Post-plus languages

S.S Yu *

Department of Applied Mathematics, National Chung-Hsing University, Taichung 40227, Taiwan

Received July 1997; revised April 1999
Communicated by M. Ito

Abstract

In this paper, a post-plus language is defined as a language which contains all the catenations of each word in the language with a proper suffix of this word. The set of all d-primitive words is a natural post-plus language. The family of all post-plus languages is a subfamily of all non-counting languages. Some basic properties and characterizations of post-plus languages are investigated. We obtain that a post-plus language spanned by a word over an alphabet with two letters is context-free if and only if the language is regular. Some general properties of post-plus languages related to code, dense property and formal language theory such as the nature of context-free, context-sensitive languages are also studied in this paper. © 2001 Elsevier Science B.V. All rights reserved.

1. Introduction

A language $L$ is a post-plus language if $uv \in L$ implies $uv^2 \in L$ for $u, v \in X^+$. The purpose of this paper is to study the family of post-plus languages. In [12], it is shown that if $u, v \in X^+$ and $uv$ is a d-primitive word, then for any $k \geq 2$, the word $uv^k$ is a d-primitive word. Thus, the set of all d-primitive words is a post-plus language. This indicates that natural post-plus languages exist.

Properties of the set of all primitive words and properties of primitive words are studied frequently for the reason that they are concerned with the basic compositions of words. At the same time, properties of d-primitive words are also investigated in many researches. For instance, d-primitive words are also called non-overlapping words in [12], dipolar words in [13], or unbordered words in [1]. Recently, Shyr and others wrote many papers related to d-primitive words ([5, 13, 15], etc.)

There are many other languages consisting of some kinds of word powers. In [2], power absorbing languages are investigated; right $k$-dense languages are studied in [6];
the $k$-catenation of words is introduced in [9]; left non-counting languages and power-separating languages are studied in [12]; $e$-closed and $c$-closed languages are discussed in [13]; and so on.

There are several reasons that make languages consisting of word powers and the generating operations an important research topic. Firstly, a set of word powers could have some special mathematical properties and could be a useful model for pattern combinations. Secondly, the inserting and deleting procedures could affect the observation and the characterization of languages of word powers. Thirdly, catenation is a basic insertion operation in formal language theory. The $k$-catenation operation is studied in [9]; the insertion and deletion operations are investigated in [8]; $e$-insertion and $c$-deletion operations are studied in [13]; shuffle relations of words are proposed in [14]; etc. In this paper, we focus our investigation on the catenation operation.

In Section 3, we investigate the general properties of post-plus languages. Some special kinds of post-plus languages are given as examples. We show that the post-plus languages $(\mathcal{P}, \subseteq, \cap, \cup)$ forms a distributive lattice. The smallest post-plus language containing the language $L$ is called the post-plus language spanned by $L$. Some basic properties of post-plus languages spanned by languages are studied in Section 3. Section 4 is devoted to the investigation of the decompositions of post-plus languages into disjoint unions of post-plus subsets. This is somewhat similar to the concept of the decomposition of a vector space into a direct sum of subspaces studied in linear algebra [3].

Same as the concept of the irreducible generating set of a semigroup, we define the generating set of a post-plus language $L$ as the set of words in $L$ which are not contained in the post-plus language spanned by any other word in $L$. Some properties of generating sets of post-plus languages are studied in Sections 4 and 5. The concept of generating sets is used in the studying of post-plus languages spanned by words or languages in Sections 5 and 6. A non-empty word is called max-cyclic if it is not contained in the post-plus language spanned by any other non-empty word. The set of all max-cyclic words is exactly the generating set of the largest post-plus language $X^+$. The max-cyclic words are characterized in Section 5. We show that a non-empty word $w$ is max-cyclic if and only if $w \neq uv^2$ for any two non-empty words $u$ and $v$. Thus a word $w \neq v^k$ for any word $v$ and $k \geq 3$ implies that there is a cyclic permutation of $w$ being a max-cyclic word. In Section 6, we also investigate the case that the catenation of a word with a post-plus language forms a post-plus language. Section 7 deals with post-plus languages being suffix codes or dense languages.

According to the Chomsky hierarchy of formal languages, there are regular languages, context-free languages, context-sensitive languages and type-0 languages. In application, when we design a compiler to compile a language, it is very important to know what kind of grammars can be used. In Sections 5 and 6, we investigate the regular, context-free and context-sensitive post-plus languages. For definitions and properties of context-free and context-sensitive languages, the reader is referred to the book [4].
2. Definitions

Let $X$ be a finite alphabet with more than one letter and let $X^*$ be the free monoid generated by $X$. Elements of $X^*$ are called words. Any subset of $X^*$ is said to be a language. For $u \in X^*$ and $L \subseteq X^*$, let $|u|$ denote the length of the word $u$ and let $|L|$ denote the cardinality of the language $L$. Let $X^+ = X^* \setminus \{1\}$, where 1 is the empty word. For $L, L' \subseteq X^*$, let the catenation of $L$ and $L'$ be the set $LL'$ defined by $LL' = \{ uv \mid u \in L, \ v \in L' \}$. Let $L^1 = L$ and $L^n = L^{n-1}L$ for $n \geq 2$.

Recall that a language $L \subseteq X^+$ is a post-plus language if $w \in L$, $u \in X^+$, $v \in X^+$ implies that $uv^2 \in L$. In this case, the empty set and languages $\{a\}$ for $a \in X$ are also considered as post-plus languages. For $u, v \in X^+$, the partial order $\leq_d$ is defined as:

$$v \leq_d u \iff u = xv = vy \text{ for some } x, y \in X^* \quad [12].$$

A word $u \in X^+$ is said to be d-primitive if for $v \in X^*$, $v \neq u$, $v \leq_d u \Rightarrow v = 1$ [12]. Let $D(1)$ be the set of all d-primitive words over $X$. A word $u \in X^+$ is a primitive word if $u = v^n$ for some $v \in X^+$ implies that $n = 1$. Let $Q$ be the set of all primitive words over $X$. It is known that every word $u \in X^+$ is a power of a unique primitive word (see [12]).

Beside the partial order $\leq_d$, we still need the following two partial orders. For $u, v \in X^+$, the partial orders $\leq_p$ and $\leq_s$ are defined as

$$v \leq_p u \iff u = vw \text{ for some } w \in X^*,$$

$$v \leq_s u \iff u = vw \text{ for some } w \in X^*.$$

For a word $u \in X^+$, we define the following sets:

$$\text{Per}(u) = \{ vw \in X^+ \mid u = vw \},$$

$$\text{Pre}(u) = \{ v \in X^+ \mid v \neq u, v \leq_p u \},$$

$$\text{Suf}(u) = \{ v \in X^+ \mid v \neq u, v \leq_s u \}.$$  

For $L \subseteq X^+$, let

$$\text{Suf}(L) = \{ u \in \text{Suf}(w) \mid w \in L \} = \bigcup_{w \in L} \text{Suf}(w).$$

For $u \in X^*$, a word $v \in X^*$ is said to be a prefix (suffix) of $u$ if $v \leq_p u$ ($v \leq_s u$). A suffix $v$ of a word $u$ is called a proper suffix if $v \neq 1$ and $v \neq u$. A non-empty language $L \subseteq X^+$ is called a prefix code (suffix code) if $LX^+ \cap L = \emptyset$. A prefix code (suffix code) $L$ is said to be maximal if $L \cup \{ w \}$ is no more a prefix code (suffix code) for every $w \notin L$.

In this note, we still need the following definition: For $L \subseteq X^*$, the equivalence relation $P_L$ is defined by

$$u \equiv v(P_L) \text{ if and only if } (xuy \in L \iff xyv \in L \text{ for all } x, y \in X^*).$$
We call a language \( L \subseteq X^* \) regular if the index of \( P_L \) is finite.

Items not defined in this section or in the rest of this paper can be found in books [4, 12] which are used as standard references.

3. Some basic properties of post-plus languages

This section is devoted to the investigation of the properties of post-plus languages. Moreover, properties of generating functions, generating sets and the post-plus languages spanned by given languages are studied too. From the definition, the following properties can be obtained immediately.

**Lemma 1.** A language \( L \subseteq X^+ \) is a post-plus language if and only if for \( u, v \in X^+ \), \( uv \in L \) implies \( uv^+ \subseteq L \).

**Proposition 2.** Let \( P \) be the family of all post-plus languages. Then, for \( A, B \in P \), \( A \cup B \in P \) and \( A \cap B \in P \). That is, \( (P, \subseteq, \cap, \cup) \) forms a distributive lattice.

If \( \{L_i \mid i \in A\} \) is a family of post-plus languages, where \( A \) is an index set, then the intersection \( \bigcap_{i \in A} L_i \) is a post-plus language. With this fact we see that for any subset \( L \subseteq X^+ \), the smallest post-plus language containing the set \( L \) exists. Thus for any set \( L \subseteq X^+ \), we let \( \circ(L) \) be the smallest post-plus language containing the set \( L \), called the post-plus language spanned by \( L \). If \( L \) is a singleton set \( \{w\} \), then we call \( \circ(w) \) the post-plus language spanned by the word \( w \). If \( w = a \in X \), then we define \( \circ(w) = \{a\} \).

