Lecture 2:
Introduction, continued

CS 490A, Fall 2020

Applications of Natural Language Processing

https://people.cs.umass.edu/~brenocon/cs490a_f20/

Brendan O'Connor

College of Information and Computer Sciences

University of Massachusetts Amherst

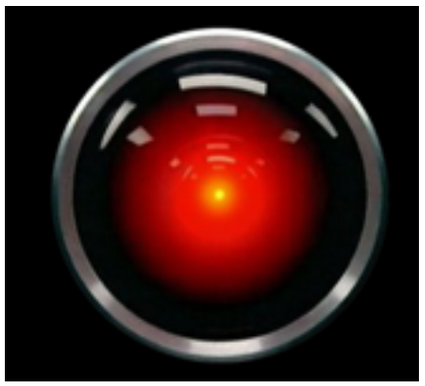
- HW0 posted, due 9/4. Probability & linear algebra basics
- "Quiz 0" survey will be posted today - please finish over weekend
- Office hours start next week

- Today:
  - NLP at a very high level
  - [Ran out of time]
    Probability and Linear Algebra review

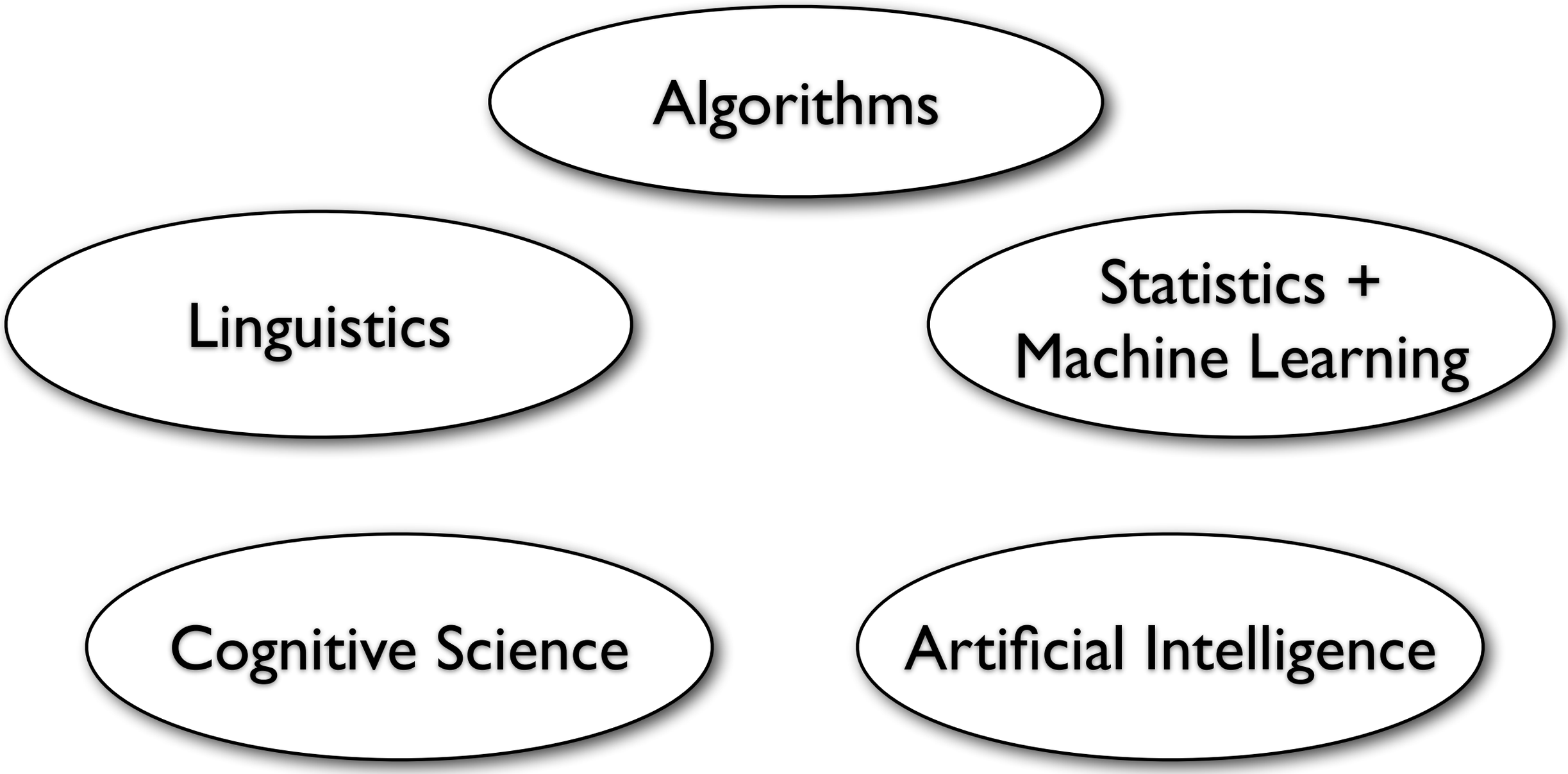
“Can Machines Think?”

