LETTER

On Non-overlapping Words

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

SUMMARY Let Q be the set of all primitive words over a finite alphabet having at least two letters. In this paper, we study the language $D(1)$ of all non-overlapping (d-primitive) words, which is a proper subset of Q. We show that $D(1)$ is a context-sensitive langauage but not a deterministic context-free language. Further it is shown that $[D(1)]^n$ is not regular for $n \geq 1$.

key words: *primitive word, non-overlapping word, d-primitive word, context-sensitive, context-free*

1. Introduction

The notion of primitivity of words plays a central role in algebraic coding theory and combinatorial theory of words (See [8], [10], [11]). Many studies have been done on context-freeness for the language Q of all primitive words, and furthermore context-freeness for some subsets of Q has been shown in [2], [3], [7].

On the other hand, various research efforts have been done on the langaluge $D(1)$ of all non-overlapping words among subsets of Q. (See, [1], [4], [9], [13], for example.) However there are very few results of $D(1)$ with respect to Chomsky-hierarchy except for [9].

In this paper, we discuss $D(1)$ and related lanaguaes with respect to the Chomsky-hierarchy. In Sect. 2, some basic definition and results are presented. In Sect. 3, we show that $D(1)$ is context-sensitive but not deterministic context-free. In Sect. 4, we show that $[D(1)]^n$ is not regular for $n \geq 1$. For the cases of $n = 1$ and $n = 2$, the authors have already proved this in [9].

2. Preliminaries

Let Σ be an 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. A *language* over Σ is a set $L \subseteq \Sigma^*$.

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.*). 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 $u = f^n$, $f \in \Sigma^+$, $n \geq 1$ 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 $x, y \in \Sigma^+$ always implies that $v = \epsilon$. Let $D(1)$ be the set of all non-overlapping words over Σ. A word in $D(1)$ is also called a *d-primitive word* (See [1] and [12]). The comlement of $D(1)$ with respect to Σ^* is denoted by $\overline{D(1)}$, that is $\overline{D(1)} = \Sigma^* - D(1)$. Define $[D(1)]^1 = D(1)$ and $[D(1)]^n = [D(1)]^{n-1}D(1)$ for $n \geq 2$.

Fact 1: Let $a, b \in \Sigma$, and $n \geq 1$.
(1) For $k_1 > k_2 \geq 0$, $w_1 = a^{n+k_1}b^{n+k_2}a^nb^n \in D(1)$.
(2) For $k_2 > k_1 \geq 0$, $w_2 = a^{n-k_1}b^{n-k_2}a^nb^n \in D(1)$.
(3) For $0 < i, j < n$, and $l \geq 0$, $a^{n+l}b^ia^jb^na^nb^n \in D(1)$.
(4) For $0 < i, j < n$, and $k \geq 0$, $a^nb^ia^jb^{n+k}a^nb^n \in D(1)$

3. $D(1)$ and $\overline{D(1)}$ in the Chomsky-hierarchy

In this section, we show that $D(1)$ is context-sensitive but not deterministic context-free.

Proposition 2: $\overline{D(1)}$ is not a context-free language.

Proof.
Suppose that $\overline{D(1)}$ were a context-free language. Then $L = \overline{D(1)} \cap a^+b^+a^+b^+ = \{a^ib^{j+k}a^{i+m}b^j \mid i \geq 1, j \geq 1, k, m \geq 0\}$ would also be a context-free language. We shall prove that L is not context-free using the pumping lemma. (see [5], for example). Let n be an integer sufficient large for L and pick $z = a^nb^na^nb^n$. We may write $z = uvwxy$ subject to the usual constrains $|vwx| \leq n$ and $vx \neq \epsilon$. There are several cases to consider, depending where uvw is in z.
(Case 1) $|uvwx| \leq n$; $vx \in a^+$.
Let $|vx| = k$, where $k > 0$. Then $uv^2wx^2y = a^{n+k}b^na^nb^n$ is in $D(1)$.
(Case 2) $|uvw| < n < |uvwx|$
Set $|v| = l, |x| = k_1 + k_2, x = a^{k_1}b^{k_2}$, where $0 \leq l, k_1, k_2 < n$, and $0 < l + k_1 + k_2 \leq n$.
Since $k_1 \neq 0$, $k_2 \neq 0$, and $k_1, k_2 < n$, we have $uv^2wx^2y = a^{n+l}b^{k_2}a^{k_1}b^na^nb^n$ is in $D(1)$ by Fact1 (3).
(Case 3) $|uv| \leq n \leq |uvw|$
Set $v = a^{k_1}$ and $x = b^{k_2}$, where $k_1, k_2 \geq 0$.
If $k_1 \neq 0$, then $uv^2wx^2y = a^{n+k_1}b^{n+k_2}a^nb^n$ is in $D(1)$ by Fact 1 (1). If $k_2 \neq 0$, then $uwy = a^{n-k_1}b^{n-k_2}a^nb^n$ is in $D(1)$ by Fact 1 (2).
(Case 4) $|u| \leq n < |uv|$
Set $|v| = l_1 + l_2$, $|x| = k$, $v = a^{l_1}b^{l_2}$, where $0 \leq l_1, l_2, k < n$, and $0 < l_1 + l_2 + k \leq n$. If $l_1 \neq 0$, then $uv^2wx^2y =$

