

《全唐詩》的分析、探勘與應用—風格、對仗、社會網路與對聯

Textual Analysis of Complete Tang Poems for Discoveries and Applications — Style, Antitheses, Social Networks, and Couplets

劉昭麟[†] 張淳甯[‡] 許筑婷[§] 鄭文惠[†] 王宏甦[§] 邱偉雲[‡]
Chao-Lin Liu Chun-Ning Chang Chu-Ting Hsu Wen-Huei Cheng Hongsu Wang Wei-Yuan Chiu

^{†§}國立政治大學資訊科學系、[†]國立政治大學語言學研究所

^{†‡}國立政治大學中國文學系、[§]美國哈佛大學 CBDB 計畫辦公室

{[†]chaolin, [‡]101703004, [§]104753021, [†]whcheng}@nccu.edu.tw,

[§]hongsuwang@fas.harvard.edu, [‡]acwu0523@gmail.com

摘要

唐詩是中國文學極重要的一部分，由清代官方力量所編纂、收錄兩千餘位詩人所著，內容四萬多首詩歌、包含超過三百萬字的《全唐詩》，無疑是研究唐詩最重要的資源之一。本文作者採取共現詞彙和 distributional semantics 的分析角度，利用計算語言學領域所發展的軟體工具，分析《全唐詩》的內容；就作者風格、詩歌內容，特別是唐詩中的顏色詞彙深入探索。同時我們也利用資訊技術，發掘唐詩內容所攜帶的唐代文人的社會網路，將研究成果擴大到歷史領域。另外，我們也藉由探勘唐詩中詞彙的共現、搭配、對仗關係，發展一個簡單的對對聯的應用。透過這一系列的工作，我們實踐了數位人文領域的初步理想，數位技術雖然尚且不足以直接被用來建立深度的人文論述，但是透過相關的資訊檢索、文本分析和資料整合的服務，數位技術讓專家可以比過去更加專注於深度議題的研究，而不需要花很多時間來蒐集基礎的研究資料。

Abstract¹

The *Complete Tang Poems (CTP)* is the most important collection for studying Tang poetry, which in turn is arguably a very influential part of the Chinese literature. Our analyzing the CTP from the perspectives of antithesis², collocation and distributional semantics offers some interesting overviews of the styles and imageries embedded in the works of some representative Tang poets. Our analyses include (1) a quantitative comparison of the uses of “wind” and “moon” in Li Bai’s and Du Fu’s works and (2) the functions of colors in Tang poems. In particular, we explored the appearances of “white” color, which is the most frequent color in Tang poems. Colors in static poems are like audios in motion pictures, so we thought the analyses could lead us to an important facet of the poems. In addition, we extracted social networks of poets from the poems, and built a simple couplet suggestion kit based on the textual analysis of the poems.

關鍵詞：數位人文、中國文學、全唐詩、詞彙語意、共現分析、文本分析、語料庫分析、中國歷代人物傳記資料庫

Keywords: Digital humanities, Chinese literature, Quan-Tang-Shi, Distributional semantics,

¹ A similar English version of this paper will be published in the proceedings of the 29th PACLIC conference (Liu et al. 2015).

² 英文的 antithesis 並不是對仗的完美翻譯，antithesis 除了帶有相對的意義、也帶有相反詞的意味，中文的對仗的平仄雖然要相反，但是在詞意上並不需要相反。

1 緒論

以單一文本集成來說，《全唐詩》³是研究唐詩的最集中的來源。《全唐詩》於清康熙時期完成編輯，據《欽定四庫全書》中〈御製全唐詩序〉所述，當時蒐集了超過二千二百位詩人的四萬八千九百餘首作品。唐詩在中國文學史中佔有極重要的地位，唐詩的風格與內容的影響延續至今，而《全唐詩》更是文學家、語言學家研究的重要文獻 (Fang et al. 2009, Lee and Wong 2012)。文學作品的分析是計算語言學(computational linguistics)文獻中相對少見的嘗試，在數位人文(digital humanities)於學術領域快速崛起的今天 (項潔及涂豐恩 2012)，我們嘗試以語文分析的角度來探索唐詩的內容。

羅鳳珠等學者(羅鳳珠等 1997, 羅鳳珠 2000)很早就提倡古籍的數位化工作，並且持續推動關於文學作品的數位化和相關應用。胡俊峰及俞士汶 (2001) 對於《全唐詩》與一批宋詞進行了基本詞彙分析，以建立深入分析的基礎，並且以愁、苦、恨、悲、哀、憂這一些單字詞的時序分析作為範例。儘管如此，我們還沒有能發現許多以數位技術處理唐詩作品的中文著述。蔣紹愚 (2003) 利用資訊檢索軟體與《全唐詩》相關資料庫的協助，比較了李白與杜甫作品中的“風”和“月”。

黃居仁等學者以《唐詩三百首》(Huang 2004) 加上羅鳳珠所整理的蘇軾作品(Chang et al. 2005)為基礎，研究蘇軾作品之中的本體論(ontology)。羅鳳珠繼續擴大建構詩詞之中的名詞分類體系(羅鳳珠 2008)，並且以所建構的分類體系作為賞析唐詩作品的基礎 (Fang et al. 2009)。

李思源等學者從計算語言學的角度切入唐詩的分析，針對唐詩中的詞類標記(Lee 2012)與語法相依(dependency trees) (Lee and Kong 2012)均有所論述。他們也從唐詩的一些特定詞彙類別，例如，名詞分類、季節、方向、顏色，為基礎來分析《全唐詩》中的作品 (Lee and Wong 2012)；並且整合唐詩的分析與計算語言學的教學(Lee et al. 2013)。

除了語言學教學之外，唐詩的內容分析也有一些現代化的應用。周明等學者利用《全唐詩》作為基礎語料，搭配統計式翻譯的技術 (Jiang and Zhou 2008, Zhou et al. 2009)，建立了一個自動對聯的「電腦對聯」系統⁴。在研究和比較台灣和中國的傳統和現代的中文詩詞作品時，Voigt 和 Jarafsky (2013) 也以唐詩作為研究材料。Chen (2010) 則從唐太宗的唐詩研究唐朝的政治問題。

在這一篇文章中，我們從許多方面探索唐詩的分析工作。我們嘗試以文學領域所熟知的唐詩一般構詞原則來擷取詞彙，因此沒有引入正統的中文斷詞、詞類標記和語法分析的技術。以此原則分析《全唐詩》所得詞彙為基礎，再進行以 distributional semantics (Harris 1954, Miller and Walter 1991) 為基礎的分析來觀察詩人的風格(style)。我們不僅分析詞彙的共現關係(collocation)，也利用律詩的對仗(antithesis)規則，分析了唐詩中顏色詞彙的使用情形。詩歌中的顏色可以比喻為電影中的聲光，對於詩歌所營造的意境有重要的影響。詩歌也傳遞了唐代文人之間的人際關係，可以透過詩歌的標題和內容，發

