

第九屆（2008年）海峽兩岸圖書資訊學學術研討會

2008年7月3-6日，武漢大學主辦

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

The research of building a semantic category system based on the
language characteristic of Chinese poetry

羅鳳珠，gefjulo@saturn.yzu.edu.tw

台灣元智大學中語系講師、台灣大學圖書資訊學研究所博士生、湖北大學客座教授

摘要

電腦處理有規則及合邏輯的資料，其效率高於人腦何止千萬倍，但是電腦只能分辨字形的符碼，無法解讀字詞含義的意碼，所以引用電腦做為偏重主觀抽象感知的詩詞文學教學研究輔助工具，比相對客觀的自然科學領域困難得多。

電腦縱然有強大的記憶與搜尋能力，因為只能分辨字形，無法理解語意內涵，不具有思考能力，不具有人的認知、感知能力，不具備人的知識體系，以之建構詩詞網站，做為詩詞研究或教學的輔助工具，無法滿足文學領域使用者的需求。

本文以中國古典詩詞資料庫為標的，分析詩詞語言特性，建立詩詞語意標記資料庫，使電腦從只能辨認字形提升到能理解語意。以語意標記資料庫，建構語意概念分類體系，除了能提升分類的準確性與完整性，也能使電腦從索引式的搜尋提升到更符合人類大腦以思維聯繫尋找資料的模式，使之更接近人的知識體系，以之建立全文檢索系統，從只能分辨字形的檢索，提升到分辨語義的檢索，提升全文檢索的準確性與完整性，滿足不同使用者的需求。

詩詞內容所描寫的元素包括「人、事、時、地、物」五種，本研究以此五類作為最上層的總類，另以「其他」概括無法包含在這五項元素之內的概念，以符合詩詞文學的內涵。總類之下再分為大類、中類。所建立的詩詞語意概念分類體系包含二部分：其一

是以詩詞作品詞彙經過詞彙控制所建立的「詩詞詞彙索引表」，其二是詩詞語意概念分類體系，以索引詞連結「詩詞詞彙」與「詩詞語意概念分類體系」二者之間的聯繫關係。本研究總計分為 54 大類，372 中類，與歷代類書相比，上下層級從二層增加到三層，大類新增 25 類，中類新增 73 類。

關鍵詞：中國古典詩詞、詩詞語言、語意標記、語意概念分類、詩詞語言索引典

全文

一、前言：

詩詞的本質與電腦存在著基本上的差異，電腦理性客觀，處理有規則而合邏輯的資料；詩詞感性主觀，講究「只可意會而不可言傳」（劉大魁〈論文偶記〉）、「不著一字而盡得風流」（司空圖《詩品》）、「言有盡而意無窮」（嚴羽《滄浪詩話·詩辨》），容許反理而合情。

詩詞所使用的語言與其他文體不同，北京大學中文系袁行霈教授在《中國詩歌藝術研究·原序》論及：「如果從語言學的角度給詩歌下一個定義，不妨說詩歌是語言的變形，他離開了口語和一般的文書語言，成為一種特異的語言形式。」¹。袁行霈教授的論述是從語意、語用及語境的角度提出詩歌語言與其他文體語言的不同，包含詩歌語言的多義性、意象與意境的創造等等。²葉聖陶從詩歌語言形式的角度，認為詩歌的語言純出於人工。³由此可知，無論從語意或語言形式的角度，詩歌非但與一般口語、文書語言不同，與其他文體也存在著差異，再加上詩歌的形式受到詩歌格律之平仄聲調及字數、句法上的限制，語言更精鍊，語意更豐富，修辭更講究，語境更複雜，創作時需要更多藝術的創造，解讀時需要更多聯想，解讀更困難。

電腦處理有規則及合邏輯的資料，其效率高於人腦何止千萬倍，但是電腦只能分辨字形的符碼，無法解讀字詞含義的意碼，所以引用電腦做為偏重主觀抽象感知的詩詞文學研究與教學的輔助工具，比客觀知識的自然科學困難得多。

¹. 袁行霈，《中國詩歌藝術研究》，五南圖書出版公司印行，1989年5月台灣初版，見〈原序〉第2頁。

². 參見袁行霈，《中國詩歌藝術研究》之〈中國古典詩歌的多義性〉，第3-24頁、〈中國古典詩歌的意境〉第25-56頁、〈中國古典詩歌的意象〉第57-72頁。

³. 參見王力，《王力詩論》之〈略論語言形式之美·詩的語言〉，引葉聖陶給王力的書信：「詩之句型，大別為二。一為平常的句型，與散文及口頭語言大致不異。一為特殊句型，散文絕不能如是寫，口頭亦絕無此說法，可謂純出於人工。」

文學的研究，涉及龐大的資料，詩詞亦然。在浩瀚的詩詞文獻中要具備旁徵博引，互相引證發明的能力，往往需要經年累月，皓首窮經才有可能。電腦強大的記憶能力及快速的搜尋檢索、分析統計能力，可以彌補人腦記憶與搜尋能力的不足。但是，電腦只能分辨字形形符，無法解讀字詞語意的意符。旁徵博引及引證發明涉及知識的內涵，而不是文字的字形，電腦縱然有強大的記憶與搜尋能力，因為只能以字形分辨，無法辨別語意內涵，不具有思考能力，不具有人的認知、感知能力，不具備人的知識體系，無法滿足使用者的需求。

本研究將以詩詞資料庫為例⁴，探討如何建構植基於語意學及符合詩詞語言特性的語意概念分類體系，使之更接近人的知識體系，滿足不同使用者的需求。

二、詩詞文體語言的特性與資訊行為的模式

就如詩評家及學者們所提出的觀點，詩詞是特殊的語言形式，以不同的修辭技巧達到精緻凝練、語意無窮，情境高遠、言意與形神兼備的境界。透過修辭的技巧，將作者主觀內化的情思意念，投注在客觀外顯的事物上，作為作者與讀者之間引發共鳴的溝通媒介。詩詞所使用的詞彙，有一些是客觀知識的描述闡釋，有一些是主觀情意的引導感發。以楊花為例，蘇軾〈少年遊〉：「去年相送，餘杭門外，飛雪似楊花。今年春盡，楊花似雪，猶不見還家。」描寫冬季飛雪與春末飛花的景象，將「楊花」與「飛雪」對比呈現，讓只見過其中一種物象的讀者，多出一項理解詞義的管道，這裡的楊花是純粹客觀的景物，不含作者的情意。同是寫楊花，蘇軾寫楊花的名作〈水龍吟〉：「似花還似非花，也無人惜從教墜。拋家傍路，思量卻是，無情有思。縈損柔腸，困酣嬌眼，欲開還閉。夢隨風萬里，尋郎去處，又還被鶯呼起。不恨此花飛盡，恨西園、落紅難綴。曉來雨過，遺蹤何在？一池萍碎。春色三分，二分塵土，一分流水。細看來、不是楊花，點點是離人淚。」多重意象交互呈現，幽居空閨的女子，自傷處境如同隨風飄盪的楊花，拋家傍路，無人憐惜，不知情歸何人，不知郎行何處，多少幽微的心事無處訴說，隨風飄落的片片楊花，看在傷心人眼中，就不再只是楊花，盡是點點離人淚，這是心境、情境、意境。⁵詞中的楊花，也跳脫了物象本身的含義，增加了作者所附加的情意，語意也擴充了，從楊花的例子可知，「飛雪似楊花」的「楊花」是形符，代表的是楊花的物象本義，但是「細看來、不是楊花，點點是離人淚。」的「楊花」是意符。

主觀情意的感發，往往因為讀者的學養、經驗、際遇、年齡、心理氣質、思想境界等各種背景差異而使感受各有不同，因此相同的詞彙，相同的作品，感受上存在者個別

⁴. 參見羅鳳珠、張智星、許介彥，【新詩改罷自長吟—唐詩吟唱虛擬實境教學環境建置計畫】，國科會數位典藏國家型科技計畫：95年度數位典藏創意學習計畫，2006年3月1日至2007年2月28日。計畫編號：NSC95-2422-H-155-001，網址：<http://cls.hs.yzu.edu.tw/tang/index.html>。

⁵. 參見拙作〈虛擬與實境的交錯：多媒體在文學教學的設計與應用〉，文學數位製作與教學研討會，成功大學主辦，2005年6月12日，收入《文學數位製作與教學》，五南圖書公司，P107~P130，2007年1月出版。

的差異。即便是同一個人，經過不同時空，重讀相同的作品，感受也會不同，這是客觀認知與主觀感知的問題。

資訊的瀏覽、搜尋、檢索等行為，牽涉到行為人的認知，Belkin, N. J. 在 1990 年發表的 “The cognitive viewpoint in information science”⁶ 提出以認識論為基礎的認知觀點；Ingwersen, P. 在 1992 年出版的 “Information Retrieval Interaction”⁷ 書中的 “The Cognitive View and Information”、” IR Interaction ?the Cognitive Turn” 二個章節論及認知觀點；在 1996 年發表的 “Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory”⁸ 以及 2001 年發表的 “Cognitive information retrieval.”⁹ 進一步將認知觀點與資訊檢索結合，提出資訊檢索的交互作用以認知理論的元素為基礎的觀點。Ingwersen, P. 以 “General model of cognitive information seeking and retrieval” 架構圖反映出對資訊搜尋和檢索 (IS&R) 的資訊搜尋之行為者或在搜尋情境裡的群組的認知過程，牽涉到認知和感知的表述，如人類認知、反應、感情、意見的表述，這些都是資訊搜尋和檢索構成的一部份，也是資訊交互影響情境的一種。¹⁰ 架構圖如下：

