

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

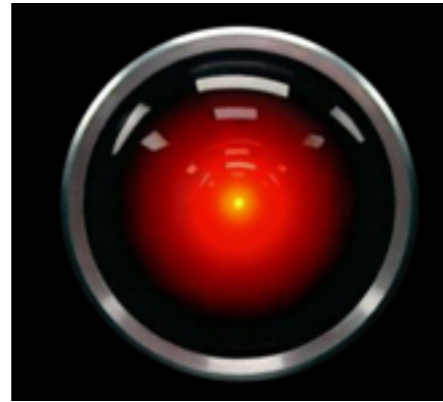
- HW0 posted, due 9/4. Probability & linear algebra basics
- “Quiz 0” survey will be posted today - please finish over weekend
- Office hours start next week
- Today:
 - NLP at a very high level
 - [Ran out of time]
Probability and Linear Algebra review

“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

Algorithms

Linguistics

Statistics +
Machine Learning

Cognitive Science

Artificial Intelligence

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

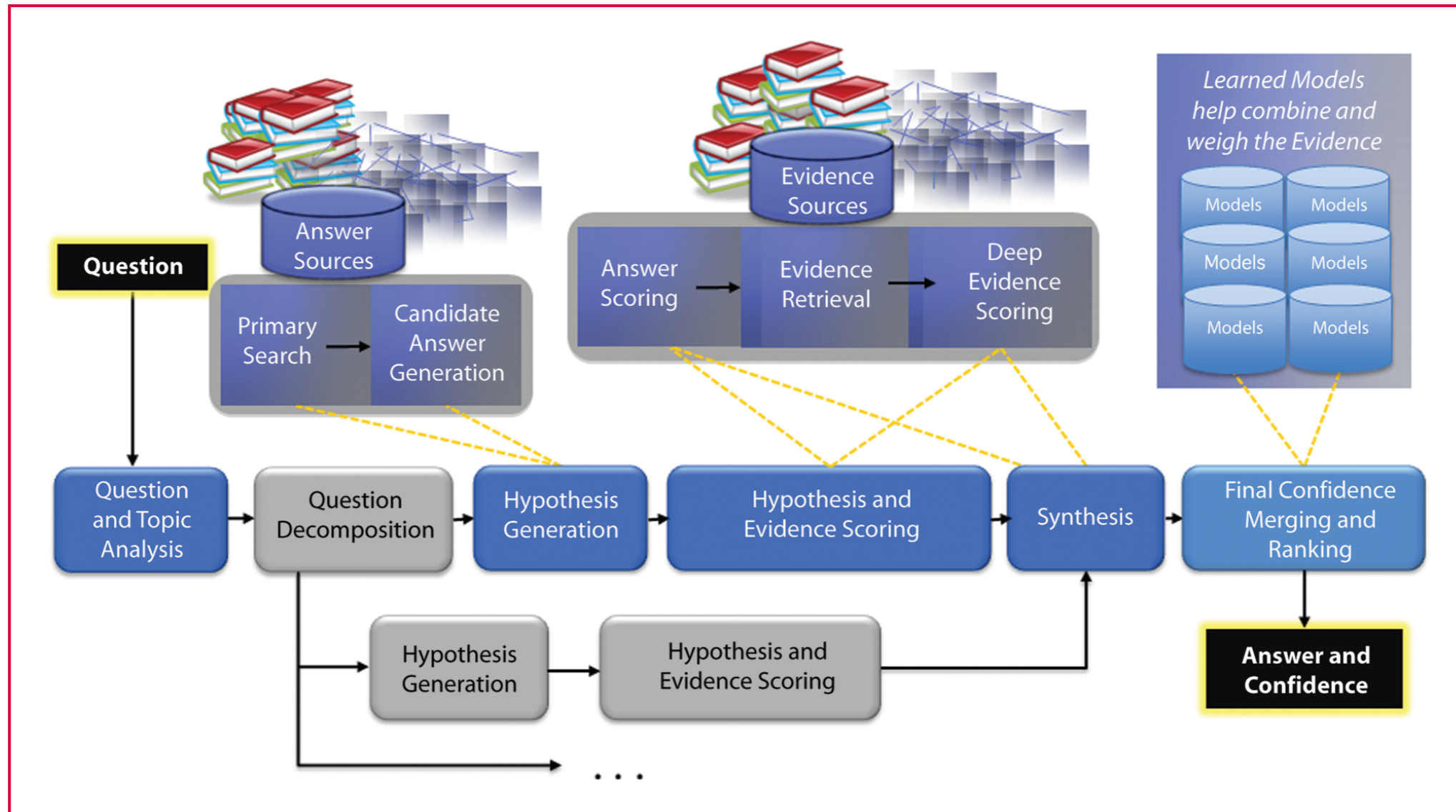


IBM Watson

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.)

