LETTER

d-Primitive Words and D(1)-Concatenated Words

Itaru KATAOKA[†] ***and*** **Tetsuo MORIYA**[†a)], ***Members***

SUMMARY In this paper, we study d-primitive words and D(1)-concatenated words. First we show that neither $D(1)$, the set of all d-primitive words, nor $D(1)D(1)$, the set of all $D(1)$-concatenated words, is regular. Next we show that for $u, v, w \in \Sigma^+$ with $|u| = |w|$, $uvw \in D(1)$ if and only if $uv^+w \subseteq D(1)$. It is also shown that every d-primitive word, with the length of two or more, is $D(1)$-concatenated.
key words: *primitive word, d-primitive word, regular component*

1. Introduction

The notion of primitive words plays an important role in algebraic theory of codes and formal languages.([7], [8], and [10]) A lot of studied have been done on subsets of the language Q of all primitive words (See [1], [2], and [3], for example). Recently, attention has been paid to the language $D(1)$ of all d-primitive words, which is a proper subset of Q [4]–[6].

In this paper, we study languages $D(1)$ and $D(1)D(1)$, the set of all $D(1)$-concatenated words. We consider the regularity of $D(1)$ and $D(1)D(1)$, a regular component contained in $D(1)$, and the inclusion relation between $D(1)$ and $D(1)D(1)$.

In Sect. 2 some basic definitions and results are presented.

In Sect. 3, the following (1) and (2) are proved.
(1) Neither $D(1)$ nor $D(1)D(1)$ is regular.
(2) For $u, v, w \in \Sigma^+$ with $|u| = |w|$, $uvw \in D(1)$ if and only if $uv^+w \subseteq D(1)$.

In Sect. 4, we consider the inclusion relation between $D(1)$ and $D(1)D(1)$, over a binary alphabet. It is proved that for a word w in $D(1)$, with the length of two or more, w is in $D(1)D(1)$.

2. Preliminaries

Let Σ be a finite alphabet consisting of at least two letters. Σ^* denotes the free moniod generated by Σ, that is, the set of all finite words over Σ, including the empty word ϵ, and $\Sigma^+ = \Sigma^* - \epsilon$. For w in Σ^* , $|w|$ denotes the length of w. Any subset of Σ^* is called a *language* over Σ. For a word $u \in \Sigma^+$, by u^+ we mean the set $\{u\}^+$.

For a word $u \in \Sigma^+$, if $u = vw$ for some $v, w \in \Sigma^*$, then v (w) is called a *prefix (suffix)* of u, denoted by $v \leq_p u$ ($w \leq_s u$, *resp.*). If $v \leq_p u$ ($w \leq_s u$) and $u \neq v (w \neq u)$, then v (w) is called a *proper prefix (proper suffix)* of u, denoted by $v <_p u$ ($w <_s u$, resp.). For a word w, let $Pref(w)$ ($Suff(w)$) be the set of all prefixes (suffixes, resp.) of w.

A nonempty word u is called a *primitive word* if $u = f^n$, for some $f \in \Sigma^+$, and some $n \geq 1$ always implies that $n = 1$. Let Q be the set of all primitive words over Σ. A nonempty word u is a *non-overlapping word* if $u = vx = yv$ for some $x, y \in \Sigma^+$ always implies that $v = \epsilon$. Let $D(1)$ be the set of all non-overlapping words over Σ. A words in $D(1)$ is also called a *d-primitive word*. For $u \in \Sigma^+$, u is said to be $D(1)$-concatenated if there exist $x, y \in D(1)$ such that $xy = u$, i.e., $u \in D(1)D(1)$. (See [4] and [11]).

For $x, y \in \Sigma^+$, if $(Pref(x) - \{\epsilon\}) \cap (Suff(y) - \{\epsilon\}) = \phi$, then (x, y) is said to be a non-overlapping pair (n-o. pair).

For $u, v, w \in \Sigma^*$, the language of the form uv^+w is called a *regular component*.

We have the following property concerning d-primitive words.

