Syntax

John Goldsmith

October 12, 2011

1 Syntax

It has long been recognized by linguists that the construction of a sentence is more than stringing a set of words together: there is a structure to it, one which is not usually indicated in the written form of the language but which is there for us to analyze. Starting in the 1940s, American linguists used ambiguous sentences — strings of words with two obviously different analyses—to drive this point home. Here are some examples of that; headlines are particularly good sources of funny ambiguous sentences:

1

Thanks to the morphology book by Mark Aronoff and Kirsten Fudeman.

British Left Waffles on Falkland Islands.
Miners Refuse to Work after Death.
Eye Drops Off Shelf.
Local High School Dropouts Cut In Half.
Reagan Wins on Budget, But More Lies Ahead.
Squad Helps Dog Bite Victim.
Juvenile Court to Try Shooting Defendant.
Kids Make Nutrious Snacks.

We will develop a method that will generate two analyses for these sentences, like the two below for the first example above:

Phrase structure rules (PSR)

The goal of syntax is to understand how we put words together to create well-formed, and meaningful, sentences. It is clear right from the start that we are looking at sequences of words: words occur one after another, in sequence. What are the principles governing the relative order of words in sentences?
Until the middle of the 20th century, thinking about this problem divided into two methods: in the first, individual words would be identified in the sentence by the role they played in a sentence. For example, in the sentence *Lee sent a birthday present to Kim*, *Lee* is the subject, *present* is the direct object, and *sent* is the verb. In the second approach, the sentence would be broken up into smaller and smaller pieces.

In the mid 1950s, this second analytic approach was stood on its head, and linguists began to write synthetic rules that generated pieces of sentences. These pieces could be as simple as a word, or it could be very complex. These rules were formulated—first by Noam Chomsky—in a way that was inspired by mathematical logic. For example,

(1) \( S \rightarrow NP \ VP \)

is a rule that says that an S[entence] can be expanded as an NP (a Noun Phrase) followed by a Verb Phrase. And we will have to immediately write some other rules to provide an answer to what those things are. We will expand VP in this way:

(2) \( VP \rightarrow \text{verb} \ NP \)

and we will expand NP in this way:

(3) \( NP \rightarrow \text{det} \ adj \ noun \)

We will distinguish between lexical categories, such as noun, adjective, and det, and phrasal categories, such as S, NP, or VP. Lexical categories are the most specific things that our syntax will delve into, at least at the beginning; and our phrase structure rules begin with an initial symbol (for now, S), which is expanded by means of phrase-structure rules, until the bottom categories of the tree that is created consists entirely of lexical categories; these lexical categories then are filled out with lexical items of the appropriate category (nouns, adjectives, and so on).

We will use lower case letters to specify lexical categories: this is not standard notation, but it is convenient.

We could write successive expansions in this way:

<table>
<thead>
<tr>
<th>expansion</th>
<th>the operative rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>( S \rightarrow NP \ VP )</td>
</tr>
<tr>
<td>NP VP</td>
<td>( S \rightarrow NP \ VP )</td>
</tr>
<tr>
<td>det adj noun VP</td>
<td>( NP \rightarrow \text{det adj noun} )</td>
</tr>
<tr>
<td>det adj noun verb NP</td>
<td>( VP \rightarrow \text{verb} \ NP )</td>
</tr>
<tr>
<td>det adj noun verb det adj noun</td>
<td>( NP \rightarrow \text{det adj noun} )</td>
</tr>
</tbody>
</table>

but it is much more common to draw this as a tree:

(4)
And this tree represents many millions of sentences, two of which are drawn here:

Big Idea: the motivation for positing the rule \( NP \rightarrow \text{det adj noun} \) is that this sequence appears several times in the description of the English sentence, and we can make the overall description more compact if we posit this entity, the ‘NP’.

The more times we are able to simplify our overall description by re-using a phrasal (non-lexical) category like \( NP \), the better we believe our analysis is motivated. So, for example, there is another VP-expansion that is motivated by examples like send a big present to the new teacher. Instead of accounting for this with a new VP-expansion rules

(5) \( VP \rightarrow NP \text{ prep det adj noun}, \)

we write instead:

(6) \( VP \rightarrow NP \text{ PP} \)

(7) \( PP \rightarrow \text{prep NP}, \)
where *prep* is a lexical category of prepositions that includes such words as *to*, *for* and *with*, and ‘PP’ marks a prepositional phrase. Thus the tree structure is not:

(8)

```
S
  /\
 NP  VP
 /  \
 det adj noun verb NP
    \
     det adj noun
```

but rather:

(9)

```
S
  /\ \
 NP  VP
 /  \
 det adj noun verb NP PP
     det adj noun
```

3 Alternative expansions of phrasal categories

We have just noted that there are two possible expansions for VP: (i) verb + NP and (ii) verb + NP + PP. In general, phrasal categories do have a lot of different, but related, ways of being expanded, and this fact is a central part of the motivation for talking about phrasal categories in the first place. Let us explore this.

