KYLE JOHNSON

TRANSFORMATIONAL GRAMMAR

LINGUISTICS DEPARTMENT, UNIVERSITY OF MASSACHUSETTS AT AMHERST

Contents

Preface

These are the always evolving notes from an introductory course on syntactic theory taught at the University of Massachusetts at Amherst. Its target audience is first-year graduate students. No background exposure to syntax is presupposed.

The Subject

Linguistic theory, and so syntactic theory, has been very heavily influenced by learnability considerations in the last fifty-some years, thanks largely to the writings of Noam Chomsky. If we decide that syntactic theory is charged with the duty of modeling our knowledge of language, then we can make some initial deductions about what this knowledge, and therefore our model of it, should look like from some simple observations. This knowledge must interact efficiently with the cognitive mechanisms involved in producing and comprehending speech, for instance. It must also be acquirable by any normal infant exposed to speakers of the language over six or so years. A number of considerations combine to make the task of acquiring knowledge of a language look very difficult indeed: the complexity of the acquired grammar, the amount of information that needs to be acquired, the attenuated nature of the information available to the child, etc. It is made even more puzzling by the fact that children appear to complete this task with relative ease in a comparatively short period of time and that the course of acquisition appears to go through a set schedule of stages. There is clearly a problem: If languages are as complex as we think they are, then how can these impossibly complex objects possibly be learned?

Linguistics as learning theory

It is Chomsky's proposal that Syntactic Theory itself should contribute to solving this dilemma. The classical formulation of this idea (see *Aspects* and *The Sound Pattern of English*) characterizes the situation as follows. Think of a grammar of L (G_L) (this is what Chomsky (1986b) calls "I-Language") as a set of rules that generates structural descriptions of the strings of the language L (Chomsky (1986b)'s E-language). Our model of this grammar is descriptively adequate if it assigns the same structural descriptions to the strings of L that G_L does. We can think of the learning process as involving a selection from the Universe of G_Ls the very one that generates these structured strings of the L to be acquired.

The learning problem can now be stated in the following terms: how is it that the learning procedure is able to find G_L when the universe of Gs is so huge and the evidence steering the device so meager.

One step towards solving this problem would be to hypothesize that the universe of Gs has a structure that enables convergence on G_L given the sort of information that the child is exposed to. This is Chomsky's proposal. It amounts to the claim that there are features of Gs which are built-in: certain properties which distinguish the natural class of Gs from the rest. There is a kind of meta-grammar of the Gs, then, which is sometimes referred to with the label Universal Grammar. Chomsky further hypothesizes that these properties are biologically given: that it is something about the construction of the human brain/mind that is responsible for the fact that the class of Gs are the way they are. This argument, the one that leads from the observation that G_Ls have features that are too complex to be learned to the conclusion that the universe of Gs is constrained is often called "The Poverty of the Stimulus" argument. It is a classic from Epistemology, imported with specific force by Chomsky into linguistics.

This way of setting up the problem, note, allows for the Universe of Gs to be larger than the learnable Gs. There could, for instance, be constraints imposed by the parsing and production procedures which limit the set of Gs that can be attained. And it's conceivable that there are properties of the learning procedure itself — properties that are independent of the structure of Gs imposed by Universal Grammar — that could place a limit on the learnable Gs. Universal Grammar places an outside bound on the learnable grammars, but it needn't be solely responsible for fitting the actual outlines of that boundary. It's therefore a little misleading to say that the set of "learnable Gs" are those characterized by Universal Grammar, since there may be these other factors involved in determining whether a grammar is learnable or not. I should probably say that Universal Grammar carves out the "available Gs," or something similar. But I will instead be misleading, and describe Universal Grammar as fixing the set of learnable Gs, always leaving tacit that this is just grammar's contribution to the learnability question.

Chomsky proposes, then, that a goal of syntactic theory should be to contribute towards structuring the universe of Gs. He makes some specific proposals about how to envision this in *Aspects of The Theory of Syntax*. He suggests that syntactic theory should include an evaluation metric which "ranks" Gs. A syntactic theory that has this feature he calls explanatory. Thus "explanatory theory" has a specific, technical, sense in linguistic theory. A theory is explanatory if and only if it encapsulates the features that ranks Gs in such a way that it contributes to the learnability problem, distinguish the learnable Gs from the unlearnable ones. This criterion can help the syntactician decide whether the model of G_L he or she has proposed corresponds exactly to G_L. In particular, the many descriptively adequate models of G_L can be distinguished on this basis: we should select only those that are ranked highly by the evaluation metric. These grammars meet the criterion of explanatory adequacy.

A very important role, therefore, is played by the evaluation metric. At the time of *Aspects*, the learning procedure was conceived of as a process very much like that which the linguist goes through. The child builds a battery of rules which generate the strings of L. The evaluation metric steering this process was thought to have essentially two parts: a simplicity metric, which guides the procedure in its search through the space of grammars, and inviolable constraints, which partitions the set of Gs into the learnable ones and the unlearnable ones. Thus, for example, we might imagine that rules which used fewer symbols could be defined as "simpler" than ones that used a greater number of symbols. Inviolable constraints might be those, for example, expressed as part of the principles which place constraints on the way that strings can be partitioned into groups, and therefore simply removes from the universe of Gs a great many possible Gs. Let's call these models of Gs "rule based," because the simplicity metric is defined as a procedure that constructs rules, and the companion picture of the acquisition process the "Little Linguist" model.

To take a concrete example, imagine that the principles which limit how words are strung into groups — one particular version of which goes by the name "$\overline{\text{X}}$ Theory" — imposes the following constraints.

$$\text{XP} \rightarrow \{\,(\text{ZP}), \overline{\text{X}}\,\}$$
$$\overline{\text{X}} \rightarrow \{\,\overline{\text{X}}, (\text{YP})\,\}$$
$$\overline{\text{X}} \rightarrow \{\,\text{X}^0, (\text{WP})\,\}$$

Understand "$\{\alpha, \beta\}$" to signify that α and β are sisters, and "(α)" to indicate that α is optional. Let W, X, Y and Z range over kinds of lexical items (e.g., "noun," "verb," "preposition," and so on). And, finally, let "$\rightarrow$" mean: "consists of." The groups here, known as phrases, are the XP and $\overline{\text{X}}$ in the formulas. These constraints, then, leave to the learner only the matter of filling in the variables W, X, Y and Z, and discovering their linear order. As the child goes from step to step in matching the grammar he or she is constructing with the information coming in, these are the only decisions that have to be made. If we imagine that this set of options were to be operationalized into a concrete decision tree, then we could see this as constituting a kind of "simplicity metric." It would constitute a procedure for searching through the space of learnable grammars that imposes an order on the grammars, enabling a deterministic method for converging at a particular grammar when exposed to a particular linguistic environment. Additionally, $\overline{\text{X}}$ Theory provides an absolute cap on the possible phrases and, in this respect, constitutes an inviolable constraint as well. If every language learner is equipped with this $\overline{\text{X}}$ Theory, then they will converge more or less on the same G_L when presented with the information that being in the environment of speakers of L provides. If there are differences in the G_L's that learners converge on, these will trace back to different decisions these learners have made about the identity of W, X, Y and Z, or how their linear order is determined. If the rest of a model that incorpo-

rates these constraints is correct, then, it should allow any language learner to pick out a G_L very close to the G_L giving shape to the speech in that learner's environment.

Let's consider another example involving transformational rules, one that Chomsky often points to. Transformational rules map one syntactic representation, D-structure, to another, S-structure, typically by way of moving constituents. Interestingly, it appears that all such rules are "structure dependent." That is, they make reference to the relative structural positions of the moved thing and the position it is moved to. They don't, for example, make reference to points in a string on the basis of their position relative to some numerical count of formatives. Thus "Wh-Movement" moves maximal projections that meet certain criteria to particular positions in a phrase marker. And this operation is governed by a set of constraints that make reference to the relation between these points solely in terms of structure. There is no rule, for example, like Wh-Movement but which affects terms based on how far apart they are numerically. Thus, the learning procedure will never have to entertain the hypothesis that G_L should contain such rules.

In both cases, the classic argument for distinguishing the inviolable constraint from the simplicity metric follows very closely the logic of the poverty of stimulus argument. Because it is difficult to see (maybe even provably impossible) how such things as $\overline{X}$ Theory or structure dependence could be learned, they must belong to the features that define the universe of Gs. And because they are overarching properties of the rules in some G_L, they also have the right form to be inviolable constraints.

There is another argument towards the same end which has gained increasing influence in the last couple decades; and this one comes to us through the narrowly linguistic study of language typology, and only tangentially from learnability considerations. I will call it "Humboldt's argument," though it no doubt has an earlier champion. Humboldt's argument is based on the observation that there are certain properties that appear to hold true of all G_Ls. This can be explained, Humboldt argues, only if the universe of Gs is constrained to just those which have the relevant, universal, properties. Like Chomsky, Humboldt relates this to the construction of the mind, and uses the language of learnability in his account. He puts it this way:

> Since the natural inclination to language is universal to man, and since all men must carry the key to the understanding of all languages in their minds, it follows automatically that the form of all languages must be fundamentally identical and must always achieve a common objective. The variety among languages can lie only in the media and the limits permitted the attainment of the objective.
>
> (von Humboldt 1836)

(One might read the last sentence of this passage as making the distinction, touched on above, between aspects of Universal Grammar ("the media") and the limits our cognition places on exploiting UG ("the limits permitted the

attainment of the objective").) So, like Chomsky, he supposes that there is a Universal Grammar, a feature of the mind, which constrains the form that languages may have. But his perspective is different from Chomsky's. He expresses the notion of Universal Grammar not in terms of learning theory, or through the glass of the Poverty of the Stimulus argument, but from the perspective of language variability. He links limits on language variability to a universal ability he sees in human psychology to acquire a language.

Humboldt's goal is an explanation for the observed limits in variability of the grammars of extant languages. One might imagine that there are explanations for these limits that do not involve, as Humboldt proposes, constraints imposed by human psychology. Similarities in extant languages might reflect their common ancestry: if all languages descend from a common one, then features that are shared among them could simply be vestiges of the ancestral language that historical change has left untouched. This is the thesis of monogenesis. I think it's possible to read Sapir as advancing this alternative. Sapir is commonly associated with the position exactly opposite to Humboldt's; in Sapir's words:

> Speech is a human activity that varies without assignable limit as we pass from social group to social group, because it is a purely historical heritage of the group, the product of long-continued social usage.
>
> (Sapir, 1921, p. 4)

But, perhaps because of his vagueness, it's possible to credit Sapir with a more sophisticated view. One that assigns the universal properties of languages to the detritus of historical change:

> For it must be obvious to any one who has thought about the question at all or who has felt something of the spirit of a foreign language that there is such a thing as a basic plan, a certain cut, to each language. …Moreover, the historical study of language has proven to us beyond all doubt that a language changes not only gradually but consistently, that it moves unconsciously from one type towards another, and that analogous trends are observable in remote quarters of the globe.
>
> (Sapir, 1921, pp. 120-121)

Perhaps the common properties of extant (and known) languages are a function of two facts: all languages descend from a common language, and the forces that cause languages to change are not fully random — they preserve certain features and change others only according to some "basic plan." If historical relatedness is to explain the common traits that extant languages have, some limit must be placed on how languages change and diverge. Otherwise, language change would act as a kind of randomizer that, over time, would destroy the limits in variability that we observe. Mongenesis needs to be coupled, then, with a theory of diachrony that characterizes the limits it imposes on change. Could it be, then, that the similarities in languages are all due to these laws of diachrony?

This seems to me to be a coherent account for language variability. But it may be just a disguised version of the Chomsky/Humboldt hypothesis that the limits of human cognition are responsible for the constraints on linguistic variation. The thesis of monogenesis entails that language variation is solely the product of historical change, as Sapir's quotes makes clear. So we expect that languages vary in features which historical change can affect, but will remain similar in those ways that are immutable. Which of the features appear as language universals, then, is determined by the internal mechanisms of historical change, and the limits thereon. What are the internal mechanisms of historical change? The only proposal I know of is that historical change is a by-product of language acquisition. It is the accumulation of the small mismatches in G_Ls that successive generations of language acquirers select. Language acquisition, the poverty of the stimulus argument tells us, is guided by Universal Grammar. So even granting the diachronic argument for language universals, we see that as historical change weeds out the mutable properties from the immutable ones, the properties it leaves are those that characterize Universal Grammar. The antidote for the argument I have blamed on Sapir, then, involves bringing the poverty of the stimulus argument into play. I don't know if Humboldt's argument can stand against this alternative unaided.

But even if it can't, it provides us with another way of viewing how to factor out the components of the evaluation metric. Following the logic of Humboldt's argument, what we expect is that language comparison should give us a means of separating inviolable constraints from the evaluation metric. The inviolable constraints will be (among) those things found in all languages; the differences in languages are to be credited to the evaluation metric. Put somewhat differently, an explanatory theory is to give us both how languages cannot be constructed, and how their construction can vary. The data it must fit, then, emerges only once languages are compared: for not only does this allow the universals to be clearly discerned, but it is only through this means that the particulars of language variation are known.

When this method of factoring out the universals in Gs is followed in earnest, a rather different picture of various G_Ls emerges; and a very different conception of the language acquisition procedure becomes available. This course is meant to illustrate these emerging pictures in detail.

The evidential basis of syntactic theory

If linguistics is one part of the study of human cognition, in the sense just described, then syntax can be described as that subdiscipline of linguistics which seeks to discover what speakers know about how to arrange the words of their language into meaningful sentences. Because speakers are not conscious of the principles that characterize this knowledge, the syntactician must make recourse to indirect means of determining these principles. The syntactician's first task, then, is to determine how to find evidence that reflects the nature of this knowledge.

One plausible source of relevant information comes from observing how speakers put this knowledge to use. We could, for instance, collect the utterances from some speaker and look for generalizations in these utterances from which evidence about the underlying knowledge-base can be gleaned. This is rarely done, however, as there are few instances of such collections that arise naturally, and to assemble them from scratch is onerous enough to have been avoided. With the exception of studies of prodigious literary figures, there are vanishingly few attempts at linguistic studies that go this route.

More common is to study the linguistic utterances of a group of speakers. This is standardly done by using the dictionary maker's device of combing texts and newspapers for examples. There are several excellent "parsed" corpora of this sort,[1] and even corpora of spoken utterances[2] can be found. With the advent of the World Wide Web, it has become possible to search a very large collection of sentences, and more and more linguists are availing themselves of this resource. This technique has the unique advantage of allowing one to determine frequencies as well. It is possible, for example, to judge how rare some particular arrangement of words is relative to some other, or to find statistically significant correlations between, say, the position of an argument relative to its predicate and the person or number marked on that argument. Some linguistic theories are specifically designed to model these sorts of frequency data.

There are some serious pitfalls to using group corpora, however. One is simply that it obliterates differences among speakers and treats the data as if it were all manufactured by the same grammatical system. Since nothing is known about the producers of these sentences – they may include speakers of different dialects and speakers for whom the language in question is non-native or has been influenced by another language, for instance – this could be a serious source of error. Without some measure of the heterogeneity of the speakers who produced the corpus, it is very difficult to judge how faithfully it represents the syntactic knowledge of any one of those speakers.

Another shortcoming is that linguistic behavior, even of one individual, is not a faithful projection of the knowledge that that individual has of his or her language. People say sentences whose syntactic form is at odds with what they would otherwise deem well-formed. A significant proportion of any corpus could be made up of such "mistakes," and indeed it would be prudent to assume so, given the degree to which misshapen sentences populate the utterances of such well-placed contributors to corpora as George W. Bush. There is a distinction between a speaker's linguistic "performance" and his or her linguistic "competence," to use the names Chomsky gives to this distinction. Corpora level this distinction.

For these reasons, then, group corpora contain an unknown amount of data that should be weeded out. They contain examples of sentences that are produced by speakers whose grammatical systems differ, and they contain sentences that are not representative of any grammatical system. But group

[1] See Marcus et al. (1993), for example.
[2] See Godfrey et al. (1992).

The papers in Bod et al. (2003) have some recent examples of statistically based corpora studies, and the work of Paul Boersma (e.g., Boersma and Hayes 2001) contains a theory that is designed to model statistical data of this sort.

corpora are not only noisy with error, they are also mute about certain kinds of information.

One important piece of evidence that corpora cannot provide concerns where speakers draw the line between impossible and possible forms in their language. This distinction is easiest to elicit in linguistic domains where there are a comparatively small number of relevant forms. For example, the morphological and phonological inventories of any one speaker at any one time is reasonably small and it is therefore salient when a novel morphological or phonological form is introduced. For many such novel forms, speakers are capable of distinguishing those that are admissible members to their languages and those that are not. Most English speakers I have asked, for instance, can tell that *blick* ([blɪk]) is an admissible addition to their lexicon but that *bnick* ([bnɪk]) is not. Presumably this ability to distinguish admissible from inadmissable forms is due to the knowledge speakers have of their language, and so it is an important piece of information about how that knowledge is constituted. A typical way of characterizing this distinction goes as follows. The phonology of a language permits many forms that are not exploited by the lexicon of that language (e.g., [blɪk]). Which of these forms are used and which are not is completely extragrammatical. By contrast, because the phonology of a language limits the forms that are available to that language (e.g., English prevents the onset cluster [bn]) these forms (e.g., [bnɪk] in English) will be blocked from its lexicon. The absence of these forms is determined by the grammar; they are said to be "ungrammatical," and when they are cited, they are prefixed with the diacritic "*" to indicate their status.

The same distinction can be elicited for sentences, although because of the larger number of forms involved it is more difficult to recognize a novel sentence. Consider, by way of illustration, the pair of sentences in (1).

(1)   a. Whenever the earth revolves around its equator, the moon begins to rotate about its axis.

    b. Whenever the earth revolves around its equator, the moon begins itself to rotate about its axis.

