

31

A)

從人工智慧角度談
中國文字蘊含的知識結構

謝清俊

中央研究院 資訊科學研究所 研究員
兼 中央研究院 計算中心 主任

中華民國 七十六年 九月 卅日

機器中的知識

目前，機器中的知識均由制式系統(formal system)產生。這種知識比知識論或認知科學中的要狹義得多。機器中的知識在於「能行」。換言之，凡能通過塗林測試(Turing Test)者 [Hofstadter, 1980]，皆稱具有人工知識或智慧。

機器究竟能知道些什麼？依Barr和Feigenbaum [1981]的說法，大約可分為四類型：

- 物(objects)
- 事(events)
- 為用之知(performance)
- 形上之知(meta-knowledge)

此四類型中，前兩者是對自然世界和人文社會的觀察與體認。「物」指對實體真像的體認，除本身外，還包括其分類，及物性之描述等。「事」指對行為或事件的了解，除本身外，還涉及時序，以及因果關係等。「為用之知」是指如何做事的知識，如造句、解題等所需的知識。「形上之知」是指在既有知識之外，相關的知識；在機械化的知識系統中常指：關於「對既有知識的了解」的知識，如對事物(既有知識)的評估。它常用在系統的控制部份。

根據以上的界定，機器擁有知識智慧並非不可能之事，真正的問題在於：對較複雜的知識系統，能不能在機器裡有效地表達(knowledge representation)。

此所謂的表達，有兩個要件：

- 是否能設計一種資料結構(data structure)，蘊含全部要處理的知識？
- 是否能構思一個程序(interpretive procedure)，可以有效的把想要得到的知識，從上述的資料結構中導出來？

目前，所有人工智慧的研究，首先遭遇到的，也是最困難的部份，就是知識表達問題。時下常見基本的知識表達方法有 [Barr 和 Feigenbaum Vol.1, PP.141-222, 1981]：

- 邏輯(logic)
- 程序(procedural representation)
- 語意網路(semantic network)
- 產生系統(production systems)
- 類比(analogical representation)
- 劇本框(script and frame)，等。

在實用的系統中，常不只用一種表達方法，而是綜合數種表達方法以應所需。

文字的產生與孳乳

古時候，「文」和「字」有不同的定義。東漢許慎〈說文解字敘〉中云：「倉頡之初作書，蓋依類象形，故謂之文；其後形聲相益，即謂之字」。由此可知，所謂「文」是指漢字在最初造字時，所產生的一群基本符號，它是漢字最基本的元素(primitives)，有「字母」的功能，可用以組成許許多多的「字」。

「依類象形」的方式有兩種：一是象「具體」之形，謂之「象形」，一是象「抽象」之形，謂之「指事」。換言之，六書中「象形」和「指事」是造「文」的法則。當「文」緣事物構形之後，再以「音」「義」相附，便完整地擁有文字「形、音、義」的要件。漢字中「文」的數目不多，根據〈說文解字〉記載的只有 489個而已。這些造字之初的文，我們有時特別名之曰「初文」以資識別。

到了宋朝，鄭樵對「文」另下了一個定義，他說：「獨體為文，合體為字」。這個定義的精神和許慎的並無不同，都是持「字」由「文」生的觀點，然而做法和結果上卻截然不同。根據相關資料對〈象類書〉之描述(該書已失傳)，鄭樵定義了 870個以聲為主的「文」和 330個以形為主的「文」，合計共1200個，是為所有「字」孳生的基礎。顯然，鄭樵的重點是放在字的孳乳上，才得到這樣個結果。當然，文字在字體演變過程中，經歷筆劃的嬗變，可能是導致此二者不同的原因之一。可惜已無法從原來的「文」印證了。

根據許慎的說法：字是「形聲相益」產生的，其實就是六書中的「形聲」和「會意」二者。會意是形與形的相益，形聲是形與聲的組合。六書有「四體二用」之說：「象形」、「指事」、「會意」、和「形聲」是造文字的法則，是體；而「轉注」和「假借」是文字運用的法則，是用。「轉注」是指互訓的字，約略可視為一群近義字之間，其義「相轉相為」的關係。「假借」是指一字數用的方法，約略可視為一個字兼具數音時，其「依聲託事」而引申其義的關係。對於一些不認識的文字，可以「望文生意」是有道理的，因為中文構字之中，已經蘊含了許多知識的線索，而字與字之間也頗有形、音、義交互的關連供人尋思。關於文字孳乳的過程，鄭樵在〈六書略〉序中云：「象形為本；形不可象，則屬諸事(指事)；事不可知，則屬諸意(會意和形聲)；意不可會，則屬諸聲(形聲、轉注、和假借)，聲則無不諧矣」，這段話把文字孳乳綿延寫得十分明晰。

為了彌補古書失傳之憾，並因應中文資訊處理的需要，自民國六十年以來，今人亦花費頗多心血在漢字構字的問題上。這些研究可分為兩方面：其一是為解決漢字檢索和字形產生的問題，純由目前的文字外觀著手，分析字形的構成；其二仍是由文字學著手，探求漢字的字根。

前者的研究，最早見於交大發表的 496 個字根 [謝清俊等，1972]。這組字根和一個制式的產生系統 (formal production system) 可以產生中文大辭典 49905 字中的 48713 字。此後，這種做法引起共鳴，許多不同的「字根集合」陸續出現 (雖然所用之名稱並不一致)。這些理論上的探討，使有今日各種輸入方法、各種字形產生器的實現。這方面的努力，亦引導著以標準鍵盤取代大鍵盤的發展歷程。

