

國立政治大學資訊科學系
Department of Computer Science
National Chengchi University

碩士學位論文
Master's Thesis

由華語流行歌詞探勘歌詞的特徵樣式
Mining Patterns from Lyrics of
Chinese Popular Music



研究生：周晏如
指導教授：沈錕坤 博士

中華民國一〇五年七月
July, 2016

由華語流行歌詞探勘歌詞的特徵樣式

Mining Patterns from Lyrics of
Chinese Popular Music

研究生：周晏如 Student: Yen-Ju Chou

指導教授：沈錕坤 Advisor: Man-Kwan Shan

國立政治大學

資訊科學系

碩士論文

A Thesis

submitted to Department of Computer Science

National Chengchi University

in partial fulfillment of the Requirements

for the degree of

Master

in

Computer Science

中華民國一〇五年七月

July, 2016

致謝

與政治大學的關係，似乎緣起不減，十多年前就讀韓國語文學系，十多年後再次返回政大就讀資訊科學研究所。然而，跨領域的學習，絕非易事，但就像跑馬拉松一樣，知道自己目標為何，並一步一腳印地完成每個過程，終究可達到目標。

特別感謝我的指導教授沈錕坤老師，常常一對一地指導我，花費了許多時間，且教學功力非凡，常常能以深入淺出地方式，讓我快速理解許多理論與實作的方法。在此，對老師致上最誠摯的謝意，並祝福老師事事順心。

此外，也特別感謝我的先生，一路支持我、鼓勵我、幫助我、體諒我，讓我能無後顧之憂地學習。

最後，非常感謝實驗室的同學和學弟妹們及公司同事，包括佩雯、韋光、凡耘、逸群、鼎崴、冠嶽及 Yoco，一路走來，有你們的幫助與陪伴，真好！



由華語流行歌詞探勘歌詞的特徵樣式

摘要

華語流行歌詞一直是語言、文學、音樂或是文化研究等相關科系赤手可熱的研究題目，內容包含作詞者、修辭分析、風格、用韻及語言表達等，然由於歌詞數量龐大，難以全部以人工分析。

近年來，資訊科技日新月異不斷地進步，隨著 Big Data 議題受到注目，Data Mining 在近年來相當熱門，然而針對華語流行歌詞的巨量資料探勘與分析研究並不多。因此，本論文研究以程式來自動化分析歌詞的樣式與特性，包括詞彙頻率、詞彙相鄰關係分析、歌名分析、使用語系分析、舊曲新唱、歌詞風格自動分類、用韻及修辭等，而研究資料係透過網路擷取知名網站內容，包含魔鏡歌詞網 (Mojim.com)、臺北之音 HitFM 聯播網 (www.hitoradio.com) 及教育部重編國語辭典，透過分析規則及以 Non-Trivial Repeating Pattern 等方法，來完成分析及系統實作。

透過華語流行歌詞的大量分析，探勘及了解各種歌詞的風格與特性，將可了解各種歌詞、作詞者的風格與特色，進而應用在歌詞資料的管理與查詢。

此外，本研究將八萬多首歌詞的各種分析資料設置成網站，提供予學術研究使用，希冀此研究資料能使華語流行歌詞相關研究研究，進行更深入地探討。

關鍵字: 巨量資料、資料探勘、重複樣式探勘、流行音樂、歌詞

Mining Patterns from Lyrics of Chinese Popular Music

Abstract

Chinese popular music lyrics has been a popular topic for researchers who major in languages and literature, music or culture. Related studies include of lyricists, rhetoric methods, styles, rhyme and language expression. However, all these studies were performed by manual analysis. It is difficult to analyze large amount of lyrics manually.

With advances in computer technology, big data and data mining techniques have been widely used in different kinds of data. However, to the best of our knowledge, none have been done on pattern mining from big data of lyrics of Chinese popular music. Therefore, the objective of this thesis is to discover patterns from tremendous lyrics data based on data mining techniques. We use data downloaded from www.mojim.com, <http://dict.revised.moe.edu.tw/cbdic/> and <http://www.hitoradio.com> (Hit FM). Data mining methods are employed to find lyrics' patterns and features, including frequent words, word adjacency, analysis of hit songs' names, lyrics' language studies, cover song research, automatic style prediction, rhyme and rhetoric patterns.

With the analysis of tremendous lyrics and data, the developed approaches of this thesis will be helpful for discovering distinguishing styles of lyrics and lyricists.

Key Words: Big Data, Data Mining, Repeating Pattern Mining, Popular Music, Lyrics

目錄

第一章	緒論	1
1.1	研究背景與動機	1
1.2	研究內容及方法	2
1.3	論文貢獻	2
1.4	論文架構	3
第二章	相關研究	4
2.1	流行音樂歌詞在各時期的特色之相關學術研究	4
2.2	流行音樂歌詞押韻相關學術研究	6
2.3	修辭學相關學術研究	6
第三章	研究方法	10
3.1	資料來源	10
3.1.1	魔鏡歌詞網 (Mojim.com)	10
3.1.2	臺北之音 HitFM 聯播網 (www.hitoradio.com)	11
3.1.3	教育部重編國語辭語修訂本 (Dict.revised.moe.edu.tw)	12
3.2	資料前處理	13
3.2.1	發音為非國語的歌詞處理	13
3.2.2	斷詞處理	14
3.2.3	重複歌詞判斷	14
3.3	歌名字數分析	16
3.4	使用語系分析	17
3.5	舊曲新唱分析	17
3.6	詞彙分析	18
3.6.1	詞彙頻率分析	18
3.6.2	詞彙相鄰關係分析 (Word Adjacency)	18
3.7	押韻分析	19
3.7.1	押韻類別定義：一韻到底、換韻、無韻、句首韻	19
3.7.2	利用 Non-Trivial Repeating Pattern 演算法判斷押韻	21
3.8	修辭分析	23
3.8.1	明喻、類字、疊字、疊句及對偶之定義	23
3.8.2	利用 Non-Trivial Repeating Pattern 演算法判斷類字、疊字及疊句	27
3.9	歌詞風格自動分類	28
第四章	實驗	30
4.1	網站實作	30
4.1.1	架構	30
4.1.2	工具	31
4.1.3	資料庫	31
4.1.4	資料來源與資料的統計分析	33
4.2	詞彙分析結果	33

4.2.1 歷年歌詞之高頻辭	33
4.2.2 1960~2015 逐年的高頻詞詞彙	38
4.2.3 特別詞彙之頻率分析	41
4.2.4 詞彙相鄰關係分析結果	42
4.4 使用語系分析結果	45
4.5 舊曲新唱分析結果	46
4.6 風格分類結果與分析	49
4.7 押韻分析結果	51
4.8 修辭分析結果：明喻、類字、疊字、疊句及對偶	53
4.8.1 明喻	54
4.8.2 類字	55
4.8.3 疊字	56
4.8.4 疊句	57
4.8.5 對偶	58
4.9 作詞者的詞頻與詞彙相鄰分析：方文山、李宗盛、林夕	60
4.9.1 詞頻	60
4.9.2 詞彙相鄰 (Word Adjacency)	67
4.10 年代對使用語系的交叉分析	69
4.11 年代對作詞者作品數量的交叉分析：方文山、李宗盛及林夕	71
4.12 年代對押韻的交叉分析	73
4.13 年代對風格的交叉分析	73
4.14 風格對歌名字數的交叉分析	80
4.15 風格對使用語系的交叉分析	81
4.16 風格對押韻的交叉分析	81
4.17 風格對修辭的交叉分析	83
第五章 結論與未來研究方向	84
5.1 結論	84
5.2 未來研究方向	85
參考文獻	86

圖目錄

圖 3-1 魔鏡歌詞網網站示意圖	10
圖 3-2 魔鏡歌詞網歌詞查詢結果示意圖	11
圖 3-3 臺北之音 HitFM 聯播網網站示意圖	12
圖 3-4 教育部重編國語辭語修訂本網站示意圖	13
圖 3-5 詞彙相鄰關係示意圖	19
圖 3-6 「如」的斷詞及用法比較	24
圖 3-7 「若」的斷詞及用法比較	25
圖 3-8 SVM 分割示意圖	29
圖 4-1 全部歌詞的詞彙的文字雲	34
圖 4-2 名詞詞彙的文字雲	35
圖 4-3 動詞詞彙的文字雲	36
圖 4-4 形容詞詞彙的文字雲	37
圖 4-5 副詞詞彙的文字雲	37
圖 4-6 1960~1969 年華語流行歌詞的文字雲	38
圖 4-7 1970~1979 年華語流行歌詞的文字雲	38
圖 4-8 1980~1989 年華語流行歌詞的文字雲	39
圖 4-9 1990~1999 年華語流行歌詞的文字雲	39
圖 4-10 2000~2009 年華語流行歌詞的文字雲	40
圖 4-11 2010~2015 年華語流行歌詞的文字雲	40
圖 4-12 相鄰詞彙(Word Adjacency) Top 10 使用的詞彙	43
圖 4-13 相鄰詞彙(Word Adjacency) 超過 5,000 次的相鄰詞彙	43
圖 4-14 魔鏡歌詞網歷年歌詞與 Hit FM 年度百首單曲歌名字數比例比較	44
圖 4-15 使用雙語歌詞的歌詞比例比較	45
圖 4-16 歌詞中出現英文的高頻詞文字雲	46
圖 4-17 被舊曲新唱 16 次以上的歌詞一覽	46
圖 4-18 訓練資料風格類型數量比較	50
圖 4-19 一韻到底的韻腳一覽 (僅列出現 100 次以上的韻腳)	51
圖 4-20 換韻的韻腳一覽 (僅顯示超過 5,000 次以上)	52
圖 4-21 句首韻的韻腳一覽 (僅列出超過 100 次以上的韻腳)	53
圖 4-22 類字高頻詞文字雲	55
圖 4-23 疊字高頻詞文字雲	57
圖 4-24 疊句高頻詞文字雲	58
圖 4-25 作者使用對偶修詞的歌詞數量佔其整體作品數量之比例	60
圖 4-26 方文山全部用詞高頻詞的文字雲	61
圖 4-27 方文山的名詞高頻詞的文字雲	62
圖 4-28 方文山的動詞高頻詞的文字雲	62
圖 4-29 方文山的形容詞高頻詞的文字雲	63
圖 4-30 李宗盛全部用詞高頻詞的文字雲	63

圖 4-31 李宗盛名詞高頻詞的文字雲	64
圖 4-32 李宗盛動詞高頻詞的文字雲	64
圖 4-33 李宗盛形容詞高頻詞的文字雲	65
圖 4-34 林夕全部用詞高頻詞的文字雲	65
圖 4-35 林夕的名詞高頻詞的文字雲	66
圖 4-36 林夕的動詞高頻詞的文字雲	66
圖 4-37 林夕的形容詞高頻詞的文字雲	67
圖 4-38 李宗盛的詞彙相鄰文字雲	68
圖 4-39 林夕的詞彙相鄰文字雲	68
圖 4-40 方文山的詞彙相鄰文字雲	69
圖 4-41 1965~2015 歷年使用英文歌詞的比例趨勢圖	69
圖 4-42 2008~2015 歷年使用韓文歌詞的比例趨勢圖	70
圖 4-43 方文山歷年作品數量	71
圖 4-44 李宗盛歷年作品數量	72
圖 4-45 林夕歷年作品數量	72
圖 4-46 愛情風格歷年作品數量	75
圖 4-47 愛情風格佔該年歌詞數量的比例	75
圖 4-48 勵志風格歷年作品數量	76
圖 4-49 勵志風格佔該年歌詞數量的比例	76
圖 4-50 友情風格歷年作品數量	77
圖 4-51 友情風格佔該年歌詞數量的比例	77
圖 4-52 親情風格歷年作品數量	78
圖 4-53 親情風格佔該年歌詞數量的比例	78
圖 4-54 其他風格歷年作品數量	79
圖 4-55 其他風格佔該年歌詞數量的比例	79
圖 4-56 各類風格類型使用各字部佔該風格之比例圖	80

表目錄

表 2-1 華語流行音樂各時期之歷史背景、特色、代表人物及代表作(1950~1980).....	5
表 2-2 華語流行音樂各時期之歷史背景、特色、代表人物及代表作(1980~2010).....	5
表 3-1 字串‘ai-ao-ao-ai-ai-ao-ao-ai-o-ai-er-ai’的關係矩陣.....	22
表 3-2 字串‘ai-ao-ao-ai-ai-ao-ao-ai-o-ai-er-ai’的所有 Non-trivial Repeating Pattern 相關資料.....	22
表 3-3 以 Non-Trivial Repeating Pattern 判斷押韻方式.....	23
表 3-4 類疊的定義及範例.....	26
表 4-1 網站內容一覽表.....	30
表 4-2 系統實作工具一覽.....	31
表 4-3 資料庫的格式.....	32
表 4-4 實驗歌詞筆數的分析.....	33
表 4-5 特別詞彙之頻率分析 I.....	41
表 4-6 特別詞彙之頻率分析 II.....	42
表 4-7 被翻唱的歌詞 超過 50 首的作詞者列表.....	48
表 4-8 非愛情的詞總集合方式及準確率.....	50
表 4-9 風格分類結果.....	51
表 4-10 方文山、李宗盛及林夕的高頻詞彙用詞一覽表.....	61
表 4-11 李宗盛、林夕及方文山詞彙相鄰 Top 10 的列表.....	67
表 4-12 1960~2015 各時期使用的 Top 10 韻腳.....	73
表 4-13 1960 年代華語流行歌曲與 1970 初期現代民歌的比較.....	74
表 4-14 各類風格使用各字部的一覽表.....	80
表 4-15 各類風格使用語系一覽表.....	81
表 4-16 各類風格修辭使用比例一覽表.....	83

第一章 緒論

1.1 研究背景與動機

華語流行音樂歌詞的相關研究一直以來為文學系、音樂系、語言學系等相關科系所熱愛探討的主題之一，因為歌詞可以反應時代的流行思潮、生活環境、文學、甚至政治經濟情況，所以華語流行歌詞的研究係文化研究非常重要的一環，且其能為探知該時代人們生活百態的第一手資料。

然而，過去的華語流行音樂歌詞相關研究，多以人工的方式，逐首分析歌詞，內容包括高頻詞、風格、押韻、修辭方式等，除少數的論文研究能人工分析超過 500 首，「例如吳嫩婉的台灣國語流行歌曲的修辭藝術[10]分析了 700 餘首華語流行歌詞」，大部分的研究多侷限於 500 首內，在此情況下，利用程式自動化分析歌詞，就更顯重要。

近年來，隨著硬體設備的快速進步，資料蒐集、資料儲存、資料分析與資料探勘 (Data Mining) 的技術也得以大放異彩，使 Big Data 及 Data Science 成為學術界及產業界熱愛研究及探討之議題。

目前 Big Data 的研究範圍涵蓋商業數據、社交網絡數據及西洋流行音樂歌詞等。其中，西洋流行音樂歌詞的 Big Data 相關研究亦逐年增加，主題包括歌詞的情意或情緒相關分析，例如 Semantic Analysis of Song Lyrics[3]、Music Mood Classification Using Intro and Refrain Parts of Lyrics [5] 及 Lyric Text Mining in Music Mood Classification[2]等，除此之外，也有押韻及風格特徵的相關分析，例如 Rhyme and Style Features for Musical Genre Classification by Song Lyrics[4]，但卻少有針對華語流行歌詞進行巨量資料研究。

至於資料探勘是一種資料分析的方法，在既有的大量資料中，進行更深層地分析，且將這些資料轉成有用、易讀的資訊和知識。因此，本論文期望能透過資料探勘技術，除探勘歌詞本身外，亦探勘其相關資料，包括作詞者、年代、歌名等，並建立交叉查詢系統，以有助於華語流行音樂歌詞未來的研究，使其不再侷限以人工方式費時費力地分析，而是能以藉由系統更快速、更準確的方式來分析資料，期能引起大家對於華語流行歌詞的興趣，且更了解歌詞創作的奧秘及各時代歌詞背後的意義與價值。

1.2 研究內容及方法

網路上流行音樂歌詞多為網站提供，較少具有交叉查詢歌詞的風格、年代、作詞者等功能。本論文以中文系相關論文之研究為基礎，利用程式自動判斷詞彙分析、歌名字數分析、風格、年代、及多種修辭方式等，並建立交叉查詢系統，讓使用者可以方便進行二維、三維等資料查詢，如此一來，可使流行音樂歌詞之相關研究，以此系統為基礎，進行更深入地探討。

本研究將針對下列主題，進行華語流行音樂歌詞資料探勘：

第一部分，以一維方式進行分析：

- (1) 歌名字數分析
- (2) 發表年代分析
- (3) 使用語系分析
- (4) 舊曲新唱分析
- (5) 詞彙分析
- (6) 聲韻分析
- (7) 修辭分析
- (8) 風格分類與分析



第二部分，以二維方式進行，即可交叉查詢分析：

- (1) 作詞者的詞頻與詞彙相鄰分析。
- (2) 年代與各特徵的交叉分析，包括使用語系、作詞者作品數量、押韻與風格等。
- (3) 風格與各特徵的交叉分析，包括歌名字數、使用語系、押韻與修辭等。

1.3 論文貢獻

本論文貢獻在於使用魔鏡歌詞網等大量資料來進行分析歌詞，並提出以程式自動判斷歌詞風格、押韻、修辭、使用語系等多樣項目的方法。而後透過探勘各項目之資料，尋找重要的規則。過去流行音樂歌詞之相關研究多以人工方式進行判斷，本研究則以自

動化程式及資料探勘的方法來分析大量歌詞，並提供交叉查詢系統。

1.4 論文架構

本論文共分為五個章節，首先為研究概論，說明本論文之動機及目的；第二章為相關研究，主要介紹華語流行音樂歌詞的相關研究；第三章為研究之核心，包含資料來源介紹、資料前處理的方法、詞彙關係、歌名字數的分析、使用語系的分析、舊曲新唱的分析、歌詞風格自動分類、押韻分析、修辭分析；第四章則是網站實作及實驗結果；最後在第五章的結論做總結及說明未來研究方向。



第二章 相關研究

2. 流行音樂歌詞相關研究

關於華語流行音樂歌詞的相關研究非常豐富，在這章節中，將針對「流行音樂歌詞在各時期的特色之相關學術研究」、「流行音樂歌詞押韻的研究」及「修辭學相關學術研究」等過去的研究進行了解，以便得知本研究之研究項目的定義。

此外，本研究亦針對其他流行音樂的相關研究進行探討，包括論敘事型音樂錄影帶和歌詞的改編與互文性、華語通俗音樂模組化數位音樂的構成、分析及應用、作詞人性別主體性對作品觀點差異之研究等等。

2.1 流行音樂歌詞在各時期的特色之相關學術研究

2000 年台北教育大學社區與區域發展學系曾慧佳教授，出版臺灣流行歌曲的歌詞研究專書—《從流行歌曲看臺灣社會》[6]，其以社會學者的研究角度，研究了 1940 年至 1990 年臺灣流行歌曲的歌詞對於當代社會價值觀、政治局勢是如何的呈現與反省，並藉由本書對台灣流行歌曲各個階段的分析，不僅能探索音樂大潮流的改變，也得以一窺時代面貌改變的徵兆。

2010 年謝櫻子《方文山華語詞作主題研究》[15]，雖主題定為研究方文山的歌詞，共 265 首，但其亦研究華語流行音樂歌曲各時期的特色，其將 1950~2010 年的華語流行音樂歌曲依年代整理為 6 個時期：

- 華語流行歌曲萌發初期：1950—1960 年代
- 華語流行歌曲的成長時期：1960—1970 年代
- 華語流行歌曲的蓬勃發展期與民歌運動時期：1970—1980 年代
- 華語流行歌曲的工業化時期：1980—1990 年代
- 華語流行歌曲的全球化時期：1990—2000 年代
- 2000 年後台灣流行音樂市場概況：2000—2010 年代

根據謝櫻子的研究，依其敘述的時期之歷史背景、流行歌詞特色、代表人物及代表作，整理如表 2-1 及表 2-2。

表 2-1 華語流行音樂各時期之歷史背景、特色、代表人物及代表作(1950~1980)

(整理自謝櫻子的方文山華語詞作主題研究)

年代	1950—1960	1960—1970	1970—1980
時期	萌發	成長	蓬勃發展 & 民歌運動
歷史背景	1949 大遷徙	1. 戒嚴時期歌詞審查制度 2. 1960 年代中期，美軍介入越戰，駐台美軍增加	1. 文藝電影當道 2. 流行歌曲有「靡靡之音」的非議
特色	1. 上海、香港派（海港派） 2. 反共及恐共的意識型態	1. 男女情愛的歌詞（相對安全，易通過審查） 2. 西洋流行歌曲	1. 電影主題曲風行 2. 反思與覺醒的大潮：民歌運動
代表人物/ 代表作	王琛：家在山那邊	今宵多珍重 我的一顆心不了情 月亮代表我的心 往事只能回味	人物：李建復、陶曉清、王夢麟 作品：浮雲遊子、木棉道、讓我們看雲去

表 2-2 華語流行音樂各時期之歷史背景、特色、代表人物及代表作(1980~2010)

(整理自謝櫻子的方文山華語詞作主題研究)

年代	1980—1990	1990—2000	2000—迄今
時期	工業化	全球化	2000 年後概況
歷史背景	1. 民歌走向商業化 2. 外商進駐台灣	1. 禁歌規定取消 2. 港星來台發展	網路歌曲盛行是新興的傳播方式
特色	比民歌時期嗆辣，更能反映社會各階層的心聲與社會現象	1. 華語、台語、英語多語交雜的曲風 2. 偶像級明星會帶動音樂風尚潮流	歌詞內涵仍以情歌為主流，間或出現其他主題如懷舊、勵志或親情等的歌詞。
代表人物/ 代表作	羅大佑、蘇芮、黃舒駿、鄭智化、陳昇、王笛、馬玉芬	L.A.Boys、小虎隊、紅孩兒、張雨生、張學友	方文山、周杰倫、林俊傑、SHE、五月天

2013 年馬占山的臺灣流行歌詞主題類型與語言表達研究（1999~2008）——以「Hit-FM 年度百首單曲」[16]，研究範圍為 1999 年~2008 年由聽眾上網票選的「Hit-Fm 年度百首單曲」，因此具備流行歌曲得「深受群眾喜愛」的流行元素，其研究的歌詞數量共 546 首。該研究的目的乃透過對「臺灣流行歌曲的歌詞」的爬梳，研究流行歌詞的主題類型與語言表達，進一步探討這些深獲人民喜愛並傳唱的作品，其與時代、社會的關係，及在該時期文學、教育上的特色與價值。

2.2 流行音樂歌詞押韻相關學術研究

2010 年劉祐銘的臺灣國語流行歌曲歌詞用韻研究(1998~2008)[11]，針對華語流行歌詞的用韻，提出如何判定押韻的方法，首先，先提及如何判斷華語流行歌曲歌詞的押韻面臨的大問題，即找不到創作規則，此乃由於歌詞不像古代的詩詞一樣，有著必須嚴格遵守的格律等規則，相反的，它的創作相當自由、隨性。此外，市面上有關談論歌詞寫作的書是相當罕見的，因此，始終缺乏明確的規範來告訴我們如何填寫歌詞，這也使得在判斷華語流行歌曲歌詞押韻時，沒有任何規則、依據可供參考，因而面臨了很大的瓶頸。

然而，透過前人判斷韻文韻腳的方法，包括研究「過去學者對《詩經》韻例、韻腳之研究」、「葉秋鳳在《臺灣因仔歌之詞彙與句式分析》」、「林香薇曾於〈論路寒袖臺語詩《春天个花蕊》的用韻與用字〉」，以此歸納出判斷華語流行歌詞用韻的方法。劉祐銘針對曾入圍第 10 至第 19 屆金曲獎流行音樂類獎項的 500 首華語流行歌曲進行韻腳的分析歸納後，大致可以歸納出下列八種韻例：一韻到底、換韻、無韻、句首韻、節奏停頓處押韻、因為節拍拉長而造成的押韻、交韻及同音重複。

2.3 修辭學相關學術研究

2001 年台灣國立師範大學國文系的蔡宗陽教授，整理了過去在「國語天地」撰寫的專欄文章，出版「應用修辭學」一書[7]，內容除包括 19 個重要修辭方式之理論與應用外，又談了兼格的修辭現象，並將容易混淆的 3 個辭格：譬喻與轉化、譬喻與象徵、對偶與排比，作了相當精細的比較分析。該書透過深入淺出的方式，讓一般人能快速理解

各種修辭方式。

2004 年台灣國立師範大學國文系的黃麗貞教授，出版「實用修辭學」（增訂版）一書[8]，乃研究如何修飾語言與文章之美，書中的每種修辭皆是先立論、後舉例，並針對範例的含義、用法及文章的佳妙處深入闡析說明，依其他修辭學者的分類，加上自己建立的新增辭格，共研究 39 種修辭。

2005 年台灣國立師範大學國文系的黃慶萱教授，出版「修辭學」（增訂三版）一書[9]，將古今七百多位作家的作品及社會生活用語中，選出精闢生動的句子，加以分析比較，歸納出三十種的修辭方法，幾乎佔全書百分之九十以上的篇幅。其認為修辭是一種「方法」，而且是一種調整語文表意之法。

2005 年吳嫩婉《臺灣國語流行歌曲的修辭藝術》[10]，蒐集 1949 年~2000 年臺灣流行音樂歌詞，以此為基礎，再搭配市面上各家出版社出版的歌本、書籍，找出較膾炙人口的國語流行歌曲做為材料，共 755 首。

其從修辭藝術的角度切入，以音樂性、主題修辭藝術及辭格藝術三部分分析歸納，發現國語流行歌曲的歌詞大多能符合平仄相間、音節錯落、押韻與歌詞內容情意相融的修辭藝術，而最常運用的辭格是類疊、譬喻、轉化、設問、對偶與排比等；其中，又以類疊、對偶等辭格出現在流行歌曲中的頻率最高。除修辭外，吳嫩婉將這些歌曲依主題分為 9 類：「愛情、生活哲理、友情、親情、人生百態、田野風光、社會批判、愛國、鄉情」。

