

中央研究院中英雙語知識本體詞網 (Sinica BOW)： 結合詞網，知識本體，與領域標記的詞彙知識庫

張如瑩 黃居仁

中央研究院語言學研究所

ruyng@gate.sinica.edu.tw churen@gate.sinica.edu.tw

Abstract. 論文中介紹了由中研院語言所和資訊所從語言工程之角度建置的中央研究院中英雙語知識本體詞網 (The Academia Sinica Bilingual Ontological Wordnet, Sinica BOW) [1]，希冀能滿足電腦更多需求以處理更多變的問題。它以WordNet[2]為基礎，加上台灣地區所使用的中文，搭配領域以及SUMO (Suggested Upper Merged Ontology, 建議上層共用知識本體)，並以WordNet1.6版offset為延伸的識別碼作為媒介連結了領域詞彙庫和領域知識本體。該系統為了便於使用者查詢，藉由多元、友善的介面，將功能切割為詞網、知識本體以及索引三個主要單元，提供跨語言資訊轉換、詞義的區分與詞義關係的連結、語言資訊與概念架構(知識本體)的連結以及使用領域等訊息。文中並提到，隨著WordNet版本的變更，為保留語言資訊演變，便以版本的比對，我們以WordNet所提供的1.6版和1.7.1版之名、動詞的單詞義和多詞義對應資料為基礎，輔以Sinica BOW現有資訊，將資料分為單詞義、多詞義對應資料以及1.7.1版新增詞形資料，透過同義詞集、詞形以及詞類訊息，經進一步自動和人工處理後，將Sinica BOW和WordNet1.7.1版結合，並提供版本對應與比對，建立之後Sinica BOW與WordNet各版本結合的處理模式。

1 簡介

知識經濟時代的來臨，全球資訊網的誕生，擴大人們接收資訊的觸角，於是乎能否掌握知識往往成爲致勝關鍵，以關鍵詞檢索已無法滿足使用者的需求，面對此現象，全球資訊網的提議者 Tim Banners-Lee 提出語意網 (Semantic Web) 願景，希冀電腦能真正理解人們的需求，爲達成此願景他一並提出以XML (Extensible Markup Language)、RDF (Resource Description Framework) 加上URI (Uniform Resource Identifier) 以及知識本體 (Ontology) 解決知識呈現[3]。除此之外，更有多位語言學專家提出其所面臨的挑戰應必須包括處理表達知識的基礎—語言。而語言工程的處理訊息應囊括語意，甚至概念上的處理，這是全球皆有的共識。然而其中，中文是台灣本土語言，相關訊息的研究我們必然責無旁貸，而英文爲世界共通的語言，是達成國際接軌的必經途徑，中英文語言訊息的處理和銜接是必要的任務。

語言學界、計算語言學界爲達語意網的願景作了不少努力，中央研究院中英雙語知識本體詞網 (The Academia Sinica Bilingual Ontological Wordnet, Sinica BOW) 便則以WordNet爲基礎，企圖建立中英雙語的基礎知識架構。然而隨著語言變遷，WordNet不斷進行版本修正，我們希望藉由Sinica BOW與WordNet各版本的結合，建置並保留完整中英雙語的語言訊息。這篇論文簡介Sinica BOW以及它與WordNet1.7.1版結合的模式。在第二節中我們將針對Sinica BOW製作動機、系統主要使用資源與架構以及現有系統開放的功能進行簡介。第三節描述如何運用WordNet提供的訊息及系統現有的資訊試著將Sinica BOW與WordNet1.7.1版結合。因與WordNet1.7.1版本的結合，Sinica BOW而新增的功能則在第四節中陳述。最後，則是結論以及未來工作。

2 中央研究院中英雙語知識本體詞網 (Sinica BOW)

Sinica BOW是由中央研究院語言所(文獻語料庫)和資訊所(詞庫小組)合作建置,企圖從語言工程的角度,以台灣地區的語言使用為經驗基礎,整合語言和語言,語言和概念以及語言和領域的資訊,甚至是跨語言間的訊息。近期目標為將以建立完整精確的中英對譯資料庫及檢索介面,作為數位典藏知識國際化的基礎;並逐步建立各領域之雙語領域辭典,以作為該領域/典藏雙語控制詞彙的參考標準,及具領域判斷能力資訊檢索之依據;建立帶領域標記之雙語辭典及檢索介面,以加值成為知識加值雙語電子辭典。未來則希望建立精確的領域知識架構,以作為高加值知識產業的基礎;建立完整的知識本體架構,做為下一代網路(如「語意網」)之語意骨幹以及建立以知識為經緯的中英雙語訊息交換平台,作為多語知識處理的憑藉為目標。相信藉由它提供的訊息將有助於含領域專門知識加上語意資訊之領域詞彙庫的建立、人工或自動翻譯、語言學習、自然語言處理、異質性系統間資料之交換與處理以及自動推論。[4]