從文字學方面著手的，有周何 [1984] 等。周教授依據〈中文資訊交換碼〉第一、二集中的二萬二千餘字，歸納出文字孳乳的字根 1134 個；其中聲母 869 個，形母 265 個。這個數字和鄭樵〈象類書〉中記載的相去頗近。

以上的討論說明了一件很重要的事情：漢字的數目雖然很多 (目前已蒐集到約八萬個，尚不包括書法之變異)，而且字數與日俱增 (漢字之集合是開放的，open set)，然而漢字字形構成的基本元素 (無論名之為「初文」、「文」、或「字根」)，是有限且恆定的封閉集合 (closed set)。這個現象對中文資訊處理技術的發展有極其深遠的影響。

至此，我們已經瀏覽了人工智慧系統中的知識，以及漢字構字和孳乳的法則。中文構字中所蘊含的知識和人工智慧技術究竟有什麼關係呢？在下文中將以一些題例說明。

依類象形

象形和指事所造的「文」，多緣事物造形；用簡單的幾個筆劃，將事物的特徵轉變為文字圖繪。這個過程包括人工智慧中常做的「擷取特徵」(feature abstraction)，及表達知識 (knowledge representation) 等工作，而這些工作，都必需基於對自然事物深入正確的了解。換而言之，沒有正確的知識和方法，是很難做到自古至今人人都覺得這些「文」能「心同此理」的。(請參考表一表二之舉例)。

然而，這些「文」的原形是圖繪，如果要一個程式能像人一樣的「了解」他們的話，無可避免的，這個程式要有高明的圖案處理和影像處理技術。把這些「文」的圖繪直接存入計算機中，便是前文所謂的類比表達法。由表一表二中的例子可知，以類比的方式表達字形時，若要擷取其中蘊含的知識，則程式必需要能夠：判斷輪廓，如之於「日月心貝」等字；認別特殊圖案，如「木、日、大、一」等圖形均曾出現其他字中；認別特殊的方向，如之於「人、巨、日」等；認識位置，如「夾末本上下」等字；認識字的一部份，如「片夕」等字。凡此種種，皆涉及相當複雜的圖案及影像處理技術。一般而言，這是頗耗

計算資源，也是頗不方便的。然而對一些變例，如增體的「文」，則依然可以用符號的方式處理。譬如：旦、雲，可由(日、一)，(雨、云)來分析。這是由於增體的「文」已含有會意的成份在內了。

象形字可分為依天象、地理、動物、植物、衣服、宮室、器物等所繪製者。每一個象形字都蘊藏著對該事物了解的知識，而這些知識不只用來認識此字，更用以與其他字區別，以及孳生新字之用。試觀下列：

【例一】象形字指事字與其孳乳字之例

木：此為木之原形；上像枝，下像根，丨像幹。

若明白以上的像徵意義，立刻可導出下列之知識：

- (1) 木有幹。
- (2) 幹上有枝。(枝上有葉)
- (3) 幹下有根。


而這些知識可用語意網路裡常用的 part-of 的關係，在計算機中表達中。其實，先人對木之了解不止於此。試看一些「木」孳生的字：

- (1) 本：此為木之基部，木賴以支持，並由此生長。(見表二)。
- (2) 末：此為木之末梢，是細節，是末端；(見表二)。
- (3) 果：木實也，田為果形，木可結果。
- (4) 東：日初出時(日光自木透出)之方向。
- (5) 杲：明也，日升至木上則天色大明。
- (6) 杳：冥也，日全落則在木下。
- (7) 林、森：木可叢生。


事實上，不止木利用相關的知識綿延而孳生新字，每一個漢字都可如此。

指事字所蘊藏的知識，可比照象形字的方式表達。所不同者為：指事字已開始表示抽象的概念，而象形字多表實體；是故象形字多為體詞，而指示字可為動詞，如「立、爪、杲」；可為形容詞或副詞，如「上、下、本、末」等。指事字所表達的概念 (concept) 和語意基 (semantic primitives)，或是與概念依屬 (conceptual dependency) 所設定的「基本概念集」之間，有什麼關係？漢字構成的基本概念有那些？這些基本概念與造字、孳乳有什麼關係？這些都是十分有趣，且值得研究問題。

表一 象形字舉例

古形	今形	造字之意
	<p>日 月 雨 心 貝 木 云 雲 片 兒 兒 漆</p>	<p>○像輪廓，一像其中不虧。 ㄩ像輪廓，常闕，中有陰影。 一像天，丨有貫穿之意，水自天雲下也。 像心之形。 ⊖像外殼，八像觸足。 上像枝，下像根，丨像幹。 山川氣升成云，雲之初文。 雨與云關係密切，故云加雨，此為增體像形 判木也，剖開的木頭，為省體像形。 ⊕像頭，骨未合之形，乃為人之奇字。 ⊖像面，人之面貌，貌之初文。 木汁也，漆之初文。</p>


表二 指事字舉例

古形	今形	造字之意
	<p>上 下 旦 叵 爪 𦏧 立 夾 本 末 𠂇 夕</p>	<p>居上也，高也。 居下也，底也。 像液體凝凍之形。 日出於地面，明也，亦晨也。 「可」之反形，意為不可，此為變體指事字。 覆手曰爪，抓之初文。 交積材也，象對交之形，構之初文。 人立於地上。 人腋下藏物，表竊物之意。 木下曰本 木上曰末。 倒首(𠂇)也，懸之初文。此為變體指事字。 月半見(之時)。</p>