I judge (1b) to be an impossible English sentence, and (1a) to be a possible one. Because I read very little science fiction, I think it's likely that both sentences are novel for me, but I do not have the certainty about this that I have about *blick* and *bnick*. I recognize that there are considerably more sentences that I have encountered than there are words I've encountered, and consequently I also recognize that it is likelier that I will mistake a sentence as novel than it is that I will mistake a word as novel. Nonetheless, most linguists would agree that the contrast in (1) is of the same kind that distinguishes *blick* from *bnick*. It does seem unlikely that the distinction could be reduced to one of novelty. After all, I am roughly as certain of the novelty of (1a) as I am of the novelty of (1b) and yet this does not affect the strength of my judgement concerning their Englishness. It seems probable that my ability to judge the

difference between (1a) and (1b) traces back to an ability my syntactic knowledge gives me to judge well-formedness.

This distinction between grammatical and ungrammatical forms is important because it seems to tap directly into a speaker's linguistic knowledge. Studying corpora cannot provide what is needed to see this distinction; corpora conflate ungrammatical and grammatical but non-occurring forms. For this reason, and because of its noisiness, I will not use data from corpora in these lectures. But do not forget that corpus studies, and so far as I know only corpus studies, can provide statistical data, for this might be an important resource in forming a complete model.

Instead, the central piece of evidence used in these lectures will be elicited grammaticality judgments. This has become the standard tool for syntactic analysis, and much of the literature relies on it. Elicited grammaticality judgments have their own shortcomings. There are special problems attendant with grammaticality judgments of sentences. Because sentences are very complex objects, and are frequently longer than the small memory buffer that our on-line processors are equipped with, there are failures of sentence processing that might easily be mistaken for judgments of ill-formedness. A famous example meant to be illustrative of this distinction comes from strings that are ambiguous with respect to the placement of some late occurring phrase. The pair of sentences in (2) illustrates.
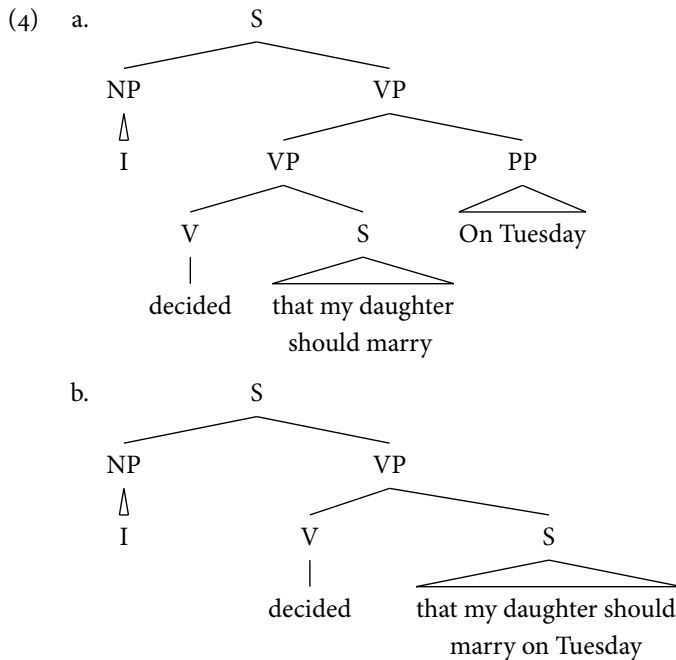
(2)   a. I decided to marry on Tuesday.

   b. I decided that my daughter should marry on Tuesday.

Upon reflection, most speakers will recognize that (2a) has two meanings. It can assert that the time of my decision to marry was Tuesday, or it can assert that what my decision was was to marry on Tuesday. As we will see, this ambiguity reflects the fact that (2) maps onto two sentences, whose difference in syntactic structure is responsible for the two meanings. The first meaning corresponds to a structure which groups the words as sketched in (3a), whereas the second interpretation corresponds to the syntactic structure shown in (3b).

(3)   a.

```
              S                          b.           S
         /         \                              /        \
       NP           VP                           NP          VP
       |                                         |          /    \
       Λ         /        \                      Λ         V        S
       I       VP          PP                    I      decided    |
            /       \     /    \                           |    /      \
       decided to marry  on Tuesday          decided  to marry on Tuesday
```
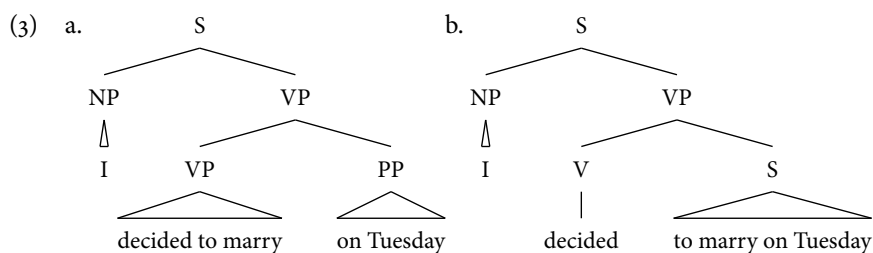
Unlike (2a), (2b) seems to have only the second of these two meanings. It can assert that my decision was for my daughter to marry on Tuesday, but it does not seem to say that the time of my decision was Tuesday. At present, this difference in (2a) and (2b) is thought to be due to constraints of sentence

processing, and not the well-formedness conditions of sentences. The relevant difference between these examples is the number of formatives between the word *decided* and the prepositional phrase *on Tuesday*. As that number grows beyond what can be held in working memory, the processor is forced to start making decisions about how to parse the initial portions of the string. These decisions favor a parse in which later material is made part of more deeply embedded phrases. Thus, in the case of (2b) it favors the structure in (4b) over that in (4a).

(4) a.

```
              S
          /       \
        NP         VP
        Λ        /     \
        I      VP        PP
             /    \     /    \
            V      S   On Tuesday
            |     /  \
         decided that my daughter
                  should marry
```

b.

```
              S
          /       \
        NP         VP
        Λ        /     \
        I       V        S
                |      /    \
             decided that my daughter should
                        marry on Tuesday
```

On this account, then, it is not that there is a difference in the syntactic well-formedness conditions which causes speakers' differing judgments about (2a) and (2b). Instead, because of the relative difficulty that (2b) presents to the on-line processor, one of the syntactic representations associated with this string (i.e., (4a)) becomes difficult to perceive. This effect of the on-line processor is what Kimball called "right association."[3]

In general, judgments of well-formedness will not be able to distinguish those sentences that do not conform to the constraints of the grammar from those that do conform to those constraints but present problems for the on-line processor.[4] There is no simple way of distinguishing these cases; they can be separated only through analysis. In the case of (2), the decision that the effect is not grammatical but, instead, the result of the processor comes partly from finding no good grammatical way of distinguishing the cases and partly from finding that manipulating factors relevant for the processor determines whether the effect materializes.

Another similar difficulty involves the fact that the meanings which sentences convey are typically bound to the context of a larger discourse. In-

[3] See Kimball (1973), Frazier (1978) and Gibson (1998).

[4] Chomsky and Miller (1963) is an early, and still useful, examination of this distinction.

evitably, then, grammaticality judgments are going to be confounded with whether or not there is a discourse in which that sentence could function. Suppose, for instance, that you are trying to determine the distribution of a process called "VP Ellipsis," which allows a sentence to go without a normally required verb phrase. VP Ellipsis is responsible for allowing the bracketed sentence in (5) to go without a verb phrase in the position marked "Δ."

(5)   Jerry annoyed everyone that [$_S$ Sean did Δ].

If you expose English speakers to the examples of VP Ellipsis in (6), you may find that they judge them ungrammatical.

(6)   a.   * Whomever she did Δ got better.

       b.   * Everything for her to Δ was hard.

One might be tempted by these examples to the hypothesis that VP Ellipsis is blocked within subjects. But if the examples in (6) are embedded into an appropriate discourse, English speakers will find (6a) well-formed while (6b) remains ungrammatical.

(7)   a.   Whomever Sally didn't tutor got worse but whomever she did Δ
             got better.

       b.   * Everything for him to do was easy and everything for her to Δ
             was hard.

The problem with (6a) is that recovering the meaning of the elided VP cannot be done without a larger context, and the grammaticality of sentences with VP Ellipsis in them depends in part on recovering the meaning of the elided VP. There is nothing syntactically ill-formed with the VP Ellipsis in (6a), however, as we see when this context is provided. By contrast, neither the context in (7b) (nor any other that I have found) improves the goodness of (6b). There is something ill-formed about the syntax of this example.

These two problems are similar. In both, the difficulty is in distinguishing judgments of ungrammaticality from other types of ill-formedness. The effect of these difficulties can be lessened if the following two practices are used in eliciting judgments.

First, embed the sentences whose well-formedness you wish to determine in discourse contexts that make the meaning these sentences should have available and salient. This helps remove the second problem.

Second, for every sentence you suspect to be ungrammatical, present your informant with a matching sentence which you suspect to be grammatical. These two sentences – the suspected grammatical and the suspected un-grammatical one – should differ minimally. Your aim should be to remove all differences between these two sentences except for the factor that you suspect is responsible for the ungrammaticality. This will help mitigate processing ef-fects, as the two sentences will end up matched in length and close to matched in complexity. It will also help remove any other confounds which might be responsible for the ungrammaticality of the sentence you wish to test.

These practices are rarely used, unfortunately. As a result, the history of syntactic research is littered with dead ends and wrong turns that have resulted from errors in the empirical base. Don't fall victim to these errors. Wherever you can, follow the Two Laws of Elicitation.

(8)   Two Laws of Elicitation

    a.   The sentences for which you elicit a grammaticality judgement should be embedded in a discourse that makes the meaning that sentence would have salient.

    b.   Every suspected ungrammatical sentence should be part of a minimal pair, the other member of which is grammatical.

In these lectures, I will sometimes violate (8a) whenever I haven't found a context that improves an ill-formedness judgement. In these cases, my guess is that the ungrammaticality of the sentence is not tied to its information content. Similarly, I will occasionally fail to give a minimal pair when I feel that the ungrammaticality of the sentence involved is dramatic enough to be obvious. For instance, examples such as (9) are so clearly violations of English sentence structure, that I cannot imagine a discourse context that could improve them, nor would minimally contrasting grammatical examples help remove the possibility of a processing effect.

(9)   a.   * Many happy the puppies barked.

    b.   * She talked people to.

    c.   * He ate should apples.

I do this partly because it will make the exposition cleaner, but obviously also because I am lazy. It would be wise to maintain a healthy skepticism about the data I present when I've taken these shortcuts.

    There is one last danger in relying on elicited grammaticality judgments, and it is the mundane and familiar one of introducing bias. It is a commonplace among experimental psychologists that eliciting psychological data can involve very subtle ways of introducing bias. Whenever the judgments are less clear than obvious cases like (9), the syntactician should clearly not rely on her or his own judgments. In these cases only judgments elicited from naïve informants will do. And in eliciting those judgments, the syntactician should adopt some of the techniques developed by experimental psychologists. Produce a survey of examples that include the sentences you wish to find judgments for but include irrelevant "fillers" as well. Those sentences should be crafted in accordance with the Two Laws of Elicitation. Then present this survey to a number of speakers native in the relevant language, controlling as best as possible for dialect variation. Finally, present the items in the survey in a randomized order, mitigating any bias that the order of presentation might introduce. When reporting data, you should also report the number of informants you have used, and make a note of any variation in the judgments you have encountered. While these safeguards wouldn't satisfy

I will also violate (8) when I am reporting data from the literature in which (8) have not been followed.

the rigorous numerical criteria of the experimental psychologist, they will go a long way towards removing error and making the data you report comparable to the data someone else gathers.

Grammaticality judgments, then, will be the central evidence used here in uncovering the principles that constitute a speaker's syntactic knowledge.

Phrase Structure

Now that we have something of a setting for our inquiry, it is time to carve out that portion of linguistic theory that belongs to syntax. I think a rather conventional definition of syntax might be (1).

(1) <u>Syntax</u> is the study of those principles and processes responsible for organizing words into sentences.

Like many definitions of areas of study, this one uses terms that are the subject matter of the study. We need definitions of "word," "sentence" and "organizing," and syntax is partly responsible for delivering these definitions. We can start with some simple concepts of what these terms refer to and work from there.

For "word," let us start with the definition that orthographic practice provides. If it's an entry in a dictionary, it'll be a word. We'll soon see that this is probably too narrow a definition, but it will get us started. For "sentence," we should aim for those sequences of words that discourses are made up of. Again, orthographic practice gives us a starting point. Those strings of words that a language pundit would acknowledge as being able to punctuate with a period will count. There are a variety of examples that have an uncertain status. For example, the utterances by "B" in the following exchanges are incomplete in other contexts.

(2) A: Jon has probably read this book.
    B: In fact, he hasn't.

(3) A: Has Jon read this book?
    B: No.

(4) A: What has Jon read?
    B: This book.

But in these contexts we will treat them as sentences. We'll need to understand, of course, why the conversations they are part of allow them to be understood as complete sentences. And finally, for "arrangement" we can begin with simple linear order. We will start, therefore, with the goal of modeling how words are linearly ordered to produce grammatical sentences.

Beyond simple grammaticality judgments, there is one other kind of datum that is important to the syntactician. As we witnessed in the previous chapter,

the syntax of sentences is intimately tied to the meanings they convey. It is the semanticist's job to discover the principles that allow users of language to extract these meanings from sentences, but their task and ours can be usefully done in tandem. Over a wide range of cases, the meanings conveyed by sentences appear to be derived, perhaps by general processes, from the meanings of the words they contain. This is a trivial observation. The kind of individual that each of the sentences in (5) is about depends on the meaning of the second word.

(5)   a.   A woman jumped.
      b.   A child jumped.
      c.   A puppy jumped.
      d.   A grasshopper jumped.

But what these sentences say about that kind of individual stays the same. It would vary if the last word varied, as you can see by replacing *jump* with some other verb. This observation about the meanings of sentences is known as the Law of Compositionality.

(6)   The Law of Compositionality
      If the meaning of a collection of formatives, α, is derivable from the meanings of those formatives then α's meaning is compositional. The meaning of sentences is compositional.

A working hypothesis is that there is a regular and productive relationship between the syntactic structure of a sentence and the ways its meanings are composed. To the extent that this is correct, then, we can use evidence about how meanings compose to investigate our hypotheses about how the syntax arranges words.

It should be noted that the "formatives" named in the Law of Compositionality do not match up with the "words" over which we have defined syntax. There are cases where strings of words seem to have a meaning that is not compositional. One famous example of that sort are idioms, such as (7).

(7)   Professor Romero is barking up the wrong tree.

bark up the wrong tree: "to promote or follow a mistaken course (as in doing research)," www.merriam-webster.com.

The string *barking up the wrong tree* has two meanings. One is compositional, and involves an action and a tree and is thoroughly uncharacteristic of Professor Romero. The other is not compositional, invokes a particular kind of mistake but no tree, and is also thoroughly uncharacteristic of Professor Romero. Here, then, we have words that syntax has arranged but from which a compositional meaning is not derived. Similarly, there are certain morphemes that do not fit our criteria for words but which compose semantically in a way that fits the Law of Compositionality. One of these is the plural suffix found on many English nouns. The meaning of *bees* is, apparently, derivable from the meanings of *bee* and *s*, for instance.

Here, then, is our specific task: discover the principles that underly language user's ability to recognize grammatical arrangements of words into sentences. What constitutes a grammatical sentence will vary from language to language, of course, and, less saliently, from one idiolect to another. We'll want a model of this variation as well, for the reasons sketched in the previous chapter. We must start with one grammar, of course, and because I can assume the readers of these notes to have some command of English, I will start with that language.

Our first observation is that we can get very far in this task using very little particular information about the words involved. A great deal about the processes that determine well-formed arrangements of words can be characterized using nothing more than the morpho-syntactic "category," or "type,"that the words belong to. This can be appreciated by virtue of the "Novel Form" argument, which is laid out in (8).

(8)   a. If enough information is introduced with a novel word to enable the individual learning that word to recognize its category, then

        b. The individual knows which arrangements it can grammatically combine in.

        c. Hence, it must be category membership to which these processes refer.

To see this argument in action, let me introduce you to the word [blow.ɹsnɪk] ('bloresnick'). Here is an example of *bloresnick* in a sentence.

(9)   Many bloresnicks are grey.

If I tell you that (9) is grammatical, you will also be able to determine that only (10b) of the examples in (10) is grammatical too.

(10)   a. It ran bloresnicks the tree.

        b. He removed the long bloresnicks.

        c. She finds Sammy bloresnicks.

        d. He made his face bloresnicks.

The position that *bloresnicks* has in (9) is enough for you to know a lot about which positions it can, and cannot have, in English. It appears, then, that words are sorted into kinds and the grammatical arrangements of words can be captured to a certain extent as the grammatical arrangements of these kinds. These kinds are morphosyntactic categories, and many of them are familiar from the folk grammar we are taught in school: noun, verb, preposition, adjective and adverb. The sentence in (9) is apparently sufficient for a knower of English to deduce that *bloresnick* is a noun. And knowing that it is a noun is sufficient to know which of the sentences in (10) is grammatical.

A slightly different way of seeing the existence of morphosyntactic category is to consider the following, closely related, experiment. I will now introduce you to another novel word "pondel," but this time by way of a definition.

(11)   *pondel:* ([pɑndl̩]) unwanted facial hair.

With this knowledge, an English speaker is now equipped with judgements about the grammaticality of many sentences containing *pondel*, including those in (12).

(12)   a.   Many pondels are grey.

   b.   He ran pondel the tree.

   c.   He made his face pondel.

   d.   He removed the long pondel.

   e.   She finds Sammy pondel.

   f.   He made his face pondel.

Only (12a) and (12d) are grammatical. Because we have here gone from information about the meaning of a word to knowledge about its possible positions in a grammatical sentence, we can conclude that there is some relationship between meaning and what I have called morphosyntactic category. At present, however, there is no consensus about how to recast morphosyntactic categories into wholly semantic terms and, given the differences in how languages seem to map meanings onto categories, there are real problems to overcome in doing so. It's also not clear that the pattern of grammaticality judgments English speakers give to (10) rests on differences in how their meanings are arrived at. Why can't the meanings of *it* and *bloresnick* in (10a) combine in the same way that they seem to in (13a), for instance? And why can't (10c) get a meaning like that assigned to (13b)?

