1. What is the goal of speech recognition? What is the goal of text-to-speech (speech synthesis)? In each case, what is the input, what is the output and what are the challenges? In which case, it is desirable to have "coarticulation" phenomenon in speech and in which case not?
2. In the lecture, we observed that in both speech recognition and speech synthesis systems each word is modeled as a sequence of "phones", also referred to as pronunciation of the word. For example, in English, BAT: /b/ /ae/ /t/, CAT: /k/ /ae/ /t/ and so on. How can these phone sequences, i.e., pronunciations, be obtained for the different words in a language? Can a language ever have a "closed set" vocabulary, i.e., the set of distinctive words does not changes over time? Justify your answer.

   Suggested link: http://www.speech.cs.cmu.edu/cgi-bin/cmudict