

語料統計分析法之例證及其警務應用

古永昌*、邱昭彰、蔡昆泓

摘要：字詞頻統計法是一種以觀察語詞出現頻次，了解語詞在社會中發展情形常用的方法。曾榮汾教授以情色文學為例，建立了以字詞頻語料統計分析的流程，並得到情色語料之關鍵詞語，用以協助偵查人員搜尋及打擊情色網站。本研究中採用此字詞頻語料統計分析流程，並改善其以人工抽取關鍵情節及切詞的方法，以及在語料統計分析上增加了「時間」屬性，使得此語料統計分析法更有效率及對語詞發展能有更多的解釋與認識。另外，由於警學是複雜的，研究的範疇幾乎涵蓋了所有的社會學問，極需要不同領域的知識融入整合，以契合社會治安的需求，故本研究亦提出了一種可行的犯罪社群語料分析模式，期能在警學研究及提昇偵防觀念與技術上有所幫助。

關鍵詞：語料分析(Linguistics Analysis)、犯罪社群語料分析(Criminal Linguistics Analysis)、語言剖繪方法(Criminal Linguistics Profiling)

綱目：

- 一、前言
- 二、字詞頻語料統計分析
- 三、考量時空因素之社群語料分析及例證
- 四、社群語料分析之警務運用
- 五、結論

一、前言

語言是一種社會現象，是人類互相溝通及進行思維交流的工具。一般而言，即便是在現今資訊網路十分便利的情況下，並非每一個人都有機會跟世界上所有人接觸。也就是說，每一個人所能碰到或交流的對象是限定在一個範圍內，其差別僅在於因個人與他人交際的程度而有範圍大小或疏密程度不同。人們因環境的阻隔，所用的語言也就因時因地而有所不同。一群人可以藉他們所說的話而彼此聯繫，便可以說是同屬一個語言群(董同龢，1989)。在一個語言群內，每一個成員在一定的情況下，能夠發出某些特殊的聲音，喚起別人的回應；同時別人所發出的某些聲音也能得到自己的回響，語言的性質及價值便表現其中。因此，在同一個範圍內，人們約定俗成用以表達的那些聲音、符號，自然而然使彼此能夠溝通聯繫，這樣一種語言即成為了這個範圍內的人們所共有使用。在社會上，各行各業或不同的社群的人都有他們自己的一套話，階層不同，語言的使用就有了分別。

*古永昌，中央警察大學教務處技士，於元智大學資訊管理學系博士班進修，unicon@mail.cpu.edu.tw。邱昭彰，元智大學資訊管理學系教授，imchiu@saturn.yzu.edu.tw。蔡昆泓，電信警察隊副中隊長，tpc1002@dgt.gov.tw。

文字則是從語言演化而來，透過文字能把聽得見的聲音，轉變成看得見的形體。事實上，文字是為了紀錄語言而發明的一種書寫符號系統，透過通過書寫符號系統把語言符號紀錄下來，使語言的口頭形式書面化。它克服了語言一發即逝的問題，讓語言轉變為文字形式，得以傳於異地、留於異時，也讓不同的人透過文字可以相互進行思想交流或協調彼此的行動(葉蜚聲與徐通鏘，1993)。語言雖和文字並非存有必然的關係，但在同一個語言群內，其成員必須對所使用的語言或文字彼此認識方能順利地溝通與交流。另一方面，由於人們可以參與不同種類的社群，如知識性社群、娛樂性社群等，雖其所屬之語言群或有不同，但社群成員要能夠互相溝通、交流亦有賴於使用共同語言或文字。為了能探索社群中語言或文字在時間或空間上遷移情況，語料統計分析恰提供一項有利的觀察。

在本研究中，將採用語料統計分析試著對社群中語料增長情形作觀察，並思考在警務上可能的應用。

二、字詞頻語料統計分析

曾榮汾教授於 1989 年進行犯罪語料庫建立之研究時，採用了字詞頻語料統計分析法進行樣本語料的特性分析，並以「網路情色文學」為例(曾榮汾，1989)，其分析方法如下：

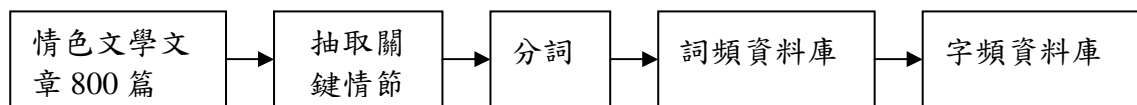


圖 1：字詞頻語料統計分析流程—以情色文學為例

資料來源：曾榮汾，1989

圖 1 中有幾個重要觀點需要說明的，包括：

1. 研究中之樣本係以網站搜尋方式尋找情色網站，並從中取出情色文章。網站可視為一種虛擬社群，透過網站將有共同興趣的特定或不特定人聚集在一起，這一群人有共同的語言及詞彙，並透過發表文章的形式交流。情色文學只是用來觀察情色網站這一個虛擬社群的例子，事實上，不同的社群有其具代表，且可被觀察的語言標的，觀察這些不同社群的語言標的，即可以了解各個社群所欲彰顯的意涵及特質。
2. 抽取關鍵情節指的將文章中較屬於情色用語的段落挑選出來，然後再由資深的語料工作者對這些關鍵情節、段落進行人工圈詞，在 800 篇的情色語料中挑選出所有可能，且合乎語法結構的語詞，再將這些語詞建置成為資料庫。在此步驟，由資深的語料工作者進行人工圈詞是重要的，但若語料量龐大，不免有耗時費力的現象。
3. 至於分詞至產出詞頻或字頻資料庫，其處理方式係進行情色文學字詞頻統計並建立統計圖表，接著與教育部年度語詞作比較，而能夠分隔出情色文學中屬於個別與共有部份的字詞。其中個別部份屬於情色樣本中之專用語詞；共有部份則屬於一般性語詞。也就是說若從社群的觀點來看，社群的語料可以如下表示：

