



國立中山大學資訊工程學系

論文計畫書

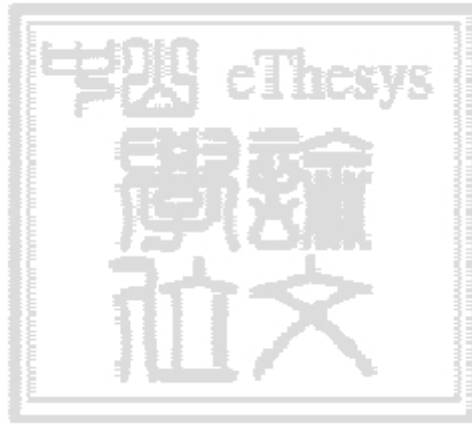
線上機器翻譯系統之整合

Combining outputs from online translation systems

研究生：陳逸昌 撰

指導教授：陳嘉平 博士

中華民國九十八年三月



© 中華民國九十八年三月

陳逸昌

All Rights Reserved

摘要

中文摘要 . . .

關鍵詞： . . .

ABSTRACT

Abstract in English . . .

Keyword: . . .

Contents

List of Figures	ii
Chapter 1 Introduction	1
1.1 Motivation	1
1.2 Introduction of the System	1
1.3 Organization of the Thesis	2
Chapter 2 Related Works	3
2.1 Sentence level combination	3
2.2 Phrase level combination	4
2.3 Word level combination	5
2.4 Discussions	6
Chapter 3 Methods	7
3.1 System architecture	7
3.1.1 Substitution	7
3.1.2 Subection	8
3.1.3 Insertion	9
3.1.4 Deletion	9

List of Figures

2.1	Sentence level combination 架構	4
2.2	Phrase level combination 架構	5
3.1	系統架構	8

Chapter 1

Introduction

1.1 Motivation

隨著科技的進步，我們能夠接觸到的眼界也越來越廣泛。在現實生活之中，所能接觸到不單只是本國人，也可能會接觸外籍人士。在這樣的情況之下，如何能夠與外國人士溝通就是所要面臨的重要課題，如果能有專業的翻譯在其間居中協調，就能使得人們的溝通更加無國界化，可惜的是傳統人工翻譯，需花費較高的成本，且不符合對話者的私密性，而其花費更不是一般社會大眾所能負擔，因此利用機器翻譯系統能夠有效節省花費成本，更能夠勝任這樣的重要角色。

如今隨著網路的日益發達，可以察覺人們對於網路日趨依賴，越來越多的網路應用隨著時代潮流逐漸興起，商用機器翻譯廠商也樂意提供網路平台，以利於人們免費使用機器翻譯系統，但是顯而易見的是不同廠商的機器翻譯系統，都有著其各自擅長的領域以及句型，但這同時也意味著沒有一個線上機器翻譯系統是能夠將翻譯原始句上傳至翻譯網頁，就能確保得到信賴的結果。在這樣的情況之下，如何能夠利用不同線上翻譯系統將之整合成單一翻譯結果，以此方法可以得到一個值得信賴的翻譯結果，可以省去使用者大部份的比對時間，因此本研究所欲探討的問題即是線上翻譯系統的整合應用。

1.2 Introduction of the System

目前有很多的研究都發現，多個機器翻譯系統的整合，相較於單一機器翻譯系統的結

果為佳。因此本研究中，將提出一個機器翻譯整合系統，針對不同的線上翻譯系統進行整合。其大致流程如下：首先利用SRILM[1]所訓練之語言模型當作評分標準，決定系統所認定最佳之翻譯，更進一步參考了語音辨識中WER(Word Error Rate)的方法，將線上機器翻譯系統錯誤分成三大類，分別是Substitution、Insertion及Deletion，分別針對這三種不同的錯誤來進行處理。

在Substitution方面，本系統引進了其他翻譯句中可能是正確翻譯的字，但是卻沒有出現在最佳翻譯，考慮該字為系統誤譯成其他字因此尋找最佳翻譯中可能的錯誤翻譯，將之進行替換。

在Insertion方面，則是利用其他翻譯假說中緊連的字考慮其為片語的可能性，但是卻沒有同時出現在最佳翻譯中，以此考慮最佳翻譯中的字並將之延伸成片語可能性。

在Deletion方面，則是進一步將翻譯假說中多餘翻譯字，或者是錯誤翻譯字進行刪除的動作。

1.3 Organization of the Thesis

以下為本研究的編排方式：在第二章，講述關於本研究中一些相關之文獻探討，第三章部份講述本研究所用翻譯系統整合流程及方法，第四節部份講述針對其所做之評估結果，第五節則做為總結。

