

## 新住民相關親權酌定裁判書的文字探勘： 對「平等」問題的法實證研究嘗試\*

邵軒磊\*\*、黃詩淳\*\*\*

<摘要>

本研究關心新住民與我國籍配偶離婚後，進入法院爭取未成年子女親權時，是否獲得平等對待之問題。本文使用資訊科學技術，驗證「法院是否會因父母的國籍而在裁判書中使用不同的詞彙」，分析 116 件親權酌定事件之裁判書。其中有 31 件之一方當事人為外籍、外裔人士（簡稱外籍裁判），85 件為雙方當事人皆為我國籍（簡稱雙臺籍裁判）。首先，將裁判書中的詞彙在分詞後，轉化為 TF-IDF 的文本詞彙矩陣，再透過主成分分析投影在二維平面上，將每則裁判書的詞彙特徵（所在位置）可視化。雙臺籍裁判與外籍裁判並未分為不同的兩群，而是位置重合，意味著法院在兩種當事人的裁判並未使用不同的詞彙，因此無法直接主張司法體系對新住民不平等。其次，再使用機器學習中的梯度提升法，發現外籍裁判中，「是否出境」以及「居住狀況」相關詞彙特徵，較雙臺籍裁判為高，表示法官在外籍裁判中較

---

\* 感謝二位審查人對本文提供諸多寶貴意見，使作者受益甚多。感謝法律扶助基金會提供相關資料，以及工作人員的詳細說明。本研究部分成果完成於筆者與臺大法律資料分析研究室成員一起參與 D4SG (Data for Social Good) 資料英雄計畫 2019 冬季班。本研究為科技部「從知識系譜到知識地圖：大數據與機器學習下的『中國研究』(MOST-107-2410-H-003 -058 -MY3)」與教育部 (NTU-109L900203) 與科技部 (MOST 109-2634-F-002-045) 高教深耕計畫之研究成果。

\*\* 國立臺灣師範大學東亞學系副教授。E-mail: hlshao@ntnu.edu.tw。

\*\*\* 國立臺灣大學法律學院教授、國立臺灣大學計量理論與應用研究中心兼任特約研究員。E-mail: schhuang@ntu.edu.tw。

• 投稿日：07/02/2020；接受刊登日：09/25/2020。

• 責任校對：何思奕、陳莉蕻、簡凱葳。

• DOI:10.6199/NTULJ.202011/SP\_49.0001

關心程序參與以及居留問題。藉由文字探勘與機器學習來分析裁判書文本，除了檢驗平等問題外，還可將抽象的法律概念（子女最佳利益）具象化為裁判書中提及之事實（出境、租屋），供實務參考，結合了資訊與法律兩個領域，呼應本刊「跨域、科技與法律」的主題。

關鍵詞：法律資料分析、自然語言處理、詞頻—倒數文件頻率法、平等、新住民、法實證研究、主成分分析

## ◆目次◆

- 壹、研究背景
- 貳、文獻回顧
- 參、研究設計
  - 一、資料來源
  - 二、研究方法
  - 三、研究步驟
- 肆、研究發現與討論
  - 一、外籍裁判與雙臺籍裁判之詞彙並無不同
  - 二、外籍裁判詞彙之特徵
- 伍、結論
  - 一、研究結論
  - 二、延伸討論與研究方法之意義

## 壹、研究背景

臺灣社會日漸開放，國人與外國人結婚的數量也不少。2019 年國人結婚計 13.2 萬對，跨國婚姻則為 2.1 萬對，換言之，該年的婚姻中有 13.73% 為國際結婚<sup>1</sup>。此 2.1 萬對的跨國婚姻，包含夫妻中一方為大陸、港澳地區人民以及外國籍人民者，其中一方為大陸地區人民者為 6,696 對、港澳人民

---

<sup>1</sup> 聯合新聞網（02/28/2020），〈跨國婚姻每年約2萬對 大陸、越南最多〉，<https://udn.com/news/story/7314/4376997>（最後瀏覽日：06/21/2020）。更細部的數據參見內政部戶政司，人口統計資料，各月人口資料，縣市結婚人數按原屬國籍（按登記），內政部戶政司網站，<https://www.ris.gov.tw/app/portal/346>（最後瀏覽日：06/21/2020）。另，2019年5月24日起同性伴侶可根據《司法院釋字第748號解釋施行法》登記結婚，至2019年底，亦有大陸籍（共2人）、港澳籍（共10人）及外國籍（共147人）人士與我國人民締結同性婚姻，惟因施行時間尚短，離婚後未成年子女的親權酌定問題也尚未大量進入法院，故本研究暫不探討。

者 1,621 對、外國籍者 12,740 對，亦即目前一方為外國籍之夫妻對數超過了一方為大陸或港澳籍之對數<sup>2</sup>。

在說明本研究的內涵之前，先釐清若干名詞用法。首先，「外籍配偶」嚴格言之，是指「具外國國籍者」與我國人民結婚後之身分。而「新住民」則是因為某些原因例如婚姻、工作或其他理由而定居在臺灣，甚至取得國籍而加入臺灣社會者<sup>3</sup>。這兩者可能重合，但細微之處還是有差異，例如非因婚姻而移民至我國者，屬於「新住民」但並非「外籍配偶」。但本文不擬深入討論，而遵循既有研究之建議<sup>4</sup>，混用二詞。此外，雖大陸籍配偶及港澳籍配偶並非嚴格意義下的「外國籍配偶」，但其與我國人民結婚後，在離婚時面臨的子女親權酌定與居留問題有類似之處，故本文亦將其納入研究範圍，統稱為「外籍配偶」或「新住民」。最後，本文所謂的「某國籍配偶」，包含現在國籍為該國、以及目前已歸化取得我國國籍但過去為其他國籍出身者；亦即「外籍配偶」或「新住民」一詞，包含了「外籍」與「外裔」的婚姻移民。

上述 2019 年的數據是該年結婚的對數。至於過去至今與我國人民結婚並居住於我國的外籍配偶人數，依照內政部之統計數據<sup>5</sup>（參照下【表一】、【圖一】），大陸籍配偶在臺灣居留的人數最多，為 349,132 人，占全體之 63%，且大部分是女性配偶（349,132 人之中的 330,157 人為女性）。其次為越南籍配偶人數 108,997 人，占全體之 20%。

---

<sup>2</sup> 在 2015 年以前則相反，向來是一方為大陸或港澳籍之對數多於一方為外國籍之對數。

<sup>3</sup> 內政部移民署定義為：「對於從國外來到臺灣結婚、移民而定居的人士稱為新住民。」參見內政部移民署網站，<https://www.immigration.gov.tw/5385/7344/70395/143257/>（最後瀏覽日：06/21/2020）。

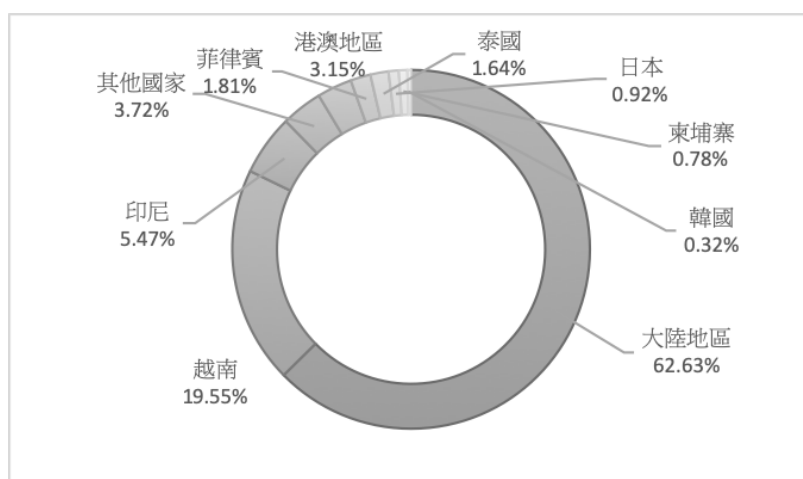
<sup>4</sup> 例如翁毓秀（2006），〈為新住民建構社會包容的家庭政策〉，《社區發展季刊》，114 期，頁 61，指出，為了避免對「外籍配偶與大陸配偶」的有色眼光與歧視態度，最近逐漸改稱為「新住民」，故基本上認為兩者指涉同一概念。

<sup>5</sup> 內政部移民署，業務統計，統計資料，外籍配偶（含大陸、港澳地區人民），外籍配偶人數與大陸（含港澳）配偶人數按證件分，內政部移民署網站，<https://www.immigration.gov.tw/5382/5385/7344/7350/8887/>（最後瀏覽日：06/21/2020）。此數據包含外國裔與外國籍配偶。

【表一】1987 年至 2019 年底外裔、外籍配偶人數（依國籍分類）

國籍	人數
大陸地區	349,132
越南	108,997
印尼	30,483
其他國家	20,749
港澳地區	17,582
菲律賓	10,102
泰國	9,126
日本	5,154
柬埔寨	4,339
韓國	1,786
總計	557,450

※ 資料來源：內政部，參見前揭註 5。作者自製。

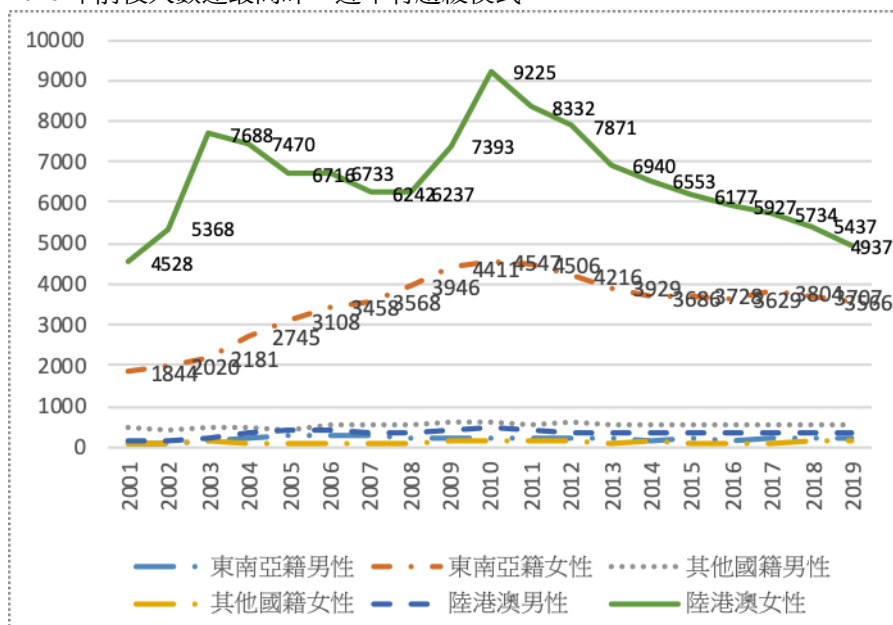


【圖一】1987 年至 2019 年底外裔、外籍配偶比例（依國籍分類）

※ 資料來源：內政部，參見前揭註 5。作者自製。

其次，觀察子女數量，2018 年母或父為新住民之嬰兒為 1.38 萬人，比重由 2003 年之 13.37% 下降至 2018 年之 7.62%<sup>6</sup>。惟國人生育率持續降低，近 10 年來全國國中、小學生人數自 2008 學年之 2,629,415 人，逐年遞減至 2018 學年之 1,782,964 人；其中新住民子女學生數卻反自 129,917 人增至 166,801 人，所占比率由 4.94% 上升至 9.36%<sup>7</sup>。這意味著因婚姻而移民至我國的新住民，婚後有相當大機率育有子女。

另一方面，外籍配偶與我國人民的離婚對數也引人注目。如下【圖二】所示，陸港澳女性配偶離婚人數最多，其次為東南亞籍女性配偶；兩者均在 2010 年前後人數達最高峰，近年有趨緩模式。



