Lecture 1:
Course Introduction

CMPSCI 585, Fall 2014
Introduction to Natural Language Processing
http://people.cs.umass.edu/~brenocon/inlp2014/

Brendan O'Connor (http://brenocon.com)

Some material borrowed from
Chris Manning and Jacob Eisenstein

Who, where, when

Instructor:		Brendan O'Connor	Office hours: Thursday after lecture (4pm–?), room 222
TA:		David Belanger	Office hours TBD (starts week of 9/8)
Lectures:	CS Building room 142, Tues and Thurs, 2:30–3:45pm		

- My email: brenocon@cs.umass.edu
- Course website:
  http://people.cs.umass.edu/~brenocon/inlp2014/
  (Reachable from my website)

2

What

- Learn fundamental principles and methods in natural language processing
  - Hands-on implementation experience
  - Appreciation of basic linguistic issues
  - Know about useful open-source NLP tools
  - Know when NLP works and when it doesn't
  - Be able to read current research papers in NLP
- "AI systems"

3

Requirements

- *[Submit all homeworks electronically via moodle.umass.edu]*

- (~15%) Short exercises (~weekly): basic linguistic analysis to sharpen intuitions, <1 hour of work. Due before class; bring paper copy to class.

- (~25%) Problem sets (several): conceptual and programming problems.

  - Python preferred. We provide starter code.

  - IPython Notebook: http://ipython.org/notebook.html

- (~20%) Midterm (in-class, mid-October)

- (~20%) Final exam

- (~20%) Group-based final projects

4

First exercise - due Thursday

- Find several ambiguous sentences from news articles, and explain.

- Exercise will be posted on website later today.

- Due this Thursday, electronic submission before class.
  Bring paper copy to class!

Tuesday, September 2, 14

Readings

- Readings will be provided -- see website.
- Suggested textbook:
  - Jurafsky and Martin, _Speech and Language Processing_
- Other great references:
  - Murphy, _Machine Learning: a Probabilistic Perspective_
  - Bender, _Linguistic Fundamentals for NLP_
  - Smith, _Linguistic Structure Prediction_ (free access from UMass campus)
  - Manning and Schütze, _Foundations of Stat. NLP_

6

Related courses at UMass

- Computational Linguistics: Ling 409, 492B (Bhatt, Dillon) Intended to prepare non-CS students for 585

  https://sites.google.com/site/umasslx409/home

  http://www.umass.edu/linguist/courses/detail.php?cid=571

- Speech: Ling 592B (Yu)

  http://courses.umass.edu/linguist592b-kmyu/category/info.html

- Information Retrieval: CS 446, 646 (Allan, Croft)

  http://ciir.cs.umass.edu/cmpsci446/

  http://ciir.cs.umass.edu/~allan/cs646/

NLP is interdisciplinary

Algorithms

Linguistics

Statistics +
Machine Learning

Cognitive Science

Artificial Intelligence

"Can Machines Think?"

- British mathematician and founding figure in computer science
- <u>Alan Turing (1950)</u>
- How do we know when we have AI?
- "Imitation Game"

NLP imagined
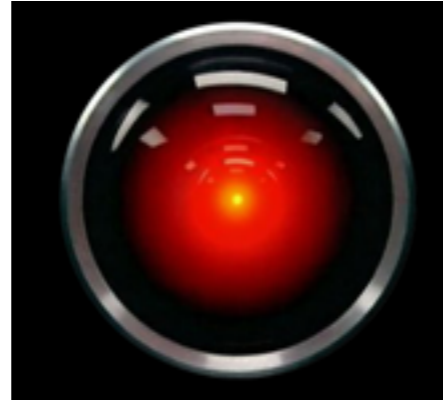
NLP today

- Speech interfaces
- Machine translation
- Sentiment analysis
- Search engines
- ...