(13)   a.   It was bloresnick.

   b.   She finds Sammy to be bloresnick.

The morphosyntactic classes that words belong to cannot be recast in wholly semantic terms then. Throughout this work, I will assume that there is some weaker correlation between semantic types and morphosyntactic categories. It is that weaker correlation that is responsible for our ability to go from the meaning of a word to knowledge about its positional distribution in grammatical sentences. What we will focus on, then, is how grammaticality judgements can be modeled by principles that make reference only to the morphosyntactic categories of the words in those sentences.

Substitution Classes

Our lexicon of words is partitioned into sets — categories — and some of our knowledge about which groups of words are grammatical is based on membership in these sets. We can use the traditional grammarian's terminology for these sets. *Bloresnick*, for instance, is a noun; *find* is a verb; *long* is an adjective, and so on. A string that is made up of a noun followed by a verb followed by a noun is judged to be a grammatical sentence in English (witness (13a))

whereas a string made up of two adjacent nouns is not (compare (10a)). If we were to look at a large assortment of strings, we would be able to discover patterns of this sort that distinguish the grammatical sentences from the ungrammatical sentences. We would discover that categories have a particular distribution in the grammatical strings.

Zellig Harris argued that morpho-syntactic category should be defined in just these terms.[5] Specifically, "noun," "verb" and so on are "substitution classes" of vocabulary items. They are substitution classes in the sense that there is a set of positions within a sentence into which any member of that class can be substituted preserving the grammaticality of the sentence. For instance, any word that can be grammatically placed in the spot marked with "___" in (14) falls within the subset of vocabulary items we know as "nouns."

(14)   the ___ exists

This is indicated by considering the lists of sentences in (15)-(20).

(15)   The lamp exists.
          The girl exists.
          The sky exists.
          The streetcar exists.
              ⋮
(16)   * The happy exists.
          * The blue exists.
          * The short exists.
          * The flat exists.
              ⋮
(17)   * The in exists.
          * The out exists.
          * The from exists.
          * The on exists.
              ⋮
(18)   * The swim exists.
          * The have exists.
          * The ate exists.
          * The broke exists.
              ⋮
(19)   * The slowly exists.
          * The apparently exists.
          * The always exists.
          * The decidedly exists.
              ⋮

[5] Harris (1946) is an accessible introduction to this procedure.

(20)   * The every exists.

          * The much exists.

          * The no exists.

          * The a exists.

                ⋮

As can be seen, this technique picks out a list of words that match what the grammar school curriculum calls nouns, and segregates them from the others. A similarly discriminating environment can be devised for each category. For each (major) word class, I've given a distinguishing environment in (21).

(21)   a. have ___ eaten: Adverb

       b. the ___ thing: Adjective

       c. dance ___ it: Preposition

       d. in ___ orange: Determiner

       e. must ___ there: Verb

Understand (21), and (14) as well, as abbreviating the following claim: there is a sentence that is grammatical which contains "X ___ Y," and for which replacing a word of category CATEGORY into "___" uniquely preserves grammaticality. So, for instance, (21a) should be understood as claiming that all the ways of completing the sentence in (22) involve filling "___" with an adverb.

(22)   They have ___ eaten rutabagas.

On this view, morpho-syntactic categories are simply partitions of the vocabulary into equivalence classes. The labels "noun," "verb" and so on are merely convenient names for the resulting subsets of vocabulary items.

There are a few things about the distinguishing environments in (14) and (21) that should be noted. First, they define substitution classes solely on the basis of adjacent items. We might elevate this to a hypothesis.

(23)   Morpho-syntactic categories can be defined on the basis of what words they can be adjacent to.

Second, the environments in (21) partition the vocabulary in ways that your language arts curriculum may not have. For instance, the Determiner class picked out by (21d) does not include *much* or *many*. There aren't grammatical sentences that contain *in much thing* or *in many thing* as a substring. One reaction to this would be to allow *much* and *many* to belong to different word classes than *every*, *the*, *a*, and so on. We could admit the two additional word classes, Det_{MASS} and $\text{Det}_{\text{COUNT}}$, defined over the environments in (24).

(24)   a. in ___ syrup: Det_{MASS}

       b. in ___ oranges: $\text{Det}_{\text{COUNT}}$

This is a straightforward application of the procedure for defining morpho-syntactic category that Harris's program offers, and it is one direction that syntactic theorists go.

There is another reaction to these data, however, and it is the one I shall follow. It's clear by comparing the environments that define Determiner and Det$_{count}$ that what distinguishes them is whether the word that follows is count or singular. The difference between singular and count is a semantic one, and so we should tie the difference between Determiners and Det$_{count}$ eventually to a semantic primitive. It is also a semantic difference, although a less familiar one, that distinguishes the Determiner and Det$_{mass}$ categories. Words such as *syrup* refer to entities which do not contain clearly delineated atomic parts, whereas words like *oranges* do. If one recursively divides a group of oranges into its parts, there will come a definitive point — when we are down to the individual oranges — that we will no longer be dividing a group of oranges. A group of oranges is made up of things that one could count. The same is not true of syrup. It is not clear how to find the things that are grouped together to make syrup. Words that refer to entities that can be piled together, or taken apart, in the way that oranges can are called "count nouns," while those that cannot are called "mass nouns." The difference between the Determiner and Det$_{mass}$ classes is just whether the term that follows them is mass or count. This is a semantic distinction. There is a clearly semantic generalization to be captured in distinguishing these classes of determiners, and we should strive to capture these generalizations in our grammar.

There are generalizations hidden in the environments in (14) and (21) as well, but it is not at all clear that these are semantic generalizations. To see these generalizations, consider the following series of distinguishing environments for the word class 'noun,' each of which is very similar to (14).

(25) a. the ___ eats

b. some ___ knows

c. a ___ exists

d. few ___ is

e. every ___ ate

f. no ___ exists

g. some ___ has

h. every ___ put

i. a ___ screamed

j. few ___ drove

k. and so on

The generalization in this list is that the words flanking the environment in which nouns are restricted are themselves of a word class; each member of this list fits the schema in (26).

(26) DETERMINER ___ VERB

Each of the environments in (21) can be similarly converted into a generalization that makes reference to morpho-syntactic category.

The mass/count distinction is, however, still one that arbitrarily maps onto lexical items, as Brendan Gillon reminds me. English decides to treat *furniture* as a mass noun, for instance, even though a group of furniture is, arguably, composed of clearly delineated parts. The point here is that there is a way of defining the mass/count distinction that is not linguistic. It doesn't need to be rendered in terms of substitution classes. About expressions like *furniture*, we could imagine that the meaning we associate with it does not entail that it refers to something that can be divided into clearly delineated parts. That is, the meaning English assigns to *furniture* is at odds with the truth in the same way that the meaning English assigns to *water* is. The difference between *furniture* and *water*, then, is that to know this fallacy about their meanings requires more education in the case of *water*.

(27)   a. verb ___ verb: Adverb

       b. determiner ___ noun: Adjective

       c. verb ___ noun: Preposition

       d. preposition ___ noun: Determiner

       e. Infl ___ preposition: Verb

(nb: The word *must* belongs to a morpho-syntactic category with a small set of members; I've labeled it infl in (27e). We'll soon encounter the evidence for this category.) At present it is not possible to reduce this generalization to a semantic one. That is, there is no known method of defining morpho-syntactic categories in semantic terms. At present, the best that can be done is to define morpho-syntactic categories in the terms that Zellig Harris gave us: substitution classes. The generalizations underlying (14) and (21) are at present irreducibly morpho-syntactic, then.

For a recent attempt to define some of the major categories in terms that verge on semantic, see Baker (2003).

Notice that converting (21) to (27) claims that the particular lexical items chosen will not matter. But, as we've just seen, it does matter: whether the noun in (27d) is count or mass or singular or plural will determine which of the Determiner, Det_{mass} and Det_{count} classes are well-formed in this position. To take the step that defines substitution classes in terms of other substitution classes, then, requires factoring out the semantic information and introducing, as a consequence, certain ungrammatical strings.

One reaction to the differences among Determiner, Det_{mass} and Det_{count}, then, is to segregate the kinds of information that together determine the distribution of words into a syntactic component and a semantic component. This is the path we shall take. We assign to the semanticist the task of explaining the wholly semantic part of this job: why, for instance, *much* can be left-adjacent to a mass noun but not a count noun. In general, it is not trivial to know when it is the semantics or the syntax that is responsible for coöccurence restrictions like those in (14) and (21), and the line is constantly being questioned. Harris, it seems, believed that virtually none of it was semantic, whereas present-day categorial grammarians push in the direction of removing an independent syntactic contribution. I'll chart a course that is somewhere in the middle.

Morpho-syntactic categories, then, are defined syntactically. They are subsets of the vocabulary that can be substituted for each other in particular positions within a grammatical sentence preserving grammaticality. Moreover, the particular positions can be characterized in terms of adjacent morpho-syntactic categories. The first step in characterizing the grammaticality judgments of some speaker is recognizing that the vocabulary of that speaker is partitioned in this way.

Phrases

We have an intuition that sentences contain subgroupings of words. That intuition lives on several vague observations. One of these is that the pronunciation of sentences comes with a prosody which, like syllables, gives the impression of subgroupings. These prosodic groupings have an effect on various aspects of the phonology of sentences. They seem to influence the pitches and volume, and also the rate, of speech. For instance, a natural pronunciation of (28) can introduce a slight pause after *cracker*, and slow the relative duration of [kɹækɹ̩].

(28)   I can't eat every cracker today.

     a.   I can't eat every CRACKER | today.

     b.   * I can't eat EVERY | cracker today.

I've indicated that pronunciation in (28a) by expressing the relatively slowed articulation of *cracker* with all caps, and the pause with a pipe. Note that (28) does not permit a natural pronunciation of the sort indicated by (28b), where *every* is slowed and followed by a pause. The slowdown plus pause is thought to be permitted only at the right edge of prosodic phrases. There is a prosodic phrase that ends with *cracker*, but not one that ends with *every*. Phenomena like these, then, suggest that there are subgroupings within sentences.

    A second vague observation that speaks on behalf of the existence of subgroupings is semantic. Certain substrings of a sentence can be associated with a meaning in isolation, and others cannot. For instance, there is a difference in the meanings attached to the two strings in (29), both parts of (28).

(29)   a. cracker today

       b. every cracker

(29a) is just a (short) list of words. (29b) has a meaning. Indeed, it seems to obey the Law of Compositionality. Some groups of words seem to be compositional in this way while others do not.

    A working hypothesis is that the subgroupings which the Law of Compositionality is sensitive to and the subgroupings that prosody is sensitive to are one and the same. There is some superficial evidence in support of this view. Consider, for example, the sentence in (30).

(30)   She ate the cracker in the box.

There are two ways this string can be semantically partitioned. In one, *the cracker in the box* is associated with a meaning; it refers to an object that (30) reports was eaten. In the other grouping, *the cracker* and *in the box* are independent; their meanings are not composed into the meaning that *the cracker in the box* has. Under this second grouping, (30) reports that the cracker was eaten in the box. These two semantic groupings can be associated with two prosodic groupings. If *the cracker in the box* makes a semantic unit,

To get a feel for what is involved in locating prosodic groupings, see Liberman and Prince (1977), Nespor and Vogel (1986), Beckman and Pierrehumbert (1986), Pierrehumbert (1980), Selkirk (1980, 1986) and Ladd (1996).

(29) can be spoken in the way indicated in (31a) but not (31b). If, by contrast, *the cracker* and *in the box* make independent semantic units, the prosody in (31b) is preferred instead.

(31)   a.   She ate the cracker in the BOX |.

b.   She ate the CRACKER | in the box.

The prosody in (31b) corresponds to one in which a prosodic phrase ends at *cracker*, whereas the one in (31a) does not. If the prosodic and semantic groupings are the same, then this is what's expected. For *the cracker in the box* to form a semantic unit, it cannot end at *cracker* as in (31b). But if *the cracker* and *in the box* do not form a semantic unit, then the juncture between these groupings in (31b) is expected.

In addition to these criteria for subgroupings, there are wholly syntactic phenomena that suggest a similar subgrouping of the words in a sentence. Discovering the principles that determine these syntactic phrases will be a central task of these lectures. As with the groupings that the semantic and prosodic phenomena pick out, a working hypothesis is that syntactic phrases are the same thing. I'll adopt this hypothesis too.

(32)   The subgroupings that the words of a sentence divide into are the same for its syntactic, prosodic, and semantic organization.

It is not at all clear that (32) is true; there are many apparent mismatches to understand if it is true. Nonetheless, this is a useful starting position. Indeed, we will adopt a slightly stronger hypothesis:

For some of these problems, see Selkirk (1984, 1996), Truckenbrodt (1995, 1999), Zec and Inkelas (1990) and the papers in Inkelas and Zec (1990).

(33)   The syntax of a sentence produces phrases, and the phonological and semantic groupings make use of these phrases.

(33) expresses a strategy for explaining (32), should it be true. Syntax will include a procedure that forms phrases, and these phrases are used by the phonological and semantic components to assign a prosody and a meaning to the sentence. That image of syntax as a intermediary between the phonological and semantic interpretations of a sentence will recur throughout these lectures. A presently popular thesis is that many of the properties that syntax has can be explained by its role as intermediary between the phonological and semantic manifestations of a sentence. In the case of phrases, for instance, one might imagine that the pronunciation of a sentence requires an ordering of the words it contains into phonological groupings. This is a requirement that the principles of phonology (and phonetics) alone impose. Similarly, one might imagine that for a sentence to convey a meaning, its words must be grouped for the semantic rules to be able to deliver that meaning. This requirement is a wholly semantic one; perhaps it is whatever underlies the Law of Compositionality. The "function" of syntax, we might imagine, is to deliver these groupings. Syntax could be defined as that portion of the phonological and semantic components that is shared. This is a view of the place of syntax

in the larger linguistic panoply that Noam Chomsky has advanced in the last twenty years. He sometimes suggests that syntax is <u>only</u> those portions of the other "interpretive" components that are shared. This is a guiding idea behind the so-call "Minimalist Program," one of several current trends in syntactic theorizing.[6] This will be an idea that figures largely in what we will do in these lectures as well.

[6] Chomsky, Noam. 1995. *The minimalist program*. Cambridge, Massachusetts: MIT Press

(34)   The Minimalist Hypothesis
       Syntax is made up only of those mechanisms necessary to mediate between the Phonology and Semantics of sentences.

What, then are the syntactic criteria for recognizing phrases? One proposal is to extend the process used to locate the morphosyntactic classes of words. Like nouns, verbs and the like, certain strings of categories also have a distribution within sentences that can be defined in terms of adjacent items. For example, the string D(eterminer)+Adj(ective)+N(oun) can appear immediately after a preposition and immediately preceding the *'s* which marks the "possessive."

(35)   ___ *'s* & P ___ : D+Adj+N
       a.   I talked to **the happy woman**.
       b.   **the happy woman**'s friend

This string can also be "coördinated" with another identical string of categories. Coördination involves the use of words called "conjuncts," words such as *and*, *or*, *nor*, etc. Thus, we find examples like (36) but not (37).

(36)   the happy woman and an unhappy man

(37)   a.   * the angry and an unhappy man
       b.   * the and an unhappy man

Finally, with respect to all these distributional tests, the strings D+N+P(re-position)+N, N+P+N, Adj+N, N, and (infinitely) many others also pass. We need some way of describing the fact that these strings are "the same," and different from, say, P+N which has a distinct distributional pattern. That is, this family of strings is a substitution class in the same sense that morpho-syntactic categories are.

Families of strings like this are called "phrases," and we can write a Phrase Structure Rule to describe which strings belong to such a family. In the case at hand, this rule might look like (38).

(38) αP → (D) (Adj) N

Understand material enclosed within "( )" to be optional; (38) therefore generates the set of strings: D+Adj+N, D+N, Adj+N and N.

This leaves out the strings D+N+P+N and N+P+N. But these strings involve another phrase, made up of the string P+N. This string, along with any string that conforms to the template P or P+αP or P+P or P+P+αP has the defining distribution in (39).

(39)   A ___ & P ___
   a. I stood **around**.
   b. I knew the man **by Mary**.
   c. I remain disliked **by Mary**.
   d. I stood next **to Mary**.

Like αPs, βPs may be coördinated with other βPs, but not with other αPs, as the following examples illustrate.

(40)   a.   Under the bed and behind the sofa are usually good places to find money in my house.

   b.   * The dining room table and behind the sofa are usually good places to find money in house.

Hence, just as with αPs, this family of strings constitutes a substitution class.

   Putting these observations together, we come up with the Phrase Structure rules in (41).

(41)   a. αP → (D) (Adj) N (βP)
   b. βP → P (αP)
   c. βP → P (βP)

It is customary to collapse the two rules in (41b,c) to (42).

(42) βP → P $\left\{ \begin{array}{c} (\alpha P) \\ (\beta P) \end{array} \right\}$

Material enclosed in "{ }" offers a set of exclusive choices: exactly one of the members of the enclosed list must occur. In (42) this means that either αP, βP or, because these are both optional, nothing may occur after P to form a PP.

   Note that (41a) and (42) together have the property of being recursive. This is an important aspect of phrase structure rules for it is the primary means by which we describe the indefinite length, and cardinality, of sentences. These two phrase structure rules are able to characterize infinitely many and infinitely long strings of words. This is a correct result, for we are, as far as our linguistic abilities go, capable of forming grammaticality judgments about an infinity of sentences and about sentences of infinite length. We have other properties that prevent us from doing this, of course. Our memories and attention are too ephemeral for even very long sentences; and even if we were able to overcome these cognitive limitations, our death will bring an eventual end to any sentence, or series of sentences, that we are evaluating. But there is no reason to think that this limitation is a linguistic one. We should let our model of grammaticality judgments characterize an infinity of sentences, as well as permit sentences of infinite length, and let the actual limits on the lengths and numbers of sentences that we evaluate be determined by other factors. The recursiveness of phrase structure rules is a step in that direction.

