

Lecture 1: Course Introduction

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

What

- Learn fundamental principles in natural language processing, and how to practically use them
- Hands-on experience: data collection and implementation
- Appreciation of basic linguistic issues
- Know when NLP works and when it doesn't
- “AI systems”

How

- Math!
- Data!
- Code!
 - Skill: translating from math to code
 - Skill: debugging math/linguistic/algorithm code
- A little bit of linguistics goes a long way

Prerequisites

- (CS220 and CS240) or (Ling 492B)
- Comfort with programming, algorithmic thinking
 - CS 220
- Comfort with probability and mathematical notation
 - CS 240
- Excitement about language!
- Willingness to learn

- “This is a HARD class”
- “The language parts are VERY INTERESTING.
The math is next to impossible”
- “The class is moving very slowly, pace can be increased”

Requirements

- (10%) Quizzes / exercises (approx. weekly)
- (30%) Problem sets
 - Written: math and concepts
 - Programs: in Python
- (20%) Midterm
- (40%) Final projects (groups of 2-3)
[Choose a topic, or select a suggested topic]
 - Project Proposal
 - Progress Report
 - In-class presentations
 - Final Report

Logistics

- Main course website
(will be linked from Piazza Resources page):
[https://people.cs.umass.edu/~brenocon/
cs490a_f20/](https://people.cs.umass.edu/~brenocon/cs490a_f20/)
- Email me and the TAs via
@googlegroups.com email (see piazza)
- Piazza for announcements & discussions
- Gradescope/Moodle for homework submissions

Readings

- Readings will be provided as PDFs on website
- Often draft chapters from Jurafsky and Martin, *Speech and Language Processing*

Related courses at UMass

- http://people.cs.umass.edu/~brenocon/complang_at_umass/

- stopped here 8/25

NLP is interdisciplinary

Algorithms

Linguistics

Statistics +
Machine Learning

Cognitive Science

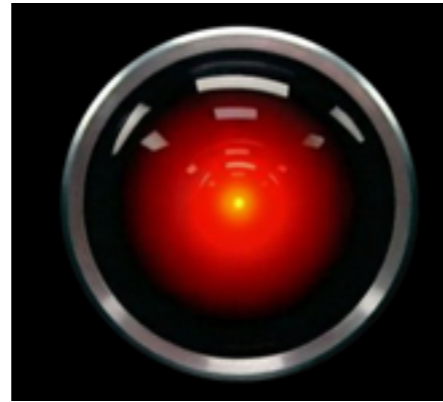
Artificial Intelligence

“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 today

- Speech interfaces
- Machine translation
- Sentiment analysis
- Search engines
- ...
- [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 1 that you use in your daily life
- 3. Nominate a person to report to the whole class

