

Learning & Teaching Expo 2022

# 數碼科技助你追上最新英語發展

---

8 Dec 2022 (Thu)

**Dr. LAU, Chaak Ming**  
**Department of Linguistics and Modern Language Studies**  
**The Education University of Hong Kong**  
[lchaakming@eduhk.hk](mailto:lchaakming@eduhk.hk)

香港教育大學  
語言學及現代語言系  
劉擇明博士

# 數碼科技與英語教學

---

## 直接

- Computer-Assisted Language Learning (CALL) applications
  - 自學、閱讀、遊戲、測試平台
  - 課堂管理、活動工具
  - 輔助學習的工具  
(詞典、文字轉語音等)
  - 提供真實材料及輔助的學習平台

## 間接

- 數據
  - 語言分析 (linguistic analysis)
  - 改良課程
  - 提供個人化學習材料

# 語言數據的兩種用法

---

- 了解最新發展 (What)：用語料庫觀察變化
  - **工具：語料庫** (Corpus, pl. corpora)  
Corpus of Contemporary American English (COCA)  
<https://www.english-corpora.org/coca/>
- 活用最新材料 (How)：分析文章難易度
  - **工具：詞彙分析器** (Vocabulary Profiler)

# 觀察

---

為何需要觀察變化？

- I. 課堂時間寶貴
- II. 重要用法要及早教
- III. 傳統分析可能忽略重要用法
- IV. 過時用法影響溝通

# 語言數據的兩種用法

---

- 了解最新發展 (What)：用語料庫觀察變化
  - **工具：語料庫 (Corpus, pl. corpora)**  
Corpus of Contemporary American English (COCA)  
<https://www.english-corpora.org/coca/>
- 活用最新材料 (How)：分析文章難易度
  - **工具：詞彙分析器 (Vocabulary Profiler)**

# 用語料庫觀察變化

---

詞彙、語音、形態、句法、語義、語用

# Let the data speak!

---

- 要尋找規律模式和典型特徵，就不能單單依靠直覺或軼事證據 (anecdotal evidence)

*Armchair Linguist*

- 使用語料庫可以預防偏頗的結論
  - 讓數據說話

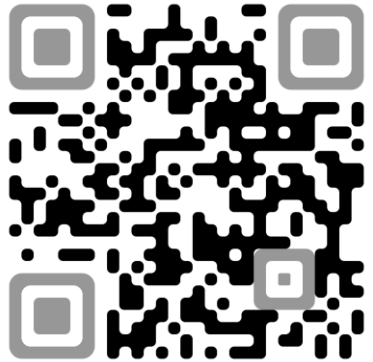


# 甚麼是語料庫？

---

- **Corpus (複數 Corpora)**
  - **結構化**的文本匯編
  - 一般可以全文搜尋
- 收集大量文字後可以做到甚麼？
  - **數據！**
    - 使用分佈
    - 比較年代差異
    - 找出文體之間的異同





# COCA語料庫

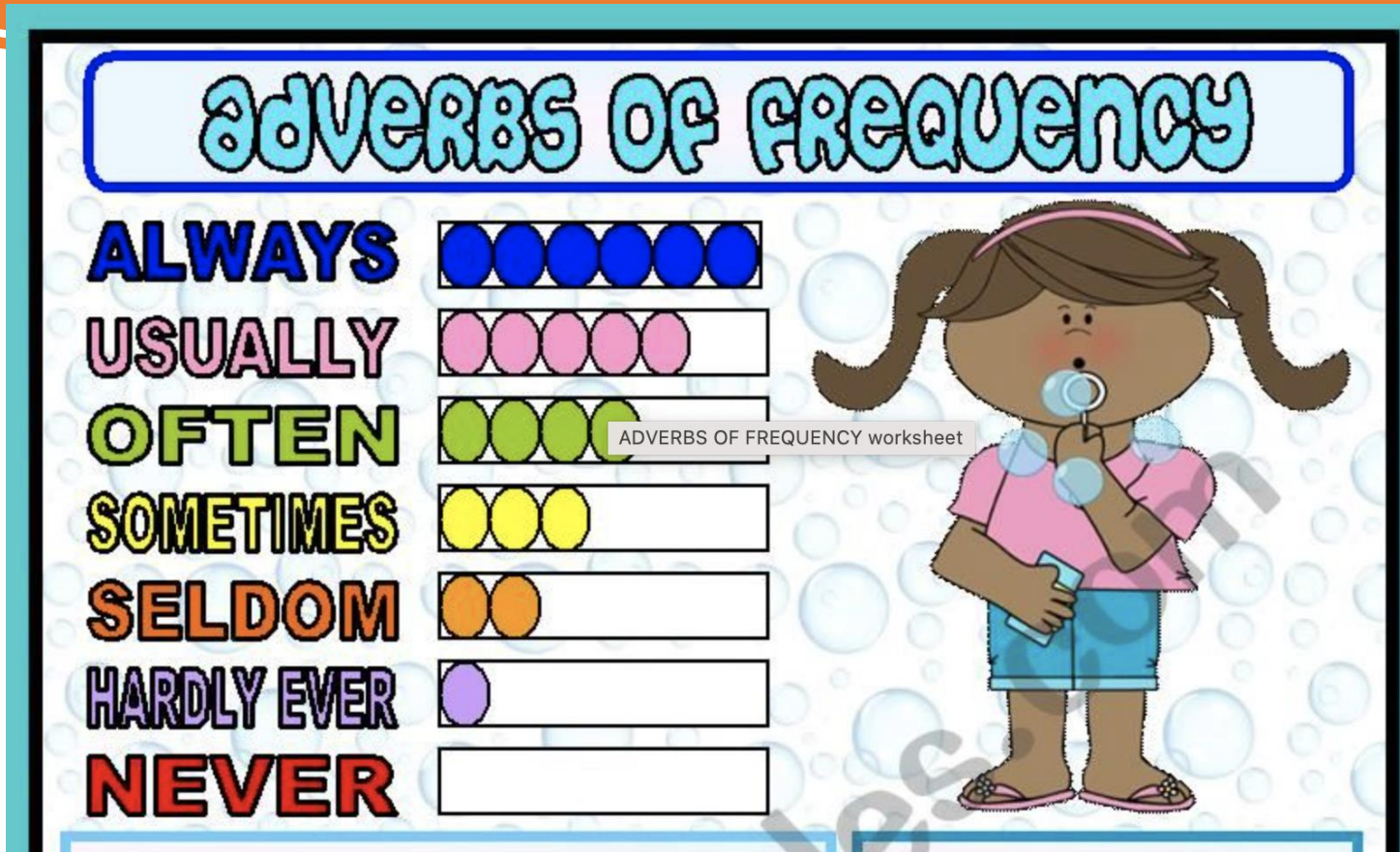
Corpus of Contemporary American English (COCA) (Davis, 2009)

<https://www.english-corpora.org/coca/>

- 平衡語料庫 (balanced corpus)
- 涵蓋各種文體：  
spoken, fiction, popular magazines, newspapers and academic texts
- 詞量超過 **3.85 億**

Davies, M. (2009) "The 385+ Million Word Corpus of Contemporary American English (1990– 2008+): Design, Architecture, and Linguistic Insights", *IJCL* 14(2): 159–190.

