

# 基於分散式主從架構之中文口語交談系統的設計

彭崇銘 王慧明 林一中 江東輝 張照煌

工業技術研究院電腦與通訊研究所

## 摘要

為了要提供多樣性的資訊存取服務，本文完成一個分散式主從架構的中文口語交談系統。在此架構下，語音辨識、語言理解、交談管制、語言生成和語音合成均分別在獨立的伺服器中執行；並且由交談管制伺服器透過資訊伺服器來存取分散在網路端的資料庫。此外，本文亦說明如何利用交談狀態網路配合所提出的工作描述表格介面及領域剖析器，使交談系統可以有效率地移植到不同的應用領域。我們已成功地建立一個分散式主從架構的中文氣象查詢口語交談系統外，也初步完成將此架構移植到鐵路資訊服務應用領域。

## 1. 前言

電腦工業的快速發展，使得電腦不僅在科學研究上佔有重要的地位，也進入到產業及家庭，在人們的日常生活中扮演不可或缺的角色。但是由於電腦一直欠缺人性化的人機介面，使得它僅能為受過專業訓練的人們所利用。長久以來人類一直期盼著電腦能夠了解人類的語言，讓每個人都能以自然的口語交談方式與電腦溝通。這樣的期盼在網際網路盛行的今日更顯其急迫性。由於網際網路超越空間限制的特性，使得參與網際網路活動的人口快速增加，網路上所提供的資訊更是呈爆炸性的成長。在這浩瀚的網路空間中，使用者往往需要不斷地重複關鍵字的搜尋及網頁的點選來表述限制的條件，篩選所需的資訊及獲得所需的服務。這使得使用者必須耗費大量時間及精力。事實上，用口語來表述限制的條件是最自然且最有效率的方式。例如、「我要一張明天上午八點台北到高雄的機票」這句話就已清楚表達使用者所需機票的起訖地點、出發時間及數量。因此以口語交談方式提供服務的電腦人機介面將是未來發展的趨勢。

口語交談系統的研究所牽涉的課題相當廣泛，包含了語音辨識(Speech Recognition) [1][2][3]、語言理解(Language Understanding) [4][5]、交談管制(Dialogue Management) [6]

[7][8][9]、語言生成(Language Generation)[10]和語音合成(Text-To-Speech)[11]。此外，當系統提供多樣性及多語言服務時，就必須發展可移植性(Portability)[12]的技術。

本文所要討論的重點便是設計一個適合多領域的口語交談系統架構。目前雖然已有許多口語交談系統問世，但大多是以單機作業方式，系統所有的資源都放在同一電腦站台內。未曾考慮模組分割及各模組間網路傳輸的問題。所以單機作業方式架構簡單，較適合於雛形系統的發展。我們也依此架構完成了一個應用於氣象資訊查詢的口語交談雛形系統[13]。

然而因系統未適當模組化及資源無法分享，上述架構會造成系統擴充至多領域時的移植困難及資源浪費。因此，本文提出以分散式主從架構(Distributed Client-Server Architecture)，做為多應用領域之口語交談系統的架構。在此架構下，語音辨識、語言理解、語言生成及語音合成等部分因與應用領域不直接相關，可做為不同領域之交談系統共享部分，因此分別獨立成為伺服器。此外，為使不同形式的資料庫能為系統所利用，必須對個別的資料庫建立伺服器，提供統一的資料查詢介面。系統則以交談管制伺服器連結各個用戶端，使用者之話語經語音辨識及語言理解伺服器分析，至相關之資訊伺服器查詢，並將結果經語言生成及語音合成伺服器以適當的方式回應使用者。

在口語交談系統中，交談管制器控制整個交談流程，為系統之核心。由於一般的交談系統在設計上將應用領域的知識與交談流程的控制緊密結合，因此當系統移植到不同應用領域時，交談管制器需要重新設計，造成移植時的負擔。例如在以語音意向型態網路(SAT Transition Network)[14]架構下之交談管制機制係透過應用領域之關鍵詞組萃取(Key-Phrase Spotting)及關鍵詞組與使用者意向的關聯性來設計控制交談流程。但是當系統移植至其他應用領域時，所需之關鍵詞組必須重新蒐集定義，關鍵詞組與使用者意向的關聯性須重新估算，而針對不同意向的回應亦須重新指定。

