

微軟亞洲研究院

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

元智大學中國語文學系・羅鳳珠

20080714

大綱

- 一. 詩詞文學的內容
- 二. 文學的表達與共鳴
- 三. 詩詞文學語言的特殊性
- 四. 概念體系
- 五. 詩詞語言的語意概念分類方法及分類體系
- 六. 語意概念分類總表
- 七. 歷代類書分類比較
- 八. 本研究新增類別
- 九. 本研究的進展
- 十. 語意概念分類：對資訊檢索質量與使用需求的影響
- 十一. 語意概念分類在教學與研究上的應用
- 十二. 遭遇的困難與未來的發展

詩詞文學的內容

文學的內容

- 一、人與外物的互動
- 二、客觀的外在景物
- 三、主觀內在的心靈活動

以賀鑄，〈眼兒媚〉為例：

蕭蕭江上荻花秋，作弄許多愁。（人與外物的互動）

半竿落日，兩行新雁，一葉扁舟。（客觀的外在景物）

惜分長怕君先去，直待醉時休。今宵眼底，明朝心上，後日
眉頭。（主觀內在心靈的活動）

詩詞文學內容的元素1/3

詩詞文學內容的元素：人、事、時、地、物
那些人在那些時間，那些地方做過那些事，
牽涉那些物。

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、詩詞因篇幅較小，其表述方式，或直
接描寫，或隱藏在字裡行間，間接表述。

詩詞文學內容的元素2/3

直接描寫：

以「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。」為例

人：蓑笠翁

事：獨釣

時：寒冬

地：江邊

物：孤舟、鳥

詩詞文學內容的元素3/3

間接表述：

雖未直接描寫，仍於詩意之中表達

以「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。」為例

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

時：三國時代

地：魚復（腹）＝今四川奉節縣

物：八陣圖

文學的表達與共鳴：形式與內容1/3

一、形式：文字、文體

二、內容：

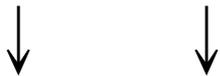
言志、言情……



主觀的情感

→ 客觀的 (媒介) →

客觀的敘述



敘事、詠史、詠物、寫景……

| 物象 → 意象

| 物象 → 意象 → 意境

| 物象 → 意象

文學的表達與共鳴：形式與內容2/3

賀鑄〈青玉案〉：試問閒愁都幾許？一川湮草，滿城風絮，梅子黃時雨。

作者抽象的情感：試問閒愁都幾許？

閒愁都幾許？如何量長短、稱輕重、衡多寡、掂厚薄？

作者與讀者共有的經驗：一川湮草，滿城風絮，梅子黃時雨
（都是具體可見可知可感的物象） → 引發意象

以具體的物象作為引發作者與讀者共鳴的媒介

媒介：文字→物象→意象

文學的表達與共鳴：以感官經驗引起共鳴3/3

人以**感官**感受外物：作者透過**感官**感受的描寫以引發讀者相同的**感官**經驗，進而體會作品，與作者共鳴。

李煜〈玉樓春〉

晚妝初了明肌雪。春殿嬪娥魚貫列（**視覺**）。鳳簫吹斷水雲間，重按霓裳歌遍徹。（**聽覺**）臨風誰更飄香屑（**嗅覺**）。醉拍闌干情味切（**味覺**）。歸時休放燭花紅，待踏馬蹄清夜月（**心境情境綜合感覺：情意感知**）。

蘇軾〈永遇樂〉：明月如霜，好風如水（**觸覺**）。

詩詞文學語言的特殊性1/6：構詞

構詞的特殊性

除了使用專有名詞或典故之外，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特殊。

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。

詩詞文學語言的特殊性2/6：構詞

1. 散文少用：

白居易：「**離離**原上草，一歲一枯榮」、蘇武：「參辰皆已沒，**去去**從此辭」、柳永：「念**去去**，千里煙波，暮靄沉沉楚天闊。」

2. 分散在上下句：

蘇軾：「詩人老去**鶯鶯**在，公子歸來**燕燕**忙」

3. 縮略倒裝語

蘇軾：「如今勝事無人共，花下**壺盧**鳥勸**提**（花下**提壺****盧**鳥勸人提起**葫蘆**喝酒）」

4. 名詞雙舉並列，以豐富詞義：

孟浩然：「把酒話**桑麻**」、蘇庠：「寄語故時**猿鶴**侶」

詩詞文學語言的特殊性3/6：語意

一般用語：只使用單一語義

所使用的詞彙含義與其他文體無異。

無論該詞彙含有幾種含義，「一般用語」一次只使用其中一種含義。

例如：

杜甫〈曲江〉：穿花蛺蝶深深見，點水蜻蜓款款飛。

李白〈登金陵鳳凰臺〉：三山半落青天外，二水中分白鷺洲。

詩詞文學語言的特殊性4/6：語意

特殊用語：雙關義、引申義、典故、象徵、比喻……，同時使用多元語義

詞義的特殊：使用雙關義、引申義以豐富詩意。（其他文體講究語意精確，詩詞正好相反）

韓翃〈章台柳〉：章台柳，章台柳，昔日青青今在否？縱使長條似舊垂，也應攀折他人手。

柳姬〈楊柳枝〉：楊柳枝，芳菲節，苦恨年年贈離別。一葉隨風忽報秋，縱使君來豈堪折。

「柳＝柳樹＝柳姬」（雙關義）＝離別折柳相贈＝柳姬身陷風塵，就如柳葉，終日送往迎來（引申義）

以文字之多義性或賦予雙關義、引申義的特性來豐富內容。

詩詞文學語言的特殊性5/6：語意

特殊用語：雙關義、引申義、典故、象徵、比喻……，同時使用多元語義

詞義的特殊：使用典故義

蘇軾：「共誰交臂論今古，只有閑心對此君。」（此君＝酒）

蘇軾：「若對此君仍大嚼，世間那有揚州鶴。」（此君＝竹子）

詞義的特殊：深層義

杜甫〈江南逢李龜年〉：「岐王宅裡尋常見，崔九堂前幾度聞。
正是江南好風景，落花時節又逢君。」

落花時節＝植物的落花時節＋人生的晚年＋國運的衰微期，有三層含義。

詩詞文學語言的特殊性6/6：語意

情境用語：詞義特殊，語義須與其他詞彙共現才能顯示

以「苔蘚」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。詩人循此「認知」所引發的「感知」如下：
- 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。」的失意文人，或是「玉階陰苔蒼蒼，君履難再得。」的失寵佳人，或是「天花寂寂香深殿，苔蘚斑斑空閨女子，或是「花木閒門苔蘚生，澹川是雪霜去得吟情。」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：「流水君恩不再使，無人前來，階閒無用而長滿綠苔。」因君恩不回，沒有

