

國家教育研究院

NATIONAL ACADEMY for EDUCATIONAL RESEARCH

華語語料庫在華語教學上之應用 —進階

李詩敏

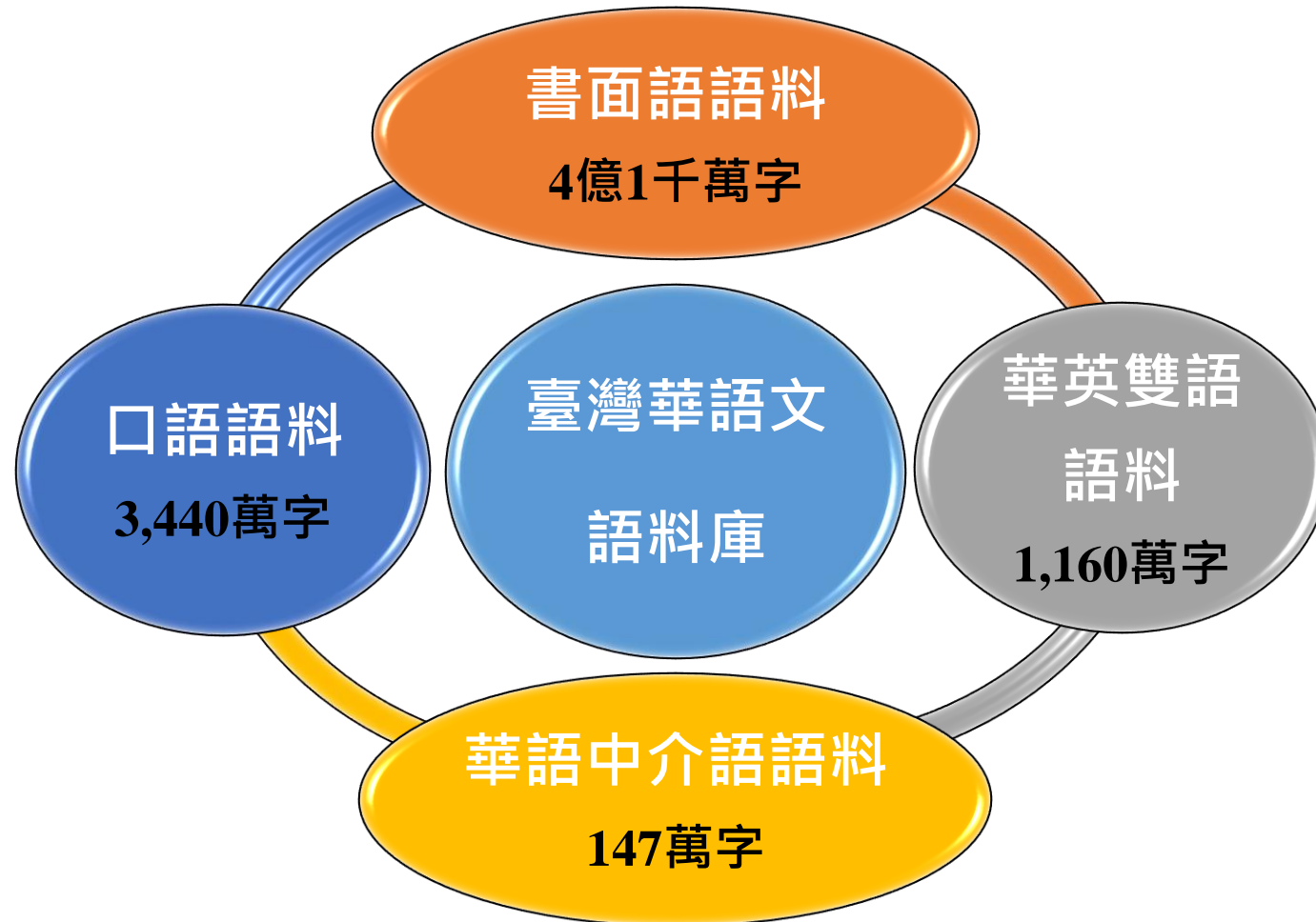
國家教育研究院語文教育及編譯研究中心

2020.06.12

大綱

- 國教院語料庫索引典系統
 - 書面語
 - 口語
 - 中介語
 - 語義場關聯詞查詢系統
- 在詞彙分析上的應用
 - 近義詞
 - 離合詞
 - 結構：A-not-A；重疊；好X/難X；動補結構
- 書面語 vs. 口語

