

中華佛學學報第 19 期 (pp.107-146)：(民國 95 年)，  
臺北：中華佛學研究所，<http://www.chibs.edu.tw>  
Chung-Hwa Buddhist Journal, No. 19, (2006)  
Taipei: Chung-Hwa Institute of Buddhist Studies  
ISSN: 1017-7132

## 巴利數位文獻資源現況評述與未來展望 ——兼談「初期佛教聖典多語多本平行語料庫」 的幾點構想<sup>[\*]</sup>

蔡奇林

南華大學宗教學研究所兼任助理教授

### 提要

近十數年來，隨著資訊科技的快速發展，巴利文獻的數位化工程也取得了大量業績。這些業績的概況，筆者曾撰文報導，本文第一部份也作簡要回顧。

接著，本文第二部分重點地省察當前這些數位資源的幾個側面，包括其發展的趨勢特點、提供的學術研究效益、以及仍需改善的地方。總括地說，這些數位資源由於在原典與譯本的「量」、校訂的「質」、以及檢索與擷取的「效能」等方面不斷昇進，因而具有提昇研究「精審度」的效益，但此種效益只達到「低精審」的規格。

本文最後部份，以案例方式，展望巴利數位資源未來的可能發展。其中提出一個以「初期佛教聖典」為範圍的擬構案例，顯示如何通過建構一個結合「多語言」、「多版本」（傳本、譯本），並標注「語言訊息」的「平行語料庫」，而突破當前巴利數位資源的規格，向「高精審」邁進。

這種「多語多本」的「平行語料庫」，除了可以提昇傳統佛教學研究的「精審度」之外，也可以支持佛典的翻譯研究、翻譯實踐，並記錄翻譯過程的語言理解。此外，它更具有「語言研究」與「語言教學」的功能特點，通過它，可以編纂更好的語言教材、語言工具書，可算是一種具備多功用、多效能的「多價資料庫」，也是數位佛典資源未來發展的一個可行方向。

p. 108 ↴

**關鍵詞：**1.巴利語 2.巴利文獻 3.數位資源 4.電子佛典 5.初期佛教 6.語料庫 7.多語平行語料庫

# 【目次】

## 一、序言

## 二、巴利數位文獻資源的現況

### (一) 巴利原典光碟版

### (二) 巴利原典網路版

### (三) 巴利原典翻譯

### (四) 其他工具資源

#### 1. 網路上的巴英辭典

#### 2. 網路上的巴利文法

#### 3. 網路上的巴利讀本

#### 4. 網路上的巴利有聲資源

## 三、略評當前的巴利數位文獻資源——其發展趨勢、效益、與限制

### (一) 當前巴利數位資源發展的幾項趨勢特點

### (二) 當前巴利數位資源所提供的學術研究效益

### (三) 當前巴利數位資源的不足之處

## 四、巴利數位文獻資源的未來展望——略談「初期佛教聖典多語多本平行語料庫」

### (一) 從「低精審」邁向「高精審」

### (二) 一個擬構案例：「初期佛教聖典多語多本平行語料庫」

#### 1. 略說「初期佛教聖典·多語多本·平行·語料庫」

[2. 必須進行的相關基礎工程](#)

[3. 能夠開發的教學研究效能](#)

[五、結語：建構一個更多效能的佛典知識寶庫](#)

p. 109

## 一、序言<sup>[1]</sup>

早在三十幾年前的七〇年代，英國巴利聖典學會（Pali Text Society，PTS）就著手從事巴利文獻的數位化工作，但當時由於人工輸入及校訂的成本太高，而光學字體辨識系統（OCR）又不夠精確（只達 95%），最後只能放棄計劃。<sup>[2]</sup>到了二十年前，泰國法身寺也開始其巴利三藏的數位化工作，可惜當時工作環境極為簡陋（只有一部電腦），進展十分緩慢。但是近十幾年來，情況已經大大改觀。隨著資訊科技與網路科技的猛迅發展，巴利數位文獻資源的建設，如雨後春筍一般，迅速勃發，並且已經取得了相當顯著的進展，成績斐然。本文將在第二節簡要回顧這些業績。<sup>[3]</sup>

其次，對於這樣快速而又大量的成績，實有必要進行一番綜合省察，以便了解其發展的趨勢特點，評估其能夠提供的學術研究效益，以及反省其存在的問題與限制，以便為未來的發展，提供一些參考座標。這是本文第三節的重點工作。

基於以上省察，本文將在第四節對於巴利數位資源的未來發展，提供一些可能的思考，焦點將放在「提高精審度」以及「開發更多價的研究教學效能」之上。其中提出一個擬構案例——「初期佛教聖典多語多本平行語料庫」，以便作為具體體現這種思惟的參考樣例。

## 二、巴利數位文獻資源的現況

### （一）巴利原典光碟版

目前巴利藏光碟共有 PTS 版（歐洲版）的 PALITEXT（Version 1.0）、泰國版的 BUDSIR IV、

p. 110

以及緬甸版的 CSCD 等三版，以下分別介紹。<sup>[4]</sup>

#### 1. PTS 版巴利光碟（PALITEXT）

【製作緣起】：(1) 如上述，英國巴利聖典學會 (PTS) 在 1970 年代，就嘗試將巴利原典電子化，但限於技術及人力，最後放棄計劃。(2) 泰國法身寺法身基金會 (Dhammakaya Foundation) 於 1989 年 9 月開始將 PTS 版巴利原典輸入電腦，並於 1996 年 4 月與 PTS 簽訂合約，正式發行此版光碟。

【光碟內容】：收錄 PTS 版巴利三藏共 53 冊——包括律藏 5 冊，經藏即五尼柯耶 37 冊，論藏七部論 11 冊；但尚未收入註釋 (aṭṭhakathā)、複註 (ṭikā)、與藏外典籍。

【功能與特色】：(1) 在版面安排上，光碟採「原貌重現」的形式，原原本本呈現原來紙本的內容與編排形式，此點大大有利於資料的參照與引用。(2) 光碟版也做了一些編排及內容的提昇，例如，標示原文行號、訂正原來紙本的打字錯誤等。(3) 提供多重搜尋功能：包括關鍵字詞、冊號頁號、文脈段落、以及詞彙序列等四種搜尋模式，並具備萬用字元 (\*) 的搜尋功能。(4) 提供雙視窗，利於原文比對。

【限制】：(1) 目前只發行 DOS 版，尚無 Windows 版。(2) 尚未收錄註釋、複註、及藏外典籍。(3) 未提供經文擷取功能。(4) 只提供點陣印表機 (dot-matrix printer) 的單頁列印模式。

## 2. 泰國版巴利光碟 (BUDSIR)

【製作緣起】：為慶祝泰王 Bhumibol (拉瑪九世) 的登基紀念大典，泰國 Mahidol 大學接受 Dr. Praves Wasee 教授建議，於 1987 年著手將泰國版巴利三藏置入電腦，以呈給泰王一個別具時代意義的獻禮。1988 年初步完成泰國字體巴利三藏的電子版 BUDSIR I；到 1994 年，又有更加完善的版本 BUDSIR IV on CD-ROM (MS-DOS 版)，此版加入了註釋書，並有羅馬字體版；1996 年又開發視窗版 BUDSIR IV for Windows；1997 年最新版本 BUDSIR IV / TT (for Thai Translation)，收入了巴利三藏的泰譯本。

【光碟內容】：BUDSIR IV 收錄了泰國版巴利原典共 115 冊，包括三藏 45 冊、註釋書 55 冊、以及藏外文獻 15 冊。BUDSIR IV / TT 另收三藏泰譯 45 冊。

p. 111

【功能與特色】：(1) 提供「關鍵字詞」與「冊號頁號段號」兩種搜尋模式，並具備萬用字元 (\*) 的搜尋功能。(2) 提供分割視窗、經註合查功能，可以經與經、註與註、或經與註比對。(3) 提供資料擷取功能 (但擷取的經文須經轉碼才能使用)。(4) 提供點陣印表機與噴墨印表機的可列印功能。(5) 1996 年的 Windows 版具有朗讀功能。

【限制】：（1）Windows 版只能在英文版 Windows 下執行，不能搭配中文 Windows 使用。（2）未提供 PTS 版頁碼對照，大大減低了版本比對效能，也降低了國際性。（3）經文擷取與轉碼程序太過繁複，資料利用不易。

### 3.緬甸版巴利光碟（CSCD）

【製作緣起】：印度政府曾委託新那爛陀大寺刊行巴利三藏，但該藏已絕版多時，難以取得。為了讓佛法重新在印度流行，S.N. Goenka 自 1985 年創立內觀研究所（Vipassana Research Institute, VRI）之後，便積極進行天城體版（Devanagari）巴利三藏、註釋、複註的刊行工作。這個版本（VRI 版）[5]以緬甸第六結集版為底本，將緬甸字體轉成天城體，並製成光碟，於 1997 年發行第一版 Chattha Savgāyana CD-ROM（簡稱 CSCD），隔年又出第二版，1999 年出了第三版。

【光碟內容】：（1）CSCD（Version 3）收錄三藏、註釋、複註、補註（anutikā）以及藏外典籍共 216 冊。（2）並收入一部巴利文法書、一部簡明巴英辭典，還有阿育王碑文。

【功能與特色】：（1）具有七種字體的顯示功能。[6]（2）強大的搜尋功能：提供字詞、文脈段落等搜尋方式，並具有「標示查詢」功能，也可直接翻查某經某品某頁。（3）可同時開啟多部典籍，並具備三藏、註釋、複註的段落關連功能，利於經文之間或經、註、複註之間的比較與對讀。（4）提供與其他版本（如 PTS 版、泰國版、原緬甸第六結集版）的對應頁碼。（5）更重要的，提供經文擷取與列印功能，並且可配合中文版 Windows 環境，利於資料的後續利用。

p. 112

## （二）巴利原典網路版

目前網路上比較完整的巴利藏原典有錫蘭版、緬甸版、以及泰國版三個版本，下面分別介紹。

### 1.網路上的錫蘭版（SLTP-BJT 版）

【製作緣起】：錫蘭政府為了紀念佛滅 2500 年（A.D. 1956 年），於 1960-1970 年代出版了一套巴利三藏，即 BJT 版，Buddha Jayanti Tipitaka Series。1991 年起，該國 Madihe Pagnnaseha Mahanayake 長老主持一個巴利三藏數位化計劃 SLTP（Sri Lanka Tripitaka Project），開始將此版三藏 58 冊輸入電腦。到 1994 年，全部輸入完成，之後，在 1996 年 3、4 月正式上網，並進行後續的校讀工作。[7]

【收錄內容】：此版內容除了三藏之外，還包括藏外典籍（如古代巴利文法、史書、古代巴利辭書、詩學與修辭學典籍等）37冊。

【功能與特色】：所有律藏和經藏內文都已附上 PTS 版對應頁碼，對照查索方便。

【限制】：目前還未提供線上瀏覽及檢索功能，必須下載之後，另以檢索器檢索。

【校讀資訊】：根據 L. S. Cousins 的附記，從 1996 年 10 月起，到 1999 年 1 月，共經過 4 次更新，目前律藏及經藏都已做過初步校讀（一次），但仍不保證完全無誤。

## 2. 網路上的緬甸版（VRI 版）

【製作緣起】：此版是基於前述緬甸版巴利光碟 CSCD 的網路版。[\[8\]](#)

【收錄內容】：同 CSCD 光碟版。

【功能與特色】：（1）提供線上瀏覽功能。（2）提供 PTS 版頁碼對照表。

【限制】：（1）只能依照藏、部、冊、章節等層次目錄翻閱瀏覽，未提供全文檢索。（2）沒有如光碟版所具有的經、註、複註的連結功能。

p. 113

## 3. 網路上的泰國版（BUDSIR on Internet）

【製作緣起】：此版是以前述泰國版巴利光碟 BUDSIR IV / TT 為基礎而製作的網路版，但收錄內容較少。[\[9\]](#)

【收錄內容】：三藏 45 冊，三藏泰譯 45 冊，註釋書 70 冊。

【功能與特色】：此版是此處介紹三個網路版中，功能最齊全的。（1）具有羅馬化、天城體、以及泰文、錫蘭文等四種字體的顯示功能。（2）提供 PTS 版頁碼對照功能。（3）最重要的，除了可線上瀏覽之外，還提供全文檢索功能。