NLP today: Speech interfaces



NLP today: Question answering

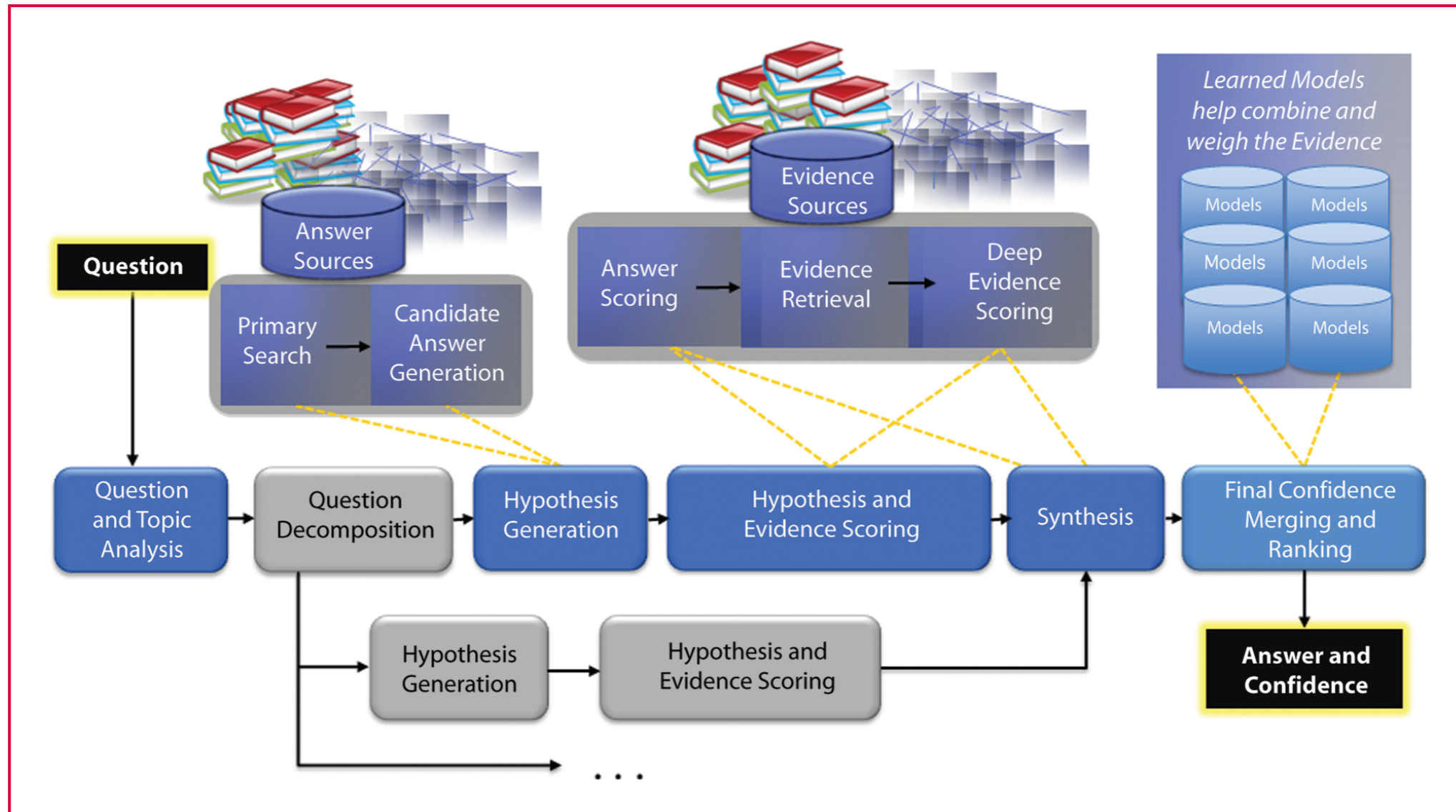


IBM Watson

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.)

<i>NLP task</i>	<i>Evaluation set</i>	<i>Project start</i>	<i>State of art</i>	<i>Watson</i>
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

