編譯論叢
第四卷 第一期（2011 年 3 月），133-160

翻譯記憶系統的發展歷程與未來趨勢

王正

本文對翻譯記憶系統的發展歷程進行了簡單回顧，指出其原型思想肇始於 20 世紀 60 年代，商用系統出現於 20 世紀 90 年代初，近 20 年來取得了迅猛發展。本文認為，當代翻譯記憶系統可以分為句級翻譯記憶和低於句級翻譯記憶的兩代模式，並批評了當代句級匹配演算法的弊病，進而指出，新興的低於句級翻譯記憶極大地提高了翻譯記憶庫的利用率，成為不可逆轉的趨勢。本文分析了當前翻譯記憶系統發展的五大趨勢，並指出低於句級翻譯記憶模式尚未成熟，在語種、系統結構等方面尚有有一些內在的缺點和不足。此外，新型翻譯記憶系統對譯者的翻譯活動和素質要求將產生何種影響，亦值得學界予以進一步關注。

關鍵詞：電腦輔助翻譯、翻譯記憶、句級翻譯記憶、低於句級翻譯記憶、機器翻譯

收件：2010 年 8 月 30 日；修改：2010 年 12 月 23 日；接受：2011 年 1 月 5 日

王正，上海外國語大學新聞傳播學院副教授，E-mail: wangzheng021@126.com

Compilation and Translation Review
Vol. 4, No. 1 (March 2011), 131-158

Translation Memory Systems: A Historical Sketch and Future Trends

Zheng Wang

This paper begins with a historical sketch of translation memory (TM) systems—though the prototype of the TM module can be traced back to the 1960s, TMs were not commercialized until almost 30 years later. Since then, they have enjoyed rapid development. This paper proposes that TM systems can be divided into two generations, namely sentential TMs and sub-sentential TMs. Based on a critique of the former, we see that a sub-sentential segment TM system, as an effective booster of TM leverage efficiency, has become an irrevocable trend. Despite the bright prospect of TMs as shown in five basic trends, the new-generation sub-sentential TM system is still premature and thus susceptible to defects and deficiencies such as the limitation of language pairs and modular structures. This study concludes that, greater attention is called for with regard to what possible impacts the new TM system will have on the translating behaviors and qualifications of translators today.

Keywords: computer-aided translation (CAT), translation memory (TM), sentence TM, sub-sentential TM, machine translation

Received: August 30, 2010; Revised: December 23, 2010; Accepted: January 5, 2011

Zheng Wang, Associate Professor, College of Journalism and Communication, Shanghai International Studies University, E-mail: wangzheng021@126.com

壹、翻譯記憶簡介

一、翻譯記憶的概念和背景

翻譯記憶（translation memory）是電腦輔助翻譯（computer-aided translation）的重要領域。翻譯記憶是指「由（劃分區段、對齊、句法處理並分類後的）多語文本組成的多語文本檔案庫，允許使用者對這些多語文本區段進行存儲並在多種搜索條件下提取（EAGLES, 1996, p.140）。」這些已翻譯好的區段仍然是人工翻譯的結果。它的特點是能夠將翻譯流程中涉及純粹記憶的活動，比如術語的匹配和自動搜索提示、高度相似句子的記憶和複現交給電腦來做，免除翻譯人員反復查找術語之苦，使其能全力對付語義的轉換和傳遞（徐彬、郭紅梅、國曉立，2007，頁 82）。總之，翻譯記憶的目的是說明譯者重複利用已經翻譯過的語料。正如 Bowker（2002, p. 93）指出，儘管語言是動態的，但重複性很強。人們在交流類似話題時，經常使用相同或相近的表達方式。隨著翻譯工作量增加，譯者有時需要重複翻譯部分相同或相似的句子或段落。在沒有電腦的時代，譯者在工作中摸索出了一套重複利用譯文的策略。他們有人把在許多文檔中重複出現的標準段落譯文抄在卡片上，有人則直接用剪刀加漿糊的方式進行粘貼。顯然，這些手工做法速度慢，又難以保證譯者之間的資源分享。而在電腦時代，利用 TRADOS 等電腦輔助翻譯工具內置的翻譯記憶系統，電腦可以自動將已經翻譯過的區段顯示在螢幕上供譯者參考，從而大大提高了翻譯效率。

翻譯記憶的雛形最早出現在上世紀 60 年代，美國自動語言處理諮詢委員會（ALPAC）在著名的研究報告《語言與機器：電腦在翻譯和語言學中的應用》（Languages and machines: computers in translation and linguistics）中提到，盧森堡的歐洲煤炭與鋼鐵聯合體術語部與布魯塞爾自

由大學聯合開發了自動詞典檢索系統，附有詞條的上下文，針對查詢的詞條，電腦輸出並列印對應的人工譯文例句（ALPAC, 1966, p. 27）。顯然，這一做法有利於譯文的重複利用，能夠有效提高翻譯效率。有鑑於此，ALPAC 雖然全面否定機器翻譯的可行性，卻在報告末尾提出了部分建議，最後一條就是：為譯者提供足夠的參考資料，包括對目前主要用於機器翻譯中自動詞典查詢的詞彙表進行改編（production of adequate reference works for the translator, including the adaptation of glossaries that now exist primarily for automatic dictionary look-up in machine translation）（ALPAC, 1966, p. 34）。

翻譯記憶的理念形成於 20 世紀 70 年代。Melby（1995, p. 187）介紹說，早在 20 世紀 70 年代他在楊百翰大學機器翻譯專案組研究互動式翻譯系統（interactive translation system）時，便已經提出了翻譯記憶的理念。1981 年 Melby 及楊百翰大學其它研究人員聯合開發 的 ALPS 機器翻譯系統成為首批商業化的機器翻譯系統之一。該系統集成了類似翻譯記憶的模組「重複處理（Repetitions Processing）」，不過其功能只限于尋找完全對應的區段。Hutchins（1998, p. 295）認為，翻譯記憶這一理念是由 Peter Arthern（1979）率先提出的。他建議將所有源文本和譯文文本存儲在電腦裡，以便對文本的任何部分快速提取，並根據需要即時插入新文本中。他將這一理念稱作「文本提取式翻譯」。在另一篇文章裡，Arthern 對這一理念做了更為清晰的表述：

> 我們一定可以製作一個程式，讓文字編輯器來“記住”剛剛鍵入的新文本中是否有任何部分已被翻譯，並將這部分連同其譯文一併嵌入文本中……
>
> 任何新文本在鍵入文本編輯工作站時，系統會將該文本與記憶體中已有文本連同其（歐洲委員會）各語種的官方譯文進行比較……與機器翻譯相比，其一大優點是，提取的所有段落均無語法錯誤。實際上，我們應

> 當執行一個電子的“剪切——粘貼”程式。據我估計，這一做法可以讓譯者在實際翻譯工作中節約至少 15% 的時間（Arthern, 1981, p. 318；引自 Somers & Diaz, 2004, p. 6）。

1980 年，Martin Kay 發表了一篇施樂公司的內部備忘錄 The Proper Place of Men and Machines in Language Translation[1]。該備忘錄對電腦輔助翻譯研究影響深遠。文中正式提出了類似翻譯記憶系統雛形的功能及其對譯者的重要性，因而 Kay 被普遍尊為翻譯記憶系統的先驅。Kay（1997, p. 19）指出，「若下一待譯文本片段不夠曉白，譯者可通過指令讓電腦顯示所有與該片段相關之文本，從而讓譯者關注到之前所做的翻譯決策，從統計資料顯示較為突出的詞彙和片語，以及他此前關注過的所有資訊之記錄。繼續翻譯前，他可以仔細閱讀過去和未來文本中包含類似材料的片段。」

