An Overview of the Grammar of English
Outline

- Grammatical, Syntactic and Lexical Categories
  - Parts of Speech
- Major Constituents
  - Noun Phrases
  - Verb Phrases
  - Sentences
- Heads, Complements and Adjuncts
Grammatical Categories

- The dimensions
  - along with constituents can vary, and
  - to which the grammar of the language is sensitive, are call grammatical categories.
- E.g., in English, nouns and demonstratives have a “number” property.
  - These have to agree (“this book”, “*these book”).
  - We must mark nouns for number, even if it is irrelevant.
- Grammatical categories tend to be grammaticized semantic/pragmatic distinctions.
  - The number across all languages is very small.
- Other frequently occurring grammatical categories are gender, case, tense, aspect, mood, voice, degree, and deictic position.
Syntactic Categories

- These are the formal objects we will associate with constituents.
- Traditionally, they are the non-terminals of our grammar.
  - As such, they are atomic, unanalyzed units.
  - However, most theories today give them some structure, making them a bundle of grammatical categories.
    » We will return to this point later.
Lexical Categories

- Most words of most languages fall into a relatively small number of grammatically distinct classes, called
  - lexical categories or
  - parts of speech (POS), or
  - word classes
- The lexical category describes the syntactic behavior of a word wrt the grammar.
- These correspond to pre-terminals in a grammar, i.e., non-terminals that appear on the left-hand side of those rules that have terminals on the right.
- Most (other) grammar rules will make reference only to POSs, and not to individual words.
Classes of Lexical Categories

- **Useful to divide POSs into two groups:**
  - **Open classes**
    - let new words into them rather casually
    - and, therefore, tend to be very large.
    - Major ones are noun, verb, adjective and adverb.
  - **Closed classes**
    - change very little
      - Indeed, to a closed class is viewed as language change.
    - include "function" words, i.e., terms of high grammatical significance
    - Examples are prepositions, pronouns, conjunctions.
**What Are They?**

- Traditional grammar tells us that European languages have eight.
- Today, a few more are generally recognized by linguists.
- There isn’t complete consensus on what these are
  - but there isn’t a large divergence either.
  - There is some disagreement about exactly what should go in which category.
- However, when we actually develop a grammar, it can be argued that we will need many more distinctions than these provide.
- And, often, pragmatically-oriented computer scientists postulate lots more POSs than would be linguistically justified.
# A More or Less Typical Modern List of (Basic) Lexical Categories

<table>
<thead>
<tr>
<th>Noun</th>
<th>Preposition</th>
<th>Foreign words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>Determiner</td>
<td>Possessive marker</td>
</tr>
<tr>
<td>Adjective</td>
<td>Pronoun</td>
<td>Punctuation</td>
</tr>
<tr>
<td>Adverb</td>
<td>Conjunction</td>
<td>Symbol</td>
</tr>
<tr>
<td></td>
<td>Subordinator</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Complementizer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intensifier</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Infinitive marker</td>
<td></td>
</tr>
</tbody>
</table>
Some of these (specifically, symbol and punctuation) are just for written language.
- Similarly, “possessive marker” is just a tokenizing artifact.

All of these have important (i.e., grammatically significant) subclasses.
- Some are true subtypes
- Some are classes we can create by deciding to include other grammatical category distinctions within the lexical category.
- Whether or how we include the subclasses is a major source of variation.
Nouns

- Nouns have a number of differentiating dimensions:
  - Proper vs common
    » Proper nouns are “Jan”, “Moscow”, “New York City”?
  - Singular vs plural (the “number” grammatical category)
    » boy, boys, man, men
  - Count vs mass
    » “too many cats”, “too much water”
    » “Wine can be red or white.”, “Tigers have stripes.”
Verbs

Types
- auxiliary (closed)
  » List: do, have
- modal (closed)
  » List: can, might, should, would, ought, must, may, need, will, shall (dare?)
  » copula (List: be)
- main (open)
Verbs (con’t)

Verbs have lots of forms:
- Finite forms:
  » Can be the only verb in a sentence
  » Tends to have lots of (morphological) markings bearing lots of information.
- Non-finite forms:
  » Doesn’t show any variation.
Finite Verb Forms