Now, there is an implicit *independence assumption* made when we posit a category such as NP or VP: no matter where that node is generated by phrase-structure rules, any of its expansions may appear in that position. There is a lot that is right about that assumption; but it is by no means the whole story, and to be perfectly blunt about it, it is far from true: it is, indeed, false. False but helpful. ²

² Perhaps the first reference to this is in Pittman 1948: if we do not view a sentence as being hierarchically broken into parts, “one is almost compelled to regard every morpheme in an utterance as pertinent to the description of every other morpheme. But a good analysis in terms of immediate constituents usually reduces the total possible environmental factors of a given morpheme or sequence of morphemes to one: in other words, it states that the only pertinent environment of a given immediate constituent is its concomitant (the other immediate constituent).” (p. 287)

For example, let us consider several possible expansions for NP in English:

(10)
(i) $\text{NP} \rightarrow \text{noun}$ \quad \text{Bananas are a good source of potassium.}

(ii) $\text{NP} \rightarrow \text{det noun}$ \quad \text{My doctor told me to exercise more.}

(iii) $\text{NP} \rightarrow \text{adj noun}$ \quad \text{Easy melodies make for good songs.}

(iv) $\text{NP} \rightarrow \text{det adj noun}$ \quad \text{The old ways are the best ways.}

(v) $\text{NP} \rightarrow \text{det noun PP}$ \quad \text{The road to Hell is paved with good intentions.}

By positing these five different, but related, rules that expand \textit{NP}, we are saying that any \textit{NP}, any place in a sentence, can have any of those five structures. To repeat: that is not entirely true, but it is a good first step to take in approximating the way words are ‘distributed’ in English and in other languages.

It is often the case that we can simplify our analysis of a phrasal category by saying that a part of its expansion is \textit{optional}. Instead of saying that we have both rules (i) and (ii) above, we say that \textit{det} is optional, and the notation for that is a set of parentheses around the optional category:

\[(11) \quad \text{NP} \rightarrow (\text{det}) \text{ noun}.\]

Looking at all of the expansions given in (12xx), we would naturally be led to the conclusion that a better form of the \textit{NP} rule would be this:

\[(12) \quad \text{NP} \rightarrow (\text{det}) (\text{adj}) \text{ noun (PP)}\]

(Discuss the consequences: more expansions predicted now.)

4 \quad \textbf{Ambiguous sentences}

In analyzing ambiguous sentences, most of the time we assign two different syntactic structures, one with each of the intended interpretations, as we did with sentences (1a) and (1b), and in most of these cases, there are two or more words which are assigned different lexical categories in the two cases. In the sentence we considered, “Left” was a noun in the intended sense—perhaps a noun derived from a verb, but in any event, it referred to a political party, or a coalition of parties. In the unintended sense, “Left” was the main verb of the sentence, the past tense of the verb \textit{leave}. Our analysis, then, predicts that if we change the word “Left” into some other word, some word that is not both a verb and a noun, the sentence should become unambiguous and not funny at all. That is true: there is no humor in \textit{British Right Waffles on Falkland Islands}, or in \textit{British Leave Waffles on Falkland Islands}. The humor of the ambiguity arises out of the totally unexpected collision between two different syntactic structures, themselves the result of simple phrase-structure rules motivated by an enormous number of simple rules.

By the way: not all ambiguities are like that; one of the most over-used ambiguous sentences, \textit{I saw the man with the telescope}, is ambiguous in a strictly structural way. Is it the man with the telescope that I claim to have seen, or am I just talking about some
man and the fact that I looked at him through the telescope? These two senses correspond to two different syntactic structures:

We do not always know when an ambiguous sentence is syntactically ambiguous. Is they are married ambiguous? If not, where does the humor come from in They’re married, but not to each other? How about Kids make nutritious snacks? That is ambiguous, but it may not be syntactically ambiguous. And what about My father always beat me… at chess, at least?

