國家教育研究院
National Academy for Educational Research

2022華語文教育課程指引與語料庫應用工作坊

語言分析應用實作

李詩敏 國家教育研究院語文教育及編譯研究中心

2022.05.14

分析面向

詞彙語意&句法分析

語體差異&語言變遷

國教院語料庫索引典系統(含國教院華語中介語索引典系統)

目前使用者: naer [登出]

Corpora available on this server (click here to view your own corpus access privileges)

中研院平衡語料庫4.0	COCT 口語語料庫2016	COCT 口語語料庫2017
COCT 口語語料庫2019	COCT 中介語語料庫2016	COCT 中介語語料庫2019
COCT 口語語料庫2021	COCT 書面語語料庫2021	COCT 書面語語料庫2016
COCT 書面語語料庫2017	COCT 書面語語料庫2019	書面語語料庫(工作坊專用)
COCT 書面語語料庫2015		

詞彙語意&句法分析

明確詞彙：「經過」

1 English/Chinese

選單

中研院平衡語料庫4.0

2 語料庫查詢

標準查詢

限制查詢

單詞查詢

詞頻列表

關鍵詞

分析語料庫

儲存查詢資料

查詢歷史

儲存查詢結果

分類查詢結果

標準查詢

3

經過

查詢模式: 簡易查詢（不區分大小寫） 指令速查表 選單參考表

每頁的查詢結果: 50

限制（檢索範圍）: 檢索整個語料庫）

4

開始查詢 重設查詢

您的查詢 "經過" 回傳 2,850 筆， 分布在2,013 相異文本中 (本語料庫總共 11,245,627 詞 [19,247 文本]; 頻率: 253.43 次/每百萬詞) [.726 秒 - 從快取中提取]

|< << >> >| 頁碼: 1 Line View 隨機排序

新查詢 Go!

- 新查詢
- 查詢結果隨機取樣...
- 頻率分解
- 詞彙分布
- 排序
- 搭配詞...
- 下載...
- 分類...
- 儲存目前的查詢結果...

序號	檔名	第 1 - 50 筆 第 1 頁/總共 57 頁		
1	100007	。中研院正如台灣的其他學術研究機構一樣，	經過	長期的努力後，眼看到我們幾乎就要追上開發
2	100007	問題。目前中研院面對的最嚴重的問題是，	經過	
3	100007	優雅、舒適、健康而美好的地方，就必須	經過	
4	100007	個所移轉到院本部。理想的做法應該是，	經過	
5	100007	中研院這學人匯集的地方，在很多事務上，	經過	
6	100007	或從更進一步的探究，我們應該能提供一些	經過	客觀分析的一些資料與看法，當做政府決策的
7	100007	有些問題，如果某一個研究所的專家們	經過	深入的討論之後，在研究所裡面建立了很強
8	100022	坑洞，凹凸不平，塵土厚積，天晴時若汽車	經過	則滿天灰塵；天雨時則污泥四濺，
9	100030	，推動生物技術在水產養殖上之應用研究，	經過	研究人員多年的辛勞研究開發，目前獲致之
10	100042	調查的核心項目在四國（地）實施之前均	經過	長時間的討論和決定，將來在進行比較研究

。_PERIODCATEGORY 中研院_Nc 正_D 如_P 台灣_Nc 的_DE 其他_Neqa 學術_Na 研究_Nv 機構_Na 一樣_VH ，_COMMACATEGORY 經過_P 長期_Nd 的_DE 努力_Nv 後_Ng ，_COMMACATEGORY 眼看到_VE 我們_Nh 幾乎_Da 就_D 要_D 追上_VC 開發_Nv

頻率分解

Your query "經過" returned 2,850 matches in 2,013 different texts (in 11,245,627 words [19,247 texts]; frequency: 253.43 instances per million words)

Showing frequency breakdown of words in this query, at the query node; there is 1 different type and 2,850 tokens at this concordance position.

|< << >> >| Breakdown position: Node Frequency breakdown of words only Go!

No.	Search result	No. of occurrences	Percent
1	經過	2850	100%

Frequency breakdown of words and annotation

|< << >> >| Breakdown position: Node Show hits sorted by node Go!

