

藏语形容词性语素研究^①

龙从军

中国社会科学院民族学与人类学研究所计算语言学研究室, 北京 100081

longcongjun@163.com

摘要

藏语形容词以二音节居多,其中以词根加词缀比较常见。带词缀的二音节形容词在句子中充当谓语、补语等语法成分时,后缀往往省略。根据这一语法现象,本文提出建立藏语形容词性语素库,旨在为藏语的机器自动处理服务。文章首先阐述了提取形容词性语素的原则与方法,然后介绍形容词性语素库的设计情况。

关键字

现代藏语 形容词 词缀 形容词性语素 语素提取 形容词性语素库

1. 引言

信息处理离不开语言知识库的支持,词汇知识库和语法信息库的建设是语言知识库的核心内容。当前自然语言处理对词汇越来越重视,较典型的范例是美国的 wordnet 项目,国内也有董振东先生研制的 hownet 和北大计算所开发的《现代汉语语法信息词典》。这些具有语言工程性的研究成果已逐渐运用于语言信息处理项目。藏语语法研究虽然起步较早,但长期囿于传统古典文法羁绊,很难在当今信息处理领域直接应用。

本文以所构建的藏语形容词性语素库为依据,尝试以现代语言学理论对藏语形容词基本特征进行探讨。具体内容包括藏语形容词的基本特点分析,提取藏语形容词性语素的基本原则与方法,以及介绍形容词性语素库构建的基本情况。

2. 藏语形容词基本特点

现代拉萨藏语形容词一般由单音节、双音节和多音节组成,尤其是双音节多为词根语素加后缀构成,词根语素义是词的核心意义,词缀只表语法意义。常见的形容词

^①本项研究获国家自然科学基金资助(项目批准号 60173024, 60473135)。本文得到江荻教授的悉心指导,提出了宝贵的意见和建议,在此表示致谢。

后缀有 པོ/བ/པ/མ/མོ (po/ba/pa/ma/mo) 其它还有 གུ/ཁ/འ/བུ (gu/kha/bu)等。在古代藏语或者其它藏语方言中, 这些后缀常常表达某种语法意义, 如 བ (bu) 表示男性, མོ (mo) 表示女性。在现代拉萨藏语中这些语法意义不一定都保存下来。只有表示级范畴的一对后缀保存得相对完整。པོ (po) 表示原级, བ/པ (ba/pa) 表示比较级, 如: དམའ་པོ (dma'po) “低”, དམའ་པ (dma'pa) “较低”; རིང་པོ (ring po) “长”, རིང་བ (ring ba) “较长”。但是这种规则也不能贯彻到整个藏语形容词中, 仅仅局限于带后缀的二音节形容词。因此, 为了处理方便, 我们把这种级的差别看作是词汇本身程度上的差别, 在处理中把它们作为一个词来处理。

非词缀的双音节、三音节及其三音节以上的其它形容词大部分由形容词词根语素的不同组合构成, 还有一些通过词根重叠构成不同音节的状态词, 如: ལྷོག་ལྷོག་སུད་སུད (lkog lkog sud sud) “鬼鬼祟祟”, སྐྱ་ཐེར་ཐེར་ (skya ther ther) “光秃秃”。三音节和三音节以上的形容词其结构大多是名词加形容词, 而名词一般是整个形容词的属性名词, 构成名词(属性)+形容词(属性值)的格式。如: གོང་ཆེན་པོ (Gong chen po) “贵”, དྲི་མ་ངན་པོ (Dri ma ngan po) “臭”。其中 གོང (Gong)、དྲི་མ (Dri) 是名词, ཆེན་པོ (chen po)、ངན་པོ (ngan po) 为形容词, 通过语料观察, 这些形容词基本上都是运用比较常用的形容词词根语素组合而成, 比如 ངན་པོ (ngan po)。由词根语素 ངན (ngan) 和后缀 པོ (po) 组成。

藏语形容词除了在构词上有许多特点之外, 句法上的特点也值得关注。一些带后缀形式的形容词当进入句子充当谓语和补语时往往只能是词根形式。如:

(2.1) རྣམ་པ་འི་མཉམ་རྒྱུད་ལ་ཡག་དག་དག་རེད་ཀྱི།

rnam-pa-ni khyed-rang -la yag-drag-drag-red.

