

語意網、詞網與知識本體： 淺談未來網路上的知識運籌

中央研究院語言學研究所研究員 黃序仁

【摘要】：本文為「佛教知識組織管理研討會」專題演講稿。演講內容分三個方向：首先介紹語意網的概念與發展，其次從語意網的發展談到需要本體論，再由本體論談到外相，需要非語言來表達，以及論述語言如何表達知識。本演講並報告在「國家數位典藏計劃」中所做的語言座標計劃，正是這方面的一些研究。

關鍵詞：語意網 (Semantic Web)；本體論 (Ontology)；詞彙網路 (WordNet)

今天很榮幸有機會來與各位講教，卻有點誠惶誠恐。因為剛剛謝教授已經說過語言是表相，表相要脫落，脫落完了之後才來講表相，恐怕有點捨本逐末。不過，剛剛謝教授講了很重要的一點，就是古知識表達上我們不能脫離語言。雖然說我個人對佛教的知識事實上談不上管理或運籌，是一個門外漢；但到底一直從事語言學的工作，最近所關心的一些議題又跟這個會議的議題是有點關係，所以特別來跟各位講教。我注意到古今天的議題裡面，有幾篇文章特別跟我今天想向各位講教的專題有關，而且有更深入的討論。譬如說阮明湖教授基本上就討論這個本體論怎麼樣去應用；歐陽彥正教授那篇文章的主題圖裡面也是一個本體論跟知識表達方式應用的一個實際例子，而且運用在「玄奘西域行」的一個網站裡面；當然還有成人的陳玉女教

授，一個很簡單、很完整的佛教本體論知識結構。由此看來，個人才疏學淺今天只能跟各位做非常粗略的介紹，事實上這幾篇文章也許能給各位更深入的討論與更貼切的啟發。

今天演講的方式大概分三個方向：第一個方向，先介紹語意網的概念與發展，從語意網的發展談到需要本體論；本體論之後，就回到本相和外相，外相需要用語言來表達，那麼語言怎麼表達知識；最後如果還有一點時間的話，就跟各位報告我們在「國家數位典藏計劃」裡面所做的一個語言座標計劃，正是這方面的一些研究。

電腦也能看懂語意網

如果各位沒有聽過語意網，它就是 Semantic Web。各位如果知道，那我是重複講了，請暫且



客忍。如果各位沒有聽過的話，這是最近一個很重要的發展，在網路上的發展。它認為電腦也能看懂語意，Semantic Web 基本上是一個新的網路內容形式，能讓電腦理解其中的語意，勢必帶來新一波的網路革命！也就是說網路走到今天，往哪裡走？大家認為就是往 Semantic Web 走。這篇文章第一次發表是在 2001 年五月的 *Scientific American* (科學美國人) 上。五月 (2001 年) 五月發表的英文原文，向全世界宣告這個新方向；今年 (2002 年) 八月由高虹博士翻譯，我做校訂，出版在今年《科學人》的八月號 (第六期) 中〈電腦也能看懂語意網〉(The Semantic Web)。所以基本上所有當今語意網的研究，所根據的就是這篇文章。作者為 Tim Berners-Lee，柏納斯。他就是發明網際網路全球資訊網 (World Wide Web, WWW) 的人。以下的介紹主要是引用這篇文章。

語意網將成為下一代的網際網路

這是一個什麼樣的發展呢？第一個，他主張並認為語意網將成為下一代的網際網路。我們剛提到他正是網際網路發明人，因此他對網路發展的宣告，不可忽視。語意網的發展方向並非沒有人批評；如科技方面的人，很多人說他講的是天方夜譚，不可能成功的，就如當初他推動全球資訊網 (WWW) 時，有人說不可能成功一樣。也許一些他個人堅持的細節是需要保留，但是他對網路發展的遠見不可忽視。把知識內容加入網路的方向是一定要走的方向，不管將來網路確切的形式為何，是不是跟他所宣告的完全一樣，我們必須要來討論說：在這個發展上面，我們網路上的人心，包括佛敎，知識典藏如何因應？如果說

再過十年之內，整個網路的面貌改變了，那麼我們還在談現在的網路的知識典藏，我們就會落伍，我們的知識就會被淘汰。因此，我們必須預期到網路將來的發展，而且有所因應。

從「全球資訊網」到「語意網」

那麼他認為全球資訊網和語意網有什麼不一樣的呢？他認為全球資訊網仍只是人們交換文件的載體 (media)，其中的資訊是機器不能自動運用的；也就是說，他認為現在的網路上，我們交換文件時，只是人把文件放上去，在網路的某一端，另一個人把文件拿下來。我們用來執行工作的電腦，它不需要了解文件的內容。事實上，它完全不了解文件的內容，但它知道這個文件的身分是什麼。它怎麼找到這個文件的呢？是根據我們的一些題目、後設資料，我們對文件其他的描述。它對文件的包裹裡面，不管是文獻、圖形、或任何檔案，裡面包裹的內容是什麼？電腦是不知道的，它是把它當做一個貨物來運送的。這個對於現在所謂的全全球資訊網做大概的描述，是正確的，是很有效的一個載體。那麼他希望什麼樣呢？他們希望說，如果我們針對電腦，增加專門提供給電腦閱讀的網頁，我們就可以把現有的網路轉換成「語意網」。所以，他對語意網的意思就是說，如果我丟進去的資料電腦看得懂，電腦就可以做別的事情，這是他的一個想法。

電腦如何閱讀語意？

電腦如何閱讀語意呢？在他的想法、架構裡，其實市場各位也有很多是對這個很熟悉的，第一個要利用資源描述架構 (RDF, Resource Description



Framework) 與通用資源標誌碼 (Universal Resource Identifier, URI) 連結到相關網頁資源。這是現在網路上很多已經用的了。大家用的 HTTP 的位址, 就是 URI 的一種。很多人除了後設資料以外, 也開始用資源描述架構來描述網頁裡的知識內容, 這是一個人的架構, 你要在網路上找到某個特定的資源, 完全沒有問題。我們規定人家用這個方式描述你的知識資源內容, 那麼你找到你要的資源, 並經由超連結連接到之後, 最重要的閱讀重點在哪裡? 是要利用知識本體 (Ontology) 來定義關鍵詞, 並做邏輯推理。

知識本體 (Ontology)

