

漢語及物化的大數據研究

A Data Scientific Study of Transitivity in Chinese

蔡維天 Wei-Tien Dylan Tsai
國立清華大學語言學研究所
Linguistics Institute
National Tsing Hua University
wtsai@mx.nthu.edu.tw

楊馨瑜 Ching-Yu Helen Yang
國立清華大學資工系
Department of Computer Science
National Tsing Hua University
chingyu@nlplab.cc

陳映竹 Chen Ying-Zhu
國立清華大學資工系
Department of Computer Science
National Tsing Hua University
jocelyn@nlplab.cc

陳志杰 Jih-Jie Chen
國立清華大學資工系
Department of Computer Science
National Tsing Hua University
jjc@nlplab.cc

張俊盛 Jason S. Chang
國立清華大學資工系
Department of Computer Science
National Tsing Hua University
jason@nlplab.cc

摘要

本文從資料科學的角度來考察漢語中一個新興的現象「及物化」：亦即原本以動前介詞組引介域內論元的謂語(如「為人民服務」)轉化為直接引介動後賓語的謂語(如「服務人民」)。我們認為這個現象其實是一種復古的趨勢，如古漢語的為動式「壯士死知己」即「壯士為知己而死」之意，是一種隱性的輕動詞用法。值得一提的是，這種趨勢仍處於變動之中：它可能逐步消亡，也可能引發爆炸性的發展(如同「語言癌」一般)；這便

是為甚麼我們需要從資料科學的角度切入為其把脈，不但能總結之前的演化歷程，更能預測未來的發展趨勢。此外，我們發現及物化現階段只在幾種文體中有高度的能產性，這也讓我們有機會一窺其使用上的限制及其背後的制約因素(如語用和韻律上的考量等)，並印證於文法搜尋引擎如 Linggle 的數據統計之上。

關鍵詞：及物化，漢語句法，輕動詞，資料科學

一、緒論：何謂及物化？

在歷史的長河中，語言的變遷無異是學者們考察關注的重點之一：其原因即在於漢語有大量的文本典籍和歷史記錄(如災荒、遷徙、戰爭、駐軍及地緣政治等)可供查對，同時還有從唐宋以下各朝各代建立起的韻書撰寫傳統，反映出多層次、多方言的動態音韻體系，豐富了我們探究漢語歷史演化的工具箱，讓專家學者得以一窺中古漢語和上古漢語的風貌，也使重建祖語(proto-language)成為可能。

而現代語言學的發展也正處在一個關鍵期：語言定義「人之所以為人」的生物本能是如何跟種種外緣因素互動(如使用、接觸、混合等)，進而驅動了演化的歷程。這些困難複雜的議題都需要有嚴謹的田野、實驗和大數據研究做為工具，才能抽絲剝繭，澄清問題的本質，並提出明確的解決之道。

根據以上理念，我們將探索重點聚焦在一個近年來興起的語法現象，可姑且稱之為「及物化」(transitivization)。漢語文獻中其實已經有了相當深入的觀察和後續討論：齊滬揚(2000)即指出「漢語述賓結構是一種優勢結構，許多原非述賓結構有向述賓結構靠攏的趨勢這樣使得一些原先不能帶賓語的動詞也逐漸可以帶賓語了，及物動詞的數量呈擴大趨勢。」他舉了以下幾個淺顯易懂的例子，可作為參考：

- (1) a. 他常[為人民]服務。
b. 他常服務[人民]。
- (2) a. 他常[用毛筆]寫字。
b. 他常寫[毛筆]字。

- (3) a. 我們[在北京]相見。
 b. 我們相見[北京]。

如(1a)中動前的由「為」引介的受惠者「人民」在(1b)中變成了動後的直接賓語；又如(2a)中動前的由「用」引介的工具「毛筆」在(2b)中變成了動後的直接賓語；最後(3a)中動前的由「在」引介的地點「北京」在(3b)中變成了動後的直接賓語。

蔡維天(2017)則注意到上述及物化現象在某些文體或地區特別顯著：比如說臺灣媒體下簡短標題時就常採用此種及物句構：表對象關係的例子如(4a-c)，可分別轉寫為(5a-c)，原本動後的直接賓語放到了動前介詞組的位置：

- (4) a. 馬林魚仍甜蜜復仇[紅雀]。
 b. 張艾亞生日告白[許孟哲]。
 c. 法院嗆聲[王寶強]。
 (5) a. 馬林魚仍[對紅雀]甜蜜復仇。
 b. 張艾亞生日[對許孟哲]告白。
 c. 法院[對王寶強]嗆聲。

