Introductory Linguistics

A draft textbook by

Bruce P. Hayes

Department of Linguistics
University of California, Los Angeles

Contents

Chapter 1: What is Linguistics?

1. What this book will be like

Linguistics is the science of language; it studies the structure of human languages and aims to develop a general theory of how languages work. The field is surprisingly technical; to describe languages in detail requires a fair amount of formal notation. A good parallel would be the field of symbolic logic, which uses a formal notation to understand the processes of reasoning and argumentation.

There are basically three things I hope you will get out of this book.

First, there is the subject matter itself, which is useful to know for people many different fields, such as education, psychology, and computation. The course is also an introduction to linguistics for those who are going to major in it.

Second, the course involves some mental exercise, involving analysis of data from English and other languages. I doubt that anyone who doesn't go on in linguistics will remember much of the course material five years after they have graduated, but the analytical skills you will get practice in will be (I hope) both more permanent and more useful.

Third, the course is intended to give a more realistic view of science and how it proceeds. The reason we can do this in linguistics is that it is a fairly primitive science, without an enormous body of well-established results. Because of this, we are less interested in teaching you a body of established knowledge; rather, our focus is on teaching you to decide what is right on your own, by looking at the data. All sciences are in this state of uncertainty at their frontiers; linguistics can give you a more authentically scientific experience in a beginning course.

2. Implicit and explicit knowledge of language; working with consultants

Linguists are constantly asked the question "How many languages do you speak?" This question is a little irritating, because it is largely irrelevant to what linguists are trying to do. The goals of linguistics are to describe and understand the structure of human languages; to discover the ways in which all languages are alike and the ways in which they may differ. The point is that even if one could speak all 8000 or so of the world's languages, one would not have solved all the problems of linguistics.

The reason is this: speaking a language and knowing its structure are two very different things. In speaking a language, one uses thousands of grammatical rules without being aware of them; they are "unconscious knowledge." Linguists attempt to make explicit this unconscious knowledge by looking closely at the data of language. That is, they attempt to make the "implicit knowledge" of native speakers into explicit knowledge.

This goal implies one of the central methods of doing linguistic research, the consultant session. Quite often, a linguist will study the structure of language she does not speak; this is done by finding a native-speaker consultant to provide the data. The linguist normally asks the

consultant a great number of questions. Some of them are simple and establish basic knowledge: "What is the word for "duck" in your language?". Others look for the various different grammatical forms of the same word: "How would you say "two ducks"? (looking, perhaps, for a plural). Others involve whole sentences and often their meanings as well.

The crucial idea in a consultant session is that the linguist is *thinking about structure*—is making and checking hypotheses. The native speaker is most often trying simply to provide an honest and accurate report of how she speaks the language, and of her intuitions about meaning and other matters.

Obviously, the lines can be blurred a bit: sometimes the consultant (especially if she knows some linguistics), may want to suggest some hypotheses herself. And linguists sometimes "work on themselves", so that the dialogue across the consultant table becomes an internal dialogue in the mind of the linguist.[1]

The following example illustrates the method: for one particular area of English grammar, we get some native speaker intuitions, and work out a series of hypotheses for what the rules of English are. I hope the analysis illustrates the following crucial point: *the native speaker doesn't know the answer*. But the native speaker has the tacit, intuitive knowledge that makes it possible to find the answer, or at least get closer to it.

3. The reference of *each other*

Supposing, then, that we have a native speaker of English present. Such a speaker is likely to tell you that in (1), *each other* refers to *we*, and that it means something like, "I like you and you like me."

(1) We like each other.
(2) John and Bill like each other.

In linguistics this is often called the **reciprocal reading**; i.e. it says we are in a state of reciprocal liking. Sentence (2) has a similar reciprocal reading.

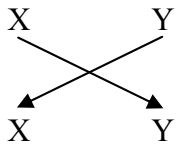
Sentence (3) is a bizarre sentence, in that *each other* cannot logically refer to *I*.

(3) *I like each other.

The native speakers responds to it by saying, "That's weird/that's bizarre/you can't say that in English." We will say for present purposes that (3) is **ungrammatical**; that is, ill-formed. Following standard practice, I will place an asterisk before sentences that are ungrammatical

In (3), the ungrammaticality can be traced to the absence of any plausible interpretation for the sentence; since *each other* describes reciprocal actions, like this:

[1] In practice I and probably other linguists find this hard to do; it's just too much going on in your head at once. More important, it poses methodological problems; the data are likely to be contaminated by wishful thinking.



Each other cannot be used unless the agent of the action is plural. But not all cases can be explained in this way. In (4), you can think of a meaning that the sentence could in principle have, but this meaning is not allowed by the rules of English grammar:

(4) *John and Bill think I like each other.

In other words, being grammatical and having a sensible meaning are two different things.

Sentence (5) shows the same thing: you can think up two logically possible meanings, but only one meaning is allowed by the rules of English.

(5) We believe they like each other.

We've now reached our basic point: there must be some rule of English that accounts for what *each other* can refer to, but it is a tacit rule. No one can look inside their mind to find out what the rule is; one can only look at the data and try to figure the rule out. Linguists have worked on this particular rule for some time, and have gradually made progress in stating the rule accurately. But we cannot claim to have a final answer.

I will present a partial answer here. We will need two preliminary definitions, both of which will come up later on in the course. Here is the first one:

A **clause** is either a whole sentence or a sentence within a sentence

You can identify clauses because they generally have a subject and a verb, and they express some sort of proposition. We depict clauses by drawing brackets around them; labeled "S" for "little sentence".

(6) [We like each other.]$_S$
(7) [John and Bill like each other.]$_S$
(8) *[I like each other.]$_S$
(9) *[John and Bill think [I like each other.]$_S$]$_S$
(10) [We believe [they like each other.]$_S$]$_S$

Note that clauses can have clauses inside them. In (11), there is a clause that expresses the content of John and Bill's thoughts (*I like each other*), and the whole thing is an (ungrammatical) clause that describes a state (John and Bill are having a particular thought.)

We also need to define **noun phrases**.

A **noun phrase** is a complete syntactic unit that refers to a thing or a set of things.

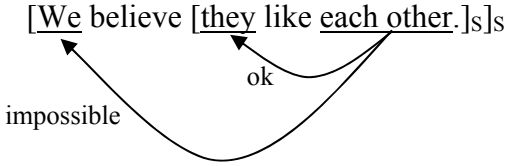
For example, in (6) and (7), the noun phrases are *we*, *each other*, *John and Bill*, and again *each other*.

With these definitions, we can write a tentative rule for what *each other* refers to:

(12) ***Each other* reference rule**

Each other can refer only to noun phrases that are inside the smallest clause containing it.

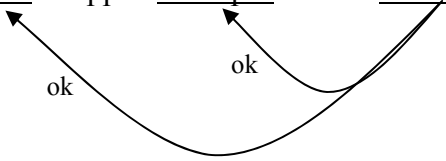
In the tricky cases, (4-5), the rule works fine: it requires that *each other* refer to *I* and *they* respectively. This can be seen graphically for (5) if we make a diagram, using arrows to show legal and illegal reference.



Cases (1)-(3) are easy: there is only one noun phrase for *each other* to refer to, and the rule permits this.

Notice that in a sentence with just one clause, but two noun phrases in addition to *each other*, there will be two possibilities for what *each other* might refer to:

(13) [We wrapped the ropes around each other.]S



This is just what our rule predicts. Because of this, the sentence has two possible meanings.

There are some further relevant data, which are perhaps syntactically the most interesting:

(14) [We consulted two detectives in order [to find out about each other]S]S

(15) [They seem to us [to like each other]S]S

These sentences are mysterious: it looks like there is no noun phrase at all that occurs inside the smallest clause containing *each other* (other than *each other* itself). But consider the meaning of the sentences: *someone* is doing the finding out in (14), namely, "we", and *someone* is doing the liking in (15), namely "they". Thus, the peculiar clauses *to find out about each other* and *to*

like each other appear to have **implicit noun phrases**. They have a meaning, but they're not pronounced.

For purposes of analyzing explicitly, let us fill them in:

(16) [We consulted two detectives in order [(*we*) to find out about each other]$_S$]$_S$

(17) [They seem to us [(*they*) to like each other]$_S$]$_S$

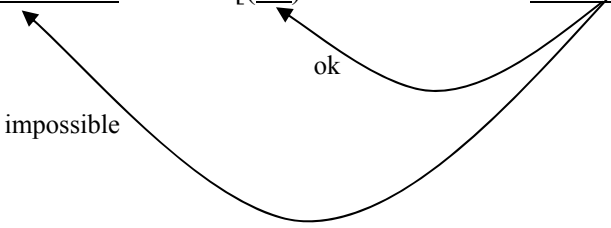
With the implicit subjects filled in, we can explain what is going on. The *Each Other* Reference Rule needs slight revising:

(18) ***Each Other* Reference Rule** (revised)

Each other can refer only to noun phrases (including implicit noun phrases) that are inside the smallest clause containing it.

Here is an analytic diagram for sentence (15):

[We consulted two detectives in order [(*we*) to find out about each other]$_S$]$_S$



We will do more on this kind of rule later. The major gap in the analysis as given so far is that we haven't said anything about what causes the implicit noun phrase to take on a particular meaning—for instance, why does the implicit noun phrase in (15) have to mean *we*, and not *two detectives*? This would lead us into a different area of English grammar, covering the implicit noun phrase behavior of *in order to* and many other grammatical constructions.

For now, the point is this: knowing English means that you "know" the *Each Other* Reference Rule, in an intuitive sense. But it does not mean that you know it explicitly. Much of the work of linguistics consists of trying to make implicit knowledge explicit. The method is much the same as in other sciences: we gather data (from the native speaker), formulate hypotheses, test the hypotheses against the data, revise the hypotheses, gather more data, and so on.

Study Exercises for "each other"

The first exercise is a very basic one: finding clauses and implicit subjects. This will matter a lot later on. The others are more directly related to the *each other* phenomenon.

Study Exercise #1

This exercise simply asks you to find the clauses. Put […]$_S$ brackets around them. Be sure to get all of the clauses, sometimes there are more than one. If the subject is implicit, put "( )" where the subject would be and say what it stands for.

Examples:

i. Alice believes that Fred sang. Answer: [Alice believes that [Fred sang]$_S$]$_S$
ii. Alice hopes to climb Everest. Answer [Alice hopes [(Alice) to climb Everest]$_S$]$_S$

Exercises:

a. I believe that turtles can swim.
b. The fact that Fred left bothers Alice.
c. Bill said that Jane sang and Fred danced.
d. I persuaded Fred to buy a telescope.
e. I promised Fred to buy a telescope.
f. To appear on television is her fondest dream.
g. Joe said that he wants to leave.
h. That Jane can sing tenor makes no difference.
i. Bill left because he was tired.
j. the idea that truth is obtainable

Study Exercise #2

This sentence is ambiguous:

My sister and I gave our parents books about each other.

Explain each possible meaning and illustrate it with a diagram (brackets and arrows) like the ones given above.

Study Exercise #3

This sentence is ambiguous:

Bill and Fred persuaded Alice and Sue ( ) to buy telescopes in order ( ) to find out more about each other.

For example, in one reading, you could continue: "In fact, as it turned out, Bill succeeded in finding out more about Fred, but Fred did not succeed in finding out more about Bill." In the other reading, you could continue, "In fact, as it turned out, Alice succeeded in finding out more about Sue, but Sue did not succeed in finding out more about Alice."

For each meaning, fill in the implicit subjects shown with ( ). Then draw diagrams for the reference of *each other*. (So you'll end up with two diagrams.)

Study Exercise #4

For this sentence:

My parents tell my sister and me every day to write books about each other.

there's only one meaning: "My parents tell my sister every day to write a book about me and tell me every day to write a book about my sister." It can't mean "My mother tells my sister and me every day to write a book about my father and my father tells my sister and me every day to write a book about my mother." Explain why, giving diagrams for both the possible and the impossible meaning.

This page is intensionally blank.

This area is intensionally blank.

4. The field of linguistics

With this background, here is a (somewhat narrow) definition of the field of linguistics: it is the study of the (largely implicit) knowledge people have when they speak a language. Some of the subfields of linguistics are the following:

syntax	study of rules for forming sentences
semantics	study of rules for meaning
morphology	study of rules for forming words
phonology	study of rules of pronunciation

In all cases, the "rules" are of the kind known implicitly by native speakers, not the kind learned in school. Linguistics has two other major subfields that also involve rules but are not as directly focused on them: **phonetics**, which studies how sounds are produced and perceived, and **historical linguistics**, which studies how languages change.

Linguists attempt to arrive at explicit knowledge of all the world's languages. I should point out that this task will never be completed. First, there are over 8000 different languages, many of which are spoken in remote areas of the world.[2] More important, the amount of explicit knowledge contained in just a single language would fill a whole library. Linguists find it both frustrating and astonishing that a small child can acquire implicitly in just a few years the same knowledge that takes decades of hard work for linguists to figure out explicitly.

Linguists are also interested in developing a general theory of language; a theory of the properties that all languages share. These are called **linguistic universals**. Finding universals is also challenging; many linguists have the experience of having proposed a linguistic universal, only to find out later on about languages that don't fit in.

[2] The best directory to the world's languages is the *Ethnologue*, on line at http://www.ethnologue.com/.

Chapter 2: Morphology

5. Orientation

In linguistics, "morphology" means "the study of word structure." We're interested in the structure of individual words, as well as the grammatical principles whereby words are formed.

Some of the terminology used here is probably familiar to you. The **stem** of a word is its core, the part that bears its central meaning. Thus in *undeniable*, the stem is *deny*; and in *insincerity* the stem is *sincere*.

Material that is added to the stem, thus modifying its meaning in some way, consists principally of **prefixes** and **suffixes**. The suffix *-able* is prefixed to *deny* to form *deniable*[3]; and the prefix *un-* is added to the result to obtain *undeniable*. Often, multiple prefixes and suffixes can be added to the same stem, producing ever longer and more elaborate words: *undeniability, hyperundeniability*.

Stem, prefixes, and suffixes are the building blocks from which words are assembled. The term used for such building blocks by linguists is **morpheme**, often defined as follows:

- A **morpheme** is the smallest linguistic unit that bears a meaning.

Thus, *un-*, *deny*, and *-able* are morphemes; *deniable* is not a morpheme because it can be split; *de* and *ny* are not morphemes because they are meaningless.

We can start with a bit of notation. Words are shown broken into their morphemes with hyphens: *un-deni-abil-ity*. And prefixes and suffixes are shown with hyphens to identify them as such: prefixes like *un-*, suffixes like *-ity*. You can think of the hyphen as the bit of imaginary "glue" with which a morpheme attaches to the stem.

6. Two kinds of morphology

Most linguists acknowledge at least a rough distinction between two kinds of morphology: **word formation** vs. **inflectional morphology**. We'll start with inflectional morphology.

Inflectional morphology is grammatical morphology. Here are some examples to start, from English:

- **tense** on verbs (present tense *jumps*, past tense *jumped*)
- **number** on nouns (singular *cow*, plural *cows*)
- a small amount of **person and number agreement** in verbs (*She sings*. vs. *They sing.*)

English is actually not a very good language for studying inflectional morphology, because it doesn't have all that much of it (Mandarin is a similar case). But other languages, such as

[3] We'll ignore the change of *y* to *i*, a convention of English spelling.

Swahili, Russian, or Turkish, have a great deal, and students of these languages can spend years getting through it all.

7. Morphological analysis

When they encounter an unfamiliar language, linguists usually begin their work by carrying out a morphological analysis. This involves gathering data, determining what morphemes are present in the data, and writing the rules that form the words from the morphemes.

There are no fancy methods for doing this; basically one must scan a collection of morphologically similar words and determine which phoneme sequences remain the same whenever the meaning remains the same.

We will do this now for a fairly simple case, namely a fragment of the nominal morphology (=morphology for nouns) in Turkish. Here are the data:

1.	el	'hand'	ev	'house'	zil	'bell'
2.	eli	'hand (object)'	evi	'house (object)'	zili	'bell (object)'
3.	ele	'to (a) hand'	eve	'to (a) house'	zile	'to (a) bell'
4.	elde	'in (a) hand'	evde	'in (a) house'	zilde	'in (a) bell'
5.	elim	'my hand'	evim	'my house'	zilim	'my bell'
6.	elimi	'my hand (object)'	evimi	'my house (object)'	zilimi	'my bell (object)'
7.	elime	'to my hand'	evime	'to my house'	zilime	'to my bell'
8.	elimde	'in my hand'	evimde	'in my house'	zilimde	'in my bell'
9.	elin	'your hand'	evin	'your house'	zilin	'your bell'
10.	elini	'your hand (object)'	evini	'your house (object)'	zilini	'your bell (object)'
11.	eline	'to your hand'	evine	'to your house'	ziline	'to your bell'
12.	elinde	'in your hand'	evinde	'in your house'	zilinde	'in your bell'
13.	elimiz	'our hand'	evimiz	'our house'	zilimiz	'our bell'
14.	elimizi	'our hand (object)'	evimizi	'our house (object)'	zilimizi	'our bell (object)'
15.	elimize	'to our hand'	evimize	'to our house'	zilimize	'to our bell'
16.	elimizde	'in our hand'	evimizde	'in our house'	zilimizde	'in our bell'
17.	eliniz	'your (plur.) hand'	eviniz	'your (plur.) house'	ziliniz	'your (plur.) bell'
18.	elinizi	'your (pl.) hand (obj.)'	evinizi	'your (pl.) house (obj.)'	zilinizi	'your (pl.) bell (obj.)'
19.	elinize	'to your (pl.) hand'	evinize	'to your (pl.) house'	zilinize	'to your (pl.) bell'
20.	elinizde	'in your (pl.) hand'	evinizde	'in your (pl.) house'	zilinizde	'in your (pl.) bell'
21.	eller	'hands'	evler	'houses'	ziller	'bells'
22.	elleri	'hands (object)'	evleri	'houses (object)'	zilleri	'bells (object)'
23.	ellere	'to hands'	evlere	'to houses'	zillere	'to bells'
24.	ellerde	'in hands'	evlerde	'in houses'	zillerde	'in bells'
25.	ellerim	'my hands'	evlerim	'my houses'	zillerim	'my bells'
26.	ellerimi	'my hands (obj.)'	evlerimi	'my houses (obj.)'	zillerimi	'my bells (obj.)'
27.	ellerime	'to my hands'	evlerime	'to my houses'	zillerime	'to my bells'
28.	ellerimde	'in my hands'	evlerimde	'in my houses'	zillerimde	'in my bells'
29.	ellerin	'your hands'	evlerin	'your houses'	zillerin	'your bells'
30.	ellerini	'your hands (obj.)'	evlerini	'your houses (obj.)'	zillerini	'your bells (obj.)'
31.	ellerine	'to your hands'	evlerine	'to your houses'	zillerine	'to your bells'
32.	ellerinde	'in your hands'	evlerinde	'in your houses'	zillerinde	'in your bells'
33.	ellerimiz	'our hands'	evlerimiz	'our houses'	zillerimiz	'our bells'

34.	ellerimizi	'our hands (obj.)'	evlerimizi	'our houses (obj.)'	zillerimizi	'our bells (obj.)'
35.	ellerimize	'to our hands'	evlerimize	'to our houses'	zillerimize	'to our bells'
36.	ellerimizde	'in our hands'	evlerimizde	'in our houses'	zillerimizde	'in our bells'
37.	elleriniz	'your (pl.) hands'	evleriniz	'your (pl.) houses'	zilleriniz	'your (pl.) bells'
38.	ellerinizi	'your (pl.) hands (obj.)'	evlerinizi	'your (pl.) houses (obj.)'	zillerinizi	'your (pl.) bells (obj.)'
39.	ellerinize	'to your (pl.) hands'	evlerinize	'to your (pl.) houses'	zillerinize	'to your (pl.) bells'
40.	ellerinizde	'in your (pl.) hands'	evlerinizde	'in your (pl.) houses'	zillerinizde	'in your (pl.) bells'

We have here three columns, indicating inflected forms of the three nouns meaning "hand", "house", and "bell". Abbreviations and grammatical conventions are as follows:

- "plur." or "pl." abbreviate "plural".
- "your (pl.)" is second person plural. Here, as a possessive, it means "belonging to you, there being more than one of you."
- (object) or (obj.) means that that form would be used as the **object** of a verb. Thus, if one were to say in Turkish something like "I saw my hand", one would use #3, *eli*.[4]

8. Breaking up the words into morphemes

The search, as always, is for invariant form paired with invariant meaning. In the first column, every single form begins with the sounds /el/ and has a meaning involving "hands". It seems inconceivable that "hand" could be anything other than /el/, or that /el/ could be anything other than "hand"—note in particular the first line, where /el/ by itself means "hand" by itself.

The columns for "house" and "bell" are completely identical to the column for "hand", except that where column has /el/, columns 2 and 3 have /ev/ and /zil/ as stems. It is plain that /ev/ means "house" and /zil/ means "bell".

Moving on, we can compare:

1. el 'hand'
2. eli 'hand (object)'
3. ele 'to (a) hand'

Subtracting out /el/ from the second and third forms, it appears that /-i/ and /-e/ must be suffixes. We can confirm this by casting an eye over the remainder of the data: /-e/ "goes together" with the English word "to" given in the translations; and likewise /-i/ with "(object)".

The /-e/ and /-i/ suffixes apparently denote the grammatical role that the noun plays in a Turkish sentence, a phenomenon called **case**. Let's briefly digress with the basics of case.

---

[4] Thus the reference source on Turkish I'm using gives the sentence

Beş	adam	heykel-**i**	kırdi
five	man	statue-**accusative**	broke

'Five men broke the statue'

- Case is fundamentally an inflectional category of nouns (though often adjectives and articles agree with their noun in case).
- Case tells us, intuitively, *who is doing what to whom* — it identifies the basic semantic roles of the participants in a clause. In many languages, "Man bites dog" is Man-nominative bites dog-accusative", and "Dog bites man" is "Dog-nominative bites man-accusative." Nominative and accusative are probably the two most common cases.
- Case is not the only way to show who is doing what to whom. In languages with no case, or ill-developed case systems (English), the work done by case is taken over by strict word order and by prepositions.[5]
- Some typical cases in languages (each language is different in its cases and their usage):

  - **Nominative**, usually for subjects of sentences or the citation form of a word
  - **Accusative**, usually for objects of verbs
  - **Dative**, conveying the notion of "to" in English: *I gave the book to the student*.
  - **Locative**, conveying the notion of "at", "in", "on", etc.

There are many other cases; Finnish is analyzed as having fifteen. This isn't really that remarkable, since many of these are simply that way of expressing notions that are expressed by prepositions.[6]

In Turkish:

- /-e/ is the suffix for the dative case
- /-i/ is the suffix for the accusative case
- /-de/ is the suffix for the locative case.

Inspecting the data in rows 21-40, it is plain that every **plural** noun has the suffix /-ler/.

Lastly, there is a set of **possessive suffixes**, which express essentially the same information as what in English is expressed by possessive pronouns like *my* and *your*.[7] There are four possessive suffixes present in the data(Turkish has more, but these are not included here.)

- -im 'my'
- -in 'your'
- -imiz 'our'
- -iniz 'your-plural'

[5] Still other ways exist—in Tagalog, much of this information is given using prefixes or suffixes on the verb.

[6] Or their counterpart, postpositions, which follow their object noun phrase.

[7] Their usage is not quite the same, because if there is a noun possessor, you use the suffix *as well*. Thus, in English, we say (for example) *Ayşe's bell*; but in Turkish *Ayşe-nin zil-si*, which is literally *Ayşe's bell-her*; similarly *biz-im zil-imiz*, literally "us's bell-our".

We can classify the possessive suffixes on the dimensions of **person** and **number**. Number is simply the distinction between singular vs. plural. Person takes (as a first approximation) three values:

- "**First person**" refers to pronouns and grammatical endings that involve the speaker, either alone or with others. Thus in English *I* is a first-person singular pronoun, *we* is first person plural.

- "**Second person**" refers to pronouns and grammatical endings that involve the hearer, either alone or with others. In Spanish *tú* is a second-person singular pronoun, used to address one person, and *vosotros* is a second-person plural pronoun, used to address more than one person.[8]

- "**Third person**" refers to pronouns and grammatical endings that involve neither the speaker nor the hearer. Thus *he/she/it* are third-person singular pronouns, *they* third person plural.

Once we've found all the parts, we can restate the original data, putting in **hyphens** to separate out the morphemes. I'll do this just for the "hand" forms. I've also add a morpheme-by-morpheme translation, also separated out by hyphens; this is called a **gloss**. Glosses are meant to clarify structure, rather than give an idiomatic reading.

	Word	*Gloss*	*Idiomatic translation*
1.	el	'hand	'hand'
2.	el-i	'hand-acc.	'hand (object)'
3.	el-e	'hand-dative	'to (a) hand'
4.	el-de	'hand-locative'	'in (a) hand'
5.	el-im	'hand-1 sg.	'my hand'
6.	el-im-i	'hand-1 sg.-acc.'	'my hand (object)'
7.	el-im-e	'hand-1 sg.-dat.'	'to my hand'
8.	el-im-de	'hand-1 sg.-loc.'	'in my hand'
9.	el-in	'hand-2 sg.'	'your hand'
10.	el-in-i	'hand-2 sg.-acc.'	'your hand (object)'
11.	el-in-e	'hand-2 sg.-dat.'	'to your hand'
12.	el-in-de	'hand-2 sg.-loc.'	'in your hand'
13.	el-imiz	'hand-1 plur.'	'our hand'
14.	el-imiz-i	'hand-1 plur.-acc.'	'our hand (object)'
15.	el-imiz-e	'hand-1 plur.-dat.'	'to our hand
16.	el-imiz-de	'hand-1 plur.-loc.'	'in our hand'
17.	el-iniz	'hand-2 plur.'	'your (plur.) hand'
18.	el-iniz-i	'hand-2 plur.-acc.'	'your (pl.) hand (obj.)'
19.	el-iniz-e	'hand-2 plur.-dat.'	'to your (pl.) hand'
20.	el-iniz-de	'hand-2 plur. -loc.'	'in your (pl.) hand'

[8] Standard English doesn't make the distinction between singular and plural in the second person; though many regional dialects have a special plural pronoun, "yall", used whenever the addressee is plural.

21. el-ler	'hand-plural'	'hands'
22. el-ler-i	'hand-plural-acc.'	'hands (object)'
23. el-ler-e	'hand-plural-dat.'	'to hands'
24. el-ler-de	'hand-plural-loc.'	'in hands'
25. el-ler-im-i	'hand-plural-1 sg.-acc.'	'my hands (obj.)'
26. el-ler-im-e	'hand-plural-1 sg.-dat.'	'to my hands'
27. el-ler-im	'hand-plural-1 sg.'	'my hands'
28. el-ler-im-de	'hand-plural-1 sg.-loc.'	'in my hands'
29. el-ler-in	'hand-plural-2 sg.'	'your hands'
30. el-ler-in-i	'hand-plural-2 sg.-acc.'	'your hands (obj.)'
31. el-ler-in-e	'hand-plural-2 sg.-dat.'	'to your hands'
32. el-ler-in-de	'hand-plural-2 sg.-loc.'	'in your hands'
33. el-ler-imiz	'hand-plural-1 plur.'	'our hands'
34. el-ler-imiz-i	'hand-plural-1 plur.-acc.'	'our hands (obj.)'
35. el-ler-imiz-e	'hand-plural-1 plur.-dat.'	'to our hands'
36. el-ler-imiz-de	'hand-plural-1 plur.-loc.'	'in our hands'
37. el-ler-iniz	'hand-plural-2 plur.'	'your (pl.) hands'
38. el-ler-iniz-i	'hand-plural-2 plur.-acc.'	'your (pl.) hands (obj.)'
39. el-ler-iniz-e	'hand-plural-2 plur.-dat.'	'to your (pl.) hands'
40. el-ler-iniz-de	'hand-plural-2 plur.-loc.'	'in your (pl.) hands'

9. Classifying the suffixes and discovering order

It is useful at this point to sort all the suffixes discovered according to their function:

Case endings

/-i/	accusative
/-e/	dative
/-de/	locative

Possessive suffixes

/-im/	'my'
/-in/	'your'
/-imiz/	'our'
/-iniz/	'your-plural'

Plural

/-ler/

In particular, if you scan the data (now greatly clarified with hyphens and glosses), you can find two important generalizations:

- No word contains more than one possessive suffix, or more than one case.
- Suffix order is invariant, and goes like this:

Plural precedes **Possessive Suffix** precedes **Case**.

With a word processor, it's not hard to prove these relationships by lining up the relevant morphemes into columns with tabs. Here the data once more, displayed in this way.

	Stem	**Plural**	**Poss.**	**Case**		
1.	el				'hand	'hand'
2.	el			i	'hand-acc.	'hand (object)'
3.	el			e	'hand-dative	'to (a) hand'
4.	el			de	'hand-locative'	'in (a) hand'
5.	el		im		'hand-1 sg.	'my hand'
6.	el		im	i	'hand-1 sg.-acc.'	'my hand (object)'
7.	el		im	e	'hand-1 sg.-dat.'	'to my hand'
8.	el		im	de	'hand-1 sg.-loc.'	'in my hand'
9.	el		in		'hand-2 sg.'	'your hand'
10.	el		in	i	'hand-2 sg.-acc.'	'your hand (object)'
11.	el		in	e	'hand-2 sg.-dat.'	'to your hand'
12.	el		in	de	'hand-2 sg.-loc.'	'in your hand'
13.	el		imiz		'hand-1 plur.'	'our hand'
14.	el		imiz	i	'hand-1 plur.-acc.'	'our hand (object)'
15.	el		imiz	e	'hand-1 plur.-dat.'	'to our hand
16.	el		imiz	de	'hand-1 plur.-loc.'	'in our hand'
17.	el		iniz		'hand-2 plur.'	'your (plur.) hand'
18.	el		iniz	i	'hand-2 plur.-acc.'	'your (pl.) hand (obj.)' evinizi
19.	el		iniz	e	'hand-2 plur.-dat.'	'to your (pl.) hand'
20.	el		iniz	de	'hand-2 plur. -loc.'	'in your (pl.) hand'
21.	el	ler			'hand-plural'	'hands'
22.	el	ler		i	'hand-plural-acc.'	'hands (object)'
23.	el	ler		e	'hand-plural-dat.'	'to hands'
24.	el	ler		de	'hand-plural-loc.'	'in hands'
25.	el	ler	im		'hand-plural-1 sg.'	'my hands'
26.	el	ler	im	i	'hand-plural-1 sg.-acc.'	'my hands (obj.)'
27.	el	ler	im	e	'hand-plural-1 sg.-dat.'	'to my hands'
28.	el	ler	im	de	'hand-plural-1 sg.-loc.'	'in my hands'
29.	el	ler	in		'hand-plural-2 sg.'	'your hands'
30.	el	ler	in	i	'hand-plural-2 sg.-acc.'	'your hands (obj.)'
31.	el	ler	in	e	'hand-plural-2 sg.-dat.'	'to your hands'
32.	el	ler	in	de	'hand-plural-2 sg.-loc.'	'in your hands'
33.	el	ler	imiz		'hand-plural-1 plur.'	'our hands'
34.	el	ler	imiz	i	'hand-plural-1 plur.-acc.'	'our hands (obj.)'
35.	el	ler	imiz	e	'hand-plural-1 plur.-dat.'	'to our hands'
36.	el	ler	imiz	de	'hand-plural-1 plur.-loc.'	'in our hands'
37.	el	ler	iniz		'hand-plural-2 plur.'	'your (pl.) hands'
38.	el	ler	iniz	i	'hand-plural-2 plur.-acc.'	'your (pl.) hands (obj.)'
39.	el	ler	iniz	e	'hand-plural-2 plur.-dat.'	'to your (pl.) hands'

40. el ler iniz de ‘hand-plural-2 plur.-loc.’ ‘in your (pl.) hands’

Study exercise #5

Reexamine these suffixes and propose a different—finer-grained—analysis.

/-im/	‘my’
/-in/	‘your’
/-imiz/	‘our’
/-iniz/	‘your-plural’

This area is intensionally blank.

---

10. Position classes in inflectional morphology

When we looked at the Turkish data, the primary finding was that the morphemes could be arranged in a linear order, which could be expressed as five slots.

Stem	*Plural*	*Possessor Person*	*Possessor Number*	*Case*
el ‘hand’	-ler	-im 1st	-iz plural	∅ nominative
ev ‘house’		-in 2nd		-i accusative
zil ‘bell’				-e dative
				-de locative

In a long word like *ellerimizde* ‘in our hands’, all five slots get filled:

Stem	*Plural*	*Possessor Person*	*Possessor Number*	*Case*
el	-ler	-in	-iz	de
hand	plural	1st	plur. poss.	locative

In analysis, words like *ellerimizde* are very useful, since they demonstrate the need for five slots. The slots in a system like this are often called **position classes**. Each position is an abstract location in the word, which can be filled by a particular morpheme or set of morphemes. In the analysis given earlier, we derived position classes using blocks of rules, one block per class.

An important check on a position class analysis is that there should be no contradictions of ordering in the data, if the analysis is correct. We can look through the data and see that (for example) *-in*, *-iz*, and *-de* never precede *-ler*; that *-iz* and *-de* never precede in; that *-de* never precedes *-iz*; and similarly with the other morphemes.

Position classes can be defined simply by looking at the morphemes and checking their ordering. But in fact, the usual picture is that the classes are related to **morphological function**.

For example, it's hardly an accident that the two suffixes in the third Turkish slot are both possessor person suffixes. The general principle is: position reflects function.

This said, it should be noted that there are exceptions; the occasional language will take the same function and put some of the morphemes into different positions; or fill a position with morphemes of variegated function. For instance, the Swahili morpheme *cho*, which means roughly "which", gets put in a different position for positive and negative verbs:

ki-tabu	a-ki-taka-**cho**	Hamisi	
book	SUBJ-OBJ-want-**which**	Hamisi	'the book which Hamisi wants'
ki-tabu	a-si-**cho**-ki-taka	Hamisi	
book	SUBJ-NEG-**which**-OBJ-want	Hamisi	'the book which Hamisi doesn't want'

11. Formalizing with a grammar

Linguists seek to make their analyses as explicit as possible, by expressing the pattern of the language with **rules**. The rules taken together form a **grammar**. We'll start with a very simple grammar for Turkish nominal inflection.

We'll assume that the stem (/el/, /ev/, /zil/, or whatever) comes with **morphological features** specifying its grammatical content. The bundle of features is called the **morphosyntactic representation**.[9] The job of our grammar will be to manifest this content with actual material. For example, we can start out with something like this for #40:

/el/:[Number:plural, PossessorNumber:plural, PossessorPerson:2, Case:Locative]

The /el/ part is the stem meaning 'hand'; it is enclosed in / / because this is the way you indicate the speech sounds (phonemes) of a word. The part in [ ] is the morphosyntactic representation. It contains four morphological features:

Number
PossessorNumber
PossessorPerson
Case

Each feature has a **value**, which is shown by placing it after a colon. So you can read the formula /el/:[Number:plural, PossessorNumber:plural, PossessorPerson:2, Case:Locative] as: "the stem /el/, with a morphosyntactic representation indicating plural Number, plural PossessorNumber, second PossessorPerson, and Locative Case". We'll return later on to the question of where these features come from.

The grammar itself consists of four rules. The order in which the rules are stated is significant and is part of the grammar. Only the first rule is stated in full.

[9] Why? We'll see later on: the morphosyntactic representation transfers information over from the syntax to the morphology.

Number Rule

Suffix /-ler/ if the morphosyntactic representation bears the feature [Number:plural]

Possessor Person Rule

Add a possessor suffix, as follows:

/-im/ if [PossessorPerson:1person]
/-in/ if [PossessorPerson:2person]

Possessor Number Rule

Add a possessor suffix, as follows:

/-iz/ if [PossessorNumber:plural]

Case Rule

Add a case suffix, as follows:

/-i/ if [Case:Accusative]
/-e/ if [Case:Dative]
/-de/ if [Case:Locative]

The reason that the rules must apply in the order given is that by doing this, we construct the word from "inside out", adding a bit more to the material we've already accumulated. This "inside out" character will be shown immediately below.

You can show how the rules apply to a particular form by giving a **derivation**. In linguistics, a derivation shows each rule applying in succession, and justifies the rules by showing that they correctly derive the observed forms.

For the Turkish form ellerinizde 'in your (plur.) hands' (#40 in the data from last time), the derivation would look like this:

/el/:[Number:plural, PossPers:2, PossNum:plural, Case:Locative]	'hand' with its morphosyntactic representation
/el**ler**/:[**Number:plural**, PossPers:2, PossNum:plural, Case:Locative]	Number Rule
/eller**in**/:[Number:plural, **PossPers:2**, PossNum:plural, Case:Locative]	Possessor Person Rule
/ellerin**iz**/:[Number:plural, PossPers:2, **PossNum:plural**, Case:Locative]	Possessor Number Rule
/ellerinizde/:[Number:plural, PossPers:2, PossNum:plural, **Case:Locative**]	Case Rule

At each stage, the relevant rule "sees" the right feature, and adds the appropriate suffix.

Study Exercise #6

Derive #34, *ellerimizi* 'our hands-accusative,' starting with an appropriate morphosyntactic representation and using the rules above.

This area is intensionally blank.

12. The bigger picture

Grammars like the one we are working on can produce a clearer understanding of large amounts of data. It's worth pondering, for instance, how many forms a Turkish noun can have. There are several choices to be made:

- Number: singular or plural, thus **two** possibilities
- Possessor Person: any of **three**: 1, 2, 3 ("his or her")
- Possessor Number: any of **two** (singular, plural)
- Case: nominative (no ending), accusative, dative, locative, plus ablative ("from"), genitive "'s", instrument ("with"), thus **seven** possibilities

Multiplying these out, every Turkish noun can appear in (at least) 2 x 3 x 3 x 7 = 84 forms, of which we covered only 40. It seems likely that Turkish speakers often must produce a new form for a noun, when they haven't heard a particular combination before.

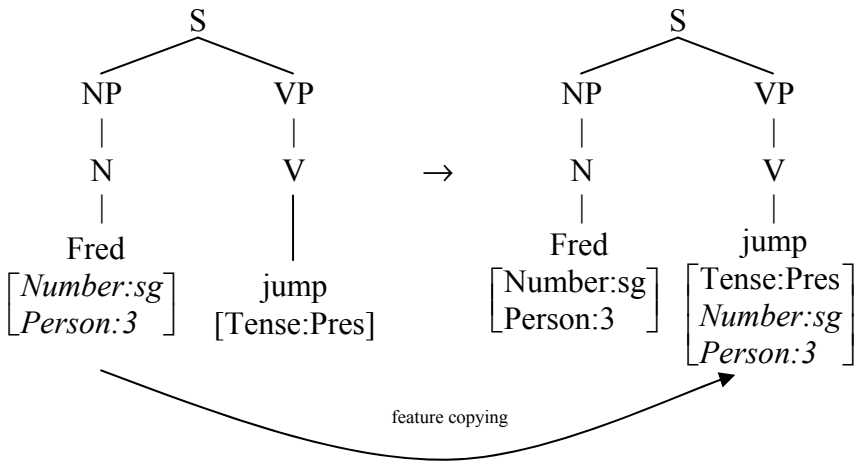
The Turkish nominal system is a fairly simple one; Turkish verbs, for instance, are quite a bit more complex. The most elaborate system I know of is the verbal system of Shona (Bantu), where (according to the linguist David Odden), the typical verb has about 10 trillion possible forms. Odden has developed a system that generates these forms using a rather complicated set of rules; most of the complication arises in getting the tones right.

It seems also likely that Turkish children or Shona children must also come up with a grammar; they could not possibly memorize every form of every word. We cannot know—yet—to what extent their grammars resemble our grammars, but the idea that through analysis and research we can get close to what they learn is a central idea of contemporary theoretical linguistics.

13. The source of morphosyntactic representations

The discussion in the last chapter showed how we can write a set of rules that create morphologically well-formed words through the successive addition of affixes by rule. But what do these rules *apply to*? There are various answers given by various linguists; here, we will examine just one fairly representative one.

The idea is that the syntax of a language builds up a feature structure for every stem that appears in a sentence. Thus, in an English sentence like *Fred jumps*, the fact that the subject of the sentence, *Fred*, is in the third person means that the rules of the syntax cause the feature [Person:3, Number:Singular] to appear on the stem *jump*; this is so-called "subject-verb agreement". Looking ahead to syntax, we can draw a syntactic structure[10] and the process of agreement:



We can assume that *Fred* is inherently [Number:sg, Person:3], since it is a proper name. The [Tense:Pres] must be assumed at the start as well, since it is part of the meaning of the sentence.

The operation above is part of syntax. Once the rules of the morphology get apply, the presence of these feature will cause a suffixation rule to apply, which attaches the suffix that we spell *-s*. Here is a sample rule:

3rd Sing. Present Rule

Suffix *-s* when the morphosyntactic representation contains [Tense:Pres,Person:Sing.,Number:3].

In sum, we have quite a bit of descriptive work to do in a complete grammar: the **syntactic component** arranges words in correct order and builds up the morphosyntactic representations, while the **morphological component** refers to the morphosyntactic representation in order to add the appropriate affixes.

14. German inflection: more than one feature per morpheme

Consider the person-number endings of German, in the present and past:[11]

[10] This is looking ahead, so don't be alarmed if the diagrams aren't clear. To clue you in a bit: S = Sentence, NP = Noun Phrase, VP = Verb Phrase, N = Noun, V = Verb, vertical line means "is part of".

[11] I'm glossing over some inessential complications arising from the fact that the stem *wart* ends in a [t].

Present			**Past**	
1 sg.	ich warte	'I wait'	ich wartete	'I waited'
2 sg.	du wartest	'you-sg. wait'	du wartetest	'You waited'
3 sg.	sie/er wartet	'she/he waits'	sie/er wartete	'she/he waited'
1 pl.	wir warten	'we wait'	wir warteten	'we waited'
2 pl.	ihr wartet	'you-plur. wait'	ihr wartetet	'you-plur. waited'
3 pl.	sie warten	'they wait'	sie warteten	'they waited'

Things here are a bit tricky: is the stem *warte*, with endings like -(zero), -*st*, -*t*, -*n*, -*t*, -*n*; or is it *wart*, with endings like -*e*, -*est*, -*et*, -*en*, -*et*, -*en*? Further evidence[12] indicates that the second is correct. Here are the forms broken up into position classes (shown with vertical alignment):

Present				**Past**			
1 sg.	ich	wart	e	ich	wart	et	e
2 sg.	du	wart	est	du	wart	et	est
3 sg.	sie/er	wart	et	sie/er	wart	et	e
1 pl.	wir	wart	en	wir	wart	et	en
2 pl.	ihr	wart	et	ihr	wart	et	et
3 pl.	sie	wart	en	sie	wart	et	en

The first thing to notice here is that unlike in Turkish, we are not going to be able to put forth an analysis in which the inflectional rules mention just one feature each—that is, with endings for person, endings for number, and endings for tense. Rather, German "bundles" features, in the sense that one single suffix manifests more than one feature at a time. Thus, for instance, the suffix -*est* is simultaneously the realization of second person and singular number. As a result, in the analysis below, I have mostly written rules that mention more than one feature at a time. For the six person/number *combinations*, one needs (at least six rules). Here is a grammar:

I. Tense Marking

Suffix -et when the morphophosyntactic representations contains [Tense:Past]

II. Person/Number Marking

Suffix:

-*e* if [Person:1, Number:Singular]
-*st* if [Person:2, Number:Singular]

[12] Notably, the imperative is just the plain stem: *Wart!* (wait).

-*e* if [Tense:Past, Person:3, Number:Singular]
-*et* if [Tense:Present, Person:3, Number:Singular]

-*en* if [Person:1, Number:Plural]
-*et* if [Person:2, Number:Plural]
-*en* if [Person:3, Number:Plural]

In fact, things are even more complicated than this. In precisely one place in the system—the 3rd person singular—the person-number ending is different in the past than in the present. The analysis takes account of this with the rules in boldface, which mention three features at once. Systems of inflectional morphology are well known for including asymmetries of this kind.

English has a very similar case: the -*s* of *jumps* simultaneously manifests [Number:Singular, Person:3, Tense:Present]. In fact, such "tangling" is found in languages all over the world.

Linguists speak of charts like the one at the top of this section as **paradigms**: a paradigm consists of all, or a systematic portion of, the inflected forms of a particular stem. We can also speak of things like the "present paradigm" (left column above) or the "past paradigm" (right column).

Subparadigms often involve **partial overlap**: thus, the German present and past verb paradigms overlap in all but the third singular.

15. What are the characteristic inflectional categories?

Every language has a set of inflectional categories, though the sheer amount of inflection can vary quite a bit. Mandarin Chinese has very little; Turkish and Finnish are quite richly inflected; English is closer to the Mandarin end of the scale.

Each inflectional category is expressed (in the theory we are using) as a feature within the morphosyntactic representations.

Here is a quick survey of some inflectional categories.

15.1 Nominal Inflection

Nouns and pronouns are often inflected for **number** (singular, plural, and occasionally dual, meaning exactly two; or even trial, exactly three). Pronouns are in addition inflected for **person** (first = includes speaker; second = includes hearer; third = neither).

15.1.1 Gender

In a number of languages nouns are inflected for **gender**; for instance, in German nouns can be masculine, feminine or neuter (as we can tell by the definite articles they take). In some cases, gender is semantically quite sensible:

der Mann 'the-masculine man'
die Frau 'the-feminine woman'

Extraordinarily, this system carries over—often quite arbitrarily—to the whole vocabulary of nouns, irrespective of meaning. Thus each of the three common items of silverware is a different gender in German:

der Löffel 'the-masculine spoon
die Gabel 'the-feminine fork'
das Messer 'the-neuter knife'

Thus gender is for the most part a purely formal device, not an expression of meaning.

Gender involves many other semantic correlations that have nothing to do with biological sex. From a web page intended to help learners of German[13] I quote the following rules:

60. Fabrics are predominantly masculine (*der Gingham*, *der Kaschmir*).
61. Heavenly bodies are predominantly masculine (*der Mond* [moon], *der Stern* [star]).
62. Forms of precipitation are predominantly masculine (der *Regen* [rain], *der Schnee* [snow]).
63. Bodies of water (restricted to inland streams, currents, and stagnant bodies) are predominantly masculine (*der See* [sea], *der Teich* [pond]).
64. Words denoting sound or loud noise or phonetic speech sounds are masculine (*der Donner* [thunder], *der Dental* [dental sound], *der Diphthong*).
65. Dance steps and popular music forms are masculine (*der Jazz*, *der Tango*).

Such generalizations are pervasive in gender languages. However, since there are usually exceptions of various sorts, it seems that people who know gender languages have memorized the gender of every word.

Gender is not just a property of familiar European languages; it is also found in Semitic, and a kind of system rather like gender (but with at least a dozen types) is found in Bantu languages.

15.1.2 Case

Nouns, and the syntactic phrases they occur in, are marked for **case**, which marks their role in the sentence. (See p. 15 above for discussion of case.)

15.2 Verbal Inflection

Very common is **tense**, which gives the time of action relative to the present: past (*I jumped*), present (*I jump*), future (*I will jump*), and other (for example, "remote past") tenses.