除了以上三個面向外，我們亦針對與流行歌詞相關的研究進行探討，分述如下。

2008 年張雯禎的台灣流行歌詞中的隱喻：以愛情為主題(1990-2008)[11]，該研究主要為分析以「愛情」為主題的流行歌詞，分析資料數量共 94 首歌詞。作者提出歌詞中的隱喻是介於詩中隱喻與日常生活語言中隱喻之間的語言，其運用兩種方式來探討如何判斷隱喻：由 Lakoff-Johnson 提出的「概念譬喻理論」及由張榮興、黃惠華提出的「融合理論」，此兩人將融合理論精闢且深入運用在中文語料上。

以下例子為解釋「概念譬喻理論」和「融合理論」：

「這個外科醫生是屠夫」。如果從概念譬喻理論分析，「屠夫」映射至「外科醫生」，「動物」映射至「病人」，「切肉刀」映射到「手術刀」，「屠宰場」映射到

「手術室」等一系列固定對應；而從融合理論來看，除了屠宰與外科手術兩個輸入空間之外，還有一個包含兩個輸入空間共有結構的類屬空間，即為一個人運用銳利工具對另一個人進行手術。經過融合後，融合空間承襲了輸入空間的某些結構，並由於不同心理空間之間的互動而湧現了新結構，如屠宰輸入空間所投射的「手段—目的」關係（屠宰的目的是將動物殺死，並把其骨肉分離），與外科手術的輸入空間所投射的「手段—目的」關係（外科手術的目的則是治癒病人）不相容，也就是在融合空間中融合出以屠宰的手段欲達到治療的目的之間的不相容，因而產生了醫生不能勝任其職務這個原本沒有的新結構。像這類的隱喻就必須依靠融合理論來分析，倘若僅是運用概念譬喻理論則無法顯現此隱喻的關鍵。

最後，張雯禎透過此兩種理論分析的愛情隱喻方式結果，發現可分為 4 種概念來比喻愛情：以食物概念、以旅程概念、以數學、化學、生物中的概念及以動物比喻情人。

2009 年施啟智的華語通俗音樂模組化數位音樂的構成、分析及應用[12]，該研究主要分析的範圍為 2007 年 7 月至 2008 年 6 月 KKBOX 華語流行音樂排行榜前 20 名，原為 240 首，扣除重覆進榜歌曲為 107 首，作為研究內容的歌曲。其根據音樂理論及聆聽音色判斷樂譜和弦進行是否正確，來建置流行音樂的資料庫。此研究目的是以最快的速度，模組化各樂句相關構成要件，再組裝成一首通俗音樂，最後聆聽作品成果，修正旋律與和弦表達感受衝突的地方，再微修樂句與樂句連接的和弦進行，以最大的效率完成一首通俗音樂。

2011 年朱宜秦的流行音樂中作詞人性別主體性對作品觀點差異之研究[14]，主要係以臺灣流行音樂文本為研究對象，採訪 4 位作詞者，包括林尚德、路寒袖、黃婷及陳韋伶，除先了解形成這四位作詞者性別主體性的成長環境外，再以巴特勒(Butler)所提出的性別表演模式(performative model of gender)來檢視四位作詞者的作品，探究性別主體性與作品之間的關聯性，包括唱片工業中的性別形象、性別與作品議題的關連性等，該研究發現早期的作品較符合傳統二元性別表演模式，而資深詞人較近期的作品和新一代詞人的作品，則已漸漸脫離性別表演模式限制的藩籬且有鬆動的跡象，尤其在二元性別系統中，甚至有「女性—剛強；男性—陰柔」的狀態出現，也反映出社會文化中性別概念的改變。

2013 年王奕舜的論敘事型音樂錄影帶和歌詞的改編與互文性-以金曲獎音樂錄影帶獎項為例(2010-2012)[17]，所謂的互文性係指 1967 年由法國後結構主義批評家克莉思蒂娃（Julia Kristeva）所提出的，主要強調任何一個單獨的文本都是不自足的，其意義是在與其他文本交互參照、交互指涉的過程中產生的，因此，任何文本都是一種互文，在一個文本中，不同程度地以各種能夠辨認的形式存在著其他的文本。在極端的意義上，甚至可以說，任何文本都是過去的引文的重新組織。

該研究以 2010 年~2012 年金曲獎音樂錄影帶獎入圍者中，選出 5 個敘事型音樂錄影帶作為研究對象，對音樂錄影帶和歌詞之間的關係進行討論，關注意音樂錄影帶如何針對歌詞內容加以重新建構，並適度地加入非文本脈絡因素，藉此找出音樂錄影帶如何將歌詞以影像與之詮釋，同時又兼顧商業與藝術的結合。

以上敘述的相關研究，有的是研究華語流行歌詞各個時期的背景及風格，有的則係研究判斷華語流行歌詞押韻的方式等，有的是著重於修辭學，但就我們所知，關於華語流行歌詞的大量歌詞及其資料探勘非常稀少。故本論文針對大量華語流行歌詞及相關資料，以程式進行自動化分析，找出其特徵樣式。

第三章 研究方法

3.1 資料來源

本研究利用網站的公開資料，進行資料分析，包括魔鏡歌詞網、臺北之音 HitFM 聯播網、及教育部重編國語辭典修訂本，分述如下：

3.1.1 魔鏡歌詞網 (Mojim.com)

魔鏡歌詞網提供中文、日文、韓文及西洋歌曲歌詞[19]，可利用歌曲名稱、歌手、專輯名稱或歌詞來進行搜尋。此外，魔鏡歌詞網擁有專輯資訊，包括發行日期、歌手、專輯名稱，首頁顯示的資訊是依照最新的發行時間來做排序。

除了被動接收網頁資訊外，使用者也可以依照歌手、專輯、歌名及歌詞，搜尋所需的資訊。其中，最吸引人的部份係用歌詞來做搜尋，倘若使用者在廣播聽到一首歌，但不知道歌名是什麼或者不知演唱者是誰的情況下，然而可能會記得其中一兩句的歌詞，使用者便可用該網站的歌詞搜尋，就能得知歌名、歌手、專輯。圖 3-1 為魔鏡歌詞網網站的示意圖。



圖 3-1 魔鏡歌詞網網站示意圖

(Retrieved June 1, 2016, <https://mojom.com/twznew.htm>)

至於魔鏡歌詞網呈現歌詞的資料如下，包含演唱者、歌名、作詞者、作曲者及編曲等資料，而歌詞部分，並非以句分行，而是加入一空格，而主歌重覆部分，有些歌詞會有，但有些不會。以下為魔鏡歌詞網的例子，歌詞為由梁靜茹演唱、由李正帆作詞的寧夏，如圖 3-2。

專輯介紹	歌詞	專輯列表
新聞	歌手介紹	相關影音

梁靜茹 => 演唱者

寧夏 => 歌名

作詞：李正帆
作曲：李正帆
編曲：五月天

寧靜的夏天 天空中繁星點點
心裡頭有些思念 思念著你的臉
我可以假裝看不見 也可以偷偷地想念
直到讓我摸到你那溫暖的臉

寧靜的夏天 天空中繁星點點
心裡頭有些思念 思念著你的臉
我可以假裝看不見 也可以偷偷地想念
直到讓我摸到你那溫暖的臉

知了也睡了 安心的睡了
在我心裡面寧靜的夏天
知了也睡了 安心的睡了
在我心裡面寧靜的夏天

[更多更詳盡歌詞 在 ※ Mojim.com 魔鏡歌詞網](#)

在我心裡面寧靜的夏天

寧靜的夏天 天空中繁星點點
心裡頭有些思念 思念著你的臉
我可以假裝看不見 也可以偷偷地想念
直到讓我摸到你那溫暖的臉

那是個寧靜的夏天
你來到寧夏的那一天

知了也睡了 安心的睡了
在我心裡面寧靜的夏天
知了也睡了 安心的睡了
在我心裡面寧靜的夏天

評論 (14)

修改

小

大

巨

現在開始我愛你(戀愛的力量 - 下集 新歌 + 精選)

所有歌曲

1. 比較愛
2. 情歌
3. 會呼吸的痛
4. 可惜不是你
5. 親親
6. C'est La Vie
7. 別再為他流淚
8. 崇拜
9. 失憶
10. 絲路
11. 屬於
12. 暖暖
13. 沒有如果
14. 燕尾蝶
15. 我都知道
16. 寧夏 (原唱)
17. 小手拉大手
18. 滿滿的都是愛
19. 今天情人節
20. 蔚藍海岸

同專輯的歌曲列表

梁靜茹

所有專輯

- > 可以的話
- > 在愛裡等你
- > 電影 極光之愛
- > 愛久見人心
- > 現在開始我愛你(戀愛的力量 - 下集 新歌 + 精選)
- > 情歌沒有告訴你
- > 敗犬女王之情歌無雙 影音雙冠
- 原聲天作之盒 (CD+DVD)
- > 靜茹 & 情歌 - 別再為他流淚
- > 幸福的抉擇EP
- > 愛情左燈右行
- > 今天情人節
- > 崇拜
- > 親親愛的大遊行

專輯列表

魔鏡歌詞網的標注
會穿插在歌詞中

副歌重覆的歌詞

圖 3-2 魔鏡歌詞網歌詞查詢結果示意圖

(Retrieved July 7, 2016, <http://mojim.com/twy100463x160x12.htm>)

3.1.2 臺北之音 HitFM 聯播網 (www.hitoradio.com)

臺北之音 HitFM 聯播網[26]，定位以播放華語流行音樂為主，但除華語外尚有相當豐富的東洋、西洋音樂資訊，是一相當全方位的音樂電台。

本研究所截取的資料 1998 年至 2015 年 Hit-Fm 年度百首單曲，該活動是該電台每年固有的傳統活動，由聽眾上網票選出該年度最 Hito 的前一百首歌曲，因此，這些歌曲能代表歌迷喜愛的歌曲類型。圖 3-3 為臺北之音 HitFM 聯播網網站的示意圖。



圖 3-3 臺北之音 HitFM 聯播網網站示意圖

(Retrieved June 1, 2016, http://www.hitoradio.com/newweb/chart_2.php)

3.1.3 教育部重編國語辭語修訂本 (Dict.revised.moe.edu.tw)

本研究於教育部重編國語辭語修訂本所截取的資料[23]，主要係用於建立判斷譬喻法中的「明喻」的喻詞表，利用此網站「相似詞」的功能，可以找出更多辭彙，如此一來，可使喻詞表更加完善。

該網站具「基本檢索」、「進階檢索」、「注音索引」、「筆畫索引」、「部首索引」及「學習筆記」等功能，且查詢結果含有「相似詞」、「相反詞」等內容。圖 3-4 為教育部重編國語辭語修訂本網站示意圖。



圖 3-4 教育部重編國語辭語修訂本網站示意圖
(Retrieved June 2, 2016, <http://dict.revised.moe.edu.tw/>)

3.2 資料前處理

在資料探勘的過程中，為了確保能有高品質的探勘結果，資料前處理是非常重要的步驟及過程。檢視我們由網站取回的資料中，資料前處理包含重複歌詞判斷、發音為非國語的歌詞處理、斷詞處理。

3.2.1 發音為非國語的歌詞處理

由網站取回的資料中，須將歌詞判斷為是否為以國語發音來歌唱，此可避免對於押韻、對偶等修辭學相關的研究判斷錯誤。魔鏡歌詞網的歌詞有標籤為粵語、台語、英語、日語、韓語、客家語、原住民語等，可將這些非以國語為發音的歌詞移除。

3.2.2 斷詞處理

詞是最基本、有意義、且可以自由使用的語言單位。任何文字處理，都必須先能分辨文章中的詞，才能進行進一步的處理，例如程式自動翻譯、語言分析、資訊抽取等等。因此，斷詞的工作成了語言處理不可或缺的技术。

由於本研究許多相關研究，例如高頻詞、詞彙相鄰關係、對偶或比喻等，須先進行斷詞，方能分析，所以在獲得資料後，我們須先將歌詞進行斷詞處理。

斷詞處理可採用由中央研究院的資訊科學研究所和中文詞知識小組研發的「中文斷詞與剖析系統」[21][22]。

3.2.3 重複歌詞判斷

檢視由網站取回的資料中，有許多歌詞被重複翻唱，即作詞者為同一人，但演唱者不同，此會影響最後各項目研究所呈現的比例失真，為避免此現象，我們先進行各首歌詞的比對，判斷是否為同首歌詞。

然而，判斷歌詞是否為同首歌詞，若直接判斷整首歌詞是否與其他首的歌詞相同，會遇到此問題：當主歌或副歌重覆，僅管演唱時，真的有重覆演唱，但歌詞資料卻沒有重覆寫，便會造成不正確的比對結果。

因此，本研究的比對方法分為兩步驟：

1. 找出所有相同的歌名
2. 比較步驟一找出的歌詞，內容是否相同。

即找出相同的歌名後，將歌詞進行斷詞，然後以程式產生為 List 及 Set，List 的內容包含了歌詞內所有的詞，而 Set 則是會移除 List 內容中重覆的詞，例子如下：

List =

['南風','吻','臉','輕','輕飄過來','花香','濃','南風','吻','臉','輕輕星','已','稀月迷朦','我倆','
'緊偎','親親','說','不','完','情意','濃','我倆','緊偎','親親','句句','話','都','由衷','不管','明
天','到','明天','要','相','送','戀','著今','宵','把','今','宵','多','珍重','我倆','臨別依依','怨','
'太陽','快','昇東','我倆','臨別依依','要','再見','在','夢','中','我倆','臨別依依','要','再見','
在','夢','中']

Set =

{'臨別依依','珍重','相','不管','都','在','濃','輕輕星','緊偎','輕飄過來','花香','話','今','我
倆','中','夢','送','太陽','情意','南風','臉','說','由衷','不','吻','昇東','怨','明天','快','要','
宵','到','完','已','多','稀月迷朦','親親','再見','句句','輕','戀','把','著今'}

然而，即使是同樣的兩首歌，歌詞卻有些許的不同，可能是誤植或是使用同義詞，
例如今宵多珍重的歌詞，有以下兩種版本：

版本一：

南風吻臉輕輕 飄過來花香濃

南風吻臉輕輕 星已稀月迷朦

我倆緊偎親親 說不完情意濃

我倆緊偎親親 句句話都由衷

不管明天 到明天要相送

戀著今宵 把今宵多珍重

我倆臨別依依 怨太陽快昇東

我倆臨別依依 要再見在夢中

我倆臨別依依 要再見在夢中

版本二：

南風吻臉輕輕 飄過來花香濃
南風吻臉輕輕 星已稀月迷濛
我們緊偎親親 說不完情意濃
我們緊偎親親 句句話都由衷
不管明天 到明天要相送
戀著今宵 把今宵多珍重
我倆臨別依依 怨太陽快昇東
我倆臨別依依 要再見在夢中

這兩個版本中的「我倆」和「我們」、「迷濛」和「迷濛」，對於歌詞而言，是可互相通用的，但對於程式卻是不同的字。因此，在程式的處理上，必須設定一個門檻值來允許誤差，本研究設定的門檻值為 90% 的詞相同，即為同首歌詞。

3.3 歌名字數分析

歌詞的歌名字數是否會影響受歡迎程度呢？我們可藉由歌名字數分部情況，與臺北之音 HitFM 聯播網年度百首單曲排行的歌詞的歌名字數，進行交叉分析。

分析所有歌詞的歌名字數與臺北之音 HitFM 聯播網年度百首單曲排行的歌詞歌名字數，步驟如下：

(1) 全部都中文，則直接計算字數。

例如："短歌行" = 3 個字

(2) 中文和數字間含有空白或特殊字元，則不計算數字，因為通常為資料誤值年代。

例如："短歌行(1998)" = 3 個字

(3) 含有數字，則一個數字為一個字

例如："分手第 7 天" = 5 個字

(4) 英文字以空白或特殊字元做分隔進行字數計算，如's、_、+

例如: "It's my life" = 4 個字

(5)中英混合的歌名，以空白做分隔進行字數計算

例如: "大家一起 High 翻天" = 7 個字

3.4 使用語系分析

華語流行音樂歌詞常常會出現夾雜外來語的現象，例如李玟的冥想愛，歌詞中的「Try Try 你和我溝通」，就夾雜著英文一起演唱。而根據馬占山，臺灣流行歌詞主題類型與語言表達研究（1999~2008）——以「Hit-FM 年度百首單曲」為對象，「異國語言的使用」在流行歌詞中已極為常見，歌詞中夾雜英文、日文等情形，在 1999 年至 2008 年的「Hit-FM 年度百首單曲」中均可發現，在 546 首華語流行歌曲中即有 129 首具有此項特色。

本研究則以程式語言判斷歌詞中是否有夾雜英文、日文、韓文等語言，步驟為以 Unicode 範圍判斷該句是否為英文、日文、韓文。

Unicode 範圍:

英文: 0041~005a, 0061~007a

日文: 30a0~30ff, 3040~309f

韓文: 1100~11FF, 3130~318F, AC00~D7AF

3.5 舊曲新唱分析

舊曲新唱得宜，可以使得該歌再次大放異彩，此研究可讓我們明白哪些歌詞曾經被翻唱，進而了解流行音樂歌詞市場方向。

判斷舊曲新唱的步驟如下：

- (1) 是否為相同歌名
- (2) 若步驟 1 為真，繼續判斷是否為「相同歌詞」
- (3) 若步驟 2 為真，繼續判斷「作詞者相同但演唱者不同」
- (4) 若步驟 3 為真，則為舊曲新唱

3.6 詞彙分析

本章節探討各詞彙出現頻率，包括各時期常出現的高頻辭、各詞性的高頻辭。此外，在本章節也將分析辭彙間的相鄰關係。

3.6.1 詞彙頻率分析

了解詞彙頻率，可讓我們了解各時期、各風格、各作詞者或各種詞性等的常用字，藉此我們可更深入這些歌詞的創作的奧秘之處。

本研究使用的方法係將由魔鏡歌詞網獲得的歌詞，每一句以中央研究院的「中文斷詞與剖析系統」進行斷詞，接著步驟如下：

- (1) 將各首歌詞依照詞性分開寫入至各檔案。例如：將每首的歌詞的名詞寫入名詞檔案、動詞寫入動詞檔案，以此類推，我們便可獲得歌詞為名詞的文字檔、歌詞為形容詞的文字檔等等。
- (2) 將各詞性的檔案以程式統計出現次數，即可得知各詞彙頻率結果。

至於各年代時期常出現的高頻詞，步驟如下：

- (1) 將歷年來歌詞以年代區分至各檔案。
- (2) 將已斷詞的歌詞以年代為區別寫進各年代的文字檔，例如：1990 年全部歌詞的斷詞寫入同一個檔案，1991 年全部歌詞的斷詞寫入另一個檔案，如此一來，我們便有 1990 年和 1991 年兩個文字檔。
- (3) 以程式統計出現次數，即可得知各年代的詞彙頻率結果。

文字雲部分，則係將統計各詞彙的數量的文字檔，上傳至由簡冠庭所製作的 HTML5 Word Cloud 工具(<http://timdream.org/wordcloud/>)[20]，便可獲得以文字雲方式的檔案。

3.6.2 詞彙相鄰關係分析 (Word Adjacency)

了解詞彙的相鄰關係，可讓我們熟悉哪些詞彙容易一起出現，藉此更加了解詞與詞之間的關係。

本研究詞彙間的相鄰關係是指每一首歌詞中，每一句的詞彙的前後相鄰的歌詞為何，

但記錄時，並無順序關係，例如「我的」和「的我」，因為不管是在「我」之前出現或之後出現的「的」，都是相鄰的關係，所以會記錄在同一筆。

整體例子如下：由周華健演唱及作詞的「明天我要嫁給你啦」，其中兩句歌詞：
「我\的\眼淚\閃爍\閃爍\好\空洞」、「我\的\心跳\撲通\撲通\地\陣陣\悸動」，
可以產生以下圖表，如圖 3-5。

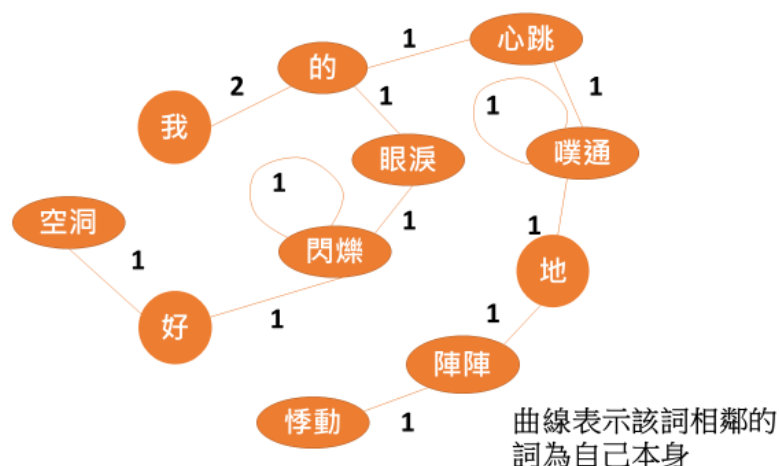


圖 3-5 詞彙相鄰關係示意圖

3.7 押韻分析

華語流行音樂歌詞可說是現今時代的韻文，因此，歌詞作品中或多或少保留了當今口語語言的語感，透過分析、研究歌詞的押韻，可以發掘當代語言的語音特色。

押韻判斷方式通常以判斷韻腳的音節為主，韻腳係指該句句末的音節，例如在一首押韻的歌詞中，每句末使用了同樣韻母的字，這就叫做該歌詞的韻腳。

根據劉祐銘的臺灣國語流行歌曲歌詞用韻研究(1998~2008)，流行音樂歌曲的歌詞，並沒有限制哪一句不可以押韻，哪一句一定要押韻；也沒有限制偶數句或單數句一定要押韻或不押韻的規定，因此，基本上國語流行歌曲歌詞的用韻是相當彈性的。

3.7.1 押韻類別定義：一韻到底、換韻、無韻、句首韻

根據劉祐銘的臺灣國語流行歌曲歌詞用韻研究(1998~2008)，押韻用例可分為一韻到底、換韻、無韻、句首韻、因為節拍拉長而造成的押韻、節奏停頓處押韻等方式，而本

研究以歌詞為主，無法考慮節拍、節奏部分，所以無判斷「因為節拍拉長而造成的押韻」、「節奏停頓處押韻」這兩部分。

一韻到底：

所謂一韻到底，也就是整首歌所押的韻都一樣，並未換韻，例如由黃霑作詞，劉德華演唱的〈最愛上海灘〉，該歌詞如下，整首歌詞為ㄣ韻：

難忘燦爛誰會忘得了上海灘
難忘浩瀚一浪一浪的奇談
絕色超凡只會出現在上海灘
一身是膽一起轟轟烈烈地干
別問是愛還是仇都在滔滔裡轉彎
翻千番拐千灣你和我是不見不散
無憾無愁我的最愛上海灘
同步灘頭和你好好地幹一番

換韻：

在同一段歌詞中，會有換韻的現象發生，即歌詞中的前幾句是押某種韻，到某幾句後又改押另一種韻，例如由沈光遠作詞，由巫啟賢演唱的〈我終於失去了你〉的第一段歌詞，歌詞韻腳一開始為ㄨ韻，中間轉為ㄣ韻，第三句轉回ㄨ韻，最後又轉為ㄣ韻：

當所有的人離開我的時候 妳勸我要耐心等待
並且陪我度過生命中最長的寒冬 如此的寬容
當所有的人靠緊我的時候
妳要我安靜從容 似乎知道我有一顆永不安靜的心 容易蠢動

無韻：

華語流行歌詞中有些歌詞是無韻的，例如由娃娃作詞，張鎬哲演唱的〈渡河〉，在分析過這段歌詞各句的最後一字後，發現並沒有押韻的跡象。

人群都散了 風裏傳來聲音 天際孤星已升起

河水一樣的寂寞 黑暗中流過 我的心就在其中沉浮
愛人也走了 黑夜不再溫柔 心中感覺是虛空
河水一樣的愛情 生命中流過 我的愛就在其中失落
我知道愛情是一種奢求 我總是無法找到渡頭
我知道一切都不該等候 卻只能隨著命運而漂流

句首韻：

押韻通常是押在一個句子的句尾、句末處，但其實也會產生在句首，意即在每一句歌詞的首字押韻，例如由張思爾作詞，由王心凌演唱的〈當你〉這首歌其中一段歌詞，歌詞如下，「當、想」為這段歌詞的句首韻：

當你的眼睛 眯著笑
當你喝可樂 當你吵
我想對你好 你從來不知道
想你想你 也能成為嗜好
當你說今天的煩惱
當你說夜深 你睡不著
我想對你說 卻害怕都說錯 好喜歡你 知不知道

3.7.2 利用 Non-Trivial Repeating Pattern 演算法判斷押韻

判斷押韻如果只判斷韻母出現的次數出現在該首歌的比例，會無法了解韻腳進行的模式，從而判斷其押韻方式，因此，為了更準確地能判斷韻腳、了解韻腳進行的方式及判斷押韻的方式，本研究以 Non-Trivial Repeating Pattern 進行判斷。

Non-Trivial Repeating Pattern 係一種演算法[1]，2001 年由 Jia-Lien Hsu、Chih-Chin Liu 及 Arbee L.P. Chen 等人提出。利用此演算法，可在一連串的字串中，找出 Non-Trivial Repeating Pattern。

例如，假設有一首歌詞的句末字為：「愛—好—腦—待—愛—好—傲—愛—喔—待—而—開」，首先將中文拆成漢語拼音的方式以 Python 的 pinyin 模組，與注音符號對照

的部分可參考教育部提供的注音符號及漢語拼音對照表，將字串轉為拼音字串”ai-ao-ao-ai-ai-ao-ao-ai-o-ai-er-ai’後，用以下步驟進行分析：

