



2024華語文語料與能力基準應用參考指引工作坊(一)

華語文語料庫與能力基準 整合應用系統實作B

白明弘

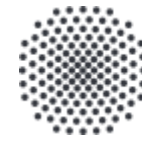


2024.03.20

國家教育研究院語文教育及編譯研究中心

壹

索引典系統



CWB (1990)

- 德國斯圖加特大學自然語言處理研究所(IMS)
- 開發 The IMS Open Corpus Workbench (CWB)
- 查詢引擎 (Corpus Query Processor, CQP)
 - [word="chair.*" & pos!="N.*"]
- 發展出一套查詢語法：
 - CQP
 - CQL
- 缺點
 - 完全命令列模式
 - > [pos="DT"] [pos="JJ"]{2,} [pos="NNS?"];
 - > sort by word %cd on match[1] .. matchend[-1];

BNCweb (2000年)



- 瑞士蘇黎世大學 (University of Zurich)
 - Hans Martin Lehmann, Sebastian Hoffmann and Peter Schneider
- 為 BNC 語料庫提供查詢介面
- 系統
 - 2000 年完成系統，2002年開源，SARA伺服器軟體
 - 2006 年改以 CWB 做後端
- 缺點
 - BNCweb 無法使用在 BNC 以外的語料庫

CQPweb (2008)

- 英國蘭卡斯特大學 Andrew Hardie
- 開發動機
 - 語料庫系統介面不一致是學習障礙。
 - BNCweb → WordSmith, AntConc
 - 需要花相當多時間適應系統
- 目標
 - 開發新系統，模仿 BNCweb 的所有功能
 - 可以應用在任何語料庫上
 - 使用者只需要熟悉一套系統

Main menu

BNCweb (CQP-Edition)

Query options

Standard query

[Written restrictions](#)

[Spoken restrictions](#)

User-specific functions

[User settings](#)

[Query history](#)

[Saved queries](#)

[Categorized queries](#)

[Make/edit subcorpora](#)

[Upload external data file](#)

Additional functions

[Browse a text](#)

[Scan keywords/titles](#)

[Explore genre labels](#)

[Frequency lists](#)

[Keywords](#)

Query mode: [Simple Query S](#)

Number of hits per page:

Restriction:

BNCweb (CQP-edition) © 1996-2008

You are logged in as use:

Figure 1. The welcome-page of BNCweb (compare Figure 2)

Menu

Corpus queries

Standard query

Restricted query

Word lookup

Frequency lists

Keywords

User controls

User settings

Query history

Saved queries

Categorised queries

Create/edit subcorpora

Corpus info

View corpus metadata

Corpus documentation

Oxford Simplified Tags

Lemma/OST

CI AWS7 Tanset

Brown Family of Corpora: powered by CQPweb

Standard Query

Query mode: [Simple query language syntax](#)

Number of hits per page:

Restriction:

System messages

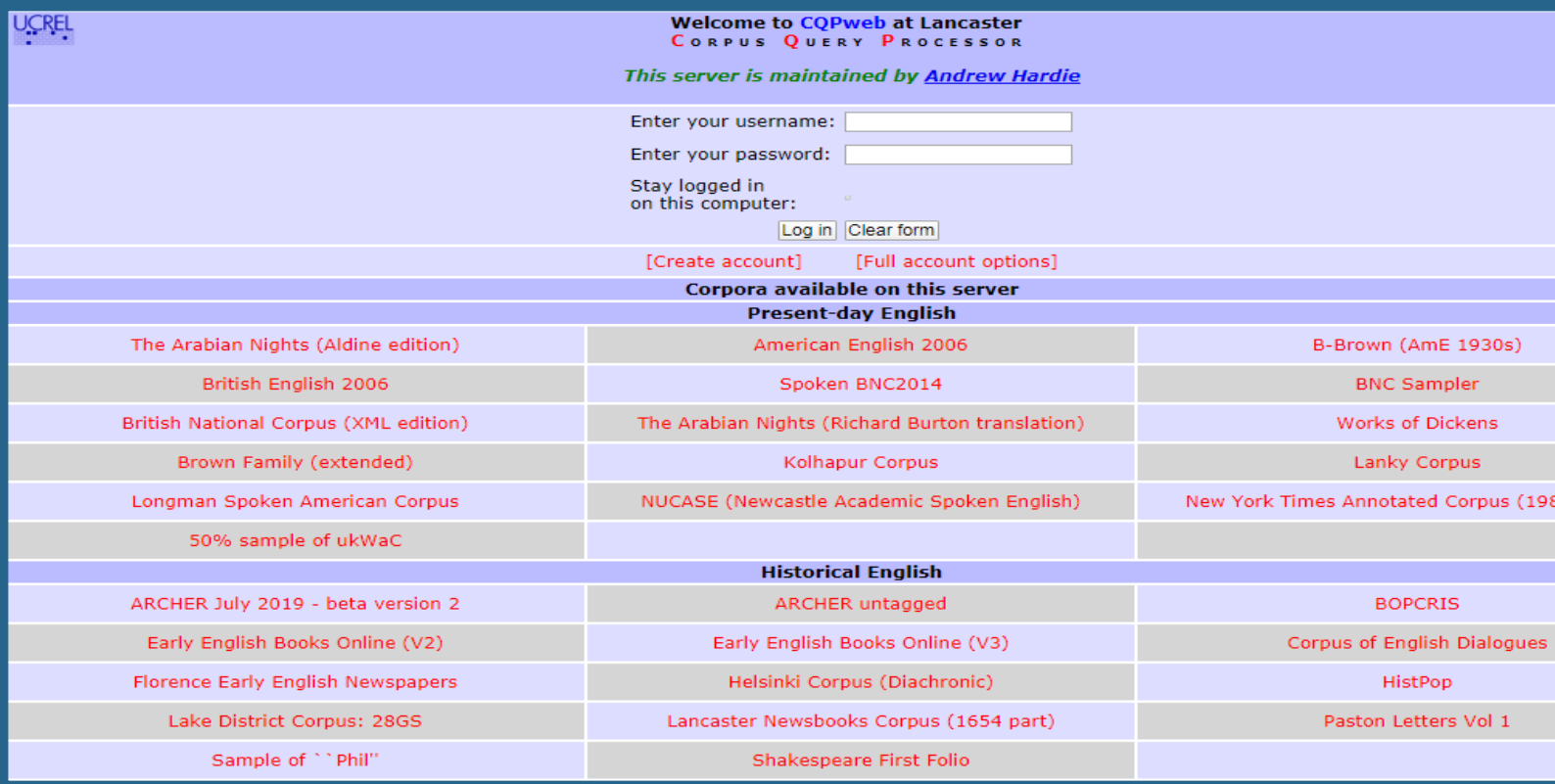
Database problems

2010-11-30 We are having database problems, you might not be able to connect... if not, please wait a while and try again. [x]

Figure 2. The welcome-page of a corpus indexed in CQPweb

CQPweb

- 英國蘭卡斯特大學 <https://cqpweb.lancs.ac.uk/>
- 提供數10種語言，100多種語料庫



UCLREL

Welcome to CQPweb at Lancaster
CORPUS QUERY PROCESSOR

This server is maintained by Andrew Hardie

Enter your username:

Enter your password:

Stay logged in on this computer:

[\[Create account\]](#) [\[Full account options\]](#)

Corpora available on this server

Present-day English

The Arabian Nights (Aldine edition)	American English 2006	B-Brown (AmE 1930s)
British English 2006	Spoken BNC2014	BNC Sampler
British National Corpus (XML edition)	The Arabian Nights (Richard Burton translation)	Works of Dickens
Brown Family (extended)	Kolhapur Corpus	Lanky Corpus
Longman Spoken American Corpus	NUCASE (Newcastle Academic Spoken English)	New York Times Annotated Corpus (198
50% sample of ukWaC		

Historical English

ARCHER July 2019 - beta version 2	ARCHER untagged	BOPCRIS
Early English Books Online (V2)	Early English Books Online (V3)	Corpus of English Dialogues
Florence Early English Newspapers	Helsinki Corpus (Diachronic)	HistPop
Lake District Corpus: 28GS	Lancaster Newsbooks Corpus (1654 part)	Paston Letters Vol 1
Sample of "Phil"	Shakespeare First Folio	

檢索語言 vs 選單式檢索

簡易查詢

■ 感染_VJ

CQP

■ [word="感染"&pos="VJ"]

選單式檢索

中央研究院平衡語料庫

請輸入你所要查詢的關鍵字

關鍵詞或重疊詞	詞類	特徵
感染	VJ (狀態及物動詞)	

檢索語言的優點

■ 查詢彈性非常大：

■ 詞構檢索：

■ 三?兩? → 三言兩語, 三三兩兩, 三天兩頭, ...