概念體系 1/6

George Lakoff:

- 何謂概念體系？他是如何組成的？
 - 所有的人是否都有同樣的概念系統？
 - 人的客觀認知與主觀感知的區別
 - 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重要的。
-
- ~Women, Fire, and Dangerous Things: What Categories Reveal about The Mind~

概念體系 2/6

人的感官經驗和運用想像機制

以前述「苔蘚」為例

對苔蘚的客觀認知：生長在潮濕而人跡罕至之處

人的經驗：看過道上、院中、門前、階下長滿苔蘚之處所，長滿苔蘚之景象。

人的想像機制：此處住著失意文人、失寵佳人、塵外僧人、空閨女子、山林雅士。

概念體系3/6

人的客觀認知與主觀感知

人的「客觀認知」+「主觀感知」+「思維體系」=構成「概念系統」

人以肉體經驗（感官系統：耳鼻口眼皮膚心智）認知客觀外物。

因為經驗、情感、性格的不同而產生不同的感知，所引發的感知是主觀的，可以分為情感感知與情境感知二種，後者受環境因素影響。

客觀事物透過想像的機制而創造不同的主觀感知。

客觀認知是理性的，個別差異比較少。

主觀感知是感性的，個別差異比較多。

概念體系4/6

所有的人是否都有同樣的概念系統？

人以生理經驗（感官系統：耳鼻口眼皮膚心智）認知客觀外物，所產生的概念體系，有共通性（多數人相同）。

但是因某些個別的差異而有例外。

- 例如：「花月人酒」被多數詩人認為是自然界與人世間美好的事物：
- 李白：「花間一壺酒，獨酌無相親。舉杯邀明月，對影成三人。」
- 李商隱：「縱使有花兼有月，可堪無酒又無人。」
- 李煜：「春花秋月何時了」（亡國的經驗）

概念體系5/6

人的概念系統

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（ontology）解構文學作品，重新建構成為電腦可以讀懂的知識，便可以使電腦具備人的知識體系。

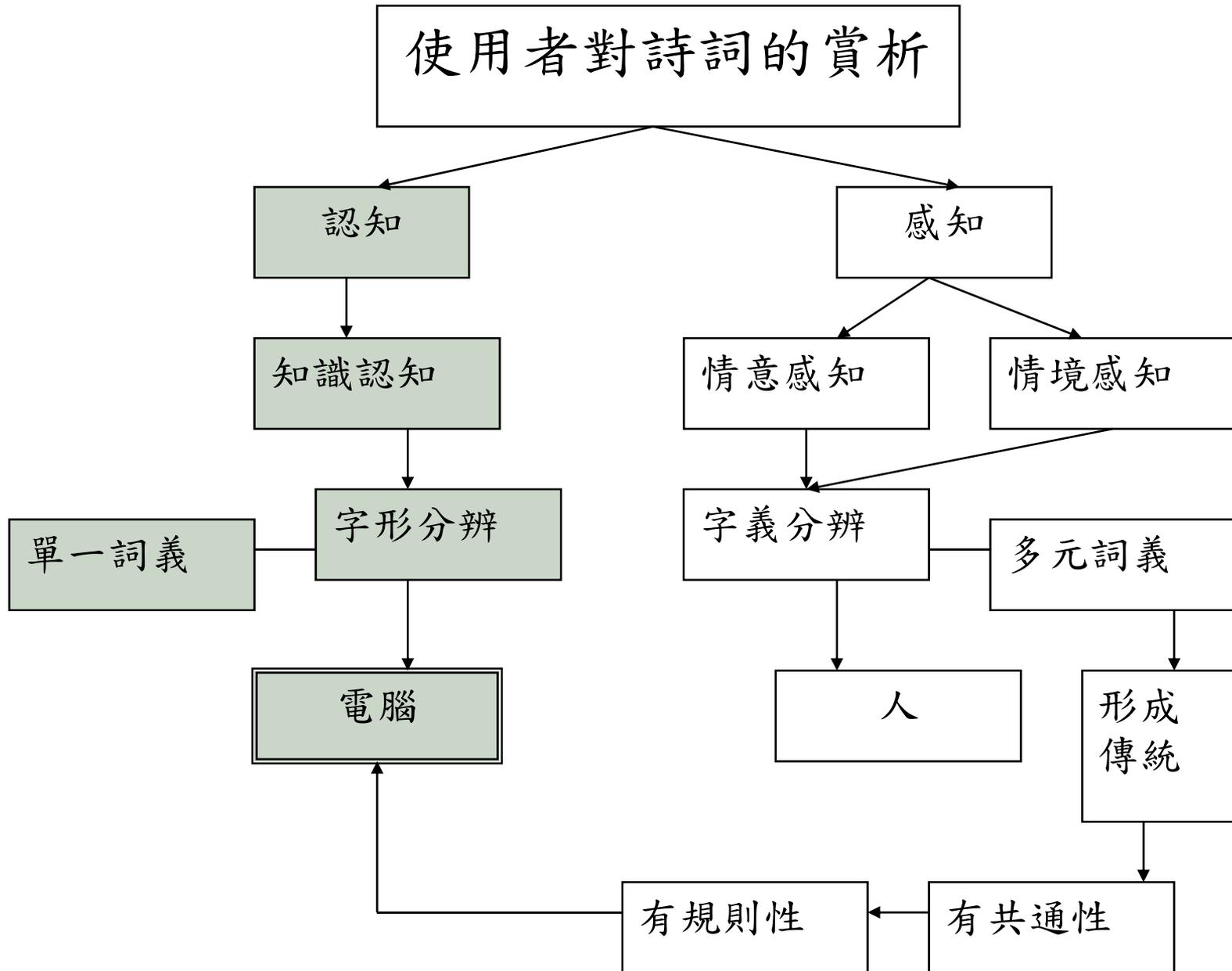
概念體系 6/6

本體論

用本體知識的基本元素（概念）描述知識表達的結構（概念之間彼此的關聯）。

知識本體論是用於描述及表達特定領域知識的一組概念或術語，它可以用來研究特定領域知識的類別、類別的屬性，以及類別與類別間的關係，進一步達成概念語意的資訊檢索。

詩詞語言的語意概念分類方法及分類體系 1/13



詩詞語言的語意概念分類方法及分類體系 2/13

- 使用者的**認知行為**，除了前述所論及對詞彙的**知識認知、情意與情境感知**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**理解與感發**能力。