Note that for any languages \( A \) and \( B \), \( A \subseteq B \) if and only if \( \circ(A) \subseteq \circ(B) \). It is clear that for \( L \subseteq X^+ \), \( \circ(L) = \bigcup_{w \in L} \circ(w) \). It is also clear that \( \circ(\circ(L)) = \circ(L) \), for any \( L \subseteq X^+ \).

These observations yield:

**Remark 3.** Let \( L \subseteq X^+ \). Then the following statements are equivalent:

1. \( L \) is a post-plus language;
2. \( \circ(L) = L \);
3. \( \circ(w) \subseteq L \) for every \( w \in L \).

Next, we investigate the relations between the post-plus language spanned by a word and its subsets, and between two languages and the post-plus languages spanned by them.

**Proposition 4.** Let \( u, w \in X^+ \) and let \( A, B \subseteq X^+ \). Then the following statements hold true:

1. \( u \in \circ(w) \) if and only if \( \circ(u) \subseteq \circ(w) \);
2. If \( A \subseteq B \), then \( \circ(A) \subseteq \circ(B) \);
3. \( \circ(A \cup B) = \circ(A) \cup \circ(B) \);
4. \( \circ(A \cap B) \subseteq \circ(A) \cap \circ(B) \).
Corollary 7. If a language \( L \subseteq X^+ \) is a post-plus language, then \( X^+L \) is also a post-plus language.
Corollary 8. Let $L \subseteq X^+$ and let $m(L) = \min\{|x| \mid x \in L\}$. If $L$ contains the set $X^{m(L)}$, then $X^+L$ is a post-plus language.

Proof. It is obtained from the fact that $\text{Suf}(\circ(L)) \cap X^{m(L)} \subseteq X^{m(L)} \subseteq L$ directly. □

$X^{m(L)}$, given in Corollary 8, is a maximal suffix code for every $L \subseteq X^+$ with $L \neq \emptyset$. The following example shows that there is a maximal suffix code $L$ such that $X^+L$ is not a post-plus language. Let $X = \{a, b\}$ and let $L = \{a, ab, a^2b^2, bab^2, b^3\}$. Then $L$ is a maximal suffix code. Since $a^2b \in X^+L$ and $a^2b^2 \notin X^+L$, $X^+L$ is not a post-plus language.

4. The decomposition of post-plus languages

In this section, we are going to investigate decompositions of post-plus languages into disjoint unions of post-plus languages. First, we decompose the largest post-plus language $X^+$ into disjoint union of two post-plus languages. For $L \subseteq X^+$, the complement $\tilde{L}$ of $L$ is defined by $\tilde{L} = X^+ \setminus L$. To investigate post-plus languages $L$ having post-plus complements $\tilde{L}$, we give the following definition: a post-plus language $L \subseteq X^+$ is full if for any $w \notin L$, $\circ(w) \cap L = \emptyset$. For example, the post-plus language $\circ(ab) = ab^+$ is full while $\circ(ab^2) = ab^+$ is not.

Lemma 9. Let $L \subseteq X^+$ be a post-plus language. Then $\tilde{L}$ is a post-plus language if and only if $L$ is full, where $\tilde{L} = X^+ \setminus L$.

Proof. $\tilde{L}$ is a post-plus language if and only if $\tilde{L} = \bigcup_{w \in L} \circ(w)$. That is, $\circ(w) \cap L = \emptyset$ for every $w \notin L$. Therefore, $\tilde{L}$ is a post-plus language if and only if $L$ is full. □

There are languages which can be expressed as disjoint unions of full languages. For example, let $X = \{a, b\}$. Then the language

$$L = aX^* \cup baX^* \cup b^2X^*$$

is a disjoint union and each component is full. Similarly,

$$L = a^2X^* \cup abX^* \cup baX^* \cup b^2X^*$$

is also a disjoint union of full languages. Here we give another two definitions concerning full languages. Let $L \subseteq X^+$ be a post-plus language. A post-plus subset $A \subseteq L$ is full in $L$ if for any $w \in L \setminus A$, $\circ(w) \cap A = \emptyset$. Similar to Lemma 9, a post-plus subset $A$ is full in a post-plus language $L$ if and only if $L \setminus A$ is a post-plus language. For a language $L \subseteq X^+$, we define the set $G(L)$ as $G(L) = \{w \in L \mid \text{for}\ u \in L, \text{ if } w \in \circ(u), \text{ then } u = w\}$. Then clearly, $G(L) \subseteq L$ and $L \subseteq \circ(G(L))$ for every $L \subseteq X^+$. If $L$ is a post-plus language, then $G(L)$ is said to be a generating set for the post-plus language $L$. Clearly, for a
language $L$, the set $G(L)$ is unique. And, $L = \circ(G(L))$ if and only if $L$ is a post-plus language.

**Lemma 10.** Let $L \subseteq X^+$ be a post-plus language. For any subset $A \subseteq L$, $G(L) \subseteq G(A) \cup G(L \setminus A)$.

**Proof.** Let $w \in G(L)$. Then $w \in A$ or $w \in L \setminus A$. If $w \in A$ and $w \notin G(A)$, then there exists $u \in A \setminus L$ such that $w \in \circ(u)$ and $w \neq u$. This implies that $w \notin G(L)$, a contradiction. Thus if $w \in A$, then $w \notin G(A)$. Similarly, if $w \in L \setminus A$, then $w \in \circ(L \setminus A)$. Therefore, $G(L) \subseteq G(A) \cup G(L \setminus A)$. □

For any post-plus language $L$, there are subsets $A$ of $L$ such that $G(L) \neq G(A) \cup G(L \setminus A)$. For example: let $X = \{a, b\}$, $L = ab^+$ and $A = ab^2$. Thus $G(L) = \{ab\}$, $G(A) = \{ab^2\}$ and $G(L \setminus A) = \{ab\}$. It is clear that $G(L) \neq G(A) \cup G(L \setminus A)$. The following lemma is concerned with the case $G(L) = G(A) \cup G(L \setminus A)$.

**Proposition 11.** Let $L \subseteq X^+$ be a post-plus language and let $A \subseteq L$ be a post-plus subset of $L$. Then the following statements are equivalent:

1. $A$ is full in $L$;
2. $G(L) = G(A) \cup G(L \setminus A)$ and $\circ(G(A)) \cap \circ(G(L \setminus A)) = \emptyset$.

**Proof.** First, $A$ is full in $L$ if and only if $A$ and $L \setminus A$ are post-plus languages. That is, $A$ is full if and only if for every $x \in L \setminus A$ and $y \in A$, $\circ(x) \cap A = \emptyset$ and $\circ(y) \cap (L \setminus A) = \emptyset$.

(1) $\Rightarrow$ (2): Let $A$ be full in $L$. Consider $w \in G(A)$. If $w \notin G(L)$, then there is $u \in L$ such that $w \in \circ(u)$ and $w \neq u$. By the definition of $G(A)$, $u \notin A$. Thus $u \in L \setminus A$ and $w \in \circ(u) \cap A$, a contradiction. Therefore, $w \in G(A)$ implies that $w \in G(L)$. That is, $G(A) \subseteq G(L)$. Similarly, $G(L \setminus A) \subseteq G(L)$. By Lemma 10, $G(L) \subseteq G(A) \cup G(L \setminus A)$. Thus $G(L) = G(A) \cup G(L \setminus A)$. Now, assume that $w \in \circ(G(A)) \cap \circ(G(L \setminus A))$. Then there exist $u \in G(A)$ and $v \in G(L \setminus A)$ such that $w \in \circ(u) \cap \circ(v)$. If $w \in A$, then $\circ(v) \notin L \setminus A$ and $L \setminus A$ is not a post-plus language, a contradiction. Similarly, $w \in L \setminus A$ implies that $A$ is not a post-plus language, a contradiction! Thus $\circ(G(A)) \cap \circ(G(L \setminus A)) = \emptyset$.

(2) $\Rightarrow$ (1): Let $G(L) = G(A) \cup G(L \setminus A)$ and $\circ(G(A)) \cap \circ(G(L \setminus A)) = \emptyset$. We are going to show that $L \setminus A$ is a post-plus language. Let $u \in G(L \setminus A)$. Then $u \notin A$ and $u \in G(L)$. If there exists $w \in \circ(u) \cap A$, then there is $v \in G(A)$ such that $w \in \circ(v)$. This implies that $w \in \circ(G(A)) \cap \circ(G(L \setminus A))$, a contradiction. Thus $\circ(u) \cap A = \emptyset$ and $\circ(u) \subseteq L \setminus A$. Therefore, $\circ(G(L \setminus A)) \subseteq (L \setminus A)$ and then, $L \setminus A$ is a post-plus language. □

Now, it suffices to show some properties of the post-plus languages which can be decomposed into disjoint unions of post-plus languages.

