

中央研究院平衡語料庫詞集及詞頻統計

1 簡介

「中央研究院平衡語料庫詞集及詞頻統計 (Word List with Accumulated Word Frequency in Sinica Corpus 3.0)」是根據五百萬詞(token)，帶有標記的中央研究院平衡語料庫所計算出的詞頻統計資料，收錄了語料庫中所有的詞項(type)共 146,876 個詞。

2 排序說明

本詞頻統計是根據每一個詞項在 5 百萬的語料庫中出現次數的多寡，由高至低依序排列。

3 標記說明

每一個詞項皆標示 3 項訊息：

- (1) 詞類標記 -- 是依據中研院詞庫小組所架構的分類系統，訂有非謂形容詞(A)、體詞(N)、動詞(V)、介詞(P)、連接詞(C)、副詞(D)、感歎詞(I)、語助詞(T)等八大類，詳細標記請參閱表一；
- (2) 詞頻 -- 在語料庫中出現的次數；
- (3) 累積頻率 -- 照詞頻排序至此詞的累積頻率(100 為總頻率和)。

表一、詞類標記

標記	說明
A	非謂形容詞
Caa	對等連接詞，如：和、跟
Cab	連接詞，如：等等
Cba	連接詞，如：的話
Cbb	關聯連接詞
D	副詞
Da	數量副詞
DE	的, 之, 得, 地
Dfa	動詞前程度副詞

Dfb	動詞後程度副詞
Di	時態標記
Dk	句副詞
FW	外文標記
I	感嘆詞
Na	普通名詞
Nb	專有名稱
Nc	地方詞
Ncd	位置詞
Nd	時間詞

Nep	指代定詞
Neqa	數量定詞
Neqb	後置數量定詞
Nes	特指定詞
Neu	數詞定詞
Nf	量詞
Ng	後置詞
Nh	代名詞
P	介詞
SHI	是
T	語助詞
VA	動作不及物動詞
VAC	動作使動動詞
VB	動作類及物動詞

VC	動作及物動詞
VCL	動作接地方賓語動詞
VD	雙賓動詞
VE	動作句賓動詞
VF	動作謂賓動詞
VG	分類動詞
VH	狀態不及物動詞
VHC	狀態使動動詞
VI	狀態類及物動詞
VJ	狀態及物動詞
VK	狀態句賓動詞
VL	狀態謂賓動詞
V_2	有

表二、範例 (頻率最高的前 100 個詞)

詞項	詞類	頻率	累積頻率	詞項	詞類	頻率	累積頻率
的	DE	285826	5.82	著	Di	11026	23.14
是	SHI	84014	7.53	她	Nh	10776	23.36
一	Neu	58388	8.72	那	Nep	10740	23.58
在	P	56769	9.88	上	Ncd	10619	23.79
有	V_2	45823	10.81	但	Cbb	10242	24.00
個	Nf	41077	11.64	年	Nf	10127	24.21
我	Nh	40332	12.47	還	D	9698	24.41
不	D	39014	13.26	可以	D	9671	24.60
這	Nep	33659	13.95	時	Ng	9565	24.80
了	Di	31873	14.59	最	Dfa	9416	24.99
他	Nh	30025	15.21	自己	Nh	9069	25.17
也	D	29646	15.81	爲	P	9026	25.36
就	D	29211	16.40	來	D	8992	25.54
人	Na	24269	16.90	所	D	8873	25.72
都	D	20403	17.31	他們	Nh	8818	25.90
說	VE	19625	17.71	兩	Neu	8692	26.08
而	Cbb	18452	18.09	各	Nes	8651	26.26
我們	Nh	18152	18.46	上	Ng	8650	26.43
你	Nh	17298	18.81	可	D	8508	26.60
了	T	15958	19.14	爲	VG	8369	26.77
要	D	15955	19.46	或	Caa	8317	26.94
之	DE	15893	19.78	好	VH	8304	27.11
會	D	14066	20.07	等	Cab	8070	27.28
對	P	13944	20.36	又	D	8037	27.44
及	Caa	13758	20.64	將	D	7858	27.60
和	Caa	13585	20.91	後	Ng	7752	27.76
與	Caa	13445	21.19	因爲	Cbb	7592	27.91
以	P	13172	21.45	於	P	7395	28.06
很	Dfa	13013	21.72	由	P	7344	28.21
種	Nf	12263	21.97	從	P	7303	28.36
中	Ng	12231	22.22	更	D	7298	28.51
的	T	11580	22.45	被	P	7272	28.66
大	VH	11577	22.69	才	Da	7266	28.81
能	D	11125	22.92	已	D	7256	28.95

詞項	詞類	頻率	累積頻率	詞項	詞類	頻率	累積頻率
者	Na	7221	29.10	台灣	Nc	6414	31.42
每	Nes	7207	29.25	卻	D	6388	31.55
次	Nf	7087	29.39	地	DE	6329	31.68
把	P	7024	29.54	並	Cbb	6171	31.80
三	Neu	6954	29.68	位	Nf	6015	31.92
什麼	Nep	6729	29.81	得	DE	5969	32.05
問題	Na	6683	29.95	想	VE	5898	32.17
其	Nep	6667	30.09	去	D	5748	32.28
讓	VL	6624	30.22	呢	T	5577	32.40
此	Nep	6599	30.36	學生	Na	5523	32.51
做	VC	6597	30.49	表示	VE	5504	32.62
再	D	6563	30.62	到	P	5468	32.73
所以	Cbb	6529	30.76	公司	Nc	5421	32.84
只	Da	6521	30.89	將	P	5365	32.95
與	P	6519	31.02	如果	Cbb	5336	33.06
沒有	VJ	6510	31.15				
則	D	6476	31.29				