中文詞彙語意資料的整合與擷取：詞彙語意學的觀點

高照明
zmgao@ntu.edu.tw
台灣大學外國語文學系

摘要

本文從詞彙語意學理論的觀點整合知網(Hownet) 、現代漢語分類辭典、教育部國語辭典等資源，並利用 Wordnet 和漢英辭典，擷取上述不同來源的中文詞彙語意訊息。我們透過整合後的訊息發展一套系統，使用者輸入兩個詞可以找出兩個詞之間的詞彙語意關係包括（一）同義關係（二）反義關係（三）上下位關係（四）部件與整體關係 （五）相同事件（六） 相同領域(domain) （七）相同語意特徵（八）相同的語意類別（九）事件與語意角色。

關鍵詞：詞彙語意關係、詞彙知識庫、知網(Hownet)、義元、語意特徵、語意角色、事件角色轉換、Wordnet、現代漢語分類辭典、重編國語辭典修訂本、同義詞、反義詞、上位詞、下位詞、全體詞、部分詞

一 前言

詞彙語意學的發展與資訊科學及人工智慧有相當密切的關係。六零年代語言學家 Fillmore (Fillmore 1968)提出語意角色的理論架構格理論(case theory)對於語意學及句法學產生深遠的影響，同一時期 Wilks (Wilks 1968)從人工智能的角度研究語意知識的表達。七零年代 Shank (Shank 1975)提出腳本理論將詞彙知識與常識具體化程序化，作爲自然語言理解的基礎。而 Sowa 等人(Sowa 1984)則從事 conceptual graph 的研究。七零年代末期 John Sinclair (參考 Sinclair 1987) 首創以語料庫及計算機研究詞義和搭配語並編纂辭典（Collins Cobuild English Dictionary）。八零年代，利用機讀辭典研究語法與詞彙語意開始興起，其中最多研究人員使用的資源是 Longman Dictionary of Contemporary English(LDOCE) (參考 Boguraev and Briscoe (eds) 1989)。九零年代隨著英國國家語料庫(http://www.natcorp.ox.ac.uk/)及相關檢索軟體(SARA, Xaira) 的完成，研究人員開始有龐大的語料庫及檢索工具研究詞彙語意。而 Wordnet 計畫(http://wordnet.princeton.edu/)推出(semantic concordancer)以 Wordnet 詞項的意義標示語料庫中的詞的詞義，爲計算詞彙語意學奠定了深厚的基礎。近年來越來越多標注詞彙語意訊息的語料庫出現，如標記論元結構（argument structure）及語意角色訊息的 FrameNet、 VerbNet 、PopNet。計算詞彙語意學研究的重心轉爲利用語料庫及統計演算法，例如 Church 首創以互見訊息(mutual information)和 t-score 來擷取搭配語(參考 Church and Hanks 1990) Church et al. (1991) Church et al. (1994)。Hearst (1992)透過句型擷取上下位詞。Grefefenstette (1994)以語法剖析器和統計擷取同義詞。Jones (2002)透過語料庫擷取反義詞。Turney (2006), Girju 等 (2007) 更進一步以統計及機器學習演算法研究詞彙語意關係，這與傳統透過詞彙知識庫擷取與判定詞彙語意關係的方法大異其趣。以大量語料結合統計或機器學習演算法的優點是不需要詞彙知識庫即可從語料中擷取一些語意關係，缺點是擷取的資料不夠精確與完整必須透過專家來校對與補充。本文的目的在於整合現有的各種中文詞彙知識庫，並利用這些資料庫截長補短

來擷取最多的語意關係，作爲未來評估機器學習演算法擷取詞彙語意關係研究的平台。

詞彙語意關係與語意網(semantic web)及本體論(ontology)息息相關。Tim Berners-Lee (2000)提出語意網的概念描繪了下一代網際網路的遠景。語意網的成功仰賴本體論，也就是必須能清楚的表達通用或某一特定領域知識的詞彙的意義並顯示相關概念之間的語意關係。目前絕大部分的本體論都是透過人工建立及人工標示。如果能夠半自動建立及標記對於語意網的發展將有很大的幫助。建立大規模詞彙語意關係的資料庫可以大幅提升本體論發展的時間及品質。

詞彙語意關係的研究也能夠廣泛應用於資訊檢索與擷取，機器翻譯系統等，例如，資訊擷取時需要辨識文章裡面人、事、時、地、物。檢索時需要利用同義詞，而自然語言理解更必須判斷語意角色，

本研究利用現有的中英文詞彙知識庫發展一個能夠截長補短自動找出中文詞彙語意關係的系統。英文部分我們主要利用 Wordnet 以及漢英辭典，中文方面我們使用知網(Hownet)，現代漢語分類辭典，教育部重編國語辭典修訂本等資源。我們的系統能夠判斷下列的詞彙語意關係（一）同義關係（二）反義關係（三）上下位關係（四）部件與整體關係 （五）相同事件（六） 相同領域(domain) （七）相同語意特徵（八）相同的語意類別（九）事件與語意角色。

二 語意關係的理論

傳統上詞彙之間的關係可以分成組合的關係(syntagmatic relation)及聚合的關係(paradigmatic relation)。組合的關係就是構成詞組的關係(水平的關係)，依據語法依存關係可以分成：修飾與被修飾的關係、主詞與謂語的關係、動詞與受詞的關係。在 Chomsky 的句法理論中，主詞與謂語的關係、動詞與受詞的關係是透過動詞來選擇它的主詞或受詞的語義 (請參考 Pustejovsky 2000 對組合關係的理論)。聚合的關係可以看成在某一種句式裡面代換的關係（垂直的關係），通常是同義詞、反義詞的關係、上下位詞的關係。