Still another phrase structure rule is required to account for the fact that the family of strings that include V, V+αP, V+βP, V+αP+βP, and an infinite set of other such strings is a substitution class. The environment that defines them is (43).

(43)   Infl ___

    a. I should **eat rutabagas**.

    b. I will **talk to Mary**.

    c. I will **tell Mary about rutabagas**.

"Infl" is a morpho-syntactic category that includes *should, will, must, would, can, could* and a few other words. It's an abbreviation for "inflection." The words that belong to it are tied to the inflectional classes that verbs belong to, as we shall see. Like αPs and βPs, coördination treats members of this family as equivalent and distinct from αPs and βPs.

(44)   a.   Mary walked and talked.

    b.   Mary visited Paul and kissed Barry.

    c.   Mary talked to Paul and met with Barry.

These facts call for a Phrase Structure rule like the following:

(45) γP → V (αP) (βP)

We have now arrived at the three rules in (46).

(46)   a. γP → V (αP) (βP)

    b. βP → P $\left\{ \begin{array}{c} (\alpha\text{P}) \\ (\beta\text{P}) \end{array} \right\}$

    c. αP → (D) (Adj) N (βP)

There is a common property to all these rules. In each case, all of the constituents are optional, except one. Thus, a verb is the only necessary member of a γP, a noun the only requisite member of an αP and a preposition is all that's required to make a βP. This is just another way of observing that the environments that define these phrases are also environments in which a word class is defined. Further, the converse also turns out to be true: whenever there is a preposition, there is a βP, wherever a noun is found, there is an NP, as so on. More precisely, the environments that define a phrase will always include an environment that defines some category. Thus, nouns and αP, prepositions and βP, verbs and γP are in one-to-one correspondence. This is a very pervasive property of Phrase Structure rules. Phrase Structure rules vary to a considerable degree across languages, but this property of them seems to always hold. We'll confront two apparent counterexamples from English shortly, but these are probably only apparent counterexamples. So far as I am aware, there is no clear counterexample to this generalization. This property of Phrase Structure rules is known as *endocentricity*. The word that must be a

member of the phrase is its *head*. Finally, it is common practice to name the phrases after their heads, so we'll rename αP, NP, βP PP and γP VP. Thus, we now have the rules in (47).

(47)   a.   NP → (Det) (Adj) N (PP)

   b.   PP → P $\left\{ \begin{array}{c} \text{(NP)} \\ \text{(PP)} \end{array} \right\}$

   c.   VP → V (NP) (PP)

In addition to these three Phrase Structure Rules, we'll need quite a few others. Indeed, the principle of endocentricity leads us to expect that for every category, there will be a Phrase Structure rule that builds a phrase headed by that category. For example, corresponding to the category Adjective, there is a rule that builds adjective phrases; (48) is a good first approximation.

(48)   AP → A (PP)

The presence of PPs within Adjective phrases is supported by the existence of strings like:

(49)   a.   She is **interested in syntax**.
            She is **interested**.

   b.   He seems **happy with linguistics**.
            He seems **happy**.

The coördination test also treats A and A+PP strings as being the same, as (50) indicates.

(50)   a.   She is happy and interested in syntax.

   b.   He seems bored but happy with linguistics.

We'll also need a Phrase Structure rule that tells us how these various phrases are put together to form a sentence. (51) looks roughly right.

(51)   S → NP Infl VP

The morpho-syntactic category that sentences are in a one-to-one relation with is Infl,[7] and so in keeping with the convention of naming phrases after their heads, we should change (51) to (52).

(52)   IP → NP Infl VP

With this rule we have finally come to the task of characterizing the grammaticality judgments of English speakers. For any speaker of English whose vocabulary has been partitioned into noun, verb, preposition, adjective, determiner and Infl, (52), with the rules in (47), characterizes those strings of words that will be judged grammatical.

This is just a first step, of course. We have hundreds of pages left. In fact, it's possible to see something wrong with (52) right away. It says that no sentence can fail to have an Infl between NP and VP, but if Infl are just words such

[7] In this case, however, unlike what we found for the VP, NP and AP rules, Infl is not the only obligatory member of a sentence. It is presently controversial whether sentences are the only phrases that have this property. We will see in later lectures that there are syntactic reasons for the obligatory presence of NP, and likely semantic ones for the obligatory presence of VP.

as *can*, *could*, *will*, and so on this is obviously wrong. There are grammatical sentences aplenty that fail to have these words in them. (53) is an example.

(53)   Jerry walked.

Where is the Infl between *Jerry* and *walked* in this sentence?

If we look hard, we find that sentences are, in fact, in a one-to-one correlation with a category, but that category includes not just words, but bound morphemes as well. Consider the sentences in (54).

(54)   a.   Jerry leave**s**.
       b.   Sally lef**t**.
       c.   Sam ha**s** left.
       d.   Sarah ha**d** left.
       e.   Martha **should** leave.
       f.   George **might** have left.
       g.   Laura desires [Sal **to** leave].
       h.   Larry remembers [Jim leav**ing**].

The boldfaced terms have similar distributions: they are found either immediately preceding the verb (if they are free) or affixed onto the following verb (if they are bound). Every sentence has one of these, and so these terms meet the criteria of being the head of a sentence. To explain how it is that those Infls which are bound morphemes materialize affixed onto the following verb, we will have to invoke a process that goes beyond phrase structure rules. Let us put off doing this.

As we gather more detail about the shapes of grammatical English sentences, we will need to make quite a number of additions to these rules. In fact, to be exhaustive about this proves to be a task beyond what we can manage here; we should consider this an open-ended process. Nonetheless, I want to gather a little more detail than we now have.

I'll begin by adding a couple of phrases to our inventory. One of these is a sort of "sentence" found in examples like (55).

(55)   a.   Mary said **that John likes chocolate**.
       b.   Mary recalled the rumor **that John likes chocolate**.
       c.   **That John likes chocolate** bothers Mary.
       d.   Jerry is angry **that John likes chocolate**.

Note that the strings following the word *that* meet the conditions imposed by the rule that builds IPs. The word *that* is called a "complementizer" and it is the head of the phrase found in these sentences. This phrase, or clause as sentence-like phrases are often called, is a "Complementizer Phrase" (CP). CPs conform to the requirements of the following Phrase Structure rule.

(56)   CP → C IP

The conclusion that sentences are headed by Infl, and that the subordinate sentences that include *that* are headed by complementizers, is reached in Chomsky (1986a). In the generative literature that precedes this work, these two types of sentences were given various treatments. Bresnan (1972) suggested that *that*-clauses were of the same category as sentences; she called them S̄s. Jackendoff (1977) argued that sentences are larger VPs, headed by a verb. The IP/CP model is now standard, though the issue is periodically revisited.

Other complementizers are *if* and *whether*, as found in the following examples.

(57)   a.   I wonder if Mary likes chocolate.

       b.   I asked whether Mary likes chocolate.

Having introduced this constituent, we will now need to revise our previous Phrase Structure rules to include the positions where they may lie. This yields the following battery of rules.

(58)   a.   IP → $\left\{\begin{array}{c} \text{NP} \\ \text{CP} \end{array}\right\}$ I VP

       b.   NP → (D) (AP) N (PP) (CP)

       c.   VP → V (NP) (PP) (CP)

       d.   AP → A (PP) (CP)

Note the option of having a CP in place of an NP at the beginning of a sentence. Note too that I've brought the NP rule into conformity with the principle of endocentricity. Our earlier rule (i.e., NP → (D) (Adj) N (PP)) permitted an adjective without an adjective phrase. I've replaced "(Adj)" with "(AP)" to correct for this. We'll see the empirical support for that change shortly.

The second phrase we'll need are ones headed by adverbs. Adverbs are a word class that is sometimes defined on the position left adjacent to adjectives, as in (59).

(59)   a.   A **deliberately** angry child

       b.   The **noticeably** large boat

They can also be found left adjacent to verbs, as in (60).

(60)   a.   I have **deliberately** misled.

       b.   I have **noticeably** erred.

We should modify our rules for VPs and APs to reflect these possibilities, then.

(61)   a.   VP → (Adv) V (NP) (PP) (CP)

       b.   AP → (Adv) A (PP) (CP)

Once we've admitted the word class "Adverb," we might expect that there will be a phrase that has the same distribution. Anticipating the existence of these phrases, let's convert the rules in (61) to those in (62).

(62)   a.   VP → (AdvP) V (NP) (PP) (CP)

       b.   AP → (AdvP) A (PP) (CP)

       c.   AdvP → Adv

There is a special class of adverb-like words that can be found in construction with some adjectives, but with no verbs. Two examples are given below.

(63) a. A **very** angry child

 b. The **extremely** large boat

(64) a. * I have **very** misled.

 b. * I have **extremely** erred.

These words indicate an "extent" or "measure" of the thing they are modi-
fying. So, for instance, *very* indicates that the angriness of the angry child is
high. If the adjective that is to be modified is not "gradable" it cannot combine
with these kinds of modifiers.

(65) * a very absolute position

I will call this class of words "degree words." Degree words can be found inside
Adverb phrases as well as Adjective phrases. So we shall want to change our
rules to those in (66).

(66) a. AP → $\left\{ \begin{array}{c} \text{(AdvP)} \\ \text{(DegP)} \end{array} \right\}$ A (NP) (CP)

 b. AdvP → (DegP) Adv

 c. DegP → Deg

There are some other changes to the VP rule that are necessary. Note, for
instance, that VPs may occur immediately following a verb, as in (67).

(67) a. Mary has walked.

 b. Mary has talked to John.

 c. Mary has visited Gary.

Interestingly, if the verb heading a VP is followed by another VP, nothing
else may follow the head verb. For instance, *Mary has on the platform walked*
is ungrammatical. We need, therefore, to modify the VP Phrase Structure
rule in such a way that the head verb is followed by a VP, or by the expansion
previously arrived at, but no combination thereof. This can be done with the
aid of curly brackets in the following way:

(68) VP → (AdvP) V $\left\{ \begin{array}{ccc} \text{(NP)} & \text{(PP)} & \text{(CP)} \\ & \text{VP} & \end{array} \right\}$

Further, it is possible to find APs embedded within VPs; (69) provides
some examples.

(69) a. Sally remains angry at Jim.

 b. Frank is happy with himself.

When APs follow verbs, they may be preceded by, at most, a PP, as in (70).

(70) Jerry seems [PP to Bill] [AP happy with his rutabagas].

So we change the rule that characterizes VPs to:

$$(71) \quad \text{VP} \rightarrow (\text{AdvP}) \ \text{V} \left\{ \begin{array}{ccc} (\text{NP}) & (\text{PP}) & (\text{CP}) \\ & \text{VP} & \\ (\text{PP}) & \text{AP} & \end{array} \right\}$$

Finally, consider that part of the NP rule that introduces determiners. Determiners include words like *the*, *a*, *that* (not to be confused with the complementizer *that*), *every*, *some*, *all*, etc. Interestingly, it's very rare that we find determiners combining with other words to form a phrase that combines with a following noun. A couple of examples which might constitute cases of this sort are given in (72).

(72)   a.   all but three dogs

      b.   more than most people

I don't know precisely what the Phrase Structure rule is that determines which strings may stand in this position. Nonetheless, one common approach to these cases is to imagine that determiners head their own anemic phrases, which are then positioned within NPs. We will revisit this idea, but for now let's imagine that determiner phrases are made up of nothing but determiners.

(73)   DP → Det

We'll therefore need to update the phrase structure rule that forms NPs. But before we do this, let's consider strings like those in (74).

(74)   a.   Mary's book

      b.   the man's toy

      c.   the man on the table's nose

These examples involve a possessive or genitive phrase. This phrase is an NP with the morpheme *'s* appended to the end. Further, note that this genitive phrase never co-occurs with a DP, as (75) illustrates.

(75)   a.   * the Mary's book

      b.   * the the man's toy

      c.   * a the man on the moon's nose

One very typical explanation for this is to understand determiners and possessives as competing for the same position. That can be done by rigging the NP phrase structure rule in such a way that it either produces a DP or a genitive phrase in the same position. This is done with the curly braces abbreviation in (76).

$$(76) \quad \text{NP} \rightarrow \left\{ \begin{array}{c} (\text{DP}) \\ (\text{NP's}) \end{array} \right\} (\text{AP}) \ \text{N} \ (\text{PP}) \ (\text{CP})$$

One final Phrase Structure rule is required by the sorts of examples we've so far reviewed. This is the Phrase Structure rule that generates coördinated phrases. This can be done with the following.

(77) $\alpha \rightarrow \alpha$ Conj α

This rule says that a phrase of any category can be made up of two other such phrases with a conjunct stuck between them. Conjuncts, recall, are *and, or* and *but*.

Summarizing, we've now introduced the following battery of Phrase Structure rules:

(78)   a.   IP $\rightarrow \left\{ \begin{array}{c} \text{NP} \\ \text{CP} \end{array} \right\}$ I VP

b.   NP $\rightarrow \left\{ \begin{array}{c} \text{(DP)} \\ \text{(NP's)} \end{array} \right\}$ (AP) N (PP) (CP)

c.   VP $\rightarrow$ (AdvP) V $\left\{ \begin{array}{ccc} \text{(NP)} & \text{(PP)} & \text{(CP)} \\ & \text{VP} & \\ & \text{(PP) AP} & \end{array} \right\}$

d.   DP $\rightarrow$ Det

e.   DegP $\rightarrow$ Deg

f.   AdvP $\rightarrow$ (DegP) Adv

g.   AP $\rightarrow \left\{ \begin{array}{c} \text{(AdvP)} \\ \text{(DegP)} \end{array} \right\}$ A (PP) (CP)

h.   CP $\rightarrow$ C IP

i.   PP $\rightarrow$ P $\left\{ \begin{array}{c} \text{(NP)} \\ \text{(PP)} \end{array} \right\}$

j. $\alpha \rightarrow \alpha$ Conj α

An interesting property of the phrases defined in (78), a property which Harris discussed, is that they all distinguish the phrases they are defining from the word class which that phrase matches. One might wonder why there is this distinction. We have decided that the way to define category is as classes of words that can substitute into a certain collection of positions. We've also used this very definition for phrases. And the principle of endocentricity tells us that there is a one-to-one correspondence between category and kind of phrase. What we end up with, then, is a situation where the positions that define a phrase are always a proper subset of the positions that define a category. We might wonder why the positions defining a phrase are always a proper subset of those that define a word class; why aren't they the very same positions? Why, for instance, isn't the rule for noun phrases something like (80) rather than (80)?

(79)   NP $\rightarrow$ (DP) (AP) N (PP)

(80)   N $\rightarrow$ (DP) (AP) N (PP)

Harris argues that phrases need to be distinguished from word class, and that this is a general property of phrase structure rules. He points out, for example, that while singular nouns are in the same substitution class as are plural ones, a plural noun cannot substitute for a singular one when it combines

with the plural morpheme. He assumes that phrase structure rules control inflectional morphology, and therefore, that in addition to the rules we have discovered, there is also a rule such as (81) that produces plural nouns.

(81)   NP → N s

(We will adopt something similar down the road.) His point, then, is that this rule should not be (82) because that would wrongly produce doubly pluralized nouns such as (83).

(82)   N → N s

(83)   kitses [kɪtsɨz]

This is the reason, then, why our phrase structure rules look like (79) and not (80). Or, to put it somewhat differently, we do not want these rules to be recursive with respect to their head. The phrases we've encountered so far all have this property.

But interestingly, it turns out that not all phrases do. Some phrases are headed by other phrases. And these phrases, it turns out, are identical. We turn to these cases next.

Recursive Phrases

There are substitution classes that pick out strings which are recursive on themselves. These phrases are headed by themselves and they are found inside those we've identified so far. For example, in the position marked by "___" in (84), we find the family of strings in (85). Some examples are in (86).

(84)   Det ___ V

(85)   {N, AP N, N AP, N PP, AP N PP, N PP AP, AP AP N, N PP PP, AP AP
        N PP, AP AP N PP PP, …}

(86)   the woman left.
        the happy woman left.
        the woman unhappy with the lecture left.
        the happy woman with a hat left.
        the woman with a hat unhappy with the lecture left.
                         ⋮

Coördination also reveals that this set of strings forms a phrase.

(87)   The woman and happy man left.
        The happy woman and man with a hat left.
                         ⋮

This family of strings does not appear to be the family we have called NP. There are two, related, reasons for this. First, there are grammatical strings from the second family which cannot be substituted for instances of the first family, as (88) indicates.

(88) a.   The woman left.

  b.   * Woman left.

Second, a close inspection of the set that the second family is made up of indicates that it does not share Harris's property. This family is recursive with respect to itself. Unlike NPs, which can have only one instance of a DP inside them, the phrase we've discovered here can contain any number of strings of the same kind as itself. A common convention is to represent these smaller, self-recursive, phrases with a bar, or prime. The self-recursive phrase within NPs, for instance, can be represented as $\overline{\text{N}}$, or N′. So, we set up something like (89).

$\overline{\text{N}}$ is pronounced [ɛn bɑɹ] ('N bar').

(89) a. NP → $\left\{ \begin{array}{c} \text{(NP's)} \\ \text{(DetP)} \end{array} \right\} \overline{\text{N}}$

  b. $\overline{\text{N}} \rightarrow \text{AP } \overline{\text{N}}$

  c. $\overline{\text{N}} \rightarrow \overline{\text{N}} \text{ AP}$

  d. $\overline{\text{N}} \rightarrow \overline{\text{N}} \text{ PP}$

  e. $\overline{\text{N}} \rightarrow \text{N}$

Note how these rules encode the "optionality" of AP and PP differently than the optionality of DP. And note, further, that they are all endocentric on N. They also leave out the position of CP; this is because fitting CPs into this structure poses a problem. We will return to it in just a moment.

We find the existence of very similar subphrases within VPs as well. Consider, for instance, the environment in (90), which permits the family of strings in (91), as (92) exemplifies.