【限制】：（1）未提供經、註關連及經文比對功能。（2）未提供經文下載功能。

### (三) 巴利原典翻譯

#### 1. 巴利原典英譯

【英譯網路版】：目前網路上似乎只見律藏及經藏的部分英譯，論藏尚未收入。其中收錄比較完整的有斯里蘭卡 MettaNet（慈網）、[10]美國 Access to Insight（內觀之道）、[11]以及馬來西亞 Mahindarama（瑪辛達寺）[12]等三個網站。收錄情形詳見筆者前述文章，此處不再贅述。

【英譯光碟版】：美國 Access to Insight 網站站長 John Bullitt 從 1993 年起，逐步收集巴利三藏英譯，到 1998、1999 年，累計取得經文超過 800 篇，並製成光碟流通（名為 A Handful of Leaves，一掌樹葉），但目前已絕版，預計 2006 年將再發行新版。

#### 2. 其他各國語譯[13]

p. 114

【錫蘭語語譯】：網路版可由 <http://www.metta.lk/> 進入 MettaNet 首頁，再點選，“Tipitaka”，其中收有巴利原典（即 SLTP-BJT 版）、以及英譯、錫蘭語譯。此外，MettaNet 也提供這些資料的光碟備索，但僅限於斯里蘭卡地區讀者。

【泰語語譯】：網路版可由 <http://budsir.mahidol.ac.th/> 進入泰國 Mahidol 大學網站，註冊或登錄後，在巴利三藏網頁的“Types”欄中，點選 Thai Tipitaka。此外，Mahidol 大學 1997 年的新版巴利光碟 BUDSIR IV / TT（for Thai Translation）也收有 45 冊巴利三藏的泰譯。

【德語語譯】：可由 <http://www.palikanon.com/> 進入。除了經藏、律藏翻譯之外，還有部分論藏及藏外典籍翻譯。

【法語語譯】：可由 <http://www.canonpali.org/tipitaka.html> 進入。目前只收錄部分經藏的翻譯，律藏、論藏翻譯尚缺。

### (四) 其他工具資源

由於後文評述對象主要為巴利原典與翻譯等核心文獻，因此其他相關的工具資源只羅列其名目，不再詳述。[14]

#### 1. 網路上的巴英辭典

(1) PTS 版巴英辭典 The Pali Text Society's Pali-English Dictionary。[15]

(2) A. P. Buddhadatta 長老的 Concise Pāli-English Dictionary（簡明巴英辭

典)。[16]

(3) G. P. Malalasekera 的 Dictionary of the Pāli Proper Names (巴利佛教專有名詞辭典)。[17]

(4) Nyanatiloka 長老的 Buddhist Dictionary (巴利佛教教理辭典)。[18]

p. 115

(5) Access to Insight 的 A Glossary of Pāli and Buddhist Terms (巴利佛教名相彙編)。[19]

## 2. 網路上的巴利文法

(1) Lily de Silva 的 The Pāli Primer (巴利語入門)。[20]

(2) Nārada 長老的 An Elementary Pāli Course (基礎巴利語教程)。[21]

(3) Charles Duroiselle 的 A Practical Grammar of the Pāli Language (巴利語實用文法)。[22]

(4) 蔡奇林編《實用巴利語文法》。[23]

## 3. 網路上的巴利讀本

(1) 「佛學數位圖書館暨博物館」所收的《法句經》(Dhammapada)詳解。[24]

(2) 護法法師選文，蔡奇林譯解《巴利語輕鬆上路》(Enjoying Pāli)。[25]

## 4. 網路上的巴利有聲資源

(1) 字母、單詞、句子讀誦。[26]

(2) 巴利佛教課誦。[27]

(3) 現代巴利經文唱誦。[28]

p. 116

# 三、略評當前的巴利數位文獻資源——其發展趨勢、效益、與限制

上面舉列並簡要回顧了當前巴利數位資源的概況。以下將以上述巴利藏光碟版、網路版、以及原典翻譯為主，考察、評析這些資源目前建構發展的趨勢特點、提供的學術研究效益、以及還待改善或開拓的空間。

## (一) 當前巴利數位資源發展的幾項趨勢特點

### 1. 收錄原典的「數量」不斷擴增

目前巴利原典收錄的總體趨向是，以三藏為核心基礎，逐步往藏外典籍、以及註釋、複註、補註不斷擴充。

以 PTS 版光碟（1996 年版）而言，目前只收錄巴利三藏 53 冊，未來第二版 MS-Windows 版則計劃收入主要的註釋書；接著第三版麥金塔（Macintosh）相容版，將納入所有註釋書及複註。

泰國版光碟 BUDSIR I-III（1988-1990 年）只收錄巴利三藏 45 冊，到了 BUDSIR IV（1994 年）則收入了註釋書 55 冊，以及藏外典籍 15 冊，共達 115 冊巴利原典。

較晚發行的緬甸版（VRI 版）光碟 CSCD，其第一版（1997 年）收錄原典 146 冊，第二版（1998 年）收錄原典 183 冊，第三版（1999 年）收錄原典已達 216 冊。

錫蘭版（SLTP）目前有 BJT 版三藏典籍 58 冊，未來計劃繼續收入 SHB 版（Simon Hewavitarne Bequest edition）註釋書 50 冊。

### 2. 原典校訂的「品質」不斷提昇

由於電子版原典是以紙本為底本而輸入製作的，因此對於先前紙本所沒有校出的錯誤，即可利用製作電子版而重新輸入及校對的過程，將之訂正。因此一般而言，電子版有很好的機會，可以提昇原典校訂的品質。[\[29\]](#)

在先前多篇文章中，筆者曾提到過，PTS 版原典的文字校對並不理想，以筆者的閱讀經驗而言，保守估計，平均一頁一般有 3、4 處以上的錯誤。這一問題，PTS 先前會長 K.R. Norman 也曾對學者提出警告，但限於經費及專家人力，

p.117

紙本的重校與新版有其一定的困難。[\[30\]](#)

據 PTS 版光碟「使用者手冊」所言，[\[31\]](#)泰國法身寺在製作此版光碟時，校訂工作十分謹慎從事。首先由兩組人員分別輸入巴利原文兩遍，而後透過機器比對，先改正大部分打字錯誤，然後再進行後續多組、多次人工校訂，務求萬無一失，因此校出並訂正了許多先前紙本的明顯錯誤。所以這個光碟版，在精確度上已經超越了原來的紙本。[\[32\]](#)

p. 118

### 3. 朝向「雙語／多語資料庫」發展

這裡的「雙語／多語資料庫」是指除了巴利原典之外，還包含原典的其他語言翻譯。這種雙語／多語資料庫對於文獻（佛法）的跨文化傳播，尤其具有重大的影響力，因此各國無不努力建構。由於國家的差別，各網站或光碟所收錄的譯本語言也就不同。以下舉出幾個具有「雙語／多語」特點的資源。

1997 年最新版的泰國版巴利光碟 BUDSIR IV/TT (for Thai Translation)，已收錄了巴利三藏的泰譯本 45 冊。此版的網路版 BUDSIR IV on Internet 也收有三藏的泰譯。這是「雙語資料庫」的例子。

斯里蘭卡 MettaNet 網站除了收錄錫蘭版巴利原典之外，還收有原典的英譯及僧伽羅語（錫蘭語）翻譯，是「三語（多語）資料庫」的例子。

此外，有些網站雖未收錄原典，而只收錄英譯（如 Access to Insight），或德譯、法譯等，但由於巴利原典在網路上可即時取得，並且彼此的目次架構基本一致，因此通過適當連結，這些資源也具廣義「雙語／多語資料庫」的功能。

### 4. 納入輔助原典解讀的「工具資源」

為了幫助讀者解讀原典，目前已有多個網站或光碟開始收錄相關文法書、辭典等工具資源。

泰國版巴利光碟 BUDSIR IV / TT (1997 年版) 收有一部「巴泰辭典」，未來還計劃收入「巴英辭典」。

緬甸版巴利光碟 CSCD Version 1.1 (1997 年版) 收有一部「巴印辭典」，Version 3 (1999 年版) 則附上一部簡明的巴英辭典，以及一部巴利語入門的文法書 (Lily de Silva 的 The Pāli Primer)。

斯里蘭卡 MettaNet 網站中的“Pali Utilities” (Dictionary and Grammar Books)，收有多部巴英辭典及巴利文法 (參前面介紹)。

PTS 版光碟 (Version 1.0) 目前尚未收錄任何工具資源，但計劃在 Version 3 附上一部交談式巴英辭典。

### 5. 提供「多重對照」功能

這裡「多重對照」意指不同版本巴利原典之間的頁碼參照，或者同一版本內部的經文與經文、經文與註釋、註釋與複註之間的參照。

關於不同版本的頁碼對照，由於 PTS 版在國際學界流行較廣且較久，因此幾乎所有重要文獻引用原典都以之為標準，有鑑於此，目前無論緬甸版 CSCD、

泰國版 BUDSIR IV（網路）、或者是錫蘭版（SLTP）都提供 PTS 版頁碼參照。使用者可以通過此版頁碼作為中介，進行各版本之間的原文比對。

至於同一版本內部的對照功能，泰國版光碟 BUDSIR IV 提供經、註合查，可以達到關鍵詞句的經與經、經與註、註與註比對的功能。緬甸版 CSCD 光碟則進一步做到經、註、補註之間相應段落的連結功能。PTS 版也提供切割視窗的經文比對功能。

## 6.提供「多字體顯示」功能

巴利語作為一個語言，歷來並沒有專屬的文字記錄系統，而佛教聖典初始之時，也是口傳文獻。目前所知，巴利三藏最初是在紀元前一世紀於錫蘭紀錄下來的，當時是用錫蘭文字拼寫，後來聖典傳到緬甸、泰國、高棉等南方國家，也同樣採用這些國家本身的文字拼寫保存。因此，使用多種字體「音寫」（音譯）保存巴利聖典就成了南傳佛教的一個傳統特點。

現在由於電腦轉換字體容易，許多巴利網站或光碟都紛紛朝向多字體顯示功能發展，這樣，不同國家的讀者將更容易閱讀巴利聖典。

目前 PTS 版光碟只提供羅馬化字體。而泰國版 BUDSIR I（第一版，1988 年）是泰國字版，第二版（1989 年）開始有了羅馬化字體，到了 1997 年的 BUDSIR IV / TT 版，就有 6 種字體的顯示功能——可以顯示泰國字、羅馬字、緬甸字、錫蘭字、高棉字、以及天城體等 6 種文字的巴利原文。BUDSIR IV 的網路版，也提供 4 種字體的選擇（上面 6 種中除去高棉字、緬甸字）。

緬甸版光碟 CSCD 1.1 版（1997 年）提供羅馬字、緬甸字、天城體等 3 種字體，2.0 版（1998 年）開始，便提供 7 種字體的顯示功能（即上述 6 種加上蒙古字）。

多字體功能與雙語／多語功能一樣，都能大大提昇文獻的跨國、跨文化傳播效力，對於巴利文獻的普及化有著莫大助益。

## 7.朝向「有聲資料庫」發展

上面提到，巴利佛典最初便是口誦傳承，而巴利語作為一個語言，口誦、耳聞更是自然而親切的傳播方式。因此「有聲化」不管是對這部聖典或這個語言，都是一種重要而有力的載錄與傳播模式。

目前已有一片光碟具備「讀誦」功能，就是 1996 年泰國 Mahidol 大學的 BUDSIR IV for Windows，讀者可以標示經文段落，要求「讀誦」。但可惜的是，這片光碟只能搭配英文版 Windows，無法在中文版下正常顯示。相信隨著電腦科技的發展，

未來在更大容量與與更高處理速度的支援下，會有更多巴利網站或資料庫朝向「有聲化」發展，提供讀者更多樣的閱聽模式。

## 8.適合「中文系統」的資料庫

由於目前巴利學數位資源幾乎都由南傳佛教國家以及西方國家所開發建構，因此一般鮮少考慮到中文系統的執行環境。例如前述泰國版 **BUDSIR IV for Windows** 就只能在英文版 **Windows** 下執行，而其他網路版即便安裝了字型之後，還是經常遇到字碼無法正常顯示的情形，這對於廣大的中文地區的使用者而言，無疑是一大缺憾。