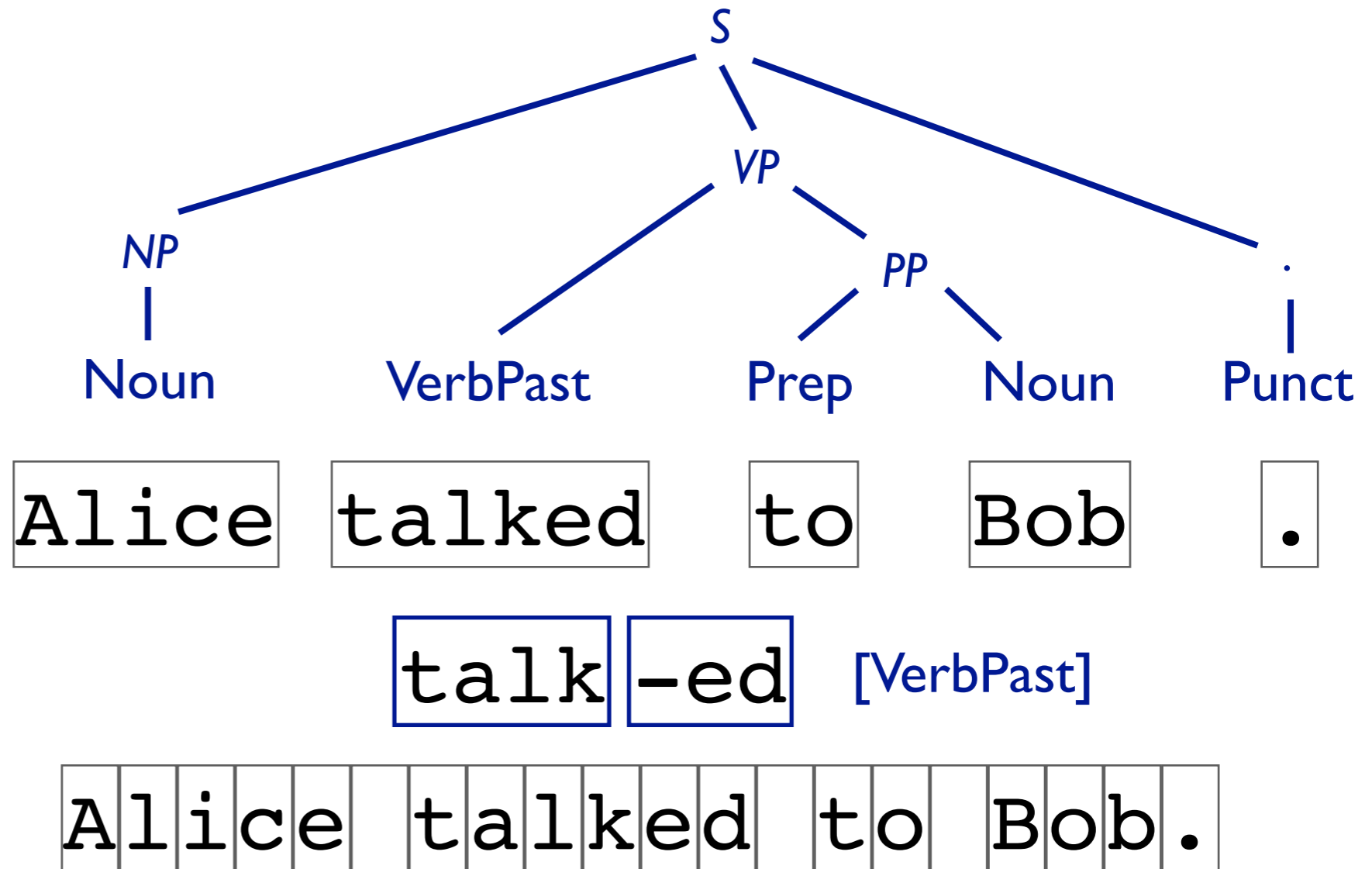
Syntax: Part of Speech

Words

Morphology

Characters

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



NLP today: Machine translation

The screenshot shows the Google Translate interface. At the top, the Google logo is on the left, and a 'Sign in' button is on the right. Below the logo, the word 'Translate' is displayed in red. The interface features two language selection menus: the left one is set to 'Chinese - detected' and the right one is set to 'English'. A blue 'Translate' button is positioned between the two menus. Below the menus, there are two columns of text. The left column contains the original Chinese text, and the right column contains the machine-translated English text. The Chinese text is a news report from Xinhua dated September 2, 2013, reporting on a meeting between Liu Yunshan, a member of the Politburo Standing Committee of the CPC, and a delegation of the European Socialist Party in Beijing. The English translation follows the same structure, detailing the meeting and Liu Yunshan's statements on China-EU relations.

