NI
Natural language

Natural Language Understanding

CSE 415: Introduction to Artificial Intelligence
University of Washington
Spring, 2019

© S. Tanimoto and University of Washington, 2017
[Some of these slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley.

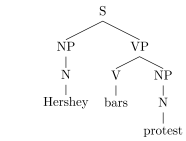
What is NLP? (Natural Language Processing)

- 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…

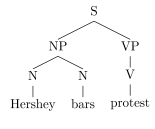
Problem: Ambiguities

- 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

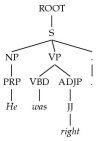


Hershey bars protest



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

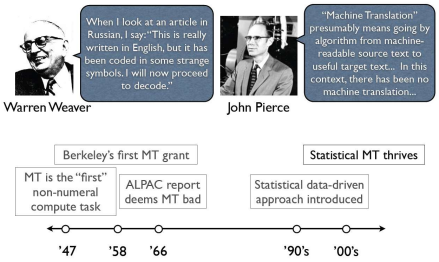
- Translate text from one language to another
- Recombines fragments of example translations
- Challenges:
  - What fragments? [learning to translate]
  - How to make efficient? [fast translation search]

The Problem with Dictionary Lookups

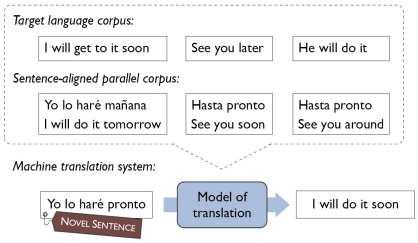
顶部	/**top**/roof/
顶端	/summit/peak/**top**/apex/
顶头	/coming directly towards one/**top**/end/
盖	/lid/**top**/cover/canopy/build/Gai/
盖帽	/surpass/**top**/
极	/extremely/pole/utmost/**top**/collect/receive/
尖峰	/peak/**top**/
面	/fade/side/surface/aspect/**top**/face/flour/
摘心	/**top**/topping/

Example from Douglas Hofstadter

MT: 60 Years in 60 Seconds



Data-Driven Machine Translation



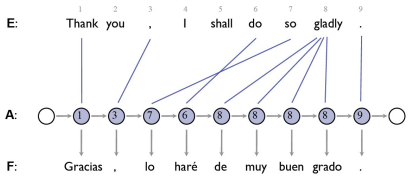
Learning to Translate

CLASSIC SOUPS

		Sm.	Lg.
57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
58.	Chicken Rice Soup	1.85	3.25
59.	Chicken Noodle Soup	1.85	3.25
60.	Cantonese (Wonton) Soup	1.50	2.75
61.	Tomato Clear Egg Drop Soup	1.65	2.95
62.	Regular (Wonton) Soup	1.10	2.10
63.	Hot & Sour Soup	1.10	2.10
64.	Egg Drop Soup	1.10	2.10
65.	Egg Drop (Wonton) Mix	1.10	2.10
66.	Tofu Vegetable Soup	NA	3.50
67.	Chicken Corn Cream Soup	NA	3.50
68.	Crab Meat Corn Cream Soup	NA	3.50
69.	Seafood Soup	NA	3.50

Example from Adam Lopez

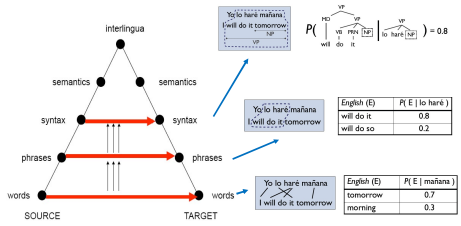
An HMM Translation Model



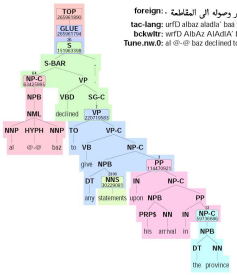
Model Parameters

Emissions: $P(F_1 = \text{Gracias} \mid E_{A_1} = \text{Thank})$ *Transitions:* $P(A_2 = 3 \mid A_1 = 1)$

Levels of Transfer



Example: Syntactic MT Output



foreign: ورفض الباز الادلاء باى تصريحات فور وصوله الى المقاطعة .

tac-lang: urfD albaz aladla' baa tSryHat fur uSulh ala almqaT'e .

bckwltr: wrfD AlbAz AlAdlA' bAY tSryHAt fwr wSwlh AlY AlmqATEp .

Tune.nw.0: al @-@ baz declined to make any statements upon his arrival in the province .

[ISI MT system output]

Nat. Lang. Understanding/Processing