總的來說，到了上世紀 80 年代初，翻譯記憶的基本原則已經確立。Hutchins（1998, p. 3）指出，「來自不同背景的不同人士在不同時間、電腦的不同發展階段獨立提出了相似的觀點。」

首批商業化的翻譯記憶工具成型於 1988 年。這一年，IBM 日本分公司的 Sumita 和 Tsutsumi 發佈了 ETOC（Easy TO Consult）工具。該工具實質上是一個改良過的電子詞典。傳統的電子詞典是按照單詞進行查詢，無法自動查詢兩個單詞以上的短語或句子，而 ETOC 則提供了靈活的解決方案：將待查詢的句子輸入後，系統在詞典中進行提取。如果找不到，則對句子進行語法分析，剝離部分實詞，僅保留負責句型結構的虛詞和形容詞等，然後再將這些與詞典庫中的雙語句子進行比較，找到具有相同結構的句子後即顯示出來供譯者選擇。譯者將對應的譯文句子複製並粘貼到編輯器中，即可進行修改並完成譯文。例如輸入日語句子“君は水泳が大変うまい”，電腦詞典中顯示無法找到對應譯文，於是 ETOC 程式對句子進行結構分析，去除部分實詞「君」、「水泳」、「大變」，即：

～は～が～つまい

然後將這句話與系統中的雙語譯文進行比照，得到三個結果：

彼【は】トランベヅ蔔を吹くの【が】【うまい】

He plays the trumpet well.

君【は】全く芝居【が】【うまい】

You're a great actor.

彼女【は】ケーキを焼くの【が】とても【うまい】

She is very good at baking cakes.

將其中的第二句複製並粘貼在編輯器中，用 swimmer 替代 actor，最終譯文為：

You’re a great swimmer.（Sumita & Tsutsumi, 1988, p. 2）

這一系統並未使用 translation memory 這一術語，而且將譯文資料庫依然稱作「詞典（dictionary）」，但顯然它已經基本具備了當代翻譯記憶的基本特徵。當然，其缺陷是需要對語法進行分析，從而加大了程式開發難度，也限制了程式的可擴展性。假如需要增加一個新的語種，就需要對該語種編寫專門的語法分析模組才能使用。而且該程式只採用了完全匹配（perfect match）演算法，不支援模糊匹配，因而大大縮小了譯文的可適用性。

同年，當時依然默默無聞的 TRADOS 公司發佈了 TED（Translation Editor）。作為該公司銷售的 Ink 公司電子詞典產品 Texttools 的外掛程式，為譯者提供了分屏編輯視窗。1992 年，TRADOS 發佈了 DOS 版電腦輔助翻譯軟體 Translator’s Workbench II，其中的 TW Editor 模組即源於 TED，但增加了一大重要模組，即翻譯記憶。這款翻譯記憶工具已經具備了分割區段（segmentation）、自動檢索、模糊匹配（fuzzy match）等基本功能。系統對當前待譯區段進行自動匹配，如果在記憶庫中發現 100% 匹配的譯文，則用譯文區段替代源文本區段，如果發現相似區段，則將該區

段的原文和譯文共同顯示供譯者參考，並將該區段與待譯區段的差異用不同顏色標明，提醒譯者注意。由於該系統依然基於對句子的語法分析（parsing），因此僅支持德語、英語、法語、義大利語和西班牙語[2]。

1991 年德國 Star 公司發佈了 Star Transit，1992 年 IBM 德國分公司也發佈了內置翻譯記憶模組的電腦輔助翻譯軟體 Translation Manager/2，1993 年西班牙 Atril 公司發佈了支援首個 Windows 版本的電腦輔助翻譯系統 Déjà Vu。自此，以翻譯記憶為核心的電腦輔助翻譯軟體已經初步形成。隨著微軟公司 Windows 95 的發佈，各公司紛紛發佈基於該系統的軟體版本，功能更加豐富，翻譯記憶的發展走上了快車道。

二、翻譯記憶的工作原理

翻譯記憶系統本質上是基於自然語言中語料的複現，利用電腦強大的記憶功能，對待譯文本中已經翻譯過的部分予以自動提示或替換，從而減輕了譯者的工作負擔，提高效率。翻譯記憶系統的運行主要基於以下工作原理：

（一）劃分區段（segmentation）

在大多數翻譯記憶系統中，翻譯的基本單位是句子。一旦導入待譯文本後，系統即自動根據句號、分號、問號、回車符等分隔符號號，將文本劃分為獨立的區段，這些區段除了完整的句子外，尚包括標題、清單和表格單元內的文字等常見的區段形式。由於部分標點符號可能在使用中存在歧義現象，導致區段劃分錯誤，例如 Mr. 和 e.g. 和英語的句號在形式上無法區分，因此常規做法是採用非索引字表（stoplist）來將這些例外情況排除開（Bowker, 2002, p. 95）。此外，數字後面的小數點也常常給劃分區段造成困難，再加上中文的標點符號與英文差異較大，因此不同軟體的劃分區段方式不盡相同。以筆者的使用經歷來看，Déjà Vu 在小數點和句號的區分方面做的不夠好，有時會誤判，將中文文本中帶小數點的數位分割在兩個句子當中。

（二）自動匹配

這一技術可以分為完全匹配和模糊匹配。前者是指待譯區段與記憶庫中現有區段 100% 無論在語言還是格式上均完全相同，包括拼寫、標點符號、數位甚至格式（如斜體、加粗等）[3]。而模糊匹配則是指待譯區段與記憶庫中現有區段較為相似，但並非 100% 匹配。翻譯記憶系統會把這些模糊匹配的結果顯示在螢幕上供譯者參考，並對兩者之間的差異突出顯示，提醒譯者注意。譯者一般可以在翻譯記憶系統裡設置匹配率的閾限（threshold），常規的比例是 60—70%。低於 60%，則其相似性基本不具備參考意義，不作為模糊匹配結果出現在螢幕上。如果待譯區段在翻譯記憶中找不到相似內容，則視為零匹配。此時系統要求譯者將該句的譯文填入翻譯區。

（三）預翻譯

早在 1992 年，IBM 德國分公司發佈的 Translation Manager/2 即具有了預翻譯功能。執行該命令後，系統將待譯文本中完全匹配的區段自動替換為相對應的譯文區段，從而大大節約了譯者的時間。而對模糊匹配的區段則不予替換，在人工翻譯過程中，則在該區段旁的視窗顯示模糊匹配的譯文供譯者參考（Alesiani, 1994）。在當今的翻譯記憶系統中，使用者可以對預翻譯功能自行設定，不僅對完全匹配的區段進行替換，而且對待譯文本中的術語和詞彙替換為術語庫中的對應譯文，亦可對模糊匹配的區段進行替換，並用醒目的方式提示使用者差異所在以及相應的匹配率。

（四）語料對齊（alignment）

對齊是指將源文本的區段與目標文本中的相應區段關聯起來，旨在創建新的翻譯記憶資料庫或為現有記憶庫增加資料。對齊的雙語區段稱作「翻譯單位」（Quah, 2006, p. 100）。語料對齊是建立翻譯記憶的最基本步驟，也可以用來建立雙語平行語料庫。黃俊紅、範雲、黃萍（2007，頁 21）指出，雙語平行語料庫的對齊方式研究主要基於三種方法 : 一是基於統計的方法，即通過雙語對譯句 / 詞的同現概率，建立句子 / 詞彙對齊的