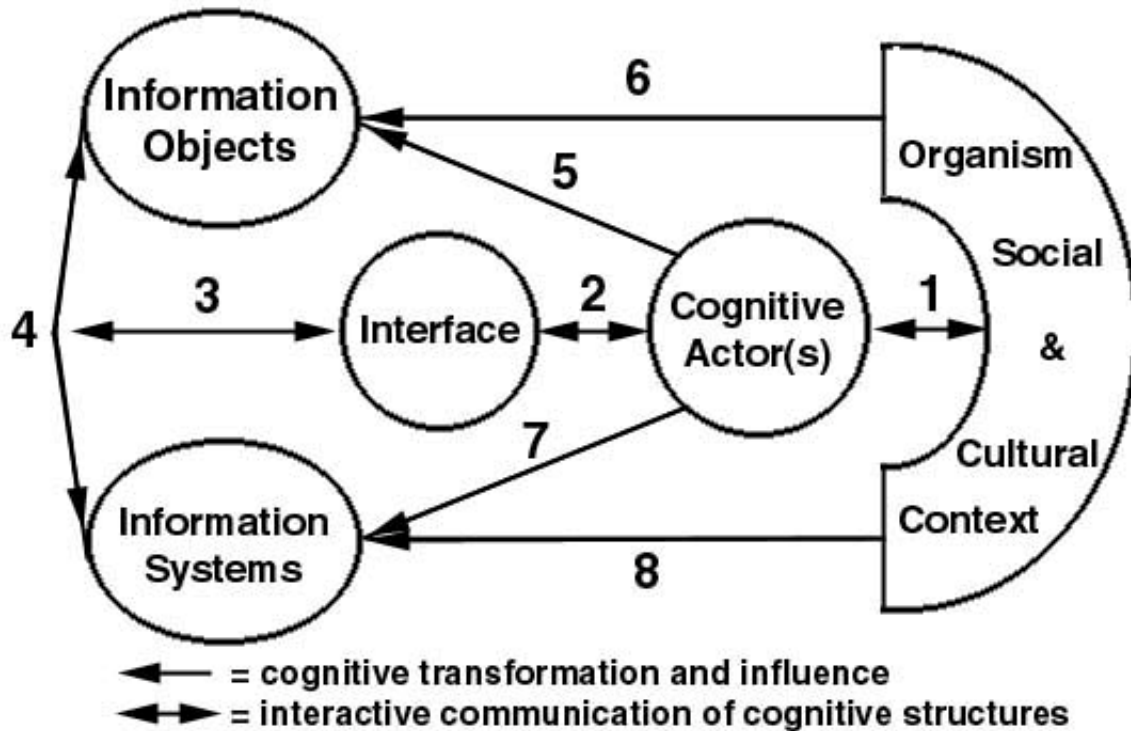
⁶. Belkin, N. J., “The cognitive viewpoint in information science”, *Journal of Information Science: Principles and Practice*, 16(1), 11-15. 1990

⁷. Ingwersen, P. “Information Retrieval Interaction”, London: Taylor Graham, 1992. X, 246 p.

⁸. Ingwersen, P. “Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory”, *Journal of Documentation*, 52(1), 3-50. 1996.

⁹. Ingwersen, P. “Cognitive information retrieval”, In: M. Williams(Ed.), *Annual Review of Information Science and Technology*, 34, 3-51. 2001.

¹⁰. 引自 Ingwersen, P. “Integrative Framework for Information Seeking and Interactive Information Retrieval”：「The framework reflects the understanding that IS&R is a process of cognition for the information-seeking actor(s) or team in context. Algorithmic and IIR, as well as information seeking (IS), involve cognitive and emotional representations are seen as manifestations of human cognition, reflection, emotion, or ideas forming part of IS&R components and kinds of interaction in context」, In K. E. Fisher, S. Erdelez & L. McKechnie (Eds.), *Theories of information behavior (The Theories 35, pp215-220)*. Medford, N. J.: ASIST



從這個架構圖可知，組織、社會和文化背景，會影響到資訊搜尋及檢索行為人的認知，行為人的認知又會影響到所有的資訊行為過程。余光中的〈隔水書〉詩，正可以用來說明形符與意符的差別，以及對行為人（閱聽者）所產生的認知落差。說明如下：

以文學作品為例，余光中的〈隔水書〉詩：「長風遠雲，七百里一望是水程，隔水宛轉望一個仙人。島是蓬萊，鷗非青鳥，空郵箋傳幾日的苦眺。委委曲曲一句話，紙摺幾摺就話摺幾摺就心折幾折，要拆要輕輕的拆。那仙人歎口氣，也罷。你心焦水西湄，我心焦東浦，念來逼人似烽火，隔水傳郵隔一宿，看你煉晚霞對我燒日出。」這一首表達愛慕之情的抒情詩，詩中描寫有一個人寫一封信向心中的仙人宛轉表達愛慕之意，將信託給海鷗傳送，幾日的苦眺，盼回的卻是一紙空郵箋。寫信人或因含蓄，或因唯恐唐突佳人，不敢將情思直言明說。當對方回以空郵箋，寫信人擔心對方無法理解自己信中曲折委婉之言，於是再寫一封信告訴對方「紙摺幾摺就話摺幾摺就心折幾折，要拆要輕輕的拆。」

在這一段文字裡，寫信的人將「紙」、「話」、「心」三種由具體到抽象的詞彙排比呈現，試圖引導讀信人從「紙摺幾摺」去推測「話摺幾摺」，再進一步理解「心折幾折」。「紙」是具體客觀的物，相當於形符，不同人之間的認知不會有落差；「心」是主觀抽象的感受，相當於意符，不同人之間容易產生感受及理解上的差異；「話」介於「紙」與「心」之間，一段話或一段文字，說者與聽者之間，雖有語言或文字作為具體客觀的表述媒介，但是說者的原意與聽者的感受難免有客觀上的認知與主觀上感知的落差。寫信的人以這種方式排比呈現，是要告訴仙人，我的「心」折幾折，雖然你看不到，你可

以聽（看）「話」摺幾摺，如果還是不夠明確，再看「紙」摺幾摺，便可以循著原路線逐一回溯我的「話」摺幾摺，我的「心」折幾折。「紙」是具體客觀的，你看我看他看都不會有誤差，寫信的人以「紙摺幾摺就話摺幾摺就心折幾折」引導讀信的仙人，循此路線追溯寫信人千迴百轉的曲折心事。

文學表達的情感通常是主觀而抽象的，即便是藉助客觀的物象作為作者與讀者之間溝通的媒介，仍然存在著感知層次的不同，因此，引用資訊技術應用在文學上，其困難度比自然科學領域高得多，應用在特殊語言形式的詩詞，困難度又高於小說、戲劇、散文等文體。

語意的標記與語意概念分類，可以解決電腦只能分辨字形，無法理解字詞義的問題，但仍然與人的判斷解讀存在著相當大的差距。將文本經過詞彙切分及語意標記，使電腦能理解詞彙結構及語義，建立語意概念的分類及上下位詞層級的結構，使之更接近人的知識體系，才能更接近使用的需求，提升分類的準確性與完整性。

三、詩詞文獻的分類

在還沒有資訊工具之前，書籍作者或編纂者因應不同的使用需求，將書籍的內容做不同的分類，來幫助使用者快速方便查檢所需要的資料。孔子刪詩書，訂禮樂，將《詩經》按照風、雅、頌分類，可說是文本篇什分類的起源，《佩文韻府》以韻分類，《唐詩別裁》、《清詩別裁》以詩體分類，分類的用意不外乎為了方便查檢瀏覽。

迄今為止，詩詞文獻的分類側重在整首詩的主題分類，以唐詩為例，宋代趙孟堅曾為唐詩做過分類，可惜「佚闕不完，世無刊本」¹¹，明代張之象「因復有此作（《唐詩類苑二百卷》）」¹²，1992年四川辭書出版社以明萬曆刻本的《唐詩類苑二百卷》為底本，「編纂時……以原書為基礎，充分考慮原來的版本、體例；詩文內容有疑誤處參校《全唐詩》等書和作家專集；部類上進行必要的分合、增減、調整，部類名稱既尊重傳統，又兼顧現代，重新編制索引。……得四十部，一千零八十三類」¹³，名之為《唐詩分類大辭典》，收錄唐詩10,197首。¹⁴分類以題材為標準¹⁵，其大類類目為：「天文部、四時節氣部、地理部、山部、河川部、京都部、州郡部、邊塞部、帝王部、皇室部、職官部、治政部、禮部、樂部、文藝部、軍事部、人事部、儒家

¹¹ . 參見《武英殿本四庫全書總目提要》，集部總集類存目二，192卷，第5冊，第156頁上，《唐詩類苑二百卷》條。

¹² . 同註11。

¹³ . 參見《唐詩分類大辭典》，馬東田主編，四川辭書出版社，1992年成都出版社出版，第6頁，前言。

¹⁴ . 清代彭定求奉敕編纂的《全唐詩》加上陳尚君輯校的《全唐詩補編》，總計收有2,821位詩人，51,170首詩，《唐詩分類大辭典》收錄分類的詩作只佔全唐詩的百分之二十。

¹⁵ . 參見《唐詩分類大辭典》，馬東田主編，四川辭書出版社，1992年成都出版社出版，第9頁，凡例五「分類以題材為標準，不按詩的體裁排列。」

部、釋佛部、道家道教部、居處部、寺觀部、祠廟部、產業部、器用部、服飾部、飲食部、玉帛部、技藝部、方術部、花部、草部、果部、木部、鳥部、獸部、鱗介部、虫豸部、祥異部、雜部」。¹⁶

