語料庫應用於華語文教學

白明弘 2018/3/28

2018語料庫應用於華語文教學工作坊

大綱

- 壹、統計語料庫詞頻
- 貳、詞彙觀察與例句選擇
- 參、詞彙的難度觀察與替換
- 肆、漢字與構詞
- 伍、語法點查詢
- 陸、CQP 查詢語言

壹、統計語料庫詞頻

如何知道語料庫所有的詞及詞頻？

- 目的：
  - 統計語料庫中所有詞彙的詞頻
- 步驟：
  1) 網址：http://coct.naer.edu.tw/cqpweb
  2) 選擇語料庫
  3) 選擇 Frequency lists
  4) 按下 Show frequency list
  5) 選擇 Download whole list
  6) 按下 Go!

詞頻下載頁面

Menu

Corpus queries

- Standard query
- Restricted query
- Word lookup
- Frequency lists ①
- Keywords
- Analyse corpus
- Export corpus

Saved query data

- Query history
- Saved queries
- Categorised queries
- Upload a query
- Create/edit subcorpora

Corpus info

- View corpus metadata
- *No corpus documentation available*

中研院平衡語料庫4.0

Frequency lists

You can view the frequency lists of the whole corpus and frequency lists for subcorpora you have created. Click here to create/view subcorpus frequency lists.

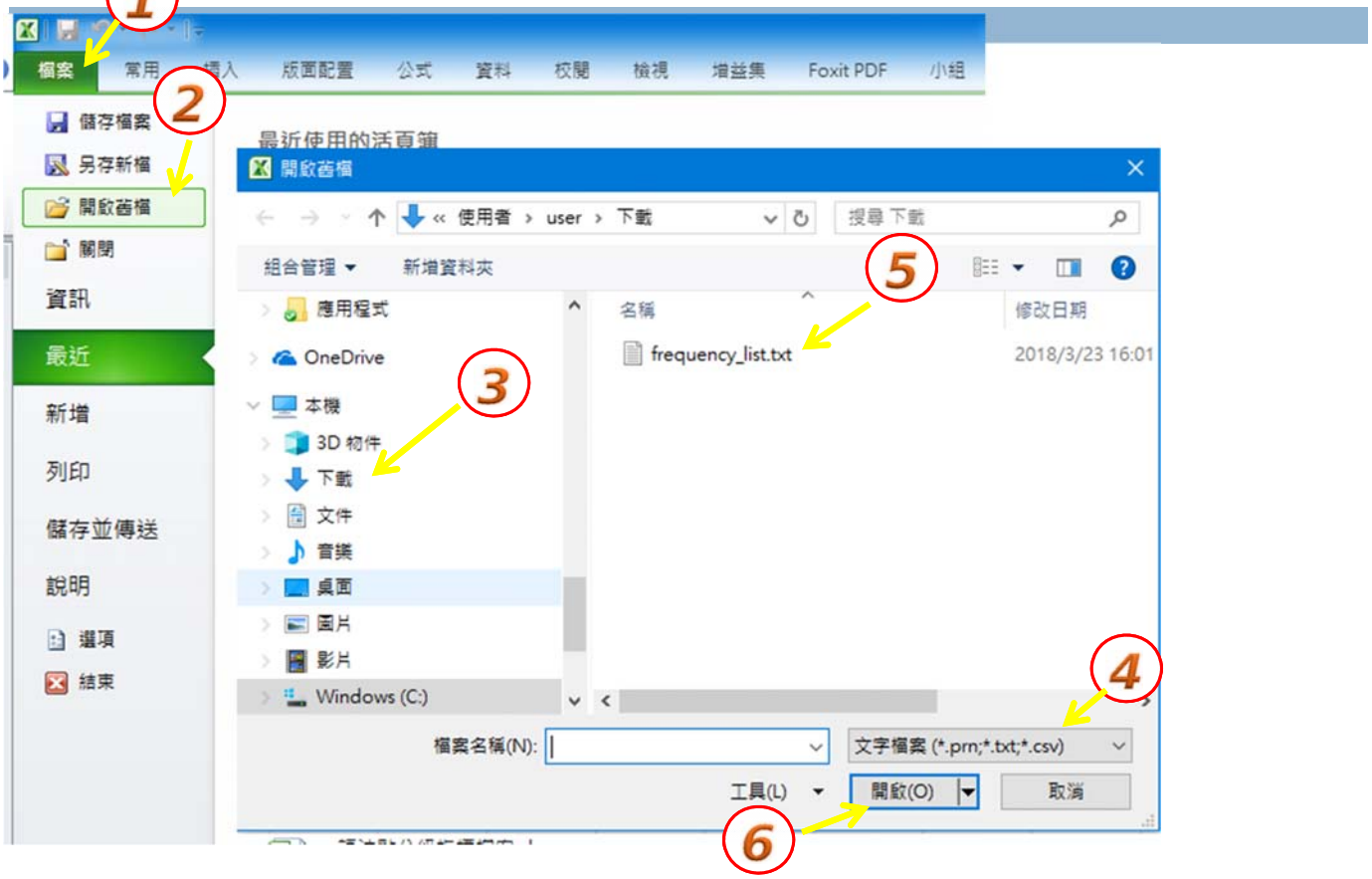
View frequency list for ...	Whole of 中研院平衡語料庫4.0 ②
View a list based on ...	Word forms