- Frequency breakdown of words only
- Frequency breakdown of annotation only
- Frequency breakdown of words and annotation
- Download frequency breakdown table (for words)
- Show hits sorted by node
- New query

No.	Search result	Percent
1	經過	100%

Your query "經過" returned 2,850 matches in 2,013 different texts (in 11,245,627 words [19,247 texts]; frequency: 253.43 instances per million words)

Showing frequency breakdown of both words and annotation in this query, at the query node; there are 5 different types and 2,850 tokens at this concordance position.

|< << >> >| Breakdown position: Node Frequency breakdown of words and annotation Go!

No.	Search result	No. of occurrences	Percent
1	經過_VCL	2208	77.47%
2	經過_P	511	17.93%
3	經過_Na	112	3.93%
4	經過_Nv	16	0.56%
5	經過_VC	3	0.11%

明確詞彙或詞性的查詢方式

明確詞彙：「經過」

- 標準查詢—簡易查詢
  - 經過
- 標準查詢—CQP語法
  - [word="經過"]
- 單詞查詢—完全符合
  - 經過

明確詞性：VCL

- 標準查詢—簡易查詢
  - _VCL
- 標準查詢—CQP語法
  - [pos="VCL"]
- 詞類列表—Part-of-speech tag—完全符合
  - VCL

限制詞頭或詞尾：作家、音樂家、語言學家

- 二字詞且詞尾為家，如：作家、專家、史家
  - 簡易查詢： ?家
  - CQP語法：[word=".家"]
- 詞尾為家的詞，且包括家，如：家、作家、藝術家、物理學家
  - 簡易查詢： *家
  - CQP語法： [word=".*家"]
- 詞尾為家的詞，且不包括家，如：作家、藝術家、物理學家
  - 簡易查詢： +家
  - CQP語法： [word=".+家"]
- 單詞查詢─結束於

Q 詞首是「老」的詞

- 二字詞，如：老師、老人、老闆
  - 老?
  - [word="老."]
- 詞的字數不限，如：老、老師、老人家
  - 老*
  - [word="老.*"]
- 二字詞以上，如：老師、老人家、老百姓
  - 老+
  - [word="老.+"]
- 單詞查詢—始於

限制條件：詞彙&詞性

- 「老」的詞性

1	老_VH	1946	90.93%
2	老_D	119	5.56%
3	老_Na	65	3.04%
4	老_A	4	0.19%
5	老_Nh	3	0.14%
6	老_Nv	2	0.09%
7	老_VC	1	0.05%

- 直接查詢「老」：VH
  - 老_VH
  - [word="老" & pos="VH"]
- 直接查詢「老」：動詞
  - 老_V*
  - [word="老" & pos="V.*"]

No.	Search result	No. of occurrences	Percent
1	老_VH	1946	100%

No.	Search result	No. of occurrences	Percent
1	老_VH	1946	99.95%
2	老_VC	1	0.05%

多重限制

1	老_VH	1946	90.93%
2	老_D	119	5.56%
3	老_Na	65	3.04%
4	老_A	4	0.19%
5	老_Nh	3	0.14%
6	老_Nv	2	0.09%
7	老_VC	1	0.05%

- **「老」：動詞或名詞**
  - (老_V*|老_N*)
  - [word="老" & pos="V.*"][word="老" & pos="N.*"]
  - [word="老" & pos="V.*|N.*"]
- **「老」：不是非謂形容詞（A）**
  - (老_V*|老_N*|老_D)
  - [word="老" & pos="V.*|N.*|D"]
  - [word="老" & pos!="A"]
  - [word="老" & !(pos="A")]

若要排除，僅能採用CQP語法

Q1 「把」：量詞（Nf）

- 把_Nf
- [word="把" & pos="Nf"]

