

## **LIN313, Language and Computers**

Fall 2010

### Instructor Contact Information

Jason Baldrige

office hours: Mon 1-3, Tue 10-11

office: Calhoun 510

phone: 232-7682

fax: 471-4340

email: [jbaldrid@mail.utexas.edu](mailto:jbaldrid@mail.utexas.edu)

### TA Contact Information

Evelyn Richter

office hours: Tue 1:30-3:30, Wed 3:15-4:15

office: TBA

email: [evelyn.richter@gmail.com](mailto:evelyn.richter@gmail.com)

### Prerequisites

None.

### **Text**

There is no official course text book for this course since the topic is quite new. Some readings will be assigned and made available for download or copying.

### **Exams and Assignments**

There will be one mid-term exam and one final exam. The midterm will consist of the material covered in the first half of the class, and the final will be comprehensive, but with a greater emphasis on the contents covered in the second half of the class.

Assignments will be updated on the assignments page. A tentative schedule for the entire semester is posted on the schedule page. Readings and exercises may change up one week in advance of their due dates.

To see your grades go to eGradebook.

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

### Philosophy and Goal

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.

This course uses 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.

Evaluation will be based on the exams, homeworks, and class participation.

This course is based on the Language and Computers course taught in the Linguistics Department of the Ohio State University and which satisfies OSU's GEC category 2B (Mathematical and Logical Analysis) requirement. It will cover much of the same content, plus additional topics.

Content Overview

Topics include:

- Storing language on the computer: Text and speech encoding. Writing systems used for language. Representing text on the computer. Digital representations of speech.
- Classifying text: Is a piece of text about sports, politics, finance, etc? Does a sentence indicate positive or negative sentiment by the speaker/writer toward the thing being discussed? Are statistical techniques better than rule-based ones, or not? When will the techniques fail?
- Question answering: automatic analysis of questions and the search for answers to them Difference between search and question answering. Techniques for analyzing what is being asked. Obtaining answers.
- 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?
- 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?
- Cryptography: scrambling natural language Ciphers, including substitution ciphers (ROT13) and public-private key cryptography. Breaking ciphers. Enigma. The Voynich Manuscript and analysis of mysterious texts. Gsrkvexypexmsrw jsv higtmlivmrk xlmw. Xeoi xli gsyvwi!

- Forensic linguistics: Can computers help spot patterns that can identify who is the actual author of a text or speech segment? How does this play out in court? What kind of evidence is admissible?
- 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?
- 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.
- World knowledge: Does a cup fall down when you let go of it? Attempts to encode common sense and world knowledge in computers. Manual and automatic approaches.

There will be six assessed assignments, one essay, and two exams.  
Course Requirements

### **Assignments (45%):**

A series of six assessed assignments will be assigned during the semester. The lowest grade will be dropped, so each homework that counts is worth 9%.

### **Essay (15%):**

A 1000-1500 word essay on a topic dealing with the social implications of computational applications for language.

### **Mid-term Exam (20%):**

There will be a mid-term exam on October 15 over the material covered in class up to October 6.

### **Final Exam (20%):**

The final exam will be given during finals week and will cover all course material.

The course will use plus-minus grading, using standard scales. Attendance is not required, and it is not used as part of determining the grade.

### **Extension Policy**

Homework must be turned in on the due date in order to receive credit. Extensions will be considered on a case-by-case basis and only if the student asks for the extension before the deadline. In most cases they will not be granted.

Points will be deducted for lateness (unless an extension has been granted). By default, 10 points (out of 100) will be deducted for lateness, plus an additional 5 points for every 24-hour period beyond 2 that the assignment is late. For example, an assignment due at 11am on Tuesday will have 10 points deducted if it is turned in late but before 11am on Thursday. It will have 15 points deducted if it is turned in by 11am Friday, etc.

Late submissions will not be accepted if they are more than one week past the deadline. No points will be received in this case.

The greater the advance notice of a need for an extension, the greater the likelihood of leniency.

#### Academic Dishonesty Policy

You are encouraged to discuss assignments with classmates. But all written work must be your own. If in doubt, ask the instructor.

Students who violate University rules on academic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since such dishonesty harms the individual, all students, and the integrity of the University, policies on academic dishonesty will be strictly enforced. For further information please visit the Student Judicial Services Web site:

<http://deanofstudents.utexas.edu/sjs>.

Notice about students with disabilities

The University of Texas at Austin provides appropriate accommodations for qualified students with disabilities. To determine if you qualify, please contact the Dean of Students at 512-471-6529 or UT Services for Students with Disabilities. If they certify your needs, we will work with you to make appropriate arrangements.

UT SSD Website: <http://www.utexas.edu/diversity/ddce/ssd>

Notice about missed work due to religious holy days

A student who misses an examination, work assignment, or other project due to the observance of a religious holy day will be given an opportunity to complete the work missed within a reasonable time after the absence, provided that he or she has properly notified the instructor. It is the policy of the University of Texas at Austin that the student must notify the instructor at least fourteen days prior to the classes scheduled on dates he or she will be absent to observe a religious holy day. For religious holy days that fall within the first two weeks of the semester, the notice should be given on the first day of the semester. The student will not be penalized for these excused absences, but the instructor may appropriately respond if the student fails to complete satisfactorily the missed assignment or examination within a reasonable time after the excused absence.

**This syllabus can also be found at**

**<https://sites.google.com/site/languageandcomputersfall2010/syllabus>**