

## 《澳門語言學刊》主辦單位變更公告

依據 2022 年 3 月 24 日澳門大學人文學院與澳門語言學會簽訂的合作協定書，《澳門語言學刊》主辦單位由原澳門語言學會正式變更為澳門大學人文學院，協辦單位為澳門語言學會。本刊主辦單位變更後，將以成為海內外語言學研究的高水平成果的發表平台與信息交流樞紐為目標，積極推動中國語言研究的發展。竭誠歡迎海內外專家學者賜稿。

# 《澳門語言學刊》

2022 年第 1 期 總第 59 期

主辦  
協辦  
定價

澳門大學人文學院  
澳門語言學會  
2022 年 6 月出版  
澳門幣 30 元 / 港幣 30 元 / 人民幣 26 元

主任委員  
委員

程祥徽 (澳門科技大學)  
蔡維天 (清華大學)  
鄧思穎 (香港中文大學)  
馮勝利 (北京語言大學)  
胡 波 (澳門城市大學)  
黃 翊 (澳門理工大學)  
金鉉哲 (延世大學)  
劉街生 (中山大學)  
劉樂寧 (哥倫比亞大學)  
羅言發 (澳門大學)  
彭 睿 (新加坡國立大學)  
齊 冲 (巴黎第七大學)  
司富珍 (北京語言大學)  
汪國勝 (華中師範大學)  
王宇嬰 (澳門理工大學)  
魏慧萍 (澳門科技大學)  
徐 丹 (法國東方語言文化學院)  
徐 杰 (澳門大學)  
嚴翼相 (漢陽大學)  
楊凱榮 (東京大學)  
袁毓林 (澳門大學)  
周 荐 (北京師範大學)  
左思民 (高麗大學)

榮譽主編  
顧問  
主編  
副主編  
助理主編  
刊名題簽

吳志良、程祥徽  
徐 杰  
羅言發  
胡 波  
劉 銳  
程祥徽

投稿信箱  
本刊網址

linguistics@um.edu.mo  
[https:// linguistics.fah.um.edu.mo](https://linguistics.fah.um.edu.mo)



# 時間方位短語語法規則在中葡機器翻譯後處理中的應用

## Application of Temporal Localizer Phrase Rules in Chinese-Portuguese Machine Translation Post-processing

◎ 盧春暉 / 澳門大學葡文系

**提 要：**機器翻譯是計算語言學的重要領域之一，融合了計算機科學、語言學、統計學等多門學科，已經走過了六十多年的歷程。本文以基於同步約束語法的中葡翻譯系統 PCT 為研究對象，展示了時間方位短語語法規則在機器翻譯後處理中的應用，並通過 BLEU 評測證實其有效性。我們認為，語言學研究和規則的制定構成優化機器翻譯的重要路徑，對計算語言學的發展亦具有重要意義。

**關鍵詞：**計算語言學 機器翻譯 規則 中葡翻譯 同步約束語法

**Key words:** computational linguistics; machine translation; rules; Chinese-Portuguese translation; Constraint Synchronous Grammar

### 一、引言

計算語言學 (Computational Linguistics) 起源於 20 世紀 50 年代，是採用計算機技術來分析、研究和處理人類自然語言的學科。計算語言學的研究首先是從機器翻譯開始的 (馮志偉，2019)。

以 1954 年美國喬治敦大學與 IBM 公司合作完成人類歷史上第一次機器翻譯實驗為起點，機器翻譯的研發

已經經歷了六十多年。馮志偉將其發展過程分為三個時期：草創期 (1954-1970)、復蘇期 (1970-1976) 和繁榮期 (1976 至今)。中國的機器翻譯發展則包括四個階段：草創期 (1956-1966)、停滯期 (1966-1975)、復蘇期 (1975-1987) 和繁榮期 (1987 至今) (馮志偉，1997)。機器翻譯技術在發展過程中發生了巨大的變化，主要體現在應用範圍大大延伸、計算機輔助翻譯在市場上的成功及應用形式更加多樣化 (劉群，2012)。