統計模型，用來判斷句子 / 詞的對譯關係；二是基於詞彙 / 詞典的方法；三是把統計手段和詞彙 / 詞典結合起來。TRADOS Translator's Workbench 的元件之一 WinAlign 提供了多種額外指標加強對齊演算法的準確性，如標記顯著性（Tags Significance）、數位顯著性（Numbers Significance）、字串顯著性（Expectations Significance）和格式顯著性（Formatting Significance）。例如，如果某一文本中包含大量數位元，這些數位在目標譯文中往往完整保留下來，在對齊時將「數字顯著性」滑竿移到“high”，就可以說明系統根據數位的對應來對齊譯文，增加了對齊準確率（王正、孫東雲，2009a，頁 82）。當然，對齊的過程中本身也包含了劃分區段的操作，將對齊的結果按照區段（主要是句子）的形式保存在記憶庫中。

貳、當代翻譯記憶系統的發展

一、翻譯記憶系統的發展階段之爭

業界對翻譯記憶系統的分代說法尚未達成共識，Lagoudaki（2006）認為，基於句級區段的、使用字串匹配相似度進行查詢的翻譯記憶系統均為第一代系統，而語言學增強的（linguistically enhanced）翻譯記憶例如 Similis 是第二代系統。Gotti et al.（2005）、Kavak（2009）和 Elita 與 Gavrila（2006）則將翻譯記憶分為三代：第一代翻譯記憶僅支持完全匹配（full match），翻譯過的句子不需要翻譯第二遍，其缺點是完全相同的句子數量很少，對翻譯記憶的重複利用價值很低；第二代翻譯記憶支持模糊匹配（fuzzy match），如果待譯區段與記憶庫中存儲的區段存在些許差異，則提供匹配供譯者參考，這一做法大大提高了翻譯記憶的利用價值；第三代翻譯記憶則在低於句級（sub-sentential）的層面提供匹配。

Lagoudaki 的「二代說」忽略了從基於句級的區段匹配到低於句級的區段匹配這一重要發展趨勢，其實語言學增強型翻譯記憶本身也是基於低

於句級的區段匹配模式，因此「三代說」更加清晰地表現了翻譯記憶系統的發展軌跡。但僅支持完全匹配的「第一代翻譯記憶」很難作為一個獨立的發展階段來看待，因為 Somers 與 Diaz（2004, p. 11）提到，早在 1991 年，*Language International* 雜誌就在一篇報導仲介紹了「模糊匹配」的概念，而且 1992 年推出的 TRADOS Translator's Workbench II 和 IBM Translation Manager 均具有模糊匹配的功能。而上文所述的 ALPS 翻譯工作站的「重複處理」模組雖然是以句子作為完全匹配的單位，不支援模糊匹配，但由於整體建構是基於機器翻譯，所以很難被看做真正的翻譯記憶系統。因此，本文將翻譯記憶系統分為兩代，即句級翻譯記憶系統和低於句級翻譯記憶系統。其中前者經過近 20 年的發展，依然風頭正勁，而後者則在實驗室條件下取得了良好的效果，代表著未來的發展趨勢。

二、句級翻譯記憶系統缺陷分析

句級翻譯記憶系統是當今的主流翻譯記憶系統，TRADOS、SDLX、Déjà Vu、Wordfast 等世界領先的電腦輔助翻譯系統均基於這一原理。這些軟體功能強大，除了提供翻譯區段的匹配外，還可以保留原文檔格式、支援翻譯記憶的跨系統交互、網路協作等諸多功能，已經成為專業譯者不可或缺的利器。然而這一系統也存在著諸多缺陷，因而備受詬病。要充分瞭解其價值和局限性，就必須對劃分區段和匹配這兩個方面進行探討（Macken, 2009, p. 195）。

（一）匹配中相似性演算法的限制

匹配演算法是翻譯記憶系統的核心技術之一。句級翻譯記憶系統在匹配演算法方面存在著諸多缺陷，因而其適用性受到了很大限制。目前大多數常規翻譯記憶系統的匹配檢索技術為基於字串的匹配檢索[4]。借用自然語言處理領域中的「編輯距離（Levenshtein distance）」比較字串之間的相似程度。該演算法由原蘇聯科學家 Vladimir Levenshtein 提出，通過計算將一個字串轉換成另一個字串所需要進行插入、刪除和替換操作的

次數，與整個字串的長度相比，得出的百分數。例如，要將單詞 magazine 轉換成 magazines，需要鍵入一個字母“s”，差別為 1，與整個字串（通常為源字串）之比為 12.5%。使用這種方法，翻譯記憶系統認為兩者之間的相似度為 87.5%。

這一演算法簡便易行，廣泛用於自然語言處理，但就翻譯記憶而言，其缺陷也非常明顯。一方面，字串匹配演算法對對字元的順序變化視為差異，未能考慮到並列關係的順序差異。以下面兩句為例，人工翻譯可以看出以下兩句的意思非常相似：

句 1："Ice cream: chocolate and vanilla"

句 2："Ice cream: vanilla and chocolate"

然而翻譯記憶系統對將第一句轉換為第二句所需的擊鍵次數進行統計，從而認為兩句差別甚大，不會從翻譯記憶庫中提取相應區段進行匹配。可見。這一演算法僅考慮到語言的形式而不涉及語言的意義，因而檢索精確度不高（Raya, 2007）。

另一方面，由於字串匹配演算法不涉及句法處理，因此對語言中的屈折變化和結構轉換無能為力。以下面的三個句子為例：

句 3：Oracle® is a registered trademark of Oracle Corporation.

句 4：Java® is a registered trademark of Sun Microsystems Inc.

句 5：Unix®, X/Open®, OSF/1®, and Motif® are registered trademarks of the Open Group.（Macken, 2008, p. 196）

顯然，人工翻譯會一眼看出，這三句話在結構上極為相似，但翻譯記憶系統給出句 3 和句 4 的相似度為 61%，而句 3 和句 5 的相似度則低於 30%。在翻譯句 4 和句 5 時，翻譯系統很可能不會將句 3 作為匹配項自動顯示出來，這顯然是不合理的，也大大降低了翻譯記憶的利用率。

有鑑於此，部分學者對編輯距離的演算法進行了反思。在自然語言理

解和機器翻譯等領域廣泛使用的 n-gram 不失為一條便捷的方法。n-gram 是指 n 個字符的連續序列，當 n 取 1、2、3 時，n-gram 模型分別稱為 unigram、bigram 和 trigram 語言模型。n 越大，模型越準確，也越複雜，需要的計算量越大，因此 n ≥ 4 的情況較少。當 n=2 時，下面的句子可以分割成下列 bigram：