2.1 主要資源和架構

Sinica BOW主要使用的資源包含WordNet、ECTEC (English- Chinese Translation Equivalents Database) 以及SUMO (Suggested Upper Merged Ontology, 建議上層共用知識本體)。

「WordNet是結合辭典和知識本體的文字百科全書」在自然語言處理、資訊檢索等的相關研究中常見其身影。[5]1985年普林斯頓大學認知科學實驗室以現代心理語言學理論所述的人類詞彙記憶為啟發,作出語意式電子字典— WordNet (<http://www.cogsci.princeton.edu/~wn/>),他以每個同義詞集表達一種詞彙概念,將其區分為四種英文詞類:名詞、動詞、形容詞、副詞,並以二十幾種詞義關係組織同義詞集。[2]因WordNet的出現,使電腦擁有更豐富的信息可處理各式問題。

由中研院資訊所詞庫小組所建構的ECTEC是以WordNet為基礎,經由現有英中或中英電子辭典的詞形對應,替每個同義詞集的詞義找出可能相對應的中譯詞組,再經由人工檢驗。尋找對譯的過程中,盡可能的以詞彙而非描述性短語表達,目的在於讓每個同義詞集都有最適當的一至三個左右的中文對譯。其中有5%特殊領域詞彙無法在現有電子資源中找到,翻譯者也無法填入適當的中譯,藉於此,我們又花了兩年的時間參考特殊領域辭典將其完成。[6]

SUMO則是由IEEE標準上層知識本體工作小組所建置,其目的在於促使自然語言處理、資訊檢索、自動推論以及資料互通性等工作的進行。知識本體類似於字典或詞彙表,但訊息更豐富,以便於電腦處理其內容。知識本體以格式化的方式表達概念(Concept)、關係(relation)以及公理(axioms)。上層知識本體是將一般性、後設性(meta)、摘要性以及哲學類的概念指出,所以特殊領域的概念可由其中的概念所涵蓋,但特殊領域概念的知識本體則期許由各領域自行制訂。[7][8] (Niles and Pease, 2001) 依據SUMO2002年版資料,我們將系統介面以及概念節點進行中文化,其涵蓋11大類的概念,每大類又分為二至五個類別,總共囊括3,912個概念。日前SUMO已經與WordNet1.6以及2.0版本結合,且以同義(synonymy)、上位(hypernym)、體例(instantiation)這三種類別顯示同義詞集和SUMO概念間的對應關係,例如:同義詞集cell(細胞)與細胞概念(cell)是同義。hockey(曲棍球)屬於運動概念(sport),兩者間的關係為上位,也就是說運動涵蓋hockey(曲棍球)。China(中國大陸)屬於國家(nation)這概念的體例。[9]

除此,我們以「中國圖書分類法」為基準,並參考各知識分類與實際研究經驗,提出:包含九大類的知識分類(Knowledge Content),涵蓋427個領域。另外,並因應語言資源特性加入下列語言使用(Language Usage)的各類訊息:專名(說明文字符號的指涉)(Proper Name)、語體(說明文字符號的使用)(Genre/Strata)、各種語言/詞源(Language/Etymology)、各國地名(Country Name)。領域階層的建立在於替不同詞義中的詞彙項目區別其使用的領域,例如:stock作「股票」和「家畜」兩個不同解釋時,分屬於財政學裡的資本以及動物學的脊椎動物學。加註領域信息可降低詞彙歧異性,增加資料交換時的互通性,輔助領域詞彙庫之建構等。