- Always marked for tense.
- May carry other “agreement markers”
  - E.g., person, number
- Tenses
  - Present
    - Examples:
      - {I/we/you/the girls/they} {hit, go, cry};
      - {He/the girl} {hits, goes cries}
      - I am; {You, we, they, the boys} are; He is.
  - Past
    - Examples:
      - {I/we/you./the girls/he/the boy} {hit, cried, went}
      - {I,he, the boy} was; {We, you, the girls} were
Non-Finite Verb Forms

- Infinitive
  - The “base”, in English.
  - E.g., be, go, hit, cry

- Participles: Verbs qua modifiers (or to make an aspect)
  - Present (imperfective) participle
    » He {is, was, has been, will be} crying
    » The woman lighting the cigarette ...
  - Past (passive) participle
    » The boy rescued from the well....
    » The man, {exhausted, gone for three weeks,}
  - Perfect participle (not quite the same thing)
    » He {has, will have, had} {cried, been, gone}
    » Always the same as the passive participle in English.
Gerunds, BTW

- Note that you can use the imperfective participle as a so-called “verbal noun”:
  Throwing stones at glass houses can be hazardous.

- This is called a gerund.
  - It looks like a verb internally, but a noun externally.

- Note there is an “more nominal” form:
  The throwing of stones at glass houses ...
  - This uses the same base form, but internally it looks just like any other NP.
Determiners

- **Types**
  - articles: the, a, (unstressed) some
  - demonstratives: this, that
  - possessives: my, your
  - quantifiers: many, few, no, some
  - misc.: either, both, and maybe, which:
    - No matter which door you chose, you lose.
    - The plane landed, at which time, the passenger disembarked.

- Some propose that quantifiers are a separate lexical category.
Pronouns

Types:
- Personal (you, she, I, it, me)
- Reflexive (herself)
- Demonstrative (this)
- Indefinite (something, anybody)
- Wh-pronouns (what, who, whom, whoever)
  » which sometimes divided into interrogative (when used in questions) and relative (e.g., which, in relative clauses)

Note that so-called “possessive pronouns” (my, your, his, her, its, one’s, our, their) are more properly regarded as determiners
- Sometimes called possessive adjectives
One commonly distinguish a class called particles.

In English, these combine with verbs to make so-called phrasal verbs:

- Jan threw up
- made up that story
- looked the word up
- put me down.

However, they are identical with the set of English prepositions.

So it is appealing to think of these as prepositions without complements.
Adverbs

- Types
  - manner (quickly, rarely, never)
  - directional/locative (here, home, downtown)
  - temporal (now, tomorrow, Friday)
  - WH-adverbs (when, where, why)

- The different subtypes have very different syntactic properties.

- Traditionally, there is another subtype:
  - degree (very, extremely, so, too, rather)

- Most linguists prefer to have a degree modifier or intensifier word class, rather than include these as adverbs.
Conjunctions

- Traditionally, the following distinctions were made:
  - Coordinating conjunctions (and, or, but) join elements of equal status.
  - Subordinating conjunctions (or subordinators) introduce adverbial clauses (before, after, when, while, if, although, because, whenever)
    » Many regard these as specialized prepositions.
  - Complementizers (that, whether)
- Most linguists today prefer to give subordinators and complementizers their own categories.
Outliers?

Some regard the following as separate categories:
- politeness markers (please, thank you)
- greetings (hello, goodbye)
- “Existential there“:
  There is only one even prime number.
  There are a couple of points I’d like to make.
POS Tag Sets

- While these are the distinctions that are linguistically justified, we sometimes make up “tag sets” that are much larger.

- The justification is pragmatic.
  - The tags will often be used just by themselves, and for some kind of task, so one is free to make what distinctions one finds useful.