English Spanish French Chinese - detected ↕ English Spanish Arabic Translate

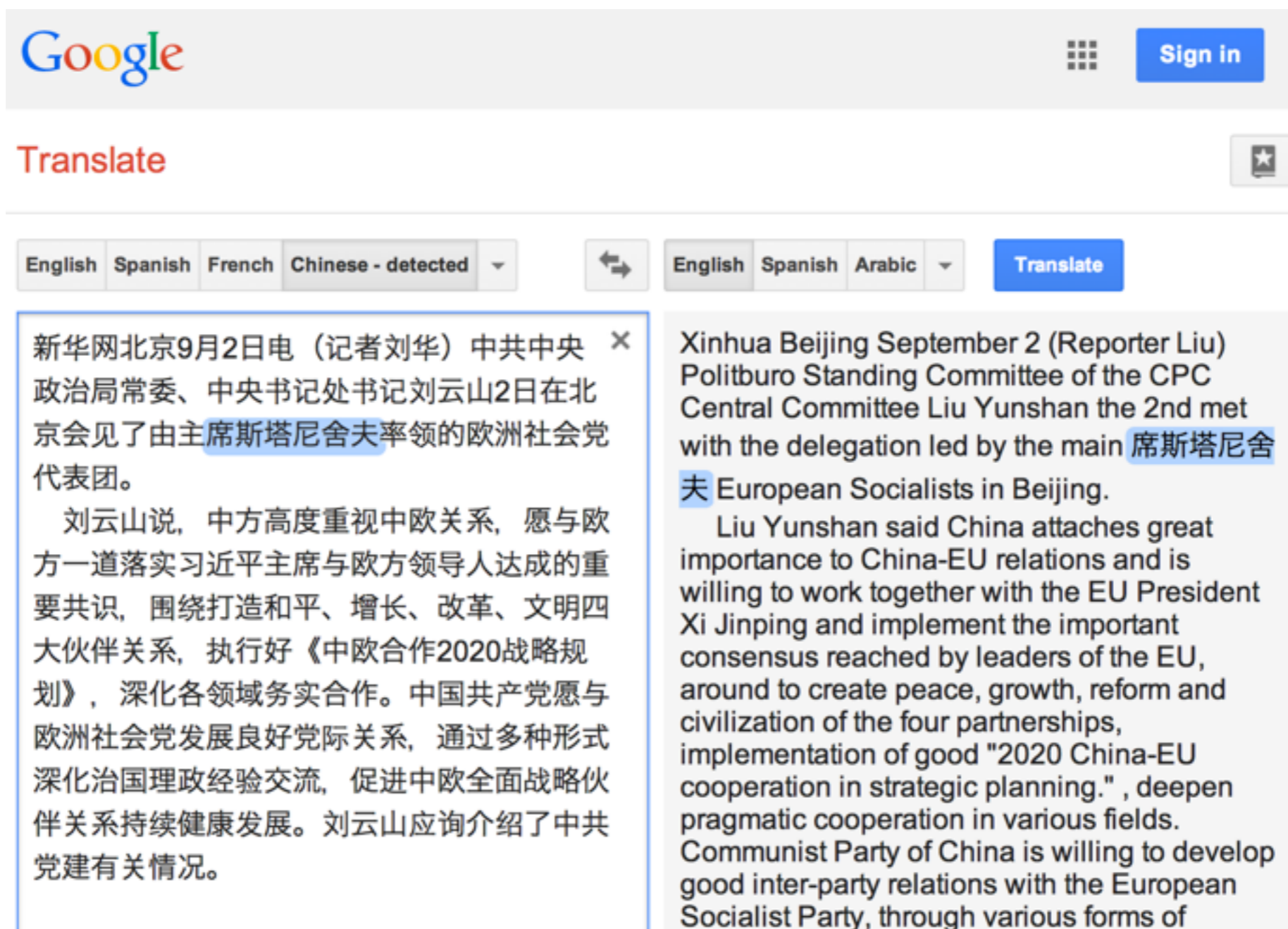
新华网北京9月2日电（记者刘华）中共中央政治局常委、中央书记处书记刘云山2日在北京会见了由主席斯塔尼舍夫率领的欧洲社会党代表团。

刘云山说，中方高度重视中欧关系，愿与欧方一道落实习近平主席与欧方领导人达成的重要共识，围绕打造和平、增长、改革、文明四大伙伴关系，执行好《中欧合作2020战略规划》，深化各领域务实合作。中国共产党愿与欧洲社会党发展良好党际关系，通过多种形式深化治国理政经验交流，促进中欧全面战略伙伴关系持续健康发展。刘云山应询介绍了中共党建有关情况。

Xinhua Beijing September 2 (Reporter Liu) Politburo Standing Committee of the CPC Central Committee Liu Yunshan the 2nd met with the delegation led by the main 席斯塔尼舍夫 European Socialists in Beijing.

Liu Yunshan said China attaches great importance to China-EU relations and is willing to work together with the EU President Xi Jinping and implement the important consensus reached by leaders of the EU, around to create peace, growth, reform and civilization of the four partnerships, implementation of good "2020 China-EU cooperation in strategic planning.", deepen pragmatic cooperation in various fields. Communist Party of China is willing to develop good inter-party relations with the European Socialist Party, through various forms of

NLP today: Machine translation



The screenshot shows the Google Translate web interface. At the top left is the Google logo, and at the top right is a 'Sign in' button. Below the logo is the word 'Translate' in red. The interface has two language selection menus. The left menu shows 'English', 'Spanish', 'French', and 'Chinese - detected'. The right menu shows 'English', 'Spanish', and 'Arabic'. A blue 'Translate' button is positioned between the two menus. Below the menus, there are two text boxes. The left box contains the original Chinese text, and the right box contains the translated English text. The Chinese text is a news article from Xinhua dated September 2, 2014, reporting on a meeting between Liu Yunshan, a member of the Politburo Standing Committee of the CPC, and a delegation of European Socialists led by the main spokesman. The English translation follows the same structure, detailing the meeting and Liu Yunshan's statements on China-EU relations.

