

Professor: Dr. Dan Parker
Email: dparker@wm.edu
Office: Tucker Hall 036

Class location: Tucker 220
Class time: 9:00 – 9:50 AM MWF
Office Hours: 1 – 3:00 PM MW / by appt.

Course description

If a computer is smart enough to beat humans at Jeopardy, drive a car, and pen a haiku, does it mean that machines can think? Is it possible to predict the spread of the flu, or the success of Hollywood's next film based on patterns in Google searches? Did Shakespeare really write that sonnet?

Scientists use patterns in language to answer these questions, using the same concepts that underlie everyday applications like search engines, automatic translators, speech recognition systems, spell-checkers, and auto-correction tools. This course takes you on a tour of these applications, focusing on the technological ideas behind them, through a combination of seminar-style discussions and practical hands-on projects.

Since this course is a linguistics course, we take seriously the goals of linguistic inquiry – to uncover the structures of human language and the mental mechanisms for language understanding. In this course, we will achieve this goal in a novel way, by taking a reflective, analytic, and hands-on approach to study how computers use and support language. Over the course of the semester, we will see that by looking at how different technologies succeed and fail at human language understanding, and by examining the ways in which their behavior diverges from that of a human, we can learn gain new insights about the human linguistic system.

This course aims to foster:

- A solid theoretical and practical understanding of how computers support language, how they succeed and fail at understanding human language, and what those successes and failures reveal about the human linguistic system.
- Hands-on experience with various language technologies
- The ability to use computers to supplement linguistic research
- The ability to read, understand, and critique primary research papers in the field
- The ability to effectively communicate the goals and methods of the field

These goals will be achieved through readings, lectures, in-class discussions and demonstrations, written homework assignments, and lab projects. The course is designed to get you to think critically about how technology and language works, focusing on the connections between linguistics, computer science, and mathematics. You will encounter essential concepts relating to language, representation, and processing. However, rather than just describing the important concepts, we will gain hands-on experience with different language technologies, including the Natural Language Toolkit (NLTK) in Python, to illustrate them as well.

Prerequisites:

- LING 220 Study of Language
- Curiosity about language and everyday experience with computers
- No programming experience required

Readings

There is no textbook for this course. But the course will be reading-intensive. Readings will be posted to Blackboard. Some of the readings will seem dense depending on your background. My job is to facilitate your understanding of the articles and to help you make the connections between them. But it is ultimately your responsibility to read the material. You are responsible for your own learning. Get started early, stay on top of the readings, and discuss the articles with me and your peers.

Evaluation

Reading responses	15%	
Presentations	15%	
Assignments	45%	
Final project	25%	(no midterm or final)

Reading responses: I'm not interested in quizzing you about the readings. I want you to read, think, critique, and ask questions. To achieve that, for each reading, you will respond to three questions: (1) What are two things that you learned from this reading? (2) What challenges does the reading face, e.g., what flaws can you find in the logic, design, execution, analysis, conclusions? and (3) What are two things that you did not understand in the reading? Responses must be posted to blackboard by 9 PM the night before class.

Presentations: Throughout the semester, you will give one in-class presentation of a reading, with a partner. The purpose of these presentations is for you to develop your skills in communicating different types of research to an audience in a clear, concise, and accessible fashion.

Assignments: Five assignments/labs will be assigned. Assignments are due at the start of class. These assignments are designed to help you gain a good grasp and hands-on experience with the concepts discussed in class. Assignments are to be completed individually, but discussion with your classmates is encouraged.

Final project: You will design your own final project and present your project to the class. Projects must involve some computational component, and can involve either a critical evaluation of some computational concept or tool, or the implementation of a computational concept or tool. Projects may be completed in teams of no more than 2 students.

Grade scale: I do not grade on a curve (you're not competing against your peers), or offer extra credit.

A	94 – 100	C+	77 – 79	D-	60 - 62
A-	90 – 92	C	73 - 76	F	0 – 59
B+	87 – 89	C-	70 – 72		
B	83 – 86	D+	67 - 69		
B-	80 – 82	D	63 – 66		

Requirements, policies, & guidelines

This is a seminar-style course that will focus heavily on readings, discussion, writing, and doing. If you want to get the maximum benefit from this class, i.e., learn a lot, have fun, and have a grade to show for it, here are my suggestions:

How to succeed in this course ...

- Read the readings. An important skill to develop is the ability to efficiently extract ideas and information from writing. This won't involve skimming, rather targeted reading.
- Attend class. Come prepared. Participate. Engage. Participating makes you an active learner and greatly increases the likelihood that you'll understand and retain the material. Share ideas, including points of confusion. Speak up!
- Avoid disruptive behavior. Hey you, in the back. It's clear when you're on Facebook.
- Come see me. No appointment needed during office hours. Stop in if you would like to talk about anything related to class or further interests. Let me know what you're thinking, what's coming easy, what you're struggling with. If you find yourself struggling seek my help sooner than later!

