

語料庫為本的現代漢語結構助詞統計分析*

阮家慶

國立中正大學語言學研究所

摘要

本研究以語料庫語言學的角度，嘗試透過大量語料庫語料，以 CART 決策樹分類 (Breiman et al. 1984) 配合 VARBRUL 變異數規則分析 (Rand and Sankoff 1990; Robinson, Lawrence, and Tagliamonte 2001; Sankoff, Tagliamonte, and Smith 2005) 等方法，證實了漢語結構助詞「的」、「得」、「地」的使用規則與實際上的語言使用是大致符合的。在使用結構助詞的認知上，因前後環境的交互作用，產生了選用結構，最後以「的」、「得」、「地」作為標記 (徐陽春 2004)。量化資料清楚說明「的」/「地」，「的」/「得」較不常混用，較常混用的是「得」/「地」。對於此三個漢語結構助詞的分用與合用的建議，應以整體社會語言使用者使用情形來決定。本研究建議，分用比較符合現階段整體的使用方式。本研究亦提出認知概念模型來描述三種不同的漢語結構助詞與前後環境的關聯性。本研究結果預期對於漢語辭典、華語教學、語言政策、自然語言處理做出參考與貢獻。

關鍵詞：漢語 結構助詞 語料庫 變異數規則分析 決策樹分類

1. 前言

在現代漢語裡面，結構助詞「的」、「得」、「地」的使用是最為常見的，例如：紅色的汽車、吃得很飽、慢慢地接近。此三者的使用方法在漢語研究、漢語母語教學或對外華語教學都有廣泛的研究與討論，像是「的」應該使用在定語（形容詞）之後、中心語之前；「得」應該使用在謂語（動詞）之後、補語之前；「地」應該使用在狀語（副詞）之後、動詞之前（方麗娜 1994；徐陽春 2004；劉雪梅 2006；江

* 本文承蒙本期刊兩位匿名審查人給予之寶貴建議，使本文內容更加完善，特此感謝！本文內容及所有研究，本文作者皆負有其責。

勝利 2008；馮元玫 2009）。就句法結構來看，「的」、「地」是為偏正結構，「得」則為動補結構（方麗娜 1994；徐陽春 2004；江勝利 2008；馮元玫 2009）。就漢語母語教學或對外華語教學來說，此三種規範性語法提供了一個很清楚的辨認方法。然而，語言使用者或語言學習者往往混（誤）用此三種結構助詞。在過去三、四十年來，有許多文獻都在討論此三者的混用情形，甚至也探討到有關此三者的應該合用或者是分用的問題（嚴戎庚 1999；徐陽春 2003；田野 2004；徐陽春 2004）。本研究嘗試從語料庫語言學的角度，使用語料庫大量的語料，對於漢語結構助詞進行分析研究，進而探討多個前人文獻尚未注意到的問題，包括了理論、規範的驗證及結構助詞使用狀況及混用程度，也另外以構式語法 (Construction Grammar) (Goldberg 1995) 的觀點，綜合統計分析結果，提出了相關的認知概念模型。

近二十年來，還是有些文獻在探討有關「的」、「得」、「地」這三個結構助詞的使用、辨析與分合問題¹。方麗娜 (1994) 以文法分析的觀點，利用了圖解法、歸納法、演繹法、比較法等方式，來辨析「的」、「得」、「地」結構助詞的使用差異，以提供國語文中小學教師在教學上的參考。「的」、「得」、「地」在文法結構裡面分別有不同的用途：「的」用在標示定語，「得」用在標示補語，「地」用在標示狀語。其中，「的」及「地」是屬於偏正片語結構，然而「得」是屬於謂補片語結構。因為「的」及「地」都是屬於同樣結構的關係，所以致使一些語言使用者容易誤用。然而，「的」、「得」、「地」三者要視其表達重點不同而使用，在一個句子裡面沒有絕對的正確與否。方麗娜 (1994) 提出了例子：「死的很難看」，「死得很難看」，「死地很難看」。在「死 de 很難看」這個句子中，「的」、「得」、「地」都可以使用，並非只能使用某一個才稱為正確，而是要視語境表達不同語義而使用不同的結構助詞。

嚴戎庚 (1999) 認為「的」、「得」、「地」在分用時因為混亂不清，所以有語言使用者將三者合併，甚至以「de」的形式將其寫出。但是，「de」在書面語的三種不同形式具有區別意義的功能，是「語言精密化的表現」（嚴戎庚 1999：94）。嚴戎庚認為如果只是說「『的』」用在定、中之間，「『地』」用在狀、中之間，「『得』」用在中、補之間；或者進而只是就整個偏正片語的功能作出說明指出體詞性偏正片語中用「『的』」，謂詞性偏正片語中用「『的』」²，都是不能解決問題的」（嚴戎庚 1999：94）。

¹ 因為過去文獻對於「的」、「得」、「地」提供的語法介紹分析已經非常詳盡，如馮元玫 (2009)，故本文不細究、重述此三者結構助詞的語法分析，而以語料庫的分析來對應過去文獻的理論結果，觀察整個語言使用者群體是如何使用「的」、「得」、「地」，以及分佈情形為何。

² 原著此處有誤，應為「『地』」。

嚴戒庚進一步認為，區分「的」、「得」、「地」得從「語法、語義、語用這三個平面出發才能講得清楚、實用」(嚴戒庚 1999: 96)。

徐陽春 (2003) 探討「的」、「得」的分合問題。徐陽春認為結構助詞「的」、「得」不能合併有兩大原因：一旦合併，會產生無法處理的歧義問題。其次，「的」、「得」在句法結構上和語義上都是不一樣的，和「的」、「地」之間是不相同的情形。但在「『的』替代句末隱去補語的『得』」(徐陽春 2003: 154) 的情形，則是可以合併使用。

田野 (2004) 討論了「的」、「得」、「地」的合用與分用問題。田野認為，雖然「的」、「得」、「地」在歷史上是合用的，但是合用在表達的精確度方面會有所欠缺。因此在 1956 年左右開始，此三者的分用開始有了提倡，但是在 1984 年後又有合用的情形(像是使用「的」代替「地」)，因為雖然「的」、「得」、「地」在使用上有明顯的規範，但是使用的情形卻是混在一起。而這混用的最大原因，則是從「的」代替「地」的使用開始的。

徐陽春 (2004) 探討了「的」、「地」的分合問題。兩者雖然是不同性質的偏正結構(「的」是體詞性，「地」是謂詞性)，在語言使用上面也常常混用，但是「的」、「地」是可以合併使用的。徐陽春進一步指出，「的」、「地」一般是使用為定語、狀語的標記，但是識別體詞性結構或謂詞性結構反而成為使用「的」、「地」的關鍵。「當整個偏正結構是體詞性結構(語義上具有指稱性)時，該結構的修飾語是定語；當整個偏正結構是謂詞性結構(語義上具有陳述性)時，該結構的修飾語是狀語」(徐陽春 2004: 152)。一旦語言使用者得知結構為何時，「的」、「地」的標記才會出現，而不是根據「的」、「地」來決定是哪一種偏正結構。

以「的」用在定語後，「得」用在補語前，「地」用在狀語後的標準基礎，劉雪梅 (2006) 以語料庫量化、質化研究法，發現中小學學生「的」、「得」、「地」有交叉誤用的偏誤情形，另外發現結構助詞的偏誤率佔所有類型偏誤率的大多數(28.2%)，在助詞誤用裡面，結構助詞的偏誤率則是佔 86.5%。

江勝利 (2008) 認為「的」、「得」、「地」具有不同的性質與特色。在「的」方面，「『的』為定語的標誌，表示它前邊的詞或片語做定語」，「從語義角度講，『的』具有區別定語的修飾性和領屬性」，「『的』還有幫助本不可以充當定語的詞語做定語的作用」，以及「『的』可以用在動詞、形容詞，特別是動賓片語後構成『的』字結構，用以概括地表示人或事物的類別。」(江勝利 2008: 101)。在「得」方面，「『得』為補語的標誌，表示『得』後邊成分是補語。」(江勝利 2008: 102)。在「地」方面，「『地』為狀語的標誌，表示它前面的詞或片語是狀語。由『地』構成的片語也

是偏正詞性，但屬於動詞性偏正片語」(江勝利 2008:101)。因此，「的」、「得」、「地」三者的不同，是在於語法標記上的不同。然而，「用『的』與否還具有區別片語性質的作用，「得」、「地」就沒有這種區別作用。」(江勝利 2008:102)。在這三個詞的使用方面，歷史以來「這三個詞的使用常常是隨意的、混亂的，而人們經常地用『的』替代『地』、『得』使用是有其道理的」(江勝利 2008:103)。

