漢字意符知識結構的建立

周亞民　　黃居仁
台灣大學資訊管理研究所　　中央研究院語言學研究所
milschou@hotmail.com　　churen@gate.sinica.edu.tw

摘要－漢字與其它書寫系統最大的差異，在於漢字有很好的表意能力，其表意的能力來自於意符。漢字有很多的意符，但是這些意符所形成的知識結構是什麼，到目前沒有完整的輪廓，本研究的目的即是建立漢字意符的整體知識架構。我們以說文的五百四十部首作為漢字意符的基礎，分析它們作為意符所要表達的概念，並利用 IEEE SUMO 進行分類和對映。本研究發現漢字意符已經能夠表達廣泛的概念，我們所建立的意符架構打破自說文以來以字形與聲韻的漢字分類，以漢字的概念建立新的分類架構，另外，本研究的結果支持衍生詞彙理論(generative lexicon theory)，證明漢字意符具有與衍生詞彙相似的衍生能力。

1. 前言

1.1 研究動機

漢字與其它書寫系統(writing systems)，最大的差異是漢字有很好的表意能力，字形(glyphs)與概念(concepts)的連結強，有時候不一定要知道發音，也可以從字形知道大概的意義，本身就構成一個知識結構，但拼音文字和音節文字與概念的連結，是透過發音對映到意義，字形與概念的連結弱，漢字的表意特性來自於字形結構中的表意符號，而且這些意符的表意能力，經歷數千年後仍保存著相當好。漢字表達的意義都與所使用的意符有概念上的連繫，而大部份漢字的意義，都是意符所表達概念的延申，或是詞義的擴展或縮小，因此，如果掌握意符的知識結構，等於掌握了大部份漢字意義，我們希望建立的是整體的知識架構，而不是各別意符的片斷，因為意符概念彼此是有關係存在的，而非互相獨立，但是漢字意符所表達的概念知識結構是什麼，至今仍然沒有整體的架構，這是我們研究這個問題的主要動機。

1.2 定義

我們將漢字的字形依其功能可以分為意符與聲符，意符表示字形與意義有關連，而聲符則表示字形與語音有關連，如果有的字形同時有兩個功能，則為意符兼聲符，這個看法跳離造字的方法，而只看字形的功能，可以大幅減少六書和三書論的歧義所造成漢字分類上的困擾。文字學家唐蘭認為漢字可以分為象形(形符)、象意(意符)和形聲，形符是藉形表意，即形符是藉由其象形表達意義，而意符是以本身的字義表達意義[3]，事實上，此定義也會有難以分類的問題，例如「木」是象形字表示樹木，當「木」出現在其它漢字之中時，我們可以說它是藉形表意，也可以說是藉字義表達意義，那麼「木」是形符或是意符？裘錫圭也認為區分為形符與意符有這方面的問題[7]，因此，我們並不區分形符與意符，只要其功能與意義連結即是意符(形符)，因為形符與意符都是藉由形式表達概念，故形符就是意符，意符也是形符。

2.研究方法

本研究的第一步要先找出漢字的意符，但漢字字形中有那些意符，不是一個容易回答的問題，因為這個問題如同漢字有多少字一樣難以回答，不過我們不一定要找出所

有的意符才能建立漢字意符的概念，而且漢字雖然可以不斷的創造新字，但是造字時大部份都是使用原有的意符或聲符加以組合，而且這些意符很多都是說文解字的部首，而說文部首本來大部份就是初文，作為創造其它漢字的基礎。在一千九百年前的許慎將說文的 9353 個小篆字分類，並建立五百四十個部首，事實上，這些部首都具備很好的表意功能，也是構成漢字字形的主要意符，每個意符都將相關的漢字連繫起來，建立彼此之間的連結，因此，本研究以說文的五百四十個部首作為漢字的核心意符，分析每個部首的本義，以及部首與從屬字的關係，再利用 IEEE SUMO(Suggested Upper Merged Ontology)建立漢字意符概念的知識結構[14]。