形聲相益

如果所有的「文」都可以用符號表達在計算機裡，處理「字」時就方便多了；至少可以省掉處理類比資料的複雜技術。在這一節裡，將舉例討論形聲字、會意字、轉注字、假借字的構字與知識表達問題。

【例二】會意字所蘊含的知識，多可直接的以語意網路表達。

「」：會意字，今之「愛」字。

圖繪之意：(一隻)手持(抓)心給(另一隻)手(之狀態)。

知識表達：①圖繪之意可用語網路表示，如圖一〔王申培，1987〕。

②手是人體的一部份，又「給」是人才能做主的行為，不是手；是故在此手代表人，而圖繪之意可以解釋為：人將心給他人(之狀態)。


③由②，愛是人與人之間的事。

④將心給人是不合理之事，所以它是指情。若是指情，則主詞(人)與受詞(人)並不重要，要點是在這件事。所以可將愛之語意解釋為：付出(給)心(之感情)。

字之語意：如付出心一般的感情。

有些會意字依然因襲圖繪的特色：以字根之間相對的位置來表示一些訊息。譬如表三中的：益、安、黑、聳、寒等字。這些字都是一幅圖畫，而圖中所繪的都是在特定時空下，某一個狀態的「快照」。為了要使人人懂，且都能心同此理的認識這些字、用這些字，圖繪所表達的「快照」就必須要通俗、客觀，據實，而且容易使每一個人的第一感，就會聯想到該字的語意。(談到聯想，以接近律、類似律、因果律、反對律來核對會意字造字的想法，應該是件有趣的事)。就觀察會意字所得，語意可就圖繪之原意往事物、情、理、時間、空間等不同方向的延申。讓我們再看一個語意延申的例子。