為切斷交談管制器與應用領域間之關聯，本文提出一個以交談狀態網路(Dialogue State Network)來描述交談流程，用以驅動交談管制器。網路中每一狀態(State)表示交談流程不同之對答狀態，並指定回應之方式。因此，當系統增加應用領域時，只須增加交談狀態網路，無須更動交談管制器。根據上述架構，我們已成功地完成一個分散式主從架構之氣象資訊查詢的中文口語交談系統，其交談管制伺服器是以交談狀態網路所驅動。系統目前已初步完成移植至鐵路訂票資訊服務之應用領域。

## 2. 分散式主從系統架構

工研院電通所自 86 年 7 月開始執行智慧型語音及語言處理技術計劃，已完成了一個應用於氣象資訊查詢的口語交談雛形系統(Chiang et al, 1998)。此系統採單機作業方式，系統架構如圖一。系統將使用者的問話先經語音辨認模組轉換成文字，之後語言理解模組利用詞彙、語法及語意的知識將文句表達之意念以語意框架(Semantic Frame)表示。交談管制模組再根據語意框架作回應，並將回應以語意框架的形式交由語言生成模組產生文字，最後由語音合成模組合成語音對使用者做出回應。

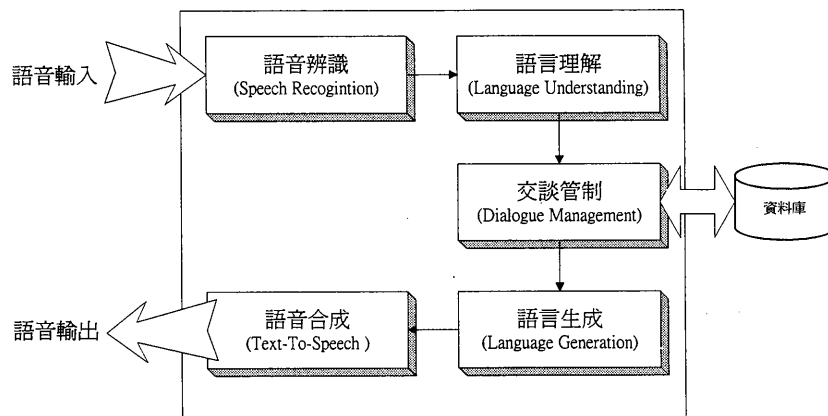
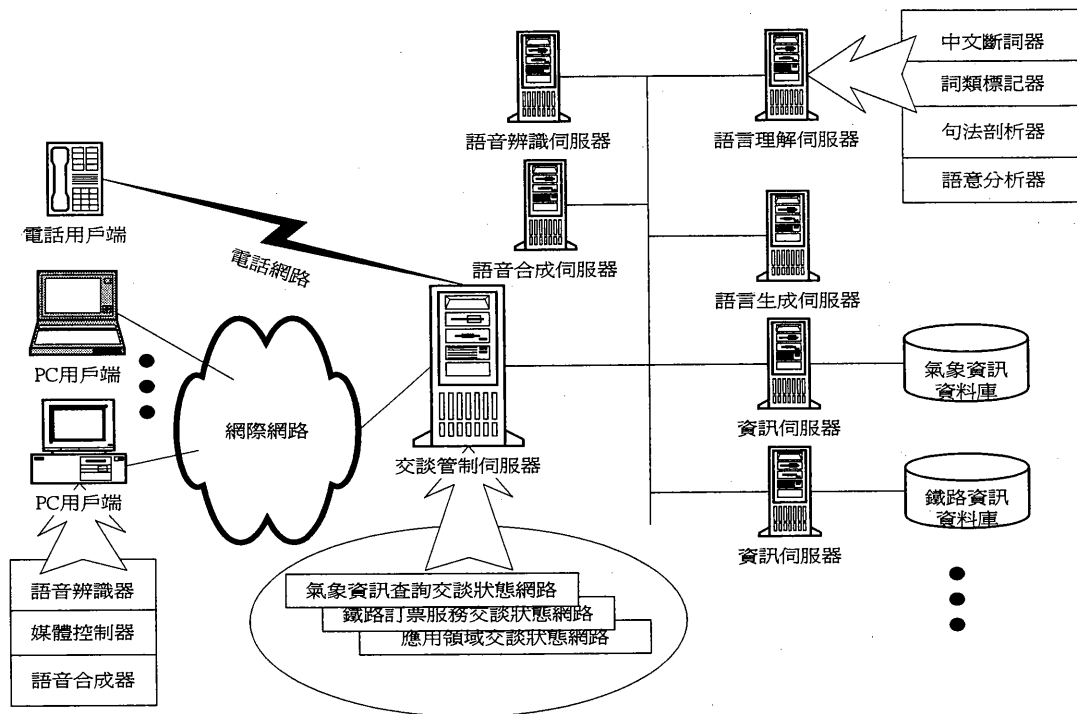


