



Stage de développement en Informatique

Implémentation d'une mesure de richesse du discours

Contexte

Dans le cadre du projet de recherche ALAIA¹ (Apprentissage des Langues Assisté par Intelligence Artificielle), nous travaillons actuellement sur l'implémentation d'un moyen d'évaluation de la compréhensibilité de la parole de locuteurs non natifs. La compréhensibilité de la parole est souvent définie comme étant « la facilité de compréhension du locuteur » dans le domaine de l'apprentissage des langues. Elle représente ainsi un objectif capital pour des apprenants d'une langue seconde, et repose sur l'utilisation de divers paramètres phonético-phonologiques et linguistiques, tels que la prononciation, la fluence, la richesse lexicale, la complexité syntaxique, la cohésion du discours [1] et la richesse du discours.

La richesse du discours désigne le fait qu'un énoncé est riche et sophistiqué. Elle peut être considérée « basse » si la narration est simple, non nuancée et manque d'idées sophistiquées ou de détails, et « haute » si le locuteur produit plusieurs idées distinctes ou détails, rendant ainsi le contenu développé et sophistiqué [2].

Objectif du stage

L'objectif de ce stage est de développer une mesure de richesse du discours qui pourra ensuite être utilisée pour prédire la compréhensibilité de la parole.

Dans un premier temps, l'étudiant devra s'approprier le sujet en effectuant un état de l'art sur la définition de la richesse du discours et les manières de la mesurer dans différents domaines (parole non native, parole pathologique, ...). Ensuite, la mesure sera implémentée et testée sur un corpus oral de description d'images dont dispose l'équipe SAMoVA.

Compétences scientifiques et techniques

Traitement automatique des langues, programmation (Python), gestion de version de code (Github). Des connaissances en traitement automatique de la parole seraient un plus.

Autres compétences

Aptitude pour le travail en équipe, bonnes pratiques de développement informatique, rigueur dans la lecture d'articles, bon niveau en anglais.

Date de début du stage : à partir de mars 2023

Durée : 4 à 6 mois suivant niveau (M1 ou M2)

Rémunération : gratification selon les tarifs en vigueur

¹ <https://www.irit.fr/SAMOVA/site/projects/current/labcom-alaia/>

Encadrement : Verdiana De Fino (IRIT)

Candidatures : envoyer un mail avec CV détaillé et lettre de motivation à verdiana.defino@irit.fr

Références

- [1] Verdiana De Fino, Lionel Fontan, Julien Pinquier, Isabelle Ferrané, Sylvain Detey. Prediction of L2 speech proficiency based on multi-level linguistic features. *23rd INTERSPEECH Conference : Human and Humanizing Speech Technology (INTERSPEECH 2022)*, The Acoustical Society of Korea, Sep 2022, Incheon, South Korea. (hal-03775950)
- [2] Dustin Crowther, Pavel Trofimovich, Talia Isaacs, Kazuya Saito (2015). Does a Speaking Task Affect Second Language Comprehensibility?. *Modern Language Journal*. 99. 80-95. 10.1111/modl.12185.