**Corollary 12.** Let $L \subseteq X^+$ be a post-plus language and let $L = \bigcup_{i \in A} A_i$ be a disjoint union. Then, $A_i$ is a post-plus language for every $i \in A$ if and only if $A_i$ is full in $L$ for every $i \in A$. 
Proof. Since $L$ is a post-plus language, $\circ(L) = L$. If for every $i \in A$, $A_i$ is full in $L$, then clearly, $A_i$ is a post-plus language. Now let every $A_i$ be a post-plus language, that is, $\circ(A_i) = A_i$. Since $A_i = \circ(A_i)$ and $A_i \cap (L \setminus A_i) = \emptyset$, $\circ(L \setminus A_i) = \bigcup_{j \in A, j \neq i} \circ(A_j) = \bigcup_{j \in A, j \neq i} A_j = L \setminus A_i$. Thus $L \setminus A_i$ is a post-plus language and hence, $A_i$ is full in $L$ for every $i \in A$. □

Corollary 13. Let $A \subseteq X^+$ be a prefix code and let $L \subseteq X^*$. If $AL$ is a post-plus language, then $AL$ is a disjoint union of post-plus subsets $wL$, where $w \in A$.

Proof. Let $A \subseteq X^+$ be a prefix code. Clearly, $AL = \bigcup_{w \in A} wL$ is a disjoint union. Let $AL$ be a post-plus language. Then $AL = \circ(AL)$. For every $w \in A$, we want to show that $wL$ is a post-plus language. For $u \in L$, if $\circ(wu) \cap (AL \setminus wL) \neq \emptyset$, then there exist $x \in A \setminus \{w\}$ and $v, y \in L$ such that $wv = xy \in \circ(wu) \cap (AL \setminus wL)$. Thus $x \leq_p w$ or $w \leq_p x$ and $x \neq w$. This implies that $A$ is not a prefix code, a contradiction. Hence, $\circ(wL) \cap (AL \setminus wL) = \emptyset$. As $\circ(wL) \subseteq \circ(AL) = AL$, $\circ(wL) \subseteq wL$. Since $wL \subseteq \circ(wL)$, $\circ(wL) = wL$. Therefore, $wL$ is a post-plus language. □

By Corollaries 12 and 13, a prefix code $A$ such that $AL$ is a post-plus language for a language $L \subseteq X^*$ if and only if $wL$ is full in $AL$ for every $w \in A$. For a post-plus language $L$, the following proposition shows that the decomposibility of $L$ into a disjoint union of post-plus languages is related to the set $G(L)$.

Proposition 14. Let $L \subseteq X^+$ be a post-plus language and let $G(L)$ be finite. Then $L$ can be expressed as a disjoint union of at most $|G(L)|$ post-plus subsets.

Proof. Let $L \subseteq X^+$ be a post-plus language. Then $\circ(G(L)) = L$. Assume that $L$ can be expressed as a disjoint union $L = \bigcup_{i=1}^m A_i$ of $m$ post-plus languages $A_i$ with $m > |G(L)|$. Then by Corollary 12, $\circ(A_i) = A_i$ for every $1 \leq i \leq m$. By $m > |G(L)|$ and the equation $L = \circ(G(L)) = \bigcup_{w \in G(L)} \circ(w) = \bigcup_{i=1}^m A_i$, there exist $w \in G(L)$ and $1 \leq i \neq j \leq m$ such that $w \in A_i$ and $\circ(w) \cap A_j \neq \emptyset$. Thus, there exists $u \in \circ(w) \cap A_j \subseteq \circ(A_i) \cap \circ(A_j) = A_i \cap A_j$. This contradicts that $A_i$ and $A_j$ are disjoint for $i \neq j$. □

Consider a post-plus language $L$. If $\circ(w) \cap \circ(G(L) \setminus \{w\}) = \emptyset$ for every $w \in G(L)$, then by Corollary 12, $L = \bigcup_{w \in G(L)} \circ(w)$ is a disjoint union of $|G(L)|$ post-plus languages. Now, let $X = \{a, b\}$ and let $L = \circ(aba) \cup \circ(abab^2ba)$. Then $G(L) = \{aba, abab^2ba\}$ and $abab^2ba \in \circ(aba) \cap \circ(abab^2ba)$ from the proof of Proposition 14, it follows that $L$ cannot be decomposed as a disjoint union of two or more post-plus languages.

5. Post-plus languages spanned by a word

Words $w \in X^+$ such that $\circ(w)$ are regular or context-free are investigated in this section. Languages $\circ(ab) = ab^+$ and $\circ(ab^n) = ab^nb^+$ for some $n \geq 2$ are examples of post-plus languages being regular. The case of post-plus languages $\circ(w)$ being full are
Lemma 15. Let $a \neq b \in X$ and let $w = b^{m_0}a^{n_1}b^{m_1}a^{n_2}b^{m_2} \cdots a^{n_r}b^{m_r}$ where $m_0 \geq 0$, $n_i, m_i \geq 1, 1 \leq i \leq r$ for some $r \geq 1$. Let $m \geq |aw|$. Then

1. $awab^{k}b^{m-s_1-m}wab^{k}ab^{k_2} \notin (aw)$ for every $0 < s_1 \leq m$ and $k, k_2 \geq 0$;

2. $awab^{k}b^{m-s_0-m}wab^{k}ab^{m-s_2-m} \notin (aw)$ for $s_0, s_1, s_2 \geq 0$ with $s_0 \leq m$ and $0 < s_1 + s_2 \leq m$;

3. $awab^{k}b^{m-s_2-m}wab^{k}ab^{m} \notin (aw)$ for $s_1, s_2 \geq 0$ with $0 < s_1 + s_2 \leq m$.

Proof. (1) Let $w_1 = awab^{k}b^{m-s_1-m}wab^{k}ab^{k_2}$ for $0 < s_1 \leq m$ and $k, k_2 \geq 0$. Suppose $w_1 \notin (aw)$. Then there exist $u_1v_1, u_2v_2, \ldots, u_nv_n \in (aw)$ such that $u_1v_1 = aw$, $u_i v_i = u_{i-1}v_i^2$ for $i = 2, 3, \ldots, n$ and $u_nv_n = w_1$. From the definition of post-plus languages and Remark 3, it follows that $u_nv_n \in (u(v_i))$, $1 \leq i \leq n$. Since $4m \geq |aw|$ and $4m - s_1 - m_0 - m_r \geq 2m \geq |wa|$, there exists $1 \leq i < n$ such that $u_i v_i = awab^{j_i}$ and $u_i v_i^2 = awab^{k_i}ab^{k_i}$ for some $j_i > 0$. By the observation of $u_i v_i = awab^{j_i}ab^{m_0}a^{n_i}b^{m_1}a^{n_2}b^{m_2} \cdots a^{n_r}b^{m_r}$, one must have that $j_i \geq 4m - m_r$. Since $u_nv_n \in (u_i v_i^2)$,

\[ u_nv_n = awab^{k}b^{m-s_1-m}wab^{k}ab^{k_2} = awab^{k}ab^{j_i}wab^{k}ab^{k_i} \]

for some $j_i \geq j_1 = 4m - m_r$. This implies that $4m - s_1 - m_0 - m_r > 4m - s_1 - m_0 - m_r$ and $k_3, k_4 \geq 0$. This implies that $4m - s_1 - m_0 - m_r \geq 4m - s_1 - m_0 - m_r$ with $s_1 > 0$, a contradiction! Therefore, $w_1 \notin (aw)$.

(2) Let $w_2 = awab^{k}b^{m-s_0-m}wab^{k}ab^{m-s_2-m}$ for $s_0, s_1, s_2 \geq 0$ with $0 < s_1 + s_2 \leq m$. By (1), $s_0 = 0$. Suppose $w_2 \notin (aw)$. Then same as (1), there exist $u_1v_1, u_2v_2, \ldots, u_nv_n \in (aw)$ such that $u_1v_1 = aw$, $u_i v_i = u_{i-1}v_i^2$ for $i = 2, 3, \ldots, n$ and $u_nv_n = w_2$. And, there exists $1 \leq i < n$ such that $u_i v_i = awab^{j_i}$ and $u_i v_i^2 = awab^{k_i}ab^{k_i}$ with $j_1 \geq 4m - m_r$. As $u_nv_n \in (u_i v_i^2)$, by the observation of $u_i v_i = awab^{k}b^{m-s_1-m}wab^{k}ab^{m-s_2-m}$, $j_i = 4m - m_r$, and $u_i v_i = ab^{k_3}ab^{k_4}ab^{k_5}$ with $j_1 \geq 4m - m_r$. Thus, $u_i v_i = awab^{k}b^{m-s_1-m}wab^{k}ab^{m-s_2-m}$.

\[ u_nv_n = awab^{k}b^{m-s_0-m}wab^{k}ab^{m-s_2-m} = awab^{k}ab^{j_i}wab^{k}ab^{k_i} \]

for some $j_i \geq j_1 = 4m - m_r$. This implies that $4m - s_1 = 4m$ and $j_2 = 4m - s_2 - m_r$. By $s_1 + s_2 > 0$, either $4m - s_1 \neq 4m$ if $s_1 > 0$ or $j_2 \geq 4m - m_r > 4m - s_2 - m_r$ if $s_2 > 0$, a contradiction! Therefore, $w_2 \notin (aw)$.