如果要分析漢字結構的表意特性，必須由小篆著手，因為它仍然保留大部份最初的字形結構，比隸或楷體更能夠反應造字的本義，雖然小篆並不是最初的文字，從小篆探求字形的本義，可能會發生錯誤，而必須從甲骨文或金文中找答案，但是這種情形在小篆並不多。現在有些字形由於隸變的關係，看不出它與部首的關係，因此，也要從小篆著手，才能充分的掌握部首表達的概念與正確的解釋。

3.漢字意符知識結構的建立

3.1 意符概念的分析

許慎對於部首的釋義，並不一定是從屬字運用部首所要表達的意義，例如說文：「羊(羊)，祥也，从丷，象頭足尾之形」，依說文的解釋「羊」為「吉祥」，實際上，許慎的釋義是「羊」的假借義，並不是本義，「羊」的本義是哺乳動物羊，「羊」的甲骨文和金文皆象羊頭[5]。如果直接採用許慎的解釋，將「吉祥」作為「羊」統領從屬字的概念，這麼做並不恰當，因為「羊」的從屬字利用這個意符表達的概念不是「吉祥」，而是哺乳動物的「羊」。許慎的釋義既使是本義，當它作為意符時也不一定用本義，例如說文：「  (虫)，一名蝮，博三寸，首大如擘指，象其卧形」，依說文的解釋「虫」為「蛇」的象形字，但從屬字有昆蟲和其它動物，如：「  (蠵)，大龜也」、「  (蠻)，閩蠻也」、「  (蛺)，蛺蝶也」、「  (蠽)，丁蠽也」等都不是用「虫」表達蛇的概念，因此，「虫」的表意概念應該是更廣。古代「虫」和「蟲」不分，禮記皆作「蟲」，爾雅釋「蟲」所屬字，說文皆入「虫」部[6]，而說文：「有足謂之蟲，無足謂之豸」， 只要有腳的動物都是「蟲」，又「蟲」就是「虫」，再由「虫」的從屬字分析，「虫」的意符概念是動物，而不是說文解釋的蛇。

上述我們討論的問題，說文對部首的解釋並沒有不對，有時候就已經無法直接將其釋義作為意符概念，說文中的不當解釋，更無法作為意符概念，例如說文：「  (有)，不宜有也，春秋傳曰日月有食之，从月，又聲」，說文認為「有」是不宜有或不當有而有，「有」的從屬字只有兩字：「  」和「  」，說文：「  ，有文章也，从有彧聲」，「  ，兼有也，从有，龍聲」，如果「有」指不宜有，有文章（即文采）為不宜有文采，兼有為不應兼有，段玉裁認為許慎對「有」的解釋，出自於春秋的用例，但章季濤認為多數的情況「有」並沒有這種特殊含義，例如詩經：「女子有行」，「有行指出嫁，出嫁為常有應用之事，解釋為不應有不直有是不對的[11]，李孝定則認為「有」應從「手」從「肉」，而非從「月」[5]。因此，「有」我們將其意符概念修正為有，而非不宜有。

由這些分析可以知道為什麼不能直接以說文對部首的解釋作為表意的概念，而必須先找出部首較為可信的解釋，並且還要分析從屬字中所含部首所要表達的概念，互相比較分析後，才能決定部首作為意符時所要表達的概念。我們盡力作好這些研究工作，但是仍然可能無法分析出來部首作為意符所要表達的概念，這種情況我們直接採用說文對該部首的釋義，作為該部首的表意概念。