Q2 「把」：名詞

- 把_N*
- [word="把" & pos="N.*"]
  - ➔ Nf, Na

Q3 「把」：不是動詞

- [word="把" & pos!="V.*"]
- [word="把" & !(pos="V.*")]
  - ➔ P, Nf, Na

Q4 「把」：不是動詞，也不是名詞

- [word="把" & pos!="V.*|N.*"]
- [word="把" & !(pos="V.*|N.*")]
  - ➔ P

「打」的搭配詞

搭配詞

選單

語料庫查詢

標準查詢
限制查詢
單詞查詢
詞頻列表
關鍵詞
分析語料庫

儲存查詢資料

查詢歷史
儲存查詢結果
分類查詢結果

中研院平衡語料庫4.0

標準查詢

打

查詢模式： 簡易查詢（不區分大小寫） 指令速查表 選單參考表

每頁的查詢結果： 50

限制（檢索範圍）： 無（檢索整個語料庫）

開始查詢 重設查詢

您的查詢 "打" 回傳 2,695 筆， 分布在1,482 相異文本中 (本語料庫總共 11,245,627 詞 [19,247 文本]; 頻率: 239.65 次/每百萬詞) [0.72 秒 - 從快取中提取]

|< << >> >| 頁碼: 1 Line View 隨機排序 新查詢 Go!

新查詢
查詢結果隨機取樣...
頻率分解
詞彙分布
排序
搭配詞...
下載...
分類...
儲存目前的查詢結果...

序號	檔名	第 1 - 50 筆 第 1 頁/總共 54 頁		
1	100195	用完 為止 。 「 九 條 好漢 在 一 班 ， 說	打	就 打 ， 說 幹 就 幹 ， 管 它 流血 流
2	100195	。 「 九 條 好漢 在 一 班 ， 說 打 就	打	， 說 幹 就 幹 ， 管 它 流血 流汗 ！
3	100351	用完 為止 。 「 九 條 好漢 在 一 班 ， 說	打	就 打 ， 說 幹 就 幹 ， 管 它 流血 流
4	100351	。 「 九 條 好漢 在 一 班 ， 說 打 就	打	， 說 幹 就 幹 ， 管 它 流血 流汗 ！ 」
5	101673	爸爸 愛 看 報紙 ， 媽媽 愛 喝 茶 ， 哥哥 喜歡	打	躲避球 ， 我 喜歡 下 跳棋 。 爸爸 點頭 說 ： 文句
6	101673	一樣 ， 最 喜歡 一邊 喝 茶 ， 一邊 說話 。 哥哥	打	躲避球 ， 勝 了 很 高興 ， 敗 了 就 一聲不響 。
7	100486	我 趕 著 似的 走 在 前面 ； 好像 是 一 個	打	了 敗仗 的 士兵 ， 拖 著 一 把 生 了 鏽
8	100500	批閱 每 天 疾馳 送來 的 奏章 文書 。 康熙 一生 身先士卒	打	過 許多 著名 的 仗 ， 但 在 晚年 ， 他 最
9	100513	對方 的 表情 ， 這 多麼 悲哀 啊 ！ 我 想 先	打	聲 招呼 卻 又 不 知 如何 開口 ， 真 複雜 的
10	100528	我 小時候 很 調皮 ， 有 一 回 在 木頭 地板 上	打	了 十几 個 蛋 ， 在 上面 溜冰 。 當時 雞蛋 可

Choose settings for proximity-based collocations:

Include annotation:	Feature tag	○ Include ● Exclude
	Part-of-speech tag	● Include ○ Exclude
Maximum window span:	+ / - 5	

Create collocation database

Collocation controls

Collocation based on:	Word form	Statistic:	Log-likelihood
Collocation window *from*:	3 to the Left	Collocation window *to*:	3 to the Right
Freq(node, collocate) at least:	5	Freq(collocate) at least:	5

A-B-C-打-X-Y-Z

No.	Word	Total no. in whole corpus	Expected collocate frequency	Observed collocate frequency	In no. of texts	Log-likelihood
1	電話	1,883	2.708	97	75	511.034
2	得	14,357	20.644	173	155	433.91
3	了	81,367	116.997	390	292	398.707
4	籃球	253	0.364	47	30	373.018
5	打	2,695	3.875	86	31	371.862
6	招呼	125	0.18	37	32	332.792
7	折	188	0.27	37	24	298.348
8	高爾夫球	65	0.094	27	26	265.301
9	場	3,606	5.185	62	50	195.161
10	著	23,227	33.398	140	119	189.275
11	過	10,432	15	92	74	180.666
12	個	67,430	96.957	250	183	169.322
13	被	16,272	23.397	105	82	152.899
14	零工	18	0.026	13	12	148.895
15	武松	30	0.043	14	7	141.875

