



CS 521: Statistical Natural Language Processing

Natalie Parde, Ph.D.

Department of Computer Science
University of Illinois at Chicago

Spring 2020



Who is teaching this class?

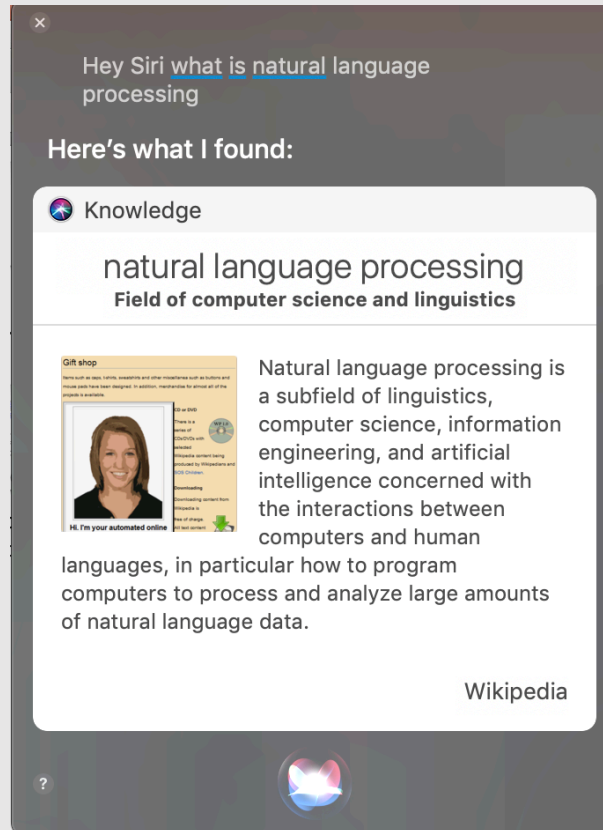
- **Assistant Professor**
- **Co-Director**, Natural Language Processing Laboratory: nlp.lab.uic.edu
- Research Interests:
 - Interactive systems
 - Multimodality
 - Creative language
 - Grounded language learning
- Joined UIC ~15 months ago



Tell me about yourself!

- M.S. or Ph.D.?
- What made you interested in this class?
- What are you hoping to get out of it?

What is natural language processing?



- Subfield of **artificial intelligence**
- Focus: Automatically **interpreting** and **generating** natural **language**
- Most popular examples: Alexa, Siri, Google Assistant, Cortana....

**Early NLP
was rule-
based.**

Rule-based NLP

- Manually create rules to interpret and generate language

Why rule-based?

- Much less computing power
- Much less data

Recently, most areas of NLP have shifted towards using statistical techniques.

- Statistical NLP
 - Automatically learn rules from labeled and/or unlabeled datasets to interpret and generate language
- Why move to statistical methods?
 - Better able to generalize
 - Often better overall performance
- Statistical methods:
 - **Classical machine learning** (usually for smaller datasets)
 - **Deep learning** (usually for larger datasets)

Topics We'll Cover This Semester

Language modeling

- N-gram
- Neural

Data collection

Logistic regression

Vector semantics

- Word2Vec
- GloVe
- Contextual word embeddings

Feedforward neural networks

Convolutional neural networks

Recurrent neural networks

Encoder-decoder models

Attention

Coreference resolution

Discourse coherence

Natural language generation

Multimodal NLP

Figurative language

Healthcare applications

Syllabus

Contact Info

Professor: Natalie Parde

Email: parde@uic.edu

Office: SEO 1132

Office Hours: Tuesday 1:30 - 2:30
p.m. / Thursday 3:00 - 4:00 p.m.

Course Communications

- **Piazza:** <https://piazza.com/uic/spring2020/cs521>
 - Also linked directly from Blackboard
- **Blackboard:** CS 521 Stat Natural Lang Processing (43232) 2020 Spring
- **Gradescope:** <https://www.gradescope.com/>
- **Course Website:** http://www.natalieparde.com/teaching/cs521_spring2020.html
- Try to keep email to a minimum! Instead, post questions on **Piazza**.
- All assignments should be submitted on **Blackboard**.
- Exam grades will be posted on both **Blackboard** and **Gradescope**.

Prerequisite

- This course assumes proficiency with the topics covered in **CS 421**
- Haven't taken CS 421?
 - Check out the course material to determine whether you are adequately prepared:
http://www.natalieparde.com/teaching/cs421_fall2019.html

What will we be reading?

- First part of the semester:
 - Daniel Jurafsky and James H. Martin. **Speech and Language Processing** (3rd Edition). Draft, 2019:
<https://web.stanford.edu/~jurafsky/slp3/>
 - Still being written! The draft can be freely accessed at the link above.
- Second part of the semester:
 - **Research papers**

What will be graded?



Paper Critiques

Eight total

Short (~ one page) critiques of papers being discussed



Project

Includes a proposal, presentation, implementation, and write-up



Exam

One total, after the first part of the semester

**How much
will each of
these items
contribute
to the final
grade?**

- Paper Critiques: 28%
 - $7 \times 4\%$ per paper critique
- Exam: 20%
- Paper Presentation: 10%
- Project: 42%
 - Proposal: 7%
 - Presentation: 10%
 - Implementation: 10%
 - Write-Up: 15%

Projects can be completed individually or in pairs!