魯川(2001)將語意組合關係分成語意關係和語意依附兩種。語意關係分爲並列關係、選擇關係、等同關係、同現關係、加合關係、配合關係、接合關係。而語意的依附關係分成爲事件的依附（包括語氣、話語、與情態），述謂的依附(時態、動貌、語態、程度)，指稱意結的依附（數）。

Koenig (1999)將詞彙的關係概分成兩種，一種是分類的關係(classificatory)，另一類是構詞的關係(morphological)，例如相同詞性的關係就是分類的關係，而 bird 和 birds 之間屬於構詞的關係。Koenig 利用中心語驅動詞組語法(Head-driven Phrase Structure Grammar, HPSG)的架構提出 Type Underspecified Hierarchical Lexicon)解釋詞的癖性(idiosyncrasy)與滋生性(productivity)。

Chaffin and Herrmann (1988)認爲語意關係有兩種理論，網路理論（network theory）及關係元素理論(relation element theory)，處理時前者計算連結兩個概念節點的路徑，而後者計算所表示的元素的異同。他們歸納出五大類共三十一種的語義關係，這五大類分別是對比(contrasts) 、相似(similars) 、類別包含(class inclusion) 、格關係(case relation)（亦稱爲語意角色關係）、部分全體關係(part whole)。

Calzolari (1988)建議應該合併辭典與同義詞辭典成爲一個大的詞彙知識庫，這裡面應該包含下列幾種語意關係：上下階層關係、同義關係、IS-A 關係、論元關係、詞彙場(lexical fields)，搭配語關係、術語庫、衍生詞的關係。

當代語言學理論中對詞彙語意關係發展最詳盡者是 Melcuk、Zholkvsky、Apresyan 等人所提出的 Explanatory Combinatorial Dictionary (ECD) (Melcuk 1988)其中提出數十個詞彙函數

(lexical functions)。以 Magn 這個詞彙函數爲例，它的功能是[intensifier]。輸入一個詞就可以得到這個函數對應的值，如 Magn(to condemn) = strongly。

Byrd (1994)描述 I.B.M.發展的詞彙知識庫 ComLex 及相關工具，可以根據一個詞的詞義找出各種關係，包括拼字相同、上下位詞、同義詞、典型的論元(argument) 、例如：領養 adopt 典型的受詞是別的父母親的孩子(a child of other parents)，及語意選擇限制(selectional restrictions)，例如領養的主詞及受詞都是人(person)。這些關係都由現有的辭典裡面（如 Merriam Webster Dictionary, Longman Dictionary of Contemporary English (LDOCE)的訊息抽取出來。Klavans (1994)也以及 Merriam Webster Dictionary 裡面的定義及同義詞辭典抽取出來相關詞的語意網(semantic net)。

近年有關於詞彙語義表徵最重要的理論架構堪稱 Jackendoff 與 Pustejovsky 的著作。Jackendoff (1983, 1990)提出抽象的概念結構(conceptual structure)用以解釋許多的語言現象。

Pustejovksy (1995)提出衍生詞彙理論(Generative Lexicon)，理論有四層表徵，論元結構、(argument structure) 、事件結構(event structure) 、 屬性結構(qualia structure)、及詞彙繼承結構(lexical inheritance structure)。論元結構註明事件參與者的語義角色.。事件結構(event structure)將事件分成狀態(state)過程(process)過渡(transition)及更小結構。屬性結構(qualia structure)將名詞屬性分成 formal(語意類別) 、 constitutive(與組成這個東西的關係) 、 telic (東西的功用) 、agentive(東西如何產生) 、詞彙繼承結構(lexical inheritance structure)。透過 type coercion、 selective binding、 co-composition 等理論的機制衍生詞彙理論可以有系統解釋許多詞彙語意的現象，例如：enjoy a book 和 enjoy a meal 前者是 enjoy reading a book，後者是 enjoy eating a meal. Jackendoff 和 Pustejovsky 的理論，由於相當抽象，並不容易實做。

中文有關詞彙語意關係的專著除了魯川(2001)的「意合網路」之外，以董政東與董強(Dong and Dong 2006)所發展的知網(Hownet) 最重要。我們將在後面詳細介紹 Hownet。

從上述各種詞彙語意學理論的介紹不難發現語意的類別、語意特徵、同義詞、反義詞、上下位詞、部分與全體關係、語意角色、事件關係，這些是絕大部分詞彙語意學理論和詞彙知識庫都包括的訊息。

三國內外著名詞彙知識庫的介紹

最著名的詞彙知識庫首推普林斯頓大學所發展出 Wordnet (http://wordnet.princeton.edu/)，有大量的自然語言處理研究依靠這個知識庫。透過 Wordnet 可以查出詞的定義，例句，同義詞，上位關係詞，下位關係詞，部分關係詞，全體關係詞等。Euro Wordnet 是歐盟幾個國家以 Wordnet 爲基礎合作發展多語詞彙知識庫。

美國南加大整合許多的知識庫和發展了 Ontosaurus (http://www.isi.edu/isd/ontosaurus.html) 如同 Wordnet 它註明了詞項的定義，同義詞，及語意的類別。

