

中國外交部用詞分析：以耿爽為例
以及與臺灣外交部用詞之比較

R 語言與資料科學導論 期末專案 G06

陳柔安 B06302345

陳敬宜 B06102106

呂曉鈞 B06102086

黃山 B03611008

一、主題

耿爽為現任中國外交部發言人之一，因為其用詞往往重複性高，且有許多特殊的疊字語，以及強勢、咄咄逼人的形象深植人心，因此被網友整理出了常用字，並寫出了一個專屬於其的發言模擬器，只要輸入文字，便可獲得被耿爽「譴責」的文章。我們在思考期末專案的主題時，恰好在無意間在網路看到了這個中國網友製作的「耿爽模擬器」，且覺得這個現象非常有趣，因此決定以此為發想，將專案主軸定為「分析耿爽之言語」，來一探究竟這位佔盡版面的發言人是否真有其特定、固定之使用辭彙？

此外，在分析中國外交部發言人的資料時，我們亦瀏覽了一些台灣外交部的對外聲明做為對照，一併將其納入分析，希望能藉此觀察兩國對外聲明的差異之處，以及探討背後可能的原因。

二、假設

我們認為耿爽的特別形象之所以深植人心，是因為談吐中有許多「強烈形容詞」配上「動詞」的言語，例如：強烈譴責、堅決反對、堅定維護等。基於以上的觀察，我們假設，在他的發言（新聞聲明、例行記者會對談）中，這樣的詞語會佔大多數比例。

在與臺灣外交部的用詞比較上，我們認為也許台灣外交部的用詞會較為平和且中性，避免與中國當局正面起衝突。

三、資料取得

耿爽的文字資料取自中華人民共和國外交部網站（將另外兩位發言人剔除）（<https://www.fmprc.gov.cn/web>），使用了從其 2016 年 8 月上任後至今的對外發言資料。

另一方面，臺灣的資料則是從中華民國外交部網站（<https://www.mofa.gov.tw>）取得。

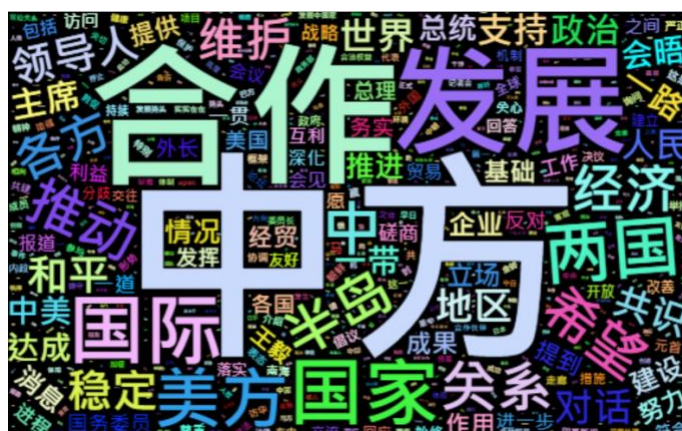
四、分析方式

將資料爬下來後，我們將取得的耿爽資料做初步的清洗：

1. 排除另外兩位發言人的資料
2. 將記者會中記者提問的部分刪除
3. 設定 stopwords（我、你、了、在等）

另外，我們也收集了一系列常用字，建立了辭典，避免在稍後使用 jieba 套件斷詞時，將重要的詞語截斷（例：「臺灣地區」不可被斷成「臺灣」、「地區」）。

取得斷詞結果，並取得詞頻後，我們將獲得的資料以時間軸做區分，2016 年 8 月至 2019 年 12 月共分為四個階段，並以文字雲（wordcloud 套件）呈現其字頻。在分析中，為了查看四字字詞是否真的有較高的出現頻率，我們也使用 stringr 套件，將耿爽的資料以兩字字詞、四字以上字詞來查看（如圖）。



而在臺灣外交部方面，因為自 1984 年以來，僅發佈過 95 則聲明，但中國方面單單耿爽上任後已經有 900 多篇聲明，因此將 95 則聲明皆納入本次分析的資料中，並未僅抓取 2016（耿爽上任）年以後的文章。以下為本次分析結果臺灣外交部的字頻呈現（如圖）



（未將我國、中華民國等字詞列為 stopwords 為本次之疏忽）



五、結果

根據分析結果，在耿爽發言方面，四個階段的字頻呈現上均沒有出現我們預期的「常用情緒性四字詞語」，反而是較中性的四字詞語（例：相互尊重、共同努力、友好合作）等較常出現。我們不禁想，那是什麼原因造成耿爽的發言令人感到壓迫、強勢、且具有攻擊性呢？

對此，我們認為，一連串的四字詞堆疊可能是原因之一，在排比式的句型當中，很容易產生韻律感，再搭配上堅定且強勢的表情、肢體、語氣，會

使聞者感受到發言者的強勢與不退讓感。在展覽當中，來自中國同學也表示，在其政治學課程上，的確要背誦許多四字用詞，以便在政治、外交上達到強勢與據理力爭的優勢。

在臺灣外交部方面上，用詞的確與我們的假設相距不遠，雖然偶有情緒性四字詞語（例：嚴正聲明、深表遺憾等），字頻最高者皆較為中性且溫和的詞語。在發言的內容上，相較於中國外交部，臺灣外交部的內容較為限縮在與中國、鄰近國家與邦交國家之議題上。此特性或許與臺灣在國際上因為中國打壓而不利的外交地位有關。在四字詞語的文字雲中，「自即日起」為最高頻的字詞，也許因為數年來與他國斷交的數量漸多，該內容多為官方宣布與他國斷交有關。

最後，在此次的專案當中，我們雖然初步分析了兩者外交部的用詞差異，但此專案還有許多能夠更深地發展的方向，例如：兩者的情緒用詞正負向差異、在特定議題上的用詞差異等，都是非常值得探討的面向。