Chapter 2

Related Works

機器翻譯是利用電腦自動模擬產生人們文本翻譯的結果，本研究將目前的機器翻譯系統分成三大類：Rule-based machine translation, Example-based machine translation及Statistical machine translation分別敘述如下：

Rule-based machine translation：這類方法是將翻譯的規則以條列式的方式列出，翻譯的過程中遇到何種情況，則依照規則採取不同的行動。

Example-based machine translation：這類方法則是利用訓練語料中有出現過最符合的句子當作樣本來進行翻譯，因此如果是出現在訓練語料中的句子，此種方法相較於其他方法正確性更高。

Statistical machine translation：這類方法則是利用訓練語料，訓練出不同的翻譯模型，之後利用這些翻譯模型來對翻譯原始句進行翻譯。

以上三種不同的方式，各自有其優點及缺點，因此才會有機器翻譯整合系統的延伸，希望能夠合併各種不同系統的優點，來補足單一系統的不足，而在[2]中，將機器翻譯整合系統分成三個不同的層級分別是：Sentence level combination, Phrase level combination 及Word level combination，以下將依照這三種不同的分類，進行深入的探討。

2.1 Sentence level combination

Sentence level combination，是利用各種不同的特徵函數，利用Minimum Error Rate Training[3]來取得各個特徵之間的權重，並以此評估各個翻譯假說之好壞，輸出最佳

者當作其翻譯結果[4]，這樣子的系統好處是不需要各個翻譯系統提供任何資料，所有的特徵函數都是利用整合系統自身訓練出來的模型來當作一致的評分標準。而這類系統的特性為所產生的句子，並沒有經過重組，因此必定與其中一個翻譯系統的句子相同。

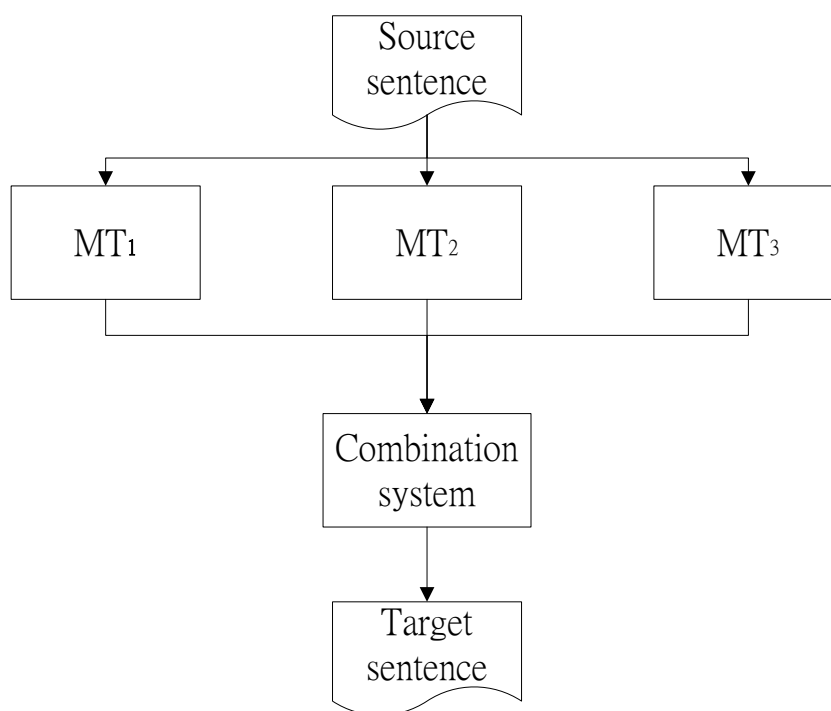


Figure 2.1: Sentence level combination 架構.

2.2 Phrase level combination

Phrase level combination，則是取得各個翻譯系統的片語對，如果無法由各個系統取得片語對，則利用GIZA++[5]對翻譯假說及翻譯原始句進行對齊來取得片語對，之後再利用整合過的片語表來重新進行翻譯[6]，此類方法可以解決訓練語料中，片語對不足的問題。而且其翻譯結果並不一定會與先前的翻譯系統類似，因此有機會會輸出比任何一個翻譯系統都還要良好的結果。