重點是, 每一個網頁, 每一個資源上面要有一個自己定義的 Ontology, 就是知識本體。為什麼要這樣做? 同一個詞, 在不同的領域裡, 不同的時代裡, 不同的用法上, 它的意義就不一樣。所以, 當下網路的搜尋常會發生錯誤。基本上, 譬如你找一個芥子, 那麼也許講的是佛教的「須彌入芥子」, 你要講的是佛教的東西, 但你找到的可能是植物的東西, 因為它用的是同樣的語詞。網路並不知道每個網頁資源用的詞屬於哪個領域。所以不但搜尋者必須要定義這個詞代表什麼意義, 屬於哪個領域。在每一個網頁裡面, 每一個詞的定義是什麼, 代表的知識範圍與架構是什麼, 就是這個網頁的知識本體要告訴你的。所以, 他的想法是, 如果網頁上所有的資源都有一個宣告, 告訴每一部來訪的機器, 這個網頁裡面知識的定義是什麼, 知識的架構是什麼, 那麼電腦就可以閱讀每個網頁。

我們剛才已經稍微提到知識本體是什麼; 它是

對任何一個網頁資源知識內容及資訊架構的描述和定義。而且語意網的構想希望知識本體是用 RDF 的型式, 或類似語言寫成之文件, 清楚定義概念關係和推理的邏輯規則。我們注意到這裡, 他認為有兩個架構, 第一你怎麼描述完整的知識呢? 你要告訴機器, 他認為這是給機器看的而不是給人看的, 所以你告訴機器, 所有在這個網頁或在這個資源裡面, 有哪些概念或全部的概念放進來。第二是電腦如何在概念與概念之間進行邏輯推理? 有什麼什麼規律? 有哪些邏輯關係存在? 這是要給它的一個知識本體。在這裡同時提醒人家, 請注意資訊學中的 Ontology 當成知識、訊息的基底架構, 與哲學中的本體論 (一切基本義) 的原定義大不相同。所以, 不談本的問題, 只談相的問題。

知識的演化

根據相綱李的說法, 知識的演化更為重要。知識本體除了在網路資訊上明顯的應用外, 他認為更重要的是有一個知識演化的意義存在; 認為如果設計得當, 語意網將有助於人類的總體知識演化。為什麼語意網會有助於人類總體知識的演化呢? 因為網頁的知識本體提供了不同知識體系的完整描述; 也就是說, 當我們在做一個佛教知識網路的時候, 我們把佛教知識的一個關係和知識架構作一個描述。如果說今天假設人家已經對所謂的本體論的作法、語意網本體論的作法很習慣的時候, 一些不懂佛教的人, 他到佛教的網頁, 進入之後, 經過機器的關係幫忙了解整個知識架構後, 當然就會了解佛教的知識, 就可以進入佛教知識領域。或者再以《六祖壇經》為例, 光看《六祖壇經》有很多人看不懂, 我把它的知識



本體架構建起來，不管是人或機器進入到那個知識體系裡面，當然會更容易理解。因此，他認為有同樣的語言，就是用 URI 去每個知識體系中明確描述了每個概念與語意的關係。那麼，語意網將有助於語意概念的溝通與知識體系的整合。因為每個知識體系已經有一個架構在「這裡」了，那這些本來雞同鴨講的，你講的話我聽不懂，你認為你講得很很有禮貌，我認為你去跟我吵架；但如果我把我的知識體系告訴你，我的語意是這樣了，我的推理是這樣了得來的，你先得我的知識體系之後，各方面應該可以更好溝通，這是他的理想。也就是說，雖然原始的設計，強調知識本體是給電腦看的；更遠大的目標，卻是希望這也成為人類知識的系統化整理。讓知識本體可被人類所閱讀使用，成為人們知識融通的橋樑。

這個當然非常棒。所以，儘管在技術上或在種種具體細節上，大家有各種不同看法，但這個方向是正確的，大概人家也會往這個方向走。大家如果想要得到語意網這方面資料，這裡提供各位幾個簡單的入口網站：第一個網站是 <http://www.w3.org/2001/sw>，我想大家很熟悉，就是網際網路的管理機構；接下來 <http://www.semanticweb.org>，有關語意網的個別標準在這裡可以找到，這已經是一個語意網的獨立組織，它來告訴你什麼標準、什麼想法。另外；可以從《科學美國人》(Scientific American) 網址 (<http://www.sciam.com>) 找到 2001 年五月完整原文下載；但《科學人》中文版這部分，我們雖然有翻譯，但並沒有在網頁上提供下載 (<http://www.sciam.com.tw>)。



▲「佛教知識組織管理研討會」黃居仁教授專題演講 (編輯組提供)

關心的問題之一：語意網將會使用什麼語言？

我們既然發現網路上有這樣的一個發展，而且是針對知識的內容表達方式和形容方式發展，我們會關心一些問題，我們想問語意網往哪裡走？關心的第一個問題就是「語意網將會使用什麼語言？」。在 WWW 上，雖然有許多語言在使用；但資料最多、通用的習慣語言是英文。語意網用什麼語言呢？我給兩個答案。第一個答案，用英文。為什麼這樣講呢？很簡單，現在網際網路上大部分資料都用英文嘛！其實，你要選一個全世界共用的一個表達語言，那用英文就好了。因為現在都習慣了，網際網路已經教人家英文了。可是，如果我們想這個問題：十年後的網路大家是否還是都用英文呢？因為大家看得到網路上的資料大部分還是使用英文，在語意網蔚為風氣之前，似乎不可能有任何語言在網路上的資源



量會超過英文。可是還要考慮到幾個問題，並非這麼單純。比如對網路使用人口的預測。以下這幾個預測是我們最近剛剛在台北主辦的 COLING2002 國際計算語言學會，其中有一場中文研討會，請大家預測十年後的網路與中文處理。朱邦復先生說：「我不要十年之內，要在五年內讓九億農民上網。」微軟研究院自然語言組經理周明的預測是：「我們比較保守，認為十年之內將有五億中國人上網。」我個人的預測是：「我不知道有多少人上網，但我想在十年之內網上的入口裡面大概有四分之一是講中文的。」如果十年內這是個事實，如果朱邦復先生的九億農民上網，那這九億農民裡面有多少人是講英文的？一億了不起吧！我們以往在看網路發展的時候，因為我們看到的是西方的發展、以及科技化國家中的發展，譬如歐洲。歐洲再多兩倍的人上網，還是用英文，因為他們會學英文，會用啊！今天如果我們產生這個質變，真的讓九億農民或五億農民上網的時候，他不會英文啊！所以，網路上不管是他製造的資料或尋找選用的資源，中文就變成很重要了。▽假設再過十年，印度人口比中國人口多，有九億印度人上網。因此，剛剛相納拿的前提假設說：「下一個時代的網路，一定用英文。」是不見得正確的。雖然現有的東西都是用英文，但如果你增加進來的新手使用人數夠多，而這些人是不懂英文的，怎麼辦？當然會用他自己的語言，而所製造的資料當然就不是用英文，那麼就會產生問題了。

