

国家教委人文社会科学研究“八·五”规划项目

电脑辅助

汉藏语词汇和语音研究

COLLECTED PAPERS ON SINO-TIBETAN LANGUAGES:

COMPUTER-ASSISTED

LEXICAL AND PHONETIC STUDIES



戴庆厦 黄布凡等著

DAI QINGXIA, HUANG BUFAN, ET AL.

电脑辅助汉藏语
词汇和语音研究

Collected Papers on Sino - Tibetan
Languages: Computer - assisted Lexical
and ~~Phonetic~~ Phonetic Studies

戴庆厦 黄布凡等著

By

Dai Qingxia, Huang Bufan, et al.

中国藏学出版社

责任编辑 安才旦

封面设计 李建雄

图书在版编目(CIP)数据

电脑辅助汉藏语词汇和语音研究/戴庆厦等著. —北京:中国藏学出版社,1996. 8 ISBN 7—80057—291—9

I. 电… I. 戴… III. ①汉藏语系—词汇—研究—计算机应用

②汉藏语系—语音—研究—计算机应用 N. H4

中国版本图书馆 CIP 数据核字(96)第 14603 号

电脑辅助汉藏语词汇和语音研究

戴庆厦 黄布凡等著

中国藏学出版社出版

新华书店北京发行所发行

中国藏学出版社微机室排版

北京大中印刷厂印刷

开本:787×1092 1/32 印张:4.625 字数:80千字

1996年8月第1版第1次印刷 印数:1000册

ISBN 7—80057—291—9/Z · 147

定价:5.80元

前 言

语言学同自然科学相结合产生的语言学新分支,是现代语言学发展的重要方面,也是现代语言学发展的必然趋势。特别是将电脑应用于语言文字的研究,使这项研究进入了精密化、程式化、系统化的新阶段,开辟了语言学研究的广阔新天地,大大地推动了语言学的发展。近期,在诸多的语言学分支中,统计语言学和实验语音学发展迅速,引人注目。

统计语言学(statistical linguistics),又称计量语言学(quantitative linguistics),是将数学统计方法应用于语言研究而产生的一门新学科。它的任务是运用统计程序对语言文字进行定量分析,从数量的分布及其相互关系中探求其特点。语言文字的定量分析使用范围很广,包括音素、语素、音节、词、词类、词组、语义、句型、篇章、字型等频率的统计。实验语音学(experimental phonetics)则是运用现代化的实验仪器分析语音特征,使语音研究具有物理实验的依据,开辟了传统的“口耳辨音”所未能触及的新领域。

在我国,统计语言学和实验语音学应用于汉语的研究,已取得了不少成绩,但应用于少数民族语言研究,历史很短,只能说是刚刚起步。不过,现有的研究成果已经证明,二者应用于少数民族语言研究,大有可为。我国是一个多民族、多语种、多文种的国家,语言文字的情况十分复杂,语言文字的研究任务十分繁重,因而,运用现代化手段研究我国

少数民族语言文字势在必行。这样做,不仅能加速民族语文研究的进程,而且其研究成果还能为丰富和发展普通语言学增添新的内容。

中央民族大学中国少数民族语言文学学院担负着少数民族语言研究的重要任务。为了提高语言研究水平,学院建立了康赛电脑语言学研究中心,逐步开发电脑和其它语音实验仪器在少数民族语言文字研究中的运用。虽然我们在仪器设备和知识结构上还存在许多不足,还有不少困难,但我们深知这个工作的重要意义。1992年,我们提出了“汉藏语统计分析和语音实验研究”课题,得到国家教委博士点社会科学基金资助。这本论文集,就是这一课题的初步研究成果。我们希望这些研究能成为今后使用现代化手段研究汉藏语系语言的良好开端。在这里,我们要感谢国家教委博士点社科基金会的鼎力支持。