機器翻譯主要經歷了兩大發展階段，即理性主義方法主導時期和經驗主義方法主導時期。理性主義方法指基於規則的機器翻譯 (Rule-based Machine Translation)；經驗主義方法包括統計機器翻譯 (Statistical Machine Translation) 以及 2014 年以來快速發展的神經機器翻譯 (Neural Machine Translation)。理性主義方法依靠語言學家，而經驗主義方法主張以數據為中心，通過數據模型描述自然語言的轉換過程 (劉洋，2017)。從上世紀九十年代開始，統計機器翻譯逐漸進入主導地位，在機器翻譯領域語言學和語言學家在逐漸淡出視線。然而，完全依賴統計技術的發展是畸形的。統計機器翻譯消耗資源巨大，對語料庫質量的要求極高，對源語言和目標語言的語法相似度存在一定依賴性。在到達了一定精度後，統計機器翻譯遇到了瓶頸，而突破瓶頸的方法之一就是重新回歸語言學知識，將規則與統計技術相結合 (張瓏馨，2013)。而如今處於主流地位的神經機器翻譯亦面臨譯文不夠忠實、罕見詞處理困難等問題。在機器翻譯模型中融入語言學信息再次成為一個熱門話題 (郭望皓等，2021)。

我們的研究對象為澳門大學自然語言處理與中葡機器翻譯實驗室研發的中葡翻譯系統 PCT。PCT 為混合系統 (hybrid system)，採用基於語料和基於規則兩種策略。規則的實現依靠同步約束語法 (Constraint Synchronous Grammar) (Wong et al., 2012)。本文研究、建立新的時間方位短語語法規則，將其應用於 PCT 系統翻譯的後處理階段，以此為例展示語言學在提升機器翻譯質量中的作用。

## 二、概念介紹

### 2.1 同步約束語法

如上文所述，PCT 採用混合機制，結合了基於語

料的機器翻譯和基於規則的機器翻譯兩種方法。規則即同步約束語法，建立在上下文無關語法 (Context Free Grammar) 及同步的基礎上 (Oliveira et al., 2010; Wong et al., 2006; Wong et al., 2005; Wong et al., 2011)。同步約束語法的基本形式如下所示 (Wong et al., 2007)：

$$S \rightarrow \text{source sentential pattern} \{ [\text{target sentential pattern; control conditions}], [\text{target sentential pattern; control conditions}], \dots \}$$

左側 S 為縮略符 (reduced syntactic symbol)，箭頭指向來源句型 (source sentential pattern)，代表原文句法結構；右側括號中包括兩部分內容，即目標句型 (target sentential pattern) 和控制條件 (control conditions)。目標句型即輸出句型，代表了譯文的結構。同步約束語法不同之處在於譯文生成的豐富性和多樣性。在原文句法結構相同的情況下，系統可以根據控制條件的不同，生成相應不同的譯文。在同一條規則中，目標句型和控制條件組合的數量可以無限增加，即保證了翻譯的準確性和多樣性，又避免了規則過於臃腫和龐大。見以下的例子 (Wong et al., 2007)：

$$S \rightarrow NP_1 VP^* NP_2 PP NP_3 \{ [NP_1 VP^1 NP_3 VP^2 NP_2; VP_{\text{category}} = vb1, VP_{\text{sense of subject}} = NP_1 \text{sense}, P_{\text{sense of indirect object}} = NP_2 \text{sense}, VP_{\text{sense of object}} = NP_3 \text{sense}], [NP_1 VP NP_3 NP_2; VP = vb0, VP_{\text{sense of subject}} = NP_1 \text{sense}, VP_{\text{sense of indirect object}} = NP_2 \text{sense}] \}$$

在以上例子中，原文對應的句型為  $NP_1 VP^* NP_2 PP NP_3$ 。在實際操作時，PCT 中的句法分析器 (parser) 首先對待譯的語句進行句法分析 (parsing)，若與該句型對應，則意味著該條規則將得以應用。接著，系統將結合控制條件對句子進行分析，若與第一組條件 ( $VP_{\text{category}} = vb1 \dots VP_{\text{sense of object}} = NP_3 \text{sense}$ ) 相符，則按照第一組譯文結構 ( $NP_1 VP^1 NP_3 VP^2 NP_2$ ) 輸出譯文；若與第二組控制條件相符，則按照第二組譯文結構輸出譯文。