- British mathematician and founding figure in computer science
- Alan Turing (1950)
- How do we know when we have AI?
- “Imitation Game”

NLP imagined



NLP is interdisciplinary



NLP today

- Language technologies
  - Speech interfaces
  - Machine translation
  - Sentiment analysis
  - Search engines
  - Relevance/recommendation ranking
  - …
- On what devices?
- For what users?

- [This course: document text analysis]

Breakout activity

- For 10 minutes in breakout rooms:
- 1. Introduce yourselves
- 2. Brainstorm 3 different NLP applications or technologies, including one that you use in your daily life
- 3. Nominate a person to report your group's ideas the whole class

NLP today: Speech interfaces

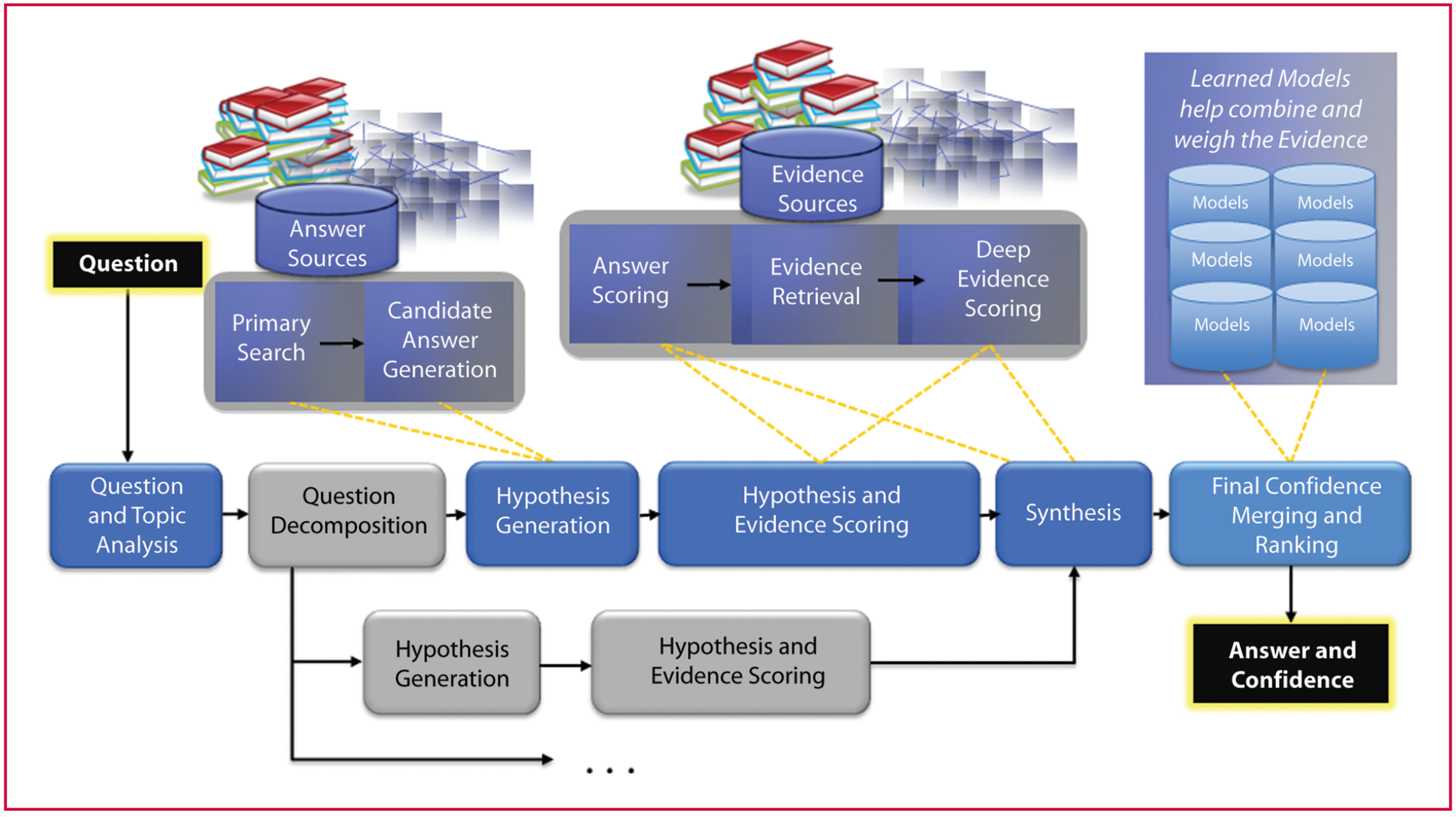
NLP today: Question answering

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 IBM Journal of Research and Development, 2012



IBM Watson

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

NLP today: Question answering

From IBM Journal of Research and Development, 2012

Table 1 DeepQA technology performance on public benchmark sets. (ACE: automatic content extraction; RTE: recognizing textual entailment.)

NLP task	*Evaluation set*	*Project start*	*State of art*	*Watson*
Parsing	Wikipedia** accuracy	84.4	81.1 Charniak parser [19]	88.7
Entity disambiguation	Wikipedia disambiguation F_1	72.5	81.9 Hoffart et al. [42]	92.5
Relation detection	ACE 2004 F_1	45.8	72.1 Zhang et al. [43]	73.2
Textual entailment	RTE-6 2010 F_1	34.6	48.0 PKUTM [44]	48.8

IBM Watson

Imperfect NLP is still useful

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.