A note about programming assignments: This class will involve learning new skills on a computer. No programming experience is required. It's my job to make learning as easy and painless as possible. Please note that mastering these skills involves trial and error ... really just a lot of error. Based on my own experience, very rarely does my code work correctly on the 1st, 2nd, 3rd, ... 10th trial. Assume that your code won't work the first time, and you'll never be disappointed. Usually, the problem is just a typo. Sometimes is a larger issue. These issues are to be expected. Do not get frustrated with yourself, the computer, or me. It's part of the learning experience. Get started early on assignments. Utilize office hours. Not getting things to work the night before the assignment is due is not an excuse for not doing the assignment. I'm here to help and I'm friendly!

Lateness policy: Please stop by office hours or email me questions about assignments in sufficient time! Not understanding the assignment or not being able to access the assignment (without seeking my help in advance) is not an acceptable excuse. Reading responses will not be accepted, at all, after the 9 PM deadline. Assignments turned in late will receive a 10% penalty unless due to a *documented* issue, and will not be accepted more than 2 days after the due date. I do understand, though, that there are extenuating circumstances, and that we all have other responsibilities competing for our time. Therefore, I will allow everyone up to two missed reading responses.

Communicate. If you have any trouble with the course, if you have a personal emergency where you need to miss class (especially more than one class), or if you are experiencing an illness or any other situation which is interfering with your ability to participate fully in the course, make sure to contact me in a timely manner. There can be no adjustments to grades or accommodations made that will affect your grade after the semester is completed. Keep me informed well in advance if possible.

Accommodations: Any student who needs accommodation should consult with Student Disability Services and the professor as early as possible. All discussions will remain confidential.

Honor code: I will follow the College's policies on academic honesty and penalize dishonesty of any kind. Please review the code, which states: "As a member of the William and Mary Community, I pledge on my honor not to lie, cheat, or steal, either in my academic or personal life. I understand that such acts violate the Honor Code and undermine the community of trust, of which we are all stewards."

My responsibilities: I also have a number of responsibilities, the most important of which is encouraging interest in, stimulating discussion about, and offering clear explanation of course topics (and beyond if you're interested). I too must come prepared for class, return assignments in a timely manner (no later than one week), and provide reasonable personal assistance outside of class.

Introduction

W 01/20	Syllabus and course logistics
F 01/22	Setting the stage ...

Language and computers

M 01/25	SNOW DAY		
W 01/27	What does a computer need to know?	Marr_1982; Chomsky_1965	
F 01/29	Representing language	MITTechReview	

Algorithms and Language as a Process

M 02/01	Encoding language on computers		
W 02/03	What are algorithms?	AutomateThis	
F 02/05	Language as an algorithm?		HW 1 assigned

Acquiring language: The successes and failures of statistical approaches

M 02/08	Statistical language learning	LewisElman_2001	
W 02/10	Against statistical learning	FrankEtAl_2013	
F 02/12	Lab		HW 1 due / HW 2 assigned

The Big Data revolution

M 02/15	What is Big Data?		
W 02/17	Linguistic corpora		
F 02/19	Building a corpus	CorporaCh1-2	HW 2 due / HW 3 assigned
M 02/22	Applications of corpus linguistics	CorporaLaw	
W 02/24	Lab		
F 02/26	Lab		HW 3 due

Putting Big Data to work: classifying text and sentiment analysis

M 02/29	Automatic text classification		
W 03/02	No class: Prof Parker @ CUNY conference		
F 03/04	No class: Prof Parker @ CUNY conference		
M 03/07	No class: Spring Break		
W 03/09	No class: Spring Break		
F 03/11	No class: Spring Break		
M 03/14	Lab		
W 03/16	Lab		
F 03/18	Lab		
M 03/21	Lab		
W 03/23	Lab		
F 03/25	Lab		HW 4 assigned
M 03/28	Lab		Project abstracts due
W 03/30	Applications of text classification	NYT-SentimentAnalysis1	
F 04/01	Applications of text classification	NYT-SentimentAnalysis2	HW 4 due

Parsing text

M 04/04	Parsing text	Abney-Parsing	
W 04/06	Lab		
F 04/08	Lab		HW 4 due / HW 5 assigned

Machine Translation & Automatic Dialogue

M 04/11	The art and math of translation	NYT-Translation	
W 04/13	Automatic Dialogue	JurafskyMartin-Dialogue	
F 04/15	The rise of Chat Bots	ChatBots	HW 5 due

Tying it all together

M 04/18	Singing the praises of Big Data	BigData	
W 04/20	The pitfalls of Big Data	Hornstein; Marcus	
F 04/22	Final remarks		

Final Presentations

M 04/25	Final Project Presentations	
W 04/27	Final Project Presentations	
F 04/29	Final Project Presentations	