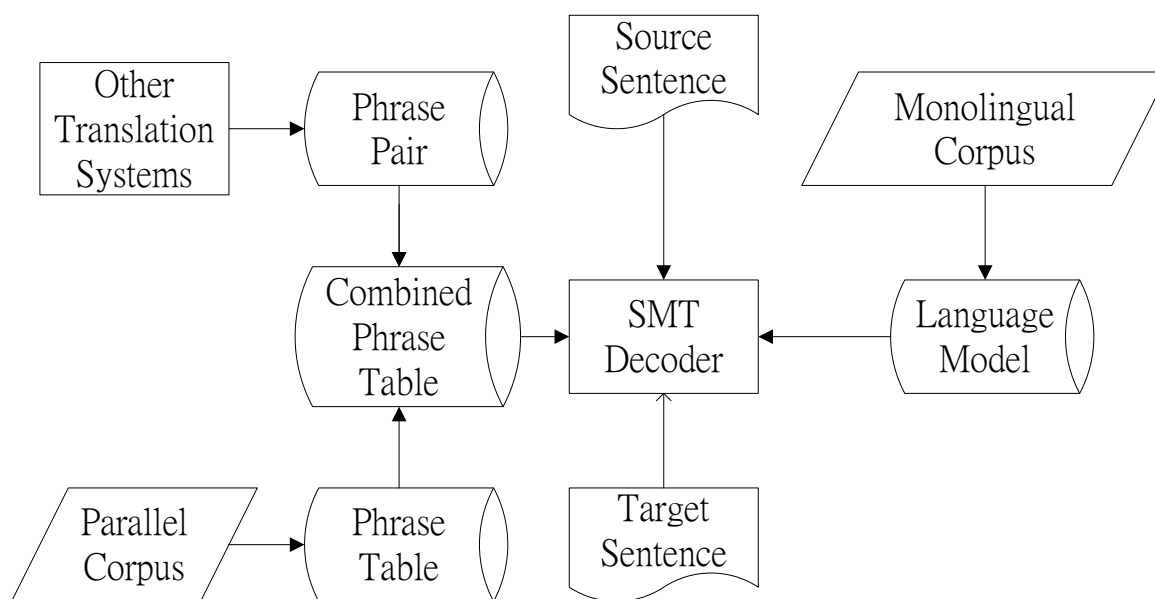


Figure 2.2: Phrase level combination 架構.

2.3 Word level combination

Word level combination方面，主要是將所有的假說，建立成一個confusion network，之後在對這個confusion network計算出最佳的假說，而其一開始的應用並不是被利用於機器翻譯的系統整合上面，而是應用在語音辨識系統整合[7],而此方法應用在機器翻譯的整合系統上時，相較於語音辨識的整合系統，所需要克服的困難點就是字的排列順序，由於語音辨識系統所輸出的假說只會有一種順序，而機器翻譯系統所輸出的假說所輸出的順序可能有極大的不同，因此我們必須先選擇一個翻譯假說當作對齊的標準，之後將其他翻譯假說對齊到這個標準，並建構出一個confusion network，再由這confusion network選擇出最佳的翻譯假說[8]。

而在[2]中，更進一步的以confusion network所得到的翻譯假說當作skelton重新進行alignment，得以進一步提昇正確率。

另外在[9]的研究，採用了另一種Word level的合併方式，這個方法利用了句子中各個字的詞幹資訊以及part-of-speech(POS)的資訊來進行對齊，而這個方法定義了其最佳對齊是兩個句子之間具有最少的交叉邊，並以此建構成一個word lattice，而系統最後搜尋這個lattice，根據language model以及其他的評分標準產生了最佳的翻譯假說。

2.4 Discussions

以上幾種不同層級的合成方法，根據[10]研究嘗試了三種不同層級的結合方法，發現了採用Word level combination的方法能夠得到較佳的結果，這是因為Word level combination其對於翻譯假說的修正較其他層級的整合系統較廣，因此能夠取得相對較高的改進。

Chapter 3

Methods

本章將針對系統的各個流程進行深入的探討

3.1 System architecture

本系統將翻譯原始句利用三個不同的線上翻譯系統取回翻譯假說，進一步利用了語言模型的分數決定了初步的最佳翻譯，之後再利用Substitution模組來取代錯誤翻譯的字，並進一步的利用Insertion模組將沒有被翻譯的字添補進來，最後則是利用Deletion模組將多餘的字刪除，3.1為本系統之系統架構圖。

3.1.1 Substitution

本研究中利用SRILM[1]工具來訓練出一個5-gram的語言模型，而最佳的翻譯假說(E')是根據以下選擇方法，針對top-one的假說來進行選擇。

$$E' = \arg \max_E \frac{1}{I} \sum_{i=1}^I \log P(e_i | e_{i-4}^{i-1}) \quad (3.1)$$