「被+X+Y+打」或「打+X+Y+被」

13	被	16,272	23.397	105	82	152.899
14	零工	18	0.026	13	12	148.895
15	武松	30	0.043	14	7	141.875

序號	檔名	第 1 - 50 筆　第 1 頁/總共 3 頁		
1	100576	神力 ， 憤 而 提議 以 兵器 比武 ， 不料 **被** 元霸	打	得 口 吐 鮮血 差點 送命 ， 幸 其 父 請 煬帝
2	100990	我們 覺得 愉快 。 但是 半夜 的 時候 一 條 狗 **被**	打	了 ， 發出 怪 聲 ， 那 就 不 一樣 ， 那
3	101219	， 可是 他 從未 想 過 法律 應當 保護 人民 不 **被**	打	。 作為 共產黨 的 主席 ， 明知 黨員 幹部 以 老爺 對
4	101222	叫 營長 的 弟弟 ， 一 天 沒 幹活 ， 去 食堂	打	飯 ， **被** 哥哥 看見 揍 了 一 頓 。 當時
5	102530	， 陳某 胡說八道 ， 並 無 挨打 情事 ， 只有 幹員 **被**	打	。 兩 方 說詞 不一 ， 事態 已 有 擴大 之 勢
6	102600	痊癒 證明書 ， 請求 南投縣 政府 准予 提前 複檢 ， 但 **被**	打	了 回票 ， 蕭雨郎 在 僵硬 的 法令 規定 下 遂 被
7	105437	行得通 ， 兒子 對 他 說 ， 六十 分 得 **被** 老師	打	好多 下 ！ 為 了 孩子 ， 三 年 前 他 就
8	105439	， 和 群眾 打成 一 團 ， 執事 人員 **被** 罵 被	打	， 辦公室 被 砸 等 風風雨雨 的 過程 。 然而 在 黃種煌
9	105439	， 和 群眾 打成 一 團 ， 執事 人員 被 罵 **被**	打	， 辦公室 被 砸 等 風風雨雨 的 過程 。 然而 在 黃種煌
10	105439	， 和 群眾 打成 一 團 ， 執事 人員 被 罵 被	打	， 辦公室 **被** 砸 等 風風雨雨 的 過程 。 然而 在 黃種煌

Within the window -3 to 3, *被* occurs 105 times in 82 different texts (expected frequency: 23.397)

被打

Distance	No. of occurrences	In no. of texts	Percent
-3	6	6	5.7%
-2	20	20	19%
-1	70	57	66.7%
1	3	3	2.9%
2	1	1	1%
3	5	5	4.8%

「打」+受詞

Collocation controls

Collocation based on:	Word form	Statistic:	Log-likelihood
Collocation window *from*:	1 to the Right	Collocation window *to*:	5 to the Right
Freq(node, collocate) at least:	5	Freq(collocate) at least:	5
Filter results by:	specific collocate:	and/or tag: (none)	Submit changed parameters — Go!

1 2 3

Extra information: **Log-likelihood** scores collocations by significance: the higher the score, the more evidence you have that the association is not due to chance. More frequent words tend to get higher log-likelihood scores, because there is more evidence for such words.

There are 5,872 different words in your collocation database for "[word="打"%c]}". (Your query "打" returned 2,695 matches in 1,482 different texts) [0.284 seconds - retrieved from cache]

No.	Word	Total no. in whole corpus	Expected collocate frequency	Observed collocate frequency	In no. of texts	Log-likelihood
1	電話	1,883	2.256	107	78	623.123
2	了	81,367	97.498	397	294	523.727
3	打	2,695	3.229	97	63	476.493
4	得	14,357	17.203	165	147	453.647
5	籃球	253	0.303	46	30	379.598
6	招呼	125	0.15	36	31	334.549
7	折	188	0.225	37	24	311.768

Collocation contro

Collocation based on:	Word form	Statistic:	Log-likelihood
Collocation window *from*:	1 to the Right	Collocation window *to*:	5 to the Right
Freq(node, collocate) at least:	5	Freq(collocate) at least	5
Filter results by:	specific collocate:	and/or tag: Na — Na	Submit changed parameters — Go!

