Discrete Mathematics 40 (1982) 225–234
North-Holland Publishing Company

ON THE NUMBER OF WORDS IN THE LANGUAGE $\{w \in \Sigma^* \mid w = w^R\}^2$

R. KEMP

Fachbereich 20 (Informatik), Johann Wolfgang Goethe Universität, 6000 Frankfurt a.M., Federal Republic of Germany

Received 29 September 1980

Let S be the set of all palindromes over Σ^*. It is well known, that the language S^2 is an ultralinear, inherently ambiguous context-free language. In this paper we derive an explicit expression for the number of words of length n in S^2. Furthermore, we show, that for $\text{card}(\Sigma) > 1$ the asymptotical density of the language S^2 is zero and that, in the average, each word w of length n in S^2 has exactly one factorization into two palindromes for large n; the variance is zero for large n. Finally, we compute an expression for the structure-generating-function $T(S^2; z)$ of the language S^2; it remains the open problem, if $T(S^2; z)$ is a transcendental or an algebraic function.

1. Introduction

Let Σ^* be the free monoid generated by a fixed alphabet Σ and let $S \subseteq \Sigma^*$ be the formal language defined by $S := \{w \in \Sigma^* \mid w = w^R\}$, where w^R is the reversal of the word w. Thus, S is the set of all palindromes over Σ^*. In [4, 5], Crestin introduced the language $S^2 = SS$ and he showed that S^2 is an ultralinear, inherently ambiguous context-free language if $\text{card}(\Sigma) > 1$.

In [2], Berstel introduced the notion of the asymptotical density $d(L)$ of a formal language $L \subseteq \Sigma^*$. If $d(L)$ exists, this number is defined by $d(L) = \lim_{n\to\infty}(d_n(L)/d_n(\Sigma^*))$, where $d_n(F) = \text{card}\{w \in \Sigma^* \mid w \in F \setminus \Sigma^*\Sigma^{n+1}\}$ for $F \subseteq \Sigma^*$. Berstel showed in his paper that $d(L)$ is rational if L is regular, and that $d(L)$ is algebraic if L is an unambiguous context-free language. Using the notion of the structure-generating-function $T(L; z)$ of a formal language $L \subseteq \Sigma^*$ (see [7]) which is defined by $T(L; z) = \sum_{n\geq 0} \text{card}(L \cap \Sigma^n) z^n$, Berstel has implicitly shown in his paper that the fact '$T(L; z)$ is an algebraic (resp. a rational) function' implies '$d(L)$ is an algebraic (resp. a rational) number', provided that $d(L)$ exists. It was an open problem, if there are inherently ambiguous context-free languages with a transcendental asymptotical density. Recently, the author has given the first example of such a language with a non-algebraic asymptotical density and a non-algebraic structure-generating-function [6]. In search of such an example, Berstel had proposed to regard the language S^2.

In this paper, we shall derive several enumeration results describing the distribution of the number of words of length n in the language S^2. Among other things, we shall compute the structure-generating-function $T(S^2; z)$ and the asymptotical density $d(S^2)$. The author is, however, unable to decide if $T(S^2; z)$ is a transcendental or an algebraic function.

2. Preliminaries

Let $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$, where ε is the identity in the monoid Σ^*. The length of $w \in \Sigma^*$ is denoted by $l(w)$. As usual, a word $w \in \Sigma^*$ is said to be *primitive* if $w = v^n$, $v \in \Sigma^*$, $n \in \mathbb{N}_0$ implies $w = v$. The set of all primitive words in Σ^* is denoted by PRIM(Σ). Two words $w, v \in \Sigma^*$ are said to be *conjugates* if there exist $\alpha, \beta \in \Sigma^*$ and $p, q \in \mathbb{N}_0$ such that $w = (\alpha\beta)^p$ and $v = (\beta\alpha)^q$. In this case, we write $w \sim v$. The following lemma is shown in [4].

Lemma 1. *Let* $w \in \Sigma^+$, $v \in \Sigma^*$ *and* $w \sim v$. *There is exactly one quadruple* $(\alpha, \beta, n, m) \in \Sigma^* \times \Sigma^+ \times \mathbb{N} \times \mathbb{N}_0$ *with* $w = (\alpha\beta)^n$, $v = (\beta\alpha)^m$ *and* $\alpha\beta \in \mathrm{PRIM}(\Sigma)$.