幸而已有一些建置單位開始注意到這個問題，而開發配合中文 **Windows** 的相容版。緬甸版 **CSCD Version 3** 便是一個好例子，不只可以在中文 **Windows** 環境下正常顯示，並且資料下載之後，還可直接配合中文檔案，進行後續的編輯利用，讓資源的效益發揮到最大。相信以中文地區佛學研究、以及佛學數位資源建置的蓬勃發展，當會促使越來越多巴利資源的建構單位，往中文相容的系統去發展。

### （二）當前巴利數位資源所提供的學術研究效益

#### 1.更充分的研究樣本

更多更廣的研究樣本與資訊，經常可以讓研究者獲得更整全的觀察與更客觀的理解。目前光碟版及網路版，共有四套巴利原典——錫蘭版、緬甸版、泰國版、歐洲版（**PTS** 版），且以緬甸版而言，當中就包含了許多西方尚未出版的原典。就筆者所知，目前國內即使是最專業的佛學研究機構，也沒有任何一個單位在其圖書館或資料中心當中，收齊這四個版本的紙本文獻，最好的恐怕只擁有其中的一半或三分之一，甚至更少。但現在，任何個人研究者，都可以在其即使是窮鄉僻壤的斗室當中，敲幾個按鍵，就佔有比專業圖書館更加充分的原典資源。

除了原典之外，原典翻譯也是研究者重要的參考資料，網路上收錄了許多尚未出版或不易取得的譯本。例如 **MettaNet** 所收的 **Bhikkhuni Uppalavana** 的英譯、或 **Access to Insight** 所收的 **Bhikkhu Thanisaro** 的英譯，藉此可以觀摩更多譯者以及不同傳統對巴利原典的不同理解，增益原典解讀的深廣度。

#### 2.更精良的原文校訂

好的校訂本，能夠給研究者提供更加可靠的資訊來源，這是一切研究最根本而重要的基石。

p.

121

巴利數位原典，一則由於各版本在數位化過程中，重新校正了一些過去錯誤；

再則，更重要的，由於數位化以及各版本開放使用所帶來的利益，使得廣大的研究者，能夠即時、方便地取用與比對更多版本，通過各版所提供的「頁碼對照」功能（如前述，以 PTS 版頁碼為中介），很容易對校不同版本的巴利原文。因此，在底本文獻的運用上，一般而言，可以取得較之過去更好的讀法。

### 3.更適切的分段、斷句與標點

以「文字」為媒體而載錄的「文本」本身，由於難以一五一十地記錄原來對話現場的全般訊息——諸如，講話者的表情、手勢、姿態，語調的高、低，語氣的緩、急、輕、重，停頓的久、暫等等，因此必須通過適切的分段、斷句與標點等輔助手段，多少填補這些失落的訊息，以便能夠更加清楚、形象、而逼真地傳達文本（或事件）的結構脈絡、意義內涵、以及情境氣氛，這對研究者解讀、運用這些基礎文獻而言，將有莫大助益。

現代編輯出版的巴利原典，以 PTS 版而言，雖說大部分都已進行了不錯的分段、斷句、與標點，但仍有不少缺失。[\[33\]](#)

除了標點問題之外，PTS 版也有部分分段與斷句明顯錯誤的地方（下文將舉實例說明）。筆者曾在一些地方談過，斷句錯誤，不僅會造成義理的誤解，同時也可能造成修行法門的歧出。[\[34\]](#)

幸而，上述 PTS 版各種分段、斷句、及標點問題，現在在數位文獻取用容易的效益下，可以通過緬甸版（或錫蘭版）而進行校正，[\[35\]](#)從而可以避免一些誤讀的危險。

p. 122

### 4.更快速而趨近窮盡的資料檢索

以上第 1 點是研究資源「數量」的更加豐富，第 2、3 點則是研究資源「品質」或「可靠度」的更加提昇，這些都是提昇研究成果必需依賴的基礎。然而，數位資源所具備的最大效能或特點卻在於：通過檢索工具的幫助，讓研究者能夠用更少的時間、精力，進行盡可能「窮盡而沒有遺漏」的觀察。

在過去，面對浩瀚的紙本文獻，研究者或者只能選取有限的樣本，進行片面或部分的觀察；或者需要憑藉記憶或印象之所及，而進行難免主觀而不盡可靠的論述；或者必須經年累月地製作無數卡片，尚難保沒有任何遺漏。但現在對於某些問題，研究者可以在檢索工具的協助下，取得趨近窮盡的樣本，進行更整全的觀察與更客觀的論述。

### 5.提昇巴利研究的「精審度」

總的來說，上述四項效益都指向一個共同目標，就是提昇巴利研究的「精審度」。而這個精審度的等級，就是一般所謂的「低精審」（lower criticism）

——也就是不管對於文本本身的讀法、或者對於文本所記載的內容的各種研究，都盡可能「窮盡地」運用「巴利內部」的所有文獻，以取得更加可靠或確當的結果。通過文本的數位化以及搜尋引擎的幫助，這樣的理想與目標變得更加可能。

### (三) 當前巴利數位資源的不足之處

儘管上一節談到，目前的巴利數位資源不管是文本的「質」、「量」、或檢索擷取的「運用效能」，都已大為提高，但這並不意味這些資源已經臻於「完善」的境地。相反的，這些數位化工程，其實不過十數年歷史，儘管進展神速，但畢竟還在一個起步階段，因此還有諸多缺失及問題仍待不斷改善提昇。以下舉其綦綦大者。

#### 1. 原文校訂仍不夠完善

在原文校訂方面，上面說過，PTS 版光碟版雖已改正一些原來紙本錯誤，但恐怕還有更多沒有訂正過來。錫蘭版的經藏、律藏目前只經過初步校訂（一次），未來還須再作校訂。緬甸版一般而言三藏（及其註釋、複註等）校對比較精良（但仍有少數錯誤），但藏外典籍則大多只經過機器校訂（約可除去 98% 錯誤），還未經過人工審校，可靠度較為不足。

在筆者部分的閱讀經驗中，緬、泰、錫、歐（PTS）四版當中，PTS 版的校對算是很不精良的，

p.123

而錫蘭版由於只經初步校訂，因此問題也不少。為了比較具體的說明問題，以下隨舉二經為例，顯示各版目前編校的大略情形。此處舉的是 S.12.35 Avijjāpaccayā 經（緣於無明經），以及 S.35.133 Verahaccāni 經（毗紐迦旃延經）。[36]這兩經的文句與字詞在 Nikāya 中算是比較平常、單純，也較簡單的。

首先，看 Avijjāpaccayā 經。

表（一）五個版本《S. 12. 35 Avijjāpaccayā 經》「文本編校情形」之比較  
（本經 PTS 版共 2.66 頁巴利原文）

差誤\版本	字詞拼讀錯誤	異讀選擇問題	分段問題	斷句問題	語詞順序錯置
PTS 版紙本	19 處	2 個（乘 4 次）	1 處（乘 4 次）	1 個（乘 4 次）	0
PTS 版光碟	14 處	2 個（乘 4 次）	1 處（乘 4 次）	1 個（乘 4 次）	0

錫蘭版 (網路)	14 處 另 3 處 (乘 4 次)	0	0	甚多斷句及標點 尚未規範化	1 個 (乘 8 處)
泰國版 (網路)	1 處	0	0	0 (但斷句及標點 過少)	0
緬甸版 (光碟)	0	0	0	1 個 (乘 4 次)	0

如表一所顯示，在此經短短 2 頁的巴利原文當中，PTS 版共有 19 處拼讀錯誤（即平均每頁超過 7 處錯誤），[37]還有 2 個異讀選用不佳，[38]以及 1 個分段錯誤及 1 個斷句錯誤（但這種情況不是通例）。

p.

124

[39]值得注意的是，泰國法身寺在製作此版光碟時，儘管運用多種電腦軟體進行比對、除錯，以及多次的人工審校，但只改正其中 5 處打字錯誤（約 26%），還留下大部分錯誤；而異讀及分段、斷句問題，也未加以處理。

錫蘭版問題也不少，有 14 處拼讀錯誤，另 3 處（乘 4 次）文法錯誤，[40]以及 1 個（乘 8 處）語詞順序顛倒的地方。[41]相較而言，泰、緬兩版的情況就好了很多。泰國版只有 1 處拼讀錯誤，[42]緬甸版只有 1 個斷句問題。[43]

其次，看 Verahaccāni 經。（見下頁）

如表二所示，此經五個版本的編校情形與前經大致相仿。在 3 頁的巴利原文當中，PTS 版有 22 處拼讀錯誤（平均一頁超過 7 處），[44]另有 2 個異讀可

p. 125

表（二）五個版本《S. 35. 133 Verahaccāni 經》「文本編校情形」之比較  
（本經 PTS 版共 3 頁巴利原文）

差誤\版本	字詞拼 讀 錯誤	異讀選擇 問題	分段 問題	斷句 問題	語詞順 序 錯置
PTS 版紙本	22 處	1 個 另 1 個 (乘 17 次)	0	0	0
PTS 版光碟	18 處	同上	0	0	0

錫蘭版（網路）	17 處	同 PTS 版	0	甚多斷句及標點尚未規範化	1 段
泰國版（網路）	4 處	1 個	0	0（但斷句及標點過少）	0
緬甸版（光碟）	1 處	同 PTS 版	0	1 個	0

再斟酌。[\[45\]](#)類似上一經，PTS 光碟版只改正了少數錯誤（4 處），比起紙本，提昇並不顯著（不到 19%）。錫蘭版有 17 處拼讀錯誤，[\[46\]](#)還有 1 段文句有脫字及倒亂情形，

p. 126

[\[47\]](#)問題還是不少。相較之下，泰、[\[48\]](#)緬[\[49\]](#)兩版顯然校讀較精，但也還有改善空間。

綜觀這五個版本，PTS 版紙本編校問題似乎是最大的，而其光碟版比起紙本，略有改善，但總體而言，幅度並不大。不過在機器與人工的多重校對、層層把關之下，似乎也未見產生新的問題，這一點是難得的。錫蘭版因為還在試用階段，問題也不少。除了一般常見的打字拼讀錯誤之外，這一版可能還有電子文件複製過程所衍生的新問題，以及在展開原本省略段落時所造成的文法錯誤，這是使用此版時必須特別留意的。

相較之下，泰國版與緬甸版顯然較為精良，但如同上面經例所顯示，並非完全沒有問題。因此，要取得比較好的原文讀法，就不能獨據任一個版本，而必須同時參校各版，小心取擇其間的差異。

總的說，這些版本的情況與前述 K. R. Norman 及水野弘元對 PTS 版問題的警告，以及水野弘元對泰國版評價較高的情況略相符合。Bhikkhu Bodhi 在從事《相應部》新譯之時，以緬甸版（第六結集版）作為基礎底本，但也同時參校其他各版的讀法，不偏好任一傳統，而取其中最好的。[\[50\]](#)這種取向或採行版本的策略，從以上的調查分析看來，顯然是極為明智的。[\[51\]](#)

p. 127

## 2.分段、標讀的精緻度仍待提昇

在原文的分段、斷句方面，如以上經例所示，PTS 版還有一些錯誤仍需訂正。[\[52\]](#)此外，甚至也有少數「分經」錯誤的例子。[\[53\]](#)標點方面，前面提到，整個《相應部》只有“||”及“|| ||”兩種分隔或停頓符號，尚未採用現代標點。還有《中部》第一冊，有必要重新標點；而其他各冊的標點問題，也還需要修正。

錫蘭版雖採用可讀性較高的現代標點，但由於目前還在校讀試用階段，因此還可見到許多標點不一致、以及標點錯誤的地方。

泰國版（BUDSIR IV）儘管字詞校訂頗為精良，但在原文編輯上，只有分段（標以段號），而未如其他三版，適當的分經並標上經名、經號，閱讀上頗為不便。此外，泰國版也是各版當中，句讀最為簡陋、最為難讀的。全編除了「省略號」（...）之外，竟只用到一種標點符號——句號（.）。因此，文中所見，盡是長長的幾個句子連串成段成篇，才有一個句點，這大大增添了讀者的閱讀負擔。