臺灣華語文語料庫 (Corpus of Contemporary Taiwanese Mandarin)



國教院語料庫索引典系統

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)

帳號:

密碼:

保持登入狀態:

訪客使用: 帳號:naer 密碼:naer

登入

清除

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](#)

[國教院華語中介語索引典系統2016](#)

[國教院華語中介語索引典系統2019](#)

[COCT 書面語語料庫2016](#)

[COCT 書面語語料庫2017](#)

[COCT 書面語語料庫2019](#)

[COCT 書面語語料庫2015](#)

簡易查詢 Simple query

Menu

1 Corpus queries

Standard query

Restricted query

Word lookup

Frequency lists

Keywords

Analyse corpus

Saved query data

Query history

Saved queries

Categorised queries

Upload a query

COCT 書面語語料庫2019

Standard Query

2 輸入查詢字詞

Query mode:

Simple query (ignore case) ▼

[指令速查表](#) [選單參考表](#)

Number of hits per page:

50 ▼

Restriction:

None (search whole corpus) ▼

3

Start Query

Reset Query

簡易查詢&萬用字元(1)

符號	意義	範例	符合之搜尋結果
?	1個中文字元	?書	讀書，秘書，圖書，……
*	0~N個中文字元	*書	書，讀書，保證書，……
+	1~N個中文字元	+書	讀書，保證書，……
\	特殊字元本身	\?	所有有「?」的句子
[X,Y]	[]內的字詞符合任一即可 (X或Y，符合任一即可)	[台,臺]北	台北，臺北
(X Y)	()內的所有詞 (類) (X或Y，符合任一即可)	喜歡(我 他)	喜歡我，喜歡他
		喜歡(_V* _N*)	喜歡這，喜歡自己，喜歡吃

- 台北，臺北

☺指令：[台,臺]北

☹指令：(台|臺)北

[]和()的區別：()視為一個分詞單位

- 喜歡我，喜歡他

☺指令：喜歡(我|他)

☺指令：喜歡[我,他]

☺指令：喜歡(我|他)

☹指令：喜歡[我,他]

簡易查詢&萬用字元(2)

範例	意義	符合之搜尋結果
把_Nf	詞類標記為Nf (量詞) 的「把」	把 (Nf)
把_N*	詞類標記為N (名詞) 的「把」	把 (Nf) , 把 (Na)
?把_N*	詞類標記為N且詞尾為「把」的雙字詞	火把, 掃把, 門把
把_N	詞類標記為N且詞尾為「把」的詞	把, 火把, 一把把, 郭威把, 雅朗阿把
+把_N*	詞類標記為N且詞尾為「把」的非單字詞	火把, 一把把, 郭威把, 雅朗阿把

- 一把火, 兩把鑰匙, 一把好弓

☺指令: 把_Nf

Collations (搭配)