Let’s consider another ambiguous sentence:
The second structure arises unambiguously if we put in some words that allow no other analysis — for example, if the sentence had been *squad helps dog find master*.

5 Constituents

Any string of words that is generated by a single phrasal node in a given sentence is called a *constituent*. To analyze a sentence is to assign a tree structure to it, and by doing so, to analyze a set of constituents in the sentence. A good part of syntactic analysis is finding the right constituency structure for a sentence (we sometimes say, the right tree structure).

The most direct way to apply tests for constituency is to use the independence assumption that I mentioned earlier: if a string of words is a constituent – an NP, let’s say – then it ought to be possible to use that string of words in other sentences that seems
structurally rather different. If a string of words if a direct object NP (the price of tea in Japan in the sentence we compute the price of rice in China), then it ought to be possible to put the same string of words in places where we are already pretty sure that NPs can appear, such as in subject position of a simple sentence, or as the object of a preposition:

(14) The price of tea in Japan drives economic conditions there.

(15) I don’t know much about the price of tea in Japan.

or other constuctions, such as the pseudo-cleft:

(16) What they study is the price of tea in Japan.

or the pseudo-cleft:

(17) It was the price of tea in Japan that was the most important factor, not the temperature in Seattle.

What does this test suggest about the constituency of The congregation sent the family flowers? Is the family flowers a constituent? The fact that the following strings of words are not good sentences suggests strongly that it is not a constituent.

(18)(a) *What they sent was the family flowers.

(b) *It was the family flowers that they sent.

6 More examples

A simple example illustrating constituent structure ambiguity:
Fireproof clothing factory burns to ground.

This headline is funny because there are two interpretations of fireproof clothing factory, and the more natural one (more natural if we only consider that phrase) is contradicted by the larger context, the sentence. The more natural interpretation is that it concerns a
clothing factory that is fireproof: fireproof then modifies (adds additional information to) clothing factory; clothing factory is a constituent in which clothing modifies factory, and together, clothing factory refers to the same kind of thing that the word factory does.

In short, when we analyze a noun phrase (roughly, a referring expression), one of the words within it expresses the type of thing that is referred to (here, factory). Typically, if any or all of the modifying material is be removed, the larger sense is vaguer but still roughly the same: factory burns to ground. Factory is said to be the head of the phrase Fireproof clothing factory: it is the element whose removal would most change the meaning of the phrase. The non-head element of a constituent is often called the modifier, or satellite. We know which structure is which in fireproof clothing factory because a non-head (or satellite) of a constituent C is not semantically modified by an element outside of that constituent. Structure (i) can be used to indicate a fireproof factory because factory is the head; that structure cannot be used to express a situation in which fireproof semantically modifies clothing.

English is relatively unusual in how poorly it marks nouns and verbs as distinct from a morphological point of view, and this can lead to multiple syntactic analyses. Time flies is famously ambiguous.

Verbs may take several arguments, and usually we can identify the different roles played by the arguments: consider I saw the man with the telescope.
A verb such as *see* has two arguments: roughly, the sighted person and the beheld object. In (i), the object is expressed with a 5-word expression, while in (ii) it is expressed with a 2-word expression. In (ii), however, an instrument, *the telescope*, appears, which modifies the *seeing* (rather than the object that is seen). It is freer to appear in different syntactic positions: *With his telescope, Galileo saw the craters on the moon.*

The interest of the headline: **GRANDMOTHER OF EIGHT MAKES HOLE IN ONE** relies on a structural difference: is [hole in one] a single item, or does it form two “sister constituents” in the verb phrase, as in *she put it in the bag* (or “...puts beans in nose”)?
Another nice way to sensitise oneself to syntactic structure is to look at garden-path sentences, like

- Fat people eat accumulates.
- The cotton clothing is usually made of grows in Mississippi.
- The girl told the story cried.
- The horse raced past the barn fell.
- I know the words to that song about the queen don’t rhyme.
7 Auxiliary verbs

One of the most impressive and influential of the early generative analyses of English was Chomsky’s analysis of the English auxiliary. Let’s consider a range of possible auxiliary verb combinations.

