

Invitation à la soutenance publique de thèse Pour l'obtention du grade de Docteur en Langues, lettres et traductologie

En cotutelle avec la KU Leuven

Madame Anaïs TACK

MARK MY WORDS! ON THE AUTOMATED PREDICTION OF LEXICAL DIFFICULTY FOR FOREIGN LANGUAGE READERS



Le vendredi 25 juin 2021
à 14H00

La soutenance sera retransmise en direct via Teams

*Les modalités pratiques seront notifiées
ultérieurement sur le [site](#)*

L'objectif de cette recherche doctorale est la prédiction automatique des mots difficiles dans un texte pour les locuteurs non natifs. Cette prédiction est cruciale car une bonne compréhension d'un texte est fortement déterminée par le vocabulaire. Si un texte contient un pourcentage élevé de mots inconnus, le lecteur aura probablement des difficultés à comprendre le texte. Afin de fournir un bon soutien au lecteur de langue étrangère, nous devons d'abord être en mesure de prédire le nombre de mots difficiles. En général, nous le faisons manuellement en nous basant sur notre expertise ou sur des tests de vocabulaire antérieurs. Cependant, ces méthodes ne sont pas pratiques lorsque nous lisons dans un environnement informatique tel qu'une tablette ou une plateforme d'apprentissage en ligne. Dans ces cas, nous devons automatiser correctement les prédictions.

La thèse est divisée en trois parties. La première partie contient un examen systématique de la littérature scientifique pertinente. La synthèse comprend 50 ans de recherche et 140 publications évaluées par des pairs. Les analyses mettent en évidence certaines limites cruciales, notamment le fait que le champ scientifique est divisé en deux domaines de recherche peu liés. La deuxième partie examine deux mesures de la difficulté lexicale pour les lecteurs non natifs. D'une part, les résultats montrent qu'il existe des incohérences importantes dans la manière dont les mots sont introduits dans les matériels de lecture pour le néerlandais et le français étiquetés avec les niveaux du CECR. Par conséquent, cette mesure de difficulté ne semble pas encore valable comme base pour un système automatisé. D'autre part, des données ont été recueillies sur la façon dont les locuteurs non natifs perçoivent les mots difficiles pendant la lecture. Cette mesure de difficulté est appropriée pour développer un système personnalisé et contextualisé. La dernière partie s'intéresse à deux types de modèles prédictifs développés sur ces données, à savoir les modèles à effets mixtes et les réseaux de neurones artificiels. D'une part, les résultats montrent clairement qu'un modèle personnalisé fait des prédictions nettement meilleures qu'un modèle non personnalisé. D'autre part, les résultats montrent qu'un modèle contextualisé peut mieux discriminer la difficulté, bien que ces améliorations ne soient pas toujours significatives pour chaque apprenant.

Le jury est composé de Mesdames et Messieurs les professeur-es :

Magali PAQUOT (UCLouvain), Présidente du jury et lectrice
Cédric FAIRON (UCLouvain), Promoteur et Secrétaire du jury
Thomas FRANCOIS (UCLouvain), Promoteur
Piet DESMET (KU Leuven), Promoteur
Maribel MONTERO PEREZ (UGent), Lectrice
Núria GALA (Aix-Marseille Université), Lectrice extérieure
Detmar MEURERS (Universität Tübingen), Lecteur extérieur