根據《唐詩分類大辭典》凡例所言「分類以題材為標準」¹⁷，然而檢視全書，有幾個問題值得商榷，舉其要者說明如下：

(一) 分類與詩作內容不合：

依凡例所言「分類以題材為標準」，檢視全書，絕大多數都以詩題之字面含義分類，不是以詩作內容之主題分類，使得有些分類與詩作內容的主題未必相關，產生分類類別與主題不合的現象，例如將杜牧〈金谷園〉：「繁華事散逐香塵，流水無情草自春。日暮東風怨啼鳥，落花猶似墜樓人。」歸在地理部的「園」類，與詩的主題不合。金谷園是石崇所建，這首詩寫的是綠珠為石崇墜樓而死的殉情故事，不是金谷園的園景，歸類於此，並不適當。又如〈關山月〉是漢樂府〈橫吹曲〉曲調名，多用來寫邊塞士兵因久戍不歸，思念家人，和家人互傷離別之情的作品，《唐詩分類大辭典》所收十一首〈關山月〉，其中儲光羲〈關山月〉：「一雁過連營，繁霜覆古城。胡笳在何處，半夜起邊聲。」的「繁霜」，可解為月光或霜雪的霜，如解為月光，這十一首詩都寫到月亮，在戰地的征人雖因見到月亮而想念家鄉，但主要的重點是描寫邊塞征人思念家人之苦，月亮只是引發思鄉情懷的觸媒劑，依該書分類凡例，這些詩應該歸入「人事部」。事實上《全唐詩》共收有二十一首〈關山月〉，只有皎然及霍總二人所寫的〈關山月〉與邊塞征戰無關，皎然的詩寫月亮¹⁸，霍總的詩寫男女之情¹⁹。

(二) 標舉中類的體例不一

有些大類之下的中類，有普通名詞、專有名詞及其他詞類，含多種詞類的類別，有些只列出普通名詞，有些增列專有名詞，體例不統一，例如「河川部」的詩，有些詩題上只泛寫普通名詞，如劉長卿〈湖上遇鄭田〉、李德裕〈潭上喜見新月〉，有些以河川的專有名詞當作詩題，如宋之問〈洞庭湖〉、杜甫〈登白馬潭〉，「洞庭湖」與「白馬潭」都是專名詞，所以「河川部」的中類，列有專有名詞「淮水、濟水、涇水、洛水……」及普通名詞「湖、池、津、潭、湍、灘……」等，「山部」亦如是。「居處部」的詩，與「山部」、「河川部」一樣，有些詩題直接以專有名詞為題，如李白〈長信宮〉、崔顥〈黃鶴樓〉等，有些只泛寫亭臺樓閣，如劉長卿〈同諸公登樓〉、邱為〈題農父廬舍〉，

¹⁶ . 參見《唐詩分類大辭典》，馬東田主編，四川辭書出版社，1992年成都出版社出版。

¹⁷ . 同註15。

¹⁸ . 皎然〈關山月〉：「家家望秋月，不及秋山望。山中萬境長寂寥，夜夜孤明我山上。海人皆言生海東，山人自謂出山中。憂虞歡樂皆占月，月本無心同不同。自從有月山不改，古人望盡今人在。不知萬世今夜時，孤月將□誰更待。」

¹⁹ . 霍總〈關山月〉：「珠璣翡翠床，白皙侍中郎。五日來花下，雙童問道傍。到門車馬狹，連夜管弦長。每笑東家子，窺他宋玉牆。」

「居處部」的中類卻只列出「宮、殿、闕、樓、臺、榭……」等普通名詞，未列出「長信宮」、「黃鶴樓」等專有名詞，「寺觀部」情況也相同，可見體例不統一。

(三) 專名與通稱、俗名混合

除了大類之下的中類列舉體例不一，在同一中類中，也有將專有名詞與通俗名稱混合的情況。該書的分類，既不是將所有詩題裡出現的專有名詞一一列舉，又未曾將通俗專名做詞彙控制，如「草部」列有「苔」及「垣衣」、「瓦松」，全唐詩詩題有「苔」字者共十七首，寫「蘚」者二首，「苔、蘚」並列者一首，寫「垣衣」者一首，寫「瓦松」者一首。事實上「垣衣」是苔的一種，《廣群芳譜》之〈苔譜〉：「苔，一名圓蘚，……苔在牆曰垣衣，在屋曰瓦松」。²⁰但中類中列了「苔」、「瓦松」及「垣衣」，未列「蘚」。在「木部」列有「情深樹」，李白有〈題情深樹寄象公〉詩：「腸斷枝上猿，淚添山下樽。白雲見我去，亦為我飛翻。」李白詩題裡的「情深樹」只是李白看到的某一棵樹，李白隨意名之曰「情深樹」，在全唐詩之詩題及詩句裡，也只有李白這首詩這麼寫，當作一個中類，顯然不合適。又如「木部」列有「端正樹」，全唐詩收有溫庭筠〈題端正樹〉、貞元文士〈題端正樹〉二首詩，貞元文士詩之詩序曰：「《酉陽雜俎》云：長安西端正樹，去馬嵬一舍之程，迺德宗皇帝幸奉天，睹其蔽芾，錫以美名。有文士經過，題詩逆旅，不顯姓名。」可見「端正樹」是唐德宗錫以美名的樹，不是該樹種的學名，與本中類其他以「柏、杉、檜、楠……」等樹種的學名為類別的體例不同。此項分類，另列有「雜樹」，應將「情深樹」、「端正樹」列在「雜樹」。其他如「鳥部」所列「反舌」即「百舌」鳥，未做詞彙控制而將二者並列為中類，徒增繁瑣與混淆。

從以上所分析的情況看來，《唐詩分類大辭典》所區分的大類或中類，其分類的方法及訂定詞目的方式，與語意的概念及使用者的認知不盡相同，也不符合學術分類的規則與慣例，使用者無論從「客觀知識性取向」或「主觀情意性取向」判斷，《唐詩分類大辭典》的分類，都不盡符合語意概念及使用者認知，再加上分類的結果與凡例所說的「分類以題材為標準」不盡相合，使用者的認知與該書分類有落差。如能以詩作之內容作語意標記與分類，作為判斷詩作主題的依據，其準確度將會比《唐詩分類大辭典》高。

本研究將透過對唐宋詩詞作內容的語意標記及分類，嘗試從不同角度建立詩詞語意概念分類，以與歷代類書及《唐詩分類大辭典》的分類相比較，並期待在未來的研究裡，能進一步從語意標記及概念分類、詞彙共現三個角度分析、彙整、歸納唐宋詩詞主題，建立唐宋詩詞主題分類，以滿足使用需求。

四、 中文語意概念分類研究回顧

²⁰ . 參見清汪灝等撰，《佩文齋索引本廣群芳譜》第九冊卉譜：苔譜，新文豐出版社，1980年出版，第5054-5055頁。

中文詞語的語意分類研究成果，紙本文獻方面以梅家駒主編，上海辭書出版社出版的《同義詞詞林》最具代表性。本書第一版發表於1983年，第二版發表於1996年，收錄詞彙54,235筆，將詞義分為12大類，95中類，1,428小類，3,925詞群。每一次類所包含的詞彙以及該詞彙的同近義詞依序列出，成為研究詞彙語意學及語意概念分類的重要工具書。電子本文獻方面有董振東先生於1988年創建的知網(HowNet Knowledge Database)，2003年中央研究院資訊所詞庫小組與董振東先生展開建構繁體字知網的合作計畫，建立「概念網」，收錄55,912筆，分析歧義後包含概念60,482筆。²¹，以上語料都是以現代漢語為範圍。以文學文本建立語意網，目前有中央研究院藉由「數位典藏國家型計畫」透過其中的「語言座標——參考資源建置與服務」²²與「溫柔在頌——唐詩三百首」網站²³，參考「中央研究院中英雙語知識本體詞網」(Sinica BOW)²⁴和SUMO (Suggested Upper Merged Ontology，建議上層共用知識本體)²⁵資料所合作建置的「唐詩三百首知識本體」網站²⁶以及「蘇軾詩知識本體」網站²⁷。這二個網站僅先將文本中的動物、植物、人造物分類。筆者於2002年7月所主持的「以XML可延伸式標注語言建立文章標記(Content Markup)系統研究——以蘇軾詩詞為範圍」計畫網站²⁸，將蘇軾詩全面標記與分類；爾後再以蘇軾詩所建置的詞彙語意標記資料庫為基礎，於2005年3月以所主持的「淺斟低唱——宋詞古唱虛擬實境教學環境建置計畫」²⁹選出柳永(218闕詞)、蘇軾(379)、秦觀(148)、周邦彥(187)、李清照(53)、姜夔(87)、辛棄疾(627)等七家詞為範圍做語意標記與分類；2006年1月所主持的「智慧型全台詩知識庫計畫第

