Dr Chiodo's Sheet 4 2017, Question 4, starred part

Thomas Forster

February 11, 2018

Show that $\{a, b\}^* \setminus \{ww : w \in \{a, b\}^*\}$ is context-free.

Dr Chiodo gives a grammar:

$S \to A|B|AB;$
$A \to CAC|a;$
$B \to CBC|b;$
$C \to a|b$

He supplies us this grammar, but I think a determined student would probably be able to work out for themselves that something along those lines might work. The hard part comes in showing that it not only *might* work, but that it does *in fact* work.

Every string corresponding to an A or a B (let's call them A *strings and* B *strings*) is of odd length and therefore can't be of the form ww. However we *do* have to show that every string AB is not of the form ww. Every A-string is of odd length and has an 'a' at its heart; every B-string is of odd length and has a 'b' at its heart. In fact the A-strings are *precisely* the set of those strings of odd length with an 'a' in the middle and the B-strings are *precisely* the set of those strings of odd length with an 'b' in the middle. We want the set of AB strings and BA strings to be precisely our putatively context-free language, and if the A string and the B string that go into our AB string are the same length we get what we want. However in an AB string (*mutatis mutandis* a BA string) the A and B moieties might be of different lengths. But this is OK! Suppose A has become the three string $(\cdot a \cdot)$ and B has become the five-string $(\cdot \cdot b \cdot \cdot)$. Now comes the clever bit. AB is the 8-string $(\cdot a \cdot \cdot \cdot b \cdot \cdot)$, and you think of it as the concatenation of two 4-strings. Now the first 4-string has 'a' as its second member and the second 4-string has 'b' as its second member—so they are distinct!!

Let's write this out properly for the general case. Suppose we have a string s of even length, that is an AB string or a BA string, wlog an AB string. It's of length $2n + 1 \; + \; 2k + 1$, where the first $2n + 1$ characters are an A string and the following $2k + 1$ characters are a B string. s is of the form ww' where w and w' are both of length $n + k + 1$. w is a string whose nth element is 'a' and w' is a string whose nth element is 'b', giving $w \neq w'$.