除了準確、多樣和簡潔以外，同步約束語法還可以



避免語言差異性造成的各種問題 (Wong et al., 2012)。中文和葡文分屬漢藏語系和印歐語系，句法差異非常大，舉例如下：

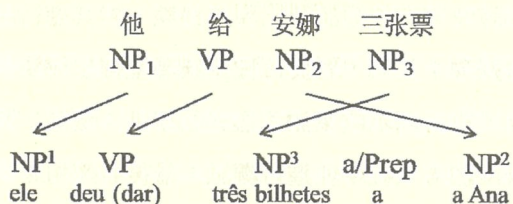
(1) 不對等。葡文中的一個動詞在中文裡可能對應的不是一個詞，而是一個框架結構。例如：Ela emprestou ao Pedro uma caneta (她把一支筆借給佩德羅)，動詞為 emprestou (借)，而到了中文裡對應的則是“把……借給”把字句型的框架結構。

(2) 順序改變。中文的習慣是形容詞在前，名詞在後；而葡文則相反。例如“漂亮的房子”在葡文中是 casa bonita (casa 房子；bonita 漂亮的)。

(3) 缺省和增加。例如，“澳門大學”在葡文中為 Universidade de Macau，增加了前置詞<sup>[1]</sup> de；葡文句子 Nasci em 1991 (我出生；在；1991) 翻譯成中文是“我出生在 1991 年”，增加了“年”字。

例如，將“他給安娜三張票”翻譯成葡文為 Ele deu três bilhetes à Ana，各成分對應關係如下：

圖 1 中葡句法差異舉例



我們發現，其中出現了順序改變（“安娜”與“三張票”）以及內容增加（葡文中出現了前置詞 a，在中文中無此對應）。同步約束語法可以非常靈活地處理這些語言差異：

$$S \rightarrow NP_1 VP NP_2 NP_3 \{ [NP_1 VP NP_3 a/Prep NP_2], \dots \}$$

因此，無論原文和譯文之間的結構差異有多大，同步約束語法都可以通過句法結構對其進行描述，並建立起對應關係。

## 2.2 時間方位短語及葡文中的對應

基於規則的機器翻譯依託語言學知識，即需要對語言現象和句法結構有一定程度的瞭解。翻譯規則實質上可以認為是對兩種語言結構的分別描述和聯繫性的建立。

時間方位短語的翻譯是本次實驗的研究範圍。方位短語的核心元素是方位詞，方位詞是漢語中特有的詞類，不存在於英文中 (蔡永強, 2010: 51)，亦不存在於葡文中。國內關於漢語方位詞的研究非常豐富，涉及詞類、句法結構、語義、語用、隱喻認知、英漢對比等多個方面。劉月華對方位詞的定義如下：方位詞是指表示方向和相對位置關係的名稱的詞。方位詞按其結構可分為兩種：單純方位詞和合成方位詞。單純方位詞都是單音節的，有東、南、西、北、上、下、前、後、左、右、裡、外、內、中、間、旁；單純方位詞前邊加上“以”或“之”或者後邊加上“邊”、“面”、“頭”就構成合成方位詞，表示方位、處所或時間 (劉月華等, 2004: 50-55)。蔡永強表示，方位詞是一個獨立的詞類，和處所詞、時間詞共同構成方位範疇。處所詞只能表示空間位置，時間詞只能表示時間位置，而方位詞既可以表達空間位置也可以表達時間位置 (蔡永強, 2010: 8-10)。

關於方位詞的句法功能，袁毓林認為方位詞是黏著詞，不能單獨使用，它可以黏附在名詞性成分後面，構成方位短語 (袁毓林, 2010: 58)。例如：五天前、會議後、商場邊。除了名詞性成分以外，方位詞還可以與動詞、動詞短語或主謂短語共同構成方位短語，如“上班前”、“完成項目後”、“活動閉幕後”等。我們將此類如“前”、“後”、“之前”、“之後”、“以後”等可以表示時間的方位詞與名詞、動詞或小句組成的結構稱為時間方位短語，即以下三種結構 (呂叔湘, 1999: 448, 269-270)：①名詞 / 名詞段 + 前 (之前、以前) / 後 (之後、以後)。例如：運動會前；登記日之後；②動詞 / 動詞段 + 前 (之前、以前)



