

# PROJET DE FIN DU MODULE « NATURAL LANGUAGE PROCESSING »

## MINI PROJET: TRANSLATION

MASTER : Ingénierie de Données et Développement  
Logiciel TA 2021

MODULE : Natural Language Processing and Applications

Réalisé par : Zakaria Benzeroual - Yassine Timmi

Encadré par : Pr. Abdelhak Mahmoudi

# Introduction

L'un des premiers objectifs des ordinateurs était la traduction automatique de texte d'une langue à une autre.

La traduction automatique est peut-être l'une des tâches d'intelligence artificielle les plus difficiles étant donné la fluidité du langage humain. Classiquement, des systèmes basés sur des règles étaient utilisés pour cette tâche, qui ont été remplacés dans les années 1990 par des méthodes statistiques. Plus récemment, les modèles des réseaux de neurones.

Dans une tâche de traduction automatique, l'entrée consiste déjà en une séquence de symboles dans une langue, et le programme informatique doit la convertir en une séquence de symboles dans une autre langue.

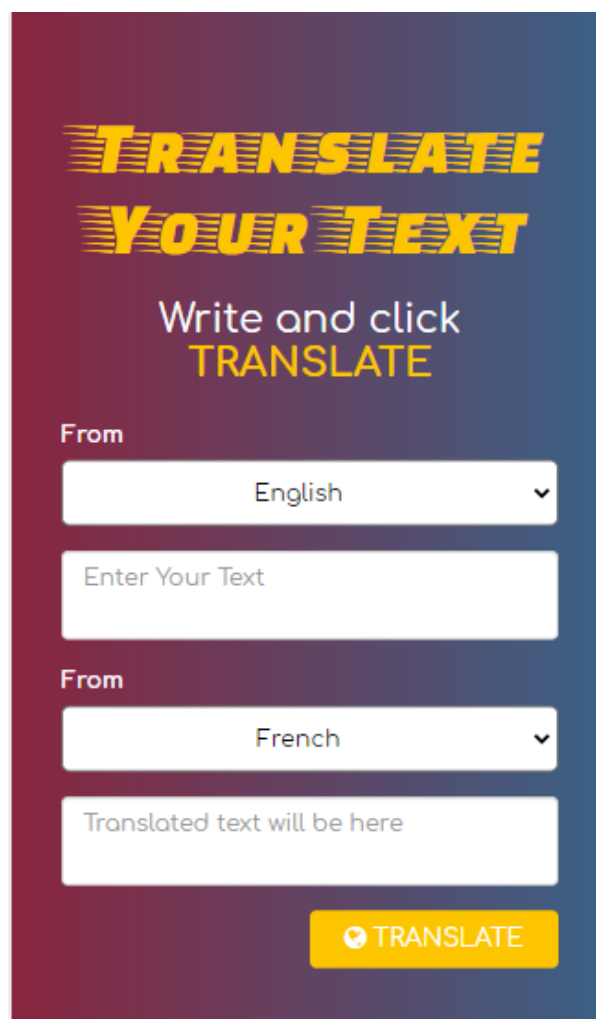
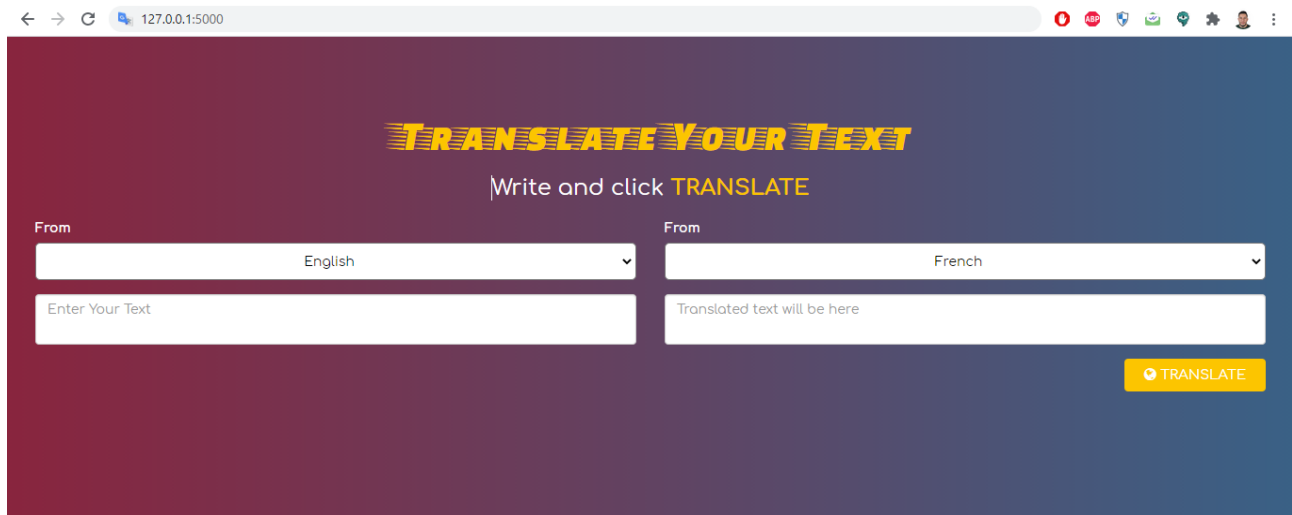
Étant donné une séquence de texte dans une langue source, il n'existe pas de meilleure traduction de ce texte dans une autre langue. Cela est dû à l'ambiguïté et à la flexibilité naturelles du langage humain, en particulier avec la plupart des langues indiennes qui sont rarement parlées dans d'autres pays.