■ 短語

■ [pos="V.*"] [word="感染"] →

■ 受到 感染、遭到 感染、降低 感染、.....

■ 輸入速度快

■ 表達清晰

■ 沒有模糊空間

■ 不會迷失在數層的選單裡

■ 使用不同語料庫不必從頭學習

Word Sketch Engine

- 英國詞彙計算公司 (Lexical Computing Ltd.)
- <https://www.sketchengine.eu/>
- CQL : Corpus Query Language

The screenshot displays the Word Sketch Engine interface. On the left is a navigation menu with options like 'Concordance', 'Word List', 'Word Sketch', 'Thesaurus', 'Find X', 'Sketch-Diff', 'Sketch-Eval', and 'Corpus Info'. The main area shows a search query for 'colour' with 16,486 results (147.0 per million). The results are displayed in a table with columns for the query, the word 'colour', and the context. The word 'colour' is highlighted in red in the original image. The interface includes pagination controls (Page 1 of 825) and navigation links (Next, Last). The Lexical Computing logo is visible in the bottom right corner.

Query **colour** 16,486 (147.0 per million)

Page 1 of 825 Go Next Last

J2L	It would be tedious to list the types and	colours	of stone, ceramic etc. used at each site	
J2L	types of stone used for various shades of	colour	are predictable and limited in number.	
J2L	Birdcombe Avon. Here, sandstone furnished a buff	colour	, pennant stone a blue, liar the white for	
J2L	most mosaics comprise three to six basic	colours	, a work of good quality will include many	
J2L	therefore, to note ten or twelve different	colours	of tesserae in one pavement. In some, such	
J2L	the Woodchester Orpheus mosaic. </p> 3.2 The	colour	of Tesserae <p> Sensitive use of shading	
J2L	1976, 9). Elsewhere, intelligent use of	colour	is responsible for the blue shading which	
J2L	are notable. </p><p> Whilst considering the	colour	of tesserae it is also pertinent to mention	
J2L	: 0.5 cm. sq. and 1.5 cm. sq. </p><p> Like	colour	, the size of the tesserae affects the perspective	
J2L	fairly dark tesserae (deep red is a favourite	colour	, so producing a stronger" proximity effect	
J2L	panels (pl. 5b). At Leicester the rosettes -	coloured	(from the edges inwards) red, yellow and	
J2L	be cramped (although" loose"). There are	colour	contrasts however: the simple guilloche	
J2L	former. However, the-more subtle use of	colour	in the latter also produces a less contrived	
J2L	angular appearance. An overall poverty of	colour	, and the use of slightly larger (but still	
J2L	mosaic A). Although including the same basic	colours	, as well as tesserae of a similar size,	
J2L	blending of many tones of five or six basic	colours	, is notable in both designs. It is a sensitivity	
J2L	shows a generally consistent interlace of	colour	, one in every four tongues of the latter	
J2L	Oceanus panel (contrast the confusion of	colour	around the heads of the lion and stag)	
J2L	However, on balance, the use here of similar	colours	(red, yellow, grey, pale-blue, brown) and	
J2L	Street mosaic, the presence there of a richly	coloured	figured panel (enclosed by a chain-guilloche	

Page 1 of 825 Go Next Last

Lexical Computing

工作坊查詢系統分流

■ 投影片下載

■ <https://coct2.naer.edu.tw/download/2024/>



■ 系統忙碌時，請自動使用不同機器查詢：

■ <https://coct.naer.edu.tw/cqpweb>



■ <https://coct2.naer.edu.tw/cqpweb>



■ <https://coct3.naer.edu.tw/cqpweb>



國教院索引典系統

- <https://coct.naer.edu.tw/cqpweb>
- <https://coct2.naer.edu.tw/cqpweb>
- <https://coct3.naer.edu.tw/cqpweb>

The screenshot displays the NAER website interface, featuring a header with the logo and the text "國家教育研究院 National Academy for Educational Research 華語文語料庫與能力基準整合應用系統". Below the header, there are several main sections:

- 最新消息** (Latest News): A section for news updates.
- 臺灣華語文能力基準** (Taiwanese Chinese Language Proficiency Standards): A section detailing language standards, including a table for "第1級 TBCL 1" (Level 1 TBCL 1) with columns for "General", "Description", and "Listening".
- 國教院索引典系統** (NAER Index System): A section for the index system, highlighted with a red dashed circle. It includes a table for "國教院索引典系統(含檢索與內容) (國教院索引典系統)" with columns for "詞類", "詞性", "詞義", and "詞源".
- 華英雙語索引典系統** (Chinese-English Bilingual Index System): A section for a bilingual dictionary.
- 教材編輯輔助系統** (Textbook Editing Assistance System): A section for editing textbooks.
- 語義場關聯詞查詢系統** (Semantic Field Related Word Query System): A section for querying related words in semantic fields.
- 漢字分級標準檢索系統** (Character Level Standard Search System): A section for searching characters by level.
- 詞語分級標準檢索系統** (Word Level Standard Search System): A section for searching words by level.
- 語法點分級標準檢索系統** (Grammar Point Level Standard Search System): A section for searching grammar points by level.

登入

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)

Enter your username:

Enter your password:

Stay logged in
on this computer:

Guest account username:

password:

語料庫列表

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)

[\[English\]](#)

Welcome back to the CQPweb server.
You are logged in as [naer](#). [\[Log out\]](#)

Corpora available on this server

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 口語語料庫2021	COCT 中介語語料庫2016
COCT 中介語語料庫2019	COCT 書面語語料庫2016	COCT 書面語語料庫2017
書面語語料庫(工作坊專用)	COCT 書面語語料庫2019	COCT 書面語語料庫2020

語料庫查詢主畫面

選單	書面語語料庫(工作坊專用) [English]
語料庫查詢	標準查詢
標準查詢	<div style="border: 1px solid black; height: 150px; width: 100%;"></div>
限制查詢	
單詞查詢	
詞頻列表	
關鍵詞	
分析語料庫	
Saved query data	查詢模式: <input type="text" value="簡易查詢 (不區分大小寫)"/> 指令速查表 (English) 選單參考表
查詢歷史	每頁的查詢結果: <input type="text" value="50"/>
儲存查詢結果	Match strategy: <input type="text" value="Standard"/>
分類查詢結果	限制 (檢索範圍): <input type="text" value="None (search whole corpus)"/>
上傳查詢指令	<input type="button" value="開始查詢"/> <input type="button" value="重設查詢"/>
建立/編輯子語料庫	
語料庫資料	
查看語料庫後設資料	
<i>Corpus manual unavailable</i>	

指令速查表

基本詞彙查詢

- 查詢詞彙：在查詢框中輸入詞彙並按下 [Start Query]
 - 容易 → 容易
- 使用圓括弧及直線表「或者」。例如：
 - (自行車|腳踏車|單車) → 找出包含其中任一個詞的句子
 - 註：符號必須是「半形」
- 使用萬用字元查詢:

? 代替 1 個中文字元

心無旁? → 心無旁騖, 心無旁騖, ...