／後（之後、以後）。例如：啟程前；宣告破產後；  
③小句＋前（之前、以前）／後（之後、以後）。例如：澳門回歸前；計劃實施後。

葡語中不存在方位詞，因此也不存在方位短語。趙元任認為，當漢語方位詞翻譯成外文時，往往與一個介詞（preposition）相當。由於方位詞往往位於短語尾部，也有人將其稱之為“後置詞”（postpositon）（趙元任，1979：278）。

葡語前置詞分為簡單前置詞和複合前置詞兩種。簡單前置詞如 a, com, de, desde, entre；複合前置詞如 dentro de, debaixo de, antes de, depois de 等。葡語前置詞在功能上與漢語方位詞類似，都具有指明空間和時間位置的功能。表示時間前後關係的前置詞有 antes de（在……之前），após, depois de（在……之後）。

### 2.3 BLEU 評測方法

自上世紀八十年代末機器翻譯興起以來，機器翻譯評測的重要性日益突出。從用戶角度看，可以更好地選擇適合自己的翻譯工具；對開發者而言，可以通過評測發現當前版本中存在的問題和漏洞，從而有效地進行更新迭代。因此，機器翻譯評測的功能性具有兩個維度：一是比較不同翻譯產品的質量，二是檢測同一產品不同版本之間的區別。

機器翻譯評測可以通過兩種方式進行：人工評測和自動評測。人工評測一般由語言專家完成，根據規定的指標對翻譯文本進行評價和打分；自動評測則依賴算法由計算機完成。兩種評測方式各有所長。如人工評測較為靈活，能夠考慮語言的多樣性和可能性，缺點是人力成本和時間成本較大，處理量小且具有一定主觀性。機器評測時間成本和經濟成本較低，處理量大，可反復使用而無需重複佔用資源，但評價機制較為單一和固定。

BLEU（Bilingual Evaluation Understudy）是機器翻譯領域廣為使用的一種自動評估方法，由 Papineni 等於

2002 年提出（Papineni et al., 2002），基本原理是測量機器譯文與人工譯文之間的“距離”和“相似度”。機器譯文與人工譯文的相似度越高，意味著其質量越好。本次實驗中，我們使用 iBleu 在線評測系統（Madnani, 2011）比較 PCT 在更新前後翻譯質量的變化。

## 三、實驗方法、過程及結果

### 3.1 實驗方法

馮志偉認為，計算機對自然語言的研究和處理一般應經過四個過程：第一、形式化（formalism），將需要研究的問題從語言學角度進行形式描述，建立語言的形式化模型；第二、算法化（algorithm），把經過語言特徵工程處理的數學形式表現為計算機的算法；第三、程序化，根據算法編寫、調試程序；第四、實用化，建立實用的自然語言處理系統，並對其進行評測和不斷完善（馮志偉，2019）。

我們的研究基本遵循以上的過程。首先，識別系統在翻譯時間方位短語時出現的錯誤，對其進行語言學分析並歸類；其次，對於每種錯誤類型提出解決方案，並轉寫成符合同步約束語法形式的規則；接著，將新建立的規則加入系統中並進行調試；最後，使用新的語料進行評估。為了提高測試的可靠性，我們採用機器算法 BLEU 和人工檢驗兩種方法進行雙重驗證。

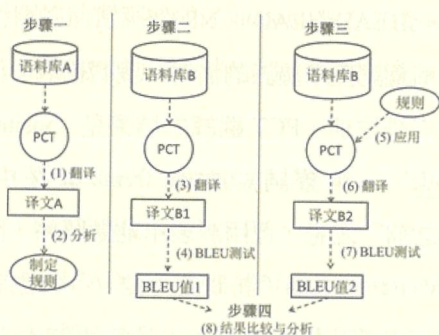
實驗中我們建立了兩個小型平行語料庫，分別命名為語料庫 A 和語料庫 B，共由 130 組中葡文對照的時間方位詞組構成。其中中文語料為原文（source text），即譯出語；葡文語料為對應的人工譯文（human translation）。語料的主要來源如下：第一，澳門特別行政區政府門戶網<sup>[2]</sup>；第二，Macahub 平臺<sup>[3]</sup>；第三，《亞洲週刊》節選部分新聞及人工譯文<sup>[4]</sup>。語料庫中的文本以新聞為主，涉及金融、醫藥、政治、教育、法律、