表示與同(comitative)關係的例子如(6a-c)，可分別轉寫為(7a-c)，亦即動前不及物用法的例子：

- (6) a. 伊朗斷交沙國。
 b. 蔡康永牽手小 S，...
 c. 蕭亞軒分手百億男友。
 d. 秦凱求婚何姿。
 (7) a. 伊朗跟沙國斷交。
 b. 蔡康永跟小 S 牽手，淚灑錄影現場。
 c. 蕭亞軒跟百億男友分手。
 d. 秦凱跟何姿求婚。

最後是表達蒙受關係的(8a,b)，可分別用動前的「給」、「把」轉寫為(9a,b)：

- (8) a. 法陸空罷工添亂歐國杯。
- b. 美國聯邦法官打臉川普。
- (9) a. 法陸空罷工給歐國杯添亂。
- b. 美國聯邦法官把川普打臉。

相對而言，這種新興的及物化在敘事文本中則較少見。更有趣的是，這類用法在口語中反而能產性頗高，而且愈來愈普遍：在動後賓語位置引介對象論元的有(10a-d)，可分別改寫為(11a-d)：

- (10) a. 有沒有人發問[你]？
- b. 他常輕挑[女生]。
- c. 你別刻薄[人家]。
- d. 但我絕對一定要來靠北[她]。
- (11) a. 有沒有人[對你]發問？
- b. 他常[對女生]輕挑。
- c. 你別[對人家]刻薄。
- d. 但我絕對一定要來[對她]靠北。

此外，引介受惠者或原因論元當直接賓語的有(12a-d)，可分別改寫為(13a-d)中的不及物用法(也有人稱為半及物)：

- (12) a. 報紙一直吹牛[他]。
- b. 薇如緊張[我]，...
- c. 他很傷心[這件事]。
- d. 他很高興[這件事]。
- (13) a. 報紙一直[為他]吹牛。
- b. 薇如[為我]緊張，就一直說你怎麼了？

- c. 他[為這件事]很傷心。
- d. 他[為這件事]很高興。

最後，引介蒙受論元當直接賓語的則是(14a,b)，可分別改寫為 (15a-d)：

- (14) a. 他還想要索賠[對方]。
- b. 你怎麼可以放鳥[人家]？
- (15) a. 他還想要[跟對方]索賠。
- b. 你怎麼可以[把人家]放鳥？

二、研究方法

我們更進一步使用語料庫驗證上述漢語及物化演變的趨勢（參閱蔡維天 2017），考量及物化可能因年代與語體而有不同的變化，本研究選用不同年代的新聞資料作為研究標的，並區分標題與內文以對比語體。我們分別將這三份資料建構成以依存關係（Universal dependency）（Nivre et al., 2015, 2016）為框架的語料庫，並抽取所需資料做統計，方法步驟如下：

1. 重新斷句、為標題補上標點符號
2. 使用中研院詞庫小組 CKIP CoreNLP 系統斷詞與做詞性剖析¹
3. 使用 Chinese Stanford Parser 剖析依存關係²

本節將介紹使用的語料庫、進行的預處理與運用的剖析器（Parser）。

¹ <http://ckip.iis.sinica.edu.tw/service/corenlp/>

² https://github.com/UniversalDependencies/UD_Chinese-GSD

(一)資料介紹

我們將 LDC Tagged Chinese Gigaword³ 的標題與內文分開儲存為兩份資料庫，涵蓋的新聞來源有：中央通訊社、新華通訊社與聯合早報，資料年份為 1991 年到 2004 年，新聞約 181 萬 6 千多篇。第三份資料為 2004 年到 2017 年間的新聞內文，主要新聞來源為：聯合報、聯合晚報、經濟日報，約有 177 萬多篇新聞。

1. 新聞資料_標題：1991 年至 2004 年，176 萬 8 千多則
2. 新聞資料_內文一：1991 年至 2004 年，176 萬 8 千多則
3. 新聞資料_內文二：2004 年至 2017 年，177 萬多則