Frequency list option settings

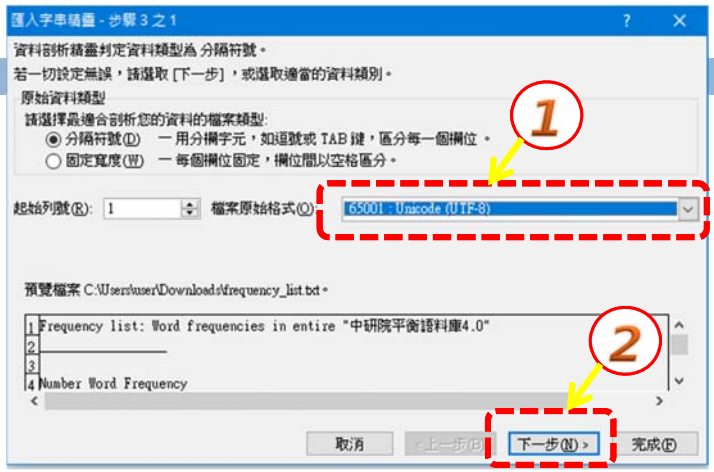
Filter the list by *pattern* - show only words/tags ...	starting with
Filter the list by *frequency* - show only words/tags ...	with frequency between ... and ...
Number of items shown per page:	50
List order:	most frequent at top ③