馮元玫 (2009) 以漢語語法及第二語言習得理論研究的觀點角度，對「的」、「得」、「地」提出綜合性的整理、辨析，提供對外華語教學的參考。馮元玫 (2009) 綜合呂叔湘 (1980)、胡裕樹 (1992) 及劉月華 (2001) 的分析得出「的」、「得」、「地」是「漢語句式結構中最核心的語用意義和規範」，也是「第二語言學習者最早習得、頻繁應用，但卻也始終是混淆不清偏誤不斷的難題所在」(馮元玫 2009:19)³。對於「的」、「得」、「地」的使用，「『的』作定語標誌，表明句法結構是定中式偏正結構」，「『得』作補語標誌，表明句法結構是中補結構」，「『地』作狀語標誌，表明句法結構是狀中式偏正結構」(馮元玫 2009:3)。第二語言學習者在華語助詞的學習中錯誤最多的情形就是「的」、「得」、「地」的混用。

由前人的文獻裡面可以看的出來，結構助詞「的」、「得」、「地」有各自的使用方式及各自不同的結構。也提及此三者具有混用、誤用的情形發生。過去也許因為電腦語料庫不發達，在這些前人的文獻裡面，大多數是透過舉例法來解析「的」、「得」、「地」的使用情形及其句法、語義上的表現，對於此三者分用與合用的討論也是如此。教學用的規範性語法似乎也沒有其他文獻透過大量語料來驗證這樣分類的適當性。前人文獻所採用的分析研究法為「由上而下」(top-down) 的觀點，這樣的分析方法有其缺點，像是容易落入自行造例、易見樹而不易見林等問題。本研究以語料庫語言學的角度，使用大量語料 (bottom-up) 自行說話的方式，來驗證、比對前人文獻對於漢語結構助詞的理論與規範。因此，前人文獻中似乎沒有關注到以下的問題，而本研究的目的就是希望經由統計分析的方式來進行回答：

- (1) 漢語結構助詞「的」、「得」、「地」的理論是否與實際上的使用相符合？
- (2) 使用分佈？混用的程度？多少精確數字百分比？
- (3) 以整體而言，語言使用者是如何使用漢語結構助詞？

提出這三個問題的原因在於，語言會隨著時間而改變，語言使用者在使用結構助詞時，是否還能對應到「的」、「得」、「地」相關的理論與規範？像是「的」、「地」在

³ 對於「的」、「得」、「地」的歷史與語法化過程，可詳參馮元玫 (2009) 的整理與分析。

台灣語言政策下合用後（方麗娜 1994），是否已經悄悄改變了「的」、「地」使用方法？而這個政策是否能對應到真實語言使用的情形？多數文獻都有提及此三者結構助詞混用的情形，有些論點甚至相悖。像是方麗娜（1994）提及「的」、「地」因其結構相似較易混用，嚴戎庚（1999）提及「得」、「地」較易混用。現有文獻似乎也無法明確說明其混用程度為何。如果無法得到混用程度的精確計算，分用合用的語言政策制定也將無法取得較為客觀、大量的依據。另外，語言使用者是如何使用漢語結構助詞？是依據助詞的前面的環境，還是根據後面的環境，或是兩者皆參考？

本研究以漢語結構助詞「的」、「得」、「地」的混用假設其為同質變異⁴，嘗試以 CART 決策樹分類方法（Breiman et al. 1984）配合 VARBRUL 變異數規則分析方法（Rand and Sankoff 1990; Robinson, Lawrence, and Tagliamonte 2001; Sankoff, Tagliamonte, and Smith 2005），來呈現漢語結構助詞「的」、「得」、「地」的分佈情形、驗證文獻上漢語結構助詞「的」、「得」、「地」規範性語法的適當性、提供合用或分用的意見參考。本研究預期對於漢語辭典、華語教學、語言政策（language policy）、自然語言處理（像是 word choice）做出參考與貢獻。

本文結構編排如下：第二節提及研究方法，包括語料來源、語料處置、統計分析架構設計與分析工具，第三節呈現統計分析的結果，第四節將分析的結果對照前人文獻分析結果，第五節則是結論。

2. 研究方法

本節提及本研究的語料來源、統計分析架構設計與語料處置、分析工具的說明。

2.1 語料來源

本研究的語料來源是來自於臺灣「中央研究院現代漢語平衡語料庫」（以下簡稱 Sinica Corpus）第三版⁵（詞庫小詞 1995）。這個漢語語料庫是全世界第一個具有斷詞（segmentation）、詞類標記（POS-tagging）的平衡語料庫，收錄各領域：文學、生活、社會、科學、哲學、藝術等的書面語語料約 500 萬詞⁶。本研究從其線上網頁介

⁴ 「的」、「得」、「地」有各自的不同使用方法，但是在目前許多文獻提及此三者有混用情形，本研究是在這樣的基礎之下，假設此三者為自由變異（free variation），投入統計分析。

⁵ 語料來自於線上版本：<http://app.sinica.edu.tw/cgi-bin/kiwi/mkiwi/kiwi.sh>，檢索日期：2014/4/1

⁶ 本研究完成分析時，第四版 1000 萬詞詞庫已上線：<http://app.sinica.edu.tw/cgi-bin/kiwi/mkiwi/kiwi.sh>，沒有繼續更新的原因主要是語料量已經足以投入統計分析。

面取得「的」、「得」、「地」各 5000 筆語料，經過保留詞類為結構助詞部份，分別取得「的」4591、「得」2966、「地」3553，共 11110 筆語料，如表 1。

表 1：結構助詞「的」、「得」、「地」頻率次數比例表

結構助詞	出現頻率	出現頻率百分比
的	4591	41.32%
得	2966	26.70%
地	3553	31.98%
總計	11110	100%

因為語料後續將使用統計工具分析，所以本研究在統計架構設計與語料的處置有做出一些調整，主要是以結構助詞的前後詞類作為分佈分析，如下節所述。

2.2 統計架構設計與語料處置

本研究結構助詞「的」、「得」、「地」的分析主要是參考 Ruan (2014) 的方法。主要是藉由此三者的「前面的詞類標記」與「後面的詞類標記」的分佈（以下簡稱「前面」、「後面」）進行統計分析。比方說，「工作 (Na) 的 (DE) 熱情 (Na)」裡，「的」前面「工作」的詞類是 Na（普通名詞），後面「熱情」詞類也是 Na（普通名詞），「工作」這個 Na 當作定語來使用，「熱情」這個 Na 則是這個「的」的結構裡面的中心語。在統計架構方面，就是把「前面」、「後面」當作兩個自變數 (Independent variable)，「的」、「得」、「地」當作依變數 (dependent variable) 來進行統計分析。另外，在詞類標記的部份，Sinica Corpus 使用的現代漢語詞類標記共有 46 個。投入統計分析時使用更精簡的詞類標記⁷，所以本研究使用其普及化標記共 13 個⁸，如表 2：

表 2：詞類標記表⁹

普及化標記	現代漢語詞類標記	說明
A	A	（非謂）形容詞
ADV	D, Da, Dfa, Dfb, Dk	副詞

⁷ 為避免統計分析過度零散致使分析結果困難解讀，故使用最精簡的詞類標記。

⁸ 因為 NAV（名謂詞）沒有對應的現代漢語詞類標記，所以本研究沒有將其算入，對分析不會造成影響。

⁹ 取自：http://app.sinica.edu.tw/kiwi/mkiwi/modern_c_wordtype.html，檢索日期：2014/4/1

ASP	Di	時態標記
C	Caa, Cbb	連接詞
DET	Nep, Neqa, Nes, Neu	定詞
FW	FW	外文標記
M	Nf	量詞
N	Na, Nb, Nc, Ncd, Nd, Nh	名詞
P	P	介詞
POST	Cab, Cba, Neqb, Ng	後置詞
T	DE, I, T	語助詞
Vi	VA, VB, VH, VI	不及物動詞 ¹⁰
Vt	SHI, VAC, VC, VCL, VD, VE, VF, VG, VHC, VJ, VK, VL, V_2	及物動詞
NAV		名詞

依照設計的方式所得到的描述性資料 (descriptive data) 如表 3：

表 3：結構助詞「的」、「得」、「地」頻率次數細目表（處理前）

自變數 ¹¹	項目 ¹²	依變數			總和
		的	得	地	
前面	1（非謂）形容詞	113	0	27	140
	2 不及物動詞	1107	958	2521	4586
	3 介詞	0	5	0	5
	4 及物動詞	622	1969	281	2872
	5 外文標記	29	1	0	30
	6 名詞	2478	23	205	2706
	7 定詞	10	1	24	35
	8 後置詞	141	2	23	166
	9 時態標記	9	0	0	9
	10 副詞	25	5	426	456
	11 量詞	57	1	46	104
	12 語助詞	0	1	0	1
總和		4591	2966	3553	11110

¹⁰ 包括 VH「狀態不及物動詞」（可當謂語使用的形容詞）。

¹¹ Independent variable.