(二) 資料處理與剖析

為提升 Chinese Stanford Parser 的依存關係剖析效果，我們先進行以下預處理：將標題（資料一）結尾補上句點符號，解決 Stanford Parser 傾向將最後一個詞剖析成標點符號的問題（如表一）。內文部分（資料二、資料三），以句點、問號、分號、驚嘆號斷句，以避免句子不完整而造成依存關係剖析錯誤（如表二）。將處理好後的資料，以中研院詞庫小組 CKIP CoreNLP 系統斷詞與做詞性剖析。最後，我們使用 Chinese Stanford Parser (UD_Chinese-GSD) 做依存關係剖析。處理後，語料庫大小，條列如下：

1. 新聞資料_標題(1991 年至 2004 年)：176 萬 8 千多則，181 萬 6 千多句，1650 萬 2 千多詞。
2. 新聞資料_內文一(1991 年至 2004 年)：176 萬 8 千多則，1699 萬 8 千多句，4 億 6293 萬多個詞。
3. 新聞資料_內文二(2004 年至 2017 年)：177 萬多則，1665 萬 5 千多句，5 億 2 千多萬詞。

³ <https://catalog.ldc.upenn.edu/LDC2007T03>

表一、添加標點符號

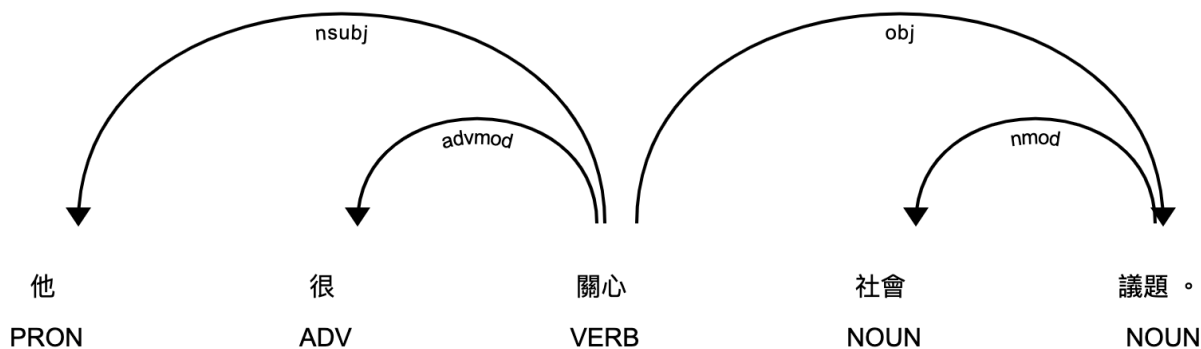
| 處理前 | 處理後 |
|---|--|
| 環保署推出環保建設六年計畫 法國不在意日本對其核試的杯葛 豐原醫院貼心手推車服務解決住出院行李煩惱 | 環保署推出環保建設六年計畫。 法國不在意日本對其核試的杯葛。 豐原醫院貼心手推車服務解決住出院行李煩惱。 |

表二、將 Tagged Chinese Gigaword 重新斷句

| 處理前 | 處理後 |
|--|--|
| 戈巴契夫曾表示， 他相信國家安全委員會前首腦克留契柯夫是政變主謀。 愛滋病患者對他社交圈內別人對他的看法， 與對待他的方式， 非常在意。 | 戈巴契夫曾表示，他相信國家安全委員會前首腦克留契柯夫是政變主謀。 愛滋病患者對他社交圈內別人對他的看法，與對待他的方式，非常在意。 |

三、 抽取依存關係的規則

近年來，很多自然處理的研究使用依存句法分析樹（Universal Dependencies Treebank）來訓練句法剖析器。Universal Dependencies (Nivre et al., 2015, 2016) 是一個跨語言語法標註的框架，涵蓋的語言超過 70 種，中文也包含其中，相關研究學者訂定了跨語言標註分析的準則 (Nivre et al., 2015, 2016)，超過 70 個語言。在依存句法分析的架構下，直接標註詞與詞之間的依存關係，結構較為扁平，如圖一所示。



圖一、依存句法分析

一個句子只有一個獨立成分，不必依附於（depend on）其他詞，圖一的「關心」即為此例，標示的依存關係為“root”，其餘詞彙則須依附在其他的詞彙上，以箭頭表示，如「他」、「很」與「議題」皆依附於「關心」，依存關係分別為主詞（nsubj）、副詞修飾語（advmod）以及受詞（obj），而「社會」則依附於「議題」，依存關係為名詞修飾語（nmod）。依存句法樹直接標註詞與詞的關係，而非詞與詞組之間的關係，如圖一中依附在「關心」的詞為「議題」，而非「社會 議題」，如此一來，可直接得知與考察目標詞彙最直接相關的詞有哪些，以及依存關係為何，方便語法的抽取與搭配詞的計算。