English Spanish French Chinese - detected

English Spanish Arabic Translate

新华网北京9月2日电（记者刘华）中共中央政治局常委、中央书记处书记刘云山2日在北京会见了由主席斯塔尼舍夫率领的欧洲社会党代表团。

刘云山说，中方高度重视中欧关系，愿与欧方一道落实习近平主席与欧方领导人达成的重要共识，围绕打造和平、增长、改革、文明四大伙伴关系，执行好《中欧合作2020战略规划》，深化各领域务实合作。中国共产党愿与欧洲社会党发展良好党际关系，通过多种形式深化治国理政经验交流，促进中欧全面战略伙伴关系持续健康发展。刘云山应询介绍了中共党建有关情况。

Xinhua Beijing September 2 (Reporter Liu) Politburo Standing Committee of the CPC Central Committee Liu Yunshan the 2nd met with the delegation led by the main 席斯塔尼舍夫 European Socialists in Beijing.

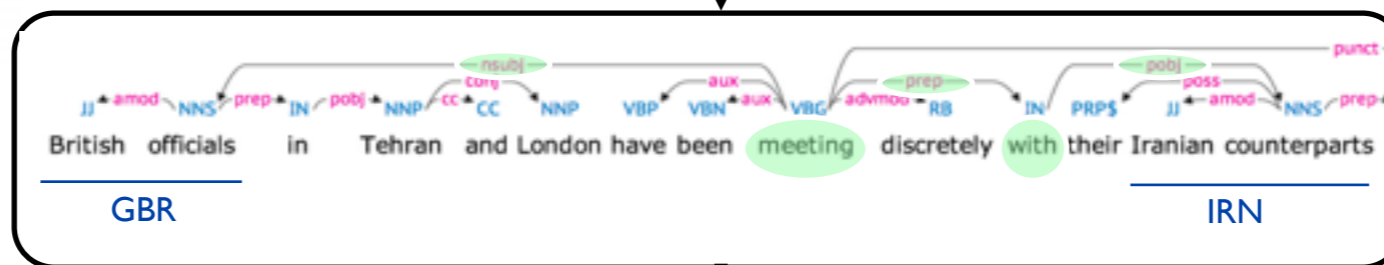
Liu Yunshan said China attaches great importance to China-EU relations and is willing to work together with the EU President Xi Jinping and implement the important consensus reached by leaders of the EU, around to create peace, growth, reform and civilization of the four partnerships, implementation of good "2020 China-EU cooperation in strategic planning.", deepen pragmatic cooperation in various fields. Communist Party of China is willing to develop good inter-party relations with the European Socialist Party, through various forms of

NLP today: Trend analysis



Data: news articles

Dependency parsing to identify events



Machine learning from text:

(1) Event class dictionaries

(2) Political dynamics

“diplomacy”