詩詞語言的語意概念分類方法及分類體系

詞彙切分^{3/13}

- 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。（詞彙語意標記→建構語意概念→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。

詩詞語言的語意概念分類方法及分類體系

詞彙切分^{4/13}

- 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，下三部分作上二下一，或作上一下二。
- 七言詩的句法多數是二二三，下三部分作上二下一，或作上一下二。

詩詞語言的語意概念分類方法及分類體系

詞彙切分^{5/13}

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；（柳永：對／瀟瀟暮雨灑江天，一番洗清秋。漸／霜風淒緊，關河冷落，殘照當樓。）

三、典故。

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。

詩詞語言的語意概念分類方法及分類體系

詞彙切分^{6/13}

一、建立人名、地名……等各種專有名詞詞彙資料庫（辛棄疾〈玉樓春〉：提壺沽酒已多時，婆餅焦時須早去。婆餅焦是飛禽，提胡盧也，杜鵑也。）

二、建立詩詞典故詞彙資料庫

三、建立詩譜、詞譜資料庫，分析其句法，並將領字位置標示出來。

詩詞語言的語意概念分類方法及分類體系

詞彙切分^{7/13}

四、詞彙切分方法及步驟：

1. 第一優先比對**專有名詞**資料庫，以專有名詞做為詞彙切分之最優先順序。
2. 第二優先之切分順序是**領字**：依據所建立的宋詞詞譜優先將領字切分，其餘詞句依「宋詞詞譜」之步驟切分，例如：「**漸**／霜風／淒緊／，關河／冷落／，殘照／當樓／。」（只有詞需要此步驟）
3. 第三優先比對典故詞彙資料庫

詩詞之句首，除了使用專有名詞、典故或領字之外，絕大多數都是雙字詞。

詩詞語言的語意概念分類方法及分類體系

詞彙切分 8/13

4. 其餘句子之切分步驟

- 句子字數為偶數句之句子，兩兩切分為雙字詞，例如：「離恨／恰如／春草／，更行／更遠／還生／」。
- 句子字數為奇數句之句子，末三字之外的句子，兩兩切分為雙字詞，例如：「菡萏／香消／翠葉殘／，西風／愁起／綠波間／，還與／韶光／共憔悴／，不堪看／。」。
- 奇數句句子之末三字，其句法或作上二下一，或作上一下二，並無準則，切分的方法需仰賴詞庫以提高切分之正確率。切分時以前述步驟所獲得之雙字詞比對句末之三字詞，若這三個字之前二字或後二字是前述步驟所獲得之雙字詞，則優先取為雙字詞，予以切分。

不過也有末三字之前二字或後二字都是詞彙的句子，此時便需以人工校正，例如：「歸心正似**三春草**」（蘇軾〈虞美人〉）之「三春」、「春草」都是宋詞常用的詞彙，遇到這種情況，便需以人工校正。

詩詞語言的語意概念分類方法及分類體系 10/13

控制詞彙	同近義詞詞群（詩詞詞彙）
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、崙崙山、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

詩詞語言的語意概念分類方法及分類體系 11/13

- 本研究將詩詞語意概念分為四層：總類、大類、中類、控制詞彙。
- 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。
- 總計於6「總類」之下再分為54「大類」、372「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：

詩詞語言的語意概念分類方法及分類體系 12/13

- 詞類是按詞的語法性質和功能分類，可以分為實詞和虛詞兩大類，**實詞**表示比較實在的意義，能夠單獨作句子成分；**虛詞**一般沒有實在的意義，不能單獨作句子成分，只是在句子的結構上起語法作用，協助實詞表達意義，使句子的變化更多和表達的意思更豐富。
- 予以標記語意及分類的詞彙**以實詞**(名詞、動詞、形容詞、數詞、量詞、代詞)**為原則**，虛詞(副詞、介詞、連詞、助詞、嘆詞)的語意需附著在實詞上，詞義才能完整顯示，因此這些詞彙不予分類。
- 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。

詩詞語言的語意概念分類方法及分類體系 13/13

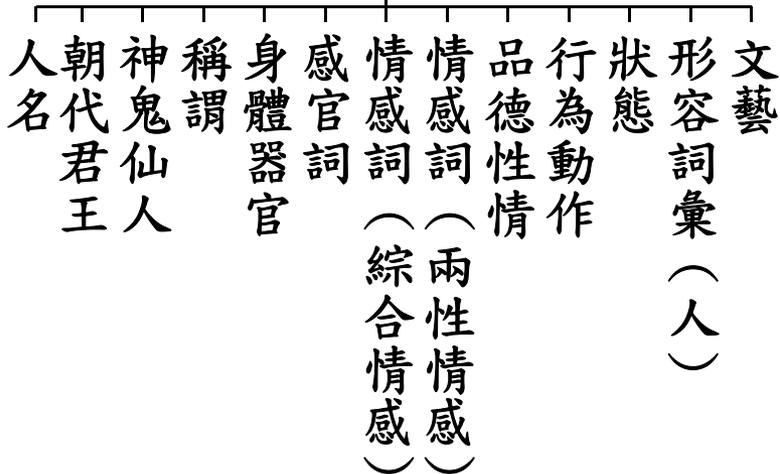
本研究以《樂府詩集》5,156首、《全唐詩》51,171首、《唐宋詞》24,226首、宋名家詩44,314首、《全台詩》23,066首，總計147,933首為範圍，建立詩詞語意標記與分類架構。

