

普通话声母和韵母的统计特性

孙金城 倪 宏 莫福源 李昌立

(中国科学院声学研究所)

1994 年 1 月 18 日收到

摘要 本文对普通话书面语中声母、韵母的动态与静态分布特性及其差异作了统计分析,结果表明:普通话声母间的、韵母间的动态与静态的相对分布关系一致,语音间的相对分布主要与发声系统有关,不受频度的影响。普通话声母、韵母的动态与静态的出现率差异,与声母发音方法和韵母组合结构、声母发音部位与韵母四呼的配合关系、音节的成字率和字的频度有关,主要受送气与不送气声母、零声母和单韵母、鼻韵母的分布变化的影响。这说明语言和人类其他行为一样,也受最小用力原理制约。普通话声母、韵母的动态与静态的分布差异,主要反映在单音节词上,单音节词的声母、韵母的动态与静态的出现率差异最大。多音节词中的韵母的动态出现率与多音节词的音节数量有关,多音节词中声母的动态出现率似乎与多音节词的音节组合结构有关。

关键词 普通话,声母韵母,统计特性

The Statistical Distribution of Standard Chinese Initials and Finals

Sun Jincheng, Ni Hong, Mo Fuyuan, Li Changli

(Institute of Acoustics Academia Sinica, Beijing 100080)

Abstract This paper presents a statistical analysis of the dynamic (in usage) and static (in

选用较高频率和较长声程的探头,以缩小测量范围和提高灵敏度及精度。

限于试验条件,本文未系统地进行定量的气泡粒径对含气量测量的影响研究。针对我们所建立的试验容器,其底部出口气泡直径为1mm,随着气泡上浮,到达水面附近,气泡直径已增至约5mm左右。在测量段范围内,气泡变化大约在2—4mm内。从目前的试验情况看,在这个气泡大小范围,气泡粒径对含气量的测量无明显的影响。

水中含沙量大小,对超声衰减有一定的影响,对于天然沙($\rho = 2.65\text{g/cm}^3$),每 8kg/m^3 含沙量约折合1%的含气量。

水温对含气量测量的影响非常小。从理论上说,水温不同,粘滞系数不同,因而将导致声吸收系数的不同。但是,由于气泡对超声衰减异常敏感,相比之下,水温变化($5^\circ\text{C}—30^\circ\text{C}$)对衰减的影响则可忽略不计。

掺气水流下,流速变化对衰减测量的影响,还有待于进一步的试验研究。从前期的“超声衰减法含沙量测试研究”课题^[5]的试验情况来看,流速对含沙水流的声衰减影响不大。由于目前室内不易实现高流速下自然掺气模拟野外实际情况,因此,此类试验只能在野外观测中逐步收集资料进行研究。

(下转第47页)

lexicon) distribution characters as well as their differences of the initials and the finals in Chinese writing language. The results show that the dynamic and static relative distributions are congruent among the initials and among the finals in Mandarin. The phonetic relative distributions are mainly concerned with the manner of pronunciation and are not influenced by the occurrence frequency. There difference is related to following factors: the manner of initial articulation and followed final, the combination of the initial articulatory position and four kinds pronunciation of the final, the frequency of construction of syllable into a word and the word occurrence. The difference is mainly influenced by the distribution variation of aspirated and unaspirated initial, zero initial, monofinal and nasal final. This shows that behaviour in speech is also ruled by less-effort principle like other human behaviours. Prominent difference between the dynamic and static distribution of the initial and the final in Mandarin is shown mainly in monosyllabic words. In polysyllabic words, the dynamic occurrence frequency of the final is related to the number of syllables; for the initial, it seems to be related to the structure of the syllable combination.