若以 CTs (Community Terms) 為社群語料的集合，以 STs (Special Terms) 代表社群中專用語，而以 GTs (General Terms) 代表社群中一般性語詞，故其關係式可為：

$$CTs = STs + GTs$$

(1)

由(1)可知，一般性語詞如同篩選器般，透過此篩選器，社群專用語則隨之而來。故建立一個完善標準的一般性語料庫對語料分析尤為重要。

經由前述的說明，可以再將圖 1 的分析流程抽象化，並得到圖 2 之社群語料分析流程。

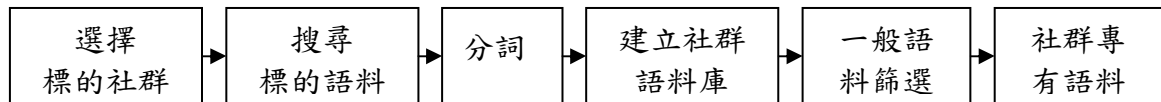


圖 2：社群語料分析流程

資料來源：本研究整理

三、考量時空因素之社群語料分析及例證

從上述的語料分析方法中，不難觀察出尚有可以改善的地方，包括：

1. 語料「分詞」：

「分詞」的目的在讓語料從「句子」結構轉變成「詞組」的結果，情色文學採用人工切詞或要藉助資深的語料工作者方能精確地圈選出具有代表意義的情色語詞。但若資料量龐大時，這種方法則顯得不經濟。故若能找到一個演算法，透過程式語言設計，讓電腦協助把可能的語詞挑選出來，雖不免有非語詞的情況發生，但這時再加上專業人員的觀察，或許提高分詞的效率。

2. 語料僅具空間性

先前的作法係蒐錄情色文章，並將其作語料整理，所得結果僅是語料的數量，這屬於觀察語料在空間上的累積現象。而事實上，語料原本即具時間性和空間性。故在社群語料分析方法中若能加上對時間的觀察，將能對社群語料有著更完整的理解。

在本研究中，將驗證社群語料分析法仍屬可行，並期能改善原先分析法中「人工分詞」及「語料僅具空間性」的問題，除能在「分詞」上提高效率，並提供對語料在時間和空間上的完整觀察。研究設計及步驟將參照圖 2。

(一)選擇標的社群

在標的社群的選擇上，將以頗負盛名的「資訊管理季刊(MIS Quarterly, MISQ)¹」為例子。從圖得知，在近 10 年的發展中，MISQ 為資訊管理領域排名第一的學術期刊，如圖 3。它由美國明尼蘇達大學(Carlson School of Management, University of Minnesota)出版，內容主要探討資訊科技在管理或組織議題上的發展，刊載其上的文章或多或少都指引了資訊管理領域的發展，因此也吸引了資訊管理領域的學者、研究人員們的將最新的研究成果藉由這本期刊發表。

¹ MISQ website, <http://www.misq.org/>

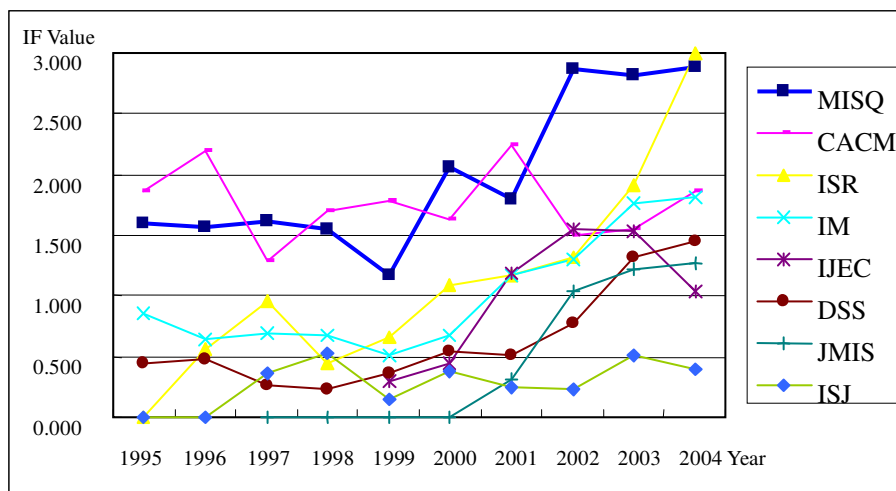


圖 3: 資訊管理領域中主要期刊影響指數(Impact Factor)趨勢圖
資料來源：整理自 SSCI/SCI Cited Index

MISQ 的作者係由一群、分處世界各地的資訊管理領域學者及研究人員所組成，經過謹慎審查後，以每三個月一期的方式刊載傑出的研究成果，並藉以知識交流。因為有共同的平台、共同的語言(以英文發表)，故將 MISQ 視為一虛擬社群。

(二)搜尋標的語料

在研究中，搜集了 MISQ 自 1977 年(1:1)至 2006 年(29:4)共 30 年，計 770 篇的文章。每一篇文章將紀錄其刊載時間(以年為單位)、文章標題及摘要內容，並將這些資料建立為 MISQ 語料庫。其中「文章標題」為作者對此一文章的命名，並為了讓其他的讀者能看到「文章標題」便能連結到文章可能的內容，故此「文章標題」隱含了更多資訊管理領域的專用術語及意涵；而「文章摘要」則是作者表明其研究意涵及與其他術語、知識的連結，在簡短、扼要的表達下仍能讓讀者清楚明白該研究的過程及欲解決的問題。因此，採取「文章標題」及「文章摘要」足可表示該文章的內容，亦能清楚地把作者的思考與讀者連結在一起。

(三)分詞

「分詞」的目的在於挑選出所有可以的語詞，同時為了提高分詞的效率，在作法則採用 PAT tree 結構進行分詞的工作。在資訊擷取技術中，PAT tree 是種有效率的資料結構，它是由 Gonnet 改良自 Morrison' s PATRICIA 演算法(一種用來實際擷取數字或符號資訊碼的演算法)，用來將字串中連續的資料予以索引化(indexing)，而且找出當字串每一個字移為字首時可能的語詞組合(Gonnet, Baeza-Yates, and Snider, 1992)。PAT tree 最大的特徵在它所處理字串的結果是屬於一種 semi-infinite (Manger and Baeza-Yates, 1991)，並把相對於原字串的次級字串的值(substream value)儲存在 PAT tree 的資料結構裡。在文本全文索引中使用 PAT tree 資料結構，所有可能的字串特徵(包括它們在文件中的次數)，都能以一種很有效率的方式被擷取出來(Chien, 1997)。以下以「**語詞自動抽取**」作為 PAT tree