說文五百四十部並不是每個部首能生產力都相同，其中有三十六個部首為空立的部首，大部份都是天干地支和數字，各部首的生產力也不同，只有一個從屬字的部首有 153

字，只有二個從屬字的部首有 106 字，如果將二個從屬字以下的部首累計，即 540 部首中有一半的部首只有兩個以下的從屬字，相反的，從屬字最多的前二十個部首，依序是「(水)」、「(艸)」、「(木)」、「(手)」、「(心)」、「(言)」、「(糸)」、「(人)」、「(女)」、「(金)」、「(邑)」、「(口)」、「(虫)」、「(竹)」、「(肉)」、「(土)」、「(玉)」、「(辵)」、「(衣)」、「(馬)」，這些高頻的部首若計算從屬字累計有 4425 字(不含重文)，已佔說文 9353 字(不含重文)的一半。

當我們建立意符的知識結構時，空立的部首由於沒有從屬字，是否也應該放入意符的知識結構？我們認為還是必須納入，因為除了部首「(凵)」和「(㒳)」沒有出現在說文，漢語大字典 54678 字中也找不到使用「(凵)」和「(㒳)」作為意符的字，其它空立部首雖然在說文沒有從屬字，但是的確有漢字用它們作為意符，如：「(四)」出現在「牭 (四歲的牛) 」和「駟 (四匹馬或四馬的車)」,「三」出現在「叁」和「弎 (表示數字三)」，漢語大字典收「仨」字，表示三個(後面不需接量詞)[10] ，「(燕)」出現在「 (燕子) 」，部首「萈(萈)」出現在「 (細角的羊) 」等，因此對於說文中的空立部首的概念，我們還是加入漢字知識本體中。

表一、 說文的空立部首

說文部首(小篆)	說文部首(楷書)
[illegible]	萈丁丙久甲才七六五乇四㸚克彔凵亥三燕𡿨寅卯它戌㞷能丏冄㒳易幵未凵庚壬癸率

有些部首雖然不是空立部首，但是找不到部首與從屬字的關係，或是沒有字義字形的解說(領頭字下釋義「闕」)，由於沒有證據可以說明部首和從屬字關係，因此，我們暫時直接以說文的解釋作為表意的概念，例如說文：「(白)，此亦自也，省自者，詞言之气从鼻出，與口相助也」，依說文的解釋，「」為鼻子，其從屬字有「(皆)，俱詞也，从比从白」、「(魯)，鈍詞也，从白羨省聲」、「(者)，別事詞也」、「 ，詞也，从白𠷎聲𠷎與疇同」、「(智)，識詞也，从白从亏从知」、「(百)，十十也，从一白數十百為一貫相章也」，若分析從屬字，無法得知為什麼「」在這些字當中表達什麼概念，故我們將「」的初義作為其意符概念。事實上，這是因為許慎不當分類，「皆」、「魯」、「者」、「 」、「智」此五字皆應从(口)，後又增繁為(甘)，此字應从(白)[1]，可能因為字形相似，因此許慎將這些字做為從屬字。

3.2 意符概念與 SUMO 的連結

我們找出五百四十部首與從屬字的關係後所得到的意符概念，再利用 SUMO 進行分類，藉由 SUMO 將意符歸類，這麼作不僅可以找出特定概念的相關意符，更重要的是了解意符的概念與其它知識概念的關係，還可以了解意符概念與當代科學知識的關係。我們允許意符概念的分類對映到兩個以上的 SUMO 概念，而非只能對映到一個 SUMO 概念，另外由於 SUMO 為階層式的知識結構，階層式結構越接近終端節點分類越細，可以反應出越細微的差異，相反的，上層的節點分類較粗，無法反應細微的差異，所以，我們在分類時盡可能的放到終端節點，如果無法找到適當的節點，則向上尋找是否有適當的節點，直到找到適當的節點， 為了區分 SUMO 概念與意符概念的差異，進行分類時我們考慮兩種不同的分類情形：

