

2019語料庫及標準體系應用工作坊

# 語料庫工具介紹

張俊盛、楊馨瑜

[jason@nlplab.cc](mailto:jason@nlplab.cc) , [chingyu@nlplab.cc](mailto:chingyu@nlplab.cc)

國立清華大學自然語言實驗室

2019-0711 16:40-17:30 正修科技大學圖書科技大樓6樓0603階梯教室

# 語言學習是重要問題

- 全球化：幾乎人人都要學習、使用通用語（英語、華語）
- 到了 2020 年，20 億人使用英語、學習英語
- 非母語者 (L2) 多於母語者 (L1)：3 比 1
- 寫作、口說比讀、聽困難，所以最好有 L1 老師、編輯
- L1 老師遠遠不足以應付 L2 學習者的需要、數量
- 什麼是大問題？
- L2 學習者寫說時，好像文法沒問題（Word 檢查沒錯）
- 但是溝通沒效率（不夠流暢、不夠正確）
- L1 老師、編輯當然可以幫上忙，但是供應不足、成本太高
- 大數據、人工智慧是「絕佳解決方案」



# 非母語常常有問題

but there are just not enough L1 teachers around to help out

Discuss 'about'  
the issue  
*Do I need 'about' here?*

'Accept'  
higher education  
*Am I using the right verb?*

Graduation 'threshold'  
*Does that sound right?*



# 目前的解決方案

- (Mostly non-L1) Language teachers
- Grammar books
- Print or online dictionaries
- Grammar checker
- *Google*<sup>3, 4</sup> or *Google Translate*

## Limitations of these traditional approaches

1. Low in availability
2. Low in speed
3. Unreliable at times
4. Difficult to find answers directly

# 數位詞彙學習

- 英語詞典、華語詞典

- Collins Dictionary, Oxford Dictionary, Cambridge Dictionary, Macmillan Dictionary, etc.

- 《國語小字典》 (4000字)、《重編國語辭典修訂本》 (16.7萬字)、《國語辭典簡編本》 (4.5萬字)、《成語典》 (5000條)、《中華語文大辭典》 (10.2萬字)

中華語文大辭典

音讀

ㄅㄧˋ ㄇㄧㄢˇ

bìmiǎn

臺灣標準字

避免

大陸規範字

避免

設法不讓某種情況（多指不好的情況）發生。㉔避免直接衝突 | 避免發生事故 | 不可避免地發生了這場爭執。

≡ 協助校正



# 數位詞彙學習

- 中文詞彙網路 <http://lope.linguistics.ntu.edu.tw/cwn/query/>

## 1. 不及物動詞 (VH) 展開 收起

形容人或物的外表好看使觀者感到愉悅而讚賞的。

同義詞：水<sup>2</sup> 美麗 美<sup>1</sup> 亮<sup>1</sup> 靚 漂<sup>3</sup> 秀<sup>1</sup> 麗<sup>1</sup>

