Syntax: Sentence Structure

Notes for SYCCL Day 3

1 Formal Grammars

Review of *Grammar Rules!*

In this activity we used a kind of **formal grammar** to model (a fragment of) English. We asked whether the grammar could **generate** various sentences, some of them real English sentences and some not. There are four possible outcomes:

1. The sentence can be generated and is a real sentence → true positive
2. The sentence can be generated but is *not* a real sentence → false positive
3. The sentence *cannot* be generated but is a real sentence → false negative
4. The sentence *cannot* be generated and is *not* a real sentence → true negative

Obviously, we want the overlap to be as great as possible!

The grammar we used is called a **context free grammar** (CFG). The name comes from the fact that we can rewrite a symbol using any rule with that symbol on its left-hand side — we cannot look at anything else when choosing what rule to use. We saw how that can lead to strange results with **subject–verb agreement**. Here is a tiny grammar that demonstrates the problem:

Grammar	**(Partial) Language**
S → NP VP	✓ the cat sits
NP → NP and NP	✓ the cat and the dog sit
NP → D N	✗ the cat sit
VP → V	✗ the cat and the dog sits
D → the	✓ the cat and the cat and the cat sit
N → {cat, dog}	✗ the cat and the cat and the cat sits
V → {sit, sits}	…

It's actually quite difficult to write a CFG that handles agreement correctly. This is a good exercise, but we won't do that today. An alternative is to factor out these patterns and deal with them separately. This is probably the most common approach.

Some of the other problems in the grammar from the previous activity could be solved pretty easily. For example, we could replace the rule 'NP → N', which is intended for personal names, with a more specific rule 'NP → Name', so that we don't generate ungrammatical sentences like *Dog sits*.

Are CFGs a good model of syntax?

A CFG makes precise, clear-cut predictions about what sentences are and are not possible. This is both a good and a bad thing. The advantage to making clear predictions is obvious. The disadvantage is that the distinction between good and bad sentences is not always so clear.

In linguistics, we distinguish **grammatical** sentences from **acceptable** sentences. A sentence is grammatical iff it is generated by the (human) grammar. A sentence may be unacceptable for many reasons:

- ungrammatical
- meaningless
- hard to process
- pronunciation is awkward
- socially inappropriate
- ...

So it's okay if our model of syntax generates some bad sentences, since they may be filtered out by other factors, which may be less black-and-white than syntax.

2 Phrase Structure

Another name for a CFG is a **phrase structure grammar** (PSG). The idea of this name is that each rule application should correspond to a meaningful unit: a **phrase**. For linguists, the goal is not only to get the right words in the right order — the **structure** should be correct too. In a nutshell, we want to know what the basic units are, and how they combine to form larger units.

Phrases, heads, and complements

One of the most important principles of syntax is that the properties of a phrase are derived from the properties of its parts — this is the principle of **compositionality**.

The most important element of a phrase is its **head**, which is a single word. The category of the head determines the category of the phrase — the head of a verb phrase (category VP) is a verb (category V), the head of a prepositional phrase (PP) is a preposition (P), and so on.

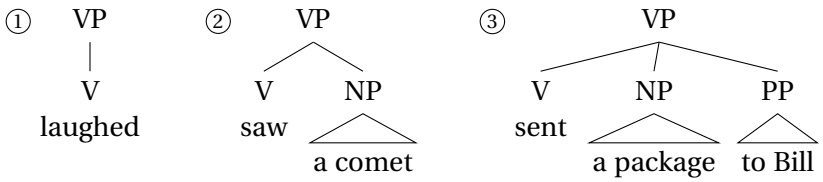
Besides the head, a phrase may contain one or more other phrases, which are called **complements**. The head determines the number and categories of its complements. For example, some verbs form a VP all on their own – these are called **intransitive verbs**. Others require a single complement, usually a noun phrase – these are called **transitive verbs**. Finally, some verbs take two complements – these verbs are **ditransitive**. Examples of each are given below.

