

單飛紅什麼： 歌手在團體中與個人的歌曲用詞變遷

語言所二 r07142006 許家誠

介紹

在樂壇中，有不少的歌手曾是在歌唱團體中的一員，而後單飛，邁向個人化的發展。有些人可能是因為與團體成員不合而離開團體，有些人可能是因為與唱片公司或經紀公司合約的問題，不得不離開原本的環境，也有些人是為了想發展出個人的風格，而告別團體的形象。

本文將以台灣的三位華語歌手為例，觀察該歌手單飛之前所在的團體、以及單飛之後所推出的歌曲作品之中，在特定主題上的用詞上有什麼樣的變化，藉以推論該歌手有哪些風格上的轉變。

方法

本次報告將會以 **word2vec** 的方式，把各個歌手的單飛前後的特定詞彙分別標記，以觀察該詞彙在兩個時期在詞向量空間中的位置差異。

第一位要觀察的歌手是楊丞琳，於 2000 年加入 4 in Love 組合出道，發行 2 張專輯後解散，2005 年發行第一張個人專輯，即活躍在華語歌壇至今。第二位蘇見信，2002 年擔任信樂團主唱，於 2007 年單飛。第三位是田馥甄，於 2001 年在團體 S.H.E 發行首張專輯《女生宿舍》正式出道，2010 年「單飛不解散」，推出首張個人專輯《To Hebe》，並同時在團體中及個人都有新作品產出。依發行歌曲與活躍年份來區分，楊丞琳為出道前期單飛，蘇見信為出道中期單飛，田馥甄則為出道後期單飛。

資料取得

魔鏡歌詞網 (<https://mojim.com>) 是一個歌詞資料網站，於 2005 年 8 月成立，涵蓋亞洲、歐洲、美洲、日本、韓國歌手等歷年的專輯歌詞，並有詳細分類，同時歌詞亦會

每天由網友提供並更新。雖然在此網站的歌詞內容大多是由網友提供及維護，歌詞的格式與正確性可能會有偏差，但是在網址結構上簡潔，html tag 的格式也相當公整，是適合做網路爬蟲的材料之一。

前處理

用 python 的 request 和 BeautifulSoup4 module 來獲取各個歌手所有歌曲的歌詞，用 regular expression 的搜尋法清理資料，再使用 jieba 的精確模式進行斷詞。得到了 4 In Love 共 20 首歌曲、2100 個 tokens、974 個 types，楊丞琳共 141 首歌曲、26250 個 tokens、4675 個 types，信樂團共 74 首歌曲、9318 個 tokens、2454 個 types，蘇見信共 130 首歌曲、19228 個 tokens、4552 個 types，S.H.E 共 252 首歌曲、41726 個 tokens、7393 個 types，田馥甄共 88 首歌曲、15415 個 tokens、3353 個 types。

詞向量表徵

在設定好要觀察的詞，分別對單飛前和單飛後的歌詞中做不同的標記，讓 word2vec 模型將他們視為不同的詞，以觀察該詞在單飛前和單飛後分別出現在詞向量空間中的位置。加上前綴 b-表示在單飛前的團體中的觀察詞，前綴 a-表示單飛之後個人推出的歌曲作品。

而要觀察的詞，將以「朋友」、「愛」、「人生」三個現在華語歌曲常見的主題來看三位歌手在這些主題有什麼改變。

參數設定，則以 min_count=5（其中在 4 in love vs. 楊丞琳「人生」組因詞數量不足而將參數改為 1）、size=300、workers=32、hs=1 來訓練模型。

結果

4 in love vs. 楊丞琳

首先，觀察「朋友」這個詞在 4 in love 與楊丞琳共同組成的歌詞語料中，與屬於 4 in love 的「朋友」一詞位置相近的詞有「走進」、「消遣」、「人們」、「悲傷」、「人們」、「黏」等詞，而與屬於楊丞琳的「朋友」一詞位置相近的詞有「做」、「一點」、「心裡」、「喜歡」、「想像」、「以為」、「笑」等詞。

