Chapter 11
Analyzing Linguistic Differences in Academic English: A Graph-Based Machine Learning Approach

Krishna Kumari R.
https://orcid.org/0000-0002-1802-628X
Department of Mathematics, SRM Institute of Science and Technology, Kattankulathur, India

Ami Femila P.
Department of English and Foreign Languages, SRM Institute of Science and Technology, Kattankulathur, India

ABSTRACT

This study employs graph-based machine learning techniques to explore and analyze the linguistic differences between the English used by Indian and British academicians. By constructing language graphs from academic texts, we identify distinct patterns, syntactic structures, and semantic variations. These graphs represent sentences as nodes and relationships, such as co-occurrences or syntactic dependencies, as edges. Our approach leverages advanced community detection algorithms to uncover clusters of linguistically similar nodes and node centrality measures to identify key terms or structures that dominate regional writing styles. Additionally, we incorporate semantic network analysis to understand contextual nuances and word usage. By combining these techniques, we reveal underlying linguistic characteristics and trends, offering novel insights into regional variations in academic writing, which can have significant implications for cross-cultural communication, education, and

DOI: 10.4018/979-8-3693-7863-2.ch011

global academic collaborations.

1. INTRODUCTION

Linguistic variations within English, especially in academic writing, offer critical insights into how regional differences shape communication and comprehension. English, as a global language, has evolved into various forms influenced by cultural, historical, and social contexts. Among these, Indian English and British English stand out as two prominent varieties with distinctive characteristics.

Indian English has developed as a nativized variant, heavily influenced by the multilingual landscape of India. This variety reflects a blend of English with local languages, resulting in unique syntactic structures, vocabulary, and idiomatic expressions. For example, Indian English often incorporates terms and expressions derived from regional languages, and its syntax may differ from the standard British English patterns. These adaptations serve to address the specific communicative needs and cultural contexts of Indian speakers, resulting in a distinct variant that is both reflective of and adapted to its socio-linguistic environment. In contrast, British English adheres more closely to traditional norms and standards of the English language. It represents the standardized form of English used in the United Kingdom and often serves as the reference point for various aspects of English grammar, spelling, and pronunciation. British English maintains a level of uniformity that is less influenced by external languages compared to Indian English. As a result, the linguistic features of British English tend to be more consistent with historical and formal norms of English usage.

The differences between these two varieties of English are not merely academic curiosities; they have practical implications for communication, particularly in academic settings. Academic writing in Indian English may exhibit variations in clarity, tone, and readability compared to British English. These variations can affect how scholarly work is perceived, interpreted, and evaluated across different linguistic and cultural contexts. For instance, differences in idiomatic expressions or syntactic structures might influence the readability and comprehension of academic texts among international audiences. Traditional linguistic analysis has often relied on qualitative methods to study these variations, focusing on descriptive and interpretive approaches. However, the advent of machine learning and computational linguistics has introduced new opportunities for more systematic and quantitative analysis. Graph-based machine learning techniques represent a particularly promising approach for this type of analysis. By employing graph representations, these techniques enable researchers to model and analyze complex linguistic structures and patterns in a more comprehensive and scalable manner.