Key word Mandarin, the initial and the final, statistical character

1 引言

定量分析普通话语音的分布特性,可为定性的语音评价提供量的依据,对汉语人机对话系统研究、推广普通话和对外汉语教学等都有实际意义。例如,可为设计普通话语音试验材料或评估语音通讯系统质量的语音测试材料等提供量的依据。

普通话有 22 个声母(包括零声母)和 39 个韵母。韵母中的 er 自成音节。韵母中的 ê 只在单音字“诶”(极少使用)和 e 前是韵头 i 和 ü 时才出现,如,jiě(姐)。普通话中“可能构成的基本音节数目应该有 814 个,但是实际能够出现的音节只有 410 个左右”,“普通话音节总数”分四声计算“约有 1250 个左右”^[1]

汉语是声调语言,音节由声母、韵母、声调三个部分构成。音节是最小的语音结构单位,同时也是听感上最自然的单位。在自然语言中,人们听不到、说不出不带声调的普通话音节,但在分析声母和韵母的配合关系时,可以不包括声调。为讨论方便,我们暂将“声母和韵母的搭配组合”称为“声韵组合”,以免与音节混淆。

散见于各家著述中有关声韵组合的数目说法不一:如“现代汉语”,张志公,400;《现代汉

语词典》,419;《新华词典》,410。其中有 20 多个声韵组合是彼收此舍或是彼舍此收,至今尚无统一的说法。好在这些有争议的声韵组合多在方言、口语、古汉语、语气词等中出现,出现率也低,对声韵组合分布的统计影响甚微。如:fiao、cei、nun、yo、chua 等,《现代汉语常用词频词典》^[2]中出现了 399 个声韵组合,这些有争议的声韵组合大都没有出现。

已有文章发表过 18300 个词条的声母、韵母的动态分布^[3]。本文以《现代汉语常用词频词典》全部 47016 个词条为对象,统计普通话声母、韵母的动态与静态分布及其差异,并比较它们在单音节词、多音节词中的分布规律。

2 材料

《现代汉语常用词频词典》(以下简称词典)是从三亿原始母体语料中抽样 2500 万汉字后的词频统计成果,共收录 47016 个词条(含多音字或词):单音节词 5416、双音节词 31333、三音节词 5129、四音节词 4566、五音节词 382、六音节词 144、七音节词 46。词典中词条频度是指某个词条在 2500 万汉字语料中的统计次数;词典中词条频率即出现率,是指某个词条的频度占词典词表总频度的百分比。利用这些语料对声母、韵母的分布特性所做的统计,可较为

准确地反映它们在普通话书面语中的分布规律。

语音的静态分布、动态分布是指语音在不带频度参数和带有频度参数的字表或词条中的分布,二者有区别,语音的动态分布反映它们在实际语言使用中的状况。词典中所有词条的频度>5,是动态语料。本文在作声母、韵母的动态与静态分布统计时,令词典中所有词条频度为1,这使动态语料成为不带频度的静态语料,以便在同一语料的基础上考察声母、韵母的动态、静态分布及其差异。例如:词典中“的”和“灯具”的频度分别为849232和5,表明它们在词频统计语料中的实际出现次数。当令“的”和“灯具”以及其它词条频度都为1,即不计频度时,便只说明这些词条在2500万汉字统计语料中出现。这时《现代汉语常用词词频词典》成了不带频度参数的“现代汉语常用词词典”,跟普通的非词频词典一样,是静态语料。

词典中双音节词有31333个,数量最多,占词条总数的67%,其出现率为39.25%。单音节词有5416个,比双音节词少得多,只占词条总数的12%,而单音节词多为常用字,出现率却高达57.53%。单音节词、双音节词占了全部词汇出现率的97.18%,因此,我们在探讨普通话语音分布规律时,不仅要考察语音静态数据,更要重视动态数据的作用;在言语工程、语音教学等实际应用中,应对单音节词、双音节词予以同样的重视。

