

# 結合漢典古籍虛詞常見字與統計量化分析進行

## 漢譯佛典譯者風格辨別

謝承恩\*、洪振洲\*\*、馬德偉\*\*\*

### 摘要

佛教於東漢由印度傳入漢地之時，記載佛教義理的佛經主要仍為印度的梵語或是中亞語言所撰寫。由東漢至唐中葉的數百年間，佛教盛行於中國，因而有了大規模的佛經翻譯活動。透過這些佛經翻譯的活動，不僅產生了巨量的漢譯佛典，也進一步影響了中國的文化，甚至是語言的發展。但是，早期的佛經翻譯由於受到傳抄、戰亂、偽託等現象影響，使得譯者記錄出現許多錯誤，這也連帶造成相關研究者的困擾。為找出正確的佛典翻譯者，許多佛學研究者使用傳統文獻學之方式，提出新的證據與看法。然而傳統文獻學之研究方式十分倚賴人工判斷與處理，不僅耗時費工，且無法處理大量文獻資料。

在現今資訊科學的幫助下，以數位化資料及統計量化方法進行資料的比對分析，已是現今人文資訊學的一大趨勢。此類方法不僅可以進行大量資料比對，也能夠找出譯者風格的潛在因子，這是傳統文獻研究方式難以達成的工作。而此方面的研究，目前仍以英文文獻為主要研究對象，以類似方法運用於古代中文文獻的效果仍未被正確的評估。因此本研究採用常見的古代中文的虛詞作為風格特徵之評斷指標，並搭配多變量統計分析手法中的主元素分析法，進行譯者之風格分析。為評估本方法之效用，本研究設計多種不同的實驗情境。根據我們的實驗結果，此種方式能成功的區分出作譯者的翻譯風格，協助研究者進行譯者紀錄的判斷。

**關鍵字：**漢譯佛典、譯者風格量化分析、虛詞、主元素分析法

---

\* 法鼓佛教學院佛學資訊組研究生，E-mail: chengen.xie@ddbc.edu.tw。

\*\* 法鼓佛教學院圖書資訊館館長，E-mail: jenjou.hung@ddbc.edu.tw。

\*\*\* 美國天普大學助理教授，E-mail: <http://mbingenheimer.net>。

# **Authorship Attribution of Early Chinese Buddhist Translations: Using Principal Component Analysis with Commonly Used Ancient Chinese Empty Words**

Cheng-en Xie<sup>\*</sup> 、 Jen-jou Hung<sup>\*\*</sup> 、 Marcus Bingenheimer<sup>\*\*\*</sup>

## **Abstract**

The Taishō edition of the Chinese Buddhist canon (1924-1932) contains ca. 1000 Indian texts translated into Chinese between the 2<sup>nd</sup> and the 11<sup>th</sup> century CE. Up to 153 of these texts are marked as 失譯, indicating that the name(s) of the translator(s) are unknown. Consequently, translator attributions from the beginning of the Tang dynasty, i.e., from the early 7<sup>th</sup> century onwards, are relatively reliable, but those for texts translated between the 2<sup>nd</sup> and the late 6<sup>th</sup> century have uncertain, problematic, or wrong attributions. We are reasonably confident that the 49 sutras ascribed to Kumārajīva (344-413 CE) were actually produced by him and his team (although this needs further research), but the attributions for many of the other texts translated before the late 6<sup>th</sup> century likely wrong.

Over the years, Buddhist scholars have leveraged traditional text-critical methods to corroborate or dispute traditional attributions. Although these methods can produce high quality results, they often rely heavily on the intuition of a single scholar honed over many years of research.

Information technology offers an alternative vector of inquiry that aims to complement rather than supersede the more traditional approaches. To accomplish this, statistical methods, quantitative methods, and artificial intelligence algorithms were adopted to analyze ancient Buddhist texts

---

\* Graduate Research Assistant, Dharma Drum Buddhist College, New Taipei City, Taiwan. E-mail: chengen.xie@ddbc.edu.tw

\*\* Assistant Professor, Director of Library and Information Center, Dharma Drum Buddhist College, New Taipei City, Taiwan. E-mail: jenjou.hung@ddbc.edu.tw

\*\*\* Assistant Professor, Temple University, US. E-mail: <http://mbingenheimer.net>

translated into Chinese to discover new evidence to address translator attribution.

The major advantage of stylometrics and quantitative authorship attribution is the capability to discover hidden patterns that cannot be discerned through traditional approaches. In the past four decades, considerable attention has been focused on the quantitative authorship attribution of literature in Western languages. However, few attempts have focused on texts written in classical Chinese, much less to the “Indian Buddhist Chinese” of early-translated texts. In the present paper, the focus is on the grammatical particles (*xuci* 虛詞) widely used in classical Chinese to express grammatical relations. After measuring their occurrence in Indian Buddhist Chinese, principle component analysis is employed to determine how their use reflects the authorship of some selected sutras, especially the three sutras attributed to Zhu Fahu 竺法護 (trad. 231-308) the *Xuzhen tianzi jing* 須真天子經 (T. 588), a version of the famous Lotus Sutra the *Zhengfahua jing* 正法華經 (T. 263), and the *Puyao jing* 普曜經 (T. 186). We have developed an algorithm to help describe and distinguish the translation style of different translators. The analysis explores different scenarios that need to be accounted for, such as the changes in translation style during the course of a translator’s career, an understanding of the commonalities between contemporaneous translations, and quantification the differences between different translations of the same sutra.

**Keywords: Chinese translation of Buddhist text, quantitative authorship attribution, empty words, Principle Component Analysis**