

介詞翻譯法則的自動擷取

張俊盛、徐瑞宏、陳惠群
國立清華大學資訊科學學系
新竹市 30043 光復路二段 101 號
Email: jschang@cs.nthu.edu.tw

摘要

介詞在自然語言的處理上一直扮演著重要的角色；而介詞所表示的不同語意關係，更直接影響了介詞的翻譯。我們將設法由機讀語料中擷取出介詞的翻譯法則。

我們將以介詞 with 為例，先依其不同的翻譯特性加以分析，共分成片語、直接受詞、功能、原因、情狀、屬性、共事、對象、關係等九類。其中對直接受詞類，我們利用了句法的次分類資訊；功能及屬性兩類，我們則提出先分類再利用辭彙語意結構來選擇翻譯時所選用的辭彙。我們引入屬性和關係兩類是用以描述 V-O-P-N 結構中的 O 和 N 兩者的關係，以彌補格位分析的不足。

實驗以朗文當代英漢雙語詞典中的雙語例句為訓練語料，抽取句子的 V-O-P-N 結構，觀察其同義詞詞林的類語意碼，再利用決策串列(decision list)來產生法則。外部測試的結果獲得超過 75% 的高正確率。

一、前言

介詞是虛詞(function word)，在句中表示詞與詞或詞與句子間的關係，因此介詞本身與其所構成的短語往往有重要作用。其種類不多，但使用量很大。編譯館國中三年級英文課本有 36.4% 的句子至少包含一個介詞(Chen,1991)。介詞的用法也相當複雜，多數的介詞有多個詞義：

(1-1) He bound it on with rope. (他用繩子把它捆綁起來)

(1-2) Will you walk up to the shop with me? (你和我一塊到那店鋪好嗎?)

朗文當代英漢雙解詞典(Longman English-Chinese dictionary of contemporary English)中，with 就有二十個詞義。此外，介詞形成的結構分歧，例如介詞組連繫歧異的解決(PP attachment disambiguation)也仍是研究重點。

學習外國語時，介詞用法往往令人困惑。因為母語中介詞的用法及現象，常跟外國語的介詞有很大的差異。以中文和英文為例，主要有以下幾點(Tang,1979)：

1. 英文介詞是虛詞，而中文裡，其語意卻常帶有動詞功能，甚至根本就是動詞。如 at (在)、to (到)、into (進)。一旦轉化為不同詞性，更增翻譯的困難。
2. 英文介詞組由介詞及介詞賓語的名詞片語組成。而中文常要再加上一個像「上面」、「裡面」的定位語 (localizer)。例如 on the desk (在桌子上面)。
3. 英文介詞組的語意差異由介詞本身即可反應出。中文卻需要定位語的幫忙。例如：beside the desk (在桌子旁邊)、in the desk (在桌子裡面)。
4. 英文介詞的選擇有時是由其介詞賓語的名詞片語決定。例如：at 6 o'clock、on Friday。有時則由動詞或形容詞決定。例如：arrive at、angry

with。此種搭配 (collocation)現象在中文中不存在。

5. 英文介詞常有多個意思，但不同的介詞有時卻表示相同的意義。例如：10 minutes before / to 6 o'clock。中文少有類似的情形。

由此可見，不解決介詞的問題，翻譯品質難以提升。

二、有關介詞翻譯的研究

(一)介詞的分析

(A.)格位語法(case grammar)

格位語法結合句法及語意解釋，期能了解深層及表面結構(Cook, 1989)(Rich and Knight, 1991)。Ravin (1990)用字典中「to VERB with NP」形式的解釋，解決介詞定義歧異。他依介詞組不同的事態(states of affairs)，對with的意義予以分類：

1. USE：相當於「by means of」「using」。如「to obscure with a cloud」「to surround with an army」。又分為USE-of-Instrument等五小類。
2. MANNER：此類用法相當於副詞。如「to anticipate with anxiety」即相當於「to anticipate anxiously」。又分為Intention-as-MANNER等五小類。
3. ALTERATION：相當於「make」「put into/onto」。如「to mark with a bar」「to impregnate with alcohol」。又分ALTERATION-by-Marking等四小類。
4. CO-AGENCY 或 PARTICIPATION：相當於「and」。例如「to combine with other parts」。
5. PROVISION：相當於「give」。例如「to fit with clothes」。
6. PHRASAL：即片語。

Ravin 分析的沒有考慮連繫到名詞的可能性。此外，分類也有許多模糊不清之處。如USE-of-Substance跟ALTERATION的分界就不明顯。

Chen(1991)大致遵循格位語法概念。介詞組做修飾語時，有多種論旨角色

(thematic role)。依其角色，分為 Agent、Source、Goal、Location、Instrument、Path、Benefactive、End、Extent Time、Point Time 十類。Durand(1993)用修飾語的語意關係(semantic relation for modifier)對介詞組分類；其依據不脫格位語法範疇。

雖然格位分析對介詞組做了區分，卻無法和實際翻譯有較直接的關聯：

(2-1) The cars were all bedecked with flowers. (車子都裝飾上花)

(2-2) She covered her ears with her hands. (她用手捂住耳朵)

這兩句都歸在 Object 類，但一譯「上」，一譯「用」。又如：

(2-3) Her hair became grey with the passing of the year.

(2-4) a child with a dirty face

格位分析並沒有分類可反應出(2-3)become 和 pass 間的關係。(2-4)child 跟 face 間的關係，和名詞間的語意網路有關，同樣也無法依格位分析加以分類。

(B.)朗文詞典的分析

在朗文當代英漢雙解詞典中，對 with 一字的分析極細微，有 20 個不同的意義：

1. 跟、同、和...在一起、帶著、連、加上：staying with a friend (待在朋友那兒)、living with one's children (跟自己的孩子住在一起)
2. 有、顯出：a book with a green cover (綠色封面的書)，a child with a dirty face (臉髒的小孩)，a factory with its chimney smoking (煙囪在冒煙的工廠)
3. 用：fight with a sword (用劍打鬥)，to hear with one's ears (用耳朵聽)
4. 用；拿(指材料、內容)：a cake made with eggs (雞蛋做的糕點)
5. 支持；贊成；向著：to vote with the government (投政府的票)
6. 跟；與；和(指對抗)：to compete with foreign company (跟外商競爭)
7. 順著；跟著：to sail with the wind(順著風航行)、carried along with the crowd (隨著人群移動)
8. 隨著...；與...同時：Her hair became grey with the passing of the year (隨著歲月的消逝，她的

頭髮變白了)

