1. Constituency

The notion of constituency is central to modern syntax. It is best to think of constituency by looking at a tree derivation:

```
A
  p          B         C
    q          r        D    u
      s        t
```

The above tree derivation is for the string "pqrstu". However, as the tree shows, the string is not simply a flat sequence of symbols but contains structure. If we choose any node on the tree and treat that particular node as the root of the subtree, the substring that we obtain is considered a constituent.

For example, we choose the node B and make it the root of the subtree that it dominates (the portion thus selected is highlighted below):

```
A
  p          B
    q          r
```

Thus the string "qr" is a constituent of the string "pqrstu". Likewise, if we instead made C the root of another subtree, it would dominate the substring "stu", which forms yet another constituent:
The set of all constituent substrings in the above tree is:
\{ p, q, r, s, t, u, qr, st, stu, pqrstu \}
Note that individual terminal symbols count as constituents, as does the entire tree itself.

String like the following are NOT constituents because there is no single node that will
dominate all the symbols of the string to the exclusion of other symbols:

pqr, qrst, rs, tu, .....

**Constituency and Natural Language**

Why is the notion of constituency important in syntax? Primarily because we do not
simply think of sentences in terms of a flat sequence of words, but in terms of constituent
structure. What does this mean? Consider the following sentence:

John eats pizza.

There are three possible ways to group the words in this sentence. For example, we could
assume that "John eats pizza" has no internal structure whatsoever, in which case, one
word bears no hierarchical relation to any other and a flat tree structure like the following
will suffice:

```
S
  /   \\
John  eats pizza
```

Case 1

We can say that the above tree representation in Case 1 assumes no intermediate
customents between the full sentence "John eats pizza", and the individual words "John",
"eats" and "pizza".

Alternatively, we could assume that "John" and "eats" group together first into a
constituent, and then this intermediate constituent "John eats" is paired off with "pizza".
(See Case 2 below).

A third possibility is that "eats" and "pizza" are paired off first, and then the intermediate
constituent "eats pizza" combine with "John", as shown in Case 3 below.
In Case 2, "John eats" form an intermediate constituent between the full sentence "John eats pizza" and the individual words "John", "eats" and "pizza". In Case 3, "eats pizza" is the intermediate constituent. In Case 1, there are no intermediate constituents.

Which of the above three structural possibilities seem to be the salient one as far as our knowledge of the English language is concerned?

There are several ways that we could test for constituenthood.

**Substitution**

The most basic constituenthood test is the substitution test. The reasoning behind the test is simple. A constituent is any syntactic unit, regardless of length or syntactic category. A single word is the smallest possible constituent belonging to a particular syntactic category. So if a single word can substitute for a string of several words, then that's evidence that the single word and the string are both constituents of the same category.

For example, we know that pronouns can substitute for noun phrases. Some examples are given in (1).

(1) a. The little boy fed the cat. ---> He fed her.
     b. Black cats detest green peas. ---> They detest them.

It is important to understand that a particular string of words can be a noun phrase in one syntactic context, but not in another. For instance, the substitution test tells us that the underlined strings are noun phrases in (1), but not in (2).

(2) a. The little boy from next door fed the cat without a tail. ---> * He from next door fed her without a tail.
     b. These black cats detest those green peas. ---> * These they detest those them.

Rather, in these sentences, it is the longer underlined strings in (3) that are noun phrases.

(3) a. The little boy from next door fed the cat without a tail. ---> He fed her.
     b. These black cats detest those green peas. ---> They detest them.

Pronouns are not the only placeholder elements, or pro-forms. For instance, adverbs such as here or there can substitute for constituents that refer to locations or directions. As in the case of noun phrases, whether a particular string is a constituent depends on its syntactic context.

(4) a. Put it on the table. ---> Put it there.
b. Put it over on the table. --- Put it over there.
> 

c. Put it over on the table. --- Put it there.
> 

(5) a. Put it on the table that's by the door. --- * Put it there that's by the door.
> 

b. Put it over on the table that's by the door. --- * Put it over there that's by the door.
> 

c. Put it over on the table that's by the door. --- * Put it there that's by the door.
> 

The word so can substitute for adjective phrases (here, the most natural-sounding results are obtained in contexts of comparison). As usual, the same string sometimes is a constituent and sometimes isn't.

(6) a. I am very happy, ... ... and Linda is so, too.
   b. I am very fond of Lukas, ... ... and Linda is so, too.
   c. I am very fond of my nephew, ... * ... and Linda is so of her niece.