* 代替 0~n 個中文字元

一心* → 一心, 一心一意, 一心一德, 一心誠敬, ...

+ 代替 1~n 個中文字元

一心+ → 一心一意, 一心一德, 一心誠敬, ...

??+ 代替 3~n 個中文字元

??+家 → 心理學家, 物理學家, 經濟學家, 生物學家, ...

簡易查詢：第一次查詢就上手

COCT 書面語語料庫2017

[\[English\]](#)

標準查詢

一味

查詢模式:

簡易查詢 (不區分大小寫) ▾

[指令速查表](#) (English) [選單參考表](#)

每頁的查詢結果:

50 ▾

Match strategy:

Standard ▾

限制 (檢索範圍):

None (search whole corpus) ▾

開始查詢

重設查詢

簡易查詢：第一次查詢就上手

Your query “一味” returned 4,645 matches in 3,924 different texts ^[English]
words [159,051 texts]; frequency: 15.291 instances per million words
[retrieved from cache]

Solution 1 to 50

Page 1 / 93

抱著隨它，或等以後再說，或	<u>一味</u>	禁止，或閉口不談，或破口大罵，或
責備自己。所謂逃避現實是指，	<u>一味</u>	夢想比現在更好的時空，想像不可能
少年不問我要什麼、我能什麼，而	<u>一味</u>	接受父母師長的安排，他們沒有度過
消除貧窮，為了在同事面前炫耀，	<u>一味</u>	強迫黃黃念醫科，到五年級時，黃黃
著隨它去、或等以後再說、或	<u>一味</u>	禁止、或閉口大罵、或鼓掌叫好的態
不可以呢？中國人的智慧中國人所以	<u>一味</u>	學理工，也可能和中國人的智慧有關
灣似乎少有這種情形，強調的只是	<u>一味</u>	的讀書。由吳靜吉個人作出發點，對
或來撒野的。自己心理不正，怎可	<u>一味</u>	假設別人也心不正呢？這種溝通方式
當做成熟的成人時，服務者任性或	<u>一味</u>	可愛，或擺出母夜叉的雄姿，或表

簡易查詢：(A|B)

標準查詢

(一味|一昧)

走太遠。面對這個世界，你不能一味地迎合，要敢於任性地加入一點新
天，不管什麼事業，不問可行性，一味地堅持，真的好嗎？有時候放棄也
（流行、時尚、品牌）為主臬，一味的追求名牌與高檔貨，讓人看了不
效為父母如果能幫助孩子，而不是~~一味~~地斥責、拒絕孩子的要求，卻從不
事端，但萬一遭受挑釁我也不認為一味忍氣吞聲就能夠大事化小、小事化
並由斯年催促無數次。乃余所長一味推諉，率領全所研究員來總處與其

頻率分解(frequency breakdown)

隨機排序

Choose action... ▾

Choose action...

新查詢

查詢結果隨機取樣...

頻率分解

詞彙分布

分散度

排序

搭配詞...

分類...

下載...

Tabulate...

Expand or narrow...

Save this query result...

0 Page 1 / 2

認定，如果今天守寡的是自己，一定

討好他人情緒易暴走的人無法與人對

討好對方，放大檢視他的言行態度，

討人歡心。我的一個女客戶就是

「忍受」，這對妳來說是種「

地想要尋求一條簡單預測漲跌的公式

但想這個事實已經完成，世界上的所有工作要

頻率分解(frequency breakdown)

No.	Query result	No. of occurrences	Percent
1	一味	66	97.06%
2	一味	2	2.94%



練功時間

- 請同時查出「儘管」和「儘管」的例句
- 請比較這兩個詞的頻率

使用 CQP 查詢一定要記得！

中

查詢模式：
每頁的查詢結果：
限制（檢索範圍）：

簡易查詢（不區分大小寫）
CQP語法
簡易查詢（不區分大小寫）
簡易查詢（區分大小寫）

無（檢索整個語料庫）

指令速查表 選單參考表

開始查詢 重設查詢

CQP：方框代表一個詞

■ [word= "笑死"]

面有一盤大陸妹這個菜擺著五十塊	笑死	我了快要在家裡切辣的東西這張
當剪接師沒有啦你不要在鏡頭前說	笑死	很花時間耶晚上叫她去睡覺都不
辦我沒有帶觀星盤然後老師就快要	笑死	了他就說老師說我只是想要你告
像有一點模糊不行這個出去要被人	笑死	TALVEN 剛剛那一個是N啦N寫了
月就鹹的我猜這個人一定不會游泳	笑死	了啦不對這可能是某個人的邪惡
質疏鬆拜託這一定是寫錯啦對不對	笑死	人了啦這個報導一點都沒有錯啊
士你講的好像卡通影片喔對啊真是	笑死	人了你知道珊瑚礁就是海底的建
不是有人坐飛機在天上放的」	笑死	我了坐飛機放煙火怎麼可能太好
繞得能看嗎不能看播出會害怕給人	笑死	當然是得獎是最好重點我們努力不
間開始訓練體能小波你體力很差耶	笑死	人了你還不是一圈半就掛了真是
心大意來敲門結果他就摔了一大跤	笑死	人了方丈來了方丈好好你們那麼

CQP 為什麼需要方框？

- 為何不用 笑死
- 而用 [word="笑死"]
- 或 "笑死" ？



因為 CQP 提供你更多的功能

如果你需要限制詞性

- [word="委屈"&pos="Na"]
- [word="委屈"&pos="V.*"]

如果你需要查同義詞(或異形詞)

- [word="委屈" | word="委曲"]
- 配合詞頻分解可以瞭解使用頻率

No.	Search result	No. of occurrences	Percent
1	委屈	142	94.67%
2	委曲	8	5.33%



練功時間

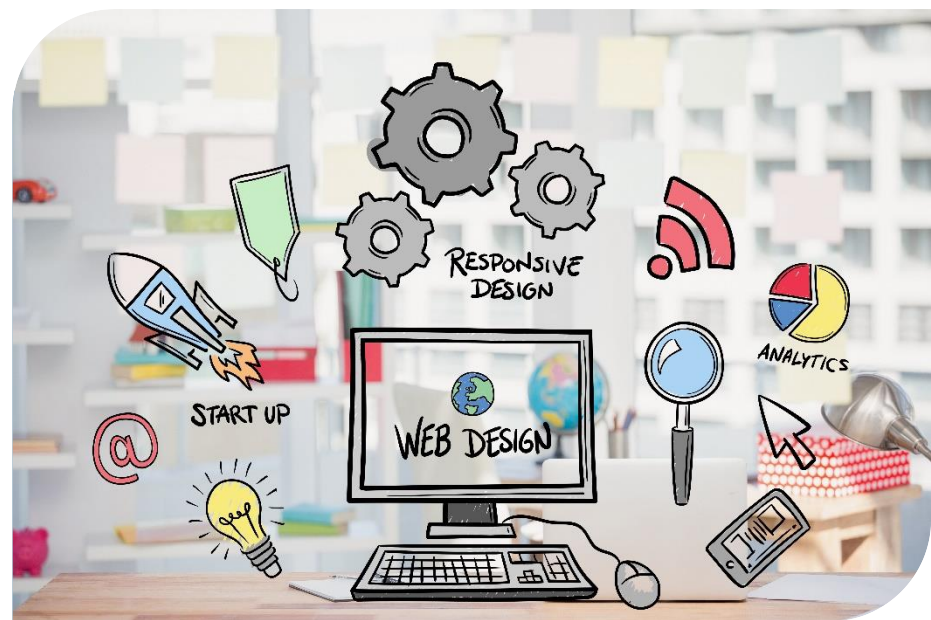


練習

■ 你平常用

- 「不只」
- 「不止」

- ### ■ 請利用 CQP 查詢， 觀察這兩種寫法在語料庫中的 使用比例



CQP：2個方框代表2個詞

■ [word="感到"] [word="委屈"]