- E.g., the Penn Treebank has 45; the C7 tag set 146.
# The Penn Treebank Tag Set

<table>
<thead>
<tr>
<th>tag</th>
<th>description</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>coord. conjunction</td>
<td>and, but, or</td>
</tr>
<tr>
<td>CD</td>
<td>cardinal number</td>
<td>one, two, three</td>
</tr>
<tr>
<td>DT</td>
<td>determiner</td>
<td>a, the</td>
</tr>
<tr>
<td>EX</td>
<td>existential there</td>
<td></td>
</tr>
<tr>
<td>FW</td>
<td>foreign word</td>
<td>a propos</td>
</tr>
<tr>
<td>IN</td>
<td>preposition/sub-conj</td>
<td>of, in, by, if</td>
</tr>
<tr>
<td>JJ</td>
<td>adjective</td>
<td>small</td>
</tr>
<tr>
<td>JJR</td>
<td>adj., comparative</td>
<td>smaller</td>
</tr>
<tr>
<td>JJS</td>
<td>adj., superlative</td>
<td>smallest</td>
</tr>
<tr>
<td>LS</td>
<td>list item marker</td>
<td>1, one</td>
</tr>
<tr>
<td>MD</td>
<td>modal</td>
<td>can, should</td>
</tr>
<tr>
<td>NN</td>
<td>noun, sing. or mass</td>
<td>sand, car</td>
</tr>
<tr>
<td>NNS</td>
<td>noun, plural</td>
<td>cars</td>
</tr>
<tr>
<td>NNP</td>
<td>proper noun, sing.</td>
<td>Jan, Mt. Etna</td>
</tr>
<tr>
<td>NNPS</td>
<td>proper noun, pl.</td>
<td>Giants</td>
</tr>
<tr>
<td>PDT</td>
<td>predeterminer</td>
<td>all, both</td>
</tr>
<tr>
<td>POS</td>
<td>possessive ending</td>
<td>'s</td>
</tr>
<tr>
<td>PP</td>
<td>personal pronoun</td>
<td>I, me, you, he</td>
</tr>
<tr>
<td>RB</td>
<td>adverb</td>
<td>oddly, ever</td>
</tr>
<tr>
<td>RBR</td>
<td>adverb, comparative</td>
<td>quicker</td>
</tr>
<tr>
<td>RBS</td>
<td>adverb, superlative</td>
<td>quickest</td>
</tr>
<tr>
<td>RP</td>
<td>particle</td>
<td>up, on</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>tag</th>
<th>description</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYM</td>
<td>symbol</td>
<td>+, %, \</td>
</tr>
<tr>
<td>TO</td>
<td>“to”</td>
<td></td>
</tr>
<tr>
<td>UH</td>
<td>interjection</td>
<td>hmm, tsk</td>
</tr>
<tr>
<td>VB</td>
<td>verb, base form</td>
<td>bite</td>
</tr>
<tr>
<td>VBD</td>
<td>verb, past tense</td>
<td>bit</td>
</tr>
<tr>
<td>VBG</td>
<td>verb, gerund</td>
<td>biting</td>
</tr>
<tr>
<td>VBN</td>
<td>verb, past participle</td>
<td>bitten</td>
</tr>
<tr>
<td>VBP</td>
<td>verb, non-3sg pres</td>
<td>bite</td>
</tr>
<tr>
<td>VBZ</td>
<td>verb, 3sg pres.</td>
<td>bites</td>
</tr>
<tr>
<td>WDT</td>
<td>Wh-determiner</td>
<td>which, that</td>
</tr>
<tr>
<td>WP</td>
<td>Wh-pronoun</td>
<td>who, what</td>
</tr>
<tr>
<td>WP</td>
<td>possessive wh-</td>
<td>whose</td>
</tr>
<tr>
<td>WRB</td>
<td>Wh-adverb</td>
<td>how, where</td>
</tr>
<tr>
<td>$</td>
<td>dollar sign</td>
<td></td>
</tr>
<tr>
<td>#</td>
<td>pound sign</td>
<td></td>
</tr>
<tr>
<td>&quot;</td>
<td>left quote</td>
<td></td>
</tr>
<tr>
<td>&quot;</td>
<td>right quote</td>
<td></td>
</tr>
<tr>
<td>(</td>
<td>left paren</td>
<td></td>
</tr>
<tr>
<td>)</td>
<td>right paren</td>
<td></td>
</tr>
<tr>
<td>,</td>
<td>comma</td>
<td></td>
</tr>
<tr>
<td>:</td>
<td>sentence-final punc.</td>
<td>&gt; ! ?</td>
</tr>
<tr>
<td>:</td>
<td>mid-sentence punc.</td>
<td>; ... -- -</td>
</tr>
</tbody>
</table>
The Major Constituents