(90) NP ___ CP

(91) {V, AdvP V, V AdvP, V PP, AdvP V PP, V PP AdvP, V AdvP PP, AdvP V PP PP, AdvP AdvP V, AdvP AdvP V PP, V PP PP, …}

(92) Sally said that Jerry left.
Sally quickly said that Jerry left.
Sally quickly said to Peter that Jerry left.
Sally said to Peter quickly that Jerry left.
Sally said quickly to Peter that Jerry left.
Sally carefully said to Peter on Tuesday that Jerry left.
⋮

And, as before, coördination recognizes this family.

(93) Sally **shouted** and **whispered** that Jerry left.
Sally **loudly shouted** and **whispered** that Jerry left.
Sally **shouted to Peter** and **quietly whispered** that Jerry left.
⋮

This subphrase is self-recursive and headed, just like $\overline{\text{N}}$. So we have something like (94).

(94)   a. VP → $\overline{V}$

        b. $\overline{V}$ → AdvP $\overline{V}$

        c. $\overline{V}$ → $\overline{V}$ AdvP

        d. $\overline{V}$ → $\overline{V}$ PP

        e. $\overline{V}$ → V

These rules leave out the expansions of VP which introduce NPs, CPs, APs, and VPs. Moreover, the first of these rules says that VPs are $\overline{V}$s and nothing more, which raises the obvious question why we posit $\overline{V}$s here at all. We would get the same result by dispensing with the first of these rules, and replacing $\overline{V}$ with VP throughout the remainder. We will soon see, however, that in certain situations there is a term which appears to be dominated by VP but not $\overline{V}$. I'll keep these rules in anticipation of that situation.

    A similar situation arises in Adjective Phrases. If we examine the environment in (95) we discover that it characterizes the set of strings in (96).

(95)   V ___ CP

(96)   {A, Deg A, Deg Deg A, A PP, Deg A PP, Deg A PP PP, …}

(97)   Sean is **happy** that syntax is cool.
       Sean was **happy on Tuesday** that syntax is cool.
       Sean was **very happy on Tuesday in this class** that syntax is cool.
$$\vdots$$

As before, this family is self-recursive and headed. It is visible to coördination as well.

(98)   A child **happy with her guardian** and **well-rested** is unlikely to cause trouble.
       A child **happy with her guardian** and **completely well-rested** is unlikely to cause trouble.
       A child **thoroughly unhappy in a zoo** and **angry at her guardian** is likely to cause trouble.
$$\vdots$$

We need to revise the AP rule to something like (99).

(99)   a. AP → $\overline{A}$

        b. $\overline{A}$ → DegP $\overline{A}$

        c. $\overline{A}$ → AdvP $\overline{A}$

        d. $\overline{A}$ → $\overline{A}$ PP

        e. $\overline{A}$ → A

Note that I have left out CP, as in the other rules. And, like the VP rule, these rules characterize AP as consisting of just an $\overline{A}$ and nothing else. Both matters we'll take up shortly.

There is a feature of this method of representing these subfamilies that I would like to draw attention to now. It allows for two separate parses of examples such as (100).

(100)   the considerate gift and donation

It is possible to produce this string either by grouping *considerate* and *gift* into one $\overline{\text{N}}$ and conjoining that with an $\overline{\text{N}}$ consisting of just *donation*, or it is possible to conjoin *gift* and *donation* into one $\overline{\text{N}}$ and then group that phrase with *considerate* into an $\overline{\text{N}}$. It is easy to represent these two parses by way of "phrase marker trees," which graphically elucidate the constituent structure of strings. The two ways of producing (100) are represented by the trees in (101).

(101)   a.

```
                    NP
          ┌──────────┴──────────┐
         DP                      N̄
         │              ┌────────┴────────┐
         D̄             AP                 N̄
         │              │         ┌────────┼────────┐
         D              Ā         N̄   and   N̄
         │              │         │         │
        the             A         N         N
                        │         │         │
                  considerate   gift     donation
```

b.

```
                    NP
          ┌──────────┴──────────┐
         DP                      N̄
         │              ┌────────┼────────┐
         D̄             N̄   and   N̄
         │         ┌────┴────┐        │
         D        AP         N̄        N
         │         │         │        │
        the        Ā         N     donation
                   │         │
                   A       gift
                   │
              considerate
```
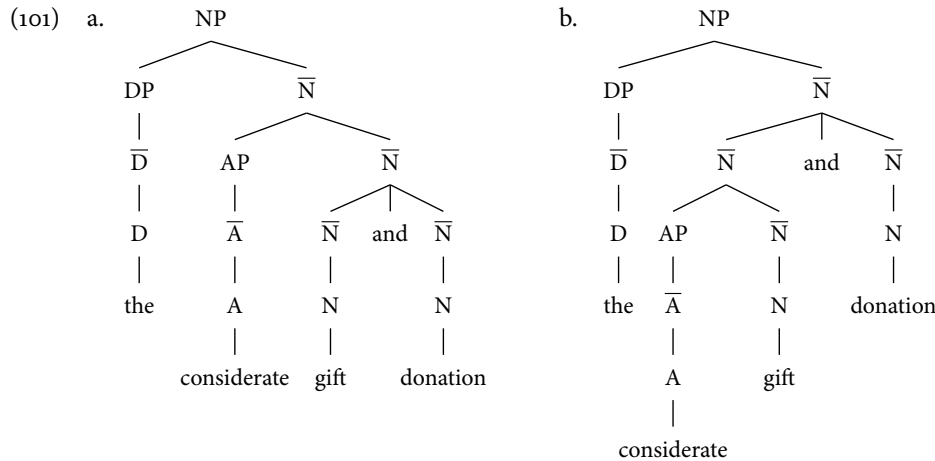
We might note that there are two meanings attached to this string as well, having to do with how the meaning of *considerate* is combined with the meanings of the rest of the parts. A loose paraphrase of these two meanings might be as given in (102).

(102)   a. the things which are considerate and which are, first, a gift and, second, a donation

b. the things which are, first, a considerate gift and, second, a donation

The difference in these meanings tracks the semantic groupings we discussed earlier. In (102a), the meaning of *considerate* is combined with the meaning of *gift and donation*. In (102b), the meaning of *considerate* instead combines with the meaning of *gift*. These semantic groupings mirror the syntactic groupings in (101).

There are a variety of reasons for thinking that syntactic phrasing always maps onto semantic groupings in the way that (101) and (102) illustrate. For one thing, the number of meanings and the number of parses matches. For instance, if we add one more adjective to the left of the coördinated nouns, as in (103), our rules allow for a total of three parses, shown in (104), and there are three meanings as well, as indicated in (105).

(103)   the considerate big gift and donation

(104)
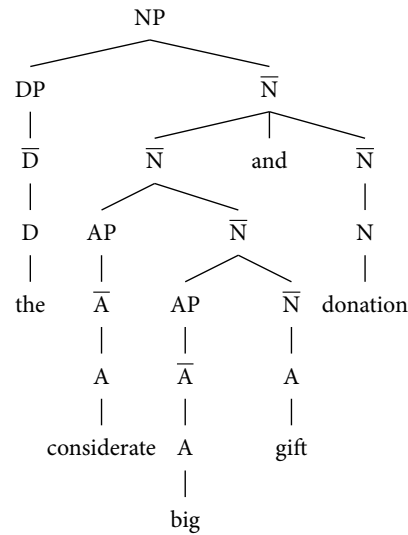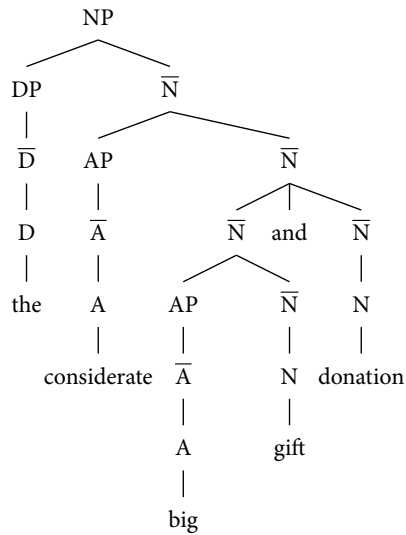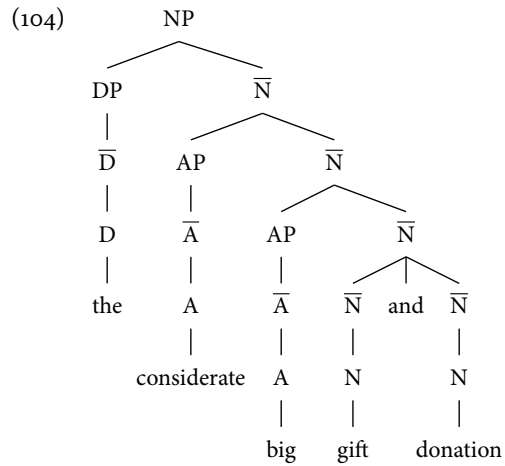
(105)   a. the things which are considerate and big and are also a gift and a
            donation.

        b. the things which are considerate and are also a big gift and a
            donation

        c. the things which are a considerate big gift and a donation

There is, then, a correspondence between the number of groupings that these
rules afford and the number of meanings that can be associated with the
relevant strings. This suggests that the syntactic phrases these rules produce
correspond to the semantic groupings.

   Furthermore, the meanings vary in a predictable way with the linear order
that these terms are arranged in. Thus, for instance, putting the second ad-
jective to the left of the coördinated nouns creates the three meanings listed

in (105), whereas putting the second adjective to the left of just the rightmost noun, as in (106), produces just two readings: they are paraphrased in (107).

(106)   the considerate gift and big donation

(107)   a. the things which are considerate and both a gift and a big donation

   b. the things which are a considerate gift and a big donation

This is predictable in the sense that our syntactic characterization of these strings would deliver just the two parses for (106) shown in (108).

(108)   a.
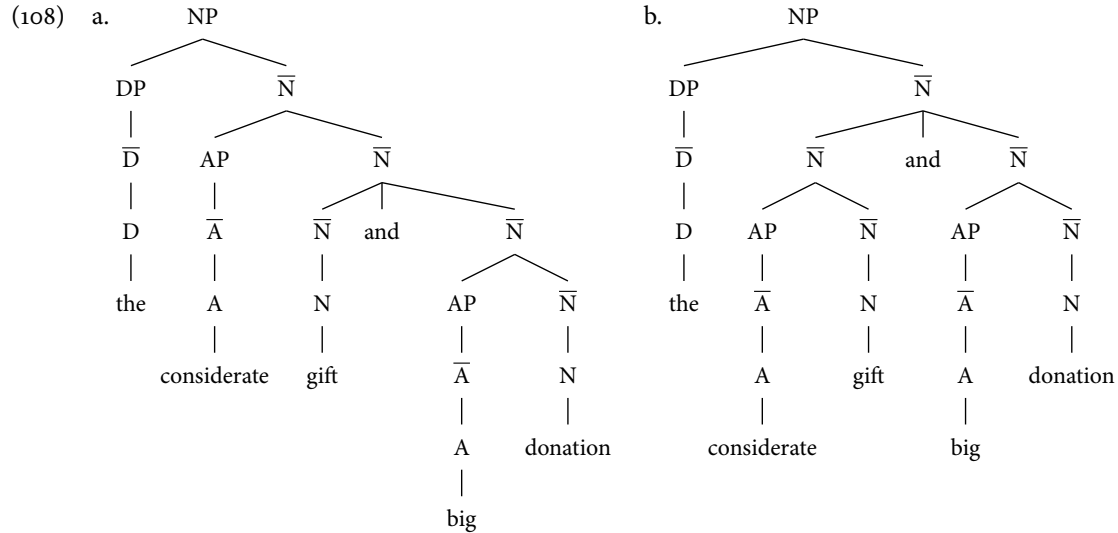
This correspondence should give us some courage that we are on the right track in characterizing the infinite strings under discussion in terms of self-recursive phrases. It provides a set of structures that are in correspondence with what look like a parallel set of meanings. Our next step should be to flesh out this correspondence, but we have some work still to do in characterizing these basic facts about grammaticality judgments. So let's return to that task.

   The strings belonging to Adverb Phrases are so simple that it is difficult to know whether they contain the substructure we've found in the other phrases. Nonetheless, they do have a recursive part and this might be construed, on analogy with the other cases, as evidence for substructure:

(109)   Sally carefully spoke.
   Sally very carefully spoke.
   Sally very, very carefully spoke.
   ⋮

The coördination phenomenon also seems to suggest subphrases, at least if our decision about the meaning-form mapping made above is correct.

(110)   Sally spoke [almost [very rapidly] and [quite softly]].

So, let's convert the AdvP rule to (111).

(111) a. AdvP → $\overline{\overline{\text{Adv}}}$

     b. $\overline{\overline{\text{Adv}}}$ → DegP $\overline{\text{Adv}}$

     c. $\overline{\overline{\text{Adv}}}$ → Adv

Like the AP and VP rules, this battery of rules equates AdvP with $\overline{\overline{\text{Adv}}}$ and so makes mysterious why they are called different things.

    The rule building sentences, IPs, is similarly meager. But it too shows some signs of the subfamilies which we have discovered in NPs, APs and VPs. This is indicated by coördination in examples such as (112).

(112) Jerry [can speak loudly] but [can't speak clearly].

And, when we add to our observations that adverbs can fall to the left of Infl, we discover the self-recursive flag of these intermediate families:

(113) Jerry evidently won't speak.

       Jerry evidently deliberately won't speak.

       Jerry evidently won't speak deliberately.

       Jerry evidently occasionally deliberately won't speak.

       Jerry evidently won't speak occasionally deliberately.

(These are all somewhat strained, I grant you, but I think still grammatical.) This calls for a change along the lines in (114).

(114) a. IP → $\left\{ \begin{array}{c} \text{NP} \\ \text{CP} \end{array} \right\} \bar{\text{I}}$

     b. $\bar{\text{I}}$ → AdvP $\bar{\text{I}}$

     c. $\bar{\text{I}}$ → $\bar{\text{I}}$ AdvP

     d. $\bar{\text{I}}$ → I VP

Note how in this battery of rules, unlike the others we've formulated, the $\overline{\text{X}}$ rule that terminates the recursion has more than just the "head" of the phrase in it. In this case it also introduces the VP. This is required because VPs are not recursively introduced, and the method we have adopted of representing recursion in these phrases is built into the structure of the substitution classes.

    Actually something similar is true for the rules that build APs, NPs and VPs as well. In the case of VPs, the NP and CP parts of their family are not recursively introduced. So we should change the terminal expansion to:

(115) $\overline{\text{V}}$ → V (NP) (CP)

And similarly, the CP parts of the AP and NP families are not recursively introduced, so the terminal expansions of these families should be changed to:

(116) $\overline{\text{A}}$ → A (CP)

      $\overline{\text{N}}$ → N (CP)

So this corrects the omission of CP and NP in our original formulation of these rules, though, as foreshadowed above, this will produce a difficulty.

But haven't we already characterized strings like the third and the fifth in (113) as adverbs introduced by a recursive $\overline{\text{V}}$ rule? Do we really need to also let sentence-final adverbs be introduced by a recursive $\bar{\text{I}}$ rule? The answer typically given is: yes. But you might want to decide for yourself what the answer to this should be.

To see this difficulty, consider how our structural method of stopping the recursion relates the terms that are within some phrase. We expect that those terms which are introduced in the terminal expansion "$\overline{X} \rightarrow X \ldots$" (that is, the non-recursively introduced terms) will form the most inclusive substitution class of the phrase involved. There are some kinds of phenomena which suggest that this expectation is fulfilled. There are processes, for example, in which a rather surprisingly short string can substitute for one or another of the families we have discovered. This happens under conditions of anaphora.

For example, the $\overline{V}$ family can be anaphorically connected to other $\overline{V}$s. One way an anaphoric $\overline{V}$ is signaled is, interestingly enough, by not pronouncing it. Such an $\overline{V}$ is said to be "elided," and sometimes the process that permits anaphoric phrases to go unpronounced is called "ellipsis." I will mark phrases that are elided with "Δ."

> "Anaphora" refers to processes in which a phrase in one position refers, in some fashion or other, to the same things that another phrase, in a different position, refers to. For instance, in the sentence: *Mary knows that she is smart*, it is possible for *she* to refer to the same individual that *Mary* refers to. In such a case, we say that *she* is <u>anaphoric</u> to, or with, *Mary*.

(117)   a.   Although Sally shouldn't Δ, Jerry must leave town.
             Δ = "leave town"

        b.   Although Sally has carelessly Δ, Jerry has carefully read *Aspects*.
             Δ = "read *Aspects*"

        c.   Because Jerry frantically read *Aspects* after dinner, Sally did Δ just before class.
             Δ = "frantically read *Aspects*"

This process of anaphora — called "VP Ellipsis," — reveals that the non-recursive parts of the VP family are trapped within the smallest subfamily.

(118)   a.   * Although Sally shouldn't Δ Chicago, Jerry must leave New York.
             Δ = "leave"

        b.   * Although Sally didn't Δ that she was tired, Jerry will say that she should sleep.
             Δ = "say"

> These examples are very carefully crafted in order to make them illustrate the properties of phrases I want to discuss here. There are other examples of ellipsis which seem to be at odds with the phrases our rules define. We'll take a longer, more careful, look at ellipsis later where I'll try to convince you that these rogue examples of ellipsis involve a complication that obscures their compliance with our phrase structure rules.

Notice, incidentally, that these instances of ellipsis provide an account for one of the problems we looked at in characterizing "sentence" during our attempt to define syntax (see page 15).

Another way an anaphoric $\overline{V}$ can be expressed is with the words *do so*. Just as with VP Ellipsis, when a $\overline{V}$ is made up of *do so*, its meaning derives from a $\overline{V}$ spoken elsewhere. This is illustrated in (119).

(119)   a.   Jerry must leave town, but Sally mustn't **do so**.
             do so = "leave town"

        b.   Jerry should eventually read *Aspects*, and Sally should immediately **do so**.
             do so = "read *Aspects*"

From the examples in (120), we learn that CPs and NPs cannot be a sister to **do so**.

(120) a.   * Jerry must leave Chicago and Sally must **do so** New York.

  b.   * Jerry must acknowledge that he will read *Aspects* and Sally must
       do so that she will read *Syntactic Structures*.