- ²¹ . 參見中央研究院資訊科學研究所陳克主持，「概念網網站」(<http://ckip.iis.sinica.edu.tw/CKIP/conceptnet.htm>)，以及陳克健、黃淑齡、施悅音、陳怡君，2005，多層次概念定義與複雜關係表達——繁體字知網的新增架構，漢語詞彙語意研究的現狀與發展趨勢國際學術研討會，北京大學。
- ²² . 中央研究院語言研究所所執行的「數位典藏國家型科技計畫」語言座標計畫，網址：<http://linganchor.sinica.edu.tw/>
- ²³ . 元智大學羅鳳珠、胡順萍、歐慶亨主持，【數位博物館計畫——中小學語言文學知識網(搜文解字二)——生活語文與唐宋流行歌——唐宋流行歌——溫柔在頌——唐詩三百首】網站，網址：<http://cls.hs.yzu.edu.tw/300/HOME.HTM>
- ²⁴ . 中央研究院語言研究所與遠見科技股份有限公司共同開發，「中央研究院中英雙語知識本體詞網」(The Academia Sinica Bilingual Ontological Wordnet)，網址：<http://bow.sinica.edu.tw/>
- ²⁵ . 參見中央研究院中英雙語知識本體詞網，網址：http://bow.sinica.edu.tw/ont/bow_sumo_cont.html；SUMO 簡介：SUMO (Suggested Upper Merged Ontology，建議上層共用知識本體)，網址：<http://www.ontologyportal.org/>，由IEEE標準上層知識本體工作小組所建置。這工作小組的目的是發展標準的上層知識本體，這將促進資料互通性、資訊搜尋和檢索、自動推理和自然語言處理。
- ²⁶ . 「唐詩三百首知識本體」網站，網址：http://bow.sinica.edu.tw/ont/ts300_ont.html
- ²⁷ . 「蘇軾詩知識本體」網站，網址：http://bow.sinica.edu.tw/ont/ShuShi_ont.html
- ²⁸ . 羅鳳珠、李錫捷，【以XML(eXtensible Markup Language)可延伸式標注語言建立文章標誌(Content Markup)系統研究——以蘇軾詩詞為範圍】，國科會，2002年7月1日至2003年6月30日(計畫編號：91-2422-H-155-3401-)，網址：<http://cls.hs.yzu.edu.tw/CM/login.htm>
- ²⁹ . 羅鳳珠、張智星、許介彥，【淺斟低唱——宋詞古唱虛擬實境教學環境建置計畫】，國科會數位典藏國家型科技計畫：94年度數位典藏創意學習計畫，2005年3月1日至2006年2月28日。計畫編號：NSC 94-2422-H-155-001，網址：<http://cls.hs.yzu.edu.tw/CSP/index.html>。

二階段建置計畫」³⁰，以國家台灣文學館委託施懿琳教授編纂的《全台詩》前三年成果為範圍，建置全台詩語意標記與分類；2006年3月繼續擴充，於「淺斟低唱—宋詞古唱虛擬實境教學環境建置計畫」³¹，以唐宋詞為範圍；於「新詩改罷自長吟—唐詩吟唱虛擬實境教學環境建置計畫」³²，以全唐詩為範圍，建置唐宋詩詞詞彙語意標記與分類，並增加詩的主題分類。

由前述可知，目前有關於語言之自動化處理技術及語意概念、語意標記等相關研究，現代漢語的研究比較多，語料的來源雖也包含了文學典籍的語料，但是以語體文為主，並未包含韻文。由於韻文的語言表達及構詞方式與語體文不同，目前所發展的語言處理工具及語意標記及分類項目，勉強套用於韻文，除了扞格不入之外，也抹煞了韻文的語言特徵。好的語言處理工具，需要兼顧其使用目的，用之於圖書資訊檢索，用之於語言學研究、用之於文學語言研究，用之於文學研究，都會因為需求不同而有不同的分類方法。

五、 詩詞語言的語意分類的困難

中國的韻文發展到唐宋代，可說是達到了顛峰，成為中國文學史上燦爛的一顆明珠。清代彭定求奉敕編纂的《全唐詩》以明代胡震亨《唐音統籤》及季振宜《唐詩》為底本編纂而成，全書分為九百卷，將帝王后妃的作品列於首，其次是樂章、樂府，接著依作者時代先後編排，附以作者小傳，最後是聯句、逸句……等等。作者之後繫上詩題，使用者只能依照作者或詩題逐一查檢資料，非常不便。《唐詩分類大辭典》雖然存在著上述問題，但是初步達到了提供使用者從內容主題尋找資料的功能，比起只從作者、詩題尋找資料的編輯方式方便許多。但是儘管《唐詩分類大辭典》聲稱是以主題分類，事實上有很多詩作除了文字表面的含義，還含有「雙關義」、「情韻義」、「象徵義」、「深層義」、「言外義」³³，不是單一主題，使得主題分類的困難度非常高，《唐詩分類大辭典》無法兼顧詞彙本義之外的含義，是分類的準確性及完整性相對不足的原因，加上《唐詩分類大辭典》處理的內容只佔全唐詩的百分至二十，涵蓋面不足。

前已述及，詩是特殊的語言，特別講究聲律及形式之美，受到字數及平仄押韻等格律的限制，詩的用語與語體文不同。因字數的限制，一首詩短則20字（五言絕句），多

³⁰ . 羅鳳珠，【智慧型全台詩知識庫計畫第二階段建置計畫】，國家台灣文學館，2006年1月1日至2006年12月31日。網址：<http://cls.hs.yzu.edu.tw/TWP/index.asp>

³¹ . 羅鳳珠、張智星、許介彥，【淺斟低唱—宋詞古唱虛擬實境教學環境建置計畫】，國科會數位典藏國家型科技計畫：94年度數位典藏創意學習計畫，2005年3月1日至2006年2月28日。計畫編號：NSC94-2422-H-155-001。

³² . 羅鳳珠、張智星、許介彥，【新詩改罷自長吟—唐詩吟唱虛擬實境教學環境建置計畫】，國科會數位典藏國家型科技計畫：95年度數位典藏創意學習計畫，2006年3月1日至2007年2月28日。計畫編號：NSC95-2422-H-155-001。網址：<http://cls.hs.yzu.edu.tw/tang/index.html>

³³ . 參見袁行霈著，〈中國古典詩歌的多義性〉，收於《中國詩歌藝術研究》，五南圖書出版公司印行，1989年5月台灣出版，第6-23頁。

則不過數百字（排律或古體詩），短短篇幅要包含豐富的內涵，修辭的技巧自然比語體文更為講究，因此也存在著更多一詞多義，一義多詞的現象。除了一般用語之外，詩詞使用了更多典故等特殊用語，使得詞彙的含義比語體文更豐富，更曲折多變，甚至於脫離了原有的語意而由詩人創造新的語意。

本研究嘗試把詩的語言，區分為一般用語、特殊用語、情境用語三種。³⁴一般用語指的是詞彙的語意與其他文體類似，多數是客觀知識性的描述用語，講究語義精確，一次只使用一種語義；特殊用語指的是含有典故、象徵、比喻、意象等等修辭技巧含義的詞彙，多數具有主觀情意性的引導與感發功能，詞彙的含義不只是字面的含義，可能同時含有多種語義，詩人刻意引發讀者想像空間，達到「言有盡而意無窮」的境界；情境用語指的是不同詞彙共現所烘托出來的情境，這些詞彙的本身，也許是一般用語，也許是特殊用語，但是因為在同一作品共現而產生的情境，往往就是詩的主題，這些詞彙的語義必須與詩中其他詞彙共現才能顯示。由此可知，植基於語意學的詞彙語意標記，可以使電腦對文本內容含義的理解與判斷更接近人的認知觀點。如前所述，詩詞語言是特殊的語言形式，比起其他文體，存在著更多一詞多義，一義多詞的現象，加上詩句語意或詩作主題不一定在單一詞彙顯示，而是以共現的詞彙整體烘托出來，增加語意標記的困難，但相對而言，也可以從中歸納、推斷詩的主題。

六、 概念體系與語言知識本體

George Lakoff在” *Women, Fire, and Dangerous Things: What Categories Reveal about The Mind*” 一書的〈緒言〉提出：「甚麼是概念體系（Conceptual system）以及它是如何組織的？所有的人是否都用同樣的概念系統？」³⁵又說：「人的肉體經驗和運用想像機制的方式，對於人如何建構範疇和理解經驗是至關重要的。」³⁶

本文以詩詞為例，說明人類是否具有相同或相類似的概念系統？是否具有類似的情境認知、感知與感發能力？詩詞的內容以人為軸心，描寫人主觀內在的心靈活動、人與外物的互動及對客觀外在景物的敘述。內在的心靈活動是主觀的情志，表現在言志與言情的作品；客觀的敘述是對客觀外物的描寫，表現在敘事、詠史、詠物、寫景……等作品；物我互動的作品內容，是作者內在心靈的活動隨著外物的改變而產生的情意波動，即劉勰《文心雕龍·物色第四十六》所說的「春秋代序，陰陽慘舒，外物之動，心亦搖

³⁴ . 參見拙作〈植基於語意學及使用者認知觀點的詩詞資訊檢索系統設計：以全唐詩網站為例〉，第三屆文學與資訊科技國際研討會，日本學藝大學主辦，2007年3月19-20日。

³⁵ . Lakoff, G. (1987)著，梁玉玲等譯，《女人、火與危險事物：範疇所揭示之心智的奧秘》（*Women, Fire and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press. ），桂冠圖書公司，1994年出版，〈緒論〉，xvii。

³⁶ . 同註35，xviii。

焉。」³⁷以及「情以物遷，辭以情發」³⁸。與外物互動的描寫，往往是作者透過感官感受以引發讀者相同的感官經驗而產生認知，引發感知，進而體會作品，與作者共鳴。共鳴的基礎來自於人以感官系統（耳鼻口眼皮膚心智）感知客觀外物，所產生的概念體系，具有共通性（多數人相同）。人的「客觀認知」加上「主觀感知」、「思維體系」構成「概念系統」，當作者與讀者之間具有共同的知識基礎、情感經驗及概念體系，便能夠產生感同身受的共鳴。