- These syntactic categories are may be thought of as “bigger” versions of lexical categories:
  - Noun phrase (NP)
  - Verb phrase (VP, S)
  - Prepositional phrase (PP)
  - Adjective phrase (AP)
  - Adverbial phrase (ADVP)
The Noun Phrase

- We can build NPs by
  - preceding a N, recursively, with different constituents
  - following an NP with other constituents.
Noun Phrase: Preceding the Noun

- We can build NPs by preceding a N with
  - one or more APs:
    small apple, very small apples, small green apples
  - one or more NPs (nominal compounds):
    heavy [cigar smoker]
    [Cuban cigar] smoker
    [gas meter] [turn-off valve]
  - quantifiers, determiners, predeterminers:
    a book, the books, that book, my book, few books
    those few books, the many books
    the books
    very many books
    all the gold, half the books, quite a few silver coins
Need to Capture Some Ordering Constraints

- We can say things like
  “two small cigars”
  “first constitutional amendment”
  “most small cigars”

but not
  “*small two cigars”
  “*constitutional first amendment”
  “*small most cigars”

- Let’s create a syntactic category Q for things like “many”, “very many”, “two”, and “more than two but less than three”, etc.

Note also that “the smallest(er) two cities” is okay, so we have to handle these elsewhere!

- We can create a lexical category, predeterminer, to accommodate “half the gold”, “all the books”, and “quite a few silver coins”.
  - Or make determiners more structured.
An Approximate Grammar (so far)

- The following captures what we have said thus far:
  \[ NP \rightarrow (PDT) (D) (Q) AP^* NP^* N \]

- Note that
  - “\( X^* \)” is just a shorthand for
    \[ Xs \rightarrow \varepsilon \]
    \[ Xs \rightarrow X \]
    \[ Xs \rightarrow Xs X \]
  - “\( X \rightarrow (Y) Z \)” is an abbreviation for
    \[ X \rightarrow Z \]
    \[ X \rightarrow YZ \]
An Approximate Grammar, Redux

- However, most analyses have more embedded constituent structure.
- So, a somewhat better set of rules might be the following:

\[
\begin{align*}
\text{NP}_{\text{min}} & \rightarrow \ N \mid \text{NP}_{\text{int}} \text{NP}_{\text{min}} \mid \text{PP} \text{NP}_{\text{min}} \\
\text{NP}_{\text{int}} & \rightarrow \ (Q) \ A\text{P}^{*} \text{NP}_{\text{min}} \\
\text{NP}_{\text{max}} & \rightarrow \ ( (PDT) \ DP \ ) \text{NP}_{\text{int}}
\end{align*}
\]
Noun and PP Compounds

- We allow NPs to be modified by PPs, especially particles:
  - “up elevator button”
  - “elevator up button”

  and more speculatively:
  - “a special [up] to the roof button”
  - “those in the bag deals”
A Possible “Determiner Phrase”

- DP ⟷ D | NP_{max} Poss-marker | D (Q) (Comparative* | Superlative*)

- E.g.:
  - “the”, “that”, “my”
  - “John’s”, “college professor’s (law suit)”
  - “the two smallest/smaller (big cities)”
  - maybe a few others...
Is * Really CFG?

- Note that with *, a single node can have an indefinite number of children.
- With pure CFG, this is not the case.
- So, this is an instance in which the notations are weakly, but not strongly, equivalent!
In addition to being able to generate “two man blobsled event”
the grammar also generates “most men blobsled event”
Whether this sort of thing is a syntactic or semantic/pragmatic issue is the subject of debate.

