

國立新竹教育大學台灣語言與語文教育研究所碩士論文

語料庫的建置與台灣客家語 VP 研究



指導教授：羅肇錦 博士、呂菁菁 博士

研究生：謝杰雄

2006 年 6 月 8 日

本論文獲行政院客家委員會 96 年
客家研究優良博碩士論文獎

語料庫的建置與台灣客家語 VP 研究

| | |
|----|---|
| 摘要 | I |
|----|---|

第一章 序論

| | |
|-----------------|---|
| 第一節 研究動機與緣由 | 1 |
| 第二節 研究目的 | 3 |
| 第三節 本研究的語料範圍和限制 | 3 |
| 第四節 研究方法 | 8 |

第二章 文獻探討

| | |
|------------------|----|
| 第一節 客家語構詞法研究回顧 | 11 |
| 第二節 客家語短語研究回顧 | 13 |
| 第三節 客家語的詞類體系研究回顧 | 13 |

第三章 「T3」語料庫的建置與規則

| | |
|-----------------|----|
| 第一節 語料庫的取樣原則和方法 | 16 |
| 第二節 語料的整理 | 21 |
| 第三節 語料的存儲形式 | 25 |

第四章 語料的分詞規範與說明

| | |
|----------------|----|
| 第四節 「詞」的定義 | 28 |
| 第五節 確認「詞」的方法 | 30 |
| 第六節 各詞類的切分具體說明 | 37 |

第五章 客家語詞類劃分與標記說明

| | |
|----------------|----|
| 第一節 詞類的劃分與標記方法 | 53 |
| 第二節 詞類鑑別的工作守則 | 57 |

第六章 客家語短語結構分類與標記說明

| | | |
|-----|------------|----|
| 第一節 | 短語的分類與標記符號 | 73 |
| 第二節 | 短語性與標記符號 | 75 |
| 第三節 | 短語的標記與示例 | 77 |
| 第四節 | 短語類型的判定 | 81 |
| 第五節 | 句子的層次關係與分析 | 92 |

第七章 建立台灣客家語的剖析樹語料庫

| | | |
|-----|--------------------|-----|
| 第一節 | 標記 (tag) 符號與規則說明 | 96 |
| 第二節 | 檔案格式說明 | 97 |
| 第三節 | 介紹 Dotted Tag | 99 |
| 第四節 | 句法結構分析編輯工具 | 102 |
| 第五節 | 「T3lan」語料檢索工具 | 104 |
| 第六節 | 語料標記樣例 | 108 |
| 第七節 | 剖析樹的產生步驟與實例 | 111 |

第八章 「T3」語料庫語料統計與分析

| | | |
|-----|-----------|-----|
| 第一節 | 各詞類使用頻率分析 | 113 |
| 第二節 | 短語統計與分析 | 115 |

第九章 客家語動詞研究

| | | |
|-----|-------------------|-----|
| 第一節 | 動詞的音節統計與分析 | 117 |
| 第二節 | 動詞的構詞形式 | 119 |
| 第三節 | 客家語的核心動詞 | 124 |
| 第四節 | 四縣客家語和海陸客家語動詞對比分析 | 127 |

| | | |
|-----|--------------|-----|
| 第五節 | 華語、客家語動詞對比分析 | 140 |
|-----|--------------|-----|

第十章 動詞短語 (vp) 的結構與關係判定

| | | |
|-----|-------------------|-----|
| 第一節 | 動詞短語 (VP) 的內部構造 | 153 |
| 第二節 | 動詞短語 (VP) 的外部功能 | 166 |
| 第三節 | VP 內部結構關係判定的問題 | 168 |
| 第四節 | PP VP VP | 170 |
| 第五節 | VP VP 不同結構關係的限制條件 | 182 |

第十一章 結語

| | | |
|------|------|-----|
| 第一節 | 回顧 | 187 |
| 第二節 | 研究成果 | 188 |
| 第三節 | 研究展望 | 190 |
| 參考書目 | | 191 |
| 附 錄 | | 198 |

國立新竹教育大學

語料庫的建置與台灣客家語 VP 研究

論文精要

本論文以建置一個單機型語料庫為主，做為嘗試規劃和發展一個 WEB 語料庫的基礎。

第一章緒論，說明本研究的動機、目的和方法，說明語料的範圍和來源。本研究的目的是透過對客家語語料的分析，以面向語料庫加工處理為目標導向，主要目的有以下幾點：

1. 建立一個適切而可操作的台灣客家語語料的分詞標準規範。
2. 建立一個可驗證、可操作的台灣客家語語詞歸類操作標準。
3. 建立一個台灣客家語「詞類劃分體系」與「詞類標記」的標記，以進行語料庫的加工工作。
4. 分析台灣客家語的「短語結構」，並對各種結構予以「標記」，用以標記語料庫中的客家語語料的句法信息。

本論文則進一步利用該語料庫針對客家語的動詞和動詞短語，透過分析與考察、以呈現客家語動詞短語內部和外部的結構。我們總計分析了總計 602326 個漢字的台灣客家語語料，其中 432969 個漢字，全部做了詞類標記，另 162347 個漢字的語料，我們做了結構分析。

第二章 文獻探討，說明有關客家語在構詞法、短語、詞類體系等三方面前人研究的成果與值得借鑒之處。

第三章 「T3」語料庫的建置與設計，說明設計該語料庫的緣由，語料的取樣原則和方法，語料的整理與存儲形式。

第四章 語料的分詞規範與說明，本章說明我們的分詞方法和定義，提出一個規範來作為操作的標準，並具體舉例說明切分的原則。

第五章 客家語詞類劃分與標記說明，本章說明劃分客家語的基本詞類為

十八類，並為每一詞類建立詞類標記符號與示例，我們劃分客家語的詞計十八個基本詞類，分別為：名詞(/nn)、時間詞(/nt)、方位詞(/lc)、處所詞(/si)、動詞(/vv)、形容詞(/jj)、副詞(/rb)、數詞(/cd)、介詞(/pa)、感嘆詞(/ij)、連接詞(/cc)、擬聲詞(/on)、助詞(/ax)、區別詞(/dis)、量詞(/m)、語氣詞(/sf)、狀態詞(/zht)、代詞(/pn)。而基於標記的需要，對「名詞」劃分出「專有名詞(nr)」；對「數詞」，劃分出「序數詞(/od)」。另對「系動詞」單獨給予標記「/vs」；領屬標記「介」單獨給予標記「/kai」；外來語(_fw)、成語(_id)；標點符號(/pu)等，總計我們採取了 25 個標記符號。同時提出每一詞類鑑別工作守則，做為標記時的操作標準。

第六章 客家語短語結構分類與標記說明，本章說明劃分客家語的短語為二十二種類型，並為每一短語建立短語標記符號與示例。短語類型及標記符號如下：

主謂短語(cp)、述賓短語(vp)、述補短語(vp)、連動短語(vp)、狀中短語(vp)、聯合短語(vp)、附加短語(vp)、聯合短語(np)、方位短語(np)、數量短語(np)、同位短語(np)、定中短語(np)、介(kai)字短語(np)、時間短語(np)、處所短語(np)、狀中短語(jp)、聯合短語(jp)、介(kai)字短語(jp)、形補短語(jp)、比況結構(jp)、介詞短語(pp)、副詞短語(rp)、

第七章 建立台灣客家語的剖析樹語料庫，本章說明採取 Dotted Tag 的方式，把客家語的語法單位分為「語素、詞、短語、子句、句子」五級單位，設計 T3 語料庫編輯程式和 T3lan 檢索軟體，透過對客家語的詞、短語、句子的層次標記，建立剖析樹語料庫 (treebank)。

第八章 「T3」語料庫語料統計與分析，本章針對已建置完成的語料庫語料統計與分析各詞類和短語的使用頻率。經過現有的語料的分析，我們一共得到 1299 個單音節動詞，而就詞類統計的結果，名詞佔最大一部份，有 64.43%強，其次是動詞佔 20.43%，兩者合計就佔去了 84.86%強，再次是形容詞佔 7.91%，其他各類詞僅佔非常小的比例。另針對短語加以統計結果，客家語的使用者，主要使用的是「動詞性短語」，佔 52.74%，其中又以述賓 (vo)、述補 (vc)、狀中

(zd) 和附加 (fg) 爲主，連動和聯合短語相對偏低。

第九章 客家語動詞研究，本章說明抽取語料庫中所有客家語的動詞進行統計與分析的結果。就數量而言，客家語動詞是雙音節爲主，但單音節詞動詞的使用總頻率卻遠高於雙音節動詞。雙音動詞中又以述賓式和述補式最爲普遍。另就文本分析的結果，客家語有 56 個核心動詞，其中與現代華語不同的計 23 個：「講 1、食 1、分 1、無、喊、愛 1、係、知、行、噉、著 1、轉 1、走 1、驚、倚、落、尋、loi53、做 3、愁、刷、囡、嫖」，佔全部核心動詞的 41.1%。這 23 個與現代華語不同的核心動詞總使用詞次爲 7850 次，佔 56 個核心動詞總使用詞次的 44.58%，高頻的現象，顯示這是核心中的核心，相對的穩定性也非常高。

第十章 動詞短語 (vp) 的結構與關係判定，本章集中討論有關客家語的動詞短語 (VP) 的結構和結構關係的判定問題，歸納出台灣客家語各種動詞短語 (VP) 的類型，並以客家語的「分」爲例，探討客家語 PP VP VP 序列切分的規則。我們就可以判定：

- (1) 若後一個 VP 可以和「分」字組合，整個結構是向右切分，構成

a、(PP <VP VP>)模式

- (2) 若後一個 VP 不能和「分」字組合，整個結構是向左切分，構成

b、(<PP VP> VP)模式

- (3) 若按 a、(PP <VP VP>)模式切分，兩項 VP 之間可以是述補、聯合、連動三種結構關係，而述補和聯合關係都有嚴格的限制條件。

- (4) 若按 b、(<PP VP> VP)模式切分，整個結構可能是聯合或連動關係，但都有許多限制性條件。

第十一章 結語，透過這個語料庫的建置，我們找出了 1299 個客家語的單音節動詞；詞類統計結果，名詞佔最大一部份，有 64.43%強，其次是動詞佔 20.43%；客家語的使用者，主要使用的是「動詞性短語」，佔 52.74%，其中又以述賓（vo）、述補（vc）、狀中（zd）和附加（fg）為主；而客家語動詞是雙音節為主，但單音節詞動詞的使用總頻率卻遠高於雙音節動詞；同時我們描述了客家語 VP 的基本結構和分布面貌，並以 PP VP VP 的組合為例，考察被動式的「分」字和 VP 組合的各種條件限制。



關鍵字：語料庫、客家語、動詞、vp

第一章 序論

第一節 研究動機和緣由

喜歡玩電腦的我，在一次呂菁菁老師上課的內容中接觸到「中央研究院平衡語料庫」，那種豐富而涵蓋面大的語料庫，讓我興起一種衝動，假如客家話也有這麼一個語料庫，一定可以提供更多、更好、更細緻的研究出來，所以我決定接觸這方面的知識，嘗試建置一個語料庫，了解得深入些才發現原來這不是件單純語料集合、堆疊的事。程式問題還好解決，最難的是語料的取得、版權問題、文字化問題、標音問題、語料整理和語料庫規劃問題等等。

正好在我不知所措之餘，清華大學統計所江永進老師也準備為閩南語建置這麼一個語料庫，在獲得呂老師的牽引之下，筆者也加入這個團隊，專門負責製作「客家語」部分，於是我們透過每週一次的研討和實作，逐步製作了「T3」語料庫。編輯程式由江永進老師撰寫，提供我們建置語料庫的多種功能和方便性。其間羅肇錦老師以其豐厚的、精鍊的客家語能力，提供筆者關於客家語的詞彙、短語、語法結構等各方面的知識，為建置這個語料庫挹注非常大的心力；呂菁菁老師則在語料庫語言學方面，協助筆者規劃和編輯與建置這個語料庫，並在運用語料庫做進一部研究方面提供許多寶貴意見。

語料庫語言學（corpuslinguistics）是指利用語料庫內的語料對語言的語音、詞彙、語法，甚至語用、篇章等進行分析研究，對所研究的語言對象提出觀點或理論的研究手段。語料庫語言學為語言學研究提供了一種全新的研究思路，透過對大量語料的分析，從中找語言使用的規律，對自然語言進行定量的分析，提高研究結果的可靠性和真實性。

客家話的研究長期以來所偏向的是以研究「語音」為主軸的研究，特別是客家語的調查、記音及與中古音的比較方面做了大量的工作，相對於對客家語的詞彙、構詞、詞義、語義、句法和語法的分析、比較研究，則是一直處於有待開發的狀態。

雖然客家語的研究在「語音」研究方面取得了一定的進展和成就，研究方法以「田野調查法」為主，投注了不少的人力、金錢和精力，但是對調查所得的原始資料，卻僅掌握在調查人的手裡，很不便於他人取得以進行進一步的分析和運用，而調查人自身也無法以透過某種媒介的方式，將調查資料分享出來，即使讀者在研讀其所做的研究報告時，也沒有一個可以做為判定其所述內容真偽的標準尺。調查結果也往往受限於研究者和受訪者的種種因素影響了「客觀性」的有效度。

基於以上的種種原因和緣由，讓我體認到建置一個語料庫的重要性，唯有透過大量蒐集真實語言的材料，才能做到更深入的分析，拓寬客家語的研究視野，當你說別人的說法好怪時，你必須有大量的或一定明確的語料來支持並解釋「怪」在哪裡的證據，而不是只有個人感覺或少數人決定式的證據。當別人說海陸客家話比較像閩南話時，我想問的是「比率是多少？可以給我一個數據或更明確的證據嗎？」。基於以上種種的原因，筆者在指導老師的協助下，決定嘗試建置一個結合客家話、閩南話、華語三者的「平行語料庫」，以做為繼續研究的基礎。而基於必須對語料庫中的生語料進行加工「標記」工作的需求，我們必須做一些前置性的基礎研究工作，建立各種規範。在加工工作過程當中，筆者深切的體認到要建置一個語料庫並為語料標記做好適切規劃的迫切需求，在這樣的需求下，讓筆者產生了很強烈的動力，期望對自己的母語能盡一點微薄的心力。

第二節 研究目的

本研究的目的是透過對客家語語料的分析，以面向語料庫加工處理為目標導向，主要目的有以下幾點：

1. 建立一個適切而可操作的台灣客家語語料的分詞標準規範。
2. 建立一個可驗證、可操作的台灣客家語語詞歸類操作標準。
3. 建立一個台灣客家語「詞類劃分體系」與「詞類標記」的標記，以進行語料庫的加工工作。
4. 分析台灣客家語的「短語結構」，並對各種結構予以「標記」，用以標記語料庫中的客家語語料的句法信息。
5. 本論文則進一步利用該語料庫針對客家語的動詞和動詞短語，透過分析與考察、以呈現客家語動詞短語內部和外部的結構。

第三節 本研究的語料範圍和限制

本研究的語料從語料腔的角度來分可以分為「台灣四縣腔客家語」和「台灣海陸腔客家語」；從語料分類的角度來分可以分為「文本語料」和「口語語料」；從語料性質來分類可以分為「散文、小說、雜文、口述」等多種語料，以下是本研究的語料範圍，我們用語料腔調來區分：

一、「台灣四縣腔客家語」語料

- (1) 《苗栗縣民間文學集 4》《苗栗縣客家語故事集》(一)

本故事集屬文本式口述語料，1998年由台灣苗栗縣立文化中心出版，羅肇錦、胡萬川總編輯。本故事集計傳說八篇、民間故事十篇，由劉桂妹女士和賴阿桶先生口述，分別由羅肇錦、陳秀琪、陳裕美三位整理，並附有客家語和華語對照。

(2)《苗栗縣民間文學集 6》《苗栗縣客家語故事集》(二)

本故事集也屬文本式口述語料，1998年由台灣苗栗縣立文化中心出版，羅肇錦、胡萬川總編輯。本故事集計二十四篇語料，由涂敏松等七位講述，年齡層從五十四歲到七十二歲，分別由羅肇錦、陳秀琪二位整理，並附有客家語和華語對照。

(3)《阿啾箭个故鄉》

這是一本散文集，屬文本語料。2004年由作者龔萬灶先生發行，行政院客家事務委員會贊助出版。這本散文集一如作者所述是用四縣客家語的思維撰寫，全部語料含前言計四十四篇文章。語料內容以描寫早期客家人台灣的生活情景為主，並增添一些民間故事內容，並以保留傳統客家話的口語、用詞為主。是一份很有價值的語料。

(4)《細老鼠同番豆》

這是一本用客家語寫的童話故事書，屬文本語料，2004年由桃園縣政府文化局出版，作者張捷明先生，用一隻聰明的客家老鼠，一面說一面演，運用客家話裡傳統的詞彙，栩栩如生的表達客家話的對話，全部語料計十九章。

(5)《台灣客家文學選集 2》

這是一本小說集，屬文本語料，李喬先生主編，龔萬灶先生用客家語改寫，2004 年由行政院客家事務委員會發行，改寫自吳濁流、鍾理和等作者的小說原著，總計九篇，含每篇前面的導讀，經過整理後總計約有 87542 個漢字。

(6)《山歌 8072 首》

這是一份收集來的山歌詞，屬文本語料，文學性較強，每首 28 字，計 226016 字。由於是山歌詞，當中有不少是代用字或非口語詞，我們尊重原山歌詞，不予更動，只在用字上予以一致性，如「心舅」改為「新婦」；「嘴」改為「喙」。

(7)《客家語實用手冊》

這是一本收集客家話特有詞彙、師父話、鈴仔（謎語）、童謠等，同時兼具一些研究性資料如量詞、語氣詞等綜合性的文本語料，基於分析的需要，我們捨去詞彙部分，取出其中的例句和師父話、鈴仔（謎語）、童謠等納入語料庫中。

以上就是本研究中已整理好的四縣客家語語料，山歌 8072 首我們依據一般社會的印象，歸入「台灣四縣腔客家語」語料，這實在是現階段不得已的做法，將來也許找到更合理的方法時再進行更動。

(二)「台灣海陸腔客家語」語料

(8)《桃園縣民間文學集(6)》：《龍潭鄉客家語故事(一)》

這是一本 2003 年由桃園縣政府文化出版，胡萬川教授總編，李敏惠、曾新

蔘、黃英偉等採錄，葉瑞娟、陳素主整理的文本語料，包括民間故事十則和笑話五則，講述者為張良溪、葉義妹、甘必開三位，年齡層從七十一歲到七十七歲。是屬於一種綜合性的文本語料。

(9)《桃園縣民間文學集(16)》：《新屋鄉客家語歌謠謎諺(16)》

這是一本 2003 年由桃園縣政府文化局出版，胡萬川教授總編，姜通男先生採錄，古國順教授審稿的文本語料，包括「童謠、一般歌謠、四句、笞頭歌、謎語、諺語」六種，是屬於一種綜合性的文本語料，講述者有二十四位，年齡層從四十三歲到九十三歲。

(10)《桃園縣民間文學集(17)》：《新屋鄉客家語故事(一)》

和前面《新屋鄉客家語歌謠謎諺(16)》一樣，是桃園縣政府文化出版的桃園縣民間文學集系列之一，包括故事二十則和笑話十三則，屬於文本語料，講述者有十八位，年齡層從四十五歲到一百一十歲，採集範圍為桃園縣新屋鄉和觀音鄉，2003 年由桃園縣政府文化出版。

(11) 現代漢語八百詞翻譯

這是一份翻譯自《現代漢語八百詞》例句的翻譯語料，《現代漢語八百詞》一書是呂叔湘先生主編，用以說明現代漢語用法的一本書，我們採用的是《現代漢語八百詞》增訂本，2002 年商務印書館出版的版本。由筆者根據書中的例句，逐一翻譯成對應的客家話，採取的翻譯方法是語義對譯。由於筆者的母語是台灣海陸客家話（花蓮吉安），所以這份翻譯語料只有海陸客家話的思考模式。

在翻譯的過程中，筆者捨去全部的華語成語不譯外，其他則一邊譯一邊請

五位同樣以海陸客家語為母語的同事和親友討論檢驗，共得到 113216 個漢字的對譯語料。要特別說明的是會轉譯《現代漢語八百詞》一書中的例句，是應建置平行語料庫（parallel corpora）的需要，同時我們認為《現代漢語八百詞》內的「例句」具備各式各樣的句型，便於各種分析與對比，這是我們建庫之初的想法，由小而逐步擴大，但是由於筆者認為這份翻譯語料覆蓋面實在太小，所以筆者才又增加了其他語料。

這份翻譯語料筆者深知還有很多問題，有很大的改進空間，所以我們在舉語料例時盡量不舉這部份的語料，而用其他語料替代，這份翻譯語料的位階只定為參考用語料。

（12）280 分鐘的口語語料

為了反應語言使用的實際，筆者平日即錄下和家人、親友等的對話，再逐一譯成文字稿，現在譯好做為本研究的語料約有 280 分鐘，計 23427 個漢字語料。

以上是現階段「台灣海陸腔客家語」語料的概況，由於海陸客家語的文本有限，所以筆者採取這種翻譯和轉錄口語的做法，補足一部份。這些錄音語料裡包括四縣、海陸，甚至華語、閩南語都有¹，筆者全部把他們轉譯成海陸客家語，並直接把他們歸入海陸客家語語料檔中。

綜合以上的說明，我們一共取得 387542 個漢字的台灣四縣客家語語料和 214784 個漢字的台灣海陸客家話語料，總計 602326 個漢字的台灣客家語語料。這 602326 個漢字語料，我們實際上只處理了 432969 個漢字，全部做了詞類標記，當中有 264317 個字的語料我們歸入四縣客家語，而 168652 個字的語料我們歸入海陸客家語，相差約 100000 萬字，但是只有 162347 個漢字的語料，我們做了結構分析，所以本研究的語料範圍，將以這些已做詞類標記和結構標記的語料為限。

¹ 因為在實際溝通時各種語言的人都會適度的加入，所以這份錄音語料包含了四縣、海陸，甚至華語、閩南語都有，當然其中仍以海陸客家話為主。

這些語料有諸多的限制：

(1) 語料量較小而且涵蓋面不足，這是本研究上很重要的一個限制，雖然有 432969 個漢字的語料，與一般採用田野調查法的研究者，僅使用幾千個詞彙或單字調查表為依據的研究相比，語料量方面雖然大許多，但就語料庫語言學的角度而言，仍是有很大努力空間的。

(2) 本研究的發音人扣除山歌詞不計，總計六十四位，年齡層涵蓋六歲到一百一十歲，這種量與範疇對語料庫語言學而言，仍是有很大努力空間的。

(3) 語言知識的不足和經驗的缺乏

本研究在規劃語料庫和標記語料的過程當中，由於筆者的能力和對客家語語言知識的不足，相信對詞性和結構的誤判及標記錯誤勢必難免，當然也就造成統計結果的偏差亦不能免，這是本研究很需要努力進步的空間。

第四節 研究方法

國立新竹教育大學
National Hsinchu University of Education

本研究的重點是在於語料庫的建置及建置後的語料庫應用，語料庫的建置著重於規劃、設計和語料標記；建置後的語料庫應用則選取客家語的「動詞與動詞短語」進一步研究，嘗試用較大面積的語料，以語言的角度來觀察並解釋一些客家話的問題。研究的步驟大致如下：

- (一) 文本語料電子化：首先將收集到的語料予以電子化，主要是用人工打字或利用掃描機（scanner）等，將文本處理成文字檔。
- (二) 語料一致性處理：針對文本的用字予以一致性處理，關於一致性的方法和原則詳見如後。
- (三) 語料分詞與詞類標記加工。
- (四) 語料結構標記加工。

(五) 抽取「動詞與動詞短語」進一步分析比較。

除以上所述，我們從語料庫的角度來進行研究之外，由於本語料庫尚有許多待改進的問題，為使本研究語料和分析內容更加符合客家語言的事實，我們輔以「田野調查法、描寫語言法」來進行研究。

(一) 田野調查法 (field work)

本文所做的田野調查法和多數學者所做的田野調查法不同，多數學者以中國社會科學院所制定的《方言調查字表》為工具，或增刪單字，或增刪詞彙等方式建立調查詞表，以記錄和分析語音，並建立該受調查方言音系、分析比較詞彙為主，這是傳統上很多學者在使用的方法²。

本文則是自語料庫建成之後，抽取語料庫中的「動詞」構成進一步調查的詞表，是針對單一詞類所做的調查。調查重點在動詞的詞義和用法方面，特別是個別動詞的短語結構合法性調查。對調查結果的分析則重於詞義和語法分析，採用田野調查法的原因是我們必需仔細區分每一個動詞的義項分合問題。要讓計算機可以自動處理語料，相對而言對義項的分合要求是比一般要求還要更為嚴苛的，更為嚴苛的目的是建立「明確」的指標。

(二) 描寫語言法 (descriptive linguistics)

本文因為是客家方言的研究，對建置語料庫而言，現階段客家方言最大的

² 客家研究裡的田野調查法多數是走這個路線，如徐貴榮(2000)、賴文瑛(2004)、黃雯君(2005)等，基本上所使用的調查模式都是相同的。這種調查規模可以取得一定量的詞彙，並了解其音系狀況，但實際上對欲交給計算機處理而言是不足的。例如調查者可能調查出海陸客家話的「走」，記錄其語音為「tseu24」，也可能調查出與「走」相關的詞彙或短語等，如「走了、走忒、走出去、走醒、走淨淨」等，但是並沒有或缺乏描述每一個「走」的語法功能和屬性分析，更不可能將語言知識形式化，這是不可能交給計算機處理的。

問題是文字化和標準化的問題，在沒有標準化之前，我們僅能採取部份描寫語言法的方法，包括用記音的方式來處理，而對「動詞與動詞短語」進一步分析比較的工作，我們也較多地採取描寫語言法的方法。

本文寫作時採用由國際語音學會(The International Phonetic Association)製定的國際音標(The International Phonetic Alphabet, 簡稱IPA)紀錄並描寫語音，以寬式記音。³對客家話聲調的標註筆者採標調值法；而對引用語料的聲調標註則採原資料的標註法，除非有特別需要時才將其改爲相對應的「調值」標法。本文所使用的客家話的聲調符號表如下：

| | 陰平 | 上聲 | 陰去 | 陰入 | 陽平 | 陽去 | 陽入 |
|------|----|----|----|----|----|----|----|
| 調號 | 1 | 2 | 3 | 4 | 5 | 7 | 8 |
| 四縣調值 | 24 | 31 | 55 | 20 | 11 | 55 | 50 |
| 海陸調值 | 53 | 24 | 11 | 50 | 55 | 33 | 20 |
| 例字 | 三 | 嫂 | 曬 | 穀 | 無 | 盡 | 力 |

(表一，本文的客家話的聲調對照表)

³ 本文寫作時採用IPA記音，但是語料庫內的語料標音則是採行政院客家事務委員會94年客語能力認證初級考試所使用的標音系統，主要原因是因為IPA不支援我們的編輯程式，而且輸入不便。

第二章 文獻探討

客家語的研究，早期研究者關注的是詞彙的收錄與本字的考求，所以編制了許多方志、字典和辭典，例如：溫仲和（1898）《光緒嘉應州志》卷七方言部分考證了若干本字；黃釗（1900）《石窟一徵》卷七、卷八有零星的本字考證；楊恭桓（1907）《客家語本字》考證了一千四百多個有音無字的客家特殊詞；羅翹雲（1922）的《客方言》則是仿《爾雅》體例，收錄興寧客家語的詞彙；西方傳教士則編有《客英大字典》《客法大字典》，近期則有黃雪貞《梅縣方言詞典》、楊政男等（1998）《客家語字音詞典》、詹益雲編（1998）《海陸腔客家字典》、古國順等《現代客家語詞彙彙編、續編》等，詞彙對比研究的則有溫昌衍〈客家話特徵詞研究〉。

另有一批學者則關注在各次方言之間的調查與紀錄，特別是強調客家語和中古音或古漢語及中古漢語之間的關係與演變，例如：董同龢（1948）〈華陽涼水井客家話記音〉；楊時逢（1957）《台灣桃園客家方言》；羅肇錦《瑞金方言》、《台灣的客家話》；李如龍、張雙慶（1992）《客贛方言調查報告》；劉綸鑫（1999）《客贛方言比較研究》；藍小玲（1999）《閩西客家方言》；其他如廣東惠陽（周日健 1987）、大埔（何耿鏞 1993）、東勢（江俊龍 1996）、饒平方言（呂嵩雁 1993、徐貴榮 2000）、豐順（賴文英 2003）等，以上這些研究者或製有字音表、詞彙表、與中古音演變對照表，偶或涉及詞彙或用字的考釋。整體而言，對客家語的研究，以語音的描述與分析成果較為豐碩。

有關本文各主題的相關文獻，在各主題之下另有說明，這裡僅就（一）客家語構詞法研究（二）客家語短語研究（三）客家語詞類體系探索三個大項的相關文獻做一探討。

第一節 客家語構詞法研究回顧

對客家語的構詞方式有系統性研究的學者筆者認為最早是羅肇錦（1988）的《客家語語法》，羅肇錦定客家語構詞為九個類型：（1）並列格（2）向心格（3）主謂格（4）動賓格（5）後補格（6）重疊格（7）前加成份（8）後加成份（9）內部曲折。

接著較有系統的是張玲瑛（1992）的《客家語的構詞與語法研究》，集中討論的主題是關於客家語的複合詞、補語結構，和句尾助詞，並將客家語與國語和閩南語作一比較。

其次是張雁雯（1997）《臺灣四縣客家話構詞研究》，張雁雯主要是以臺灣苗栗四縣客家話為主，用結構描述法，對詞彙的結構作平面的分析，探討客家語的構詞現象，藉由對構詞形式的觀察，探討其功能及語法意義。

鍾榮富（2000）《福爾摩沙的烙印---台灣客家話導論》，則主張客家語和其他漢語方言一樣，可以「音節與詞素」為辨認標準的「單音節詞」和「多音節詞」的劃分法，並同時提出「重複、附加、複合及合併」四種客家語的構詞手段。

而江俊龍（1996）《台中東勢客家方言詞彙研究》，以東勢客家話為討論對象，分別討論客家話的單音詞、複合詞和派生詞，並做了部份詞源及外來詞等的探討，最有特點的是針對同義詞運用義素分析法進行分析。

古國順 2002〈客家語的詞彙特色〉，分別從構詞、詞類、形義配合及語言風格上四個角度，分析客家語的構詞法和詞類問題，全文簡要，重點是在說明「客家語的詞彙特色」，但仍對本文有參考價值，特別是構詞及詞類兩個部份。

以上諸位學者的研究主要是面向以人理解為主的分析，學者之間對客家語的「詞」這個語言單位的定義邊界持不同的看法，但卻又都沒有提出一個可以操作的和驗證的分詞規範，這種以面向人為主的構詞規律分析，雖然對筆者的分詞工作仍有重要的參考價值，然而對語料庫的建置和計算機自動化的要求，顯然還

是比較粗略了一點，需要更明確、更形式化的規範。

第二節 客家語短語研究回顧

「短語」是由詞或詞組依一定的語法規則組合而成的語法單位，客家語的短語研究，就筆者所知目前並沒有文獻，雖然羅肇錦（1988）《客家語語法》在第三章第二節提及關於「字詞與短語」的區別，但其目的主要是為說明「短語」這級語法單位的界線，並沒有針對短語這一級語法單位提出客家語的現象分析，也沒有對客家語的短語進行分類和歸類研究。這正是筆者希望能處理的面向，以利語料庫的建置和標記工作。

第三節 客家語詞類體系研究回顧

客家語的詞類問題文獻主要見於幾部相關的客家語語法研究著作：

羅肇錦（1988）《客家語語法》、張玲瑛（1987）《客家語的構詞與語法研究》、何耿庸（1993）《客家方言語法研究》、項夢冰（1997）《連城客家話語法研究》，其他散見的則有藍小玲（1999）《閩西客家方言》有部分針對長汀話的語法舉例，鍾榮富（2000）《福爾摩沙的烙印--台灣客家話導論》則針對客家的疑問句和否定句進行討論。以上這些著作中針對客家語的「詞類」問題提出討論和規劃的，筆者認為嚴格說來只有羅肇錦（1988）的《客家語語法》、項夢冰（1997）《連城客家話語法研究》。

羅肇錦（1988）的《客家語語法》參考漢語的馬建忠、黎錦熙、王力、呂叔湘、朱德熙、趙元任等諸家之說，分合去取得三大類十一小類，另加上象聲詞合為十二類；何耿庸（1993）的《客家方言語法研究》則依照詞和詞的結合關係，詞在句子中的職能以及詞的形式和形式變化這幾項標準，將客家方言的詞分為十一類；項夢冰（1997）《連城客家話語法研究》則是以朱德熙《語法講義》為參

照，以連城客家話為素材進行的分類體系，並沒有羅列出全部的體系，而是進行某些單一詞類構詞、語法功能或次類劃分的研究；盧彥杰（1999）《新竹海陸客家話詞彙研究》碩士論文第五章分客家語詞類為十類，其他如藍小玲（1999）《閩西客家方言》僅則在討論語法舉例時列有動詞、形容詞、時間副詞三類詞，謝永昌（1994）《梅縣客家方言志》列有分類詞彙二十八類，是採取語義和語法功能分類兼混的分類法；林立芳（1997）《梅縣方言語法論稿》雖未正式列出詞類，然而根據書中所列可知作者列有副詞、動詞、形容詞、代詞、介詞等詞類，劉丁衡（2002）《客家語文法》則劃分為九大詞類¹，僅先將各家劃分的情況表列如下，以利比較和說明：

| | 羅肇錦 | 項夢冰 | 何耿庸 | 盧彥杰 | 藍小玲 | 謝永昌 | 林立芳 | 劉丁衡 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 名詞 | + | + | + | + | | | | + |
| 代詞 | + | + | + | + | | + | + | + |
| 定詞 | + | | | | | | | |
| 量詞 | + | + | + | + | | + | | |
| 動詞 | + | + | + | + | + | | + | + |
| 形容詞 | + | | + | + | + | + | + | + |
| 副詞 | + | + | + | + | + | + | + | + |
| 介詞 | + | + | + | + | | + | + | + |
| 連詞 | + | | + | + | | + | | + |
| 助詞 | + | | + | + | | + | | + |
| 嘆詞 | + | | + | | | | | + |
| 象聲詞 | + | + | | + | | | | |

¹另有羅慶武（1997）的《客語詞類講解》，書中將客家語詞的詞性分為形容詞篇、動詞篇、民俗篇三類。這本書是基於保存和推廣客家話的角度來做分類，分為形容詞、動詞篇、民俗篇，並根據詞句的句數分為二至十五字部，羅列許多客語的詞彙和語句，雖然書名為《客語詞類講解》然而其所謂「詞類」的意義實際上是「詞或語句的意義講解與使用說明」，與語法研究上所謂的「詞類」意義無關，所以本文不予計入。

| | | | | | | | | |
|-----|--|---|---|--|--|---|--|--|
| 數詞 | | + | + | | | + | | |
| 方位詞 | | + | | | | | | |
| 處所詞 | | + | | | | | | |
| 時間詞 | | + | | | | | | |

(表二，客家語詞類劃分比較表)

以上各家對客家話所劃分的詞類體系，從劃類數量來說差異非常大；就劃類標準而言，各家各有取捨權衡，筆者認為諸家大體上是以側重語法功能、兼顧意義和形態的方式來進行分類。

總的說來，各家詞類劃分體系最大的難點是並未給出一個明確的、可操作的、可驗證的標準，並建立一致性的詞類體系劃分標準，以提供客家語句法、語法研究方面的分析。



第三章 「T3」¹語料庫的建置與設計

自二十世紀八十年代以來，計算機語料庫的建置與應用，逐漸受到許多國家資訊界和語言學界的重視，在話語識別、人機對話、機器翻譯、文字處理等各方面均具有相當的重要性和潛力，同時計算機語料庫以其語料的充分性、客觀性、可驗證性、使用的便捷性，使語言學的研究產生質與量的改變，基於語料庫的研究方法已經成為當代語言學和應用語言學研究中的一種主流。

本章主要在說明我們如何歸類與整理語料，分為三小節（一）、語料庫的取樣原則和方法（二）、語料的整理（三）、語料的存儲形式

第一節 語料庫的取樣原則和方法

語料庫是語言採樣的樣本集合，為使「充分性、客觀性、可驗證性、使用的便捷性」等目標得以實現，首先遇到的問題便是語料的「代表性」問題，語料有代表性才能對目標語言整體做出合乎實際的推理和判斷，而語料庫的代表性則取決於語料庫中的文本類別、量、長短、版權等問題，這些問題歸結起來就是「取樣」問題。所以考慮語料庫中需包含多少文本、文本的性質、每個文本需包含多少詞次等各方面必須花費相當多的時間和精力。

「T3」語料庫客家語部分在設計與建置過程中，考慮目前客家語文本產生的情況和文本版權等問題，為兼顧反應語言的實狀，「T3」語料庫除將現有蒐集到的部份文本電子化以外，還包括了二百八十分鐘的對話語料錄音檔轉錄成文字檔，做為本研究的範圍。取樣原則以「口語」化為優先考量，口語形式比書面形式更能體現語言的基本組成與特點，是許多研究者普遍的共識，所以「T3」語料

¹ 「T3」是指Taiwan（台灣）的三種語言，即閩南語、客家語和華語而言，這個名稱是由江永進老師取名的。

庫在選擇客家語樣本時首先著重「口語」化。所謂「口語」化，是指文本的紀錄是否如實地反映「口語」而言。

取樣的方法則是將搜集到的文本不加以分類，不考慮文章的量、長短、流通度、標準差等因素²，將文本篇章分別加以編號後，用亂數表選出二十篇文章，分別找兩位以海陸客家語為母語和兩位以四縣客家語為母語的人³為受測人，逐一閱讀文本，同時我們設計了一個「口語」化量表，由受測人依自己的感覺填寫，我們定義受測人認可平均「口語化值」⁴需達八十五分以上者才列入合乎「口語」化的標準，並採納為本研究的語料。

下表是接受判定語料樣本別及口語化值的人：

| 姓名 | 性別 | 學歷 | 年齡 | 客語腔調別 |
|-----|----|----|----|-------|
| 蔡華廷 | 男 | 無 | 65 | 海陸 |
| 徐朝龍 | 男 | 無 | 84 | 海陸 |
| 彭明宇 | 男 | 國中 | 57 | 四縣 |
| 周美嬌 | 女 | 國中 | 62 | 四縣 |

(表三，判定語料樣本別及口語化值人)

因為客家語的文本並沒有註明是屬於四縣腔調或海陸腔調，所以在這個量表裡頭，我們特別要求受測人判定語料樣本是「海陸客家語或四縣客家語」以觀察這兩種客家語之間的相通度，並要求受測人將文本與各自口語差異之處標示出來，做為判斷的依據。

我們用亂數表選出的二十篇文章，請四位受測人加以判定，根據四位受測人所填寫的「口語」化量表值，統計結果如下：

² 若要考慮這些因素，需要很大的各方面、各領域的樣本才能做得到，以現階段客家語的樣本而言，實在是差距很遠，所以我們略去這些因素，只以是否「口語」化為唯一的取樣考量。

³ 這四位受測人完全是以海陸或四縣客家話為母語的人，海陸客語的兩位受測人能聽簡單的四縣客家語，但不會說四縣客家語；而四縣客家語的兩位受測人則對海陸客家語聽說都有困難。

⁴ 對「口語化值」的量表值，我們不細分計分項次，而採取「總合」值的方式。

| 序號 | 篇章名 | 口語 化值 | 腔調 | 腔調 | 鑑別詞 | 原文出處 |
|-----|-----------------------|----------|----|----|--------------|-------------------|
| 1. | 阿啾箭个故鄉 | 97 | 四縣 | 四縣 | 落雨、揚尾仔、 | 阿啾箭个故鄉 |
| 2. | 大安河个水鬼 | 93 | 四縣 | 四縣 | 圓身 | 苗栗縣客家語故事集 4 |
| 3. | 對面山排伙房屋 | 94 | 四縣 | 四縣 | 斷真 | 阿啾箭介故鄉 |
| 4. | 祠堂 | 91 | 四縣 | 四縣 | 撿採、譴、雨點 仔 | 台灣客家文學選集 2 |
| 5. | 百歲婆 | 92 | 海陸 | 四縣 | | 台灣客家文學選集 2 |
| 6. | 伯公出巡 | 95 | 海陸 | 四縣 | | 楊梅鎮客家語故事（一） |
| 7. | 雜文（一） | 87 | 海陸 | 海陸 | | 翻譯自漢語八百詞 |
| 8. | 豺食先生 | 93 | 海陸 | 四縣 | | 楊梅鎮客家語故事（一） |
| 9. | 特效藥 | 94 | 海陸 | 四縣 | | 楊梅鎮客家語故事（一） |
| 10. | 山歌 50 首 | 92 | 四縣 | 四縣 | | 亂數選出的山歌 50 首 |
| 11. | 賭狗三 | 94 | 海陸 | 四縣 | | 龍潭鄉客家語故事（一） |
| 12. | 防癌，低脂高纖食 物 m-he 萬能 | 72 | 海陸 | 四縣 | | 客台語專刊第五十三期 |
| 13. | 火焰蟲（二） | 94 | 海陸 | 四縣 | | 新屋鄉客家語歌謠謎諺 （一） |
| 14. | 人生有高低 | 91 | 海陸 | 四縣 | | 新屋鄉客家語故事（一） |
| 15. | 一下仔就大仔十過 歲 | 92 | 海陸 | 四縣 | | 細老鼠同番豆 |
| 16. | 其姆平常時介吩咐 | 91 | 海陸 | 四縣 | | 細老鼠同番豆 |
| 17. | 白頭公 | 91 | 海陸 | 四縣 | | 苗栗縣客家語故事集 6 |

| | | | | | | |
|-----|-------|----|----|----|--|-------------|
| 18. | 三娘教子 | 92 | 海陸 | 四縣 | | 新屋鄉客家語故事（一） |
| 19. | 阿四妹 | 92 | 海陸 | 四縣 | | 苗栗縣客家語故事集 6 |
| 20. | 雜文（二） | 90 | 海陸 | 四縣 | | 翻譯自漢語八百詞 |

（表三，口語化量表統計結果）

表三顯示筆者所提供的客家語文本資料，除「防癌，低脂高纖食物 m-he 萬能」這篇文章以外，其他各文經受測人判定語料樣本的結果是口語化程度相當高，但是在語料樣本是屬於四縣客家語或是海陸客家語的「腔調」鑑別方面，除有很明顯的詞彙可據以判定屬於哪一種腔調之外，其他則無從判斷，這顯示四縣客家語或是海陸客家語的相通度非常高。

這便引起我們語料歸類上的困擾，另外使用海陸客家語的兩位鑑別人表示，「山歌五十首」他們列入「四縣」客家語的原因是一般而言客家山歌使用「四縣腔調」來唱，所以他們將「山歌五十首」列入「四縣」客家語，所以社會習慣，也是影響歸類的因素。為解決這些問題，我們另外提出十個鑑別詞，作為鑑別的參考準則：

| 序號 | 四縣 | 海陸 | 說明 |
|----|----|----|--|
| 1 | 个 | 介 | 領屬格 ⁵ 四縣話讀為「ke ⁵⁵ 」，海陸話讀為「kai ¹¹ 」，一般四縣文本多做「个」，海陸文本則作「介」。 |
| 2 | 同 | 搵 | 連接詞方面四縣客家語一般用「同」，海陸客家語用「搵」。 |
| 3 | 落雨 | 落水 | 「下雨」四縣客家語用「落雨」，海陸客家語用「落水」。 |
| 4 | 譴 | 闕 | 「生氣」四縣客家語一般用「譴」，海陸客家語用 |

