

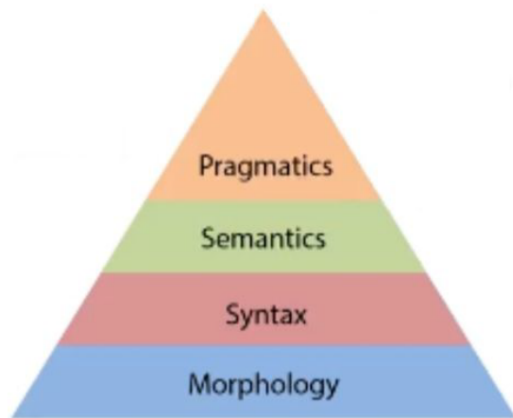
Plan of the Tutorial

- 1 Plan of the Tutorial
- 2 Introduction to NLP**
- 3 Overview of Distributional Representation Learning for NLP
- 4 Overview of Transformer based Language Model
- 5 Overview of Large Language Models
- 6 Concept of in-context learning and its application
- 7 Conclusion

Natural Language Processing (NLP)

- What is NLP?
- What makes NLP difficult?

NLP Pyramid



Ambiguity in Natural Language

- Lexical Ambiguity
- Structural Ambiguity
- Semantic Ambiguity

Lexical Ambiguity

- When a single word possesses more than one interpretation
 - Morphology can be ambiguous
 - does drier mean 'more dry' or 'machine which dries'?
 - Words can represent different parts of speech
 - E.g., ship, round, etc.
 - Words can have multiple meanings
 - E.g., bank, letter, ball
- How to disambiguate?
 - Look at the context.
 - "John kicked a ball."
 - "The ball lasted until midnight."
 - "When you hold a ball. . ."

Structural Ambiguity

- When multiple syntactic structures can be associated with a sentence.
- E.g.,
 - “Flying planes can be dangerous.”
 - “Visiting relatives can be boring.”
 - “Time flies like an arrow.”
 - “Fruit flies like banana.”
 - “The boy saw the girl with a telescope.”

Semantic Ambiguity

- How about meanings of polysemous words?

He deposited money in a bank account
with a high interest rate.

Sitting on the bank of the Mississippi,
a passing ship piqued his interest.

- Phrasal Verbs

- “John eats in on Sundays.”
- “John eats in a restaurant on weekdays.”
- “John fell for Mary.”
- “John fell for a lie.”

- Idioms

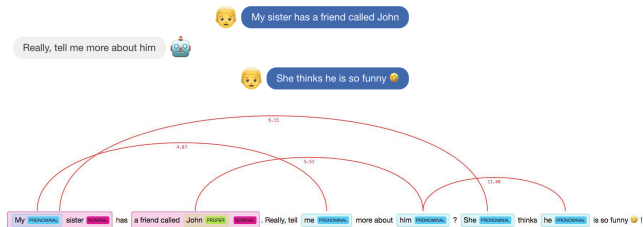
- kick the bucket
- let your hair down

Coreference Resolution

- Example

- “The police opened fire at the crowd, and some of them fell.”
- “The police opened fire at the crowd, and some of them missed.”
- Winograd Schema Challenge: “The city councilmen refused the demonstrators a permit because they [feared/advocated] violence”

- Utility in Chatbots



Source:

<https://medium.com/huggingface/state-of-the-art-neural-coreference-resolution-for-chatbots-3302365dcf30>

Different Levels of Ambiguity

- E.g.,
 - “He noticed her shaking hands.”
 - “I like tennis.”
 - “John is eager to please.”
 - “John is easy to please.”
 - “The chicken is ready to eat.”
 - NLP?

Ambiguity is Pervasive

- Do you see any ambiguity in this?
 - I made her duck
 - I cooked waterfowl for her benefit (to eat)
 - I cooked waterfowl belonging to her
 - I created the (plaster?) duck she owns
 - I caused her to quickly lower her head or body
 - I waved my magic wand and turned her into undifferentiated waterfowl

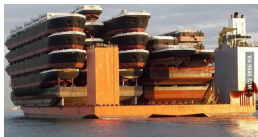
Ambiguity is Pervasive

- Do you see any ambiguity in this?
 - I made her duck
 - I cooked waterfowl for her benefit (to eat)
 - I cooked waterfowl belonging to her
 - I created the (plaster?) duck she owns
 - I caused her to quickly lower her head or body
 - I waved my magic wand and turned her into undifferentiated waterfowl

Ambiguity in Speech Processing

- Phonetics! – I made her duck
 - I mate or duck
 - I'm eight or duck
 - Eye made her duck
 - Aye mate, her duck
 - I maid her duck
 - I'm aid her duck
 - I mate her duck
 - I'm ate her duck
 - I'm ate or duck
 - I mate or duck

More Ambiguity



- Difficult to solve
 - “Teacher Strikes Idle Kids”
 - “oil prices rose gas prices rose up”
 - A ship-shipping ship ships shipping-ships.
 - Will, will Will will Will Will's will?
 - “Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo”

Why else is NLP difficult?

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 ❤️

segmentation issues

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

idioms

dark horse
get cold feet
lose face
throw in the towel

neologisms

unfriend
Retweet
Bromance
Netizens

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
...

But that's what makes it fun!

To name a few. . .

- Information Retrieval and Extraction
- Web Search
- Text Categorization
- Text Summarization
- Sentiment Analysis
- Math Word Problems
- Recommendation Systems
- Knowledge Extraction
- Fact Checking

Applications of NLP

To name a few...

- Machine Translation
- Question Answering
 - On Unstructured data (i.e., Text)
 - On Structured data
 - On Database (NL query \rightarrow SQL)
 - On Knowledge Base (NL query \rightarrow SPARQL)
- Text Normalization
 - Spell checker
 - Grammar Correction
 - Disfluency Detection in Natural Speech

To name a few. . .

