Succinctness and choosing between grammars (Solutions)

Exercise 1.
For $n \geq 2$, no grammar ever needs to contain every member of Σ_E^n. Explain why.

Solution

There are two answers here. First, if Σ_E^n is taken to contain useless n-grams like $\ltimes\rtimes$ or $\ltimes ab$, then these do not need to be included in the grammar because no string will ever have $\ltimes$ preceding another symbol that is not $\ltimes$.

Second, it is also true that there is always a more compact alternative to using a grammar that contains every member of Σ_E^n. Suppose that you have a positive grammar that contains every possible n-gram over Σ_E, where $n \geq 2$. In this case, the grammar generates Σ^*, the set of all possible strings over Σ. But Σ^* could have just as well been defined by a negative grammar G that forbids no n-grams at all (formally, $G = \emptyset$). Similarly, suppose that you have a negative grammar that contains every possible n-gram over Σ_E, with $n \geq 2$. In this grammar, nothing is allowed at all, not even the empty string, so the grammar generates the empty language $\emptyset$. But the empty language could also be generated by the positive n-gram grammar G that contains no n-grams at all (formally, $G = \emptyset$).

The same strategy of switching polarities can be used to show that for every n-grammar G that contains more than half of all possible n-grams (rounded up) there is a smaller grammar of opposite polarity that generates the same string language as G.