³ 《全唐詩》漢語拼音譯為 Quan-Tang-Shi，英文常翻譯為 Complete Tang Poems

⁴ 微軟亞洲研究院的「電腦對聯」系統：<http://couplet.msra.cn/>

掘文人之間的社會網路。最後我們以《全唐詩》為基礎，建構一個類似微軟亞洲研究院的「電腦對聯」系統，另將唐詩作品分為初唐、盛唐、中唐、晚唐詩人作品，做為尋找對聯詞彙時的基礎資料，藉此比較使用不同時期作品的影響。

我們在第二節說明所使用的《全唐詩》版本和一些相關的基本分析。在第三節，我們以 distributional semantics 的角度分析和比較李白與杜甫作品中的“風”和“月”，並且以“白”所構成的許多詞彙為基礎，例如“白雲”、“白頭”、“白玉”、“白馬”，觀察詩人用到這一些詞彙的比例以比較詩人的風格。在第四節中，我們以對仗規則為基礎，更加深入分析詩人作品中的顏色。在第五節中，我們利用人名和動詞的相關資訊發掘詩歌裡面所暗藏的人際關係。在第六節中，我們利用不同時期的七言唐詩，以作品中詞彙搭配狀況作為找尋對聯的匹配詞彙的基礎，建構了一個小規模的對對聯工具軟體。在第七節中，我們討論一些唐詩作者的問題；最後提出簡短的結語和檢討。

2 語料來源與基本分析

雖然文獻之中已經有一些關於全唐詩的論述，可是關於《全唐詩》的版本並沒有詳加敘述。依照《欽定四庫全書》中《御製全唐詩》的內容，《全唐詩》中的唐詩並沒有一個最權威的版本，而現在公開可得的《全唐詩》版本與《御製全唐詩》的內容也不完全一致，因此我們先說明我們所處理的《全唐詩》版本和相關的基本分析。

2.1 語料來源

要以程式分析《全唐詩》的首要工作，就是獲得《全唐詩》的文字版本檔案。雖然近幾年國內推動數位典藏計畫，數位人文研究在國際學術界也已經受到極高的重視，但是要取得《全唐詩》的文字版本並不容易。

在網際網路上，有一些公開的版本，例如「維基文庫」⁵、「文學 100」⁶、「蕭堯藝文網界」⁷和「中國哲學書電子化計畫」⁸。透過中央研究院傅斯年圖書館網站⁹查詢，可以查到元智大學羅鳳珠教授的檢索服務¹⁰、故宮博物院所建置的「寒泉」系統¹¹和「中國古籍全錄」¹²三個來源。不過我們無法從元智大學的網站或者「寒泉」系統取得全部的《全唐詩》文本。

我們可以透過直接下載或者逐頁人工下載，取得「文學 100」、「蕭堯藝文網界」、「中國哲學書電子化計畫」和「中國古籍全錄」所公開的《全唐詩》版本。

2.2 版本問題

我們正在以程式進行比對，希望獲得一份有高度共識的版本。我們已經完成「文學 100」和「中國哲學書電子化計畫」兩版本的初步比對，以目前的結果來說，這一些可以取得

⁵ 維基文庫：<https://zh.wikisource.org/zh-hant/全唐詩>

⁶ 文學 100：<http://www.wenxue100.com/>

⁷ 蕭堯藝文網界：<http://www.xysa.com/>

⁸ 中國哲學書電子化計畫：<http://ctext.org/zh>

⁹ 中央研究院傅斯年圖書館網站：http://lib.ihp.sinica.edu.tw/pages/02-aboutfsn/af01-library_8.htm

¹⁰ 元智大學羅鳳珠教授的檢索服務：<http://cls.hs.yzu.edu.tw/tang/Database/index.html>

¹¹ 寒泉查詢系統：<http://210.69.170.100/s25/>

¹² 中國古籍全錄：<http://guji.artx.cn/>

表一、《全唐詩》最高頻率二字字串的頻率統計

二字字串	頻率	二字字串	頻率	二字字串	頻率	二字字串	頻率	二字字串	頻率
何處	1669	無人	881	青山	662	流水	550	落日	498
不知	1469	風吹	834	少年	634	回首	544	不如	497
萬里	1455	惆悵	780	相逢	629	可憐	539	歸去	496
千里	1305	故人	778	平生	597	如此	526	日暮	496
今日	1165	秋風	749	年年	593	白髮	520	不能	481
不見	1158	悠悠	740	寂寞	592	主人	517	別離	481
不可	1148	相思	733	黃金	589	今朝	516	何時	478
春風	1128	長安	722	天子	588	月明	515	此時	477
白雲	1108	白日	697	人不	587	從此	509	洛陽	476
不得	947	如何	687	天地	586	日月	508	天下	472
明月	896	十年	678	何事	579	行人	507	芳草	472
人間	890	何人	663	江上	553	將軍	499	歸來	471

唐詩的五言與七言絕句和律詩作品的斷詞，目前仍難完全自動進行。實際上，五言作品的斷詞大都有固定的規律（羅鳳珠 2005），採用 2+2+1 或者 2+1+2 的句法，例如“白日依山盡”¹⁴是“白日”+“依山”+“盡”的 2+2+1 的句法，而“感時花濺淚”¹⁵則是“感時”+“花”+“濺淚”的 2+1+2 的句法。七言的作品的句法，常見的則是 2+2+1+2 和 2+2+2+1，例如“晉代衣冠成古丘”¹⁶是“晉代”+“衣冠”+“成”+“古丘”，而“東風不與周郎便”¹⁷是“東風”+“不與”+“周郎”+“便”。

雖然有以上的規律可循，可是要精確統計唐詩裡面的詞彙的使用狀況，目前仍得依賴人工的確認，因此相當不容易。我們以 PAT Tree (Chien 1997) 的技術，先行統計《全唐詩》中常見的二字字串，再以「臺灣數位人文小小讚」¹⁸ 網站上的軟體工具統計其中高頻率的二字詞的出現頻率，可以得到表一的數據。表一裡面雖然大多是真實的二字詞，但是“人不”並不是真正的二字詞。在《全唐詩》裡面，跟在“人”之後的“不”大都是另一動詞的一部份，例如：“盡日傷心人不見”¹⁹和“雖病人不知”²⁰

3 風格比較

Distributional semantics (Harris 1954, Miller and Walter 1991) 方法中透過與一個詞彙周邊出現的詞彙的分佈狀況來定義詞彙的語意。這種以語境來定義詞彙語意的想法由來已久，Firth (1957) 所說的“You shall know a word by the company it keeps”是經常被研究者提及的簡要說明。

¹⁴ 王之渙〈登鶴雀樓〉：白日依山盡，黃河入海流。欲窮千里目，更上一層樓。

¹⁵ 杜甫〈春望〉：國破山河在，城春草木深。感時花濺淚，恨別鳥驚心。烽火連三月，家書抵萬金。白頭搔更短，渾欲不勝簪。

¹⁶ 李白〈登金陵鳳凰臺〉：鳳凰臺上鳳凰遊，鳳去臺空江自流。吳宮花草埋幽徑，晉代衣冠成古丘。三山半落青天外，二水中分白鷺洲。總為浮雲能蔽日，長安不見使人愁。

¹⁷ 杜牧〈赤壁〉：折戟沉沙鐵未銷，自將磨洗認前朝。東風不與周郎便，銅雀春深鎖二喬。

¹⁸ 臺灣數位人文小小讚：<https://sites.google.com/site/taiwandigitalhumanities/>

¹⁹ 李商隱〈遊靈伽寺〉：碧煙秋寺泛湖來，水打城根古堞摧。盡日傷心人不見，石榴花滿舊琴臺。

²⁰ 白居易〈讀史五首〉：含沙射人影，雖病人不知。巧言構人罪，至死人不疑。

表二、《全唐詩》中李白所使用的月(頻率大於二者)