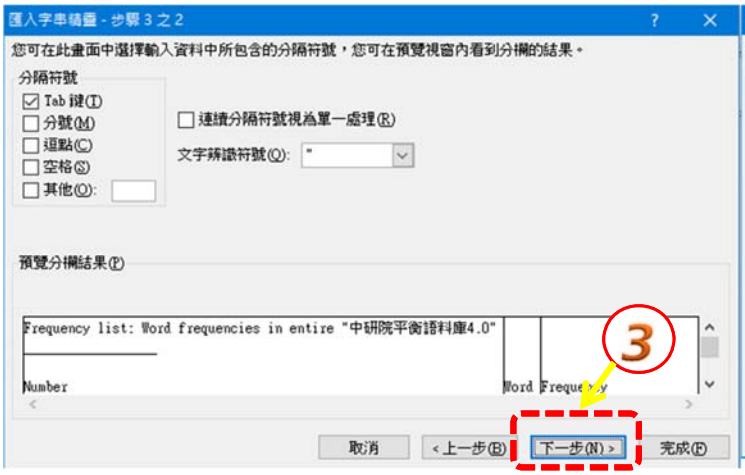
④ Show frequency list Clear the form

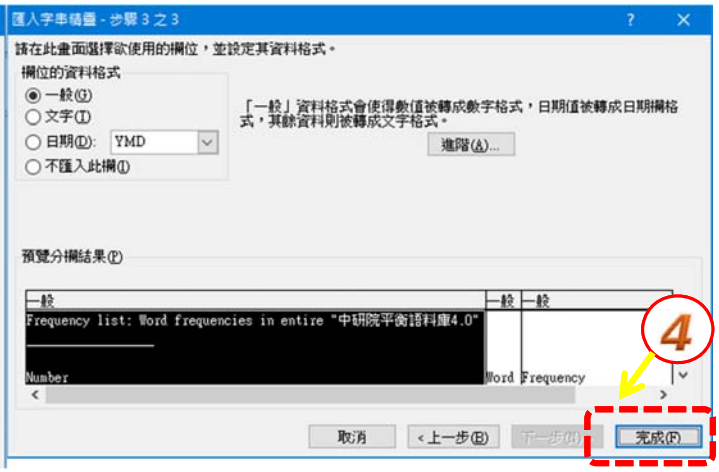
在 Excel 中匯入詞表



Excel 匯入精靈

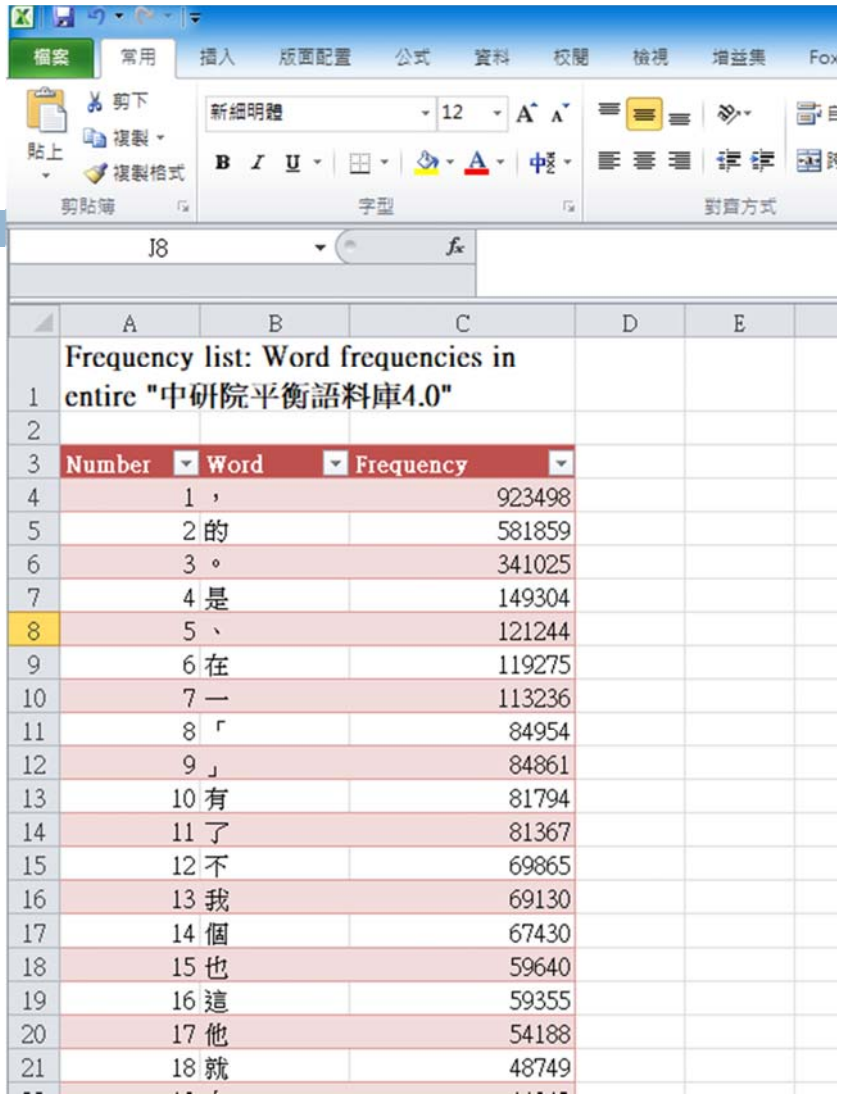






詞表匯入另存新檔

- 完成後，調整一下格式
- 另存成 **excel** 格式



	A	B	C	D	E
1	Frequency list: Word frequencies in entire "中研院平衡語料庫4.0"				
2					
3	Number	Word	Frequency		
4	1	，	923498		
5	2	的	581859		
6	3	。	341025		
7	4	是	149304		
8	5	、	121244		
9	6	在	119275		
10	7	一	113236		
11	8	「	84954		
12	9	」	84861		
13	10	有	81794		
14	11	了	81367		
15	12	不	69865		
16	13	我	69130		
17	14	個	67430		
18	15	也	59640		
19	16	這	59355		
20	17	他	54188		
21	18	就	48749		

練習：下載詞表

統計選項說明

- View frequency list for …
  - 選擇語料庫或子語料庫
- View a list based on …
  - Word form：輸出詞
  - Feature tag：輸出特徵標記
  - Part-of-speech: 輸出詞類
- Filter the list by pattern – show only tags/words
  - starting with：設定詞首
  - ending with：設定詞尾
  - containing：設定包含
- Filter the list by frequency – show only tags/words
  - 設定詞頻區間
- List order
  - most frequent at top：高頻先
  - least frequent at top：低頻先
  - alphabetical order：依字母順序(內碼順序)

貳、詞彙觀察與例句選擇

搭配詞表的再說明

There are 13,381 different words in your collocation database for "[word="高興"%c]". (Your query "高興" returned 14,368 matches in 9,573 different texts) [1.14 seconds - retrieved from cache]

No.	Word	Total no. in whole corpus	Expected collocate frequency	Observed coll frequency	In no. of texts	Log-likelihood
1	很	258,295	225.077	4,592	3770	19261.847
2	地	216,746	188.871	1,437	1236	3361.223
3	非常	56,292	49.053	843	757	3225.823
4	我	1,034,152	901.154	2,919	2208	2877.925
5	極了	2,668	2.325	320	290	2557.054
6	不	739,692	644.563	2,214	1861	2357.53
7	得	188,944	164.645	1,089	973	2280.531
8	感到	21,726	18.932	338	310	1316.126
9	十分	18,162	15.826	317	300	1303.977
10	見到	9,129	7.955	253	234	1267.803

