COMP 472: Artificial Intelligence
Natural Language Processing
Introduction

part 5

video 1

- Russell & Norvig: Sections 23.5, 23.6

Today

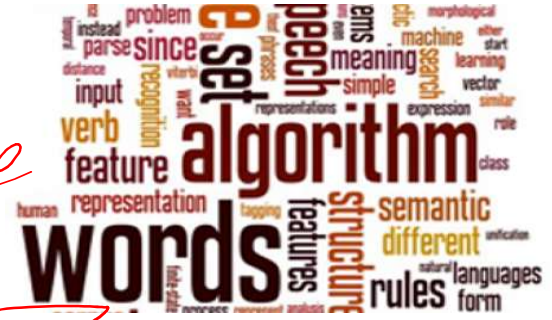
1. Introduction
2. Bag of word model
3. n-gram models
4. Deep Learning for NLP
    1. Word Embeddings
    2. Recurrent Neural Networks

NLP vs Speech Processing

- Natural Language Processing

  = automatic processing of written texts

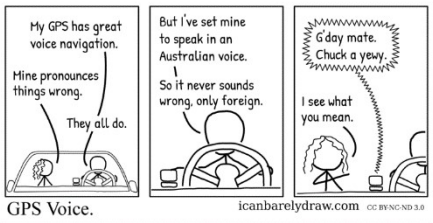
  1. Natural Language Understanding
     - Input = text
  2. Natural Language Generation
     - Output = text



- Speech Processing

  = automatic processing of speech

  1. Speech Recognition
     - Input = acoustic signal
  2. Speech Synthesis
     - Output = acoustic signal



Question Answering: IBM's Watson

WATSON VS. HUMANS			
Round	Watson	Rutter	Jennings
1 (Mon.)	$5000	$5000	$200
2 (Tues.)	$35,734	$10,800	$4,800
3 (Wed.)	$77,147	$21,600	$24,000
Final prize	$1,000,000	$200,000	$300,000

- Won Jeopardy on February 16, 2011!

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE PRINCIPALITIES OF WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL

→ Who is Bram Stoker? (Dracula)

Information Extraction

Subject: **curriculum meeting**

Date: January 15, 2012

To: Dan Jurafsky

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

It will be in Gates 159 tomorrow from 10:00-11:30.