以詩詞裡寫及苔蘚的作品為例，苔蘚植物是由水生生活趨向陸生生活的植物族群，生長於極端潮濕的向陽地區或是陰暗的環境中，以配合其親水性結構，不然便很快枯死。這些地方往往也是人跡罕至之處，這是我們對於「苔蘚」的認知基礎。³⁹由於苔蘚性喜陰暗潮濕的特性，潮濕陰暗在感官上給人清冷幽深，人跡罕至的感覺，因此詩詞中寫到苔蘚時，往往是「江邊道路多苔蘚，塵土無由得上衣。」（姚合，〈送崔約下第歸揚州〉）的失意文人，或是「玉階陰陰苔蘚色，君王履綦難再得。」（沈佺期，〈古歌〉）的失寵佳人，或是「天花寂寂香深殿，苔蘚蒼蒼闕虛院。」（劉長卿，〈齊一和尚影堂〉）的塵外僧人，或是「肌膚銷盡雪霜色，羅綺點成苔蘚斑。」（嚴郾，〈望夫石〉）的空閨女子，或是「花木閒門苔蘚生，滄川特去得吟情。」（張喬，〈城東寓居寄知己〉）的山林雅士。「苔蘚」若是長於門前或台階，詩人將之與「門」、「階」並列呈現時，更核心的集中在描寫因失寵而幽居空閨女子的淒清冷落。鄭谷〈長門怨，二首之二〉：「流水君恩共不回，杏花爭忍掃成堆。殘春未必多煙雨，淚滴閒階長綠苔。」以不再使用而長滿綠苔的閒階，婉轉呈現幽居於此的空閨女子，因君恩不回，沒有人前來，階閒無用而長滿綠苔的蒼涼景象。⁴⁰從這個例子可知，多數人會因為具有共同的知識基礎，而有共同的認知，進而引發感同身受的感知。

但是，感知也會因為個人的特殊經驗、遭遇、情性而有差異。從李白〈月下獨酌〉：「花間一壺酒，獨酌無相親。舉杯邀明月，對影成三人。」李商隱〈春日寄懷〉：「縱使有花兼有月，可堪無酒又無人。」等作品，可知「花、月、酒、人」被多數詩人認為是自然界與人世間美好事物，但是在亡國之君李煜筆下的〈虞美人〉：「春花秋月何時了」，春花秋月成為不堪忍受，不忍面對的事物，這是因境遇不同而產生的個別差異。

從「苔蘚」及「花、月、酒、人」的例子，正可以印證 Lakoff 所提出的說法，不同的文人以其所對苔蘚的客觀認知：「苔蘚生長在潮濕而人跡罕至之處」以及肉體經驗：「看到長滿苔蘚的道上、院中、門前、階下」，運用想像，想像這裡住著失意文人、失寵佳人、塵外僧人、空閨女子、山林雅士。

³⁷ . 劉勰著，范文瀾注，《文心雕龍注·物色第四十六》，香港商務印書館，1960年出版，第693頁。

³⁸ . 同註37。

³⁹ . 引自延陵科學綜合室建置的「延陵植物標本閣」網站：<http://www.ngensis.com/flora/fn01.htm>，苔蘚類單元：<http://ngensis.com/flora-b/L&M.htm>

⁴⁰ . 同註34。

如果人的概念系統是有規則的，如果所有的人都有同樣的概念系統，如果多數人的多數概念是相同的，那麼，以本體知識解構文學作品，再重新建構成為電腦可以讀懂的知識體系，便可以使電腦具備接近人的知識體系。

詩詞的語意概念體系建立在詩詞語言知識本體上。本體論(ontology)在十七世紀以後成為哲學論述的課題，資訊科學引用本體論，起始於人工智慧領域的研究，引用作為組織、分類、管理及描述知識的基礎，並且成為語意網的基本理論，本體論也開始被應用在不同的學門，本研究不僅是建立詩詞語意網，還進一步將語意概念予以分類，建立語意概念分類知識體系。

七、 詩詞語言的語意概念分類方法及分類體系

使用者的認知行為，除了前述所論及對詞彙的知識認知、情意感知、情境感知之外，使用者還具備由各種單一詞彙的知識所建構的知識體系以及理解感受詩詞作者在詩詞作品內容所創造的豐富語意、語境與情境，這是只能分辨字形的電腦所無法具備的。詞彙經過語意標記，雖然已經能夠使電腦初步具有分辨語意的能力，但也只是零星片段沒有經過組織的語意，不符合人的知識體系，難以觸引對詩詞的感發能力。

本研究以詞彙為單位，先建立詞彙語意標記，再依語意類別建構語意概念，依語意概念分類。電腦除了無法分辨字詞義，也無法判斷詞組的構詞單位，因此，正確的詞彙切分是語意標記重要的關鍵。

詩詞是特殊的語言形式，其特殊性包含語意的特殊、構詞的特殊、篇章結構的特殊。除了使用專有名詞或典故之外，詩詞文體以雙音節及單音節詞為主，尤以雙音節詞居多。因遷就雙音節及平仄格律，使得詩詞語言的構詞方式與其他文體不同，有些詞彙只用於詩詞，不用於其他文體。或該詞彙在詩詞裡是通順而有意義的詞彙，但是當抽離該句詩詞時，這個詞彙往往不通順或無法顯示完整語義。舉例說明如下：

- (一) 構詞特殊，其他文體罕用詞彙：白居易〈賦得古原草送別〉：「離離原上草，一歲一枯榮」的「離離」；蘇武〈結髮為夫妻〉：「參辰皆已沒，去去從此辭」、柳永〈雨霖鈴〉：「念去去，千里煙波，暮靄沉沉楚天闊。」的「去去」，在其他文體很少使用
- (二) 詞彙分散在上下句：蘇軾〈張子野年八十五尚聞買妾述古令作詩〉：「詩人老去鶯鶯在，公子歸來燕燕忙」，「鶯鶯燕燕」詞分散在上下句。
- (三) 使用縮略倒裝語：蘇軾〈和子由柳湖久涸忽有水開元寺山茶舊無花今歲盛開，二首之一〉：「如今勝事無人共，花下壺盧鳥勸提」，下句完整的詞句是：「花下提壺盧鳥勸人提起葫蘆喝酒」，全句使用了倒裝句、雙關語、縮略語。

(四) 名詞雙舉並列，以豐富詞義：孟浩然〈過故人莊〉：「把酒話桑麻」、施肩吾〈山中喜靜和子見訪〉：「每逢猿鶴即相親」。這二句的句法分別是「把酒／話／桑麻」、「每逢／猿鶴／即／相親」，「桑麻」與「猿鶴」都是名詞，理應切分為單字詞，但是在詩詞裡常有將不同名詞並列的寫作方式，作者將某些名詞並列，除了詞彙本身的語意之外，往往以並列方式彰顯其言外之意，如果予以切分為單字詞，不但因並列所烘託的語意將隨之消失，也不符合詩詞雙音節詞的構詞特性。

詩詞的句法有固定的規則，五言詩之前二字，七言詩之前四字，除非使用專有名詞或典故，以雙音節詞為主。唐宋詞除了使用單字領字的句子第一個詞彙是單字詞，其他句子及扣除單字領字之後的句子，也以雙音節詞為主。五言詩的句法多數是上二下三（夜來／風雨聲，花落／知多少），下三部分作上二下一（風雨／聲），或作上一下二（知／多少），而成為二二一（夜來／風雨／聲）或二一二（花落／知／多少）的句法；七言詩前四句是二組雙音節詞，多數是二二一二或二二二一句法，五、七言詩之末三字，是上二下一或上一下二沒有規則。五言詩的前二個字及七言詩的前四個字，除非是專有名詞或典故，鮮少使用單字或三字詞，因此依據詩的句法做適切的斷詞，切分的優先順位是：一、專有名詞；二、詞的領字；三、典故。⁴¹末三字在名詞及典故優先的前提下，比對已切分出來的詞彙，相符合者，則先認定為詞彙，最後再以人力校正。

詩詞內容所描寫的元素包括「人、事、時、地、物」五種，寫哪些人在哪些時間，哪些地方做過、發生過哪些事，牽涉哪些物。小說、戲曲等多數文學作品的內容都包含這五項元素，且都是直接描寫。散文、詩詞因篇幅較短，其描寫方式，或直接表述，或隱藏在字裡行間間接表述。直接描寫者，如柳宗元〈江雪〉：「千山鳥飛絕，萬徑人蹤滅。孤舟蓑笠翁，獨釣寒江雪。」詩句裡直接書寫：人（蓑笠翁）、事（獨釣）、時（寒冬）、地（江邊）、物（孤舟）。間接表述者，如杜甫〈八陣圖〉：「功蓋三分國，名成八陣圖。江流石不轉，遺恨失吞吳。」詩句裡的文字，雖不是直接寫出全部的人、事、時、地、物，但仍隱藏在字裡行間，包含：人（諸葛亮、劉備、孫權、周瑜、曹操）、事（吳、蜀、魏三分天下之爭戰，劉備遺恨失吞吳之史事）、時（三國時代）、地（魚復（腹），今四川奉節縣）、物（八陣圖）。

