

# 声母统计分析的软件实现

郑林啸 苏彦华

南京大学中文系, 江苏 南京 210093 zhenglinxiao@263.net;

中国人民解放军 91404 部队 310 所, 河北 秦皇岛 066000 su\_y\_h@263.net

**摘要:**本文详细阐述了一种现代语言学中进行声母统计分析的方法——几遇数统计法, 并介绍了利用这种方法进行计算分析的软件及其具体实现, 最后给出了利用这个软件对《篆隶万象名义》中部分声母进行计算的结果, 并做出了分析结论。

**关键词:**声母; 几遇数统计法; 统计软件

## Statistic And Analyze Initial By Designing A Software

Zheng Linxiao Su Yanhua

Chinese Department of Nanjing University, Nanjing, Jiangsu, 210093 zhenglinxiao@263.net;

No.310 Graduate School, No.91404 PLA Unit, Qinhuangdao Hebei 066000 su\_y\_h@263.net

**ABSTRACT:** A kind of method about statistics and analyzing initial— value of meet statistics is discussed in this paper, and how to design software using this method is introduced. At the end, a few of initials in *Zhuanli Wanxiang Mingyi* are calculated and analyzed by using the designed software.

**Keyword:** Initial, value of meet statistics, statistic software

### 1. 引言

现代语言学研究比较注重信度和效度, 信度主要是指研究的稳定性和可重复性, 效度主要是指研究的解释性和可推广性, 缺乏这些特性的研究是不可信和无效的。而这些特性的检验, 是以统计学为基础的。现代音韵学是现代语言学的一项重要内容, 统计法利用精确的数字使研究资料量化, 又有音韵学的各种方法作为辅助, 从而可以使音韵研究更加有效和可信。因而用统计法研究音韵学既是音韵学创新性的要求, 也是音韵学发展的必然结果。

由于统计法涉及大量的数学知识, 并且要进行复杂而繁琐的机械计算, 工作量比较大, 为了完成题目为《〈篆隶万象名义〉唇音、舌音、半齿音研究》的硕士毕业论文, 笔者从开始收集资料到最后计算出数据用了将近九个月, 尤其感到困难的是计算列表和卡片检索, 由于卡片比较多, 在分类时难免出错, 或是在点数时点错, 同时几遇数的计算又比较繁琐, 若数出的数据有错, 则计算的几遇数也就失去了意义, 还要重新计算。为了保证数据的准确, 笔者曾将卡片用多种方法点了多次, 又将《名义》每一页中属于这十七声母的字分别列表统计, 进行验证, 然后才开始计算。如今, 利用计算机软件可以很快完成这种繁琐、单调而又易出错的工作。

为了更好地应用统计法进行音韵学的研究，笔者利用 C++ 语言开发了一个专门用来进行声母统计的计算软件，软件的主要功能如下：

- 1、建立数据库，提供输入资料的界面，储存搜集的资料
- 2、将数据资料进行报表打印输出，可以进行三种方式的报表：卡片报表、分组报表、表格报表。
- 3、提供功能强大的数据库编辑、浏览、检索功能。
- 4、根据用户的需要，进行声母的统计分析，提供总数计算、对折计算、几遇数计算、比值计算等功能。
- 5、将计算的结果存储在文件中，随时可以打开、编辑、打印输出。

## 2. 几遇数统计法

软件中主要利用了几遇数统计法来进行资料的计算分析。

几遇数统计法是在音韵学研究中处理大规模数据时，运用几遇相逢数（简称几遇数）公式（即陆志韦先生在《古音说略》中提出的公式），计算出一个客观的标准，再与实际相逢数进行比较：凡是实际相逢数与几遇数的比值大于 1，则考虑当分；小于 1，则考虑当合，并用音韵学中的语音理论进行解释。这种方法的关键在于几遇数公式：

$$\frac{AB}{N(N-1)} \times \frac{N}{2} = \frac{AB}{N-1},$$

其中 A、B 代表两个声母（或韵部、声调）的统计次数，N 代表所有声母（或韵部、声调）的统计次数。例如，在研究《篆隶万象名义》一书的声母时，帮母与非母的几遇数为 26.6，其计算为：A（帮母）=745，B（非母）=310，N=8688， $N-1=8687$ ，利用公式求得几遇数为： $745 \times 310 \div 8687 \approx 26.6$ 。该公式的原理为：若从 A 声母和 B 声母中各取一个字共有  $C_A^1 \cdot C_B^1 = AB$  种情况，从 N 个字中任取两个字共有  $C_N^2 = \frac{N!}{2!(N-2)!} = \frac{N(N-1)}{2}$  种情况，

