CMPU 366

Computational Linguistics

Spring 2021
It’s Thursday, the 354th day of March, 2020.
'I wish it need not have happened in my time,' said Frodo. 

'So do I,' said Gandalf, 'and so do all who live to see such times. But that is not for them to decide. All we have to decide is what to do with the time that is given us.'

Today, we’ll talk a bit about

Course logistics,

Language, and

Using computers to work with language.
Course information
Prerequisites

Two 200-level CS courses

Strongly recommended:

CMPU 240: Theory of Computation
Course website

cs.vassar.edu/~cs366
On the course website will be links to the other two sites we use:

**CampusWire**

Use for general discussion about the course content
Use for all questions about the course

- You can post anonymously.
- You can send private questions to me.

**Gradescope**

Use for submitting assignments and receiving feedback
There will be several exercises for practice. They will only be graded for completion.
Grading

Approx. four larger programming assignments
The other half of the grade is the final project and presentation.
Syllabus

Read for more details on the course.
Student survey

https://forms.gle/c8weWkoepuwhXvCu9
Finally, an advertisement

I’m also teaching

**CMPU 187: Introduction to Digital Humanities,**

a six-week course in the second half of the semester. It’s 0.5 credit, pass/fail.
What is computational linguistics?
Computational linguistics is the study of language, using computers.

Natural language processing (NLP) is the study of how to make computers use language.

In this course, we’ll do a bit of each, but the emphasis will really be on NLP.
Natural language?
Natural language?

Mandarin Chinese, Spanish, English, Hindi, …
Natural language?

Mandarin Chinese, Spanish, English, Hindi, …

We’ll often focus on English because it’s the common language for people taking this class.
The linguists strike back

Marine biologist? Wow, so how many dolphins do you own?

Anesthesiologist, eh? So do you actually, like, feel anything?

So you're an ophthalmologist? Fascinating, so how many eyes do you have?
Computer models of human language understanding and generation can lead to insights about human processing.

But the emphasis on computational, not cognitive plausibility.
What’s natural language processing for?
Solving authorship mysteries!
What can I help you with?

October 2011
A major goal for NLP is the creation of *conversational agents* – computers we can talk with like we do with people.
DAVE BOWMAN: Open the pod bay doors, HAL.

HAL: I’m sorry, Dave, I’m afraid I can’t do that.

Stanley Kubrick and Arthur C. Clark, 2001: A Space Odyssey
What does it take to create an agent like HAL?

It would have to be able to

Understand via *speech recognition* and *natural language understanding*

Communicate via *natural language generation* and *speech synthesis*

---

Open the pod bay doors, HAL.

understanding

generation

computer

I'm sorry, Dave, I'm afraid I can't do that.
If Alexa were HAL...
If Alexa were HAL...
If Alexa were HAL...
Where else have you seen NLP in your life, in the news, or elsewhere?
Today’s applications

Conversational agents (Siri, Alexa)
Information extraction and question answering (Watson)
Machine translation (Google Translate)
Opinion and sentiment analysis
Social media analysis
Rich visual understanding
Essay evaluation
Mining legal, medical, or scholarly literature
Summarization
Spelling and grammar checking
Many of these problems have been worked on for a long time, but have improved dramatically in recent years.
Web searches
Web searches

Big improvements due to better NLP!
Google Translate, a few years ago

La mancanza di risorse linguistiche (RL) adeguate viene riconosciuta come uno dei principali ostacoli al successo delle attività di ricerca e sviluppo nella LC. Diverse agenzie governative statunitensi (NSF, ARPA, NSA, ecc.) e la CEE hanno indicato RL e standard come i temi di massima priorità. La loro produzione richiede uno sforzo cooperativo di competenze, finanziamenti, soggetti. Per questi motivi, abbiamo promosso attività internazionali (progetti CEE e CEE-NSF) e nazionali (MIUR) attraverso le quali i progetti di questa linea possono contare, oltre che su finanziamenti esterni, sulla cooperazione di molti soggetti pubblici e privati in Italia e all'estero (Europa, America, Giappone, paesi asiatici).