(1) 意符概念與 SUMO 概念為同義

例如說文：「(死)，澌也，人所離也，从歺从人」,「死」的初義依說文的解釋為死亡，可以在 SUMO 的實體/物質/歷程/內在改變/生物歷程/生理歷程/有機歷程/死亡，找到同義的概念。又如說文：「(骨)，肉之覈也，从冎有肉」,「骨」的初義依說文的解釋即為骨骼，可以對映 SUMO 概念的實體/物質/物體/自體連結物/物質/混合物/體物質/組

織/骨骼。這些意符在說文的解釋不僅是本義，也是作為從屬字的意符所表達的概念。

另外一種情形是說文對意符的解釋，不能找到同義的 SUMO 概念，但是意符所表達的概念可以對映同義的 SUMO 概念。如說文：「  (隹)，短尾總名也，象形」，又「  (鳥)，長尾禽總名也，象形，鳥之足似匕，从匕」，依說文的解釋尾羽長的飛禽為「鳥」，尾羽短的飛禽為「隹」，但是如果分析從屬字找出「鳥」和「隹」所表達的概念，並沒有長尾和短尾的差別，如「雉」與「雞」為「隹」的從屬字，但並不是短尾，而「鶴」、「鷺」、「鴻」為「鳥」的從屬字，亦非長尾。再依段玉裁注:「短尾名隹，長尾名鳥，析言則然，渾言則不別也」，故「鳥」與「隹」所表達意符概念皆為鳥之通名，可以對映 SUMO 的實體/物質/物體/行為主體/生物體/動物/脊椎動物/温血脊椎動物/鳥類。

(2) SUMO 節點的概念為廣義

由於 SUMO 為各個領域知識的共同上層結構，其概念分類並不會很細，用來涵蓋各種領域知識本體，如果分類太細就無法作為各種知識本體的上層知識，因此，可以預期的是 SUMO 節點大部分皆為其它概念的廣義概念。例如：說文：「  (鬲)，鼎屬，實五穀，斗二升曰穀，象腹交文，三足」，依漢語大字典的解釋，「鬲」為古代炊具，有陶製與金屬製兩種，圓口，三足[10] 。再進一步分析「鬲」的從屬字：「  (鬻)，釜屬，从鬲，  聲」、「(鬳)，鼎屬，从鬲，虍聲」等，可以推論「鬲」的意符概念就是炊具，我們將它對映到 SUMO 的兩個概念：實體/物質/物體/自體連結物/微粒子物體/人造物/裝置和實體/物質/歷程/內在改變/產生/製作/烹煮，這兩個 SUMO 概念與炊具並不同義，而是廣義概念。

每個部首的初義本類可以同時歸到二個或以上的 SUMO 節點，如說文：「  (耑)，物初生之題也」，我們同時對應到未完全形成的和植物，又如說文：「  (且)，薦也，从几，足有二橫，一其下地也」，則同時對應到裝置(device)和宗教歷程，因為依許慎的解釋，「且」是祭神的禮器，而祭祀在 SUMO 屬宗教歷程，禮器則屬 SUMO 人造物下的裝置，但依李孝定和蔡信發等人的看法，「且」指的是神主的牌位，後來才加上示[5][8]，也可能是生殖器[1]，不過如前所述，我們主要還是許慎的看法為主。

