

基于藏缅语的音位结构和负担量计算方法研究*

孔江平

1. 引言

随着人们利用基因研究人类演化的进展，已有比较充分的证据说明人类起源于非洲，如果这个结论正确，目前世界上说不同语言的民族就具有同一个祖先，从这个角度看，人类语言的演化经历了相同的时间和同样的演化路线，因此我们可以提出一个问题，语言的本质和演化的本质是什么？

从古人类学的研究成果看，相对于人类生理的进化，语言形成和演化的时间要短得多。虽然从目前古人类学的研究成果看语言产生的时间还很难确定，但现代言语科学技术的进步，利用古人类化石经过声道复原，最终合成出语音已经成为可能。因而，怎样从语言学的角度利用现代活的语言来研究语言进化的基本性质则是语言学家面对的重要课题。

郑锦全先生的“词涯八千”（郑锦全，1999, 2006）很巧妙地证明了人类掌握一种语言的基本能力。如果排除掉五千年文字的影响，现在比较封闭的社会中，语音的音位通常有几十个，基本语素在八百到到一千左右，常用词汇在三千左右，基本句法结构在二百左右。然而人类的发音器官能发出一两千个用于语言的音素，但人类只选了几十个能区别语义的音位，这说明了人类大脑目前处理语言音位的能力和水平。因此怎样从音位数量、音位结构、音位负担量研究语言的本质是本文的出发点。

过去对音位负担量的研究主要是通过大文本的计算，由于世界上大多数语言没有文字，因此，研究只能在个别有文字的语言中进行，这无疑极大地限制了这一领域的发展。根据多年的研究，我们发展了一种计算词汇层面音位负担量的方法，从而可以对每一种语言进行信息量和结构的计算。本文以藏缅语为例，介绍基于小词汇量的音位负担量、信息量和结构的基本计算方法，同时从音位负担量角度讨论的藏缅语语音结构、音位负担量及其演变的本质。

2. 音位负担量研究

在语音息量的研究方面，该研究可以追溯到早期的布拉格学派时期，当时主要注重于音位学的二元对立。50年代的研究主要有霍凯特(Hockett, 1955, 1951)的研究和格林博格(Greenberg, 1959)的研究。霍凯特认为：功能负担的重要性在于它对描写音韵系统有重要的价值，从而使我们可以有一个尺度来认识语言信息、语言冗余度和言语识别。格林博格认为：功能负担以通用的方式反映了一组音位或一组对立特征各成员之间的对有区别意义信号的贡献。60年代的研究，论文主要介绍了赫厄希斯瓦尔德(Hoenigswald, 1960)关于功能负担和音变的研究，他认为：功能负担和语言的音变有关，并提出了一个假说，即，在一种语言里，如果一种对立用的很少，它的消失对系统造成的危害要小于功能负担大的对立。京·罗伯特(King R. D., 1965)将音变和功能负担一同进行研究，并着重研究了音位功能和语音音变的关系，发现在日耳曼语中，功能负担和历史音变的关系不大。

在60年代，王士元教授有两篇重要的论文(王士元, 1960, 1967)。第一篇针对美国英语辅音出现频率的统计差异进行了研究，研究结果表明美国英语辅音的频率受文献风格、方言差异和样本数量的影响不大。其差异主要是来自于不同的统计词表(电子字典)和版本。第二排篇关于“音位功能信息量”研究的经典文章，论文首先讨论了音位功能负担的概念，在前人研究的基础上，王士元先生首次实现了功能负担的计算和指出了计量功能负担的困难，并给出了解决这些困难的方法。首先他讨论了音位系统中常见的三种分布、霍凯特与格林博格的测量方法以及这些方法和香农(Shannon, 1948, 1951)的通信理论及各种语言学概念的关系。其次在这些背景知识的基础上，王士元教授讨论了功能负担计量必须满足的五个条件。最后他系统地发展了四种计量功能负担的方法。另外，王士元先

