CMPE 409 - Machine Translation

Final Project

Deadline: 23:00, Jun 12

In your assignment,explain your codes with *comments*. Without *comments*, your assignment will not be marked.

Problem

In this project you are asked to write a spell correcting project with using language models such as bigram and trigrams. To do this you may improve your second assignment with necessary data. First of all, make a corpus which contains about 100,000 more sentences (you may use news page such as "hurriyet", "milliyet") etc. Then make your table.
You may make a group with **maximum 4 students** and make the presentation together. It doesn't mean all student will get the same mark. Still you may get different marks according to you performance.

- There are 9222 sentences in the corpus.
- Raw biagram counts of 8 words (out of 1446 words)

	i	want	to	eat	chinese	food	lunch	spend
i	5	827	0	9	0	0	0	2
want	2	0	608	1	6	6	5	1
to	2	0	4	686	2	0	6	211
eat	0	0	2	0	16	2	42	0
chinese	1	0	0	0	0	82	1	0
food	15	0	15	0	1	4	0	0
lunch	2	0	0	0	0	1	0	0
spend	1	0	1	0	0	0	0	0

Figure 1: Biagram Table

In your code:

- You may use NLTK package
- See the lecture slides of week-10 how to calculate biagrams.
- At last, make a raw bigram table for a sentence (10 words long).

- Calculate 2 or 3 sentences' probability with using your table.
- Write methods that correct miss splled words:
  - Should find possible wrong words
  - Results about correcting
  - Prints out related messages (about calculation)
  - NOTE: you should able to explain the results during presentation
- Write full report that contains data and test report of your examples. Test considering all possible cases. During presentation your project will be tested by random sentences.

Submission

- Submit your source code.
- readme.txt file. This file includes information such as configuration and setup and execute...
- Submit your report: It contains description of the project and analysis of the result. You may support your explains with screen shot pictures.

Hint: Look at the lecture notes and examples as references