反義詞：醜<sup>1</sup> 陋

下位詞：正<sup>2</sup>

附註：下位詞：「正2(2920)」。

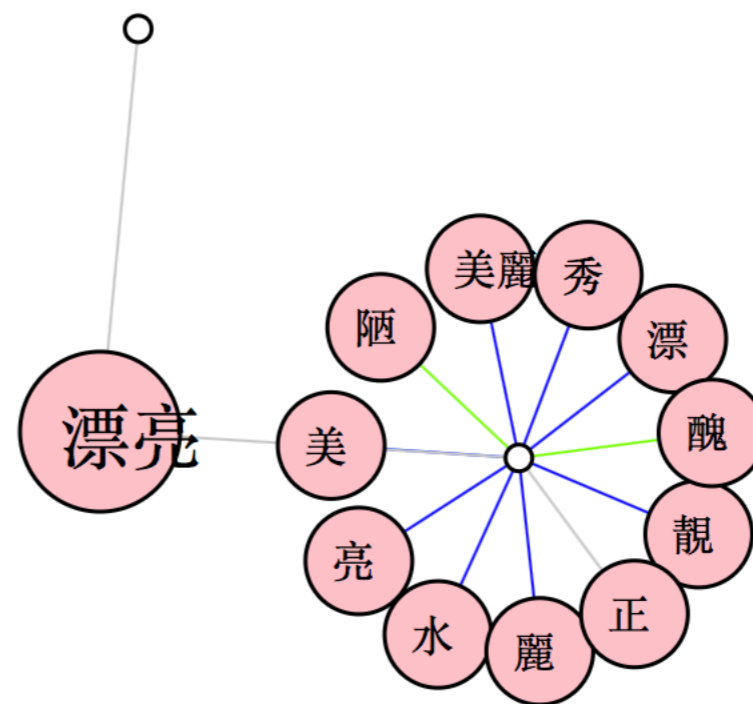
隱藏例句

例句

當年，你媽咪是一流演員，長得又甜蜜又漂亮。

立美也看到了一個漂亮的皮包，可是價錢貴了一點兒。

那時她二十多歲，漂亮而纖瘦，隨時帶著一種謙虛的笑容。



## 2. 不及物動詞 (VH) 展開 收起

形容事情做得相當出色或有技巧。

# 數位詞彙學習

- 語義場關聯詞查詢系統 <http://lingcorpus.iis.sinica.edu.tw/cgi-bin/kiwi/mkiwi/kiwi.sh>

## 國家教育研究院 - 語義場關聯詞查詢系統(雛型)

**選擇語料庫**

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

**比較詞(組)關聯度**    **查詢詞(組)語義場關聯詞**

**輸入**

**正關聯詞(組) :**

走私

**負關聯詞(組) :**

請輸入負關聯詞，例如：櫻桃

輸出詞數: 10

送出

**輸出**

**語義場關聯詞:**

私運 0.7931156158447266  
私梟 0.7482520341873169  
偷渡 0.7080168724060059  
夾藏 0.7033454179763794  
私貨 0.7008394002914429  
走私案 0.684169352054596  
毒梟 0.6788852214813232  
走私貨 0.677647590637207  
運毒 0.6690158843994141  
香菸案 0.661353349685669

# 詞彙句法數位學習

## • 中央研究院現代漢語標記語料庫 4.0 版

<http://lingcorpus.iis.sinica.edu.tw/cgi-bin/kiwi/mkiwi/kiwi.sh>

搜尋範圍-

設定行寬 (50 ~ 119) : 78

詞類候選單

特徵候選單

關鍵詞 或 重疊詞	詞類	特徵
<input checked="" type="radio"/> 避免  <input type="text"/> 或 <input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="radio"/> <input type="text"/> 或 <input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="radio"/> <input type="text"/> 或 <input type="text"/>	<input type="text"/>	<input type="text"/>

重疊詞示例

AAB：試試看、走走路  
 ABB：試看看、亮閃閃  
 AABB：高高興興、平平安安  
 ABAB：高興高興、研究研究

執行 清除

1. 作(VC)整體(Na)的(DE)規劃(Na)，不然(Cbb)很(Dfa)難(VH) 避免(VE) 那(Nep)「大街小巷(Nc)停滿(VC)車(Na)」，「狗年(Nd)

2. 了(Di)這些(Neqa)建議(Na)後(Ng)，使(VL)我們(Nh)能夠(D) 避免(VE) 浪費(VH)走入(VCL)歧途(Na)。那麼(Dk)他們(Nh)對(P)社會(Na)

3. ，中央(Nc)研究院(Nc)的(DE)公信力(Na)，無可(D) 避免(VE) 的(DE)將(D)會(D)受到(VJ)這些(Neqa)個人(Na)意見(Na)的(DE)

4. 的(DE)參考(Na)。但(Cbb)中央(Nc)研究院(Nc)絕對(D)要(D) 避免(VE) 捲入(VCL)政府(Na)的(DE)決策圈(Na)內(Ncd)。不然(Cbb)

5. 發展(VC)他們(Nh)人格(Na)的(DE)過程(Na)上(Ng)應該(D) 避免(VE) 的(DE)一些(Neqa)事情(Na)。第一(Neu)、要(D)溝通(VC)，

6. 發展(VC)他們(Nh)人格(Na)的(DE)過程(Na)上(Ng)應該(D) 避免(VE) 的(DE)一些(Neqa)事情(Na)。第一(Neu)、要(D)溝通(VC)，

7. 核心(Na)的(DE)訴求(Na)：(1) (Neu)他(Nh)刻意(D)在(D) 避免(VE) 西方(Ncd)傳統(VH)繪畫(Na)裡面(Ncd)，通過(VC)對於(P)

8. (如(P)好壞(Na)、美醜(Na))等(Cab)因素(Na)，甚至於(D) 避免(VE) 將(P)觀眾(Na)囿限(VJ)在(P)他(Nh)預設(VC)的(DE)「作品(Na)

9. 莖(Na)葉(Nb)或(Caa)澆灌(VC)土(Na)中(Ncd)，但(Cbb) 避免(VE) 與(P)根部(Nc)直接(VH)接觸(VC)。速效(Na)化學(Na)肥料(Na)

10. 速效化學肥料可用點狀，條狀或環狀施肥法，施用時應 避免 太接近根部或植株基部。可以使用噴洒葉面吸收的肥料



# 詞彙句法數位學習

- 國教院 : [https://coct.naer.edu.tw/cqpweb/yl\\_a/](https://coct.naer.edu.tw/cqpweb/yl_a/)
- - 前字後字查詢: 一心一? 、一心\* → 一心一意
- - 詞類查詢: \_V\* 到 → 增加到, 送到

使用說明

<http://coct.naer.edu.tw/cqpweb/doc/指令速查表.pdf>

Your query “一 (\_V\*|\_N\*) 就” returned 909 matches in 718 different texts (in 11,202,927 words [19,247 texts]; frequency: 81.14 instances per million words) [1.893 seconds]

|< << >> >| Show Page: 1 Line View Show in random order New query Go!

