Editorial

We are pleased to present the third issue of the seventeenth volume of the ***International Journal of Computational Linguistics Research*****,** featuring the papers detailed below.

In the opening article, "**Demographic Inequality and Linguistic Diversity in Indian States and Union Territories**," the authors propose an integrated analytical framework to examine demographic inequality and linguistic diversity across India's 36 States and Union Territories. Using a comprehensive dataset, they create network based visualizations such as state language bipartite networks and co-occurrence maps to elucidate the structural patterns of multilingual coexistence. Furthermore, their linguistic analysis identifies 43 distinct languages exhibiting high lexical richness and equitable distribution, demonstrating minimal dominance by any single language.

The second paper, "**Analyzing Lexical Difficulty and Thematic Distribution in the TOEFL-CEFR Vocabulary Dataset: Implications for NLP and ESL Curriculum Design**," examines the comparative efficacy of semantic versus thematic lexical clustering on ESL learners' vocabulary acquisition, alongside the application of AI-driven lexical metrics. Additionally, the study explores the integration of NLP-based interventions designed to identify and disrupt habitual procrastination patterns, there by reducing stress among secondary and post secondary ESL students. By bridging psycholinguistic theory with modern educational technology, this research provides a robust framework for developing adaptive, evidence based vocabulary assessments and instructional tools in contemporary language education.

Finally, in "**Vocabulary Distribution, Lexical Diversity, and Statistical Properties of the Fact Verification Corpus,"** the authors present a comprehensive corpus linguistic analysis of a fact verification dataset. They examine its vocabulary distribution, lexical diversity, and adherence to natural language statistical regularities. Employing a multi metric framework, the study evaluates lexical richness using indices such as the Shannon and Simpson Diversity Indices, Type Token Ratio, MTLD, and HDD. The results indicate substantial lexical heterogeneity, with high diversity indices confirming a rich vocabulary inventory suitable for complex semantic reasoning. These findings validate the dataset's linguistic coherence and statistical suitability for downstream natural language processing applications, including transformer based language modeling, information retrieval, and retrieval augmented generation.

We hope these contributions will significantly advance the literature in computational linguistics.

Editors