Now, let ${}^+S^2 = S^2 \setminus \{\varepsilon\}$, ${}^+S = S \setminus \{\varepsilon\}$ and $w \in {}^+S^2$. The tupel (f, g) is called a *factorization of* w if $w = fg$ with $f \in S$ and $g \in {}^+S$. The number of all factorizations of $w \in {}^+S^2$ is denoted by $F(w)$. We have $F(w) \geq 1$ for all $w \in {}^+S^2$ by definition. The following lemma is a direct implication of [5, Lemma 4.2] and [4, Corollary 2.2.3].

Lemma 2. *Let* $n \in \mathbb{N}$, $w \in \mathrm{PRIM}(\Sigma)$ *and* $w^n \in {}^+S^2$. *We have for all* $m \in \mathbb{N}$:

(a) $w^m \in {}^+S^2$ *and*

(b) $F(w^m) = m$.

Now, let $S_k = \{w \in {}^+S^2 \mid F(w) = k\}$ be the set of all words in ${}^+S^2$ having exactly k factorizations. We prove the following:

Lemma 3. *The set* S_1 *is a subset of* PRIM(Σ).

Proof. Let $u \in S_1$. There is exactly one factorization (f, g) of u. We have $u = fg$ and $u^R = gf$ and therefore $u \sim u^R$. By Lemma 1 there is exactly one quadruple $(\alpha, \beta, p, q) \in \Sigma^* \times \Sigma^+ \times \mathbb{N} \times \mathbb{N}_0$ with $u = (\alpha\beta)^q$, $u^R = (\beta\alpha)^p$ and $\alpha\beta \in \mathrm{PRIM}(\Sigma)$. Using Lemma 2 with $m = 1$ we obtain $q = F((\alpha\beta)^q) = F(u) = 1$. Hence, $u = \alpha\beta \in$ Prim(Σ). This completes the proof.

The preceding lemmata enable us to give a characterization of ${}^+S^2$.

Theorem 1. *Let* $n, d \in \mathbb{N}$ *with* $d \mid n$. *We have*:

$$ {}^+S^2 \cap \Sigma^n = \bigcup_{d \mid n} \{u^d \mid u \in S_1 \cap \Sigma^{n/d}\}. $$

Proof. (a) Let $w \in {}^+S^2 \cap \Sigma^n$. There is at least one factorization (f, g) of w. Since $w = fg$ and $w^R = gf$, we have $w \sim w^R$. Using Lemma 1 there is exactly one quadruple $(\alpha, \beta, p, q) \in \Sigma^* \times \Sigma^+ \times \mathbb{N} \times \mathbb{N}_0$ with $w = (\alpha\beta)^p$, $w^R = (\beta\alpha)^q$ and $\alpha\beta \in \mathrm{PRIM}(\Sigma)$. By Lemma 2 we obtain $(\alpha\beta)^m \in {}^+S^2$ and $F((\alpha\beta)^m) = m$ for all $m \in \mathbb{N}$. Choosing $m = 1$ we get $\alpha\beta \in {}^+S^2$ and $F(\alpha\beta) = 1$. Hence, $\alpha\beta \in S_1$. Since $l(w) = n$ and $w = (\alpha\beta)^p$, we have further $p \mid n$. Therefore, $w \in \{u^p \mid u \in S_1 \cap \Sigma^{n/p}\}$ with $p \mid n$.

(b) Let $w \in \{u^d \mid u \in S_1 \cap \Sigma^{n/d}\}$ where $d \mid n$, that is $w = u^d$ for some $u \in S_1 \cap \Sigma^{n/d}$. There is exactly one factorization (f, g) of u. Choose $a := f$ and $b := g(fg)^{d-1}$. A simple calculation shows that $ab = u^d = w \neq \varepsilon$, $a^R = a$ and $b^R = b$. Hence, $w \in {}^+S^2$. Furthermore, $l(w) = dl(u) = d(n/d) = n$. Therefore, $w \in {}^+S^2 \cap \Sigma^n$.

This completes the proof of Theorem 1.

