Natural Language Processing

114

Why is NLP hard?

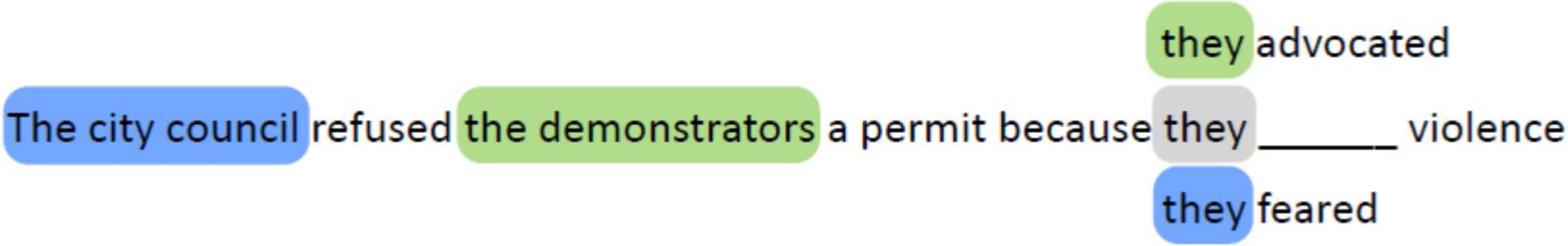
Silly Sentences

- Children make delicious snacks
- Stolen painting found by tree
- I saw the Rockies flying to San Francisco
- Court to try shooting defendant
- Ban on nude dancing on Governor’s desk
- Red tape holds up new bridges
- Government head seeks arms
- Cameron wins on budget, more lies ahead
- Local high school dropouts cut in half
- Hospitals are sued by seven foot doctors
- Dead expected to rise
- Miners refuse to work after death
- Patient at death's door - doctors pull him through
- In America a woman has a baby every 15 minutes. How does she do that?

The Winograd Schema Challenge

The city council refused the demonstrators a permit because they _____ violence

The Winograd Schema Challenge



More Classic Examples

- Beverly Hills
- Beverly Sills
- The box is in the pen
- The pen is in the box
- Mary and Sue are mothers
- Mary and Sue are sisters
- Every American has a mother
- Every American has a president


SADANDUSELESS.COM
Coffee or tea?
Coffee.
Wrong. It's tea.

Ambiguous Words

- ball, board, plant
  - meaning
- fly, rent, tape
  - part of speech
- address, resent, entrance, number, unionized
  - pronunciation – give it a try

Answer to the quiz

- address
  - The stress can be on either syllable. Compare with transport, effect, outline
- resent
  - As a verb infinitive or as “re-sent” a letter
- entrance
  - As a noun or as a verb meaning to put someone in a trance
- number
  - As a noun but also as the comparative of the adjective “numb”

Syntax vs. Semantics

** Little a has Mary lamb.*

? Colorless green ideas sleep furiously.

[Chomsky 1957]

Syntactic Amgbiguity

- How many different interpretations does the above sentence have?
- How many of them are reasonable/grammatical?

Time flies like an arrow.

