website: http://web.mit.edu/norvin/www/LSA/home.html 32-144
preliminaries: intro means intro, please be vocal!,
weekly writing assignments, no textbook, but I'll put up lecture notes and readings.

We seem to know more about language than we have any business knowing.

First example: a case where this isn't true; data undetermine the answer, and people just guess. Tagalog –um-: gumradweyt, grumadweyt.

but:

(1) English, Japanese, Bulgarian multiple-wh
concentrate on Bulgarian; infinitely many possible hypotheses.
everyone converges on the same one, even though there's very little evidence and what there is underdetermines the solution.

(2) The man is sick
The man is claiming that Syntax is lucrative
The man who is talking is bearded
(and this is perverse: computer programming, Lardil)

(3) Mary kept the car in the garage (ambiguous)
Which car did Mary keep in the garage? (unambiguous)

(4) Will the men expect to see them? (men≠them)
Who will the men expect to see them? (men can be them)

**poverty of the stimulus** argument for **innateness**

(Legate and Yang, Sat, Jul 16, "Happy Golden Anniversary, Generative Syntax")

despite the lack of relevant data, everybody converges on the same answers.
Individual variation in lots of domains, but not this.
not an argument that learning plays no role;
just that we're structured to only consider some of the logically possible hypotheses; we're 'born knowing' certain things about how language works. This innate knowledge is **Universal Grammar** (sort of unfortunate term; it's not a grammar of a particular language).

So what is this knowledge? and what's the content of the grammars that we end up with as adults? This is our quest.

We're going to be studying syntax specifically. That means making some simplifying assumptions:

**autonomy of syntax**
colorless green ideas not just "meaningful"

decoff caffee, put up it not just "understandable"

**competence vs. performance**:
what you accept as grammatical vs. what you actually say.

filtering out speech errors, mortality (utterances are potentially infinite in length: house-that-Jack-built sentences, mother's mother's mother's mother)

…. hypothesis is that there's an autonomous syntactic component that in an ideal world would produce perfectly grammatical utterances all the time, which is interacting with other systems (memory, attention, etc.) to give you actual performance.

(how do we figure out what belongs to syntax? we won't know for sure until we have complete theories of the different components contributing to performance; for now, we do what we can with what we understand about these components.)

One thing we know about the syntactic component is that we want it to be **generative**; since the utterances are in principle infinitely long, it can't just be (say) a list of all the sentences you've heard. Has to be a set of procedures for generating sentences, "making infinite use of finite means" (von Humboldt)

So what are these procedures? What's the nature of the initial state, and of the final state?

consider a sentence:

I will find the red book

Lots of evidence that this sentence isn't just a string of words; there are substrings (**phrases**) that are privileged as objects that the syntax can manipulate.

1. topicalization:
   The red book, I will find
   …and find the red book, I will
   *find the red, I will book

   (*=ungrammatical)

2. pseudoclefts:
   what John will find is [the red book]
   what John will do is [find the red book]
   *what John will do book is [find the red]

3. sentence fragments:
   (what did you find?) The red book
   (what will do you?) Find the red book

(???) Find the red

what part of "the red book" is necessary for it to act this way? Just the noun, looks like ("the book" and "books" behave the same way). So we'll call that the "noun phrase", and by similar reasoning, "find the red book" is a "verb phrase". "I" is also a noun phrase (consider some examples).

and behold the prepositional phrase: Mary will put the book on the table.

(run through your tests, and you'll see that 'on the table' is a phrase, and part of the verb phrase)

phrase structure rules:

NP→ (Det) (A) N
VP → V (NP) (PP)
PP→ P (NP) ("I walked in")

Almost allows for complete diagrams of our sentences; only "will" is missing. All our rules so far are **endocentric** (phrases consisting of a **head** after which the phrase is named, plus some other optional stuff), so let's make the sentence a TP ("S" in older terminology).

TP→ NP T VP

Now all we need is a **lexicon**, and we have a very simple fragment of English.

V=buy, eat, put…
N=John, book, chocolate…

Add embedded clauses:

I think <u>that</u> John likes chocolate
I wonder <u>whether</u> Mary likes chocolate

CP→ C TP (S' in older terminology)
VP→ V (NP) (PP) (CP)

now it's **recursive**, hence generates potentially infinite sentences:

I think that John knows that Mary wonders whether…

Similarly, PP should be an option in NP:

John is tall
the man is tall
the man [with the knife] is tall

NP→ (Det) (A) N (PP)

and since PPs contain NPs, we have recursion again:
the man with the knife with the handle with stars of ivory from elephants with…

Our rules also let us do attachment ambiguities:

Mary saw the man with the binoculars.

(2 readings; diagram, do tests to show that these are the right constituent structures).

now these rules have two problems:
(1) largely redundant with stuff we're going to have to state in the lexicon anyway (**selection**)
(2) insufficient internal structure (more about this next time)

Consider the VP rule:

VP → V (NP) (PP) (CP)

We said before that we're going to combine these rules with lexical entries to get fragments of English. Lots of lexical choices won't work out, of course:

the dragon
- …devoured the villagers
- *…devoured
- …rejoiced
- *…rejoiced his victory

John…
- …put the book on the table
- *…put the book
- *…put on the table

classic observation, kicked up a notch; some verbs are <u>transitive</u>, requiring NP objects, and others are <u>intransitive</u>, disallowing them; and in fact verbs can require other things, like PPs (cf. *put*).

So in the lexical entry for *put* we'll need to have the information "needs an NP and a PP". Once we have that information there, do we really want to restate it in a phrase structure rule?

**selection**, which we'll be talking about more later. Heads get to **select** for certain types of phrases.