圖 一、單機作業模式之口語交談系統架構圖。

由於上述單機作業架構下系統各個處理模組及資料庫皆集中於同一電腦內，各用戶必須有完整之系統及資料庫方能運作，當系統擴充至多領域時會造成資源的大量浪費。因此，本文提出利用分散式主從架構(Distributed Client-Server Architecture)做為多應用領域之口語交談系統的架構，使系統各處理模組及資料庫能分散於網路並為各用戶端所共享。本系統之架構如圖二所示。系統連線方式分為兩種，一為電話連線另一為網際網路連線(採 TCP/IP 協定)。系統流程部分首先將用戶端的輸入轉換成文字字串，其後送到語言理解伺服器作語言分析，其所得結果以語意框架形式傳送給交談管制伺服器。此時交談管制器需判斷使用者所提問題之應用領域，再根據對應之交談狀態網路對使用者作出回應。其回應動作可能為向使用者要求澄清或為對資訊伺服器提出查詢請求。最後將回應交由語言生成伺服器產生語句並傳回到用戶端。



圖二、分散式主從架構之中文口語交談系統。

基本上系統在考量分割用戶端和伺服器端時，是以資源分享及網路傳輸兩點為依據。在網路連線的情況下，我們希望在用戶端和伺服器端間儘量以傳遞資料量小的文字資料來作溝通，故將掌管影像及聲音之媒體控制器、語音辨識器及語音合成器納入用戶端。而可共享之資源包含交談管制器，語言理解器及語言生成器獨立成為伺服器端。但在電話連線的狀況下，系統於用戶端須以語音訊號直接溝通，因此另須建立語音辨識伺服器及語音合成伺服器，做為與用戶端電話連線之用。此外、為使不同格式的資料庫能為系統利用，須對個別的資料庫建立伺服器，提供統一的資料查詢介面。而由網路上提供的資訊的不同網站可視為不同形式的資料庫。以下就用戶端、語言理解伺服器及語言生成伺服器詳細論述。交談管制伺服器則於第三節介紹。

## 2.1 用戶端之組成

用戶端為面對使用者的最前線，本系統提供包括語音、鍵盤和滑鼠等多型態(Multi-modal)的輸入模式，及文字、圖形、影像視訊和語音之多媒體(Multi-media)輸出模式。本系統之語音辨識器採用 CCLMDS'96 System[3]，具非特定語者(Speaker-Independent)及大中文字彙庫之語音辨識器。語音合成器採用 CCL/ITRI 中文語音合成器[11]。另外為了讓使用者對查詢的結果能一目了然，也加入多媒體的輸出，此項功能由媒體控制器達成，而其

多媒體資料須更新者將由伺服器端傳送，其它則可由用戶端本機取得。為因應用戶直接以電話交談之需要，系統亦建立語音辨識伺服器及語音合成伺服器，做為電話網路連線之用。

## 2.2 語言理解伺服器

語言理解伺服器之功能為分析並了解語句，藉由不同階層的語言分析(包含斷詞、詞類標記、語句剖析及語意分析)，將文字符號最後轉換成代表意念的語意框架(Semantic Frame)。圖三為本伺服器計劃中之架構圖，目前則是以中文斷詞器、詞類標記器及語意分析器來抽取關鍵詞彙形成語意框架。

