Corpus Linguistic Analysis for Language Planning

Joel Ilao
De La Salle University
`joel.ilao@dlsu.edu.ph`

Abstract

The Filipino language, the Philippines' national language, is based on the lingua franca of Manila, the nation's capital. It has shown rapid and significant changes in its vocabulary, orthography and grammar, thanks to the Philippines' rich colonial history and the conscious efforts in the national and institutional levels to standardize the grammar and orthography of the language. This fact is proof that the socio-political landscape of the times shapes the language of its people. In this talk, I will describe my forays in the analysis and processing of text corpora largely written in the Filipino language. Through corpus linguistic analysis of a historical text collected from various publication domains and genres, and covering approximately one century, I will illustrate some observable long-term and short-term trends in Filipino writers' writing styles, and conduct correlational analysis with some notable Philippine Socio-linguistic trends. Acknowledging that language plays a vital role in the formation of a national identity, and hence, should be cultivated towards intellectualization, I propose that ICT could effectively be used to monitor language usage and provide added insights for language planners, in understanding the interplay between language and socio-political developments. Thus, for this talk, I will also describe various ways by which corpora could automatically be harnessed and analyzed, and used to identify areas of language use that could be highlighted or further improved. Practical insights in large scale text analysis would also be provided; the depictions would all be drawn from my experiences as a corpus linguistic researcher.