詞彙	頻率	詞彙	頻率	詞彙	頻率	詞彙	頻率	詞彙	頻率
明月	57	溪月	9	有月	5	湖月	3	夜月	3
秋月	40	八月	9	轉月	4	漢月	3	夕月	3
五月	28	雲月	9	曉月	4	樓月	3	喘月	3
日月	23	花月	8	孤月	4	新月	3	向月	3
海月	14	見月	7	台月	4	待月	3	古月	3
上月	13	江月	6	落月	3	弄月	3	十月	3
三月	13	蘿月	5	片月	3	如月	3	二月	3
山月	10	素月	5	滿月	3	好月	3	乘月	3

表三、《全唐詩》中杜甫所使用的月(頻率大於二者)

詞彙	頻率	詞彙	頻率	詞彙	頻率	詞彙	頻率	詞彙	頻率
日月	20	明月	7	落月	4	正月	3	從月	3
歲月	14	江月	6	秋月	4	星月	3	九月	3
十月	10	五月	6	漢月	4	新月	3		
三月	9	夜月	5	門月	3	四月	3		
八月	8	二月	5	素月	3	六月	3		

表四、《全唐詩》中李白所使用的風(頻率大於三者)

詞彙	頻率	詞彙	頻率	詞彙	頻率	詞彙	頻率	詞彙	頻率
春風	72	松風	17	南風	8	悲風	6	高風	4
清風	28	隨風	14	北風	8	飄風	5	西風	4
秋風	26	香風	11	涼風	8	胡風	5	扶風	4
東風	24	天風	10	狂風	7	從風	5	屏風	4
長風	22	英風	8	雄風	6	巖風	5	動風	4

表五、《全唐詩》中杜甫所使用的風(頻率大於三者)

詞彙	頻率	詞彙	頻率	詞彙	頻率	詞彙	頻率	詞彙	頻率
秋風	30	朔風	8	高風	6	江風	4	南風	4
春風	19	微風	8	清風	6	驚風	4	涼風	4
北風	14	隨風	7	天風	6	山風	4	東風	4
悲風	10	回風	7	長風	5	多風	4		
裡風	8	臨風	7	陰風	4	含風	4		

以類似的道理，我們研究和比較詩人使用某一特定類別詞彙的整體表現，再以這些詞彙的使用狀況充作定義詩人風格的“語境”，我們以“You shall know a poet's style by the words s/he uses”，亦不失為一種 distributional semantics 的創新詮釋。

3.1 李白與杜甫的風月

從文學專業來看，詩人的風格是多面向，不能只看一個面向以管窺天。比較詩人使用特定詞彙的不同方式，提供了比較詩人風格的一種選項。北京大學中文系教授蔣紹愚(2003)利用檢索系統找出李白與杜甫用到“風”和“月”的作品，然後透過人工分析與專業精讀和鑒賞來比對兩位著名詩人的風格。

除了基本的資訊檢索之外，自然語言處理技術提供了許多基本分析和進階分析的機會，可以進一步提供不同的分析角度。我們可以計算詩人一些特定的二字詞的使用狀況，然後以 distributional semantics 分辨詞彙語意的類似觀點，比較兩位詩人使用“風”和“月”的統計數據來比較兩位詩人的風格。這樣的分析角度，不同於個別詩歌特例的比較，而提供一個比較是計算語言學的分析角度。

在「文學 100」版本的《全唐詩》中，李白和杜甫分別有 896 和 1158 首作品。表二和表三列示這兩位詩人使用“月”的方式中頻率大於兩次者的資訊。雖然我們可以逐一把這一些作品列出來，然後精讀，但是光看這兩個表格的內容，就相當區分了李白和杜甫使用“月”的風格。李白的“月”明顯地有較多的變化，講到多種“月”的樣貌。相對地，杜甫經常講到月份，一月到十月之中只有七月沒有列入表三。

表四和表五列示這兩位詩人使用“風”的方式中頻率大於三次者的資訊。李、杜兩人使用“風”的狀況相仿，不再像使用“月”的狀況的差異。如果“春風”比較適用於喜悅的氣氛、而“北風”比較適用於較為悲傷的氣氛；表四與表五中兩位詩人最常用的前五名的“風”的數據突顯了兩位詩人筆下的“風”的明顯差異。

以上兩位詩人使用“風”與“月”的統計，多少呼應了一般人將李白歸類為浪漫詩派，而將杜甫歸類於社會詩人。

3.2 《全唐詩》的白色詞彙

“白”是《全唐詩》中最常見的顏色字，以“白”為基礎可以找到諸如“白日”、“白髮”、“白雲”、“白頭”、“白玉”、“白馬”、“白帝”、“白露”、“白石”等白色詞彙。表六列出了 46 個這樣的白色詞彙。我們可以利用這一些白色詞彙，從不同角度來比較表六裡面 13 位詩人的特色。

表六是一個複雜且巨大的表格，因此必須放置於本文末尾；該表以詩人的姓名作為橫軸，以白色詞彙作為縱軸。個別白色詞彙的左側，列出這一些 13 位詩人用到該詞彙的總和頻率。由上而下，白色詞彙是以他們出現在這 13 位詩人的作品中的總合頻率來排序。個別詞彙(以 T 代表)的右側與個別詩人姓名(以 N 代表)之下所列的數字，是詩人 N 作品中使用 T 的比例。以李白為例，李白有 896 份作品，其中出現過 62 次的“白日”，因此李白的作品中有 6.92%的機會看到“白日”。表六的大部份數據，都是詩人(N)作品中看到某一詞彙(T)的比率。

針對單一詩人之下，被標記為紅色²¹的數字者，是該詩人最常用的白色詞彙的比率。藍色標示者是第二、第三或者第四常用的白色詞彙，有一些詞彙因為有同樣的比率，因此會有多於一個紅色數字或者多於三個藍色數字的情況。

這一些數據可以用來找尋詩人特殊的取向。例如，賈島的作品之中有幾乎 1%會用到“白衣”，這是 13 位詩人之中的特例。杜甫使用“白帝”²²的比率非常高，甚至超過李白，背後的緣由特別值得探索。王維一般被歸類為田園詩人，其作品有 7.41%用到“白雲”，也算是名符其實的一項表徵。

在表六的上方有兩項比率，比率 A 是詩人作品中有白色詞彙的比率的總和。這一項總和是直接把詩人姓名之下的所有數字全部加總，並沒有考慮到某一些作品可能同時有多個白色詞彙時的交集情況，因此總和比率有可能高於實際的比率。不過詩人作品之中，通常講究避免重複相同的字，因此我們預期實際的誤差不大。