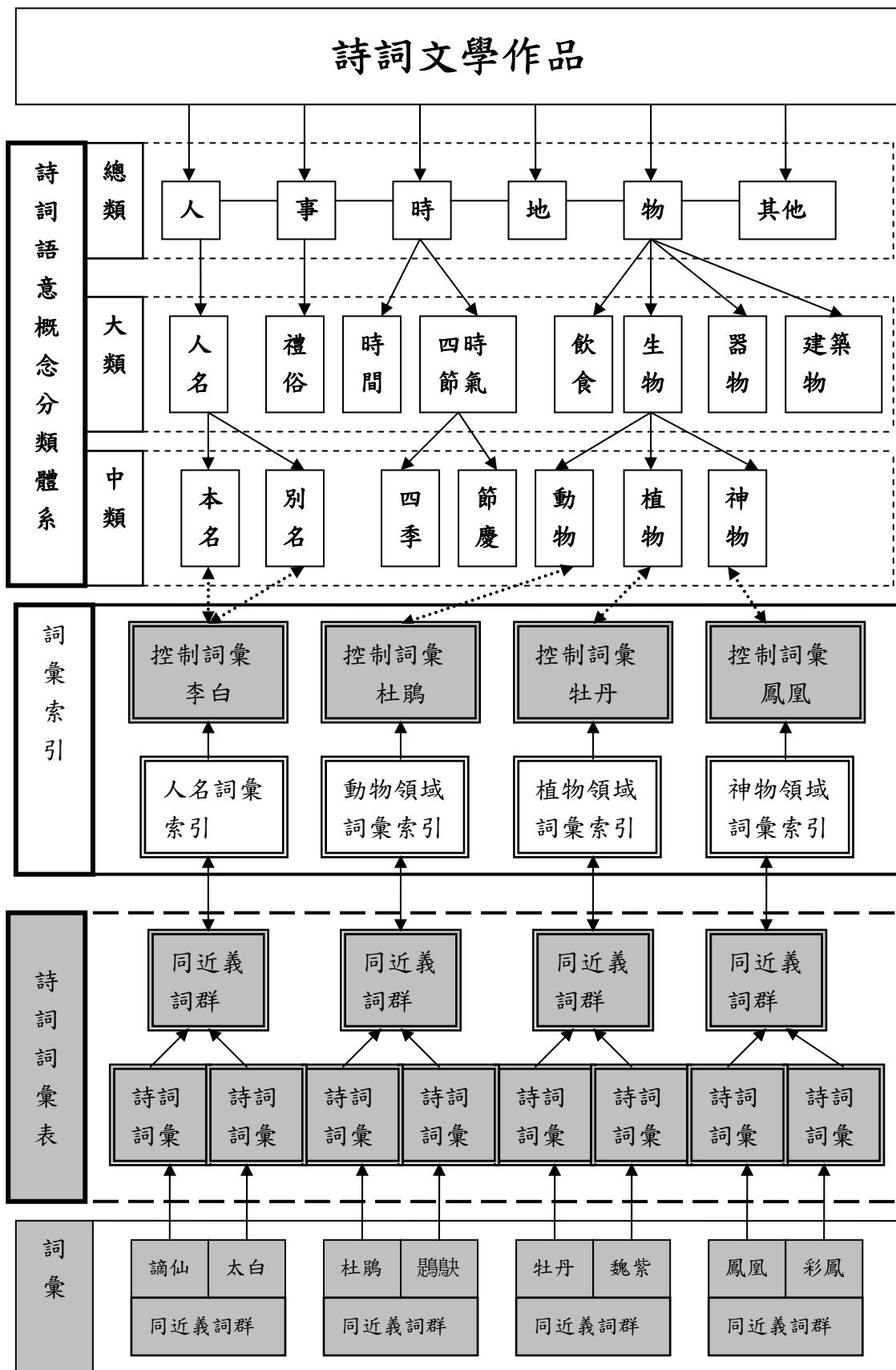
本研究將詩詞作品之詞彙以前述方法切分之後，以詞彙為單位進行語意標記，並將同近義詞詞彙建立「同近義詞詞群」，再針對每一組「同近義詞詞群」建立「控制詞彙」（同近義詞詞彙控制），藉以統領相同語意概念的詞彙。詞彙語意標記所建立的語意概念予以分類，建立分類體系架構。分類的考慮以符合詩詞語言特性、保留詩詞語意內涵、滿足文學研究需要為主，其次考慮分類學的意義。包含詩詞全部內容的「詩詞詞彙表」經過語意概念分類之後，便可以成為詩詞詞彙的各種「領域詞彙索引」，例如「人名詞彙索引」、「動物領域詞彙索引」、「植物領域詞彙索引」等等。

⁴¹ . 參見拙作，〈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，中央研究院主辦，2005年9月1-2日。（收於網路展書讀網站：<http://cls.hs.yzu.edu.tw>）

詩詞詞彙表每一組同近義詞詞群以一個詞彙作詞彙控制，舉例如下：

控制詞彙	同近義詞詞群（詩詞詞彙）
李白	李白、太白、李太白、李謫仙、謫仙李、謫仙、騎鯨公子、謫仙人、李青蓮
一年	一載、一歲、一稔、一春、一年、一秋
浣花溪	花溪、浣花溪、浣沙、浣紗、浣沙溪
崑崙山	崑崙、昆侖、昆崙、昆岡、崑崙、崑崙山、崑崙山、崑岡、崑山、崑崗
月	月、月華、太陰、玉兔、玉鉤、玉蟾、玉盤、玉輪、玉蟾、生魄、冰壺、冰輪、冰鏡、冰鑑、兔影、兔魄、孤輪、明蟾、金波、金輪、秋蟾、飛鏡、桂宮、桂窟、桂魄、桂蟾、海蟾、涼蟾、圓魄、圓蟾、銀蟾、嬋娟、輪影、濯魄、蟾、蟾光、蟾宮、蟾窟、蟾蜍、蟾影、蟾輝、蟾魄、蟾闕、寶鏡、寶鑑、懸魄、顧兔

本研究因詩詞文體內容所包含的元素，以「人、事、時、地、物」作為最上層的總類，無法包含在這五項元素之外的概念另以「其他」概括之。總類之下再分為「大類」、「中類」。所建立的詩詞語意概念分類體系包含二部分：其一是以詩詞作品詞彙經過詞彙控制所建立的「詩詞詞彙表」，其二是詩詞語意概念分類體系，以「控制詞彙」連結「詩詞詞彙表」與「詩詞語意概念分類體系」二者之間的聯繫關係，關係表如下：



詩詞語意概念分類體系在第一層「人、事、時、地、物、其他」等六大總類之下，總計分為 54 大類，372 中類。予以標記語意及分類的詞彙以實詞為原則，助詞、介詞、連詞、嘆詞等虛詞的語意需附著在實詞上，實詞裡的副詞也須和實詞連結，詞義才能完整顯示，因此這些詞彙暫時不分類。但是虛詞裡的擬聲詞所代表的含義，有些是某些物象的聲音，故予以分類。本研究分類的標的是中國詩詞文獻，所分類別是依據詩詞文獻所切分之詞彙，依據前述之原則予以分類，所得結果是詩詞內容所包含的語意，不是完整的人類知識體系，因此有些古典詩詞缺乏的內容，其知識類別便不會出現在詩詞語意概念分類體系裡。分類表如下：

總類	大類	中類
人	人名	本名、別名字號、人名他稱、法號、人名（合稱）、外國人名、虛構人名、待考人名
人	朝代君王	君王帝號、君王年號、君王陵墓
人	神鬼仙人	神鬼仙人（專稱）、神鬼仙人（泛稱）、菩薩尊號
人	稱謂	人稱代名詞、自稱謙稱、尊稱美稱、貶稱惡稱、親人眷屬稱謂、指代稱謂、一般稱謂、職業身份、職官爵位、皇室稱謂、宗教稱謂
人	身體器官	五官顏面、鬚鬢髮膚、身體四肢、五臟器官、分泌物、身體器官（合稱）
人	感官詞	嗅覺、聽覺、味覺、視覺、觸覺
人	情感詞（綜合情感）	情感（概稱）、正面情感（喜悅欣樂）、正面情感（自我肯定）、正面情感（景仰羨慕）、正面情感（悠然自適）、正面情感（憐惜愛寵）、負面情感（自我否定）、負面情感（惱怒怨妒）、負面情感（悲哀傷痛）、負面情感（感懷惆悵）、負面情感（憂愁掛慮）、負面情感（厭煩憎惡）、負面情感（懊惱憾惜）、負面情感（擔驚受怕）、負面情感（辛酸艱苦）、負面情感（慚愧自責）、負面情感（驕矜自恃）、負面情感（思鄉念舊）、負面情感（尷尬為難）、綜合情感（相反合稱）、綜合情感（相近合稱）
人	情感詞（兩性情感）	傾心愛戀、相思怨懟
人	品德性情	正面品性、負面品性
人	動作（人）	一般行為（人部）、一般行為（刀部）、一般行為（口部）、一般行為（土部）、一般行為（彳部）、一般行為（手部）、一般行為（日部）、一般行為（水部）、一般行為（糸部）、一般行為（衣部）、一般行為（艸部）、一般行為（言部）、一般行為（貝部）、一般行為（走部）、一般行為（足部）、一般行為（馬部）、一般行為（宀部）、一般行為（支部）、一般行為（辵部）、一般行為（其他部）、言語動作、生活活動、官場活動、治國之道、藝文活