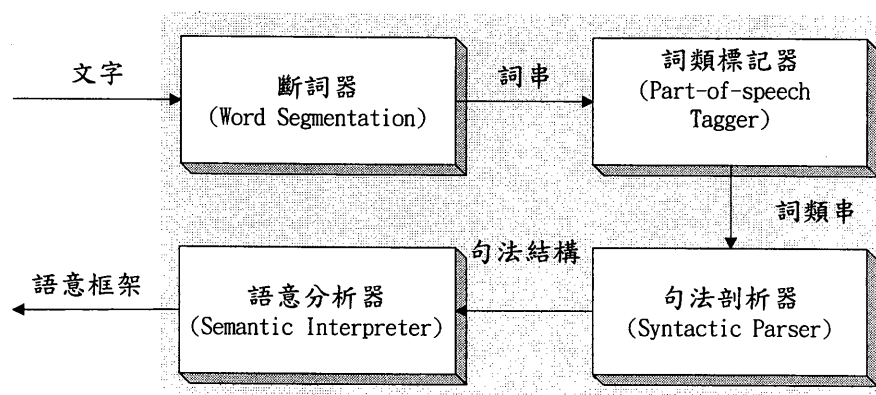


圖 三、語言理解伺服器之架構圖。

### 中文斷詞及詞類標記

詞是語言中攜帶語意的基本單位。而中文在一般語音或文字輸入的狀況下，詞與詞並無明顯的停頓或間隔。因此理解語句的首要步驟就是先將文字分隔成詞串。本系統之中文斷詞器採用詞頻乘績[15]來選擇最有可能詞串。文字經斷詞後，詞類標記器再根據 Trigram 模型[16]來標記各詞詞類，判定各詞之語法功能。隨後，句法剖析器再根據語法規則分析語句之語法結構[17]。由於本系統之句法剖析器尚在發展之中，目前系統直接將詞類標記結果交由語意分析器分析語句之意念。

## 語意分析

語意分析器應根據語句之語法結構及各詞之語意特徵(Semantic Feature)將語句表達的意念以語意框架(Semantic Frame)表示[4]。由於目前句法剖析器尚未完成，系統目前的語意分析器是以語句中各詞之詞類及其可能之語意特徵來抽取關鍵詞彙填入適當的語意框架。以「請問 台北 明天 的 氣溫」為例，其中「台北」、「明天」及「氣溫」皆為名詞，其語意特徵分別註記為地方名詞、時間名詞及氣象名詞。該句之語意框架則為

```
[
  CASE_FRAME: Query_Weather
  Location: 台北
  Date: 明天
  Topic: 氣溫
]
```

透過這個語意框架，系統能夠了解使用者所欲查詢氣象資料的地點、時間及主題。

### 2.3 語言生成伺服器

語言生成伺服器的功能是根據交談管制伺服器傳來的回應語意框架(Response Semantic Frame)來產生適當的文字用以回應用戶端。目前系統的語言生成伺服器是以樣本導向(Template-based)設計，於眾多的訊息樣本(Message Templates)中，語言生成伺服器根據系統回應的語意框架，選出適當的訊息樣本作為生成語句的依據。為提供生成語句的多樣性，對於代表某一意念的語意框架可能有不同的訊息樣本可供選擇，例如適合語意框架

```
[
  CASE_FRAME: Query_Response
  Location: 台北市
  Date: 今天
  Topic: 天氣
  [
    Status: 晴天
  ]
]
```

之訊息樣本有

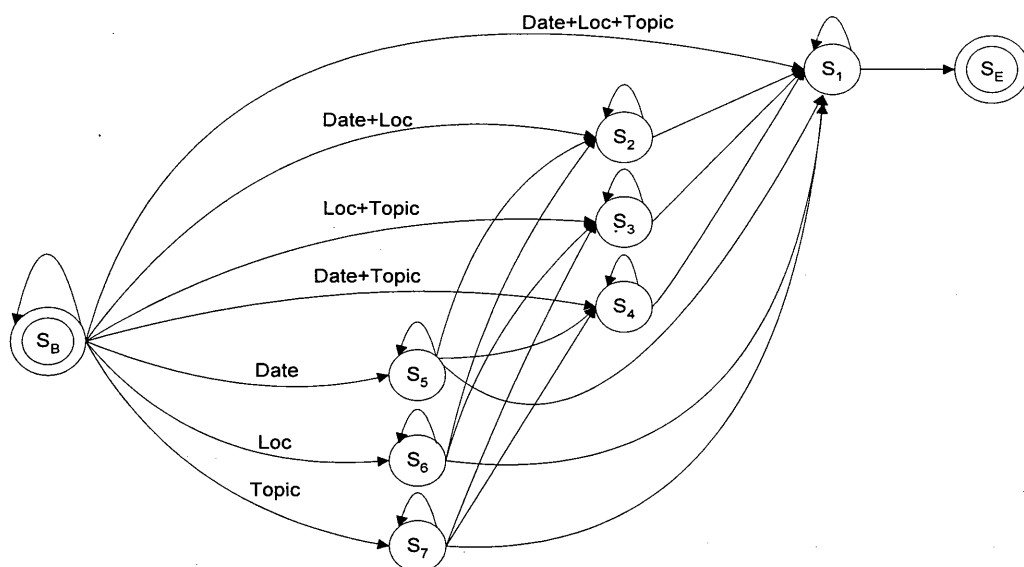
```
Query_Response: Location Date Topic "是" Topic.Status
Query_Response: Date Location Topic "是" Topic.Status
```