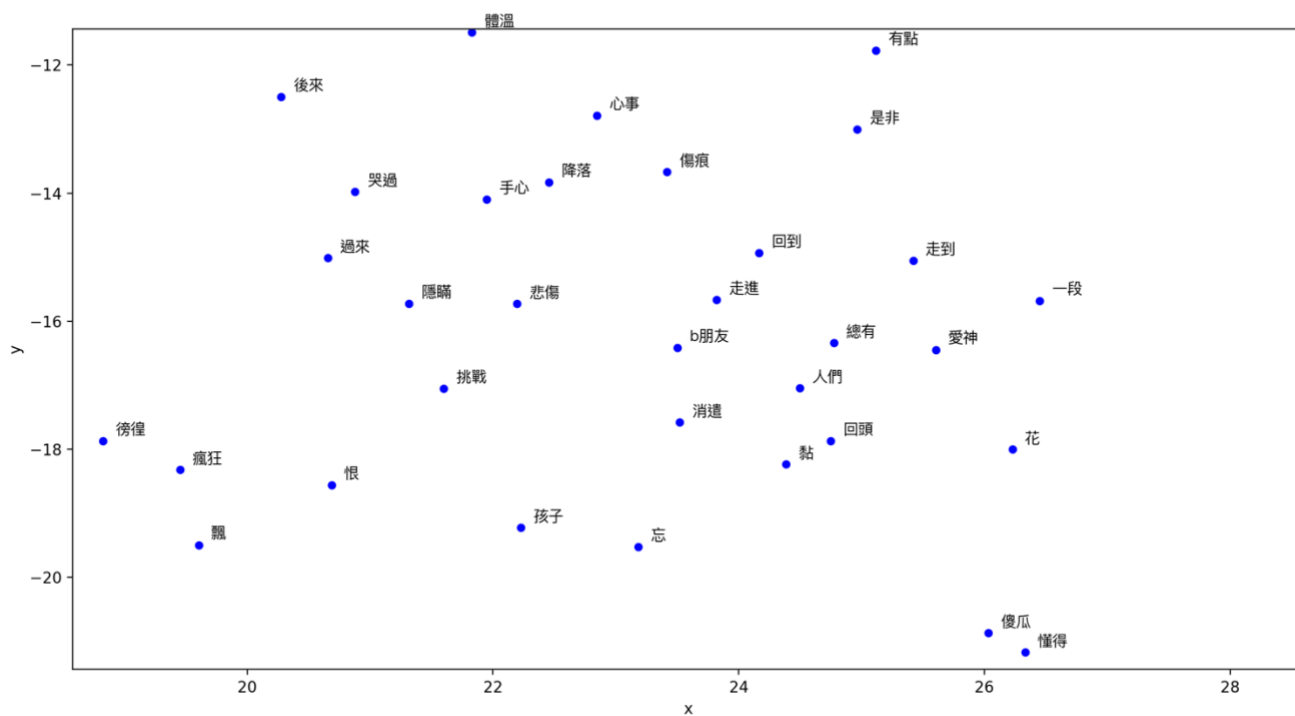


圖1：「4 in love」vs.「楊丞琳」word2vec 2D-視覺化散布圖—b 朋友附近

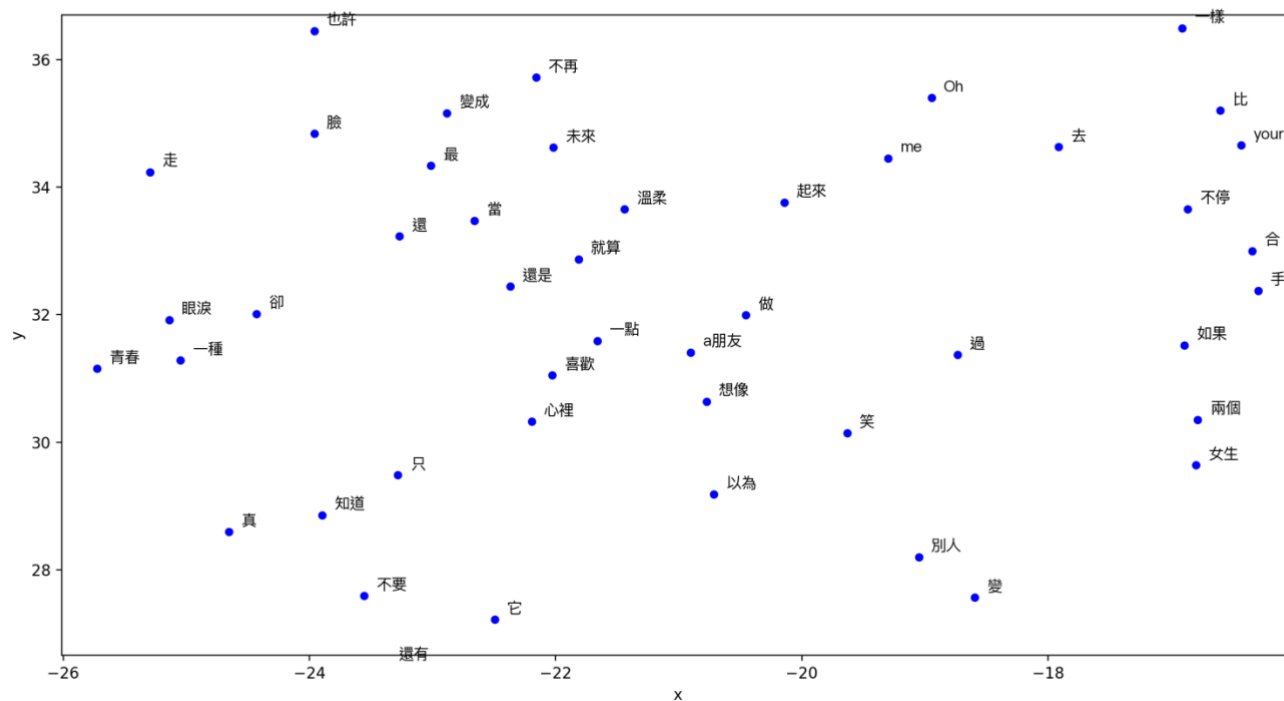


圖2：「4 in love」vs.「楊丞琳」word2vec 2D-視覺化散布圖-a 朋友附近