I 為各個翻譯假說的總長度， e_i 為假說中的第 i 個字， e_{i-4}^{i-1} 為 e_i 前面四個字，之所以後來會再以 $\frac{1}{I}$ 對其Normalize，是考量到單純以語言模型分數來當作考量對長句子而言較顯不利，而Normalize過後的分數為各個字平均的分數，以此為基準下去比較，免除了長句不利的考量。

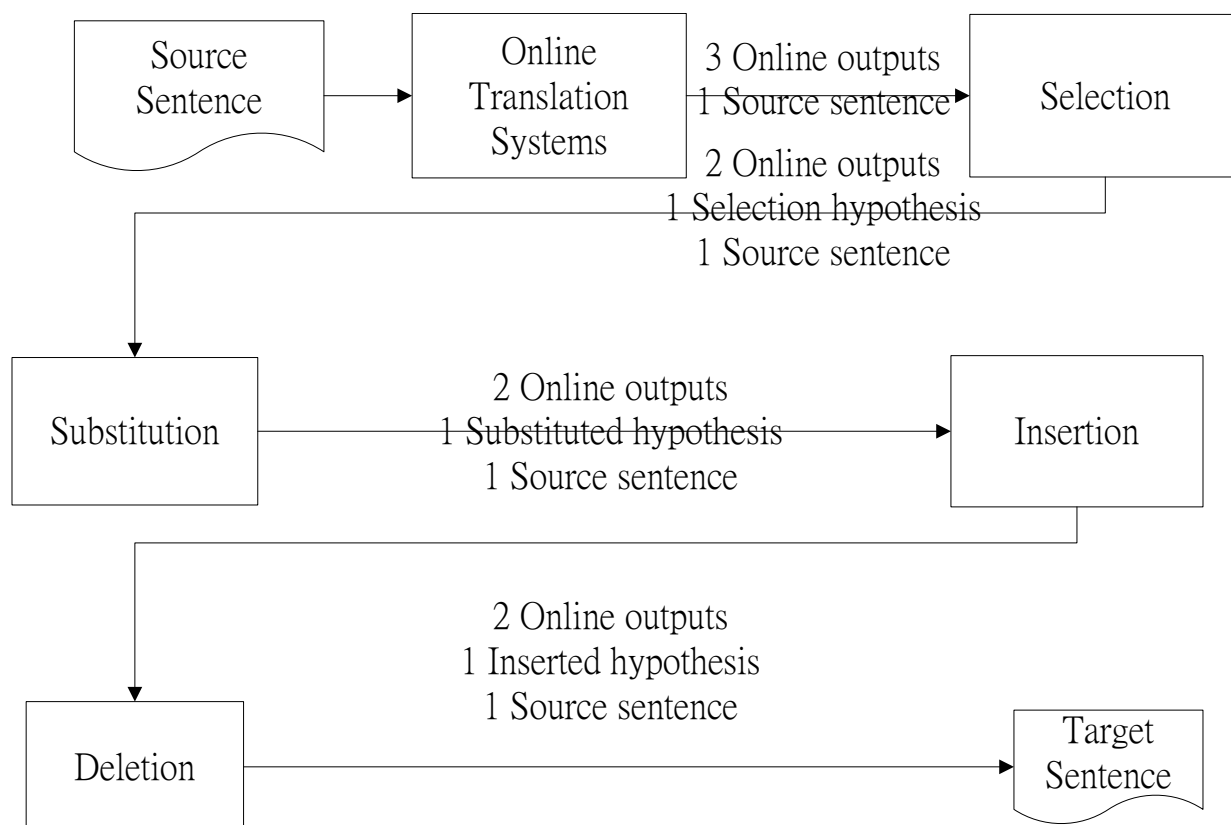


Figure 3.1: 系統架構.

3.1.2 Subection

另外由於觀察發現，相同翻譯原始句所得到的翻譯假說中，同一個字出現在多個翻譯系統的翻譯假說中，則代表這個字為正確翻譯的機率較高，如果沒有出現在選擇的翻譯假說之中，則有可能是所選擇的翻譯假說沒有正確翻譯出來，因此給予一特定的門檻值，如果有超過半數的系統翻譯出相同的字(R)，而選擇的最佳翻譯假說中沒有出現該字，則考慮該字引入置換翻譯假說中字的可能性：

