Linguistics 384: Language and Computers

Course goals: In the past decade, the widening use of computers has had a profound influence on the way ordinary people communicate, search and store information. For the overwhelming majority of people and situations, the natural vehicle for such information is natural language. Text and to a lesser extent speech are crucial encoding formats for the information revolution.

In this course, you will be given insight into the fundamentals of how computers are used to represent, process and organize textual and spoken information, as well as tips on how to effectively integrate this knowledge into working practice. We will cover the theory and practice of human language technology. Topics include text encoding, search technology, tools for writing support, machine translation, dialog systems, computer aided language learning and the social context of language technology.

GEC: The course satisfies the GEC category 2B (Mathematical and Logical Analysis). It does so by using natural language systems to motivate students to exercise and develop a range of basic skills in formal and computational analysis. The course philosophy is to ground abstract concepts in real world examples. We introduce strings, regular expressions, finite-state and context-free grammars, as well as algorithms defined over these structures and techniques for probing and evaluating systems that rely on these algorithms. The course goes beyond merely subjective evaluation of systems, emphasizing analysis and reasoning to draw and argue for valid conclusions about the design, capabilities and behavior of natural language systems.

Instructor: Scott Martin (http://www.ling.ohio-state.edu/~scott/)

- *Office:* 204 Oxley Hall
- *Phone:* 614-688-3108
- *Mailbox:* 225 Oxley Hall (end of the hall, by the kitchen)
- *Web Page:* http://www.ling.osu.edu/~scott/teaching/2008/spring/384/
  This course web page contains digital copies of this syllabus, homeworks, etc. The class schedule is maintained on the course Carmen web page, including dates for quizzes, homeworks, and exams.
- *Office hours:* Mondays and Wednesdays 1:30–2:30 PM (in general, arranging an appointment by email works best)
- *TA Coordinator:* Dr. Hope Dawson, dawson.165@osu.edu, 222a Oxley Hall, 688-3109.

Course meets: Mondays at 11:30 AM in 185 Mendenhall Labs and Wednesdays at 11:30 AM in 29 Derby Hall. **Note: There will be no class on Monday, May 26th (Memorial Day).**

Carmen: We'll be using the Carmen course management tool for the course, which is accessible at http://carmen.osu.edu/. You'll use it to

- do the on-line quizzes
- submit your homeworks in the electronic dropbox
- locate/read the updated syllabus, schedule, slides, handouts, reading material, etc.
- send email to the class

Note that email from Carmen is sent to the official email addresses (`Name.Number@osu.edu`) of the students enrolled in the class and the instructor. Because of this, you should read email sent to your official OSU account frequently.

Carmen and privacy: Be aware that the Carmen system as it is set up at OSU keeps detailed logs of your interaction with the system, for example: when you log in, how long you take to complete which question of the quiz, etc.

Feedback: If you have comments, complaints, or ideas you'd like to send me, please feel free to contact me via email or in person. If you prefer to remain anonymous, you can contact me anonymously from my web page (http://www.ling.osu.edu/∼scott/contact/) or leave a note for me in my mailbox in the linguistics department (225 Oxley hall).

Readings: There is no textbook for this course (the topic is very new, at OSU and elsewhere). There will be some readings assigned periodically throughout the course.

I will distribute slides in class for each unit. These will also be available on the web after the class in which they are first distributed. **These slides are only a skeleton of the material covered; they cannot replace actually being in class.**

Course requirements: The basic requirement is **regular attendance in class** and **active participation**. There will be roughly one online **quiz** per topic, to ensure the material covered in class is mastered. And there will be one **homework** (exercise sheet) per topic, which is intended to give the opportunity to explore new aspects of the topics discussed in class. The **midterm** will consist of the material covered in the first half of the class, and the **final** will cover the contents covered in the second half of the class. Both the midterm and final will be held in class and will be closed book. **The midterm will be given Wednesday, May 7th 2008 during the regular class period and the final will be held Thursday, June 5th 2008 11:30 AM–1:18 PM in 50** Jennings Hall.

Grading: Grades will be based on participation in classroom discussion and group work, quizzes, homeworks, a midterm exam, and the final examination, using the following scheme:

Participation	10%
Quizzes	20%
Homeworks	30%
Midterm	20%
Final	20%

- Given that homeworks, quizzes, and the exams address the material covered in class, **attendance is essential for doing well in this class.**

- Homeworks are administered on-line through Carmen, using its dropbox feature. Please submit homeworks in plain text, pdf, or html format, which you can either write as such or save in that format from the editor you use (for example, the "Save as Web Page" option of Word produces html output). **No formats other than plain text, pdf, or html will be accepted.** If you are unable to produce documents in one of these formats, see me outside of class.

