2.1 is in strict triangular form. To each such transducer is associated a morphism $\mu : A^* \to B_n(\mathbb{F}_p)$, called its *linear representation*, and defined as follows. For each letter $a \in A$,

$$\mu(a)_{p,q} = \begin{cases} 0 & \text{if there is no transition of label } a \text{ from } p \text{ to } q \\ r & \text{if } p \xrightarrow{a|r} q \text{ is the unique transition of label } a \text{ from } p \text{ to } q \end{cases}$$

On our example, we obtain

$$\mu(a) = \begin{pmatrix} 2 & -1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad \mu(b) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix} \quad \mu(abab) = \begin{pmatrix} -1 & 2 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

The linear presentation gives an easy way to compute the function realised by the transducer, since $\tau(u) = \mu(u)_{1,n}$ (see [4] for details). For instance, on our example, $\mu(abab)_{1,3} = 1$.

We can now state our last characterisation of the variety of languages $\mathcal{U}_p$.

Theorem 2.13 *A language belongs to $\mathcal{U}_p(A^*)$ if and only if it is a Boolean combination of languages of the form $\tau^{-1}(r)$, where $r \in \mathbb{F}_p$ and $\tau : A^* \to \mathbb{F}_p$ is a function realised by some transducer in strict triangular form.*

Proof. Let $\mathcal{B}$ be the Boolean algebra described in the statement of the theorem. We want to show that $\mathcal{B} = \mathcal{U}_p(A^*)$.

Consider a function $\tau : A^* \to \mathbb{F}_p$ realised by a transducer in strict triangular form and let $\mu : A^* \to B_n(\mathbb{F}_p)$ be its linear representation. Let $r \in \mathbb{F}_p$. We claim that the language $\tau^{-1}(r)$ is recognized by $B_n(\mathbb{F}_p)$. Indeed, since $\tau^{-1}(r) = \{u \in A^* \mid \mu(u)_{1,n} = r\}$, one has $\tau^{-1}(r) = \mu^{-1}(R)$ where R is the set of all matrices m of $B_n(\mathbb{F}_p)$ such that $m_{1,n} = r$. This proves the claim and shows that the languages of the form $\tau^{-1}(r)$ are in $\mathcal{U}_p(A^*)$. The inclusion $\mathcal{B} \subseteq \mathcal{U}_p(A^*)$ follows, since both $\mathcal{B}$ and $\mathcal{U}_p(A^*)$ are Boolean algebras.

Conversely, since by Theorem 1.2, the variety $\mathbf{G}_p * \mathbf{Ab}(p-1)$ is generated by the groups $B_n(\mathbb{F}_p)$, the Boolean algebra $\mathcal{U}_p(A^*)$ is generated by the languages recognized by $B_n(\mathbb{F}_p)$, for some $n > 0$. Consider a language L of A^* recognized by $B_n(\mathbb{F}_p)$. By definition, there exists a morphism $\eta : A^* \to B_n(\mathbb{F}_p)$ and a subset P of $B_n(\mathbb{F}_p)$ such that $L = \eta^{-1}(P)$. We claim that L belongs to $\mathcal{B}$. Since $\eta^{-1}(P) = \bigcup_{m \in P} \eta^{-1}(m)$, it suffices to establish the result when P contains a single matrix m. Observe that

$$\eta^{-1}(m) = \bigcap_{1 \leqslant i,j \leqslant n} L_{i,j} \quad \text{where} \quad L_{i,j} = \{u \in A^* \mid \eta(u)_{i,j} = m_{i,j}\}$$

Put $t = j - i + 1$ and let $\mu : A^* \to B_t(\mathbb{F}_p)$ be the morphism defined, for all $a \in A$, by

$$\mu(a)_{k,\ell} = \eta(a)_{i+k-1,i+\ell-1} \qquad \text{for } 1 \leqslant k, \ell \leqslant t$$

Thus $\mu(a)$ is the submatrix of $\eta(a)$ whose right top element is $\eta(u)_{i,j}$ and bottom left element is $\eta(u)_{j,i}$. It follows that, for all $u \in A^*$, $\mu(u)_{1,t} = \eta(u)_{i,j}$. Setting $m_{i,j} = r$, one sees that $u \in L_{i,j}$ if and only if $\mu(u)_{1,t} = r$. Therefore L is of the form $\tau^{-1}(r)$, where τ is the function realised by the transducer in strict triangular form defined by μ.

