Post Correspondence Problem: Words Possible as Primitive Solutions *

Marjo Lipponen

Department of Mathematics, University of Turku
20500 Turku, Finland
e-mail: marlip@polaris.cc.utu.fi

Abstract. Three types of "primitive" or "prime" solutions for PCP have earlier been investigated. The sets of words belonging to one of the three types (for some instance of PCP) form an increasing hierarchy. In our main result we show that the hierarchy is proper, except the binary case. We also give a sharpened characterization of the finite type.

1 Introduction

The Post Correspondence Problem, [8], is one of the "cornerstones of undecidability". Reduction from the Post Correspondence Problem is one of the most common undecidability arguments in many areas of computer science, [9].

It is well known that the set of solutions for an instance of the Post Correspondence Problem is a star language: whenever \( w_1 \) and \( w_2 \) are solutions, then so is \( w_1w_2 \). It is natural to consider \( w_1 \) and \( w_2 \) to be "simpler" solutions than \( w_1w_2 \). There are also other ways to define what it means for a solution to be "simple", "prime" or "primitive". Such considerations have turned out to be theoretically important in various contexts. For example, in the purely morphic characterization of recursively enumerable languages given in [1], attention is restricted to solutions \( w \) such that no proper prefix of \( w \) is a solution. Because of this restriction the use of an intersecting regular language becomes unnecessary in the characterization. (The intersection is needed in the normal construction to exclude the garbage produced by unrestricted equality sets.) Solutions restricted in this way are in the sequel referred to as \( F \)-primes: no (nonempty) final subword can be removed in such a way that the remainder is still a solution. Similarly, a solution \( w \) for an instance of the PCP is termed an \( S \)-prime if no nonempty subword can be removed, and a \( P \)-prime if no nonempty scattered subword can be removed. These three different types of prime solutions were first introduced in [10]. In [7] the investigation was extended to concern analogous problems for the target alphabet.

In this work we take another point of view by studying words that can appear as a prime solution, for some instance of PCP. Such a study of primitive words and languages was initiated in [6].

* Work supported by Project 11281 of the Academy of Finland
Our main result in Section 3 studies the two crucial inclusion problems. We know, by definition, that every P-prime is an S-prime, and every S-prime is an F-prime. Thus the set of P-words is included in the set of S-words which, in turn, is included in the set of F-words. Our main result consists in showing that both of these inclusions are proper, except in the case of a binary alphabet.

Section 4 consists of a more detailed study of P-prime solutions. We consider two natural subtypes of the type of P-prime solutions and show that all type combinations are indeed possible in this framework.

The suitability of the Post Correspondence Problem for reduction arguments is due to the fact that in some sense the essence of computations is captured by PCP. Thus simple solutions of PCP mean simplifications of computations and the results contribute on an abstract level to the understanding of computations.

The study of P-words, S-words and F-words can also be viewed as a contribution to the general combinatorics on words. The property of being an S-word, for instance, is an abstract property of words that is so far rather poorly understood: of two similar-looking words, one can be an S-word, whereas the other fails to be an S-word.

2 Definitions and earlier results

Let \( g \) and \( h \) be nonerasing morphisms of \( \Sigma^* \) into \( \Delta^* \), where \( \Sigma \) and \( \Delta \) are finite alphabets. The equality set between \( g \) and \( h \) is defined by

\[
E(g, h) = \{ w \in \Sigma^+ \mid g(w) = h(w) \}.
\]

The pair \((g, h)\)=PCP is also referred to as an instance of the Post Correspondence Problem. Words in \( E(g, h) \), if any since the empty word \( \lambda \) is not included, are called solutions of PCP.

For a word \( w \) over \( \Sigma^* \), we now consider the sets of words obtained from \( w \) by removing a final subword, a subword or a scattered subword, respectively. Define

\[
\text{fin} (w) = \{ v_1 \mid w = v_1 x, \text{ for some } x \in \Sigma^* \},
\]

\[
\text{sub} (w) = \{ v_1 v_2 \mid w = v_1 x v_2, \text{ for some } v_1, v_2, x \in \Sigma^* \},
\]

\[
\text{scatsub} (w) = \{ v_1 \cdots v_k \mid w = x_1 v_1 \cdots x_k v_k x_{k+1}, \text{ for some } x_i, v_i \in \Sigma^* \}.
\]

(Observe that \text{fin} is not the same here as in [3].)

We can now determine three further sets, as follows:

\[
F(g, h) = \{ w \in E(g, h) \mid \text{fin} (w) \cap E(g, h) = \{ w \} \},
\]

\[
S(g, h) = \{ w \in E(g, h) \mid \text{sub} (w) \cap E(g, h) = \{ w \} \},
\]

\[
P(g, h) = \{ w \in E(g, h) \mid \text{scatsub} (w) \cap E(g, h) = \{ w \} \}.
\]

Words in the three sets are called \textit{F-prime}, \textit{S-prime} and \textit{P-prime solutions} for the instance \( \text{PCP}=(g, h) \), respectively.