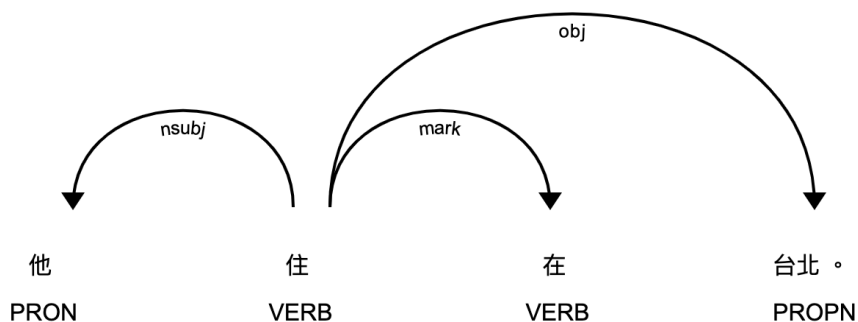
本研究目標為考察動詞及物與不及物用法在不同語料庫的對比，需抽取以下兩個結構：(1) 動詞 受格論元; (2) 介系詞 旁格論元 動詞。我們依照表三、表四的規則抽取依存關係，組成標的結構，留下例句並統計次數。

表三、「動詞 受格論元」抽取規則

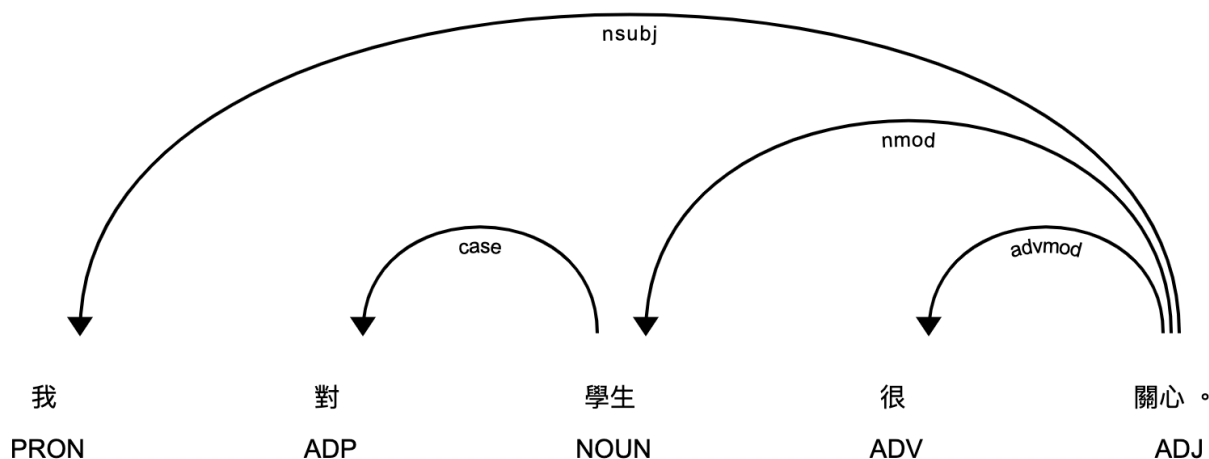
| | |
|------|---|
| 抽取規則 | 1. 目標動詞為 root，受格論元依存於 root，兩者依存關係為 obj。 2. root 與其他詞的依存關係不包含依存關係 mark。 |
| 抽取範例 | (1) 他 很 關心 社會 議題。 (符合抽取規則，參見圖一) 抽取詞彙：關心(root) 議題(obj) |
| 濾除範例 | (2) 他 住 在 台北。(不符合抽取規則 2，參見圖二) 住(root) 在(mark) 台北(obj) |