Lemma 1: ([5]) Let $u \in \Sigma^+$. Then $u \notin D(1)$ iff there exists a unique word $v \in D(1)$ with $|v| \leq (\frac{1}{2})|u|$ such that $u = vwv$ for some $w \in \Sigma^*$. ::

Remark 1: Let $u, v \in \Sigma^+$. Obviously $uv \in D(1)$ implies that (u, v) is a n-o. pair. The converse does not hold; for $u = abbbba$, and $v = bb$, (u, v) is a n-o. pair but uv is not in $D(1)$. However, in the next Proposition, we show the above two are equivalent on the condition that u and v are in $D(1)$.

Proposition 2: Let u and v be in $D(1)$. The following are equivalent.
(1)Both uv and vu are in $D(1)$.
(2)Both (u, v) and (v, u) are n-o. pairs.

[Proof] (1) $\Rightarrow$ (2) : Obvious.
(2) $\Rightarrow$ (1) : Suppose that (2) holds but $uv \notin D(1)$ or $vu \notin D(1)$. It suffices to show the result only for the case of $uv \notin D(1)$. We can write $uv = zwz$ for some $z \in \Sigma^+$, $w \in \Sigma^*$.

Since (u, v) is n-o.pair, obviously $|u| \neq |z|$.
(Case 1) $z <_p u$
(1.1) $zw <_p u$. We have that $u = zwy$ and $z = yv$ for some $y \in \Sigma^+$. Thus $u = yvwy \notin D(1)$.
(1.2) $u \leq_p zw$. We have that $u = zw_1$ and $v = w_2z$ for some $w_1 \in \Sigma^+$, $w_2 \in \Sigma^*$ with $w = w_1w_2$. Thus (u, v) is not n-o.

Manuscript received February 2, 2009.
Manuscript revised March 31, 2009.
[†]The authors are with School of Science and Engineering, Kokushikan University, Tokyo, 154–8515 Japan.
a) E-mail: moriya@kokushikan.ac.jp
DOI: 10.1587/transinf.E92.D.1577

pair.
(Case 2) $u <_p z$
We have that $v = xwz, z = ux$ for some $x \in \Sigma^+$. Thus $v = xwux \notin D(1)$. ::

The next lemma is immediate by Lemma 1.

Lemma 3: (1) For a n-o. pair (x, y) and $c \in \Sigma$, with $|x| = |y|$, both xy and xcy are in $D(1)$.
(2) Let $w \in D(1)$. For every $x \in Pref(w) - \{\epsilon\}$ and $y \in Suff(w) - \{\epsilon\}$ with $(x, y) \neq (w, w)$, (x, y) is a n-o.pair. ::

3. Regularity of D(1) and $D(1)D(1)$

In this section first we prove that neither D(1) nor $D(1)D(1)$ is regular.

Proposition 4: $D(1)$ is not regular.

[Proof] Suppose that $D(1)$ were a regular set. Then there would be an integer n satisfying the conditions of the pumping lemma. Let x be $a^{n+1}b^na^nb^n$. Then the word x can be written as uvw for some $u, w \in \Sigma^*$, $v \in \Sigma^+$. Since $|uv| \leq n$, uv is in a^+. Moreover, we have that uw is in $D(1)$ by the pumping lemma. It follows that $uw = a^mb^na^nb^n$ for some $m \leq n$. However, uw is not in $D(1)$ since a^mb^n is in $Pref(uw) \cap Suff(uw)$. This is a contradiction. ::

Proposition 5: $D(1)D(1)$ is not regular.

[Proof] Suppose that $D(1)D(1)$ were a regular set. Let n be the integer in the pumping lemma, and let x be $a^{n+1}b^na^nb^na^nb^n$. Then the word x can be written as uvw for some $u, w \in \Sigma^*$, $v \in \Sigma^+$. Since $|uv| \leq n$, uv is in a^+. Moreover, we have that uw is in $D(1)D(1)$ by the pumping lemma. It follows that $uw = a^mb^na^nb^na^nb^n$ for some $m \leq n$. However, uw is not in $D(1)D(1)$ since neither $a^nb^na^nb^n$ nor $a^mb^na^nb^n$ is in $D(1)$. This is a contradiction. ::

Next we study a regular component contained in D(1).