<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

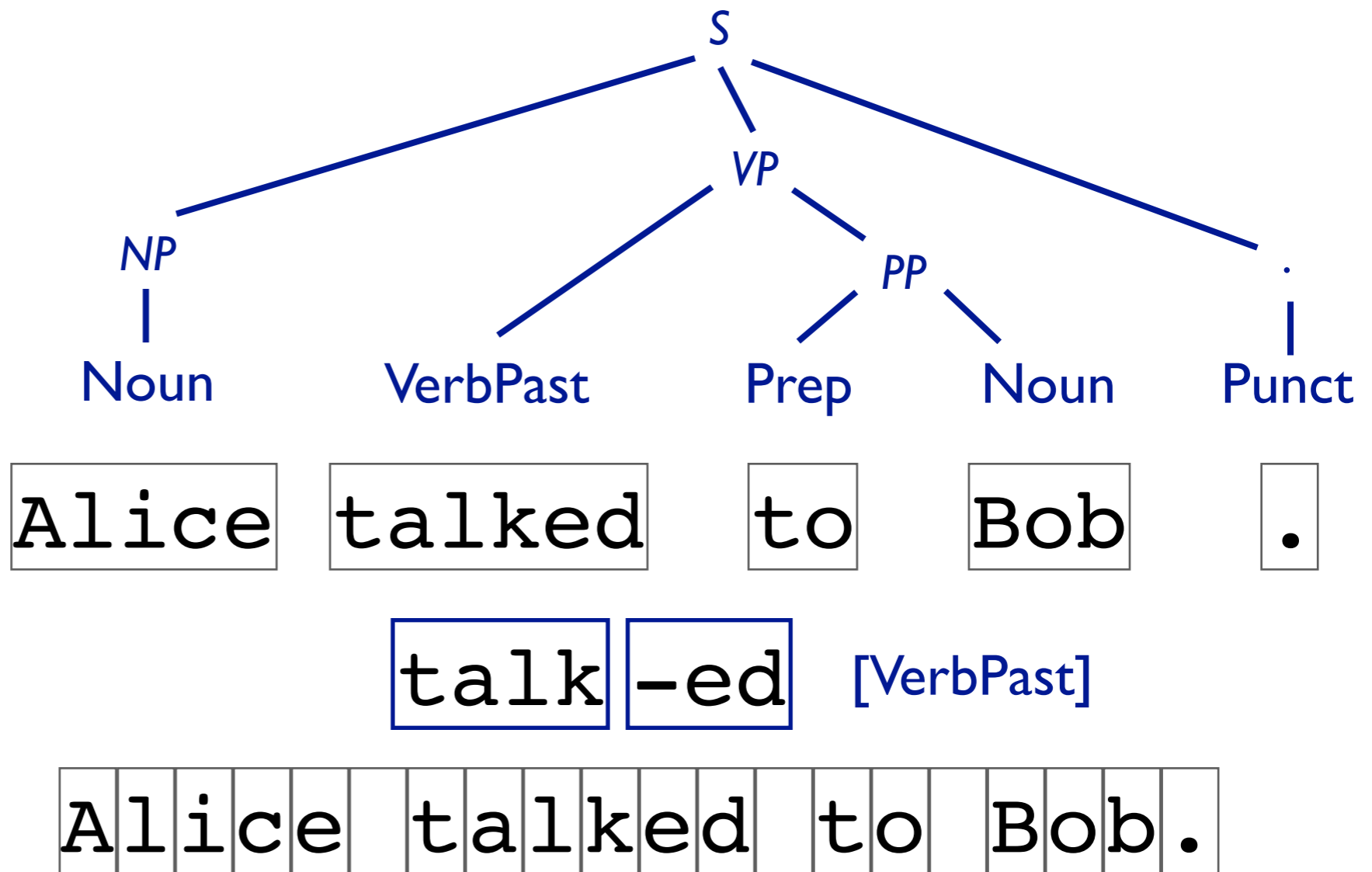
Syntax: Part of Speech

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 web 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, the original Chinese text is shown in a light blue box, and the translated English text is shown in a light grey box. 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 led by the main chairman 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.