*本研究得到了国家社会科学重大项目“中国有声语言及口传文化保护与传承的数字化方法研究和基础理论研究”的支持，批准号：10&ZD125。

生还指出：“音变严重受到其他许多因素的影响，如音位之间语音的相似度和语言的接触等，但正如许多历史语言学家相信的那样，如果功能负担在音变中确实起作用的话，那么用量化的解释至少可以从一个方面阐明音变这一难题”。关于音位系统的分布，王士元教授指出：“对任意一个语音序列，有三种相互关联的分布，即相似性分布、交叉性分布和互补性分布。当语音序列中每一个音位的排列都共有一组相同的环境时，它就处于相似性分布中；当音位的排列共有一些相同的环境，而不是所有的环境时，该序列处于交叉性分布中；当音位排列没有任何共有环境时，该序列处于互补性分布中”。王士元教授的研究为后来功能负担的研究建立了一个理论上的基本框架。实际上现代语音识别技术中常用的双音子和三音子的概念就起源于音位功能负担量的研究。

3. 研究方法

根据以上两节的讨论，我们首先确定在本研究中只用单音节语素，主要是单音节词。计算采用声、韵和调为音位单位。两个单音节词之间信息量设定为一个信息量单位。声、韵和调在区别两个词时各自的负担为三分之一。根据这些定义，下面介绍一下基本语素音位负担量的计算方法。

我们以彝语喜德话为例，选了6个单音节词和其相关数据，见表一。表一第一列是序号，第二列是汉译，第三列是国际音标，第四列是声母的国际音标，第五列是韵母的国际音标，第六列是声调的调值。从表一可以看出，声调有3个，声母有5个，韵母有3个。

表一、喜德彝语例词表

序号	汉义	彝(喜德)	声母	韵母	声调
1	天	mu33	m	u	33
2	山	bo33	b	o	33
3	肝	si21	s	i	21
4	胃	hi55	h	i	55
5	汗	ku21	k	u	21
6	士兵	mo55	m	o	55

在计算时取此表中的第一个词，将其声韵母分别和其他所有词对比，如果是最小对立对就得一分，

具体方法是：如果是声母对立就给这个词的声母加一分，如果是韵母对立就给韵母加一分，如果是声调对立就给声调加一分，为了方便计算，一分的数值是6，这样可以避免小数。表二是只6个词声韵调的分数，本文将这种对立称为单项对立。从表二可以看出，在单项对立上，声母的得分较高，韵母和声调较低。

表二、单项对立表

声母对立	韵母对立	声调对立
192	90	12
408	30	24
246	30	12
246	18	42
120	24	24
216	18	36

众所周知，在语言的音位系统中冗余度是普遍存在的，在词汇这一层面往往体现为二项对立和三项对立。二项对立本文定义为两个词之间，对立是由声母、韵母和声调中的两项来完成，如两个词是靠声韵来完成对立就分别给声母和韵母各加二分之一，其数值为3。如果是由声调两项完成对立，就分别给声母和声调各加二分之一，其数值为3。如果是由韵调两项完成对立，就分别给韵母和声调各加二分之一，其数值为3。见表三。从表三可以看出，声韵对立的数值最大，声调对立的数值比较小，韵调对立的数值最小。

表三、声韵调二项对立表

声韵(声)	声韵(韵)	声调(声)	声调(调)	韵调(韵)	韵调(调)
1161	1161	171	171	15	15
1077	1077	195	195	33	33
297	297	330	330	54	54
459	459	300	300	42	42
357	357	201	201	66	66
489	489	297	297	45	45

第三种情况是三项对立，即两个词之间声韵调都不同，对比词的声韵调各加三分之一，其数值是2。因此在三项对立中，声韵调的得分完全相同，见表四。从表四可以看出，6个词条声母、韵母和声调的得分数值虽然有差别，但每一词条的数值是相同的。

表四、声韵调三项对立表

声韵调（声）	声韵调（韵）	声韵调（调）
594	594	594
562	562	562
1044	1044	1044
948	948	948
1118	1118	1118
950	950	950

从表三可以看出，而相对立的情况比较复杂，二单项对立和三相对立的情况比较简单，为了更加清楚的查看结果，我们而相对立的六中情况合为声韵调三种情况，见表五。从表五可以看出，表中只有声、韵和调三列数值。从具体的数值来看，声母的得分数值最大，韵母次之，二声调的得分数值最小。因此我们可以得知，在两项对立中，声母的音位负担是最大的，韵母的音位负担次之，而声调的音位负担最小。