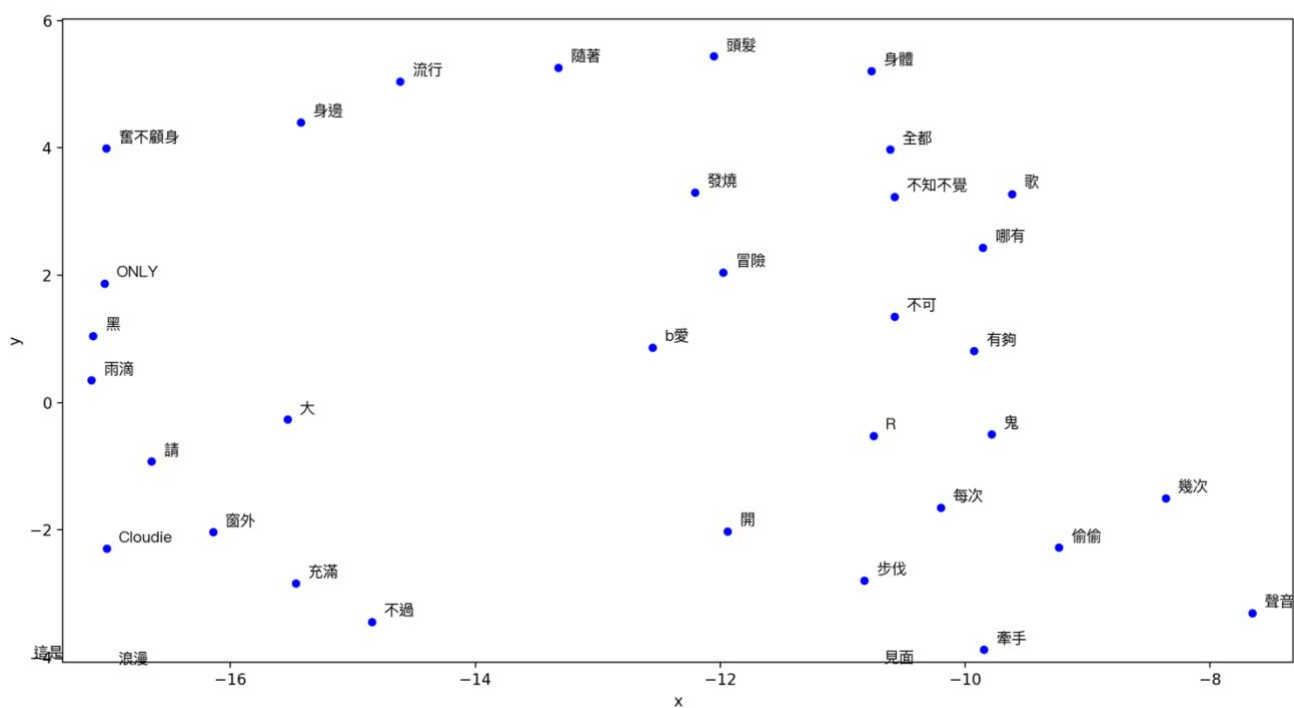


圖 3：「4 in love」vs. 「楊丞琳」 word2vec 2D-視覺化散布圖—b 愛附近

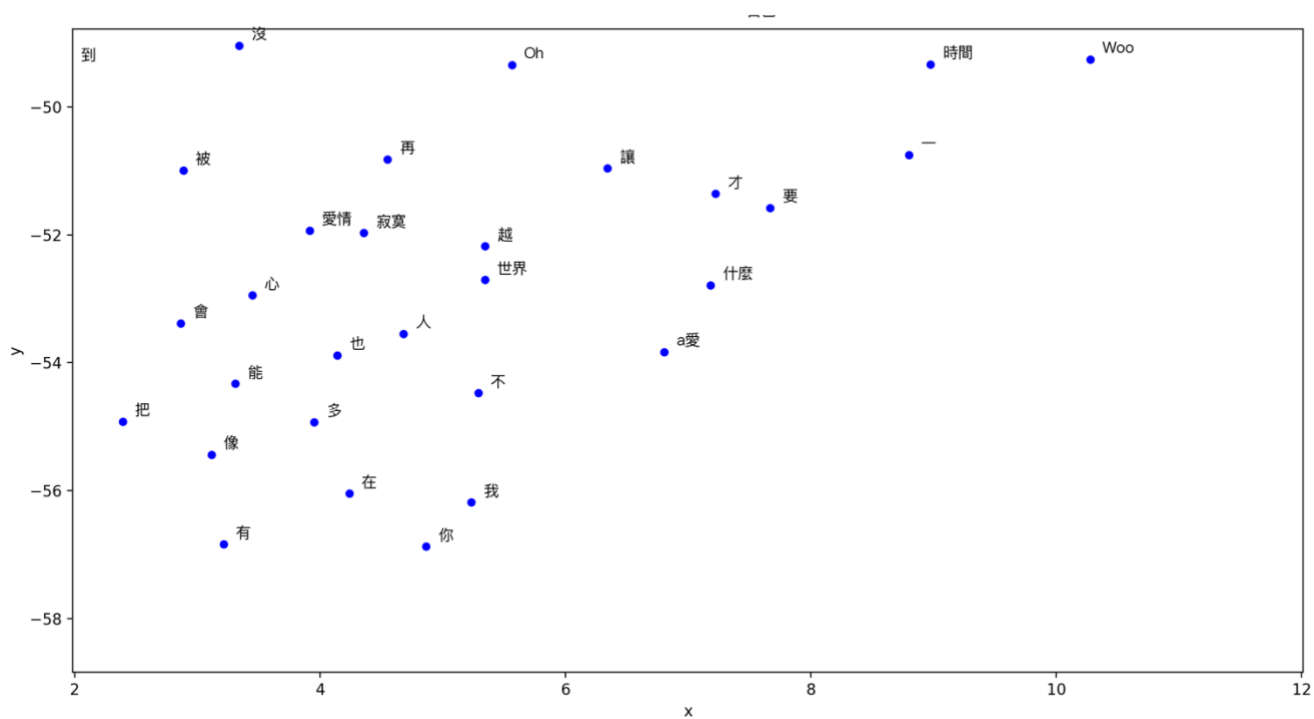
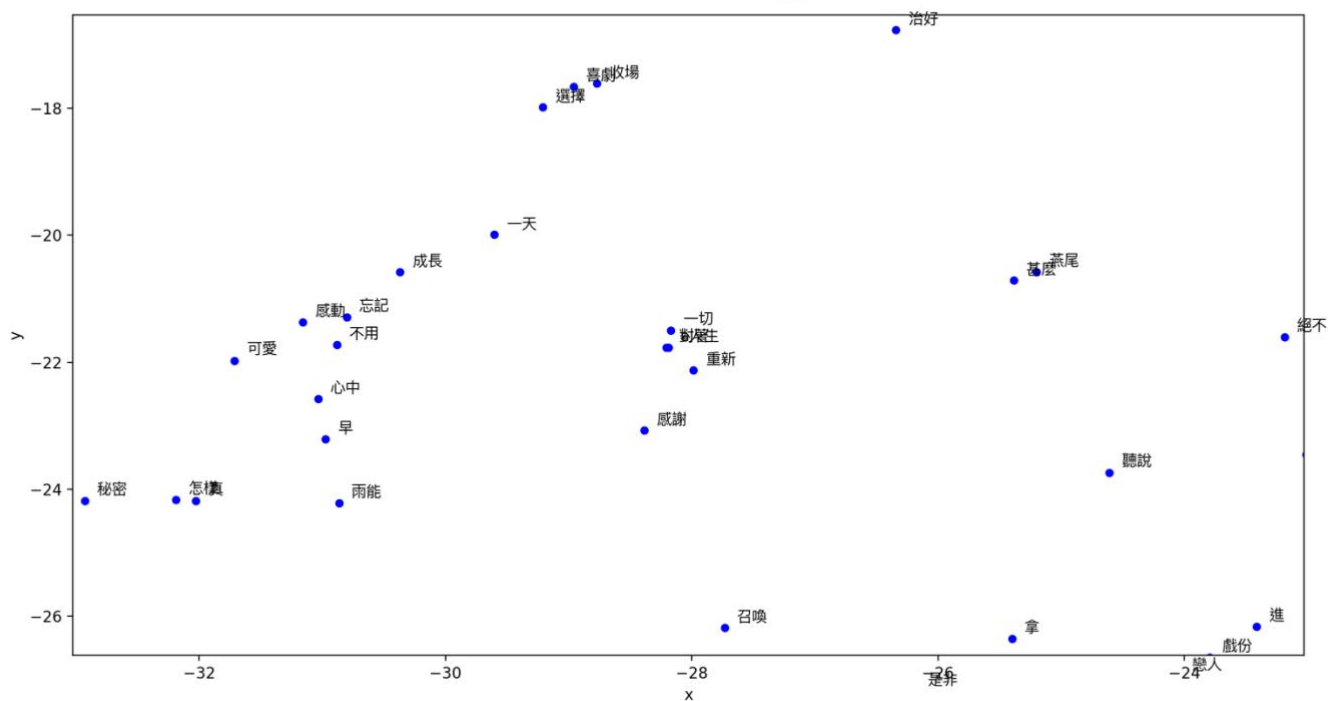
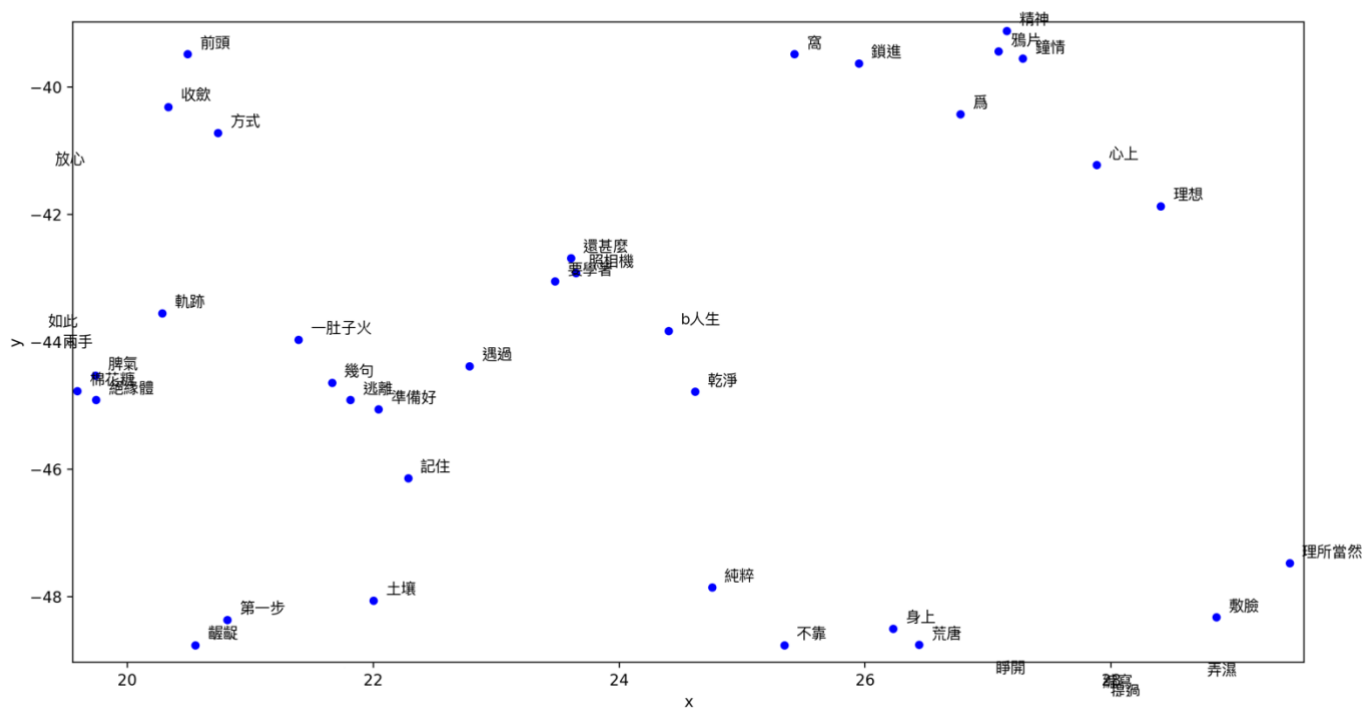


