ON A GENERALIZATION OF THE DYCK-LANGUAGE OVER A TWO LETTER ALPHABET

Helmut PRODINGER

Institut für Mathematische Logik und Formale Sprachen, Technische Universität Wien, Wien, Österreich

Received 30 May 1978
Revised 23 April 1979

Some properties of the language \( \{ w \in \{a, b\}^* | (\binom{w}{a}) = (\binom{w}{b}) \} \), which can be regarded as a generalization of the (unrestricted) Dyck-language, are given. \((\binom{w}{a})\) are the binomial coefficients for words.)

1. Introduction

Let \( \Sigma^* \) be the free monoid generated by the alphabet \( \Sigma \) with unit \( \varepsilon \). The binomial coefficients for words are defined as follows: For \( x, y \in \Sigma^* \) let \( \binom{x}{y} \) be the number of factorizations \( x = x_0c_1x_1 \cdots x_{n-1}c_nx_n \) where \( y = c_1 \cdots c_n \), \( c_i \in \Sigma \). They appear for the first time in [1] within the context of \( p \)-groups. They can be used in order to embed the monoid \( \Sigma^* \) in the ring of all formal power series in the noncommuting variables \( \sigma \in \Sigma \) with real coefficients by means of

\[
w \mapsto \sum_{z \in \Sigma^*} \binom{w}{z} z.
\]

See also the reference given in [5]. Since they are a generalization of the ordinary binomial coefficients \((\binom{w}{a})\) for \( \Sigma = \{\sigma\} \) and with the identification \( \sigma^n = n \), they seem to be important from a combinatorial point of view.

In the sequel it is assumed that \( \Sigma \) is the two letter alphabet \( \{a, b\} \).

The (unrestricted) Dyck-language \( D \) (cf. [2]) can be expressed as

\[
D = \left\{ w \in \{a, b\}^* \left| \binom{w}{a} = \binom{w}{b} \right. \right\}.
\]

This leads to the following generalization: For \( x, y \in \{a, b\}^* \) let

\[
D(x, y) = \left\{ w \in \{a, b\}^* \left| \binom{w}{x} = \binom{w}{y} \right. \right\}.
\]

In this paper the case \( x = ab, y = ba \) will be considered. For sake of convenience \( D(ab, ba) \) is shortly denoted by \( A \) in the sequel.
It is necessary to give few additional definitions: For \( w \in \{a, b\}^* \) let \( |w| \) denote the length of \( w \) and \( w^R \) the mirror image.

\[
\Delta(w) := \begin{pmatrix} w \\ ab \end{pmatrix} - \begin{pmatrix} w \\ ba \end{pmatrix}.
\]

Clearly \( A = \{w \in \{a, b\}^* | \Delta(w) = 0\} \). Finally let \( \sigma(a) = 1 \) and \( \sigma(b) = -1 \).

The structure generating function of a language \( L \subseteq \Sigma^* \) is the formal power series \( \sum_{n=0}^{\infty} u_n z^n \), where \( u_n = |L \cap \{a, b\}^n| \). (Cf. [6].) For \( L \subseteq \Sigma^* \) the syntactic congruence \( \sim_L \) is defined by \( x \sim_L y \) iff for all \( u, v \in \Sigma^* \) \( uvx \in L \) holds exactly if \( uyv \in L \) holds (cf. [1]).

This paper gives the following results about the language \( A \): Differently from \( D \) \( A \) is not contextfree. A submonoid of \( 3 \times 3 \) matrices with integer coefficients which is isomorphic to the syntactic monoid \( \Sigma^*/\sim_A \) of \( A \) will be given. The coefficients \( u_n \) of the structure generating function of \( A \) are examined. It turns out that \( u_n \) is the number of solutions of

\[
\sum_{k=1}^{n} \varepsilon_k (n + 1 - 2k) = 0 \quad (\varepsilon_k \in \{-1, +1\}).
\]

The asymptotic behaviour of \( u_n \) will be established by a method similar to that of Van Lint [4].

2. Results

Theorem 1. \( A \) is not contextfree.

Proof. It is sufficient to prove that \( A' := A \cap R \) is not contextfree, where \( R \) is the regular language \( a^+b^+a^+b^+ \).

For \( i \in \mathbb{N}_0 \)

\[
\begin{pmatrix} a^ib^{2i}a^3ib^i \\ ab \end{pmatrix} = i \cdot 2 \cdot i + i \cdot i + 3 \cdot i \cdot i = 6i^2 = 2 \cdot i \cdot 3 \cdot i = \begin{pmatrix} a^ib^{2i}a^3ib^i \\ ba \end{pmatrix}.
\]

Therefore \( a^ib^{2i}a^3ib^i \in A' \). Assuming \( A' \) to be contextfree the \( uvwxy \)-theorem (cf. [3]) guarantees a factorization \( a^ib^{2i}a^3ib^i = uvwx \), where \( i \) is large enough and \( ux \neq e, |uw| \leq m \), such that \( uv^nwx^n y \in A' \) for all \( n \in \mathbb{N}_0 \). It is a simple calculation to show that all possible factorizations lead to a contradiction by taking a suitable \( n \).