表五、声韵调二项对立综合表

二声母 总值	二韵母 总值	二声调 总值
1332	1176	186
1272	1110	228
627	351	384
759	501	342
558	423	267
786	534	342

将以上的各种情况简单总结一下可以看出，以声、韵和调为音位单位的对立在单音节词层面，总共有 7 中情况，分别是：1) 声母对立；2) 韵母对立；3) 声调对立；4) 声韵对立；5) 声调对立；6) 韵调对立；7) 声韵调对立。从彝语者六个例词的数据来看，三项对立的数值最大，其次是两项对立，最小是单项对立。从本文使用的藏缅语所有数据来看，情况也是如此。从这一分析的结果可以看出，音位学中的对立原则在汉藏语以单音节和声韵调为基本语音和音位单位的语言中，实际上是以三相对立和而相对立为主，最小对立对气的作用很小，并不是音位对立的主体。

表六、声韵调音位负担量总数值表

声母总值	韵母总值	声调总值
2118	1860	792
2242	1702	814
1917	1425	1440
1953	1467	1332
1796	1565	1409
1952	1502	1328

为了方便分析，表六给出了声母、韵母和声调音位负担量总数值，声母为 11978、韵母为 9521 和声调为 7115。从这些数值可以看出彝语声韵调的音位负担量是由差别的，而且声母最大，声调最小。这一结果使得我们可以来研究整个藏缅语声韵调的音位负担量，因而为定量评价藏缅语声韵调各自的功能和类型提供证据。也开辟了定量分析不同语言音位系统和语言演化程度的方法。

4. 藏缅语的音位负担量

根据以上对音位负担量的定义和计算方法，我们计算了部分藏缅语的音位负担量，藏缅语的数据库是根据《藏缅语族语言词汇》（黄布凡，1992）建立，在本项研究中只用了其中的单音节词，由于每种语言的单音节数量不同，本文的计算结果除了实际的数值外，还计算了声韵调比值，这样就可以对所有语言进行对比研究，见表七。表中第一列是序号，序号是根据声调音位负担量的大小排序而成，数值小的序号小，数值大的序号大；第二列是语言或某语言的方言名称；第三列是声母音位负担量总值；第四列是韵母音位负担量总值；第五列是声调音位负担量总值；第六列是声韵调音位负担量总值；第七列是声母音位负担量比值；第八列是韵母音位负担量总值；第九列是声调音位负担量总值。其中第三列至第六列的数据要处理词汇的总数才可以使用在比较研究方面，不能简单地单独使用。每种语言具体声韵调音位负担量的研究结果将另文要论文。