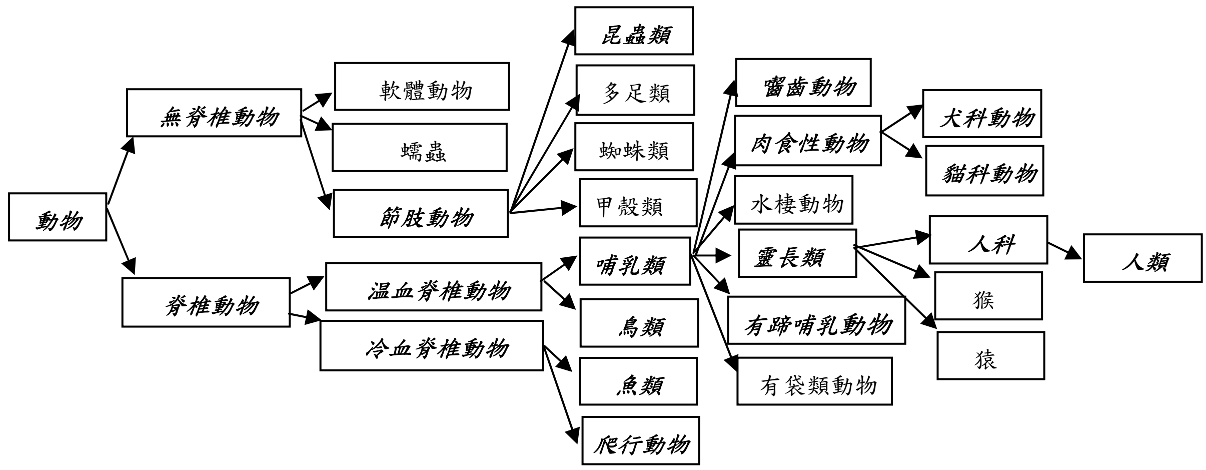
對映到多個 SUMO 概念的原因之一是由於古漢語單字詞較多，一個漢字所表達的概念較為豐富，如「  (駣)，馬飽也」指馬肥壯的樣子，「  (耄)，年八十曰耄」指八十歲的人、「  (豯)，生三月豚腹豯豯皃」指三個月的小豬，「  (驃)，黃馬發白色，一曰白髦尾也」指黃色有白斑或黃身白鬃尾的馬，這些字都有很豐富的意義，而現代漢語則多使用雙字詞和多字詞表示，如此豐富的意義，包含了很多的概念，只使用一個 SUMO 的概念無法表達完整的概念，所以對於這些字，我們將它對映到多個 SUMO 概念。對映到一個以上的 SUMO 概念。另一個原因則是因為部首使用一個以上的概念統領從屬字，如說文：「  (止)，下基也」，依許慎的解釋「止」為腳，但作為意符時則還有行走的概念，因此，「止」的表意概念同時歸類到軀幹部件和行走。

4.研究結果

我們運用 IEEE SUMO 將說文部首重新分類後，可以發現大部份的意符表達的概念都是具體的，與漢字的特性相符，也與其它書寫系統的發展相符。漢字構造理論，無論是六書或三書，都是將文字起於象形。三書的象形象意形聲，以象形最早，由象形發展出象意，再發展出形聲。唐蘭最畫出像事物的形狀，就是象形，如果還有某種意義就是象意，他說人像人形是象形字，尸像人蹲形是象意字，事實上象意字也是象形字，只是象意字可以表達除了象形所指物以外的意義。再從書寫系統的發展來看，埃及文字和蘇美文字也都是起源於象形，只是後來的發展與漢字不同，逐漸走向拼音，而漢字也在文字中加入表音的部份，但是並沒有捨棄表意的意符，而發展出形聲字，且仍然保留意符

直到現在。象形字必定像具體的事物，因此，意符反應在 SUMO，大部份都是具體的概念。

文字學家認為甲骨文已是相當成熟的文字，而小篆較甲骨文更成熟，主要的證據是看形聲字的比例，這個角度是由文字的結構出發，我們認為應該還可以從另一個角度來看文字是否成熟，若某書寫系統發展成熟，應該能夠描述所需傳達的基本概念，換句話說如果大部份的基本概念可以被記錄，則該書寫系統已經成熟。我們從這個角度分析意符所呈現的知識結構，可以發現意符所表達的概念涵蓋了動物、植物、基本物質、心理、疾病、宗教、飲食、交通、生死、數字、時間、交易、衣物、顏色、形狀、量度單位、天體、地理、人造物、人際互動、音樂、人和身體部位、器官、感官、性別等，五百四十個意符就能表達這麼廣泛的概念，而說文部首很多都能找到甲骨文字，所以，從這個角度也可以證明甲骨文非常成熟，已經能夠表達很多概念，而不只是單純的圖畫。由於 SUMO 概念節點共有九百多個，我們以 SUMO 中動物有關的概念，討論有關的意符知識結構，圖一表示的是 SUMO 動物的概念結構，表二是 SUMO 概念與對映的意符概念。漢字意符概念有脊椎動物與無脊椎動物皆有，但以前者較多。無脊椎動物中，漢字意符沒有軟體動物和蠕蟲有關的概念，只有節肢動物，且只有節肢動物的昆蟲類。脊椎動物的子類別温脊椎動和冷血脊椎動物都有從屬的意符概念，但是沒有水棲動物與有袋類動物的從屬意符。