3 声母、韵母在普通话词汇中的分布特性

表1、表2是声母、韵母在普通话词汇中的出现率排序。

图1是按声母发音方法和韵母结构归类的声母、韵母出现率的动态与静态分布图。

图2是按声母发音部位和韵母四呼归类的声母、韵母出现率的动态与静态分布图。

图1、图2是归纳表1、表2的数据得到的,图中序号是归类后的声母、韵母的动态与静态的出现率的降序排列序号,用来表示声母间的、

表1 声母在普通话词汇中的动态、静态出现率*

序号	声母	动态出现率%	声母	静态出现率%
1	零*	14.016963	零	12.372531
2	d	11.345391	j	7.744040
3	sh	7.520778	sh	6.817770
4	j	7.289815	d	6.462906
5	zh	6.864079	x	6.454321
6	l	5.288978	l	6.088964
7	x	5.145222	zh	5.864789
8	g	4.811546	h	5.192265
9	b	4.470702	b	4.937565
10	h	4.467683	g	4.792567
11	z	3.754430	t	4.100964
12	t	3.413512	q	3.848172
13	q	3.204953	ch	3.695542
14	ch	2.966524	f	3.594425
15	m	2.957014	m	3.526696
16	f	2.786016	z	2.980091
17	r	2.052069	s	2.190234
18	n	1.822816	k	2.070992
19	s	1.816471	r	1.993723
20	k	1.622787	p	1.991815
21	c	1.306396	n	1.704681
22	p	1.075855	c	1.574946

* 表中序号是声母或韵母出现率的降序排列序号。

* 零指零声母

* 单韵母 i 中包括:

	动态出现率%	静态出现率%
舌面元音 i	9.71	7.72
舌尖前元音 [ɿ]	1.49	3.36
舌尖后元音 [ʅ]	5.56	3.97

韵母间的语音相对分布关系。

普通话是以北京语音为标准音。统计普通话语音的分布特性,实际上就是统计北京话的语音分布特性。“北京人读零声母音节时,在前面也加上喉音性辅音或半元音”^[4]。鉴于零声母的出现率较高,为全面反映普通话语音的分布特性,我们也把它放在图1或图2中与其它声母一道讨论。

图2中韵母四呼的动态和静态的出现率排序都是开口呼韵母>齐齿呼韵母>合口呼韵母>撮口呼韵母。图2中除唇音与零声母的动态和静态的出现率的序号有交叉外,其它声母的

表 2 韵母在普通话词汇中的动态、静态出现率

序号	韵母	动态出现率%	韵母	静态出现率%
1	i*	16.759741	i	14.874701
2	e	10.588680	u	8.549161
3	u	6.897552	ian	5.181772
4	ian	4.285902	an	4.675233
5	ong	4.090159	ing	4.622767
6	uo	3.737683	ong	4.114320
7	ing	3.584688	ao	3.580116
8	ai	3.507278	e	3.333047
9	an	3.289141	eng	3.204266
10	a	3.284073	ang	3.143214
11	eng	3.138550	a	3.091702
12	ui	2.993787	ui	3.009663
13	en	2.968725	ai	2.875159
14	ao	2.638198	en	2.777857
15	ang	2.593114	iao	2.683418
16	iu	2.555809	in	2.677694
17	ü	2.269335	uo	2.532696
18	iao	2.217050	ou	2.495493
19	ie	2.192076	ü	2.382928
20	in	2.042616	uan	2.181648
21	iang	2.036479	iang	2.141583
22	ou	1.797602	ie	2.081485
23	uan	1.486331	iu	1.902145
24	ei	1.299320	ei	1.502447
25	ia	1.161175	ia	1.353633
26	üan	0.964759	üan	1.327877
27	un	0.930493	uang	1.322153
28	üe	0.916280	un	1.305936
29	uang	0.739083	üe	1.092255
30	er	0.653284	ua	1.069361
31	iong	0.595723	o	0.793674
32	ua	0.584336	ün	0.787950
33	ün	0.513511	uai	0.448349
34	o	0.355125	iong	0.445487
35	uai	0.330234	er	0.429271
36	ueng	0.002106	ueng	0.009539