Sumo Search Tool (http://sigma.ontologyportal.org:4010/sigma/WordNet.jsp)這個工具將 Wordnet 的 sense 直接對應到 IEEE Suggested Merged Upper Ontology (SUMO)。優點是 Wordnet 的詞義類別更清楚的表示出來，例如 buffalo 這個詞對應到 SUMO 有 buffalo 水牛，city 城，meat 肉。透過 SUMO ontology 我們可以得到許多詞彙之間的關係，例如 SUMO 裡面包括 subclass 和 disjoint 這類的關係，在 SUMO 裡面有諸如 Buffalo is a subclass of hoofed mammal Buffalo is disjoint from DomesticAnimal 這類對於自然語言理解和推論非常重要的知識。DOLCE 與 SUMO 類似，也是一個非常重要的 ontology，並提供與 Wordnet 的對應。

卡內基美侖大學利用 Wordnet 及其它包括百科全書在內的許多電子資源，發展 Lexical Freenet (http://www.cinfn.com/doc/)。這是到目前爲止我們所知最最完整的詞彙關係資料庫。能夠找出的詞彙關係遠遠超過 Wordnet。例如，我們輸入 Taipei, Taiwan 兩個詞得到一些相當

有趣的結果，包括 Taipei 是 Taiwan 的一部份，而反義詞是中國的 China's。

柏克萊加州大學所發展的 FramNet (http://framenet.icsi.berkeley.edu/)所發展的檢索介面，依據不同的語意框架 frame 詳細探討每一個語意框架常用的詞彙(lexical element)語意角色(semantic role)和對應的語法功能(grammatical function)。

具有語意訊息的中文資料庫有中研院語言所黃居仁教授的中文詞彙網路，這個系統結合了 Wordnet，SUMO ontology，以及中研院詞彙知識庫裡面的詞性標記。

大陸商務出版社所出版的「同義詞詞林」編排的方式是按照語意階層由大類到小類分類。類似同義詞詞林但是分類更細的是大陸出版的現代漢語分類辭典。共有三層的語意分類，最上層的總目有十二類分別是 A 人 B 物 C 時間空間 D 抽象事物 E 特徵F 動作 G 心理活動 H 活動 I 現象與狀態 J 關聯 K 助詞 L 敬語。總類是最大語意類，下面有次類，次類下有次分類。

教育部重編國語辭典修訂本(http://140.111.34.46/dict/)，除了解釋，並有例句，相似詞，相反詞。

大陸董振東先生獨力發展出來的知網 Hownet (http://www.keenage.com)也是一個非常重要的詞彙知識庫(參考 Dong and Dong (2006)。知網 Hownet 包含的訊息相當的多，是一個雙語的知識庫，可以表達概念與概念之間的常識關係。細節將在下一節介紹。

四 Hownet 語意訊息的表達與擷取

我們逐一介紹如何利用知網(Hownet)，,現代漢語分類辭典，重編國語辭典修訂本等資源，並利用 Wordnet 和漢英辭典這些知識庫來找出語意關係，及每一個知識庫的限制。最後我們將這幾個詞彙資源整合成一個綜合性詞彙語意關係知識庫來找出具有語意關係的詞。由於 Hownet 的結構最複雜，語意的關係也最詳細，我們先探討如何應用這個具備雙語詞彙語意表達模式。Hownet2002 比 Hownet 2000 的語意表達方式更豐富，特別是語意角色的部分更清楚，對自然語言理解更有幫助，但是因爲結構相當複雜處理不易，所以我們針對新舊版的不同語意表達模式設計不同的應用程式。我們分別建立 Hownet 2000 及 Hownet2002 的資料庫。這兩版的基本差異可以從下列的表達模式看出來。

{human|人, #occupation|職稱,*cure|醫治, medical|醫}
{human|人:HostOf={Occupation|職位},domain={medical|醫},{doctor|醫治:agent={~}}}

第一個是舊版的 Hownet 的表示法，其中的詞由義元表達，義元本身沒有內部結構，第二個是新版的表示法，義元有內部結構。舊版的 Hownet 語意角色不清楚，例如＊符號可能表示在某個事件擔任 agent, experiencer，或其它的語意角色，但在新版的 Hownet 裡面哪個義元在哪個事件中擔任什麼語意角色非常明確。換言之，新版詞與詞之間的語意關係相當明確。舊版 Hownet 的義元沒有內部結構缺點是對語意角色的判定較不容易，但相對的程式處理起來較簡單，檢索義元的速度也較快速。我們利用這個特色設計一個工具程式。輸入一個中文或英文詞，程式會將這個詞的所有義元列出來，使用者可以選擇其中一個或多個義元，並且可以用 AND 或 OR 來查詢包含這些義元的詞。

例如；輸入車這個中文詞，程式會顯示這個詞 Hownet 義元的表示法，有交通工具、切割、人的姓等幾個不同的意義。使用者選擇其中一個意義，再選擇要檢索一個或多個義元並選擇這些義元之間的關係是 AND 或 OR，程式就會列出包含這些義元的相關中文詞。由於 Hownet 是中英雙語，我們的程式也可以輸入英文用相同的方法找出相關的英文詞及其中文翻譯。這個工具對於全自動或半自動建立中英雙語或中英跨語言的本體論(ontology)系統非常有用，例如輸入車選擇 LandVechicle|車這個義元可以得到不同的車如板車，叉車，餐車，彩車，巴士，長途汽車等。

詞彙與詞彙之間的關係透過 Hownet 的意元表達模式，原本不清楚的詞彙語意關係變得比較清楚。例如：餐車的義元是 {LandVehicle|車，@eat|吃}透過 Hownet 的知識表達法，餐車與吃的語意關係得以連結。但是@究竟代表什麼意義或哪一個語意角色在 Hownet2000 裡面並沒有明確的交代。而在 Hownet 2002 裡面就比較清楚，{LandVehicle|車:{eat|吃:location={~}}}顯示餐車是吃的地點。

