

# 學術論文簡介的自動文步分析與寫作提示

黃冠誠 吳鑑城 許湘翎 顏孜曦 張俊盛

## Abstract

近年來，英文逐漸變成全世界學術研究最主要的溝通的媒介。而學術英文寫作，也成為非常重要的研究與教學的領域。學者也很重視，如何透過電腦的輔助，幫助一般性的語言學習，甚或特定性的學術論文寫作。學術寫作包含許多的文節類型，包括研究論文、計畫申請書、回顧與評論文節等 (Swales, 1990)。其中，研究論文占有最重要的角色。

在學術論文中，「簡介」是絕大部分論文都有的第一個節。現今，幾乎沒有學術論文，沒有「摘要」與「簡介」，而直接詳細地描述研究的目的、方法、結果。而且，對寫者和讀者而言，「簡介」在學術論文中都扮演非常重要的角色。一篇好的簡介，要能為整篇論文定調，抓住讀者的興趣，提供論文的扼要資訊。換言之，「簡介」肩負重大責任——吸引讀者注意，讀完全文。

因此，有一些研究開始分析論文簡介如何達成其溝通的任務。Graetz (1985) 發現論文簡介似乎有共同的「問題—解法」修辭結構，依序包括問題 (problem)、方法 (solution)、評估 (evaluation)、結論 (conclusion) 等部分。Swales (1990) 分析大量的論文簡介，分析歸納出一套修辭的動機與模式：「創造研究空間」(Create A Research Space, CARS)。Swales 認為論文爭取研究得到讀者的認同，有如環境中生物爭取生存空間。為此，大部分作者依循三個修辭的步驟——也就是文步 (moves)——來說服讀者。

CARS 文步	子文步與資訊內容
文步 I 界定範圍	1. 聲明研究領域的重要性，及/或 2. 聲明研究課題的廣泛性與普及性，及/或 3. 回顧與評論前人研究
文步 II 建立利基	1A. 提出與前人不同的聲明，或 1B. 指出前人研究的缺口 (gap)，或 1C. 提出本論文的研究議題 (research question)，或 1D. 說明本研究所根據的典範與傳統
文步 III 佔據利基	1A. 概述本論文的目的，或 1B. 概述本論文的方法 2. 宣布本論文的主要結果與發現 3. 指出本論文的結構

圖 1. Swales (1990) 提出的 CARS 模式的文步與資訊內容

如圖 1 所示，這三個文步包括了「界定研究範圍」、「建立利基」、「佔據利基」。在每一個文步下，又需要描述若干必要或選項的內容。另外，美國國家醫學圖書館，也主張醫學論文作者，應提供分段有標題（labeled sections）的結構化摘要（structured abstract）。

目前已經有許多學術寫作教材，透過文步分析來教導英文非母語的學生，如何寫作學術論文（如 Swales and Feak, 2004; Glasman-Deal, 2010）。也有研究者開發軟體系統（例如，Marking Mate: [writingtools.xjtlu.edu.cn:8080/mm/markingmate.html](http://writingtools.xjtlu.edu.cn:8080/mm/markingmate.html)），分析學生的作文，自動產生批改的建議與評分。但是很少有系統能夠在學生寫作中，依照文步的推進，適時地提供寫作提示與輔助。直覺上，如果我們能將大量的論文簡介加以處理，自動化分析其中每句的文步，繼而分析特定文步句子的常見片語或句型，我們將可以在寫作的過程，有效地協助學生。

然而，過去所提出的自動文步分析方法，都需費時費工標註大量論文。有鑑於此，我們提出新方法，以降低人工標註的工作量，運用於訓練統計式分類器，來預測論文簡介的句子文步，並藉以開發一個線上輔助寫作系統 *WriteAhead*。在 *WriteAhead* 的開發過程，我們採用了比 *CARS* 更簡單的文步分類，如圖 2 所示。用了這一套分類方式，系統容易自動分類文步，而使用者比較容易掌握使用於寫作過程。

我們期望此一自動文步分析工具，以及 *WriteAhead* 系統，有助於提升英文非母語者（non-native speakers, NNS），寫作學術論文的能力。在本論文中，我們提出了一套監督式機器學習的方法，能夠自動地學習如何將語料庫內的簡介句子，大略地分類為幾個文步。有了分類的句子之後，我們就可以統計各文步的連續詞頻率（ngrams）。在 *WriteAhead* 系統，就參考使用者選擇的文步，以及游標之前的幾個字，提示接續片語。

*WriteAhead* 能夠提供與排列這些提示，是因為 *WriteAhead* 透過大量的論文原始資料以及少量的人工標示，學習如何辨識 *OWN* 文步的句子，以及這些句子內的常見片語及其頻率。我們將在第三節詳述 *WriteAhead* 所運用的文步分類器的訓練過程。

致謝詞

本研究承蒙科技部補助研究經費，計畫標號 NSC 100-2511-S-007 -005 -MY3