动态与静态的出现率的序号是一致的。图 1 中除韵母中的舌尖鼻韵母与舌根鼻韵母的动态和静态的出现率的序号有交叉（二者的出现率相差无几）以及声母中的不送气塞音与不送气塞擦音出现率的序号有交叉外，其它声母、韵母的动态与静态的出现率的序号一致。上述情况表明：普通话声母间的、韵母间的动态和静态的

相对分布关系基本一致，反映了普通话语音间的相对分布主要与发声系统有关，而不受频度的影响。

从表 1、表 2 可以看到：声母或韵母在普通话词汇中的动态与静态的出现率有的一致，有的不同。如：表 1 中声母动态、静态出现率累计高达 57.47%、51.80% 的，是 7 个相同的声母，但排列次序不尽一致；表 2 中韵母动态、静态出现率累计高达 53.45%、48.93% 的 8 个韵母中，有 2 个韵母是不同的，而且除韵母 i 外，另外 7 个韵母排列次序完全不同；零声母和 i 韵母的出现率最高，其动态、静态出现率均排在第一位；声母 d、韵母 e 的动态值比静态值分别高出 4.88、7.26 个百分点，它们在动态、静态中的排列次序相差甚远。声母、韵母的动态与静态的出现率差异反映了声母、韵母在词频统计语料中的实际出现（计频度）与出现（不计频度）之间的差异。造成差异的因素有：

1. 声母发音方法和韵母组合结构

由图 1 发现：声母中的不送气塞音、不送气塞擦音、零声母和韵母中的单韵母的动态值高于静态值，其它相反。不送气塞音、不送气塞擦音、零声母和单韵母较之其他声母或韵母发音省力，它们的动态值自然要比静态值高。但是，并非所有的不送气塞音、不送气塞擦音或单韵母都如此，不送气塞音 b、单韵母 u 的动态值就比静态值小，而声母中擦音 sh 的动态值却比静态值大。这是因为声母、韵母出现率不仅与自身属性有关，还跟声母与韵母的搭配组合有关。

2. 声母发音部位与韵母四呼的配合关系

图 2 中，舌尖中音、舌尖后音的动态值大于静态值，主要跟声韵组合 de、shi 的出现率有关，de 和 shi 的动态出现率分别为 43.10%、47.15%，比各自的静态出现率高出 38.85%、19.60%。

普通话中声母能否与韵母配合取决于声母的发音部位和韵母开头元音的性质，它们之间的配合关系有很强的规律性。例如：声母 b 和 d 同是不送气塞音，都能与开口呼、齐齿呼、合