1. 依照GIZA++所訓練出來雙向Translation table，考慮此字(替換字)所對應到翻譯原始句中的字(C_j)(雙向機率最大者)，而再由(C_j)尋找其最有可能對應到翻譯假說中的字 T_i 。
2. 比較 $P(R|C_j)$ 以及 $P(T_i|C_j)$ 彼此之間的機率大小，若 $P(R|C_j) > P(T_i|C_j)$ 則考慮進行替換。
3. 替換的評估如下：假設被替換字(T_i)出現在翻譯假說中第 i 個位置，嘗試以替換字 R 代入後，考慮與翻譯假說中 $i - 1$ 位置的字(T_{i-1})及 $i + 1$ 位置的字(T_{i+1})，利

用Bigram的language model分數來進行比較，比較方式如下：

$$h(x) = \begin{cases} 1 & \text{if } \log P(T_i|T_{i-1}) + \log P(T_{i+1}|T_i) > \log P(R|T_{i-1}) + \log P(T_{i+1}|R) \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

4.若回傳結果為1則以替換字(R)取代被替換字(T_i)

3.1.3 Insertion

假使翻譯句已經包含了共同的翻譯字(R)時，而(R)位於其他翻譯假說(j)之位置時，令 o_j 代表出現在其他假說中的 R ，考慮 o_j 與其前後各一個字(o_{j-1}, o_{j+1})分別為片語的可能性，考慮將之引入翻譯假說中，詳細步驟如下：

1.找到 R 在最佳假說中的位置 i ，以 e_i 表示出現在最佳假說的 R ，計算以下機率考慮是否將 o_j 插入最佳假說中。

$$h(x) = \begin{cases} 1 & \text{if } \frac{\log P(o_{j-1}|e_{i-1}) + \log P(e_i|o_{j-1})}{2} > \log P(e_i|e_{i-1}) \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

2.若回傳結果為1則將 o_{i-1} 插入至最佳假說 $i-1$ 的位置，若是回傳結果為0，則進一步考慮以 o_{i-1} 替換 e_{i-1} 的可能性，以下列式子判斷：

$$h(x) = \begin{cases} 1 & \text{if } \frac{\log P(o_{j-1}|e_{i-2}) + \log P(e_i|o_{j-1})}{2} > \frac{\log P(e_{i-1}|e_{i-2}) + \log P(e_i|e_{i-1})}{2} \\ 0 & \text{otherwise} \end{cases} \quad (3.4)$$

3.若回傳結果為1則以 o_{j-1} 置換最佳假說中的 e_{i-1} ，若是回傳結果為0，則不進行任何動作。

4.同理以類似的方式考慮 o_{j+1} 引進的可能性。

3.1.4 Deletion

Deletion的步驟最主要是為了刪除假說中可能出現的錯誤翻譯，主要是參考

Bibliography

- [1] A. Stolcke, “SRILM-an extensible language modeling toolkit,” in *Seventh International Conference on Spoken Language Processing*, ISCA, 2002.
- [2] N. Ayan, J. Zheng, and W. Wang, “Improving alignments for better confusion networks for combining machine translation systems,” in *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pp. 33–40, 2008.
- [3] F. Och, “Minimum error rate training in statistical machine translation,” in *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pp. 160–167, Association for Computational Linguistics Morristown, NJ, USA, 2003.
- [4] A. Stolcke, “Combination of machine translation systems via hypothesis selection from combined n-best lists,” in *Proceedings of the Eighth Conference of the Association for Machine Translation*, pp. 254–261, 2008.
- [5] F. J. Och and H. Ney, “A systematic comparison of various statistical alignment models,” *Computational Linguistics*, vol. 29, no. 1, pp. 19–51, 2003.
- [6] F. Huang and K. Papineni, “Hierarchical system combination for machine translation,” in *Proc. of the Conf. on Empirical Methods in Natural Language Processing (EMNLP2007)*, 2007.
- [7] J. Fiscus, “A post-processing system to yield reduced word error rates: Recognizer output voting error reduction (ROVER),” in *1997 IEEE Workshop on Automatic Speech Recognition and Understanding, 1997. Proceedings.*, pp. 347–354, 1997.

- [8] B. Bangalore, G. Bodel, and G. Riccardi, “Computing consensus translation from multiple machine translation systems,” in *IEEE Workshop on Automatic Speech Recognition and Understanding, 2001. ASRU’01*, pp. 351–354, 2001.
- [9] S. Jayaraman and A. Lavie, “Multi-engine machine translation guided by explicit word matching,” in *ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, vol. 43, p. 101, 2005.
- [10] A. Rosti, N. Ayan, B. Xiang, S. Matsoukas, R. Schwartz, and B. Dorr, “Combining outputs from multiple machine translation systems,” in *Proceedings of NAACL HLT*, pp. 228–235, 2007.