4 5

Extra information: **Log-likelihood** scores collocations by significance: the higher the score, the more evidence you have that the association is not due to chance. More frequent words tend to get higher log-likelihood scores, because there is more evidence for such words.

There are 5,872 different words in your collocation database for "[word="打"%c]}". (Your query "打" returned 2,695 matches in 1,482 different texts) [0.263 seconds - retrieved from cache]

No.	Word	Total no. in whole corpus	Expected collocate frequency	Observed collocate frequency	In no. of texts	Log-likelihood
1	電話	1,883	2.256	107	78	623.123
2	籃球	253	0.303	46	30	379.598
3	高爾夫球	65	0.078	26	24	262.441
4	仗	95	0.114	20	19	171.513
5	零工	18	0.022	14	13	169.267
6	主意	132	0.158	17	13	127.609
7	折扣	185	0.222	18	18	124.52

近義詞：「切」、「割」

簡易查詢

- 切_V*
- 割_V*

CQP

- [word="切" & pos="V.*"]
- [word="割" & pos="V.*"]

搭配詞

- 5 to the left, 5 to the right

- **切**

物體：菜，蔥，蛋糕，肉
結果：塊，絲，片
工具：刀

菜刀、美工刀、手術刀、鐮刀

- **割**

物體：雙眼皮，草
結果：得零零碎碎，割得很漂亮
工具：刀

割肉：力道，方向

Q 「製造」、「生產」

- **受詞—成品**
  - 「製造」+髒亂、垃圾、問題、歡樂
  - 「生產」+主機板、藥品
- **和後面名詞的關係**
  - **「生產」較常做定語，修飾名詞**
    - 「生產」+毛額、成本、技術、效率、過程、方式、規模
    - 「製造」+日期、過程

字詞形

- 依循傳統，義民廟在農曆7月最後一天舉辦**普度**《聯合報》
- 農曆7月民間有**普渡**拜拜的習俗，是對歷代祖先的無限追思《中國時報》
- (普渡|普度)
- [word="普渡|普度"]
- 普[渡,度]
- [word="普[渡度]"]
- 普渡：86.96%；普度：13.04%
  - 中元+普渡 / 普度
  - 普渡 / 普度+眾生
  - 普渡 / 普度+大學

Q 臺北市 / 台北市

- (臺北市|台北市)
- [word="臺北市|台北市"]
- [臺北,台北]市
- ! [word="[臺北台北]市"]
- [word="(臺北|台北)市"]

Q「掛勾」、「掛鉤」

- 財團與利益～
- 購買無痕～

- (掛勾|掛鉤)
- [word="掛勾|掛鉤"]
- 掛[勾,鉤]
- [word="掛[勾鉤]"]

- **掛勾**：掛鉤
  - 平衡語料庫 ─ 54.41% ：45.59%
  - 書面語2015 ─ 57.02%：42.98%
  - 書面語2019 ─ 56.51%：43.49%

- 不同詞類的分布情形
  - 平衡語料庫 ─ 掛勾(V)35、(N)2；掛鉤(V)19、(N)12
  - 書面語2015 ─ 掛勾(V)50、(N)2；掛鉤(N)39、(V)30
  - 書面語2019 ─ 掛勾(V)225、(N)32；掛鉤(N)193、(V)141

重疊：A not A，AABB

僅能採用CQP語法

- **A not A：「是不是」，「對不對」，「好不好」**
  - a:[] [word="不"] b:[] :: a.word=b.word
  - 說明
    - a:[] ➔ 第1個詞命名為變數 a
    - b:[] ➔ 第3個詞命名為變數 b
    - :: a.word=b.word ➔ 限制a和b相同
- **AABB：「清清白白」，「堂堂正正」，「乾乾淨淨」**
  - 舊方法：[word="...." & char(word, 0)=char(word, 1) & char(word,2)=char(word, 3)]
    ➔ 3,246筆
  - 新功能：[word=@AABB]
    ➔ 3,160筆

新功能：
- [word=@AABB]
- [word=@AAB]
- [word=@ABAB]
- [word=@ABB]

