

回歸前後《澳門日報》用語的量化分析

摘要：語言是不斷在變遷的，影響語言變化的因素有很多，歷史因素，時代因素，文化因素等諸多因素都有可能對一個地區的語言產生影響。澳門作為我國特別行政區之一，因其特殊的歷史、地理、經濟、政治等因素，成為了中西文化交融的典範，形成了特有的“三文四語”語言景觀，為語言的發展變遷提供了天然的土壤和研究樣本。此次研究中，通過《澳門日報》這一媒介，以澳門回歸前後不同階段的新聞文本為語料，搜集並建立語料庫。之後通過數據科學與電腦技術，利用分詞和演算法找到高頻詞和重要性高的詞，來對語料進行歷時性量化分析，並且結合時代因素、歷史因素、文化因素等方面，來分析回歸前後《澳門日報》用語的變化情況，從而探究中文對澳門地區語言變化的影響，並且通過量化分析，從數據科學的角度來驗證語言在會隨著兩個地區交流融合的增加而走向趨同的態勢這樣的觀點。

關鍵字：語言變化；澳門；數據科學；語料庫；量化分析

A Quantitative Analysis of Language Use of Macau Daily Before and After Macau' s Return

Abstract: Language is constantly changing. A lot of factors can attribute to language change, such as historical background, different time era throughout history, cultural development. All these factors can have an impact on the language change of a certain region. Macau as a Special Administrative Region of China, has become a model example of the fusion of eastern and western culture due to its specific historical background, geographical location, economical system and political status within China. Its “ Three Written Languages and Four Spoken Languages ” policy is the natural soil and study example of language change. In this project, we use Macau Daily as a media and the news text in throughout different stages of Macau' s Return as corpora. We start from collecting and constructing the corpora, and then conduct a diachronic quantitative analysis by using data science methods and computer science technology. We implemented tokenization and algorithm to find out words with high term frequency and high importance in the corpora. Combined with time era, historical and cultural factors, we can analyze how language use has changed in the news text in Macau Daily before and after Macau' s Return, therefore we can try to explore how Chinese Language has impacted on the language change in Macau and show how the increase of communication and interaction of two regions resulted in the convergence of their languages through data science methods.

Keywords: language change; Macau; data science; corpus; quantitative analysis