Next the syntactic congruence \( \sim_A \) is characterized.

Theorem 2. \( x \sim_A y \) if and only if \( \Delta(x) = \Delta(y) \), \( (y) = (\xi) \) and \( (\xi) = (\xi) \).

Proof. First it should be noted that \( w = w^R \) implies \( \Delta(w) = 0 \).
Let be \( x \sim_A y \) and \( u \in \{a, b\}^* \). Then
\[
xu(xu)^R \sim_A yu(xu)^R \quad \text{and} \quad (xu)^Rxu \sim_A (xu)^Ryu.
\]
Since \( xu(xu)^R \in A \) ((\( xu \)^R \( xu \) \( \in \) \( A \))) it follows that \( yu(xu)^R \in A \) ((\( xu \)^R \( yu \) \( \in \) \( A \))). Therefore
\[
0 = \Delta(yu(xu)^R) = \Delta(yu) - \Delta(xu) + \binom{yu}{a} \binom{xu}{b} - \binom{yu}{b} \binom{xu}{a}
\]
and
\[
0 = \Delta((xu)^Ryu) = \Delta(yu) - \Delta(xu) + \binom{xu}{a} \binom{yu}{b} - \binom{xu}{b} \binom{yu}{a}.
\]
Adding these equations
\[
\Delta(xu) = \Delta(yu) \quad \text{and} \quad \binom{xu}{a} \binom{yu}{b} = \binom{xu}{b} \binom{yu}{a}
\]
for each \( u \) is obtained. Setting \( u = \varepsilon \) yields
\[
\Delta(x) = \Delta(y) \quad \text{and} \quad \binom{x}{a} \binom{y}{b} = \binom{x}{b} \binom{y}{a}.
\]
Setting \( u = a \) yields
\[
\binom{xa}{a} \binom{ya}{b} = \binom{xa}{b} \binom{ya}{a}
\]
or equivalently
\[
\left( \binom{x}{a} + 1 \right) \binom{y}{b} = \left( \binom{x}{b} + 1 \right) \binom{y}{a}
\]
from which \( (x) = (y) \) follows. For \( u = b \) \( (x) = (y) \) is obtained in a similar way.

A simple calculation gives the second part of the proof.

**Remark.** Since
\[
\Delta(w) = 2 \binom{w}{ab} + \binom{w}{a} + \binom{w}{b} - \binom{w}{2}
\]
the condition
\[
\Delta(x) = \Delta(y) \quad \text{and} \quad \binom{x}{a} = \binom{y}{a} \quad \text{and} \quad \binom{x}{b} = \binom{y}{b}
\]
is equivalent to
\[
\binom{x}{ab} = \binom{y}{ab} \quad \text{and} \quad \binom{x}{a} = \binom{y}{a} \quad \text{and} \quad \binom{x}{b} = \binom{y}{b}.
\]
Now the syntactic monoid of $A$ can be described. For this purpose let $M$ be the submonoid of the (multiplicative) monoid of $3 \times 3$-matrices with integer coefficients which is generated by

$$m_1 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad m_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$  

**Theorem 3.** \(\{a, b\}^*/\sim_A \) is isomorphic to $M$.

**Proof.** It is easy to see that

$$\phi(w) = \begin{pmatrix} 1 & \binom{w}{a} & \binom{w}{ab} \\ 0 & 1 & \binom{w}{b} \\ 0 & 0 & 1 \end{pmatrix}$$

is the unique homomorphism from $\{a, b\}^*$ onto $M$ for which $\phi(a) = m_1$ and $\phi(b) = m_2$.

By Theorem 2 and the remark $\phi(x) = \phi(y)$ if and only if $x \sim_A y$. Hence $\sim_A$ is the congruence induced by $\phi$.

Let $\sum_{n=0}^{\infty} u_n z^n$ be the structure generating function of $A$. To study the asymptotic behaviour of $u_n$ some preparations are made.

**Lemma 1.** For each word $w = a_1 \cdots a_n$ ($a_i \in \{a, b\}$)

$$2\Delta(w) = \sum_{k=1}^{n} \sigma(a_k)(n+1-2k).$$

**Proof.** By induction on $n$.

(i) For $n = 0$, i.e. $w = \varepsilon$ the statement is obvious.