詞表比較工具

離合詞的中插成分

- 「結婚」
  - [word="結"][][word="婚"]
    - 結了婚，結過婚，結次婚，結個婚，結的婚，結完婚
  - [word="結"][]{1,5}[word="婚"]
    - 結過一次婚，結幾次婚，結過兩次婚
  - [word="結"][]+[word="婚"]
    - 正是她婚姻中不可 解之結。我永遠.....自動刪節了部分當時還十分敏感的婚外情欲等部分。
  - 查詢結果不要夾有標點符號
    - [word="結"][word!="。|，|!|？"]+[word="婚"]
    - [word="結"][pos!="PUNC"]+[word="婚"]

平衡語料庫標點符號的詞性標記和國教院語料庫不同

Q「睡覺」的中插成分

- [word="睡"][]{1,10}[word="覺"]
- [word="睡"][]+[word="覺"]
  - 睡，一覺
  - 睡不著了。李文秀這一覺

- **刪除夾帶標點符號**
  - [word="睡"][word!="，|；|。|!|？"]+[word="覺"]
    - 睡了一個很久以來難得的無夢的好覺

動補結構

- **X死**
  - ?死_V*
  - [word=".死" & pos="V.*"]
  - +死_V*
  - [word=".+死" & pos="V.*"]
- **X掉**
  - +掉_V*
  - [word=".+掉" & pos="V.*"]
  - _V* 掉
  - [pos="V.*"][word="掉"]

➔ 殺死，餓死，累死，……

➔ 殺死，餓死，累死，折磨死，……

➔ 330 types：吃掉，拿掉，犧牲掉，處理掉，……

➔ 41 types：革 掉，剝 掉，擦拭 掉，排除 掉，……

中動結構/中間結構 (middle construction)

- **中動結構：好X，難X**
  - **這枝筆很好寫**
  - (好?_V*|難?_V*)
  - [word="好.|難." & pos="V.*"]
  - 限制第一個字不等於第二個字：[word="好.|難." & pos="V.*" & char(word,0)!=char(word,1)]

➔ 好吃，難聽，好賺，難學，好好……

語法點：一……就……

簡易查詢	意義	符合之搜尋結果
一 (+)* 就	「一」和「就」中間夾很多個詞	她一走出門妝就變花了
一 (+){2,5} 就	「一」和「就」中間夾2~5個詞	它一動，就啄它的眼睛 一有閒暇她就想到外面
一 <<3>> 就	「一」「就」中間最多3個詞， 「一」「就」順序不定	火炭一聽就火了 我走了就當一場夢
一 >>3>> 就	「一」「就」中間最多3個詞， 「一」「就」順序固定	火炭一聽就火了 他一開始就用這種態度

*一 股 難忘 的 溫馨 。……。 我 就 以 「 中央 研究院 未來 的 展望 」 為 題目…

- 限制標點符號：[word="一"][word!="。|？|!"][word="就"] → 一年就，一輩子就
- 限制詞性：[word="一" & pos="D"][word!="。|？|!"][word="就"]

Q1 一邊……一邊……

- 一邊
- [word="一邊"]
- [word="一邊"][word!="。|？|！"][word="一邊"]

搭配詞　一邊

- 一邊做飯一邊唱歌，一邊煮一邊攪拌，……

Q2 被……V

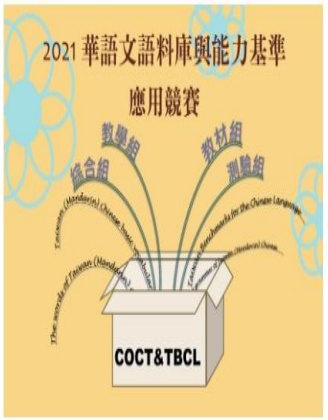
- [word="被"][word!="。|？|！"]{0,5}[pos=".V*"]
  - 被佔住了，被上帝眷顧，被漢人所同化，被畫在畫布，……