等兩條，分別可產生「台北市今天天氣是晴天」及「今天台北市天氣是晴天」。目前針對氣象資訊查詢領域所設計之訊息樣本數共有三十八條，系統以隨機方式在適合的訊息樣本中擇一產生語句。

### 3. 非限定領域之交談管制伺服器

透過前述的資訊伺服機制，可以讓使用者查詢系統中不同應用領域的資料；同時，在這樣主從的架構中，系統可以增加其資訊變動的彈性。然而在一般的口語交談系統中，交談管制器往往與應用領域密不可分。領域移植時，系統同時須更換交談管制器的設計，是系統移植之一項負擔。

為了解決領域可移植性的問題，如圖一所示，我們利用一個以交談狀態網路(Dialogue State Network)來描述交談流程，用以驅動交談管制器。一個交談狀態網路代表一個特定應用領域，網路中每一個狀態(State)代表目前對話進行的狀況。根據系統對特定領域之對話策略及流程，系統發展者可在每個狀態中設計應對的動作(Action)，當交談進行到某個狀態時，交談管制器必須接受該狀態所指定的動作作回應，例如跟使用者要求更多的資訊，或查詢資料庫等。如此系統與使用者一來一往的交談中，交談狀態網路同時會依交談的過程不斷轉移(State transition)其目前狀態，直到到達最終狀態(Final



圖四、交談狀態網路(Dialogue State Network)。

state)，即完成一個特定需求的對話。因此在特定應用領域移植過程中，只需更換交談狀態網路而不需更動整個交談管制器。使得交談管制器與應用領域彼此間互相獨立，增進系統之可移植性(Portability)及可擴充性。

圖四是氣象資訊查詢應用領域的交談狀態網路的例子。本例子中狀態的描述是以使用者是否提供時間、地點及主題三項目之資訊來訂定，總共有八個狀態，每個狀態包含系統對應的反應要求使用者提供所缺少之資訊的動作等。此八個狀態的轉移會由使用者輸入的資訊來決定，例如狀態  $S_3$  表示系統已取得地點及主題兩項資訊，所以會要求交談管制器詢問使用者欲查詢之時間；狀態  $S_1$  則表示系統已取得所需資訊可進行氣象資料庫之查詢。

對於一個較複雜的應用領域，交談狀態網路的設計也會變得複雜，例如前述的氣象資訊查詢，包含三個須使用者輸入的變數；而本系統目前移植的鐵路訂票服務，則包含九個需使用者輸入的變數，設計上的複雜度及困難度大大的增加。為解決這樣的問題，本系統提供了一個「工作描述表」(Task Description Table, TDT)的介面，幫助系統發展者以填寫表格方式設定想要處理的交談狀態及相對之回應動作，以便快速定義出該領域之交談狀態網路。另外本系統提供領域剖析器(Domain Parser)將發展者填寫之表格編譯成交談狀態網路，以便交談管制器使用。如此便能降低系統在領域移植時所需花費的代價。

|       | Date | Loc | Topic | ACTION              |
|-------|------|-----|-------|---------------------|
| $S_B$ | -    | -   | -     | Ask(Date,Loc,Topic) |
| $S_1$ | +    | +   | +     | Query()             |
| $S_2$ | +    | +   | -     | Ask(Topic)          |
| $S_3$ | -    | +   | +     | Ask(Date)           |
| $S_4$ | +    | -   | +     | Ask(Loc)            |
| $S_5$ | +    | -   | -     | Ask(Loc,Topic)      |
| $S_6$ | -    | +   | -     | Ask(Date,Topic)     |
| $S_7$ | -    | -   | +     | Ask(Date,Loc)       |
| $S_E$ |      |     |       | Bye()               |