Manuscript received July 16, 2010.
Manuscript revised October 28, 2010.
[†]The authors are with the School of Science and Engineering, Kokushikan University, Tokyo, 154–8515 Japan.
a) E-mail: moriya@kokushikan.ac.jp
DOI: 10.1587/transinf.E94.D.707

$a^n b^{l_2} a^{l_1} b^{n+k} a^n b^n \in D(1)$ by Fact 1 (4). If $l_1 = 0$, then $uwx = a^n b^{n-l_2-k} a^n b^n \in D(1)$ by Fact 1 (2).
(Case 5) $n < |u| < 2n$ and $|uv| < 2n$
(5-1) $|uvwx| < 2n$; $vwx \in b^+$ Let $|v| = l$, $|x| = k$ with $0 < l + k \leq n$. $uwy = a^n b^{n-k-l} a^n b^n \in D(1)$ by Fact 1 (2).
(5-2) $|uvwx| > 2n$; $v \in b^+$. Let $|v| = l$.
(5-2-1) $|uvw| < 2n$

Set $|x| = k_1 + k_2$, and $x = b^{k_1} a^{k_2}$, where $k_1, k_2 > 0$. We have $uwy = a^n b^{n-l-k_1} a^{n-k_2} b^n \in D(1)$.
(5-2-2) $|uvw| \geq 2n$

Set $|x| = k$, and $x = a^k$, where $k > 0$. We have $uwy = a^n b^{n-l} a^{n-k} b^n \in D(1)$.
(Case 6) $n < |u| < 2n$ and $|uv| \geq 2n$

Let $|x| = k$, $|v| = l_1 + l_2$, $v = b^{l_1} a^{l_2}$. We have that $uwy = a^n b^{n-l_1} a^{n-l_2-k} b^n$.
(Case 7) $|u| \geq 2n$

The argument is symmetric to the above cases. ::

Corollary 3: $D(1)$ is not deterministic context-free.

Proof. The result holds since the class of deterministic context-free languages is closed under complement (See [5].) ::

Proposition 4: $\overline{D(1)}$ is context-sensitive.

Proof. Let M be a Turing machine with an input tape T_1 and a working tape T_2. M operates the following steps for a given word w on T_1: Let $|w| = n$.
(1) For $k = 1$ to $n/2$, M iterates (1.1)–(1.2).
(1.1) M puts the prefix x and suffix y with $|x| = |y| = k$ of w on T_2
(1.2) M checks whether x is equal to y; if x is equal to y, then M halts in a state "Yes"; otherwise, M turn back to step (1.1) (for the next k) again.
(2) M halts in a state "No"

Since the length of a given word w is finite, one needs only $|w|$ space on both of the tapes. Therefore M is a linear bounded automaton which accepts $\overline{D(1)}$. ::

Corollary 5: $D(1)$ is context-sensitive.

Proof. The result holds since the class of context-sensitive languages is closed under complement. [6]

4. Regularity of $[D(1)]^n$

In this section, we show that $[D(1)]^n$ is not regular for $n \geq 1$. First we present some lemmas necessary for the proof of the aimed proposition as follows.

Lemma 6: Let $m \geq 1$, $x, x' \in \Sigma^+$ with $x' \leq_s x$, $u', u", y \in \Sigma^*$. For any $u \in D(1)$, if $x'(yx)^m = u'uu"$, then $|u| \leq |yx|$.

Proof. Suppose that $|yx| < |u|$.
(Case 1) $u = zxyw$ for some $z, w \in \Sigma^*$, with $|zw| > 0$. Since $z \leq_s y$ and $w \leq_p x$, we have that $u = zwvzw$ for some $v \in \Sigma^*$. Thus $u \notin D(1)$.
(Case 2) $u = z'yxw'$ for some $z', w' \in \Sigma^*$, with $|z'w'| > 0$. By the same argument as in Case 1, $u \notin D(1)$.
(Case 3) $u = y'xy"$ for some $y' \leq_s y$ and some $y" \leq_p y$. Since $|y'y"| > |y|$, we have that $u = y_1 v' y_1$ for some $v' \in \Sigma^*$ and some $y_1 \in Pref(y') \cap Suf(y") - \{\epsilon\}$. Thus $u \notin D(1)$
(Case 4) $u = x'yx"$ for some $x' \leq_s x$ and some $x" \leq_p x$. By the same argument as in Case 3, $u \notin D(1)$. ::

Lemma 7: Let $m \geq 1$, $r \geq 1$, $f, f' \in \Sigma^+$, with $f' \leq_s f$. If $f' f^m = u_1 u_2 \ldots u_r$ with $u_i \in D(1)$, $i = 1, \ldots, r$, then $|u_i| \leq |f|$.