社會等領域，均為已公開發表的文本；人工翻譯經過母語或專業人士審核及修改，質量較高。為避免干擾，提高實驗準確性，我們僅摘取了語料中時間方位短語的部分進行研究，例如：“競選活動開始前 / antes do inicio da campanha eleitoral”。

我們從中挑選 30 組平行語料構成語料庫 A，剩下的 100 組平行語料組成語料庫 B，兩者功能不同：前者用於規則學習，後者用於規則測試。實驗主要分為四個步驟，分別為：規則學習、初次翻譯及質量評測、二次翻譯及質量評測、兩次翻譯質量評測結果對比。具體流程如下：

圖 2 實驗方法流程圖



步驟一為規則學習。我們使用 PCT 系統將語料庫 A 中的 30 個時間方位短語從中文翻譯成葡文，得到譯文 A。以語料庫中的人工譯文作為參考，對機翻譯文進行檢查，找出其中前置詞或前置詞短語使用錯誤的情況，對其進行語言學分析並歸類，最後制定出同步約束語法形式的規則。

步驟二為初次翻譯及質量評測。我們使用 PCT 系統將語料庫 B 中的 100 個時間方位短語從中文翻譯成葡文，得到譯文 B1。使用 BLEU 算法對該譯文進行評估和打分，得到 BLEU 值 1，該數值代表了本次實驗中譯文 B1 的質量。

步驟三為二次翻譯及質量評測。我們將步驟一中制定的新規則導入 PCT 系統，使用該更新後的系統對語

料庫 B 中的 100 個時間方位短語進行再次翻譯，得到譯文 B2。使用 BLEU 算法對該譯文進行評估和打分，得到 BLEU 值 2，該數值代表了本次實驗中譯文 B2 的質量。

步驟四為兩次翻譯質量評測結果對比。BLEU 值 1 與 BLEU 值 2 分別代表了 PCT 系統加入新規則前後的翻譯質量，通過數值的對比，可以對新規則的有效性進行基本的判斷。最後，我們對兩次譯文進行人工核查，確保結論的準確性。

### 3.2 實驗過程

本次實驗的重點和難點在於步驟一，即通過分析語料庫 A 的機器譯文制定新規則。語料庫 A 中含有 30 個時間方位短語，其中包括 15 個帶有“前/之前”的短語及 15 個帶有“後/之後”的短語。我們使用 PCT 系統將這些方位短語從中文翻譯成葡文，結合人工譯文進行比對、分析，對漢語方位詞的翻譯錯誤類型進行了分類和總結，如下表所示：

表格 1 語料庫 A 中方位詞翻譯錯誤類型

類別	數量 / 比例	舉例（原文 / PCT 譯文）
前置詞位置不當	24 / 80%	整合完成後 / a conformidade depois de acabar
前置詞使用錯誤	21 / 70%	六月前 / junho em frente de
漏譯 / 未識別	2 / 6.7%	計劃實施後 / o plano está a entrar em vigor hou
正確	2 / 6.7%	

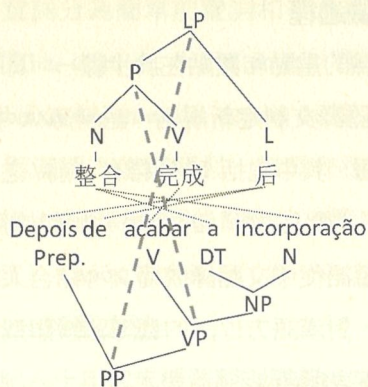
由此可見，PCT 系統在翻譯漢語方位詞時主要存在三個問題：前置詞位置不當、前置詞使用錯誤和漏譯 / 未識別。因此，我們分別從這三個方面進行分析和規則制定，以改善機器翻譯的質量。由於涉及的語料較多，我們選取兩例加以說明。

首先是前置詞位置不當的問題。我們以“整合完成後”這一短語為例，機器翻譯的結果為 a conformidade depois de acabar。前置詞 depois de（在……之後）對應



中文方位詞“後”，詞翻譯正確，但位置錯誤。原文“整合完成後”是一個由主調小句(整合/完成)和方位詞“後”構成的表示時間的方位詞短語，對應的葡文翻譯應該是一個由前置詞 após 或 depois de 引導的前置詞詞組。然而，PCT 系統在翻譯過程中將 depois de 置於譯文中間，導致句意不通。我們將原文和人工參考譯文 (depois de acabar a incorporação) 以語法樹的形式表達如下：