[13] http://montgomery.cas.muohio.edu/meyersde/kitchensink/german-gender/. Unfortunately, since I wrote this it seems to have been taken down (9/2010).

Aspect sets the boundaries of the action of the verb time, for instance, completed vs. non-completed action.

Verbs often agree with their subjects (and sometimes their objects as well) in features for nouns (as shown above in section 10). These features include person (*I am*, *you are*, *she is*), number (*I am, we are*), gender.

Verbs, particularly second person forms (see below) can also be inflected for the **degree of familiarity** of the addressee; thus English used to make a distinction between (say) *thou believest*, addressed to intimates, children, and animals; and *you believe*, for less familiar addressees. Most European languages, Javanese, Persian, and Japanese have such systems today.

In various languages verbs are inflected for **degree of belief**; thus from my German textbook:

Er sagte, dass er krank ist.
He said that he sick is-indicative
"He said he is sick" (acknowledging a belief held by all)

Er sagte, dass er krank sei.
He said that he sick is-weak subjunctive
"He said he is sick" (and it's not necessarily true)

Er sagte, dass er krank wäre.
He said that he sick is-strong subjunctive
"He said he is sick" (and the speaker doubts it)

Related to this is the category of verbal inflection in many languages which marks information known only by hearsay rather than by direct witness; this is common in American Indian languages.

15.3 Adjectival Inflection

Adjectives typically don't have their own inflectional categories, but acquire inflection by agreeing with the nouns they modify; thus German:

ein guter Löffel 'a-masculine good-masculine spoon'
eine gute Gabel 'a-feminine good-feminine fork'
ein gutes Messer 'a-neuter good-neuter knife'

16. The principle of obligatory expression

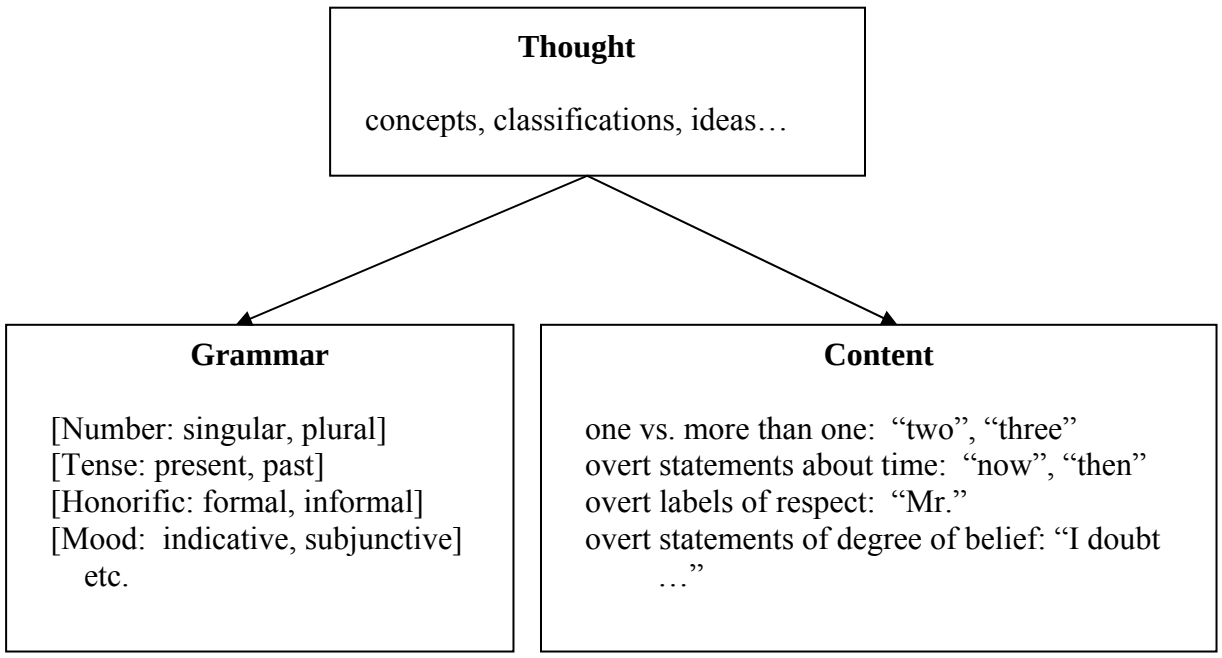
An important aspect of inflectional morphology is that it often involves *obligatory choices*. When in English one says "I bought the book", it specifically means "one book", not "any old number of books". Likewise, "books" necessarily implies the plural. To avoid the obligatory

choice, one must resort to awkward circumlocutions like "book or books". There are other languages (for example, Mandarin) that work quite differently. Thus, the following sentence:

wɔ^{3}	mai^{3}	ʂu^{1}
I	buy	book

is quite noncommittal about how many books are bought. (It is also noncommittal about when the buying takes place.) Thus an important aspect of the grammar of languages is the set of choices they force speakers to make when speaking; this is determined by their systems of inflectional morphology.

Fundamentally, there is a bifurcation between the two ways that thought is embodied in language. The following diagram tries to make this clearer.



Languages differ: each one takes a subset of the fundamental ideas, and *grammatically codifies* them. By this I mean that in some particular language, a particular concepts get expressed as grammatical features, and that these features are included in the morphosyntactic representations and thus integrated into the grammar. Whenever this happens, the expression of the concept in question becomes obligatory—since you have to obey the grammar of your language when you speak. Alternatively, a concept can remain uncodified grammatically, and the speaker is free to express it or not as she chooses, through choice of words and other means.

On the whole, the forms of thought that can get integrated into grammar are, as we might expect, the ones that are most omnipresent in our lives: time, number, belief vs. doubt, and the fundamental aspects of conversations (speaker/hearer/other and their social relations.)

17. The typology of inflection

A rough way of characterizing languages for their system of inflectional morphology is the following three way split:

isolating / agglutinative / inflecting

- A language is **isolating** to the extent that it has little or no inflectional morphology. Examples: English, Chinese.

- A language is **agglutinating** if it has a rich inflectional morphology, and each morpheme (tends to) expresses a single morphosyntactic feature. Thus, words tend to be long but have a very clear structure. Examples: Turkish, Swahili.

- A language is **inflecting** (bad term, since it's ambiguous) if it has a rich morphology, and morphemes express *multiple* features. Example: Latin.

  somnus, somni. **nm.**, sleep.

Case	**Singular**	**Plural**	**Meaning of case**
Nominative	*somn-us*	*somn-i*	for subjects
Genitive	*somn-i*	*somn-orum*	for possessors
Dative	*somn-o*	*somn-is*	"to"
Accusative	*somn-um*	*somn-os*	for objects
Ablative	*somn-o*	*somn-is*	"from"
Locative	*somn-i*	*somn-is*	"at, on, in"
Vocative	*somn-e*	*somn-i*	for calling to someone

The point of the example is that (for instance) *-us* packs a considerable bundle of information: it tells us that *somnus* is nominative, that it is singular, and (with a few exceptions we will ignore) that it is masculine. We could write the rule like this:

Add -us if [Case:Nominative, Number:singular, Gender:masculine]

In general, the agglutinative languages will have just one, or a few, features mentioned in each rule, whereas the inflecting languages tend to have more. (This is just a more formal way of characterizing the basic distinction.)

All else being equal, inflecting languages will tend to have shorter words than agglutinating languages. However, there is usually a "cost" to this terseness: typically, in an inflecting the same ending often serves multiple purposes, so words tend to be inflectionally ambiguous.

WORD FORMATION

[xxx note for next time: this is missing a discussion of ambiguous words with two derivations]

18. Rules of word formation

The other function of morphology is to expand the stock of words in the language, by forming new words from old. Often linguists refer to this process as **derivational morphology**; I will try to stick to the term **word formation** since it is more precise.

For example, given that *identify* is an existing word of English, a rule of English word formation can create a new word, *identifiable*. From this another rule can provide *identifiability*, and from this yet another rule can create *unidentifiability*.

18.1 Rules of Word Formation

Consider some words formed with the English suffix *-able*:

-able: washable lovable thinkable growable doable

We wish to write the **word formation rule** that attaches *-able* to an existing word to form a new one. There are three kinds of information that must be included in the rule.

First, there is a change of **form**; the existing word is augmented by the suffix. This could be expressed with the formalism:

$$X \rightarrow X + \text{-able}$$

Second, there is a change of **meaning**: *Xable* means "able to be Xed". We will not formalize this, since the task of representing meaning is far too big to take on in this context. Finally, there is often a change in **part of speech**. *-able* attaches to Verbs (e.g. *wash*, *love*, *think*, etc.) and forms Adjectives.

All three aspects of the rule can be expressed more compactly in the following abbreviated form:

-able Rule

$[X]_{Verb} \rightarrow [[X]_{Verb}$ -able $]_{Adjective}$ Meaning: "able to be V'ed"

You can read this as follows:

> "Take some string X that is a Verb. Add to it the string əbl. Classify the resulting string as an Adjective."

Rules of word formation can be shown applying in a formal derivation. As before, we label each line of the derivation according to the rule that applies. Thus, for instance, here is a derivation for *washable*:

[wash]$_{Verb}$	Stem
[[wash] $_{Verb}$ able]$_{Adjective}$	-able Rule

Here are some further word formation rules of English. To express the derivation of words in *-ity*, (for example, *divinity*, *obscurity*, *obesity*, *insanity*, *sensitivity*), we could write the rule

-ity Rule

[X]$_{Adjective}$ → [[X]$_{Adjective}$ ity]$_{Noun}$ Meaning: "quality of being Adjective"

This creates structures like [[obes]$_{A}$ ity]$_{N}$.

To handle words formed with the prefix *un-*, (*unfair*, *unkind*, *unjust*, *unspoken*, *unattested*, *unidentifiable*) we could write the following rule:

un- Rule

[X]$_{Adjective}$ → [un [X]$_{Adjective}$]$_{Adjective}$ Meaning: "not Adjective"

The rule creates structures like [un [kind]$_{A}$]$_{A}$.

19. Stacked derivation

At least in English, the idea of the position class, covered above for inflection, is not relevant for derivation. Rather, the rules of derivation can apply freely, provided their requirements are met. For example, we can derive the long word *unmindfulness* by applying the following rules in succession:

[mind]$_{Noun}$	stem
[[mind]$_{Noun}$ ful]$_{A}$	**-ful Rule:** [X]$_{N}$ → [[X]$_{N}$ ful]$_{Adj}$
[un[[mind]$_{Noun}$ ful]$_{Adj}$]	**un- Rule**
[[un[[mind]$_{Noun}$ ful]$_{Adj}$]ness]$_{Noun}$	**-ness Rule:** [X]$_{Adj}$ → [[X]$_{Adj}$ ness]$_{Noun}$

With a bit of strain, it's possible even to have the same inflectional rule *apply twice in the same form*. Here is an outline derivation for the (novel) word *industrializational*.

industry
industrial
industrialize
industrialization
industrializational

Although the last word is a bit of stretch, you can see that the result has "double application" of the rule that attaches *-al*.

Study Exercise #7: give the rules and derivation for *industrializational*.

This area is intensionally blank.

The repetition of the same suffix in the word is fairly good evidence that English word formation does not involve position classes. The multiple appearances result from the inherent property of word formation, that the rules apply where they can. In contrast, in the position-class systems seen in inflection, the rules apply in a strict arrangement defined by blocks.

20. What meanings are expressed by word formation rules?

The short answer here is "almost anything," as we'll see shortly. But there are some core meanings.

20.1 Changing syntactic category

Perhaps the most common purpose of word formation rules is to **change syntactic category**; we may want to say pretty much the same idea, but using the stem as a noun instead of a verb:

He had trouble concentrating.	(verb)
He had trouble with his concentration.	(noun)

[14] We may ignore the spelling change, assuming our focus is on spoken English.

In English, there are word formation processes that can change between any pair of the three major syntactic categories of verb, noun, and adjective:

Verb to noun: -ation, as above
usual meaning: "the process of Verbing"
Noun to verb: -ify, -ize (as in *classify*, *demonize*)
usual meaning: "to do something involving Noun"

Adjective to noun: -ness, -ity
usual meaning: "the quality of being Adjective"
Noun to adjective: -ish, -esque
usual meaning: "resembling Noun"

Verb to adjective: -ent (as in *effevescent*)
usual meaning: "tending to Verb"
Adjective to verb: -ify (as in *clarify*, *humidify*)
usual meaning: "render Adjective"

20.2 Changing the number of participants in a verb

Verbs often have rules of word formation that change the number of participants. Consider the Persian verbs below:

res-idan	'reach-infinitive'
res-aːn-idan	'send-infinitive'
xaːb-idan	'to sleep'
xaːb-aːn-idan	'to put to sleep'[15]

Here, we can take a verb that has just one participant (the one who is reaching, or sleeping), and make from it a verb that has an additional participant (the one who causes to reach, or causes to sleep). This is called a **causative verb**. English has no such word formation process, and uses syntactic constructions to express causation ("He made them sleep").

For Persian, the rule could be expressed as:

-aːn Rule

$[\ X\ ]_{Verb} \rightarrow [\ [\ X\ ]_{Verb}$ aːn $]_{Verb}$ Meaning: "to cause to Verb"

[15] In the International Phonetic Alphabet, the symbol [ː] designates a long vowel.

20.3 The grand miscellaneous

Although the two purposes of word formation rules just given are probably the most common across languages, individual languages can include word formation rules of marvelous specificity. Among my favorites is one in Ilokano (Philippines), with a process that derives from a verb a new verb meaning "to pretend to be verbing"

daʔit	'to sew'
ʔagindadaʔit	'to pretend to be sewing'
saʔit	'to cry'
ʔaginsasaʔit	'to pretend to cry'
d͡ʒanitor	'to work as a janitor'
ʔagind͡ʒad͡ʒanitor	'to pretend to be a janitor'

English has some very specific word formation processes:

bowl-arama, *carpet-arama*	'grand emporium for X or X-ing'
Stalin-ism, *Mao-ism*	'doctrine propounded by X'
pay-ola, *shin-ola*, *plug-ola*	'bribery involving X'

21. The ordering of word formation and inflection

It is at least a strong cross-linguistic tendency—perhaps a universal of language—that rules of word formation apply before inflectional rules. Thus, for instance, in English it is possible to have words like *nullifies*, which are derived as follows (I'll use IPA transcription to duck issues involving spelling):

$[nʌl]_{Adj.}$	root: 'null'
$[[nʌl]_{Adj.}ɪfaɪ]_{Verb}$	Word formation rule: Adjective + /-ɪfaɪ/ → Verb
$[[nʌl]_{Adj.}ɪfaɪ+z]_{Verb}$	Inflectional rule: X → Xz when [Verb, +3rd person, +singular, +present]

The opposite rule ordering would have derived *[nʌlzɪfaɪ], so that the inflectional suffix would appear "inside" the derivational suffix. Cases of this sort are rare at best.

This has implications for when you analyze a new language: typically it is possible to work out the inflection—appearing on the "outside" of the word, and then work with the leftover material and find the word formation rules.

22. Compounding

A widespread view of compound words is that they are a form of word formation. They differ in that rather than attaching an affix to a stem, they concatenate (chain together) two stems.

Here is a same rule of compounding for English nouns:

$[X_1]_{Noun} + [X_2]_{Noun} \rightarrow [[X]_{Noun} [X]_{Noun}]_{Noun}$
Meaning: "an X_2 that has something to do with X_1."

Thus: *boat house* (structure: $[[\text{boat}]_{Noun} [\text{house}]_{Noun}]_{Noun}$) is a house that has something to do with boats (for example, you keep boats inside it). A *houseboat* is a boat that functions as a house.

The word *tigerbird* is probably not familiar to you, but you can guess part of its meaning simply by knowing how to speak English: you know it is a kind of bird (and not a kind of tiger!), and that it has something to do with tigers (perhaps it is striped like a tiger, or it likes to roost on top of sleeping tigers, or that it fights like a tiger, and so on).

Compounds like *houseboat, boathouse*, and *tigerbird*, derived by the rule given above, are said to be **headed**: the "head" of *houseboat* is *boat*, because *a houseboat is a boat*. Likewise *house* is the head of *boathouse*, because a boathouse is a house, and *bird* is the head of *tigerbird*.

In English, most compounds have at most one head, but other languages allow "double-headed" compounds, for instance when "mother-father" is used to mean "parents." One possible English example is *Austria-Hungary*, which designated the country of the 19th century that included both Austria and Hungary. You'll find some double-headed compounds on this week's homework. You can derive them with a rule that's exactly like the compound rule given above, except that the meaning has to be stated differently.

It is possible to form a compound from two words one of which is itself a compound. For example, we can combine the compound *law degree* with the word *requirement* to get the complex compound *law degree requirement*. This compound can in turn be combined with *changes* to get *law degree requirement changes*; and so on. The following example suggests that the process is essentially unlimited:

eggplant	'plant shaped like an egg'
eggplant plant	'factory for manufacturing eggplants'
eggplant plant plant	'factory for manufacturing factories for manufacturing eggplants'

Thus compounding is like other forms of word formation in that it applied freely, rather than in the strict "assembly line" fashion of inflectional rules.

22.1 The spelling of compounds

The spelling system of English is inconsistent with regard to compounds; some are spelled without a space between the component words and some are spelled with a space. It is important to realize that an expression spelled with a space can still be a compound.

One can argue for this in two ways. First, consider German: it is customary in German to spell all compounds without a space between the component words. That is, the English practice is more or less an accident; given that other languages go the other way.

More important, there are linguistic arguments that compounds spelled with spaces are just like compounds spelled without them. Note first that, in the case of a genuine NP of the form Adjective + Noun, it is possible to insert an extra adjective between the adjective and the noun. For example, we can take the NP *large cake* and add an additional adjective to get *large round cake*. But if we start with a compound, it is impossible to get an additional adjective in the middle. For example, starting from *pancake*, we cannot get **pan round cake*. The basic point is the while the noun of a NP can be modified by an additional immediately preceding adjective, a noun that is the second word of a compound cannot.

This fact provides us with a test to determine whether an expression really is a compound, even if it is spelled with a space. For instance, we can show that *carrot cake* is a compound by trying to place an adjective in the middle: **carrot large cake*. Other examples also show that expressions spelled with a space can be compounds:

coal scuttle	*coal dirty scuttle
lap dog	*lap slobbery dog
can opener	*can handy opener

23. Summing up so far

The picture of morphology thus far drawn is like this.

First, languages have means of expanding their inventory of words (more precisely: of stems). The rules of word formation add affixes to stems to derive new stems, which have new meanings. These meanings can be common, characteristic ones (like "the quality of being Adjective", "to cause to Verb"), or exotic ones (like "emporium for selling Noun"). Compounding likewise expands the stock of stems, creating either single-headed compounds (like *boathouse*) or, in some languages, two-headed ones (like *Austria-Hungary*).[16] There is in principle no limit to "when" a derivational rule can apply; it simply looks for the right kind of base form and applies optionally.

The stems that result, whether they are basic or derived, are used in sentences. In a sentence context, the rules of the syntax (as yet undiscussed) provide each stem with a morphosyntactic representation, that is to say, a bundle of inflectional features. These features are specific to a particular language, although a number of features like [Case:Accusative] or [Number:Plural] occur repeatedly in languages. The features are referred to by the rules of inflectional morphology, which add affixes in order to express their content overtly. It is generally possible to arrange the affixes of an inflectional system into "slots", where each word has at most one affix per slot. In terms of rules, the slots are expressed by having one rule per slot; each rule

[16] For thoroughness: there are also compounds with **implied** heads, like *airhead*. These typically have an unstated head, usually meaning "person" or "thing". Thus *airhead* means, essentially, "air-headed person", "person with head filled with air". Similarly: *pick-pocket* "person who picks pockets"; *stopgap* "thing that stops gaps".

attaches the affix that corresponds to the features given in the morphosyntactic representation of the stem.

As a consequence of this scheme, inflectional morphology, being attached by rules that apply "later", occurs on the "outside" of a word; that is to say, further from the stem than inflectional morphology.

24. Phonological realization in morphology

When I say "phonological realization", I mean the arrangement of the phonological material (speech sounds) that realizes the morphological categories, whether they be derivation or inflectional. I would guess that a large majority of all morphology (in the narrow sense that excludes compounding) is prefixation, suffixation, or compounding. All three are **concatenative**, in the sense that they string together sequences of speech sounds. They are the meat and potatoes of morphology, and are found in most languages.

But concatenation is not the only way you could carry out an inflectional or derivational process: segments can be interpolated, or copied, or altered in their phonetic content. Below, I will give some cases, and present ways that explicit rules can be written for them.

Note that all of these "fancy" forms of morphology can be used for both inflection and derivation—on the whole, the *functions* of morphology (grammatical or derivational) can be studied independently of the changes in phonological material that carry out these functions.

24.1 Infixation

The following data from Bontoc (Philippines) illustrate **infixation**, which can be defined as insertion of segments into some location inside the base:

fikas	'strong'	fumikas	'he is becoming strong'
kilad	'red'	kumilad	'he is becoming red'
bato	'strong'	bumato	'he is becoming stone'
fusul	'enemy'	fumusul	'he is becoming an enemy'

It's reasonably clear that this is a derivational process, and that the brackets we'll need are something like this:

$[\ X\ ]_{Adj} \rightarrow [[\ X\ ]_{Adj}]_{Verb}$
Meaning: "become Adjective"

$[\ X\ ]_{Noun} \rightarrow [[\ X\ ]_{Noun}]_{Verb}$
Meaning: "become Noun"

But how to express the infixation? The important part here is to be precise about just where the infixed material should be inserted. We will use here a method that makes uses of **variables** and **subscripts**.

The variables we have seen already with simple rules in prefixation and suffixation, as in [X]$_{Adj}$ → [[X]$_{Adj}$ ness]Noun there re are various methods proposed; we will follow a rather simple one. Instead of simply expressing the speech sounds of the base with a simple variable X (meaning: any sequence), we will give this part of the rule more structure, sufficient structure to specify where the infix goes. Doing just the adjective case, we have:

[	C	V	X	]$_{Adj}$	→	[[	C	um	V	X	]$_{Adj}$	]$_{Verb}$
	1	2	3				1		2	3		

Meaning: "become Adjective"

You can read the rule above as follows:

> "An adjective base consisting (precisely) of a consonant, followed by a vowel, followed by anything, is converted to an verb by inserting the sequence *-um-* after the consonant."

Some details: the numbers under the terms of the rule are included to make sure we are clear on what matches up with what (important if, for example, a rule contains more than one C). "C" and "V" are very commonly used in linguistics as abbreviations for "consonant" and "vowel". (The vowels in the examples above are [i, a, o, u].)

Applying the rule to the form, we have the following matchup:

[	f	i	kas	]$_{Adj}$		[[	f	um	i	kas	]$_{Adj}$	]$_{Verb}$
[	C	V	X	]$_{Adj}$	→	[[	C	um	V	X	]$_{Adj}$	]$_{Verb}$
	1	2	3				1		2	3		

Infixation is not common in English. You are probably familiar with the colloquial expression *fan*[ˈfʌkən]*tastic*,[17] in which a taboo word is placed in the middle of the stem, as a kind of infix. Cockney dialect recognizes similar constructions like *abso-bloomin'-lutely*. From time to time linguists have proposed analyses that predict, for any given word, where the expletive can be infixed; this turns out to be a surprisingly difficult area for analysis.[18]

Infixes are normally written with both preceding and following hyphens, since they have two "joining points": *-um-*.

---

[17] IPA symbols: [ʌ] is the vowel of *cut*, [ə] is the second vowel of *taken*.

[18] A pretty good analysis appears in John McCarthy (1981) "Prosodic structure and expletive infixation," *Language* 58, 574–590, available at http://people.umass.edu/jjmccart/prosodic_structure_and_expletive_infixation.pdf

A caution concerning infixes: not all morphemes in the middle of a word are infixes. Many of them are prefixes/suffixes that happen to have had additional material added to their left/right: in *ex-vice-president*, *vice-* is a prefix, not an infix. You can identify the infixes by their ability to occur in the middle of a morpheme.

Infixes are normally considered to be affixes (like prefixes and suffixes); the English cases above, a curious sort of "compounding infixation", are a curious exception.

24.2 Reduplication

Reduplication is a morphological process in which all or part of a word is copied. For example, in Yidiɲ, an aboriginal language of Australia, the intensive form of verbs is created by adding to the beginning of a word a copy of the first two syllables of the word:

mad^{j}indan	'walk up'
mad^{j}inmad^{j}indan	'keep walking up'
d^{j}ad^{j}aman	'jump'
d^{j}ad^{j}ad^{j}ad^{j}aman	'jump a lot'[19]

Here is a rule

$[\ X\ ]_{Verb} \rightarrow [\ [\ X\ ]_{Verb}\ [\ X\ ]_{Verb}\ ]_{Verb}$ Meaning: 'to Verb repeatedly'
1 1 1

In Samoan (S. Pacific), the plural form of a verb is formed by reduplicating the second-to-last syllable of a verb:

nofo	'he sits'	nonofo	'they sit'
pese	'he sings'	pepese	'they sing'
savali	'he walks'	savavali	'they walk'
atamaʔi	'he is wise'	atamamaʔi	'they are wise'

One might think of reduplication as a morpheme whose content varies, dependent on the segments that it is copied from.

We can use our numerical subscript notation to express the Samoan process above unambiguously:

Samoan Plural Subject Reduplication

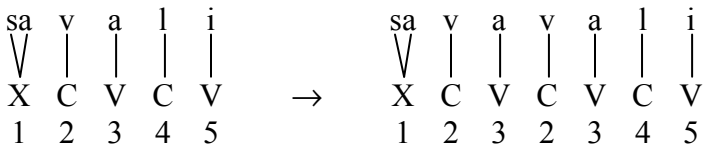
X	C	V	C	V	→	X	C	V	C	V	C	V
1	2	3	4	5		1	2	3	2	3	4	5

[19] d^{j} is the IPA for a voiced lamino-palatal stop, similar to the English "j" sound.

when the morphosyntactic representation contains [Number:Plural]

The rule tell us to count off the final CVCV of a word, and copy its first CV sequence (what is numbered "23" in the rule). Here is a derivation for *savavali*

savali[Number: Plural]



We can't formalize the Yidiɲ rule (yet) because we haven't yet covered the theory of syllables.

The meanings expressed by reduplication are often "symbolic" in nature; languages often use reduplication to express plurals, intensiveness, repeated action, and the like.

Study Exercise #8

Write the rule for forming causatives in Ateso (Nilotic family, Uganda).

duk	'to build'	tuduk	'to cause to build'
lel	'to be glad'	telel	'to gladden'
ɲam	'eat'	taɲam	'feed'
wadik	'write'	tawadik	'cause to write'
cak	'throw'	tacak	'cause to throw'

This area is intensionally blank.

24.3 Segment Mutation

Morphology sometimes is carried out by finding a particular segment (perhaps in a particular context) and changing it into something else. For instance, a fair number of English verbs form their past tense by changing the stem vowel [ɪ] into [ʌ]:

Present	Past
cling	*clung*
sling	*slung*
fling	*flung*
string	*strung*
wring	*wrung*

While these are irregular verbs (and thus are probably memorized), the process is nevertheless a little bit productive: forms have arising in dialects like *sing - sung*, *ring - rung*; and many children explore the possibility of *bring - brung*. In experiments, people asked to provide a past tense for the made-up verb *spling* often volunteer *splung*.

We can state this rule as follows, noting that a crucial element in (most of) these verbs is the presence of a following [ŋ], the "ng" sound:

[	X	ɪ	ŋ	]	→	[	X	ʌ	ŋ	]
	1	2	3				1	2	3	

when the morphosyntactic representation contains [Tense:Past]

24.4 Morphological Conversion

Consider:

Fred likes to jump.	His last jump was 20 feet.
Chomsky likes to talk.	His last talk was attended by 500 people.
We need to think.	We had a good long think.

These cases have simplest possible string operation of all; that is, nothing changes. Such rules can be expressed as follows:

Answer: **Verb-Noun Conversion**

$[\ X\ ]_{Verb} \rightarrow [\ [\ X\ ]_{Verb}\ ]_{Noun}$
meaning: "an instance of Verbing"

There is also a rule that goes in the opposite direction, for data like these:

Noun	Verb
a mop	I mopped the floor.
a fax	I faxed the message.
a hammer	I hammered the nail in.

The rule needed is something like this:

Verb-Noun Conversion

$[\ X\ ]_{Noun} \rightarrow [\ [\ X\ ]_{Noun}\ ]_{Verb}$
meaning: "do something crucially involving Noun"

These are simply word formation rules that carry out no affixation (or any other change). There is no reason to exclude them from the theory, and indeed they seem to be pretty common among languages. The usual term for rules of derivational morphology that do nothing but change category is **conversion**.

One might ask why we want rules going in both directions. The best answer, perhaps, is that the morphological base form in each case is somehow semantically primary: a *jump* is what happens when you engage in jumping (rather than: "jumping is what happens when you execute a jump"); *mopping* is the activity you do with a mop (rather than: a mop is the device you mop with).

Occasionally in older linguistic works one will find the claim "Language X lacks a distinction between nouns and verbs." This is currently viewed as rather implausible; instead, one could say that in Language X, morphological conversion between nouns and verbs is highly productive, so most nouns stems can be used as verb stems and vice versa. In any language, there are good *syntactic* reasons to want to have a distinction between nouns and verbs.

Chapter 4: Syntax I — Phrase Structure

1. Knowledge of syntax

A theme of Chapter 1 was *implicit knowledge*: people show they possess such knowledge in that it is reflected in the patterning of their language, but they cannot directly intuit the form of that knowledge. Here, we will focus on the kinds of implicit knowledge encountered in studying **syntax**, which is the study of sentence structure. What do speakers known when they know the syntax of a language?

(1) They have intuitions about **grammaticality**. A sentence is grammatical if it is syntactically well-formed; if it counts as "part of the language." Grammaticality is distinct from merely making sense. Consider, for example, the following series of sentences:

She wonders if Alice is going to like Bill.	
Who does she wonder if Alice is going to like?	(answer: *Bill*)
*Who does she wonder if is going to like Bill?	(answer: *Alice*)

As far as meaning goes, the third sentence is as sensible as the second. It is only ungrammatical. Similarly, sentences like **Bill and Fred think that I like each other* (p. 5) have a perfectly sensible interpretation, but are ungrammatical. Sentences like *Colorless green ideas sleep furiously*, however, are quite grammatical but are nonsense.

(2) Our implicit knowledge of syntax cannot possibly take the form of a list of sentences. No such list could be stored in a finite mind, as there are an infinite number of grammatical sentence in English (or any other human language). It is easy to show this. A list of sentences like the following:

Alice likes Fred
John said that Alice likes Fred
Bill believes that John said that Alice likes Fred

can be extended onward to infinity.

Since syntactic knowledge cannot take the form of a list, we are led to the more reasonable hypothesis that we implicitly know a set of syntactic **rules**; the rules enable us to create novel sentences (a potentially infinite supply of them) on the spot. Just what sort of rules could do this will become clear later on.

(3) Speakers have the ability to recognize and manipulate systematic relations among sentences. For example, the following set of four sentences:

Bill shaved Fred	(active statement)
Did Bill shave Fred?	(active question)
Fred was shaved by Bill	(passive statement)
Was Bill shaved by Fred?	(passive question)

forms a clear pattern that can be duplicated by a speaker of English for an indefinite number of other sentences.

(4) Sentences are not simply strings of words; they also involve grouping of words into larger units. The easiest way to show this is with sentences that have two meanings, traceable to two different groupings of the words:

There were (old)(men and women)
There were (old men)(and women)

They (danced) and (sang the first number)
They (danced and sang)(the first number)

Sue saw (the man)(with the telescope)
Sue saw (the man with the telescope)

Bill (gave)(the Chinese vases)	(...to someone)
Bill (gave)(the Chinese)(vases)	(...even though they already had vases)

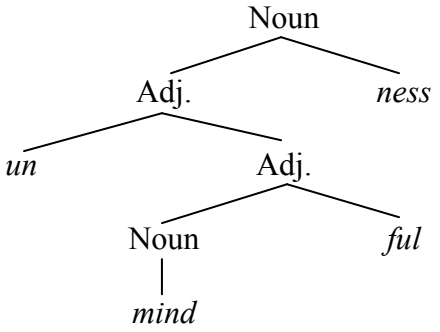
2. Constituent structure

The first step in developing a syntactic theory is to devise a formal notation for the structure of sentences. We wish to express the fact that the words of a sentence form groups of various kinds; that the groups are themselves grouped into larger units, so that a sentence forms a single complex structure. Linguists normally use a *tree* notation to do this.

Trees are actually applicable to morphology as well as syntax, so I'll illustrate the idea with a morphological example done earlier. We gave a derivation for the word *unmindfulness*, as follows.

$[\text{mind}]_{\text{Noun}}$	stem
$[[\text{mind}]_{\text{Noun}}\ \text{ful}]_{\text{A}}$	**-ful Rule:** $[\ X\ ]_N \rightarrow [[\ X\ ]_N\ \text{ful}\ ]_{\text{Adj}}$
$[\text{un}[[\text{mind}]_{\text{Noun}}\ \text{ful}]_{\text{Adj}}]$	**un- Rule**
$[[\text{un}[[\text{mind}]_{\text{Noun}}\ \text{ful}]_{\text{Adj}}]\text{ness}]_{\text{Noun}}$	**-ness Rule:** $[\ X\ ]_{\text{Adj}} \rightarrow [[\ X\ ]_{\text{Adj}}\ \text{ness}\ ]_{\text{Noun}}$

The output, $[[\text{un}[[\text{mind}]_{\text{Noun}}\ \text{ful}]_{\text{Adj}}]\text{ness}]_{\text{Noun}}$, can be shown more clearly with a tree, as follows:

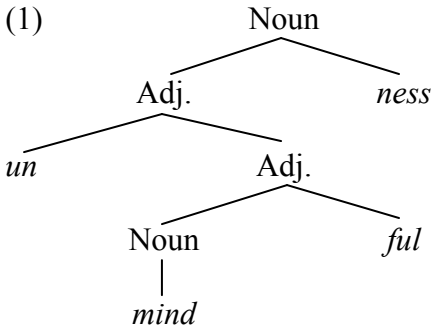


As you can see, the tree metaphor is a bit odd, since linguistic trees are virtually always drawn upside down relative to biological ones. The virtue of the term "tree" is that it is briefer than "root system", 'tree located in Australia", or whatever...

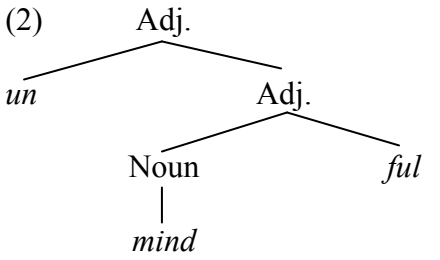
Definition: any unit in a tree is called a **node**. The nodes in the tree above are { Noun, Adj., -*ness*, Adj. (again), Noun, -*ful*, and *mind*.

When you combine a node with *all the material you can reach by going "downhill" from that node*, the result is called a **constituent**. The constituents of the tree just given are:

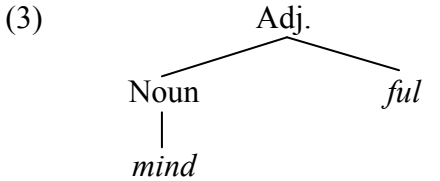
(1)



(2)



(3)



(4) Noun
|
mind

In addition, the elements *un*, *mind*, *ful*, and *ness*, each of which is at the "bottom" of the tree, are called **terminal nodes**. The terminal nodes are constituents, too, though in informal practice they are usually left out of a list of constituents.

If you compare the tree with the bracketed version of *unmindfulness* given above, you'll see that every constituent that isn't a terminal node corresponds to a bracketed unit.

(1) = $[[un[[mind]_{Noun}\ ful]_{Adj}]ness]_{Noun}$
(2) = $[un[[mind]_{Noun}\ ful]_{Adj}]$
(3) = $[[mind]_{Noun}\ ful]_{Adj}]$
(4) = $[mind]_{Noun}$

So the two notations are equivalent. For syntax, we'll mostly use trees, because syntactic structures tend to be quite a bit more complex than morphological structures, and the tree notation is much more readily apprehended by the eye.

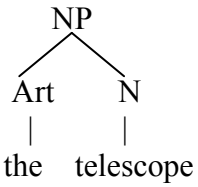
You can *name* a constituent by pronouncing its terminal nodes in order. So, for example, you can say things like: "in the word *unmindfulness*, *unmindful* is a constituent, and *mindfulness* is not a constituent."

3. Trees in syntax

Drawing the syntactic trees for sentences depends in part on our knowledge of the meaning of the sentence, and in part on our knowledge of the grammar (the syntactic part of the grammar) of the language. The idea is to think through the meaning, and locate the syntactic units.

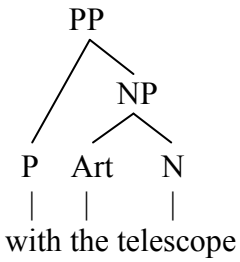
Consider the sentence *Sue saw the man with the telescope*. This sentence actually has two meanings (either Sue used a telescope for her observations, or the man was carrying one). Often, different meanings correspond to different trees, so let us for present purposes assume the meaning in which the man was carrying the telescope. I will build the tree from the bottom up.

I believe it is pretty intuitive that *the telescope* is a linguistic unit. We show this with a tree diagram.



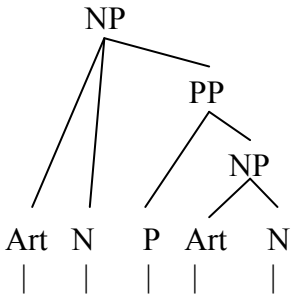
What does this diagram mean? The basic idea is that *the* is classified as an Article, and *telescope* as a Noun, and the entire unit is a **Noun Phrase**, abbreviated NP.[24] This NP can stand alone, for instance as the answer to the question "What did the man have with him?"

Let's move on to the next larger unit. If want the answer to "which man", we could say (rather tersely):



This is a **prepositional phrase** (PP), with the preposition *with* (P). The constituent *the telescope* is contained within the constituent *with the telescope*. One can also say it like this: *the telescope* is **embedded in** *with the telescope*.

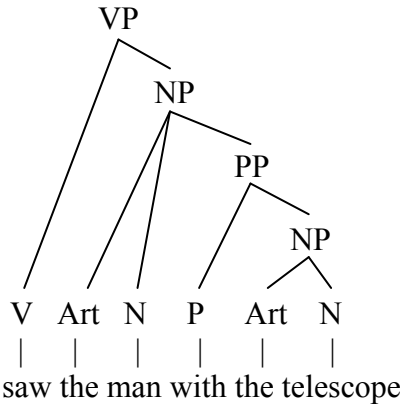
We can continue, building up the structure as follows:



(Which man are we speaking of?)

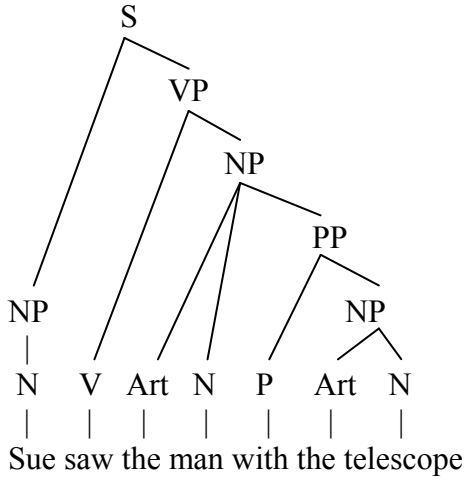
This is a bigger Noun Phrase, involving a *man*, further identified with the article *the* and the Prepositional Phrase *with the telescope*. It could answer the question, "Which man are we speaking of?

[24] I'll assume you learned in school how to identify articles, nouns, verbs, helping verbs, adjectives, and prepositions. If you'd like to review this material, please consult this help page: http://www.linguistics.ucla.edu/people/hayes/20/resources/CheckingPartsOfSpeech.pdf.



This is a **Verb Phrase**, whose verb is *saw*. What we created before can now be seen to be the object of this verb. The Verb Phrase could answer the question, "What did Sue do?".

Ultimately we arrive at a structure for the complete **Sentence**, abbreviated S:



Here, we have a **subject**, in the form of the NP *Sue*, and a **predicate**, in the form of the VP *saw the man with the telescope*.[25]

Trees of this kind will be the formalism we will be using for syntactic structure. Trees in syntax are also referred to as **phrase structure diagrams**.

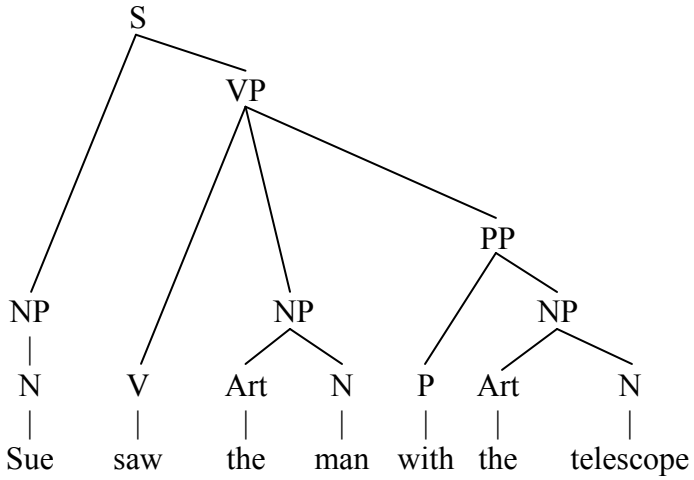
4. Phrase structure and ambiguity

As mentioned above, one of the first and most obvious descriptive benefits of constituent structure is that it provide a clear account of the ambiguity of many sentences and phrases.

For example, with the tree just given, the meaning we had in mind was that "with the telescope" identifies the particular man that Sue saw (for example, he was walking down the

[25] For why we are treating *Sue* as a full NP, not just an N, see below.

street holding the telescope in its carrying case). For the (probably more obvious) meaning that Sue used the telescope to see the man, we would have:



N = Noun
V = Verb
Art = Article
P = Preposition

What is at issue is where "with the telescope" is *attached* in the tree: is it part of VP or of NP? We can clarify this concept a bit further with some terminology.

5. Heads and modifiers

Many (but not all) syntactic constituents possess a **head**. In a Noun Phrase (NP), the head is a Noun, and similarly the head of a Verb Phrase is a Verb, of a Prepositional Phrase is a Preposition, and (as we'll see) of an Adjective Phrase is an Adjective. Intuitively, the head is the "core" of a constituent, what expresses the essence of its meaning.

You can think of heads either formally (as a property of trees), or semantically. Semantically, the thing denoted by NP *is a* Noun, where Noun is the head of NP; thus, *the tall boy* is a *boy*. The action denoted by VP "is an" instance of Verb-ing, where Verb is the head of VP. Thus, in the VP "slowly eat pies", the action described is necessarily an act of eating.

Everything within a phrase that is not the head can be termed a **modifier**, so long as we are willing to use the word "modifier" in a rather loose sense. This terminology may differ from what you learn in later linguistics courses, but it will be useful for our purposes.

Getting the concept of head and modifier right is, in my teaching experience, on of the trickier parts of learning syntax, so here are some examples.

(a) *tall women*

The head of this NP is the N *women* (tall women are instances of women). The word *tall* is a modifier, specifying what kind of women.

(b) *the book*

The head of this NP is the N *book*; when we say *the book* we are speaking of a book. The meaning of *the* is somewhat elusive, but essentially its purpose is to tell the listener that the speaker expects that she will be able to know (through overt presence, prior discourse, or reasoning) which book is being discussed—it says, "You know, somehow, which book I am talking about". The "opposite" of *the* is *a*, which signals that a book of which the listener is not necessarily aware is under discussion.

(c) *the man with the telescope*

The head of this NP is the noun *man*, and both the article *the* and the PP *with a telescope* are modifiers.

(d) *read the book*

The head of this VP is the V *read*; the VP describes an instance of reading, and *the book* is in some sense a modifier; it indicates what sort of reading-event took place by specifying one of the participants.

(d) *on Sepulveda*

The head of this PP (prepositional phrase) is the P(reposition) *on*. The meaning or function of the PP is to express location, and the word *on* serves to express this core meaning (*Sepulveda* has no inherent locative meaning; one can say "Sepulveda is a busy street", "They are repaving Sepulveda", and so on.)

(e) *very tall*

Looking ahead a bit, this is an Adjective Phrase, with an Adjective head *tall*, preceded by an Adverb modifier *very*.

6. Parsing sentences

The starting point for syntactic analysis of a language is to **parse** (provide a parse for; find the tree structure of) a variety of sentences. In the theory taught here, the basic principles of parsing are quite simple.

Principles of Parsing

- For the phrases NP, VP, PP, AP, locate the head, and include all its modifiers in the same phrase.
- Sentences (including sentences inside sentences; see Chapter 1) are assumed to consist of a **subject**, which is an NP; and a **predicate**, which is a VP.

Just as in traditional school grammar, the subject indicates what the sentence is about, and the predicate says something about the subject.

The hard part seems to be to make sure you find *all* the modifiers of each head, and include them in the phrase of which it is the head; so exercise care here.

Returning to the two structures of *Sue saw the man with the telescope*, the crucial distinction is what the PP *with a telescope* is a modifier of: in one reading, it modifies *man* (that is, it specifies which man), and thus belongs as part of NP; in the other reading, it modifies *see* (that is, it specifies what kind of act of seeing took place), and thus belongs as part of VP.

Study Exercise #9

Diagram (that is, parse) both readings of the sentence *Bill gave the Chinese dishes*.

This area is intensionally blank.

The example illustrates the point that differing parses of the same string are only one source of ambiguity in language. To mention some others in passing:

[26] How do we get *Chinese* as a Noun? The analysis I assume here uses the morphology, with a rule of morphological conversion:

$[\ X\ ]_{Adj} \rightarrow [[\ X\ ]_{Adj}]_{Noun}$ meaning: "person who is X"

A curious property of the rule is that the output can only be used in the plural; hence it should also attach the inflectional feature [Number:plural] to the morphosyntactic representation of its output.

- Homophonous words: *We walked to the bank*.
- Multiple quantifier words: *Three men were examined by each doctor.*
- Phonological merger: *We were patting/We were padding*, which are identical in many North American English dialects

Study Exercise #10

Here is one more ambiguity with its two parses (check that you know the answer before you look). The sentence is: *The hungry bear fishes*

This area is intensionally blank.

7. Possessive constructions

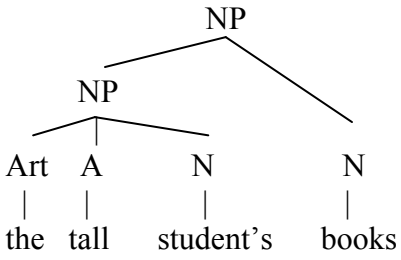
People are taught in school that adjectives are words that modify nouns. I think this is basically true; provided that you don't say they are the *only* words that modify nouns; there are quite a few other possibilities.

One very common noun modifier is the possessive construction, as in *the tall student's books*. *The tall student's* modifies *books*, but in its internal structure it looks just like an NP. (except for the extra material *'s*). It couldn't possibly be an Adjective; an Adjective is a word, but *The tall student's* is a whole phrase.