-Chris

Create new Calendar entry

Event: Curriculum mtg

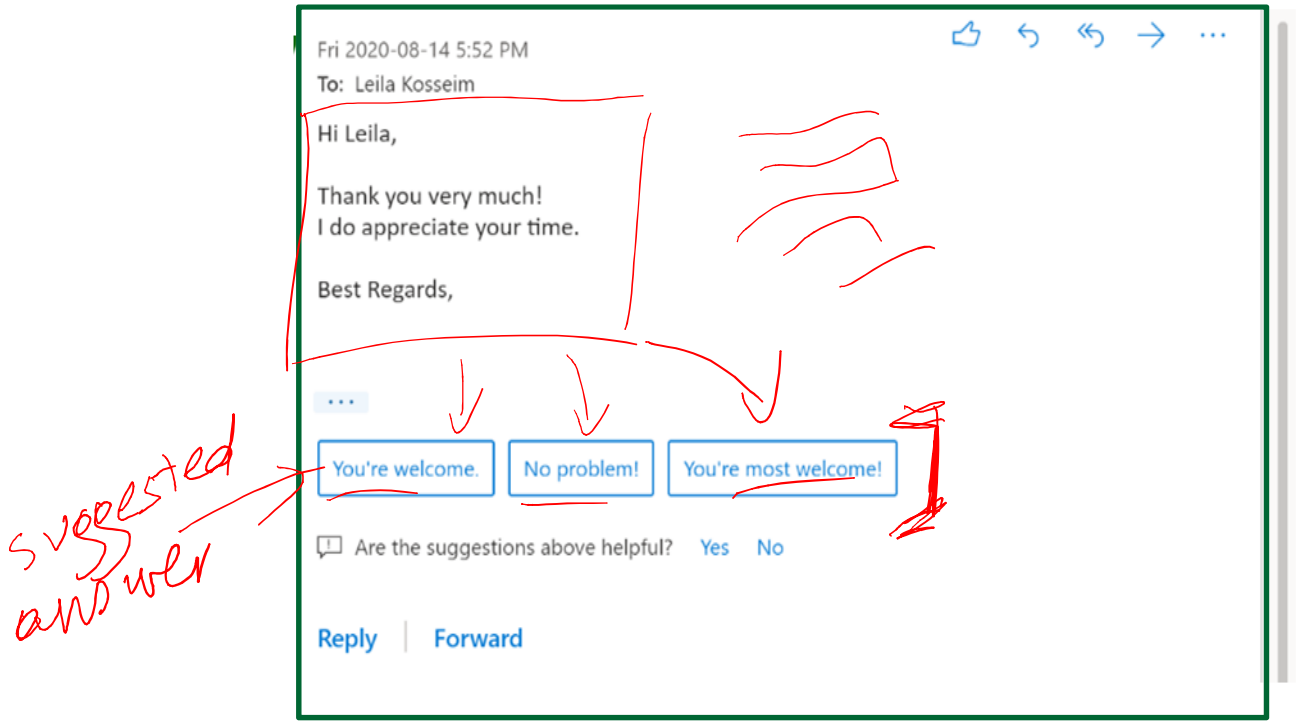
Date: Jan-16-2012

Start: 10:00am

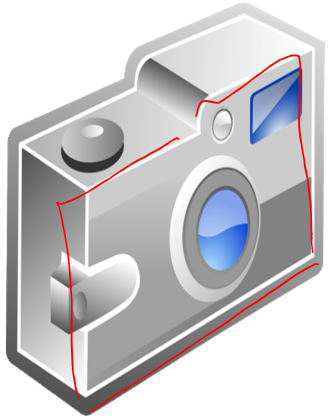
End: 11:30am

Where: Gates 159

Email Answering

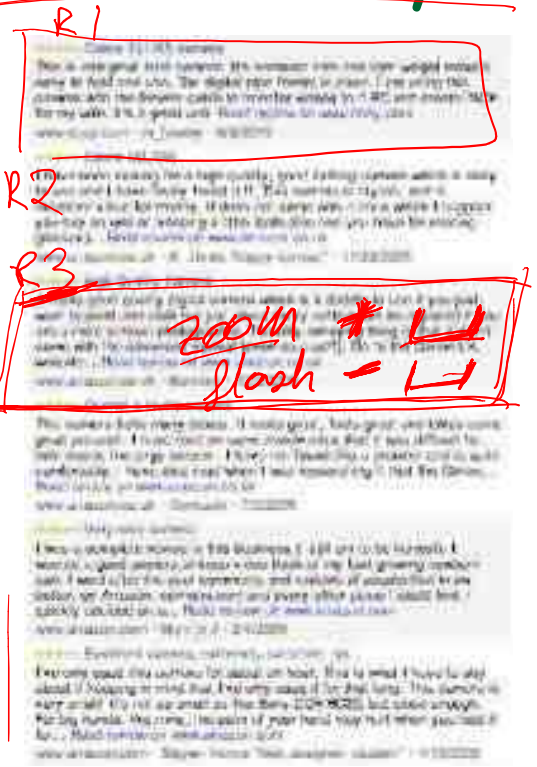


Information Extraction & Sentiment Analysis



Attributes:

- zoom
- affordability
- size and weight
- flash
- ease of use



Size and weight

✓ 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

slide from Olga Veksler (U. Western Ontario)

Machine Translation

Fully automatic

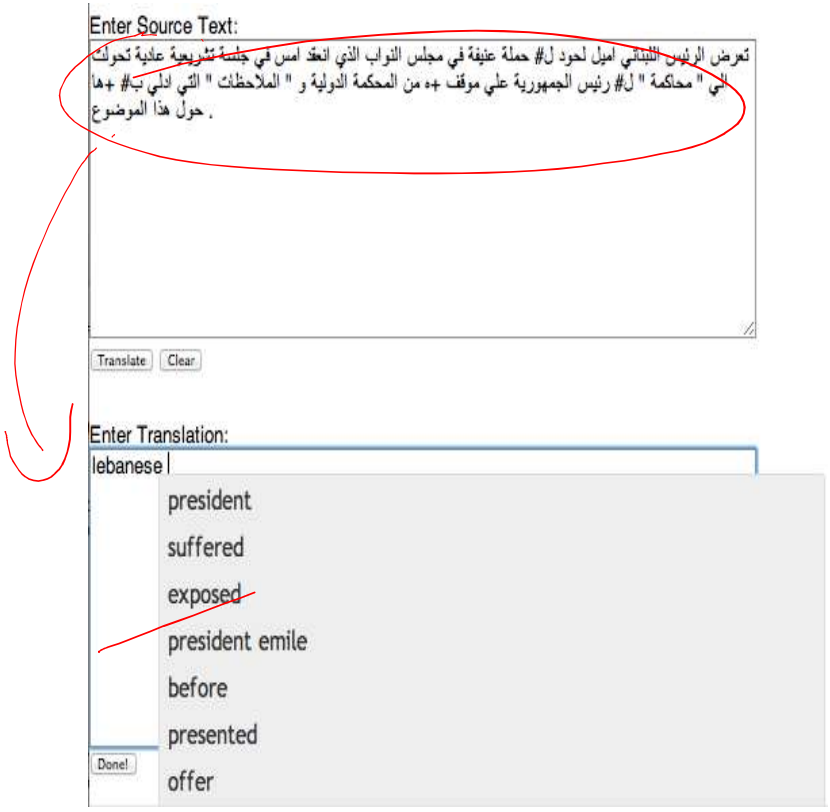
Enter Source Text:

这 不过 是 一 个 时间 的 问题 .