產生出 semi-infinite 的說明：

原始字串	語詞自動抽取
semi-infinite	語詞自動抽取 0 0
(字串逐字往句 首推移)	詞自動抽取 0 0 0
	自動抽取 0 0 0 0
	動抽取 0 0 0 0 0
	抽取 0 0 0 0 0 0
	取 0 0 0 0 0 0 0

PAT tree 資料結構處理文本全文檢索時尚有幾點需要考慮的，即是：

1. 文本中的「字串」

字串指的是：一段敘述，並以標點符號作分隔。

2. 字串中「單元完整性(unithood)」與「主題代表性(termhood)」(Kageura and Umino, 1996)

「單元完整性」指的是做為語詞的字串是否屬於語言結構(linguistic structure)上的完整單位，如詞(words)或詞組(phrases)；「主題代表性」則是指此一語詞是否具代表性並和其他語詞作區別。

3. 停用詞(stop word)

一般來說，停用詞指的是介詞、連接詞或定詞，較常出現在字串的首尾，一旦出現了停用詞，那這一個語詞將被刪去。但如果停用詞出現在字串的中間時(part of MIS)，只要這一個語詞出現的頻率夠高，這一個語詞仍是重要的(林頌堅，2004)。在本研究所用的停用詞如附錄。

(三)建立字詞頻資料庫及社群專用語料庫

當社群的語料經過 PAT tree 資料結構的分詞處理後，所有可能的語詞就很容易地從字串中挑選出來了，同時在同一個字串中，愈長的語詞包含了較小的語詞，也解釋了更多的訊息，如：

例 1：

“management information systems (管理資訊系統)” 包含了
 “management information (管理資訊)” 及 “information systems (資訊系統)”，也包含了
 “management (管理)”、“information (資訊)” 及 “systems (系統)”

例 2：

“technique acceptance model (科技採用模式)” 包含了
 “technique acceptance (科技接受)” 及 “acceptance model (接受模式)”，也包含了
 “technique (科技)”、“acceptance (接受)” 及 “model (模式)”

在例 1 及例 2 中，若以資訊管理領域角度來看，“management information systems”、

“management information”、“information systems”、“management”、“information” “systems”、“technique acceptance model”、“technique”及“model”具有資訊管理知識中的完整意涵，而“technique acceptance”、“acceptance model”及“acceptance”則屬於訊息隱晦不明或屬於一般性的用語。而事實上能夠這樣判讀，須具備幾個要素，即是：

- 1.語詞的出現的頻次(即同樣語詞重複出現的次數)
- 2.透過一般性語料庫的篩選，在本研究中則是以 6000 個常用英文單字作一般語料庫
- 3.以及，對社群所屬某領域的專業知識

通過前述步驟後，以下則分別顯示 MISQ 字詞頻語料統計(完整資料請參閱附錄 2) 及社群專用語料，說明如下：

- 1.表 1 為 MISQ 語料詞組統計。從中可以清楚看出資訊管理的內涵，能透過這些語詞的彙整，了解資訊管理社群涵括的範疇。其中「詞組頻次」指的是此一詞組出現的次數；「文章頻次」指的是此一詞組在 770 篇文章中出現的次數，如「INFORMATION SYSTEMS」即在 243 篇文章中出現。
- 2.表 2 為 MISQ 語料縮寫字詞統計。縮寫字更為代表資訊管理社群專用語詞，並能與其他領域作區別，屬資訊管理領域中之專用字，亦屬此領域中較為重要的概念。
- 3.表 3 為就 MISQ 以時間和空間的觀察角度的語料統計彙整表。以每 3 年作一觀察區間，可以看出每一個重要的概念是在何出現，及其在各區間的變化情形，以此可以了解資訊管理領域的發展或研究趨勢。
- 4.表 4 為 MISQ 語料單字統計。為資訊管理領域中較常使用的單字。

表 1：MISQ 語料詞組統計

序號	詞組	中譯	詞組 頻次	文章 頻次
1	INFORMATION SYSTEMS	資訊系統	350	243
2	INFORMATION TECHNOLOGIES	資訊技術	193	127
3	CASE STUDY	個案研究	66	48
4	DECISION MAKING	決策制訂	68	40
5	DECISION SUPPORT SYSTEMS	決策支援系統	46	37
6	SYSTEMS DEVELOPMENT	系統發展	33	30
7	MANAGEMENT INFORMATION SYSTEMS	管理資訊系統	32	29
8	DECISION MAKERS	決策制訂者	45	28
9	DATA PROCESSING	資料處理	33	25
10	INFORMATION MANAGEMENT	資訊管理	29	24
11	DATA COLLECTION	資料收集	29	24
12	COMPETITIVE ADVANTAGE	競爭優勢	29	21
13	DEVELOPMENT PROCESSES	發展流程	28	21

14	EXPERT SYSTEMS	專家系統	43	19
15	SOFTWARE DEVELOPMENT	軟體發展	26	19
16	SYSTEMS ANALYSTS	系統分析師	27	18
17	TOP MANAGEMENT	高階管理	24	16
18	ORGANIZATIONAL CONTEXTS	組織脈絡	20	16
19	USER INVOLVEMENT	使用者涉入	31	15
20	USER SATISFACTION	使用者滿意度	30	15
21	DECISION SUPPORT	決策支援	20	15
22	MIS RESEARCH	資訊管理研究	19	15
23	EASE OF USE	系統易用性	29	14
24	NEW TECHNOLOGIES	新科技	17	14
25	PLANNING PROCESSES	計畫流程	16	14
26	END-USER COMPUTING	端末使用者運算	22	13
27	CRITICAL SUCCESS FACTORS	關鍵成功因素	20	13
28	INFORMATION SYSTEMS MANAGERS	資訊系統管理者	17	13
29	COMPUTER SYSTEMS	電腦系統	15	13
30	USE OF INFORMATION TECHNOLOGY	資訊科技使用	14	13