表七、藏缅语声韵调音位负担量及声韵调比值表

序号	方言点	声母总值	韵母总值	声调总值	声韵调总值	声母比例	韵母比例	声调比例
1	嘉戎	122112	112248	0	234360	0.52	0.48	0.00
2	藏(夏河)	2157414	2043690	0	4201104	0.51	0.49	0.00
3	藏(书面语)	2576946	2491050	0	5067996	0.51	0.49	0.00
4	羌	1686158	1668194	3020	3357372	0.50	0.50	0.00
5	藏(阿力克)	1059990	1006410	2376	2068776	0.51	0.49	0.00
6	博嘎尔珞巴	848218	904144	2278	1754640	0.48	0.52	0.00
7	墨脱门巴	2565864	2576694	1024866	6167424	0.42	0.42	0.17
8	独龙	1150068	1160004	462984	2773056	0.41	0.42	0.17
9	义都珞巴	101120	89210	42794	233124	0.43	0.38	0.18
10	却域	2460136	2368204	1208860	6037200	0.41	0.39	0.20
11	阿依怒	87334	84742	45316	217392	0.40	0.39	0.21
12	彝(南华)	1630160	1564388	847544	4042092	0.40	0.39	0.21
13	彝(喜德)	1618072	1371046	827170	3816288	0.42	0.36	0.22
14	克伦	474610	465448	263122	1203180	0.39	0.39	0.22
15	贵琼	371858	352190	202964	927012	0.40	0.38	0.22
16	仙岛	1994042	2039090	1133312	5166444	0.39	0.39	0.22
17	阿昌	2517026	2583032	1440146	6540204	0.38	0.39	0.22
18	格曼登	435402	440184	250686	1126272	0.39	0.39	0.22
19	浪速	2297168	2384534	1353014	6034716	0.38	0.40	0.22
20	波拉	2088520	2173318	1231678	5493516	0.38	0.40	0.22
21	错那门巴	665470	676474	396448	1738392	0.38	0.39	0.23
22	载瓦	2242938	2331534	1353168	5927640	0.38	0.39	0.23
23	达让登	198802	193300	116794	508896	0.39	0.38	0.23
24	哈尼(绿春)	1058744	998612	617960	2675316	0.40	0.37	0.23
25	景颇	1000370	1043324	618854	2662548	0.38	0.39	0.23
26	藏(巴塘)	1490360	1434542	895682	3820584	0.39	0.38	0.23
27	彝(巍山)	1713606	1579020	1009098	4301724	0.40	0.37	0.23
28	史兴	495886	466402	296812	1259100	0.39	0.37	0.24
29	勒期	1677070	1772518	1072228	4521816	0.37	0.39	0.24
30	哈尼(墨江)	1021232	975956	622436	2619624	0.39	0.37	0.24
31	傈僳	1479748	1387264	951208	3818220	0.39	0.36	0.25
32	白	1185304	1231198	833434	3249936	0.36	0.38	0.26
33	纳木兹	4148886	4108596	2855610	11113092	0.37	0.37	0.26
34	缅(书面语)	2345622	2272662	1599816	6218100	0.38	0.37	0.26
35	藏(拉萨)	1428336	1421790	1017570	3867696	0.37	0.37	0.26
36	彝(武定)	2809544	2553182	1921742	7284668	0.39	0.35	0.26
37	土家	664624	620026	470794	1755444	0.38	0.35	0.27
38	扎坝	4729342	4709830	3527248	12966420	0.36	0.36	0.27
39	吕苏	2481422	2414234	1885064	6780720	0.37	0.36	0.28
40	彝(撒尼)	1695308	1447904	1219292	4362504	0.39	0.33	0.28
41	基诺	767482	694054	571300	2032836	0.38	0.34	0.28

42	缅(仰光)	4555552	4465258	3541774	12562584	0.36	0.36	0.28
43	普米(兰坪)	6272854	6186160	4963510	17422524	0.36	0.36	0.28
44	纳西	6410460	6170514	5016138	17597112	0.36	0.35	0.29
45	怒苏怒	6144024	5974662	4955754	17074440	0.36	0.35	0.29
46	普米(九龙)	6123764	6051278	5104298	17279340	0.35	0.35	0.30
47	拉祜	1420182	1253196	1202202	3875580	0.37	0.32	0.31
48	嘎卓	886158	804126	765120	2455404	0.36	0.33	0.31
49	木雅	3599292	3606120	4308216	11513628	0.31	0.31	0.37

数据表是按照声调的音位负担量进行的排序，从数据可以看出，前六个语言或方言是：1) 嘉戎；2) 藏夏河话；3) 藏书面语；4) 羌；5) 藏阿力克话和 6) 博嘎尔洛巴。这六个语言或方言没有声调，因此声调的音位负担量为零，排在表的最前面。这六种语言或方言声母和韵母的音位负担量比较大，其和等于 1。声调音位负担量最大的六个语言是：1) 纳西；2) 怒苏怒；3) 普米九龙话；4) 拉祜；5) 嘎卓和 6) 木雅，其比值分别是 0.29、0.29、0.30、0.31、0.31 和 0.37。从这些基本数据可以看出，藏缅语声韵调的音位负担量是不同的，他们反映了各自语言的音位系统的差别和发展的不同程度，这为我们研究藏缅语音位体系、音位机构和音位负担量和语言信息之间的关系奠定了基础。本项爱那个研究的计算结果也可以按声母或韵来排序，从而进一步研究声母和韵母的特性。