如果不是用英文的話，第二個答案是任何語言都可以。為什麼呢？因為語意網的前提：「是靠知識溝通，而不是靠語言溝通。」為什麼一定

要用英文，知識本體誰規定一定要用英文寫的呢？沒有人規定啊！任何語言只要可以表達知識的體系，當然都可以。只要有知識體系在這裡的時候，可以讓不同語言之間，有更好的溝通，知識的交往。當然是有規定的，如有關 OWL、XML……，是大家都很熟悉的。大家都認為，語意網的內容標示會用 XML，這大概是沒問題的。但有規定一個 Ontology Language，是最近發展中的，就是描述本體論時，應該用什麼語言。有人認為用 OWL，可以在 Semantic Web 的網站看到。所以，這問題的兩個答案都有可能，但是我們希望它是第二個答案。

關心的問題之二：既然語意網主要依靠知識本體；特定語言與本體的知識管理還有需要嗎？

既然語意網主要是依靠知識本體，是知識為重，那麼特定語言或本體的知識管理還有需要嗎？也就是說，既然「已經脫落了」，「已經看到知識了」，還要表相做什麼呢？答案甲：當然是不需要。因為知識本體的存在是獨立於特定語言與本體之外的。那我們今天就不需要討論什麼佛教知識管理啦，什麼中文知識管理啦，我們只要討論本體論就好了，我只要把本體論統一就可以了。可以做得到嗎？答案乙是當然還需要。其實剛才謝老師在第一場演講已經講得非常清楚了，所以事實上我在這裡是重複了。因為每個特定語言或本體都是一個獨特的知識體系。也就是說，我們習慣用語言來表達一些東西，人是著重於表相的，必須靠這些表相才能表達我們的知識。選擇一個表相表達知識的時候，就受到表相的限制；



除非把所有的人類都「經了」知了，都不用語言就可理解了，可以溝通了。古還沒有到達這個地步之前，當然所有的資料還是古不同的語言裡面，因此還是需要正確整理分析一個語言所表達的知識內容，才能夠建立知識本體，還需要語言這個工具，才能達到這個知識。我個人當然還是比較趨向於第二個答案。

知識本體的變遷

知識的豐富性從何而來？各位應該都很熟悉，從文化、領域、環境、族群、社會階層、媒體、學科、時代等……，不同文化的背景。同樣講一個佛經的故事，如佛陀講的故事與禪宗大師所講的故事，他用的語意、背景、故事對象都是不一樣的啊！也許所表達的意義是一樣的，但用的語言、譬喻都是不一樣的。領域、族群、環境、社會階層、媒體、學科、時代等等，你到每個不同的地方，它就用不同的方式講。剛才謝老師「給我一些例子，你古兩、三千年前怎麼可能講碳、氫、氧，因為古語言知識裡面，並沒有碳、氫、氧這三樣東西，只有講風火水土。古一個時代裡面語言有它的限制，會有不同的表達方式。組成知識本體的元素不同，就影響到元素間的意義與推論關係，進而影響知識的架構。

知識的豐富性如何體現呢？我們面對的知識真的還古變，每個時代變，每個環境變，每個語境變，每個學科變。如何表現其豐富變異，又維持內部知識表達的一致性呢？答案就是用每個特定環節中共同的語言與詞彙。我們古圖書館界、資訊界講所謂的次語言（sub-language），或講白話就是行話嘛。每一行有行話，每一個時代有它自己的行

話；領域詞典，就是專業詞典。由外部看，去看行話，看語言的表達方式，看它的專門詞典，你就可以知道它的知識體系是不一樣的。但由內部看，行話正是行內共享與溝通的依據與標準。

由（知識的）本體到語言的本體

我們知道也看到整個網路的趨向，是心願要走到一個以知識本體為中心設計的網路。整個網路要更好用，每個個別資源要更能流通應用；針對下一個時代的網路，心願對每一個資源，每一個網頁，有知識本體的描述。但第一個需要跨越的門檻是不可能有一個人類共通的知識架構，即大家都同意的知識本體表達方式。因此，我認為第一步需要建構的就是語言的本體。為什麼這樣說呢？因為每個語言都有其架構完善的知識本體。為什麼說每個語言有架構完善的知識本體呢？因為語言包括次語言，可以表達所有（領域內的）知識。譬如我們古這個領域裡，古談佛教知識用的一些術語，人家都能彼此了解，可以表達領域內的知識。我們所用的語言，幫助我們來溝通，說話者與聽話者之間知識的交換，通常是有效而正確的。當你碰到一個史瓦濟蘭來的佛教徒，如果要跟他用英文談佛教的時候，會有一點點問題，但古佛教知識小領域裡面還是可以談得很好。同樣的，如果說古台灣的居士之間，都用台灣的語言當然沒問題；如果到九華山去跟大陸高僧談佛教領域內的東西，談的時候還不錯，但當談到一些比較灵活的東西時，似乎不太一樣了。相對的，與台灣同胞談靈活頭頭是道，但如果是和不懂佛理的人談佛，卻又會很吃力。所以，每個小詞彙，每個小的語言、次語言，古大語言



中間有一個共通的知識架構，在這個知識架構範圍之內，交換知識是沒問題的。既然語言已經給我們一個自然知識架構，為什麼還要跳出語言找其他的東西呢？我們可不可以回來見山是山？所以，既然我們不知道這個世界通用的知識本體怎麼樣，我們先從語言開始，起碼知道，在經驗中告訴我，一個語言裡面，一個次語言裡面，它的知識本體架構，是大家可以用通的。

第二，就是任何人用知識本體的使用者，不論數目或準確性，都不會超過語言本體的使用者。我想到目前為止，這是正確的。今天有很多人認為 Ontology 要怎麼表達，剛才講表達得很漂亮的佛勃本體論，你說用這樣的本體論來描述知識的人有多少？可以表達出來的有多少？這個數目絕對不會比講中文的人多，那他所表達的知識範圍，絕對不會比中文語言豐富。所以，這些是我們利用本體架構或本體的理論，用來溝通的一個手段，但是否還沒有得到一個完整的理論之前，語言給我們的是最完整的、最有效的知識體系。

最後，語言本體本來就是文本典藏處理的必要資源。為什麼這麼說？因為要做任何一個文本典藏，如要做詞典，要做同義詞典，要做專門詞典、雙語辭典等，別人都做了。工作都已經做了大半了，從這些基本的語言資源到現有的本體論，還是需要工作，但事實上已經是為了山見山，已經到最後一步了。