本书内容主要是通过通过对藏、白、布依、苗等汉藏语的词汇和语音结构进行统计分析和语音实验,探讨词汇演变的速度和方式、文献中古词的意义、语音演变的轨迹、从语音结构的统计结果看语言的亲属关系等问题。对舌面元音共振峰的实验则是对普通语音学理论问题的探讨。此外,还介绍了近期对汉语以外的汉藏语进行语音实验研究的成果及其理论价值;介绍了统计分析语言的几种电脑软件和有关的语音实验仪器等。

本书编辑工作由戴庆厦、黄布凡负责。

感谢吴和得先生将书名和目录译成英文。

作者

1994.11.11

目 录

前言

藏语词汇演变的速率和方式

- 对敦煌藏文写卷抽样的电脑统计分析…黄布凡 1
- 从电脑统计结果看白语语音特点……戴庆厦 赵富芬 33
- 布依族摩经中非口语词的电脑统计分析……周国炎 63
- 黔东南苗语声母和声调关系的统计分析……李炳泽 86
- 从共振峰看元音音区……………罗安源 97
- 湘西苗语一组声母语音实验分析……………关辛秋 107
- 汉藏语语音实验研究及其理论贡献……………李锦芳 114

附录一:康赛电脑语言学研究中心的几种软件简介

……………周国炎 127

附录二:康赛电脑语言学研究中心简介……李锦芳 136

Contents

Preface

1. Huang Bufan Manner and pace of lexical change in Tibetan: A computer - assisted statistical study of a sample of Dunhuang manuscripts
2. Dai Qingxia & Zhao Fufen Characteristics of the sound system of the Bai language: A statistical approach
3. Zhou Guoyan A computer - assisted statistical analysis of words in Buyi scriptures not used in oral speech
4. Li Bingze A statistical analysis of initial consonants and tones in the Miao dialect spoken in eastern Guizhou
5. Luo Anyuan The relationship between the first and second formant frequencies and the partition of the traditional vowel diagram

6. Guan Xinqiu An acoustic phonetic study of a group of initial consonants in the Miao dialect spoken in western Hunan
7. Li Jinfang An overview of recent experimental phonetic studies of Sino - Tibetan languages and their contribution to linguistic theory

Appendices:

- I. Major software used by the CUNSIL Computer Center for Linguistic Research: A description (Zhou Guoyan)
- II. The CUNSIL Computer Center for Linguistic Research (Li Jinfang)

藏语词汇演变的速率和方式

——对敦煌藏文写卷抽样的电脑统计分析

黄布凡

【提要】 本文通过对古今藏语词汇的抽样比较分析，探寻藏语词汇演变的速率和方式。古藏语词汇取材于敦煌藏文写卷中关于吐蕃赞普传记的两个长卷，其写作年代据推断为9世纪初；今藏语词汇取材于拉萨话。方法是将写卷切分成词后，用电脑 Fling 软件统计词频，列出全部词表及词的上下文，再逐一与拉萨话比较，求出两卷实词的保留率与变化率，并按斯瓦迪士(M. Swadesh)百词表求出基本词汇词根的保留率与变化率，同时求出一般词汇与基本词汇词根语素千年演变速率之间的比例。演变速率的统计结果对语言年代学基本词根语素保留率提供了新情况，对其常数的确定提出了新问题。分析古词的演变方式有消失、转义、改造、沿用等4种，并分析研究了古词消失的外部 and 内部原因。

对于藏语自古至今的演变，语音方面多有研究，语法方面也时有文章涉及，而词汇方面却少有专门的研究文章。词汇的演变也是藏语史研究的一个重要方面。本文内容是通过对话古今词汇的抽样比较分析，探讨藏语词汇演变的速率和方式。

一、材料和方法

词汇是开放性体系，很难精确统计某一语言在某一时期的词汇总量。古代语言的词汇材料更是难以全面统计和搜集。因此，作古今词汇比较，只能划定一定范围，抽取一定数量的词语，作抽样定量分析。这样做，虽不能作到“以偏概全”，但可达到窥斑见豹的目的。