⁵ 這裡所謂「領屬格」即相當於華語的「的」，客語文本中，一般四縣文本多做「个」，海陸文本則作「介」。

| | | | |
|----|-----|----|---|
| | | | 「闕」。 |
| 5 | 吾 | 偈 | 表示領屬「我的」，四縣客家語用「吾ŋa ¹¹ 」，海陸客家語用「偈gai ⁵⁵ 」，如「我的爺爺」，四縣客家語說「吾公」，海陸客家語說「偈公」。 |
| 6 | 麼个 | 麼介 | 疑問詞「什麼」，四縣客家語說「麼介mak ²⁰ ke ⁵⁵ 」，海陸客家語說「麼介 mak ²⁰ kai ¹¹ 」。 |
| 7 | 刷 | 刷 | 四縣客家語說「cii ¹¹ 」，海陸客家語說「chi ⁵⁵ 」，像這種「精莊知章」分合的現象，我們列為區分的標準。 |
| 8 | 該 | 該 | 表示遠指的「該」，四縣客家語說「ke ¹¹ 」，海陸客家語說「kai ⁵⁵ 」。 |
| 9 | 恁仔細 | 承蒙 | 表示「謝謝」，四縣客家語用「an ³¹ tsii ³¹ se ⁵⁵ 」，海陸客家語說「in ⁵⁵ muŋ ⁵⁵ 」。 |
| 10 | 筷子 | 箸 | 筷子，四縣客家語說「k' uai ⁵⁵ e ³¹ 」，海陸客家語說「t' u ¹¹ 」。 |

國立新竹教育大學 (表四，文本腔調別鑑別參考詞)

National Hsinchu University of Education

除此之外我們還參考黃雯君（2004）的比較，來確定文本的腔調別，同時若能找到原作者的，我們直接向其詢問清楚。

根據以上的幾個特性，我們將文本重新分類，但前揭鑑別詞也並不是絕對的標準，例如作家龔萬灶先生的「阿啾箭个故鄉」一書，在序言中作者宣稱該作品是用「四縣客家話个思維」，但是在作者文中頁 131：「阿松伯姆見擺想著這個妹仔，心肝就盡闕。」，同樣使用了「闕」這個詞彙，但四縣客家語的「闕」是指「生悶氣」而言；屬於四縣腔的「苗栗縣客家語故事集 4」也使用了海陸腔的「拞」這個連接詞，如頁 146「錢財如糞土 仁義值千金」，這篇文章中「拞」這個詞彙也出現了 18 次之多，如「就去拞錢財公做長年」、「五六介家丁就拞蛇哥扛到海唇」、「也有人拞飯菜煮好好」等等，但是我們根據其他的鑑別詞，仍將

「苗栗縣客家語故事集 4」列入「四縣腔」的語料範疇中。

而楊梅客家語故事集中也使用了「吾」這個詞彙，例如第 32 頁「²，這係吾爸⁵，樣還有阿爸？」，根據原書的標音是「ŋa⁵」⁶，這是語言接觸所產生的「四海話」特徵，「聲、韻」是四縣腔的說法，而調卻是海陸腔的調，但是根據其他的鑑別詞，如「麼介」仍做「mak⁴ kai³」、「刷」仍做「chi⁵」、「該」做「kai³」，我們仍將龍潭客家語故事集歸入「海陸腔」的語料範疇中，對已成爲四海腔的詞彙，則維持其原來的語言事實面貌，不加以變動。

第二節 語料的整理

客家語由於長期以來並沒有發展出標準化、規範性的文字化文本，這是很不利於語料庫建置、研究和推廣的，更導致相關研究軟體的研發和設計上的困擾，爲解決這些問題，使本研究得以進行並擴大語料庫的用途，我們針對語料文本進行整理，以取得語料的一致性，我們的語料整理可分爲以下四方面：

(一) 用「字」的一致性：

客家語的文字化問題一直是很困擾的問題，各作者或各教材之間用字上的出入造成語料庫建置上很大的困擾。對語料庫而言「一致性」是非常重要的，語料庫語言學強調對語料樣本的統計分析，並對所取得的數據加以解釋和判斷，用字的「一致性」不足，將導致統計結果的偏差，造成推理和結論的錯誤。所以我們在建置語料庫的同時花費相當多的時間在維護語料用字的一致性上。關於用字的一致性方面，以下分四點說明：

⁶ 這裡尊重原作的標音，標調號。

(1) 同音、同義者取其一：

一個詞出現在不同的文本或一個文本的重複出現，根據上下文判定為「同詞、同義」者，我們擇一入庫，如客家人說「媳婦」是「𪗇𪗇⁵³ 𪗇𪗇⁵³」，有做「新白、心白、心舅、新舅、辛白、新婦、辛婦、新kiu1、薪白、新婦」等各種寫法，我們擇一統一改為「新婦」。至於根據什麼標準選定哪一個，並沒有一定的標準，原則上以能辨義、電腦輸入能處理及用字普遍通用等為基本原則，非謂所選即是本字。

(2) 同音異義者分別標示，視為不同的詞：

一個詞的詞形相同、語音相同，但不同義者，我們根據上下文，視為不同詞處理，如客家話的「走」(tseu³¹ / tseu²⁴)有許多義項，分舉例如下：

1. 跑：毋好用走介，會跌 (𪗇𪗇³¹ 𪗇𪗇³¹) 倒。
2. 離開：時間差毋多咧，毋走會赴無著火車哦！
3. 死亡：厥爸好命人，歲數恁大了，平平安安，去睡一覺目，無痛著斯走了。

像這種情形，我們分別用走₁、走₂、走₃來標示不同的義項，一義一詞，其語法功能亦各自不同，所以分開處理，視為不同的詞。

這個工作就是「詞的歸併問題」，是相當費神的工作，對於義項的分合，基於我們的研究方向需要，我們暫時把重點放在「動詞」這個詞類上，其他則比較粗略些。

(3) 異音、異義卻同形者，我們採取標注的方法做為區分，如：

「降下來」如果是讀成「𪗇𪗇³¹ 𪗇𪗇³¹ 𪗇𪗇³¹」，「降」是指「降落、

本乾淨、清楚，便於搜尋和分析。如下面這段語料：

「毋過，講正講，阿乾哥係毋領頭³，麼人就奈毋何，任在你有嘴講到無口湍又乜無彩工。到落尾，有兜細賴仔想愛又愛毋著，就冷冷故故⁴講麼个：「三擇四擇，擇著爛瓠杓⁵。」」

原作者有為「領頭³、冷冷故故⁴、爛瓠杓⁵」三個詞做注釋說明，我們去除這些「注釋」，使語料成爲：

「毋過，講正講，阿乾哥係毋領頭，麼人就奈毋何，任在你有嘴講到無口湍又乜無彩工。到落尾，有兜細賴仔想愛又愛毋著，就冷冷故故講麼个：「三擇四擇，擇著爛瓠杓。」

所以進入語料庫查詢所得的語料沒有註解標記，如要了解其原意，須查考原文本，而「註解」一般多用華語註解說明，更不適宜和客家語文本混合進入語料庫中，所以我們全部刪除，不予入庫。

(四) 將「國語詞彙」改回「客家語詞彙」

有些作者或許因爲先寫完國語文本稿，再轉譯爲「客家語」文本，這種文本會有漏譯的情形，或者因當時的研究水平，客家語詞彙的文字尙不確定等因素，而暫時用華語詞彙代替等情形，我們根據上下文推斷，如能找到原作者的，向其詢問，經確認之後改回「客家語」詞彙，無法找到原作者的，則依上下文判斷，改回「客家語」詞彙，如：「苗栗縣客家語故事集 4」頁 148，原文做「我倆儕就結婚、我無爺無哀、我來若介屋家分你請好無」，其中「我」這個詞彙應做「偌」，根據上下文及同一篇文章中「我、偌」並存的事實，我們全部改回合乎客家語語言事實的文本。

以上是關於語料的整理工作，如果不加以整理，而直接入庫，對語料的統計、分析和研究，將造成深遠的影響，所以我們花費了許多時間整理這些語料，雖然如此，仍不免會有遺珠之處。

第三節 語料的存儲形式

在計算機中存儲語料，現在的計算機儲存媒體不論在方式、量、技術等各方面均已不是問題，所以對語料的儲存形式要考慮的是如何標示「語料的時間、語料腔調、語體的分類、檔名的邏輯性」等問題，在建庫之初就必須加以確認這些問題。

(一) 生語料儲存的設計

生語料是指語料的原始電子文本，即不帶任何語言訊息的文本的電子化檔案，生語料又分為兩種，一種是「原始文本生語料」，一種是「整理過後具有一致性的生語料」。原始文本生語料是依據原始文本錄入的生語料；整理過後具有一致性的生語料是根據前面所述對用字、標音、注釋、詞彙較正等處理過的生語料，整理過後具有一致性的生語料是進一步語料加工的原始文本。

儲存的檔名非常重要，需具有邏輯性，並要能帶上語料的相關訊息，便於檢索、統計、分析等作業，我們的設計是採取六碼設計的形式，分述如後：

第一碼代表「語料時間」：定義為原始語料採集時間，如果不知採集時間的，則以出版時間為準，如果任何時間都沒有標示的，則以入庫時間為準，時間的設計是用西元年為代表。如 1993_，2001_

第二碼代表「文本名」：目前的作法是採取原文本名前兩字的客家語拼音做爲標示，如龔萬灶先生所寫的「阿啾箭介故鄉」，第二碼就定義成「aziuzian_」

第三碼是「文本序」：如龔萬灶先生所寫的「阿啾箭介故鄉」的第一篇散文是「阿啾箭介故鄉」，則編碼成「01_」

第四碼是「語料腔調」：S 表四縣腔、H 表海陸腔、D 表大埔腔、R 表饒平腔、Z 表詔安腔、SH 表四海腔、O 其他腔調

第五碼是「語體分類」：暫時只分四類，N 表「小說」、V 表「散文」、P 表「詩詞」、C 表「雜文」。

第六碼是「生語料標示」：org 表「原始文本生語料」，orgok 表「整理過後具有一致性的生語料」。

根據以上的六碼設計，呈現「語料的時間_文本名_篇章序_語料腔調_語體的分類_生語料別.txt」的儲存格局，以龔萬灶先生所寫的「阿啾箭介故鄉」的第一篇散文「阿啾箭介故鄉」爲例，原始生語料表述爲儲存形式則是：2004_aziuzian_01_S_V_org.txt；如果是整理過後具有一致性的生語料表述爲儲存形式則是：2004_aziuzian_01_S_V_orgok.txt。

（二）對「標記語料」儲存的設計

生語料要進一步成爲有用的語料，就必須進行標記，「標記語料」分爲兩種，一種是「詞類」標記，一種是「句法」標記。標記詞類是朝向句法標記的一個階段，是其他一系列任務的基礎，對標記了詞類的文本儲存的設計除前面的五碼具

備之外，第六碼改爲「le」，如龔萬灶先生所寫的「阿啾箭介故鄉」的第一篇散文「阿啾箭介故鄉」標記了詞類之後，表述爲儲存形式則是：2004_aziuzian_01_S_V_le.txt；如果是標記了「句法」信息，則第六碼改爲「str」，如龔萬灶先生所寫的「阿啾箭介故鄉」的第一篇散文「阿啾箭介故鄉」標記了句法之後表述爲儲存形式則是：2004_aziuzian_01_S_V_str.txt



第四章 語料的分詞規範與說明

第一節 「詞」的定義

要進行語料加工，做語料詞性標注、句法標注等工程，首先碰到的問題便是對一個成分是否「成詞」的認定問題，客家語文本和漢語文本一樣，使用的是一個個的漢字，詞和詞之間沒有明顯的標記，文本的分詞對客家語的分析和理解、檢索、統計、教材編訂、甚至將來有可能的機器翻譯等各領域，都具有重要的實用價值。根據語言學的討論，語言學者一般定義「一個能獨立運用且不能再被分割，具有獨立意義，而扮演特定語法功能的字串視為一個詞。」，雖然這仍是見仁見智，未能充分解決的問題，但是基於研究的需要，我們參考華語的研究，並考量客家語的特性，提出一個客家語的「分詞規範」，做為操作的依據和程序。

什麼是「詞」？這是一個東西方學者討論多年、傾注心力研究至今仍然懸而未決的問題。研究分詞問題必須先知道：什麼是詞？知道什麼是詞，就可以製定辨認法則，根據法則把它們找出來，這就是分詞；知道怎樣分詞，就可以進行分詞操作，分出的單位就是詞。要讓計算機做自動分詞，首先必須告訴計算機什麼是詞？需要製作一個分詞詞表使之匹配或智能化，而製定這樣一個分詞詞表又需要一個分詞規範，做為判斷詞與非詞的標準。

我們不去爭辯相關的標準問題，而以提出我們的工作方案，即「可操作的分詞標準」替代爭辯，同時我們容許也承認各級語法單位之間可以有交集或重疊的事實。

詞有兩個基本特性：「備用性和獨立運用性」。從「備用性」來看，「詞」是儲存在大腦裡的備用零件，是獨立單一、不可再更加細分的個體，隨時可以取用以組成短語或句子等更大一級語法單位。當零件被取用，並以一定合理的關係結

合、分離、移位等時，才會產生「短語或句子」等。「詞」有一個很重要的特性就是「有限性」，但「短語或句子」等則具有「無限性」。「詞」的「有限性」表現在「數量有限、長度有限」這兩個方面，「短語或句子」的「無限性」也表現在「數量無限、長度無限」這兩個方面，這樣便形成了「互斥關係」。

| | 長度有限 | 數量有限 | 長度無限 | 數量無限 |
|----|------|------|------|------|
| 詞 | + | + | - | - |
| 短語 | - | - | + | + |
| 句子 | - | - | + | + |

「詞」做為零件備用，這便呈現出詞的「獨立運用性」。我們可以根據「有限或無限」這兩個互斥的特點，做為「詞和短語或句子」之間的區別性特徵，將備用性的「詞」分離出來，進而做為建置語料庫時進行語料加工的詞性標注工作的零件。

前面說「我們容許也承認各級語法單位之間可以有交集或重疊的事實。」，在我們的工作方案中對這類問題以特別定義的方式來解決，例如客家語：

第一、五點七二、一百五、百五、千七、萬六、十過、百過、七十零、百零、萬零

這些結構是詞還是短語？就可以用「有限與無限」這兩個標準來檢驗，很清楚的是他們在數量和長度上都是無限的，所以他們是「短語」，也就是說，一個「數」可以用一個數字詞表示，如「一、二、三」，也可以用一個短語的形式表示，如「第一、五點七二、一百五」，客家語還可以省去位數詞前面的數字詞，成為「位數詞+係數詞」的結構，如「百五、千七、萬六」，但這種情形只限超過

一百一十以上的數，且「位數詞+係數詞」的結構中的「係數詞」，必定限為「十」的倍數才可省略位數詞前的數詞，也就是說，如「一百二十一，一千七百二十三」這種結構不能省略成爲「百二十一、千七百二十三」；「十過、百過、七十零、百零、萬零」，這種結構也是客家語數詞的一種表示法，這些都是「短語」的形式，內部組成成分其實都是「備用零件」，組成的結構體是短語，關於這種短語結構的切分和標記方式請參考分詞規範第九點。

這裡所要說明的是本語料庫的標記基本上是針對「詞」這個單位級進行標記，但在標記如上各例時，依據「有限與無限」的標準，雖應爲「短語」型態，但筆者容許這種模糊性，將其視爲一個切分單位來進行標記，不予做切分，以免分割得支離破碎。

第二節 確認詞的方法

影響分詞的因素很多，而這些因素都可以成爲分詞的依據和標準。呂叔湘說，判定一個語素組合體是詞還是短語，主要涉及五個因素：「第一、這個組合的成分能不能單用；第二、這個組合能不能拆開，也就是這個組合的成分能不能變換位置或讓別的語素隔開；第三、這個組合的成分能不能擴展；第四、這個組合的意義是不是等於他的成分的意義的總和；第五、這個組合包含多少個語素，也就是它有多長」¹。根據呂叔湘的意見我們採用（一）單用及語義標準（二）擴展標準（三）音節標準（四）頻率標準（五）互信息標準，這五項來辨認一個字串是否爲一個詞，在逐步對文本分詞過程中修正我們的分詞規範。

（一） 單用及語義標準

¹ 呂叔湘，漢語語法分析

一般語法書把「單用」和「語義」標準分開為兩個標準，我們認為這是必須結合的一個標準。如果只從「單用」與否的標準來看，顯然「草鞋」中的「草」和「鞋」這兩個成分都是可以單用的，但是「草鞋」既非「草」，也非「鞋」，更多例子如客家話的「~仔、~板、~盒、~儕」都是能產性很強的結構體，如果把「單用」和「語義」標準分開就會被切割得支離破碎，所以我們把這兩個標準合起來考量，稱為「單用及語義標準」。

所謂「單用及語義標準」，是指一個結構體，不論它包含多少語素，如 AB、ABC、ABCD 只要其中的任一語素在結構體中被獨立出來時，會破壞該結構體的整體意義者，我們就視為不能單用，則該結構體就是一個詞。如：

~仔：客家話有許多的「仔」尾，能產性非常強。

A：賴仔、妹仔、凳仔、兵仔、細人仔、細猴仔、閒閒仔、石跳仔、阿叔仔、笑笑仔、三月仔、笑咪咪仔、臨暗仔、豆仔

B：烏豆仔、大屋仔、長豆仔、大賴仔、長凳仔、

~板：客家話有許多的「板」尾詞，能產性非常強。

C：水板、紅板、菜頭板、甜板、發板、艾板、九層板、糯米板

D：炊板、做板、打板、cip²⁰板、揆板、切板、爇板、煎板、

以上 A、B、C、D 四組是詞還是短語，一般而言 A、C 組是詞，不太有爭議，但是 B、D 組就有爭議，有的字典或研究者把「大賴仔、長凳仔、炊板、爇板」等列為詞，有的不認同，如果從「單用標準來看」，B 組是短語，如「大賴仔」這個結構體的組成份，可以再拆成表序數、修飾功能的「大」和被修飾的名詞成分「賴仔」，這兩個成分都可以被單用，成為組詞或構句的零件，如「大車、大牛、大伯、厥賴仔、降賴仔、做賴仔」，但是像 D 組，一般詞典都把它們列為「詞」，從結構上看，組成成分之間是可以被拆開的，而且都是「動詞+名詞」的述賓結構，但是同樣是述賓結構的 E 組，一般人都不會將它們列為詞，會認為是短語。

E：爌菜頭粿、煎甜粿、炊發粿、炒豬肉、做紅豆粿、按粿圓

像 E 組這種原結構體都是能被拆開成部分組件，而且部份組件都是具有組詞或構句能力的成分的，這時候我們引入「音節標準」，考慮詞彙的雙音化傾向，將「音節標準」做為「單用標準」的輔助標準，規定凡是雙音節的結構體，視為一個詞，其他則予以切分。

語義標準方面我們採取：「語義無法由表層組成分直接相加而得到之字串應該合為一分詞單位。」的理念做為分詞的標準，例：下列客家語字串，語素組合後語義改變，故應視為一個分詞單位。

- a. 風水（是「墳墓」，不是「風和水」。）
- b. 水毛（是「毛毛雨」，不是「水的毛」。）
- c. 眉毛水（是「毛毛雨」，不是「眉毛上的水」。）
- d. 紅毛泥（是「水泥」，不是「泥上長紅毛」。）
- e. 三十暗晡（是「除夕」，不是「三十個晚上」）
- f. 雞毛掃（是「雞毛撻子」，不是「拿雞毛來掃」，這裡的「掃」，已完全失去動詞義，與「雞毛」結合成「雞毛掃」這個名詞。）
- g. 拿阿等（是「一拿著就~」的意思，不是單純的「拿著」或「拿著等」。）
- h. 阿婆官（是「沒有魄力的閒官」，不是「阿婆當官」。）

（二） 擴展標準

「擴展標準」是指一個結構體是否能被其他成分插入，導致原結構體成分之間分離的情形，凡是能擴展的就依據擴展情形切分，不能擴展的參酌其他標準處理，如：

「洗身、降子、剃頭、落水、食茶」可以擴展為「洗一擺身、降豬子、剃過頭、落三日水、食下茶」當其為擴展結構時，根據擴展結構切分。

但「擴展標準」有其侷限性，對於區別複合詞和短語既不是充分條件也不是必要條件。如：

方位結構：桌頂、背囊項、箱肚、肚屎肚、河壩唇、飯碗肚

本+名詞結構：本人、本國、本報、本莊、本市

滿+名詞結構：滿山、滿圓身、滿街、滿頭林、滿姑、滿女、滿叔

都是不能擴展的結構，而且能產性很強，所以如果我們將凡是不能擴展的就全部訂為詞而收入詞表，顯然是不合理的，這時候我們引入「音節標準」，仍考慮詞彙的雙音化傾向，將「音節標準」做為「擴展標準」的輔助標準，規定凡是雙音節的結構體，視為一個詞，其他則予以切分。

(三) 音節標準

以音節做為分詞標準，實際上是很不可靠，而且無法訂為規則操作的，如果客家語的詞全部是固定的雙音節，自然就沒有分詞的問題，然而並非如所願。音節只是一種語感，有很多的因素會導致語感的不同，所以音節標準並不是一個非常科學性可以操作的分詞標準，只能做為其他標準的輔助標準。例如：行政院客家委員會，讓不同的人來讀，會有不同的音節結構，可以是「行政院 / 客家 / 委員會」，也可以是「行政院 / 客家委員會」，也可以是「行政 / 院 / 客家 / 委員 / 會」。

所以音節標準只是一個參考標準，無法形式化、規則化。我們定義以音節為參考標準時，分別依「雙音節、三音節、四音節、……………」的順序為考量標準，斟酌其他標準來分詞，因此前舉「行政院客家委員會」，以分為「行政院 / 客家 / 委員會」為合理。

(四) 頻率 (frequency) 標準

使用頻率也是影響人們定義一字串是否成詞的因素，比如客家話的「V 毋得」結構，到底是詞還是短語？查閱台灣客家話辭典有「行毋得、除毋得絕、過毋得、講毋得、怪毋得、嫌毋得、看毋得、耐毋得、摸毋得、豺毋得、動毋得、倚恃毋得、做毋得」這些都列入詞，可是為什麼沒有收「聽毋得、問毋得、拿毋得、放毋得、食毋得、吞毋得、打毋得」等，其他字典則一定把「做毋得」列入，顯然這和頻率有關。

「做毋得」是一個高頻字串，表面上和上舉各例是同一個結構，都是「V+毋得」結構，但加以分析之後會發現「行毋得、過毋得、講毋得、怪毋得、嫌毋得、看毋得、耐毋得、摸毋得、豺毋得、動毋得、倚恃毋得、聽毋得、問毋得、拿毋得、放毋得、食毋得、吞毋得、打毋得」這些都可以在「V+毋得」中插其他成分，如：

「行佢毋得、過佢毋得、講佢毋得、怪佢毋得、嫌佢毋得、看佢毋得、耐佢毋得、摸佢毋得、豺佢毋得、動佢毋得、倚恃佢毋得、聽佢毋得、問佢毋得、拿佢毋得、放佢毋得、食佢毋得、吞佢毋得、打佢毋得」

也可以在「V+毋得」之後加入補語或賓語成分，如：

「行毋得過、過毋得去、講毋得出、怪毋得佢、嫌毋得著、看毋得出、耐毋得荷、摸毋得著、豺毋得落、動毋得佢、倚恃毋得佢、聽毋得落耳、問毋得出、拿毋得出、放毋得入、食毋得落、吞毋得落、打毋得死」。

所以如果把全部的「V+毋得」結構全部收入詞表，顯然是不合理的，因此我們規定這種「V+毋得」的結構體，一律切分為「V/ 毋得」。如果是「中插其

他成分」的一律予以分詞，參考其他標準做切分。

而「做毋得」是不能中插成分的，加上使用頻率相當高，所以將其列為詞，相同的還有「使毋得」，「使毋得」這一字串有兩種情形，一種是表示「不可以」，如「你想愛恁樣做，**𠵼**看使毋得」，這種情形是一個詞，不予切分；但如果是指「不得使用或花費的意思」時，則是可以用中插其他成分的，與前舉各例相同，則參考其他標準予以切分。如「厥介錢使毋得。」、「著大獎，恁多錢，仰會使毋得忒？」、「使厥錢毋得，莫 ham⁵³ ts'e⁵³！」

「頻率」在分詞中究竟應該起什麼作用？我們認為「頻率」不應該成為決定一個字串是否成詞的標準，因為使用頻率低，不能就說他是「非詞」，例如「出息」(tj'ut⁵⁰ sit⁵⁰) 意指「發芽」，這個詞在語料庫中只出現一次，在龍潭客家語故事集(一)第四頁第六行：「噯哦，**𠵼**菜又吉好賣咧，種落去項項都吉出息咧。」

根據筆者探訪八位老人家，只有一位九十三歲的老人家知道「出息」還有「發芽」的意思，足見這個詞使用的頻率低，但不能因為它使用頻率低，就認為它是「非詞」，既然「頻率標準」不能成為決定「詞或非詞」的標準，那為什麼還需要列入參考標準呢？

我們的想法是可以把「頻率標準」當作是一個實用性的參考標準，把使用頻率高、搭配強度高的短語也直接列入詞的層次來進行分詞，如客家話的「厥公、厥婆、厥姆、若叔、正來僚、恁仔細」及對「成語」的看法等等，就是基於這種思考。從單用及語義標準、擴展標準的角度來看，這些都應該是短語，但是他們是高頻的固定搭配詞，所以我們把他們列入已詞化了的短語，也就是實際上等同於詞的層次了，不再進行切分。

(五) 互信息標準

有時候我們斟酌再三，仍難決定某一字串是否成詞，於是我們引入「互信息標準」，我們先以 123457 字的語料做為計算互信息的基準，共得到 49784 詞次，

假設某一字串AB，先計算出如果字串AB成詞的詞頻 F_w ，再計算AB的互信息 MI_w 。詞頻 F_w 用以衡量該字串是否是一使用頻繁、搭配度高的字串，互信息值 MI_w 做為判斷該字串是否是一結合緊密量化指標。

我們以三字字串為例，假設該三字字串為 $W=ABC$ ，詞頻的計算是先假設該字串是一個詞的情況下，在 49784 詞次中出現的頻率，比如我們要決定「拿毋得」算不算一個詞？我們發現「拿毋得」在 49784 詞次裡出現 2 次，則其頻率估計值 F_w 是 $2 \div 49784 = 0.004\%$ ；接著再考慮互信息值 MI_w ，互信息值由下式定義：

$$P(ABC)$$

$$MI_w = \log \frac{P(ABC)}{P(A) \times P(B) \times P(C)}$$

例如「拿毋得」，「拿」出現 23 次、「毋」出現 187 次、「得」出現 132 次，而「拿毋得」同時出現 2 次，「毋得」同時出現 14 次，所以我們便可以透過計算得到下列數據：

$$F_{\text{拿}} = 0.046\% \quad , \quad F_{\text{毋得}} = 0.028\% \quad , \quad F_{\text{拿毋得}} = 0.004\%$$

$$MI_{\text{毋得}} = -2.9774 \quad MI_{\text{拿毋得}} = -5.4531$$

上面的數據告訴我們，「拿」和「毋得」單獨被使用的頻率，遠高於「拿毋得」三者共現使用的頻率，「拿」和「毋得」的搭配度低，互信息值來看也是「毋得」高於「拿毋得」，也是「拿」和「毋得」的搭配度低。

根據這些明確的數據結果，我們再參考其他標準，做為是否成詞的判斷依據。這樣以計算量化為依據的方式需要有更大量的語料做為基礎，會更札實，但是目前我們沒辦法處理或取得更大量的語料來進行分析。

根據以上的分析我們採取（一）單用及語義標準（二）擴展標準（三）音

節標準（四）頻率標準（五）互信息標準，這五項做為分詞的基本原則。有了這些原則和方法之後，我們便可以建立一套分詞規範並具體落實到文本的分析，進行分詞工作。

另外加以說明的是「成語」不切分，但是「諺語、老古人言」等我們根據分詞規範予以切分，這是因為我們認為「成語」是已詞化了的單位，成語的組成份子已成為固定且高頻的搭配關係，所以不予切分；而諺語或老古人言，往往是一個表達含意或概念完整的句子，所以我們予以切分。

例如：「伯公翻崩崗」這是一句歇後語，並不是真的神明「伯公」翻落「崩崗」下，而是另有所指，意在言外表示「贊成」（戰神的諧音）的意思，但是我們看「伯公翻崩崗」，這是一個「主語、謂語、賓語」結構完整的字串，與句子無別，況且每一個組成成分都是可以再進入組詞構句的單位，所以我們視為句子予以切分。



國立新竹教育大學
National Hsinchu University of Education

第三節 各詞類的切分具體說明

根據上面的分析和我們的工作經驗，下面具體提出我們如何來進行分詞的具體操作。以下分別依據動詞、形容詞、狀態詞、名詞、量詞、方位詞、時間詞、處所詞、區別詞、數詞、副詞、擬聲詞、介詞、連詞、語氣詞、助詞、嘆詞、代詞等及成語、俗語、師父話等各方面具體舉例說明其切分原則。

（一）、動詞：

（1）**動詞的重疊**：單音節動詞重疊不予切分，雙音動詞重疊則予以切分。

1、AA：切分為 AA

- 2、AA 啊、AA 啊哩：切分為 AA/ 啊、AA/ 啊哩
- 3、A 啊 A、A 啊 A 仔：視為一個切分單位。
- 4、由一個單音節語素並列重疊者全數切分。
- 5、由雙音節動詞，並列重疊者，予以切分。
- 6、由二個單音節語素重疊的格式有 AABB、ABAB 格式，亦視為一個切分單位，不予切分。

示例：

AA：洗洗、收收、挖挖

AA 啊：寫寫/啊、咬咬/啊、剝剝/啊

AA 啊哩：食食/啊哩、割割/啊哩

A 啊 A：行啊行、食啊食

A 啊 A 仔：跳啊跳仔、行啊行仔

AAAA：掙掙/掙掙、摸/摸/摸

AABB：來來去去、出出入入、行行坐坐

ABAB：參詳/參詳、詐意/詐意、

(2) 動詞的否定：動詞的否定格式一律單獨切分。

示例：

無：無/拿、無/食、無/咬

毋：毋/拿、毋/食、毋/咬

莫：莫/拿、莫/食、莫/咬

A 毋 A：食/毋/食、信/毋/信、

A 啊毋 A：行/啊/毋/行、去/啊/毋/去

唯一例外處理的是「係毋係」，這是客家話裡構成正反問句的標記，我們將它視為一個已固化了的短語，當成詞的單位來處理，不再切分。

(3) **AD+V**：這個結構是指並列結構的雙音節詞中，前一語素修飾或說明後一語素的動詞，我們視前一語素詞性如副詞，這種結構為雙音節時，視為雙音節動詞，不予切分，其他則切分。

示例：

亂講、亂食、亂衝、亂打、亂走、亂使、亂剝、亂拂
去割、去喊、去尋、去換、去買、去量、去探、去問
去/上班、去/打票

(4) **V+N**：對述賓結構的 V 和 N 之間：

- 1、如果 V-N 成詞，而 V-N 之間有中插成份隔開 V 和 N，則切分。
- 2、如果 V-N 之間是沒有中插成分的離合詞，則視 V-N 為一個分詞單位。

示例：

中插：洗 / 下 / 身 / 來、食 / 碗 / 滷肉飯、

V-N 緊密：關心、開會、讀書、出血、行棋、撩人、捉人

(5) **V+R**：對動補結構的 V 和 R 分合定義如下：

- 1、如果 V 和 R 都是單音節，且 V-R 之間沒有中插其他成分隔開的情形，則合併 V-R 為一個切分單位。
- 2、如果 V 和 R 其中之一是雙音節者，且 V-R 之間沒有中插其他成分隔開的情形，則切分 V 和 R。
- 3、如果 V-R 成詞，而 V 和 R 之間有中插成分，則予以切分。
- 4、如果 V-R 不成詞，而 V 和 R 之間有中插成分，則予以合併。

示例：

V-R 一個分詞單位：食飽、掙淨、淳出、趨走、拿來、落去

V-R 二個分詞單位：走/過來、拿/出來、咬/核核、弓/弦弦

V-R 成詞有中插：食/咁/飽、趨/咁/走、拿/得/出來、食/得/落

V-R 不成詞：買得起、買咁起、堵毋著、

(6) **V+AX**：對動詞後粘著的時態標記「起來、了、等、過、忒、去了、去、過了、忒了」等，一律切分。

示例：

走/了、食飽/了、行/等/過、坐/等/了、食/等/飯、行/過

(7) **V+TO**：對動詞後粘著的「著 (to11)、分、爲、成、向、入」等，切分方式如下：

示例：

著：食/著/盡/飽、行/著/遠遠、反對/著/無葛無煞。

分：送分、拿分、交分、踢分、退轉/分、拿出/分

爲：做爲、成爲、認爲

成：變成、促成、構成、養成、落成、釀成、完成、收成

向：面向、轉向、偏向、換向

入：敲入、撞入、侵入、收入、放入、

(8) **並列結構**：對並列結構的單音節動詞，視爲一個切分單位。

建立、宣傳、出入、行行

(二)、形容詞：

(1) **形容詞的重疊式**：形容詞的重疊式 AA、AAA、AABB、ABB、ABAB，一律爲分詞單位，不予切分；重疊式加「仔」尾時，視「仔」尾爲一粘著語素，故與詞根合併爲一個分詞單位。

示例：

- 1、AA：細細、綿綿、皺皺、高高、幼幼
- 2、AAA：烏烏烏、輕輕輕、硬硬硬
- 3、AABB：正正經經、偷偷掠掠、隨隨便便、厝厝不不
- 4、ABB：硬确确、肥鼓鼓、紅噹噹、肚挺挺、肚蠟蠟
- 5、ABAB：熟面熟面、熟絡熟絡、
- 6、~+仔：硬确确仔、肥鼓鼓仔、牙豺豺仔、嘴扁扁仔

(2) **雙音節形容詞**：雙音節形容詞，不論其語素之詞性，一律合併為一個分詞單位。

沙鼻、高毛、癡哥、花頭、衰過、孤老

(三) **狀態詞**：狀態詞不切分。

示例：

恬恬靜靜、大方、白雪雪、雪白、清清楚楚、笑眯眯

(四)、**名詞**：

(1) **親屬關係詞**：親屬關係詞根據客家語的音節結構和切分點前後組成部分之是否能單用，決定其分合。

1、「阿+N」的切分：「阿」一律前置，視為一個切分單位，不予切分。

示例：

阿太、阿公、阿婆、阿叔、阿姨、阿舅、阿伯、阿憨伯、阿田哥、阿青叔

2、「大、細、孫、外、厝、表、堂、侄+N」的切分：

「大、細、孫、外、滿、表、堂」等一律前置，如果是「1+1」的單音節語素所構成的親屬詞，視為一個切分單位，不予切分；如果是「1+2」的語素所構成的親屬詞，則予以切分。這種「1+2」結構因組成部分的N，更常用於口語，是一個完整的單位，根據詞的獨立運用性原則，這種結構予以切分。

示例：

大叔、細舅、滿姨、堂叔、孫女、大孀、外甥、大/賴仔、大/妹仔、外/新婦、細/新婦、表/姊妹、侄/孫女、堂/兄弟

3、「叔伯+N」的親屬詞切分：這種結構因組成部分的N，更常用於口語，是一個完整的單位，根據詞的獨立運用性原則，這種結構予以切分。

示例：

叔伯 兄弟、叔伯 姊妹、叔伯 阿哥、叔伯 老弟

4、「CD+N」的親屬詞切分：這種結構是「數詞+N」的結構：

如果是「1+1」的單音節語素所構成的親屬詞，視為一個切分單位，不予切分；如果是「1+2」的語素所構成的親屬詞，則予以切分。這種「1+2」結構因組成部分的N，更常用於口語，是一個完整的單位，根據詞的獨立運用性原則，這種結構予以切分。

示例：

二叔、三舅、二伯、三姨、二貨、三姊、二/姑丈、三/舅姆、二/伯婆、二/姨婆、三/姑婆

5、「PN+N」的親屬詞切分：這種結構是「代詞+N」的親屬詞結構，只有在「吾、若、厥、佢」與單音親屬詞並用時會產生，就語義上說「吾、若、厥、佢」是「你介~、佢介~」的省略說法，由「吾、若、厥、佢+N」所組成的親屬詞，平時也可以單獨運用，根據詞的獨立運用性原則，這種結構

不予以切分。但如果是「1+2」的語素所構成的親屬詞，則予以切分。

示例：

若爸、若姐、若姊、若嫂、厥舅、厥哥、厥公、**偃公、偃爸、偃嫂** 若
/老弟、厥/老妹、**偃/老弟**

6、「N+仔」的親屬詞切分：這種結構的親屬詞，就語意上說具有特指或貶低的語義，這種結構是一個完整的分詞單位，不予以切分。

示例：

賴仔、妹仔、叔仔、舅仔、阿叔仔、

7、「其他」親屬詞的切分：其他親屬詞一律不切分。

示例：

姐公、姐婆、老姑婆、丈人老、丈人哀

(2)「CD+M+N」：「數詞+量詞+名詞」的結構，一律切分。

示例：

一 / 儕 / 人、三 / 枚 / 針、二 / 枝 / 斧頭、一 / 雙 / 鞋、斤 / 二

(3)「DT+N」：「定詞+名詞」結構，如果「DT和N」都是單音節，則視為一個分詞單位，如果不都是單音節，則予以切分。

示例：

本人、全球、本/單位、歸/屋下、歸/間肚

(4)「PN+N」：這種結構是「人稱代詞+N」的結構，客家話「PN+N」的結構，除前述親屬詞部分有1+1的單音節結構外，其他都是「1+2或1+3」的結構，一律切分。另一種「人稱代詞+N」的結構，是「若、厥、偃」等這三個「人稱代

詞+N」的情形，「若、厥」會變讀成陰平調²，根據詞義的分合，予以切分。

示例：

偈/學校、厥/老屋、偈/學生仔、厥/筆、若/車

(5)「名詞標誌」的切分：許多語法書著作，都認為名詞標誌不外乎「前綴、後綴」及一些「類前綴、類後綴」之類，並以「虛與不虛」或「全虛與半虛」或「結合緊密、使用穩定與否為檢定標準」，其實這些都是難以判定的標準，無法成為操作的準則。

這裡我們跳脫詞綴的概念，不以全虛或半虛，附著或不附著為分合的標準，而以語塊（chunk）的概念來定義。單音節的標誌，一律合併不予切分，雙音節以上的標誌成分則予以切分。

示例：

化：專業化、工業化、社會化、

非：非賣品、非正式、

老：老虎、老猴、老牛、老阿伯、老伯公、老人家、老貨仔

環：牛環、杓環、笠環、鑊環、刀環、板環、福佬環、

阿：阿沙布魯、阿里不答、阿公婆、阿婆官、阿啾箭、阿財伯

室：辦公室、工程師室、電腦室、資訊室

局：教育局、財政局、公賣局、警察局

仔：桌仔、凳仔、鳥仔、星仔、妹仔、賴仔

頭：日頭、芋頭、拳頭、肩頭、擔頭

公：雷公、蟻公、蝦公、石頭公、伯公

牯：牛牯、槌牯、石牯、屎牯

局長：教育局 局長；公賣 局長

² 關於這部份詳細的說明，請參考下文。

- (6) 「**形容詞語素+名詞語素**」：這種結構是單音節的「形容詞語素」+單音節的「名詞語素」的結構，合成詞的詞義並不是由「形容詞語素和名詞語素」相加而得者，則 AN 合為一個分詞單位，其他則予以切分為兩個詞。

示例：

一個分詞單位：鮮花、老屋、紅花、硬頸、靚肉

二個分詞單位：肥/肉、矮/屋

- (7) **N1+N2**：這種結構是指 N1 修飾 N2 的情形，N1 可能是一個或兩個以上的語素，而 N2 則是一個語素的情形時，N2 視為接尾詞，合併 N1+N2 為一個分詞單位，如果 N2 是二個以上語素，且不能獨立運用，則一樣視 N2 為接尾詞，合併 N1+N2 為一個分詞單位，如果二個以上語素的 N2 可以獨立運用，則予以切分為兩個詞。

示例：

一個分詞單位：台北市、研究生、語言所、法律系、物理學家、

二個分詞單位：電腦/工廠、台灣/大學

- (8) **區別詞+名詞**：這種結構是指「區別詞+名詞」的結構，區別詞可能是單音節或雙音節或是一個短語。如果「區別詞+名詞」都是單音節，則「區別詞+名詞」合併為一個切分單位；如果是雙音節，則予以切分為兩個詞。

示例：

一個分詞單位：婦人家、男仔人、前人子、後哀、

二個分詞單位：共同/利益，老牌/電腦、前/總統

- (9) **簡稱**：簡稱詞一律合併為一個分詞單位。

示例：

一個分詞單位：中小學、國中小、

(10) 人名：「人名」，姓和名，一律合併為一個分詞單位，姓氏併入人名專有名詞，不另外增設姓氏標記切分；帶詞綴的人名，一律合併為一個分詞單位；帶職稱或頭銜的人名則予以切分。

示例：

一個分詞單位：張庭妹、謝英財、卡爾馬克思、老王³

二個分詞單位：羅/教授、古/教授、張/局長、羅/古/兩/位教授

(11) 國名、組織名、鄉鎮名、學校名、、、、：「國名、組織名、鄉鎮名、學校名等」，如果是 N1+N2 的結構，而 N1 不論其音節數多寡，只要 N2 是單音節時，則一律合併 N1+N2 為一個分詞單位；如果 N2 是雙音節以上時，則按 N1+N2 的組成是否能獨立運用的標準來切分，不能獨立運用則合併，能獨立運用則切分。但是如「市政府」這種結構，依能否「獨立運用」的標準，本應予切分為「市/政府」，然而我們不這樣來認定，我們認為這種結構應歸入帶接頭詞的專有名詞結構看待，予以合併為一個切分單位。

示例：

一個分詞單位：台灣省、台北市、吉安鄉、宜昌村、阿美族、
長江、吉祥四街、三號、師院路、頭前溪

二個分詞單位：新竹/師院、台北市/政府、中國/國民黨、民主/
進步黨、基隆河、淡水河

視為接頭詞的一個切分單位：市政府、鄉公所、鎮公所、秘書室

³ 像「老王」這種「詞綴+姓氏」的結構，傳統客家話不使用，但已逐漸有人使用中，我們還是列入分詞規範來定義。

(12) **專名簡稱**：專名的簡稱詞一律為分詞單位。

示例：

一個分詞單位：中/美、台/日、蘇/花、高/屏、花/東

(五)、**量詞**：量詞，均視為一個分詞單位。

示例：

一個分詞單位：一/ 斤/ 肉、一/ 間/ 牛欄、三/ 頂/ 眠床

(六)、**方位詞**：X1+X2 的結構裡，如果 X1、X2 都是單音節語素，則合併 X1+X2 為一個切分單位；並列結構的方位詞則給予合併為一個切分單位；其他則予以切分。