9. 跟；與；和(對同等的比較)：to compare chalk with cheese (拿乾酪來和粉筆比較)、to match a coat with a skirt (裙子配外套)、level with the street (與街成一水平)
10. 跟；與；和(指分離)：to part with money (掏錢)、to break with the past (和過去斷絕關係)
11. 雖然...可是；儘管...可是：With the best will in the world, I can't make her like me (我好意去接近她，但總是得不到她的歡心)
12. 因為；...得...；因為有：singing with joy (高興地歌唱)、grass wet with rain (被雨淋濕的草地)、eyes bright with excitement (眼睛閃著興奮)、With 3 children we cannot afford new furniture (因為有三個孩子，我們買不起新傢俱)
13. 交給...看：to trust someone with a secret (把秘密告訴某人)
14. 有關；至於；對於：Be careful with that glass (那個玻璃杯要小心使用)
15. 跟；與；和(指連接)：connect with
16. (用於命令)：Down with the school (打倒學校)
17. 由...選出：The decision rests with you (由你決定)
18. in with (與朋友鬼混)
19. with it (指穿著、思想、行為等方面很入時)
20. with me/you (聽懂我的/你的話)：Are you still with me? (你還明白我的話嗎)

針對介詞的機器翻譯，我們認為上述分類需要做調整：

1. 相同翻譯的解釋合併：以 6.及 9.為例，都譯成「跟」，似乎沒有必要區分。
2. 不同翻譯的解釋應區分：以 12.為例，當副詞的「得、地」和說明原因的「因」，翻譯上有明顯差異。加上「得」可對應格位語法的「Object」及 Ravin 所提的「Manner」，而「因」卻找不到對應的分類，證實有細分的必要。

(二)介詞翻譯知識的來源

(A.)手工撰寫的法則

Chen (1991)的作法中，動詞組、介詞組、及名詞組須先給予語意特徵 (semantic feature)，例如 the man who he met in the park 中的 met、in、park 的語意特徵分別為

met :	[subj :	[cat : tv]	,	park :	[cat : noun]	,	in :	[cat : prep]
				feature : animate					num : singular					cl : 在	
				case : agent					feature : location					chinese : c2 : X	
				feature : animate					cl : 座					c3 : 裡	
				case : patient					chinese : n : 公園					n : X	
				chinese : 看見											

動詞組除語意特徵外，又分 23 次分類(subcategory)，應用在 \bar{X} 理論(Sells,1985)，決定介詞組的用法。依此撰寫句法規則(syntactic rule)和轉換規則(transfer rule)。

(B.)例句資料庫(example database)

例句為本的作法，知識來源是例句資料庫及同義詞典(thesaurus)。Sumita 和 Iida(1992)用英日文 ATR 語料庫，17,000 句(270,000 字)，限定在研討會註冊的對話。

(C.)目標語語料

許(1994)先用雙語語料窮舉目標語的可能翻譯，再用一千萬字中文目標語語料，經斷詞及詞性標注，統計中文詞頻，計算考慮距離的二元接續表(distance bigram)，輔以同義詞詞林的語意碼，彌補取樣不足，最後以機率模型解決辭彙選擇。

(D.)語意網路(semantic network)

語意網路如 WordNet，儲存名詞，動詞，形容詞及副詞同義群(synonym group)間的反義詞(antonym)、下義詞(hyponym)、同義詞(synonym) …等關係。朗文英漢雙解多功能分類詞典(Longman Lexicon of Contemporary English)及中文的同義詞詞林，將每個字分別給予語意碼，建構成階層式結構，也視為語意網路。

(E.)中英文語意碼

中英文語料在訓練時，不易涵蓋所有辭彙，導致翻譯時不易找到對應辭彙。因此此在語料為本的機器翻譯中，適當的把辭彙分類，以分類來處理，有助於解決一詞多義問題，降低參數空間，也降低儲存空間而提升搜尋速度(柯,1993)。Sekine 和 Tsujii(1995)曾提到，辭彙根據語意分類，特別是根據主題的分類，有助於解決句法歧異。類似作法如 Guthrie(1991)、Yarowsky(1992)、Chen 和 Chang(1995)。

本實驗先透過辭彙語意分析前處理，獲得實詞(content word)的英文語意碼及中文語意碼，這些碼分別取自朗文英漢雙解多功能分類辭典和同義詞詞林。朗文以 14 個主題(subject)來編排，每個主題分 7 至 12 個標題(title)，每個標題含 10 至 50 個相關字的集合(set)。總計在朗文中有 2504 個集合。詞林的編排，則以詞義為分類原則。將語意碼分成大、中、小三級，包括 12 大類，94 中類，1428 小類。以「狗」為例，有兩個語意碼：Bi07(B 類屬「物」類，Bi 屬「動物」類，Bi07 則是「豬狗兔」)和 Dd15(D 類屬「抽象事物」類，Dd 屬「性能」類，Dd15 屬「名稱姓名生肖性別」)。

(F.)衍生辭彙理論 (Theory of Generative Lexicon)

Pustejovsky(1991)提出辭彙語意結構(qualia structure)，每個字的語意就其組成(Constitutive role)、形態(Formal role)、功能(Telic role)、生成(Agentive role)四方面加以討論；以「書」為例，其辭彙語意結構為 (Pustejovsky,1993)：

$$\left[\begin{array}{l} \text{book}(x,y) \\ \text{CONST} = \text{information}(y) \\ \text{FORMAL} = \text{physobj}(x) \\ \text{TELIC} = \text{read}(T,W,y) \\ \text{AGENTIVE} = \text{write}(T,z,y) \end{array} \right]$$

物體又分容器(container)、工具(instrument)、空間物體(figure-ground object)等型

態。然後又分別定義詞項概念範例(Lexical Conceptual Paradigms, LCP)，使名詞的辭彙語意結構可預設常用的功能及生成動詞。如 door 是空間物體，因此可優選 paint 及 scrub 等作用於具象物體的動詞，也可優選 walk through 和 fill 之類的空間動作。

(G.)辭彙句法資訊

前面提到 Chen(1991)將動詞分成 23 個子分類，用以決定介詞組該連繫到名詞還是動詞。由其實驗中可發現，辭彙在句法上的次分類資訊，對翻譯有一定程度的幫助。在分析介詞 with 時，我們就利用了朗文詞典中的句法分析代碼。

(三)介詞翻譯的計算模式

(A.)法則式的作法

Chen (1991)的作法是用字的語意特徵等資訊以及人工撰寫的法則，而在翻譯時應用符合條件的法則。這要花費很多時間在撰寫法則以及語意特徵分析上，法則涵蓋範圍也有限，而且未解決一詞多義的問題。Durand(1993)的作法跟 Chen 很接近。

(B.)例句式的作法

Sumita 和 Iida(1992)的例句為本翻譯系統，英譯日的名詞性(adnominal)介詞正確率 87%，副詞性(adverbial)介詞更達 90%。但用英日文 ATR 語料庫範圍較小，因此且沒有對應例句時，就容易錯誤。且其對介詞連繫等問題，採前處理方式事先解決。

(C.)語料式的作法

許(1994)由目標語出發，從所有可能的翻譯中，找出最可能的組合。例如：

(2-5) I ate a fish with a fork.