圖一、SUMO 動物概念階層結構與意符概念的對映(黑體字表示從屬的意符)

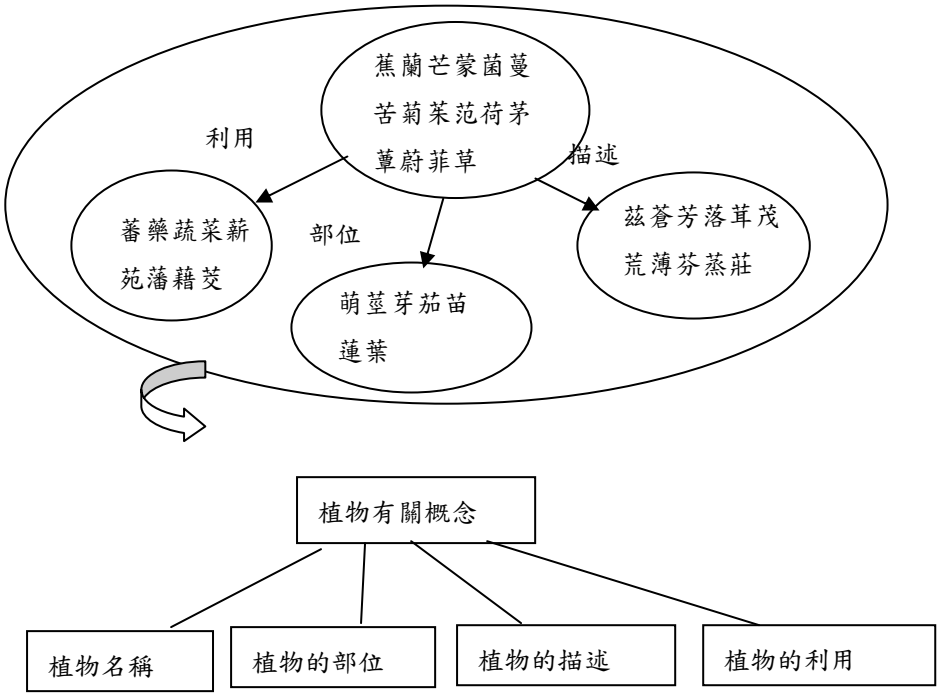
表二、SUMO 動物概念與從屬的漢字意符(底線表示沒有出現在下層概念)

動物	人儿尸夫了馬牛羊羴犛豩豕鹿彑萈舄㣇犬犾虤虎虍鼠兔象能熊瞿鳥烏隹雥萑雔燕乙魚鱻它易巴龜豸禸㲋蚰蟲虫
無脊椎動物	蚰蟲
脊椎動物	人儿尸夫了馬牛羊羴犛豩豕鹿彑萈舄㣇犬犾虤虎虍鼠兔象能熊瞿鳥烏隹雥萑雔燕乙魚鱻它易巴龜豸禸㲋
温血脊椎動物	人儿尸夫了馬牛羊羴犛豩豕鹿彑萈舄㣇犬犾虤虎虍鼠兔象能熊瞿鳥烏隹雥萑雔燕乙
冷血脊椎動物	魚鱻它易巴龜
哺乳類	人儿尸夫了馬牛羊羴犛豩豕鹿彑萈舄㣇犬犾虤虎虍