1

2

「很」出現位置分布

- 從位置分布可知「很」常出現在「高興」的前一個詞

Within the window -3 to 3, 很 occurs 4,592 times in 3,770 different texts (expected frequency: 225.077)			
Distance	No. of occurrences	In no. of texts	Percent
-3	42	42	0.9%
-2	405	387	8.8%
-1	4,037	3,393	87.9%
1	7	7	0.2%
2	56	56	1.2%
3	45	44	1%

「很」和「高興」一起出現的例句

我認識一些義大利的朋友，他們高興或不	高興	表現得很明顯，心理好像沒有很多壓抑的東西
的動作。在校園打籃球，投球進籃後很	高興	，於是他們手跟手就互相碰一下，我們
會非常高興地叫出名字，然後說：「很	高興	看到你！今天過得好嗎？」一邊說
也有的朋友嘴裡說著：「我今天很	高興	，心情很好。」可是腳步沉重，步履蹣跚
也有的朋友嘴裡說著：「我今天很	高興	，心情很好。」可是腳步沉重，步履蹣跚
九三年二月發出同樣的警語。他說：我很	高興	，《資本主義與自由》的中文版能在台灣發行
藍先生同意，作為本書附錄一。我很	高興	，《資本主義與自由》的中文版能在台灣發行
、擴大干預範圍，弗利曼說得好：我很	高興	，《資本主義與自由》的中文版能在台灣發行
，就能完成一大片，那時你會很	高興	、很有成就感。這樣就叫做有步驟、有
，就能完成一大片，那時你會很	高興	、很有成就感。這樣就叫做有步驟、有
拉菲爾先生遺贈給依瑟的一筆錢。她很	高興	拉菲爾先生後來並沒有改變心意。「我想你
，」瑪波小姐說。「沒錯。我很	高興	他那樣做。我就在想，很可能他
一七八○年，」格林太太說。格林太太很	高興	瑪波小姐有鑑賞眼光。她把瑪波小姐帶進客廳
的敲門聲，克羅蒂走進房間，看起來很不	高興	。「瑪波小姐，樓下有個年輕人要見你
壞事的壞人。」「聽你這麼說我很	高興	，」汪斯岱教授說。「你簡直不會相信，
的是，你知道，我覺得她反而看起來很	高興	，就像是——呃，就是很高興。這不

語料庫的句子適合教學嗎？

- 語料庫研究者對傳統例句的批判
  - 人造例句不自然、不具代表性
- 傳統例句編輯者的回應
  - 語料庫的句子不夠典雅(不是出自名家手筆)
  - 語料庫的句子未必正確
- 語料庫的價值何在？

下列的例句適合詞彙教學嗎？

- **感冒**：
  - 我對他的態度很**感冒**！
- **高興**：
  - 你就讓他**高興**個夠吧！
- **代溝**：
  - 我不知道什麼是**代溝**！
- **無地自容**：
  - 林布蘭幾乎**無地自容**
- **感冒**：
  - 已經多年了，我沒再罹患**感冒**。
- **創業**：
  - 把**創業**精神運用在自己的生活中。

什麼樣的例句適合教學？

- 適合教學的例句應具備哪些條件？(Atkins&Rundell, 2008)
  - 自然而典型(naturalness and typicality)
    - 常見的搭配詞、語法模式、前後文等等
    - 偏好的時態、數字、心情、態度等等
  - 包含有用的訊息(informative)
    - 有助於了解詞彙的意思
  - 易於理解(intelligible)
    - 避免不必要的困難用詞
    - 避免不必要的複雜結構

重新檢視例句

- **用法不典型**
  - **感冒**：我對他的態度很**感冒**！
  - **高興**：你就讓他**高興**個夠吧！
- **缺乏訊息**
  - **代溝**：我不知道什麼是**代溝**！
  - **無地自容**：林布蘭幾乎**無地自容**
- **不易理解**
  - **感冒**：已經多年了，我沒再罹患**感冒**。
  - **創業**：把**創業**精神運用在自己的生活中。

如何從語料庫找到適合教學的例句？

- 常見**搭配詞**找例句
  - 高興
    - 很/非常 + 高興
    - 高興 + 地
    - 覺得 + 高興
    - 感到 + 高興
  - 無地自容
    - 羞愧 + 無地自容
    - 感到 + 無地自容
    - 使/令/讓人 + 無地自容
- 常見**搭配詞性**找例句
  - 高興
    - Dfa + 高興
    - 高興 + T
    - D + 高興
    - VK + 高興
  - 無地自容
    - VL +無地自容
    - VH +無地自容
    - VK +無地自容
    - DE + 無地自容