本研究分類的標的是中國詩詞文獻，所建立的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。

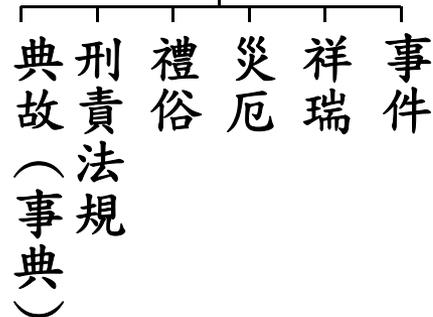
所得結果是詩詞內容所包含的語意，不是完整的人類知識體系，因此有些古典詩詞缺乏的內容，其知識類別便不會出現在詩詞語意概念分類體系裡。

語意概念分類總表

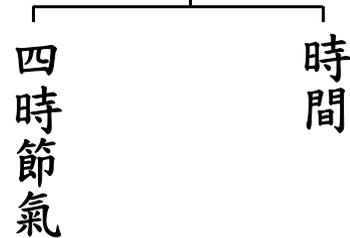
人



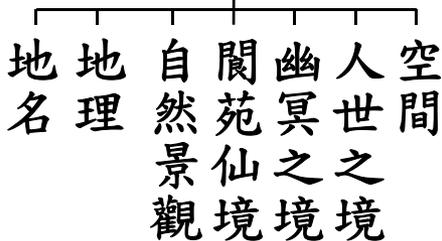
事



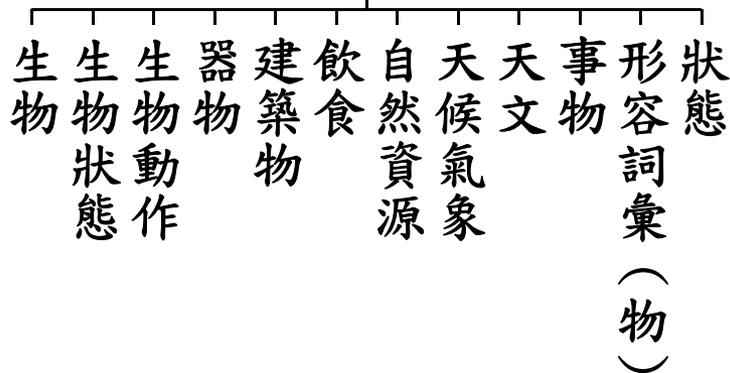
時



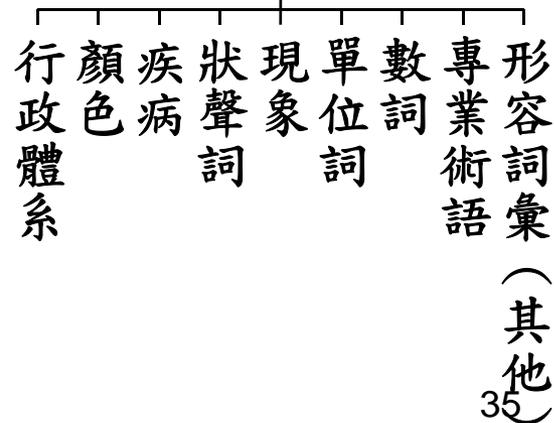
地



物



其他



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

36

形容詞彙（人）

↓ 氣質內涵、人生機遇、官場境遇

狀態

生理狀態、心智狀態、**心理能願**、生活狀態、心神氣力、合稱（相反詞、相近詞、相異詞）

行為動作

一般行為（**人部、刀部、口部、土部、彳部、日部、水部、糸部、衣部、艸部、言部、貝部、走部、足部、馬部、宀部、支部、走部、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

品德性情

↓ 正面、負面

情感詞（兩性情感）

↓ 傾心愛戀、相思怨懟

情感詞

正面情感（喜悅欣樂、自我肯定、景仰羨慕、悠然自適、憐惜愛寵）、

（綜合情感）

負面情感（自我否定、惱怒怨妒、悲哀傷痛、感懷惆悵、憂愁掛慮、厭煩憎惡、懊惱憾惜、擔驚受怕、辛酸艱苦、慚愧自責、驕矜自恃、思鄉念舊、尷尬為難）、合稱（相反詞、相近詞）

