

Introduction to Natural Language Processing

Dr. Hamada Nayel
AIx51, Lecture 01

1

Outline

- ▶ What is NLP?
- ▶ Why NLP?
- ▶ Applications of NLP.
- ▶ Tasks of NLP.
- ▶ Challenges of NLP

2

Natural Language Processing (NLP)

A field in Artificial Intelligence (AI) devoted to creating computers that use natural language as input and/or output.

```
graph LR; A[language] --> C[computer]; C --> B[language]; C <--> D[understanding<br/>(NLU)]; C <--> E[generation<br/>(NLG)];
```

3

Natural Language Processing (NLP)

- Identify the structure and meaning of ***words, sentences, texts and conversations.***
- Deep understanding of broad language
- NLP is all around us

4

Why NLP?

- To interact with *computing devices* using human (natural) languages. For example,
- Building intelligent robots (AI).
- Enabling voice-controlled operation.
- To access (large amount of) information and knowledge stored in the form of human languages quickly.

5

Early Conversational Programs

ELIZA (by Joseph Weizenbaum),

1966

- A psychotherapist
- No real understanding; simple pattern-matching to respond to user input -> canned responses

```
>> Hello.  
How do you do. Please state your problem.  
>> Men are all alike.  
In what way ?  
>> They're always bugging us about something or other.  
I'm not sure I understand you fully.  
>> They're always bugging us about something or other.  
Can you think of a specific example ?  
>> Well, my boyfriend made me come here.  
Your boyfriend made you come here ?  
>> He says I'm depressed much of the time.  
I am sorry to hear that you are depressed.
```

6

Modern NLP: NLP in the Days of Big Data

Three trends:

1. An **enormous amount of information** is now available (newspapers, web pages, medical records, financial filings, product reviews, discussion forums, etc.)
2. Conversational agents are becoming an important form of human-computer **communication**
3. Much of human-human interaction is now mediated by computers via **social media**

7

NLP Applications

➤ NLP breaks down language into shorter, called *tokens* (words, periods, etc.), and attempts to understand the relationships of the tokens. This process often uses *higher-level features*, such as:

8

NLP Applications (Content Classification)

<https://www.analyticsvidhya.com/blog/2018/04/a-comprehensive-guide-to-understand-and-implement-text-classification-in-python/>

www.wired.com

<https://www.masrawy.com/>

9

NLP Applications (Machine Translation)

- ▶ One of the oldest non-numerical applications in CS.
- ▶ MT has gone from a niche academic curiosity to a robust **commercial industry**.
- ▶ Tourism – Multi-language Community (India)

<https://analyticsindiamag.com/deep-dive-in-datasets-for-machine-translation-in-nlp-using-tensorflow-and-pytorch/>

10

NLP Applications (Document Summarization)

https://github.com/nandini1999/text-summarizer_TF-IDF

<https://medium.com/@prashasti.sonal/brief-history-of-text-summarization-9d1b3787a5707>

- ▶ Condensing a piece of **text** to a shorter version:
 - ▶ Media monitoring.
 - ▶ Newsletters.
- ▶ Financial research.
- ▶ Legal contract analysis.
- ▶ Social media marketing.
- ▶ Question answering and bots.

11

NLP Applications (Sentiment Analysis)

- ▶ Deals with categorization (or classification) of opinions expressed in textual documents
- ▶ Helps businesses to monitor brand and product sentiment in customer feedback and understand customer needs.

<https://analyticsindiamag.com/deep-dive-in-datasets-for-machine-translation-in-nlp-using-tensorflow-and-pytorch/>

12

NLP Applications (Conversational Agents)

- ▶ Examples (Siri, Amazon Alexa..., Chatbots.) and combine
 - ▶ Speech recognition/synthesis
 - ▶ Question answering
 - ▶ From the web and from structured information sources
 - ▶ Simple agent-like abilities
 - ▶ Create/edit calendar entries
 - ▶ Reminders
 - ▶ Directions
 - ▶ Invoking/interacting with other apps