1. 建立成關係矩陣，如表 3-1。

表 3-1 字串‘ai-ao-ao-ai-ai-ao-ao-ai-o-ai-er-ai’的關係矩陣

	ai	ao	ao	ai	ai	ao	ao	ai	o	e	er	e
ai	-			1	1			1		1		1
ao		-	1			2	1					
ao			-			1	3					
ai				-	1			4		1		1
ai					-			1		1		1
ao						-	1					
ao							-					
ai								-		1		1
o									-			
ai										-		1
er											-	
e												-

2. 考慮 Non-Trivial Repeating Pattern 四種情況，即：

Case 1: $Q_{i,j} = 1$ and $Q_{(i+1),(j+1)} = 0$

Case 2: $Q_{i,j} = 1$ and $Q_{(i+1),(j+1)} \neq 0$

Case 3: $Q_{i,j} > 1$ and $Q_{(i+1),(j+1)} \neq 0$

Case 4: $Q_{i,j} > 1$ and $Q_{(i+1),(j+1)} = 0$

4. 移除 $\text{rep_count} = \text{sub_count}$ 項目，例如(“ai-ao”,1,1)，(“ai-ao-ao”,1,1)，(“ao-ao”,1,1)，(“ao-ao-ai”,1,1)及(“ao-ai”,1,1)均須被移除，因為它們都屬(“ai-ao-ao-ai”,1,0)的子字串

5. 最後，以公式獲得實際的出現次數，如表 3-2。

公式：

$$f = \frac{1 + \sqrt{1 + 8 \times \text{rep_count}}}{2}$$

表 3-2 字串‘ai-ao-ao-ai-ai-ao-ao-ai-o-ai-er-ai’的所有 Non-trivial Repeating Pattern 相關資料

Repeating Pattern	Pattern 長度	Repeating Pattern 頻率
ao (ㄠ)	1	4
ai (ㄞ)	6	6
ai-ao-ao-ai (ㄞ-ㄠ-ㄠ-ㄞ)	4	2

接著，以 Non-Trivial Repeating Pattern 的片段、長度及頻率來判斷押韻方式，如表 3-3。

- 片段：即以上述方式找到的片段，例如 ao 或 ai-ao-ao-ai。
- 長度：例如 ai-ao-ao-ai 有 4 個音，所以長度為 4。
- 頻率：例如 ao 出現 4 次，ai-ao-ao-ai 則出現 2 次。

表 3-3 以 Non-Trivial Repeating Pattern 判斷押韻方式

	一韻到底	換韻	無韻	句首韻
片段	1 種音	1 種以上	0 種	1 種(含)以上
長度	總長度-1	2 (含)以上	-	2 種(含)以上
頻率	2 次(含)以上	2 次(含)以上	-	2 次(含)以上

3.8 修辭分析

修辭是文學的美容師，修辭可以美化文學，文學可以美化人生，流行歌詞亦須要修辭，來使得歌詞更生動，更能扣人心弦。

3.8.1 明喻、類字、疊字、疊句及對偶之定義

修辭方式非常多種，根據相關研究，修辭方式可多達 30 種，每種都可豐富歌詞的內容，且可使歌詞更朗朗上口或更使人感同身受。根據吳嫩婉的臺灣國語流行歌曲的修辭藝術，以類疊、對偶等辭格出現在流行歌曲中的頻率最高，而本研究除了分析類疊、對偶外，亦分析隸屬於譬喻法中的明喻。

明喻定義：

根據蔡宗陽的《應用修辭學》一書，譬喻法包括明喻、隱喻、略喻、借喻及假喻五類，明喻的結構主要分為三部份：

- 喻體，指作者欲描述之主題。
- 喻依，指的是用來說明喻體的事物。
- 喻詞，指用來連結喻體和喻依的補助詞。

以程式分析明喻，可以判斷該句是否含有喻詞，喻詞列表產生的方式如下：

- (1) 根據蔡宗陽《應用修辭學》一書，喻詞有：如同、恰似、宛如、像、彷彿、似、好比、如、若。
- (2) 利用教育部重編辭語修訂本，找出與第一步驟喻詞的同義詞，包括猶如、好像、好似、類似、相像、相類、類似、如同、相似。
- (3) 利用教育部重編辭語修訂本，找出與第二步驟喻詞的同義詞，包括恰似、鬚鬚、近似、形似、相仿、相像、相似、類似。
- (4) 利用教育部重編辭語修訂本，找出與第三步驟喻詞的同義詞，但這時發現同義詞皆為步驟 1~步驟 3 囊括的詞彙。
- (5) 將喻詞列表整理為：「如同，猶如，好像，恰似，好似，宛如，彷彿，類似，似乎，似，相像，相類，類似，好比，如，如同，好像，若，像，相似，鬚鬚，近似，形似，相仿，相像，相似，類似，雷同」。

然而，以中研院斷詞系統，無法針對「如」進行完整斷詞，舉例如圖 3-6：

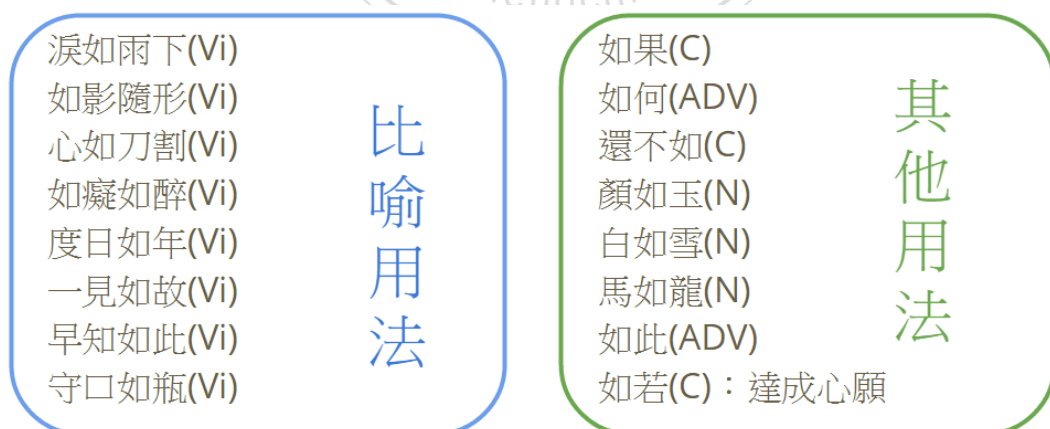


圖 3-6 「如」的斷詞及用法比較

因此，本研究將全部歌詞含有「如」的句子整理出來，並進行以下步驟：

- (1) 將該詞以人工判斷，分析是否為相似的意思。
- (2) 發現有使用「如」的句子多為明喻的用法。
- (3) 若該詞非相似的意思，則列入「負面」表列。

除此之外，以中研院斷詞系統，無法針對「若」進行完整斷詞，舉例如圖 3-7：

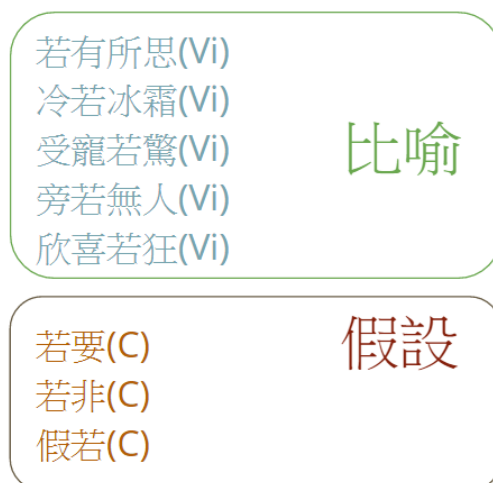


圖 3-7 「若」的斷詞及用法比較

因此，本研究將全部歌詞含有「若」的句子整理出來，並進行以下步驟：

- (1) 將該詞以人工判斷，分析是否為相似的意思。
- (2) 發現有使用「若」的句子多為假設的用法。
- (3) 若該詞非相似的意思，則列入「正面」表列。

接著，即可以程式開始判斷該首歌詞是否有使用明喻修辭法：

- (1) 該首歌詞是否含有喻詞列表的詞
- (2) 若該首歌詞含有「如」，則判斷該詞是否有在「如的負面表列」，若有，則非明喻，反之，則判斷為明喻。
- (3) 若該首歌詞含有「若」則判斷該詞是否有在「若的正面表列」，若有，則為明喻，反之，則判斷為非明喻。

類字、疊字及疊句的定義及分析方式：

根據國立師範大學國文系的蔡宗陽教授所著的《應用修辭學》，「類字」、「疊字」及「疊句」的比較及例子如表 3-4。

值得注意的是，類字須排除停用字(stop word)，包括：我、你、妳、他、她。

疊字的判斷須排除專有名詞，例如：媽媽、姐姐、姊姊、爸爸、星星、公公、哥哥、弟弟、叔叔、伯伯、奶奶、爺爺、弟弟、妹妹、謝謝、寶寶、娃娃、婆婆、嫂嫂等。

表 3-4 類疊的定義及範例（整理自蔡宗陽的應用修辭學一書）

項目	定義	舉例及解釋
類字	同句，同一字詞間的間隔使用。	<p>例：燕去燕來還過日，花開花落即經春。</p> <p>解釋：間隔使用「燕」、「花」字，為此句的類字。</p>
	不同句，但有使用相同的字。	<p>例：你愛穿什麼就穿什麼；扮一個牧童，扮一個漁翁，裝一個農夫，裝一個走江湖的吉普賽，裝一個獵戶。</p> <p>解釋：間隔中使用兩個「穿」、「扮」，三個「裝」字，「穿」、「扮」及「裝」為這段的類字。</p>
疊字	同句，同一字的連續使用。	<p>例：由於這些花，我自然而然地想起北平公園裡的花花朵朵，與這些簡直沒什麼兩樣。</p> <p>解釋：其中「花花朵朵」，不是指一花一朵，而是形容花朵眾多。「花」、「朵」為這句的疊字。</p>
疊句	不同句，但有連續使用同一語句。	<p>例：不要哭，不要哭；一會兒就不痛了。</p> <p>解釋：重複使用不要哭兩次，表示加強作用，屬於疊句。</p>

對偶的定義及分析方式：

根據黃麗貞著《實用修辭學》（增訂版），對偶從形式上可分為兩種：

- (1) 嚴式對偶：即「工對」，要求上下兩句字數相等，結構相同，詞性相對，平仄相對，不能重複用字。
- (2) 寬式對偶：學者黃麗貞認為「寬式對偶」：「就是在對偶的上、下句（或段）之中，『字數不必絕對相等；也不求兩兩相對；也不避字同意同』。由於內容的限制或時間的緊迫，採用工對很難，那就不用寬對。

此外，根據蔡宗陽的應用修辭學，對偶與排比的辨別，若依句型而言，對偶是兩兩相對，但排比經常是三項或三項以上。

因此，本研究在定義對偶時，以寬式對偶為主，且兩句為兩兩相對，無三句或三句以上。

程式分析方法：該句與下一句進行詞性比對，若兩句所用的詞性皆相同，則判斷為對偶，且第三句與這兩句不同，舉例如下：

誰(N) 能(ADV) 了解(Vt) 我(N) 的(DE) 情(N),

誰(N) 能(ADV) 接受(Vt) 我(N) 的(DE) 愛(N)

那(DET) 是(Vt) 怎樣(Vi) 的(DE) 一(DET) 種(M) 生活(N),

那(DET) 是(Vt) 怎樣(Vi) 的(DE) 一(DET) 種(M) 罪過(N)

3.8.2 利用 Non-Trivial Repeating Pattern 演算法判斷類字、疊字及疊句

類字、疊字及疊句程式部分，以 Non-Trivial Repeating Pattern 方式進行分析。

步驟如下：

- (1) 判斷該句是否為中文，若不是，則移除非中文字部份。
- (2) 將該首歌的全部歌詞的每個字，建立成關係矩陣。
- (3) 以 Non-Trivial Repeating Pattern 方式記錄 pattern，並記錄 pattern 在歌詞的何處，便可得知 pattern 在哪一句，方便之後查詢是否在同一句或是相鄰句。
- (4) 判斷 pattern 是否在同一句，若為同句，且 pattern 都是相同字、字是相鄰的，則為

疊字，例如「天天想你」的 pattern 為「天」，「天」出現的地方皆為同句，且是相鄰的，因此為疊字。

(5) 判斷 pattern 是否在同一句，若為同句，且 pattern 的字不相鄰，則為類字。例如「若隱若現」的 pattern 為「若」，且「若」被記錄的位置在同一句，但兩字不相鄰，因此為類字。

(6) 判斷 pattern 是否為相鄰句，若為相鄰句，且 pattern 都是相同字、pattern 長度等於該句字數，則為疊句，例如「無敵鐵金鋼;無敵鐵金鋼;無敵鐵金鋼」的 pattern 為「無敵鐵金鋼」，且長度等於該句字數，因此這三句為疊句。

(7) 判斷 pattern 是否為相鄰句，若為相鄰句，且 pattern 長度不等於該句字數，亦為類字的一種。例如「我的身體在等待自由、我的腳趾在等待自由、我的手指在等待自由」的「我的」和「在等待自由」為 pattern，且「我的」「在等待自由」不等於該句字數，所以為類字。

3.9 歌詞風格自動分類

歌詞的風格可以分為好幾種，例如抒情歌、民謠、勵志及搖滾等，聽眾會依自己的心情選擇想要聆聽的歌曲，藉此抒發情緒或是沉浸於歌詞描述的感覺。

本研究的歌詞風格分類是根據歌詞的用詞將其分類。分類技術是按照分析對象的屬性，分門別類並加以定義。利用已知的資料及其分屬性質，來建立資料分類的模型，此分類模型可用來預測未知的新資料。

我們從流行音樂歌詞的相關研究，獲得標示有風格類別的資料，資料來源包括吳嫩婉的台灣國語流行歌曲的修辭藝術〈1949-2000〉、馬占山的臺灣流行歌詞主題類型與語言表達研究（1999~2008）——以「Hit-FM 年度百首單曲」為對象、九酷音樂[24]及酷我音樂[25]，並以此做為訓練資料 (Training Data)。

歌詞的特徵值部分，本研究實驗了兩種方法，以提升準確率：

(1) 將訓練資料的歌詞的斷詞做計數統計，並保留計數出現 100 次以上的詞彙，產生為「詞總集合」。

(2) 將訓練資料的歌詞的斷詞做計數統計，然後僅保留形容詞及計數出現 100 次以上的詞彙，產生為「詞總集合」。

接著將每首歌詞，根據詞總集合，轉為向量陣列。例如：詞總集合為「天: 100，愛: 250，氣球: 50」三個詞，冒號後面數字表示在訓練資料中出現的次數，而若該首歌詞只有「我愛你」，那傳回值就是一個陣列「0,0.004,0」，因為「愛」出現了一次，除以詞總集合中「愛」出現次數的 250 次，得到 0.004 的值，而「天」、「氣球」出現了零次，因此再向量陣列中為 0。

透過此方式，我們可以獲得每首歌詞的向量陣列，並以分類器進行自動分類。分類是資料探勘的一種非常重要的方法，分類的概念是在既有資料之基礎上，建立起模型，從而將未歸類的資料分門別類。資料探勘演算法有 Support Vector Machine (SVM)、貝氏網路 (Bayesian network) 及決策樹(decision tree)等，而本研究主要使用 SVM，乃因其特點為它能夠同時最小化經驗誤差及最大化幾何邊緣區，對於解決小樣本、高維度和非線性有較好的解決能力。

SVM 通常應用在機器學習(Machine Learning)，係監督式學習(Supervised Learning)的方法之一，可廣泛應用於統計分類與回歸分析。SVM 的核心思想可概括分為：(1) 將訓練資料視為多維的向量資料，針對線性可分條件下進行資料分析，對於非線性可分之原始資料，使用非線性轉換演算法來升維，使得在高維度空間可使用線性演算法來分析；(2) 在特徵空間中，找出最優分割超平面(hyper-plane)，使得分格的兩邊邊界越寬越好，如圖 3-8。

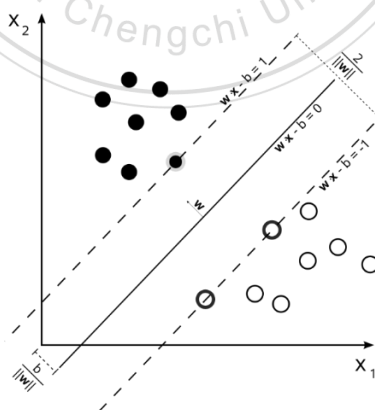


圖 3-8 SVM 分割示意圖

(Retrieved June 10, 2016, https://en.wikipedia.org/wiki/Support_vector_machine)

第四章 實驗

4.1 網站實作

網站架構可分為詞庫建造、特徵擷取、以 Non-Trivia Repeating Pattern 判斷押韻及類疊、風格分類器、使用者輸入及結果顯示等。

4.1.1 架構

本研究將各類研究項目的結果儲存在資料庫，並以網站方式呈現，如此一來，便可讓使用者進行以下項目的交叉查詢，如表 4-1 的網站內容一覽表。

表 4-1 網站內容一覽表

類別	項目	
修辭	明喻	疊句
	類字	對偶
	疊字	
押韻（尾韻）	無韻	
	一韻到底（每句字尾一定都有押該韻）	
	一韻到底（不一定每句字尾都有押韻）	
	換韻	
押韻（頭韻）	無押韻	有押韻
使用語系	純中文	中日英
	中英	中韓
	中日	中韓英
風格	愛情	勵志
	親情	其他
	友情	
歌名字數	1~10 字部或 10 字部以上	
作詞者	曾獲金曲獎的 19 位作詞者或可自行輸入	
年代	1956~2015	

4.1.2 工具

斷詞處理採用中央研究院的「中文斷詞與剖析系統」。該系統係由中央研究院的資訊科學研究所和中文詞知識小組研發，由於該系統僅支援 BIG5 與 UTF-16LE，所以須要先將歌詞轉換為此兩種格式，接著進行斷詞轉換，最後，再將歌詞格式轉為 UTF-8，方便日後做分析。

至於整個系統實作工具如表 4-2。

表 4-2 系統實作工具一覽

Web Framework	Django
Database	SQLite3
Front-End	HTML
	CSS
	JavaScript
3rd Party API	Google Chart API

網站部分建置是基於 Django 的 Framework，依照表 4-3 的資料庫格式，將所有歌詞的資料填入，並匯入 Django 的 ORM 系統，接著以即可使用 Django 的搜尋功能進行交叉查詢，並以網頁方式呈現。此外，歌詞數量年代圖則是以 Google Chart API 繪製而成。

4.1.3 資料庫

本研究使用的資料庫係以 Django 指定 database engine 為 sqlite3，資料格式如表 4-3，以此格式匯入資料庫，如此一來，得以透過 Django Model 來跟資料庫做互動。

表 4-3 資料庫的格式

欄位	內容
ID	資料庫的 ID
演唱者類型	男生/ 女生/ 團體/ 其他
演唱者	演唱者的姓名
歌名	歌名
專輯	專輯
使用語言	純中文/ 中英/ 中日/ 中日英/ 中韓/ 中韓英
年代	年代
作詞者	作詞者姓名
作曲者	作詞者姓名
歌名字數	歌名的字數
使否為舊曲新唱	是/ 否
使否使用明喻修辭法	是/ 否
明喻喻詞	喻詞的詞彙
使否使用類字修辭法	是/ 否
使用的類字詞彙	類字的詞彙
使否使用對偶修辭法	是/ 否
使用的對偶句子	對偶的句子
使否使用疊字修辭法	是/ 否
使用的疊字詞彙	疊字的詞彙
押韻型態	一韻到底/ 換韻/ 無韻/ 句首韻
押韻的韻腳	韻腳的音
歌詞	歌詞
風格	愛情/ 親情/ 友情/ 勵志/ 其他

4.1.4 資料來源與資料的統計分析

本研究使用的歌詞係由知名的歌詞網站魔鏡歌詞網（Mojim.com）而來，截取 1956 年至 2015 年 10 月的歌詞，共 191,626 首。移除歌詞內容完全相同的歌曲 46,382 首，及移除非國語的歌詞，包含粵語、台語、純英文、純日語、純韓語、客家話、原住民語等 56,596 首歌詞，可進行研究的歌詞有 88,648 首，如表 4-4。

表 4-4 實驗歌詞筆數的分析

	歌詞數量
資料前處理	191,626 首
移除重複的歌詞	46,382 首
移除非華語的歌詞	56,596 首
可研究的歌詞	88,648 首

4.2 詞彙分析結果

本研究將常用的詞彙統整，並以文字雲方式呈現，藉此可以得知何種詞彙較容易被使用。文字雲工具部分，採用由簡冠庭製作的 HTML5 Word Cloud 工具，將高頻詞以文字雲方式呈現。

4.2.1 歷年歌詞之高頻辭

全部歌詞：

華語流行音樂歌詞全部歌詞中，最常出現的詞彙前五名為：我們 (62,597 次)、沒有 (53,001 次)、什麼 (47,663 次)、自己 (41,715 次) 及世界 (30,329 次)。原本預期第一人稱用詞「你、我、他」會比較常出現，然而，結果是第一人稱複數代名詞的「我們」最常出現，第一人稱反而未出現在 Top 5。文字雲部分參考圖 4-1。

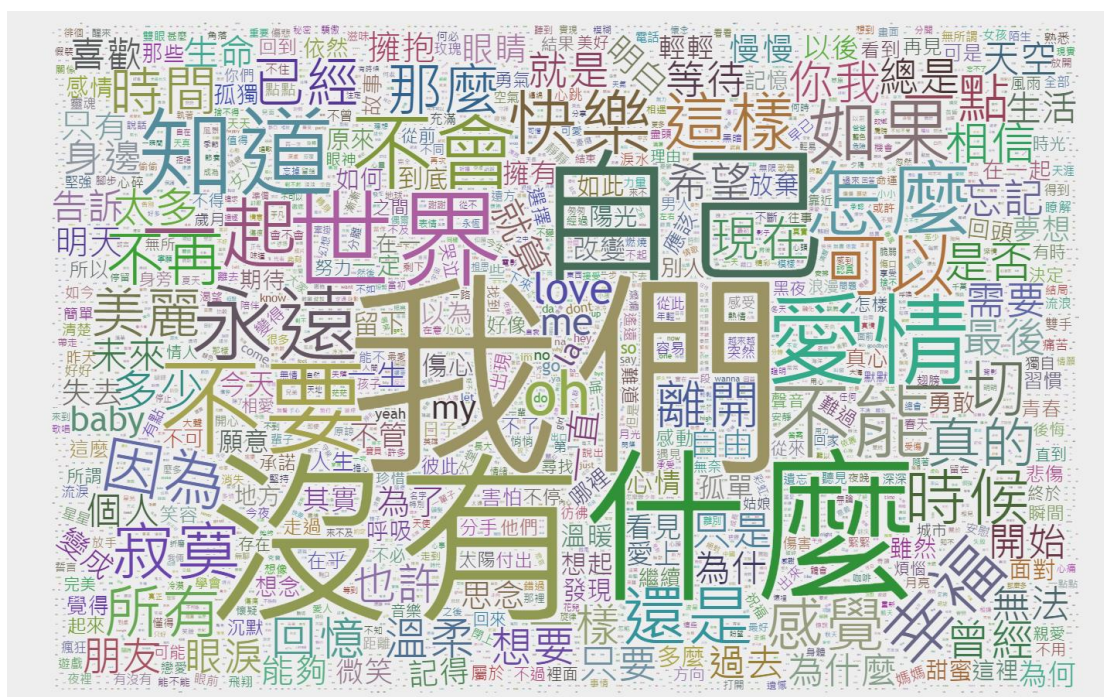


圖 4-1 全部歌詞的詞彙的文字雲

名詞：

華語流行音樂歌詞「名詞」詞彙中，排除你、我、他三個詞彙，最常出現的詞前五名為：我們 (62,596 次)、自己 (41,712 次)、世界 (30,328 次)、愛情 (28,987 次)及時候 (15,774 次)，文字雲部分參考圖 4-2。高頻詞中，「我們」和「自己」都與自身較相關，世界則有種較大我的感覺，含有世界的歌詞數量比例為 18.8%，然進一步了解，歌詞中出現「世界」的歌詞為愛情風格的比例達 58.8%，代表這此詞彙非大我的意思，而係指愛情中的世界。

例如由鄧麗君演唱、翁炳榮作詞的「戀愛的路多甜」，歌詞中出現「世界」的歌詞：

這個世界只能容納我們兩個人

只有我們才能體會相互的情意



圖 4-2 名詞詞彙的文字雲

動詞：

華語流行音樂歌詞「動詞」詞彙中，最常出現的詞前五名為：沒有 (35,417 次)、知道 (27,923 次)、離開 (14,787 次)、相信 (12,945 次) 及想要 (11,902 次)，文字雲部分參考圖 4-3。「沒有」為最常出現的動詞，其雖也有副詞的意義，但中研院的斷詞系統，可明確判斷出歌詞中的沒有的詞性為動詞或是副詞，例如「相愛沒有那麼容易」，此處的沒有為副詞用法，斷詞系統會顯示為副詞；至於「沒有你怎麼辦」，此處的沒有為動詞用法，斷詞系統則為顯示該詞為動詞。若將第一人稱為受詞，發現「沒有你」的比例為 9.3%，「沒有我」的比例為 2.1%，「沒有他」的比例則為 0.4%，由此可見，描述沒有對方(你)的情況較多。