Sinica BOW透過WordNet1.6 offset延伸所產生的識別碼作為媒介,進行串連,將每個資源以及各類訊息連結。因WordNet1.6 offset延伸的識別碼可獲得原本WordNet存在的詞類、解釋、英文例句、同義詞集、各同義詞集間的詞義關係及其所屬詞彙。而SUMO概念與WordNet的連結,使得可透過該識別碼獲取詞義與概念搭配的訊息。以WordNet為基礎所建置的ECTED與針對WordNet同義詞集的各詞彙項目所給予的領域值,也是透過該識別碼獲取。如果是特殊領域詞彙庫,加上相對應的Sinica BOW識別碼,也可保留原始

資源的資料庫格式和WordNet連結。又，領域知識本體則是在SUMO某些概念下進行延伸發展。每個特殊領域詞彙庫中的詞彙一樣具有所屬的概念，其所屬概念可能是SUMO或特殊領域知識本體的某一概念，特殊領域詞彙庫和領域知識本體的結合，使得透過該識別碼又串起所有的訊息。Sinica BOW的資源和架構如圖 1所示。由於透過WordNet可以和同是以WordNet為基礎架構所建置的其他語系WordNet資源加以連結，例如：EuroWordNet[10]，以此作基礎架構可編製成多語的詞彙網路，成為多語環境中所需之語言知識結構的基礎資料。

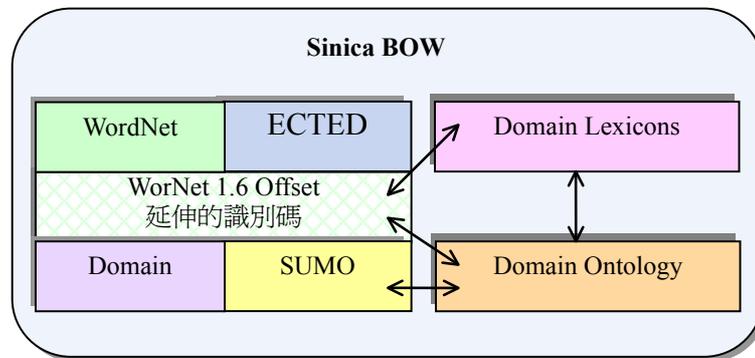


圖 1. Sinica BOW主要資源及架構[11]

2.2 現有功能

Sinica BOW的主要資源和架構中，使用者可獲得跨語言資訊轉換、詞義區分與詞義關係連結、語言資訊與概念架構（知識本體）連結以及使用領域的訊息。為了讓使用者透過友善介面很容易地可擷取所需訊息，系統將介面切割為詞網、知識本體以及索引。

1. 詞網

又分成「中文查詢」、「英文查詢」以及「專門領域」三個單元。「中文查詢」和「英文查詢」中使用者可輸入中或英文詞形，系統將顯示詞形搭配詞類在各資源中的情況，包括出現與否以及分部頻率，所比對的資源是以具有代表性且容易獲取電子檔者為考量。點選其他各資源的超連結都將連結該資源原始或相關網頁進行進階查詢。但點選「WordNet1.6英中對譯」將出現該詞形所有的詞義，每個詞義以表格呈現該詞義下的所有訊息，包括所屬的領域、詞類、解釋、翻譯、同義詞集、各詞義關係詞、SUMO概念以及英文例句。使用者可再點選任一詞彙項目進行再查詢。

目前的領域訊息主要有三個來源1.藉由領域階層的中英文值找出詞形相符之中譯或同義詞集的詞彙項目，再透過WordNet本身同義和上位詞義關係，假設獲得之同義詞集和上位詞屬於該領域。[12]2.透過遠見科技提供的電子辭典，針對其中擁有領域的詞形推論至WordNet[13]，以及3.線上「使用者建議領域」功能所得之訊息。綜合上述三項結果，目前針對同義詞集之中譯詞彙總共有18,562筆資料有領域值，針對同義詞集之詞彙項目總共有24,229筆資料有領域值，這相當於有14,396筆同義詞集具領域訊息。

「專門領域」可查詢其他領域詞彙庫與WordNet連結的訊息。除了繼承領域詞彙庫中原始的領域訊息，其他則保留WordNet和中譯等訊息。目前可在線上查到中研院資訊所詞庫小組財經辭典和WordNet連結的資訊。

2. 知識本體

由「SUMO」和「領域知識本體」這兩單元組成。都可以以中、英詞形或概念查詢，並以樹狀結構呈現各概念間的關係。提供的訊息包括以詞形查詢所屬概念、概念的定義、公理、詞彙和概念連接之關係以及相連結的WordNet之所有訊息，其中概念訊息皆已中文化。

