

北方人口较少民族语言词汇缺失现象探究

——以语保工程民语资源调查材料为例*

丁石庆

[提要] 基于语保工程民族语言调研专项任务语料的相关数据统计分析,发现北方人口较少民族语言存在着程度不同的词汇缺失现象,主要表现为绝对缺失与相对缺失、义类缺失与义项缺失。词汇缺失在整体上呈现复杂性、不平衡性及非线性等特征,导致词汇缺失的原因可归结为自然环境与传统生计生产方式、语言结构特点、有无文字以及居住格局和宗教信仰等的不同。

[关键词] 北方人口较少民族 词汇缺失 绝对缺失

中国北方人口较少民族主要分布于黑龙江、内蒙古、甘肃、青海、新疆等省(自治区),据全国第六次人口普查(2010)数据,按人口数量多少分别为:土族 289565 人、锡伯族 190481 人、柯尔克孜族 186708 人、达斡尔族 131992 人、撒拉族 130607 人、塔吉克族 51069 人、鄂温克族 30875 人、保安族 20074 人、俄罗斯族 15393 人、裕固族 14378 人、乌孜别克族 10569 人、鄂伦春族 8659 人、赫哲族 5354 人、塔塔尔族 3556 人。

中国北方人口较少民族使用的语言共计有 15 种,阿尔泰语系 13 种,其中,突厥语族语言 5 种,蒙古语族语言 4 种,满-通古斯语族语言 4 种;印欧语系 2 种。其中,含跨境语言 8 种。拥有传统文字的语言 3 种,有新创文字的语言 1 种。

2015 年以来,我们开展了北方人口较少民族语言资源保护专题的深度调查研究工作。先期主要涉及达斡尔、保安、鄂温克、鄂伦春、锡伯、撒拉等阿尔泰语系语言和中国境内印欧语系斯拉夫语族的俄罗斯语。调查进程中,适逢中国语言资源保护工程(简称“语保工程”)的实施,借助其东风和相关调查资金支持,我们陆续扩大到了东部裕固、土族、赫哲等蒙古语族、满-通古斯语族语言,柯尔克孜、乌孜别克等突厥语族语言及印欧语系伊朗语族的塔吉克语,基本上实现了北方人口较少民族语言全覆盖。

在对调查材料整理研究过程中,我们陆续发现一些问题,其中最引人关注的是人口较少,尤其是具有濒危迹象的民族语言或方言的“词汇缺失”现象。如何认识“词汇缺失”现象,如何科学采集语料并精准保护民族语言,是我们一直在深入思考并探索的重要论题。实际上,词汇缺失也是较为普遍存在的现象。对词汇缺失现象的研究,有助于积累语言活力、语言资

* 本文为国家社科基金重点项目“中国北方人口较少民族语言资源保护的理论与实践研究(15AYY012)”、语保工程民语调研专项任务“民族语言管理项目(YB1826B026)”的阶段性成果。曾在“《民族语文》创刊 40 周年学术研讨会”(北京 2019.10.11-13)上作过主题发言。

源类型差异的个案样本，于语保工程民族语言资源调查及语料整理也具有实践参考意义。

一 相关概念及说明

词汇缺失，是指在一种语言中原本不存在或由于某种原因出现空缺词语的总称。根据缺失的情况，可分为绝对缺失词汇和相对缺失词汇两类。根据其类别和义项，缺失情况可分为词汇义类缺失和义项缺失等。

（一）绝对缺失

绝对缺失或称自然缺失、客观缺失，指缺失的词汇系是非传统文化背景下产生的词汇。也就是一个语言原本没有的词，或不可能产生的词。如游牧民族语言缺失农耕民族的农业词汇，农耕民族语言缺失牧业词汇等。又如南方生长的很多植物、水果、蔬菜等的名称，在北方很多民族语言中都缺失等。也可能由语言结构本身及其演变特征而造成，如汉语有数量庞大且丰富的量词系统，但很多北方民族语言均绝对缺失此类词。另外，很多无文字民族语言缺失某些抽象词语的现象也较为普遍。

（二）相对缺失

相对缺失指某些固有词或原有的一些词汇，因各种原因导致流失或忘却，或称发展缺失、衰变缺失。也就是在民族语言中有据可考，原本应有的词语，但因诸多因素而没能采集到的词汇，或因语言社会功能整体衰变而导致常用词长期不用而逐渐淡出人们的记忆而缺失的词汇等。词汇的相对缺失或可和词汇流失等概念相提并论。