- Text Similarity
 - Plagiarism Detection
 - Answer / Essay grading
- Readability Assessment and Text Simplification
- Event extraction and Causality
- Natural Language Generation
- Domain-specific NLP
 - FinNLP, BioNLP, LegalNLP
- Multimodal Processing

Text Classification

- Assigning subject categories, topics, or genres
- Spam detection
- Authorship identification
- Age/gender identification
- Aggression/Hate Speech detection
- Personality identification
- Language identification
- Sentiment analysis
- ...

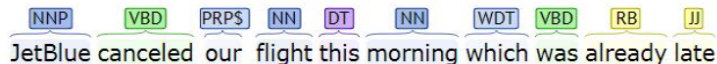
To name a few...

- Speech Processing
 - Core Tasks
 - Speech Recognition (ASR)
 - Speech Synthesis
 - Additional Tasks
 - Speaker Identification
 - Speech Diarization
 - Normalization of ASR o/p
 - Disfluency Detection
 - Dialog Systems

- Basic Text Processing
 - Sentence segmentation
 - Word Tokenization
 - Normalization
- Lexical Analysis
 - Lemmatization (e.g., comes/coming/came → come)
 - Morphological Analysis (e.g., unsuccessfully → un+success+ful+ly)

Syntactic Analysis

- POS Tagging
 - JetBlue canceled our flight this morning which was already late



Core NLP Tasks

Syntactic Analysis

- POS Tagging
 - JetBlue canceled our flight this morning which was already late

(ROOT

(S

(NP (NNP JetBlue))

(VP (VBD canceled)

(NP (PRP\$ our) (NN flight))

(NP (DT this) (NN morning))

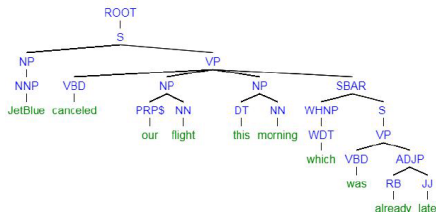
(SBAR

(WHNP (WDT which))

(S

(VP (VBD was)

(ADJP (RB already) (JJ late))))))

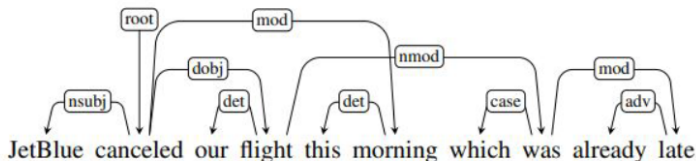
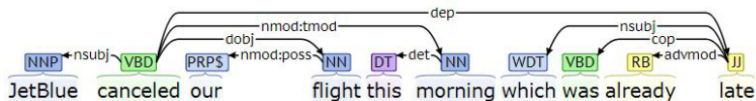


Core NLP Tasks

Syntactic Analysis

- POS Tagging

- JetBlue canceled our flight this morning which was already late



Semantic Analysis

- Word Sense Disambiguation (WSD)
- Semantic Similarity
- Word Sense Induction

Distributional models of meaning

- Also called vector-space models of meaning
- Zellig Harris (1954): “oculist and eye-doctor . . . occur in almost the same environments. . . .
 - If A and B have almost identical environments we say that they are synonyms.
- Firth (1957): “You shall know a word by the company it keeps!”

Intuition of Distributional Semantics

- Nida example: tesgüino

A bottle of tesgüino is on the table

Everybody likes tesgüino

Tesgüino makes you drunk

We make tesgüino out of corn.

- From context words humans can guess tesgüino
 - means: an alcoholic beverage like beer
- Intuition for the algorithm:
 - Two words are similar if they have similar word contexts.

More Tasks

- Named Entity Recognition
- Anaphora and Coreference Resolution
- Textual Entailment / Natural Language Inferencing (NLI)
- Paraphrasing
- Entity disambiguation
- . . .

Named Entity Recognition (NER)

A very important sub-task: **find** and **classify** names in text, for example:

- The decision by the independent MP Andrew Wilkie to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the 2010 election, Wilkie, Rob Oakeshott, Tony Windsor and the Greens agreed to support Labor, they gave just two guarantees: confidence and supply.

Named Entity Recognition (NER)

A very important sub-task: **find** and **classify** names in text, for example:

- The decision by the independent MP **Andrew Wilkie** to withdraw his support for the minority **Labor** government sounded dramatic but it should not further threaten its stability. When, after the **2010** election, **Wilkie, Rob Oakeshott, Tony Windsor** and the **Greens** agreed to support **Labor**, they gave just two guarantees: confidence and supply.

Named Entity Recognition (NER)

A very important sub-task: **find** and **classify** names in text, for example:

- The decision by the independent MP **Andrew Wilkie** to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the **2010** election, **Wilkie**, **Rob Oakeshott**, **Tony Windsor** and the **Greens** agreed to support **Labor**, they gave just two guarantees: confidence and supply.
- **Person**, **Date**, **Location**, **Organization**.

Entity Resolution & Linking

...during the late 60's and early 70's, **Kevin Smith** worked with several local...
...the term hip-hop is attributed to **Lovebug Starski**. What does it actually mean...



Like Back in 2008, the Lions drafted **Kevin Smith**, even though Smith was badly...
... backfield in the wake of **Kevin Smith**'s knee injury, and the addition of Haynesworth...



The filmmaker **Kevin Smith** returns to the role of Silent Bob...
Nothing could be more irrelevant to **Kevin Smith**'s audacious "Dogma" than ticking off...



The Physiological Basis of Politics," by **Kevin Smith**, Douglas Oxley, Matthew Hibbing...

