

Université des Sciences et de la Technologie HOUARI BOUMEDIENE

Faculté d'Informatique

Département d'Intelligence Artificielle et Sciences des Données

Traitement Automatique du Langage Naturel (TALN)

Introduction

Séance N° : 1

Ilyes KHENNAK

Maitre de Conférences Classe A

Laboratoire de Recherche en Intelligence Artificielle (LRIA)

Bureau 212, Département IASD, Faculté d'Informatique

Introduction au TALN

Motivation

“ ... *Natural language processing (NLP) or **computational linguistics** is one of the **most important technologies of the information age** ... ”*

“ ... *In the last decade, **deep learning** (or neural network) approaches have obtained very **high performance** across many different **NLP tasks** ... ”*



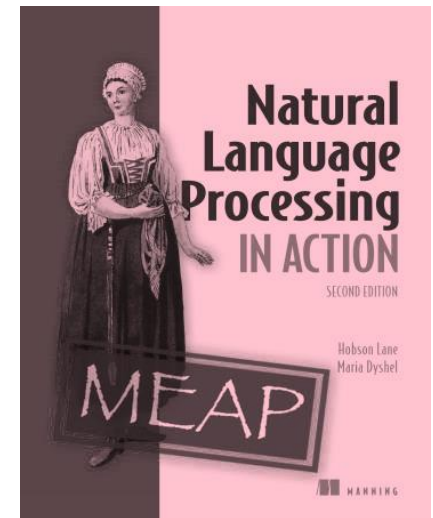
*From **Natural Language Processing with Deep Learning** course, 2023, Stanford University.*

Introduction au TALN

Motivation

*“ ... Natural Language Processing may be the **fastest-developing** and **most important field of Artificial Intelligence** and **Data Science** ...”*

*“ ... If you **want to change the world** you will need to **understand how machines read** and **process natural language text** ...”*



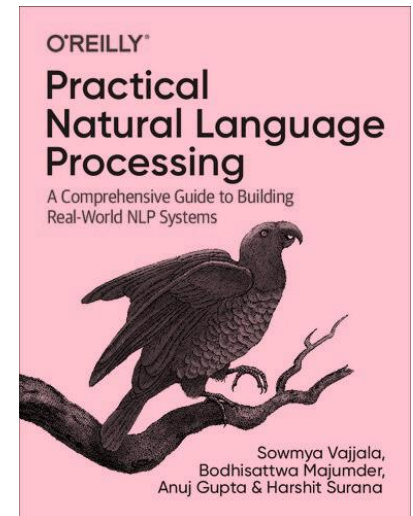
*From **Natural Language Processing in Action**, 2023, Manning Publications.*

Introduction au TALN

Motivation

“... The field of natural language processing (NLP) has undergone a dramatic shift in recent years, in terms of the applications supported ...”

“ ... Every intelligent application involving human language has some NLP behind it ...”



*From **Practical Natural Language Processing**, 2020, O'Reilly Media.*

Introduction au TALN

Motivation

- Des **algorithmes plus rapides**, plus **puissants** et des **applications plus prosociales** pour le **TALN** sont publiés chaque année.

Publié en **2019**



Modèle de langage

Publié en **2020**

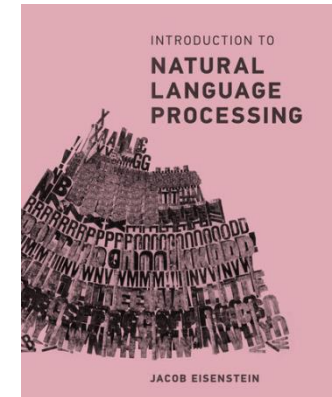


Agent conversationnel ou assistant virtuel (chatbot)

Introduction au TALN

Définitions

*“ ... Natural language processing is the **set of methods for making human language accessible to computers ... ”***



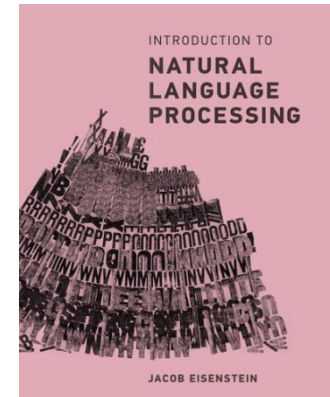
Jacob Eisenstein

*From **Introduction to Natural Language Processing**, 2019.*

Introduction au TALN

Définitions

*“ ... the goal of natural language processing is to provide new **computational capabilities** around **human language** ... for example, **extracting information** from texts, **translating** between languages, **answering questions, holding a conversation** ... ”*



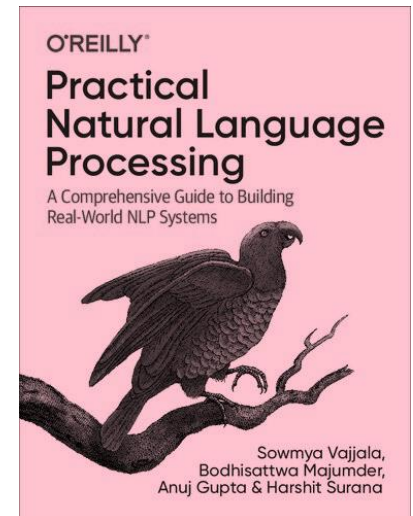
Jacob Eisenstein

*From **Introduction to Natural Language Processing**, 2019.*

Introduction au TALN

Définitions

*“ ... Natural language processing (NLP) is a field at the **intersection of computer science, artificial intelligence, and linguistics**. It concerns **building systems that can process and understand human language ...** ”*