11

NLP today: Speech interfaces

"Rao's coffee in Amherst, Massachusetts"
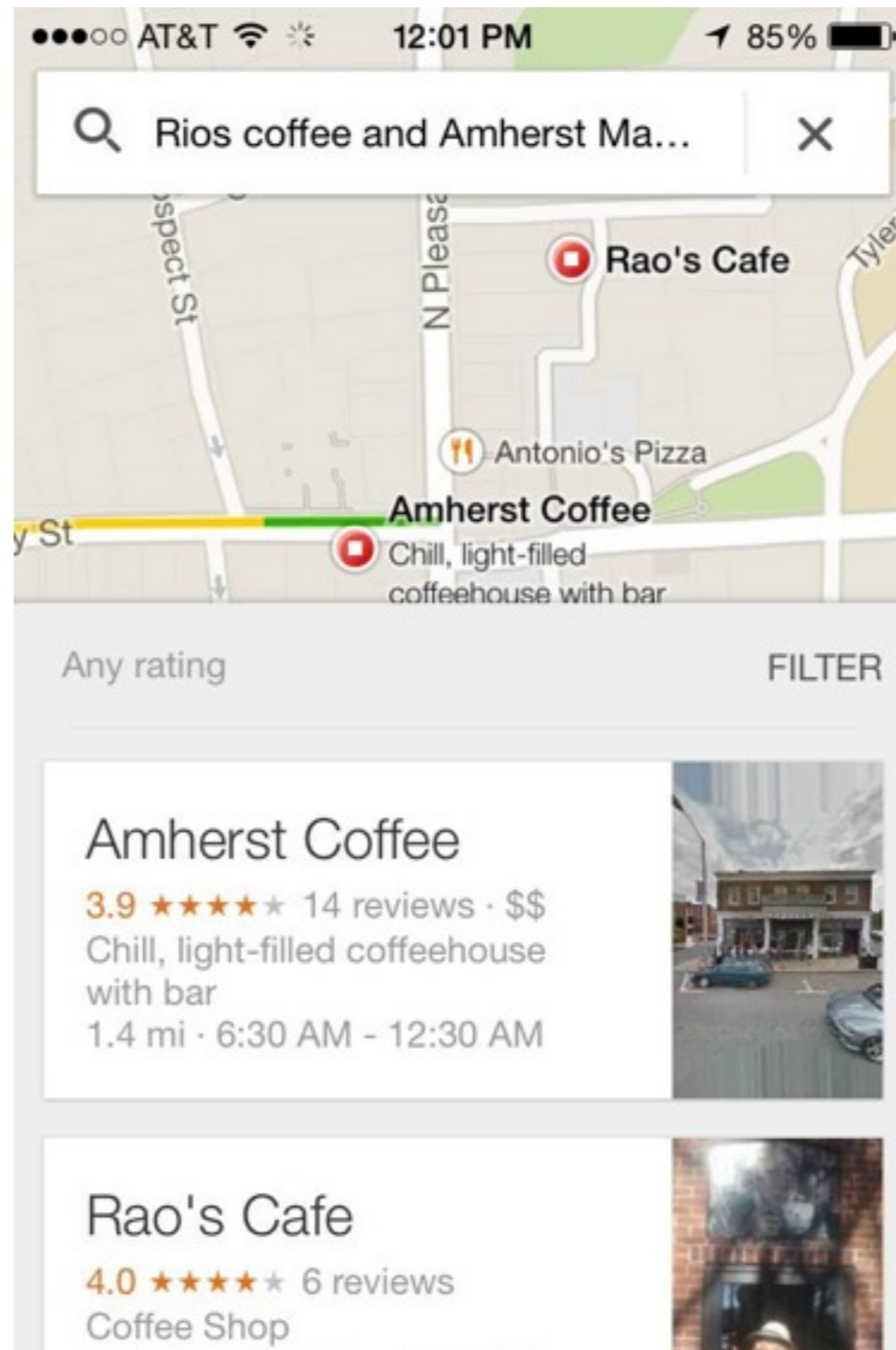
NLP today: Question answering