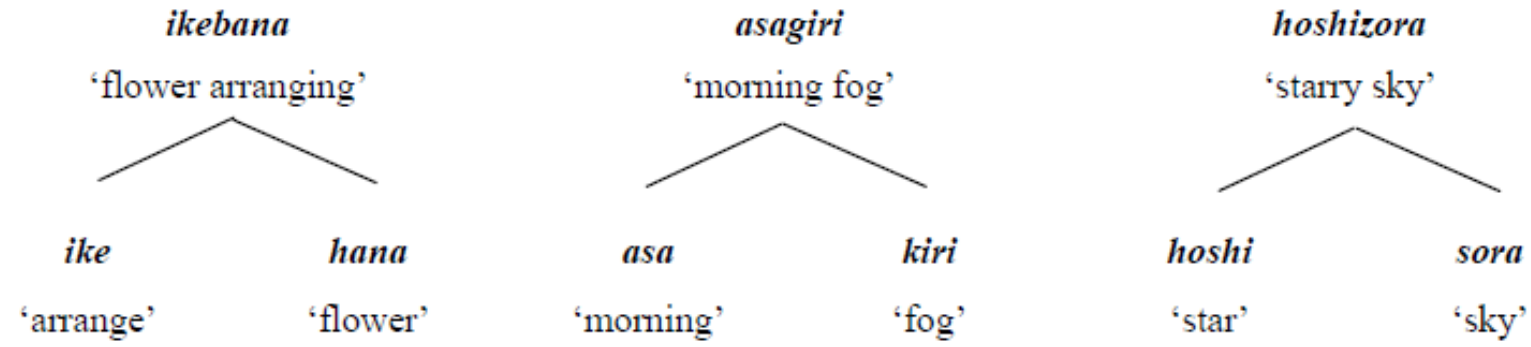
Quiz Answer

- The most obvious meaning is
  - time flies very fast; as fast as an arrow.
- This is a metaphorical interpretation.
  - Computers are not really good at metaphors.
- Other interpretations:
  - Flies like honey -> flies like an arrow -> fruit flies like an arrow
  - Take a stopwatch and time the race -> time the flies

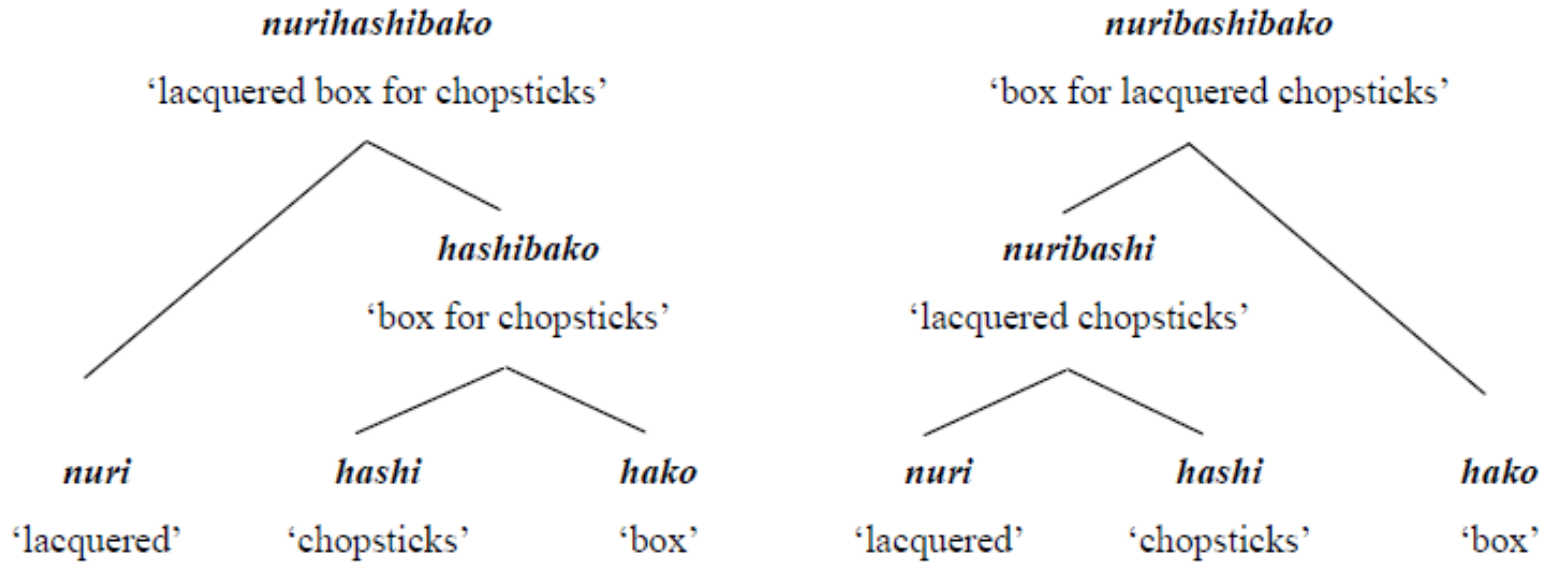
NACLO Problem

- Fakepapershelfmaker, by Willie Costello
  - http://www.nacloweb.org/resources/problems/2008/N2008-F.pdf

In English, we can combine two nouns to get a compound noun, such as in ‘mailbox’ or ‘sandcastle’. We can do this in Japanese as well, but just sticking the two words together isn’t enough. Instead, the words themselves undergo predictable changes:



Compound words can then be compounded again, creating compounds with three or more members. Study the diagrams below carefully. You’ll notice that the order in which the compound is built affects both the meaning and the final form of the word.