表 一、工作描述表(Task Description Table, TDT)。

以氣象資訊查詢應用領域為例，系統發展者可利用 TDT 介面，建構一個「工作描述表」，如表一。其中每一欄代表本領域所關切的資訊，有時間(Date)、地點(Loc)及主題(Topic)三項目。每一列代表一個狀態，每個狀態的描述可以是對每個項目的不同運算(如有(+)、無(-)、不理會(Don't care))或限制(如數值的比較或字串的比較)作設定；而最後一欄是該狀態對應的回應動作。此表格再經領域剖析器編譯成交談狀態網路。根據上述的 TDT 介面與領域剖析器目前已成功地使得交談管制器可以提供鐵路訂票之作業。

#### 4. 系統評估

為測試目前系統之效能，我們以 Wizard of OZ 方式蒐集 20 位使用者，共 322 組氣象查詢問答語料。其中 266 組於第一次問答完成查詢，其餘 56 組查詢則於第二至第三次問答完成。我們利用文字輸入方式，以 322 句使用者第一次詢問語句測試系統，其中 239 句獲得系統正確回應，正確率為 74.2%(詳見表二)。表三則列示語言理解伺服器之效能。於 322 句測試語句中，共有 793 個格位(Case)，而系統分析輸入的 322 句得到 793 格位，其中正確之格位有 735 個，召回率(Recall Rate)為 88.3%，精確率(Precision Rate)為 92.7%。

|       |       |
|-------|-------|
| 輸入句數  | 322   |
| 正確回應數 | 239   |
| 正確率   | 74.2% |

表 二、使用者第一回問話文字輸入測試結果。

|        |       |
|--------|-------|
| 輸入之格位數 | 832   |
| 產生之格位數 | 793   |
| 相符之格位數 | 735   |
| 格位召回率  | 88.3% |
| 格位精確率  | 92.7% |

表 三、語言理解伺服器效能測試結果。

| 錯誤型態 | 數目 | 比例    |
|------|----|-------|
| 插入   | 7  | 6.7%  |
| 刪除   | 46 | 44.2% |
| 取代   | 51 | 49.0% |

表 四、錯誤型態及其統計。

基本上格位錯誤的樣式可分為三類，分別為插入性(Insertion)錯誤、刪除性(Deletion)錯誤及取代性(Substitution)錯誤，如表四所示。其中大部分為刪除性錯誤及取代性錯誤，共佔 93.3%。刪除性錯誤肇因於間接的詢問，例如使用者要問明天的氣溫常以「請問明天熱不熱」來代替「請問明天的氣溫」。解決這種錯誤須從延伸多樣的語意框架著手。取代性錯誤則常發生於長問句，因為長句的詞數多，增加正確安插格位的困難，針對此種錯誤我們將以更精細的語言分析來克服，例如分析語句之語法結構。

目前我們除了已成功建立一個分散式主從架構之氣象資訊查詢的中文口語交談系統外。並利用工作描述表及領域剖析器成功產生鐵路訂票及資訊查詢之應用領域之交談狀態網路，已初步完成將系統移植至鐵路訂票及資訊查詢之應用領域。

## 5. 結論及未來發展方向

本文提出以分散式主從架構(Distributed Client-Server Architecture)，做為多應用領域之口語交談系統的架構。在此架構下，語音辨識、語言理解、語言生成及語音合成等部分分別獨立成為伺服器，為不同應用領域之用戶端所分享。

此外，由於一般的交談系統在設計上將應用領域的知識與交談流程的控制緊密結合，增加系統移植時的負擔。因此本文亦提出以交談狀態網路(Dialogue State Network)來描述交談流程，用以驅動交談管制器。當系統增加應用領域時，只須增加交談狀態網路，無須更動交談管制器。配合 TDT 介面及領域剖析器的設計，讓系統發展者更容易完成領域移植的工作。

