

Linguistics 445/515

The Computer and Natural Language

Autumn 2008

Course goals Present-day computer systems work with human language in many different forms, whether as stored data in the form of text, typed queries to a database or search engine, or speech commands in a voice-driven computer system. We also increasingly expect computers to produce human language, such as user-friendly error messages and synthesized speech. Through selected readings, exercises, demonstrations and Python programming, this course will: a) survey a range of issues relating natural language to computers, covering real-world applications, b) provide practical experience about representation and use of natural language on computers, and c) illustrate key principles of natural language processing through programming. Emphasis will be put on basic natural language processing strategies and technologies using linguistic theory

Topics include text encoding, search technology, tools for writing support, machine translation, dialogue systems, computer-aided language learning, and the social context of language technology.

Instructor: Markus Dickinson

Office: Memorial Hall (MM) 317

Phone: 856-2535

E-mail: md7@indiana.edu

Office hours: (at least for the first week)

T 11:00am

R 1:00pm

or by appointment

Meeting time: MW 9:30-10:45am

Classroom: Lindley Hall (LH) 030

Course website: <http://jones.ling.indiana.edu/~mdickinson/08/445/>

Assignments, slides, etc. will be posted here.

Credits: 3

Course prerequisites: None. That means that no prior programming experience is expected.

Course requirements: There will be various reading selections throughout the quarter, but most of the material will be introduced solely in the classroom. There will be approximately one exercise sheet, or homework, every two weeks. These assignments give you the opportunity to explore new aspects of the topics discussed in class, as well as to ensure that you are comprehending the material covered in class. These assignments will occasionally also give you the opportunity to practice your programming skills. Additionally, there will be in-class exercises which are included in your participation grade.

Readings: There is no textbook for this course, but there will be readings assigned periodically throughout the course.

For each unit, slides will be available from the webpage before class. These slides are meant to aid classroom discussion and cannot replace actually being in class.

Grading: Grades will be based on classroom discussion/participation, homeworks, a midterm exam, and a final examination. For 515, there will be an additional final project.

L445 grading:

PARTICIPATION	8%	
HOMEWORKS	42%	(7@6% each)
MIDTERM	25%	Wednesday, October 20 @ 9:30am
FINAL	25%	Monday, December 15 @ 8:00am

L515 grading:

PARTICIPATION	5%	
HOMEWORKS	35%	(7@5% each)
MIDTERM	20%	Wednesday, October 20 @ 9:30am
FINAL	20%	Monday, December 15 @ 8:00am
FINAL PROJECT	20%	Wednesday, December 17, by 5:00pm

- Homework assignments are due by the beginning of each class (9:30am)—you may hand them in or e-mail them to me. You can work together on the homework assignments but write out your own answers. Your homework grade will be based on both quality and effort.

– Late Homework Policy: You will lose one (1) letter grade (10 percentage points) for every 24 hour period after the turn-in time.

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

Grading scale: (Scores in percentages)

A+	99-100	B+	87-89	C+	77-79	D+	67-69	F	0-59
A	93-98	B	83-86	C	73-76	D	63-66		
A-	90-92	B-	80-82	C-	70-72	D-	60-62		

Make-up Policy: If you plan on missing either the midterm or final, you will have to provide extensive documentation for your excuse. See me immediately if this is the case.

Final Project (L515): For those enrolled at the 515 level, there is a final project requirement, the topics of which will be discussed individually with the instructor (beginning in October). The projects will generally be papers extending discussion of specific topics touched on in class, although they may also be implementations of a specific natural language processing algorithm (documented and evaluated) or evaluation of existing algorithms and software systems. The projects will be due on Wednesday, December 17 at 5:00pm.

Programming: To assist you in learning how to think logically & algorithmically, you are going to be taught some fundamentals of programming, using the Python programming language. We will include this in various class sessions (not always listed on the syllabus). I expect that most of you have absolutely no experience in programming and might be a little (or a lot) scared of it, and so I want to be clear about a few points:

- The programming material is not the main course material! It is meant as an aid to help think logically and for you to come to a better understanding about all the details that go into a program.
- You will not learn to be a programmer in this class. In other words, this is not a programming class, only a class where we occasionally use programming.

Academic Misconduct: Academic misconduct is not allowed in this course. The Indiana University *Code of Student Rights, Responsibilities, and Conduct* (<http://dsa.indiana.edu/Code/>) defines academic misconduct as “any activity that tends to undermine the academic integrity of the institution . . . Academic misconduct may involve human, hard-copy, or electronic resources . . . Academic misconduct includes, but is not limited to . . . cheating, fabrication, plagiarism, interference, violation of course rules, and facilitating academic misconduct” (II. G.1-6).

Students with Disabilities: Students who need an accommodation based on the impact of a disability should contact me to arrange an appointment as soon as possible to discuss the course format, to anticipate needs, and to explore potential accommodations.

I rely on Disability Services for Students for assistance in verifying the need for accommodations and developing accommodation strategies. Students who have not previously contacted Disability Services are encouraged to do so (812-855-7578; <http://www.indiana.edu/~iubdss/>).

Class etiquette: I expect you to respect one another, to respect me, and to respect yourself.

- Participate: share experiences, ask questions, express your opinions. Ask me to provide more information and see me during office hours for help or clarification or recommendations for further research.
- Do not sleep, pack up early, take phone calls, etc.
- If you ever have a problem with the course or with me, I prefer you see me during office hours (or schedule an appointment).

Computational Linguistics: If you find yourself loving this material, I encourage you to come see me or Professor Sandra Kübler for more information about computational linguistics.

Schedule: Links to notes and homeworks will be posted on the course website.

Month	Date	Topic	Assignments
Sep.	3	Intro to class	
	8	Text & speech encoding: text	
	10	Text & speech encoding: speech	
	15	Programming basics	HW1 due
	17	Searching	
	22	Searching: internals	
	24	Searching: regular expressions	
	29	Corpus annotation	HW2 due
Oct.	1	Text classification (TC)	
	6	TC: Spam filtering	
	8	TC: Spam filtering	
	13	Spelling & grammar correctors	HW3 due
	15	Spelling correctors	
	20	MIDTERM	MIDTERM
	22	Spelling correctors for the web	
	27	Grammar correctors: parsing	HW4 due
	29	Grammar correctors: n-grams	
Nov.	3	N-grams in Python	
	5	Machine Translation (MT)	
	10	Symbolic MT	HW5 due
	12	Statistical MT	
	17	Dialogue systems: dialogue	
	19	Dialogue systems: chatterbots	
	24	Dialogue systems: modern systems	HW 6 due
	26	NO CLASS, THANKSGIVING BREAK	
Dec.	1	Computer-aided language learning (CALL)	
	3	CALL: authentic text CALL	
Dec.	8	CALL: parser-based CALL	HW7 due
	10	Social context of language technology use	
	15	FINAL EXAM, 8:00-10:00am	FINAL
	17	FINAL PROJECT (L515)	due by 5pm

Disclaimer This syllabus is subject to change. All important changes will be made in writing, with ample time for adjustment. (Midterm and final dates, however, will not change.)