因此，比率 A 是個別詩人作品中使用到白色詞彙的總和比率。這其中最值得注意的是李白，李白的作品之中超過 46%使用到“白”字。

²¹ 為了方便黑白印刷時閱讀，不管是紅色或者藍色標示的數字，都搭配了粗線條的框線。

²² 唐詩中的大都是用於“白帝城”地名

比率 B 的計算跟比率 A 類似，但是並不是把所有詩人名字之下的所有數字加總。比率 B 只包含“白髮”、“白頭”、“白首”、“白鬚”、“白骨”、“白髭”這一組可以分析文人心理情緒的核心詞彙。表六把這六個詞彙的相關都加上淡淡的底色以便查詢。以李商隱為例，1.8%是來自於 0.54%、0.72%、0.18%、0.18%、0.18%、0.00%的總和。

當我們說一位詩人屬於某一詩派，應該是說詩人的作品呈現某一類的意境的比率較高，而不會說詩人的每一首詩總是有某一類的意境。如果我們認定以上這一組六個白色詞彙是偏向於負面情緒的詞彙的話，觀察個別詩人比率 B 的大小，我們可以發現孟浩然、李商隱和溫庭筠三人的比率 B 都明顯地低。這一些數據可以用來支持孟浩然經常被歸類為田園詩派；而李商隱和溫庭筠的評論則有“多綺麗濃艷之作”(李瑋質 2009)²³。杜甫和白居易的比率 B 則明顯高出其他人，這跟兩人常被分為社會詩派應該也有相當的關聯。

4 共現與對仗

共現(collocation)分析是計算語言學領域的基本分析，計算在特定文字範圍之內，詞彙一起出現的狀況。除了一起出現之外，對仗的詞彙有空間上的額外限制；對仗的詞彙必須位於所對仗的兩句中相同的相對位置。“白日當空天氣暖，好風飄樹柳陰涼”²⁴這一句對中，“白日”和“好風”是一組對仗的詞彙。在 15 個字的範圍之內，“白日”和“樹柳”雖然共現，但是兩者沒有對仗關係。在律詩之中，第三句和第四句(又稱第三聯)必須對仗，第五和第六句也必須對仗。

除了詞彙在句子中相對位置的限制之外，對仗的兩個詞彙還有聲韻上的限制，受到詩歌的平仄韻律的限制²⁵。現代漢語的發音並不足以了解唐詩中詞彙的平仄關係，有許多漢字的現代發音已經和古代不同，韻書、例如《唐韻》、《廣韻》、《平水韻》是研究古代漢字發音的重要參考工具。

詞彙的共現和對仗關係提供了文字探勘(text mining)的機會，我們可以從既有的唐詩之中抽取有助於詩歌創作的資訊。

4.1 共現與對仗詞組

針對我們有興趣研究的詞彙，例如“白雲”；我們可以先擷取“白雲”的語境，例如前後 n 個字，或者是一首作品。然後利用 PAT Tree 或者適當技術，找出這一些語境中的有意義詞彙，利用程式計算《全唐詩》中與“白雲”前後 n 個字範圍之內出現的詞彙的頻率。

表七列示依照這樣程序，將 n 設定為 30²⁶，所找到的一些跟“白雲”、“白日”和“白髮”共現的詞彙和頻率。這類的資訊可有一些可能的應用。例如可以用於詩歌的賞析教學，也可以用於詩歌創作的課程。如果可以搭配一些類似 E-HowNet²⁷所提供的詞彙的

²³ 多個評論來源的一般說法，例如，宜蘭縣教育支援平台、新北市樟樹詩詞網：

http://ostube.tctes.ntpc.edu.tw/poetry/index.php?option=com_content&view=article&id=51&Itemid=63

²⁴ 元稹〈清都春霽，寄胡三、吳十一〉：蕊珠宮殿經微雨，草樹無塵耀眼光。白日當空天氣暖，好風飄樹柳陰涼。蜂憐宿露攢芳久，燕得新泥拂戶忙。時節催年春不住，武陵花謝憶諸郎。

²⁵ 元智大學羅鳳珠教授提供的簡要原則：<http://cls.hs.yzu.edu.tw/300/all/primary1/DET4.htm>

²⁶ 透過我們所建構的軟體工具，這是一個可以自行改變的數字，可以變大、也可以變小。

²⁷ 中研院廣義知網：<http://ehownet.iis.sinica.edu.tw/>

表七、《全唐詩》中所有作品的一些共現詞組頻率 (n=30)

白雲				白日		白髮			
詞彙	頻率	詞彙	頻率	詞彙	頻率	詞彙	頻率	詞彙	頻率
明月	61	清露	10	青春	32	青山	38	丹砂	7
流水	40	青壁	7	青山	21	青雲	27	黃河	6
芳草	29	秋草	7	清風	18	朱顏	16	清光	4
滄海	28	丹灶	5	紅塵	15	青春	15	丹霄	4
紅葉	17	青鏡	2	黃河	15	黃金	13	黃衣	3
黃葉	16	青玉	2	滄江	6	滄洲	8	紅塵	3
青草	14	皇道	1	青蓮	3	青衫	7	紅旗	3
				青霄	3				
				青楓	2				

taxonomy 關係，我們還有可能可以研究不同作者的作品共現和對仗詞組的使用是否攜帶個人風格的資訊。

把計算共現詞彙的程式稍加改進，多考慮律詩中對仗的規則，就可以用來擷取唐詩中的對仗詞組。以下是一些關於白色詞彙的對仗例子。抽取完整的詩歌作品，可以讓一般人欣賞詩歌，也有助於專業研究人員找到所要研究的特定語料。

白居易〈北窗閑坐〉

虛窗兩叢竹，靜室一爐香。門外紅塵合，城中白日忙。
無煩尋道士，不要學仙方。自有延年術，心閑歲月長。

盧綸〈九日奉陪侍郎登白樓〉

碧霄孤鶴發清音，上宰因添望闕心。睥睨三層連步障，茱萸一朵映華簪。
紅霞似綺河如帶，白露團珠菊散金。此日所從何所問，儼然冠劍擁成林。

元稹〈清都春霽，寄胡三、吳十一〉

蕊珠宮殿經微雨，草樹無塵耀眼光。白日當空天氣暖，好風飄樹柳陰涼。
蜂憐宿露攢芳久，燕得新泥拂戶忙。時節催年春不住，武陵花謝憶諸郎。

在特別針對中唐詩人作品的研究中，我們(鄭文惠等 2015)更進一步分析哪一些對仗是普遍地為詩人所採用？例如，“白髮”和“青山”的對仗出現在盧綸、司空曙、李端、白居易、耿漳、賈島和顧況的作品中。“白雲”和“流水”的對仗出現在劉禹錫、姚合、皇甫冉、皇甫曾、賈島和錢起的作品中；“白雲”和“青草”的對仗則出現在劉長卿、司空曙、姚合、張籍、李端和郎士元的作品中。