Proposition 6: Let $|u| = |w|$ for $u, v, w \in \Sigma^+$. Then $uvw \in D(1)$ if and only if $uv^+w \subseteq D(1)$.

[Proof]
[if] Trivial.
[Only if] Suppose that uvw is in $D(1)$ for $u, w, v \in \Sigma^+$, with $|u| = |w|$. We shall show that for every $n \geq 2$, uv^nw is in $D(1)$ by induction.
(Basis) $n = 2$. By Lemma 3(2), (uv, vw) is n-o.pair. Then we have that $uvvw$ is in $D(1)$ by Lemma 3(1). ::
(Induction) Assume that uv^nw is in $D(1)$, for $n \geq 2$.
(Case 1) n is even
(1.1)$|v|$ is even. By Lemma 3(2), $(uv^{\frac{n}{2}}v_1, v_2v^{\frac{n}{2}}w)$ is a n-o.pair, for $v_1, v_2 \in \Sigma^+$, with $v = v_1v_2$ and $|v_1| = |v_2|$. Then $uv^{\frac{n}{2}}v_1v_2v^{\frac{n}{2}}w = uv^{(n+1)}w$ is in $D(1)$ by Lemma 3(1).
(1.2)$|v|$ is odd. $(uv^{\frac{n}{2}}v_1, v_2v^{\frac{n}{2}}w)$ is a n-o.pair, for v_1, v_2, with $v = v_1cv_2$, $|v_1| = |v_2|$ and $c \in \Sigma$. Then $uv^{\frac{n}{2}}v_1cv_2v^{\frac{n}{2}}w = uv^{(n+1)}w$ is in $D(1)$ by Lemma 3(1).
(Case 2) n is odd. By Lemma 3(2), $(uv^{(\frac{n}{2}+1)}, v^{(\frac{n}{2}+1)}w)$ is n-o.pair. Then $uv^{(\frac{n}{2}+1)}v^{(\frac{n}{2}+1)}w = uv^{n+1}w$ is in $D(1)$ by Lemma 3(1). Note that $\frac{n}{2} + \frac{n}{2} = n - 1$ for n odd. ::

Remark 2: Unfortunately, the previous proposition does not hold without the condition $|u| = |w|$. For example, let $u = babaa$, $v = ba$, and $w = a$. Then $uvw = babaabaa \in D(1)$, but $uv^2w = (babaa)^2 \notin D(1)$.

4. d-Primitive Words and D(1)-Concatenated Words

In this section we consider an inclusion relation between $D(1)$ and $D(1)D(1)$.

Lemma 7: Let $zxyx$ be in $D(1)$ for $z, x \in \Sigma^+$, $y \in \Sigma^*$. If z is in $D(1)$, then zx is also in $D(1)$.

[Proof] Suppose that zx is not in $D(1)$. Let $zx = uvu$ for $u \in \Sigma^+$ and $v \in \Sigma^*$.
(Case 1)$|u| \leq |x|$
We can write as $x = x'u$ for some $x' \in \Sigma^*$. Then we have that $zxyx = uvuyx'u$. This contradicts the assumption that $zxyx$ is in $D(1)$.
(Case 2)$|u| > |x|$
We can write as $u = u'x$ for some $u' \in \Sigma^+$. Then we have that $z = u'z'u'$ for some $z' \in \Sigma^*$. This contradicts the assumption that z is in $D(1)$. ::

Proposition 8: Let $|w| \geq 2$ for $w \in \Sigma^+$, with $|\Sigma| = 2$. † If $w \in D(1)$, then w is a $D(1)$-concatenated word. In other words, for a word w in $D(1)$, with the length of two or more, w is in $D(1)D(1)$.

[Proof] Let $\Sigma = \{a, b\}$. It suffices to show the result for the case $zxyx \in a^+\Sigma^*b^+$. From now on, a word $a^{i_1}b^{j_1}\ldots a^{i_k}b^{j_k}$ is denoted by $< i_1, j_1 > \ldots < i_k, j_k >$.