示例：

一個分詞單位 (X1+X2)：頭前、旁唇、以東、之內、國內

一個分詞單位 (並列結構)：東西南北

二個分詞單位：西方 之前、左片/析

(七)、**時間詞**：從結構上看，客家話的時間詞結構有：

(1)「數詞+時間詞」：如果是計數時則切分，非計數時則予以合併。

(2)「數詞+ (量詞) +時間詞」：這種結構則一律切分。

(3)「前、後、上、下、大前、大後」+時間詞：採取合併的方式。

(4)「初+十以內的數字」：採取合併的方式。

(5) 一週七天均合併為一個分詞單位。

(6) 其他時間詞一律不切分。

示例：

「數詞+時間詞」：1998年/三月/二十三⁴/五點/三分/七秒

「數詞+量詞+時間詞」：三/只/月

「前、後、大前、大後」等+時間詞：前日、後年、大前年、大後年

「上、下」+時間詞：上禮拜、下禮拜、上只禮拜、

「初+十以內的數字」：初一、初二、初三

一週七天：拜一、拜二、拜三、拜四

其他時間詞：朝晨、下晝、暗夜、當今、正月、立冬、春天、後日

(八)、**處所詞**：如果是 X1+X2，其中 X1 和 X2 都是單音節時，則合併為一個切分單位，其他則予以切分。

示例：

一個分詞單位：天頂、室內、路頂、街項

二個分詞單位：左片 肩頭、田頭 田尾、田頭 田角

(九)、**區別詞**：區別詞，均視為一個分詞單位。

示例：

現任/總統、深色、穀/胖心、高等、慢性/難醫了

(十)、**數詞**：客家語的數詞從結構和功能上來看可以分為四類：

(1) 係數詞：0、一、二、三、四、五、壹、貳、叁、肆、伍、陸

(2) 位數詞：個、拾、百、千、萬、拾萬、百萬、千萬、億、兆、

(3) 限定數詞：盡多、好多、益多、幾多、幾、大多數、大部分。

(4) 助數詞：分之、點、百分之、零、上下、空、過、

⁴ 客家話通常不把「數詞+日」中的「日」說出來，可能會被誤認為是數字詞，這是被允許的錯誤。

(5) 數量詞：百年、百日、半日、一生、一點、一下、一截、一埕

(4) 序數詞：第一、第二、第三、第五十、初一、初二、初三

數詞的組合，應看成是短語，切分原則如下。

(1) 如果 X_1+X_2 的結構裡， X_1 是 0 到無限大的係數詞， X_2 是位數詞或概數詞，而且 X_1+X_2 的結構合理存在時，則合併 X_1+X_2 為一個切分單位。

(2) 如果 X_1+X_2 的結構裡， X_1 是「非 0 到無限大的係數詞」， X_2 是位數詞或數量詞等，而且 X_1+X_2 的結構合理存在時，則切分 X_1+X_2 為二個分詞單位。

(3) 如果 X_1+X_2 的結構裡， X_1 是概數詞， X_2 是係數詞或位數詞，而且 X_1+X_2 的結構合理存在時，則切分 X_1+X_2 為二個分詞單位。

(4) 如果 $X_1+X_2+X_3+\dots+X_n$ 的結構裡， X_1 、 X_2 、 X_3 、 X_n 可能是係數詞、位數詞、限定數詞、數量詞，交互結合而成 $X_1+X_2+X_3+\dots+X_n$ ，而且 $X_1+X_2+X_3+\dots+X_n$ 的結構合理存在時一律切分。

(5) 序數詞不予切分。

(6) 省略說法不切分

(7) 助數詞的切分：

a. 「分之、點、百分之」所構成的結構，切分。

b. 「零、上下、空、過、」所構成的結構，不切分。

(8) 數量詞的切分：

c. 「數量詞」的結構，如果是計數時則切分，是語義決定。

d. 「數量詞」的結構，如果非計數則不切分。

示例：

一個分詞單位：九百、五千萬、三百年、三千上下、1300 萬、五百零、
叁拾零萬、50.1、50%、九百五十三萬

二個分詞單位：差𠵼多/五千萬、大約/三千上下、

多個分詞單位：三 分 之 一、四/點/五、百 分 之 五 十

序數詞：第一、第二、差𠵼多/第三、大約/第五十

省略說法：十過、百過、千過萬、三千零、

數量結構（計數）：三 百 加 一 百、五 垓 減 一 垓

數量結構（非計數）：加 兩 禮 拜 阿 爸 做 百 日；一 生 人 去 美 國 一
擺；一 垓 一 垓 放 好 來。

(十一)、**副詞**：副詞一律為分詞單位，而使用穩定，起副詞作用的短語，不予切分。

示例：

當 好、盡 遠、忒 肥、一 等 慶、共 下 去、常 透 食

(十二)、**擬聲詞**：擬聲詞一律為分詞單位，不予切分。

示例：


𠵼𠵼𠵼 一 聲、*𠵼𠵼𠵼 一 聲、𠵼𠵼𠵼 𠵼𠵼𠵼 𠵼𠵼𠵼、
𠵼𠵼𠵼 𠵼𠵼𠵼 𠵼𠵼𠵼

(十三)、**介詞**：介詞不切分。

示例：

照 你 講 介 來 做。分 人 看 著。佢 比 你 高。爲 著 十 萬 元

(十四)、**聯接詞**：聯接詞不切分。

示例：

因爲 你 艱苦，故所 愛 食 藥仔。你 既然 講 了，斯 愛 做到。

佢 噢 你。鄉村 或者 都市。再過講 你身體無好。

(十五)、**語氣詞**：語氣詞不切分。

示例：

你田去啊？、樣會無早講呢？落水了。敢講算了係無？

紅色也好，青色也好。

(十六)、**助詞**：助詞不切分。

示例：

像牛樣仔？、外背落等水哦！你介。食 過了？

對佢來講，無爭差到。

(十七)、**嘆詞**：嘆詞不切分。

示例：

ha24，你樣會講出喺？

oi53，有人在無？

o11，原來係佢。

(十八)、**代詞**：代詞不切分，但反身代詞則切分。

示例：

佢、佢 自家、佢 兜

這位 有花。仰般 做好？

(十九)、**成語**：成語不切分。

示例：

包山塞海、伯公生鬚、巴掌生毛、嫌七嫌八、一家一務、寅食卯糧。

(二十)、**俗語**：俗語依分詞規範進行切分。

示例：

鴨嫲 絡食 毋 知 坑窮。

斧頭 削 自家 个 柄 毋著。

不孝 新婦 三餐 燒 ， 有孝 妹仔 路上 遙。

(二十一)、**師父話**：師父話依分詞規範進行切分。

示例：

阿婆 嫁 孫女

巴掌 生毛

花鉢仔 種菜

屨水 上天



第五章 客家語詞類劃分與標記說明

就語言研究和語言工程研究來說，生語料庫是一個語言資料的集合罷了，侷限性是不言而喻的。例如客家語：「踢爛球仔」，可能是指「踢 爛球仔」，也可能是指「踢爛 球仔」，究竟是哪一個的訊息才正確，這些訊息集中在人們表達、交流、傳播的知識上面，對這種意義的理解，以使用該母語者來說是沒有理解上的阻礙的，但是對計算機處理來說卻是困難重重，不易實現。

當然生語料庫在實際應用中所造成的麻煩還不只這方面，例如英文 fly 是動詞（飛翔）還是名詞（蒼蠅），沒有適當的語言知識告訴計算機，計算機是無法處理的。基於提高機器學習和重複處理的計算機處理能力，人們透過各種方法來指導機器學習，其中最主要的手段就是對文本進行語言信息標注，使之成為具有豐富信息的熟語料庫。所謂熟語料庫就是給原始文本添加各類信息的語料庫，添加信息的過程稱為標注，所添加的符號稱為標記。其中以添加詞法、句法、語意、語用信息最為常見。

對語料的標注分析訊息，都是人們理解話語的底層知識或背景知識，如語法規則、語境蘊含、語用概念等等，如果所標注的範例類型夠齊全、數量夠多，則愈有利於機器的模仿和自動分析，在語料標記的過程中，一般都需要建立一定的標注符號，這些符號要能有效地表示各語言單位及語言單位之間的關係，還具有方便的可操作性。

對於客家語的標注，我們現在研究的水平和經驗都非常少，現階段我們採取一邊研究一邊分析，一邊逐一手工標記的方式來進行，期望將來有一天可以進入自動標記的里程。

第一節 詞類的劃分與標記方法

(一) 詞類的劃分

爲詞標記詞類，其實是標記「詞性」，相同或詞性近似的詞，歸爲一類，便是「詞類」，詞性標注和歸詞類是語言工程重要基礎。標記詞類首先遇到的第一個困難是「什麼是詞？」的問題，我們已在上一章提出現階段操作的分詞規範，做爲工作的準則。但是分出詞之後又有更進一步的困難等待解決，那便是詞類的判定，我們先根據一般客家語及漢語語法書的看法，制訂一份較爲簡易的客家語詞類體系表，並定義了相關詞類的標記符號，一邊工作，一邊進行修正，以下是我們給本語料庫使用的詞類體系表和標注集。

我們劃分客家語的詞計十八個基本詞類，分別爲：名詞、時間詞、方位詞、處所詞、動詞、形容詞、副詞、數詞、介詞、感嘆詞、連接詞、擬聲詞、感嘆詞、助詞、區別詞、量詞、語氣詞、狀態詞。

而基於標記的需要，對「名詞」劃分出「專有名詞」；對「數詞」，劃分出「序數詞」。

對「系動詞」單獨給予標記「/vs」，如「係、等於、姓、安到、變成、做₃、當、當做、算、像、像係、稱呼」等，因爲「系動詞」本身不表示動作、行爲或具體某種狀態，在句法上只有關聯主語和賓語的作用，主賓語之間是等同關係，如「佢係屋下介單登子。」，「你當佢係細人仔，騙人毋識。」，「佢認佢做賴仔。」。

對「領格標記」單獨給予標記「kai」，領格標記「介」，相當於華語的「的」；對「標點符號」不再細加分類，全部標示爲「/pu」。

對「外來語」我們依據其句法功能增加「_fw」的方式標注，如「照相機」的客家話一般口語說「k'a33 me55 la33」，是源自日語「カメラ」的音譯詞，因爲「k'a33 me55 la33」符合「名詞」的語法功能，所以我們標記爲

「k'a33 me55 la33/mn_fw」。

對「成語」我們也採取與外來語相同的處理方式，增加「_id」為標記代號，如「閒神野鬼/jj_id」，表示「閒神野鬼」是形容性的成語。

以下是本語料庫使用的標記符號代號表：

| 序號 | 詞性名稱 | 標記 | 序號 | 詞性名稱 | 標記 |
|----|------|-----|----|-------------|-----|
| 1 | 名詞 | nn | 14 | 副詞 | rb |
| 2 | 專有名詞 | nr | 15 | 介詞 | pa |
| 3 | 時間詞 | nt | 16 | 連接詞 | cc |
| 4 | 方位詞 | lc | 17 | 助詞 | ax |
| 5 | 處所詞 | si | 18 | 語氣詞 | sf |
| 6 | 數詞 | cd | 19 | 擬聲詞 | on |
| 7 | 序數詞 | od | 20 | 感嘆詞 | ij |
| 8 | 量詞 | m | 21 | 外來語 | _fw |
| 9 | 代詞 | pn | 22 | 成語 | _id |
| 10 | 動詞 | vv | 23 | 其他 1---係 | vs |
| 11 | 形容詞 | jj | 24 | 其他 2---介 | kai |
| 12 | 狀態詞 | zht | 25 | 其他 4---標點符號 | pu |
| 13 | 區別詞 | dis | | | |

(表四，本語料庫使用的標記符號表)

(二) 詞類標記的方法與示例：

由於客家語的文本是採取漢字連寫的，詞與詞之間並沒有分隔符號，所以我們進行語料的標記，實際上就是為詞與詞之間增加分隔符號，我們

採取在一個詞的後邊帶上「/和詞類標記」的方式，同時為詞與詞之間帶上「分隔符號與詞類標記。」，以下分就各類詞的標記方式進行示例，我們用表列的方式說明。

| 序號 | 詞性名稱 | 標記 | |
|----|------|-----|--|
| 1 | 名詞 | nn | 梨/ nn、事情/ nn、日仔/ nn、河壩/ nn、精神/ nn |
| 2 | 專有名詞 | nr | 宗教名：基督教/nr、印度教/nr、佛教/nr 國名：米國/nr、日本/nr 山名：中央山脈/nr、大雪山/nr 組織名：教育部/nr、客委會/nr 地名：新竹/nr、台北/nr 人名：媽祖婆/nr、關公/nr、觀音/nr |
| 3 | 時間詞 | nt | 朝晨頭/nt、今晡日/nt、清朝/nt、1992 年/nt、花期/nt、驚蟄/nt |
| 4 | 方位詞 | lc | 肚/ lc、頂高/ lc、正手片/ lc、以下/ lc |
| 5 | 處所詞 | si | 天頂/ si、庄下/ si、田肚項/ si、國內外/ si |
| 6 | 數詞 | cd | 一/ cd、二十/cd、七十零/ cd、百五/ cd |
| 7 | 序數詞 | od | 第一/ od、初一/od、/ cd、初二/od |
| 8 | 量詞 | m | 日/m、粒/m、歲/m、儕/m |
| 9 | 代詞 | pn | 佢/pn、你/pn、佢/pn、佢兜/m |
| 10 | 動詞 | vv | 打/vv、走/vv、行/vv、做大水/vv |
| 11 | 形容詞 | jj | 靚/jj、軟/jj、細/jj、煞猛/jj |
| 12 | 狀態詞 | zht | 氣噴噴仔/ zht、燒爌爌仔/ zht、 |
| 13 | 區別詞 | dis | 高級/dis、一流/dis、烏刻刻仔/dis、迪迪去去/dis |
| 14 | 副詞 | rb | 無/rb、就/rb、堵堵/rb、盡/rb |

| | | | |
|----|-----------------|-----|-----------------------------------|
| 15 | 介詞 | pa | 在/pa、同/pa、爲/pa、從/pa |
| 16 | 連接詞 | cc | 因爭/cc、故所/cc、下後/cc、搵/cc |
| 17 | 助詞 | ax | 來講/ax、到/ax、了/ax、忒/ax |
| 18 | 語氣詞 | sf | 啊/sf、no/sf、無/sf、呢/sf |
| 19 | 擬聲詞 | on | 哈哈滾/on、叮叮噹噹/on、咚咚咚咚/on、k'iaŋ55/on |
| 20 | 感嘆詞 | ij | 啊/ij、哦/ij、唉/ij、唉呀/ij |
| 21 | 外來語 | _fw | k'a33 me55 la33/nn_fw |
| 22 | 成語 | _id | 半出半入/zht_id、變鬼變怪/vv_id |
| 23 | 其他 1--- 係 | vs | 係/vs、係毋係/vs |
| 24 | 其他 2--- 介 | kai | 買/vv 介/kai、桌仔/nn 介/kai |
| 25 | 其他 4--- 標點符號 | pu | ，/pu。/pu；/pu、/pu：/pu！/pu？/pu |

國立新竹教育大學 (表五, 各類詞的標記範例表)
National Hsinchu University of Education

第二節 詞類鑑別的工作守則

畫分詞類的本質依據是詞的語法功能，爲對數以萬計的詞進行歸類工作，便要給出每一種詞類的鑑別準則，也就是要考察各個詞類的語法功能。而在進行將具體的一個個的詞劃歸某個詞類的實際操作時，我們採取的是大致的分類，強調抓住主要的語法特點即可，而不是滿足全部的語法功能，我們給出每一種詞類的鑑別的工作守則，做爲語料加工的操作依據。

(一)、動詞：

- (1)「〈盡〉V 賓」：指一個動詞可以帶上賓語而言，以能不能帶賓語將動詞分爲及物動詞和不及物動詞，而可以受副詞修飾後還能再帶賓語也列入動詞，如「恁希望你來、當想佢會遠遠走」，所以用規則「〈盡〉V 賓」表示。「盡」只是代表副詞類，並不是只有副詞「盡」一個。
- (2) 不能帶賓語的包括形容詞和不及物動詞，但是形容詞可以受副詞修飾，而不及物動詞不能受副詞修飾，以此標準便可以把形容詞和不及物動詞分開。
- (3) 動詞可以受「無、毋、旨」其中之一個否定詞否定，如「無拿、毋食、旨寫」。
- (4) 動詞可以構成「V 毋」格式，如「拿無著、走毋出、歇暍毋得」。
- (5) 動詞能帶「了、等、過、著」表示完成或進行等，如「食了、行等過、睡過、咬著」
- (6) 動詞可以帶「啊」，構成「V 啊 X」格式，如：「行啊出、拿啊著、食食啊哩」。
- (7) 動詞前可以加副詞修飾，如「盡打、盡食、恁希望你來」
- (8) 動詞可以帶「賓語、補語」，如「打著了、打佢、食飽過」
- (9) 動詞可以做「黏合式補語」，如「送來、趨出去、擲去」
- (10) 不能做「主語」。
- (11) 有一些動詞例如「影響、教育、調查」等具有名詞的功能，如可以作動詞「有」的賓語，構成「有影響、有調查」；也可以受「數量」結構修飾，構成「一種調查、三項教育」，這種情形我們處理爲兼類。

根據以上的定義，我們可以製定一個辨認動詞的操作表來操作：

| 詞例 | 無~ | 毋~ | 旨~ | ~毋 | ~了 | ~等 | ~過 | ~著 | ~啊 X | 盡~ | 盡~(賓) | ~(賓) | ~(補語) | 黏合 式補 語 | 做主語 | 做謂語 | 狀~ |
|----|----|----|----|----|----|----|----|----|---------|----|-------|------|-------|---------------|-----|-----|----|
| 拿 | + | + | + | + | + | + | + | + | + | + | - | + | + | + | - | + | + |
| 食 | + | + | + | + | + | + | + | + | + | + | - | + | + | + | - | + | + |
| 嗽 | + | - | + | + | + | + | + | + | + | + | - | - | + | - | - | + | + |
| 歇暍 | + | + | + | + | + | - | + | - | + | - | - | - | - | - | - | + | + |
| 合意 | + | + | - | - | + | - | + | + | - | + | + | + | - | - | - | + | + |

(表六, 鑑別「動詞」操作表)

(二) 形容詞：

- (1) 形容詞是可以受「盡」修飾，但不能帶賓語的一個集合，這裡的「盡」是指程度副詞而言，包括「當、盡、忒、一等、較、十分、非常、特別」等，這些程度副詞我們叫「盡」類，如「當肥、一等靚、忒大箍」，但是並不是所有形容詞都可以受「盡」類中的每一個直接修飾，也不是所有形容詞都可以受程度副詞修飾。
- (2) 形容詞可以受否定副詞「無、毋、旨」修飾，如「無靚、毋澎湃」
- (3) 我們定義形容詞不能帶賓語，像「輕兜、短一截」這種情形，是賓語還是補語眾說紛紛，我們暫時定義為「補語」，因此形容詞可以帶補語，不必再另立一條形容詞可以帶賓語的規則。

根據上面的定義正好與動詞互斥，可以把動詞和形容詞分開來。我們可以製定一個辨認形容詞的操作表來操作：

| 詞例 | 盡~ | 當~ | 忒~ | 無~ | 毋~ | 旨~ | 作定語 |
|----|----|----|----|----|----|----|-----|
| 靚 | + | + | + | + | + | + | + |
| 癩 | + | + | + | + | + | + | + |
| 大籊 | + | + | + | + | - | + | + |
| 弦 | + | + | + | + | + | + | - |
| 痾糟 | + | + | + | + | + | - | + |

(表七, 鑑別「形容詞」操作表)

(三) 狀態詞

- (1) 狀態詞是不能受程度副詞「盡」類修飾的，這是區分形容詞和狀態詞的標準，如「肥鼓鼓、烏梭梭、靚妮妮仔」，形容詞「肥、烏、靚」可以受程度副詞「盡」類修飾，構成「盡肥、十分烏、忒肥、當靚」，但是「肥鼓鼓、烏梭梭、靚妮妮仔」不能和程度副詞結合構成「*忒肥鼓鼓、*盡烏梭梭、*當靚妮妮仔」。
- (2) 狀態詞不能受否定副詞「毋」修飾，也不能帶賓語，這是區分狀態詞和動詞的標準。如不能說「*毋肥鼓鼓、*毋烏梭梭、*毋靚妮妮仔」，也不能說「*肥鼓鼓人、*烏梭梭車、*靚妮妮仔細阿姊」。
- (3) 有些單音節形容詞重疊後加上「介」，才能成為狀態詞，我們把重疊後的形容詞視為一個分詞單位，如「臭臭介圓身、肥肥介鴨仔」
- (4) 狀態詞可以做補語，而且一定是構成述補結構的補語，如「扳核核、搯核核、倚 ten55 ten55、食到肥鼓鼓」

根據以上的定義，我們可以製定一個辨認狀態詞的操作表來操作：

| 詞例 | 盡~ | 忒~ | 毋~ | ~賓語 | ~介 | 作補語 |
|------|----|----|----|-----|----|-----|
| 白雪雪 | - | - | - | - | + | + |
| 烏梭梭 | - | - | - | - | + | + |
| 歪歪 | - | - | - | - | + | + |
| 鬱鬱卒卒 | - | - | - | - | + | + |
| 光光光 | - | - | - | - | + | + |

(表八, 鑑別「狀態詞」操作表)

(四) 量詞

- (1) 量詞最主要是能受數詞或數詞短語修飾, 如「一斤、十五間、三點鐘」。
- (2) 量詞可以和指示詞結合, 而且當所指為「一」時, 常省去「一」, 構成「這種、該埕、這間」的結構; 量詞也可以和動詞結合, 構成「V(一)量」的結構, 也常省略「一」, 構成「買張、拿疊、拿條」的結構, 我們定義這種結構裡的「種、埕、間、張、疊、條」仍為量詞。
- (3) 量詞不能做主語, 一般而言也不做賓語, 只有下列三種情況例外:
 1. 在動詞「論、算」之後做賓語: 如「論斤來賣、算分算錢」。
 2. 在「**拷**~算做/看做/換做~」的結構裡做賓語: 如「**拷**斤看做兩、**拷**斤換做兩」。
 3. 在「用~算」的結構裡做賓語: 如「用題算、用點鐘算」

因此我們可以根據量詞不能做主語又能受數詞修飾的特點, 把量詞和名詞分開。
- (5) 有些量詞可以受形容詞修飾, 如「一大輪, 一長料、三大團」。
- (6) 有些單音節的量詞可以重疊, 量詞重疊後有些變成可以做定語、主語甚至狀語, 如「條條大路透台北」(做定語)、「儕儕全當乖」(做主語)、「擺擺去乜無在」(做狀語), 重疊後的量詞從意義上看其實不再表示單位, 而是

表示數量，而且從可以做主語、定語甚至狀語來看，似乎失去量詞的語法功能特徵，但現階段我們仍將這些重疊的量詞，定義為量詞。

根據以上的定義，我們可以製定一個辨認量詞的操作表來操作：

| 詞例 | 一~ | 幾(多)~ | 這(一)~ | V(一)~ | 與形容詞 結合 | 做賓語 | 重疊 |
|----|----|-------|-------|-------|------------|-----|----|
| 寶 | + | + | + | + | + | + | + |
| 袋 | + | + | + | + | + | + | + |
| 擔 | + | + | + | + | + | + | + |
| 針 | + | + | + | + | + | + | + |
| 擺 | + | + | + | + | - | + | + |

(表九，鑑別「量詞」操作表)

(五) 方位詞

- (1) 方位詞需要有參照物，最重要的是指能進入「在/到+名詞/動詞」結構中的一些後附的成分，如「在桌頂、到佢面前來」，這些如「頂、面前」的後附成分，表示一種空間關係或時間關係，「在」為介詞，構成介賓短語，「到」為動詞，構成述賓短語。
- (2) 像「三月底、十年之間、三年以來、九月初」這種結構裡的「底、之間、以來、初」等，也表示一種時間關係，所以也劃入方位詞。
- (3) 表示處所的合成方位詞，可以受副詞修飾，如「忒頂高了、盡後背、最南片」。
- (4) 有些方位詞可以直接充當句法成分，如「前蹶毋當後踐」，「前、後」充當主語，還有些直接作狀語，如前舉「前、後」或如「底背講，外背莫講」。

根據以上的定義，我們可以製定一個辨認方位詞的操作表來操作：

| 詞例 | 在~ | 到~ | 向~ | (在/到+名詞~) |
|----|----|----|----|-----------|
| 背後 | + | + | - | + |
| 底 | - | - | - | + |
| 之間 | - | - | - | + |
| 以南 | - | - | - | + |
| 北 | - | - | + | + |

(表十，鑑別「方位詞」操作表)

(六) 時間詞

- (1) 可以做「在、到、等到」的賓語的一類詞，如「在宋朝加下、到掛紙該日、等到韶早」。「在」為介詞，構成介賓短語，「到、等到」為動詞，構成述賓短語。
- (2) 可以用「奈時」、「麼介時節」提問，也可以用「這時(節)」、「該時(節)」、「該量時」等指代。如「麼介時節轉？回答：五月節」、「立冬該時會去」。

根據以上的定義，我們可以製定一個辨認時間詞的操作表來操作：

| 詞例 | 在~ | 到~ | 等到~ | 提問：奈時 | 指代：這時 | 指代：該時 |
|-----|----|----|-----|-------|-------|-------|
| 晚晡頭 | + | + | + | + | + | + |
| 五月節 | + | + | + | + | + | + |
| 立冬 | + | + | + | + | + | + |
| 過年 | + | + | + | + | + | + |
| 唐朝 | + | + | + | + | + | + |

(表十一，鑑別「時間詞」操作表)

(七) 處所詞

- (1) 處所詞表示一個空間位置，能做「在、到」的賓語，「在」為介詞，構成介賓短語，「到」為動詞，構成述賓短語，如「在門口、到天頂去了」。
- (2) 處所詞還可以用「奈位」來提問，也可以用「這位」或「該位」來指代。
- (3) 有些處所詞可以當狀語，這可以明確地做為區分名詞和處所詞的依據，因為名詞絕不做狀語，如「到屋肚講、去伯公該講」

根據以上的定義，我們可以製定一個辨認處所詞的操作表來操作：

| 詞例 | 在~ | 到~ | 提問：奈位 | 指代：這位 | 指代：該位 |
|-----|----|----|-------|-------|-------|
| 街路項 | + | + | + | + | + |
| 門口 | + | + | + | + | + |
| 耳空唇 | + | + | + | + | + |
| 天頂 | + | + | + | + | + |
| 圓身項 | + | + | + | + | + |

(表十二，鑑別「處所詞」操作表)

(八) 名詞

- (1) 不受副詞修飾，如「*忒電腦、*盡火車」是不合語法的。
- (2) 大多數名詞可以受「數量詞或數量短語」修飾，如「二斤柑仔、二頭禾、四十五尺布仔」。
- (3) 名詞多可構成「在~頂」的結構，即帶上方位詞後可以做「在」的賓語（方位詞「頂」只是一個代表），如「在車頂、在火車頭項、在拖箱肚」。
- (4) 名詞多可作主語和賓語，但不做狀語。
- (5) 名詞不能帶「了、等、過」這種表時態的助詞。

以上標準只是一個大概，我們也發現有些名詞並不能做主語或賓語，也有些名詞不能受數量短語修飾，更有許多名詞不能做定語，現階段我們採取寬鬆的標準，並靠語義來協助劃分。

根據以上的定義，我們可以製定一個辨認名詞的操作表來操作：

| 詞例 | 主 | 賓 | 數量~ | 數~ | 定 | 在~頂 |
|----|---|---|-----|----|---|-----|
| 桌仔 | + | + | + | - | + | + |
| 公婆 | + | + | - | + | + | - |
| 禾 | + | + | + | - | + | - |
| 方面 | - | - | - | + | + | - |
| 電火 | + | + | + | - | + | + |

(表十三，鑑別「名詞」操作表)

(九) 擬聲詞

- (1) 可以構成「~~滾」、「~聲仔」的結構，如「咚咚滾、p'iak50一聲、tiaŋ55聲仔」。
- (2) 大部分的擬聲詞都可以帶上「介」(的)做定語，但也可以不帶「介」(的)，如「麼人k'iaŋ55介聲？、滴答聲」。
- (3) 擬聲詞可以獨立成句法成分，如「嗚-----，仰會恁樣？、哈哈！佢跌倒了！」
- (4) 擬聲詞不能受否定副詞「毋」否定，也不能受「盡、忒、非常」等副詞修飾，如「*毋咩咩滾、*盡ve55ve55滾」

根據以上的特點，我們可以製定一個辨認擬聲詞的操作表來操作：

| 詞例 | ~~滾 | ~聲仔 | ~介 | 毋~ | 盡~ |
|---------|-----|-----|----|----|----|
| 哈哈 | + | - | + | - | - |
| p'iak50 | + | - | + | - | - |
| k'iaŋ55 | + | + | + | - | - |

| | | | | | |
|-------|---|---|---|---|---|
| kut20 | + | - | + | - | - |
| 噹 | + | + | + | - | - |

(表十四，鑑別「擬聲詞」操作表)

(十) 數詞

客家話的數詞是一個封閉類，我們採取列舉的方式做為判斷準則。

- (1) 係數詞：零、一、二、三、四、五、壹、貳、叁、肆、伍、陸
- (2) 位數詞：個、拾、百、千、萬、拾萬、百萬、千萬、億、兆、
- (3) 限定數詞：盡多、好多、益多、幾多、幾、大多數、大部分。
- (4) 助數詞：分之、點、百分之、零、上下、空、過、
- (5) 序數詞：第一、第二、第三、第五十、初一、初二、初三
- (6) 概數詞：零、上下、大零、零零

(十一) 區別詞

- (1) 區別詞主要作定語，通常只出現在名詞或助詞前的黏著詞，如「初等教育、良性瘤、高速公路、正式通知單、急性介斯麻煩了」。
- (2) 區別詞數量詞或數量短語直接結合時，也必須構成「~〈介〉名詞」的結構，如「一種慢性介癌、一張正式介通知單」，不能說成「*一種慢性、*一條高速」，這樣就把區別詞和名詞分開了。
- (2) 區別詞不能獨立做「主語和賓語」，也不能受「盡、毋」修飾，如不能說「*盡公共、*毋高等」。這就把區別詞和形容分開了。
- (3) 區別詞就意義而言是一種分類詞，而且經常成對或成組出現，如「定期、長期；公立、私立；公家、私人；慢性、急性」，但成對或成組不見得就是區別詞，例如「慢性、急性、良性、彈性」中的「慢性、急性、良性」是區別詞，但是「彈性」卻是名詞，因為可以說「有彈性、彈性很好」。

根據以上的定義，我們可以製定一個辨認區別詞的操作表來操作：

| 詞例 | ~〈介〉名詞 | 盡~ | 毋~ | ~名 |
|----|--------|----|----|----|
| 公立 | + | - | - | + |
| 高等 | + | - | - | + |
| 定期 | + | - | - | + |
| 公家 | + | - | - | + |
| 良性 | + | - | - | + |

(表十五，鑑別「區別詞」操作表)

(十二) 副詞

(1) 副詞是只能做狀語，修飾動詞和形容詞的成分，如「忒高、當燒、毋使（不必）加講、識看過」。

(2) 能做狀語的卻不一定是副詞，要和其他詞類標準綜合來考察。如：

目瞠瞠仔him11人（狀態詞）

認真讀書（形容詞）

恬恬仔行過來（狀態詞，恬是形容詞）

(3) 少數副詞可以修飾方位詞、數量短語等，如：「大約三十斤仔」、「最東片該位」。

(4) 少數副詞可以放在主語前面，如「好得你來了。」、「敢講你毋識佢？」、「到底你有拿無？」

(5) 少數副詞可以獨立成句或帶上語氣詞成句，如「無哦！」、「隨下哦！」、「遽遽！」、「毋哦！」

根據以上的定義，我們可以製定一個辨認副詞的操作表來操作：

| 詞例 | ~動詞 | ~形容詞 | ~名詞 | ~哦 | 語料例 |
|----|-----|------|-----|----|----------|
| 較 | + | + | - | - | 較食乜無幾多了。 |
| 正 | + | + | - | - | 正轉來定。 |
| 堵堵 | + | + | - | - | 堵堵來。 |
| 好得 | + | - | + | + | 好得拿著了。 |
| 滿奈 | + | - | - | + | 滿奈去告人。 |

(表十六，鑑別「副詞」操作表)

(十三) 介詞

- (1) 介詞不能單用，介詞一定要和後面的成分結合才可以進入句法結構中做補語或狀語等，如「分佢拿走去了。」、「在桌頂項睡忒了」、「搵佢借三千銀」。
- (2) 介詞不能單獨做補語，這就把動詞和介詞區分開來，如「該本書分佢」中的「分(給)」是動詞，「分」可以和其他動詞結合構成「拿分、擊分、踢分」的述補結構，也可以有重疊(「分分啊」、「分啊分)或與時態助詞結合「分過、分忒」等形態變化，但是介詞完全沒有形態變化也不能重疊。下面各例中左邊的都是動詞，右邊的才是介詞。

火車到台北了

到台北去療三日

二儕來比看啊

佢比佢高點仔

佢識分佢一支筆

分佢看一下

- (3) 「同、搵 (lau53)」是介詞或是連詞的區分方法是：如果在「同、搵」的前後詞可以互換位置且意思不變，而且這些詞前面不再有修飾成分，同時又

是中置的成分，則這就是「連詞」，否則就是介詞。下面各例中左邊的都是連詞，右邊的才是介詞。

阿婆**搵**孫仔共下行等轉

搵孫仔借錢

書同衫全部收好了

同衫收好來

根據以上的定義，我們可以製定一個辨認介詞的操作表來操作：

| 詞例 | 前後成分不具 交換性 | 沒有形態 變化 | 在句首 | 語料例 |
|----|---------------|------------|-----|-----------------------|
| 分 | + | + | + | 分 俾 看一下。 |
| 佇 | + | + | - | 你到屋下來，係無？ |
| 在 | + | + | + | 在桌頂項放等。 |
| 向 | + | + | + | 向門背看。 |
| 爲著 | + | + | + | 爲著將來，千辛萬苦也愛 kai53。 |

(表十七，鑑別「介詞」操作表)

(十四) 連詞

- (1) 連詞是連接「詞、短語及句子」的黏著成分，和介詞的區別參考前面說明。
- (2) 連詞可以分爲兩小類，一類是連接「詞或詞組」，構成「聯合結構」的連詞，這種連詞有一特性，就是連詞前後的成分具有交換性，常見的如「同、**搵**」。
- (2) 另一類連詞是用來連接「分句、句子或句群的」，這種連詞常成對出現，如「雖然-----毋過」、「毋單淨-----同時」、「因爭-----故所」、「算來-----抑係」

等等。

- (3) 有些連詞一定要和副詞配合使用，才起連接作用，如「若係-----就」、「斯有-----正」、「就算-----也」

根據以上的定義，我們可以製定一個辨認連詞的操作表來操作：

| 詞例 | 前後成分具 交換性 | 中置成份 | 成對性 | 語料例 |
|-----|--------------|------|-----|------------------------|
| 搵 | + | + | - | 佢搵偈共下轉屋下。 |
| 毋單淨 | - | + - | + | 毋單淨佢滿奈講定，同時還會 添油加火。 |
| 因爭 | - | + | + | 莫因爭食就相打 |
| 同 | + | + | - | 夢著書生來同佢同床共枕 |
| 若係 | - | + - | + | 若係你毋來，就無機會了。 |

(表十八，鑑別「連詞」操作表)

(十五) 助詞

- (1) 助詞是黏著詞，黏著於詞、短語或句子上，標示著某種語法意義，例如表動詞時態的時態助詞「了、等、過、忒」及結構助詞「介」等。

- (2) 結構助詞「介」我們獨立給予標記「/kai」。

- (3) 「到」也是黏著的，我們獨立給予標記「/to」，如「拿到恁重」、「做到當好」、「咬到當痛」。

- (4) 助詞是可以列得完的集合，我們採列舉方式來定義：

結構助詞：介、到（後置），連、所（前置）

時態助詞：了、等、過、忒（後置）

比況助詞：共樣、樣仔

數助詞：分之、點、百分之、零、上下、空、過

其他助詞：等、等等、來講、來

(十六) 語氣詞

- (1) 語氣詞置於句尾，後面總是跟著標點符號，一般多是句號、問號或驚嘆號，如「無早囉！」、「唔來 ne33！」、「佢大了！」、「麼人去 no55？」、「生到還靚啊！」。
- (2) 表完成的「了」雖位於句末，但我們全部將「了」劃入「助詞」，表語氣的我們用「哩」這個漢字，這個語氣詞有些人會發音成「le3」，和表完成的「了」同音，應該是上聲「le24」。
- (3) 兩個語氣詞連用，並不是合成語氣詞，如「食飽哩 ka24？」，其結構是（食飽哩）ka24？
- (4) 「啊哩、也好、乜好、好哩」是合成語氣詞，如「食看啊哩」「有好食也好，無好食也好」、「佢無來乜好哦！」、「來 (le53) 走好哩！莫加想佢多。」
- (5) 語氣詞也是個封閉集合，我們採用列舉的方式處理。

(十七) 嘆詞

- (1) 嘆詞是可以獨立使用的，從結構來看，嘆詞多數和後面的句子沒有關係，嘆詞後面一定有個停頓，書面上多用標點符號表示，如「！、。」
- (2) 嘆詞必定前置，語氣詞必定置於句尾，兩者就判然分開了。如下面各例中左邊的都是嘆詞，右邊的才是語氣詞。

啊！毋記得了。

毋記得了啊！

嗯，做得。

做得哩哦！

哦！知了。

知了哦！

(十八) 代詞

(1) 凡是具有臨時稱代功能的詞都劃入代詞，嚴格說來代詞其實是各種詞類的集合，並不能算是一個詞類，我們根據其所稱代的詞類來分：

- 1、名詞性代詞：如「**佢**、你、佢、**佢兜**、你兜、佢兜、麼人、麼介、這、該」等，這類代詞可以作主、賓語。
- 2、動詞性代詞：如「**恁樣**、**仰般**、**仰般形**」等，這類代詞可以受副詞「無」修飾，可以做謂語、補語，但不能受副詞「盡」修飾。動詞性代詞必定置於動詞之後，如「講恁樣、寫到仰般形」。
- 3、處所性代詞：如「**這位**、**該位**、**這**、**該**、**奈位**、**奈項**、**該項**、**這項**、**該位項**」等。
- 4、時間性代詞：如「**這時**、**該時**、**該量時**、**這量時**、**加下**」
- 5、數詞性代詞：如「**幾多**、**kio53**、**多少**」
- 6、副詞性代詞：「**恁樣**、**恁樣形**、**仰般**、**仰般形**」，副詞性代詞必定置於動詞前，如「**恁樣做**、**仰般行**、**仰般形扛**、**恁樣形食**」

第六章 客家語短語結構分類與標記說明

第一節 短語的分類與標記符號

短語的分類我們採取依據字串的「內部結構關係」為分類標準，我們劃分客家語的短語為二十二種類型，分別是：主謂短語、述賓短語、述補短語、連動短語、狀中短語（動詞性）、聯合短語、附加短語、聯合短語、方位短語、數量短語、同位短語、定中短語、介（kai）字短語、時間短語、處所短語、狀中短語（形容性）、聯合短語、介（kai）字短語、形補短語、比況結構、介詞短語、副詞短語，下面是我們語料庫裡頭所使用的各種短語的標記表和相應的實例。

| 序號 | 短語性標記 | 短語標記 | 短語名稱 | 概念 | | 實例 |
|----|-------|------|------|-----|-------------|--------------------------|
| 1 | cp | zu | 主謂短語 | 主語 | 謂語 | 天空落水、雞啼、狗吠 |
| 2 | vp | vo | 述賓短語 | 述語1 | 賓語 | 食一碗飯、想打佢、送佢一包菸 |
| 3 | vp | vc | 述補短語 | 述語2 | 補語 | 洗淨、做著益好、拿出來、走一輪、請阿婆來、趨佢走 |
| 4 | vp | ld | 連動短語 | 前項 | 後項 | 打門出去、打電話請醫生、笑等講 |
| 5 | vp | zd | 狀中短語 | 狀語 | 中心語 (vp) | 遽走、盡打、分佢趨走、艱苦賺、屋肚僚 |

| 序 號 | 短語 性標 記 | 短語 標記 | 短語名稱 | 概念 | | 實 例 |
|--------|---------------|----------|---------------|-----------|-------------|--|
| | | | | 前項 | 後項 | |
| 6 | vp | bl | 聯合短語 | 前項 | 後項 | 唱歌 搵 跳舞、又 噉 又笑、 參詳參詳、泗水打球 |
| 7 | vp | fg | 附加短語 | 動詞 | 時 態 標 記等 | 食飽了、行等出去、咬斷 去了、試食看啊 |
| 8 | np | bl | 聯合短語 | 前項 | 後項 | 民族英雄、白菜 搵 豬肉 |
| 9 | np | fu | 方位短語 | | | 大海唇、桌頂、台北到桃 園之間 |
| 10 | np | dm | 數量短語 | 數 | 量 | 一張、二十零頭、三擺、 一本 |
| 11 | np | tu | 同位短語 | 前項 | 同後項 | 噉倆儕、你這兜後生人 |
| 12 | np | dd | 定中短語 | 定語 | 中心語1 | 一斤白菜、先生介目汁、 紅噉噉介燈籠、削梨介 刀、冰涼介菜 |
| 13 | np | kai | 介(kai) 字短語 | 名詞或 動詞 | 介 | 買菜介、先生介、阿公介 |
| 14 | np | ts | 時間短語 | | | 今晡日下晝、過年、二十 年 |
| 15 | np | ss | 處所短語 | | | 灶下項、廳下肚、洗身間 肚、 |
| 16 | jp | zd | 狀中短語 | 狀語 | 中心語2 | 當靚、盡姣、 |
| 17 | jp | bl | 聯合短語 | 前項 | 後項 | 又高又靚；高矮長短 |

| 序號 | 短語性標記 | 短語標記 | 短語名稱 | 概念 | | 實例 |
|----|-------|------|-----------|-------------|-----|---------------------|
| 18 | jp | kai | 介(kai)字短語 | 狀態詞 或形容詞 | 介 | 高高介、肥肥介、淨淨介、屙糟介、顛東介 |
| 19 | jp | jc | 形補短語 | 形容詞 | 補語 | 鬧熱到無葛無煞、大多了、屙糟到無人識 |
| 20 | jp | pk | 比況結構 | 前項 | 似後項 | 發癲樣仔、像蜂仔樣、啞狗共樣 |
| 21 | pp | pb | 介詞短語 | 介詞 | 介賓語 | 在桌頂、因為欠水、揸大家參詳、為人民 |
| 22 | rp | bl | 副詞短語 | 前項 | 前項 | 隨下斯~、絕對毋~、定著旨~ |

國立新竹教育大學 (表十九, 短語標記與實例)
National Hsinchu University of Education

第二節 短語性與標記符號

關於「短語性」及其範圍，我們的定義是：

(一) 主謂短語 (cp)

主謂短語實際上並沒有「短語性」可言，但是為了標記上一致性的需要，同時根據我們的考察，主謂短語加上適當的語調或在具體的語境下，能非常自由的成為「句子」，我們稱為「子句」。根據這樣的特性，我們增加一個專門給主謂