IBM Watson

Wanted for general evilness, last seen at the Tower of Barad-Dur. It's a giant eye, folks, kinda hard to miss

NLP today: Question answering

From <u>IBM Journal of Research and Development, 2012</u>
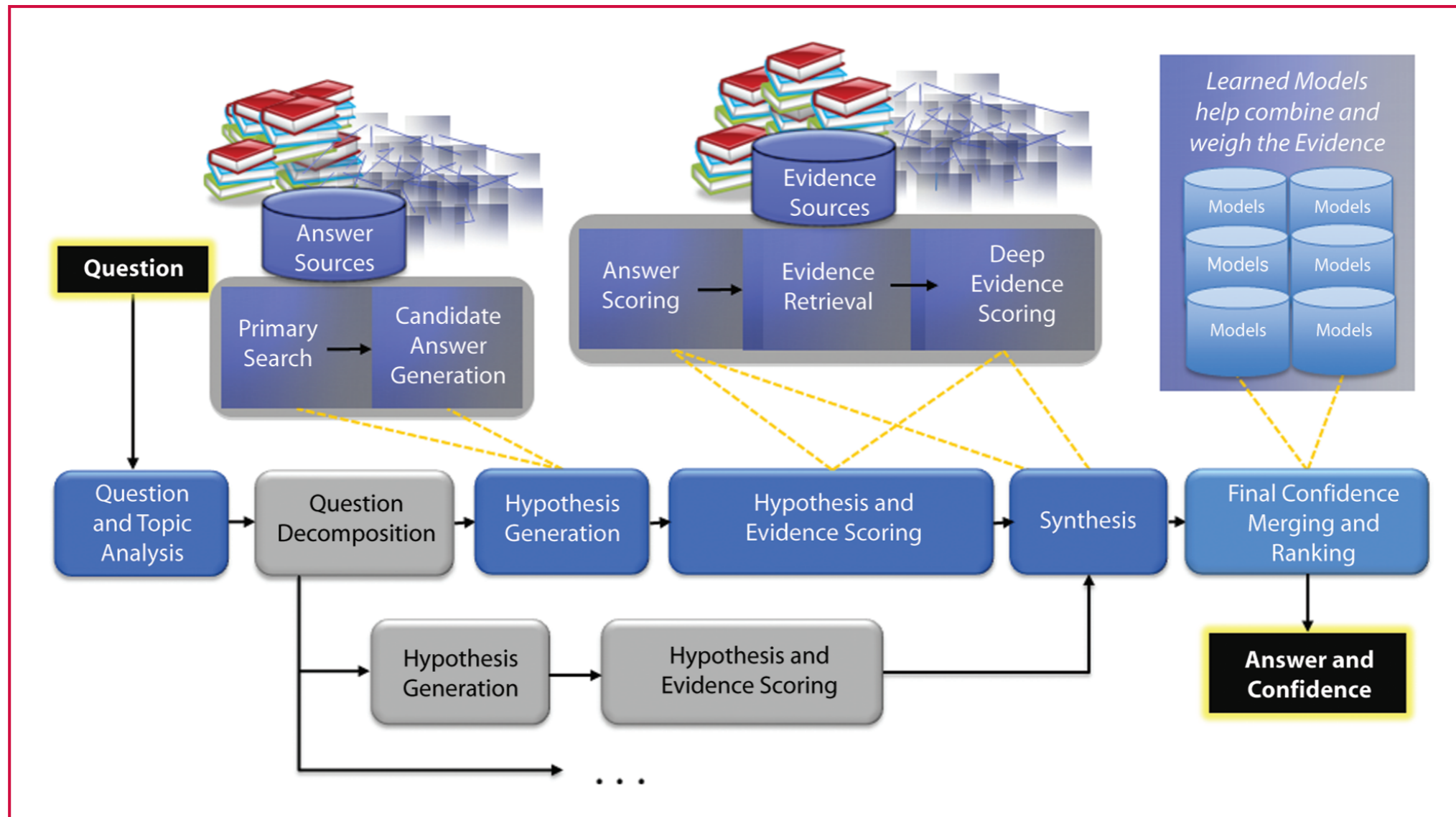
IBM Watson