The lack of linguistic resources (RL) adequate is recognized as one of the main obstacles to the success of research and development in the LC. Several U.S. government agencies (NSF, ARPA, NSA, etc..) And the EEC indicated RL and standards such as the issues of highest priority. Their production requires a cooperative effort of skills, financing and subjects. For these reasons, we have promoted international activities (projects EEC and EEC-NSF) and national (MIUR) through which projects of this line can count not only on external funding, on the cooperation of many public and private entities in Italy and 'abroad (Europe, America, Japan, Asian countries).
Google Translate, more recently

The lack of linguistic resources (RL) adequate is recognized as one of the main obstacles to the success of research and development in the LC. Several U.S. government agencies (NSF, ARPA, NSA, etc.) and the EEC indicated RL and standards such as the issues of highest priority. Their production requires a cooperative effort of skills, financing and subjects. For these reasons, we have promoted international activities (projects EEC and EEC-NSF) and national (MIUR) through which projects of this line can count not only on external funding, on the cooperation of many public and private entities in Italy and abroad (Europe, America, Japan, Asian countries).

The lack of adequate linguistic resources (RL) is recognized as one of the main obstacles to the success of research and development activities in the LC. Several US government agencies (NSF, ARPA, NSA, etc.) and the EEC have indicated RL and standards as the top priority issues. Their production requires a cooperative effort of skills, financing and subjects. For these reasons, we have promoted international activities (CEE and CEE-NSF projects) and national (MIUR) projects through which the projects of this line can count, as well as external financing, on the cooperation of many public and private entities in Italy and in Italy, abroad (Europe, America, Japan, Asian countries).
NLP of the (near) future

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Example from Semantic Machines
NLP of the (near) future

Information retrieval / question answering

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

OK, let’s book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Example from Semantic Machines
NLP of the (near) future

Can you get me a table at O Ya tomorrow at 7?
AI: How many?
Four of us.
AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.
Ugh, that's sort of late. What about Oishii?
AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?
AI: OK, email sent.
AI: Liz says it's fine.
OK, let's book the table at Oishii.
AI: OK, all set for 4 people at 9:30 PM.
Some time later...
You know what, just take the O Ya one at 10 PM.
AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Example from Semantic Machines
NLP of the (near) future

**Ambiguity resolution**

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

- **AI:** OK, email sent.
- **AI:** Liz says it's fine.
- **OK, let's book the table at Oishii.**
- **AI:** OK, all set for 4 people at 9:30 PM.
- **Some time later...**
- **You know what, just take the O Ya one at 10 PM.**
- **AI:** Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.
NLP of the (near) future

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Example from Semantic Machines
NLP of the (near) future

Example from Semantic Machines

Word sense disambiguation

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.
NLP of the (near) future

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

AI: OK, let's book the table at Oishii.

Four of us.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

Ugh, that's sort of late. What about Oishii?

You know what, just take the O Ya one at 10 PM.

AI: Oishii has an opening at 9:30 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.
NLP of the (near) future

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?
AI: OK, email sent.
AI: Liz says it’s fine.
OK, let’s book the table at Oishii.
AI: OK, all set for 4 people at 9:30 PM.

You know what, just take the O Ya one at 10 PM.
AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Can you get me a table at O Ya tomorrow at 7?
AI: How many?
Four of us.
AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.
Ugh, that's sort of late. What about Oishii?
AI: Oishii has an opening at 9:30 PM.

Example from Semantic Machines
NLP of the (near) future

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Example from Semantic Machines
NLP of the (near) future

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there's nothing at 10 PM.