There is one thing that separates this data from the kind of data we have considered up to now. In the earlier examples, the choice of words that we made was essentially irrelevant; we included words by selecting nouns where the phrase structure rules generated “noun”, and likewise for the other categories. But here – each word or morpheme acts differently and uniquely. Why would we expect phrase-structure rules to work here? Either we will have actual words in our phrase-structure rules, or we will have to create categories that contain only a single item. The two pretty much boil down to the same thing.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Verb</th>
<th>Tense</th>
</tr>
</thead>
<tbody>
<tr>
<td>You</td>
<td>walk</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>walk-s</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>walk-ed</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>walk</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>walk-ed</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>has</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>walk-ed</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>walk-ing</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>walk-ing</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>may have walk-ed</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>has be walk-ing</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>may have be-en walk-ing</td>
<td></td>
</tr>
</tbody>
</table>

**Sentences with -ed:**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Verb</th>
<th>Tense</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>may have</td>
<td>walk-ed</td>
</tr>
<tr>
<td>John</td>
<td>has</td>
<td>walk-ed</td>
</tr>
<tr>
<td>John</td>
<td></td>
<td>walk-ed</td>
</tr>
</tbody>
</table>

**Sentences with -ing:**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Verb</th>
<th>Tense</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>is</td>
<td>walk-ing</td>
</tr>
<tr>
<td>John</td>
<td>may be</td>
<td>walk-ing</td>
</tr>
<tr>
<td>John</td>
<td>may have be-en</td>
<td>walk-ing</td>
</tr>
</tbody>
</table>

**Sentences with 3rd p. sg -s:**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Verb</th>
<th>Tense</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>walk-s</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>is</td>
<td>walk-ing</td>
</tr>
<tr>
<td>John</td>
<td>has</td>
<td>walk-ed</td>
</tr>
</tbody>
</table>

**Sentences with -do:**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Verb</th>
<th>Tense</th>
</tr>
</thead>
<tbody>
<tr>
<td>You</td>
<td>do</td>
<td>walk</td>
</tr>
<tr>
<td>John</td>
<td>does</td>
<td>walk</td>
</tr>
<tr>
<td>*John</td>
<td>does</td>
<td>walk-s</td>
</tr>
<tr>
<td>*John</td>
<td>does may have</td>
<td>walk-ed</td>
</tr>
<tr>
<td>*John</td>
<td>does has/ have</td>
<td>walk-ed</td>
</tr>
<tr>
<td>*John</td>
<td>does is/be</td>
<td>walk-ing</td>
</tr>
<tr>
<td>*John</td>
<td>does may be</td>
<td>walk-ing</td>
</tr>
<tr>
<td>*John</td>
<td>does may have be-en</td>
<td>walk-ing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>Verb</th>
<th>Tense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do</td>
<td>you</td>
<td>walk</td>
</tr>
<tr>
<td>Does</td>
<td>John</td>
<td>walk</td>
</tr>
<tr>
<td>May</td>
<td>John</td>
<td>walk</td>
</tr>
<tr>
<td>May</td>
<td>John have</td>
<td>walk-ed</td>
</tr>
<tr>
<td>Has</td>
<td>John</td>
<td>walk-ed</td>
</tr>
<tr>
<td>Is</td>
<td>John</td>
<td>walk-ing</td>
</tr>
<tr>
<td>May</td>
<td>John be</td>
<td>walk-ing</td>
</tr>
<tr>
<td>May</td>
<td>John have be-en</td>
<td>walk-ing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>Verb</th>
<th>Tense</th>
</tr>
</thead>
<tbody>
<tr>
<td>*You</td>
<td>not</td>
<td>walk</td>
</tr>
<tr>
<td>You</td>
<td>do not</td>
<td>walk</td>
</tr>
<tr>
<td>*John</td>
<td>not</td>
<td>walk-s</td>
</tr>
<tr>
<td>John</td>
<td>does not</td>
<td>walk</td>
</tr>
<tr>
<td>John</td>
<td>may not</td>
<td>walk</td>
</tr>
<tr>
<td>John</td>
<td>may not have</td>
<td>walk-ed</td>
</tr>
<tr>
<td>John</td>
<td>has not</td>
<td>walk-ed</td>
</tr>
<tr>
<td>John</td>
<td>is not</td>
<td>walk-ing</td>
</tr>
<tr>
<td>John</td>
<td>may not be</td>
<td>walk-ing</td>
</tr>
<tr>
<td>John</td>
<td>may not have be-en</td>
<td>walk-ing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>Verb</th>
<th>Tense</th>
</tr>
</thead>
<tbody>
<tr>
<td>You</td>
<td>were amaze-d</td>
<td></td>
</tr>
</tbody>
</table>

---
John was amaze-d.
John may be amaze-d.
John may have be-en amaze-d.
John has be-en amaze-d.
John is be-ing amaze-d.
John may be be-ing amaze-d.
John may have be-en be-ing amaze-d.