# 英語的變化：詞彙



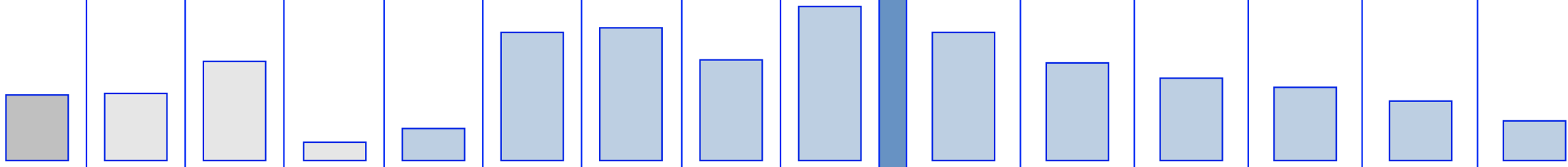
是否都是常用字詞？

[https://www.eslprintables.com/grammar\\_worksheets/adverbs/adverbs\\_of\\_frequency/ADVERBS\\_OF\\_FREQUENCY\\_777260/](https://www.eslprintables.com/grammar_worksheets/adverbs/adverbs_of_frequency/ADVERBS_OF_FREQUENCY_777260/)

# *seldom*

從體裁 (genre) 角度觀察

從時間 (time) 角度觀察

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD		1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	8562	809	1168	215	378	1449	1605	1173	1765		1703	1372	1146	1013	809	542
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8		139.1	147.8	146.6	144.9	145.3	144.7
PER MIL	8.62	6.29	9.40	1.68	3.00	12.25	12.73	9.64	14.73		12.25	9.28	7.82	6.99	5.57	3.74
SEE ALL SUB-SECTIONS AT ONCE																