原句：他們為年輕付出的代價很大。

Bigram 序列：< 他們 >< 們為 >< 為年 >< 年輕 >< 輕付 >< 付出 >< 出的 >< 的代 >< 代價 >< 價很 >< 很大 >。

比較：他們付出的代價很大。

Bigram 序列：< 他們 >< 們付 >< 付出 >< 出的 >< 的代 >< 代價 >< 價很 >< 很大 >。

從 bigram 角度對上述兩句進行對比，按照 bag of words 演算法查看匹配比率，即可計算其相似度。n-gram 在比較語句相似度方面已經得到廣泛應用，是機器翻譯自動評價方法 BLEU（Papineni & Roukos, 2002）和 NIST 等的基礎。Feiliang 與 Shaoming（2006, p. 454）介紹了富士施樂公司應用的基於 n-gram 的翻譯記憶，其原理是，第一步，對待譯區段通過 hash 檢測是否有完全匹配；第二部，若無完全匹配，則以 n-gram 為單位在 TM 當中尋找模糊匹配。這一演算法具有結構簡單、速度快等優勢。Baldwin（2010）則通過實驗指出，在翻譯記憶的語句相似度計算當中，基於 n-gram 的演算法與傳統的編輯距離演算法精確度基本相同，而在速度上佔據絕對優勢。這一進步令人鼓舞，但 n-gram 能否彌補編輯距離演算法的上述缺陷，有待進一步研究。

（二）區段的劃分方式與匹配檢索。

正如上文所述，句級翻譯記憶系統中的區段大多以句子為單位，但在實際工作中，整句重複或相似的可能性僅限於特定的文檔類型。因此它更適用於具有內部重複的文檔如表格、說明書等，不適用於全新的、毫不相

干的文檔；適用於修訂的文本或更新後的文檔（updates）例如新版的產品說明書，其中有著大量重複或模糊匹配的字句，而不適用於完全不同的文檔；適用於長期服務的客戶，以適應該客戶特定的術語和風格要求，而不適用於新的客戶要求（Bowker, 2002, p. 112-114）。而在常規的翻譯過程中經常出現的區段匹配存在於短語或小句層面，例如中文「建設和諧社會」，英文 grow by leaps and bounds 等短語即不存在於術語庫，也不存在於翻譯記憶庫，卻在翻譯過程中經常出現，是目前此類翻譯記憶系統的一大盲點。如果能夠將這些短語在翻譯中重複利用，將會大大提高匹配率，減輕譯者負擔。

為了提高重複利用的匹配率，部分翻譯記憶系統採用了各種各樣的手段作為翻譯記憶的輔助。一種常見的解決方案是將某一專案特定的部分語塊（chunk）或稱片斷（fragment）放在區段裡面，加入翻譯記憶庫。如果較長的區段中包含這一片斷時，則部分翻譯記憶系統如 MemoQ 可以自動對該區段做出匹配提示。但這一做法需要譯者在翻譯過程中，手工選中雙語對應片斷並加入記憶庫。再如 Déjà Vu 使用 lexicon 來作為記憶庫和術語庫（terminology database）之間的中間品。區段作為匹配的單位有時較大，造成匹配率偏低。而術語則單位太小，而且一般較為穩定而嚴謹。Lexicon 作為當前翻譯項目專用的詞彙表，由譯者手工選擇雙語對應片斷並加入，在系統自動檢索時優先於術語庫，且在項目結束時可以生成專用的詞彙表提交給客戶。但這兩種做法都需要涉及手工操作，不僅複雜，而且數量非常有限，如果沒有長期的人工積累，則對翻譯的實際幫助不大。

三、低於句級翻譯記憶系統

（一）低於句級翻譯記憶系統的工作原理

由於句級翻譯記憶系統在劃分區段和匹配演算法方面存在著種種缺陷，因而效果是高精度、低利用率，僅提供最佳匹配的區段，而且要和

待譯區段極為相似（Simard & Langlais, 2001, p. 1）。正因為這些不足，部分研究者開發了號稱「第二代」（Planas, 2005）或「第三代」（Gotti et al., 2005）翻譯記憶系統的新系統。這些系統採用了較小的匹配單位，有的稱作「短語詞彙（Phrasal lexicon）」(Schäler, 1996)，有的稱作「語塊」（Planas, 2000），統稱為低於句級的翻譯記憶（sub-sentential translation memory）。

第二代翻譯記憶的一大特點是突破了以句子為基礎的區段匹配模式，使用了低於句級（sub-sentential）的對齊模式，大大提高了匹配效果。正如 Bowker 與 Barlow（2004, p. 4）指出，「在整句和專業術語之間還有一個語言複用的層面——短語或習語層面。其實這才是語言複用發生最多的層面。」要做到短語層面的匹配檢索，首先要做好短語層面的對齊（alignment）。對此，統計機器翻譯已經取得了長足的進展，基於短語的統計機器翻譯模型（phrase-based statistical machine translation）從大規模語料庫中抽取大量對齊短語片斷，利用這些短語片斷來匹配組合要翻譯的句子，取得了令人滿意的對齊效果。目前，低於句級的對齊模組已經成為統計機器翻譯的重要模組。Gotti et al.（2005）提出，將這一模組應用在電腦輔助翻譯工具尤其是翻譯記憶系統中，定會收到良好的效果。

進入 21 世紀以來，低於句級的翻譯記憶研究得到了研究者的廣泛重視（Macklovitch & Russell, 2000; Planas, 2000; Simard & Langlais, 2001; Simard, 2003）。Simard & Langlais（2001, p. 206）提出用語塊作為翻譯記憶的匹配單位，並通過實驗表明這一做法大大提高了匹配精度和區段的複用效率。他們的研究同時證明，基於語言分析的語塊比統計機器翻譯的 n-gram 匹配率更高。當然，使用語塊作為匹配單位元，需要考慮兩個因素：複用率和精度。前者涉及到語塊作為翻譯記憶單位在翻譯記憶匹配中的使用頻率，後者則涉及到系統給出的匹配結果是否相關，對譯者是否能夠提供實質性幫助。一個低於句級的翻譯記憶系統是否優秀，要看它是否在提高複用率的同時也具有較好的精度或可用性（Colominas, 2008, p. 345）。

為了提高匹配精度，在匹配演算法方面，部分新一代的翻譯記憶系統還採用了語言學增強匹配模式，即在系統中使用了語言分析，Colominas（2008）將語言分析與基於語塊的翻譯記憶結合，使用 Phrase Tagger 提取的名詞短語語塊（NP chunks），通過語料庫實驗證明，這一做法在翻譯記憶的匹配複用性和精度方面效果均令人滿意。號稱「第二代翻譯記憶系統」的 SIMILIS 就採取了類似的演算法。在製作翻譯記憶庫時，Similis 系統先將源文本和譯文文本劃分為對齊的句級區段，然後對區段進行語言學分析，進一步將句子分割為句法單位（或稱「語塊」），對這些語塊進行語法標注（在此過程中使用了單語詞典和語法範疇識別的演算法）（Planas, 2005），然後該系統對這些翻譯區段進行索引。因此，系統每次搜索匹配項時，不僅查找句對，而且查找語塊，因此找到匹配項的可能性大大提高了。Macken（2009）通過實驗對比第一代翻譯記憶系統 TRADOS 和第二代翻譯記憶系統 Similis，結果表明 Similis 的語塊匹配確有優勢，在術語和固定搭配（或語塊）方面比 TRADOS 提供了更多有效的幫助。

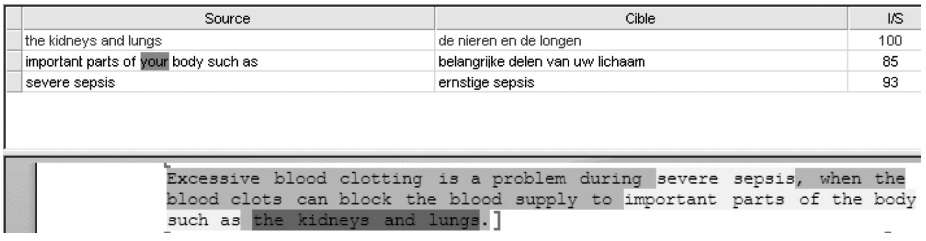

圖 1：Similis 的語塊匹配工作介面

資料來源：出自 Macken (2009: 208)