You were not amaze-d.
John was not amaze-d.
John may not be amaze-d.
John may not have be-en amaze-d.
John has not be-en amaze-d.
John is not be-ing amaze-d.
John may not be be-ing amaze-d.
John may not have be-en be-ing amaze-d.

Table 1: English auxiliary

Let’s try to extract some basic generalizations concerning this data:

• No sentence with two words from the group called modal verbs: may, can, will, would, may, should, shall is grammatical; but one word from this group can co-occur with the other auxiliary verbs, such as have, be.

• When auxiliaries appear, their left to right order is summarized by a table:

<table>
<thead>
<tr>
<th>Modal verb</th>
<th>have (perfective)</th>
<th>be (progressive)</th>
<th>be (passive)</th>
<th>verb</th>
</tr>
</thead>
</table>

• The auxiliary verb do does not appear when there is any other auxiliary present: any of the auxiliaries we are exploring. It only appears when there are no others.

• However, the auxiliary do can appear along with the possessive have and the real (not dummy) verb do: We do not have enough money to do that. Anyway, we do not do things like that.

• If the negative not is present, it appears after the left-most (i.e., the first) of all of these auxiliaries. And if we count the auxiliary do as belonging to this group (and we do!), then when there is a not, there must be an auxiliary.

Chomsky’s account in Syntactic Structures (1957) was essentially the following:
It’s a lot cleaner to the eye if we add some constituency:
What is the right way to think about this? Is it position of a morpheme in a string, or is it something else?
\[ S \rightarrow NP \ Aux \ VP \]
\[ Aux \rightarrow Tense(Modal)(have – en)(be – ing) \]

\[
\begin{align*}
\text{Tense} & \quad \{ \text{Modal} \\
\text{-en} & \quad \text{have} \\
\text{-ing} & \quad \text{be} \}
\end{align*}
\]

\[ : 1 \rightarrow 2 \rightarrow 1 \]

Figure 1: English auxiliary (after Chomsky 1957)

Figure 2: Tree generated by rules in Figure 1

Figure 3: What is responsible for affix choice?
Figure 4: After affix-hopping

```
S
  NP  Aux
  |    |    
  John Pres modal perf prog
    |      |      |      |
    may have be+en verb
      |      |      |      |
      drinki+ng noun
          |      |
          beer
```

Figure 5: When all that hops is Tense

```
S
  NP  IP  NP  IP
  John Tense VP John Tense VP
    |      |      |      |
    -s verb verb
      |      |
      walk walk-s
```
8 Constituents -2

8.1 NP Verb PP; NP Verb NP PP

The syntactic patterns NP Verb PP and NP Verb NP PP are very common patterns in English and other languages. Let’s take a look at several patterns of this general sort:

Peacock was born to hustle, bustle, jostle, and command, but he had as well a clear-eyed sense of who in the English mathematical establishment could be counted on, who counted in, and who counted out. David Berlinsky, One, Two Three, p. 93.
8.2  He climbed over the wall

(a) What did he climb over?
(b) Over what did he climb?
(maybe)
(c) Over the wall climbed the monkeys.
(d) Over the wall the monkeys climbed. (maybe)
(e) The wall was climbed over.
(maybe)
(f) This wall has never been climbed over.
(g) He climbed over it.
(h) He climbed over the wall and the hedges.

The (b) example—if it is grammatical—is evidence that over and its following object VP forms a constituent; in the metaphor of syntactic movement, a preposition would only move with its object. (c) (which is, I think, unquestionably grammatical) makes the same point, but in the context of a different construction. (e) is a passive, in which the object of over has been passivized; this suggests a tight syntactic relationship between over and the preceding verb climb, and if (e) is not great, (f) is, and it makes the same point regarding grammar. 3

8.3  She put her name on the door

(a) What did she put on the door?
(b) Where did she put her name?
(c) What did she put her name on?
(d) On the door, she put her name.
(e) On the door, she put her name; on her desk, she put her new title.

Movement:

3 The point is often made in relation to the contrast between This bed has been slept in and This bed has been slept under, where the first is much better than the second.
8.4 They turned out the light: A

Now, let’s consider the sentence They turned out the light, which is also of the form NP V P NP. Does this have the same structure? – that is, is it:

The first sign that this is not the same structure is that this structure is unavailable when we have it rather than the light (remember, this was fine with she put her name on it):

(21) • *They turned out it.

Figure 6: Wrong analysis!
• They turned it out.