表四、「介系詞 旁格論元 動詞」抽取規則

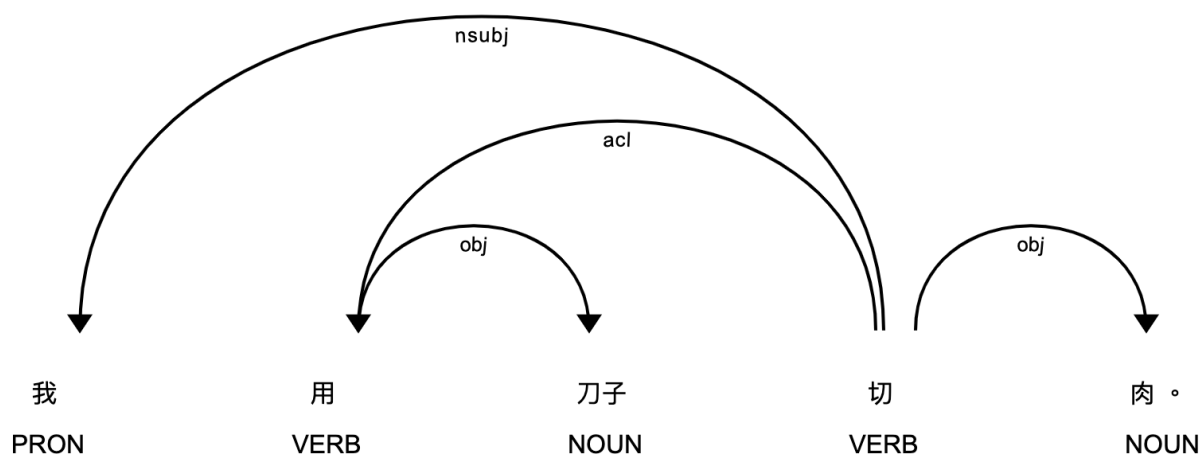
| | |
|-------|--|
| 抽取規則 | <ol style="list-style-type: none"> 1. 介系詞 <ol style="list-style-type: none"> A. 句子包含下欄的介系詞，且中研院 Core NLP 詞性分析為介系詞 (P)。 B. 依附於動詞(root)或是動詞的依附詞(見抽取範例(2))。 C. 依附關係為 case、mark 或是 acl。 2. 旁格論元 <ol style="list-style-type: none"> A. 依附於動詞(root)或是介系詞 B. 依附關係為 obj、obl 或是 nmod |
| 介系詞列表 | 在, 以, 與, 對, 從, 爲, 由, 到, 向, 於, 和, 用, 跟, 給, 至, 經, 往, 靠, 替, 藉, 朝, 按, 憑, 沿, 把, 對於 |
| 抽取範例 | <p>(1) 他 住 在 台北。(符合抽取規則, 參見圖二) 抽取詞彙: 住(root) 在(mark) 台北(obj)</p> <p>(2) 我 對 學生 很 關心。(符合抽取規則, 參見圖三) 抽取詞彙: 對(case) 學生(nmod) 關心(root)</p> <p>(3) 我 用 刀子 切 肉。(符合抽取規則, 參見圖四) 抽取詞彙: 對(case) 學生(nmod) 關心(root)</p> |



圖二、依存句法分析



圖三、依存句法分析



圖四、依存句法分析

四、統計結果與分析

我們使用大數據考察漢語及物化的趨勢，聚焦在以下兩個項目：

1. 標題比其他書面文體更傾向使用及物句構。
2. 年代較近的資料比較久遠的資料更傾向使用及物句構。

我們先呈現及物化在標題與內文的差異，再探究年代對及物化的影響。我們以蔡維天(2017)一文中提及的動詞作為起點，來觀察漢語及物化的趨勢：在意、關心、服務、相見、擔心、復仇、告白、嗆聲、斷交、牽手、分手、求婚、添亂、打臉、背信、偷情、發問、輕挑、刻薄、靠北、吹牛、緊張、傷心、高興、索賠、放鳥、混、劈腿、求救、前進、褒獎、怪罪、加害、看好、眼紅、頭痛。

(一) 標題與內文的差異與分析

資料一與資料二皆為 1991 年至 2004 年的新聞資料，前者為標題，後者為內文，可用來探究趨勢一指出的及物化情形。我們根據第三節的抽取規則，統計標的動詞的及物結構與不及物結構的次數，結果請見表五。