醫生在 Hownet2002 裡面有三個英文翻譯 doctor, surgeon, doctor，它們的義元表示都是{human|人:HostOf={Occupation|職位},domain={medical|醫},{doctor|醫治:agent={~}}}。由於義元內部具有結構，因此我們先撰寫一個剖析器將內部結構剖析再來比較義元。義元是一種表達語言知識的 meta language,醫生的義元表示醫生是一個人，具有職位，是醫學領域，且是醫治事件裡面扮演主事者的語意角色。 而醫療這個詞只有一個義元{doctor|醫治}。我們的程式（http://140.112.185.57/~denehs/compare.html）可以把具有相同的義元找出來，或是把相同事件裡面不同的語意角色找出來。例如輸入醫生和醫療會得到下面訊息。

醫生(doctor)和醫療(doctor)同屬於醫治(doctor)事件
醫生(doctor)是 agent
醫療(doctor)是事件

護士(nurse) 的義元是{human|人:HostOf={Occupation|職位},domain={medical|醫},{TakeCare|照料:agent={~}},{doctor|醫治:agent={~}}}，所以輸入醫生和護士會得到。

醫生(doctor)和護士(nurse)同屬於醫治(doctor)事件
醫生(doctor)是 agent
護士(nurse)是 agent

醫生 (physician)	護士 (nurse)	共同點：人:domain={medical|醫}}	{human|	共同語意特徵：{medical|醫}

病人 (patient) 的義元是 {human|人:domain={medical|醫},{SufferFrom|罹患:experiencer={~}},{doctor|醫治:patient={~}}}，所以輸入醫生和病人會得到。

醫生(doctor)和病人(patient)同屬於醫治(doctor)事件
醫生(doctor)是 agent
病人(patient)是 patient

醫生 (physician)	病人 (patient)	共同點：人:domain={medical|醫}}	{human|	共同語意特徵：{medical|醫}

醫院 (hospital) 的義元是 {InstitutePlace|場所:domain={medical|醫},{doctor|醫治:content={disease|疾病},location={~}}}，所以輸入醫生和醫院會得到
醫生(doctor)和醫院(hospital)同屬於醫治(doctor)事件
醫生(doctor)是 agent
醫院(hospital)是 location

醫生 (doctor)	醫院 (hospital)	共同點: 不屬於相同的語意類別	共同語意特徵：{medical|醫}

五 語意資料的整合、擷取、與評估

在本節中我們敘述如何整合知網(Hownet)，,現代漢語分類辭典，教育部國語辭典等資源，並利用 Wordnet 和漢英辭典，擷取上述不同來源的中文詞彙語意訊息。由於大規模質與量的評估非常困難，我們隨機選擇十四組詞來測試這些詞彙知識庫。

處方在 Hownet 裡面有兩個意義，一個是動詞，一個是名詞，所以有兩個義元。

處方(prescription) {document|文書:domain={medical|醫},{order|命令:ResultEvent={prepare|準備:content={medicine|藥物}},instrument={~}}}

處方(prescribe) {write|寫:ContentProduct={document|文書:{order|命令:ResultEvent={prepare|準備:content={medicine|藥物}},instrument={~}}},domain={medical|醫}}

這裡可以看出 Hownet 的一些問題。理論上 Hownet 應該將開處方這個詞與醫療這個事件相連結，並表示開處方的 agent 爲醫生，但是 Hownet 並沒有這樣的訊息，所以醫生和處方的相同特徵只有{medical|醫}。

我們再看其它的例子，藥在 Hownet 裡面有三個意義，所以有三個不同的義元，

藥(certain chemicals) {chemical|化學物}
藥(kill with poison) {kill|殺害:instrument={physical|物質:modifier={poisonous|有毒}}}
藥(medicine) {medicine|藥物}

輸入醫生和藥得到沒有共同的語意類別或特徵。原因是 Hownet 裡面並沒有連接藥{medicine|藥物}與醫療這個事件。理論上藥{medicine|藥物}應該列爲與醫療這個事件的工具 instrument。

另外 Hownet 的義元表示法常有不一致的情形，例如悲哀和痛苦，Hownet 的表示法如下：

悲哀(sorrowful) {sorrowful|悲哀}
痛苦(pain) {experience|感受:CoEvent={unfortunate|不幸}}
痛苦(agony) {unfortunate|不幸}

它們是跟感情情緒有關的近義詞，但在 Hownet 裡面卻沒有任何的關係。此外，Hownet 沒有明顯的同義和近義關係，必須完全靠義元比對，看似一個簡單的工作，實際上卻相當複雜，原因是 Hownet 裡面有相當多的義元，用來定義所有詞彙的每一個義元的重要性並不相等，而有時近義的詞卻用完全不同的義元來表示。所以直接以義元比對有時並不能找到同義詞。事實上 Hownet 的問題是所有以 meta language 來表示語意的理論所必須面對的共同問題。不管是 meta language 或語意特徵，都很難用一套有限的元素來定義所有的詞。例如輸入老闆和老闆娘可以發現 Hownet 的義元沒有表示老闆和老闆娘之間有配偶的關係。

老闆(boss) {human|人:{employ|雇用:agent={~}}}
老闆娘(shopkeeper's wife) {human|人:modifier={female|女},{employ|雇用:agent={~}}}