各版當中，緬甸版（CSCD）的斷句與標點，算是最細緻、也最精良的。CSCD 靈活運用了現代標點，如：逗號（,）、分號（;）、句號（.）、問號（?）、

p.128

驚嘆號（!）、破折號（—）、單引號（‘ ’）、雙引號（“ ”）、省略號（...）、連接號（-）等，讓原來口語的停頓、語氣能夠呈現，也讓語句的意義及前後關係更為明白，因此可讀性極高。但如同前面的經例所顯示，仍多少有改善空間。此外，經中的人名、地名等專有名詞，若能如 PTS 版採用大寫開頭，可讀性將會更高。這一點也是 PTS 版較之其他三版，更為精緻的地方。

### 3.許多原典及翻譯仍待收錄

目前 PTS 版光碟只收錄三藏原文 53 冊，其他已經出版的大量的三藏註釋、部分複註、以及藏外典籍都還未收入。儘管這些文獻絕大部分在其他版光碟中（如 CSCD）已經收錄，但由於不同校訂本所依據的底本經常不同，因此這些原典仍甚具數位化價值，以利於不同版本的比對、研究。

在當前一些網站中（例如 Access to Insight、MettaNet），原典翻譯雖已收錄不少，但距離最起碼的「完備」（律、經、論都至少一個譯本）還有一大段距離，其中尤以律藏、論藏最為缺乏。[\[54\]](#)這些譯本的收錄將是下一階段的重要工作。

### 4.多數資源「版本資訊」不夠清楚

為了讓讀者對於原典的來源有一基本了解，一般而言，至少需要說明該原典的編輯者、編輯年（或出版年）、所依據的相關底本、取擇異讀及編輯建立版本的原則等等。但目前見到的巴利數位資源，這些資訊的提供仍十分不足。

目前不管是內觀研究所（VRI）的 CSCD、泰國 Mahidol 大學的 BUDSIR IV、還是斯里蘭卡 SLTP，都只說明該版根據哪個版本進行數位化，[\[55\]](#)但並未進一步說明這些底本編輯過程所依據的本子，以及編輯校訂的原則。因此儘管原文中都附有校勘資訊，但讀者卻無從得知那些異讀所從出的那些本子的情形。[\[56\]](#)

p. 129

相較之下，PTS 版光碟的情況就好了很多，每一冊都將校訂者的「引言」（Introduction）原原本本的收錄進去，讀者可以藉此了解該冊的相關內容、底本資訊、以及編輯原則等；但美中不足的是，不少原典沒有注明其最初出版年（first published），而只列出光碟版所據的紙本的出版年（published）——經常只是重印年（reprinted），這就無法正確反映該冊原典的實際編校年份。[57]

至於網路上所收羅的譯本（英譯為主），其出版資訊就更不足了。許多譯文都未交代來源出處，是否經過授權，是否經過校對，有的甚至連譯者是誰都完全沒有交代。[58]這些都是極待改善之處，否則其價值就十分可疑。

數位資源的建構者，應假設讀者沒有（或看不到）其所根據的紙本文獻，而盡可能地將紙本原來的相關資訊充分提供，否則該資源的價值必然大打折扣——此點對於學術資源尤其重要。

## 5.部分資源尚未提供檢索功能

檢索功能是數位化資源之所以能發揮高效能的核心引擎，它可以讓資源的運用收到如虎添翼之效。

目前各版巴利光碟，都已提供相當好的檢索功能。可是網路版，除了泰國版 BUDSIR IV on Internet 具備全文檢索功能之外，緬甸版和錫蘭版都只提供使用者依藏、部、冊、相應、品、經之類的目次結構，去翻查及瀏覽經文，使其在研究上的效用幾乎折損殆盡。

## 6.資料擷取與引用功能仍待加強

目前除了 PTS 版之外，其他三版都提供資料擷取或下載，只是功能還十分簡陋。

CSCD 是以原緬甸版（Myanmar）頁次為單位來儲存檔案，其操作過程有多項不便之處：（1）儲存的檔案中常包含許多多餘資料，必須再以人工檢查、編輯。（2）更麻煩的是，巴利「語詞」（或長串複合詞組）遇到跨行時都被切為兩半，使用者必須逐一接回，或補上「連接號」（“-”），以免被誤為兩詞。（3）即使只是引用 3、5 行原文，一樣必須進行一連串確定頁次、存檔、開檔、資料篩選、語詞接合等繁複的機械動作，十分不經濟。

p. 130

泰國版 BUDSIR IV 光碟，資料擷取同樣相當不便，[59]網路版雖有改善，但仍有一些限制。網路版中，系統只提供單頁下載功能，一次只能擷取一頁，而且系統限定存檔動作最多 5 次，因此每次登錄最多只能取得 5 頁資料，而且是分在 5 個檔案中，仍須後續大量編輯工作。

相較之下，錫蘭版的擷取功能就比較簡便。SLTP 提供使用者下載整個三藏檔案，使用者只要找到經文所在，便可直接複製利用。

較為可惜的是，PTS 版目前仍未提供資料擷取功能。儘管此版還有一些錯誤，但目前還是學術界最常引用的版本。未能擷取，研究者引用文獻時，就只能重新輸入。重新輸入不只是費時費工，更大的問題在於，可能存在打字錯誤、或校對不精的危險，這使得資料的「可靠性」更沒有保障。[\[60\]](#)

原典是所有研究的第一手根據，研究者對它的「引用頻率」及「引用量」自然是最高的。因此，方便、友善的擷取與引用功能，對於學術界及研究者的幫助實在十分重大。它不但節省了眾多研究者不斷重複輸入的「勞務成本」，同時更能維護及確保基礎資料的「正確性」與「可靠性」。[\[61\]](#)

## 7.資源建構尚未達到「高精審」的規格

以上所談，相對而言是屬於比較微觀的問題。而如果從比較巨觀的角度看，目前巴利數位資源的限制在於：還沒有達到「高精審」（higher criticism）的規格。

前面談到，當前巴利數位資源建構的總體趨向就是「低精審」（lower criticism）——不管是 PTS 版、緬甸版、錫蘭版、還是泰國版——儘管傾力收羅建構所有的巴利文獻資源，但畢竟還是侷限於巴利一系當中。但目前學界普遍的認識是，巴利文獻以及其他犍陀羅語、或佛教梵語等初期佛教文獻，很可能都是從其他中古印度雅利安語轉譯過來的，

p.131

[\[62\]](#)因此巴利文獻並不能直接視為最原型的佛教文獻（包括語言與內容）。這樣，單就巴利本身的資料進行研究，事實上是有所侷限的。

要達到更高的精審度，就必須盡可能地至少納入同時期的其他傳承的其他語本，進行綜合的考察。這不只是一要對它所承載的初期佛法取得更好理解的必要手段，即便只是要對巴利文獻本身的各方面（例如語言、組織、內容、傳承間的變化等）取得更好的了解，都不得不同時參考其他語本，而進行對比的研究。

因此，從低精審的思惟與規模，開始廣納同一時期的其他語本文獻，從而邁向高精審的思惟與規模，這將是巴利數位資源的建構者下一階段必須努力的方向與目標。

## 四、巴利數位文獻資源的未來展望——略談「初期佛教聖典多語多本平行語料庫」

### (一) 從「低精審」邁向「高精審」

如前所述，巴利數位資源建構的一項重要目標，就是支持研究者提高研究的精審度。因此，當前資源的一些不足之處，未來應持續改善，這些包括：更精良的版本校訂及標讀、更充足的原典及譯本、更好的辭典及工具資源、更清楚的版本資訊、更強化的資料檢索及擷取功能等。此外，雖然已有部分資源已經具備有聲資料庫、多語資料庫的特性、以及適合中文系統的執行環境，但大多數則沒有此種功能，這也是未來必須努力的。

除了巴利內部資料的精審化，未來最需要的是：突破目前的思惟與規格，將資源建構的規模往「高精審」推進——結合同一時期的不同傳承的不同語本，將巴利資源置於整個初期佛教聖典的總集當中，讓研究者更容易進行比對研究，以便對聖典的各方面求得更切當的了解。

其實這原本就是學術界行之有年的研究方法，數位資源的建構應該反映這種事實，以便與真正研究的操作模式更緊密地結合，提供研究者更符合實務需求的幫助。下面就是一個基於這樣思考的擬構案例。

p. 132

### (二) 一個擬構案例：「初期佛教聖典多語多本平行語料庫」

本節將提供一個擬構案例，以「初期佛教聖典」為收錄文本的核心範圍，以「高精審」為指導目標，結合「多語言」、「多版本」的相關材料，做到「平行對讀」的功能，並且對於相關語料進行「語言屬性」的標記，以期能在現有研究功能之外，更進一步拓展「語言研究」與「語言教學」的功能特點。[\[63\]](#)

p. 133

#### 1.略說「初期佛教聖典·多語多本·平行·語料庫」

##### (1) 初期佛教聖典

初期佛教聖典主要指佛教最早期的文獻集成，也就是是記錄佛所教導、施設的「法」與「律」。「法」就是「阿含經」，目前數量最大的是巴利語系的五尼柯耶以及漢譯的四阿含。「律」是「律藏」，目前數量最大的也是巴利語系（上座部）的律藏，以及其他部派所傳的漢譯五部廣律——四分律、五分律、十誦律、摩訶僧祇律、以及根本說一切有部律。在經、律的核心基礎上，再附以論典及藏外文獻，乃至註釋、複註等輔助材料。

將初期聖典作為一個「整體」建構資料庫，有幾個原因：一是，這些聖典在時間上有「同期性」。儘管這些經、律經過數百年時間、在不同地區、由不同的

僧伽團體受持流傳，其內部已有一些演進改變的痕跡，但總體而言，並不巨大，反映的大致是同一時期、同一階段的佛教樣貌。二是，內容上有「一體性」。儘管經、律各有不同的意趣或性質，而四阿含的各部也有不同偏重，但總的來說，它們卻是密切相關、互為補充、不可切割的一個整體。因此將這些關係「密切度」極高的文獻作為一個整體，進行各面向的比對研究，將可獲得更高的「效度」。

## (2) 多語、多本

早期的經、律目前流傳下來的有眾多語言的眾多傳本或譯本：(A) 印度雅利安語系的有巴利語本、佛教梵語本、犍陀羅語本等。(B) 古代譯本有漢譯本、藏譯本等。(C) 現代譯本更多，如中譯、英譯、日譯、德譯、法譯等。「多語、多本」意指佛典流傳、翻譯的複雜情況，有時「一語一本」，有時「一語多本」，有時「多語一本」，但更多時候是「多語多本」，以《法句經》為例，就有巴利語本、佛教梵語本、犍陀羅語本、漢譯本、藏譯本等多種語言的多種傳本（或譯本），[\[64\]](#)而若再加上各傳本（或譯本）的現代語譯本就更多了。

p. 134

將多種語言的多個傳本（或譯本）收在一起，可以起到相當多面的研究效用，例如文本的內容問題、語言問題、翻譯問題、傳誦或傳承間的演變問題等等。雖說許多個別經典的研究，早有不少學者做過了，[\[65\]](#)但若就整個龐大的初期佛教聖典的各種各樣的問題而言，尚待掘發之處必定更多。[\[66\]](#)除了提昇研究效能之外，多種語本的總集，也能提供教學的便利，拓展教學的深度與廣度。當然，它也可以開展另一種聖典集成的方式，不只是依傳統不同語系（巴、漢、藏等）而集成聖典，同時可以透過數位資源靈活的彈性，依文本的內容為主軸，將同樣內容而散落諸方的不同語本、譯本，總集在一起。

## (3) 平行

要讓這些經、律的不同語本可以起到更好的比對、研究的效果，並非單單把它們收在一起就夠了。以 **MettaNet** 為例，雖然同時收有巴利語本、英譯本、僧伽羅語譯本（算是「一本多語」的例子），但是並沒有做到「平行」結構，也就是沒有把同一經文或同一律文的三個語言的本子連結在一起，因此也無法替讀者省去翻查目錄的機械操作。

進行平行關連的「文獻單位」不能太大，否則就沒有意義了。單位越小，就像使用更高倍率的顯微鏡，讀者將更容易進行微觀的研究，但建構者自然要更費力氣。以經藏而言，最大的平行單位至少應該是「小經」，例如「雜阿含」、「增一阿含」的一一小經，但對於「中阿含」、「長阿含」而言，因為經文長，單做到經對經，其實效用並不大，因此有必要進行更小單位的切割。