Translation from Stanford's *Phrasal*:

This is only a matter of time.

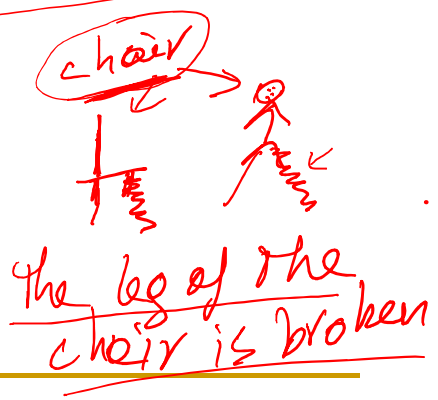
Helping human translators



slide from Olga Veksler (U. Western Ontario)

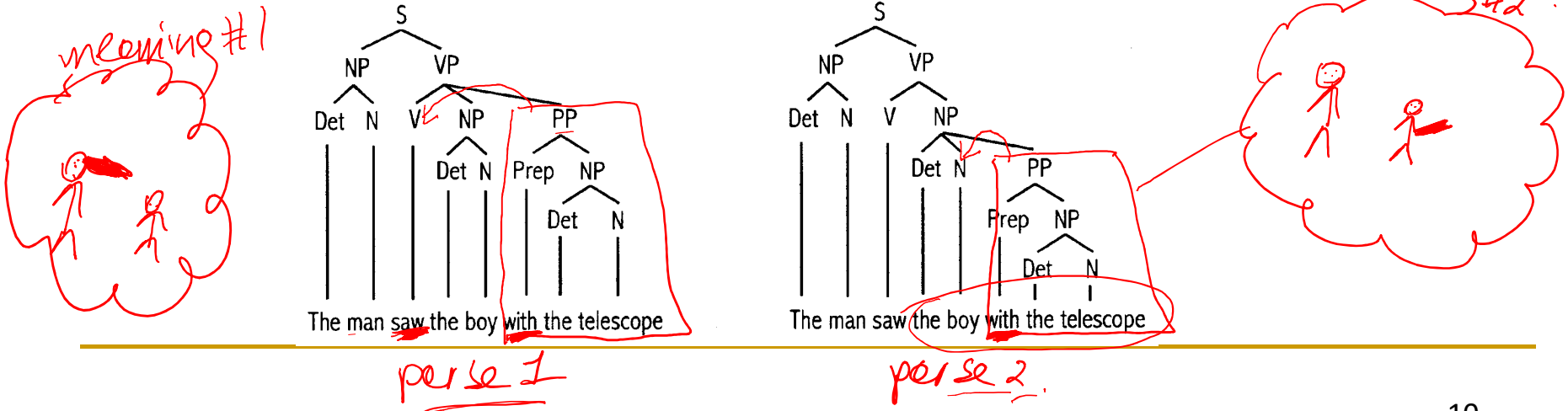
Why is NLP hard?

- Languages
  - Artificial
    - Smaller vocabulary
    - Simple syntactic structures
    - Non-ambiguous
    - Not tolerant to errors (ex. Syntax error)
  - Natural
    - Large and open vocabulary (new words everyday)
    - Complex syntactic structures
    - Very ambiguous
    - Robust (ex. forgot a comma, a word... still OK)