window from: 1 to the Right

☺指令: 把_Nf * _N*

☺指令: 把_Nf * *_N*

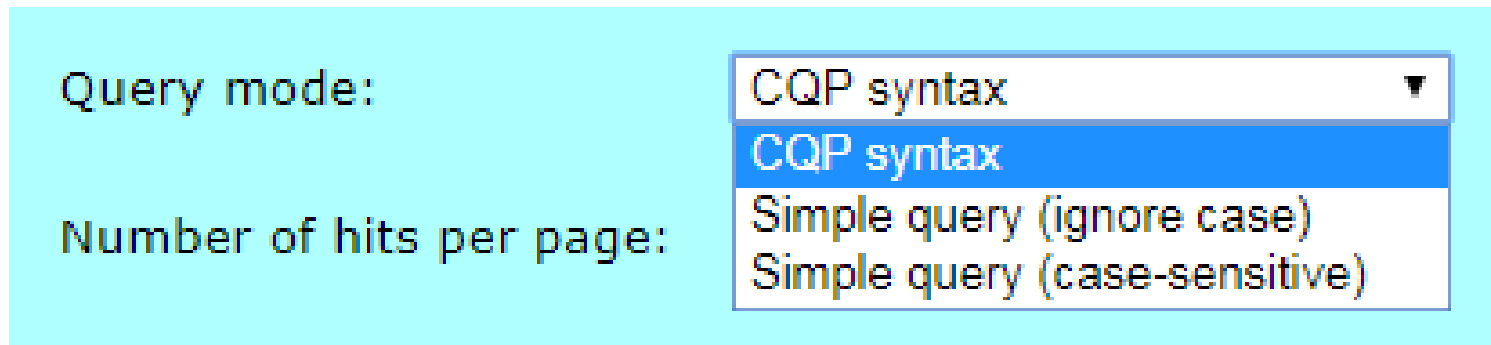
☹指令: 把_Nf + _N*

☹指令: 把_Nf + *_N*

找到: 玉米兩把, 雞蛋一個

CQP查詢

- CQP: Corpus Query Processor (Evert & Hardie, 2011)
 - 提供細緻的查詢功能
 - 指令有些複雜
- 查詢介面
 - Query mode → CQP syntax



CQP查詢式(1)

符號	意義	說明
[word="把"]	所有「把」	[] (一個詞。[]內可以空白，代表一個任意詞) word="A" (限制詞形為A)
"把"		
[pos="Nf"]	所有的量詞	pos="B" (限制詞類為B) pos!="B" (限制詞類不得為B)
[word="把" & pos="Nf"]	「把」：量詞	
[word="把" & pos!="Nf"]	「把」：非量詞	X & Y (兩個條件都要符合) X Y (任一條件符合即可)
[word="把" & !(pos="Nf")]		
[word="把" & pos="Nf Na"]	「把」：量詞或名詞	
[word="把" & (pos="Nf" pos="Na")]		

☹指令：[word="把" & pos="Nf" & pos="Na"]

指令的意思：「把」既是量詞也是名詞



- 「把」後面直接接普通名詞(Na)

提示1：[word="把"]

提示2：[pos="Na"]

☺指令：[word="把"][pos="Na"]

把犀牛骨變成石頭

坐第一把交椅

- 介詞(P)「把」後面直接接普通名詞(Na)

☺指令：[word="把" & pos="P"][pos="Na"]

- 把(P) 一封 信件 (Na) 丟進郵筒

☺指令：[word="把" & pos="P"][][pos="Na"]

☺指令：[word="把" & pos="P"][][pos="Na"]

☺指令：[word="把" & pos="P"][][1,3][pos="Na"]

把一些小事看得十分嚴重

把自己真正的意願表達出來

CQP查詢式(2)

符號	符合之搜尋結果	說明
[word="心無旁."]	心無旁驚，心無旁驚， 心無旁顧	. (代替一個字)
[word="哈*"]	哈，哈哈，哈哈哈	* (前字重複0~N次)
[word="把" & pos!="V.*"]	非動詞「把」	X.* (以X開頭的所有詞類)
[word="[台臺]北"]	台北，臺北	[AB] (AB任一字均可)
[word="一[^百千萬]二[^百千萬]"]	一清二楚，一乾二淨， 一石二鳥，一九二年	[^AB] (限制詞語不可出現的組成成分)
[word="(台北 高雄)市"]	臺北市，高雄市	(X Y)Z (XY任一組均可)

- 比較：[word="[台北高雄]市"]

找到：北市

高市

台市

多重限制詞類

- 找到：既不是介詞(P)也不是量詞(Nf)的「把」

提示1：pos!="P" 或 !(pos="P")

提示2：A & B 或 (A|B)

☺指令：[word="把" & pos!="P" & pos!="Nf"]

☺指令：[word="把" & (pos!="P" & pos!="Nf")]