除了「經對經」之外，平行單位若能更小，做到「段對段」、「句對句」甚至「詞對詞」，它的學術價值以及可以起的效用就更高。當然這裡牽涉的就不只

是電腦技術的問題，更關鍵、更費力的是在於文本的精細判讀工作。判讀錯誤，

p.135

反而提供給使用者錯誤訊息。這種工作需要大量的專業研究人力。

#### (4) 語料庫

「語料庫」是作為語料的文本集合，[67]因此原始佛教聖典的總集，自然也是語料庫的一種。將它稱做「語料庫」，是因為在這個應用案例中，將著重於這些材料在佛教學研究當中的「語言研究」與「語言教學」等方面的功能特點，並強調它作為支持佛教學研究的「基礎建設」的功能性。

雖然如上所說，聖典的總集（單純的文本堆砌）是一種語料庫，但它只能稱為「生語料庫」

p. 136

（raw corpus），也就是並沒有經過進一步的「加工處理」——標注語言訊息，因此它能提供的語言研究的效用就十分有限。[68]現代語言學意義下的語料庫，通常必須在文本中加注大量「語言訊息」（或語言知識），並且有適當的軟體可以提取及處理這些訊息，以便達到計算機輔助的語言研究的效能。

對文本加工處理的深度越深（即存放的語言知識越豐富），語料庫能夠發揮的語言研究的效能就越大。這些語言訊息的加工標注，基本的，例如音韻訊息（如連音法）、語法訊息（如詞類、性、數、格等）、句法訊息（句子結構），以及語意訊息、語用訊息、語篇訊息等等。

「標注語料庫」除了提供語言研究的效能之外，它也為原典的翻譯提供基礎並且記錄過程。古代的譯經大德，儘管業績顯赫，但可惜的是，並未將他們對於文本解讀的珍貴語言知識流傳給後人。因此，標注語料庫也有佛法傳承的重要價值；當然，對於今日及未來苦學原典語言的學子們，它更是一個輔助學習的「語言寶庫」。此外，它也是編寫更好的文法書、[69]更精良的辭典的重要資源。[70]

p. 137

總的說，語料庫的作用是多樣的，但最基本而重要的功能在於，基於附加「語言訊息」這種「超語言」（metalinguistic）的「深層加工」以及「計算機輔助訊息提取」這種「超人工」的工作效能，得以支持學者從巨量的真實語料當中，取得對該語言更客觀與平衡的了解——包括「定性的」（qualitative）與「計量的」（quantitative）。因此，這也意味著，通過新的工具與新的方法，吾人未來對這些語言的理解與掌握，有可能超越過往的文法書、辭典等等對這個語言的描述；並且基於這種更正確及更恰當的語言理解，便有可能建構更好的文本翻譯，以及基於該文本的其他種種方面的研究與詮釋。[71]

## 2. 必須進行的相關基礎工程

上面大略談過「初期佛教聖典多語多本平行語料庫」的一些基本想法，儘管整個概念是簡單的，但是當中實牽涉到大量的基礎建設，並且這些工作主要的關鍵點，不只在於資訊科技的技術問題，而更在於相關語言文獻學的基本訓練。以下舉列幾項這個案例所須進行的相關工程建設：

p. 138

- (1) 巴利語本、漢譯本等【重新精校、分段、標點】[\[72\]](#)
- (2) 巴利語本、梵語本、犍陀羅語本等【語料標注】
- (3) 巴利語本、梵語本、犍陀羅語本、藏語本【中譯】
- (4) 各語本及譯本（含現代語譯）之間的經對經、段對段、句對句、詞對詞【平行連結】

上述任何一項，或任何一項的一小部份，都是重要的學術工程，都無法由機器代勞，而需要專業人力一字一詞地校訂、標注、翻譯、判讀，可以想見此項工程之耗時費力了！但由於初期佛教文獻基本上具有「封閉性」——也就是不會有大的增殖，即使未來有新出土的材料，數量依舊是有限的。因此儘管耗時費力，終究是可能逐漸完成的。

## 3. 能夠開發的教學研究效能

如上所說，「多語多本平行語料庫」的建構是極耗時力的，但它卻可以帶來極長遠的價值與多方面的效能，並且這些效能將會隨著不同領域研究者的需求、創見與開發，而不斷增長，可以說，其潛在的學術效益將是無可限量、難以估計的。以下簡單舉示幾項：

p. 139

- (1) 語言研究[\[73\]](#)
- (2) 語言教／學
- (3) 翻譯研究
- (4) 翻譯教／學
- (5) 版本校勘／校訂
- (6) 教理研究
- (7) 經典集成史研究
- (8) 經典傳承史研究
- (9) 文法書編纂
- (10) 辭典編纂

## 五、結語：建構一個更多效能的佛典知識寶庫

佛教聖典自從 2500 年前第一次結集開始，就作為佛教文化的知識寶庫而流傳。其後隨著佛法的傳播，聖典也以種種不同語言而轉譯、翻譯、筆錄、輾轉傳抄。後來各語本大藏經的編輯與刊印，便是保存這個聖典寶庫的具體體現。

如今，文本的載體有了重大改變，一片片光碟、一個個網站，給了這個寶庫全新的樣貌，也帶來以往做不到的種種應用效能。以巴利數位資源而言，這十幾年的發展，不可謂不迅速，不只收錄原典、譯典的數量增加、版本校訂的品質提昇，並且附加了效力甚大的搜尋引擎、以及便利的資料擷取功能，部分資源甚至已經進展到有聲資料庫的階段。

然而誠如上文所言，這些發展儘管已有相當成績，但只是個起步，還有眾多不足之處仍待改善。如何繼續提高「精審度」，以及對這個知識寶庫進行各種「深層加工」，是未來需要努力的。

本文提供的「多語多本平行語料庫」的觀念，不過是以「語言研究/教學」為導向的一種加工模式。但這些佛經文本，除了承載語言訊息之外，更是一部豐富多采的「古典文化百科」，因此如何運用各不同專業學科的研究取向與方法，進行不同角度、不同類型、不同模式的加工（例如歷史的、思想的、社會的、倫理的等等），以提取潛藏在文本當中的種種不同知識，讓它發揮更高、更多樣的研究效能、教學效能、以及佛法傳承與傳播的效能，這些都有待吾人後續的努力！

## 【縮略語】

BJT = Buddha Jayanti Tipitaka Series

BUDSIR = The Buddhist Scriptures Information Retrieval

CSCD = Chattha Savgāyana CD-ROM

PTS = Pali Text Society

S = Samyutta Nikāya (PTS)

SHB = Simon Hewavitarne Bequest edition

SLTP = Sri Lanka Tripitaka Project

VRI = Vipassana Research Institute

## 【參考文獻】

### 一、原典與原典翻譯

Dhammapada, edited by O. von Hinüber and K.R. Norman, Oxford : PTS, 1994.

The Gāndhāri Dharmapada, edited by John Brough, London: Oxford University Press, 1962.

“Patna Dhammapada. Part I: Text”, edited by M. Cone, JPTS XIII, 1989, pp. 101-217.

The Samyutta-nikāya, edited by M. Léon Feer, vols. II, IV, London: PTS, 1888, 1894.

Udānavarga, Band I, II, edited by Hg. von F. Bernhard, Gottingen 1965.

Abhandlungen der Akademie der Wissenschaften in Gottingen.

Philologisch-Historische Klasse. Dritte Folge, Nr. 54.

The Connected Discourses of the Buddha : A New Translation of the Samyutta Nikāya, trans.

by Bhikkhu Bodhi, Boston: Wisdom Publications, 2000.

《法句經》（吳·維祇難等譯），《大正藏》冊4，210號。

《法句譬喻經》（西晉·法炬共法立譯），《大正藏》冊 4，211 號。

《出曜經》（姚秦·竺佛念譯），《大正藏》冊 4，212 號。

《法集要頌經》（宋·天息災譯），《大正藏》冊 4，213 號。

## 二、專書

水野弘元（1981）

《法句經 研究》。東京：春秋社。

黃昌寧、李涓子（2002）

《語料庫語言學》。北京：商務印書館。

p. 141

楊惠中主編，衛乃興等編著（2002）

《語料庫語言學導論》。上海：上海外語教育出版社。

Biber, D., S. Conrad, and R. Reppen (1998).

Corpus Linguistics: Investigating Language Structure and Use, Cambridge University Press.

Tognini-Bonelli, Elena (2001)

Corpus Linguistics at Work, J. Benjamins.

McEnery, T. and A. Wilson (1996)

Corpus Linguistics, Edinburgh University Press.

## 三、論文

莊德明（1995）

〈以《心經》為例說明如何利用計算機處理佛經的多版本〉，《佛教圖書館館訊》第 3 期。臺北：伽耶山基金會圖書資訊中心。

馮志偉（2002）

〈中國語料庫研究的歷史與現狀〉，《漢語語言與計算學報》(Journal of Chinese Language and Computing)，11(2)，頁127-136，新加坡。

黃希敏(2004)

〈語料語言學概述〉，《敦煌英語教學電子雜誌》，11月號。

(<http://cet.cavesbooks.com.tw/html/m0131100.htm>)

黃希敏(2005)

〈語料語言學研究面面觀〉，《敦煌英語教學電子雜誌》，3月號。

(<http://cet.cavesbooks.com.tw/html/m050307.htm>)

蔡奇林(2001)(翻譯·補注)

〈巴利學的現況與未來任務〉，《正觀雜誌》第18期，頁171-209。南投：正觀出版社。(譯自：K.R. Norman, "The present state of Pali studies, and future tasks", Collected Papers, vol. VI, Oxford: PTS1996, pp. 68-87)

蔡奇林(2002)

〈巴利學研究紀要：1995-2001〉，《正觀雜誌》第20期，頁227-83。

南投：正觀出版社。

蔡奇林(2004a)

〈文本的異讀與法門的轉變——從安那般那念的二個疑難句談起〉(講稿)，6月、9月講於中華佛學研究所、南華大學宗教學研究所。

蔡奇林(2004b)

〈「六群比丘」、「六眾苾芻」與「十二眾青衣小道童兒」——論佛典中「數·(群/眾)·名」仿譯式及其對漢語的影響〉，《佛學研究中心學報》第9期，頁37-72。臺北：國立臺灣大學文學院佛學研究中心。

p. 142

蔡奇林(2004c, 12月4日)

〈「原典語言」與「原典研究」的重要——從「不見水白鶴」的公案談起；兼行「無我相經」勘正〉，「宗教教育與宗教研究研討會」講稿，內政部暨南華大學宗教學研究所。

蔡奇林（2004d）

〈網海一滴：網路上的巴利教學與研究資源舉隅〉，《佛教圖書館館訊》第40期，頁21-35。臺北：伽耶山基金會圖書資訊中心。

釋惠敏、維習安、杜正民、郭麗娟、周邦信（2001）

〈漢文電子佛典製作與運用之研究——以《瑜伽師地論》為例〉，《中華佛學學報》第14期，頁43-53。臺北：中華佛學研究所。

Bingenheimer, Marcus（馬德偉）（2003）

“Issues in the Use of Electronic Markup for the Comparative Analysis of Agama Literature”，《中華佛學研究》第7期，頁361-384。臺北：中華佛學研究所。

Brand, Quentin and Joe. Lavallee（2005）

“Corpus Linguistics For Teachers”，《敦煌英語教學電子雜誌》，1月號。

（<http://cet.cavesbooks.com.tw/htm/m0140100.htm>）

#### 四、工具書

中央研究院詞庫小組（黃居仁、陳克健、賴慶雄）（1997）

《國語日報量詞典》。臺北：國語日報。

水野弘元（1955，昭和30年）

《——語文法》。東京：山喜房佛書林。

水野弘元著，許洋主譯（1986）

《巴利文法》，收於《世界佛學名著譯叢》第5冊。臺北：華宇出版社。

王還、常寶儒等（1986）

《現代漢語頻率詞典》。北京：北京語言學院出版社。

Collins Publishers and University of Birminham（1987）

Collins COBUILD English Language Dictionary, London: Collins.

Dhammakaya Foundation (1996)

Palitext Version 1.0 Buddhist Canon CD-ROM Users' Guide, Thailand:  
Dhammakaya Foundation.

