Fall Semester 2016
Iowa State University

ENGL 520 - Computational Analysis of English

Problem Set 6
NLP for CALL
(ungraded)

1. When people are learning a second language, they may still have some native language influence in their writing. Try to pick two or three such hypotheses for English learners (of any one native language background), and write a program that identifies such features in L2 writing.

2. Learn to use a pre-existing spell checker in your program and identify spelling errors in the texts given as input to the program (figure out if there are any existing spell check programs available as perl or python libraries, or those that provide a interface to access their results programmatically!)

3. Let us say a user typed "Sleep" for "Creep". How do you calculate the edit-distance between these two words using the distance matrix? Draw the matrix and calculate the distance between two words, assuming that all operations (insertion, deletion, replacement etc) have 1 point.

4. Spelling checkers typically correct words in isolation. What in your view is the best way to attempt spelling correction in context? Why do we need context at all? Write an analysis, write like a language technologist - not as a ESL teacher.

5. Write a program that takes a English word as input and converts it into Arpabet transcription format (use lowercase instead of upper case in Arpabet transcriptions). The format is described in Chapter 7 (first table) in Jurafsky and Martin textbook. The format is also accessible online at: `http://fave.ling.upenn.edu/downloads/ARPAbet.pdf`. Your program does not have to be perfect, I understand syllabic transcription for English is not a straight forward problem. But the program should demonstrate that you thought through your mapping rules, and it should show some output for any English word entered as input.

6. Take any one CALL system that you know of, and use or have access to. Write a (linguistic) analysis of where it works, in what cases it fails, and why it may be failing. Consider atleast two failed cases.

7. Assume that the CALL system you are developing expects short-answer responses from students (1–3 sentences). How will you evaluate these answers automatically if (a) you have some gold standard answers to compare against (b) you don't have a reference answer? What are the possible issues you face when you begin to develop a short answer scoring program?

8. Use the FLAIR system (`http://goo.gl/9vRpz4`), and study its functions. Now, write an analysis of where and how it could be useful for language learners, what are its shortcomings, and how it can be improved.

9. Investigate how you can accept speech input from users in your program, and convert it to text (without doing any speech processing on your own!). Look for such libraries in Perl and Python that support this function for you.

10. Let us assume you are asked to write a program to categorize English learner essays as beginner, intermediate and advanced. How will you approach the problem if a) you are given a set of 1000 human graded essays b) you are not given any such reference essays.