表 2：MISQ 語料縮寫字詞統計

序號	縮寫字	全稱	中譯	詞組 頻次	文章 頻次
1	MIS	Management Information System	管理資訊系統	325	182
2	IS	Information System	資訊系統	560	129
3	IT	Information Technology	資訊科技	499	88
4	DSS	Decision Support System	決策支援系統	115	34
5	EDP	Electronic Data Processing	電子資料處理	30	11
6	GDSS	Group Decision Support System	群體決策支援系統	48	10
7	CIO	Chief Information Officer	資訊主管	22	10
8	DP	Data Processing	資料處理	19	10
9	GSS	Group Support System	群組支援系統	43	9
10	EDI	Electronic Data Interchange	電子資料交換	29	9
11	ES	Expert System	專家系統	31	7
12	CSFs	Critical Success Factors	關鍵成功因素	18	7
13	EIS	Executive Information System	主管資訊系統	32	6
14	IBM	International Business Machine	國際商務機器公司	17	6
15	CEO	Chief Executive Officer	執行長	12	6

16	EMS	Electronic Meeting System	電子會議系統	20	5
17	ISD	Information System Development	資訊系統發展	15	5
18	ERP	Enterprise Resource Planning	企業資源規劃	16	4
19	BPR	Business Process Reengineering	企業流規再造	14	4
20	TAM	Technology Acceptance Model	科技採用模式	8	4
21	KBS	Knowledge-Based Systems	知識庫系統	20	3
22	KM	Knowledge Management	知識管理	15	3
23	SISP	Strategic Information System Planning	策略資訊系統規劃	14	3
24	SIM	Society for Information Management	社區資訊管理	11	3
25	IOS	InterOrganizational Systems	跨組織資訊系統	10	2
26	GIS	Geographical Information Systems	地理資訊系統	6	1
27	EUC	End-User Computing	端末使用者運算	4	1

表 3

表 4：MISQ 語料單字統計

序號	單字	中譯	詞組 頻次	文章 頻次	序 號	單字	中譯	詞 組 頻 次	文章 頻次
1	SYSTEM	系統	534	318	39	CHARACTERISTICS	特性	83	59
2	ORGANIZATION	組織	383	271	40	PLANNING	規劃	90	57
3	DEVELOP	發展	252	225	41	ACTIVITIES	活動	73	56
4	MODEL	模式	388	203	42	EMPIRICAL	實證的	60	56
5	TECHNOLOGY	科技	254	172	43	QUALITY	品質	72	55
6	INFORMATION	資訊	268	167	44	MANAGERIAL	管理的	62	53
7	APPROACH	方法	224	156	45	STRATEGIC	策略	84	52
8	MANAGEMENT	管理	207	143	46	METHODS	方法	83	51
9	RELATIONSHIP	關係	172	132	47	PERSPECTIVE	觀點	73	51
10	DEVELOPMENT	發展	164	126	48	ASSOCIATED	關聯	55	51
11	ORGANIZATIONAL	組織的	174	123	49	COMMUNICATION	通訊	69	49
12	APPLICATION	應用	170	122	50	RESOURCES	資源	68	49
13	PROCESS	處理	188	120	51	PERCEIVED	感知	58	49
14	ANALYSIS	分析	154	116	52	EXPERIENCE	經驗	56	47
15	USERS	使用者	210	113	53	TECHNIQUES	技術	69	46
16	SUPPORT	支援	154	111	54	PROCESSES	程序	67	46
17	IDENTIFY	定義	112	109	55	BEHAVIOR	行爲	66	45
18	MANAGERS	管理者	169	107	56	CONTROL	控制	56	45
19	EFFECT	效果	143	104	57	INDUSTRY	產業	63	44
20	INDIVIDUAL	個人	138	96	58	TECHNICAL	技術的	51	43
21	GROUP	群體	172	95	59	STRUCTURE	結構	56	41
22	DECISION	決策	133	89	60	METHODOLOGY	方法學	56	41
23	DESIGN	設計	129	82	61	COMPANY	公司	53	41
24	PROJECT	專案	148	81	62	FUNCTION	功能	56	40
25	KNOWLEDGE	知識	177	78	63	TASKS	任務	52	40
26	PERFORMANCE	效能	145	78	64	MANAGING	管理	50	39
27	FACTORS	因素	112	78	65	STRATEGY	策略	56	38
28	BUSINESS	企業	115	76	66	EFFECTIVENESS	有效的	54	38
29	FRAMEWORK	架構	111	71	67	EVALUATION	評估	53	38
30	COMPUTER	電腦	104	68	68	ELECTRONIC	電子	50	36
31	FIRMS	公司	112	66	69	EXECUTIVES	執行	56	35
32	SOFTWARE	軟體	118	65	70	PERCEPTIONS	感知	54	35
33	VALUE	價值	98	63	71	PEOPLE	人們	54	35
34	BENEFITS	效益	89	63	72	REQUIREMENTS	需求	53	35
35	EFFECTIVE	有效	71	62	73	PRODUCTIVITY	產能	54	33
36	DIFFERENCES	差異化	80	61	74	TRAINING	訓練	53	31

37	ENVIRONMENT	環境	70	60	75	INNOVATION	創新	54	28
38	IMPLEMENTATION	實作	91	59	76	NETWORK	網路	53	23

四、社群語料分析之警務運用

在本研究中，「社群」僅是一種概念，並不限定必須在現實社會中聚集在一起的團體。事實上，「社群」的概念包括「虛擬」與「實體」社群。它亦具備二種形式：

- 1.由某種意願或興趣將一群人集合在一起共同經營或參與此一社群的活動、運作，如 MISQ 社群、網路家族、讀書會、幫派、小團體等。
- 2.由不同的人在不同的時間、地方或社群從事同一類型事件、回應或觀點，可視為就某事件、回應或觀點的社群，如 MISQ 中在不同時間，由不同的作者在其文章中回應或進行有關 MIS、DSS、KM 或 TAM 的研究，或在不同時地，由不同的人犯下搶劫、擄人勒索等案件，雖犯罪者未必聚在一起就所犯下的案件，但因屬同一類型案件，其間應有相類似的特質。

