Why is NLP hard?

Lexical Ambiguity

- Will Will will Will's will?

Will	Will	will	Will's	will?
Modal verb (should/ would/ can)	Name of person	verb	Name of person	noun

Why is NLP hard?

Lexical Ambiguity

- *Will Will will Will's will?*
- *Rose rose to put rose roes on her rows of roses.*

Why is NLP hard?

There can be different interpretations of the same sentence.

Language ambiguity: Structural

- *The man saw the boy with the binoculars.*
- *Flying planes can be dangerous.*
- *Hole found in the room wall; police are looking into it.*

Language imprecision and vagueness

Why is NLP hard?

Language ambiguity: Structural

- *The man saw the boy with the binoculars.*
- *Flying planes can be dangerous.*
- *Hole found in the room wall; police are looking into it.*

Language imprecision and vagueness

- *It is very warm here.*

Why is NLP hard?

Language ambiguity: Structural

- *The man saw the boy with the binoculars.*
- *Flying planes can be dangerous.*
- *Hole found in the room wall; police are looking into it.*

Language imprecision and vagueness

- *It is very warm here.*
- *Q: Did your mother call your aunt last night?*
  A: I'm sure she must have.

But that's the fun part of it

Why is the teacher wearing sun-glasses?

...

But that's the fun part of it

Why is the teacher wearing sun-glasses?

...

Because the class is so **bright**.

Ambiguities

News Headlines

- Hospitals Are Sued by 7 Foot Doctors
- Stolen Painting Found by Tree
- Teacher Strikes Idle Kids

Ambiguity is pervasive

- Find at least 5 meanings of this sentence:
  - I made her duck

Ambiguity is pervasive

- Find at least 5 meanings of this sentence:
  - I made her duck
- I cooked duck for her
- I cooked duck belonging to her
- I created the (artificial) duck, she owns
- I caused her to quickly lower her head or body
- I waved my magic wand and turned her into a duck

Ambiguity is pervasive

Syntactic Category

- 'Duck' can be a noun or verb
- 'her' can be a possessive ('of her') or dative ('for her') pronoun

Ambiguity is pervasive

Syntactic Category

- 'Duck' can be a noun or verb
- 'her' can be a possessive ('of her') or dative ('for her') pronoun

Word Meaning

- 'make' can mean 'create' or 'cook'

Ambiguity is pervasive

Grammar

make can be

I cooked a duck belonging to her

- **Transitive:** (verb with a noun direct object)
- **Ditransitive:** (verb has 2 noun objects) I cooked a duck for her
- **Action-transitive:** (verb has a direct object + verb)

I made her lower her body

Ambiguity is pervasive

Grammar

make can be

- **Transitive:** (verb with a noun direct object)
- **Ditransitive:** (verb has 2 noun objects)
- **Action-transitive:** (verb has a direct object + verb)

Phonetics

- I'm eight or duck
- I'm aid her duck

Ambiguity is Explosive

- I saw the man with the telescope. **2 parses**

Ambiguity is Explosive

- I saw the man with the telescope. **2 parses**
- I saw the man on the hill with the telescope. **5 parses**

Ambiguity is Explosive

- I saw the man with the telescope. **2 parses**
- I saw the man on the hill with the telescope. **5 parses**
- I saw the man on the hill in Texas with the telescope. **14 parses**

Ambiguity is Explosive

- I saw the man with the telescope. **2 parses**
- I saw the man on the hill with the telescope. **5 parses**
- I saw the man on the hill in Texas with the telescope. **14 parses**
- I saw the man on the hill in Texas with the telescope at noon. **42 parses**

Ambiguity is Explosive

- I saw the man with the telescope. **2 parses**
- I saw the man on the hill with the telescope. **5 parses**
- I saw the man on the hill in Texas with the telescope. **14 parses**
- I saw the man on the hill in Texas with the telescope at noon. **42 parses**
- I saw the man on the hill in Texas with the telescope at noon on Monday. **132 parses**

