NATURAL LANGUAGE COMPUTING
https://docs.google.com/presentation/d/1UtP7nABc_DJXIi5wKwLAaje7xRu94ZHBV-c8pl0Cfrg/edit#slide=id.geeb5401fec_1_51

THE INSTRUCTIONAL TEAM
What is natural language computing?

Getting computers to understand everything we say and write.

In this class (and in the field generally), we are interested in learning the **statistics of language**.

Increasingly, computers give insight into how humans process language, or generate language themselves.
Today

• Basic definitions in **natural language processing (NLP)**.

• Applications
  • Text Classification
  • Translating between languages
  • Automatic speech recognition
  • Natural Languages Understanding
  • Information Retrieval
  • Interpretability & Large Language Models.

• Course logistics. (at the end)
What can natural language do?

The ultimate in human-computer interaction.

“translate Also Sprach Zarathustra”

“take a memo...”

“open the pod bay doors”

“how far until Jupiter?”

“Can you summarize 2001: A Space Odyssey?”

We’re making progress, but why are these things still hard to do?
A little deeper

• Language has hidden structures, e.g.,
  • How are sounds and text related?
    • e.g., why is this: not a ‘ghoti’ (enough, women, nation)?

• How are words combined to make sentences?
  • e.g., what makes ‘colourless green ideas sleep furiously’ correct in a way unlike ‘furiously sleep ideas green colourless’?

• How are words and phrases used to produce meaning?
  • e.g., if someone asks ‘do you know what time it is?’, why is it inappropriate to answer ‘yes’?

• We need to organize the way we think about language...
Categories of linguistic knowledge

• **Phonology**: the study of patterns of speech sounds.
  e.g., “read” → /r iy d/

• **Morphology**: how words can be changed by inflection or derivation.
  e.g., “read”, “reads”, “reader”, “reading”, ...

• **Syntax**: the ordering and structure between words and phrases (i.e., grammar).
  e.g., NounPhrase → article adjective noun

• **Semantics**: the study of how meaning is created by words and phrases.
  e.g., “book” → 

• **Pragmatics**: the study of meaning in contexts.
  e.g., explanation span, refutation span
Ambiguity – Phonological

• **Phonology**: the study of patterns of speech sounds.

  “read” → /r iy d/ as in ‘I like to read’
  “read” → /r eh d/ as in ‘She read a book’

  “object” → /aa¹ b jh eh⁰ k t / as in ‘That is an object’
  “object” → /ah⁰ b jh eh¹ k t / as in ‘I object!’

  “too” ← /t uw/ as in ‘too much’
  “two” ← /t uw/ as in ‘two beers’

• Ambiguities can often be **resolved** in context, but not always.
  • e.g., /h aw t uw r eh¹ k ah ?? n ay² z s (b|p) iy ch/ → ‘how to recognize speech’
    → ‘how to wreck a nice beach’
Resolution with syntax

• If you hear the sequence of speech sounds

   /b ah f ae l ow b ah f ae l ow b ah f ae l ow b ah f ae l ow ...
   b ah f ae l ow b ah f ae l ow b ah f ae l ow b ah f ae l ow /

   which word sequence is being spoken?

   → “Buff a low buff a lobe a fellow Buff a low buff a lobe a fellow...”
   → “Buffalo buff aloe buff aloe buff aloe buff aloe buff aloe ...”
   → “Buff aloe buff all owe Buffalo buffalo buff a lobe ...”
   → “Buff aloe buff all owe Buffalo buff aloe buff a lobe ...”
   → “Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo”

• It’s obvious (to us) that the last option is most likely because we have knowledge of syntax, i.e., grammar.
Ambiguity – Syntactic

• **Syntax**: the ordering and structure between words. Words can be grouped into ‘parse tree’ structures given grammatical ‘rules’.

  e.g., “I shot an elephant in my pyjamas”
Resolution with semantics

- It’s obvious (to us) that the elephants don’t wear pyjamas, and we can discount one option because of our knowledge of semantics, i.e., meaning.
Ambiguity – Semantic

- **Semantics**: the study of how meaning is created by the use of words and phrases.

