出國報告（出國類別：其他(業務接洽)）

出訪日本北海道對於 TCUS-RIBS(臺灣綜合大學中英對照法規資料庫)後續建置作業之探討執行

服務機關：國立中正大學國際事務處
姓名職稱：施國際長慧玲教授
工學院黃院長仁竑教授
資訊工程學系熊主任博安教授
派赴國家：日本北海道
出國時間：103 年 06 月 28 日至 06 月 30 日
報告日期：103 年 10 月 12 日

摘　要

2014 年 06 月 28 日至 06 月 30 日，前往日本北海道與名古屋大學團隊、韓國團隊及中國人民大學團隊研討臺灣綜合大學中英對照法規資料庫後續建置之探討執行與研討法律關鍵字英譯辭典建立分享系統，對於網站與資料庫建置以及與各法域、各校合作網絡的後續工作做更進一步的規劃，並透過工作會議整合修正各法域及各校的作業進行方法與內容及強化合作之相關事宜，希望經由定期會議的召開，強化彼此的合作關係，並在每次的例行會議交流中不斷地尋求在建置資料庫中會遇到的各種問題的解決方法，嘗試化解各項難題及保持持續地發展及合作關係。

目　　次

壹、目的

施慧玲國際長、黃仁竑教授與熊博安教授此次出國之目的為參與TCUS-RIBS(臺灣綜合大學中英對照法規資料庫)在北海道舉行之例行會議及第七屆四法域 CJKT 標準翻譯字典研討會(7th Workshop of Standard Translation Dictionary for China, Japan, Korea, and Taiwan)。

主題:目前台灣參與 TCUS-RIBS(臺灣綜合大學中英對照法規資料庫)在北海道舉行之例行會議及四法域法律用語資料建置的工作中，除針對原先日本(主辦國)以韓國官法公佈的法律用語為基礎所整理出來的資料進行台灣相對應的用語的編輯外，亦增加許多司法院、學術期刊的相關資料。在 TCUS-RIBS(臺灣綜合大學中英對照法規資料庫)與四法域進行資料交換或分享是透過計算表(excel)檔案以電子郵件交換，在資料維護與版本的一致性上，面臨了許多的問題。所以在這個研究中，我們因應資料維護的需求，開發了一套資料庫系統，可以很便利地進行資料維護，並保持資料的一致性。在此資料庫的使用者介面上，我們提供完整資料列表、依資料編號或關鍵字搜尋、資料過濾條件、資料編輯、資料匯出成計算表檔案等功能。各領域代表對於這樣的資料庫開發都感到非常便利，希望可以儘快開放提供使用。

緣起：本次會議由日本名古屋大學法律資訊中心協助辦理。TCUS-RIBS(臺灣綜合大學中英對照法規資料庫)研究團隊施慧玲國際長、熊博安主任及黃仁竑院長以上三位教授代表參加。計畫團隊參與一個建置多語法律與社會資訊分享平台的跨國長期合作計畫，逐步建置與分享資訊的法律網站平台與資料庫。

預期效益或欲達成事項：我們採用五個統計的方法，針對一個中文關鍵字，各取得前 5 個最佳的英文對應詞(複合詞，非僅僅是單字)，再利用期望值最佳化(Expectation Maximization, EM)的方法，從 25 個結果中取出最佳的 5 個做為輸出的結果。我們的實驗結果證實這樣的做法，比原來日本所提供的本文中的關鍵字(Key Word in Context, KWIC)方法正確率可提升 10~30%。我們將結合我們之前的關鍵詞自動擷取的結果，做正確率的進一步改進。我們可以得到更正確的結果主要原因是在中文詞的處理上，日本的做法是只採統計為基礎的 n 字詞法(技術用語)，而我們是採機器自動學習的方法加上中文分詞的工具，所以可以大幅降低中文詞的數量，在計算時間及正確率上來說都可以得到改善。

貳、過程

參訪單位：本次為出席會議無參訪單位。

考察單位介紹:名古屋大學日本法律信息研究中心由名古屋大學法律系松浦好治教授主持，由法律專業及資訊科學專業團隊共同組成，自 2008 年成立，在基礎研究方面著力甚深，研究方面不僅在以多語訊息提供日本法律訊息，此外並透過國際合作方式增進國際間的相互理解進而做出貢獻，未來可以有效而靈活地共享大量法律資訊的基礎研究。

訪問過程及考察經過：

本次大會議程如附錄一所示，大會安排與會人員於 2014 年 6 月 28 日入住旅館(即會議地點)，6 月 30 日會議結束。當日會議結束後施慧玲教授及黃仁竑教授與熊博安教授隨即搭機返國，於當日晚上 8:10 返抵桃園機場。施慧玲教授及黃仁竑教授與熊博安教授行程如表 1 所示。