[27] You may be wondering why we bother with a VP symbol when there is no modifier present; see below on phrase structure rules for some justification.

[28] As a student pointed out to me, there's yet a third parse: "the hungry bear-fishes", an NP modeled on *catfishes*. This involves *bearfishes* as a compound word, discussed in the earlier morphology readings. Multiple parses lurk everywhere.

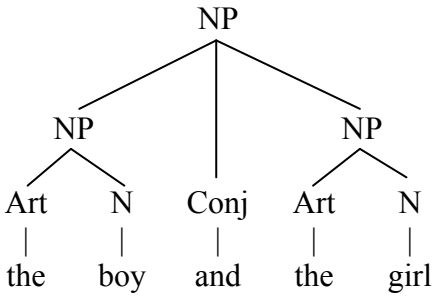
We will assume here that *the tall student's* is in fact an NP, and it sits inside the larger NP *the tall student's books*, modifying the head *books* (i.e., it says in effect, "whose books?"). Thus the structure is:



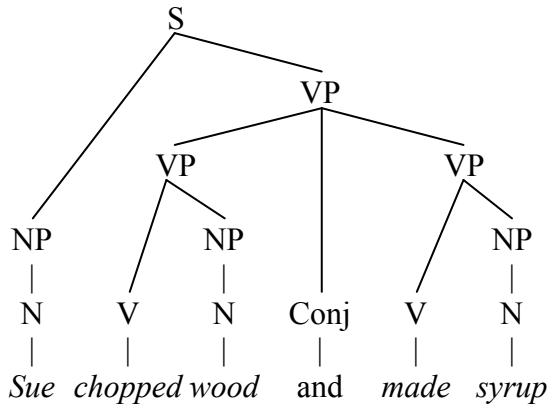
There's a debt to pay here: where does the *'s* morpheme come from, and where should it sit in the tree? We'll cover this next time. The brief answer is that the *'s* will be treated as inflectional morphology. What we need is a way to relate the inflectional morphology to the syntax, which is what we will take up later on. For now, we'll just take the -'*s* as something we need to handle eventually, but will ignore for now.

8. Conjoined structures

Conjunctions like *and* and *or* are fairly straightforward: we'll assume that they link together two identical units, forming a large unit of the same kind. Thus *the boy and the girl* is



We say that the two NP's *the boy* and *the girl* are **conjoined** by *and* into a larger NP, the entire structure. Similarly, *Sue chopped wood and made syrup* has a VP made of two conjoined VP's:



Several other categories, including Adjective Phrases, PP, and S, can participate in this construction: examples of these (same order) are *very tall but quite thin*; *over the river and through the woods*; *I like coffee and you like tea*.

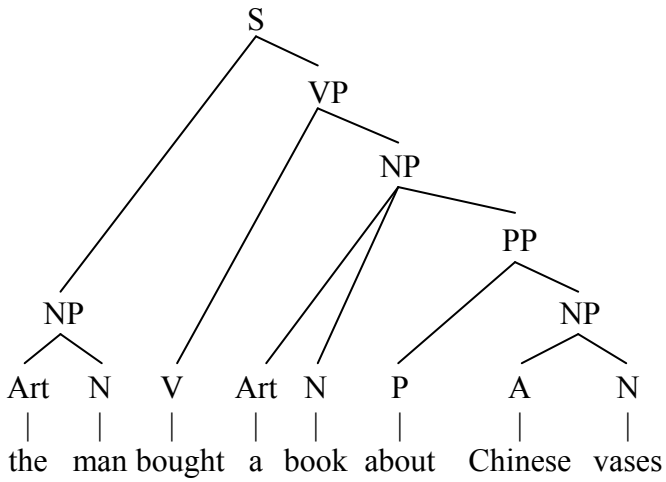
Study Exercise #11

Parse *the king and the queen's throne*.

This area is intensionally blank.

9. Terminology for trees

Here is some terminology that will be useful in referring to trees. I will use the following tree to illustrate the various terms:



Dominate: Node X dominates node Y if you can get from X to Y by going "downhill" in the tree and never uphill. For example, S dominates everything in the tree; the NP on the right dominates an A, an N, and the words *Chinese* and *vases*. The NP on the right does not dominate the VP, nor does it dominate the word *man*.

Constituent: We can redefine this formally as a node, plus all the nodes that it dominates. As noted above, one usually refers to constituents by the words they contain. Thus one can say that the following:

the man a book about Chinese vases bought bought a book about Chinese vases

are constituents (in this particular sentence). Note that sequences like

bought a book
the man bought a book
a book

are not constituents in this particular sentence, *though they could be in other sentences*.[29]

directly dominate: X directly dominates Y if Y is "one node down the tree" from X. Thus the NP *a book about Chinese vases* directly dominates the PP *about Chinese vases*.

daughter: if X directly dominates Y, then Y is X's daughter

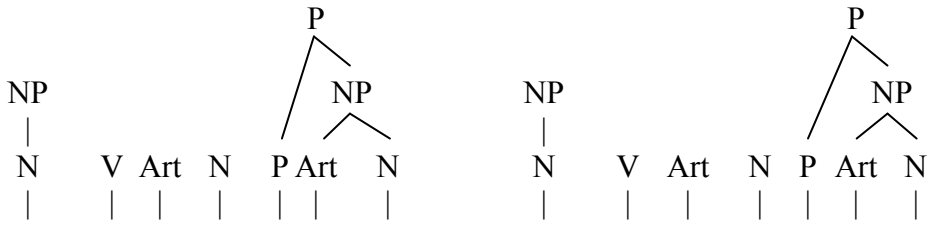
sister: two daughters of the same node are sisters.

head: We've defined this casually, but can now give the tree-based formal version: the head of an NP is the N that it directly dominates. The head of a VP is the V that it directly dominates. For example, the head of the NP *a book about Chinese vases* is *book*. The head of the VP *bought a book about Chinese vases* is *bought*. And similarly, the head of the AP *very tall* is the Adjective *tall*. (One could extend this definition to PP as well, though it won't matter for us.)

10. More on parsing

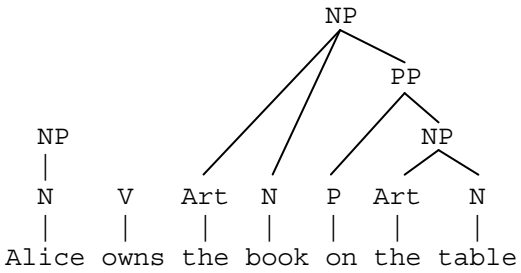
Coming up with the phrase structure tree for a sentence is a skill that is crucial for studying syntax. There are two useful principles to keep in mind. First, until you've had some practice it's easier to *work from the bottom up*: label each word with its part of speech, form small phrases out of the single words, group the small phrases into bigger phrases, and so on up the tree. For example, if you were diagramming the two sentences below, you might start like this:

[29] As mentioned above, this is one of the principal difficulties in parsing; that is, not to get distracted by mere "potential" constituents like these, and instead choose *complete* constituents.

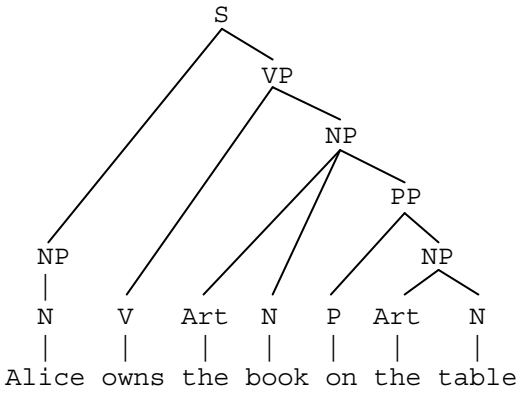


A. Alice owns the book on the table B. Alice placed the book on the table

Second, think consistently about heads, and about grouping modifiers into the same phrase as their heads. For example, in diagramming sentence A above, the crucial question is what *on the table* belongs to. If you think about the meaning of the sentence, it is clear that *on* the table *modifies* book*;* that is, it specifies *which book is* under discussion. The rest of the reasoning goes like this: 'book' is a noun; it must be the head of a NP; anything that modifies it (namely 'the' and 'on the table' must be its sister; therefore the full NP is 'the book on the table'.

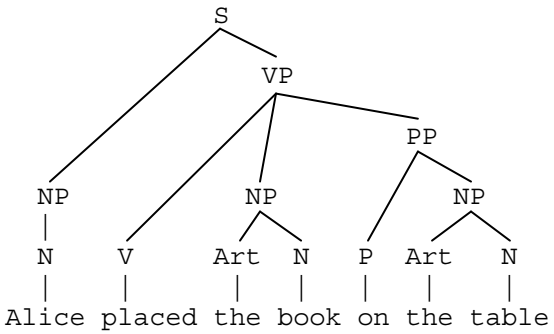


From there on, the diagramming is straightforward; you just need a VP (verb and object) plus the whole sentence:

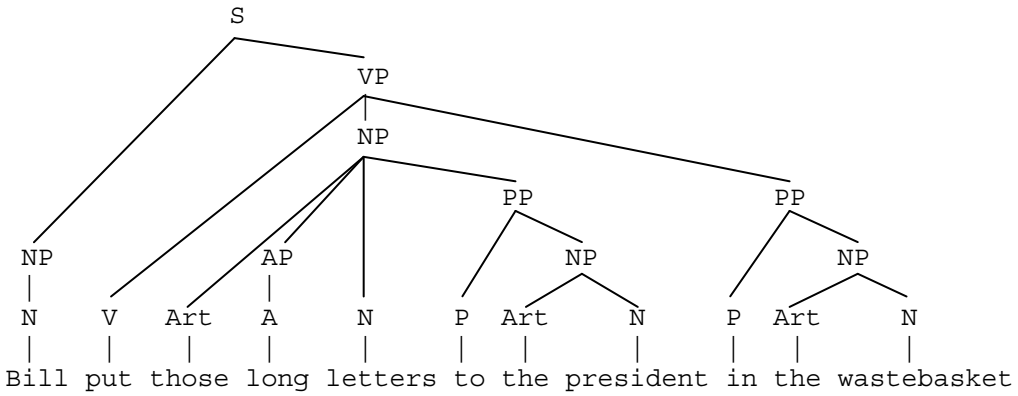


Note that *the book* is not an NP; it is only *part* of an NP because the head is missing one of its modifiers.

Suppose this time that you are diagramming sentence B above, *Alice placed the book on the table*. In this case, the PP *on the table* modifies the verb *placed* (it indicates the target of placing). Accordingly it must be the sister of the verb within the VP. *The book* is left as an NP on its own.



The following sentence has two PP's, which get placed in different positions according to what they modify:



A fundamental principle for diagramming sentences, worth memorizing perhaps, is:

The modifiers of a head must be sisters to the head.

You can see this principle in effect in all of the parsing done so far.

I have one other handy hint in parsing. It's often easiest to parse English sentences going backwards, starting with the end of a sentence.[30]

11. Constituency tests

I'm a bit uncomfortable with the discussion so far because it consists simply of directions to you, the student, on how to diagram sentences. This is merely being tyrannical unless it can be shown that the structures we're creating have some scientific purpose and validity. As at least a move in this direction, we can note the following evidence.

The following are examples of what linguists often call "cleft sentences:"

[30] Why so? If you're curious: it has to do with a property of English called "right-branchingness". When a constituent has two daughters, rather often the daughter on the left is a single word, whereas the daughter on the right has some internal structure. When you have a right-branching system, right-to-left implies bottom-up. Japanese, which is mostly left-branching, is probably easier to parse left-to-right.

It's Bill that they don't like.
It's on Mulberry street that they live.
It was the flowers that Mary sent to Bill.

Such sentences are clearly related to simpler sentences, such as

They don't like Bill.
They live on Mulberry Street.
Mary sent the flowers to Bill.

We can express the relation between simple sentences and cleft sentences by writing a syntactic rule (we'll cover this more formally later on):

Clefting Rule

To form a cleft sentence, take a simple sentence and perform the following operations on it:

1. Add *it* and an appropriate form of the verb *be* to the beginning.
2. Find a NP or PP constituent inside the sentence and reorder it so that it directly follows *be*.
3. Insert the word *that* just after the reordered NP or PP.

You can see for yourself that the cleft sentences cited above are derived from the corresponding simple sentences.

The crucial part of the rule is where it says "find an NP or PP *constituent*". It predicts that if we apply Clefting to a sequence of words that is not constituent, the result should be ungrammatical. If you look at the tree drawn earlier for 'Alice placed the book on the table', you will see that 'the book on the table' is not a constituent. The rule thus correctly predicts that if we attempt to do Clefting with this sequences of words, the result will be ungrammatical:

*It was the book on the table that Alice placed.

On the other hand, in 'Alice owns the book on the table', the sequence 'the book on the table' is a constituent, so that Clefting produces a grammatical result:

It's the book on the table that Alice owns.

In summary: neither the principle that modifiers form constituents with their heads, nor the rule of Clefting can be assumed in advance to be correct. We can only test them out against the facts. The more correct predictions they make, the greater is our confidence that they are true. If we want to be really confident about these principles, we must test them out against a much larger set of facts. We will carry out part of this task later on.

Study Exercise #12

(a) In 'They sent the king to Barataria' is 'the king to Barataria' a constituent? Support your answer with evidence from Clefting.
(b) Substitute *to* for *of* in the same sentence and answer the same question.
(c) What are the grammatical clefted versions of 'Alice put the book on the table'? (There are about four).

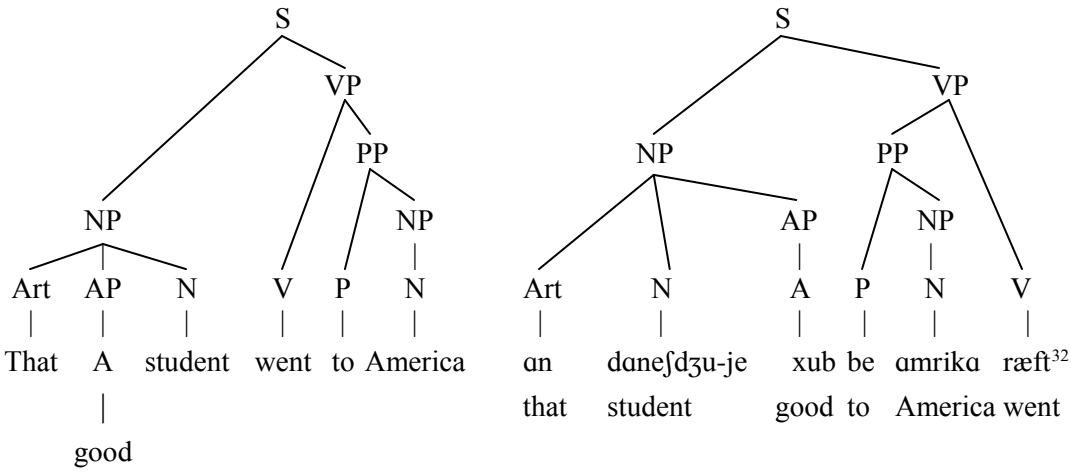
This area is intensionally blank.

12. Grammars for syntactic structure I: Phrase structure rules

The discussion so far has been about structures; now we can turn to the grammars that are responsible for these structures.

The need for such grammars should be clear: although nodes like NP and S seem to occur in all languages,[31] the actual order of the constituents of a phrase varies from language to language. For example, consider the following English sentence and its literal translation into Persian (=Farsi):

[31] It's less clear that there is a Verb Phrase in languages where the subject comes between the verb and the object (for example, Verb-Subject-Object order, as in Malagasy).



Because of this, every language must include rules that specify its grammatical word orders.

The rules that specify word order are called *phrase structure rules*. Some examples of phrase structure rules are as follows:

```
English:  NP  →  Art A N          Persian:  NP  →  Art N A
          VP  →  V PP                     VP  →  PP V
```

You can read the rules as follows: "an NP may consist of the sequence *Art, A, N*.

There is a more interesting way of interpreting phrase structure rules. If we have a complete set of them for a given language, we can think of the set of rules as an abstract machine that *generates* syntactic structures. For example, assume for the moment the following (obviously incomplete) set of phrase structure rules for English:

S → NP Aux VP (Aux = "helping verb", like *can*, *will*, *be*, etc.)
NP → Art A N
VP → V NP
AP → A[33]

In this respect, the phrase structure rules are like the rules of inflectional morphology given earlier: given a starting point, they *generate* a sentence. For inflectional morphology, the starting point is the stem with its morphosyntactic representation. For syntax, the starting point is a single symbol, such as NP or (most often) S, which designates the category that we wish to generate.

Here is the procedure.

[32] IPA symbols: ɑ = somewhat like *ah*; ʃ = *sh*, dʒ = *j*, x as in *ch* of German *Bach*, æ = somewhat like the vowel of *cat*.

[33] This rule looks trivial right now—we'll beef it up a bit later by allowing Adverbs.

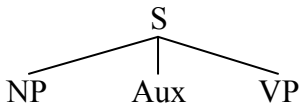
(a) Provide the rules with the symbol S (or NP, or whatever) to start out with;
(b) Whenever a symbol appears in a tree that is found on the left side of the rule, give that symbol daughters according to what the rule says;
(c) Do this over and over until you can't apply any more rules.

If we start with S, the stages of this process would be as follows:

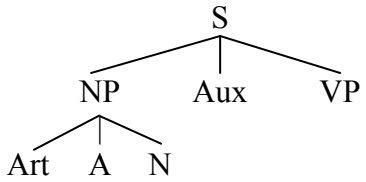
1. Starting point:

S

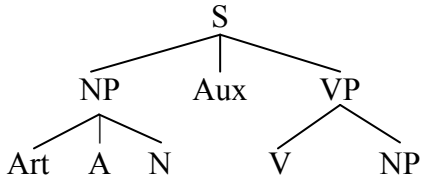
2: apply S → NP Aux VP



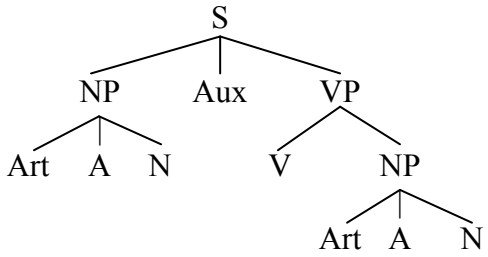
3: apply NP → Art A N



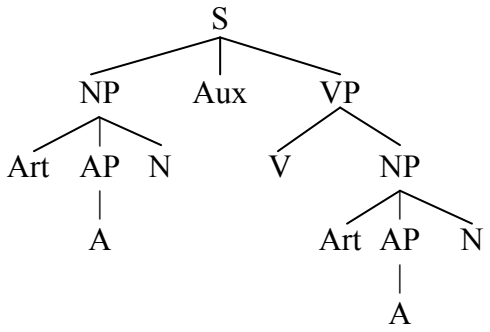
4: Apply VP→ V NP



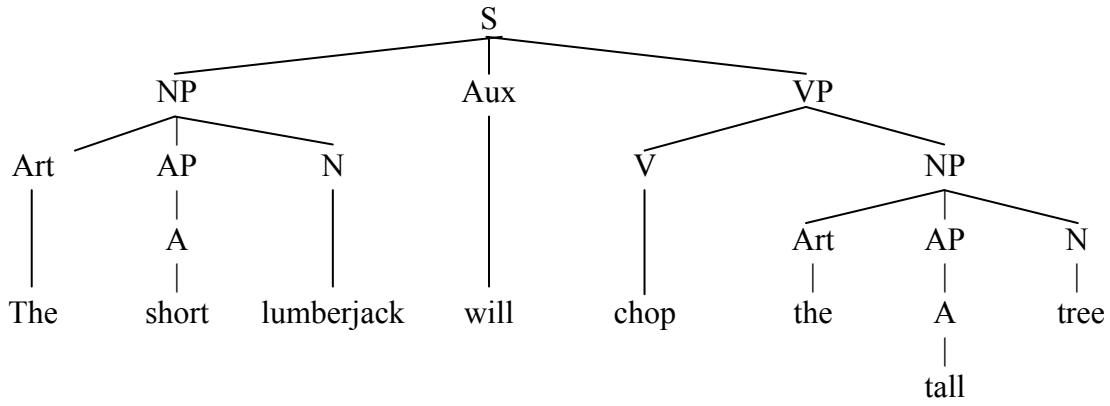
4: Apply NP→ Art A N



5: Apply NP→ Art AP N



All that remains is to insert actual words into the tree (a process called **lexical insertion**), and you get sentences:



or, with different choices for lexical insertion:

The sleepy student might ignore the noisy alarm.
The green idea will paint the blue intellect.

Note that these sentences will not necessarily be sensible.

The phrase structure rules just proposed are obviously primitive. We can improve them by observing that some of the daughters introduced by a rule are optional. In particular, the NP rule has to introduce a N, but it doesn't have to introduce an Art or an A. The standard notation in linguistics for expressing optional elements is parentheses:

S → NP (Aux) VP
NP → (Art) (A) N
VP → V (NP)

These more flexible rules can provide the syntactic structures of sentences like

The lumberjack will chop the tree

Frogs will eat flies
Fish can see
Those students read books.
Sue won

and so on. (Diagram these if it is not obvious what the structure is.)

We can also make our AP rule less trivial, so that Adverbs are allowed.

$$AP \rightarrow (Adv)\ A$$

For instance: *very tall*.

13. Curly brackets for "or"

One other complication in the notation for phrase structure rules. We find that a NP can begin *either* with an Article *or* with a possessive NP, but not both.

Article:

the book, a book, this book, those books

NP:

Fred's book, the king of England's book, my book

not both:

**the Fred's book, *the king of England's this book, *those my books*

Here is a simple way to account for this: we use curly brackets in the rules to mean "one or the other, but not both" (logicians call this "exclusive or"). The basic NP phrase structure rule for English comes out something like this:

$$NP \rightarrow \left(\begin{Bmatrix} Art \\ NP \end{Bmatrix}\right)(AP)\ N\ (PP)$$

This means that you can start out an NP with an Article, or an NP, then continue with the rest (optional Adjective, obligatory Noun, optional PP). Examples of each type:

the long book about linguistics	(beginning with Article)
the king's long book about linguistics	(beginning with NP)

14. Phrase structure rules for English sentences given so far

As we continue through syntax, we will gradually build up an ever-improving grammar of phrase structure rules. Just to catch us up so far, I believe the following set of rules can generate most of the examples given in this book so far, as well as the sentences in the Study Exercises.

S → NP (Aux) VP
NP → $\left(\begin{Bmatrix}\text{Art}\\ \text{NP}\end{Bmatrix}\right)$(AP) N (PP)
NP → Pronoun
VP → V (NP)(NP)(PP)
PP → P NP
AP → (Adv) A

S → S Conj S
NP → NP Conj NP
VP → VP Conj VP
PP → PP Conj PP
AP → AP Conj AP
V → V Conj V

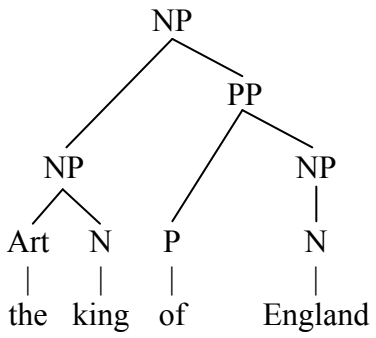
15. Parsing: using the phrase structure rules as a guide

Once you've got a grammar like this to work with, then in principle it becomes easier to diagram sentences—any particular set of rules represents a claim about the inventory of phrase types a language allows, and thus constrains what kind of structures you can set up. Thus:

- When diagramming sentences, make sure every structure you set up is licensed by the rules.

In other words, you can't set up an NP whose structure is N AP, unless there is a phrase structure rule that specifies this sequence (either directly, or by leaving out parentheses). Thus you can be guided to an answer by both the meaning of the sentence and by the rules of the grammar.

Example: if you're thinking of the structure below for *the king of England*



you can tell it's not right because the grammar above contains no rule that permits NP to dominate NP followed by PP.[34]

There actually is one way you can legitimately diagram a structure that the grammar doesn't allow—namely, *change the grammar*. In other words, you have to say something like "This sentence shows that our grammar was wrong, and has to be fixed like this [offer substitute rules here]." In this book I have included only sentences that can be parsed with the grammar given so far (unless I've made a mistake). But of course real life is different: a grammar that could parse all of English would be quite large and a real challenge to create.

16. Further details of our current grammar

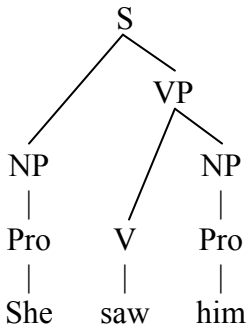
A few of our phrase structure rules need clarification.

16.1 Pronouns

The phrase structure rule above that introduces Pronouns is very simple:

NP → Pronoun

Pronoun, appearing in trees, is often abbreviated as Pro. Thus:



The reason to have a separate rule for pronouns is that, unlike nouns, they do not admit modifiers, except in special circumstances we'll defer for now.[35] This is one reason to give them their own phrase structure rule, rather than just calling them a kind of Noun. The other reason is that, later on, we will need rules of semantic interpretation that indicate what the pronouns refer to, and these rules need to identify the pronouns.

Incidentally, pronouns in English are unusual in that they are inflected for case. English has a three-way case system, with Nominative, Objective, and Genitive. Objective covers what in many other languages (including English, long ago) was Accusative or Dative. Different authors will give different names to these cases.

---

[34] Note that there is a rule NP → $\left(\begin{Bmatrix}\text{Art}\\ \text{NP}\end{Bmatrix}\right)$(AP) N (PP). But it won't help, because it *requires* there to be an N daughter.

[35] Examples: *Poor me*, a "frozen" memorized expression; *He who dares to go…*, with a relative clause.

		Nominative	Objective	Genitive
Singular	**1**	*I*	*me*	*my*
	2	*you*	*you*	*your*
	3	*he/she/it*	*him/her/it*	*his/her/its*
Plural	**1**	*we*	*us*	*our*
	2	*you*	*you*	*your*
	3	*they*	*them*	*their*

Part of what a grammar must do is ensure that the correct case form of each pronoun is used in the right context; we will turn to the sort of rules that are needed later on.

16.2 Aux

"Aux", meaning "auxiliary verb", is the "helping verb" taught in school. In our phrase structure rules, it is the optional second daughter of S (S → NP (Aux) VP). Here is a list of auxes:

"Modal" verbs: *can*, *could*, *shall*, *should*, *may*, *might*, *will*, *would*
Example: *I can go*.

Forms of *have*: *have*, *has*, *had*
Example: *I have gone*.

Forms of *be*: *be*, *am, are*, *is*, *was*, *were*
Example: *I am going.*

You can see that the choice of Aux also determines the inflectional morphology of the following verb—this involves rules we haven't yet covered.

Be aware that *have* and *be* can serve as either Auxes or main Verbs. Thus:

He is having a fit

involves *be* as an Aux and *have* as a main Verb.

He has been President.

has *have* as an Aux and *be* as a main Verb.

16.3 Complementizers and subordinate clauses

Much of the most intricate syntax arises when one "puts a sentence inside a sentence"; that is, when one uses a **subordinate clause**. This showed up early in the course when we looked at the patterning of *each other*. Thus, *[John and Bill think [I like each other.]$_S$]$_S$ is impossible, because *each other* is allowed to refer only to Noun Phrases that are within the smallest clause containing it—in this case, [I like each other]$_S$. Subordinate clauses often occur when the verb

of the main clause is a *verb of saying or belief*—the subordinate clauses serves to express the content of the thought that is said or believed. With the notions of syntax we've developed so far, we can now be much more explicit about subordinate clauses.

To analyze subordinate clauses, we need to provide a slot in phrase structure for the grammatical words that often introduce them—*that* in sentences like:

I think *that* [John and Bill like each other]$_S$

There is also *for*, as in:

I would prefer *for* [John and Bill to like each other]$_S$.

Such words are called **subordinating conjunctions** in traditional terminology. Linguists use the slightly shorter term **complementizer**, abbreviated **Comp**.

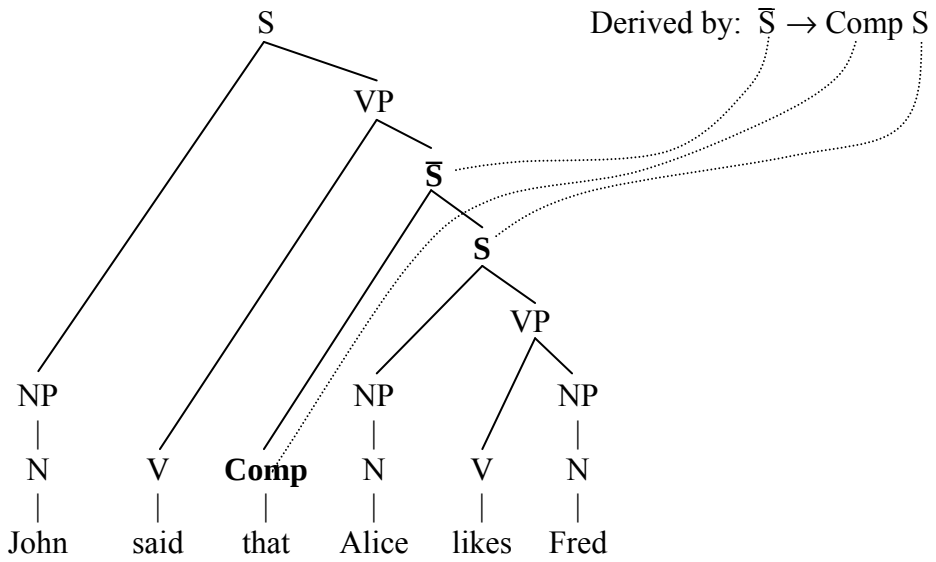
Other complementizers include *if*, *(al)though, when, whether*, and some others we'll mention later.

With this apparatus, we can set up rules like these (I'm omitting optional material; see below for the full rules):

VP → V $\overline{S}$

$\overline{S}$ → Comp S

$\overline{S}$ is read **S-bar**, and is simply the category that provides the syntactic "slot" for the complementizer. Here is an example sentence that can be generated by these rules:



Since $\overline{S}$ is a nuisance to type on a word processor; a prime (S′) or apostrophe (S') is considered an acceptable substitute.

16.4 Phrase structure rules for subordinate clauses

Subordinate clauses in English most often occur the last constituent of the VP, indicating what was said or thought. Here are some examples:

We said [that we were going]$_{\overline{S}}$
We told Alice [that we were going]$_{\overline{S}}$
We gave Bill notice [that we were going]$_{\overline{S}}$
We sent word to Jane [that we were going]$_{\overline{S}}$

From these sentences, you can see that the Verb Phrase can, in addition to its subordinate clause, include one or two NP objects and a PP, all of them preceding the $\overline{S}$. Thus the phrase structure rule needed is something like:

$VP \rightarrow V\ (NP)(NP)(PP)(\overline{S})$

Study Exercise #13

Parse the four sentences given above.

This page is intensionally blank.

This area is intensionally blank.

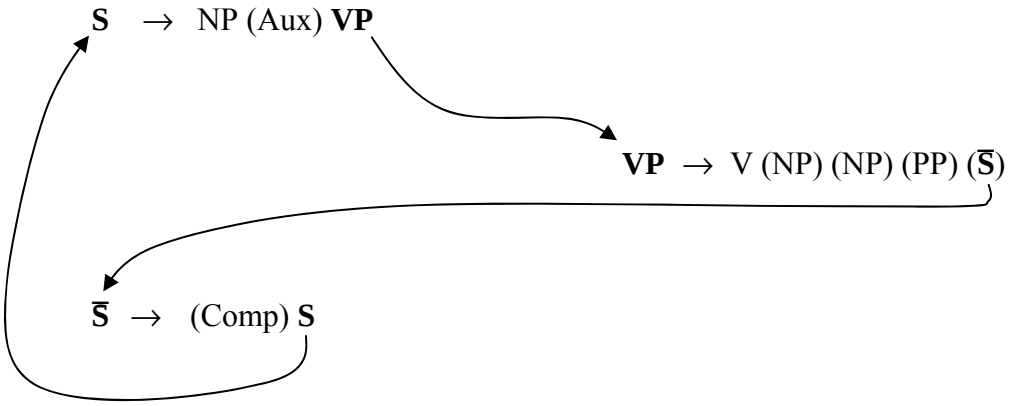
17. Recursive application of phrase structure rules

I mentioned above that the speaker's knowledge of syntax is large but finite (that is, it fits somehow encoded in a single brain), yet permits the creation of an infinite number of sentences (p. 53). The following partial list was meant as a demonstration:

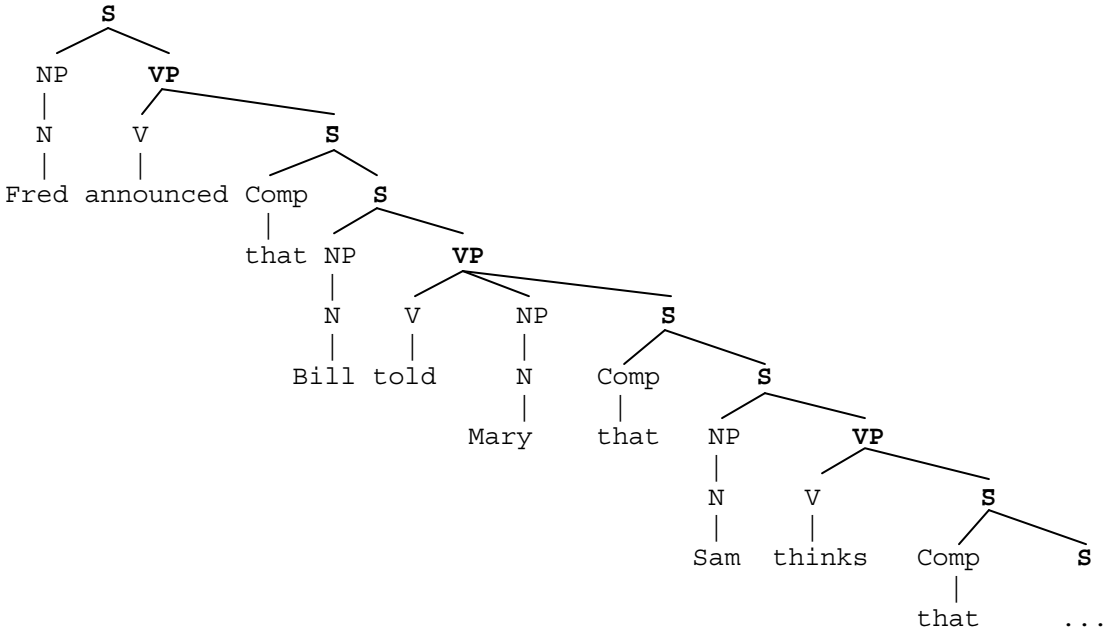
Alice likes Fred
John said that Alice likes Fred
Bill believes that John said that Alice likes Fred
…etc.

This infinity results, by and large, from a particular property of phrase structure rules, namely that they permit application in **loops**.

Below, I demonstrate one of these loops, taken from the phrase structure rules given so far. With this more interesting grammar, we arrive at a crucial property: the rules in our grammar can apply *recursively*. That is, by following the appropriate procedure it is possible to make the same rule apply over and over, "recurring" in the derivation. The procedure requires finding a repeating loop, such as the following:



If we employ this loop in deriving a sentence and lexically insert appropriate words, we can generate a sentence as long as we like:



This is because there is an infinite number of places where we could stop the loop. Thus there are an infinite number of possible sentences that the grammar can generate.

As far as we know, *every* human language allows an infinite number of sentences. In every case, the principal reason is the same: the phrase structure rules of all languages contain recursive loops, which allow infinitely long syntactic trees to be generated. The recursive loop of phrase structure rules is the device that allows a finite number of rules to generate an infinite number of structures.

18. Relating syntax to inflectional morphology

We are now in a position to tie together our two course units so far (morphology and syntax). The crucial notion is the morphosyntactic representation, covered earlier. You can think of the morphosyntactic representation as the means by which the syntax communicates with the inflectional morphology.

The features in a morphosyntactic representation can have three sources.

18.1 Inherent features

First, some features of a morphosyntactic representation are **inherent**. They are properties of particular words or stems. It is conventional to use the word **lexicon** to refer to the speaker's mental dictionary; their store of memorized stems, words, and other entities.[36] Since a feature like [Gender] on nouns is memorized, it must be listed in the lexicon.

Here are three examples of inherent inflectional features.

I. The German word *Messer* (knife) is inherently, and arbitrarily, neuter. Its lexical entry must look something like this:

Messer [Gender:Neuter]

That is, attached to *Messer* is a partial morphosyntactic representation that indicates that *Messer* is a neuter noun.

II. The English pronoun *his* is inherently [Case:Genitive,Gender:Masculine].

his [Case:Genitive,Gender:Masculine]

III. All nouns derived by the English word formation rule $[\ X\ ]_{Adj} \rightarrow [[\ X\ ]_{Adj}]_{Noun}$ (example: *The <u>French</u> care a lot about food*) are inherently [Number:Plural].[37] This is also true for a small number of words for "pairlike" things, such as *trousers*, *scissors*, and so on.

[36] We also memorize a great number of word sequences, often called **idioms**.

[37] Thus, a fully explicit version of the conversion rule would actually attach a partial morphosyntactic representation: $[\ X\ ]_{Adj} \rightarrow [[\ X\ ]_{Adj}\ ]_{Noun,[Number:Plural]}$.

18.2 Speaker-selected features

Other features of the morphosyntactic representation are meaningful; they represent **choices made by the speaker**. When we say *book* in English we are implicitly conveying the partial morphosyntactic representation [Number:singular], and similarly [Number:plural] when we say *books*. (This raises the question of how linguistic entities bear meaning, a question we will postpone to Chapter 9.)

18.3 Features derived by syntactic rules

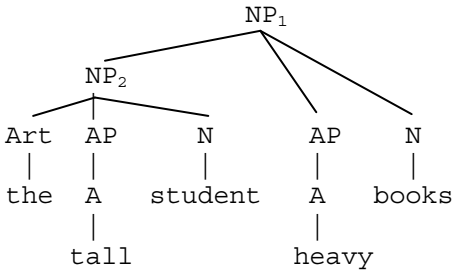
The remaining source for the features in morphosyntactic is **syntactic rules**. These attach the features that depend on what else occurs in the tree. There are (at least) two kinds of rule of this sort: rules of **case marking** and rules of **agreement**.

19. Case marking

19.1 Genitive case in English

Genitive case in English is the case that we spell with the suffix -'*s*. Semantically, it denotes the relationship of possession. To derive it, we need a syntactic case marking rule, and a morphological suffixation rule.

Here is a tree to serve as an example. The phrase structure rules given so far generate this:

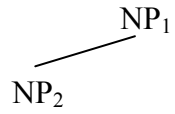


Choices employed: for NP_1: $NP \rightarrow (\left\{\begin{matrix}\underline{\text{Art}}\\ \underline{\text{NP}}\end{matrix}\right\})(\underline{\text{AP}})\ \underline{\text{N}}\ (\text{PP})\ (\bar{\text{S}})$

for NP2: $NP \rightarrow (\left\{\begin{matrix}\underline{\text{Art}}\\ \underline{\text{NP}}\end{matrix}\right\})(\underline{\text{A}})\ \underline{\text{N}}\ (\text{PP})\ (\bar{\text{S}})$

The syntactic rule of case marking that is needed is as follows:

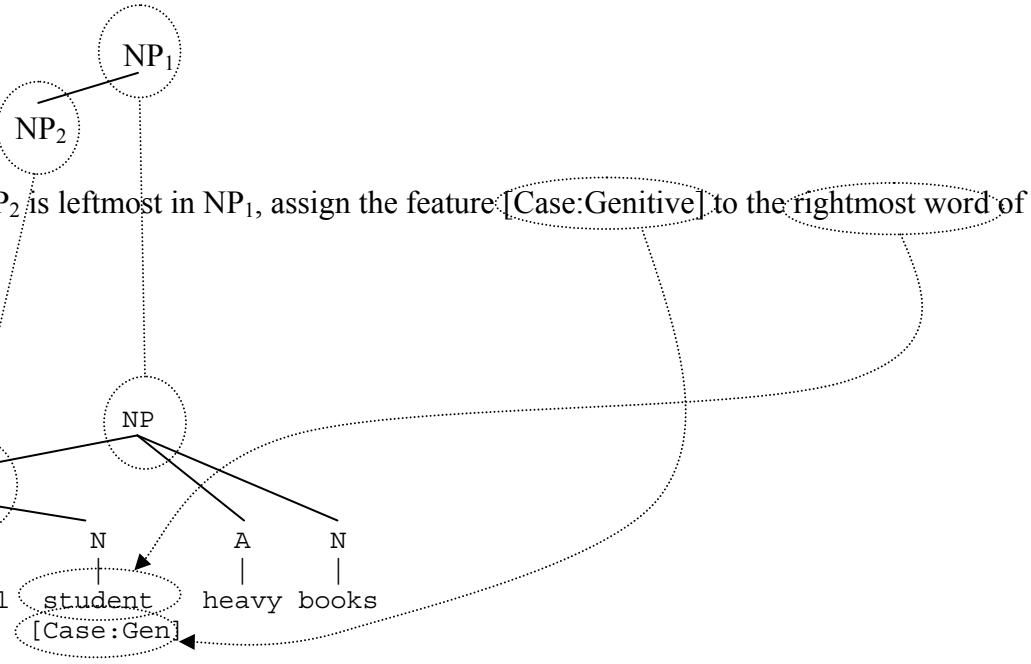
Genitive Case Marking

In 

where NP_2 is leftmost in NP_1, assign the feature [Case:Genitive] to the rightmost word of NP_2.

Genitive Case Marking can be applied to the above as shown. I use dotted lines to show what part of the rule matches up to what part of the form

Genitive Case Marking

In 

where [Case:Genitive] is the morphosyntactic representation of *student*.

That is the most complicated part. Once we have the feature [Case:Genitive] sitting on the word *student*, it is straightforward to get the suffix in place, with an ordinary rule of inflectional suffixation, as follows:

Genitive Inflection

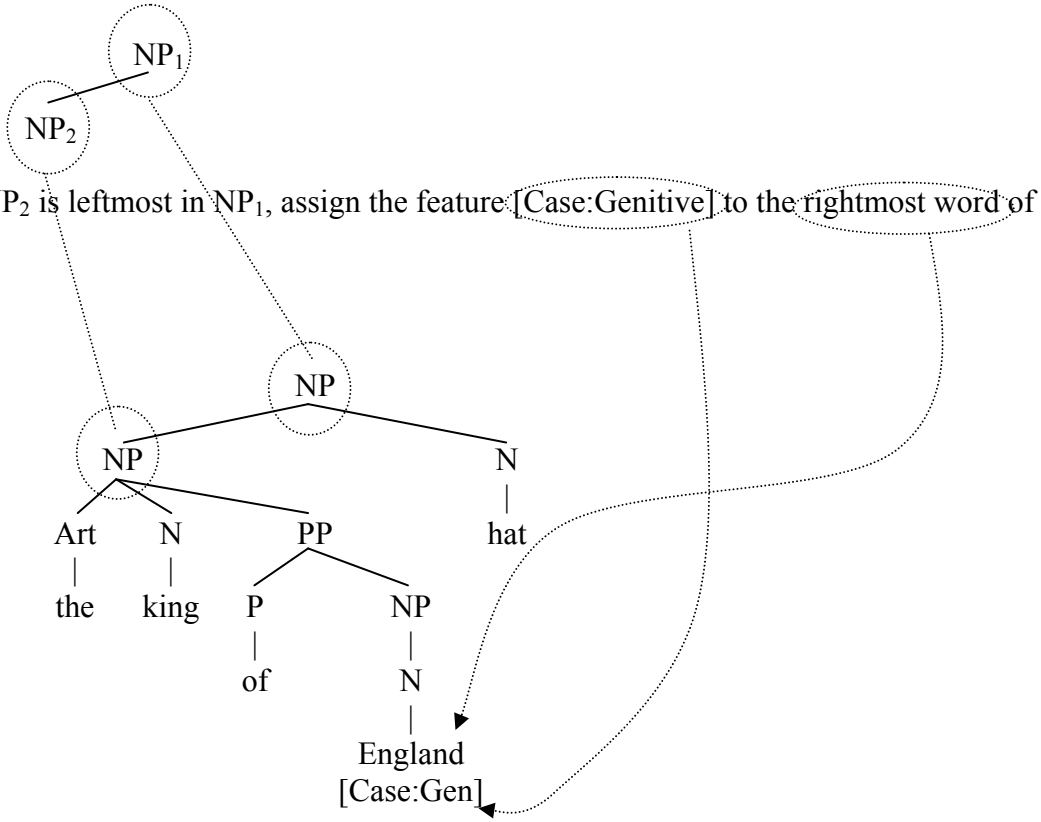
Suffix -s if [Case:Genitive].

Thus the full NP *the tall student's* is the combined result of syntactic and morphological rules.

19.2 Where to inflect? Edges vs. heads

The rule of Genitive Case Marking in English perhaps unusual for putting the relevant feature on the *rightmost word* of NP. We need this for cases like [*the king of England's*]$_{NP}$ *hat*, where *England* is the rightmost word of its NP. The matchup is shown below:

Genitive Case Marking

In 

where NP_2 is leftmost in NP_1, assign the feature [Case:Genitive] to the rightmost word of NP_2.

The other major form of case marking targets the *head* of the NP that is to bear case. Let us consider an example from German. On German Amazon I found an entry for a book with this title:

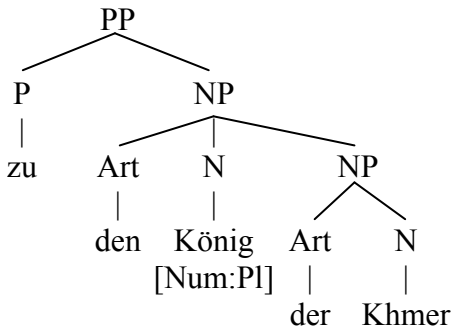
Schliemanns	Erben,	Von	den	Herrschern	der	Hethiter
Schliemann's	legacy	from	the	ruler-dative-plural	the	Hittite-genitive plural

zu	den	Königen	der	Khmer
to	the	king-**dative plural**	the	Khmer-genitive plural

'Schliemann's legacy: from the rulers of the Hittites to the kings of the Khmers'[38]

We're interested in *zu den König-**en** der Khmer*, meaning 'to the kings of the Khmers'.[39] Prior to case marking, the structure looks like this (for this particular construction, the relevant phrase structure rules of German are the same as in English):

[38] Hermann Schliemann was the archaeologist who excavated the ruins of Troy.

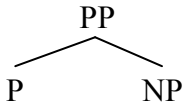


[Number:Plural] is already attached to *König* 'king'; this reflects a semantic choice made by the person who made up this title.