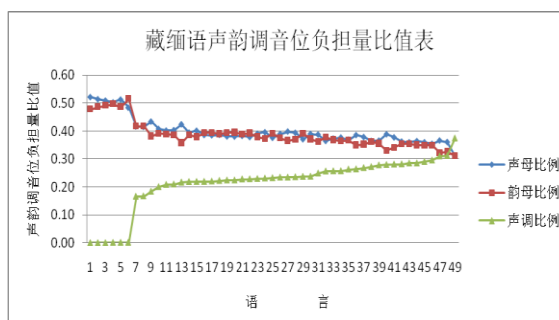
5. 研究结果

根据本文对词汇层面音位负担量的定义和对藏缅语 49 个语言和方言的计算，可以得到许多很结果和结论，限于篇幅本文在此先从宏观的角度就一些主要的结果进行讨论。有关声母、韵母和声调音位负担量的内部结构、分布以及和语言系数的关系将放在另外的文章中讨论。

首先，从图一可以看出：1) 声调从无到有，体现为声母和韵母音位负担量的下降。2) 在声调的音位负担量为 0 时，声母和韵母的负担量比有声调的声韵母负担量要高许多，两者之间有一个跳跃。3) 在有声调的语言和方言中，声调的信息量和声韵母的信息量成反比关系，也就说一种语言里，如果声调承载的信息量大，声母和韵母能承载的信息量就是小。4) 大多数语言声韵母的负担量在一个数量级上，相对来说比较大，而声调的音位负担量要比声韵母小很多。从这 49 个语言和方言上来看，

声调音位负担量在有些语言中已经很接近声母和韵母的数值，者体现了藏缅语声调音位负担量发展成都的不同。见图一。

图一、藏缅语声韵调音位负担量比值表

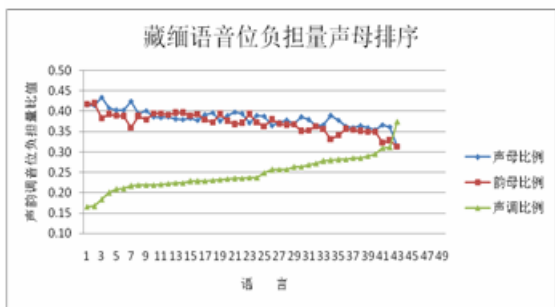


很显然，在藏缅语中，声调的音位负担量是很不同的，这反映了藏缅语声调在不同语言中发展和演化的不同阶段。最小的只有 0.17，而最大的为 0.37，这个数值大过该语言的声母和韵母。

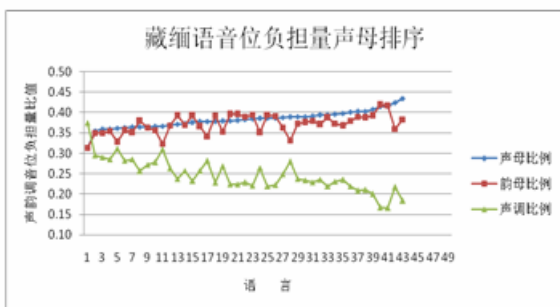
第二，宏观上声母和韵母的信息量相对于声调来说基本相等，也就是说，一种语言里，声母能承载的信息量和韵母基本相同，但和声调有很大的差别。因此从音位结构、功能和层次上可以看出，在藏缅语的发展过程中，声母和韵母在结构和功能上是一个层次，随着声调的发展和演化，声调的功能逐渐增强，而声母和韵母的功能相对减弱。

根据本文的计算方法，如果按声调排序可以发现声韵母的音位负担量成镜像分布，见图二；如果按声母排序可以发现声调和韵母的音位负担量成镜像分布，见图三；如果按韵母排序可以发现声调和声母的音位负担量呈镜像分布。但三者数值上有较大差别。