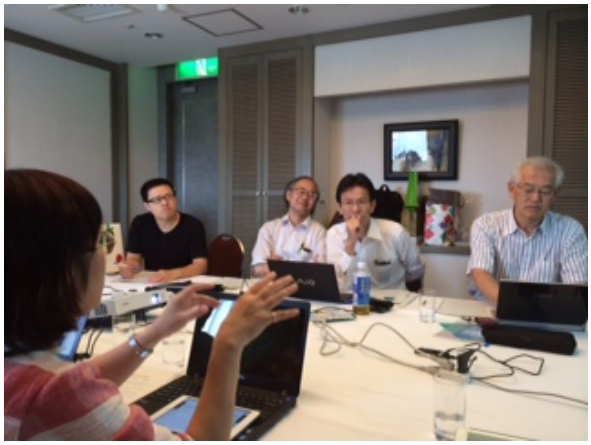
表 1 出席研討會行程表

日期/星期	活動內容	備註
6/28 (週六)	去程(桃園機場-日本札幌) (*國際長已提前前往北海道參與高中訪問之行程，之後接續本行程)	大會晚宴
6/29 (週日)	會議第一天: 黃仁竑教授報告 議題討論	參見附錄二、三的投影片
6/30 (週一)	會議第二天: 熊博安教授報告 議題討論 回程(日本札幌-桃園機場)	參見附錄四的投影片

圖一: 黃仁竑教授於會議第一天進行口頭報告

會議第一天首先由黃仁竑教授針對兩個研究主題進行報告(圖一)。第二個報告是如何從台灣現有的雙語法規資料庫中，建置雙語語料庫(bilingual corpus)，並以兩個不同的方法(使用台灣國際商業機器股份有限公司所提之模式〔IBM model〕及以統計為基礎的方法)，進行中英文關鍵字的自動對應，以協助建立雙語標準詞典(Standard Bilingual Corpus, SBD)。在此報告中，我們針對以統計為基礎的方法報告我們的研究成果。第一個報告是四法域法律用語資料庫建置成果。第一天的會議在黃仁竑教授報告完後，由日本計畫主持人松浦好治教授(Prof. Yoshiharu Matsuura)接續主持會議，開始進行法律用語資料庫的修訂與議題討論(圖二、圖三)。

圖二: 松浦好治教授主持第一天的議題討論



圖三:中正大學法律施慧玲教授對議題發表意見

會議最後一天(六月三十日)，議程只有上半天，從早上九點鐘開始到中午十二點鐘。下午即搭機返國。在這天的議程中，主要由熊博安教授報告四法域之法律用語資料庫未來發展方向中很重要的翻譯平台建置(圖四)。熊教授的報告分為兩個部分。其一，介紹台綜大四校法規雙語資料庫第二年計畫中已經開始建置的法規中英翻譯標準作業(Standard Operating Procedure, SOP)。其二，介紹一套全球專業翻譯人員廣泛使用的翻譯工具 SDL Trados Studio(一項翻譯工具軟體的名稱)。以下說明這兩個部分的主要內容以及大家討論的議題。相關簡報內容在附

錄四。

考察成果及後續辦理或推動事項：

經過約 3 個小時多方的討論，得到以下的共識與結論：

增加新的法律用語，先從各項主要法規著手。於下次會議前，提出新增清單。

(1)對於一個法律用語，是否可以從法規中找出其定義。此問題因不同法域其法規的寫法不一樣，所以不容易正確地做好。例如台灣的法規並沒有一定在法規中對某一用語進行定義，但在日本的法規中，其第二條一定是用語定義。此問題的決議是先暫緩進行。

(2)針對目前的四法域法律用語資料庫的英文譯文進行校對與統一部分用語。將先針對四法域都使用類似的漢字關鍵字進行校對，以韓國的翻譯為主，日本的為輔。

(3)是否對每一個法律用語進行分類? 例如某一用語或部分用語最常出現在民事法中。決議是需先對要如何分類進行討論，各國於下次開會前，先擬定 6 至 7 個類別，建議以六法全書為依據。

(4)目前的法律用語資料庫中是否再增加一個”外國法”的欄位?由於目前有一個”相當於”(equivalent)的欄位會與新增的”外國法”欄位有重疊或難以區分的問題，決議是先暫緩實施。

(5)對於部份重複概念的用語該如何處理? 例如在日本有”公訴人”的用語，但日本沒有自訴的制度。在台灣則有分公訴與自訴，而”檢察官”相當於”公訴人”。那麼”檢察官”是否完全相當於日本的”公訴人”? 並不盡然，因為日本沒有自訴人。決議是在備註欄中加註此一不同(即台灣有分公訴與自訴)。

(6)對於已廢止的法規中的法律用語如何處理? 決議是由日本統一在資料表中多加一行及加以註解。

(7)對於法規用語是否加註其英文詞性，如名詞、動詞、形容詞、副詞? 決議是有必要，但需先思考如何加註較合適。合併產生的問題是複合詞如何處理，例如日本有旅卷、一般旅卷、公用旅卷等(相當於台灣的護照、一般護照、公務護照、外交護照等)? 上例中，很明顯” 旅卷”是最基本的用語，而”一般”、”公用”是形容詞，所以” 一般旅卷”、”公用旅卷”是複合詞。決議是四國法律用語資料庫中只記錄” 旅卷”，將” 一般旅卷”、”公用旅卷”改記錄到”相關資訊”(related information)的欄位中。