目前我們了已成功建立一個分散式主從架構之氣象資訊查詢的中文口語交談系統外，並已初步完成將系統移植至鐵路訂票及資訊查詢之應用領域。在氣象資訊查詢應用領域的測試中，系統能對 322 句文字輸入的測試語句中的 239 句做正確回應，正確率為 74.2%。造成錯誤的語句大部分是長問句及間接的詢問句，需要更精確的語言分析才能克服。因此系統未來將著重於語法結構分析及語意分析的結合。

## 致謝

本文係工研院電通所執行經濟部委託之前瞻性資訊與網路技術研究專案 3P11200 的計劃成果之一。特別感謝語音辨識小組同仁彭吳忠謀、張森嘉、簡世杰、陳春賢、周思誠及語音合成小組同仁郭志忠、馬坤源在系統完成期間的鼎力幫忙；而網際網路瀏覽器版本的部分則是由鍾錦鈞獨自設計完成在此一併致謝。

## 參考資料

- [1] J.-W. Hung, J.-L. Shen and L.-S. Lee, "Improved Robustness for Speech Recognition Under Noisy Conditions Using Correlated Parallel Model Combination," Proc. ICASSP98, pp. 553-556.
- [2] Y.-R. Wang and S.-H. Chen, "Mandarin Telephone Speech Recognition for Automatic Telephone Number Directory Service," Proc. ICASSP98, pp. 841-844.
- [3] T.H. Chiang et al. "CCLMDS'96: Towards a Speaker-Independent Large-Vocabulary Mandarin Dictation System," Proc. ICASSP97, pp.1799-1802.
- [4] T.-H. Chiang and K.-Y. Su, "Statistical Models for Deep-structure Disambiguation," In Proceedings of the 4th Workshop on Very Large Corpora," pp. 113-124, Copenhagen, Denmark, August, 1996.
- [5] W. Minker, "Stochastically-Based Natural Language Understanding Across Tasks and Language," Proc. Eurospeech97, pp. 1423-1426.
- [6] H.C.Wang and J.F.Wang. "A Telephone Number Inquiry System with Dialogue Structure," Proc. ICASSP98, pp. 193-196.
- [7] E. Levin, R. Pieraccini and W. Echert, "Using Markov Decision Process for Learning Dialogue Strategies," Proc. ICASSP98, pp. 201-204.
- [8] J.-U. Moller, "DIA-MOLE: An Unsupervised Learning Approach to Adaptive Dialogue Models for Spoken Dialogue Systems," Proc. Eurospeech97, pp. 2271-2274.
- [9] 吳宗憲,「航空資訊服務系統之口述對話模型」,Proc. COTEC'98, pp. E2-4.1-E2-4.12. 1998.
- [10] J. Glass, J. Polifroni and S. Seneff, "Multilingual Language Generation Across Multiple Domains," Proc. ICSLP94, pp. 983-986.
- [11] 郭志忠,「文字轉語音系統之一般性架構」,電腦與通訊, Vol. 56, pp. 32-37, 1997.
- [12] C. Wang et al. YINHE : "A mandarin Chinese Version of the GALAXY System," In Proc. Eurospeech97, pp. 351-354. 1997.
- [13] T. -H. Chiang et al. "The Design of a Mandarin in Chinese Spoken Dialogue System," Proc. COTEC'98, pp. E2-5.1-E2-5.7. 1998.

- [14] 李琳山,「國語對話技術之初步研究」,交談系統暨語境分析研討會, Taipei, Taiwan, 1997。
- [15] T.-H. Chiang, J.-S. Chang, M.-Y. Lin and K.-Y. Su, "Statistical Models for Word Segmentation and Unknown Word Resolution," in Proceedings of the 5th R.O.C. Computational Linguistics (ROCLING V) Conference, pp. 123-146, Taipei, Taiwan, ROC, 1992.
- [16] K.W. Church, "A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text," in Proceedings of the 2nd Conference on Applied Natural Language Processing, pp. 299-307, February , 1988.
- [17] T.-H. Chiang and Y.-C. Lin and K.-Y. Su, "Robust Learning, Smoothing, and Parameter Tying on Syntactic Ambiguity Resolution," Computational Linguistics, Vol. 21, No. 3, pp. 321-349, 1995.