Google

Sign in

Translate

English Spanish French Chinese - detected

English Spanish Arabic Translate

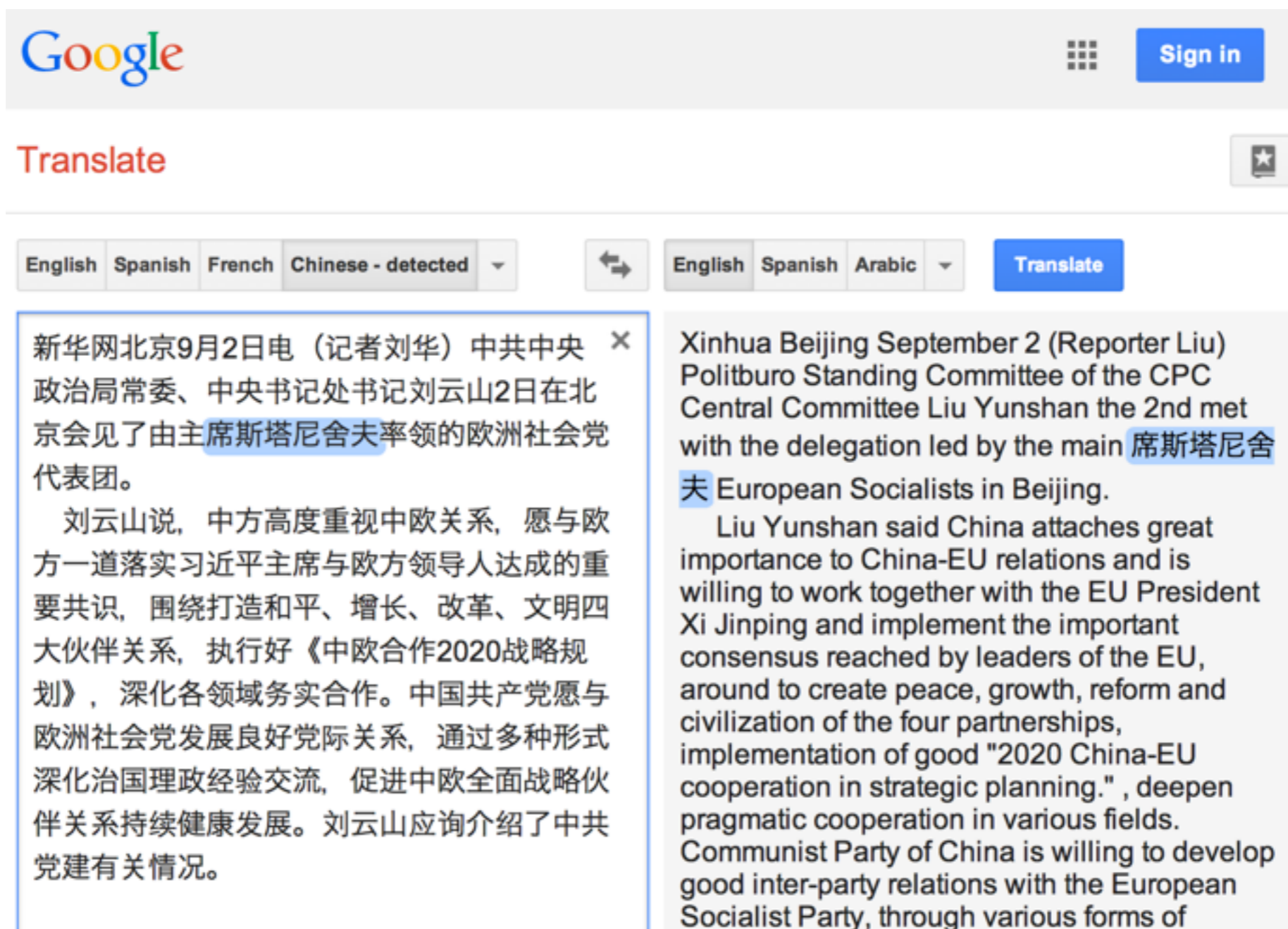
新华网北京9月2日电（记者刘华）中共中央政治局常委、中央书记处书记刘云山2日在北京会见了由主席斯塔尼舍夫率领的欧洲社会党代表团。

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 chairman of the European Socialist Party 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 features two language selection dropdown menus. The first menu is set to 'Chinese - detected' and the second to 'English'. A blue 'Translate' button is positioned between the two menus. Below the menus, the original Chinese text is displayed in a light blue box, and the translated English text is shown in a light grey box. The Chinese text is a news report 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 from the European Socialist Party led by Stanislav Mikheev. Liu Yunshan's remarks on Sino-EU relations are also included. The English translation on the right mirrors this content, mentioning the meeting with the delegation led by Stanislav Mikheev and Liu Yunshan's statement on the importance of 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.