你欺負。你有什麼地方	感到 委屈	想說的，告訴我，
無奈等情緒，而女方則	感到 委屈	、疲累、有罪惡感
鼠狼以偷雞為生，實在	感到 委屈	，便見玉皇大帝，
不便和挫折。當我為她	感到 委屈	時，女兒安慰我說
？」卡堅卡像大人似地	感到 委屈	。「這都是別人的
要宣洩而出，最讓她	感到 委屈	的就是愛達說她不
的信，雖然我也替你	感到 委屈	，但我相信十個少
認識之媒體記者當然會	感到 委屈	失望。同樣的道理
道消息。錢瓊光也許是	感到 委屈	，還想說什麼，旁
。公主對這種做法原本	感到 委屈	，但眼見宋珍一表

CQP 如何檢索「詞組」？

方框代表一個詞

- [] 匹配一個詞
- [] [] 匹配兩個詞
- [] [] [] 匹配三個詞

填入內容表示對匹配詞的「限制」

- [] [word="電話"] 限制第2詞只能是電話
- [pos="N.*"] [word="電話"] 第1詞只能是名詞

觀察詞組

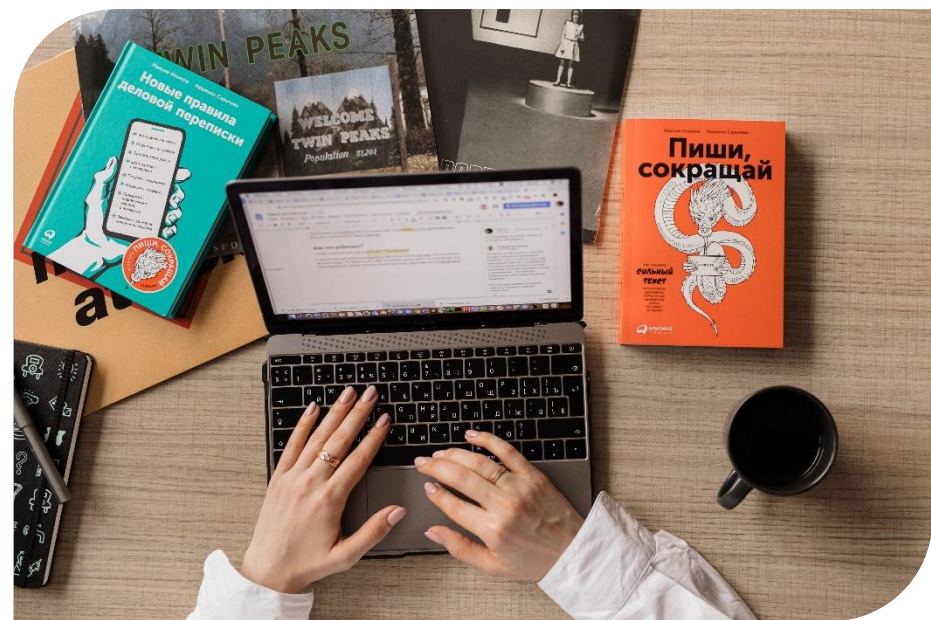
觀察量詞「把」可以接那些名詞

- [word="把"] []
- [word="把"&pos="Nf"] []
- [word="把"&pos="Nf"] [pos="Na"]

動彈，接著又點燃一	把火	，創造截止期限逼近
以我說演化思想是一	把尺	。你到底是更相信自
嗎？因此，正回饋這	把火	並不是那麼容易點起
和古人區別開來的一	把尺	。我們先不談那些大
音劍，跟了出來。這	把劍	，是用來把牛刺死的
菜與一疊餐巾布、一	把火雞	大頭針（用來代替一
而不同●染深：抓一	把核桃	殼，與兩杯水混合製
製鍋鏟。●刀具：一	把削皮	刀4、一把主廚刀5



練功時間



練習

- 請找出「很」後面接「Na」的例句
 - 例如：「很耐心」
- 請找出「很」後面接「Nc」的例句

複雜的查詢難不倒 CQP

- 語法點：到.....去/來+VP
 - 你要到哪裡去旅行？
 - 他們到臺灣來學中文。
- [word="到"] [word!="[, , °]"]{1,5} [word="來|去"]
[pos="V.*"]

到 鄉鎮 去 服務

到 各地 去 了解

到 放手 讓 弟弟 和 我 去 選擇

到 任何 詞彙 來 形容

到 法國 的 海濱 實驗室 去 做

到 陝北 的 山溝 裡 去 參加

到 船 上 來 看看

到 昆明 去 籌辦

CQP 提供你更複雜的功能

■ 簡單查詢無法達成的功能：

查出所有：一 M 一 M 結構

■ 例如：

■ 一陣 一陣

■ 一座 一座

■ 一箱 一箱

```
"—" a:[] "—" b:[] ::a.word=b.word
```

查出所有：A not A結構

■ 例如：

■ 是 不 是

■ 對 不 對

■ 有 沒 有

■ 好 不 好

■ 算 不 算

```
a:[] [word="不"|word="沒"] b:[] ::a.word=b.word
```

貳

語義場關聯詞系統

教材編輯輔助系統

最新消息

國家教育研究院
華語文語料庫與能力基準整合應用系統

最新消息

新的資訊

臺灣華語文能力基準

國家教育研究院
臺灣華語文能力基準

簡介 | 第1級 | 第2級 | 第3級 | 第4級 | 第5級 | 第6級 | 第7級

第1級 TBCL 1

整體描述
General Description

能理解並運用與個人生活相關的簡單字詞。
包含246個漢字、401個詞條、6個詞類標、15個語法點。
Able to understand and use simple words and phrases relating to one's daily life. It contains 246 Chinese characters, 401 words, 6 affixes, and 15 grammar points.

能聽懂以簡單字詞表達之個人生活相關訊息。
Able to understand information conveyed through the use of simple words and phrases and having to do with one's personal life.

聽
Listening

L-1-1 能聽懂簡單的數量名詞(數字、價錢、時間)。
Able to understand simple numbers and quantities (numbers, prices, time).

L-1-2 能聽懂以簡單字詞介紹的個人相關資訊(興趣、住址、學經歷、家庭狀況)。
Able to understand information conveyed through the use of simple words

國教院索引典系統

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)

目前使用者: naer [退出]

中研院平權語料庫4.0	COCT 口語語料庫2015	COCT 口語語料庫2017
國教院華語中介語索引典系統	COCT 書面語料庫2015	COCT 書面語料庫2017
COCT 書面語料庫2015		

華英雙語索引典系統

查詢詞: 教育 查詢

Sample Panel

1. I think that a nation can't do without elite education. But elite education can't be the only thing available. If elite education can cultivate some outstanding figures, such as Einstein, or Lee Yuan tseh, this is a fortunate thing for society, because they won't only do good things for themselves. 我認為，一個國家不能沒有菁英教育，但也不能只有菁英教育，如果菁英教育能培養出這位「偉大」人物，例如愛因斯坦、李遠哲，就是社會的福氣，因為他們不會只顧自己。

2. As dual education moved closer to mainstream education in the 1980s, education departments developed "character signs" that made teaching easier. Formed by the hands in the shape of individual Chinese characters and used sequentially in sentences that adhere to Chinese syntax, character signs differ significantly from natural signs. 民國70年代，隨著雙語教育向主流教育靠攏，教育單位研發出方便教學的「文字手語」，依著一組中文語句的結構順序，以手勢與字跡打出來，兩者並行甚大。

3. The ROC Constitution stipulates that each Chinese citizen has the right to receive an education. Principal Chou believes that the government is pushing a "look south" policy of improving relations with the nations of Southeast Asia, yet it has not solved the problem of educating the children of Taiwanese businessmen

Translation Panel

education (2403)
educational (5410)
educated (107)
school (85)
teaching (65)
training (91)

Collocation Panel

教育 [N]
[ADJ-]
[B]-
[+R-]
-下
教育 [N]
[AS-]
[A-]
[A-]
[A-]
大學... 學校... 子女...