語料庫的句子適合直接用在教學上嗎？

- 大部分的例句都需要經過編修(Kilgarriff, 2008)
- 原因：
  - 常有不相干的子句
  - 常包含太複雜的名稱
  - 常包含代名詞、指代，如果沒有足夠的前後文，將使例句不知所云。
  - 中文常常缺少主詞(零代詞)

原住民朋友們唱起歌跳起舞，似乎特別地愉快，像在豐年祭當中大家開心地喝酒，之後唱歌跳舞，**高興**得不得了。傅夏器一時羞愧萬分，**無地自容**，只好紅著臉皮，不辭而別。

例句的編修

- 選擇例句：(Kilgarriff, 2008) 的建議
  - 依據前面的建議找到例句
  - 句長建議：10~25個詞
  - 詞頻建議：不要超出最常見的 17,000 詞
- 例句編修
  - 原則：儘量不影響句子的結構、詞語搭配、語法搭配等
  - 去除不相干的子句
  - 調整太複雜的名詞
  - 指代詞調整

小結

- **使用高頻搭配詞選擇例句的好處**
  - **用法典型**
    - 聽到這個好消息，全校的師生都 **很 高興** 。
    - 後來警察真的找到了失主，我 **覺得 很 高興** ，能夠物歸原主 。
  - **訊息有助理解**
    - 這件事一旦傳播開來，我們便**羞愧**得**無地自容**啦！
    - 他幹過的壞事仍然讓他**感到無地自容**。
  - **易理解**
- **大部分語料庫的句子都需要經過編修才適用於教學**
  - 原則：儘量不影響句子的結構、詞語搭配、語法搭配等

參、詞彙的難度觀察與替換

詞彙難度的觀察

- 目的
  - 找出例句中太難的詞
  - 取代成難度較低的詞
- 方法
  - 工具
    - 詞彙分級標記：http://coct.naer.edu.tw/tools/
    - 語義場關聯詞查詢：http://coct.naer.edu.tw/word2vec


國家教育研究院語料庫工具
詞表統計工具
字表統計工具
例句編輯工具
例句編
1
自動斷詞
2
傅夏器一時羞愧萬分，無地自容，只好紅著臉皮，不辭而別。
3
送出
輸出詞彙分級訊息
4
分級詞表：華語八千詞
準備一級 準備二級 入門級 基礎級 進階級 高階級 流利級
1 2 3 4 5 6 7
傅夏器 一時 羞愧 萬分 ， 無地自容 ， 只好 紅 著 臉皮 ，
X 5 X 7 X 4 5 2,5 X
不辭而別 。
X

詞彙分級標記


輸出詞彙分級訊息
分級詞表：華語八千詞
準備一級 準備二級 入門級 基礎級 進階級 高階級 流利級
1 2 3 4 5 6 7
傅夏器 一時 羞愧 萬分 ， 無地自容 ， 只好 紅 著 臉皮 ，
X 5 X 7 X 4 5 2,5 X
不辭而別 。
X

語義場關聯詞查詢

語義場關聯詞查詢

- 羞愧
  - 羞恥 0.7027 --
  - 慚愧 0.6797 6,高階級
  - 羞慚 0.6679 --
  - 內疚 0.6390 --
  - 愧疚 0.6304 --
  - 懊悔 0.6297 --
  - 難過 0.6014 4,基礎級
  - 丟臉 0.5861 6,高階級
  - 失望 0.5858 6,高階級
  - 害怕 0.5852 5,進階級

語義場關聯詞查詢

- 操心(高階級)
  - 擔心　0.6447　4,基礎級
  - 掛心　0.5852　--
  - 擔憂　0.5694　--
  - 煩心　0.5506　--
  - 操煩　0.5397　--
  - 發愁　0.5189　--
  - 大驚小怪　0.5151　--
  - 耽心　0.5036　--
  - 傷腦筋　0.4832　6,高階級
  - 操勞　0.4809　--

練習：語義場關聯詞查詢

- 五官　X,無等級
- 周到　6,高階級
- 關照　7,流利級
- 彰顯　X,無等級
- 恭維　7,流利級
- 鼻涕　7,流利級

如何出作業或考題

- 看到同學玩得如此____，我覺得這樣的活動很有意義。
  - 高興
    - 榮幸　0.6006　6,高階級
    - 幸運　0.5118　5,進階級
    - 慚愧　0.4881　6,高階級
    - 得意　0.4841　5,進階級
    - 難過　0.4763　4,基礎級
    - 光榮　0.4542　6,高階級
    - 驕傲　0.4538　5,進階級
    - 驚訝　0.4485　5,進階級
    - 抱歉　0.4419　7,流利級
    - 懊惱　0.4385　7,流利級
    - 失望　0.4362　6,高階級

