COSC343: Artificial Intelligence

Lecture 22: Natural Language Syntax II

Alistair Knott

Dept. of Computer Science, University of Otago

A simple context-free grammar

Here are the context-free grammar rules I introduced at the end of the last lecture.

Some rules about grouping words:

S	→	NP, VP
NP	→	PN
NP	→	Det, N
VP	→	V0
VP	→	V1, NP
S	→	S, Conj, S

Some rules about word classes:

Det	→	“the” | “a”
N	→	“dog” | “cat”
PN	→	“Fred” | “Jip”
V0	→	“slept” | “ran” | “snorted”
V1	→	“bit” | “chased” | “caught”
Conj	→	“and” | “so”

Syntactic analysis

Given these rules, what is the structure of the following sentence?

The dog bit a cat

Some terminology

A phrase-structure grammar is a formal specification of how well-formed sentences can be built out of phrases.

A context-free grammar (CFG) is a grammar in which the way a phrase is constructed doesn’t depend on the context that the phrase appears in.

A CFG is typically expressed as a set of context-free rules, a.k.a. rewrite rules.

- A rule has the form $lhs \rightarrow rhs_1, \ldots rhs_n$.
- This specifies that the phrase *lhs* can be composed of the sub-phrases $rhs_1, \ldots rhs_n$ (in that order).

Some terminology

A CFG has a distinguished symbol, to represent whole sentences. (Typically, S.)

A CFG has a set of terminal symbols, which only appear on the right-hand-side of rules. The terminal symbols of the CFG are words.

The set of terminal symbols in a CFG is called its lexicon.

Every symbol in a CFG which is not a terminal symbol is a nonterminal symbol.

Grammar basics: verb types

Notice that we have two sorts of verb in our grammar.

- Intransitive verbs (V0) describe an action by themselves. E.g. *slept*, *ran*, *snorted*...
- Transitive verbs (V1) describe an action that's done on an object (e.g. *bit*, *chased*, *caught*).
  - These verbs must be followed by a noun phrase, specifying the object in question.
  - If they don't get one, the result is syntactically ill-formed.
    **The dog caught*
    **Fred chased*
    (Linguists mark ill-formed sentences with an asterisk *.)

A phrase which is required by a verb is called its complement.

Grammar basics: verb types

There are also ditransitive verbs, which have two complement NPs.

- What's an example?
- What rule can we use to introduce these?

Other verbs take whole sentences as their complements.

- What's an example?
- What rule can we use to introduce these?

Grammar basics: verb phrases

The verb and its complements join together to produce a verb phrase.

> *Notice that the subject is not included: it joins to the whole verb phrase, rather than directly to the verb.*

There are some good reasons for this:

- *All* verbs have a subject: this doesn't depend on the type of verb.
- We can replace a verb and its complement(s) with a single phrase, leaving out the subject, but we can't replace the verb and the subject, leaving out the complements. For instance:

John	*chased*	*Bill*
John	*did so*	
??		*Bill*

Grammar basics: noun types

There are lots of different ways of making a noun phrase (NP).

- Singular count nouns need a determiner.
  I like the/a/my dog.
  **I like dog.*
- Mass nouns don't need one, but can optionally take *some* determiners.
  I like sand/the sand.
  **I like a sand.*
- Plural count nouns are a lot like mass nouns.
  I like dogs.
- Proper nouns are names of objects. They make NPs by themselves.
  I like John.
  **I like the John.*

Grammar basics: prepositional phrases

A prepositional phrase describes a place, a path, or a trajectory.

- It's formed by a preposition (P) and an NP. The rule: PP $\rightarrow$ P, NP.

For instance: *to the beach*, *along the floor*, *by John*, *between the trees*.

Some verbs take a PP as a complement—for instance *went*.
I went to the beach.

Some verbs take an NP and a PP complement—for instance *put*.
I put the cup on the table.
**I put.* **I put the cup.* **I put on the table.*

Some nouns take PP complements—for instance *bunch*.
I bought a bunch of bananas today.

Complements and adjuncts

Not everything which joins to N or V is a complement.

- If it's *obligatory*, we call it a complement.
- If it's *optional*, we call it an adjunct.

NPs can have PP adjuncts:

the dog
the dog in the tree

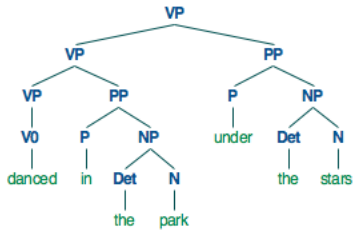
VPs can have PP adjuncts too:

danced
danced in the park
danced in the park under the stars

Rules for adjuncts

Rules for adding an adjunct can be applied many times, so they have a recursive structure.

VP $\rightarrow$ *V0* *(base case)*
VP $\rightarrow$ *VP PP* *(recursive case)*

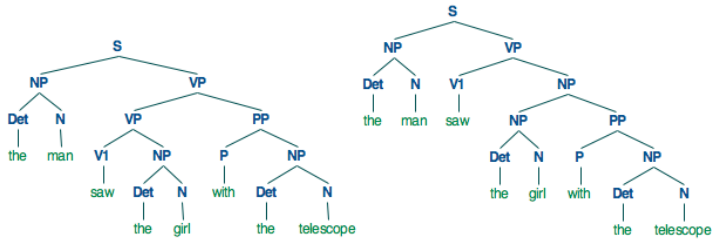


Syntactic ambiguity

There is often more than one way of analysing the structure of a sentence. For instance, how should we interpret the PP adjunct in this sentence?