教材編輯輔助系統

教材編輯工具操作區

自動學習 * 語料庫詞彙分析

好聽友: 買了「好」幾本「科」, 感「色」的「志」子「帶」你「色」的「道」, 會「知」拿「到」我的「看」好「...」

• 標音選到「注音」怎麼辦?
可將文本複製到下列ToneOO網站, 會自動標音以及提供聲母、末音、ToneOO
標音與字彙標註的詳細解說: [ToneOO 說明書](#)

• 想使用「標音字型」請點以下連結下載
注意: 標音與字型字體, [請點這裡](#)

X	1	X	X
X	1	X	X
1	X	X	X
1	7	X	X
1	7	X	X
5	X	25	1
5	X	25	1
X	1	15	3
X	X	X	X
X	X	X	X

語義場關聯詞查詢系統

查詢詞(詞)語義場關聯詞

輸入

正關聯詞(組):
感覺

負關聯詞(組):
請輸入負關聯詞, 例如: 買伯斯

輸出到數: 10
送出

語義場關聯詞:

感受 0.7342	6.高階級
覺得 0.6458	1.準備一級
感到 0.5878	6.高階級
情緒 0.5664	6.高階級
身體 0.5535	2.準備二級
心緒 0.5337	4.基礎級
恐懼 0.5082	7.流利級
反應 0.5058	5.進階級 6.高階級
舒服 0.4925	3.入門級
渴望 0.4866	5.進階級



國教院語義場關聯詞查詢系統

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

選擇分級詞表

- 國教院分級詞表

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 4 級(含)以下詞

查詢詞(組)語義場關聯詞

輸入

正關聯詞(組) :

一味

負關聯詞(組) :

請輸入負關聯詞，例如：賈伯斯

輸出詞數:

輸出

一味	0.7961	--
盲目	0.5943	6,精熟
過分	0.5713	4*,進階
過份	0.5489	4*,進階
短視	0.5254	--
刻意	0.5201	6,精熟
不能	0.5149	--
老是	0.5135	4,進階
反而	0.5135	5,進階
而是	0.5072	--
胡亂	0.5047	6,精熟
輕率	0.5041	6,精熟
捨本逐末	0.5033	--
譁眾取寵	0.4998	--
任由	0.4983	6,精熟
閉門造車	0.4977	--

語義場關聯詞:

關聯圖

複雜的查詢難不倒 CQP

查詢詞(組)語義場關聯詞

輸入

正關聯詞(組) :

一味

負關聯詞(組) :

請輸入負關聯詞，例如：賈伯斯

輸出詞數: 30

送出

輸出 ⓘ

一味	0.7961	--
盲目	0.5943	6,精熟
過分	0.5713	4*,進階
過份	0.5489	4*,進階
短視	0.5254	--
刻意	0.5201	6,精熟
不能	0.5149	--
老是	0.5135	4,進階
反而	0.5135	5,進階
而是	0.5072	--
胡亂	0.5047	6,精熟
輕率	0.5041	6,精熟
捨本逐末	0.5033	--
譁眾取寵	0.4998	--
任由	0.4983	6,精熟
閉門造車	0.4977	--

語義場關聯詞:

同義詞 與 近義詞

下列詞是同義、近義或以上皆非？

- 嘗試 vs 嚐試
- 念書 vs 唸書
- 痛哭 vs 慟哭
- 稍後 vs 稍候
- 作法自斃 vs 咎由自取
- 杞人憂天 vs 庸人自擾
- 刻舟求劍 vs 按圖索驥
- 走馬看花 vs 浮光掠影
- 蘋果鳳梨 vs 鳳梨
- 蘋果 vs 鳳梨

詞彙語義關係是連續性的

兩個詞之間的語義關係

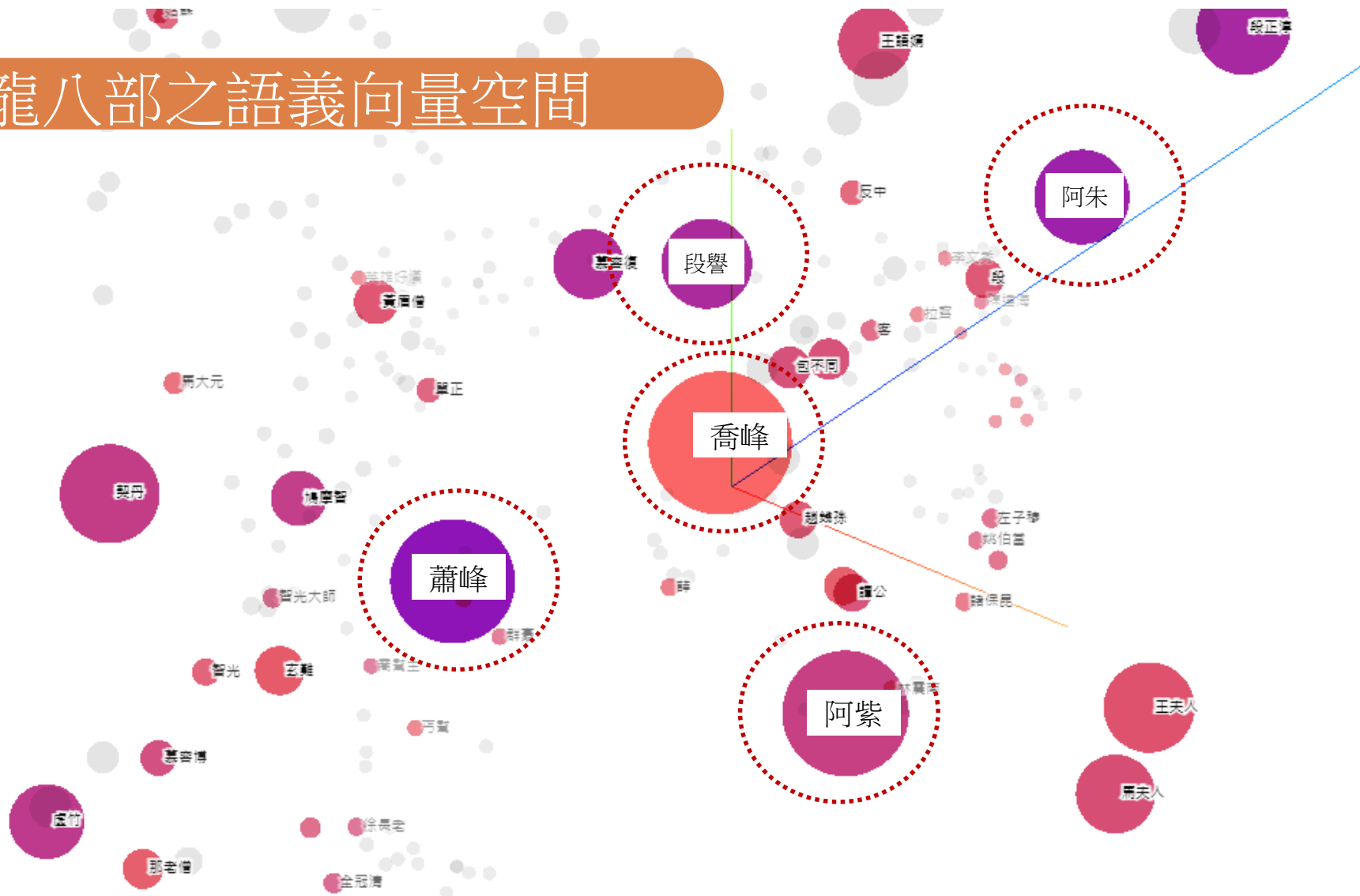
- 同義：嘗試 vs 嚐試
- 近義：杞人憂天 vs 庸人自擾
- 介於同義和近義之間：念書 vs 唸書
- 比近義關係更遠些： 棒球、足球、氣球