Levels of linguistic structure

Discourse

Semantics

Syntax: Constituents

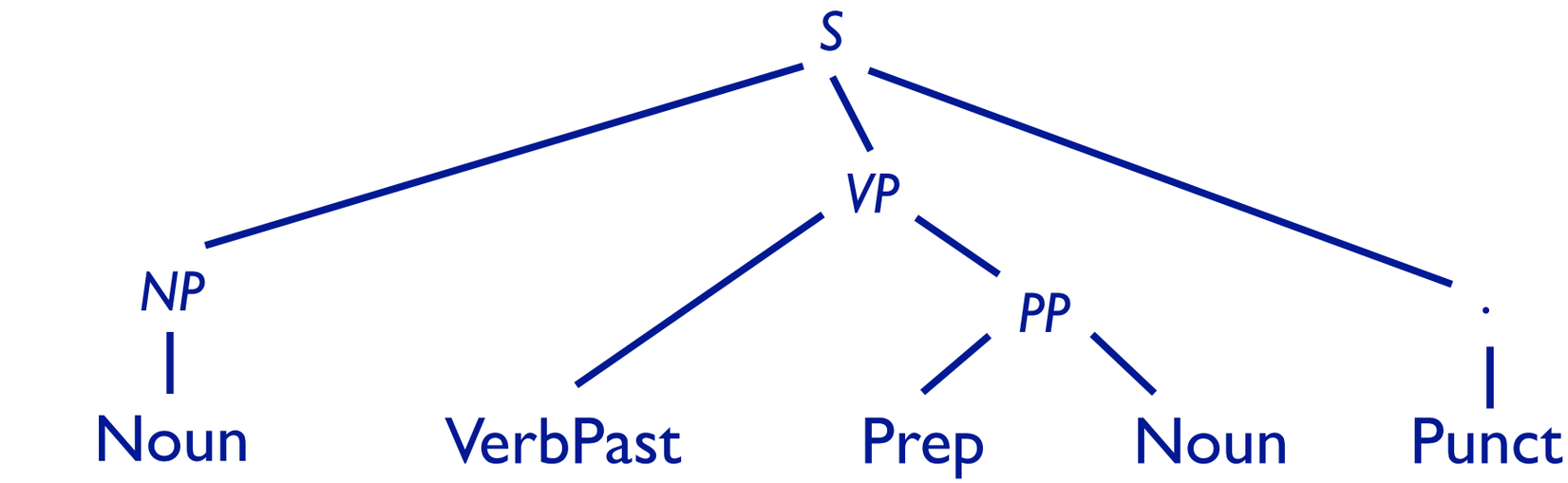
Syntax: Part of Speech

Words

Morphology

Characters

CommunicationEvent(e)
Agent(e, Alice)
Recipient(e, Bob)
SpeakerContext(s)
TemporalBefore(e, s)