老闆(boss)和老闆娘(proprietress)同屬於雇用(employ)事件
老闆(boss)是 agent

老闆娘(proprietress)是 agent

老闆(boss)	老闆娘(proprietress)	共同點：{human|人}	共同語意特徵：{employ|雇用:agent={~}

如果我們再看 Hownet 裡面男人和女人的表示，就會發現中間存在許多不一致的情形。

男人(husband) {human|人:belong={family|家庭},modifier={male|男}{spouse|配偶}}
男人(man) {human|人:modifier={male|男}}
女人(wife) {human|人:belong={family|家庭},modifier={female|女}{spouse|配偶}}
女人(women) {human|人:modifier={female|女}}

男人(husband)	女人(wife)	共同點：{human|人:belong={family|家庭}}	共同語意特徵: {family|家庭} {spouse|配偶}
男人(man)	女人(women)	共同點: {human|人}	共同語意特徵: (無)

男人與女人有配偶的意義，有共同語意特徵: {family|家庭} {spouse|配偶}，但是老闆與老闆娘同樣有配偶的意義，Hownet 卻沒有相對的義元表示。

如前所述 Hownet 並沒有明確的同義關係，同義關係必須另外寫義元比對的程式。同樣的，Hownet 也沒有明確的反義關係。例如，買與賣的 Hownet 表示法如下，兩者不但沒有任何相同的義元也沒有明確的反義關係。

HowNet 義元
買(buy) {buy|買}
賣(betray) {betray|背叛}
賣(sell) {sell|賣}

事實上，Hownet 還有獨立的檔案描述事件角色轉換關係(Event Role Shift)，例如
欠與有這兩個義元的關係是 implication，也就是如果 X 欠 Y 一樣東西(target)，Y 就是這一樣東西(target)的 possessor。

owe|欠(X) [implication]←→own|有(Y);
  target OF owe|欠=possessor OF own|有;
  possession OF owe|欠=possession OF own|有.

取與得到這兩個義元的關係是 consequence，也就取的結果是得到。而取的 agent 就是得到的 possessor.
take|取←→obtain|得到 [consequence];
  agent OF take|取=possessor OF obtain|得到;
  possession OF take|取=possession OF obtain|得到.

同理，偷與取這兩義元的關係是 hypernym 的關係，我們從這這些例子可以看出來在事件角色轉換關係裡面，語意的關係如 [implication]， [consequence] ， [hypernym]在左邊或右邊表

示不同的關係。例如 owe|欠(X) [implication]←→own|有(Y) 表示欠 imply 有，steal|偷←→take|取 [hypernym]表示偷是取的下位詞。下面是 Hownet 事件轉換關係的一些例子。

steal|偷←→take|取 [hypernym];
agent OF steal|偷=agent OF take|取;
possession OF steal|偷=possession OF take|取;
source OF steal|偷=source OF take|取.

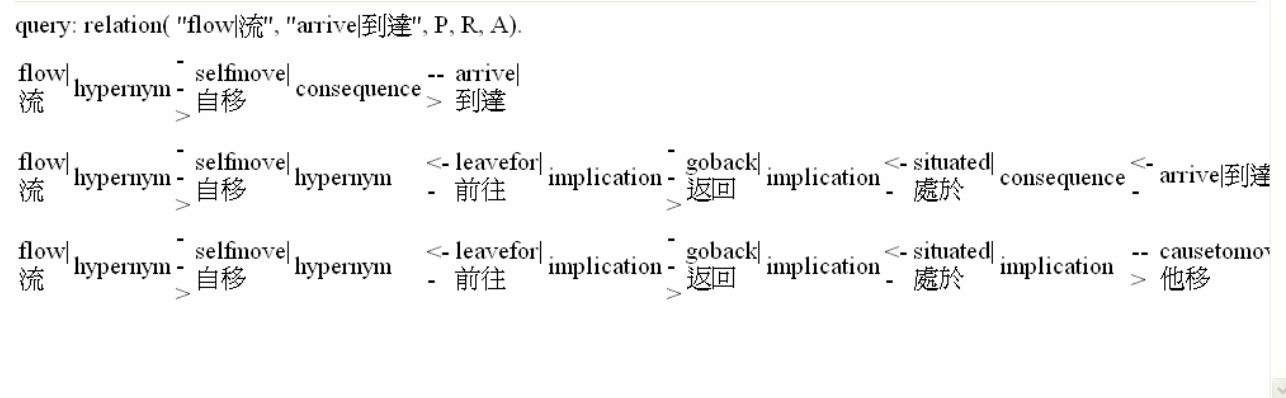
rob|搶←→take|取 [hypernym];
agent OF rob|搶=agent OF take|取;
possession OF rob|搶=possession OF take|取;
source OF rob|搶=source OF take|取.

earn|賺←→take|取 [hypernym];
agent OF earn|賺=agent OF take|取;
possession OF earn|賺=possession OF take|取;
source OF earn|賺=source OF take|取.

buy|買←→take|取 [hypernym];
beneficiary OF buy|買=agent OF take|取;
possession OF buy|買=possession OF take|取;
source OF buy|買=source OF take|取.