例如：由余天演唱、莊奴作詞的「戀愛的路多麼苦」，歌詞中出現「沒有你」的段落：

我的心裡想念你 你知道我不能沒有你
我愛你 我恨你 愛你又恨你

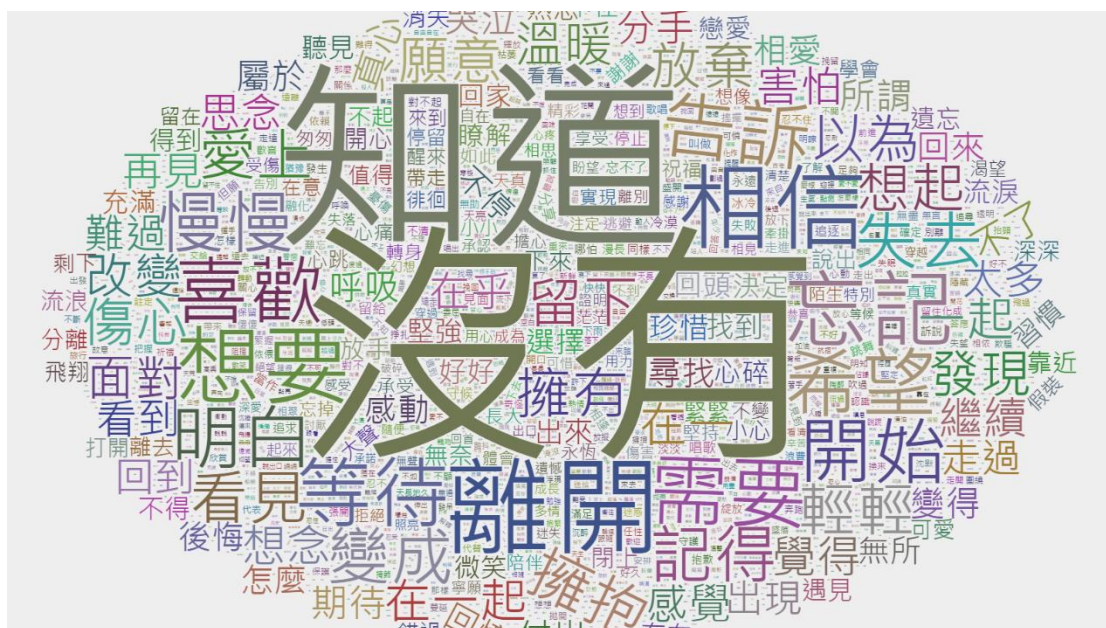


圖 4-3 動詞詞彙的文字雲

形容詞：

華語流行歌詞「形容詞」詞彙中，最常出現的詞前五名為：快樂 (18,558 次)、幸福 (16,483 次)、寂寞 (16,238 次)、溫柔 (11,453 次) 及孤單 (7,353 次)，文字雲部分參考圖 4-4。由形容詞的高頻詞可得知積極正面的詞彙出現比例較負面詞彙多，正面詞彙例如快樂、幸福、溫柔及勇敢等，負面詞彙則為寂寞、孤單及悲傷等，但這並不能完全表示正面情緒的歌較多，因為有可能歌詞主要目的為抒發感傷之心情。例如張惠妹的我要快樂，歌名和歌詞中的快樂為主要詞彙，但實則表達難過的心情，副歌歌詞如下：

我要快樂 我要能睡的安穩
 有些人不抱了才溫暖 離開了才不恨 我早應該割捨
 我要快樂 哪怕笑的再大聲
 心不是熱的 全都是假的
 只有眼淚是真的 (我的決定是對的)

4.2.2 1960~ 2015 逐年的高頻詞詞彙

本研究以每 10 年進行高頻詞詞彙分析，發現 1960~1999 主要的高頻詞彙為：我們、什麼、沒有、不要、自己、愛情，然而，2000 年後，「不要」消失了，取而代之的是「世界」。參考圖 4-6~ 圖 4-11。

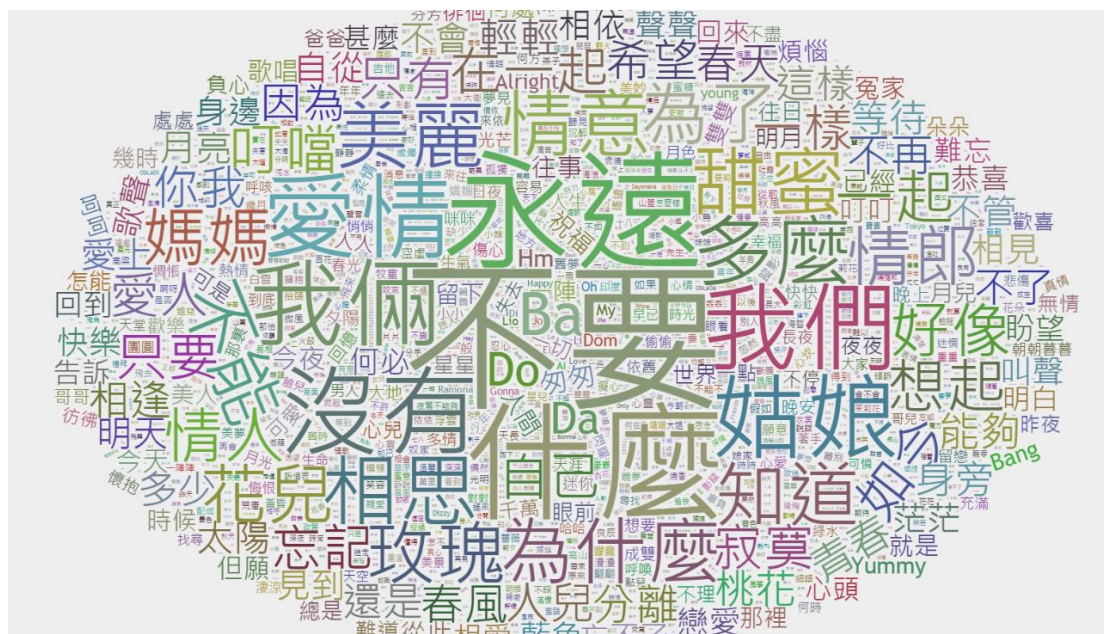


圖 4-6 1960~1969 年華語流行歌詞的文字雲

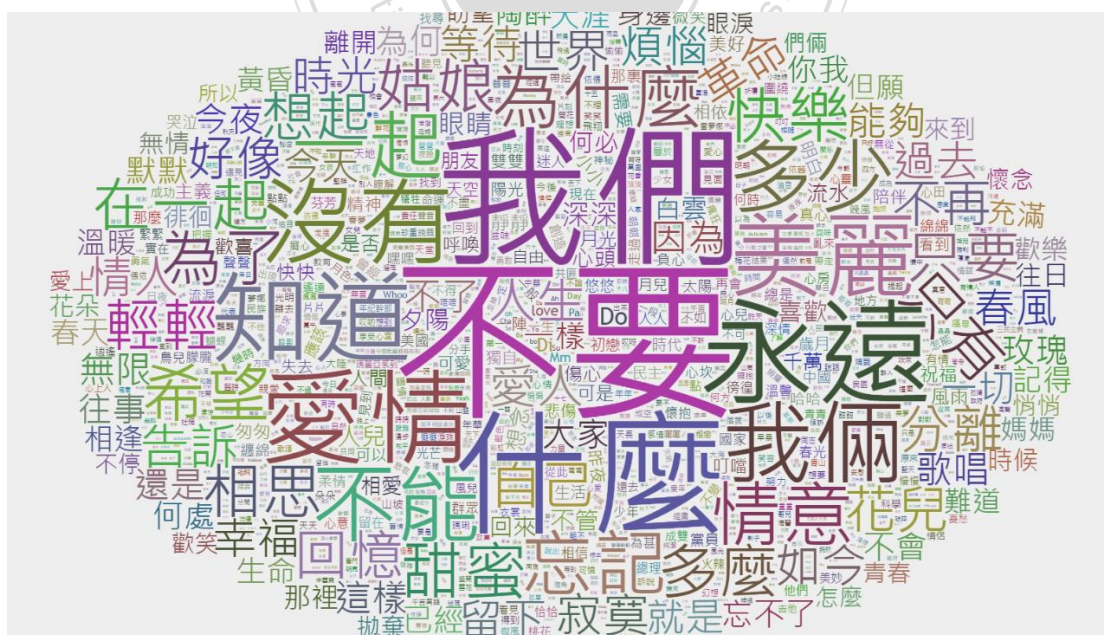
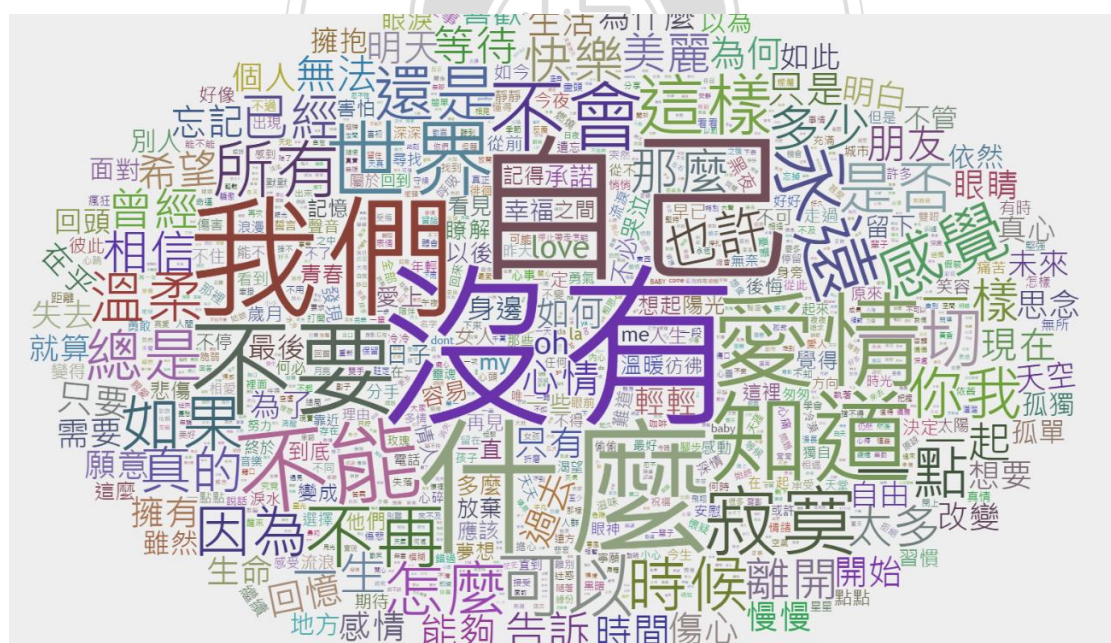
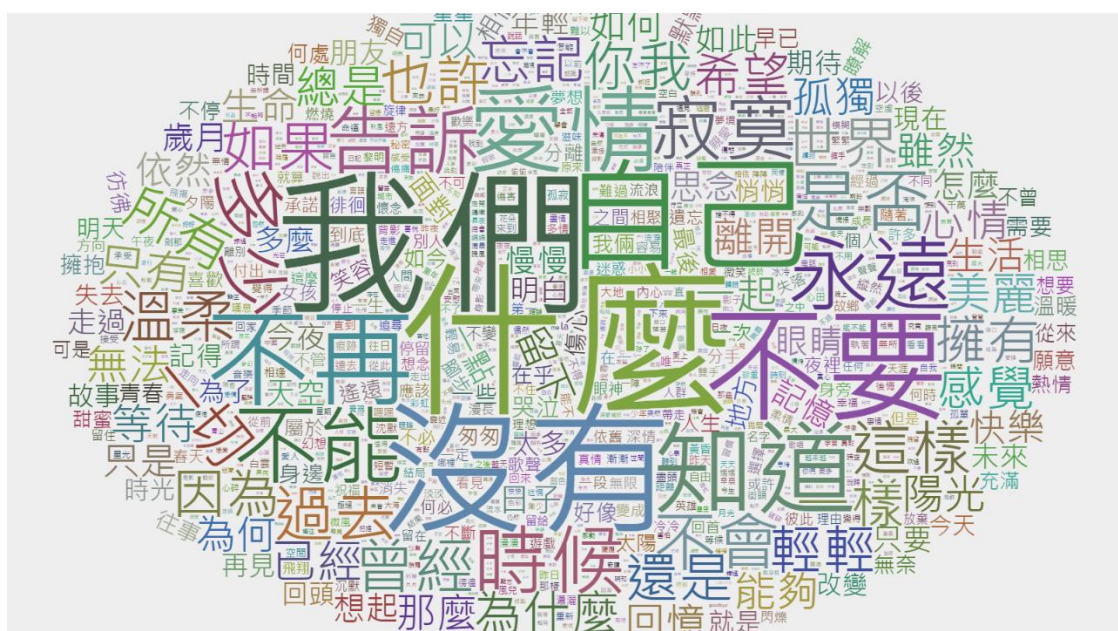
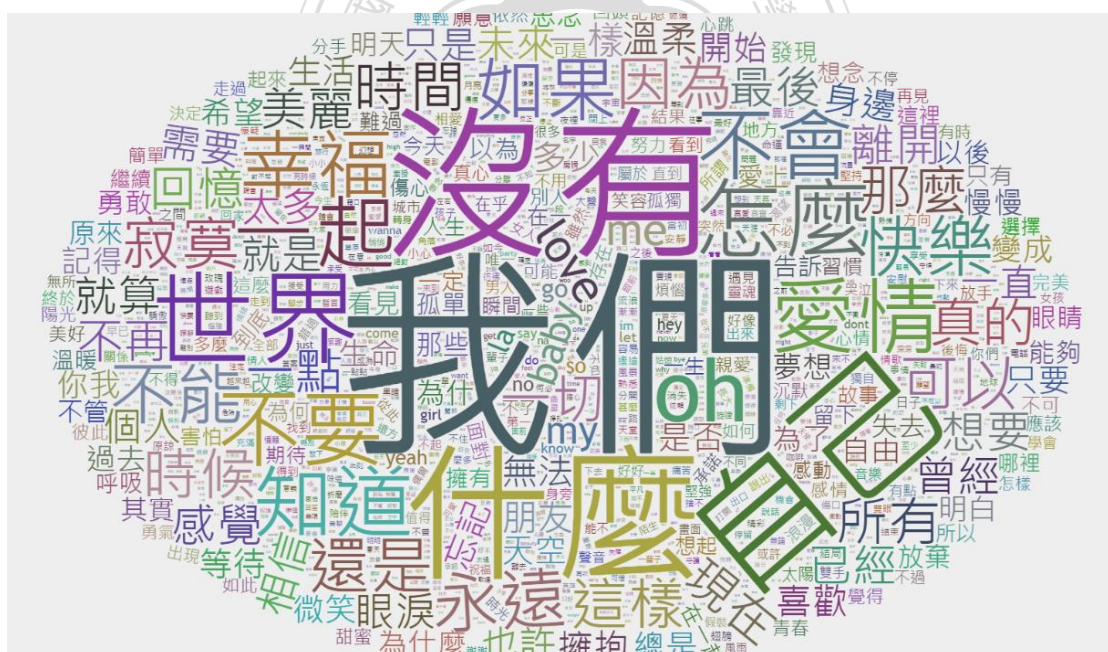
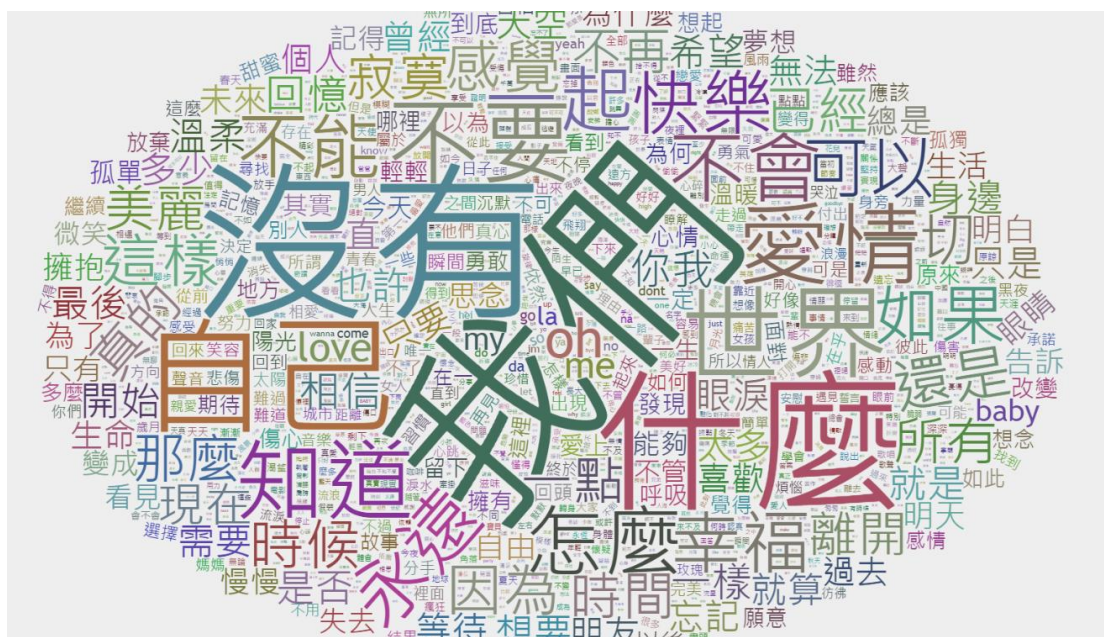


圖 4-7 1970~1979 年華語流行歌詞的文字雲





4.2.3 特別詞彙之頻率分析

在 88,648 首歌詞中，本研究選出一些詞彙，如表 4-5 及表 4-6。

四季中，最常提到的季節為「春」，佔歌詞數量 9.6%，其次為「秋」，比例為 4.0%。

至於「爸爸」和「媽媽」或其相關詞「父親」、「母親」，則以媽媽或母親居多，參考表 4-5。

至於人稱代詞部分，則以第一人稱的「我」最常出現，佔所有歌詞的 90.7%，其次為第二人稱的「你/妳」，佔 77.6%；第三人稱的「他/她」則佔 21.9%。

性別部分，「女」出現比例 7.6%，略大於「男」的 4.9%。

味道部分，包括酸、甜、苦及辣，「苦」位居首，佔 12.2%，不過值得注意的是，這四字在歌詞中，有可能意指心情，而非對食物的評論。

心情部分，愛、恨、哭、笑、哭泣及傷心，「愛」出現的比例為 62.7%，遠高於「恨」的 5.2%，「笑」和「快樂」的比例，則高於難過相關的詞彙，包括「哭」、「哭泣」及「傷心」。

天體相關用詞部分，本研究討論「星星」、「月亮」及「太陽」，「太陽」的比例比例為 3.5%，其次為「星星」，比例為 2.7%，最後為「月亮」，比例為 2.0%。

表 4-5 特別詞彙之頻率分析 I

類別	詞彙	筆數	比重
四季	春	8,519	9.6%
	夏	2,934	3.3%
	秋	3,560	4.0%
	冬	3,182	3.6%
父母相關詞	爸爸	764	0.9%
	媽媽	1,489	1.7%
	父親	119	0.1%
	母親	403	0.5%
	「爸爸」及「媽媽」兩詞彙同時出現	519	0.6%

表 4-6 特別詞彙之頻率分析 II

類別	詞彙	筆數	比重
人稱代詞	你/ 妳	68,822	77.6%
	我	80,436	90.7%
	他/ 她	19,427	21.9%
	我們	22,495	25.4%
性別	男	4,356	4.9%
	女	6,744	7.6%
味道	酸	1,976	2.2%
	甜	8,102	9.1%
	苦	10,792	12.2%
	辣	653	0.7%
心情	愛	55,586	62.7%
	恨	4,597	5.2%
	哭	9,323	10.5%
	笑	21,350	24.1%
	哭泣	3,192	3.6%
	傷心	4,480	5.1%
	快樂	10,381	11.7%
天體相關詞	星星	2,383	2.7%
	月亮	1,732	2.0%
	太陽	3,076	3.5%

4.2.4 詞彙相鄰關係分析結果

詞彙的相鄰關係分析使我們熟悉哪些詞彙容易一起出現，藉此更加了解詞與詞之間的關係。

針對 88,648 首歌曲，斷詞後進行相鄰(Adjacency) 關係分析，前 10 名常使用的詞彙

如圖 4-12，使用超過 5,000 次的相鄰詞彙如圖 4-13 的文字雲。

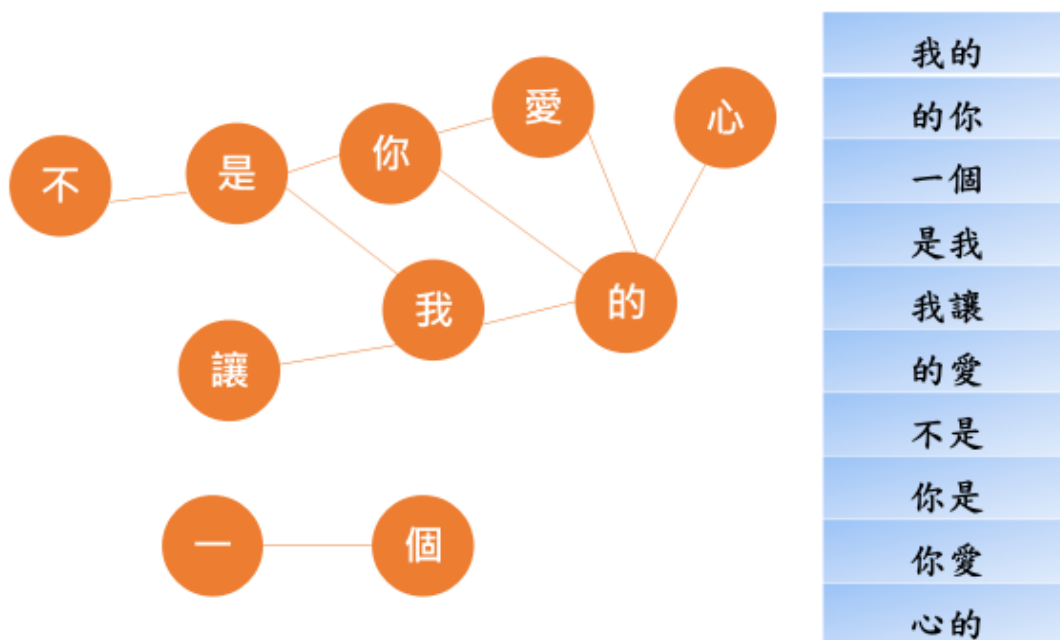


圖 4-12 相鄰詞彙(Word Adjacency) Top 10 使用的詞彙



圖 4-13 相鄰詞彙(Word Adjacency) 超過 5,000 次的相鄰詞彙

位居前兩名的詞為「我」、「的」，其雖然為一個字，但亦是詞，為單字詞。根據

中央研究院語言學研究所黃居仁博士的「漢字知識表達的幾個層面：字，詞，與詞義關係概論」：在中文裡，字與詞的最大差別，在每個詞可能帶有一個或一個以上的字。因此，意義上，「我」是一個字詞，因為該字有其意義，即表示第一人稱的代名詞。

找到這些詞的相鄰關係後，即可以這些常用的兩兩相鄰的詞彙來作詞。舉例來說，以圖 4-14 中的常用的相鄰詞彙，找出以下常用的相鄰詞彙：「我的」、「的愛」、「你的」、「的心」、「是我」、「美麗的」、「的世界」，即可組成一句歌詞：「我的愛你的心 是我美麗的世界」。

4.3 歌名字數分析結果

針對 88,648 首歌曲的歌名進行分析，歌名字數為四字部的比例達 24.3%，即每 4 首歌就有一首歌名為四個字。

而根據臺北之音 HitFM 聯播網年度百首單曲排行，最常使用的歌詞長度為二字部 (23.5%)和四字部(23.35%)，且其他字部意與我們分析結果相近，此有可能是唱片公司或作詞者認為使用二字部~四字部的歌名容易被記憶或查詢（例如 KTV、卡拉 OK），因此多創作二字部~四字部的歌名。如圖 4-14。

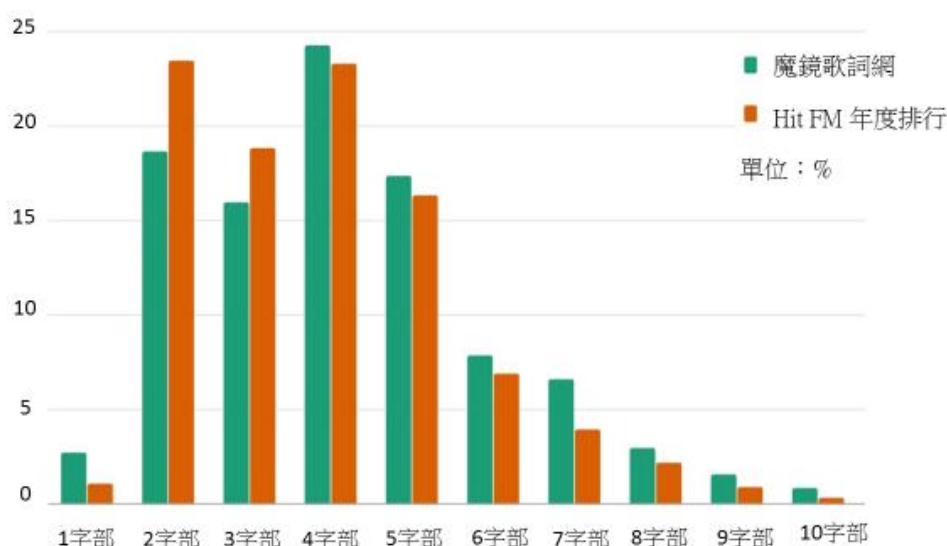


圖 4-14 魔鏡歌詞網歷年歌詞與 Hit FM 年度百首單曲歌名字數比例比較

4.4 使用語系分析結果

88,648 首歌詞中，僅使用中文而未使用他國語系的歌詞佔有 76.7%，而有使用英文的歌詞的比例達 22.8%，而使用日文或韓文的歌詞則佔不到 1%，參考圖 4-15。