此外，词汇缺失还可分义类缺失和义项缺失。义类缺失主要指原本有的某一义类的缺失现象。义项缺失则主要指多义词原本有的某一义项的缺失现象。义类缺失与义项缺失更适于在一种语言内部进行微观分析。

（三）相关说明

本文的调查研究主要根据《中国语言资源调查手册·民族语言》进行。该手册分为8个分册：藏缅语族、侗台语族（兼用于南亚语系语言）、苗瑶语族、突厥语族（兼用于塔吉克语）、蒙古语族、满一通古斯语族、俄罗斯语、朝鲜语^①。每个分册均分“工作规范”（包括调查规范、语料整理规范）和“调查表”（包括概况、音系、词汇、语法、口头文化、地方普通话）两大部分。为便于使用，各分册单独印刷装订。调查手册中的词表分各民族语言通用的1200条通用词和按照语族或语言分别设计的1800条扩展词，均分天文地理、时间方位、植物、动物、房舍器具、服饰饮食、身体医疗、婚丧信仰、人品称谓、农工商文、动作行为、性质状态、数量、代副介连词14类。其中，阿尔泰语系三个语族的1800条扩展词绝大多数基本一致，部分词汇兼顾了不同语族的特点。俄罗斯语和朝鲜语的1800条扩展词则兼顾了各自语言的词汇特点。

本文基于上述调查手册所采集到的词汇缺失材料的整理与分析：首先对语保工程民族语言调研专项任务所设的33个调研点的词汇缺失情况按照通用词和扩展词进行了初步统计，对某些语种或方言词汇记录中的问题进行了初步核对和甄别，并在此基础上选择了于后语种的

^① 笔者主持并参与了阿尔泰语系三个语族及俄罗斯语、朝鲜语五本调查手册（初定版）中词表（扩展词）和语法条目表等的编写工作。

典型代表点来进行对比分析，它们是：突厥语族的柯尔克孜语南部方言、撒拉语街子方言；蒙古语族的达斡尔语布特哈方言、土族语互助方言；满-通古斯语族的锡伯语察布查尔方言、鄂温克语敖鲁古雅、辉、陈3个方言；印欧语系斯拉夫语族的俄罗斯语伊宁、额尔古纳2个方言。而突厥语族的西部裕固语、乌孜别克语、塔塔尔语，蒙古语族的东部裕固语、保安语，印欧语系伊朗语族的塔吉克语等语言因涉及复杂的细节问题暂时排除在外，将另立专题讨论。另外，因相关统计数据工作量巨大，某些细节未及兼顾，可能在数据统计基础上的分析还不够全面详尽，其涉及到的相关论题研究也将期待后续数据的补充完善而继续推进。

二 词汇缺失调查数据统计分析

（一）词汇缺失总体数据统计分析

表1 词汇缺失总体数据统计表

语系	语族	语种	方言	通用词	扩展词	总计
阿尔泰	突厥	柯尔克孜	南部	133	17	150
		撒拉	街子	115	110	225
	蒙古	达斡尔	布特哈	126	162	288
		土族	互助	28	58	86
	满-通古斯	锡伯	察布查尔	59	56	115
		鄂温克	敖鲁古雅	487	682	1169
			辉	252	229	481
			陈	235	475	710
		鄂伦春	阿里河	337	619	956
		赫哲	同江	524	1071	1595
印欧	斯拉夫	俄罗斯	伊宁	99	36	135
			额尔古纳	170	590	760

表1数据显示，表中所列各语言普遍存在词汇缺失现象，但不同语种情况不同，甚或差异悬殊，甚至同一个语种内部的不同方言或土语间也存在数据不等的差异。从语族为单位来说，满-通古斯语族中，除了锡伯语外，其他3种语言整体上都缺失严重。其中鄂温克语的3个方言间内部缺失数量差异较大。突厥语族和蒙古语族的4种语言缺失数量一般，内部差异较小。俄罗斯语的2个方言中，额尔古纳方言缺失较多，而伊宁方言则缺失较少，形成较为明显的内部差异。再看图1：