8.5 to turn on X

(22) The lion turned on his trainer, and it was several minutes before he could be removed from the cage.

(23) (Not: ...turned his trainer on...)

(24) The detective turn on her radio, and it was several minutes before she could tear herself away from what she was hearing.

(25) (just as fine...The detective turned her radio on... )

Questions: Do we wish to assign different structures to these sentences, and if so, how? What do you notice about the stress or prominence of the word on in the two sentences?

8.6 They turned over the blanket.

Is this right?

We can still say:

(27) What did they turn over?

but not:

(28) *Over what did they turn?

or

(29) *It was over the blanket that they turned.

So there is no evidence of pied-piping, of the preposition ‘moving’ along with the following NP. So Over the blanket does not behave like a constituent. And we can say:

(30) They turned the blanket over.
What is the right structure for that sentence?

\[
S \\
\text{NP} \quad \text{VP} \\
\text{They} \quad \text{verb} \quad \text{PP} \\
\text{turned} \quad \text{NP} \quad \text{prep} \quad \text{the blanket} \quad \text{over} \\
\]

What do we find if the object is a pronoun?

4

(31) • They turned it/him over.
    • *They turned over it.

8.7 They rolled it over/they rolled over it.

(32)(a) They jumped over the box.

\[
S \\
\text{NP} \quad \text{VP} \\
\text{They} \quad \text{verb} \quad \text{PP} \\
jumped \quad \text{prep} \quad \text{NP} \quad \text{over} \quad \text{the box} \\
\]

(b) They jumped over the box, not the blanket.

\[
S \\
\text{NP} \quad \text{VP} \\
\text{They} \quad \text{verb} \quad \text{PP} \\
jumped \quad \text{prep} \quad \text{NP} \quad \text{over} \quad \text{NP} \quad \text{not} \quad \text{the blanket} \\
\text{the box,} \\
\]

4 These facts might remind us of the similar ungrammaticality of *They gave Mary it, alongside of the fine They gave Mary some.
(c) They jumped over the box, not over the blanket.

(d) They turned over the box.

(e) They turned over the box, not the blanket.

(f) **They turned over the box, not over the blanket.

8.8 They threw the garbage out the window.
(33)(a) They jumped over the box.
(b) They turned over the box.
(c) They jumped over the box, not over the shoes.
(d) **They turned over the box, not over the shoes.
(e) They turned over the box, not over the shoes.

Verbs: look Somewhere I have notes on look.
<table>
<thead>
<tr>
<th>Action</th>
<th>Object</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>put</td>
<td>the book</td>
<td>on the table</td>
</tr>
<tr>
<td>put</td>
<td>it</td>
<td>under the tree</td>
</tr>
<tr>
<td>put</td>
<td>it</td>
<td>over the sink</td>
</tr>
<tr>
<td>put</td>
<td>the coat</td>
<td>on</td>
</tr>
<tr>
<td>put</td>
<td>the coat</td>
<td>on the monkey</td>
</tr>
<tr>
<td>put</td>
<td>it</td>
<td>on</td>
</tr>
<tr>
<td>put</td>
<td>the coat</td>
<td>on the monkey</td>
</tr>
<tr>
<td>put</td>
<td>it</td>
<td>off</td>
</tr>
<tr>
<td>take</td>
<td>the coat</td>
<td>off</td>
</tr>
<tr>
<td>take</td>
<td>the coat</td>
<td>off the monkey</td>
</tr>
<tr>
<td>take</td>
<td>it</td>
<td>off</td>
</tr>
<tr>
<td>take</td>
<td>it</td>
<td>off the monkey</td>
</tr>
<tr>
<td>take</td>
<td>off</td>
<td>the coat</td>
</tr>
<tr>
<td>take</td>
<td>*off the monkey</td>
<td>the coat.</td>
</tr>
<tr>
<td>drink</td>
<td>the water</td>
<td>(all) up</td>
</tr>
<tr>
<td>drink</td>
<td>the water</td>
<td>up the water</td>
</tr>
<tr>
<td>drink</td>
<td>*all up</td>
<td>the water</td>
</tr>
<tr>
<td>drink</td>
<td>it</td>
<td>up</td>
</tr>
<tr>
<td>*drink</td>
<td>up</td>
<td>it</td>
</tr>
<tr>
<td>drink</td>
<td>the water</td>
<td>out of the bottle</td>
</tr>
<tr>
<td>?* drink</td>
<td>the water</td>
<td>up out of the bottle</td>
</tr>
</tbody>
</table>

9  **Productivity**

It is particularly striking that we can generate all day long sentences that we have never heard before, and yet which are fine sentences. We need to formulate principles that can account for that ability. The two most striking characteristics of syntax is the meaningfulness of the objects it accounts for (i.e., sentences), and the wide range of possible sentences each language generates.