圖 3 語料庫 A 例句語法樹分析 1



語法樹清晰顯示了兩個對應短語的句法結構。中文部分，“整合”與“完成”構成一個短句 (P)，再與方位詞 (L) “後”一同構成方位詞短語 (LP)。葡文部分，定冠詞 (DT) a 與名詞 (N) incorporação (合併) 構成名詞短語 (NP)，再與動詞 (V) acabar (結束) 構成動詞短語 (VP)，最後與前置詞 (Prep.) depois de 構成前置詞短語 (PP)。對應來看，中文句子中的短句 (P) 對應葡文中的動詞短語 (VP)，最頂層的方位詞短語 (LP) 則對應前置詞短語 (PP)。我們將這種對應關係用公式表達出來即是：

$$S(\text{中文}) = P(N+V)+L \quad (L = \text{後} / \text{之後})$$

$$S(\text{葡文}) = \text{Prep.}+VP(V+NP)$$

我們再將這種對應的關係用同步約束語法的格式轉寫出來即成為以下規則：

$$S = NP VP PP <\text{depois,}, \text{ADV de,}, \text{PP MG} (VP)$$

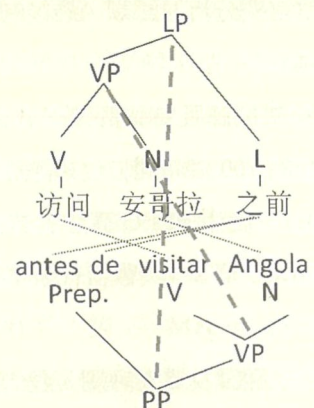
$$MG(NP) << \text{COND: WRD} (PP) = \text{"後"} | \text{WRD} (AP) = \text{"之後"} >> < \text{HEAD: HEAD} (NP) >$$

另外，我們發現，如果把中文句子中的方位詞“後”、“之後”改為“前”、“之前”，無論原文還是譯文的句子結構均不發生改變，只是葡文中的前置詞 depois de 須改為 antes de。因此，我們在上述規則的基礎上加以補充，便可以有助於機器在翻譯這兩種句子的時候提高質量：

$$S = NP VP PP <\text{depois,}, \text{ADV de,}, \text{PP MG} (NP) >< \text{COND: WRD} (PP) = \text{"後"} | \text{WRD} (AP) = \text{"之後"} >< \text{HEAD: HEAD} (NP) >, \backslash <\text{antes,}, \text{ADV de,}, \text{PP MG} (VP) > MG (NP) >< \text{COND: WRD} (PP) = \text{"前"} | \text{WRD} (PP) = \text{"之前"} >< \text{HEAD: HEAD} (NP) >$$

對於前置詞使用錯誤的情況，我們也取一例加以說明：訪問安哥拉前。PCT 翻譯的結果是“visitar Angola em frente de”。前置詞短語 em frente de 在中文裡有“在……之前”之意，但用於表示地點關係，而非時間關係。“visitar Angola (訪問安哥拉)”為動詞短語，與前置詞“之前”搭配時只能表示時間先後，因此 em frente de 在這裡使用是錯誤的。我們在建立規則之前，採取同樣的方法，將原文和人工參考譯文 (antes de visitar Angola) 以語法樹的形式表達如下：

圖 4 語料庫 A 例句語法樹分析 2





中文部分，“訪問”與“安哥拉”構成動詞短語(VP)，再與方位詞“之前”構成方位詞短語(LP)；葡語部分，visitar(訪問)與Angola(安哥拉)構成動詞短語(VP)，再與前置詞antes de構成前置詞短語(PP)。從對應關係來看，中文句子中的動詞短語(VP)與葡文句子中的動詞短語(VP)對應，結構頂層的方位詞短語(LP)與前置詞短語(PP)相對應。我們將這種關係用公式表達出來是：

$$S(\text{中文}) = \text{VP}(\text{V}+\text{N})+\text{L}(\text{L} = \text{前} / \text{之前})$$