表五、資料一與資料二及物性統計結果

| 資料一（標題 1991-2004） | | | 資料二（內文 1991-2004） | | | 比較 | | |
|-------------------|-----|-----|-------------------|----|-----|-----|------|-------|
| | 及物 | 不及物 | 及物性 | | 及物 | 不及物 | 及物性 | 1 > 2 |
| 在意 | 10 | 0 | 1 | 在意 | 29 | 37 | 0.44 | ✓ |
| 復仇 | 1 | 0 | 1 | 復仇 | 1 | 1 | 0.5 | ✓ |
| 怪罪 | 7 | 0 | 1 | 怪罪 | 22 | 17 | 0.56 | ✓ |
| 加害 | 2 | 0 | 1 | 加害 | 2 | 4 | 0.33 | ✓ |
| 關心 | 489 | 19 | 0.96 | 關心 | 481 | 230 | 0.68 | ✓ |
| 擔心 | 176 | 9 | 0.95 | 擔心 | 417 | 185 | 0.69 | ✓ |
| 看好 | 383 | 22 | 0.95 | 看好 | 325 | 310 | 0.51 | ✓ |
| 前進 | 56 | 12 | 0.82 | 前進 | 101 | 182 | 0.36 | ✓ |
| 服務 | 168 | 41 | 0.8 | 服務 | 555 | 563 | 0.5 | ✓ |
| 混 | 2 | 1 | 0.67 | 混 | 15 | 17 | 0.47 | ✓ |
| 告白 | 5 | 3 | 0.63 | 告白 | 7 | 6 | 0.54 | ✓ |
| 斷交 | 14 | 11 | 0.56 | 斷交 | 7 | 112 | 0.06 | ✓ |
| 分手 | 3 | 4 | 0.43 | 分手 | 2 | 35 | 0.05 | ✓ |
| 相見 | 3 | 5 | 0.38 | 相見 | 3 | 29 | 0.09 | ✓ |
| 求婚 | 5 | 10 | 0.33 | 求婚 | 4 | 19 | 0.17 | ✓ |
| 索賠 | 16 | 45 | 0.26 | 索賠 | 31 | 64 | 0.33 | ✗ |
| 緊張 | 3 | 15 | 0.17 | 緊張 | 2 | 81 | 0.02 | ✓ |
| 求救 | 3 | 16 | 0.16 | 求救 | 11 | 129 | 0.08 | ✓ |
| 嗆聲 | 0 | 2 | 0 | 嗆聲 | 4 | 10 | 0.29 | ✗ |
| 發問 | 0 | 2 | 0 | 發問 | 8 | 31 | 0.21 | ✗ |
| 傷心 | 0 | 1 | 0 | 傷心 | 1 | 5 | 0.17 | ✗ |
| 高興 | 0 | 26 | 0 | 高興 | 7 | 211 | 0.03 | ✗ |
| 頭痛 | 0 | 2 | 0 | 頭痛 | 0 | 18 | 0 | -- |

我們先排除資料一或資料二中及物結構「動詞 受格論元」與不及物結構「介系詞 旁格論元 動詞」筆數為零的動詞 13 筆（如：「添亂」、「打臉」），第二欄與第六欄為及物結構的筆數，第三欄與第七欄為不及物結構的筆數，第四欄與第八欄為標的動詞的及物性，計算方式為：及物結構除以總次數（不及物結構的次數加上及物結構的次數），分數越高，及物性越強，最後一欄比較標的動詞在資料一的及物性是否比在資料二高。對比 23 個動詞中，有 17 個動詞在標題資料中的及物性比在內文資料中來得高（如：「在意」、「關心」），有 6 個動詞在內文資料的及物性較高（如：「發問」、「嗆聲」）。

(二) 及物性於跨年代的演變

趨勢二的及物化情形可透過對比資料二與資料三來一虧端倪，資料二與資料三皆為新聞內文，前者整年代較久(1991 年至 2004 年)，後者年代較新。及物化在不同年代對比結果請見表六。