A crucial fact about German is that the various prepositions take (more formally: **govern**) different cases. The preposition *zu*, pronounced [tsu] and meaning 'to', is one of the prepositions that governs the dative case. A partial dative-case marking rule for German can be written as follows:

German Dative Case Marking

In the configuration shown:

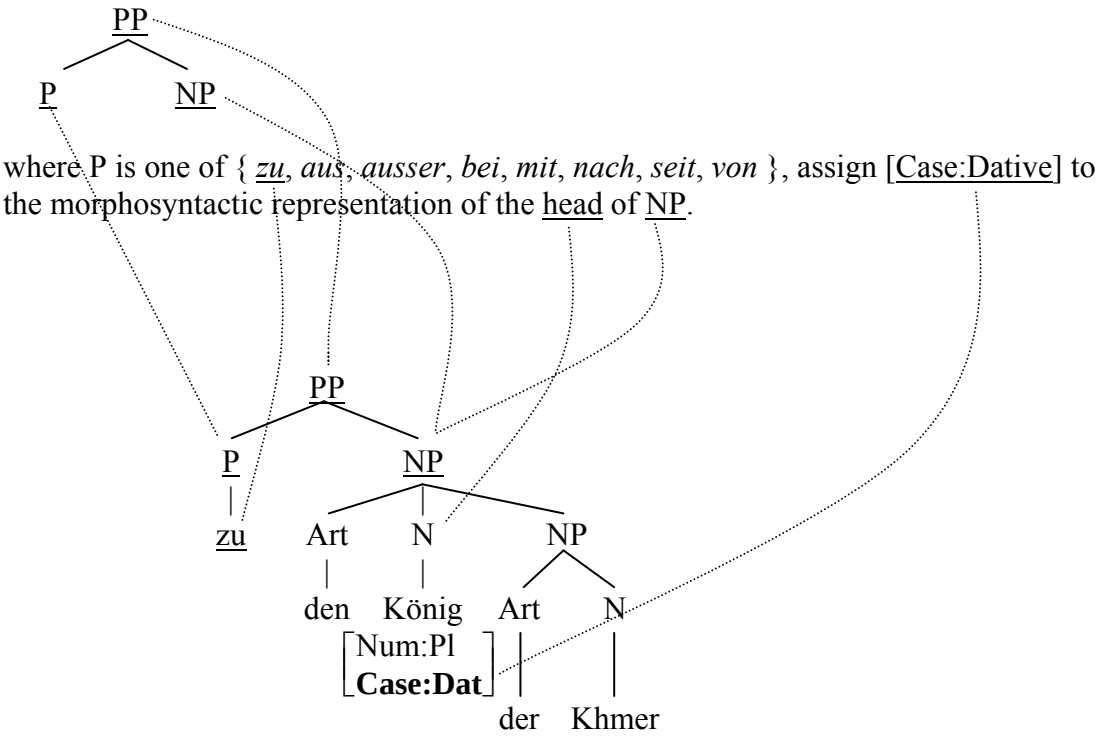


where P is one of { *zu*, *aus*, *ausser*, *bei*, *mit*, *nach*, *seit*, *von* },[40] assign [Case:Dative] to the morphosyntactic representation of the head of NP.

This rule targets the *head* of NP for dative case realization, hence applies to our example as follows:

[39] The Khmers are the Cambodians.

[40] 'to', 'from', 'except', 'at X's home', 'with', 'after', 'since', 'of'



The dative plural is then realized in the morphology with suffixation:

Inflection morphology: Dative Plural Realization (German)

Suffix *-en* if morphosyntactic representation contains $\begin{bmatrix}\text{Num:Pl}\\ \text{Case:Dat}\end{bmatrix}$,

This will derive the boldfaced material in *zu den König**en** der Khmer*.

There are further details about German we'll pass over here quickly. Case is generally also realized, through additional agreement rules (see below), on the Article beginning a Noun Phrase. Thus, *den* is in fact the dative plural form of the definite article.

The crucial distinction illustrated here is the edge-based case marking of the English genitive vs. the head-based marking of German datives. If each language used the opposite language's strategy, we'd get very different results: ***the king**'**s** of England hat* (marking of genitive on the head), and ** zu den König der Khmer**en*** (marking of dative on the rightmost word).

There are other differences between edge-based and head-based case marking. Marking on heads tends to get complicated, with different affixes for different nouns and so on; marking on edges tends to be a simple, single morpheme like English -'*s*. Marking on heads probably is more often accompanied by agreement on modifying adjectives and articles.

20. Agreement

Features also get assigned in syntax when one phrase **agrees** with another. For instance, in English we have a very simple agreement paradigm in verbs.

I jump	*we jump*
you jump	*you jump*
*he/she/it jump**s***	*they jump*

There is only one ending, *-s*, which occurs when the subject is [Person:3, Number: singular]. Note, however, that for the special verb *be* there is a richer system, with difference between all three persons in the singular and a separate form for the plural:

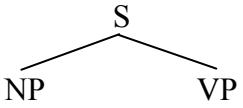
I am	*we are*
you are	*you are*
he/she/it is	*they are*

The point at hand is that agreement with the subject is inherently syntactic; the verb needs to "know," as it were, what the subject is in order to bear the right inflectional features.

Again, our strategy is to write a syntactic rule that assigns the features of the morphosyntactic representation, then a rule of inflectional morphology to add the appropriate affix.

The syntactic rule can be written provisionally as follows:

Verbal Agreement (English)

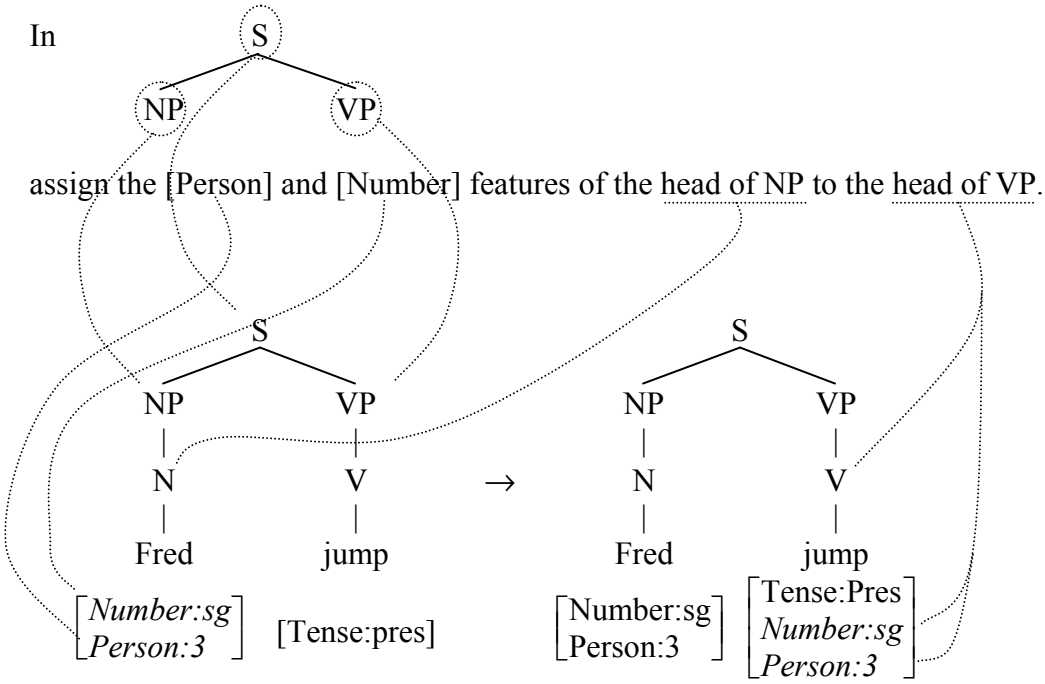
In 

assign the [Person] and [Number] features of the head of NP to the head of VP.

An application is shown as follows.

Verbal Agreement (English)

In



The rule of inflectional morphology that generates the *-s* suffix is given below:

3rd Singular Present Rule

Suffix *-s* if the morphosyntactic representation contains:
[Tense:Pres, Person:Singular, Number:3]

which will produce convert the stem *jump* in the tree above to the correct form *jumps.*

Compare: *I jump*, *they jump*, etc.

20.1.1 Agreement in general

In languages with rich inflection, agreement rules like the above copy a great deal of information around the tree: verbs agree with their subjects (and sometimes their objects, too), adjectives and articles agree with the nouns they modify, and in at least one language (Lardil, Australia) nouns agree with the verb of their clause in tense.

Summing up, agreement and case marking are the main phenomenon in which syntax determines morphosyntactic representation, and hence the inflectional form of the words of the sentence.

21. An example of phrase structure rules in another language: Hittite

Languages differ quite a bit in their word order, a fact which can be describe in grammars by writing different phrase structure rules.

One kind of analytic skill to be developed here is to formulate the phrase structure rules needed to analyze any particular language. Assuming you have a representative batch of sentences to work with, this involves two steps:

- Parse all the sentences
- Look at all the trees, and see which daughters any given type of node can have.
- Express what you find with a reasonably economical set of rules.

The data below involve sentences in Hittite, taken from an exercise created by Jay Jasanoff of Harvard University. The transcription and syntactic analysis were guided by input from my UCLA colleague Prof. Craig Melchert; both are top experts on this language.

Hittite was spoken in early ancient times in what is now Turkey. It is known from a hoard of about 25,000 cuneiform tablets discovered early in the last century and deciphered in the decades that followed. Some of the texts date back to about 1700 B.C. and thus count as the oldest attestation of any Indo-European language.[41] We accept here on Jasanoff's authority that the sentences below, which he made up, would be grammatical to real Hittite speakers if we could somehow bring them back.

Phonetic symbols are necessarily based on educated guesses. [x] is as in Spanish *jamon* or German *Bach* (voiceless velar fricative).

1. nu xassus salli parn-i anda estsi
   comp king big-dative house-dative in is
   'The king is in the big house.'

2. nu antuxsas akuwakuwan istamastsi
   comp man-nominative frog-accusative hears
   'The man hears the frog.'

3. nu antuxsas sallin akuwakuwan parn-a pexutetsi
   comp man-nominative big-accusative frog-accusative house-allative brings
   'The man brings the big frog home.'

4. nu akuwakuwas westar-i assun memijan tetsi
   comp frog-nominative shepherd-dative good-accusative word-acc. says
   'The frog says a good word to the shepherd.'

[41] Indo-European is the very large language family that includes (for example) English, Russian, Hindi, Latin, Irish, etc. See Chapter 13.

5. nu westaras sallin akuwakuwan pir-i anda xassussar-i katta istamastsi
   comp shepherd-nom. big-acc. frog.-acc. house-dative in queen-dative with hears
   'The shepherd hears the big frog in the house with the queen.'

6. nu akuwakuwas antuxsan natta istamastsi
   comp frog-nominative man-accusative not hears
   'The frog doesn't hear the man.'

7. nu xassussaras xassui piran salli akuwakuwi katta tijatsi
   comp queen-nom. king-dat. before big-dative frog-dative with comes
   'The queen comes before the king with the big frog.'

8. nu westaras assui xassui akuwakuwan pexutetsi
   comp shepherd good-dative king-dative frog-accusative brings
   'The shepherd brings the frog to the good king.'

One can do both syntactic and morphological analysis on these texts. At the level of morphology, it is possible to collect some partial noun paradigms, as follows.

xassu-s	king nominative
xassu-i	king-dative
antuxsa-s	man-nominative
antuxsa-n	man-accusative
akuwakuwa-s	frog-nominative
akuwakuwa-n	frog-accusative
akuwakuw-i	frog-dative
westara-s	shepherd-nominative
westar-i	shepherd-dative (a drops before i? not known)
memija-n	word-accusative
parn-i	house-dative
parn-a	home-allative
xassussara-s	queen-nominative
xassussar-i	queen-dative

It looks at least roughly that the nominative suffix is *-s*, the accusative suffix is *-n*, and the dative suffix is *-i*. This predicts **akuwakuwa-i* and **xassussara-i* for the datives of "frog" and 'queen'; in fact, there's a bit of phonology going on: the vowel *a* is dropped before this suffix. We expres the rules of inflectional morphology as follows.

Hittite Nominal Inflection (sketch)

Suffix -*s* when [Case:nominative]
Suffix -*n* when [Case:accusative]
Suffix -*i* when [Case:dative]

There also appears to be verbal inflection, for which we can conjecture this rule:

Suffix -*tsi* when [Person:3, Number:singular, Tense:present]

But in fact we know almost nothing about -tsi from these few data.

Turning now to the phrase structure rules, the idea is to inspect the sentences, parse them according the principles of the theory, and generalize over what we see to produce the rules.

An intriguing aspect of the sentences is that they all begin with *nu*. This is most likely a complementizers: Hittites usually spoke in $\overline{S}$'s, not S's, though it certain contexts it was possible to say just a plain S. Thus we will start our derivations with $\overline{S}$ and assume this phrase structure rule:

$\overline{S} \rightarrow$ Comp S

NP could be derived by the following rule:

NP $\rightarrow$ (A) N[42]

Probably the A should be an AP, but we will skip this for brevity.

Another simple rule is for PP, which is this language is evidently not a phrase for prepositions but for **postpositions**, which are just like prepositions but come after their NP rather than before. The phrase structure rule needed is:

PP $\rightarrow$ NP P

In sentences, the subject evidently comes before the predicate, justifying the rule

S $\rightarrow$ NP VP

The trickiest phrase structure rule to write here is for VP. Below are all of the data aligned in a format that places similar phrases vertically underneath one another.

[42] In truth we cannot know the relative order of Art and A; they never appear together in the data.

	NP	**NP**	**PP**	**PP**	**Adv**	**V**
1			salli piri anda big-dat. house-dat. in			estsi is
2		akuwakuwan frog-acc.				istamastsi hears
3	sallin akuwakuwan big-acc. frog-acc.	parn-a house-dative				pehutetsi brings
4	westari shepherd-dat.	assun memijan good-acc. word-acc.				tetsi says
5		sallin akuwakuwan big-acc. frog.-acc.	piri anda house-dat. in	hassussari katta queen-dat. with		istamastsi hears
6		antuhssan man-acc.			natta not	istamastsi hears
7			hassui piran king-dat. before	salli akuwakuwi katta big-dat. frog-dat. with		tijatsi comes
8	assui hassui good-dat. king-dat.	akuwakuwan frog-acc.				pehutetsi brings

If we collect all of the various items that evidently fit within a VP, and (going out on a limb) put them in a single rule, we get:

VP → (NP)(NP)(PP)(PP)(Adv) V

This completes the set of phrase structure rules, stated all in one place thus:

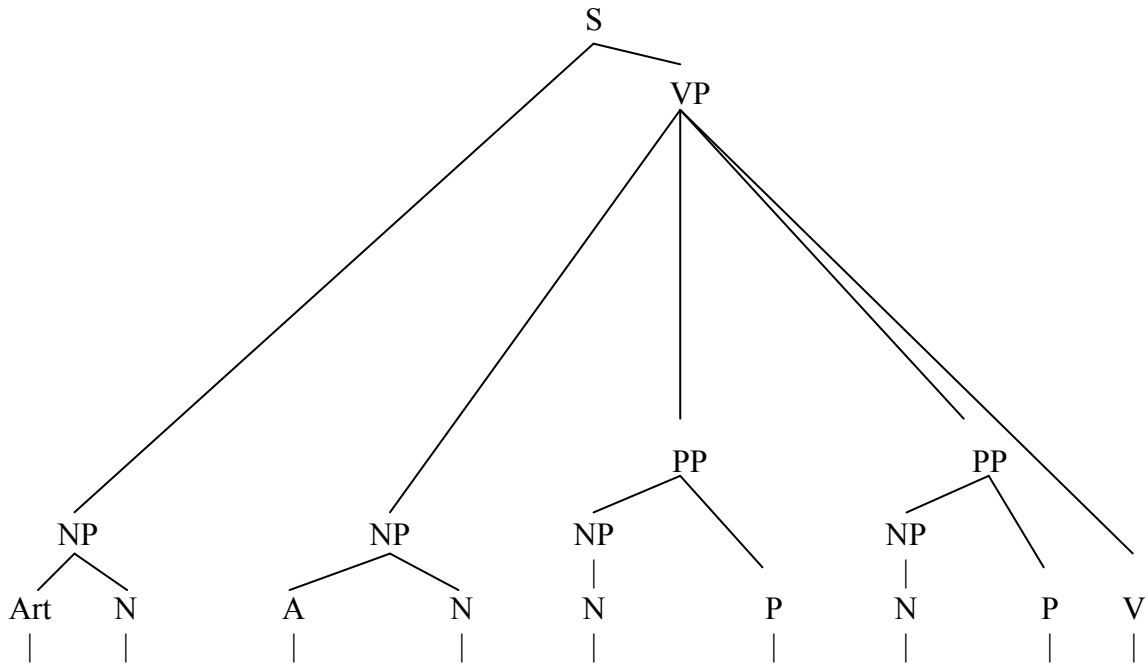
Phrase structure rules for Hittite

S → NP VP
VP → (NP)(NP)(PP)(PP)(Adv) V
NP → (Art) (A) N
PP → NP P

"P" must be read "postposition", rather than "preposition."

21.1 Example diagrammed sentence

The rules suffice to generate all the sentences; here is one particularly long example.



tsi

the shepherd-nom. big-acc. frog.-acc. house-dative in queen-dative with hears

'The shepherd hears the big frog in the house with the queen.'

Study Exercise #14

Parse the Hittite sentence

nu	akuwakuwas	westari	assun	memijan	tetsi
the	frog-nominative	shepherd-dative	good-accusative	word-acc.	says

'The frog says a good word to the shepherd.'

This area is intensionally blank.

21.2 Hittite as a head-final language

It can be seen that, at least in these data, Hittite is a **head-final** language: N is last in NP, P is last in PP, V is last in VP (and we don't know about AdjP).

Some other well-known head-final languages are Japanese, Korean, Bengali, and Turkish. The Bantu languages, such as Swahili and Zulu, tend to be strongly head-initial. English tends towards being head-initial, but is conflicted, in the sense that it puts adjectives before the head noun in NP. Hence some English noun phrases have the head noun in the middle:

[the long [**book**]$_{N}$ about linguistics]$_{NP}$

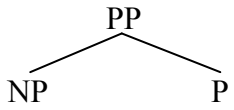
21.3 Case marking in Hittite

Hittite has a richer case system than English, with overt suffixes marking the Nominative, Dative, and Accusative. We can write syntactic rules that place the appropriate value for the the feature [Case], based on the configuration of the tree.

For instance, Dative case is assigned in Hittite by postpositions. It can be attached by a similar rule:

Dative Case Marking

In the configuration



add [Case:Dative] to the morphosyntactic representation of the head of NP.

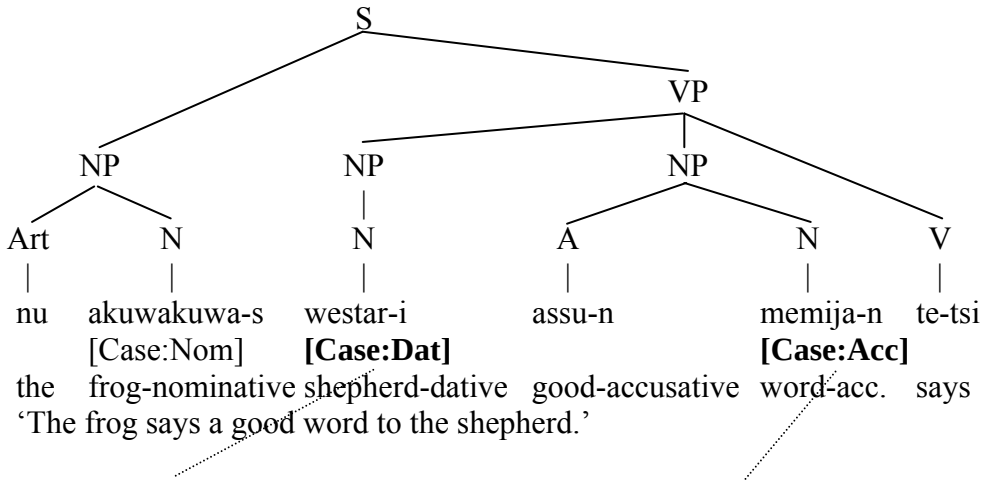
Getting Accusative and Dative objects right is trickier, and we also have very few data, so the following is really something of a guess:

Case Marking for Objects

In verb phrases containing one or more NP, then

- if there are two NP, assign [Case:Dative] to the head of the first and [Case:Accusative] to the head of the second.
- if there is just one NP, assign [Case:Accusative] to its head.

Here is an example with two NP's inside the VP:



"assign [Case:Dative] to the head of the first and [Case:Accusative] to the head of the second"

A further rule, not stated here, would cause adjectives (such as *assun* above) to agree with their head nouns in case.

22. A bit more on phrase structure rules: Kleene star

Let us beef up the system of phrase structure rules once more. Some phrase structure rules allow for any number of daughters of a certain type. For example, the rule for NP allows for an unlimited number of Adjectives preceding the noun, as in 'a long, dull,

boring movie'. A formalism for this often employed is to enclose in brackets the element that can be repeated indefinitely, and place an asterisk after the right bracket (the asterisk is known as "Kleene star", after the mathematician who proposed the notation). For example, the phrase structure rule for NP can be written as follows:

NP → (Art) (A)* N (PP)*.

An NP that uses both (A)* and (PP)* would be 'the big blue book about linguistics on the counter'.

Quite a few of the items on our previous phrase structure grammar would be more accurately depicted with Kleene star; the following is a list:

S	→	NP (Aux) VP	
NP	→	$\left(\begin{Bmatrix}\text{Art}\\ \text{NP}\end{Bmatrix}\right)$ (AP)* N (PP)*	Example: *his noble, inspiring gift of $1,000,000 to X*
NP	→	Pronoun	
AP	→	(Adv) A	
VP	→	V (NP) (NP) (PP)* ($\overline{\text{S}}$)	Ex.: *sold books to students for $50*
PP	→	P NP	
$\overline{\text{S}}$	→	(Comp) S	
NP	→	NP (Conj NP)*	Ex. *Alice and Sally and Bill left.*
VP	→	VP (Conj VP)*	Ex. *We sang the song and danced the dance.*
PP	→	PP (Conj PP)*	Ex. *We tossed it over the fence and through the window.*
S	→	S (Conj S)*	Ex. *He said that he was sick and he would go.*
$\overline{\text{S}}$	→	$\overline{\text{S}}$ (Conj $\overline{\text{S}}$)*	Ex. *He said that he was sick and that he would go.*
V	→	V (Conj V)*	Ex. *They washed and diced the vegetables.*

Chapter 5: Syntax II — Transformations

1. Syntax beyond phrase structure: the need for transformations

As seen already, our overall goal is to beef up the grammar so that it becomes an ever better approximation to the grammar internalized by speakers of English. We have done this by amplifying the system of phrase structure rules, and also by adding rules of agreement and case marking to govern the distribution of inflectional features. This section introduces the next major type of grammatical rule, the **transformation**, and argues for why it is needed.

English contains a construction called the **Tag Question**. Tag questions appear after the comma in the following examples:

Frogs can eat flies, can't they?
The president has resigned, hasn't she?
Bill was watching the stew, wasn't he?

As the data show, a tag question contains three parts in order:

- A copy of the Aux of the main sentence (*can…can*, *has … has*, *was ... was*).
- A contracted form of the word *not*
- A pronoun expressing the person and number of the subject of the main sentence.

1.1 Digression: spell-out rules

Before going on, we need a bit of clarification: we are assuming, as seems intuitively reasonable, that *can't* is the normal realization of *can not*, *hasn't* is the normal realization of *has not*, and (more interestingly) *won't* is the normal realization of *will not*. For such contractions (as school grammar calls them), we need minor morphological "spell-out" rules, of which the following are a partial list:

Some Spell-Out Rules of English

will not → *won't*
can not → *can't*
am not → *aren't*[43]
do not → *don't*
shall not → *shan't*[44]
etc.

[43] As in *I'm tall, aren't I*?, used only in vernacular speech. (Remember your white lab coat…)

[44] Archaic, at least for Americans.

1.2 Phrase structure rules inadequate for tags

It is in the nature of phrase structure rules that they *can't copy*: they specify the daughter nodes of a particular kind of mother node, as well as the order in which the daughters appear, but that it all. If we naively attempted to generate tag questions simply by extending our set of phrase structure rules, we would derive many ungrammatical instances with a mismatched Aux, because these rules lack the copying capacity. Here is the failed approach in detail:

Failed Hypothesis for Tags

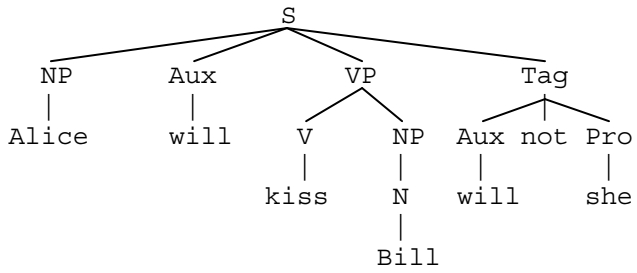
(a) Change the phrase structure for S to:

S → NP (Aux) VP (Tag)

(b) Add the phrase structure rule:

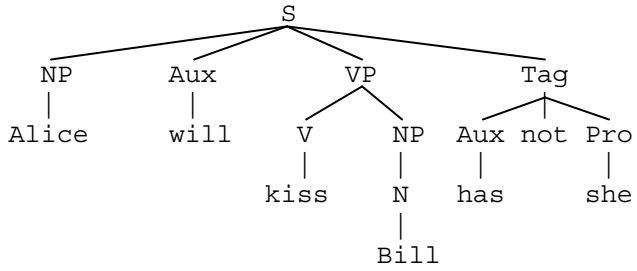
Tag → Aux *not* Pro

This hypothesis derives *Alice will kiss Bill, won't she?* as follows:



(The tree shows the pre-spelled-out version of the sentence; the spell-out rule would convert *will not* to *won't*.)

This hypothesis fails because it doesn't enforce copying. We can apply the very same rules and derive preposterous sentences:



By Spell-Out: *Alice will kiss Bill, hasn't she?

and similarly:

*Alice will kiss Bill, can't he?
*Alice will kiss Bill, hasn't it?
*Alice will kiss Bill, won't they?

1.3 Terminology of failed grammars

Linguists are always dealing with failed grammars like the one just given, taking them back to the drawing board and trying either to improve them or replace them with a better approach. Failed grammars are not a pointless activity; they lead us to explore the data more thoroughly and force us to refine or revise our thinking.

Some terminology for failed grammars that is widely used:

undergenerate: A grammar *undergenerates* when there are grammatical sentences that it can't generate.

overgenerate: A grammar *overgenerates* when it generates ungrammatical sentences

The grammar we just looked at overgenerates, as the * examples above indicate. A grammar that can't generate tag questions at all (what we had before) undergenerates.

1.4 Diagnosing failure

As suggested earlier, the failed grammar given above is failing because nothing in the rule apparatus developed so far can copy. (The failed hypothesis just given is a poor substitute for copying: it copies the structure, but not the actual words involved.) Plainly, we need more kinds of rules.[45]

More generally, phrase structure grammars don't allow for cases where the constituents present in one part of a tree depend on the constituents present in another part, which may be some distance away. In fact, tag questions are a rather out-of-the-way instance of this phenomenon; the really important cases are yet to come. The tag questions will suffice, however, to give the basic idea.

1.5 Transformations

Faced with phenomena like tag questions, linguists generally assume that phrase structure rules do not alone suffice as a grammar formalism for languages. An additional kind of rules takes as its input a sentence generated by the phrase structure rules and alters it in some way.

[45] For honesty's sake, I should add that you *could* produce a phrase structure grammar that copies Auxes, but intuitively speaking it would be a really crummy grammar. The trick is to replace S and Tag with a whole set of nodes like S_{can} "S with *can* as its Aux", S_{will} "S with *will* as its Aux", each allowing a matching daughter Tag_{can}, Tag_{will}, etc. This gets the facts but fails to characterize the tags as involving *copying in general*.

- A syntactic rule that alters a tree structure is called a **transformation**.
- A grammar that includes transformations as well as phrase structure rules is a **transformational grammar**.

The rules of case marking and agreement given earlier in this book could be considered a kind of transformation, although their effects are not as dramatic as the copying and word-movement processes we will cover in what follows; as they change only the morphosyntactic representation.

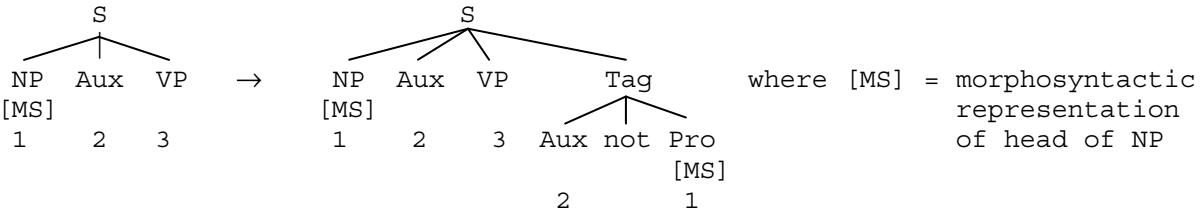
The general strategy seen in transformations is to let the phrase structure rules define the "basic inventory" of sentences in the language, and let the transformations apply to generate a wider variety of sentences that go beyond the capacity of phrase structure rules. For example, the sentence *Alice will kiss Bill* is in some sense a basic sentence (being generable by phrase structure rules alone), and *Alice will kiss Bill, won't she* is in a sense a syntactic elaboration of the simple sentence.

What can transformations do? This is a rather open question, whose answer forms a large part of the theory of syntax. At the moment, it's best to simply formulate the transformations we need and later on see what general theoretical principles are applicable.[46]

1.6 A transformation for tag questions

Here is a copying transformation that can derive tag questions. As you can see, it uses notation seen earlier in morphology, where we used numeral subscripts to make clear what changes into what for rules of infixation and reduplication. However, the syntactic transformation also contains reference to the tree structure that is manipulated.

Tag Question Transformation



Here is an explication of this rule. It assumes you have an S, consisting of an NP, an Aux, and a VP. The NP is assumed to have a morphosyntactic representation, that is, a feature bundle located on the head of the NP. These three items (NP, Aux, VP) are subscripted 1, 2, and 3.

On the right side of the arrow in the rule, the change is shown. A new daughter of S is added at the right edge, with the category Tag. Its internal content consist of an Aux, the

[46] To be honest, we're really going to stop at the first step here; for a deeper theory of transformations you'll have to take more syntax courses (the UCLA sequence is: 120B, 165B, various graduate courses…)

word *not*, and a Pronoun. The Aux is a copy of the Aux in the original sentence (this is shown by its bearing the number 2), and the Pronoun is assigned a copy of the morphosyntactic representation of the subject (this is indicated by the numerical subscript 1). Assuming that the features [Gender], [Person], [Number], and [Case] are part of the morphosyntactic representation, this will place the appropriate kind of pronoun into the tag; masculine subjects will get masculine pronoun, plural subjects will get plural pronouns, and so on.

For explicitness, here are the nominative pronouns of English with their inflectional features (for the non-nominative pronouns, see p. 80 above).

I	[Case:Nominative, Person:1, Number:Singular]	(Gender is free)[47]
you	[Case:Nominative, Person:2]	(Gender and number are free)
he	[Case:Nominative, Person:3, Number:Singular, Gender:Masculine]	
she	[Case:Nominative, Person:3, Number:Singular, Gender:Feminine]	
it	[Case:Nominative, Person:3, Number:Singular, Gender:Neuter]	
we	[Case:Nominative, Person:1, Number:Plural]	(Gender is free)
they	[Case:Nominative, Person:3, Number:Plural]	(Gender is free)

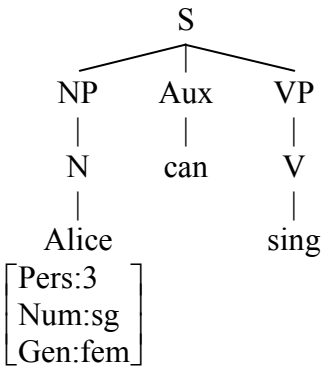
We assume, for example, that an abstract entity like

Pro
[Case:Nominative
Person:3
Number:Singular
Gender:Feminine]

would be spelled out as *she*.

Here is a derivation for the sentence *Alice can sing, can't she?*

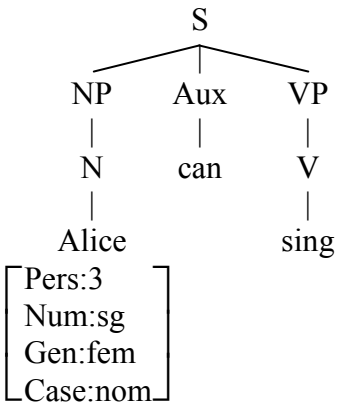
First step: application of the phrase structure rules to derive *Alice can sing*



[47] By "free", I mean that you can use this pronoun no matter what the specification in the morphosyntactic representation.

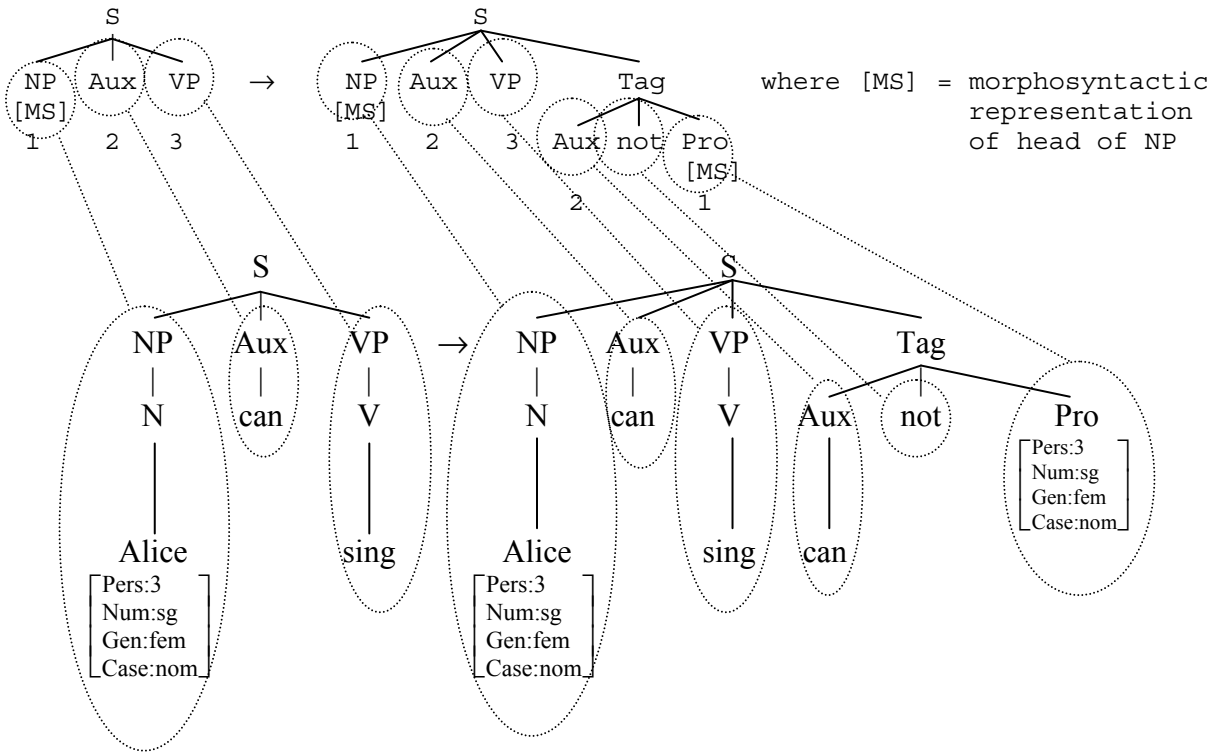
Note that *Alice* by its meaning is inherently 3rd person, singular, and feminine.

Second step: since Alice is the subject, a rule of case marking makes it Nominative:



Third step: application of Tag Question Transformation (matchup shown with dotted lines):

Tag Question Formation



Fourth step: spell out the pronoun that is [Pers:3, Num:sg, Gen:fem, Case:nom] as *she* (tree omitted):

Alice can sing, can not she?

Last step: spell out the sequence *can not* as *can't*:

Alice can sing, can't she?

Study Exercise #15

Derive *The frogs will sing, won't they?*

The steps should include Phrase Structure rules, Tag Question formation, spelling out of the pronoun *they*, spell-out of *won't*, and attachment of the plural suffix (inflectional morphology) to *frog.*

This page is intensionally blank.

This area is intensionally blank.

2. Another Transformation: Yes/No Questions

Consider a second transformation. Every sentence in English that is a statement has a corresponding question. Thus for

Bill is leaving

we have

Is Bill leaving?

and for

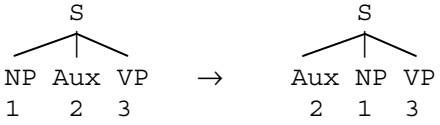
The frog might hop.

we have

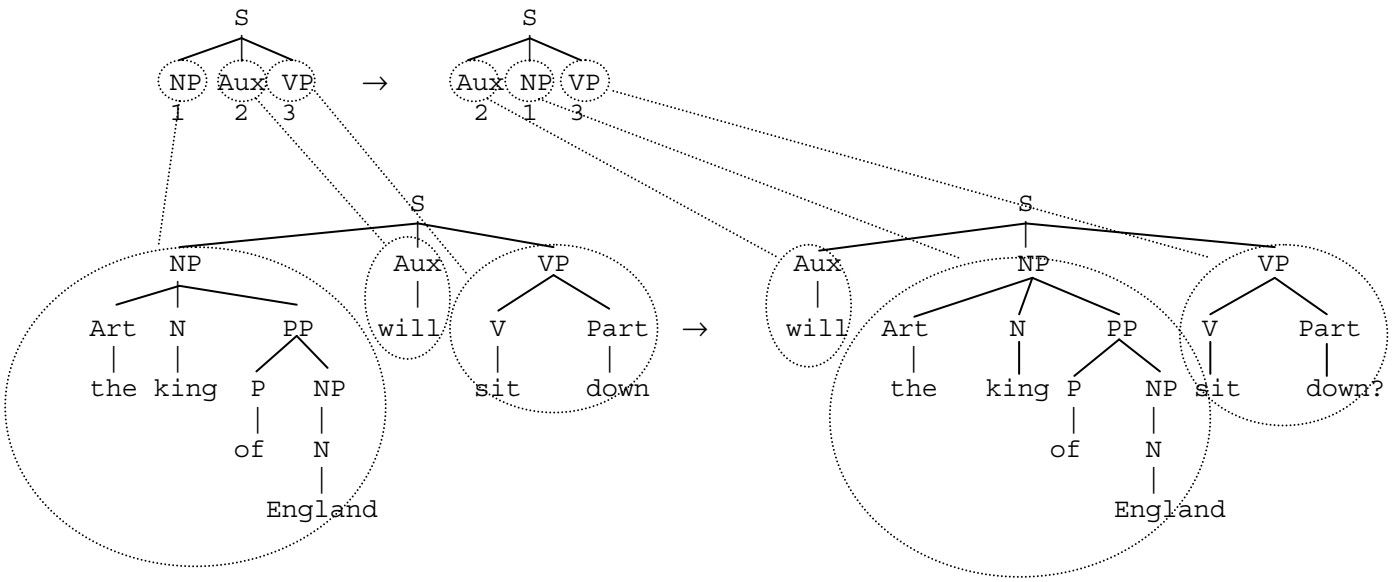
Might the frog hop?

Such questions are called **Yes/No questions**, to distinguish them from questions that begin with 'who', 'what', 'where', etc., which are called **Wh- questions**. It is plausible to regard a yes/no question as a syntactic variant of the corresponding statement; thus the phrase structure rules will derive the statement, which is converted to a yes/no question. The crucial transformation is as follows:

Yes/No Question Formation



Here is a derivation; dotted lines show the matchup between rule and form:



You may be worried at this point that we have no way of forming Yes/No questions from a sentence that lacks an Aux. This issue addressed in the next section.

For now, it's worth considering Yes/No Question Formation as a transformation. In this case (unlike for tag questions), it would be quite possible to derive the sentences just by using phrase structure rules, something along the lines of:

S → Aux NP VP

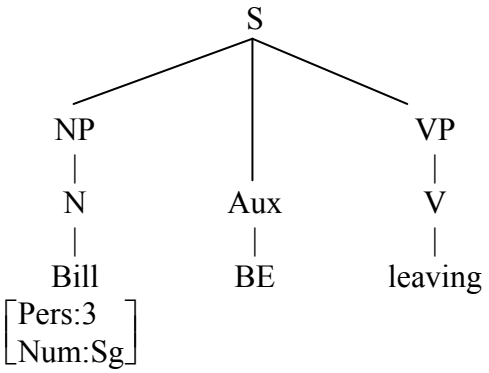
However, there seem to be at least two reasons that at least suggest that the transformational approach is better. First, speakers seem to recognize that (for example) *Is Bill leaving*? is the yes/no question that "goes with", or is appropriately paired with, *Bill is leaving*. We can characterize this sense of relatedness if we derive the question from the statement. Moreover, Auxes in English agree with their subjects (see previous readings for English agreement):

Bill is leaving.	(3rd person singular)
Bill and Fred are leaving.	(3rd person plural)

This agreement is carried over into the questions:

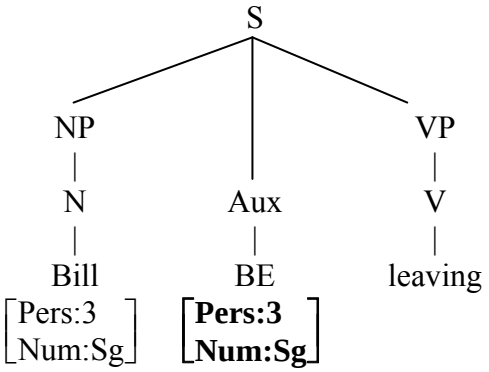
Is Bill leaving?
Are Bill and Fred leaving?

A clean analysis of this is possible, in which we only state the agreement rule once, if the questions are derived from the statements. In brief, the derivation would like this:

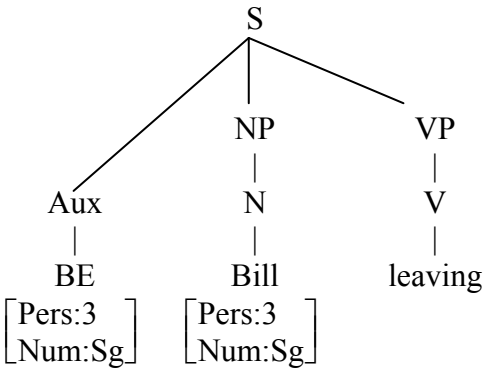


output of phrase structure rules

features of *Bill* are inherent in this noun



Agreement: copy the subject features onto the Aux



Yes/No Question formation

Is Bill leaving?

Morphology converts BE $\begin{bmatrix}\text{Pers:3}\\\text{Num:Sg}\end{bmatrix}$ into *is*.

This, then, is at least some justification for saying that Yes/No questions are formed by a transformation.

3. Inserted *do* in English

It's clear that tag questions and yes/no questions *can* be formed, even if there is no Aux in the base sentence. The method used in English is to insert the verb *do*, which could be described as the "default Aux" of the language.

For tag questions:

John likes turnips, doesn't he?
We left early, didn't we?

For yes/no questions:

Does John like turnips?
Did we leave early?

This phenomenon is an unusual aspect of English, and seems to be completely general. To give three additional examples:

- **Negation**: English negates a sentence by placing *not* directly after its Aux.

  I have not done my practicing.
  I will not take out the trash.
  He cannot play this concerto.

  Where the basic sentence has no aux (as in "He likes turnips"), *do* is inserted to provide one:

  I do not like turnips.

- **Polarity focus**: one can emphasize the truth of what one is saying (for example, to contradict someone who doubts it) by putting a strong accent on the Aux.

  I *have* done my practising.
  You *will* take out the trash.
  He *can* play "Chopsticks".

  Where there would otherwise be no Aux, *do* is provided:

  I *do* like turnips.

- **VP elision**. The second of two identical Verb Phrases can be elided, provided an Aux is left behind:

  You should take up hang gliding. Sue has. [that is, has taken up hang gliding]

This Aux will be *do* if no other Aux is present:

I wonder if there are any people who grow turnips around here. / Well, Bill does.

It would appear then, that some kind of process provide the aux *do* as the "backup Aux" whenever a syntactic transformation is applicable that requires an Aux to apply. A number of ways to formalize this idea have appeared, but I will not attempt this here, simply noting the general point that *do* is the "backup Aux" of English. We can at least state "what happens" as follows:

For all syntactic rules of English that refer to Aux, the Aux *do* is inserted prior to their application when the input sentence contains no Aux.

4. Summing up so far

Transformations have been posited to perform a variety of functions, as follows:

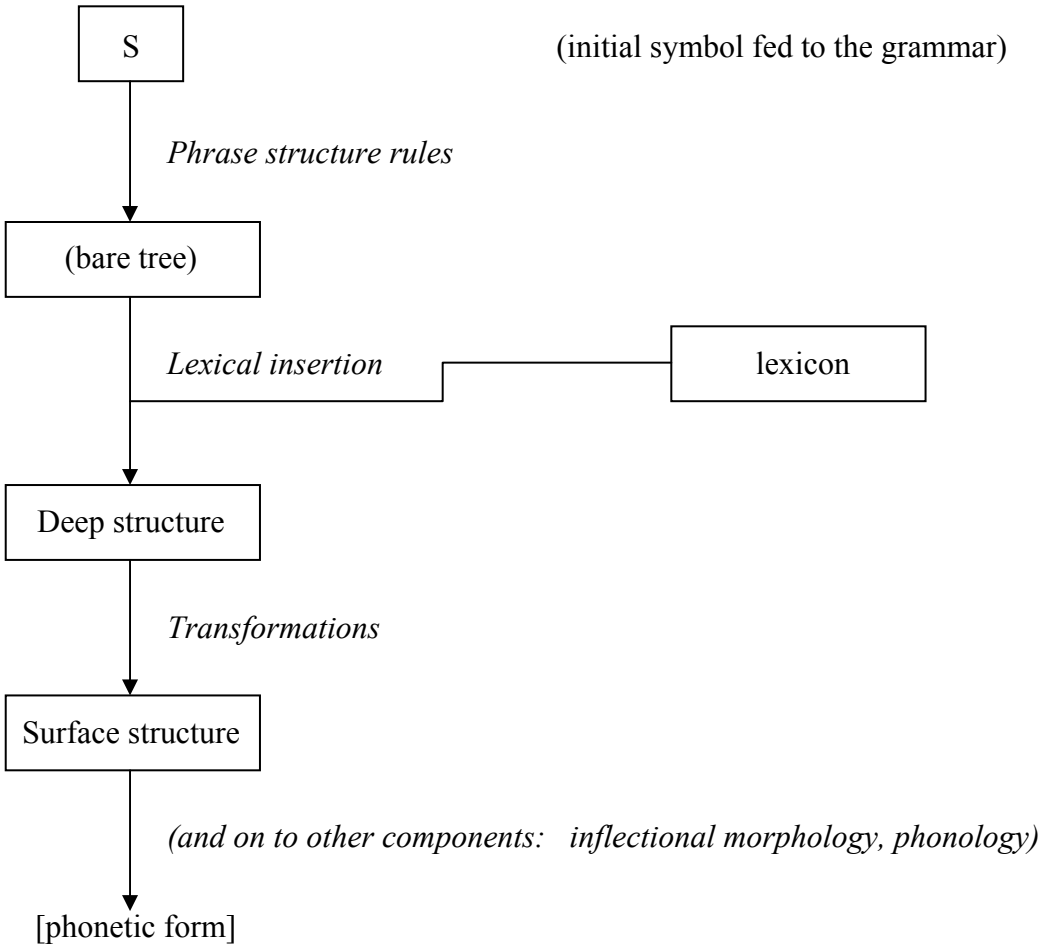
- **Assignment of inflectional features** to morphosyntactic representations (agreement and case marking)
- **Copying** of material (as in tags)
- **Movement** of parts of the tree (as in Yes/No Question Formation)
- **Insertion** (unformalized here) the of semantically empty words like the Aux *do*.