(ii) Now let $|w| = n$ be assumed.

$$2\Delta(wa) = 2\Delta(w) - 2\binom{w}{b}$$

$$= \sum_{k=1}^{n} \sigma(a_k)((n+1)-2k) - \sum_{\tau=1}^{n} \sigma(a_\tau) - 2\binom{w}{b}$$

$$= \sum_{k=1}^{n} \sigma(a_k)((n+1)-2k) - \binom{w}{a} - \binom{w}{b}$$

$$= \sum_{k=1}^{n+1} \sigma(a_k)((n+1)-2k)$$

since

$$-\binom{w}{a} - \binom{w}{b} = -n = \sigma(a)((n+1)+1-2(n+1)).$$

The calculation for $wb$ is similar.
Lemma 2. \( u_n \) is the number of solutions \( (\varepsilon_1, \ldots, \varepsilon_n) \) of
\[
\sum_{k=1}^{n} \varepsilon_k (n + 1 - 2k) = 0 \quad \varepsilon_k \in \{-1, +1\}.
\]

Proof. If \( w = a_1 \cdots a_n \in A \) then \( \Delta(w) = 0 \). By Lemma 4 \( (\sigma(a_1), \ldots, \sigma(a_n)) \) is a solution.

If conversely \( (\varepsilon_1, \ldots, \varepsilon_n) \) is a solution then \( \sigma^{-1}(\varepsilon_1) \cdots \sigma^{-1}(\varepsilon_n) \in A \). Clearly the above correspondence is 1–1.

Theorem 4.
\[
u_n = 2^{2 \cdot (n-1)/2 + 1} \left( \frac{3}{\pi} \right)^{1/2} \left[ \frac{m}{2} \right]^{-3/2},
\]
where \([x]\) denotes the greatest integer \( \leq x \).

Proof. Let \( n = 2m \). The number \( u_{2m} \) is the constant term in the expansion of
\[
\prod_{k=1}^{m} (x^{-(2k-1)} + x^{2k-1})^2
\]
which can be expressed as
\[
\frac{1}{2\pi i} \int_{\Sigma} \prod_{k=1}^{m} (z^{-(2k-1)} + z^{2k-1})^2 \frac{dz}{z}.
\]
\( (\Sigma \) is the unit circle in the complex plane.\) The substitution \( z = e^{ix} \) yields
\[
u_{2m} = \frac{2^{2m+1}}{\pi} \int_{0}^{\pi/2} \prod_{k=1}^{m} \cos^2 (2k-1)x \, dx.
\]

For \( \pi/2(2m-1) \leq x \leq \pi/2 \) is
\[
\prod_{k=1}^{m} \cos^2 (2k-1)x = O(e^{-m/6}).
\]

For \( 0 < x < \pi/2 \)
\[
\cos^2 x < e^{-x^2}
\]
holds. Therefore
\[
\int_{0}^{\pi/2(2m-1)} \prod_{k=1}^{m} \cos^2 (2k-1)x \, dx < \int_{0}^{\pi/2(2m-1)} \exp \left[ -x^2 \sum_{k=1}^{m} (2k-1)^2 \right] dx
\]
\[
= \int_{0}^{\pi/2(2m-1)} \exp \left[ -x^2 \left( \frac{4m^3}{3} - \frac{m}{3} \right) \right] dx
\]
\[
\sim \frac{(3\pi)^{1/2}}{4} m^{-3/2}.
\]
Similar to the calculation in [4] it will be shown that the symbol "<" can be replaced by "~": 

Let \(0 \leq x < \infty\), then

\[
\prod_{k=1}^{m} \cos^2(2k-1)x = \prod_{k=1}^{m} e^{-(2k-1)^2x^2} \prod_{k=1}^{m} \{1 + \mathcal{O}((2k-1)^4x^4)\}
\]

\[
= \prod_{k=1}^{m} e^{-(2k-1)^2x^2} \prod_{k=1}^{m} \{1 + \mathcal{O}(k^4x^4)\}
\]

\[
= \exp \left\{ - \sum_{k=1}^{m} (2k-1)^2x^2 + \mathcal{O}(m^{-1/3}) \right\}.
\]

Thus

\[
\int_{0}^{m/2(2m-1)} \prod_{k=1}^{m} \cos^2(2k-1)x \, dx
\]

\[
> \int_{0}^{m^{4/3}} \prod_{k=1}^{m} \cos^2(2k-1)x \, dx \sim \frac{(3\pi)^{1/2}}{4} m^{-3/2}.
\]

Hence

\[
u_{2m} \sim 2^{2m-1} \left(\frac{3}{\pi}\right)^{1/2} m^{-3/2}.
\]