此外，我們曾嘗試建置「唐詩三百首知識本體」和「魚類知識本體」兩個主題的領域知識本體。我們試著從唐詩三百首所使用的「植物」、「動物」、「人造物」主題詞彙，利用WordNet詞義和詞義關係等訊息的驗證並加以連結，在SUMO的基礎架構下進而建置唐詩三百首領域知識本體。系統介面可顯示加入WordNet語意訊息後知識本體以及詞彙分佈的差異，唐代的知識架構因此可被驗證，例如：有袋類動物並不出現在唐代，有關鳥類或有翅膀的昆蟲詞彙則廣泛分布於詩中，這正符合唐代時興的飛天觀念。[14]

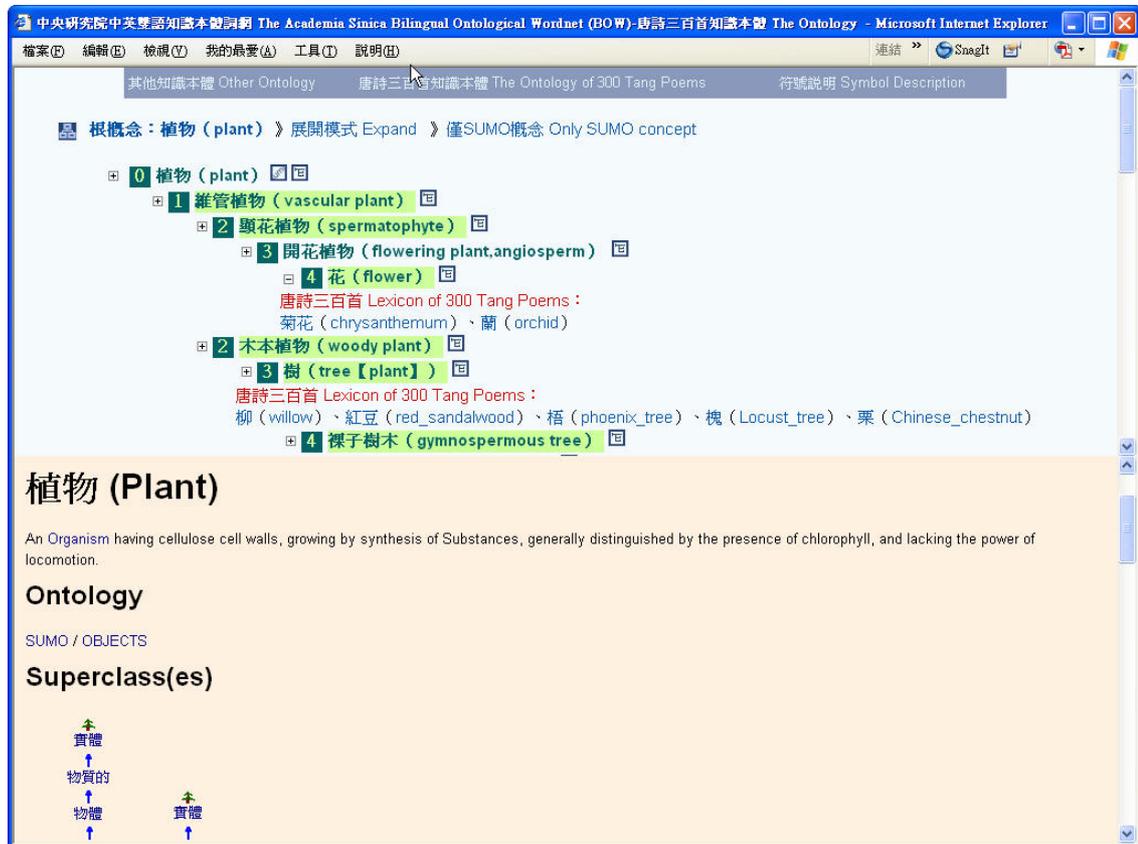


圖 2. 經WordNet語意驗證的植物類唐詩三百首知識本體

另外，跟「臺灣動物相之典藏－臺灣魚類資料庫」計畫（網址：<http://fishdb.sinica.edu.tw/>）的合作，蒐集領域相關後設資料（metadata）和分類系統，並以WordNet輔助分析上下位關係，最後，以該領域的所有詞彙觀察知識本體是否完整，而完成了魚類知識本體的初步結果。