10  **Word categories**

The classical Greeks gave us eight categories:

2. Verb (rhêma): no case inflection, but inflected for tense, person and number, indicating an activity or process
3. Participle: shares properties of verb and noun
4. Interjection
5. Pronoun: a substitute for a noun, and marked for person
6. Preposition: placed before other words
7. Adverb: not inflected, modifying a verb
8. Conjunction: a word that binds together parts of the discourse and filling gaps in interpretation.

Syntax as we know it is possible because a large number of generalizations about each language can be stated with respect to categories, rather than individual words or morphemes. No one has seriously proposed that we learn actual series of words: but do we learn what words can do in our language, or do we learn what categories of words can do? I suspect that most syntacticians would say that it is only the syntactic “behavior” (i.e., distribution) of categories of words that is of interest.

Most work in syntax is about generalizations that we can make about a given language with regard to entire categories, not individual words. And language is organized so that the same grammatical position can (most of the time) be occupied either by a single, simple word, or by an indefinitely large expansion.

10.1 Sentences

There can hardly be something called syntax if we do not recognize the existence of sentences in language: but it is difficult to define what a sentence is. Most serious efforts either approach the task distributionally (and employ the sentence as the unit that makes sense out of our intuitions of grammaticality) or semantically (a sentence is an expression of a proposition, a notion whose characterization can be passed to philosophers).

Otto Jespersen (1924): “A sentence is a (relatively) complete and independent human utterance—the completeness and independence being shown by its standing alone or its capacity of standing alone, i.e., of being uttered by itself.” Karl Sundén (1941) proposed, “A sentence is a portion of speech that is putting forward to the listener a state of things (a thing meant) as having validity, i.e., as being true.” This combines the first and the third approach (logical and social). Leonard Bloomfield (1933): “Each sentence is an independent linguistic form, not included by virtue of any grammatical construction in any larger linguistic form.”

11 Constituent structure

Constituent structure is the single most important notion in syntactic theory and analysis. It covers all relationships between the words of an utterance (typically within a sentence) that go beyond the notion of precedes and follows. We can be struck by the importance, the reality, and the significance of constituent structure when we are presented with ambiguous sentences (sometimes in
a joke). Not all ambiguity involves constituent structure, though: e.g., the ambiguity in *Kids make nutritious snacks* relies on different grammatical roles that the single constituent (*nutritious snacks*) can play in the larger sentence (*puppies make good friends, I made lunch*). Similarly, *My father always beat me...at chess, at least.* is vaguely funny, because of the switch it induces with regard to the role played by the direct object (or perhaps the meaning of the two homonyms *beat*?).

A simple example illustrating constituent structure ambiguity: *Fireproof clothing factory burns to ground.*

This headline is funny because there are two interpretations of *fireproof clothing factory*, and the more natural one (more natural if we only consider that phrase) is contradicted by the larger context, the sentence. The more natural interpretation is that it concerns a clothing factory that is fireproof: *fireproof* then modifies (adds additional information to) *clothing factory*; *clothing factory* is a constituent in which clothing modifies factory, and together, *clothing factory* refers to the same kind of thing that the word *factory* does.

In short, when we analyze a noun phrase (roughly, a referring expression), one of the words within it expresses the type of thing that is referred to (here, *factory*). Typically, if any or all of the modifying material is be removed, the larger sense is vaguer but still roughly the same: *factory burns to ground*. *Factory* is said to be the head of the phrase *Fireproof clothing factory*: it is the element whose removal would most change the meaning of the phrase. The non-head element of a constituent is often called the modifier, or satellite. We know which structure is which in *fireproof clothing factory* because a non-head (or satellite) of a constituent C is not semantically modified by an element outside of that constituent. Structure (i) can be used to indicate a *fireproof factory* because *factory* is the head; that structure cannot be used to express a situation in which *fireproof* semantically modifies *clothing*.

English is relatively unusual in how poorly it marks nouns and verbs as distinct from a morphological point of view, and this can
lead to multiple syntactic analyses. *Time flies* is famously ambiguous.