考察詳細內容說明：

就 SOP(標準作業流程)而言，我們在台綜大的計畫中設計了一套很完整的流程，可供四校使用的翻譯平台。此平台，結合三種翻譯資源，包含(一)Trados(翻譯工具軟體的名稱)工具、(二)翻譯經理、(三)翻譯專家。這三種資源的使用均納入平台的流程中。流程開始時，任何學校可將欲翻譯的中文法規以及其英文法規(若有)繳交到平台中，經註冊確認為有效的法規文件後即啟動一個新的文件翻譯流程。在此流程中，先經由 Trados(翻譯工具軟體的名稱)工具翻譯，這是第一個版本的翻譯。使用此工具前，工具必須經過訓練。此部分的訓練，將於第二部分的簡報中說明。工具翻譯出來的版本，必須經過翻譯經理修改、確認，因為工具可能會翻譯錯誤或者可能會有模稜兩可的問題發生亦或會有文件格式的問題發

生。這些問題均需要翻譯經理幫忙處理。處理完，此第二版本的翻譯，才能接下來提供給翻譯專家進行專業修飾與確認翻譯品質。專家確認後的版本即是第三版。此第三版本的翻譯，將提供給原大學，以便由原單位確認翻譯的結果是否與原文文件的語意吻合。若有任何問題，翻譯經理將與原學校進行一次以上的修改及確認。最後若無誤，翻譯好的文件才會進入匯入階段，正式匯入資料庫。在整個 SOP(標準作業流程)的流程中，若發生致命的錯誤時，文件翻譯可以直接進入錯誤狀態並終止翻譯。若原學校發現他們並不需要翻譯該份文件，流程中亦保留讓原學校直接取消該文件的翻譯流程。在取消時，翻譯經理會被通知。報告中，熊教授亦針對前端(frontend)使用者介面以及後端(backend)資料庫處理的部分進行說明與介紹。

此項 SOP(標準作業流程)報告完，各與會人員提出相關的問題。例如，日方詢問主要會有哪些錯誤。熊教授和施國際長回覆，這些都是致命的錯誤，例如中英文法規版本不一致，所以目前沒有絕對必要的再繼續進行翻譯。

圖四：熊博安教授於六月三十日早上會場進行報告

第二部分，熊教授介紹 SDL Trados Studio (翻譯工具軟體的名稱)工具。此工具有兩個重點，包含翻譯記憶(Translation Memory)及詞庫(Term base)。翻譯記憶，主要是這套工具經過訓練後可以用來翻譯新的文件的一些翻譯對照。例如，工具可以輸入中文及英文法規文件的對照組，使工具知道這些對照的文件中的中文法規應該如何正確的翻譯成英文。工具具有智慧，他可以學習翻譯過的文件內容，並且利用這些學習過的經驗，翻譯未來新的文件。這就是所謂的翻譯記憶。另外，詞庫是用來儲存兩種以上語言中重要的對照字詞。例如，中文的“大學”與英文“university”可能就是詞庫中對照的字詞。使用詞庫的好處是各校之間的翻譯會比較一致，這樣才不會有相同的中文在各校的英文版本裡翻譯成不同的英文。所以，詞庫也需要在計畫中建構。

工具介紹完，另外三個法域的與會人員均表達強烈的興趣。日本名古屋大學 Toyama 教授表示，熊教授的報告激勵他，讓他在向學校提供相關研究計畫時更具有信心與相關資料。Toyama 教授也立即請求熊教授提供簡報檔。大陸的學者丁教授也表示，人民大學的團隊也希望有一個像我們這樣的翻譯平台。丁教授甚至表示願意到台灣來一起討論方向與經驗。其中一位與會者韓國的洪博士，也表示他們希望韓國政府也可以這樣建構一個完整的翻譯平台。洪博士進一步說明希望台灣可以提供更多除了 Trados(翻譯工具軟體的名稱)以外的其他工具資訊。熊教授在現場就直接幫洪博士找到一份比較各種翻譯的報告書。洪博士亦請求熊教授能提供比較的資料以及當天簡報的資料。

在這樣熱烈的討論與切磋之後，會議主持人松浦好治教授(Prof. Matsuura)做總結論，並表示希望四法域未來可以有這樣的翻譯平台以及希望台灣團隊能陸續提供相關資料給各合作夥伴參考。

圖五：會後合影留念

參、心得

此會議對各與會人員均有很大的幫助。無論從各法域之法律的角度或大學法規的角度來講，很多各項有關語言的問題均需要解決。資料庫的建置是有必要的，若只是從 Excel 檔案中想要整理出很多的資訊不是一件簡單的事情。而且隨著資料的增加，整理的困難度也會暴增。另外，遇到各種語言在用語上的差異等問題會突顯準確翻譯法規的重要性。由於這些問題逐漸慢慢地出現，所以這次會議最大的收穫是釐清這些問題，以及訂定未來的重要發展方向。離開時，大家有一起拍照留念，如圖五所示。回國後，大家開始規畫下次會議的重點，並且鎖定改善法規用語上的種種問題。期盼我們在下一次的會議中可以增加法規翻譯的正確度。

