中研院中文口語詞頻表

本院覽號

公告日期

know-how

智財權狀態

24T-1080124

摘要

本研究成果為以中研院台灣華語口語對話語料庫(約50萬詞)做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)	累積詞類 (Accumulated Frequency)
2	的	DE	de5	1	15778	3.8916
3	我	Nh	wo3	1	13999	7.3445
4	是	SHI	shi4	1	13397	10.6488
5	你	Nh	ni3	1	7429	12.4812
6	就	D	jiu4	1	7092	14.2304
7	有	V_2	you3	1	6991	15.9547
8	對	P	dui4	1	6705	17.6085
9	不	D	bu4	1	6677	19.2554
10	個	Nf	ge5	1	6330	20.8167
11	他	Nh	ta1	1	5453	22.1616
12	_	Neu	yi1	1	5301	23.4691
13	就是	Cbb	jiu4 shi4	2	5260	24.7665
14	那	Nep	na4	1	4827	25.9571
15	在	Р	zai4	1	4694	27.1148
16	然後	D	ran2 hou4	2	4473	28.2181
17	都	D	dou1	1	4419	29.3080

圖2.中研院口語詞頻表:音節/聲調、頻率

Syllable_Tone	Frequency	Syllable	Frequency	Tone	Frequency_to	Frequency_type
shi4	27634	shi	35225	1	105168	272
de5	21398	de	22242	2	96586	220
wo3	17492	yi	19404	3	129505	250
jiu4	13865	wo	17532	4	228182	301
dui4	13478	you	15403	5	47559	31
you3	13266	jiu	14424			
ta1	11088	dui	13558			
ni3	9814	bu	12287			
bu4	9529	ta	11129			
hou4	8639	ni	9883			
zai4	8439	hou	8643			
na4	7481	zai	8507			
zhe4	6918	na	8409			
yi1	6799	zhe	7635			
ge5	6641	ge	7303			
	shi4 de5 wo3 jjiu4 dui4 you3 ta1 ni3 bu4 hou4 zai4 na4 zhe4 yj1	de5 21398 wo3 17492 jju4 13865 dui4 13478 you3 13266 ta1 11088 ni3 9814 bu4 9529 hou4 8639 zai4 8439 na4 7481 zhe4 6918 yi1 6799	shi4 27634 shi de5 21398 de wo3 17492 yi jiu4 13865 wo dui4 13478 you you3 13266 jiu ta1 11088 dui ni3 9814 bu bu4 9529 ta hou4 8639 ni zai4 8439 hou na4 7481 zai zhe4 6918 na yi1 6799 zhe	shi4 27634 shi 35225 de5 21398 de 22242 wo3 17492 yi 19404 jiu4 13865 wo 17532 dui4 13478 you 15403 you3 13266 jju 14424 ta1 11088 dui 13558 mi3 9814 bu 12287 bu4 9529 ta 111129 hou4 8639 ni 9883 zai4 8439 hou 8643 na4 7481 zai 8507 zhe4 6918 na 8409 yi1 6799 zhe 7635	shi4 27634 shi 35225 1 de5 21398 de 22242 2 wo3 17492 yi 19404 3 jiu4 13865 wo 17532 4 dui4 13478 you 15403 5 you3 13266 jiu 14424 ta1 11088 dui 13558 ni3 9814 bu 12287 bu4 9529 ta 11129 hou4 8639 ni 9883 zai4 8439 hou 8643 na4 7481 zai 8507 zhe4 6918 na 8409 yi1 6799 zhe 7635	shi4 27634 shi 35225 1 105168 de5 21398 de 22242 2 96586 wo3 17492 yi 19404 3 129505 jiu4 13865 wo 17532 4 228182 dui4 13478 you 15403 5 47559 you3 13266 jju 14424 4 4 47559 ta1 11088 dui 13558

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