(3) Let $w_3 = awab^{k}b^{m-s_0-m}wab^{k}ab^{m-s_2-m}$ for $s_1, s_2 \geq 0$ with $0 < s_1 + s_2 \leq m$. By (1), $s_1 > 0$. Suppose $w_3 \notin (aw)$. Then same as (1), there exist $u_1v_1, u_2v_2, \ldots, u_nv_n \in (aw)$ such that $u_1v_1 = aw$, $u_i v_i = u_{i-1}v_i^2$ for $i = 2, 3, \ldots, n$ and $u_nv_n = w_3$. Since $4m - s_1 \geq 3m$ and $4m - s_2 \geq 3m$, there exists $1 \leq i < n$ such that $u_i v_i = awab^{j_i}$ and $u_i v_i^2 = awab^{k_i}ab^{k_i}$ with $j_1 \geq 4m - s_1 - m_r$. As $j_1 \geq 4m - s_1 - m_r > 2m \geq |wa|$, there exists $1 \leq i \leq n$ such that $u_i v_i = awab^{m-s_1}ab^{j_i}$ for some $j_i \geq j_1$ and $|u_i v_i^2| > |awab^{m-s_1}ab^{m-s_2-m}|$. Thus $v_i \in X^*aX^*$. From $v_i \leq s, awab^{m-s_1}ab^{j_i}$, it follows that $|v_i| \geq |ab^{j_i}|$.
and hence, \( |u_i v_i^2| > |awb^{4m-s_1}ab^{4m-s_2-m_0-m}wa| \). By the observation of \( u_i v_i = awb^{4m-s_1}ab^{4m-s_2,j} \), \( v_i = b^{4m-s_1-m-r}a^{n_2}b^{m_2} \ldots a^{m_r} b^{m_r} \). That is,

\[
u_i v_i^2 = awb^{4m-s_1}ab^{4m-s_2-m_0-m}wab^{4m-s_1}ab^{j,3}\]

Since \( u_n v_n \in \circ (u_i v_i^2) \),

\[
u_n v_n = awb^{4m-s_1}ab^{4m-s_2-m_0-m}wab^{4m-s_1}ab^{4m-s_1}ab^{j,3}\]

for some \( j_3 \geq j_2 \). This implies that \( 4m - s_1 = 4m \) with \( s_1 > 0 \), a contradiction! Therefore, \( w_j \in \circ (aw) \).

For \( u \in X^+ \), if \( w \in X^+uX^+ \), then \( u \) is said to be an infix of \( w \). From Lemma 15, we now investigate the word \( w \in \{a,b\}^+ \) such that the post-plus language \( \circ (w) \) spanned by \( w \) is not context-free.

**Proposition 16.** For any \( w \in a\{a,b\}^+ \setminus (a^+ \cup ab^+) \), \( \circ (w) \) is not a context-free language.

**Proof.** Suppose \( w = ab^{m_0}a^{n_1}b^{m_1}a^{n_2}b^{m_2} \ldots a^{n_r} b^{m_r} \) for \( m_0 \geq 0 \) and \( n_i, m_i \geq 1 \), \( 1 \leq i \leq r \). Let \( w' = b^{m_0}a^{n_1}b^{m_1} \ldots a^{n_r} b^{m_r} \). Let \( L = \circ (w) \cap wab^+ ab^+ w' ab^+ ab^+ \). Then \( L \) is a context-free language if \( \circ (w) \) is a context-free language. For any \( k \geq 1 \), let \( m \geq k |w| \). Consider \( w_1 = wab^{k}ab^{4m-m_0-m}w' ab^{m_1}ab^{4m-m_r} \). Then \( w_1 \in L \). Let \( w_1 = u v x w y z \) with \( |vxy| \leq k, |v| \leq m \) and \( |v| \leq k \), \( |y| \leq m \). Then \( \circ (w) \) is a context-free language. For any \( k \geq 1 \), let \( m \geq k |w| \). Consider \( w_1 = wab^{k}ab^{4m-m_0-m}w' ab^{m_1}ab^{4m-m_r} \). Then \( w_1 \in L \). Let \( w_1 = u v x w y z \) with \( |vxy| \leq k, |v| \leq m \) and \( |v| \leq k \), \( |y| \leq m \). Then \( \circ (w) \) is a context-free language. For any \( k \geq 1 \), let \( m \geq k |w| \). Consider \( w_1 = wab^{k}ab^{4m-m_0-m}w' ab^{m_1}ab^{4m-m_r} \). Then \( w_1 \in L \). Let \( w_1 = u v x w y z \) with \( |vxy| \leq k, |v| \leq m \) and \( |v| \leq k \), \( |y| \leq m \). Then \( \circ (w) \) is a context-free language. For any \( k \geq 1 \), let \( m \geq k |w| \). Consider \( w_1 = wab^{k}ab^{4m-m_0-m}w' ab^{m_1}ab^{4m-m_r} \). Then \( w_1 \in L \). Let \( w_1 = u v x w y z \) with \( |vxy| \leq k, |v| \leq m \) and \( |v| \leq k \), \( |y| \leq m \). Then \( \circ (w) \) is a context-free language.

By Lemma 15, \( u x z \notin \circ (w) \). From the Pumping lemma of context-free languages, \( L \) is not context-free. Thus \( \circ (w) \) is not context-free.

If \( w \) is obtained by exchanging \( a \) and \( b \) of \( w' \) mentioned above, then we have \( w = a a^{m_0}b^{m_1}a^{n_2}b^{m_2} \ldots b^{m_r} a^{m_r} \), where \( m_0 \geq 0 \) and \( n_i, m_i \geq 1 \), \( 1 \leq i \leq r \). Since the above proof is independent from the first letter used in \( w \), by an analogous argument, \( \circ (w) \) is not context-free.

By exchanging \( a \) and \( b \), it is also true that \( \circ (w) \) is not a context-free language for every \( w \in b\{a,b\}^+ \setminus (b^+ \cup ba^+) \). This yields the following theorem.

**Theorem 17.** Let \( w \in \{a,b\}^+ \). Then the following are equivalent:

1. \( \circ (w) \) is context-free;
2. \( w \in (a^+ \cup ab^+ \cup b^+ \cup ba^+) \);
3. \( \circ (w) \) is regular.
Corollary 19. If the word $w$ is also a regular post-plus language. Then $w = w^2$ for any $u, v \in X^+$. 

Proof. (1) \Rightarrow (2): Let $\omega(w)$ be context-free. By Proposition 16, if $w \notin (a^+ \cup ab^+ \cup b^+ \cup ba^+)$, then $\omega(w)$ is not a context-free language, a contradiction. Thus $w \in (a^+ \cup ab^+ \cup b^+ \cup ba^+)$. 

(2) \Rightarrow (3): Let $w \in (a^+ \cup ab^+ \cup b^+ \cup ba^+)$. Then $\omega(w)$ must be one of $\{a\}$, $\{b\}$, $a^i a^j$, $b^i b^j$, $a^i b^j$ or $b^i a^j$, where $i, j \geq 0$. Thus $\omega(w)$ is a regular language. 

(3) \Rightarrow (1): Immediate. 

There are many other regular post-plus languages if we consider the post-plus languages spanned by languages. For example, consider the language $G +$. According to Proposition 4, $G +$ is also maximal. From the definition of $G(X^+)$, one must have that $G(X^+)$ is the set consisting of max-cyclic words in $X^+$. 

Proposition 18. Let $w \in X^+$. Then $w$ is max-cyclic if and only if $w \neq w^2$ for any $u, v \in X^+$. 

Proof. ($\Rightarrow$) If $w = uv^2$ for some $u, v \in X^+$, then $w = w^2 \in \omega(w)$. By (1) of Proposition 4, $\omega(w) \subseteq \omega(uv)$. We have then $\omega(w)$ is not maximal, a contradiction. 

($\Leftarrow$) Assume that $w \neq w^2$ for any $u, v \in X^+$. Suppose $\omega(w)$ is not maximal and $\omega(w) \subseteq \omega(z)$ for some $z \in X^+$, $z \neq w$. Then $w \in \omega(z)$ and by definition of the set $\omega(z)$, $w = w^2$ for some $w \in \omega(z)$ with $u, v \in X^+$. This contradicts our assumption. 

A word $w \in X^+$ is called square-free if $w \in X^+u^2X^+$ for every $u \in X^+$ [12], and is called post-square free if there exist no $u, v \in X^+$ such that $w = uv^2$. It is clear that every square-free word in $X^+$ is a post-square free word. Proposition 18 means that a word $w \in X^+$ is max-cyclic if and only if $w$ is post-square free. This yields: 

Corollary 19. If the word $w \in X^+$ is a square-free word, then $w$ is a max-cyclic word. 

In the following, we investigate properties of the set $G(X^+)$ of all max-cyclic words over $X$. 

Proposition 20. Let $u \in X^+$ and let $a \neq b \in X$. Then for every $k \geq |u|$, $ub^k ab \in G(X^+)$. 

Proof. Suppose $ub^k ab \notin G(X^+)$ for some $k \geq |u|$. Then by Proposition 18, $ub^k ab = u_1 u_2^2$ for some $u_1, u_2 \in X^+$. Since $a \neq b$, $|u_2| \geq |ab|$. As $k \geq |u|$, $|ub^k ab| \geq 2|u| + 2$. By $u_1 \in X^+$, $|u_1 u_2| > |u| + 1$. Then $u_1 u_2 = ub^i$ for some $i \geq 2$. This implies that $bb \leq u_2$ and $ab \leq u_2$. Thus $a = b$, a contradiction! Therefore, $ub^k ab \in G(X^+)$ for every $k \geq |u|$. 