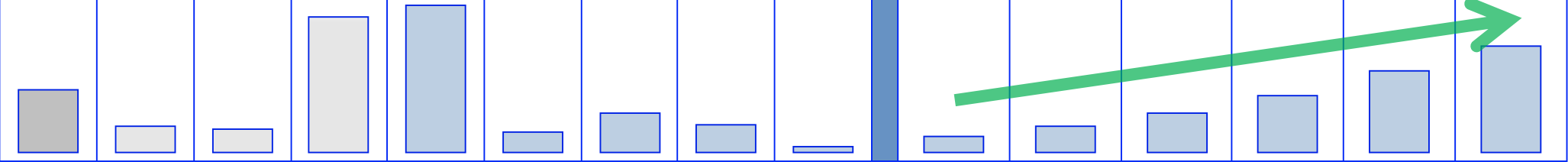
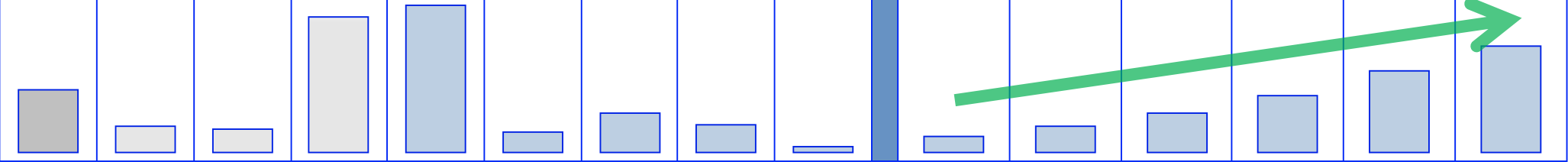
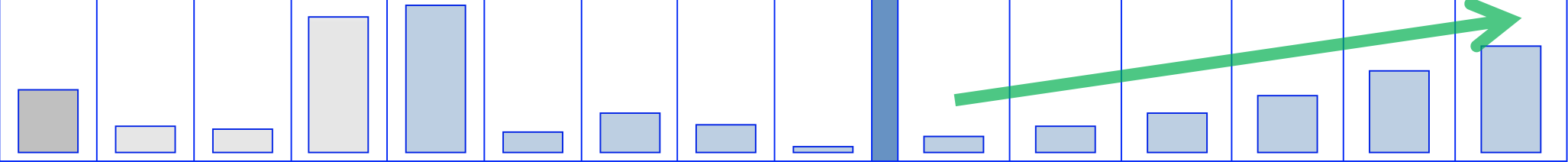
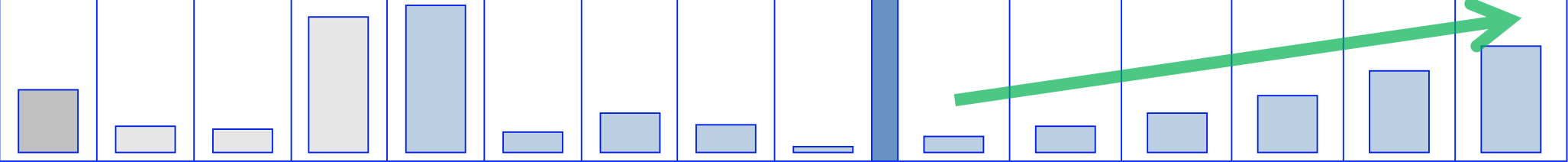
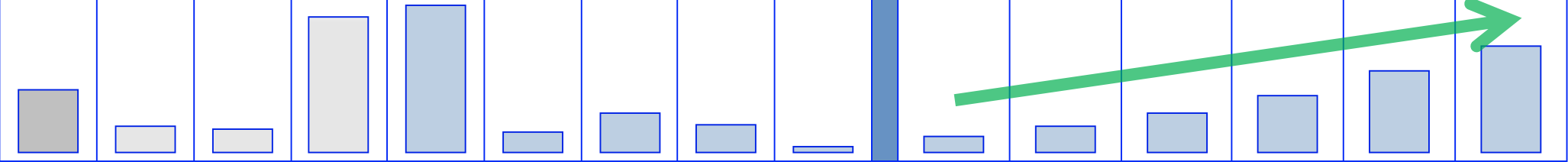
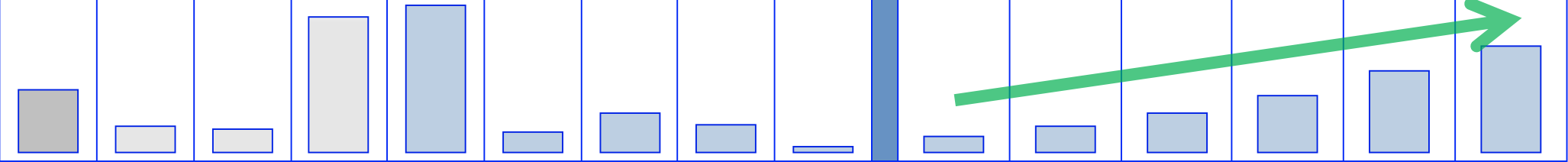
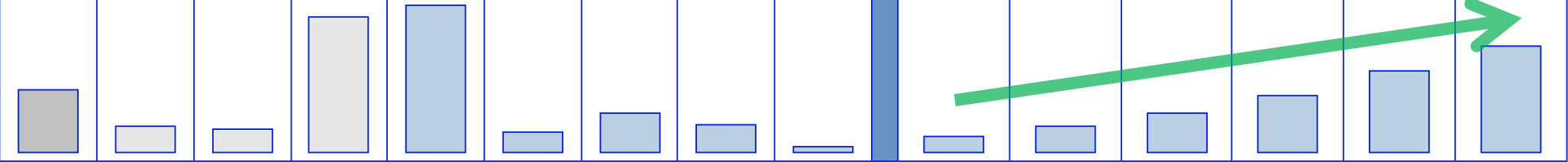
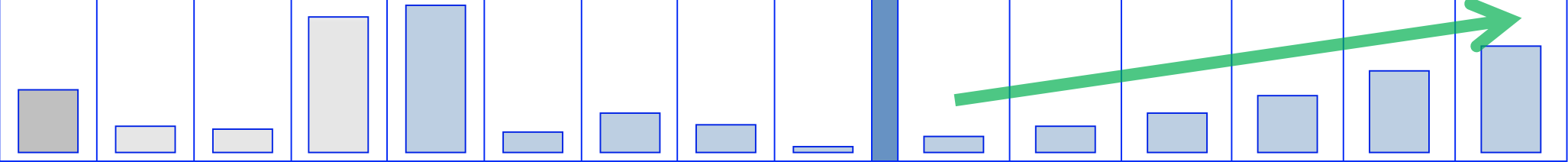
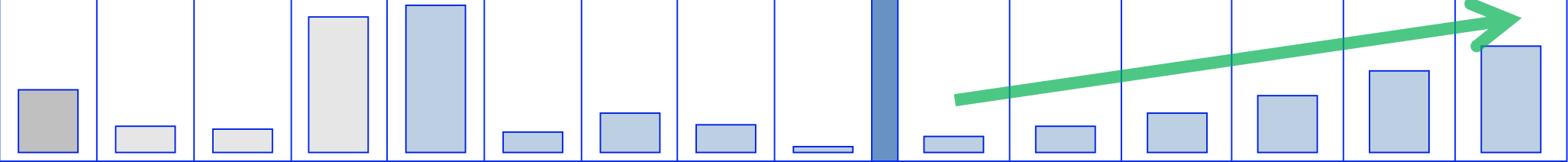
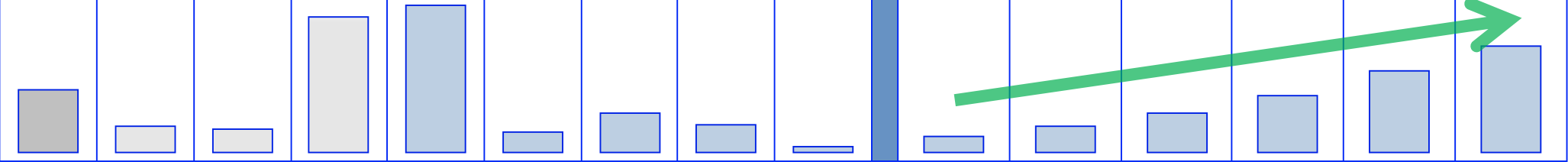
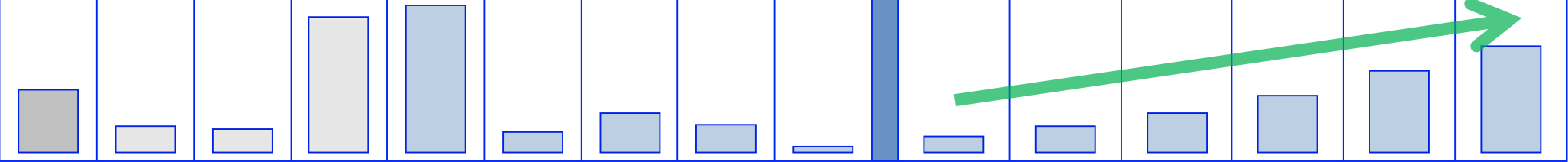
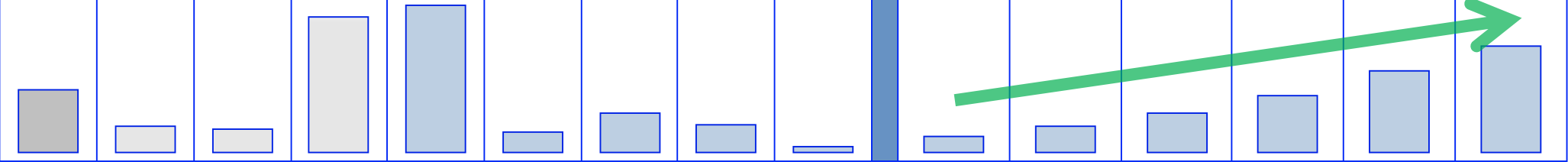
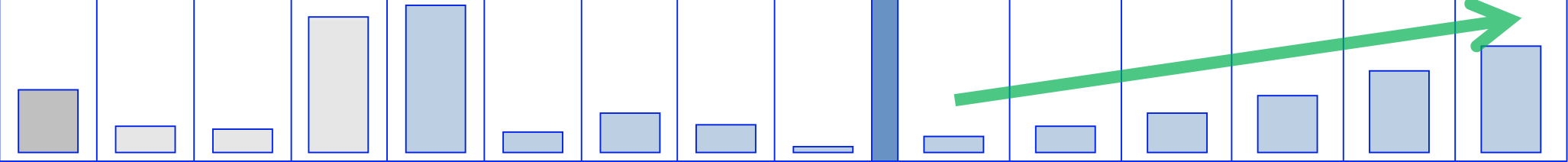
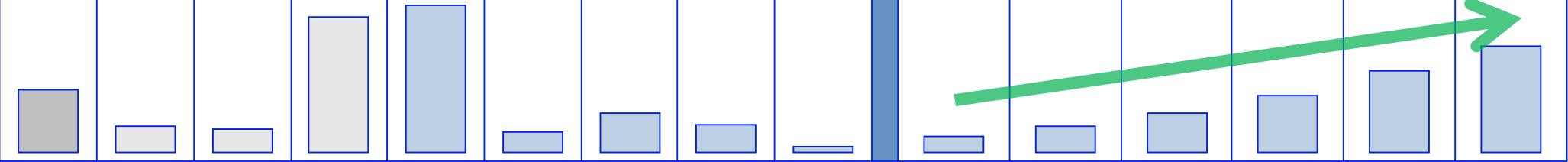
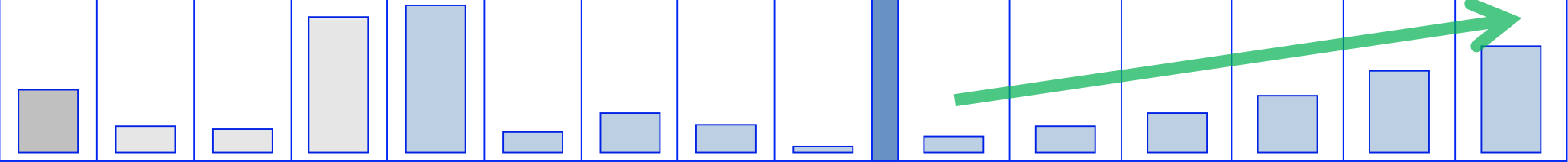
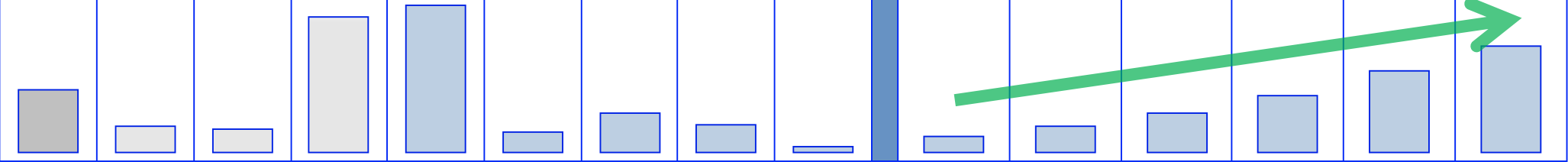
用 COCA 觀察 *seldom* 的詞頻