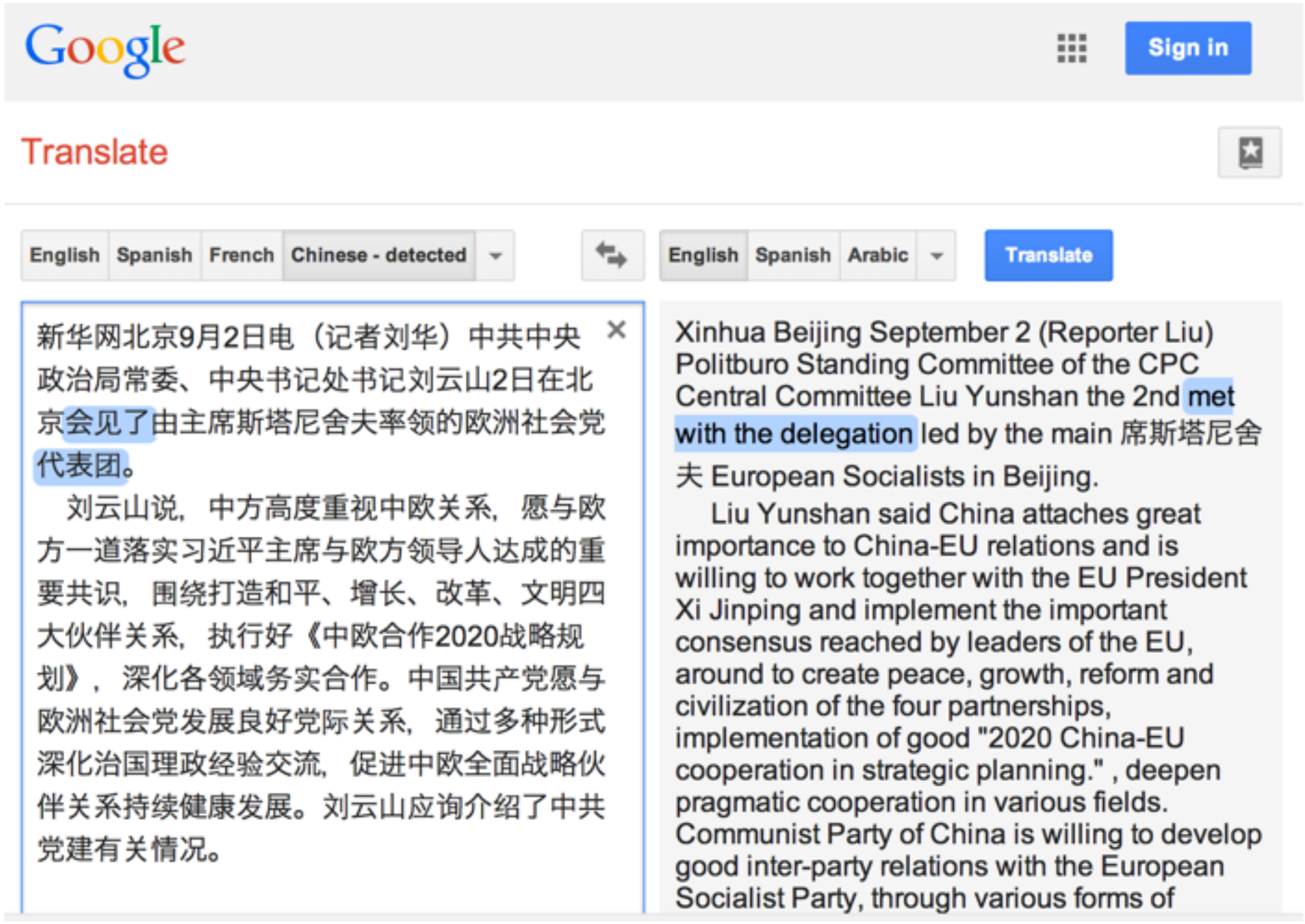


Alice talked to Bob .

