Linguistics 409 – Spring 2013
Computational Linguistics
M/W/F 9:00 - 9:50 in HUM 119

Instructor:	Dr. Kevin B. McGowan	Office Hours:	Monday 10:00 - 11:00am
	kmcgowan@rice.edu		Wednesdays 3:00 - 4:00pm
	208 Herring Hall		or by appointment

Course Description and Goals
This course is designed to do two things:

1. This course is primarily an introduction to computational linguistics (with a strong emphasis on the *linguistics*). We will work our way through regular expressions, finite state automata, finite state transducers and morphology, N-gram models, part of speech (POS) tagging, hidden Markov models, text-to-speech, automatic speech recognition, computational phonology, context free grammars (CFGs), syntactic parsing, statistical parsers, feature unification grammars, computational lexical semantics, and computational approaches to discourse. As we cover each computational topic we will also reflect on related advances in research on human language processing and cognition.

   This is pretty ambitious. We may not get to all of it, and we won't go very deeply into any one topic. It is reasonable to think of this as a survey course –like Ling 200 only much, much harder.

2. Secondarily, but just as essentially, this course will introduce (or reinforce!) a set of skills necessary to do computational linguistics. These will include basic UNIX skills, probability, phonetic transcription, phonological analysis, morphological analysis, syntactic parsing, lamda

Required Course Background Prerequisites: LING 200 or equivalent knowledge of linguistics.
What I assume you remember from this course:

- the basics of phonetics, phonology, morphology, syntax, semantics, and discourse.

- the general goals of the scientific study of language.

Required Textbooks

Daniel Jurafsky and James H. Martin. 2008. *Speech and Language Processing: An Introduction to Natural Language Processing Computational Linguistics, and Speech Recognition* **Second Edition**, ISBN 0-13-187321-0. Important: Be sure you get the second edition; the first edition is ancient and buggy.
Allen Downey. 2012. *Think Python: How to Think Like a Computer Scientist.* ISBN: 144933072X
Available free from: `http://www.greenteapress.com/thinkpython/`

The Jurafsky & Martin book should be available at the bookstore or through online retailers. You cannot get away without having and, crucially, reading Jurafsky and Martin. Any other materials we use in class –handouts, exercises, assignments– will be distributed via OwlSpace.

Course Requirements

1. **Homework Assignments & Quizzes** 70%
   There will be several homework assignments and quizzes over the course of the semester. Your lowest homework or quiz grade will be excluded from the calculation of final grades.

2. **Take-home midterm** 15%

3. **Take-home final exam** 15%

Homework Information
Homework sheets will be available for download from OwlSpace. Completed homeworks are to be uploaded to the course website. Unless otherwise indicated on the assignment, hardcopy assignments will not be accepted. I vastly prefer PDF or RTF to proprietary file formats. Please don't give me Word files. Would you give someone a can of delicious beans that could only be properly opened by a single company's proprietary opener? Would you give someone a book that could only be read under a particular brand of lightbulb? No, because those things would be ridiculous.

Late Homework Policy
I will accept late homework if you ask *in advance* for an extension. "In advance" means *more* than 24 hours before the assignment is due. I have been very forgiving about this in the past and it leads to students getting in the habit of not handing in work and me grading piles and piles of work at the end of the semester that I should have been able to do as the semester progressed. I intend not to be very forgiving this semester.

Honor Policy
Appropriating someone else's work and portraying it as your own is cheating. Collaborating with someone and portraying that work as solely your own is cheating. Obtaining answers to homework assignments or exams from previous semesters is cheating. Falsifying data or experimental results is cheating. (The foregoing is not intended to be a complete list. A complete description of Rice's Honor Code, plagiarism, and other general information can be found at the Rice Honor Council Web page at http://honor.rice.edu/). If you are caught cheating, you will be referred to the Honor Council. If you are unsure about whether a specific action is cheating, you may check with the intstructor. Some general guidelines are:

- Do not look at notes, assignments, or exams from previous semesters.

- Do not seek solutions to homework problems or exams from outside sources, including books (other than the textbook) or the internet.

- Do not copy other (current or former) students homework assignments. To minimize this temptation, always type up your homework answers by yourself, separately from your study group or other students in the class.

- Once you have started to work on your acoustics or final exam, do not discuss it with other students, until after you have turned it in and the exam time is over.