3. 索引

索引窗口的建立在於讓使用者以字母、字首字尾、詞類、來源、頻率、領域、概念以及綜合（來源、詞類、頻率）的限制條件從事進一步資訊查詢。系統將列出符合條件的詞彙列表，使用者可再進一步查詢對應SUMO、WordNet以及在各資源分佈情況。使用者可從符合條件的詞表中，進一步分析所具有之共同特性和差異。

接下來，我們將針對Sinica BOW與WordNet1.7.1版結合之步驟與提供的新功能進行簡介。

3 Sinica BOW 結合WordNet1.7.1資料

語言會隨著時空的變化而有所變更，WordNet也隨其變更不斷更新、修正其資料。直至2004年止最新推出授權的為2.0版本。Sinica BOW針對1.7.1版本進行先導測試性實驗，試圖作為之後Sinica BOW與各版WordNet資料結合時處理模式的根據。我們企圖保留、追蹤並呈現其中的演變，以豐富語言訊息，且這對於作進一步的語言變遷研究有十分的助益。

3.1 WordNet1.7.1以及和WordNet1.6對應原始資料

WordNet1.7.1版本於2001年公布，WordNet 1.7.1詞義順序和1.6版本並不相同，原本詞義第一順位的並不見得仍保持為第一順位。有很多的新詞形和新詞義，因此詞義重新整理過，1.6版中被察覺到的重複詞義已經被刪除。1.7.1版本無法和1.6版的同義詞集間之差異進行直接對應，針對此問題WordNet提供兩版本間名詞和動詞的對應，其中名詞和動詞又細分為單詞義的 (Monosemous) 和多詞義 (Polysemous) 的對應。對應準則和方法為：

1. 1.6或1.7.1版資料庫中唯一的名詞或動詞則忽略。
2. 兩版本資料庫中，若為名詞和動詞的單詞義之sense_key (同義詞集中的某一詞彙項目)和synset_offset則直接對應。論文中將分別以1.6to1.7.1.noun.mono 以及1.6to1.7.1.verb.mono表示這類對應的結果。
3. 1.6版和1.7.1版兩個資料庫中若為名詞和動詞的多詞義，WordNet則以不同的啓發方式去評估1.6和1.7.1版本之間詞義的相似性，並針對每個比對結果給予分數。針對每個詞彙 (word)，每個1.6版詞義將和1.7.1版詞義的相同詞彙進行比對，1.7.1版詞義(s)有比較高分數的將被視為最佳的對應結果。文中將這些詞義對應的結果以1.6to1.7.1.noun.poly 以及1.6to1.7.1.verb.poly示之。WordNet在進行對應時採用的啓發方式包括：sense keys比較、同義詞集詞彙項目 (term) 相似度以及相關樹位置 (relative tree location) (也就是上位詞比較)。比對時，註解 (Glosses) 的資訊不被參考，因為她們常更新。對應分數從0到100，表示經由啓發對應之可信度，分數越高可信度越高。90或100是最主要的詞義對應。其中，96%的名詞對應分數高過90，94%的動詞對應分數高過90。分數和資料分布如下：

表 1. WordNet提供1.6和1.7.1版多詞義名詞與動詞對應的分數和資料分布情況

1.6to1.7.1.noun.poly				1.6to1.7.1.verb.poly			
分數	筆數	1.6 同義詞集數	1.7.1 同義詞集數	分數	筆數	1.6 同義詞集數	1.7.1 同義詞集數
100	30,072	22,757	22,757	100	12,690	7,905	7,905
90	4,605	3,130	3,143	90	3,167	1,793	1,796
80	412	291	297	80	321	187	189
70	230	190	187	70	149	121	120
60	145	100	103	60	79	48	50
50	19	16	16	50	41	39	39
40	43	38	38	40	47	41	42
30	48	45	44	30	25	24	23
20	135	120	122	20	120	96	102
0	32	29	0	0	46	41	0
總和	35,741	26,560	26,484	總和	16,685	10,037	9,969

表 2. WordNet提供1.6和1.7.1版單詞義名詞與動詞對應的資料分布情況

資料類型	項目	詞數	1.6.同義詞集數	1.7.1同義詞集數
1.6to1.7.1.noun.mono		80,163	49,870	49,946
1.6to1.7.1.verb.mono		5,278	3,933	3,936