Fransic, W. N., H. Kucera (1967)

Frequency Analysis of English Usage: Lexicon and Grammar, Boston: Houghton  
Mifflin.

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik (1985)

A Comprehensive Grammar of the English Language, London: Longman.

p. 143

## 五、光碟與網路藏經資源

1. 漢文藏經光碟及網站 (2006)

《CBETA 電子佛典集成》。臺北：中華電子佛典協會。  
(<http://www.cbeta.org/index.htm>)

2. 巴利藏光碟：

- (1) PTS 版：Palitext version 1.0 (Pali-CD), Bangkok, Dhammakaya  
Foundation 1996.
- (2) 泰國版：BUDSIR IV on CD-ROM (The Buddhist Scriptures  
Information Retrieval Release 4.0), Bangkok, Mahidol University  
Computing Center (MUCC) 1994.
- (3) 緬甸版：Chattha Savgāyana CD-ROM (Version 3), Dhammagiri,  
Igatpuri, Vipassana Research Institute 1999.

3. 巴利藏網站：

- (1) 泰國版 (BUDSIR IV)：<http://budsir.mahidol.ac.th>
- (2) 緬甸版 (VRI)：<http://www.tipitaka.org/tipitaka/booklistframe2.html>
- (3) 錫蘭版 (SLTP)：<http://jbe.gold.ac.uk/palicanon.html>

4. 巴利藏英譯網站：

- (1) 美國 Access to Insight：<http://www.accesstoinsight.org/>
- (2) 斯里蘭卡 MettaNet：<http://www.metta.lk/>
- (3) 馬來西亞 Mahindarama：<http://www.mahindarama.com/>

5. 初期佛教文獻 (多語本) 線上對照表：<http://www.suttacentral.net/oscp/>

6. 佛教文獻百科 (多語本對照語料庫) Thesaurus Literaturae Buddhicae

(TLB) : <http://folk.uio.no/braarvig/tlb/>

## 六、語料庫

### 1. 漢語語料庫

- (1) 中央研究院「漢籍電子文獻」資料庫及「上古漢語語料庫」  
<http://www.sinica.edu.tw/~tdbproj/handy1/>
- (2) 中央研究院「現代漢語平衡語料庫」  
<http://www.sinica.edu.tw/ftms-bin/kiwi.sh>
- (3) 北京大學漢語語言學研究中心「現代漢語語料庫」、「古代漢語語料庫」及「漢英雙語語料庫」  
[http://ccl.pku.edu.cn:8080/ccl\\_corpus/jsearch/index.jsp?dir=xiandai](http://ccl.pku.edu.cn:8080/ccl_corpus/jsearch/index.jsp?dir=xiandai)  
[http://ccl.pku.edu.cn/YuLiao\\_Contents.Asp](http://ccl.pku.edu.cn/YuLiao_Contents.Asp)
- (4) 北京大學計算語言學研究所、日本富士通研究中心「現代漢語標注語料庫」
- (5) 北京語言文化大學「現代漢語研究語料庫」  
<http://xinghb.nease.net/keti/cmcr.htm>
- (6) 北京語言文化大學「現代漢語詞頻統計語料庫」

p. 144

### 2. 英美語語料庫

- (1) 布朗語料庫 (Brown Corpus)  
[http://www.essex.ac.uk/linguistics/clmt/w3c/corpus\\_ling/content/corpora/list/private/brown/brown.html](http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html)
- (2) 美國國家語料庫 (American National Corpus, ANC)  
<http://americannationalcorpus.org/>
- (3) LOB 語料庫 (The Lancaster-Oslo/Bergen Corpus)  
[http://www.essex.ac.uk/linguistics/clmt/w3c/corpus\\_ling/content/corpora/list/private/LOB/lob.html](http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/LOB/lob.html)
- (4) SEU 語料庫 (The Survey of English Usage)  
<http://www.ucl.ac.uk/english-usage/>
- (5) 國際英語語料庫 (The International Corpus of English, ICE)  
<http://www.ucl.ac.uk/english-usage/ice>
- (6) 英國國家語料庫 (British National Corpus)  
<http://www.natcorp.ox.ac.uk/>
- (7) COBUILD 語料庫 (Collins Birmingham University International Language Database)  
<http://www.collins.co.uk/books.aspx?group=140>
- (8) Longman 語料庫 (The Longman-Lancaster English Language Corpus)

p. 145

# A Discussion of the Present Situation and Future Prospects of Pāli Digital Resources:

With some Remarks on the “Multilingual Parallel Corpus of Early Buddhist Texts”

Tsai Chi-lin  
Assistant Professor, Nanhua University  
Institute of Religious Studies

## Abstract

During the last 10 years, due to the rapid developments in information technology, great achievements have been made in the digitization of Pāli texts. The author has commented on these achievements before and the first part of the present paper serves to summarize our previous findings.

The second part critically examines different aspects of these resources, including the directions and characteristics of their development, their benefits for research, and unresolved problems. In summary it can be said that, although, in order to offer a more critical basis for research, the digital resources are constantly improved in terms of their volume, their editorial criticism, and their search and extract functionality, so far only relatively low levels of refinement and detail (ie, lower criticism) have been achieved.

In the last part of this paper we will in a case study try to gauge the possibilities for future developments in Pāli digital resources. Our planned project tries to build a parallel corpus of Buddhist scriptures that is multilingual, spans different versions and translations and encodes “linguistic annotations” in order to surpass previous resources and express the texts on a higher level of refinement and detail (ie, higher criticism).

This kind of multi-lingual parallel corpus of different editions, can not only raise the level of criticism of traditional Buddhist studies, but also support research in the translation of Buddhist scriptures, as well as translation itself, and record the understanding of language in the process of translation. Moreover the corpus will offer benefits for linguistic research and language learning. With its help we will be able to produce better language reference books such as grammar books and dictionaries. It is a multi-functional, multi-valent database that shows a possible direction for the future development of Buddhist digital resources.

**Keywords :** 1.Pāli 2.Pāli texts 3.digital resources 4.electronic Buddhist scriptures 5.Early Buddhism 6.text corpus 7.multi-lingual parallel corpus

(English abstract translated by Marcus Bingenheimer)

[\*] 本文初稿曾在 2005 年 9 月 16 日於國立臺灣大學圖書館、佛學研究中心、東亞文明研究中心共同主辦的「佛學數位資源之應用與趨勢」研討會口頭發表，題作「巴利數位文獻資源的現況與未來」。今稿基本架構大體不變，但內容做了大幅度的改寫與補充。

[1] 感謝兩位審查者（甲和乙先生／女士）細心審閱本稿，並惠賜許多寶貴意見。這些意見筆者有些直接採納並更正文稿，有些則作了補充說明。此外，審查意見中，有些觀點筆者認為可以補充本文的部分看法，這些意見，文中將完整抄錄，以便提供讀者更加多元及有益的訊息。

[2] 參 K. R. Norman 著、蔡奇林譯，〈巴利學的現況與未來任務〉，頁 204-207。

[3] 有關巴利數位文獻資源的現況，筆者曾有兩篇文章談及，本節將這些內容重新作比較扼要的陳述，詳細可參原文：〈巴利學研究紀要：1995-2001〉及〈網海一滴：網路上的巴利教學與研究資源舉隅〉。其中前一文的第 6 節是有關巴利光碟的介紹，而後一文則是巴利網路資源的介紹。

[4] 為了行文方便，以下經常直接稱作 PTS 版、泰國版、緬甸版等。

[5] 這個版本精確應稱作 VRI 版，因為它不管是內文的編輯（有些段落有增添），或是相關頁碼都已不同於原來的緬甸第六結集版，但文中為了容易與錫蘭版、泰國版等對稱，仍稱之為緬甸版。

[6] 即羅馬字（Roman）、緬甸字（Myanmar）、天城體（Devanagari）、泰國字（Thai）、錫蘭字（Sinhalese）、高棉字（Khmer）、蒙古字（Mongol）等。

[7] 錫蘭版巴利藏（SLTP 版）網址：<http://jbe.gold.ac.uk/palicanon.html>。

[8] 緬甸版巴利藏（VRI 版）網址：  
<http://www.tipitaka.org/tipitaka/booklistframe2.html>。

[9] 泰國版巴利藏（BUDSIR IV）網址：<http://budsir.mahidol.ac.th>。

[10] MettaNet 網址：<http://www.metta.lk/>。

[11] Access to Insight 網址：<http://www.accesstoinsight.org/>，再點選“Tipitaka”。

[12] Mahindarama 網址：由 <http://www.mahindarama.com/> 進入首頁，在此頁最下方的“e-Service”選項欄中，點選“e-Tipitaka”即可進入三藏英譯網頁。

[13] 本節所羅列的，主要為收錄數量較多、或部帙較完整的幾個英語之外的譯本。漢譯雖有元亨寺版《漢譯南傳大藏經》，且在某些網站可以見到（簡體字版），但經筆者向該出版社查證，目前尚未提供任何光碟版或網路版，因此此處沒有列入。日譯《南傳大藏經》，筆者尚不知是否有光碟版或網路版。

[14] 以下資源大多在拙文〈網海一滴：網路上的巴利教學與研究資源舉隅〉中已有詳述，但本文也補充了幾項新資料。

[15] <http://dsal.uchicago.edu/dictionaries/pali/>。

[16] <http://www.metta.lk/pali-utils/index.html>。

[17] <http://www.metta.lk/pali-utils/Pali-Proper-Names/index.html>。

[18] [http://www.buddhanet.net/pdf\\_file/palidict.pdf](http://www.buddhanet.net/pdf_file/palidict.pdf)。此書雖名為 Buddhist Dictionary（佛教辭典），但因全書所據文獻為巴利一系列的典籍，且書中所釋，皆為佛教教理方面的詞目名相，為了比較清楚的表明其性質與內容，此處譯作《巴利佛教教理辭典》。

[19] <http://www.accesstoinsight.org/glossary.html>。

[20] <http://www.vri.dhamma.org/publications/pali/primer/index.html>。此書已有中文版，參釋性恩，〈巴利語入門〉（法雨道場，2005年）；網路版：

<http://nt.med.ncku.edu.tw/biochem/lsn/newrain/>。

[21] <http://www.metta.lk/pali-utils/ele-pali.doc>。

[22] <http://www.metta.lk/pali-utils/index.html>。

[23] 由 <http://ccbs.ntu.edu.tw/BDLM/index.htm> > 「語言教學」 > 「巴利文教學」 > 「4. 實用巴利語文法」。

[24] <http://ccbs.ntu.edu.tw/BDLM/index.htm> > 「語言教學」 > 「巴利文教學」 > 「3. 經文選讀」。

[25] <http://nt.med.ncku.edu.tw/biochem/lsn/newrain/>（法雨道場網站的「好書介紹」網頁）。

[26] <http://ccbs.ntu.edu.tw/BDLM/index.htm> > 「語言教學」 > 「巴利文教學」 > 「1. 字母與發音」或「3. 經文選讀」。

[27] [http://www.concentration.org/b\\_chants.htm](http://www.concentration.org/b_chants.htm)（泰國法身寺課誦）。

[28] <http://www.buddhanet.net/audio-chant.htm>。

[29] 由於筆者手邊沒有泰、緬、錫三版的紙本三藏可供比對，因此尚無法評估其紙本與電子版之間的品質差異，以下只以 PTS 版為例，做初步評述。

[30] (a) K. R. Norman 對 PTS 版編輯過程的種種問題，以及對於此版中存在的錯誤之警告，參拙譯〈巴利學的現況與未來任務〉，頁 191-196。茲摘錄如下：

「在其他文章，我曾就巴利原典如何被編輯出來，提供了一些訊息。有時，編輯者只是把一個或數個東方本子的讀法（readings）再版就滿足了，絲毫不想考察這些本子的根據。……有時它就直接從這個底本抄刊異讀（variant readings）。……至於那些東方本子所據為何，則隻字未提，而且也未從中引出任何異讀資料。此外，也未說明此校訂本是根據什麼原則建立起來的。……另外有些刊本，在刊印時有部份甚至全部尚未經過校對。

「我們很懷疑，許多從事上座部佛教相關著述的人，以及根據此種校訂本從事翻譯或研究的人，是否了解到這種事實。而即使對這種缺陷有所了解的人，也未能對它做出任何改善——或者因為時間不足，或者是對巴利語的掌握還不到足以訂正錯誤的地步。

「我必須澄清，巴利聖典協會對於它所出版的原典的疏缺，以及不可避免的，由此而造成的翻譯上的錯誤，知之甚詳，並且急於謀求改善。然而不幸的，錯誤的本子並不會只因想要改善的願望與念頭就自動改善，而有能力又有意願對前人校本進行訂正的人才實在難以尋得。經常是舊本須要訂正之處過多，結果是需要出版一全新的校訂本，而非單只舊本的修訂。可是巴利學領域年輕一輩的新進研究者，通常不甘於從事前人校訂過的典籍，而更樂於校訂尚未出版過的原典，即使這些新東西價值並不高。」

(b) 日本巴利學者水野弘元也指出，PTS 版訛誤相當多，相對而言，泰國版較為精良，因此從事研究及翻譯時必須詳加參照對勘，才能取得比較可靠的成果。參氏著《 — 語文法》，頁 213（中譯為許洋主譯，《巴利文法》，頁 304）。

[31] Dhammakaya Foundation, Palitext Version 1.0 Buddhist Canon CD-ROM Users' Guide, pp. 46-51.