含有英文字: 22.8%

含有日文字: 0.027%

含有韓文字: 0.028%

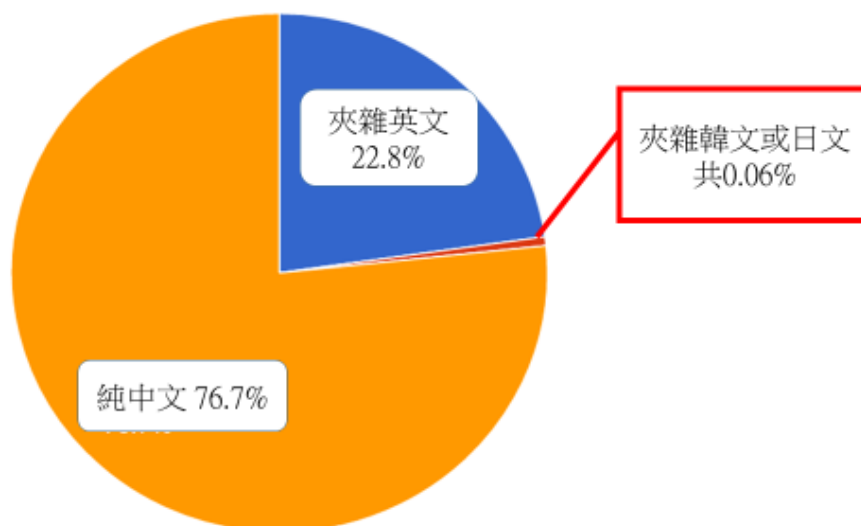


圖 4-15 使用雙語歌詞的歌詞比例比較

至於英文高頻詞，Top 10 中，主詞代名詞「You」與「I」，位居第一名和第二名，此與中文流行歌詞相同。不過，「he」則是未出現在前十名（為第 83 名），取而代之的是「baby」。「love」亦出現在 Top 10，可顯示中文流行歌詞中，會以外來語來表達對人、事、物的感情。文字雲如圖 14-16。

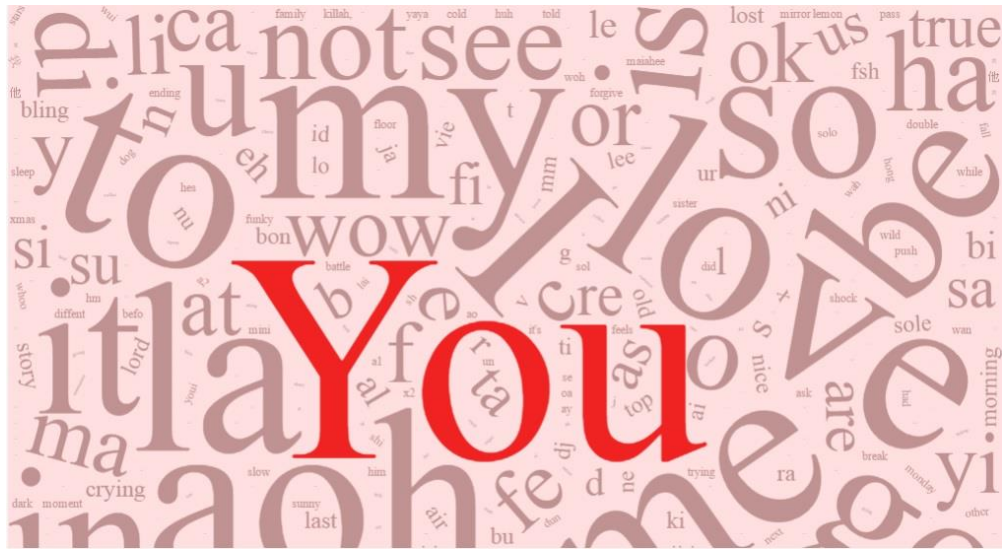


圖 4-16 歌詞中出現英文的高頻詞文字雲

4.5 舊曲新唱分析結果

135,030 首歌詞中，共有 15,896 首為舊曲新唱（翻唱），比例高達 11.77%。被翻唱 16 次（含）以上的歌詞列表如圖 4-17，表 4-7 為歌詞被翻唱超過 50 次的作詞者列表。

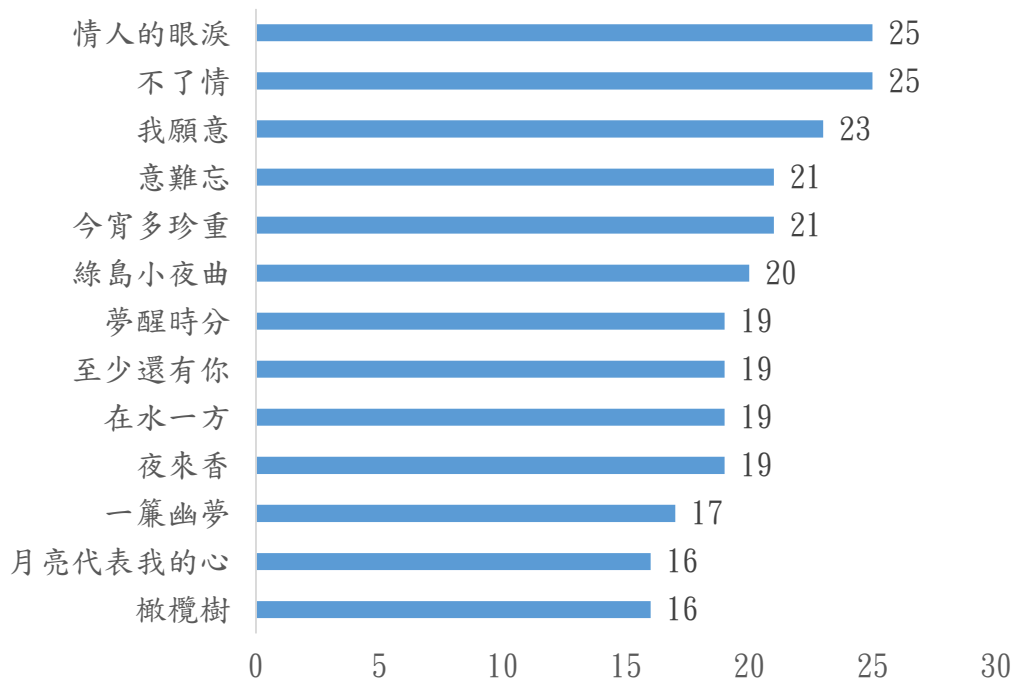


圖 4-17 被舊曲新唱 16 次以上的歌詞一覽

最常被翻唱的歌有兩首：情人的眼淚及不了情。情人的眼淚原唱是新加坡歌手潘秀琼，為 1955 年杏花溪之戀的片中歌曲，該首歌曾被齊秦、青山、余天、林承光、張學友、甄妮、蔡琴、美黛、林憶蓮及周蕙等人唱過。分析其歌詞，可發現如同本論文所分析的特性：該歌詞有押韻且使用換韻方法（ㄟ、ㄣ兩韻）；有使用疊字，例如「一顆顆眼淚」的一顆顆為疊字用法；為愛情風格的歌詞等。歌詞如下：

為什麼要對你掉眼淚 你難道不明白是為了愛
只有那有情人眼淚最珍貴
一顆顆眼淚都是愛都是愛

為什麼要對你掉眼淚 你難道不明白是為了愛
要不是有情人跟我要分開
我眼淚不會掉下來掉下來
好春常在 春花正開 你怎麼捨得說再會
我在深閨 望穿秋水 你不要忘了我情深 深如海

至於不了情部分，原唱是顧媚，為 1961 年同名電影的片中歌曲，該首歌曾被青山、羅時豐、余天、費玉清、鄧麗君、葉蒨文、潘越雲、高勝美、梅艷芳及周蕙等人唱過。分析歌詞，可發現如同本論文所分析的特性：大量使用類字和疊句，例如重覆使用「忘不了」；為愛情風格的歌詞；歌名為三字部等。歌詞如下：

忘不了 忘不了
忘不了你的錯 忘不了你的好
忘不了雨中的散步
也忘不了那風裡的擁抱

忘不了 忘不了
忘不了你的淚 忘不了你的笑
忘不了葉落的惆悵

也忘不了那花開的煩惱

寂寞的長巷
而今斜月清照
冷落的鞦韆
而今迎風輕搖
它重復你的叮嚀
一聲聲忘了 忘了
它低訴我的哀曲
一聲聲難了 難了
忘不了 忘不了
忘不了春已盡 忘不了花已老
忘不了離別的滋味
也忘不了那相思的苦惱

表 4-7 被翻唱的歌詞 超過 50 首的作詞者列表

作詞者	被翻唱的歌詞數量
林夕	339
莊奴	112
許常德	73
姚若龍	73
姚謙	70
李宗盛	70
慎芝	58

歌詞被翻唱超過 50 次的作詞者中，林夕最常被翻唱的歌詞是「至少還有你」，共被 19 位演唱者翻唱過，包括巫啟賢、張敬軒、謝霆鋒及周華健等人；莊奴最常被翻唱的歌詞是「煙雨斜陽」，分別由劉家昌、費玉清、李茂山、徐小鳳等 14 位歌手演唱過；許常德最常被翻唱的歌詞是「愛我的人和我愛的人」，分別由巫啟賢、游鴻明、迪克牛仔等 7 位歌手演唱。

4.6 風格分類結果與分析

實驗使用的 SVM 分類器，係採用台灣大學林智仁教授開發的 LIBSVM 之 open source 工具。

訓練資料來源包含吳嫩婉《台灣國語流行歌曲的修辭藝術〈1949-2000〉》的歌詞分析資料、馬占山《臺灣流行歌詞主題類型與語言表達研究（1999~2008）——以「Hit-FM 年度百首單曲」為對象》的歌詞分析資料及九酷音樂[24]、酷我音樂[25]的分類歌詞，共 3,771 首。

風格類型部分，依吳嫩婉《台灣國語流行歌曲的修辭藝術〈1949-2000〉》的歌詞分析資料，有風格 9 類，包括「愛情、生活哲理（勵志）、友情、親情、人生百態、田野風光、社會批判、愛國、鄉情」，但因人生百態、田野風光、社會批判、愛國及鄉情的歌詞數量較少，如圖 4-18，導致無法以 SVM 訓練正確的模型，因此，本實驗改由階層式作法，作法如下：

- (1) 將訓練資料分為愛情與非愛情的模型，以 5-fold 交叉驗證，準確率為 85.39%。
- (2) 非愛情部分，分為親情和勵志兩類，以 5-fold 交叉驗證，準確率為 95%。
- (3) 非愛情部分，加入友情，所以共有三類：親情、友情和勵志，以 5-fold 交叉驗證，準確率為 70.8%。
- (4) 將「人生百態、田野風光、社會批判、愛國、鄉情」定為「其他類」，並加入非愛情部分，所以非愛情部分共有四類：親情、友情、勵志和其他，以 5-fold 交叉驗證，準確率為 67.23%。

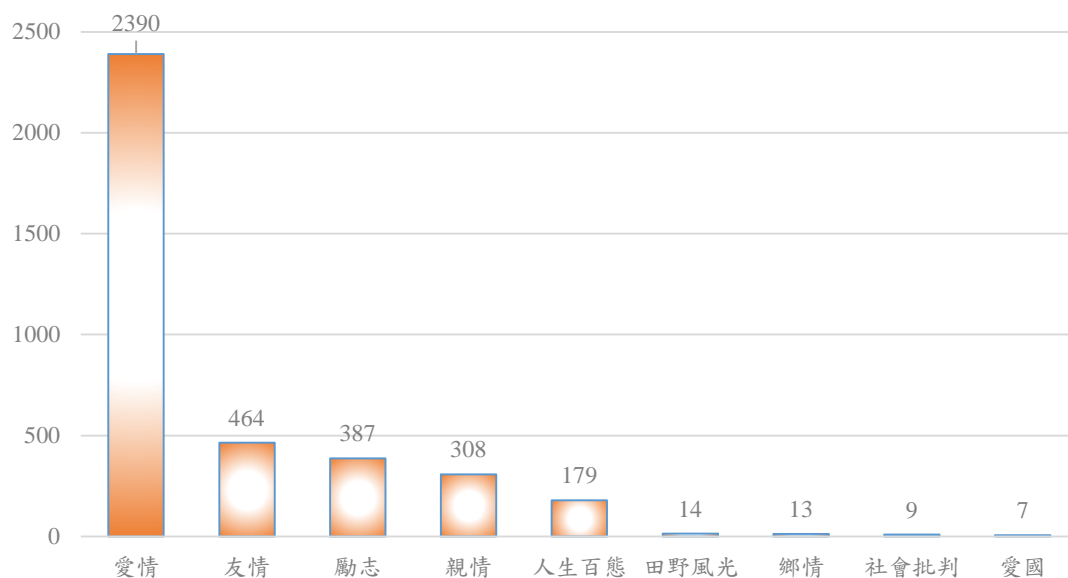


圖 4-18 訓練資料風格類型數量比較

然而，非愛情部分的風格準確率未達七成，因此本實驗分別以兩種的詞總集合方式產生向量陣列，並建立模型，以提升準確率，如表 4-8：

表 4-8 非愛情的詞總集合方式及準確率

詞總集合方式	準確率
訓練歌詞的斷詞，計數達 100 次以上的詞彙	67.23%
訓練歌詞的斷詞，計數達 100 次以上的詞彙， 但僅保留形容詞	70.21%

最後，將待預測的歌詞資料 88,648 筆進入第一層分類，即分成愛情及非愛情兩類，結果為愛情有 54,492 首，非愛情為 34,156 首。接著，將非愛情類的歌詞進入第二層分類，結果為勵志類 25,131 首、友情類 3,842 首、親情類 37 首、其他類 5,145 首。整體結果如表 4-9。

表 4-9 風格分類結果

風格分類	歌詞數量	比例
愛情	54,492	61.47%
勵志	25,131	28.35%
友情	3,842	4.33%
親情	37	0.04%
其他	5,145	5.80%

4.7 押韻分析結果

讓歌詞押韻能讓增加其易唱性及豐富歌曲的音樂性，例如根據曾慧佳在《從流行歌曲看臺灣社會》所提及：

流行歌曲為求易學、易唱，並且在只能用極少的字數來陳述一個情境的限制下，押韻與類似詩的意境，似乎是歷年來歌詞創作者，至今不敢放棄的金科玉律……。

本論文研究針對一韻到底、換韻、無韻、句首韻進行分析。

一韻到底：

88,648 筆中，一韻到底比重為 7.5%，其中，常用字為「你、了、的、子、來」；最常用的韻腳為一韻，佔一韻到底的 32.7%，常用字為「你、裡」，其次為ㄣ韻，佔一韻到底的 10.5%，常用字為「愛、來、開」。參考圖 4-19。

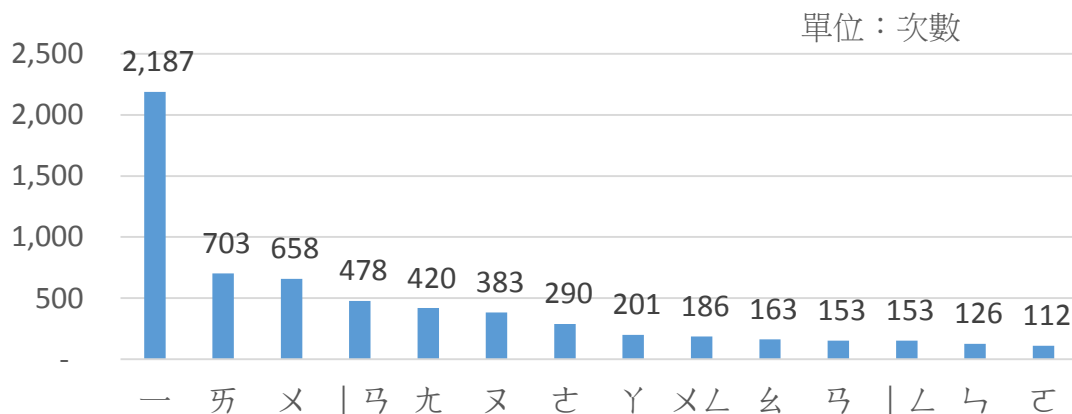


圖 4-19 一韻到底的韻腳一覽 (僅列出現 100 次以上的韻腳)

換韻：

88,648 筆中，換韻比重為 88.6%，為最常使用的押韻方式。其中，最常使用的韻腳為 | ㄋ，常用字為「年、天、見」等字；韻腳為 ㄨ 的常用字為「手、頭、後」等字。參考圖 4-20。

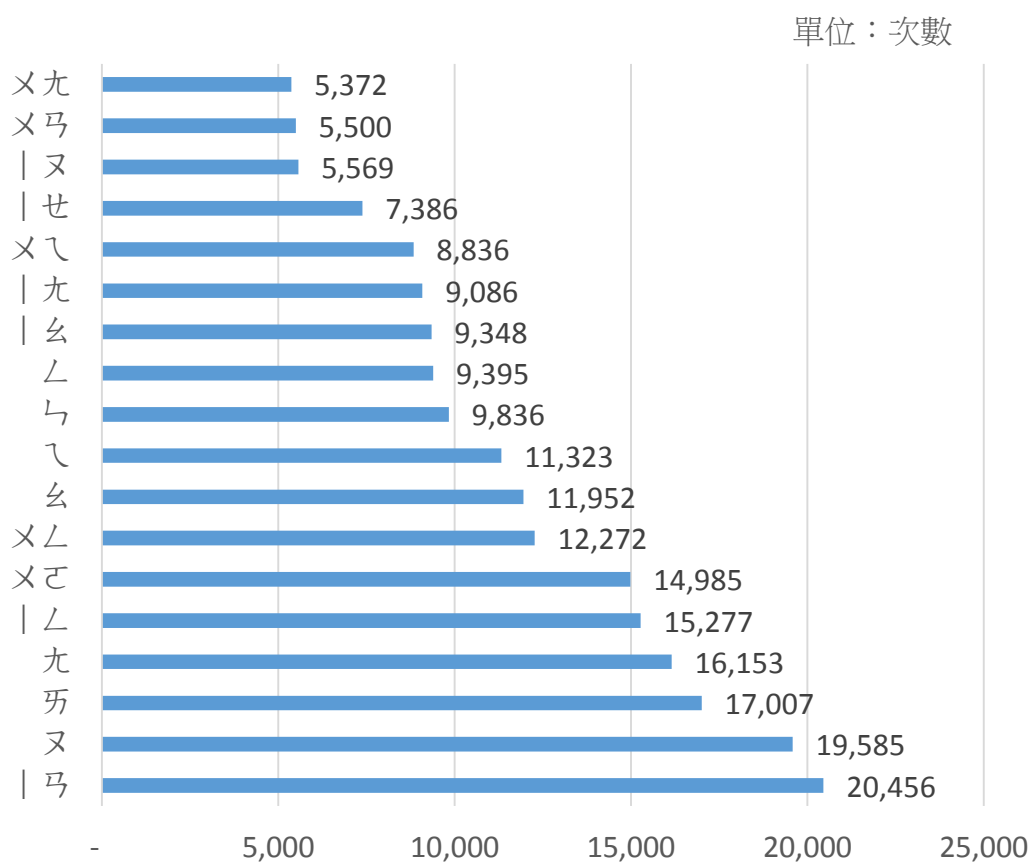


圖 4-20 換韻的韻腳一覽 (僅顯示超過 5,000 次以上)

無韻：

88,648 筆中，無韻比重為 3.9%，為最不常使用的押韻方式。

句首韻：

88,648 筆中，有使用句首韻的比重為 89.7%，可見句首韻為非常常使用的押韻方式，其中，最常使用的韻腳為 ㄟ，如圖 4-21 句首韻的韻腳一覽，最常使用韻腳字為「愛、在、

帶」等字。

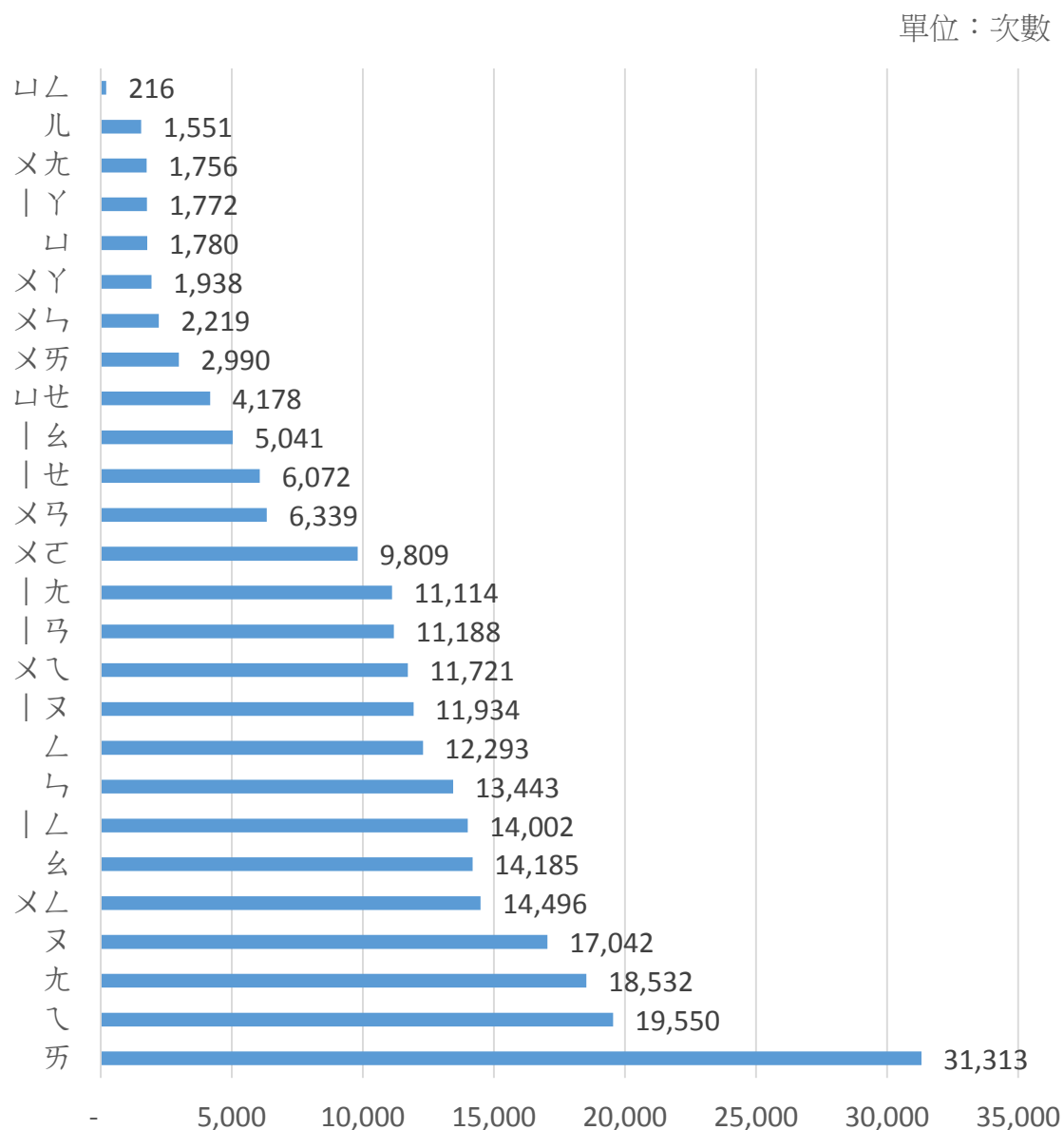


圖 4-21 句首韻的韻腳一覽 (僅列出超過 100 次以上的韻腳)

4.8 修辭分析結果：明喻、類字、疊字、疊句及對偶

修辭就好比化妝方式，高超的化妝技巧，可突顯該人的優點，並隱藏不足之處，在歌詞當中，修辭方式使用得當，除使聽眾容易了解歌詞想傳達的內容，更可產生共鳴。

本研究針對明喻、類疊及對偶進行實驗及分析，其中類疊法包括：類字、疊字、疊

句三種方式。

4.8.1 明喻

巧譬妙喻可以使文字生動、意象鮮明，讓讀者反覆咀嚼、認識新事物，而獲致無窮滋味；抽象事物以實例譬喻，更能讓人透徹理解，甚至引發讀者聯想。

根據研究方法分析，本研究的喻詞列表、「如」的負面表列、「若」的正面表列如下：

- 喻詞列表：「如同，猶如，好像，恰似，好似，宛如，彷彿，類似，似乎，似，相像，相類，類似，好比，如，如同，好像，若，像，相似，髣髴，近似，形似，相仿，相像，相似，類似，雷同」。
- 「如」的負面表列（括弧內為詞性）：「如果(C)，如(P)，如此(ADV)，如今(N)，如准(Vt)，如何(Vi)，如何(ADV)，如來佛(N)，姜怡如(N)，蘭相如(N)，果真如此(Vi)，如此(Vi)，如此費(N)，如願以償(Vi)，如實(ADV)，無如(ADV)，晨如(N)，如哇(Vi)，甲如忠(N)，古秀如(N)，如何路(N)，如期(ADV)，如過唔(Vi)，不如人意者(N)，馬如龍(N)，還不如(Vt)，如此路(N)，倒不如(ADV)，如今花(N)，不如人意(Vi)，命該如此(Vi)，假如(C)，如約(ADV)，黃倩如(N)，魏如萱(N)，原來如此(Vi)，不如歸去(Vi)，自如(Vi)，一如(P)，如一(Vi)，大不如前(Vi)，柳如(N)，如是(Vi)，如意(Vi)，如何是好(Vi)，早知如此(Vi)，突如其來(A)，如願(Vi)，如果(C)，意下如何(Vi)，石如箭(N)，清簡如(N)，如意算盤(N)，咀臉如(N)，如果說(C)，張如城(N)，如來心(N)，王欣如(N)，還不如(C)，顏如玉(N)，不如(C)，如此(ADV)，如廁(Vi)，不如(Vt)，東如(N)，藍如玉(N)，蕙如(N)，無論如何(ADV)，吳淡如(N)，徐如林(N)」
- 「若」的正面表列（括弧內為詞性）：「猶若(P)，婉若(N)，若有其事(Vi)，若即若離(Vi)，固若金湯(Vi)，門庭若市(Vi)，若愚(N)，旁若無人(Vi)，悵然若失(Vi)，判若兩人(Vi)，仿若(Vt)，若有所思(Vi)，若有所失(Vi)，冷若冰霜(Vi)，恍若(Vt)，若無其事(Vi)，風若靜雨(N)，情若夢(N)，