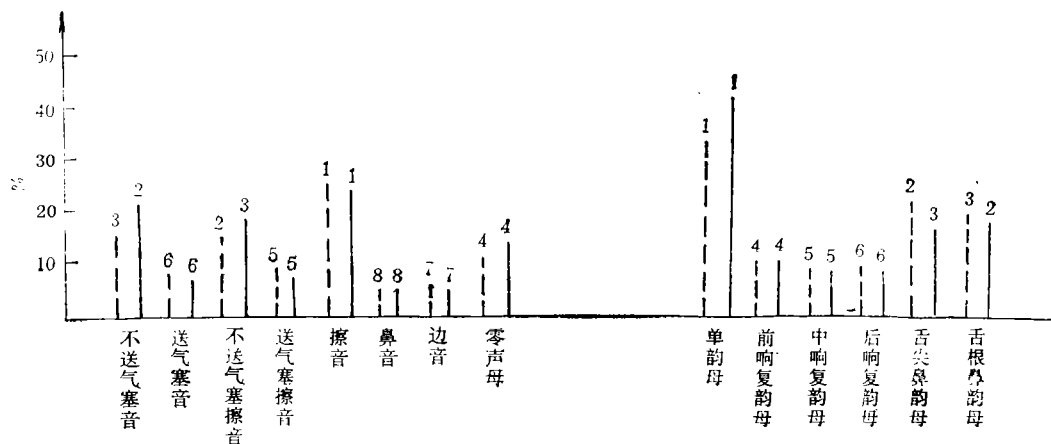


图1 按声母发音方法和韵母结构分类的声母、韵母分布特性

实线: 动态分布 虚线: 静态分布

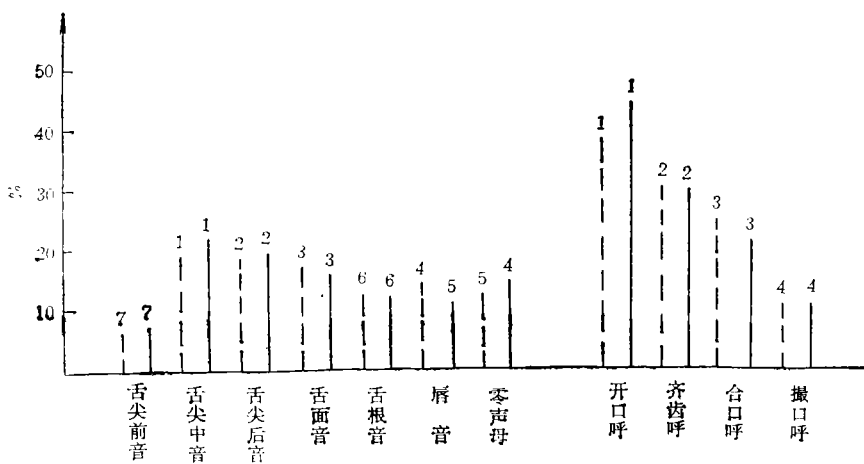


图2 按声母发音部位和韵母四呼分类的声母、韵母分布特性

实线: 动态分布 虚线: 静态分布

口呼韵母相配合。但是, 声母 b 在与合口呼韵母配合时只能与单韵母 u 相拼; 由于“唇音声母或多或少都有唇化(合口)现象”^[4], 唇音 b 与合口呼单韵母 u 的组合自然、发音省力, bu 是声母 b 与韵母配合中出现率最高(29.81%)的声韵组合。擦音 sh 的发音部位是舌尖后, 它与同部位的舌尖后元音 [ɿ] 的组合自然、发音省力, shi 的出现率为 47.15%, 是声母 sh 与韵母配合中出现率最高的声韵组合。声韵组合中声母、韵母的分布是互为影响的。例如, 随着韵母 u 的动态值减小, 与之组合的塞音 b 的动态值减小; 随着韵母 [ɿ] 的动态值增大, 与之组

合的擦音 sh 的动态值增大。

3. 音节的成字率和字的频度

声韵组合配以声调后构成普通话音节, 若按 410 个声韵组合计, 配以四声后应该有 1640 个音节, 有些音节在普通话中没有字, 如, gai。普通话中实际能够出现的音节总数约有 1280 个左右。这个音节总数如同声韵组合的总数一样, 尚无统一的说法。例如, 文献一说约有 1250 个左右, 文献二说有 1327 个。

各声韵组合配以声调构成音节后的音节成字率不同, 有些相差很多。仅以本文所用词典中的单音节的词或字为例, 声韵组合 ji 配以四

声后成字 84 个, 累计频度为 85682; 声韵组合 bi 配以四声后只成字 38 个, 累计频度为 20641。尽管声母 b 可与开、齐、合三呼的韵母相拼, 而声母 j 只与齐、撮两呼的韵母结合, 表 1 中声母 j 的静态、动态排序却都在声母 b 的前面, 说明音节成字率影响声母或韵母的出现率。

声韵组合 de 配以声调后, 只成字 6 个, 仅这六个字的累计频度就达 892193, 其中“的”一个字的频度高达 849232。表 1 中声母 d 的动态值是其静态值的近两倍, 表 2 中韵母 e 的动态值是其静态值的三倍多, 主要与“的”频度有关。可见, 字的频度大小即字的常用程度对声母、韵母的出现率有很大影响。“的”读轻声, 元音 e[ə] 发音时舌位不低不高、不前不后, 不送气塞音 d 是舌尖中音, 它与元音 e[ə] 的组合自然、发音省力, 是“的”和舌尖中音 d、元音 e[ə] 的出现率极高的根本原因。

