

機械処理のための中国語拼音の調査

5F-7

金明哲 田中 栄一 丁光躍
 長春郵電学院 宇都宮大学 西南交通大学

1. はじめに

近年、中国語の計算機処理の研究が進んでいる。中国語を拼音で計算機に入力したり、中国語音声の機械認識をするとき、中国語の言語情報を有効に利用しなければならないことは疑いない。そこで中国語の性質を知るために文献2)の中国語高頻度単語6321語について、声母、韻母の出現頻度、声調分布、字数による単語長、声母数に基づく単語の分布、同字数単語中に占める近距離単語数などの調査を行なった。

2. 拼音について

拼音 (pin yin) は中国語の字の発音を表わす記号であり、中国式ローマ字綴りで表記される。大きく声母と韻母に分けられる。一漢字の音は「韻母」(音節)あるいは「声母+韻母」(音節)と音節の音の高さ変化を表わす記号「声調」によって表わされる。同じ拼音でも声調が異なっていれば、全く違う字を意味する。

拼音の声母は24個、韻母は音形によると37個ある。しかし、i は3種類の発音があるので、音の種類は39個である。即ち i が zh, ch, sh の後に続く時 [i] と発音し、z, c, s後に続く時 [ɿ] と発音し、j, q, x などに続く時 [i] と発音する。

3. 調査結果

3.1 声母、韻母の分布

声母を出現頻度の高低順に並べると d, y, sh, j, zh, w, b, l, x, g, h, t, z, q, n, m, r, ch, f, k, c, s, p となる。文献4)では w と y はある種の韻母の前につく発音されない記号とされている。韻母を出現頻度の高低順に並べると i, e, u, uo, a, ai, ian, eng, en, ong, ao, an, ing, ui, iu, ang, ü, ie, iang, iao, ou, in, ei, uan, ia, üe, üan, un, uan, er, iong, ua, uai, ün, o, ueng となる。表1は声母の分類とその出現頻度を表わし、表2は韻母の分類とその出現頻度である。声母では無気音の出現頻度が一番高く、全声母の約40%を占め、韻母では開口韻母の出現頻度が一番高く、全韻母の約52%を占める。

3.2 単語長の分布と声調の分布

中国語の最小表記単位は字である。単語を字数によって分類した結果を図1に示す。一漢字単語(以下で一字単語という)は全辞書の約29%であるが出現頻

度は約69%を占めることが分かった。図2は声調出現頻度分布である。四声の出現頻度が一番高く、轻声の出現頻度が一番低い。

表1. 声母の分類と出現頻度

調音点	調音の仕方						合計
	無音	有気音	摩擦音と流音	鼻音	半母音		
唇音	b 4.98	p 0.72	f 1.77	m 2.54	w 5.33		15.34
舌音	d 13.88	t 3.98	l 4.72	n 2.75			25.33
齒音	z 3.92	c 1.06	s 0.96				5.94
舌音	zh 6.06	ch 1.88	sh 8.23		r 1.93		18.10
舌面音	j 6.82	q 2.81	x 4.71		y 11.28		25.60
舌根音	g 4.51	k 1.16		ng 0			5.67
喉音			h 4.02				4.02
合計	40.17	11.61	24.41	5.29	18.52		100.00

表2. 韻母の分類と出現頻度

開	口	齊	合	口	撮		
i	9.42	i	8.11	u	5.11	u	2.73
a	4.30	ia	1.04	ua	0.40		
o	0.19	uo	5.09				
e	11.45	ie	2.02		ue	0.95	
ai	3.99			uai	0.38		
ei	1.47			ui	2.87		
ao	2.97	iao	1.95				
ou	1.93	iou	2.85				
an	2.96	ian	3.92	uan	1.28	uan	0.86
ang	2.38	iang	1.99	uang	0.66		
en	3.43	in	1.88	un	0.73	un	0.35
eng	3.54	ing	2.95	ueng	0		
ong	3.28	iong	0.44				
er	0.49						
合計	51.80	27.15	16.52	4.53			

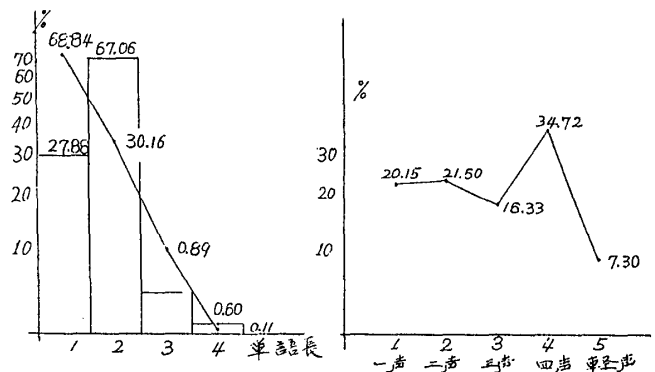


図1 単語の分布

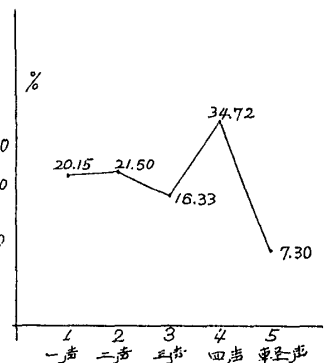


図2 声調の分布

Statistical investigation of Chinese words(pin-yin) for machine processing.
 Jing Ming Zhe¹, Eiichi Tanaka² and Ding Gaung Yao³.

¹Chang Chun Electoro Communcation Univ., ²Utsunomiya Univ. and ³South/West Jing Tong Univ.