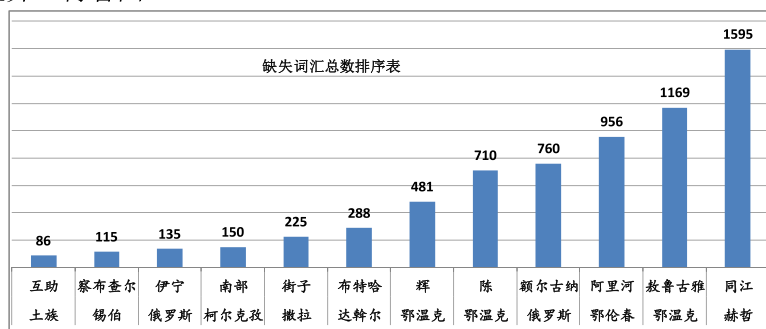


图1 词汇缺失总体数量排序表

从图1所列语种及其方言的排序上看，赫哲语缺失数量最多（1595条），缺失现象非常

突出。其他由多到少依序为鄂温克语敖鲁古雅方言（1169条）、鄂伦春语（956条）、鄂温克语陈方言（760条）、俄罗斯语额尔古纳方言（710条）、鄂温克语陈方言（710条）、鄂温克语辉方言（481条）、达斡尔语布特哈方言（288条）、撒拉语街子方言（225条）、柯尔克孜语南部方言（150条）、俄罗斯语伊宁方言（135条）、锡伯语察布查尔方言（115条）、土族语互助方言（86条）。由此形成由少到多的5个梯次：缺失200条以内的土族互助方言（86条）、锡伯语察布查尔方言（115条）、俄罗斯语伊宁方言（135条）、柯尔克孜语南部方言（150条）为第一梯次；缺失300条以内的撒拉语街子方言（225条）和达斡尔语布特哈方言（288条）为第二梯次；缺失500条以内的鄂温克语辉方言（481）为第三梯次；缺失800条以内的鄂温克语陈方言（710条）和俄罗斯语额尔古纳方言（760条）为第四梯次；缺失1000条以上的鄂温克语敖鲁古雅方言（1169条）和赫哲语（1595）为第五梯次。

（二）通用词缺失数据统计分析

因调查大纲词汇表中的通用词与汉语和其他所有少数民族语言均一致，不同语种、语族甚至语系之间均具有一定的可比性，我们将其也按照缺失数量予以了排序统计。见图2：

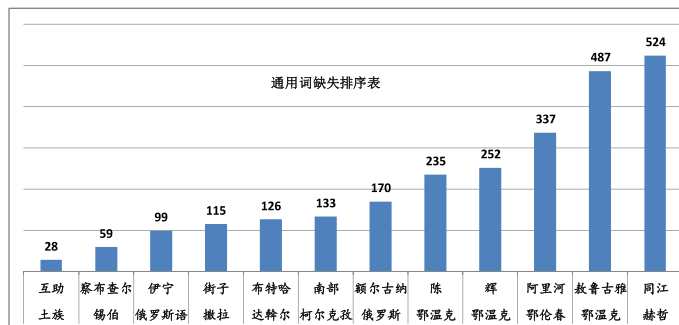


图2 通用词缺失排序表

从图2显示的数据来看，似乎和图1的由高到低的排序前4种是一样的，其余的排序情况与图1有一点小的变动，但差别不算太大。

（三）扩展词缺失数据统计分析

在制定阿尔泰语系不同语族的词汇调查大纲时，考虑到统一性和可比性，我们对1800条扩展词也进行了部分统一，3个语族约70%的词汇相同，约30%词汇略有差异。因此，扩展词也极具可比性，将其数据统计并排序如下，可为后续研究提供一定基础（见图3）。俄罗斯语与上述语言使用不同的扩展词调查大纲，但其相关数据也作为参照列入分析之列。

