

# Sujet de stage de master recherche

## Génération de paraphrases : Étude d'une mesure de distance sémantique basée sur un *LLM*

Contacts : quentin.lemesle@irisa.fr, jonathan.chevelu@irisa.fr, arnaud.delhay@irisa.fr

**Mots clefs** : Apprentissage profond ; Grands modèles de langue ; Évaluation sémantique ; Évaluation automatique ; Traitement automatique des langues ; Développement Python.

### 1 Objet du stage

Ce stage a pour but de poursuivre un travail de recherche existant sur une mesure de proximité sémantique entre deux phrases. Cette mesure, ParaPLUIE [3], repose sur l'utilisation d'un grand modèle de langue (*LLM*). ParaPLUIE a été évaluée sur deux corpus en langue anglaise de paraphrases annotées par des humains, MRPC [1] et PAWS [5]. Le modèle de langue Mistral [2], affiné en tant qu'agent conversationnel, a été utilisé pour cette évaluation. L'objectif principal de ce stage est d'évaluer ParaPLUIE avec d'autres configurations, en changeant le *LLM* utilisé et les données d'évaluation.

### 2 Contexte et motivation

Au cours de la dernière décennie, le développement de l'apprentissage profond a permis une amélioration conséquente de l'état de l'art pour de nombreuses tâches du traitement automatique des langues. C'est en particulier le cas pour de nombreuses tâches de génération de texte comme la traduction automatique, la génération de texte, l'identification d'entités nommées, *etc.*

Ces modèles semblent avoir la capacité de capturer des relations fines comme le sens. Ces relations, par nature ambiguës, ne sont pas descriptibles par un ensemble de règles. Pour évaluer la conservation du sens d'une phrase après transformation, comme dans la tâche de génération de paraphrase, l'utilisation d'un tel modèle semble être le plus pertinent (BertScore [4]).

Obtenir une mesure d'évaluation sémantique fiable a de nombreuses retombées positives. Cela permet d'évaluer la qualité d'un système de génération de paraphrase. Or ces systèmes sont cruciaux pour une tâche d'augmentation de données, ce qui est très demandé par la communauté scientifique du traitement automatique du langage (*TAL*). De plus, pouvoir identifier automatiquement des messages ayant le même sens mais écrit différemment est crucial dans le cadre de la lutte informatique d'influence (*L2I*).

### 3 Missions

Pour répondre aux objectifs de ce stage la/le candidat(e) devra :

- Prendre en main l'implémentation existante de ParaPLUIE.

- Prendre en main l'implémentation existante d'une toolchain permettant de passer une batterie de mesures automatiques sur un ensemble de corpus de couple de phrases. Possiblement, adapter et améliorer cette toolchain.
- Identifier et utiliser d'autre *LLM* en tant que modèle sémantique pour ParaPLUIE.
- Identifier, préparer et utiliser des corpus pour évaluer ParaPLUIE
- Identifier et utiliser d'autre mesures automatiques de l'état de l'art afin de les utiliser comme références d'évaluation.

*Dans l'idéal, nous aimerions travailler avec des données en français.*

- Rédiger un document récapitulatif des expériences.  
*Nous souhaiterions que cela aboutisse sur un article scientifique rédigé par la/le candidat(e).*
- Possiblement, étudier la contamination des *LLM* utilisés sur les corpus.

## 4 Compétences requises

- Connaissance du langage Python : *La/le candidat(e) devra être à l'aise avec le Python car c'est le langage que nous utilisons très majoritairement.*
- Notion en apprentissage profond : *La/le candidat(e) sera amené(e) à utiliser et à modifier légèrement le comportement de modèle d'apprentissage profond.*
- Notion en système Linux : *La/le candidat(e) sera amené à utiliser des machines de calcul distantes utilisant différentes variantes d'Ubuntu.*
- Rigueur : *La/le candidat(e) réalisera de nombreux tests et mesures. Les résultats devront être reproductibles, documentés et clairs.*
- Curiosité : *C'est un incontournable pour faire de la recherche.*

## Références

- [1] W. B. Dolan and C. Brockett. Automatically constructing a corpus of sentential paraphrases. In *Proceedings of the Third International Workshop on Paraphrasing (IWP2005)*, 2005.
- [2] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. d. I. Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, et al. Mistral 7b. *arXiv preprint arXiv :2310.06825*, 2023.
- [3] Q. Lemesle, J. Chevelu, D. Lolive, A. Delhay-Lorrain, and P. Martin. Parapluie - une mesure automatique d'évaluation de la qualité sémantique des systèmes de paraphrases. In *Actes de JEP-TALN-RECITAL 2024. 31ème Conférence sur le Traitement Automatique des Langues Naturelles, volume 1 : articles longs et prises de position*, pages 605–616, Toulouse, France, 7 2024. Association pour le Traitement Automatique des Langues.
- [4] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi. Bertscore : Evaluating text generation with bert. In *International Conference on Learning Representations*, 2020.
- [5] Y. Zhang, J. Baldridge, and L. He. PAWS : Paraphrase Adversaries from Word Scrambling. In *Proceedings of North American Chapter of the Association for Computational Linguistics*, 2019.