Week 1

What is computational linguistics?
Course overview and policies,
Quest login, regular expressions
Who are we?

Rob Voigt
robvoigt@northwestern.edu
Assistant Professor of Linguistics

Wes Orth
WesleyOrth2022@u.northwestern.edu
PhD Student in Linguistics
Who is this class for?

- Linguists, social scientists, computer scientists
- Basic programming experience required!
- Applications:
  - research in linguistics or social science
  - research in machine learning / deep learning
  - industry jobs and data science
Who is this class for? - other relevant classes

- CS 337 - NLP
- CS 348 - AI
- CS 349 - Machine Learning
- Bryan Pardo’s Deep Learning
- LING 330 - Stat methods
- Whatever seminars I teach down the road
When and where will we see each other?

Zoom at normal class times
recorded if you can’t make it

Mondays          generally lecture and discussion all together

Wednesdays      mixed, with lecture, group work, extra OH
**When and where will we see each other?**

<table>
<thead>
<tr>
<th>Office hours</th>
<th>Wes</th>
<th>Thursdays 3-4pm and by appt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rob</td>
<td>Fridays 11am-noon and by appt</td>
<td></td>
</tr>
</tbody>
</table>

Ed discussion board for questions

Be involved, be interested, help each other out!

(quick demo)
What will we learn?

Course content:

● Basic problems and algorithms for CL/NLP
● Related programming and math (nothing too crazy)
● Overview of topic areas
● Foundations for contemporary / neural
What is computational linguistics?

The use of computational methodologies to understand language and language use at a large scale

The development of algorithms and statistical tools to allow computers to process human languages

Computational Linguistics (CL)  Natural Language Processing (NLP)
Engineering, Social Science, and beyond

Deeply interdisciplinary field - questions like:

How can we build a robot that can talk to us?  
(and what would it mean if it could?)

How do we organize and access all the information  
(written in human language) on the internet?

How can computational models represent linguistic ones?

How does language use at large scales show human biases?
Historically Intertwined ➔ Now Less Clear

Originated with Machine Translation in the 50s

Political/military purposes - translate Russian into English (and vice versa)

“We’ll have this solved in a decade.” ... a few times. Turned out to be a hard problem!

ALPAC Report, 1966 - heavily critical of the progress thus far
Historically Intertwined → Now Less Clear

MT introduced numerous layers of difficulty: lexical, syntactic, pragmatic, contextual

Early work connected with theoretical linguistics more directly:

What is the computational complexity of natural languages?

Computational semantics - defining and reasoning with formal representations of meaning

(these are still relevant concerns for many)
Historically Intertwined → Now Less Clear

As CL / NLP got more applied, the link loosened:

1990s - beginning of statistical revolution in NLP

(field as a whole became much more empirical)

2010s - beginning of neural revolution in NLP

(core characteristic is uninterpretable features)
Sidenote of Social Importance

CL/NLP is probably among the most gender-diverse CS subfields.

Many important women historical (and contemporary) figures!

Margaret Masterman
Early MT

Karen Spärck Jones
Early Information Retrieval, tf-idf
Unique Properties of Language (that make it hard)

Polysemy
7 always means 7;
‘Waldorf’ can be a hotel, or a school, or a salad, or a muppet

Ambiguity and Vagueness
We saw her duck ... pet? dodgeball?
The western part of North America ... Wyoming? Manitoba?
Unique Properties of Language (that make it hard)

**Sparsity**

Many ways to say “the same thing”

Q: Where is he?  
He went to the store  
Oh, Johnny left to get groceries  
Out to grab the essentials

**Nested / Recursive / Infinite**

Stacey ate the candy that Naveen had found next to the adorable cat statue on the table that had been left out from last week’s party where Bill had unexpectedly proclaimed his love for Maurice and ...
Unique Properties of Language (that make it hard)

These properties (among others) are what make language useful for humans and interesting to study!
So in This Class? - Basic Algorithms

Algorithm?

Wikipedia - “finite sequence of well-defined, computer-implementable instructions, typically to solve a class of problems or to perform a computation”

Methods for solving problems!

Tend to address some of these common challenges

The building blocks of the CL/NLP “thought process”
So in This Class? - Programming Skills

Translating conceptual understanding, pseudocode, and math into actual, working code that we can run

More on-your-own than LING300/CS110 etc
Still some scaffolding
So in This Class? - Some Math

Much of contemporary NLP relies on probability

If you have high school math, you can learn it

Probability primer by Sharon Goldwater on course webpage!
So in This Class? - Concepts and Applications

Why are we doing what we’re doing?
   What linguistic phenomena motivate us to do it this way?

Applying methods to real-world datasets!

Error analysis! Ethical concerns!
What will we **not** explicitly cover?
(but you can learn if you’re motivated!)

*Sequence models*

- HMMs, CRFs, seq2seq, etc
  - Very important, needs its own class and more math

*Neural models*

- We will cover foundations (vector semantics) and discuss generally towards the end
How will we learn it?

Course structure and policies:

- Schedule
- Assignments
- Grading and Evaluation
- Agreements
Syllabus and Schedule

Syllabus and policies on course website:

https://faculty.wcas.northwestern.edu/robvoigt/courses/2021_spring/ling334/
Learning Structure - Assignments

Generally out on Monday, due the following Sunday night
Largely programming, some qualitative

You can work on them anyway, anyhow
Your own machine, or on Quest, or on Quest Analytics, etc

But your assignments **must** run on Quest
Autograders provided - helpful for gauging your progress!
Learning Structure - In Class

A few group work projects in class
regexes tomorrow!

A few peer evaluations

Some additional OH, working on assignments

One group presentation towards the end
Learning Structure - Final Assignment

2.5 weeks at the end of the quarter with no other out-of-class work

Very open to possibilities!

(some listed on website) (double-dipping encouraged)

Will discuss again about halfway through - but talk with me or Wes at any point about ideas.
Grading and Evaluation

Heavy emphasis on qualitative feedback

Wes will be primary grader,
I’ll read your comments and be spot-checking

Letter grades at the end based on effortful completion,
Midterm and final self-evaluations

The point of this whole thing is for you to learn, period!
What constitutes strong performance?

**Effort** and **Engagement** with learning.

Performance relative to *you*, not absolute performance.

Challenge yourself.

We have a very broad range of backgrounds and skill levels!

You are smart, you are adults -

We provide a structure, but it’s ultimately on you.
What constitutes strong performance?

There is a lower bound:
  Do basic reading, complete basic assignment (make it work)

There is no upper bound:
  Each week will have extra “relevant readings”
  Each assignment will have a number of possible extensions
  You can start working early on your final assignment
  Plus whatever you can dream up
Agreements

I see this class as entering into a set of mutual agreements, on top of the basic agreements of the university (academic honesty etc)

We’re building a community of learners interested in this topic! (I’m a learner too.)

By registering, you agree to certain things - By being the instructor, I agree to certain things.
You agree to:

Invest substantial time and effort in this course this quarter
Hold yourself accountable for your own progress
Be honest in assignments, self-evaluations
Stay on top of your work, and ask for help when needed
Be open to constructive feedback
Challenge yourself
Communicate with me when any of the above falls through
I agree to:

Invest substantial time and effort in your process of learning
Prepare well for class, construct meaningful assignments
Make myself available to help
Be open to criticism and commentary
Provide structures for learning
Communicate with you when any of the above falls through
We’re on the same team, this is not an adversarial relationship!

No such thing as a dumb question here.