No	Filename	Solution 1 to 50	Page 1 / 19
1	100042	(地) 中產階級內部分殊化的探討。本計劃從	一開始就
2	100140	工地附近道路、不於工地燃燒廢棄物等等；最後	一招就
3	100794	」。忽然間，我似乎醒悟過來。我今天寫這	一封就
4	101177	李鎮對陳中申啦，一直想寫些感想，但是	一忙就
5	101177	在音色、技巧上都很有自己的特色，	一聽就
6	100486	佇立在雨中，時間彷彿凍結了，一分	一秒就
7	100510	若對方常要保護他自己時，開放自己的	一方就
8	100609	天可以得穀一合，一月三升，	一年就

# Chinese Linggle

# Chinese Linggle

- 自定義字串查詢 (5字內)
- 支援以下功能：
  - 前後搭配字查詢 ( \_ , \* )
  - 確認是否需要某個詞 ( ? )
  - 選擇較佳用字 ( / )
  - 詞彙內部查詢 ( \$ )
  - 近似詞查詢 ( ~ )
  - 詞性查詢
- 查詢結果以頻率排序，並提供例句。
- 符號(萬用字元)以**半形**輸入

# Chinese Linggle

- 中央通訊社
  - <https://cna.linggle.com>
  - 1991-2004
  - 句數：五千四百萬； 字數：4億4千萬
- 聯合報
  - <https://udn.linggle.com>
  - 2004-2017
  - 句數：六千六百萬； 字數：4億6千萬

# 查詢前字、後字「\_」

- 「\_」：查詢與底線數目相符的數個字詞 +

linggle<sup>10</sup>

逼\_\_



N-gram	Percent	Count	Example
逼討債務	19.7%	110	Show
逼得他	7.5%	43	Show
逼他下台	6.6%	38	Hide

- 但一直試圖逼他下台的政治對手則說，
- 以及對於試圖逼他下台行動的一次公民投票，
- 那些想逼他下台的人都是反對改革的。
- 但杜氏對這股逼他下台的顛覆壓力毫無所懼，
- 但另一面卻醞釀要逼他下台。
- 今天稍早他很僥倖地在逼他下台的過程中獲勝。
- 這是他的政敵為了逼他下台而捏造出來的罪狀。
- 說它的目的在於「彈劾總統並逼他下台」。
- 國會內部以及新聞界逼他下台的壓力更是有增無減。
- 但白宮官員有意逼他下台的說法並非事實。

# 查詢前字、後字「\_」

linggle<sup>10</sup>

\_ 避免



N-gram	Percent	Count	Example
<u>為 避免</u>	26.4%	14,000	<a href="#">Hide</a>
<ul style="list-style-type: none"><li>• 新光大樓管理單位 <b>為 避免</b> 類似情形再度發生，</li><li>• 但由於降溫的目的是 <b>為 避免</b> 經濟過熱導致經濟泡沫化，</li><li>• 但 <b>為 避免</b> 廠商對外投資腳步再度受到限制，</li><li>• 准考證規定掛在左胸前也是 <b>為 避免</b> 考生在准考證上寫小抄作弊，</li><li>• 但 <b>為 避免</b> 造成學校困擾及社會大眾誤解，</li><li>• 但 <b>為 避免</b> 因交通狀況而影響七月份陸續舉行的大學、高中及五專聯招，</li><li>• 北縣議員強調拆違建是 <b>為 避免</b> 更多人受害</li><li>• 但 <b>為 避免</b> 政府對外締結條約有損國人權益，</li><li>• 行政院農業委員會 <b>為 避免</b> 基層金融問題爆發影響社會安定，</li><li>• 是 <b>為 避免</b> 對市場有太大的衝擊。</li></ul>			
<u>以 避免</u>	15.4%	8,400	<a href="#">Show</a>
<u>為了 避免</u>	9.5%	5,200	<a href="#">Show</a>

# 查詢前字、後字 「\*」

- 「\*」：查詢0-多個字詞 \*

linggle<sup>10</sup><sup>12</sup>

培養 \* 習慣



N-gram	Percent	Count	Example
培養 閱讀 習慣	28.8%	173	Show
培養 運動 習慣	15.3%	92	Show
培養 深入 閱讀 的 習慣	5.0%	30	Show
培養 孩子 閱讀 習慣	5.0%	30	Show
培養 學生 閱讀 習慣	4.8%	29	Show
培養 孩子 的 閱讀 習慣	2.8%	17	Show



# 查詢是否需要某個字詞 「？」

(字)?

linggle<sup>10</sup><sup>12</sup>

喜歡 ? 做 運動



N-gram

Percent

Count

Example

喜歡 運動

96.1%

99

Show

喜歡 做 運動

3.9%

4

Show

linggle<sup>10</sup><sup>12</sup>

打 ? 了 ? 一個 噴嚏



N-gram

Percent

Count

Example

打 個 噴嚏

57.3 %

47

Show

打了 個 噴嚏

22 %

18

Show

打 了 一 個 噴嚏

14.6 %

12

Show

打 一 個 噴嚏

6.1 %

5

Show



# 選出最佳用字「/」

(字|字)

