

中研院中文口語詞頻表

本院覽號

公告日期

智財權狀態

24T-1080124

know-how

摘要

本研究為以中研院台灣華語口語對話語料庫（約50萬詞）為基礎所製作之資料表。「中研院中文口語詞頻表」包含台灣成人對話所使用之詞彙、音節、聲調之頻率（個別頻率與累積頻率）與詞類標記等相關訊息。

技術優勢

- 本技術為目前唯一以口語對話語音庫為本之中文詞頻以及音節結構頻率表資料庫，可作為語音相關科技與應用之基礎參照。

應用範圍

- 設計聽力測試與評估材料之依據微軟正黑體
- 設計兒童構音發展測試與評估材料之依據
- 開發華語教學教材之依據

1	詞項 (Word)	詞類 (CKIP_POS)	拼音(Pinyin)	詞音節數 (Word_size)	詞頻 (Frequency)	累積詞頻 (Accumulated Frequency)
2	的	DE	de5		1	15778
3	我	Nh	wo3		1	13999
4	是	SHI	shi4		1	13397
5	你	Nh	ni3		1	7429
6	就	D	jiu4		1	7092
7	有	V_2	you3		1	6991
8	對	P	dui4		1	6705
9	不	D	bu4		1	6677
10	個	Nf	ge5		1	6330
11	他	Nh	ta1		1	5453
12	一	Neu	yi1		1	5301
13	就是	Cbb	jiu4 shi4		2	5260
14	那	Nep	na4		1	4827
15	在	P	zai4		1	4694
16	然後	D	ran2 hou4		2	4473
17	都	D	dou1		1	4419

圖2.中研院口語詞頻表：音節/聲調、頻率

1	Syllable_Tone	Frequency	Syllable	Frequency	Tone	Frequency_tc	Frequency_type
2	shi4	27634	shi	35225	1	105168	272
3	de5	21398	de	22242	2	96586	220
4	wo3	17492	yi	19404	3	129505	250
5	jiu4	13865	wo	17532	4	228182	301
6	dui4	13478	you	15403	5	47559	31
7	you3	13266	jiu	14424			
8	ta1	11088	dui	13558			
9	ni3	9814	bu	12287			
10	bu4	9529	ta	11129			
11	hou4	8639	ni	9883			
12	zai4	8439	hou	8643			
13	na4	7481	zai	8507			
14	zhe4	6918	na	8409			
15	yi1	6799	zhe	7635			
16	ge5	6641	ge	7303			

圖3.中研院口語詞頻表：言談相關詞項、詞分類、詞頻

創作人

曾淑娟



中央研究院
ACADEMIA SINICA