Finally, pronouns and sometimes the word so can substitute for subordinate clauses introduced by that, as illustrated in (7).

(7) a. I { know, suspect } that they're invited. ---> I { know, suspect } it.
   b. I { imagine, think } that they're invited. ---> I { imagine, think } so.

Movement

Substitution by pro-forms is not the only diagnostic for whether a string is a constituent. If it is possible to move a particular string from its ordinary position to another position---typically, the beginning of the sentence---that, too, is evidence that the string is a constituent. In order to make the result of movement completely acceptable, it's sometimes necessary to use a special intonation or discourse context, especially in the case of noun phrases. In the grammatical instances of movement in (8), appropriate discourse material is included in parentheses. "___" indicates the ordinary position that a constituent has moved from.

(8) a. I fed the cats. --- The cats, I fed ___. (The dogs, I didn't.)
> 

b. I fed the cats with long, fluffy tails.
   --- The cats with long, fluffy tails, I fed ___. (The other cats, I didn't.)
> 

Movement of constituents other than noun phrases is illustrated in (9).
(9) a. Prepositional phrase: The cat strolled across the porch with a confident air. -- With a confident air, the cat strolled across the porch ___.
b. Adjective phrase: Ali Baba returned from his travels wiser than before. -- Wiser than before, Ali Baba returned from his travels ___.
c. Adverb phrase: They arrived at the concert hall more quickly than they had expected. -- More quickly than they had expected, they arrived at the concert hall ___.

(10) shows the ungrammatical results of moving strings that aren't constituents.

(10) a. I fed the cats with long, fluffy tails. --- The cats, I fed ___ with long, fluffy tails.1
b. The cat strolled across the porch with a confident air. --- * With a, the cat strolled across the porch ___ confident air.
c. Ali Baba returned from his travels wiser than before. --- * Wiser than, Ali Baba returned from his travels ___ before.
d. They arrived at the concert hall more quickly than they had expected. --- * More quickly than they, they arrived at the concert hall ___ had expected.

Questions and sentence fragments

Another way to tell whether a string is a constituent is to see whether it can function as a sentence fragment in response to a question. The question itself also functions as a diagnostic test, since we can think of question formation as involving the substitution of a question word for a string and the subsequent movement of the question word. (11) illustrates this pair of tests for a variety of constituent types.

b. Prepositional phrase: How did the cat stroll across the porch? With a confident air.
e. Adverb phrase: How did they do? Not badly. Surprisingly well. Much better than they had expected.
Once again, attempting to question nonconstituents is ungrammatical.

   d. * How did they arrive at the concert hall ___ had expected? --- * More quickly than they.

So substitution for adjective phrases and subordinate clauses has a variant that is reminiscent of questions. In addition to just substituting for the string of interest, as illustrated earlier, so can move to the beginning of the sentence, which then undergoes subject-aux inversion—the same process that turns declarative sentences into yes-no questions. This variant of so substitution is illustrated in (13) and (14).

(13) a. I am very happy, ... and so is Linda.
   b. I am very fond of Lukas, ... and so is Linda.
   c. I am very fond of my nephew, ... * and so is Linda of her niece.

(14) I { imagine, think } that they're invited, ... * and so do they.

2. X-bar Theory

Above, we discussed that the lexicon provides the syntax with vocabulary items that come equipped with a syntactic category as well as a set of arguments. X-bar theory proposes that when a item from the lexicon enters the syntax, it projects an elementary tree. In this elementary tree, the projecting vocabulary item is the head, and the tree also contains appropriate locations for the arguments of the head:

```
XP
|  
|  
|  X'  
|  /  
|  X  
|  /  
specifier  
```

The leaf where the subject argument is located is called the specifier whereas the leaf where the object argument is located is called the complement. X is called the head of the projection where the projecting lexical item is located. The 'P' in XP stands for 'phrase'. Think of X as a variable - it will assume the value of a syntactic category, depending on the lexical item which projects the tree. The X-bar structure captures the
intuition that the head and its object are far more ‘closely’ related relative to the subject argument, hence the asymmetry in the relative locations of specifiers and complements.

As an example, here is the specification for the lexical item "eat". This item has a syntactic category "V" (which stands for verb) as well as two arguments a subject and an object, both of which are noun-like elements (let's call them NP, for nominal phrase). Thus, when "eat" enters the syntax, it immediately projects a structure like this:

```
VP
   NP
     V'
      V
      NP
       eat
```

Therefore, from its humble beginnings as a simple lexical item, "eat" is transformed into a full elementary tree (a VP tree, which stands for verbal phrase) when it enters the syntactic module.