Just as with VP Ellipsis, this follows if *do so* cannot stand in place of a single
verb, and NPs and CP must be positioned in the smallest $\overline{V}$, as (115) says.

There is is no parallel form of anaphora for APs which we can employ
to confirm (116), but there is a form of anaphora in NPs that is sometimes
leveraged into trying to understand where CPs fit into them. Recall that one of
the things left out when we revised the NP rule so that it includes $\overline{N}$s is where
CPs are positioned. For a certain class of CPs, it looks like they do not belong
to the recursive part of NPs.

(121) a.   The rumor that beans can explode is widespread.

  b.   * The rumor that beans can explode that natto is a bean is
       widespread.

So this supports a decision to put CPs within the nonrecursive $\overline{N}$, as in (122).

(122) $\overline{N} \longrightarrow N\ (CP)$

For some English speakers, this decision can be verified by a form of anaphora
within NPs that uses the word *one*.

The term *one* is an anaphor that derives its meaning from an $\overline{N}$ spoken
elsewhere, as the examples in (123) demonstrate.

(123) a.   I will examine the blue book about language if you will examine
       the brown one.
       one = "book about language"

  b.   I will examine the big blue book about language if you will
       examine the small one.
       one = "blue book about language"

  c.   I will examine the long book about language if you will examine
       the one about Quarks.
       one = "long book"

Now, let's consider how *one* anaphora works in cases where an NP contains a
CP. Some linguists have found that there is a contrast in examples like (124)

(124) a.   I will examine the repeated proofs that language exists if you
       will examine the isolated ones.
       ones = "proofs that language exists"

  b.   ?* I will examine the repeated proofs that language exists if you
       will examine the ones that it doesn't.
       ones = "repeated proofs"

This contrast makes sense if *ones* must "stand in" for the noun and the CP
that follows, and cannot stand in for the noun by itself. This is explained if

There are some kinds of CPs which do look as if they belong to the recursive part. These are relative clauses, illustrated by the portion of "the book **that she read**" in bold face. We'll come back to the difference between these two kinds of clauses.

What the "?*" is meant to indicate in (124b) is that I do not reliably get a judgment of ungrammatical from my consultants. I cannot tell whether the variation is a matter of differences in individual grammars (although it appears to go that way) or due to something about my presentation of the examples.

ones can be an $\overline{\text{N}}$, but not a N, and CP must be positioned as (122). It isn't, incidentally, that there is some semantic constraint on *one(s)* that prevents it from standing in for something that has the meaning of a single noun, because that is possible in cases such as (125).

(125)   I will examine the book on the shelf if you will examine the one on the table.

The difference between (125) and (124b) is just whether the material that combines with *one* is allowed to be a sister to an $\overline{\text{N}}$ or not: PPs are (look at (123c)), and CPs aren't.

Unfortunately, however, the judgements this argument relies on seem to vary considerably across speakers, and do not seem to produce the same degree of sharpness for the speakers who do get them that are found with the $\overline{\text{V}}$ anaphora examples.

In any case, let us adopt as a preliminary conclusion that the smallest expansion of $\overline{\text{V}}$ and $\overline{\text{N}}$ should have the shapes that these considerations lead us to.

(126) $\overline{\text{N}} \rightarrow$ N (PP) (CP)
        $\overline{\text{V}} \rightarrow$ V (NP) (PP) (CP)

This way of distinguishing the recursive and non-recursive parts of phrases also predicts that the non-recursive parts will always come between the head of their phrase and the recursive parts. This seems true sometimes, as in (127) and (128).

(127)   a.   Jill ate it at noon.

        b.   * Jill ate noon it.

(128)   a.   Jill ate spätzle at noon.

        b.   * Jill ate at noon spätzle.

But for other cases it seems uncertain, or downright wrong, as in (129) and (130).

(129)   a.   Jill ate the rotting kumquats.

        b.   Jill ate at noon the rotting kumquats.

(130)   a.   ?? Jill said [that you shouldn't eat kumquats] at noon.

        b.   Jill said at noon [that you shouldn't eat kumquats].

This then, is the difficulty in trying to place CPs, and certain NPs, within VP (and other phrases too, as we'll see). Let's set this problem aside momentarily. It will be the focus of a lot of our work in the chapters that follow.

There is a similarity to the organization of the family of substitution classes that make up NP, VP, AP, AdvP, and IP. The other phrases: PP, CP, DegP and DP are too anemic for us to see that structure, so we don't know, empirically, whether or not they have it. But, following Chomsky's injunction that we

We will soon see that this contrast between PP and CP illustrated here does not turn out to be a distinction that hinges on the PP/CP difference. It will emerge that the semantic function of the PP or CP determines how it behaves with regard to this test, and not simply that the phrase is a PP or CP.

We will eventually see that PP and DP do have the same structure, but it requires more exotic constructions than we are now prepared for.

contribute to solving the learnability problem, we would do well to accept as the null hypothesis that they are in fact organized along the same guidelines. This is because doing so is a step towards shrinking the space of grammars through which the learning device has to search. Here, then, is an illustration of how explanatory adequacy can help guide the inquiry. It provides a way to choose among a range of descriptively adequate grammars a 'null hypothesis': the one to start with.

If we give all phrases the shape that NPs, VPs, etc. do, we end up with a family of substitution classes like that below.

(131) $CP \rightarrow \overline{C}$ $IP \rightarrow \left\{ \begin{array}{c} (NP) \\ (CP) \end{array} \right\} \overline{I}$

$\overline{C} \rightarrow C\ IP$ $\overline{I} \rightarrow I\ VP$

$NP \rightarrow \left\{ \begin{array}{c} (DetP) \\ (NP's) \end{array} \right\} \overline{N}$ $VP \rightarrow \overline{V}$

$\overline{N} \rightarrow AP\ \overline{N}$ $\overline{V} \rightarrow AdvP\ \overline{V}$

$\overline{N} \rightarrow \overline{N} \left\{ \begin{array}{c} PP \\ AP \\ CP \end{array} \right\}$ $\overline{V} \rightarrow \overline{V} \left\{ \begin{array}{c} PP \\ AdvP \\ CP \end{array} \right\}$

$\overline{N} \rightarrow N\ (PP)\ (CP)$ $\overline{V} \rightarrow V\ (NP)\ (PP) \left\{ \begin{array}{c} (IP) \\ (AP) \\ (CP) \end{array} \right\}$

The $\overline{V}$ rule that introduces verbs does not conform to the $\overline{X}$ Skeleton below. It allows for there to be three phrases within the $\overline{V}$. In fact, however, it appears that at most two phrases are possible in this $\overline{V}$: the full three V+XP+YP+ZP sequence allowed by this rule never actually arises (although, as Laura Aldridge points out, there is the apparent counterexample in *I'll bet you $10 that I can find a counterexample*).

$AP \rightarrow \overline{A}$ $AdvP \rightarrow \overline{Adv}$

$\overline{A} \rightarrow \overline{A}\ PP$ $\overline{Adv} \rightarrow DegP\ \overline{Adv}$

$\overline{A} \rightarrow \left\{ \begin{array}{c} DegP \\ AdvP \end{array} \right\} \overline{A}$ $\overline{Adv} \rightarrow Adv$

$\overline{A} \rightarrow A\ (PP) \left\{ \begin{array}{c} (CP) \\ (IP) \end{array} \right\}$

$PP \rightarrow \overline{P}$ $DetP \rightarrow \overline{Det}$

$\overline{P} \rightarrow P \left\{ \begin{array}{c} (NP) \\ (PP) \end{array} \right\}$ $\overline{Det} \rightarrow Det$

These all conform to the following shapes.

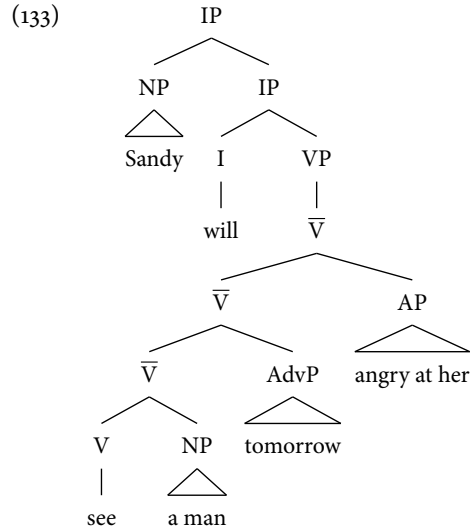(132) $\overline{X}$ Skeleton: $XP \rightarrow (ZP)\ \overline{X}$
            $\overline{X} \rightarrow QP\ \overline{X}$
            $\overline{X} \rightarrow \overline{X}\ WP$
            $\overline{X} \rightarrow X\ (YP)\ (UP)$

ZP is called the *Specifier* of XP, WP, QP are called *Adjunct*(s), and YP and UP are called the *Complements* of X.

These definitions of "Specifier," "Adjunct" and "Complement" come from Jackendoff (1977), as do most of the basic ideas sketched in this section about how phrases are organized. Jackendoff's work is a detailed working out, and considerable extension, of proposals in Chomsky (1970), where there is also a proposal about the underlying building blocks of morpho-syntactic category. Chomsky's work has its seeds in Harris (1946).

It should be said that these rules leave out considerable detail. In particular, there are a wide range of things that can stand in adjunct position which are not indicated in these rules. For example, $\overline{V}$ can have an AP adjoined to it, as in (133).

(133)

```
                        IP
              ┌─────────┴─────────┐
             NP                   IP
            ╱△╲           ┌────────┴────────┐
          Sandy          I                 VP
                         │                  │
                        will               V̄
                              ┌─────────────┴──────────┐
                             V̄                         AP
                       ┌──────┴──────┐              ╱──△──╲
                      V̄             AdvP          angry at her
                 ┌────┴────┐       ╱──△──╲
                 V         NP     tomorrow
                 │        ╱△╲
                see      a man
```

And, as noted earlier, an $\overline{N}$ can have certain kinds of CPs adjoined to them. An example of this is (134).

(134)

```
                      NP
            ┌──────────┴──────────┐
           DP                     N̄
          ╱△╲          ┌──────────┴──────────┐
         the          N̄                     CP
                ┌──────┴──────┐          ╱───△───╲
               N̄             CP      which no one will
               │          ╱──△──╲       admit writing
               N        that Mary read
               │
              book
```
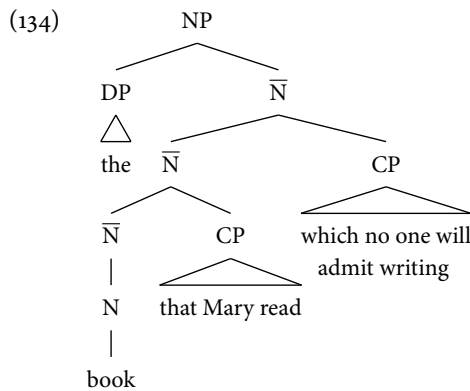
If you have a native English speaker handy, it won't take much time to discover many other kinds of combinations that have been left out. I will continue to leave out this detail, invoking it where necessary as we go along.

Arguments and Modifiers

Charting substitution classes in the way we've been doing characterizes grammatical arrangements of words solely in terms of the words' categorial status. It throws strings of words together according to their status as nouns, verbs, adjectives and the like and ignores all other properties of the particular verbs, nouns, adjectives, etc. chosen. It forms some pretty odd sentences:

(135)   a.   Jerry danced with pickles.

      b.   Jerry danced at noon at midnight.

      c.   Jerry slowly stood still.

      d.   a green idea

These are odd because the meanings they deliver are so bizarre. But they are still recognizable as grammatical strings of words.

But some combinations which these rules allow seem to go bad in a very different way; consider (136).

(136)   a.   Jerry laughed Mary.

      b.   Sam gave it at Jill.

      c.   Sally died that you should eat better.

      d.   Jim claimed to Kris.

      e.   Jerry slapped.

These don't go together into weird meanings; they don't go together at all. These examples, arguably, should be marked as syntactically ill-formed. They are ungrammatical, and so our syntax should be revised.

What's wrong, here, is that we've matched up verbs with the material that follows them incorrectly. As can be seen by comparing (136) with (137), if the verbs are followed by different material, the results are grammatical.

(137)   a.   Jerry laughed.

      b.   Sam gave it to Kris.

      c.   Sally died.

      d.   Jim claimed that you should eat better.

      e.   Jerry slapped his thigh.

Here, then, is something more particularly about the words themselves that seems to be relevant to the procedure that recognizes grammatical strings. To capture what goes on here, we must do more than know what the category of the words being combined is.

There's another respect in which the particular choices of words seems to play a role in the syntax. Consider the different semantic contributions the NP *Tuesday* makes in (138).

(138)   a.   I danced Tuesday.

      b.   I remember Tuesday.

In the first case, *Tuesday* indicates when *I danced* happened. We say in this case that *Tuesday* is a modifier. It modifies the sentence's meaning by restricting the events denoted by *I danced* to just those that transpire on Tuesday. But this is not the role it has in the second case. Here *Tuesday* refers to the thing remembered. We say in this case that it is an argument of the relation that *remember* denotes.

A similar contrast can be seen in the pair in (139).

(139)   a.   I kissed her on the bus.

   b.   I put her on the bus.

Again, *on the bus* is a modifier in the first case. It locates the event described by *I kissed her*; it indicates that this event took place on board the bus. In the second case, by contrast, it names a locations related to the referents of *I* and *her* by *put*. It is an argument.

The semantic role an argument has in a sentence is determined by the word for which it is an argument. The meaning that modifiers contribute to the sentence they're part of is much less dependent on the meanings of the words and phrases nearby.

There's a way of talking about argumenthood that is commonplace, and which we inherit from Gruber (1965). Gruber was concerned with the problem of verb meanings, and in particular with finding a theory that restricted the kinds of argument types that verbs permit. He speculated that there was a finite, in fact quite small, set of argument types, or 'roles', that could be put together by verbal meanings. He argued that the roles which verbs related were always ones that have to do with the metaphysics of motion. For example, a verb like *send* involves three terms, one that can be seen as indicating the source of the motion, another that denotes the moved term and a third that names the goal, or endpoint, of that motion. Gruber called the role borne by the term undergoing motion 'Theme.'

Fillmore (1968) makes a very similar proposal to Gruber's.

(140)   Sandy sent his book to Sean.
   Source   Theme   Goal

On Gruber's thesis, then, verbs are just words that name different kinds of motion. The arguments of verbs will therefore always be the objects, locations, and manners of movement that are involved in that motion. Arguments will have roles in the motion relation named by a verb, and what those roles are will be determined by the verb. Those roles Gruber called "theta roles," after the Theme role. You will often see "theta role" expressed as "θ-role." We say, then, that verbs assign θ-roles to their arguments.

There are cases where it is not obvious that motion underlies the relation named by a verb. Nothing has to move in the events described by (141), for example.

(141)   a. Sandy showed his book to Sean.

   b. Sandy pleases Sean.

But in these cases, one might imagine that there is a kind of abstract motion involved. The image of *his book* might be transferred from Sandy to Sean in (141a), for example; and pleasure might be thought to have moved from Sandy to Sean in (141b).

But there are cases which even metaphorical extensions of the logic of motion look unlikely to characterize. (142) are some.

(142)   Sandy finds Sean unpleasant.

Sandy is unhappy.

'Londres' refers to London in Portuguese.

This means that something is wrong.

Sean became unhappy.

They know something.

This way of constraining the meanings of verbs has been pretty firmly aban-
doned, I believe, as a consequence. In its place, a method has been pursued
that tries to see the meanings verbs have to be the result of combining a small
number of elemental predicates, like CAUSE, MOVE, and BECOME. On this
conception, the roles that arguments play in a sentence can be viewed as re-
sulting from being the arguments of these more basic predicates. We will
continue to talk about verbs and the θ-roles they assign.

Dowty (1979, 1989) are good points of entry into this literature.

The relation between verbs and their arguments that is expressed by
θ-roles can be seen as a special instance of a more general relationship which
goes under the name "selection," or sometimes "s-selection" (for "semantic"
selection). This refers to the connection between a verb's (or other similar
term's) meaning and the semantic values that its arguments deliver. θ-roles
express a similar function: they name the meaning that an argument's seman-
tic value must be compatible with. But the relation holds for other cases too,
where the language of θ-roles does not so easily extend. One of those places is
where verbs connect with clauses of various types. So, a verb's meaning deter-
mines somehow whether the clause it combines with must have the meaning
of an interrogative or a declarative, for example.

There is a subtle difference between what θ-roles are and what s-selection is. When a predicate s-selects an argument, it specifies what kind of meaning its argument must have. When a predicate assigns a θ-role to an argument, it determines how the meaning of its argument will be composed into the meaning of the sentence. In practice, these different notions cover nearly the same phenomena, since how a term's meaning is composed into a sentence is closely tied to what meanings that term may have.

(143)   a.   Martha denied that John has left.

b.   Martha said that John has left.

c.   * Martha wonders that John has left.

(144)   a.   * Martha denied whether John has left.

b.   Martha said whether John has left.

c.   Martha wonders whether John has left.

We say of these cases that verbs select or s-select a question or declarative.
Note that some verbs are compatible with either, as is *say*.

Although it is hard to see these differences as fitting the functions that
θ-roles typically name, I will use the language of θ-roles to describe these
relations too.

Now that we have in view this distinction between arguments and non-
arguments, let's return to the contrast between (136) and (137), repeated below.

(136)   a.   * Jerry laughed Mary.

b.   * Sam gave it at Jill.

c.   * Sally died that you should eat better.

d.   * Jim claimed to Kris.

e.   * Jerry slapped.

(137) a.   Jerry laughed.

       b.   Sam gave it to Kris.

       c.   Sally died.

       d.   Jim claimed that you should eat better.

       e.   Jerry slapped his thigh.

In all these examples, I've chosen phrases that cannot get an interpretation as a modifier. The contrast between (136) and (137), then, hinges on giving the verbs involved the right argument. The post-verbal phrases differ with respect to their category. What we discover from these examples is that verbs are picky with respect to what category their arguments belong to. This can also be seen in examples, such as those in (145) and (146), where the difference in the category of an argument is not tangled up so much with the meaning of that argument.