從中擷取出的 S-V-O-P-N 單元結構為

E = {I, eat, fish, with, fork}

t("I")={我, 余, 吾, 本人}, a("I")=4
t("with")={用, 和, 以, 跟, 支持, 隨著,}, a("with")=12
t("eat")={吃, 吃飯, 腐蝕, 侵蝕}, a("eat")=4

t(e)是英文詞經辭彙對應的中文集合, a(e)是指英文詞 e 譯成中文時, 所有可能的個數。另有一組規則來變化翻譯辭彙的位置。處理後, 從解答空間 T(E)中找出最佳解:

T(E)={(我, 用, 刀叉, 吃, 魚), (我, 和, 刀叉, 吃, 魚), (我, 以, 刀叉, 吃, 魚).....}

此作法的問題在於介詞對應的中文很難列舉, 參數計算空間太大, 需要大量語料。

(四)目標語的辭彙選擇

我們將辭彙選擇問題分三層次:最簡單的層次是由互等的翻譯中, 選取較佳者:

(2-6) I played a game with her.

譯成「我跟她一起玩遊戲」或「我和她一起玩遊戲」。我們也常省略中文虛字:

(2-7) Glass; handle with care.

其譯文「玻璃, 小心地搬運」中的「地」往往省略。辭彙選擇中最簡單的層次, 只要從可能的辭彙中選擇一個。選擇時, 用許(1994)的作法, 應可得到流暢的翻譯。

第二層次, 是意思正確的轉換。例如:

(2-8) Every night she finished her prayers with a chaplet.

with 在目標語中會轉為動詞。雖可譯成「她每晚用唸珠唸完禱告」, 但常會選擇和「唸珠」搭配的功能動詞 (telic verb), 而譯成「她每晚數著唸珠唸完禱告」。功能動詞可由前述辭彙語意結構中得到。這層次的辭彙翻譯, 無法簡單直接地找到。

由於語言風格(style)不同, 常為求詞句順暢, 而在目標語中配合上下文來進行詞彙修飾, 這是最複雜的層次。因來源語及目標語常會發生語意相同而用詞不盡相同、來源語衍生度(fertility) 大於一、零代詞 (zero anaphor)(陳, 1994)等等的現象:

(2-9) He acted with a Shakespeare company.

可直譯為「他跟一個莎士比亞劇團一起演出」，但此句常轉為「他是莎士比亞劇團的演員」，把動詞「act」轉為名詞「演員」。如此難由目標語中找到介詞的翻譯。

(2-10) The ladies were ablaze with jewels.

(2-11) He was ablaze with anger.

不譯為「女士們因珠寶而發光」和「他因憤怒而激動」；而是「女士們滿身珠光寶氣」和「他怒容滿面」。同時也發現在句型、用字一致下，翻譯也非一成不變；因介詞賓語不同，會有不同的翻譯。另外還有一種可能就是諺語及慣用法：

(2-12) She was born with a silver spoon in her mouth.

此句的含意非字面的「她出生時嘴裡有一支銀湯匙」，而是「她生而富貴」。這類句子，都需處理，才能得到較好的翻譯品質。這個層次並不在本文的討論範圍內。

(五)歸納及整理

根據前面幾節所述，我們得知，若依不同事態或語意關係來分類，可區分介詞的不同語意。但機器翻譯應考慮實際選字的問題。這使得我們必須做適當調整。而且知識來源應減少人工部份。若翻譯每個句子都需人參與決策過程，那只是交談式(interactive)，或機器輔助翻譯。再者須考量參數的數量，以免訓練語料不足，即使是超大型語料。我們提出語意碼的模式來降低參數數量，同時解決一詞多義問題。

三、介詞的分析—以 with 為例

(一)with 的九大類翻譯

(A.)片語 (Phrasal)

出現於片語、慣用法、及祈使句或命令句中。由機讀詞典可取得相關資訊。

(3-1) I cannot put up with your behaviour any longer. (我再也不能容忍你的行為了)

(3-2) Nothing is the matter with me. (我沒什麼)

(B.) 直接受詞 (Direct Object)

V-O-with-N 結構中的動詞 V 在朗文詞典中的語法次分類碼 (syntactic subcategorization code) 是以下其中之一時，with 即為此類：

[A] : 雙賓動詞(double-object verb)，需一直接受詞和一間接受詞。如 fill、emboss。

[B] [Wv5] : 常帶-ed 而作形容詞用之動詞。如 bedeck、ornament、pinch。

[C] [T1，通常被動] : T1 指及物動詞，且緊接在動詞後的受詞或補語位置要置一名詞。此類動詞常以被動式出現。如 beset、encompass。

格位語法分析時，這些動詞幾乎都在 Object 類。而介詞組中心語往往是直接受詞，故為「直接受詞」類。翻譯有兩種情形，一是譯成「被」，特別是用於被動句：

(3-3) The river was contaminated with waste from the factory. (河川被工廠排出的廢物污染了)

另一種因動詞性質，使 with 譯成一定位語(如「上」)或修飾語(例如「滿」)：

(3-4) A machine is clogged with dirt. (機器塞滿了髒物)

(3-5) to spray a wall with paint (將牆噴上油漆)

此種動詞，如 emboss 代碼 D1+on/with，使 with 分在「直接受詞」類。其翻譯從 on 中可得到線索，表 emboss 動作作用在物體表面上；故 with 譯「上」。又 embed 代碼 D1+in/with 表動作作用到物體內部；with 譯為「入」。emboss 的中文是「加浮雕花紋於...上」，fill 的解釋是「填滿」，也指出 with 的翻譯。

(C.) 功能 (Telic)

第三類為「功能」類，表示「用...工具」「以...方法(手段)」等概念。

(3-6) Polish your shoes with a brush. (用刷子擦亮你的鞋)
Fa Bq Bp

上肢動作 衣物 用品

(3-7) He fobbed me off with a story. (他編了一個故事來騙我)