肆、建議事項

1.在資訊方面循序漸進尋找支援完成臺灣綜合大學中英對照法規資料庫的建置，建議未來盼可增加資訊技術的相關支援及互相技術支援體系；及多方面展開與其他各校間的合作。

2.在法規方面持續加強以現有法規在新舊版本上原始分項的比照及加強與臺綜大夥伴學校在法規上的整理比對等。

伍、攜回資料名稱及內容

會議相關資料(詳見附錄)

陸、附錄

附錄一: 大會議程

The workshop will take place as follows:

1. Date: From 28 June, 2014 - 30 June, 2014
2. Place: ANA HOTEL SAPPORO http://www.anahotel-sapporo.co.jp/

   Kita 3-jo Nishi 1-chome 2-9, Chuo-ku, Sapporo 060-0003
   TEL : +81-11-221-4411 FAX : +81-11-222-7624
3. The Plan of the workshop

   28 June, 2014 Reception

   29 June, 2014 Workshop (1:00pm - 5:00pm)

   30 June, 2014 Workshop (9:00am - 12:00am)

附錄二: 黃仁竑教授報告投影片(1/2): 四國法律用語資料庫建置成果

Introduction of CJKT SBD database

Ren-Hung Hwang

Outline

- Challenge in lawyer work
- Work together with lawyer and IT
- Functionality

Challenge in lawyer work

- A large amount of data
- The same word has different meanings in different countries
- Each word has a complex information

Work together with lawyer and IT

- Lawyer provides data and information
- IT provides technical support for data management
- We work together to develop a system for lawyer to manage lawful keyword efficiently
- **CJKT SBD database**

Functionality

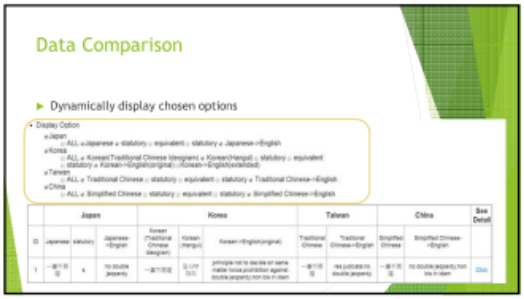
- Table Pagination
- Easy to compare data from different countries
- Search data
  - By English
  - By ID (with multi-search)
- Display Option
  - Show detail information OR
  - Display information you want to compare
- Edit data

Interface of CJKT SBD Database

Table Pagination

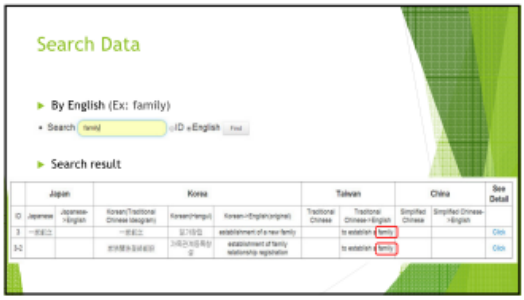
Data Comparison

- Dynamically display chosen options



Search Data

- By English (Ex: family)
- Search



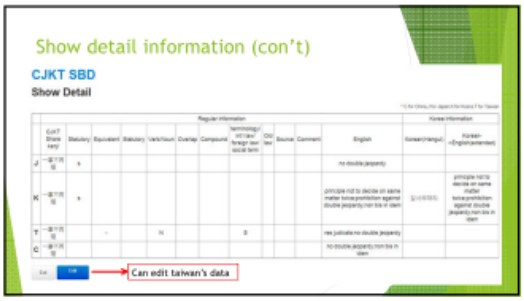
Search Data (con't)

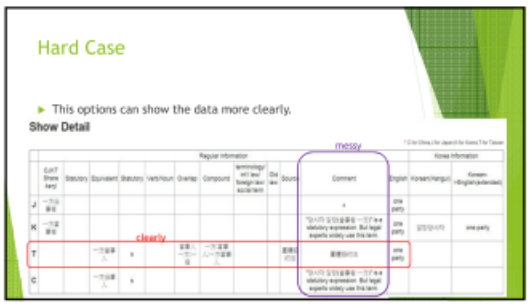
- By ID
  - Multi-search : separated by a semicolon (Ex: 1;2)
- Search

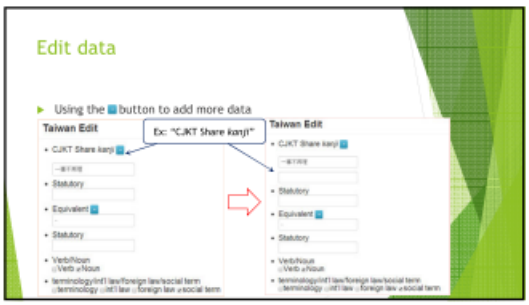
Show detail information

- Click : Open the new page to show detail information

Show detail information (con't)




Hard Case
This options can show the data more clearly.
Show Detail
messy
clearly


Edit data
Using the button to add more data
Taiwan Edit
Ex: "CJKT Share kanji"
Taiwan Edit
CJKT Share kanji
Statutory
Equivalent
Statutory
Verb/Noun
Statutory
Equivalent
Statutory
Verb/Noun

附錄三：黃仁竑教授報告投影片(2/2)：自動中英對照研究方法與成果

Automatic Word alignment of Legal Terms from Taiwan Laws

Prof. Ren-Hung Hwang
Dept. of Computer Science & Info. Eng.
National Chung Cheng University

Outline

- Introduction
- Related work
- Proposed Approach
  - Word alignment: Combined statistical methods
- Performance Evaluation
- Conclusion

Introduction

- Tasks to work on since 2012
  - Better Chinese keyword automatic extraction from Taiwan Laws
  - Better bilingual keyword alignment
  - Write PHP programs for KWIC search interface
- BTW, need JaLII's assistance to set up KWIC in our machine.