F1. The following is a list of several Japanese words with their English meanings. Use this word bank to write definitions of the Japanese compounds (a)-(f). Be very specific with how you phrase your definition. If your definition is ambiguous (has two meanings), it will not be counted.

sakura	cherry blossom	***kami***	paper	***nise***	fake
shiru	soup	***tana***	shelf	***tsukuri***	maker
iro	color(ed)	***tanuki***	raccoon	***hako***	box

(a) ***nisetanukijiru***	
(b) ***nisedanukijiru***	
(c) ***irogamibako***	
(d) ***irokamibako***	
(e) ***nisezakuradana***	
(f) ***nisesakuradana***	

F2. Match the following four-member Japanese compound words on the left with their English meanings on the right. (Some will require you to stretch your imagination a bit!) One of the Japanese words will correspond to two possible English meanings.

____ (1) a fake (fraudulent) shelf-maker made of paper	(A) ***nisegamidanadzukuri***
____ (2) a maker of fake shelves for paper	(B) ***nisekamitanadzukuri***
____ (3) a fake (fraudulent) maker of shelves for paper	(C) ***nisegamitanadzukuri***
____ (4) a shelf-maker made of fake paper	(D) ***nisekamidanadzukuri***
____ (5) a maker of shelves for fake paper	

F3. Explain your answers to F1 and F2 in the space provided below.

Solution

F1. The following is a list of several Japanese words with their English meanings; use them to write definitions of the Japanese compounds.

sakura	cherry blossom	***kami***	paper	***nise***	fake
shiru	soup	***tana***	shelf	***tsukuri***	maker
iro	color(ed)	***tanuki***	raccoon	***hako***	box

(a) ***nisetanukijiru***	fake soup made out of raccoons
(b) ***nisedanukijiru***	soup made out of fake raccoons
(c) ***irogamibako***	box for colored paper
(d) ***irokamibako***	colored box for paper
(e) ***nisezakuradana***	shelf for fake cherry blossoms
(f) ***nisesakuradana***	fake shelf for cherry blossoms

F2. Match the following four-member Japanese compound words with their English meanings; one of the Japanese words has two possible meanings.

(1) a fake shelf-maker made of paper	B: ***nisekamitanadzukuri***
(2) a maker of fake shelves for paper	D: ***nisekamidanadzukuri***
(3) a fake maker of shelves for paper	D: ***nisekamidanadzukuri***
(4) a shelf-maker made of fake paper	C: ***nisegamitanadzukuri***
(5) a maker of shelves for fake paper	A: ***nisegamidanadzukuri***

F3. Explain your answers.

When we compound two Japanese words, the first word modifies/describes the second. For example, adding *hashi* before *hako* makes a word meaning a box (*hako*) for chopsticks (*hashi*). As another example, adding *nuri* before *hashi* makes a word meaning chopsticks (*hashi*) that are lacquered (*nuri*).

Every simple (noncompound) word has two forms: the basic form, used when it occurs alone, and the variant form, sometimes used in compound words.

Basic	Variant
hako	*bako*
hana	*bana*
hashi	*bashi*
kami	*gami*
kiri	*giri*
sakura	*zakura*

Basic	Variant
shiru	*jiru*
sora	*zora*
tana	*dana*
tanuki	*danuki*
tsukuri	*dzukuri*

The variant form has a different first letter, which depends on the first letter in the basic form. Specifically, we replace the initial *h* with *b*, initial *k* with *g*, initial *s* with *z*, initial *sh* with *j*, initial *t* with *d*, and initial *ts* with *dz*. As a side note, some letters do not require replacement, but they do not occur in the problem.

NACLO Problem Solutions

- One Two Tree
  - http://www.nacloweb.org/resources/problems/2012/N2012-RS.pdf
- Fakepapershelfmaker
  - http://www.nacloweb.org/resources/problems/2008/N2008-FS.pdf

Types of Ambiguity

- Morphological:
  - Joe is quite impossible. Joe is quite important.
- Phonetic:
  - Joe’s finger got number.
- Part of speech:
  - Joe won the first round.
- Syntactic:
  - Call Joe a taxi.
- Prepositional phrase attachment:
  - Joe ate pizza with a fork / with meatballs / with Samantha / with pleasure.
- Sense:
  - Joe took the bar exam.