Spelling correction

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Example from Semantic Machines
NLP of the (near) future

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: OK, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Example from Semantic Machines
NLP of the (near) future

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Example from Semantic Machines
NLP of the (near) future

Can you get me a table at O Ya tomorrow at 7?
AI: How many?
Four of us.
AI: How many?
Four of us.
AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.
Ugh, that's sort of late. What about Oishii?
AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?
AI: OK, email sent.
AI: Liz says it's fine.
OK, let's book the table at Oishii.
AI: OK, all set for 4 people at 9:30 PM.
Some time later...
You know what, just take the O Ya one at 10 PM.
AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Example from Semantic Machines
NLP of the (near) future

Example from Semantic Machines

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.
NLP of the (near) future

Example from Semantic Machines

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.
NLP of the (near) future

Example from Semantic Machines

Summarization

Can you get me a table at O Ya at 7 PM?
AI: How many?
Four of us.
AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.
Ugh, that's sort of late. What about Oishii?
AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?
AI: OK, email sent.
AI: Liz says it's fine.
OK, let's book the table at Oishii.
AI: OK, all set for 4 people at 9:30 PM.

Some time later...
You know what, just take the O Ya one at 10 PM.
AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.
NLP of the (near) future

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there’s nothing at 7 PM. They have an opening at 10 PM.

Ugh, that’s sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it’s fine.

OK, let’s book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Example from Semantic Machines
NLP of the (near) future

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

AI: Oishii has an opening at 9:30 PM.

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

Four of us.

AI: How many?
NLP of the (near) future

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

Four of us.

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.

Example from Semantic Machines
NLP of the (near) future

Example from Semantic Machines

Translation (if we were in Osaka)

Can you get me a table at O Ya tomorrow at 7?

AI: How many?

AI: Sorry, there's nothing at 7 PM. They have an opening at 10 PM.

Ugh, that's sort of late. What about Oishii?

AI: Oishii has an opening at 9:30 PM.

Can you email Brooke, Mike and Liz and ask if 9:30 is too late?

AI: OK, email sent.

AI: Liz says it's fine.

OK, let's book the table at Oishii.

AI: OK, all set for 4 people at 9:30 PM.

Some time later...

You know what, just take the O Ya one at 10 PM.

AI: Ok, O Ya is booked at 10 PM for 4 people and I cancelled the reservation at Oishii.
How do humans do

Spelling correction,
Named entity extraction,
Question answering,
Coreference resolution,
Grounding,
Ambiguity resolution,
Summarization, and
Translation?

Long story short, we know languages!
What does it mean for a computer to understand language?
MIND
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing
“I shall replace the question ['can machines think?'] by another, which is closely related to it and is expressed in relatively unambiguous words.

“The new form of the problem can be described in terms of a game which we call the ‘imitation game’…”
“How long is your hair?”
“This is so 1950 it hurts.”

“How long is your hair?”
Big brain time
Big brain time

“How long is your hair?”
Big brain time

“How long is your hair?”

“It’s a cute pixie cut lol”
Turing suggested talking over teletypes, i.e., texting with the computer.

But we often prefer speech – and high-quality speech recognition and synthesis turns out to be one of the easier parts of making an intelligent machine!
Q: Please write me a sonnet on the subject of the Forth Bridge.
A: Count me out on this one. I never could write poetry.
Q: Add 34957 to 70764
A: (Pause about 30 seconds and then give as answer) 105621.
Q: Do you play chess?
A: Yes.

Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?
A: (After a pause of 15 seconds) R-R8 mate.
There’s a long history of making computers we can talk to, some serious, some not so serious.
ELIZA (c. 1964)

USER: You are like my father in some ways.

ELIZA: What resemblance do you see

USER: You are not very aggressive but I think you don’t want me to notice that.