- If completing the project in a pair:
 - Workload should scale accordingly
 - Include a work statement signed by both team members in the project submission, detailing which tasks were completed by each team member



Project Details



Proposal

Short (2-4 pages)
Defines research objectives
Outlines preliminary plans
(these can change as the project progresses!)



Presentation

Share your findings with the class at the end of the semester



Implementation

Source code, documentation,
results files, data



Write-Up

2500-4000 words
Conference/journal-style
paper



Paper Critiques

- Examples posted on Blackboard
- Summarize the paper, highlight strengths/weaknesses, analyze the methodology and evaluation, and explain whether conclusions are justified
- First two paper critiques:
 - Select **any paper remotely relevant** to any topic covered in the specified timeframe
- Next six paper critiques:
 - Select **one of the papers that will be discussed** in the upcoming week

Paper Presentation

Summarize and critically analyze a research paper

Select your paper anytime between now and 2/17 (first come, first served!)

- Choose a paper relevant to one of the six paper discussion topics
 - See syllabus for a list of suggested papers
 - Feel free to choose outside the syllabus as well
- Leave a comment on the “Discussion Paper Selections” Wiki page on Blackboard
- If approved, your selection will be added to the schedule
- Note: There can be a maximum of five papers selected per topic!



Exam

- Mix of question types:
 - True/false
 - Multiple choice
 - Free response
- Covers all topics in the first six weeks of class

Missed Class

Missed Presentation:

- If you discover that you'll miss your scheduled presentation, contact me as soon as possible. I'll either:
 - Reschedule it
 - Assign a video presentation due the same day
- If you don't contact me ahead of time, I'll deduct:
 - 25% as soon as class begins
 - An additional 50% at that time the following day
 - The remaining 25% at that time the day after that

Missed Exam:

- If you need to miss class on the exam date, contact me as soon as possible. I'll:
 - Reschedule it to take place under my supervision at a time earlier than the original exam date/time
- If you do not contact me early enough to reasonably reschedule the exam to an earlier time, no make-up will be given

Late Cards



- A late card allows you to turn in your assignment up to 24 hours after the deadline with no grade penalty
- 4 per semester
- It's fine to use multiple late cards for the same assignment (including the project)
- You cannot use late cards for presentations

- If you have no late cards remaining (or choose not to use them for a given assignment):
 - 25% deducted for each day the assignment is late, starting a minute after it is due

- Try to keep your late cards for unforeseen scenarios!
 - No extra late cards granted to individual students under any circumstances

Regrade Requests

- Submit regrade requests within **one week** of grades being published
- Submit request via email, including:
 - Which rubric item(s) were inconsistently graded
 - Why the grade is inconsistent with the rubric

Honor Code



You are expected to do your own work on all assignments!



If I discover that you cheated or plagiarized:

You will receive a 0 on the assignment
If a repeat offense, you will also be reported to CS Student Affairs



Examples of cheating:

Having someone complete all or part of your assignment for you
Copying someone's answers from their exam
Copying and pasting text from published material without proper attribution
Paraphrasing published material without proper attribution
Copying and pasting someone else's code without attribution

Schedule

Week	Topic	Deliverables
1/13 – 1/17	Introduction and Language Modeling	—
1/20 – 1/24	Data Collection and Logistic Regression	—
1/27 – 1/31	Vector Semantics, Word2Vec, and GloVe	1/31: Paper Critique (Topic from 1/13 – 1/31)
2/3 – 2/7	Feedforward Neural Networks and Convolutional Neural Networks	—
2/10 – 2/14	Recurrent Neural Networks, Encoder-Decoder Models, and Attention	2/14: Project Proposal
2/17 – 2/21	Coreference Resolution and Discourse Coherence	2/21: Paper Critique (Topic from 2/3 – 2/21)
2/24 – 2/28	Review and Exam (2/27)	—
3/2 – 3/6	Neural Language Modeling	3/2: Paper Critique
3/9 – 3/13	Contextual Word Embeddings	3/9: Paper Critique
3/16 – 3/20	Natural Language Generation	3/16: Paper Critique
3/23 – 3/27	Spring Break	—
3/30 – 4/3	Multimodal NLP	3/30: Paper Critique
4/6 – 4/10	Figurative Language	4/6: Paper Critique
4/13 – 4/17	Healthcare Applications	4/13: Paper Critique
4/20 – 4/24	Project Presentations	—
4/27 – 5/1	Project Presentations	5/1: Project Implementation and Write-Up
5/4 – 5/8	Finals Week	—

Classroom Environment

- Treat everyone with respect!
 - Silence all devices
 - Don't chat with one another during lectures/presentations unless explicitly asked to do so
 - Don't bring disruptive food
- Inform me and UIC's Disability Resource Center (<http://drc.uic.edu/>) about any disabilities for which you would like to request accommodation
- Let me know if you'd like me to update any of the information (e.g., name or pronouns) I've received about you from the class roster
- Feel free to reach out to me with feedback throughout the semester!





What's next?

- Check course website for this week's required reading
- Select (or start thinking about) the paper you'd like to present
- Start thinking of project ideas