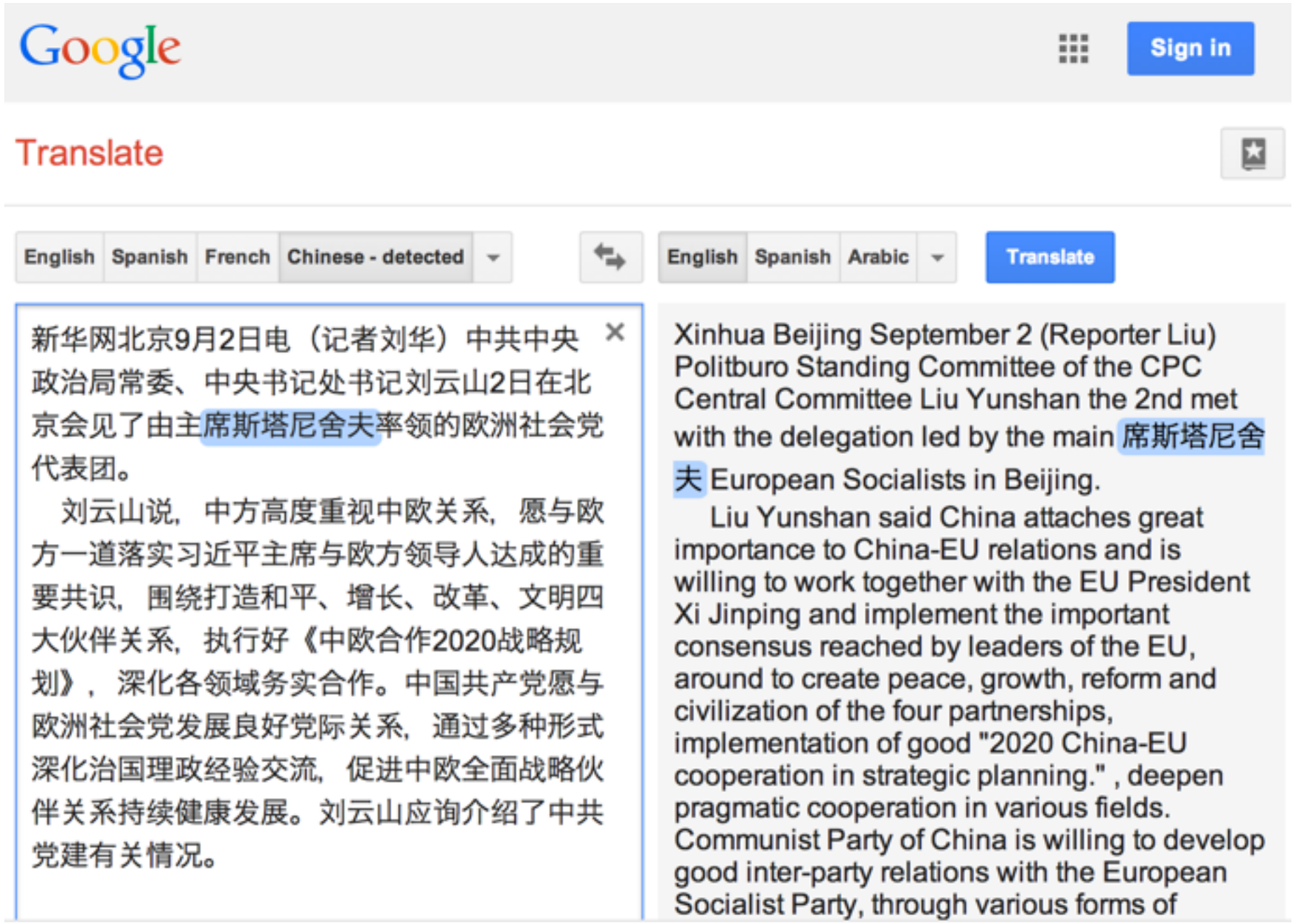
talk -ed [VerbPast]

