What is Computational Linguistics?

• Study of computer processing of natural languages

• Interdisciplinary field
  – Roots in linguistics and computer science (specifically, AI)
  – Influenced by many other fields
The field goes by various names...

• **Computational linguistics (CL)**
  – the science of doing what linguists do with language, but using computers.

• **Natural language processing (NLP)**
  – the engineering discipline of doing what people do with language, but using computers.

• Speech/language/text processing
• Human language technology/technologies
Science vs. Engineering

• What is the goal of this endeavor?
  – Understanding the phenomenon of human language
  – Building better applications

• Goals (usually) in tension
  – Analogy: flight
Today

• What is computational linguistics?

• What does it mean for computers to process natural language?

• Why is this challenging?

• Class logistics
But first....
let’s get to know each other
Today

• What is computational linguistics?

• What does it mean for computers to process natural language?

• Why is this challenging?

• Class logistics
What’s a word?

• Break up by spaces, right?
  Ebay | Sells | Most | of | Skype | to | Private | Investors
  Swine | flu | isn’t | something | to | be | feared

• What about these?
  达赖喇嘛在高雄为灾民祈福
  ليبيريا تحيي ذكرى وصول القذافي إلى السلطة
  百貨店、8月も不振 大手5社の売り上げ8〜11%減
Morphological Analysis

• Morpheme = smallest linguistic unit that has meaning

• Morphemes are combined into words
  – duck + s = \([N \text{ duck}] + [\text{plural } s]\)
  – duck + s = \([V \text{ duck}] + [\text{3rd person singular } s]\)
  – happiness = \([\text{Adj happy}] + [\text{ness}]\)
### Complex Morphology

In Turkish, from the root “uyu-” (sleep), the following can be derived…

<table>
<thead>
<tr>
<th>Form</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>uyuyorum</td>
<td>I am sleeping</td>
</tr>
<tr>
<td>uyuyorsun</td>
<td>you are sleeping</td>
</tr>
<tr>
<td>uyuyor</td>
<td>he/she/it is sleeping</td>
</tr>
<tr>
<td>uyuyoruz</td>
<td>we are sleeping</td>
</tr>
<tr>
<td>uyuyorsunuz</td>
<td>you are sleeping</td>
</tr>
<tr>
<td>uyuyorlar</td>
<td>they are sleeping</td>
</tr>
<tr>
<td>uyuduk</td>
<td>we slept</td>
</tr>
<tr>
<td>uyudukça</td>
<td>as long as (somebody) sleeps</td>
</tr>
<tr>
<td>uyumalıyız</td>
<td>we must sleep</td>
</tr>
<tr>
<td>uyumadan</td>
<td>without sleeping</td>
</tr>
<tr>
<td>uyuman</td>
<td>your sleeping</td>
</tr>
<tr>
<td>uyarken</td>
<td>while (somebody) is sleeping</td>
</tr>
<tr>
<td>uyuyunca</td>
<td>when (somebody) sleeps</td>
</tr>
<tr>
<td>uyutmak</td>
<td>to cause somebody to sleep</td>
</tr>
<tr>
<td>uyutturmak</td>
<td>to cause (somebody) to cause (another) to sleep</td>
</tr>
<tr>
<td>uyutturturmak</td>
<td>to cause (somebody) to cause (some other) to cause (yet another) to sleep</td>
</tr>
</tbody>
</table>
What’s a phrase?

• Coherent group of words that serve some function
  – Organized around a central “head”
  – The head specifies the type of phrase

• Examples:
  – Noun phrase (NP): the happy camper
  – Verb phrase (VP): shot the bird
  – Prepositional phrase (PP): on the deck
Syntactic Analysis

- Parsing: the process of assigning syntactic structure

\[
S \\
| NP \\
| | V \\
| | | NP \\
| | | | det \\
| | | | | N \\
| | | | | | I \\
| | | | | saw \\
| | | | | the \\
| | | | | man \\

[S [NP I] [VP saw [NP the man]]]
Exercise

Bracket the phrases in the following English text

“paint branch drive”
Semantic analysis

different words/structure, same meaning

– She needed to make a quick decision in that situation.
– The scenario required her to make a split-second judgment.

– I saw the man.
– The man was seen by me.
Semantic analysis

same words, different meaning

- I walked by the bank
  - ... to deposit my check.
  - ... to take a look at the river.

– Everyone on the island speaks two languages.
– Two languages are spoken by everyone on the island.
Discourse Analysis

• Discourse: how multiple sentences fit together

• Pronoun reference:
  – The dog wanted the bone, but Sam threw it away.

• Inference and other relations between sentences:
  – The bomb exploded in front of the hotel. The fountain was destroyed, but the lobby was largely intact.
Pragmatics and World Knowledge

• Interpretation of sentences requires context, world knowledge, speaker intention/goals, etc.

• Rules of conversation
  – Can you tell me what time it is?
  – Could you pass the salt?

• Speech acts change the state of the world
  – Will you marry me?
Why is CL/NLP hard?

So easy...

Ambiguity!
Ambiguity at the word level

• Part of speech
  – [V Duck]!
  – [N Duck] is delicious for dinner.

• Word sense
  – I went to the bank to deposit my check.
  – I went to the bank to look out at the river.
Ambiguity at the syntactic level

• PP Attachment ambiguity
  – I saw the man on the hill with the telescope

• Structural ambiguity
  – I cooked her duck.
  – Visiting relatives can be annoying.
  – Time flies like an arrow.
Difficult cases...

• Requires world knowledge:
  – The city council denied the demonstrators the permit because they advocated violence
  – The city council denied the demonstrators the permit because they feared violence

• Requires context:
  – John hit the man. He had stolen his bicycle.
So how do humans cope?
How do computers cope?
Today

• What is computational linguistics?

• What does it mean for computers to process natural language?

• Why is this challenging?

• Class logistics
http://www.cs.umd.edu/class/fall2015/cmsc723/
Before next class...

- Read the syllabus
  http://www.cs.umd.edu/class/fall2015/cmsc723/
- Sign up for Piazza
  https://piazza.com/umd/fall2015/cmsc723/home
- Email me dates of religious holidays you will observe this semester

- Do the readings
- Get started on HW1