Corollary 1. *Each word $w \in {}^+S^2 \cap \Sigma^*$ has exactly one representation of the form $w = u^d$ with $u \in S_1 \cap \Sigma^{n/d}$ and $d \mid n$. Furthermore, $F(w) = d$.*

Proof. Theorem 1 shows that each $w \in {}^+S^2 \cap \Sigma^n$ has at least one such representation. Assume $w = u^d = y^t$ with $u \in S_1 \cap \Sigma^{n/d}$, $y \in S_1 \cap \Sigma^{n/t}$, $d \mid n$ and $t \mid n$. We have $u, y \in \mathrm{PRIM}(\Sigma)$ by Lemma 3. Using Lemma 2 we get $F(w) = F(u^d) = d$ and $F(w) = F(y^t) = t$ and therefore $d = t$ which implies $l(u) = l(y)$. Generally, $xz = rs$ with $l(x) = l(r)$ and $x, z, r, s \in \Sigma^*$ implies $x = r$ and $z = s$ (see [8]). Choosing $x := u$, $z := u^{d-1}$, $r := y$ and $s := y^{d-1}$, we obtain $u = y$. An application of Lemma 2(b) leads immediately to $F(w) = d$.

3. Enumeration results

This section is devoted to the computation of the number of words of length n in S^2.

Lemma 4.

$$\mathrm{card}({}^+S^2 \cap \Sigma^n) = \sum_{d \mid n} \mathrm{card}(S_1 \cap \Sigma^d) \quad \textit{for } n \geq 1.$$

Proof. By Theorem 1 and Corollary 1 we have:

$$\mathrm{card}({}^+S^2 \cap \Sigma^n) = \sum_{d \mid n} \sum_{u \in S_1 \cap \Sigma^{n/d}} \mathrm{card}(u^d) = \sum_{d \mid n} \mathrm{card}(S_1 \cap \Sigma^{n/d}).$$

This expression is equivalent to our statement.

By Corollary 1, $w \in S_k \cap \Sigma^n$ has exactly one representation of the form $w = u^k$

with $u \in S_1 \cap \Sigma^{n/k}$ and $n \mid k$. Hence, we further have:

Lemma 5.

$$\operatorname{card}(S_k \cap \Sigma^n) = \begin{cases} 0 & \text{if } k \nmid n, \\ \operatorname{card}(S_1 \cap \Sigma^{n/k}) & \text{if } k \mid n. \end{cases}$$

In other words, the number of words of length n in ${}^+S^2$ with k factorizations is equal to the number of words of length n/k in ${}^+S^2$ with one factorization.

Lemma 6. *Let $a \in \mathbb{N}$ and let $R_a : \mathbb{N} \to \mathbb{N}$ be the arithmetical function defined by $R_a(n) = \frac{1}{4}n\, a^{n/2}((1+\sqrt{a})^2 + (-1)^n(1-\sqrt{a})^2)$. We have for $n \geq 1$;*

$$R_{\operatorname{card}(\Sigma)}(n) = \sum_{d \mid n} \frac{n}{d} \operatorname{card}(S_1 \cap \Sigma^d).$$

Proof. Note that $R_a(n) \in \mathbb{N}$ for $a, n \in \mathbb{N}$. Let ¢ be a new symbol not in Σ, $B := \Sigma \cup \{¢\}$ and $\boldsymbol{S}^2 := S¢^+S$. Obviously, $h(\boldsymbol{S}^2) = {}^+S^2$ where $h : (B^*, \cdot) \to (\Sigma^*, \cdot)$ is the monoid homomorphism defined by $h(a) = a$ if $a \in \Sigma$, and $h(c) = \varepsilon$ if $a = ¢$. It is not hard to see that for $w \in {}^+S^2$ there are exactly $F(w)$ words $w_i \in \boldsymbol{S}^2$ with $h(w_i) = w$, $1 \leq i \leq F(w)$, because for each factorization (f, g) of w there is exactly one corresponding word $f¢g \in \boldsymbol{S}^2$. By Corollary 1, each $w \in {}^+S^2 \cap \Sigma^n$ has a unique representation of the form $w = u^d$ with $d \mid n$ and $u \in S_1 \cap \Sigma^{n/d}$; we obtain then $F(w) = d$. Therefore

$$\begin{aligned} \operatorname{Card}(\boldsymbol{S}^2 \cap B^{n+1}) &= \sum_{w \in {}^+S^2 \cap \Sigma^n} F(w) = \sum_{d \mid n} \sum_{u \in S_1 \cap \Sigma^{n/d}} F(u^d) \\ &= \sum_{d \mid n} d \operatorname{card}(S_1 \cap \Sigma^{n/d}) = \sum_{d \mid n} \frac{n}{d} \operatorname{card}(S_1 \cap \Sigma^d). \end{aligned}$$