語義向量空間

- 語義關係不再只有同義、近義兩種
- 語義關係的遠近，是連續性的關係

詞彙語義關係是連續性的

天龍八部之語義向量空間



詞嵌入技術 (Word Embedding)

詞嵌入 (Word embedding)

- 是把詞映射為實數域上的向量。

背後原理：

- 詞的語義由所出現的語境決定

實作技術：

- 類神經網路

語義真的由語境決定嗎？

生平得意事蹟之第一項。若要問：「你	X	嗎？」周先生一定毫不遲疑地說：「
」周先生一定毫不遲疑地說：「我很	X	。」有什麼事能比得過成就了自己關心的
挖掘不盡的奇妙現象，一點發現就是一大	X	，所以沈潛其中的科學家有時候是很幸福的
來奮力墾荒，還覺得是大好天賜良機真是	X	得不得了。哇！在清華做這種事真
少。別人以為他很苦，但是他覺得很	X	。他一直找不到合用的材料，來做
我上街去逛逛。我也說：希望爸爸媽媽	X	健康，也希望妹妹活潑美麗，更希望我們一
一起唱唱歌。他們每天早睡早起，都非常	X	。24、我們都愛洗澡。廚房的洗碗槽裡
。一看到爸爸的笑容，我心裡就很	X	。2、懷念梅姊。夜深了，月光下的
人生的理想中找到為社會發展貢獻心力的	X	。四年後的今天，遠哲已結束三十二年
以定論。生活在台北，已經愈來愈不容易	X	，街道擠壓，日常休閒昂貴，生活忙又茫
是人生。也許有一天，你在裡面很	X	，不用跳出來，這也是人生。我好不容易有
事是由我所產生的，我希望她是	X	的，於是我們之間有意無意的開始躲避對方的
她，我現在也不知道怎麼辦，只希望她	X	些。我原本是怕會失去與她更深
知。難道妳已忘記當初我們在一起的種種	X	了，在那時我們都沒錢，日子過的
在那時我們都沒錢，日子過的清苦但	X	，最近常下大雨，每當我開車望著其它
??也沒啥好說的。只要妳過的	X	！我還抱怨什麼呢??中國陶瓷之發展
驗到存在於鄰人身上的神性；於是鄰人的	X	，也是我的快樂，過去那種覬覦別人
的神性；於是鄰人的快樂，也是我的	X	，過去那種覬覦別人成就的作為變得毫無
蔡：我也有同感，且認為作者很不	X	，以色彩語言來說，他們表現的卻是
善。李文秀心想：「他和她今天一定特別	X	，這麼熱鬧，這麼歡喜。」她心中的
卻沒猜對，蘇普和阿曼這時候並不特別	X	，卻是在特別的緊張。在火堆之旁
上班，晚上下班，已經很疲累了，怎會有	X	的情緒或愉快的心情？但是我回家之後聽
而不傷。」淫就是過度，很多人	X	之時不知收斂。以參加音樂會來說，

語義向量的用途

計算兩個詞的語義距離（相似度）

- distance (高興, 開心) = 0.73
- distance (高興, 難過) = 0.65
- distance (高興, 歡喜) = 0.63
- distance (高興, 雞蛋) = 0.0001

國教院語義場關聯詞查詢系統

查詢詞(組)語義場關聯詞

輸入

正關聯詞(組) :

高興

負關聯詞(組) :

請輸入負關聯詞，例如：賈伯斯

輸出詞數: 10

送出

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



練功時間



練習

- 使用語義場關聯詞找出和「**廚房**」相關的事物
- 試試看選擇不同的語料庫觀察
- 試試看限制詞彙等級

參

教材編輯輔助系統

教材編輯輔助系統

最新消息

國家教育研究院
華語文語料庫與能力基準整合應用系統

最新消息

新的資訊

臺灣華語文能力基準

國家教育研究院
臺灣華語文能力基準

簡介 | 第1級 | 第2級 | 第3級 | 第4級 | 第5級 | 第6級 | 第7級

第1級 TBCL 1

整體描述
General Description

能理解並運用與個人生活相關的簡單字詞。
包含246個漢字、401個詞語、6個詞類標、15個語法點。
Able to understand and use simple words and phrases relating to one's daily life. It contains 246 Chinese characters, 401 words, 6 affixes, and 15 grammar points.

能聽懂以簡單字詞表達之個人生活相關訊息。
Able to understand information conveyed through the use of simple words and phrases and having to do with one's personal life.

L-1-1 能聽懂簡單的數量名詞(數字、價錢、時間)。
Able to understand simple numbers and quantities (numbers, prices, time).

L-1-2 能聽懂以簡單字詞介紹的個人相關資訊(國籍、住址、學經歷、家庭狀況)。
Able to understand information conveyed through the use of simple words

國教院索引典系統

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)

目前使用者: naer [退出]

中研院平權語料庫4.0	COCT 口語語料庫2015	COCT 口語語料庫2017
國教院華語中介語索引典系統	COCT 書面語料庫2015	COCT 書面語料庫2017
COCT 書面語料庫2015		

華英雙語索引典系統

查詢功能: 查詢功能 | 查詢詞: [] | Search

Sample Panel

1. I think that a nation can't do without elite education. But elite education can't be the only thing available. If elite education can cultivate some outstanding figures, such as Einstein, or Lee Yuan tseh, this is a fortunate thing for society, because they won't only do good things for themselves. 我認為，一個國家不能沒有菁英教育，但也不能只有菁英教育，如果菁英教育能培養出這位「領袖」人物，例如愛因斯坦、李遠哲，就是社會的福氣，因為他們不會只顧自己。

2. As deaf education moved closer to mainstream education in the 1980s, education departments developed "character signs" that made teaching easier. Formed by the hands in the shape of individual Chinese characters and used sequentially in sentences that adhere to Chinese syntax, character signs differ significantly from natural signs. 民國70年代，隨著聽覺教育向主流教育靠攏，教育單位研發出方便教學的「文字手語」，依著一般中文語句的結構順序，以手勢與字跡打出來，兩者並益甚大。

3. The ROC Constitution stipulates that each Chinese citizen has the right to receive an education. Principal Chou believes that the government is pushing a "look south" policy of improving relations with the nations of Southeast Asia, yet it has not solved the problem of educating the children of Taiwanese businessmen

Translation Panel

education (2403)
educational (540)
educated (107)
school (85)
teaching (55)
training (91)

Collocation Panel

教育 [N]
[ADJ]
[V]
[P]
[C]
[S]
[A]
[P]
[S]
[A]
[P]
大學、學校、子女...

教材編輯輔助系統

教材編輯工具備註區

自動注音 * 自動詞類分析功能
不注音 * 標註聲調 * 標註詞類 * 標註詞性 * 標註詞類 * 標註詞性 * 標註詞類 * 標註詞性

好朋友: 買了1好: 幾: 件: 衣: 褲: 顏色: 的: 鞋子: 帶: 顏色: 的: 鞋子: 帶: 顏色: 的: 鞋子: 帶: ...