Word Alignment

- WA is the cornerstone of Machine Translation
- Approaches:
  - IBM model
  - Statistical approach: Get word pair from bilingual corpus by statistic method
    - Mutual information(MI)
    - Correlation coefficient(CC)
    - Likelihood ratios(LR)
    - Dice coefficient(DC)
    - Fractional count(FC)

Statistical approach: MI

- Mutual information
  - $MI(c,e) = log_2 \frac{p(c,e)}{p(c)p(e)}$
  - $p(c,e)$ = The probability of Chinese and English appear at the same time.
  - $p(c,e) = \frac{Appear\ sentence\ pair\ count}{Total\ sentence\ pair\ count}$
  - Simple but not accurate enough

Statistical approach: CC

- Correlation coefficient
  - $X^2(c,e) = \frac{(f_{11} \times f_{22} - f_{12} \times f_{21})^2}{F_c F_c^* F_e F_e^*}$
  - $f_{11} = The\ sentence\ include\ word\ pair$
  - $f_{22} = The\ sentence\ not\ include\ word\ pair$
  - $f_{12} = The\ sentence\ include\ english\ word\ only$
  - $f_{21} = The\ sentence\ include\ chinese\ word\ only$
  - $F_c = f_{11} + f_{21}; F_e = f_{11} + f_{12}$
  - $F_c^* = f_{12} + f_{22}; F_e^* = f_{21} + f_{22}$
  - Not able to identify low frequency pairs

Statistical approach: LR

- Likelihood ratios
  - $LR(c,e) = \log L(f_{11}, F_c, p(e|c)) + \log L(f_{12}, F_{c^*}, p(e|c^*)) - \log L(f_{11}, F_c, p(e)) - \log L(f_{12}, F_{c^*}, p(e))$
  - $L(k,n,x) = x^k(1-x)^{n-k}$
  - $p(e) = \frac{F_e}{N}, p(e|c) = \frac{f_{11}}{F_c}, p(e|c^*) = \frac{f_{12}}{F_{c^*}}$
  - Accurate, but high computation complexity

Statistical approach: DC & FC

- Dice Coefficient
  - $DC(c,e) = \frac{f_{11}}{(F_c+F_e)} = \frac{2f_{11}}{2f_{11}+f_{12}+f_{21}}$
- Fractional count
  - $FC(c,e) = \sum_{for\ all\ i,s.t(c,e)\in sp(i)} p_{sp(i)}(c,e)$
  - For each word pair, calculate intra-sentence alignment probability
  - sp(i) : *ith* sentence pair
  - $p_{sp(i)}(c,e)$: word pair(c,e) alignment probability of *ith* sentence pair

Expectation-Maximization (EM)

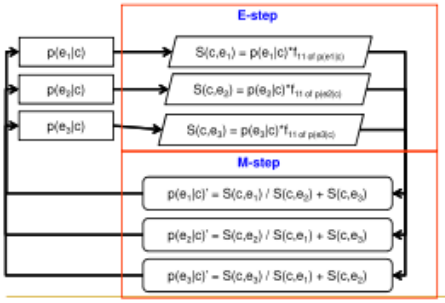
- Highlight the translation appears more times
- E-step: Cumulative conditional probability for all pair(c ,e)

$$s(c,e) = \sum_{for\ all\ i.s.t\in sp(i)} p(e|c) = p(e|c) \times f_{11}$$

- M-step: Generate new conditional probability

$$p(e|c) = \frac{s(c,e)}{\sum_v s(c,v)}$$

Expectation-Maximization (EM)



Alignment workflow

- Pre-process for bilingual corpus
  - Word segmentation
- Align word by statistical approach
  - Apply 5 statistical methods
- Decode and filter candidate
  - Merge results
  - Select top 5 candidates based on Expectation-Maximization (EM)

Pre-process of bilingual corpus

- Target: Get Chinese word and English word
- Chinese word: Word segmentation
  - Currently: use ICTCLAS toolkit
  - Next step: based on keywords derived from SVM
- English word: N-gram
  - Length: 4-grams, 8-grams
  - Remove Syntax Error phrase

Word Alignment Mechanism

- Alignment Methods
  - Mutual information(MI), Correlation coefficient(CC), Likelihood ratios(LR), Dice coefficient(DC), Fractional count(FC)
- A combined approach
  - Apply 5 statistical methods
  - Select top 5 candidates from each method
  - Merge results
  - Select top 5 candidates based on Expectation-Maximization (EM)