Alice talked to Bob.

NLP today: Machine translation



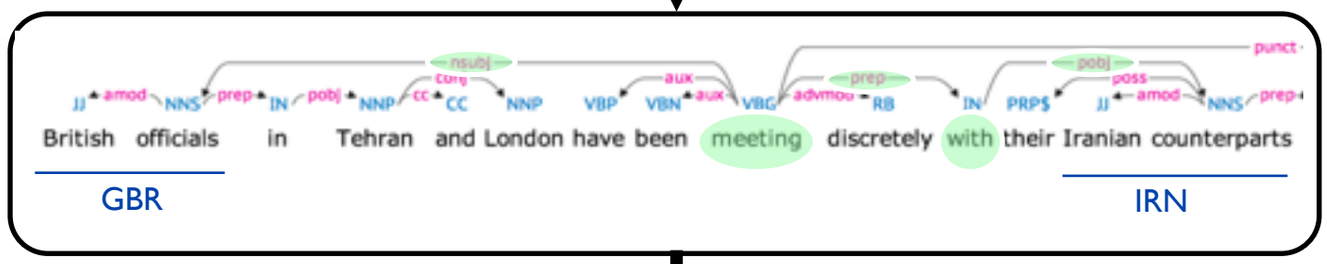
NLP today: Machine translation



NLP today: Trend analysis

Data: news articles

Dependency parsing to identify events

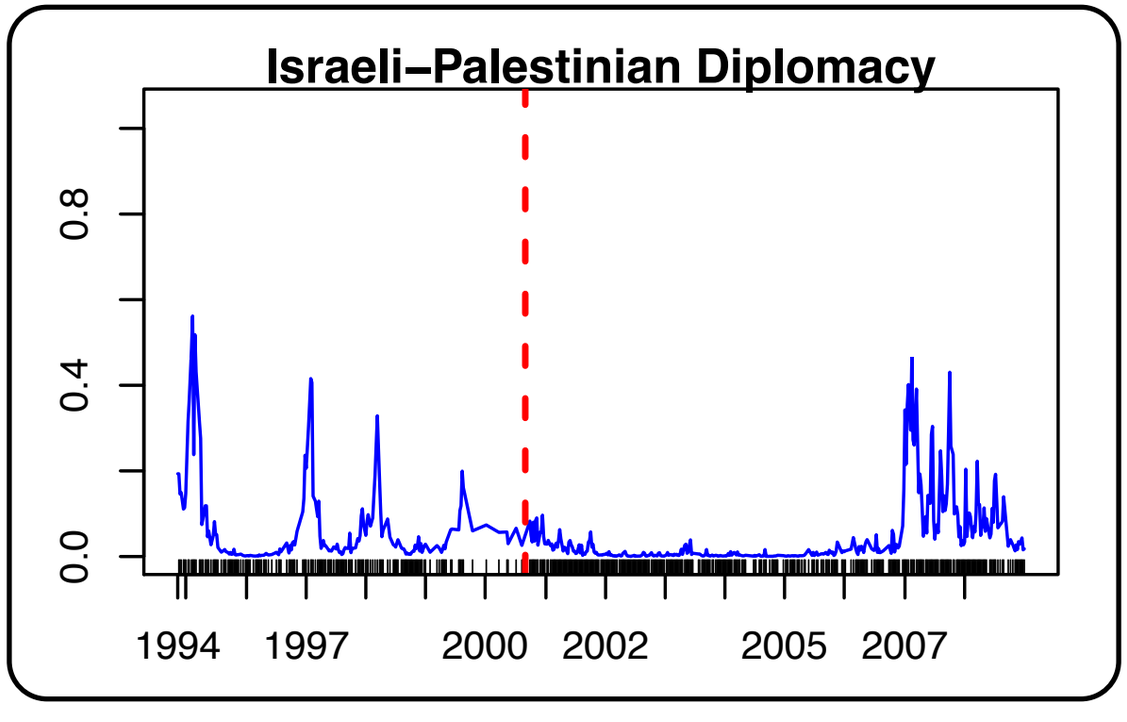


Machine learning from text:

(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