[32] 但是，這片號稱精度已有相當提昇的光碟，其實還留下大量錯誤尚未訂正——數量恐怕超過一半以上，甚至更多。這是由於巴利原文的校訂工作，往往牽涉到相當多的專業知識，機器與一般人力通常只能檢查出部分字詞拼讀問題，至於其他文法問題、斷句問題、異讀問題、文句倒亂問題等等，都須要在深入研讀經文的基礎上，才能進行簡擇與判讀，這自然是更加耗時費工的，後文將會舉一些實例說明。但是由於數位資料的訂正再版更加容易，我們可以期待未來的新版本，會不斷訂正這些錯誤。

[33] 例如，《相應部》的標點，全編只採用“||”及“|| ||”兩種分隔或停頓符號，很難傳達出更加豐富的意涵與訊息（例如引述、疑問、感嘆等等）。又如，「問答體」是初期佛教文本中極重要的特徵，因此「問句」的表現就顯得格外重要，但 PTS 版《中部》第一冊雖已採用現代標點，但全冊竟沒有使用到任何「問號」，這也是須再改善的地方。

[34] 拙稿〈文本的異讀與法門的轉變——從安那般那念的二個疑難句談起〉（講稿）。

[35] 後文將會論及，緬甸版在這方面做得比較精良；錫蘭版也可參考，但還有一些問題；泰國版斷句及標點太少，參考價值有限。

[36] 即《相應部》第 12 相應（因緣相應）第 35 經（S II 60-63），以及第 35 相應（六入處相應）第 133 經（S IV 121-124）。

[37] 在 19 處拼讀錯誤中，主要是長短音、捲舌音、喉音等特殊辨音符號（diacritical marks）的標讀錯誤（如 Savatthiyam、jivam、aññam、taṇhāti、bhikkhu、sankhārā 等，應作 Sāvattthiyam、jīvam、aññaṃ、taṇhāti、bhikkhu、saṅkhārā）；也有漏掉字母的（如 ekattam 應作 ekatthaṃ）；或者漏掉空格的（如 sarīrantivā 應作 sarīranti vā）；或是多了空隔的（如 yā nissitāni、tālavatthu katāni，應作 yānissitāni、tālavatthukatāni）。這些拼讀錯誤顯然不是底本的問題，應該也不是編校者的疏忽，而是來自打字人員的手誤，以及打字文稿未經仔細審校的緣故。

[38] 即 S II 62 第 14 行的 yānissitani 及 19 行的 Sabbānissitāni，較好的讀法是泰、緬、錫三版的 yānissa tāni 及 Sabbānissa tāni，這從構詞、句法、及前後語意判斷都更合理。因此前一讀法很可能是該版所據的底本之抄經手的手誤。

[39] 即該經第 14 段末兩行“Avijjāya tveva bhikkhu . . . kānicia kānici”應移至第 15 段開頭，類似的，第 15、16、24 段末的上述內容也應移至下一段開頭。這個錯誤是由於編校者沒有細讀及貫通全經的意義脈絡及句法結構所造成，這樣的分段及斷句，肯定會讓初學者讀得滿頭霧水，挫折連連。

[40] 這些拼讀錯誤中，除了常見的辨音符號及字母增減的問題之外（如 pahīṇāni、ekattaṃ、ucchinnaṃmūlāni、anuppādadhammāti 等，應作 pahīnāni、ekatthaṃ、ucchinnaṃmūlāni、anuppādadhammāni），也有一些可能是數位化過程中衍生的（如 Ḥkatamo、taṃkāyitāni，應作 Katamo、visūkayitāni）。但比較值得注意的是，有 3 處（乘 4 次）文法錯誤，即用 katame、ime、añña 修飾或限定女性字主格的 taṇhā 及 vedanā，這三個字應改作 katamā、imā、añña。這個錯誤出現的段落，其他三版都是省略文（pe），因此很可能是此版的抄寫者或編者在將省略文展開時的一個文法疏忽，還未被訂正過來。

[41] 這個語序問題是“aññassa bhavo añña panāyaṃ bhavo ti”，應作“añña bhavo aññassa panāyaṃ bhavo ti”。類似的，upādānaṃ、taṇhā、. . .、nāmarūpaṃ、viññānaṃ 等 7 段的該句也有同樣問題。後 7 段在其他版本都是省略文，因此很可能原本只有 bhavo 一段二詞顛倒，但抄經手或編者不察，在擴充經文時，又據此而衍生了另 7 處錯誤。

[42] 即該經 72 頁 128 段的 hotiti，應作 hotīti。

[43] 即該經 VRI 版頁 55 的“Avijjāya tveva . . . kānici kānici”與“Katamaṃ jarāmaṇaṃ, . . .’iti vā”之間不應該用句號斷開，因後一段其實正是表達前一段之先行詞的具體內容，應置於同一句之中。

[44] 拼讀問題與前經略同，如 maṇavako、hoti、oṇitapattapāṇim、paṇitena、tuṇhibhāvena 等，應作 māṇavako、bhoti、oṇitapattapāṇim、paṇitena、tuṇhibhāvena。

[45] 這兩個異讀問題，PTS、錫、緬三版都一樣，這裡一併說明。其中一個是 S IV 124 第 12 行的 ayya Udāyi（呼格），此語或許應讀如校勘注 1 以及泰國版的 bhante Udāyi，這個稱呼用法在 Nikāya 的其他地方一律讀作 bhante Udāyi，因此，ayya Udāyi 很可能是受到下文主格形 ayyo Udāyi（原作 Udāyī）的影響而來。另一個異讀是人名 Udāyi（優陀夷）的主格形，在 PTS、錫、緬三版都取 Udāyī，似乎視作 -in 結尾的詞，但這個名字（有多人同名）在經律的絕大多數版本中都視作 -i 結尾，並且那些主格取 Udāyī 的地方，也多有異讀作 Udāyi，反之，主格作 Udāyi 的地方卻絕少異讀作 Udāyī 的，因此它很可能如泰國版所判定那樣，是個短 i 結尾的詞，主格應作 Udāyi。

[46] 情形略同前經，如 brāhmaṇim、tuṇhibhāvena、paṇitena、hoti、bhuntāvīm、arahato 等，應作 brāhmaṇim、tuṇhibhāvena、paṇitena、bhoti、bhuttāvīm、arahanto；另外，同前經一樣，也發生文法問題，如 Adhivāsesi、jivhāsmim，應作 Adhivāsetu、jivhāya。

[47] 這個錯置的句子在該經 BJT 版頁 250，“ghānasmim asati . . .”之前的一句“kho bhagini sati arahanto sukhadukkhāṃ paññāpentī”，應移至後一句的“jivhāya（原作 jivāsmim）asati . . .”之前，且該句 kho 之前應補上一字“jivhāya”。

[48] 泰國版有 4 處拼讀錯誤：bhotī（2 處）、khādanimiyena、kevalaparipañṇam，此三詞應作 bhoti、khādanīyena、kevalaparipunṇam。一個異讀問題是，該經 153 頁 ācariyāya，應讀作其校勘注 1. Ma. Yu. 二版的 ācariyabhariyāya（此版 154 頁以及另三版也採此讀法）。

[49] 緬甸版一個拼讀問題及斷句問題發生在同一句中，即該經 VRI 版第 127 頁的“Adhivāsetu kira, bhavaṃ, udāyi, amhākaṃ ācariyabhariyāya . . . svātanāya bhattan’ti”。畫底線處，udāyi 採短 i 結尾且前後逗點，編者似將此詞讀作呼格，但實際上，此處應為主格，以配合動詞 Adhivāsetu。因此若照緬甸版的一致讀法，此句應讀如此版下文（頁 128）的“Adhivāsetu kira bhavaṃ udāyī amhākaṃ ācariyabhariyāya . . . svātanāya bhattan’ti”，也就是應採主格形 udāyī（長 ī 結尾，因為此版將此詞視作 -in 結尾的字），且前後不需逗號。

[50] 參 Bhikkhu Bodhi, “Preface” in *The Connected Discourses of the Buddha: A New Translation of the Samyutta Nikāya*, p. 12。

[51] (a) 對於此處的調查與評述，審查者甲提出了寶貴意見，茲抄錄如下：「雖如作者指出，PTS 版《相應部》有許多打字錯誤，又說 Bhikkhu Bodhi 在從事《相應部》新譯之時，以緬甸版（第六結集版）作為基礎底本，但 Bhikkhu Bodhi（Preface p. 12）也說：the PTS’s roman-script edition（which itself draws from older Sinhala and Burmese versions）. . . as time went on I found myself

increasingly leaning towards the older Sinhala transmission as in many respects the most reliable. 因此，PTS 版的學術價值仍不亞於甚至超過第六結集版。」

(b) 確如審查者甲所言，儘管如以上調查所顯示，以及多位學者所指出，PTS 版編校問題頗大，但這並不意味此版缺乏學術價值，相反的，此版許多校訂者採用的底本（抄本或刊本）眾多，而校勘注一般而言也詳細豐富，這對於考察、了解各底本的種種問題，確實保留了珍貴的訊息，因此其學術價值（主要在保留眾多底本的異讀資訊上）仍是無可取代的。總括的說，目前緬、泰、錫、歐（PTS）各版，各有其優劣之處，要取得較好的讀法，自然不能偏依一家，而須要博校各版，取其長者，而棄其短者。

[52] 除了上一小節所舉的 Avijjāpaccayā 經的例子之外，再如，S. 12. 20 Paccayo 經（S II 25-27），第 4 段第 1 句“Jātipaccayā bhikkhave jarāmaṇam”，應移至前一段之末，作為 āha（說）的內容。同樣的，第 5 段第 1 句也應移至上一段之末。

[53] 例如 S. 35. 189 Bāḷisiko 經（S IV 158），其開頭的第 2 段整段內容，應歸屬於前一經（188 經）之末，189 經事實上應從第 3 段開始。這個錯誤，PTS 光碟版也未作訂正。

[54] 在各種語言譯本當中，漢譯是極缺乏的。漢譯的紙本文獻雖有元亨寺的《漢譯南傳大藏經》以及其他零星翻譯，但整體而言，或者品質尚待提昇，或者數量太少，仍待未來努力。

[55] 例如：CSCD 依據緬甸第六結集版；BUDSIR IV 的三藏及註釋等分別來自 Syāmaratṭha 版及摩訶朱拉隆功大學等數個版本；SLTP 的三藏根據 BJT 版（Buddha Jayanti Tripitaka Series），未來三藏的註釋將依據 SHB 版（Simon Hewavitarne Bequest edition）。

[56] 例如，錫蘭版在前述 S.12.35 Avijjāpaccayā 經的“visūkāyitāni”一詞上，附有一條校勘注：“sūkāyitāni – sī. Mu.”，但在整個網站或檔案中卻未說明 sī. Mu. 到底指的是哪兩個本子，以及這兩個本子的相關訊息。緬甸版（CSCD）與泰國版（BUDSIR IV）也有同樣情形。