linggle<sup>10</sup>

不/沒 那麼 高



N-gram

Percent

Count

Example

沒 那麼 高

85.3 %

87

Show

不 那麼 高

14.7 %

15

Show

linggle<sup>10</sup>

寫/作/做 作業



N-gram

Percent

Count

Example

寫 作業

88.3%

1,020

Show

做 作業

11.3%

131

Show

寫 作業

0.3%

4

Show

# 詞彙內部查詢「\$」

字\*

linggle<sup>10</sup><sup>12</sup>

\$破



突破

31.2 % 37,580

Show

破

21.1 % 25,482

Show

打破

15.8 % 19,041

Show

跌破

5.9 % 7,152

Show

偵破

2.8 % 3,333

Show

衝破

2.1 % 2,557

Show

識破

1.7 % 2,017

Show

殘破

1.5 % 1,783

Show

爆破

1.2 % 1,452

Show

# 支援詞性查詢

- 簡化中研院詞彙標記集

Tag	Meaning	Examples
<b>adj.</b>	adjective	高 'tall', 紅 'red', 基本 'fundamental'
<b>adv.</b>	adverb	很 'very', 只 'only', 穩穩 'firmly', 常常 'usually'
<b>asp.</b>	aspect marker	了 perfective, 著, durative, 過 experiential
<b>cl.</b>	classifier, massifier	朵, 個, 束 'bundle', 年 'year'
<b>de.</b>	the marker DE 的	的 'DE'
<b>det.</b>	determiner	一 'one', 每 'every', 此 'this', 一些 'some'
<b>loc.</b>	localizer	上 'on', 裡 'inside', 外 'outside', 前 'in front of'
<b>n.</b>	noun	電腦 'computer', 報告 'report', 魯蛇 'loser', 康乃馨 'carnation'
<b>prep.</b>	preposition	在 preposition, 對 'to', 與 'with', 向 'to'
<b>post.</b>	postposition	時 'when', 等等 'etc.', 之外 'besides', 前後 'around'
<b>v.</b>	verb	避免 'avoid', 抵達 'arrive', 出生 'be born', 提供 'offer'

# 使用詞性查詢

linggle<sup>10</sup><sup>12</sup>

— v. 就



N-gram	Percent	Count	Example
一開始就	22.4%	3,403	Show
一看就	9.2%	1,392	Show
一出生就	3.9%	597	Show

linggle<sup>10</sup><sup>12</sup>

adv. 工作



N-gram	Percent	Count	Example
強制工作	9.1%	250	Show
一起工作	7.7%	210	Show
無法工作	7.5%	200	Show

# 查詢近似詞 「~」

linggle<sup>10</sup><sup>12</sup>

~網紅



N-gram	Percent	Count	Example
美女	18.8%	12,290	<a href="#">Show</a>
名人	17.8%	11,636	<a href="#">Show</a>
社群	14.0%	9,151	<a href="#">Show</a>
當紅	9.0%	5,891	<a href="#">Show</a>
名媛	3.5%	2,281	<a href="#">Show</a>
部落客	3.2%	2,117	<a href="#">Show</a>
宅男	3.0%	1,985	<a href="#">Show</a>
教主	2.5%	1,638	<a href="#">Show</a>

# 查詢同義詞 「~」

linggle<sup>10</sup><sup>12</sup>

~避免 疫情



N-gram

Percent

Count

Example

避免 疫情

30 %

261

Show

防止 疫情

17.8 %

155

Show

防堵 疫情

11.1 %

97

Show

防範 疫情

9.3 %

81

Show

以免 疫情

5.4 %

47

Show

遏止 疫情

4.1 %

36

Show

# 牛刀小試 (Linggle)

## • Linggle 功能：

- 「\_」：查詢前字、後字
- 「?」：檢查可否使用
- 「詞性.」：以詞性查詢
- 「\*」：查詢零到多個字
- 「/」：查詢最佳字
- 「~」：查詢同義字

### ① 查詢同義字：

魯蛇、走私、女神、臉書、LOL、老兵

### ② 查詢可否為離合詞：

睡覺、關心、幫忙、打掃

### ③ 查詢動詞後，最常接續的「得字補語」：

吃、買、看、氣

### ④ 查詢名詞的量詞：

西瓜、玫瑰花、兇殺案、豪宅

### ⑤ 查詢是否可加介系詞：

住 + 處所(台北、街頭、旅館)、跑 + 處所(廁所、台北)

### ⑥ 查詢字詞選擇：

打電動 vs. 玩電動 欣賞畫作 vs. 欣賞比賽

### ⑦ 查詢語序：

或許應該 vs. 應該或許。 是或許 [動詞] vs. 或許是 [動詞]

Tag	Meaning
adj.	adjective
adv.	adverb
asp.	aspect marker
cl.	classifier, massifier
de.	the marker DE 的
det.	determiner
loc.	localizer
n.	noun
prep.	preposition
post.	postposition
v.	verb

# 字詞變化



# 寶可夢

- 中央通訊社 (1991-2004)

linggle<sup>10</sup><sup>12</sup>

寶可夢



N-gram

Percent Count Example

No result

# 寶可夢

- 聯合報 (2004-2017)

linggle<sup>10</sup>

寶可夢



N-gram

Percent

Count

Example

寶可夢

99.2 %

1,274

Show

寶可夢

0.8 %

10

Show

# 寶可夢神奇寶貝

- 中央通訊社 (1991-2004)

linggle<sup>10</sup><sup>12</sup>

神奇寶貝|



N-gram

Percent Count Example

神奇 寶貝

100 %

166

Show

# 寶可夢神奇寶貝

- 聯合報 (2004-2017)

linggle<sup>10</sup><sup>12</sup>

神奇寶貝



N-gram

Percent

Count

Example

神奇 寶貝

100 %

575

Show

# 鐵粉

- 中央通訊社 (1991-2004)

linggle<sup>10</sup><sup>12</sup>

鐵粉



鐵粉

100 %

17

Hide

- 利用中鋼公司酸洗冷軋工場副產品 – 高純度氧化鐵粉為原料，
- 就是將純淨的鐵粉封在可以透氣的韌性薄膜裡。
- 利用中鋼鋼品生鏽後的廢棄物高純度氧化鐵粉，
- 這批廢料在提單上寫著是塑膠、油料和鐵粉廢料。
- 回收的廢酸內含副產品氧化鐵粉，
- 五年前又設廠自瑞典進口銅粉、鐵粉和不鏽鋼粉等，
- 回收的廢酸液內會含有副產品氧化鐵粉，
- 也可製成較高品級的氧化鐵粉，
- 更可進一步添加其他成分製成較高品級的氧化鐵粉，
- 將當作觸媒用的鐵粉鑲在矽基板上，