【圖二】歷年外籍、外裔人士與我國人民之離婚人數

※ 資料來源：內政部<sup>8</sup>；作者自製。

<sup>6</sup> 教育部統計處(2019)，《新住民子女就讀國中小人數分布概況統計：107學年度》，頁1，[http://stats.moe.gov.tw/files/analysis/son\\_of\\_foreign\\_107.pdf](http://stats.moe.gov.tw/files/analysis/son_of_foreign_107.pdf)（最後瀏覽日：06/23/2020）。

<sup>7</sup> 教育部統計處，前揭註6，頁2。

<sup>8</sup> 內政部戶政司，人口統計資料，各月人口資料，縣市離婚人數按原屬國籍（按登

上【圖二】之數據包含外籍以及外裔（已歸化我國國籍）之配偶，然由此數字無法得知離婚率之高低，並與雙方皆為本國籍之夫妻之離婚率相較。唯一能找到的資料是內政部關於「已歸化」之外籍配偶的相關統計調查（不包含大陸、港澳配偶），根據該報告，自 2008 年至 2016 年間，外籍配偶歸化之人數總計 55,855 人，其中有 13,769 人與我國配偶離婚，離婚人數占外籍配偶歸化人數之 24.65%，顯較國人粗估離婚率（離婚人數占有偶人數與離婚人數加總之比率）之 14% 為高<sup>9</sup>。且若再對照此份內政部報告中的每年歸化後離婚人數，與每年離婚之外籍、外裔配偶人數（即【圖二】數據除去陸港澳籍配偶人數），以 2010 年為例，該年共有 5,539 位（不包含陸港澳籍之）外籍配偶與我國人民離婚，但其中僅 2,145 人已歸化，這代表有 3,394 人並未來得及歸化即離婚，很可能面臨「若未取得子女親權則必須被迫離開臺灣」的窘境。

不同國籍的婚姻移民，在臺居留之法律依據與效果有所不同。尤其，本文欲討論「離婚後子女親權酌定」的議題，在大陸、港澳或外國籍配偶在與我國籍配偶離婚後，可否因為取得未成年子女之親權而得繼續居留臺灣，相關規定的修正歷程有相當差異。亦即，《香港澳門居民進入臺灣地區及居留定居許可辦法》（下稱「港澳辦法」）在 1999 年 5 月 31 日修正第 28 條時，即導入「取得親權得繼續居留」之規定（第 28 條第 1 項第 3 款）；《大陸地區人民在臺灣地區依親居留長期居留或定居許可辦法》（下稱「大陸辦法」）則是在 2004 年 3 月 1 日導入此規定（第 14 條第 1 項第 4 款）；而適用於外國籍配偶的《入出國及移民法》，承認「取得親權得繼續居留」的第 31 條，係在 2007 年 12 月 26 日始為修正。且三者規定內容亦有不同，以現行有效之條文為例，港澳配偶的部分，係「未辦妥定居手續前與其配偶離婚，已生產有未成年子女且離婚後任該子女權利義務之行使或負擔」（港澳辦法第

---

記），內政部戶政司網站，<https://www.ris.gov.tw/app/portal/346>（最後瀏覽日：06/21/2020）。

<sup>9</sup> 內政部統計處（2017），《外籍配偶歸化後婚姻狀況之研究》，頁 6-7，[https://www.moi.gov.tw/files/site\\_node\\_file/7471/外籍配偶歸化後婚姻狀況之研究.pdf](https://www.moi.gov.tw/files/site_node_file/7471/外籍配偶歸化後婚姻狀況之研究.pdf)（最後瀏覽日：06/23/2020）。

28 條第 1 項第 4 款但書<sup>10</sup>)；大陸配偶的部分，係「於離婚後對在臺灣地區設有戶籍之未成年親生子女，有扶養事實、行使負擔權利義務或會面交往」（大陸辦法第 14 條第 2 項<sup>11</sup>第 1 款第 3 目）；外國籍配偶的部分是「外國人於離婚後取得在臺灣地區已設有戶籍未成年親生子女監護權」（入出國及移民法第 31 條第 4 項<sup>12</sup>第 3 款）。比較上述條文之規定，大陸配偶不必然

---

<sup>10</sup> 香港澳門居民進入臺灣地區及居留定居許可辦法第 28 條第 1 項規定：「經許可在臺灣地區居留之香港或澳門居民，有下列情形之一者，得撤銷或廢止其依第二十三條規定所為之居留許可，並註銷其臺灣地區居留證或臺灣地區居留入出境證」，其中第 4 款規定：「申請居留之原因消滅。但直系血親、配偶死亡或以其配偶在臺灣地區設有戶籍申請在臺灣地區居留，未辦妥定居手續前與其配偶離婚，已生產有未成年子女且離婚後任該子女權利義務之行使或負擔，不在此限。」  
<sup>11</sup> 大陸地區人民在臺灣地區依親居留長期居留或定居許可辦法第 14 條第 2 項規定：「大陸地區人民申請依親居留，有下列情形之一者，不予許可；已許可者，撤銷或廢止其許可：

一、申請依親居留原因消失。但已許可依親居留，有下列情形之一者，不撤銷或廢止其許可：

(一) 依親對象死亡且未再婚。

(二) 於離婚後三十日內與原依親對象再婚。

(三) 於離婚後對在臺灣地區設有戶籍之未成年親生子女，有扶養事實、行使負擔權利義務或會面交往。

(四) 因遭受家庭暴力經法院判決離婚，且有在臺灣地區設有戶籍之未成年親生子女。

(五) 因依親居留許可被廢止而遭強制出境，對在臺灣地區設有戶籍之未成年親生子女造成重大且難以回復損害之虞。」（第 2 款之後的內容與本文無關，故省略之）。

<sup>12</sup> 入出國及移民法第 31 條第 4 項規定：「入出國及移民署對於外國人於居留期間內，居留原因消失者，廢止其居留許可，並註銷其外僑居留證。但有下列各款情形之一者，得准予繼續居留：

一、因依親對象死亡。

二、外國人為臺灣地區設有戶籍國民之配偶，其本人遭受配偶身體或精神虐待，經法院核發保護令。

三、外國人於離婚後取得在臺灣地區已設有戶籍未成年親生子女監護權。

四、因遭受家庭暴力經法院判決離婚，且有在臺灣地區設有戶籍之未成年親生子女。

五、因居留許可被廢止而遭強制出國，對在臺灣地區已設有戶籍未成年親生子女造成重大且難以回復損害之虞。



要取得親權才能居留臺灣，即便僅是扶養或會面交往亦可，然港澳、外國籍配偶卻必須獲取親權，三者顯然寬嚴有別。此種區別是否妥當，並非本文討論之範圍，故暫且擱置。

綜上，外籍配偶婚後有相當人數育有子女，又有相當人數在尚未歸化時即與我國籍配偶離婚，離婚時多半子女尚未成年，必須決定子女親權如何行使。且如上所述，尚未歸化取得我國國籍之外籍配偶，是否取得未成年子女之親權攸關能否繼續在臺居留。故從外籍配偶的權益保障角度言，有必要探討與分析其在離婚後對子女行使親權之情形。

民法第 1055 條第 1 項規定：「夫妻離婚者，對於未成年子女權利義務之行使或負擔，依協議由一方或雙方共同任之。未為協議或協議不成者，法院得依夫妻之一方、主管機關、社會福利機構或其他利害關係人之請求或依職權酌定之。」亦即法律先尊重父母雙方之協議，未為協議或協議不成者，始由法院酌定之<sup>13</sup>。目前關於夫妻離婚後的子女親權酌定，內政部的統計數據並未標記一方為外籍配偶之情形，故無從得知有多少外籍配偶取得子女親權。另一方面，法院的裁判則可能讓吾人一探究竟，因裁判書所記載的當事人資料可能看出外籍配偶之身分與國籍。如前述，離婚時夫妻雙方若對子女的親權行使方式無共識，始會走進法院，從而法院裁判並非外籍配偶離婚的全數案例，只是少數無法達成協議的案例，但卻是目前觀察此問題時的唯一分析素材。

我國的民法、家事事件法並未對不同國籍者設有差別待遇，不過，一些研究指出了新住民在爭取親權時較為不利的現象（將於「貳、文獻回顧」部分詳述）。例如，有認為法官在個案中指出外籍父母中文不佳，對子女學習教養，終究不如本籍父母，這顯示了法官個人或司法體系對外籍配偶的制度性歧視<sup>14</sup>。那麼，這樣的「歧視」到底是個案還是通案？我國的司法體系對外籍配偶是否存在普遍的不平等？

---

六、外國人與本國雇主發生勞資爭議，正在進行爭訟程序。」

<sup>13</sup> 林秀雄（2013），《親屬法講義》，3版，頁203，元照。

<sup>14</sup> 施慧玲、陳竹上（2011），〈論婚姻移民家庭權之平等保障：全球法本土化的考

為了探討此問題，此處先說明本文採取的「平等／不平等」之定義，係指「法院對外國人與本國人是否採取相同的審酌標準」。具體的作法是檢視法院的裁判書，在「僅涉及本國籍父母之親權酌定裁判」與「一方為外國籍父母的親權酌定裁判」之「裁判用語」是否有差異；若有，則認定法院對此二種狀況採取不同的審酌標準。倘若在涉及外籍配偶的裁判中，出現上述「外籍父母中文不佳，故不如本籍父母」的句子，由於雙方皆為本國籍配偶的裁判，顯然不會有這樣的文句，從而，在本文採用的研究方法之下，結果將會是「法院在本國籍父母與外國籍父母的裁判用語有差異」；若然，在上述「平等／不平等」之操作型定義下，研究結論應該會是「不平等」。反之亦然。在研究方法上，將使用人工智慧（artificial intelligence）<sup>15</sup>中的機器學習（machine learning）<sup>16</sup>與「文字探勘」（text mining）<sup>17</sup>的技術，分析法院的裁判書（方法將於「參、研究設計」部分詳述）。

上述研究方法呼應了本刊「跨域、科技與法律」的主題，亦即，本文使用資訊科技技術，來輔助吾人觀察司法實務的運作，回答「臺灣的法院對於新住民是否平等裁判」之疑問，此為結合資訊與法律兩個領域的研究。此外，雖然本文的研究方法與「先將概念編碼，再以統計技術計算並得出觀察結果」的法實證研究有些許不同，本文所採取的是「不經編碼，直接以文本計算」的法律資料分析作法，但均是關懷法律落實狀況的法經驗面向<sup>18</sup>。近年新任

---

察與反思》，《月旦法學雜誌》，189期，頁33。

<sup>15</sup> 關於人工智慧方面的法律議題盤點，參見林勤富、劉漢威（2018），〈人工智慧法律議題初探〉，《月旦法學雜誌》，274期，頁195-215。

<sup>16</sup> 機器學習係指能自動學習或改善其表現的電腦演算法，亦即藉由巨量數據的蒐集、模式分析，訓練機器逐漸修正並尋找其中的規則，而非仰賴人類預先設定的規則。See Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 94 (2014).

<sup>17</sup> 文字探勘係運用語言學、統計、機器學習等技術，對文件執行編輯、組織與分析的過程，目的在瞭解內容事實和趨勢，並探勘隱性特徵的有用價值。See DAN SULLIVAN, DOCUMENT WAREHOUSING AND TEXT MINING: TECHNIQUES FOR IMPROVING BUSINESS OPERATIONS, MARKETING, AND SALES 326 (2001); Xia Hu & Huan Liu, *Text Analytics in Social Media*, in MINING TEXT DATA 385, 387-88 (Charu C. Aggarwal & Chengxiang Zhai eds., 2012).