¹² Independent variable levels.

自變數 2	項目	依變數			總和
		的	得	地	
後面	1 (非謂) 形容詞	59	3	1	63
	2 不及物動詞	354	1227	558	2139
	3 介詞	2	77	386	465
	4 及物動詞	641	253	2217	3111
	5 外文標記	25	0	0	25
	6 名詞	3129	213	112	3454
	7 定詞	254	68	28	350
	8 時態標記	0	14	0	14
	9 副詞	108	1087	233	1428
	10 連接詞	15	21	17	53
	11 量詞	0	1	0	1
	12 語助詞	4	2	1	7
總和		4591	2966	3553	11110

在表 3 裡，「前面」詞類 3、5、9、12 及「後面」詞類 5、8、11 裡面的資料，在投入統計分析前已經刪除（暗底部分），因為這些項目在依變數的表現上沒有變異的情形；也就是沒有分佈的情形發生。經過整理之後，得到「的」4528、「得」2944、「地」3553 筆，表 4 為處理後的資料總表。在表 4 裡面，自變數「前面」有 8 個項目，「後面」有 9 個項目，本表是投入統計分析的基本資料。表 5 及表 6 裡面提供一些詞類標記的範例。

表 4：結構助詞「的」、「得」、「地」頻率次數細目表（處理後）

自變數 1	項目	依變數			總和
		的	得	地	
前面	1 (非謂) 形容詞	113	0	27	140
	2 不及物動詞	1104	955	2521	4580
	3 及物動詞	618	1957	281	2856
	4 名詞	2464	23	205	2692
	5 定詞	10	1	24	35
	6 後置詞	137	2	23	162
	7 副詞	25	5	426	456
	8 量詞	57	1	46	104
總和		4528	2944	3553	11025

自變數 2	項目	依變數			總和
		的	得	地	
後面	1 (非調) 形容詞	59	3	1	63
	2 不及物動詞	351	1227	558	2136
	3 介詞	2	77	386	465
	4 及物動詞	635	253	2217	3105
	5 名詞	3104	211	112	3427
	6 定詞	251	66	28	345
	7 副詞	107	1084	233	1424
	8 連接詞	15	21	17	53
	9 語助詞	4	2	1	7
總和		4528	2944	3553	11025

表 5：結構助詞「的」、「得」、「地」在 Sinica Corpus 的例句-前面¹³

前面	
1 (非調) 形容詞	研究所(Nc)，除了(P)人事費(Na)、水電費(Na)以及(Caa)日常(A)的維護(VC)費用(Na)外(Ng)，剩下來(VJ)能夠(D)用(VC)無得無 注入(VC)生命(Na)的(DE)火花(Na)，讓(VL)研究員(Na)參與(VC)院士(Na)會議(Na)的(DE)活動(Na)該(D)是(SHI)適度(A)地有(V_2)
2 不及物動詞	決定(VE)報考(VC)台大(Nc)歷史系(Nc)，我(Nh)好奇(VH)的問(VE)他(Nh)為什麼(D)學(VC)歷史(Na)？這(Nep)位(Nf)用(VC)在(P)研究(Na)上(Ng)的(DE)經費(Na)真(D)是(SHI)少(VH)得很(Dfa)難(VH)令(VL)人(Na)相信(VK)。當然(D)，我們(Nh)大(VH)步(Nf)，也(D)已(D)成長(VH)到(P)能夠(D)實事求是(VH)地認真(VH)對待(VE)我們(Nh)的(DE)任務(Na)的(DE)時候(Na)。
3 及物動詞	我(Nh)到底(D)為什麼(D)，只(Da)說(VE)大家(Nh)尊敬(VJ)的台大(Nc)校長(Na)傅斯年(Nb)教授(Na)是(SHI)一(Neu)位(Nf)都(D)彙聚(VAC)在(P)南港(Nc)，但是(Cbb)還(D)沒有(D)做(VC)得(Na)很(Dfa)好(VH)而(Cbb)必須(D)努力(VH)改善(VC)的(DE)地方(Na)而(Caa)複雜(VH)的(DE)問題(Na)，可以(D)毫無(VJ)保留(VC)地表達(VC)意見(Na)，共同(A)討論(VE)，結合(VHC)不同(VH)
4 名詞	合理(VH)的(DE)作法(Na)也許(D)是(SHI)趕快(D)把(P)機身(Na)的重量(Na)大幅(D)減少(VHC)到(P)二十五(Neu)噸(Nf)，然後(D)*程式(Na)分析(VC)的(DE)能力(Na)。三(Neu)、對(P)研究所(Nc)得分配(VD)有(V_2)興趣(Na)者(Na)。意者(Na)請(VF)近日(Nd)的(DE)傳遞(VD)[+nom]之後(Ng)，或許(D)他們(Nh)善意(Na)地表現出(VC)不(D)在乎(VK)，然而(Cbb)有(V_2)種(Nf)
5 定詞	們(Na)不僅(Cbb)是(SHI)願意(VK)，因為(Cbb)沒有(VJ)別(Nes)的課(Na)，還(D)要求(VF)吳(Nb)先生(Na)多(D)加(VC)幾(Neu)*致(VL)癌(Na)病因(Na)的(DE)可能性(Na)越(D)大(VH)，其(Nep)得癌症(Na)的(DE)機會(Na)愈(D)多(VH)。一般(D)在(P)邁入(VCL)如果(Cbb)想(VE)在(P)家(Nc)裡(Ncd)一(Neu)對(P)一(Neu)地教(VC)孩子(Na)，也許(D)不如(VJ)「易(VC)子(Na)而(Cbb)

¹³ 有灰色底的例句是 Sinica Corpus 詞類標記可能有誤的句子。因為 Sinica Corpus 裡面的詞類標記是電腦自動化的，所以有錯誤是在所難免，而且數量不高，投入統計分析不會受到太大的影響。

6 後置詞	。二十餘(Neu)年(Nf)來(Ng)，他(Nh)在(P)幾何(Na)中(Ng)的	基本(A)偏微分(Na)方程(Na)之(DE)成就(Na)對於(P)拓樸學(Na)
* 多(D)問(VE)問題(Na)，仔細(VH)作環境(VC)。必要(VH)時(Ng)，得	學會(VC)保持(VJ)寂靜(VH)。如此(VH)才(Da)能(D)瞭解(VK)	
海(Na)上(Ncd)，一(Neu)艘(Nf)帆船(Na)像(P)樹葉(Na)般(Ng)地	飄浮(VA)著(Di) (頁(Nf)一九五(Neu))，她(Nh)孤獨(VH)	
7 副詞	在(P)我(Nh)無心(D)細說(VE)時(Ng)，總(D)也(D)不經意(D)	提到(VE)，台大(Nc)的(DE)校長(Na)錢思亮(Nb)教授(Na)，
燕子(Na)，在(P)山谷(Na)中(Ncd)來回(VCL)飛行(VA)，快(D)得	像(VG)夜空(Nc)的(DE)流星(Na)。爸爸(Na)說(VE)：有時與(D)，	
的(DE)這(Nep)段(Nf)時間(Na)裡(Ng)，只有(D)馬上(D)大幅(D)地	減少(VHC)資本門(Nb)的(DE)開支(Na)才(Da)能(D)看到(VE)	
8 量詞	的(DE)大(VH)飛機(Na)，裝載(VC)負荷量(Na)四十五(Neu)噸	的引擎(Na)。也就是說(Dk)這(Nep)飛機(Na)的(DE)承載量(Na)
(Nf)的	引擎(Na)。也就是說(Dk)這(Nep)飛機(Na)的(DE)承載量(Na)	
* 防(VC)癌(Na)是(SH)當務之急(Na)；而(Cbb)第一(Neu)次(Nf)得	癌(Na)的(DE)倖存(VH)者(Na)，則(D)開始(VL)嘗試(VF)以(P)	
在(P)山(Na)上(Ncd)。時間(Na)一(Neu)天(Nf)一(Neu)天(Nf)地	過去(VA)，很(Dfa)快(VH)就(D)過(VCL)了(Di)一千(Neu)天(Nf)	