小結

- 利用「例句編輯工具」可以快速找出太難的用詞
- 利用「語義場關聯詞系統」可以：
  - 找出難詞的替換詞
  - 出考題、出作業

肆、漢字與構詞

如何觀察詞首或詞尾的使用情況？

- 目的
  - 觀察詞首或詞尾的使用
  - 例如：
    - 觀察詞首：副總統、副主席、副總裁、....
    - 觀察詞尾：藝術家、物理學家、化學家、...
  - 有沒有辦法查所有：副XX，XX家？
- 方法：簡式查詢提供的構詞萬用字：
  - ？：代替一個字
    - 心無旁? → 心無旁驚、心無旁驚
    - 副??→副總統、副主席、副總裁、....
  - * ： 代替 0~N 個字元
    - 副* → 副，副手，副作用，副總統，副院長，....
  - + ： 代替 1~N 個字元
    - 副+ → 副手，副作用，副總統，副院長，...
  - [A,B,...]：任選一個字
    - [台,臺]灣 → 台灣、臺灣

配合：Frequency Breakdown

- 詞彙分布觀察：觀察不同詞形，不同詞性的比例，例如：
- 統計詞形的分布
  - [台, 臺]灣

No.	Search result	No. of occurrences	Percent
1	台灣	28648	80.79%
2	臺灣	6812	19.21%

- 統計詞性的分布
  - 研究

No.	Search result	No. of occurrences	Percent
1	研究_Na	39974	74%
2	研究_VE	14046	26%

用途一：觀察錯別字、異體字

- 觀察錯別字
  - 心無旁? → 心無旁騖, 心無旁鶩
  - 好高?遠 → 好高騖遠, 好高鶩遠
  - 口乾舌? → 口乾舌燥, 口乾舌躁, 口乾舌噪
  - 相形見? → 相形見絀, 相形見拙
  - 練習：
    - 意興??
    - ?然欲泣
    - 篳路??
- 觀察異體字的使用
  - [台, 臺]灣
  - 練習：
    - 裡 vs. 裏
    - 蹤 vs. 踪
    - 豔 vs. 艷

用途二：觀察詞首、詞尾的用法

- 觀察詞首用法
  - 副?
  - 副*
  - 副+
  - 副**
  - 副+*
  - 副++
- 觀察詞尾用法
  - **練習詞尾觀察：**
    - 「家」、「機」、「會」、「節」、「系」

用途三：觀察字的構詞情況

- 構詞觀察：
  - ***故*** → 故事、故意、事故、故障
  - ***鶩*** → 趨之若鶩，心無旁鶩
  - ***騖*** → 好高騖遠，心無旁騖
  - ***絀*** → 相形見絀，左支右絀，短絀

用途四：易混淆字的辨析

- 用字辨析
  - 蹟 vs. 跡

奇蹟	痕跡
事蹟	跡象
古蹟	軌跡
神蹟	足跡
史蹟	蹤跡
遺蹟	血跡
奇蹟式	字跡
名勝古蹟	遺跡
墨蹟	筆跡
聖蹟	蛛絲馬跡

- 練習：
  - 作 vs. 做
  - 份 vs. 分
  - 佈 vs. 布
  - 拼 vs. 拚

用途五：觀察構詞規則

- 千？萬？
  - 千變萬化，千辛萬苦，千言萬語，千頭萬緒，千真萬確，千軍萬馬
- 相對字
  - *裡?外* → 裡裡外外, 吃裡扒外, 霧裡雲外 . . .
  - *上?下* →上上下下， 七上八下， 跳上跳下, 上吐下瀉 . . .
  - 練習：來去，前後，出入，男女
- 數字成語
  - *一?一* → 一模一樣, 一舉一動, 一點一滴 . . .
  - 練習：一二、二三、三四、四五. . . .
- 否定
  - *不?不* →不知不覺, 不慌不忙, 不折不扣 . . .
  - 練習：無無、不無、無不

用途六：觀察詞性分布

- 查詢「把」
- 選擇 Frequency breakdown
- 選擇 Frequency breakdown of words and annotation

No.	Search result	No. of occurrences	Percent
1	把_P	12001	94.64%
2	把_Nf	668	5.27%
3	把_VC	9	0.07%
4	把_Na	3	0.02%

No.	Search result	No. of occurrences	Percent
1	被_P	16249	99.86%
2	被_Na	16	0.1%
3	被_VJ	5	0.03%
4	被_VC	2	0.01%