For \(n = 2m + 1\) a similar calculation shows that

\[
u_{2m+1} \sim 2^{2m+1} \left(\frac{3}{\pi}\right)^{1/2} m^{-3/2}.
\]

The number of solutions of

\[
\sum_{k=1}^{n} \epsilon_k (n + 1 - 2k) = 0
\]

is the same as the number of solutions of

\[
\sum_{k=1}^{n/2} \zeta_k (2k-1) = 0 \quad \left(\sum_{k=1}^{(n-1)/2} \zeta_k k = 0\right), \quad \zeta_k \in \{-1, 0, +1\}
\]

for even (odd) \(n\):

To show the first statement let be \(n = 2m\).

\[
\sum_{k=1}^{2m} \epsilon_k (2m + 1 - 2k) = \sum_{k=1}^{m} \epsilon_k (2m + 1 - 2k) + \sum_{k=m+1}^{2m} \epsilon_k (2m + 1 - 2k)
\]

\[
= \sum_{i=1}^{m} \epsilon_{m+1-i} (2i-1) + \sum_{i=1}^{m} \epsilon_{m+i} (1-2i)
\]

\[
= \sum_{i=1}^{m} (\epsilon_{m+1-i} - \epsilon_{m+i})(2i-1).
\]
Defining $\xi_k = \frac{1}{2} (\varepsilon_{m+1} - \varepsilon_{m+1})$ there is a 1–1 correspondence between the two sets of solutions. The second statement can be seen in a similar way.

If in a solution all $\xi_k$ are in $\{-1, +1\}$, the corresponding word $w \in A$ has the property that it has no factorization $w = xcyz$ where $|x| = |z|$ and $c \in \{a, b\}$. Let $B$ denote the subset of $A$ which contains exactly the words with this property. Then the asymptotic behaviour of the coefficients $u_n$ of the structure generating function of $B$ can be established by methods similar to those of Theorem 4.

**Theorem 5.**

$$v_{2n} \sim \begin{cases} 2^{n-1/2} \left( \frac{3}{\pi} \right)^{1/2} n^{-3/2} & \text{for even } n, \\ 0 & \text{for odd } n, \end{cases}$$

$$v_{2n+1} \sim \begin{cases} 2^{n+1/2} \left( \frac{3}{\pi} \right)^{1/2} n^{-3/2} & \text{for } n \equiv 0, 3 \pmod{4}, \\ 0 & \text{for } n \equiv 1, 2 \pmod{4}. \end{cases}$$

**Proof.**

$$v_{2n} = \frac{1}{2\pi i} \int_C \prod_{k=1}^{n} \frac{(z^{2k-1} + z^{-(2k-1)})}{z} dz = \frac{2^n}{\pi} \int_0^{\pi} \prod_{k=1}^{n} \cos (2k-1)x \, dx$$

$$= \begin{cases} \frac{2^{n+1}}{\pi} \int_0^{\pi/2} \prod_{k=1}^{n} \cos (2k-1)x \, dx & \text{for even } n \\ \frac{2^n}{\pi} \int_0^{\pi/2} \prod_{k=1}^{n} \cos (2k-1)x \, dx & \text{for odd } n. \end{cases}$$

Now let $n$ be even: For $0 < x < \pi/2$, $\cos x < e^{-x^2/2}$ holds.

$$\int_0^{\pi/2(2n-1)} \prod_{k=1}^{n} \cos (2k-1)x \, dx < \int_0^{\pi/2(2n-1)} \exp \left[ -\frac{x^2}{2} \sum_{k=1}^{n} (2k-1)^2 \right] dx$$

$$\sim (3\pi)^{1/2} (2n)^{-3/2}.$$  

$$v_{2n+1} = \frac{1}{2\pi i} \int_C \prod_{k=1}^{n} (z^k + z^{-k}) \frac{dz}{z} = \frac{2^n}{\pi} \int_0^{\pi} \prod_{k=1}^{n} \cos kx \, dx$$

$$= \begin{cases} \frac{2^{n+1}}{\pi} \int_0^{\pi/2} \prod_{k=1}^{n} \cos kx \, dx & n \equiv 0, 3 \pmod{4} \\ \frac{2^n}{\pi} \int_0^{\pi/2} \prod_{k=1}^{n} \cos kx \, dx & n \equiv 1, 2 \pmod{4}. \end{cases}$$

Now let $n \equiv 0, 3 \pmod{4}$:

$$\int_0^{\pi/2n} \prod_{k=1}^{n} \cos kx \, dx < \int_0^{\pi/2n} \exp \left[ -\frac{x^2}{2} \sum_{k=1}^{n} k^2 \right] dx \sim (3\pi)^{1/2} (2n^3)^{-1/2}.$$  

The justification that "<" can be replaced by "~" is as in Theorem 4.
References