【例三】古今字義延申之例。

「」。會意字，今之「莫」或「暮」字。

圖繪之原意：日(落)至(地面)草隙中(之狀態)。

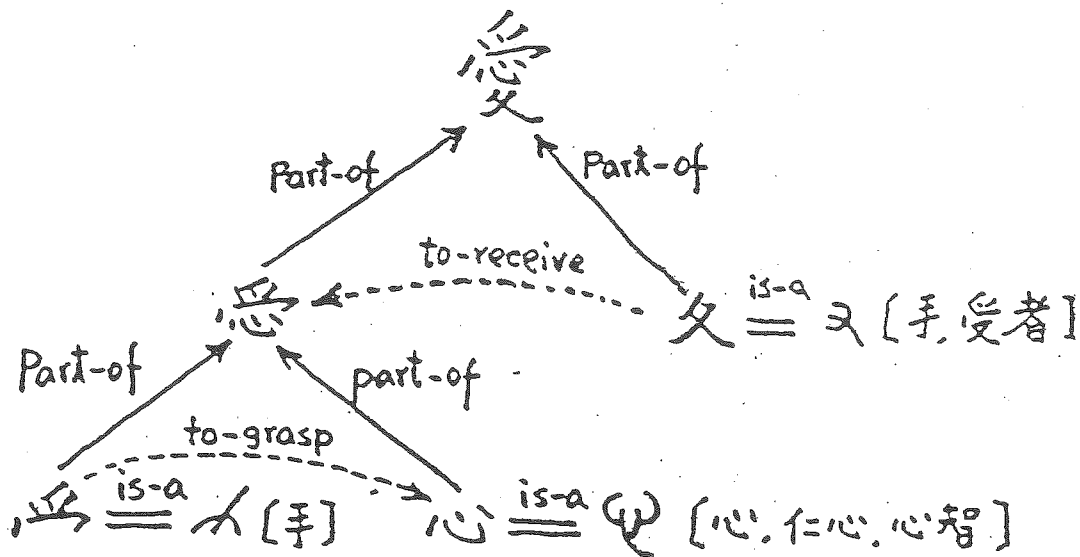
表達之概念：日落(之時)。

語意延申：請參考圖二。

表三 會意字舉例

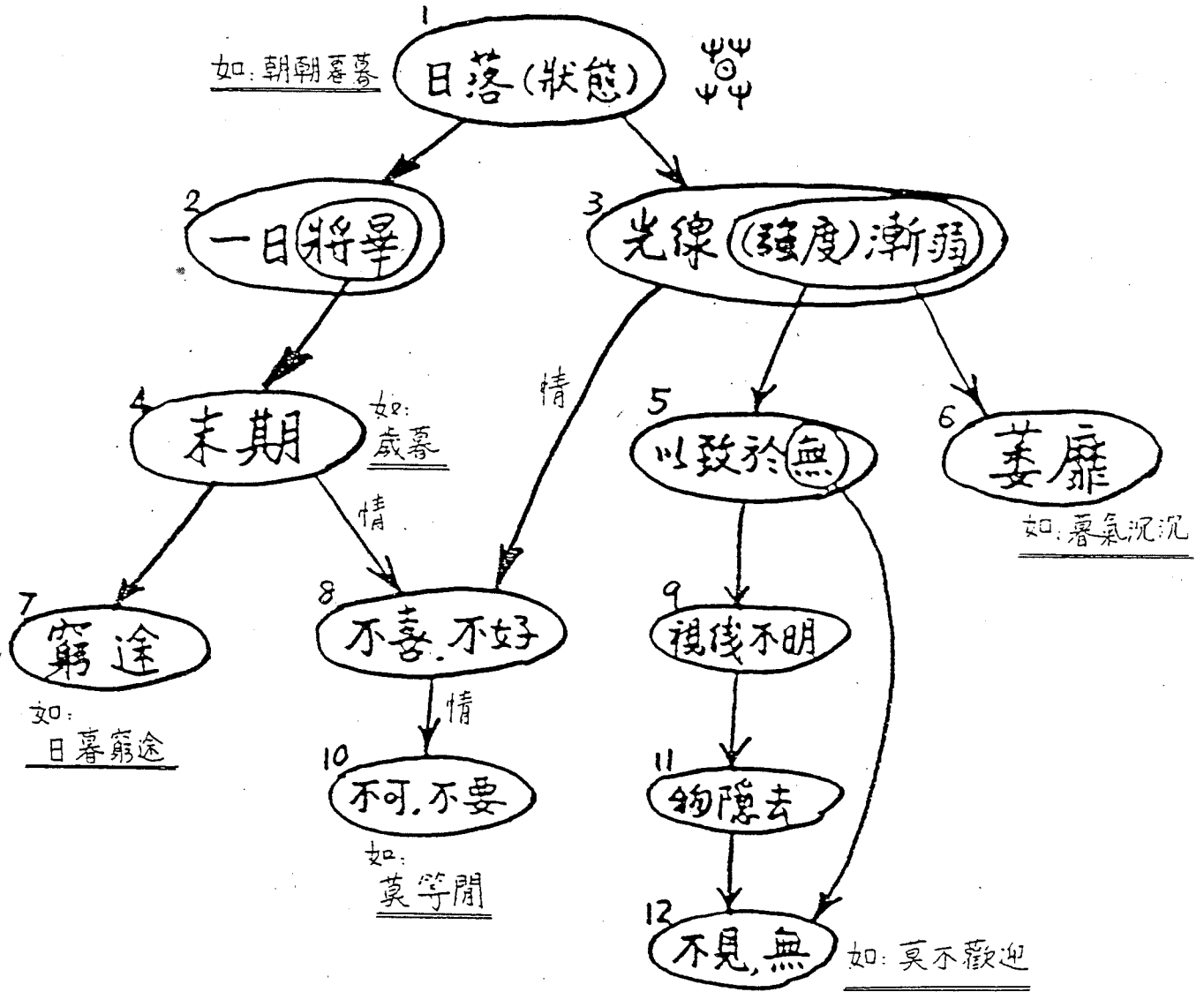
字	形	造字之意
祭	祭 古史	以手(右上又為𠄎)持肉(𠄎在左上)以享神祇(示, 在下方)。十口相傳, 識前言者也。
史	史 吏 惠 臭 好	𠄎從手(下方)持中(上方), 記事者也; 記事官必須中正。治人者也, 從一從史; 吏必須心主於一, 明法令掌故(從史)正直的心為惠, 此為古德字; 彳表行為, 德為今日之德字。自, 古鼻字, 犬, 敏於嗅。臭為古嗅字。
益	益 安	子為男子美稱, 男女結合成家, 好事也。
冰	冰 黑	水(橫)浮出皿上, 溢之古字。 女在家中, 平安狀也。
鬯	鬯 寒 漏	水之凝凍者, 冫見表一。 囟(古字為囟)在火(火 ...)上所薰出灰(中間)之顏色。
寒	寒 漏	頁(古之首字; 頭也)在四口之中, 鬯鬧也。
漏	漏 召	人在宀下下覆以草(艹), 再下有冰(冫)。𠄎, 寒也。 雨入屋中。漏之古字。
陷	陷	從人在白上, 陷之古字。卩為自, 層疊之地形也。

圖一 語意網路之例：愛



其中: predicates: {part-of, is-a, to-receive, to-grasp}
 Semantic primitives: {手, 心, 又}

圖二 語意延申之例：莫



圖中: ○ 表示一個概念。

今日字義: 根據國語日報辭典

莫: (1) 不可, 不要, (2) 無, (3) 姓

暮: (1) 日落, (2) 末期, (3) 萎靡, (4) 窮途

由人工智慧角度來看，有關日出日落一天的景像，可以用一個劇本框來表達，而日落正是其中一幕。在日落的幕中，莫、暮或其他詞彙，都可以與劇本連繫上，而構成一較複雜的知識表達體系，請參閱表四。由表可知：漢字可以和知識表達渾然結成一體，而且，此字與其字根也可作為擷取知識時的索引(index)或指標標(pointer)。再者，字本身所蘊含的知識，也可以和劇本框結合，使整個架構的知識內涵更為豐富。

表四 劇本框之例

GENERIC 白天 FRAME

TYPES : RANGE : (晴天, 陰天, 雨天,)

DEFAULT : 晴天

STYLE : RANGE : (週日, 週末, 節氣, 假,)

DEFAULT : 週日

TIME-OF-OPERATION: BEGIN : RANGE : (5AM-6AM)

DEFAULT : 5AM

END : RANGE : (6PM-7PM)

DEFAULT : 6PM

EVENT-SEQUENCE : RANGE : (日出, 晨間, 上午, 中午, 下午, 黃昏, 日落)

DEFAULT : (a function of the time-of-a-day)

劇本：日出

相關詞：(旦、破曉，早，...)