<sup>18</sup> 關於實證方法或實證研究之定義，參見王鵬翔、張永健（2015），〈經驗面向的

民逐漸參與臺灣社會，自然有使用司法制度的機會，本文亦可能提供新住民認識臺灣司法的思考框架，有助於臺灣多元群體建立共同價值觀。

## 貳、文獻回顧

關於新住民的相關研究眾多，故此處僅討論與法律相關的文獻。

首先有從憲法角度出發，建構「家庭權」概念，並論及外籍勞工應具有家庭權；至於外籍配偶，應給予平等對待，且著眼於「婚姻與家庭權」，應提供多元服務並建構支持網路<sup>19</sup>。

其次，有不少文獻探討新移民的入境、居留至歸化政策。在入境方面，2006年的文獻指出，憲法與法律保障「婚姻自由」、「婚姻制度」及「夫妻共同生活」的權利，但在外籍配偶欲入境與我國籍配偶共同生活時，各種入境許可與否的規定之法律要件模糊，加上行政機關得運用「面談」，使得駐外單位與境管單位擁有廣泛而不受約制的判斷權限<sup>20</sup>。在居留權及身分申請的問題，此文獻指出，外籍配偶只要未取得中華民國國籍並設戶籍，一旦離婚，就會因「逾期居留」而被驅逐出境，而鮮少考慮被驅逐者的處境例如離婚原因為何<sup>21</sup>。在2007年修正入出國及移民法時，於第31條增設了數款

規範意義：論實證研究在法學中的角色》，《中研院法學期刊》，17期，頁207-208。

<sup>19</sup> 李震山（2004），〈憲法意義下之「家庭權」〉，《中正大學法學集刊》，16期，頁96-97。

<sup>20</sup> 廖元豪（2006），〈全球化趨勢中婚姻移民之人權保障：全球化、台灣新國族主義、人權論述的關係〉，《思與言：人文與社會科學雜誌》，44卷3期，頁86-89。另外，夏曉鵬（2002），〈流離尋岸：資本國際化下的「外籍新娘」現象〉，頁57-81，台灣社會研究雜誌社；Pei-Chia Lan, *Migrant Women's Bodies as Boundary Markers: Reproductive Crisis and Sexual Control in the Ethnic Frontiers of Taiwan*, 33 SIGNS: J. WOMEN CULTURE & SOC'Y 833, 833-61 (2008)。均指出，移民官員在面談婚姻移民時，常涉及侵犯個人隱私的問題，在態度和行為上表露歧視，甚至認為婚姻移民是國家的恥辱。

<sup>21</sup> 王君琳（2005），〈性別與國族：從女性主義觀點解讀新移民女性現象〉，夏曉娟（主編），《不要叫我外籍新娘》，頁197，左岸文化；廖元豪，前揭註20，頁102。

規定<sup>22</sup>，考量離婚事由與未成年子女之照顧需求，得准予繼續居留。惟對於現行法仍有提出質疑，批評此規定原則上禁止離婚之新住民再行居留臺灣，將導致受家庭暴力之新住民，因害怕離婚後被廢止其居留許可並註銷其外僑居留證，故隱忍不敢訴請離婚<sup>23</sup>。又，即使獲得居留權，離婚後的單親婚姻移民女性常面臨諸多艱困的狀況，例如對法律及法庭的不瞭解往往使其在面臨離婚官司時處於劣勢<sup>24</sup>。最後是歸化時的障礙，由於我國法律規定外籍配偶申請歸化必須提出「財力證明」，在實務上造成許多跨國通婚家庭困擾<sup>25</sup>。將財力與公民身分綁在一起，隱含外籍配偶被視為「無能依賴者」的錯誤推論，有可能深化跨國婚姻家庭內的性別壓迫<sup>26</sup>。

更進一步，聚焦於外籍配偶與我國籍配偶「離婚問題」之法學相關研究，則有下述幾篇。首先，郭書琴在 2007 年搜集 2001 至 2004 年共 22 個涉及跨國婚姻的判決，從事實證分析，指出其中 16 個判決是我國籍配偶（男性）為原告，而被告（外籍配偶）則在法庭上缺席，呈現一造辯論判決之現象<sup>27</sup>。

施慧玲、陳竹上在 2011 年針對 2010 年 1 月至 3 月涉及外籍配偶離婚的 43 件判決進行實證研究，亦發現超過 9 成的原告是我國籍配偶，超過 9 成

---

<sup>22</sup> 即第31條第4項之規定，內容參見前揭註12。

<sup>23</sup> 監察院（06/02/2016），〈外籍配偶離婚後居留制度極不合理 監察院促內政部及移民署推動修法並採取補救〉，[https://www.cy.gov.tw/News\\_Content.aspx?n=213&s=7657](https://www.cy.gov.tw/News_Content.aspx?n=213&s=7657)（最後瀏覽日：10/09/2020）。另須指出的是，不僅臺灣，在其他已開發國家，外籍配偶要取得合法的居留權及公民身分，也根據是否有繼續維持婚姻關係而定，例如Caren Freeman, *Marrying Up and Marrying Down: The Paradoxes of Marital Mobility for Chosŏnjok Brides in South Korea*, in *CROSS-BORDER MARRIAGES: GENDER AND MOBILITY IN TRANSNATIONAL ASIA* 80, 80-100 (Nicole Constable ed., 2005).

<sup>24</sup> Abdul Paliwala、Anne Elizabeth Barlow、許碧純（2011），〈跨國／跨文化婚姻移民的人權保障與移民政策：法律與社會政策的跨領域對話〉，《中正大學法學集刊》，34期，頁20。

<sup>25</sup> 廖元豪，前揭註20，頁104。

<sup>26</sup> 陳雪慧（2005），〈我們會是一家人？！〉，夏曉娟（主編），《不要叫我外籍新娘》，頁178-179，左岸文化。

<sup>27</sup> 郭書琴（2007），〈逃家的妻子，缺席的被告？：外籍配偶與身分法之法律文化初探〉，《中正大學法學集刊》，22期，頁21。

是一造辯論判決，印證了郭書琴指出的外籍配偶在離婚訴訟「缺席」的現象<sup>28</sup>。即使外籍配偶參與了訴訟，但外籍配偶在語言、經濟、親屬支持系統方面，難以與夫家居於平等之立足點，在現行子女最佳利益及相關指標的運作下，常被質疑或否定擔任親權人之能力<sup>29</sup>。此研究認為，法官不應僅因一方的種族與臺灣民情習俗不同而排除其行使子女親權之權利，故提倡將多元文化理念納入司法系統，並建構更符合外籍人士需求的通知程序，減少一造辯論之比例，並將取得會面交往權納入外籍父母得繼續居留之事由<sup>30</sup>；換言之，家庭法規不但不應因「移民」身分而有歧視規定，尚且應基於其跨國婚姻而產生的特殊需求給予特別保護<sup>31</sup>。

同年陳竹上與曾竹寧的論文，研究了一個外籍母親經歷三級三審爭取親權未果的事件。本研究呼籲從事訪視調查之社工人員，應具備更深入之相關法律知能，有助於確保當事人之權益及促成法院之最佳判斷<sup>32</sup>。

2016年陳昭如的研究，並非直接關注外籍配偶的離婚，而在探討子女親權酌定與女性主義、父親權利運動之關聯。此論文選取了2000至2013年間共272件親權酌定裁判，發現一造辯論判決的數量增加，並且此現象主要來自於跨國婚姻<sup>33</sup>。

2019年姜晴文的論文選取了2009至2017年子女親權酌定之第一審裁判2,775件，共有未成年子女4,340人，發現其中母親為外籍人士之子女有1,176人，亦即從涉訟子女人數來看，外籍母親占比28.45%。此研究將法院酌定親權時的考量因素予以編碼，並以實證之方法分析，指出外籍母親獲得親權的比率，從無書狀的1.16%大幅提高到有書狀的45.38%，故參與訴訟

---

<sup>28</sup> 施慧玲、陳竹上，前揭註14，頁27-29。

<sup>29</sup> 施慧玲、陳竹上，前揭註14，頁32。

<sup>30</sup> 施慧玲、陳竹上，前揭註14，頁33-34。

<sup>31</sup> 施慧玲、陳竹上，前揭註14，頁24。

<sup>32</sup> 陳竹上、曾竹寧（2011），〈法律知能與社會工作專業：跨界整合的探索及一位外籍配偶爭取監護權案例的啟發〉，《Asian Journal of Arts and Sciences》，2卷2期，頁177-179。

<sup>33</sup> Chao-Ju Chen, *The Chorus of Formal Equality: Feminist Custody Law Reform and Fathers' Rights Advocacy in Taiwan*, 28 CAN. J. WOMEN & L. 116, 143-44 (2016).

有利於爭取權益<sup>34</sup>。但即使參與訴訟，外籍母親獲得親權的比率還是不如我國籍母親，原因是在主要照顧者、現住所、支持系統的相對弱勢，而從裁判書的內容，也看見了法官為子女最佳利益而不得已如此決定的無奈<sup>35</sup>。

以上 5 篇的既存研究，共同特色是觀察的對象是法院的裁判書。或許因為這能凸顯書本上的法與現實中的法的差異：離婚時酌定親權的相關規定並未因父母的性別或國籍而有差別，但法院實際解釋、適用法律的結果，卻是外籍母親取得親權的比率比他們的配偶（我國籍父親）低，也遠低於我國籍母親。那麼為何一致的標準卻會造成結果的不同？既存研究或以少數裁判的內容分析、或以多數裁判的審酌要因編碼，來探究造成結果差異之原因。本文則希望另闢蹊徑。

亦即，透過逐漸發展的文字探勘技術，本研究將檢驗「雙方當事人都是我國人民」（以下稱雙臺籍裁判或雙臺籍事件）與「有一方當事人是新住民配偶」（以下稱外籍裁判或外籍事件），法官在此兩類裁判書的用詞是否有差異，若有差異，則內容為何，進而回答是否兩者是否不平等。雖本文與既存研究均是以法院的裁判書為素材，但既存研究是讓人類閱讀裁判書，透過人類的文字理解、概念分類後抽取出關鍵；相對地，本文是讓機器運用文字處理技術，找出每個裁判書的位置與特徵，以回答「差異」是否存在的問題。此一方法有著跨越資訊技術與法學兩種專業領域的特質。

使用語料庫或文字技術來分析社會議題的方法正在起步，有使用媒體報導來研究公投議題<sup>36</sup>、國際關係議題<sup>37</sup>，或研究土地徵收相關之社會運動<sup>38</sup>，

---

<sup>34</sup> 姜晴文（2019），《法律資料分析的優化與應用：以離婚後未成年子女親權酌定的裁判為素材》，頁186，國立臺灣大學法律學研究所碩士論文。

<sup>35</sup> 姜晴文，前揭註34，頁192-193。

<sup>36</sup> 陳世榮（2015），〈社會科學研究中的文字探勘應用：以文意為基礎的文件分類及其問題〉，《人文及社會科學集刊》，27卷4期，頁691-692。測試機器是否能判讀某篇新聞報導是支持民進黨版或國民黨版的公投。

<sup>37</sup> 郭文平（2020），〈語料庫輔助的媒體論述分析：以台灣平面媒體中國夢報導為語料的實證研究〉，《資訊社會研究》，38期，頁51-92。

<sup>38</sup> 王章逸、關河嘉（2020），〈大埔之歌：臺灣主流報紙中的「土地徵收」〉，《資訊社會研究》，38期，頁19-50。

目前所分析的文本都集中在新聞報導。使用文字探勘以分析法院裁判書的研究則較為有限，有使用「強盜罪」與「恐嚇取財罪」的刑事判決書，訓練機器分類者<sup>39</sup>；使用毒品相關刑事判決書的資料，描繪出我國的毒品流入與流出的管道與特徵者<sup>40</sup>；以及使用本國籍配偶離婚後法院酌定親權之裁判書，訓練機器判斷裁判結果者<sup>41</sup>。以往使用個案內容分析之方法來考察社會議題，分析對象（文本數量）較為有限，新的資訊科技方法可透過機器大量處理數據的能力，擴展分析的文本數量，甚至透過機器學習方法，發覺人類以往並未察覺的新洞見（insights）。本文期望在此一脈絡下，以文字探勘與機器學習（弱 AI）探討外籍配偶離婚時的子女親權酌定，將以人工智慧協助人類考察文本，找出問題所在，進而改善制度，或能增進實質的平等並連結至人性尊嚴。