Experiment Set Up

- Data set: From Taiwan's law bilingual corpus
  - The number of laws: 186
- Experiment parameter
  - 4-gram
  - Top 5 score of statistic method

Initial Experiment (KWIC)

- Data set: From Taiwan's law bilingual corpus
- Select keywords: 1698 words (out of 2842)
- Select top 5 aligned English words for each keyword → Total have 8081 word pair
- Precision rate: 10.1225%
  - 493 keywords are aligned with the same English selected by experts
  - Precision: a keyword is correctly aligned if any one of the top 5 aligned English words is the same as selected by experts

Initial Experiment (Our Approach)

- Known word pair: 2842
  - 2037 keyword, each keyword may have more than 1 aligned pair.
- Alignment Probability threshold set to 0.05.
- Case 1: Each keyword has 5 aligned pairs
  - 544 / 2518 = 21.6044%
  - 313 words and 231 phrases
- Case 2: Each keyword has 3 aligned pairs
  - 503 / 1702 = 23.4718%
  - 308 words and 195 phrases

Comparison

- Among 1698 pairs (selected by experts)
  - 428 pairs are not found by both methods
  - 233 pairs found by KWIC, but not found by our method
  - 312 pairs found by our method, but not found by KWIC

Experiment: Random Selection (1/2)

- Randomly select 300 Chinese keywords
- Precision rate:
  - A keyword is correctly aligned if any one of the top 5 aligned candidate English is correct
  - KWIC: 67.6667% (203/300)
  - Our approach: 88.2155% (262/300)
- Remarks
  - 18 of them are not found by both methods
  - 6 found by KWIC, but not by our method
  - 47 found by our method, but not by KWIC

Experiment: Random Selection (2/2)

- Randomly select 300 Chinese keywords
- Precision rate:
  - An aligned pair is correct if the aligned candidate English is correct
  - Each keyword is aligned up to 5 English words (5 aligned pairs)
  - KWIC: 26.9416% (392/1455)
  - Our approach: 35.9717% (509/1415)
    - Can be improved to 60% if post processing is applied
    - Quite a few errors are due to missing of article, preposition.

Examples: KWIC vs. Our Approach

- 教唆犯
  - 「中華民國刑法」
    - 第29條 教唆他人使之實行犯罪行為者，為教唆犯。
    - Article 29 A person who solicits another to have committed an offense is a solicitor.
    - 教唆犯之處罰，依其所教唆之罪處罰之。
    - A solicitor shall be punished according to the punishment prescribed for the solicited offense.
  - 「人工生殖法」
    - 前項教唆犯及幫助犯罰之。
    - Whoever solicits or abets another person to commit the crimes as described in the proceeding paragraph, shall be guilty of the crime and punished.

Examples: KWIC vs. Our Approach

- Our results
  - a solicitor
  - solicitor
  - a person who solicits
  - according to the punishment
  - committed an offense
- KWIC
  - whoever solicits or abets another person to commit
  - person who solicits another to have committed
  - whoever solicits or abets another person to
  - solicits or abets another person to commit
  - whoever solicits or abets another person

Examples: KWIC vs. Our Approach

- 不正之方法
  - 訊問被告應出以懇切之態度，不得用強暴、脅迫、利誘、詐欺、疲勞訊問或其他**不正之方法**。
  - An accused shall be examined in an honest manner; violence, threat, inducement, fraud, exhausting examination or other **improper means** shall not be used.
  - 二　以恫嚇、侮辱、利誘、詐欺或其他不正之方法者。
  - (2) The examination is conducted by ways of threat, insult, inducement, fraud, or other improper means;

Examples: KWIC vs. Our Approach

- 不正之方法
  - 被告陳述其自白係出於不正之方法者，應先於其他事證而為調查。該自白如係經檢察官提出者，法院應命檢察官就自白之出於自由意志，指出證明之方法。
  - If the accused states that his confession was extracted by improper means, his confession shall be investigated prior to investigating other evidences; if the said confession is presented by the public prosecutor, the court shall order the public prosecutor to indicate the method to prove that the confession is obtained under the free will of the accused.

Examples: KWIC vs. Our Approach

- 不正之方法
  - 被告之自白，非出於強暴、脅迫、利誘、詐欺、疲勞訊問、違法羈押或其他不正之方法，且與事實相符者，得為證據。
  - Confession of an accused not extracted by violence, threat, inducement, fraud, exhausting interrogation, unlawful detention or other improper means and consistent with facts may be admitted as evidence.