對仗是比較嚴格的共現關係，但兩者都可以是詩人用來營造意象的方式。下面這一首李嘉佑的〈題游仙閣白公廟〉中，雖然“白雲”和“青山”只有共現關係，但是藉由“荔”、“竹”、“風”、“雨”、“流水”、“青山”、“焚香”和“白雲”的共同出現，這一些詞彙營造了一個特殊的氛圍。

仙冠輕舉竟何之，薜荔緣階竹映祠。甲子不知風馭日，朝昏唯見雨來時。
霓旌翠蓋終難遇，流水青山空所思。逐客自憐雙鬢改，焚香多負白雲期。

4.2 發掘唐詩的顏色關聯

顏色在詩歌中的角色可以譬如電影中的配樂，找出詩歌之中的顏色，可以建構一些詩人給詩歌上色的分析基礎。我們在分析李白和杜甫的作品時，就發現這兩位詩人最常用到顏色就是“白”，因此以“白”作為基礎來尋找《全唐詩》之中的其他顏色。

表八、《全唐詩》中顏色的搭配統計

白		青		紅		黃		綠		紫		碧		丹		赤		黑	
顏色	頻率	顏色	頻率	顏色	頻率	顏色	頻率	顏色	頻率	顏色	頻率	顏色	頻率	顏色	頻率	顏色	頻率	顏色	頻率
青	919	白	919	白	358	白	505	紅	335	青	197	紅	199	白	142	青	54	青	36
黃	505	綠	202	綠	335	青	152	青	202	黃	139	青	188	紫	70	黃	39	黃	27
紅	358	紫	197	碧	199	紫	139	黃	83	紅	107	清	100	碧	50	白	33	紅	24
清	274	碧	188	翠	139	綠	83	白	70	白	72	黃	74	青	41	紫	19	白	15
丹	142	黃	152	青	111	碧	74	清	70	丹	70	白	57	翠	35	蒼	15	明	13
滄	99	紅	111	紫	107	紅	44	丹	31	清	56	丹	50	綠	31	紅	13	清	10
朱	97	翠	54	黃	44	赤	39	朱	27	朱	41	金	42	玉	29	滄	12	丹	8
明	96	赤	54	清	36	翠	33	紫	26	金	39	紫	35	素	25	丹	10	寒	8
綠	70	明	42	素	31	清	32	碧	26	碧	35	朱	31	金	21	清	8	紫	7
玄	66	丹	41	金	21	黑	27	金	23	玄	32	寒	22	清	17	朱	7	赤	7

以“白”構成的雙字詞，加上考慮唐詩中律詩的對仗規則，我們可以透過與白色對仗的詞彙找到其他顏色，例如“白雲”和“丹井”在劉長卿的〈過包尊師山院〉的對仗²⁸；在岑參的〈號中酬陝西甄判官見贈〉²⁹則有“白髮”和“青雲”的對仗。“白髮”和“青雲”是《全唐詩》中相對常見的對仗，合計出現過 26 次。

透過以上這一程序，我們從《全唐詩》中找到諸多顏色，例如，“朱”、“丹”、“紅”、“緋”、“彤”、“青”、“翠”、“碧”、“綠”、“蒼”、“清”、“紫”、“玄”、“皂”、“黑”、“漆”、“明”、“黃”、“金”、“銀”，其中“白”出現最多次。這一些並不是《全唐詩》中所有的顏色字，例如、從“白”的對仗出發就不容易找到一樣是代表白色的“素”、“皓”、“皚”。

以這一些顏色作為基礎，我們就可以進行唐詩顏色的相關研究，例如不同顏色的搭配關係。我們以類似對仗的規則來找尋詩歌之中位置相對的顏色詞彙，位置相對是對仗的條件之一，可是嚴格的對仗還需要符合其他條件，這一些條件不容易完全自動化。表八的頻率統計只考慮到顏色詞的對仗位置，但是沒有檢查顏色詞彙是否出現在絕句、律詩、排律或者古詩，也沒有檢查平仄關係³⁰。

表八分為十欄，每一欄以一個顏色作為標題，並且再細分成兩個子欄，左邊的子欄是標題顏色的搭配顏色，右邊子欄則是兩個顏色的搭配頻率。這一個表格的資料是以《全唐詩》的全部內容做為計算基礎，但是限制所考慮的作品的每一句話必須有相同字數。表八的數據可以用於研究《全唐詩》中顏色構圖的一些基礎，例如用於協助找到更細部的語料，以連結顏色的運用和詩歌中的情感分析（鄭文惠等 2015）。

5 社會網路分析

我們可以從《全唐詩》中擷取詩人名單，然後分析姓名出現在詩人作品的標題和內容的狀況，以作為分析唐代詩人社會網路的基礎。舉例來說，我們可以發現李白在自己的作

²⁸ 劉長卿〈過包尊師山院〉：「漱玉臨丹井，圍棋訪白雲」

²⁹ 岑參〈號中酬陝西甄判官見贈〉：「白髮徒自負，青雲難可期」

³⁰ 這樣的作法雖然會引入一些錯誤，但是通常數量不致大到影響所觀察到的趨勢。杜甫的〈春日憶李白〉是一首五言律詩，其中首聯寫道「白也詩無敵，飄然思不群」。這裡的“白”並非顏色，而是指李白。如果不區分絕句律詩、首聯末聯、字面詞意，則會誤認“白”對到了“飄”。儘管如此，如果目標是要找尋“白”所對應的顏色字的話，這樣的錯誤並不會引起問題；特別是所對應的其他顏色的頻率夠高時。

品之中提到自己：「雖為李白婦，何異太常妻」、「李白乘舟將欲行，忽聞岸上踏歌聲」和「舒州杓，力士鎗，李白與爾同死生」³¹。

詩人的作品標題和作品內容可能提到他人的姓名。《全唐詩》中，至少八位詩人的15首詩歌，提到李白。其中杜甫占了七首³²。羅隱的作品之中也提到他對於杜甫的評論「洛陽賈誼自無命，少陵杜甫兼有文」和「杜甫詩中韋曲花，至今無賴尚豪家」³³。

詩人的關係可以是多面向的，詩人提到他人的時候，也不見得是直接字面提到；詩歌之中提到的姓名，也不見得屬於同一時代的人。戴叔倫在〈過賈誼宅〉有「上書憂漢室，作賦吊靈均」³⁴，透過“靈均”輾轉提到了戰國時代的屈原，而賈誼其實是西漢時代的人；賈誼因為遭到貶官到長沙，而著有〈弔屈原賦〉。劉長卿的〈長沙過賈誼宅〉是以人物姓名和地點名稱營造意象的另一個代表作。

如果是要分析詩人作品提到他人這一類的關係，只要運用名稱擷取就可以完成基本的工作。但是要進行完整、深度的社會網路分析，必須能夠妥善處理上述問題。人物的字號別名部分，需要依賴歷史人物的傳記資料庫，例如哈佛大學的「中國歷代人物傳記資料庫」³⁵。而要瞭解透過相似背景或者類似事件的人物地點所營造的意象，則需要專家專業知識或者非常高階的知識庫來支援。如果同一時代有相同名號的詩人，則更要有姓名分辨(person-name disambiguation)的工作要作。