伍、語法點查詢

短語結構查詢 I

- 使用情境：對 XXX 來 說
- 以萬用字填充不確定成分
  - 萬用字單獨使用時的意義
    - ？：代表任意單字詞
      - **原義**：代表任一字
      - **單獨使用**：代表單字詞
      - **例：對 ？ 來 說** → 對 我 來 說，對 他 來 說
    - ＋：代表任意1個詞
      - **原義**：代表 1 到多個字
      - **單獨使用**： 代表1 字詞到 n 字詞，(亦即1個詞)
      - **例：對 ＋ 來 說** → 對 我 來 說，對 我們 來 說
    - ＊：代表任意0或1個詞
      - **原義**：代表 0 到多個字
      - **單獨使用**： 0 字詞到 n 字詞，(亦即0個詞或1個詞)
      - **例：對 ＊ 來 說** → 對 我 來 說，對 我們 來 說
  - 萬用字的使用組合
    - 對 ＊ ＊ 來 說
    - 對 ＋ ＊ 來 說
    - 對 ＋ ＋ 來 說
    - 對 ＋＊＊＊＊＊＊ 來 說

短語結構查詢 II

- 短語結構中的不確定成分，可以透過萬用字填充
  - 對 ＋ ＊ 來 說
  - 對 ＋ ＊ ＊ 來 說
  - 對 ＋ ＊ ＊ ＊ 來 說
- 小括弧：結構的重覆
  - 對 ＋＊ 來 說 → 中間夾1~2詞
  - 對 (＋)＊ 來 說 → 中間夾0~n詞
  - 對 ＋＋ 來 說
  - 對 (＋)＋ 來 說
  - 對 (＋ ＋)＋ 來 說

結構的重覆：符號定義

- ? ：結構重複 0 或 1次
  - 打 (+)? 電話 → 打 電話, 打 過 電話, ...
  - 注意：與構詞的 ? 定義 不同
- * ：結構重複 0 到 n 次
  - 打 (+)* 電話 →
    - 打 電話, 打 過 電話, 打 了 通 電話,...
- + ：結構重複 1 到 n 次
  - 打 (+)+ 電話 →打 過 電話, 打 了 通 電話,...
- {x,y}：結構重複 x 到 y 次
  - 打 (+){1,4} 電話 → 打 了 好多 通 電話, ...

結構的重覆：詞性的限制

- 限制中間的詞性
  - 對 _Nh 來 說
  - 對 (_Nh)+ 來 說
  - 對 + _Nh 來 說
  - 對 +* _Nh 來 說

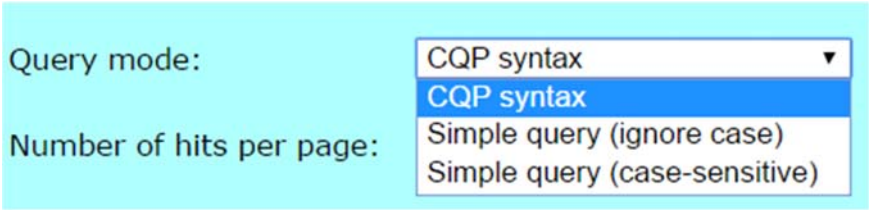
Simple Query 的極限

- 無法排除某些字
  - *一+二* → 一九二六、一百二十
- 查詢結構會跨句
  - 例：一 (+)+ 就
- 無法查詢重疊詞
  - AABB：大大小小、乾乾淨淨、清清楚楚
  - AAB：爬爬山，掃掃地，揮揮手，搖搖頭
  - ABB：一家家，一篇篇，陰沉沉，慢吞吞
  - ABAB：巴結巴結，練習練習，悠哉悠哉，研究研究
  - AA：常常，滾滾，統統，深深
- 使用 CQP 語法解

陸、CQP 查詢語言

CQP 查詢語言簡介

- CQP: Corpus Query Processor (Evert and Hardie, 2011)
  - 提供細緻的查詢功能
  - 使用上較複雜
- 使用
  - 在查詢介面中切換 Query mode → CQP syntax



CQP查詢式

- 例子：
  - [word="把"] ← 查詢「把」
  - [word="把" & pos="Nf"] ← 只要量詞的「把」
  - [word="把" & pos!="Nf"] ← 只要非量詞的「把」
  - [word="把" & !(pos= "Nf")] ←同上
  - [word="把" & (pos="Nf" | pos="Na")] ← ?
- **練習：**
  - 查詢非名詞且非動詞的「把」

CQP 萬用字

- . ：代替一個字
  - [word="好高 .遠"]
- * ： 前字重複 0~N 次
  - [word="哈*"] → 哈，哈哈，哈哈哈
  - [word="副.*"] → 副，副手，副作用，副總統，副院長，....
- + ： 代替 1~N 個字元
  - [word="副.+"]→ 副手，副作用，副總統，副院長，...
- **[AB...]：任選一個字**
  - [word="[台臺]灣"]
  - 注意：「台」和「臺」中間沒有逗號
- **[^AB...]：否定一組字**
  - [word="一 [^百千萬] 二 [^百千萬]"] →一清二楚，~~一千二百~~，一石二鳥
  - 限制一和二之間不能夾「百」、「千」、「萬」
- **"(X|Y)Z"：X,Y 任選一組**
  - [word="(台北|高雄)市"]