*From **Practical Natural Language Processing**, 2020, O'Reilly Media.*

Introduction au TALN

Définitions

*“ ... Natural language processing (NLP) is a crucial part of **artificial intelligence (AI)** ... Natural language processing is a field of science and engineering focused on the development and study of **automatic systems** that **understand** and **generate natural** (that is, human,) **languages** ... ”*



*From **Natural Language Processing with Deep Learning** course, 2023, Stanford University.*

Introduction au TALN

Définitions

HE

Could you please provide us a definition of NLP?



Certainly! NLP stands for "Natural Language Processing." It is a subfield of artificial intelligence (AI) and computational linguistics that focuses on the interaction between computers and human language. NLP aims to enable computers to understand, interpret, and generate human language in a way that is both meaningful and useful.



From **GPT-3.5**, 2023,
OpenAI.

Introduction au TALN

Natural Language Understanding

- Une **partie du TALN**. Elle consiste à **traiter automatiquement le texte** en vue d'**extraire une représentation numérique du sens** de ce texte.
- La **représentation numérique du sens** du langage naturel prend généralement la forme d'**un vecteur** appelé un **embedding**.

Introduction au TALN

Natural Language Generation

- Une **partie du TALN**. Elle consiste à **composer un texte par une machine** de manière à ce qu'il **ressemble** à celui rédigé par un **être humain** ou de **créer un texte lisible par les humains** à partir d'**une représentation numérique du sens** que vous souhaitez transmettre.

Introduction au TALN

Tâches et applications

Cette tâche consiste à **prédire quel sera le prochain mot dans une phrase** en se basant sur l'historique des mots précédents. L'objectif de cette tâche est d'apprendre la **probabilité qu'une séquence de mots apparaisse dans une langue donnée.**

MODELISATION DU LANGAGE

Language modeling

Application :

Reconnaissance vocale

Reconnaissance d'écriture manuscrite,

Traduction automatique

Correction d'orthographe

Introduction au TALN

Tâches et applications

Cette tâche consiste à **classer le texte** dans un **ensemble connu de catégories** en fonction de son contenu. La classification de texte est **la tâche la plus populaire en TALN.**

CLASSIFICATION DE TEXTE

Text classification

Application :

Identification du courrier indésirable



Analyse des sentiments

Introduction au TALN

Tâches et applications

Cette tâche consiste à **extraire** des **informations pertinentes** d'un **texte**, telles que les **événements du calendrier** à partir des e-mails ou les **noms des personnes** mentionnées dans une publication sur les **réseaux sociaux**.

EXTRACTION D'INFORMATION

Information extraction

Application :

Extraction des événements

 Gmail  Outlook

Introduction au TALN

Tâches et applications

Cette tâche consiste à **trouver** des **documents pertinents** pour une **requête utilisateur** à partir d'une **grande collection**.

RECHERCHE D'INFORMATION

Information Retrieval

Application :

Moteurs de recherche

Google  Bing

Introduction au TALN

Tâches et applications

Cette tâche consiste à **convertir** un morceau de **texte d'une langue à une autre**.

TRADUCTION AUTOMATIQUE

Machine translate

Application :

Service de traduction



Introduction au TALN

Tâches et applications

Cette tâche consiste à **construire** des **systèmes de dialogue** capables de **converser** dans des **langues humaines**.

AGENT CONVERSATIONNEL

Conversational agent

Application :

Assistant virtuel



Introduction au TALN

Tâches et applications

Cette tâche consiste **construire** un **système** capable de **répondre automatiquement aux questions posé en langage naturel.**

QUESTION-REPONSE

Question answering

Introduction au TALN

Tâches et applications

Cette tâche consiste à **créer de courts résumés de documents** tout en conservant le contenu essentiel et en préservant le sens global du **texte.**

RESUME DE TEXTE

Text summarization

Introduction au TALN

Cours



Artificial Intelligence
Graduate Certificate



*Natural Language
Processing with
Deep Learning
course*

Introduction au TALN

Cours



Artificial Intelligence
Graduate Certificate



Andrew ng

***Machine Learning &
Deep Learning***
courses

Introduction au TALN

Cours



Natural Language Processing with Deep Learning course



Chris Manning



227221 citations sur
Google Scholar

Introduction au TALN

Cours

20

The term vocabulary and postings lists

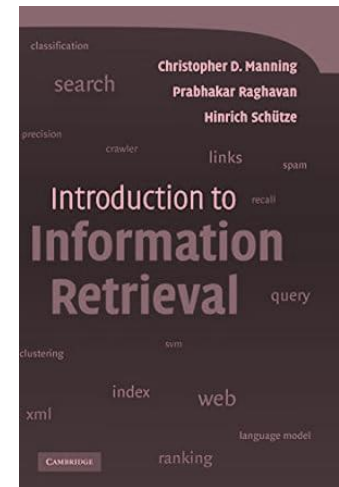
استقلت الجزائر في سنة 1962 بعد 132 عاما من الاحتلال الفرنسي.