# 鐵粉

- 聯合報 (2004-2017)

linggle<sup>10</sup><sup>12</sup>

鐵粉



鐵粉

100 %

239

Hide

- 鐵粉 顆粒在磁場分布中的受力行為」、「吸金耶
- 主要成分為 鐵粉、不明鹽類、活性碳和蛭石等
- 撕開後把暖暖包裡的 鐵粉、活性碳全倒入水中喝下
- 若銀彈充足、又有體力站全場的年輕 鐵粉
- 這兩天不少藍軍「鐵粉」因焦慮、緊張、失眠
- 後才知道這些全是「鐵粉」楊先生的珍藏
- 霹靂 鐵粉 不可錯過台中中港店的「霹靂英雄武動江湖特展」
- 低薪失業的年輕人更意外成了這位七十多歲老人的 鐵粉
- 首度入圍金鐘戲劇節目女配角的曾之喬化身超級「丸迷」出席小丸子 鐵粉 點召令活動
- 甚至有 鐵粉 要「全勤」看完台北、台中、高雄三站展覽

# 鐵粉

- 中央通訊社 (1991-2004)

linggle<sup>10</sup>

N-gram	Percent	Count	Example
液態	8.2 %	949	<a href="#">Show</a>
汞	7.7 %	892	<a href="#">Show</a>
水銀	5.1 %	591	<a href="#">Show</a>
氮	4.4 %	515	<a href="#">Show</a>
顆粒	4.3 %	495	<a href="#">Show</a>
硫	3.9 %	454	<a href="#">Show</a>

# 鐵粉

- 聯合報 (2004-2017)

linggle<sup>10</sup><sup>12</sup>

~鐵粉



N-gram	Percent	Count	Example
觀眾	27.5 %	56,254	<a href="#">Show</a>
粉絲	17 %	34,704	<a href="#">Show</a>
明星	16.6 %	33,938	<a href="#">Show</a>
歌迷	11.4 %	23,218	<a href="#">Show</a>
偶像	8.3 %	16,929	<a href="#">Show</a>
巨星	3.4 %	6,895	<a href="#">Show</a>



# 住

- 中央通訊社 (1991-2004)

linggle<sup>10</sup><sup>12</sup>

住 ?prep. 附近/台北



N-gram

Percent

Count

Example

住 台北

81.7 %

58

Show

住 附近

18.3 %

13

Show

# 住

- 聯合報 (2004-2017)

linggle<sup>10</sup><sup>12</sup>

住 ?prep. 附近/台北



N-gram	Percent	Count	Example
住在附近 	37.1 %	1,146	<a href="#">Show</a>
住台北 	27 %	835	<a href="#">Show</a>
住附近 	19.3 %	595	<a href="#">Show</a>
住在台北 	16.5 %	510	<a href="#">Show</a>
住台北 	0.1 %	3	<a href="#">Show</a>

# 跑

- 中央通訊社 (1991-2004)

linggle<sup>10</sup><sup>12</sup>

跑 n.	Q	?	
跑 票	16 %	141	Show
跑 馬燈	12.5 %	110	Show
跑 活動	10.8 %	95	Show
跑 路費	9.9 %	87	Show
跑 馬拉松	4.8 %	42	Show
跑 錯	3.3 %	29	Show
跑 公文	2.3 %	20	Show

# 跑

- 聯合報 (2004-2017)

linggle<sup>10</sup><sup>12</sup>

跑 n.	Q	?	
跑 行程	10.9 %	554	Show
跑 馬拉松	9.8 %	499	Show
跑 操場	7.9 %	400	Show
跑 馬燈	7 %	356	Show
跑 廁所	5.5 %	282	Show
跑 醫院	5.3 %	268	Show
跑 票	4.8 %	245	Show
跑 業務	4.7 %	237	Show

# 字詞差異

# 牛刀小試(字詞差異)

- 以前 vs. 從前 ‘before’ :

	以前	從前
詞頻		
前方名詞類別		

- 搖動 vs. 動搖 ‘sway’ :

	搖動	動搖
受詞種類		
修飾語種類		

- 又 vs. 再 ‘again’ :

	又	再
與模態詞的語序		
「又」和「再」的語序		
與時間詞的語序		

# 以前 vs. 從前 'before'

	以前	從前
詞頻	64000	5600
Linggle 輸入	以前/從前	

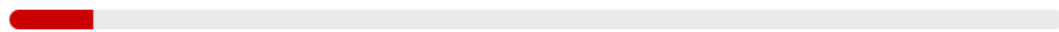
以前



92 % 64,020

Show

從前



8 % 5,587

Show

# 以前 vs. 從前 'before'

	以前	從前
前方名詞類別	人：我、他 法人：公司、台灣 時間點：月底、高中、明天中午 一段時間：三個月、半年	人：我、他 法人：公司、國家
Linggle 輸入	n. 以前	n. 從前