$$S(\text{葡文}) = \text{Prep.}+\text{VP}(\text{V}+\text{N})$$

與上一個例子一樣，此處若把方位詞(L)換成“後/之後”，句子結構和對應關係也不會發生改變，只是葡語中對應的前置詞變成了depois de(之後)。我們再將以上公式用同步約束語法的格式轉寫出來，即成為以下規則：

$$\text{VP} = \text{VP NP PP} \langle \text{depois, ADV de, PP MG}(\text{VP}) \text{MG}(\text{NP}) \rangle \langle \text{COND: WRD}(\text{PP}) = \text{" 後 " | WRD}(\text{AP}) = \text{" 之後 " } \rangle \langle \text{HEAD: HEAD}(\text{VP}) \rangle, \backslash \langle \text{antes, ADV de, PP MG}(\text{VP}) \text{MG}(\text{NP}) \rangle \langle \text{COND: WRD}(\text{PP}) = \text{" 前 " | WRD}(\text{PP}) = \text{" 之前 " } \rangle \langle \text{HEAD: HEAD}(\text{VP}) \rangle$$

我們通過這樣的方式，對語料庫A中所有的PCT譯文按照錯誤類別進行了逐條分析，最終共建立了9條同步約束語法規則。在下文中，我們將使用語料庫B對這9條規則的有效性進行檢驗，判斷是否可以真正解決機器在翻譯漢語時間方位短語中出現的問題。

### 3.3 實驗結果和數據

語料庫B中共包含100組中文方位詞短語及對應的人工參考譯文。在這一部分，我們兩次使用PCT系統將語料庫B中的中文短語翻譯成葡文。兩次實驗的唯一不同之處在於：第一次我們使用原始的PCT系統，得到譯文B1；第二次我們將規則應用於系統中，即使用改良後的系統對同樣的源文本進行翻譯，得到譯文

B2。對兩次翻譯文本分別進行BLEU測試，得到的結果是實驗有效性的重要參考依據。

如前文所述，BLEU測試需要兩項基本材料：機器翻譯文本及至少一個人工翻譯文本作為對照。本次實驗中，我們使用iBleu在線測試系統評測PCT翻譯質量。iBleu的分值範圍為0至100，分數越高，代表被測試文本與對照文本相似度越高，即翻譯質量越高。經測試，譯文B1的BLEU評分為6.28，譯文B2的評分為13.26。兩次結果對比，BLEU評分增長了111.1%，證實了PCT系統翻譯質量的提升及新加入規則的有效性。

為了進一步確保結論的可靠性，我們對兩次的譯文進行了整理和人工對比，摘取部分例子如下：

表格 2 有效性測試結果

原文	人工譯文	PCT 譯文 B1	PCT 譯文 B2
啟程前	antes de partir	partir em frente de	antes de partir
2017 年之前	antes de 2017	2017 ano antes	antes de 2017
施工前	antes da construção / antes da obra	construção em frente de	antes da construção
移居前	antes de mudar de casa / antes de imigrar	mudar de casa em frente de	antes de mudar da casa
死亡前	antes da morte	a sua morte em frente de	antes da morte
二月底前	antes do final de fevereiro	fim de fevereiro em frente de	antes do fim de fevereiro
農曆新年後	depois do ano novo lunar chinês / após o ano novo lunar chinês	o ano do calendário lunar atrás de	depois do ano novo do calendário lunar
會議後	após a reunião	reunião atrás de	após a reunião
派對後	após a festa	festa atrás de	após a festa
喪失聽力後	depois de perder a audição	perder capacidade auditiva atrás de	depois de perder a capacidade auditiva

從上述例子可以看出，PCT在加入新規則後，糾正了譯文B1中出現的前置詞使用錯誤及位置不當的問題，提升了翻譯質量。



#### 四、結語

本文從PCT系統對時間方位短語的翻譯結果出發，發現、總結譯文中出現的錯誤，制定並應用規則，優化了系統的翻譯質量。本研究具有雙重意義：在語言學的層面上，我們建立了兩個小型雙語語料庫，分別用於規則學習及規則測試。在規則制定的過程中，以語法樹的形式直觀地展示和比較了中葡兩種語言在時間方位短語中的句法特徵；在計算機科學方面，我們通過案例展示了語言規則從理論到應用的過程，證實了語言學知識在機器翻譯後處理及優化中的可行性和重要性。