- Do not falsify data or other results in your homeworks or Extra Credit project.

- Cite all sources used and cite and designate all quotations as such.

Study Groups

I encourage students to form study groups to talk about readings and lectures, and especially to discuss and work through understanding how to solve homework problems. Invite me if you want and I'll try to come. However, after you figure/argue them out together, *you must do the work and type up your homework answers entirely on your own*, separately from the other study group members. You must also list the names of all of the members of your group at the top of your assignment. Failure to list study group members is an unethical misappropriation of others' contributions without acknowledgement; here I refer you to the previous section.

Americans with Disabilities Act
If any student in the class has a documented disability needing academic adjustments or accommodations, please get in touch with me during the first two weeks of class. All discussions will remain confidential. Students with disabilities will also need to contact Disability Support Services in the Ley Student Center. I look forward to working with you to make this class enjoyable and accessible for all.

Preliminary Class & Reading Schedule

Readings are in **boldface**. Except for Chapter 1, readings should be completed prior to the first day they are listed. Please note that this schedule is preliminary: dates and topics and even readings are subject to change.

Monday	Wednesday	Friday
Jan 7th **1** Overview & What is Computational Linguistics? 1 **JM ch1**	9th **2** What is Computational Linguistics? 2	11th **3** Regular Expressions **JM ch2 pp 17 - 26**
14th **4** Regular Expressions **JM ch2 pp 26 - 44**	16th **5** Basic UNIX skills 1	18th **6** Basic UNIX skills 2
21st	23rd **7** Morphology **JM ch. 3 pp 45 - 52**	25th **8** Finite State Transducers **JM ch. 3 pp 52 - 67**
28th **9** Stemming, Spelling, Edit Distance **JM ch. 3 pp 68 - 79**	30th **10** Probability for Linguists **Abney (OwlSpace)**	Feb 1st **11** N-Gram models **JM ch. 4 pp 81 - 97**
4th **12** N-Gram models **JM ch. 4 pp 81 - 97**	6th **13** Smoothing, Backoff, & Interpolation **JM ch. 4 pp 97 - 109**	8th **14** Part of Speech Tagging **JM ch. 5 pp 123 - 139**
11th **15** Part of Speech Tagging **JM ch. 5 pp 139 - 149**	13th **16** Part of Speech Tagging **JM ch. 5 pp 149 - 157**	15th **17** Hidden Markov Models **JM ch. 6 pp 173 - 183**
18th **18** Hidden Markov Models **JM ch. 6 pp 184 - 192**	20th **19** MaxEnt **JM ch. 6 pp 192 - 200**	22nd **20** MaxEnt **JM ch. 6 pp 201 - 207**
25th	27th	Mar 1st
4th **21** MaxEnt **JM ch. 6 pp 207 - 213**	6th **22** Speech Synthesis **JM ch. 8 pp 249 - 261**	8th **23** Speech Synthesis **JM ch. 8 pp 262 - 276**

MONDAY	WEDNESDAY	FRIDAY
11th **24** Speech Synthesis **JM ch. 8 pp 276 - 284**	13th **25** ASR **JM ch. 9 pp 285 - 302**	15th **26** ASR **JM ch. 9 pp 303 - 314**
18th **27** ASR **JM ch. 9 pp 314 - 331**	20th **28** Context Free Grammars **JM ch. 12 pp 385 - 404**	22nd **29** Context Free Grammars **JM ch. 12 pp 404 - 414**
25th **30** Syntactic Parsing **JM ch. 13 pp 427 - 443**	27th **31** Syntactic Parsing **JM ch. 13 pp 443 - 457**	29th
Apr 1st **32** Statistical Parsing **JM ch. 14 pp 459 - 479**	3rd **33** Statistical Parsing **JM ch. 14 pp 479 - 486**	5th **34** Machine Translation **JM ch. 25 pp 859 - 879 (s1-4)**
8th **35** Machine Translation **JM ch. 25 pp 879 - 899 (s5-10)**	10th **36** Meaning **JM ch. 17 pp 545 - 580**	12th **37** Computational Semantics **JM ch. 18 pp 583 - 598 &605 - 607**
15th **38** Lexical Semantics **JM ch. 19 pp 611 - 633**	17th **39** Computational Lexical Semantics **JM ch. 20 pp 637 - 650**	19th **40** Computational Lexical Semantics **JM ch. 20 pp 650 - 676**