如何呈現語言的知識本體？

在前面第一部分的報告中，我希望已經說服各位：知識本體在未來網路的發展上是不可或缺的，而且希望我們著手做本體時，從語言做個媒

介。再來看語言媒介如何做？我們認為應該用詞彙網路 (WordNet) 來做。

如何呈現語言知識本體？認為詞彙網路是最直接的語言本體表徵。詞彙網路的建構與研究已經有相當的時間，自 1990 年普林斯頓大學開始規劃進行。詞彙網路構成的元素有哪些呢？第一是某個語言內所有的詞彙。請注意語言學上所謂「詞彙」的定義是：任何一個詞形 lemma 與詞義 sense 的獨特配對為一個詞彙。用中文的例子來說，比如說「鎖」，我們第一個想到鎖是一個工具，對不對？其實鎖這個字，這個詞形 (鎖 suo3，相同的音，一樣的字形) 它至少有兩個蠻清楚不同的意義：一個指各種不同形狀的鎖的實體，如喇叭鎖；另一個是鎖的動作。這兩個指的是不同的意義。所以，如果做詞彙網路及語言學分析的作法，就要變成兩個詞彙。詞彙的區辨是以意為主，以形為輔的。意義不同，不論是否共用一個詞形，一定要視為兩個詞。意義若不分，當有一個以上可互換的詞形 (如異體字) 時，則視為同一個詞彙。若是英文就更明顯了，在獻中，中文的例子比較少，英文比較多。譬如英文的 bank 這個詞，講「銀行」是寫 bank，講「河岸」也是 bank，講船跟飛機傾斜的動詞也是 bank 等等之類的，或者說你到某個銀行開戶等，也叫 banking、to bank，動詞也是 bank。這麼多意義，它每一個意義跟每一個形的組合就是一個詞彙。在詞彙網路中，bank 這個詞形共有 17 個詞義。這個就是剛剛謝老師講的「依義不依語」這個表相，根據它的意義，每個意義是一個單位，而不因為外表的相與形狀而做一個單位。不因為它是一個 bank，就認為它是一個詞，因為它的意義，就會把它區分開來。所



以，每個語言所能表達的所有概念，就是所有詞義 sense 的集合，正好在詞彙網路中表達出來。

再來最重要的就是，一組基本的詞彙語意關係。如果我再借用謝宇師的話，用表相「依法不依人」；也就是我不去描述它的本體，因為我現在很難去告訴你，它的意義是什麼？我怎麼樣去表達它？有些基本的概念幾乎是無法定義的。但是我有一個辦法告訴你，這些詞義跟詞義之間，意義與意義之間可以有哪些關係，而這個關係是固定的，有限的。也就是說，意義的本性不可說；但意義與意義間的關係是可以有規律來描述的。

詞彙網路的架構

詞彙網路架構的第一個準則是以詞義為基準，把所有相同詞義的所有詞彙放在一個同義詞集 (SynSet)。這麼一來，同義詞集即是表達相同概念的所有詞的集合。也就是說，當一個語言（或次語言）經過這樣的分析歸納之後，這個語言裡所有的概念都整理出來了。可能有一個以上的辦法來表達某些概念，但是當把所有的詞義的集合列出來之後，這個語言裡所有的概念都已經有了。除了英文的 WordNet，在歐洲的 EuroWordNet 也已經做得差不多了，認為相當完整。但其實不可能真正完備無遺，因為語言是動態、有生命力的，不但隨時會有新的意義出來，現有的語意及詞彙也可能改變。但就一般使用而言，上述的詞彙網路是相當完整的了。他們的同義詞集或概念詞集大概都有十萬個左右。

下一步就是以定義過的詞彙語意關係，連結所有的同義詞詞集。原先依法不依語，已經把概念找出來，把所有的詞義找出來之後；第二步再

來依法不依人，不依它的相，只依意義與意義之間一個法的關係歸納出來。我們如果還記得，剛剛講語意網的時候，語意網的關鍵就是，每個網路資源的知識本體裡，需要二個東西：第一就是這個網路資源裡面所有的概念，再來就是跟所有概念之間的邏輯推理關係。把所有的概念找出來，把所有的邏輯推理關係連結架構出來，就是一個完整的知識本體，整個知識架構的範圍就占裡面了。我們之所以主張要用所謂的語言本體論關係，是因為如果利用詞彙網路這樣的現有語言資源，就已經有相當多的資料。把現成語言表達的所有概念中找出每個特定網路資源引用的；然後把所有概念之間的語意關係找出來，便是建立了該網路資源所有概念間的邏輯關係的雛形。當然，每個特定資源一定有超出語言一般表達的概念或關係（或有一些不在語言中表達的概念）。但有了以語言為基礎的豐富資料與基本架構，個別知識本體的建立才有可行性。

這裡要注意到詞彙關係這個法是具相當知識內容的。譬如大家熟悉的索引典，或同義詞典為例，基本上會把同義詞抓成一組，但對詞彙網路，除了同義、反義、近義之外，更重要的是上位、下位，及功能等關係。也就是不是只講它有相同的關係，因為意義與意義之間的關係的法，並不是只有身分相同而已。它有很多的關係，這個關係是我們真正去利用推理的時候，或者是去了解知識內容的時候，所憑藉的東西。

詞彙網路提供的知識基本架構

首先再強調一次，我們認為詞彙驅動的概念 Synset 是知識的單位。也就是說我們用語言來表



達知識的時候，最小單位是一個概念。但是這個知識單位的表相是什麼呢？是一個詞彙。對任何一個概念，當然不是很容易馬上描述它的本體，它的實質是什麼。你跟我描述說概念是什麼東西呢？我利用語言來表達它，但是能夠把它切成一個詞彙的時候，知道它是一個概念的單位。所以，這是用這個表相的單位做一個驅動的概念，達到的知識。然後知識與知識間有一些關係，它是怎麼衍引出來的呢？是用語意的關係來表達出來的。所以，這個語意關係（Semantic Relation）說明了概念連結與知識衍引的基本關係。

我們回到一個簡單的問題。我們現在知道在網路上，為什麼一般搜尋不好用，會錯？是因為它是按照一個詞的形狀去找。或如 Google，去找詞的時候，是依照大多數人找的詞去搜尋的，也就是按照群眾的行為去找。事實上我們對圖書資訊界還蠻熟悉的，如果是一個專家（比如是參考圖書館員），去問一個問題的時候，要用什麼方法去找關鍵詞找不到的資料？