由於新一代的翻譯記憶系統在翻譯記憶的複用性和精度方面均較句級翻譯記憶系統有所提高，因此也獲得了 SDL、Atril、Kilgray 等各大翻譯記憶軟體企業的重視，如 SDL 的最新版 TRADOS 2009 已經內置了基於語塊匹配的 Autosuggest™ 模組。但不同的廠商採取的演算法不同，TRADOS 的 Autosuggest™ 基於自動鍵入提示功能，即譯者鍵入某字母，

則螢幕下方自動出現各種可能的語塊作為提示，可供譯者選擇使用，而 Similis 等則直接顯示符合匹配的整個區段，並用高亮部分提示語塊的匹配。

參、新興翻譯記憶系統的演化趨勢和不足

一、基於語言分析的低於句級翻譯記憶

正如 Melby（2006）指出，對於屈折型語言來說，在雙語語料庫基礎上基於特定語言構詞法的翻譯記憶處理，會說明系統找到以不同屈折形式存在的翻譯記憶語塊。因此他預測具備自動低於句級層面檢索的翻譯記憶系統將針對部分語言提供形態分析。目前僅有少數翻譯記憶工具提供了數種歐洲語言的語言分析。另一翻譯記憶系統 Fluency 提供了區段中詞彙的 POS 賦碼功能，不過僅用這一功能作為術語庫的參考，尚未拓展到翻譯記憶的匹配當中。儘管如此，多數學者同意將語言分析加入到翻譯記憶系統當中是未來不可逆轉的趨勢。

二、翻譯記憶（TM）與機器翻譯（MT）的融合

作為低於句級的匹配機制的理論基礎，雙語詞組對齊演算法已經在統計機器翻譯（SMT）和基於實例的機器翻譯（Example-based Machine Translation，EBMT）發展到了相當成熟的地步。將這一演算法移植到翻譯記憶系統當中來，便促成了新一代翻譯記憶的誕生。因此 Shuttleworth 與 Lagoudaki（2006）均認為，翻譯記憶（TM）和機器翻譯（MT）的融合已經成為未來翻譯記憶發展的潮流。Atril 公司的 Déjà Vu 從基於實例的機器翻譯（簡稱 EBMT）中借鑒技術，支援系統將兩個獨立的區段合併，一併對某一待譯區段提供匹配，類似於 EBMT 中的語塊重組。Masterin 也通過他們號稱的「知識庫」提供類似的匹配演算法。由於翻譯記憶（TM）和 SMT 以及 EBMT 均基於大規模雙語語料庫，因此 Melby（2006, p.4）

預測，未來將會出現將翻譯記憶和機器翻譯整合起來的一體化電腦輔助翻譯系統，共用同樣的雙語語料庫。

TM 和 MT 融合的又一趨勢是將 MT 直接內置在翻譯記憶系統當中。中國的雅信 CAT 在 2.5 版就內置了機器翻譯引擎，可以給使用者提供預翻譯，但翻譯品質較差，影響不大。近年來，以 Google 為代表的統計機器翻譯（Statistical Machine Translation，簡稱 SMT）異軍突起，以其優秀的翻譯品質在 NIST 評比中屢獲第一（王正、孫東雲，2009b，頁 73）。有鑑於此，一些翻譯記憶工具如 TRADOS 2009、Wordfast Pro、Fluency 已經直接將 Google 的機器翻譯引擎內置在系統當中，其翻譯結果為譯者提供了非常有用的參考。而 Google 推出的譯者工具包（Translator Toolkit）也同樣支援翻譯記憶、術語庫，並在缺乏翻譯記憶的情況下為譯者提供 Google 機器翻譯的譯文，而譯者自行上傳的術語庫可以改善機器翻譯的譯文結果。

除直接將 SMT 的翻譯結果整合入 TM 外，另一頗具吸引力的做法是「互動式機器翻譯（Interactive Machine Translation）」（Foster, Isabelle, & Plamondon, 1997）或「互動式預測機器翻譯（Interactive Predictive Machine Translation，簡稱 IPMT）」（Barrachina et al., 2009, p. 4）。加拿大蒙特利爾大學等合作建設的 TransType 專案（Langlais, Foster, & Lapalme, 2000; Langlais & Lapalme, 2002）採用了文本預測（text prediction）技術，在互動式翻譯環境下，SMT 系統根據譯者鍵入的詞彙自動判斷所需輸入的片語和句子並在鍵入行下方予以自動提示，譯者可接受提示，亦可鍵入自己的譯文。系統根據譯者鍵入的譯文自動更新翻譯記憶，不斷提高提示精度[5]。Barrachina et al. (2009, p. 4) 認為，以 TransType 為代表的 IPMT 將人機交互的重點從過去的原文消歧（disambiguation）轉移到譯文的生成上來，將 MT 嵌入到互動式翻譯環境下，取長補短，實現 MT 和 TM 的優勢互補。此外，Koehn 與 Haddow（2009）推出了基於網路的 IPMT 方案 Caitra（www.caitra.org），具有三大功能：文本預測、輸入選項和譯後

編輯。譯者在輸入譯文時，系統給予自動片語完成（unit completion）的提示，並在螢幕下方通過 SMT 引擎對原文中各片語提供了其它參考譯文，譯者亦可用滑鼠逐個點擊選中的片語譯文來完成該句的翻譯。翻譯完成後，譯者可對譯文進行譯後編輯。所有鍵盤和滑鼠操作均記錄下來，以支援翻譯過程研究。這一 TM 和 MT 融合的範例為未來的翻譯記憶提供了可行的發展方向。

三、注重語境的文本記憶

翻譯記憶發展的另一趨勢是保持文本區段的完整性，避免孤立、零散的翻譯記憶。在翻譯中，我們強調根據語境來選詞，確定文本的意義。據 Kussmaul（1995）和 Jääskeläinen（1993）通過實證研究發現，職業譯者和高級學習者注重翻譯的整體策略和語用因素，而普通學習者往往在翻譯中忽視這些因素（引自 Colina, 2003, p. 34）。而翻譯記憶的一大劣勢就是忽視翻譯的上下文語境，因而不利於從語用和語篇角度對原文進行考察並翻譯，容易使翻譯學習者養成不良的翻譯習慣（王正、孫東雲，2009c，頁 20）。尤其是在低於句級的翻譯記憶系統中，譯者在待譯區段中可能遇到許多零碎、孤立的對應語塊，如何根據語境去做出適當的判斷成為一大難題。有鑑於此，國際本地化行業標準協會 LISA 於 2007 年發佈了 XML 文本記憶（XML: TM）1.0 標準[6]。根據該標準，文本記憶（text memory）包括作者記憶（author memory）和翻譯記憶（translation memory）。前者是指對 XML 格式編碼的文檔進行劃分區段，所有區段及其修訂的全部歷史記錄都保存在這份 XML 文檔當中。後者是指當 xml:tm 格式命名空間的文檔準備開始翻譯時，命名空間自身決定了要翻譯的文本（LISA, 2007）。目前，各大公司已經紛紛推出了注重語境的翻譯記憶產品，如 TRADOS 2009、Déjà Vu、MemoQ 等，將記憶庫中的區段與原文中的相鄰區段通過設定的命名空間來組織起來。如果系統中待譯區段與記憶庫中某區段達到了 100% 的完全匹配，系統則對其相鄰區段進行