Hn Aa Dk
惡行 人的泛稱 文教

(3-8) He refreshed himself with a glass of beer. (他喝了杯啤酒提提神)

Je Aa Br
影響 人的泛稱 食品

明顯看出英文介詞在中文會轉成動詞。句中 with 都譯成「用」或「以」，意思雖不變，卻未必合習慣。因此處理時將介詞正確歸類，再依中文辭彙語意結構中的動詞加以翻譯；「故事」的生成動詞是「編」，「啤酒」的功能動詞是「喝」。

過去的分析，都有「工具」(instrument)類，跟「功能」類意義相近。不同的是，「功能」類範圍大。(3-6)的「刷子」是工具，但(3-7)的「故事」視為工具就很勉強。但兩者均可以相同方式選取正確目標語，故應屬同類。

(D.)原因 (Cause)

第四類為「因」類，表示「原因」「條件」：

(3-9) The child's eyes rounded with excitement. (孩子因興奮眼睛睜得圓圓的)

Ih Ga
變化 心理狀態

(3-10) His back was bent with age. (他因年老而背部彎曲)

Fd Ca
全身動作 時間

(3-11) Your hands are blue with cold. (你的雙手凍得發青了)

XX Ec Eb
be 動詞 顏色 表象

Durand(1993)的分析也有「原因」類。若不潤飾，幾乎都譯成「因」。然可為原因的東西很多；像是時間(Ca)、物體(Bo)、溫度(Eb)，或是情緒(Ga)。稍後實驗證明「原因」這種語意關係，不易從少量資料中擷取。

(E.)情狀 (Manner)

第五類為「情狀」類，當介詞片語用以修飾動詞，用法相當於副詞時即為此類：

(3-12) He assailed the difficulty with eagerness. (他興致勃勃地克服困難)

Jd	Da	Df
存在	事情	意識

(3-13) The boy's heart fluttered with excitement. (這男孩興奮得心砰砰地跳)

Ib	Ga
生理現象	心理狀態

不經潤飾，幾乎都譯成「地」。介詞賓語的中心語都是些與情緒或人的特質有關的辭彙，涵蓋範圍包括意識 (Df)、德才 (Ee)、心理狀態 (Ga)、心理活動 (Gb)等，分類時較明確簡單。Ravin 及 Durand 的分類，也有 Manner 類。

(F.)屬性 (Attribute)

第六類為「屬性」類。當 N 為具體的東西時我們常說「S(O)帶著 N」，而當 N 是抽象概念時，則常說是「S(O)有 N」。換言之，N 可視為 S 或 O 的一個屬性：

(3-14) some fresh blood with new ideas (有新觀念的新人)

Al	Dk
才識	文教

(3-15) people with certain diseases (患某種病的人)

Aa	DI
人的泛稱	疾病

(3-16) The servant came into the bedroom with a cup of tea. (僕人端著一杯

Hj	Bn	Br
生活	建築物	茶進入臥室) 食品

(3-17) He acted with great nobility of purpose. (他懷著崇高的目的的行事)

Hi	Db
社交	事理

這類也有譯成中文動詞的現象。因此用中文辭彙語意結構中的動詞來選字。和「功能」類不同的是，「屬性」類用形態(formal)方面的資訊，而非功能方面。不同

型態和不同動作搭配。例如目的、理想，用「懷著」來描述。若無適當動詞，則用預設的「帶」和「有」。未必流暢，但能正確傳達介詞代表的語意關係。

此外，「屬性」類描述的是介詞組連繫到名詞的關係，Durand (1993) 的 Quality 類中有類似的概念，而在 Ravin(1990) 的研究及格位語法中則找不到對應的分析。

(G.)共事 (Co-agency)

第七類為「共事」類，通常出現於「S 跟(與/和/隨)N 一起 VO」情境。基本上這類就是出現在 S 與 N 一起做 V 這件事的情形下，相當於格位語法中的 Agent 類：

(3-18) I am quits with him. (我和他互不相欠)
XX Ed Aa
be 動詞 性質 人的泛稱

(3-19) Sweeden has frontiers with Norway. (瑞典與挪威接壤)
Jd Cb Di
存在 空間 社會政法

(3-20) She is always fun to be with. (和她在一起總是很有趣)
XX
be 動詞

大致是在「跟」「與」「和」三者中選擇，輔以「在一起」等字眼。可用許(1994)目標語語料為本的作法，選擇一個最常用，最流暢的辭彙。

(H.)對象 (Concern)

第八類為「對象」類，表「對於」「關於」等概念。Durand (1993)也有此類：

(3-21) I am losing my patience with him. (我對他失去耐心了)
Jd Ec Aa
存在 德才 人的泛稱

(3-22) I have some authority with the young boy. (我對那小男孩有些影響力)
Jd Je Ab
存在 影響 男女老幼

翻譯時，如不考慮潤飾，此類的直譯為「對」或是「對於」。

(L)關係 (Relation)

第九類為「關係」類，表示 O 和 N 間的關係，例如所謂的 Part-Whole Relationship：

(3-23) He has got a good position with an oil company. (他已在油公司謀得一份好工作)

Jd Di Dm
存在 社會政法 機構

(3-24) I am back to square one with the work. (這項工作得從頭再來)

XX Da Di
be 動詞 事情 社會政法

或是 O 帶有動作意味，而 N 為其賓語，例如：

(3-25) John is a great favourite with his grandmother. (約翰是他祖母的最愛)

XX Ag Ah
be 動詞 人的狀況 親人

(3-26) I need some guidance with my studies. (我的功課需人指導)

Jc Hg Dk
配合 教衛科研 文教

這類描述 O 與 N 的關係，僅 Durand(1993)有類似分析。O 與 N 的關係，用其他方法訓練，涵蓋範圍也許較廣。如語意網路。「關係」類的翻譯，似乎沒有規則。

	朗文	格位語法	Ravin	本實驗	備註
1	跟	Agent	Co-Agency	共事	Part-Whole
2	有、顯出			屬性/關係	
3	用	Instrument	Use	功能	
4	用、拿(材料)	Object	Alteration/Provision	功能/直接受詞	
5	支持	Object	Co-Agency	共事	
6	跟(對抗)	Agent	Co-Agency	共事	While
7	順著	Agent	Co-Agency	共事	
8	隨著		Co-Agency	共事	
9	跟(比較)	Object	Co-Agency	共事	
10	跟(分離)			共事	Phrasal Constraint
11	雖然			原因	
12	因為、得	Object	Manner	原因/情狀	Cause
13	交給...看	Object	Co-Agency	共事	Phrasal Phrasal Phrasal Phrasal Phrasal Phrasal Phrasal Phrasal
14	有關、對於	Benefactive		對象	
15	跟(連接)		Phrasal	共事	
16	(命令)		Phrasal	片語	
17	由...選出		Phrasal	共事	
18	in with		Phrasal	片語	
19	with it		Phrasal	片語	