馮志偉（2019）認為，“計算語言學的出現，使得語言學在現代科學體系中的地位產生了明顯的變化，成為了人文科學發展的突破點和生長點，古老的語言學由一門傳統基礎的科學變成了一門領先的帶頭科學，獲得了與數學、哲學同等的地位，它的重要意義已經成為學術界的共識。”可以說，計算語言學對推動語言學的現代化和前沿化做出了巨大的貢獻。反過來說，計算語言學是人類自然語言與計算機聯接、互動的平臺，也是實現人機交互的學科保障，其最終目的是以最大程度及精確度識別、描述和處理自然語言，為各種真實的社會語言需求提供高質量的服務，其出發點和落腳點都是自然語言。因此，我們不應忽視語言學研究在計算語言學發展過程中的獨特意義。

#### 註釋：

[1] 前置詞即介詞 (preposition)。

[2] <https://www.gov.mo/zh-hant/> 澳門政府網站提供中、葡、英三語新聞。

[3] <https://macauihub.com.mo/zh/> 該網站由澳門特別行政區新聞局為推動中國和葡語國家經貿發展而建立，每週一至週五定期發送中、葡、英三語新聞簡訊。

[4] 其中包括 46 篇中文新聞稿件及葡語譯文。該譯文是在澳門大學 Autema Dis-II 項目中完成的 (Leal & Schmaltz 2010)。

#### 參考文獻：

- 馮志偉 1997 機器翻譯今昔談，《金秋科苑》第 4 期
- 馮志偉 2019 我國計算語言學研究 70 年，《語言教育》第 4 期。
- 郭望皓 范江威 張克亮 2021 融合語言學知識的神經機器翻譯研究進展，《計算機科學與探索》第 7 期。
- 劉月華 潘文娛 故韡 2004 《實用現代漢語語法（增訂本）》，北京：商務印書館。
- 劉洋 2017 神經機器翻譯前沿進展，《計算機研究與發展》第 6 期。
- 劉群 2012 機器翻譯技術現狀與展望，《集成技術》第 1 期。
- 呂叔湘 1999 《現代漢語八百詞》，北京：商務印書館。
- 張瓏馨 2013 消失的語言學家：機器翻譯發展的現狀與歧路，《西南農業大學學報》第 10 期。
- 蔡永強 2010 《漢語方位詞及其概念隱喻系統（第一版）》，北京：中國社會科學出版社。
- 袁毓林 2010 《漢語詞類的認知研究和模糊劃分》，上海：上海教育出版社。
- 趙元任 1979 《漢語口語語法（第一版）》，北京：商務印書館。
- Leal, A. L. and Schmaltz, M. 2010. “Um Compromisso Interdisciplinar: Tradução e Automatização da Temática Discursiva Chinês – Português”, in *XX Encontro da Associação de Universidades de Língua Portuguesa, Macau*.
- Madhani, N.. 2011. “iBLEU: Interactively Debugging & Scoring Statistical Machine Translation Systems Nitin”. *Proceedings of the Fifth IEEE International Conference on Semantic Computing*.
- Oliveira, F. et al. 2010. “Parsing Extended Constraint Synchronous Grammar in Chinese-Portuguese Machine



Translation”. *International Conference on Computational Processing of Portuguese Language*, Porto Alegre.

Papineni, K. et al. 2002. “BLEU: a Method for Automatic Evaluation of Machine Translation”, in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia.

Wong, F. et al. 2005. “Machine translation based on constraint-based synchronous grammar”. *International Conference on Natural Language Processing*. Berlin, Heidelberg: Springer.

Wong, F. et al. 2006. “Machine Translation Using Constraint-Based Synchronous Grammar”. *Tsinghua Science*

*and Technology*, 11 (3), 295–306.

Wong, F. et al. 2007. “Translation Technologies of Portuguese to Chinese Machine Translation System PCT Assistente”.

Wong, F. et al. 2011. “Handheld Machine Translation System Based on Constraint Synchronous Grammar”. *Proceedings of Machine Translation Summit XIII. Xiamen*.

Wong, F. et al. 2012. “Hybrid Machine Aided Translation System based on Constraint Synchronous Grammar and Translation Corresponding Tree”. *Journal of Computers*, 7 (2), 309–316.