In general, it is tempting to think that the grammar of noun phrases can be made simpler, and that at least some of these constraints can be explained semantically.

- Exactly how to do so is not always clear.
Personal pronouns
- can be NPs all by themselves.
  \[ \text{NP}_{\text{min}} \rightarrow \text{ProP} \]
- and can join with NPs:
  » “We few survivors”; “You worse than senseless things”
  » “All us chickens”

Perhaps include these as determiners?

Proper nouns
- can be NPs all by themselves.
- and can form some bigger NPs: “poor little Rosie” and “the Jan I knew”)

So we could add a rule such as:
  \[ \text{NP}_{\text{min}} \rightarrow \text{ProperN} \]
Gerundive phrases can also be nouns. E.g.:
I enjoy watching television.
Watching television rots your brain.

So we could just add:

\[ \text{NP}_{\text{int}} \rightarrow \text{GrvP} \]

However, recall that, in English, gerunds are identical with imperfective participles.

- Moreover, below, we will introduce an imperfective reduced relatives clause, which is internally identical to a gerundive phrase.

So, it might be better to add:

\[ \text{NP}_{\text{int}} \rightarrow \text{RC}_{\text{imperfective}} \]
Noun Phrase: Following the Noun Phrase

- We can build a bigger NP by following an NP with one of the following:
  - prepositional phrases
  - relative clauses
  - infinitive clauses
In Terms of Our Grammar

- We can add these rules:

  NP → NP PP
  "the man on the moon"

  NP → NP RC
  "the gun (that) the man shot the victim with"

  NP → NP RC_{passive}
  "the gun used in the crime"

  NP → NP RC_{imperfective}
  "the man pointing the gun at you"

  NP → NP infC
  "the guy to go to in a pinch"
Which “NP” are we talking about here?

Consider
“most baguettes from the Cheese Board”,
This should probably be analyzed as
“[most [baguettes from the Cheese Board]]”

Also
“a package from overseas delivery”
is okay.

So, this looks like “NP$_{int}$”.
Following the Noun: Odds and Ends

- **Appositionals:**
  "the Senator from Arizona, John McCain", "Jan and Pat Shmoe, 123 Euclid Avenue, Berkeley"

  So add
  \[ NP \rightarrow NP \ , \ NP \]

- **Consider also**
  "our fine resort, on the Rogue River,"

  So add
  \[ NP \rightarrow NP \ , \ PP \]

- **There are some post-nominal adjectives:**
  - "arms akimbo", "I alone", "attorneys general"

- **And a more general post-nominal adjective construction:**
  - "love false or true", "children 8 years old or younger"
And, Finally, Coordination

- **Conjunction:**
  
  Dorothy, the tin woodman, and the scarecrow

  So add

  \[ NP \rightarrow NP^+ \text{ Conj } NP \]

- **Note this allows**

  “a pig in a poke and a cat in the bag”

  as well as

  “the boy and girl”
We’ve Missed Some Important Issues, Though

- Note that some nouns can stand by themselves as a noun phrase, while others need help:
  - Jan likes (tall) boys.
  - Jan likes {a, the, that, some} (tall) boy.
  - *Jan likes (tall) boy.
  - Jan likes (vanilla) ice cream.

- I.e., NPs derived from
  - proper nouns, plurals, and mass nouns don’t need determiners
  - those derived from singular common count nouns (generally) do.
    - There are, of course, lots of oddities: “part”, unique appositionals, prototype activity nouns….

- But our rules for NPs lose this distinction.
We can differentiate our grammar rules further.  