20	with me/you		Phrasal	共事	Phrasal
----	-------------	--	---------	----	---------

表 3-1 Longman、Case Grammar、Ravin、及本實驗四種分析的對照表

跟過去比較起來，我們的分析有以下的優點及特性：

1. 最終目標是正確翻譯，所以不需分析細微的語意差異，因為即使分析正確，對實際翻譯可能沒有幫助。所以分析時應以目標語的翻譯作為依據。
2. 在「功能」跟「屬性」兩類中，提出先分類後再選字的觀念。若配合如辭彙語意結構的處理，只要分類正確，即可得到正確的介詞翻譯。
3. 引入「直接受詞」類。句法次分類可決定介詞的分析與翻譯。
4. 引入「屬性」及「關係」類，表 O 與 N 的關係。格位語法無類似分類。

(二)其他介詞的分類

本實驗的介詞分類似乎免不了人力介入。觀察那些已有對應到中文翻譯的 with，其實分類相當明顯。再者，介詞數目不多且不會增加，人力介入也不困難。

譯文經常是經過潤飾的。本實驗訓練語料，48.8%的 with 無對應中文。除去「片語」類，仍有 34.1%因目標語潤飾而無對應。目前採人工輔助的方式解決這個問題。

四、介詞分析的計算模式

(一)雙語資料的處理

我們以朗文當代英漢雙解字典含 with 的中英對照例句為訓練語料。英文句先做詞性標註(Church,1988)及原形化處理，然後由朗文英漢雙解多功能分類詞典，得到英文辭彙的英文語意碼集合。而對應的中文句斷詞後查閱同義詞詞林，得到中文辭彙及其中文語意碼集合和詞性。再做辭彙語意分析，給予每個英文字正確語意碼(Chen and Chang,1994; Chang and Chen,1996a; Chang and Chen,1996b; Chang et al.,

1996)，並完成中英文辭彙對應(柯,1993; Ker and Chang,1995; Ker and Chang, 1996)。

有關介詞的問題，多以 V-O-P-N 單元結構¹表示句子 (Hindle and Rooth, 1993) (許, 1994) (Chen and Chang, 1995)。取得所需的 V-O-with-N 結構(許,1994)方法如下：

1. with 之前第一個出現的動詞為 V-O-with-N 結構中的 V。
2. 動詞與 with 間的名詞組為 O，其最末字為中心語。若無名詞，取形容詞。
3. with 之後為介詞賓語，取名詞組的中心語為 V-O-with-N 結構中的 N。

取出 1,424 個 V-O-with-N 結構後，將句中 with 中文翻譯歸類至九大類中。

(4-1) I <u>open</u>	the <u>door</u>	with a <u>key</u> .	(我 <u>用鎖匙</u> <u>開</u> <u>門</u>)
Mb065	Db024	Hd131	Bo03 Fa31 Bn04
打開與開鎖	門及其組成	鎖與鑰匙	機件 開關 門窗

假設語意碼取第二級，動詞(open)屬手的動作，受詞(door)是建築物的一部份，介詞賓語(key)是工具。因介詞表達「關係」，手的動作跟工具間的關係，就是「手用工具作用於建築物上」。所以從句子組成的語意，可推出組成成份間可能的關係。

(4-2) I went out with a doctor. (我跟一個醫生一起出去)
Bj166 Ae15

(4-3) I went out with a lawyer. (我跟一個律師一起出去)
Ck201 Ae12

本實驗以同義詞詞林 94 個中類語意碼²，來研究 V-O-with-N 結構和 with 的翻譯。

(二)法則的產生

採用決策串列(decision list)(Rivest,1987; Yarowsky,1994a; Yarowsky,1994b)來產

1. ¹ V 表動詞組的主要動詞，O 表受詞，P 表介詞，N 表介詞賓語的中心語。若進一步推廣至 S-V-O-P-N 結構的話，其中的 S 表主詞

2. ² 若 V、O、N 其中一欄沒有字的話，則以一表之。例如：I go with you. 句中的 O 的中文語意碼即標為-。

生法則。基本上法則就是說明搭配辭和 with 翻譯的關係，以及其關係的可靠性分數。

但產生法則前，要先加以處理兩種特殊情形：

1. 特殊用法：當 with 以固定的搭配形式和某些特定字共現時，可直譯成常用中文，而不需看其語意碼或是其他。例如出現 mix with，with 直接歸到「跟」類(...跟...混合)。au fait with 常直譯成「熟悉」，with 就歸到「片語」類。
2. 「直接受詞」類：當句子中的動詞在朗文詞典中的文法分析代號為「D」、「Wv5」、「T1 通常被動」這三類時，我們將 with 分到「直接受詞」類。

實驗中的 V、O、N，所可能出現的搭配形式有以下幾種：

1. VON：是要三個都符合才算。亦即應用了所有包含在 V、O、N 這三個變數中的資訊。因為要符合所有的條件，所以接近於例句式的作法。
2. VO、VN、ON：只要二個符合。其中 VN 類代表動詞及其論元的關係，ON 類代表 Part-Whole 等字與字間的語意關係。三類間，不預設優先權，完全由分數值來決定。
3. 只要一個符合。V 類由動詞來看介詞翻譯。N 類則由修飾語(adjunct)為出發點。O 類雖無理論基礎，但 O 空白或為代名詞時，介詞組此時無法連繫到 O，將間接解決結構歧異問題，所以也有助於介詞的翻譯。而這三類間也不預設優先權，完全由分數值來決定。

表 4-1 根據可能出現的搭配形式，分別舉了幾個例子：

搭配形式	with 分類	V	O	N
VON	功能	Fa(上肢動作)	Bp(用品)	Bk(全身)
VO	對象	Hi(社交)	Aa(人的泛稱)	
VN	共事	Hj(生活)		Aa(人的泛稱)
ON	對象		Ee(德才)	Aa(人的泛稱)
V	功能	Ig(始末)		
O	功能		Bm(材料)	
N	共事			Aj(人的關係)

表 4-1 不同形式搭配的法則

以表 4-1 第一列 Fa-Bp-Bk 為例，在訓練語料中共有下表 4-2 中的四個例句：

with 分類	原句
功能	She <u>untied</u> the <u>knots</u> with dexterous <u>fingers</u> .