短語使用的「短語性」，並標記為「cp」。

(二) 動詞性短語 (vp)

動詞性短語包括以下六種，標記符號為「vp」：

- 1、述賓短語：打佢、看鬧熱、搵衰人
- 2、述補短語：看清楚、行過來、拿淨淨
- 3、連動短語：打電話請醫生、去泅水、橫等看書
- 4、附加短語：食飽了、行等出去、咬斷去了
- 5、中心語為動詞的狀中短語：恬恬仔行開、
- 6、動詞性的聯合短語：緊行緊講、又打又捶、好搞又好噉、唱歌揸跳舞

(三) 名詞性短語 (np)

名詞性短語包括以下八種，標記符號為「np」：

- 1、名詞性的聯合短語：柑仔、葡萄、弓蕉，台北揸花蓮、阿爸同阿哥
- 2、方位短語：台北到花蓮之間、桌仔頂高、拖箱底肚
- 3、數量短語：二十零頭、八九斤、這儕、該駁仔、歸條、幾尺
- 4、同位短語：你這兜後生人、市長馬英九、
- 5、中心語為名詞的定中短語：新買介書、厥介書、偲大哥、一斤白菜
- 6、名詞性的介 (kai) 字短語：買菜介、菜頭介、先生介、過路介
- 7、時間短語：今晡日下晝、過年、二十年、掛紙前、
- 8、處所短語：灶下項、廳下肚、洗身間肚

(四) 形容性短語 (jp)

- 1、中心語為形容詞的狀中短語：當高大、盡靚、忒過遽、暢過厝糟
- 2、形容性的聯合短語：又高又靚、高矮長短
- 3、形容性的介 (kai) 字短語：肥肥介、高高介、硬頸介
- 4、形補短語：暢到歸身搖、硬頸到無人識
- 5、比況短語：發癲樣仔、像臊牯共樣、啞狗樣、像飛共樣

(五) 介詞短語 (pp)

凡是帶上介詞的全部劃歸介詞短語，如：為著細人仔、在這、搵孫仔

(六) 副詞短語 (rp)

副詞並列使用時的短語，如：隨下斯~、絕對毋~、定著旨

第三節 短語的標記與示例

National Hsinchu University of Education

在我們進行分詞和詞類的確認之後，接著便是進行短語的標記工程，由於客家語的文本是採取漢字連寫的，詞與詞之間並沒有分隔符號，短語和短語之間也沒有分隔符號，所以我們進行短語的標記，實際上就是為「詞與詞」、「短語和短語」及「詞和短語」之間增加分隔符號，我們採取在一個短語的後邊帶上「/和短語標記」的方式，下面是短語標記的符號說明：

「/」：短語的標記符號。

「~」：區分短語性和短語結構的符號。

以下分就各類短語的標記方式進行示例。

| 序號 | 短語名稱 | 示例 |
|----|--------------|---|
| 1 | 主謂短語 (zu) | 天空落水/cp~zu.2si.2vv 老弟食飯/cp~zu.2nn.2vv 今晡日拜三/cp~zu.3nt.2od |
| 2 | 述賓短語 (vo) | 搵佢人/vp~vo.2vv.1pn 惜妹/vp~vo.1vv.1nn 離別爺娘/vp~vo.2vv.2nn |
| 3 | 述補短語 (vc) | 行出去/vp~vc.1vv.2vv 走轉來/vp~vc.1vv.2vv 講清楚/vp~vc.1vv.2jj |
| 4 | 連動短語 (ld) | 去泅水/vp~ld.1vv.。2vv 打門出去/vp~ld.2vv.2vv 伸箸挾菜/vp~ld.2vv.2vv |
| 5 | 動詞性狀中短語 (zd) | 煞猛抔/vp~zd.2rb.1vv 盡打/vp~zd.1rb.1vv 無降/vp~zd.1rb.1vv |
| 6 | 動詞性聯合短語 (bl) | 唱歌搵跳舞/vp~bl.2vv.1cc.2vv ; 搞、睡搵食/vp~bl.1vv.1pu.1vv.1cc.1vv ; <緊行/vp~zd.1rb.1vv 緊講 /vp~zd.1rb.1vv>/vp~bl.1vp.1vp |
| 7 | 附加短語 (fg) | 打起來/vp~fg.1vv.2ax 拿等/vp~fg.1vv.1ax 食過/vp~fg.1vv.1ax |

| 序號 | 短語名稱 | 示例 |
|----|-------------------------|--|
| 8 | 名詞性聯合短語 (bl) | 凳仔搵桌仔/np~bl.2nn.1cc.2nn 基隆、桃園、新竹/np~bl.2nr.1pu.2nr.1pu.2nr |
| 9 | 方位短語 (fu) | 大樓頂高/np~fu.2nn.2lc 九月初/np~fu.2od.1lc 盡後背/np~fu.1rb.2lc |
| 10 | 數量短語 (dm) | 一張/np~dm.1cd.1m 二十公斤/np~dm.2cd.2m 四百條/np~dm.1cd.1cd.1m |
| 11 | 同位短語 (tu) | 嚟俚客家人/np~tu.2pn.3nn 故鄉花連/np~tu.2nn.2nr 市長馬英九/np~tu.2nn.3nr |
| 12 | 定中短語 (dd) | <十介/np~dm.1cd.1m 人 /nn>/np~dd.1np.1np <桌仔介/np~kai.2nn.1kai 腳 /nn>/np~dd.1np.1np |
| 13 | 名詞性介 (kai) 字短語 (kai) | 買介/np~kai.1vv.1kai 駛車介 /np~kai.2vv.1kai 塑膠介/np~kai.2nn.1kai |
| 14 | 時間短語 (ts) | 今晡日下晝/np~ts.3nt.2nt 過年/np~ts.2nt 二十年/np~ts.2cd.1nt |
| 15 | 處所短語 (ss) | 灶下項/np~ss.2nn.1lc 廳下肚/np~ss.2nn.1lc 洗身間肚/np~ss.3nn.1lc |

| 序號 | 短語名稱 | 示例 |
|----|-------------------------|---|
| 16 | 形容性聯合短語 (zd) | <又靚/jp~zd.1rb.1jj 又高大 /vp~zd.1rb.2jj>/jp~bl.1jp.1jp ; <當尖/jp~zd.1rb.1jj 當臭 /jp~zd.1rb.1jj>/jp~bl.1jp.1jp |
| 17 | 形容性介 (kai) 字短語 (kai) | 高高介/jp~kai.2zht.1kai 硬頸介/jp~kai.2jj.1kai 屌糟介/jp~kai.2jj.1kai |
| 18 | 形補短語 (jc) | <硬頸/jj 到/to 無法度 /vp~vo.1vv.2nn>/jp~jc.1jp.1to.1vp ; <大多/jp~jc.1jj.1jj 了/ax>/jp~fg.1jp.1ax |
| 19 | 形容性狀中短語 (zd) | 盡靚/jp~zd.1rb.1jj 忒過遽/jp~zd.2rb.1jj 益豐湧/jp~zd.1rb.2jj |
| 20 | 比況短語 (pk) | 像細人樣仔/jp~pk.1vv.2nn.2ax 像細新婦樣/jp~pk.1vv.1np.1ax |
| 21 | 介詞短語 (pb) | 為著細人仔/pp~pb.2pa.3nn 在桌頂/pp~pb.1pa.2lc 搵該本書/pp~1pa.1np |
| 22 | 副詞短語 (bl) | 隨下斯/rp~bl.2rb.1rb 絕對毋/rp~bl.2rb.1rb 定著旨/rp~bl.2rb.1rb |

(表二十，各類短語的標記方式示例)

第四節 短語類型的判定

(一) 主謂短語

主謂短語是由主語和謂語兩部分組成，主語是謂語的陳述對象，謂語陳述了主語「怎麼了？」這個命題，主語和謂語的結構關係是陳述和被陳述的關係。客家話的主謂短語按照內部的結構成分，可以分為以下幾種形式：

(1) 主語+動詞及主語+動詞性短語

雞啼、阿婆來、大家唱、阿公睡當晝、老弟食乳、佢去台北、佢轉花蓮、老弟好搞又好噉、阿哥橫等看書、

(2) 主語+形容詞及主語+形容性短語

佢盡靚、阿伯當高大、飯菜恁豐湧、佢像臊牯樣、阿姨又高又靚、佢人當硬頸、該手帕屙糟、老弟啞狗樣

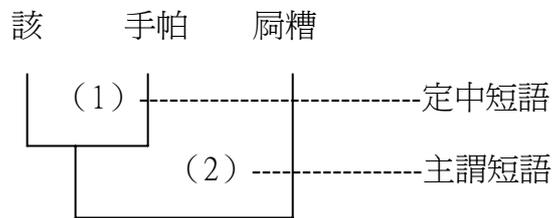
(3) 主語+名詞及主語+名詞性短語

韶早好天、後日掛紙、今晡日十五、一斤白菜十五箍、這張桌仔三隻腳、菜頭幾多條

主謂短語的判斷，可以在謂語之後用「係無？」來提問，可以用「係無？」來提問的便是主謂短語，否則則不是，如「阿公睡當晝，係無？」、「該手帕屙糟，係無？」、「今晡日十五，係無？」客家話也可以改成「主語+仰般」的問句形式，如「雞仔仰般？阿公仰般？後日仰般？」。用「係無？」來提問或改為「主語+仰般」問句形式的問題焦點都是詢問「謂語」。

主謂短語還可做謂語，如「該手帕屙糟、後日佢去掛紙」，一般語法書認為這是「主謂謂語句」，即認定這是主謂結構的一種變化格式，雖然這不

是本文討論的重點，但是我們在標記語料時根據問題焦點詢問「謂語」這一特點，我們並沒有將這種結構分析為「主謂謂語句」結構，而是分析為一般的主謂結構。例如「該手帕扇糟」我們的分析圖示如下：



根據以上的分析圖我們把這種主謂謂語結構標記為：

<該手帕/np~dd.1pn.2nn 扇糟/jj>/cp~zu.1np.1jp

<後日^僵/np~dd.2nt.1pn 去掛紙/vp~ld.1vv.2vv>/cp~zu.1np.1vp

這麼做還有一個好處是可以減少標記上的困擾，減少一些規則，試比較：「後日^僵去掛紙」這個短語的兩種分析標記結果：

A、分析為主謂謂語結構的標記：

(後日/nt <^僵/pn 去掛紙
/vp~ld.1vv.2vv>/cp~zu.1np.1vp)/cp~zu.1np.1cp

B、分析為一般主謂短語的標記：

<後日^僵/np~dd.2nt.1pn 去掛紙
/vp~ld.1vv.2vv>/cp~zu.1np.1vp

若分析為「主謂謂語結構」則必須多出一個層次和「/cp~zu.1np.1cp」規則，也就是主謂短語(1cp)又成為「1np」的謂語，層層套疊就不斷擴大，也會增加規則，如果分析為一般主謂短語結構，則不會增加規則和冗餘的套疊層次，所以我們採取這種分析法。

(二) 述賓短語

述賓短語由述語和賓語構成，賓語是述語發生關連的對象，述語一般就是動詞，特別是及物動詞，如「阿公買包子」，主語是「阿公」，述語是「買」，賓語是「包子」。賓語「包子」是和述語「買」直接發生關係，構成「述賓短語」做為主語「阿公」的謂語。

述賓短語裡的特殊結構「雙賓結構」，我們仍歸入述賓結構處理，不需再增加規則，雙賓結構的圖示是：



雙賓結構的標記都遵循這個原則，標記為：

國立新竹教育大學
National Hsinchu University of Education

(送阿公/vp~vo.1vv.2nn <一付/np~dm.1cd.1m 目鏡
/nn>/np~dd.1np.1np)/vp~vo.1vp.1np

而形容詞我們規定不能帶賓語，也就是說我們這種情況下的形容詞視為動詞和形容詞的兼類辭，這樣的好處是我們可以減少一些規則，如：大佢三歲、矮人一截。

例中的「大、矮」多做形容詞使用，但是在「大佢三歲、矮人一截。」的結構中我們視他們為動詞，試比較兩種不同的標記：

A、視為形容詞時：

<大佢/jp~vo.1jj.1pn 三歲/np~dm.1cd.1m>/jp~vc.1jp.1np

B、視為動詞時：

<大佢/vp~vo.1vv.1pn 三歲/np~dm.1cd.1m>/vp~vc.1vp.1np

當視為形容詞時，會增加「/jp~vo.1jj.1pn」、「/jp~vc.1jp.1np」兩條規則，而且與我們對「vo、vc」的定義相違背，但是當視為動詞時，則不必增加任何規則，也符合「vo、vc」的定義，所以我們定義「形容詞不能帶賓語」，把上述的情形以「述賓結構」處理，既明確而且清楚。

(三) 述補短語

述補短語是由述語和補語組成，兩者之間是補充與被補充的關係，我們仍然定義述語不能由形容詞擔任，一般語法書將形容詞帶補語也劃入「述補短語」去，我們則對「形容詞帶補語」的情形重新建立一個「形補短語」的結構，請參考以下的說明。

客家話的述補短語，在述語和補語之間常加上「到」相當於華語的「得」，我們單獨給「到」一個標記「/to」。如「洗到歸身汗」、「講到清清楚楚」、「飛得出去」。

述補短語標記為：

講到清清楚楚/vp~vc.1vv.1to.4jj、飛得出去/vp~vc.1vv.1to.2vv

客家話還有一種述補短語，構成「V 啊 X」的格式，如「拿啊著、行啊去、打啊開來」，能產性非常強，一般論者把「啊」視為一個中插成分，並定義為助詞，我們不做這樣的定義，我們基於標記上的需要，對「啊」後面的補語成分規定如果是單音節時，不再分解他們，視整個「V 啊 X」為一個動詞，當「X」為雙音節或雙音節以上時，則予以切分，「啊」視為黏著成分，黏著於動詞，這樣做的好處是不必增加一條述補結構的規則，試比較兩種分析標記的結果：

A、分析為「動詞—助詞—助詞」的標記法：

拿啊著/vp~vc.1vv.1ax.1vv、打啊開來/vp~vc.1vv.1ax.2vv

B、我們的分析

拿啊著/vv、打啊開來/vp~vc.2vv.2vv

(四) 連動短語 (ld)：

連動短語是兩個動作的連用，兩個動作之間沒有主謂、聯合、述賓、述補、狀中等關係，多數是前一動作是後一動作的方式或手段，後一動作是前一動作的目的。

去泅水/vp~ld.1vv.2vv、打門出去/vp~ld.2vv.2vv、伸箸挾菜/vp~ld.2vv.2vv

(五) 動詞性聯合短語

兩個或兩個以上的詞沒有主次之分、沒有修飾關係，只是平等地串聯在一

起，這種短語叫聯合短語。兩個或兩個以上動詞構成聯合關係所形成的短語叫「動詞性聯合短語」。如：

「唱歌**搵**跳舞」、「攞酒又買菸」、「做詩又作曲」、「座車**搵**行路」

（六）名詞性聯合短語

兩個或兩個以上名詞構成聯合關係所形成的短語叫「名詞性聯合短語」。如：
「阿爸、阿哥**搵**偈」、「紅菜頭、柑仔、芹菜」、「先生**搵**學生仔」

（七）形詞性聯合短語

兩個或兩個以上形容詞構成聯合關係所形成的短語叫「形容性聯合短語」。如：

「又靚又瘦」、「又高又靚」、「油、肥、硬」

國立新竹教育大學
National Hsinchu University of Education

（八）同位短語

同位短語是指各成分之間互相指稱，指同一個事物而言，如「若姆阿貞嫂」，「若姆」就是「阿貞嫂」，兩個成分互相指稱；又如「佢兜自家」，「佢兜」就是「自家」，像這種短語就叫同位短語。

（九）狀中短語

狀中短語由「狀語和中心語」構成，分為以動詞為中心語的狀中結構和以形容詞為中心語的狀中結構兩種。

客家語裡可以做狀語的有：

(1) 副詞、狀態詞、介詞短語、部分的量詞短語、時間詞、處所詞等，

如：

十分靚、輒輒去、恬恬睡（副詞做狀語）

恬恬仔行出去、輕輕仔放下去（狀態詞做狀語）

分佢拿去、在桌頂放、**揸**書拿分佢（介詞短語當狀語）

一擺食忒、做三擺食、一年看一擺（量詞短語做狀語）

韶早等你來、明年去日本、後日來去花蓮（時間詞當狀語）

桌頂放、菜籃肚拿出來、屋肚講（處所詞當狀語）

(2) 某些動詞或動詞性短語也可以作狀語，動詞一般需加上「等」

笑等講、仆等噉、行等出、跪等落、疊到齊齊放、糊到靚靚出去

(十) 定中短語

定中短語由定語和中心語構成，定中短語的中心語定義為都是名詞性的，可以做定語的有：形容詞和形容性短語、區別詞、數量短語、介字短語等。

以下舉一些例子：

數量短語：十儕人、一斤菜頭、三兩金仔、兩條豬仔、該本書

形容詞：舊拖箱、新電腦、細阿哥、大豬牯

形容性短語：肥介豬肉料、矮介人

介字短語：送禮來介人、豁出去介東西、佢介筆、**佢揸**阿姆介屋

(十一) 介(kai)字短語

客家話的「介(kai)字短語」，相當於華語結構助詞「的」，標記為「kai」，是用海陸客家語的發音來訂符號的。客家話可以構成「介字短語」的有：

名詞：掌牛哥仔介口潏水、耕種人介細人仔

時間詞~：頭擺介掌牛哥仔、該央時介細人仔

處所詞~：桌下介書、天頂介雲

方位詞~：該片介親戚

數詞~：三介五倍係幾多？、一斤四十介西瓜買毋起

數量短語~：一斤柑仔介錢，做得買一兩金仔。

代詞~：佢兜介錢、佢介書

動詞~：這係食介，搞毋得。佢點介。

形容詞~：高介、矮介、細介、新介

狀態詞~：烏疏疏介衫、靚妮妮仔介細阿姊

區別詞~：慢性介難醫、彩色介電視、一等介有三儕

主謂短語~：佢兜食介係柑仔、嗯俚出介係錢、阿婆想愛介係孫仔

擬聲詞~：Kut4 聲仔，斯吞落去了。

對「介(kai)字短語」我們的描述體系區分為「名詞性的介(kai)字短語」和「形容性的介(kai)字短語」兩種。除「形容詞~」歸入「形容性的介(kai)字短語」之外，其他全部歸入「名詞性的介(kai)字短語」。

(十二) 介詞短語

介詞主要是附著在其他詞類或短語前的成分，客家話的介詞沒有型態變化，介詞不能帶時態助詞「了、等、過、忒」，不能單獨使用。客家話可以構成介詞短語的有：

~名詞：在桌仔下背、在拖箱底背

~代詞：照你屋下介俗、在若爸該位、對你盡敗勢、分人罵

- ~方位詞：在頭前項、在後背
- ~時間詞：自韶朝晨八點算起、從下只月初一開始
- ~動詞：自出世到老、從耕田討食到做生理
- ~數詞：從一算到一百、從三十加到一百
- ~數量短語：分一儕人看毋起、對三個人來講
- ~聯合短語：照你**恁**介意見、該兜在學校、郵局、市政府食頭路介人
- ~同位短語：向大嫂玉貞潑過去、**恁**兜三儕人癩死去了

(十三) 方位短語

方位詞附著在詞或短語後面構成的短語叫方位短語，客家話可以構成方位短語的有：

- 1、名詞~：田肚、屋面前、凳下、車頭底肚、河壩底、風災過後
- 2、時間詞~：三月以前、四年以內、拜一前、過年後
- 3、方位詞~：桌仔面前、頂背**恁**負責，後背、下背你愛做好來
- 4、數詞~：四十以內
- 5、數量短語~：五斤以內。
- 6、代詞~：麼介唇、麼介以外
- 7、動詞~：考試前、打山豬後。
- 8、主謂短語~：佢食飽過後、**恁**出國之前、阿婆罵人前

(十四) 數量短語

數量短語是由數詞在前和量詞在後結合構成的短語，包括「數量短語、指

量短語、問量短語、定量短語」幾類：

- 1、數量短語：兩張、三擺、第二日、三四台、二十零條、
- 2、指量短語：該份、這條、該五儕、逐儕、逐份、上回、頭禮拜
- 3、問量短語：奈儕、幾多條、幾斤
- 4、定量短語：歸條、歸日、半箱、好多儕、kia11 罐

(十五) 比況短語

客家話的比況短語是由比況助詞「樣、樣仔、共樣、恁樣」等，構成的一個短語，共同充當句子的成分。客家話可以構成比況短語的有：

- 1、名詞~：啞狗樣、形仔像菅草樣仔
- 2、動詞~：像飛樣仔、吵到像相打樣、像阿爸恁樣做
- 3、定中短語~：像麼介好東西樣仔
- 4、狀中短語~：像盡艱苦樣、像當苦樣

國立新竹教育大學
National Hsinchu University of Education

(十六) 形補短語

形補短語是由形容詞帶補語構成，一般語法書把形容詞歸入「述語」，我們則把形容詞從述語分離出來，規定當形容詞帶賓語時處理成動詞，不需增加標記規則，但是如果形容詞帶補語時，仍將形容詞歸入述語的話，便會多出一些規則，試比較：

A、形容詞歸入述語時的標記：

(厠糟/jj 到/to <無人/vp~vo.1vv.1pn 識

/vv>/vp~zd.1vp.1vp)/vp~vc.1jp.1to.1vp

B、形容詞不歸入述語時的標記：

(厠糟/jj 到/to <無人/vp~vo.1vv.1pn 識

/vv>/vp~zd.1vp.1vp)/jp~jc.1jp.1to.1vp

當把形容詞歸入述語時會增加一條標記「/vp~vc.1jp.1to.1vp」，而這是與定義完全違背的，因為「/vp~vc」為述補短語，「vc」之後既出現「vv」又出現「jj」，這是不利於分析的。再則如果把這種情形也像「述賓短語」一樣處理為「動形兼類詞」，則會造成大量的兼類情形，這也是不理想的。

如果我們把形容詞從述語分離出來，並設「形補短語」便可解決這個問題，也可處理掉大量的動詞和形容詞兼類問題，使兼類情形變少，增加處理的方便。

(十七) 附加短語

附加短語是針對動詞後的「了、等、過」及句末語氣詞等而定義的，表示時態的時態助詞起來、了、等、過、忒、去了、去、過了、忒了、著、啊忒及句末語氣詞等和其前面的成分行成一個附加短語，共同做為句子的成分。

1、動詞~：食飽了

2、動詞~：行等過

3、動詞~：拿走哩

(十八) 時間短語

時間短語由時間詞實現而得到，時間詞短語作為句子的成分主要做為

主語，如：

今晡日下晝、過年、二十年、韶朝晨、三十晚晡、春夏秋冬、
舊年八月十五、前年立冬、現時、昨晡日半夜、頭擺、當日臨暗頭

(十九) 處所短語

處所短語由處所詞實現而得到，處所短語作為句子的成分主要做為主語，客家話的處所短語一般都是「名詞+處所詞」的結構，如：

窗門背、河壩唇、圓身項、門口、耳空唇、台北大橋前

(二十) 副詞短語

副詞修飾動詞和形容詞，由副詞並列構成的短語叫副詞短語，如

隨下斯~、絕對毋~、定著言~、全部毋~、輒常無~

國立新竹教育大學
National Hsinchu University of Education

第五節 句子的層次關係與分析

在一個連續的語流中，人們一個一個依次說出一個個的音節，一個個的成分也只能在一個向度上一個接一個的出現，構成線性的序列，但線性序列只是表像，句子並不是一個個詞的線性序列組合，而是按一定的規則一層層組合起來的，句子的長短並不在於組合成分的多少而已，還包括組合層次的多少和層次之間的關係。本文對句子的分析採取層次分析法。

層次分析法源出英語的 Immediate Components (簡稱 IC)，在 20 世紀 20 年代就由美國描寫語言學家布龍菲爾德系統的運用到語言分析中，這種分析法分析的對象是單句，分析的方法是對大單位進行分解，以找出更小的單位，直至分析

出的成分，不能再分解為止。

對句子的分析便是要找出比句子小一級的單位「短語」，對短語的分析便是要找出比短語更小一級的單位「詞」，對詞的分析便是要找出比詞更小一級的單位「語素」。關於層次分析的原則，語法學界儘管有不同的看法，但歸納起來不外乎結構、功能、意義三條原則。所謂結構原則，是指切分後的語言片段各自能成爲一個結構；所謂功能原則，是指切分後的語言片段可以按照漢語的語法規律重新組合起來；所謂意義原則，是指切分後的語言片段不能違背原來短語所具有的意思。¹

層次分析法主要包括兩個步驟，一是對句子成分決定切分點；二是確定成分之間的關係。

對句子成分決定切分點，需注意幾個問題：

(1) 不能憑語感和語音停頓決定，語感有很大的主觀性，因人而異。例如：

「呂先生寫介一篇文章」，這是一個定中結構，如果憑語感切分，可能有以下幾中切分法：

- 1、呂先生 / 寫介一篇文章
- 2、呂先生寫介 / 一篇文章
- 3、呂先生寫介一篇 / 文章

上面三種切法各自有其理，但第二種切法才是合語法規律的。

¹ 《現代漢語短語》，張斌主編，齊滬揚著，上海華東師範大學，2000年。

(2) 切分後所得的成分要具有可再用性，也就是說可在別的句法結構中使用，例如「食有三碗了」可能的切分法有：

1、食 / 有三碗 / 了

2、食有 / 三碗 / 了

按照第一種方法切分取得的各個成分都可以在其他句法結構裡再被運用，如：「佢食咗？、有三碗飯放桌頂。、食飽了。」。

可是如果照第二種方法切分取得的「食有」就不能再被運用，只能在「食有-----」這類結構出現，況且「V有」結構的能產性非常強，如「做有、行有、睡有、橫有、追有、、、」，要把「V有」作為一個成分也是不合「詞是一個最小音義結合體」的定義的。

(3) 切分後所得的成分要具有意義上的搭配可能性和組合性，例如「阿公介皮箱送佢」可能的切分法有：

1、阿公 / 介皮箱送佢

2、阿公介 / 皮箱送佢

3、阿公介皮箱 / 送佢

按照第一種切分方法取得的各個成分，「阿公」有意義，可是「介皮箱送佢」沒有意義，這種切分是不合理的；按照第二種切分方法取得的各個成分，與原意相違背，不具意義上的搭配性和組合性，變成不是「阿公」的皮箱，只有第三種切分是合理的。

決定句子成分切分點的同時也決定了成分之間的關係，因此可以根據前面定義的二十二種關係進行分析與標記。

第七章 建立台灣客家語的剖析樹語料庫

剖析樹語料庫（通常叫 *treebank*），是將語句做句法分析所形成的集合，分析的過程同時定義了「詞、短語、句子」三個語法單位及各單位之間的層次關係，加上適當的標記所形成的集合。剖析樹語料庫的建置是語法研究中重要的一環，包括詞頻、短語頻率、短語結構、歧義分析、句型研究乃至語言之間的對譯，都具有相當迫切的需要性，特別是對台灣客家語的研究長期偏重語音描寫和個人語感的分析，具有相當重要的應用價值。

本章共分五小節，第一節 說明標記（tag）符號與規則，第二節 檔案格式說明，第三節 介紹 Dotted Tag，第四節 句法結構分析編輯工具，第五節 語料庫檢索工具，第六節 語料標記樣例，第七節 剖析樹的產生步驟與實例。

第一節 標記（tag）符號與規則說明

以上針對客家語的詞類、短語、句子分析方法分別做出定義和說明，接著我們便可以根據這些定義和分析方法對客家語的句子進行「標記」(tag)的工程，為使「標記」(tag)工程能順利進行，我們設計了一些操作軟體，以下分別加以說明：

以下是我們所使用的標記符號與標記規則說明：

| 符號 | 標示規則說明 | 示例 |
|--------|---|---------------------|
| n=" " | 代表語料別及序號 HAK：表客家語 HD800-46-02：表「現代漢語八百詞」的「頁碼」和「例句號」 | n="HAK_HD800-46-02" |
| hkw="" | 表該客家語語料例句的「關鍵字」 | hkw="你" |
| h=" " | 表「尚未加工的客家語語料生語料例句」 | h="你講麼介？俚聽毋清楚" |

| | | |
|---------|---|--|
| | | 哦！" |
| h.t=" " | 表「已加工的客家語語料例句」加工的深度包括：詞、短語、子句、標點符號、句法分析，分別標示詞類碼；短語性、短語結構；句法結構分析 | h.t="<你/pn 講麼介 /vp~vo.1vv.2pn>/cp~zu.1np.1v p ?/pu (佢/pn <聽/vv 毋清楚 /jp~zd.1rb.2jj>/vp~vc.1vp.1jp) /cp~zu.1np.1vp !/pu" |
| /xx | 表「某詞、某短語、子句或某標點符號的標記」 | 你/pn、講麼介/vp~vo.1vv.2pn |
| ~ | 用以區分短語性和短語結構 | 參考上例 |
| . | 用於斷詞 | 參考上例 |
| <> | 詞或短語的第一層次組合 | <你/pn 講麼介 /vp~vo.1vv.2pn> |
| () | 詞或短語的第二層次組合 | (佢/pn <聽/vv 毋清楚 /jp~zd.1rb.2jj>/vp~vc.1vp.1jp) |
| { } | 詞或短語的第三層次組合 | |

(表二十一，標記符號與標記規則)

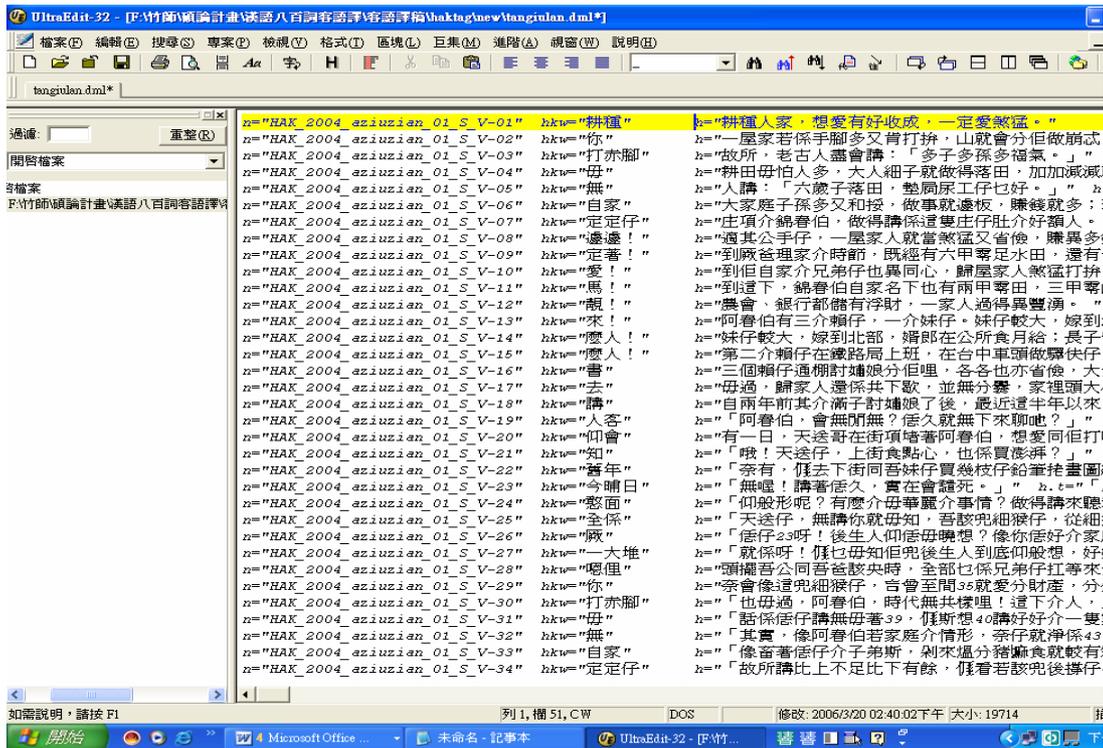
結構層次符號一共三個「<>」、「()」、「{ }」，原則上依層次標記，但有些結構大於三層時則依序重複使用。

第二節 檔案格式說明

這個語料庫建置所採用的檔案格式是.dml 格式，每一筆語料分為四個欄位，分別為「n=" "」、「hkw=" "」、「h=" "」、「h.t=" "」，相關的功能及說明請參考上表，以下我們以實例和圖一、圖二說明：

| 欄位序 | 實 例 |
|-----|--|
| 欄位一 | n="HAK_2004_aziuzian_01_S_V-01" |
| 欄位二 | hkw="耕種 |
| 欄位三 | h="耕種人家，想愛有好收成，一定愛煞猛。" |
| 欄位四 | h.t="耕種人家/np~dd.2nn.2nn ，/pu (想愛/vv <有/vv 好收成 /vp~zd.1rb.2vv>/vp~vc.1vp.1vp)/vp~vo.1vp.1vp ，/pu <一定/rb 愛煞 猛/jp~zd.1rb.2jj>/jp~zd.1rp.1jp 。/pu" |

(表二十二，語料庫檔案格式表)



(圖一 語料的欄位 1~3 格式)

構成剖析樹的可能結果就越少，而句子的文法剖析情形就越簡單。所以爲了減少這些單位數，我們利用 Dotted Tag 來處理。例如：「民族英雄」，我們用 Dotted Tag 表示則變成「民族英雄/ **np~dd.2nn.2nn**」，而不必是「<民族/nn 英雄/nn>/np~dd.1np.1np」。

(一)「Dotted Tag」標記規則說明

「Dotted Tag」一共由「短語性、~、短語結構、數字、詞類、」五個元素構成，例如：食飽了/vp~fg.2vv.1ax，「vp」表示這是一個動詞性短語；「~」用於連接「短語性和短語性」；「fg」表示這是一個附加短語；「2vv」表示這個短語裡的前兩個字（食飽）及其詞類爲「vv」；「1ax」表示接下來的一個詞是一個字及其詞類是「ax」。

再舉個例子，客家話的親屬詞像「若公、厥叔、吾爸」這種結構，是一個詞還是一個短語是有爭議的，我們可以看它在句子中的實際情形，以「若公在無？」這句爲例，以下分別是兩種分析的標記：

1、<若公/np~dd.1pn.1nn 在無/vp~fg.1vv.1ax>/cp~zu.1np.1vp ?/pu

2、<若公/nn 在無/vp~fg.1vv.1ax>/cp~zu.1np.1vp ?/pu

第一種標記方法是把「若公」視爲短語，第二種方法是把「若公」視爲詞，顯然第二種方法是比較簡潔的，而根據前面對「詞」所下的定義，若把「若公」視爲短語，會把「若」、「公」都變成語素，而不是詞，因爲客家話裡頭「若」、「公」是不能單用的語言成分，這麼一來便與詞的定義相違背了。

(二) 各語言單位之間的組合

我們定義客家語句子的語言單位是「語素、詞、短語、子句、句子」五級單位，本語料庫所標記的是「詞、短語、子句、句子」這四級單位，所以對句子的分析我們分析到「詞」這個層次，以下分別說明各級單位之間組合的表示法：

(1) 詞和詞的組合

詞和詞的組合構成各種結構的短語，已如前面所述，只要一個字串符合前面所規定的任一短語類型，則我們就將「詞和詞的組合」字串用短語結構去標記它。如：「上山打林/vp~vc.2vv.2vv」、「羊子食乳/cp~zu.2nn.2vv」、「跳下去/vp~vc.1vv.2vv」。

(2) 詞和短語的組合

詞和短語的組合產生較大層次的結構，我們採取第一級組合符號「<>」，如「<看/vv 羊子食乳/cp~zu.2nn.2vv>/vp~vo.1vp.1cp」。

上例這個結構被「<>」夾住，表示這是一個較大一級的組合，位於<>符號之後的結構表示法，仍然遵循我們所定義的短語表示法。

(3) 短語和短語的組合

短語和短語的組合產生更大層次的結構，我們採取二級組合符號「()」，如「(<鳥子出去/cp~zu.2nn.2vv 啣蟲轉來/vp~vc.2vv.2vv>/cp~zu.1cp.1vp <分鳥攔/vp~vo.1vv.2nn 食/vv>/vp~ld.1vp.1vp)/cp~zu.1cp.1vp」。

上例這個結構被「()」夾住，表示這是一個較大一級的組合，由一個「cp」

和一個「vp」組合，「cp」和「vp」的內部，還有其他的結構組合成。

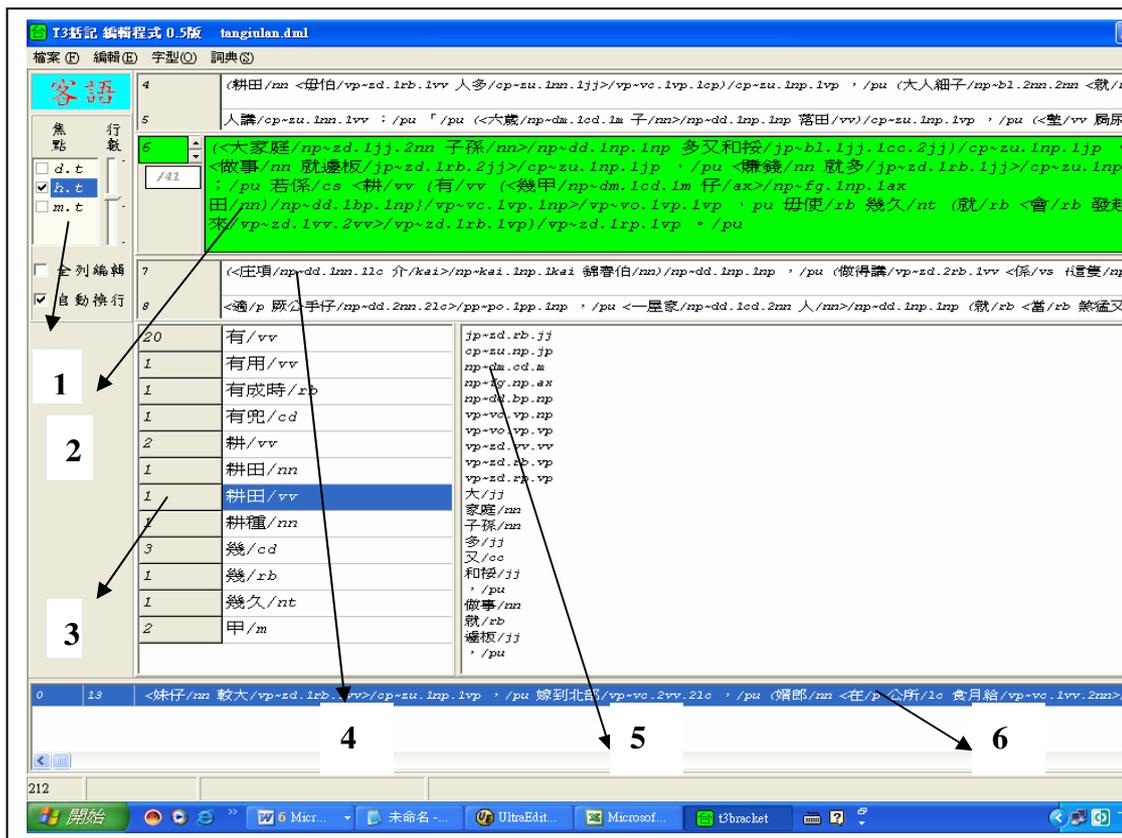
就這樣一層層的分析，一層層的組合上去，組成一個個的子句和一個個的句子。二級組合符號「()」用過之後，則使用三級組合符號「{ }」，當各級組合符號都用過之後，仍然不能標記完整個句子時，則重複交替使用。

第四節 句法結構分析編輯工具

我們設計了一個「T3bracket」編輯程式，該程式是由國立清華大學統計所江永進教授撰寫，此程式是用來標記詞類和分析句法，同時結合了電子詞典的功能，使我們在進行標記的過程中保持一致性，並且可以檢查斷詞、標記、句法分析的過程中是否有錯誤發生。

下圖三是「T3bracket」編輯程式的畫面，當我們把語料依據前面規劃好的四個欄位格式整理好，並存成.dml 檔案之後，就可以進入編輯程式進行句法樹剖析和標記工作，底下根據視窗的標號一一做說明。

- 1、是指語料樣本區，目前有「d.t」，代表閩南語；「h.t」，代表客家語；「m.t」，代表華語。當使用者要檢視哪個樣本時，只要挑選正確的樣本代號即可。
- 2、是指句法分析區，在這裡我們根據前面的各項定義和標記符號進行「詞、詞類、短語、子句、句子」等各及語法單位的分析和標記，分析和標記的結果會出現在第 5 區。



(圖三：T3bracket 編輯程式)

- 3、是詞頻顯示區，可以顯示每一個詞的詞頻。
- 4、是電子詞典區，標記過的詞及詞類會顯示出來，提供標記工作進行時參考與分析之用。
- 5、會顯示該筆語料分析的結果，包括短語結構和各各詞的詞類及分析狀況。
- 6、會顯示使用者所點選的關鍵詞所在的語料樣例，雙擊該語料樣例後會進入修正編輯區，使用者可以直接對該筆語料進行修正並存檔，圖四是修正編輯區。

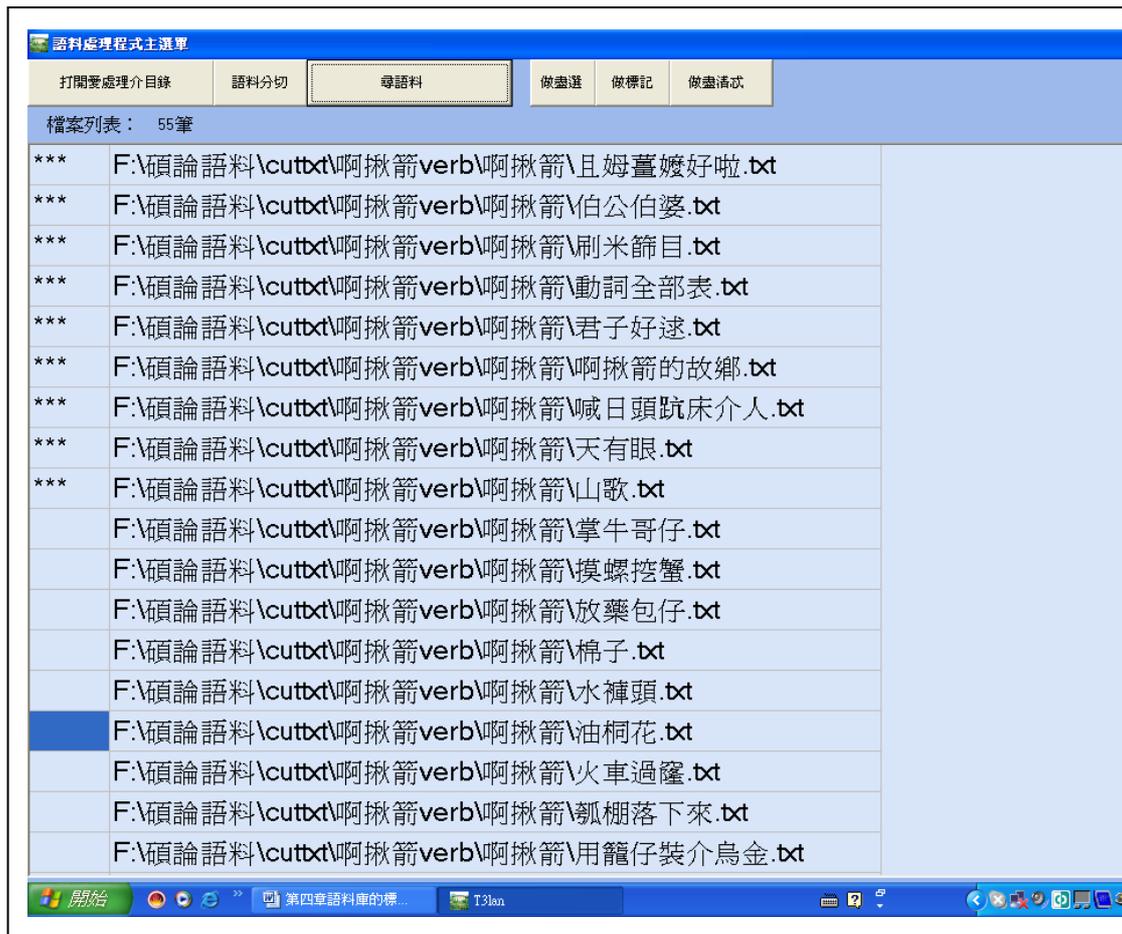


以上是關於編輯程式「T3bracket」的各項視窗和編輯功能說明，透過句法分析與標記過的語料為進一步檢索與分析，筆者另行設計了一個檢索工具「T3lan」，下一節針對這個檢索工具進行說明。