The man saw the girl with the telescope

Syntactic ambiguity



Grammatical agreement

Some determiners can only introduce singular nouns:

a dog
**a dogs*

Others can only introduce plural nouns:

**all dog*
all dogs

(Others can introduce both singular and plural nouns:

the dog
the dogs)

In general, we say that the determiner has to agree with the N it introduces.

Variables for stating agreement constraints

To capture agreement between N and Det, we could write separate rules for singular and plural NPs:

NP	→	DetSing, NSing
NP	→	DetPlur, NPlur

However, it's more economical to use a variable to denote number. The idea is to express number as an *argument*, thus:

N[Num=sing]	→	"dog"
N[Num=plur]	→	"dogs"
Det[Num=sing]	→	"a"
Det[Num=plur]	→	"all"

Now we can write a single rule for NP, using a *variable* ?n for Num:

NP	→	Det[Num=?n], N[Num=?n]

Subject-verb agreement

There are also agreement relationships between subjects and verbs.

- These are partly to do with number (singular or plural)
- And partly to do with person (1st, 2nd or 3rd).

English doesn't have a rich subject-verb agreement system, but 3rd person singular is often markedly different. Agreement also shows up in very common verbs—e.g. the verb *to be*.

Person	Number	Subject	to love	to be
1st	sing	I	love	am
2nd	sing	You	love	are
3rd	sing	He/she/it	loves	is
1st	plur	We	love	are
2nd	plur	You	love	are
3rd	plur	They	love	are

Using variables for subject-verb agreement

To capture subj-verb agreement, we must modify the rule S → NP, VP.

- Again, it's inefficient to write separate rules for each person/number. We can again use variables instead.

We begin by defining each word with appropriate arguments for num/person:

NP[Pers=1, Num=sing] → "I"	V1[Pers=1, Num=sing] → "love"
NP[Pers=2, Num=sing] → "you"	V1[Pers=2, Num=sing] → "love"
NP[Pers=3, Num=sing] → "she"	V1[Pers=3, Num=sing] → "loves"

We must alter the VP rules, to pass the variables from verbs up to VPs:

VP[Pers=?p, Num=?n] → V1[Pers=?p, Num=?n], NP

Then we can define a single rule for creating sentences:

S → NP[Pers=?p, Num=?n], VP[Pers=?p, Num=?n]

A final change to the NP rule

Note that our current determiner-noun agreement rule is expressed as follows:

NP → Det[Num=?n], N[Num=?n]

We have to amend this a little, now that our NPs have person and number arguments:

NP[Pers=3, Num=?n] → Det[Num=?n], N[Num=?n]

Note: an NP of the form Det, N is always third person:

The dog is happy.
**The dog am happy.*

A grammar with variables for agreement

Here's a small lexicon of different word types:

Det[Num=sing] → "a"	N[Num=sing] → "dog"
Det[Num=plur] → "all"	N[Num=plur] → "dogs"
Det → "the"	N → "sheep"
NP[Pers=1, Num=sing] → "I"	V1[Pers=1, Num=sing] → "love"
NP[Pers=2] → "you"	V1[Pers=2, Num=sing] → "love"
NP[Pers=3, Num=sing] → "she"	V1[Pers=3, Num=sing] → "loves"
NP[Pers=1, Num=plur] → "we"	V1[Num=plur] → "love"
NP[Pers=3, Num=plur] → "they"	
NP[Pers=3, Num=sing] → "John"	V_Sent[Pers=1, Num=sing] → "think"
NP[Pers=3, Num=sing] → "Mary"	V_Sent[Pers=2, Num=sing] → "think"
PN[Num=sing] → "Fido"	V_Sent[Pers=3, Num=sing] → "thinks"
	V_Sent[Num=plur] → "think"
P → "in"	V0 → "snorted"
P → "behind"	V2 → "put"

A grammar with variables for agreement

Here are some grammar rules to go with the lexicon:

S	→	NP[Pers=?p, Num=?n], VP[Pers=?p, Num=?n]
S	→	“if”, S, S
NP[Pers=3, Num=?n]	→	PN[Num=?n]
NP[Pers=3, Num?n]	→	Det[Num=?n], N[Num=?n]
NP[Pers=3, Num=plur]	→	N[Num=plur]
VP[Pers=?p, Num=?n]	→	V0[Pers=?p, Num=?n]
VP[Pers=?p, Num=?n]	→	V1[Pers=?p, Num=?n], NP
VP[Pers=?p, Num=?n]	→	V2[Pers=?p, Num=?n], NP, PP
VP[Pers=?p, Num=?n]	→	V_Sent[Pers=?p, Num=?n], S
VP[Pers=?n, Num=?n]	→	VP[Pers=?p, Num=?n], PP
PP	→	P, NP

Exercise: what sentences can we now produce with this grammar? Are they all good ones?

Summary

We’ve introduced:

- Some terminology about phrase-structure grammars
- Different types of verb—and the idea of a verb phrase
- Different types of noun
- Prepositional phrases
- Complements and adjuncts
- Syntactic ambiguity
- Syntactic agreement
- Phrase-structure rules using variables

Readings

For this lecture, AIMA Chapter 23 Section 1.

For next lecture, AIMA Chapter 22 Section 1, Chapter 23 Section 2.1.