Natural Language Processing

CT052-3-M-NLP

Introduction to NLP

ASIA PACIFIC UNIVERSITY
OF TECHNOLOGY & INNOVATION

LEARNING OUTCOMES

At the end of this topic, You should be able to:

Explain the meaning and fundamental concepts of Natural Language Processing

QUESTION ANSWERING: IBM'S WATSON

- Won Jeopardy on February 16, 2011!

INFORMATION EXTRACTION

Subject: **Programme Committee Meeting**

Date: October 20, 2016

To: Rad

Hi Rad, we've now scheduled the curriculum meeting.

It will be in TPM Main Building, L4, Board room, tomorrow from 10:00-11:30.

- Jacqueline

Create new Calendar entry

INFORMATION EXTRACTION & SENTIMENT ANALYSIS

Attributes:

- zoom
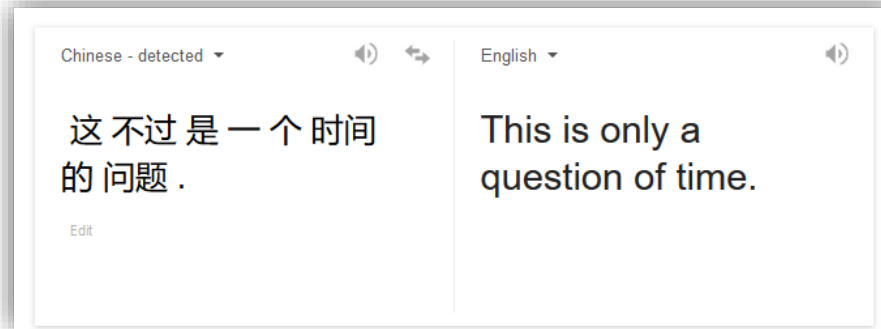- affordability
- size and weight
- flash
- ease of use
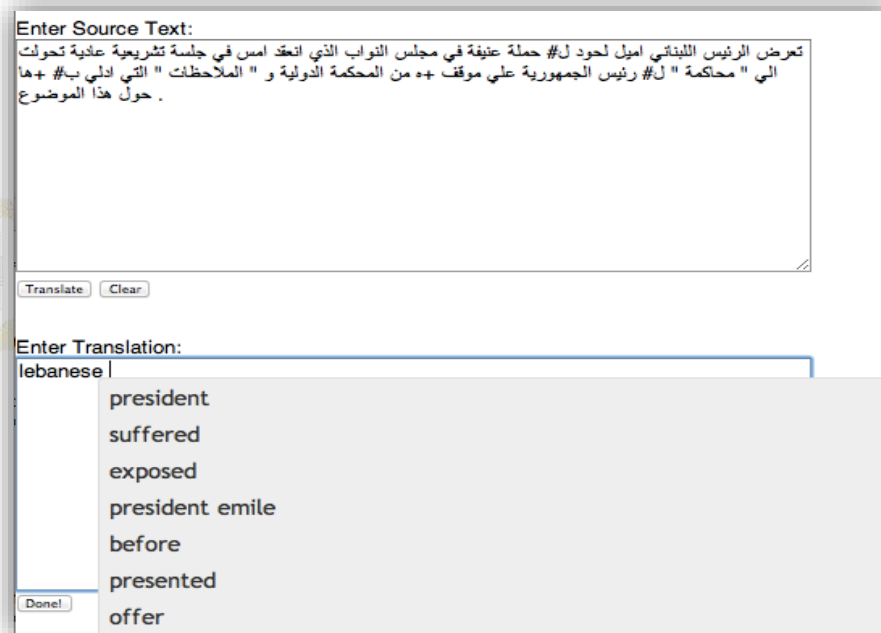
☑ nice and compact to carry!

☑ since the camera is small and light, I won't need to carry around those heavy, bulky professional cameras either!

☑ the camera feels flimsy, is plastic and very light in weight you have to be very delicate in the handling of this camera

MACHINE TRANSLATION

- Fully Automatic

- Helping human translators

LANGUAGE TECHNOLOGY

mostly solved

Spam detection
Let's go to Agra! ✓
Buy V1AGRA … ✗

Part-of-speech (POS) tagging
ADJ   ADJ NOUN VERB   ADV
Colorless green ideas sleep furiously.