  - “Every man loves a woman”
    \[ \forall x \, \text{man}(x) \land \exists y : (\text{woman}(y) \land \text{loves}(x, y)) \]
    \[ \exists y : \text{woman}(y) \land \forall x (\text{man}(x) \lor \text{loves}(x, y)) \]

  - “I made her duck”
    \[ \text{I cooked waterfowl meat for her to eat.} \]
    \[ \text{I cooked waterfowl that belonged to her.} \]
    \[ \text{I carved the wooden duck that she owns.} \]
    \[ \text{I caused her to quickly lower her head.} \]

  - “Give me the pot”
    \[ \text{It’s time to bake.} \]
    \[ \text{It’s time to get baked.} \]
Resolution with pragmatics

• It’s obvious (to us) which meaning is intended given knowledge of the context of the conversation or the world in which it takes place.

  • “Every man loves a woman”
    \[ \forall x \text{ man}(x) \exists y:\ (\text{woman}(y) \land \text{loves}(x, y)) \]
    \[ \exists y. \text{woman}(y) \land \forall x (\text{man}(x) \rightarrow \text{loves}(x, y)) \]

  • “I made her duck”
    \[ \rightarrow \text{I cooked waterfowl meat for her to eat.} \]
    \[ \rightarrow \text{I cooked waterfowl that belonged to her.} \]
    \[ \rightarrow \text{I carved the wooden duck that she owns.} \]
    \[ \rightarrow \text{I caused her to quickly lower her head.} \]

  • “Give me the pot”
    \[ \rightarrow \text{It’s time to bake.} \]
    \[ \rightarrow \text{It’s time to get baked.} \]
Ambiguity – miscellaneous

• Newspaper headlines (spurious or otherwise)

  Kicking Baby Considered to be Healthy
  ...
  Squad Helps Dog Bite Victim
  ...
  Canadian Pushes Bottle Up Germans
  ...
  Milk Drinkers are Turning to Powder
  ...
  Grandmother of Eight Makes Hole in One
  ...
  Kids Make Nutritious Snacks
  ...
  Juvenile Court Tries Shooting Defendant
  ...
  Local High School Dropouts Cut in Half
  ...
NLP as machine learning

• Modern NLP increasingly ignores linguistic theory in order to obtain models directly from data (visualized here)

• We still use linguistic theory to interrogate (or ‘probe’) the resulting models.
NLP as artificial intelligence

• NLP involves resolving ambiguity at all levels.
  • **Reasoning with world knowledge.**
    • In the early days knowledge was explicitly encoded in artificial symbolic systems (e.g., context-free grammars) by experts.

• We tend to use probabilities (or pseudo-probabilities) to distinguish subtly different competing hypotheses.
  • E.g., is Google a noun or a verb?
  • Examples where Google ∈ Nouns (“Google makes Android”), does not mean that Google is never a verb (“Go Google yourself”).

  • \( P(\text{Google } \in \text{ Nouns}) > P(\text{Google } \in \text{ Verbs}) > 0 \)
The Turing Test

• First and most fundamental test of machine intelligence.

• A machine (A) imitates a human using nothing but a text-based instant messenger.
  • If a human interrogator (C) cannot reliably differentiate a real human (B) from the machine, that machine is said to be ‘intelligent’.

Aside – Chatbots

• ELIZA (Weizenbaum, 1966): simple pattern matching to imitate a psychiatrist.
• Surprisingly effective despite no linguistic knowledge.
• e.g.,
  User: Men are all alike.
  ELIZA: In what way?
  User: They’re always bugging us about something or other.
  ELIZA: Can you think of a specific example?
  User: My boyfriend made me come here.
  ELIZA: Your boyfriend made you come here.  
  (Jurafsky and Martin, 2009)
Course outline (approximate)

• Introduction, linguistic data, language models (3 lectures)
• Features and classification (2 lecture) *
• Entropy and information theory (2 lectures) *
• Neural language models (3 lectures) *
• Machine translation (3 lectures) **
• Hidden Markov models (3 lectures) *
• Natural Language Understanding (2 lectures) *
• Automatic speech recognition (2 lectures) **
• Information retrieval (1 lecture) **
• Interpretability and Large Language Models (2 lectures) *
• Review (1 lecture)

* techniques    ** applications
Preview: Machine translation

• One of the most prized applications in NLP.
• Requires both interpretation and generation.
• Over $100B spent annually on human translation.
According to the data provided today by the Ministry of Foreign Trade and Economic Cooperation, as of November this year, China has actually utilized 46.959B US dollars of foreign capital, including 40.007B US dollars of direct investment from foreign businessmen.