第五節 「T3lan」語料檢索工具

「T3lan」語料檢索工具是為因應對語料進一步研究和分析而設計的檢索軟體，雖然著重在檢索與統計功能，但同時也對語料標記的校正具有修正功能，以下是這個檢索軟體和運用方法的說明。

「T3lan」語料檢索工具可以同時對單一筆或多筆語料甚至是多個目錄裡的語料同時進行檢索，圖五是「T3lan」語料檢索程式主畫。



(圖五 語料檢索程式主畫)

不論是生語料或已標記的熟語料，使用者均可將語料存成文字檔(.txt)，即可進入 T3lan 程式進行標記和檢索與分析工作。切分的方式可以依使用者需要選擇「，、。；？！」幾種方式切分，也可以對已標記的語料或特別的符號進行切分，我們一般都是將語料以「。」切分，這是我們認為文本裡的「。」代表該語言使用者語意表達完整的句子，所以我們用「。」做切分點。

進入檢索畫面（如圖六），可以輸入「關鍵字或詞」檢索，同時也可以針對「兼類」情況進行檢索，例如「打」同時具有「動詞」和「量詞」兩種詞性，程式除可以讓使用者輸入關鍵字詞直接檢索，並將檢索結果全部顯示出來以供比對外，使用者也可以根據需要同時輸入「關鍵字或詞」和相應的「詞類」代碼，這

樣便能縮小範圍，取得更精確的檢索結果。

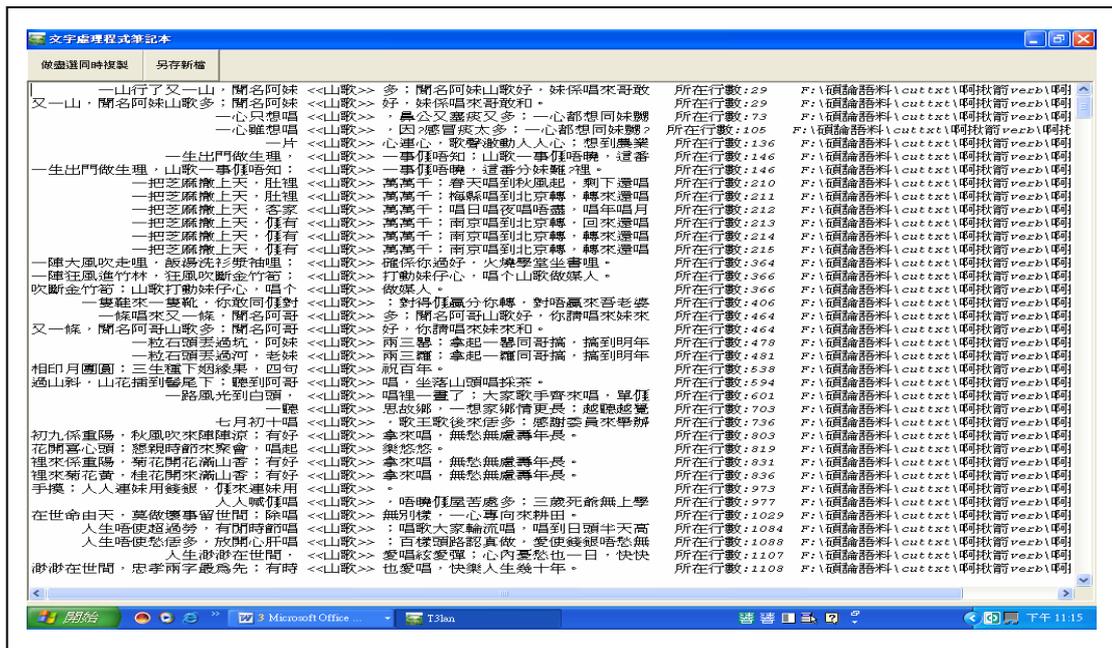


(圖六：T3lan 檢索程式檢視畫面)

檢索的結果顯示畫面（如圖七），我們以筆記本方式呈現，程式會將所檢索的「關鍵字或詞」用「<< >>」標示，並對齊，並顯示「關鍵字或詞」的資訊，包括所在的行數、檔案位置等，以方便研究者進行觀察和對比，同時使用者也可以複製到其他文書處理軟體進行其他編輯工作。

「輸入詞類或短語」搜尋（如圖八），是針對已標記「詞類」或短語及句法訊息的語料，不能用於生語料，您可以輸入相應的詞類或短語結構或句法訊息來進行檢索，例如研究者需要將語料中全部的「動詞」列出進行研究，則只需要打入動詞的代號「/vv」即可；若要檢索語料庫中的「述補短語」則只要打入「/vp~vc」即可，檢索的結果，我們一樣提供筆記本，同時具有統計功能，可以統計詞頻、

詞類統計報表、短語頻率等訊息。



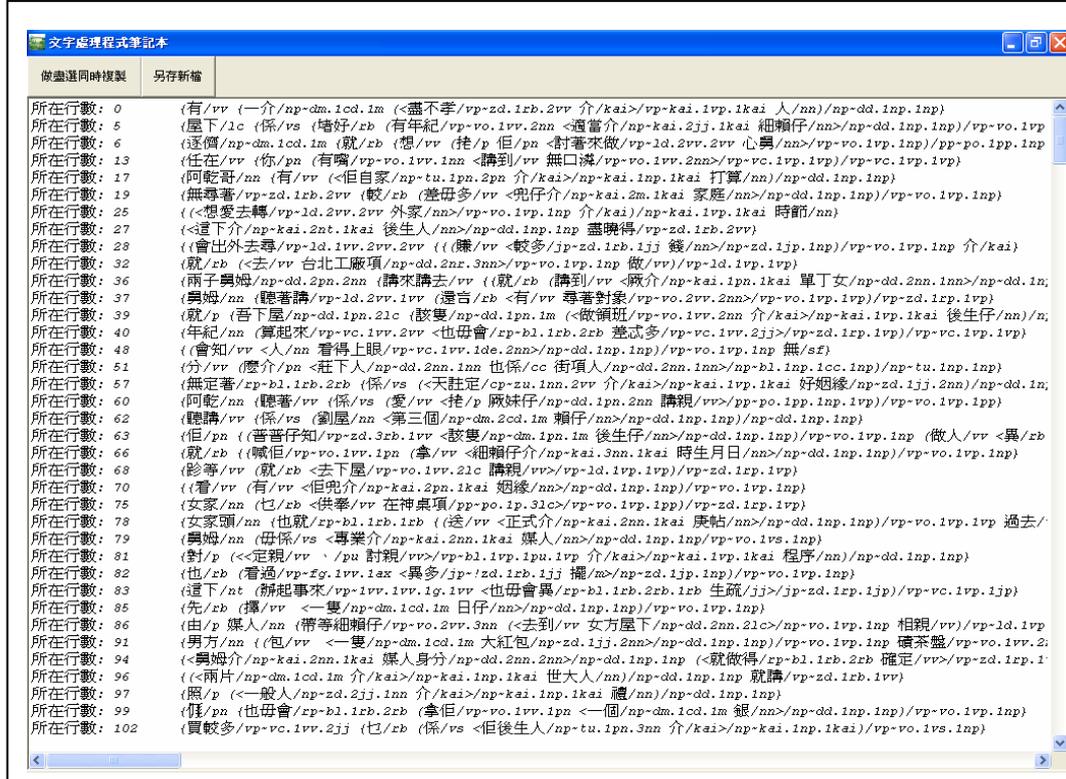
(圖七：T3lan 語料檢索結果顯示畫面)

| 序號 | 詞類名稱 | 標記符號 | 分類筆數 | 分類詞次 | 頻率 |
|----|-------|------|------|------|---------|
| 1 | 名詞 | /n | 413 | 810 | 28.581% |
| 2 | 專有名詞 | /nr | 3 | 3 | 0.1058% |
| 3 | 時間詞 | /nt | 25 | 32 | 1.1291% |
| 4 | 方位詞 | /lc | 31 | 53 | 1.8701% |
| 5 | 處所詞 | /si | 0 | 0 | 0.0352% |
| 6 | 數詞 | /cd | 20 | 94 | 2.9640% |
| 7 | 序數詞 | /od | 0 | 0 | 0.0352% |
| 8 | 量詞 | /m | 47 | 117 | 4.1284% |
| 9 | 代詞 | /pn | 14 | 61 | 2.1524% |
| 10 | 動詞 | /vv | 348 | 786 | 27.734% |
| 11 | 形容詞 | /jj | 62 | 135 | 4.7635% |
| 12 | 狀態詞 | /zht | 10 | 11 | 0.3881% |
| 13 | 區別詞 | /dis | 3 | 3 | 0.1058% |
| 14 | 副詞 | /zb | 90 | 448 | 15.808% |
| 15 | 介詞 | /pa | 0 | 0 | 0.0352% |
| 16 | 連接詞 | /cc | 18 | 43 | 1.5172% |
| 17 | 從屬連接詞 | /cs | 0 | 0 | 0.0352% |

(圖八：T3lan 詞類統計報表介面)

當研究者需要了解句法的層次結構時，可以依據我們現有的句法層次標記

「<>、()、{ }」三者來進行搜尋（如圖九），研究者也可以自訂其他的語法層次標記，本程式也提供研究者自訂的語法層次標記搜尋模式，以符合個人的需求。



(圖十：T3lan 語法層次檢索)

第六節 語料標記樣例

經過以上的詞類規劃、短語標記規劃、句法層次分析，我們就可以開始來進行語料標記工作，標記的工作分為詞類標記和句法層次標記，這僅需要一次在 T3bracket 編輯程式上即可完成，分析的結果可以取出分為詞類標記檔和句法層次結構檔兩種，以下是新屋客家語故事集的文章經過分析標記的結果。

(一) 詞類標記樣例：

H_01 頭擺/nt ,/pu 有/vv 一/cd 介/m 盡/rb 不孝/vv 介/kai 人/nn ,/pu 每
/cd 日/m 上山/vv 打林/vv ,/pu 逐/cd 日/m 愛/vv 厥姆/pn 攞/vv
飯/nn 分/vv 佢/pn 食/vv 。/pu

H_02 忒早/rb 去/vv 分/pa 佢/pn 打/vv ,/pu 忒慢/rb 去/vv 又/rb 分/pa
佢/pn 打/vv 。/pu

H_03 有/vv 一/cd 日/m ,/pu 歇暍/vv 時/nn ,/pu 看/vv 羊子/nn 食
乳/vv ,/pu 四/cd 向/nn 跪拜/vv 好/jj 正/rb 食/vv 。/pu

H_04 又/cc 一/cd 擺/m ,/pu 看著/vv 鳥竇/nn 肚/lc ,/pu 鳥攞/nn 啣蟲
/vv 分/vv 鳥子/nn 食/vv ,/pu 等著/vv 鳥子/nn 大/vv 咧/sf ,/pu 鳥
攞/nn 毛/nn luk20/vv 忒/ax ,/pu 毋會/rb 飛去/vv 討食/vv ,/pu 鳥
子/nn 出去/vv 啣/vv 蟲/vv 轉來/vv 分/vv 鳥攞/nn 食/vv 。/pu

H_05 看著/vv 這/pn 情形/nn ,/pu 感覺/vv 自家/pn 毋/rb 著/vv ,/pu 平
時/nn 自家/pn 對/pa 阿姆/nn 不孝/vv 。/pu

H_06 一/cd 日/m 咧/sf ,/pu 轉想/vv 過來/vv ,/pu 看著/vv 阿姆/nn
kun11/vv 飯/nn 來/vv ,/pu 愛/rb 去/vv 接/vv 。/pu

H_07 阿姆/nn 看著/vv 丁蘭/nn 來/vv ,/pu 驚/vv 又/rb 分/pa 佢/pn 打
/vv ,/pu 跳落/vv 唇頭/lc 介/kai 深潭窟/nn 。/pu

H_08 丁蘭/nn ten55 著/cc 也/rb 跳/vv 下去/vv ,/pu 撈/vv 當/rb 久/jj
撈無/vv 。/pu

H_09 撈到/vv 一/cd 埕/m 枋仔/nn ,/pu 轉去/vv 刻/vv 厥姆/pn 介/kai
名/nn ,/pu 逐/cd 日/m 拜/vv ,/pu 這/pn 就係/vv 神主牌/nn 介/kai
由來/nn 。/pu

(二) 句法層次標記樣例

h.t="頭擺/nt ，/pu {有/vv {一介/np~dm.1cd.1m (<盡不孝/vp~zd.1rb.2vv 介
 /kai>/np~kai.1vp.1kai 人
 /nn)/np~dd.1np.1np}/np~dd.1np.1np}/vp~vo.1vp.1np ，/pu <每日
 /np~dm.1cd.1m 上山打林/vp~vc.2vv.2vv>/cp~zu.1np.1vp ，/pu {逐日
 /np~dm.1cd.1m (<愛厥姆/vp~vo.1vv.2pn kuan11 飯
 /vp~vo.1vv.1nn>/vp~ld.1vp.1vp <分佢/vp~vo.1vv.1pn 食
 /vv>/vp~ld.1vp.1vp)/vp~ld.1vp.1vp}/cp~zu.1np.1vp 。/pu "

h.t="忒早去/vp~zd.2rb.1vv 分佢打/pp~po.1pa.1pn.1vv ，/pu 忒慢去/vp~zd.2rb.1vv
 又/cc 分佢打/pp~po.1pa.1pn.1vv 。/pu "

h.t="<有/vv 一日/np~dm.1cd.1m>/vp~vo.1vp.1np ，/pu 歇暍時/vp~vo.2vv.1nn ，
 /pu <看/vv 羊子食乳/cp~zu.2nn.2vv>/vp~vo.1vp.1cp ，/pu 四向
 /np~dm.1cd.1nn 跪拜好/vp~vc.2vv.1jj 正食/vp~zd.1rb.1vv 。/pu "

h.t="又/cc 一擺/np~dm.1cd.1m ，/pu <看著/vv 鳥糞肚
 /np~dd.2nn.1lc>/vp~vo.1vp.1np ，/pu (<鳥攞/nn 啣蟲
 /vp~vo.1vv.1nn>/cp~zu.1np.1vp <分鳥子/vp~vo.1vv.2nn 食
 /vv>/vp~ld.1vp.1vp)/cp~zu.1cp.1vp ，/pu (等著/vv <鳥子/nn 大咧
 /vp~fg.1vv.1sf>/cp~zu.1np.1vp)/vp~vc.1vp.1cp ，/pu <鳥攞毛/np~dd.2nn.1nn
 luk20 忒/vp~fg.1vv.1ax>/cp~zu.1np.1vp ，/pu <毋會/rb 飛去討食
 /vp~ld.2vv.2vv>/vp~zd.1rp.1vp ，/pu (<鳥子出去/cp~zu.2nn.2vv 啣蟲轉來
 /vp~vc.2vv.2vv>/cp~zu.1cp.1vp <分鳥攞/vp~vo.1vv.2nn 食
 /vv>/vp~ld.1vp.1vp)/cp~zu.1cp.1vp 。/pu "

h.t="<看著/vv 這情形/np~dd.1pn.2nn>/vp~vo.1vp.1np ，/pu (感覺/vv <自家/pn 毋
 著/vp~zd.1rb.1vv>/cp~zu.1np.1vp)/vp~vc.1vp.1cp ，/pu 平時/nn 自家/pn 對
 阿姆不孝/pp~po.1p.2nn.2vv 。/pu "

h.t="<一日/np~dm.1cd.1m 咧/sf>/np~fg.1np.1sf ，/pu 轉想過來/vp~vc.2vv.2vv ，
 /pu (看著/vv <阿姆/nn kun11 飯來

/vp~ld.2vv.1vv>/cp~zu.1np.1vp)/vp~vo.1vp.1cp ， /pu <愛/rb 去接

/vp~ld.1vv.1vv>/vp~zd.1rp.1vp 。 /pu"

h.t="(阿姆/nn <看著/vv 丁蘭來/cp~zu.2nn.1vv>/vp~vo.1vp.1cp)/cp~zu.1np.1vp ，

/pu 驚/vv 又/rb 分佢打/pp~po.1pa.1pn.1vv ， /pu (跳落/vv <唇頭介

/np~kai.2lc.1kai 深潭窟/nn>/np~dd.1np.1np)/vp~vo.1vp.1np 。 /pu"

h.t="丁蘭/nn ten55 著/cc <也/rb 跳下去/vp~vc.1vv.2vv>/vp~zd.1rp.1vp ， /pu (<撈

/vv 當久/jp~zd.1rb.1jj>/vp~vc.1vp.1jp 撈無/vv)/vp~vc.1vp.1vp 。 /pu"

h.t="(撈到/vv <一埕/np~dm.1cd.1m 枋仔/nn>/np~dd.1np.1np)/vp~vo.1vp.1np ， /pu

(轉去刻/vp~ld.2vv.1vv <厥姆介/np~kai.2pn.1kai 名

/nn>/np~dd.1np.1np)/vp~vo.1vp.1np ， /pu <逐日/np~dm.1cd.1m 拜

/vv>/cp~zu.1np.1vp ， /pu 這/pn 就係/cc <神主牌介/np~kai.3nn.1kai 由來

/nn>/np~dd.1np.1np 。 /pu"

第七節 剖析樹的產生步驟與實例(Parse Tree)

國立新竹教育大學
National Hsinchu University of Education

經過上面的各種規劃、分析和標記之後，我們便可以得到一棵棵的句法剖析樹 (Parse Tree)，下面我們以「佢看著鳥竇肚介鳥子。」這句話為例，逐步說明標記的過程。

步驟一：

首先當研究者拿到一筆語料後，如「佢看著鳥竇肚介鳥子。」，先要根據「詞」的定義決定切分點，將「詞」這一級的語法單位分析出來，上例分析結果是：

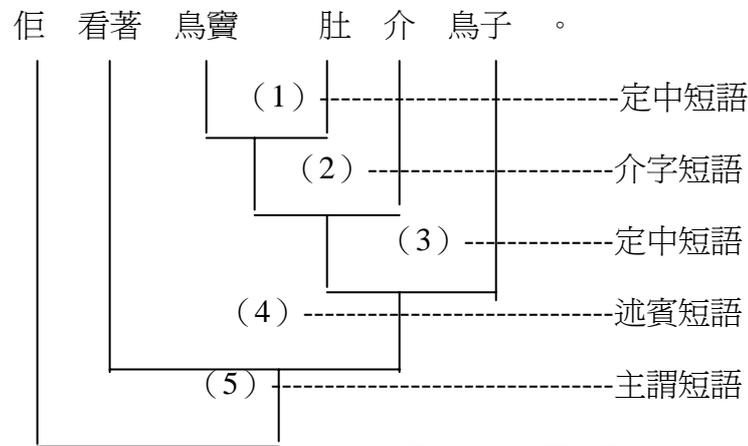
佢 看著 鳥竇 肚 介 鳥子 。

上例一共具有六個詞和一個標點符號。

步驟二：

「詞」這一級的語法單位分析出來後，接著便要決定「詞和詞」之間的關

係，並將「詞和詞」化爲短語結構，上例短語分析結果是：



步驟三：

根據分析的結果，我們便逐層進行標記，上例標記的結果是

{佢/pn {看著/vv (<烏竇肚/np~dd.2nn.1lc 介/kai>/np~kai.1np.1kai 鳥子
 /nn)/np~dd.1np.1np}/vp~vo.1vp.1np}/cp~zu.1np.1vp 。/pu

用 Dotted Tag 和各種括號表示是給電腦方便處理，但是當括號越來越多時，一般人並不容易看懂。我們可以把這些標記好的語料畫成如上的樹圖 (parse tree)，便於觀看其中的層次關係。

第八章 「T3」語料庫語料統計與分析

「T3」語料庫，目前客家語部份計有 432,187 個漢字和 782 個有音無字或目前無法輸出的標音字，總計 432,969 個字已完成詞類標記，共得 181,172 個詞次，平均 2.389 個漢字就構成一個詞，這個數據表明客家話的構詞是趨向雙音節化的，並不是單音節取向。雖然這和我們對「詞」的定義有關，但至少在我們的定義下，這個數據是指向雙音節化的結論。

其次我們共完成了 162,347 個漢字（含標音字）文本的句法結構標記，並完成 432969 個漢字詞類標記，以下我們僅針對對這些已完成的語料的詞類和短語進行一些統計分析。

本章分為二小節，第一節 各詞類使用頻率分析，第二節 短語統計與分析。

第一節 各詞類使用頻率分析

運用「T3」語料庫，作者在寫本文時計有 181172 個詞次，總計 32619 個詞，平均每個詞出現的頻率是 5.55 次，以下我們根據這 32619 個詞進行詞類分佈統計，結果如下表：

| 序號 | 詞類名稱 | 標記符號 | 分類個數 | 百分比 |
|----|------|------|-------|--------|
| 1 | 名詞 | /nn | 21334 | 64.43% |
| 2 | 動詞 | /vv | 6764 | 20.43% |
| 3 | 形容詞 | /jj | 2618 | 7.91% |
| 4 | 副詞 | /rb | 621 | 1.88% |
| 5 | 區別詞 | /dis | 123 | 0.37% |
| 6 | 狀態詞 | /zht | 158 | 0.48% |
| 7 | 量詞 | /m | 136 | 0.41% |

| 序號 | 詞類名稱 | 標記符號 | 分類個數 | 百分比 |
|----|-------|------|------|-------|
| 8 | 時間詞 | /nt | 187 | 0.56% |
| 9 | 方位詞 | /lc | 123 | 0.37% |
| 10 | 代詞 | /pn | 136 | 0.41% |
| 11 | 數詞 | /cd | 104 | 0.31% |
| 12 | 專有名詞 | /nr | 89 | 0.27% |
| 13 | 處所詞 | /si | 73 | 0.22% |
| 14 | 連接詞 | /cc | 24 | 0.07% |
| 15 | 介詞 | /pa | 52 | 0.16% |
| 16 | 助詞 | /ax | 16 | 0.05% |
| 17 | 擬聲詞 | /on | 24 | 0.07% |
| 18 | 句末語氣詞 | /sf | 15 | 0.05% |
| 19 | 感嘆詞 | /ij | 16 | 0.05% |
| 20 | 序數詞 | /od | 6 | 0.02% |

國立新竹教育大學 (表二十三, 詞類分佈統計)
National Hsinchu University of Education

上面這個統計表是根據我們現有且已做好詞類標記的語料量來進行統計的，實際語言裡的數據當更大於這些數據，例如「口罩」，有人說「口落仔」或有人說「喺落仔」，這並未出現在語料庫裡，當然這和我們的語料樣本有關，又如擬聲詞「嘻嘻滾、咻咻滾」等，這些也都沒有出現，雖然如此，但已可看出一個大體的趨勢，同時也表示我們的語料量還不夠大、樣本仍不夠全面，有待繼續努力。

大體而言，名詞佔最大一部份，有 64.43%強，其次是動詞佔 20.43%，兩者合計就佔去了 84.86%強，再次是形容詞佔 7.91%，其他各類詞僅佔非常小的比例。

第二節 短語統計與分析

| 序號 | 短語性標記 | 短語標記 | 短語名稱 | 數量 | 百分比 | 短語性百分比 |
|----|-------|------|-----------|------|--------|--------|
| 1 | cp | Zu | 主謂短語 | 4328 | 10.20% | 10.20% |
| 2 | vp | vo | 述賓短語 | 8746 | 20.61% | 52.74% |
| 3 | | vc | 述補短語 | 4227 | 9.96% | |
| 4 | | ld | 連動短語 | 674 | 1.59% | |
| 5 | | zd | 狀中短語 | 4653 | 10.96% | |
| 6 | | bl | 聯合短語 | 214 | 0.50% | |
| 7 | | fg | 附加短語 | 3872 | 9.12% | |
| 8 | | np | bl | 聯合短語 | 547 | |
| 9 | fu | | 方位短語 | 48 | 0.11% | |
| 10 | dm | | 數量短語 | 2374 | 5.59% | |
| 11 | tu | | 同位短語 | 234 | 0.55% | |
| 12 | dd | | 定中短語 | 5479 | 12.91% | |
| 13 | kai | | 介(kai)字短語 | 3264 | 7.69% | |
| 14 | ts | | 時間短語 | 437 | 1.03% | |
| 15 | ss | | 處所短語 | 234 | 0.55% | |
| 16 | jp | zd | 狀中短語 | 1324 | 3.12% | 3.57% |
| 17 | | bl | 聯合短語 | 48 | 0.11% | |
| 18 | | kai | 介(kai)字短語 | 68 | 0.16% | |
| 19 | | jc | 形補短語 | 54 | 0.13% | |
| 20 | | pk | 比況結構 | 23 | 0.05% | |
| 21 | pp | pb | 介詞短語 | 1352 | 3.19% | 3.19% |

| 序號 | 短語性標記 | 短語標記 | 短語名稱 | 數量 | 百分比 | 短語性百分比 |
|----|-------|------|------|-----|-------|--------|
| 22 | rp | bl | 副詞短語 | 243 | 0.57% | 0.57% |

(表二十四 , 短語分佈統計)

根據上表統計結果來分析，可知客家語的使用者，主要使用的是「動詞性短語」，佔 52.74%，其中又以述賓 (vo)、述補 (vc)、狀中 (zd) 和附加 (fg) 為主，連動和聯合短語相對偏低，這和動詞的分布是相呼應的，雙音節動詞中也是以述賓和述補式構詞為主的。這樣的分布意義顯示客家語的動詞是以帶賓語、補語和附加時態標記及受狀語修飾等為主要的語法功能，而連動和聯合只是動詞或動詞短語的一種變化運用方式，並不是主要的構句方式。

就名詞性短語來看，佔 29.72%，其中以定中 (dd)、介 (kai) 字短語和數量短語 (dm) 為主，這三者也和名詞的主要語法功能相呼應。

就形容性短語來看，佔 3.57%，以狀中短語為主，佔 3.12%，顯示形容詞以主要受狀語修飾為主。

第九章 客家語動詞研究

本章我們將語料庫中所有客家語的動詞，抽取出來，進行統計與分析，抽取出來的結果是具有不平衡性的，並不是每一個詞都會在四縣或海陸的語料檔案中出現，這和語料的量及取樣有關，我們現有的語料量是四縣客家語大於海陸客家語，在總計 432,969 個字已完成詞類標記的語料中，有 264,317 個字的語料我們歸入四縣客家語，而 168,652 個字的語料我們歸入海陸客家語，相差約 100,000 萬字，本章我們抽取語料庫中的「動詞」（包含係動詞），再配合口語及田野調查做對比與分析。¹

對客家語動詞的研究，羅肇錦（1984、1999）最為詳盡，根據音節多寡的角度將客家語動詞分為「單音動詞和複音動詞」兩類；依據動作的性狀分及物動詞、不及物動詞和助動詞三部份，羅肇錦更進一步提出：詞序相反、前加成份、後加成份、內部曲折、代詞結構、動詞重疊、形容詞等級、量詞用法等八個部份來說明客家語構詞的特點。本章則根據語料分析所得，我們著重在對客家語動詞的特點分析和特點的量化上，並和華語、閩南語及客家語周邊語言進行一些比較和分析，從中找出客家語動詞的特點。

本章分為五小節，第一節 動詞的音節統計與分析，第二節 動詞的構詞形式，第三節 客家語的核心動詞；第四節 四縣客家語和海陸客家語動詞對比分析，第五節 華語、客家語動詞對比分析。

第一節 動詞的音節統計與分析

根據現有已標記的語料抽取出來的結果，單音節動詞計 1,299 個，佔全部動

¹ 這些是指已標記好的語料而言，並不是目前語料庫裡全部的語料量。

詞的 19.20%；雙音節動詞計 5,264 個，佔全部動詞的 77.82%；三音節動詞計 127 個，佔全部動詞的 1.88%；四音節動詞計 74 個，佔全部動詞的 1.10%，這個數據表明客家語的動詞是雙音節化的居多。由於大部分雙音節動詞都是由單音節的動詞和另一成分，以一定的方式組合而成的複音詞，同時限於篇幅的因素，以下我們僅列出全部所得到的單音節動詞和部分非由單音節動詞為語素構成的雙音節動詞。

就使用頻率來看，單音節詞動詞的使用總頻率 33808 次，使用頻率為 66.89%；雙音節動詞的使用總頻次 15882 次，使用頻率為 31.51%；三音節動詞的使用總頻次為 591 次，使用頻率為 1.17%；四音節動詞的使用總頻次為 218 次，使用頻率為 0.43%。雖然雙音節動詞的數量遠大於單音節數量近 5 倍之多，但就使用頻率看單音節動詞的使用頻率遠高於雙音節動詞達 2 倍之多，三、四音節動詞的使用頻率相對偏低。

根據分析結果，我們製成下表，以利分析和比較：

| | 詞數（個） | 佔全部動詞比 | 使用頻次（次） | 使用頻率比 |
|-------|-------|--------|---------|--------|
| 單音節動詞 | 1299 | 19.20% | 33808 | 67.07% |
| 雙音節動詞 | 5264 | 77.82% | 15882 | 31.51% |
| 三音節動詞 | 127 | 1.88% | 591 | 1.17% |
| 四音節動詞 | 74 | 1.10% | 123 | 0.25% |
| 總計 | 6764 | 100% | 50404 | 100% |

(表二十五，動詞音節統計)

一般客家語的研究者多認為客家語的單音詞多於華語，但實際上多出多少？或單音詞是哪些？分布情形如何？為什麼客家語會有較多的單音詞？這個

結論的真偽是否禁得起檢驗等問題，長期以來沒有得到解答。亢世勇（2004）針對 43,224 個華語詞彙進行的統計分析結果是單音節詞共 2,479 個，在 10,283 個動詞中，單音動詞 1208 個；雙音節動詞 8,918 個；三音節動詞 157 個，並統計出單音節詞中動詞的數量遠大於其他各類詞，佔絕對優勢，佔所有單音節詞的 48.73% 強，佔全部動詞的 11.75%。

在客家語的統計方面，單就動詞這個詞類而言，單音動詞有 1278 個，以目前我們的描述體系來看的多於華語的單音節動詞，而且從整個動詞類的音節方面也可看出華語和客家語的動詞都是以雙音節為主的。

第二節 動詞的構詞形式

客家語動詞的構成，除單音節動詞由單一語素構成之外，對於客家語動詞構詞的形式，羅肇錦（1984）提出雙聲疊韻、並列複合、主從複合、主謂複合、動賓複合、動補複合六種來說明，這幾種構詞形式都是屬於雙音複合動詞結構，其中「雙聲疊韻」屬修辭格式，與其他四種並不在同一平面，故筆者略去「雙聲疊韻」的分析，而根據我們在語料庫中的定義，並參考其他語法書的看法，就構詞形式來看客家語的雙音節動詞，我們將客家語的雙音節動詞分為主謂式、述賓式、聯合式、述補式、偏正式、重疊式這五種來加以說明和統計，同時就能力所及，筆者用亂數表採樣選出雙音節動詞 1,000 個進行分析，雖然與全部雙音節動詞 5264 個差距很遠，但這已是目前能力所極了，留待日後繼續努力完成。

(1) 主謂式雙音節動詞

由主語和謂語複合而成的雙音節動詞，一般主語由名詞性構成，謂語則由動詞性或形容性成分構成，我們所考察的 1,000 個雙音節動詞中，只有三例被歸入主謂式複合，「天光、肚渴、酒醉」，比例是非常少的，佔 0.03%。

(2) 述賓式雙音節動詞

由一個述語成分（動詞）和賓語成分（名詞）構成的雙音節動詞，這種述賓式雙音節動詞，多半可以變成離合詞，如：「煮菜、攪水、泡茶、敬神、供豬、畜雞、嘍狗、挖泥、發病、獵雞、斬頭、添飯」等，以上各例都可以變成離合詞，如「煮一盤菜、攪一桶水、泡三兩茶、敬過神、供過豬、畜當多雞、嘍一擺狗、挖一窟泥、發一擺病、獵三隻雞、斬二只頭、添一碗飯」等。

但根據我們的分詞標準定義，凡是述賓結構的 V-N 之間，如果沒有中插成分，則視 V-N 為一個分詞單位。所以我們會發現述賓結構的雙音節動詞數量非常龐大，列舉如下：

「淋水、涿雨、唸喙、牽魚、岐毛、岐皮、賴人、賸田、賸屋、碰門、入屋、入屋、分爨、出桌、叩頭、打胎、打索、打魚、吊頸、扛茶、扛飯、瀉屎、搞水、搭話、搶光、煖酒、睺人、落肉、落嘴、出丁、割禾、犁田、逕人、洗身、蕩喙」等，不勝枚舉。

我們所考察的 1,000 個雙音節動詞中計有 458 個「述賓式結構」的動詞，佔 45.8%。

(3) 述補式雙音節動詞

由一個述語成份（動詞）和補語成份（形容詞或動詞）構成的雙音節動詞，這種述補式雙音節動詞，多半也可以被其他成分中插，變成離合詞，如：

「食飽、掙淨、淳出、趨走、拿來、落去、趨出、煮熟、打走、切幼、吊死」等，以上各例都可以變成離合詞，如「食毋飽、掙毋淨、淳毋出、趨毋走、拿得來、落毋去、趨毋出、煮毋熟、打毋走、切毋幼、吊毋死」等。

根據我們的分詞標準定義，凡是述補結構的 V-R 之間，如果沒有中插成分，則視 V-R 為一個分詞單位。所以我們會發現述補結構的雙音節動詞數量也非常龐大，能產性也非常之高，列舉如下：

「掙開、吹開、扯開、擊開、避開、奮開、搥開、收淨、洗淨、舐淨、祛淨、搓淨、撚出、暴出、讓出、鑽出、看差、做差、賴差、抹平、耙平、掌平、楔平、糊平、扼死、啄死、翕死、闕死、譴死、tak20 弦、弓弦、攪礮、撈礮、撇礮、透涼、撥涼、遮涼、燎涼、轉涼、放恬、煞恬、頓恬、擋恬、^ㄉ來、杯來、剝來、渡落、跌落、敵落、跳落、跪落、驕落、炙燥、晒燥、焙燥、燥透、淋透、結透、撈透、捫爛、擊爛、捶爛、走醒、喊醒、歎冷」。

我們所考察的 1,000 個雙音節動詞中計有 372 個「述補式結構」的動詞，佔 37.2%。

(4) 聯合式雙音節動詞

兩個語素之間的關係是平等的，由意義項同相反相對的成分並列所形成的

詞，語素之間不能顛倒。如「出入、承蒙、收留、跪拜、發作、省儉、施捨、選舉、聯想、應承、遶療、擔保、遷徙」等。

我們所考察的 1,000 個雙音節動詞中計有 83 個「聯合式結構」的動詞，佔 0.83%。

(5) 偏正式雙音節動詞

兩個語素之前一個語素限制或修飾後一個語素，詞的意義以後一個語素為主，前一個語素只起附加作用，前為偏，後為正的合成詞。如「轉來、賺食、折算」，我們所考察的 1,000 個雙音節動詞中計有 12 個「偏正式結構」的動詞，佔 0.12%。

(6) 重疊式雙音節動詞

指由兩個單音節語素重疊而成的雙音節動詞，客家語的單音動詞可重疊，重疊的方式也很多種，這裡我們指的是 AA 式的重疊雙音節動詞，如「看看、行行、食食、掌掌、擗擗、杯杯、粹粹、拿拿、副副、灑灑、豁豁、祛祛、捋捋、離離、咬咬、播播、舞舞、搥搥、fung11fung11」，重疊式雙音節動詞就語意來看仍是動詞，這裡特別說明的是，我們所考察的 1,000 個雙音節動詞中計有 72 個重疊式結構雙音節動詞，佔 0.72%，這些重疊式雙音節動詞，除「看看、行行」外，其他全部是進入「VV 忒」的格式。

| 構詞形式 | 詞數 | 百分比 | 詞例 |
|-------|-----|-------|----------|
| 主謂式複合 | 3 | 0.03% | 天光、喙燥、酒醉 |
| 述賓式複合 | 458 | 45.8% | 扛飯、搞水、搭話 |
| 聯合式複合 | 83 | 0.83% | 應承、遶療、擔保 |
| 述補式複合 | 372 | 37.2% | 舐淨、捫爛、撚出 |
| 偏正式複合 | 12 | 0.12% | 轉來、賺食、折算 |
| 重疊式複合 | 72 | 0.72% | 孛孛、擗擗、抔抔 |

(表二十六，雙音節動詞構詞統計)

這個分析結果顯示，客家語的雙音動詞以述賓式和述補式最為普遍，能產性也最強，主謂式複合動詞的比率相對非常的少，這是因為主謂式複合動詞很容易加上適當的語氣之後實現為句子，而主謂式複合動詞是否成詞問題一直頗有爭議，我們處理的方式則是看它在句子中的語法現象來決定，例如：

A、天光了

B、還言天光。

國立新竹教育大學
National Hsinchu University of Education

這兩個例句都有「天光」的字串，例 A 的主語是「天」，動詞是「光」，例 A 是一個句子，例 A 可以插入否定副詞「言」，構成「天言光」，這是一個主謂句；而例 B「天光」則是動詞，受否定副詞「言」修飾，副詞「言」不能移動構成「*還天言光」，所以類似這種主謂式複合的字串，必須根據在句子中的語法現象來決定該字串是否成詞。

第三節 客家語的核心動詞

筆者針對語料庫中的每一種文本的動詞進行統計，選出每一個文本中出現頻次最高的動詞 100 個進行對比分析，發現 56 個動詞在任一個文本中均出現，這 56 個動詞僅佔全部 32619 個詞的 0.17%，但實際使用出現的次數卻高達 17607 次，佔全部動詞使用詞次的 34.93%，相當於說客家語言的使用者在使用動詞時有 34.93% 的機率會使用到這 56 個動詞，我們把這樣的情形稱為「核心動詞」。這 56 個動詞是：

| 序號 | 詞例 | 詞頻 | 詞義 |
|----|-----|------|-----------|
| 1 | 講 1 | 2183 | 說 |
| 2 | 有 1 | 1013 | 表領有、具有 |
| 3 | 去 | 876 | 前往 |
| 4 | 看 1 | 775 | 注視、觀賞 |
| 5 | 來 | 666 | 與「去」相反 |
| 6 | 想 4 | 624 | 思憶、思念 |
| 7 | 做 1 | 612 | 進行某種工作或行爲 |
| 8 | 食 1 | 601 | 吃 |
| 9 | 分 1 | 551 | 給 |
| 10 | 無 | 511 | 缺少、沒有 |
| 11 | 像 | 469 | 相似 |
| 12 | 喊 | 397 | 呼叫 |
| 13 | 愛 1 | 393 | 要 |
| 14 | 到 1 | 381 | 前往 |
| 15 | 用 1 | 370 | 使用 |

| | | | |
|----|-----|-----|---------------|
| 16 | 係 | 361 | 是 |
| 17 | 知 | 351 | 洞悉、了解 |
| 18 | 行 | 342 | 走 |
| 19 | 噉 | 319 | 哭 |
| 20 | 問 1 | 278 | 請人解答 |
| 21 | 著 1 | 250 | 穿戴 tsok20 |
| 22 | 打 1 | 224 | 擊 |
| 23 | 賣 | 222 | 出售 |
| 24 | 罵 | 204 | 用不好聽的話指責或侮辱別人 |
| 25 | 轉 1 | 197 | 回去 |
| 26 | 走 1 | 194 | 跑 |
| 27 | 聽 1 | 193 | 用耳朵收受聲音 |
| 28 | 過 1 | 184 | 時、空的經歷 |
| 29 | 坐 2 | 183 | 搭乘 |
| 30 | 坐 1 | 182 | 臀部依附在物體上 |
| 31 | 等 | 178 | 待、候 |
| 32 | 拿 1 | 175 | 提供 |
| 33 | 想 1 | 175 | 思索、思考 |
| 34 | 寫 | 167 | 用筆書寫 |
| 35 | 睡 | 157 | 閉著眼睛休息 |
| 36 | 出 1 | 154 | 由內向外 |
| 37 | 拿 2 | 154 | 持住 |
| 38 | 買 | 146 | 以金錢換取物品 |
| 39 | 請 1 | 146 | 懇求、乞求、希望對方做某事 |
| 40 | 驚 | 145 | 心中害怕而不安 |

| | | | |
|----|-------|-----|------------------|
| 41 | 死 | 143 | 去世 |
| 42 | 倚 | 140 | 站立 |
| 43 | 落 | 136 | 掉下、降 |
| 44 | 放 1 | 135 | 解脫、鬆開 |
| 45 | 裝 | 132 | 置入 |
| 46 | 尋 | 125 | 找 |
| 47 | loi53 | 124 | 去 |
| 48 | 做 3 | 123 | 當、身為 |
| 49 | 笑 1 | 121 | 因內心欣喜而露出快樂的表情或聲音 |
| 50 | 好 1 | 109 | 恢復健康 |
| 51 | 帶 | 107 | 隨身攜著、拿著 |
| 52 | 愁 | 105 | 憂慮、擔心 |
| 53 | 會 | 102 | 能 |
| 54 | 刷 | 102 | 宰殺 |
| 55 | 园 | 101 | 躲藏 |
| 56 | 僚 | 99 | 遊玩 |

(表二十七，客家語的核心動詞)

這裡要特別說明的是出現頻率高不見得就是核心動詞，例如「連」是高頻詞，但它出現在山歌詞的比率相對非常之高，只能說它是山歌詞中的核心動詞，並不是一般口語中的核心動詞，所以筆者也將它排除。

這五十六個核心動詞具有以下幾點特徵：

(1) 除連係動詞「係類」²、助動詞「會」、動詞「僚」不能帶賓語之外，其他

² 連係動詞，如「係、姓、安、安做、變做、當做、當、像」等，我們的描述體系叫「係類」。

全部都是可帶賓語的及物動詞。

- (2) 核心動詞出現的頻次相對也是使用頻次較高動詞。
- (3) 核心動詞都是單音節詞，而且具有較寬泛的搭配範圍，例如與賓語的搭配，可以是名詞或名詞片語或時間詞、處所詞等。
- (4) 這五十六個核心動詞與現代華語不同的計二十三個：「講 1、食 1、分 1、無、喊、愛 1、係、知、行、噉、著 1、轉 1、走 1、驚、倚、落、尋、loi53、做 3、愁、刷、園、撩」，佔全部核心動詞的 41.1%。這二十三個詞當中有些是保存自古漢語，如「講、食 1、無、喊、知、行、噉、轉 1、走 1、驚、倚、尋、愁」等，有些是客家語本有的「底層詞」，如「刷、園、撩」。
- (5) 這二十三個與現代華語不同的核心動詞總使用詞次為 7,850 次，佔五十六個核心動詞總使用詞次的 44.58%，高頻的現象，顯示這是核心中的核心，相對的穩定性也非常高，目前尚未發現這些詞有被新詞或其他語言詞彙取代的現象。