我 以前	22.1 %	1,578
月底 以前	4 %	284
我們 以前	3.5 %	251
爸爸 以前	1.2 %	84
台灣 以前	1.2 %	83
15日 以前	1.1 %	76

我 從前	17.4 %	69
他 從前	16.2 %	64
從前 從前	11.4 %	45
她 從前	8.8 %	35
台灣 從前	2.3 %	9



# 以前 vs. 從前 'before'

	以前	從前
詞頻 Linggle 輸入	64000	5600
前方名詞類別 Linggle 輸入	人：我、他 法人：公司、台灣 時間點：月底、高中、明天中午 一段時間：三個月、半年 n. 以前	人：我、他 法人：公司、國家 n. 從前
分析	1. 在 <b>現在</b> 之前（過去） 2. 在 <b>過去</b> 、 <b>未來時間點</b> 之前 3. 在 <b>一段時間</b> 之前	1. 在 <b>現在</b> 之前（過去）
<p>「以前」、「從前」皆表「在...之前」的意思，但「從前」的參考時間點為「現在」，而「以前」的參考時間點的種類較為彈性。</p>		

# 牛刀小試(字詞差異)

- 以前 vs. 從前 ‘before’ :

	以前	從前
詞頻		
前方名詞類別		

- 搖動 vs. 動搖 ‘sway’ :

	搖動	動搖
受詞種類		
修飾語種類		

- 又 vs. 再 ‘again’ :

	又	再
與模態詞的語序		
「又」和「再」的語序		
與時間詞的語序		

# 又 vs. 再 'again'

	又	再
與模態詞的語序	又 > 模態詞; 模態詞 > 又	模態詞 > 再
Linggle 輸入	又可能; 可能又 又會; 會又 又得; 得又 又/可能 又/可能	可能再 會再 得再 再/可能 再/可能

可能 又      67.8 %      883



又 可能      29.2 %      380



又 又      2.1 %      27



可能 再      95.3 %      2,171



再 再      4.6 %      104

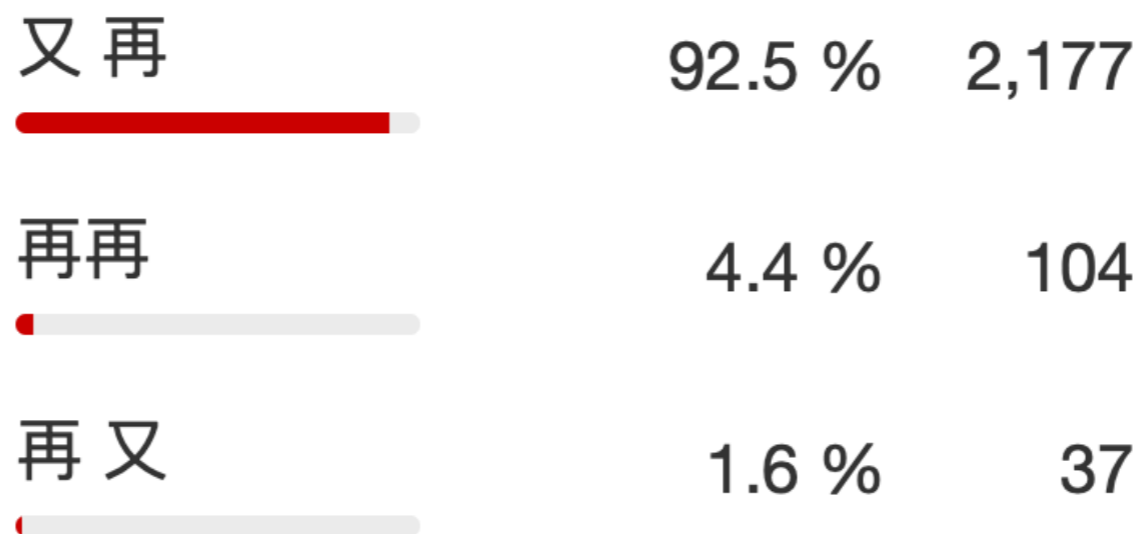


可能 可能      0.1 %      3



# 又 vs. 再 'again'

	又	再
「又」和「再」的語序 Linggle 輸入		又 > 再 又/再 又/再



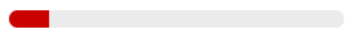
# 又 vs. 再 'again'

	又	再
與時間詞的語序	時間詞 > 又 明天又 昨天又	時間詞 > 再 明天再 昨天再
Linggle 輸入	又/明天 又/明天	再/明天 再/明天

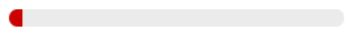
明天 又      82.2 %      185



又 又      12 %      27



又 又      4 %      9



明天 再      83 %      526



再 再      16.4 %      104



# 又 vs. 再 'again'

	又	再
與模態詞的語序 Linggle 輸入	又>模態詞;模態詞 > 又 又可能;可能又 又會;會又 又得;得又 又/可能 又/可能	模態詞 > 再 可能再 會再 得再 再/可能 再/可能
「又」和「再」的語序 Linggle 輸入		又> 再 又/再 又/再
與時間詞的語序 Linggle 輸入	時間詞 > 又 明天又 昨天又 又/明天 又/明天	時間詞 > 再 明天再 昨天再 再/明天 再/明天