则从 A 声母中和 B 声母中各取一个字的概率为： $P = \frac{C_A^1 \cdot C_B^1}{C_N^2} = \frac{AB}{\frac{N(N-1)}{2}}$ 。现在我们求的

是出现的次数而不是概率，故而要用任意两个字出现的总次数乘以这个概率。任意两个字同时出现的总次数为  $N/2$ ，则可以得到以下的几遇数公式：

$$J = \frac{AB}{N(N-1)} \times \frac{N}{2} = \frac{AB}{N-1}.$$

然后用实际相逢数与几遇数 J 的比值，即

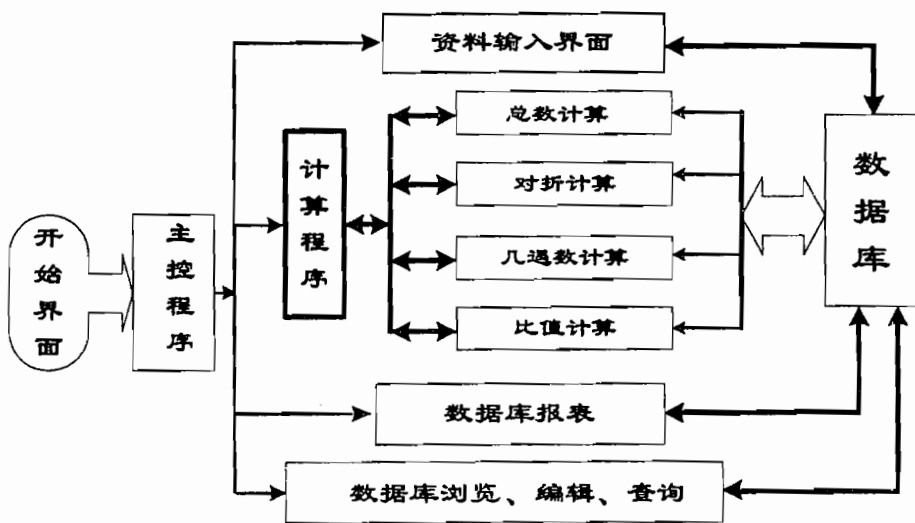
$$I = \frac{S}{J} = \frac{S}{\frac{AB}{N-1}} = \frac{S(N-1)}{AB},$$

得到的值 I 如果小于 1, 则可以肯定 A、B 两声母是两个独立的声母; 如果 I 大于或者等于 1, 则可以认为 A、B 两声母已经相合。

此方法主要适用于通过考查大量的反切材料, 来寻求音韵演变的大势。若是研究诗文用韵, 则只可用于划分韵辙, 对于各辙内部的差异则无能为力, 即不能用于分韵。因为押韵只求韵母相近, 即使相押多次也不一定韵母相同。

### 3. 软件的总体设计

软件主要由四大模块组成: 数据输入模块、数据库操作模块、数据库报表模块、统计计算模块, 组成框图如下:



数据输入模块主要解决语音资料的输入; 数据库操作模块主要解决用户对数据库中资料的编辑修改、浏览、检索等功能; 数据库报表模块主要解决数据库中资料的输出问题, 如卡片报表是将数据资料用卡片的形式打印出来; 统计计算模块是本软件的核心, 主要实现对数据库中资料的统计, 进行声母相逢几遇数及比值的计算, 存储计算结果, 方便用户对结果进行分析, 得出统计结论。

整个数据库系统由四大部分组成: 数据管理、数据交换、数据源和数据源。其中数据源即音韵数据库。

#### (1) 数据源 (音韵数据库) 的具体结构:

我们知道, 由于《切韵》原本早已失传, 现存最早且最全的《切韵》系韵书就是宋濂跋本《王仁昫刊谬补缺切韵》(又称《王三》), 据此书为起点研究此前的语音情况以及以后的语音发展, 是一种重要的研究方法。我们这个软件就是据此原则设计的: 以《王三》的反切作为中古音的标准, 据以标出中古音韵地位。“记录号”是输入材料时自动添加的, 以便于检索。

