Chapter 9

Using Prolog Grammar Rules

9.1 The Parsing Problem

Sentences in a language such as English are much more than just arbitrary sequences of words. We cannot string together any set of words and make a reasonable sentence. At the very least, the result must conform to what we consider to be grammatical.

A grammar for a language is a set of rules for specifying what sequences of words are acceptable as sentences of that language. It specifies how the words must group together into phrases and what orderings of these phrases are allowed. Given a grammar for a language, we can look at any sequence of words and see whether it meets the criteria for being an acceptable sentence. If the sequence is indeed acceptable, the process of verifying this will have established what the natural groups of words are and how they are put together. That is, it will have established something of the underlying structure of the sentence.

A particularly simple kind of grammar is known as a "context free" grammar. Rather than give a formal definition of what such a thing is, we will illustrate it by means of a simple example. The following might be the start of a grammar of English sentences:

\[
\begin{align*}
\text{sentence} & \rightarrow \text{noun_phrase}, \text{verb_phrase}.
\text{noun_phrase} & \rightarrow \text{determiner}, \text{noun}.
\text{verb_phrase} & \rightarrow \text{verb}, \text{noun_phrase}.
\text{verb_phrase} & \rightarrow \text{verb}.
\text{determiner} & \rightarrow [\text{the}].
\text{noun} & \rightarrow [\text{apple}].
\text{noun} & \rightarrow [\text{man}].
\text{verb} & \rightarrow [\text{eats}].
\text{verb} & \rightarrow [\text{sings}].
\end{align*}
\]
The grammar consists of a set of rules, here shown one to a line. Each rule specifies a form that a certain kind of phrase can take. The first rule says that a sentence consists of a phrase called a noun_phrase followed by a phrase called a verb_phrase. These two phrases are what are commonly known as the subject and predicate of the sentence:

```
sentence
   noun_phrase   verb_phrase
        the man           eats the apple
```

To see what a rule in a context free grammar means, read "X → Y" as saying "X can take the form Y", and read "X, Y" as saying "X followed by Y." Thus, the first rule can be read as:

A sentence can take the form: a noun_phrase followed by a verb_phrase.

This is all very well, but what is a noun_phrase and what is a verb_phrase? How are we to recognise such things and to know what constitute grammatical forms for them? The second, third and fourth rules of the grammar go on to answer these questions. For instance,

A noun_phrase can take the form: a determiner followed by a noun.

Informally, a noun phrase is a group of words that names a thing (or things). Such a phrase contains a word, the "noun", which gives the main class that the thing belongs to. Thus "the man" names a man, "the program" names a program and so on. Also, according to this grammar, the noun is preceded by a phrase called a "determiner":

```
noun_phrase
  determiner   noun
       the     man
```

Similarly, the internal structure for a verb_phrase is described by the rules. Notice that there are two rules for what a verb_phrase is. This is because there are two possible forms. A verb_phrase can contain a noun_phrase, as in "the man eats the apple", or it need not, as in "the man sings."