Prolog and its Applications
Lecture 9: Natural Language Grammar I

Ian Pratt-Hartmann

Room KB2.38: email: ipratt@cs.man.ac.uk

2016-7
Outline

Grammars and languages

The grammar of English
Some simple observations regarding English:

Who did Mary telephone? (Answer: John)
Who telephoned Mary? (Answer: Peter)

Who do you think Mary telephoned? (Answer: John)
Who do you think telephoned Mary? (Answer: Peter)

Who do you think that Mary telephoned?
* Who do you think that telephoned Mary?

Similarly:
* Who do you wonder whether Mary telephoned?
Another phenomenon you probably have not noticed:

The police sped up the traffic
The police sped (=went) up the alley

The police sped the traffic up
# The police sped the alley up
• The really fascinating thing about grammar is its systematicity:

The police sped up the alley
Which alley did the police speed up?
Up which alley did the police speed?

The police sped up the traffic
Which traffic did the police speed up?
* Up which traffic did the police speed?
• We can see this systematicity best by comparing phenomena across languages.

• Compare the following Italian and English examples:
  arrivato Gianni
  * Has arrived Gianni

  chiaro che Louisa non partir
  * Is clear that Louisa will not leave

  Chi credi che abbia telefonato?
  * Who do you think that has telephoned?

• Italian is known as a *Pro-drop* language, English is *non-Pro-drop*.
• Another example: word-order in English and German.
    Hast du [den Postboten gesehen]?
    Have you [seen the postman]?
    Er ging [die Treppe hinauf]
    He went [up the stairs]

• The head (main word) in the boxed phrase often comes at the end in German and at the beginning in English: German is (predominantly) head-final, and English predominantly head-initial.
• (Natural language) **Syntax** is the study of sentence structure
• Its most basic task is to explain why some sentences sound acceptable and others unacceptable.
• Actually, that is a simplification: there may be all sorts of reasons why sentences are unacceptable:
  Colourless green ideas sleep furiously.
(We will see another example at the end!)
Children master the basic syntax of their native language by the age of six or so.

That this is a puzzle. Note that (apparently) very little negative feedback is given.

A popular view is that certain basic principles of syntax are universal (hence, presumably, innate); languages vary in the ‘settings’ of certain parameters.

The task of theoretical linguistics is to elicit these principles and parameters.
There are many different formalisms for describing natural language grammar. Here are a few:

- Transformational grammar
- Head-driven phrase-structure grammar
- Tree-adjoining grammar
- Type-logical grammar
- Model-theoretic syntax
In the early days of theoretic linguistics, interest centered on the expressive power over different grammar-formalisms:

- Finite-state automata
- Context-free grammars
- Context-sensitive grammars
- Turing machines.

It can be shown that these formalisms allow the definition of a strictly increasing hierarchy of languages.

Which formalism is the correct one for natural language?
• For example, the language \( \{ a^{2n} \mid n > 0 \} \), i.e.

\[
\text{aa, aaaa, aaaaaa, aaaaaaaa, \ldots}
\]

is recognizable by a finite state automaton, thus:

![Finite state automaton diagram]

• Languages recognizable by FSAs are called regular languages.
• The language \( \{ a^n b^n \mid n > 0 \} \), i.e.

\[
\begin{align*}
ab, & \quad aabb, & \quad aaabbb, & \quad aaaaabbbb, & \ldots
\end{align*}
\]

cannot be recognized by an FSA; however, it can be recognized by a context-free grammar:

\[
S \rightarrow ab \quad S \rightarrow aSb,
\]

thus:

\[
\begin{align*}
S & \Rightarrow aSb \\
& \Rightarrow aaSbb \\
& \Rightarrow aaabbb
\end{align*}
\]

• Languages recognizable by CFGs are called context-free languages.
- The language $\{a^n b^n c^n \mid n > 0\}$, i.e.

$$abc, aabbcc, aaabbbccc, aaaaabbbcccc, \ldots$$

cannot be recognized by a context-free grammar; however, it can be recognized by a context-sensitive grammar:

$$S \rightarrow abc \quad ab \rightarrow aaXbb$$

$$Xb \rightarrow bX \quad Xc \rightarrow cc,$$

thus:

$$S \quad \Rightarrow \quad abc \quad \Rightarrow \quad aaXbbbc$$

$$\quad \Rightarrow \quad aabXbc$$

$$\quad \Rightarrow \quad aabbXc$$

$$\quad \Rightarrow \quad aabbcc.$$

- Languages recognizable by CSGs are called context-sensitive languages.
Outline

Grammars and languages

The grammar of English
• Over time, linguists lost interest in Formal Language Theory, however. The dominant formalism became **Transformational Grammar**, and its variants, which have very high expressive power.