功能	He <u>drummed</u> on the <u>table</u> with his <u>fingers</u> .
功能	Hit the <u>ball</u> with a long free swing of the <u>arm</u> .
功能	If you <u>pick up</u> the <u>ball</u> with your <u>hand</u> in golf, you suffer a penalty.

表 4-2 訓練語料中 Fa-Bp-Bk 的例句

我們假設 VON 都符合者所含資訊較精確，優先權最高；只符合一個的優先權最低。而對每個可能的搭配 C，對應的 with 分類 P，其分數值為

$$Score(C, P) = \frac{Count(C, P)}{Count(C, \bar{P})}$$

$Count(C, P)$ 表示搭配 C 且 with 的分類為 P 的次數，而 \bar{P} 表示其他可能的 with 分類。分數高則正確率高，會優先使用。分數值大於一時，表示對的次數比較多；因此可為門檻值。若不設門檻值(或為 0 時)，表示只要是訓練語料中取出的訊息，就儘量使用。再以表 4-2 為例，此時搭配 C 為 Fa-Bk-Bp，P 為「功能」，規則的分數值為

$$Score(Fa - Bk - Bp, 功能) = \frac{4}{0}$$

即 Fa-Bk-Bp 出現四次，with 皆屬「功能」類，且找不到非功能類的例句。這表示此法則可靠度極高。

但為解決資料稀疏(sparse)的問題，還要做修勻(smoothing)。假設所有的零值都是由於抽樣不足，用簡單方式來修勻：在 $I \times J$ 表格中，每個方格加上 $1/J$ (Agresti, 1990)。如表 4-2，VON 類法則可視為在 $J = 7$ 的表格中³，因此搭配為 Fa-Bp-Bk 且介詞 with 屬「功能」類的這條法則，分數值為

$$Score(Fa - Bk - Bp, 功能) = \frac{4 + \frac{1}{7}}{0 + \frac{7-1}{7}} = 4.8333$$

3. ³ 全部可能的介詞分類共有九類，可是由於「片語」類跟「直接受詞」類的資訊明確，在之前就先處理掉了，所以在產生法則時只考慮剩下的七類。

五、實驗結果與討論

(一)實驗結果

表 5-1 到 5-3 是產生出來的法則。在此我們只列舉各種形式分數最高的前十名。

分類	V	O	N	分數	計數值	例句
共事	Hj	--	Aa	28.714	25	You must <u>come</u> with <u>us</u> , I insist.
共事	XX	--	Aa	11.571	10	The poor <u>are</u> always with <u>us</u> .
情狀	Fc	--	Ga	7.0000	6	The children <u>squealed</u> with <u>delight</u> .
功能	Fb	--	Bk	7.0000	6	One old lady <u>walked</u> with heavy <u>foot</u> .
功能	Ig	--	Dk	5.8571	5	The party <u>finished</u> with a <u>song</u> .
對象	XX	Ee	Dj	4.7143	4	He <u>is</u> <u>carefree</u> with his <u>money</u> .
對象	XX	Ee	Ab	4.7143	4	They <u>are</u> very <u>strict</u> with their <u>children</u> .
對象	XX	Ee	Aa	4.7143	4	Are you <u>being</u> <u>straight</u> with <u>me</u> ?
共事	Hj	--	Aj	4.7143	4	She is <u>shacking up</u> with her <u>boyfriend</u> .
共事	Hi	--	Aa	4.7143	4	You must not <u>joke</u> with <u>him</u> .

表 5-1 VON 類法則的前十名

1,424 個 V-O-with-N 結構中，「共事」類最多，佔 361 句(25.3%)，其次為「功能」類，324 句(22.8%)；可見 with 的分析相當困難。只取最高頻對應，準確率極低。

我們以應用率(applicability)及準確率(precision)作為評估實驗結果的依據：

$$\text{應用率} = \frac{\text{完成猜測的VOPN數}}{\text{所有測試的VOPN數}} \quad \text{準確率} = \frac{\text{正確猜測的VOPN數}}{\text{完成測試的VOPN數}}$$

分類	V	O	N	分數	計數值	例句
共事	Hj		Aa	44.573	39	It is a pleasure to <u>do</u> business with <u>you</u> .
功能	Fa		Bo	21.716	19	<u>Cut</u> it with the <u>scissors</u> .
共事	XX	--		16.002	14	She <u>was</u> here with her betrothed.
功能	Ig	--		13.716	12	The story <u>opens</u> with a snowstorm.
功能	Fa		Bp	13.716	12	<u>Bind</u> the prisoner with <u>rope</u> .
功能	Fa		Bk	13.716	12	<u>Hit</u> him with your <u>right</u> .
屬性	--	Aa		10.287	9	<u>Someone</u> with creativity is needed.
屬性	--	Ab		9.1444	8	The <u>man</u> with the big dog came in.
功能	Fa	Aa		9.1444	8	I <u>cut myself</u> free with an axe.
功能	Ig		Dk	9.1444	8	The party <u>finished</u> with a <u>song</u> .

表 5-2 VO/VN/ON 類法則的前十名

表 5-4 是 1,424 句內部測試。外部測試用牛津高級英英英漢雙解辭典以及 Collins Cobuild English Language Dictionary 的 109 個 with 的解釋例句。實驗結果如表 5-5。

分類	V	O	N	分數	計數值	例句
功能	Fa			10.043	59	He <u>splashed</u> his face with cold water.
功能			Fa	5.7143	5	He set the machine going with a <u>push</u> .
功能	Ig			4.9032	19	Let's <u>top off</u> the evening with a drink.
功能	Hd			4.8696	14	They <u>feed</u> the pig with the parings.
功能		Fa		4.5714	4	He gave me a <u>rap</u> with her pencil.
情狀			Gb	4.5217	13	Cross the road with <u>care</u> .
共事			Af	4.2667	8	He arrived with several <u>attendant helper</u> .
共事	Ja			3.7333	7	The decision <u>lies</u> with you.
共事			Aj	3.4872	17	I'm storing my television with a <u>friend</u> .
功能		Bm		3.4286	3	He probed the <u>mud</u> with a stick.