WordNet1.6版有116,317名詞詞義，其中115,904可對應到1.7.1版WordNet。殘餘WordNet1.6表達唯一詞義的413名詞詞義。在1.6to1.7.1.noun.poly中被對應的多詞義名詞全部有35,741詞義，在1.6to1.7.1.noun.mono中對應的單詞義名詞有80,163個。

WordNet1.6版有24,169動詞詞義，其中21,963可對應到1.7.1版WordNet。殘餘WordNet1.6表達唯一詞義的2,206動詞詞義。在1.6to1.7.1.verb.poly中被對應的多詞義動詞全部有16,685詞義，在1.6to1.7.1.verb.mono檔案中對應的單詞義動詞有5,278個。（WordNet, 2001）

表 3. WordNet提供1.6和1.7.1版名詞與動詞對應的資料分布情況

資料類型	項目	對應1.7.1總詞義／1.6原詞義	1.6.同義詞集數	1.7.1同義詞集數
1.6to1.7.1.noun		116,317 / 115,904	65,946	66,039
1.6to1.7.1.verb		24,169 / 21,963	12,113	12,065
1.6to1.7.1.noun+verb		140,486 / 137,867	77,987	78,041

3.2 進一步處理WordNet提供的WordNet1.7.1和1.6版對應訊息

根據WordNet提供的資料特性，為了便於與Sinica BOW資料的結合，我們分別針對WordNet提供的「單詞義」以及「多詞義」資料透過暨有的1.6版本中譯資訊的輔助作進一步的處理。又，因1.7.1版新增詞形並不會有比對訊息，當然更無法從WordNet1.6暨有的中譯訊息中獲得任何輔助，因此1.7.1版中新增的英文詞形又另作處理。這三類資料的個別處理方式與結果如下：

1. WordNet 1.6版中對應到WordNet 1.7.1版的「單詞義」資料
不論名詞或動詞，假設直接繼承1.6版本的翻譯，且不再進行人工檢查。
2. WordNet 1.6版中對應到WordNet 1.7.1版的「多詞義」資料
 - (1) 不論名詞或動詞，由於WordNet對應多詞義資料的方式是以兩版同義詞集中的某一詞彙項目進行比對，並給與對應分數，若對應分數為100分，則假設1.7.1版本中對應的該同義詞集下的所有中譯詞組，並不需要變更，直接繼承1.6版相對應的翻譯。
 - (2) 其他對應分數小於100的多詞義資料，則保留原WordNet提供的對應模式，也就是以兩版同義詞集中的某一詞彙項目比對，我們並提供1.6版本的中譯，再針對1.7.1版本中該詞彙項目是否仍適用於自動對應到之1.6版中譯詞組進行人工檢視，若否，則變更為校正值。

表 4. WordNet提供其對應分數小於100的多義詞人工校對結果

資料類型	項目	總筆數	1.6沒中譯	1.7.1有校正	結合後有中譯	結合後沒中譯
1.6to1.7.1.noun.poly.70-20		1,064	3	109	1,064	0
1.6to1.7.1.noun.poly.80		412	0	9	412	0
1.6to1.7.1.noun.poly.90		4,605	0	53	4,605	0
1.6to1.7.1.verb.poly.70-20		461	7	61	459	2
1.6to1.7.1.verb.poly.80		321	2	8	319	2
1.6to1.7.1.verb.poly.90		3,167	18	48	3,158	9

3. WordNet 1.7.1新英文詞形

我們依據1.6版本將1.7.1版本中新增的英文詞形另外找到1.7.1版本中相對應的同義詞集，其中新增了18,969英文詞形，若去掉大小寫因素則新增18,461英文詞形，共佔13,708個1.7.1版的同義詞集。對於這類相對於1.6版本而新增的英文詞形，我們假設新英文詞形在1.6版本中找不到對應的同義詞集，於是我們直接進行人工檢驗，由人工新增相對應的中譯詞組。人工翻譯的結果，找不到中譯的同義詞集仍有17筆，多為特殊專有名詞（名詞14筆，形容詞3筆）。

3.3 綜合處理對應訊息的結果

接著我們試著將上述步驟所產生的資料綜合，其中若1.7.1版本的中譯內容沒有校正值，則假設繼承1.6版本中相對應的同義詞集之中譯詞組。所有單詞義或多詞義之名詞或動詞的對應資料類型中，只要針對同一個同義詞集進行對譯，不論針對該同義詞集下哪一個英文詞彙項目的中文對譯，都視為該同義詞集的中譯詞組的成員。