# 英語的變化：新用法

- 語料庫有助我們決定是否需要納入某些通俗用法 (colloquial usage)
  - *He's like*
  - *I was like*
- 透過時間軸 (time axis) 和體裁分類 (genre categorization), 我們亦可以得知一個趨勢的持續發展

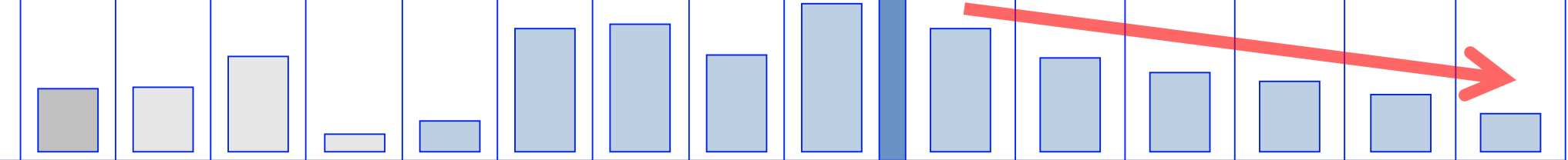
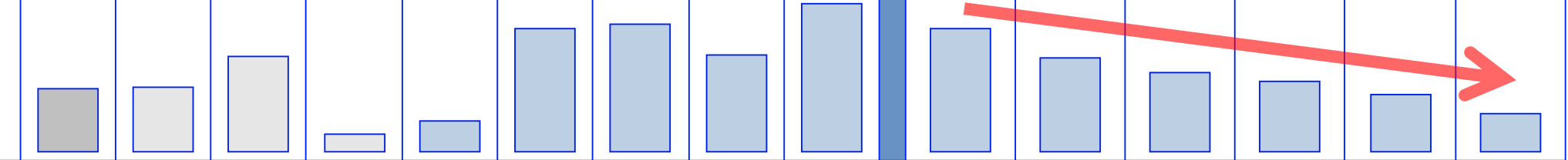
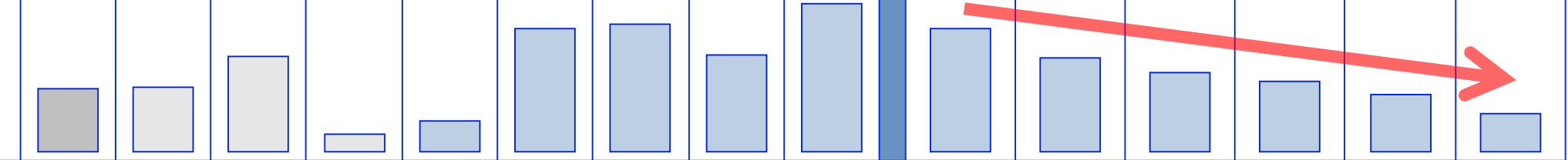
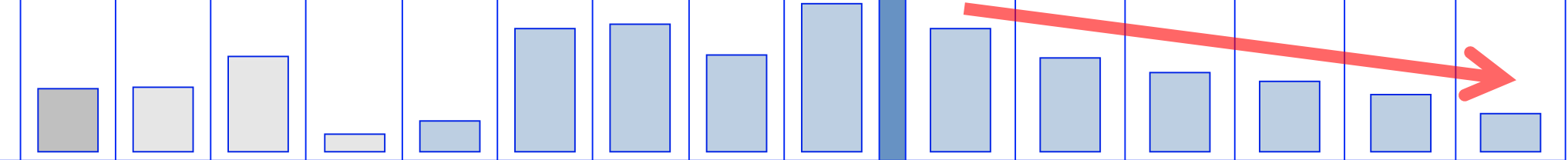
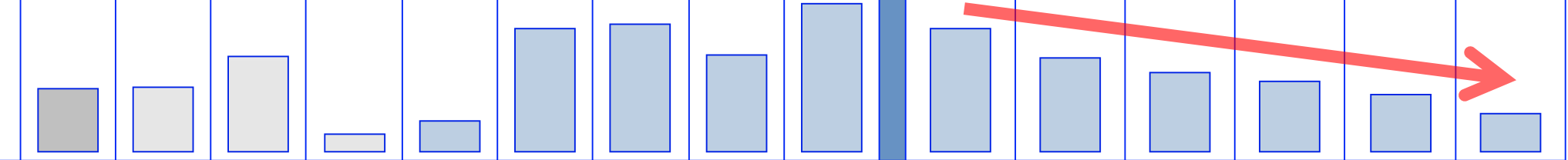
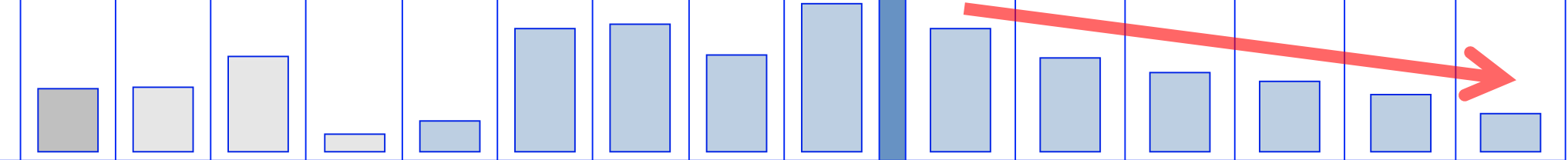
