# Spring 2006 L44 Ling 317: Introduction to Computational Linguistics

Use of computers to analyze, understand, and generate human language. Emphasis on appreciating practical applications such as text analysis, search and creation of dictionaries and corpora, information retrieval, machine translation, and speech interfaces. Survey of rule-based and statistical techniques. Students will acquire programming skills appropriate for solving small- to medium-scale problems in linguistics and text processing, using a language such as Perl. Students will have regular programming assignments and will complete a semester project. No previous knowledge of programming required. Prerequisites: L44 Ling 170D or permission of instructor.

| | |
|---|---|
| *Venue* | Lopata Hall 103, Tue, Thur 2:30–4:00 |
| *Instructor* | Brett Kessler |
| *Web* | http://brettkessler.com/CompLing |
| *E-mail* | bkessler@wustl.edu |
| *Phone* | 314-935-8839 [but prefer e-mail] |
| *Office* | Psychology 235A |
| *Visitation hours* | Wednesday 1:00–3:00 or by appointment |

*Required Text:* Jurafsky, Daniel, & Martin, James H. 2000. *Speech and language processing*. Upper Saddle River, NJ: Prentice Hall.

*Recommended Text:* Schwartz, Randal L., Phoenix, Tom, & Foy, Brian D. (2005). Learning Perl. – 4th ed. – Sebastopol, CA: O'Reilly.

*Web Site:* The web site for the course will frequently contain supplemental and updated material. Logon at http://brettkessler.com/CompLing and give CompLing for the user name. Password will be given in class.

*Assignments:* There will be approximately 8 assignments, mostly small programming exercises. You may collaborate on assignments, but you should turn in your own, unique, solutions, and you must disclose who you collaborated with. Unless otherwise provided, assignments should be submitted electronically.

*Project:* There will be a semester project. This will involve a class presentation in the last week or two of class, and also a final, written report, due May 10. You are encouraged to select your own project, but it needs to be approved by me, preferably in the first few weeks of the course. The project will normally involve programming, but other possibilities will be considered. You are encouraged to collaborate in small groups, and may turn in a single co-authored report. However, each student is expected to make a substantial contribution to the project and to be able to answer questions about the project as a whole.

*Participation:* Participation in class discussions is important, and therefore you should come to class. Excessive unexcused absences (more than 10%) may negatively impact your grade.

*Grading:* Assignments will be worth 40% of the semester grade; participation, 10%; and the semester project, 50%. Letter grades will be assigned on the following schedule:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 98 | A+ | 88 | B+ | 78 | C+ | 68 | D+ |
| 93 | A | 83 | B | 73 | C | 63 | D |
| 90 | A− | 80 | B− | 70 | C− | 60 | D− |

Midterm grades will be pro-rated based on participation and assignments completed before spring break. Grades of Pass will be assigned under the Pass/Fail option if the student receives 60 points or more. No grades of Incomplete will be assigned.

*Other policies:* The policies of Washington University and the College of Arts and Sciences apply except for defaults overridden by this syllabus or the agreement of the class.

*Schedule:* This timeline is subject to change at the instructor's discretion. Reading assignments should be read before the start of class on the day they are listed. It is recommended that you read them again after class and work through some exercises as well.

| **Date** | | | **Chapter** | **Topic** |
|---|---|---|---|---|
| Tu | Jan | 17 | | What is this course? |
| Th | Jan | 19 | | Programming environments |
| Tu | Jan | 24 | 1 | Programming languages |
| Th | Jan | 26 | | Asst: Get Perl running |
| Tu | Jan | 31 | 2; App A | Regular expressions and finite state automata |
| Th | Feb | 2 | | Asst: Using regular expressions to search text |
| Tu | Feb | 7 | | Frequency analyses and concordances |
| Th | Feb | 9 | | Asst: Counting wordform frequency |
| Tu | Feb | 14 | | Reporting results with XML and the Web |
| Th | Feb | 16 | | Asst: Publish frequencies on the WWW |
| Tu | Feb | 21 | 3; App B | Computational morphology |
| Th | Feb | 23 | | Asst: Word stemming |
| Tu | Feb | 28 | 4 | Handling IPA and foreign-language characters with Unicode |
| Th | Mar | 2 | | Asst: Foreign-language word stemming |
| Tu | Mar | 7 | 5 | Computational phonology and orthography |
| Th | Mar | 9 | | Asst: Text-to-phonemes or vice versa |
| Tu | Mar | 21 | | Information retrieval and mining |
| Th | Mar | 23 | | Asst: Quantifying linguistic structure |
| Tu | Mar | 28 | 8 | Syntax |
| Th | Mar | 30 | | |
| Tu | Apr | 4 | | Phrase structure rules |
| Th | Apr | 6 | | Asst: Generating sentences |
| Tu | Apr | 11 | 14 | Semantics |
| Th | Apr | 13 | 21 | Machine translation |
| Tu | Apr | 18 | 6 | Stochastic techniques |
| Th | Apr | 20 | 7 | Speech recognition |
| Tu | Apr | 25 | | Project presentations |
| Th | Apr | 27 | | Project presentations |
| We | May | 10 | | Final report due, my office |