  You generally will have a week for each homework and a clear deadline will be given for each homework. **No late homeworks will be accepted.**

  I encourage group work on the homework assignments, but each of you has to write and submit your own answers, i.e., you have to use your own words – copying part or all of the answer of another student obviously is unacceptable. Note that group work means that everyone in the group contributes and fully understands what you turn in—where in doubt, I may ask you individually to explain to me in person what you turned in.

- Quizzes are administered on-line through Carmen (http://carmen.osu.edu/). Each of them takes about 5–20 minutes, but you'll have two hours to complete the entire quiz. After two hours, or after the deadline for doing the quiz has elapsed, it's automatically switched off. You will generally have a week to complete a quiz, so **do not put it off until the last minute!** The quizzes naturally are open book, so you should view them as an opportunity for reviewing the material covered, based on the handouts.

  Note: I will not remind you when you have a quiz due. It is your responsibility to keep up with the syllabus/calendar of the course.

- If you feel that I have graded anything incorrectly or improperly, please contact me outside of class. I will be happy to discuss your concerns.

- The OSU Standard Grading Scheme is used, which maps percentages to grades in the following way:

		B+	87–89	C+	77–79	D+	67–69	E	0–59
A	93–100	B	83–86	C	73–76	D	60–66		
A−	90–92	B−	80–82	C−	70–72				

Make-up Policy: If you know you won't be able to make a deadline or exam, please see me **before** you miss the deadline or exam. If you miss the midterm or final, you will have to provide extensive written documentation for your excuse.

As you generally will have a week to take them, there are no make-ups for the quizzes.

Academic Misconduct: To state the obvious, academic dishonesty is not allowed. Any suspected academic misconduct or violation of the code of student conduct **must**, by university rules, be reported to to the University Committee on Academic Misconduct. The most common form of misconduct is plagiarism. Remember that any time you use the ideas or the materials of another person, you must acknowledge that you have done so in a citation. This includes material that you have found on the Web or given to you by another student by email, telephone, in person, etc. The University provides guidelines for research on the Web at http://gateway.lib.ohio-state.edu/tutor/ and you can find the Student Code of Conduct at http://studentaffairs.osu.edu/resource_csc.asp.

Class etiquette: I expect you to respect one another, to respect me, and to respect yourself. To that end, I expect you to obey the following rules:

- Participate: share experiences, ask questions, express your opinions. Ask me to provide more information, send me emails or see me during office hours for help, clarification, or recommendations for further research.

- Do not read newspapers, materials from other classes, etc. in class. When in the computer lab, only use the computers when I ask you to do a specific activity – do not read email or browse the web. Do not pack up early. Switch off your cell phone. If for some reason, you must leave early or you have an important call coming in, notify me before class.

Topics:

1. Storing language on the computer: Text and speech encoding

Writing systems used for language. Representing text on the computer. Digital representations of speech.

2. Searching: web, library catalogs, and other language-based databases

    What facilities exist for searching for language-based information? Different query languages and what they allow you to do. Differences between specific and general queries. How to evaluate the results of a search.

3. Classifying documents: Language identification and spam filtering

    Techniques for classifying documents. What language(s) are they written in? Are they junk mail? Are statistical techniques better than rule-based ones, or not? When will the techniques fail?

4. Writer's aids: Spelling and grammar correction

    What do so-called "grammar checkers" and "spelling correctors" do? What do such programs base their advice on? When does it make sense to use such tools and what kind of errors are to be expected?

5. Machine translation

    What do the free internet-based translation services manage to do—and where do they fail? For what purposes can automatic machine translation work reliably? What translation support functions can a computer provide? A closer look at what makes machine translation such a hard task. Is it the grammar, the meaning, the culture, all three, or something else?

6. Dialog systems

    Eliza and its surprising success in engaging people in conversation. When are dialog systems used, for what purpose? A closer look at the components of a dialog system. Where is what kind of knowledge needed to make it work?

7. Computer-Aided Language Learning

    What is involved in learning a foreign language? What role in language learning can computers play: from vocabulary training, via presentation of learning material, to providing feedback on learner errors and progress.

8. Social context of language technology use

    How do we react to computers that make use of language? What does it mean for the way we see ourselves? What assumptions do we make about every user of language, be it a human or a machine.

Students with Disabilities: Students who need an accommodation based on the impact of a disability should contact me as soon as possible to discuss the course format, to anticipate needs, and to explore potential accommodations. I rely on the Office of Disability Services for assistance in verifying the need for accommodations and developing accommodation strategies. Students who have not previously contacted the Office for Disability Services are encouraged to do so (292-3307; http://www.ods.ohio-state.edu).