表 5-3 V/O/N 類法則的前十名

門檻值	0	1	2	3
總數	1424	1424	1424	1424
正確	1397	1396	1174	1005
錯誤	27	26	106	48
未知	0	2	144	371
應用率	100%	99.9%	89.9%	74.0%
準確率	98.1%	98.2%	91.7%	95.4%

表 5-4 內部測試的結果

門檻值	1	2	3
總數	109	109	109
正確	82	72	54
錯誤	27	21	9
未知	0	16	46
應用率	100%	85%	58%
準確率	75%	77%	86%

表 5-5 外部測試的結果

外部測試可看出，門檻值越高，應用率越低，準確率越高。若以應用的法則種類來觀察(門檻值為 1)，可獲表 5-6 的結果：

種類	總數	正確	錯誤
直接受詞	7	7(100%)	0(0%)
片語	21	19(90%)	2(10%)
VON	27	23(85%)	4(15%)
VO/VN/ON	49	33(67%)	16(33%)
V/O/N	5	0(0%)	5(100%)

表 5-6 外部測試結果依法則種類加以分類

這和我們之前「符合的欄位數越多，可靠度越高」的假設完全符合。只符合一個欄位的五句全部都錯，雖說是資訊太少，所以可靠度較低，但不代表這些法則沒有利用價值。以只看 N 的欄位為例，部份的修飾語其實已可左右介詞 with 的翻譯：

(5-1) Do your work with care. (小心做你的工作)

Hj Dk Gb
生活 文教 心理活動

(5-2) Cross the road with care. (小心過馬路)

Hf Bn Gb

交通運輸 建築物 心理活動

(5-3) Wash the cups with care. (小心洗杯子)

Fa Bp Gb
上肢動作 用具 心理活動

由這三例可發現，雖然動詞受詞不同，但大部份“with care”都屬「情狀」類（訓練語料中 7 句皆屬「情狀」類）。只是由於字典偏差，導致法則有過份特殊(over-specialization)的現象，使得在應用時，較傾向使用符合兩個以上欄位的法則。

(二)錯誤結果分析

觀察外部測試(門檻值 1)，若以 with 分類來看，可得到如下表 5-7 的數據：

分類	總數	正確	錯誤
片語	14	14(100%)	0(0%)
直接受詞	7	7(100%)	0(0%)
情狀	4	4(100%)	0(0%)
功能	16	15(94%)	1(6%)
共事	36	27(75%)	9(25%)
對象	8	6(75%)	2(25%)
原因	7	3(43%)	4(57%)
關係	5	2(40%)	3(60%)
屬性	12	4(33%)	8(66%)

表 5-7 依 with 翻譯分類結果分類

前六類正確率在 75%以上，後三類卻不到 50%。我們的解釋是：

1. 「原因」：Yarowsky(1992)在解決字義的歧異時提到，若一個字可出現在多種上下文中(如 interest)，就難由鄰近字的標題(topic)推測其含意。本實驗中也有類似現象。一件事的原因可能是一個人，一個東西，甚至只是一個人的心理狀態。光從 V、O、N 似乎無法看出介詞賓語是用來表示原因的。也許要設法從句子的語意中觀察出介詞賓語是用來說明前面的動詞或全句，才可能將其正確譯成「因為」。例如“Metal expands with heat.”，需

要現實世界的知識—東西因熱會膨脹；否則「原因」和其他類並不易區別。

2. 「關係」：「關係」這類出現於描述 O 與 N 的關係。然而 with 連繫到動詞的機率原本就較高(Sumita and Iida,1992)，使得從訓練語料中得到的資訊有限。Iris(1988)曾將 Part-Whole 關係分四個模型來解決，其中，有些要有現實世界知識(如 Valentine's Day-February)，有些在用字上有特殊的搭配(如 sheep-flock)。這些資訊龐大，從實驗中的 1,400 句中，是取不到的。
3. 「屬性」：在錯誤的八句中，有六句的介詞片語是用來主詞的述語，而主詞在此並未加以考慮。將實驗擴大至 S-V-O-with-N，也許可以解決。

而其他的錯誤原因包括：

1. 語料不足：只有一千四百多個訓練句，部份的現象無法涵蓋。
2. 同義詞詞林的分類：在同一個同義詞詞林的中類裏，仍有很多語意分歧。這也就是 Yarowsky(1992)的「同一分類中的些微差異」。以“Cut the cloth with the knife.” 以及 “Put the cloth with the knife.” 為例，我們用刀子來剪 (Fa28)，但也會放 (Fa13) 一樣東西跟刀子在一起。可以改用小類來做，或搭配其他相關的資訊(如英文的語意碼)，但先決條件是要有更大的語料。
3. 前處理的錯誤：前處理的錯誤會間接影響結果。詞性標注，以布朗語料庫測試，有 97.495% 的準確率。解決字的歧義，有 94% 的準確率 (Chang and Chen, 1996)。辭彙對應，是 93.3% (Ker and Chang, 1995)。中英文語意碼對應，則為 93.4% (Ker and Chang, 1996)。改進前處理，也是未來研究的重點。
4. 機讀字典的缺點：字典列舉所有可能，所以用於處理片語等特殊用法時，

不易因語料不足而取樣不到。但字典通常偏好特殊現象，加上編纂者的個人觀點(Zernik, 1991)，因此也不易反應正常的語言現象。Hearst (1991)提到，機讀字典對某些語意結構的偏好，會導致不平衡的共現資訊。

六、結論與未來展望

根據實驗的結果與分析，我們提出一些有待加強之處及可能的發展方向：

1. 推廣至其他的介詞，甚至是其他的虛詞，以期得到更好的機器翻譯品質。
2. 解決需人工分類訓練語料的問題，使整個流程能做到完全自動。
3. 收集目標語中的辭彙相關資訊，用以輔助辭彙選擇的問題。對中文做較深入的語法分析，應該可能對介詞在中文中扮演的角色有更進一步的認識。
4. 利用其他機讀字典中可以擷取的資訊，例如領域碼 (domain code)，定義，語意碼間的階層關係，以得到更多有用的訊息。
5. 利用其他雙語語料，配合現有機讀字典，以得到較一般性、平衡性的法則。

致謝

本實驗獲得行政院國科會計畫編號 NSC85-2213-E-007-042 贊助，特此致謝。

參考文獻

1. Agresti, Alen, *Categorical Data Analysis*, New York: John Wiley & Sons, 1990.
2. Chang, Jason J.S., and J.N. Chen, "Acquisition of Computational-Semantic Lexicons from Machine Readable Lexical Resources", in *Proceeding of ACL SIGLEX Workshop*, 1996.
3. Chang, Jason J.S., and J.N. Chen, "Word Sense Division Based on Dictionary and Thesaurus", unpublished manuscript, 1996.
4. Chang, Jason J.S., H.H. Sheng, J.N. Chen, S.J. Ker, "Combining Machine Readable Lexical Resources and Bilingual Corpora for Board Word Sense Disambiguation", unpublished manuscript, 1996.
5. Chen, Hsin-Hsi, "The Transfer of Prepositional Phrase in English-Chinese Machine Translation System", *Conference Book, ACH/ALLC*, 1991.
6. Chen, Jen-Nan, and Jyun-Sheng Chang, "Towards Generality and Modularity in