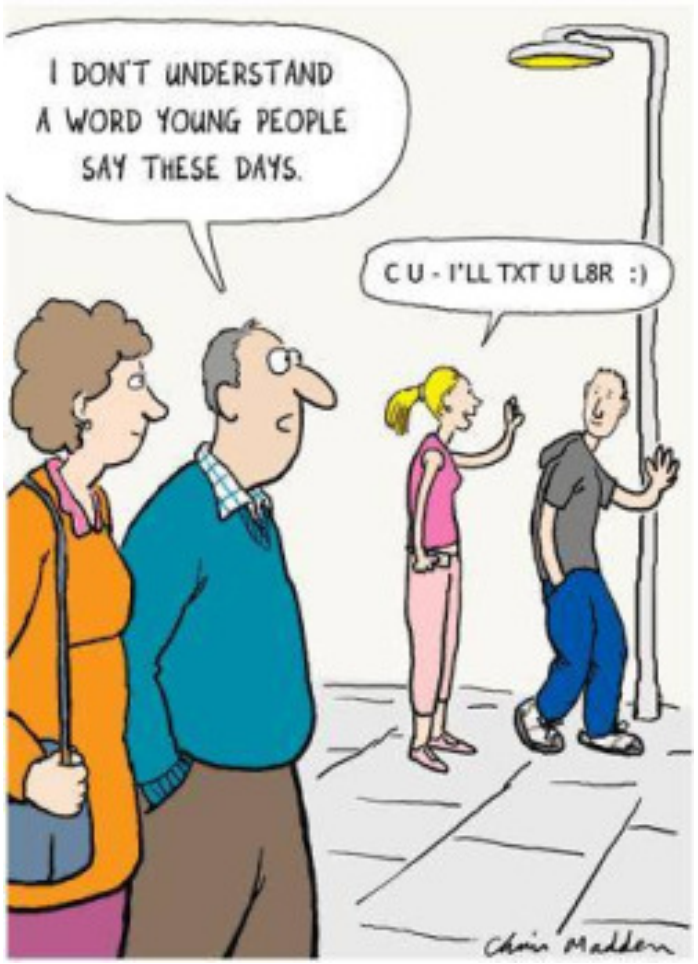
Why is Language Ambiguous?

- The goal in the production and comprehension of natural language is *efficient* communication.
- Allowing resolvable ambiguity
  - permits shorter linguistic expressions
  - avoids language being overly complex
- Language relies on people's ability to use their knowledge and inference abilities to properly resolve ambiguities

Natural Languages vs. Computer Languages

- Ambiguity is the primary difference between natural and computer languages.
- Formal programming languages are designed to be unambiguous
  - *Formal programming languages can be defined by a grammar that produces a unique parse for each sentence in the language.*
- Programming languages are also designed for efficient (deterministic) parsing.

Why else is NLP hard?



Why else is NLP hard?

Non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either

Why else is NLP hard?

Non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either

Segmentation Issues

the New York-New Haven Railroad

the [New] [York-New] [Haven] [Railroad]

the [New York]-[New Haven] [Railroad]

Different ways to segment a given text

Why else is NLP hard?

Idioms

- dark horse
- Ball in your court
- Burn the midnight oil

Cannot capture the meaning of the phrase by the meaning of the individual words

neologisms

- unfriend
- retweet
- Google/Skype/photoshop

New words getting added to the vocabulary

Why is NLP hard?

New Senses of a word

- That's *sick* dude!
- Giants

Why is NLP hard?

New Senses of a word

- That's *sick* dude!
- Giants ... *multinationals, conglomerates, manufacturers*

Why is NLP hard?

New Senses of a word

- That's *sick* dude!
- Giants ... *multinationals, conglomerates, manufacturers*

Tricky Entity Names

- Where is *A Bug's Life* playing ...
- *Let It Be* was recorded ...

What we do in NLP?

Tools Required

- Knowledge about language
- Knowledge about the world
- A way to combine knowledge resources

What we do in NLP?

Tools Required

- Knowledge about language
- Knowledge about the world
- A way to combine knowledge resources

How is it generally done?

- Probabilistic models built from language data
  - P(“maison” $\rightarrow$ “house”) is high

What we do in NLP?

Tools Required

- Knowledge about language
- Knowledge about the world
- A way to combine knowledge resources

How is it generally done?

- Probabilistic models built from language data
  - P("maison" $\rightarrow$ "house") is high
  - P(I saw a van) > P(eyes awe of an)
- Extracting rough text features does half the job.