CQP查詢式的語法

- 符號
  - 方括號 []：一組括號代表一個詞
- 括號中的條件：限制該詞的性質
  - **word="A"**：限制詞形為 A
  - **pos="B"**：限制詞性為 B
  - **pos!="B"**：限制詞性不得為 B
  - **X & Y**：兩個條件都要符合
  - **X | Y** ：任意條件符合即可
- 註：若沒有限制條件，代表任意詞
  - 例如： [word="對"] [] [word="來說"]
  - 相當於Simple Query： 對 + 來說

CQP 語法於構詞上的觀察

- 用途一：觀察錯別字、異體字
  - [word="心無旁."] → 心無旁騖, 心無旁鶩
- 用途二：觀察詞首、詞尾的用法
  - [word="副.*"] → 副總統, 副主席 ...
- 用途三：觀察字的構詞
  - [word=".*驁.*"]
- 用途四：易混淆字的辨析
  - 蹟 vs. 跡
- 用途五：觀察規則性構詞
  - [word="千.萬."]
- 用途六：觀察詞性分布
  - [word="把"]

CQP 語法於結構的重覆

- ? ：結構重複 0 或 1次
  - [word="打"] []? [word="電話"] →
    - 打 電話, 打 過 電話, ...
- * ：結構重複 0 到 n 次
  - [word="打"] []* [word="電話"] →
    - 打 電話, 打 過 電話, 打 了 通 電話,...
- + ：結構重複 1 到 n 次
  - [word="打"] []+ [word="電話"] →
    - 打 過 電話, 打 了 通 電話,...
- {x,y}：結構重複 x 到 y 次
  - [word="打"] []{1,4} [word="電話"] →
    - 打 了 好多 通 電話, ...

CQP 語法於詞性的限制

- 限制結構中間詞語的詞性
  - [word="對"] [pos="Nh"] [word="來"] [word="說"]
  - [word="對"] [pos="Nh"]+ [word="來"] [word="說"]
  - [word="對"] [] [pos="Nh"] [word="來"] [word="說"]
  - [word="對"] []* [pos="Nh"] [word="來"] [word="說"]

CQP 語法簡化

- 若 CQP 語法只限制詞語時，指令可簡化
- 例如：
  - [word="對"]
    → "對"
  - [word="對"] [pos="Nh"] [word="來"] [word="說"]
    →"對" [pos="Nh"] "來" "說"
  - [word="對"] [pos="Nh"]+ [word="來"] [word="說"]
    →"對" [pos="Nh"]+ "來" "說"
  - [word="對"] [] [pos="Nh"] [word="來"] [word="說"]
    → "對" [] [pos="Nh"]+"來" "說"
  - [word="對"] []* [pos="Nh"] [word="來"] [word="說"]
    → "對" []* [pos="Nh"]+ "來" "說"

以 CQP 語法解決Simple Query 極限 I

- Simple Query無法排除某些字
  - *一+二*
    → 一清二楚，一千二百，一石二鳥
  - CQP等效查詢：[word=".*一.+二.*"]
  - CQP排除查詢： [word="一 [^百千萬] 二 [^百千萬]"]
    → 一清二楚，~~一千二百，~~ 一石二鳥
- Simple Query查詢結構會跨句
  - 一 (+)+ 就
    → 一 台 汽車。...。 就
  - CQP 等效查詢："一" []+ "就"
  - CQP 排除查詢：
    - "一" [word!="，|。|！|？|；|："]+ "就"
    - "一" [word!="[，。！？；：]"]+ "就"
    - "一" "[^，。！？；：]"+ "就"

以 CQP 語法解決Simple Query 極限 II

- 查詢重疊詞
  - AABB：大大小小、乾乾淨淨、清清楚楚
  - [word="...."&char(word,0)=char(word,1)&char(word,2)=char(word,3)]
    - word="...." →四字詞
    - char(word,0)=char(word,1)→1,2字相同
    - char(word,2)=char(word,3)→3,4字相同
    - & → 條件都要符合
  - 練習：
    - AAB：爬爬山，掃掃地，揮揮手，搖搖頭
    - ABB：一家家，一篇篇，陰沉沉，慢吞吞
    - ABAB：巴結巴結，練習練習，悠哉悠哉，研究研究
    - AA：常常，滾滾，統統，深深

參考資料

- 指令速查表：
  - http://coct.naer.edu.tw/cqpweb/doc/指令速查表.pdf
- CQP Query Language Tutorial
  - http://cwb.sourceforge.net/files/CQP_Tutorial.pdf
- CQP User's Manual
  - http://corpora.dslo.unibo.it/TCORIS/cqpman.pdf
- CQP 教學影片(英文)
  - https://www.youtube.com/user/CorpusWorkbench