標註: 詞類 聲調 詞性 詞類

標註選到注音字怎麼辦?
可將文本複製到下列ToneOZ網站，會自動注音以及提供聲調，來源: ToneOZ
標註詞類詞性的外網網站: ToneOZ 查詢功能

想使用標註字型請點以下連結下載
注意: 標註注音字型: 查詢功能

字彙標註 詞類標註 詞性標註 詞類標註 詞性標註

X	1	X	X
X	1	X	X
1	X	X	X
1	7	1	X
1	7	1	X
5	X	25	1
5	X	25	1
X	1	15	3
X	X	X	X
X	X	X	X

語義場關聯詞查詢系統

查詢功能: 查詢功能 | 查詢詞: [] | Search

查詢(詞)語義場關聯詞

輸入

正確關聯(詞):
感覺

負關聯(詞):
請輸入負關聯詞，例如: 買、的、斯

輸出

語義場關聯詞:

感受 0.7342	6.高階級
覺得 0.6458	1.學舊一級
感到 0.5878	6.高階級
情緒 0.5664	6.高階級
身體 0.5535	2.學舊二級
心緒 0.5337	4.基礎級
恐懼 0.5082	7.高階級
反應 0.5058	5.進階級 6.高階級
舒服 0.4925	3.入門級
渴望 0.4866	5.進階級

輸出詞數: 10
退出

教材編輯輔助系統

國教院教材編輯輔助系統

選擇關聯替換詞搜索
語料庫

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

關聯替換詞語等級限制

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

教材編輯工具編輯區

- 自動斷詞
- 國教院詞語分級表
- 不標音
- 標注音字型
- 標漢語拼音字型

[送出](#) [複製](#) [擷圖](#) [清空](#) [重置](#)

- 若分級標級的級別顯示「X」，則表示目前TBCL詞

[分級標記](#) [詞彙等級分布](#) [詞表](#)

[關聯替換詞語選擇器](#)

教材編輯輔助系統

教材編輯工具編輯區

- 自動斷詞 國教院詞語分級表
 不標音 標注音字型 標漢語拼音字型

今天我很忙，很累。早上七點半起床，洗了澡以後就吃早飯。我一邊吃飯，一邊聽錄音。九點到教室去上課。

第一節課是中文，老師教我們發音、生詞和語法，也教我們寫字，還給了我們一篇新課文，這篇課文很有意思。第二節課是電腦課，很難。

中午我和同學們一起到餐廳去吃午飯。我們一邊吃，一邊練習說中文。下午我到圖書館去上網。四點王朋來找我打球。五點三刻吃晚飯。七點半我去白英愛的宿舍跟她聊天（兒）。到那兒的時候，她正在做功課。我八點半回家。

教材編輯輔助系統

分級標記

詞彙等級分布

詞表

關聯替換詞語選擇器

分級詞表：國教院詞語分級表

今天 我 很 忙 ， 很 累 。 早上 七 點 半 起 床 ，
 1 1 1 1* 1 1*,7 1 X 1*

洗 了 澡 以 後 就 吃 早 飯 。 我 一 邊 吃 飯 ， 一 邊
 2 X X 1* 1 1 1* 1 X 1 X

聽 錄 音 。 九 點 到 教 室 去 上 課 。
 1 5 X 1 1* 1 1*

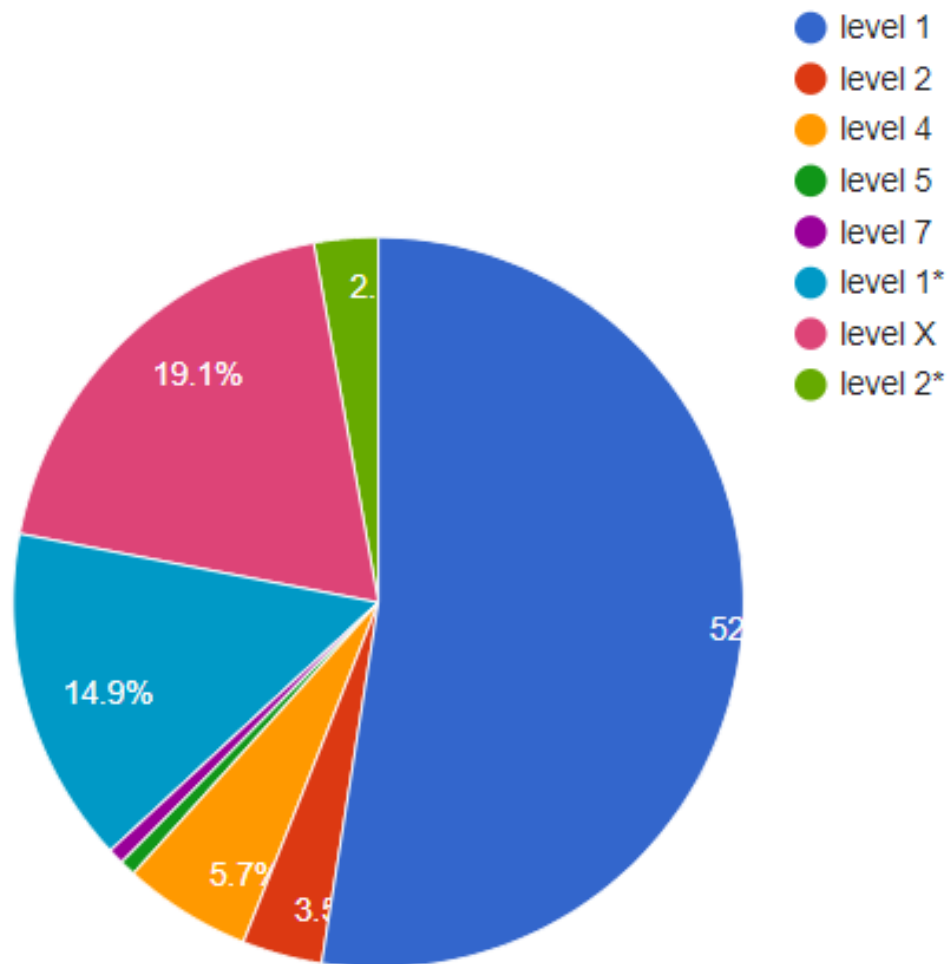
第 一 節 課 是 中 文 ， 老 師 教 我 們 發 音 、
 2 4 1* 1 1 1 1*,4 1 4

生 詞 和 語 法 ， 也 教 我 們 寫 字 ， 還 給 了
 4 1,6 7 1 1*,4 1 X X X

我 們 一 篇 新 課 文 ， 這 篇 課 文 很 有 意 思 。
 1 1 4 1 4 1 4 4 1 2*

第 二 節 課 是 電 腦 課 ， 很 難 。
 X 4 1* 1 X 1 1,6

詞彙等級分布



教材編輯輔助系統

分級標記 詞彙等級分布 詞表

關聯替換詞語選擇器

詞語	次數	等級
我	11	基礎
很	4	基礎
了	4	未收錄
吃	4	基礎
一邊	4	未收錄
到	4	基礎
去	4	基礎
我們	4	基礎
節	2	進階
課	2	基礎
是	2	基礎
中文	2	基礎
教	2	基礎
和	2	基礎
一	2	基礎
篇	2	進階
課文	2	進階

1 今天	1 我	1 很
1	1	1
1* 忙	X ,	1 很
1*	X	1
1*,7 累	X .	5 清晨
1*,7	X	5 清晨
X 六點半	1* 起床	4* 早晨 0.75 5 傍晚 0.73 6 凌晨 0.72 1 下午 0.7 1 早上 0.69 X 午後 0.69 5 黃昏 0.66
X	1*	
2 洗	X 了	
2	X	

肆

華英雙語索引典

雙語語料庫

取自台灣光華雜誌 2021/6月

海在那裡，人生就在那裡

蘇達貞的海洋夢想

四面環海的台灣，島國居民坐享近1,600公里的海岸線，與海的距離僅有一步之遙，但長久以來的恐海教育，卻讓人民懼海、退避三舍。拖鞋教授蘇達貞用「認識、親近、喜愛」的步驟循序打破民眾的恐海藩.....