從上述的觀點來看，本研究所進行之以時間、空間為考量基礎的社群語料分析似可應用於對幫派或同一類型犯罪的語言特徵分析上，從選定的幫派或某類型的犯罪案件中，收集有關的語料(如蒐集相關的刑案紀錄、訪談犯罪者等)，再透過本研究所採取的語料分析方法進行關鍵語詞萃取，進而建立各類型語料庫。

另一方面，由於人們將其內心表現於外的方式係透過「行為」或「語言」，而犯罪偵查的概念中，即從「行為」去追溯可能的犯罪者，而當「行為」被隱匿的很好時，追查的行動似乎就陷入膠著；倘「語言」也能協助犯罪偵查，亦即使用語料庫中的關鍵語詞，將可能的犯罪者、時間、地點或情境連結起來，那麼追查犯罪者的方式便多了一種選擇，尤其在資訊科技，網路無遠弗屆的今日，新型態的犯罪愈來愈多是透過網路來傳遞訊息，如何能夠新型態犯罪層出不窮的情況下仍能有效打擊犯罪，光是依賴過去的偵查方法及作為應是不足的。

故若從這樣的角度出發，蒐集犯罪相關社群並進行其語料分析的模式，可修改如下：

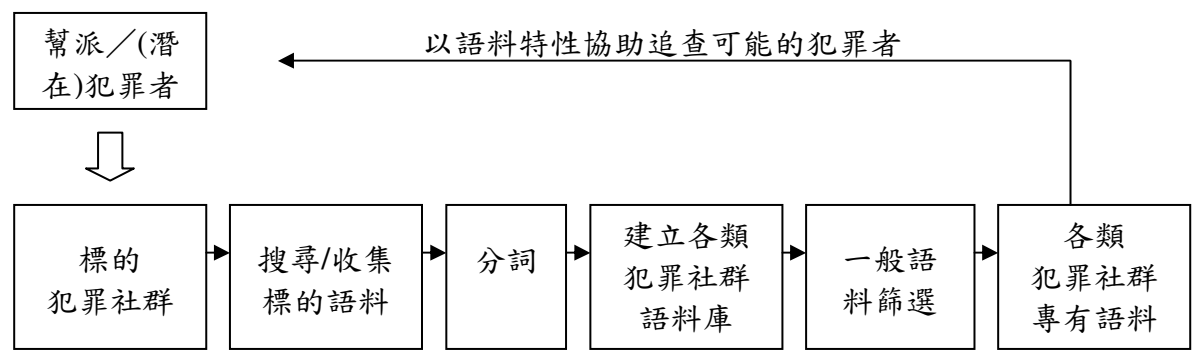


圖 4：犯罪社群語料分析流程及語料特性應用
資料來源：本研究整理

在此模式下，篩選出的社群專有語詞，為此社群有別於其他社群的特徵，可以用來作為社群間的明顯區別，更甚者，可以成為社群成員在既有屬性(如年齡、性別、教育程度、

族群等)外另增一項可用來識別成員身份的屬性。基此，語言應用於實務上的功用及其重要性便因應而生。將語言視為犯罪預防及偵查的要項，不但是賦予語言在應用上的新義，更是偵防觀念上的一種進步。當然，要將語言的功用落實到實務上，前述語料整理和分析步驟是不可或缺的。因應變遷快速的社會環境，犯罪型態趨於複雜與多變，警務運用的利器當是越多越廣越好，而就人的行為而論，最多最廣的行為不就是語言？

因此結合語言學學理去協助警察實務的需求，確實是今日警學研究者該面對的嚴肅課題。筆者此處所提出的犯罪社群語料的研究也正是面對此課題所提出的核心構思。這個核心構思以語言學為基礎，資訊管理學為手段，犯罪學為目的，是標準的科際整合的學識。警學是複雜的，研究的範疇幾乎涵蓋了所有的社會學問，極需要不同領域的知識融入整合，以契合社會治安的需求，若以此觀點來看，社群語料分析之警務運用將是值得期待的。

五、結論

本研究將「語料字詞頻統計法」應用為以社群為範疇的「社群語料字詞頻統計法」，並從「時間」(即各時間點出現的頻次)、「空間」(即總頻次)的角度作觀察，能夠對語料的認識、解釋與應用更充份、更豐富，就結果來說，社群語料分析能夠透過對語料的彙整與分析，增加對某一社群的認識與瞭解，或從語詞中推測出可能的所屬社群。

在警學應用上，「犯罪社群語料分析」透過結合語言學、資訊管理學與犯罪學作科際整合，提供警學發展的參考。在前述例證中所提之「犯罪社群語料分析」僅只是對分析模式的初探，未來仍有許多方面可以繼續發展的，包括：

- 1.由於在語料蒐集與分析上，仍花費了大量的時間在語料的初步整理，並將語料轉換成事先定義好的格式，之後再以語詞擷取演算法(如 Pat-tree)進行分詞，在後續的發展上，應可朝向簡化語料整理的過程，即不須事先定義語詞格式而直接進行「分詞」的步驟，如此將能大大地節省語料初步整理的時間。
- 2.若以語料分析的方法來看，字詞頻統計是其中一種常用的方法，尚可採語意或語法分析，或從不同的年齡、性別、教育程度、職業或語言群等屬性進行語言層的分析，使各個社群或語言群有更明顯的區隔，進而實現建立社群語料庫的可能。
- 3.無庸置疑地，警學的研究必須超越社會或犯罪現狀作前瞻思考，在資訊社會儼然形成並趨於成熟的今日，發生在現實社會中的犯罪已逐漸在資訊網絡中顯露並作了難以預測的變異，而資訊網絡中人與人交流的媒介多數的成份仍是透過「語言」，因此，從語料分析來看與犯罪現象或犯罪事件間的關聯應也是必須且迫切發展的重點。