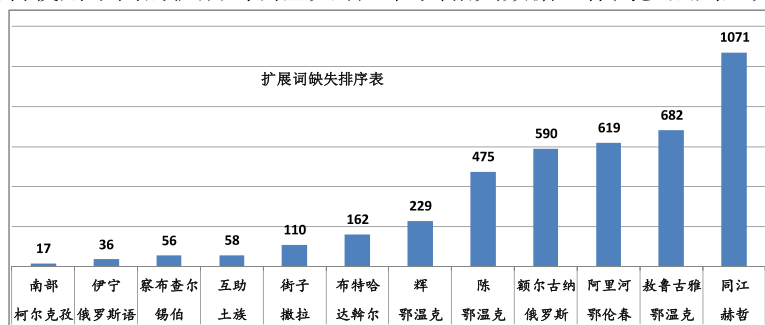


图3 扩展词缺失排序表

图3显示，在图1中前6位的排序基本没有变，其后的排列顺序有一些小的变动，但总体上与图1排序差别不大。

(四) 义类缺失数据统计分析

表2 义类缺失数据统计表

语种	方言	天文地理	时间方位	动植物	房屋器具	服饰饮食	身体医疗	婚丧嫁娶	人品称谓	农工商文	动作行为	性质状态	数量	代副介连	总计
土族	互助	5	2	16	11	9	5	5	6	20	3	3	1	0	86
锡伯	察县	2	0	48	10	7	0	2	2	8	0	0	33	3	115
俄罗斯	伊宁	3	12	18	2	16	3	14	7	5	1	6	43	5	135
柯尔克孜	南部	5	11	36	21	6	0	12	0	8	8	0	33	10	150
撒拉	街子	10	8	20	10	13	7	17	18	36	13	9	41	23	225
达斡尔	布特哈	24	16	10	34	29	3	20	21	64	14	4	44	5	288
鄂温克	辉	27	3	150	60	30	21	7	19	32	44	21	46	13	473
鄂温克	陈	26	18	149	63	39	24	33	40	122	62	53	62	19	710
俄罗斯	额尔古纳	43	30	93	47	59	29	30	72	185	82	19	62	9	760
鄂伦春	阿里河	44	41	177	140	57	46	43	44	170	44	31	89	30	956
鄂温克	敖鲁古雅	54	85	238	106	55	55	55	84	148	72	74	74	2	1102
赫哲	黑龙江	39	49	186	174	90	113	74	89	147	319	183	71	60	1594

表2数据显示，不同语种或方言各义类分布缺失的数量不等，但绝大多数集中于动植物、农工商文及数量三类。其中，动植物类词汇中缺失的是由于环境差异而缺少的一些动植物名称，也在某种程度上反映了不同语言与方言分布于不同自然环境特点的差异。数量词的缺失主要集中于1200条通用词中的量词上。农工商文类比较集中地缺失其中的抽象词、文化词、工业和商业类词语。图4将词汇缺失较多的赫哲语、鄂伦春语、鄂温克语敖鲁古雅方言与辉方言、俄罗斯语的伊宁方言与额尔古纳方言缺失词汇的义类分布数据进行了统计展示。其中，俄罗斯语伊宁方言和鄂温克语辉方言仅作为横向参照数据。