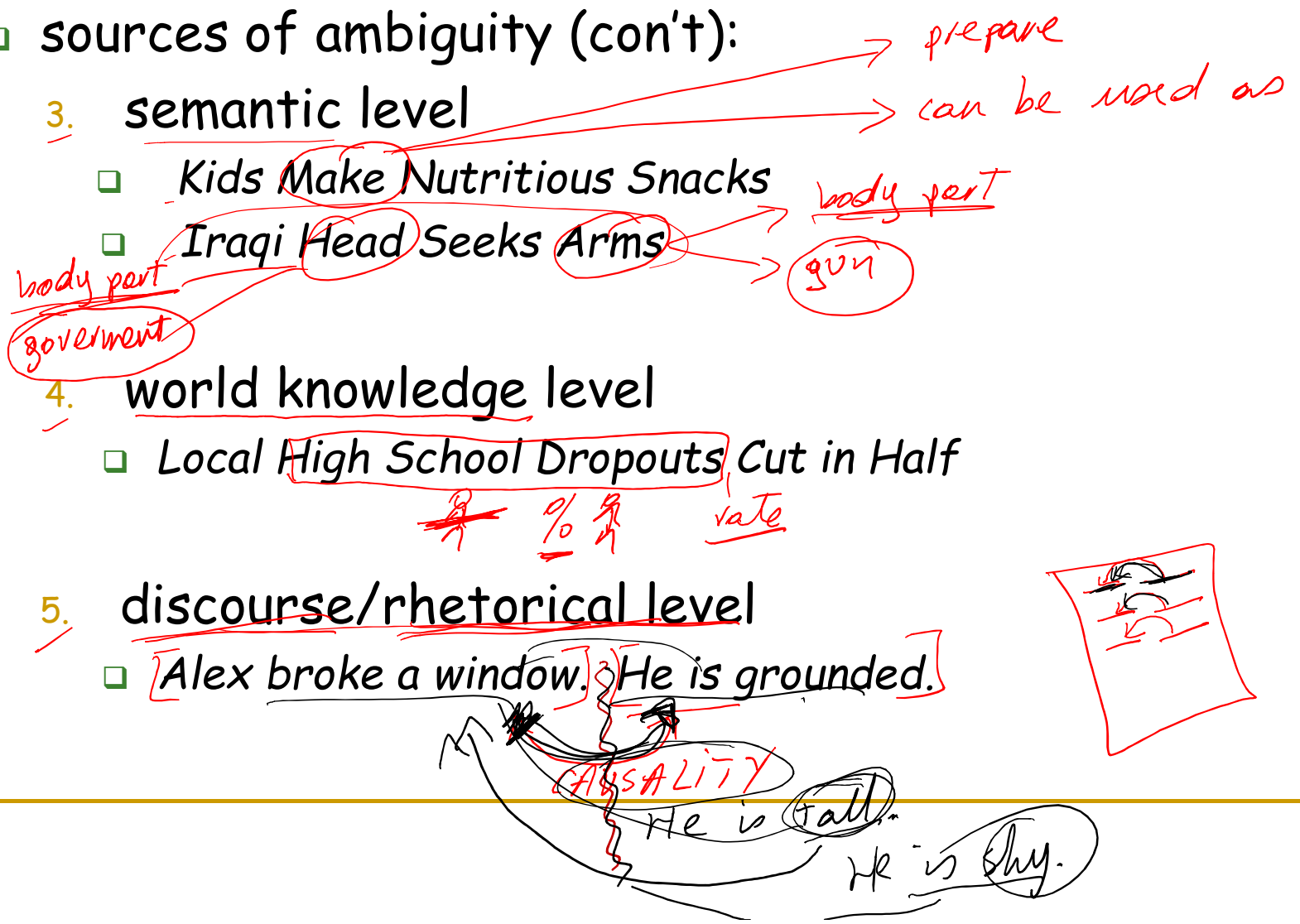
Ambiguity

- Even simple sentences can be highly ambiguous at different levels
- sources of ambiguity:
  1. lexical level
     - Can I offer you a glass of *airag*?
  2. syntactic level
     - *The man saw the boy with the telescope.*



Ambiguity

- sources of ambiguity (con't):
  3. semantic level
     - *Kids Make Nutritious Snacks*
     - *Iraqi Head Seeks Arms*
  4. world knowledge level
     - *Local High School Dropouts Cut in Half*
  5. discourse/rhetorical level
     - *Alex broke a window. He is grounded.*



Remember these slides?

History of AI

- Another big "hype" … Expert Systems (70s - mid 80s)
  - people realized that general-purpose problem solving (weak methods) do not work for practical applications
  - systems need specific domain-dependent knowledge (strong methods)
  - development of knowledge-intensive, rule-based techniques
  - major expert systems
    - MYCIN (1972): expert system to diagnose blood diseases
  - In the industry (1980s): First expert system shells and commercial applications.