語體差異&語言變遷

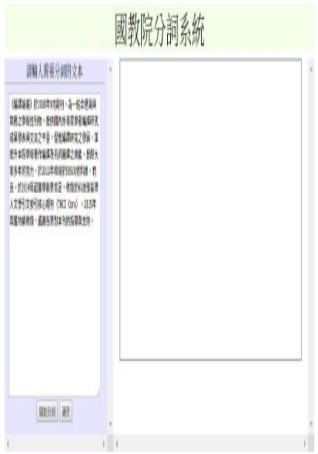
語料庫覆蓋率統計系統

歷年工作坊講義及錄影

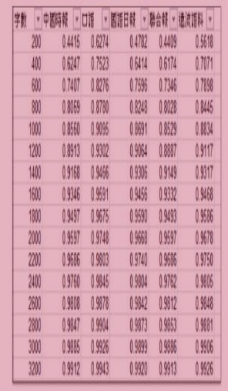
華語文語料庫與能力基準應用競賽



國教院分詞系統



語料庫覆蓋率統計系統



1

2

3

4 計算完成！字表下載

中國時報、口語語料、國語日報、
聯合報、遠流語料、教材語料

序號	漢字	序位	字頻	累積字頻	覆蓋率
1	你	681	349346	349346	0.000341
2	我	127	1754367	2103713	0.002051
3	他	24	3677400	5781113	0.005636
4	一	2	10432208	16213321	0.015806
5	琵	2810	12356	16225677	0.015818
6	琶	3918	2721	16228398	0.015821
7	葡	2225	28293	16256691	0.015848
8	萄	2286	26442	16283133	0.015874

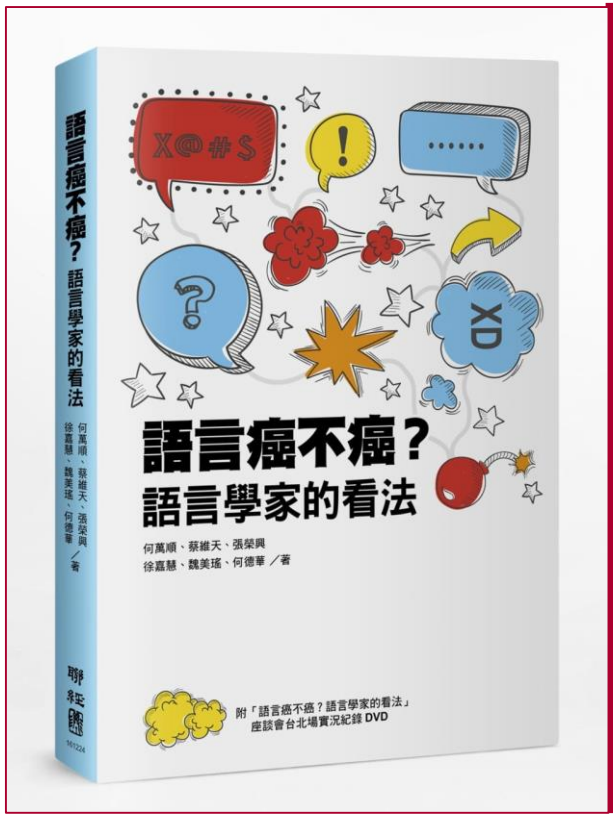
不同類型語料，代表的意義

	你	我	他	的	嗎
中國時報序位	681	127	24	1	1207
口語語料序位	20	3	21	1	164
國語日報序位	501	20	44	1	1102
聯合報序位	549	60	18	1	1062
遠流語料序位	30	5	10	1	351
教材語料序位	17	6	10	1	80

	記者	今天	學生	上人	說	書寫	課
中國時報序位	372	472	116	40934	20	5457	2385
口語語料序位	1412	119	365	519	21	5921	1136
國語日報序位	1616	367	7	87135	12	2085	608
平衡語料庫序位	788	256	112	53591	26	2892	1426
聯合報序位	619	271	69	46883	9	5108	1772
遠流語料序位	1695	384	256	43261	17	3038	1810
教材語料序位	3078	178	269	20846	19	3316	880

書面語 vs 口語

- **語言癌不癌**
  - 王品服務生：讓我為您做一個上菜的動作
  - 阿基師：有做一個擁抱的動作
  - 消防隊員要進行一個滅火的動作



圖片來源：https://www.books.com.tw/products/0010703220

- **做 / 進行……動作**
  - COCT書面語語料庫2019 vs COCT口語語料庫2019
    - [word="進行|做"][pos!="PUNC"]{1,10}[word="動作"]
    - "進行|做" [pos!="PUNC"]{1,10} "動作"
      - 回傳 2,509 筆；頻率: 7.85 次/每百萬詞
      - 回傳 1,430 筆；頻率: 66.38 次/每百萬詞

