We'll be writing grammars, then: models of our syntactic knowledge. More specifically, we'll be formulating models of our ability to distinguish those strings of words that make sentences, from those strings of words that don't. These models will be rather "high level," that is, they will express the relations, the functions, patterns, etc., that characterize those judgments.

We will often be confronted with the "problem" of having too many ways of describing what we see. It would be nice to have a way of choosing among the different ways of characterizing the facts that we are confronted with. It'd be nice if we could find a way of deciding which of the various ways of characterizing our facts is right.

There is a way of thinking about how to approach this problem that is very common. That way involves taking into account the fact that this knowledge is learned. The grammar that constitutes our knowledge of English is not something we were born with, of course. It is something that we acquired by being present in an English speaking environment. So one thing that must be true of this knowledge is that it is something that could be learned. This, it turns out, is a deeply mysterious fact about language. Not only do we not know how children manage to acquire the knowledge of the language that is spoken around them, we do not know even how it is possible. It's clear that children are not explicitly taught this knowledge. And it's also clear that the grammars of languages are very complex, subtle, things. There's an awful lot that must be learned, and yet there is very meager teaching involved. The distance, then, between what is acquired and the information made available about what is acquired, is gigantic. We should discover the secret to language acquisition. How is it possible?

We can use this mystery to help us decide which of our equivalent ways of describing some phenomenon is the best way. If we had a way of telling which of two otherwise equivalently good grammars is more easily learnable, then we'd have grounds for choosing that grammar. If we are attentive to the learnability issues involved, we might be able to see a way of favoring one grammar over another. We shall have occasion to use this consideration. Let me present the idea in a slightly more specific way.

We may not know how kids acquire languages, but we can nonetheless give a schema that brings out what happens. A child, presumably, gathers information about the language that she is trying to acquire and formulates a hypothesis about what the grammar of that language is based on that information. We could represent that hypothesis with the same sort of high-level characterization of the relations, functions, patterns, what have you, that our own hypotheses about the grammars we'll study are. As the child gathers more information about the language, it is likely that she will encounter a reason to change her hypothesis. So, we could diagram that as follows.

(1) Hypothesis 1 → Hypothesis 2

And of course, this process could continue. At some point it will stop, and that will be the grammar that constitutes the adult grammar of the language.

(2) Hypothesis 1 → Hypothesis 2 → Hypothesis 3 ... → Adult grammar

We know that this chain of hypotheses exists because we know: (1) children do not know the language they are going to acquire at the beginning, (2) they do know the language at the end, and (3) there are intermediate states of knowledge. We also know that this process is deterministic and reasonably easy. Under rather different conditions, a child is able to go through this process and arrive at a grammar that is usually indistinguishable from the one used by the adults around them. We may not know how children form the hypotheses they do, or how they transition from one hypothesis to another, but we know that they do, and that the process (almost) always ends with the same grammar. We can expect, therefore, that the grammar acquired will have properties that reflect this process. The grammar acquired must be one that could be arrived at through this process. It must be acquirable. Because the process is deterministic, the initial hypothesis, and the transitions from one hypothesis to the next, must not vary too much from child to child. The acquired grammar, then, should be in the space of possible hypothesized grammars, and those cannot be too large a set to begin with, or this process would not be deterministic. There must be properties, then, that are found in all adult grammars. Those properties are the ones that distinguish the set of hypotheses a child entertains from all those hypotheses the child doesn't. We can look for these properties in our own models of adult grammars. We can ask ourselves: is there some property that is found in only one of the two grammars we are comparing that would...
make that grammar belong to a narrower class of possible grammars than the other. If so, we should choose that grammar, because that grammar belongs to a sequence like that in (2) that could be bounded by that property.

One knows a lot about how a collection of words can be arranged grammatically just by knowing what morpho-syntactic “category” that word belongs to. We can see that with the following “experiments.”

(3) Many bloresnicks are grey.