(1) The children [$_{\text{VP}}$ laughed]. (intransitive)

(2) I [$_{\text{VP}}$ saw [$_{\text{NP}}$ a comet]]. (transitive)

(3) Ann [$_{\text{VP}}$ sent [$_{\text{NP}}$ a package] [$_{\text{PP}}$ to Bill]]. (ditransitive)

In these examples, I added labeled brackets showing some of the phrase structure. Another way to convey the same information is with **tree diagrams**. Below are three trees corresponding to the same three VPs.



The edges (lines) of the diagram connect the label for each phrase to the words and phrases that it is composed of. A triangle means that the inner structure of the phrase is not specified. You may also

encounter tree diagrams with lines from the word categories to the words themselves, but I won't draw those. After all, *laughed* isn't part of a verb, it *is* a verb!

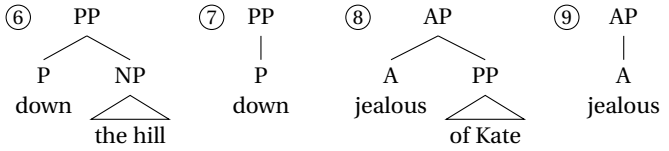
Some heads are flexible in the numbers and kinds of complements they may combine with. For example, some verbs which are normally transitive allow their complement to be omitted . Additionally, many ditransitive verbs that allow an NP and PP complement also allow a pair of NP complements.

(4) a. We [$_{VP}$ ate [$_{NP}$ cookies]].
    b. We [$_{VP}$ ate].

(5) a. Ann [$_{VP}$ sent [$_{NP}$ a package] [$_{PP}$ to Bill]].
    b. Ann [$_{VP}$ sent [$_{NP}$ Bill] [$_{NP}$ a package]].

Other common word categories work much like verbs. Prepositions, for example, also come in both intransitive and transitive types, as do adjectives. Many work either way.

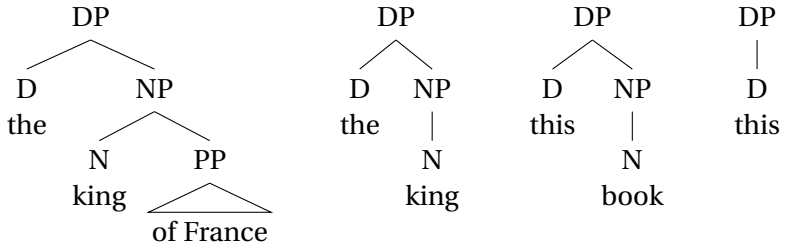
(6) The ball rolled [$_{PP}$ down [$_{NP}$ the hill]].
(7) Mary fell [$_{PP}$ down].
(8) Ann is [$_{AP}$ jealous [$_{PP}$ of Kate]].
(9) Ann is [$_{AP}$ jealous].



Noun phrases and determiner phrases

You may have noticed that in all of the examples so far, the head is the first element of the phrase. In general, this is true in English — we say that English syntax is **head-initial**. So what about noun phrases? Doesn't the noun come in the middle, after the determiner and before any PPs?

Actually, there is a theory that what we traditionally call a noun phrase is actually a **determiner phrase** (DP). The real noun phrase (NP) is a subpart of the DP. One cool aspect of this theory is that it allows either the determiner or noun to be transitive. A pronoun is just an intransitive determiner!

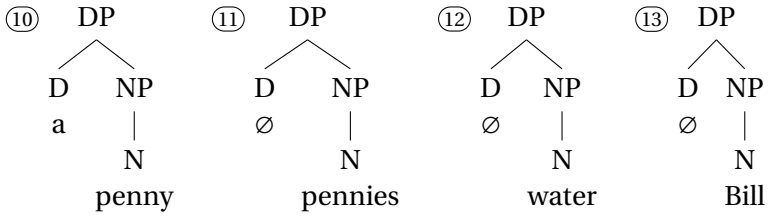


From now on, we'll assume the DP theory.

But what about noun phrases that aren't preceded by a determiner?