(145) a.   Jerry pleases Mary.

       b.   * Jerry pleases to Mary.

(146) a.   * Jerry talks Mary.

       b.   Jerry talks to Mary.

We say that verbs are *subcategorized* by the category of their argument. Or — this term has been relexicalized — that verbs subcategorize their arguments. Sometimes this is also described as a verb *c-selecting* its argument.

The terms *s-select* and *c-select* come from Pesetsky (1982).

   Jackendoff (1977) argues that arguments (when they follow the head they are an argument of) are necessarily in complement position. This is supported by contrasts like those in (147) and (148).

(147) a.   Although Sally didn't Δ Tuesday, she will dance Monday.

       b.   * Although Sally didn't Δ Tuesday, she will remember Monday.

These sentences should be pronounced with contrastive focus on the terms in small caps.

(148) a.   Although Kylia won't Δ on the bus, she will hug me in the car.

       b.   * Although Kylia won't Δ on the bus, she will put me in the car.

Because the phrase following the verb is an argument in (147b) and (148b), it must be within the $\overline{V}$ which elides, whereas in (147a) and (148c), the phrase following the verb is a modifier and can therefore remain outside the ellipsis. This thesis is also supported by similar contrasts involving *do so* anaphora.

(149) a.   ?* Sam talked to Mary and Sally *did so* to George.

       b.   * Gerry eats chocolate, and Sandy *does so* marzipan.

       c.   * Mag proved that she loved chocolate, and Holly *did so* that she loved marzipan.

The anaphora tests we used previously to discover what is in the smallest $\overline{V}$ are indicating that this position is reserved for arguments.

   To the extent that we can tell, the same is true with respect to arguments of other classes of words. It's difficult to be certain of the argument-status of

terms which combine with nouns, for instance, but Jackendoff's claim seems correct here as well.

(150)   a.   I'll listen to your long, careful discussion of it, if you'll listen to my short *one*.
one = "careful discussion of it"

   b.   * I'll listen to your long, careful discussion of it, if you'll listen to my short *one* of it.
one = "careful discussion"

   c.   I'll listen to your long, careful discussion in class, if you'll listen to my short *one* in the office.
one = "careful discussion"

The contrast between (150a) and (150b) is expected if *of it* must be positioned within the smallest $\overline{\text{N}}$, and *one* stands in place of an $\overline{\text{N}}$. The contrast between (150b) and (150c) corresponds to the differing argument-status of the PPs involved: *of it* is more strongly perceived as an argument of *discussion* than is *in class*. As with the *do so* and $\overline{\text{V}}$ Ellipsis facts, then, this contrast supports the hypothesis that arguments and modifiers are fit into phrases in different positions.

Okay, to summarize: we're looking for a way to factor into our procedure for recognizing grammatical sentences enough of the meanings of the words involved to guarantee that Verbs and Nouns (and perhaps other words) combine with the arguments they select and subcategorize. Moreover, when these arguments follow them, we must find a way of guaranteeing that they are in the non-recursive $\overline{\text{X}}$: the complement position.

We can ensure that these arguments are in the non-recursive part of the $\overline{\text{X}}$ if we force them to bear a θ-role, and allow θ-roles to be assigned only to complement positions. We need also to describe the fact that when a verb has a θ-role, there must be an argument present in the syntax which bears that θ-role. It is customary to divide this task into two parts, which can be expressed as follows:

(151)   The Theta Criterion

   a.   For every θ-role associated with a word, there is a position to which that θ-role is assigned.

   b.   For every θ-position, there is something with an appropriate semantic value that occupies that position (i.e., the argument).

It is usual to strengthen the Theta Criterion to a bijection between arguments and θ-roles, because of cases like (152).

(152)   Sally showed John *doesn't mean* Sally showed John himself.

Without constraining the Theta Criterion to a bijection, we might expect (152) to get the disallowed interpretation since presumably the NP *John* could name the object which bears both the Theme and Goal θ-roles. So I'll change (151) to (153).

There is evidence that nouns assign θ-roles only when they have a verb-like use; that is, when they are used to describe processes or events, and not when they are used to name things (see Grimshaw, 1990). The examples in (150) are constructed with this in mind. We return to this aspect of nouns in a later chapter.

Baker (1978) is, perhaps, the first to argue from *one* anaphora for this conclusion about where arguments are positioned within NPs.

The Theta Criterion comes from Chomsky's important *Lectures on Government and Binding*.

(153)   THE THETA CRITERION

    a. For every θ-role there is exactly one position to which that θ-role is assigned.

    b. For every θ-position, there is exactly one thing with an appropriate semantic value (i.e., an argument) that occupies that position.

We need to worry about cases like the following, of course, in which there appears to be an optional argument.

(154)   a.   Martha ate (pie).

    b.   It seems (to me) that Marty left.

Here we might imagine either that there actually is an object in these cases that bears the θ-role, but that argument is unpronounced. Alternatively, we might conjecture that something relaxes the condition which forces every θ-role to be assigned to a position holding an argument. The common wisdom is that both possibilities exist — we will return to this issue in some detail later. For now, let us imagine that there is a lexically determined process which allows θ-roles for certain predicates to not be assigned.

The Theta Criterion glues the positions that arguments can be in to the positions that θ-roles are assign to. Our next step should be to specify where those positions are. But we might also linger for a moment and ask whether we should write the Theta Criterion in terms of c-selection instead of θ-roles, or "s-selection." Might we say, for instance, that if a word c-selects a certain kind of phrase, there must be a single position that is occupied by a phrase of that type. There is an intimate connection between the kind of phrase c-selected, and the θ-role it is assigned, so it is not easy to tell which of these is relevant.

In fact, so intimate a connection is there between s-selection and c-selection, that some have suggested that they should be collapsed. It is sometimes suggested that the categorial type some argument has can be determined from its θ-role. Grimshaw (1979) provides a way of viewing this hypothesis which has gained some popularity. Her idea is that one of the functions that makes up the learning device assigns a categorial status to arguments on the basis of their θ-role. She calls this function "Canonical Structural Realization" (CSR). She sketches how this function might work in the special case of CPs and NPs.

So let's look as some of the facts she considers. Note first that CPs may distinguish themselves as according to whether they have meanings that we might describe as "Propositions," "Exclamatives" and "Questions." NPs can denote things of these types too, sometimes. Some examples are in (155).

(155)   a.   John asked me $\begin{cases} \text{what the time is} \\ \text{the time} \end{cases}$ (Question)

    b.   I'll assume $\begin{cases} \text{that he's intelligent} \\ \text{his intelligence} \end{cases}$ (Proposition)

    c.   Bill couldn't believe $\begin{cases} \text{how hot it is} \\ \text{the heat} \end{cases}$ (Exclamative)

The Theta Criterion is also often formulated in terms of a bijective relation between θ-roles, or θ-positions, and arguments. That is, it is sometimes written to say: "For every θ-role (or θ-position) there is exactly one argument and for every argument there is exactly one θ-role (or θ-position)." (In *Lectures on Government and Binding* it is formulated in various ways, including these two.) The difference between this alternative formulation and the one I've given here is that mine does not force every argument to receive a θ-role, whereas the alternative does. I've decided to place this requirement in another principle, which we'll come to shortly.

In these cases, then, the verbs s-select either Q(uestion), P(roposition) or E(xclamative) and c-select either an NP or CP.

There are other verbs, however, which s-select these very same meanings, but c-select only CPs. Some examples are in (156).

$$
\begin{array}{lll}
(156) \quad \text{a.} & \text{John wondered} \left\{ \begin{array}{l} \text{what the time was} \\ \text{*the time} \end{array} \right. & \text{(Question)} \\[2em]
\text{b.} & \text{I'll pretend} \left\{ \begin{array}{l} \text{that he's intelligent} \\ \text{*his intelligence} \end{array} \right. & \text{(Proposition)} \\[2em]
\text{c.} & \text{Bill complained} \left\{ \begin{array}{l} \text{how hot it was} \\ \text{*the heat} \end{array} \right. & \text{(Exclamative)}
\end{array}
$$

Here then, we have a special instance of the difference in s-selection and c-selection that needs to be overcome if one is to be derived from the other.

Grimshaw's suggestion is that the CSR of Questions, Propositions and Exclamatives is CP and that those verbs which s-select NPs with these meanings are learned on a case-by-case basis. Thus, this is a partial collapse of c-selection to s-selection. It predicts that every verb which s-selects a Question, Proposition or Exclamative will c-select a CP; that is, there should be no verbs that c-select only NPs with this meaning. This seems to be correct.

Whether or not this project can be maintained for the situation involving the relation between CPs and NP and the meanings they have, I don't think a parallel story holds for the complements of other categorial type. Moreover, the scheme Grimshaw proposes won't help determine which verbs select nonfinite as opposed to finite clauses, which also seems to be a rather language particular fact. The problem we face here is the same problem we face when we try to understand whether morphosyntactic word-class can be expressed in solely semantic terms. It remains unknown what the relationship between word-class and semantic type is, and consequently, we are not in a position to yet understand what the connection, if any, is between c-selection and s-selection. From now on let us assume that c-selection is at least in part independent of s-selection, and determined on a verb-by-verb basis.

Interestingly, however, it looks like the thesis that c-selection can be derived from s-selection fares better when external arguments are concerned. To begin with, the range of categories that serve as external arguments looks somewhat less varied; to a large extent, only NPs and CPs seem to be clausal subjects in English.[8] And second, when a θ-role is consistent with either NP and CP, any kind of CP is possible as is an NP:

$$
(157) \quad \left\{ \begin{array}{l} \text{That John left} \\ \text{To have to leave} \\ \text{Leaving} \\ \text{The fact} \end{array} \right\} \quad \begin{array}{l} \text{bothers Mary.} \\ \text{makes Mary happy.} \end{array}
$$

By contrast, when the subject θ-role is incompatible with the meanings that CPs yield they are banned from Specifier position:

[8] With the exception of cases like "Under the bed is a slipper," plausibly instances of impersonal constructions with inversion; see Stowell (1981) and Rochemont and Culicover (1990).

(158) $\left\{\begin{array}{l} *\text{That John left} \\ *\text{To have to leave} \\ *\text{Leaving} \\ \text{John} \end{array}\right\}$ $\left.\begin{array}{l} \text{kisses Mary.} \\ \text{likes Mary.} \end{array}\right.$

Subject arguments do not seem to be c-selected by the verbs involved. The θ-role they bear is sufficient to determine what category they can be. Only complements are c-selected.

This is a reason, then, for writing the Theta Criterion in terms of θ-roles, and not c-selection. We want to link both subject and object arguments to positions relative to the word they are arguments of, but only the object argument will be c-selected.

We can summarize what we've discovered so far with (159).

(159)   a. If a verb has a θ-role, then there is exactly one syntactic position to which that θ-role is assigned.

b. A θ-marked position must be occupied by something with the appropriate semantic value.

c. A verb c-selects only its object arguments.

The statements in (159a) and (159b) are the Theta Criterion, whereas those in (159c) and (159d) concern the relation between c-selection and s-selection which we've just reviewed. The Theta Criterion insists that for every θ-role that some term has, there will be a unique position occupied by an argument bearing that θ-role. (159c) determines whether that argument will be c-selected or not.

To force arguments to be within the smallest $\overline{X}$, it will now be sufficient to force the θ-position for that argument to be within the smallest $\overline{X}$. We want this effect for complement arguments only — we don't want to force "subject" arguments into $\overline{X}$ — so one way of doing this would be to restrict those positions that are c-selected to just those within the smallest $\overline{X}$. This would mean that we'd have two principles: one that determines the c-selected position for verbs, and another, yet to be determined, which locates the s-selected position for subjects. We're going to see, however, that the procedure for locating the θ-positions for both subject and object arguments is the same, or very nearly so, and so we won't take precisely this course.

Instead, we will follow a popular view of these principles that is first found in Chomsky's *Lectures on Government and Binding*. He formulates there what he calls "The Projection Principle," which is responsible for mapping the argument structure of a verb — or a argument taking word more generally — into a syntactic representation. I will formulate his principle as (160).

Chomsky's own formulation builds in various other properties that we will encounter later on; compare (160) to Chomsky (1981, pp. 34–48).

(160)   The Projection Principle

If α assigns a θ-role to β, then α and β must be sisters.

X and Y are *sisters* if every phrase including one includes the other.

This forces arguments of a verb to be in the lowest $\overline{V}$, for only in that position will it be a sister to the verb. Note that this principle is not restricted to verbs

and their projections, it spreads what we've discovered about VPs to all other categories. This, so far as I know, is correct.

As presently formulated, the Projection Principle wrongly forces subjects into the smallest $\overline{V}$ of the verb that assigns it a θ-role. We will see, however, that this problem is only apparent. Once we discover what is truly responsible for assigning the subject its θ-role, this problem is resolved (or, depending on how certain particulars play out, mutated into a different problem).

The first part of the Projection Principle is just (159c). It has the interesting consequence of preventing non-arguments from standing in the smallest $\overline{X}$. Thus, the Projection Principle not only has the effect of forcing arguments into the smallest $\overline{X}$, but also of forcing non-arguments out of this position. Whether this stronger result is correct is rather difficult to determine.

Deriving Phrase Structure Rules

An interesting consequence of the Projection Principle is that it factors into the lexical specification of the verbs everything needed to know what sort of phrases will be found in the non-recursive part of $\overline{X}$s. Tim Stowell, in his dissertation,[9] argues that the information about what resides in the lowest $\overline{V}$ should only be found as part of the verb's lexical content. This sort of information doesn't properly reside in the phrase structure rules, since it is information that is tied to the particular choice of word and not the pure form that sentences may take. In fact, the information phrase structure rules give about the contents of the smallest $\overline{X}$ can now be seen as merely a summation of what is possible across particular lexical items filling the head slot. Thus, we should factor out of the phrase structure rules information which concerns the categorial nature of the complements involved. We can do the same for the subject arguments as well, since their categorial nature is derived from their θ-role.

What we have seen, then, is that the phrase structure rules can be stripped of a great deal of their information. Indeed, what is left is largely what the $\overline{X}$ skeleton expresses and the categorial specification of non-arguments. We could complete factoring out the phrase structure rules into general statements, then, by simply listing the class of phrase-types that can adjoin to various phrases. We could add to the $\overline{X}$ Skeleton a language particular set of statements of the form in (161).

(161)   a.   If α is adjoined to $\overline{N}$, then α must be …

b.   If α is adjoined to $\overline{V}$, then α must be …

$$\vdots$$

The "…" will carry lists of category types. We might imagine doing the same thing with respect to Specifiers. We could envision the grammar of a language having a set of statements like (162).

(162)   a.   If α is in the Specifier of IP, then α is an NP or CP.

b.   If α is in the Specifier of NP, then α is a DP or NP's.

[9] Stowell, Timothy. 1981. Origins of phrase structure. Doctoral Dissertation, Massachusetts Institute of Technology
A very similar conclusion is reached in Ann Farmer's dissertation.

X is <u>adjoined to</u> Y, if X is a sister to Y, and X's (immediate) mother of X and Y is another thing that is the same type as Y.

If this project is successful, then the Phrase Structure rules of English collapse in full to the $\overline{\text{X}}$ Skeleton, the c-selection requirements of heads and statements like (161) about what kind of phrase can modify another. Some have argued for a picture of language variation that makes the hierarchical arrangements of constituents that the $\overline{\text{X}}$ Skeleton, together with the PROJECTION PRINCIPLE and THETA CRITERION and whatever yields (161), completely immutable.[10] All that varies across languages is the linear order in which the terms that follow the arrows in the $\overline{\text{X}}$ Skeleton may have. So, the phrase structure component of the grammar might have nothing more than (163) in it, where "$\{\alpha, \beta\}$" should be understood as representing both the string $\alpha+\beta$ and the string $\beta+\alpha$.

(163)   a. $\text{XP} \to \{(\alpha\text{P}), \overline{\text{X}}\}$
         b. $\overline{\text{X}} \to \{\overline{\text{X}}, \beta\text{P}\}$
         c. $\overline{\text{X}} \to \{\text{X}^0, (\gamma\text{P}), (\psi\text{P})\}$

What the morpho-syntactic category of α, β and γ are is fully determined by the c-selection properties of X^0 and the language particular principles governing modifier types (i.e., (161)).

Note, incidentally, that I've somewhat arbitrarily set the highest number of things that can share the head with the smallest $\overline{\text{X}}$ to two. That is, I've decided that there can be at most two complements. That seems to be supported for verbs and adjectives, when can have at the same time a PP and CP complement, for instance, but no combination of complements that goes beyond two. In later chapters we will revisit this issue.

The linear arrangements of these constituents must then be determined by the language particular part of the grammar. There is evidence from language typology that whatever it is that determines the order of Specifier and $\overline{\text{X}}$ is independent of what determines the order of heads and their complements. There is no widely agreed upon account of what is responsible for this factor, so let's leave this for the future. There is also typological evidence[11] that the order or complement and head correlates with the order of modifier and $\overline{\text{X}}$. One way of expressing this is with a "headedness parameter" which specifies whether the head of a phrase may come initially or finally in its (immediate) projection. This predicts that complements will either all precede or follow their heads, and not come among them. While this is not superficially true (German/Dutch, for example, look like counterexamples), it does look like this could be true of the underlying arrangements of these constituents and it does seem to capture a tendency that languages have.

As a starting point, then, let's take the view that languages linearize their phrasal constituents by way of setting separately the linear order of the immediate constituents of XP and $\overline{\text{X}}$. This can be achieved by letting languages pick the values "first" and "last" for the terms in (164).

[10] See Travis (1984), for example.

The expression "X^0" refers to the position in a phrase-marker that only a lexical item can be fit into. It is the position of the "head" of the phrase.