圖 4：「4 in love」vs. 「楊丞琳」 word2vec 2D-視覺化散布圖—a 愛附近



而「愛」這個詞在 4 in love 與楊丞琳共同組成的歌詞語料中，與屬於 4 in love 的「愛」一詞位置相近的詞有「冒險」、「發燒」、「不可」、「開」等詞，而與屬於楊丞琳的「愛」一詞位置相近的詞有「什麼」、「才」、「要」、「世界」、「越」、「不」、「讓」等詞。

最後，「人生」這個詞在 4 in love 與楊丞琳共同組成的歌詞語料中，與屬於 4 in love 的「人生」一詞位置相近的詞有「照相機」、「還甚麼」、「要學著」、「乾淨」、「遇過」等詞，而與屬於楊丞琳的「人生」一詞位置相近的詞有「對著」、「重新」、「感謝」、「一切」、「一天」、「成長」等詞。

這三個詞在單飛前和單飛後的餘弦相似度分別為 0.8509、0.9510 和 0.5117。從數據上可以發現，「人生」一詞在單飛前和單飛後的差別是比較大的。「人生」一詞相較於單飛前與「遇過」位置相近，單飛後與「感謝」、「成長」位置相近，似乎對人生有新的體悟，使其成長。

信樂團 vs. 蘇見信

信樂團與蘇見信的部分，觀察「朋友」這個詞在信樂團與蘇見信共同組成的歌詞語料中，與屬於信樂團的「朋友」一詞位置相近的詞有「竟」、「陪伴」、「難過」、「不放」等詞，而與屬於蘇見信的「朋友」一詞位置相近的詞有「精彩」、「煩惱」、「心還」、「向」、「表情」、「頭」等詞。

而「愛」這個詞在信樂團與蘇見信共同組成的歌詞語料中，與屬於信樂團的「愛」一詞位置相近的詞有「才」、「想」、「只」、「中」、「也」、「擁抱」、「青春」等詞，而與屬於蘇見信的「愛」一詞位置相近的詞有「很」、「才」、「只」、「自己」、「沒」、「生命」、「未來」等詞。

最後，「人生」這個詞在信樂團與蘇見信共同組成的歌詞語料中，與屬於信樂團的「人生」一詞位置相近的詞有「夜夜」、「春天」、「祢」、「每」、「情人」、「珍惜」等詞，而與屬於蘇見信的「人生」一詞位置相近的詞有「只是」、「遙遠」、「兄弟」、「早」、「這一刻」、「為什麼」、「並肩」等詞。

這三個詞在單飛前和單飛後的餘弦相似度分別為 0.8202、0.9459 和 0.9704，都很相近。其中在「人生」一詞雖然十分相近，但是從圖中觀察還是有些不同。單飛前的「人生」似乎偏向「情人」，而單飛後似乎比較偏向「兄弟」。