E.g., instead of

\[ \begin{align*}
\text{NP}_{\text{min}} & \rightarrow N \mid \text{NP}_{\text{int}} \text{NP}_{\text{min}} \mid \text{PP} \text{NP}_{\text{min}} \\
\text{NP}_{\text{int}} & \rightarrow (Q) \text{AP}^* \text{NP}_{\text{min}} \\
\text{NP}_{\text{max}} & \rightarrow ((\text{PDT}) \text{DP}) \text{NP}_{\text{int}}
\end{align*} \]

we could have

\[ \begin{align*}
\text{NP}_{\text{min/scc}} & \rightarrow N_{\text{scc}} \mid \text{NP}_{\text{int}} \text{NP}_{\text{min/scc}} \mid \text{PP} \text{NP}_{\text{min/scc}} \\
\text{NP}_{\text{int/scc}} & \rightarrow (Q) \text{AP}^* \text{NP}_{\text{int/scc}} \\
\text{NP}_{\text{max}} & \rightarrow ((\text{PDT}) \text{DP}) \text{NP}_{\text{int/scc}} \\
\text{NP}_{\text{min/ppm}} & \rightarrow N_{\text{ppm}} \mid \text{NP}_{\text{int}} \text{NP}_{\text{min}} \mid \text{PP} \text{NP}_{\text{min/ppm}} \\
\text{NP}_{\text{int/ppm}} & \rightarrow (Q) \text{AP}^* \text{NP}_{\text{min/ppm}} \\
\text{NP}_{\text{max}} & \rightarrow ((\text{PDT}) \text{DP}) \text{NP}_{\text{int}}
\end{align*} \]
But There’s More Like This

- Other grammatical categories of the lexical items need to “shine through” to the NPs.
- E.g.:
  - “Most little girls like ice cream.”
  - “*That little boy like ice cream.”
  - “*Most little girls likes ice cream.”
  - “*Those little boy likes ice cream.”

So, would we would have to differentiate our NPs for “number” as well.

- And, similarly, for “person”:
  - “I like ice cream.”
  - “He likes ice cream.”

although this isn’t as bad, as everything is 3\textsuperscript{rd} person except a few pronouns.
In duplicating the rules, we lose important generalizations.
- E.g., one can make an NP by adding an adjective, but this fact is now replicated several times in the grammar.

However, there is no other solution if we stick to CFGs.
- Indeed, it is exactly the context-free-ness of the rules that causes the problem!

Note that this is a “strong adequacy” objection.
- It’s not that we can’t write down the grammar; it’s that we can’t write down a satisfying one.
The Verb Phrase

- **Main clauses, e.g.,**
  "Pat baked Jan cookies"

  are typically analyzed as

  \[
  [S [NP Pat] [VP [V baked] [NP Jan] [NP cookies]]]
  \]

  as opposed to

  \[
  [S [NP Pat] [V baked] [NP Jan] [NP cookies]]
  \]

- **I.e., the basic general structure is**
  - "NP VP",
  - with the VP having the further structure of "V NP NP"

  rather than the flatter

  - "NP VP NP NP"

- **But why?**
In general, we have to look for evidence that that structure can appear in different contexts.

Some useful sorts of tests involve
- Substitution
- Question and fragment response
- Coordination
- “Movement”
- Ellipsis
- Asymmetric c-command

Note: These are generally revealing, but don’t always agree with each other, leaving lots to debate about the particulars.
Constituent Structure Analysis Examples

- **Substitution**
  - Pat [baked Jan cookies] → Pat [did so], Pat [ran]
  - Pat baked [Jan cookies] → Pat baked [???].

- **Question and fragment response**
  - What did Pat do? → Bake Jan cookies

- **Coordination**
  - Pat [baked Jan cookies] and [put them on the stove to cool].

- **“Movement”**
  - What Pat did was [bake Jan cookies].

- **Ellipsis**
  - Pat [baked Jan cookies] and so did Lynn/Lynn did too.

- **Asymmetric c-command**
  - Pat and Jan [baked each other cookies].
  - *Each other baked Pat and Jan cookies.
As we said, these are sometimes conflicting. E.g., note that coordination allows the following:

Pat baked and Jan iced a chocolate layer cake, which suggests that [Pat baked] and [Jan iced] are constituents.

