1. This is the same sentence in English and another language. Draw lines connecting words that you think are probably translations of each other. (Pay attention to repetition in the foreign language . . . )

   Jacob had some lentils that Esau wanted.

   us manchat-aa e yakab, uss manchat-aa karpuo esa

   “yakab” and “esa” translate to “Jacob” and “Esau.” A literal word-by-word translation of the whole sentence is “Some lentils had Jacob, the lentils wanted Esau.” The word order is different from English. Notice that the foreign language quite logically mentions “manchat-aa” twice, since both brothers are interacting with the lentils. But the English version uses the special word “that” to avoid having to repeat “lentils.” (To be precise, the English uses a relative clause; many but not all languages allow this option.)

2. For each of the following ambiguous sentences, underline the most ambiguous part, and indicate whether the ambiguity is lexical, syntactic, semantic, or pragmatic. (Note that ambiguity at one level will lead to ambiguity at later levels, so specify only the lowest level at which the ambiguity arises. You are welcome to write a few words justifying your answer.)

   _pragmatic_ Are you wearing a watch?
   _semantic_ Don’t put no pennies in my cup.
   _lexical_ Leia threw up her hands and Luke his dinner.
   _syntactic_ But delegates dropped language condemning Israel as a racist state, which prompted a U.S.-Israeli walkout. [this sentence from CNN.com]

The answers above were what I had in mind (though other answers could get credit if you explained them).

- **Lexical ambiguity** means it’s not clear what words you’re dealing with—“arms” the body part, or its homonym, “arms” the munition? “throw up” meaning vomit, or “throw” meaning toss into the air?

- **Syntactic ambiguity** is ambiguity in the sentence diagram. It arises if you know what all the content words mean, but it’s not clear what the phrases are, or which phrases modify one another. Thus, what is the implied subject of “prompted a U.S.-Israeli walkout”? A computer would have to know a lot about the context to figure out that it was the language that prompted the walkout, and not the dropping of the language!

- **Semantic ambiguity** arises if the sentence diagram is clear, but the literal meaning of the sentence is still not clear. (Our example in class was “A woman gives birth every 15 minutes”—is it the same woman?)
To see the ambiguity in this example, try stressing different words when you read it aloud. “Don’t put no pennies in my cup” could be a double negative: “Don’t put NO [= zero] pennies in my cup, put SOME pennies in.” But in many Southern and African-American dialects of English, the two negative words are interpreted as reinforcing each other rather than canceling each other out: “Don’t put no [= any] PENNIES in my cup, put QUARTERS in instead.”

(Before you say this latter interpretation is illogical, notice that it is exactly how French, Spanish, and Italian work! All languages and dialects—including Standard English—have their own quirky internal rules, which are neither logical nor illogical. And they are all equally interesting to linguists who want to understand the miracle of language—and to computer scientists who want to build working systems for processing language as it is actually used.)

- Even if the literal meaning is clear, you still have to figure out why the speaker said it. This is a matter of pragmatics (i.e., conversational conventions). The first sentence could be meant literally, or it could be a polite way of asking for the time.

3. The lecture suggested that this field goes through the following development cycle:

(a) Formalize some insights that might help system performance.
(b) Study the formalism mathematically.
(c) Develop and implement algorithms.
(d) Test on real data.

Label each of the following tasks as a, b, c, or d, according to the development stage that it is most likely to fall in.

- **d** parsing sentences
  (i.e., try the parser out once you built it)
- **a** studying the parsed sentences
  (what’s wrong with the output? how can you do better next time?)
- **c** writing grammars
  (part of the job of implementing an English parser, or an Arabic parser)
- **d** multiplying probabilities
  (a parser does this to try to figure out the overall probability of a possible parse)
- **b** finding a grammar that can be written down in one notation but not in a different notation
  this proves the two notations are inequivalent—they represent different formalisms
- **d** catching up on sleep (why? you have to wait for the program to run!)

4. Can you think of a field outside computer science that goes through the same kind of development cycle, trying to produce elegant theories that approximate hopelessly complicated data? Explain your answer in a sentence or two.

There are lots of possible answers here. My favorite is economics, which does a passable job of explaining all buying and selling behavior—from eBay to the stock and bond markets—by reducing human psychology to selfish preferences and a small set of strategies. In fact, some of the same statistical techniques used in natural language processing and speech recognition are also useful for predicting the market!

---

1For example, isn’t Standard English “illogical” because “-s” makes nouns singular but verbs plural?