On the other hand, we have

$$\operatorname{card}(\boldsymbol{S}^2 \cap B^{n+1}) = \sum_{0 \leq k \leq n} \operatorname{card}(S \cap \Sigma^k) \operatorname{card}({}^+S \cap \Sigma^{n-k}).$$

Now, it can be easily shown, that the number of palindromes of length m over Σ^* is given by

$$\operatorname{card}(S \cap \Sigma^m) = \operatorname{card}(\Sigma)^{\lfloor (m+1)/2 \rfloor}.$$

Using this result, we obtain finally, by an elementary computation,

$$\begin{aligned} \operatorname{card}(\boldsymbol{S}^2 \cap B^{n+1}) &= \sum_{0 \leq k \leq n-1} \operatorname{card}(\Sigma)^{\lfloor (k+1)/2 \rfloor + \lfloor (n-k+1)/2 \rfloor} \\ &= R_{\operatorname{card}(\Sigma)}(n), \end{aligned}$$

where $R_{\operatorname{card}(\Sigma)}$ is the function defined in our lemma. This completes the proof.

For sake of convenience, henceforth we use the notation $f * g$ for the Dirichlet convolution of two arithmetical functions f, g (see [1]), that is $h = f * g$ stands for $h(n) = \sum_{d|n} f(d)\,g(n/d)$. The ordinary product fg of two arithmetical functions f and g is defined by the usual formula $(fg)\,(n) = f(n)\,g(n)$. Furthermore, let $u(n)$ be the arithmetical function such that $u(n) = 1$ for all $n \in \mathbb{N}$, $N(n)$ be the function such that $N(n) = n$ for all $n \in \mathbb{N}$, $I(n)$ be the function such that $I(n) = \delta_{n,1}$ for all $n \in \mathbb{N}$, where $\delta_{n,i}$ is Kronecker's symbol, $\mu(n)$ be the Möbius function and $\varphi(n)$ be Euler's totient function [1]. We now prove the following:

Theorem 2. *The number of words of length n in the language S^2 is given by* $\mathrm{card}(S^2 \cap \Sigma^n) = 1$ *for $n = 0$ and by*

$$\mathrm{card}(S^2 \cap \Sigma^n) = \sum_{d|n} \varphi^{-1}(d) R_{\mathrm{card}(\Sigma)}(n/d)$$

for $n \geqslant 1$. Here, $R_{\mathrm{card}(\Sigma)}$ is the function given in Lemma 6 *and φ^{-1} is the Dirichlet inverse of Euler's totient function given by*

$$\varphi^{-1}(1) = 1, \qquad \varphi^{-1}(n) = \prod_{\substack{p|n\\ p \text{ prime}}} (1-p) \quad \textit{for } n \geqslant 2.$$

Proof. The case $\mathrm{card}(S^2 \cap \Sigma^0) = 1$ is obvious. Now, let $n \geqslant 1$. With the notation $H(n) := \mathrm{card}({}^+S^2 \cap \Sigma^n)$ and $L_k(n) := \mathrm{card}(S_k \cap \Sigma^n)$ Lemmas 4 and 6 can be stated in the form: $H = u * L_1$ and $R_{\mathrm{card}(\Sigma)} = N * L_1$. Since $*$ is associative and the Dirichlet inverse φ^{-1} exists, we can make the following computation:

$$\begin{aligned}
H &= u * L_1 && \text{(by Lemma 4)},\\
&= u * L_1 * I && (\text{since } f * I = I * f = f \text{ for any } f)\\
&= u * L_1 * \varphi * \varphi^{-1} && (\text{since } f * f^{-1} = f^{-1} * f = I \text{ if } f^{-1} \text{ exists}),\\
&= u * \varphi^{-1} * L_1 * N * \mu && (\text{since } * \text{ is commutative and } \varphi = N * \mu,\\
& && \quad \text{see [1, p. 29]}),\\
&= \mu * u * \varphi^{-1} * R_{\mathrm{card}(\Sigma)} && (\text{since } * \text{ is commutative and Lemma 6}),\\
&= I * \varphi^{-1} * R_{\mathrm{card}(\Sigma)} && (\text{since } I = \mu * u, \text{ (see [1, p. 31])},\\
&= \varphi^{-1} * R_{\mathrm{card}(\Sigma)}.
\end{aligned}$$

Hence, $H = \varphi^{-1} * R_{\mathrm{card}(\Sigma)}$. This relation is equivalent to our statement.

