

蒙古语族和突厥语族 关系词的词阶分布分析

孟和达来 黄 行

[提要] 本文通过蒙古语族和突厥语族核心词中关系词的词阶分布, 考察了语族内部语言之间和语族之间的语言关系。第一部分从语言标记性的角度论证了核心词词阶划分的内在依据。第二部分统计描述了这两个语族关系词的分布, 发现有些语言呈有阶分布, 有些语言呈无阶分布。第三部分讨论了有阶分布和无阶分布的关系和机制。文章的结论是, 关系词在词汇中的分布可以是有阶的, 也可以是无阶的, 词阶分布理论对证明词汇有阶分布的语言关系有效, 但不能说明词汇无阶分布的语言是同源关系还是接触关系。

根据语言学理论划分基本词汇和非基本词汇的原理, 基本词汇是稳定的, 很少从其他语言借用词汇, 非基本词汇相对来说不那么稳定, 语言之间容易相互借用和渗透。这种差异表明, 词汇是有阶的, 即固有词处于靠近核心的较高的词阶, 借词处于远离核心的较低词阶。如果语言之间的关系词(有音义对应关系的词)在高词阶中的比例大于在低词阶中的比例, 就可以推断它们可能具有同源关系; 反之, 如果关系词在低词阶中的比例大于在高词阶中的比例, 它们就可能没有同源关系而是相互接触的关系。

陈保亚先生的一项关于壮侗语和汉语关系的研究,^① 就是根据这种原理, 通过核心词中关系词在不同词阶分布比例的差异, 即壮侗语之间是高词阶的关系词比例大于低词阶, 壮侗语和汉语之间是低词阶的关系词比例大于高词阶, 证明壮侗语族内部的语言具有同源关系, 壮侗语和汉语之间为接触关系。这项研究采用的两个词阶, 是斯瓦迪士研究语言年代学提出的200个核心词的前100词(1阶词)和后100词(2阶词)。这个核心词系统及其两个词阶的划分已被多项研究证明, 对各种语言都具有相当普遍的适用性。

阿尔泰语系各语族内部的语言具有清楚的同源关系, 但是语族之间的谱系关系学术界历来有分歧, 分歧的症结也在于对有对应关系的词汇及形态是同源还是借用的不同看法。本文拟尝试运用上述理论(以下称词阶分布理论), 通过语言之间核心词中的关系词(以下称关系词)的词阶分布, 考察蒙古语族和突厥语族内部及语族之间的语言关系。在分析关系词的词阶分布之前, 有必要对这个核心词系统划分两个词阶的内在依据做一些探讨。

① 参见陈保亚《从核心词分布看汉语和侗台语的语源关系》, 《民族语文》1995年第5期。

一 词阶的标记性差别

既然核心词也是有阶的, 1 阶词和 2 阶词必然存在某些本质的差异。根据词的意义和功能可以把 200 词分为 15 个词类 (不是语法的词类)。各类词在 1 阶词和 2 阶词中的词数分布如下。

表 1 词类和词阶的词数分布

词 类	1 阶词	2 阶词	合计
1 人称词	4	2	6
2 指示词	4	4	8
3 指人名词	4	5	9
4 自然名词	14	9	23
5 身体名词	21	3	24
6 动物名词	9	4	13
7 植物名词	5	4	9
8 时间词	1	2	3
9 方位词	0	2	2
10 工具食物词	0	3	3
11 数词	2	3	5
12 颜色词	5	0	5
13 性状词	11	19	30
14 动作词	19	36	55
15 虚词	1	4	5
合 计	100	100	200

两阶词在词类上的数量差异反映出这样一种现象, 1 阶词数多的词类表示的概念多与自然事物或属性相联系, 例如表示自然现象、身体部位、动物、植物和颜色的词; 2 阶词数多的词类表示的概念多与社会化了的事物或属性相联系, 例如指人名词中的亲属词、时间词、方位词、工具食物词、数词、虚词等。因此可以说, 词项意义或功能社会化程度的高低是词阶间的一个重要差异。

概念的社会化程度只是影响语言结构的间接因素。如果从词汇系统内部分析, 可以发现词项的词阶差异和词项的标记性之间有着明显的相关关系。这种关系表现为, 在词义对立的两个或几个词项中, 无标记词项不会低于有标记词项所处的词阶。以标记性对立最明显的性状词的词阶分布为例。