我們利用資訊科學常用的 acylic graph (http://en.wikipedia.org/wiki/Directed_acyclic_graph)來表示這些關係。我們利用 Prolog 程式能夠很方便的使用 predicate calculus 自動推論的優點結合 Perl 程式字串處理能力，透過 perl 的 Prolog 模組 (Fandino 2006)，以 perl 程式處理字串後直接在 perl 程式內呼叫 Prolog 程式。如下圖輸入兩個詞彙選擇其中的意義後會找出兩個詞彙之間的語意關係。兩個詞之間的語意的路徑不是唯一的。

圖一 輸入兩個詞並選擇意義後可以得到兩個詞之間的語義關係的路徑



總之，Hownet 裡面沒有明確的同義詞和反義詞。同義的關係必須靠義元比對。反義的關係則被事件角色轉換關係裡面的 mutual precondition 等所部分取代。上下位關係也在事件角色轉換關係裡面記載。至於部分與全體的關係在 Hownet 的裡面透過 whole，part 等義元來表示。例如，輪胎和汽車的義元分別爲

輪胎(tire) {part|部件:whole={part|部件:PartPosition={leg|腿},whole={LandVehicle|車}}}

汽車(automobile) {LandVehicle|車}

輪胎是汽車的一部份這個關係可以透過剖析 Hownet 義元的結構得到。

上面花了相當多的篇幅探討如何利用 Hownet 找出詞彙語意的關係。Howent 雖然提供許多的語義訊息，但是有些地方不一致，而且對於同義詞，反義詞並沒有清楚的記載。英文 Wordnet 裡面則詳細記載了同義關係(synset),反義關係，部分關係，全體關係，上位關係，下位關係。我們利用 Wordnet ::QueryData 和 Wordnet::Similarity 這兩個 Perl 模組及漢英辭典，輸入兩個中文詞，利用漢英辭典將中文詞轉換成英文詞，再利用前述兩個 Perl 模組，即可得到兩個詞之間是否爲同義關係,反義關係，部分關係，全體關係，上位關係，下位關係。例如輸入男人與女人可以找到兩者之間是反義詞。

男人 = boy/dick/joe/man/buck/hombre/blighter/menfolk/husband
女人 = jane/dame/women/judy/donah/wife/womenfolk/frow/hen/tomato/frau/woman/female
女人(woman) is 男人(man)'s antonyms
男人(man) is 女人(woman)'s antonyms
女人(wife) is 男人(husband)'s antonyms
男人(husband) is 女人(wife)'s antonyms

輸入汽車與救護車可以得到汽車是救護車的上位詞，救護車是汽車的下位詞。

汽車 = auto/car/machine/motorcar/motor/autocar/automobile
救護車 = ambulance
汽車(car) is 救護車(ambulance)'s hypernyms
救護車(ambulance) is 汽車(car)'s hyponyms

透過 Pedersen Wordnet::Similarity 這個模組，我們可以得到兩個詞之間語意相似度。如同 Hownet 的事件角色轉換關係，兩個詞的語義路徑不止一條，通常最短的那一條較符合我們的直覺。

汽車 = auto/car/machine/motorcar/motor/autocar/automobile
救護車 = ambulance

WordNet::Similarity

auto#n#1 - ambulance#n#1 : 0.96
auto#n#1 -- motor_vehicle#n#1 -- car#n#1 -- ambulance#n#1

car#n#1 - ambulance#n#1 : 0.96
car#n#1 -- ambulance#n#1

car#n#2 - ambulance#n#1 : 0.782608695652174
car#n#2 -- wheeled_vehicle#n#1 -- self-propelled_vehicle#n#1 -- motor_vehicle#n#1 -- car#n#1 -- ambulance#n#1

car#n#3 - ambulance#n#1 : 0.5
car#n#3 -- compartment#n#2 -- room#n#1 -- area#n#4 -- structure#n#1 -- artifact#n#1 --

instrumentality#n#3 -- container#n#1 -- wheeled_vehicle#n#1 -- self-propelled_vehicle#n#1 -- motor_vehicle#n#1 -- car#n#1 -- ambulance#n#1

但 Wordnet 本身的限制造成應該找到的關係卻沒有找到，例如醫生和病人，員工和雇主並沒有找到反義的關係。

醫生 = medic/aesculapius/physician/surgeon/hakeem/medico/doctor/housestaff
病人 = in-patient/invalid/valetudinarian/patient/inpatient/case
No relationship

員工 = staff/personnel
雇主 = hirer/employer/gaffer
No relationship

翻譯可能出錯及 Wordnet 本身的不一致，即使找到語意關係不見得是正確的。例如我們輸入汽車和輪子，應該找到汽車是輪子的全體詞，卻找到汽車是輪子的上位詞。

汽車 = auto/car/machine/motorcar/motor/autocar/automobile
輪子 = wheel
汽車(machine) is 輪子(wheel)'s hypernyms
輪子(wheel) is 汽車(machine)'s hyponyms

而輸入汽車和輪胎卻找不到任何關係。

汽車 = auto/car/machine/motorcar/motor/autocar/automobile
輪胎 = tire
No relationship

由於 Wordnet 也有一些不一致的地方，我們進一步整合其它的資源。將教育部重編國語辭典修訂本的內容剖析後找出包括解釋、同義詞、反義詞等三類重要語意訊息。教育部重編國語辭典修訂本裡面的相似詞和相反詞事實上就是同義詞和反義詞，不過各個詞條之間內容並不一致，例如輸入醫生得到下面的同義詞與反義詞。

醫生的同義辭 : 醫師,大夫,
醫生的反義辭 : 病人,

但是輸入大夫卻沒有反義詞。我們將所有的同義詞，反義詞連結起來，去掉不一致的地方，重新建立一個資料庫。

除了教育部重編國語辭典修訂本外，我們也將現代漢語分類辭典建成資料庫。現代漢語分類辭典提供比同義詞詞林更詳盡的語義類別訊息。例如：犯罪，犯法的上一層語意是罪刑，刑罰，再上一層是政治，法律，軍事。同屬最下一層的詞是同義或近義詞。如果我們輸入犯罪和法律可以得到這兩個詞共有的類別。

分類辭典: 不屬於相同的語義類別, 共同的部份: 政治●法律●軍事
犯罪 : 政治●法律●軍事→罪行●刑罰→犯罪・違法

法律 : 政治●法律●軍事→法律（一般）→法（一般）

由於每一個詞彙知識庫都有優缺點我們把 Hownet、Wordnet、教育部國語會辭典、現代漢語分類辭典這四個資源完全整合起來。程式在 (http://nlp2.csie.org/~denehs/word_relation.html)。

輸入桌子，椅子，我們得到下面結果。

教育部重編國語辭典修訂本: 不是同義詞
桌子的同義辭 : Not found.
椅子的同義辭 : Not found.
教育部重編國語辭典修訂本: 不是反義詞
桌子的反義辭 : Not found.
椅子的反義辭 : Not found.