感官詞

↓ **嗅覺、聽覺、味覺、視覺、觸覺、綜合感覺詞**

身體器官

五官顏面、鬚髮髮膚、身體四肢、五臟器官、分泌物、合稱

稱謂

人稱代名詞、自稱謙稱、尊稱美稱、貶稱惡稱、親人眷屬、指代稱謂、一般稱謂、職業身份、職官爵位、皇室稱謂、宗教稱謂

神鬼仙人

↓ 神鬼專稱、神鬼泛稱、菩薩尊號

朝代君王

↓ 君王帝號、君王年號、君王陵墓

人名

↓ 本名、別名字號、他稱、法號、合稱、外國人名、虛構人名、待考

事

事件

↓ 史事、事情概稱、世局、**國**
家氣運

祥瑞

↓ **福泰祥安、祥瑞預兆**

災厄

↓ 旱災、水災、地震、兵災、
其他災害、災厄預兆

禮俗

↓ 祭祀、婚喪喜慶、其他

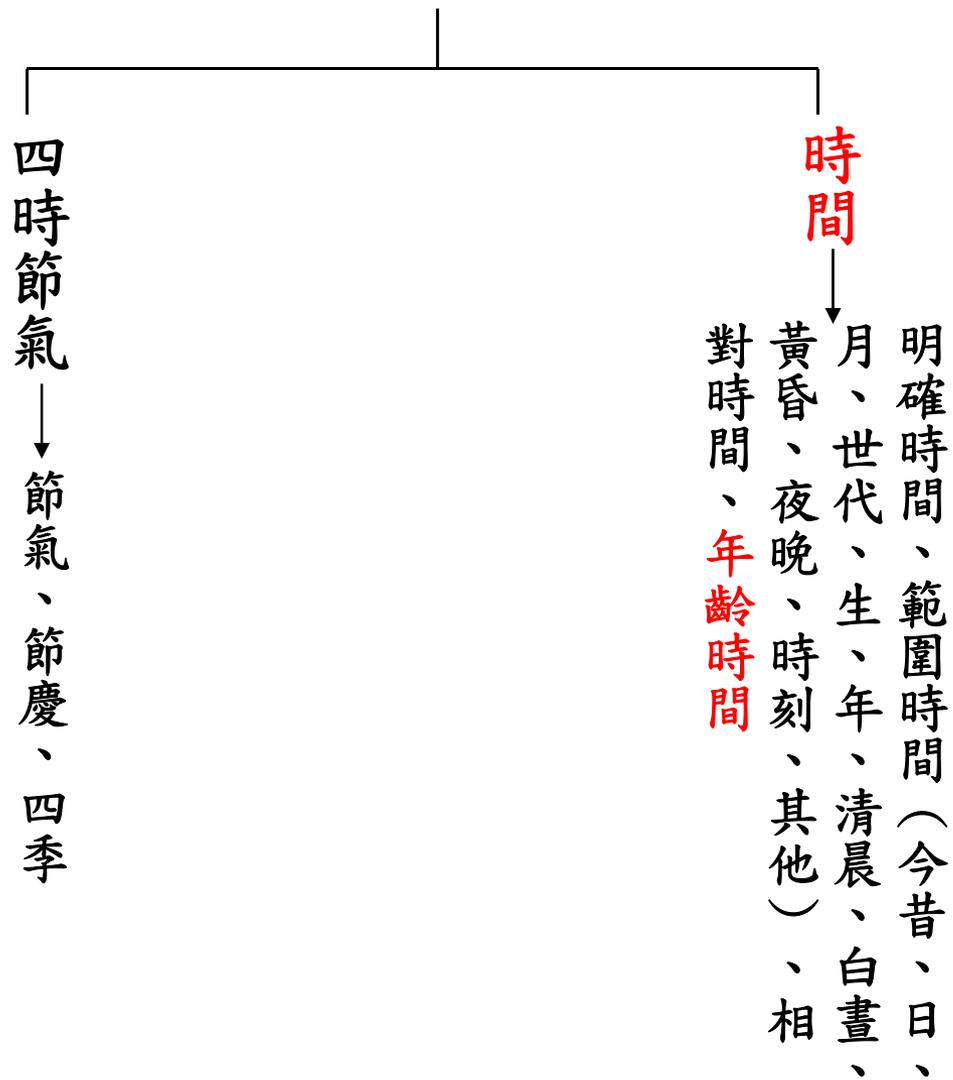
刑責法規

↓ 刑責、法規

典故

↓ 佛典、神話典故、其他

時



地

空間 → 方向、位置、距離

人世之境 → 人間

幽冥之境 → 幽冥

閬苑仙境 → 閬苑、仙境

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、津堤、合稱)、風景勝蹟(自然景物、季節景物、人為景物)

地理

全域範圍、境外區域範圍、境內區域範圍
(全域、泛稱、東方、西方、南方、北方)、邦國都城、郊原村野、街道巷弄、特殊場域、場域概稱

地名

行政地名、街坊地名、山嶺地名、河湖地名
(江河溪流、湖海池潭、沙洲)、島嶼地名、
關津地名(關隘、門、津堤、港口、橋)、
虛構地名、外國地名、地名別稱(行政地名、
街坊地名、山嶺地名、河湖地名、島嶼地名、
關津地名)、地名合稱

物

狀態 → 對比狀態

形容詞彙（物） → 景物形態、植物形態、動物形態、
物品形態、環境形態

事物 → 事物概稱

天文 → 日、月、星、天空、銀河、合稱（同類、異類）

天候氣象 → 風霜雪露、雲霧煙霞、雷電電霓、冷寒、熱
早、晴天、陰雨、合稱（相反詞、相近詞、
相異詞）、概稱

自然資源 → 玉石、火、金屬、礦物、燃料、合稱（同類、
異類）

飲食 → 飲品（茶、酒、其他）、食品、食材、糕點、調
味品、菸草、宴席、合稱

建築物 → 建築專名（宮室屋廬、亭臺樓閣、園林院落、寺廟道
觀、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

器物 → 生活用品（飲食器具、烹飪器具、裝置容器、起居用
品、其他用品、梳妝用品）、交通工具（水路、陸路、
空路、合稱）、工具用品（兵器單稱、兵器合稱、戰
爭物品、農牧漁獵工、生活工具單稱、生活工具合稱、
刑具）、絲帛服飾（衣冠腰帶、布料、鞋襪、配飾、
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、合稱、
泛稱）、文具（單稱、合稱、泛稱）、器物合稱、器
物泛稱

生物動作 → 動物動作、植物動作

生物狀態 → 動物生命狀態、植物生命狀態

生物 → 神物（動物、植物、合稱、泛稱）、動物專名（走獸、
飛禽、昆蟲爬蟲、水生兩棲、合稱）、動物指稱（皮
毛、泛稱、其他、合稱）、植物專名（木本、草本、
禾本、苔菌、瓜果、合稱）、植物泛稱（花、木、葉、
枝、果、草、莖、根、苗、合稱）、異類合稱

其他

形容詞彙（其他）→ 對比詞

專業術語 → 佛家語、作戰語、法術語、道家語

數詞 → 定量數詞、概量數詞、等級數詞

單位詞 → 文藝單位、長度單位、重量單位、面積單位、容量單位、量詞單位

現象 → 光芒、陰影、自然現象

狀聲詞 → 自然界聲音、動物聲音、人類聲音、器物聲音

疾病 → 疾病療法、疾病類別

顏色 → 原色、借色、間色、顏色概稱

行政體系 → 部門名稱、朝代名號、朝代名號合稱、
民族邦國名、民族邦國名（合稱）、國名

歷代類書分類比較^{1/7}

編者	項目	大類數	層次
唐代，歐陽詢等	《藝文類聚》	46	2
唐代，徐堅等	《初學記》（敘事、事對之詞彙、詩賦文章）	26	2
明代，張之象	《唐詩類苑》	38	2
明代，張之象	《古詩類苑》	44	2
清代，張英、王士禎、王愔等	《淵鑑類函》（事對之詞彙、詩賦文章）	45	2
當代，馬東田	《唐詩分類大辭典》	40	2
本研究分類		54	4

歷代類書分類比較^{2/7}

一、分類標的不同：

1. 《唐詩類苑》 《古詩類苑》 《唐詩分類大辭典》：依詩題字面含義分類
2. 《藝文類聚》 《初學記》 《淵鑑類函》：兼顧題目、內容、詞彙
3. 本研究以詩句詞彙語意分類。