表 6：結構助詞「的」、「得」、「地」在 Sinica Corpus 的例句-後面

後面		
1	(非調) 形容詞	
。二十餘(Neu)年(Nf)來(Ng)，他(Nh)在(P)幾何(Na)中(Ng)的	基本(A)偏微分(Na)方程(Na)之(DE)成就(Na)對於(P)拓樸學(Na)	
把(P)在(P)這(Nep)組織體(Na)下(Ng)的(DE)人(Na)，變(VH)得	真正(A)的(DE)生機化(VHC)，本來(D)組織(Na)與(Caa)生機	
的(DE)二十餘(Neu)幢(Nf)高樓大廈(Na)，霸氣(Na)凜凜(VH)地	雄(A)據(VC)各(Nes)方(Nf)，永(D)不(D)忘記(VK)隨時(D)	
2	不及物動詞	
苦心(Na)，心(Na)裡(Ng)總(D)有(V_2)一(Neu)股(Nf)難忘(VK)的	溫馨(VH)。我(Nh)今天(Nd)能夠(D)在(P)這(Nep)紀念(VJ)	
很(Dfa)大(VH)的(DE)地方(Na)。研究(Na)工作(Na)做(VC)得	出色(VH)的(DE)研究員(Na)，研究群(Na)該(D)受到(VJ)應(D)	
大(VH)步(Nf)，也(D)已(D)成長(VH)到(P)能夠(D)實事求是(VH)	認真(VH)探討(VE)我們(Nh)的(DE)任務(Na)的(DE)時候(Na)。	
3	介詞	
寡核鏈(Na)為(VG)探針(Na)。如果(Cbb)探針(Na)成功(VH)的	與(P)基因(Na)結合(VHC)，那麼(Dk)基因(Na)上(Ncd)這(Nep)	
座(Nf)又(D)一(Neu)座(Nf)的(DE)摩天(A)大廈(Na)建(VC)得	如(P)阿後春筍(Na)般(Ng)地(DE)熱鬧(VH)，被(P)那麼多(Neqa)	
維護(VC)[+nom]上(Ng)的(DE)理由(Na)，洞燭機先(VH)地	為(P)本(Nes)隊(Na)潛滋養(VC)向(P)各(Nes)所(Nc)求借(VD)	
4	及物動詞	
決定(VE)報考(VC)台大(Nc)歷史系(Nc)，我(Nh)好奇(VH)的	問(VE)他(Nh)為什麼(D)學(VC)歷史(Na)？這(Nep)位(Nf)	
生存(VH)的(DE)空間(Na)，它(Nh)無所不在(VH)，大(VH)得	充滿(VJ)宇宙(Na)，小(VH)得(DE)如(VG)空中(Nc)一(Neu)縷	
學術(Na)的(DE)艱巨(VH)任務(Na)，但是(Cbb)滿懷(VJ)信心(Na)地	舉(VC)了(Di)芝加哥(Nc)大學(Nc)與(Caa)約翰(Nb)霍浦金斯	
5	名詞	
一九七八年(Nd)獲有(VJ)數學(Na)諾貝爾(Nb)之(DE)稱(Na)的	費爾茲獎(Nb)，八十年代(Nd)獲(VJ)麥克萊瑟(Nb)獎金(Na)。	
鋪上(VC)薄(VH)被(Na)。深白(VK)自己(Nh)病(VH)了(T)：病	自己(Nh)無法(D)在(P)軌道(Na)上(Ncd)行走(VA)。今年(Nd)	
(VH)得	基礎(Na)，同時(Nd)學校(Nc)也(D)教授(VC)職業(Na)技能(Na)，	
修整(VC)花園(Nc)、清理(VC)環境(Na)、課業(Na)則(D)相當(VH)地		
6	定詞	
、偏微分(Na)方程(Na)等(Cab)數學(Na)和(Caa)物理(Na)中(Ng)的	多(Neqa)個(Nf)領域(Na)都(D)有(V_2)重大(VH)貢獻(Na)，	
淹沒(VC)大(VH)片(Nf)森林地(Na)，也(D)不會(D)逼(VF)得	成千(Neqa)的(DE)原住民(Na)無家可歸(VH)。由於(Cbb)不斷	
，有(V_2)片刻(Nd)的(DE)安靜(VH)[+nom]，接著(D)轟(VC)地	(VH)	
7	副詞	
的(DE)支持(VC)與(Caa)富有(VJ)理想(Na)的(DE)校長(Na)的	全權(D)計畫(VF)下(Ng)，能夠(D)在(P)開學(VH)的(DE)	
用(VC)在(P)研究(Na)上(Ng)的(DE)經費(Na)真(D)是(SH)少	很(Dfa)難(VH)令(VL)人(Na)相信(VK)。當然(D)，我們(Nh)	

	(VH)	
	的(DE)人(Na)時(Ng)，我(Nh)鐵定(D)會(D)努力(VH)認真(VH)地	狠狠(D)瞪(VC)他(Nh)——(Neu)眼(Nf)；當(P)我(Nh)有所(VJ)
8	耐性(Na)尤其(D)不可或缺(A)，何(Nb)先生(Na)從中(D)熟練(VC)的	連接詞 的不僅(Cbb)是(SHI)技術(Na)經驗(Na)的(DE)累積(VC)
	常(D)可(D)發現(VE)，很多(Neqa)時候(Na)，他們(Nh)表現(VC)得	「既(Cbb)保守(VH)又(Caa)激進(VH)」，「既(Cbb)含蓄(VH)
	；三(Neu)只(Nf)鳥(Na)，濃墨(Na)自足(VH)，意態(Na)輕鬆(VH)地	或(Caa)啄(VC)或(Caa)望(VC)向(P)中(Ncd)軸(Na)行列(Na)
9	，這樣(VH)的(DE)父母(Na)就(D)是(SHI)非常(Dfa)成功(VH)的	語助詞 了(T)。第二(Neu)、不要(D)勉強(VH)孩子(Na)去(D)做(VC)
	袋子(Na)，要不然(Cbb)這麼多(Neqa)資料(Na)怎麼(D)拿(VC)得	了(T)? 噢(I)，這(Nep)本(Nf)小冊子(Na)是(SHI)什麼(Nep)?
*	考卷(Na)。青青(Nb)停住(VC)，在(P)心(Na)中(Ng)輕蔑(VH)地	哼(I)了(Di)——(Neu)下(Nf)。地理(Na)老師(Na)進來(VA)，

2.3 分析工具

本研究應用了兩個統計分析工具來分析結構助詞「的」、「得」、「地」。首先，本研究應用了軟體 Rbrul (Johnson 2009) 來執行「變異規則分析」(Variable rule analysis; VARBRUL analysis)¹⁴。VARBRUL analysis 是一種邏輯斯回歸 (logistic regression) 的變體，但是其分析結果比原來的邏輯斯回歸還要方便解讀。由變異規則分析的結果可以看見此三個結構助詞在相對性的統計分析中，比較常與哪些詞類放在前後環境、是否達到統計上的顯著以及此三者混用的情形。因為變異規則分析相關軟體 Rbrul 或 GOLDVARB 皆無支援多項式分析 (multinomial)，只有二項式分析 (binomial)；另也沒有自動找出交互作用關係 (interaction) 的功能。因此，在變異規則分析中，依變數（本研究有三個項次：「的」、「得」、「地」）只能兩兩配對，會做出三個模組。這樣的情形難免會產生誤差（對比多項式分析）。所以本研究另外使用 SPSS 統計軟體執行決策樹分析 (CART (Classification and Regression Tree) analysis (Breiman et al. 1984)) 進行各方面的補足。CART 是一種機器學習 (machine learning) 常用的統計分類模型，可以用來補足變異規則分析的不足，像是資料視覺化、支援多項式分析、自動找出所有自變數的相互影響關係 (interaction) 等¹⁵。在下節，統計分析的結果將會分兩部分呈現，第一是「CART 決策樹分析」，第二是「變異規則分析」。

¹⁴ 有關 VARBRUL analysis 的相關文獻，請參見 Rand and Sankoff (1990), Robinson, Lawrence, and Tagliamonte (2001), Sankoff, Tagliamonte, and Smith (2005) 等文章。另外，「變異規則分析」一般使用的軟體為 GOLDVARB，本研究使用的是 Rbrul，主要原因是其便利性較高，詳細原因請參閱 Ruan (2014)。