Traditionnellement, le traitement du langage naturel de la langue parlée et écrite passe par les étapes suivantes :

- Phonologie et phonétique : (Traitement du son)
- Morphologie : (Traitement des formes de mots)
- Lexique : (Stockage des mots et connaissances associées)
- Analyse : (Traitement de la structure)
- Sémantique : (Traitement du sens)
- Pragmatique : (Traitement de l'intention utilisateur, modélisation, etc.)
- Discours : (Traitement du texte connecté)

# Création de l'application web

Pour arriver à notre objectif, qui a la réalisation d'une application de traduction nous avons pensé à la création d'une application web connecté à notre librairie qui sert à la traduction d'un texte en français vers l'anglais et vice versa.



# Notebook

Dans cette étape nous allons projeter la partie Flask, Flask est un micro framework open-source de développement web en Python.

## *Templates en Flask*

- Flask utilise le système de *templates* [Jinja2](#), qui permet de générer n'importe quel format textuel (HTML, CSS...).
- Les templates sont stockés dans un répertoire `templates`, situé dans le même répertoire que le fichier Python définissant l'application.

Pour appliquer un *template*, il suffit d'appeler la fonction `flask.render_template` en lui passant + le nom du template (relatif au répertoire `templates`), et + la liste des variables utilisées par le *template* :

```
translated.py > process
1  from flask import Flask, render_template, request, url_for
2  from transformers import MarianTokenizer, MarianMTModel
3
4  app = Flask(__name__)
5
6  @app.route('/')
7  def index():
8      return render_template('index.html')
9
10 @app.route('/translated', methods=['POST', 'GET'])
11 def process():
12     if request.method == "POST":
13
14         original_select = request.form['original_select']
15         original_text = request.form['original_text']
16         translated_select = request.form['translated_select']
17
18         if original_select != translated_select and original_text != "" :
```

# Hugging Face Translation models: Helsinki-NLP

Plus récemment, **Huggingface** a publié plus de 1 000 modèles linguistiques pré-entraînés de l'Université d'Helsinki . Pour tous ceux qui cherchent à créer leur propre API AWS ou Google translate, cela n'a jamais été aussi simple. Donc, j'ai pensé que je capitaliserais sur le travail acharné des autres. C'est l'équivalent fonctionnel de « enveloppons l'apprentissage automatique dans une API.

Traduction automatique avec des transformateurs;

**Huggingface** a fait un travail incroyable en rendant des modèles SOTA (état de l'art) disponibles dans une simple API Python pour les codeurs copier + coller comme moi. Pour traduire du texte localement, il vous suffit de `pip install transformers` puis d'utiliser l'extrait ci-dessous à partir de la documentation des transformateurs .

[https://huggingface.co/transformers/model\\_doc/marian.html](https://huggingface.co/transformers/model_doc/marian.html)

```
#bibliotheque Hugging Face

mname = f'Helsinki-NLP/opus-mt-{src}-{trg}'

model = MarianMTModel.from_pretrained(mname)
tok = MarianTokenizer.from_pretrained(mname)

gen = model.generate(**tok.prepare_seq2seq_batch(translated_text, return_tensors="pt"))
words = tok.batch_decode(gen, skip_special_tokens=True)

# translated Text
return render_template('index.html', translated_text=words[0], original_text=original_text)
```

## Helsinki-NLP/opus-mt-en-fr

← → ↻ 127.0.0.1:5000/translated

**TRANSLATE YOUR TEXT**

Write and click **TRANSLATE**

From	From
English	French
Natural language processing is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and	Le traitement du langage naturel est un sous-domaine de la linguistique, de l'informatique et de l'intelligence artificielle qui s'intéresse aux interactions entre les ordinateurs et le langage humain
<b>TRANSLATE</b>	

## Helsinki-NLP/opus-mt-fr-en

← → ↻ 127.0.0.1:5000/translated

**TRANSLATE YOUR TEXT**

Write and click **TRANSLATE**

From	From
French	English
A quoi sert le Traitement Automatique du Langage Naturel (TAL) ?	What is the Language Processing Natural (NLP)?
<b>TRANSLATE</b>	

## Helsinki-NLP/opus-mt-ar-fr

← → ↻ 127.0.0.1:5000/translated

**TRANSLATE YOUR TEXT**

Write and click **TRANSLATE**

From	From
Arabic	French
هل تتحدث اللغة العربية؟	Parlez-vous la langue arabe?
<b>TRANSLATE</b>	

127.0.0.1:5000/translated

TRANSLATE YOUR TEXT

Write and click TRANSLATE

From

Arabic

هل تتحدث اللغة العربية؟

From

English

Do you speak Arabic?

TRANSLATE

6

## Conclusion

La traduction automatique est principalement un équipement qui aide les spécialistes ou les traducteurs à accomplir un motif. Cependant, ce n'est pas un remplacement pour les anciens systèmes de traduction, c'est plutôt une modification au fond. De nombreux traducteurs automatiques offrent considérablement leurs services gratuitement, ce qui les rend beaucoup plus attrayants.