不难看出, 普通话声母、韵母的动态与静态的出现率差异与声母发音方法和韵母组合结构、声母发音部位与韵母四呼的配合关系、音节的成字率和字的频度有关, 主要受送气与不送气声母、零声母和单韵母、鼻韵母的分布变化的影响。这反映了语言和人类其他行为一样, 受最小用力原理制约, 结构简单、组合自然、发音省力的语音多用, 其它少用。

4 声母、韵母在单、双、三、四音节词中的分布特性

图 3 是按声母发音方法归类的单、双、三、四音节词的声母出现率的动态与静态分布图。

图 4 是按韵母结构归类的单、双、三、四音节词的韵母出现率的动态与静态分布图。

图 3、图 4 中的序号是归类后的普通话词汇中声母、韵母的动态与静态出现率的降序排列序号, 用来表示多音节词间的声母或韵母的语音相对分布关系。由图发现:

1. 多音节词间的(双、三、四音节词间的)声母、韵母的动态与静态的相对分布基本一致。例如: 不送气塞擦音的动态和静态的出现率排列次序都是四音节词>双音节词>三音节词; 单

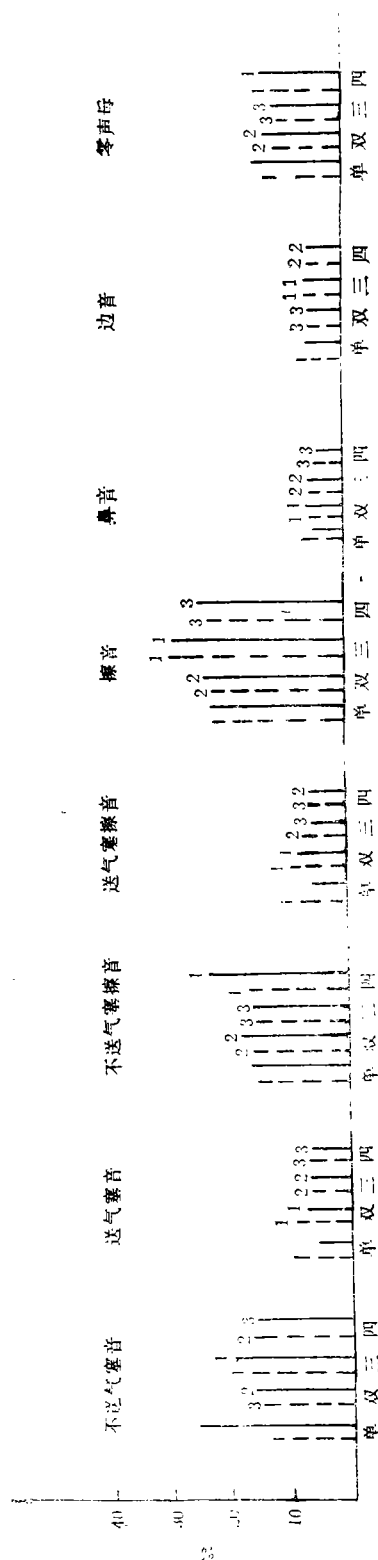


图 3 按声母发音方法分类的单、双、三、四音节词声母的分布特性

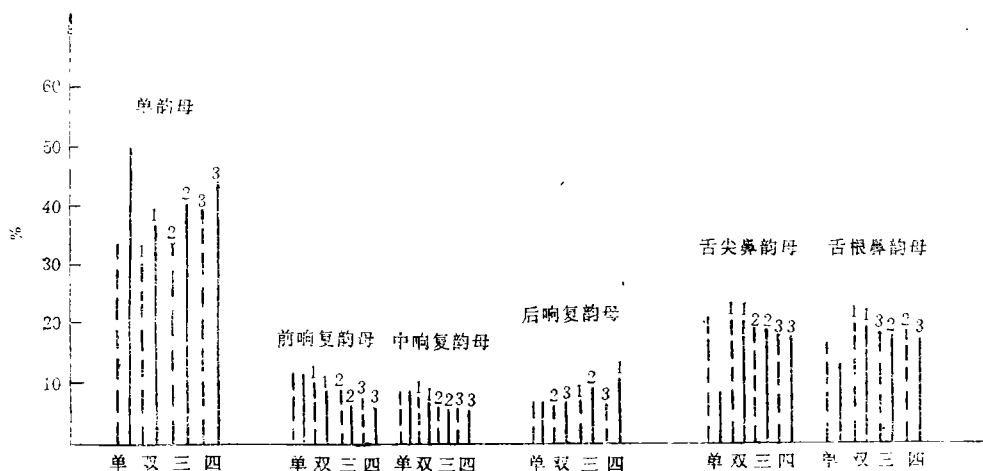


图4 按韵母结构分类的单、双、三、四音节词韵母的分布特性

实线: 动态分布, 虚线: 静态分布, 单: 单音节词, 双: 双音节词, 三: 三音节词, 四: 四音节词

韵母的动态和静态的出现率排列次序都是四音节词>三音节词>双音节词。尽管双、三、四音节词的不送气塞音、送气塞擦音、后响复韵母、舌尖鼻韵母的出现率的序号有些交叉, 但其差值甚微, 对多音节词间的语音相对分布关系影响并不大, 似乎可以说, 多音节词语音间的相对分布不受频度影响。

