Instructor: Prof. Nianwen Xue (email: xuen@brandeis.edu)  
Teach Assistant: Alex Plotnick  

Course Description  

This is an introductory graduate-level course on the computer processing of natural language text with the Python programming language. Python has quickly become the most popular programming language in Natural Language Processing (NLP) since it first came into existence because it has built-in data structures that allow natural language text to be manipulated with ease and elegance. Python has a relatively short learning curve compared with other high-level programming languages such as Java and beginners in Python can build up their programming skills fairly quickly. In addition, a large number of Python modules already exist for language processing purposes so that linguistically oriented Python programmers can start to write practically useful programs within a relatively short period of time.  

The main objectives of the course are:  

- Provide an introduction of the fundamental data structures and programming constructs in Python, and,  
- Provide an introduction of the basic programming techniques and linguistic resources that contribute to solving natural language problems, and,  
- Help students acquire solid Python programming skills by manipulating natural language data, and,  
- Provide an introduction of the fundamental techniques used in lexical, morphological, syntactic, and semantic analysis.  

The key to being successful in the course is to get your hands dirty and write a lot of code. By taking this course you have shown a commitment to become proficient in programming. If you have never written any code before you may need to adapt to a new learning style that is practice-oriented rather than reading-oriented.  

This course is a recommended class for both the MA in Computational Linguistics (CL) and the undergraduate major and minor in Language and Linguistics. It is a required course for students with no Computer Science background, and an elective for students with Computer Science background, in the newly approved BA/MA and Two-Year MA Programs in Computational Linguistics. The course is also designed to help Linguistics majors to acquire programming skills to manipulate large amount of linguistic data, a necessary skill to conduct linguistic analysis with a wide empirical basis. The course is also intended to help CL MA students with little computer science background to build up programming skills necessary for more advance Computational linguistics courses.  

Course Content  

Weeks 1-2: Python preliminaries  
Weeks 3-4: Language processing and Python
Weeks 5-6: Accessing text corpora and lexical resources
Weeks 7-8: Processing raw text
Weeks 9-10: Writing structured programs
Weeks 11-12: categorizing and tagging words
Weeks 13-14: Analyzing syntactic structure

Required Reading

Recommended Reading
*Learning Python*, by Mark Lutz and David Ascher, O’Reilly Media Inc. 2004

Prerequisites
The prerequisite for this course is LING 100a or the approval of the instructor.

Grading
- 5% class participation
- 50%: regular exercises and projects (approximately 5 programming assignments)
- 20%: in-class quizzes
- 20%: final project
- Assignments and projects need to be turned submitted on time to receive full credit.

Academic Integrity:
Each student in this course is expected to abide by the Brandeis University Student Development and Conduct (http://www.brandeis.edu/studentaffairs/sdc) Handbook on Rights and Responsibilities, with particular emphasis on section 4. Any work submitted by a student in this course for academic credit will be the student’s own work.