依據類字所使用的高頻詞，與「我、你」相關的即佔了前四名，包括「我的、你的、我們、我不」，顯示使用人稱的類字方法更可渲染氣氛。舉例如下：

W.S.S.(草莓救星)的「自在」：

我的腳趾在等待自由

我的手指在等待自由

我的冰箱在等待自由

我的衣服在等待自由

原諒我不能擁有自由

我的床單在等待自由

我的男朋友也在等待和我的自由

我不自在

我不自在

這種感覺你們男人怎麼會明白

「我的」表達出一種對自由的渴望，接著再以「這種感覺你們男人怎麼會明白」渲染氣氛，也表現了敘述者心中對「不可能之可能」深切的期望與希冀。

4.8.3 疊字

88,648 首歌詞中，有使用疊字的歌詞佔 63.6%，高頻詞如圖 4-23。

疊字的高頻詞多為副詞，例如：慢慢、輕輕、深深、默默等，使用此方法，則可使副詞更具有強化或淡化的效果，舉例如下：

例一：龍軍的「來不及說對不起」

我的天空慢慢失去色彩

不能再見我最愛的你

「慢慢」可委婉地表現失去的感覺，接著再以「不能再見我最愛的你」渲染氣氛，以表現出敘述者的心中的落寞感。

例二：韓志堅的「思念的痛」

只能輕輕祈禱你來入夢

[illegible]

圖 4-23 疊字高頻詞文字雲

在下雨天 空氣特別的新鮮

敘述者因想要鼓勵對方不管流言蜚語，僅管努力往前，以「向前 向前」、「一步一步」表達出走向前的熱烈期待與陪伴。

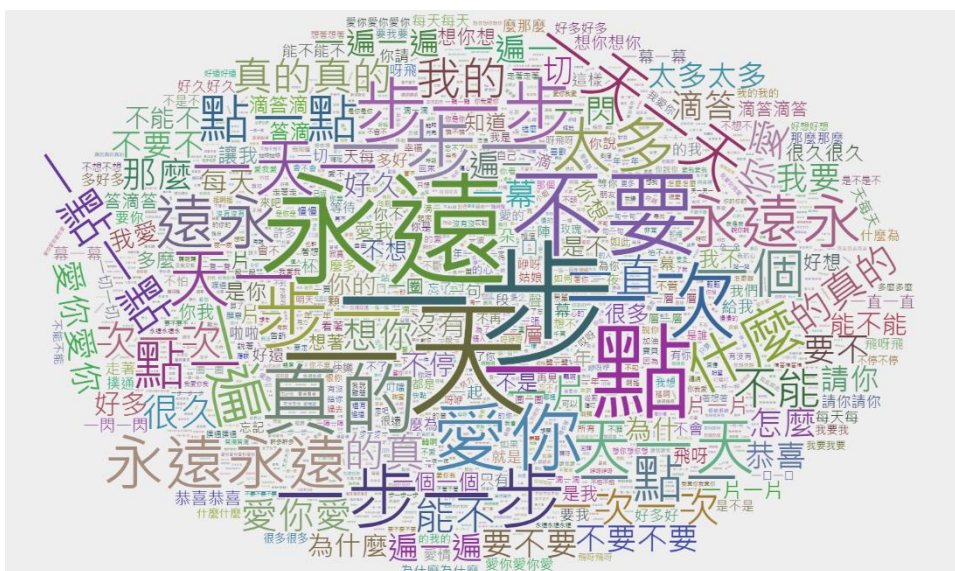


圖 4-24 疊句高頻詞文字雲

歌詞要易記易唱，要琅琅上口，因此，在不少作品中，刻意融入淺易的詩詞或對偶，可令聽眾一聽難忘。

月朦朧
鳥朦朧
螢火照夜空
山朦朧

樹朦朧

秋蟲在呢嚟

花朦朧

夜朦朧

晚風叩簾籠

燈朦朧

人朦朧

但願同入夢慢慢失去色彩

不能再見我最愛的你

例二：

十二恨/彭資閔的「革命」：

狼煙(N) 燒焦(Vt) 了(ASP) 天空(N)

鮮血(N) 染痛(Vt) 了(ASP) 湖泊(N)

至於作詞者使用對偶修辭的比例情形，參考圖 4-25。使用對偶的歌詞數量佔其整體作品數量比例，最高的是阿弟仔，雖其作品不多，僅 116 首，然其使用對偶修辭的歌詞數量佔其整體作品數量比例卻高達 65.5%，第二名為五月天的阿信，使用比例達 60.2%；林夕雖為詞曲之王，作品高達 1,138 首，然使用對偶的歌詞數量只佔其整體作品數量的 33.3%。

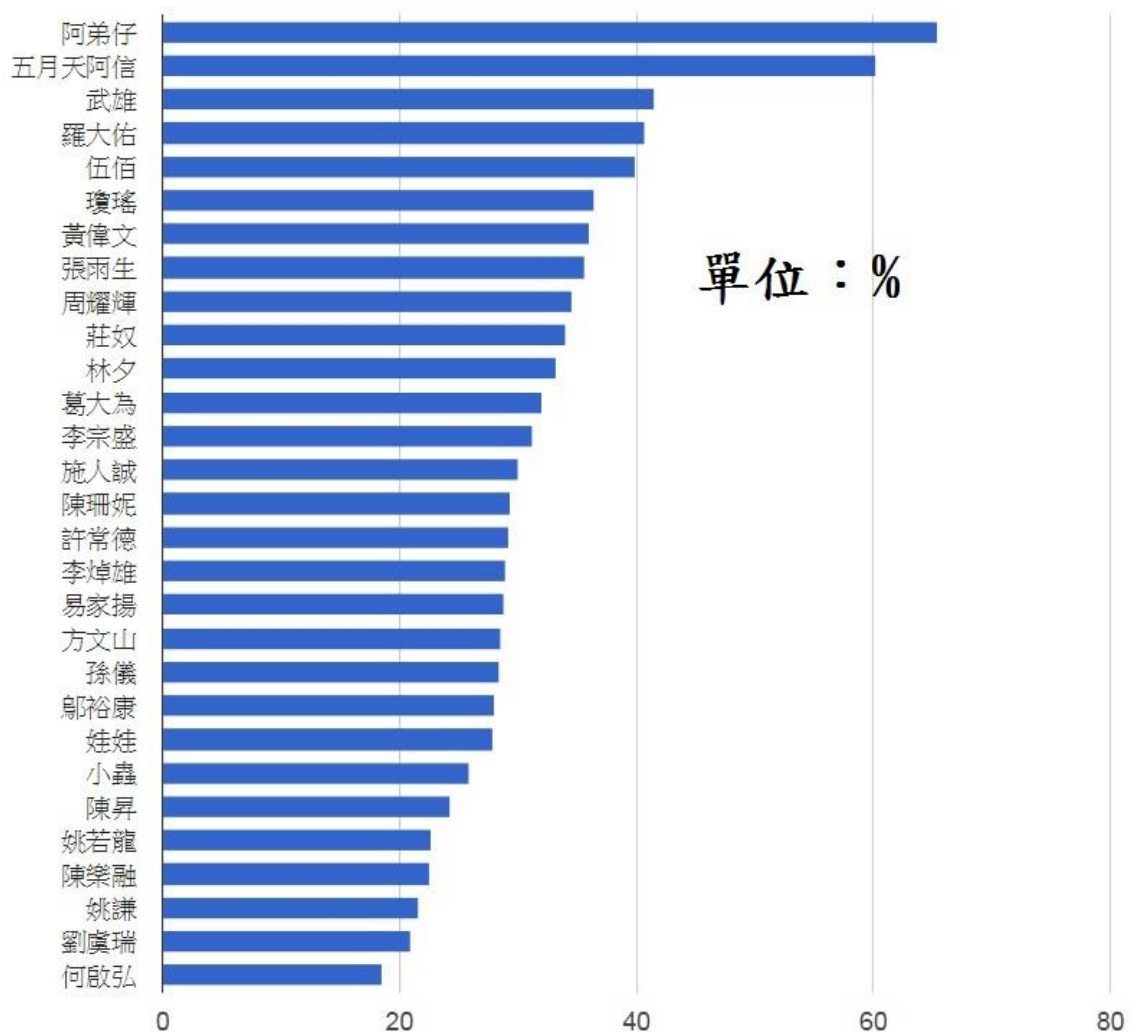


圖 4-25 作詞者使用對偶修詞的歌詞數量佔其整體作品數量之比例

4.9 作詞者的詞頻與詞彙相鄰分析：方文山、李宗盛、林夕

分析作詞者的詞頻，可快速掌握作詞者常用的詞彙，並了解其風格。

4.9.1 詞頻

本實驗以中央研究院的中文斷詞與剖析系統，將各作詞者的歌詞斷詞後，依照其詞性寫成檔案，接著將檔案匯入簡冠庭所製作的 HTML5 Word Cloud 工具，以進行各詞計數及繪製文字雲。參考表 4-10 方文山、李宗盛及林夕的高頻詞彙用詞一覽表，及圖 4-26~ 圖 4-37 的方文山、李宗盛及林夕的高頻詞文字雲圖。

表 4-10 方文山、李宗盛及林夕的高頻詞彙用詞一覽表

作詞者	方文山	李宗盛	林夕
作詞數量	430	400	1163
名詞	故事 感覺 回憶 畫面 最後 離開 開始 安靜 慢慢 微笑 錯過 回憶 等待	女人 過去 人生 現在 時候	個人 感情 眼淚 朋友
動詞		以為 在乎 願意 面對 覺得	需要 一樣 擁抱 忘記 得到
形容詞	沉默 遙遠 成熟	容易 努力 纏綿 不同 孤單	浪漫 重要 愉快



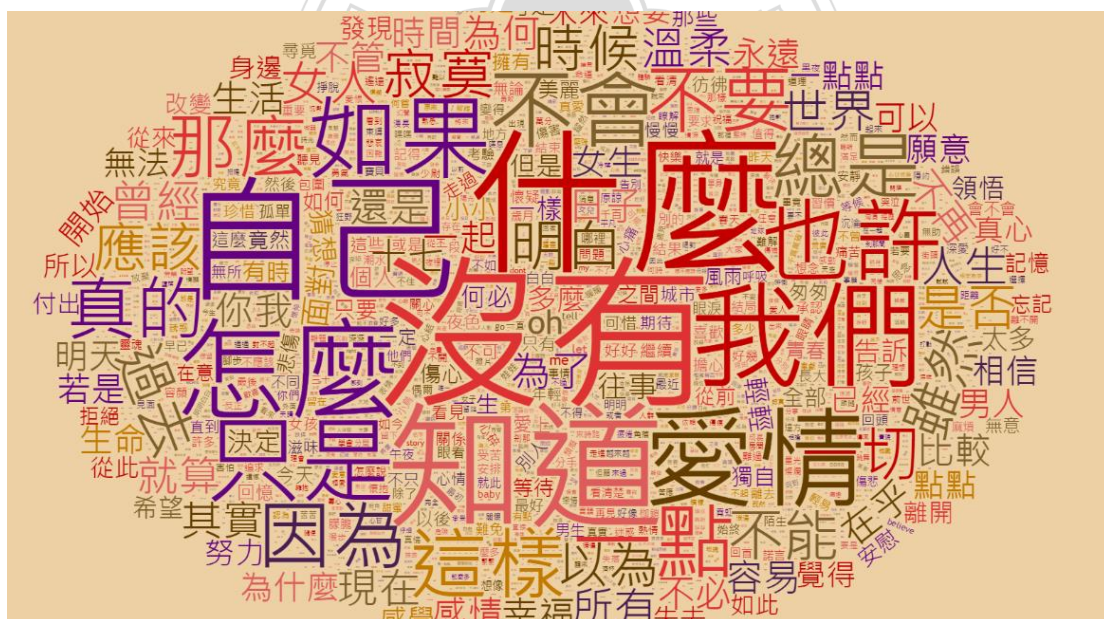
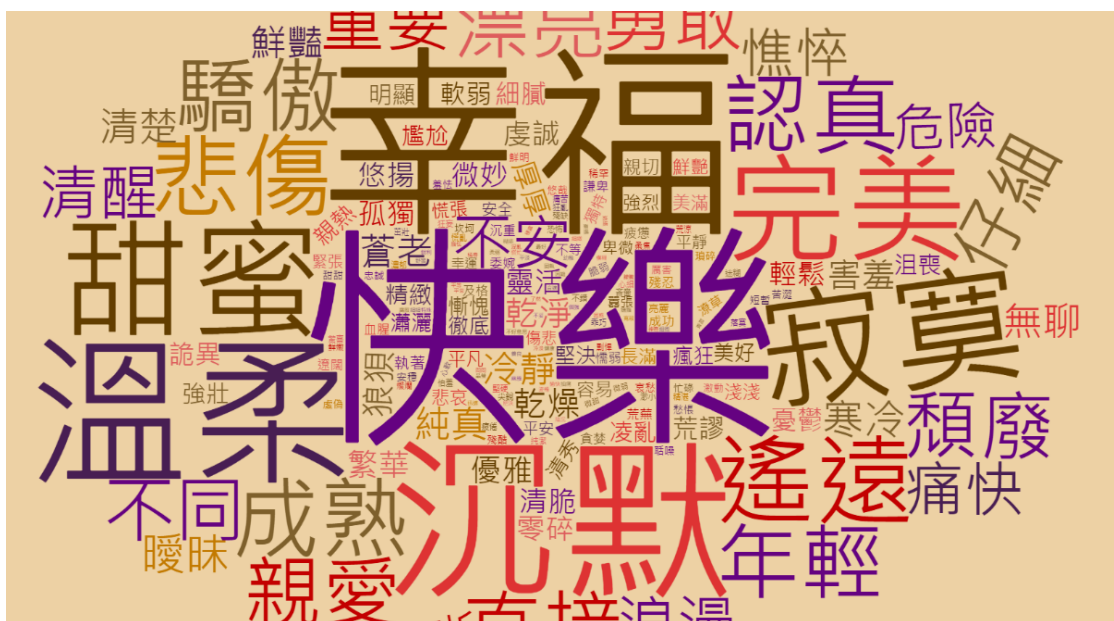
圖 4-26 方文山全部用詞高頻詞的文字雲



圖 4-27 方文山的名詞高頻詞的文字雲



圖 4-28 方文山的動詞高頻詞的文字雲



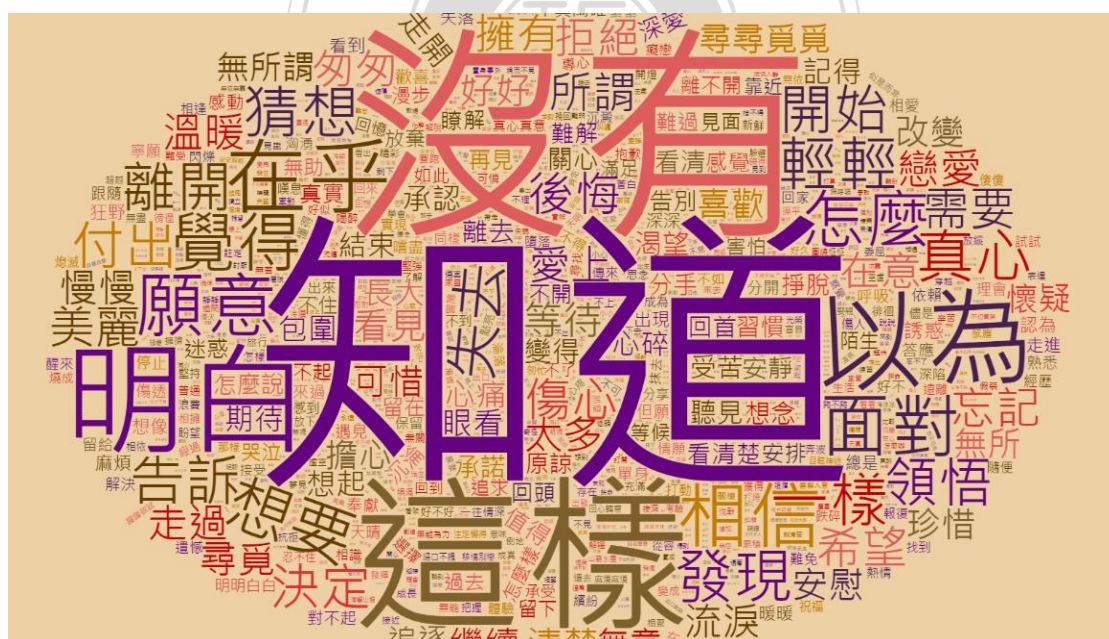
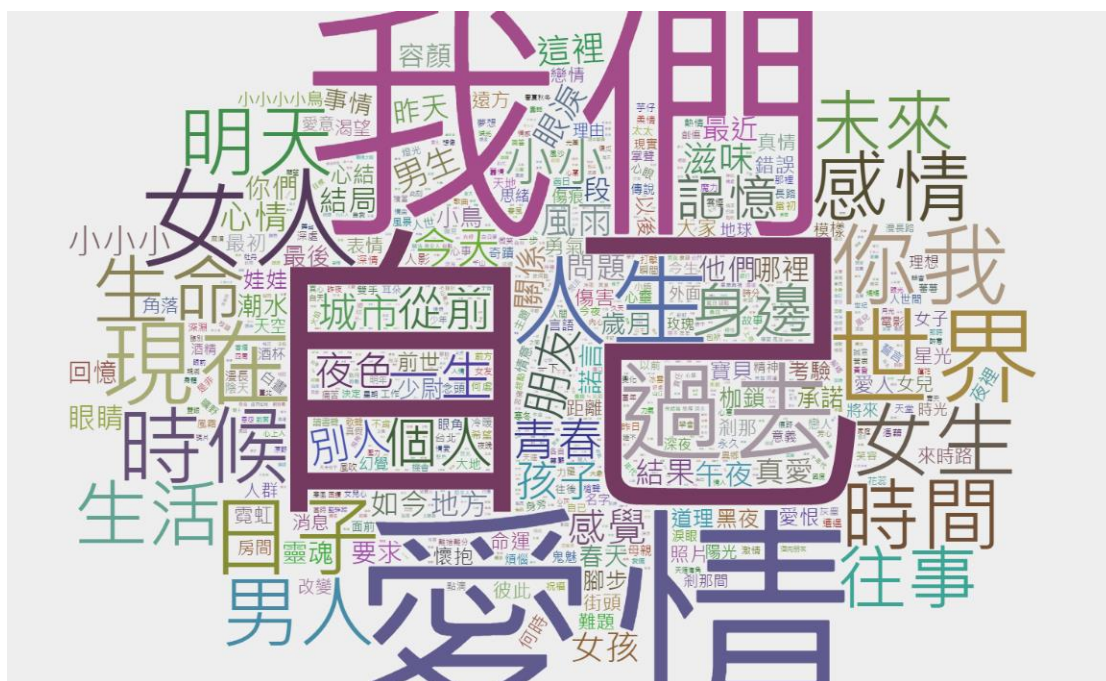






圖 4-38~圖 4-40 為李宗盛、林夕及方文山詞彙相鄰的文字雲，僅列出使用 100 次以上的詞彙。



圖 4-38 李宗盛的詞彙相鄰文字雲

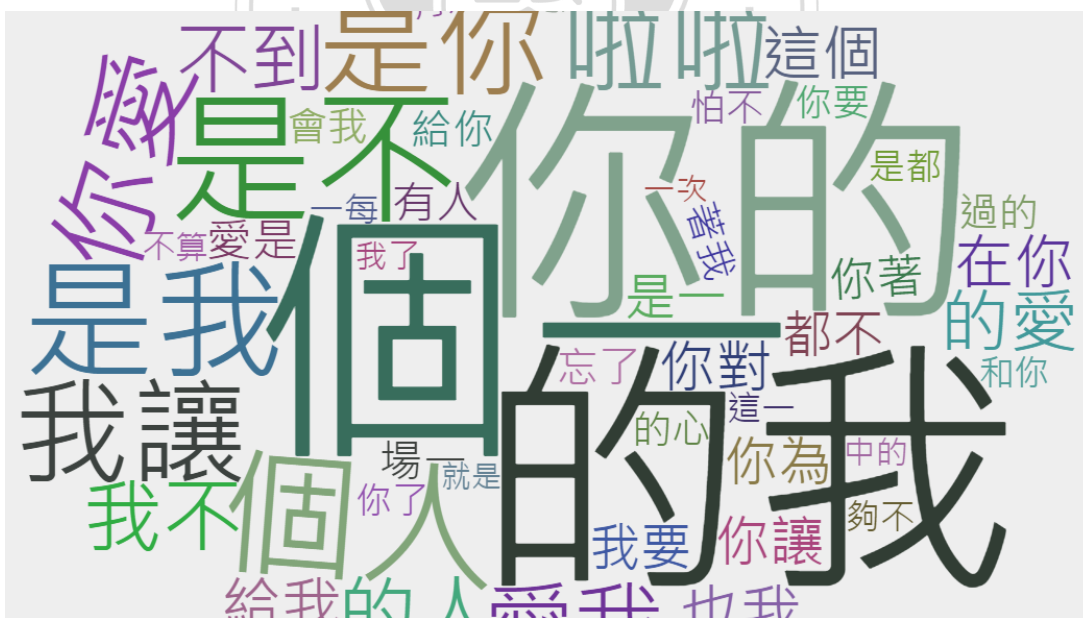


圖 4-39 林夕的詞彙相鄰文字雲

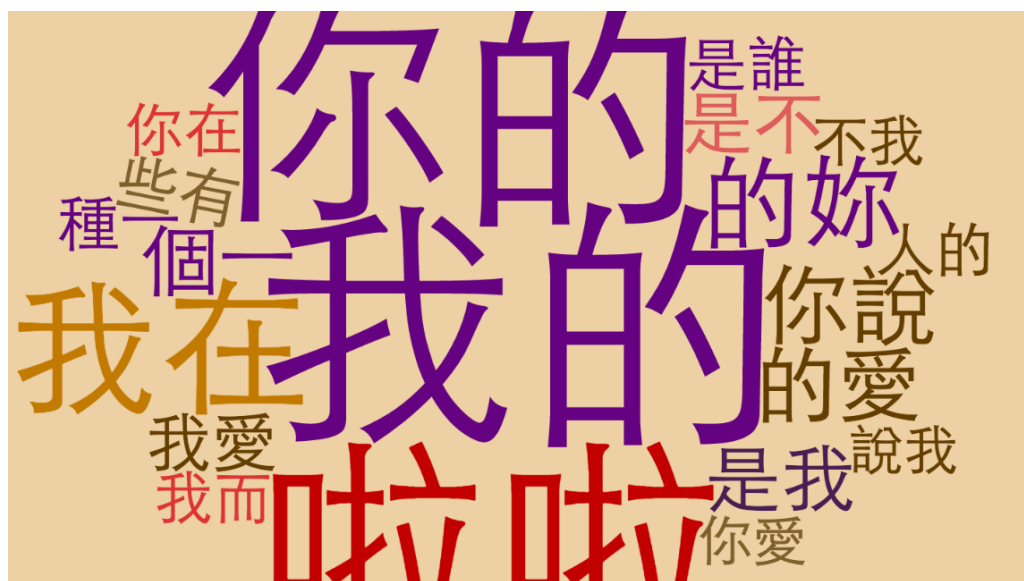


圖 4-40 方文山的詞彙相鄰文字雲

4.10 年代對使用語系的交叉分析

根據4.4使用語系分析研究結果，得知使用英文在華語流行音樂歌詞的比例達22.8%，而使用日文或韓文的歌詞佔不到1%，因此，本章節將以此為基礎，進而分析1965~2015歷年使用英文歌詞的比例。圖4-41為1965~2015歷年使用英文歌詞的比例趨勢圖。



圖 4-41 1965~2015 歷年使用英文歌詞的比例趨勢圖

由圖 4-41 可得知早在 1966 年就開始有夾雜英文歌詞的現象，1974 年則突然大幅增加，然 1975~1990 年則又突然減少，直至 1991 年才又逐年增加。

然而，1966~1974 年魔鏡歌詞網所獲得的歌詞僅 1,546 首，而有英文夾雜的僅約 120 首，無法就此斷定歌詞夾雜英文的情況大幅盛行，且根據華語流行音樂歌詞相關研究，較少解釋這段時間有受到西方文化影響的情況，但這部分仍有繼續深入研究的價值。

至於 1991 年後，歌詞夾雜英文情況大幅增加，且逐年提升，至 2015 年，該年歌詞夾雜英文的比例達 36%，此符合過去華語流行歌詞相關研究的結果。例如根據曾慧佳的《從流行歌曲看台灣社會》及謝櫻子的方文山華語詞作主題研究：「華語流行歌曲的全球化時期為 1990—2000 年代，1990 年後，台灣意識抬頭，流行音樂界新人輩出，團隊型歌手興起，外來歌手增加，流行歌曲時興「雙語」，甚至華語、台語、英語等多語交雜的曲風，L.A.Boys 則是這波風潮的先驅。」

至於使用中韓文交雜的部分，由於韓流自 2002 年開始持續在亞洲升溫，舉凡韓劇、韓國流行音樂，均受亞洲地區觀眾喜愛，2010 年開始，韓文歌詞開始夾雜於華語流行音樂，雖數量不多，但仍顯示韓流仍有些許影響華語流行歌詞。圖 4-42 為 2008~2015 歷年使用韓文歌詞的比例趨勢圖。

