

法鼓佛教學院佛教學系碩士論文

**A Computer-Based Approach for Predicting the
Translation Period of the Early Chinese Buddhist
Translations:**

**Texts from the Eastern Han, Three Kingdoms
and Western Jin Periods**

利用分類演算法來判別早期漢譯佛典之朝代
—以東漢、三國、西晉為對象

指導教授：洪振洲博士

研究生：郭捷立

中華民國 一〇一年六月

Buddhism has been spreading in China for more than two thousand years since its first introduction in the Eastern Han dynasty (C.E. 25-221), and has become an important part of daily life and culture at large in China. A great number of Buddhist scriptures were translated from Indian originals starting from the Eastern Han dynasty to the Tang dynasty (C.E. 618-907) and beyond. Scholarship has become increasingly aware, over the last few decades, that traditional authorship and translatorship attribution of the early Chinese works is often unreliable. The current reference edition of the Chinese Buddhist canon (Taishō shinshū daizōkyō (abbr.: T.) 大正新修大藏經), collated 1924-1934) contains 3053 works in 85 volumes, including about 1000 texts of Indian (or alleged Indian) provenance. However, ca. 150 of these texts are marked as shiyi 失譯, indicating that the name(s) of the translator(s) are unknown. In addition to such unknown cases of attribution, for the texts that were translated between the 2nd and the late 6th century, many attributions are uncertain, problematic or simply incorrect.

Text-critical and philological studies have brought a significant advancement of the status of the research in the field. However, traditional philology has its scale limit. The research project the present thesis stems from has thus designed a statistic model employing variable-length n-gram, with Fisher Linear Discriminant, to establish a highly accurate classification mechanism for predicting the translation period of Chinese texts. The time brackets we focus on in the present study include three early Chinese dynasties: the Eastern Han (C.E. 25-220), the Three Kingdoms (C.E. 220-280) and the Western Jin (C.E. 266-316). These three dynasties constitute the earliest phase of Buddhist translation history and most of the translations from these periods present attribution problems. In this research, we build up classification mechanisms for each of the three dynasties. These can be used to test whether the translation style of a text is similar to the one prevalent during a certain period. According to the output of our experiment, all of the three classifications for three dynasties have an accuracy rate of more than 89%. Also, by examining the classification result, we extract the special translation usages of Chinese sutras in different time period. With the help of statistic information bearing on the characteristics and features of Chinese texts, this approach can not only provide new evidence relevant to uncertain authorships but also encourage Buddhist scholars and scholars of linguistics to do further studies.

Keywords: Ancient Chinese, Chinese Translation, Authorship Attribution, Fisher Linear Discriminant Analysis, Variable Length n-Gram Feature Extraction