6.11 Expressiveness of grammatical formalisms and efficiency of syntactic processing

Nederhof M.J.

One of the problems studied in computational linguistics is the automatic analysis of the syntax of sentences in a language, as a first step towards determining the meaning of sentences. Typically, a computer program for syntactic analysis operates on the basis of a description of the syntax of a language developed by a linguist. Such a description is often called a (formal) grammar, and is written in a formalism, called the grammatical formalism, that can be interpreted by the computer program doing the syntactic processing.

The choice of the grammatical formalism is subject to two constraints. On the one hand, the grammatical formalism must be powerful, in order to allow linguists to express their knowledge about a language in a convenient way. On the other hand, we would like the syntactic processing to be fast, so that we can build it into practical systems where speed is very important, as for example in speech recognizers.

These two constraints are often in conflict. Powerful formalisms tend to lead to slow syntactic analysers, whereas formalisms that are less expressive often allow fast syntactic processing. It therefore appears that we should either restrict the power of the grammatical formalism, which means that the process of writing grammars becomes more difficult, or accept that our systems will be slower.

An alternative is offered by the observation that often linguists do not use the full power of grammatical formalisms, and the power that is used typically corresponds to a simpler formalism that can be processed more efficiently. This observation motivates an approach as exemplified in the figure: instead of doing syntactic processing on the basis of a context-free grammar, we may find that the power actually used can also be expressed by a finite automaton, and this is used for a cheaper kind of syntactic processing. The outcome has to be translated back to what context-free analysis would have produced.

Figure 17 ▸