Canad. Math. Bull. Vol. 17 (2), 1974.

ON FREE SEMIGROUPS AND RAMSEY NUMBERS

BY
GERARD LALLEMENT

ABSTRACT. If the length of a word w in a free semigroup $F(X)$ satisfies $l(w) \geq pn^k$, then for every partition of $F(X)$ into k classes, w has n consecutive factors of length $\geq p$ in the same class. As a consequence, the diagonal Ramsey numbers $R(pn+1, p+1, k)$ have $1+pn^k$ as lower bound.

1. The free semigroup $F(X)$ on the alphabet X is the set of all non-empty words in the letters of X with the usual concatenation operation. The length of a word $w \in F(X)$, denoted by $l(w)$, is the total number of occurrences of letters of X in w. A factor (resp. left, right factor) of a word w is a word w' such that $w=uw'v$ (resp. $w=w'v$, $w=uw'$) for some $u, v \in F(X)$. Congruences on $F(X)$ of finite index are of interest in language theory, especially in the study of recognizable subsets of $F(X)$ (also called regular events, see e.g. [5] Theorem 2.1.5). Herein, we are concerned with partitions of $F(X)$ into a finite number of classes and we prove the following results.

THEOREM. *Let $A_1, A_2, \ldots, A_k$ be a partition of the free semigroup $F(X)$ into k classes.*

(a) *For every integer n, there exists a smallest integer $r_k(n, p)$ such that every word $w \in F(X)$ of length $l(w) \geq r_k(n, p)$ has n consecutive factors of length $\geq p$ in a single class A_i of the partition.*

(b) $r_k(n, p) = pn^k$.

For $p=1$, this theorem is established in [11], and in sections 2 and 3 we present an adaptation of the proof in [11] to the case of an arbitrary p. The connection with a theorem of Van der Waerden are explained in [9].

Part (a) of the Theorem is a direct consequence of a theorem of Ramsey: Given a set E and a partition θ of the set $P_r(E)$ of all r-subsets of E into k classes, then for every integer q there is a smallest integer $R(q, r, k)$ such that card $E \geq R(q, r, k)$ implies that there is a q-subset F of E such that $P_r(F)$ is contained in a single class mod θ. In the notation of [10], p. 39, we have

$$R(q, r, k) = N(q_1, q_2, \ldots, q_k, r) \quad \text{with } q_1 = q_2 = \cdots = q_k = q$$

and the numbers $R(q, r, k)$ appear as the "diagonal" Ramsey numbers. The theorem above has the following

COROLLARY. $1+pn^k \leq R(pn+1, p+1, k)$.

REMARKS. For the case $p=1$, the inequality $1+n^k \leq R(n+1, 2, k)$ appears in [4] (Proposition 3.5.3 and Remarque 4.2a) as a consequence of a Ramsey type theorem on partitions of $P_2(E)$ respecting a linear ordering of E. It has been improved to

$$\frac{(2n-1)^k+3}{2} \leq R(n+1, 2, k)$$

(see [6], [8]). For $R(n+1, 2, 2)$ the lower bound $1+n^2$ is better than $(\sqrt{2})^{n+1}$, [3], up to $n=15$, while other methods (see e.g. [1], [7]) produce better lower bounds in particular cases. For example

$$R(pn+1, p+1, k) \geq ((pn+1)!)^{\frac{1}{pn+1}} k^{\frac{\binom{pn+1}{p+1}-1}{pn+1}}$$

proved in [1] (Corollary 2B), gives a better lower bound only in case k is small with respect to n and p.

2. **Proof of part (a) and the Corollary.** Let $w=x_1x_2 \ldots x_t \in F(X)$. To any sequence of $p+1$ integers $i_0, i_1, \ldots, i_p$ such that $0 \leq i_0 < i_1 < \cdots i_p \leq t$ we associate the word

$$b(i_0, i_1, \ldots, i_p) = |x_{i_0+1} \cdots x_{i_1} \mid x_{i_1+1} \cdots x_{i_2}| \cdots |x_{i_{p-1}+1}, \ldots, x_{i_p}|$$

in $F(X \cup \{ \mid \})$ and we call $b(i_0, i_1, \ldots, i_p)$ a p-block of w. Letting $E=\{0, 1, 2, \ldots, l(w)\}$, there is a 1–1 correspondence between the set of all $(p+1)$-subsets of E and the set of all p-blocks of w. The partition $\theta=\{A_1, A_2, \ldots, A_k\}$ of $F(X)$ induces a partition π of the set of all p-blocks of w defined as follows

$$b(i_0, i_1, \ldots, i_p) \, \pi \, b(j_0, j_1, \ldots, j_p) \Leftrightarrow (x_{i_0+1} \cdots x_{i_p}) \, \theta \, (x_{j_0+1} \cdots x_{j_p})$$

In turn, π defines a partition (also denoted π) of the set of all $(p+1)$-subsets of E which has at most k classes. By Ramsey's theorem if card $E=1+l(w) \geq R(pn+1, p+1, k)$ there is a $(pn+1)$-subset F of E having all its $(p+1)$-subsets in a single class of π. Let $F=\{l_0, l_1, \ldots, l_{pn}\}$ with $l_1 < l_{i+1}$. The p-blocks of w

$$b(l_{kp}, l_{kp+1}, \ldots, l_{(k+1)p}) \qquad 0 \leq k \leq (n-1)$$

are all in the same class of π. It follows that the n consecutive factors of w of the type

$$x_{l_{kp}} \cdots x_{l_{kp+1}} \cdots x_{l_{(k+1)p}} \qquad 0 \leq k \leq (n-1)$$

are all of length $\geq p$ and contained in the same class of θ. This proves part (a) of the theorem and also $r_k(n, p) \leq R(pn+1, p+1, k)-1$.