表六、資料二與資料三及物性統計結果

| 資料三 (內文 2004-2017) | | | 資料二 (內文 1991-2004) | | | 比較 | | |
|--------------------|------|-----|--------------------|----|-----|-----|------|-------|
| | 及物 | 不及物 | 及物性 | | 及物 | 不及物 | 及物性 | 3 > 2 |
| 擔心 | 923 | 265 | 0.78 | 擔心 | 417 | 185 | 0.69 | ✓ |
| 關心 | 919 | 287 | 0.76 | 關心 | 481 | 230 | 0.68 | ✓ |
| 吹牛 | 5 | 1 | 0.83 | 吹牛 | 3 | 0 | 1 | ✗ |
| 怪罪 | 28 | 11 | 0.72 | 怪罪 | 22 | 17 | 0.56 | ✓ |
| 在意 | 188 | 84 | 0.69 | 在意 | 29 | 37 | 0.44 | ✓ |
| 看好 | 674 | 309 | 0.69 | 看好 | 325 | 310 | 0.51 | ✓ |
| 服務 | 1201 | 743 | 0.62 | 服務 | 555 | 563 | 0.5 | ✓ |
| 加害 | 3 | 2 | 0.6 | 加害 | 2 | 4 | 0.33 | ✓ |
| 背信 | 3 | 2 | 0.6 | 背信 | 1 | 0 | 1 | ✗ |
| 索賠 | 64 | 60 | 0.52 | 索賠 | 31 | 64 | 0.33 | ✓ |
| 混 | 115 | 109 | 0.51 | 混 | 15 | 17 | 0.47 | ✓ |
| 傷心 | 12 | 12 | 0.5 | 傷心 | 1 | 5 | 0.17 | ✓ |
| 前進 | 339 | 426 | 0.44 | 前進 | 101 | 182 | 0.36 | ✓ |
| 偷情 | 3 | 5 | 0.38 | 偷情 | 1 | 0 | 1 | ✗ |
| 刻薄 | 1 | 2 | 0.33 | 刻薄 | 0 | 1 | 0 | ✗ |
| 復仇 | 4 | 10 | 0.29 | 復仇 | 1 | 1 | 0.5 | ✗ |
| 發問 | 8 | 31 | 0.21 | 發問 | 8 | 31 | 0.21 | -- |
| 緊張 | 19 | 108 | 0.15 | 緊張 | 2 | 81 | 0.02 | ✓ |
| 高興 | 25 | 147 | 0.15 | 高興 | 7 | 211 | 0.03 | ✓ |
| 頭痛 | 5 | 29 | 0.15 | 頭痛 | 0 | 18 | 0 | ✓ |
| 牽手 | 2 | 13 | 0.13 | 牽手 | 1 | 3 | 0.25 | ✗ |
| 求婚 | 27 | 198 | 0.12 | 求婚 | 4 | 19 | 0.17 | ✗ |
| 分手 | 26 | 204 | 0.11 | 分手 | 2 | 35 | 0.05 | ✓ |
| 告白 | 8 | 81 | 0.09 | 告白 | 7 | 6 | 0.54 | ✗ |
| 斷交 | 3 | 29 | 0.09 | 斷交 | 7 | 112 | 0.06 | ✓ |
| 相見 | 10 | 97 | 0.09 | 相見 | 3 | 29 | 0.09 | -- |
| 嗆聲 | 29 | 279 | 0.09 | 嗆聲 | 4 | 10 | 0.29 | ✗ |

| | | | | | | | | |
|----|----|-----|------|----|----|-----|------|---|
| 求救 | 40 | 647 | 0.06 | 求救 | 11 | 129 | 0.08 | * |
|----|----|-----|------|----|----|-----|------|---|

資料二或資料三中及物結構「動詞 受格論元」與不及物結構「介系詞 旁格論元 動詞」筆數為零的動詞有 8 筆（如：「添亂」、「打臉」），暫不列入考量，表六表格安排與及物性按表五方式計算，分數越高，及物性越強，對比結果為，28 個動詞中，有 16 個動詞的及物性有隨年代而遞增的趨勢（如：「擔心」、「關心」），有 2 個動詞的及物性（如：「發問」、「相見」）未應年代而產生變化，而有 10 個動詞的及物性不增反減（如：「牽手」、「刻薄」）。

（三）結果分析

資料一與資料二的對比反應出標題比內文文體更傾向使用及物句構（趨勢一），23 個動詞中有 16 個動詞支持這個趨勢。資料二與資料三的差異呈現，在年代較近的語料有較多及物結構，28 個動詞中有 16 個動詞有及物化的趨勢，些微反應出年代對及物化的影響（趨勢一）。從數據上看來，年代對及物化的影響沒有文體差異顯著，可能句法上的新興演變會優先反應在口語上，需要一段時間才會反應在書面語上，因此動詞有隨時間演進有及物化的趨勢，但仍不顯著。而媒體下標題需簡短、吸引讀者注意力，不及物動詞使用及物化結構，可達到媒體因此可因語用因素而違反句法規則。