		動、娛樂活動、戰爭活動、行蹤消息
人	狀態 (人)	生理狀態、心智狀態、心理能願、生活狀態、心神氣力、人類狀態 (相反合稱)、人類狀態 (相近合稱)、人類狀態 (相異合稱)
人	形容詞彙 (人)	氣質內涵、外表形貌、人生機遇、官場境遇
人	文藝	書籍 (泛稱)、書籍 (專名)、書篇名、文體、引文、樂曲 (泛稱)、樂曲 (專名)、書畫 (泛稱)、書畫 (專名)、書法、書法 (字帖專名)、舞蹈、戲劇
人	專業術語	佛家語、作戰語、法術語、道家語
事	典故	佛學典故、神話典故、一般典故
事	刑責法規	刑責、法規
事	禮俗	祭祀、婚喪喜慶、其他禮俗、節慶
事	災厄	旱災、水災、地震、兵災、其他災害、災厄預兆
事	祥瑞	福泰祥安、祥瑞預兆
事	事件	史事、事件 (概稱)、世局、國家氣運
事	行政體系	部門名稱、朝代名號 (單稱)、朝代名號 (合稱)、民族邦國名 (單稱)、民族邦國名 (合稱)、國名
事	疾病	疾病療法、疾病類別
時	四時節氣	節氣、四季
時	時間	明確時間、範圍時間 (今昔)、範圍時間 (日)、範圍時間 (月)、範圍時間 (世代)、範圍時間 (生)、範圍時間 (年)、範圍時間 (清晨)、範圍時間 (白晝)、範圍時間 (黃昏)、範圍時間 (夜晚)、範圍時間 (時刻)、範圍時間 (其他)、相對時間、年齡時間
地	地名	行政地名、街坊地名、山嶺地名、河湖地名 (江河溪流)、河湖地名 (湖海池潭)、河湖地名 (沙洲)、島嶼地名、關津地名 (關隘)、關津地名 (門)、關津地名 (津堤)、關津地名 (港口)、關津地名 (橋)、虛構地名、外國地名、地名別稱 (行政地名)、地名別稱 (街坊地名)、地名別稱 (山嶺地名)、地名別稱 (河湖地名)、地名別稱 (島嶼地名)、地名別稱 (關津地名)、地名 (合稱)
地	地理	全域範圍、境外區域範圍、境內區域範圍 (全域)、境內區域範圍 (泛稱)、境內區域範圍 (東方)、境內區域範圍 (西方)、境內區域範圍 (南方)、境內區域範圍 (北方)、邦國都城、郊原村野、街道巷弄、特殊場域、場域 (概稱)
地	自然景觀	山峰崖嶺、水澤湖泊、沙洲島嶼、洞穴坑谷、林原莽漠、山川泛稱、關津渡口 (關隘)、關津渡口 (津堤)、關津渡口 (合稱)、風景勝蹟 (自然景物)、風景勝蹟 (季節景物)、風景勝蹟 (人為景物)

地	閨苑仙境	閨苑、仙境
地	幽冥之境	幽冥
地	人世之境	人間
地	空間	方向、位置、距離
物	生物	神物（動物）、神物（植物）、神物（合稱）、神物（泛稱）、動物專名（走獸）、動物專名（飛禽）、動物專名（昆蟲爬蟲）、動物專名（水生兩棲）、動物專名（合稱）、動物指稱（皮毛）、動物指稱（泛稱）、動物指稱（其他）、動物指稱（合稱）、植物專名（木本）、植物專名（草本）、植物專名（禾本）、植物專名（苔菌）、植物專名（瓜果）、植物專名（合稱）、植物泛稱（花）、植物泛稱（木）、植物泛稱（葉）、植物泛稱（枝）、植物泛稱（果）、植物泛稱（草）、植物泛稱（莖）、植物泛稱（根）、植物泛稱（苗）、植物泛稱（合稱）、生物（異類合稱）
物	狀態(生物)	動物生命狀態、植物生命狀態
物	動作(動物)	動物動作
物	動態(植物)	植物動態
物	器物	生活用品（飲食器具）、生活用品（烹飪器具）、生活用品（裝置容器）、生活用品（起居用品）、生活用品（其他用品）、生活用品（梳妝用品）、交通工具（水路）、交通工具（陸路）、交通工具（空路）、交通工具（合稱）、工具用品（兵器單稱）、工具用品（兵器合稱）、工具用品（戰爭工具）、工具用品（農牧漁獵工）、工具用品（生活工具單稱）、工具用品（生活工具合稱）、工具用品（刑具）、絲帛服飾（衣冠腰帶）、絲帛服飾（布料）、絲帛服飾（鞋襪）、絲帛服飾（配飾）、絲帛服飾（合稱）、宗教用品、喪葬祭祀、娛樂器具、金帛錢幣、公堂器物、旗幟、帳幕、神器、樂器（單稱）、樂器（合稱）、樂器（泛稱）、文具（單稱）、文具（合稱）、文具（泛稱）、器物（合稱）、器物（泛稱）
物	建築物	建築專名（宮室屋廬）、建築專名（亭臺樓閣）、建築專名（園林院落）、建築專名（寺廟道觀）、建築專名（碑塔陵墓）、宮室屋廬、亭臺樓閣、園林院落、寺廟道觀、碑塔陵墓、公署建築、建築部件、軍事攻防
物	飲食	飲品（茶）、飲品（酒）、飲品（其他）、食品、食材、糕點、調味品、菸草、宴席、飲食（合稱）
物	自然資源	玉石、火、金屬、礦物、燃料、自然資源（同類合稱）、自然資源（異類合稱）
物	天候氣象	風霜雪露、雲霧煙霞、雷電電霓、冷寒、熱旱、晴天、陰雨、天候氣象（相反合稱）、天候氣象（相近合稱）、天候氣象（相異合稱）、

		天候氣象（概稱）
物	天文	日、月、星、天空、銀河、天文（同類合稱）、天文（異類合稱）
物	事物	事物（概稱）
物	形容詞彙 （物）	景物形態、植物形態、動物形態、物品形態、環境狀態
物	顏色	原色、借色、間色、顏色（概稱）
其他	狀態（其他類）	對比狀態（其他類）
其他	形容詞彙 （其他類）	形容對比詞（其他類）
其他	狀聲詞	自然界聲音、動物聲音、人類聲音、器物聲音
其他	現象	光芒、陰影、自然現象
其他	單位詞	文藝單位、長度單位、重量單位、面積單位、容量單位、個體量詞、集合量詞、不定量詞、動量詞
其他	數詞	定量數詞、概量數詞、等級數詞

八、本研究的分類與歷代類書分類的異同

將本研究所做的分類與《藝文類聚》、《初學記》、《唐詩類苑》、《古詩類苑》、《淵鑑類函》、《唐詩分類大辭典》等六部類書比較，其分類類別數與層次數比較如下：

編者	項目及分類標的	層次數	大類 類別數	中類 類別數
唐代，歐陽詢等	《藝文類聚》	2	46	757
唐代，徐堅等	《初學記》 （敘事、事對之詞彙、詩賦文章）	2	26	313
明代，張之象	《唐詩類苑》	2	38	1089
明代，張之象	《古詩類苑》	2	44	779
清代，張英、王士禛、王愔等	《淵鑑類函》 （事對之詞彙、詩賦文章）	2	45	2806
當代，馬東田等	《唐詩分類大辭典》	2	40	1083
本研究分類		3	50	368

歷代類書用以分類的標的不同，但是都分為二層。《唐詩類苑》、《古詩類苑》、《唐詩分類大辭典》依標題或詩題字面含義分類；《藝文類聚》、《初學記》、《淵鑑類函》兼顧標題、內容主題、詞彙含義；中類之類別數量差異比較大，主要是因為不同類書內容數量及涵蓋範圍之多寡不同，分類之詳略精粗也不同，以致造成差異。《唐詩類苑》、《古詩類苑》、《唐詩分類大辭典》依標題分類，難以兼顧多重主題，其分類結果，或只取一種主題，或將同一作品分列在不同類別項下。也容易發生因為主題與詩題字面含義不相符而錯分的情形，或分類與詩作內容不合的情形，從前述所舉《唐詩分類大辭典》的例子即可知。

本研究以詩詞內容（詩句）詞彙語意分類，優點是：分類可以更切合主題含義並兼顧多元主題、可以從所分類別建立詩詞韻文之知識組織及知識體系架構、較能符合文學及文學語言學研究之文學專業使用。缺點是資料數量太大，難以單憑人力處理、雖可顯示多元主題，但主題不易集中，較難呈現主從關係。

比較各家分類，各家共有的類別如下：天部、歲時部、地部、山部、水部、州郡部、帝王部、職官部、治政部、禮部、樂部、文部、武部、人部、居處部、器用部、服食部、玉帛部、草部、果部、木部、鳥部、獸部、鱗介部、虫豸部等二十五種。各家不同的類別（含詳略不同、涵蓋範圍不同）如下：中宮部、五穀部、內典部、方術部、火部、古諺部、巧藝部、布帛部、刑法部、后妃部、存逸部、寺觀部、百穀部、舟車部、衣冠部、技藝部、災異部、京都部、明良部、封爵部、帝戚部、皇室部、軍器部、食物部、祠廟部、產業部、祥異部、祥瑞部、符命部、菜蔬部、飲食部、補亡部、道部、道釋部、儀飾部、儒家部、儲宮部、雜部、藥部、邊塞部、釋佛部、讖數部、靈異部等四十三種。

本研究分類與歷代類書比較，新增大類 25 類，中類 73 類，類別的增減分合，以詩詞語言特性及提供文學研究為主要考量，限於篇幅，不在本文討論。新增類別表列如下：

總類	大類	中類
人	感官詞	嗅覺、聽覺、味覺、視覺、觸覺
事	行政體系	部門名稱、朝代名號（單稱）、朝代名號（合稱）、民族邦國名（單稱）、民族邦國名（合稱）、國名
地	地名	外國地名、地名（合稱）
	地理	特殊場域、場域（概稱）
	自然景觀	山川泛稱、風景勝蹟（自然景物）、風景勝蹟（季節景物）、風景勝蹟（人為景物）、
	閬苑仙境	閬苑、仙境
	幽冥之境	幽冥
	人世之境	人間