Other Sources of Difficulty

- Subjectivity:
  - Joe believes that stocks will rise.
- Cc attachment:
  - Joe likes ripe apples and pears.
- Negation:
  - Joe likes his pizza with no cheese and tomatoes.
- Referential:
  - Joe yelled at Mike. He had broken the bike.
  - Joe yelled at Mike. He was angry at him.
- Reflexive:
  - John bought him a present.
  - John bought himself a present.
- Ellipsis and parallelism:
  - Joe gave Mike a beer and Jeremy a glass of wine.
- Metonymy:
  - Boston called and left a message for Joe.

Other Sources of Difficulties

- Non-standard, slang, and novel words and usages
  - A360, 7342.67, +1-646-555-2223
  - “spam” or “friend” as verbs
  - yolo, selfie, chillax – recently recognized as dictionary words
  - www.urbandictionary.com – (Parental Warning!)
- Inconsistencies
  - junior college, college junior
  - pet spray, pet llama
- Typoes and gramattical erorz ☺
  - reciept, John Hopkins, should of
- Parsing problems
  - Selbständigkeit (self-reliance)
  - cup holder
  - Federal Reserve Board Chairman

Other Sources of Difficulties

- Complex sentences
- Counterfactual sentences
- Humor and sarcasm
- Implicature/inference/world knowledge:
  - I was late because my car broke down.
  - Implies I have a car, I use the car to get to places, the car has wheels, etc.
  - What is not explicitly mentioned, what is world knowledge?
- Semantics vs. pragmatics
  - Do you know the time?
- Language is hard even for humans
  - Both first language and second language

What do we need in order to understand language?

- World knowledge: have access to information beyond the training data

DOJ *greenlights* *Disney* - *Fox* *merger*

Department of Justice

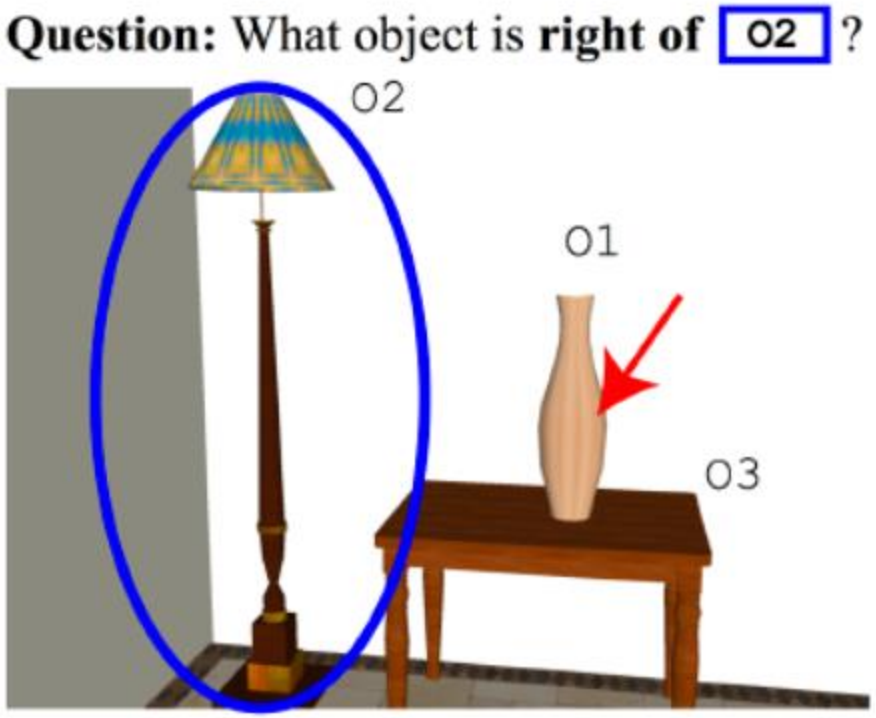
metaphor;
"approves"

- What is a green light? How do we understand what "green lighting" does?
- Need commonsense knowledge