除了以姓名字號作為發掘社會網路的起點之外，也可以利用適當的動詞來找詩人的人際關係的資訊。例如，在詩歌的標題中找尋“賜”，就可以找到許多唐朝皇帝賜詩的對象，其他的贈與的動詞也可以是很好的線索，例如“送”。以此類視角出發我們可以找到李世民的〈賦秋日懸清光賜房玄齡〉、〈賜蕭瑀〉、〈賜房玄齡〉、〈賜魏徵詩〉，李隆基的〈賜道士鄧紫陽〉、〈集賢書院成，送張說上集賢學士，賜宴得珍字〉、〈賜崔日知往潞州〉，李亨的〈賜梨李泌與諸王聯句〉、李昂的〈上巳日賜裴度〉等。其他如“讀”、“寄”、“見”“懷”、“憶”、“夢”、“贈”等，都是有用的動詞。

要表列這一些有用的動詞，可以依賴專家的專業知識，也可以利用詞夾子的技術(張尚斌 2006)，從一些已知的人名資訊來著手，找到人名之前的相關動詞。從動詞著手所獲知的人名，不見得是詩人的名字，例如魏徵和房玄齡就是政治家，這一類的資訊跟前面所提到的方式所找到的人際關係有互補的效用。如果我們在唐詩之中搜索「中國歷代人物傳記資料庫」中的唐代人物資料，或許也可以發現一些先前所未知的人際關係。

6 對聯

對對聯是一種比較貼近民間的文藝活動，以唐詩中對仗或者搭配的詞彙來對對聯，是一

³¹ 這三片段分別來自李白的〈贈內〉、〈贈汪倫〉和〈襄陽歌〉。

³² 這15首詩歌包含杜甫的〈贈李白〉(第216、224卷各一首)、〈送孔巢父謝病歸游江東，兼呈李白〉、〈夢李白二首〉、〈春日憶李白〉、〈冬日有懷李白〉、〈天末懷李白〉、任華的〈寄李白〉、白居易的〈李白墓〉、項斯的〈經李白墓〉、鄭穀的〈讀李白集〉、徐鉉的〈寄饒州王郎中效李白體〉、齊己的〈讀李白集〉、許宣平的〈見李白詩又吟〉、作者不詳的〈李白名許雲封謎〉

³³ 這兩片段分別來自羅隱的〈湘南春日懷古〉和〈寄南城韋逸人〉。

³⁴ 戴叔倫〈過賈誼宅〉

³⁵ 中國歷代人物傳記資料庫 <<http://isites.harvard.edu/icb/icb.do?keyword=k35201>> 是哈佛大學 Peter K. Bol (包弼德) 教授所主持的研究計畫所提供的開放資料庫。

個有趣的可能應用。我們以 128 位中唐詩人的七言作品當作基礎語料，嘗試回答填空式的對對聯問題。我們可以從網路上找到一些通俗的對聯資料³⁶，下面是一道填空式的對對聯問題，空白底線之上需要填入適當文字。

上聯：楊柳染綠芳草地；下聯：__ __ 映紅豔陽天

在這一問題中，我們首先從唐詩的資料中找到與“楊柳”搭配的詞彙。主要的方法就是去找唐詩之中，用來與“楊柳”搭配或者對仗的詞彙，例如，「楊柳青青鳥亂吟，春風香靄洞房深」和「故人相憶僧來說，楊柳無風蟬滿枝」³⁷。以這一個簡化的例子來說，我們先檢查“春風”和“故人”的平仄³⁸是否和“楊柳”搭配，如果不能搭配，則會被排除。以此例來說“春風”和“故人”皆能和“楊柳”搭配，便再計算唐詩中“春風”和“故人”的頻率。然後推薦兩者頻率較高者。因此，推薦的下聯是「春風映紅豔陽天」。

以上的方法不見得可以在基礎語料之中找到所有的詞彙。中唐詩人的作品中，僅僅在「國泰事留侯，山春縱康樂」³⁹提到“國泰”，可是這是一首五言作品，所以我們目前的推薦機制認定中唐詩人沒有用過“國泰”。因此，處理用到“國泰”的對聯時，我們先找到基礎語料中分別與“國”和“泰”對應的單一漢字群，例如“國”對應到{“青”，“陽”}、“泰”對應到{“山”，“春”}。然後把這一些漢字群組合成一些二字詞，得到“青山”，“陽山”“青春”，“陽春”。我們同樣會檢查這一些候選項目的平仄，然後再在基礎語料之中計算這一些二字詞的頻率，選擇其中頻率最高者。因此在處理下面這一對聯時，我們會推薦「青山兩順頌年華」。

上聯：國泰民安達盛世；下聯：__ __ 兩順頌年華

以上這樣的對對聯的機制，並不如周明等學者所採用的機器翻譯機制複雜(Jiang and Zhou 2008, Zhou et al. 2009)，程序簡單許多，但有不同的成效，可以找到不同意境的建議詞彙。針對上面兩個例子來說，微軟亞洲研究院的「電腦對聯」系統，首選的推薦分別是「畫眉映紅豔陽天」和「風調兩順頌年華」⁴⁰。這兩組對聯原本是分別用「桃李映紅豔陽天」和「風調兩順頌年華」。

以上面第一組對聯來說，我們所建議的「春風映紅豔陽天」和「電腦對聯」系統所建議的「畫眉映紅豔陽天」，各有自己可以想像的情境和缺點。「電腦對聯」系統是一個開發相當時日的系統，可以蒐集相當多的相關語料。“國泰民安”跟“風調雨順”是現代人常常用到的組合，而這樣的組合在唐詩之中並不會經常出現，所以光是靠唐詩的對仗頻率，不見得總是可以找到預想的組合。

改變基礎語料直接影響了詞彙頻率的計算基礎，因此也影響了所推薦的詞彙。如果我們改以盛唐詩人的作品作為基礎語料，則針對上面兩組對聯，我們的程式就會分別推薦「長安映紅豔陽天」和「青春兩順頌年華」，除了又是不同的想像意境之外，也反映了時代的背景。在唐代對上述的第一組對聯，使用“長安”或許比使用“桃李”要恰當。

³⁶ 對聯大全：<http://duilian.51240.com/>

³⁷ 分別出自沈宇的〈代閩人〉和賈島的〈酬姚合〉

³⁸ 元智大學羅鳳珠教授的網站：<http://cls.hs.yzu.edu.tw/300/all/primary1/DET3.htm>

³⁹ 盧綸〈奉陪侍中游石筍溪十二韻〉

⁴⁰ 微軟亞洲研究院的「電腦對聯」系統是提供多個有排序、可選擇的詞彙，並沒有直接建議特定詞彙。

7 唐詩作者

以電腦程式計算，在「文學 100」版本的《全唐詩》之中，合計收錄由 2517 位相異姓名的作者(包含“不詳”)；在「中國哲學書電子化計劃」版本的《全唐詩》之中，則表列 2523 位相異姓名的作者(包含“不詳”)。這一些作者數目比〈御製全唐詩序〉中所記錄的詩人人數還要多！