What remains to be covered are the most dramatic of transformations, the so-called long-distance movements. These will be studied in the next lecture, but to present them clearly, a preliminary concept, subcategorization, is needed.

5. Architecture of the theory: deep structure and surface structure

It may be useful at this point to back off and consider the architecture of the theory as developed so far. By this I mean the various kinds of rules and the order in which they are arranged; or the "direction of information flow" that the theory assumes. Such information can be expressed with diagrams containing boxes and arrows, and indeed is sometimes jokingly referred to as the "boxology" of the theory.

The following diagram of this sort incorporates the terms **deep structure** and **surface structure.**



The terms "deep" and "surface" involve no notion of profundity or superficiality. Deep structure is simply the output of the phrase structure rules with words plugged in by lexical insertion. Surface structure is the output of the syntax as a whole. In a sentence in which no transformations have applied, the deep and surface structures are the same.

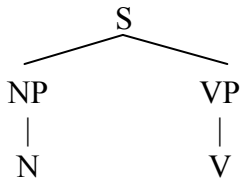
A caution to bear in mind is that a diagram of this sort is simply depicting the logical structure of the model; we are not (necessarily) making any claim that this represents the time course of sentence production in humans; but rather a claim about the structure of the language; that what we observe can be described in terms of a fixed number of perturbations of a simple structure that is generable by a phrase structure grammar.

Chapter 6: Syntax III — Subcategorization and Wh- Movement

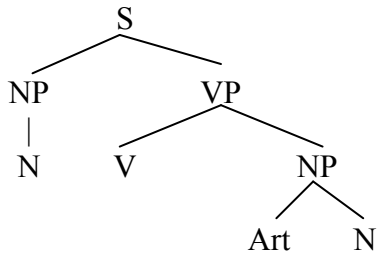
1. Lexical insertion and subcategorization

Our phrase structure rules generate, among many others, the following trees:

A.



B.



We have so far assumed that words are inserted whose part of speech matches up to the appropriate node in the tree. However, closer inspection shows that this procedure frequently overgenerates. Thus, for instance, a verb like *sigh* may appear in tree A but not tree B:

Fred sighed.
*Fred sighed his fate.

A verb like *destroy* behave in the reverse fashion: it can appear in B but not in A:

Bill destroyed his car.
*Bill destroyed.

Verbs like *destroy*, which must take an object, are called **transitive** verbs; verbs like 'sigh', which must not take an object, are called **intransitive**. Some verbs, such as *eat*, fit into both categories; they can be called "optionally transitive".

To avoid overgenerating in the way just shown, the theory needs a means of specifying the requirements of particular words for what tree structures they may appear in.

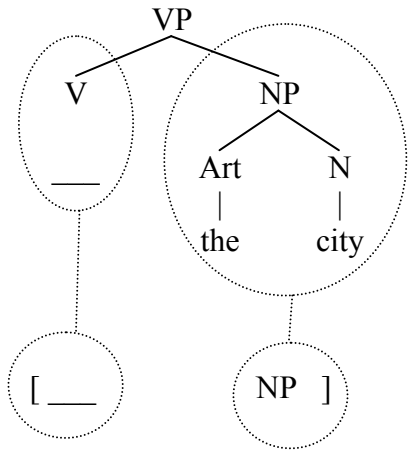
The process of "inserting words into the tree" is called **lexical insertion**. The idea given here is that speakers possess a mental dictionary, generally referred to as the **lexicon**. Lexical insertion consists of extracting a word from the lexicon and inserting it into a syntactic tree. The entries in the lexicon contain the crucial information about what kinds of tree the words can be inserted into, in the form of a **subcategorization frame**.

Thus, the lexical entry for *destroy* would be like this:

destroy:

Pronunciation: /dɪsˈtrɔɪ/ (more on this later)
Meaning: "violently cause no longer to exist" (we lack a better way to represent meaning)
Syntactic category: Verb
subcategorization frame: [___ NP].

The subcategorization frame indicate the sisters that must be present in order for the word to be legally inserted into the tree. *Destroy*, being a verb, will be inserted as the head of a VP. The subcategorization frame says that for insertion of *destroy* to be legal, the VP must contain an NP, occurring immediately to the right of V within VP. The diagram below is meant to explicate this notation:



Where a subcategorization frame does *not* include some particular phrase, then lexical insertion is impossible for where that phrase is present. Thus, for instance, the intransitive verb *sigh* would have the following subcategorization frame:

Here is a lexical entry for *sigh*:

sigh: pronunciation: /saɪ/
meaning: "exhale loudly to express sorrow"
syntactic category: verb
subcategorization: [___]

The frame [___] indicates that 'sigh' may not have sisters in the VP.

Optionally transitive verbs like *sing* (*John sang*, *John sang the song*.) have subcategorization frames that employ parentheses to show the optionality:

sing: pronunciation: /sɪŋ/
meaning: "use one's voice to produce music"
syntactic category: verb
subcategorization: [___ (NP)]

Verbs of saying and belief often subcategorize for an $\overline{S}$. For example, *say* has the subcategorization [___ (PP) $\overline{S}$] and *tell* has the subcategorization [___ NP ($\overline{S}$)]. This can be justified by the following sentences:

*Alice said.
Alice said to Bill that she would be going.
Alice said that she would be going.

*Fred told.
*Fred told that he would be going.
Fred told us.
Fred told us that he would be going.[48]

Nouns have subcategorization frames as well. For example, the subcategorization frame of *gift* is [___ (PP) (PP)], as in *a gift of $10 to the Red Cross*. The subcategorization frame of *picture* is [___ (PP)], as in *picture of Alice*. The subcategorization of *dog* is [___] (there are no noun phrases like **dog of teeth*). When a rule of word formation derives a noun from a verb that subcategorizes for $\overline{S}$, the resulting noun tends also to subcategorize for $\overline{S}$:

They believe that Sue left
their belief that Sue left

They assert that Sue left.
their assertion that Sue left

They claim the Sue left.[49]
their claim that Sue left.

Such constructions require us to modify the phrase structure rule for NP to the following:

$$\text{NP} \rightarrow \left(\begin{Bmatrix}\text{Art}\\ \text{NP}\end{Bmatrix}\right) \text{(AP)* N (PP)*}(\overline{\text{S}})$$

For discussion, see section 13.1 below.

[48] For completeness, observe that *tell* also has the subcategorization [___ (NP)(NP)]: *Fred told us his sorrows*, *Alice told them her name*.

[49] The word formation rule here is conversion, which adds no suffix or prefix; see p. 45 above.

1.1 Items not included in the subcategorization frame

Some constituents evidently get to appear "for free" in the syntactic tree; they don't have to be subcategorized. This is true for PP's with general adverbial meaning of place, time, or manner can occur with virtually any verb:

John sighed on Tuesday.
John sighed in the garden.
John sighed with great feeling.

The general practice for subcategorization is this: if any element *is always able to occur* as a sister, then we don't bother to mention it in the subcategorization frame. Basically, we are interested only in the restrictions that hold of individual words. This aspect of the grammar will not be formalized in this book.[50]

What is true of verbs is also true of nouns: PP's of place, time, and manner are ignored in determining noun subcategorizations, so cases like

dog in the garden
the party on Tuesday
a person in a good mood

would not justify a frame like [___ PP] for their nouns. Likewise, articles and possessors are not considered in the subcategorization frame, since they are possible for any noun (*the dog*, *Alice's dog*).

2. Subcategorization and meaning

Error! Bookmark not defined. It's a somewhat vexed question to what extent subcategorization should be treated (as it is above) as a straightforward matter of syntax. An alternative view is that heads occur in particular syntactic locations simply because of what they mean. For example, the verb *say* is entitled to occur in the syntactic frame [___ PP $\overline{S}$] because an act of saying generally has someone who is being spoken to (in *I said to Fred that I was leaving*, this is *Fred*), and a thing which is said (*I was leaving*). Similarly, *put* occurs [___ NP PP] because it involves a thing that is put, and location into which the thing is put. *Sigh* occurs [___] because nothing is affected when you sigh.

Although there is probably a grain of truth to this "semantics, not subcategorization" view, there are also reasons to treat it with skepticism.

First, there are cases of verbs that have very similar meanings, but different patterns of occurrence. Consider for instance *say* and *tell*.

[50] In a more thorough grammar, we place the subcategorized PP under the VP node, and the unsubcategorized PP for place/time/manner under S. This permits us to say that subcategorization is (normally) reserved for sister nodes.

I told Bill that I was leaving.
*I told to Bill that I was leaving.
*I said Bill that I was leaving.
I said to Bill that I was leaving.

It's not clear how semantics alone could tell us which verb requires an NP object and which a PP. Likewise, the pattern below:

I like jumping.	I like to jump.
I prefer jumping.	I prefer to jump.
I enjoy jumping.	*I enjoy to jump.

where only one of the three similar verbs can't take an infinitive subordinate clause (see Chapter 1), suggests that meaning won't suffice to tell us everything about subcategorization.

The verbs *give* and *donate* are semantically similar, but have different syntactic behavior:

She gave the library $1,000,000.	She gave $1,000,000 to the library.
*She donated the library $1,000,000.	She donated $1,000,000 to the library.

There is one more phenomenon that suggests that subcategorization has a slightly arbitrary character. Consider verbs like:

He ate.
She sang.
We raked.

These have what are sometimes called "implicit arguments"—it's understood that "he" ate *something*; and that likewise she sang something (song unspecified), and we raked (leaves or grass unspecified). In other words, the syntax does not always have to provide overt expression for all the participants in an act.

Yet in other cases, an implicit argument evidently is not allowed:

*We took.
*We own.

3. Solving subcategorization problems

The best method seems to be the following:

- Sit and think of lots of words and sentences that include the word you're considering. [51]

[51] I admit that this is harder for non-native speakers, a problem hard to avoid in linguistics teaching. If you don't have native intuitions in English, I suggest doing one of two things when you solve subcategorization

- Look at the phrase structure rule that introduces the word (for example, if you're dealing with a noun, look at the phrase structure rule NP → (Art)(A) N (PP)* ($\overline{S}$)). This will tell you the sisters that at least *might* be present.
- Remember that a subcategorized expression usually has a kind of intimate relation to the meaning of the word that subcategorizes it. The noun *claim* subcategorizes for an $\overline{S}$ because the $\overline{S}$ is used to designate the conceptual content of the claim.
- Try collecting as many individual frames for the word that you can, then use parentheses to collapse them into one or more simpler expressions.

Study Exercise #16

Give subcategorizations for the following words, justifying them with example sentences.

a. Verbs: elapse, award, tell, shout, die
b. Nouns: turtle, bowl, announcement, reason

problems on your assignments: either find a native speaker consultant and get their intuitions, or else add verbal discussion to your answer, with wording like "assuming that xxx is grammatical in English; I'm not sure." It would be fair to grade your answer based on the facts as you give them.

This page is intensionally blank.

This area is intensionally blank.

4. Wh-Movement

4.1 Backdrop

This section returns to the topic of transformations. Thus far, we've seen two reasons to move beyond simple phrase structure grammars to transformational grammars:

- Phrase structure rules cannot copy material—only a copying transformation can generate the legal array of tag questions.
- Phrase structure rules cannot relate sentences to one another (for example, simple statements to yes-no questions).

We now move on to what many linguists would probably agree is the most important basis for transformations, sometimes called "long distance gap-filler dependencies". The first example of such a case will be Wh- Movement.

4.2 Basics and terminology

A ***wh-Word*** is one of a fixed inventory of words used for asking questions. They are so called because most of the wh-words in English begin with these letters.

The *wh-* words of English can be various parts of speech:

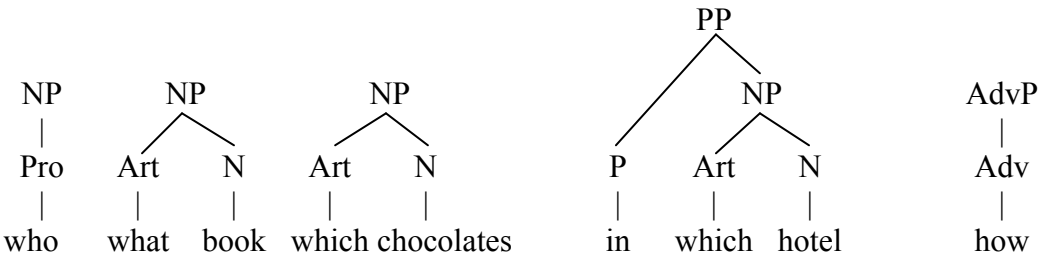
which	Article
whose	Article
who	Pronoun
whom	Pronoun
what	either an Article or Pronoun
how	Adverb
when	Adverb
why	Adverb
where	Adverb

A ***wh-* question** is a question that involves a *wh*-word. For example, the following are wh-questions:

Who did you see?
What book did you read?
Which chocolates did you like?
In which hotel are you staying?
How do you feel?

You can see that the wh- word usually comes at or near the beginning of the sentence. It constitutes, or is part of, a phrase that (intuitively), the sentence is about; i.e. the focus of the questioning.

A **wh- phrase** is an NP, PP, or AdvP (Adverb Phrase) that contains a *wh-* word and is placed at the beginning of a clause. In the wh- questions just mentioned, the wh- phrases are



This permits a more precise definition of **wh- question**; it is a question that begins with a wh- phrase.

4.3 Wh- questions, subcategorization, and gaps

Wh- questions are interesting in that they appear to violate otherwise-valid principles of subcategorization. Here is an example. The verb 'put' has the subcategorization [___ NP PP]. Because of this a sentence like the following:

Fred will put the chicken in the oven

is grammatical; the subcategorization of *put* is satisfied. But

*Fred will put in the oven

is ungrammatical because of the missing NP, and

*Fred will put the chicken

is ungrammatical because of the missing PP.[52]

It is a bit surprising, then, that the following sentences should be grammatical:

Into what oven will Fred put the chicken?
What chicken will Fred put into the oven?

These sentences, which are wh- questions, contain what are commonly called **gaps**. Instead of the NP or PP that the subcategorization calls for, one finds — nothing. The gaps are shown below, denoted with a ___:

What chicken will Fred put ___ into the oven?
Into what oven will Fred put the chicken ___ ?

Most people who ponder the question will judge that these gaps are (intuitively speaking) "filled" by the wh- phrase. We understand *what chicken* to be the object of *put* in the first sentence, and in the second sentence we understand *into what oven* to be the PP indicating where Fred put the chicken.

Let us define "gap", for precision:

- A **gap** is a location in syntactic structure where the subcategorization requirements would lead one to expect a phrase, but none occurs.

Such gaps are widely observed in English and in many other (not all) languages.

[52] As elsewhere we are ignoring extended uses of verbs, which often change the subcategorization. *John put the chicken* is fine in a fantasy world in which Olympic medals are awarded in the chicken-put.

There is an intimate connection between wh- phrases[53] and gaps: to a rough approximation, gaps are allowed only when a wh- phrase is present; recall

*Fred will put in the oven
*Fred will put the chicken

Moreover, most people who ponder the question will judge that gaps are somehow "filled" by the wh- phrase. In

What chicken will Fred put ___ into the oven?

we understand *what chicken* to be the object of *put*, and in

Into which oven will Fred put the chicken ___ ?

we understand *into which oven* to be the PP indicating where Fred put the chicken.

The two questions that demand to be answered here, then, are

- Why should wh- questions, and only wh- questions, permit gaps?
- How do we account for the fact that the wh- phrase at the beginning of the sentence intuitively fills the gap?

As you might be imagining already, the answer will involve a transformation.

5. Further background: echo questions

Before we proceed to the analysis, let us ponder a further phenomenon of English syntax, the so-called **echo question**. These are questions that contain a Wh- phrase, but have no gap; the Wh- phrase occurs in the ordinary position for its type, and satisfies the subcategorization requirements of the relevant head. Echo questions are not all that common, because they can only be used to offer an astonished replies to a parallel statement:

The Romans destroyed the television set.
The Romans destroyed *what*?

I saw Marilyn Monroe in Westwood last Saturday.
You saw *who*?

Fred will put the chicken in the Socratic Oven.
Fred will put the chicken in *what* oven?

Echo questions make an important point: it is possible to generate a wh-phrase in the 'normal' position for an NP or PP; wh- phrases do not always have to appear at the beginning of sentence.

[53] And, as we'll see later on, phrases that behave quite similarly to wh- phrases.

A bit of terminology: the wh- phrases of echo questions are sometimes said to be **in situ**, which is Latin for "in its original position". [54]

6. A transformation for Wh- questions

The grammatical problem at hand is that Wh- questions have subcategorization gaps that match up with the initial wh- phrases. This is a dependency that cannot be expressed with the phrase structure rules we have been using. These rules can only say what daughters a node may have, and thus they have no ability to regulate matchups between elements in the tree that are far apart. A transformation is needed.

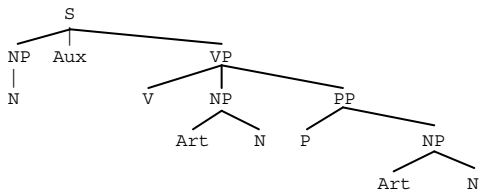
The intuitive idea behind our transformation analysis will be to let normal questions be derived from deep structures that look like echo questions. That is, we will have a transformation that will move the wh- phrase out of its *in situ* deep structure position (where it satisfies the subcategorization of the verb) to the beginning of the sentence. As a first approximation:

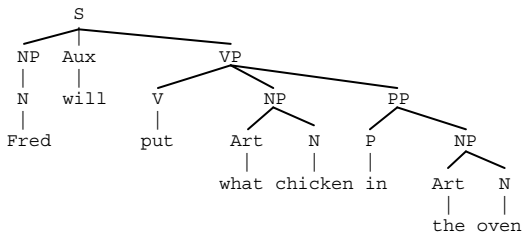
Wh- Movement

Move a wh- phrase to the beginning of the sentence, leaving a trace.

The term "trace" will be defined shortly.

Here is a derivation to illustrate the analysis. We begin with the phrase structure rules (on the left), then do lexical insertion (on the right):



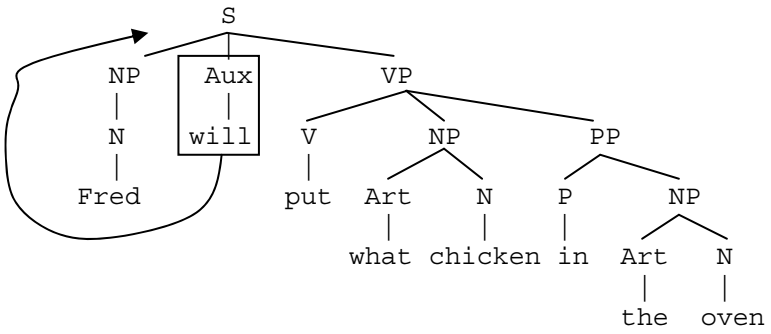


This creates the stage of deep structure, with *what chicken* in situ. The crucial point at this stage is that we have not violated the subcategorization of 'put', which in deep structure does have the required NP and PP sisters. In fact, with the theory we are working on, ultimately this will be seen to be true even in surface structure (more on this below).

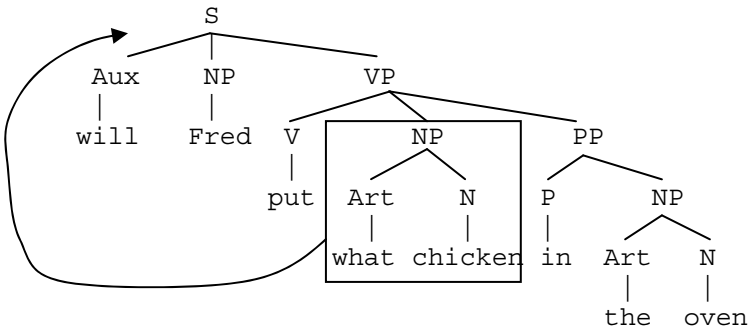
To derive surface structure, we apply two transformations: Subject/Aux Inversion (which, as mentioned earlier, is applicable to questions in general) and Wh- Movement. I show this below first by drawing arrows to show what moves where, then showing the surface structure that results. A caution: the destination of *what chicken* is provisional; we will change the analysis a bit below.

[54] *Situ* is an inflected form ([Case:ablative, Number:singular]) of *situs* 'place'.

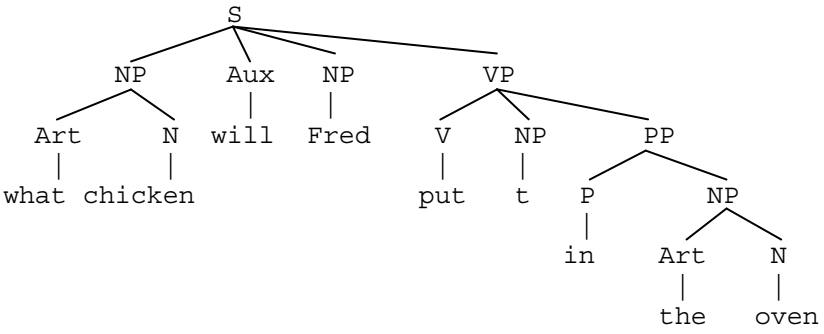
(1) Deep structure, showing Subject/Aux Inversion:



(2) Output of Subject/Aux Inversion, showing Wh- Movement:



(3) Surface structure:



As stated in the Wh- Movement rule, the movement of *what chicken* is assumed to leave a **trace**. A trace is essentially an empty copy of what got moved; it has the same category, but contains no phonetic material. To show that a trace is empty, we use the letter *t*, as the daughter of the trace's category. For now, the trace is just an arbitrary choice, but we will see later on that it plays an important role when we do the semantics of wh- questions and similar constructions. For now, we can observe that the trace NP means that the subcategorization requirements of *put* are satisfied (albeit by an empty, abstract entity) at surface structure as well as deep structure.

To review the general purpose of this transformational analysis: it offers an account of subcategorization gaps that does better justice to the facts than a phrase structure

grammar could. In this theory, gaps only arise from movement,[55] so the fronted wh-phrase will always match the gap. This ability to capture a **long-distance dependency** ("X here only if Y there") is a common justification for a transformational analysis.

Study Exercise #17

Explain how our grammar predicts that *'What city have the Romans destroyed Carthage?' is ungrammatical.

Study Exercise #18

Explain how our grammar predicts that *'Who will the princess sigh?' is ungrammatical.

Study Exercise #19

Derive the question 'Who will leave?'.

[55] A caution: there are many other sources of gaps, such as the dropped subject pronouns of Spanish, Persian, and many other languages. But these tend to have a special distribution, so the general point still holds.

This page is intensionally blank.

This area is intensionally blank.

7. The "landing site" of Wh- Movement

Wh- Movement doesn't always move words to the beginning of the sentence. In so-called **embedded Wh- questions**, movement is to the beginning of a subordinate clause. A wh-question is a subordinate clause that is itself a wh- question, as in the following examples.

I wonder what city the Romans destroyed.
We asked for whom the bell tolls.

They are found when the main clause has a verb like *wonder* and *ask*, which takes a question as its sister node. I'll assume that these verbs have a special categorization, not formalized here, that they take an $\overline{S}$ which is not a declarative (the usual case), but a wh-question.

An intriguing aspect of embedded questions is that they don't occur with the complementizer *that*:

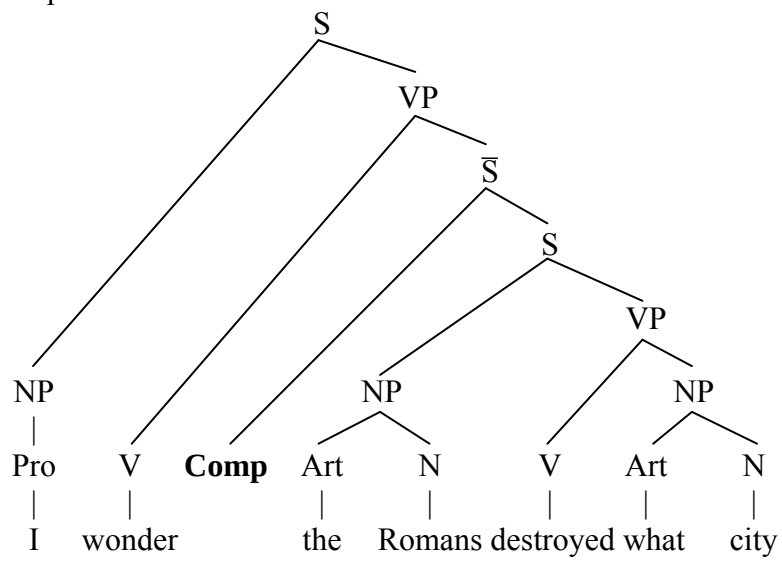
*I wonder what city that the Romans destroyed.
*We asked for whom that the bell tolls.

*I wonder that what city the Romans destroyed.
*We asked that for whom the bell tolls.

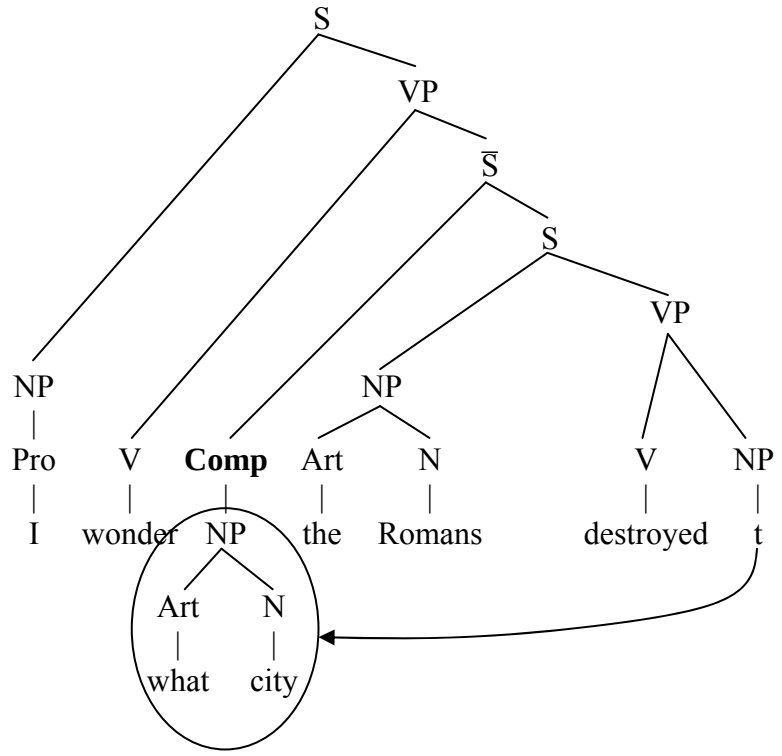
The interaction of such movement with complementizers (such as *that*) requires us to refine the analysis somewhat. An influential idea in syntactic theory that the order of words in sentences can be explicated in terms of **slots**, which the words compete to fill. We've already said that the Complementizer *that* occupies the position Comp, a daughter of $\overline{S}$. The idea to be developed here is that in an embedded Wh- question, the moved Wh-phrase actually *occupies* the Comp slot. When Comp is thus occupied, there is no room for *that*, which is semantically empty in any event.

Under this view, we can arrange lexical insertion simply to leave Comp empty for embedded clauses introduced by verbs like *wonder* and *imagine.* Then, Wh- Movement acts to fill the empty slot by moving the wh- phrase into it, as follows:

Deep structure:



Wh- Movement and surface structure:



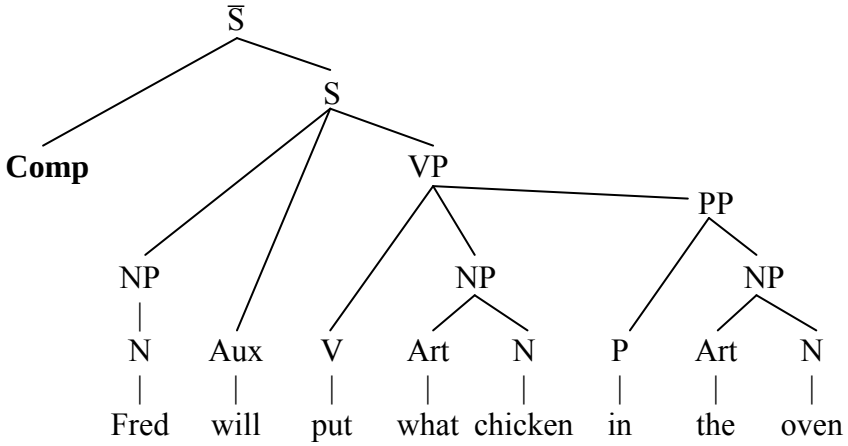
In this view, the empty Comp node provides a "landing site" for the moved Wh- phrase.

If this theory is going to work, we also need to cover the wh- questions that are *not* embedded, that is, the ones we started out with. There is a fairly reasonable tack that can be taken here, namely that *these also have Comp*, which provides the landing site for the sentence-initial wh-phrase. Specifically, the assumptions we need to make are as follows:

- Wh- questions are not instances of S, but of $\overline{S}$.
- They require (by means not stated here) that the initial Comp be empty in Deep Structure.[56]

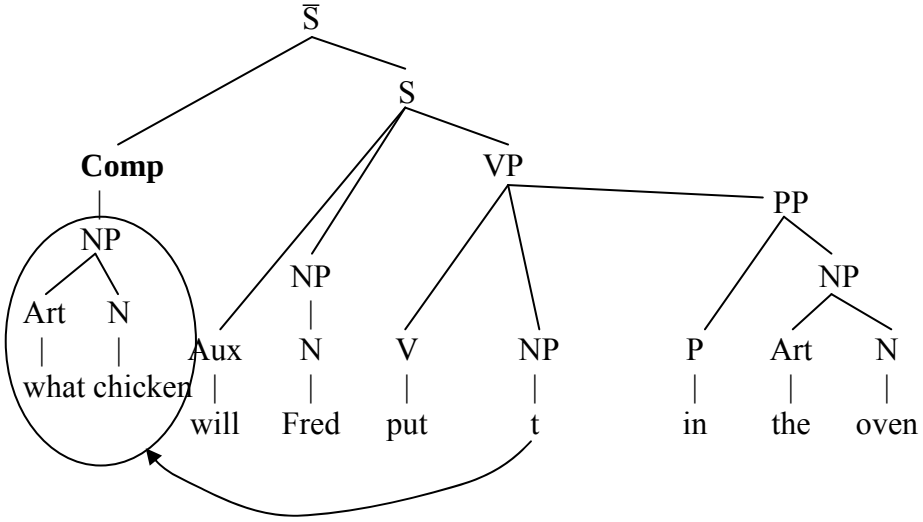
Under this analysis, the derivation of "What chicken will Fred put in the oven?" comes out slightly differently:

Deep structure:



[56] There are alternatives to this, for instance letting the moved Wh- phrase displace a *that*, and adding a transformation that deletes *that* from the topmost complementizer of the sentence.

Wh- movement (and Subject-Aux inversion), yielding surface structure:



With this in mind, we can express the Wh- Movement transformation more explicitly. To keep the notation from becoming rather messy, I will define a preliminary notation:

- Let wh- denote an NP, PP, or AdvP (Adverb Phrase) containing a wh- word.

This said, Wh- Movement can be stated as follows.

Wh- Movement

W	[	[	X]$_{Comp}$	X	wh-	Y]$_{\bar{S}}$	Z	→	W	[[	wh-]$_{Comp}$	X	t	Y]$_{\bar{S}}$	Z
1		2		3	4	5	6		1		4	3	4	5	6

where t is an empty element of the same category as wh-.

That is to say: when a sentence contains an $\bar{S}$, and the $\bar{S}$ begins with a Comp and contains a wh- phrase, the wh- phrase is moved to occupy the Comp position, leaving behind a trace of the same category.

7.1 Slots in syntax

This sort of analysis, in which an empty position is available for anything that moves (or, as we'll see, is copied), has been extended by linguists to a consistent, across-the-board practice, essentially "a place for everything and everything in its place." Thus, in more refined theories, there is a slot into which the Aux moves in questions, and many others. You will encounter this approach further if/when you study more syntax.[57]

[57] The next course in the Linguistics Dept. undergraduate sequence is Linguistics 120B, offered Winter 2009.

8. Typology of Wh-movement

Many languages other than English form Wh- questions by moving the wh- phrase to the beginning of the sentence. Here are three examples:

French:

Tu	as	vu	Paris			(normal statement)
you	have	seen	Paris			
Quelle	ville	as-	tu	vu	*t* ?	(wh- question)
what	city	have	you	seen	(trace)?	

Chamorro (South Pacific):

Hafahan	si-Maria	i-sanhilo	gi	tenda		(normal statement)
bought	Maria	the-blouse	at	the-store		
‘Maria bought the blouse at the store’						
Hafa	hafahan	si-Maria	*t*	gi	tenda	(wh- question)
what	bought	Maria	(trace)	at	store	
‘What did Maria buy at the store?’						

Vata (Ivory Coast, West Africa) Tones: 1 = highest, 4 = lowest

4 1	3	2 1			
Kofi	le	saka			(normal statement)
Kofi	ate	rice			
3	41	3		2	
yi	Kofi	le	*t*	la	(wh- question)
what	Kofi	eat	(trace)	question-particle	
‘What did Kofi eat?’					

Many other languages work in the same way; for example Modern Hebrew, Russian, and Spanish.

However, a large number of languages do not have Wh-Movement. These languages form Wh- questions simply by leaving the Wh- phrase in situ. An example of a non-Wh-Movement language is Persian:

Ali	a:n	keta:b-ra:	xa:nd
Ali	that	book	read
‘Ali read that book’			

Ali tʃe ketaːb-i xaːnd?[58]
Ali what book read
'What book did Ali read?'

*tʃe ketaːb-i Ali xaːnd?
what book Ali read

Japanese is similar:

John-wa **naze** kubi-ni natta no?
John-Topic why was fired question particle
'Why was John fired?'

Bill-wa [John-ga **naze** kubi-ni natta tte] itta no?
Bill-TOP John-nom why was fired Comp said question particle
'Why did Bill say that John was fired?'[59]

It's striking that the languages seem to pattern together; for instance, unbounded movement to the *right* is apparently exceedingly rare in language. Moreover, there are logical possibilities for Wh- movement that seem to be unattested:

- *Move a wh- phrase to the exact middle of a sentence.
- *Move a wh- phrase so that it follows the second word of a sentence.
- *Move all the words that precede the wh- phrase so that they follow the wh- phrase, and move all the words that follow the wh- phrase so that they precede the wh- phrase.

No such rules have been found in any language. We will discuss such cross-linguistic patterns in greater detail later on.

9. Why Wh- Movement?

It's something of a puzzle why languages have Wh- Movement at all—why not adopt the sensible Persian/Japanese/Chinese strategy, and just leave your Wh- words where they inherently belong? Surely it would be clearer for the listener to interpret the wh- word in its proper syntactic location.[60]

[58] If you're thinking about case marking here, the answer to your question is that the Accusative suffix *-raː* only attaches to *definite* Noun Phrases, the kind that would be translated with *the* in English.

[59] An odd custom of linguists writing in English about Japanese syntax is to use English first names. I don't know why they do this.

[60] Indeed, experimental work by psycholinguists has documented the increased cognitive load and memory burden that listeners experience when they have heard a wh- phrase and are "looking for" the corresponding gap.

A clue, I think, can be found in pairs of sentences that have the same gap, but where the Wh- phrase appears in a different location:

[What song] can Sue imagine that Bill sang *t* ?
Sue can imagine [what song] Bill sang *t* ?

Such pairs are often said to illustrate a difference of **scope**: the location of the wh-phrase indicates the domain in which the wh- phrase is acting as a logical operator. Thus, in the first sentence above, the wh- phrase *what song* is used to ask something about the content of Sue's imaginings—its scope is the entire sentence. The second sentence reports a thought of Sue's. Within this thought, *what song* is being used to ask something about Bill's singing (that is, Sue is mentally answering the question, "What song did Bill sing?"). Therefore, the scope of *what song* in the second sentence is just the subordinate clause. It can be seen, then, that the linear position of the wh- phrase is suited to expressing a distinction of scope.

What emerges, if this speculation is correct, is that there's no perfect design available. Languages without wh- movement make it clear where the inherent location of the wh-phrase is, but are less clear in indicating scope; languages with wh- movement mark scope clearly, but impose a burden on their listeners in the form of gap detection.

10. The unbounded nature of Wh- Movement

An important aspect of Wh- Movement is that it can move a wh- phrase over very long stretches of syntactic structure. Consider the following deep structures and corresponding surface structures:

You have seen who.
[Who] have you seen *t* ?

Joan thinks that you have seen who.
[Who] does Joan think that you have seen *t* ?

Bill would imagine that Joan thinks that you have seen who.
[Who] would Bill imagine that Joan thinks that you have seen *t* ?

Sally believes that Bill would imagine that Joan thinks that you have seen who.
Who does Sally believe that Bill would imagine that Joan thinks that you have seen?

Study Exercise #20

Provide a syntactic derivation (that is, deep structure, arrows showing what moves where, surface structure) for the sentence 'What city will Fred say that Judy thinks that you live in?'.

This page is intensionally blank.

This area is intensionally blank.

11. Another transformation: Topicalization

English has a number of transformations similar to Wh- Movement. Perhaps the simplest is the so-called Topicalization rule, used to account for sentences like these:

Linguistics, I can teach.
Those guys we would never give our credit cards to.
In that oven you should never put a chicken.

The name of the rule is from that fact that the fronted NP serves as the "topic" of its sentence; what it is about. These sentences have a distinctly rhetorical character, and often sound best if you imagine that the topic is being contrasted with some other topic:

Postmodernism, I'm clueless about, but linguistics, I can teach.

The "landing site" for fronted topics is *not* Comp, since you can get both *that* and the fronted topic in sequence:

I'd say that linguistics, I can teach.

Thus I will state the rule as simply moving a phrase to the left edge of S, as follows:

Topicalization

$$\begin{array}{lllllll} V & [& X & \left\{\begin{array}{l}NP\\PP\end{array}\right\} & Y &]_S & Z \\ 1 & & 2 & 3 & 4 & & 5 \end{array} \quad \rightarrow \quad \begin{array}{lllllllll} V & [& \left\{\begin{array}{l}NP\\PP\end{array}\right\} & X & t & Y &]_S & Z \\ 1 & & 3 & 2 & 3 & 4 & & 5 \end{array}$$

Topicalization, like Wh- movement, appears to be unbounded, though the examples that show this tend to be a bit less natural:

John, I don't think a lot of people would like.
Fred, I'd imagine that you'd think that a lot of people wouldn't like.

As unbounded transformations, Wh- Movement and Topicalization (as well as others to come) have some crucial similar behaviors, which we'll examine next week in discussing "islands".

12. Syntactic derivations

In a "derivation", one applies the rules in order. In the theory we are working with, sentences are derived by first creating a deep structure with the phrase structure rules and lexical insertion, then by applying the transformations in order.

Here is a formalized version of Subject Aux Inversion.

Subject-Aux Inversion

$$\begin{array}{lllllll} X & [& NP & Aux & VP &]_S \\ 1 & & 2 & 3 & 4 & \end{array} \quad \rightarrow \quad \begin{array}{lllllll} X & [& Aux & NP & VP &]_S \\ 1 & & 3 & 2 & 4 & \end{array}$$

In words, it says, essentially: "Switch the order of Aux and NP, under S."

Wh-Movement is stated as follows:

Wh- Movement

$$\begin{array}{lllllllll} W & [& [& X &]_{Comp} & X & wh\text{-} & Y &]_{\bar{S}} & Z \\ 1 & & 2 & & & 3 & 4 & 5 & & 6 \end{array} \quad \rightarrow \quad \begin{array}{lllllllll} W & [[& wh\text{-} &]_{Comp} & X & t & Y &]_{\bar{S}} & Z \\ 1 & & 4 & & 3 & 4 & 5 & & 6 \end{array}$$

where

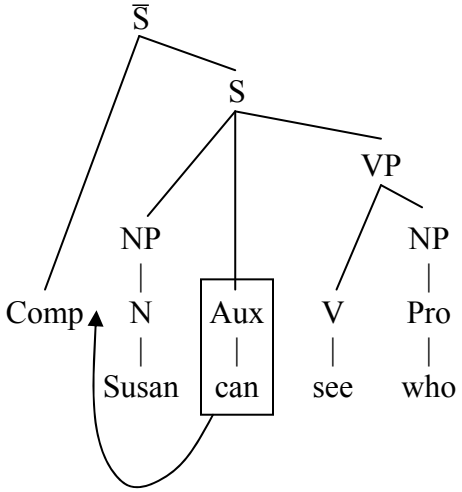
- *wh-* is a phrase containing a *wh-* word
- t is an empty element of the same category as *wh-*.
- W, X, Y, and Z are variables, standing for any word sequence.

This, too, is a bit more intuitive expressed in words:

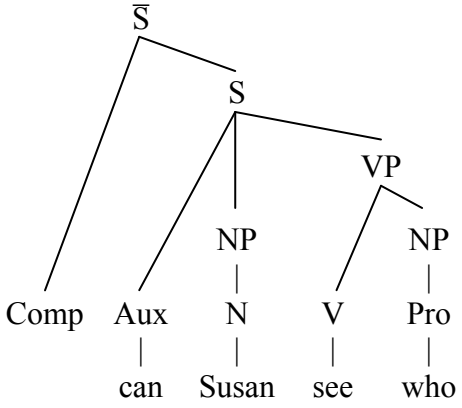
"Move a wh- phrase into Comp, leaving a trace."

In many English questions, both Subject-Aux Inversion and Wh- Movement both apply. Here is the sentence *Who can Susan see*.

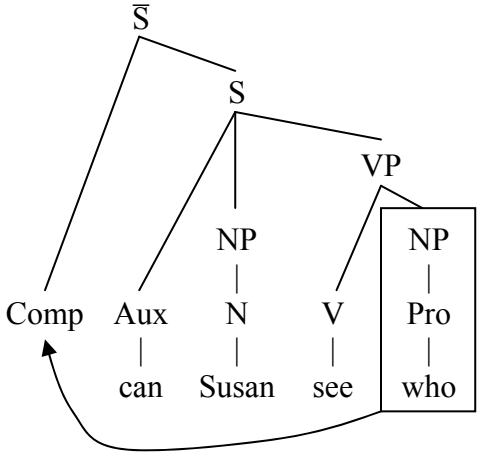
Deep structure, with box around the Aux that is going to move by Subject-Aux Inversion:



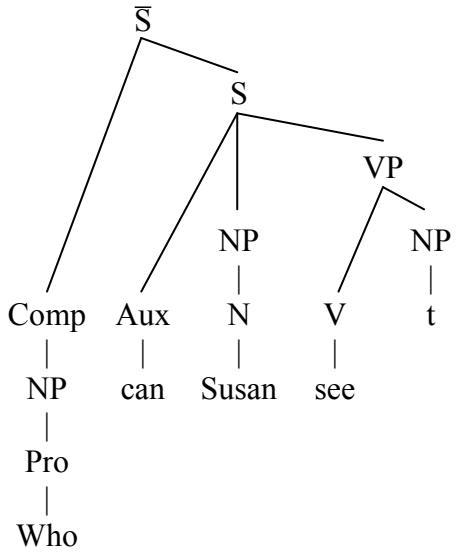
Result of moving the Aux:



Now showing how Wh- Movement is going to work:



Surface structure:



Study Exercise #21

Provide a syntactic derivation (that is, deep structure, arrows showing what moves where, surface structure) for the following sentences:

a. What city will Fred say that Judy thinks that you live in?
b. Which book will Sue ask that we study?
c. Sue will ask which book we should study.

Assume the transformations of Subject/Aux Inversion and Wh- Movement.

This page is intensionally blank.

This page is intensionally blank.

This area is intensionally blank.

This page is intensionally blank.

This area is intensionally blank.

This area is intensionally blank.

13. Update of Phrase Structure Rules

[note to self: this belongs a few pages earlier, under subcategorization. 2010: better late than never.]

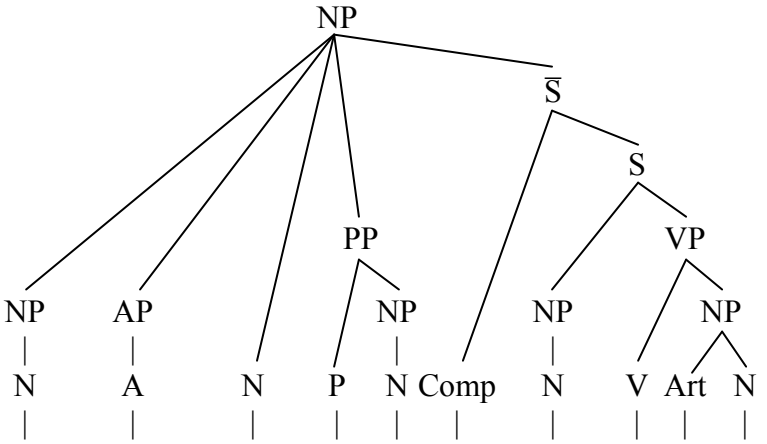
Here are the last two steps in our gradual update of the phrase structure rules.

13.1 $\overline{S}$ as daughter of NP

In the discussion of subcategorization above (p. 118), we found that there are a few nouns that subcategorize for $\overline{S}$. Typically these are nouns that express beliefs or statements, such as *belief*, *claim*, *assertion*, and so on. Thus, the NP phrase structure rule has to include the option of an $\overline{S}$ under NP.

$$NP \rightarrow \left(\begin{Bmatrix} Art \\ NP \end{Bmatrix}\right) (AP)^* \; N \; (PP)^* (\overline{S})$$

A long NP that includes every possibility for the main NP phrase structure rule is the following:



Bill's forthright assertion to Fred that television has no future.

Further example sentences showing the $\overline{S}$ option are as follows:

Fred's *belief* [$_{\overline{S}}$ that he is a genius]
George's *insistence* [$_{\overline{S}}$ that he be included]
the *claim* [$_{\overline{S}}$ that this structure is an island]
Sally's *assertion* [$_{\overline{S}}$ that we should eat pasta]

Note that all of these nouns are created by a rule of word formation from verbs that themselves categorize for $\overline{S}$ (*believe*, *insist*, *claim*, *assert*). (Others are not, however: *hypothesis*, *hunch.*)

In passing, observe that there is another kind of $\overline{S}$ that can occur as part of an NP. It is *not* subcategorized. You can tell that it is different because it includes a gap.

the turtles [$_{\overline{S}}$ that we caught ___ in the pond]
the beliefs [$_{\overline{S}}$ that we hold ___]
the king [$_{\overline{S}}$ who the peasant deposed ___ last year]

This is called a **relative clause**; we won't have the time to analyze them here.