語意網路與知識運籌

在網際網路的時候，雖然它的媒體、載體改變了，但它所觸及的對象還是文件，或是一個書本，一個圖形檔案。如果進入語意網的時候，任何要處理知識運籌的範圍時，會遇到兩個人問題，一是知識如何計畫（How to Itemize Knowledge?）；二是知識如何編目（How to Catalogue Knowledge?）。我可以告訴你要找哪一本書，要找一個實體、一個文件、一個 document 的哪一個 title。這是一種實體，是在全球資訊網可以做的事情，因為它就是一個單位。我規定它

是什麼單位，給它一個 metadata 控制資料，按照這個資料去找。但是如果進入語意網的時候，這些假設都被打破了，因為它有效的單位定義不再重要，都被打破了，只剩下知識的本體而已。我要找的東西是一個知識，不是一個特定的實體。那怎麼計畫知識？怎麼知道知識有多少？怎麼去編目知識？怎麼對知識分類？

語意網路與(圖書館學中的)索引引

所以，稍微退一步回想，我們原來一直很熟習的做法，圖書館學或資訊界所建立的「索引引」。 「索引引」的著眼點在資料與檢索結果的標準化，因此建立分類別 equivalence class 的概念上；並輔以分類為單位的（概念）層級架構。其目的是要可以很快找到預先設定的固定類別。因此原則上只標註同義關係（synonymy）。如同個名詞，有稱為「佛陀」，也有叫佛祖的，等等之類。按照步驟去找，也許可以把跟英文不同的其他語言，如華文、梵文等等。這麼一來可以很快找到預先設定的固定類別，也可以處理部分多語問題。只要規定它是這個類的，當然可以找到，而且找得很好。

但很有可能，我們要的那一個類並沒有被定義過；而且環境會改變，如果分類定義因時制宜、因地制宜，就會不一樣。舉一個是前一陣子蠻具政治性的說法，「番薯」代表台灣本地人，外省人是「芋頭」。所以要找這個同義詞集的時候，找本地東西索引引可能會告訴電腦也要找番薯，要找外來人的資料也要找芋頭。這是政治語言上的一個東西。可是，如果你想要做植物學呢？芋頭是本土植物，番薯是外來植物；番薯是從美洲印地安人來的，芋頭是本土的，那要找的東西不是



完全相反了嗎？這個分類定義法的問題，就是迷於語言的表相，但是實質呢？因為知識的內容跟意義，在不同的時、地裡面不一樣，我們要怎麼樣找到知識的本質呢？非用語言不可。網路資源，所有表達方式，以及使用者去檢索的時候，所用的可能是同樣的表相（詞形），但他們可能分屬不同知識本體，或在不同的架構節點上。電腦怎麼能知道呢？答案就在每個網頁上，知識的表達上面；告訴你，我的知識本體是怎麼樣的架構，用了哪些概念。在植物的知識本體裡面，當然，原生植物跟外來植物、引進植物，是分在不同的本體架構裡面。在一個政治語言裡，它又是一個不同的本體架構。只要標示它，就不會有上述的問題了。因為使用者不是因為簡單的下一個「番薯」的檢索或本地的檢索而已。他是在一個語境底下，一個知識範圍底下的一個檢索。但是前提是，要有這麼多知識嗎？有多少知識才是夠呢？譬如「響尾蛇」跟「飯匙倩」是相同的東西，在索引典裏是相同的。但使用者下的指令是搜尋「響尾蛇」還是（台灣的）「飯匙倩」，它的意義上又不一样了。飯匙倩它是一個台灣地區的響尾蛇，那麼它一定有特別的意義聯想。比如心會在某一個詩裡面，我如果寫的是飯匙倩，不是響尾蛇，那麼可能是描述我在台灣出長的經驗；若我寫響尾蛇，我可能是寫我到美國亞利桑那州旅行的經驗，那個經驗上的層次是不一樣的，它勾起了知識的聯想是不一樣的。那要不要去找這個勾起的知識聯想，在描述語言這個表相給你內在反應的時候，要不要這些東西呢？所以同義詞永遠不同義，因為在某個知識本體之下，實地的東西要知道，實地範圍要知道。

第三個的例子就是說在不同的語言裡，即使是被認為是對等的詞彙，它表達的概念及關係很可能還是不一樣。剛剛提到中文的桌子，翻譯成英文是 table，還是 desk 呢？在做跨語言的檢索時，當指令是用中文下的，要找桌子；資料檢索時，table、desk 都要全部找回來；要找椅子時，有 stool、chair、armchair、sofa 等，全部都要找。或者雖講 sofa，但在這個語境裡面如果只是要有一個舒服的椅子坐（比如說在網路訂旅館時指定要房間內有 sofa）。假設某個高級人飯店房間內有很舒服的皮椅子，但他們叫它是 armchair（因為是單人座），那怎麼辦？這符合你的要求嗎？即使在同一個語境裡，在跨語言的時候，跟跨領域一樣，甲語言意義在這裡，乙語言的定義在那裡，我們是不是能假設兩個詞意義完全相等，而進行搜尋與比對呢？我們的答案是單詞對單詞的基礎上是不行的，應該建立每個語言共享的知識體系才行。跨語言的知識處理，必須利用兩個共享知識體系的對應才能正確又有效率。

Equivalence Classes vs. Relational Classes

同樣的，當我們做一個對等類（equivalence class）的話，規定它是索引典裡同一類的時候，就規定死了。在這一般的本體架構，世界上的東西有兩種，實體與非實體。實體裡有人造和非人造，非人造物有自然物，裡面有植物、動物等等。類似的架構很多，大家都很熟悉了，也常常發現在分析時，有些東西就是哪裡也丟不進去；或者說在不同的語境裡面，它會有不同的意義，而且屬於不同的幾個類。另外，有時候一組概念或事物有同樣的行為或分佈，又似乎該訂為一類。譬