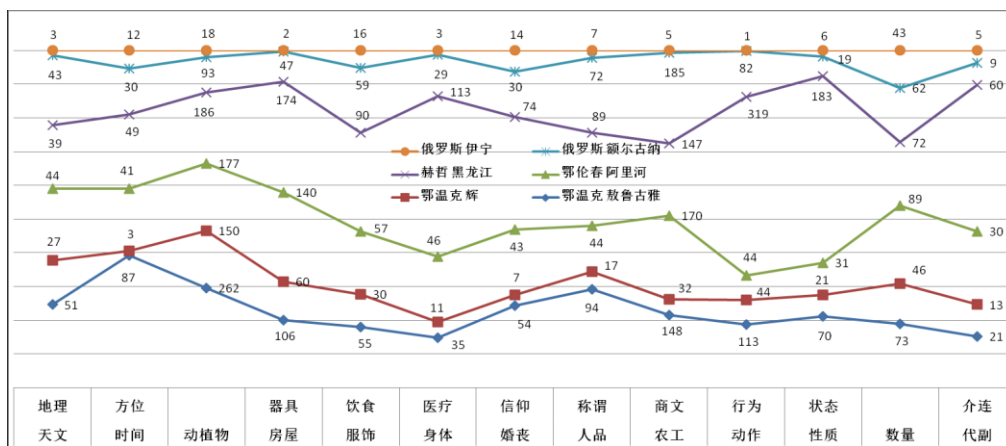


图4 词汇缺失较多的数种语言的义类缺失情况统计表

图4数据显示，几种词汇缺失较多的语种除了与其他所有语言一样集中于动植物、人品

称谓、农工商文、数量等几个义类的缺失外，还缺失器具房屋、行为动作两个义类的词汇。

三 讨 论

（一）词汇缺失成因

上节各图表的统计数据凸显了词汇缺失现象较为复杂的特征，而同时普遍存在于各语族、各语种内部的不平衡性和非线性特点则更令人关注，而这些可能均源自以下不同因素：

其一，自然环境、传统生计生产方式不同。语言本身也是客观世界的反映，属于一种社会现象。人们在什么样的环境里就会产生什么样的语言。如果某一事物或者物象在人们所生活的客观世界中不存在，那么折射在语言上就表现为语言要素中的固有词乃至一般词的缺失。诸如由于地域环境等差异导致的动植物类词汇的缺失显然属于绝对缺失类型。

其二，因语言结构及特点差异所导致的某些词汇，诸如量词等的缺失或不甚发达也应属于语言绝对缺失的一种表现形式。

其三，人口数量多少也可以体现一种语言的活力和社会功能，相对来说，人口数量多的民族语言，因母语人实际基数较大，可能会更多地保留并使用本民族语言中的固有词和基本词，而人口数量更少的民族的语言则可能词汇缺失程度更为严重。

其四，有无文字也是一种语言词汇缺失与否的一个重要原因。一般来说，有文字或书面语形式的语言，其词汇缺失程度相对要低于无文字的语言，尤其是某些抽象词语的缺失可能是无文字民族语言最易缺失的细类。

此外，不同的母语个体也会因生理、心理及精神等原因出现各种词汇缺失现象。

（二）词汇资源与语言活力

孙宏开、黄行曾就中国少数民族的语言活力进行排序。孙宏开（2006）将中国境内的少数民族语言分为充满活力、有活力、活力降低、活力不足、活力很差5种状态，每一类中包含若干语种。黄行（2000:184）以语言活力的各类相关数据作为依据，并根据数据特征将中国少数民族的语言活力分为三类五等。我们认为，词汇缺失论题的研究可为语言资源保持类型或语言活力排序研究提供更为客观和实际的数据支持，其研究结果不仅可以反映一种语言整体功能情况，也可反映一种语言及方言更为细微的差异。

（三）关于借词资源

我们在词汇缺失统计过程中发现，调查者对借词有标注为“无”或不标注为“无”两种处理方法，这也是导致词汇缺失数据统计有所差异的原因之一。对此，我们认为，从语言资源视角来看，一种语言从其他语言借入相关词汇也应看作是一种分享本民族没有但其他民族却拥有的语言资源现象。借用强势语言或邻近语言词汇或其他语言成分，是一种正常的语言交际活动，抑或是一种文化交往活动。但借入的语言资源毕竟是母语中所没有的外来成分。因此，无论借入的词汇有多么古老，有多么常用，它们仍不是本民族固有的语言资源。从这个角度来说，调查采集此类词汇的过程中应予以特别关注或特殊标记，并对其是否绝对缺失或是相对缺失甚至是否已经转化为本语言资源的性质进行深入研究。而在类似借词等词汇的调查中，绝对缺失词汇和相对缺失词汇的调查方法也应有所不同。绝对缺失词汇涉及到明显差异，可能会容易调查一些。而相对缺失的词汇则在调查时需要增加反复询问的环节，以确认某些发音合作人临时因各种原因未能说出的词汇是否是缺失层级的，并允许发音合作人有