ELIZA: What makes you think I am not aggressive

USER: You don’t argue with me

ELIZA: Why do you think I don’t argue with you

USER: You are afraid of me

ELIZA: Does it please you to believe I am afraid of you
“You are like $x$”
→ “What resemblance do you see”

“You are $x$”
→ “What makes you think I am $x$”
→ “Does it please you to believe I am $x$”

“You don’t $x$”
→ “Why do you think I $x$”

Note: Need to substitute “you” for “me”!
Levels of linguistic knowledge
Language processing applications use knowledge about language.

A simple word counting program must have knowledge about what it means to be a word.

More sophisticated programs like HAL require much broader and deeper knowledge of language.
**Phonetics**: the study of language sounds (*phones*), how they are physically formed.

/\l/ = alveolar lateral approximant  

/lace/

/\ɾ/ = alveolar tap  

/race/

/\ɾ/ = alveolar trill  

/rey (sp.)/

**Phonology**: systems of meaningfully distinct sounds (*phonemes*)

English: /\ɾ/ and /\ɾ/ are conflated

Japanese: /\ɾ/, /\ɾ/, and /\l/ are conflated

“shallow”

“deeper”
**Phonetics:** the study of language sounds (*phones*), how they are physically formed.

/\l\/ = alveolar lateral approximant; does

/\ɾ\/ = alveolar tap

/\r\/ = alveolar trill

**Phonology:** systems of meaningfully distinct sounds (*phonemes*)

English: /\ɾ\/ and /\r\/ are conflated

Japanese: /\ɾ\/, /\r\/, and /\l\/ are conflated

**MORAL:**

It is very hard to recognize speech.

It is very hard to wreck a nice beach?
ลูกศิษย์วัดกระหังยังอีกตอนทางเข้าไปมั่งการพระบาทเขาคิชฌกุฎ หัวดิ่งแท่นกับเจ้าถิ่นที่ออกมาเผชิญหน้าเพราะเดือดร้อนสัญจรไม่ได้ หวั่นเร่งทุกฝ่ายเจรจา ก่อนที่ชื่อเสียงของจังหวัดจะเสียหายไปมากกว่านี้ พร้อมเสนอหยุดจัดงาน 15 วัน....
Producing and recognizing variations of words requires knowledge of how words are formed in a given language, e.g.,

- talk +s
  present, 3rd person sing.
- talk +ed
  past

“shallower”

“deeper”

Speech

Text

Phonetics & Phonology

Orthography

Morphology

Lexemes

Syntax

Semantics

Pragmatics

Discourse
There are morphological rules:

To make a word plural, add “s”
   E.g., “dog” → “dogs”
If the word ends with “y”…
   E.g., “baby” → “babies”
If the word ends with “h”…
   E.g., “dish” → “dishes”

And there are exceptions:

The plural of “goose” is “geese”
The plural of “child” is “children”
“Did” is the past tense of “do”
uygarlaştıramadıklarımızdan mı zcasına

“(behaving) as if you are among those whom we could not civilize”

**TIFGOSH ET HA-LELED BA-GAN**

“you will meet the boy in the park”
chillax
unfriend
Obamacare
Twitterati
Manfuckinghattan

“shallower”
“deeper”
The challenges of words

Segmenting text into words (e.g., Thai example)

Morphological variation (e.g., Turkish and Hebrew examples)

Words with multiple meanings: bank, mean

Domain-specific meanings: latex

Non-compositional multiword expressions, e.g., New York or take out.
Example: Part-of-speech tagging

ikr   smh   he   asked   fir   yo   last   name

so   he   can   add   u   on   fb   lololol
Example: Part-of-speech tagging

I know, right  shake my head  for  your

ikr  smh  he  asked  fir  yo  last  name

so  he  can  add  u  on  fb  lololol

Facebook  laugh out loud
Example: Part-of-speech tagging

I know, right shake my head for your

ikr smh he asked fir yo last name

! G O V P D A N

interjection acronym pronoun verb prep. det. adj. noun

so he can add u on fb lololol

P O V V O P ^

proper noun

Facebook laugh out loud you

you
How do words fit together?