2. 单音节词声母、韵母的动态与静态的出现率相差悬殊, 而且改变了它与双、三、四音节词间的语音相对分布关系。前述普通话词汇中声母、韵母的动态与静态的出现率差异主要反映在单音节词上, 为送气与不送气声母、零声母和单韵母、鼻韵母的出现率变化所左右。图3中单音节词的送气塞音、不送气塞音、送气塞擦音和图4中单音节词的单韵母、鼻韵母, 它们的动态与静态的出现率相差悬殊, 并改变了它们与双、三、四音节词的相对分布关系。例如, 在单、双、三、四音节词中, 单音节词的不送气塞音的静态出现率最低而动态出现率变为最高, 相反, 单音节词的舌尖鼻韵母的静态出现率最高而动态出现率变为最低。

3. 声母在多音节词中的动态分布似乎与多音节词的音节组合结构有关。如图3中, 依单、双、三、四音节词的顺序看, 除零声母和擦音外, 声母出现率呈现出高低或低高交错的态势。最明显的是不送气塞音、塞擦音在单、双、三、四

音节词上的出现率呈互补态势: b、d、g 在单、三音节词上的出现率高、在双、四音节词上出现率低且相近, j、z、zh 的出现率刚好与其相反; b、d、g 出现率在单音节词上居高, 降至四音节词最低, 而 j、z、zh 出现率却在四音节词上居高, 降至单音节词最低。

4. 韵母在多音节词中的动态分布与多音节词的音节数量有关。韵母在多音节词中的出现率随词的音节数量增加或递升或递减, 而且不受单音节词韵母的动态与静态的出现率的巨大反差的影响。如图4中, 双、三、四音节词的单韵母和后响复韵母的出现率随词的音节数量增多而递升; 双、三、四音节词的前响、中响复韵母、鼻韵母的出现率随词的音节数量增多而递减。

至于为什么多音节词中韵母的动态出现率与多音节词的音节数量有关, 多音节词中声母的动态出现率似乎与多音节词的音节组合结构有关, 尚待进一步研究。

参 考 文 献

- [1] 林焘, 王理嘉, 语音学教程, 北京大学出版社, 1992, 119—130.
- [2] 刘源等, 现代汉语常用词频词典, 宇航出版社, 1990.
- [3] 马大猷, 语言信息和语言通信, 知识出版社, 1987, 82—83.
- [4] 林焘, 王理嘉, 语音学教程, 北京大学出版社, 1992, 106.
- [5] 薛凤生, 北京音系解析, 北京语言学院出版社, 1986, 31.