☹ HUMANS need to write the rules by hand…

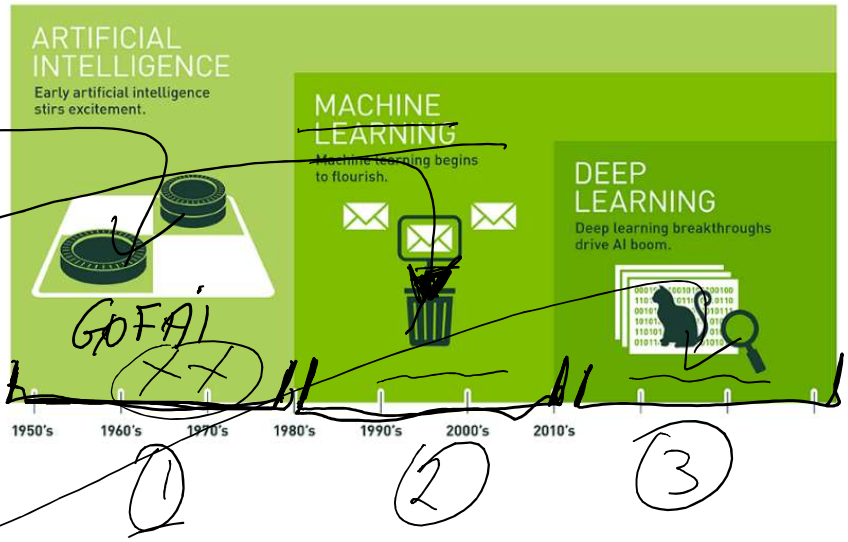
GOFAI

History of AI

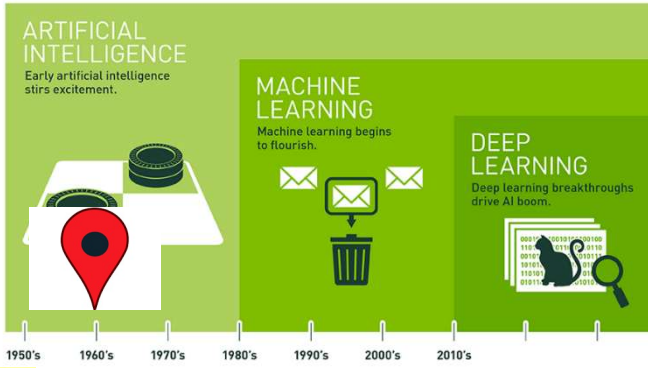
- The rise of Machine Learning (1980s – 2010)
  - More powerful CPUs-> usable implementation of neural networks
  - Big data -> Huge data sets are available to learn from
    - document repositories in NLP, datasets in ML, billions on images for image retrieval, billions of genomic sequences, …
  - ☺ Rules are now learned automatically !
  - AI adopts the Scientific Method

History of AI

- The era of Deep Learning (2010-today)
  - Development of "deep neural networks"
  - Trained on massive data sets
  - Use of GPU for computations
  - Use of "generic networks" for many applications



The Ancient Land of NLP (aka GOFAI)
(circa A.D. 1950...mid 1980)

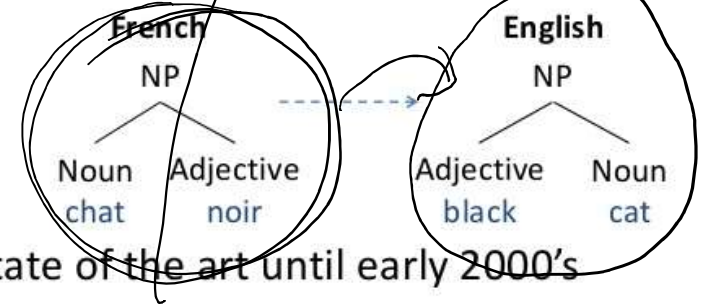


Rule-based NLP
(circa A.D. 1950...mid 1980)

```
s       --> np, vp.
vp      --> v, np.
vp      --> v.
np      --> n.
n       --> [john].     n       --> [lisa].
n       --> [house].
v       --> [died].     v       --> [kissed].

?- s([john, kissed, lisa], []).
   yes
?- s([lisa, died], []).
   yes
?- s([kissed, john, lisa], []).
   no
```

- Rules hand-written by linguists



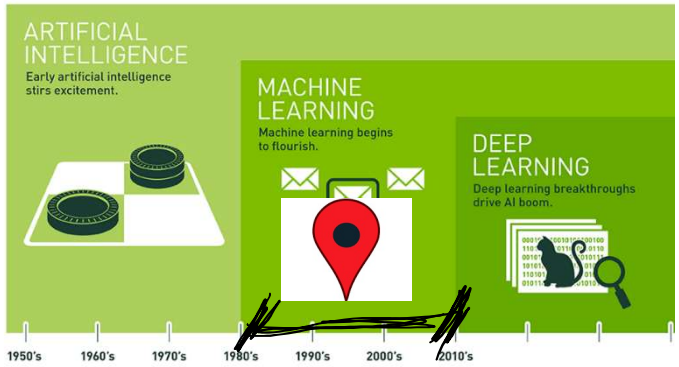
- State of the art until early 2000's
  - e.g. Systran
- Expensive to create maintain and adapt