Proof. Immediate by Lemma 6. ::

Lemma 8: Let $m \geq n \geq 1$. For any $f, f' \in \Sigma^+$, with $f' \leq_s f$, $f' f^m \notin [D(1)]^n$.

Proof. Suppose that $f' f^m \in [D(1)]^n$. Then there exist $u_1, \ldots, u_n \in D(1)$ such that $f' f^m = u_1 \ldots u_n$. By Lemma 7, for $1 \leq i \leq n$, $|u_i| \leq |f|$. However, $|u_1 \ldots u_n| \leq n|f| \leq m|f| = |f^m| < |f' f^m|$. This shows that $u_1, \ldots, u_n \in D(1)$ is not true. ::

Lemma 9: For any $f \in D(1), f' \in \Sigma^+$, with $f' \leq_s f$ and $n \geq 1$, $f' f^n \notin [D(1)]^n$.

Proof. Immediate by Lemma 8. ::

Corollary 10: For $k, n \geq 1, m \leq k$, $a^m b^k (a^k b^k)^n \notin [D(1)]^n$.

Proof. Put $f = a^k b^k, f' = a^m b^k$ in Lemma 9. ::

Fact 11: For $k \geq 1$, $a^{k+1} b^k a^k b^k \in D(1)$. ::

Proposition 12: For $n \geq 1$, $[D(1)]^n$ is not regular.

Proof. By Fact 11, $x = (a^{k+1} b^k a^k b^k)(a^k b^k)^{n-1} = a^{k+1} b^k (a^k b^k)^n \in [D(1)]^n$, since $a^k b^k \in D(1)$.

Let k be the integer in the Pumping Lemma.

Then the word x can be written as uvw for some $u, w \in \Sigma^*$, $v \in \Sigma^+$. Since $|uv| \leq k$, uv is in a^+. Moreover, we have that uw is in $[D(1)]^n$ by the pumping lemma. However, since $uw = a^m b^k (a^k b^k)^n$ for some $m \leq k$, uw is not in $[D(1)]^n$ by Corollary 10. This is a contradiction. ::

Concluding Remark 1: We proved that $D(1)$ is context-sensitive but not deterministic context-free. The following problem is still unsolved.
"Is $D(1)$ context-free ?"

References

[1] C.-M. Fan, H.J. Shyr, and S.S. Yu, "D-words and d-languages," Acta Informatica, vol.35, pp.709–727, 1998.
[2] P. Domosi, S. Horvath, and M. Ito, "Formal languages and primitive words," Publ. Math. Debrecen, vol.42, no.3-4, pp.315–321, 1993.
[3] P. Domosi, S. Horvath, M. Ito, L. Kazonyi, and M. Katsura, "Formal languages consisting of primitive words," Lect. Notes Comput. Sci., vol.710, pp.194–203, 1993.
[4] S.C. Hsu, M. Ito, and H.J. Shyr, "Some properties of overlapping order and related languages," Soochow J. Mathematics, vol.15, pp.29–45, 1989.
[5] J.E. Hopcroft, R. Motwani, and J.D. Ullman, Introduction to Automata Theory, Languages, and Computation, Addison Wesley Longman, 2001.
[6] N. Immerman, "Nondeterministic space is closed under complementation," SIAM, J. Comput., vol.17, no.5, pp.935–938, Oct. 1998.
[7] L. Kaszonyi and M. Katsura, "On the context-freeness of class of

primitive words," Publ. Math. Debrecen vol.51, no.1-2, pp.1–11, 1997.

[8] M. Lothaire, Combinatorics on Words, Addison-Wesley, Reading, MA, 1983.

[9] I. Kataoka and T. Moriya, "D-primitive words and D(1)-concatenated words," IEICE Trans. Inf. & Syst., vol.E92-D, no.8, pp.1557–1579, Aug. 2009.

[10] R.C. Lyndon and M.P. Sh*ü*tzenberger, "The equation $a^M = b^N c^P$ in a free group," Michigan Math. J., vol.9, pp.289–298, 1962.

[11] H.J. Shyr, "Disjunctive languages and codes," Proc. FCT77, Lect. Notes Comput. Sci., vol.56, pp.171–176, Springer, Berlin, 1977.

[12] H.J. Shyr, Free monoids and languages, Hon Min Book Company, Taichung, Taiwan, 2001.

[13] Z. Li and Y.S. Tsai, "Three-element codes with one d-primitive word," Acta Informatica, vol.41, no.2/3, pp.171–180, 2004.