By a dense language \( L \) we mean that \( L \subseteq X^* \) and for any \( u \in X^+ \), there exist two words \( x \) and \( y \) in \( X^* \) such that \( xuv \in L \), that is, \( L \cap X^*uX^+ \neq \emptyset \) for every \( u \in X^+ \). Proposition 20 yields the following result directly.

**Corollary 21.** \( G(X^+) \) is dense.

Next, we characterize the max-cyclic words being primitive.

**Proposition 22.** \( G(X^+) \cap Q = \text{Suf}(G(X^+)) \).

**Proof.** For \( v \in \text{Suf}(G(X^+)) \), there exists \( u \in X^+ \) such that \( uv \in G(X^+) \). By Proposition 18, \( uv \) is post-square free; and then, \( v \) is a primitive post-square free word. Again by Proposition 18, \( v \in G(X^+) \). Thus \( \text{Suf}(G(X^+)) \subseteq G(X^+) \cap Q \). Now, we show that \( G(X^+) \cap Q \subseteq \text{Suf}(G(X^+)) \). Let \( v \in G(X^+) \cap Q \). Suppose \( av \notin G(X^+) \) for some \( a \in X \). Then, by Proposition 18, \( av \) is not a post-square free word. Since \( v \) is a post-square free word, the only case which can hold is that \( v = p^2 \) for some \( p \in Q \). Thus \( v \notin Q \), a contradiction! Therefore, \( av \in G(X^+) \), i.e., \( v \in \text{Suf}(G(X^+)) \). \( \square \)

A word \( w \in X^+ \) being a max-cyclic word does not imply that \( \circ(w) \) is full in \( X^+ \). For example, let \( a \neq b \in X \). The word \( aba \) is a max-cyclic word while \( X^+ \setminus \circ(w) \), the complement of \( \circ(aba) \), is not a post-plus language. This can be seen from the fact that the word \( aba^2ba^2 \) is in \( X^+ \setminus \circ(aba) \), while \( aba^2ba^2 \) is in \( \circ(aba) \). In the following, we investigate the case of intersections of post-plus languages spanned by distinct words in \( X^+ \) being empty or infinite.

**Proposition 23.** For any distinct words \( w, v \in X^+ \), either \( \circ(w) \cap \circ(v) = \emptyset \) or \( |\circ(w) \cap \circ(v)| = \infty \).

**Proof.** Suppose \( \circ(w) \cap \circ(v) \neq \emptyset \) for some \( w \neq v \in X^+ \). Let \( u \in \circ(w) \cap \circ(v) \). Then \( u \in \circ(w) \) and \( u \in \circ(v) \). By (1) of Proposition 4, \( \circ(u) \subseteq \circ(w) \) and \( \circ(u) \subseteq \circ(v) \). Since \( w \neq v \), \( u \notin X \). Thus \( \circ(u) \) is infinite. This shows that \( \circ(w) \cap \circ(v) \neq \emptyset \) implies \( |\circ(w) \cap \circ(v)| = \infty \). \( \square \)

**Proposition 24.** For \( u, v \in X^+ \), if \( \circ(u) \cap \circ(v) \neq \emptyset \), then \( u \leq_p v \) or \( v \leq_p u \).

**Proof.** Suppose \( \circ(u) \cap \circ(v) \neq \emptyset \). Then there exists \( w \in \circ(u) \cap \circ(v) \). It is clear that \( u \leq_p w \) and \( v \leq_p w \). Therefore, \( u \leq_p v \) or \( v \leq_p u \). \( \square \)

Now, we consider words \( w \) such that the post-plus languages \( \circ(w) \) are full.

**Lemma 25.** If \( w \in X \cup X^2 \), then \( \circ(w) \) is full.
Proof. For $w \in X \cup X^2$, clearly, $\circ(w)$ is a post-plus language. Since $\circ(a) = \{a\}$ for every $a \in X$, $\circ(a) = X^+ \setminus \{a\} = \circ(X^+ \setminus \{a\})$ is a post-plus language. Thus $\circ(a)$ is full for every $a \in X$.

Now, let $w \in X^2$. Then $w = a_1a_2$ for some $a_1, a_2 \in X$. It follows that $\circ(w) = a_1a_2^+$. If $u \in \circ(w)$, then $u = a_3u_1$ or $u = a_1u_2$, where $a_3 \neq a_1 \in X$, $u_1 \in X^+$ and $u_2 \in X^+ \setminus a_2^+$. Clearly, $\circ(u) \cap \circ(w) = \circ(u) \cap a_1a_2^+ = \emptyset$. Thus, $\circ(\circ(w)) \cap \circ(w) = \bigcup_{u \in \circ(w)} \circ(u) \cap \circ(w) = \emptyset$. That is, $\circ(\circ(w)) \subseteq \circ(w)$. As $\circ(w) \subseteq \circ(\circ(w))$, one must have that $\circ(\circ(w)) = \circ(w)$. Therefore, $\circ(w)$ is a post-plus language too. □

Let $a \neq b \in X$ and let $u = aaba$. Then $u \in G(X)$. But by Proposition 18, $u(ba)^3aba = u(ba)^2b(aba)^2 \notin G(X^+)$. In the following proposition, we consider the case of distinct max-cyclic words $u, v$ being such that $\circ(v) \cap \circ(u) \neq \emptyset$.

Proposition 26. Let $u = u_1u_2u_3 \in G(X^+)$ for some $u_1, u_2, u_3 \in X^+$ and let $v = uu_1^k u_2 u_3 \in G(X^+)$ for some $k \geq 1$. Then $v \notin \circ(u)$ and $\circ(v) \cap \circ(u) \neq \emptyset$.

Proof. Let $u = u_1u_2u_3 \in G(X^+)$ for some $u_1, u_2, u_3 \in X^+$ and let $v = uu_1^k u_2 u_3 \in G(X^+)$ for some $k \geq 1$. Since $v \neq u$ and $u, v \in G(X^+)$, by the definition of $G(X^+)$, $v \notin \circ(u)$. As $uu_1^k u_2 u_3 \in \circ(v) \cap \circ(u)$, $\circ(v) \cap \circ(u) \neq \emptyset$. □

From Propositions 20 and 26, we have the following corollary:

Corollary 27. For every $u \in G(X^+)/(X \cup X^2)$, there are infinitely many $v \in G(X^+)$ such that $\circ(v) \cap \circ(u) \neq \emptyset$.

Proof. Let $u \in G(X^+)/(X \cup X^2)$. Then by Proposition 18, $u$ is post-square free. Thus there are $a \neq b \in X$ such that $u = u'ab$ for some $u' \in X^+$. By Proposition 20, $ub^k ab \in G(X^+)$ for every $k \geq |u|$. From Proposition 26, it follows that $\circ(ub^k ab) \cap \circ(u) \neq \emptyset$ for every $k \geq |u|$. □

There are many other cases except the case proposed in Proposition 26 such that $\circ(u) \cap \circ(v) \neq \emptyset$ for two words $u, v \in G(X^+)$. For example, let $X = \{a,b,c\}$, let $u = abcaaca$ and let $v = abcaacaaba; u, v \in G(X^+)$. Then $a(bca^2ca^3)^2 \in \circ(u) \cap \circ(v)$. Let $w = aabca$. Then $bca \leq_s w$ and $bca \notin_s v$. Clearly, $v \neq uu_1^k u_2 u_3$ for every $u_2, u_3 \in X^+$ and $k \geq 1$ with $u = u_1 u_2 u_3$ for some $u_1 \in X^+$.

Theorem 28. Let $u \in X^+$. Then $\circ(u)$ is full if and only if $|u| \leq 2$.

Proof. If $u \notin G(X^+)$, then there exists $w \in G(X^+)$ such that $\circ(u) \subseteq \circ(w)$. By the definition of $G(X^+)$, $w \neq u$ and $w \in \circ(u)$. This implies that $\circ(u) \cap \circ(\circ(u)) \neq \emptyset$. Thus $\circ(\circ(u)) \neq \circ(u)$ and $\circ(u)$ is not full. For $u \in G(X^+)$, if $u \notin X \cup X^2$, then, by Corollary 27, $\circ(u)$ is not full. Thus $\circ(u)$ is full implies that $|u| \leq 2$.

Conversely, if $u \in X^+$ with $|u| \leq 2$, then, by Lemma 25, $\circ(u)$ is full. □
6. Some more properties of post-plus languages

By Theorem 17, a post-plus language spanned by a word is regular if and only if it is context-free. This is not true for the post-plus languages spanned by a language which is not a singleton set. For example, the language $L = \{a^n b^n a \mid n \geq 1\} X^*$ where $a \neq b \in X$ is a context-free post-plus language which is not a regular language. Post-plus languages being context-sensitive and catenations of words with post-plus languages spanned by words will be investigated in this section. To progress our investigation step by step, we give the following definition.

