A Note on Encodings for Context-Free Languages

William E. Marsh

a note on encodings for context - free languages

william e marsh

This note describes a problem that came up in an introductory mathematics class at Hampshire College, and is presented in the hope of generating amusing exercises rather than useful consequences.

In giving an example of a grammar for a context-free language, we numbered the re-write rules and listed the numbers of the rules in their order of use in different derivations. We spoke of the sequence of numbers as an "encoding" of the derivation. The class asked the following

<u>QUESTION</u>: Is the set of encodings of derivations from a grammar for a context-free language itself context-free ?

We eventually found an example which answered the question in the negative, but I will present here a quicker proof due to Stan Peters.

EXAMPLE: Let G be a grammar with the rules

1. S → S A B
2. S → a b
3. A → a
4. B → b (S,A and B are non-terminal; a and b are terminal)

It is immediately clear that the set E of encodings of derivations from G is characterized by $\alpha \varepsilon$ E, iff

(i) α contains the same number of 1's, 3's and 4's;
(ii) α contains exactly one 2 and it occurs to the right of all 1's;
(iii) no initial segment of α has more 3's or more 4's than 1's.

Peters pointed out that the intersection of E with the regular language $\{1^n 2 3^m 4^k \mid n,m,k > 0\}$ is the language $\{1^n 2 3^n 4^n \mid n>0\}$, which is not context-free (proof via the "xuwvy theorem" (cf Ginsburg, 1966, p 84)). Since the intersection of a context-free and a regular language is context-free, E cannot be such.

As an example of the kind of exercise that might be suggested, recall that a group is called "complete" if it is isomorphic to its automorphism group, and ask which grammars G are complete in the sense that the encodings of derivations from G and the language generated by G are letter-by-letter substitutions of each other.

BIBLIOGRAPHY.

Ginsburg, S. The Mathematical Theory of Context-Free Languages, 1966, New York: McGraw-Hill.