Verbs may take several arguments, and usually we can identify the different roles played by the arguments: consider *I saw the man with the telescope.*

A verb such as *see* has two arguments: roughly, the sighted person and the beheld object. In (i), the object is expressed with a 5-word expression, while in (ii) it is expressed with a 2-word expression. In (ii), however, an instrument, *the telescope,* appears, which modifies the *seeing* (rather than the object that is seen). It is freer to appear in different syntactic positions: With his telescope, Galileo saw the craters on the moon.

The interest of the headline: GRANDMOTHER OF EIGHT MAKES HOLE IN ONE relies on a structural difference: is *hole in one* a single item, or does it form two “sister constituents” in the verb phrase, as in *she put it in the bag* (or “...puts beans in nose”)?
Another nice way to sensitize oneself to syntactic structure is to look at garden-path sentences, like
• Fat people eat accumulates.
• The cotton clothing is usually made of grows in Mississippi.
• The girl told the story cried.
• The horse raced past the barn fell.
• I know the words to that song about the queen don’t rhyme.

\[
\begin{array}{c}
S \\
NP \quad VP \\
\quad NP \quad S \quad verb \\
\quad adj \quad NP \quad VP \\
\quad fat \quad noun \quad verb \\
\quad people \quad eat
\end{array}
\]

12 Some of the basic phenomena of interest to syntactians

1. Cases where word-order interacts clearly with logical scope of operators such as negation. For example, in English: Liberman 1975
   • i. With no job, John would be happy. If he had no job (= if he were unemployed), John would be happy.
   • ii. With no job would John be happy. There is no job such that it would make John happy (if it were given to him).

2. Basic word order: SVO and its permutations;
   Joseph Greenberg in 1966 drew attention to the fact that the order of constituents in sentences was not uniformly distributed among all the logical possibilities. Focusing on subject (S), object (O), and verb (V), studies (such as Ruhlen 1975) have found distributions along these lines: www.hku.hk/linguist
   VOS: Malagasy, Seediq (Austronesian)
   OSV: Kabardian (Northwest Caucasian)
   OVS: Apalai, Hixkaryana (Carib)

Pullum 1981

13 English: SVO

Subject-Verb-Object

*The police arrested E. Howard Hunt.*
14 **Japanese: SOV**

Japanese is a strictly verb-final language, with massive pro-drop and topic-marking (-wa). This combination is of great interest to many linguists.

*Tanaka-san wa ringo -o tabemasu*

Mr. Tanaka TOPIC apple DO eat

Mr. Tanaka eats the apple.

The preceding sentence would be a reasonable answer to the question: What does Tanaka-san eat? To answer, Who eats the apple?, you might say:

*ringo -wa Tanaka-san ga tabemasu*

apple TOPIC Mr. Tanaka SUBJ eat

Mr. Tanaka eats the apple.

Consider: 5

*Tanaka-san ga kono ie ni sunde imasu.*

Mr. Tanaka SUBJ this house in living is.

Mr. Tanaka is staying in this house.

*Tanaka-san wa sensei desu.*

Tanaka TOPIC teacher is.

Tanaka is a teacher.

sunde ← sum+te.

15 **German: mixed SVO, SOV**

**First approximation:** In main clauses, the finite verb appears in second position, and a major syntactic constituent precedes it. A separable prefix does not appear in second position, even if it is lexically associated with the verb that is in second position. When a series of verbs occurs in a single clause, the logically highest one is that which appears in second position. None of
this occurs in embedded clauses – or rather, in sentences with overt complementizers.

```
S
  NP  VP
  Er  V  NP  V  NP
  heisst  Rolf  heisst  Rolf
```

Roughly: The old man comes today home.

```
S
  NP  VP
  Der alte Mann  V  heute  nach Hause
  kommt
```

```
S
  NP  ist  VP
  Der alte Mann  heute  nach Hause  an - ge - kommen
```

Der alte Mann ist gestern angekommen.
Der alte Mann will heute nach Hause kommen.
Heute kommt der alte mann anch Hause.
Ich wiess nicht, wann er heute ankommt.

3. Movement, especially in formation of questions, and what is (and is not) possible

15.1 Question movement

Questions are used by speakers for particular ends: they request an appropriate answer. We distinguish yes/no questions from content questions (or in English wh-questions):
You have gone to court. Have you gone to court? When have you gone to court? Who did you talk to? Who did most of the talking? Why were you there?
4. Special syntactic positions of pronouns and short, unstressed elements

5. Different word orders in main and embedded clauses

6. Languages in which subjects (or arguments, more generally) may be left implicit if context permits (pro-Drop languages).