比較，確保它們擁有同樣的上下文語境，因而達到了所謂 101% 的匹配率。

四、網路大規模翻譯記憶庫的發展。

資料驅動的翻譯記憶系統必將在未來獲得更大的競爭優勢。Google 公司一篇著名的文章 “The Unreasonable Effectiveness of Data” 指出，即使語言資料魚龍混雜且未加處理，只要達到海量的規模，依然比人工「淨化」過的語料庫更有說服力（Halevy, Norvig, & Perira, 2009）。同樣，自然語言在實際使用中，如果記憶庫不夠大，則包含的區段在待譯文本中重複出現的幾率偏小。就特定的項目而言，已有的翻譯記憶往往數量有限。正所謂「眾人拾柴火焰高」。如果分佈在全球各地的譯者將翻譯記憶庫彙集起來，則記憶庫可能達到驚人的數量級，其匹配率也會大大提高。結合基於語塊的翻譯記憶系統，必將為譯者提供諸多便利。因此基於網路共用的超大規模翻譯記憶庫已經蔚然成風。Wordfast 公司在其伺服器上推出了超大規模翻譯記憶（Very Large Translation Memory，簡稱 VLTM），包含了海量各語種翻譯記憶，由用戶自願分享並經專人審核過，以保證品質，而 Google 公司也在譯者工具包裡提供了全球共用 TM 的服務，譯者可以自行選擇共用自己的翻譯記憶與否，也可以接受網路上其它網友共用的翻譯記憶。

此外，翻譯自動化用戶協會（Translation Automation User Society，簡稱 TAUS）也推出了他們的線上超大規模翻譯記憶庫 TAUS DATA。該記憶庫由自願加入協會的會員單位簽署協定並共用，訂閱該資料庫，即可根據不同許可權在一年時間內下載一定數量的翻譯記憶句對。由於加入協會的均為大型跨國企業如 Dell、Oracle、HP、CA、Symantec 等，因此資料的權威性有了保證，而且資料的版權問題也得到了有效解決。目前，GlobalSight 已經與 TAUS Association 合作，成功地將 TAUS 27 億詞的翻譯資料納入其翻譯記憶系統當中[7]，採用 XLIFF 作為資料交換模式，使用者可通過 WebServices API 在 TM 系統當中直接使用這一超大規模翻譯記

憶，因而有望大大提高 TM 的匹配精度和利用率，減輕工作強度。這一模式勢必將對未來的翻譯行業產生深遠的影響。

五、翻譯記憶的分句標準和交換標準更加普及

長期以來，各翻譯記憶系統的開發廠商各自為政，使用各自獨有的檔案格式來存儲翻譯記憶，各系統之間存在著嚴重的技術壁壘，致使某一翻譯記憶系統的使用者無法將現有的翻譯記憶移植到其它翻譯工具當中，也部分造就了 TRADOS 等軟體巨頭的壟斷地位。即使這些翻譯記憶可以用純文字格式匯出，但其存儲的格式符號和其它重要資訊均無法保留。在全球化的今日，翻譯記憶交換的通用格式成為迫在眉睫的一大問題。

有鑑於此，國際本地化行業標準協會 LISA 推出了基於 XML 技術推出了一批翻譯交換標準，包括翻譯記憶交換格式 TMX（Translation Memory eXchange）、術語交換標準 TBX（Term Base eXchange）、斷句規則交換標準 SRX（Segmentation Rule eXchange）等，而資訊標準組織 OASIS（Organization for the Advancement of Structured Information Standards）則推出了本地化文檔交換標準 XLIFF（XML Localization Interchange Format）等等。目前，絕大多數翻譯記憶系統如 TRADOS、Wordfast、Heartsome 等已經為上述標準提供了支援，這給翻譯記憶和術語的交換共用帶來了很大方便。上文提到的 TAUS 為加入協會的會員或訂戶提供的翻譯記憶下載服務，就是使用了 TMX 標準，而翻譯記憶製作工具 AlignFactory 也提供了 TMX 格式的檔輸出。而 SRX 的實施，則為翻譯記憶的劃分區段提供了標準，基於該標準進行劃分區段，有利於翻譯記憶中區段的對應，從而有效提高了區段複用的幾率。

六、新興翻譯記憶系統的缺陷與不足

目前，業界普遍認為低於句級的翻譯記憶前景看好，並紛紛推出了相應的產品。但作為新興事物，低於句級的翻譯記憶如何進行匹配尚無固定

成熟的模式，有的採用術語提取的方式提取獨立的專門記憶庫來存儲語塊區段，有的採取臨近字元段匹配的方式逐個比較，反映出這一新生事物有著其內在缺陷和不足。

（一）語言局限性

由於採用通用的字串演算法和劃分區段方式，句級翻譯記憶系統不受語種限制，只要支援國際標準編碼方式的語言就可以在系統中進行處理。因此，目前 TRADOS 等主流翻譯記憶軟體依然採取了字串匹配的演算法，但以語言學增強為特色的 Similis 等新型翻譯記憶的語言分析模組則在匹配準確度和結構分析方面更具優勢。當然，要具備這一優勢，則必須針對特定的語言制訂語法分析規則和標注模式，這大大限制了其可用性。Similis 它僅支持英語、法語、德語、義大利語、荷蘭語、西班牙語等少數幾種歐洲語言，而另一代表性產品 Masterin 僅支援英語、芬蘭語的互譯。眾所周知，這些歐洲語言在結構上與漢語等亞洲語言迥異，要將這一模式移植到漢語等語言上，尚需時日，其效果如何也有待驗證。此外，翻譯記憶在限定了語種的同時，也失掉了一大競爭優勢。正如 Benito（2009）指出，翻譯記憶優於機器翻譯的一個方面就是可以用於小語種的翻譯工作，因為傳統翻譯記憶工具不需進行任何語言分析。因此，在匹配效果和語種限制之間必須有所取捨。

（二）匹配結果超量

Melby（2006, p. 4）認為，基於語塊的翻譯記憶系統面臨的一大挑戰是符合匹配區段的結果太多，使得譯者無所適從。看完某個語塊卻無法提取任何有用的譯文，只會浪費譯者的時間。這一點在翻譯記憶庫建設的初期並不明顯，但隨著記憶庫不斷增大，可匹配的區段不斷增多，符合匹配的區段反而造成了瓶頸。因此翻譯記憶系統不應簡單地將大量相關區段列出，而是根據區段中其它詞彙的相似性數量等語境資訊，將可能性最大的譯文語塊區段列出（Simard & Langlais, 2001）。要做到這一點，還需要開發人員不斷的努力探索。

（三）結構局限性

低於句級的翻譯區段並不能替代句級區段，而是作為句級區段的有益補充。區段越長，包含的語境資訊越多，匹配也就更加準確，但匹配率也相應下降；以語塊為單位的區段越短，就越容易與待譯區段相匹配，但符合匹配要求的區段也可能越多，過多的資訊也會讓譯者吃不消。此外，如果僅在系統中保存句級區段，而對待譯的句級區段與記憶庫中所有句級區段進行 n-gram 分析可能的語塊匹配，則會佔用大量系統資源，拖慢反應速度。因此經過語言分析後提取的語塊區段應存入翻譯記憶庫，才能提高翻譯記憶系統的反應速度，在匹配中提高系統運行效率。蘇明陽、丁山（2009，頁 88）建議將翻譯的「轉換單位」（基本等於語塊）作為較為穩定的部分存入術語庫待用，但術語的定義和特性均與語塊有著較大差異，過多的語塊會降低術語庫的品質。正如 Bowker（2002, p. 104）指出，“disk space”、“drive” 可以算是術語庫中的術語，但 “check for disk space on the drive” 只能看做語塊而不是術語。因此在語塊的儲存和提取方式上，SDL TRADOS 2009 採用了獨立的翻譯記憶庫模式即 AutoSuggest™ 模組從現有翻譯記憶中提取語塊，簡便易行，但缺點是缺乏上下文語境的提示。如何將獨立的語塊與相關上下文結合起來，根據上下文自動匹配並顯示相關資訊，將成為下一步翻譯記憶系統研究的努力方向。