如下載、拷貝、錄製，下載、拷貝、錄製的對象是什麼？它們在我們的概念裡面，在我們日常生活語言使用的經驗裡面，它是一個類。這類東西都是可以下載、拷貝、錄製的。不管科技如何進步，CD 取代了錄音機，將來有新的機器取代 CD，這一類事物的概念不會因而改變。它當然包括數位檔案、電子檔、機讀檔、程式。但是可不可以下載音樂呢？可不可以下載相片呢？可不可以錄製小說呢？當然可以。在哪一個索引典裡面會告訴你它是同一類？沒有吧！那就完蛋了。因為，這樣所有的東西都是同一類。可不可以下載貝多芬呢？可不可以拷貝李玟呢？可不可以錄製 Stephen King 呢？當然可以啊！所以產生一個問題了，絕大部分索引典裡的 equivalence classes，定義固定一個類的東西，在實際使用時都被打破了。語言裡面產生的一個因為事件關係而定義的類，在概念裡面使用的時候，說它是同一類，並不需要另外去描述它是同一類。因為你講的是一個有知識內容 (content) 的東西。因此，在下载音樂的時候，對音樂不是一個抽象的概念，是下載某一段實際的音樂，是下載某一個相片，是下載貝多芬的某一作品，一個交響曲，不是下載貝多芬這個人。那我怎麼知道不是下載貝多芬這個人呢？因為，我知道我下載的對象。這裡有一個關係存在，這個動作所針對的關係，是自然的、有內容的類。所以，怎麼樣從表相得到本體？語言是個表相，但它的表相是因為很多知識本體中概念與關係體現出來了，才長成這樣的。我們常使用的一個例子，書、書本，這類東西最好玩了，你說書本這個文章很有趣，是講內容；講它很長，三萬多字，不是內容，是用字來表達；這書是

綠色的，綠色的是指它的外觀；這個書是精裝的，是指它的裝訂方式；這本書花了我十年的時間，是指我寫作的過程。請問書是一個概念，還是以上所提的幾個概念？知識是一個本體，表相上面如果要区分，它就是一個或多個。所以，有時候不要迷於相，有時候不要迷於實質，因為根據你的理論，給你實質的定義，很可能讓你錯過本來相是一致的這個事實。因為書就是這麼一個東西，它是這麼多概念組合，因為書的定義對我們的知識上，就是有內容、有字、有印刷方式等等之類的。它是一本電子書的時候，它就是沒有紙了，它就是一個電子檔案，但是它是不是失去了書的本質呢？對我們來講，對使用語言的人來說，它的本質是存在的，它還是書。所以，在這個語言裡，我們把這一組概念的結合看成一個單位。從人與語言使用的觀點出發，讓我們知道這些在某個知識分類上跨類，因而似乎必須區分的東西，在我們的知識架構中結合成了一個單一的概念單位。是語言告訴我的，是表相告訴我，說它是一組概念，而不是一群散開的不同概念，它又是實體，又是電子檔案，又是字本等等之類的。類似的，當我提到可以下載、拷貝、錄製的一組事物，根據我們的語言使用與經驗，可以預測大概也可以閱讀、欣賞、買之類的。這樣的事件行為與經驗，又在知識架構上產生了一個新的概念群組。可以這樣產生的類似乎無窮。但是，在知識本體上，其實有很簡潔的處理方法。也就是說，這些事件 (event) 的分類是必須且無可避免的。但因參與事件相同而群聚的事物，則不必另立一類。他們可藉著參與事件的關係而被定義。也就是說，他們的類似關係，是在參與同一類事件的



條件下產生的，而不是概念上的必然。剛提到最重要，在詞彙網路裡面，我們所看到的，不只是同義與反義的關係，很重要的一些事件的關係性的東西指的就是這個。事件關係的定義，使知識本體上的概念關係簡潔化。

語義關係

詞彙網路的特色，也是最重要的標準，是在每個同義詞集（即概念）上標記一組經過嚴謹定義的詞義關係。這些詞義關係的連結，構成了以詞彙為節點的網路。以下先列舉語意學理論中最典型的九種詞義關係，及其標準中譯，並加上非典型但對知識表徵非常重要的涉入（Involved）關係加以解釋：

antonymy	反義關係
hypernymy	上位關係
hyponymy	下位關係
holonymy	整體—部分關係
meronymy	部分—整體關係
metonymy	轉指關係
near-synonymy	近義關係
synonymy	同義關係
troponymy	方式關係

同義／反義關係 Synonymy/Antonymy

同義／反義關係。反義詞，同義詞。同義詞當然大家已相當熟悉，不另多說。反義詞在語意關係上的地位，倒是值得討論。一般人通常以為反義詞是意義相差距離最遠的兩個詞。其實，從語言學的觀點來講，反義詞是語意距離最近的兩個詞。打個比方，通常與最親密的人吵架是最激烈的；比如巴勒斯坦和以色列打得你死我活，他

們同是閃族人（Semite），民族血緣其實是最近的。共同的生活環境與資源需求造成利益的競爭與衝突；在需求相同後造成競爭，特性上些微的不同便造成對立。在詞語中，反義詞根據語意學的定義是，兩個詞除了對比的語意特徵外，其他所有語意特徵完全一樣。也就是說，反義詞的功能是在某個概念關係上的對比，而要建立這個對比，其他（語意）條件必須相同。另一個說法，對這個關係也許有幫助的，就是我們常說的「蘋果怎麼跟橘子比？」反義的對比，是建立在其他語意特質相同的條件上。

以下以幾個例子說明討論。比如買跟賣是一對反義詞，所牽涉的動作一樣，只是方向不一樣而已。胖與瘦，講體重，一個體重的數量大，一個數量小，只有一個特質不一樣。所以，從觀念來講，比如從搜尋，或其他從知識的處理來講，會不會想到用反義詞來搜尋，是成功的關鍵之一。假設說你要找「紅」的東西，你會想到綠；紅跟綠根據定義是反義詞，因為它只是描述顏色的色譜不一樣而已。因此，當你來搜尋紅的時候，人是會有時候一個詞想不出來，一個概念想不出來，會用類似的概念，會用反義詞。如找叔叔，找不到，會找伯伯、或找姪子，這樣的關係反義詞。但是，通常的搜尋介面裡是不會提供這樣的功能的，因為它誤認為反義詞代表著完全不同的概念，忽視了反義詞才是詞義距離最近的。這是一個反義詞的概念。

上位關係／下位關係 Hypernymy/Hyponymy

具上位關係的詞叫「上位詞」，又稱泛稱詞。如說汽車是一種交通工具（vehicle），這便是一個上位關係。上位關係與下位關係（即 is a



kind of的關係)有對應性,是成雙出現的。這個關係有時候是遞移性的,如吉普車是一種汽車,當然也是一種交通工具。但是有時候是非遞移性的;如佛教是宗教的一種,禪宗又是佛教的一種,那麼禪宗是不是宗教的一種呢?古大部分的知識體系裡,禪宗、天台宗都是佛教的一個宗派,但這個宗派的從屬觀念是只限於佛教的,因此不能往上傳。所以,討論上位、下位的關係,有時候還是要考慮關係本身的特性。也就是說,詞義關係的界定,並不是只有一種關係存在與否那麼單純。因此,我們有時候會把一種關係做更詳細的分類,像以下介紹的整體—部分關係。