	空間	方向、位置、距離
物	生物	神物（動物）、神物（植物）、神物（合稱）、神物（泛稱）、生物（異類合稱）
	狀態（生物）	動物生命狀態、植物生命狀態、
	動作（動物）	動物動作
	動態（植物）	植物動態
	器物	交通工具（合稱）、器物（合稱）、器物（泛稱）
	建築物	建築專名（碑塔陵墓）、碑塔陵墓
	自然資源	礦物、自然資源（同類合稱）、自然資源（異類合稱）、
	事物	事物（概稱）
	形容詞彙（物）	景物形態、動物形態、物品形態、環境狀態
	顏色	原色、借色、間色、顏色（概稱）
其他	狀態（其他類）	對比狀態（其他類）
	形容詞彙（其他類）	形容對比詞（其他類）
	狀聲詞	自然界聲音、動物聲音、人類聲音、器物聲音
	現象	光芒、陰影、自然現象
	單位詞	文藝單位、長度單位、重量單位、面積單位、容量單位、個體量詞、集合量詞、不定量詞、動量詞
	數詞	定量數詞、概量數詞、等級數詞

九、結論與未來的發展

詩詞語言的特殊性增加了詩詞詞彙切分及語意標記、語意分類的困難，使得詩詞語意概念的分類，除了必須符合知識認知層次的語言知識分類，還需顧及情感認知層次的意象取向分類。詩詞的詞彙以雙字詞居多，但是詩詞句子的字數多數是奇數，所以多數奇數字數的句子之中都會有一個單字詞，單字詞的詞義比雙字詞更豐富多變，這也增加了詞義標記的複雜性。

詩詞裡典故用語的標記方式比一般用語困難，以某些詞彙共現以烘托某種情境的情境用語，其詞義無法在單一詞彙之詞義顯示或涵蓋，這些特殊詞彙的語意，需要仰賴更多相關的研究，用更多分析、統計、歸納的方法尋找出規則。

詩詞語意標記的困難，還包括詩詞喜用多義字來創造聯想的空間，以引出無窮的趣味及意在言外的情境。袁行霈教授〈中國古典詩歌的多義性〉文中即闡明：「一個多義詞固然有多種意義，歧義是一般情況下使用語言時需要特別避忌的毛病。但是在詩詞裡，恰恰要避免詞義的單一化，總是盡可能地使用詞語帶上多種意義，以造成廣泛的聯

想，取得多義的效果。中國古典詩詞的耐人尋味，就在於這種複合的作用。」⁴²，文中袁教授並將詩詞語言的多義性分為宣示義及啟示義二種，啟示義包含雙關義、情韻義、象徵義、深層義、言外義五種⁴³。宣示義是借助語言明確表達給讀者的語意，如王維〈積雨輞川莊作〉：「漠漠水田飛白鷺，陰陰夏木轉黃鸝。山中習靜觀朝槿，松下清齋折露葵。」只有字面含義。雙關義、情韻義、象徵義能夠從統計學歸納出規則，能夠從單一詞彙標記其多重含義，如前述的「苔、蘚」，但是深層義與言外義的含義隱藏在字詞表面含義之外，以整首詩來烘托，這些詩即是所謂「言有盡而意無窮」、「不著一字而盡得風流」者，讀者要了解這些深層含義，必須要了解作者及寫作對象的背景資料，必須要具備更豐富的學養及詩詞鑑賞能力才能夠體會，以詞彙為單位的語意標記無法完整標記這一類詩詞的完整含義，想要讓電腦理解，誠非易事。

宋代以後，受到理學發達及禪宗思想的影響，有很多的哲理詩，如蘇軾〈題沈君琴〉：「若言琴上有琴聲，放在匣中何不鳴？若言聲在指頭上，何不於君指上聽？」楊萬里〈戲筆〉：「野菊荒苔各鑄錢，金黃銅綠兩爭妍。天公支與窮詩客，只買清愁不買田。」這些哲理詩正是劉大魁所說的「只可意會而不可言傳」的作品，需仰賴讀者的哲理思維領悟，當然也不是單獨的詞義標記所能夠清楚描述的，想要借助詞義標記或語意概念分類讓電腦能理解，其難度又比深層義及言外義更高，在分類上，只能以詩的主題作分類。

從以上所述可知，可以透過語意標記及語意概念分類，讓電腦從分辨字形提升到理解詞義，使電腦的認知能力更接近人的知識體系。人的知識體系分為「對知識的認知」、「對情意的感知」、「對哲理的辨知」，認知是表述的層次，感知需要聯想輔助，哲理詩的解讀體會，卻不是單純的邏輯思維能解決，也就更難以用語意標記處理。

唐宋詩詞語意標記及分類的架構是以唐宋詩詞的文本內容為範圍，並不包含唐宋詩詞內容之外的其他詞彙或知識。本研究進行的過程，最初是以蘇軾全部詩作為基礎，建立蘇軾詩作的詞彙庫及語意標記及分類資料庫，其後加入了宋代七家詞人⁴⁴、四百萬字的全台詩⁴⁵，以這些文本的詞彙庫、語意標記及分類資料庫為基礎，進行唐宋詩詞的詞彙切分、語意標記及分類。整合不同文體（詩、詞）、不同時代（唐、宋、清末至民國）、不同寫作背景（中國大陸及台灣），但是構詞方式相近的韻文文本，所包括的詞彙也因為文體、時代、寫作背景的差異而有些微不同，因此需要再進行人工校對。

⁴² . 引自袁行霽，〈中國古典詩歌的多義性〉，收於《中國詩歌藝術研究》，五南圖書出版公司印行，1989年5月台灣初版，第6頁。

⁴³ . 同註33。

⁴⁴ . 七家詞人包含柳永（218闕詞）、蘇軾（379）、秦觀（148）、周邦彥（187）、李清照（53）、姜夔（87）、辛棄疾（627）等，選擇的標準考量北宋與南宋的時代均衡及婉約派與豪放派的風格均衡。

⁴⁵ . 參見羅鳳珠主持，【智慧型全台詩知識庫計畫第二階段建置計畫】，國家台灣文學館，2006年1月1日至2006年12月31日，網址：<http://cls.hs.yzu.edu.tw/TWP/index.asp>。

唐宋詩詞網站所包含的文本數量龐大，人工校對所需要花費的人力很多。本研究的語意標記及分類只針對表中項目所訂的類別分類，撰寫本文時，還來不及完成全部文獻的人工校對，語意的分類的完整性及準確性必定還有不足之處，也勢必會隨著文本的增加而調整，尚待後續的研究修正及補足。本研究目前只完成上下階層的樹狀結構，尚未處理橫向階層的相關性結構，這些都有待後續的研究補足。

語意標記及語意概念的分類，除了可以據以建立詩詞語言語意資料庫、詩詞語言索引典、詩詞語意概念分類知識體系，以及文學及語言學研究使用，除此之外，還可以提升資訊檢索的質量。資訊檢索所涉及的相關議題，包括認知的元素、語意與多重表述的認知觀點，詞語概念及表述的關聯性，資訊行為的研究、使用者的交互作用，檢索的策略、搜尋引擎、檢索方法的應用等等。被尊為資訊科學之父的Vannevar Bush於1945年所發表的“*As We May Think*”一文提出：記錄思想的文獻儲存在圖書館，圖書館的資料按照字母或數字排列的索引，只能一層一層往下尋找，不符合人類大腦以思維聯繫尋找資料的方式，因此，Bush便構思著建立一個以聯繫（association）代替索引（indexing）以選擇資料的機械化個人文件夾和圖書館（mechanized private file and library），Bush將之命名為Memex。Bush期待在這個機器裡可以儲存所有的文獻資料，檢索時能如同人類的思維邏輯，建立知識連結的軌跡以進行聯繫索引，需要引用時，所有的資料都可以隨意檢索出來，並且可以快速的找到另一個相關的資料，也可以二個條件同時並行檢索，這就是Memex的基本概念。⁴⁶ Bush的這篇文章，被認為是第一篇論及超文件基本觀念的文獻，啟發資訊科學家對資訊檢索技術之研究。依據語意概念分類知識體系建立能分辨語意及含有上下層級關係，整體與部分關係等等的語意概念分類知識體系，正符合「以聯繫取代索引」的資訊搜尋方法。

本研究從語意學及使用者的認知觀點，以唐宋詩詞網站的設計作為實證的研究，初步建立唐宋詩詞語意標記資料庫及唐宋詩詞語意概念分類表，期待後續的研究，能進一步彙整、分析、歸納唐宋詩詞語意類別，再與共現詞彙比對，期望能從中建立唐宋詩詞主題分類的方法與規則，以彌補《唐詩分類大辭典》等前代類書分類的缺失與不足。

後記：

本研究感謝國科會：【新詩改罷自長吟—唐詩吟唱虛擬實境教學環境建置計畫】，國科會數位典藏國家型科技計畫：95年度數位典藏創意學習計畫，2006年3月1日至2007年2月28日執行。計畫編號：NSC95-2422-H-155-001。以及【淺斟低唱—宋詞古唱虛擬實境教學環境建置計畫】，國科會數位典藏國家型科技計畫：94年度數位典藏創意學習計畫，2005年3月1日至2006年2月28日。計畫編號：NSC 94-2422-H-155-001

⁴⁶ . V. Bush, “As we may think” *The Atlantic Monthly*, Vol.176, No.1 pp.101~108, 1945。本段文字從“*As we may think*”原文摘錄重點。

等二項研究計畫經費資助，感謝計畫共同主持人張智星、許介彥教授。感謝參與計畫之助理：程式設計陳駿銘、林宜嫻、李元萍、曹偉政；資料分析整理：劉怡姍、吳侑倣、詹錦麗。研究過程中，得到中研院鄭錦全院士、業師黃啟方教授指導，北京大學蔣紹愚教授、袁行霈教授、俞士汶教授提出訂正意見，台灣大學圖書資訊研究所陳雪華教授、藍文欽教授、黃慕萱教授給予很多指導，獲益匪淺，一併致謝。