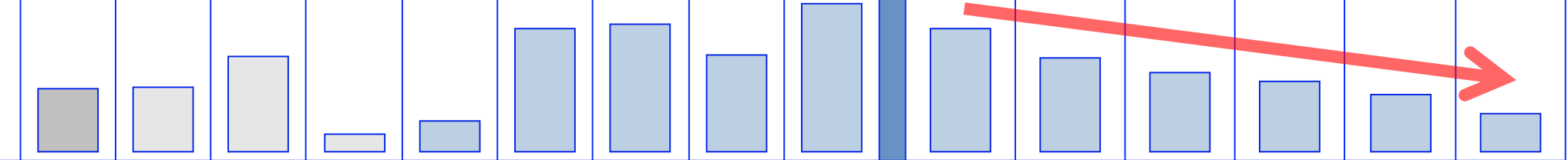
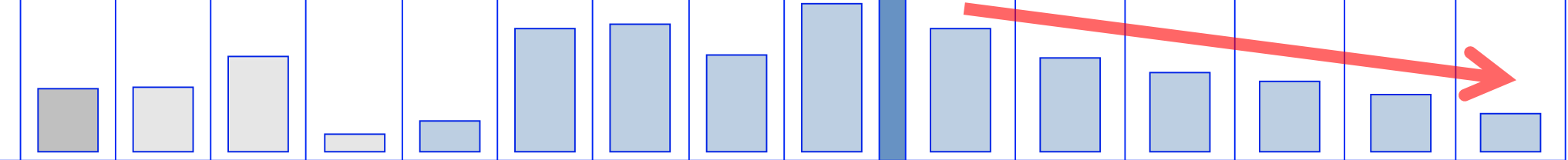
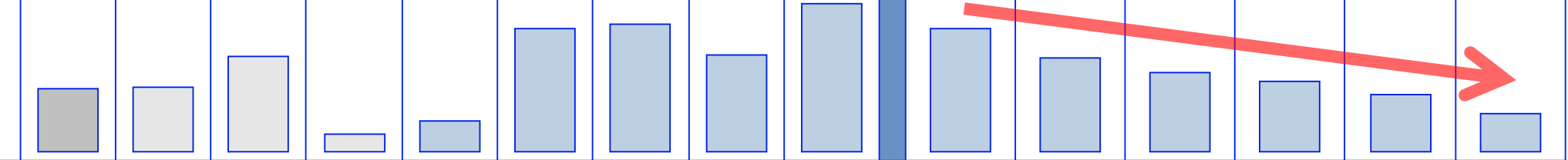
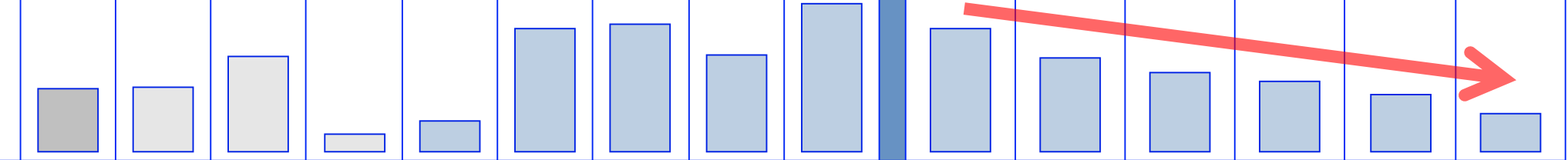
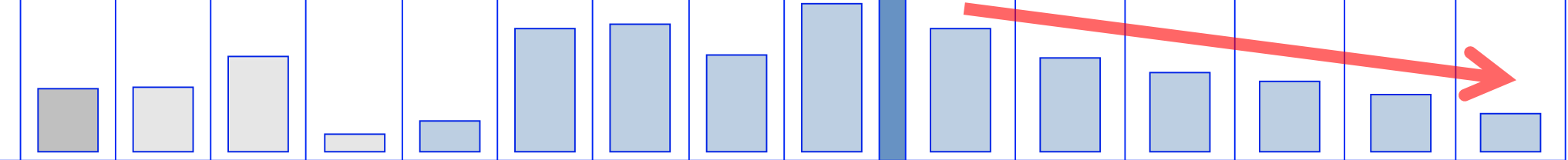
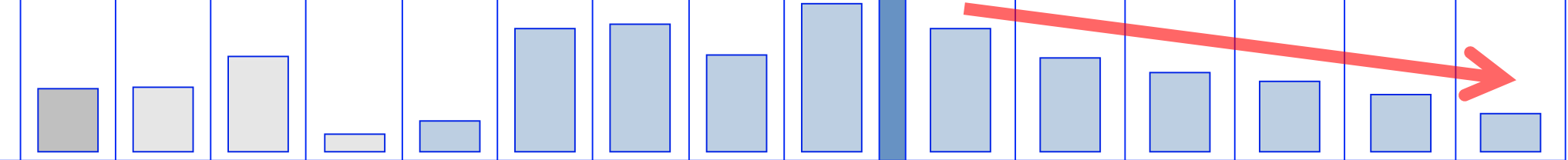
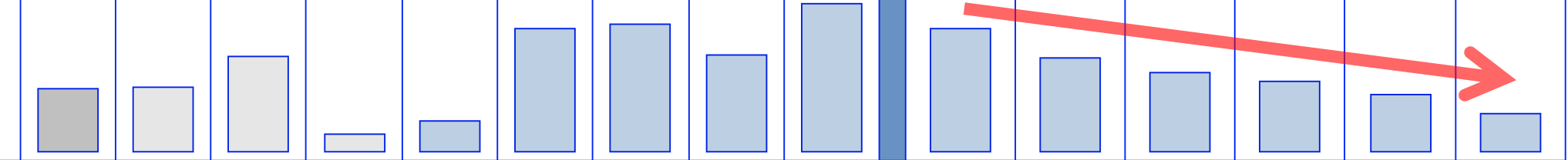
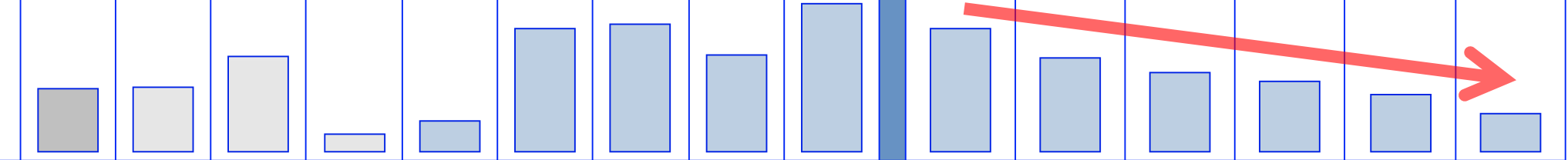
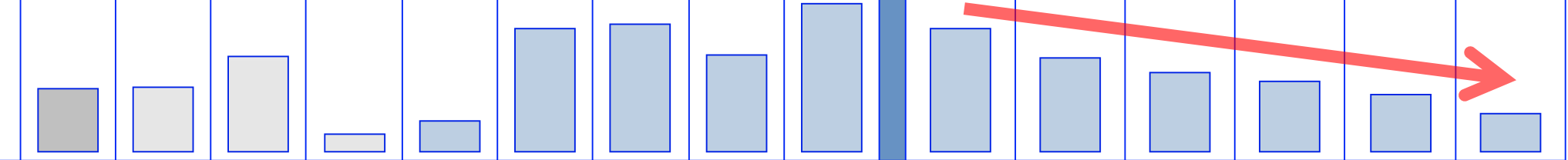
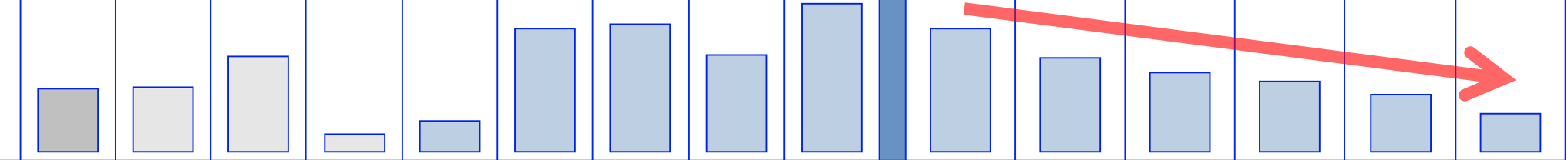
# I am/was/be like

