# Use of Linguistic Features in Automatic Readability Assessment

## **Prof. John Sie Yuen LEE**City University of Hong Kong

Date: April 18, 2023 (Tuesday)

**Time:** 4:30pm - 6:15pm (Hong Kong Time, UTC +8)

Venue: Lee Shau Kee Building 302,

The Chinese University of Hong Kong



### **Abstract**

Recent studies in automatic readability assessment have shown that hybrid models --- models that leverage both linguistically motivated features and neural models --- can outperform neural models. However, most evaluations on hybrid models have been based on in-domain data in English. This paper provides further evidence on the contribution of linguistic features by reporting the first direct comparison between hybrid, neural and linguistic models on cross-domain data. In experiments on a Chinese dataset, the hybrid model outperforms the neural model on both in-domain and cross-domain data. Importantly, the hybrid model exhibits much smaller performance degradation in the cross-domain setting, suggesting that the linguistic features are more robust and can better capture salient indicators of text difficulty.

#### **Speaker**

Dr. Lee is an Associate Professor at the Department of Linguistics and Translation at City University of Hong Kong. He received his PhD from the Massachusetts Institute of Technology (MIT) in Computer Science in 2009. Dr Lee's research focus is on natural language processing (NLP) and computational linguistics, especially their applications in computer-assisted language learning. His recent projects include automatic readability assessment, reading material recommendation and exercise generation for language learning, and Cantonese chatbot development. His research projects have been supported by the Innovation and Technology Fund, Health and Medical Research Fund, General Research Fund, and Language Fund from the Standing Committee on Language Education and Research.

#### All are Welcome