參考文獻

- G.H. Gonnet, R.A. Baeza-Yates, T. Snider, “New indices for text: PAT Trees and PAT arrays,” *Information retrieval: data structures and algorithms*, Prentice-Hall, pp. 66-82, 1992.
- U. Manber and R. Baeza-Yates, “An algorithm for string matching with a sequence of don't cares,” *Information Processing Letters*, Vol. 37, No. 3, pp. 133-136, 1991.
- L.F. Chien, “PAT-tree-based keyword extraction for Chinese information retrieval,” *Proceedings of SIGIR’ 97*, PP. 55-58, 1997.
- K. Kageura and B. Umino, “Methods of Automatic Term Recognition-A Review,” *Terminology*, Vol. 3, No. 2, pp. 259-289, 1996.
- 董同龢，語言學大綱，台北：東華書局，1989。
- 葉蜚聲與徐通鏞，語言學綱要，台北：書林，1993。
- 謝國平，語言學概論，台北：三民書局，2004。
- 曾榮汾，刑事案件資料庫系統之建立與綜合應用 - 犯罪語料庫建立之研究，行政院國科會專題研究計畫(NSC89-2420-H-015-001-QA)，1989。
- 林頌堅，「基於術語抽取與術語叢集技術的主題抽取」，中文計算語言學刊，第9卷第1期，頁97-111，2004。

附錄 1：停用詞列表

a	between	he'll	me	perhaps	therefore	whence
about	beyond	he's	meantime	rather	therein	whenever
above	both	hence	meanwhile	recent	thereupon	where
according	but	her	might	recently	these	where's
across	by	here	miss	same	they	whereafter
actually	can	here's	more	seem	they'd	whereas
after	can't	hereafter	moreover	seemed	they'll	whereby
afterwards	cannot	hereby	most	seeming	they're	wherein
again	could	herein	mostly	seems	they've	whereupon
against	couldn't	hereupon	much	several	this	wherever
all	did	hers	must	she	those	whether
almost	didn't	herself	my	she'd	though	which
alone	do	him	myself	she'll	through	while
along	does	himself	namely	she's	throughout	whither
already	doesn't	his	neither	should	thru	who
also	don't	how	never	shouldn't	thus	who'd
although	down	however	nevertheless	since	to	who'll
always	during	i'd	next	so	together	who's
among	each	i'll	nobody	some	too	whoever
amongst	eg	i'm	none	somehow	toward	whole
an	either	i've	nonetheless	someone	towards	whom
and	else	if	noone	something	under	whomever
another	elsewhere	in	nor	sometime	unless	whose
any	end	indeed	not	sometimes	unlike	why
anyhow	ending	instead	nothing	somewhere	unlikely	will
anyone	enough	into	now	still	until	with
anything	etc	is	nowhere	stop	up	within
anywhere	even	isn't	of	such	upon	won't
are	ever	it	off	taking	us	would
aren't	every	it's	often	than	used	wouldn't
around	everyone	its	on	that	using	without
as	everything	itself	once	that'll	very	yet
at	everywhere	last	one	that's	via	you
be	except	later	one's	that've	was	you'd
became	few	latter	only	the	wasn't	you'll
because	for	latterly	onto	their	we	you're

become	former	less	or	them	we'd	you've
becomes	formerly	let	other	themselves	we'll	your
becoming	found	let's	others	then	we're	yours
been	from	like	otherwise	thence	we've	yourself
before	further	likely	our	there	well	yourselves
beforehand	had	ltd	ours	there	were	的
begin	has	made	ourselves	there'd	weren't	個
beginning	hasn't	make	out	there'll	what	們
behind	have	makes	over	there're	what'll	之
being	haven't	many	overall	there've	what's	與
below	he	may	own	thereafter	whatever	本文
beside	he'd	maybe	per	thereby	when	可以
besides						

資料來源：林頌堅，2004

附錄 2：資訊管理領域常用語詞

序號	詞組	詞組 頻次	文章 頻次
1	INFORMATION SYSTEMS	350	243
2	INFORMATION TECHNOLOGIES	193	127
3	CASE STUDY	66	48
4	DECISION MAKING	68	40
5	DECISION SUPPORT SYSTEMS	46	37
6	SYSTEMS DEVELOPMENT	33	30
7	MANAGEMENT INFORMATION SYSTEMS	32	29
8	DECISION MAKERS	45	28
9	DATA PROCESSING	33	25
10	INFORMATION MANAGEMENT	29	24
11	DATA COLLECTION	29	24
12	COMPETITIVE ADVANTAGE	29	21
13	DEVELOPMENT PROCESSES	28	21
14	EXPERT SYSTEMS	43	19
15	SOFTWARE DEVELOPMENT	26	19
16	SYSTEMS ANALYSTS	27	18
17	TOP MANAGEMENT	24	16
18	ORGANIZATIONAL CONTEXTS	20	16
19	USER INVOLVEMENT	31	15
20	USER SATISFACTION	30	15
21	DECISION SUPPORT	20	15
22	MIS RESEARCH	19	15
23	EASE OF USE	29	14
24	NEW TECHNOLOGIES	17	14
25	PLANNING PROCESSES	16	14
26	END-USER COMPUTING	22	13
27	CRITICAL SUCCESS FACTORS	20	13
28	INFORMATION SYSTEMS MANAGERS	17	13
29	COMPUTER SYSTEMS	15	13
30	USE OF INFORMATION TECHNOLOGY	14	13
31	PERCEIVED USEFULNESS	25	12
32	PROJECT TEAMS	17	12
33	SYSTEM USAGE	16	12
34	INFORMATION SYSTEMS RESEARCH	12	12
35	EMPIRICAL EVIDENCE	12	12
36	KNOWLEDGE MANAGEMENT	19	11
37	ROLE OF INFORMATION	13	11