← → ← →

← START

‘Algeria achieved its independence in 1962 after 132 years of French occupation.’

Figure 2.2 The conceptual linear order of characters is not necessarily the order that you see on the page. In languages that are written right to left, such as Hebrew and Arabic, it is quite common to also have left-to-right text interspersed, such as numbers and dollar amounts. With modern Unicode representation concepts, the order of characters in files matches the conceptual order, and the reversal of displayed characters is handled by the rendering system, but this may not be true for documents in older encodings.



*From Introduction to
Information Retrieval, 2009,
By Chris Manning.*

24581 citations (Google Scholar)

Introduction au TALN

Cours



Natural Language Processing with Deep Learning course

Speech and Language Processing (3rd ed. draft)

Dan Jurafsky and James H. Martin.

Natural Language Processing

Jacob Eisenstein.

A Primer on Neural Network Models for Natural Language Processing

Yoav Goldberg.

Deep Learning

Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

Natural Language Processing with PyTorch

Delip Rao and Brian McMahan..

Natural Language Processing with Transformers

Lewis Tunstall, Leandro von Werra, and Thomas Wolf.

Introduction au TALN

Cours



Natural Language Processing with Deep Learning course

Speech and Language Processing (3rd ed. draft)

Dan Jurafsky and James H. Martin.

Natural Language Processing

Jacob Eisenstein.

A Primer on Neural Network Models for Natural Language Processing

Yoav Goldberg.

Deep Learning

Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

Natural Language Processing with PyTorch

Delip Rao and Brian McMahan..

Natural Language Processing with Transformers

Lewis Tunstall, Leandro von Werra, and Thomas Wolf.

Introduction au TALN

Cours

Daniel Jurafsky

Stanford University



James H. Martin

University of Colorado



Speech and Language Processing

An Introduction to Natural Language Processing,
Computational Linguistics, and Speech Recognition

Third Edition draft

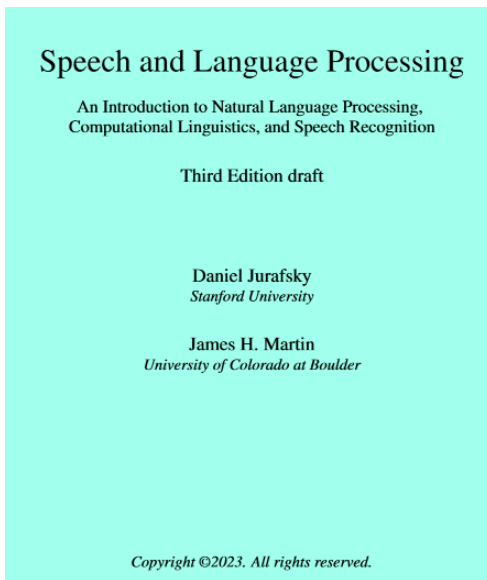
Daniel Jurafsky
Stanford University

James H. Martin
University of Colorado at Boulder

Copyright ©2023. All rights reserved.

Introduction au TALN

Cours



I. Fundamental Algorithms for NLP

1. Introduction

2. Regular Expressions, Text Normalization, Edit Distance

3. N-gram Language Models

4. Naive Bayes, Text Classification, and sentiment

5. Logistic Regression

6. Vector Semantics and Embeddings

7. Neural Networks and Neural Language Models

8. Sequence Labeling for Parts of Speech and Named Entities

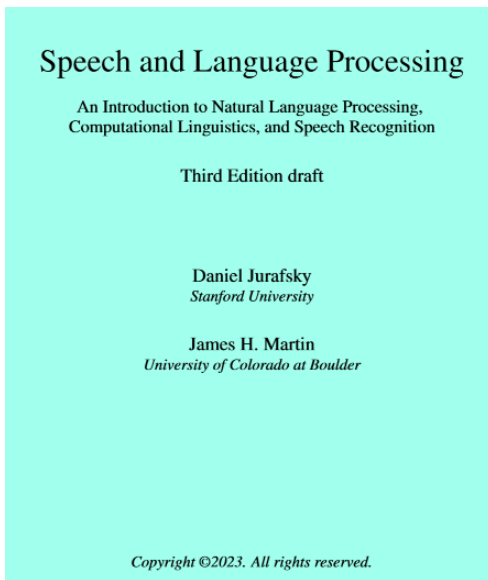
9. RNNs and LSTMs

10. Transformers and Pretrained Language Models

11. Fine-Tuning and Masked Language Models

Introduction au TALN

Cours



I. Fundamental Algorithms for NLP

1. Introduction

2. Regular Expressions, Text Normalization, Edit Distance

2.1. Regular Expressions

2.2. Words

2.3. Corpora

2.4. Text Normalization

2.5. Minimum Edit Distance

Fin de la séance N° 1