分析 「又」和「再」都表again的意思，兩者修飾動詞組，皆時間副詞之下，但「又」的位置較「再」的位置高。

# 搖動 vs. 動搖 ‘sway’

	搖動	動搖
詞頻	760	2800
Linggle 輸入	搖動/ 動搖	
受詞種類	實體名詞 搖動身體、搖動車身、 搖動手機、搖動鍋子、 搖動樹枝、搖動瓶子 搖動n.	抽象名詞 動搖國本、動搖軍心、 動搖信心、動搖根本、 動搖信搖念、動搖地位 動搖 n.
Linggle 輸入		
修飾語種類	動作副詞 上下搖動、輕輕搖動、 左右搖動、用力搖動 迅速搖動、緩緩搖動 _ 搖動	否定副詞、模態詞、動量副詞 絕不動搖、無法動搖、 從未動搖、勢必動搖、 有點動搖、有些動搖 _ 動搖
Linggle 輸入		
分析	「搖動」、「動搖」都有搖擺不定的意思，「搖動」的對象是具體名詞，表物體的擺動。「動搖」的對象則是抽象名詞，表抽象事物的不穩固、不堅定。	

**SmartWrite**



# SmartWrite

- <https://whisky.nlplab.cc/smartwrite/>
- 聯合報資料
- 自動產生文法
- universal dependency parser + 人工抽取規則
- <http://t.linggle.com> (中文 SmartWrite+Linggle)

# SmartWrite

住

## 文法規則

[V] 住在 n 24574

住在 [附近, 宿舍, 這裡, 地區, 臺北]

搭配字

## 例句

發現歹徒以徒步方式偷車，證實他也住在附近。  
後來只剩4、5戶住在這裡，宿舍老舊、  
由公司安排在臺北租屋，一年中有大半年住在臺北，徹底融入臺灣食衣住行，  
明明我生在臺灣、住在臺灣，只因為生父是日本人，  
但願我住在學校附近。

[V] subj 住 13205

[家, 人, 我, 他, 她] 住

現年十九歲的周海洋，家住武漢市偏遠的漢南區，父親靠加工棉被謀生，  
但一人住大屋很孤獨，她養了一只黑色貴賓犬取名「  
在我住的家裡，有很多的不可以，  
陳姓男子告訴指出，他住的這棟大樓，糞水管線老舊失修，  
下次要拒絕就很難，何況她住朴子，開車回家要1小時，

[V] 住 obj 7944

住 [附近, 校, 人, 飯店, 晚]

但為能專心唸書，她請求住校。  
勸說不要因愛犬咬傷而對簿公堂，大家都是住附近的鄰居，見面也難免尷尬，  
而家住臺北、搭火車回蘇澳過端午節死亡的李姓婦人，  
等於現賺2萬3100元，若無法續住一晚，飯店特別提供住宿券，  
歐美人士消費力強，住飯店常用迷你吧的紅酒、飲料，

# 對比近義詞

## ● 燒 vs. 烤

[V] 燒 obj 4219

燒 [錢, 紙錢, 金紙, 燙傷, 王船]

[V] subj 燒 1962

[蠟燭, 柴, 火, 窯, 頭] 燒

[V] 燒 VP 357

燒 [起來, 出來, 起, 去, 好]

[V] 用 n 燒 139

用 [火, 打火機, 柴, 瓦斯, 柴火] 燒

[V] 燒 num m 94

燒 [一, 兩, 多, 幾, 3] [元, 天, 次, 小時, 年]

[V] 烤 obj 2752

烤 [肉, 地瓜, 麵包, 香腸, 番薯]

[V] subj 烤 415

[火, 鹽, 窯, 自己, 他] 烤

[V] 烤 VP 376

烤 [出, 出來, 到, 上, 起來]

[V] 用 n 烤 161

用 [火, 烤箱, 炭火, 木炭, 鹽] 烤

[V] 烤 num m 276

烤 [20, 10, 15, 一, 40] [分鐘, 元, 次, 串, 只]

[V] 以 n 烤 72

以 [火, 鐵板, 炭火, 焗, 烤箱] 烤

# 對比近義詞

## ● 幫 vs. 幫助 vs. 幫忙

[V] 幫 obj 31113

幫 [他, 我, 她, 你, 忙]

[V] subj 幫 4267

[他, 我, 她, 星光, 人] 幫

[V] 幫 clause 3198

幫 [解答, 做, 買, 找, 到]

[V] 在 n 幫 obj 390

在 [家, 時候, 旁邊, 期間, 現場]

幫 [他, 忙, 人, 我, 她]

[V] 用 n 幫 172

用 [方式, 手, 科技, 手機, 錢] 幫

[V] subj 幫忙 5610

[他, 人, 我, 她, 大家] 幫忙

[V] 幫忙 obj 3451

幫忙 [家計, 處理, 工作, 拉票, 家務]

[V] 幫忙 clause 2549

幫忙 [照顧, 做, 帶, 推, 找]