Theorem 3. *The number of all words of length n in the language ${}^+S^2$ with k factorizations is given by*

$$\mathrm{card}(S_k \cap \Sigma^n) = \begin{cases} 0 & \textit{if } k \nmid n,\\ \displaystyle\sum_{d|(n/k)} \frac{n}{kd}\,\mu\!\left(\frac{n}{kd}\right) R_{\mathrm{card}(\Sigma)}(d) & \textit{if } k \mid n. \end{cases}$$

Proof. Let again $H(n)=\mathrm{card}({}^+S^2\cap\Sigma^n)$ and $L_k(n)=\mathrm{card}(S_k\cap\Sigma^n)$. We get:

$$\begin{aligned}
L_1 &= L_1 * I && \text{(since } f*I=I*f=f \text{ for any } f)\\
&= L_1*\mu*u && \text{(since } I=\mu*u, \text{ see [1, p. 31]),}\\
&= \mu*H && \text{(since } * \text{ is commutative and Lemma 4)}\\
&= \mu*\varphi^{-1}*R_{\mathrm{card}(\Sigma)} && \text{(by Theorem 2),}\\
&= \mu*\mu^{-1}*N^{-1}*R_{\mathrm{card}(\Sigma)} && \text{(since } \varphi^{-1}=\mu^{-1}*N^{-1}, \text{ see [1, p. 37]),}\\
&= I*\mu N*R_{\mathrm{card}(\Sigma)} && \text{(since } \mu*\mu^{-1}=I \text{ and } N^{-1}=\mu N,\\
& && \quad\text{see [1, p. 37]),}\\
&= \mu N*R_{\mathrm{card}(\Sigma)}.
\end{aligned}$$

Hence, $L_1=\mu N*R_{\mathrm{card}(\Sigma)}$. Now, an application of Lemma 5 leads directly to our statement.

By inspection of Theorems 1 and 2 we further obtain the following:

Corollary 2. *Let p be a prime. We have*

(a) $$\mathrm{card}(S^2\cap\Sigma^p)=\begin{cases}\mathrm{card}(\Sigma)^2 & \text{if } p=2,\\ p\,\mathrm{card}(\Sigma)^{(p+1)/2}-(p-1)\mathrm{card}(\Sigma) & \text{if } p\neq 2;\end{cases}$$

(b) $$\mathrm{card}(S_k\cap\Sigma^p)=\begin{cases}\mathrm{card}(\Sigma)^2-\mathrm{card}(\Sigma) & \text{if } k=1 \text{ and } p=2,\\ p[\mathrm{card}(\Sigma)^{(p+1)/2}-\mathrm{card}(\Sigma)] & \text{if } k=1 \text{ and } p\neq 2,\\ \mathrm{card}(\Sigma) & \text{if } k=p,\\ 0 & \text{otherwise.}\end{cases}$$

Corollary 3. *Let* $\mathrm{card}(\Sigma)\geq 2$. *We have the asymptotic* formula

$$\mathrm{card}(S^2\cap\Sigma^n)=R_{\mathrm{card}(\Sigma)}(n)\Phi(n),$$

where $R_{\mathrm{card}(\Sigma)}$ is the function given in Lemma 6 and $\Phi(n)$ is given by $\Phi(n)=1+O(\sqrt{n}\,\mathrm{card}(\Sigma)^{-n/4})$.

Proof. Obviously, $R_a(n)\leq\frac{1}{2}n(a+1)a^{n/2}$ for all $a\geq 0$ and $|\varphi^{-1}(n)|\leq n$ for all $n\in\mathbb{N}$. Let $d(n)$ be the number of all positive divisors of the natural number n. Since surely $d(n)=O(\sqrt{n})$, we can make the following estimations:

$$\begin{aligned}
\left|\sum_{\substack{d\mid n\\ d\geq 2}}\varphi^{-1}(d)R_{\mathrm{card}(\Sigma)}(n/d)\right| &\leq \sum_{\substack{d\mid n\\ d\geq 2}}|\varphi^{-1}(d)|\,R_{\mathrm{card}(\Sigma)}(n/d)\\
&\leq \tfrac{1}{2}n(1+\mathrm{card}(\Sigma))\sum_{\substack{d\mid n\\ d\geq 2}}\mathrm{card}(\Sigma)^{n/2d}\\
&\leq \tfrac{1}{2}n(1+\mathrm{card}(\Sigma))\,\mathrm{card}(\Sigma)^{n/4}[d(n)-1]\\
&= O(n\sqrt{n}\,\mathrm{card}(\Sigma)^{n/4}).
\end{aligned}$$