第四節 四縣客家語和海陸客家語動詞對比分析

本節將全部單音節動詞加以分析和對比，分離出（1）兩種腔調相同單音節動詞（2）僅用於四縣客家語的單音節動詞（3）僅用於海陸客家語的單音節動詞三種成分，並以「語言接觸的影響」角度說明這些動詞交互使用的情形。

兩種腔調均相同的單音節動詞計 1269 個，佔全部單音節動詞的 97.69%；僅用於四縣客家語的單音節動詞計十九個，佔全部單音節動詞的 1.46%；僅用於海陸客家語的單音節動詞計五個，佔全部單音節動詞的 0.39%。

（一）兩種腔調均相同的單音節動詞：

在台灣地區四縣和海陸這兩種較優勢的客家語腔調，在單音節動詞方面有許多的相同之處，本文所謂相同是指在語法功能和詞義兩方面均相同的意思，這些相同的單音節動詞計 1269 個：

講 1、有 1、去、看 1、來、想 4、做 1、食 1、分 1、無、像、喊 1、愛 1、到 1、用 1、係、知、行、噉、問 1、著 1、連 1、打 1、賣、罵、轉 1、走 1、聽 1、過 1、坐 2、坐 1、等、拿 1、想 1、寫、睡、出 1、拿 2、買、請 1、驚、死、倚、落、放 1、p'iak50、裝、尋、loi53、做 3、笑 1、好 1、帶、愁、會、刷 1、園、燎、照、種、辦 1、學、降、燒、走 2、嫁、le53、放 3、擊 2、討 3、煮、跌 1、留、攬 1、算 1、求、加、送 1、開 1、畜、搞、跳、核、在 1、捉、換 1、趨、變、擊 1、包 1、改 2、擗、借、教、k'u55、惜、肯、接 1、起 1、愛 2、摘、拜、跔、攞、整、纏、跌 2、揸、憑、入、吊 1、好 2、點 3、裝 2、幫、搵、贖、傳 1、搖、讀、補 1、吶、岐、撓、順、拐 2、敬、偷、拈、害 2、割、想 3、跪、戴 a、獵、p'an53、冷 1、發 1、歇 2、k'em55、mak20、敵 1、牽 1、橫 2、姓、揩 1、拉 1、管、撥 1、追 1、停、唱、嫌、踏、mu24、輸、弄 2、退 4、想 2、監、擘、還、下 1、貼 1、lap20、笑 2、救、喊 2、試、算 2、搥、屙、lak20、撈、應、罇 1、挖、流、割 1、掌、結 1、駛、磧、上 2、伸、忍、咬 1、唸、蒔、靠、嚇、識、降、脫、話 1、跣、退 2、蓋、趕、安 1、冷 2、拆 1、炒、背 1、害 1、惇、搭 2、躡、播 1、跔、摸、磕 1、衝、賭、大 1、扞、扛、見、刮 2、晒、兜、跋、暢、遮、醒、賺 1、聽 2、lau11、moi24、躡、ok20、tam53、t'eu11、斗、打 10、吊 2、含 2、使 1、剝、宕、封、洗、飛、倒 2、耕、退 5、插 1、搶、腫、搵 1、撈、撥 2、擋 1、橫 1、搭、緝、揸、合、敢、運、敲 1、撿、上 3、召、生 2、扶、走 4、使 4、爭 2、捲 2、涿、超、敵 2、當 1、搵 2、領、綯 1、撐 2、湯、拂 1、動 1、粹 1、騎 1、欠 1、收 1、渡、溜 1、載、漏、歎、贏、鑽、中、切、反 2、打 2、拖 1、挾、浸、接 2、掃 1、粹 3、隔 2、認 1、熟、束、罰、

hoŋ53、nen11、光、按 1、梟、喝、擗 2、惹、搬、請 3、豁、lai55、唛、tem53、
生 1、吹 1、的 2、看 4、屛 2、望、透、報、捶、焙、睐、隔 1、煲、認 2、綯 2、
潦、擋 2、歸、關 1、颯、落 b、fun11、抨、辽、爭 1、省、看 3、穿、弇、剝 1、
推 b、圍 1、量、滷、撐 1、磬、k'at50、la11、t'iu11、分 2、出 2、收 2、吞、使
2、到 2、定、拚、拆 2、信 1、咬 3、做 2、探 3、掛 2、排、覓、塞 1、滾、攙
2、暴、請 2、搗 3、選 1、頷、醫、噍、k'iam11、piak20、代、叩、安 2、埋、
按 2、接 4、貪、暗、撩 2、鬧、舖、瞞、響、顧、關 2、la11、lai55、lep20、lu11、
mi55、naŋ11、瞞、suk20、tem55、vuk20、叨、沉 a、使 3、拉 3、拂 2、拍、放
2、爬 2、指、看 2、泄、剝 2、退 3、啣、掀、袋、逕、粹 2、焗、換 2、翕 2、
進、搗、傳 2、搜、對、撞、撥 3、撩 1、揪、磕 2、談、調 1、離 1、簽、鬥 1、
拏 2、趨 2、跟、紮、眨、vak50、丟、成、抵、拖 2、昂、閃、犁、釣、拏 1、
掙、斟 a、煖、趨 1、踢、勸、聾、必、串、走 3、爬 1、捫、討 2、添、舂、焗、
飲、楔、過 3、慢、慣、稱、箍、彈、暴、毆、糊、練、闌、黏、搨、k'ie11、揆、
少、扣 1、盯、受、眨、退 1、釘、假、得、斬、提、皺、編、奮、餵、擺、蹶
2、攪、顯、鑪、fin11、ŋiap20、tep20、伏、印、尖、考、含 1、花、兼、啄、
掛 1、啼、跛、傷、痺、逼、嘍 2、幹、貌、擗、潑、擇、翻 2、轉 3、壞、蹶 1、
騙、sia55、tap50、t'iau55、佔、抔、刺、招、歿、怨、拚 1、訂、赴、修、展、
捏、鬥 3、捲 1、探 1、接 3、掛 3、赦、厠、畫、開 5、閒、煎、落 1、墊 1、摺、
漆、鼻、熅、劈、熱、擔、嚐、聲、擲 1、瀉、織、醬、爆、躑、hng、當 2、輪
2、賸、p'un55、p'en11、vok20、上 1、爪、出 3、打 7、托、收 3、老、扠、告、
吠、弄 1、抓、刻、承、抹 2、放 4、泡、辱、拚 2、屛 1、剖、梳、舀、除、高、
堵 1、探 2、推 a、敗、蛀、通、揣、散、發 2、翕 3、開 4、傾、剷、搽、暈、
溼、解、頓 1、鼓、嘮、燂、畫 2、嗽、奪、摻、蝕、撚、瘦、碾、賞、焗、曉、
燉、賴 1、點 1、攤、疊、ts'am55、lien53、挑 2、ŋien11、piu55、ŋiap50、oŋ55、
伺、了、包 2、打 6、回、吵、孝、扭、找、扳、扯、投、沉 1、車、取、呸、
呻、拉 2、拋 2、抽 2、拐 1、旺、迎、咬 4、待、架、洩、紅、耐、倒 1、凍、

捎1、晃、浮、耙、掩1、推1、捩、清、喊3、圍2、揀、揪、答、翕1、跌3、
開3、煨、墊2、齊、隨、輾、點2、翹1、翹2、轉2、離2、爛、躍、鑄、kiuk20、
激、調2、選2、搥、him11、k'ap20、k'iuk20、搥、搵4、弓、付、包3、卡、失、
打3、交2、守、扣2、低、完、刷、和、念2、怪、拂3、抹1、拔、拋1、抽1、
拗、服、注、泌、的1、糾、昇、咬2、恬、挑1、派、相、苦、食3、捆、挨3、
祝、納、討1、託、做4、偎、側、啜、堆、捧1、掛4、陪、滄、祛2、割2、
喪、結2、裂、開2、開6、集、黃、亂、搭3、損、督、矮、聘、嘔、嘍1、摧、
漲、搏、墜、影、撈2、謎、箭、蔭2、論、賠、熠、播4、謀、辦2、燥、點4、
點5、戳、擲2、繡、鎖、題1、蹬、願、邏、讓、遞、肖、鬆1、膨、k'uak50、
tun53、撬、tiak20、縫、插3、臭、浮、堵2、閉、sio11、賒、摺、化1、出4、
凸、打11、打4、划、同1、行2、判、扼、念1、抵、押2、汨、炊、直、保、
削、門、挨2、病、眠、秤、缺、記、豺、送2、追3、鬥2、淳、做5、偏、圈、
崩、淹、移、責1、祛1、惱、摒1、減、發4、絞、絕、絮、著2、詐、鈕、頓
2、嘮、搨、嫖、漂、舞、標、靚、凝、播2、鋸、擦、臨、蹶、題2、攀、護、
齧、彎、捋2、餓、剪1、訛1、噎、k'uak55、p'at20、ŋioŋ24、k'ak50、沖、遷、
勾2、休、離2、溜2、牽2、打9、fen1、焯、綦、p'u11、ven24、大2、分3、
化2、刊、佔、打5、打8、交1、合2、吮、局、戒、批、泐、沒、巡、供、刺
2、咄、咚、奈、帕、征、押1、升、肥、咬5、拱、柵、炸1、炸2、祈、負、
食2、香、值、准、哺、射、座、挽、挨1、捍、烏、舐、配、噉、捋1、剪2、
強、徙、掩2、採、淋、設、揆、傍、測、發3、結4、絡、著3、裁、貯、貼2、
搭1、搏、溶、滅、滑、落a、蹠、過2、過4、飼、損、嗽1、嘆、摟、敲2、
槌、綻、攪1、層、標、潤、磅、蔭1、輪、燒、壅、戰、燙、爛、瞠、縈、錐、
霑、遶、壓、擠、攔、聯、謝1、齋、蹠、膠、薰、覆、闖、雜、騎2、嚨、攔、
攝、灌、露、鑲、糶、躡、鑿、丙、亾、爍、漼、拖、鬆2、訛2、精、ti55、蕪、
koi55、掃2、出5、p'ak20、koŋ55、siu11、kiu55、稱2、遏、蕩、捎2、kio53、
lak20、nak20、tap50、t'e24、劉、tut20、勻、勾1、犯、生3、交3、同2、吐、

刮、免、助、呆、改 1、防、刮 1、咕、昏、沾、祀、侵、恨、挖 1、施、查、
 畏、枷、框、栽、消、起 2、迷、追 2、哈、祛、祭、副、掣、插 2、揚、湯、
 登、硬、結 3、貸、越、跑、森、圖、滴、滬、熔、聚、辣、醉、篩、礪、應 2、
 盪、隆、縱、謝、鋏、罇 2、辭、躅、鑄、灑、嚙、湊、voj53、lem53、禁、驕、
 拑、狲、刨、k'au55、k'au55、piak20、鰻、鬱、k'iet20、合 3、合 4、激 2、激 3、
 掄 1、掄 2、掄 3、掄 4、拉 4、扭 2、扭 3、竭、孔、kam24、賺 2、掇 3、刻 2、
 拉 5、割 4、sot、賴 2、oj11、kiuk20、kiu55、熟、揩 2、挾 2、拏 1、拏 2、哽、
 裝 3、啣、話 2、破、mo11、起 3、結 5、在 2、鬥 4、擯 3、piang11、捉 2、捉
 3、多、浪、連 2、連 3、食 4、當 3、掉、發 5、起 4、放 5、放 6、有 2、有 3、
 有 4、欠 2、好 3、算 3、問 2、問 3、問 4、寫 2、信 2、用 2、講 2、補 2、補 3、
 攬 2、乾、kien55、搭 4、荊、會 2、貸、vut20、ɲiap20、塞 2、責 2、爭 3、擊 3、
 im11、染、動 2、刷 2、鬥 5、綿、打 12、tam53、p'un55、稅、租、荒、枷 2、
 沖 2、過 5、包 4、修 2、喁、行 3、巴、貸、噉、擔、掙 3、

兩種腔調相同單音節動詞的高比率相同，顯示這兩種客家語腔調的溝通度
 非常高。這些相同的單音節動詞總使用詞次為 33,339，佔全部動詞使用詞次的
 66.13%。

(二) 僅用於四縣客家語的單音節動詞

對於四縣和海陸話動詞的比較研究，黃雯君（2004）比較了「譴和闕」、「餵
 和供」、「歇和戴」、「飢和枵」這四組單音動詞，以下我們根據語料庫查詢所得，
 並透過田野調查法訪問發音人，分別檢驗出兩種腔調中各自使用的單音節動詞。
 筆者所謂僅用於「某一腔調」的意思是指根據文本和訪問發音人所得的結果來定
 義的，並不是符合每一位客家語言使用者的每一種情形，同時也不代表這是結

論，筆者了解應該還有一些單音節動詞並未錄入語料庫中，加上義項的分合及個人語感上的差異等因素，均會造成結果的偏差。

關於黃雯君（2004）所比較的「譴和闕」、「餵和供」、「歇和戴」、「飢和枵」這四組單音動詞，根據筆者的標準，我們認為只有「譴和闕」、「歇和戴」這兩組單音動詞是四縣和海陸腔的區別，「供和餵」、「飢和枵」這組我們加以排除，這裡先敘述排除「供和餵」的理由，關於排除「飢和枵」的理由見後。

關於「餵豬」，海陸客家語並不全然使用「供」這個動詞，也使用「餵」這個動詞，根據海陸客家語的發音人表示這兩個意思是一樣的，使用「供」相同的還有「供雞、供鴨、一樣飯供百樣人」，使用「餵」的有「餵投生、餵雞鴨、餵豬仔、餵奶、餵狗」，四縣客家語也是同時使用這兩個動詞，查《梅縣客家語志》有「供豬、一樣飯供百樣人」，《台灣客家語辭典》也有「供豬仔」，根據四縣客家語的發音人表示「供和餵」可以交互使用，但使用「餵」較多，可見「供」不見得是海陸客家語使用的動詞，「供」的對象也不是只有動物，人也可以用「供」，如「一樣飯供百樣人」，而「餵」則是普遍用於餵養動物，沒有用於「人」的情形，根據這些資料和說明我們認為「供和餵」並不能成為四縣客家語和海陸客家語的區別性動詞，因此我們將「供和餵」這一組予以排除。

下表是僅用於四縣客家語的單音節動詞，總計十九個：

| 序號 | 詞例 | 音 | 詞頻 | 詞義 | 語料例 |
|----|----|---------|-----|------|------------|
| 1 | 歇1 | hiet20 | 167 | 住 | 庄肚項歇有一個惡霸 |
| 2 | 譴 | k'ien31 | 53 | 生氣 | 一下就譴起來 |
| 3 | 養 | ioŋ31 | 20 | 生養 | 貓環養介貓子死忒兩隻 |
| 4 | 恹 | men31 | 18 | 推想、想 | 頭下恹恹過哩。 |

| | | | | | |
|----|----|---------|----|---------|------------------|
| 5 | 繳 | kieu31 | 12 | 供應 | 刻奈仔乜會繳你讀 |
| 6 | 拚 | p'in24 | 6 | 拚毛辮 | 佢毛辮拚起來，加益靚。 |
| 7 | 反1 | p'on24 | 6 | 歐吐 | 下擺仔還會呃呃滾像愛反樣仔 |
| 8 | 炙 | tsak20 | 4 | 烘烤 | 炙火窗 |
| 9 | 搓 | ts'ok50 | 4 | 搓牙齒 | 搓牙齒 |
| 10 | 綁 | pon31 | 3 | 綁住 | 就 lau53 佢捉來樹仔項綁等 |
| 11 | 瞞 | liu31 | 2 | 欺騙 | 瞞走了 |
| 12 | 掇 | ŋau11 | 2 | 轉向 | 掇過該片去 |
| 13 | 下2 | ha55 | 2 | 脫下 | 帽仔到屋家就好下哩，戴等做麼个 |
| 14 | 弛 | ie24 | 2 | 滑動 | 弛往上 |
| 15 | 沿 | ien11 | 2 | 順著 | 原住民介社庄部落沿等河霸設立 |
| 16 | 撮 | ts'ot20 | 2 | 用耍騙方法偷走 | 撮人幾擺了 |
| 17 | 薰 | hiun24 | 2 | 煙薰 | 毋使吊蚊帳，毋使薰蚊草。 |
| 18 | 攣 | lion11 | 2 | 縫製 | 阿姆先攣較狹兜仔， |
| 19 | 斫 | ts'ok20 | 1 | 砍伐 | 斫樵 |

(表二十八，僅用於四縣客家語的單音節動詞)

(1) 歇 四縣客家語說「hiet2」，相當於華語的「住」，如：「庄肚項歇有一個惡霸」(阿啣箭个故鄉__天有眼)、「吾屋家這下歇个屋仔」(阿啣箭个故鄉__油桐花)。海陸客家語在這個義項是說「戴」，音「tai11」。四縣和海陸客家語說「歇」的另一義項是指「休息片刻」，如：「歇睏一下，加下繼續上課。」。

(2) 譴 四縣客家語說「kien31」，相當於華語的「生氣」，如：「一下就譴起來」(阿啣箭个故鄉__阿爸个畚箕篤)。海陸客家語在「生氣」方面，則說「闕」，音「at20」，以現有的文本來看，海陸客家語的文本很一致的使用「闕」這個詞彙，如《新屋客家語故事集》p60、p190。四縣客家語「譴」是指生氣，「闕」是

指生悶氣，我的發音人表示海陸客家語沒有這種區分，「譴」只用於四縣客家語。

(3) 養 四縣客家語說「iong24」，是指「動物產子」的意思，如：「貓攞養介貓子死忒兩隻」(台灣客家文學選集)。我的發音人表示「動物產子」海陸客家語說「降」，音「kiung11」，如「厰屋下狗嬭這擺降六條狗子」。

(4) 恁 四縣客家語說「men31」，相當於華語的「想」之意，如：「佢兜歸屋下人就分人恁毋解」(台灣客家文學選集 p76)、「恁事情」、「恁心事」、「恁一下仔」、「恁真兜仔」。我的發音人表示海陸客話不說「恁」，只有「想」，音「siong24」，四縣客話也說「想」，音「siong31」，用法與海陸客家語無別，如「無想到」、「想毋出來」。從頻率來分析，「想」的頻率為 175 次，分布在四縣客家語文本 98 次，在海陸客家語文本 77 次；而「恁」的頻率為 18 次，全部在四縣客家語文本。

(5) 繳 四縣客家語說「31」，相當於華語的「供應、提供、繳納」之意，如：

你愛煞猛讀，係讀得來，刻苦仔乜會繳你讀，下二擺畢業，再過去讀中學、讀大學，正會有春光个日仔。(阿啾箭个故鄉__石跳仔路)

下後，佢讀中學、大學，吾姆還盡煞猛、省儉，畜豬仔分佢繳註冊費。(阿啾箭个故鄉__細心舅仔)。

我的發音人表示海陸客家語不說「繳」，而說「交」(kau53)，如：

細人仔嚴時愛開學了，愛交註冊錢。

這只月愛交屋稅錢，下只月愛交地價稅，逐只月交無停。

(6) 拚 四縣客家語說「24」，相當於華語的「編製」之意，如「拚毛辮仔」，指「編絞毛髮使成辮子。」的意思。我的發音人表示海陸客家語不說「拚」而說「編」，音「pian53」，或說「綯」，音「55」，如「編毛辮仔」、「綯毛辮仔」。

(7) 反 1 四縣客家語說「𠵼𠵼𠵼24」，相當於華語的「嘔吐」之意，如「下擺仔還會呃呃滾像愛反樣仔」（阿啾箭个故鄉 p134）。我的發音人表示海陸客家語這個義項不說「反」，而說「嘔」，音「eu24」；四縣客話也說「嘔」，音「eu31」，用法與海陸客家語無別，如「嘔出」、「嘔血」、「又屙又嘔」。我的發音人表示海陸客家語說「𠵼𠵼𠵼53」是用於「翻面」的意思，如「𠵼𠵼𠵼53 雞腸」是指將雞腸子翻面以利清洗之意。

(8) 炙 四縣客家語說「𠵼𠵼𠵼20」，相當於華語的「煮」之意，如「炙火」、「炙茶」、「炙燥來」，「炙」廣韻「之石切，音隻」。「炙茶」應是煮開水用來泡茶的意思，又可泛稱煮開水，與「炙滾水」同義。海陸客家語不說「炙」，而說「焯」，音「𠵼𠵼55」，如「焯茶」、「焯燥忒」、「焯滾水」。

(9) 搓 四縣客家語說「𠵼𠵼𠵼55」，相當於華語的「搓揉去除」之意，如「搓 man55」、「搓牙齒」、「搓落下來」、「搓衫褲」。海陸客家語不說「搓」，而說「洗」，音「se24」，所以「刷牙」不說「搓牙齒」而說「洗喙」（se24 tsoi11），而「洗衣服」不說「搓衫褲」，而說「洗衫褲」，四縣客家語也用「洗」，音「se31」，用法與海陸客家語無別，如「洗身腳」、「洗面」、「洗衫褲」。從語料來看，「搓」只出現 4 次，而且都在四縣客家語的文本中的幾個固定詞語，「搓 man55」、「搓牙齒」、「搓衫褲」、「手一搓就揸佛字搓忒了」（苗栗縣民間文學集 4，p104），而使用「洗」的頻率高達 53 次，顯示「洗」還是主要的用詞，筆者的 4 位四縣客家語發音人中在「搓衫褲」方面，只有一位有使用「搓」，其他使用「洗」；在「搓牙齒」方面倒是很一致，綜合研判「搓」或可解釋為華語詞彙逐漸擴散進入四縣客家語詞彙中的一個例子。

(10) 綁 四縣客家語說「𠵼𠵼𠵼31」，相當於華語的「綁」之意，如「就揸佢捉來樹仔項綁等」（苗栗縣民間文學集 4，p128）。海陸客家語只說「綯」，音「𠵼𠵼55」，或「綯」，音「𠵼𠵼𠵼50」。四縣客家語的「綁」在語料中只出現五

次，分別出現在「苗栗縣民間文學集 4 和 6」兩種文本中，根據筆者向原紀錄人詢問，四縣客話的確有這種講法，而筆者的六位發音人全部表示沒有這種說法，只和海陸客家語一樣有「綯」或「綯」、「綯」三種說法，從使用頻率來看「綁」這個說法，在台灣地區的客家語並不是主流的說法，詳見下文分析。

(11) 罨 四縣客家語說「liu31」，相當於華語的「欺瞞」之意，如「罨人」。海陸客家語只說「lau11」或「騙」。這個「罨」是個低頻詞，只出現 1 次。我的四縣客家語的發音人表示「罨」有點近似「耍騙」的意思，比「騙」輕微，有點類似「耍詐」之意，如「講好愛來，仰會無來，實在罨人。」，我的發音人表示海陸客話沒有「罨」這個詞彙。

(12) 掇 四縣客家語說「掇」，相當於華語的「轉向」之意，如「掇過該片去」、「掇轉頭」、「掇過來」。我的發音人表示海陸客家語沒有「掇」這個詞彙，只用「轉」這個詞，如「轉過該片去」、「轉過來」。

(13) 下 2 四縣客家語說「ha24」，相當於華語的「取下」之意，特別是指戴在頭上的帽子脫下來的意思，如「帽子到屋家就好下哩，戴等做麼个？」。我的發音人表示海陸客家語的「下」沒有這個義項，帽子脫下來一般海陸客家語用「剝」或「拿」。

(14) 弛 四縣客家語說「ie24」，相當於華語的「鬆落、滑動」之意，如「弛走」是向外移動，「弛胸 lut20 胛」是鬆落衣服，露出胸脯與肩胛，猶言「袒胸露乳」之意。我的發音人表示海陸客家語沒有「弛」這個詞彙。

(15) 沿 四縣客家語說「ian11」，相當於華語的「順著」之意，如「原住民介社裝部落沿等河霸設立」。我的發音人表示海陸客家語沒有「沿」這個詞彙，而用「順」這個詞。

(16) 撮 四縣客家語說「20」，相當於華語的「欺騙」之意，如「撮人」是欺騙人，「撮假」是做假騙人之意。我的發音人表示海陸客家語沒有「撮」這個詞彙。

(17) 薰 四縣客家語說「fun55」(hiun24)，相當於華語的「煙燻」之意，如「毋使吊蚊帳，毋使薰蚊草。」，這個用法在語料庫中僅出現一次，即出現於《台灣客家文學選集 2》頁 170：「毋使吊蚊帳，毋使薰蚊草。蚊仔像摻佢打過契約，毋嚙巨，毋食其血，嫌其血臭腥。」，我的發音人表示海陸客家語沒有使用「薰」這個詞彙。

(18) 攣 四縣客家語說「li11」，相當於華語的「縫製」之意，如「攣衫褲」，海陸客家語說「聯」，音「lien55」，如「褲腳聯好了。」。

(19) 斫 四縣客家語說「ts'o」，相當於華語的「砍」之意，如「斫柴、斫豬肉、斫樵、斫蔗、斫斷、鷓鴣滿山啼，放心去斫樵、斫柴入山」等，這個詞彙發音人認為可能與閩南語有關³，閩南語讀「ts'o」，根據《台灣話大辭典》做「剉」，釋義：「斬也。斬截也。發木曰剉」，並有「剉竹、剉柴、剉樹」等詞彙。我的發音人表示海陸客家語沒有「斫」這個詞彙，只用「剉」(tok50)，如「剉豬肉、剉樵、剉斷、剉綿綿、剉盤」。

經查《梅縣客家方言志》頁 98 有「斫蔗」，《客家方言》頁 130 有「斫樹」，頁 129 有「斫」(tsok1)，釋意為「用刀斧砍」，可見「斫」原為梅縣客家方言的用詞，姑不論其來源，至少台灣四縣客家語的「斫」，並不一定是來源於閩南語。

(三)、僅用於海陸客家語的單音節動詞

| 序號 | 詞例 | 詞頻 | 詞義 | 語料例 |
|----|------|----|-----|----------------|
| 1 | 戴 1 | 82 | 住 | 佢戴奈你知無？ |
| 2 | 闕 | 41 | 生氣 | 厥姆聽著當闕。 |
| 3 | li11 | 12 | 拉起 | 衫介拉鍊 li11 起來旨？ |
| 4 | 柺 | 2 | 使飢餓 | 柺等轉 |

³ 發音人表示，所謂「與閩南語有關」，並非意味就是吸收自閩南語。

| | | | | |
|---|---|---|----|----------|
| 5 | 聯 | 2 | 縫製 | 爛空介位仔旨聯好 |
|---|---|---|----|----------|

(表二十九，僅用於海陸客家語的單音節動詞)

(1) 戴¹ 海陸客家語說「tai11」，相當於華語的「居住」之意，指長久居住而言，亦相當於四縣客家語的「歇¹」之意，如「若屋下戴奈？」、「偈戴台北幾十年了。」。

(2) 闕 海陸客家語說「at50」，相當於華語的「生氣」之意，亦相當於四縣客家語的「譴」之意，如「心肝就盡闕」、「人係毋會發闕，該正怪！」。

(3) li11 海陸客家語說「li11」，主要用於「拉」拉鍊，如：「拉鍊 li11 起來，毋好食著風。」，這個漢字為何？筆者不知，暫時用標音表示。關於拉拉鍊這個動作，海陸客家語還讀「lo53、lai53」這兩個音，我們都記為漢字「拉」。根據發音人表示四縣客家語「拉」拉鍊也可說「lai24」或「*~~lai~~24」，而海陸客家語的「li11」，則是四縣客家語不說的。

(4) 枵 海陸客家語說「iau53」，用於動詞表示「使飢餓」的意思，但也用「餓」(ngo11) 這個動詞，如：「該憨狗，餓佢兩餐仔，看佢還敢吠無？」。四縣客家語也有用「枵」，但不用作動詞，而是用作形容詞，如「枵鬼假細義」、「肚枵」⁴，《客話辭典》認為是「「枵」iau24 音來自閩南語。」。四縣客家語用於表示飢餓的形容詞是「飢」，如「一飽忘百飢、肚飢」等，而用於「使飢餓」的還是「餓」(ng055) 這個動詞，所以「枵」的動詞義只有海陸話在使用。黃雯君(2004)所比較的「飢和枵」這一組是形容性的比較，黃雯君(2004)的原比較詞是「肚飢」和「肚屎枵」，不是動詞性的，所以我們加以排除，動詞性的是「餓和枵」這一組。

(5) 聯 海陸客家語說「lien55」，相當於華語的「縫製」之意，如「聯衫褲」，四縣客家語則說「攣」，音「~~li~~11」，如「褲腳爛一空，攣好了。」。

⁴ 根據筆者訪問的四縣客家語發音人表示，他們同時使用「肚飢」和「肚枵」，沒有差別。

以上我們分析了四縣和海陸腔單音節動詞方面的差異，我們把分析的結果

做成如下表，以利比較：

| 序號 | 詞例 | 四縣 | 海陸 | 對應詞 | 詞頻 | 詞義 |
|----|-----|-----|----|-----------|-----|---------|
| 1 | 歇 1 | + | — | 戴 1 | 167 | 住 |
| 2 | 譴 | + | — | 闕 | 53 | 生氣 |
| 3 | 養 | + | — | 降 | 20 | 生養 |
| 4 | 恹 | + | — | 想 | 18 | 推想 |
| 5 | 繳 | + | — | 交 | 12 | 供應 |
| 6 | 拚 | + | — | 編、綯 | 6 | 編髮辮 |
| 7 | 反 1 | + | — | 嘔 | 6 | 嘔吐 |
| 8 | 炙 | + | — | 焗 | 4 | 烘烤 |
| 9 | 搓 | + | — | 洗 | 4 | 搓牙齒 |
| 10 | 綁 | + | — | 綯、絹、 網 | 3 | 綁住 |
| 11 | 瞞 | + | — | lau11、騙 | 2 | 欺騙 |
| 12 | 掇 | + | — | 轉 | 2 | 轉向 |
| 13 | 下 2 | + | — | 剝、拿 | 2 | 脫下 |
| 14 | 弛 | + | — | — | 2 | 滑動 |
| 15 | 沿 | + | — | 順 | 2 | 順著 |
| 16 | 撮 | + | — | lau11、騙 | 2 | 用耍騙方法偷走 |
| 17 | 薰 | + | — | 薰 | 1 | 煙薰 |
| 18 | 攣 | + | — | 聯 | 2 | 縫製 |
| 19 | 斫 | + | — | 剝 | 1 | 砍伐 |
| 20 | 戴 1 | — | + | 歇 1 | 82 | 住 |
| 21 | 闕 | — + | + | 譴 | 41 | 生氣 |

| 序號 | 詞例 | 四縣 | 海陸 | 對應詞 | 詞頻 | 詞義 |
|----|------|-----|----|-----|----|------|
| 22 | li11 | — | + | 拉 | 12 | 拉起拉鍊 |
| 23 | 枵 | — + | + | 餓 | 2 | 使飢餓 |
| 24 | 聯 | — | + | | 2 | 縫製 |

（「+—」表示四縣客話也使用的意思。）

（表三十，海陸客家語的單音節動詞比較表）

第五節 華語、客家語動詞對比分析

本節我們將動詞和華語進行比較⁵，針對單音節動詞分析出華客家語之間的（一）相同度（二）相異度，針對雙音節動詞我們探討（三）華客家語之間的詞序問題

（一）相同度

在 1299 個單音節動詞中，只有一個「跑」是引述華語，出現在作家龔萬灶先生的《阿啾箭个故鄉》：「我就跑給他趨」，華語的「跑」相當於客家語的「走 1」，這裡我們略去這個動詞，以 1298 個動詞為分析對象。

經過分析我們發現總計有 633 個單音節動詞與華語相同，我們的驗證方式是進入「中央研究院現代漢語語料庫」去查核，凡能在該語料庫中查核出來並與客家語在詞義、音節方面相同者，我們便視為「華客家語相同的動詞」，反之在

⁵ 請注意這裡比較的也是華語該詞的某一義項的意思，並不是該詞義的全部，根據我們「一詞一義」的原則，實際上就是該同形詞的某一義項，但是華語部分我們不再一一標示其義項分合關係，以下均同。

詞義、音節方面，其中有一項不同者我們便視為不相同的動詞。經過查核後，發現華客家語相同的單音節動詞佔全部單音節動詞的 48.73%，使用頻次計 17898 次，佔全部單音節動詞使用頻次的 52.94%。這 633 個動詞是：

有 1、去、看 1、來、想 4、做 1、像、到 1、用 1、問 1、打 1、賣、罵、聽 1、過 1、坐 2、坐 1、等、拿 1、想 1、寫、睡、出 1、拿 2、買、請 1、死、放 1、裝、笑 1、好 1、帶、會、照、種、辦 1、學、燒、走 2、嫁、放 3、煮、留、求、送 1、開 1、跳、在 1、捉、換 1、變、包 1、借、教、肯、接 1、愛 2、摘、拜、纏、吊 1、好 2、點 3、傳 1、搖、讀、補 1、敬、偷、害 2、想 3、跪、戴 a、牽 1、姓、揸 1、拉 1、管、追 1、停、唱、嫌、踏、退 4、想 2、還、貼 1、笑 2、救、試、算 2、挖、流、結 1、上 2、伸、忍、咬 1、唸、靠、嚇、脫、退 2、蓋、趕、冷 2、拆 1、炒、背 1、害 1、搭 2、孵、摸、磕 1、衝、賭、大 1、扛、見、刮 2、晒、遮、醒、賺 1、聽 2、吊 2、含 2、剃、封、洗、飛、倒 2、耕、退 5、插 1、搶、腫、撈、擋 1、敢、運、敲 1、上 3、扶、捲 2、超、當 1、領、撐 2、動 1、騎 1、欠 1、收 1、溜 1、載、漏、贏、鑽、中、切、挾、浸、接 2、掃 1、認 1、熟、罰、光、惹、搬、請 3、吹 1、看 4、望、報、捶、隔 1、煲、認 2、關 1、飆、爭 1、省、看 3、穿、剝 1、推 b、圍 1、量、滷、撐 1、分 2、出 2、吞、到 2、定、拚、拆 2、信 1、咬 3、做 2、排、塞 1、滾、暴、請 2、選 1、醫、代、叩、安 2、埋、接 4、貪、暗、鬧、舖、瞞、顧、關 2、拉 3、拍、放 2、爬 2、指、看 2、退 3、啣、掀、焗、換 2、進、傳 2、對、撞、談、調 1、離 1、簽、跟、紮、丟、成、拖 2、閃、犁、釣、踢、勸、聾、走 3、爬 1、添、舂、過 3、慢、慣、稱、彈、暴、毆、練、闖、黏、少、扣 1、受、眨、退 1、釘、假、得、斬、提、皺、編、餵、擺、顯、伏、考、含 1、花、啄、啼、傷、逼、轉 3、壞、騙、怨、訂、修、捏、鬥 3、捲 1、探 1、接 3、掛 3、赦、屙、畫、開 5、煎、摺、漆、劈、熱、擔、嚐、織、爆、當 2、上 1、出 3、收 3、告、吠、刻、抹 2、放 4、泡、剖、梳、吞、高、堵 1、敗、蛀、通、散、開 4、剷、搽、暈、溼、

解、畫2、奪、摻、瘦、碾、賞、曉、燉、點1、攤、疊、包2、回、吵、孝、扭、找、扯、沉1、取、拉2、拋2、抽2、拐1、旺、迎、咬4、待、架、洩、紅、耐、凍、浮、耙、清、圍2、揪、答、開3、墊2、齊、隨、輾、點2、翹1、轉2、離2、爛、鑄、調2、選2、付、包3、卡、失、交2、守、扣2、低、完、和、念2、怪、抹1、拋1、抽1、服、咬2、挑1、派、苦、捆、祝、納、討1、託、做4、側、啜、堆、捧1、掛4、陪、裂、開2、集、黃、亂、損、矮、聘、嘔、摧、漲、搏、墜、撈2、謎、論、賠、謀、辦2、點4、繡、鎖、題1、願、讓、膨、臭、賒、化1、打11、判、抵、押2、炊、直、保、削、門、病、眠、秤、缺、記、送2、追3、偏、圈、崩、淹、移、發4、絞、絕、著2、詐、鈕、嫖、漂、標、鋸、臨、護、彎、餓、沖、休、溜2、打9、分3、化2、刊、佔、交1、戒、批、巡、征、押1、升、沿、肥、咬5、拱、炸1、炸2、祈、負、香、值、准、射、座、挽、挨1、捍、烏、配、剪2、掩2、採、淋、設、測、發3、結4、裁、貼2、搭1、溶、滅、過2、嘆、敲2、槌、標、輪、戰、燙、爛、壓、擠、謝1、薰、闖、雜、攔、灌、露、鑲、糶、鑿、勾1、犯、交3、吐、免、助、改1、防、刮1、昏、祀、侵、恨、挖1、施、查、畏、枷、框、迷、追2、祭、插2、揚、登、硬、貸、越、跑、滴、熔、聚、辣、醉、篩、縱、謝、辭、灑、湊、禁、刨、合3、合4、拉4、扭2、拉5、哽、啣、起3、結5、在2、多、連2、連3、當3、掉、放5、放6、有2、有3、有4、欠2、好3、算3、問2、問3、問4、寫2、信2、用2、講2、補2、補3、剃、會2、貸、染、租、荒、沖2

在這 633 個華客家語相同的動詞中，「沿」這個動詞是海陸客家語不使用的，其他四縣和海陸都相同。

(二) 相異度

華客家語相異的單音節動詞計 665 個，佔全部單音節動詞的 51.27%，使用詞次計 15910 次，佔全部單音節動詞使用詞次的 47.06%。這 665 個華客家語不相同單音節動詞是：

食 1、分 1、無、喊 1、愛 1、係、知、行、噉、著 1、連 1、轉 1、走 1、歇 1、驚、倚、落、p'iak50、尋、loi53、做 3、愁、刷 1、囡、僚、降、le53、擊 2、討 3、跌 1、攪 1、算 1、戴 1、加、畜、搞、核、趨、擊 1、改 2、撻、k'u55、惜、起 1、跣、攪、整、跌 2、揸、憑、入、裝 2、幫、譴、搥、賸、覘、敲、撓、順、拐 2、拈、割、獵、p'an53、冷 1、發 1、歇 2、k'em55、mak20、敵 1、橫 2、闕、撥 1、mu24、輸、弄 2、監、擊、下 1、lap20、喊 2、搥、扇、lak20、撈、應、罇 1、割 1、掌、駛、磧、蒔、識、降、話 1、跣、安 1、惇、搥 1、跔、扞、兜、跋、暢、lau11、moi24、躡、ok20、tam53、t'eu11、斗、打 10、使 1、宕、搥 1、撥 2、橫 1、揸、緝、揸、合、撿、召、生 2、走 4、使 4、爭 2、涿、敵 2、搥 2、絢 1、湯、拂 1、揸 1、養、渡、歎、反 2、打 2、拖 1、揸 3、恹、隔 2、束、hoŋ53、nen11、揸 1、梟、喝、擗 2、豁、lai55、味、tem53、生 1、的 2、扇 2、透、焙、睽、蒸、絢 2、潦、搥 2、歸、落 b、fun11、揸、辽、舛、磬、k'at50、la11、t'iu11、收 2、使 2、探 3、掛 2、覓、搥 2、搥 3、頷、噉、k'iam11、piak20、揸 2、揸 2、響、la11、lai55、lep20、li11、lu11、mi55、naŋ11、躡、suk20、tem55、vuk20、叨、沉 a、使 3、拂 2、泄、剝 2、袋、逕、揸 2、翁 2、搥、搜、撥 3、揸 1、揸、磕 2、繳、鬥 1、掌 2、趨 2、眨、vak50、抵、昂、掌 1、揸、斟 a、煖、趨 1、必、串、搥、討 2、煇、飲、楔、箍、糊、搥、k'ie11、揸、盯、奮、蹶 2、攪、鑪、fin11、ŋiap20、tep20、印、尖、兼、掛 1、跛、痺、噉 2、幹、貌、搥、潑、擇、翻 2、蹶 1、sia55、tap50、t'iau55、佔、杯、刺、招、歿、拚 1、赴、展、閒、落 1、墊 1、鼻、煖、聲、擗 1、瀉、醬、蹶、hng53、輪 2、

拚、賾、p'un55、p'en11、vok20、反1、爪、打7、托、老、扱、弄1、抓、承、
 扞、拚2、扞1、除、探2、推a、揣、發2、翕3、傾、頓1、鼓、嘮、燂、嗽、
 蝕、撚、焮、賴1、ts'am55、lien53、挑2、ŋien11、piu55、ŋiap50、oŋ55、冏、
 了、打6、扱、投、車、呬、呻、倒1、揷1、晃、掩1、推1、振、喊3、揀、
 翕1、跌3、煨、翹2、躍、kiuk20、激、搥、him11、k'ap20、k'iuk20、搥、搥4、
 弓、打3、刷、拂3、拔、拗、注、泌、炙、的1、糾、昇、恬、相、食3、挨3、
 俛、滄、祛2、割2、喪、結2、開6、搓、搭3、督、嘍1、影、箭、蔭2、熠、
 搗4、燥、點5、戳、擲2、蹬、邏、遞、肖、鬆1、k'uak50、tun53、擣、tiak20、
 縋、插3、浮、堵2、聞、sio11、播、出4、凸、打4、划、同1、行2、扼、念
 1、泓、挨2、豺、鬥2、淳、做5、責1、祛1、惱、摒1、減、絮、綁、頓2、
 嘮、搨、舞、靚、凝、搗2、擦、蹶、題2、攀、齧、拵2、剪1、訛1、噎、k'uak55、
 p'at20、ŋioŋ24、k'ak50、遷、勾2、離2、牽2、fen1、焯、綦、罾、撇、p'u11、
 ven24、下2、大2、打5、打8、合2、弛、吮、局、泔、沒、供、刺2、咄、咚、
 奈、帕、柵、食2、枵、哺、舐、嘜、拵1、強、徙、揆、傍、絡、著3、貯、
 搏、滑、落a、蹠、過4、飼、損、嗽1、摟、綻、摟1、層、撮、潤、磅、蔭1、
 爐、壅、瞠、綦、錐、霑、遶、摺、聯、齋、蹠、朦、覆、騎2、嚨、攝、攣、
 躡、丙、亾、燂、漵、拙、鬆2、訛2、精、ti55、黹、koi55、掃2、出5、p'ak20、
 koŋ55、siu11、kiu55、稱2、盪、蕩、揷2、kio53、lak20、nak20、tap50、t'e24、
 剗、tut20、勻、生3、同2、剗、呆、咕、沾、斫、栽、消、起2、哈、祛、副、
 掣、湯、結3、焱、圖、滄、礪、應2、盪、隆、鍬、罇2、擻、躡、鑄、嚙、voŋ53、
 lem53、驕、拑、獅、k'au55、k'au55、piak20、鰻、鬱、k'iet20、激2、激3、掄
 1、掄2、掄3、掄4、扭3、謁、孔、kam24、賺2、揆3、刻2、割4、sot20、
 賴2、oŋ11、kiuk20、kiu55、熟、措2、攞、挾2、拏1、拏2、裝3、話2、破、
 mo11、鬥4、摒3、piaŋ11、捉2、捉3、浪、食4、發5、起4、攬2、乾、kien55、
 搭4、vut20、ŋiap20、塞2、責2、爭3、擊3、im11、動2、刷2、鬥5、綿、打
 12、tam53、p'un55、稅、枷2、過5、包4、拏、修2、喞、行3、巴、貸、噉、