¹⁵ Eddington (2010) 比較了 GOLDVARB 及 CART 的優缺點，並提及可以使用 CART 來輔助 GOLDVARB。

3. 分析結果

本節呈現統計分析結果，分為「CART 決策樹分析」以及「變異規則分析」(VARBRUL analysis)。

3.1 CART 決策樹分析

CART 決策樹分析可以用來檢視描述性的統計資料結構，可以表格（表 7）及樹狀圖（圖 1）的方式來呈現。

表 7：結構助詞「的」、「得」、「地」的 CART 決策樹分析表（暗底表示終端點）¹⁶

樹狀結構表格												
節點	的		得		地		總數		預測	父節點	主要的自變數項目	
	百分比	個數	百分比	個數	百分比	個數	個數	百分比			變數	分割值
0	41	4528	27	2944	32	3553	11025	100	的			
1	89	3418	7	282	4	142	3842	35	的	0	後面	定詞; (非謂) 形容詞; 名詞; 語助詞
2	15	1110	37	2662	47	3411	7183	65	地	0	後面	及物動詞; 不及物動詞; 副詞; 連接詞; 介詞
3	13	473	65	2332	22	808	3613	33	得	2	後面	不及物動詞; 副詞; 連接詞
4	18	637	9	330	73	2603	3570	32	地	2	後面	及物動詞; 介詞
5	4	68	93	1601	3	56	1725	16	得	3	前面	及物動詞
6	21	405	39	731	40	752	1888	17	地	3	前面	名詞; 後置詞; 不及物動詞; 副詞; 量詞; (非謂) 形容詞
7	8	228	11	322	82	2426	2976	27	地	4	前面	不及物動詞; 及物動詞; 副詞; 定詞; 量詞
8	69	409	1	8	30	177	594	5	的	4	前面	名詞; 後置詞; (非謂) 形容詞
9	5	78	49	717	46	675	1470	13	得	6	前面	不及物動詞; 副詞
10	78	327	3	14	18	77	418	4	的	6	前面	名詞; 後置詞; 量詞; (非謂) 形容詞
11	6	75	53	717	42	568	1360	12	得	9	前面	不及物動詞
12	3	3	0	0	97	107	110	1	地	9	前面	副詞

從表 7 及圖 1 來看，「後面」的訊息量最多，所以是第一個分割值節點分割依據。節點 1 的解讀為：當「後面」為「定詞」、「(非謂) 形容詞」、「名詞」或「語助詞」時，「的」佔有較大次數比例，因此選擇「的」作為預測值。節點 5 的解讀為：當「後面」為「不及物動詞」、「副詞」或「連接詞」且當「前面」為「及物動詞」時，「得」佔有較大次數比例，因此選擇「得」作為預測值。節點 7 的解讀為：當「後面」為「及物動詞」或「介詞」且當「前面」為「不及物動詞」、「及物動詞」、「副詞」或「定詞」時，「地」佔有較大次數比例，因此選擇「地」作為預測值。

¹⁶ 特別注意的是，本決策樹信度 (Validity) 為 81.7%，意即此決策樹的分類結果，有約八成二的預測是正確的，有一、二成的預測不符合實際資料。

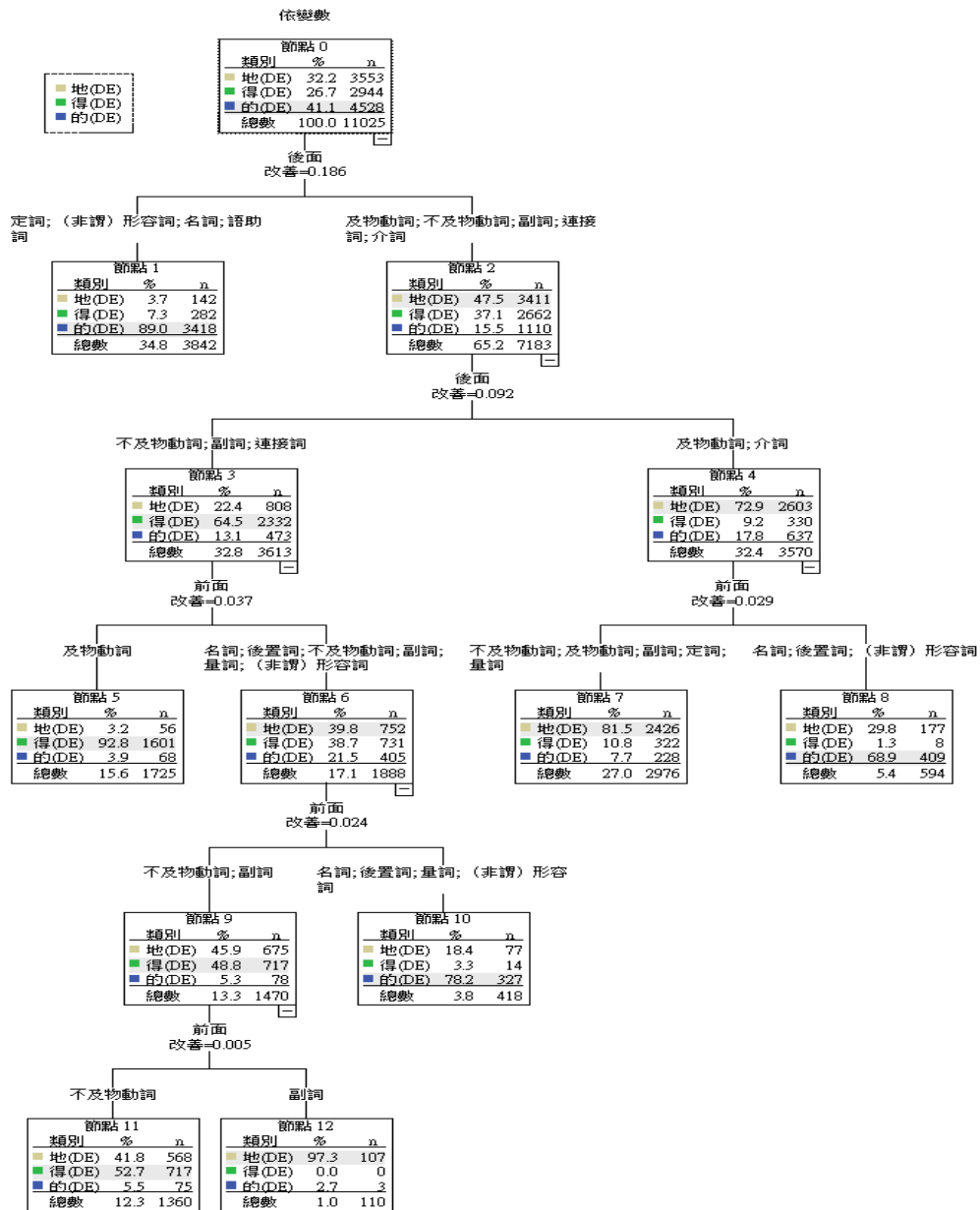


圖 1：結構助詞「的」、「得」、「地」的 CART 決策樹狀圖

CART 的分析結果可以用來預測在什麼樣的情形之下，語言使用者會使用較多

的何種結構助詞。讀者可以在圖 1 看見表 7 裡的描述性資料是有其特殊樹狀結構。然而，CART 的分類結果沒有提供統計顯著性考驗 (Significance Test)，所以本研究再另以變異規則分析來進一步檢視。

3.2 變異規則分析¹⁷

在 CART 分析中發現了「前面」、「後面」具有潛在的交互作用關係，所以在變異規則分析中，本研究使用完全回歸模型 (saturated model)¹⁸，在自變數的分析列表分為「前面」、「後面」、「前面後面交互關係」。因為是二項式 (binomial) 分析的緣故，所以在表 8、表 9、表 10 裡面各有三個模型，分別是「的/的+得+地」、「得/的+得+地」、「地/的+得+地」，各自代表「的」、「得」、「地」為依變數的分析對象 (application value)¹⁹。舉一個解讀的例子：在表 8 裡面，「名詞」對於「的」有正向 (positive) 的影響 (center factor weight ≥ 0.5)；意即，在「前面」是「名詞」時，有相對性較高的加權機率 (center factor weight) 會選擇「的」作為結構助詞。統計結果顯示，「前面」、「後面」未達統計顯著性，然而「前面」、「後面」的交互關係則有達到統計顯著性。

所以在表 8 裡面，結構助詞「前面」為詞類「名詞」、「量詞」、「後置詞」、「及物動詞」、「不及物動詞」、「(非謂) 形容詞」，並且不考慮「後面」詞類時，會有相對性較高的加權機率選擇「的」。「前面」為詞類「及物動詞」、「不及物動詞」時，會有相對性較高的加權機率選擇「得」。「前面」為詞類「定詞」、「副詞」、「(非謂) 形容詞」時，會有相對性較高的加權機率選擇「地」。

¹⁷ 變異規則分析和 CART 決策樹分析不同地方在於變異規則分析是以整個模型來去計算每一個加權機率值 (centered factor weight) 來呈現正向 (centered factor weight ≥ 0.5)、負向 (centered factor weight < 0.5) 影響程度來做為預測值的選擇。然而在 CART 裡是以呈現的次數大小做為預測值的選擇，故二者有可能在一些資料的選擇上會有不同的結果。本研究還是以變異規則分析為主，而以 CART 做為描述性統計 (descriptive statistics) 的結構分析。