	鼠兔象能熊冒
鳥類	鳥烏隹雥萑雔燕乙
魚類	魚鱻
昆蟲類	虫蟲
人類	人儿尸夫了
肉食性動物	犬㹜虤虎虍
囓齒動物	鼠兔
有蹄哺乳動物	馬牛羊羴犛豩豕鹿彑萈舄㣇
貓科動物	虤虎虍
犬科動物	犬㹜
爬行動物	它易巴龜

這些出現在 SUMO 結構中的意符概念，是非常基本的核心概念，因為其它漢字都是由這些概念而擴展，例如由人的概念(SUMO 概念為人類)，擴展到描述人的行為和德性，從手的概念(SUMO 概念為軀體部件)，擴展到描述手的動作。呈現在 SUMO 的意符概念，其生產力並非完全相同，因為有些意符只有幾個從屬字甚至沒有從屬字，有些則有數以百計的從屬字，我們希望能夠反應意符概念在造字時的重要性，因此，將每個意符的從屬字數做為意符概念的權重，從屬字多的意符概念權重較高，反之，如果從屬字少，則意符概念的權重低。意符概念中最有生產力的是植物有關的概念，其次是軀體部件，第三是人類有關的概念。這些意符因為使用頻次較高，一定是生活上有實際的需要所致，例如 SUMO 動物的相關概念中，意符可以直接對映動物概念的包括人類(意符:人)、犬科動物(意符:犬)、鳥類(意符:鳥)、魚類(意符:魚)，沒有直接對映動物概念的包括貓科動物、有蹄哺乳動物、爬行動物、囓齒動物、脊椎動物、哺乳類。動物概念相關的意符中，人以外最有生產力的意符是有蹄哺乳動物，屬於此概念的意符包括：牛、羊、豕、馬、豩、萈、羴、鹿、彑、舄、犛、㣇。為什麼有蹄哺乳動物的意符生產力高，因為馬、羊、牛、豕皆為六畜，在中國都是很早就被眷養的動物，六畜與飲食、宗教儀式、交通、戰爭等都有密切的關係，由於這些動物被利用的很頻繁，自然有描述它們的需要，所以這些意符才有很高的生產力。

我們還發現意符與從屬字所形成的知識結構，是一個高度衍生力的知識體系，以生產力最高的意符之一的艸為例，意符艸的衍生字和衍生概念有植物的專名、描述植物的部位、描述植物的屬性和外觀，以及描述植物的功用(圖二)，這個知識體系符合由 Pustejovsky 提出的新語言學理論－衍生詞彙理論，此理論解釋了詞彙衍生的環境－經驗結構(qualia structure)，此結構分為四個面向：物質(formal)、組成(constitutive)、功用(telic)、產生(agentive)，由意符艸的衍生概念來看，正好符合物質(植物)、組成(植物的部份)、功用(植物的利用)，而植物非人造物，所以沒有「產生」此面向的衍生概念．由研究結果來看，我們所建立的意符概念結構可以證明意符具有與衍生詞彙相似的衍生能力[12][13]



圖二、意符艸的衍生字與衍生概念

5.結論

本論文所作的研究工作，事實上是對漢字意符概念進行分類，我們從這個角度出發與之前的文字分類加以比較，過去的字書對文字的分類，除了以字音加以分類外，都是以說文部首作為基礎，唐蘭認為說文部首的分類法，不能看出文字的發生和演變，又不能藉以作同類文字的比較研究，也不能給一般人檢查的便利，因此提出自然分類法，將甲骨文字依次歸類，打破長期以來使用說文部首的架構[2][3][4]。唐蘭與我們所建立的漢字知識本體共同之處是字義進行分類整個自然分類法的分類架構，可以分為兩層，第一層共有四大類，第二層為二百三十一部(部首)。四大類依唐蘭的定義分別為：