動詞有 16 組未支持及物化趨勢，其中一個可能的原因是動詞的詞頻較低（如：「偷情」、「刻薄」、「復仇」、「嗆聲」等），使用度較低的動詞，對語法變化會較保守，未來研究可先以詞頻較高的動詞為研究標的，以排除詞頻的干擾因素。另一個可能為，資料的差異性不夠高，資料一、資料二的文體不同，一為標題，一為內文，但仍同屬新聞文體，未來關於文體差異的研究，可對比書面語與口語資料，反應出的及物化趨勢會更加明顯。關於年代的對比，統計方式可改為統計所有標的動詞的逐年及物化變化，或許更適合考察及物化歷年演變的趨勢。

五、結論與未來展望

前述觀察與研究顯示語法研究應該和資料科學充份結合，取得實證面上的數據支持，並在語言教學和人工智能等面向上發揮其潛在應用價值。事實上，若從中英文平行語料庫做初步觀察，及物化的用法很難找到跨語言的對應，但卻凸顯了中英文在語言類型上的差異：以(16)中的對譯為例，首先是語序上英語的介詞結構在動後而非動前，這與其「中心在前」(head-initial)的特色一致。其次是英語名動互轉的機制非常發達（亦即 *quality* ⇒ *qualify*），而漢語只能靠引介輕動詞(light verb)如「取得」來翻譯。

(16) Andrew **qualified as** a teacher in 1995.

安德魯 於 1995 年 取得 教師 資格 。

這點在(17)、(18)的對譯中也得到印證：相當於英語中一個簡單的介詞 *for*，中文翻譯卻需要添加「喝」、「找」等動詞，才能形成動後的及物用法：其實此處 *for* 的性質非常接近漢語中語意泛化的輕動詞；一旦翻成中文便需要照顧到賓語的選擇限制，必須用更明確的動詞才行：

(17) It was freezing outside and Marcia **longed for** a hot drink

外面 很 冷 ， 瑪西雅 很 想 喝 一 杯 熱 飲 。

(18) She **groped for** her glasses on the bedside table

她 在 床 頭 櫃 上 摸 索 著 找 眼 鏡 。

最後是 *qualify* 其實是一種使動動詞，相當於 *cause someone to have the quality/certificate of a teacher*。由於漢語沒有類似用法，也不太能說「取得資格成教師」，因此「教師」便從補語轉化為名前定語，相當英語中名後的 *of a teacher* 或是動後的 *as a teacher*。這些現象若有足夠多的語料可供大數據研究，相信一定能在華語教學、文法寫作、機器翻譯等面向上開啟更具突破性的發展，讓文法理論和應用做更完美的結合。

此外，我們可以開始發展「客製化」的中文文法搜尋引擎，用大數據理念來研究各種不同文體的語法差異和通則，進而應用到修辭學、文體學、高級寫作、辭典編纂以至符號學、社會語言學等做具有前瞻性的跨界研究。一旦我們有了夠大的資料庫，便可將觸角延伸至認知結構、本體論(ontology)、歷史語法、語言類型學及普遍語法的研究，讓不同領域的學者既能各取所需又可相互支援。

參考文獻

- [1] 蔡維天. 2017. 〈及物化、施用結構與輕動詞分析〉，〈《現代中國語研究》〉，第19期，1-13頁。
- [2] 齊滄揚. 2000. 《現代漢語短語》，張斌主編，上海：華東師範大學出版社。
- [3] Chang, P.-C., Tseng, H., Jurafsky, D., and Manning, C. D. Discriminative reordering with Chinese grammatical relations features. In *Third Workshop on Syntax and Structure in Statistical Translation*, pages 51–59, 2009.
- [4] de Marneffe, M.-C., Connor, M., Silveira, N., Bowman, S. R., Dozat, T., and Manning, C. D. More constructions, more genres: Extending Stanford dependencies. In *DepLing*, 2013.
- [5] Elming, J., Johannsen, A., Klerke, S., Lapponi, E., Martinez, H., and Søgaard, A. Down-stream effects of tree-to-dependency conversions. In *NAACL HLT*, 2013.
- [6] De Marneffe, Marie-Catherine, and Christopher D. Manning. *Stanford typed dependencies manual*. Technical report, Stanford University, 2008.
- [7] de Marneffe, M.-C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., and Manning, C. D. Universal Stanford Dependencies: A cross-linguistic typology. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)*, 2014.
- [8] McDonald, R., Nivre, J., Quirnbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K., Petrov, S., Zhang, H., Tačkstrõm, O., Bedini, C., Bertomeu Castelló, N., and Lee, J. Universal dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2013.
- [9] Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic̃, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., and Zeman, D. Universal dependencies v1: A multilingual tree- bank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC)*, 2016.