Symbolic methods / Linguistic approach / Knowledge-rich approach

- Cognitive approach
- Rules are developed by hand in collaboration with linguists

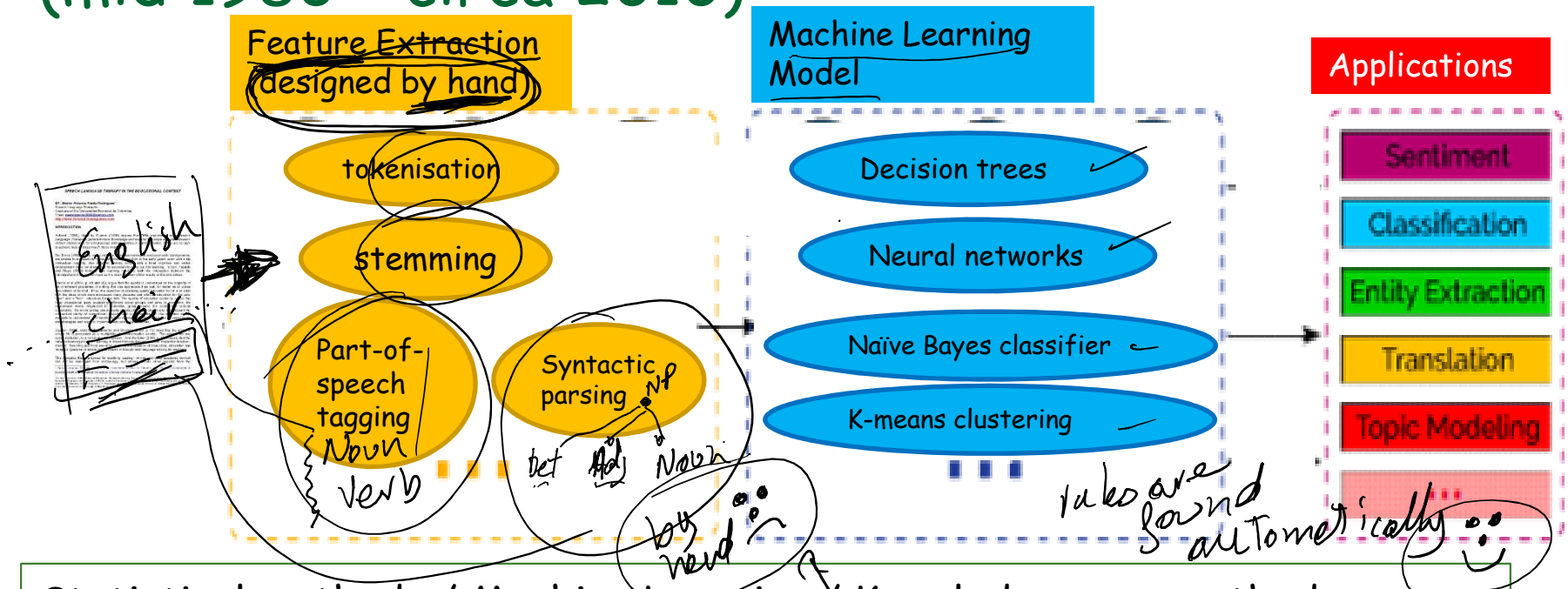
https://image.slidesharecdn.com/maryamsaihbani-161004094341/95/lefttoright-hierarchical-phrasebased-translation-and-its-application-in-simultaneous-speech-translation-maryam-siahbani-4-638.jpg?cb=1477004310

1st Invasion of NLP, from ML
(mid 1980 – circa 2010)



Statistical NLP
(mid 1980 – circa 2010)

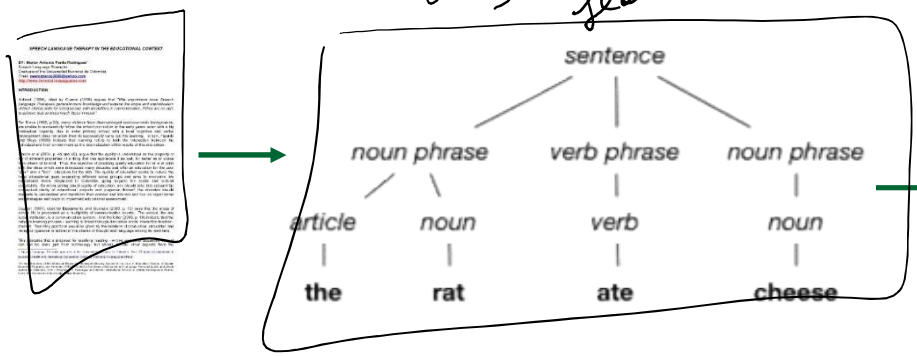
multinomial NB classifier for spam filtering.