图二、藏缅语音位负担量声调排序图



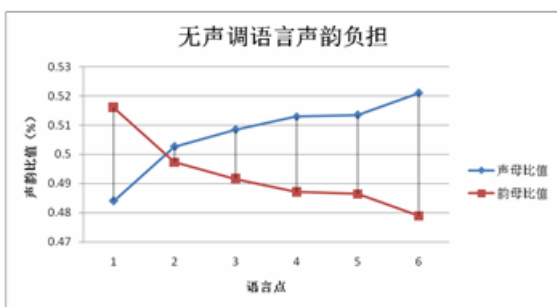
图三、藏缅语音位负担量声母排序图



图四、藏缅语音位负担量韵母排序图



图五、无声调语言声韵母音位负担量表



第三，在无声调的语言里，声母和韵母在信息量的常在上成反比关系，也就说声母承载的信息量大，韵母承载的信息量就小，否则相反，见图二和表七。这些语言呈现出镜像的分布，但总的来说，声母和韵母的数值差别不太大，基本在 0.5 左右。这一点可能和我们采用了声母和韵母做基本单位有关。可以看出前面 6 种语言是无声调的语言，其声调的音位负担量为 0，所以音位负担量有声母和

韵母承载，声韵母的负担量比有声调的声韵母要高许多，因为其总和等于一。

第四，对于声母和韵母来说，音位数量的增减不太影响声韵母总的音位负担量。以藏语为例，从古藏语到现在大多数方言，复辅音声母大量脱落，但声母总的音位负担量并没有减少，而是由剩余的声母承载，直到声调产生，整体音位负担量下降。这无疑从一个语言内部的音位负担量为研究语言的演化提供了线索。

6. 讨论

本文的研究表明，语言音位负担的语言学研究能解释许多语言信息量的结构、承载和演变的性质和规律。从 49 个藏缅语语言和方言的情况看，有声调语言 and 无声调语言之间存在有一个跳跃，声韵调音位负担量的曲线不是平滑的，这存在两种可能：1) 我们在调查语言的过程中，声调产生的过程没有调查出来。如，藏语有许多方言正处于声调的产生过程中，而我们目前采用的结构主义音位学的调查方法不能很好地或者说精确地描写声调产生的过程；2) 声调的产生确实是突变产生的。另外，在计算方法上，由于不同音位对立类型的数量级差别较大，也有可能掩盖了一些内部的规律，因此在计算上做进一步的加权算法研究是很必要的。最后希望这项研究能成为研究语言学和语言学量研究的一个很好的切合点，形成科学地研究语言演化的一个新的方向。

7. 参考文献

- [1] 黄布凡主编，1992，《藏缅语族语言词汇》，中央民族学院出版社，北京。
- [2] 郑锦全，2006，从词语八千到学海无涯，国立中山大学中文系，高雄，讲座。
- [3] Greenberg, H.H., 1959, A method of measuring functional yield as applied to tone in African languages. Georgetown University Monograph Series on Languages and Linguistics 12: 7-16.
- [4] Hockett, C.F., 1955, A Manual of Phonology, p.218, Baltimore Waverly Press.
- [5] Hockett, C.F., 1961, The quantification of functional load. Rand Report, p.2338, Santa Monica.
- [6] Hoenigswald, H.M., 1960, Language change and linguistic reconstruction, pp.79-80, University of Chicago Press, Chicago.

- [7] King, R.D., 1955, Functional load: its measure and its role in sound change, University of Wisconsin PhD dissertation. A version of this will appear in 'Language'.
- [8] Shannon, C.E., 1951, Prediction and entropy of printed English. Bell System Technical Journal, 30: 50--64.
- [9] Shannon, C.E. and Weaver, W., 1949, The mathematical theory of communication, University of Illinois Press, Urbana 1949.
- [10] Wang, W. S-Y., 1967, Phonemic Theory A (with Application to Midwestern English), Doctoral dissertation. The University of Michigan, -Stress in English. Language
- [11] Wang, W. S-Y. and Crawford, J., 1960, Frequency studies of English consonants, Language and Speech 3: 131-139.
- [12] Cheng, C. C., 1999, 词涯八千 (Active vocabulary upper limit 8000), Graduate Institute of Linguistics, National Taiwan University, Taipei, May 10. Invited lecture.