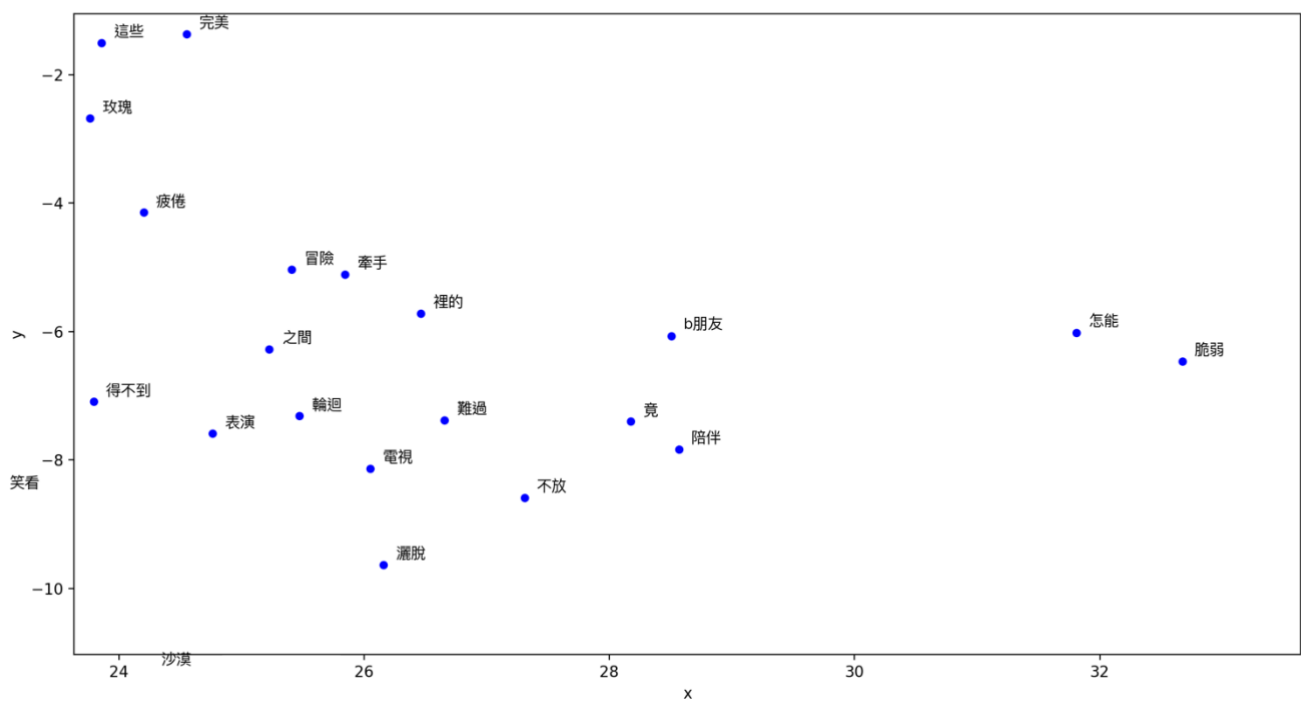


圖 7：「信樂團」vs.「蘇見信」word2vec 2D-視覺化散布圖—b 朋友附近

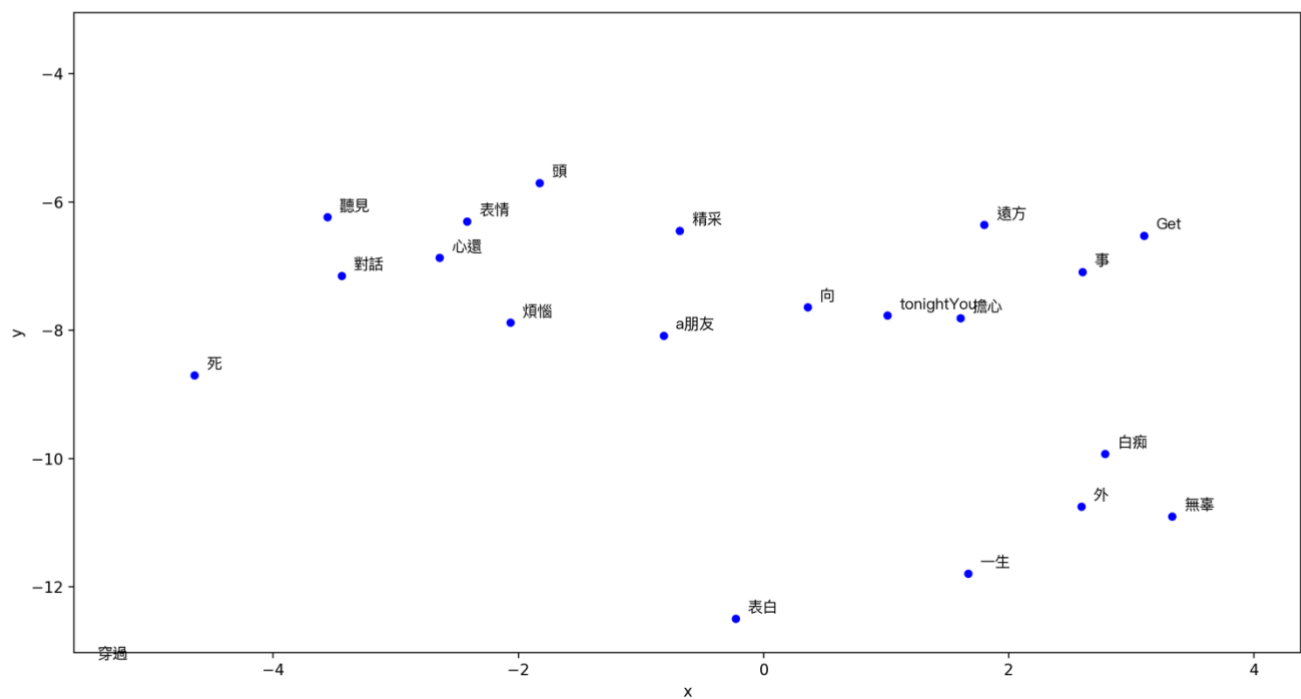


圖 8：「信樂團」vs.「蘇見信」word2vec 2D-視覺化散布圖—a 朋友附近

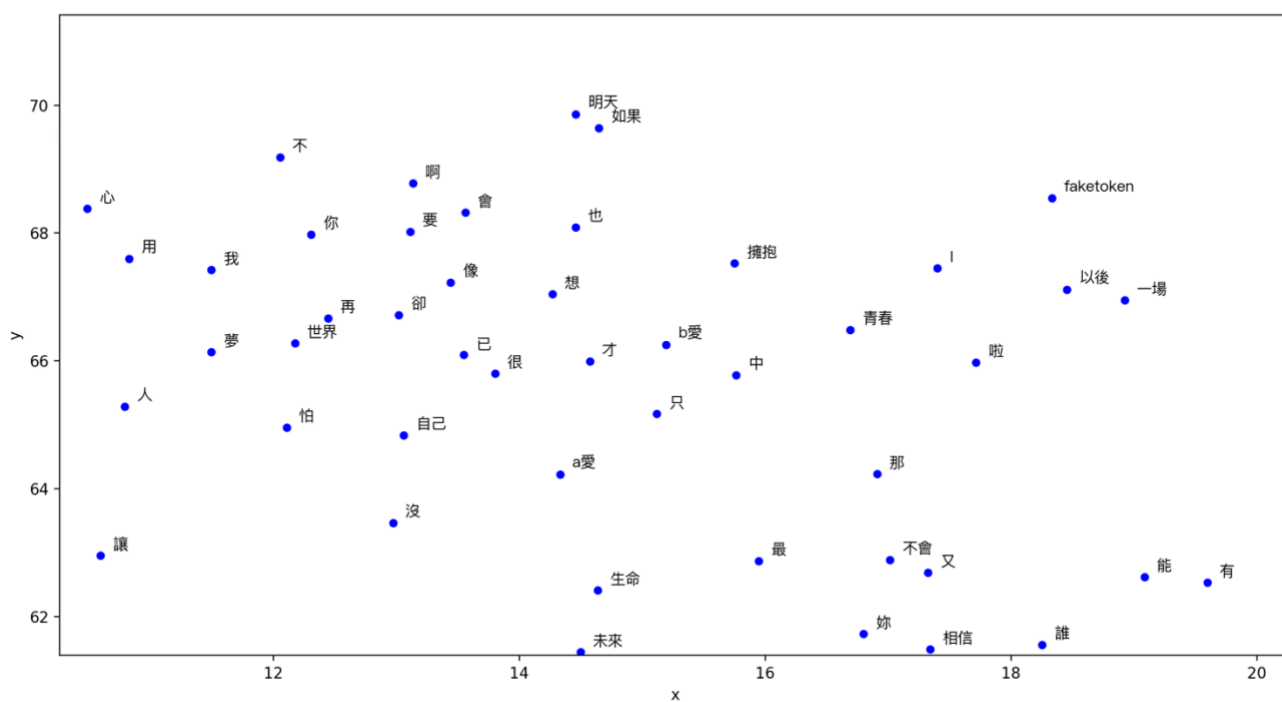


圖9：「信樂團」vs.「蘇見信」word2vec 2D-視覺化散布圖—b 愛附近

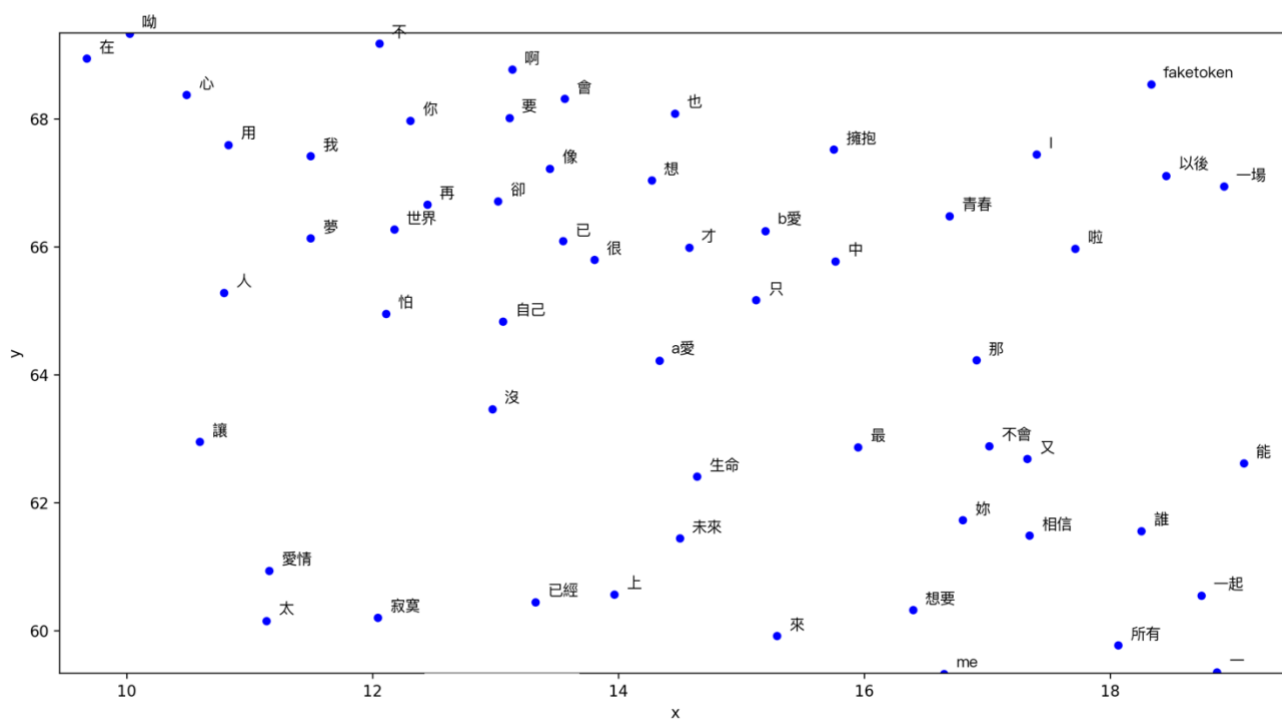


圖10：「信樂團」vs.「蘇見信」word2vec 2D-視覺化散布圖—a 愛附近

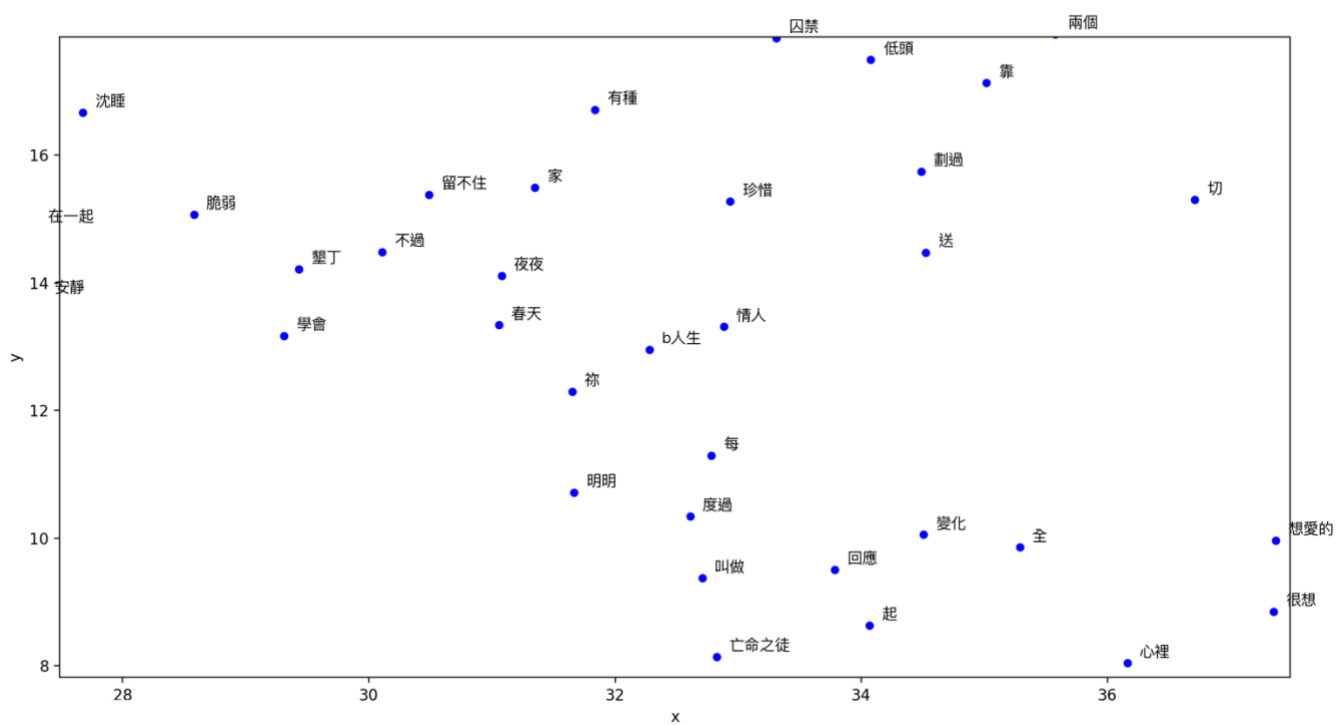