斯瓦迪士(M. Swadesh)选取 200 个基本词根语素，编制了一个词表，后来又修正为一百词表，用以观察一种语言在一定的时间长度(例如一千年)内的变化速度，这对于观察语言变化的速率或保留率的系数，并以此来测算语言发展的年代，具有较好的效果。在探寻藏语词汇变化速率时，我们也要使用这种方法。但探讨藏语词汇变化方式时，一百或二百个词就显得太少了。因此，需要扩大范围，在取材上需要大量能涵盖基本词汇的使用频率较高的一般词汇。这些词应该是能够确定使用年代的词。我们也同时用这份材料求出一般词汇的变化速率。

藏语各类词书为数不少，其中，收录古词最多的是张怡荪主编的《藏汉大辞典》(北京民族出版社，1985 年版)，达 53000 余条；收录今词较多的主要有于道泉主编的《藏汉对照拉萨口语词典》(北京民族出版社，1983 年版)和华侃、龙博甲合编的《安多藏语口语词典》(甘肃民族出版社，1993 年版)等。但这类词书都不能作为此项比较研究的取材样本，例如《藏汉大辞典》虽然收词偏古，但也兼收今词，且古词大多未注词的出处，无法断定使用年代。又如《拉萨口语

词典》，虽然收词大部分是拉萨口语，但也收进一部分书面语的词，如 la la“有些”、du ma“若干、许多”、kun“都、一切”、rjod“说、解说”、bsgo“命令、下谕”、rna“耳朵”等；还有一些在古代可以单独成词，而在现代口语里只作构词语素的，也作为单独的词条，如 sna“鼻子”、brang“胸”、rkyal“游、渡”、brtan“坚固”等（这些词在拉萨话里分别说成 sna khug、brang khog、rkyal rgyag、brtan po 等），由此可见，它不是一部纯粹的拉萨口语词典。

本文用于比较研究的古词材料，取自敦煌藏文 P. T. 1286 号和 P. T. 1287 号两份长卷。^{〔1〕}选择这两份写卷作为古词取材样本，主要基于下列原因：

1、这两份长卷的内容是吐蕃时期著名赞普（藏王）和论布（大臣）的传记以及历代赞普和论布世系，文学性、故事性较强，常用的生活词语多，能基本涵盖基本词汇。

2、从写卷内容和字体特征可推断出相对准确的写作年代，即语言年代学所谓的可能性系年。赞普传记记载了第一代赞普聂赤赞普至赤松德赞（公元 755—797 年在位）之间八位著名赞普的事迹。赞普世系终记于吾益董木丹赞普（即达磨赞普，公元 836—842 年在位）。由此可推断写卷书写年代的上限不会早于达磨时期，至少不会早于赤松德赞时期。从各传记的语言风格和写作体例来看，似出自同一时代甚

〔1〕 Choix de documents Tibétains Conservés à la Bibliothèque National complète par quelques manuscrits de L'India office et du BM, par A. Macdonal et Yoshiro Imaeda, Tom 1er II, Paris, 1979.（《巴黎国家图书馆藏藏文选集》）影印了这两份写卷。

至是同一人之手，不像是历代分期记载的累积。其文字特征与有明确纪年的长庆会盟碑（立于公元 823 年）碑文非常接近。值得注意的是，赞普世系记到吾益董木丹时不是一个完整的句子，文章也没有结尾，写卷有可能因人截断而残缺了；另有一种可能，即达磨上台后采取“禁佛焚经”的做法，使吐蕃社会的内部矛盾日趋激化，全盛的吐蕃王朝开始走向崩溃，吐蕃社会处于内部分裂、兵荒马乱、人心惶惶的混乱状况，因而史记官顾不得往下写史了。因此，这两份写卷的写作年代，不会晚于达磨赞普在位时期，大约书写于 9 世纪初。