(10) [$_{DP}$ a penny]
(11) [$_{DP}$ ∅ pennies]
(12) [$_{DP}$ ∅ water]
(13) [$_{DP}$ ∅ Bill]

All of these can be used in the same positions as singular noun phrases, so they must have the same category: DP. The solution is to add a **null determiner** (∅) to our lexicon which combines with plurals, mass nouns, and proper names, and has no pronunciation of its own.



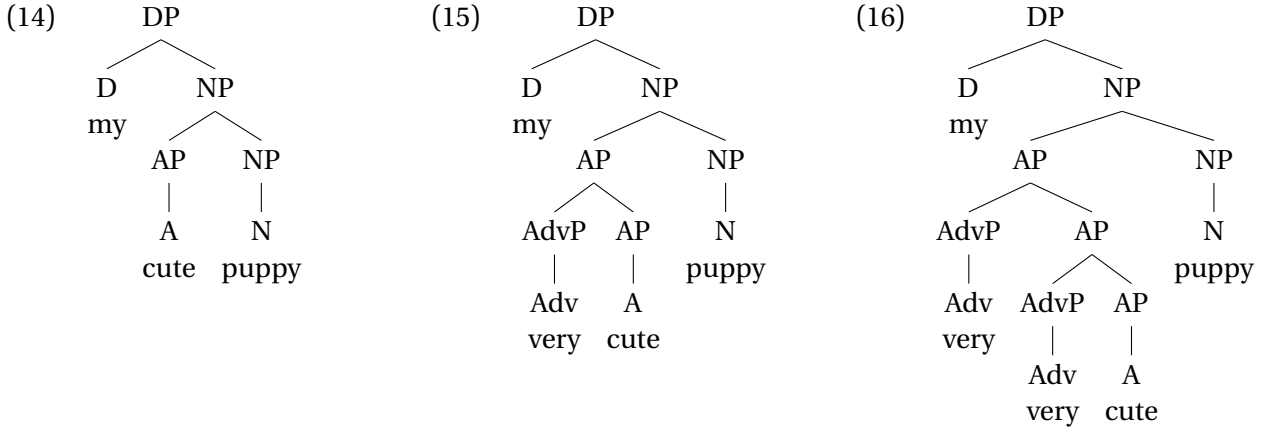
The usage of null determiners varies between languages. In French, for example, plural and mass DPs have a pronounced determiner in most cases. In Greek, personal names take a determiner. In many languages, all indefinite determiners are null — there was a time when English worked like this!

Adjuncts

Adjuncts, put simply, are optional modifiers of phrases. When an adjunct is added to a phrase, the result is a phrase of the same category. The stereotypical adjunct is the adjective, or rather, **adjective phrase** (AP), which combines with an NP to produce another NP (ex. 14)

The head of a phrase has no control over its adjuncts, unlike its complements — any AP can adjoin to any NP. You can also add an adjunct to an adjunct. For example, we can add an **adverb phrase** (AdvP) to an AP, then add the whole thing to an NP (ex. 15).

Furthermore, you can always add more adjuncts (ex. 16).



Sentences

Here is a list of all the head-complement rules we have seen so far:

- VP → V (DP) (DP) (PP)
- PP → P (DP)
- DP → D (NP)
- NP → N (PP)
- AP → A (PP)

- AdvP → Adv

Each of these rules is a shorthand for multiple rules with and without each element in parentheses. Next, here are the adjunction rules:

- NP → AP NP
- AP → AdvP AP

All of the rules look very similar, and this is by design. Our theory of phrase structure only includes heads, phrases, complements, and adjuncts — it doesn't allow random rules like XP → Y Y ZP W.

But we don't yet have a rule for an entire sentence. Revising the rule from *Grammar Rules!* to use DP rather than NP, we get:

- S → DP VP

This rule violates the theory that every phrase must have a head. But what kind of phrase is a sentence, and what could its head be? For now, we will have to let this remain a mystery.

Now we can draw trees for entire sentences, like the following:

(17)