<https://www.frontiersin.org/research-topics/24418/socially-intelligent-conversational-agents-co-evolving-with-humans>

13

IBM Watson



https://www.youtube.com/watch?v=WFR3lOm_xhE

14

NLP Applications

mostly solved	making good progress	still really hard
Spam detection OK, let's meet by the big ✓ GTF too small? Buy VAGA... X	Sentiment analysis The pho was authentic and yummy! Waiter ignored us for 20 minutes.	Semantic search people protesting globalization demonstrations offices
Text categorization Philips cuts down Rangers 2-0 SPORTS Jobless rate hits five-year low BUSINESS	Coreference resolution Obama told Mueller he shouldn't run again.	Question answering (QA) What currency is used in China? A: The yuan
Part-of-speech (POS) tagging AG AGI NOUN VERB ADV Colored, green, sleep, sleep furiously.	Word sense disambiguation (WSD) I need new batteries for my mouse.	Textual inference & paraphrase Three soldiers lost their lives. Several houses were affected. YES
Named entity recognition (NER) PERSON ORG LOC Obama met with UAW leaders in Detroit...	Syntactic parsing house	Summarization Sheen continues her against Sheen is nuts
Information extraction (IE) You're invited to our bunga bunga party, Friday May 27 at 8:30pm in Condura Hall Party May 27 add	Machine translation (MT) Our specialty is panda fried rice!	Discourse & dialog Where is Thor playing in '97? Metron at 4:30 and 7:30

Source: Computational Linguistics, Bill McCartney, Indiana University, 2003.

15

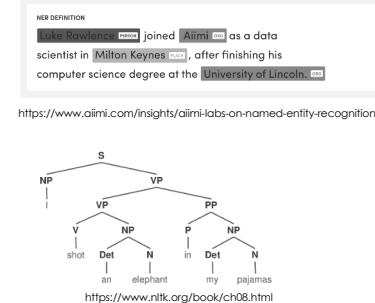
NLP Tasks

- ▶ NLP applications require several NLP analyses:
 - ▶ Word tokenization
 - ▶ Sentence boundary detection
 - ▶ Part-of-speech (POS) tagging
 - ▶ To identify the part-of-speech (e.g. noun, verb) of each word

16

NLP Tasks

- ▶ Named Entity Recognition (NER)
- ▶ to identify proper nouns (e.g. names of person, location, organization;)
- ▶ Parsing
- ▶ to identify the syntactic structure of a sentence



17

But NLP very is hard..

- ▶ Understanding natural languages is hard ... because of inherent *ambiguity*
- ▶ Engineering NLP systems is also hard ... because of:
 - ▶ Huge amount of data resources needed (e.g. grammar, dictionary, documents to extract statistics from)
 - ▶ Computational complexity (intractable) of analyzing a sentence

18

Challenges (Ambiguity)

- ▶ "Get the cat with the gloves."



19

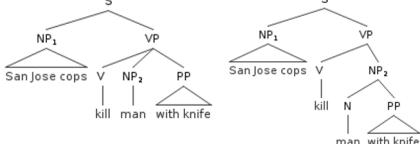
Challenges (Ambiguity)

- ▶ Word sense / meaning ambiguity



20

Challenges (Ambiguity)



Credit: Mark Liberman, <http://languagelog.ldc.upenn.edu/?p=17711>

21

Challenges – language is not static

Language grows and changes

LOL	Laugh out loud
G2G	Got to go
BFN	Bye for now
B4N	Bye for now
ldk	I don't know
FWIW	For what it's worth
10Q	Thank You
2m2h!	Too much to handle
?4U	I have a question for you

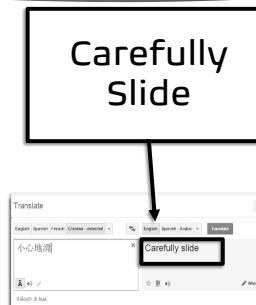
23

Challenges (Context Awareness)