Since $\varphi^{-1}(1)=1$ we obtain with Theorem 2,

$$\mathrm{card}(^{+}S^2 \cap \Sigma^n) = R_{\mathrm{card}(\Sigma)}(n) + \mathrm{O}(n\sqrt{n}\,\mathrm{card}(\Sigma)^{n/4}).$$

This relation is equivalent to our statement.

The number of words in $S^2 \cap \Sigma^n$ for some n and card(Σ) is given in Table 1.

Table 1. The numbers card($S^2 \cap \Sigma^n$) for some n and some card(Σ)

n \ card(Σ)	1	2	3	4	5
0	1	1	1	1	1
1	1	2	3	4	5
2	1	4	9	16	25
3	1	8	21	40	65
4	1	16	57	136	265
5	1	32	123	304	605
6	1	52	279	880	2 125
7	1	100	549	1 768	4 345
8	1	160	1 209	4 936	14 665
9	1	260	2 127	9 112	27 965
10	1	424	4 689	25 216	93 025
20	1	30 136	2 356 737	52 402 336	585 842 065
30	1	1 469 632	860 825 439	$8.052959968_{10}10$	$2.746575977_{10}12$

4. Some statistical results

In this section, we shall compute the asymptotical density $d(S^2)$ of the language S^2 and the average number of factorizations of a word in S^2.

Theorem 4. *The asymptotical density $d(S^2)$ of the language S^2 over the alphabet Σ is given by*

$$d(S^2) = \delta_{1,\mathrm{card}(\Sigma)}$$

where $\delta_{i,k}$ is Kronecker's symbol.

Proof. We consider the quotient $\rho_n(S^2) = \mathrm{card}(S^2 \cap \Sigma^n)/\mathrm{card}(\Sigma)^n$. First, let card($\Sigma$) = 1. In this case we have $R_{\mathrm{card}(\Sigma)} = n$ for all $n \in \mathbb{N}$. Since $\varphi^{-1} = \mu^{-1} * N^{-1}$ (see [1, p. 37]), we obtain further $N * \varphi^{-1} = \mu^{-1} = \mu$ (see [1, p. 31]). Hence with Theorem 2, $\mathrm{card}(S^2 \cap \Sigma^n) = u(n) = 1$ for all $n \in \mathbb{N}$. Therefore $\rho_n(S^2) = 1$ for all $n \in \mathbb{N}$.

Next, let card(Σ) $\geqslant 2$. In this case we obtain with Corollary 3, $\rho_n(S^2) = \mathrm{O}(n\,\mathrm{card}(\Sigma)^{-n/4})$. Therefore, $\eta = \lim_{n\to\infty} \rho_n(S^2) = \delta_{1,\mathrm{card}(\Sigma)}$. Now, the same calculation as in [6] shows that $d(S^2) = \eta$. This completes the proof of our theorem.

Assuming that all words of length n in S^2 are equally likely, the quotient $p(n,k)=\text{card}(S_k\cap\Sigma^n)/\text{card}(S^2\cap\Sigma^n)$ is the probability that a word $w\in S^2$ of length n has exactly k factorizations. The sth moment about origin is defined by $m_s(n)=\sum_{1\le k\le n}k^s p(n,k)$. We prove the following:

Lemma 7. *The* sth *moment about origin is given by*

$$m_s(n)=[\text{card}(S^2\cap\Sigma^n)]^{-1}\sum_{d\mid n} dJ_{s-1}(d)\,R_{\text{card}(\Sigma)}(n/d)$$

where $R_{\text{card}(\Sigma)}$ *is the function given in Lemma* 6 *and* J_s *is Jordan's totient function defined by*

$$J_s(1)=1 \quad \text{and} \quad J_s(n)=n^s\prod_{\substack{p\mid n\\ p\ \text{prime}}}(1-p^{-s}).$$