(1)象人：即鄭樵所謂人物之形，易經：「近取諸身」。
(2)象物：凡自然界的一切所能畫出的象形字。
(3)象工：一切人類文明所製成的器物。
(4)象事：凡是抽象的形態和數目等屬之。

第二層的分部方式也是利用部首的概念，將三千多個甲骨文分為二百三十一部，其中大部份都是說文的部首，由於第二層的分類方式與過去字書的分類方式沒有太大的差異，我們只將討論的重點放在第一層的分類方式。唐蘭的分類方法，可以視為將說文部首打散後重新依象人、象物、象工和象事加以歸類，這點與我們的做法是共通的，但是我們分類的依據是漢字透過意符所表達的概念，再依 SUMO 的結構加以組織。本研究與唐蘭的分類相比較，有下列的優點：

(1)分類的互斥性較佳

好的分類架構應該要滿足互斥性與完整性，如果類別定義沒有重疊或模稜兩可即滿足互斥性，完整性則要求任何待分類的對象，都可以分到其中一類，不會找不到適當的類別加以歸類。自然分類的象人與象物並不滿足互斥性，依唐蘭對象物的定義為凡自然界的一切所能畫出的象形字，人也是自然界的一部份，也能被畫出象形字，因此與象人產生重疊。

(2)分類較為詳細

唐蘭的分類只有四大類 ，而 SUMO 則有九百多個類別，可以作更細微的區別，例如植物、動物、河川、山石等概念的部首，依自然分類都被歸類到象物，無法區分開來，但是，如果用 SUMO 加以分類，將會歸類到不同的類別，可以將它們區別開來以增加鑑別率。

(3)知識的分享力較佳

如果要增加知識的分享力，必須有一個大家接受的知識描述架構，所有的知識都利用這個架構為基礎，對於計算機而言，更需要共同的知識結構才能讓不同的計算機共享資源，目前 IEEE SUMO 即扮演這個角色，許多領域知識都開始與它連接，使不同的領域知識得以共享，而自然分類的四種類別並不是共同的知識描述架構，我們以 SUMO 作為知識描述的基礎，可以得到較高的知識的分享力。

本研究主要的貢獻是建立漢字新的分類架構，突破了傳統文字學對漢字的分類限制，建立了意符的知識架構，並發現意符具有與衍生詞彙類似的概念衍生能力。未來我們將會建立不同斷代的概念結構，以本研究為基礎，進一步比較不同時間的概念變化。

參考文獻

[1]. 季旭昇，說文新證，台北，藝文印書館，初版，2002。
[2]. 唐蘭，古文字學導論，台北，樂天出版社，初版，1960。
[3]. 唐蘭，中國文字學，香港，太平書局，1963。
[4]. 唐蘭，甲骨文自然分类简编，山西教育出版社，第一版，1999。
[5]. 李孝定，讀說文記，中央研究院歷史語言研究所，初版，1992。
[6]. 徐復、宋文民，說文五百四十部首正解，江蘇古籍出版社，第一版，2003。
[7]. 裘錫圭，文字學概要，臺北，萬卷樓圖書有限公司，3 月初版 ，1994。
[8]. 蔡信發，說文部首類釋，台灣學生書局，2002。
[9]. 許慎，說文解字，徐鉉校定，中華書局影印。
[10]. 徐中舒主編，漢語大字典，台北，建宏出版社 ，1992。
[11]. 章季濤，怎樣學習說文解字，群玉堂出版公司，初版，1991。
[12]. 黃居仁，漢字知識表達的幾個層面：字、詞、與詞義關係概論，漢字與全球化國際學術研討會，台北，2005。
[13]. Pustejovsky, J. The Generative Lexicon, The MIT Press, 1995.
[14]. Niles, I., and Pease, A. “Towards a Standard Upper Ontology”, Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), Ogunquit, Maine, October 17-19, 2001.