肆、結論

經過數十年的發展，翻譯記憶系統愈發成熟，在匹配演算法方面，從最初的完全匹配發展到如今的模糊匹配、低於句級匹配乃至語言學匹配和語境資訊匹配，大大提高了匹配精度和利用率；在使用者介面上，從當初的單一支援 Word 文檔（如 TRADOS、Wordfast 等）到如今的獨立編輯介面（如 TRADOS 2009 和 Wordfast Pro），全面支援各類檔案格式，介面更加友好；在記憶庫存儲方面，從當初的小規模、單機記憶庫發展到如

今的基於網路的大規模共用記憶庫，真正體現了互聯網的共用精神，促進了譯者的共同進步；在記憶交換格式方面，由當初的各自獨立的格式發展到如今的國際標準交換格式和劃分區段格式，為使用不同工具、不同系統平臺的使用者提供了交換資訊的便利；在譯文生成方面，由過去單一的譯員人工翻譯發展到人工翻譯與機器翻譯整合，高品質的機器翻譯引擎自動為譯者提供參考譯文供譯者進行譯後編輯（post-editing），進一步減輕了譯者的工作負擔；在作業系統上，由過去的 DOS 等作業系統過渡到 Windows 一統天下，再發展到如今基於 JAVA 的跨平臺使用，為用戶提供了更多靈活選擇。同時，電腦軟硬體技術的飛速進步使得各種創新理念成為可能。因此，我們完全有理由相信，翻譯記憶系統在未來會獲得更大的發展，得到越來越多的重視。

當然，翻譯記憶系統在教育界和翻譯產業界尚未得到足夠的重視，為翻譯記憶的推廣和普及帶來了諸多不利因素。究其原因，首先是對其重視不夠，概念混淆。部分翻譯界人士對電腦輔助翻譯瞭解甚少，對翻譯記憶系統甚至聞所未聞，常常誤作機器翻譯，認為「電腦永遠無法代替人腦」，甚至說 “Garbage in, garbage out”。目前，中國大陸地區開設電腦輔助翻譯或以翻譯記憶系統為基礎的相關課程的高校寥寥無幾。其次，翻譯記憶系統的應用還受制於譯者電腦使用水準（computer literacy）的限制。如今的翻譯記憶系統功能不斷增強，其模組也越發複雜，普通譯者往往需要一定學時的專門培訓才能夠基本上手，靠自學能夠精通系統操作的人少之又少。

另外一個很少有人關注的問題是，隨著翻譯記憶的發展，譯者的創造性勞動如何體現？著名文藝批評家瑞恰慈曾說過：「翻譯很可能是宇宙演化過程中誕生的迄今最為複雜的活動（Translation may very probably be the most complex type of event ever produced in the evolution of cosmos.）（Richards, 1953）。」但隨著翻譯記憶的發展，基於低於句級匹配的翻譯記憶使得譯者在面對特定表達方式時無需發揮其創造性思維，只需選擇接

受現有的結果即可，或對已有的模糊匹配結果稍加修改。此外，高品質機器翻譯的出現也使得機器翻譯加人工譯後編輯的翻譯模式漸漸成為可能。那麼譯者的語言和專業技能「門檻」是否會漸漸降低？未來的翻譯行業更需要的是精通兩種語言和文化的翻譯家，還是會電腦、會外語的「譯匠」？隨著翻譯記憶系統的蓬勃發展，這些問題也許將無法回避。Garcia（2007, p. 55）認為，隨著基於網路的翻譯記憶的發展，譯者的專業技能在就業中的重要地位逐漸喪失，而譯者喪失了他們自己創造的基於網路的翻譯記憶庫，其話語權進一步削弱，從本地化或翻譯行業寶貴的合作夥伴地位淪為了新技術的僕人。這種言論或許有點危言聳聽，但新技術在給譯者帶來便利的同時，又會帶來哪些不利因素？譯者在技術進步面前應當怎樣適應轉變？這些還有待進一步研究，讓我們拭目以待。

註釋

1. 作為施樂公司內部文檔，該文一直透過非公開管道流傳，直至 1997 年才在 Machine Translation 第 12 期正式發表。
2. 見 Language Industry Monitor 雜誌 1992 年 8/9 月刊專題報導 TRADOS: Smarter Translation Software，www.mt-archive.info/LIM-1992-11-3.pdf。
3. 在 20 世紀 90 年代早期，大多數翻譯記憶系統僅支援純文字文檔，因而不涉及格式問題。後來，各電腦輔助翻譯系統陸續支援 MS-Word 等流行的文本編輯格式，因而格式一致性也成為衡量完全匹配的指標之一。
4. 有趣的是，Somers and Diaz（2004）指出，部分廠商在推廣其翻譯記憶系統產品時也拒絕承認其匹配演算法僅基於字串的編輯距離，而是使用了“神經網路”、“複雜矩陣”（Heyn, 1998, p. 127）等術語來蒙混用戶，可見這些廠商也認識到了這一演算法本身的局限性和不足。
5. 目前該專案已推出 TransType2（Macklovitch, 2006），使用介面見 http://rali.iro.umontreal.ca/Traduction/TransTypeSession.gif
6. http://www.xml-intl.com/docs/specification/xml-tm.html#tm
7. http://www.tausdata.org/index.php/news/144-globalsight-a-taus-data-association-complete-largest-globalsight-implementation-on-record-a-ground-breaking-tm-matching

感謝詞

作者衷心感謝匿名審稿人和美國楊百翰大學 Alan K. Melby 教授的幫助。本文得到上海外國語大學 211 工程三期重大課題子專案“基於技術的翻譯研究”課題資助，課題編號 211YYWZ001。

參考文獻

黃俊紅、範雲、黃萍（2007）。雙語平行語料庫對齊技術述評。**外語電化教學，6**， 21-25。

蘇明陽、丁山（2009）。翻譯單位研究對電腦輔助翻譯的啟示。**外語研究，6**，84-89。

王正、孫東雲（2009a）。利用翻譯記憶系統自建雙語平行語料庫。**外語研究，5**，80-85。

王正、孫東雲（2009b）。統計機器翻譯系統在網路翻譯教學中的應用。**上海翻譯，1**，73-77。

王正、孫東雲（2009c）。翻譯記憶在翻譯教學中的優勢和局限性。**外語界，2**，16-22。

徐彬、郭紅梅、國曉立（2007）。21世紀的電腦輔助翻譯工具。 **山東外語教學，4**，79-86。

Alesiani, E. (1994). Translation tool technology –a WHA evaluation report. *The Globalization Insider, 7*. Retrieved August 10, 2010, from http://www.lisa.org/globalizationinsider/1994/07/translation_too.html.

ALPAC. (1966). *Languages and machines: Computers in translation and linguistics.* A report by the Automatic Language Processing Advisory Committee. Division of Behavioral Sciences, National Academy of Sciences, National Research Council, Washington, D.C.