¹⁸ 如果只是一般的 ENTER 法，其模型為：依變數~前面+後面。如果是完全回歸模型，其模型為：依變數~前面+後面+前面與後面的交互關係[依變數~自變數 1+自變數 2+(自變數 1:自變數 2)]。完全回歸模型分析出來的結果較為完整。

¹⁹ 三模型的解釋變數 (Nagelkerke R^2) 如下：「的/的+得+地」(76.9%)、「得/的+得+地」(66.6%)、「地/的+得+地」(66.6%)，皆有大於 15%，因此具有解釋力。

表 8：自變數「前面」的變異規則分析²⁰（不考慮「後面」）， $p>.05$

factor	logodds	tokens	的/的+得+地	centered factor weight
名詞	21.207	2692	0.915	>0.999
量詞	5.740	104	0.548	0.997
後置詞	4.641	162	0.846	0.99
及物動詞	3.630	2856	0.216	0.974
不及物動詞	2.937	4580	0.241	0.95
（非謂）形容詞	1.161	140	0.807	0.762
factor	logodds	tokens	得/的+得+地	centered factor weight
及物動詞	13.701	2856	0.685	>0.999
不及物動詞	13.296	4580	0.209	>0.999
factor	logodds	tokens	地/的+得+地	centered factor weight
定詞	28.795	35	0.686	>0.999
副詞	13.328	456	0.934	>0.999
（非謂）形容詞	9.397	140	0.193	>0.999

在表 9 裡面，結構助詞「後面」為詞類「（非謂）形容詞」、「定詞」、「語助詞」，並且不考慮「前面」詞類時，會有相對性較高的加權機率選擇「的」。在「後面」為詞類「名詞」時，會有相對性較高的加權機率選擇「得」。在「後面」為詞類「介詞」、「及物動詞」、「不及物動詞」、「副詞」、「連接詞」、「語助詞」時，會有相對性的機率，會有相對性較高的加權機率選擇「地」。

²⁰ 因為分析結果太長，所以表 8、表 9、表 10 只列出自變數項目對於依變數有正向影響（centered factor weight ≥ 0.5 ）的項目。

表 9：自變數「後面」的變異規則分析（不考慮「前面」）， $p>.05$

factor	logodds	tokens	的/的+得+地	centered factor weight
（非謂）形容詞	15.061	63	0.937	>0.999
定詞	13.962	345	0.728	>0.999
語助詞	0.199	7	0.571	0.55
factor	logodds	tokens	得/的+得+地	centered factor weight
名詞	12.547	3427	0.062	>0.999
factor	logodds	tokens	地/的+得+地	centered factor weight
介詞	19.016	465	0.830	>0.999
及物動詞	3.877	3105	0.714	0.98
不及物動詞	2.450	2136	0.261	0.921
副詞	1.351	1424	0.164	0.794
連接詞	1.351	53	0.321	0.794
語助詞	1.351	7	0.143	0.794

在表 10 裡面，「前面」為「定詞」且「後面」為「及物動詞」時，會有相對性較高的機率選擇「的」²¹。「前面」為「名詞」且「後面」為「副詞」時，會有相對性較高的機率選擇「得」。「前面」為「及物動詞」且「後面」為「定詞」時，會有相對性較高的機率選擇「地」。

表 10：自變數「前面」、「後面」交互關係的變異規則分析， $p<.05$

factor	logodds	tokens	的/的+得+地	centered factor weight
定詞・及物動詞	34.154	5	0.4	>0.999
定詞・名詞	29.97	29	0.276	>0.999
副詞・及物動詞	16.155	258	0.05	>0.999
副詞・名詞	14.839	23	0.348	>0.999
副詞・不及物動詞	14.344	86	0.035	>0.999
不及物動詞・介詞	14.323	313	0.006	>0.999

²¹ 只舉第一項目，其餘各項目類推，不再一一細舉。

(非謂)形容詞·名詞	6.788	83	0.988	0.999
(非謂)形容詞·及物動詞	6.437	33	0.606	0.999
(非謂)形容詞·不及物動詞	5.138	11	0.636	0.997
(非謂)形容詞·定詞	4.579	1	1	0.994
(非謂)形容詞·(非謂)形容詞	3.48	3	1	0.983
後置詞·及物動詞	3.288	22	0.682	0.98
不及物動詞·名詞	2.743	928	0.894	0.965
後置詞·不及物動詞	2.277	34	0.765	0.946
後置詞·名詞	1.872	82	0.951	0.921
不及物動詞·及物動詞	1.808	1903	0.081	0.917
及物動詞·及物動詞	1.769	343	0.146	0.914
後置詞·定詞	1.099	13	1	0.844
及物動詞·名詞	0.827	601	0.714	0.805
不及物動詞·不及物動詞	0.244	836	0.072	0.697
不及物動詞·連接詞	0.238	32	0.188	0.696
factor	logodds	tokens	得/的+得+地	centered factor weight
名詞·副詞	15.239	90	0.089	>0.999
後置詞·及物動詞	14.522	22	0.045	>0.999
名詞·不及物動詞	13.869	248	0.024	>0.999
副詞·介詞	13.472	61	0.016	>0.999
副詞·及物動詞	13.415	258	0.016	>0.999
名詞·及物動詞	13.291	510	0.014	>0.999
名詞·定詞	12.639	139	0.007	>0.999
及物動詞·副詞	1.664	807	0.936	0.841
及物動詞·不及物動詞	1.496	903	0.925	0.817
定詞·名詞	0.331	29	0.034	0.582
不及物動詞·副詞	0.024	492	0.652	0.506

不及物動詞・語助詞	0	4	0.25	0.5
及物動詞・連接詞	0	15	0.733	0.5
及物動詞・語助詞	0	3	0.333	0.5
量詞・名詞	0	40	0.025	0.5
factor	logodds	tokens	地/的+得+地	centered factor weight
及物動詞・定詞	28.387	118	0.025	>0.999
名詞・定詞	27.807	139	0.014	>0.999
不及物動詞・（非謂）形容詞	15.719	8	0.125	>0.999
不及物動詞・定詞	15.704	64	0.297	>0.999
名詞・介詞	15.467	20	1	>0.999
名詞・副詞	15.109	90	0.189	>0.999
及物動詞・及物動詞	14.257	343	0.554	>0.999
及物動詞・名詞	13.998	601	0.018	>0.999
名詞・不及物動詞	13.484	248	0.121	>0.999
及物動詞・副詞	12.991	807	0.027	>0.999
名詞・及物動詞	12.957	510	0.253	>0.999
名詞・名詞	12.527	1641	0.004	>0.999
及物動詞・不及物動詞	12.226	903	0.038	>0.999
（非謂）形容詞・副詞	3.931	4	1	0.966
不及物動詞・語助詞	2.197	4	0.25	0.833
定詞・定詞	2.197	1	1	0.833
後置詞・副詞	2.197	4	0.5	0.833
副詞・定詞	2.197	4	0.75	0.833
副詞・副詞	2.197	23	1	0.833
副詞・連接詞	2.197	1	1	0.833
量詞・不及物動詞	2.197	18	0.5	0.833
量詞・介詞	2.197	6	1	0.833
量詞・及物動詞	2.197	31	0.806	0.833
量詞・名詞	2.197	40	0.075	0.833

量詞・副詞	2.197	4	0.75	0.833
不及物動詞・連接詞	1.099	32	0.5	0.625

在表 8、表 9、表 10 裡面不難發現有重疊的項目出現，這表示「的」、「得」、「地」在這些情形之下，有相對性較高機率會產生混用，重疊混用的情形整理如表 11。

表 11：結構助詞「的」、「得」、「地」變異規則分析自變數重疊

混用項目 自變數	的/得	的/地	得/地
前面	及物動詞 不及物動詞	(非謂) 形容詞	無
後面	無	語助詞	無
前面：後面	定詞.名詞 副詞.及物動詞 後置詞.及物動詞	及物動詞.及物動詞 及物動詞.名詞 不及物動詞.連接詞	名詞.副詞 名詞.不及物動詞 名詞.及物動詞 名詞.定詞 及物動詞.副詞 及物動詞.不及物動詞 不及物動詞.語助詞 量詞.名詞