歷代類書分類比較^{3/7}

二、以詩題分類的缺點：

1. 依詩題字面含義分類，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。也容易發生主題與詩題字面含義不相符的情形。

2. 分類與詩作內容不合：以《唐詩分類大辭典》為例

有些琴歌琴曲題名文字字義與內容無關，如〈關山月〉是漢辭重懷部。然詩
樂府〈橫吹曲〉和一首征人書共寫男女之情。
思念家人，十邊塞依唐人所寫的詩
是描寫齊《全唐》的詩
點觸媒上《全唐》的詩
事實及寫月，霍總
寫月，霍總

杜牧〈金谷園〉歸在地理部的「園」類，與詩的主題不合。

歷代類書分類比較^{4/7}

三、其他缺點：

1. 標舉小類的體例不一：有些類別只列專有名詞，有些類別專有名詞與普通名詞混合。
2. 專名與通稱、俗名混合，未作詞彙控制，以電腦處理時，電腦無法分辨。

四、依詩句分類的優缺點

1. 缺點：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，無法呈現主從關係。
2. 優點：分類可以更切合主題含義並兼顧多元主題。
3. 優點：提供文學及文學語言學研究及教學使用
4. 從所分類別建立詩詞韻文之知識組織及知識體系架構。

歷代類書分類比較^{5/7}：各家類別異同

一、相異

1. 類別數量不同：因各家所分類別粗細不同
2. 標題不同：標題使用詞彙不同，但含義相同或相近
3. 分類標的不同：有些以題目分類，有些以內容，有些以詞彙分類

二、相同

1. 分類層級二層，以篇題名為分類基礎
2. 《初學記》及《淵鑑類函》多了詞彙分類，其他以篇題名或內容分類

歷代類書分類比較^{6/7}：各家共有的類別

- 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

註：標題名稱採《唐詩類苑》 標題₄₇

比較歷代類書分類^{7/7}：各家不同的類別

以下各類別，各家詳略不同，整合如下：

- 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

本研究新增類別：總類、大類、中類、小類

- 人世之境、方言俚語、生物、生物狀態、
- 生物動作、同近義詞、地名、地理、
- 自然景觀、自然資源、行政體系、
- 形容詞彙、事物、狀態、狀聲詞、
- 空間、幽冥之境、建築物、時間、
- 現象、單位詞、虛詞、感官詞、
- 感嘆詞、數詞、複疊詞、閨苑仙境、
- 雙字詞、顏色（以筆畫順序排列）

本研究新增類別：總類、大類、中類、小類

- 人間、人類聲音、文藝、方向、方言、仙境、外國地名、
- 民族邦國名、民族邦國名（合稱）、光芒、同近義詞、
- 合稱（同類）、合稱（異類）、年齡時間、自然界聲音、
- 自然現象、位置、事物概稱、味覺、定量數詞、
- 物品形貌、物態形貌、長度、俚語、幽冥、重量、
- 面積、風景勝蹟、借色、原色、容量、特殊場域、
- 動物生命狀態、動物形貌、動物動作、動物聲音、國名、
- 異類合稱、部門名稱、陰影、朝代名號、朝代名號合稱、
- 植物生命狀態、植物動作、等級數詞、虛詞、虛構地名、
- 視覺、距離、量詞、間色、嗅覺、感嘆詞、概量數詞、
- 碑塔陵墓、對比狀態、對比詞、綜合感覺詞、複疊詞、
- 閨苑、器物聲音、雙字詞、顏色概稱、礦物、觸覺、聽覺。

本研究的進展1/2

一、在前人分類的基礎上增進

二、以蘇軾詩詞及《全宋詩》、宋代名家詞為範圍。初步建立分類架構，再從所建立的分類架構及所抽取及標註的詞彙語意資料庫，標註其他韻文，再進行人工校正。

三、切分詩文內容之詞彙，進行詞彙語意標記，再依詞彙語意類別分類，並作詞彙控制。

本研究的進展2/2

三、以詩文內容詞彙語意概念分類

四、分類層級由二層增為三層，並建立「詞彙控制」項。

1. 總類：人、事、時、地、物、其他等六類
2. 大類：目前分為54大類
3. 中類：目前分為372小類
4. 控制詞彙：同義詞詞彙之控制詞彙，共3,165種（不含重複）

說明：

1. 將語意概念分為那些類別，純粹從文本內容分析所得結果分類，不是以學科知識體系所包含的知識類別分類。
2. 大類、中類與控制詞彙類別將隨著文本的擴充而調整，上層所需要的調整機率比下層小。

語意概念分類

對資訊檢索質量與使用需求的影響

一、對資訊檢索質量的影響

1. 提高全文檢索的完整性與正確性
2. 使資料的搜尋從索引式變為關聯式

二、對文學研究與教學的影響

1. 提供詩詞韻文教學及習作使用
2. 拓展的文學研究的議題
3. 提供詩詞語言演變的研究
4. 電腦運算功能提供科學的數據作為文學研究的立論基礎

語意概念標記應用於全文檢索之驗證： 以唐宋詞網站為例

語意概念分為總類、大類、中類、控制詞彙四層。

往上層擴大搜尋範圍。

往下層縮小搜尋範圍。

同一層為同一知識領域範圍。（物→生物→動物、植物、神物）

語意概念檢索

有布林功能的欄位可使用「&」=AND、「/」=OR、「-」=NOT

作者：	<input type="text"/>	(含別名字號/布林)
朝代：	請選擇-	
詞牌：	<input type="text"/>	
詞題、序：	<input type="text"/>	(含詞序)
詞文：	<input type="text"/>	(字形/布林檢索)
詞句語意 類別：	總類：請選擇	(測試中)
	大類：請選擇	
	次類：人事	
	元素：時地物其他	
宮調：	請選擇-	
目前只有蘇軾 (1036~1101) 詞含有作品時空標記資料 (1060~1101)		
註解：	<input type="text"/>	
寫作時間：	<input type="text"/>	(紀元年ex.宋仁宗嘉祐五年)
寫作時間：	<input type="text"/>	(西元年ex.1060)
寫作年齡：	<input type="text"/>	
寫及地點：	<input type="text"/>	
寫作地點：	<input type="text"/>	
<input checked="" type="radio"/> 字形檢索		
<input type="button" value="開始檢索"/> <input type="button" value="取消"/>		