表 2 性状词标记性的词阶分布

无标记词项	词阶	有标记词项	词阶
大	1	小	1
长	1	短	2
热	1	冷	1
好	1	坏	2
新	1	旧 (老)	2

干	1	湿	2
远	2	近	2
多	1	少	2
宽	2	窄	2
厚	2	薄	2

各组对立的性状词都有这样一些共同的分布特性,即在陈述式中各词项对立,如:很长:很短,不长:不短(汉语);ix ɔrt: ix ɔxɔr, ɔrt biʃ: ɔxɔr biʃ(蒙古语);bek uzun: bek qisqa, uzun emes: qisqa emes(维吾尔语),在疑问式和中和位置中各词项不对立,如:多长:*多短,长度:*短度(汉语);xədi: ɔrt: *xədi: ɔxɔr, ɔrt ni: *ɔxɔr ni(蒙古语);uzunlixi qantʃe: *qisqilixi qantʃe, uzunluq: *qisqiliq(维吾尔语)。根据标记性原理,各种分布均能出现的词项是无标记词项,只能在部分分布中出现的词项是有标记词项,因此表2左列和右列每组对立的词项分别为无标记和有标记词项。其他词类的词如果存在某种意义上的对立,通常也是无标记和有标记的差异。例如:我、你(1阶):他(2阶),我们(1阶):他们(2阶)是第一、第二人称和第三人称的对立;这、那(1阶):这里、那里(2阶)是单纯指示词和复合指词的对立;人(1阶):孩子(2阶)是成年与未成年的对立;男人、女人(1阶):父亲、丈夫、母亲、妻子(2阶)是人的自然属性和社会属性的对立;鸟(1阶):翅膀(2阶)是本体和属性的对立;树(1阶):树林(2阶)是单数和复数的对立;一、二(1阶):三、四、五(2阶)是小数和大数的对立等。这些方面的对立均为标记性差异,^①即对立的前项为无标记项,后项为有标记项,并且无标记项不会低于有标记项所处的词阶。

以上分析说明,核心词的词阶划分从语言的外部关系看,与词项概念的社会化程度有关;从语言结构内部看,与词项的标记性程度有关。这些相关因素是导致不同词阶稳定性不同的主要原因。一般说来,词阶越高的词项,其概念的社会化程度和词汇的标记性程度越低,因此在语言中的稳定性越强,不容易在语言接触中受到干扰和影响;词阶越低的词项,社会化程度和标记性程度越高,因此稳定性不强,在语言接触中容易发生互相借用。

二 蒙古语族、突厥语族语言关系词的词阶分布

我们选择了国内6种蒙古语族语言^②和8种突厥语族语言^③进行了语族内部和语族之间关系词词阶分布的分析和统计。具体统计结果见下表。

表3 蒙古语族语言关系词的词阶分布

语言	200词	1阶词	2阶词	1阶词>2阶词
蒙(古)-达(斡尔)	166	90	76	14
蒙-土(族)	165	91	74	17
蒙-东(乡)	148	87	61	26

① 参见 Greenberg, J. H. Language Universals, Hague, 1966。

② 选自孙竹《蒙古语族语言词典》,青海人民出版社,1990。

③ 选自陈宗振等《中国突厥语族语言词汇集》,民族出版社,1990。

蒙-保(安)	138	85	53	32
蒙-东(部)裕(固)	166	91	75	16
达-土	155	87	68	19
达-东	145	84	61	23
达-保	136	82	54	28
达-东裕	152	84	68	16
土-东	156	89	67	22
土-保	148	87	61	26
土-东裕	160	88	72	16
东-保	154	92	62	30
东-东裕	144	83	61	22
保-东裕	140	84	56	28
平 均	151. 53	86. 93	64. 60	22. 33

表4 突厥语族语言关系词的词阶分布^①

语 言	200 词	1 阶词	2 阶词	1 阶词>2 阶词
维(吾尔)-哈(萨克)	184	93	91	2
维-柯(尔克孜)	186	95	91	4
维-乌(孜别克)	194	97	97	0
维-塔(塔尔)	179	88	91	-3
维-图(瓦)	139	73	66	7
维-撒(拉)	149	83	66	17 *
维-西(部)裕(固)	151	80	71	9
哈-柯	191	98	93	5
哈-乌	179	91	88	3
哈-塔	184	94	90	4
哈-图	142	74	68	6
哈-撒	158	87	71	16 *
哈-西裕	148	79	69	10
柯-乌	184	94	90	4
柯-塔	184	92	92	0
柯-图	144	73	71	2
柯-撒	154	87	67	20 *
柯-西裕	150	80	70	10
乌-塔	177	87	90	-3
乌-图	134	70	64	6
乌-撒	149	84	65	19 *