在表 11，因為「前面」、「後面」在整個完全回歸模型裡面沒有達到統計上的顯著性，所以沒有辦法很有力的提出對於這三個結構助詞選擇的影響性，但是仍然可以看出一個相對性的分佈重疊。如果只探討「前面」，「的」、「得」在「及物動詞」、「不及物動詞」上，「的」、「地」在「(非謂) 形容詞」上可能有混用的情形。如果只探討「後面」，「的」、「地」在「語助詞」上可能有混用的情形。唯上述二者沒有達到統計顯著性。如果探討的是「前面」、「後面」交互關係，因為有達到統計上的顯著性，所以可以提及「的」／「得」、「的」／「地」、「得」、「地」皆有混用現象²²。表 12 為其重疊混用的比例計算整理結果。在表 12，混用代表重疊的項目數（即「定詞.名詞」，「副詞.及物動詞」，「後置詞.及物動詞」，3 個），依變項 1 或依變項 2 代表

²² 本研究有以完全回歸模型投入多項式邏輯回歸 (multinomial logistic regression) 再進行確認。經過分析、考驗，「前面」或「後面」依舊是未達顯著 ($p < .05$)，然而「前面」、「後面」交互關係則有達到統計上的顯著 ($p < .05$)，信度 (Validity) 或正確分類率為 82%。

的是相對性機率高於 0.5 的項目個數。由表 12 可知，「的」／「得」相對性機率高於 0.5 的項目平均重疊個數比例有 17.14%，「的」／「地」為 12.91%，「得」／「地」為 42.05%。顯然，「得」、「地」較為混用²³（嚴戎庚 1999）。

表 12：結構助詞「的」、「得」、「地」在「前面」、「後面」交互關係的重疊比例

依變項 1/依變項 2	混用	依變項 1	依變項 2	混用/依變項 1	混用/依變項 2	平均
的/得	3	21	15	14.29%	20.00%	17.14%
的/地	3	21	26	14.29%	11.54%	12.91%
得/地	8	15	26	53.33%	30.77%	42.05%

4. 討論

本研究旨在以語料庫搭配統計分析的方式進行觀察語言使用者如何使用漢語結構助詞「的」、「得」、「地」的情形。依照分析的結果來回答以下問題：

1. 漢語結構助詞「的」、「得」、「地」的理論是否與實際上的使用相符合？
2. 使用分佈？混用的程度？多少精確數字百分比？
3. 以整體而言，語言使用者是如何使用漢語結構助詞？

4.1 漢語結構助詞「的」、「得」、「地」的理論與實際使用

根據表 4 描述性統計分析的結果發現，「的」、「得」、「地」整體的使用大致與其使用規則（方麗娜 1994；徐陽春 2004；劉雪梅 2006；江勝利 2008；馮元玫 2009）相符合：

體詞性偏正結構	[定語+的+中心語]
謂補結構	[謂語+得+補語]
謂詞性偏正結構	[狀語+地+中心語(謂詞性)]

從表 4 可以看見，結構助詞「前面」的環境裡，出現在「的」前面較高頻率為「名詞」、「不及物動詞」、「及物動詞」；「得」為「及物動詞」、「不及物動詞」；「地」為「不及物動詞」、「副詞」、「及物動詞」、「名詞」。以「的」來說，「名詞」當作定

²³ 整體信度為 70%， $(=(76.9+66.6+66.6)/3)$ 。

語使用，像是「機身的重量」，是最為常見的。其次，「不及物動詞」²⁴當作定語使用，像是「新的原子」。部份「不及物動詞」當作定語來使用則是可能與「地」的用法混用，像是「好奇的（地）問他」。再者，「及物動詞」用作「定語」，像是「大家尊敬的台大校長」。以「得」來說，「及物動詞」當作謂語使用，像是「做得很好」，是最為常見的。其次，「不及物動詞」，像是「變得更成熟」、「少得令人難以相信」²⁵。以「地」來說，「不及物動詞」當作狀語使用，像是「獨立地發展」、「不停地下著」，是最為常見的。「副詞」當作狀語使用，像是「大幅地減少」。「及物動詞」當作狀語使用，像是「關心地問」。「名詞」當作狀語使用，像是「善意地表現」。

另一方面，從表 4 可以看見，結構助詞「後面」的環境裡，出現在「的」後面較高頻率為「名詞」、「及物動詞」、「不及物動詞」；「得」為「不及物動詞」、「副詞」、「及物動詞」；「地」為「及物動詞」、「不及物動詞」、「介詞」。以「的」來說，其中心語是比較名詞（或體詞）性的。「的」後面的「名詞」、「及物動詞」、「不及物動詞」，依序舉例像是：「諾貝爾獎之稱的費爾茲獎」、「世界的國際化」、「難忘的溫馨」。以「得」來說，其後的補語說明前面的謂語。「得」後面的「不及物動詞」、「副詞」、「及物動詞」，依序舉例像是：「做得出色」、「做得很好」、「痛得拋下長刀」。其中，「副詞」補語的中心語在「副詞」之後。以「地」來說，其中心語是比較謂語性的。「地」後面的「及物動詞」、「不及物動詞」、「介詞」，依序舉例像是：「適當地擴充某些研究所」、「大方地一筆勾銷」、「認真地把眼淚往肚裡吞」。其中，「介詞」例子的真正中心語在「吞」這個動詞。

4.2 漢語結構助詞「的」、「得」、「地」使用分佈、混用情形

在變異規則分析的部分（表 8、表 9、表 10），「前面」、「後面」皆無達到統計上的顯著，只有「前面：後面」交互作用關係有達到統計上的顯著。這意味著，語言使用者在使用結構助詞「的」、「得」、「地」時，其前後的交互作用關係會影響到結構助詞的選擇。交互作用達統計上的顯著即是識別整個結構的最主要關鍵，也就是語言使用者一如同徐陽春（2004）所提及的——是先識別出結構後，才決定使用哪一個結構助詞。這也說明了，結構不同帶來標記不同，進而形容不同的意義表達，如同方麗娜（1994）提到的例子：「死的很難看」、「死得很難看」、「死地很難看」，三者不同結構、不同標記、也不同意義。

²⁴ 這類的詞即是一般可以當作謂語使用的「形容詞」，不同於「（非謂）形容詞」。Sinica Corpus 將此類的詞歸納在「不及物動詞」裡面。

²⁵ 「少」為可當作謂語使用的形容詞，同批註 24。

在表 11、表 12 方面，「的」、「得」、「地」雖然有前後環境使用重疊的現象產生，但是重疊的情形除了「得」／「地」較為明顯外，其餘「的」／「得」、「的」／「地」重疊現象不甚明顯。這代表著，此三者在全體的語言使用中，是相對性的傾向分用的。

4.3 漢語結構助詞「的」、「得」、「地」分用與合用的決策

過去文獻提出以語言結構（語義、語用）的許多例證、合用的歧義性、分用的不必要性來探討是否分用或合用這三者。然而，以大量客觀的語料庫語料來直接說明與描述應為更好的依據辦法來說明分用與合用的問題。根據本研究所有的統計分析指出，目前現階段語言使用者對漢語結構助詞「的」、「得」、「地」是呈現分用的情形。語言使用也受到媒體、報章雜誌的影響甚深，語言教學也應順應社會大眾的使用方式。本研究建議，此三者應順應整體語言使用情形，採取分用。

4.4 漢語結構助詞認知模型與漢語教學上的建議

綜上所述，本研究以統計分析及前人文獻提出一漢語結構的認知模型：語言使用者在使用漢語結構助詞「的」、「得」、「地」時，首先要(1)考量語境 (context)，(2)考量結構助詞前後的關聯性 (relation)，(3)決定其使用結構，最後選用漢語結構助詞的標記 (marker choice) (徐陽春 2004)，而此模型將產出三種構式 (construction)。以 Goldberg (1995) 構式語法(Construction Grammar)的觀點，每一種構式都有其本身的結構意義，是一種「形式-語義」(form-meaning) 的配對。因此，「的」一架構指出「從屬狀態」的意義，「得」一架構指出「包含狀態」的意義，而「地」一架構指出「狀態進行」的意義。再以方麗娜 (1994)「死 de 很難看」例子說明，「死」與「很難看」放置「的」一架構下，「很難看」從屬於「死」，因為「很難看」不可能從屬在「死」，一定會是跟「死」有關的名詞，亦即「死」的人、樣子等等「很難看」，故能得知其名詞有省略情形。如果放置在「得」一架構下，「死」是一個動作，「很難看」是包含動作整體的狀態（相對地，也就是透過「死」表達「很難看」的狀態），亦即「死」是在「很難看」的狀態下。如果放置在「地」一架構下，「死」是一個動作初始的狀態，而「很難看」是後續進行的動作，也就是「極度的狀態下進行很難看」，亦即「極度很難看」，如圖 2 的圖解：