以唐宋詞網站為例

語意概念標記應用於全文檢索之驗證：

語意概念標記應用於全文檢索之驗證： 以唐宋詞網站為例（物→天文→星）

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

← 上一頁 → 搜尋 我的最愛

網址(D) http://cls.hs.yzu.edu.tw/CSP/W_DB/index.htm 移至 連結 >>

語意概念檢索

有布林功能的欄位可使用「&」=AND、「/」=OR、「-」=NOT

作者：	<input type="text"/>	(含別名/字號/布林)
朝代：	請選擇-	
詞牌：	<input type="text"/>	
詞題、序：	<input type="text"/>	(含詞序)
詞文：	<input type="text"/>	(字形/布林檢索)
詞句語意類別：	總類：物 大類：天文 次類：星 元素：請選擇	(測試中)
宮調：	請選擇-	
註解：	<input type="text"/>	

目前只有蘇軾 (空標記資料 (1060~1101))

三星
三臺
女宿
太白
太白、歲星、辰星、熒惑
太微
太歲
太微

完成 網際網路

以語意概念檢索「星」結果：2,076筆

—提高完整性

http://cls.hs.yzu.edu.tw/CSP/W_DB/index.htm

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H) 連結 >>

【淺顯低唱】若有知音見採—唐宋詞全文資料庫

共計2076筆

... 41 42 43 44 45 46 47 48 49 50 ...

序號	詞牌	作者	朝代	詞題(序)	宮調	詞文	詞中辭彙	辭彙總類	辭彙大類	辭彙次類	辭彙元素
401	阮郎歸	王之道	宋代	二首之一：和廣濟王宰		玉繩低轉斗闌干。	玉繩	物	天文	星	玉繩
402	醉蓬萊	王之道	宋代	代人上高御帶，時在太學		長庚誕貴。	長庚	物	天文	星	長庚
403	好事近	王之道	宋代	何希淵生日		曾見少微初降，	少微	物	天文	星	少微
404	好事近	王之道	宋代	董令升生日		歸／去騎箕尾。	箕尾	物	天文	星	尾星、箕星
405	臨江仙	王之道	宋代	和劉南伯		疏星明耿耿，	疏星	物	天文	星	星

以字形比對檢索「星」：1,427筆

http://cls.hs.yzu.edu.tw/CSP/W_DB/index.htm

Live Search

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H) 連結 >>

【淺顯低唱】若有知音見採—唐宋詞全文資料庫

工具(O) >>

共計1427筆

... 31 32 33 34 35 36 37 38 39 40 ...

序號	詞牌	作者	朝代	詞題(序)	宮調	詞文
351	浣溪沙	曾覿	宋代	櫻桃		滿林璀璨綴繁星。
352	好事近	王之望	宋代	和榮太監		佳句麗星日。
353	減字木蘭花	王之望	宋代	恭人生日		與爾尊前作壽星。
354	醉花陰	王之望	宋代	生日		休畫老人星，
355	風流子	王之望	宋代	范覺民生日		中有麗天星斗，
356	錦堂春	葛立方	宋代	正旦作		星翁說道宜官。
357	滿庭芳	葛立方	宋代	五姪將赴當塗，自金壇來別		別來幾換星霜。
358	水調歌頭	王識	宋代	觀星		初出極星未遠，
359	水調歌頭	毛并	宋代	上元郡集		十里寒星相照，
360	漁家傲引	洪适	宋代	伏以黃童白叟，皆是煙波之釣徒；青笠綠蓑，不識衣冠之盛事。長浮家而醉月，更輟棹以吟風。樂哉生涯，翻在樂府。相煩女伴，漁父分行。		碧落雲高星爛爛。

... 31 32 33 34 35 36 37 38 39 40 ...

58

語意概念標記應用於全文檢索之驗證： 以唐宋詞網站為例（物→天文→月）

http://cls.hs.yzu.edu.tw/CSP/W_DB/index.htm

Live Search

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H) 連結 >>

【淺甄低唱】若有知音見採一唐宋詞全文資料庫

Home RSS Print 網頁(P) 工具(O) >>

語意概念檢索

有布林功能的欄位可使用「&」=AND、「/」=OR、「-」=NOT

作者：	<input type="text"/>	(含別名字號/布林)
朝代：	請選擇-- <input type="button" value="v"/>	
詞牌：	<input type="text"/>	
詞題、序：	<input type="text"/>	(含詞序)
詞文：	<input type="text"/>	(字形/布林檢索)
詞句語意 類別：	總類： <input type="button" value="v"/> 物	(測試中)
	大類： <input type="button" value="v"/> 天文	
	次類： <input type="button" value="v"/> 請選擇	
	元素： <input type="button" value="v"/> 請選擇	
宮調：	請選擇-- <input type="button" value="v"/> 日	
目前只有蘇軾		空標記資料 (1060~1101)
註解：	<input type="text"/>	
寫作時間：	<input type="text"/>	(紀元年ex:宋仁宗嘉祐五年)
寫作時間：	<input type="text"/>	(西元年ex:1060)
寫作年齡：	<input type="text"/>	
寫及地點：	<input type="text"/>	
寫作地點：	<input type="text"/>	

合稱 (同類)

合稱 (異類)

星

銀河

以語意概念檢索「月」結果：6,734筆

—提高正確性

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

← 上一頁 → 搜尋 我的最愛 移至 連結 >>

網址(D) http://localhost/CSP/W_DB/index.htm

搜尋網頁...

共計6734筆

... 21 22 23 24 25 26 27 28 29 30 ...