---

<sup>39</sup> 林琬真、郭宗廷、張桐嘉、顏厥安、陳昭如、林守德（2012），〈利用機器學習於中文法律文件之標記、案件分類及量刑預測〉，《中文計算語言學期刊》，17卷4期，頁49-67。

<sup>40</sup> 邵軒磊、吳國清（2019），〈法律資料分析與文字探勘：跨境毒品流動要素與結構研究〉，《問題與研究》，58卷2期，頁96-97。

<sup>41</sup> 黃詩淳、邵軒磊（2020），〈以人工智慧讀取親權酌定裁判文本：自然語言與文字探勘之實踐〉，《臺大法學論叢》，49卷1期，頁195-224。

## 參、研究設計

### 一、資料來源

本研究之資料來源有二。首先是法律扶助基金會（以下簡稱「法扶」）提供了 2015 至 2017 年的「親權及監護權」事件<sup>42</sup>，共 4,248 件。

此 4,248 件准予扶助的親權事件中，父母雙方皆為本國籍人士之雙臺籍事件的女性當事人（即法扶之申請人）人數為男性之 3.2 倍，父母其中一方為外籍人士之外籍事件的女性當事人人數則為男性的 25.8 倍，可發現使用法扶服務的外籍配偶是女性壓倒性地多。但因本文的研究對象是裁判書，故不擬對法扶事件本身過多著墨。

法扶資料除了記載當事人及事件特徵、案件概述外，亦有法扶律師提供的法院裁判字號（但並無裁判全文）。藉此，本文再使用司法院的法學資料檢索系統查詢，獲得裁判內容與結果。不過，因為親權酌定事件涉及未成年人，相關裁判公開不多<sup>43</sup>，順利取得之公開裁判書僅有 116 件，其中的外籍裁判有 31 件，其餘 85 件為雙臺籍裁判，且均為地方法院或少家法院的第一審裁判。此 116 件的裁判書語料全文，構成了本研究的資料庫，作為後續文字探勘的對象。另須留意者，此 116 件裁判均為當事人（父母）之一方，向

---

<sup>42</sup> 法扶將家事事件分成數種類別，其中與本研究相關者為「監護權」及「親權」二種。至於此二類的區別，經詢問法扶工作人員，表示並無明確區分基準，而僅是接案律師選填之類別。由於實務經常混用「監護權」與「親權」二詞，本研究只得將此二類別的事件均納入分析對象。以 2017 年數據為例，家事事件最多之類別為給付扶養費 3,164 件，其次為離婚 1,919 件，第三位即為監護權 1,057 件，第四為親權 555 件。參見法律扶助基金會（2018），《2017 年度報告書》，頁 11，<https://www.laf.org.tw/upload/2018/05/20180528114440.pdf>（最後瀏覽日：06/29/2020）。

<sup>43</sup> 且每個法院公開與否並無一定之標準，相關說明參見黃詩淳、邵軒磊（2019），〈人工智慧與法律資料分析之方法與應用：以單獨親權酌定裁判的預測模型為例〉，《臺大法學論叢》，48 卷 4 期，頁 2041。



法扶申請扶助之事件。也就是說，這 116 件多是申請人對爭取親權有意願，獲得法扶律師協助。申請人多為女性，少數為男性。關於其中 31 件外籍裁判的原被告國籍分布以及親權酌定之結果，詳如後述【表 7】所示。

之所以執著於使用法扶與司法院兩方「交集」的事件／裁判，是因為法院裁判書並不會詳細記載每位當事人的各種資料，例如年齡、收入、國籍、法扶律師等<sup>44</sup>，且一般透過司法院系統檢索到的外籍裁判，如既存研究指出，多數都是一造辯論，如此，缺席的外籍配偶的相關資料，更難在裁判書裡看出。因此，本研究所觀察的乃是「少數」確定積極利用法律扶助與司法機關的外籍或外裔配偶，亦即本文主要探討的是「法官在外籍、外裔配偶有意願爭取親權時，相較於父母雙方皆為臺籍配偶之裁判，是否會考量不同的因素」。

## 二、研究方法

本文使用「法律資料分析」(legal analytics)的方法，也就是運用資訊科學技術如文字探勘、自然語言處理、機器學習等，使得電腦能夠直接從法律文本(legal texts)例如裁判書或法規中，萃取出有意義的資訊，並依此進行運算、推理<sup>45</sup>。法律資料分析乃涉及對於法律文書和資料的探勘，並將探勘得到的資料合併觀察，以獲致關於法律人(如律師和法官)、法律組織(如法院或法律事務所)、或特定法律主題(如專利)方面原先未知的知識<sup>46</sup>。換言之，法律資料分析是「法學」與「資料科學」的跨領域研發之研究

<sup>44</sup> 同樣的問題已在過去的實證研究中被指摘，參見張永健、李宗憲(2015)，〈身體健康侵害慰撫金之實證研究：2008年至2012年地方法院醫療糾紛與車禍案件〉，《臺大法學論叢》，44卷4期，頁1829。

<sup>45</sup> Daniel Martin Katz, *Quantitative Legal Prediction - or - How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry*, 62 EMORY L. J. 909, 930-932 (2013); Julie Sobowale, *How Artificial Intelligence Is Transforming the Legal Profession*, ABA J. (Apr. 1, 2016), [http://www.abajournal.com/magazine/article/how\\_artificial\\_intelligence\\_is\\_transforming\\_the\\_legal\\_profession](http://www.abajournal.com/magazine/article/how_artificial_intelligence_is_transforming_the_legal_profession) (last visited June 20, 2020).

<sup>46</sup> Owen Byrd, *Legal Analytics vs. Legal Research: What's the Difference?*, L. TECH.

方法。以往法學實務工作非常依賴「經驗」，資深律師或許有豐富的經驗、專業訓練與直覺，但一個好的機器學習演算法，可能作為彌補律師經驗不足的工具<sup>47</sup>。

本文欲探討的「歧視／平等」概念非常抽象，即使裁判所依據的法規看似並無歧視，即使子女最佳利益的判準在每個裁判中都一體適用，有可能因裁判結果使外籍母親難以獲得親權，而無法脫免「歧視」的批判之聲<sup>48</sup>。本文嘗試提出一個客觀測量「平等」與否的方法，也就是說，本文先將「平等」定義為「法院對外國人與本國人是否採取相同的審酌標準」，至於判斷方式，不是使用人類閱讀判決再獲取「感想」，而是運用文字探勘、機器學習之資訊技術，分析雙臺籍裁判與外籍裁判之內容，觀察此兩種類型的裁判書用語是否有不同之處；如果有，內容為何；藉此來獲得「我國法院是否對外籍配偶平等對待」之結論。

### 三、研究步驟

整體的研究流程，是先蒐集了上述共 116 件親權酌定裁判，再經過自然語言處理、文字探勘與「梯度提升演算法」（gradient boosting），試圖發現雙臺籍裁判與外籍裁判是否有不同之處。

文字探勘係運用語言學、統計、機器學習等技術，對文件執行編輯、組織與分析的過程，目的在瞭解內容事實和趨勢，並探勘隱性特徵的有用價值<sup>49</sup>。文字探勘主要程序包含資料檢索與處理、分詞、特徵萃取、分類與集群、

---

TODAY (June 12, 2017), <https://www.lawtechnologytoday.org/2017/06/legal-analytics-vs-legal-research/> (last visited June 20, 2020).

<sup>47</sup> John O. McGinnis & Russell G. Pearce, *The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services*, 82 *FORDHAM L. REV.* 3041, 3052 (2014).

<sup>48</sup> 例如施慧玲、陳竹上，前揭註14，頁33，認為法官在個案的見解隱含了法官個人層次的種族歧視，並反映了臺灣社會對越南移民的文化歧視，以及司法體系對越南被告的制度性歧視。

<sup>49</sup> SULLIVAN, *supra* note 17, at 326.

文本表示與詮釋等<sup>50</sup>。文字探勘可以比較不同文本的相似度，本研究運用此技術來比較 116 件裁判書的相似程度。

在作法上，將每則裁判書作為一筆資料，接著將裁判全文做分詞（word segmentation）處理。此際，係使用裁判書全文，不須由人工挑選特定段落。這是因為本文目的並不是讓電腦猜測（分類）親權酌定的結果是判給父親或母親，所以不需要遮去裁判主文，而是希望分析雙臺籍裁判與外籍裁判的用語差異之故。分詞的範例如下【表二】所示：

【表二】本研究分詞範例表（以臺灣士林地方法院 104 年度婚字第 93 號判決為例）

分詞前	分詞後
<p>本院於民國 104 年 11 月 10 日言詞辯論終結，判決如下：主文准原告與被告離婚。對於兩造所生未成年子女丙○○（男，民國八十四年十二月十七日生，身分證統一編號：000000000 號）、甲○○（女，民國九十一年十月十二日生，身分證統一編號：000000000 號）權利義務之行使或負擔，由原告任之。訴訟費用由被告負擔。</p> <p>並聲明：（一）准原告與被告離婚。（二）對於丙○13○、甲○○權利義務之行使或負擔由原告任之。三、被告未於言詞辯論期日到</p>	<p>本院 於 民國 104 年 11 月 10 日 言詞辯論 終結 判決 如下 主文 准 原告 與 被告 離婚 對於 兩造 所生 未成年 子女 丙 男 民國 八十四年 十二月 十七日 生 身分證 統一 編號 000000000 號 甲 女 民國 九 十一年 十月 十二日 生 身分證 統一 編號 000000000 號 權利義務 之 行使 或 負擔 由 原告 任 之 訴訟費用 由 被告 負擔 並 聲明 一准 原告 與 被告 離婚 二 對於 丙 13 甲 權利義務 之 行使 或 負擔 由 原告 任 之 三 被告 未 於 言詞辯論 期日 到場 亦 未 提出 書狀 作何</p>

<sup>50</sup> 曾元顯（2002），〈文件主題自動分類成效因素探討〉，《中國圖書館學會會報》，68期，頁62-83；RONEN FELDMAN & JAMES SANGER, THE TEXT MINING HANDBOOK: ADVANCED APPROACHES IN ANALYZING UNSTRUCTURED DATA 13-17 (2006).