表 5. 經自動及人工檢驗後的WordNet提供兩版本對應對應資料以及1.7.1版新增詞形的中譯詞組統計

資料類型	項目	有中譯的同義詞集筆數	有中譯的同義詞集百分比	產生的中譯詞形	產生的 WN1.6 對應 WN1.7.1 數
1.6to1.7.1.noun.mono		459,94	41.3530%	55,922	80,163
1.6to1.7.1.noun.poly.70-0		500	0.4495%	745	620
1.6to1.7.1.noun.poly.80		298	0.2679%	473	412
1.6to1.7.1.noun.poly.90		3,143	2.8259%	4,636	4,604
1.6to1.7.1.noun.poly.100		22,758	20.4616%	27,131	30,072
1.6to1.7.1.verb.mono		3,904	3.5101%	5,847	5,278
1.6to1.7.1.verb.poly.70-0		368	0.3309%	547	461
1.6to1.7.1.verb.poly.80		190	0.1708%	339	321
1.6to1.7.1.verb.poly.90		1,797	1.6157%	2,688	3,167
1.6to1.7.1.verb.poly.100		7,906	7.1082%	9,036	12,691
1.7.1新增詞形的同義詞集		13,708	12.3248%	17,258	
總和		85,619	76.9796%	96,378	137,789

表 6. WordNet1.6版Sinica BOW資訊和1.7.1版利用WordNet提供對應資料以及新增詞形產生中譯詞組之訊息的比對

項目	版本	WordNet1.6	WordNet 1.7.1
同義詞集數		99,642	111,223
有中譯的同義詞集數		99,642	85,619
沒中譯的同義詞集數		0	25,604
同義詞集的中譯總詞彙數		149,780	128,873
平均每個同義詞集的中譯詞數		1.5032	1.15587
同義詞集的英文詞總數		17,401	195,817
平均每個同義詞集的英文詞數		1.7463	1.7606
平均每個英文對應的中譯數		0.8608	0.6581
中譯總詞形數		109,982	96,378
同義詞集英文總詞形數		122,864	140,488

綜合WordNet提供的兩版對應資料所做的中譯資料校正，以及針對1.7.1版新增詞形所對應的同義詞集進行人工對譯，上述兩大類所產生的中譯資料如上表 5和表 6所示，發現尚有25,604筆同義詞集尚未有任何中譯。除了多詞義資料在人工校對時即無法找到適合對譯的，其他我們假設可能是WordNet1.6和1.7.1 版並未針對形容詞和副詞進行對應。於是我們透過1.6版和1.7.1版本的資料再作第二次的進一步比對處理：

1. 若兩版本的同義詞集完全相等，則中譯詞組直接繼承，但發現並無此類同義詞集。
2. 具有相同詞形，且詞類相同的資料，就給與1.6版本相對應的同義詞集的中譯詞組。這類資料有15,942筆同義詞集，佔尚缺中譯資料的62.2637%，相對於1.6版本的33,791筆同義詞集。
3. 其餘則繼續比對，只要具有相同詞形，一樣給與1.6版本相對應的同義詞集的中譯詞組，找到9,645筆這類同義詞集，佔尚缺中譯資料的37.6699%，相對於1.6版本的48,121筆同義詞集。

5 結論

Sinica BOW以WordNet為基礎，在其資訊上融合中英跨語言資訊轉換，使用者可以中英文的詞形查詢，獲得詞義區分與詞義關係連結、語言資訊與概念架構（知識本體）連結以及所屬領域的資訊。除此，我們在上層共用知識本體SUMO下延伸領域知識本體，並以WordNet1.6版Offset所延伸的識別碼串連上述訊息及各類領域知識本體和領域詞彙庫。並以多樣化友善介面讓使用者從詞形、領域、詞類、概念、出處等角度切入，交錯查詢各類綜合訊息。

論文中並針對WordNet提供的1.6版和1.7.1版單詞義及多詞義的名、動詞對應，以及利用現存於Sinica BOW中WordNet1.6版的各類訊息，輔助進行Sinica BOW與WordNet1.7.1版的結合，並新增各版本資訊查詢、版本間同義詞集以及同義詞集詞彙項目對應以及版本間同義詞集詞彙項目、中譯詞及各詞義關係成員的版本比對功能。希冀以此嘗試作為與WordNet2.0版甚至與其他版本擴增的基礎模式。