分類辭典: 不屬於相同的語義類別, 共同的部份: 飲食●衣服●居住●財產→家用器物

桌子 : 飲食●衣服●居住●財產→家用器物→桌子
椅子 : 飲食●衣服●居住●財產→家用器物→坐具・椅・凳

Hownet

桌子 V.S. 椅子

HowNet 義元
桌子(table) {furniture|家具:{put|放置:location={~}}}
椅子(chair) {furniture|家具:{sit|坐蹲:location={~}}}

Events
(No Event Match)

Relationship
桌子(table) 椅子(chair) 共同點: {furniture|家具} 共同語意特徵: (無)

Wordnet
桌子 = desk/table
椅子 = chair
No relationship

我們再輸入桌子，家具測試這幾個詞彙知識庫。

教育部重編國語辭典修訂本: 不是同義詞
桌子的同義辭 : Not found.
家具的同義辭 : Not found.
教育部重編國語辭典修訂本: 不是反義詞
桌子的反義辭 : Not found.

家具的反義辭 : Not found.

分類辭典: 不屬於相同的語義類別, 共同的部份: 飲食●衣服●居住●財產→家用器物

桌子 : 飲食●衣服●居住●財產→家用器物→桌子
家具 : 飲食●衣服●居住●財產→家用器物→家具
Hownet

桌子 V.S. 家具

HowNet 義元
桌子(desk) {furniture|家具:{put|放置:location={~}}}
家具(furniture) {furniture|家具}

Events
(No Event Match)

Relationship

桌子(desk) 家具(furniture) 共同點: {furniture|家具} 共同語意特徵: (無)

Wordnet
桌子 = desk/table
家具 = furniture/movable
家具(furniture) is 桌子(table)'s hypernyms
桌子(table) is 家具(furniture)'s hyponyms

我們可以從不同的詞彙知識庫得到不同的訊息，但是對於因果關係，目前我們建立的四個詞彙知識庫仍然無法找到。例如輸入犯罪和入獄兩個詞得到下列輸出結果。

教育部重編國語辭典修訂本: 不是同義詞
犯罪的同義辭 : 違法,違警,坐法,犯法,犯科,犯警,
入獄的同義辭 : 下獄,坐牢,
教育部重編國語辭典修訂本: 不是反義詞
犯罪的反義辭 : 立功,犯案,
入獄的反義辭 : 出獄,

分類辭典: 不屬於相同的語義類別, 共同的部份: 政治●法律●軍事→罪行●刑罰
犯罪 : 政治●法律●軍事→罪行●刑罰→犯罪・違法
入獄 : 政治●法律●軍事→罪行●刑罰→關押・監禁

Hownet
HowNet 義元
犯罪(commit a crime) {do|做:content={fact|事情:modifier={guilty|有罪}}}
入獄(put in prison) {suffer|遭受:cause={guilty|有罪},content={detain|扣住},domain={police|警}}

Events
(No Event Match)

Relationship

犯 罪 (commit crime)	a	入 獄 (put in prison)	共同點: 不屬於相同的語意類別	共同語意特徵: (無)

Wordnet
犯罪 =
crime/malefaction/misdeed/sin/maleficent/commitment/misdoing/transgress/perpetration/delinquency/guilt/guilty/wrongdoing/trespass
入獄 = be jailed
No relationship

四個詞彙知識庫只有在現代漢語分類辭典裡面找到兩個詞有相同的語意類別政治●法律●軍事→罪行●刑罰。但是兩者之間因果的關係並無法得到。

上面的 14 組測試資料顯示雖然有不少關係可以透過我們整合後的資料擷取出來，但是扔然有不少關係無法得到。

六 結論

我們利用了四個大規模的詞彙知識庫，並開發了不少的工具程式，輸入任兩個詞彙可以找出下列的語義關係（一）同義關係（二）反義關係（三）上下位關係（四）部件與整體關係 （五）相同事件（六） 相同領域(domain) （七）相同語意特徵（八）相同的語意類別（九）事件與語意角色。

從我們的測試資料顯示單純只靠詞彙知識庫無法得到所有的詞彙語義關係。下一階段將會進一步加入用現有的各種資源如中研院句法樹庫資料，FrameNet，VerbNet，PopNet，Sketch Engine，充分結和語料庫和詞彙知識庫的優點，將統計，語意，語法結合起來擷取更多的詞彙知識。

致謝

本研究得到國科會計畫「詞彙語意關係之自動標注─以中英平行語料庫爲基礎(I)（II）（III）」NSC91-2411-H-002-080 NSC92-2411-H-002-061 NSC93-2411-H-002-013 經費補助，特此致謝。本研究建構的系統由台大資工系高紹航，黃子桓，江加恩，台大資管系戴士強程式設計一併致謝。

參考文獻

Berners-Lee, Tim. (2000) Weaving the Web : the original design and ultimate destiny of the World Wide Web by its inventor. New York : HarperBusiness.