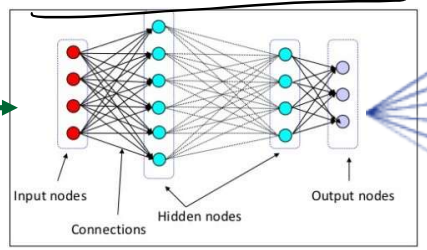


Statistical methods / Machine Learning / Knowledge-poor method

- Engineering Approach
- Rules are developed automatically (using machine learning)
- But the linguistic features are hand-engineered and fed to the ML model
- Applications: Information Retrieval, Predictive Text / Word Completion, Language Identification, Text Classification, Authorship Attribution...

Statistical NLP (mid 1980 – circa 2010)





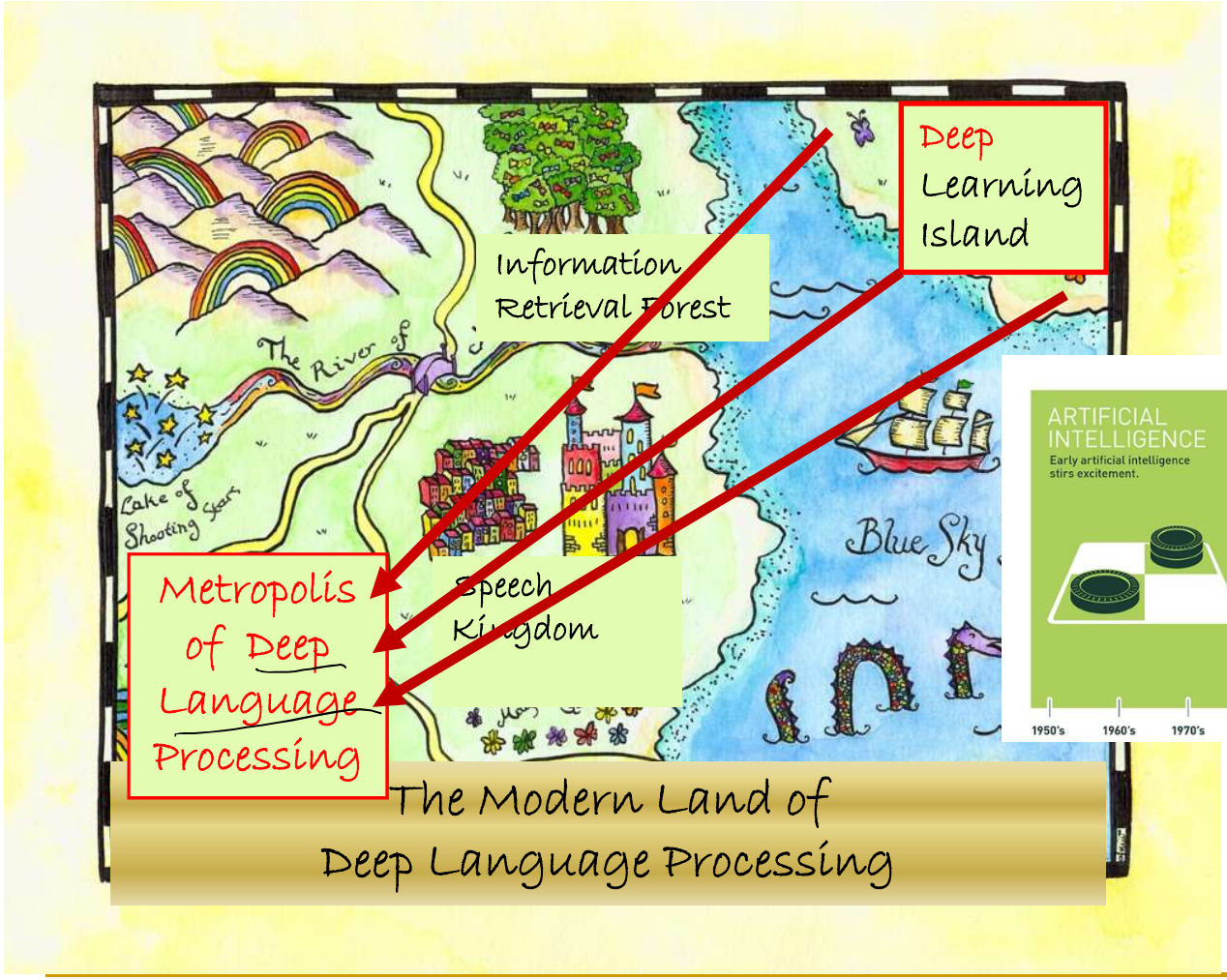
Applications

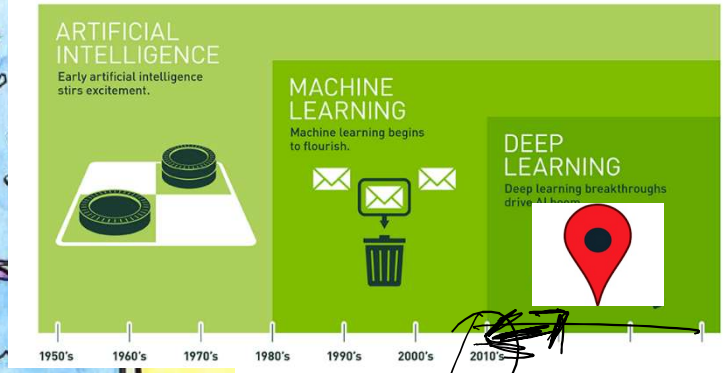
- Sentiment
- Classification
- Entity Extraction
- Translation
- Topic Modeling
- ...

Google translate based on statistical Machine Translation