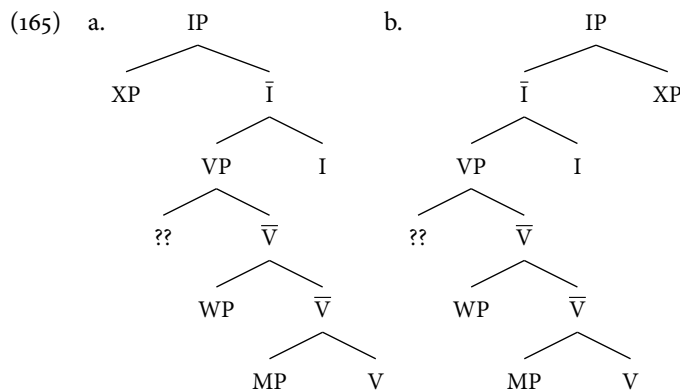[11] See Greenberg (1963) and Dryer (1992).

(164)   If σ is an immediate daughter of π, π a projection of α, then choose a value in "$<\ >$" for:

   a. σ = Specifier of π: <first in π, last in π>

   b. σ = the term π is a projection of: <first in π, last in π>, modulo (164a)

> β is a "projection of α" iff β dominates α and they have the same head.

This connects the linear order of head and complement to the linear order of head and adjunct, which Dryer's work suggests might be correct.[12] So, for instance, a language that sets Specifier to "first" and Projection-of-α to "last" will license sentence structures like those in (165a), whereas a language that sets both terms to "last" will produce structures like those in (165b).

> [12] See also Saito (1985) and Saito and Fukui (1998).

(165) a.

```
            IP                              IP
          /    \                          /    \
        XP      Ī                        Ī      XP
              /   \                     /   \
            VP     I                  VP     I
           /  \                      /  \
         ??    V̄                   ??    V̄
              /  \                      /  \
            WP    V̄                   WP    V̄
                 /  \                      /  \
               MP    V                   MP    V
```
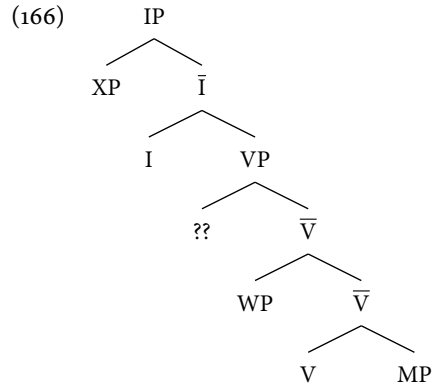
b.

The categorial values for MP will be determined by the c-selection specification of the verb involved. The categorial values for XP will be determined by the θ-role it receives. And the categorial values for WP will be whatever (163) for the language in question allows to modify $\overline{V}$s. We haven't yet discovered what sits in the Specifier of VP, so this spot is marked with "??."

The linearization parameters in (164) produce these phrase markers in the following way. Setting Specifier to "first" in (165a) linearizes XP and ?? so that they precede $\overline{I}$ and $\overline{V}$ respectively. Setting Projection-of-α to "last" makes every other $\overline{I}$ and $\overline{V}$, as well as I and V, follow the phrase they are sisters to. As a consequence WP, MP and VP precede the phrase they are complements to or modifiers of. In (165b), by contrast, Specifier is set to "last," which linearizes XP and ?? so that they follow $\overline{I}$ and $\overline{V}$ respectively. As with (165), Projection-of-α is set to "last" in (165b) and the consequence for the position of WP, MP and VP is the same.

Restricting the linearization options to just those in (164) blocks certain phrase markers. It blocks languages, for instance, in which the complement to a verb falls on a different side of that verb than does a complement to a noun (or any other category). That is, it forces languages to unify the linearization of Specifier, Complement and modifier across phrase types. It is not hard to find languages that seem to violate this restriction, but as Greenberg and Dryer find, there is a tendency for languages to avoid this type. Similarly, (164) prevents all languages that put modifiers to one side of the $\overline{X}$ they modify but
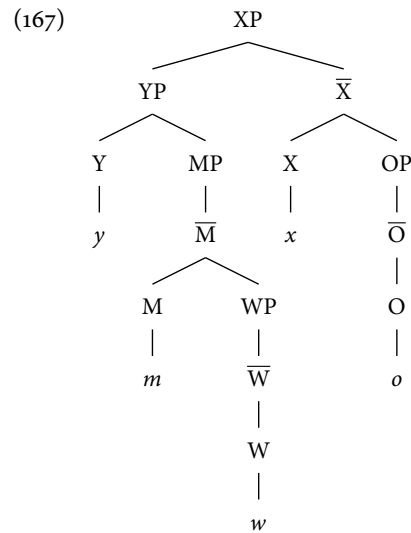
put complements to the other side. For instance, phrase markers like (166) are prevented.

(166)

```
              IP
            /    \
          XP      Ī
                /   \
               I     VP
                    /   \
                  ??     V̄
                        /  \
                     WP     V̄
                           /  \
                          V    MP
```

This phrase marker linearizes $\overline{V}$ (a projection of V) "last" relative to WP, but linearizes V (also a projection of V) "first" relative to its complement. Clearly there are languages of this unexpected type. English, for instance, seems to look precisely like (166).

This proposal, then, seems clearly too restrictive. Nonetheless, it will be our starting point. In the chapters that follow we will explore ways of loosening this model so that it is enabled to account for the range of language types we do see without losing the trends in linear organization that Greenberg and Dryer have discovered. What we have now is not yet complete enough to really engage this problem.

In fact, the linearization scheme in (164) is itself not yet complete enough to generate the strings we want to associate with the phrase markers it allows, for example those in (165a) and (165b). All (164) does is linearize the phrases within a sentence. It does not determine how the strings of words within those phrases are linearized relative to the other phrases. To see this, consider a phrase marker like that in (167), in which lower-cased letters should be understood as representing words.

(167)

```
                          XP
                     /          \
                   YP            X̄
                 /    \        /    \
               Y      MP      X      OP
               |       |      |       |
               y       M̄      x       Ō
                      /  \            |
                    M     WP          O
                    |      |          |
                    m      W̄          o
                           |
                           W
                           |
                           w
```

This phrase marker arises by imposing the $\overline{\text{X}}$ Skeleton and setting Specifier to "first" and Projection-of-α also to "first." What we would like is for this to be sufficient to generate the string *ymwxo*. Instead, however, all that these settings give us is the information in (168).

(168) a. Y precedes MP

b. M precedes WP

c. YP precedes $\overline{\text{X}}$

d. X precedes OP

What's required is something to determine how the information in (168) determines the linear order of *y* relative to the words within MP and $\overline{\text{X}}$, the linear order of *m* relative to the words in WP, and the linear order of *x* relative to the words within OP. Let's turn to that now.

Recall that in defining morpho-syntactic category, we entertained the hypothesis that looking at only adjacent terms would be sufficient for defining the relevant substitution classes. As it happens, in defining phrases we have also obeyed this constraint. As a result, phrases are always strings of adjacent terms. Let's elevate this too to an hypothesis:

(169) Contiguity

Let $\tau = \{ \alpha_1, \alpha_2, \ldots, \alpha_n \}$ be terminals dominated by ζ. The string formed from τ cannot contain β if β is not dominated by ζ.

If Contiguity holds, it is possible to determine from (168) what the consequent linearization for all the words in (167) is. If the words in MP must be adjacent to each other, then (168a) is enough to know that *y* precedes all those words (i.e. *m* and *w*). Similarly, if the all the words in YP must form a contiguous string, and all the words in $\overline{\text{X}}$ must too, then from (168c) it is possible to deduce that every word in YP (= *y*, *m* and *w*) must precede every word in $\overline{\text{X}}$ (= *x* and *o*). All that is required is an explicit statement that the words within a phrase, α, are linearized with respect to the words in phrase β in the same way that α is linearized to β. This can be done with (170).

(170) $\alpha < \beta =_{\text{def.}} \alpha$ precedes β.

$\{_\delta \alpha, \beta\} =_{\text{def.}} \quad \delta \quad \text{or} \quad \delta \;.$

$$\begin{array}{cc} \wedge & \wedge \\ \alpha \; \beta & \beta \; \alpha \end{array}$$

a. For all words, *x* and *y*, within a phrase marker, either $x < y$ or $y < x$.

b. Let X and Y be points on a phrase marker. If X < Y, then $x < y$ for all *x* dominated by X, and all *y* dominated by Y.

(170a) merely makes explicit that all the words in a phrase marker must have a linear relation to every other word in a phrase marker. (170b) determines how these linear relations are derived from the language particular orderings imposed upon phrases. It also derives Contiguity.

The relationship between phrase markers and the strings they map onto is investigated in a vast literature, where a wide array of different formalizations can be found. Two widely read classics are Chomsky (1985) and Lasnik and Kupin (1977). For a simple introduction to the formalisms of graph theory (from which phrase markers are drawn) and their projection onto strings, see Partee et al. (1990).

On this view, then, the phrase structures of languages are the result of four fixed universals — the $\overline{X}$ Skeleton, the Theta Criterion, the Projection Principle, and the linearization principles in (170) — plus the language particular pieces of information in (171).

(171) a. Specifying the categories of adjuncts (i.e., (161)).

   b. Setting the "headedness parameter" (i.e., (164)).

   c. A vocabulary of lexical items s-select and c-select arguments.

There's a sense, then, in which languages do not actually have Phrase Structure rules. They are merely the epiphenomena that emerge when the various factors of Universal Grammar and language particular information are combined. This theory, if correct, meets the criterion of explanatory adequacy. It provides both inviolable constraints (i.e., $\overline{X}$ Theory, the Theta Criterion, the Projection Principle and (170)) and an evaluation metric (i.e., (171) and the language particular vagaries of vocabulary).

Notice how the evaluation metric this proposal embraces is quite different from the "simplicity" metric suggested in Chomsky's early work. The evaluation metric here involves learning the word-by-word selection requirements and fixing parameter values in the headedness linearization procedure in (171). This proposal has the following form: inviolable constraints come in the form of immutable principles, while the evaluation metric (once lexical idiosyncrasies are removed) consists of principles with a menu of parameters that are set on a language particular basis. Theories that have this general form are said to belong to the "Principles and Parameters" framework. This conception of what explanatory grammars might look like was suggested by Noam Chomsky and his collaborators in the late 1970's, and much of the work of the 80's and early 90's has this form. In 1981, Chomsky published an ambitious book in which he organized much of the work of that time into a principles and parameters form. This book, *Lectures on Government and Binding*, serves as a rough starting point for much of my exposition in these lectures.

In moving from a battery of English specific phrase structure rules to the more explanatory interaction between $\overline{X}$ Theory and the c-selection requirements of verbs, language particular settings of modifier types, etc., we have lost some information. Because that transition removed any reference to categories, it is no longer possible to order complements in situations, like (172), where there are more than two.

(172) Sheila put this on the table.
   compare: *Sheila put on the table this.

This information was conveyed in the phrase structure rules by way of referencing category type. We had rules such as (173) for instance which ensure that if a verb is followed by two complements, the first will be the NP.

(173) $\overline{V} \rightarrow$ V (NP) (PP) (CP)

Because the $\overline{\text{X}}$ Skeleton does not have information about category type in it, it is not possible to use the $\overline{\text{X}}$ Skeleton to order complements. Nor would we want to rely on the c-selection requirements of verbs to do this. That would amount to the claim that the order of complements varies as a function of the verb involved. But the fact of English is that no matter what verb is selected, the complements line up in the way that (173) requires.

There must be another component of the grammar which expresses this information. Indeed, we have several outstanding problems concerning the position that arguments take. For instance, we have the paradoxical behavior of CP objects yet to understand. How are object CPs at once within the lowest $\overline{\text{X}}$ with respect to constituency tests like *do so* anaphora, but also linearized so that they appear after modifiers that are not within this $\overline{\text{X}}$? Among the loose-ends fraying at the edges of the proposals presented in this chapter, the correct characterization of argument placement looms large. This will be the subject of the next chapter.

Bibliography

Baker, C. L. 1978. *Introduction to generative transformational syntax*. Englewood Cliffs, New Jersey: Prentice-Hall.

Baker, Mark C. 2003. *Lexical categories*. Cambridge: Cambridge University Press.

Beckman, Mary, and Janet Pierrehumbert. 1986. Intonational structure of Japanese and English. *Phonology* 3:255–309.

Bod, Rens, Jennifer Hay, and Stefanie Jannedy. 2003. *Probabilistic linguistics*. Cambridge, Massachusetts: MIT Press.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32:45–86.

Bresnan, Joan. 1972. Theory of complementation in English syntax. Doctoral Dissertation, Massachusetts Institute of Technology.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, Massachusetts: M.I.T. Press.

Chomsky, Noam. 1970. Remarks on nominalization. In *Readings in English transformational grammar*, ed. J. Jacobs and P. Rosenbaum, 184–221. Waltham, Massachusetts: Ginn.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht, The Netherlands: Foris Publications.

Chomsky, Noam. 1985. *The logical structure of linguistic theory*. Chicago: University of Chicago Press.

Chomsky, Noam. 1986a. *Barriers*. Cambridge, Massachusetts: MIT Press.

Chomsky, Noam. 1986b. *Knowledge of language*. New York, New York: Praeger Publishers.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, Massachusetts: MIT Press.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English*. New York: Haper & Row, Publishers.

Chomsky, Noam, and George Miller. 1963. Finitary models of language users. In *Handbook of Mathematical Psychology* 2, ed. Luce, Bush, and Galanter, volume Handbook of Mathematical Psychology 2, 419–491. New York, New York: Wiley and Sons.

Dowty, David. 1989. On the semantic content of the notion of 'thematic role'. In *Properties, types and meaning: Semantic issues*, ed. Gennaro Chierchia, Barbara H. Partee, and Raymond Turner. Dordrecht, The Netherlands: Kluwer Academic Press.

Dowty, David R. 1979. *Word meaning and Montague grammar*, volume 7. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68:81–138.

Farmer, Ann K. 1980. On the interaction of morphology and syntax. Doctoral Dissertation, Massachusetts Institute of Technology.

Fillmore, Charles. 1968. The case for case. In *Universals in linguistic theory*, ed. Emmon Bach and Richard Harms, 1–90. New York, New York: Holt, Rinehart and Winston.

Frazier, Lyn. 1978. On comprehending sentences: Syntactic parsing strategies. Doctoral Dissertation, University of Connecticut, Storrs.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68:1–76.

Godfrey, J., E. Holliman, and J. McDaniel. 1992. SWITCHBOARD: telephone speech corpus for research and development. In *Proceedings of ICASSP 92*, 517–520. San Francisco.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of language*, ed. Joseph Greenberg. Cambridge, Massachusetts: MIT Press.

Grimshaw, Jane. 1979. Complement selection and the lexicon. *Linguistic Inquiry* 10:279–326.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, Massachusetts: MIT Press.

Gruber, Jeffrey S. 1965. Studies in lexical relations. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Harris, Zellig. 1946. From morpheme to utterance. *Language* 22:161–183.

von Humboldt, Wilhelm. 1836. *über die Verschiedenheit des Menschlichen Sprachbaues und Ihren Einfluss auf die Geistige Entwickelung des Menschengeschlechts*. Berlin: Royal Academy of Sciences of Berlin.

Inkelas, Sharon, and Draga Zec, ed. 1990. *The phonology-syntax connection*. Chicago Illinois: University of Chicago Press.

Jackendoff, Ray. 1977. *X′ Syntax*. Cambridge, Massachusetts: MIT Press.

Kimball, J. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2:15–47.

Ladd, Robert D. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Lasnik, Howard, and J. Kupin. 1977. A restrictive theory of transformational grammar. *Theoretical Linguistics* 4:173–196.

Liberman, Mark, and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8:249–336.

Marcus, Mitchell, Beatrice Santorini, and May Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn treebank. In *Computational Linguistics 19*, 313–330.

Nespor, Marina, and Irene Vogel. 1986. *Prosodic phonology*. Dordrecht, Holland ; Riverton, N.J.: Foris.

Partee, Barbara Hall, Alice G. B. ter Meulen, and Robert Eugene Wall. 1990. *Mathematical methods in linguistics*. Dordrecht ; Boston: Kluwer Academic.

Pesetsky, David. 1982. Paths and categories. Doctoral Dissertation, Massachusetts Institute of Technology.

Pierrehumbert, Janet. 1980. The phonetics and phonology of English intonation. Doctoral Dissertation, Massachusetts Institute of Technology.

Rochemont, Michael, and Peter W. Culicover. 1990. *English focus constructions and the theory of grammar*. Cambridge: Cambridge University Press.

Saito, Mamoru. 1985. Some asymmetries in Japanese and their theoretical implications. Doctoral Dissertation, Massachusetts Institute of Technology.

Saito, Mamoru, and Naoki Fukui. 1998. Order in phrase structure and movement. *Linguistic Inquiry* 29:439–474.

Sapir, Edward. 1921. *Language: an introduction to the study of speech*. New York: Harcourt, Brace and Co.

Selkirk, Elisabeth O. 1980. The role of prosodic categories in English word stress. *Linguistic Inquiry* 11:563–605.

Selkirk, Elisabeth O. 1984. *Phonology and syntax : the relation between sound and structure*. Cambridge, Mass.: MIT Press.

Selkirk, Elisabeth O. 1986. On derived domains in sentence phonology. *Phonology* 3:371–405.

Selkirk, Elisabeth O. 1996. Sentence prosody: intonation, stress and phrasing. In *The handbook of phonological theory*, ed. J. A. Goldsmith. London: Blackwell Publishers.

Stowell, Timothy. 1981. Origins of phrase structure. Doctoral Dissertation, Massachusetts Institute of Technology.

Travis, Lisa. 1984. Parameters and effects of word order variation. Doctoral Dissertation, Massachusetts Institute of Technology.

Truckenbrodt, Hubert. 1995. Extraposition from np and prosodic structure. In *Proceedings of the North East Linguistic Society 25*, ed. Jill Beckman, 503–518. University of Pennsylvania: Graduate Linguistic Student Association.

Truckenbrodt, Hubert. 1999. On the relation betwen syntactic phrases and phonological phrases. *Linguistic Inquiry* 30:219–255.

Zec, Draga, and Sharon Inkelas. 1990. Prosodically constrained syntax. In *The phonology-syntax connection*, ed. Sharon Inkelas and Draga Zec, 365–378. Chicago: University of Chicago Press.