To see the difference between subcategorized $\overline{S}$'s and relative clauses, it may help to observe that you can get both of them in the same NP:

the assertion [$_{\overline{S}}$ that we should eat pasta] [$_{\overline{S}}$ that you made ___]

head of NP

Subcategorized $\overline{S}$ $\overline{S}$ forming a relative clause

13.2 AP as daughter of VP

The Persian homework had the VP *raftani ast*, 'go-incipient adjective be-3 singular present'. This implies a phrase structure rule VP → AP V. English has the same thing, only in verb-initial order:

V → V AP

Fred is sick.
Bill looks tired.
Alice seems very friendly.

These three examples make use of three of the (fairly unusual) verbs of English that subcategorize for an AP.

13.3 Final phrase structure rule list

Putting all of these together, we have the phrase structure rules shown:

S → NP (Aux) VP
NP → $\left(\begin{Bmatrix}\text{Art}\\ \text{NP}\end{Bmatrix}\right)$ (AP)* N (PP)*($\overline{S}$)
NP → Pronoun
VP → V (NP) (NP) (PP)* ($\overline{S}$)
VP → V AP
PP → P NP
$\overline{S}$ → (Comp) S

NP → NP (Conj NP)*
VP → VP (Conj VP)*
PP → PP (Conj PP)*
S → S (Conj S)*
$\bar{S}$ → $\bar{S}$ (Conj $\bar{S}$)*
V → V (Conj V)*

14. Another transformation: Topicalization

English has a number of transformations similar to Wh- Movement. Perhaps the simplest is the so-called Topicalization rule, used to account for sentences like these:

Linguistics, I can teach.
Those guys we would never give our credit cards to.
In that oven you should never put a chicken.

The name of the rule is from that fact that the fronted NP or PP serves as the "topic" of its sentence; what it is about. These sentences have a distinctly rhetorical character, and often sound best if you imagine that the topic is being contrasted with some other topic:

Postmodernism, I'm clueless about, but linguistics, I can teach.

The "landing site" for fronted topics appears *not* to be Comp, since you can get both *that* and the fronted topic in sequence:

I'd say <u>that</u> <u>linguistics</u>, I can teach.

Thus we should state the rule as simply moving a phrase to the left edge of S, as follows:

Topicalization

W	[	X	NP	Y	]$_S$	Z	→	W	[	NP	X	t	Y	]$_S$	Z				
1		2	3	4		5		1		3	2	3	4		5				

"Move an NP to the beginning of S."

Topicalization, like Wh- movement, appears to be unbounded, though the examples that show this tend to be a bit less natural:

John, I don't think a lot of people would like.
Fred, I'd imagine that you'd think that a lot of people wouldn't like.

The justification for Topicalization is much the same as that for Wh- Movement: the presence of a topicalized element is correlated with a subcategorization gap later in the sentence.

Study Exercise #22

Derive *I'd say that linguistics, I can teach*.

This area is intensionally blank.

15. It-Clefting

We will cover one more long-distance movement rule, one which was briefly discussed above in Chapter 4, section 11, under the name "Clefting". In this context we will use its more specific name, **It Clefting**.[61] The sort of data that justify the rule are given below.

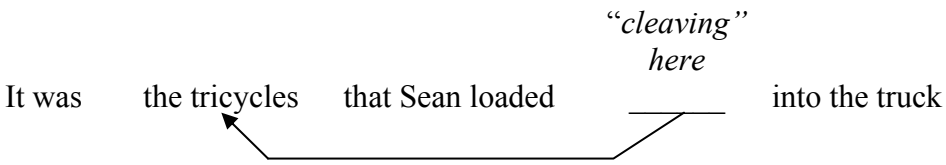
a. *Sean loaded the tricycles into the truck.*

It was Sean that loaded the tricycles into the truck.
or: It was the tricycles that Sean loaded into the truck.
or: It was into the truck that Sean loaded the tricycles.
or: It was the truck that Sean loaded the tricycles into.

b. *Fred thinks Alice climbed up Mt. Everest.*

It is Fred that thinks Alice climbed up Mt. Everest.
or: It is Alice that Fred thinks climbed up Mt. Everest.
or: It is up Mt. Everest that Fred thinks Alice climbed.
or: It is Mt. Everest that Fred thinks Alice climbed up.

The idea is that the transformation "cleaves" the sentence, by moving one of its constituents into a high clause containing *it* plus *BE*. Intuitively it works like this:



The second through fifth sentences in each group are all clearly related to the first sentence, and can be derived from it with a transformation.

The transformation needed is pretty complicated and so I give below first an abbreviated version showing the essentials, then a more careful version, and lastly a prose version.

It Clefting

a. Abbreviated version:

$$[_S\ X \begin{Bmatrix} NP \\ PP \end{Bmatrix} Y\ ] \quad \rightarrow \quad [_S\ it\ be \begin{Bmatrix} NP \\ PP \end{Bmatrix} [_{\bar{S}}\ that\ [_S\ X\ t\ Y\ ]]]$$
$$1 \quad 2 \quad 3 \qquad\qquad 2 \qquad\qquad 1\ 2\ 3$$

[61] The other kind of clefting in English is often called the "wh- cleft", and is found in sentences like *What Bill needs is a vacation*; "vacation" is clefted here.

b. Full version:

$$[_S\ X\ \begin{Bmatrix}NP\\PP\end{Bmatrix}_{2}\ Y\ ]_S \quad \rightarrow \quad [_S\ [_{NP}\ [_{Pro}\ it\ ]\ ]\ [_{VP}\ [\ be\ ]_V\ \begin{Bmatrix}NP\\PP\end{Bmatrix}_{2}\ [_{\overline{S}}\ that\ [_S\ X_1\ t\ Y_3\ ]]]]]$$

c. In words:

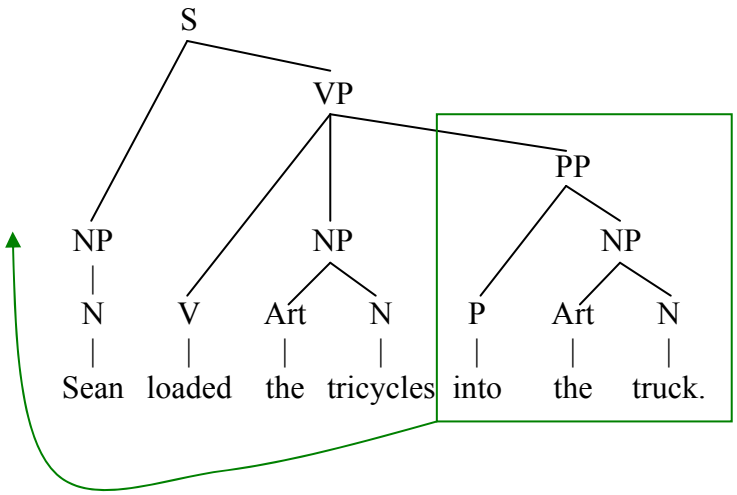
- Take a Sentence containing an NP or PP.
- Construct "on top" of it a new Sentence, with subject *it*, verb *be*, and an $\overline{S}$ sister to V within its VP. The new $\overline{S}$ contains the original S.
- Move the NP or PP to a position following *be*.

Even this is not quite enough; we need to specify how a Tense feature gets into the morphosyntactic representation of the *be*. We could do this by copying the morphosyntactic representation of the original main verb onto the inserted *be*. We'll skip this part in the interest of brevity.

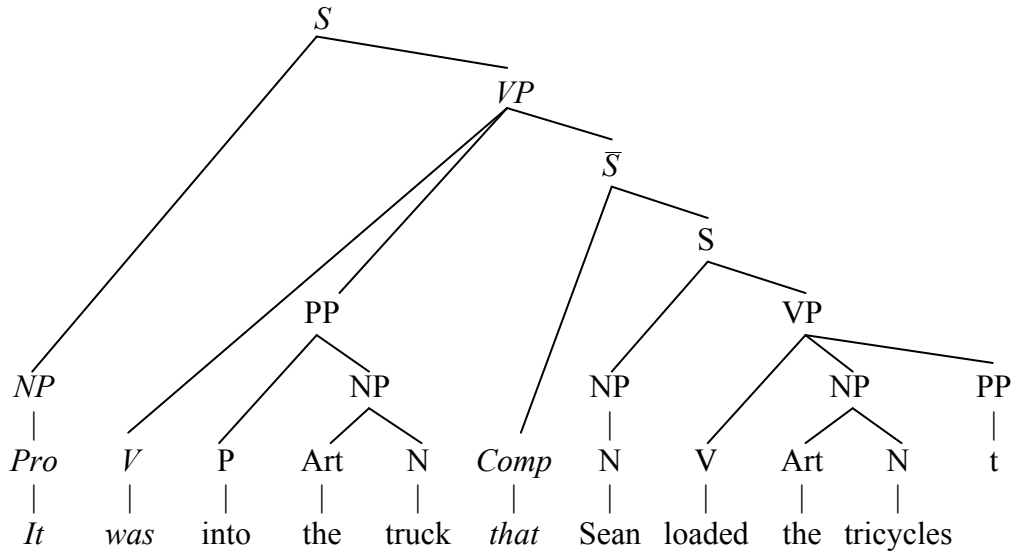
It Clefting is another instance of an unbounded dependency, and for the same reason as in Wh- Movement requires a movement analysis.

Here is an example of how It Clefting applies. Structure added by the rules is shown in italics.

Deep structure and movement:



Surface structure, with trace:



Like Wh- Movement and Topicalization, It Clefting is an unbounded rule:

It was [the tricycles]$_{NP}$ that Tom thinks Sue knows that Bill loaded ___ onto the truck.

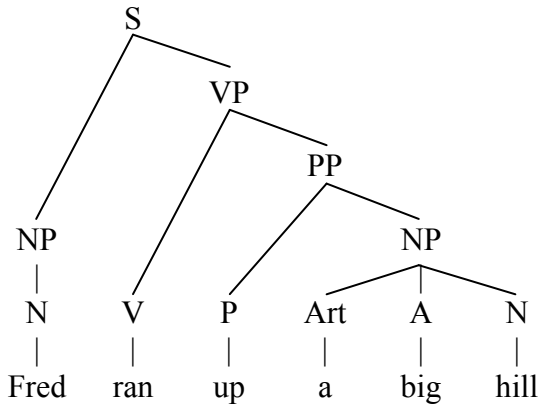
16. Constituency testing

As noted in Chapter 4, section 11, It Clefting can be used as a **constituency test**. Here is the rationale: the rule requires there to be an NP or PP to move; so if you're not sure whether a particular sequence of words is an NP or a PP, try It Clefting it. (The other transformations also can be used, but It Clefting often seems to work smoothest, since the grammaticality judgment for It Clefting seem to be less sensitive to other factors.)

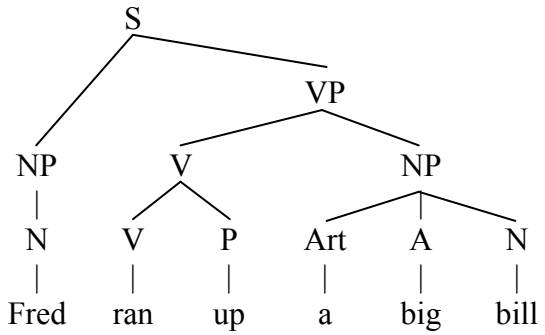
One example is the following pair:

Fred ran up a big hill.
Fred ran up a big bill.

The correct structures are evidently as follows:



(Prepositional phrase *up a big hill*, indicating the direction Fred ran)



The structure here has a compound verb of the structure V + P. *A big bill* is the object of this verb, indicating what Fred ran up.

If we do constituency testing, we ought to find that:

- *up a big hill* is a constituent (PP)
- *a big hill* is a constituent (NP)
- *up a big bill* is **not** a constituent
- *a big bill* is a constituent (NP)

Applying It Clefting to the relevant bits of both sentences, we get:

It was <u>up a big hill</u> that Fred ran.	(Fine, clefting PP)
It was <u>a big hill</u> that Fred ran up.	(Fine, clefting NP)
*It was <u>up a big bill</u> that Fred ran.	(Bad, result of trying to It-Cleft a nonconstituent)
It was <u>a big bill</u> that Fred ran up.	(Fine, clefting NP)

As you can see, the pattern of grammaticality perfectly matches the constituency that we proposed—the result is grammatical only when a constituent is moved.

Constituency testing is commonly used in syntactic investigation, particularly for nonobvious cases. The most common other constituency test other than movement is probably conjunction (X and Y), since only constituents join together in conjoined structures. Thus, a parallel test for the same sentences would be:

Fred ran up a big hill and down a big mountain.	(Conjoining two PP's)
Fred ran up a big hill and a big mountain.	(Conjoining two NP's)
*Fred ran up a big bill and over a cat.	(Conjoining nonconstituents)[62]
Fred ran up a big bill and a big tab	(Conjoining two NP's)

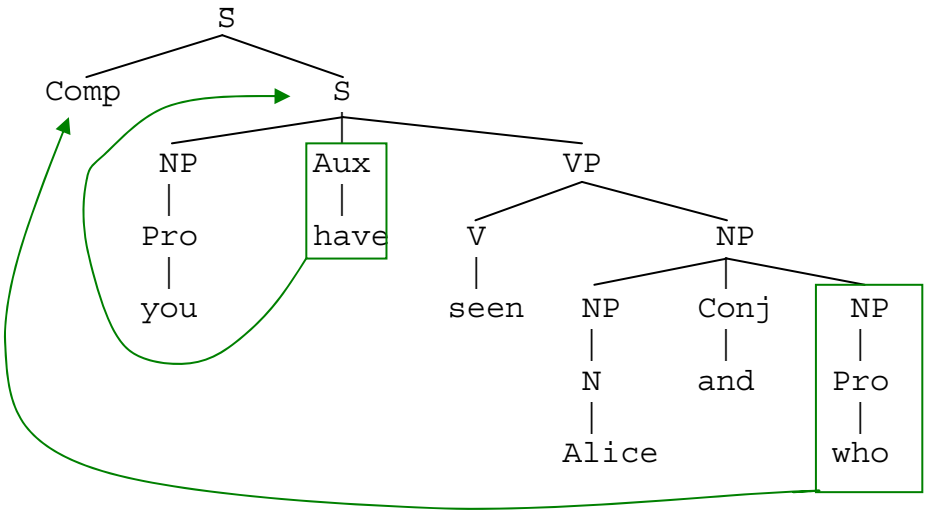
17. Island constraints

Consider the following deep structure:

You have seen Alice and who?

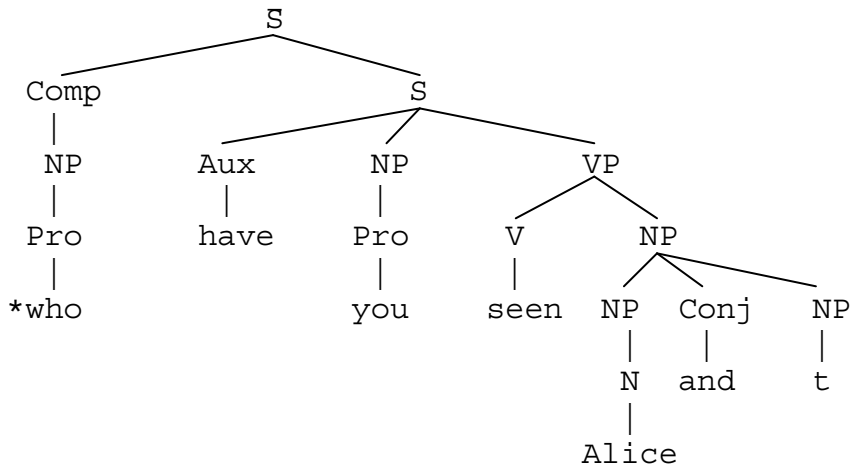
This is clearly a possible deep structure, as it can be an echo question if nothing applies to it. However, if make it part of an $\overline{S}$, in anticipation of making it into a Wh- question (see previous notes for why), and then apply Wh- movement to this deep structure, the result is unexpectedly ungrammatical:

Deep structure, Subject/Aux Inversion, and Wh- Movement:



[62] If the example isn't clear to you, imagine: "Fred ran up a big bill and (then) ran over a cat (with his car)."

Surface structure:



Note that we really are dealing with ungrammaticality rather than nonsense; the question is perfectly reasonable, and could mean roughly 'Who did you see Alice with'.

The ungrammaticality of this sentence is a serious problem for our grammar: we have applied the rules of the grammar in a perfectly legitimate way, but have derived a bad result. Here are some further data of the same sort:

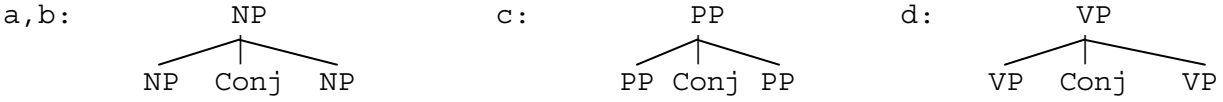
a. You have seen who and Alice (okay as echo question)
   *Who have you seen and Alice?

b. Bill will take pictures of Fred and Alice (not the same deep structure, but close enough)
   *Who will Bill take pictures of and Alice?
   *Who will Bill take pictures of Fred and?

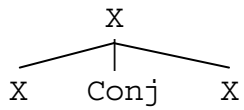
c. Jay jumped onto the trampoline and into the pool.
   *What did Jay jump onto the trampoline and into? (moving NP)
   *What did Jay jump onto and into the pool? (moving PP)

d. Phil loves Coke and abhors Pepsi.
   *What beverage does Phil love Coke and abhor?
   *What beverage does Phil love and abhor Pepsi?

The generalization here is that Wh- Movement produces an ungrammatical result if it tries to move a wh- phrase outside a structure in which two constituents are joined by a conjunction. In the four groups the sentences above, the structures are as follows:



Structures of this sort are called **coordinate structures**. A general notation for coordinate structures is as follows:[63]



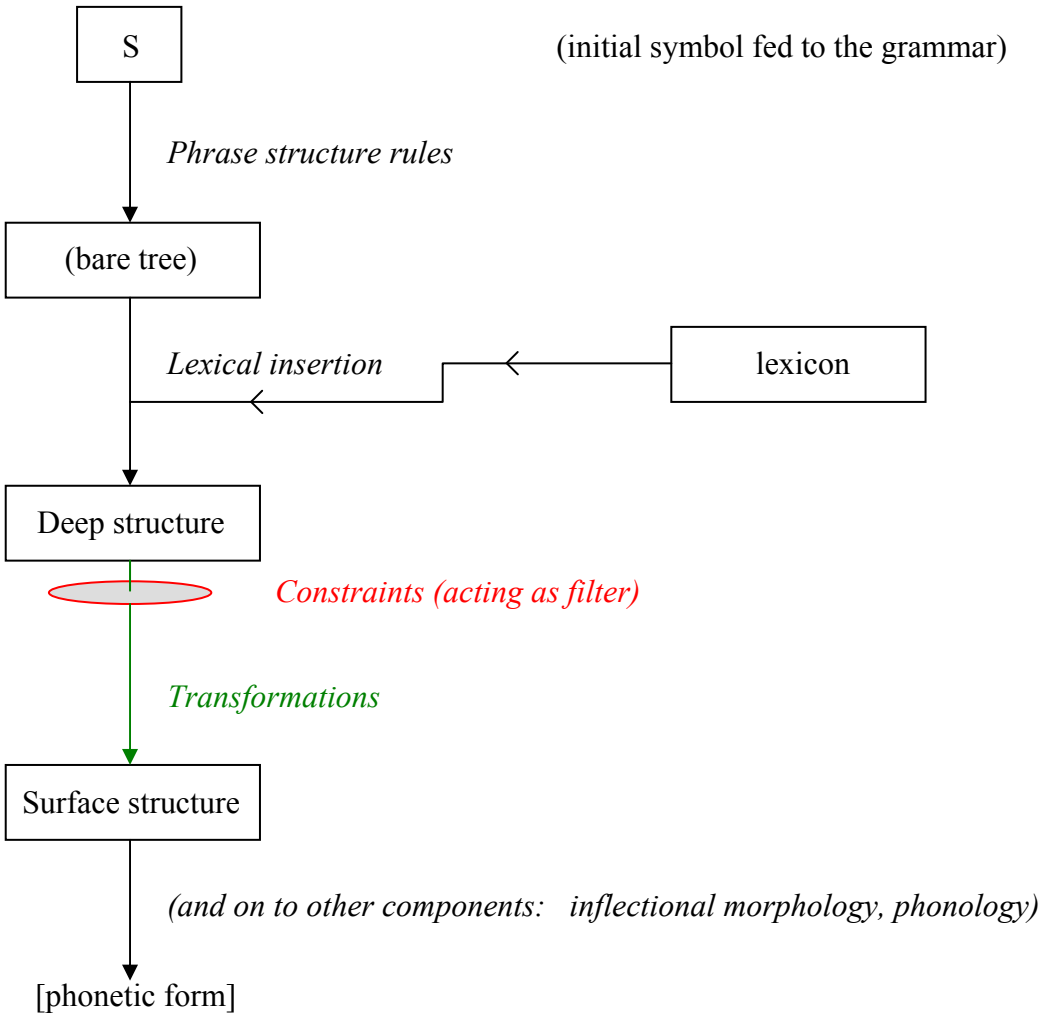
The next step is to fix the grammar so that it will no longer generate sentences in which a coordinate structure has been extracted from. The most obvious move would be to add a complication to the Wh- movement rule that would simply block the rule from doing this. However, we will see later on that all the other long-distance transformations are blocked in exactly the same way. If we added exactly the same complication to all the other rules, we would be missing a generalization.

18. Constraints in grammar

A better solution is to add to grammar a **constraint** on possible derivations. A constraint could be thought of as a "filter" on the operation of the grammar: if the derivation of a sentence violates the constraint, then the constraint marks the sentence as ungrammatical, and it is eliminated from the (infinite) set of sentences that the grammar generates.

The general organization of such a grammar can be imagined as follows:

[63] We could generalize this to cover multiple conjuncts (NP *and* NP *and* NP...), but won't take the time.



This conception includes three of (what I take to be) the four basic formal mechanisms of linguistic theory: (a) generative rules (here, phrase structure rules); (b) transformations (converting one structure to another); (c) filters (throwing out the result of a derivation).[64]

For the data under discussion, the constraint we need is the following:

[64] The fourth rule type is interpretation, which we will cover when we get to semantics.

Coordinate Structure Constraint

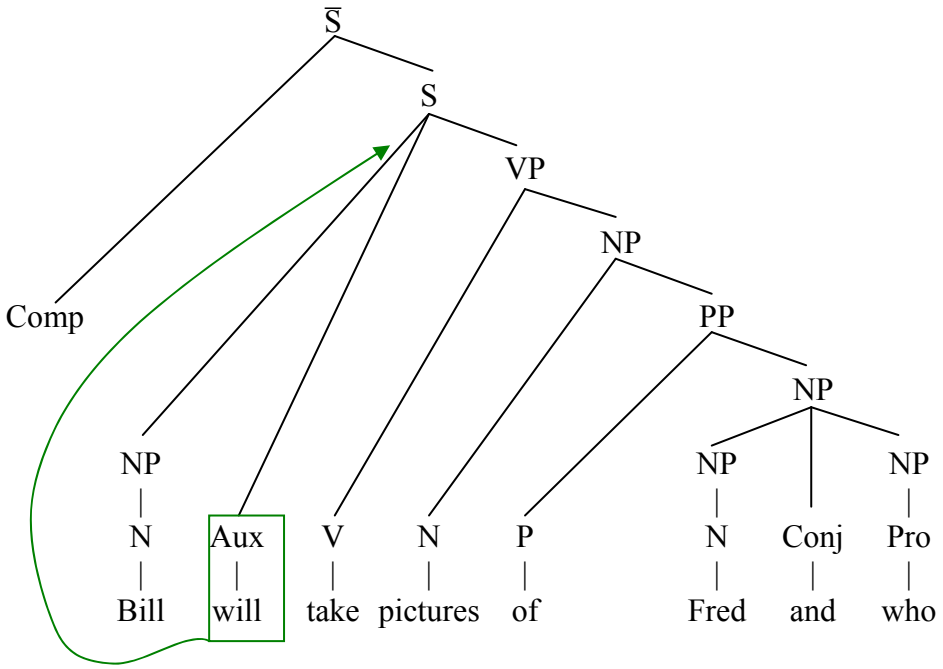
Mark as ungrammatical any sentence in which a constituent has been extracted from inside a coordinate structure.

The notation of the triangle seen here means, "any structure dominated by X".

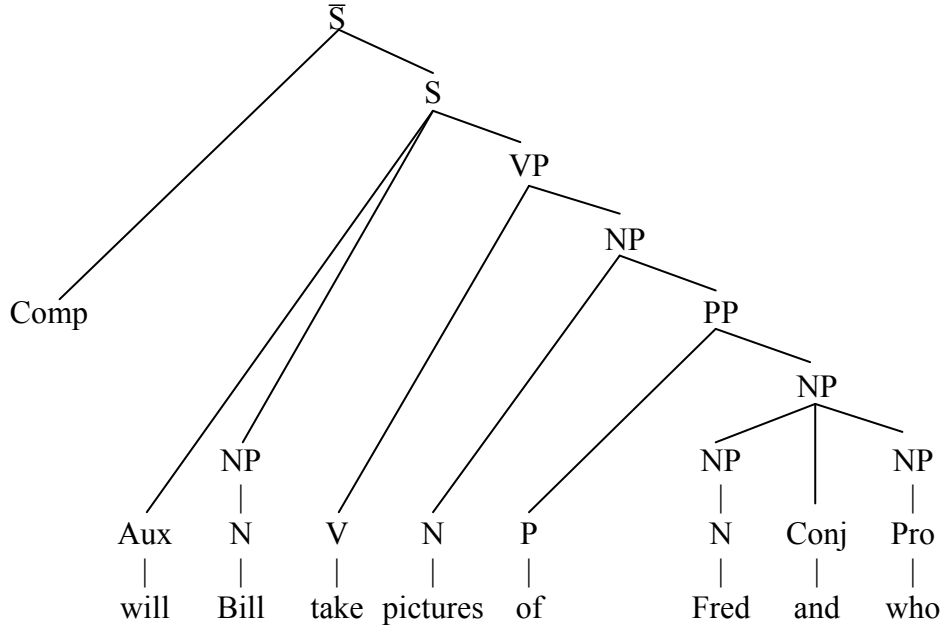
The Coordinate Structure Constraint is called an **island constraint**. A coordinate structure acts as an "island," in that it is inaccessible to the efforts of transformations to remove things out of it. The Coordinate Structure constraint will correctly rule out the ungrammatical sentences given earlier.

Here is one way (which you can use on the homework) to demonstrate how a constraint works: you draw the deep structure of a sentence, outline the constituent that moves, outline the island that contains it, and draw an arrow showing that the movement does indeed move a constituent outside of the island. (One also adds an arrow, to indicate that this movement results in ungrammaticality.)

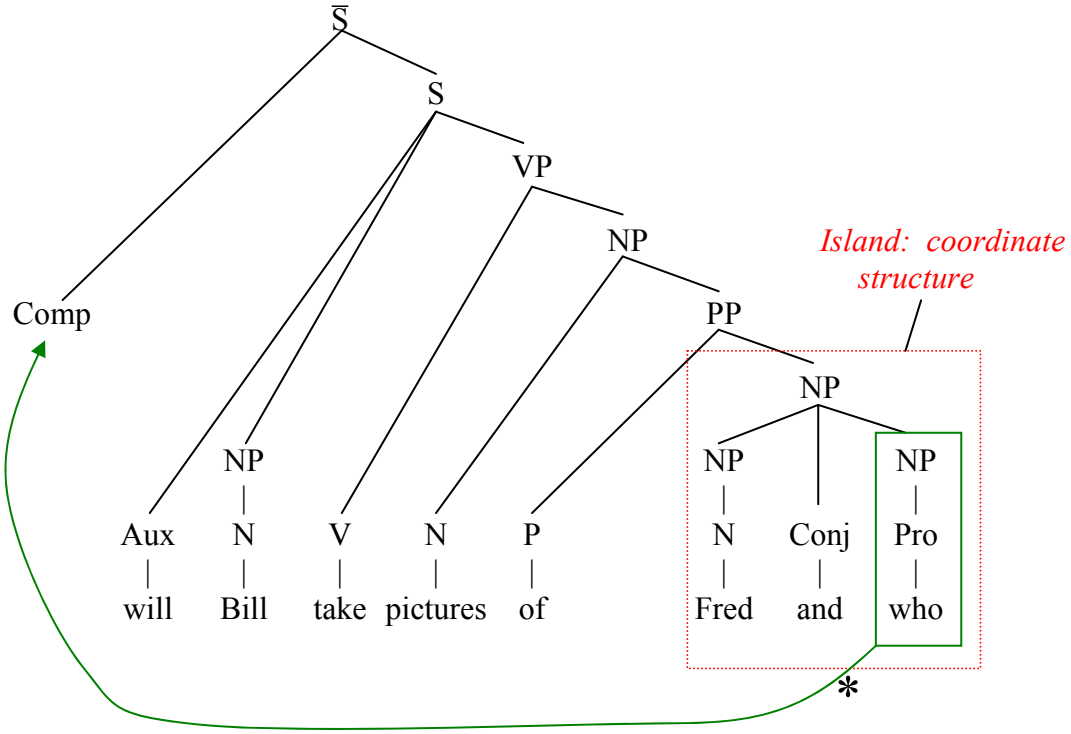
Deep structure (empty Comp is the landing site for Wh- Movement)
Arrow shows application of Subject/Aux Inversion.



Output of Subject/Aux Inversion:



2nd transformation (attempted): Wh-Movement:



Result: *Who will Bill take pictures of Fred and?

Study Exercise #23

Explain why the sentence

*What city have the Romans destroyed and attacked Athens?

is ungrammatical. Illustrate with a derivation.

This page is intensionally blank.

This area is intensionally blank.

Study Exercise #24

Explain why the sentence 'What city have the Romans attacked and destroyed?' is grammatical. Illustrate with a derivation. Hint: recall that V → V Conj V is a phrase structure rule of English.

This area is intensionally blank.

Study Exercise #25

Explain why the sentence ‘Which city and which province will the Romans destroy?’ is grammatical. Illustrate with a derivation.

This area is intensionally blank.

19. The universality of the Coordinate Structure Constraint

One of the goals sought by linguistics in writing formalized grammars is to locate universals of language. A linguistic universal is a property shared by all human languages. The explanation of linguistic universals is one of the central tasks of linguistic theory.

Linguistic universals are proposed and tested against data from the languages of the world; there are thought to be about 8000 of them.[65] No universal has been checked against all 8000, however, at least some proposed universals

Some universals that have been proposed are fairly superficial, for example:

- All languages have nouns and verbs.
- All languages have wh- questions.
- All languages have consonants and vowels.
- All languages use the vowel [a][66] or something phonetically close to it.

[65] This number is, alas, declining steadily; probably the best list of languages is the *Ethnologue*, at http://www.ethnologue.com/.

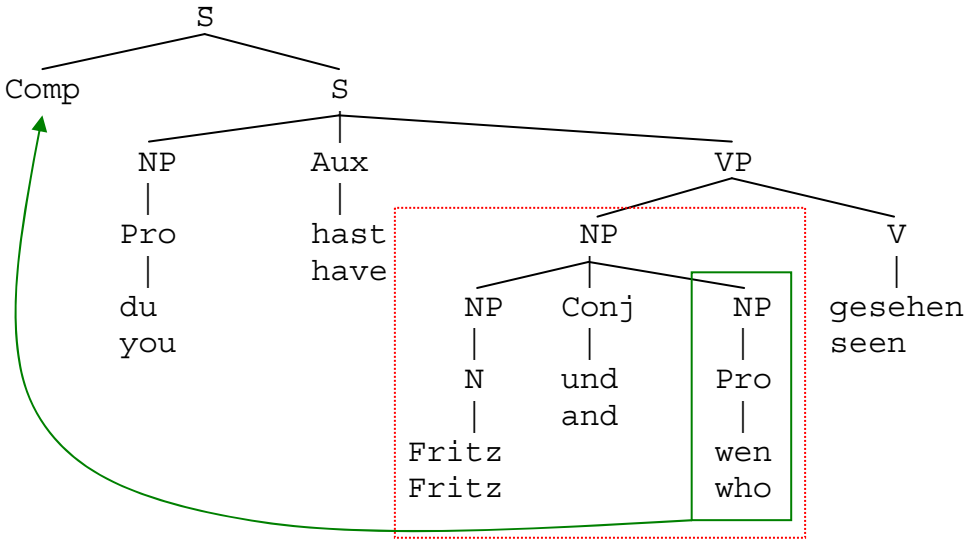
[66] IPA [a] is more or less the [a] vowel of Spanish, or in some dialects of English the vowel of *hot*.

Others are more subtle, and emerge only when we have submitted a large number of languages to formal analysis—that is, have constructed grammars for them.

As you might expect, it is common for linguists to propose universals, then be forced to abandon or modify their proposal in the face of falsifying evidence. This is only natural, and indeed it is often felt that the job of the linguist is to be a bit "out on a limb", creating hypotheses about language that are interesting enough to be worth checking.

The Coordinate Structure Constraint was first noticed and proposed as a universal by the linguist John R. Ross, who pioneered the study of syntactic islands in the mid 1960's. The phenomenon of islands attracted a great deal of attention and has been extensively studied and analyzed since then. Today, there seems to be a consensus, based on study of a fair number of languages, is that the Coordinate Structure Constraint is universal. (The doubtful cases are instances in which we're not sure that the structure in question is really a coordinate structure.) To be more precise: in all languages that can be tested (because they have wh- movement; *in situ* languages don't count), extraction from coordinate structures is impossible. Here are some sample data from other languages:

German	Simple sentence:	Du hast Fritz gesehen you have Fritz seen 'You have seen Fritz.'
	Wh- question (with gap):	Wen hast du ___ gesehen? who have you (gap) seen
	Extraction from coordinate structure:	*Wen hast du Fritz und ___ gesehen? who have you Fritz and (gap) seen *'who have you seen Fritz and?'



French: *Quelle journal as- tu lu ___ et ce livre?
what newspaper have you read (gap) and this book

*Quelle livre as- tu lu ce journal et ___?
what book have you read this newspaper and (gap)

Russian: *Kovo Ivan videl Petra ili ___?
who Ivan saw Peter or (gap)
*'Who did Ivan see Peter or?'

*Kovo Ivan videl ___ ili Petra?
who Ivan saw (gap) or Peter

Formal universals like the Coordinate Structure Constraint have inspired a fair amount of theorizing about language and language learning, which we'll take on in the next chapter.

19.1 Looking ahead

The combination of long-distance transformations and islands leads to a fair amount of analytical work: for each transformation, you want to show that it respects the island, and for each island (we'll cover more) you want to show that the transformations all respect it.

20. The transformations restated intuitively

In previous readings, I've given formalized versions of the three transformations Wh- Movement, Topicalization, and It-Clefting. Here are informal working versions, which may be useful for problem solving:

Subject-Aux Inversion

In a yes/no question or a Wh- question, move the highest Aux in the sentence to the left of the subject, keeping it a daughter of S, and leaving no trace.

Wh- Movement
Move a wh- phrase into Comp, leaving a trace.

Topicalization
Move an NP or PP to the beginning of S, making it a daughter of S, and leaving a trace.

It-Clefting
Add *it be that* to the top of the sentence, giving it the expected parse (new S, $\overline{S}$, NP, Pro, V, and Comp). Then move some NP or PP from the original sentence so it comes after *be*. Leave a trace.

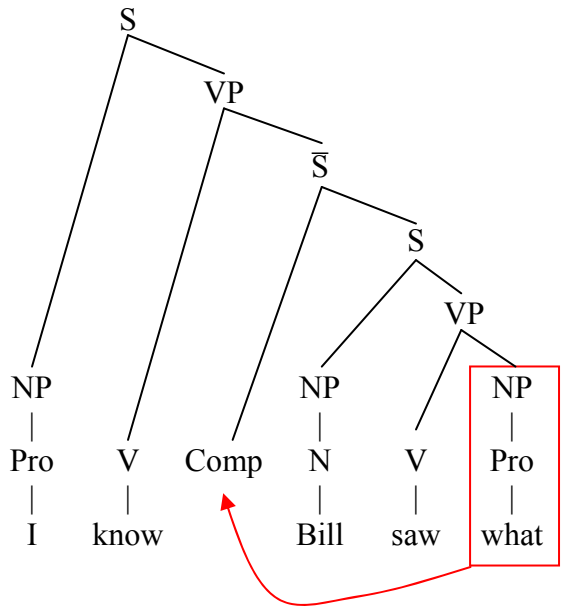
21. Two more islands

To complete our general account of long-distance transformations and island constraints, here are two more islands (there are quite a few more, varying from language to language, but we will stick with just three total). The general point that emerges is that all the long-distance transformations obey all of the island constraints (since there will be three of each, we will need to check a total of nine cases).

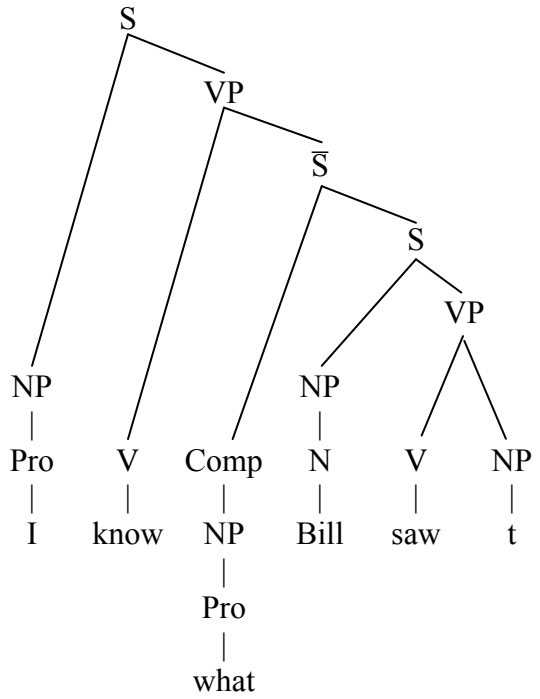
21.1 The Wh- Island Constraint

Recall embedded wh- questions, like *I know what Bill saw*. We already have the means to derive this, and the example is reviewed below.

Deep structure and Wh- Movement:



Surface structure:



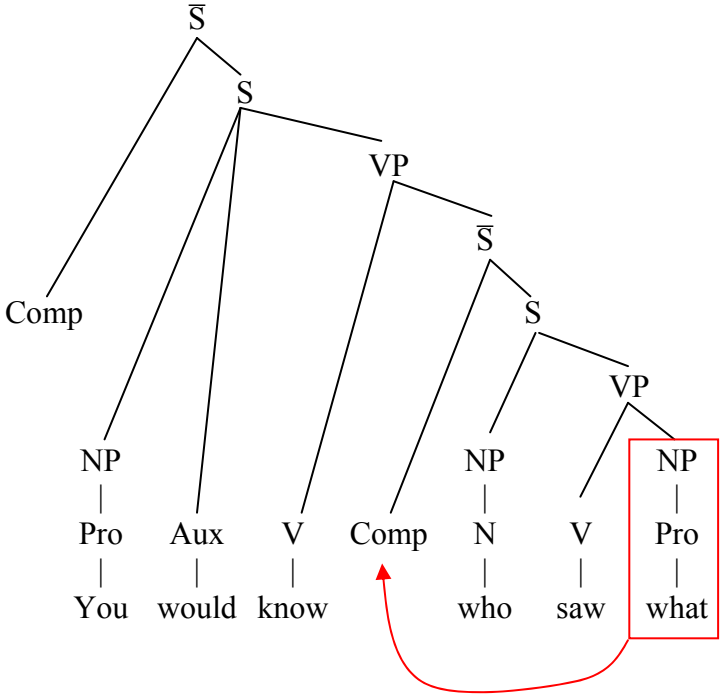
But now consider the following scenario: what if, at the level of deep structure, there were *two* wh- phrases in the same clause? This is not so absurd, since we actually have sentences like the following:

I know [who saw what]$_{\bar{S}}$

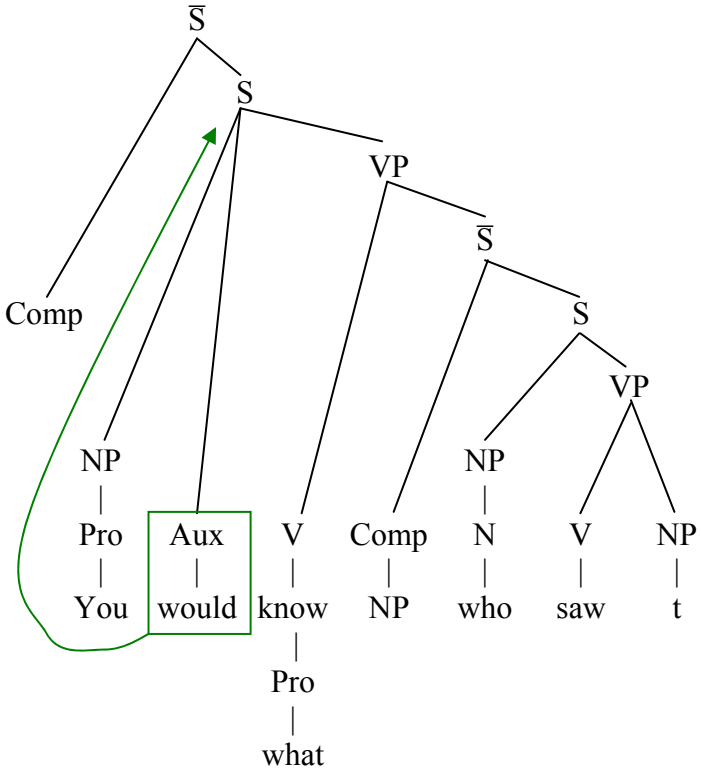
Here, the wh- word *what* remains in situ, as the object of *say*. We won't be able to cover here just what circumstances permit a wh- phrase to remain in situ in English, but for now this sentence suffices to show that it is quite possible to have a clause with two wh- phrases.

Now, consider this scenario: we take the above sentence as a deep structure, move *what* into the "lower" Comp, and *who* into the "higher" Comp, as follows:

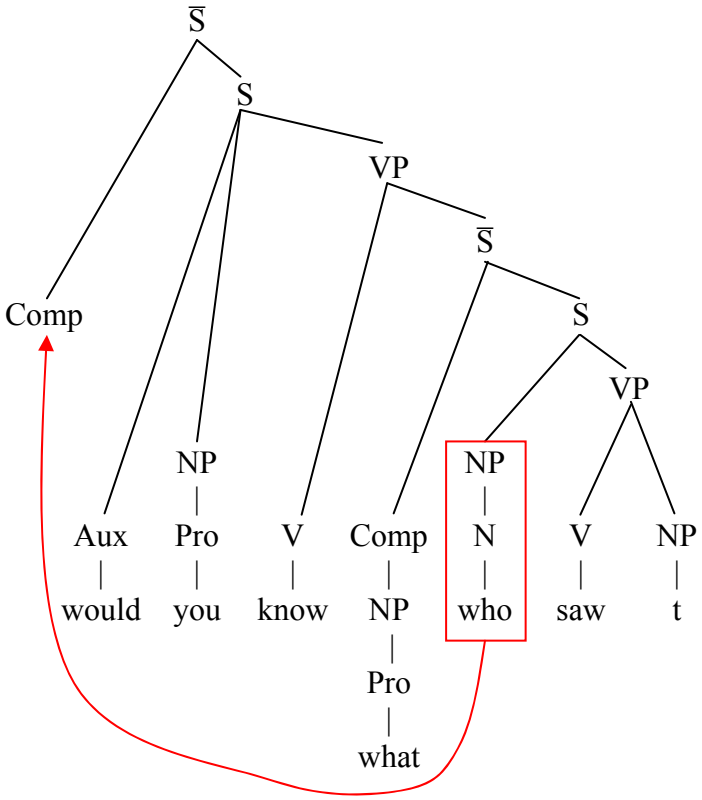
Deep structure and first application of Wh- Movement:



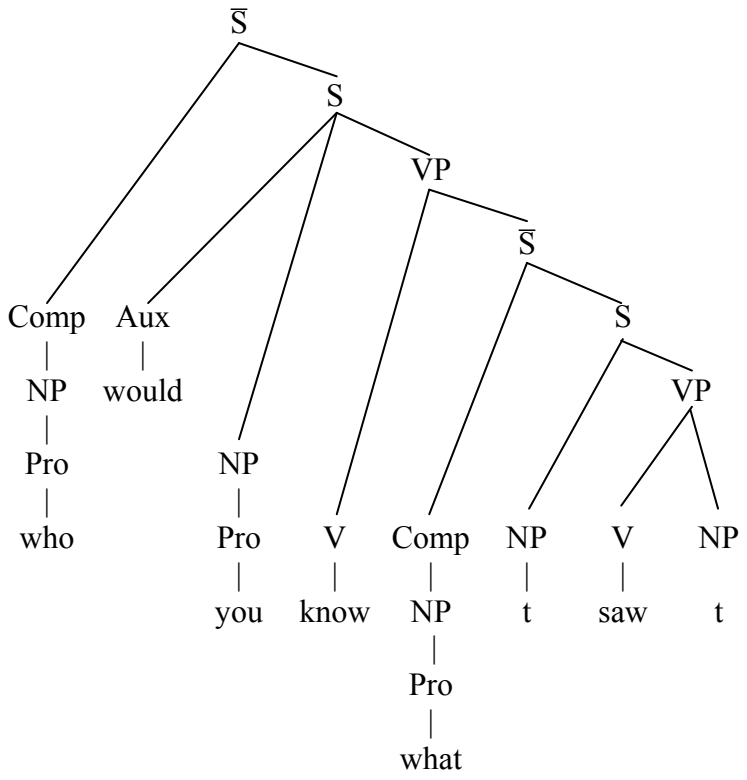
Result of first application of Wh- Movement, with arrow showing Subject-Aux inversion:



Result of Subject-Aux Inversion, showing second application of Wh- Movement:



Surface structure:



The result is **Who would you know what saw*?, which most speakers find crashingly bad. It is worth emphasizing that this is not due to its lacking a meaning; it's clear that it should mean the following:

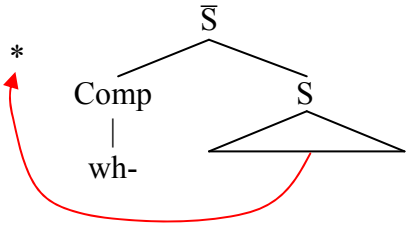
"What is the person such that you know what that person saw?"

The meaning is hard to access, given the extreme ungrammatically of the sentence.