(4) a. It ran bloresnick the tree.
   b. He removed the long bloresnick.
   c. She finds Sammy bloresnick.
   d. He made his face bloresnick.

(5) pondel: (⟨pondel⟩) unwanted facial hair.

(6) 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.

The linguist Zellig Harris went with the first experiment’s results, and this is the usual method now.

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

For example, “nouns” are defined by what can stand in “___” below and return a grammatical judgment on the resulting sentence.

(8) ___ exists.

This is indicated by considering the lists of sentences in (9)-(14).
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 (17).

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 (17).

(15)  
  a. have ___ eaten: **Adverb**  
  b. the ___ thing: **Adjective**  
  c. dance ___ it: **Preposition**  
  d. in ___ orange: **Determiner**  
  e. must ___ there: **Verb**

(16)  
  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. and so on

(17) **Determiner** ___ **Verb**

Each of the environments in (15) can be similarly converted into a generalization that makes reference to morpho-syntactic category.

(18)  
  a. **Verb** ___ **Verb**: **Adverb**  
  b. **Determiner** ___ **Noun**: **Adjective**  
  c. **Verb** ___ **Noun**: **Preposition**  
  d. **Preposition** ___ **Noun**: **Determiner**  
  e. **Aux** ___ **Preposition**: **Verb**

In some cases, we can also use morphology to define categories. So, for instance, only nouns combine with the plural suffix -s. And only verbs combine with the past tense suffix d.

With the recognition that our vocabulary can be divided into classes in this way, we might take as our first idea that our knowledge of syntax consists of a list of strings of categories. Something like (19) for instance.

(19)  
  a. D A N V  
  b. D A N V D N  
  c. D N V P D N  
  d.  

That won't do, however, because this list would have to be infinite in length and if what we are characterizing is knowledge that is represented in an individual's mind, something of infinite length is too long. Why do I think this list is infinite in length? Because of series like (20).

(20)  
  a. The dog barked.  
  b. The big dog barked.  
  c. The big black dog barked.  
  d. The big noisy black dog barked.  
  e. The big noisy hungry black dog barked.  
  f. The big noisy hungry scary black dog barked.  
  g.  

My guess is that where this series ends is not determined by English. That is, it's not something that is in the grammar of English that fixes the end point of this series. Instead, it's some non-linguistic thing — like good sense. If that guess is correct, then English should not end this series and it should, therefore, be infinite.

So a simple list of strings will not do. A slightly more sophisticated grammar that can handle infinity employs the machinery of a finite state automaton. A finite state automaton is a model that can handle very simple calculations, and it is capable of characterizing strings of infinite length. One way of describing a finite state automaton is as follows.

(21)  
  a. There a finite number of states  
  b. One state is designated as the start state  
  c. Some states are designated as the final states  
  d. There are transitions that connect states  
  e. Any number of transitions may exit or enter a state  
  f. Each transition is associated with writing a symbol
Here's a way of representing such an automaton that will produce the series in (20).

The states are numbered circles, and the transitions are represented by arrows connecting states. Each transition is associated with writing a symbol that represents one of our categories: "D" for determiner, "N" for noun, and so on. The final state is indicated by a double circle. Each string you can imagine being paired with a particular "derivation." To get "D N V," you'd transit through the automaton passing through each state once. To get "D A N V," you'd transit through state 2 twice.

One of the things we will learn in this class is that there are properties of the models we will use to characterize syntax, limits, if you like, to their expressive power, and one of our jobs is to match those limits to those that we find actually obtain of syntax. What we're trying to do is build into our model of syntactic knowledge the very properties and characteristics that our syntactic knowledge has.

So, if finite state automata are the right way of modeling what our syntactic knowledge is, then we should learn something about the expressive capabilities of finite state automata. That's partly what assignment one does. But let's play around a little bit with some examples to see what you can do with a finite state automaton.

What are the strings that are produced by (20)?

Are these strings of English?

Make the determiners in the strings produced by (22) optional. What are the strings that (23) produces?

Are these strings of English?