光線強度：事前：暗

當時：黑暗至光明

事後：逐漸明亮

時程：幾分鐘

方向：東

溫度：事前：低

當時：低

事後：逐漸昇高

筍是曲竹做的捕魚具；鈎是彎曲之金屬物；趺是天寒曲足之狀；胸是曲胸狀；羽是曲羽狀；痾是曲脊；耄是因老而僵僵；絢是繩曲狀；枸是車下之曲木；枸是木曲狀； 是曲刀。鑣也；苟是草之曲也；雉是雄雉鳴時曲頸狀；媼是媼也。以上諸字皆從句音，並有曲之意，凡形聲多兼會意，此為一例。

轉注和假借之孳乳，多半與聲有關。僅再舉一與義相關之假借，一與音相關之轉注，作為此節之結束。

【例五】因義借假之例。

年：說文：「穀熟也，從禾，千聲」，穀(在中原)一年一熟，是故其週期為一年，故借為曆中之年。

月：月之圓缺以29日為週期，故借為曆中之月，表29日也。

日：日之運行以一日為週期，故借以計時表示一日。

在例五中日月的知識，可與表一中的相輔相成。這些與字有關的知識，都可以經由文字關係而相互引得，這樣的結構也是我國文字獨有的。

【例六】轉注混同之例。

爾雅釋詁：「初、哉、首、基、肇、祖、元、胎、俶、落、權輿；始也」

初：說文：「裁衣之始」。

哉=才(古文作才)說文：「才：草木之初」。尚書康誥「三月哉生魄」。

首：首先也。

基：說文：「牆始築」。

肇：說文作，「始開」之意。

祖：宗廟之始。

元：說文：「始也」。九家易：「氣之始也」。

胎：人之始也。

俶：動作之始。詩大雅既醉：「令終有俶」。

落：木葉隕墜之始。詩周頌訪落：「訪予落止」。

權輿：天地之始。詩秦風權輿：「胡不承權輿」。

在例六中的各個字，其語意中均含有「始」的成份。這個「始」應可視為一語意基。換言之，這些字其含義的交集便是「始」。類同這樣結構，「明」字的語意亦取自日月共同之特徵。這種關係完全合乎集合(sets)之運算法則，實為漢字有趣之知識表達法。例六亦可作為文字義意演變「混同」的例子。

分化與孳乳

在此節中，將舉例說明一些較為複雜的例子：一個訓詁中字義分化(義訓)之例，一個文字孳乳字系之例，以及一個現代辭典中同部首裡，文字結構與知識表達間的關係。

【例七】義訓，分化之例。

「立」，說文：「從大在一上」指事字。立之初文。

圖繪狀態：人站在(地)平面上。

原字意：「站立」之意。

語意分化：參考圖三。圖三(A)表示訓詁中之解說，圖三(B)表示語意中概念之延申關係。

在例七中，有二點值得說明；其一是語意中重點概念的觀念，其二為概念延申之法則。立字原始圖繪表示的是「人立於地」。真正的重點在「立」，因為不一定只有人可以立，再者，也不一定要立於地，只要是平面(此為地之性質)就好。是故可緣「人是動物」推廣為「動物皆可立」；又此觀念亦可不局限於動物，植物、物品皆亦可立，因此，字義又分化轉變了一層；再說，這個立亦可借用到名譽、品德層面，故更進至抽象層面去了。這種推演過程即是人工智慧中所討論的普遍化(generalization)的問題。可以自辭彙的索引典(thesaurus)按層次翼索推廣。因此，詞彙之類屬關係，亦可與構字合而為一，結合為一更威力的知識表達形態。

【例八】孳乳字系之例

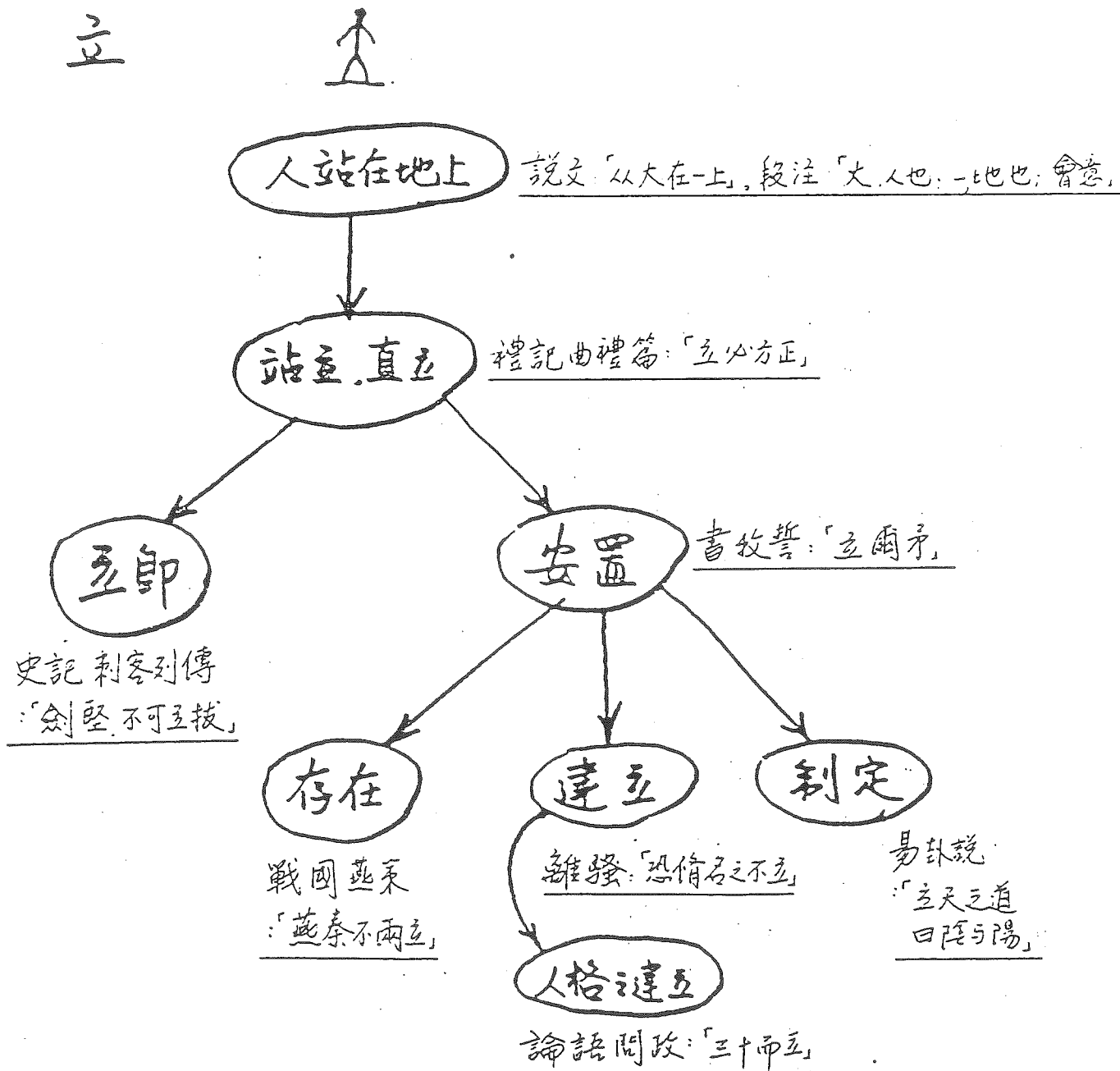
缶字孳乳字系，請參考圖四。缶之本義是上面有蓋，有頸，可有手持，下端為容體的容器，是一個象形字。由於缶是瓦器，因此缶、瓦、甗是屬同義可通用義之字。