圖 11：「信樂團」vs.「蘇見信」word2vec 2D-視覺化散布圖—b 人生附近

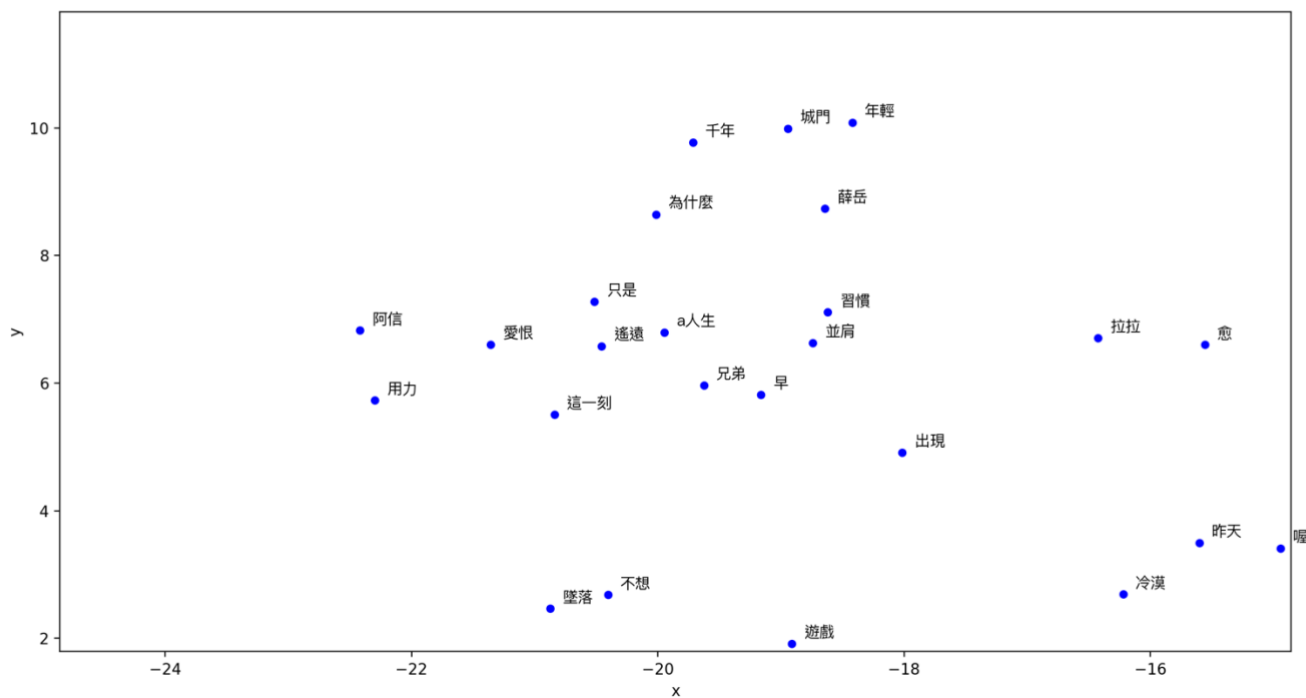


圖 12：「信樂團」vs.「蘇見信」word2vec 2D-視覺化散布圖—a 人生附近

S.H.E vs.田馥甄

S.H.E 與田馥甄的部分，觀察「朋友」這個詞在 S.H.E 與田馥甄共同組成的歌詞語料中，與屬於 S.H.E 的「朋友」一詞位置相近的詞有「開始」、「也許」、「今天」、「明天」、「溫柔」等詞，而與屬於田馥甄的「朋友」一詞位置相近的詞有「不用說」、「芬芳」、「起伏」、「刺激」、「流淚」等詞。

而「愛」這個詞在 S.H.E 與田馥甄共同組成的歌詞語料中，與屬於 S.H.E 的「愛」一詞位置相近的詞有「能」、「快樂」、「多」、「做」、「哭」等詞，而與屬於田馥甄的「愛」一詞位置相近的詞有「她」、「自己」、「能」等詞。