在「文學 100」版本的《全唐詩》之中，這些相異姓名有一些是在文字檔案中名字不全者，以第 204 卷中的〈懷素上人草書歌〉為例，有兩位可能的作者：在《全唐詩》中列為王顥作品，但是《御製全唐詩》注釋也說明可能是王邕的作品。在許多文字版本之中，過去因為沒有“顥”的電腦字形，所以就把“顥”以空白取代，產生了一位只是叫作“王”的作者。

另外有一些作品可能是以字號等的別名來記錄詩人。例如第 823 卷中的〈龍潭〉的作者是應物，在中唐詩人之中有一位韋應物，這兩人是否是同一人，需要一些考證的功夫。另外類似第 807 卷中的拾得和第 862 卷中的樵夫，這一些姓名是本名或者字號，也是需要求證的。

8 結語

我們利用語文分析工具從不同的角度分析了《全唐詩》的內容，以作為一些專業研究的基礎。我們透過觀察李白和杜甫筆下的“風”、“月”詞彙，來比較兩人的差異；也透過“白日”、“白髮”、“白雲”、“白頭”等許多白色詞彙來觀察多位詩人的風格。我們以分析詞彙的對仗、共現和一般搭配關係為出發點，進一步探討了唐代詩人運用顏色詞彙的一些現象。社會網路的分析讓我們有機會一窺唐代詩人的一些人際關係，而對對聯的應用則聯結了唐詩研究與現代生活。

在目前的工作中，我們明顯未能就唐詩之中的音韻面向進行深入分析。平仄押韻等相關的資訊，未能被完美處理會牽涉一些問題，例如、我們沒有能夠全面自動化地分辨古詩和近體詩。納入韻書的資訊是我們的當務之急。詩歌的情感分析需要對於文字有高度的敏感度，如果系統設計足夠細膩，電腦軟體還是有機會對於深度情感的分析有所貢獻(鄭永曉 2012)。

誌謝

本研究承蒙科技部人文及社會科學研究發展司數位人文專題計畫透過研究計畫 MOST-102-2420-H-004-054-MY2 與之補助，謹此致謝。由於論文頁數的限制，我們僅能在口頭報告時回應評審先進的寶貴問題和建議。

參考文獻

李瑋質 (2009) 《晚唐「溫李」作品對南朝宮體詩之承傳與創變》，國立中央大學，中國文學系，碩士論文；指導教授：王力堅。

- 胡俊峰及俞士汶 (2001) 唐宋詩之計算機輔助深層研究,《北京大學學報(自然科學版)》,37(5):725-733。
- 張尚斌 (2006) 詞夾子演算法在專有名詞辨識上的應用-以歷史文件為例,國立臺灣大學,資訊工程學系,碩士論文;指導教授:項潔。
- 項潔及涂豐恩 (2011)〈導論-什麼是數位人文〉,《從保存到創造:開啟數位人文研究》,項潔編,9-28,臺北:國立臺灣大學出版中心。
- 蔣紹愚 (2003) 李白杜甫詩中的"月"和"風"-電腦如何用於古典詩詞鑒賞,《第一屆文學與資訊科技國際會議論文集》。
- 鄭文惠、劉昭麟、許筑婷及邱偉雲 (2015) 情感現象學與色彩政治學:中唐詩歌白色抒情系譜的數位人文研究,《第六屆數位典藏與數位人文國際研討會論文集》。
- 鄭永曉 (2012) 情感計算應用於古典詩詞研究芻議,《科研信息化技術與應用》,3(4):59-66。
- 羅鳳珠、李元萍及曹偉政 (1997) 古詩詞研究的電腦支援環境的實現。《中文資訊學報》,1:27-36
- 羅鳳珠 (2000) 台灣地區中國古籍文獻資料數位化的過程與未來的發展方向,《五十年來台灣人文學術研究叢書---文獻學與圖書資訊學》,臺北:學生書局。
<<http://cls.hs.yzu.edu.tw/present/tarcf.htm>>
- 羅鳳珠 (2005) 詩詞語言詞彙切分與語意分類標記之系統設計與應用,《第四屆數位典藏技術研討會論文集》。
- 羅鳳珠 (2008) 植基於中國詩詞語言特性所建構之語意概念分類體系研究,《第九屆海峽兩岸圖書資訊學學術研討會論文集》。
- Chang, Ru-Yng, Chu-Ren Huang, Feng-Ju Lo and Sueming Chang (2005) From general ontology to specialized ontology: A study based on a single author historical corpus, *Proc. of the Workshop on Ontologies and Lexical Resources*, 16-21.
- Chen, Jack Wei (2010) *The Poetics of Sovereignty: On Emperor Taizong of the Tang Dynasty*, Harvard University Asia Center, 2010.
- Chien, Lee-Feng (1997) PAT-tree-based keyword extraction for Chinese information retrieval, *Proc. of the 20th Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*, 50-58.
- Fang, Alex Chengyu, Fengju Lo, and Cheuk Kit Chinn (2009) Adapting NLP and corpus analysis techniques to structured imagery analysis in classical Chinese poetry, *Proc. of the Workshop on Adaptation of Language Resource and Technology to New Domains*, 27-34.
- Firth, John Rupert (1957) A synopsis of linguistic theory 1930-1955, *Studies in Linguistic Analysis*, 1-32.
- Harris, Zellig (1954) Distributional structure, *Word*, 10(2-3):1456-1162.
- Huang, Chu-Ren (2004) Text-based construction and comparison of domain ontology: A study based on classical poetry, *Proc. of the 18th Pacific Asia Conf. on Language, Information and Computation*, 17-20.
- Jiang, Long and Ming Zhou (2008) Generating Chinese couplets using a statistical MT approach, *Proc. of the 22nd Int'l Conf. on Computational Linguistics*, 377-384.
- Lee, John (2012) A classical Chinese corpus with nested part-of-speech tags, *Proc. of the 6th EACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 75-84.
- Lee, John, Ying Cheuk Hui, Yin Hei Kong (2013) Treebanking for data-driven research in the classroom, *Proc. of the 4th Workshop on Teaching Natural Language Processing*, 56-60.
- Lee, John and Yin Hei Kong (2012) A dependency treebank of classical Chinese poems, *Proc. of the 2012 Conf. of the North Chapter of the Association for Computational Linguistics: Human Language Technologies*, 191-199.
- Lee, John and Tak-sum Wong (2012) Glimpses of ancient China from classical Chinese poems, *Proc. of the 24th Int'l Conf. on Computational Linguistics*, posters, 621-632.
- Liu, Chao-Lin, Hongsu Wang, Wen-Huei Cheng, Chu-Ting Hsu, and Wei-Yun Chiu (2015) Textual analysis of complete Tang poems for discoveries and applications: styles, colors, and social networks, *Proc. of the 29th Pacific Asia Conf. on Language, Information and Computation*.
- Miller, George and Walter Charles (1991) Contextual correlates of semantic similarity, *Language and Cognitive Processes*, 6:1-28.
- Voigt, Rob and Dan Jurafsky (2013) Tradition and modernity in 20th century Chinese poetry, *Proc. of the 2nd Workshop on Computational Linguistics for Literature*, 17-22.