Life Is Where the Sea Is

Su Tar-zen's Ocean Dreams

Surrounded on all sides by water, Taiwan's residents enjoy nearly 1600 kilometers of coastline, and few live more than a short distance from the sea. But Taiwanese people have long been taught about the dangers of the ocean, which makes them keep their distance from it. Professor Su Tar-zen br.....

統計式雙語句對應

海在那裡，人生就在那裡

Life Is Where the Sea Is

蘇達貞的海洋夢想

Su Tar-zen's Ocean Dreams

四面環海的台灣，島國居民坐享近1,600公里的海岸線，與海的距離僅有一步之遙，但長久以來的恐海教育，卻讓人民懼海、退避三舍。

Surrounded on all sides by water, Taiwan's residents enjoy nearly 1600 kilometers of coastline, and few live more than a short distance from the sea.

But Taiwanese people have long been taught about the dangers of the ocean, which makes them keep their distance from it.

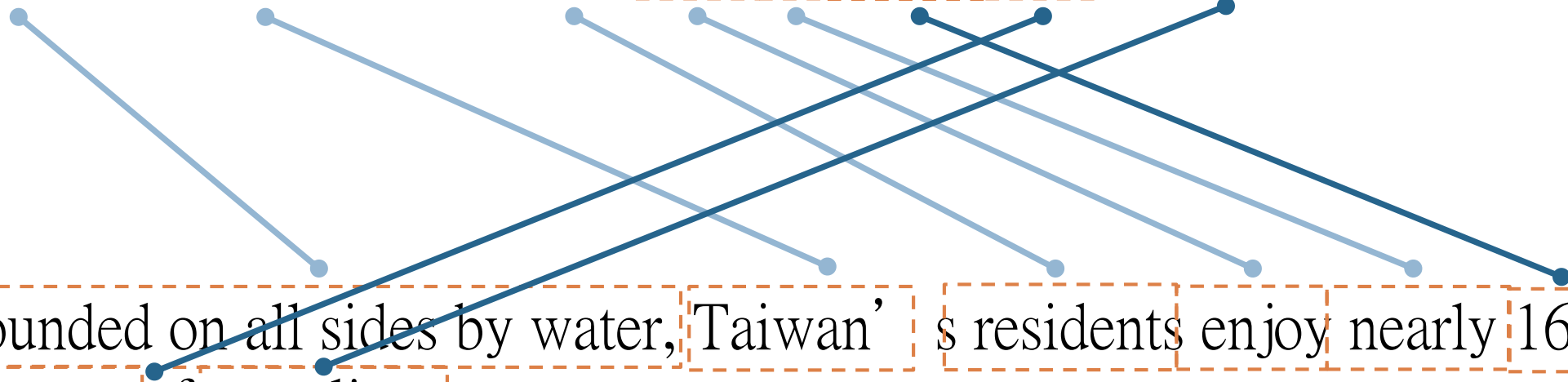
拖鞋教授蘇達貞用「認識、親近、喜愛」的步驟循序打破民眾的恐海藩.....

Professor Su Tar-zen br.....

統計式雙語詞對應

■ 四面環海的台灣，島國居民坐享近1,600公里的海岸線

■ Surrounded on all sides by water, Taiwan's residents enjoy nearly 1600 kilometers of coastline



國教院華英雙語索引|典系統

雙語例句

自動對應翻譯

語料庫 光華雜誌

關鍵詞

總共找到 1275 句

<< 上一頁 1 下一頁 >>

1. However, while **improvements** in the police personnel and promotion system [which provided much of the focus of the article] proved the important factor more important in relieving the problem, the improvement in society 's behavior. 但警察人事制度的**改善**與升遷管理減低我們壓力、**改善**我們工作環境與犯罪率的減低。
2. " The Mini-Europe Plan " is the result of research on how to **reduce** the trade deficit with Japan , commissioned by the Ministry of Economic Affairs at the end of 1989 and carried out by the American management consulting firm of Mckinsey & Co .. Its principal method is to struggle to get European brand names to manufacture in Taiwan or to produce and sell with the ultimate aim of opening up the Japanese market .
「小歐洲計畫」是民國七十八年底，經濟部委託美國麥肯錫企管公司就**改善**我國對日貿易逆差所提出的研究結果。它的作法主要是以爭取歐洲名牌商品來台灣生產，或是生產兼銷售為手段，而目的是在於拓展日本市場。
3. From Candles to Flashlights : The shortage of nurses is a serious problem with no signs of **improvement** .
現今護理人力不足的情況，一直沒有獲得**改善**。

自動抽取搭配詞

自動對應翻譯

improve (269)
improved (139)
improving (107)
improvement (93)
improvements (57)
better (8)

自動抽取搭配詞

改善 [Vt]
[ADV~]
大為~, 大幅~, 有~, 限期~,
[C~]
以~,
[N~]
情況~,
[Vt~]
致力~, 設法~,
[~N]
~生活, ~環境, ~體質,
~品質, ~交通, ~設施,

1. However , while **improvements** in the police personnel and promotion system [which provided much of the focus of the article] would be important to us , even more important in relieving the pressure on ordinary policemen would be an improvement in society ' s behavior and a lowering of the crime rate .
 但警察人事制度的**改善**與升遷管道的暢通，對基層員警雖然重要，真正能減低我們壓力、**改善**我們工作環境的，卻是我們社會更根本的人心良善的提昇與犯罪率的減低。
2. " The Mini-Europe Plan " is the result of research on how to **reduce** the trade deficit with Japan , commissioned by the Ministry of Economic Affairs at the end of 1989 and carried out by the American management consulting firm of Mckinsey & Co .. Its principal method is to struggle to get European brand names to manufacture in Taiwan or to produce and sell with the ultimate aim of opening up the Japanese market .
 「小歐洲計畫」是民國七十八年底，經濟部委託美國麥肯錫企管公司就**改善**我國對日貿易逆差所提出的研究結果。它的作法主要是以爭取歐洲名牌商品來台灣生產，或是生產兼銷售為手段，而目的是在於拓展日本市場。
3. From Candles to Flashlights : The shortage of nurses is a serious problem with no signs of **improvement** .
 現今護理人力不足的情況，一直沒有獲得**改善**。
4. National and city governments have been hardpressed to keep up with urban demand for public utilities and traffic control . High-rise buildings have proliferated . Beautification and **pollution** control are getting high priority from planners .
 政府為求都市有計畫的發展，早已製定都市發展計畫，改良公共給水，興建國民住宅，**改善**交通秩序，美化都市環境，管制空氣污染，以及合理

自動對應翻譯

自動對應翻譯

improve (269)
improved (139)
improving (107)
improvement (93)
improvements (57)
better (18)
improve their (17)

自動對應搭配

自動抽取搭配詞

改善 [Vt]

[ADV~]

大為~, 大幅~, 有~, 限期~,

[C~]

以~,

[N~]

情況~,

[Vt~]

致力~, 設法~,

[~N]

~生活, ~環境, ~體質,

~品質, ~交通, ~設施,

~逆差,

改善 [Vt]

雙語索引典：輔助寫作教學

「設法改善」

- We have to **come up with ways to improve** all these .
這些都亟需 **設法改善** 。
- We can 't avoid them for this reason ; we should **find a way to improve** election practices . "
顯然我們不能因這個理由而逃避，而是應該 **設法** 去 **改善** 選舉風氣。」
- he really wished he could **find a way to improve** the lives of those on the farms .
真希望將來能 **設法改善** 農民生活。



練習

- 請查詢「討論」
 - 觀察「討論」的翻譯
 - 觀察「討論」的搭配詞
 - 查詢「討論」與其搭配詞的英文用法

雙語索引典的限制

句對應、詞對應、自動對應翻譯、搭配

- 都是 NLP 技術自動產生
- 無法達到百分之百正確率
- 使用者必須自行判斷正確性



**THANK
YOU**