Let $w =< n_1, m_1 > \ldots < n_k, m_k >$ for some $k \geq 1$.

If $k = 1$, that is, $w =< n_1, m_1 >$, it is obvious that w is in $D(1)D(1)$ since both a and b are in $D(1)$, and $< i, j >$ is in $D(1)$ for $i, j \geq 1$.

Let $k \geq 2$. We have that $z =< n_1, m_1 >\in D(1)$. Suppose that $w' =< n_2, m_2 > \ldots < n_k, m_k >\notin D(1)$. Then we can write as $w' = xyx$ for some $x \in \Sigma^+$, and $y \in \Sigma^*$.

There exists an integer $i \geq 2$ such that $w_1 =< n_2, m_2 > \ldots < n_{i-1}, m_{i-1} >< n_i, 0 > <_p x \leq_p < n_2, m_2 > \ldots < n_i, m_i >= w_2$. Note that $w_1 =< n_i, 0 >$ for $i = 2$. By the previous lemma, $zx \in D(1)$. If $x <_p w_2$, then $yx =< 0, m_i - m'_i >< n_{i+1}, m_{i+1} > \ldots < n_k, m_k >$ for some $m'_i \geq 1$. By the lemma again, $< n_1, m_1 > \ldots < n_i, m_i >\in D(1)$. If $< n_{i+1}, m_{i+1} > \ldots < n_k, m_k >\notin D(1)$, then $< n_1, m_1 > \ldots < n_j, m_j >\in D(1)$ for some $i < j < k$, by the lemma. Repeating this process, we have that w is in $D(1)D(1)$ since $< n_k, m_k >$ is in $D(1)$. ::

†The cardinality of a set X is denoted by $|X|$

Concluding Remark 1: As one of our results, we proved that for $n = 1$ and 2, $[D(1)]^n$ is non-regular. As to Q, it is known from [9] that Q^n is regular for $n \geq 2$. The following problem is still unsolved.

"Is $[D(1)]^n$ non-regular for $n \geq 3$?"

References

[1] P. Domosi, S. Holvath, M. Ito, "Formal languages and primitive words," Publ. Math. Debreen, vol.42, nos.3-4, pp.315–321, 1993.
[2] P. Domosi, S. Holvath, M. Ito, L. Kaszonyi, and M. Katsura, "Formal languages consisting of primitive words," Proc. FPCTS93, Lecture Notes in Computer Science, vol.710, pp.194–203, Springer, Berlin, 1993.
[3] P. Domosi, S. Holvath, M. Ito, L. Kaszonyi, and M. Katsura, "Some results on primitive words," Proc. Conf. Semigroupes, Automata and Languages, Proto, 1994.
[4] C-M. Fan, H.J. Shyr, and S.S. Yu, "d-words and d-languages," Acta Informatica, vol.35, pp.709–727, 1998.
[5] S.C. Hsu, M. Ito, and H.J. Shyr, "Some properties of overlapping order and related languages," Soochow Journal of Mathematics, vol.15, pp.29–45, 1998.
[6] Z-Z. Li and Y.S. Tsai, "Three-element codes with one d-primitive word," Acta Informatica, vol.41, pp.171–180, 2004.
[7] M. Lothaire, Combinatorics on words, Addison-Wesley, Reading, MA, 1983.
[8] R.C. Lyndon and M.P. Sh*ü*tzenberger, "The equation $a^M = b^N c^P$ in a free group," Michigan Mathematical J., vol.9, pp.289–298, 1962.
[9] C.M. Reis and H.J. Shr, "Some properties or disjunctive languages on a free monoid," Information and Control, vol.37, no.3, pp.334–344, June 1978.
[10] H.J. Shyr, "Disjunctive languages and codes," Proc. FCT77, Lecture Notes in Computer Science, vol.56, Springer, Berlin, pp.171–176, 1977.
[11] H.J. Shyr, Free Monoids and Languages, Hon Min Book, 2001.