But the other tests don’t bear this out:

*What was done to the cake was Pat baked.
*Pat baked a cake and so did frost.
The Verb Phrase

Here are some common structures, and phrases that conform to them:

- \( VP \rightarrow V \)
  - walked
- \( VP \rightarrow V \ NP \)
  - shot the gun
- \( VP \rightarrow V \ NP \ PP \)
  - put the book on the shelf
- \( VP \rightarrow V \ NP \ NP \)
  - baked Jan a cake
- \( VP \rightarrow V \ PP \)
  - leave for New York
- \( VP \rightarrow V \ S \)
  - think I would like to leave now
As we saw, we should have a VP coordination rule as well:

\[ \text{VP} \rightarrow \text{VP Conj VP} \]

And we need to allow for
- adverbials
- auxiliaries

which we will skip for now.
Note, however, that within the basic VP, which structure you use depends heavily on the verb.

- Traditionally, we have the transitive/intransitive distinction.
- But here we see that particular verbs subcategorize for a variety of different structures.
- This is the principle area in which syntax has to come to grips with the properties of individual words.
Solutions?

- We really only have one trick. 😊
- Let’s introduce syntactic categories $V_i$, $V_t$, $V_{do}$, $V_{o[to]}$, $V_{to-inf}$, etc., and then write special rules for each one:
  
  $\begin{align*}
  VP & \rightarrow V_i \\
  VP & \rightarrow V_t \text{ NP} \\
  VP & \rightarrow V_{nppp} \text{ NP PP} \\
  VP & \rightarrow V_{do} \text{ NP NP} \\
  VP & \rightarrow V_{pp} \text{ PP} \\
  VP & \rightarrow V_{to-inf} \text{ S}
  \end{align*}$

  which is in fact what some approaches do.

- Again, it has been argued that one can’t capture certain regularities this way.
  - E.g., “Jan verbed Pat a book.” $\leftrightarrow$ “Jan verbed a book to Pat.” (sometimes)
Sentence Level Constructions

- Sentences are generally regarded as a bigger form of VP, just as we had different forms of NP.
- But, traditionally, we use the separate symbol “S” anyway.
- Here are some common sentence types:
  
  \[ S \rightarrow NP \text{ VP} \]
  
  Jan put the book on the shelf.

  \[ S \rightarrow Aux \text{ NP} \text{ VP} \]
  
  Did Jan put the book on the shelf?

  \[ S \rightarrow Wh-\text{NP} \text{ VP} \]
  
  Which suspects may have put the book on the shelf?

  \[ S \rightarrow Wh-\text{NP} \text{ Aux} \text{ NP} \text{ VP} \]
  
  Which book did Jan put on the shelf?

- And we can conjoin sentences as well:

  \[ S \rightarrow S \text{ Conj } S \]
Complications

- This analysis is incomplete in lots of ways.
- Consider, for example, the last sentence type, a so-called “non-subject wh-question”:
  Which book did Jan put on the shelf?
- Note that its VP is
  put on the shelf
  which is not a valid according our analysis so far.
  - I.e., it is “missing” the NP, which is now part of the S.
- There are other constructions that similarly leave “gaps”:
  Whichever toy you pick Eli will want to play with.
- Dealing with gaps is a major cottage industry.
And We Have the Second Half of Our NP Problem

- We noted that NPs had to export the “number” (and “person”) properties of their lexical start.
  - In particular, subject NPs have to agree with Vs along these dimensions.
  - However, the V has long since been abstracted away by the time we get to a VP.

- So, once again, we have no choice but to “version” all of our VP rules, to show all possible combinations of number and person.
Comment

- An ugly solution just got uglier.
Heads, Complements and Adjuncts

- For most constituents, there is a syntactically central part, and some less central parts.
- For example, consider:
  - “the conservative senator”
    - This is a noun phrase whose head is the noun phrase “conservative senator”.
    - This noun phrase in turn has the head “senator”.
    - We further say that “senator” is the lexical head of both NPs.
- In almost all theories of grammars today, almost all constituents are regarded as projections of lexical heads.
- I.e., we start with a noun, and build up noun phrases, start with verbs, build up verb phrases, etc.
Terminology

- The other items in the constituent besides the head are either complements or adjuncts.
  - A complement is something that the head subcategories for;
  - An adjunct is anything else.
- E.g., in
  "Jan put the can on the shelf yesterday in her apartment in New York."
  - the NP “the can” and the PP “on the shelf” are complements of the verb “put”;
  - “yesterday” and “in her apartment…” are adjuncts.
- Note that the subjects are always required, but are not part of the same constituent as the verb.
  - Sometimes these are called “distant complements” (but this usage doesn’t seem widespread).
Projections and Syntactic Categories