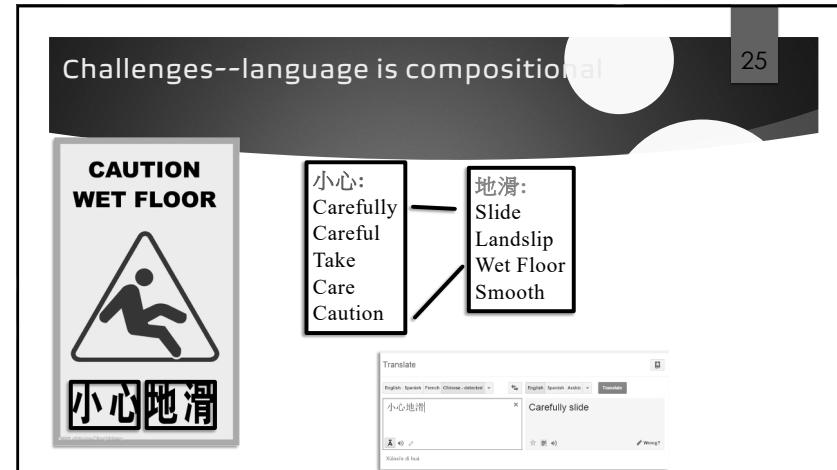
- ❖ I am a talented individual working for Microsoft Windows core OS team
 - ❖ I am a janitor, I wash windows at Microsoft.

22

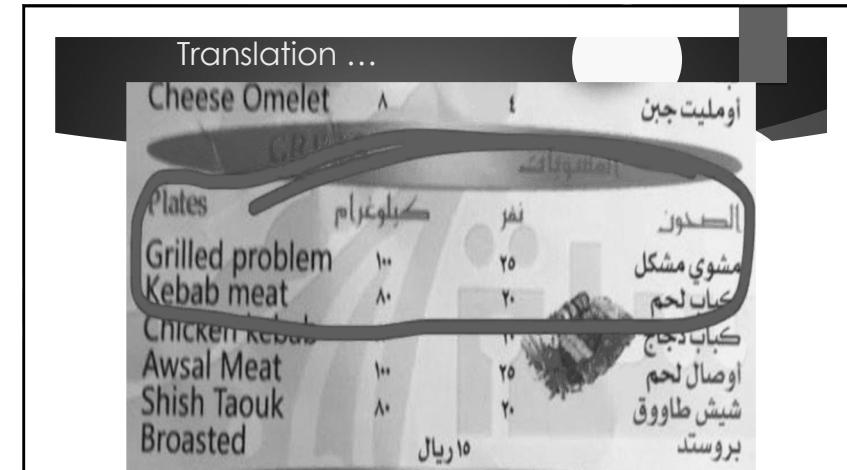
Challenges--language is compositional



24



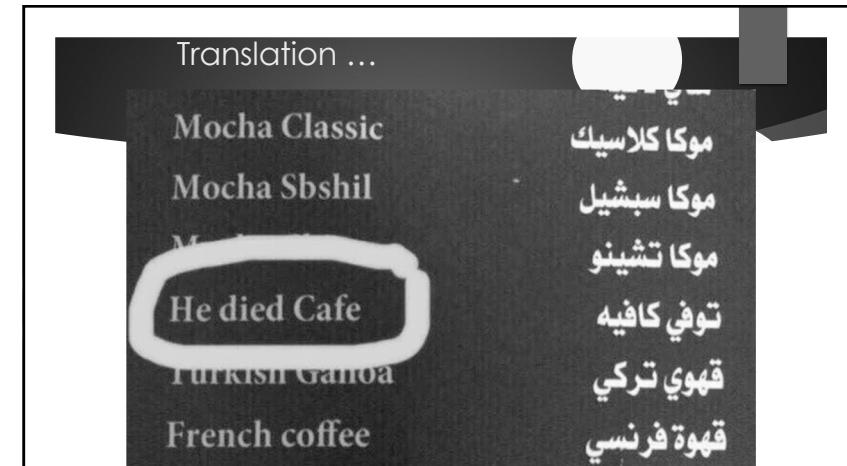
25



26



27



28