25 engineers, 4 years, 200 subsystems,
2,880 CPU cores, 15 TB storage

NLP today: Machine translation

NLP today: Machine translation

NLP today: Trend analysis

Data: news articles

Dependency parsing to identify events

British officials in Tehran and London have been meeting discretely with their Iranian counterparts

GBR   IRN
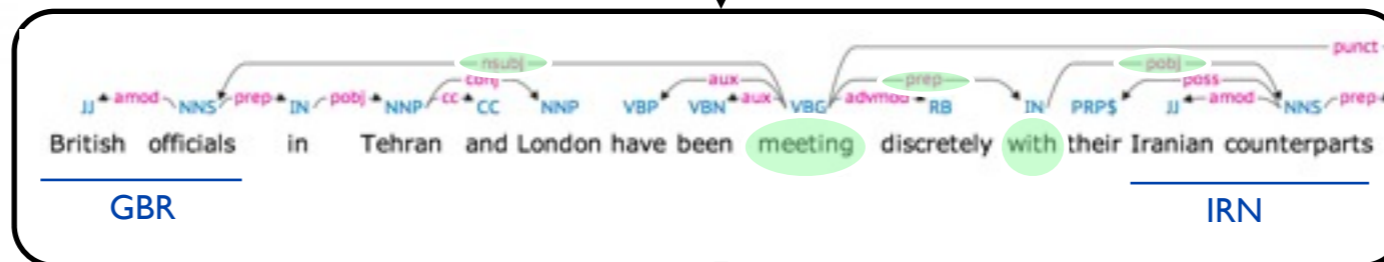
Machine learning from text:

### (1) **Event class dictionaries**

"diplomacy"
> arrive in, visit, meet with, travel to, leave, hold with, meet, meet in, fly to, be in, arrive for talk with, say in, arrive with, head to, hold in, due in, leave for, make to, arrive to,

"verbal conflict"
> accuse, blame, say, break with, sever with, blame on, warn, call, attack, rule with, charge, say←ccomp come from, say ←ccomp, suspect, slam, accuse government ←poss,

"material conflict"
> kill in, have troops in, die in, be in, wound in, have soldier in, hold in, kill in attack in, remain in, detain in, have in, capture in, stay in, about ←pobj troops in, kill, have troops