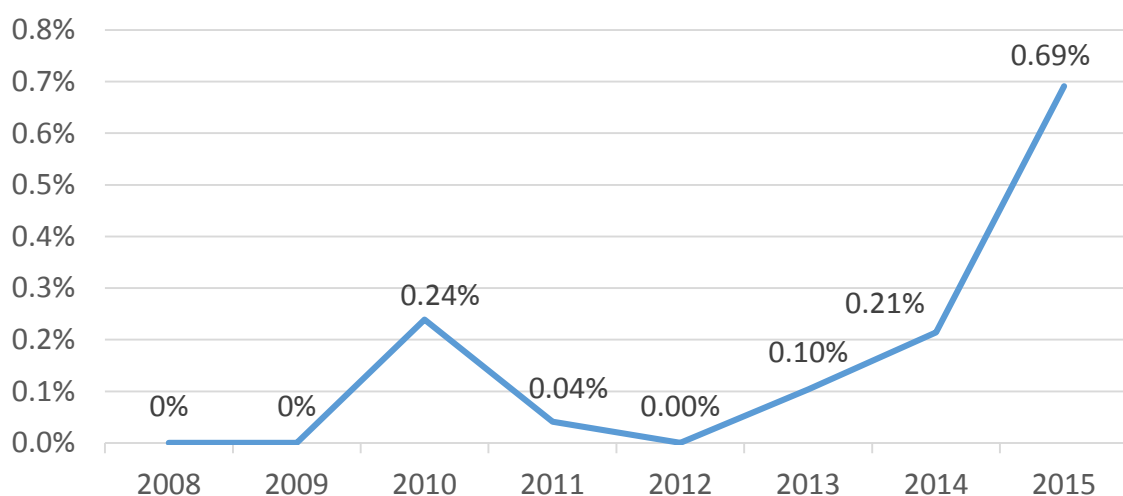


圖 4-42 2008~2015 歷年使用韓文歌詞的比例趨勢圖

4.11 年代對作詞者作品數量的交叉分析：方文山、李宗盛及林夕

本研究資料為方文山 393 首、李宗盛 31 首及林夕 1069 首的作品，將其與年代進行交叉分析，參考圖 4-43~圖 4-45。由圖表可看出，李宗盛最早於 1982 年就開始在市場上有華語流行音樂歌詞作品，林夕則為 1987 年，而林夕每年作品非常多，2003 年曾高達 93 首，此還不含粵語等作品，可見林夕不愧為作詞之王，至於方文山則曾於 2005 年作品量高達 50 首。

而這三位作詞家作品數量最多均為 2000 年~2010 年，2013 年後，則成遞減現象。

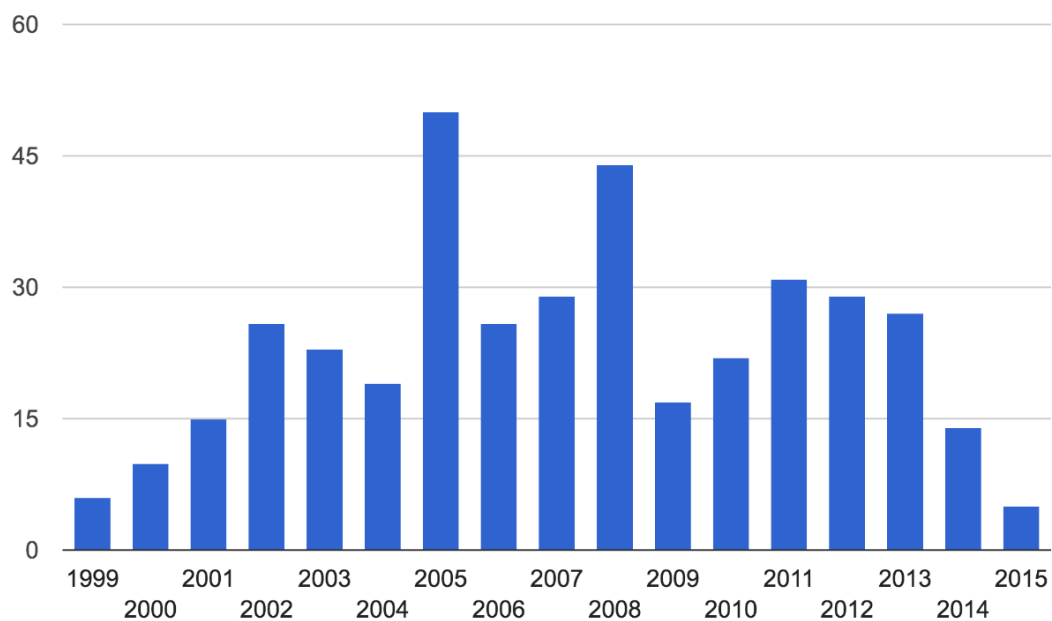


圖 4-43 方文山歷年作品數量

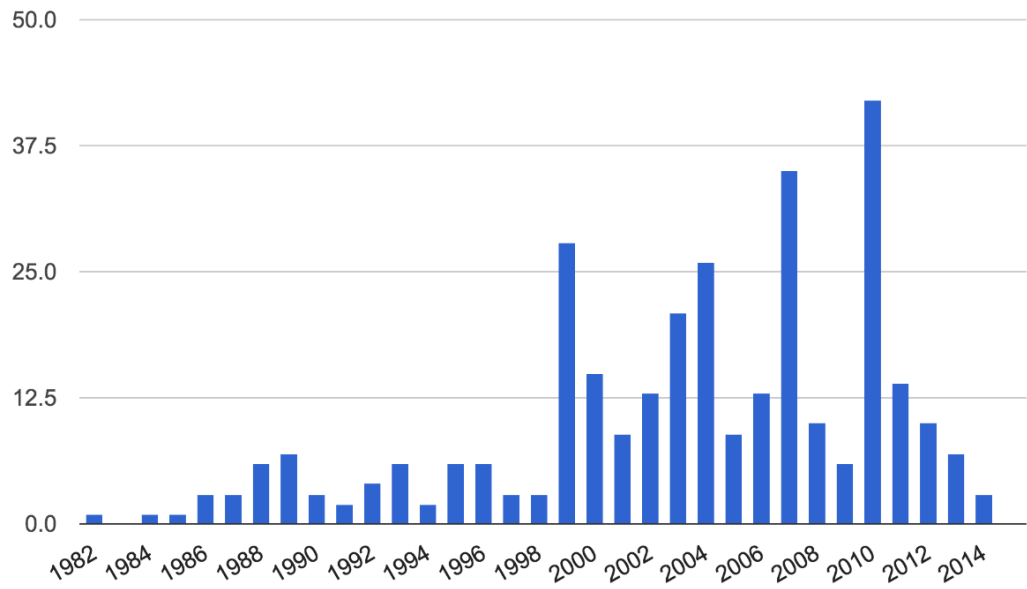


圖 4-44 李宗盛歷年作品數量

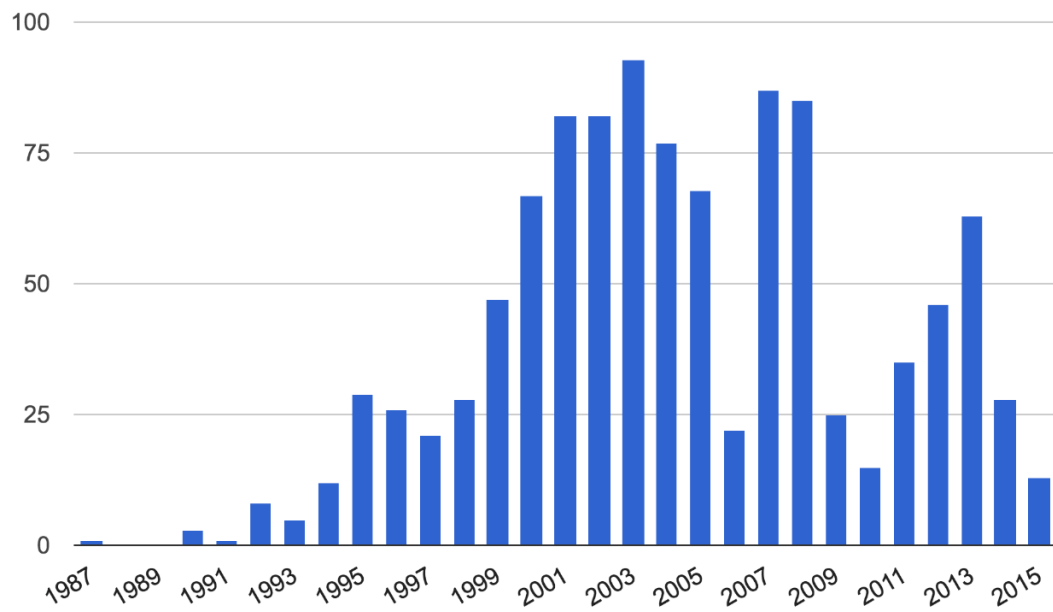


圖 4-45 林夕歷年作品數量

4.12 年代對押韻的交叉分析

1960~2015 年，以每十年為單位，取各年代前十名最常使用的韻腳分部，發現各個時期均以一韻為最常使用的韻腳，其次為尤，至於丫韻則為 1960~1979 年常使用的韻腳，但 1980 年後就未列入前十名常使用的韻腳，取而代之的是又韻；此外，乚韻為 2000 年以前未出現過的 Top 10 韻腳，於 2000~2015 年才開始進入前十名常使用的韻腳名單內，且比例為前十名最常使用的韻腳的 9~10%。參考表 4-12。

𠂔和𠂔韻則為各個年代均有使用的韻，且𠂔韻腳比例在各時期均很呈現很平均的狀態，可能與常用字「愛、在、待」等字常使用有關。

表 4-12 1960~2015 各時期使用的 Top 10 韻腳

	1960~1969	1970~1979	1980~1989	1990~2000	2000~2009	2010~2015
■ ㄛ	0%	0%	0%	0%	9%	10%
■ ㄥ	0%	0%	0%	8%	7%	9%
■ ㄨ ㄛ	0%	0%	7%	8%	0%	0%
■ ㄛ	0%	0%	9%	0%	0%	10%
■ ㄨ	0%	0%	10%	11%	10%	10%
■ ㄛ	0%	7%	9%	9%	8%	9%
■ ㄨ	0%	7%	0%	0%	0%	0%
■ ㄛ	0%	8%	7%	8%	9%	9%
■ ㄥ	5%	8%	10%	9%	9%	9%
■ ㄥ	6%	0%	0%	0%	0%	0%
■ ㄨ	7%	0%	0%	0%	0%	0%
■ ㄨ	7%	0%	0%	0%	0%	0%
■ ㄛ	8%	8%	0%	10%	11%	0%
■ ㄥ	9%	8%	9%	0%	0%	0%
■ 𠂔	9%	7%	7%	8%	9%	9%
■ 丫	12%	10%	0%	0%	0%	0%
■ 尤	14%	12%	8%	8%	8%	8%
■ 一	22%	26%	24%	22%	19%	18%

4.13 年代對風格的交叉分析

本研究針對年代及愛情 49,961 首、勵志 22,250 首、友情 2,954 首、親情 30 首及其他 4,567 首進行交叉分析。

愛情風格的每年作品數量佔該年作品總數量 40% 以上，然 1956 年及 1964 由於透過魔鏡歌詞網所獲得的資料僅 1 首和 4 首，所以不列入比較範圍。至於 1975 年~1983 年，有可能正好是民歌時代崛起，所以整體比例降至 40% 以下，1979 年甚至僅有 25.2% 的歌詞為愛情風格，參考圖 4-46 及圖 4-47。

根據馬占山，臺灣流行歌詞主題類型與語言表達研究（1999~2008）——以「Hit-FM 年度百首單曲」為對象，針對民歌時期背景及風格的敘述，節錄如下：

1971 年我國被迫退出聯合國，再加上釣魚台的歸屬問題和引發的保釣運動，催化了當時臺灣年輕人對自我定位的認同疑惑，民族認同的問題成為文化運動的焦點。西洋歌曲、日本歌謠、海港派歌詞，此時成了無病呻吟、無益民生的歌詞，被知識分子批評為「靡靡之音」。政府於 1976 年推行「淨化歌曲運動」，希望透過文化上的重整，塑造一個團結安定的社會，企圖淨化瀰漫於 1960 年代日趨奢靡的國語流行歌曲。

1973 年，楊弦在胡德夫的演唱會上發表了以余光中的詩所譜成的歌曲〈鄉愁四韻〉，「民歌運動」於焉展開。1975 年 6 月，楊弦在臺北中山堂舉行了「現代民謠創作演唱會」，掀起了民歌運動的序幕，此期標榜著自己的創作、唱自己的歌，歌曲多具風雅與書卷氣味，較少商業氣息，形成此時期民歌的特色，為國語歌壇帶來清新淳樸的新感受。同年楊弦出版首張專輯《中國現代民歌集》引起熱烈迴響。

另外，節錄謝櫻子的方文山華語詞作主題研究的比較表，如表 4-13，可得知 1970 時期的民歌風格及資訊。

表 4-13 1960 年代華語流行歌曲與 1970 初期現代民歌的比較

（整理自謝櫻子的方文山華語詞作主題研究）

	1960 年代華語流行歌曲	1970 初期現代民歌
創作者	專業詞曲作者	知識青年
演唱者	明星藝人	知識青年
消費階層	成年人	青年學生
伴奏配器	樂隊繁複編曲	吉他簡易伴奏
歌詞內涵	愛情的悲歡離合	民族情緒或年輕人心聲
傳播管道	廣播、電視等媒體	西餐廳、校園
評價	靡靡之音	清新純樸

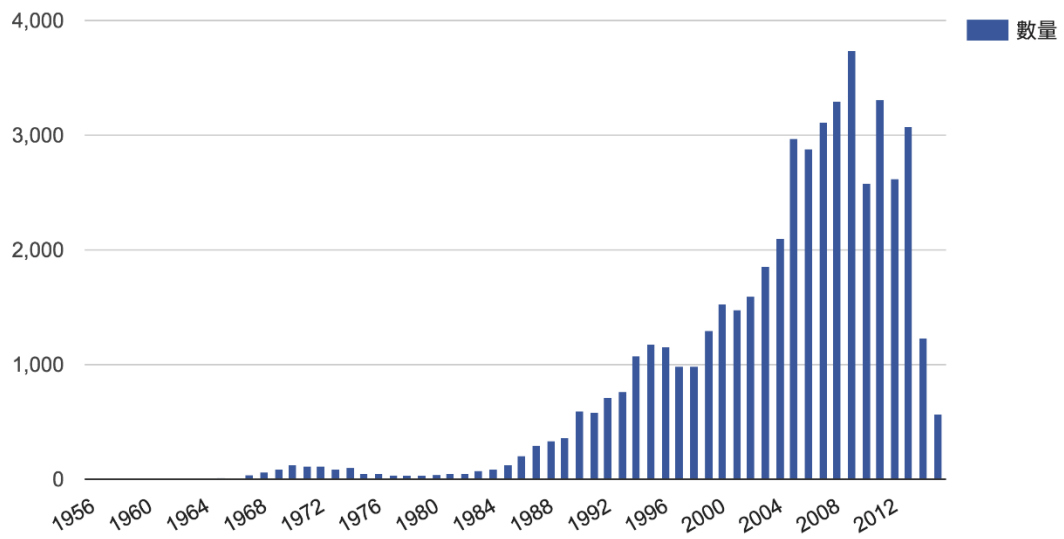


圖 4-46 愛情風格歷年作品數量

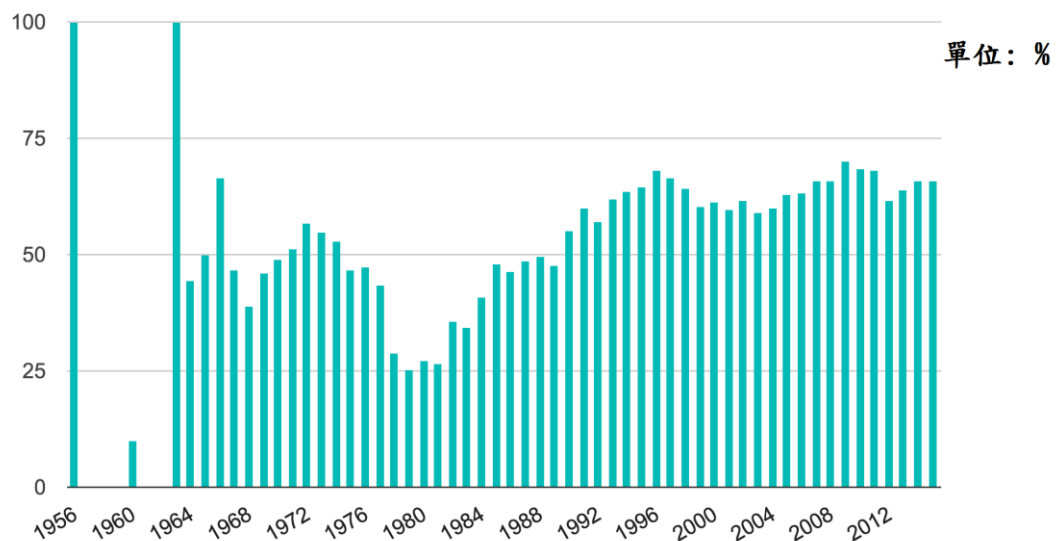


圖 4-47 愛情風格佔該年歌詞數量的比例

勵志風格的每年作品數量佔該年總數量 27~46%，1979~1989 年十年間，曾 7 年超過 40% 以上，接著之後的每年也維持在 26%~29%，可見勵志風格的歌詞為除愛情風格外，最多的風格，參考圖 4-48 及圖 4-49。

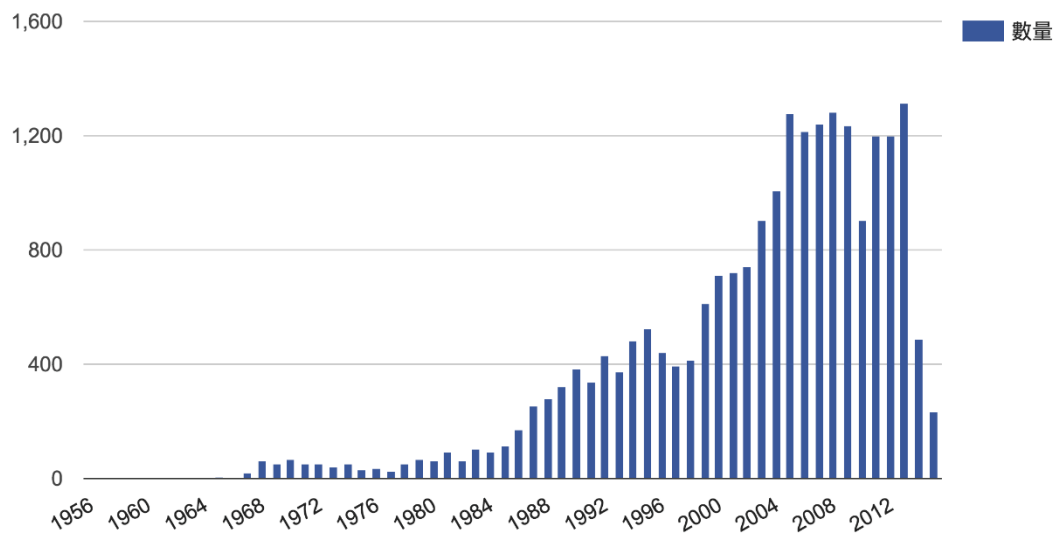


圖 4-48 勵志風格歷年作品數量

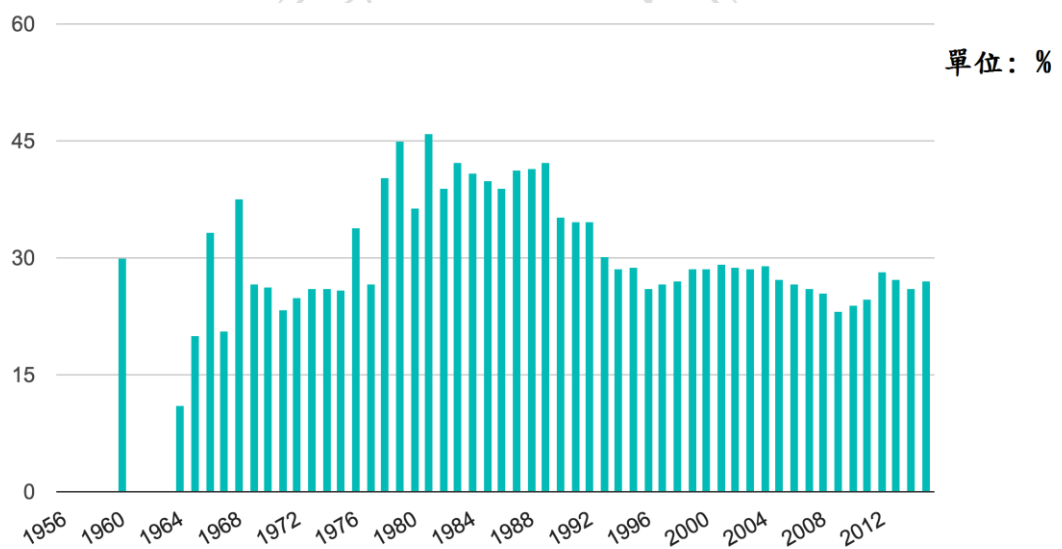


圖 4-49 勵志風格佔該年歌詞數量的比例

友情風格部分，除 1960 年、1979 年及 1980 年曾佔該年歌詞比重達 30%、21.9% 及 21.2% 外，其他年則皆低於 20%，由比例圖可得知友情風格數量逐年下降，不及 1975 年~1983 年超過 10% 的盛況。參考圖 4-50 及圖 4-51。

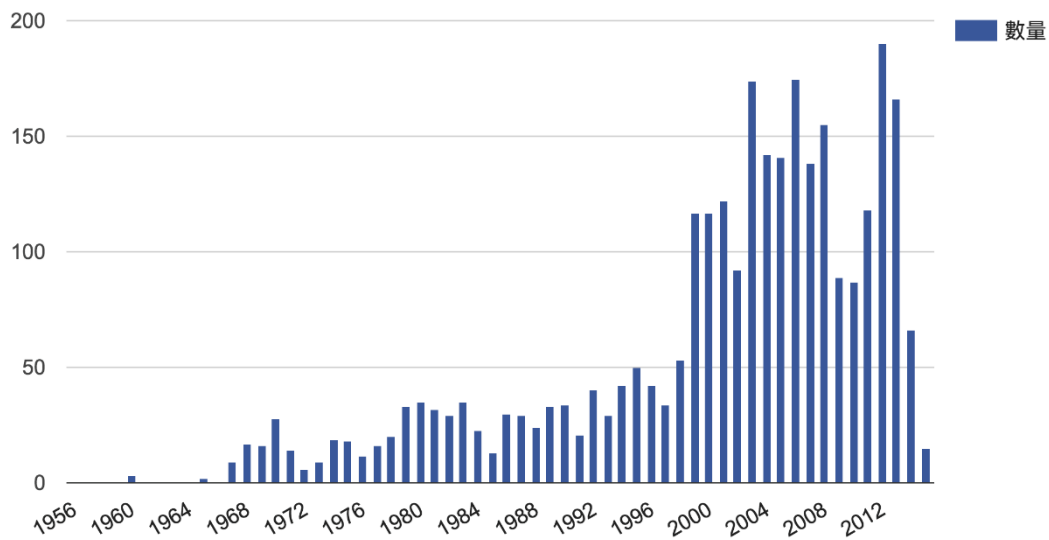


圖 4-50 友情風格歷年作品數量

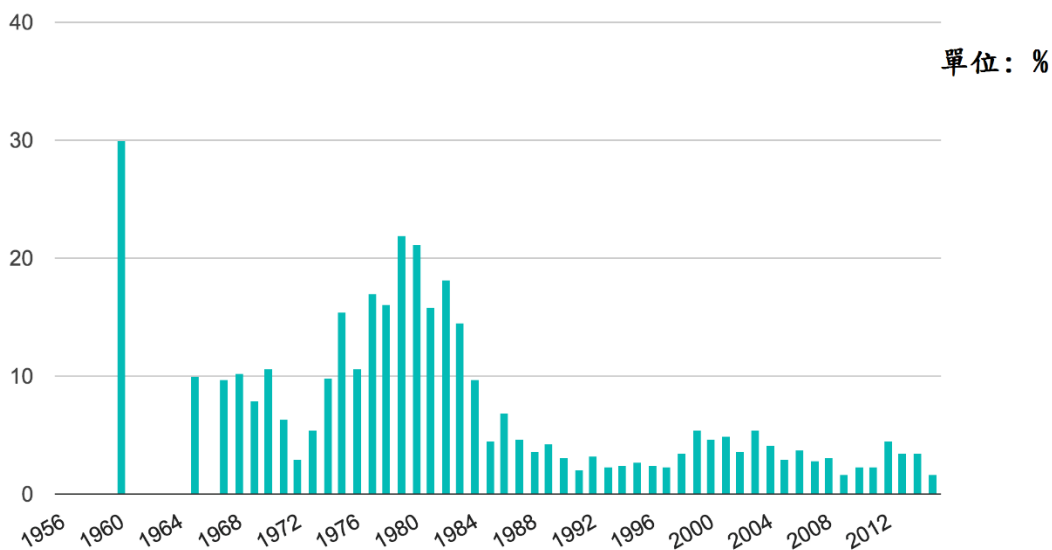


圖 4-51 友情風格佔該年歌詞數量的比例

親情風格的歌詞則較少，僅 37 首，除 1977 年佔該年歌詞比重 1.1% 外，其他年則皆低於 0.5%。參考圖 4-52 及圖 4-53。

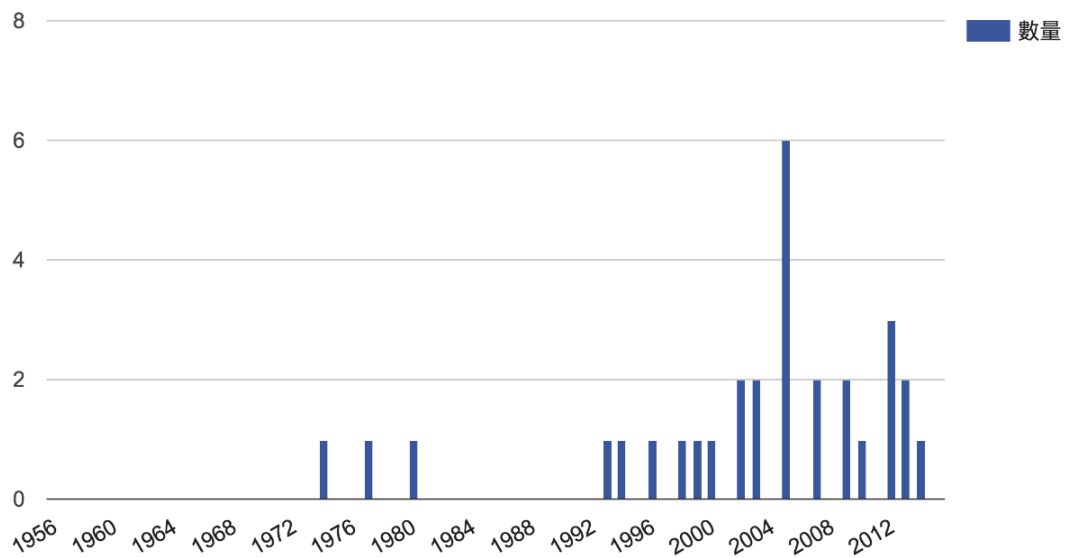


圖 4-52 親情風格歷年作品數量

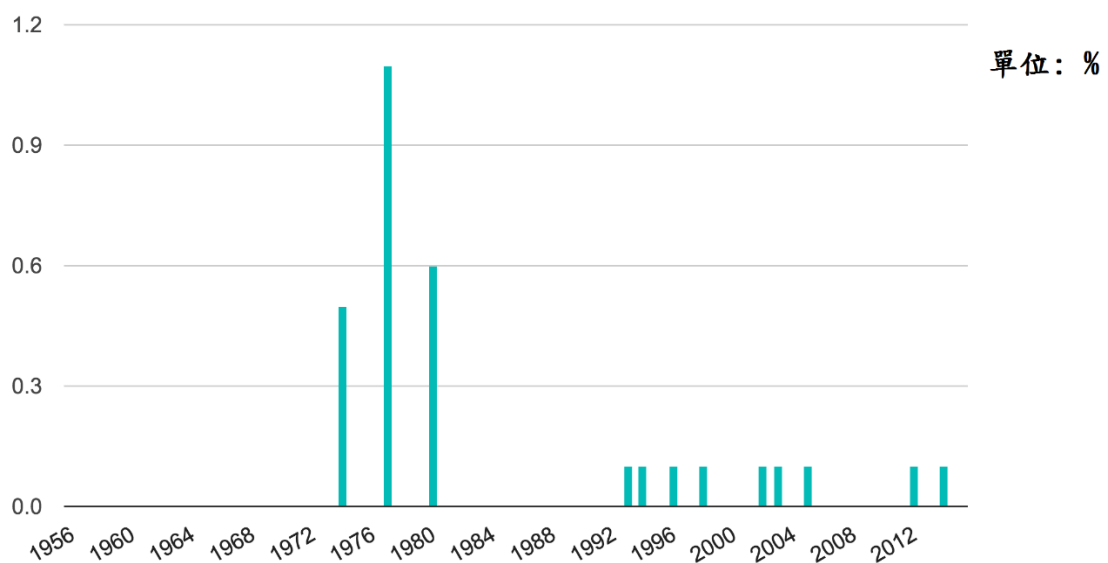


圖 4-53 親情風格佔該年歌詞數量的比例

其他風格部分，包含人生百態、田野風光、社會批判、愛國、鄉情等 5 種風格，由於 1960 年、1964 年及 1965 年由魔鏡歌詞網僅獲得 3 首、4 首及 4 首歌詞，所以這三年不列入比較範圍。

