

# Université de Montréal

---

IFT6285 (TALN) — A2023  
Plan

---

Contact :  
**Philippe Langlais**  
RALI/DIRO  
Université de Montréal

+1 514 343 61 11 ext: 47494  
[felipe@iro.umontreal.ca](mailto:felipe@iro.umontreal.ca)  
<http://www.iro.umontreal.ca/~felipe/>

■ dernière compilation : 21 septembre 2023 (11:23)

## Modus operandi

---

- Les ressources seront déposées sur Studium / *resources will be available on Studium* : <https://studium.umontreal.ca/course/view.php?id=247156>
- À priori, le cours sera donné en français, mais tous les énoncés de sujets (devoirs, projets et examen) seront également disponibles en anglais. / *I will likely teach in French, but all assignments (devoirs, projects and exams) will also be available in English.*
- Le cours du mercredi sera décomposé en deux parties : une présentation invitée dans le cadre des [séminaires RALI-OLST](#) (attention la salle peut changer d'une présentation à l'autre et se tient habituellement en salle 3195, Pav. André-Aisenstadt, ou en salle C-9019, Pav. Lionel-Groulx), suivie du cours en salle B-3270 Pav. 3200 J.-Brillant. Les semaines où il n'y a pas de séminaire, nous commençons le cours à 11h30. / *Wednesday's class will be divided into two parts : an invited presentation as part of the [RALI-OLST seminars](#) (note that the room may change from one presentation to the next, and is usually held in room 3195, Pav. André-Aisenstadt, or room C-9019, Pav. Lionel-Groulx), followed by the course in room B-3270 Pav. 3200 J.-Brillant. Course starts at 11h30 on Wednesdays without seminars.*
- Un espace de travail slack a été créé afin que vous puissiez communiquer avec moi : <https://ift6285-A23.slack.com/>. Je suis également joignable par [mail](#) si vous prenez soin de préfixer l'objet de votre message de IFT6285. Je ne lirai que sporadiquement les messages envoyés sur Studium. / *A slack workspace has been created, which is the best way to reach me : <https://ift6285-A23.slack.com/>. I can also be reached by [mail](#) if you prefix the subject of your message with IFT6285. I will only sporadically read messages sent to Studium.*
- Les remises des travaux sont à faire sur Studium (des consignes seront données à cet effet). Les devoirs sont à remettre en groupe de deux ou seuls, les projets seront remis en groupes de deux ou trois personnes. / *All assignments will have to be submitted on Studium (more about this*

later). *Devoirs must be done individually or in groups of two. Projects should be done in groups of 2 to 3 personnes.*

— Une connaissance de base de Python est attendue. / *A familiarity with Python is expected.*

## Matière / *Agenda*

---

Plan approximatif de notre session. Pour chaque sujet, différentes approches seront discutées (pas seulement neuronales) ainsi que leurs limites ; les principaux jeux de données seront présentés. Le titre des devoirs et projets est indicatif seulement. / *Approximative plan of this semester. For each topic, different approaches (not only neural) and their limitations will be discussed ; the main datasets will be presented. The titles of the assignments and projects are indicative only.*

1. Introduction
2. Mains dans le cambouis / *Dirty hands*
  - quels langages de programmation / *which programming language ?*
  - compter des mots / *counting words*
3. Modèles de langue / *Language models*
  - modèles discrets / *count-based models*
  - modèles neuronaux / *neural models*
  - **devoir 1** : training and testing a model with kenlm
4. Approches distributionnelles / *Distributional Approaches*
  - vector space model, plongement de mots / *word embeddings*
  - **devoir 2** : Word embeddings with gensim

**Projet 1** : Kaggle like task

5. Algorithmique du texte / *Text Algorithmic*
  - comparaison de chaînes / *comparing strings* : distance d'édition / *edit-distance*, soundex, etc.
  - recherche rapide / *fast search* : tables de suffixes / *suffix arrays*, FM-index
6. Étiquetage séquentiel / *Sequential Labelling*

- étiquetage morpho-syntaxique / *Part-of-speech (POS) tagging*, chunking
- reconnaissance d'entités nommées / *named-entity recognition (NER)*
- **devoir 3** : NER with Spacy

7. Analyse syntaxique / *Parsing*

- théorie des langages / *language theory*
- analyse en constituant et en dépendance
- **devoir 4** : parsing with NLTK

**Projet 2** : On fair analysis of NLP performance

8. Extraction d'information / *Information Extraction*

- extraction de patrons
- extraction ouverte de relations

9. Traduction automatique / *Automatique translation*

- alignement de phrases
- systèmes de traduction
- évaluation
- **devoir 5** : Translating with Huggins Face

10. Répondre à des questions / *Question Answering*

- techniques
- biais

## Évaluation

---

- **5 devoirs** / *5 assignments* : 5 x 6
  - remise environ une semaine après l'énoncé en groupe d'au plus deux personnes / *to be completed within a week or so by groups of maximum 2 persons.*
  - des consignes claires accompagneront chaque devoir. Le respect de ces consignes conditionnera votre note. / *Respect to specific requirements will condition your score.*
- **2 projets** / *2 projects* : 20 x 2
  - remise sur Studium en groupe de 2 à 3 personnes / *to be submitted on Studium in group of 2 to 3 persons*
  - un mois environ / *a month or so*
  - Vos projets seront notés d'après un rapport (pdf) de quelques pages. La note dépendra notamment : / *Projects will be scored thanks the companion report. We will search for :*
    - de votre curiosité / *curiosity*
    - de la clarté de votre rapport / *clarity*
    - du contenu informationnel / *substance*
- **examen** final / *final exam* : 30
  - question de compréhension sur des concepts vus en cours ou abordés dans les séminaires / *Comprehension questions on subjects discussed in course or seminars*