linguistic features are hand-engineered and fed to the ML model

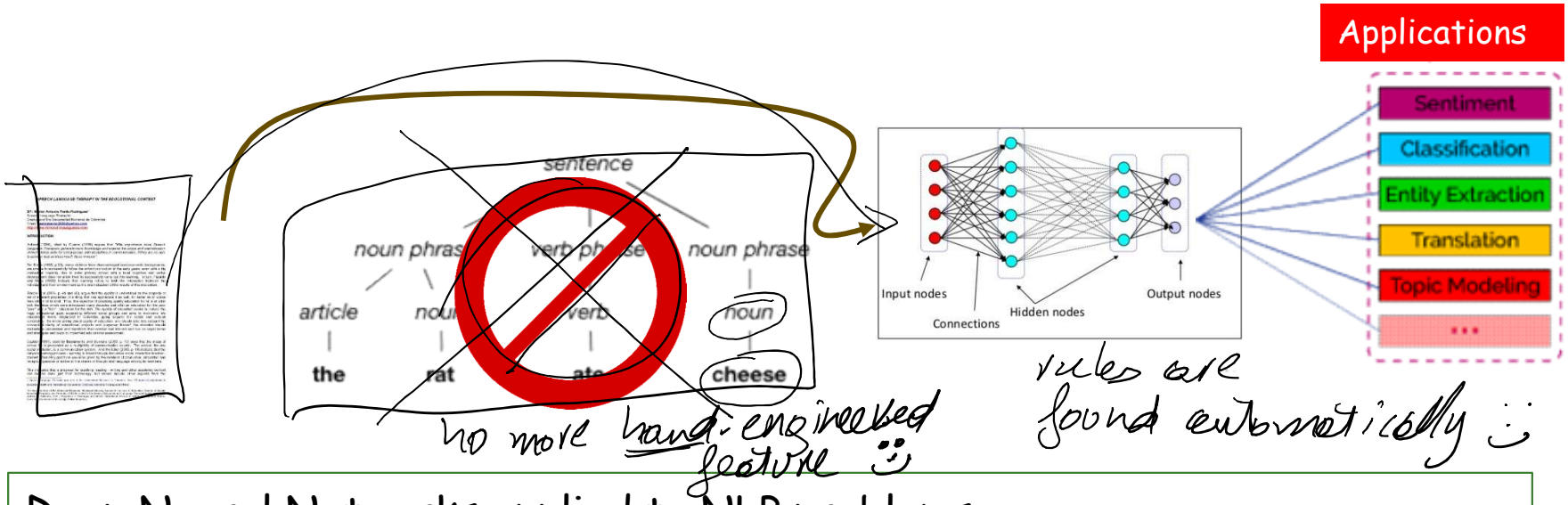
2nd Invasion of NLP, by Deep Learning (circa 2010-today)





2015

Deep Language Processing (3)
(circa 2010-today)



Deep Neural Networks applied to NLP problems

- Rules are developed automatically (using machine learning)
- And the linguistic features are found automatically!

Today

1. Introduction 
2. Bag of word model
3. n-gram models
4. Deep Learning for NLP
    1. Word Embeddings
    2. Recurrent Neural Networks

Up Next

1. Introduction
2. Bag of word model
3. n-gram models
4. Deep Learning for NLP
    1. Word Embeddings
    2. Recurrent Neural Networks