“字头”指要考查的字，“反切”是该字在要考查的书中的反切，“上声”“上韵”“上调”“上摄”“上等”“上呼”是该反切的反切上字在中古的音韵地位，“下声”“下韵”“下调”“下摄”“下等”“下呼”是该反切的下字在中古的音韵地位。“《王三》反切”是被考查的字在《王三》中的反切，下面的“声”“韵”“调”“摄”“等”“呼”则是该字在中古的音韵地位。最后“声母比较”是将“上声”与“声”进行比较。其实，若是考查声母，似乎仅输入“字头”“反切”“上声”“中古声”就够了，可是语音是一个完整的系统，在考查音变时，只考虑某一个因素是不行的，因此，虽然这个软件是考查声母的，但也必须研究其反切上下字的中古音韵地位。

音韵数据库的具体结构如下：

字段名	数据类型	长度(字节)	字段名	数据类型	长度(字节)
记录号	Numeric	8	下摄	Character	2
字头	Character	4	下等	Character	2
反切	Character	6	下呼	Character	2
上声	Character	2	《王三》反切	Character	6
上韵	Character	2	声	Character	2
上调	Character	2	韵	Character	2
上摄	Character	2	调	Character	2
上等	Character	2	摄	Character	2
上呼	Character	2	等	Character	2
下声	Character	2	呼	Character	2
下韵	Character	2	声母比较	Character	6
下调	Character	2			

## (2) 数据库管理系统

由于数据库在整个软件系统中占有重要的地位，因此做好数据的维护管理就显得非常有必要。在本软件中，笔者对数据库管理系统进行了深入的开发，基本上建成了一个功能比较齐全、界面简单易用、性能安全可靠的数据管理系统。所有的基本维护，如数据浏览、数据添加、数据查找、数据删除、结构修改、文件备份等等均可在本系统中完成，而不必外挂专门的数据库开发软件。

利用查找记录菜单，我们可以进行音韵数据库中所有字段以及各种逻辑关系组合的检索及筛选，便于我们从语音的各个角度进行分析。例如，对于查找满足“上等为三等，下等为一等”这类条件的所有记录时，就可以在软件中很容易地实现。

## 4. 统计计算程序

计算程序的界面设计如下图所示：

	幫	滂	並	明	非	敷	奉	微	端	透	定	泥	知	徹	澄	合計	總數
幫	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5
滂	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
並	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	3
明	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
非	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	2
敷	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
奉	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
微	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
端	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	2
透	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	2
定	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	2
泥	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
知	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	2	2
徹	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	2	2
澄	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
計	3	1	1	0	1	0	0	0	1	1	1	0	1	1	0	11	22

在右侧的编辑框内写入要统计的声母，然后按“增加声母”按钮，左侧窗体中则向右向下各增加一行表格，利用下面的四个计算按钮即可进行相应的统计计算，计算的对象为音韵数据库中所有已输入的记录。上图所示为库中全部记录只有 39 条时进行总数计算的结果。计算结束后，计算结果自动存入文件中，可随时调出查看打印。