Search string: I BE like

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD		1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	15755	989	800	5170	5561	667	1443	991	134		617	1094	1655	2439	3543	4618
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8		139.1	147.8	146.6	144.9	145.3	144.7
PER MIL	15.87	7.69	6.44	40.37	44.09	5.64	11.44	8.14	1.12		4.44	7.40	11.29	16.83	24.39	31.91
																

# seldom

為何小學英文教科書都會教“seldom”，而不教“I’s like”？

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD		1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	8562	809	1168	215	378	1449	1605	1173	1765		1703	1372	1146	1013	809	542
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8		139.1	147.8	146.6	144.9	145.3	144.7
PER MIL	8.62	6.29	9.40	1.68	3.00	12.25	12.73	9.64	14.73		12.25	9.28	7.82	6.99	5.57	3.74
SEE ALL SUB-SECTIONS AT ONCE																

# 英語的變化：語法

- 我們真正認識英語語法嗎？

– *ing*

**進行式 (progressives)：單一的連續事件**

# -ing

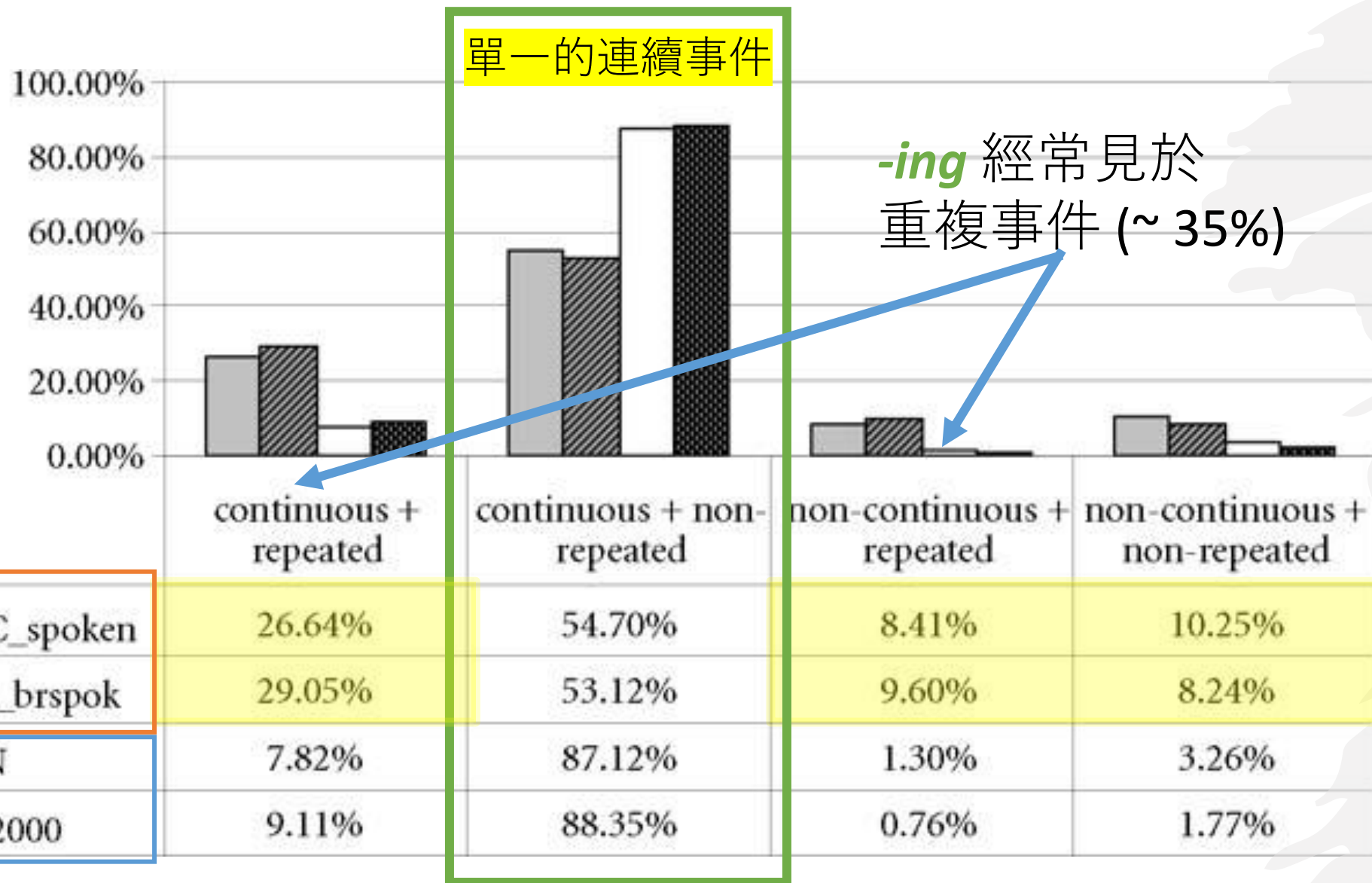
## 最常見用法

- **單一**的連續事件  
(Continuous single events)
  - *I was eating when you called.*

## 其他用法

- **重複**而連續的行為  
(Repeated, continuous actions)
  - *I was eating too much.*
- **重複**而非連續的行為  
(Repeated, non-continuous actions)
  - *I've been buying all the bike magazines lately.*

Examples modified from BNC\_spoken & BoE\_brspek data



實際語料

BNC\_spoken,  
BoE\_brspok

■ BNC\_spoken

▨ BoE\_brspok

教科書例子

GLN, EG 2000

□ GLN

■ EG 2000

英文進行式 (progressives) 的情況 (Römer, 2005)



# 小結 從數據認識英語

---

- 從頻率轉變觀察變化
  - 過時的用法（如 *seldom*）
  - 新近出現的用法（如 *I'm like*）
- 從數據分析找出教科書中未能充分體現的用法 (under-representation)
  - 如 *-ing* 的重覆用法 等 ...

# 語言數據的兩種用法

---

- 了解最新發展 (What)：用語料庫觀察變化
  - **工具：語料庫** (Corpus, pl. corpora)  
Corpus of Contemporary American English (COCA)  
<https://www.english-corpora.org/coca/>
- 活用最新材料 (How)：分析文章難易度
  - **工具：詞彙分析器** (Vocabulary Profiler)

# 活用最新材料

---

客觀方法分析文章程度

# 詞頻為何重要？

- **齊夫定律 (Zipf's Law)** 描述了單詞頻率的冪律分佈 (power-law distribution)
- 排名序號與出現頻率成反比
  - 只有極少數的詞被經常使用，而絕大多數詞很少被使用
- 基於這種分佈，先學習出現頻率最高的單詞並遵循此順序，對學習者極其有益

