

## A NOTE ON THE SUFFICIENCY OF SOKOLOWSKI'S CRITERION FOR CONTEXT-FREE LANGUAGES

Anton NIJHOLT

*Faculty of Science, Department of Informatics, Nijmegen University, The Netherlands*

Received 10 May 1982

*Keywords:* Context-free languages, pumping lemmas

Sokolowski [5] introduced a condition on languages which can be used to show that a language is not context-free. This condition is the following: Let  $L \subseteq \Sigma^*$  be a context-free language. For every subset  $\Sigma'$  of  $\Sigma$  containing at least two distinct letters and for every  $u, v, w$  in  $\Sigma^*$ , if  $\{uxvxw \mid x \text{ in } \Sigma'^+\} \subseteq L$ , then there exist two different words  $x', x''$  in  $\Sigma'^+$ , such that  $ux'vx''w$  is in  $L$ . In [5] the non-context-free language

$$L = \{xc^kx \mid k > 0 \text{ and } x \in \{a, b\}^+\} \cup \{a, b\}^+$$

is used as an example to show that the "Classical Pumping Lemma" [1] cannot be used to determine that  $L$  is non-context-free, while the above given criterion simply leads to the conclusion that  $L$  is not context-free.

In [5] the question is raised whether the criterion is not only a necessary, but also a sufficient one. The answer is negative and this can be understood with the help of a simple variation of  $L$ . Consider the language

$$L' = \{xcxx \mid x \in \{a, b\}^+\} \cup \{xcx \mid x \in \{a, b\}^+\}.$$

For  $u = \varepsilon$ ,  $v = c$  and  $w = \varepsilon$  we have  $\{xcx \mid x \in \{a, b\}^+\} \subseteq L'$ . However, there exist two different words  $x$  and  $xx$  such that  $xcxx$  is also in  $L'$ . Hence, the criterion is satisfied while  $L'$  is not context-free.

It is interesting to note that the Ogden-like

languages (cf. [3]) of the form  $B_p = A_p \cup X^*\{aa, bb\}X^*$  with  $X = \{a, b\}$  and  $A_p = \{(ab)^n \mid n \in P\}$ , where  $P$  is a subset of the natural numbers, also satisfy Sokolowski's criterion. Notice that  $P$  can be chosen in such a way that  $A_p$  and therefore  $B_p$  is not context-free, not context-sensitive, not recursive or not recursively enumerable. The languages  $B_p$  illustrate that although Sokolowski's criterion is a very useful tool in determining non-context-free-ness, it only considers one of the many conditions of a language being non-context-free.

### References

- [1] Y. Bar-Hillel, M. Perles and E. Shamir, On formal properties of simple phrase structure grammars, *Zeitschr. Phonetik, Sprachwiss. und Kommunikationsforsch.* 14 (1961) 143-172.
- [2] L. Boasson, A remark on Ogden's lemma, *EATCS-Bull.* 4 (1978) 3-4.
- [3] L. Boasson and S. Horvath, On languages satisfying Ogden's lemma, *RAIRO Inform. Theorique* 12 (1978) 201-202.
- [4] W. Ogden, A helpful result for proving inherent ambiguity, *Math. Systems Theory* 2 (1968) 191-197.
- [5] S. Sokolowski, A method for proving programming languages non-context-free, *Inform. Process. Lett.* 7(3) (1978) 151-153.
- [6] D.S. Wise, A strong pumping lemma for context-free languages, *Theoret. Comput. Sci.* 3 (1976) 359-369.