Note that the complement and specifier branches in an X-bar structure are optional. Their presence or absence depends on the argument structure of the projecting lexical item. For example, the verb "sleep" only takes one subject argument, which is a nominal phrase. Its projection into the syntax will therefore look like this:

```
VP
   NP
     V'
      V
      sleep
```

Nouns often don't have arguments. For example, the lexical item "John" is listed as a noun (denoted 'N') which takes no argument. When John enters the syntax, it just projects an X-bar tree with no branches:

```
NP
   | 
   N'
   | 
   N
   | 
John
```
Substitution

We build larger units out of these elementary trees just as we did with Regular Grammars or Context-Free Grammars (in fact, X-bar structures are a type of CFG), that is by substituting at the appropriate nodes. So, to derive the sentence "John sleeps", we plug in the NP projection of "John" into the specifier location of VP:

```
  VP
   | 
  NP V'
   | 
 N'  V
   | 
   N  sleeps
     | 
     John
```

Auxiliaries

Auxiliaries are lexical items like "will", "might", "do", "may" which often precede another verb. For example, "John will eat pizza". These auxiliaries often carry information about tense, aspect (indicating whether an action is ongoing or completed) or some modality (indicating the possibility of something happening). Consistent with every other lexical item, an auxiliary also projects an elementary tree in X-bar theory. For example, an auxiliary like "will" is labeled I (for "Inflection", since tense, aspect and mood markers are often heavily inflected in many languages) and often take a VP as a complement:

```
  IP
   | 
   I'
   | 
   I VP
   | 
   will
```

When we stick another elementary tree projected as VP, we get tree derivations of the following sort:

```
  IP
   | 
   I'
   | 
   I VP
   | 
   will
```
Now, this gives us the derivation "Will John eat pizza". This is fine as a question, but we were really trying to derive a declarative sentence here. What has gone wrong??

This is where parameters come into play!

The Subject Movement Parameter

We first note that a string like "Will John eat pizza" is perfectly grammatical as a declarative sentence in a language like Welsh. For example:

Naeth y dyn brynu car
Did the man buy car
"The man did buy a car."

Other languages like Welsh include the other Celtic languages (Irish, Breton, etc.) as well as far-flung non-related ones like Classical Arabic, Chamorro (spoken in the Pacific), Salish (spoken in British Columbia), etc..
Others, like English or French or Indonesian (spoken in South-East Asia) or Edo (spoken in Ghana) seem to do something with the subject argument of their verbs... the proposal is that the subject is moved to the specifier of IP, leaving a trace (t) in its original position:

One thing to note about the auxiliary head at this point. In English, sometimes there is no overt auxiliary expression. While the future tense is expressed as "John will eat pizza", the present is simply "John eats pizza" with no auxiliary preceding the verb. In this case, we still assume that there is an auxiliary tree present, except that it contains a silent auxiliary inflection which we will denote as something like [+pr] for "present tense":

We maintain that there exists a silent [+pr] form projecting an IP elementary tree for reasons of uniformity. Note too that this [+pr] inflection can have an equivalent overt expression in English: John DOES eat pizza, in which case "does" goes under I. A similar
reasoning holds for past tense, we assume a silent [+pt] "past tense" inflection for the sentence "John ate pizza". Note too the equivalent expression "John did eat pizza"....

Back to subject movement. Why movement? Is this not dubious? Syntacticians have constructed some reasonable arguments that motivate the movement of the subject to the Specifier of IP in languages that have such movements (for reasons related to case and agreement, etc.). We will not get into this elaborate topic in this class, but it is good to keep this in mind and explore the validity of this issue if you have the interest and time.

Another simple argument why movement might be plausible is simply that things DO move in language. We do it all the time in speech for effect, "I like this book" can be transformed to, "Well, THIS BOOK I like!". Also, patterns change over time... verbs used to come AFTER objects in Old English but not anymore in Modern English. So having a theory of movement in syntax does not seem so out of place or ad hoc given the empirical evidence.

We therefore have our first parameter, which I will simply call the SUBJECT MOVEMENT PARAMETER - languages either move subjects to specifier of IP or keep them inert in the specifier of VP where they were initially projected....