最後，「人生」這個詞在 S.H.E 與田馥甄共同組成的歌詞語料中，與屬於 S.H.E 的「人生」一詞位置相近的詞有「不如」、「為何」、「臉」、「靠」、「一遍」、「甜」等詞，而與屬於田馥甄的「人生」一詞位置相近的詞有「總有」、「甚至」、「了解」、「換個」、「當你」等詞。

這三個詞在單飛前和單飛後的餘弦相似度分別為 0.6816、0.9424、0.8051。從數據上可以發現，「朋友」一詞在單飛前和單飛後的差別是比較大的。「朋友」一詞相較於單飛前與「開始」、「也許」、「明天」等具有可能性的詞位置相近，單飛後與「不用說」等有確性、及「芬芳」、「刺激」、「流淚」、「起伏」等心情五味雜陳的詞語位置相近。

在樂壇定位上，S.H.E 是華語女子團體中姊妹情深的代表，而「單飛不解散」的走向，也讓有田馥甄在內的 S.H.E 與田馥甄同時「存在」，風格的差異可能就必須更大，才容易被看見。

信樂團雖然在蘇見信的退團之後，和蘇見信同時「存在」於華語樂壇，但兩方於樂壇中受眾路線本身有所不同，加上蘇見信也將他在信樂團作品收錄在演唱會實錄專輯中，同質性可見一斑。