整體—部分關係 Holonymy

回到剛剛那一個問題,當一個專家,去尋找一個概念、一樣事物的時候,就跟你自己想東西想不出來的時候,叫不出名字的時候,會用什麼方法去解決問題,逼近要找的對象?事實上就是靠語意關係。譬如你可能要找一個門,一時門講不出來,你講房子(整體詞,因門是房子的一部分),講窗戶(反義詞);利用不同的語意關係概念,想要去接近那個詞,不一定是用同義詞,不一定是用同形詞。人們最常用到的詞義關係之一就是部分—整體關係。房間有門,手有手指頭,字有筆畫,這些都是部分—整體關係。

我們剛提到部分—整體關係,可以做更細的分類。譬如整體跟組成元件 HAS_HOLO_PART、HAS_MERO_PART,專指部分是可拆成元件的。例如:桌腳跟桌子,可以一塊一塊被拆出來的。第二個是集合跟成員(SET-MEMBER)的關係,例如:內政部是內閣的一部分,球員是球隊的一部分;他們都是一個集合裡面的一部分。第三個是整

體與組成成分(PORTION)間的關係,如飯與飯粒,河流/海洋與水滴。組成部分,古英文常是一個不可數名詞。下一個是部分—整體關係牽涉到材料。例如:紙張與書籍,這個也是。組成成分與材料不同的地方在於組成成分可以由整體分割回復(海可以分成無數小水滴);材料無法分割回復(書不能再變回紙)。古地理區域上也有部分—整體關係,比如綠洲是沙漠的一部分。古概念上來講,這是常用的例子。怎麼去定義一個綠洲,而不定義沙漠呢?不能說有水的地方就叫綠洲,有水的地方要在沙漠裡面才叫綠洲。所以,這個詞義關係就變成了詞義的基本定義。

我們之所以會分這些關係,是因為我們回到剛剛提到的,就是這些細分的關係各有不同的邏輯推理關係。古部分—整體關係的基礎上,有些(語意)特性是可以繼承的。比如說桌子是木頭的,那麼可以斷定桌腳也是木頭的;如果桌子是方形的,卻不能斷定桌腳也是方形的。所以,不同的部分—整體關係會影響不同的概念繼承與邏輯推理。這帶出了詞義關係的標幟的另一個重要知識處理功能,就是定義各種可能的概念繼承與邏輯推理。再以組成成分關係為例:若說飯很好吃,飯粒當然也很好吃;但說這碗飯煮得很鬆,卻不能斷定飯粒很鬆吧!原來沒有想這個事情之前,會認為綠洲一定存在沙漠中這些概念屬於常識(real world knowledge)。事實上他們古每個人的詞彙知識裡確實存在,更支持詞彙是語言知識本體的想法。

轉指關係 Metonymy

轉指關係,這是古現代語言裡常用的。譬如「思科(Cisco)宣佈裁員」,這個公司,它是一個無生命的東西,會宣佈嗎?其實是思科某個發言人宣



市的，或者是以我個人經驗而言，說「台北主辦 COLING2002」，台北這個城市會主辦嗎？當然是位於台北裡面的某個成員來辦這個研討會。用一個組織／法人／地區等的名稱來代替實際執行工作的一個人，這就是轉指關係。有時候用大代小，有時候用小代人，是我們常用的一個語言上的東西，它的表相如此。但處理知識時，我們必須知道真正產生這樣關係，真正做這個動作的是誰？

方式關係 Troponymy

方式關係，主要是用在動詞上。在早期的理論裡面，是把它當做上、下位，但是基本上現在都把它稱為「方式關係」。如在物品中，桌子是一種傢俱，是一種上、下位關係。但如煎、炒、炸、燴各是一種煮食的動作，這時候它與上下位關係不一樣的地方在哪裡呢？在語意與概念的對比裡，它主要是方式不一樣。譬如移動 move，移動的方式詞有走、跑、慢跑等等。所有這些動詞移動的事件是一樣的，只是事件進行的方式不一樣。這不是像物品分類的關係，而是對某一種行為的方式（manner）上做更細微的區分。概念上的差異，可能在於速度，可能是工具（如手）擺動的幅度等等之類的。那麼，煎、炒、炸、燴可能是使用的工具不一樣，火的溫度不一樣，這是事件經驗的一部分。在語言表達上，就把同類的事件，用方式區分。

涉入 (Involved)

普林斯頓的語意網路，最早的設計中並無涉入關係；主要原因之一是因為他的設計把不同詞類的詞分開處理，並不允許跨詞類的語意關係。但是，人類的知識系統中並沒有這樣的限制。比

如說，事件與參與者／涉入者的概念關係，並不一定因為他們一個用動詞，一個用名詞表達，而從知識本體中消失。更重要的，當我們考慮到事件與事物這兩類實體的系統，其實很多是要靠涉入的關係來連接，來實現邏輯推理與概念繼承的。因此比較晚開始的歐語詞網中就加入了涉入關係。我們也採取這個較新的架構。

涉入角色，有涉入施事者、涉入經歷者等。譬如廚子跟烹調、做飯。我想個人會有這樣經驗，你想說大廚，大廚這個名詞想不出來，你會說煮飯（的人）。這個事件是一類實體（或人物）的語意關係，在語言裡和在知識本體裡面都有很重要的地位。這個關係是經由如何定義一個廚師而建立的，因為語言裡定義廚師是煮飯、烹飪的人。但在傳統語意架構裡面不來表達這種關係，這是我們比較需要詞彙網路中加進來的。另外，使用什麼工具也是很重要的涉入關係。譬如我們為什麼說犁地呢？犁是一個工具，在犁地的事件中，犁這個工具是最顯著的概念特徵，語言上給我們這樣的證據了。

涉入的工具，有場所、動向……等，就不一一列舉了。詞彙網路用到的其他涉入角色，可參見 EWN 網頁上或「語言座標」網頁上上獻的說明。

語彙網路 (WordNet) 網路資源現況

1990 年普林斯頓 (Princeton) 開始進行做的英語詞網 (<http://www.cogsci.princeton.edu/~wn/>) 是詞彙網路的濫觴。目前已經是 1.7 版了，可以上網去檢索。這是英語單語。另外，歐語詞網 (EuroWorNet, EWN) 是在 1996 年開始做的 (<http://www.hum.uva.nl/~ewn/>)。歐盟意識到如