Proof. An application of Theorem 3 leads directly to

$$m_s(n)\text{card}(S^2\cap\Sigma^n)=\sum_{k\mid n}k^s\sum_{d\mid(n/k)}\frac{n}{kd}\,\mu\left(\frac{n}{kd}\right)R_{\text{card}(\Sigma)}(d).$$

A simple rearrangement of the terms on the right side yields to

$$m_s(n)\,\text{card}(S^2\cap\Sigma^n)=\sum_{d\mid n}\frac{n}{d}R_{\text{card}(\Sigma)}(d)\sum_{k\mid(n/d)}\mu(k)\left(\frac{n}{kd}\right)^{s-1}.$$

Now, Jordan's totient function has a representation of the form $J_s(n)=\sum_{k\mid n}\mu(k)\,(n/k)^s$ (see [1, p. 48]). Using this relation we get our lemma.

Theorem 5. *The* sth *moment about origin is asymptotically given by*

$$m_s(n)=\begin{cases} n^s & \text{if } \text{card}(\Sigma)=1,\\ 1+O(n^\alpha\,\text{card}(\Sigma)^{-n/4}) & \text{if } \text{card}(\Sigma)\geq 2\end{cases}$$

where $\alpha=\max(\frac{1}{2},s-1)$.

Proof. If $\text{card}(\Sigma)=1$, then we have $R_{\text{card}(\Sigma)}(n)=n$ for all $n\in\mathbb{N}$. In this case we obtain with Lemma 7 and Theorem 2,

$$m_s(n)=\sum_{d\mid n}dJ_{s-1}(d)\frac{n}{d}=n^s$$

because in general $n^s=\sum_{d\mid n}J_s(d)$ (see [1, p. 48]).

Now, let $\text{card}(\Sigma)\geq 2$. Since $J_s(n)$ and $R_{\text{card}(\Sigma)}(n)$ are always positive and

$R_a(n) \leqslant \frac{1}{2}n(a+1)a^{n/2}$ for $a \geqslant 0$, we can make the following estimations:

$$\begin{aligned}\left|\sum_{\substack{d\mid n\\ d\geqslant 2}} d\,J_{s-1}(d)\,R_{\mathrm{card}(\Sigma)}(n/d)\right| &\leqslant \tfrac{1}{2}(1+\mathrm{card}(\Sigma))n \sum_{\substack{d\mid n\\ d\geqslant 2}} \mathrm{card}(\Sigma)^{n/2d} J_{s-1}(d)\\ &\leqslant \tfrac{1}{2}(1+\mathrm{card}(\Sigma))n\,\mathrm{card}(\Sigma)^{n/4}[n^{s-1}-1]\\ &= \mathrm{O}(n^s\,\mathrm{card}(\Sigma)^{n/4}).\end{aligned}$$

Using this relation we obtain our statement by inspection of Lemma 7 and Corollary 3.

Since $m_1(n)$ is the average number of factorizations of a word $w \in S^2$ of length n and the variance $\sigma^2(n)$ is given by $\sigma^2(n) = m_2(n) - m_1^2(n)$, we obtain immediately:

Corollary 4. *Assuming that all words of length n in S^2 are equally likely, the average number of factorizations of a word $w \in S^2 \cap \Sigma^n$ is asymptotically given by*

$$m_1(n) = \begin{cases} n & \textit{if } \mathrm{card}(\Sigma) = 1,\\ 1 + \mathrm{O}(\sqrt{n}\,\mathrm{card}(\Sigma)^{-n/4}) & \textit{if } \mathrm{card}(\Sigma) \geqslant 2.\end{cases}$$

The variance is asymptotically given by

$$\sigma^2(n) = \begin{cases} n(n-1) & \textit{if } \mathrm{card}(\Sigma) = 1,\\ \mathrm{O}(n\,\mathrm{card}(\Sigma)^{-n/4}) & \textit{if } \mathrm{card}(\Sigma) \geqslant 2.\end{cases}$$

Corollary 4 shows that in the average all words $w \in S^2$ have exactly one factorization into two palindromes provided that $\mathrm{card}(\Sigma) \geqslant 2$.

