

Ch. 9: CONTEXT-FREE GRAMMARS FOR ENGLISH

We looked at this a little bit in the first week:

- "I made her duck."
- "I saw the man in the park with the telescope."

Let's go back to this area of computational linguistics and look at it with more rigor.

Section 9.1: CONSTITUENTS

Syntax = how words are arranged in a sentence

constituent = a group of words that can behave as a single word

Michael	[proper noun]
she	[pronoun]
the house	[noun phrase]
Russian Hill	
a well-weathered three-story structure	

Context-free grammars = a formalism that will allow us to model constituents.

CFGs have a hierarchy (tree structure), so we can do more sophisticated modeling than with the strings (sounds, words, etc.) we saw earlier.

HOW DO WE KNOW WHEN A STRING IS A CONSTITUENT?

- When they can appear in the same syntactic environment, e.g., before a verb (see p. 325).
- When they can be preposed or postposed.

On Sept. 17th, I'd like to fly from A to D.
I'd like to fly on Sept. 17th to from A to D.
I'd like to fly from A to D on Sept. 17.

*On Sept., I'd like to fly 17th from A to D.
*On I'd like to fly Sept. 17 from A to D.
*I'd like to fly on Sept. from A to D 17th.

These are heuristic rules, i.e. not black-and-white rules.

Section 9.2: Definition of CFG

CFG = phrase-structure grammar

rules

lexicon

terminal

non-terminal

derivation

parse tree

start symbol

Fig. 9.1

Fig. 9.2

Fig. 9.3

Fig. 9.4

- Bracketed notation is an alternative to drawing a tree. See example on bottom of p. 330.
- Note that whether a string of words is a sentence or not depends on the grammar.

IMPORTANT THEORETICAL QUESTIONS ABOUT ENGLISH

- Is English context-free? I.e, can English be represented by a context-free grammar?
- Can English be represented by a finite-state machine? By a set of finite-state machines?

"CORRECT" GRAMMAR

- There is no one "correct" grammar of English.

Data from sociolinguistics:

- People disagree on whether a given sentence is OK.
- Also: intra-national variants, international differences, register (formal/informal, spoken/written).

- People aren't 100% consistent in their speech.
- People don't always use the forms they think they use (i.e. people will often tell you the written form they learned in school).

SENTENCES (section 9.3)

Syntactically, there are three kinds of sentences:

- declarative: NP VP
- imperative: no initial NP
- interrogative
 - wh-question: starts with wh-word (who, what, where, when, why, how)
 - yes/no question: inverted word order (we can recognize these by the fact that expected answer is yes/no, but that's not the kind of syntactic definition we want)

From the point of view of pragmatics, there are 3 kinds of sentences:

- statement
- question
- command

Any of the 3 pragmatic types can be expressed with any of the 3 syntactic types.

E.g.: Why don't you finish that pie?

Syntactically, this is an interrogative. (has question mark and inverted word order).

But pragmatically, is this really a question?

It could be:

A: Because it's spoiled.

But more likely...

- This means that having a parser (i.e. a syntactic analyzer) is not sufficient to be able to figure out the meaning of a sentence.
- This can be used for hints, jokes, and other indirect forms of expression.

--

NOUN PHRASES (section 9.4)

Premodifiers:

Advantage of having many kinds of premodifiers (see p. 336): can get some control over order, i.e. we can use the grammar to block invalid sequences of premodifiers.

But the rule on p. 336 is still too broad because you can't have all of those types in the same NP.

3 main kinds of postmodifiers:

- PPs
- non-finite clauses = gerundives
- relative clauses (that, who, which, etc.)

PPs often add time or location information to an NP.

Gerundives and relative clauses can add information or restrict the possible objects (in the real world) to which an NP can apply.