Examples: KWIC vs. Our Approach

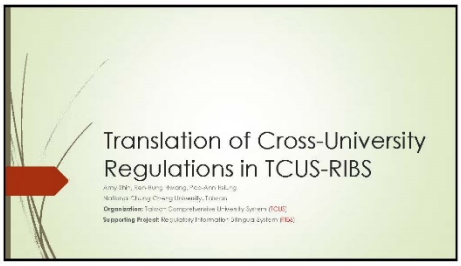
- Our results
  - improper
  - improper means
  - fraud
  - inducement
  - threat
- KWIC
  - violence threat
  - extracted by
  - exhausting
  - extracted
  - or other improper means

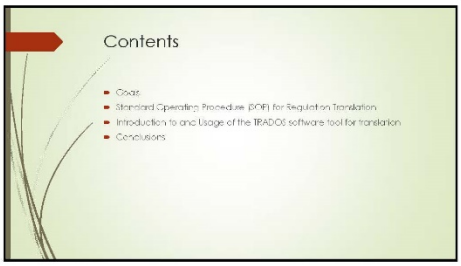
Conclusion

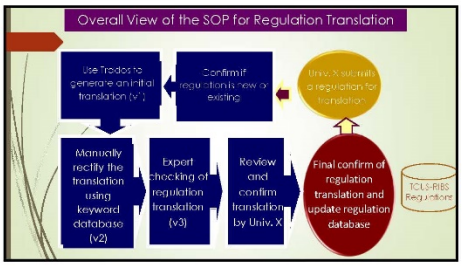
- Reported results of word alignment based on statistical methods
- Will report results of IBM model next time
- Will use SVM instead of word segment tool next time
- Phrase alignment is still a big challenge

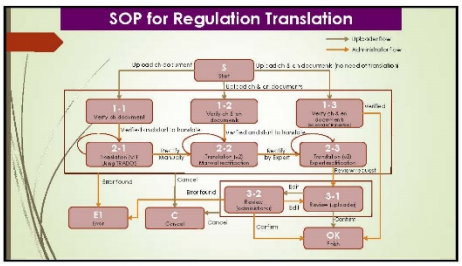
附錄四：熊博安教授報告投影片

2014/7/4

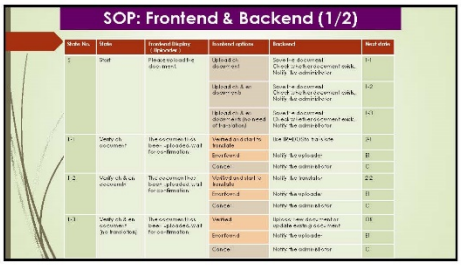


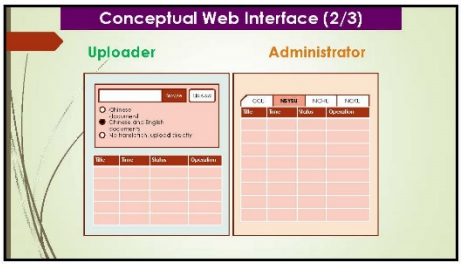


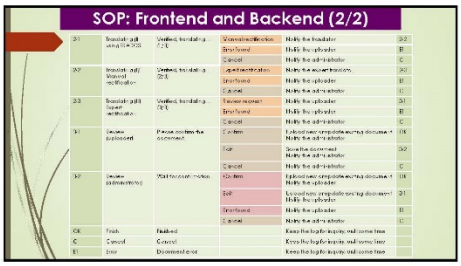


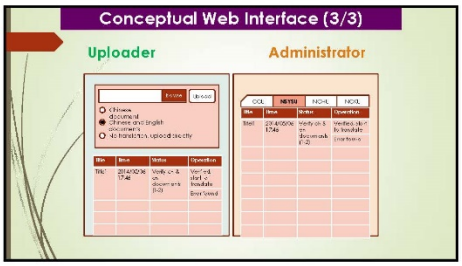


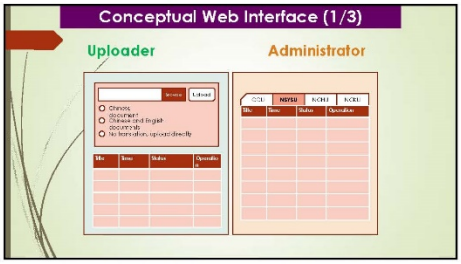
1

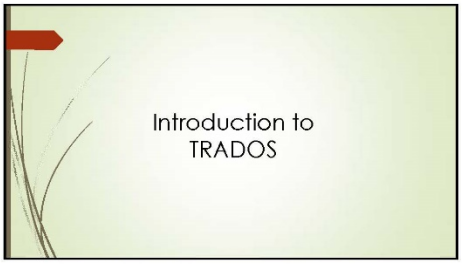

SOP: Frontend & Backend (1/2)


Conceptual Web Interface (2/3)
Uploader
Administrator


SOP: Frontend and Backend (2/2)


Conceptual Web Interface (3/3)
Uploader
Administrator


Conceptual Web Interface (1/3)
Uploader
Administrator


Introduction to
TRADOS

Introduction to TRADOS

- Created initially by Trados GmbH (Germany) in 1990
- Available through SDL International (Translationzone.com) from 2005
- Latest version: **Trados Studio 2014**
  - TCUS-IITBS project TCUS has 2011/2014 versions of Trados Studio
  - A Trados expert is assigned to TCUS-IITBS Project
- A computer-assisted translation software suite
  - SDL Trados Studio
  - SDL MultiTerm

Training Trados Translation Software

Develop Trados English Translation Memory System

1. Trados uses translation memory technology to effectively manage translation data.
2. During translation, the system will automatically search for the same or similar translation sources to generate translated text as reference, thus avoiding unnecessary repeated labor.

