



Introduction to Natural Language Processing

Chloé Clavel,
chloe.clavel@telecom-paristech.fr,

Telecom ParisTech, France



Natural Language processing ?

In French : Traitement Automatique du Langage Naturel (TALN)

Crossroads of :

- ▶ Artificial Intelligence
- ▶ Linguistics
- ▶ Machine learning

Natural Language processing ?

Objectives :

- ▶ Extract meaning from textual data
- ▶ Speech synthesis, natural language generation

https://www.youtube.com/watch?v=Ea_ytY0UDs0 Luc Steels -
BREAKING THE WALL TO LIVING ROBOTS. How Artificial
Intelligence Research Tries to Build Intelligent Autonomous
Systems - 1 min 52



NLP applications

?

NLP applications

- ▶ Automatic translation (Google translate)
- ▶ Textual data mining/ document classification / information extraction
- ▶ Spell-checkers
- ▶ Automatic summary
- ▶ Human-Computer interactions
- ▶ Speech recognition
- ▶ Speech synthesis
- ▶ Opinion analysis (from social media)

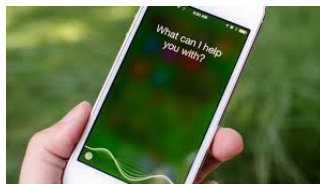
The textual Data

Natural language is everywhere on the web :



in companies : chatbots, callcentres, mails, etc.

at home : amazon echo, siri, google home



amazon echo

Always ready, connected, and fast. Just ask.




The textual Data and its challenges


First Challenge : build models that represent language that is far from academic writing : spontaneous expressions (abbreviations, hashtags, acronyms, typos/mistakes, oral transcript)

 **Agence France-Presse** @afpr 5 Sept
Incident chimique à Fessenheim: 2 personnes légèrement brûlées, selon EDF bit.ly/TmN97R #AFP
Ouvrir

 **Stéphane GRAND** @Stephane_Grand 5 Sept
Incident à #Fessenheim : deux personnes ont été légèrement brûlées, selon EDF
Ouvrir

 **Mediapart** @mediapart 4 Sept
Nucléaire: les déboires anglo-saxons d'EDF avec son EPR
bit.ly/R3z2EY
Ouvrir

 {breath} bonjour Madame . C'est bon madame , vous n'y êtes pour rien , mais je vais passer ma colère sur vous .

 D'accord .

The textual Data and its challenges

Second challenge : build model from small labelled dataset when labels are difficult to obtain (ex : opinions)

Samples

“ The acting is terrible, the plot is ridiculous but no one took it seriously ”



“madame, vous n’y êtes pour rien mais je vais passer ma colère sur vous”



“lol, A +, mouhahah a”



→ Labels?

The textual Data and its challenges

Third challenge – explainability : build transparent and explainable models



MDI341 and INF344 Pedagogical team

NLP lectures and labs during

- ▶ MDI341 (P3) : focus on Machine/Deep learning for NLP
- ▶ INF344 (P4) : focus on linguistic approaches

Lectures given by **Matthieu Labeau** and **CLAVEL Chloe**

Labs given by **Matthieu Labeau**

MDI341 : focus on Machine Learning

2 Lectures

- ▶ Word embeddings and Deep learning for NLP **CLAVEL Chloe** (hybrid)
- ▶ Sequential models for NLP **Matthieu Labeau** (remotely)

2 Labs

- ▶ Word embeddings **Matthieu Labeau** (remotely)
- ▶ Sequential models **Matthieu Labeau** (remotely)

INF344 : focus on linguistic approaches

2 Lectures (**C. Clavel**)

- ▶ Pre-processing, syntactic analysis and resources for Natural Language Processing
- ▶ Sentiment Analysis

3 Labs (**M. Labeau**)

- ▶ Bert representation for information extraction
- ▶ Linguistic approaches for sentiment analysis
- ▶ Pre-processing and machine learning for sentiment analysis

At the end of the course...

- ▶ You will be able to describe and implement the different methods for text representation into vectors
- ▶ You will master the main linguistic issues for NLP
- ▶ You will be able to build a text classification framework