- Above, we stipulated quite a few NP syntactic categories.
- However, it might be that we can get away with fewer if we understood the relation of each of these to the lexical head.
- Indeed, there are theories that postulate that there are only fixed number of projection types for all syntactic categories. These are usually:
  - the lexical item itself (e.g., an N)
  - a “maximal projection” (e.g. an NP that can be a complement elsewhere)
  - an intermediate projection
- These were written, for a given lexical category $X, X, X',$ and $X''$ (but pronounced “$x$ bar” and “$x$ double bar”).
X-bar Theory

In such theories:
- Complement is daughter of X', sister of X.
- Adjunct is daughter of X', sister of X'.
- Specifier is daughter of X'', sister of X'.
Comments

- S is usually regarded as a V”.
- Lots of versions, controversy on the details.
- However, most theories today incorporate some notion of head + projections.
- Note that syntactic categories are no longer atomic.
  - What we have been called “NP” is now “N with bar feature = 2” or some such.
- BTW, our analysis of NP doesn’t quite fit into this model.
  - But it’s close, and can probably be made to fit.
Confusion About Heads

- There are some cases where what the head is may not be entirely clear.
  - Expressions like “hunter gatherer” has been analyzed as dual-headed.
  - Some analyses consider coordinate structures as having as many heads as elements they coordinate.

- There is some disagreement as to what is the head of a given constituent type.
  - E.g., some linguists have argued that phrases like “the little girl” are really determiner phrases, rather than noun phrases.
Note

- We posited (deep) cases only for (possibly distant) complements.
- Semantically, adjuncts describe more general aspects of a situation, and syntactically, are probably “further away” a lexical item.
Adding Clausal Modifiers

- Prepositional and adverbial adjuncts are okay before an S:
  - In the morning, Jan left.
  - Oddly, Jan sang folks songs.

So we might add

\[
S \rightarrow AA^* S
\]

\[
AA \rightarrow PP \mid \text{AdvP}
\]

- You can also get these at the end, but then they are best analyzed as part of the VP:
  - Jan left in the morning/quickly.
  - Jan sang folks songs oddly.
  - Jan quickly left the meeting

So one might add

\[
\text{VP} \rightarrow AA^* \text{VP AA}^*
\]
An Approximate Grammar, Redux

- However, most analyses have more embedded constituent structure.
- So, a somewhat better set of rules might be the following:

  \[
  \begin{align*}
  \text{NP} \_{\text{bare}} & \rightarrow \text{N} \\
  \text{NP} \_{\text{bare}} & \rightarrow \text{NP} \_{\text{small}} \text{ NP} \_{\text{bare}} \\
  \text{NP} \_{\text{adj}} & \rightarrow \text{NP} \_{\text{bare}} \\
  \text{NP} \_{\text{adj}} & \rightarrow \text{AP} \text{ NP} \_{\text{adj}} \\
  \text{NP} \_{\text{small}} & \rightarrow \text{Num} \text{ NP} \_{\text{adj}} \mid \text{PP} \text{ NP} \_{\text{adj}} \\
  \text{NP} \_{\text{small}} & \rightarrow \text{NP} \_{\text{adj}} \\
  \text{NP} \_{\text{q}} & \rightarrow \text{Q} \text{ NP} \_{\text{small}} \\
  \text{NP} \_{\text{q}} & \rightarrow \text{NP} \_{\text{small}} \\
  \text{NP} \_{\text{d}} & \rightarrow \text{D} \text{ NP} \_{\text{q}} \\
  \text{NP} \_{\text{d}} & \rightarrow \text{NP} \_{\text{q}} \\
  \text{NP} & \rightarrow \text{PDT} \text{ NP} \_{\text{d}} \\
  \text{NP} & \rightarrow \text{NP} \_{\text{d}}
  \end{align*}
  \]