1. Parsing

If trees are useful, how do we get them?

Parsing is the process of taking a string and a grammar and returning one or more parse trees for that string.

Analogous to running a finite-state transducer over a tape:

- non-terminals (e.g., words)
- terminal symbols (e.g., words)

The parsing process is likewise more complicated as well as a bit more challenging.

2. Exploring syntax

By grammar, or syntax, we have in mind the kind of implicit knowledge of your native language that you had mastered by the time you were 3 years old without explicit instruction.

The internal structure of NPs varies from language to language:

- The approach we'll explore isn't exactly "cutting-edge"
- The Cambridge Grammar of the English Language

3. Syntax (or Grammar)

Refers to the way words can be arranged in a given language:

- Grammar checkers
- Dialogue management
- Information extraction
- Machine translation

5. Constituency

Groups of words can be drawn to act as single units, called constituents.

In a given language, these units form coherent classes that behave in similar ways, with respect to:

- Internal structure
- External behavior

We can describe an internal structure for the class:

- Paradigmatic
- Syntagmatic

For example, English noun phrases consist of a pronoun, a proper noun, or a complex phrase including a common noun.

6. Constituency, cont'd: Noun Phrases

We can observe some commonality over the behavior of the following English phrases:

- They relate to other units in the language
- How they relate to other units in the language

7. Noun phrases in other languages

The internal structure of NP varies from language to language:

- English: This three expensive books (Dem Num Noun Adj nouns)
- French: Ces trois livres chers
- Chinese: 这些昂贵的书
- Thai: นี้สามหนังสือที่昂貴

These rules describe two kinds of NPs:

- Nominal → Noun | Nominal Noun
- NP → Det Nominal
- NP → ProperNoun

10. Some preliminary NP Rules

Some very simple rules for noun phrases:

<table>
<thead>
<tr>
<th>English</th>
<th>Thai</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>นี้แม่ของฉัน</td>
</tr>
<tr>
<td>VP</td>
<td>ที่ทำให้ฉัน</td>
</tr>
<tr>
<td>Nominal</td>
<td>a</td>
</tr>
<tr>
<td>NP</td>
<td>ของ</td>
</tr>
</tbody>
</table>

These rules describe two kinds of NPs:

- One that consists of a determiner followed by a nominal
- Another that says that proper nouns are NPs.

The third rule illustrates:

- A disjunction
- We disallow a phrase that contains a sequence of non-terminals
- We can short-circuit for two rules
- A recursion definition
- We can have a left-hand side of the rule
- We can have a right-hand side of the rule
- We can see how this works if we consider a noun phrase such as

11. A bit more detail on English Grammar

Huddleston and Pullum's The Cambridge Grammar of the English Language is 1860 pages long.

So we won't cover all of English by a very long way.
Just enough to uncover some key shortcomings of CFGs

We'll look briefly at

• Sentences
• Noun phrases
  ◦ Agreement
• Verb phrases
  ◦ Subcategorisation

12. Sentence Types
Declaratives
A plane left
Imperatives
Leave!
Yes-no questions
Did the plane leave?
WH questions
When did the plane leave?

13. Noun Phrases, more carefully
We can identify three quite distinct types of noun phrases:

Pronouns
she, he, we, ...

Proper Nouns
Edinburgh, Star Wars, the Eiffel Tower, ...

Complex noun phrases
the next prime minister after Thatcher

Consider the following moderately complicated noun phrase:
the first three morning flights from Denver to Tampa leaving before 10

S → NP VP
S → VP
S → Aux NP VP
S → WH-NP Aux NP VP
NP → Pro
NP → PropN
NP → CNP

14. NP Structure
That big NP is really about flights
• ... possessives
simple Robin’s car
complex Robin’s youngest child’s toy

Det → Art
Det → PropN ’s
Det → CNP ’s

15. Before the nominal: Determiners
Complex noun phrases can start with determiners

CNP → Det CNP

Determiners can be
Simple lexical items
the, this, a, her

(Recursive) possessives
simple Robin’s car
complex Robin’s youngest child’s toy

Det → Art
Det → PropN ’s
Det → CNP ’s

16. Before the nominal: Other premodifiers
Other premodifiers include

• Quantifiers, cardinals, ordinals:
every flight
every flight
three flights
first flight

• Adjectives and adjective phrases:
large cars
extremely sleepy baby

There are constraints we haven’t captured as the order of pre-modifiers:
• Between adjectives and quantifiers:
every eligible candidate
every eligible candidate
eligible every candidate

• Between one adjective and another:
big red bus
red big bus

• Following a common linguistic convention, I’m using an initial asterisk to indicate a word sequence which is not in a natural language, or cannot be accepted by a formal grammar.

• Likewise an initial question mark for a borderline in/out word sequence

17. The nominal: the head and its postmodifiers
Eventually (or even right away), we get to the Nominal
• Including the head, with or without compounding

The postmodifiers which stack up behind the head may include

• Prepositional phrases:
  flight from Seattle
• Non-finite clauses (gerundive, infinitive):
  flights arriving before noon
  flights to depart
• Relative clauses:
  flights that serve breakfast
  people whom the pilot knows

Similar general (recursive) rules to handle these

Nominal → Nominal PP
Nominal → Nominal GerundVP
Nominal → Nominal InfVP
Nominal → Nominal RelClause