Linguists have proposed to explain the ill-formedness of sentences like *Who do you known what saw*? by positing yet another island, along the following lines:

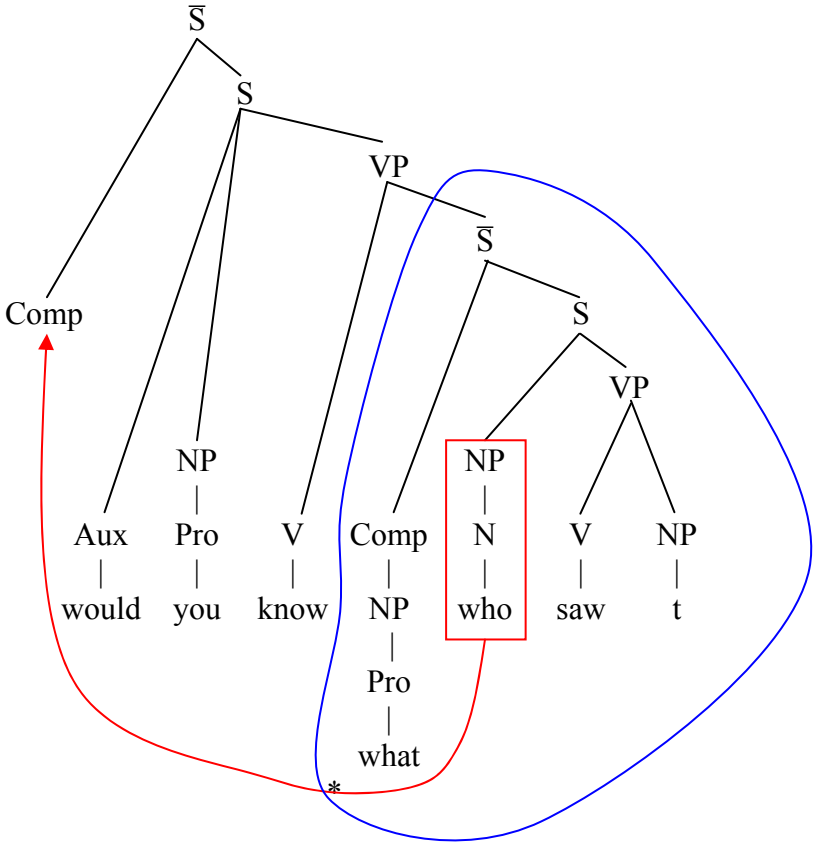
Wh- Island Constraint

Mark as ungrammatical any sentence in which a constituent has been extracted from inside an $\overline{S}$ whose Comp contains a wh- phrase.



This island constraint is slightly different from the Coordinate Structure Constraint, because the island is actually created by a transformation. The "lower down" Wh-Movement forms an island that blocks any further Wh- movements higher up in the tree.

To illustrate: returning to the derivation given above, but this time drawing in the island, we can see that it is correctly excluded by the Wh- Island Constraint. The sequence *what who saw* is covered by the description of the island, and thus the sentence is ruled out.



The Wh- Island constraint covers a fair amount of data; here are some other sentences that it excluded. I've put brackets in to illustrate the $\overline{S}$ that begins with a wh-phrase and thus forms a Wh- Island.

Bill admitted who committed this crime.
*[What crime] did Bill admit [who committed ___]$_{\overline{S}}$?

Fred doesn't care how long you take on this exam.
*[Which exam] doesn't Fred care [how long you take on ___]$_{\overline{S}}$?

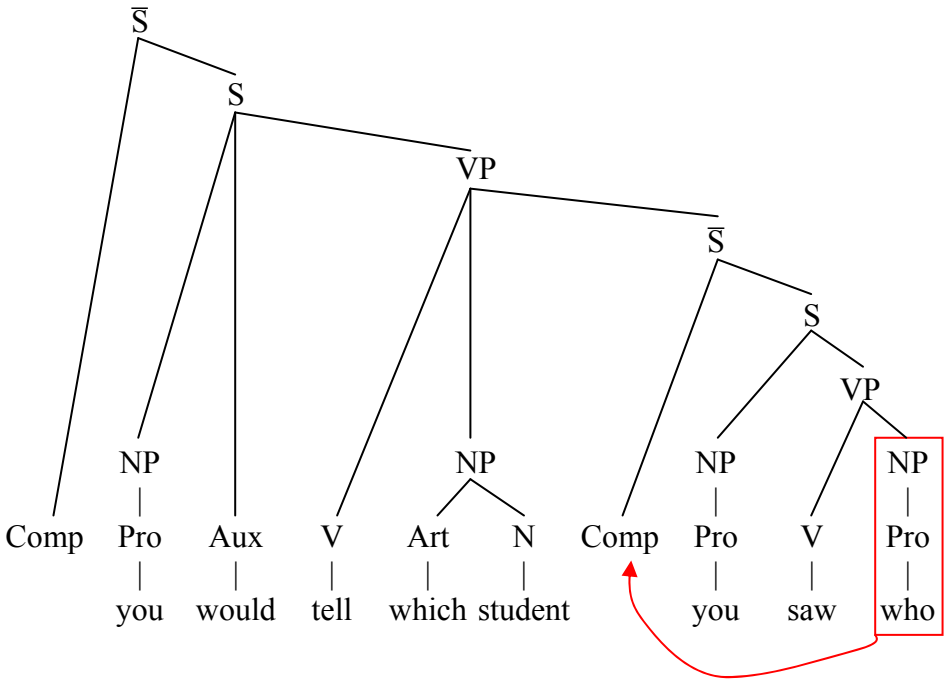
Alice doesn't care which exam you take a long time on.
*[How long] doesn't Alice care [which exam you take ___ on]$_{\overline{S}}$?

Observe further that there is nothing wrong with having two wh- phrases in the same sentence. It's only when one wh-phrase is moved *out of the $\overline{S}$ that the other one begins* that you get a bad result. Here is an example. In the sentence

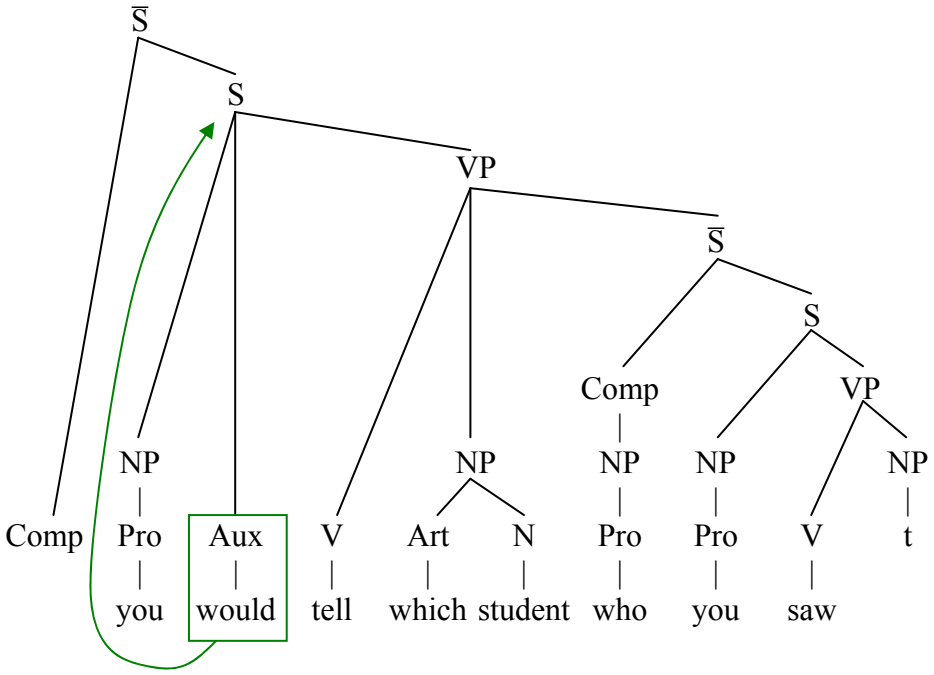
Which student would you tell who you saw?

the two instances of wh- movement are non-overlapping. The movement that goes to the higher Comp is *not out of the island*, so the sentence comes out fine. Here is the full derivation.

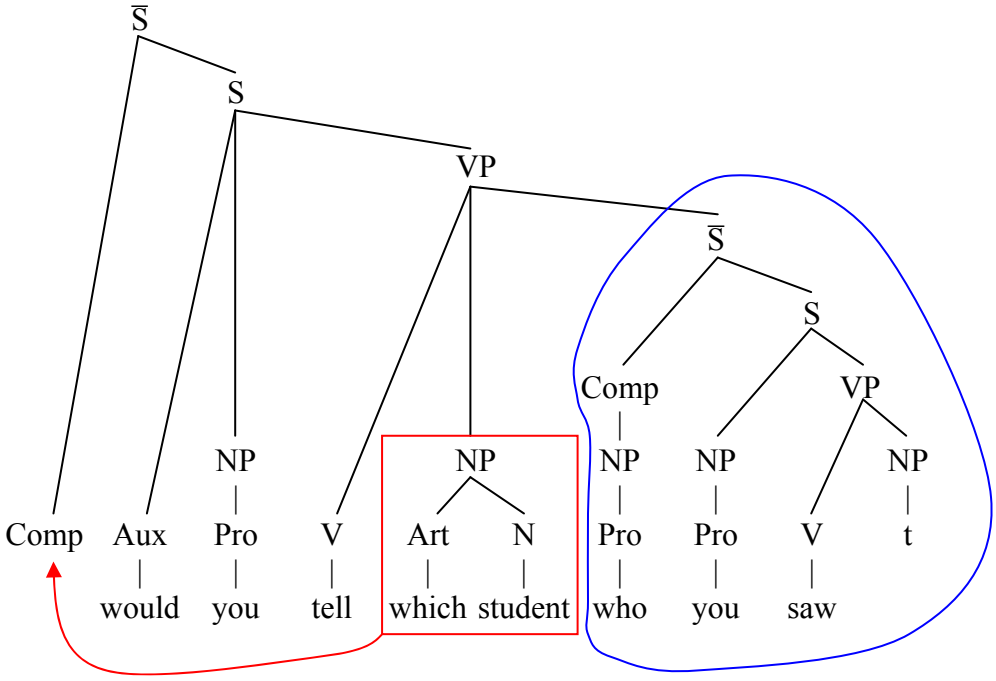
Deep structure with lower wh- movement:



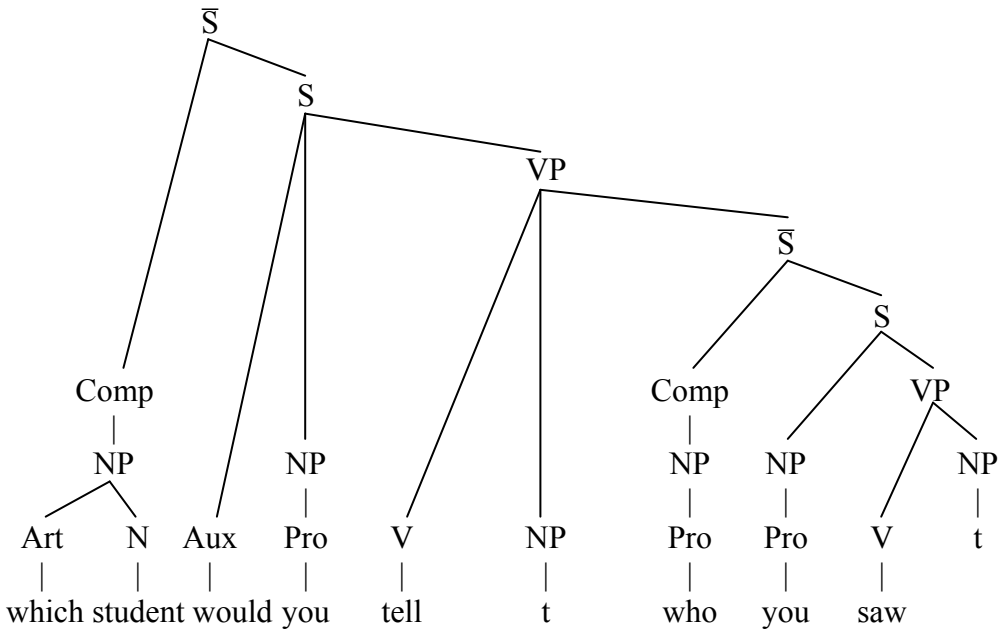
Output of lower wh- movement, showing Subject-Aux Inversion:



Output of upper wh- movement. Note that this is *not* movement outside of the wh-island, shown in red.



Surface structure:



You can see this all at once if we put the material on just one line, showing only the two instances of Wh- Movement and the island:

[Who] would you tell ___ [[who] you saw ___]$_{\bar{S}}$?

Note finally that Topicalization and It-Clefting obey the Wh- Island Constraint:

Topicalization:

*[Kate], I realized [to whom they would send ___]$_{\bar{S}}$.

It-Clefting:

*It was [Oliver] that I wondered [which book ___ would read]$_{\bar{S}}$.

21.2 The Complex NP Constraint

Another kind of island is the so-called "complex noun phrase". Recall the main phrase structure rule in English for NP, the one to which we have just added a possible $\bar{S}$ daughter:

$$NP \rightarrow (\{^{NP}_{Art}\}) (AP)^* \; N \; (PP)^* \; (\bar{S})$$

A complex NP is an NP having $\overline{S}$ as a daughter (there may also be other modifiers). You get a complex NP if you include the boldface items below in applying the rule.

$$NP \rightarrow (\{\begin{matrix}NP\\Art\end{matrix}\}) (AP)^* \mathbf{N} (PP)^* (\mathbf{\overline{S}})$$

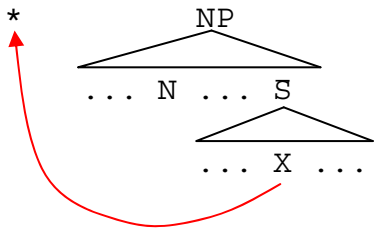
Some examples of complex NP's include

Sue's belief that Sam is leaving Omaha
Alice's hunch that the burglar used this window
Bill's inane hypothesis that Frieda saw Jack[67]

The island constraint for complex NP's, called the Complex NP Constraint, is stated as follows:

Complex NP Constraint

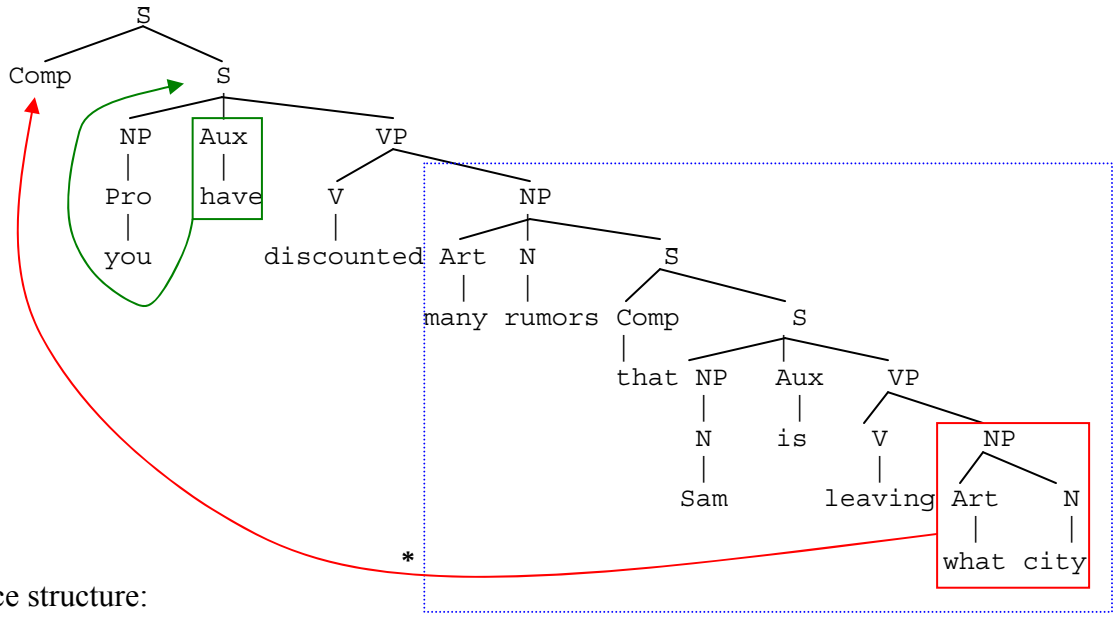
Mark as ungrammatical any sentence in which a constituent has been extracted from inside a complex NP.



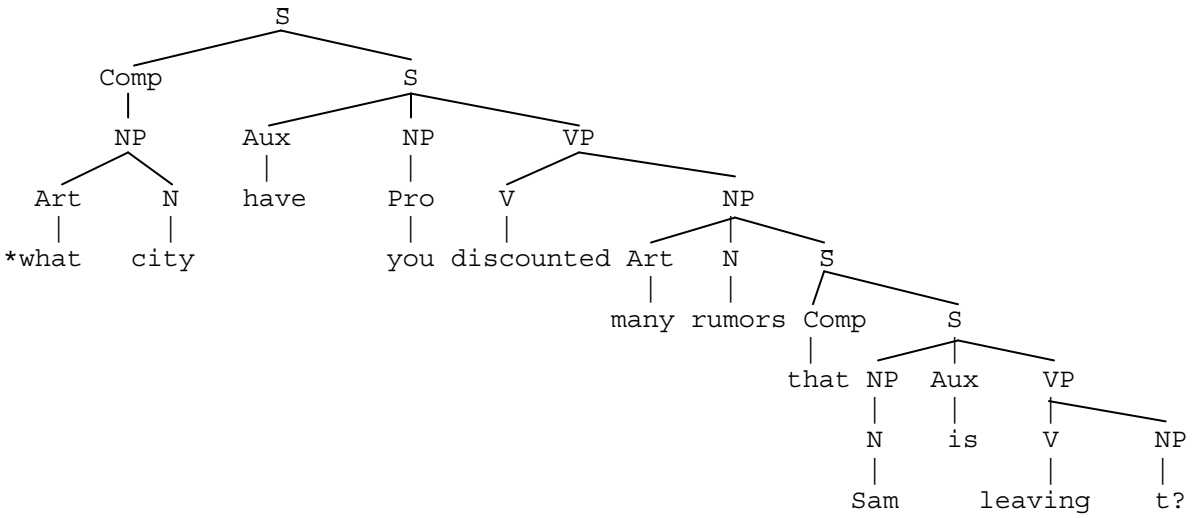
To demonstrate that complex NP's are islands, one does the following. (a) Set up a deep structure that contains a complex NP; (b) make sure that in this deep structure, there is a wh-phrase contained within the complex NP; (c) apply Subject Aux Inversion and Wh- Movement to the deep structure and see if the result is grammatical. I have done this in the following example. The Complex NP is circled, and the arrows show what moves where.

[67] The relative clauses mentioned above (p. 150) are also islands; for instance: **What apples will you see the man who picked ___?*

Deep structure, Subject-Aux Inversion, and Wh- Movement:



Surface structure:



The fact that the surface structure is ungrammatical supports the existence of the Complex NP Constraint. Similar ungrammatical sentences would be

*Which window would you disagree with Alice's hunch that the burglar used?
*Who might you hear Bill's inane hypothesis that Frieda saw?

Study Exercise #26

Give a derivation, with boxes, arrows for movement, and a circled island, for the two sentences just given.

This area is intensionally blank.

This page is intensionally blank.

This area is intensionally blank.

Study Exercise #27

Why is the sentence

Whose theory that Sam is crazy could you believe?

grammatical? (You have to imagine a scenario in which all sorts of people are presenting theories that Sam is crazy.)

This area is intensionally blank.

21.3 The Complex NP Constraints and other syntactic transformations

As the following labeled sentences show, the Complex NP Constraint is obeyed by Topicalization and by It-Clefting:

Topicalization

*[Kate], I discounted [many rumors that they would elect ___]$_{NP}$.

It-Clefting

*It was [Kate] that I discounted [many rumors that they would elect ___]$_{NP}$.

21.4 The complex NP constraint is not universal

When it was noticed and first formalized by the linguist John R. Ross in the 1960's, it was thought that the Complex NP Constraint is a linguistic universal, just like the Coordinate Structure Constraint is. Shortly thereafter, however, Scandinavian linguists began studying the island constraints of their native languages, and noticed that neither Norwegian nor certain dialects of Swedish and Danish respect the constraint. The linguist Jens Allwood offers the following data from Swedish,[68] which he checked with a number of speakers; the complex NP is shown in brackets

Simple sentence:

Herodes	levde	i	[hopp-et	om	att	Salome	skulle	förföra	den	mannen.]$_{NP}$
Herod	lived	in	hope-the	of	that	Salome	should	seduce	that	man

'Herod lived in the hope that Salome should seduce that man.'

Wh- Movement out of Complex NP:

[Vem]	levde	Herodes	i	[	hopp-et	om	att	___	skulle	förföra	den	mannen]$_{NP}$?
Who	lived	Herod	in		hope-the	of	that		should	seduce	that	man

[Vem]	levde	Herodes	i	[	hopp-et	om	att	Salome	skulle	förföra	___	]$_{NP}$?
Who	lived	Herod	in		hope-the	of	that	Salome	should	seduce		

Topicalization out of Complex NP:

[Salome]	levde	Herodes	i	[	hopp-et	om	att	___	skulle	förföra	den	mannen.]$_{NP}$
Salome	lived	Herod	in		hope-the	of	that		should	seduce	that	man

[Den	mannen]	levde	Herodes	i	[	hopp-et	om	att	Salome	skulle	förföra	___]$_{NP}$.
That	man	lived	Herod	in		hope-the	of	that	Salome	should	seduce	

This is unusual; most languages that have these rules do respect complex NP's. Thus, here are some French data, very much like English:

*[Qui]	as-	tu	proposé	[l'idée	que	Marie	aime ___]$_{NP}$?

'Who have you suggested the-idea that Marie loves (gap)?

[68] The reason for the Biblical subject matter is not clear to me. You can make up example sentences about whatever you like, of course.

22. Why these islands?

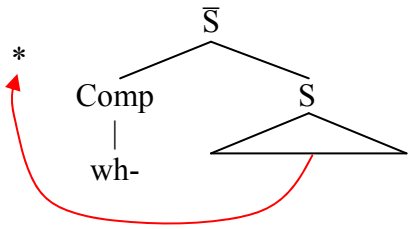
To review, we have discussed three separate island constraints:

Coordinate Structure Constraint

Mark as ungrammatical any sentence in which a constituent has been extracted from inside a coordinate structure.

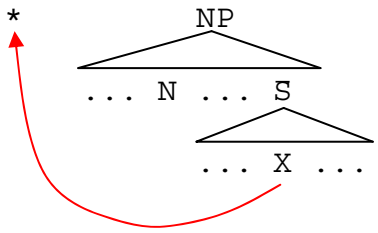
Wh- Island Constraint

Mark as ungrammatical any sentence in which a constituent has been extracted from inside an $\overline{S}$ whose Comp contains a wh- phrase.



Complex NP Constraint

Mark as ungrammatical any sentence in which a constituent has been extracted from inside a complex NP.



Of these, the Coordinate Structure Constraint seems to be a good candidate for being a linguistic universal; the other two are probably not universal but seem to be found in many languages.

The question that arises when one lines up the islands in a row like this is: "**Why these islands**?" That is, why should island-hood be found for just this particular

configuration of syntactic structures? The three islands seem to have little in common with each other.

The view of most linguists who consider this question is that the islands as formulated above are a first-pass approximation. That is, it's a good idea to formulate the islands in this way, for the sake of explicitness of analysis, but in the long term it seem desirable to seek more abstract principles to explain the data.

One approach that seems fruitful is to invert the problem: one specifies what places it is *legal* to extract from rather than what places it is illegal. You may encounter approaches of this type if you study syntax in future course work.

A final point is that the islands may be in some sense useful to the speakers. Psycholinguistic experimentation (including with brain-scanning devices) suggests there is a cognitive burden for the listener whenever the sentence heard involves a filler-gap constructions such as those created in the transformations described here. When a language has island constraints, they in effect tell the language user, "don't bother to look for gaps here."—perhaps this reduces the burden on speech perception, and thus reflects a principle of good "language design".

Chapter 9: Semantics

1. Goals of semantics

Semantics is the branch of linguistics that studies meaning, particularly meaning as it is conveyed by language. We can start out by asking what meaning is.

Meaning is a characteristic of symbolic systems; language is by far the most elaborate and powerful symbolic system that has ever been found. Our sentences are complex symbols, physically realized in speech or writing, which bear meanings and thus express our thoughts.

Clearly, there is more to thought than the language that expresses it. Thought can exist in the absence of language, since many animals can behave in a sophisticated and rational fashion without having anything like human language.[84] It also seems clear that we experience thought in ways that are very direct and not linguistic.[85] There is no need for thought to occur in a linear sequence, as our words must; and moreover that our *visual* thoughts are not particularly expressible in language.

The development of a theory of thought is at present an active but speculative activity, involving psychologists, philosophers, cognitive scientists, and scholars in the field of artificial intelligence. One vindication for a proposed theory of thought would be if it could be embodied in a system that could think and reason like a person. This remains a distant goal.

Our focus in semantics is not quite as grand; we just want to know how language expresses thought. The problem faced by semanticists is to study the ways in which language embodies thought, without a well-developed theory of thought to go by. This problem has not stymied research, however, because there are plenty of ways to conduct careful research that don't require a full theory of thought to make progress. For instance, one strategy that has been followed (it originates in the field of philosophy) is to develop formal systems that determine the **truth conditions** of sentences (properties of the world that must hold for sentences to be true), often in a small, artificially-constructed world. This kind of approach requires a fair amount of development and will not be taught here; instead, in the interest of a unified course I want to cover aspects of semantics that interact most closely with syntax.

2. Some aspects of linguistic meaning

We will cover three aspects of linguistic meaning.

- predicate-argument structure

[84] A book on this topic I have enjoyed, written from a sober but exploratory viewpoint, is *Animal Minds*, by Donald Griffin (University of Chicago Press, 1992).

[85] It's probably unnecessary to give an example, but for what it's worth: imagine a parent who sees his toddler put in danger from an unleashed dog: the experience is direct and primal, and depends in no way on an internalized utterance "That dog is a threat to my child" or the like. The dog, the child, the teeth are all part of the thought, but the thought is probably complete before it is ever embodied in NP's, VP's, etc.

- anaphora
- operators and scope

3. Predicate-argument structure

A **predicate-argument structure** is a kind of semantic representation that limits itself to depicting "who is doing what to whom." For example, in a sentence like

John cooked the egg.

an act of cooking is described. We could characterize this act with the following predicate-argument structure.

COOK (($_{\text{Cooker}}$ John), ($_{\text{Cook-ee}}$ the egg))

In this structure, COOK is a "predicate", which has "arguments", in this case filling the slot of Cooker and Cook-ee. The labels for the argument slots are arbitrary, and in fact I will sometimes be choosing slightly silly ones, simply because they are short and clear.[86]

Predicate-argument structure contains both more and less information than a syntactic tree.

Predicate-argument structure contains **more** information than syntactic structure because it identifies the argument slots for each participant in the action. Syntactic structure instead places the participants (designated by NP's) in syntactic roles, such as subject (daughter of S) or object (daughter of VP)—and, as we will see, there is only a loose connection between thematic roles and syntactic roles.

Predicate-argument structure contains **less** information than syntactic structure for various reasons. Most notably, predicate-argument structure is not meant to convey any sense of linear order; COOK "has" the two arguments given, but there is nothing in the thought being expressed that requires this order. The order that appears on the page is selected purely for convenience.

More generally: linear order is a property of language, not of thought. Different languages have idiosyncratic orders, including all six logically possible orders for simple two-NP sentences like *John cooked the egg*. Here are all six, with examples of each.

SVO	English, Spanish, Swahili
SOV	Japanese, Korean, Turkish
VSO	Tagalog, Biblical Hebrew, Irish
VOS	Fijian
OSV	Xavante (Brazil)
OVS	Hixkaryana (Brazil)[87]

[86] More ambitious theories try to *generalize* over slots, with widely-applicable terms. For instance, Agent is used for any slot occupied by an entity that controls the action, Theme is used for objects in motion, and so on. This kind of generalization is not agreed upon by all linguists and will not be used here.

[87] The last three of these are rare.

The point is that languages do use linear order to convey predicate-argument structure (as we saw in studying phrase structure rules, but that is a matter of grammar, and not of thought.

3.1 Variation in how syntax manifests predicate-argument structure

Very often, the grammars of languages offers more than one syntactic means of expressing the same predicate-argument structure. A well-known example of this kind is the **passive** construction, found in many languages:

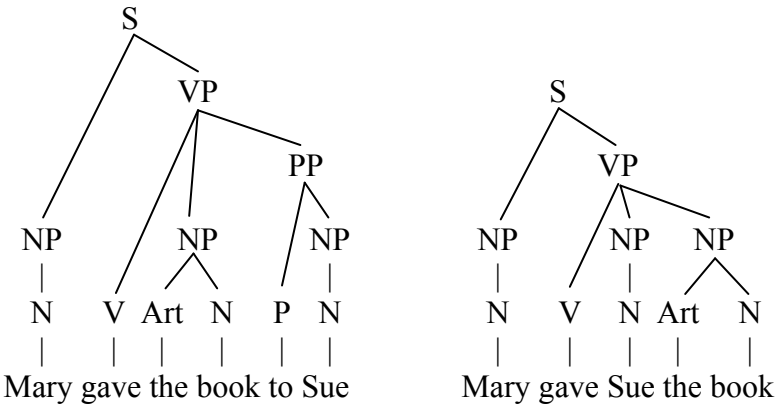
The doctor examined John.
John was examined by the doctor.

The first of these sentences is said to be in the "active voice" and the second in the "passive voice." For both sentences, the predicate-argument structure is something like this:

EXAMINE (($_{\text{Examiner}}$ Doctor), ($_{\text{Examinee}}$ John))

Why might languages offer more than one way to connect up the thematic roles with the grammatical positions? One view is that these variations are related to **discourse structure**: when we converse or tell a story, we are not producing sentences in isolation; rather, each sentence builds on a body of information that already exists and adds a new bit of information. Quite often, the subject NP embodies the pre-existing information, and the VP is what adds something new. Thus, *The doctor examined John* is most naturally used where one is already talking about the *doctor*, and *John was examined by the doctor* is most naturally used when one is already talking about *John*. Thus, the passive construction permits the speaker to organize information in a dialogue or narrative in a coherent way that builds on older information, by making the old information the subject.

Here is another instance in which the same predicate argument structure has more than one syntactic expression. It and occurs with verbs of giving. Here is an example:



The first tree illustrates the NP PP construction, in which the item given is the NP object and the recipient is in the PP. The second tree illustrates the NP NP construction, in which the

recipient is the first NP and the item given is the second NP. Both have the same predicate-argument structure:

GIVE (($_{\text{Giver}}$ Mary), ($_{\text{Gift}}$ book), ($_{\text{Recipient}}$ Sue))

As with passive, the variation seems to be related to the form of a discourse: the first sentence would be more natural when one is already talking about *the book*, the second would more natural when one is already talking about *Sue*. As with passive, the new information comes later in the sentence.

3.2 Unfilled argument slots

Passive sentences often lack a PP with *by*. Such sentences plausibly involve an argument slot that goes unexpressed, along these lines:

John was examined.
EXAMINE (($_{\text{Examiner}}$ ∅), ($_{\text{Examinee}}$ John))

Here, zero with a slash means "unspecified". That is, clearly there was someone who did the examining, but the sentence does not say who. German allows this sort of construction even when the verb is intransitive:

Es wurde getanzt.
It was danced
'There was dancing, people danced.'

DANCE (($_{\text{Dancer}}$ ∅))

3.3 Propositions as filling argument slots

The following sentence has a predicate-argument structure in which one of the participants is a Proposition — depicting an event.

Mary had John cook the egg.
CAUSE (($_{\text{Causer}}$ Mary), ($_{\text{Proposition}}$ COOK (($_{\text{Cooker}}$ John), ($_{\text{Cook-ee}}$ the egg))))[88]

To treat such a case, we need a kind of nested structure, similar to the multi-clause structure of syntax. In this sentence, Mary, the agent, caused the state of events described in Proposition to come into being.

Here is another sentence whose predicate-argument structure involves a proposition, here, the content of Mary's thoughts:

[88] In this and some later predicate-argument structures, I've used color to make sure that brackets match up correctly. For correct bracket structure: every argument is surrounded by parentheses, and every list of arguments is surrounded by parenthesis (even if there is just one argument).

Mary thinks that Bill jumped.
THINK ((Thinker Mary), (Proposition JUMP ((Jumper Bill))))

3.4 Cases of mismatch between syntax and predicate-argument structure

Consider the following sentence, shown with a proposed predicate-argument structure.

It rained.
RAIN

What is special about such a case is that there are no arguments—*raining* is a thing that just happens (nobody rains!).[89] The *it* we get in syntactic structure is meaningless, and is evidently present simply to satisfy the grammatical requirement (S → NP VP) that sentences must have subjects. Such semantically empty elements are a mismatch between syntax and predicate-argument structure. They illustrate that syntax involves demands of "pure form" that have nothing to do with expression.

The *it* that occurs as the subject of *rain*, *snow*, etc. is sometimes called "weather *it*".

Here is another such case:

It seems that Mary gave the book to Sue.
SEEM((Proposition GIVE((Giver Mary), (Gift book), (Recipient Sue))))

Here again we have a semantically empty *it*, present to give the main clause a subject. This *it* is sometimes called **pleonastic** it.[90]

A related construction gives the main clause a subject by taking the logical subject of the embedded clause and expressing it "in the wrong position":

Mary seems to have given the book to Sue.
SEEM ((Proposition GIVE((Giver Mary), (Gift book), (Recipient Sue))))

In this grammatical construction, often called "Subject Raising", the NP *Mary* occurs in a syntactic location that is intuitively "higher" than its location in predicate-argument structure.

3.5 Predicate-argument structure in linguistic theory

There are two possibilities for integrating predicate-argument structure into linguistic theory. One possibility is to find a set of rules that inputs syntactic trees and derives the predicate-argument structure from them. Another approach that has been taken is to let the predicate-

[89] Observe that this is different from *John was examined* and *Es wurde getanzt*, discussed above. Some really did examine John, and someone really did dance, but no one rains.

[90] "Pleonastic" comes from the Greek for "superfluous"; the *it* is felt to be somehow unnecessary (though it's necessary for the sentence to be grammatical!)

argument structure be the starting point of the derivation—embodying the message the speaker wishes to communicate—and let the grammar find an appropriate tree structure or structures for communicating this message. We will not pursue this question any further here.

Study Exercise #28

Give predicate-argument structures for the following sentences. Be brave about labeling the argument slots; in a problem set we would necessarily be flexible about the grading of such labels.

a. John appears to have been given a book by Sue.
b. It seems to have rained.
c. It seems that it rained.
d. It is felt that Bill rants.

Study Exercise #29: The predicate-argument structure of *As*-phrases

The particle *as* has interesting syntactic and semantic behavior, in which the phrase structure again mismatches the semantics. Some sample sentences:

0. We consider him as being eccentric.
1. They regard him as praiseworthy.
2. We judge it as unfortunate that he visited Mary.
3. We regard him as appearing to be sick.

We could accommodate *as* phrases with the following phrase structure rules:

$$\text{VP} \rightarrow \text{V (NP) } \textit{As}\text{P } (\overline{\text{S}})$$

$$\text{AsP} \rightarrow \textit{as} \begin{Bmatrix} \text{NP} \\ \text{PP} \\ \text{VP} \end{Bmatrix}$$

Furthermore, we must add rules of inflectional morphology that would ensure that the VP that is part of an *as*-phrase, the verb is marked to be a present participle (V-*ing*). Only a few verbs such as *regard* and *consider* subcategorize for *as*-phrases.

What is interesting semantically is that as-phrases express propositions without including any $\overline{\text{S}}$. For example, in the first sentence above, *we* are not doing anything to *him*; rather, we are holding a belief about him, that is, we are the mental experiencers of a proposition involving him. This idea could be expressed with the predicate-argument structure below:

CONSIDER (($_{\text{Experiencer}}$ we) ($_{\text{Proposition}}$ ECCENTRIC (($_{\text{Characterized}}$ he)))

The proposition is, essentially, "he is eccentric", without any verb or $\overline{\text{S}}$ encoding this proposition.

Assign predicate-argument structures to sentences #1-3 above. Note that the *it* in #2 is pleonastic.

ANAPHORA

4. Defining Anaphora

All languages have pronouns. For example, these are the pronouns of English in their various forms:

Nominative		*Objective*		*Genitive*	
I	we	me	us	my	our
you	you	you	you	your	your
he/she/it	they	him/her/it	them	his/her/its	their

Predicative	*Genitive*[91]	*Reflexive*	
mine	ours	myself	ourselves
yours	yours	yourself	yourselves
his/hers/—	theirs	himself/herself/itself	themselves

Pronouns are like nouns, but they get their reference from context—either the linguistic context, or the situational context of speech. As already noted, the English pronouns are distinguished by morphosyntactic features of Number, Case, and Person, and in the third person, for gender. Their meanings are determined entirely by these features.

There are also pro-forms for other parts of speech. The phrases *do it* and *do so* are pro-forms for Verb Phrases:

I wanted to [teach Linguistics 865]$_{VP}$ but was too busy with other courses to [do so]$_{VP}$.
I had to [teach Linguistics 497]$_{VP}$ because no one else would [do it]$_{VP}$.

Thus is a somewhat archaic proform for Adverb Phrases:

He did it thus.

[91] Used after *be*, as in *It is mine*. There is no 3rd pers. singular inanimate form; for example, you can't say **That fuel pump is its*, referring to a particular car.

The term **anaphora** refers, in linguistics, to the process whereby one word or phrase gets its reference from the meaning of another phrase; thus in:

Bill thinks he's a genius.

we say that *he* makes anaphoric reference to *Bill*; likewise, above *do so* makes anaphoric reference to *teach Linguistics 865*.

5. The Pronominalization Hypothesis and why it fails

A tempting analytical option, since we need transformations anyway, is to suppose that pronouns are the result of applying a "Pronominalization" transformation.

Consider a sentence like

Alice told Sue that she was a genius.

Here, the pronoun *she* can refer either to *Alice* or to *Sue*. The sentence is therefore ambiguous. The Pronominalization theory would say that when *she* means *Alice*, then the deep structure is:

Alice told Sue that Alice was a genius.

and analogously when *she* means *Sue*, then the deep structure is:

Alice told Sue that Sue was a genius.

We can then adopt a transformation that converts the second instance of two identical noun phrases into the appropriate pronoun.

Pronominalization

X	NP	Y	NP	Z	→	X	NP	Y	Pro	Z
1	2	3	4	5		1	2	3	4	5

where X stands for any sequence of words, and Pro retains its original values for the morphosyntactic features [Number], [Animacy], and [Gender].

It is easy to see that the Pronominalization transformation will convert the two deep structures

Alice told Sue that Alice was a genius

and

Alice told Sue that Sue was a genius

into the same surface structure, namely

Alice told Sue that she was a genius

If this aspect of meaning is read off of deep structure, then we have accounted for the ambiguity of the sentence. Let us call this the Pronominalization Hypothesis.

Although this proposal looks reasonable, in fact it suffers from several problems. First, there are sentences in which the deep structure that the Pronominalization Hypothesis provides doesn't mean what we want it to. If all pronouns are derived from full noun phrases, then the deep structure of

Everyone thinks he is a genius

would be

Everyone thinks everyone is a genius.

But this deep structure obviously means something quite different from the surface structure. We'll return to this problem later on.

A second problem with the Pronominalization Hypothesis is that it can't cover all of the cases at hand. There instances in which one uses a pronoun without even *knowing* the full noun phrase that is its antecedent. The following example was invented by the linguist Howard Lasnik. Imagine a cocktail party at which a man arrives, a stranger to all, who starts drinking heavily and getting into heated, unpleasant discussions with all he encounters. After an hour of unpleasantness, he storms out of the room, slamming the door behind him. At this point one could, without knowing the man's name, say:

Well, he's left.

The point is that if *some* pronouns are interpreted as referring to a salient person in the context (that is, the pragmatic, real-life context), then we should consider the possibility that even the *she* in *Alice thinks she's a genius* is similarly interpreted—*Alice* is a plausible person for *she* to refer to, since, after all, we're talking about her.

A final problem with the Pronominalization Hypothesis is that, curiously enough, it appears to lead us to infinite deep structures.[92] Here is an example:

The girl who deserves it will get the prize she wants.

This sentence contains two pronouns, *it* and *she*. According to the Pronominalization Hypothesis, we can get the deep structure by replacing these pronouns with copies of the full NP's to which they refer. Doing this yields:

[92] The problem was noticed in the 1960's by Emmon Bach, of the University of Massachusetts and Stanley Peters of Stanford, and is sometimes called the Bach-Peters paradox.

> The girl who deserves [the prize she wants] will get the prize [the girl who deserves it] wants.

But this sentence also contains pronouns! Thus, to arrive at the true deep structure we will have to substitute for these as well:

> The girl who deserves [the prize [the girl who deserves it] wants] will get the prize [the girl who deserves [the prize she wants]] wants.

And we are still not done, so:

> The girl who deserves [the prize [the girl who deserves [the prize she wants]] wants] will get the prize [the girl who deserves [the prize [the girl who deserves it] wants]]] wants.

No matter how long we keep going, we are still going to have uninterpreted pronouns in our representation, so it's clear that this process is never going yield an interpreted representation. The upshot is that "spelling out" pronouns as their full noun phrases does not seem promising as an account of their semantics.

6. Interpretive rules

Given what we've just seen, one might think that the right way to handle the meaning of pronouns would be just to let them be pronouns; that is, nouns whose meaning is determined by referring to a salient (highly noticeable) entity in the context (either linguistic context, or real-life context), which matches the requirements of number (*she* vs. *they*), gender (*she* vs. *he*), and animacy (*she* vs. *it*). In this approach, interpreting pronouns is relegated largely to the domain of thought, not language—pretty much every sentence would be interpreted the way we interpret the sentence *Well, he's left* given above.

This is an appealingly simple theory, but it likewise cannot work. Research on the possibilities of how pronouns refer has shown that there is a heavy *linguistic* contribution to their interpretation.

Consider the following very simple sentence:

John likes him.

Fluent speakers of English will assert pretty firmly that *him* cannot refer to *John*, even though there is no logical reason why it could not. Similar sentences are:

He likes John.
He likes John's brother.
He thinks John is a genius.

The reason why *he* cannot refer to *John* in these sentences turns out, as we'll see shortly, to be linguistic; that is, grammatical. Curiously, there seem to be linguistic rules that tell you what

certain pronouns *cannot* refer to. In what follows, we will work out the basics of these rules, and find that they are very much dependent on syntax.

Our rules will not change the syntactic structure or words of sentences in any way; they simply specify possible (or impossible) meanings. Thus, are called **interpretive rules**.

We have already covered, informally, an interpretive rule for English, the Each Other Reference rule from Chapter 1. Here, we will cover further rules, with a more serious formalization of them.

7. Formal preliminaries

7.1 Subscripts and coreference

In what follows, we will use a standard notation for designating what pronouns refer to, namely, subscripting. When I write:

Bill_i thinks he_i is a genius.

I will mean: the reading of this sentence in which *it* is understood as referring to *Bill.* This is denoted by the use of identical letters as subscripts.

When I write

Bill_i thinks he_j is a genius.

the nonidentical subscripts should be taken to mean that *he*, in this reading, refers to someone other than Bill.

It will important later on to suppose that the indices are attached to the NP node, not further down (like to the Pronoun or Noun node).

Terminology: in the first sentence above, *Bill* and *he* are said to be **coreferent**, meaning that they refer to the same thing. In the second sentence, *Bill* and *he* are not coreferent. Also, in the first sentence, *Bill* is taken to be the **antecedent** for *he*, which means that it supplies the information about what *he* refers to.

7.2 Types of noun phrases

We will also make use of a three-way distinction:

- **Reflexive pronouns** are members of the set {myself, yourself, ourselves, ...}
- **Regular pronouns** are members of the set {I, me, you, he, them, ... }
- **Full noun phrases** are Noun Phrases that are neither reflexive pronouns or regular pronouns; such as *Sue*, *the president*, *my brother*, etc.

7.3 C-command

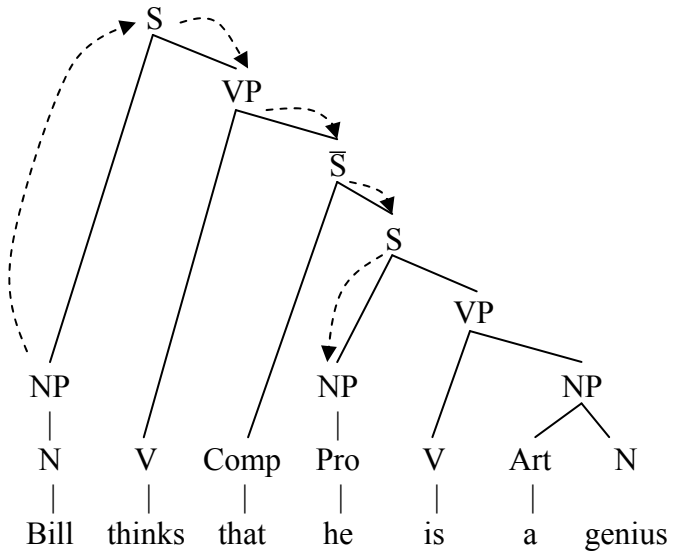
Lastly, a technical definition:

In a syntactic tree, constituent A **c-commands** constituent B if the mother of A dominates B.

In other words, A c-commands B if there is a path that

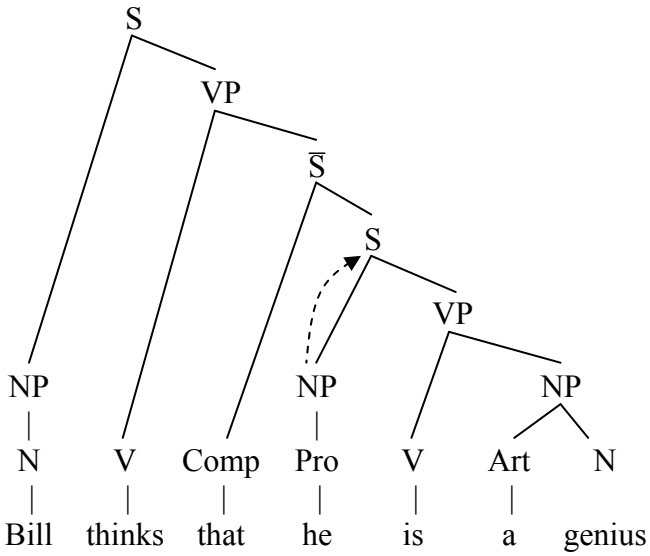
- starts at A
- moves up one node from A to its mother node
- travels exclusively downward through the tree and arrives at B.

Thus in the following example:



the NP *Bill* c-commands the pronoun *he* because you can go upward by one from the NP Bill, arrive at S, then move downward through VP, $\overline{S}$, S, and thence to the NP *he*. See dotted arrows.

