Unavoidability of Primitive and Palindromic Words

Peter Leupold

Institut für Informatik,
Universität Leipzig,
Leipzig, Germany
Peter.Leupold@web.de

1 Primitive and Palindromic Words

A word is *primitive*, iff it is not a non-trivial (i.e. with exponent one) power of any word. Thus u is primitive, if $u = v^k$ implies $u = v$ and $k = 1$. We denote the language of all primitive words by Q. It is a well-known fact that for every non-empty word w there exists a unique primitive word p such that $w \in p^+$; this primitive word is called the *(primitive) root* of w and we will denote it by $\sqrt{w}$. The unique integer i such that $\sqrt{w}^i = w$ is called the *degree* of w. The notion of root is extended to languages in the canonical way such that $\sqrt{L} := \bigcup_{w \in L} \sqrt{w}$.

Primitive Words play an important role in the Theory of codes [1]. The most important open question concerning primitive words is, whether the language of all primitive words is context-free. This question was first raised by Dömösi, Horváth and Ito [2], and has so far resisted all attempts to answer it.

A second concept central to the work presented here is *palindromicity*. First off, for a word w by w^R we denote its reversal, that is $w[|w| \dots 1]$. If $w = w^R$, the word is called a palindrome. For words of even length, this means there is some word u such that $w = uuR$; these are called *even* palindromes. On the other hand *odd* palindromes are of odd length, and they are of the form $w = uxuR$ with a letter x at their center. The set of all palindromes of a language L is denoted by $\mathcal{P}al(L) = \mathcal{P}al \cap L$.

2 Unavoidable Languages

Unavoidability of languages formalizes the following intuitive concept: if a language L shares some words with every language from a given class $\mathcal{C}$, then it is unavoidable in $\mathcal{C}$, because parts of it appear in some sense everywhere. Depending on the size of these parts we define also a strong version of unavoidability.

Definition 1 [5] A language $U \subseteq \Sigma^*$ is called *unavoidable* in the language class $\mathcal{C}$, iff $U \cap L \neq \emptyset$ for all infinite languages $L \in \mathcal{C}$. U is *strongly unavoidable*, iff $U \cap L$ is infinite for all infinite languages $L \in \mathcal{C}$.

Notice that this concept is different from unavoidable sets or languages as they are used in Combinatorics on Words [6]; there, a set of words U is unavoidable, if there exists an integer k such that every word longer than k must have a word from U (or a morphic image of such a word) as a factor. Thus unavoidability is an absolute property of languages, not one relative to a language class as in our case. A further difference is that we demand that words of U be elements of all languages in $\mathcal{C}$, and not just that they occur as factors.

Trivially, Σ^* is strongly unavoidable for all possible language classes over the alphabet Σ, because it has an infinite intersection with any infinite language. Two less trivial examples can be derived from the Pumping Lemmata for regular and context-free languages.

Example 2 Let L_{sq} be the language of all words that contain a square. From the two Pumping Lemmata we can see that every infinite regular language has a subset of the form $w_1 w_2 w_2^+ w_3$ and that every infinite context-free language has a subset of the form $\{w_1 w_2^i w_3 w_4^i w_5 : i \geq 2\}$. Both sets contain only words with squares and are thus infinite subsets of L_{sq}. Thus the latter is strongly unavoidable for regular and context-free languages.

□

The central result from the first work on unavoidability is the following:

Theorem 3 *[5] The language of primitive words is strongly unavoidable for $CF \setminus LIN$.*

Like Example 2 this follows from the pumping properties, but in a much less direct manner. Basically one needs to use the fact that pumping a factor of a word either produces powers of the same word or infintely many primitive words, see for example the results of Shyr and Yu [8], Kászonyi and Katsura [4] or Păun et al. [7].

When we additionally require palindromicity of words a careful analysis leads to the following result:

Theorem 4 *[3] The language $Q^{(2)}$ is strongly unavoidable for palindromic context-free languages that are not regular and only have finitely many primitive words.*

Unfortunately, these results have not yet helped to get closer to a solution for the question about the context-freeness of Q. For example, the avoidability of primitive words for non-deterministic and/or inherently ambiguous context-free languages would provide a negative answer.

References

1. J. Berstel and D. Perrin. *Theory of Codes.* Academic Press, Orlando, 1985.
2. P. Dömösi, S. Horváth, and M. Ito. On the connection between formal languages and primitive words. *Analele Univ. din Oradea, Fasc. Mat.*, pages 59–67, 1991.

3. S. Z. Fazekas, P. Leupold, and K. Shikishima-Tsuji. On non-primitive palindromic context-free languages. *International Journal of Foundations of Computer Science*, 23(6):1277–1290, 2012.
4. L. Kászonyi and M. Katsura. On the context-freeness of a class of primitive words. *Publicationes Mathematicae Debrecen*, 51:1–11, 1997.
5. P. Leupold. Primitive words are unavoidable for context-free languages. In A. H. Dediu, H. Fernau, and C. Martín-Vide, editors, *LATA*, volume 6031 of *Lecture Notes in Computer Science*, pages 403–413. Springer, 2010.
6. M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics and Its Applications*. Addison-Wesley, Reading, Massachusetts, 1983.
7. G. Păun, N. Santean, G. Thierrin, and S. Yu. On the robustness of primitive words. *Discrete Applied Mathematics*, 117:239–252, 2002.
8. H. Shyr and S. Yu. Non-primitive words in the language p^+q^+. *Soochow J. Math.*, 4, 1994.