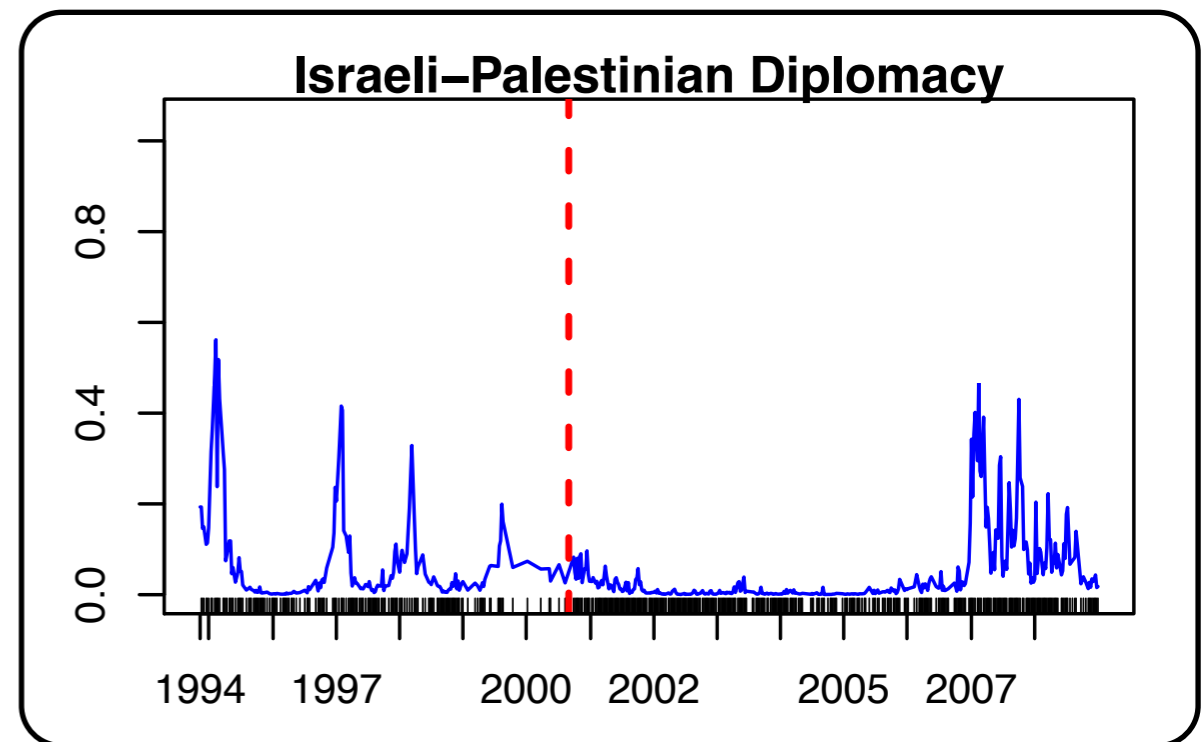
arrive in, visit, meet with, travel to, leave, hold with, meet, meet in, fly to, be in, arrive for talk with, say in, arrive with, head to, hold in, due in, leave for, make to, arrive to,

“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



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

The screenshot shows the Yelp interface for 'Oriental Flavor'. At the top, there's a search bar with 'Find tacos, cheap dinner, Max's' and 'Near amherst, ma'. Below the search bar are navigation links: Home, About Me, Write a Review, Find Friends, Messages, Talk, Sign Up, and Log In. The restaurant name 'Oriental Flavor' is prominently displayed, followed by a 4.5-star rating (5 stars) and '54 reviews'. There are buttons for 'Write a Review', 'Add Photo', 'Share', and 'Bookmark'. The price range is '\$\$' and the cuisine is 'Szechuan, Dim Sum, Cantonese'. A map shows the location in Amherst, MA, near the Amherst Center. Below the map are several photos of food items, including dumplings and other dishes.



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

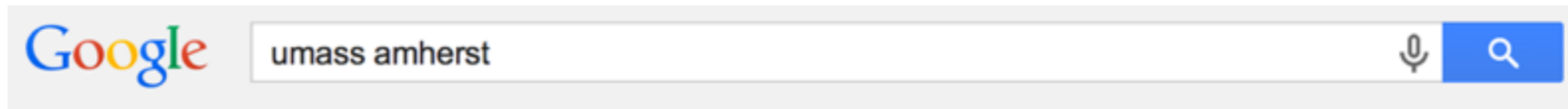


“In addition to having the best pork buns in the area, they also have the best **scallion pancakes** and xiaolongbao.” in 8 reviews



“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



Web **News** Images Maps Videos More ▾ Search tools

About 8,510 results (0.42 seconds)



UMass Amherst breaks a world record

[wwlp.com](#) - 17 hours ago

AMHERST, Mass. (WWLP) – **UMass Amherst** has broken a world record by serving over 3,000 people a New England clambake in one and a ...

UMass Amherst starts semester with giant clambake

[SouthCoastToday.com](#) - 1 minute ago

UMass creates, sets new record for most clambake meals served ...

[The Republican - masslive.com](#) - 14 hours ago

UMass gets \$37.5 million for environmental projects

[The Recorder](#) - 14 hours ago

No cheers here for **UMass** football program

Opinion - [Boston Globe](#) - 10 hours ago



[wwlp.com](#)



[Boston Globe](#)



[The Massac...](#)



[WCVB Boston](#)

[Explore in depth](#) (49 more articles)



UMass football players struggle academically


[Boston Globe](#) - Aug 30, 2014

The **University of Massachusetts Amherst** football team has struggled not only on the field, but in the classroom as well, leaving it barely above ...

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

- This week we'll post HW0 - probability & linear algebra basics/review
- See you on Thursday