<p>場，亦未提出書狀作何聲明或陳述。四、按夫妻有民法第 1052 條第 1 項各款以外之重大事由，難以維持婚姻者，夫妻之一方得請求離婚，但其事由應由夫妻之一方負責者，僅他方得請求離婚，民法第 1052 條第 2 項定有明文。而所謂「難以維持婚姻之重大事由」，係指婚姻是否已生破綻而無回復之希望，應依客觀之標準進行認定，審認是否已達於倘處於同一境況，任何人均將喪失維持婚姻意欲之程度（最高法院 94 年度台上字第 115 號判決意旨參照）。又婚姻係以夫妻相互間之感情為立基，並以經營夫妻之共同生活為目的，故夫妻自應誠摯相愛，並互信、互諒，協力保持婚姻共同生活之圓滿及幸福。倘上開基礎已不復存在，夫妻間難以繼續共同相處，彼此間無法互信、互諒，且無回復之可能時，自無仍令雙方繼續維持婚姻形式之必要，此時即應認有「難以維持婚姻之重大事由」。</p>	<p>聲明或陳述。四、按夫妻有民法第 1052 條第 1 項各款以外之重大事由，難以維持婚姻者，夫妻之一方得請求離婚，但其事由應由夫妻之一方負責者，僅他方得請求離婚。民法第 1052 條第 2 項定有明文。而所謂「難以維持婚姻之重大事由」，係指婚姻是否已生破綻而無回復之希望，應依客觀之標準進行認定，審認是否已達於倘處於同一境況，任何人均將喪失維持婚姻意欲之程度。最高法院 94 年度台上字第 115 號判決意旨參照。又婚姻係以夫妻相互間之感情為立基，並以經營夫妻之共同生活為目的，故夫妻自應誠摯相愛，並互信互諒，協力保持婚姻共同生活之圓滿及幸福。倘上開基礎已不復存在，夫妻間難以繼續共同相處，彼此間無法互信互諒，且無回復之可能時，自無仍令雙方繼續維持婚姻形式之必要，此時即應認有難以維持婚姻之重大事由。</p>
---	--

※ 資料來源：司法院。作者自製。

說明：因篇幅關係本處僅提供部分判決內容，但在實際分詞時是以裁判全文為對象。

由於使用了自動分詞系統，有時分詞結果無法盡如人意，例如上【表二】右欄中，出現了「條第」、「項定」、「人均」、「倘上」等，違反法律人直覺的分詞。但多數詞彙例如「權利義務」、「行使」、「負擔」、「喪失」、「維持」、「婚姻」、「意欲」，已正確斷開，少數錯誤的分詞不致影響後續運算的結果。

接著計算出每件裁判中，每個詞彙的 TF-IDF (term frequency-inverse document frequency) 值，並將該值製成文本詞彙矩陣 (document term matrix, DTM)，以向量空間來呈現資料 (vector-based representation)。

在一般想法中，在一個文本內，出現次數 (詞頻) 愈高的詞彙愈重要，也就是愈能代表該文本的特徵。然而，實際上卻未必如此，大量的常用字，例如「得」、「的」、「很」等詞彙，又例如裁判書中的「原告」、「被告」、「上訴人」、「被上訴人」等詞彙，都是高詞頻詞彙，但這些詞彙由於太常出現，反而沒有太大意義。更精確地說，由於每一份裁判書中都有原告，所以「原告」一詞並非該裁判書的「特徵」。為了解決文本中出現的無意義頻繁詞的問題，並取出文本特徵，就有了 TF-IDF (詞頻-逆向文件頻率) 的算法。它的基本概念是：一個詞在一個文件中出現次數愈多，同時，在所有文件 (語料庫) 中出現次數愈少，該詞愈能代表該文件的特徵<sup>51</sup>，換句話說，這樣的詞彙愈能將該文件與其他語料庫中的文獻區別開來<sup>52</sup>。作法是將每個文件中特定詞彙的詞頻值，再乘上一個文件頻率倒數 (取對數)，這樣有助於消解掉「每份文件都有的詞」的影響力。詳細說明如下：

TF (詞頻, term frequency) 的算法：分子是  $n_{i,j}$ ，是該詞  $t_i$  在檔案  $d_j$  中的出現次數，而分母則是在檔案  $d_j$  中所有字詞的出現次數之和。使用「比率」而非次數，可以避免長文本與短文本出現的天然差異。

<sup>51</sup> Karen Spärck Jones, *A Statistical Interpretation of Term Specificity and Its Application in Retrieval*, 28 J. DOCUMENTATION. 11, 17 (1972).

<sup>52</sup> Gerard Salton & Christopher Buckley, *Term-Weighting Approaches in Automatic Text Retrieval*, 24 INFO. PROCESSING & MGMT. 513, 516 (1988).

$$\text{tf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

而 IDF (inverse document frequency, 逆向文件頻率) 則是

$$\text{idf}_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|}$$

分子 $|D|$ : 語料庫中的檔案總數。分母 $|\{j : t_i \in d_j\}|$ : 包含詞語 $t_i$ 的檔案數目(即 $n_{i,j} \neq 0$ 的檔案數目)。如果該詞語不在語料庫中, 就會導致分母為零, 因此有時使用 $1 + |\{j : t_i \in d_j\}|$ 。相對地, 如果這個詞越「普遍」則分母就越大, 所以整體數值就小; 最極端的情形, 每個文本都有某個詞(如「的」), 這樣 IDF 數值就會是零( $\log 1=0$ ), 進而, TF-IDF 整體數值也會是零。

舉例而言, 下【表三】比較了 9 個文本(虛構文本)的原始 DTM 與 TF-IDF。在左邊表格也就是原始的 DTM 中, 「原告」與「被告」幾乎在每件文本都出現。在其中的 c5、m2、m3 文本中, 「原告」的詞頻高過我們想知道的一些特徵(「租屋」、「到庭」), 但法律人的常識應可理解, 「原告」一詞是每個裁判都有的詞彙, 它的存在無法說是某個裁判的特徵。所以, 如右方表格, 在詞頻加入 IDF 的參數後, 數值的大小比較能反映出該文本的詞彙特徵。

【表三】DTM 與 TF-IDF 之比較

DTM	原告	被告	租屋	到庭	tf*idf(j)	原告	被告	租屋	到庭
c1	1	1	0	1		0.00	0.04	0.00	0.50
c2	1	1	2	2		0.00	0.02	0.20	0.50
c3	1	1	2	0		0.00	0.03	0.29	0.00
c4	1	1	3	0		0.00	0.02	0.35	0.00
c5	2	1	1	0		0.00	0.03	0.15	0.00
m1	1	1	0	0		0.00	0.06	0.00	0.00
m2	3	1	0	0		0.00	0.03	0.00	0.00
m3	4	1	0	0		0.00	0.02	0.00	0.00
m4	1	0	1	0		0.00	0.00	0.29	0.00
df(j) <sup>a</sup>	9	8	5	2					
idf(j) <sup>b</sup>	0.00	0.12	0.59	1.50					

a : df(j)意指出現該詞彙的文本數量

b : idf(j) 逆向文件頻率

※ 資料來源：作者自製。

在本研究中，由於分詞後的詞彙數量極大，我們先獲得 TF-IDF 值後再做後續運算，較諸僅使用 DTM 會得到較好的效果。在其他資訊科學領域使用 TF-IDF 來做文本資訊提取相當常見<sup>53</sup>。

<sup>53</sup> Ho Chung Wu, Robert Wing Pong Luk, Kam Fai Wong & Kui Lam Kwok, *Interpreting TF-IDF Term Weights as Making Relevance Decisions*, 26 ACM TRANSACTIONS ON INFO. SYS. 13:1, 13:13-13:36 (2008).

計算出每個詞彙的 TF-IDF 值後，再將每件裁判書每個詞彙的 TF-IDF 製成 DTM。本研究共有 116 個裁判，故矩陣有 116 列。至於欄，因有的裁判有字元漏失導致部分語料喪失，最終由 2,097 個詞彙所組成矩陣的欄，如下【圖三】所示。

A	B	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	1. 上林04審43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2. 上林04審19	0	0.04851	0	0	0.02644	0.02544	0	0.01961	0	0.14792	0.01374	0	0	0	0	0	0.01588	0.02341	0.07187	0.01588	0	0	0.01938	0
3	3. 上林04審3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03636	0	0	0	0	0
4	4. 上林03審30	0	0	0	0	0	0	0	0	0.02813	0	0	0	0	0	0	0	0.0193	0.0144	0	0	0	0	0	0.02385
5	5. 上林03審60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03782	0	0	0	0	0	0	0.02523
6	6. 上林03審13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04283	0	0	0	0	0	0	0
7	7. 上林03審6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02157	0	0	0	0	0	0	0
8	8. 上林03審72	0	0	0	0	0	0	0	0	0.03544	0	0.01461	0	0	0	0.04954	0.01668	0.05702	0.01668	0.0308	0	0	0	0	0
9	9. 上林03審8	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0276	0	0	0.0134	0.01	0.02879	0	0	0	0	0
10	10. 宜蘭03審0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02598	0	0	0.02588	0.0238	0.03532	0.0238	0	0	0.04305	0
11	11. 花蓮03審1	0.05789	0	0	0	0	0	0	0.05772	0.05789	0	0.04465	0	0.02747	0.01833	0	0	0.0238	0.03532	0.0238	0	0	0	0	0
12	12. 南投04審10	0.03841	0.03924	0	0.03924	0	0.03288	0	0	0	0	0	0	0.02301	0	0	0	0.01553	0	0.0196	0.04401	0	0	0	0
13	13. 南投04審21	0	0	0	0.0485	0	0	0	0	0	0	0	0	0.02301	0	0	0	0.0196	0.04401	0	0	0	0	0	0
14	14. 南投04審3	0	0	0	0.0779	0.02314	0	0	0	0	0	0	0	0.01946	0	0.03166	0	0	0.03005	0	0.02027	0	0	0	0.02514
15	15. 南投04審2	0	0	0	0	0	0	0	0	0	0	0	0	0.0178	0	0.03042	0	0	0.03005	0	0.02027	0	0	0	0
16	16. 南投03審04	0	0.04002	0	0	0	0	0	0	0	0	0	0	0.01689	0	0	0	0	0.03166	0	0	0	0	0	0.02822
17	17. 屏東03審169	0	0	0	0	0	0	0	0	0.010725	0.02537	0.01596	0	0	0	0	0	0	0.0131	0.03911	0.01998	0.0282	0	0	0
18	18. 屏東03審33	0	0	0.02026	0	0.02119	0.0283	0	0.01946	0	0	0	0	0	0	0.03893	0	0	0.0131	0.03911	0.01998	0.0282	0	0	0
19	19. 屏東03審45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02318	0	0	0	0	0	0	0
20	20. 屏東03審51	0.01716	0.01753	0.01437	0	0	0	0	0.01385	0.01716	0.04649	0.03176	0	0.030814	0	0	0	0	0.03929	0.04162	0	0.01716	0.01681	0.02297	0
21	21. 屏東03審27	0	0	0	0	0	0	0	0	0.02822	0	0	0	0.01977	0	0	0	0	0.03349	0	0.02257	0	0	0	0
22	22. 屏東03審112	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07023	0	0	0	0	0	0
23	23. 屏東03審113	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03972	0.02703	0	0	0	0	0.02118
24	24. 屏東03審135	0	0	0	0	0	0	0	0	0.04655	0.04653	0.01154	0	0	0	0.05336	0	0	0.03972	0.02703	0	0	0	0	0.02118
25	25. 屏東03審55	0	0	0	0	0	0	0	0	0.05171	0	0	0	0	0	0	0	0	0.04517	0	0	0	0	0	0
26	26. 屏東03審35	0	0	0	0	0.03857	0	0	0	0.02831	0.03859	0	0.01983	0	0	0	0	0	0.0169	0	0.02284	0	0	0	0
27	27. 屏東03審43	0	0	0	0	0	0.01834	0.01884	0	0	0	0	0	0	0	0	0	0.02157	0	0.01157	0.02155	0	0	0.01443	
28	28. 屏東03審48	0.02349	0	0.02721	0	0	0	0	0	0	0	0	0	0.01154	0	0	0	0.01799	0.01313	0	0.02349	0	0	0	0
29	29. 屏東03審4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04116	0	0	0	0	0	0
30	30. 屏東03審30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	31. 屏東03審22	0	0	0	0	0	0	0	0	0	0.02817	0.02516	0	0.02674	0	0	0	0	0	0	0.02278	0	0	0	0
32	32. 屏東03審7	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07888	0.03663	0	0	0	0	0	0	0	0	0
33																									

【圖三】本研究 TF-IDF 文本詞彙矩陣示意圖

※ 資料來源：作者自製。



經上述方法生成的文本詞彙矩陣，因經由 TF-IDF 處理，相較於單純的詞頻，將出現大量的空值（零）。詳言之，造成零的原因大致有二：第一，由於 TF-IDF 的計算，是將詞頻值乘以文件頻率倒數，若某個詞彙從未出現在某個裁判中，則詞頻值是零，自然 TF-IDF 值亦為零。第二，如上所述，若每則裁判都出現的詞彙，例如「的」、「中華民國」、「原告」、「法官」、「書記官」，IDF 值為零，則 TF-IDF 值亦為零。這樣的詞彙矩陣維度高達 2,097（每個詞彙就是一個維度），零值很多，相當稀疏，因此單一特徵詞，即使影響相當大，但最後能提供資訊數量的絕對值也較小，不容易觀察。又，因為文本詞彙矩陣維度過高，人類難以想像每個文本在該高維空間中的位置，故接下來使用主成分分析（Principal Component Analysis，以下稱 PCA）來降低維度，幫助使用者觀察每個文本（亦即裁判）的特徵所在。