Named entity recognition (NER)
PERSON   ORG   LOC
Einstein met with UN officials in Princeton

making good progress

Sentiment analysis
Best roast chicken in San Francisco! 👍
The waiter ignored us for 20 minutes. 👎

Coreference resolution
Carter told Mubarak he shouldn't run again.

Word sense disambiguation (WSD)
I need new batteries for my *mouse*.

Parsing
I can see Alcatraz from the window!

Machine translation (MT)
第13届上海国际电影节开幕…
The 13th Shanghai International Film Festival…

Information extraction (IE)
You're invited to our dinner party, Friday May 27 at 8:30
Party May 27
add

still really hard

Question answering (QA)
Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

Paraphrase
XYZ acquired ABC yesterday
ABC has been taken over by XYZ

Summarization
The Dow Jones is up
The S&P500 jumped
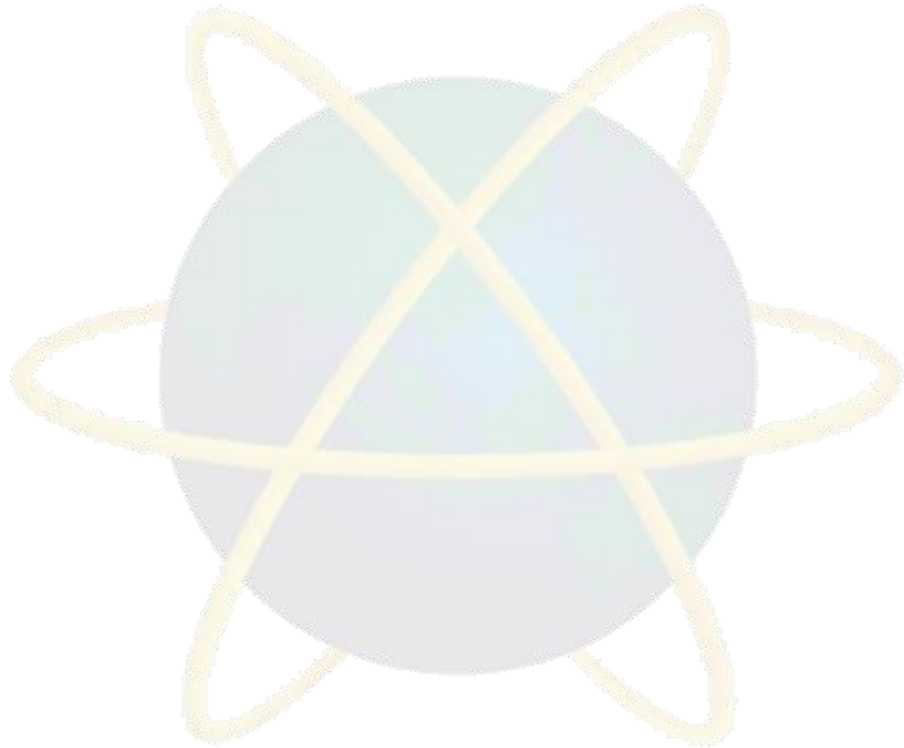Housing prices rose
→ Economy is good

Dialog
Where is Citizen Kane playing in SF?
Castro Theatre at 7:30. Do you want a ticket?

DIFFICULTIES IN NLP

AMBIGUITY ("CRASH BLOSSOMS")