But in the same example, the NP *he* does not c-command the NP *Bill* because once you've gone uphill once from *he*, you can't get to *Bill* by going just downhill:



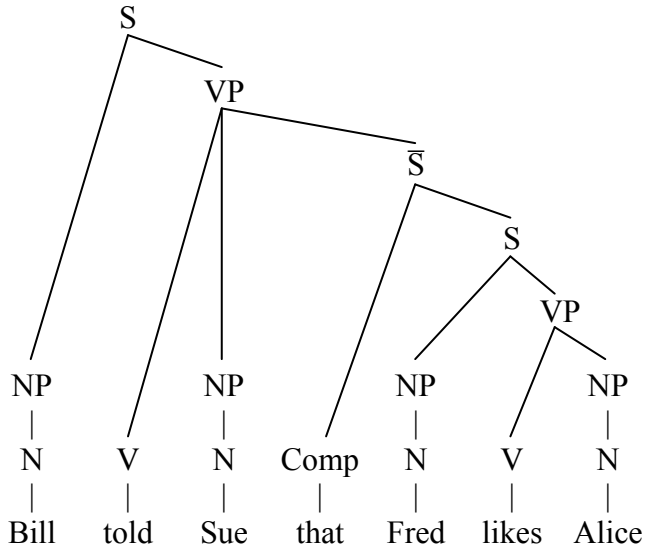
He does c-command *genius*, however.

In general, we will speak of c-command only for NP's. In drawing these little arrows, you want to start with the NP node, or you'll have problems…

The term c-command apparently means "constituent-command". It emerges from a period of syntactic research that tried out a number of similar definitions, of which c-command appears to have been both the simplest and most effective. We'll see the relevance of c-command to pronouns shortly.

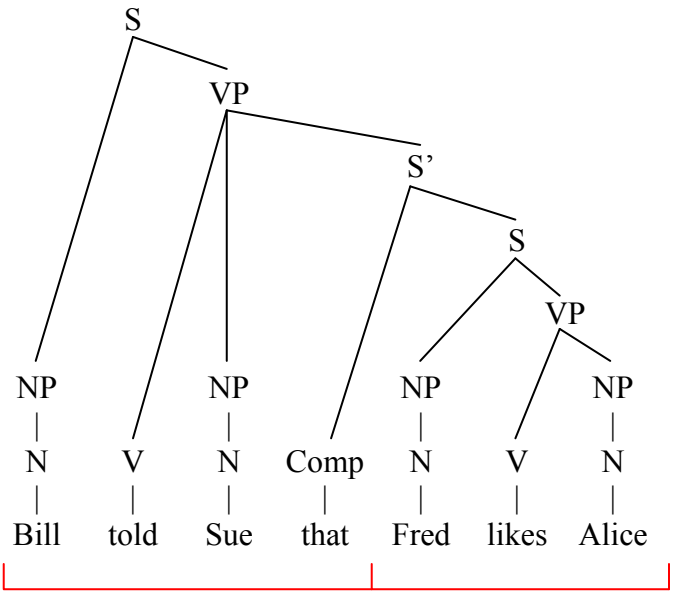
7.4 Clausemates

Following up on the discussion in Chapter 1, we will say that constituents X and Y are **clausemates** if every S node that dominates X also dominates Y, and vice versa. Clausemates are often said to be **in the same clause**, which means the same thing.



Clausemates: *Bill-Sue*, *Fred-Alice*. Non-clausemates: *Bill-Alice*, *Bill-Fred*, *Sue-Alice*, *Sue-Fred*.

A quick and easy way to show clausemates is to bracket the sentences into domains of clausematehood, like this:



Within brackets, any two NP are clausemates.

Study Exercise #30: in

Mary assumes that Fred will tell Sam that Alice saw Tom.

8. An interpretive analysis for reflexive pronouns

For reflexives, we can write the following rule of interpretation:

Reflexive Interpretation[93]

A reflexive pronoun must be coreferent with an NP that

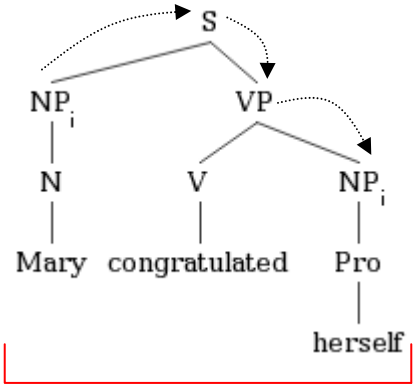
(a) is its clausemate; and
(b) c-commands it.

Here are examples, labeled for how the rule works. As you read these examples, I suggest you draw the tree, consult the definitions of c-command and clausemate, and check the rule is working correctly.

**Himself sings*.

Here, there's no NP for *himself* to be coreferent with, so it receives no interpretation. The standard assumption is that a sentence with an uninterpretable pronoun is ungrammatical.

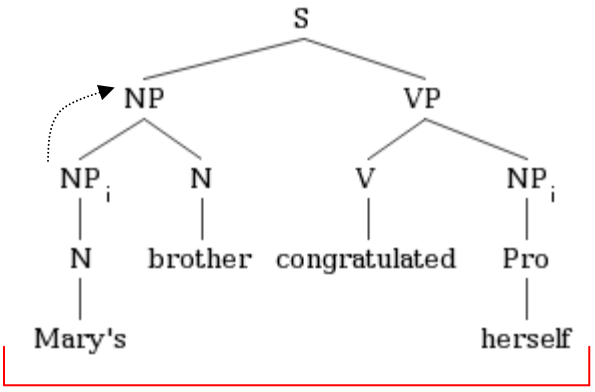
Mary$_i$ congratulated herself$_i$.



This one is fine; the NP *Mary* c-commands the NP *herself* and, since there is just one clause, the two are clausemates. The correct indexation (note: on the NP's, not lower down) is shown in the tree above.

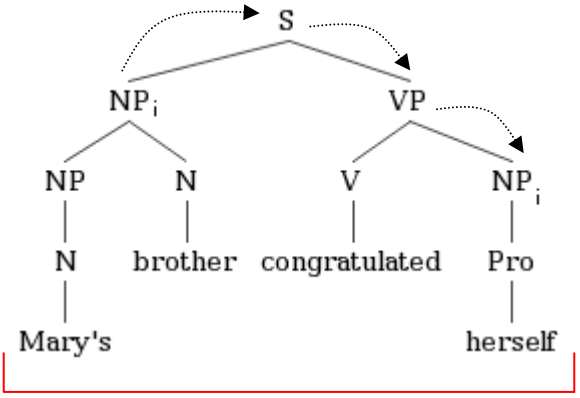
*[Mary$_i$'s brother] congratulated herself$_i$.

[93] In the linguistics literature this rule, in a slightly generalized version, is called "Principle A." I'm deviating from standard terminology here, taking the view that rules should be given names that correspond to their function.



Here, *Mary* is a clausemate of *herself*, but doesn't c-command it—the mother of *Mary* is the higher NP *Mary's brother*; so *Mary* is not "high enough" in the tree to c-command *herself*.

*[Mary's brother]$_i$ congratulated herself$_i$.

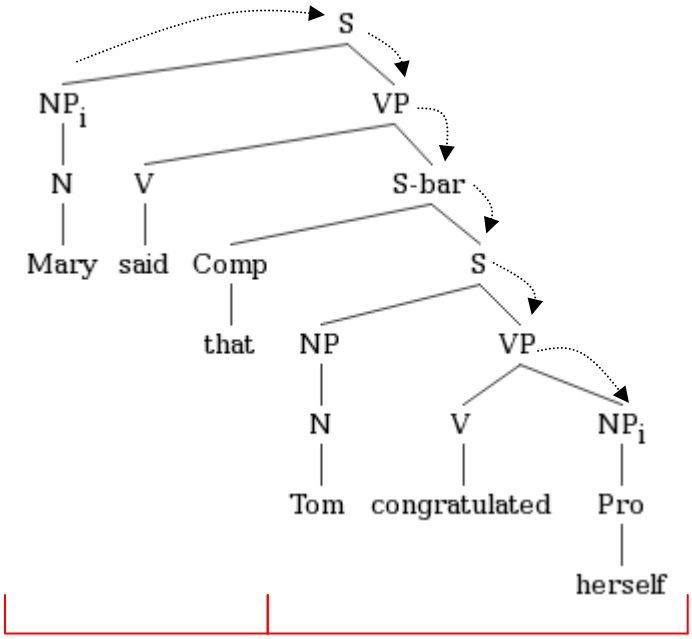


Same tree, but different indices. Here, the NP *Mary's brother* does c-command the NP *herself*, and is a clausemate, but there's a gender mismatch (brothers are always male, and *herself* is [Gender:feminine]), so the sentence is still ungrammatical.

[Mary's brother]$_i$ congratulated himself$_i$.

This one matches all requirements (gender match, c-command, clausemate condition), and is fine.

Mary$_i$ said that Tom congratulated herself$_i$.



Bad: *Mary* is not the clausemate of *herself* (*herself* is in the lower S, *Mary* is not).

Mary said that Tom$_i$ congratulated herself$_i$.

Bad: gender mismatch (unless, of course, *Tom* designates a female person; who knows...)

Study Exercise #31

The fact that Mary$_i$ lost the race surprised herself$_i$.

This one is bad; you give the explanation.

8.1 Each other

The phrase *each other* is a **reciprocal** pronoun, not a reflexive, but it works essentially like a reflexive and is normally analyzed using the same sort of rule. Thus:

[John and Mary]$_i$ like [each other]$_i$.	Ok.
*[Each other]$_i$ like [John and Mary]$_i$.	Bad: c-command condition violated
*[John and Mary]$_i$ think I like [each other] $_i$.	Bad: clausemate condition violated

For further relevant data see Chapter 1.

9. An interpretive analysis for regular pronouns

The regular pronouns (like *she*, *him*, *us*, *our*, etc.) are used quite differently from reflexives. For one thing, they can be used without any linguistic Noun Phrase to refer to at all—as in the the "He left" example on p. 244 above.

The key to these pronouns, in the view of many linguists, is that you to specify not what they *can* refer to, but rather what they *cannot* refer to. Here is a version of the rule commonly proposed:

Regular Pronoun Interpretation[94]

A regular pronoun cannot be coreferent with a c-commanding clausemate.

Some examples follow.

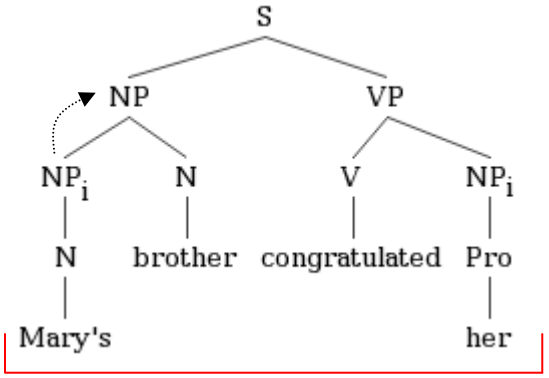
9.1 Pronouns alone

He$_i$ *sings.*

This is fine: there is no NP in the sentence that *he* is required to be non-coreferent with, and the sentence is freely usable whenever there's an obvious enough male entity available for *he* to refer to. This could be someone mentioned in a previous sentence, or someone noticed in the physical surroundings.

9.2 No c-command, coreference ok

[Mary$_i$'s brother] congratulated her$_i$.



This is fine, because *Mary* doesn't c-command *her*.

9.3 No coreference — always ok

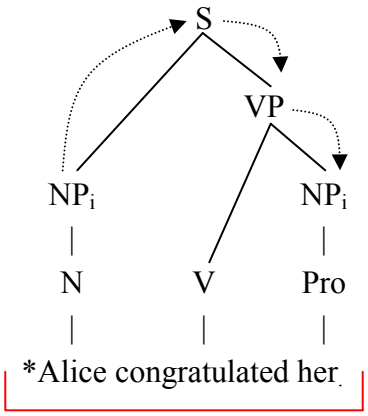
[Mary$_i$'s brother] congratulated her$_j$.

[94] In the linguistics literature this rule is called "Principle B."

The subscript *j* means that the *her* refers to a female person other than *Mary*. This is fine, too—Regular Pronoun Interpretation doesn't actually *require* that pronouns be coreferent with any other NP in the sentence. Thus, this sentence could appear in a context like this:

Alice sang incredibly well, enough to convince her sternest critics. In fact, even Mary's brother congratulated her.

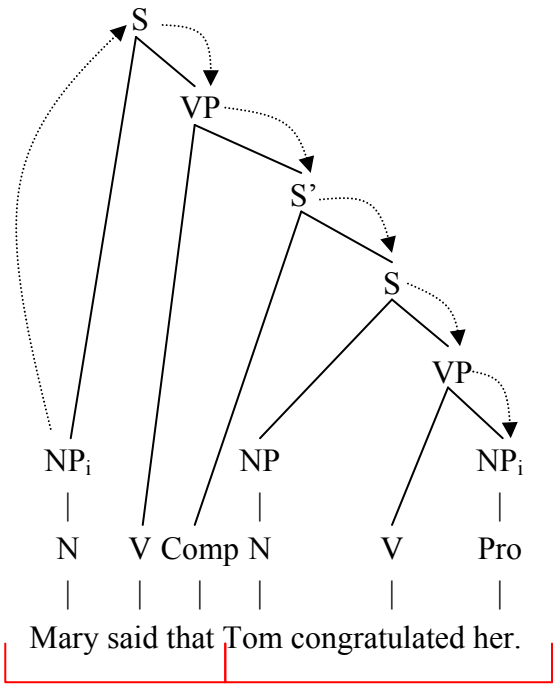
9.4 C-commanding clausemate: coreference impossible



This one is no good: *Alice* is the clausemate of *her*, and also c-commands *her*, so it can't be coreferent. However, with distinct reference, the following reading is ok:

Alice$_i$ congratulated her$_j$.

9.5 C-command but not clausemate: coreference ok



This one is fine: *Mary* does c-command *her*, but it is not the clausemate of *her*, so Regular Pronoun Interpretation doesn't rule out this reading.

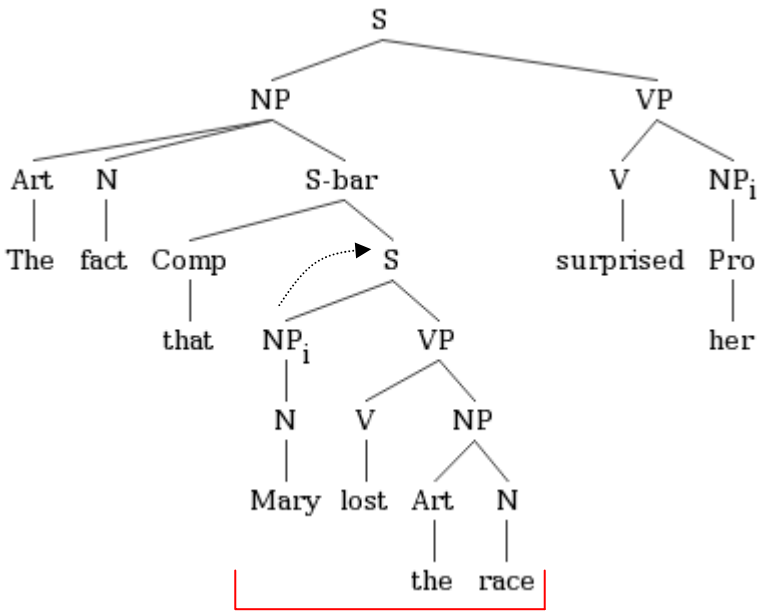
9.6 C-command but not clausemate: non-coreference ok

Mary$_i$ said that Tom congratulated her$_j$.

This is likewise fine, *her*$_j$ refers to some female person mentioned earlier or physically present.

9.7 No c-command, not clausemate: coference ok

The fact that Mary$_i$ lost the race surprised her$_i$.

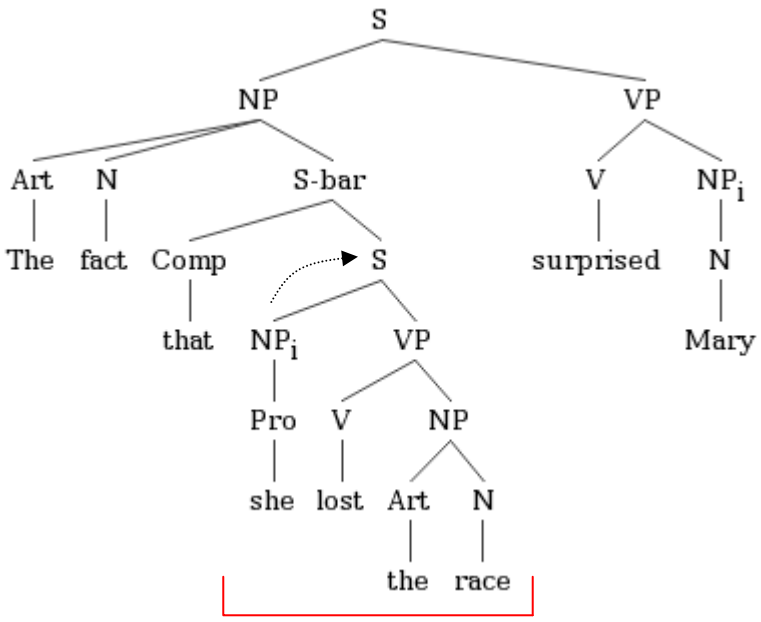


Ok, *Mary* is neither a clausemate of *her*, nor does *her* c-command *Mary*, so the coreference is allowed.

9.8 Backwards coreference

An intriguing prediction of the analysis is that you could, in principle, get sentences in which the pronoun actually comes before the full NP with which it is coreferent. These do in fact arise, though because of additional factors they won't be found in all places you would expect them. Here is an example:

The fact that she$_i$ lost the race surprised Mary$_i$.



This sounds best only under particular conditions of emphasis and intonation. In particular, you can't utter *Mary* with a full phrasal stress, as if the name were being introduced to the conversation for the first time—if *Mary* were new information, you wouldn't have been referring to her with a pronoun. The sentence sounds ok if you say:

The fact that she$_i$ lost the race SURPRISED Mary$_i$.

Of course, since Regular Pronoun Interpretation only forbids coreference, the following reading is also acceptable:

The fact that she$_j$ lost the race surprised Mary$_i$.

10. An interpretive analysis for full noun phrases

One wouldn't think that there need to be any rules for the meaning of full noun phrases, but these are in fact needed. Consider a sentence like:

*He$_i$ thinks Bill$_i$ is a genius.

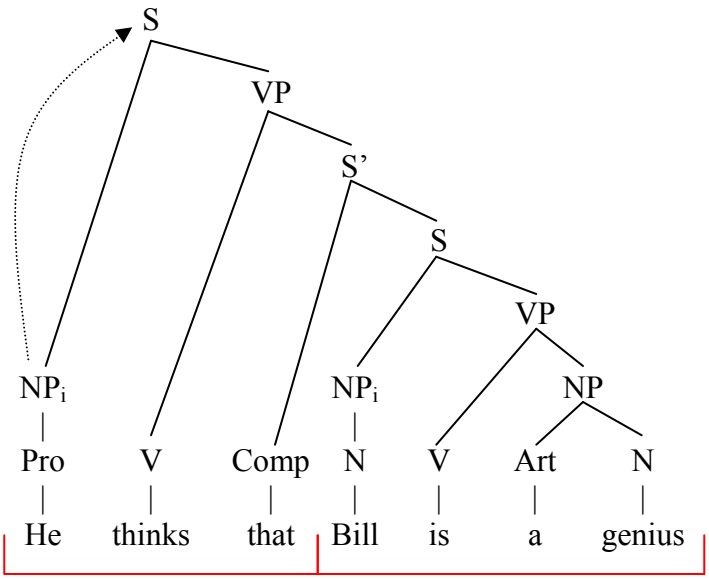
The coreference shown is impossible, even though nothing we've said so far rules it out. The rule commonly used is this one:

Full Noun Phrase Interpretation[95]

A full noun phrase cannot be coreferent with a c-commanding Noun Phrase.

[95] In the linguistics literature this rule is often called "Principle C."

This rules out **He*$_i$ *thinks that Bill*$_i$ *is a genius* because *he* c-commands *Bill* and *Bill* is a full NP.



Indeed, the same rule predicts that in

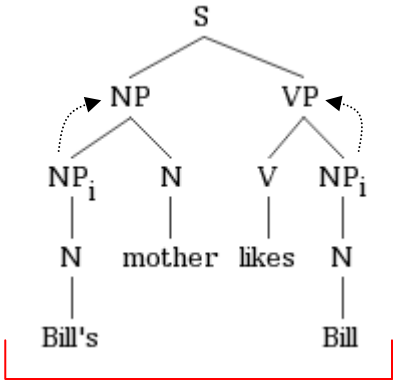
Bill thinks that Bill is a genius.
Bill saw Bill.

we must interpret the two *Bill*'s as being different people; that is, these sentences must be interpreted:

Bill$_i$ thinks Bill$_j$ is a genius.
Bill$_i$ saw Bill$_j$.

If neither copy of Bill c-commands the other, then coreference becomes more or less ok:

[well, at least] Bill$_i$'s mother likes Bill$_i$.



Study Exercise #32: in the following, why can the two *Bill*'s be the same person?

The idea that Bill might have the lowest score bothers Bill.

11. Summary

We've now done a particular corner of English semantics, setting out rules of semantic interpretation for anaphoric elements. Dividing all NP's into the categories of Reflexive Pronouns, Regular Pronouns, and Full NP's, we developed three rules, one of which requires coreference in certain contexts, the other two of which forbid it:

Reflexive Interpretation

A reflexive pronoun must be coreferent with an NP that (a) is its clausemate; and (b) c-commands it.

Regular Pronoun Interpretation

A regular pronoun cannot be coreferent with a c-commanding clausemate.

Full Noun Phrase Interpretation

A full noun phrase cannot be coreferent with a c-commanding Noun Phrase.

Chapter 15: More review problems

Study Exercise #50: Phonetic Dictations

southern
myrrh
corpulent
whether
multiple
coinage
parameter
ostentatious
turmoil
trapezium

Answer to Study Exercise #50

southern	[ˈsʌðɚn]
myrrh	[ˈmɚ]
corpulent	[ˈk{oʊ,o,ɔ}ɹpj{u,ʊ,ə}lənt]
whether	[ˈwɛðɚ]
multiple	[ˈmʌlt{ɪ, ə}p{l̩, əl}]
coinage	[ˈkɔɪn{ə,ɪ}dʒ]
parameter	[pəˈɹæmərɚ] [ɚ] for first [ə] or [əɹ] ok
turmoil	[ˈtɚmɔɪl]
ostentatious	[ɔstɛnˈteɪʃəs]
trapezium	[tɹəˈpiziəm]

Study Exercise #51: Historical Linguistics

Here are matched sets from three dialects of English. Apply the Comparative Method, forming correspondence sets and positing sound changes. Here, is it best to compare sequences rather than sounds. Do: [juɹ, uɹ, oɹ].

		Dialect A	*Dialect B*	*Dialect C*
1.	*Muir*	[ˈmjuɹ]	[ˈmjuɹ]	[ˈmjoɹ]
2.	*moor*	[ˈmuɹ]	[ˈmuɹ]	[ˈmoɹ]
3.	*more*	[ˈmoɹ]	[ˈmoɹ]	[ˈmoɹ]
4.	*cure*	[ˈkjuɹ]	[ˈkjuɹ]	[ˈkjoɹ]
5.	*Coors*	[ˈkuɹz]	[ˈkuɹz]	[ˈkoɹz]
6.	*core*	[ˈkoɹ]	[ˈkoɹ]	[ˈkoɹ]
7.	*Buhr*	[ˈbjuɹ]	[ˈbjuɹ]	[ˈbjoɹ]
8.	*boor*	[ˈbuɹ]	[ˈbuɹ]	[ˈboɹ]
9.	*bore*	[ˈboɹ]	[ˈboɹ]	[ˈboɹ]
10.	*endure*	[ɛnˈdjuɹ]	[ɛnˈduɹ]	[ɛnˈdoɹ]
11.	*dour*	[ˈduɹ]	[ˈduɹ]	[ˈdoɹ]
12.	*door*	[ˈdoɹ]	[ˈdoɹ]	[ˈdoɹ]
13.	*Turing*	[ˈtjuɹɪŋ]	[ˈtuɹɪŋ]	[ˈtoɹɪŋ]
14.	*tour*	[ˈtuɹ]	[ˈtuɹ]	[ˈtoɹ]
15.	*tore*	[ˈtoɹ]	[ˈtoɹ]	[ˈtoɹ]

16.	*inure*	[ɪˈnjuɹ]	[ɪˈnuɹ]	[ɪˈnoɹ]
17.	*Koh-i-noor*[187]	[ˈkoʊhɪnuɹ]	[ˈkoʊhɪnuɹ]	[ˈkoʊhɪnoɹ]
18.	*nor*	[ˈnoɹ]	[ˈnoɹ]	[ˈnoɹ]

[187] A famous diamond, from the Persian for “mountain of light”.

Answer to Study Exercise #51

Correspondence sets:

Proto	A	B	C	Examples
*juɹ	juɹ	juɹ	joɹ	1, 4, 7
*juɹ	juɹ	uɹ	oɹ	10, 13, 16
*uɹ	uɹ	uɹ	oɹ	2, 5, 8, 11, 14, 17
*or	oɹ	oɹ	oɹ	3, 6, 9, 12, 15, 18

The proto-language is identical to Dialect A.

B and C have both undergone:

*j > ∅ / [+alveolar] ___

See *endure*, *Turing*, and *inure*

C has also undergone:

*u > o / ___ ɹ

which has merged *moor* with *more*, *boor* with *bore*, and so on.

Study Exercise #52: Wh- Movement and islands

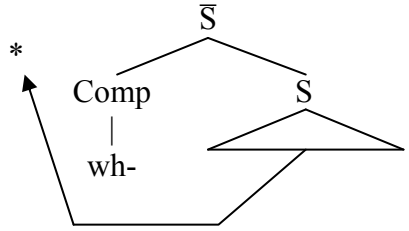
Show why

**What donor might Sue wonder what books donated to the library?*

is ungrammatical, given the Wh- Island Constraint below. In particular, first extract *what books* to the lower Comp, then extract *what donor* to the higher Comp, showing the island violation graphically.

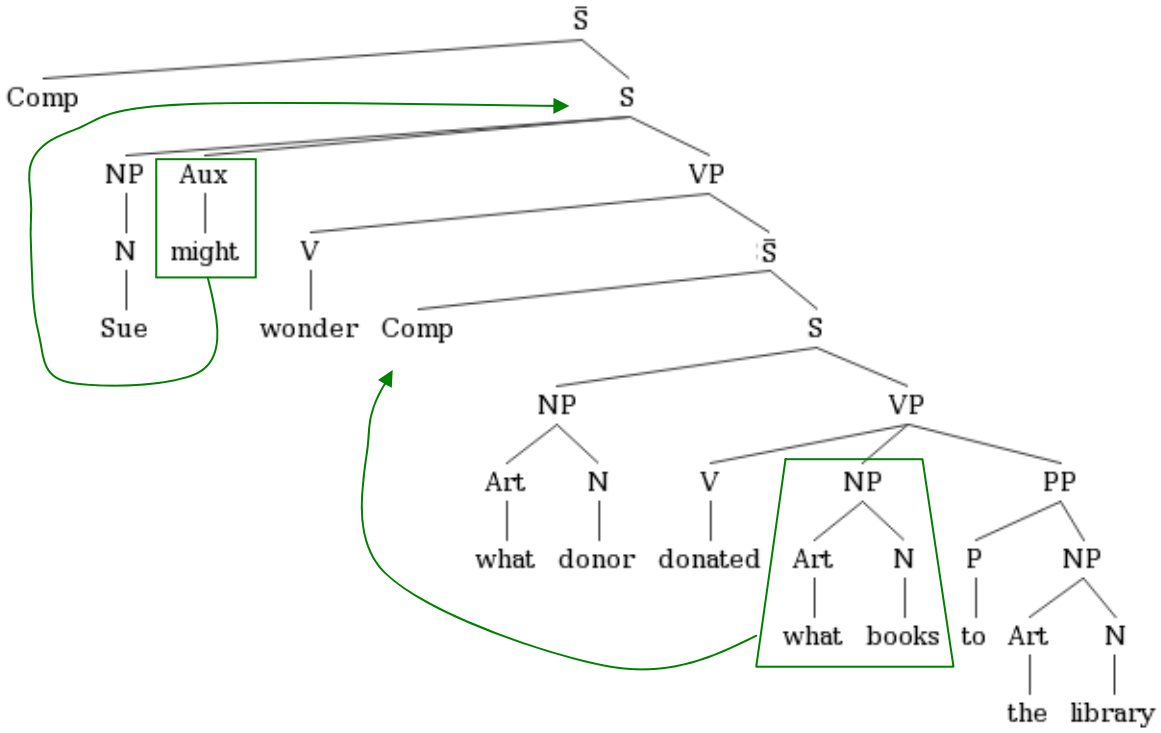
Wh- Island Constraint

Mark as ungrammatical any sentence in which a constituent has been extracted from inside an $\overline{S}$ whose Comp contains a wh- phrase.

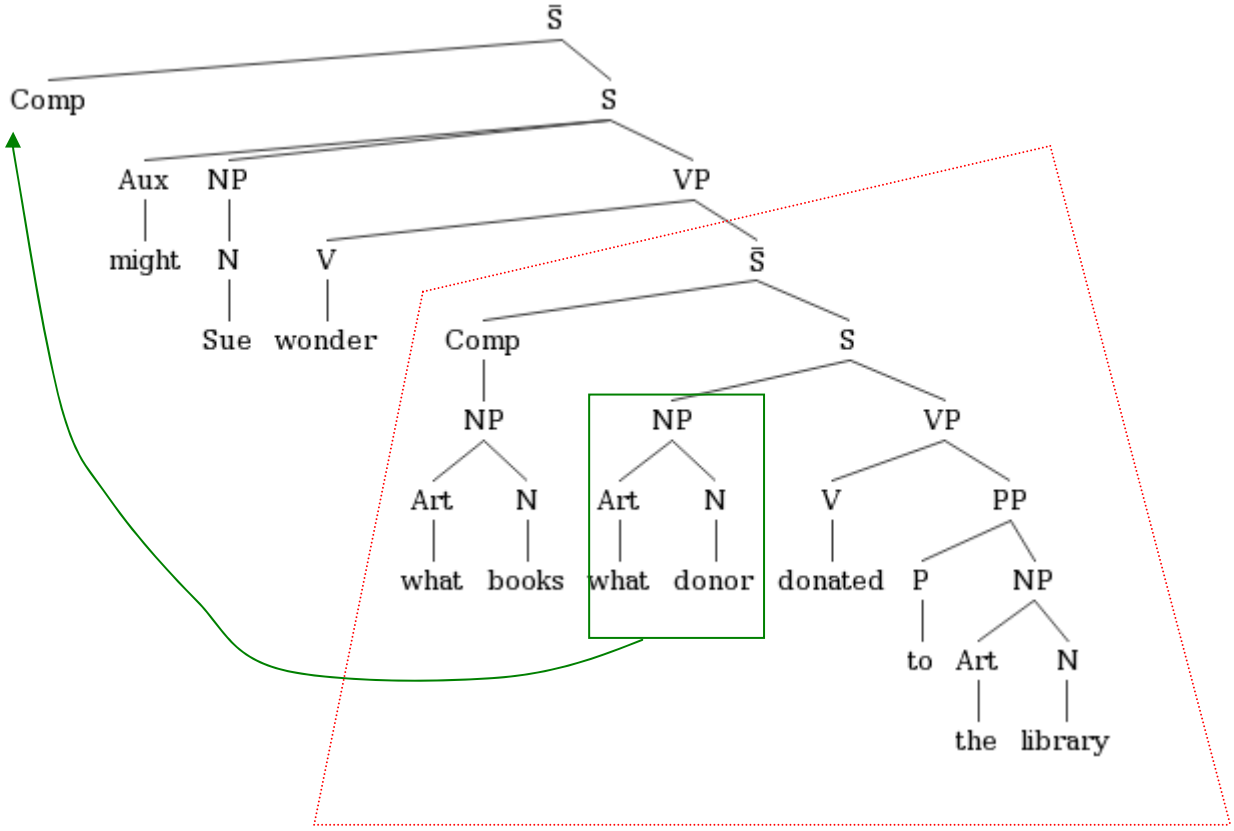


Answer to Study Exercise #52

Deep structure (all wh- phrases *in situ*), with lower instance of Wh-Movement; also Subject-Aux Inversion in upper clause:



Resulting tree, with subsequent movement of *what donor* into the higher Comp. This violates the Wh-Island Constraint; the island is enclosed in a dotted box:



Since a wh-phrase is moved out of the island, the resulting sentence is ungrammatical.

Study Exercise #53: Phonology

This is an imaginary language but the rules it has are found in real languages.
[β, ð, ɣ] are voiced fricatives (bilabial, dental, velar). [t̪, d̪, n̪] are dental.

	'Noun'	*'the Noun'*	*'two Nouns'*	*'five Nouns'*	
1.	[pama]	[la bama]	[d̪ue bamas]	[kwin̪d̪o bamas]	'tuna'
2.	[peli]	[la beli]	[d̪ue belis]	[kwin̪d̪o belis]	'swordfish'
3.	[t̪ube]	[la d̪ube]	[d̪ue d̪ube]	[kwin̪d̪o d̪ube]	'mackerel'
4.	[t̪azo]	[la d̪azo]	[d̪ue d̪azo]	[kwin̪d̪o d̪azo]	'cod'
5.	[kame]	[la game]	[d̪ue game]	[kwin̪d̪o game]	'mahi mahi'
6.	[koli]	[la goli]	[d̪ue goli]	[kwin̪d̪o goli]	'carp'
7.	[bafi]	[la βafi]	[d̪ue βafi]	[kwin̪d̪o βafi]	'catfish'
8.	[belu]	[la βelu]	[d̪ue βelu]	[kwin̪d̪o βelu]	'pollock'
9.	[d̪aba]	[la ðaba]	[d̪ue ðaba]	[kwin̪d̪o ðaba]	'yellowtail'
10.	[d̪azo]	[la ðazo]	[d̪ue ðazo]	[kwin̪d̪o ðazo]	'sturgeon'
11.	[gele]	[la ɣele]	[d̪ue ɣele]	[kwin̪d̪o ɣele]	'halibut'
12.	[gova]	[la ɣova]	[d̪ue ɣova]	[kwin̪d̪o ɣova]	'salmon'

a) Produce consonant and vowel charts, labeling the rows and columns with features. You may assume [+dental] is a feature.
b) Do the stems alternate? Explain
c) Give rules, naming them.
d) Is any rule ordering required?
e) Give right order/wrong order derivations for *la dazo* and *la ðazo*.

Answer to Study Exercise #53

a) Produce consonant and vowel charts.

		[+bilabial]	[+labiodental]	[+dental]	[+alveolar]	[+velar]
[+stop]	[−voice]	p		t		k
	[+voice]	b		d		g
[+fricative]	[−voice]		f		s	
	[+voice]	β	v	ð	z	ɣ
[+nasal]		m		n		
[+liquid]					l	
[+glide]		w				

	[–back, –round]	[+back, –round]	[+back, –round]
[+high, –low]	i		u
[–high, –low]	e		o
[–high, +low]		a	

b) Do the stems alternate? Explain

Yes, for example the stem for "tuna" has the two allomorphs [pama] and [bama].

c) Give rules, naming them.

Intervocalic Voicing

[+stop] → [+voiced] / [+syllabic] ___ [+syllabic]

This voices any stop occurring between vowels. It can be applied harmlessly to [b, d, g], since they are already voiced, so I left out [−voice] from the left side of the arrow.

Intervocalic Spirantization[188]

[+stop, +voice] → [–stop, +fricative] / [+syllabic] ___ [+syllabic]

This turns any voiced stop between vowels to its fricative counterpart, thus [b, d, g] → [β, ð, ɣ].

[188] Standard terminology for a rule that creates fricatives. "Spirant" is an old-fashioned synonym for "fricative."

d) Is any order required?

Intervocalic Spirantization must precede Intervocalic Voicing, to keep the voiced stops that derive from voiceless from turning into fricatives—we want Intervocalic Frication to apply "too late" to affect those stops.

e) Give right order/wrong order derivations for la dazo and la ðazo.

Correct:

/la tazo/	/la dazo/	phonemic forms
—	ð	Intervocalic Spirantization
d	—	Intervocalic Voicing
[la dazo]	[la ðazo]	phonetic forms

Incorrect:

/la tazo/	/la dazo/	phonemic forms
d	—	Intervocalic Voicing
ð	ð	Intervocalic Spirantization
*[la ðazo]	[la ðazo]	phonetic forms

Study Exercise #54: Semantics; anaphora

The wizards believe that the witches turned the girls into copies of each other.

a. Produce the phrase structure tree.
b. Show clausemates with brackets, as in xxx the readings, show c-command with arrows.
c. Explain with reference to rule given below the possible reference of *each other*.

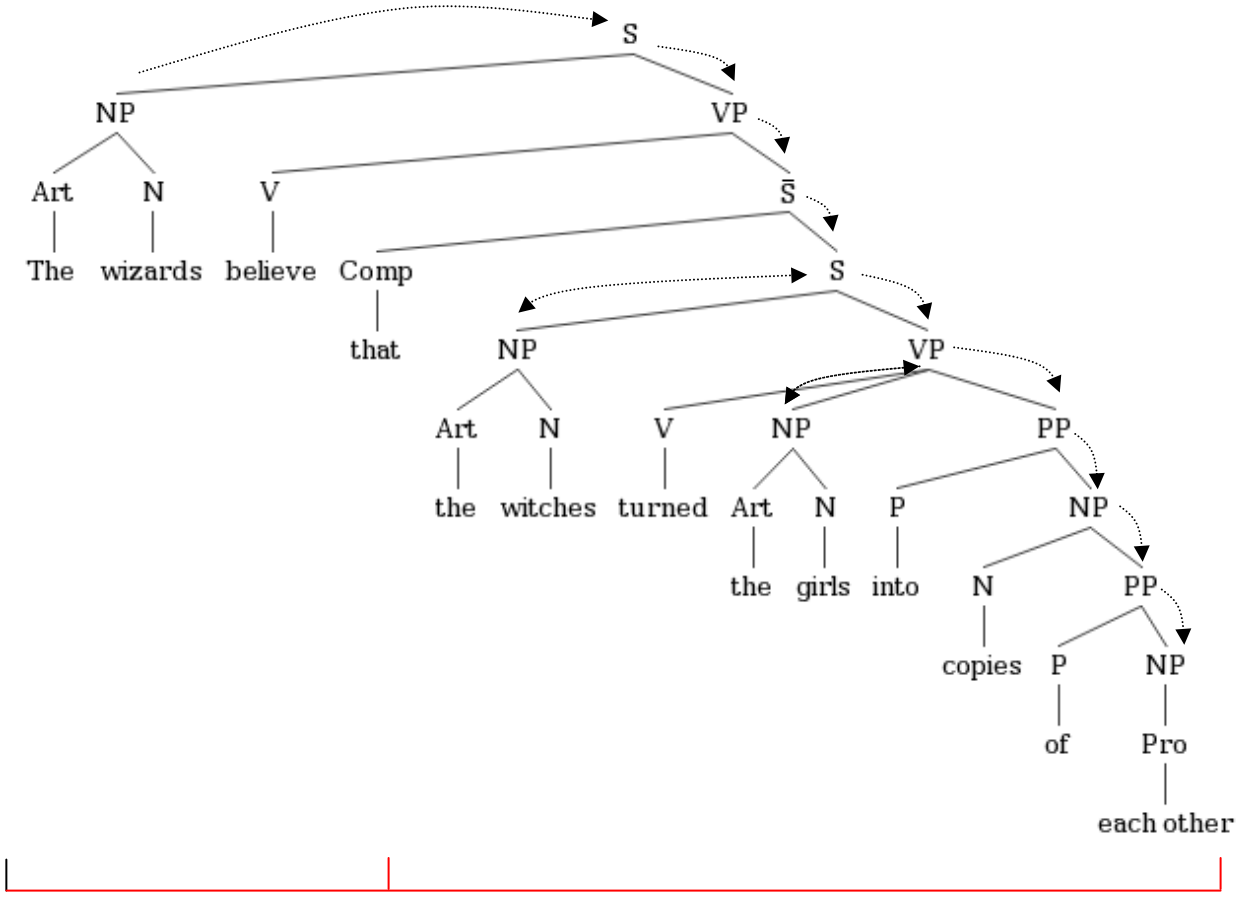
Each Other Reference

Each other may refer only to a c-commanding clausemate.

Answer to Study Exercise #54

The wizards believe that the witches turned the girls into copies of each other.

a. Diagram/b. Show clausemates with brackets, show c-command with arrows.



The witches, *the girls*, and *each other* are all clausemates, but *the wizards* is not clausemates with any of them.

Looking at the tree and the crucial NP's, we see the following relations of c-command:[189]
the wizards c-commands the other three NP's
the witches c-command *the girls* and *each other*
the girls c-commands *each other*

[189] Recall how this is determined: go up one node from any NP, and anything dominated by this node is dominated by this NP.

c. Putting it all together, we see that:

- *the girls* c-commands and is a clausemate of *each other*, and so can be coreferent with *each other*

  Scenario: The wizards believe that the witches turned Sue into a copy of Ellen, and turned Ellen into a copy of Sue.

- *the witches* c-commands and is a clausemate of *each other*, and so can be coreferent with *each other*

  Scenario: The wizards believe that Alice, a witch, turned the girls into copies of Miriam (another witch), and that Miriam turned the girls into copies of Alice.

- While *the wizards* c-commands *each other*, it is not a clausemate of *each* other, and so it cannot be coreferent with *each other*

  Scenario: Bob, a wizard, believes the witches turned the girls into copies of Ted, another wizard; and Ted believes the witches turned the girls into copies of Bob. Logically possible, but evidently not available linguistically.

Study Exercise #55: Semantics: scope

This sentence has a scope-based ambiguity.

Many people visit two islands.

i. Describe clearly in words the two meanings of these sentence.
ii. Make up a scenario of which it could hold true.
iii. Using Quantifier Translation and Quantifier Raising, derive the logical forms for each meaning.

Quantifier Translation

Replace

[every N]$_{NP}$	with	[for every x, x an N]$_{NP}$
[some N]$_{NP}$	with	[for some x, x an N]$_{NP}$
…		

and similarly for other quantified expressions. If the variable *x* is already in use, use *y* instead; etc.

Quantifier Raising
Left-adjoin a quantified NP to S, leaving behind a variable in its original location.

Answer to Study Exercise #55

Describe clearly in words the two meanings of this sentence. Give a scenario of which it could hold true.

Many people visit two islands.

i. Describe clearly in words the two meanings of this sentence.

(a) It is true of many people that they visit two islands (not necessarily the same two).
(b) It is true of two islands that many people visit them (not necessarily the same people).

ii. Give scenarios.

(The possibilities are infinite, but here is one:)

(a): A travel agency offering tours of the Hawaiian Islands offers a great number of package tours:

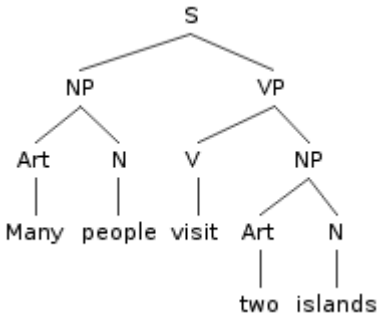
One-Island tours: just Oahu, or Big Island, or Maui, or Molokai, etc.
Two-Island tours: Oahu and Big Island, or Oahu and Maui, or Big Island and Kauai, etc.
Three-Island tours: Oahu, Big Island, and Maui; or Kauai, Oahu, and Maui; etc.

Many customers select a Two-Island Tour.

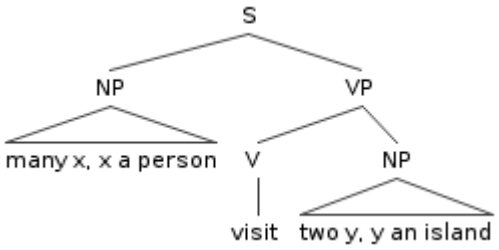
(b) In this scenario: many people visit Oahu, many people visit the Big Island, but hardly anyone ever visits Kauai, Maui, Molokai, or any of the other islands.

iii. Derivation of logical forms

Surface structure:



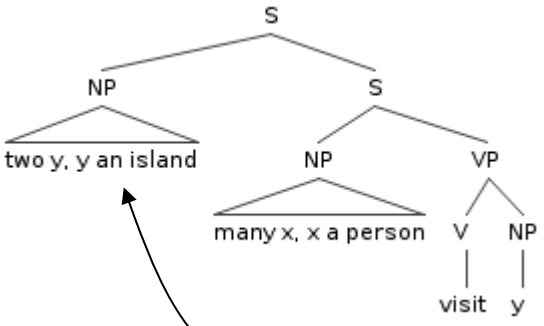
Quantifier Conversion:



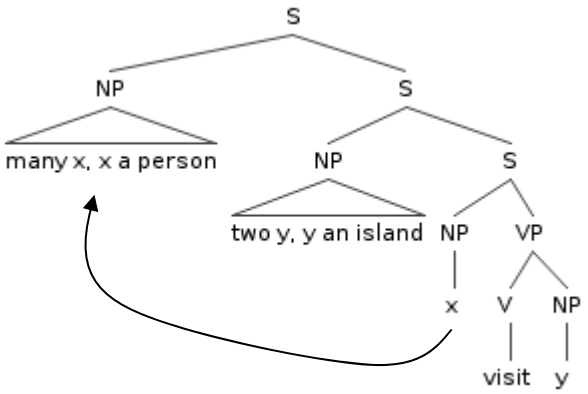
At this point, the meanings depend on the order in which the quantifier operators are raised.

(a)

Quantifier Raising I



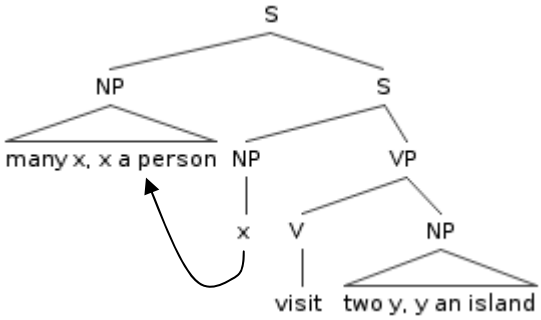
Quantifier Raising II



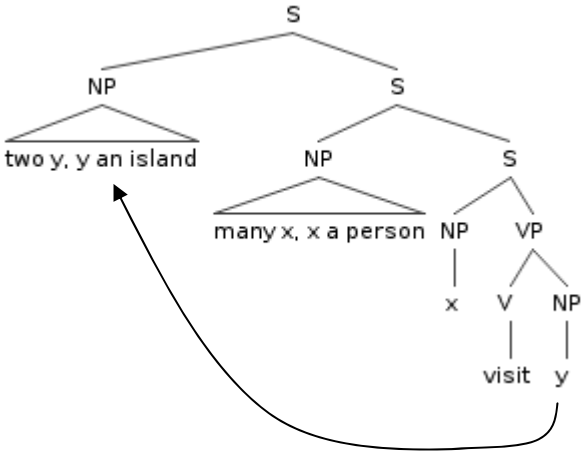
"It is true of many people that they visit two islands"

(b)

Quantifier Raising I



Quantifier Raising II



"It is true of two islands that many people visit them."