PCA 是一種資訊處理方法，可簡化變數。其原理是使用一個新的平面或直線，將原有的資訊點「投影在新的平面」上，從而將原來所需要的描述維度（dimension）減少，一般稱為降維（dimensionality reduction）<sup>54</sup>。PCA 的特色是，降維後的各點還是保應該有原來特性，而在維度減少之後，其分群特性會更加明顯。其算法是求一個投影面，最小化各點到這個投影面的投影誤差。在高維運算中，可將數百維降維成數十維，但由於人類視覺的限制，以二維來呈現較為適當。本文將原本 2,097 維的數據，經由 PCA 聚合成 2 個維度後，製成如下述【圖四】，其意義將在下述第肆部分再為檢討。

## 肆、研究發現與討論

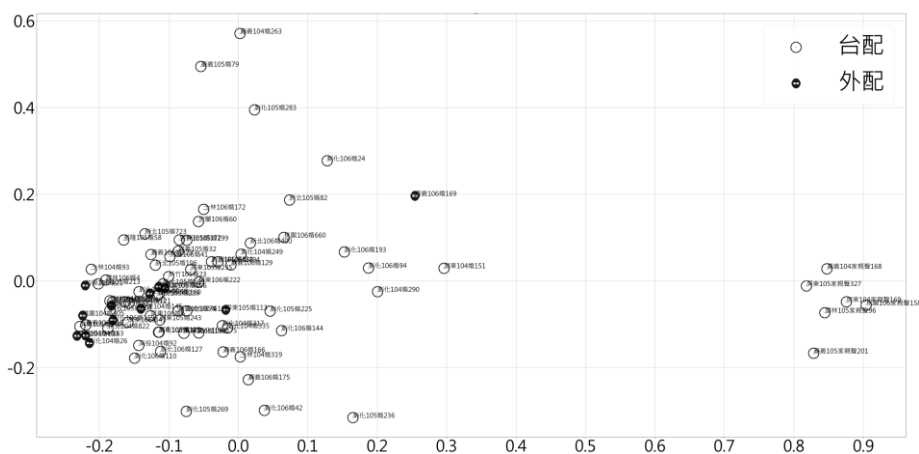
### 一、外籍裁判與雙臺籍裁判之詞彙並無不同

本文將 2,097 維所代表的裁判詞彙特徵數據，以 PCA 投影在二維平面上，一個點代表了一則裁判（見下【圖四】），距離愈近的點，代表其文本

---

<sup>54</sup> I. T. JOLLIFFE, PRINCIPAL COMPONENT ANALYSIS 1 (1986).

特徵愈相似。為區辨哪一點是哪一件裁判，在每一點標上了裁判字號，其中，白色圓圈是「雙臺籍裁判」，黑色圓圈是「外籍裁判」。



【圖四】親權酌定裁判之用詞特徵 PCA 投影圖

(X、Y 軸之意涵詳如本文下述)

※ 資料來源：作者自製。

由【圖四】可看出，大多數的裁判都位於左下角，不論雙臺籍裁判或外籍裁判均如此。且外籍裁判與雙臺籍裁判並無特別分開為二群的狀況。這意味著，外籍裁判與雙臺籍裁判中的用語與特徵並無顯著差異。此外，外籍裁判的位置較集中（嘉義 106 婚 169 例外），相較之下，雙臺籍裁判雖多數亦在左下角，但亦有一些位於左上角與右下角，這說明了雙臺籍裁判可能較有歧異性與多樣性，外籍裁判則有彼此相似的用語特質。

PCA 可讓吾人更進一步查閱【圖四】中的二維軸線代表的意涵（詞彙）。水平的 X 軸是「聲請人」、「相對人」、「交往」、「扶養費」、「會面」、「給付」、「扶養」、「調解」、「每月」等詞彙的群組。垂直的 Y 軸代表「長女」、「次女」、「案主」、「大陸」、「給付」、「地區」、「父親」、「母親」、「人民」、「未成年」的詞彙群組。雙臺籍裁判有 6 件位在【圖四】的右下角，X 值較大的地方，可能意味著這些裁判中，法官較多提到了會面交往以及按月給付扶養費的問題。實際上查閱了這些右下角的裁

判後，發現均為「家親聲」字號，亦即雙方已經離婚，僅聲請酌定親權、扶養費、會面交往之案例。也因為這類事件是家事非訟事件，這些裁判中的當事人是「聲請人」、「相對人」；其他全部裁判則都是家事訴訟事件，當事人是「原告」、「被告」；因此，在 TF-IDF 的算法下，這 6 件裁判中的「聲請人」、「相對人」就會被認定為重要的詞彙，代表了這 6 個文本的特徵。換言之，這 6 件裁判是非訟程序，且無離婚相關詞彙，故 X 軸特徵的「聲請人」、「相對人」、「交往」、「扶養費」、「會面」、「給付」、「扶養」、「調解」等詞彙重要性，自然較其他離婚裁判高。這也表示機器所為的分類有一定的道理及可解釋性。

## 二、外籍裁判詞彙之特徵

PCA 呈現出「整體而言」雙臺籍裁判與外籍裁判之用詞並無顯著差異，但並不意味「沒有差異」。為了吹毛求疵，也就是找出兩類裁判書的差異，本文使用機器學習中的「梯度提升法」(gradient boosting)<sup>55</sup>，從單一模型中計算複數的決策樹，將每一個決策樹疊加起來的結果作為最後檢定成果，而回推出某種分類能否增加「模型的正確率」。而不斷反覆演算的過程中，若某些樹不能「增加」模型正確率，就會停止運算，而發展另外的決策樹。而也由於這個模型計算出複數的決策樹，因此在所有的因素(即 2,097 個詞彙)中，能計算出每個節點(node)對結果的「貢獻」(gain)。本研究之目的是發現雙臺籍裁判與外籍裁判的用詞差異，因此，結果(目標值)將是雙臺籍裁判或外籍裁判，而決策樹中的節點則是特定的詞彙。換言之，對機器而言，這是一個二元分類(binary classification)問題。梯度提升法幫助我們找出「能最佳區分」雙臺籍裁判與外籍裁判的詞彙。愈常在反覆運算的決策樹中出現的節點(詞彙)，對於總體模型的貢獻就越大，我們就認為是區分兩類裁判的關鍵詞彙。下【表四】是貢獻值為前 15 名之詞彙。

<sup>55</sup> 梯度提升法 (gradient boosting) 又可稱為 gradient tree boosting 或 gradient boosting machine, see Jerome H. Friedman, *Greedy Function Approximation: A Gradient Boosting Machine*, 29 ANNALS STAT. 1189, 1189-1232 (2001).

【表四】區分雙臺籍裁判與外籍裁判之關鍵詞表

貢獻排名	關鍵詞	貢獻值	可能法律意義
1	人民	0.32	「印尼國人民」等，表彰國籍
2	地區	0.08	外國地區、桃園地區物價指數
3	同住	0.05	居住環境、狀態
4	自無	0.05	
5	條定 <sup>56</sup>	0.04	
6	喜歡	0.04	
7	支持	0.03	社會支持系統
8	法律	0.03	
9	到庭	0.03	通常是「未到庭」或是「到庭證稱」，表彰參與程序狀態或是否有證人
10	項定 <sup>57</sup>	0.03	
11	租屋	0.03	生活狀態、居住型態
12	一年	0.02	
13	出境	0.02	通常是「入出境紀錄」、「出境未歸」、「不得攜未成年子女出境」，表彰居留狀態、生活狀態
14	上列	0.01	
15	大陸	0.01	國籍或居住狀態（赴大陸經商）

※ 資料來源與製表：作者自製。

<sup>56</sup> 原文應為「第N條定有明文」，然機器自動分詞效果不佳，將「條」與「定」當成了同一個詞彙，此為自動分詞系統的問題。未來可能必須自建辭典才能解決。

<sup>57</sup> 同上，「項定」應該也是「第N條第N項定有明文」分詞錯誤的結果。

以下進一步探討，上【表四】中貢獻值較高的詞彙，為什麼能夠較好地區分出雙臺籍裁判與外籍裁判（成為決策樹中的節點），以及使用情境。

### （一）「出境」、「到庭」

如下【表五】所示，「出境」一詞在外籍裁判中的 TF-IDF 平均值遠大於雙臺籍裁判，或許是事理之必然。因為通常一方為外籍配偶時，即使居住於臺灣，有意願爭取親權，我國籍配偶會要求外籍方「未經同意，不得帶子女出境」<sup>58</sup>。另外更為常見者，則是其中一方「出境迄今未歸」，可能是我國籍配偶出境<sup>59</sup>，更多則是外籍配偶出境<sup>60</sup>。

「到庭<sup>61</sup>」的值在外籍裁判則略大於雙臺籍裁判。在外籍裁判中，「到庭」一詞使用的場合，除了證人「到庭證稱」之狀況外，「到庭」多數指涉的是被告「未到庭抗辯」、「並未到庭或具狀表示具體意見」，也就是並未參與程序，通常會得到不利也就是無法獲得親權之結果。不過，未到法庭現場者不一定是外籍配偶方<sup>62</sup>，也有一些是我國籍配偶<sup>63</sup>。

<sup>58</sup> 例如臺灣南投地方法院104年度婚字第63號民事判決。

<sup>59</sup> 臺灣士林地方法院105年度婚字第96號民事判決、臺灣屏東地方法院105年度婚字第64號民事判決、臺灣嘉義地方法院105年度婚字第96號民事判決。

<sup>60</sup> 臺灣屏東地方法院104年度婚字第145號民事判決、臺灣屏東地方法院105年度婚字第113號民事判決、臺灣苗栗地方法院104年度婚字第45號民事判決、臺灣苗栗地方法院106年度婚字第21號民事判決、臺灣桃園地方法院104年度婚字第709號民事判決、臺灣高雄少年及家事法院105年度婚字第205號民事判決。

<sup>61</sup> 「到庭」一詞原為刑事訴訟程序之用語，例如刑事訴訟法第31條第2項「於審判期日無正當理由而不到庭者」、第116條之2第3項「命被告於宣判期日到庭」。民事訴訟法中並無「到庭」用語，均使用「到場」一詞，例如民事訴訟法第385條第1項「當事人之一造不到場者」；家事事件法亦同，例如該法第13條第1項「命當事人或法定代理人本人到場」。然而，在親權酌定裁判中仍經常看到「未到庭抗辯」、「並未到庭或具狀表示具體意見」之語句，可見法院在用字上常混淆二者，並不精確。

<sup>62</sup> 臺灣士林地方法院104年度婚字第143號民事判決、臺灣苗栗地方法院104年度婚字第45號民事判決、臺灣苗栗地方法院106年度婚字第21號民事判決。

<sup>63</sup> 臺灣士林地方法院105年度婚字第130號民事判決、臺灣士林地方法院105年度婚字第96號民事判決、臺灣桃園地方法院104年度婚字第405號民事判決、臺灣嘉義地方法院105年度婚字第96號民事判決、臺灣彰化地方法院106年度婚字第159號