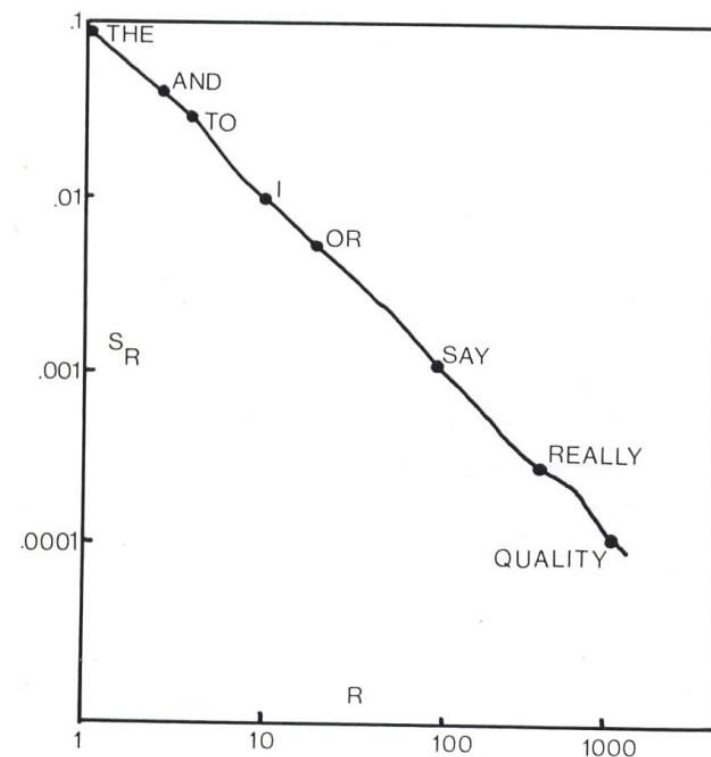
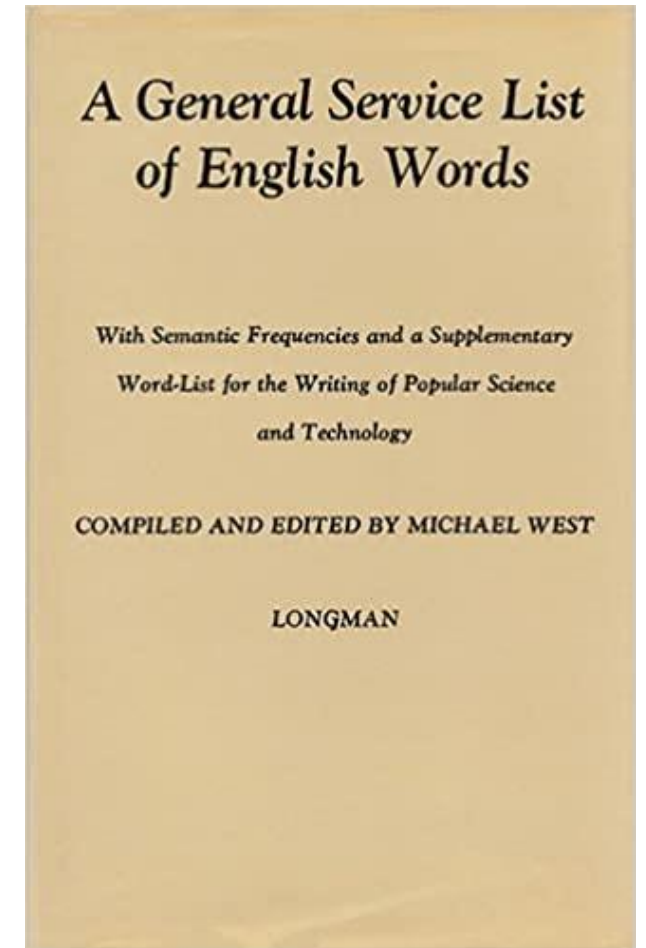


Figure 1 Frequency of word usage in English

# Michael West (1953)

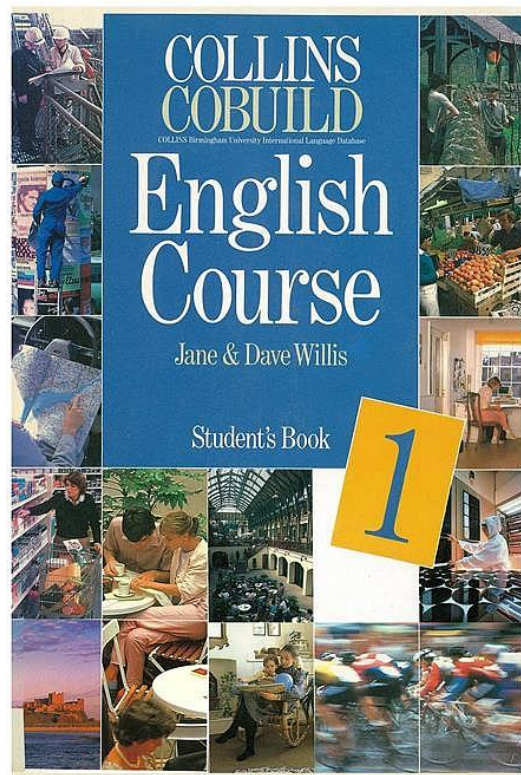
- **General Service List of English Words (GSL)**
  - 語言教學中使用頻率數據的早期嘗試
  - 包含2,000個詞族 (word families)



<https://www.amazon.com/General-Service-List-English-Words/dp/0582525268>

# 現代例子 ①

- **Collins COBUILD English Course (CCEC)** (Willis & Willis, 1989)
  - 基於語料庫數據建構而成



## 1 The syllabus

The main innovation is the lexical syllabus on which the books are based. The Collins–Birmingham University database of authentic text has been used to identify the most useful words and patterns of contemporary English, from which 700 appear at Level 1, 850 at Level 2, and 950 at Level 3. The fact that the course is linked with the *Collins–COBUILD* dictionary project means that the writers have been able to include a Lexicon at Level 2 and authentic dictionary extracts at Level 3. These are used as a basis for exercises such as the following:

### Informal and other uses

**informal** /ɪnfɔːməl/. You use **informal** to describe behaviour or speech that is relaxed and casual rather than correct and serious. In this dictionary, language of this kind is indicated by the word 'Informal' in the Extra Column. EG ...*a relaxed and quite informal discussion*. ◇ **informally**. EG ◇ ADV ... *people talking informally together*.

brainy dim dumb thick loaded smart

- a Look up these words and find which have informal uses. Which word has a different use in American English? Which can be offensive?

(Level 3: 70)

Taylor, L (1991). Reviews, *ELT Journal*, 45 (1) 74–77.  
<https://doi.org/10.1093/elt/45.1.74>

## 現代例子 ②

- Paul Nation為ESL/EFL學習者彙編了BNC/COCA的單詞清單  
<https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists>
- 此清單可與vocabulary profiler一起使用
- Nation, I.S.P. (2017). The BNC/COCA Level 6 word family lists (Version 1.0.0) [Data file]. Available from  
<http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>

# 詞彙分級

---

- 一定要每次參考語料庫嗎？
  - 採用真實語料
  - 但用語料庫相關工具，確認材料深淺程度
- **詞彙分析器 (vocabulary profiler)** 在與分為多個級別的詞表進行比較的基礎上，視覺化地呈視指定文本的程度和複製性



# 分級讀本 (graded readers)

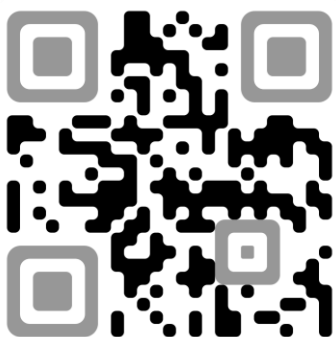
- 語料庫單詞表是分級讀本的基礎
- 受控詞表 (controlled vocabularies) 編寫而成的材料



<https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/readers>



<https://freegradedreaders.com/wordpress/>



# Lextutor

## VocabProfiler

<https://www.lex tutor.ca/vp/eng/>

	Families	Types	Tokens	Percent
<b>K1 Words (1-1000):</b>	86	96	<b>172</b>	<b>85.15%</b>
Function:	...	...	(115)	(56.93%)
Content:	...	...	(57)	(28.22%)
> Anglo-Sax	...	...	(29)	(14.36%)
<b>K2 Words (1001-2000):</b>	7	7	<b>7</b>	<b>3.47%</b>
> Anglo-Sax	...	...	(4)	(1.98%)
1k+2k			...	(88.62%)
<b>AWL Words:</b>	4	4	<b>4</b>	<b>1.98%</b>
> Anglo-Sax	...	...	( )	(0.00%)
<b>Off-List Words:</b>	<u>?</u>	<u>16</u>	<u>19</u>	<u>9.41%</u>
	97+?	123	202	100%

Page menu: Jump down to ~ | [Types](#) | [Families](#)

Note: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns and program decisions); punctuation is eliminated except for 'a' and 'I.'