☺指令：[word="把" & pos!="P|Nf"]

☺指令：[word="把" & !(pos="P" | pos="Nf")]

☹指令：[word="把" & (pos!="P" | pos!="Nf")]

找到：把_P，把_Nf，把_Na，把_VA，把_VC



限制詞語的組成成分

- [word="一[^百千萬]二[^百千萬]"]

找到：一清二楚_VH
一乾二淨_VH
一石二鳥_Na
~~一九二三年_Nd~~
~~一九二三_Neu~~

☺指令： [word="一[^百千萬]二[^百千萬]" & pos!="Nd|Ne.*"]

找到：一清二楚
一乾二淨
一石二鳥
一夫二主_Na
一七二街_Nc

近義詞辨析：切；割

- ☺指令(Simple query)：切_V*
割_V*

- ☺指令(CQP syntax)：[word="切" & pos="V.*"]
[word="割" & pos="V.*"]

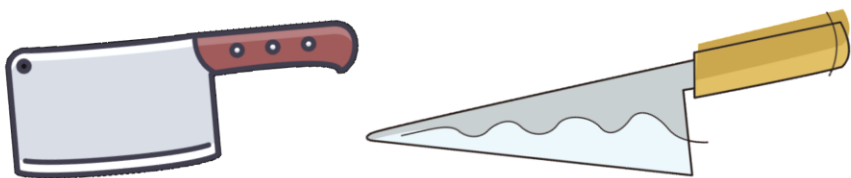
- 搭配 Collocations

切

結果：塊，絲，(薄)片，細，末

對象：食物

工具：刀，菜刀



力道；方向

割

結果：割去，割掉，割得支離破碎

對象：身體部位或器官，肉，韭菜，草

工具：刀，刀片，鐮刀



☹指令：[word="切"][pos="V.*"]

☹指令："切" [pos="V.*"]

找到：切+動詞

語義場關聯詞查詢系統

A semantic field is a set of words grouped semantically (by meaning) that refers to a specific subject.

1 選擇語料庫

- 遠流語料
- 中國時報
- 國語日報
- 平衡語料庫
- Chinese Giga Words

2 選擇分級詞表

- 國教院分級詞表
- 華測八千詞

3 選擇詞彙等級

- 不限制
- 查 7 級(含)以下詞
- 查 6 級(含)以下詞
- 查 5 級(含)以下詞
- 查 4 級(含)以下詞
- 查 3 級(含)以下詞
- 查 2 級(含)以下詞
- 查 1 級詞

查詢詞(組)語義場關聯詞

4 輸入

正關聯詞(組)：

請輸入正關聯詞，例如：蘋果

負關聯詞(組)：

請輸入負關聯詞，例如：賈伯斯

輸出詞數：

輸出

語義場關聯詞：

語義場關聯詞查詢系統：應用

- 應用
 - 考題 (選項誘答力)
 - 教材 (詞語難易度 ; 詞語聯想)
 - 研究 (辨析)
- 例：看到同學玩得如此_____，我覺得這樣的活動很有意義。
 - 高興
 - 開心
 - 難過
 - 歡喜
 - 興奮
 - 欣慰
 - 著急
 - 驚奇
 - 樂意
 - 感激
 - 吃驚

語義場關聯詞：

開心 0.7349	3,基礎
難過 0.6556	2,基礎
歡喜 0.6312	5,進階
興奮 0.5932	4,進階
欣慰 0.5866	6,精熟
著急 0.5853	4,進階
驚奇 0.5843	5,進階
樂意 0.5792	5,進階
感激 0.5782	5,進階
吃驚 0.5745	5,進階
惱火 0.5744	7,精熟
驚訝 0.5636	4,進階
難為情 0.5625	5,進階
懊喪 0.5620	--
興高采烈 0.5600	6,精熟
得意 0.5600	5,進階
氣憤 0.5548	6,精熟
過意不去 0.5538	7,精熟
難受 0.5492	5,進階
不好意思 0.5410	2,基礎

JUST
DO IT!