- Violinist Linked to JAL Crash Blossoms

Violinist linked to JAL crash blossoms

By May Masangkay

The requested article has expired, and is no longer available. Any related articles, and user comments are shown below.
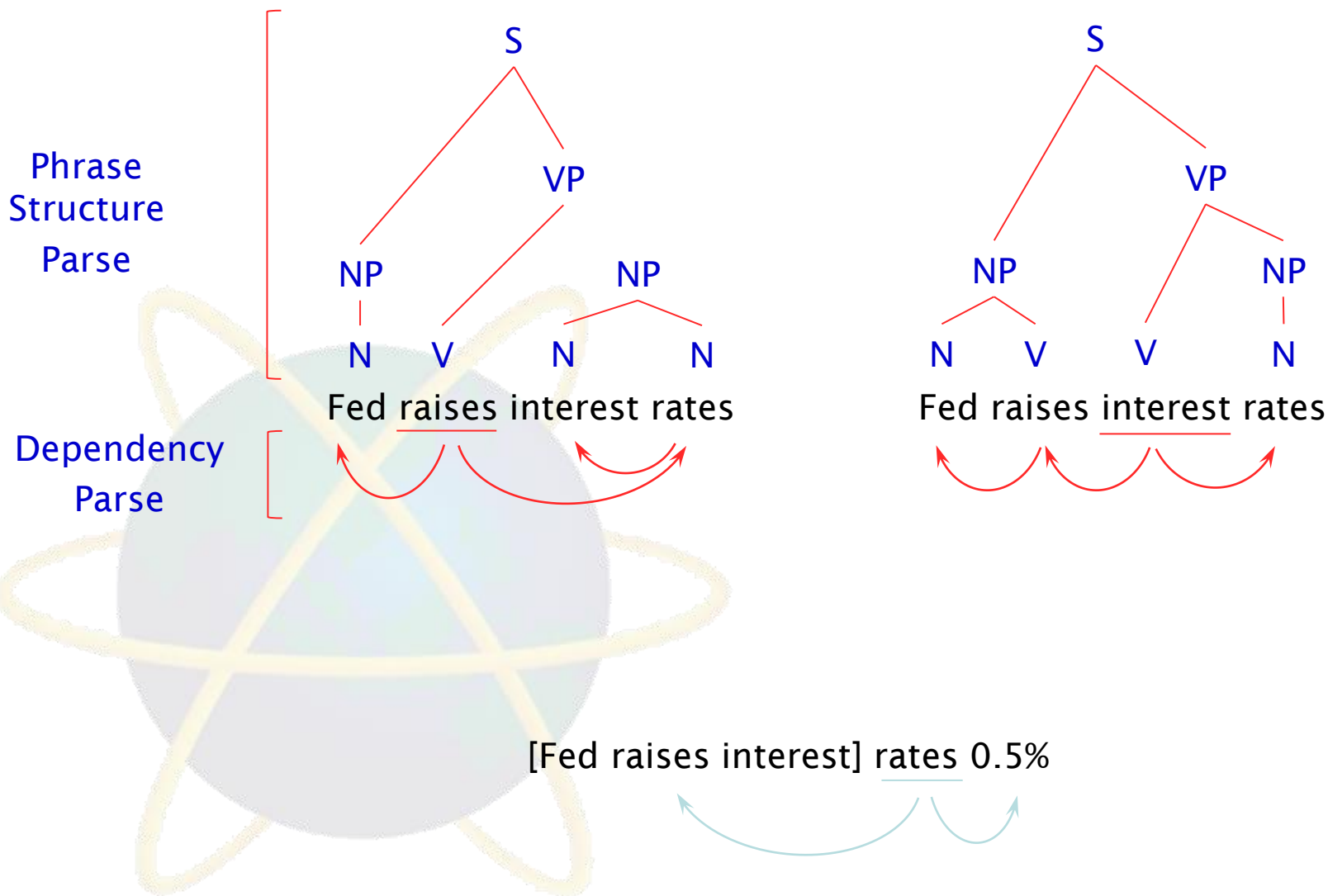
- Teacher Strikes Idle Kids

- Red Tape Holds Up New Bridges

- Hospitals Are Sued by 7 Foot Doctors

- Juvenile Court to Try Shooting Defendant

- Local High School Dropouts Cut in Half

AMBIGUITY IS PERVASIVE

New York Times headline (17 May 2000)

Phrase Structure Parse

Dependency Parse

S
VP
NP NP
N V N N
Fed raises interest rates

S
VP
NP NP
N V V N
Fed raises interest rates

[Fed raises interest] rates 0.5%

OTHER DIFFICULTIES

non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either♥

neologisms

unfriend
Retweet
bromance

idioms

dark horse
get cold feet
lose face
throw in the towel

segmentation issues

the New York-New Haven Railroad
the New York-New Haven Railroad

world knowledge

Mary and Sue are sisters.
Mary and Sue are mothers.

tricky entity names

Where is *A Bug's Life* playing …
Let It Be was recorded …
… a mutation on the *for* gene …

Neologism → a newly coined word or expression
Idioms → a group of words established by usage as having a meaning

WHAT TO DO?

- The task is difficult! What tools do we need?

  – Knowledge about language

  – Knowledge about the world

  – A way to combine knowledge sources

- How we generally do this:

  – probabilistic models built from language data

    - P("maison" $\rightarrow$ "house")   high

    - P("L'avocat général" $\rightarrow$ "the general avocado")   low

  – Luckily, rough text features can often do half the job.

Q & A

WHAT WE WILL COVER NEXT..!

Introduction to Basic Text Processing