"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



NLP today

INVESTING 8/28/2014 @ 10:00AM | 121 views

Earnings for OmniVision Technologies Expected to Fall

By Narrative Science

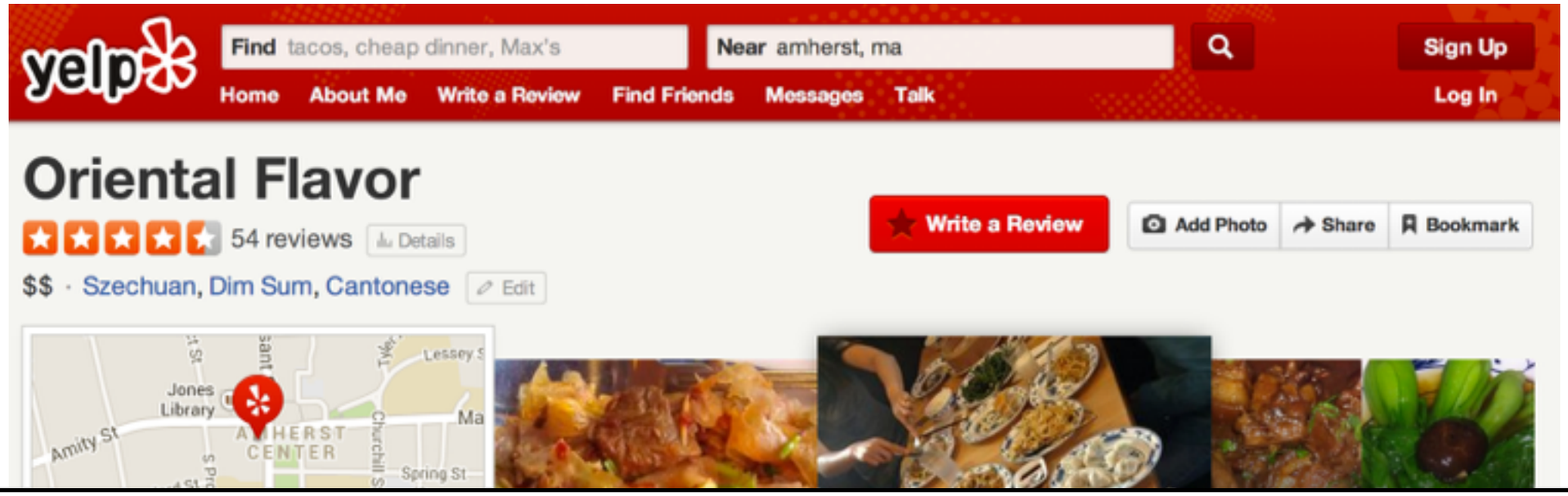
+ Comment Now + Follow Comments

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.