For a language $L \subseteq X^+$, the post-power set $P(L)$ of $L$ is defined to be $P(L) = \{w^2 \mid w \in L, u \in X^+, v \in X^+\}$. If $L = \emptyset$ or $L \subseteq X$, then let $P(L) = \emptyset$. Here, we also consider $P(L)$ as the set obtained by applying an operation $P$ on the language $L$. Let $P^0(L) = L$, $P^1(L) = P(L)$, and let $P^n(L) = (P^n-1(L))$ for $n \geq 2$. By the definition of post-plus languages, $L \subseteq X^+$ is a post-plus language if and only if $P(L) \subseteq L$. For $w \in X^+$ and $L = \{w\}$, let $P_m(w)$ denote $P_m(L)$ for $m \geq 0$. For example, $P(aba) = \{abaa, ababa\}$. Let $L = ab^+ \subseteq \{a, b\}^+$. Since $P(L) \subseteq L$, $L$ is a post-plus language. Usually $P_m(L) \cap P_n(L) \neq \emptyset$ for $m \neq n$. For example, $aba^3 \in P^3(ab^3) \cap P^2(ab^3)$.

This operation $P$ is an insertion generating function defined on languages and is similar to the generating function for L-languages (see [11]). Both of them operate directly on some given words and all the generated words are in the language generated.

**Lemma 29.** Let $L, L' \subseteq X^+$. Then the following statements hold true for every $k \geq 1$:

1. If $L \subseteq L'$, then $P^k(L) \subseteq P^k(L')$.
2. $P^k(L \cup L') = P^k(L) \cup P^k(L')$.
3. $P^k(L \cap L') \subseteq P^k(L) \cap P^k(L')$.

**Proof.** Let $L, L' \subseteq X^+$. Consider the following three steps:

(a) Assume that $L \subseteq L'$. If $L = \emptyset$, then $P(L) = \emptyset$. Clearly, $P(L) \subseteq P(L')$. Now, let $L \neq \emptyset$. Then, for every $w \in P(L)$, there must exist $u, v \in X^+$ such that $uv \in L$ and $w^2 = w$. As $L \subseteq L'$, $w \in L'$ and then $w^2 \in P(L')$. Statement (1) holds true when $k = 1$.

(b) Next, as $L \subseteq L \cup L'$ and $L' \subseteq L \cup L'$, by step (a), $P(L) \subseteq P(L \cup L')$ and $P(L') \subseteq P(L \cup L')$, that is, $P(L) \cup P(L') \subseteq P(L \cup L')$. If $w \in P(L \cup L')$, there must exist $u, v \in X^+$ such that $uv \in L \cup L'$ and $w^2 = w$. Thus $w \in L$ or $w \in L'$. This implies that $w = w^2 \in P(L)$ or $w^2 \in P(L')$. Hence, $P(L \cup L') \subseteq P(L) \cup P(L')$. Statement (2) holds true for $k = 1$.

(c) As $L \cap L' \subseteq L$ and $L \cap L' \subseteq L'$, by statement (1), $P(L \cap L') \subseteq P(L)$ and $P(L \cap L') \subseteq P(L')$. Thus $P(L \cap L') \subseteq P(L) \cap P(L')$. Statement (3) holds true when $k = 1$.

By applying the operation $P$ to $L$ $k$ times, from the definition of $P^k$, it follows that statements (1)–(3) hold true for every $k \geq 1$. □

There exist languages $L, L' \subseteq X^+$ such that $P(L \cap L') \neq P(L) \cap P(L')$. For example, let $L = \{ab^3\}$ and let $L' = \{ab^4\}$. Then $L \cap L' = \emptyset$. But $ab^5 \in P(L) \cap P(L')$. 


The following proposition concerns the relations between the operation $P$ and post-plus languages.

**Proposition 30.** For a language $L \subseteq X^+$, the following two statements hold true:

1. $P(L) \subseteq L$ if and only if $\circ(L) = L$;
2. $\circ(L) = \bigcup_{k \geq 0} P^k(L)$.

**Proof.** (1) $P(L) = \{uv^2 \mid uv \in L, u, v \in X^+\} \subseteq L$ if and only if $L$ is a post-plus language. By Remark 3, $L$ is a post-plus language if and only $\circ(L) = L$.

(2) Clearly, $L = P^0(L) \subseteq \bigcup_{k \geq 0} P^k(L)$. If $uv \in \bigcup_{k \geq 0} P^k(L)$, $u, v \in X^+$, then $uv \in P^i(L)$ for some $i$. Thus $u^2v \in P^{i+1}(L) \subseteq \bigcup_{k \geq 0} P^{k}(L)$. That is, $\bigcup_{k \geq 0} P^k(L)$ is a post-plus language containing $L$. From the definition of $\circ(L)$, we have that $\circ(L) \subseteq \bigcup_{k \geq 0} P^k(L)$. Now show that $\bigcup_{k \geq 0} P^k(L) \subseteq \circ(L)$. Since $\circ(L)$ is a post-plus language, if $uv \in \circ(L)$, $u, v \in X^+$, then $u^2v \in \circ(L)$. As $L = P^0(L) \subseteq \circ(L)$, by the definition of $P^k(L)$ and by mathematical induction, $P^k(L) \subseteq \circ(L)$ for every $k \geq 0$. □

For a language $L \subseteq X^+$, define $m(L)$ to be $m(L) = \min\{|x| \mid x \in L\}$. Consider a word $x \in X^+$. If $|x| = 1$, then $P(x) = \emptyset$. If $|x| > 1$, then $x \notin P(x) = \{w^2 \mid x = uw, u, v \in X^+\}$ and $|x| < m(P(x))$. Now, we show the following theorem concerning context-sensitive post-plus languages.

**Theorem 31.** Let $L \subseteq X^+$ be a context-sensitive language. Then $\circ(L)$ is a context-sensitive language.

**Proof.** Let $L \subseteq X^+$ be a context-sensitive language. Then, there is a linear-bounded automaton $M_1$ which puts $u$ in tape $T_1$ for any given $u \in X^+$, checks whether there is any prefix $w$ of $u$ in $L$ and puts $w$ on tape $T_2$ if $w$ is in $L$. For any given word $u$, the length of $u$ is finite and then, there are only finitely many words on tape $T_2$.

Let $M_2$ be a Turing machine which checks whether there is any word on tape $T_2$. If there is no word on tape $T_2$, then $M_2$ enters a state ‘No’ and halts. Otherwise, $M_2$ checks whether the first word $x$ on tape $T_2$ is the word $u$. $M_2$ halts in a final state ‘Yes’ if $x = u$. When $x \neq u$, $M_2$ generates $P(x)$, puts all words $y \in P(x)$ with $|y| \leq |u|$ on tape $T_2$ and deletes $x$ from tape $T_2$. Then $M_2$ goes back to check whether tape $T_2$ is empty again. Since $|x| < m(P(x))$ and every word generated is bounded by the length of $u$, $M_2$ will always finish checking a word $x$ and all the possible words generated from $x$ and either find a word equal to $u$ or turn to the next word on tape $T_2$.

As $|X|^{|u|}$ is finite for every $u \in X^+$, one needs only a bounded space for each $u \in X^+$ and $M_2$ will always halt in a final state ‘Yes’ or a state ‘No’. That is, $u \in \circ(L)$ or $u \notin \circ(L)$. And, $M_2$ is a linear-bounded automaton. If we construct a linear-bounded automaton $M$ which simulates $M_1$ followed by $M_2$, then $\circ(L)$ is accepted by $M$. Therefore, $\circ(L)$ is a context-sensitive language. □

Theorems 17 and 31 yield:
Lemma 33 (Lyndon and Schützenberger [10]). If \( uv = x \) and \( v \in X^* \) and \( u \neq 1 \), then \( u = v \cdot y \) for some \( x, y \in X^* \) and \( k \geq 0 \).

The rest of this section deals with properties of concatenations of words with post-plus languages spanned by words. First, we consider the following lemma:

Lemma 34. Let \( a \neq b \in X \) and let \( w \in X^* \). Then \( wb^kub^k \notin \omega(w) \) for every \( w \in X^* \) and \( k \geq |w| \) with \( w < u \).

Proof. Suppose there is \( wb^kub^k \in \omega(w) \) for some \( u \in X^+ \) and \( k \geq |w| \) with \( w < u \).