3、这些写卷除个别地方残缺不清外，大部分完整、清晰。而且因为是吐蕃王室的重要正式文件（即使是抄本也不能排除这一性质）在遣词造句和使用正字（即规范字）方面都比较认真、讲究，不象有些写卷那样错漏屡见白字较多。这两份写卷可以看作是基本反映当时藏文所依据的基础方言面貌的代表作。我们不否认那时可能也有文字和口语的差异，写卷的词中就可能有更早期古词的沉积，但一般认为，藏文创制于 7 世纪松赞干布时期，从藏文创制到 9 世纪初不过只有一百多年，语言的变化不会太大；再加那时没有木刻印刷，文字记录、传播和固定词语形式的功能都远远小于有了藏文印刷以后的时期。因此，即使写卷的词中可能会保留一些脱离当时口语的古成分，其所占比例可能也是很小的。

4、这两份写卷经过半个世纪以来国内外学者的研究，大部分词语得到可信的解释（本人也研究多年，有一份尚未发表的注释和译文，本文所注词语的释义即根据上述成

果)。

因此可以说,从这两份写卷分解出来的,是词义和年代都比较准确、清晰的词语。

用以比较的今词材料,取自现代拉萨口语。《拉萨口语词典》因有前已述及的缺陷,只能作为重要参考。对其中的非口语成分,通过与拉萨人核对,予以排除或校正。藉此,向热心帮助我的朗珍女士顺致谢意。

本文采用的比较方法是,首先将两卷通篇切分成词,输入电脑,使用美国 SIL(暑期语言学院)白默翰(John Alsop)教授设计并提供的一种专门用于话语材料统计分析的电脑软件 Fling 进行词频统计,并按字母顺序列出全卷词表,然后与拉萨口语词相对应,逐个注明消失、转义、词转语素、保留等情况,最后再根据这份对照词表进行各种分析。用 Fling 软件排出的词表,不仅可做到一个也不遗漏,一个也不重复,而且还可反映各个词出现的上下文语句,有利于分析和把握词义,有利于区分同音词和多义词。通过这个课题,第一次尝到运用电脑统计分析话语材料的甜头。在此,要向热心指导和帮助我们的白默翰教授致以谢意!

二、词汇变化的速率

1. 两卷词汇分析统计结果

P. T. 1286 和 P. T. 1287 两卷共 594 行(P. T. 1287 卷中 526 - 536 行因残缺较多,影响对词义的理解,未入统计),从中共切分出不同的词 2187 个。除去专名(包括人名、地名、部族名、国名等),共有词 1646 个。虚词无实在意义,

主要起语法作用,也将它们除去(指格助词、连词、语气词、叹词等,副词计入实词内),共得实词 1590 个。其中,有形态变化的词,不同的形态按一个词计算,如 lta(现在时)、blta(未来时)、bltas(过去时)“看”。另外,某些因当时自由变读(如塞音、塞擦音的送气和不送气交替)而拼写不同的词和因正字规划尚未厘定而书写形式不同的词,也只按一个词计算,如 bcings~bching“束缚”、brgya~brgyav“百”等。文中有关的拉萨话的今词,统一用代表它的藏文正字标记,不标实际音值;无正字的则用相应的藏文拼音结构标记。〔1〕

比较统计时,以写卷中的古词为出发点,按照它们与现代拉萨话的对应情况,划分成消失、转义、词根保留、保留等 4 种类型。

消失,指在拉萨话里已消失不用或以今词替换的词,如 blon che“大论”、khyo mo(古)→skye dman(今)“妻”等。

转义,指古今词的形式相同、意义发生变化的词,如 lto“肚子、腹”(古)→“食物”(今)。

词根保留,有两种情况。一是古词的词根变为今词,如 rdo ba(古)→rdo(今)“石头”;二是古词作为主要语素保留在复合词中,如 char(古)→char pa(今)“雨”等。