[57] 例如 Saṃyutta Nikāya（相應部）第一冊，初版應是 1884 年，但光碟版列的卻是 1973 年。

[58] 例如馬來西亞 Mahindarama 網站。

[59] BUDSIR IV for Windows 無法在中文 Windows 系統下使用；而早先的 MS-DOS 版，要擷取利用經文，比 CSCD 更加繁複，不僅資料存檔後須重新編輯（因為會有一些段落重複），並且還須經過數道轉碼程序，才能在 Windows 下使用。

[60] 筆者曾經校訂過一些出版社的書籍，有時薄薄一冊不算多的巴利引文當中，就有上百處錯誤。

[61] 筆者所見，目前擷取及引用功能最強大、最友善的，要屬國內「中華電子佛典協會」所開發的《CBETA 電子佛典集成》。若有一天巴利光碟也能達到這樣的境界，當是學界的一大福音。

[62] 參 K. R. Norman 著、拙譯，〈巴利學的現況與未來任務〉，頁 195-196。

[63] (a) 在 2002 年初，國立中正大學中國文學系為了替該年年底的「漢文佛典語言學國際學術研討會」暖身，舉辦了一次「佛經語言研究座談會」，邀請國內外相關研究者討論佛經語言研究的種種問題與前景，筆者有幸與會，在會中曾提出這個擬構案例的雛形。

(b) 用計算機軟體技術來處理佛典「多語多本」的問題，在國內外已有一些實務案例。例如：莊德明，〈以《心經》為例說明如何利用計算機處理佛經的多版本〉；釋惠敏、維習安、杜正民、郭麗娟、周邦信，〈漢文電子佛典製作與運用之研究——以《瑜伽師地論》為例〉；Marcus Bingenheimer（馬德偉），“Issues in the Use of Electronic Markup for the Comparative Analysis of Āgama Literature”。另外，據筆者所知，中華佛學研究所目前正執行一個專案計劃，是以漢譯《別譯雜阿含經》為主的多語本對照研究計劃（計劃中尚包含將漢譯本英譯），計劃名稱為「漢文古籍譯註與數位編輯的研究——以巴利文與漢文《別譯雜阿含經》（T.100）的版本比對與英譯為例」。可見佛典的多語本處理，已越來越受到重視。就在此文完稿之際，筆者有幸從上述中華佛學研究所「別譯雜阿含經計劃」小組成員戎錫琴小姐處得知，在泰國出家的澳洲籍法師 Ajahn Sujato 正進行「線上藏經對照索引計劃」（Online Sutta Correspondence Project），建構初期佛教各種語本文獻之間的「對照表」，以利學者研究比對。該計劃先期階段以巴利 Nikāya 及漢譯《阿含經》為主，兼及梵、藏等其他語本或譯本，未來擬擴大到整個初期佛教所有語本文獻。（參 <http://www.suttacentral.net/oscp/>）此外，審查者乙也提供筆者一個重要資訊，即挪威奧斯陸大學（University of Oslo）的 Jens Braarvig 教授帶領一個研究團隊，自 2002 年開始，建構網路上的佛典多語本對照語料庫：Thesaurus Literaturae Buddhicae（TLB 佛教文獻百科，見 <http://folk.uio.no/braarvig/tlb/>）。這個語料庫可同時呈現一段本文（經文、律文等）的梵／巴、漢、藏、英等四種語本的內容（若相關語本存在的話），這樣便提供讀者或研究者，除了從宗教、哲學的向度看待或研究佛教之外，更可以從文化、翻譯、語文學、語言學、辭典編纂等各個面向去利用這一批珍貴的佛典文獻。這個語料庫的構想和旨趣，與本文所提，極為一致。

[64] 《法句經》的種種語本及傳本，例如，O.von Hinüber 及 K.R. Norman 編輯的 Dhammapada（巴利本）；John Brough 編輯的 Gāndhārī Dharmapada（犍陀羅語本，並附巴利本、其他語本對照），M. Cone 編輯的 Patna Dhammapada（佛教梵語本，巴特那法句經），Hg. Von F. Bernhard 編輯的 Udānavarga（梵語本，優陀那集）；古代漢譯本有吳·維祇難等譯《法句經》，西晉·法炬共法

立譯《法句譬喻經》，姚秦·僧伽跋澄與竺佛念譯《出曜經》，以及宋·天息災譯《法集要頌經》等。

[65] 例如水野弘元及 J. Brough 對各語本《法句經》的對照研究。參水野弘元《法句經 研究》；John Brough, *The Gāndhārī Dharmapada*。

[66] 筆者曾就《雜阿含·1006 經》的偈頌譯詞問題，比對巴利本及《別譯雜阿含·232 經》，確認其中在翻譯之前，已經存在底本傳誦或傳抄的錯誤；此外，關於「無我相經」的問題，筆者透過此經的不同傳本及語本，以及類同經的考察，確認《雜阿含·86、87、33、34 經》（無我相經）經文有誤傳，從而指出水野弘元對此經不甚確當的詮解。參拙稿〈「原典語言」與「原典研究」的重要——從「不見水白鶴」的公案談起；兼行「無我相經」勘正〉。

[67] (a) 語料庫 (corpus) 的建構，在國內外已有許多案例，國外的案例，例如：二十世紀六〇年代，Francis 與 Kucera 在美國布朗大學 (Brown University)，建構了以研究當代美國英語為目的的「布朗語料庫」(Brown Corpus)；七〇年代初由 Leech 倡議，S. Johansson 主持，建構以研究當代英國英語為目的的「LOB 語料庫」(The Lancaster-Oslo / Bergen Corpus)；1988 年，為了對同樣講英語的不同國家的英語進行對比研究，Greenbaum 提議研製的「國際英語語料庫」(The International Corpus of English, ICE)；1991-1995 年間，由英國政府與牛津大學出版社、朗文 (Longman) 集團、錢伯斯 (Chambers) 出版社、英國國家圖書館、牛津大學、蘭卡斯特大學等多個單位合作建構了「英國國家語料庫」(The British National Corpus, BNC)；由於美語與英語的許多差異，為了有利於美語的研究，1999 年起，美國也開始建構「美國國家語料庫」(American National Corpus, ANC)，並於 2003 年完成 1100 萬詞的初步規模。國內的案例，例如：中央研究院所建構，超過 1 億 3 千 400 萬字的「漢籍電子文獻」資料庫 (內含「上古漢語語料庫」)，以及 1991-1997 年完成的 500 萬詞「現代漢語平衡語料庫」。中國大陸如：北京大學漢語語言學研究中心超過 1 億 3 千 100 萬字的「現代漢語語料庫」及「古代漢語語料庫」，北京語言文化大學的「現代漢語詞頻統計語料庫」，北京大學計算語言學研究所與日本富士通研究中心合作建構的「現代漢語標注語料庫」等。

(b) 有關利用語料庫進行語言學各面向研究的介紹，可參考黃昌寧、李涓子著，《語料庫語言學》；楊惠中主編、衛乃興等編著，《語料庫語言學導論》；Tony McEnery & Andrew Wilson, *Corpus Linguistics*; Douglas Biber, Susan Conrad, Randi Reppen, *Corpus Linguistics: Investigating Language Structure and Use*; Elena Tognini-Bonelli, *Corpus Linguistics at Work* 等。比較簡要的介紹，如黃希敏，〈語料語言學概述〉、〈語料語言學研究面面觀〉；馮志偉，〈中國語料庫研究的歷史與現狀〉；以及 Quentin Brand & Joe Lavallee, “Corpus Linguistics For Teachers”。

[68] 儘管「生語料庫」使用起來比較費力，功能也比較侷限，但依舊可以給研究帶來甚大幫助。筆者曾利用巴利藏光碟、《CBETA 電子佛典集成》、以及中央研究院「漢籍電子文獻」資料庫等三個非平行的生語料庫，進行印度語言通過佛典翻譯而影響漢語「數—量結構」的研究。參拙稿〈「六群比丘」、「六

眾苾芻」與「十二眾青衣小道童兒」——論佛典中「數·（群／眾）·名」仿譯式及其對漢語的影響）。

[69] 運用語料庫編寫文法書，其優點在於，透過大量真實語料的觀察，一方面可以減少編纂者的主觀臆度，再則在計量分析的支持下，對於哪些用法最為典型，可以得到比較客觀的了解。這種立基於經驗主義（empiricism）的方法，越來越受到語言學者的重視，而成為現代語言研究的主流。英國語言學家 Quirk 從 1959 年起，著手建立「英語用法調查語料庫」（SEU, The Survey of English Usage），就在 SEU 語料庫的支持下，1985 年由朗文（Longman）出版社出版了一部英語語法大全 A Comprehensive Grammar of the English Language。而目前根據 SEU 語料庫所編寫出的書籍，已經超過 200 部，可見語料庫具有十分可觀的「加值性」與極為長遠的「潛在效益」。

[70] 運用（計算機）語料庫從事辭典編纂，是現代辭典學發展的一大趨勢。1967 年，美國布朗大學出版社根據前述的「布朗語料庫」（Brown Corpus），出版了一部英語用法頻率詞典 Frequency Analysis of English Usage: Lexicon and Grammar。1980 年代，英國柯林斯出版社（Collins Publishers）和伯明翰大學（University of Birmingham）合作，在 John Sinclair 教授帶領下，建立了一個 2000 萬詞次的現代英語語料庫 COBUILD（Collins Birmingham University International Language Database）。在這個語料庫的支持下，1987 年英國柯林斯出版社出版了著名的 Collins COBUILD English Language Dictionary，讓全世界辭典編輯界耳目一新。1988-1990 年，朗文（Longman）語料庫小組建構完成一個全新的英語語料庫「朗文－蘭卡斯特語料庫」（The Longman-Lancaster English Language Corpus），用以編纂辭典和提供學術界使用。1979-1983 年，北京語言文化大學建立了一個規模 200 萬詞次的「現代漢語詞頻統計語料庫」，在這個語料庫的支持下，於 1986 年出版一部《現代漢語頻率詞典》。1997 年中央研究院詞庫小組黃居仁、陳克健和賴慶雄在中央研究院「現代漢語平衡語料庫」的支持下，編纂完成一部《國語日報量詞典》，解釋漢語「量詞」以及量詞和名詞搭配的用法。

[71] 所有藏經光碟的製作（如巴利、漢文、藏文等），除了大大有利於這些典籍的流通與典藏之外，其最重要的功用與目的，其實都在於提供學者作為研究的重要輔助工具。但如前所說，由於這些資料庫都屬於「生語料庫」，尚未進行語料的語言加工，因此對於語言研究的幫助仍大受限制。這也是此處之所以要特別說明「語料庫」的相關觀念與作用，並舉示當前語言學界紛紛建構「標注語料庫」以作為語言研究基礎的用意所在。

[72] (a) 有關數位化過程中，對於原來紙本重新校訂及改訂一事，審查者乙提供了寶貴意見，茲抄錄如下：「作者自頁 120 以降多處提及數位原典校訂了過去的錯誤而取得更好的讀法。不過在嚴謹的版本學研究中，一般會更同時重視未經更改的原抄本的原文表現。因為即使是現在看起來是錯誤者，有可能保存了語言發展上珍貴的訊息，更何況有些異讀是值得商榷的。電子版有意無意的更動了來源抄本的原來面貌，雖然利益現在的讀者，卻有可能遺失許多資訊，久而久之順理成章取代原文。是否可以考慮區分兩種電子版本：一者完整而忠實地保存各種抄本，另一種才是所謂的現代校訂本？」

(b)誠如審查者乙所言，重新校訂固然重要，但原抄本的保存同樣不可忽視，二者自可並行而不悖。而此處所謂重新精校，自然須要謹慎從事，任何改訂之處，必須保留原來底本內容，並詳加說明改訂的理由，如此，讀者在新校版當中，同時可知舊版內容，並了解改訂之原委，相關學者也可在這些資訊的基礎上，進行考訂及抉擇。總之，以吾人今日所處的時代，各種語本及研究資源的取得，比起古代更加容易，因此，在數位化的過程中，除了保留古代的底本之外，實可藉由多語本互校的機會，進一步「改善」現存古本的某些問題，以取得一個比較精良的本子。而既然任何改訂，都難免存有疏闕，因此，保留並提供充分訊息「以待來者」，就十分重要了。

[73] 例如音韻學、構詞學、語義學、句法學、語用學、語言風格學、語篇分析等各方面。