TCUS

SDL Trados Studio (with MultiTerm)

- Translating files
- Creating and managing **translation memories**
- Terminology management using **TermBase**

Training Trados Translation Software

Develop Trados English Translation Memory System

3. At the same time, the translation memory database keeps learning and auto-saving the new translated texts, thus becoming more and more "smart"!

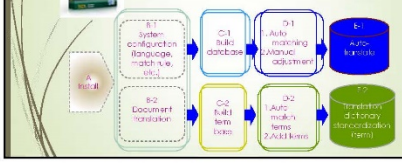
TCUS

Trados Translation Software

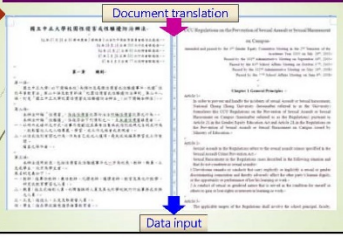
Trados can be used for:

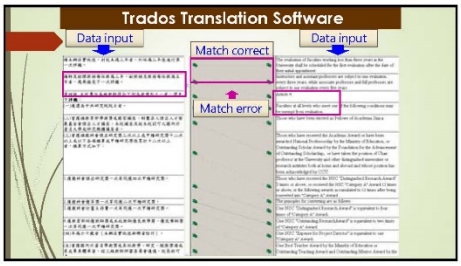
(a) Translating files using **translation memory**

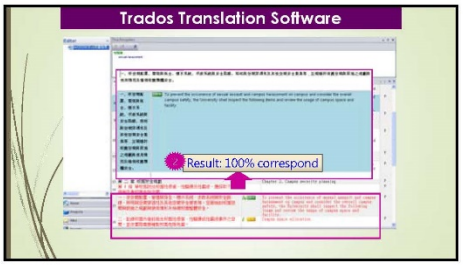
(b) Keeping terms consistent using **termbase**



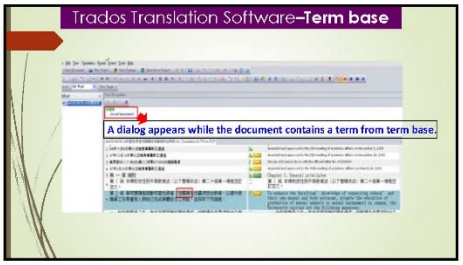
Trados Translation Software

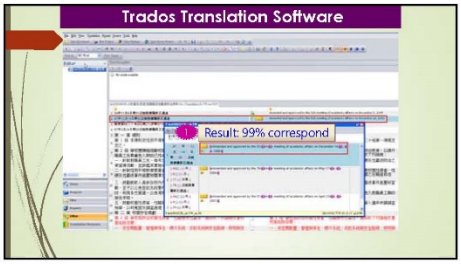


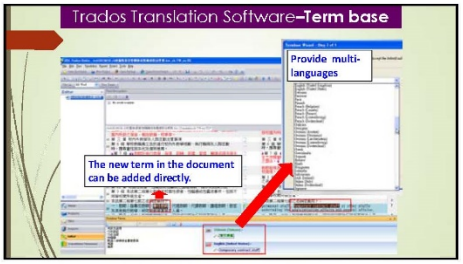

Trados Translation Software
Data input
Match correct
Data input
Match error

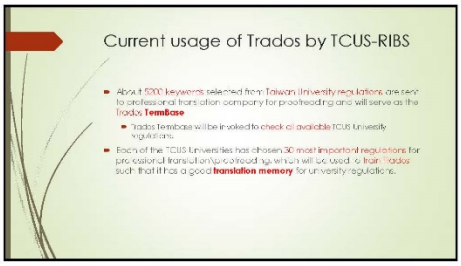

Trados Translation Software
Result: 100% correspond

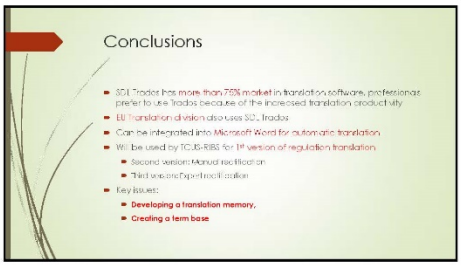

Trados Translation Software
New translation
Auto match


Trados Translation Software–Term base
A dialog appears while the document contains a term from term base.


Trados Translation Software
Result: 99% correspond


Trados Translation Software–Term base
Provide multi-languages
The new term in the document can be added directly.


Current usage of Trados by TCUS-RIBS
About 5200 keywords selected from Taiwan University regulations are sent to professional translation company for proofreading and will serve as the Trados TermBase
Trados Termbase will be invoked to check all available TCUS University regulations
Each of the TCUS universities has chosen 30 most important regulations for professional translation/proofreading, which will be used to train Trados such that it has a good translation memory for university regulations.


Conclusions
SDL Trados has more than 75% market in translation software, professionals prefer to use Trados because of the increased translation productivity
EU Translation division also uses SDL Trados
Can be integrated into Microsoft Word for automatic translation
Will be used by TCUS-RIBS for 1st version of regulation translation
Second version: Manual rectification
Third version: Expert rectification
Key issues:
Developing a translation memory,
Creating a term base


Thank you! Any Questions?