• Grammaticality is accounted for by supposing that every grammatical sentence has a **phrase structure**, e.g.:

\[
S \leftarrow NP \rightarrow VP
\]

\[
NP \rightarrow John\rightarrow V \rightarrow NP
\]

\[
V \rightarrow hit
\]

\[
NP \rightarrow Charlie
\]

\[
S \leftarrow NP \rightarrow VP
\]

\[
NP \rightarrow Det \rightarrow N \rightarrow VP
\]

\[
V \rightarrow sat
\]

\[
P \rightarrow PP
\]

\[
NP \rightarrow the \rightarrow Det \rightarrow N \rightarrow mat
\]
The sentence is divided into phrases, corresponding to nodes of the tree.

Each phrase has a category (N, V, P, NP, VP, PP, S etc.)

There is considerable dispute among linguists as to what categories there are.

We shall take a fairly traditional approach.
• Common **terminal** categories (parts of speech) are:
  • N – noun
  • V – verb
  • Adj — adjective
  • Adv — adverb
  • P – preposition
  • Det – determiner
• Common nonterminal (phrasal) categories are:
  • S – sentence
  • NP – noun phrase
  • VP – verb phrase
  • PP – Preposition phrase
These structures are not arbitrary
• We are interested in a descriptive theory of language

• Phrase structure is assigned—at least in theory—on the basis of evidence, not tradition.

• In these lectures we will not discuss this evidence in any detail.

• Rather, we will be interested in using sentence structure to compute sentence-meaning.
- We can derive the structure of some simple English sentences using a context-free grammar, e.g.:

<table>
<thead>
<tr>
<th>Rule</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>NP,VP</td>
</tr>
<tr>
<td>VP</td>
<td>V,NP</td>
</tr>
<tr>
<td>NP</td>
<td>PN</td>
</tr>
<tr>
<td>V</td>
<td>hit</td>
</tr>
<tr>
<td>V</td>
<td>kissed</td>
</tr>
<tr>
<td>PN</td>
<td>John</td>
</tr>
<tr>
<td>PN</td>
<td>Charlie</td>
</tr>
<tr>
<td>PN</td>
<td>Mary</td>
</tr>
</tbody>
</table>
• This grammar generates the following phrase-structures

\[
S \\
\quad \rightarrow NP \quad VP \\
\quad \quad \rightarrow PN \quad V \quad NP \\
\quad \quad \quad \rightarrow John \quad hit \quad PN \\
\quad \quad \rightarrow Charlie \\
\]

\[
S \\
\quad \rightarrow NP \quad VP \\
\quad \quad \rightarrow PN \quad V \quad NP \\
\quad \quad \quad \rightarrow Charlie \quad kissed \quad PN \\
\quad \quad \quad \quad \rightarrow Mary \\
\]
Here is a more complicated grammar:

\[
\begin{align*}
S & \rightarrow \text{NP}, \text{VP} \\
\text{VP} & \rightarrow \text{V}, \text{NP} \\
\text{NP} & \rightarrow \text{Det}, \text{N} \\
\text{V} & \rightarrow \text{loves} \\
\text{N} & \rightarrow \text{boy} \\
\text{N} & \rightarrow \text{girl} \\
\text{Det} & \rightarrow \text{some} \\
\text{Det} & \rightarrow \text{every}
\end{align*}
\]
This grammar generates the following phrase-structures:
• In addition, it is natural to suppose that sentences arise through 
  *transformations*, in which material is moved around:
  
  I like bananas
  Bananas, I like

  It is clear that Louise will not leave
  That Louise will not leave is clear
  Mary has seen the postman
  Who has Mary seen? (Answer: the postman)

• We will make use of transformations in this course.
• This suggests a programme of work:
  • Write a (transformational) grammar which generates exactly the acceptable sentences of English
  • or German, Welsh, Chinese, Hebrew, Walpiri …

• It also suggests some questions:
  • Do the grammars of these very different languages have anything in common?
  • How is it that children (not adults!) learn them so easily?

• But we shall be interested in a more technical question:
  • How can we implement grammars for natural language in Prolog?
Finally, the promised example of an unacceptable but (arguably) grammatical sentence:

The ship sank.
Finally, the promised example of an unacceptable but (arguably) grammatical sentence:

The ship sank.

The ship the sailor built sank.
Finally, the promised example of an unacceptable but (arguably) grammatical sentence:

The ship sank.

The ship the sailor built sank.

The ship the sailor the dog bit built sank.

**Grammaticality** is a term of art.