1967 年~1981 年，其他風格的佔該年歌詞比重較常出現超過 10% 的情況，然 1986 年

後，鮮少有出現 7% 以上的比重，這或許與 1987 年 7 月台灣政府宣佈「解除戒嚴」有關係，因此社會批判、愛國、鄉情這三類的歌詞量也相對減少許多。

根據謝櫻子，方文山華語詞作主題研究，1985 年後，台灣經濟結構改變，導致華語流行音樂歌詞的風格也開始由民歌風格轉為探討都市中產的愛情生活，例如張艾嘉演唱的《忙與盲》及陳淑樺演唱的《夢醒時分》。參考圖 4-54 及圖 4-55。

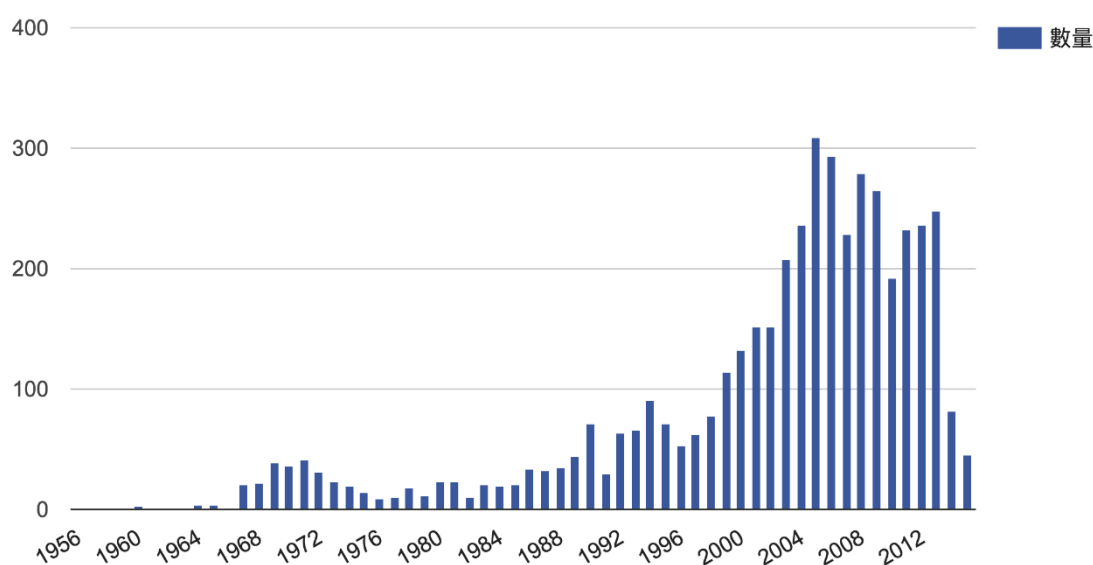


圖 4-54 其他風格歷年作品數量

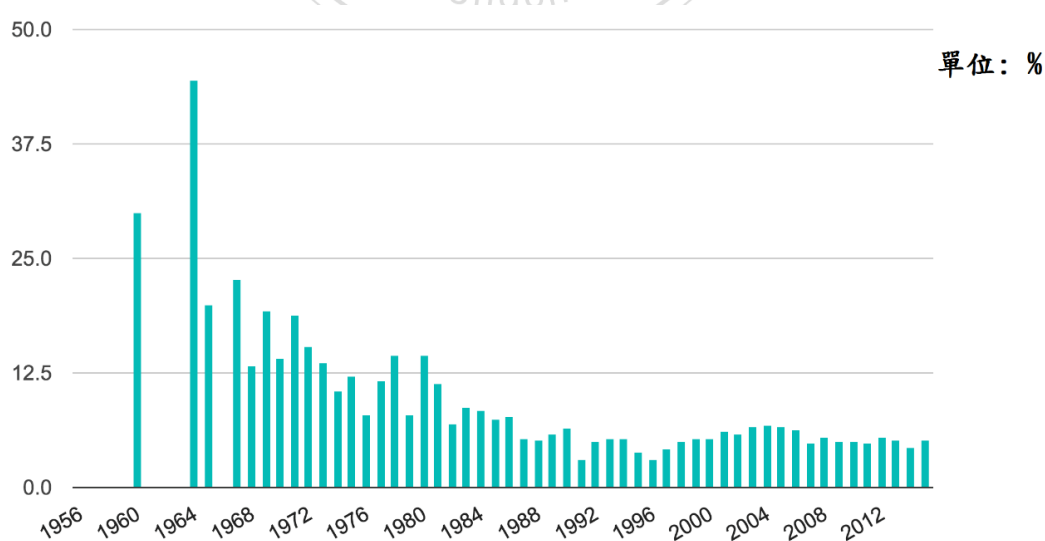


圖 4-55 其他風格佔該年歌詞數量的比例

4.14 風格對歌名字數的交叉分析

各類風格的歌名大多為二字部~五字部，特別的是親情風格特別偏向三字部和四字部；至於愛情、友情及勵志風格的歌名字數，則多集中於四字部。參考表 4-14 各類風格使用各字部的一覽表及圖 4-56 各類風格類型使用各字部佔該風格之比例圖。

表 4-14 各類風格使用各字部的一覽表

	愛情	親情	友情	勵志	其他
一字部	1,304	1	107	852	142
二字部	10,195	3	652	4,874	801
三字部	8,038	11	835	4,107	1,200
四字部	12,974	12	945	6,421	1,213
五字部	9,664	5	559	4,368	834
六字部	4,561	1	341	1,780	340
七字部	3,844	4	215	1,505	332
八字部	1,827	0	74	617	152
九字部	1,046	0	34	293	57
十字部	569	0	40	151	27
十字部以上	468	0	40	163	47

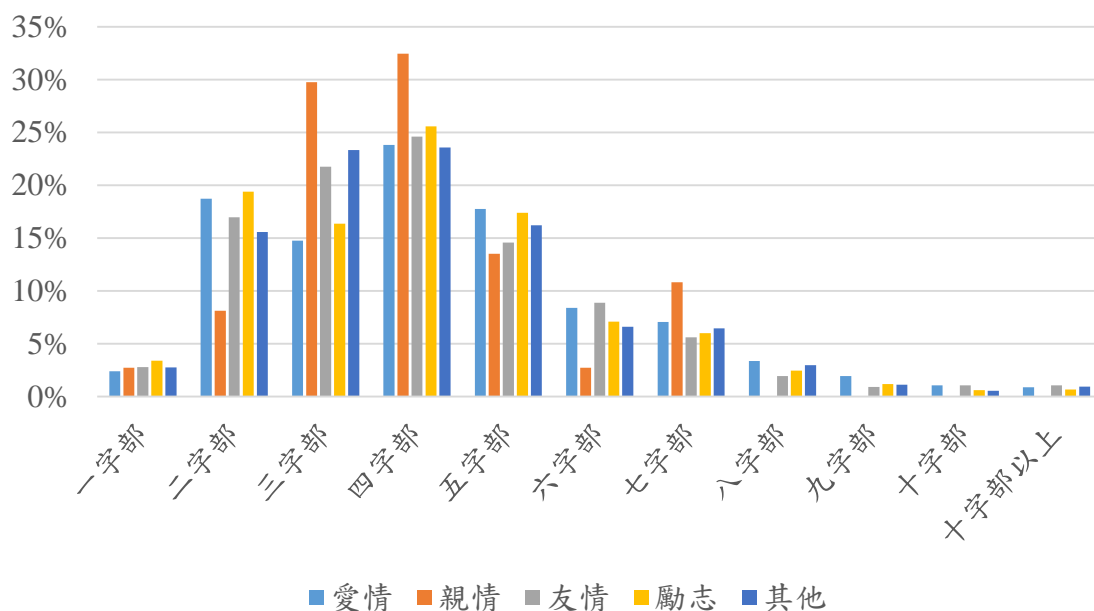


圖 4-56 各類風格類型使用各字部佔該風格之比例圖

4.15 風格對使用語系的交叉分析

各類風格還是使用全部都中文的情況較多，而使用外來語的情況，則以英文為主，勵志風格有使用中英夾雜的比例達 30.9%，而愛情風格有使用中英夾雜的比例達 29.9%，其次是韓文及日文。參考表 4-15 各類風格使用語系一覽表。

表 4-15 各類風格使用語系一覽表

	愛情	親情	友情	勵志	其他
只有中文	41,935	36	3,229	19,128	4,103
中英夾雜	12,542	1	605	5,906	1,041
中日夾雜	7	0	4	3	0
中韓夾雜	16	0	2	7	0

4.16 風格對押韻的交叉分析

以風格來分析韻腳，取每個風格的 Top 10 韻腳，可發現各風格最常使用韻腳皆為「一」韻，常用字為「你、氣、離」等字。而除親情外，其他四韻皆常使用「ㄣ」韻，常用字為「天、見、面、現」等字，如圖 4-57。

值得注意的是在愛情風格中，有使用「ㄛ」韻及「ㄣㄛ」韻，而其他四類風格前十名常用的韻腳則無這兩韻。ㄛ韻常用字為「冥、喔」等字，ㄣㄛ韻常用字為「過、說、火」等字。

勵志和親情風格使用的 Top 10 韻腳中，「ㄣ」韻只有該兩風格使用，未被其他風格所使用，該韻代表字為「晚、然、慢」等字。而勵志和其他風格使用的 Top 10 韻腳中，「ㄩ」韻只有該兩風格使用，未被其他風格使用，該韻代表字為「他、它、啊、吧」等字。

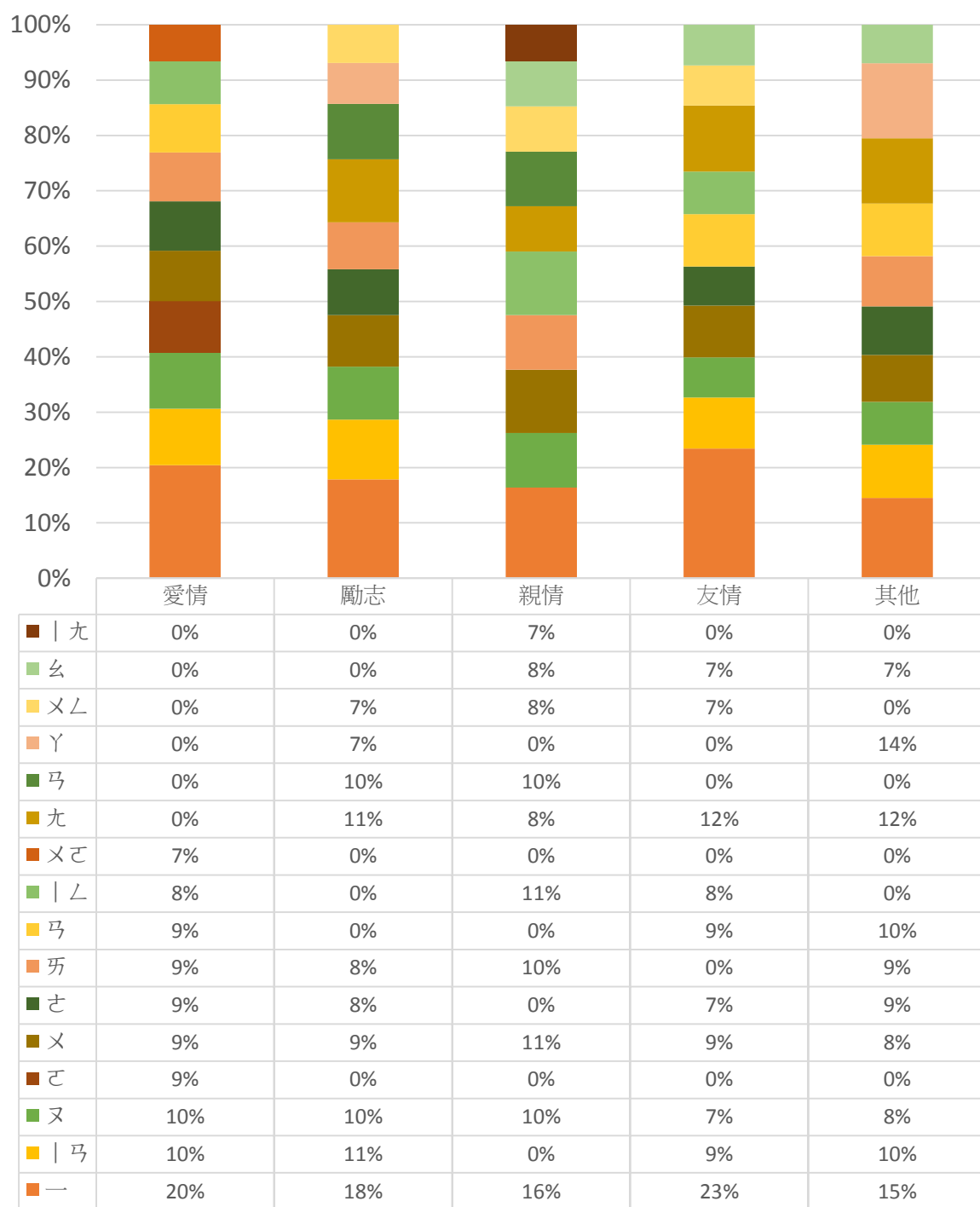


圖 4-57 各風格押韻 Top 10 一覽表

4.17 風格對修辭的交叉分析

明喻部分以愛情風格最常使用，恰當的比喻，不但能使人恍然大悟，明白所指，且能讓人驚佩、滿足、信服及印象深刻，例如由李宗盛作詞的「連夢也沒有」歌詞：「他的愛情彷彿如同一場夢，他和他的女人如此短暫的相逢，但是我連夢也沒有」，以夢來比喻這段愛情之短暫及飄渺須無之感。

至於類字，各類風格使用的比例非常高，均達 79% 以上，愛情風格的歌詞更是高達 79.9% 的歌詞都有使用類字；其次，疊字的使用比例也非常高，均超過 54% 以上；對偶部分，雖本研究以採用寬式對偶，但使用情況還是較其他修辭來得少。

至於疊字及疊句的修辭法，則以親情風格最多；對偶部分則以其他風格類使用狀況最多，達 21.7%。參考表 4-16。

表 4-16 各類風格修辭使用比例一覽表

	愛情	親情	友情	勵志	其他
明喻	34.1%	27.0%	16.4%	28.6%	27.2%
類字	97.9%	89.2%	79.0%	96.7%	97.4%
疊字	64.4%	75.7%	54.2%	62.1%	69.2%
疊句	16.3%	24.3%	18.5%	16.7%	26.6%
對偶	18.4%	10.8%	17.6%	18.7%	21.7%

第五章 結論與未來研究方向

5.1 結論

本論文針對由知名的歌詞網站「魔鏡歌詞網」取得大量歌詞資料後，利用資料探勘的技術，先進行文字規則分析，除找出歌詞中的高頻詞及相鄰關係，也進行歌名字數及使用語系的探討。

至於歌詞風格分析方面，我們進行以往研究較少使用的階層式分類，除了提高分類準確性的目標外，階層式分類的上層分類器的準確率會高於底層分類器，此外，階層式分類的好處是可以依照不同需求，進行中間層或底層分類器。在風格分類實驗中，相較於單一階層的平面分類器，階層式分類器準確率也的確較為增加，此乃因階層式分類器將複雜的分類問題分開來處理，支分類器僅處理分支下的結點，而其他分支下的節點則不列入考慮，如此一來，所處理的問題較單純，準確率因而得以提升。

此外，本研究有兩方面的貢獻：1. 應證了相關研究的結果及 2. 發現新結果。例如 1975~1983 年愛情風格的歌詞比例降低，對照流行歌詞時期相關研究，這時期剛好為為民歌興盛時代，故被認為靡靡之音的愛情風格的歌詞比例降低。再者，1990 年後，全球化時期，所以歌詞內夾雜英文歌詞的比例增加；流行歌詞使用的修辭法為類疊法較多，符合吳嫩婉的台灣國語流行歌曲的修辭藝術的研究，且透過程式分析，可更得到更精細的結果，即歌詞中類字使用方法即高，達 96.7%，疊字為 63.6%，然疊句較少，只有 17.1% 的歌詞有使用該修辭法。

至於新發現部分，舉例如下：探勘出高頻詞除了「我們」等稱謂詞外，亦發現「世界」為非常常用的詞；此外，在押韻部分，Y 在 1980 年後不在 Top 10 韻腳，而ㄊ 韻在 2000 年後才興起，而押韻與風格部分，僅愛情風格經常使用 ㄊ、ㄣ 韻等，這些項目都值得研究發音的學者進一步探討。

此外，系統實作完成後，我們得以進行各項目交叉分析，包括年代、風格與各項目的交叉分析，我們得以發現在不同時期，歌詞風格數量、作詞者作品數量、韻腳及押韻風格的變化，同時，也得知不同風格的歌名、使用語系、押韻及修辭法使用的狀況，進而更了解華語流行歌詞的規則與奧妙之處。

5.2 未來研究方向

本研究處理比譬喻法時，發現目前程式只能分析明喻，但還有更多係屬於隱喻、略喻及借喻的方式，然這些方式因為沒有喻詞，較難以程式直接分析，可是隱喻、略喻、借喻及假喻經常出現在文章，故在華語流行歌詞的使用率應該不少，值得進一步研究自動化分析的方法。

至於風格方面，愛情風格為最常使用該明喻法的風格，該用法佔愛情風格比例達 34%，而其他風格的歌詞使用明喻法的比例都低於 30%，這或許也可以進一步分析探討其原因。

至於本研究的交叉查詢系統，在未來可增加更多功能，包括自動將最新歌詞自動導入系統作分析、自動產生詞彙相鄰的圖等功能。此外，過去也曾有自動作詞的相關研究，例如 2015 年國立政治大學葉佩雯的「為你寫一首歌：以資料探勘為基礎的自動作詞系統」[18]，未來亦可將本研究的歌詞特徵研究與自動作詞的研究結合，自動創作出更觸動人心的歌詞。

在過去十年間的資料科學數據分析，已經徹底改變許多企業的經營模式，例如電子商務以資料挖掘技術找出顧客的習性、購買商品間的關聯性等，來建立客戶關係管理，並進而規劃出一個完善的企業資源規劃，幫助企業留住顧客；而未來歌詞科學很可能是下一個因 Big Data 研究的受益者之一，使得更美妙的歌詞因而產生。

參考文獻

- [1] Jia-Lien Hsu, Chih-Chin Liu, and Arbee L.P. Chen, Discovering Non-Trivial Repeating Patterns in Music Data, IEEE Transactions on Multimedia, Vol. 3, No. 3, Sep. 2001.
- [2] Xiao Hu, J. Stephen Downie and Andreas F. Ehmann, Lyric Text Mining in Music Mood Classification, 10th International Society for Music Information Retrieval Conference (ISMIR), Oct. 2009.
- [3] Beth Logan, Andrew Kositsky, and Pedro Moreno, Semantic Analysis of Song Lyrics, IEEE International Conference on Multimedia and Expo, Vol. 2, June 2004.
- [4] Rudolf Mayer, Robert Neumayer and Andres Rauber, Rhyme and Style Features for Musical Genre Classification by Song Lyrics, 9th International Conference on Music Information Retrieval (ISMIR), Sep. 2008.
- [5] Seungwon Oh, Minsoo Hahn and Jinsul Kim, Music Mood Classification Using Intro and Refrain Parts of Lyrics, IEEE International Conference on Information Science and Applications, June 2013.
- [6] 曾慧佳，從流行歌曲看台灣社會，桂冠圖書公司，2000。
- [7] 蔡宗陽，應用修辭學，萬卷樓圖書公司，2001。
- [8] 黃麗貞，實用修辭學（增訂版），國家出版社，2004。
- [9] 黃慶萱，修辭學（增訂三版），三民書局出版社，2005。
- [10] 吳嫩婉，台灣國語流行歌曲的修辭藝術(1949-2000)，國立臺北教育大學語文教學碩士班論文，2005。
- [11] 張雯禎，台灣流行歌詞中的隱喻：以愛情為主題(1990-2008)，國立中正大學碩士班論文，2008。
- [12] 劉祐銘，臺灣國語流行歌曲歌詞用韻研究(1998~2008)，靜宜大學碩士班論文，2009。
- [13] 施啟智，華語通俗音樂模組化數位音樂的構成、分析及應用，樹德科技大學碩士班論文，2009。
- [14] 朱宜秦，流行音樂中作詞人性別主體性對作品觀點差異之研究，國立臺南藝術大學碩士班論文，2011。
- [15] 謝櫻子，方文山華語詞作主題研究，國立新竹教育大學碩士班論文，2010。
- [16] 馬占山，臺灣流行歌詞主題類型與語言表達研究（1999~2008）——以「Hit-FM 年度

百首單曲」為對象，國立臺灣師範大學碩士班論文，2013。

[17] 王奕舜，論敘事型音樂錄影帶和歌詞的改編與互文性-以金曲獎音樂錄影帶獎項為例(2010-2012)，國立臺灣藝術大學碩士班論文，2013。

[18] 葉佩雯，為你寫一首歌:以資料探勘為基礎的自動作詞系統，國立政治大學資訊科學系大專生專題研究計畫報告，2015。

[19] 魔鏡歌詞網，Retrieved Oct. 1, 2015, from the World Wide Web，<https://mojim.com/>。

[20] 簡冠庭，HTML5 Word Cloud，Retrieved Jan. 10, 2016, from the World Wide Web，<http://timdream.org/wordcloud/>。

[21] 中央研究院資訊科學研究所，Retrieved Jan. 10, 2016, from the World Wide Web，http://www.iis.sinica.edu.tw/index_zh.html。

[22] 中文詞知識庫小組，Retrieved Mar. 15, 2016, from the World Wide Web，<http://godel.iis.sinica.edu.tw/CKIP/>。

[23] 教育部重編辭語修訂本，Retrieved Mar. 1, 2015, from the World Wide Web <http://dict.revised.moe.edu.tw/>

[24] 九酷音樂，Retrieved Mar. 10, 2016, from the World Wide Web，<http://www.9ku.com/zhuanji/taste.htm>。

[25] 酷我音樂，Retrieved Mar. 10, 2016, from the World Wide Web，<http://yinyue.kuwo.cn/yy/category.htm>。

[26] 臺北之音 HitFM 聯播網，Retrieved June. 13, 2016, from the World Wide Web，<http://www.hitoradio.com/>。