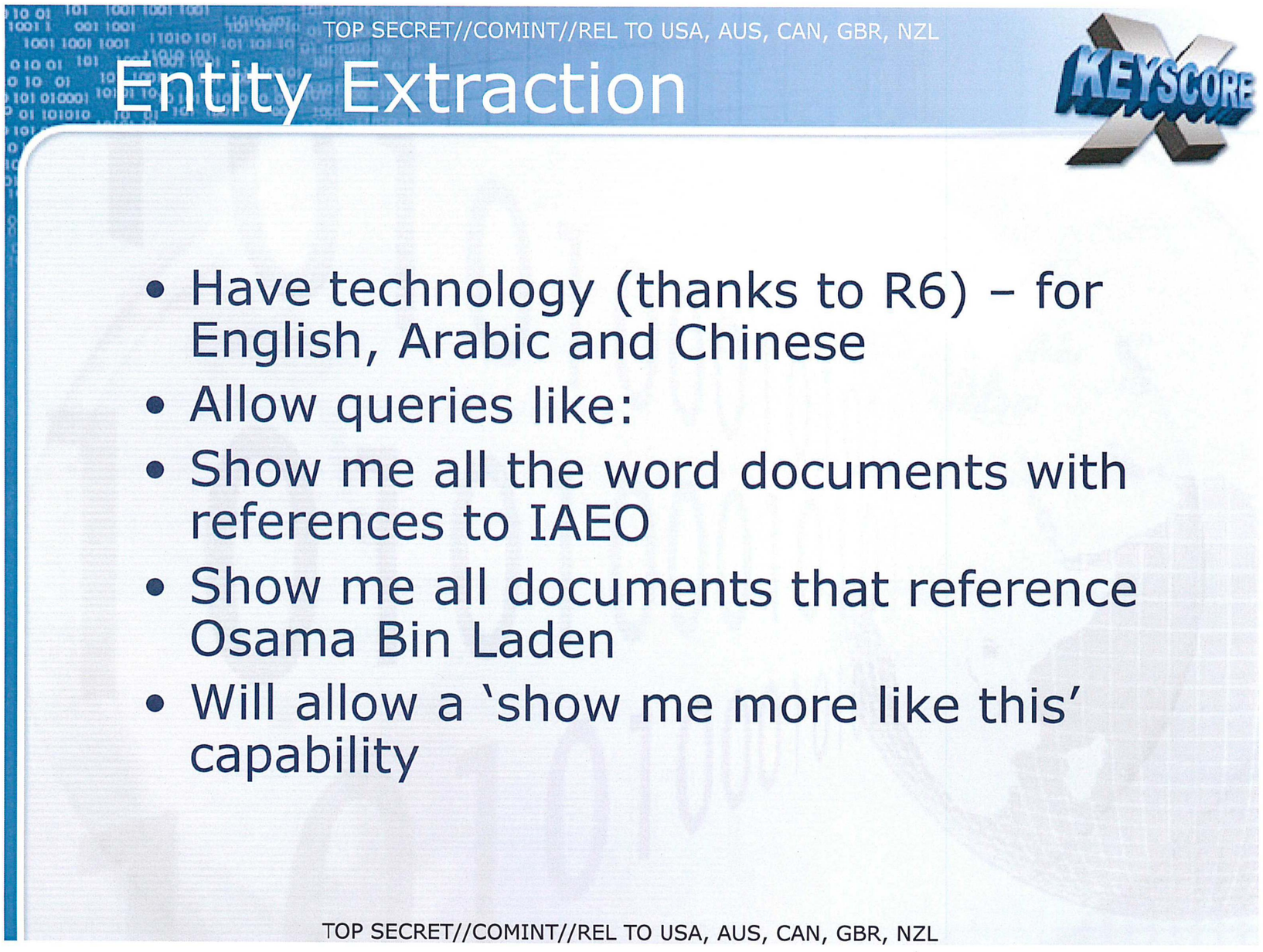
http://www.forbes.com/sites/narrativescience/

NLP today: Search/summarization



"We had the **crystal shrimp dumplings** that is seen wowing all yelpers, and they were superb." in 5 reviews

"In addition to having the best pork buns in the area, they also have the best **scallion pancakes** and xiaolongbao." in 8 reviews

"Usually I eat at the large dim sum restaurants in Boston with the rolling **carts**, but this experience was as good as any." in 5 reviews

NLP today: Search/summarization