缶有六個主要的性質，其一是盛酒「漿」用，此「漿」之意被借為重點而孳生出「菊」字，它是「菓之含漿者」之意。由於「缶」需開蓋使用，故覆以丩(參考表三)而「𠩺」，此為「缶」之重文。又缶有「可握持」之意，借於「劔」字，則表示持刀之意。

缶之顏色有灰有黃，借這顏色而產生「駒」字，表示色如「缶」之馬也。又缶為瓦器，故「匍」象燒缶之「窑」，因而「匍」亦與「缶」通。在中缶多「堆積」而燒，故生「陶」字，阜者阜也，阜者為土「堆積」之狀。

又缶之形狀渾圓宏張，囊之大者如缶，故有「囊」字。再者蟻封形狀如缶，合「囊」與「虫」產生「蠶」字，表蟻封中之虫，即通「蟻」字。

圖三(A) 義訓分化之例



訓詁：義訓「分化」之例

圖三(B) 語意中概念之延申

圖繪：人站在地上

立：站立 [直立，易於立刻行動]

直立 立錐，立杆見影

立刻 立候，~~立要~~ (立談，立) (立馬，立言)

[豎立，安置，放置]物品

豎立 立眉，立眼
[立伴]

置放

設置 立廟，設立，建立

[存在，建立秩序，完成一件事]

立伴
立方，立伴，立樞，立廟

存在
並立，立春，立夏，立秋，立冬

建立抽象事 立誓，立志，立名，立異

[抽象事：{成就，品格，立場}，制定條約]

建立成就 立功，立業

建立品格 立德，立名，立名

建立立場 立志，立場，立言，立身，立意

制定條約
立法，立契，立約，立憲

缶中空，因而擊之有「聲」借重「聲」之意而生「馨」字，鼓擊之聲也。又因有「聲」故可為擊節為「歌」，借「歌」之意而生「謠」字。為歌則需「張口舒氣」，借其意生「歛歛」之詞表示出氣貌。又聞歌則「喜」，借而生「嘍」字，喜也。又歌則諠譁吵人，故生「詢」字。

缶在古時甚為貴重，故「寶」字中亦有「缶」焉，而缶亦為寶之聲符。

這個缶字孳乳的字系是據杜學知「文字孳乳考」[1979]而引來的，雖然杜教授的論點在文字學界尚有爭論，圖四所表達的知識和字的關係，卻是很適合人工智慧技術對文字中知識的處理。

這個孳乳之字系，可用人工智慧中的語意網路，或是框架結構 (frame) 來表示。除了字的意義所含的知識以外，像燒窯時坯是成堆放置的、中空之容器擊之有聲、唱歌時須張口舒氣....等等知識，盡含於此結構之中。若能設計有效之譯出程序，則可在此資料結構中尋出上述之知識。

【例九】現代辭典(國語日報詞典)部首之例。

缶部字系之例，請參考圖五。

根據國語日報辭典：缶部共10字：其中2字「缺」與「罅」，表示瓦器之「狀態」或「性質」，其餘字為器物名稱。根據圖五，我們可一眼看出，此圖與圖四(同是缶)所含的知識差異甚大。在此圖中，應可推導出瓦器有易「缺」及易生「罅」之性質，若不然，此二現象(狀態)之字不會從「缶」孳生，而在缶部出現。同理，瓦器可盛酒漿和固體，亦可為樂器，此等性質均可由缶部器物類八字之字意中推導出來。

