

## New Language Resources for Cantonese Linguistics Research

Andy C. Chin

Department of Linguistics & Modern Language Studies

Cantonese is the lingua franca in Hong Kong, Macau and most part of Guangdong Province of China. This vernacular language however has not developed a formal and proper writing system. The study of Cantonese thus poses a lot of challenges and has to rely on spoken data.

Currently, there is a relatively rich source of Cantonese language materials published between the 19<sup>th</sup> and mid-20<sup>th</sup> centuries. These materials allow us to reconstruct the Cantonese language of about 200 years ago, and reveal a number of significant linguistic changes that had taken place. Unfortunately, there is a critical gap in the provision of comparable language data after the 1950s, both qualitatively and quantitatively, for further examining the development of the Cantonese language.

To bridge this gap in Cantonese diachronic research, an initial attempt was thus made to construct an annotated corpus and an online search engine by transcribing one type of authentic and natural spoken data that has not received serious attention in previous Cantonese linguistics research - *early Cantonese movies* produced in Hong Kong between the 1950s and the 1970s. Considering the production practice of Hong Kong movie industry in the mid-20<sup>th</sup> century, the spoken data collected from these early Cantonese movies can be claimed to largely reflect and represent the actual use of the language of the period concerned.

Besides providing additional language resources for Cantonese research, this corpus also serves the purpose of documenting a *spoken* language of the past, which has not been attempted before. The value of the corpus can be further enhanced not only by including more language data (e.g. movies with different genres), but also annotating the spoken data at linguistic and non-linguistic levels so that it can benefit research in other disciplines, such as discourse analysis, conversation analysis, language and gender, as well as the inter-relationship between language, society and culture.