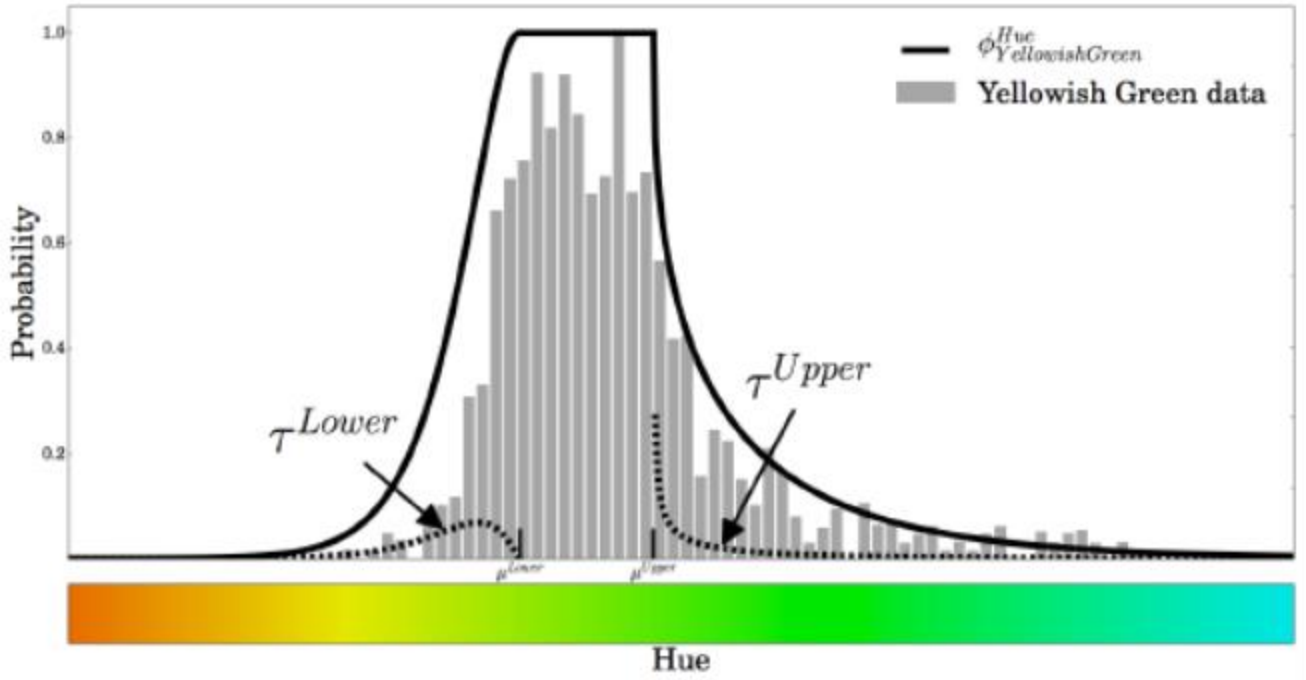
[Slide from Greg Durrett]

What do we need in order to understand language?

- Grounding: learn what fundamental concepts actually mean in a data-driven way



Golland et al. (2010)



McMahan and Stone (2015)

[Slide from Greg Durrett]

What do we need in order to understand language?

- Linguistic structure
- …but computers probably won't understand language the same way humans do
- However, linguistics tells us what phenomena we need to be able to deal with and gives us hints about how language works

a. John has been having a lot of trouble arranging his vacation.

b. He cannot find anyone to take over his responsibilities. (he = John)
C_b = John; C_f = {John}

c. He called up Mike yesterday to work out a plan. (he = John)
C_b = John; C_f = {John, Mike} (CONTINUE)

d. Mike has annoyed him a lot recently.
C_b = John; C_f = {Mike, John} (RETAIN)

e. He called John at 5 AM on Friday last week. (he = Mike)
C_b = Mike; C_f = {Mike, John} (SHIFT)

Centering Theory
Grosz et al. (1995)

[Slide from Greg Durrett]

What is needed to build a robust NLP system

- Lots of data
- Linguistic intuition
- Appropriate representation
- Robust algorithms
- World knowledge
- Grounding

SOURCE	Cela constituerait une solution transitoire qui permettrait de conduire à terme à une charte à valeur contraignante.
HUMAN	That would be an interim solution which would make it possible to work towards a binding charter in the long term .
1x DATA	[this] [constituerait] [assistance] [transitoire] [who] [permettrait] [licences] [to] [terme] [to] [a] [charter] [to] [value] [contraignante] [.]
10x DATA	[it] [would] [a solution] [transitional] [which] [would] [of] [lead] [to] [term] [to a] [charter] [to] [value] [binding] [.]
100x DATA	[this] [would be] [a transitional solution] [which would] [lead to] [a charter] [legally binding] [.]
1000x DATA	[that would be] [a transitional solution] [which would] [eventually lead to] [a binding charter] [.]

[Example from Dan Klein]