- Statistical Word Sense Disambiguation”, in *Proceedings of the Asian Conference on Language, Information and Computation*, 1994.
7. Chen, Mathis H.C., and Jason J.S. Chang, “Structural Ambiguity and Conceptual Information Retrieval”, in *Proceedings of the 10th Pacific Asia Conference*, HongKong, 1995.
 8. Church, Kenneth Ward, “A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text”, in *Proceeding of 2nd Conference on Applied Natural Language Processing*, Austin, Texas, 1988.
 9. Cook, Walter A., *Case Grammar Theory*, Washington, D.C.: Georgetown University Press, 1989.
 10. Dagan, Ido, and Alon Itai, “Word Sense Disambiguation Using a Second Language Monolingual Corpus”, *Computational Linguistics* 20(4), 1994.
 11. Durand, Jacques, “On the Translation of Prepositions in Multilingual MT”, In Frank Van Eynde, editor, *Linguistic Issues in Machine Translation*, London: Pinter Publishers, 1993.
 12. Gale, William, and Kenneth Church, “What is wrong with adding one?”, *AT&T Bell Laboratories Statistical Research Report No.90*, 1989.
 13. Hearst, Marti A., “Noun Homograph Disambiguation Using Local Context in Large Text Corpora”, in *Proceedings of the 7th Annual Conference of the UW Centre for the New OED and the Text Research: Using Corpora*, 1991.
 14. Hindle, Donald, and Mats Rooth, “Structural Ambiguity and Lexical Relation”, *Computational Linguistic* 19(1), 1993.
 15. Hutchins, W. John, and Harold L. Somers, *An Introduction to Machine Translation*, London: Academic Press, 1992.
 16. Iris, Madelyn Anne, Bonnie E. Litowitz, and Martha Evens, “Problems of the Part-Whole Relation”, in Martha Walton Evens, editor, *Relational Models of the Lexicon: Representing Knowledge in Semantic Networks*, Cambridge: Cambridge University Press, 1988.
 17. Kennedy, Graeme, “Between and Through: The Company They Keep and Functions They Serve”, in Karin Aijmer and Bengt Altenberg, editors, *English Corpus Linguistics*, New York: Longman Inc., 1991.
 18. Ker, Sur-Jin, and Jason J.S. Chang, “Automatic Acquisition of Class-based Rules for Word Alighment”, in *Proceedings of the 10th Pacific Asia Conference*, HongKong, 1995.
 19. Ker, Sur-Jin, and Jason J.S. Chang, “Aligning More Words with High Precision for Small Bilingual Corpora”, in *Proceedings of 31th Annual Meeting of the Association for Computer Linguistics*, 1996.
 20. Longman, *Longman English-Chinese Dictionary of Contemporary English*, Hong Kong: Longman Group(Far East) Ltd., 1992.
 21. McArthur, Tom, *Longman Lexicon of Contemporary English (English-Chinese Edition)*, Hong Kong: Longman Group(Far East) Ltd., 1993.
 22. McRoy, Susan W., “Using Multiple Knowledge Source for Word Sense Disambiguation”, *Computational Linguistics* 18(1), 1992.
 23. Pustejovsky, James, “The Generative Lexicon”, *Computational Linguistics* 17(4), 1991.

24. Pustejovsky, James, "Lexical Semantic Techniques for Corpus Analysis", *Computational Linguistics* 19(2), 1993.
25. Ravin, Uael, "Disambiguating and Interpreting Verb Definitions", in *Proceedings of 28th Annual Meeting of the Association for Computer Linguistics*, Pittsburgh, PA, 1990.
26. Rich, Elaine, and Kevin Knight, *Artificial Intelligence*, Singapore: McGraw-Hill, Inc., 2nd ed., 1991.
27. Rivest, Ronald L., "Learning Decision List", in *Machine Translation*, 1987.
28. Sekine, Satoshi, and Jun-Ichi Tsujii, "Automatic Acquisition of Semantic Collocation from Corpora", in *Machine Translation*, Netherland: Kluwer Academic Publishers, 1995.
29. Sells, Peter, *Lectures on Contemporary Syntactic Theories*, United States: Center for the Study of Language and Information, 2nd Ed., 1985.
30. Sinclair, John, *Collins Cobuild English Language Dictionary*, Great British: William Collins Sons & Co Ltd., 1988.
31. Sumita, Eiichiro, and Hitoshi Iida, "Example-Based NLP Techniques - A Case Study of Machine Translation", in *Statistically-Based Natural Language Programming Techniques – Papers from the 1992 AAI Workshop Technical Report W-92-01*, Menlo Park, California: AAI Press, 1992.
32. Tang, Ting-Chi, "Contrastive Approach in Biligual Education", *Studies in Chinese Syntax: Monographs on Modern Linguistics*, Taipei: Students Book Co., 1979.
33. Yarowsky, David, "Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora", in *Proceedings of 30th Annual Meeting of of the Association for Computational Linguistics*, Nantes, 1992.
34. Yarowsky, David, "Decision List for Lexical Ambiguity Resolution: Application to Accent Restoration in Spanish and French", in *Proceeding of 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, NM, 1994.
35. Yarowsky, David, "A Comparison od Corpus-based Techniques for Restoring Accents in Spanish and French Text", in *Proceeding of 2nd Annual Workshop on Very Large Corpora*, Kyoto, Japan, 1994.
36. Zernik, Uri, "Introduction", in Uri Zernik, editor, *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.
37. 柯淑津, "英中辭彙對應與語意分析之研究", 行政院國家科學委員會專題研究計劃成果報告 NSC82-0408-E-007-195, 1993.
38. 張芳杰, 牛津高級英英英漢雙解辭典, 台北: 東華書局, 6 版, 1989.
39. 梅家駒, 竺一鳴, 高蘊琦, 殷鴻翔, 同義詞詞林, 台北: 東華書局, 1993.
40. 許曙峰, "英中機器翻譯—介詞組分析、轉換、生成", 清華大學資訊科學研究所碩士論文, 1994.
41. 陳正誼, "英中機器翻譯—中文補詞增刪之研究", 清華大學資訊科學研究所碩士論文, 1994.