【表五】「出境」與「到庭」在二類裁判的 TF-IDF 平均值對照表

	Mean(TF-IDF_出境)	Mean(TF-IDF_到庭)
外籍裁判	0.0371	0.0339
雙臺籍裁判	0.000642495	0.0208

※ 資料來源：作者自製。

## (二) 「租屋」、「同住」

另一個發現是「租屋」、「同住」的 TF-IDF 平均值在外籍裁判較雙臺籍裁判高。通常指涉的情狀是，夫妻感情不睦後，外籍配偶離開我國籍配偶家庭，攜帶子女在外租屋，與子女同住。租屋是一種居住環境，與子女同住與否則是主要照顧者原則及繼續性原則重視的事實，兩者都是子女最佳利益之判斷標準。而外籍裁判中，「租屋」的 TF-IDF 值遠較雙臺籍裁判為高（如下【表六】所示），凸顯出多數外籍配偶在台灣沒有原生家庭或親戚家庭可以「借住」，而不得不在外「租屋」的特性。實際上，在本研究的樣本中，這些攜帶子女在外租屋的外籍配偶<sup>64</sup>，幾乎都取得了子女親權。這或許能提供一定的啟示：即使新住民的親屬支持網絡較為薄弱，但若能透過外在的資源適當補強，使其能夠租屋而有穩定的居住環境照顧子女，將有很大機率被法官認為是合適的親權人。

---

民事判決。

<sup>64</sup> 臺灣南投地方法院104年度婚字第63號民事判決、臺灣桃園地方法院104年度婚字第405號民事判決、臺灣嘉義地方法院104年度婚字第119號民事判決、臺灣嘉義地方法院104年度婚字第192號民事判決、臺灣嘉義地方法院106年度婚字第169號民事判決。其中，臺灣嘉義地方法院106年度婚字第169號民事判決的外籍母親是自己搬出夫家在外租屋，子女仍住在本國籍父親家中，但外籍母親仍獲得共同親權並擔任主要照顧者。

【表六】「租屋」與「同住」在二類裁判的 TF-IDF 平均值對照表

	Mean(TFIDF_租屋)	Mean(TFIDF_同住)
外籍裁判	0.0155	0.0515
雙臺籍裁判	0.00495	0.0341

※ 資料來源：作者自製。

## 伍、結論

### 一、研究結論

本研究試圖提出並實踐一個判斷多數裁判「平等」與否的方法，也就是先定義所謂的平等是指「在父母之一方為外國籍的親權酌定事件，與父母雙方皆為本國籍的親權酌定事件，法院是否採取相同的審酌標準」；具體的驗證方法，是看法院的裁判書，在「一方為外國籍父母的親權酌定裁判」與「僅涉及本國籍父母之親權酌定裁判」之「裁判用語」是否有差異。亦即使用人工智慧的初級技術，包括自然語言、文字探勘與機器學習的方式，比較跨國婚姻與本國婚姻之夫妻，在離婚後酌定親權之 116 件裁判。首先，將裁判書中的詞彙轉化為 TF-IDF 的文本詞彙矩陣，以向量空間表示每則裁判的詞彙特徵，再透過主成分分析將之分類，並投影在二維平面上，將每則裁判所在位置可視化。在該平面上距離愈近之裁判，代表其文本相似度愈高。本研究發現，外籍裁判與雙臺籍裁判並未分布在投影平面上的不同位置而成為兩群，相反地，此二種裁判呈現大圓（雙臺籍裁判）包小圓（外籍裁判）的關係。也就是說，雙臺籍裁判彼此之間的詞彙特徵有較大差異；外籍裁判分布較密集，彼此之間的詞彙特徵差異較小，而其位置幾乎都與雙臺籍裁判重合。綜上，透過資訊技術精密的詞彙計算方式，我們發現法官在跨國婚姻與本國婚姻的親權酌定事件，使用的詞彙基本上相同，且跨國婚姻裁判之間的

差異性比本國婚姻之間的差異更小。從而，本文認為法官在親權酌定的裁判中，並未對跨國婚姻或本國婚姻採取相異的論述方式，也就是從裁判書的文本詞彙並未發現因為當事人的國籍差異而用語不同的「不平等」跡象。

雖然整體而言外籍裁判與雙臺籍裁判並無用語差異，但吾人仍想了解外籍裁判的用語特徵。本文藉由梯度提升的機器學習演算法，要求電腦將外籍裁判與雙臺籍裁判分成兩類，並計算出造成此二類裁判差異的用語為何。發現結果是，相較於雙臺籍裁判，外籍裁判當中的「出境」、「到庭」、「租屋」、「同住」的 TF-IDF 值較高，也就是這些詞彙可說是外籍裁判的特徵。這凸顯了法官比較關心跨國婚姻的當事人的出境、到場之狀況，以及外籍配偶若要離婚，可能得在外租屋，才能跟子女同住，進而取得親權之現實。

綜上，在本研究所設定的定義與觀察方法之下，認為外籍配偶在我國司法體系中沒有遭受不平等對待。若欲保護教養未成年子女，外籍配偶可試著勇敢使用法律扶助以及法院。亦即外籍配偶爭取親權的必要條件是到場，以及租屋等有安定之居住環境。惟，在親屬支持系統較薄弱的現狀下，外籍配偶能否有資力、時間、精神出庭以及租屋，則有疑問，此超出司法系統之所能，須賴社福或其他團體的連結相關的資源，以協助外籍配偶在親權爭取時能獲得較佳的結果。現實中，法扶基金會提供的服務至關重要，本研究的 31 件獲扶助的外籍裁判樣本中，女性外籍配偶採取主動姿態（擔任原告）且獲得親權者共有 14 件，未獲得者僅 1 件；相較之下，男性我國籍配偶擔任原告且獲得親權者為 12 件，未獲得者亦為 1 件（參見下【表七】）。可見女性外籍配偶或許有資源上的弱勢，但在積極參與且武器平等的原則下，其獲得親權的比例並未低於我國籍男性配偶。



【表七】本研究的 31 件外籍裁判之原告與親權酌定之結果

原告／聲請人之性別與國籍	裁判總數	原告／聲請人獲得親權之裁判數
女性（我國籍）	4	4
男性（我國籍）	12	11
女性（外籍）	15	14

※ 資料來源：作者自製。

本文研究限制在於：多數親權裁判並未公開，以及裁判中資訊不足，致使本次使用法扶與司法院「交集」的親權酌定之公開裁判，僅有 116 筆，資料數較少。期待未來資訊更公開，筆數更多時，將能提供更有代表性的分析結果。其次，應再提升文字探勘中之分詞技術，自建專用的辭典；日後也可採取相似詞合併等方式，使研究結果更有意義。第三，其中一方為外籍配偶的裁判，與雙方皆為我國籍配偶的裁判書，即使用詞相近，是否就代表「平等」，這個定義仍可能被挑戰。本研究僅以裁判書為分析對象，例如新住民使用法律服務的便捷度、居留權的障礙、家庭支持系統的差異等裁判書記載的以外的事項，即無法觀察。在閱讀本研究的「平等」結論時，須留心這是某種限定素材、文脈之下的考察結果。

## 二、延伸討論與研究方法之意義

更深一步來說，「平等／不平等」本身是個哲學問題，很難有客觀定義。一般而言，我們不會用當事人的主觀想法來判定某個裁判是否違反了平等。若由研究者這樣的第三人，來認定某一個裁判「不平等」，雖較當事人來得客觀，但亦難免夾雜著研究者的價值判斷。其次是所謂的平等，到底是指涉「立足點的平等」抑或「結果平等」，本身也是一個哲學問題，每個人的看法都不同。因此，本文嘗試先給「平等」一個定義，也就是法院是否會因父母的國籍而使用不同的判準、語彙，論證的依據，是使用裁判書本身，用資訊技術的方法來檢驗裁判是否「有差異地」使用詞彙，來回答平等與否的問題。讀者並不一定要認同本文所給定的定義或使用的方法，但這個方法的特

徵是，不會因為閱讀裁判之人的價值觀，而導出不同的結果（平等或不平等的判斷），即使換了一台電腦，還是會得到相同的結論，這是與以往的內容分析法不同之處。

結合科技與法律的「法律資料分析」方法，在蒐集更多數裁判的前提下，期望做出更細緻的分析，提出具體建議。比如，從本文的分析，可得知「有沒有出現在法院、有沒有安定住處」會是爭取親權的考量；新住民如果失去本國籍配偶的社會支撐，更該好好照顧自己，而後進入司法系統，取得親權。這比起停留在評價或爭執這個司法體系「平等／不平等」，或是悲觀地認為新住民都無法爭取到親權，本文使用資訊方法，給出一個可能具體努力的方向：證明了司法體系對於「外籍事件」並無不平等，此外，應更致力於投入資源，提供法律扶助及居住地點確保之資源，促進新住民的實質權益。

最後，本研究採的方法在也有法學研究上之意義。一般認為，法院在為子女酌定親權人時，應以「子女之最佳利益」為審酌時之最高指導原則；由於「子女最佳利益」為不確定之法律概念，為了讓法院有所參考，故於民法第 1055 條之 1 明定具體客觀事由。也就是說，民法第 1055 條之 1 的各款事由例如子女之意願、父母之生活狀況、父母保護教養子女之態度等，皆為法官可審酌的事實，以資判斷如何符合子女最佳利益。以往的實證研究，皆是就第 1055 條之 1 的各款事由（或學說上認為重要的事項，例如主要照顧者、撫育環境、撫育時間、親職能力）加以編碼，希望得知法院比較重視哪一款事由<sup>65</sup>。惟這些事由雖然已較具體化了「子女最佳利益」的觀念，但有些仍有其抽象性。例如，法院會以什麼樣的事實認定父或母的撫育環境之好壞？以往的研究可能訴諸研究者之「專業判斷」，如看到「公寓」、「租屋」、

---

<sup>65</sup> 劉宏恩（2011），〈「子女最佳利益原則」在臺灣法院離婚後子女監護案件中之實踐：法律與社會研究（Law and Society Research）之觀點〉，《軍法專刊》，57卷1期，頁84-106；鄭諺霓（2015），〈離婚後未成年子女親權酌定之實證研究〉，國立臺灣大學法律學研究所碩士論文；曾鈞玟（2017），〈離婚事件中親權酌定之實務考察〉，《全國律師》，21卷10期，頁68-89；黃詩淳、邵軒磊（2018），〈酌定子女親權之重要因素：以決策樹方法分析相關裁判〉，《臺大法學論叢》，47卷1期，頁299-344；姜晴文，前揭註34；黃詩淳、邵軒磊，前揭註43，頁2023-2073。

「居住環境整潔」等描述，透過直覺，歸納為「撫育環境」之事由，但其判斷方式可能人言言殊，而且每件裁判中的具體事實過於繁複，千變萬化。

因此，本文的貢獻在於縮小「不確定法律概念」到「具體情狀」之間差距，採取文字探勘方法後，我們探索出更具體、特定的事實：「租屋」、「同住」對新住民親權酌定裁判有重要性；「租屋」與否可能是法官判斷撫育環境所憑藉的事實，「同住」與否可能是判斷繼續性原則所憑藉的事實。換言之，文字探勘與機器學習來分析裁判書文本，可對抽象的法律概念進行抽絲剝繭，化為具體、特定的事實，並告訴我們哪些事實在程序中發揮了作用。相較於過去係由專家的知識與經驗來判斷「何為具體情狀」、「此些情狀如何連結到不確定法律概念」，本文係交由機器處理而更加客觀，具有可重複性。研究者或實務專家未必能在理論中充分意識到這些，但透過機器學習發現的這些事實，可逐步具象化這些「不確定法律概念」，切合社會的需要。