3. **Proof of part (b).** To every word $w \in F(X)$ we associate a k-tuple of integers $(a_i(w))$, $i=1, 2, \ldots, k$ where $a_i(w)$ is the largest integer m such that a right factor of w consists of m consecutive factors of length $\geq p$ in A_i. Let $\mu: F(X) \to N^k$ be the mapping $\mu(w)=(a_i(w))$.

Suppose that $w=uv$ with $l(v)\geq p$ and that $\mu(w)=\mu(u)$. Then for every $i=1, 2, \ldots, k$ we have $w=u'_i u''_i v$ where u'' is a product of $a(u)=a(w)$ words of length $\geq p$ all contained in A_i. In particular if $v\in A_{i_0}$, then $w=u'_{i_0}u''_{i_0}v$ and w has a right factor $u''_{i_0}v$ having $a_{i_0}(w)+1$ consecutive factors of length $\geq p$ all contained in A_{i_0}. This contradicts the definition of $a_{i_0}(w)$. Therefore $\mu(w)\neq\mu(u)$. Considering two factorizations of w, say $w=u_1v_1$, $w=u_2v_2$ with $l(u_i)=k_ip$, $l(v_i)\geq p$ for $i=1, 2$ and $k_1>k_2$ we have, by Theorem 9.6 [2], $u_1=u_2v$ with $l(v)\geq p$. The same argument as above shows that $\mu(u_1)\neq\mu(u_2)$.Thus all the factorizations $w=uv$ with $l(u)=kp$ $(k\geq 1)$ and $l(v)\geq p$ give rise to words u that are mapped onto distinct points of N^k by μ. If $l(w)\geq pn^k$, $\mu(w)$ and the $\mu(u)$'s from the various factorizations of w constitute a set of n^k distinct points in N^k. Since there are only n^k points in N^k having all their coordinates $<n$ and since $\mu(w)$ or $\mu(u)\neq(0, 0, \ldots, 0)$ for any left factor u of w, it follows that w has n consecutive factors of length $\geq p$ contained in an A_i for some $i=1, 2, \ldots, k$. Therefore $r_k(n, p)\leq pn^k$.

To show that $r_k(n,p)\not< pn^k$ we construct counterexamples by induction on k. Define in $F(x_1, x_2, \ldots, x_p)$

$$w_1(n, p) = (x_1x_2\cdots x_p)^{n-1}x_1x_2\cdots x_{p-1}.$$

For $k>1$ define $w_k(n,p)$ in $F(x_1, x_2, \ldots, x_{p+k-1})$ by

$$w_k(n, p) = [w_{k-1}(n, p)x_{p+k-1}]^{n-1}w_{k-1}(n, p)$$

By induction on k one checks that $l(w_k(n,p))=pn^k-1$. On $F(x_1, x_2, \ldots, x_{p+k-1})$ we define the following partition into k classes

$$A_1 = F(x_1, x_2, \ldots, x_p)$$

$$A_i = F(x_1, x_2, \ldots, x_{p+i-1})-F(x_1, x_2, \ldots, x_{p+i-2}) \quad \text{for } 1 < i \leq k$$

Then, by induction on k, one shows easily that $w_k(n,p)$ has at most $n-1$ consecutive factors of length $\geq p$ in a single A_i. This completes the proof of $r_k(n,p)=pn^k$.

Acknowledgment. The author wishes to express his thanks to the referee for providing most of the references contained in the remarks at the end of Section 1.

References

1. V. Chvátal, *Hypergraphs and Ramseyian theorems,* Proc. Amer. Math. Soc. **27,** (1971), 434–440.

2. A. H. Clifford and G. B. Preston, *The algebraic theory of semigroups,* Vol. II, 1967, Amer. Math. Soc., Providence, R.I.

3. P. Erdös, *Some remarks on the theory of graphs,* Bull. Amer. Math. Soc. **53,** 1947, 292–294.

4. C. Frasnay, *Quelques problèmes combinatoires concernant les ordres totaux et les relations monomorphes,* Ann. Inst. Fourier, Grenoble, **15,** 1965, 415–524.

5. S. Ginsburg, *The mathematical theory of context-free languages,* 1966, McGraw-Hill.

6. G. R. Giraud, *Une généralisation des nomblres et de l'inégalité de Schur,* C.R. Acad. Sc. Paris, t. 266 (1968), 437–440.

7. R. E. Greenwood and A. M. Gleason, *Combinatorial relations and chromatic graphs*, Can. Journ. Math., **7**, 1955, 1–7.

8. P. Hell, *Ramsey Numbers*, M.Sc. thesis, McMaster University, 1969.

9. J. Justin, *Généralisation du théorème de Van der Waerden sur les semi-groupes répétitifs*, J. Combinatorial Theory (A) **12**, (1972), 357–367.

10. H. J. Ryser, *Combinatorial mathematics*, 1963, Carus Math. Monographs number 14.

11. M. P. Schützenberger, *Quelques problèmes combinatoires de la théorie des automates*, Cours de l'Institut de Programmation 1966–1967 rédigé par J. F. Perrot.

Pennsylvania State University,
Pennsylvania 16802