① 突厥语族语言关系词词阶分析统计结果承蒙赵明鸣先生提供, 特此致谢。

乌 - 西裕	147	79	68	11
塔 - 图	139	72	67	5
塔 - 撒	156	86	70	16 *
塔 - 西裕	148	77	71	6
图 - 撒	127	70	57	13
图 - 西裕	128	66	62	4
撒 - 西裕	138	79	59	20 *
平 均	158. 68	83. 14	75. 54	7. 61

带 * 的数字表示通过 Z 值假设检验 (设显著度 $p \leq 0.05$), 词阶的词数分布有显著差异。

从表 3 和表 4 的统计结果可以看出, 蒙古语族和突厥语族内部都有很高的关系词比例, 并且关系词在 1 阶词中的比例均高于 2 阶词。根据词阶分布理论, 可以证明语族内部的语言具有同源关系, 同时也证明 1 阶词比 2 阶词具有更大的稳定性。不过我们也注意到, 在突厥语族中, 多数语言之间关系词在两个词阶中的比例经过假设检验统计后并无显著差异。据分析, 越是结构相近 (可以互相通话) 的语言词阶差异越小; 而 1 阶词显著高于 2 阶词的, 却是语言结构较远 (通话有困难) 的语言。因此对于没有显著差异的无阶分布语言, 不能用词阶分布理论证明其语言关系。

为了检验词阶分布理论对证明系属关系不清楚的蒙古语族和突厥语族语言之间关系的有效性, 我们进一步考察了蒙古语和 4 种突厥语之间关系词的分布情况。表 5 是统计结果。

表 5 蒙古语和 4 种突厥语族语言关系词的词阶分布

语 言	200 词	1 阶词	2 阶词	1 阶词 > 2 阶词
蒙古 - 维吾尔	79	40	39	1
蒙古 - 哈萨克	78	39	39	0
蒙古 - 图瓦	80	41	39	2
蒙古 - 西部裕固	76	39	37	2
平 均	78. 25	39. 75	38. 50	1. 25

蒙古语和各突厥语之间的关系词平均为 78. 25 个, 明显少于蒙古语族或突厥语族内部语言之间的词数; 关系词在 1 阶词和 2 阶词中分别平均为 39. 75 和 38. 50 个, 没有显著差异。这种无阶分布状况用词阶分布理论仍然不能断定两个语族之间是同源关系, 还是借用关系。问题在于如何认识关系词的有阶分布和无阶分布的性质和区别。

三 关系词分布的有阶性和无阶性

无论是语言同源关系表现出的关系词词阶分布的下降趋势, 还是语言接触关系表现出的关系词词阶分布的上升趋势, 都有一种共同的性质, 就是词汇分布的有阶性。我们在上面分析蒙古语和突厥语的关系词分布时, 通过假设检验, 证实了突厥语族除少数几种语言外, 多数语言之间关系词的词阶分布没有显著差异; 蒙古语和突厥语之间的词阶分布差异更小。这种现象说明, 无阶分布也是词汇的正常分布特性, 是我们研究词汇分布时不能回避的。

为了突出关系词有阶分布和无阶分布的差异, 我们把蒙古语族和突厥语族的语言各分为

两组,第1组是受外部影响较小,保留本语族固有词较多的语言,对蒙古语族来说有蒙古语、达斡尔语、土族语和东部裕固语,对突厥语族来说有维吾尔语、哈萨克语、柯尔克孜语、乌孜别克语和塔塔尔语。第2组是受其他语言接触和影响较多,固有词比例相对较低的语言,蒙古语族中有东乡语和保安语,突厥语族中有图瓦语、撒拉语和西部裕固语。分组后关系词的词阶分布%统计如下。

表6 蒙古语族、突厥语族语言分组关系词的词阶分布平均%

语 言	200 词	1 阶词	2 阶词	1 阶词>2 阶词
蒙古语族	75. 77	86. 93	64. 60	22. 33
蒙古 1 组	80. 33	88. 50	72. 17	16. 33
蒙古 2 组	77. 00	92. 00	62. 00	30. 00
突厥语族	79. 34	83. 14	75. 54	7. 61
突厥 1 组	92. 10	92. 90	91. 30	1. 60
突厥 2 组	65. 50	71. 67	59. 33	12. 33
蒙古语与突厥语	39. 13	39. 75	38. 50	1. 25
蒙古语与突厥 1 组	39. 25	39. 50	39. 00	0. 50
蒙古语与突厥 2 组	39. 00	40. 00	38. 00	2. 00