擔、捋 3

這些華客家語不相同的單音節動詞就不同的情況還可以分爲幾類，有些詞在客家語是單音節的，華語也是單音節的，如：

| 客家語 | 華語 | 語料例 |
|---------|----|--------------|
| 講 | 說 | 你有講出來無？ |
| 食 1 | 吃 | 覓兜轉來剝酒食 |
| 分 1 | 給 | 包一隻紅包分老弟 |
| 喊 | 叫 | 喊若哥好轉了 |
| 愛 1 | 要 | 大家就愛適該過 |
| 係 | 是 | 佢係客家人 |
| 行 | 走 | 行過來 |
| 噉 | 哭 | 目汁流來噉心肝 |
| 著 1 | 穿 | 著等尖尖幼幼介鞋 |
| 走 1 | 跑 | 佢走第一名 |
| 歇 1 | 住 | 庄肚項歇有一個惡霸 |
| 驚 | 怕 | 盡驚鷓婆來吊 |
| 倚 | 站 | 倚在大路唇 |
| 落 | 下 | 看會落一陣大雨 |
| p'iag20 | 打 | 拍蚊子 |
| 尋 | 找 | 就走去尋阿桂伯公 |
| 刷 | 殺 | 刷投生 |
| 囤 | 藏 | 滿姑囤在間肚介時間 |
| 撩 | 玩 | 本來打算愛撩乸十頭八日仔 |

| 客家語 | 華語 | 語料例 |
|--------|-----|---------------------|
| 降 | 生 | 豬攞降十三條豬子 |
| 擎 | 拿 | 擎等一支兔骨在該 kie11 |
| 討 | 娶 | 討轉來做婦娘 |
| 攬 | 抱 | 佢當好攬細妹仔，人喊佢「豬哥伯」 |
| 戴 1 | 住 | 佢戴奈你知無？ |
| 加 | 多、增 | 賺加一息仔外水 |
| 畜 | 養 | 逐家人就有畜雞鴨。 |
| 搞 | 玩 | 係講著搞就一條線 |
| 趨 | 趕 | 趨出大門去 |
| 擎 | 拿 | 去擎箸食飯 |
| k'u55 | 蹲 | k'u55 下來，毋好分佢看著。 |
| 起 1 | 蓋 | 起一間大屋 |
| 跲 | 跟 | 甘願跲其哥去工廠學 |
| 攞 | 提 | 使女仔攞等一隻竹籃仔。 |
| 跌 2 | 跌 | 田項割芥菜又跌一下 |
| 搥 | 弄 | 無搭無碓搥兜空頭出來 |
| 賸 | 剩 | 面賸兩指大 |
| lien11 | 吐 | 牛捶，無愛食了斯緊 lien11 出來 |
| 順 | 順 | 你做得順自家个意 |
| 拐 2 | 哄 | 拐厥姆出去曬日頭花 |
| 拈 | 撿 | 帽仔拈起來 |
| 獵 | 趕 | 佢 liap20 牛仔去食草。 |
| k'em55 | 蓋 | k'em55 塞塞 |
| mak20 | 打 | mak20 落去 |

| 客家語 | 華語 | 語料例 |
|------|-----|--------------------|
| 撥 1 | 拍 | 塵灰撥下來 |
| mu24 | 搞 | 佢常透 mu24 到天天光正有睡。 |
| 輸 | 輸 | 輸淨淨 |
| 屙 | 排、拉 | 屙一大堆屎、痢肚屙歸日，歸身軟軟軟。 |

(表三十一，華語、客家語的單音節動詞)

有些詞在客家語是單音節的，華語卻是多音節的，如：

| 客家語 | 華語 | 語料例 |
|------|-------|---------------|
| 無 | 缺少、沒有 | 還無一餐半餐過來 |
| 知 | 洞悉、了解 | 佢知了！ |
| 連 | 愛戀 | 一妹莫連兩介郎 |
| 轉 1 | 回去 | 秋娘還旨轉 |
| 做 3 | 當、身為 | 做婦娘人，愛守婦得 |
| 愁 | 憂慮、擔心 | 一隻恁分人愁介細人仔 |
| le53 | 來去 | 恁晝了，來去轉！ |
| 跌 1 | 遺失 | 錢跌忒去了 |
| 擗 | 搔癢 | 爸爸擗佢三擺。 |
| 整 | 修理、整治 | 耕種阿伯整屋舍 |
| 憑 | 倚靠 | 細人好憑厥婆共下睡 |
| 入 | 進入 | 佢直直入其婦娘介間房 |
| 裝 2 | 處理 | 裝到盡淨利 |
| 幫 | 擦塗 | 兩隻袖仔幫凖水幫到烏金烏金 |
| 譴 | 生氣 | 一下就譴起來 |

| 客家語 | 華語 | 語料例 |
|--------|-------|-----------------------|
| 撓 | 搔癢 | 爸爸緊 neu53 人，當好撓 |
| p'an53 | 摔倒 | 聽著這消息，佢歸隻人p'an53地泥下去。 |
| 冷 1 | 受寒、感冒 | 冷啊著 |
| 發 1 | 發生 | 該擺係發風災 |
| 歇 2 | 休息 | 歇一下，加下再過上課。 |
| 歇 1 | 休息 | 歇一口大氣 |
| 橫 2 | 倒下 | 係痛這頭樹仔橫忒 |
| 闕 | 生氣 | 厥姆聽著當闕。 |
| 弄 2 | 逗弄 | 細人仔分佢弄到吃吃笑。 |
| 監 | 強迫 | 監佢同列祖列宗歇在該間爛寮仔肚涿雨 |
| 下 1 | 下去 | 裝好斯好下哩 |
| lak50 | 分開 | 布仔拿剪刀 lak50 開來 |
| 應 | 回覆 | 佢無應佢一聲。 |
| 割 1 | 收成 | 翻耕禾割起來 |
| 跼 | 蹲下 | 該日在灶空前跼恁久 |
| 暢 | 高興 | 第一特獎還較暢 |

(表三十二，客家語單音節，華語多音節)

有些詞在客家語是詞義是無法用華語的一個詞彙對應的或華語並沒有相應的說法的，這些只能勉強用一個接近的華語詞或用解釋的方式來說明，如：

| 客家語 | 華語 | 語料例 |
|--------|-------------|---------------------|
| tiak20 | 以手指彈物 | 來比看啊，看麼儕 tiak20 較遠？ |
| sio11 | 用腳踢或以工具攻擊的一 | sio11 界死 |

| | | |
|---------|---------------------|----------------------------------|
| | 種方法 | |
| 凸 | 指從平面突出 | 凸啊出去 |
| 划 | 吃飯的動作 | 盡探划也兩碗 |
| 頓 2 | 撞到 | 手骨頓一下，恁痛！ |
| 嘜 | 打氣 | 嘜風嘜到無人識 |
| 靚 | 長漂亮 | 禾仔毋會靚 |
| 凝 | 凝固 | 等泥漿凝忒再來蒔田 |
| 搨 2 | 近似耍賴、賴皮 | 做毋得斯係做毋得，毋好緊搨人腳髀。 |
| 捋 2 | 打 | 佢捋分你死，試看啊！ |
| 剪 1 | 刺激 | 日頭恁 lat50，會剪人目珠。 |
| 眈 1 | 光亮刺眼 | 日頭眈目珠，看毋清楚。 |
| 噎 | 打嗝 | 歸日噎毋會停，肚屎又脹脹，險會死了。 |
| k'uak55 | 用食指和中指彎曲成半拳狀擊打對方的頭部 | 厥頭那分佢 k'uak55 啊去，一只 leu55，隨下膨起來。 |
| p'at20 | 披掛 | 面帕 p'at20 上去 |
| ŋion24 | 抖動 | 該領被毋好緊ŋion24，盡多塵灰。 |
| k'ak50 | 打 | 五公 ke55 k'ak50 啊分你 |
| 焗 | 一種烘培法 | 正焗一下仔定 |
| 核 | 以肩承物 | 該擔恁重佢核毋贏 |

(表三十三)

有些詞在客家語和華語裡都是單音節詞，如果把雙方的單音節詞組合起來，恰好可以組成華語的一個雙音節詞。不過這種雙音節詞一般都只用在書面語。如：

| 客家語單音詞 | 華語單音詞 | 華語書面語詞 |
|--------|-------|--------|
| 尋 | 找 | 尋找 |
| 燥 | 乾 | 乾燥 |
| 摺 | 疊 | 摺疊 |
| 拐 | 哄 | 哄拐 |
| 落 | 下 | 落下 |
| 降 | 生 | 降生 |
| 驚 | 怕 | 驚怕 |
| 行 | 走 | 行走 |
| 著 | 穿 | 穿著 |
| 光 | 亮 | 光亮 |
| 擇 | 選 | 選擇 |
| 揀 | 選 | 揀選 |
| 出 | 刊、登 | 登出、刊出 |
| 割 | 收 | 收割 |

(表三十四)

以上各種情況，筆者沒有對他們進行量化分析，原因是他們的變異情形太多，筆者暫時略去。例如客家語的「整」是單音節的，表示東西壞掉了加以整理、修理的意思，相對於華語既可以是單音的「修」，也可以是雙音節的「修理」，如「時鐘~~佢~~整好了。」是「時鐘我修好了。」，又如「整屋舍」可以是「修房子」也可以是「修理房屋」，類似這種情形是不少的，所以筆者暫時略去。

(三) 詞序相反

和華語比較，客家語有與華語詞彙詞序相反的現象，但究竟有多少一直也是個謎？羅肇錦（1984）特別指出客家語方言在構詞上比較突出的是複合詞詞序常有顛倒的現象，它常發生的構詞類形是並列格及主從格，羅肇錦並列出 11 個例子來說明，其所列計形容詞「緊要、鬧熱、齊整、歡喜」四個，名詞「氣力、塵灰、康健、人客、牛公、雞公」六個，方位詞「頭前」一個。以下是筆者在語料庫中找出的雙音節詞序相反動詞，總計 19 個，佔全部動詞的 0.06%，可見「詞序相反」並不是客家語動詞構詞的主要手段，所佔比率非常小，但卻是重要的特點。

| 序號 | 客家語詞 | 華語詞 | 詞頻 | 語料例 |
|----|------|-----|----|-------------------------|
| 1 | 崩山 | 山崩 | 2 | 舊年大風災，崩山、發大水，當得人驚。 |
| 2 | 上北 | 北上 | 4 | 佢坐火車上北去尋厥賴仔。 |
| 3 | 實落 | 落實 | 1 | 這事情若無實落核核，毋會有個好个結果。 |
| 4 | 落著 | 著落 | 2 | 錢銀到這下旨有落著，仰葛煞好？ |
| 5 | 裁剪 | 剪裁 | 4 | 佢裁剪衫褲个功夫無人比得。 |
| 6 | 面會 | 會面 | 3 | 韶早先面會講好來，後日正毋會出問題。 |
| 7 | 受領 | 領受 | 1 | 受領你這恩情， 佢 試著益敗勢。 |
| 8 | 買收 | 收買 | 5 | 政府官員分財團買收去了 |
| 9 | 缺欠 | 欠缺 | 2 | 若有麼介缺欠，毋使細意哦！ |
| 10 | 容縱 | 縱容 | 3 | 容縱細人仔，大了你斯知死！ |
| 11 | 收沒 | 沒收 | 2 | 該片土地無納稅分政府收沒去了。 |
| 12 | 紹介 | 介紹 | 4 | 佢 紹介學校同事分你。 |
| 13 | 下南 | 南下 | 2 | 這班火車要下南去嘉義。 |

| | | | | |
|----|----|----|---|------------------------------|
| 14 | 棄嫌 | 嫌棄 | 1 | 若你無棄嫌，今晡日戴佢這。 |
| 15 | 加添 | 添加 | 1 | 加添碗飯分佢。 佢加添一息錢，鬥五千好了 |
| 16 | 量度 | 度量 | 1 | 量度一個人，做毋得打淨看外表。 |
| 17 | 運搬 | 搬運 | 1 | 你想仰搬運搬這兜磚角？韶早正講。 |
| 18 | 合適 | 適合 | 3 | 無合適斯莫強共下，監人也分人監著做 盡係盡艱苦个。 |
| 19 | 歡喜 | 喜歡 | 3 | 你來佢盡歡喜，毋好攞等路。 |

(表三十四，詞序相反)



第十章 動詞短語 (vp) 的結構與關係判定

客家語的研究有許多面向，從語音的紀錄、對比、到用字的考釋、詞彙的整理、對比，理論上和實踐上都取得了一定的成績，許多研究團隊在語音的紀錄、對比方面，做了許多工作，但是對客家語「短語」這一級語法單位的研究目前則處於有待開發的狀態，自然語言裡存在大量的歧義造成短語界定的困難，也造成短語和短語之間關係判定的困難，要解決這些難題，首先人們要對短語的系統先有個清楚的認識，並把他們描述清楚，加以形式化，然後才能以一定的方式交給電腦來處理。

本章只討論有關客家語的動詞短語 (VP) 的類型和 PP VP VP 切分的規則。

在我們的描述體系中，一般而言短語均可由單個詞直接上升得到，例如動詞「翕」，可以直接視為動詞短語「翕」；短語也可以由詞和詞的組合上升而得到，例如動詞「食」和名詞「飯」，可以組合得到動詞短語「食飯」，但像主謂短語、介詞短語、比況短語、附加短語等必定是以組合型式才能充當句子的成分，不能由詞類直接上升而構成短語，例如「韶早拜三愛考試了」，主謂短語「韶早拜三」在句中作主語，必定是以組合的方式構成；又如介詞「照、分、對」，必定不能單用作句子的成分，一定得構成介詞短語才能做句子的成分，如「照你講介、對佢來講、分佢罵」。

第一節 動詞短語 (VP) 的內部構造

考察一個語言單位的結構體可以從兩方面來看，一是向內觀察，即看該結構體內部的構造和組成成分之間的關係；二是向外觀察，考察其他成分和該結構體發生組合的關係。向內觀察無法了解該結構體在句法上的功能和限制，只能了

解該結構體內部成分之間的關係，向外觀察則能更全面地概括出該結構體的句法位置和功能。

下面我們就這兩種角度來考察客家語的動詞短語（VP），我們除了從語料庫中列出各種的 VP 加以考察之外，爲了盡量描述全部的 VP 情況，筆者採取個人內省（introspective judgement）的方式，必要時另外再造一些短語或句子來說明。

根據語料庫抽取出來的動詞短語（VP），我們觀察其內部組合的形式特徵，將 VP 分爲下面三組八個小類：（一）、帶有顯性標記的 VP（二）、固定組合的 VP（三）、不定組合的 VP。

（一） 帶有顯性標記的 VP

所謂顯性標記的 VP，是指語言中的某些或某類成分，必定以一定的方式與動詞發生結合所構成的 VP 而言，從功能上看「顯性標記」與動詞結合後主要用以表達「動詞」的某種語法意義，根據客家語的特性，我們把顯性標記的 VP，細分爲三個小類：

（1）、附加式 VP

指由「動詞或動詞短語」與用以分別動詞時態的成分「起來、了、等、過、忒、去了、去、過了、忒了」等所形成的 VP，因爲「起來、了、等、過、忒、去了、去、過了、忒了、著、啊忒」是個封閉集合，可以看做是客家語動詞短語 VP 的顯性標記，這些顯性標記一定是後置的，所以這種 VP，我們可以用規則描述出來。

| 序號 | 語法意義 | 組合規則 | 語料例 |
|----|------|-------------|---|
| 1 | 表起始 | vp~fg.vp.ax | 打/vp 起來/ax、笑/vp 起來/ax、啼/vp 起來/ax、罵/vp 起來/ax |
| 2 | 表進行 | vp~fg.vp.ax | 拿/vp 等/ax、擊/vp 等/ax、坐/vp 等/ax、 囤/vp 等 ax、放/vp 等/ax |
| 3 | 表完成 | vp~fg.vp.ax | 食/vp 過/ax、去/vp 過/ax、看/vp 過/ax、 做/vp 過/ax、罵/vp 過/ax |
| 4 | 表完成 | vp~fg.vp.ax | 看/vp 著/ax、敲/vp 著/ax、搨/vp 著/ax、 同/vp 著/ax |
| 5 | 表完成 | vp~fg.vp.ax | 食/vp 了/ax、拿/vp 了/ax、拂/vp 了/ax、 橫/vp 了/ax、斷/vp 了/ax |
| 6 | 表完成 | vp~fg.vp.ax | 食/vp 過了/ax、去/vp 過了/ax、打/vp 過了/ax、 豁/vp 過了/ax |
| 7 | 表完成 | vp~fg.vp.ax | 食/vp 忒/ax、過/vp 忒/ax、脫/vp 忒/ax、 豁豁/vp 忒/ax、擰擰/vp 忒/ax |
| 8 | 表完成 | vp~fg.vp.ax | 嚙斷/vp 忒了/ax、脫/vp 忒了/ax、辭/vp 忒 了/ax、割斷/vp 忒了/ax |
| 9 | 表完成 | vp~fg.vp.ax | 食/vp 啊忒/ax、摔/vp 啊忒/ax、放/vp 啊忒 /ax、擊/vp 啊忒/ax |
| 10 | 表過去 | vp~fg.vp.ax | 咬斷/vp 去了/ax、行遠/vp 去了/ax、拂忒 /vp 去了/ax |
| 11 | 表過去 | vp~fg.vp.ax | 打爽/vp 去/ax、食忒/vp 去/ax、毒死/vp 去 /ax、敲斷/vp 去/ax |
| 12 | 表完成 | vp~fg.vp.ax | 擄/vp 忒去/ax、走/vp 忒去/ax、死/vp 忒去 /ax、睡死/vp 忒去/ax、橫/vp 忒去/ax |

(表三十五，表動詞時態式的附加式 vp)

以上十一種是客家話裡用以表示動詞時態的附加式 VP，全部只需要一個規則「vp~fg.vv.ax」即可表述，另外還有四種並不表示動詞時態的附加式 VP，爲了使描述規則簡潔明白，我們也把他們歸入這種附加式 VP，這種附加式 VP 主要是由四種附加成分「看、啊、啊了、來」附著於動詞後構成：

| | | | |
|----|-----|-------------|---|
| 13 | 表嘗試 | vp~fg.vv.ax | 問先生/vp 看/ax、試食/vp 看/ax、摸/vp 看/ax |
| 14 | 表嘗試 | vp~fg.vv.ax | 划划/vp 看啊/ax、試食/vp 看啊/ax、摸/vp 看啊/ax |
| 15 | 表嘗試 | vp~fg.vv.ax | 問先生/vp 看啊嘞/ax、試食 vp 看啊嘞/ax、摸/vp 看啊嘞/ax |
| 16 | 表期望 | vp~fg.vv.ax | 帶著/vp 來/ax、食飽/vp 來/ax、拿便/vp 來/ax、掙平/vp 來/ax |

(表三十六，不表動詞時態式的附加式 vp)

以上四種客家語的附加式 VP 並不表示動詞的時態，但他們的內部結構關係我們認爲與規則「vp~fg.vv.ax」無異，所以併入附加式 VP 裡。

這些附加式 VP，有非常清楚的結構邊界可以辨識，這些結構邊界是「起來、了、等、過、忒、下去、去了、去、過了、忒了、看、啊、看啊嘞、來」，我們全部把他們劃入「助詞」詞類裡，他們並沒有實際的詞彙意義，只能與動詞結合產生特定的語法意義。

還有一種由動詞帶上方位詞，表示動詞發生的時間的 VP，如「泅水前、買

菜過後」等，這種「動詞和方位詞」的組合，我們的描述體系也畫入「附加式 VP」，單立一個規則來表述。

| 序號 | 組合規則 | 語料例 |
|----|-------------|-----------------------------------|
| 17 | vp~fg.vp.fu | 泅水/vp 前/fu、買菜/vp 過後/fu、楔入/vp 後/fu |

(表三十七，動詞帶方位詞的附加式 vp)

動詞後帶上的語氣詞，我們也畫入「附加式 VP」，用一個規則來表述。

| 序號 | 組合規則 | 語料例 |
|----|-------------|-------------------------------------|
| 18 | vp~fg.vp.sf | 食飽/vp 啊/sf、愛出門/vp 哩哦/sf、無早講/vp 呢/sf |

(表三十八，動詞帶語氣詞的附加式 vp)

「方位詞和語氣詞」都是有限的集合，動詞只能在後面帶上方位詞及語氣詞，構成「vp~fg.vp.fu」、「vp~fg.vp.sf」組合結構，所以我們可以根據這個特性，把方位詞及語氣詞作為結構邊界，據以辨認屬於哪一種 VP。

(2) 重疊式 VP

指由單一動詞以特定的方式重疊形成的 VP，「重疊格式」是一個顯性的格式，構成的 VP 也有很明確的結構邊界可以辨識，所以這種 VP，我們也可以用規則描述出來。

| 序號 | 組合規則 | 語料例 |
|----|------|-----|
|----|------|-----|

| | | |
|---|-------------------|---|
| 1 | vp~bl.vp | 行行/vp、看看/vp、嫵嫵/vp |
| 2 | vp~bl.vp | 行啊行/vp、瞧啊瞧/vp、祛啊祛/vp、摔啊摔/vp |
| 3 | vp~bl.vp | 跳啊跳仔/vp、行啊行仔/vp |
| 4 | vp~bl.vp | 來來去去/vp、出出入入/vp、行行坐坐/vp |
| 5 | vp~bl.vp.vp | 參詳/vp 參詳/vp、詐意/vp 詐意/vp |
| 6 | vp~bl.vp.vp.vp.vp | 揆/vp 揆/vp 揆/vp 揆/vp、划/vp 划/vp 划/vp、 罵/vp 罵/vp 罵/vp |

(表三十九，重疊式 vp)

客家話動詞的重疊根據我們的分詞標準定義一共有七種重疊格式，但是我們在進行字串內部構造分析時把「AA 看、AA 啊、AA 啊嘞」歸入「附加式 VP」，其他則歸入「重疊式 VP」，對客家語的研究，一般語法書對重疊的定義是限於前舉序號 1、4、5、6 這四種格式，但我們根據語料事實及規則清晰簡化的要求，將序號 2、3 這兩種格式也歸入「重疊式 VP」。

國立新竹教育大學
National Hsinchu University of Education

這些重疊式 VP 的各成分之間，我們定義他們是以「聯合（並列）」的方式組合，除序號 5、6 予以根據分詞標準切分之外，其他序號 1~4 這四種格式，我們視他們為一個動詞的形態變化，而「動詞的形態變化」仍是「動詞」，所以由詞類直接上升而構成短語。

(3) 離合式 VP

指某些動詞內部的語素間關係較為鬆散，可以插入「助詞、數量詞」等成分，而使原有的語素造成分離形式的 VP。這種 VP 具有很明確的結構邊界可以辨識，我們也可以用規則描述出來。

| 序號 | 組合規則 | 語料例 |
|----|-------------|---|
| 1 | vp~fg.vv | 洗過身/vp、寫過字/vp、翻等菜/vp、食了飯/vp |
| 2 | vp~fg.vv | 祛啊好/vp、粹啊淨/vp、晃啊來/vp、豁啊去/vp、行遠 啊去/vp |
| 3 | vp~vo.vp.np | 睡/vp 一覺目/np、著/vp 二領衫/np、撈/vp 一頭草/np、 拏/vp 一日草/np |

(表四十，離合式 vp)

客家話的離合式 VP 根據語料來看似乎只有這三種，原式動詞兩語素之間若是「動賓」格式者，如「洗身、寫字、翻菜、食飯」等，中間只能插入助詞「了、等、過」三者之一，構成原式動詞兩語素之間分離，我們視這種被助詞分離的結構體仍為一個「動詞」。

而原式動詞兩語素之間若是「述補」格式者，如「祛好、粹淨、晃來、豁去」等，中間只能插入語素「啊」，構成原式動詞兩語素之間分離，表達與原動詞完全不同的語義，我們視這種被助詞分離的結構體仍為一個「動詞」。

另外原式動詞兩語素之間若是「動賓」格式者，如「睡目、著衫、撈草、拏草」等，中間只能插入與賓語素相應的「數量」結構，如「著二領衫、撈一頭草」，數量結構與賓語素先結合構成一個 NP 結構體後，再與動詞或動詞短語結合，構成一個「vv-cd-m-nn」順序的字串序列，形成一個被數量結構分離的 VP。

以上大體而言就是台灣客家話帶有顯性標記的 VP 的內部構造，從內部構造

規則來看只需要四個規則便可以管住這些帶有顯性標記的 VP：「vp~fg.vv.ax、vp~bl.vv、vp~fg.vv、vp~vo.vv.np」，再嚴格定之只需要兩個規則「vp~fg.vv.ax、vp~vo.vv.np」即足足有餘，因為「vp~bl.vv、vp~fg.vv」這兩者實際上在我們的描述體系裡被視為一個詞看待，這種動詞短語是被允許由一個詞類上升而得到的。

在規則「vp~fg.vv.ax」裡，我們需要確認和辨認的是構成「ax」（助詞）的元素是哪些？又 ax 如何與 vv 發生關係？需要描述發生了哪些關係？更要描述不能發生哪些關係？把這些知識描述清楚便能交給計算機進行我們的工作方案、達成我們的目標。而在規則「vp~vo.vv.np」裡，我們則需要描述「數量結構」裡的量詞和賓語素之間的選擇性限制，即「vv 和 np」之間的關係，但這些都超出本節甚至本文的範圍，筆者暫時略去，僅就結構體的組成份來探討內部構造。

（二） 定項式 VP

這類 VP 內部組合成分的項數是有限的，都為兩項的組合，有些在兩項之間可以有「得、到、毋」等，包括下面三類結構方式。

（4） 述補式

動詞帶補語一般語法書分為「黏合式述補」、「組合式述補」兩種，單個動詞帶補語的屬黏合式述補 VP，是無標記式的；而有標記式的屬組合式述補 VP，主要的標記是「得、到、毋、啊」等。

| 序號 | 組合規則 | 語料例 |
|----|-------------|--|
| 1 | vp~vc.vp.vp | 擎/vp 出去/vp、綯/vp 起來/vp、扛/vp 轉去/vp、豁/vp 過去/vp、行/vp 落去/vp |

| | | |
|---|-------------|---|
| 2 | vp~vc.vp.jp | 掙淨/vp、撇/vp 弦弦/jp、撈/vp 淨淨/jp、糊/vp 塞塞/jp、 擇/vp 淨淨/jp |
| 3 | vp~vc.vp.pp | 囡/vp 在後背/pp、行/vp 在頭前/pp、睡/vp 到地泥項 /pp、掛/vp 到壁頂/pp |
| 4 | vp~vc.vp.np | 讀/vp 一遍/np、行/vp 三擺/np、戴/vp 兩年/np |

(表四十一，述補式 vp)

序號 1 的規則是「vp~vc.vv.vv」，定義以「趨向動詞和動詞」充當補語的 VP。客家話的述補短語 V-V 關係，做為補語的 V 由「趨向動詞和動詞」構成，根據我們的分詞標準雙音節的 V-V，我們定義為一個詞，如「打走、擊來、拿去、走開、搥開、撚死、奮開」等，這些雙音節詞的構成成分也都是由一個單音節動詞和一個單音節的「趨向動詞或動詞」組成，依據動詞短語 VP 可以直接由動詞上升得到的做法，並不違背整個短語系統的體系。

客家話的趨向動詞是一個有限的集合，可分為簡單的單音節趨向動詞，如：「來、去、上、下」；及由單音趨向動詞複合而成的複合趨向動詞，如：「上來、下來、入來、落來」等，如下表：

| | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|----|
| | 上 | 下 | 入 | 落 | 出 | 轉 | 過 | 開 | 到 | 起 |
| 來 | 上來 | 下來 | 入來 | 落來 | 出來 | 轉來 | 過來 | 開來 | 到來 | 起來 |
| 去 | 上去 | 下去 | 入去 | 落去 | 出去 | 轉去 | 過去 | | | |

(表四十二，客家話的趨向動詞表)

這些趨向動詞便可視為 VP 的短語邊界，作為區分的準則，而一般以動詞為補語構成的述補式 VP，則用直接由動詞上升得到的做法去規範，這麼一來我們只需要一個規則「vp~vc.vv.vv」便可管住客家話 V-V 式的述補短語。

序號 2 的規則是「vp~vc.vv.jp」，定義以「形容詞」充當補語的 VP。同樣根據我們的分詞標準把雙音節的述補結構一律定義為一個「詞」的層次，並可直接上升為短語 VP 的做法，我們可以得到「vp~vc.vv.jp」規則便可管住這類的短語，並不違背整個短語系統的體系。

序號 3 是以介詞短語為補語的情形，這些介詞短語一般都是表示空間位置的，所以整個述補短語也大都表示空間位置，我們只需要一個規則

「vp~vc.vv.pp」，便可描述這種短語 vp。

以上是無標記的黏合式補語述補短語組合規則，其中序號 1、2 也可以在述語之後加入標記「得、到、毋、啊」之一，構成「組合式述補」。

| 序號 | 組合規則 | 語料例 |
|----|----------------|---|
| 1 | vp~vc.vp.ax.vv | 擎/vp 得/ax 出去/vv、綯/vp 毋/ax 起來/vv、扛/vp 毋/ax 轉去/vv、豁/vp 啊/ax 過去/vv、行/vp 啊/ax 落去/vv |
| 2 | vp~vc.vp.ax.jp | 摔/vp 得/ax 淨/jj、脫/vp 到/ax 光光/jj、黏/vp 到/ax 核核/jj、撇/vp 毋/ax 弦/jj、擄/vp 毋/ax 淨/jj |
| 3 | vp~vc.vp.ax.vp | 打/vp 到/ax 喊毋敢/vp、噉/vp 到/ax 無人識/vp |
| 4 | vp~vc.vp.ax.cp | 打/vp 到/ax 佢喊毋敢/cp、咬/vp 著/ax 佢歸身血/cp |

以上是客家話裡以「得、到、毋、啊」為組合標記的述補式 VP，只需要「vp~vc.vv.ax.vv」、「vp~vc.vv.ax.jp」、「vp~vc.vv.ax.vp」、「vp~vc.vv.ax.cp」四個規則並以「得、到、毋、啊」做為結構邊界，考察後邊的字串狀況，便可定義出相應的述補短語 VP。尚有待努力的是需要對述補結構裡的補語部分和述語部分的選擇性關係加以描述清楚，除需要描述發生了哪些關係之外，更要描述不能發生哪些關係等。

(5) 述賓式 VP

指由述語 VP 帶上 np、jp、vp、cp 等不同類型的賓語所構成的 VP 而言，例如：

| 序號 | 組合規則 | 語料例 |
|----|--------------|--|
| 1 | vp~vo.vp.np | 講/vp 麼介/np、擇/vp 蕃薯葉/np、招/vp 佢/np、破/vp 西瓜/np、擊/vp 三斤/np、刷/vp 一尾蛇/np、拂/vp 三碗飯/np |
| 2 | vp~ vo.vp.jp | 想/vp 清楚/jp、好/vp 淨利/jp、搵/vp 清楚/jp、抹/vp 屙糟/jp |
| 3 | vp~ vo.vp.cp | 想/vp 阿公來嫖/cp、拜託/vp 若舅過去/cp、答應/vp 阿哥去大陸/cp、趨/vp 阿叔去/cp |
| 4 | vp~vo.vp.vp | 想/vp 咬芭樂/vp、想/vp 食鮮糜/vp、肯/vp 借繫/vp |

客家話的述賓式 VP 中的賓語，以名詞性的名詞或名詞短語為主，其他如 jp、vp、cp 也都可以做爲述賓式 VP 中的賓語，但不若名詞性的名詞或名詞短語普遍。

(6) 狀中式 VP

指中心語是 VP，狀語由 pp、rp、jp、np (ti、si) 等充當，所構成的 VP 而言，其中 np，僅限時間詞短語 (ti)、處所詞短語 (si) 例如：

| 序號 | 組合規則 | 語料例 |
|----|-------------|---|
| 1 | vp~zd.pp.vp | 分佢/pp 趨走/vp、搵書/pp 拿走/vp、分阿叔/pp 擊去/vp、搵蛇/pp 打死去/vp |
| 2 | vp~zd.rp.vp | 堵/rp 走/vp、當/rp 食等飯/vp、這下還冇/rp 來/vp |
| 3 | vp~zd.ts.vp | 晚晡頭/ts 看/vp、韶下晝/ts 去/vv、明年/ts 掛紙/vv |

| | | |
|---|-------------|--|
| 4 | vp~zd.jp.vp | 艱苦/jp 賺/vp、認真/jp 學/vp、好/jp 搞/vv、大/jp 譴/vv |
| 5 | vp~zd.ss.vp | 灶下角/ss 放/vp、廳下/ss 坐/vp、屋肚/ss 僚/vp、間肚/ss 搞/vp |

以上是客家話的定項式 VP 的分類和組合規則，大體可以分為述補式、述賓式和狀中式三種，各種之間又可以細分為幾個小類，這些小類之間的分合和組合成分之間的關係，都很值得繼續做深入的研討，礙於時間和本文的範圍，留做日後繼續努力。

(三) 不定多項組合 VP

這類 VP 的內部構造是由非固定向次成分所組成的 VP，我們的描述體系把動詞性聯合短語和連動短語畫入「不定多項組合 VP」，從技術處理的角度，我們用兩項組合來表述他們的組合規則，也就是說我們把多項並列分析為兩項組合的套疊構造。

(7) 聯合式 VP

聯合式 VP 包括兩種，一種是以「搵、抑、或、或者」等連接詞為標記的聯合式 VP；另一種是無標記的聯合式 VP，例如：

| 序號 | 組合規則 | 語料例 |
|----|----------------|---|
| 1 | vp~bl.vp.cc.vp | 唱歌/vp 搵/cc 跳舞/vp、搵/vp 搵/cc 敲/vp、去/vp 抑/vv 轉/vp、 |

| | | |
|---|-------------|---|
| 2 | vp~bl.vp.vp | 唱歌跳舞/vp 泅水打球/vp、 <去台北/vp 去花蓮/vp>/vp 去台中/vp |
|---|-------------|---|

(8) 連動式 VP

從結構上看，連動式 VP 和無標記的聯合式 VP 是同形的：

| 序號 | 組合規則 | 語料例 |
|----|-------------|--------------------------------|
| 1 | vp~ld.vp.vp | 打電話/vp 請醫生/vp，出去/vp 啣蟲/vp 轉/vp |

上面對客家語的 VP 內部構造做了一個大致的、粗略的分類描寫，這是根據我們分析的需要及分析的結果所得到的，當然不同的目的和分析方法會得到不同的結果，這是不言自明的。其實每個類都可以在細分為更小的類，而這裡我們也不對每一個類內部的差別作描寫，僅就「組合規則」這方面做分析。

國立新竹教育大學
National Hsinchu University of Education

第二節 動詞短語 (VP) 的外部功能

上一節描述了客家語動詞短語 (VP) 的內部構造，也涉及了一些關於 VP 的句法位置，例如 VP 會在述補和述賓構造中佔據「述語中心」這個位置，而在狀中結構的 VP 中佔據中心語位置等，這裡接下去要對 VP 的外部功能加以描寫，這裡所謂外部功能是指 VP 能佔據哪些句法位置的意思，我們仍然用「組合規則」來表述。

(一) VP 佔據主謂短語裡的謂語位置

| 序號 | 組合規則 | 語料例 |
|----|-------------|---|
| 1 | cp~zu.np.vp | 阿公/np 去台北/vp、 先生/np 盲轉/vp、 佢兜/np 無行了/vp |
| 2 | cp~zu.ts.vp | 今晡日/tp 係拜幾/vp、 八月十五/tp 愛轉家鄉去/vp |
| 3 | cp~zu.ss.vp | 灶下/ss 毋好罔/vp、 洗身間肚/ss 做毋得放瓦斯桶/vp |
| 4 | cp~zu.jp.vp | 靚/jp 係儕儕愛/vp、高毛/jp 係你/vp |
| 5 | cp~zu.cp.vp | 手帕屙糟/cp 放洗衫機肚去/vp、 手烏烏/cp 遽遽去洗/vp |

國立新竹教育大學
National Hsinchu University of Education

(二) VP 佔據主謂短語的主語位置

| 序號 | 組合規則 | 語料例 |
|----|-------------|------------------------------|
| 1 | cp~zu.vp.ap | 做飯菜/vp 當麻煩/ap、打拼/vp 盡重要/ap |
| 2 | cp~zu.vp.vp | 洗身/vp 毋好搞水/vp、去睡目/vp 毋好吵人/vp |

(三) VP 佔據名詞短語的定語位置

| 序號 | 組合規則 | 語料例 |
|----|------|-----|
|----|------|-----|

| | | |
|---|-------------|-------------------------------------|
| 1 | np~dd.vp.np | 駛車/vp 技術/np、泅水/vp 功夫/np、掛紙/vp 日仔/np |
|---|-------------|-------------------------------------|

(四) VP 佔據名詞性的「介」字結構的核心位置：

| 序號 | 組合規則 | 語料例 |
|----|---------------|-----------------------------------|
| 1 | np~kai.vp.kai | 食/vp 介/kai、搥斷/vp 介/kai、擗/vp 介/kai |

以上是客家語 VP 參與與其他成分組合時的情況，在上一節我們描述 VP 的內部結構時都是用 vp 做為組合規則描述的起點，而這裡描述外部功能時都是用 cp、np 等，把這兩部分組合起來就大致可以反應客家語 VP 的面貌。例如：「一橫下去斯睡忒了」，這是由兩個 VP 構成的，我們用以上的規則來表述「一橫下去/vp 斯睡忒了/vp」，再觀察 vp 的內部結構，即可得到「一橫下去/vp~zd.1rp.3vp 斯睡忒了/vp~1rp.3vp」；再繼續往下一層觀察即可得到整個 VP 的組合結構：

(<-/rb 橫下去/vp~vc.1vv.2vv>/vp~zd.1rp.1vp <斯/rb 睡忒了
/vp~fg.1vv.2ax>/vp~zd.1rp.1vp)/vp~ld.1vp.1vp

又如：

「阿公一橫下去斯睡忒了」，這句便是「cp~zu.np.vp」的組合結構，是「<阿公/np 一橫下去斯睡忒了/vp>/cp~zu.1np.1vp」的結構。

第三節 VP 內部結構關係判定的問題

(一) 賓語和補語的識別方法

客家語的賓語和補語都位於述語之後，分別構成述賓短語和述補短語，而且多有相似之處，如「讀三本」和「讀三擺」，這兩個短語都是「動詞—數詞—量詞」的序列結構，而實際上「讀三本」是述賓短語，「讀三擺」卻是述補短語，區分賓語還是補語可以用二種方法來檢測：

(1)、用功能來區分

一般而言作為賓語的大多是體詞或體詞短語，而做補語的大多數是謂詞或謂詞短語。所以像這樣「動詞—數詞—量詞」的序列結構，如果「數詞—量詞」是指稱事物的數量時，則它是「賓語」，如果是指稱動作發生的次數時，則它是「補語」。

例如「讀三本書」，「三本」是指稱書，所以是「賓語」，用規則來表述便是「vp~vo.vv.np」，這種屬賓語式的「數詞—量詞」結構，一定只能前置於名詞之前；而「讀三擺」中的「三擺」是指稱「讀」的動作發生三次，所以是「補語」，用規則來表述便成為「vp~vc.vv.np」，這種屬賓語式的「數詞—量詞」結構，一定只能前置於述語動詞之後，兩者之差是「vo」還是「vc」。

(2) 以能否用介詞「搵」把述語後的詞提前來判定

這個方法可以區分「數詞—時量詞」是賓語還是補語的問題，例如「打爽兩點鐘」和「歇睏兩點鐘」這兩個短語：

「打爽兩點鐘」可以變換成「搵兩點鐘打爽忒去了」；但是「歇睏兩點鐘」卻不能變換成「*搵兩點鐘歇睏忒去了」。所以「打爽兩點鐘」是述賓短語，用規

則來表述是「vp~vo.vv.np」；而「歇睏兩點鐘」是述補短語，用規則來表述是「vp~vc.vv.np」。

(二) 關於「vp~ vo.vv.cp」組合規則

「vp~ vo.vv.cp」組合規則，一般語法書多列為「兼語短語」，在我們的描述體系裡取消了兼語短語，一般語法書對兼語短語的分析是：

動詞 1-----名詞（賓語）

名詞（主語）----動詞 2

把兼語短語視為由一個述賓短語和一個主謂短語套疊而成的更大單位的短語，但是我們不做這樣的分析，我們的分析圖式如下：

動詞 1-----名詞

名詞（主語）----動詞 2（主謂、賓語）

把「名詞和動詞 2」的組合視為一個結構完整的主謂短語，同時是「動詞 1」的賓語，試比較兩種不同的分析結果的標記，以「喊佢討婦娘」為例：

1、如果分析為兼語短語時的標記是：

<喊佢/vp~vo.1vv.1pn 討婦娘/vp~vo.1vv.2nn>/vp~ky.1vp.1vp

2、如果分析為述賓短語時的標記是：

(喊/vv <佢/pn 討婦娘/vp~vo.1vv.2nn>/cp~zu.1np.1vp)/vp~vo.1vp.1cp

分析爲述賓短語時雖然增加了一個層次，但是卻維持了述賓短語的組合規則，不必另立一個兼語短語的組合規則，同時也符合主謂短語可以做賓語的功能，但是如果照一般語法書的兼語短語的分析，則我們必須另增一條

「vp~ky.lvp.lvp」的組合規則，增加分析上的困擾與不便，所以我們對兼語短語的組合規則一律遵循「vp~ vo.vv.cp」的組合規範來做標記。

第四節 PP VP VP

因爲我們還沒有能力研究與描述清楚關於客家語全部的VP，所以本節我們舉客家語PP VP VP的組合例子來歸納客家語VP的一小部分語言知識，嘗試用這個例子來說明結構層次、結構關係判定和歧義的問題，限於能力及時間，我們把範圍縮小到PP只限於「分」字的情況，以下如果沒有特別說明時，凡是PP處都是指「分」¹字而言，請看語料例句：

- 1、分人/pp 開來/vp 種香茅/vp
- 2、分法院/pp 罰錢/vp 拍賣/vp
- 3、分阿叔/pp 偷搬/vp 出去/vp
- 4、分佢/pp 刷著/vp 出大血了/vp
- 5、分佢/pp 奮走/vp 無拂著/vp

上面 5 個例句，都是 PP VP VP 模式排列的字串，我們只考慮這 5 種例句的情況，其他的情況我們在語料庫裡也尚未發現，這 5 種例句的情況當然也不代表客家話 PP VP VP 模式排列的全部，但爲使問題簡化，並基於爲將來計算