We can say

```
DT JJ NN
the blue boat
```

but not

```
NN JJ DT
boat blue the
```

That’s similar to in programming languages!

“shallower”

“deeper”
Example: Syntactic ambiguity

```
NP vs NP
  | NP
  |   | Noun
  |   | processing
  | NP
  |   | Noun
  |   | language
  | Adj.
  |   | natural
  |   | language

NP
  | Adj.
  |   | natural
  |   | language
  | NP
  |   | Noun
  |   | processing
```
HAL needs syntax to determine Dave’s utterance is a request for action:

Open the pod bay doors, HAL.

vs a statement:

HAL, the pod bay door is open.

or a question:

HAL, is the pod bay door open?
What does a sentence mean?

**Lexical semantics:** What do the words mean?
- *pod*
- *bay*
- *doors*

**Compositional semantics:** What does it mean to put them together in a particular way?
- *pod bay doors*
How sentences are used in different situations and how language use affects the interpretation of the sentence.

That is, not what the sentence means, but what the speaker means.

“shallower”

“deeper”

Speech  |  Text

Phonetics & Phonology  |  Orthography

Morphology
Lexemes
Syntax
Semantics
Pragmatics
Discourse
How sentences are used in different situations and how language use affects the interpretation of the sentence.

That is, not what the sentence means, but what the speaker means.

Can you get me a table at O Ya?
How sentences are used in different situations and how language use affects the interpretation of the sentence.

That is, not what the sentence means, but what the speaker means.

Can you get me a table at O Ya?

Yes, I have the ability to book reservations.
How sentences are used in different situations and how language use affects the interpretation of the sentence.

That is, not what the sentence means, but what the *speaker* means.

Can you get me a table at O Ya?

Yes, I have the ability to book reservations.
Concerns how the preceding sentences affect the interpretation of the next sentence – the thread of meaning.

“shallower”

“deeper”
Is NLP linguistics?

To be successful, a machine learner needs bias/assumptions; for NLP, that might be linguistic theory/representations.

NLP must contend with NL data as found in the world.

Linguistics has begun to use tools originating in NLP!
Is NLP machine learning?

Many NLP problems are reduced to ML problems, and this works better than anything that came before.

However, $\mathcal{R}$ is not directly observable.

Early connections to information theory (1940s)

Symbolic, probabilistic, and connectionist ML have all seen NLP as a source of inspiring applications.
Fields with connections to NLP

Machine learning

Linguistics (include psycho-, socio-, descriptive, and theoretical)

Cognitive science

Information theory

Logic

Data science

Political science

Psychology

Economics

Education

Theory of computation
The engineering side

Application tasks are difficult to define formally; they are always evolving.

Objective evaluations of performance are always up for debate.

People who succeed in NLP for long periods of time are foxes, not hedgehogs.
The history of NLP in one slide

1960s: Pattern-matching and rules; huge (unrealistic) ambitions

1970s–1980s: Linguistically rich, logic-driven systems; labor-intensive successes on a few narrow tasks.

1990s–2000s: Statistical modeling revolution, machine learning becomes a central part of NLP, systems start to be deployed for practical tasks.

2010s: Deep learning (neural networks) takes off, accelerates progress on most tasks.

Factors changing the NLP landscape

Increases in computing power

The rise of the web, then the social web

Advances in machine learning

Advances in understanding of language in social context

Hirschberg & Manning, 2015
Goals for this course

Introduce you to computational linguistics problems and solutions

Ultimate focus on handling ambiguity by probabilistic integration of evidence

Teaches key theory and methods for language modeling, tagging, parsing, etc.

Give you some hands-on practice with data and a handful of methods
Acknowledgments

This lecture incorporates material from:

Sam Bowman, New York University
Nancy Ide, Vassar College
Dan Jurafsky & James Martin, *Speech and Language Processing*
Jon May, University of Southern California
Noah Smith, University of Washington