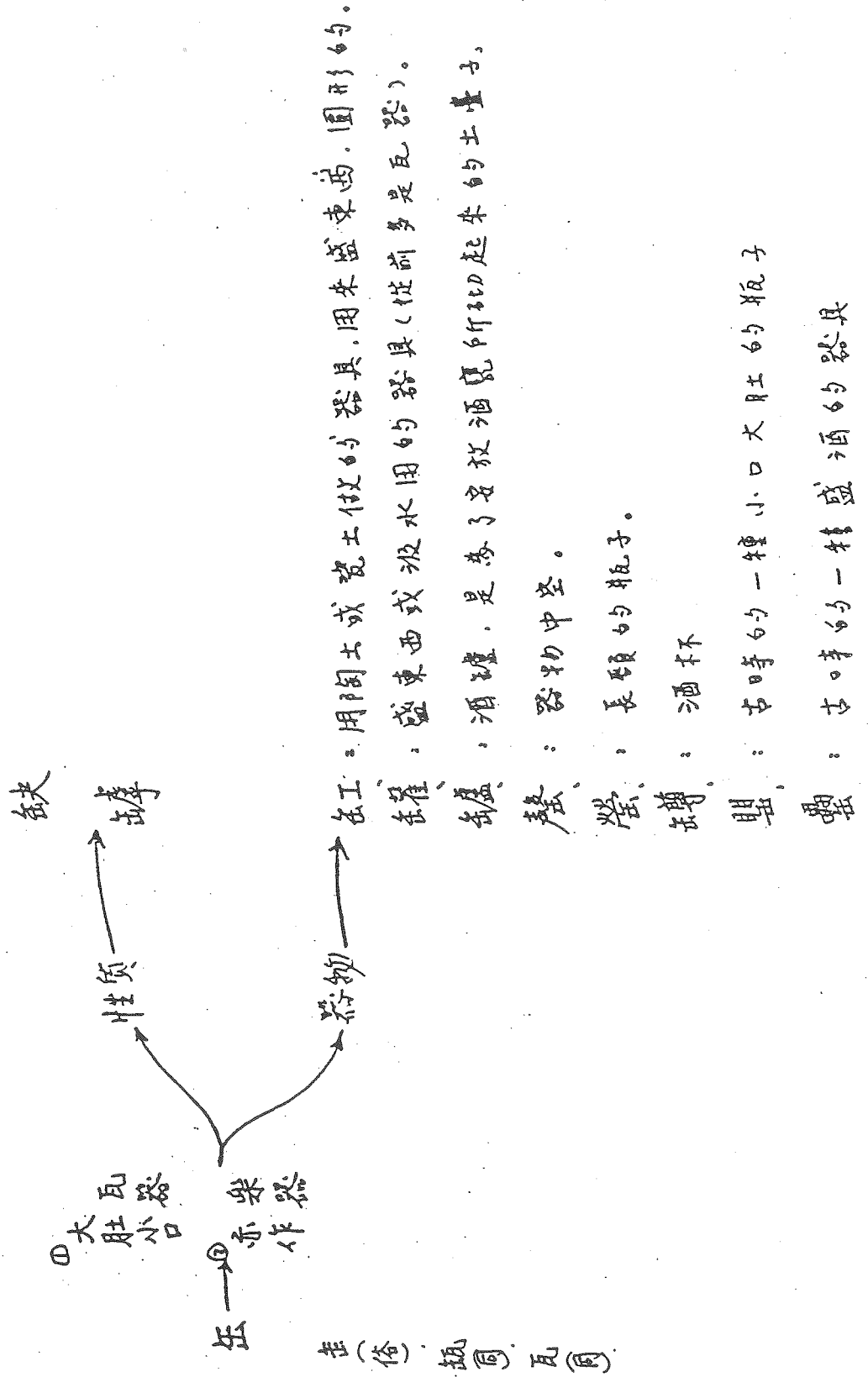
在表五中列有中華大辭典中木字部 336字之分類表。此分類表雖現在甚為粗略，然而木材之主要功用卻不言可知。此外，表示木製器物之字中，有關建築者39字，有關交通者18字(其中13字與舟楫有關)，此外，家具13字、容器12字、文具9字、敲打工具9字、棍棒7字、刑具及棺具各6字、武器4字。由此分類結構應可推知木材之主要用途。

表五：中華大辭典中木字部統計。

樹木名稱	:141字
樹木部名稱	:22字
樹木形貌	:16字
木製器物名稱	:138字
其他	:19字

國語日報解典：

圖五 部首字系之例：缶部



在木字部中有「柴」字，其意為作為燃料之木材，是否由此可推知木之可燃性呢？是否可由此推知「有些家具可燃」的知識呢？當然，這些知識均已蘊含在構字和字典的部首結構中，人可以明瞭，可是電腦也聰明的推導出這些知識嗎？這是仍須努力的研究題目。

討論至此，我們已將在各層面文字中蘊含之知識，包括：造字之始，六書、文字孳乳過程，和現代字書等一一討論；而在各例中，亦分別以類比，語意網路，劇本框等主要的知識表達方式討論其可行性。文中沒有把常用的知識表達法一一舉例，因為這是不必要的，知識表達法並無排他性，當用一種可以表達時，常常也可以用另一種方法來表達。

系吉

語

先賢造字，肇始於對自然現象之觀察，此中涉及人體、天文、物理、化學、動物、植物、等等科學領域。以現代科學的水準來觀察千百年前文字的演進，我們幾找找不到違反科學的痕跡，這是多麼令人驚訝的事！由此可知，先賢造字均能摒除巫迷，在文字的結構裡表現出對事物真象客觀正確的體認。這種具體而微，詳盡周延的觀察功夫，深深契合科學方法的要領。對於人倫、文化、社會方面的觀察，也充份表現出持正不阿，善體天心的人本文化精神。在文字孳乳方面，則思路之明晰，推論之邏輯，觀念之延申，聯想之法則，重點之把握，以及由實物進而引申到抽象意義的創意過程等等，精闢絕倫，無不表現出卓越的睿智和極其科學的內涵。這種成就，發生在數千年前的上古時代，實在令人難以想像！現謹就本文討論所得綜合之，分列如下，作為本文之結語。