[Edit-to-a-Profile](#)

Original text [\[Lengthen\]](#)[\[Shorten\]](#)[\[Narrow\]](#)[\[Widen\]](#) space]

RECATS POST NONE

Profiled text Edit, Check with [RE-VP](#), and [SAVE](#)

The grown-ups' response, this time, was to advise me to lay aside my drawings of boa constrictors, whether from the inside or the outside, and devote myself instead to geography, history, arithmetic, and grammar. That is why, at the age of six, I gave up what might have been a magnificent career as a painter. I had been disheartened by the failure of my Drawing Number One and my Drawing Number Two. Grown-ups never understand anything by themselves, and it is tiresome for children to be always and forever explaining things to them.

So then I chose another profession, and learned to pilot airplanes. I have flown a little over all parts of the world; and it is true that geography has been very useful to me. At a glance I can distinguish China from Arizona. If one gets lost in the night, such knowledge is valuable.

Current profile

%	Cumul.
<b>85.15</b>	85.15
<b>3.47</b>	88.62
<b>1.98</b>	90.60
<b>9.41</b>	100.00

the grown ups response this time was to advise me to lay aside my drawings of boa constrictors whether from the inside or the outside and devote myself instead to geography history arithmetic and grammar that is why at the age of six i gave up what might have been a magnificent career as a painter i had been disheartened by the failure of my drawing number one and my drawing number two grown ups never understand anything by themselves and it is tiresome for children to be always and forever explaining things to them

so then i chose another profession and learned to pilot airplanes i have flown a little over all parts of the world and it is true that geography has been very useful to me at a glance i can distinguish china from arizona if one gets lost in the night such knowledge is valuable

# 總結

## 👍 數碼轉型全新世代

將數碼科技概念融入英語專業💖



- 「數據」在英語教學中的角色
- 語言學家  
透過數據分析變化、整理成統計數據
- 前線教師
  - ① 用語料庫掌握語言變化
    - 🔑 有懷疑時，檢查語料庫
  - ② 多採用真實語料，用詞語分級的工具，判斷或調整材料深淺

# 學習資源

- 關於語料庫的學習資源



## The Corpus-Aided Platform for Language Teachers (CAP)

Home News **Teacher Training** Workshop Series Teaching Activities Resources Contact Us

### How To Use Corpora?

This session covers some selected useful corpora, concordance tools, and corpus tools for teachers and provides them with the skills needed to explore corpus data and discover patterns of language use.

1. The British National Corpus (BNC)
2. BYU Corpora
3. Corpus Of Contemporary American English (COCA)
4. Word And Phrase
5. Compleat Lexical Tutor
6. AntConc

### Quick Links

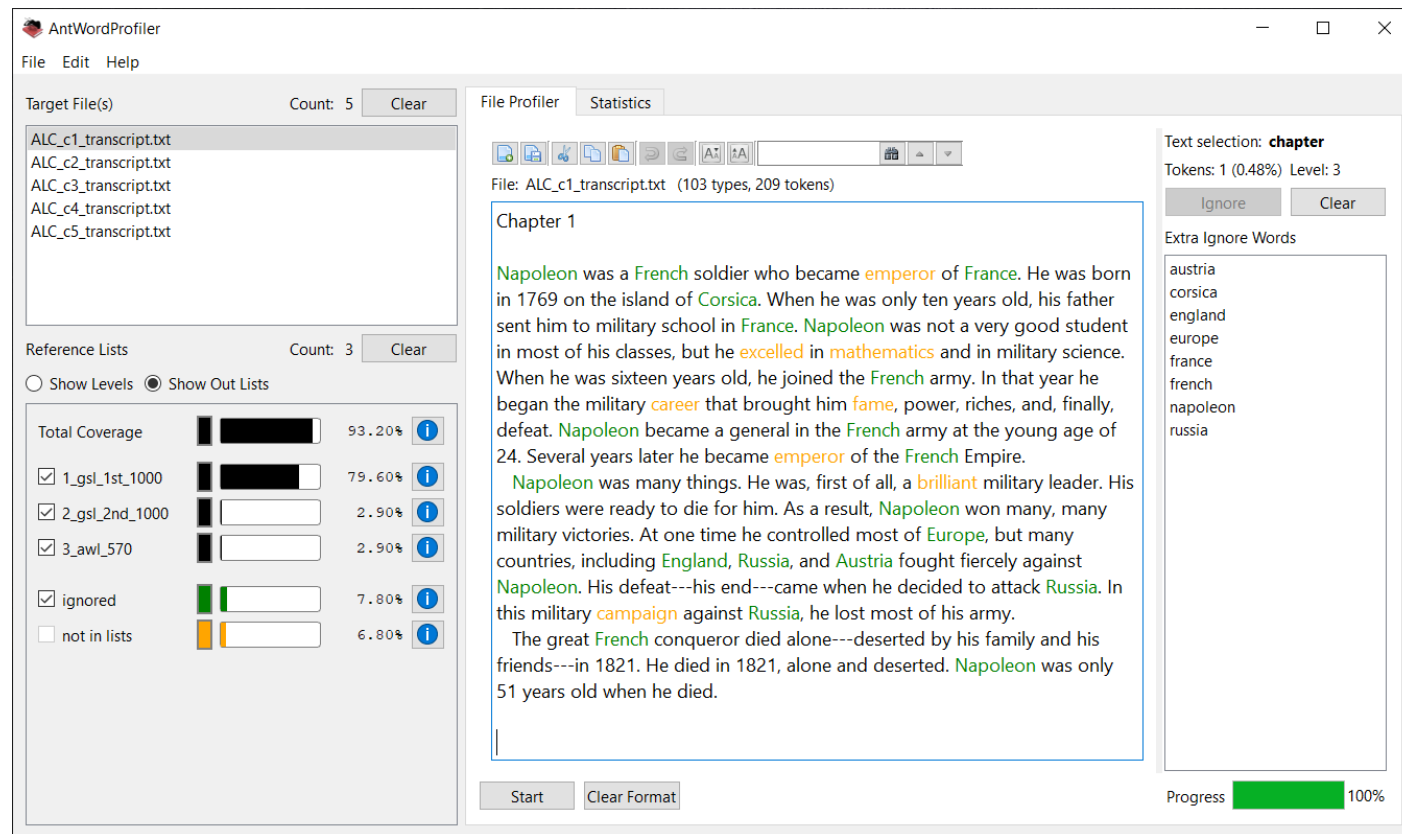
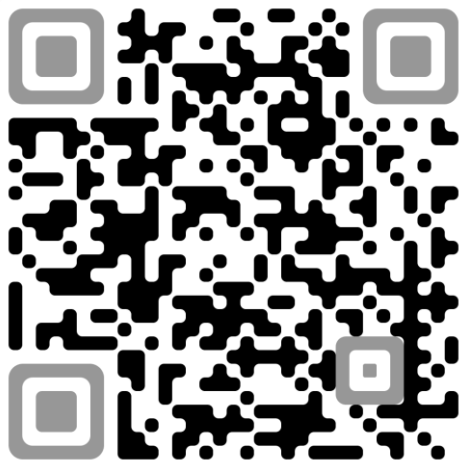
- Crash Course >
- How To Use Corpora? >
- How To Create Corpus-Based Materials For Classroom Use >
- Videos For How To Learn Or Teach With Corpora >
- Workshop Series – Vocabulary >
- Workshop Series – English-Chinese Parallel >
- Workshop Series – Grammar >
- Workshop Series – Pronunciation >

<https://corpus.eduhk.hk/cap/how-to-use-corpora/>

# 其他工具



**Ant Word Profiler**  
*by Anthony Lawrence*



<http://www.laurenceanthony.net/software/antwordprofiler/>