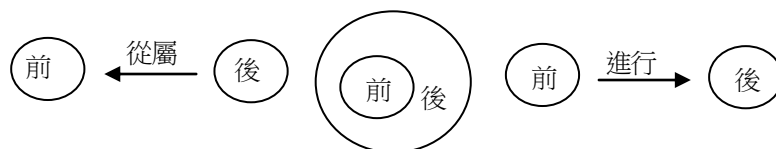


圖 2：「的」（左）、「得」（中）、「地」（右）結構助詞認知概念圖

由本研究統計分析的結果來看，母語漢語教學或對外漢語教學在漢語結構助詞的教學上，不建議教學者只採用結構助詞的看前看後，像是「的」使用在定語與中心語之間（偏正結構），「得」使用在謂語與補語之間（謂補結構），「地」使用在狀語與謂詞性中心語之間（偏正結構），而建議除以上的提示之外，更應該強調結構助詞的「前後」關聯性（前後環境有顯著性交互作用）。這個結果也如同嚴戒庚（1999）所提，若只單論以上詞性辨別角度的方法是無法解釋此三者的使用，而有可能在教學上使學習者混淆更甚。另外，教材編輯者亦可以本研究列出的語料分析結果作為參考，編寫出較為適合、清楚的用例給予教學者或學習者應用。

5. 結論

本研究以統計分析的方式證實了「的」、「得」、「地」的使用規則與實際上的使用是相符合的。在使用結構助詞的認知上，因前後環境的交互作用，產生了選用結構，最後以「的」、「得」、「地」作為標記。量化資料清楚說明「的」／「地」，「的」／「得」混用尚輕，較為混用的是「得」／「地」。對於此三個漢語結構助詞的分用與合用的建議，應以整體社會語言使用者使用情形來決定。本研究建議，分用比較符合現階段整體的使用方式。本研究也綜合了統計的分析結果、Goldberg (1995) 的構式語法與前人文獻，提出了漢語結構助詞認知概念模型，並給與在漢語教學上些許建議，應該強化漢語結構助詞前後關聯的教學，不能只是給與詞性及結構。

本研究的研究限制在於語料庫只提供詞類標記，沒有剖析結構的資料，因此未考慮語法結構。此外，此三者的使用情形應該還有一些非語言的因素造成影響，未來的研究可以再納入其它因素再細為研究。

引用文獻

- Breiman, Leo, Jerome Friedman, Charles. J. Stone, and R.A. Olshen. 1984. *Classification and Regression Trees*. Chapman and Hall.
- Eddington, David. 2010. A comparison of two tools for analyzing linguistic data: logistic regression and decision trees. *Italian Journal of Linguistics* 22: 265-286.
- Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed effects variable rule analysis. *Language and Linguistics Compass* 3.1: 359-383.

- Rand, David and David Sankoff. 1990. *GoldVarb: A variable rule application for the Macintosh*. Montreal, Canada: Centre de recherches mathématiques, Université de Montréal [Mathematics Research Center, University of Montreal.]
- Robinson, John, Helen Lawrence and Sali Tagliamonte. 2001. *Goldvarb 2001: A Multivariate Analysis Application for Windows*. University of York, Department of Language and Linguistic Science. Available at:
http://www.romanistik.uni-freiburg.de/pusch/Download/variacionismo/GoldVarb2001_User_manual.pdf.
- Ruan, Jia-cing. 2014. A corpus study on 「begin」「start」 in academic writing: A VARBRUL approach. *Open Journal of Modern Linguistics* 4: 260-274.
- Sankoff, David, Sali Tagliamonte, and Eric Smith. 2005. *GoldVarb X*. Department of Linguistics, University of Toronto, Toronto, Canada.
- 方麗娜。1994。〈結構助詞「的」、「地」、「得」用法解析〉，《台南師院學報》，第27期，95-108。[Fang, Li-na. 1994. The study of utilization of 結構助詞－「的」「地」and「得」。 *Journal of National Tainan Teachers College* 27: 95-108.]
- 田野。2004。〈「的」「地」「得」的合用與分用〉，《鹽城師範學院學報（人文社會科學版）》第24卷第3期，76-78。[Tian, Ye. 2004. On the usage of the three cognate Chinese characters – 「的」, 「地」 and 「得」. *Journal of Yancheng Teachers College (Humanities and Social Sciences)* 24.3: 76-78]
- 呂叔湘。1980。《現代漢語八百詞》。（第一版）。北京：商務印書館。[Lü, Shu-xiang. 1980. *The 800 Words of Modern Chinese (1st ed.)*. Beijing: The Commercial Press.]
- 胡裕樹。1992。《現代漢語》。（增訂版）。台北：新文豐出版公司。[Hu, Yu-shu. 1992. *Modern Chinese (Rev. ed.)*. Taipei: Shin Wen Feng Publishing Company.]
- 詞庫小詞。1995。《中央研究院平衡語料庫的內容與說明（修訂版）——技術報告95-02/98-04》，臺北：中央研究院資訊科學研究所。[CKIP. 1995. *The Revised Technical Report to the Sinica Corpus of Academia Sinica 95-02/98-04*. Taipei: Institute of Information Science, Academia Sinica.]
- 馮元玫。2009。《現代漢語結構助詞「的、地、得」的語法分析及教學應用》，未出版碩士論文，臺灣中國文化大學。[Feng, Yuan-mei. 2009. *The Grammatical Analysis and Pedagogical Applications of the Structural Particles 「的」, 「地」 and 「得」 in Mandarin Chinese*. Unpublished Master Thesis. Taipei: Chinese Culture University.]

- 江勝利。2008。〈能否把結構助詞「的、地、得」看成一個詞〉，《黃山學院學報》，第 10 卷第 6 期，101-104。[Jiang, Sheng-li. 2008. Whether 的, 地, 得 can be regarded as one structural particle. *Journal of Huangshan University* 10.6: 101-104.]
- 徐陽春。2003。〈結構助詞「得」與「的」的分合〉，《南昌大學學報（人社版）》，第 34 卷第 3 期，153-156。[Xu, Yang-chun. 2003. Structural particle 「得」 should not be replace by「的」. *Journal of Nanchang University (Humanities and Social Sciences)* 34.3: 153-156.]
- 徐陽春。2004。〈結構助詞「的」與「地」的分合問題〉，《南昌大學學報（人社版）》，第 35 卷第 6 期，151-155。[Xu, Yang-chun. 2004. Structural particle 「地」 should be replace by「的」. *Journal of Nanchang University (Humanities and Social Sciences)* 35.6: 151-155.]
- 劉月華。2001。實用現代漢語語法。（第一版）。北京：商務出版社。[Liu, Yue-hua. 2001. *Practical Modern Chinese Syntax* (1st ed.). Beijing: The Commercial Press.]
- 劉雪梅。2006。〈結構助詞「的」「地」「得」用法分析〉，《現代語文（語言研究版）》，第 12 期，98-99。[Liu, Xue-mei. 2006. The analysis of usage of the structural particles 「的」, 「地」 and 「得」. *Modern Chinese* 12: 98-99.]
- 嚴戎庚。1999。〈「的、地、得」新論〉，《新疆大學學報（哲學社會科學版）》，第 27 卷第 4 期，94-96。[Yan, Rong-geng. A new theory to 「的」, 「地」 and 「得」. *Journal of Xinjiang University (Social Sciences)* 27.4: 94-96.]

[審查：2016.03.25 修改：2016.07.21 接受：2016.09.07]

阮家慶

Jia-cing RUAN

62102 嘉義縣民雄鄉大學路一段 168 號 國立中正大學語言學研究所

No.168, Sec. 1, University Rd., Min-Hsiung Township, Chia-yi County 62102, Taiwan

ac27037@gmail.com

A Corpus-based Quantitative Study of the Structural Particle in Modern Chinese

Jia-cing RUAN

**Graduate Institute of Linguistics
National Chung Cheng University**

Abstract

This study, with a perspective of Corpus Linguistics, tries to utilize a CART analysis (Breiman et al., 1984) and a VARBRUL analysis (Rand and Sankoff, 1990; Robinson, Lawrence, and Tagliamonte, 2001; Sankoff, Tagliamonte, and Smith, 2005) with a large number of corpus data to prove that the grammar rules of the three structural particles “de” in Chinese are consonant with its real use by language users, and to prove that language users are aware of the structure before selecting an appropriate particle (徐陽春, 2004). Moreover, the quantitative analyses show the lower degree of misuse between 「的」／「地」 or 「的」／「得」, but higher degree of misuse between 「得」／「地」. This study suggests that independent usage for each structural particle is required, since the quantitative analyses of corpus data show the consonance with the real use of language users. Additionally, three conceptualized models for the three structural particles have been proposed. This study expects to contribute to the composition of Chinese dictionary, Chinese language teaching, language policy, and natural language processing.

Keywords: Chinese, structural particle, corpus, VARBRUL, CART