(2) **Political dynamics**

Israeli–Palestinian Diplomacy
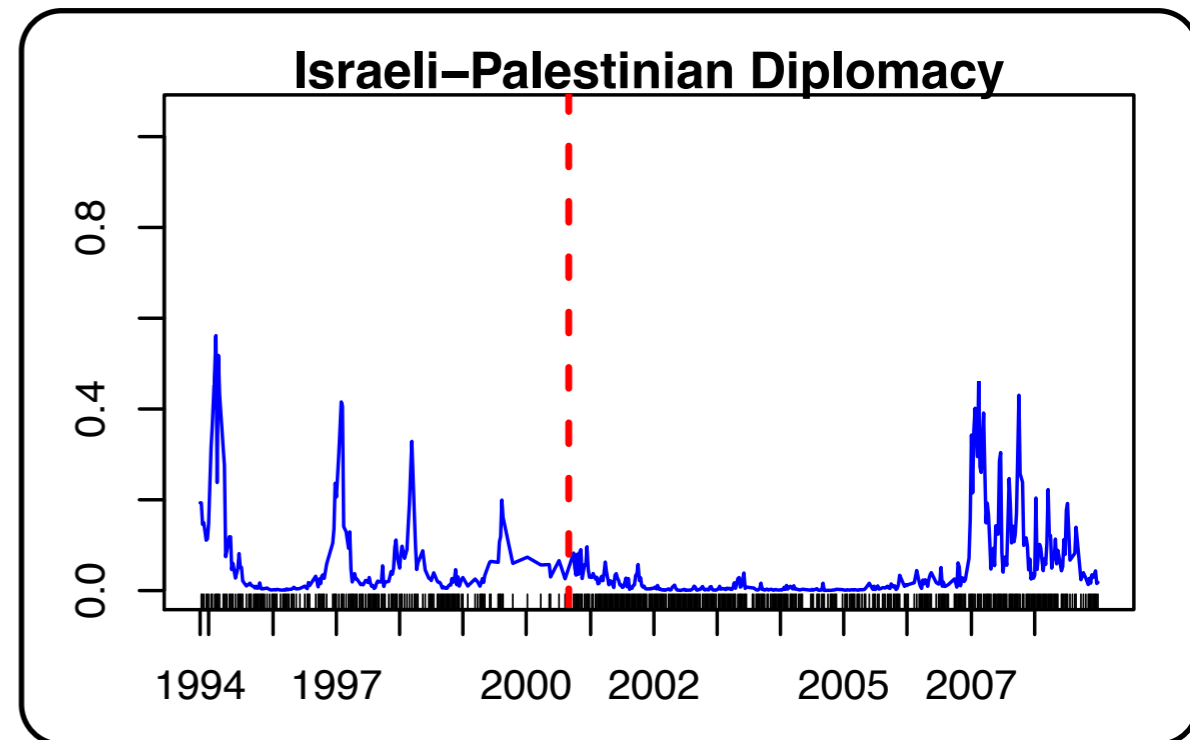
NLP today: Story generation

Earnings for OmniVision Technologies Expected to Fall

Narrative Science

FOLLOW

full bio →

Opinions expressed by Forbes Contributors are their own.

By Narrative Science

+ Comment Now   + Follow Comments

Wall Street is expecting lower profit for **OmniVision Technologies** when the company reports its first quarter results on Thursday, August 28, 2014. Analysts are expecting earnings per share of 39 cents after the company booked a profit of 42 cents a share a year earlier.

The consensus estimate has risen from 16 cents over the past three months. Analysts are expecting earnings of 99 cents per share for the fiscal year. Revenue is projected to eclipse the year-earlier total of $373.7 million by 2%, finishing at $381.5 million for the quarter. For the year, revenue is projected to come in at $1.39 billion.

http://www.forbes.com/sites/narrativescience/

NLP today: Search/summarization

NLP today: Search/summarization

NLP today: Search/summarization

TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL

Entity Extraction

- Have technology (thanks to R6) – for English, Arabic and Chinese
- Allow queries like:
- Show me all the word documents with references to IAEO
- Show me all documents that reference Osama Bin Laden
- Will allow a 'show me more like this' capability

TOP SECRET//COMINT//REL TO USA, AUS, CAN, GBR, NZL

NSA slides from Snowden leaks

http://www.theguardian.com/world/interactive/2013/jul/31/nsa-xkeyscore-program-full-presentation

Ambiguity: why NLP is hard

- Juvenile Court to Try Shooting Defendant
- Hospitals Are Sued by 7 Foot Doctors
- Alice saw Bob with a telescope.
- Our company is training workers.
- They found that in order to attract settlers -- and make a profit from their holdings -- they had to offer people farms, not just tenancy on manorial estates.

Natural interfaces?

Where is A Bug's Life playing in Mountain View?

A Bug's Life is playing at the Century 16 Theater.

When is it playing there?

It's playing at 2pm, 5pm, and 8pm.

OK. I'd like 1 adult and 2 children for the first show.
How much would that cost?

But we need domain knowledge, discourse knowledge, world knowledge, linguistic knowledge.

24

Demo

- Many constraints for meaning
  We use probability to combine
- Full language understanding is a long ways off
- But it is surprising how far simple methods can get you