序號	詞牌	作者	朝代	詞題(序)	宮調	詞文	詞中辭彙	辭彙總類	辭彙大類	辭彙次類	辭彙元素
221	漁家傲	歐陽修	宋代			月下風前愁不少。	月下	物	天文	月	月
222	御街行	歐陽修	宋代			落星沈月，	月	物	天文	月	
223	御街行	歐陽修	宋代			落星沈月，	月	物	天文	月	月
224	臨江仙	歐陽修	宋代			待得月華生。	月華	物	天文	月	月華
225	臨江仙	歐陽修	宋代			玉鉤垂下廉旌。	玉鉤	物	天文	月	玉鉤
226	嘉山溪	歐陽修	宋代			三五銀蟾滿。	銀蟾	物	天文	月	銀蟾
227	浣溪沙	歐陽修	宋代			一尊風月爲公留。	月	物	天文	月	月
228	浣溪沙	歐陽修	宋代			燈燼垂花月似霜。	月	物	天文	月	月
229	浣溪沙	歐陽修	宋代			薄簾映月兩交光。	映月	物	天文	月	月
230	御帶花	歐陽修	宋代			月淡寒輕，	月淡	物	天文	月	月

... 21 22 23 24 25 26 27 28 29 30 ...

近端內部網路

以字形比對檢索「月」檢索：7,705筆

http://cls.hs.yzu.edu.tw/CSP/W_DB/index.htm

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H) 連結 >>

【淺甄低唱】若有知音見採—唐宋詞全文資料庫

文字檢索

回檢索畫面

共計7705筆

... 11 12 13 14 15 16 17 18 19 20 ...

序號	詞牌	作者	朝代	詞題(序)	宮調	詞文
161	謁金門	晏殊	宋代			風月宛然無異。
162	破陣子	晏殊	宋代			人間好月長圓。
163	浣溪沙	晏殊	宋代			月好謾成孤枕夢，
164	浣溪沙	晏殊	宋代			三月和風滿上林。
165	浣溪沙	晏殊	宋代			鬢禪欲迎眉際月，
166	鵲踏枝	晏殊	宋代			明月不諳離恨苦。
167	紅窗聽	晏殊	宋代			風前月底，
168	採桑子	晏殊	宋代			陽和二月芳菲遍，
169	採桑子	晏殊	宋代			淡月朧明。
170	喜遷鶯	晏殊	宋代			堯冀隨月欲團圓。

... 11 12 13 14 15 16 17 18 19 20 ...

語意概念分類在教學上的應用1/6

提供詩詞教學及習作使用

- 一、解讀詩詞最常遇到的問題是無法理解語意，若以翻譯解決，侷限了讀者想像與感受的空間；以註解輔助，又難以對全部詩詞註解。
- 二、語意概念分類幫助讀者了解語意，進而幫助理解詩詞內容。
- 三、習作時最常遇到的困難是「詞窮」的問題，語意概念分類所建立的同近義詞詞群可以解決「詞窮」的問題。

語意概念分類在教學上的應用2/6

- 今人創作詩詞時，不熟悉詩詞格律所包含之平仄押韻問題，因為平仄押韻都是固定而有規則的，以電腦輔助，可以做得比人好。
- 參見「依韻作詩格律自動檢測索引系統」
(<http://cls.hs.yzu.edu.tw/tang/PoemTone/index.asp>)、
- 「倚聲填詞格律自動檢測索引系統」
(http://cls.hs.yzu.edu.tw/CSP/W_stave/index.htm)
- 功能：自動檢查平仄格律及押韻

語意概念分類在教學上的應用3/6

局限

- 電腦無法理解語意，無法判斷多音字，因此前述作詩填詞系統，可以做出格律完全相符，但可能完全不通的作品。
- 詞窮時沒有輔助工具
- 無法處理對仗問題

語意概念分類在教學上的應用4/6

語意概念可以解決的問題

- 一、詞窮問題：經過語意概念分類的詞群，可以作為選詞的輔助工具。
- 二、以詞彙為單位所建立的讀音資料庫可以解決電腦無法判斷多音字的問題（音隨義轉），因此可以提高判斷平仄的正確率。
- 三、對仗問題：律詩中間二聯及對聯之上下聯必須對仗，語意概念分類及詞彙語音資料庫可以解決對仗問題。

語意概念分類在教學上的應用5/6

語意概念無法解決的問題：情意與情境的感受

- 詩詞以言情、寫景、敘事為主，言情與寫景時常彼此襯托，具有相關性。
- 語意概念分類使電腦具備初步理解詞彙語意的能力，比較容易解決敘事詩的寫作，但是電腦不具有「感發」能力，對於理解情感、情境仍有一段距離。

語意概念分類在教學上的應用6/6

語意概念無法解決的問題：情意與情境的感受

未來可以努力解決的方法與方向

以詩詞的「共現詞彙」分析，嘗試歸納出一些原則，比如寫離別感傷時，常會使用那些詞彙一起烘托出離別的情意與情境，以所歸納的結果作為提供選擇的輔助工具。

語意概念分類在研究上的應用1/3

1. 拓展的文學研究的議題

有些文學研究議題，需要有大量的資料分析才能進行，是以人力所做的傳統研究無法做到的。

例如：詩詞典故用語研究、文學風格的比較研究、文學與地理環境相互影響的研究等等。

語意概念分類在研究上的應用2/3

2. 提供詩詞語言演變的研究

詩詞詞彙的增減，語意的改變（消失、新增等改變）研究。

語言風格的時代環境的關係，與作者個人風格的關係，與作品風格（豪放詞與婉約詞）的關係，與文體的關係等等研究議題。

語意概念分類在研究上的應用3/3

3. 電腦運算功能提供科學的數據作為文學研究的立論基礎

電腦對大量數據的分析、統計、歸納能力，幫助文學進行需涉及大量統計的研究，並且提出科學的數據作為立論的基礎。

例如：羅鳳珠、曹偉政，〈唐宋詞單字領字研究〉，《語言暨語言學》，第九卷第二期，中央研究院語言學研究所，2008年4月出版，第189～220頁。

遭遇的困難與未來的發展1/6

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。

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。（幸好出現在句子中間的單字詞多數是虛詞）

遭遇的困難與未來的發展2/6

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，**無法在單一詞彙單獨標註詞義**。

有些詩詞詞句切分之後，無法顯示意境，如陸游：
「小樓一夜聽春雨，深巷明朝賣杏花。」

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、歸納的方法尋找規則。

遭遇的困難與未來的發展3/6

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。

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，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。

遭遇的困難與未來的發展4/6

宋代哲理詩

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。

遭遇的困難與未來的發展5/6

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，也就難以用語意標記處理。

遭遇的困難與未來的發展6/6

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。

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。

謝謝聆聽

敬請指教