一个回忆、询问或查阅资料的过程。

（四）关于词汇调查大纲的制订

中国语言资源保护民族语言调研专项任务中所采用的调查手册，因其科学性、规范性和实用性等整体特点，在调查过程中起到了重要的指导和规范作用。但仍有几个问题需要探讨：其一，调查手册 1200 条通用词中，对汉语方言或部分南方少数民族语言来说，的确具有适用性，而部分蔬菜、水果及动植物名称，尤其是量词等词汇对北方少数民族语言来讲，可能就不甚合理（丁石庆 2018）。又如以语族为单位编写的调查手册，貌似具有一定的合理性。但从调查实际效果来看，有些同语族语言，甚至同语种的不同方言，有时反倒适用其他语族的调查手册，尤其是不同语系的语言通用一个调查手册的做法更值得研究。是否可考虑制定从生计生活方式如游牧、农耕、渔猎等先行进行归类，并同时考虑自然环境、文化特征等因素编制更为适用的调查手册则成为新的规划思路。其二，每个民族都有一些独具本民族文化特色的词汇或称特色词汇，而目前所使用的所有语言资源专题的调查手册中往往缺少了这样一个重要内容的调查，建议在语保工程的后续推进工作中应尽快弥补此类遗憾。其三，我们在语言资源专题调查过程中，发现语料的采集质量和记录研究方法，包括对这些语料的使用在很大程度上取决于调查研究者的学术素养和治学态度。从大的方面讲，对语言资源的理念和认识水平，都会在一定程度上影响调查者的作为。从小的方面讲，调查过程中的调查者的一举一动也在一定程度上决定了所采集语料的客观性、真实性和可靠性。如在调查者提问时是采用启发式还是简单命令式都会对采集的词汇是否缺失或义类分布产生影响。如提问时采用“××词怎么说”和“××意思有几种说法”或“这个词有几种说法”所取得的实际效果则大相径庭。

四 结 语

北方人口较少民族语言普遍存在的程度不同的词汇缺失现象主要分为绝对缺失与相对缺失两类，同时也包括词汇的义类缺失或义项缺失等细类。一种语言的词汇缺失与其所处自然环境、语言结构、语言功能等诸多因素相关，也与有无文字及宗教信仰等因素有关联。词汇缺失数量或可作为界定一种语言活力及语言资源类型的重要参考依据。核心词汇与特色词汇的保护是词汇资源保护乃至语言资源保护的重要任务。

参考文献

- [1] 丁石庆. 2018. 《语言资源保持类型与数据采集层次——以北方民族语言为例》，载张世方主编《语言资源》（第一辑）第 21-36 页，北京：商务印书馆。
- [2] 黄 行. 2000. 《中国少数民族语言活力研究》，北京：中央民族大学出版社。
- [3] 孙宏开. 2006. 《中国少数民族语言活力排序研究》，《广西民族大学学报》第 5 期。
- [4] 张 岩. 2018. 《北方五种民族语词汇缺失问题研究——以语言资源调查数据为例》，中央民族大学硕士学位论文。

Lexical Gaps in Languages of Ethnic Minority Groups with Small Populations in Northern China: A Study Based on Ethnic Minority Language Resources Data Collected in the Language Protection Project

DING Shiqing

[Abstract] Based on statistics and analysis of data acquired from the Special Task on Ethnic Minority Language Investigation in China's Language Resources Protection Project (hereinafter as "the Language Protection Project"), this paper proposes that lexical gaps exist to varying degrees in the languages of ethnic minority groups with relatively small populations in northern China, which manifest mainly as absolute gaps and attritional gaps, involving gaps in semantic categories or gaps in meanings or senses. Generally speaking, lexical gaps are characterized by complexity, imbalance and nonlinearity. Variances in lexical gaps can be attributed to differences in natural environment and traditional means of livelihood and production, characteristics of linguistic structures, presence or absence of a writing system, as well as inhabiting patterns and religious beliefs.

[Keywords] ethnic minority groups with small populations in northern China lexical gap absolute gap

(通信地址: 100081 北京 中央民族大学中国少数民族语言研究院)

【本文责编 木再帕尔】

《民族语文》第十四届学术研讨会延期举行

原定于2020年7月11-12日由《民族语文》编辑部和贵州民族大学文学院联合主办、贵州民族大学文学院承办的“《民族语文》第十四届学术研讨会”，因疫情延期举行，待确定学术研讨会会议日期后，我们将于《民族语文》采编系统网、《民族语文》学术会议微信群、中国民族语言学网、中国民族语言学会QQ群与微信群发布信息。

学术研讨会仍以《民族语文》征稿为目的，领域包括以少数民族语言为研究对象的描写语言学、历史语言学、语言类型学、纪录语言学、接触语言学、语法化、实验语音学、社会语言学、生态语言学、经济语言学、语言政策与语言规划、濒危语言与语言资源保护、汉语与少数民族语言关系等的成果。

期待届时专家学者提交高质量的全文学术论文。

本刊编辑部