(一)中文構字以知識為基礎，構詞亦然。從知識或認知的角度 來研究文字學應是值得重視的方向。

由本文的討論可知中文構字以知識為基礎。構詞亦復如是。這種方式和拼音語系以音為主的做法截然不同。正因如此，中國文字蘊含的知識成分，就遠較拼音文字為多。試比較：「冰」與ice，「愛」與love，就可了然。

中文構字與孳乳，凡六書、訓詁、音韻諸學，已有相當系統化的架構。從知識的角度來看，這就是一種結構化的知識。似乎前人很少有從知識的角度來了解中國文字的體系，若由此嚐試，或可發前人之未見。

若將構字知識妥為電腦化，則其蘊藏之知識，實為極佳之常識(知識)庫(common-sense knowledge base)，而常識庫正是目前人工智慧中急待突破的問題。如果我們的文字結構真的較西方語文更適合做常識庫，那麼是名符其實託先祖的福蔭了。

(二)在中文構字中，形音義三者密切相關，無親疏之分。

有些語言學者認為「音」與「義」是語言最主要的兩大要素，至於文字(形)則是次要的，它只是紀錄語言的符號罷了。文盲亦可流利的使用語言，就是常用的佐證。這種說法，在西方拼音語系中尤顯得有力，因為拼音文字的「形」與「音」的關係是直接的，而「形」與「義」的關係就比較疏遠；「形」常須經由「音」的媒介與「義」相關連。然而，我們文字「形」與「義」的關係就不盡然如此。在中國文字結構中蘊藏了豐富的知識，這些知識和字「義」緊密的結合在一起。換言之，我們的文字，「形」「音」「義」三者是密切相關無分親疏的。其實，世界上任何一種語言皆有豐富的知識內涵，然而中西語言表達知識的層次不相同，在我們的語言裡，即使在構字的層面上已經把知識和語言緊密的結合在一起了，更不用說構詞，構句和文章了。

(三)我國文字結構，是一個典型人用的知識表達系統，

它與人工智慧技術可相輔相成

從人工智慧的角度來看，我國文字的結構，實是一個給人用的「最佳知識表達系統」。文字中的「字根」是知識檢索的鑰匙，構字所呈現的是知識的骨架和要點(key-note)，而六書、訓詁、音韻的知識即可作為知識的譯出程序。由本文所討論的各個例子可知：它們所用的基本技術和人工智慧中常用的一一對應。此二者的結合，是極為自然的事情。由此觀之，人工智慧技術極適合作為文字學研究的工具。若能善用此工具，不僅可以提昇語文學的研究，而善用我們文字的特質，亦將全面提升計算機處理中國語文研究的進展。

(四)未來的工作

如何將蘊藏在文字中的知識整理出來，再設計為實用的知識庫系統，仍然是未解決的問題。然而本文指出的觀念和論點，則是一個嶄新的研究領域。

本文所談的構字與知識的關係，也是一條學習中文的途徑。從此著手應該可更有系統，更有效率地講授中文。循此發展，或可改善我們的國語教學。

總之，由於語文的用途廣泛，計算機的功能強大，本文所提的方向：結合文字學與計算機科學的研究，或將對許多語言問題的研究產生影響，若能有所助益，正是本文之所企盼也。

參考資料(姓名筆劃序)

除文中標註之處以外，其餘皆為通用之參考資料，如下表

1. 王申培〔1987〕，Intelligence in Chinese Character, Proceedings of 1987 International Conference on Chinese and Oriental Language Computing (pp.85-88), June 15-17, Illinois Institute of Technology, Chicago, Illinois, USA
2. 王初慶〔1983〕，中國文字結構析論(再版)，文史哲出版社，台灣，中華民國
3. 杜學知〔1979〕，文字孳乳考(再版)，世界書局，台灣，中華民國。
4. 林 尹〔1971〕，文字學概說，正中書局，台北，中華民國。
5. 周 何〔1983〕，中文字根孳乳表稿，文建會國字整理小組，台北市南港區一之十三信箱，中華民國。
6. 齊佩瑢〔1985〕，訓詁學概說(台初版)，漢京公司，台灣，中華民國。
7. 謝清俊〔1972〕，中文字根之分析，交大學刊(第六卷，第一期，PP112-121,)新竹市交通大學，中華民國。
8. 謝清俊〔1987〕，談中國文字中的知識結構，中央研究院史語所語言組五月卅日例會之講稿；(可與本文作者連絡)。
9. Barr, A and Feigenbaum, E.A.〔1981〕，The Handbook of Artificial Intelligence(vol.1); William Kaufmann Inc., Los Altos, Calif. USA
10. Hofstadter, D.R.〔1980,Sept.〕，Godel, Escher, Bach: an Eternal Golden Braid(vintage books edition), Basic Books Inc., New York, USA

誌

謝

本文之能寫出，首先要感謝中研院史語所丁邦新所長給作者登門求教的機會，經史語所語言組同仁們若干斧正之後，修正如本文。其次要感謝謝瀛春教授更正若干錯別字、曾士熊先生的討論、顧秋芬、楊淑吟兩位小姐幫忙稿件整理。沒有他們的協助，本文無法如期完稿。謹在此一並致謝。