表6统计结果表明,对蒙古语族和突厥语族来说:1. 因为两个语族的1组语言关系词比例均高于2组语言,所以与外部接触少的语言(1组)同源程度要高于外部接触较多的语言(2组)。2. 关系词的词阶分布比例在不同语言组中差别较大,都是2组语言的词阶差异大于1组语言。也就是说,同源程度越高的语言,关系词词阶分布差异(无阶性)越小;同源程度越低的语言,词阶分布差异(有阶性)越大。3. 突厥1组语言之间、蒙古语和突厥语言之间关系词没有词阶分布差异,词汇分布是无阶的,因此不能用词阶分布理论来推断各突厥语言是否有同源关系,更不能因为诸突厥语言事实上有同源关系,而蒙古语、突厥语之间又具有和突厥诸语之间相同的词阶分布状态,而断定它们也有同源关系。

亲属关系确凿、关系词(同源词)比例很高的突厥语言之间和亲属关系不清楚、关系词比例不高的蒙古语和突厥语之间的关系词都呈无阶分布状态,因此很难直观地看出无阶分布和什么样的语言关系有联系。如果把无阶分布和有阶分布结合起来,就会发现一种至少对阿尔泰语言来说很有趣的分布或变异过程:同源程度很高,受外部语言影响小的语言(如多数突厥语族语言)词汇分布是无阶的→有同源关系,同时也受到其他语言一定影响的语言(如蒙古语族语言);或者没有同源关系,但相互有深刻接触的语言(如壮侗语和汉语),固有词和借词呈有阶分布状态→语言关系疏远,关系词比例较低的语言(如蒙古语族和突厥语族语言),词汇又呈无阶分布状态。也即无阶分布处于两端,有阶分布介于中间。这种分布过程可以解释为,当语言高度同源时,词汇的词源差异极小,固有词和借词随机地(无阶地)分布在各个词阶;当语言因外部接触而出现分化或混合时,借词需要遵照一定的顺序(有阶地)从外缘进入词汇系统;当语言分化或混合到相当大的程度时,词汇系统中的固有词和借词因词项身分的逐渐模糊而又回归到新的无阶分布状态。从更抽象的角度说,语言处

于初始的同质状态时,语言成分的分布是随机和平衡的;随着语言接触造成语言的同质成分和异质成分混杂时,异质成分要按特定的顺序或等级进入语言系统,因此出现了语言系统的不平衡状态;当异质成分被语言系统吸收而达到新的同质状态时,语言成分在系统中又趋向随机和平衡的分布。随机和平衡是无序的表现,差异和非平衡是有序的表现,因此无阶分布和有阶分布也可以称无序分布和有序分布,它们是语言系统及任何系统都具有的两种基本的分布状态。

回到本文的题目,我们可以得出这样的结论:仅根据词阶分布理论,因关系词呈有阶分布,因此可以证明蒙古语族语言之间及部分突厥语族语言之间有同源关系;多数突厥语言以及蒙古语族和突厥语族之间关系词呈无阶分布,因此仍不能确定它们之间是什么关系。因为这种理论仅对有阶分布语言有解释力,对突厥语言那样亲缘关系十分密切的语言,或阿尔泰语系各语族那样关系已经相当疏远的语言来说,词汇(不论是固有词还是借词)分布基本是随机的,因此要证明它们的语言关系(阿尔泰语系各语族内部语言无需证明),还需要通过其他更为有效的理论和方法来解决。

(通信地址:100081 北京 中国社会科学院民族研究所)

~~~~~

## 金铜古名小议

早在夏代,黄河流域的居民已使用和冶炼铜。春秋战国以前“金”一般指的是铜。《左传·僖·十八》:“郑伯始朝于楚,楚子赐之金,……故以铸三钟。”金子称为黄金,后来黄金用途日广,金和黄金名称易混淆才有“铜”之名。“金”字见于周中期金文,“铜”字最早见于战国金文。古代“金”又为金属总称。《说文》:“金,五色金也,黄为之长,久藏不生衣,百炼不轻,从革不违。”“铜,赤金也。”《尔雅·释器》:“黄金谓之璠,其美者谓之璆。”“璆”上古音\*glau(群幽)。而中原以外古代西北、东北地区的语言可能把铜叫做“璆”。“铜”缅文kre,景颇语mākzi——mā-为前缀,k3-<\*kr-。印夏曼尼普尔的藏缅语族那加语(Tangkhu Naga)“铜”kori。朝鲜语kuri,鄂温克语gooli,西部裕固语gula,蒙古语yauli(黄铜)。汉语“金”和“铜”分别指黄金和青铜约始于战国时代,侗台、苗瑶语的“金”和“铜”应是古汉语借词。岭南地区铜的使用远晚于中原地区,大约始于西周。而早在4千年前,藏缅先民在黄河上游已有铜的冶炼和冷锻。齐家文化的遗址中出土有铜斧、铜镜。“铜”\*daŋ(定东)古名与“璠”daŋ²(定阳)相近。

吴清河