## 參考文獻

### 一、中文部分

- 內政部統計處（2017），《外籍配偶歸化後婚姻狀況之研究》，載於：  
[https://www.moi.gov.tw/files/site\\_node\\_file/7471/外籍配偶歸化後婚姻狀況之研究.pdf](https://www.moi.gov.tw/files/site_node_file/7471/外籍配偶歸化後婚姻狀況之研究.pdf)。
- 王君琳（2005），〈性別與國族：從女性主義觀點解讀新移民女性現象〉，收於：夏曉娟（主編），《不要叫我外籍新娘》，頁 192-205，左岸文化。
- 王章逸、闕河嘉（2020），〈大埔之歌：臺灣主流報紙中的「土地徵收」〉，《資訊社會研究》，38 期，頁 19-50。  
[http://doi.org/10.29843/JCCIS.202001\\_\(38\).0004](http://doi.org/10.29843/JCCIS.202001_(38).0004)
- 王鵬翔、張永健（2015），〈經驗面向的規範意義：論實證研究在法學中的角色〉，《中研院法學期刊》，17 期，頁 205-294。
- 李震山（2004），〈憲法意義下之「家庭權」〉，《中正大學法學集刊》，16 期，頁 61-104。  
<http://doi.org/10.30094/NCCULJ.200407.0002>
- 林秀雄（2013），《親屬法講義》，3 版，元照。
- 林琬真、郭宗廷、張桐嘉、顏厥安、陳昭如、林守德（2012），〈利用機器學習於中文法律文件之標記、案件分類及量刑預測〉，《中文計算語言學期刊》，17 卷 4 期，頁 49-67。  
<https://doi.org/10.30019/IJCLCLP.2012.12.0004>
- 林勤富、劉漢威（2018），〈人工智慧法律議題初探〉，《月旦法學雜誌》，274 期，頁 195-215。  
<https://doi.org/10.3966/102559312018030274010>
- 法律扶助基金會（2018），《2017 年度報告書》，載於：  
<https://www.laf.org.tw/upload/2018/05/20180528114440.pdf>。

- 邵軒磊、吳國清（2019），〈法律資料分析與文字探勘：跨境毒品流動要素與結構研究〉，《問題與研究》，58 卷 2 期，頁 91-114。  
[https://doi.org/10.30390/ISC.201906\\_58\(2\).0003](https://doi.org/10.30390/ISC.201906_58(2).0003)
- 姜晴文（2019），《法律資料分析的優化與應用：以離婚後未成年子女親權酌定的裁判為素材》，國立臺灣大學法律學研究所碩士論文（未出版）。  
<https://doi.org/10.6342/NTU201904293>
- 施慧玲、陳竹上（2011），〈論婚姻移民家庭權之平等保障：全球法本土化的考察與反思〉，《月旦法學雜誌》，189 期，頁 22-37。
- 夏曉鶯（2002），《流離尋岸：資本國際化下的「外籍新娘」現象》，台灣社會研究雜誌社。
- 翁毓秀（2006），〈為新住民建構社會包容的家庭政策〉，《社區發展季刊》，114 期，頁 61-76。
- 張永健、李宗憲（2015），〈身體健康侵害慰撫金之實證研究：2008 年至 2012 年地方法院醫療糾紛與車禍案件〉，《臺大法學論叢》，44 卷 4 期，頁 1785-1843。  
<https://doi.org/10.6199/NTULJ.2015.44.04.03>
- Abdul Paliwala、Anne Elizabeth、許碧純（2011），〈跨國／跨文化婚姻移民的人權保障與移民政策：法律與社會政策的跨領域對話〉，《中正大學法學集刊》，34 期，頁 1-44。  
<https://doi.org/10.30094/NCCULJ.201107.0001>
- 教育部統計處（2019），《新住民子女就讀國中小人數分布概況統計：107 學年度》，載於：[http://stats.moe.gov.tw/files/analysis/son\\_of\\_foreign\\_107.pdf](http://stats.moe.gov.tw/files/analysis/son_of_foreign_107.pdf)。
- 郭文平（2020），〈語料庫輔助的媒體論述分析：以台灣平面媒體中國夢報導為語料的實證研究〉，《資訊社會研究》，38 期，頁 51-92。  
[https://doi.org/10.29843/JCCIS.202001\\_\(38\).0005](https://doi.org/10.29843/JCCIS.202001_(38).0005)
- 郭書琴（2007），〈逃家的妻子，缺席的被告？：外籍配偶與身分法之法律文化初探〉，《中正大學法學集刊》，22 期，頁 1-40。  
<https://doi.org/10.30094/NCCULJ.200705.0001>

- 陳世榮（2015），〈社會科學研究中的文字探勘應用：以文意為基礎的文件分類及其問題〉，《人文及社會科學集刊》，27 卷 4 期，頁 683-718。
- 陳竹上、曾竹寧（2011），〈法律知能與社會工作專業：跨界整合的探索及一位外籍配偶爭取監護權案例的啟發〉，《Asian Journal of Arts and Sciences》，2 卷 2 期，頁 162-184。  
<https://doi.org/10.30168/AJAS.201112.0004>
- 陳雪慧（2005），〈我們會是一家人?!〉，收於：夏曉娟（主編），《不要叫我外籍新娘》，頁 170-191，左岸文化。
- 曾元顯（2002），〈文件主題自動分類成效因素探討〉，《中國圖書館學會會報》，68 期，頁 62-83。
- 曾鈞玟（2017），〈離婚事件中親權酌定之實務考察〉，《全國律師》，21 卷 10 期，頁 68-89。
- 黃詩淳、邵軒磊（2018），〈酌定子女親權之重要因素：以決策樹方法分析相關裁判〉，《臺大法學論叢》，47 卷 1 期，頁 299-344。  
[https://doi.org/NTULJ.201803\\_47\(1\).0005](https://doi.org/NTULJ.201803_47(1).0005)
- （2019），〈人工智慧與法律資料分析之方法與應用：以單獨親權酌定裁判的預測模型為例〉，《臺大法學論叢》，48 卷 4 期，頁 2023-2073。  
[https://doi.org/10.6199/NTULJ.201912\\_48\(4\).0005](https://doi.org/10.6199/NTULJ.201912_48(4).0005)
- （2020），〈以人工智慧讀取親權酌定裁判文本：自然語言與文字探勘之實踐〉，《臺大法學論叢》，49 卷 1 期，頁 195-224。  
[https://doi.org/10.6199/NTULJ.202003\\_49\(1\).0004](https://doi.org/10.6199/NTULJ.202003_49(1).0004)
- 廖元豪（2006），〈全球化趨勢中婚姻移民之人權保障：全球化、台灣新國族主義、人權論述的關係〉，《思與言：人文與社會科學雜誌》，44 卷 3 期，頁 81-129。  
<https://doi.org/10.6431/TWJHSS.200609.0081>
- 劉宏恩（2011），〈「子女最佳利益原則」在臺灣法院離婚後子女監護案件中之實踐：法律與社會研究（Law and Society Research）之觀點〉，《軍法專刊》，57 卷 1 期，頁 84-106。

鄭彥寬 (2015), 《離婚後未成年子女親權酌定之實證研究》, 國立臺灣大學法律學院法律學研究所碩士論文 (未出版), 臺北。  
<https://doi.org/10.6342/NTU.2015.02749>

## 二、英文部分

Byrd, O. (2017, June 12). *Legal Analytics vs. Legal Research: What's the Difference?* [Online forum comment]. Law Technology Today. <https://www.lawtechnologytoday.org/2017/06/legal-analytics-vs-legal-research/>

Chen, C.-J. (2016). The Chorus of Formal Equality: Feminist Custody Law Reform and Fathers' Rights Advocacy in Taiwan. *Canadian Journal of Women and the Law*, 28(1), 116-151. <https://doi.org/10.3138/cjwl.28.1.116>

Feldman, R., & Sanger, J. (2006). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511546914>

Freeman, C. (2005). Marrying Up and Marrying Down: The Paradoxes of Marital Mobility for Chosŏnjok Brides in South Korea. In N. Constable (Ed.), *Cross-Border Marriages: Gender and Mobility in Transnational Asia*, (pp. 80-100.) University of Pennsylvania Press. <https://doi.org/10.9783/9780812200645.80>

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189-1232. <https://doi.org/10.1214/aos/1013203451>

Hu, X., & Liu, H. (2012). Text Analytics in Social Media. In C. C. Aggarwal & C. Zhai (Eds.), *Mining Text Data* (pp. 385-414). Springer. [https://doi.org/10.1007/978-1-4614-3223-4\\_12](https://doi.org/10.1007/978-1-4614-3223-4_12)

Jolliffe, I. T. (1986). *Principal Component Analysis*. Springer.

- Jones, K. S. (1972). A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*, 28(1), 11-21. <https://doi.org/10.1108/00220410410560573>
- Katz, D. M. (2013). Quantitative Legal Prediction - or - How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry. *Emory Law Journal*, 62(4), 909-966.
- Lan, P.-C. (2008). Migrant Women's Bodies as Boundary Markers: Reproductive Crisis and Sexual Control in the Ethnic Frontiers of Taiwan. *Signs: Journal of Women in Culture and Society*, 33(4), 833-861. <https://doi.org/10.1086/528876>
- McGinnis, J. O., & Pearce, R. G. (2014). The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services. *Fordham Law Review*, 82(6), 3041-3066.
- Salton, G., & Buckley, C. (1988). Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing and Management*, 24(5), 513-523. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)
- Sobowale, J. (2016, April 1). *How Artificial Intelligence is Transforming the Legal Profession*. ABA Journal. [http://www.abajournal.com/magazine/article/how\\_artificial\\_intelligence\\_is\\_transforming\\_the\\_legal\\_profession](http://www.abajournal.com/magazine/article/how_artificial_intelligence_is_transforming_the_legal_profession)
- Sullivan, D. (2001). *Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing, and Sales*. John Wiley & Sons.
- Surden, H. (2014). Machine Learning and Law. *Washington Law Review*, 89(1), 87-116.
- Wu, H. C., Luk, R. W. P., Wong, K. F., & Kwok, K. L. (2008). Interpreting TF-IDF Term Weights as Making Relevance Decisions. *ACM Transactions on Information Systems (TOIS)*, 26(3), 13:1-13:37. <https://doi.org/10.1145/1361684.1361686>



## Applying Texts Mining on Child Custody Cases Regarding New Immigrants in Taiwan: An Empirical Study on Equality

*Hsuan-Lei Shao*<sup>\*</sup> & *Sieh-Chuen Huang*<sup>\*\*</sup>

### Abstract

There are more and more transnational (international) marriage cases in Taiwan today. As an ethnic group, the New Immigrants are facing multiple issues regarding equality/ inequality in daily life. This article attempts to find if a foreign spouse is treated equally in child custody proceedings in Taiwan. In detail, we studied 116 real legal texts (judgements) by Natural Language Processing (NLP) and Machine Learning (ML) in which 31 foreign cases are included. Firstly, after word segmentation, the document-term-matrix of TF-IDF is produced and then put into principle component analysis so that a two-dimensional plot is formed. It is thus found that 85 domestic marriage cases and 31 international marriage cases are almost located at the same place, meaning that the court does not use different vocabularies between these two types of cases. It indicates that the family judges did not treat or judge the New Immigrants differently from domestic couples. Secondly, using gradient boosting of machine learning, it is found that the vocabulary characteristics of cases involving New Immigrants are: “whether one party appears at court”, “renting a house”, and “living together”, indicating that judges are more concerned about the participation of procedures and residence of New Immigrant parents. If a foreign spouse obtains legal aid and participates in the judicial process, since the standard adopted by the judge on the child custody is the same on both New Immigrant cases and domestic marriage cases, it is very possible that the foreign spouse will receive custody. However, the foreign spouse may

---

<sup>\*</sup> Associate Professor, Department of East Asian Studies, National Taiwan Normal University. E-mail: hlshao@ntnu.edu.tw

<sup>\*\*</sup> Professor, College of Law, National Taiwan University. E-mail: schhuang@ntu.edu.tw

need to acquire other social support on housing as it is a factor that family judges tend to investigate in New Immigrant cases. In this paper, the authors adopt disciplinarity of information science and law, showing one of the possibilities in which technology helps us to deepen the discussion of social responsibility.

**Keywords: Legal Analysis, Natural Language Processing (NLP), TF-IDF, Equality, New Immigrants, Empirical Legal Studies, Principle Component Analysis**