The Ministry of Foreign Trade and Economic Cooperation, including foreign direct investment 40.007B US dollars today provide data include that year to November China actually using foreign 46.959B US dollars and

Today’s available data of the Ministry of Foreign Trade and Economic Cooperation shows that China’s actual utilization of November this year will include 40.007B US dollars for the foreign direct investment among 46.959B US dollars in foreign capital.
Preview: Machine translation

- In the 1950s and 1960s direct word-for-word replacement was popular.
- Due to semantic and syntactic ambiguities and differences in source languages, results were mixed.

“The spirit is willing, but the flesh is weak”

The vodka is good, but the meat is rotten

US English

Russian
Preview: Machine translation

- One problem is disparity of meanings in languages.

**nation n.** a large body of people, associated with a particular **territory**, that is sufficiently conscious of its **unity** to seek or to possess a **government** of its own.

**nation n.** an aggregation of persons of the same **ethnic family**, often speaking the same **language** or cognate **languages**.

Stephen Harper
Former Prime Minister of Canada

Pauline Marois
Former Première Ministre du Québec
Preview: Machine translation

- **Solution**: automatically learn statistics on parallel texts

  ... citizen of Canada has the right to vote in an election of members of the House of Commons or of a legislative assembly and to be qualified for membership ...

  e.g., the *Canadian Hansards*: bilingual Parliamentary proceedings

  ... citoyen canadien a le droit de vote et est éligible aux élections législatives fédérales ou provinciales ...
Statistical machine translation

• Much of modern statistical machine translation is based on the following perspective...

When I look at an article in Russian, I say: ‘This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.’

Warren Weaver March, 1947

Claude Shannon July, 1948
Preview: Speech recognition

Buy ticket...
AC490...
yes

Multimodal interaction

Telephony

Dictation
Speech waveforms

“Two plus seven is less than ten”
Spectrograms

• Speech sounds can be thought of as overlapping sine waves.
  • Speech is split apart into a 3D graph called a ‘spectrogram’.
  • Spectrograms allow machines to extract statistical features that differentiate between different kinds of sounds.
Preview: Information retrieval

Google

WolframAlpha

what woman won more than one nobel prize

Marie Curie won the Nobel prize in 1903 for Physics and 1911 in Chemistry; Linus Pauling in 1954 (for Chemistry) and 1962 (for Peace); John Bardeen in 1956 (for Physics) and 1972; Frederick Sanger in Chemistry in 1958 and 1980. Who has won more than one Nobel prize? Apr 1, 2007

Who has won more than one Nobel prize? - Times of India
timesofindia.indiatimes.com/home/...won-more-than-one-Nobel-prize/...1839923.cms

People also ask
Who has won Nobel Prize twice?
What women won the Nobel Prize?
How many women have won the Nobel Prize?
How many women have been awarded the Nobel Peace Prize?
Answer questioning?

- **Retrieving information** can be a **clever combination** of many very **simple concepts** and algorithms.

\[
\cos (\mathbf{q}, \mathbf{d}) = \frac{\sum_{i=1}^{n} q_i d_i}{\sqrt{\sum_{i=1}^{n} q_i^2} \sqrt{\sum_{i=1}^{n} d_i^2}}
\]
Overview: NLP

• Is natural language processing (the discipline) hard?
  • **Yes**, because *natural language*
    • is highly ambiguous at all levels,
    • is complex and subtle,
    • is fuzzy and probabilistic,
    • involves real-world reasoning.
  • **No**, because *computer science*
    • gives us many powerful statistical techniques,
    • allows us to break the challenges down into more manageable features.

• Is Natural Language Computing (the course) hard?
  • More on this soon...
NLP in Industry
Natural language computing

- **Instructor:** En-Shiun Annie Lee, Raeid Saqur, Zining Zhu ([csc401-2021-01@cs](mailto:csc401-2021-01@cs))
- **Meetings:** MW (lecture), F (tutorial) at 10h and 11h
- **Languages:** English, Python.
- **Website:** Quercus
- **You:** Understand basic **probability**, can **program**, or can pick these up as we go.
- **Syllabus:** Key **theory** and **methods** in statistical natural language computing. Focus will be on **Markov and neural models**, **machine translation**, and **other topics**.
Office hours

- **Time:** TBD
- **Location:**
  - Zoom or In-Person in Bahen 7th floor
Evaluation policies

- **General**: Three assignments: 15%, 20%, 25% (ranked by your mark)  
  Final ‘assessment’: 40%

- **Lateness**: 10% deduction applied to electronic submissions  
  that are 1 minute late.  
  Additional 10% applied every 24 hours up to 72 hours total, at which point grade is zero.