Then there exist \( u_1v_1, u_2v_2, \ldots, u_nv_n \in \omega(w) \) such that \( u_1v_1 = w \), \( u_2v_2 = u_1v_1 \), \( i = 2, 3, \ldots, n \), \( u_nv_n = wb^kub^k \). As \( k \geq |w| \) and \( v, u \in X^+ \), by the observation of post-plus languages, \( b^{i+} \leq v \) for every \( v \in \omega(u_i) \). As \( u \in X^+ \) and \( j_i \geq |u| \), \( b^{i+} \neq s u \). One must have that \( u_iv_i = wb^{j_i} \) for some \( j_i \geq j_1 \) and \( v_i = w_i b^{j_i} \) for some \( w_i, v_i \in X^+ \) such that \( v = w_v v_1 \). Thus \( u_2v_2 = wb^{j_2}w_1b^{j_2} = wb^kub^k \) for some \( j_3 > 0 \). As \( k \geq |w| \), if \( j_3 < j_2 \), then \( b^{j_3-k}w_1b^{j_3} = w \) when \( j_2 \geq k \) or \( w_1b^{j_3} = b^{j_3}u \) when \( k \geq j_2 \). Since \( w_1 \leq |w| < u \), \( w_1 \leq |w| < u_3b^{j_3} \). Thus \( w_1b^{j_3} = xw_1 \) for some \( x \in X^+ \). By Lemma 33 and \( b^{j_3} = b^{j_3}b^{j_3} \), \( u_nv_n = wb^kub^k \) for some \( j_3 \geq j_3 > k \). This contradicts the fact that \( w < u \) and \( w' \in X^* \). Hence, \( j_3 \geq j_2 \). Since \( x \in \omega \), \( |x| = |w| \), \( k < j_2 \leq j_3 \). As \( u_nv_n = wb^kub^k \) for some \( j_3 \geq j_3 > k \). This contradicts the fact that \( w < u \) and \( w' \in X^* \). Therefore, \( wb^kub^k \notin \omega(w) \).

Proposition 35. Let \( w \in X^+ ab^k \) where \( a \neq b \in X \). Then for every \( u \in X^+ \), \( \omega(w) \) and \( \omega(w) \) are not post-plus languages.

Proof. First, we show that \( \omega(w) \) is not a post-plus language. Suppose \( w \in X^+ \) and \( a \neq b \in X \). Then \( wb^kab^k \in \omega(w) \) for some \( k \geq 1 \). Let \( u \in X^+ \). Then \( wb^kab^k \in \omega(w) \) and \( \omega(w) \) is a post-plus language.

Now, we show that \( \omega(w) \) is not a post-plus language. Let \( w = w^kab^k \) for some \( k \geq 1 \). Then \( \omega(w) \) is not a post-plus language. We consider the following three cases: (a) If \( u = b^m \) for some \( m \geq 1 \), then \( \omega(w) \subseteq b^m \). Thus \( \omega(w) \subseteq b^m \in \omega(w) \). Suppose \( u = u^m \) for some \( u^m \subseteq X^+ \), \( m \geq 0 \). (b) If \( m = 0 \), then \( \omega(w) \subseteq b^m \in \omega(w) \). By Lemma 34, \( \omega(w) \subseteq b^m \in \omega(w) \). (c) If \( m \geq 1 \), then \( \omega(w) \subseteq b^m \in \omega(w) \). By Lemma 34, \( \omega(w) \subseteq b^m \in \omega(w) \). That is, \( \omega(w) \subseteq b^m \).
$b^{[\omega(u)]} \notin w(\omega(u))$. Each of the above three cases implies that $\omega(w \circ u) \neq w(\omega(u))$. By Remark 3, $w(\omega(u))$ is not a post-plus language. □

For words $w, u \in X^+$, we now consider the case $\omega(w)u \subseteq \omega(w)$.

**Lemma 36.** Let $w \in X^+b$ where $b \in X$. For $u \in X^+$, $\omega(w)u \subseteq \omega(w)$ if and only if $u = b^m$ for some $m \geq 1$.

**Proof.** Let $w \in X^+b$ and let $u \in X^+$. Then clearly $\omega(w) \subseteq X^+b$. Assume that $\omega(w)u \subseteq \omega(w)$. Consider $u = vab^m$ for some $m \geq 0$, $v \in X^+$ and $a \neq b \in X$. If $m = 0$, then $wu = wwa \in \omega(w)u \setminus \omega(w)$, a contradiction! Now, let $m > 1$. Suppose $w^2b^{[\omega(w)]}u \in \omega(w)$. Then there exist $u_1v_1, u_2v_2, \ldots, u_nv_n \in \omega(w)$ such that $u_1v_1 = w$, $u_iv_i = u_{i-1}v_{i-1}^2$, $i = 2, 3, \ldots, n$, $u_nv_n = w^2b^{[\omega(w)]}u$. As the power of $b$ is $2[\omega(w)]$, there exists $1 \leq i < n$ such that $u_iv_i = w^b, j_1 \geq |\omega(u)|$. By the definition of $\omega(u_i, v_i)$, for every $x \in \omega(u_i, v_i)$, $b^j \leq x$. Since $u_nv_n = w b^2b^{[\omega(w)]}u \in \omega(u_i, v_i)$, $m \geq j_1 > |u|$. This contradicts that $|u| = |\omega(w)| + m + 1 > m$. Thus $w^2b^{[\omega(w)]}u = w^2b^{[\omega(w)]}vab^m \in \omega(w)u \setminus \omega(w)$, a contradiction! Hence, $\omega(w)u \subseteq \omega(w)$ implies that $u = b^+$.

Conversely, let $u = b^m$ for some $m \geq 1$. For every $w' \in \omega(w)$, $w' = X^+b$. Thus $w'u = w^b^m \in P^{m}(w')$. By (2) of Proposition 30, $\omega(w) = \bigcup_{k \geq 0} P^k(w)$. Since $w' \in \omega(w)$, there exists $k \geq 0$ such that $w' \in P^k(w)$. By (1) of Lemma 29, $w'u \in P^{m}(w') \subseteq P^{m}(P^{k}(w)) \subseteq \omega(w)$. Therefore, $\omega(w)u \subseteq \omega(w)$. □

From Lemma 36, we have the following property:

**Corollary 37.** Let $w \in X^+b$ where $b \in X$. For $L \subseteq X^+$, $\omega(w)L \subseteq \omega(w)$ if and only if $L \subseteq b^+$.

**Lemma 38.** Let $w \in Xb^+$ and let $u \in X^+$. Then $\omega(w)u$ is a post-plus language if and only if $u \in b^+$.

**Proof.** Let $w \in Xb^+$ and let $u \in X^+$. It is clear that $\omega(w) \subseteq Xb^+$.

Suppose $u \in X^+ \setminus b^+$. Say $u = u'a^{b^k}$ for some $k \geq 0$, $u' \in X^*$ and $a \neq b \in X$.

**Case (1):** If $k = 0$, then, for any $v \in \omega(w)$, $vu = vu'a \notin \omega(w) \subseteq Xb^+$, i.e., $vu^2 \notin \omega(vu) \cap (w)u$.

**Case (2):** If $k \geq 1$ and $u \leq_s ub$, then by Lemma 33, $u \in b^+$, a contradiction! Thus, if $k \geq 1$, then $u \notin_s ub$. This implies that $vu \in \omega(vu) \cap (w)u$. By Remark 3, languages $\omega(w)u$ in both of the above two cases are not post-plus languages.

Conversely, suppose that $u = b^k$ for some $k \geq 1$. Since every $v \in \omega(w)$ is in this form $v = ab^i$ for some $i \geq 1$ and $a \in X$, $vu = ab^{i+k}$. It is clear that $\omega(vu) = ab^{i+k}b^* = (ab^*)^+ b^k = \omega(v)u \subseteq \omega(w)u$. From Remark 3, we have that $\omega(w)u$ is a post-plus language. □

Before we consider the case of $w(\omega(u))$ being a post-plus language for $w, u \in X^+$, we show the following lemma.
Lemma 39. For $u \in X^+b$ with $b \in X$, $ub^2[u]ub^2[u] \in \omega(u)$ if and only if $u \in bb^+$.  

Proof. Let $(ub^2[u])^2 \in \omega(u)$. Then, clearly, $u \neq b$. Assume that $u \in X^+ab^j$ for some $j \geq 1$ and $a \neq b \in X$. If $u = ab^j$, then $\omega(u) = ab/b^*_b$. This implies that $(ub^2[u])^2 = (ab^2[u]+)^2 \notin \omega(u)$, a contradiction! If $u \in X^+ab^j$, then by Lemma 34, $ub^2[u]ub^2[u] \notin \omega(u)$, a contradiction! Thus $u \in bb^+$. Conversely, suppose $u = b^m$ for some $m \geq 2$. Then $(ub^2[u])^2 = b^{6m} \in b^m b^*_b = \omega(u)$. □

Proposition 40. Let $w \in Xb^+$ and let $u \in X^+$. Then $w(\omega(u))$ is a post-plus language if and only if $u \in bb^+$.  