Boguraev, Branimir. and Briscoe, Ted. (1989) Computational Lexicography for Natural Language Processing. Longman: Harlow.
Boguraev, Branimir and Pustejovsky, James (eds.) (1996) Corpus Processing for Lexical Acquisition, MIT Press.
Chaffin, Roger and Illerrmann, Douglas. (1988) The Nature of Semantic Relations: a Comparisons of Two Approaches. In Evens (eds) (1988), pp. 289-334.
Church, K. and Hanks, P. (1990) "Word Association Norms, Mutual Information, and Lexicography." Computational Linguistics, Vol. 16, No. 1, pp. 22-29.
Church, K. et al. (1991) "Parsing, Word Associations, and Typical Predicate-Argument Relations." In Tomita (ed) *Current Issues in Parsing Technology*, Kluwer.
Church, Kenneth, William Gale, Patrick Hanks, and Donald Hindle. (1994) 'Lexical Substitutability,' in Atkins and Zampolli (eds.) Computational Approaches to the Lexicon, pp. 153- 177. Oxford, Oxford University Press.
Cruse, Allan. (1986) Lexical Semantics. Cambridge: Cambridge University Press.
Dong, Zhendong and Dong, Qiang. (2006) Hownet and the Computation of Meaning. World Scientific.
Evens, Martha. (eds.) (1988) Relational Models of the Lexicon: Representing Knowledge in Semantic Networks. Cambridge University Press.
Fillmore, Charles. (1968) The Case for Case. In E. Bach and R. T. Harms, eds., Universals in Linguistic Theory, Holt, Riinehart and Winston, New York, 1-88.
Koenig, Jean-Pierre. (1999) Lexical Relations. CSLI , Stanford University.
Girju, R., Nakov, P., Nastase, V., Szpakowicz, S., Turney, P., and Yuret, D. (2007), SemEval-2007 Task 04: Classification of Semantic Relations between Nominals, *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval 2007)*, Prague, Czech Republic, pp. 13-18.
Grefefenstette, Gregory. (1994) Explorations in Automatic Thesaurus Discovery. Kluwer Academic Publishers.
Hearst, M.A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*, pages 539–545, Nantes, France.
Jackendoff, Ray. (1983) Semantics and Cognition. Cambridge, Mass.: MIT Press.
Jackendoff, Ray. (1990) Semantic Structures. Cambridge, Mass.: MIT Press.
Jones, Stevens. (2002). Antonymy: A Corpus-based Perspective. London ; New York : Routledge, 2002
Levin, Beth. (1985) 'Introduction,' in B. Levin (ed.) Lexical Semantics in Review, Lexicon Project Working Papers 1, Center for Cognitive Science, MIT, pp. 1-62.
Melcuk, Igor. (1988) 'The Explanatory Combinatory Dictionary,' in M. Evens (ed.) (1988), pp. 41 - 74.

Pedersen, Patwardhan, and Michelizzi (2004) WordNet::Similarity - Measuring the Relatedness of Concepts - Appears in the *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-04)*, pp. 1024-1025, July 25-29, 2004, San Jose, CA (Intelligent Systems Demonstration)
Pustejovsky, James, Sabine Bergler, and Peter Annick (1993) 'Lexical Semantic Techniques for Corpus Analysis,' Computational Linguistics, Vol. 19, No. 2, pp. 331 - 358.
Pustevojsky, James. (1995)The Generative Lexicon. The MIT Press.
Pustevojsky, James. (2000) Syntagmatic Processes. in Handbook of Lexicology and Lexicography**,** de Gruyter, 2000.
Resnik, Phillip. (1992) 'WordNet and Distributional Analysis: A Class-based

Approach to Lexical Discovery,' in Workshop Notes, Statistically-Based NLP Techniques, American Association for Artificial Intelligence, pp. 109 - 113.
Schank, Roger. (1975) Conceptual Information Processing. Amsterdam: North-Holland.
Sinclair, John. (eds). (1987) Looking up. Glasglow: Collins.
Sowa, John F. (1984) Conceptual Structures: Information Processing in Mind and Machine. Addison-Wesley.
Turney, P.D. (2006), Expressing implicit semantic relations without supervision, *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (Coling/ACL-06)*, Sydney, Australia, pp. 313-320.
Wilks, A. Yorick (1968) On-line Semantic Analysis of English Texts. Machine Translation, Vol. 11, pp. 59-72.

董大年(主編)(1998)現代漢語分類辭典。上海：漢語大辭典出版社。

魯川 (2001) 漢語語法的意合網路。北京：商務印書館。

梅家駒(主編)(1984)同義詞詞林。北京：商務印書館。

軟體

British National Corpus http://www.natcorp.ox.ac.uk/

DOLCE ontology http://www.loa-cnr.it/DOLCE.html

FrameNet

Hownet http://www.keenage.com/

Language::Prolog::Yaswi

http://search.cpan.org/~salva/Language-Prolog-Yaswi-0.14/Yaswi.pm
Lexical Freenet http://www.cinfn.com/doc/
Ontosaurus http://www.isi.edu/isd/ontosaurus.html
PopNet http://verbs.colorado.edu/~mpalmer/projects/ace.html

Sketch Engine http://www.sketchengine.co.uk/
Wordnet http://wordnet.princeton.edu/

VerbNet http://verbs.colorado.edu/~mpalmer/projects/verbnet.html
Wordnet::Similarity http://www.d.umn.edu/~tpederse/similarity.html
Wordnet ::QueryData http://people.csail.mit.edu/jrennie/WordNet/
中文詞彙網路 Chinese Wordnet (CWN) http://cwn.ling.sinica.edu.tw/
教育部重編國語辭典修訂本 http://140.111.34.46/dict/