- 「蘋果」的正關聯詞
- 「黃牛」，使用正負關聯詞加以分類

從蘋果到奇異果，詞彙夥伴在哪裡？
語料庫科技在語文教學的應用(II)
https://epaper.naer.edu.tw/edm.php?edm_no=192&content_no=3416

遠流語料

中國時報

國語日報

平衡語料庫

Chinese Giga Words

輸出

語義場關聯詞:

藍莓	0.6044	--
檸檬	0.5820	4,進階
橘子	0.5804	3,基礎
柳橙	0.5728	--
Apple	0.5664	--
梨子	0.5659	--
桃子	0.5615	5,進階
草莓	0.5603	3,基礎
柳丁	0.5554	--
葡萄	0.5551	3,基礎

輸出

語義場關聯詞:

APPLE	0.6890	--
橘子	0.6484	3,基礎
Apple	0.6471	--
iPhone	0.6454	--
黑莓	0.6450	--
iPad	0.6439	--
IPHONE	0.6381	--
IWATCH	0.6265	--
IPAD	0.6240	--
IPHONE5	0.5997	--

輸出

語義場關聯詞:

iPhone	0.5140	--
賈伯斯	0.4725	--
木瓜	0.4480	3,基礎
香蕉	0.4421	2,基礎
水果	0.4375	1,基礎
番茄	0.4290	3,基礎
西瓜	0.4260	2,基礎
火腿	0.4260	--
肉類	0.4216	4,進階
切片	0.4202	6,精熟

輸出

語義場關聯詞:

馬鈴薯	0.4875	5,進階
柿子	0.4693	--
橘子	0.4676	3,基礎
櫻桃	0.4540	5,進階
刨冰	0.4400	--
起司	0.4361	7,精熟
榴槤	0.4348	--
梨子	0.4319	--
芒果	0.4315	3,基礎
苦瓜	0.4266	--

輸出

語義場關聯詞:

蘋	0.5130	--
櫻桃	0.5032	5,進階
柑橘	0.4871	--
梨子	0.4830	--
橘子	0.4727	3,基礎
香蕉	0.4659	2,基礎
荔枝	0.4639	--
葡萄	0.4607	3,基礎
奇異果	0.4415	--
馬鈴薯	0.4413	5,進階

離合詞：結婚

- 書面語語料
 - ☺指令：`[word="結"][][word="婚"]`
 - 找到：結了婚 找不到：結了一次婚
 - 結什麼婚
 - 結個婚
 - ☺指令：`[word="結"][]{1,5}[word="婚"]`
 - 找到：結了婚
 - 結過好幾次婚
 - 結個幸福的婚
 - ☹指令：`[word="結"][][word="婚"]`
 - 找到：但卻發現她早已另結新歡。兩人後來雖結了婚
- 中介語語料
 - ☺指令：`結婚_N*`
 - 找到：我媽媽在高中教國語以前她結婚我爸爸。

離合詞：睡覺

- 找到：睡了一覺
睡不著覺
睡個安穩香甜的好覺

提示：`[word="睡"][][word="覺"]`

☺指令：`[word="睡"][]{1,5}[word="覺"]`

找不到：睡了很長、很長一覺
睡了一個沒有惡夢的好覺

☺指令：`[word="睡"][]{1,10}[word="覺"]`

找到：睡了。但願這一覺

☺指令：`[word="睡"][pos!="PUNC"]{1,10}[word="覺"]`

找到：睡個解除疲勞的好覺
睡前做伸展操有助於晚上睡個好覺



JUST
DO IT!