保留,指古词的意义和形式都与今词相同,如 rta“马”;

〔1〕为适应 Fling 软件能处理的 Ascll 字符,即拉丁字母,本文所引写卷内的藏文词语和标记拉萨话的藏文都用拉丁字母转写。转写依藏文字母表顺序分别为:k kh g ng;c ch j ny;t th d n;p ph b m;ts tsh dz w;zh z v y;r l sh s;h a. y 带前加字 g 时前加-,作 g 的下加字时不加-。

有的古今拼写不同但意义相同、语音有渊源关系或对应关系的词也属于这一类,如:ral gyi(古)与 ral gri(今)“宝剑”、myi(古)与 mi(今)“人”等。

比较分析统计的结果是:消失的古词 791 个,占比较词的 50%;古词转义的 37 个,占比较词的 2%;词根保留的 212 个,占比较词的 13%;完全保留的 550 个,占比较词的 35%。按照斯瓦迪什百词表的分类法,前两项应算变化率,合起来是 52%,后两项是保留率,合起来是 48%。

2. 按斯瓦迪什百词表统计结果

斯瓦迪什百词表所选基本词,大部分在这两份写卷中可以找到。尚缺 28 个,从大致是同一时期的其他敦煌写卷中补充了 23 个,还缺 5 个,一时不易从浩繁的写卷中找到,便另选了几个基本词(顺序号上带圆圈者)。^[1]换用的词与原词表上的词基本上是词性相当、类别相当,如以“草”换“(树)皮”,都是名词,都是植物类,而且在藏语里“草”是较之“(树)皮”,更为基本的词。统计对象以基本词根语素为准,不算词缀,古今同源的成分用“+”号表示,不同源的用“-”号表示,词义有变化的也用“-”号表示。在对应的一组词中,古或今若有两个以上的词,以第一行的对应为准。藏语的形容词词根具有自动动词性质,如写卷中具有同一词根的形容词和动词,则两个词都列出。(下表中未标明出处的,其古词都出自上述两份写卷)

[1] 换用的词如下:22 原为“虱”,27 原为“树皮”,45 原为“瓜子”,68 原为“坐”,98 原为“圈”。

序号	古藏语	今拉萨话	汉义	同源	古词出处
1	nga bdag	nga	我	+	
2	khyed(敬) khyod	khyed rang(敬) khyod ra rang	你	+	
3	bdag chag	nga tsho	我们	-	
4	vdi vung	vdi	这	+	
5	de	de	那	+	
6	su	su	谁	+	
7	gang ji	ga re	什么	-	
8	myi	ma ⁽¹⁾	不	-	
9	kun tham shad ~ tham chad	tshang ma	全部	-	
10	mang po mang	mang po mang	多的 多	+	
11	gchig	gcig	一	+	
12	gnyis	gnyis	二	+	

(1) ma 只在 -gi ma red 或 -gi min(<ma yin) 这种格式中才有“不”的含义, -gi 是附加成分, 加在动词未来时形式后, red 或 yin 原义为“是”, 在此格式中已虚化, 如 kho(他)vgro(去) -gi ma red(不)“他不去”; ma 的另一义位是“没有”, 如 ma bltas“没看”。

13	chen po ~ched po che	chen po che	大的 大	+	
14	ring po ring	ring po ring	长的 长	+	
15	chung ngu chung ~cung	chung chung chung	小的 小	+	
16	bud myed myi smos myi mo	bud med	女人	+	P. T.1057-191 ⁽¹⁾ P. T.1057-35
17	khyo myi pho ~myi po	khyo ga	男人	+	P. T.1057-182 P. T.1057-35, -36
18	myi	mi	人	+	
19	nya	nya	鱼	+	
20	bya	bya	鸟、禽	+	
21	khyi	khyi	狗	+	
㉔	srin bu	srin vbu	虫(牙虫、蛔虫)	+	P. T.1057-19

(1) 短横后面的数码是原卷行数。