未來，除了以此模式繼續融合WordNet其他具有代表性之新版本的資訊，也希望與透過相同架構，以不變更各資源原格式為原則，進一步結合其他資源，甚至是完整的中文詞網，藉以豐富並建構成完整的知識網路，作為電腦多語訊息處理基礎架構。

References

- [1] 中央研究院中英雙語知識本體詞網 The Academia Sinica Bilingual Ontological Wordnet (Sinica BOW), <http://BOW.sinica.edu.tw>
- [2] WordNet, <http://www.cogsci.princeton.edu/~wn/>
- [3] Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001. <<http://www.sciam.com/2001/0501issue/0501berners-lee.html>>
- [4] 黃居仁、張如瑩、蔡柏生（民93）。「語意網時代的網路華語教學：兼介中英雙語知識本體與領域檢索介面」（Chinese Language Education and the Developing Semantic Web: An Introduction to Chinese-English Bilingual Ontology Interface）。在黃一農、張寶塔總編輯、羅鳳珠執行編輯，資訊與社會叢書系列之三：語言文學與資訊科技（卓越計畫）（頁443-467）。新竹市：清華大學出版社出版。
- [5] 張俊盛，「喬治米勒在聖塔菲 WordNet：結合辭典和本體論的文字百科全書」，科學人雜誌（2004年5月），<<http://www.sciam.com.tw/forum/forumshow.asp?FDocNo=444&CL=16>>。
- [6] Chu-Ren Huang, Elanna I. J. Tseng, Dylan B. S. Tsai, and Brian Murphy. Cross-lingual Portability of Semantic relations: Bootstrapping Chinese WordNet with English WordNet Relations. *Language and Linguistics*. 4.3, pp.509-532. 2003.
- [7] Suggested Upper Merged Ontology, <http://www.ontologyportal.org/>
- [8] Niles, I., and Pease, A. "Toward a Standard Upper Ontology". In *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*. Chris Welty and Barry Smith, eds, Ogunquit, Maine, October 17-19, 2001.
- [9] Niles, I., and Pease, A. "Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology". In *proceedings of the IEEE International Conference on Information and Knowledge Engineering. (IKE 2003)*. Las Vegas, Nevada, June 23-26, 2003.
- [10] EuroWordNet: Building a multilingual database with wordnets for several European languages., <http://www.ilc.uva.nl/EuroWordNet/>
- [11] Chu-Ren Huang, Ru-Yng Chang, and Shiang-Bin Lee. Sinica BOW (Bilingual Ontological Wordnet): Integration of Bilingual WordNet and SUMO. 4th International Conference on Language Resources and Evaluation (LREC2004). Lisbon, Portugal. 26-28 May, 2004.
- [12] Chu-Ren Huang, Xiang-Bing Li, and Jia-Fei Hong. Domain Lexico-Taxonomy: An Approach Towards Multi-domain Language Processing. The 1st International Joint Conference on Language Language Processing (IJCNLP-04) Asian Symposium on Natural Language Processing to Overcome Language Barriers. Sanya City, Hainan Island, China. 25-26 March, 2004.
- [13] Echa Chang, Chu-Ren Huang, Sue-Jin Ker, and Chang-Hua Yang. Induction of Classification from Lexicon Expansion: Assigning Domain Tags to WordNet Entries. 19th COLING 2002 Post-Conference Workshop--SEMANT: Building and Using Semantic Networks Processing. Center of Academia

Activities, Academia Sinica. Taipei. Taiwan. September 1, 2002.

- [14] Chu-Ren Huang, Feng-ju Lo, Ru-Yng Chang, and Sueming Chang. Reconstructing the Ontology of the Tang Dynasty: A pilot study of the Shakespearean-garden approach. 4th International Conference on Language Resources and Evaluation (LREC2004) Workshop on Ontologies and Lexical Resources in Distributed Environments (OntoLex 2004). Lisbon. Portugal. 29 May, 2004.
- [15] WordNet: 1.6 to 1.7.1 Sense Mapping Package,
<ftp://ftp.cogsci.princeton.edu/pub/wordnet/1.7.1/WNsnsmap-1.7.1.tar.gz>