Corollary 2.14 *A language belongs to $\mathcal{U}(A^*)$ if and only if it is a Boolean combination of languages of the form $\tau^{-1}(r)$, where $r \in \mathbb{F}_p$, p is a prime number and $\tau : A^* \to \mathbb{F}_p$ is a function realised by some transducer in strict triangular form.*

Acknowledgements

The authors would like to thank Adolfo Ballester-Bolinches and the referees of the first version of this article for various suggestions.

References

[1] J. Almeida, S. W. Margolis, B. Steinberg and M. V. Volkov, Representation Theory of Finite Semigroups, Semigroup Radicals and Formal Language Theory, *ArXiv Mathematics E-Prints*, Fév. 2007.

[2] J. L. Alperin and R. B. Bell, *Groups and representations*, *Graduate Texts in Mathematics* vol. 162, Springer-Verlag, New York, 1995.

[3] K. Auinger and B. Steinberg, Varieties of finite supersolvable groups with the M. Hall property, *Math. Ann.* **335**,4 (2006), 853–877.

[4] J. Berstel, *Transductions and Context-Free Languages*, Teubner, 1979.

[5] R. Carter and T. Hawkes, The $\mathcal{F}$-normalizers of a finite soluble group, *J. Algebra* **5** (1967), 175–202.

[6] K. Doerk and T. Hawkes, *Finite soluble groups*, *de Gruyter Expositions in Mathematics* vol. 4, Walter de Gruyter & Co., Berlin, 1992.

[7] S. Eilenberg, *Automata, languages, and machines. Vol. B*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976. Pure and Applied Mathematics, Vol. 59.

[8] J.-E. Pin, Hiérarchies de concaténation, *RAIRO Informatique Théorique* **18** (1984), 23–46.

[9] J.-E. Pin, Topologies for the free monoid, *J. of Algebra* **137** (1991), 297–337.

[10] J.-É. Pin, Syntactic semigroups, in *Handbook of formal languages*, G. Rozenberg and A. Salomaa (éd.), vol. 1, ch. 10, pp. 679–746, Springer Verlag, 1997.

[11] J.-E. Pin and J. Sakarovitch, Some operations and transductions that preserve rationality, in *6th GI Conference*, Berlin, 1983, pp. 277–288, *Lect. Notes Comp. Sci.* n° 145, Springer.

[12] J.-E. Pin and J. Sakarovitch, Une application de la représentation matricielle des transductions, *Theoret. Comput. Sci.* **35** (1985), 271–293.

[13] J.-E. Pin and P. Weil, The wreath product principle for ordered semigroups, *Communications in Algebra* **30** (2002), 5677–5713.

[14] H. Straubing, Families of recognizable sets corresponding to certain varieties of finite monoids, *J. Pure Appl. Algebra* **15**,3 (1979), 305–318.

[15] H. Straubing, The wreath product and its applications, in *Formal properties of finite automata and applications (Ramatuelle, 1988)*, pp. 15–24, *Lecture Notes in Comput. Sci.* vol. 386, Springer, Berlin, 1989.

[16] D. Thérien, Subword counting and nilpotent groups, in *Combinatorics on words (Waterloo, Ont., 1982)*, pp. 297–305, Academic Press, Toronto, ON, 1983.

[17] P. Weil, An extension of the Schützenberger product, in *Lattices, semigroups, and universal algebra (Lisbon, 1988)*, pp. 315–321, Plenum, New York, 1990.

[18] P. Weil, Products of languages with counter, *Theoret. Comput. Sci.* **76** (1990), 251–260.