Zhou, Ming, Long Jiang, and Jing He (2009) Generating Chinese couplets and quatrain using a statistical approach, *Proc. of the 23rd Pacific Asia Conf. on Language, Information and Computation*, 43–52.

表六、《全唐詩》中 13 位詩人所使用的白

比率 A		8.96	18.41	9.73	46.65	23.83	12.55	26.94	15.67	18.80	17.37	15.19	16.30	10.48
比率 B		1.87	5.72	1.80	5.92	2.13	4.66	7.94	1.99	2.28	7.19	4.54	3.70	3.23
詞類	詞彙	孟浩然	孟郊	李商隱	李白	李賀	杜牧	杜甫	溫庭筠	王維	白居易	許渾	賈島	韓愈
217	白日	0.75	4.73	1.62	6.92	2.98	1.01	2.42	0.00	1.14	2.04	1.18	2.22	3.23
164	白髮	1.12	3.73	0.54	2.34	1.28	1.62	1.99	0.00	0.85	2.50	1.58	2.22	0.54
158	白雲	2.99	1.99	0.54	3.79	0.85	1.42	0.86	0.28	7.41	0.95	2.96	4.44	0.27
149	白頭	0.00	0.75	0.72	0.67	0.43	2.23	3.37	1.14	0.57	2.23	2.37	0.49	1.61
86	白首	0.75	1.00	0.18	1.56	0.43	0.20	1.99	0.85	0.85	1.02	0.59	0.25	0.81
74	白玉	0.00	0.50	2.34	3.01	0.85	0.81	0.60	0.00	0.85	0.53	0.20	0.00	0.27
74	白馬	0.37	0.50	0.00	2.34	4.68	0.00	1.38	2.85	0.85	0.30	0.39	0.00	0.00
63	白雪	0.37	0.25	0.36	2.34	0.00	0.40	1.04	0.28	0.00	0.68	0.79	0.00	0.27
59	白帝	0.00	0.00	0.18	1.00	0.43	0.00	3.54	0.28	0.00	0.08	0.39	0.00	0.54
58	白露	0.00	0.50	0.18	1.56	0.43	0.00	0.86	0.28	0.28	0.79	0.39	0.99	0.27
54	白石	0.00	1.00	0.90	1.12	0.43	0.00	0.26	0.57	0.57	0.68	0.20	1.23	0.81
38	白蘋	0.37	0.75	0.18	0.22	1.28	0.20	0.52	2.85	0.00	0.30	0.20	0.00	0.54
32	白水	0.00	0.25	0.00	0.89	1.28	0.00	1.12	0.00	0.57	0.08	0.39	0.00	0.27
31	白蘋	0.00	0.00	0.18	0.11	0.00	0.20	0.00	0.00	0.00	0.98	0.00	0.49	0.00
30	白鷺	0.00	0.25	0.00	1.79	0.00	0.40	0.26	0.00	0.85	0.15	0.00	0.25	0.00
25	白壁	0.37	0.25	0.18	1.79	0.85	0.40	0.00	0.28	0.00	0.00	0.00	0.25	0.00
23	白楊	0.00	0.00	0.36	1.12	0.00	0.00	0.26	0.00	0.00	0.26	0.20	0.00	0.00
22	白蓮	0.00	0.00	0.00	0.00	0.00	0.40	0.00	0.57	0.00	0.61	0.39	0.00	0.00
21	白羽	0.37	0.25	0.00	0.67	0.00	0.20	0.52	0.28	0.85	0.08	0.00	0.00	0.00
21	白骨	0.00	0.25	0.18	1.23	0.00	0.00	0.60	0.00	0.00	0.00	0.00	0.00	0.27
19	白鷗	0.00	0.00	0.00	0.78	0.00	0.20	0.69	0.00	0.00	0.11	0.00	0.00	0.00
19	白屋	0.00	0.25	0.18	0.00	0.43	0.20	0.86	0.00	0.00	0.11	0.20	0.25	0.00
19	白鶴	0.37	0.25	0.00	0.33	0.00	0.00	0.43	0.00	0.57	0.15	0.39	0.00	0.27
19	素琴	0.00	0.00	0.18	0.78	0.00	0.00	0.00	0.57	0.28	0.19	0.39	0.25	0.00
19	素手	0.00	0.00	0.00	1.56	0.00	0.00	0.00	0.85	0.00	0.04	0.20	0.00	0.00
18	白浪	0.37	0.00	0.00	0.56	0.00	0.00	0.17	0.00	0.00	0.23	0.20	0.74	0.00
17	白衣	0.00	0.00	0.18	0.22	0.00	0.00	0.17	0.00	0.57	0.19	0.20	0.99	0.00
16	白鹿	0.00	0.00	0.00	0.89	1.28	0.20	0.09	0.00	0.00	0.11	0.00	0.00	0.00
15	白波	0.00	0.25	0.00	0.78	0.43	0.00	0.09	0.00	0.00	0.08	0.59	0.00	0.00
15	白鬢	0.00	0.00	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.45	0.00	0.25	0.00
15	皓齒	0.00	0.00	0.00	0.78	0.85	0.00	0.26	0.85	0.00	0.00	0.00	0.00	0.00
14	白沙	0.00	0.00	0.00	0.33	0.00	0.20	0.43	0.00	0.57	0.11	0.00	0.00	0.00
14	白鳥	0.00	0.00	0.18	0.00	0.00	0.61	0.35	0.28	0.28	0.08	0.00	0.49	0.00
14	白花	0.00	0.00	0.00	0.22	0.00	0.40	0.26	0.57	0.00	0.15	0.00	0.00	0.27
12	白社	0.75	0.00	0.18	0.00	0.00	0.20	0.00	0.57	0.28	0.08	0.39	0.25	0.00
12	白龍	0.00	0.25	0.00	0.89	0.00	0.00	0.00	0.28	0.00	0.08	0.00	0.00	0.00
12	白紵	0.00	0.00	0.18	0.89	0.00	0.00	0.00	0.28	0.28	0.04	0.00	0.00	0.00
12	白晝	0.00	0.00	0.00	0.11	1.70	0.20	0.09	0.00	0.00	0.15	0.20	0.00	0.00
11	白如	0.00	0.00	0.00	0.33	0.43	0.20	0.00	0.57	0.00	0.11	0.00	0.00	0.27
11	素書	0.00	0.00	0.00	0.33	0.43	0.00	0.26	0.00	0.00	0.08	0.20	0.25	0.00
10	素月	0.00	0.25	0.00	0.56	0.43	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00
10	白魚	0.00	0.00	0.00	0.00	1.28	0.00	0.52	0.00	0.00	0.04	0.00	0.00	0.00
10	白刃	0.00	0.25	0.00	0.56	0.00	0.00	0.26	0.00	0.00	0.04	0.00	0.00	0.00
10	素絲	0.00	0.00	0.00	0.33	0.43	0.00	0.17	0.28	0.28	0.08	0.00	0.00	0.00
10	白家	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00	0.00	0.00
10	白猿	0.00	0.00	0.00	0.89	0.00	0.20	0.00	0.00	0.00	0.04	0.00	0.00	0.00