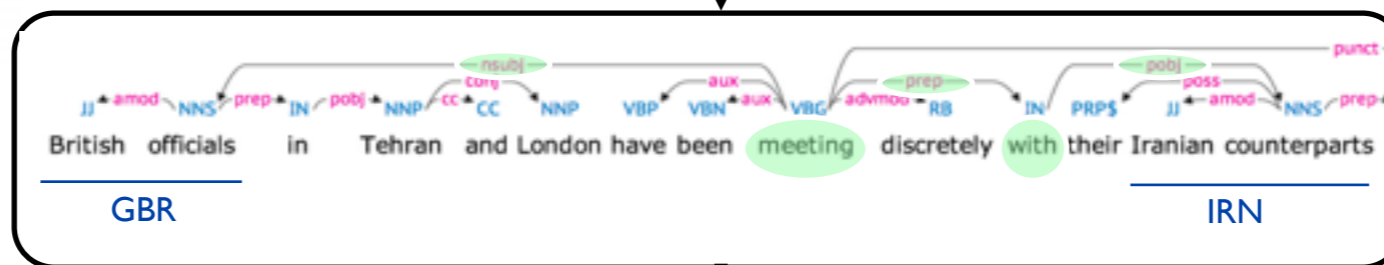
Liu Yunshan said China attaches great importance to China-EU relations and is willing to work together with the EU President Xi Jinping and implement the important consensus reached by leaders of the EU, around to create peace, growth, reform and civilization of the four partnerships, implementation 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”

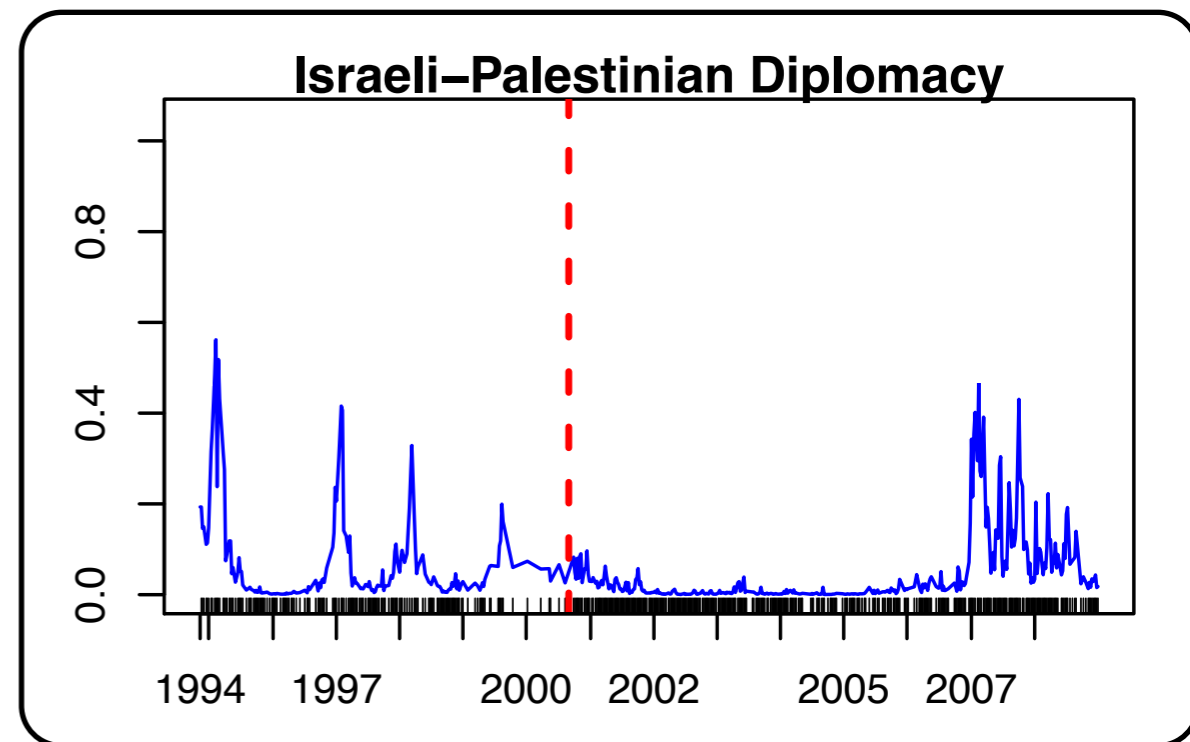
arrive in, visit, meet with, travel to, leave, hold with, meet, meet in, fly to, be in, arrive for talk with, say in, arrive with, head to, hold in, due in, leave for, make to, arrive to,

“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, and Talk. On the right, there are 'Sign Up' and 'Log In' buttons. The restaurant name 'Oriental Flavor' is prominently displayed, followed by a 4.5-star rating (5 stars with the last one half-filled) and '54 reviews'. There are buttons for 'Write a Review', 'Add Photo', 'Share', and 'Bookmark'. Below the rating, it says '\$\$ · Szechuan, Dim Sum, Cantonese' with an 'Edit' button. A map shows the location in Amherst, MA, near the Amherst Center. Three food photos are shown: a close-up of a dish, a table with several plates of food, and a dish with green vegetables.



“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

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