Adjunction

Verbal phrases can be modified by adverbs, for example, one can say "John rarely eats pizza". These adverbs are normally considered *adjuncts* in the sense that they do not intuitively seem to be part of the argument structure of the verb "eat", but is added on to modify the meaning of the verbal phrase.

How is modification by an adjunct represented in X-bar syntax? The proposal is to introduce a new process called adjunction, with elementary trees of the following type:

\[
\begin{array}{c}
X' \\
\text{adjunct} \\
X'
\end{array}
\]  

(in some cases, the adjunct could follow the lower X’ too)

Presumably, the lexicon stores the information for each adjunct regarding the type of syntactic categories which it modifies. So when an adverbial adjunct like "rarely" is sent from the lexicon into the syntax, it projects a structure like the following:

\[
\begin{array}{c}
V' \\
rarely \\
V'
\end{array}
\]

An adjunction structure like the above is *spliced* into the appropriate location in a verbal phrase:

\[
\text{IP}
\]
For this class, we will also assume that a negated element like 'not' is an adjunct, so it also splices itself onto some \( X' \). Here's some more information about adjunct vs argument distinctions: http://www.ling.upenn.edu/~beatrice/150/ch4.html#comp-adj

**Verb Movement Parameter**

Let us now discuss a second parameter - in some languages, the verb moves to the I node and in others, the verb stays inert. The initial argument for this parameter came from a comparison between English and French. In English, as the example above showed, we say stuff like: "John rarely eats pizza".

The equivalent expression in French is:

Jean mange rarement le pizza.
John eat rarely pizza
"John rarely eats pizza."

Notice the location of the adverb. The adverb preceding the verb is ungrammatical in French. This variation has been adduced as evidence of verb movement in French:
One way to think about verb movement is that the verb moves to 'merge' with the inflection node to check its inflectional properties. Verbs in French are more inflected (conjugated) compared to English verbs, which lends some weight to the verb movement hypothesis.

A nice empirical consequence of this is to look again at Welsh data. Earlier we saw a sentence in Welsh like the following:

Naeth y dyn brynu car  
Did the man buy car  
"The man did buy a car."

Like English, Welsh can also drop the overt auxiliary ("did" in this case) and simply inflect the verb "buy" itself for the past tense. But if the overt auxiliary is dropped, it now means that there is an option for the main verb in V to move to the phonologically empty I. What does Welsh do? According to the verb-movement parameter, Welsh essentially has two options, it can move the verb to I, or keep it in V.

Here is what the sentence without the overt auxiliary looks like:

Bryn-odd y dyn gar  
Buy-PAST the man car

The verb moves to I and merges with the past tense marker ‘odd’. Remember too that Welsh has a different setting in its subject-movement parameter compared to English. The net effect of all this is a verb-subject-object order:
VERB MOVEMENT PARAMETER: languages either move a verb to the I node or keep
the verb inert in its original location

Head Directionality Parameter

A final parameter that we will discuss is the Head Directionality Parameter. Unlike the
previous two parameters which we looked at, this one does not concern movement
possibilities. It is a statement about variations in the projection of the X-bar structure. It
was noted earlier that an X-bar tree has the following structure:

```
XP
   /|
  / \\
subject X' complement
   |
   X

X
  /|
 / |
head object argument
```

The head-directionality parameter suggests that the precedence relation between the head
X and its complement is variable.
In languages like English, French or Welsh, the head precedes the complement. These
languages are known as head-initial languages.
On other other hand, a number of languages seem to project a slightly different but
analogous X-bar structure:

```
XP
   /|
  / \\
subject X' head
   |
   X

X
  /|
 / |
object complement argument
```

In the above tree, the head follows the complement. These languages are known as head-
final languages and include the likes of Japanese, Basque, Navajo, Turkish, Hindi, etc..
On average, there is an equal number of head-initial and head-final languages.
Here are some comparisons of headedness between English and Japanese:

<table>
<thead>
<tr>
<th>Verb-Object</th>
<th>Object-Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>John eats apples</td>
<td>John-wa ringo-o taberu</td>
</tr>
<tr>
<td></td>
<td>John        apple eat</td>
</tr>
</tbody>
</table>
"John eats apples"

*Verb-Prepositional phrase*  
Preposition-Noun  
John goes to school  
John-wa *gakkoo-ni iku*  
John  
*school-to*  
*go*

*Auxiliary-Verb Phrase*  
Verb Phrase-Auxiliary  
John may go to school  
John-wa *gakkoo-ni iku-kamoshirenai*  
John  
*school-to*  
*go*  
*may*