果要處理知識的話，必須要用到詞彙網路上這樣的架構。然而，國際資訊界的主要研究機構，由於是企業（如 IBM, AT&T, Microsoft）所支持，大概都專供有利可圖的英文。然而，為歐洲的競爭力著想，又不能沒有歐洲各語言的詞彙網路資源。所以，歐盟組織規劃並補助了多種歐洲語言詞彙的建構。而且把多語的資料處理分到相關的各國，技術也是共同開發。當然，與英語的連結不可少。所以，歐語詞彙與普林斯頓的英語詞彙網路是相容的。歐語詞彙包括巴斯克語（Basque）、卡達蘭語（Catalan）、捷克語（Czech）、荷語（Dutch）、法語（French）、英語（English）、愛沙尼亞語（Estonian）、德語（German）、義大利語（Italian）、西班牙語（Spanish）、瑞典語（Swedish）、挪威語（Norway）、丹麥語（Danish）、希臘語（Greek）、葡萄牙語（Portuguese）、羅馬尼亞語（Romanian）、立陶宛語（Lithuan）、俄羅斯語（Russian）、保加利亞語（Bulgarian）、斯洛文尼亞語（Slovene）等等。整個歐洲語言大部分的詞彙網路已經做好了。因此，我們可以這樣說：西方語言大部分的詞彙網路已經做好了。

在中心方面，也已有了一個類似詞彙網路的資源，叫做「知網」（HowNet, <http://www.keenage.com>）。由大陸的董振東先生於 1995 年自力著手進行。它是中英／英中的一個雙語詞彙網路。早期版是開放不用收費的。2002 起新版改由中國科學院軟件所管理後，就需要付費使用了。

「知網」作法的特色是獨樹一幟；不採用英文詞彙網路的架構，只採他自建的架構。而且他先把世界知識本體做個定義，在這定義裡再

去做區分。這個由上而下的方法，與英語與歐語詞彙網路由下而上的方法不同，當然有其可取之處。可惜的是，由於當年資源與訊息的限制，董振東教授與他的兒子董強，基本上是憑著信念與熱誠完成「知網」的，過程中絕少外界的奧援，也並未與世界相關研究接軌。他跟他兒子花了約有七、八年的功夫來做這個事。但是，基本上跟其他語言的詞彙網路連結，並無架構上的基礎，而其上層知識分類，也是兩人的自由心證，不能說錯，卻也缺乏理論的基礎，面臨一些與其他系統互通性（inter-operability）的問題。

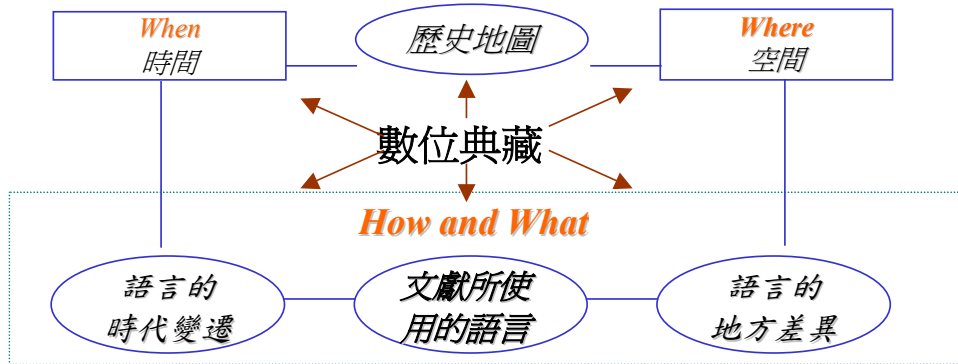
因此，現在中央研究院由 2000 年開始做，做法是先做英中的雙語詞彙網路。前面的報告說英文的詞彙網路把歐語的詞彙網路跟普林斯頓詞彙網路結合起來，大概有十萬個概念的單位，我們把十萬個概念的單位先一個個的看，找出最好（兼顧概念表達與語言使用）的中心翻譯。這個找出最好的中心翻譯，最重要的一點是雙語對譯不見得可找到兩個表達概念完全相同的詞。剛才給的這例子：table 跟 desk 這兩個英語，可能最直接的翻譯就是桌子。但事實上它不是一個同義詞。桌子是 desk 的上位詞，因為桌子包括了 desk 跟 table 這兩類。所以，好的翻譯有時候是有上下位等的關係，必須要把語意的關係表達出來。又如 orange 翻譯為橘子，其實橘子不只是 orange，英語中還有 tangerine 等等之類都是橘子。我們不能說橘子不是好的翻譯，卻也不能說橘子這概念／知識內容上等同於 orange。當做一個雙語的知識轉換時，必須要知道對譯詞間的語意關係。所以，我們把十萬個 WordNet 的詞做出來，現在正在標註，把語意關係建立起來。也就是說，我們把雙



語的對應關係，當成語意關係處理。部分資料可參見我們的網址 <http://corpus.ling.sinica.edu.tw/>

project/LanguageArchive。

語言座標——參考資源建置與服務概念圖



結語：歡迎參加「語言座標」的建置工作

以上的圖是「語言座標」這個計劃的概念與關係圖。這篇文章中談到的研究，將在「語言座標」計劃中實現。這個計劃是古謝孝師負責的「數位典藏國家型科技計劃」中的一個分支計劃。基本的想法是，數位典藏包羅萬象；但是要去解釋典藏內容知識的時候，卻一定要知道原始典藏的語言、領域，時空變遷都會影響內容的詮釋。因此，我們必須在時間、空間上面，根據語言來做一個標識，針對某個主題、某個領域使用它的內容是什麼。這就是語言座標。

語言座標是要描述任何典藏內容的 how 與 what，因為要知道知識是怎麼來的；並輔以時空的 when 與 where，語言座標協助典藏內容知識的投射與解讀。針對每個時代、每個主題、每個典藏，先建不同的詞彙庫，在不同的詞彙庫上建不

同的詞彙網路，這就是 Ontology 的基礎。就是在語意網路的未來發展中，數位典藏必須要經知識本體 (Ontology) 的媒介才能為其所用。語言座標提供了由典藏中抽取本體知識，以及在不同知識本體中建立概念轉換的必要基礎架構。如果慢慢建起來了，也就是在建典藏的過程，也建立了完整的知識架構。這將是一個浩大的知識工程，需要大家共襄盛舉！

後記

語言座標的「領域與知識本體雙語檢索介面雛型」剛在 2002 年底完成：<http://ckip.iis.sinica.edu.tw/CKIP/ontology/>，除了可以檢索中英對譯外，更可以檢索 SUMO 通用上層知識本體，英語詞網的詞義關係，及部分領域標記。使用者若能回饋新的領域標記，或其他資料，並將有禮品相贈，歡迎使用。

【編者按】本文為錄音謄稿，經講者撥冗審閱刊載。