38	COMPUTER-BASED INFORMATION SYSTEMS	13	11
39	RESEARCH IN INFORMATION SYSTEMS	12	11
40	IMPLEMENTATION PROCESS	12	11
41	LABORATORY EXPERIMENT	11	11
42	FACTORS AFFECTING	11	11
43	FIRM PERFORMANCE	21	10
44	CHARGEBACK SYSTEMS	17	10
45	INFORMATION SERVICE	16	10
46	SYSTEM DEVELOPMENT	14	10
47	ORGANIZATIONAL CHANGE	12	10
48	STRATEGIC PLANNING	11	10
49	GROUP DECISION SUPPORT SYSTEMS	10	10
50	IT INFRASTRUCTURE	16	9
51	BUSINESS UNITS	14	9
52	ORGANIZATIONAL ENVIRONMENTS	12	9
53	INFORMATION NEEDS	12	9
54	DECISION-MAKING PROCESSES	12	9
55	INFORMATION SYSTEMS DEVELOPMENT	11	9
56	INFORMATION RESOURCE	11	9
57	SENIOR EXECUTIVES	10	9
58	GROUP PROCESS	10	9
59	EMPIRICAL STUDIES	10	9
60	DEVELOPMENT LIFE CYCLE	10	9
61	BUSINESS PROCESSES	10	9
62	ORGANIZATIONAL BEHAVIOR	9	9
63	DEVELOPMENT METHODS	9	9
64	SERVICE QUALITY	31	8
65	MIS MANAGERS	16	8
66	ELECTRONIC MARKETS	16	8
67	INFORMATION REQUIREMENTS	15	8
68	IMPLEMENTATION SUCCESS	15	8
69	PERCEIVED EASE OF USE	13	8
70	DESIGN PROCESS	12	8
71	SUCCESSFUL IMPLEMENTATION	11	8
72	MIS RESEARCHERS	11	8
73	GROUP SUPPORT	11	8
74	INFORMATION SYSTEMS DEPARTMENTS	10	8
75	INFORMATION PROCESSING	9	8
76	CHIEF INFORMATION OFFICERS	9	8
77	LARGE ORGANIZATIONS	8	8
78	INFORMATION CENTER	17	7

語料統計分析法之例證及其警務應用

79	MANAGEMENT SUPPORT	12	7
80	LEARNING PROCESSES	11	7
81	KNOWLEDGE CREATION	11	7
82	STRATEGIC INFORMATION SYSTEMS	10	7
83	SENIOR MANAGERS	9	7
84	PROBLEM SOLVING	9	7
85	DECISION QUALITY	9	7
86	MIS LITERATURE	8	7
87	KNOWLEDGE-BASED SYSTEMS	8	7
88	ELECTRONIC DATA INTERCHANGE	8	7
89	THEORETICAL MODEL	7	7
90	SYSTEM DESIGN	7	7
91	ACTION RESEARCH	16	6
92	VIRTUAL TEAMS	11	6
93	PATH ANALYSIS	11	6
94	KNOWLEDGE WORKERS	10	6
95	PROJECT MANAGEMENT	7	6
96	MANAGEMENT ISSUES	7	6
97	INFORMATION FUNCTION	7	6
98	GROUP SUPPORT SYSTEMS	7	6
99	EXECUTIVE INFORMATION SYSTEMS	7	6
100	DATA ANALYSIS	7	6
101	BUSINESS VALUE	7	6
102	TYPES OF INFORMATION	6	6
103	THEORETICAL AND PRACTICAL	6	6
104	STRATEGIC BUSINESS	6	6
105	PERFORMANCE MEASURES	6	6
106	ORGANIZATIONAL KNOWLEDGE	6	6
107	NEW SYSTEM	6	6
108	LARGE FIRMS	6	6
109	DEVELOPMENT APPROACH	6	6
110	DATABASE MANAGEMENT SYSTEMS	6	6
111	COMPUTER TECHNOLOGY	6	6
112	PERSONAL COMPUTING	20	5
113	USER PARTICIPATION	18	5
114	PUBLIC SECTOR	14	5
115	SMALL FIRMS	10	5
116	COMPUTER-BASED INFORMATION	10	5
117	USER ACCEPTANCE	9	5
118	JOB CHARACTERISTICS	9	5
119	INVESTMENT DECISIONS	8	5

120	HUMAN RESOURCES	8	5
121	TEAM MEMBERS	7	5
122	SURVEY INSTRUMENT	7	5
123	PROCESS MODEL	7	5
124	ORGANIZATIONAL STRUCTURE	7	5
125	INFORMATION PRESENTATION	7	5
126	TECHNOLOGY PROFESSIONALS	6	5
127	SYSTEM IMPLEMENTATION	6	5
128	RESEARCH QUESTIONS	6	5
129	REQUIREMENTS ANALYSIS	6	5
130	MODEL PROVIDES	6	5
131	MIS IMPLEMENTATION	6	5
132	MANAGEMENT PRACTICES	6	5
133	INFORMATION TECHNOLOGY INVESTMENTS	6	5
134	GENERAL MANAGERS	6	5
135	EMPIRICAL ANALYSIS	6	5
136	DISCRIMINANT VALIDITY	6	5
137	CLASSIFICATION SCHEME	6	5
138	CHIEF EXECUTIVE	6	5
139	VALUE OF INFORMATION	5	5
140	USE OF INFORMATION	5	5
141	UNIVERSITY OF MINNESOTA	5	5
142	UNDERSTANDING OF INFORMATION	5	5
143	TECHNICAL ISSUES	5	5
144	SYSTEMS DESIGN	5	5
145	RESEARCH APPROACH	5	5
146	PLANNING AND CONTROL	5	5
147	ORGANIZATIONAL ISSUES	5	5
148	ORGANIZATIONAL CAPABILITIES	5	5
149	NO SIGNIFICANT DIFFERENCES	5	5
150	NO SIGNIFICANT	5	5
151	MANAGERIAL IMPLICATIONS	5	5
152	MANAGEMENT INFORMATION	5	5
153	INFORMATION SYSTEMS RESEARCHERS	5	5
154	INFORMATION SYSTEMS MANAGEMENT	5	5
155	INFORMATION SYSTEMS EXECUTIVES	5	5
156	INFORMATION SYSTEMS DESIGN	5	5
157	GROUP DECISION SUPPORT	5	5
158	FACTOR ANALYSIS	5	5
159	EMPIRICAL SUPPORT	5	5
160	ELECTRONIC COMMERCE	5	5