.....的部分 / 部份

- 湯的部分建議攪拌均勻再做享用喔
- 您好，這是沙拉的部分，這邊可以加上我們的醬汁，淋在沙拉上做享用喔！
- 中研院平衡語料庫4.0 vs COCT 口語語料庫2019
  - 的 (部分|部份)
  - 回傳 675 筆，頻率: 60.02 次/每百萬詞 vs 回傳 6,274 筆，頻率: 291.26 次/每百萬詞
    - 你 就 跟 老闆 講 我 要 讓 水 只 能 流過去 那 實作 的 部份 讓 他 們 要 先 來 現場 看
    - 接下來 來 介紹 像 我們 今天 主要 的 食材 首先 是 大 番茄 的 部分 大 番茄 它 裡面 富含 了 茄紅素
- COCT 書面語語料庫2019：回傳 18,830 筆，頻率: 58.90 次/每百萬詞

語言變遷：感恩

- **搭配詞**
  - 1 to the Right
  - Freq (node, collocate) at least: 1
  - Freq (collocate) at least: 1

中研院平衡語料庫4.0
COCT書面語語料庫2019
COCT口語語料庫2019

謂語

No.	Word
1	您
2	科技
3	感恩
4	謝謝
5	師姊
6	你
7	師兄
8	大家
9	尊重
10	善解
11	上人
12	靜慧
13	靜原
14	惜福
15	我們
16	阿芳
17	知足
18	靜映
19	蔣
20	慈濟

定語

No.	Word
1	的
2	禮拜
3	水上
4	惜福
5	祭
6	。
7	他人
8	之
9	聖祭
10	PARTY
11	禮讚
12	茶會
13	祭禮
14	園遊會
15	餐會
16	過去
17	晚會
18	師父
19	懷念
20	晚餐

定語

No.	Word
1	戴德
2	之
3	的
4	不盡
5	。
6	戴
7	，
8	與
9	惜福
10	、
11	心情
12	報德
13	母親
14	和
15	日記
16	party
17	之餘
18	禱告
19	地
20	日誌

語言變遷：暖~

- **暖心；暖+心**
  - (暖心|暖 心)
    - 平衡語料庫 — 0
    - 書面語2016 — 0.02 次/每百萬詞（感覺暖心，自在暖心，紅酒暖心胃，文字讓您暖心）
    - 口語2016 — 0.23 次/每百萬詞（暖心的一個低氣壓）
    - 書面語2019 — 0.07 次/每百萬詞（暖心的肯定，暖心的話）
    - 口語2019 — 0.28 次/每百萬詞（暖胃暖心，暖心茶會，暖心的火爐）
- **暖+X**
  - 暖 _N*
    - 平衡語料庫 — 冬，空氣，風，場，瓶
    - 書面語2016 — 手，酒，人心，你
    - 書面語2019 — 手，男，人，他，人心

語言變遷：可愛死

- 柴柴的小眼神，真是**可愛死**了。
  - [word="..死" & pos="V.*"]
    - 平衡語料庫 ─ 7 types（生病死，安樂死，折磨死，感動死，過勞死，自然死，見光死）
    - 書面語2019 ─ 301 types（悟徹死，鞭打死）
  - 可愛 死
  - ❗可愛死 **尚未詞彙化**
    - 平衡語料庫 ─ 0
    - 書面語2016 ─ 0.04 次/每百萬詞（那條小狗可愛死了，左邊算起第三個有瀏海的可愛死了）
    - 口語2016 ─ 0
    - 書面語2019 ─ 0.03 次/每百萬詞
    - 口語2019 ─ 0

研究目的 vs 語料庫

- 平衡語料庫
  - 小而美
  - 人工分詞
  - 1981 ~ 2007 年
- COCT語料庫
  - 數大便是美
  - 自動分詞
  - 2007 年 ~

語料庫指令

天下合久必分，分久必合


謝謝
Thank You