4 in love 則是在楊丞琳推出個人作品之前就因為經紀公司的原因而被迫解散，推出的歌曲數量也不多，很難推測出楊丞琳歌曲風格的轉變是否與脫離團體有關。

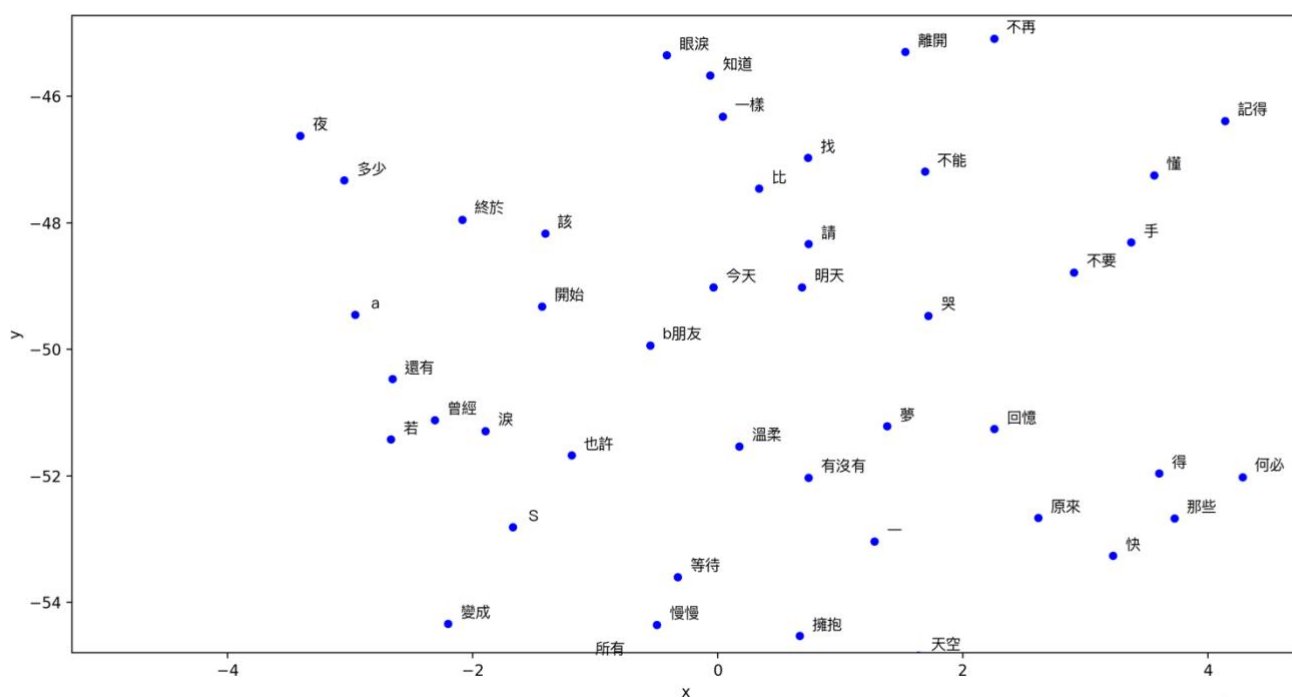


圖 13：「S.H.E」vs. 「田馥甄」 word2vec 2D-視覺化散布圖—b 朋友附近

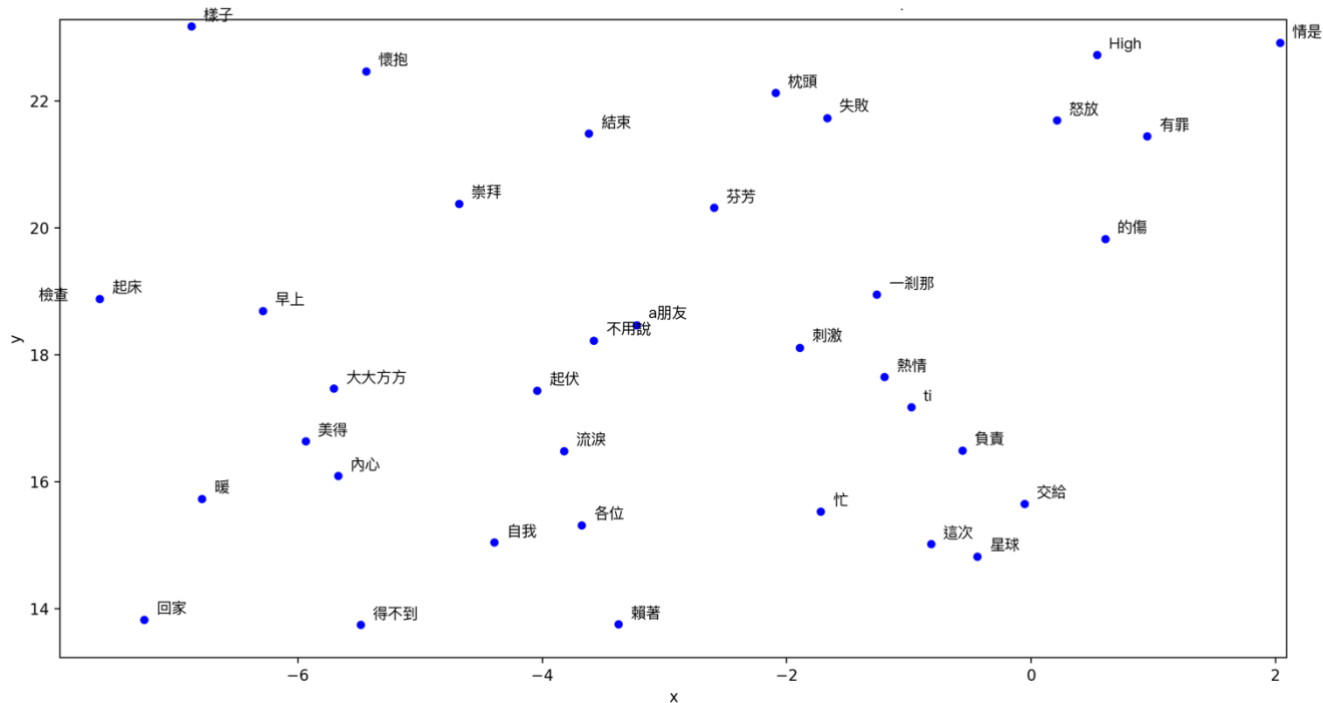


圖 14：「S.H.E」vs. 「田馥甄」 word2vec 2D-視覺化散布圖—a 朋友附近

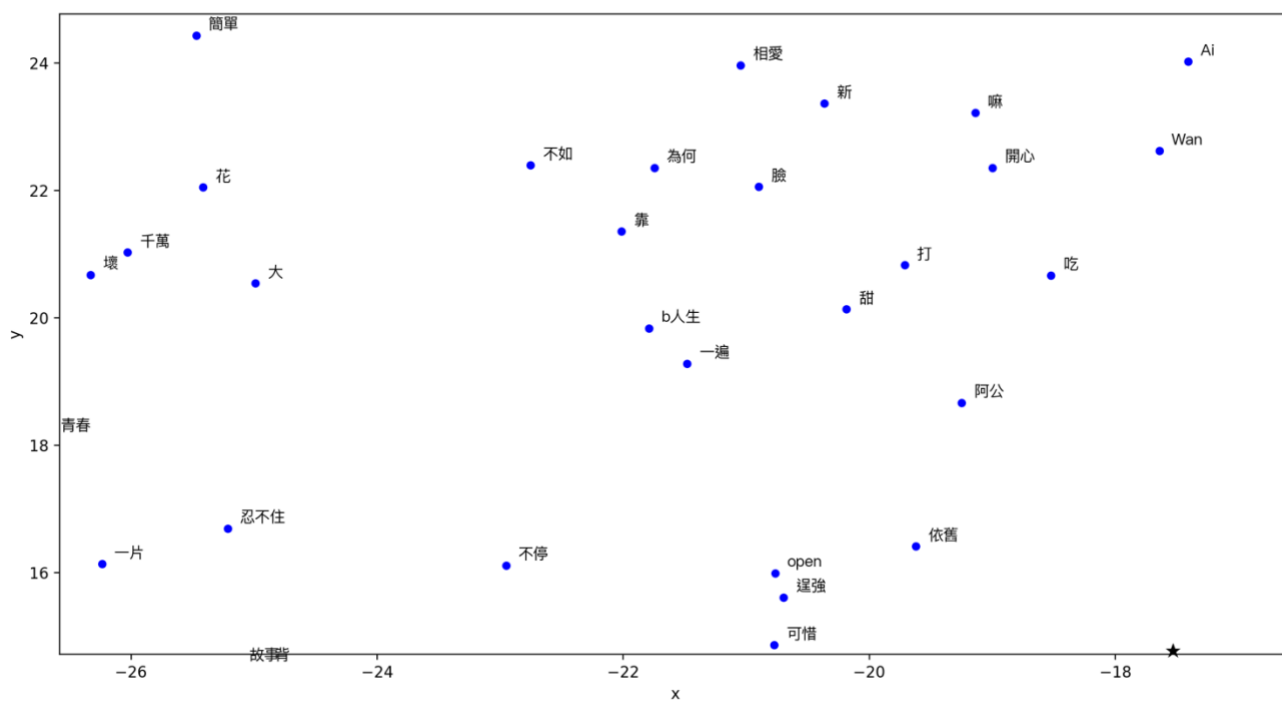


圖 17：「S.H.E」vs. 「田馥甄」 word2vec 2D-視覺化散布圖—b 人生附近

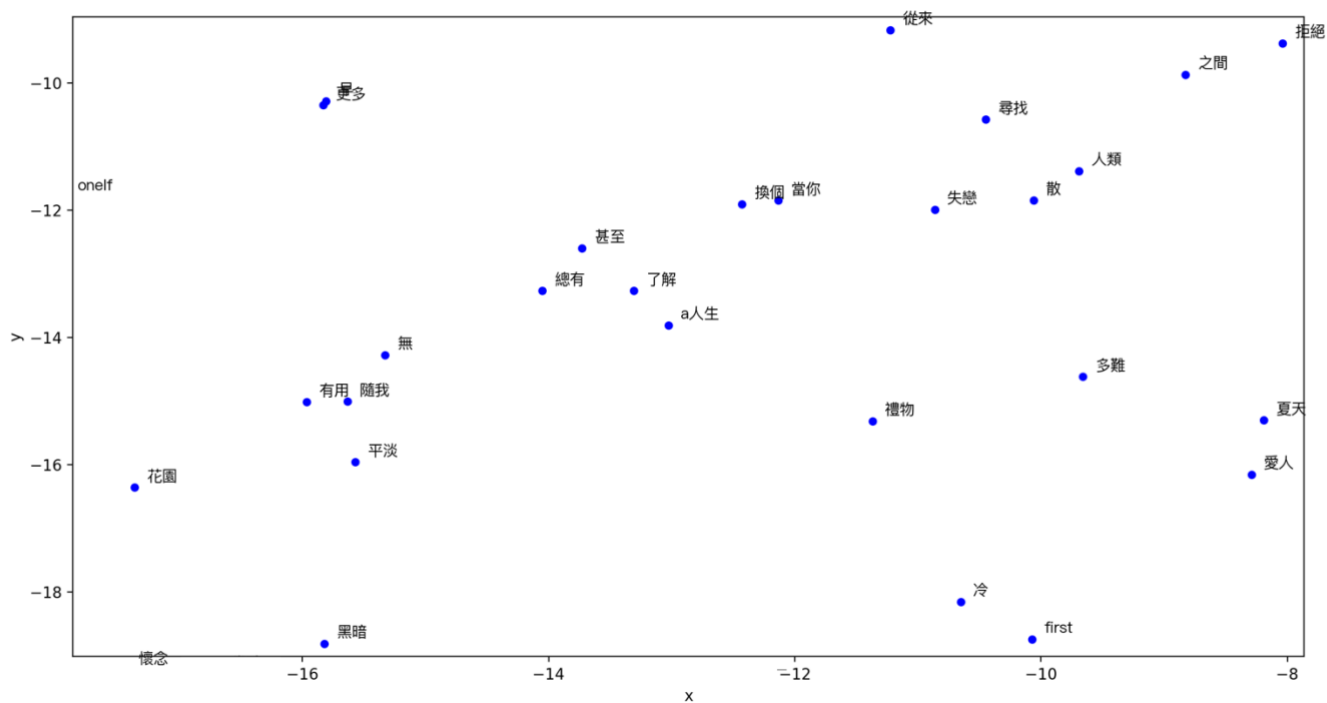


圖 18：「S.H.E」vs. 「田馥甄」 word2vec 2D-視覺化散布圖—a 人生附近

討論及限制

在詞向量空間的視覺化中，從 300 維的向量降至二維平圖時，難免會有一些偏差，所以從詞向量分布圖上可以看到，圖 15 與圖 16 中，「b 愛」與「a 愛」的位置其實非常相近，但餘弦相似度卻不如「信樂團」與「蘇見信」的「人生」組。

word2vec 的運作上由於加上了隨機的雜湊數，加上歌詞的資料量並不是很足夠，以目前調的參數來說，做出來的結果每次都會不盡相同，效果不是很好。

而即使以一個歌手來說，人也可能因為成長，或有了新的人生經驗，對事物有不同的看法，也會有風格上改變的可能性，要從這樣的結果看出是單飛的改變，或是歌手隨著時間的改變，可能還需要更多的數據和比較。