Proof. Let $w \in Xb^+$ and let $u \in X^+$. Then $w = w'b$ for some $w' \in X^+$. From Proposition 4, it follows that $\omega(wu) \subseteq \omega(w(\omega(u)))$. Suppose $w(\omega(u))$ is a post-plus language. Then $\omega(w(\omega(u))) = w(\omega(u))$. Since $w(\omega(u))$ is a post-plus language, $u \in XX^+$, that is, $u = u'a$ for some $u' \in X^+$ and $a \in X$. As $wua^2[ua^2[u] \subseteq \omega(wu) \subseteq \omega(w(\omega(u))) = w(\omega(u))$, $uab^2[u]uab^2[ua^2[ua^2[u] \in \omega(u)]$. By Lemma 39, $u \in aa^+$. Say, $u = a^k$ for some $k \geq 2$. Clearly, $\omega(u) \subseteq a^+$. If $a \neq b$, then $wu = w'b^*a$. This implies that $wuba^k = w'ba^kba^k \in \omega(wu) \subseteq w(\omega(u))$. Thus $uba^k \in \omega(u) \subseteq a^+$, a contradiction. Hence, $a = b$, that is, $u \in bb^+$. Conversely, suppose $w = ab^j$ and $u = b^k$ for some $a \in X$, $i \geq 1$ and $k \geq 2$. Then, clearly, $w(\omega(u)) = ab^j(b^k b^*_b) = ab^{i+k}b^*_b = \omega(w(\omega(u)))$. That is, $w(\omega(u))$ is a post-plus language. □

We conclude the results obtained in Propositions 35 and 40, Lemmas 36 and 38 as follows:

Theorem 41. Let $w \in X^+b$ with $b \in X$. Then for $u \in XX^+$, the following statements are equivalent:

1. $\omega(wu)$ is a post-plus language;
2. $w \in Xb^+$ and $u \in bb^+$;
3. $w \in Xb^+$ and $\omega(wu) \subseteq \omega(w)$;
4. $w \in Xb^+$ and $w(\omega(u))$ is a post-plus language.

7. Post-plus languages, codes and dense languages

This section is devoted to the study of relations between post-plus codes, dense languages and the post-plus languages with finite generating sets. Recall that $D(1)$ is the set of all d-primitive (i.e., non-overlapping) words over $X$. From [5], we have the following property concerning overlapping words.

Lemma 42 (Hsu et al. [5]). Let $w \in X^+$. Then $w \notin D(1)$ if and only if there exists a unique word $z \in D(1)$ with $|z| \leq \frac{1}{2}|w|$ such that $w = wz'z$ for some $w' \in X^+$. 
Let Proposition 43. Let \( w \in X^+ \). Then \( \sigma(w) \) is a suffix code if and only if there exist no \( u \in X^+ \) and \( v \in X^* \) such that \( w = uvu \).

**Proof.** Let \( w \in X^+ \) and let \( \sigma(w) \) be a suffix code. Suppose \( w = uvu \) for some \( u \in X^+ \) and \( v \in X^* \). Then \( \{w, uvvu, (uvwu)(uvu)\} \subseteq \sigma(w) \) which is not a suffix code, a contradiction!

For the converse, assume that \( w \neq uvu \) for every \( u \in X^+ \) and \( v \in X^* \). Suppose \( \sigma(w) \) is not a suffix code. Then \( w \notin X \) and there are \( x, y \in \sigma(w) \) such that \( x <_s y \). Let \( x = wx_1 \) and let \( y = wy_1 \) for some \( x_1, y_1 \in X^* \). Then \( x <_s y \) implies that there exists \( y_2 \in X^+ \) such that \( x <_s y = y_2x_1 \). That is, \( y_2 = y_2w_1 \). Thus there exist \( y_2 \in X^+ \) such that \( y_2w_1 \) is a suffix code. Then \( w = u(iviz) \) for some \( u \in X^+ \) and \( v \in X^* \) such that \( w = u(iviz) = w_1 \). By Lemma 33, there are \( u, v \in X^* \) such that \( w = u(iviz) = w_1 \). If \( k > 0 \), then \( w = u(iviz) = w_1 \), a contradiction! Let \( k = 0 \). Then \( w = y_2w_1x_1 = wuvx_1 = wuvx_1 \) is a suffix code. There exist \( u_1v_1, u_2v_2, \ldots, u_nv_n \in \sigma(w) \) such that \( u_1v_1 = w, u_1v_1 = u_1v_2^{2}, i = 2, 3, \ldots, n, u_nv_n = y \). This implies that there exists \( 1 \leq i < n \) such that \( |u_1v_1| \leq |w| \) and \( |u_1v_2| > |w| \). It is clear that every \( v_1 \) is a concatenation of proper suffixes of \( w \). Thus \( v_1 = z_1z_2 \cdots z_m \) for some \( m \geq 1 \) and \( z_j \in X^+ \) with \( z_j <_s w \). Since \( |u_1v_1| \leq |w| \) and \( |u_1v_2| > |w| \), there is \( 1 \leq i \leq m \) such that \( |u_1v_2z_1z_2 \cdots z_{i-1}| < |w| \) and \( |u_1v_2z_1z_2 \cdots z_{i-1}| > |w| \). Thus, either \( z_i <_s w \) or there exists \( z' \in X^+ \) such that \( z' <_s z_i <_s w \) and \( z' <_p w \). Both cases imply the same result that \( w \notin D(1) \). By Lemma 42, there is \( z \in D(1) \) such that \( w = zw'z \) for some \( w' \in X^* \), a contradiction! \( \square \)

**Example 44.** \( \sigma(abab^2) \) is a suffix code. For if there exist \( u, v \in \sigma(abab^2) \) such that \( u <_s v \), then there exist \( x, y \in X^+ \) such that \( u = xabay \). This contradicts the fact that every word in \( \sigma(abab^2) \) can have subword \( aba \) only as a prefix.

A language \( L \subseteq X^+ \) is said to be a **solid code** if \( L \) is an infix code, every word in \( L \) is \( d \)-primitive and \( \text{Pre}(u) \cap \text{Suf}(v) = \emptyset \) for every two words \( u, v \in L \). Properties of solid codes can be found in [7] and [12]. There exists a language \( L \) which is not a solid code such that \( \sigma(L) \) forms a suffix code. For example: let \( L = \{aba^2b^2, a^2b^2a^3b^3\} \). By Proposition 43, \( \sigma(aba^2b^2) \) and \( \sigma(a^2b^2a^3b^3) \) are suffix codes. Since every word in \( \sigma(aba^2b^2) \) can not have the subword \( a^3 \) and every word in \( \sigma(a^2b^2a^3b^3) \) cannot have the subword \( aba \), \( \sigma(L) \) is a suffix code. However, when we consider the post-plus languages spanned by a solid code, we have the following property:

**Proposition 45.** Let \( L \subseteq X^+ \) be a solid code. Then the language \( \sigma(L) \) is a suffix code.

**Proof.** Let \( L \subseteq X^+ \) be a solid code. Then, by Proposition 43, \( \sigma(u) \) is a suffix code for every \( u \in L \). Suppose that \( \sigma(L) \) is not a suffix code. Then there exist two distinct words \( u, v \in L \) such that there are \( u_1 \in \sigma(u) \) and \( v_1 \in \sigma(v) \) with \( u_1 <_s v_1 \). As
Proposition 47. For any finite language $L \subseteq X^+$. Thus $u$ is a subword of $v$. It implies that either $u$ is a subword of $v$ or $\text{Pre}(u) \cap \text{Suf}(v) \neq \emptyset$. This contradicts the fact that $L$ is a solid code. Therefore, $\sigma(L)$ is a suffix code. □

For $a \in X$ and $w \in X^*$, the number $M_d(w)$ is defined by $M_d(w) = \max\{|y| \mid w \in X^*, y \in X^*aX^* \}$, that is, $M_d(w)$ is the maximal length of subwords of $w$ which do not contain the letter $a$.

Now, we are going to investigate the case of post-plus languages being not dense.

Lemma 46. Let $w \in X^*$ and let $a, b \in X$ with $a \neq b$. Then $X^*b^{\|w\}|X^* \cap \sigma(wa) = \emptyset$.

Proof. Let $a, b \in X$ with $a \neq b$ and let $w \in X^*$. Then clearly, $M_d(wa) < |wa|$. For $y \in \sigma(wa)$, there exist $u_1, v_1, u_2, v_2, \ldots, u_n, v_n \in X^+$ such that $u_1v_1, u_2v_2, \ldots, u_nv_n \in \sigma(wa)$, $u_1v_1 = wa, u_iv_i = u_{i-1}v_{i-1}^2$, $i = 2, 3, \ldots, n$, and $u_nv_n = y$. Clearly $M_d(v_1) \leq M_d(wa)$. Since $v_1$ ends by the letter $a$ and $u_1v_1 = wav_1$, $M_d(u_1v_1) = M_d(u_1v_1^2) = M_d(wa)$. Similarly, $M_d(u_iv_i) = M_d(u_{i-1}v_{i-1}^2)$, $i = 2, 3, \ldots, n$. Thus $M_d(y) = M_d(wa)$. As $M_d(wa) < |wa|$, $X^*b^{\|w\}|X^* \cap \sigma(wa) = \emptyset$. □

Proposition 47. For any finite language $L \subseteq X^+$, the language $\sigma(L)$ is not dense.

Proof. Suppose that $L \subseteq X^+$ is a finite language. Let $m = \max\{|w| \mid w \in L\}$. Consider the word $v = a^{2m}b^{2m}$ for $a \neq b \in X$. From Lemma 46, we have that $X^*vX^* \cap \sigma(wa) = \emptyset$ for every $w \in L$. That is, $X^*vX^* \cap \sigma(L) = \emptyset$. Thus $\sigma(L)$ is not dense. □

Clearly, if $L$ is dense, then $\sigma(L)$ is dense. We conjecture that $\sigma(L)$ being dense implies that $L$ is dense, which we do not intend to prove in this paper.

Acknowledgements

The author would like to thank the referee for his reading of the manuscript and valuable suggestions.

References