- **Final**: If you fail the final ‘assessment’, then you fail the course.

- **Ethics**: Plagiarism and unauthorized collaboration can result in a grade of zero on the homework, failure of the course, or suspension from the University.  
  See the course website.
Theme – NLP in a post-truth society

• The **truth** is the most important thing in the Universe.
  • At the very least, the truth allows us to rationally **optimize** legal, political, and personal decisions.
• The truth can sometimes be obscured deliberately via **deception**, or inadvertently through **bias**, **fallacy**, or intellectual **laziness**.
  • Nowhere is this perhaps more obvious than on **social media** or in **pseudo-journalism**.
• Natural language processing **may** give us **tools** to combat this scourge.
Assignments

● Assignment 1: Corpus statistics, sentiment analysis
  task: analyze bias on Reddit
  learn: statistical techniques, features, and classification.

● Assignment 2: Neural machine translation
  task: translate between languages
  learn: neural seq2seq and language models.

● Assignment 3: Automatic speech recognition
  task: detect lies in speech
  learn: signal processing, phonetics, and hidden Markov models.
Projects – graduate students only

• Graduate students can *optionally* undertake a full-term *project* worth *60%* of their grade *instead* of the assignments.
  • Good for those, e.g., who prefer to work in teams.

• Teams must consist of 1 or 2 humans (*no more, no fewer*).
• Projects must contain a significant *programming* and *scientific* component.
• Projects must be *relevant* to the course.
Projects – graduate students only

- Some possible ideas for projects include:
  - A deception filter for news media online.
  - A novel method of using data in language $A$ to train a classification system in language $B$ for $A \neq B$.

- If you decide to take this option, you have to notify us by email about your team by **18 January**!

- You will need to periodically submit checkpoints that build on their antecedents.
  - See course webpage for detailed requirements!
Reading

https://search.library.utoronto.ca/details?10552907

SPEECH AND LANGUAGE PROCESSING
An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition

DANIEL JURAFSKY & JAMES H. MARTIN
Stats from 2017-2019

2017

<table>
<thead>
<tr>
<th>Grade</th>
<th>2017 Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>37.2%</td>
</tr>
<tr>
<td>B</td>
<td>24.1%</td>
</tr>
<tr>
<td>C</td>
<td>17.9%</td>
</tr>
<tr>
<td>D</td>
<td>9%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>2017 Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+</td>
<td>15.2%</td>
</tr>
<tr>
<td>A</td>
<td>9.7%</td>
</tr>
<tr>
<td>A-</td>
<td>12.4%</td>
</tr>
<tr>
<td>F</td>
<td>11.7%</td>
</tr>
<tr>
<td>OTH</td>
<td>0%</td>
</tr>
</tbody>
</table>

Class average excluding exam no shows: 75.20%
Fails excluding exam no shows: 3.79%

2019

<table>
<thead>
<tr>
<th>Grade</th>
<th>2019 Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>49.7%</td>
</tr>
<tr>
<td>B</td>
<td>27%</td>
</tr>
<tr>
<td>C</td>
<td>9.2%</td>
</tr>
<tr>
<td>D</td>
<td>3.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>2019 Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+</td>
<td>15.3%</td>
</tr>
<tr>
<td>A</td>
<td>16%</td>
</tr>
<tr>
<td>A-</td>
<td>18.4%</td>
</tr>
<tr>
<td>F</td>
<td>10.4%</td>
</tr>
<tr>
<td>OTH</td>
<td>0%</td>
</tr>
</tbody>
</table>

Class average excluding exam no shows: 77.52%
Fails excluding exam no shows: 4.58%

Consider the waitlist!
Assignment 1 – Bias in social media

• Involves:
  • Working with social media data
    (i.e., gathering statistics on some data from Reddit),
  • Part-of-speech tagging (more on this later),
  • Classification.

• **Announcements**: Piazza forum, email.
• You should get an early start.
Assignment 1 and reading

- **Assignment 1** available by Friday (on course webpage)!
  - Due 10 February
  - **TAs:** J Chen;
    KP Vishnubhotla.

- **Reading:**
  - Manning & Schütze: Sections 1.3—1.4.2,
    Sections 6.0—6.2.1.