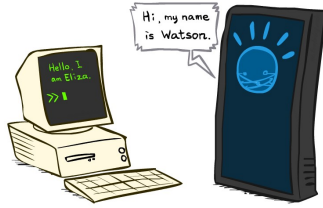


CSE 473: Artificial Intelligence

Advanced Applic's: Natural Language Processing



Steve Tanimoto --- University of Washington

[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at <http://ai.berkeley.edu>.]

What is NLP?



- Fundamental goal: analyze and process human language, broadly, robustly, accurately...
- End systems that we want to build:
 - Ambitious: speech recognition, machine translation, information extraction, dialog interfaces, question answering...
 - Modest: spelling correction, text categorization...

Problem: Ambiguities

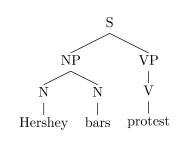
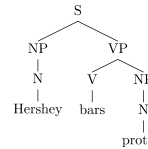
Headlines:

- Enraged Cow Injures Farmer With Ax
- Hospitals Are Sued by 7 Foot Doctors
- Ban on Nude Dancing on Governor's Desk
- Iraqi Head Seeks Arms
- Local HS Dropouts Cut in Half
- Juvenile Court to Try Shooting Defendant
- Stolen Painting Found by Tree
- Kids Make Nutritious Snacks



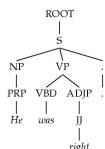
Why are these funny?

Parsing as Search



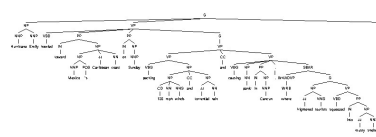
Grammar: PCFGs

- Natural language grammars are very ambiguous!
- PCFGs are a formal probabilistic model of trees
 - Each "rule" has a conditional probability (like an HMM)
 - Tree's probability is the product of all rules used
- Parsing: Given a sentence, find the best tree – search!



ROOT → S	375/420
S → NP VP	320/392
NP → PRP	127/539
VP → VBD ADJP	32/401
.....	

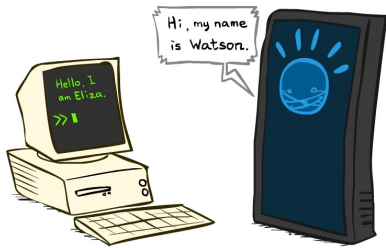
Syntactic Analysis



Hurricane Emily howled toward Mexico's Caribbean coast on Sunday packing 135 mph winds and torrential rain and causing panic in Cancun, where frightened tourists squeezed into musty shelters.

[Demo: Berkeley NLP Group Parser <http://tomato.banatao.berkeley.edu:8080/parser/parser.html>]

Dialog Systems



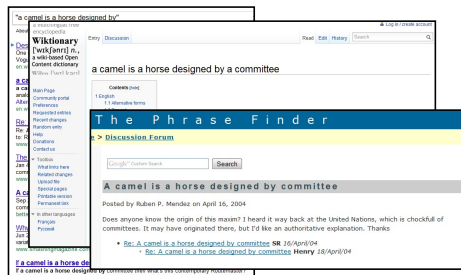
ELIZA



- A “psychotherapist” agent (Weizenbaum, ~1964)
- Led to a long line of chatterbots
- How does it work:
 - Trivial NLP: string match and substitution
 - Trivial knowledge: tiny script / response database
 - Example: matching “I remember ___” results in “Do you often think of ___”?
- Can fool some people some of the time?

[Demo: <http://nlp-addiction.com/eliza>]

Watson



What's in Watson?

- A question-answering system (IBM, 2011)
- Designed for the game of Jeopardy
- How does it work:
 - Sophisticated NLP: deep analysis of questions, noisy matching of questions to potential answers
 - Lots of data: onboard storage contains a huge collection of documents (e.g. Wikipedia, etc.), exploits redundancy
 - Lots of computation: 90+ servers
- Can beat all of the people all of the time?



Machine Translation



Machine Translation



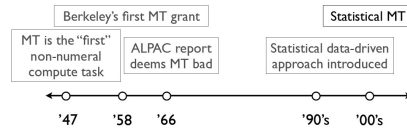
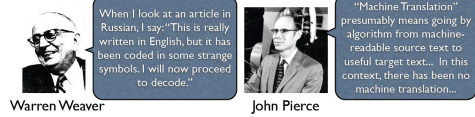
- Translate text from one language to another
- Recombines fragments of example translations
- Challenges:
 - What fragments? [learning to translate]
 - How to make efficient? [fast translation search]

The Problem with Dictionary Lookups

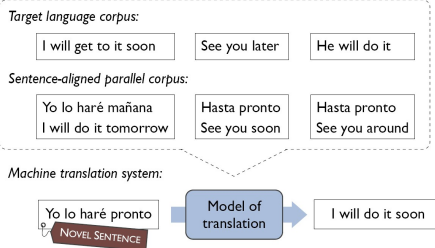
- 顶部 /**top**/roof/
- 顶端 /summit/peak/**top**/apex/
- 顶头 /coming directly towards one/**top**/end/
- 盖 /lid/**top**/cover/canopy/build/Gai/
- 盖帽 /surpass/**top**/
- 极 /extremely/pole/utmost/**top**/collect/receive/
- 尖峰 /peak/**top**/
- 面 /fade/side/surface/aspect/**top**/face/flour/
- 摘心 /**top**/topping/

Example from Douglas Hofstadter

MT: 60 Years in 60 Seconds



Data-Driven Machine Translation

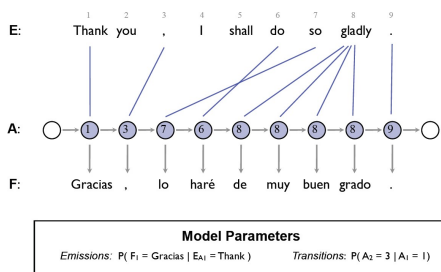


Learning to Translate

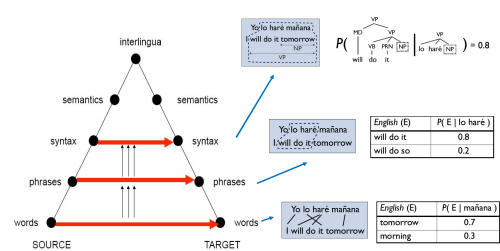
CLASSIC SOUPS		Sm.	Lg.
清 燉 雞 湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50 2.75
雞 飯 湯	58.	Chicken Rice Soup	1.85 3.25
雞 麵 湯	59.	Chicken Noodle Soup	1.85 3.25
廣東 湯	60.	Cantonese (Wonton) Soup	1.50 2.75
番茄 湯	61.	Tomato Clear Egg Drop Soup	1.65 2.95
粟 米 湯	62.	Regular (Wonton) Soup	1.10 2.10
鹹 辣 湯	63.	Hot & Sour Soup	1.10 2.10
香 花 湯	64.	Egg Drop Soup	1.10 2.10
雲 吞 湯	65.	Egg Drop (Wonton) Mix	1.10 2.10
豆 腐 菜 湯	66.	Tofu Vegetable Soup	NA 3.50
雞 玉 米 湯	67.	Chicken Corn Cream Soup	NA 3.50
蟹 肉 玉 米 湯	68.	Crab Meat Corn Cream Soup	NA 3.50
海 鮮 湯	69.	Seafood Soup	NA 3.50

Example from Adam Lopez

An HMM Translation Model



Levels of Transfer



Example: Syntactic MT Output