¹ 這裡所指的「分」字，是指構成被動式的「分」，相當於華語的被動式的「被」，我們說「相當」，意味者只是某些部分的相同，並不等於完全相同，詳見下文分析。

機自動處理提供規則的目的，我們先考慮普遍性和一般性，對於個別的或整體的均有待長遠繼續努力。

以上面 5 個例句為例，他們的結構層次和關係的語言知識，究竟是 a 或 b，哪一種模式？

a、(PP <VP VP>)模式

b、(<PP VP> VP)模式

對人而言例句 1、2、3，採用 a 模式切分，4、5 用 b 模式切分，理解上是沒有問題的，但其間的關係其實也有差異。例如例 1 中的 vp 和 vp 是連動關係；例 2 中的 vp 和 vp 是聯合關係；例 3 中的 vp 和 vp 是述補關係；例 4 中的 vp 和 vp 是連動關係；例 5 中的 vp 和 vp 是述補關係，所以不論是用 a 模式還是 b 模式切分，即使是用相同的切分方式去定義他們，結構關係也還是不相同，這些複雜的切分和關係之間的對應要讓計算機理解是件不容易的事，需要人類提供更為嚴苛的限制條件，而嚴苛的條件其實意味著精準判斷的提高。

我們認為在語言結構中，構成短語兩項組合是最普遍的方式，多項組合其實可以分解為兩項組合的擴大結構，所以我們在分析 PP VP VP 的組合模式時採取兩項組合的看法，當然在 PP VP VP 的組合裡只有 PP VP 兩項組合和 VP VP 兩項組合，所以我們必須把能和不能夠成 PP VP 兩項組合及 VP VP 兩項組合的限制條件歸納出來，這麼一來便可以解決 PP VP VP 三項組合的歧義問題。所以就 PP VP VP 的組合而言，我們的問題其實是：

- (1) 若是 PP VP 兩項先組合，則 VP 必須滿足哪些條件？
- (2) 若是 VP VP 兩項先組合，構成不同結構關係的限制條件是什麼？

(一) PP VP 先組合中 VP 必須滿足的條件

考察「分」字和 VP 的組合有什麼條件限制，可以從兩個方向出發，一個是考察什麼樣的 VP 不能和「分」字組合，另一個是看什麼樣的 VP 可以和「分」字組合，下面分別從這兩個角度來考察「分」字和 VP 的組合的條件限制。

(1) 不能和「分」字直接組合的動詞

我們一共考察了 1299 個單音節動詞和 1000 個雙音節動詞，發現不論單音節或雙音節均不能直接和「分」字直接組合，構成 PP VP 組合模式，例如：

「*分罵、*分打 1、*分丟、*分敲、*分調整、*分恐嚇、*分了解」等這些全部是不合客家語語法的，但前舉各例在現代華語都是可以直接和相當於客家語的「被」字組合的，如「被罵、被打、被丟、被敲、被調整、被恐嚇、被了解」等。

我們在這裡只考察一個動詞是否有跟「分」字組合的能力，而不對這種能力作出的解釋，根據考察的結果，我們可以為客家語的每一個單音節動詞貼上一個標籤，來標記它和「分」字組合的能力，如果計算機也掌握了這種知識自然就不會對單音節動詞做出「<PP VP> VP」這種模式的切分，因為這種與單音節動詞的組合在客家語根本就不存在。下面我們簡單分類，用列舉的方式列出不能和「分」字直接組合的動詞。

| | | |
|-----|-------|-------------------------|
| (1) | 連系動詞： | 係、姓、安、安做、變做、當做、當、像 |
| (2) | 助動詞： | 會、肯、做得、可能、可以、應該、需要、愛、甘願 |

| | | |
|------|-----------|----------------------|
| (3) | 趨向動詞： | 落、出、轉、上去、下去、入去、落去、出去 |
| (4) | ~得： | 看得、曉得、扛得、擎得、打得、敲得 |
| (5) | 主語為複數的動詞： | 會面、相罵、開扁、交換、 |
| (6) | 有 | 有、無 |
| (7) | 心理動詞 | 好、合意、憨想、想、嫌、保證、愛、愁 |
| (8) | 某些動賓格動詞 | 擦面、行禮、出力、吊頸、回神、守節 |
| (9) | 某些不及物動詞 | 行、走、噉、睡、撩、恹、驚、徙 |
| (10) | 其他 | ten11 手、收淨、洗好、承蒙 |

以上的分類並不是依據什麼標準來分類的，只是列舉以說明不能和「分」字組合的動詞，使用「某些」來表述，也是因為我們並沒有全面考察。從上面所列來看不能和「分」字組合的動詞範圍很廣，我們只單純描寫一個動詞是否跟「分」字有組合的能力，並不打算對這些動詞做進一步的概括和提出解釋。

從這個考察裡我們可以看出，如果能精確地判斷一個 VP 能否跟「分」字組合，那對判定「PP VP VP」組合的結構成次是相當有用的。接著我們便需來一一考察各種複雜 VP 結構和「分」字組合的限制。

(2) 不能和「分」字組合的 VP

「分」字不僅僅對動詞有限制作用，對於作為句子成份的 VP 也有一些限制，如果我們能把這些限制找出來，對結構的切分和結構成分關係的判定很有裨益。

1、狀中式 VP

客家話的「狀中式 VP」，根據我們的組合規則一共有「vp~zd.pp.vp、vp~zd.rp.vp、vp~zd.ts.vp、vp~zd.jp.vp、vp~zd.ss.vp」五種表述式，但只有「vp~zd.pp.vp」組合規則可以和「分」字組合，構成「PP VP VP」，其他均無法構成「PP VP VP」。請看以下各例：

| | |
|-------------|--------------------|
| 分佢/pp 趨走/vp | 分佢/pp 趨走/vp 追出去/vp |
| 盲/rp 來/vp | *分盲/rp 來/vp |
| 韶下晝/ts 去/vv | *分韶下晝/ts 去/vv |
| 大/jp 譴/vv | *分大/jp 譴/vv |
| 廳下/ss 做/vp | *分廳下/ss 做/vp |

我們也發現狀中式 VP，如果狀語由否定副詞「無、毋、盲、莫」擔任時，所構成的狀中結構，完全不能被「分」字修飾，如：

毋打 *分毋打 ； 盲食 *分盲食

客家話「分」字的否定是在整個結構之外做否定，如「無分佢打著、盲分佢食著」，不能在結構之內。

2、述補式 VP

客家語的「述補式 VP」根據我們的組合規則無標記的黏合式述補短語一共有「vp~vc.vp.vp、vp~vc.vp.jp、vp~vc.vp.pp、vp~vc.vp.np」四種表述式，有標記的組合式述補短語有「vp~vc.vp.ax.vp、vp~vc.vp.ax.jp、vp~vc.vp.ax.vp、vp~vc.vp.ax.cp」，全部都可以和「分」字組合構成「PP VP VP」，請看以下各

例：

| | | |
|-------------------|--------|-------------------------|
| 擎/vp 出去/vp p | -----> | 分佢/pp 擎/vp 出去/vp |
| 揸/vp 弦弦/jp | -----> | 分佢/pp 揸/vp 弦弦/jp |
| 掛/vp 到壁頂/pp | -----> | 分佢/pp 掛/vp 到壁頂/pp |
| 看到/vp 三擺/np | -----> | 分佢/pp 看到/vp 三擺/np |
| 豁/vp 啊/ax 過去/vp | -----> | 分佢/pp 豁/vp 啊/ax 過去/vp |
| 脫/vp 到/ax 光光/jj | -----> | 分佢/pp 脫/vp 到/ax 光光/jj |
| 打/vp 到/ax 喊毋敢/vp | -----> | 分佢/pp 打/vp 到/ax 喊毋敢/vp |
| 打/vp 到/ax 佢喊毋敢/cp | -----> | 分佢/pp 打/vp 到/ax 佢喊毋敢/cp |

但是我們也發現述補短語 VP，一但在述語和補語之間帶上否定副詞「無、毋、旨」或在述語之前帶上「莫」，就不能和「分」(被)字組合，也就是說否定式的述補式 VP，不能和「分」字組合，如：

分佢/pp 揸/vp 弦弦/jp *分佢/pp 揸/vp 毋/rp 弦弦/jp

(1) 述賓式、連動式、聯合式 VP 中的任意一項，只要含有否定標記嵌套在 VP 裡，就不能和「分」(被)字組合，例如：

述賓式：*分佢/pp 擇/vp 無蕃薯葉/np

聯合式：*分佢/pp 毋唱歌/vp 無跳舞/vp

連動式：*分佢/pp 無打電話/vp 請醫生/vp

綜合上面的分析可以看出，客家語裡頭做為被動的「分」字，只要 VP 中內

嵌套有否定成分，它就不能和「分」字組合，當然也就構不成「PP VP VP」。
 所以我們很清楚的便可以判定像：

「分大門/pp 擋等/vp 無射著目珠/vp」

這個「PP VP VP」結構，只能按 b 式切分為：「(<分大門擋等> 無射著目珠)」。
 這些分析知識清楚的告訴我們如果人們能掌握究竟哪些形式的 VP 不能和
 「分」字組合，對確定「PP VP VP」組合的結構層次是非常有幫助的。

(二) 滿足一定條件下才能和「分」字組合的 VP

(1) 附加式 VP

在我們的描述體系裡共有 18 種附加式 VP，當中只有以「過、過了、忒、忒了、著、啊忒、下去、去了、去」為 VP 標記的附加式 VP 可以和「分」字組合，
 例如：

| | | |
|-------------|--------|-------------------|
| 罵/vp 過/ax | -----> | 分佢/pp 罵/vp 過/ax |
| 豁/vp 過了/ax | -----> | 分佢/pp 豁/vp 過了/ax |
| 擰擰/vp 忒/ax | -----> | 分佢/pp 擰擰/vp 忒/ax |
| 割斷/vp 忒了/ax | -----> | 分佢/pp 割斷/vp 忒了/ax |
| 掙/vp 啊忒/ax | -----> | 分佢/pp 掙/vp 啊忒/ax |
| 舞/vp 下去/ax | -----> | 分佢/pp 舞/vp 下去/ax |
| 咬斷/vp 去了/ax | -----> | 分佢/pp 咬斷/vp 去了/ax |
| 打爽/vp 去/ax | -----> | 分佢/pp 打爽/vp 去/ax |

| | | |
|-----------|--------|-----------------|
| 看/vp 著/ax | -----> | 分佢/pp 看/vp 著/ax |
|-----------|--------|-----------------|

以上我們僅列出可以和「分」字組合的附加式 VP，這似乎是和動詞本身的性質及意義有關，而不是和附加式標記有關，但我們並沒有深入考察，比較可能而準確的做法是設立一個「分」字組合的屬性標籤，考察每一個動詞在附加式結構中與「分」字組合的能力。

(2) 重疊式 VP

在我們的描述體系裡共有 6 種重疊式 VP，但只有 3 種重疊式 VP 可以和「分」字組合，例如：

| | | |
|----------------|--------|----------------------------|
| 祛啊祛/vp | -----> | 分佢/pp 祛啊祛/vp 祛淨淨去/vp |
| 行啊行仔/vp | -----> | 分佢/pp 行啊行仔/vp 行到台北了 /vp |
| 划/vp 划/vp 划/vp | -----> | 分佢/pp 划划划/vp 划著淨淨/vp |

重疊式 VP 和「分」字組合後一般都是黏著的 VP，後面必須帶著另一個成分（一般也是一個 VP），做為補語，構成「PP VP VP」。

(3) 離合式 VP

可以離合的動詞一般是由「述語動詞和賓語」所構成的述賓構造，離合式 VP 一般也都有與之相應的緊湊（指未被離合前）動詞形式，這類的述賓構造動詞因為有賓語成分存在，所以不太能跟「分」字組合，例如

| | | |
|-------|--------|--------------|
| 洗身/vp | -----> | *分佢/pp 洗身/vp |
| 搽淨/vp | -----> | *分佢/pp 搽淨/vp |
| 著衫/vp | -----> | *分佢/pp 著衫/vp |

但是離合之後的離合式 VP 卻可以和「分」字組合，例如：

| | | |
|-------------|--------|--------------------|
| 洗過身/vp | -----> | 分佢/pp 洗過身了/vp |
| 搽啊淨/vp | -----> | 分佢/pp 搽啊淨/vp 就~ |
| 著/vp 二領衫/np | -----> | 分佢/pp 著二領衫/vp 轉/vp |

離合式 VP 和「分」字組合似乎沒有什麼結構上的特別要求，我們比較可能的做法是設立一個「分」字組合的屬性標籤，考察每一個述賓構造的動詞在離合之後與「分」字組合的能力。

(4) 述補式 VP

述補式 VP 內部構造與「分」字組合的情形，已如上文所述，上文我們已將帶否定標記的述補式 VP 和「分」字組合的可能性排除出去了，而根據我們的考察一般簡單形式的述補式 VP 動詞，大都不能和「分」字直接組合，如：

| | | |
|-------------|--------|----------------|
| 洗淨/vp | -----> | *分洗淨/vp |
| 糊紅紅/vp | -----> | *分糊紅紅/vp |
| 著/vp 二領衫/np | -----> | *分佢/pp 著二領衫/vp |

對於述補式 VP，值得一提的是一般都能和「分」字組合，特別是在表示趨向位移義時。但如果不是表示實際的趨向位移義，而是表示動作的持續義時，我們發現跟「分」字組合的能力就會受影響，比如「分佢兜拖下去」，這是一個歧義短語，既可表示位移義，也可以表示動作持續，當表示動作持續的時候，這個「分」的詞義相當於華語的「讓」而不是被動式的「被」。這說明「分」字的組合對趨向述補 VP 是有選擇性限制的，導致這樣結果的因素，似乎和動詞及趨向動詞的語義有關，關於這種歧義的現象，本文留待日後做詳細討論。

(5) 述賓式 VP

述賓式 VP 不能直接和「分」字組合，要夠成被動式時，採取的是把賓語提前的方式，當賓語被提前放在主語位置後，一般而言述語動詞不再帶賓語，但有幾種例外，如：

| | | |
|-------|--------|--------------|
| 買菜/vp | -----> | *分佢/pp 買菜/vp |
| 出汗/vp | -----> | *分佢/pp 出汗/vp |
| 洗衫/vp | -----> | *分佢/pp 洗衫/vp |

以上都是述賓式 VP 不能直接帶上「分」字，組合成被動式情形，例外的情形有：

首先如果賓語是表示「數量成分」時，一般都能跟「分」字組合。如：

「十粒芭樂分佢食二粒去」、「分先生看著二擺」

當賓語成分的語義和「分」字前面的主語成分有領屬或整體與部分關係時，一般也可以和「分」字組合。如：

「台灣銀行分賊仔搶忒當多錢」（「銀行」領屬「錢」）

「厥公分福佬攞改忒好多錢銀」（「厥公」領屬「錢銀」）

「厥胃分醫生割忒一大坵」（「整個胃」和「一部分胃」是整體和部分關係）

「三斤分佢分忒兩斤較加」（「叁斤」和「兩斤」是整體和部分關係）

綜合上面粗略的分析可知述賓式 VP 要跟「分」字組合，在句法和語義上都要受到一些限制條件的要求，這些條件是什麼？比較可能的做法是設立一些屬性值去規範述語動詞，並一一加以考察。

(6) 連動式 VP

連動式 VP 要能與「分」字組合形成被動式，我們考察的結果是必須滿足一個條件，即連動式裡的後項 VP 必須能跟「分」字組合，那麼後項裡的 VP 的各種限制條件就可以按照前面所探討的種種情況來加以約束和判斷，例如：

「該憨狗/np 分佢/pp 拿鍊仔/vp 綯等了/vp」

後項「綯等了/vp」可以滿足與「分」字組合的條件，所以這個連動式 VP 才可以和「分」字組合。

上面我們對各類型的 VP 能和「分」字組合的各種選擇和約束條件進行了粗略的分析，歸納出來的語言知識，我們可以將之形式化，幫助計算機對 PP VP VP 做出正確的判斷。例如：

(a) (該憨狗/np) 分佢/pp 拿鍊仔/vp 綯等了/vp

上面這個例句是 PP VP VP 序列的結構，要按「<PP VP> VP」還是「PP <VP VP>」切分，我們立刻可以根據上面的分析判斷出來，因為「拿鍊仔/vp」這個 VP 是個述賓短語，不能和「分」字組合；而後一個「綯等了/vp」是個述補式的附加結構 VP，滿足和「分」字組合的制約條件，所以很清楚的是應該按「PP <VP VP>」切分，因此它的結構層次是：

分佢/pp <拿鍊仔/vp 綯等了/vp>，根據標記規範標出的結果是：

(分佢/pp~po.1pa.1pn <拿鍊仔/vp~vo.1vv.2nn 綯等了
/vp~fg.2vv.1ax>/vp~ld.1vp.1vp)/vp~zd.1pp.1vp

再如：

(b) 分佢/pp 奮走/vp 無拂著/vp

「分佢/pp 奮走/vp 無拂著/vp」這個例子，也是 PP VP VP 序列的結構，要按「<PP VP> VP」還是「PP <VP VP>」切分，我們也可以立刻判定，因為「奮走/vp」這個 VP 是個述補式 VP，可以和前面的 PP 直接組合；而後項「無拂著/vp」這個 VP 帶有否定標記「無」，我們已確認它不能與「分」字組合，所以必須依據「<PP VP> VP」來切分，它的結構層次是：

<分佢/pp 奮走/vp> 無拂著/vp，根據標記規範標出的結果是：

(<分佢/pp~po.1pa.1pn 奮走/vv>/vp~zd.1pp.1vp 無拂著
/vp~zd.1rb.2vv)/vp~ld.1vp.1vp

至此我們基本上已經回答了 PP VP VP 的組合結構層次問題及各種限制上的制約條件，但若要完全無誤地判斷內部結構關係，似乎需要更嚴格的條件，所以我們必須了解 VP VP 組合時形成的不同結構關係的各種制約條件。

第五節 VP VP 不同結構關係的限制條件

vp和vp組合時，根據我們的描述體系一共可以構成「述賓、述補、聯合、連動、主謂」五種結構關係，為了方便敘述，我們把兩個VP分別標為VP₁和VP₂，因為前四種構成的關係仍然是VP，所以我們只討論前四種情形，暫時把「主謂」關係這一種略去，下面先給出幾個例子：

- (1) 述賓式：同意/vp 去台北/vp
- (2) 述補式：擊/vp 落來/vp
- (3) 連動式：橫下來/vp 睡目了/vp
- (4) 聯合式：讀書/vp 寫字/vp

(一) VP₁和VP₂構成述賓關係

上文已經說明過，根據我們目前的考察，VP₁ 和 VP₂ 構成的述賓關係不能直接帶上「分」字組合成被動關係。

(二) VP₁和VP₂構成述補關係

這種述補關係我們採取簡化的做法，規定兩個動詞性成分直接結合，不依靠「得」字或「到」字結合的情形，要構成述補關係的VP₁一般而言是能帶補語的動詞，而VP₂是個封閉集合，如「死、走、忒、橫」等等，以及趨向動詞「上、下、來、落、出來、過去、轉」等。

我們所考察的 1000 個雙音節動詞中計有 372 個述補結構的動詞，當中能與「分」字組合成被動關係有 354 個，如「拉開、拆開、的開、挖開、閃走、剪開、徒開、剝開、奮開、蹬開、搥開、搥開、拈開、收淨、洗淨、舐淨、掃淨、清淨、祛淨、瀝淨、罇淨、掙淨、毒死、害死、崁死、浸死、笑死、啄死、寒死、痛死、翕死、註死、搥死、撚死、罵死、踏死、餓死、燒死、激死、磧死、闕死、嚇死、打爛、扯爛、拖爛、的爛、著爛、捋爛、啄爛、涿爛、蛀爛、揉爛、敲爛、彈爛、踏爛、擊爛、捶爛、吵醒、喊醒、嚇醒、驚醒、喝醒」等，少部分仍然是不能與「分」字組合成被動關係的，如「哽死、詐死、跌死、睡死、熱死、走醒」等。這似乎是和動詞的語義有關，目前我們沒有做這方面的考察，有待進一步研究。

其次如果我們能夠把VP₂這個封閉集合的元素認識清楚，則對VP₁和VP₂構成述補關係就能清楚的描述出來。

(三) VP₁和VP₂構成聯合關係

VP₁和VP₂要能構成聯合關係，其實很複雜，和VP₁和VP₂本身的語義有很大的關聯，有些還必須使用連詞連接，我們看來基本上VP₁和VP₂必須同屬同一個動詞次類且詞義會相關，所以現階段爲了簡化處理，我們只考慮這種聯合關係，同時排除使用連詞連接的狀況，如：

「弄鬼/vp 弄怪/vp」、「做賊/vp 偷人/vp」、「擊刀/vp 擊棍/vp」、「食好/vp 食壞/vp」

這種聯合關係可以和 PP 結合，構成「PP VP VP」，但似乎也不是每個都可以，而且會有歧義現象，如：

「食好/vp 食壞/vp」可以構成「<分佢/pp 食好/vp>食壞/vp」，但卻有歧義，因為也可以是「分佢/vp<食好/vp 食壞/vp>」，我們認為這和動詞的語義有關，需要更細緻的考察。

(一) VP₁和VP₂構成連動關係

上文我們已經分析出連動關係裡的VP₁和VP₂必須VP₂能和PP結合者，整個VP₁VP₂連動關係才有可能和PP結合，這是一個很具體的規則，對客家語來說，能由VP₁和VP₂構成連動關係的，有幾種限制：

- (1) 只有VP₁是「來(le1)、來去(loi1)、去(hi3)」時，可以是單音節動詞，其他都必須是雙音節的述賓或述補式動詞，而且「來(le1)、來去(loi1)、去(hi3)」只能佔據VP₁的位置，如：

「去/vp1 泅水/vp2」、「來去/vp 洗身/vp」、「來/vp 走了/vp」

這種情形的VP₁和VP₂連動關係，不能和PP結合，所以我們便可得到要構成「PP VP VP」裡的連動關係VP₁不能是單音節動詞的限制。

- (2) 正反兩個動作的陳述，有互為補充的作用，如：

「坐等/vp1 無起來/vp2」、「倚等/vp1 毋出去/vp2」、「園等/vp1 毋出聲/vp2」

這種情形的VP₁和VP₂連動關係，不能和PP結合，因為VP₂帶有否定標記。

(3) 先後有連續關係的兩個或幾個動作行爲，如：

「食飽/vp1 出去/vp2 行行/vp3」、「打門/vp1 出去/vp2」、「走等/vp1 講/vp2」
 「拿等書/vp1 上課/vp」、「打電話/vp1 喊車/vp2」、「有權力/vp 簽名/vp」

這種情形的VP₁和VP₂連動關係，不能和PP結合，除非VP₂滿足能和PP結合的規則，例如：

| | | |
|-----------------|--------|-----------------------|
| 打開門/vp1 出去了/vp2 | -----> | 分佢/pp 打開門/vp1 出去了/vp2 |
| 有權力/vp 簽名了/vp | -----> | 分佢/pp 有權力/vp 簽名了/vp |

國立新竹教育大學

從上面的分析來看，「PP VP VP」要能成立，連動式的VP₁和VP₂中的VP₂一定要具有能和PP結合的能力，否則「PP VP VP」不能成立。這種「PP VP VP」，必須按a式切分，即「PP <VP VP>」。

(二) VP₁ 是狀中式 VP

因為我們只考察PP VP VP 的情形，而且只限PP是「分」字，所以我們只簡單討論這一種狀況。如果VP₁是狀中式時，整個「PP VP VP」必須按照b式切分，VP₁不能再帶賓語和補語，則VP₁和VP₂之間也就不會構成述賓和和述補結構，所以只有聯合和連動兩種關係了。

當 VP1 是狀中式 VP 時，如果要讓這種情形的 VP1 和 VP2 的聯合式關係成立，則 VP2 也必須由「分」字充當狀語，也就是說 VP2 也必須是狀中式 VP。如：

「分阿爸打/vp 分阿姆罵/vp」、「分阿婆拿走/vp 分阿舅搶去/vp」

如果要讓這種情形的 VP1 和 VP2 的連動式成立，那麼 VP2 就必須是不能與「分」字結合的動詞，否則就會變成以 a 式切分了，下面我們舉實例說明，並作為本章考察之小結。

「分佢擎棍仔mak20 著」這個短語我們先分析出是「PP VP₁ VP₂」的結構，即「分佢/PP 擎棍仔/VP₁ mak20 著/VP₂」，接著分析VP₁是述賓構造，不能直接和PP組合，而VP₂是可以和PP組合的動詞，可以構成「分佢/PP mak20 著/VP」，那麼可知本例必須按a式切分，是「PP <VP₁ VP₂>」的結構。

而「分佢喊去改蕃薯」這個短語我們一樣先分析出是「PP VP₁ VP₂」的結構，即「分佢/pp喊去/VP₁改蕃薯/VP₂」，接著分析VP₁是述補構造，可以和PP組合成狀中式VP，而VP₂是述賓構造，是不可以直接和PP組合的VP，那麼可知本例必須按b式切分，是「<PP <VP₁> VP₂」的結構，「<PP <VP₁>」和「VP₂」構成連動結構。

第十一章 結語

第一節 回顧

語料庫是語言採樣的樣本集合，著名的語言學家兼語料庫語言學家 Leech 深知收集和建設語料庫是一項非常艱鉅的工程，他指出（Leech 1998：xvii）：

Only those who have first-hand knowledge of this work can fully appreciate that the compilation of a corpus (with proper to quality, design and so on) always takes twice as long as one thought, and sometimes ten times as much effort.

（只有對收集與建置計算機語料庫有第一手實踐經驗的人，才能充分理解建庫過程的艱苦。建立一個對質量、設計標準等給予恰當注意的語料庫，比起預先估計的程度，總是要多花一倍的時間，有時甚至多花十倍的努力。）

筆者選擇走這條路，不敢說花費十倍的努力，但的確花費許多倍的時間在整理語料和標記工作上，對這樣的一個工作筆者感觸很多也獲益良深。

（一）客家語言「文字、拼音系統」標準化迫切的需求，這已是客家研究界多年的要求，在筆者建置這個語料庫時，這個問題的確非常困擾，只能期待這一天早日到來。

（二）語料庫的建置工作，需要投注大量的人力和物力之外，筆者更期待以客家語為母語的人，能盡量產出「語料」，因為現有的語料量和涵蓋面都太小，口語對話式語料更是少，要做好更細緻的研究工作，沒有適當的語料做為依據，只能永遠在研究者自行造句的模式下進行，是阻礙研究進步的絆腳石，希望大家合力來搬走這些絆腳石。

（三）在建置語料庫的過程中，大量閱讀與分析文本，讓我找回許多小時候的記憶，也讓我常想起父親的話語，這是獲益最良深之處。

(四) 就漢語的書寫而言是漢字連寫，當中沒有分詞符號，如果客語的文本書寫者可以採取「分詞書寫」，如：

你 愛 去 柰 ？ 今 晡 日 愛 去 台 北 。

這對語料庫的建置會有相當大的幫助，但由於長期以來的習慣，要改變恐不易。

第二節 研究成果

(一) 在經過語料分析之後，我們一共得到 1299 個單音節動詞，以往客家研究者多只推估式的陳述「客家語的單音節詞較豐富」這個命題做為結論，但究竟多多少少？一直沒有解答，以我們目前的語料量和描述體系來看，的確比華語的單音動詞多。

(二) 詞類統計結果，名詞佔最大一部份，有 64.43%強，其次是動詞佔 20.43%，兩者合計就佔去了 84.86%強，再次是形容詞佔 7.91%，其他各類詞僅佔非常小的比例。

(三) 客家語的使用者，主要使用的是「動詞性短語」，佔 52.74%，其中又以述賓 (vo)、述補 (vc)、狀中 (zd) 和附加 (fg) 為主，連動和聯合短語相對偏低，這和動詞的分布是相呼應的，雙音節動詞中也是以述賓和述補式構詞為主的。這樣的分布意義顯示客家語的動詞是以帶賓語、補語和附加時態標記及受狀語修飾等為主要的語法功能，而連動和聯合只是動詞或動詞短語的一種變化運用方式，並不是主要的構句方式。

(一) 就數量而言，客家語動詞是雙音節為主，但單音節詞動詞的使用總頻率卻遠高於雙音節動詞。雙音動詞中又以述賓式和述補式最為普遍。

(二) 就文本分析角度的結果，客家語有 56 個核心動詞，其中與現代華語不同的計 23 個：「講 1、食 1、分 1、無、喊、愛 1、係、知、行、噉、著 1、轉 1、走 1、驚、倚、落、尋、loi53、做 3、愁、刷、罔、僚」，佔全部核心動詞的 41.1

。這 23 個與現代華語不同的核心動詞總使用詞次為 7850 次，佔 56 個核心動詞總使用詞次的 44.58%，高頻的現象，顯示這是核心中的核心，相對的穩定性也非常高。

(三) 台灣客家語中四縣和海陸兩種腔調均相同的單音節動詞計 1269 個，佔全部單音節動詞的 97.69%；僅用於四縣客家語的單音節動詞計 19 個，佔全部單音節動詞的 1.46%；僅用於海陸客家語的單音節動詞計 5 個，佔全部單音節動詞的 0.39%。

(四) 在 1283 個客家語單音節動詞中，有 630 個與華語相同，佔全部單音節動詞的 49.10%，與華語相異的單音節動詞計 653 個，佔全部單音節動詞的 50.90%。

(五) 客家語的動詞（甚至是整個詞彙系統）是多來源，多層次性的，要釐清這些來源和層次，需要有適當方法和標準，建立這些方法和標準的目的與功能就是要把這些要釐清的元素找出來，我們建置這個語料庫的目的就是透過較大量文本的分析，把「元素找出來」的一個方法和過程。

(六) 本研究抽取了語料庫中的動詞，並結合田野調查等方法，我們描述了客家語 VP 的基本結構和分布面貌，並以 PP VP VP 的組合為例，考察被動式的「分」字和 VP 組合的各種條件限制，根據這些條件限制，我們就可以判定：

(1) 若後一個 VP 可以和「分」字組合，整個結構是向右切分，構成

a、(PP <VP VP>)模式

(2) 若後一個 VP 不能和「分」字組合，整個結構是向左切分，構成

b、(<PP VP> VP)模式

(3) 若按 a、(PP <VP VP>)模式切分，兩項 VP 之間可以是述補、聯合、

連動三種結構關係，而述補和聯合關係都有嚴格的限制條件。

- (4) 若按 b、(<PP VP> VP)模式切分，整個結構可能是聯合或連動關係，但都有許多限制性條件。

第三節 研究展望

雖然我們建置了這個小型語料庫，目的是要離析出客家語裡的各個元素，做為進一步研究和比較的基礎，但是有許多工作卻是有待繼續完成的：

(一) 語料量不足與涵蓋面偏差

語料量不足與涵蓋面偏差的問題，已如上述，雖然這是一個永遠沒有終點的問題，但還是期待以客家語為母語者，能產出更多的客家語文字化文本，更期待電子檔能無條件提供給本語料庫使用，也期待本語料庫在未來的日子裡能網路化，提供給更多有志客家研究者使用。

(二) 本研究雖然集中討論了客家語 VP 的問題，但在實際分析語料時，則是全面牽涉到其他的短語類別，語言觀會影響分析的結果，而要構築一個全面而系統性的分析系統，必需對每個及每類詞和每類短語的分布和句法限制條件做調查與分析總結，這個工作無疑是長期而艱鉅的，需要投注大量的人力和時間才能完成，本文對 VP 的研究只是其中很小的一部份，而且只做了句法層面上的分析，在語義及語用等方面因素幾乎都沒有考慮，縱橫方面都有許多待將來進一步研究的空間和需求。

參考書目

一、專書

- 羅肇錦 1988 《客家語語法》。台北：學生書局。
- 羅肇錦 2000 《台灣客家族群史·語言篇》。南投：台灣省文獻委員會。
- 羅肇錦 1990 《講客家語》。台北市：自立晚報社文化出版部。
- 羅肇錦 1996 《臺灣的客家話》。台北市：台原出版社。
- 羅肇錦 1998 《客家話的字詞與音義析論》。台北：紅葉文化事業公司。
- 鍾榮富 2001 《福爾摩沙的烙印：臺灣客家話導論》【語言篇】上、下冊。
臺北：行政院文化建設委員會。
- 橋本萬太郎 1972 《客家語基礎語彙集》。東京：東京外國語大學。
- 藍小玲 1999 《閩西客家方言》。廈門：廈門大學出版社。
- 何耿庸 1993 《客家方言語法研究》。廈門：廈門大學出版社。
- 項夢冰 1997 《連城客家話語法研究》。北京：語文出版社。
- 謝永昌 1994 《梅縣客家方言志》。廣州：暨南大學出版社。
- 林立芳 1994 《梅縣方言語法論稿》。廣州：暨南大學出版社。
- D. MacIver 編著 1992 《客英大辭典》。南天書局。
- 詹益雲 1998 《海陸腔客家字典》。台北：南天書局。
- 黃雪貞 1998 《梅縣方言詞典》。北京：江蘇教育出版社。
- 龔萬灶 2003 《客家語實用手冊》。作者自行發行。
- 龔萬灶 2004 《阿啾箭个故鄉》。作者自行發行。
- 張維耿 1995 《客家話辭典》。北京：廣東人民出版社。
- 徐運德 1989 《客家話詞彙》。台北：中原週刊社。
- 徐兆泉 2001 《台灣客家話辭典》。台北：南天書局。
- 饒秉才 2000 《客家音字典》。北京：廣東人民出版社。
- 何石松等 2002 《現代客家語詞彙彙編》。台北：台北市政府客家事務委員會。

- 何石松等 2004 《現代客家語詞彙續編》。台北：台北市政府客家事務委員會。
- 楊政男等 1998 《客家語字音詞典》。台北：商務印書館。
- 俞士汶 2004 《計算語言學概論》。北京：商務印書館。
- 馮志偉 2001 《計算語言學探索》。北京：黑龍江教育出版社。
- 馮志偉 1995 《自然語言機器翻譯新論》。北京：語文出版社。
- 林杏光 1999 《詞彙語義和計算語言學》。北京：語文出版社。
- 靳光瑾 2003 《計算語言學視窗》。北京：北京廣播學院出版社。
- 翁富良 1996 《計算語言學導論》。北京：中國社會科學出版社。
- 侯 敏 1999 《計算語言學與漢語自動分析》。北京：北京廣播學院出版社。
- 劉開瑛 2000 《中文文本自動分詞與標注》。北京：商務印書館。
- Jenny Thomas 2001 《Using Corpora for Language Research》。北京：外語教學與研究出版社。
- 胡明揚主編 1997 《詞類問題考察》。北京：北京語言文化大學出版社。
- 胡明揚主編 2004 《詞類問題考察續集》。北京：北京語言文化大學出版社。
- 邢福義 2003 《詞類辨難》。北京：商務印書館。
- 陳光磊 1994 《漢語詞法論》。上海：學林出版社。
- 陸志偉 1957 《漢語的構詞法》。北京：科學出版社。
- 竺家寧 1999 《漢語詞彙學》。台北：五南圖書出版社。
- 邢福義 1996 《漢語語法學》。吉林：東北師範大學出版社。
- 韓秀英編 1997 《後綴形容詞例解詞典》。南京：江蘇教育出版社。
- 呂叔湘 2002 《現代漢語八百詞》。北京：商務印書館。
- 朱德熙 1982 《語法講義》。北京：商務印書館。
- 朱德熙 1985 《語法答問》。北京：商務印書館。
- 朱德熙 1980 《現代漢語語法研究》。北京：商務印書館。
- 胡裕樹、范曉合著 1995 《動詞研究》。開封市：河南大學出版社。
- 張 斌 1989 《漢語語法研究》。北京：商務印書館。

- 張斌主編 2002 《現代漢語語實詞》。北京：華東師範大學出版社。
- 張斌主編 2002 《現代漢語語虛詞》。北京：華東師範大學出版社。
- 張斌主編 2002 《現代漢語語短語》。北京：華東師範大學出版社。
- 張斌主編 2002 《現代漢語語句子》。北京：華東師範大學出版社。
- 張斌主編 2002 《現代漢語語語法分析》。北京：華東師範大學出版社。
- 范曉、張豫峰等 2002 《語法理論綱要》。北京：上海譯文出版社。
- 馬慶珠 2002 《漢語動詞和動詞性結構》。北京：北京大學出版社。
- 徐 杰 2004 《普遍語法原則與漢語語法現象》。北京：北京大學出版社。
- 王理嘉等 2004 《現代漢語專題教程》。北京：北京大學出版社。
- 李 珊 2003 《動詞重疊式研究》。北京：語文出版社。
- 陸儉明 2003 《現代漢語語法研究教程》。北京：北京大學出版社。
- 張誼生 2001 《現代漢語副詞研究》。北京：學林出版社。
- 鄒韶華 2001 《求真集》。北京：新華書店。
- 李 泉 2001 《漢語語法考察與分析》。北京：北京語言文化大學出版社。
- 邵敬敏 2001 《漢語語法的立體研究》。北京：商務印書館。
- 周 荐 1994 《詞語的意義和結構》。中國大陸：天津古籍出版社。
- 季羨林主編 2001 《二十世紀現代漢語語法八大家選集---黎錦熙選集》。北京：東北師範大學出版社。
- 季羨林主編 2001 《二十世紀現代漢語語法八大家選集---黎錦熙選集》。北京：東北師範大學出版社。
- 季羨林主編 2001 《二十世紀現代漢語語法八大家選集---王力選集》。北京：東北師範大學出版社。
- 季羨林主編 2001 《二十世紀現代漢語語法八大家選集---呂叔湘選集》。北京：東北師範大學出版社。
- 季羨林主編 2001 《二十世紀現代漢語語法八大家選集---胡裕樹、張斌選集》。北京：東北師範大學出版社。

季羨林主編 2001 《二十世紀現代漢語語法八大家選集---朱德熙選集》。北京：東北師範大學出版社。

季羨林主編 2001 《二十世紀現代漢語語法八大家選集---陸儉明選集》。北京：東北師範大學出版社。

劉月華等 1996 《實用現代漢語語法》繁體版。台北：師大書苑。

繆錦安編著 1998 《漢語的語義結構和補語形式》。上海：上海外語教育出版社。

屈承熹著、紀宗仁協著 1999 《漢語認知功能語法》。台北：文鶴出版社。

袁毓林 1998 《語言的認知研究和計算分析》。北京：北京大學。

三、單篇論文：

羅肇錦 1996a 〈四縣客家語附著成分結構〉，《『臺灣客家語概論』講授資料彙編》。臺北：臺灣語文學會出版 頁 117-153。

羅肇錦 1996b 〈四縣客家語虛詞的功能結構〉，《『臺灣客家語概論』講授資料彙編》。臺北：臺灣語文學會出版 頁 154-187。

羅肇錦 1998 〈客家語字線索與非本字思索〉，《國文學誌》。第 2 期 頁 383-413。

盧彥杰 1999 〈新竹海陸客家詞彙研究〉，國立新竹師範學院台灣語言與語文教育研究所碩士論文。

徐子晴 2000 〈客家諺語的取材和修辭研究〉，國立新竹師範學院台灣語言與語文教育研究所碩士論文。

彭美賢 2002 〈客家語動物名稱詞彙初探-以蟲仔、蛇、蛙為例〉，國立新竹師範學院台灣語言與語文教育研究所碩士論文。

黃佳文 2004 〈台灣東勢客家語表性狀詞的語義分析〉，國立新竹師範學院台灣語言與語文教育研究所碩士論文。

邱燕美 2004 〈客家語肩肘手指動作詞的語音詞義探討〉，國立新竹師範學院台灣語言與語文教育研究所碩士論文。

強舒嫩 2002 〈客家語打類、忍類及促類動詞中動力學之體現〉，國立政治大學

語言學研究所碩士論文。

張雁雯 1997 〈臺灣四縣客家話構詞研究〉，國立台灣大學中國文學系碩士論文。

吳莉雯 2000 〈台灣四縣客家語量詞系統研究〉，國立中正大學語言學研究所碩士論文。

陳韻予 2002 〈台灣閩南語及客家語量詞系統比較研究〉，國立中正大學語言學研究所碩士論文

江俊龍 1996 〈台中東勢客家方言詞彙研究〉，國立中正大學中國文學研究所碩士論文。

江俊龍 1988 〈論東勢客家話的單音詞〉。載於：《台灣源流八十七年秋季刊》，105-118。

江俊龍 1998 〈台中東勢客家方言中的外來語〉，載於：《第二屆台灣語言國際研討會論文選集》，585-598。

徐光榮 1987 〈臺灣客家話同義詞比較研究〉，輔仁大學中國文學系碩士論文。

張玲瑛 1992 〈客家語的構詞與語法研究〉，輔仁大學語言學研究所碩士論文。

李靜延 2003 〈從「形容詞狀語」探討「漢語狀語」〉，國立台灣師範大學華文所碩士論文語文教學研究所。

徐貴榮撰 2002 〈臺灣客家語表「手動作行為」的詞義分析〉，《客家方言研究》第四屆客方言研討會論文集 廣州：暨南大學出版社。

洪成玉 1994 〈語法分析和語義分析〉，載於：首都師範大學學報社會科學版 1994年 第3期，1-7。

方琴 1994 〈狀態形容詞的語法特徵及相關問題〉，載於：徐州師範學院學報哲學社會科學版 1994年第2期，119-123。

王素梅 1998 〈雙音節狀態形容詞的 ABAB 式重疊〉。載於：漢語學習 1998年 第2期，53-54。

朱德熙 1956 〈現代漢語形容詞研究〉。載於：語言研究 1956年第一期。

金昌吉 1994 〈方位詞的語法功能及其語義分析〉。載於：內蒙古民族師院學報

哲社版 1994 年第 3 期，22-26。

鍾守滿、王偉 2000 〈言語行為動詞語義分析及架構〉。載於：外語學刊 2000 年第 2 期，62- 64。

高永希、鄭獻芹 1998 〈對擬聲詞歸類問題的思考〉。載於：殷都學刊 1998 年第 2 期，75-78。

王桂安 1995 〈擬聲詞的探討〉。載於：廣東民族學院學報〈社會科學版〉。1995 年第 2 期，p84-87。

金昌吉 1994 〈方位詞的語法功能及其語義分析〉，《內蒙古民族師院學報》。第 3 期，22-26。

陳宇萍 .1987. 〈現代漢語名詞類的擴大----現代漢語動詞和名詞分界線的考察〉，《中國語文》第五期。

郭 銳 1999 〈語文詞點的詞性標注問題〉，《中國語文》第二期。

郭 銳 2001 〈漢語形容詞的劃界〉，《中國語言學報》第十期。北京：商務印書館。

李宇明 1996 〈非謂形容詞的詞類問題〉，《中國語文》第一期。

劉丹青 1991 〈從漢語特有詞類問題看語法的宏觀研究〉，《江蘇社會科學》第二期。

張伯江 1994 〈漢語詞類歸屬的理據〉，《國外語言學》第四期。頁 29-36。

陸儉明 1991 〈現代漢語時間詞說略〉，《語言教學與研究》第一期。

陸儉明 1993 〈關於漢語詞類的劃分〉，《人文科學》第 69、70 合輯期。

陸儉明 1994 〈關於詞的兼類問題〉，《中國語文》第一期。

張伯江 1994 〈漢語詞類歸屬的理據〉，《國外語言學》第四期。頁 29-36。

馬慶珠 1999 〈詞組的研究〉，《語法研究入門》。北京：商務印書館

周明強 1999 〈「的」字短語的界定及其主要特點〉，《內蒙古師大學報》第 28 卷第 3 期