詞彙結構：A-not-A

- 是不是
對不對
一樣不一樣

☺指令：a:[] [word="不"] b:[] :: a.word=b.word

找到：，不，
不不不
父不父

☺指令：a:[pos!="PUNC"] [word="不"]
b:[pos!="PUNC"] :: a.word=b.word

找到：不不不
父不父

☺指令：a:[pos!="PUNC|N.*"] [word="不"]
b:[pos!="PUNC|N.*"] :: a.word=b.word

找到：不不不

詞彙結構：重疊

- 乾乾淨淨
漂漂亮亮
平平安安

☺指令：[word="..." & char(word, 0)=char(word, 1) & char(word, 2)=char(word, 3)]

找到：許許多多_Neqa
時時刻刻_D
慢慢慢慢_VH

☹指令：[word="..." & char(word, 0)=char(word, 1) & char(word, 2)=char(word, 3)
& pos="VH.*"]

(錯殺過多，限縮太窄)

找到：慢慢慢慢_VH

找不到：搖搖晃晃_VA，恭恭敬敬_VI

☺指令：[word="..." & char(word, 0)=char(word, 1) & char(word, 2)=char(word, 3)
& char(word, 1)!=char(word, 2) & pos="V.*"]

其他類似的詞彙結構

- A非A

- 提示：`a:[] [word="不"] b:[] :: a.word=b.word`
- ☺指令：`a:[] [word="非"] b:[] :: a.word=b.word`
- 找到：魔 非 魔
霧 非 霧

- 不X不Y

- 提示：`[word="..." & char(word, 0)=char(word, 1) & char(word, 2)=char(word, 3)]`
- ☺指令：`[word="不.不." & char(word, 1)!=char(word, 3)]`
- 找到：不理不睬
不三不四

- 其他：半X半Y；X里XY；ABAB



詞彙結構：好X；難X

- 好看
好吃
難聽
難寫

☺指令：[word="好.|難." & pos="VH"]

找到：好好
好奇
難得
好看
好玩
難看

☺指令：[word="好.|難." & pos="VH" & char(word,0)!=char(word,1)]

動補結構：V死；V滿；V完；V壞

☺指令：[word=".死" & pos="V.*"]

找到：枯死
累死

找不到：鞭打死
笨死

☺指令：[pos="V.*"][word="死"]

(加入斷詞可能有誤的情形)

找到：睡死
笨死
裝死
快死

☺指令：([word=".+死" & pos="V.*"]|[pos="V.*"][word="死"])

(一網打盡，寧可錯殺也不可放過)

☺指令：[word=".+死|.滿|.完|.壞" & pos="V.*"]

(小滿貫)

☺指令：([word=".+死|.滿|.完|.壞" & pos="V.*"]|[pos="V.*"][word="死|
滿|完|壞"])

(大滿貫)

動補結構：和其他結構的搭配

- 和「把」、「被」、「將」字句的搭配：以「V滿」為例
 - 提示：[word=".+滿" & pos="V.*"]
 - ☹指令：[word="把|被|將"][*][word=".+滿" & pos="V.*"]
 - ☹指令：[把,被,將](+)* +滿_V*
(無法阻止跨句出現→必須使用CQP指令)
找到：將處處均是喜悅。先知書中充滿
 - ☺指令：[word="把|被|將"][pos!="PUNC"]*[word=".+滿" & pos="V.*"]
找到：被 擠滿
把 空間 塞滿
將 光復南路 地下道 填滿
將 造成 公司的 自滿
- 和副詞(D)搭配的情形：以「V滿」為例
 - [word=".+滿" & pos="V.*"] → Collocations → and/or tag: D
 - +滿_V* → Collocations → and/or tag: D