語料統計分析法之例證及其警務應用

161	DEVELOPMENT AND MAINTENANCE	5	5
162	DEVELOPMENT ACTIVITIES	5	5
163	DESIGN AND IMPLEMENTATION	5	5
164	CONVERGENT VALIDITY	5	5
165	CONSTRUCT VALIDITY	5	5
166	COMPUTER-BASED SYSTEMS	5	5
167	COMPUTER AND COMMUNICATIONS	5	5
168	COMPLEX ORGANIZATIONS	5	5
169	BUSINESS FUNCTIONS	5	5
170	APPLICATION DEVELOPMENT	5	5
171	IT USAGE	14	4
172	JOB SATISFACTION	13	4
173	MIS DESIGN	10	4
174	CUSTOMER SERVICE	10	4
175	COMPUTER SELF-EFFICACY	10	4
176	COGNITIVE STYLES	10	4
177	MANAGEMENT PROCESSES	9	4
178	END-USER COMPUTING SATISFACTION	9	4
179	ELECTRONIC MAIL	9	4
180	COGNITIVE FIT	9	4
181	TASK COMPLEXITY	8	4
182	PROJECT MANAGERS	8	4
183	DSS DEVELOPMENT	8	4
184	COMPETITIVE STRATEGY	8	4
185	WORK ENVIRONMENT	7	4
186	USER ATTITUDES	7	4
187	ORGANIZATIONAL LEARNING	7	4
188	OFFICE AUTOMATION	7	4
189	KNOWLEDGE SHARING	7	4
190	GROUP DECISION MAKING	7	4
191	FIRM SIZE	7	4
192	DATA WAREHOUSING	7	4
193	WORK SYSTEMS	6	4
194	ORGANIZATIONAL CHARACTERISTICS	6	4
195	LIFE INSURANCE	6	4
196	INFORMATION SYSTEM DEVELOPMENT	6	4
197	FINANCIAL PERFORMANCE	6	4
198	DESIGN PRINCIPLES	6	4
199	DATA QUALITY	6	4
200	COMPUTER SUPPORT	6	4
201	USER PERCEPTIONS	5	4

202	USER INTERFACE	5	4
203	TECHNOLOGY MANAGEMENT	5	4
204	SYSTEM PERFORMANCE	5	4
205	SYSTEM DESIGNERS	5	4
206	SUPPLY CHAINS	5	4
207	SUCCESS OR FAILURE	5	4
208	RESEARCH METHODS	5	4
209	MIS DEVELOPMENT	5	4
210	INFORMATION TECHNOLOGY TO SUPPORT	5	4
211	INFORMATION SHARING	5	4
212	INFORMATION AND DECISION SUPPORT	5	4
213	HIGH QUALITY	5	4
214	GROUP MEMBERS	5	4
215	ELECTRONIC MEETING SYSTEMS	5	4
216	ECONOMIC THEORY	5	4
217	BUSINESS OBJECTIVES	5	4
218	BUSINESS INFORMATION	5	4
219	ANALYZING DATA	5	4
220	SYSTEM EFFECTIVENESS	11	3
221	SUPPLY CHAIN	11	3
222	LOGIC PROGRAMMING	11	3
223	MIDDLE MANAGERS	10	3
224	STRATEGIC MANAGEMENT	9	3
225	SHARED KNOWLEDGE	9	3
226	JOB DESIGN	9	3
227	DATABASE DESIGN	9	3
228	SYSTEM SUCCESS	8	3
229	PRIVATE SECTOR	8	3
230	PERSONAL INFORMATION	8	3
231	FINANCIAL MODELING	8	3
232	PERFORMANCE GAINS	7	3
233	LIFE CYCLE	7	3
234	BELIEFS AND ATTITUDES	7	3
235	USER MANAGERS	6	3
236	SYSTEM COMPLEXITY	6	3
237	SOCIAL PRESENCE	6	3
238	MEDIA SELECTION	6	3
239	INTERNAL AUDIT	6	3
240	DATA MANAGEMENT	6	3
241	CONTEXTUAL FACTORS	6	3
242	SUBJECTIVE NORM	5	3

語料統計分析法之例證及其警務應用

243	LEARNING STYLES	5	3
244	JOB PERFORMANCE	5	3
245	INTANGIBLE BENEFITS	5	3
246	HUMAN RESOURCE MANAGEMENT	5	3
247	CORPORATE STRATEGY	5	3
248	COMMUNICATION MEDIUM	5	3
249	BUSINESS FIRMS	5	3
250	WORD PROCESSING	14	2
251	KNOWLEDGE TRANSFER	11	2
252	DECISION MODELS	9	2
253	NETWORK EFFECTS	8	2
254	MIS EMPLOYEES	7	2
255	COLLABORATIVE LEARNING	7	2
256	SEARCH COSTS	6	2
257	IDEA GENERATION	6	2
258	ELECTRONIC BRAINSTORMING	6	2
259	DATA DICTIONARY SYSTEMS	6	2
260	DATA ADMINISTRATION	6	2
261	CASE TOOLS	6	2
262	ANALYSTS AND USERS	6	2
263	SYSTEMS DESIGNERS	5	2
264	SOFTWARE REUSABILITY	5	2
265	PROCESS PERFORMANCE	5	2
266	PROCESS IMPROVEMENT	5	2
267	PERCEIVED PERFORMANCE	5	2
268	INDIVIDUAL PERFORMANCE	5	2
269	ENGINEERING SYSTEMS	5	2
270	DSS SUCCESS	5	2
271	DESIGN PHASE	5	2
272	ATTITUDES AND PERCEPTIONS	5	2
273	APPLICATIONS PORTFOLIO	5	2