Baldwin, T. (2010). The hare and the tortoise: speed and accuracy in translation retrieval. *Machine Translation*, *23*(4), 195-240.

Barrachina, S., Bender, O., Casacuberta, F., Civera, J., Cubel, E., Khadivi, S., Lagarda, A., Ney, H., Tomás, J., Vidal, E., & Vilar, J. (2009). Statistical approaches to computer-assisted translation. *Computational Linguistics*, *35*(1), 3-28.

Benito, D. (2009). Future trends in translation memory. *Revista Tradumàtica*, *7*, 1-8.

Bowker, L., & Barlow, M. (2004). Bilingual concordancers and translation memories:

A comparative evaluation. In E. Yuste Rodrigo (Ed.), *Proceedings of the Second International Workshop on Language Resources for Translation Work, Research and Training* (pp.70-83). Geneva: Switzerland.

Bowker, L. (2002). *Computer-aided translation technology: a practical introduction.* Ottawa: University of Ottawa Press.

Colina, S. (2003). *Translation teaching: from research to the classroom: A handbook for teachers.* New York: McGraw-Hill.

Colominas, C. (2008). Towards chunk-based translation memories. *Babel, 54*(4), 343-354.

EAGLES. (1996). *EAGLES evaluation of natural language processing systems*. Final Report. EAGLES Document EAG-EWG-PR.2. Center for Sprogteknologi, Copenhagen. Retrieved August 10, 2010, from http://issco-www.unige.ch/projects/ewg96/ewg96.html.

Elita, N., & Gavrila, M. (2006). Enhancing translation memories with semantic knowledge. In *Proceedings of the First Central European Student Conference in Linguistics* (pp. 24-26). Budapest.

Feiliang, R. & Shaoming L. (2006). Building Translation Memory System by N-Gram. In *The 20th Pacific Asia Conference on Language, Information and Computation* (pp.452-458). Wuhan, China.

Foster, G., Isabelle, P., & Plamondon, P. (1997). Target-text mediated interactive machine translation. *Machine Translation*, *12*(1-2), 175-194.

Garcia, I. (2007). Power shifts in web-based translation memory. *Machine Translation*, *21*, 55-68.

Gotti, F., Langlais, P., Macklovich, E., Bourigault, D., Robichaud, B., & Coulombe, C. (2005). 3GTM: a third-generation translation memory. In *3rd Computational Linguistics in the North-East (CLiNE) Workshop* (pp. 26–30). Gatineau, Québec.

Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *Intelligent Systems, 24*(2), 8-12.

Heyn, M. (1998). Translation Memories: Insights and Prospects. In L. Bowker, M. Cronin, D. Kenny & J. Pearson (Eds.), *Unity in Diversity? Current Trends in Translation Studies* (pp. 123-136). Manchester: St. Jerome Publishing.

Hutchins, J. (1998). The origins of the translator's workstation. *Machine Translation, 13*(4), 287-307.

Kavak, P. (2009). *Development of a translation memory system for Turkish to English*. Unpublished Master dissertation. Boğaziçi University, Turkey.

Kay, M. (1997). The proper place of men and machines in language translation. *Machine Translation*, *12*(1/2), 3-23.

Koehn, P., & Haddow, B. (2009). Interactive Assistance to Human Translators using Statistical Machine Translation Methods. In *Machine Translation Summit XII* (pp. 73-80). Ottawa, Canada.

Lagoudaki, E. (2006). Translation Memories Survey 2006: Users perceptions around TM use. In *Proceedings of ASLIB International Conference 'Translating and the Computer 28'* (pp. 15-16). London: Aslib.

Langlais, P., & Lapalme, G. (2002). TransType: Development-evaluation cycles to boost translator's productivity. *Machine Translation (Special Issue on Embedded Machine Translation Systems*, *17*(2), 77-98.

Langlais, P., Foster, G., & Lapalme, G. (2000). Unit completion for a computer-aided translation typing system. *Machine Translation*, *15*(4), 267-299.

LISA. (2007). *XML Text Memory (xml: tm).* Retrieved November 28, 2010, from http://www.lisa.org/XML-Text-Memory-xml.107.0.html.

Macken, L. (2009). In search of the recurrent units of translation[J]. *Linguistica Antverpiensia New Series—Themes in Translation Studies*, *7*, 195-212.

Macklovitch, E. (2006). TransType2: The last word. In *Proceedings of LREC 2006*. (pp.167-172). Genoa, Italy.

Macklovitch, R., & Russell, G. (2000). What's been forgotten in translation memory. Envisioning Machine Translation in the Information Future. *Lecture Notes in Computer Science*, 205-207.

Melby, A., & Warner, C. T. (1995). *The possibility of language: A discussion of the nature of language, with implications for human and machine translation.* Amsterdam/Philadelphia: John Benjamins.

Melby, A. K. (2006). MT+TM+QA: The Future is Ours. *Tradumatica, 4*, 1-6.

Papineni, K., & Roukos, S. (2002). Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp.311-318). Philadelphia.

Planas, E. (2000, May). Extending Translation Memories. In J. Hutchins, A. Clarke & G. Ansparch (chair), *Harvesting Existing Resources.* Symposium Conducted at EAMT Machine Translation Workshop, Ljubjana, Slovenia.

Planas, E. (2005). SIMILIS: Second-generation translation memory software. *In Proceedings of the 27th International Conference on Translating and the Computer (TC27)* (pp.331-339). London: Aslib.

Quah, C. K. (2006). *Translation and technology.* Hampshire/New York: Palgrave McMillan.

Raya, R. (2007). XML in Localisation: Reuse translations with TM and TMX—Reduce translation time and effort with the aid of XML standards. *Globalization Insider*

,5. LISA. Retrieved August 1,2010, from http://www.lisa.org/globalizationinsider/2007/05/xml_in_localisa.html.

Richards, I. A. (1953). Towards a theory of translating. In A. F. Wright (Ed.), *Studies in Chinese Thought* (pp. 247-262). Chicago: University of Chicago Press.

Schäler, R. (1996). Machine translation, translation memories and the phrasal lexicon: The localisation perspective. *Proceedings of TKE-96: EAMT Workshop on Machine Translation*, 21-33. Vienna, Austria.

Shuttleworth, M., & Lagoudaki, E. (2006). *Translation memory systems: technology in the service of the translation professional*. Retrieved August 10, 2010, from http://translation.hau.gr/telamon/files/MarkShuttleworth_ElinaLagoudaki_PaperAICTI.pdf.

Simard, M., & Langlais, P. (2001). Sub-sentential exploitation of translation memories. In *Proceedings of the Machine Translation Summit VIII* (pp. 335-340). Santiago De Compostela, Spain.

Simard, M. (2003). Translation Spotting for Translation Memories. In *Proceedings of the HLT-NAACL 2003 Workshop on Building and using parallel texts: data driven machine translation and beyond*, Vol. Ⅲ, 65 – 72. Edmonton, Canada.

Somers, H., & Diaz., G. F. (2004). Translation Memory vs. Example-based MT: What is the difference? *International Journal of Translation, 2* (16), 5-33.

Sumita, E., & Tsutsumi, Y. (1988). Translation aid system using flexible text retrieval based on syntax-matching. In *Proceedings of The Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*. Pittsburgh, Pennsylvania: CMU. Retrieved September 30, 2010, from http://www.mt-archive.info/TMI-1988-Sumita.pdf.