文類比較：書面語；口語

- 語言癌

- 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

- 做.....動作；進行.....動作

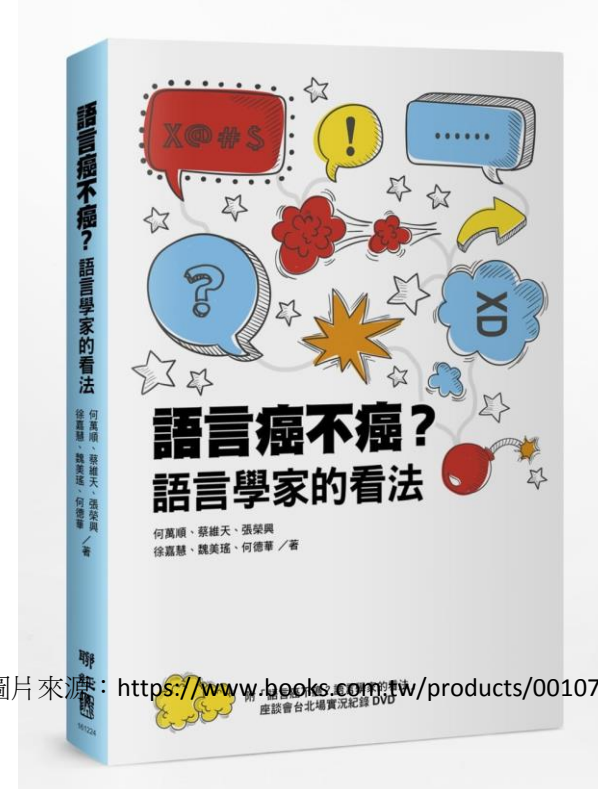
- 😊指令："進行|做" [pos!="PUNC"]{1,10} "動作"

- COCT書面語語料庫2019：2,509 matches；frequency: 7.85 instances per million words

- COCT口語語料庫2019：1,403 matches；frequency: 66.38 instances per million words

- 😞指令："進行|做" [pos!="PUNC"]+ "動作"

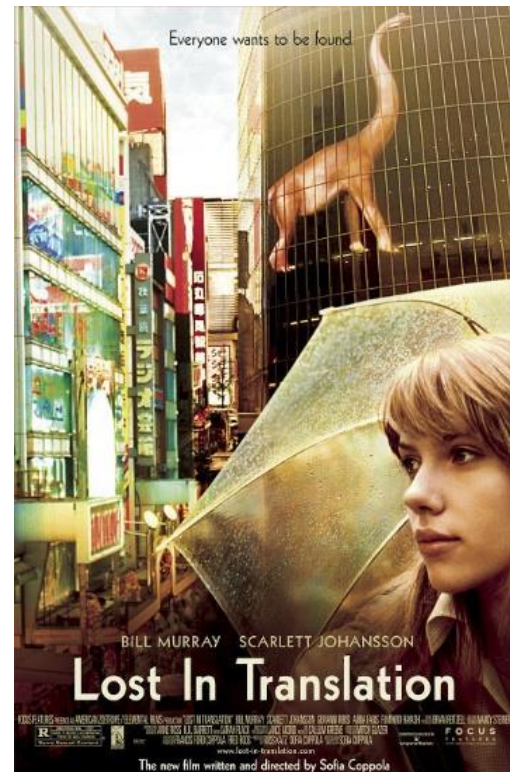
(無法處理沒有標點符號的口語語料)



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

文類比較：書面語；口語

- 不必 vs. 不用
 - ☺指令：(不必|不用)
 - ☺指令：[word="不必|不用"]
 - 書面語：55.81% vs. 44.19%；口語：7.34% vs. 92.66%
 - 不必
 - 書面語：29,161 matches；frequency: 91.21 instances per million words
 - 口語：23,091 matches；frequency: 72.22 instances per million words
 - 不用
 - 書面語：510 matches；frequency: 23.68 instances per million words
 - 口語：6,441 matches；frequency: 299.01 instances per million words
-的部分/部份
 - ☺指令：的(部分|部份)
 - ☺指令："的"[word="部分|部份"]
 - ☺指令：[word="的"][word="部分|部份"]
 - 書面語：18,830 matches；frequency: 58.90 instances per million words
 - 口語：6,274 matches；frequency: 291.26 instances per million words
 - 找到：所以說其實媽媽們我們可以多多運用這些菇類還有素火腿的部分來提升爸爸對於這種番茄的一個攝取的一個狀況



英文片名：Lost in Translation
臺灣片名：愛情，不用翻譯



sml@mail.naer.edu.tw

請幫忙填答問卷

華語文語料庫及標準體系
整合應用工作坊 

2020.08.07 (五)

09:00~17:30

國家教育研究院臺北院區 11樓大禮堂

系統基礎與進階實作

&

華語文教學應用實作



國家教育研究院
National Academy for Educational Research
www.naer.edu.tw

臺北市大安區和平東路一段179號

報名網址