## 5. 计算实例及结论

为了对软件进行验证，笔者将《篆隶万象名义》中的部分声母材料利用这一软件输入电脑，分别进行了总数计算、对折计算和几遇数计算，最后得到的计算结果如下：

### 1、总数计算结果：

	幫	滂	並	明	非	敷	奉	微	端	透	定	泥	知	徹	澄	娘	日	合計
幫	311	7	28	0	30	5	2	0	0	0	0	0	0	0	0	0	0	383
滂	15	151	14	1	3	25	5	0	0	0	0	0	0	0	0	0	0	214
並	28	5	323	0	2	5	30	0	0	0	0	0	0	0	0	0	0	393
明	0	0	0	313	0	0	0	177	0	0	0	0	0	0	0	0	0	490
非	7	1	0	0	122	8	2	0	0	0	0	0	0	0	0	0	0	140
敷	0	11	1	0	8	83	8	0	0	0	0	0	0	0	0	0	0	111
奉	1	3	9	0	5	3	181	1	0	0	0	0	0	0	0	0	0	203
微	0	0	0	28	0	0	0	89	0	0	0	0	0	0	0	0	0	117
端	0	0	0	0	0	0	0	0	317	3	9	0	5	1	0	0	0	335
透	0	0	0	0	0	0	0	0	4	186	21	1	0	64	4	0	0	280
定	0	0	1	0	0	0	0	0	10	14	528	1	1	6	11	0	0	572
泥	0	0	0	0	0	0	0	0	1	0	131	0	0	1	12	14	0	159
知	0	0	0	0	0	0	0	0	24	2	3	0	171	1	8	0	1	210
徹	0	0	0	0	0	0	0	0	2	8	4	0	1	132	9	0	0	156

澄	0	0	0	0	0	0	0	0	1	1	23	0	3	2	217	0	1	248
娘	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	86	8	101
日	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	12	215	232
合计	362	178	376	342	170	129	228	267	358	215	588	145	181	206	250	110	239	8688

## 2、几遇数计算结果:

帮	滂	並	明	非	敷	奉	微	端	透	定	泥	知	彻	澄	娘	日	合计	
帮	63.9	33.6	65.9	0	26.6	37	0	0	0	0	0	0	0	0	0	0	745	
	滂	17.7	34.7	37.5	14	10.8	19.4	0	0	0	0	0	0	0	0	0	392	
		並	68	0	27.4	21.2	38.1	0	0	0	102.1	0	0	0	0	0	769	
			明	79.7	0	0	0	36.8	0	0	0	0	0	0	0	0	832	
				非	11.1	8.6	15.4	0	0	0	0	0	0	0	0	0	310	
					敷	6.6	11.9	0	0	0	0	0	0	0	0	0	240	
						奉	21.4	19.1	0	0	0	0	0	0	0	0	431	
							微	17	0	0	0	0	0	0	0	0	384	
								端	55.3	39.5	92.5	0	31.2	28.9	39.7	0	693	
									透	28.2	66.1	17.3	22.3	20.6	28.4	0	495	
										定	154.9	40.6	52.2	48.3	66.5	0	1160	
											泥	10.6	0	0	17.4	7.3	304	
												知	17.6	16.3	22.4	0	391	
													彻	15.1	20.8	0	362	
														澄	28.5	0	498	
															娘	5.1	11.4	211
																日	25.5	471
																	合计	8688

这与以前手工计算的结果一致，这样很容易就可以得出结论：《篆隶万象名义》音系中唇音有：帮、滂、並、明（含非、敷、奉、微）；舌音有：端、透（含彻）、定、知、澄、泥（含娘、日）。至于其中详细的论述，由于已经发表在《中国音韵学研究会第十一届学术讨论会、汉语音韵学第六届国际学术研讨会论文集》中，此处不再赘述。

从以上的分析可以看出，在没有这个软件时，我们要用统计法研究音韵问题，不仅要甄别材料、摘抄卡片，还必须进行卡片分类和各种计算。利用本软件不仅使人从繁琐的机械计算中解脱出来，从而可以节省大量的时间，专注于结果的分析，而且可以根据需要对材料进行各种分类，使我们从各个角度进行分析，这无疑会加快音韵研究的进程。

这个软件是我们计划开发的音韵学系列软件中的第一个，只研究声母，在今后的工作中，我们将一方面完善这一软件，另一方面继续这个系列软件的开发。毕竟利用电脑将数学方法应用于音韵学和其它语言学科中，是语言学自身发展的必由之路。

## 参考文献

- [1] Charlie Calver, et al. 《C++Builder 应用开发大全》，清华大学出版社，1999年6月。
- [2] 陆志韦：“古音说略”，《陆志韦语言学著作集》（一），中华书局，1985年5月。
- [3] 郑林啸：“《篆隶万象名义》唇音、舌音及半齿音研究”，《中国音韵学研究会第十一届学术讨论会、汉语音韵学第六届国际学术研讨会论文集》，香港文化教育出版社有限公司，2000年8月。