5. Concluding remarks

In this paper we have derived several enumeration results describing the distribution of the number of words of length n in the inherently ambiguous context-free language S^2. Using Theorem 2 we can also compute the structure-generating-function $T(S^2; z)$; we obtain

$$\begin{aligned} T(S^2; z) &= \sum_{n\geqslant 0} \mathrm{card}(S^2 \cap \Sigma^n) z^n\\ &= 1 + \sum_{n\geqslant 1} z^n \sum_{d\mid n} \varphi^{-1}(d) R_{\mathrm{card}(\Sigma)}(n/d).\end{aligned}$$

A simple rearrangement of the terms in the last sum shows that an equivalent expression is given by

$$T(S^2; z) = 1 + \sum_{j\geqslant 1} \varphi^{-1}(j) \sum_{\lambda\geqslant 1} R_{\mathrm{card}(\Sigma)}(\lambda) z^{\lambda j}.$$

Using the definition of $R_{\mathrm{card}(\Sigma)}$ from Lemma 6, the second sum can be calculated explicitly for $|z|<\mathrm{card}(\Sigma)^{-1/2}$; we get

$$T(S^2;z)=1+\tfrac{1}{4}\sqrt{\mathrm{card}(\Sigma)}\sum_{j\geq 1}\varphi^{-1}(j)z^j$$
$$\times\left[\frac{(1+\sqrt{\mathrm{card}(\Sigma)})^2}{(1-\sqrt{\mathrm{card}(\Sigma)}z^j)^2}-\frac{(1-\sqrt{\mathrm{card}(\Sigma)})^2}{(1+\sqrt{\mathrm{card}(\Sigma)}z^j)^2}\right]$$

or, equivalently,

$$T(S^2;z)=1+\sum_{j\geq 1}\varphi^{-1}(j)\frac{\mathrm{card}(\Sigma)z^j(1+z^j)(1+\mathrm{card}(\Sigma)z^j)}{(1-\mathrm{card}(\Sigma)z^{2j})^2}$$

where $|z|<\mathrm{card}(\Sigma)^{-1/2}$.

If card $(\Sigma)=1$, we obtain immediately for $|z|<1$:

$$T(S^2;z)=1+\sum_{j\geq 1}\varphi^{-1}(j)\frac{z^j}{(1-z^j)^2}=\frac{1}{1-z},$$

because it is well known that the last sum is equal to $z/(1-z)$ for $|z|<1$. In this case, we obtain the expected result that $T(S^2;z)$ is rational function. In the general case, that is for $\mathrm{card}(\Sigma)>1$, the author is unable to give a simple expression or a functional equation for $T(S^2;z)$. In order to prove that $T(S^2;z)$ is a transcendental function, it is sufficient to show, that there is no linear recurrence relation for the numbers $\mathrm{card}(S^2\cap\Sigma^n)$ with polynomials in n as coefficients, because such a recurrence always exists for the Taylor coefficients of any algebraic function [3].

It is not hard to show that there is no recurrence relation of the form $P(n)\,\mathrm{card}(S^2\cap\Sigma^{n+1})+Q(n)\,\mathrm{card}(S^2\cap\Sigma^n)=T(n)$ where $P(n)$, $Q(n)$ and $T(n)$ are polynomials in n, but it seems that there is no obvious generalization.

References

[1] T.M. Apostol, Introduction to Analytic Number Theory (Springer, New York, 1976).
[2] J. Berstel, Sur la densité asymptotique des langages formels, in: M. Nivat, Ed., Automata, Languages and Programming (North-Holland, Amsterdam, 1973) 345–358.
[3] L. Comtet, Calcul pratique des coefficients de Taylor d'une fonction algébrique, Enseign Math. 10 (1964) 267–270.
[4] J.P. Crestin, Sur un langage quasi-rationnel d'ambiguité inherénte non bornée, Thèse de 3° Cycle, Faculty of Science, University of Paris (1969).
[5] J.P. Crestin, Un langage non ambigu dont le carré est d'ambiguité non bornée, in: M. Nivat, ed., Automata, Languages and Programming (North-Holland, Amsterdam, 1973) 377–390.
[6] R. Kemp, A note on the density of inherently ambiguous context-free languages, Acta Inform. 14 (1980) 295–298.
[7] W. Kuich, On the entropy of context-free languages Inform. and Control 16 (1970) 173–200.
[8] F.W. Levi, On semigroups, Bull. Calcutta Math. Soc. 36 (1944) 144–146.