[V] 在 n 幫忙 938

在 [旁, 家, 店裡, 廚房, 工廠] 幫忙

[V] 由 n 幫忙 280

由 [老師, 人員, 志工, 她, 他] 幫忙

[V] 靠 n 幫忙 235

靠 [守備, 隊友, 大家, 人, 打擊] 幫忙

[V] 透過 n 幫忙 232

透過 [朋友, 立委, 友人, 協會, 基金會] 幫忙

[V] 幫助 obj 26686

幫助 [人, 別人, 孩子, 學生, 他們]

[V] 幫助 clause 2861

幫助 [有, 需要, 吸收, 瞭解, 度過]

[V] subj 幫助 1218

[他, 她, 才能, 人, 我們] 幫助

[V] 以 n 幫助 obj 324

以 [行動, 方式, 經驗, 力量, 專業]

幫助 [人, 學生, 球隊, 家庭, 孩子]

[V] 用 n 幫助 obj 305

用 [方式, 力量, 行動, 方法, 能力]

幫助 [人, 別人, 球隊, 他們, 他人]

[V] 透過 n 幫助 obj 302

透過 [方式, 力量, 活動, 教育, 民]

幫助 [孩子, 人, 學生, 他們, 民]

# 跨語言查詢

<https://x.linggle.com>

# x.Linggle

- 跨語言查詢: 打中查英、打英查中
- 支援所有Linggle的功能

接受 education



EN

receive education

40.9 % 26,013

Show

受教育

12 %

Show

接受教育的

10 %

Show

offer education

31.9 % 20,303

Show

receiving education

13.2 % 8,391

Show

obtain education

7.3 % 4,663

Show

## 受 教育

12 %

Hid

- 「人是什麼？」，「我從何而來？」這些都是很多小孩子常會提出的問題，因此我們如認為兒童有**受教育**的權利，則這權利亦應包括性教育。

questions like " what are mankind ? " , " where did i come from ? " have always been raised by kids . hence , if it is considered that children have the right to **receive education** , then sex education should also be included as part of the curriculum .

- 由於港府嚴格限制外來移民政策、節制生育及由於經濟好轉、使青少年**受教育**機會提高等因素，勞工市場來源減弱。

the fact that the government of hong kong imposes strict limitation on the number of immigrants into hong kong and factors like birth control , economic upturn and greater opportunity for youngsters to **receive education** jointly reduced our labour supply .



cultivate the friendship

23.3 %

1,133

Show

strengthen the friendship

20.4 %

991

Show

promote the friendship

12.3 %

596

Show

enhance the friendship

9.8 %

478

Show

deepen the friendship

9.1 %

443

Show

change jobs



51.9 % 134,772

Hide

轉業

50 %

Show

轉職

30.8 %

Show

轉工

7.7 %

Show

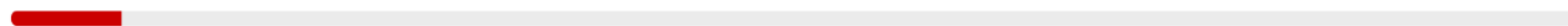
change efforts



10.4 % 26,930

Show

change work



7.1 % 18,463

Show

# Supporting all Linggle Functions



接受 \* education prep. \_



N-gram	Percent	Count	Translation
obtain an education to pursue	14 %	2,198	<a href="#">Show</a>
offer continuing education for psychologists	13.2 %	2,068	<a href="#">Show</a>
receive education in the	5.1 %	796	<a href="#">Show</a>
receive the education to help	4.7 %	730	<a href="#">Show</a>
offer continuing education for National	3.9 %	612	<a href="#">Show</a>

# 切換語言

xlinggle

solve 問題



N-gram	Percent	Count	Translation
解決 問題	80.2 %	13,484	Show
處理 問題	16.1 %	2,703	Show
根治 問題	0 %	6	Show
化解 問題	0.5 %	89	Show
克服 問題	0.6 %	93	Show

# 切換語言

xlinggle

solve problem|



CH

N-gram

Percent

Count

Translation

解決 問題

60.6 %

13,484

Show

處理 問題

12.1 %

2,703

Show

解決 難題

0.6 %

144

Show

解決 危機

1.7 %

375

Show

根治 問題

0 %

6

Show

解決 困難

2.6 %

584

Show

# 其他系統介紹

# NLPLab

- **Linggle 系列**

- <https://linggle.com>
- <https://coca.linggle.com> (COCA)
- <https://zh.linggle.com> (中文 Linggle 中央通訊社)
- <https://udn.linggle.com> (中文 Linggle 新聞)
- <https://bi.linggle.com> (TotalPhraseBook)
- <https://x.linggle.com> (雙語Linggle)

- **自動產生文法提示系列**

- <http://writeahead.nlpweb.org> (英文 WriteAhead)
- <https://whisky.nlplab.cc/smartwrite/> (中文 SmartWrite)
- <http://whisky.nlplab.cc:2345> (中文 SmartWrite+Linggle)

# NLPLab

- 文法改錯與升級系列
  - <https://whisky.nlplab.cc/> (英文改錯 Cool English)
  - <http://feedback.linggle.com> (Linggle Write)
  - <http://up.linggle.com> (Level Up)
- 閱讀輔助系列：
  - <https://b.linggle.com> (Linggle Booster)
  - <https://read.linggle.com> (Linggle Booster + wikification)
- 中英 **WordNet** 詞義、例句查詢 LanguageNet
  - <https://whisky.nlplab.cc/langnet/>