相貌-PAP 你-OBT 点 是

对你来说, 相貌好了点

(2.2) ཡག་ཤོས་བྱང་ན་ཀྱི་ཆུང་ཆུང་ལྗོད་པོ་དེ་འདས་ཀྱིས་གཙོག་དགོས་རེད་ཀྱི།

yag-shos-byang-na chung-chung-ljid-po-de-vdras-kyis gcog-dgos-red.

最好 的话-POC 小小 重 这样-INS 打断 要

最好的话, 要用这种轻点的东西打断。

(2.3) ང་ཚོའི་ཐོབ་ལས་ཐོན་སྐྱེད་ཡག་ཅུ་ཡག་ཅུ་འགྲོ་གོས་ཀྱི།

nga-thovi-zhing-las-thon-skyed yag-ru-yag-ru vgro-gis.

我们-POS 农业 生产好-COP好-COP 去-

我们的农业生产越来越好。

(2.4) ག་འདས་བྱས་ན་ཡག་གི་རེད་ཀྱི།

ga-vdras-byas-na yag-gi-red.

怎样做-POC 好

怎样做才好。

以上句子中的形容词都只用形容词词根形式 ཡག (yag), 而不采用带词缀标记的词

汇形式 ཡག་པོ་ (yag po)。以 ཇེན་པོ་ (che po) 为例, 从《拉萨口语会话》中统计, 词汇形式入句的有 43 个, 而以词根形式入句的反倒有 48 个。显然, 对文本中词根形式形容词的处理不能简单利用收录词汇形式的形容词词典进行。在可选的其他方法中, 如建立形容词词法规则, 或建立形容词词根语素库, 本文认为后者更易于实现形容词的文本自动处理。

3. 形容词语素提取原则及方法

从形容词词条中抽取形容词性语素可以遵循霍凯特对语素的话语切分方法, 切分语素就是切出话语中以大致相同的意义复现于不同的话语的最小部分。切分和验证的手续是 (1) 先从一段话语中切分出以大致相同的意义复现于不同的话语的部分 (2) 再检验这些切分出的部分能再分为大致相同的意义复现于不同的话语中的更小的部分。我们把霍凯特对一种语言的语素切分方法运用到藏语形容词语素的切分中, 从不同的形容词中切分出同型的部分, 然后再从这些同型的语素中归纳出不同的语素意义。

由于藏语形容词性语素要通过归纳整理, 因此必须要有一定的标准, 但是这种主观的标准不好确定, 只有对藏语词汇全面系统的研究后, 再确定标准应该说比较科学, 目前我们大多采用内省的方法并遵循下面的原则来提取形容词性语素:

第一, 明确性原则

明确性是指提取的形容词性语素必须有实际意义, 即有确定的语素意义, 基本规则是同一个形式在不同的形容词中多次出现, 就可以把其提取出来。如: མཉམ་པོ་ (khyab) 出现在 མོངས་མཉམ་ (yongs khyab) “普及; 普遍”, མཉམ་རྒྱལ་ (khyab rgyas) “宽广”, མཉམ་ཅུང་ (khyab chung) “不普遍; 范围狭小”等词中, 基本意义为“遍、广”, 因此可以把它作为一个基本语素提取出来。

第二, 常用性原则

常用性指的是提取的形容词性语素比较常用。从统计的情况看, 藏语形容词约 2000 个, 其中双音节约 1400 个, 三音节及三音节以上的约 600 个。但是从拉萨口语材料看, 常用的也就 200 个左右。这些常用语素反复出现在其它多音节形容词中, 有些加后缀构成带后缀的双音节形容词, 如 དཀའ་བ་ (dkav ba) “困难; 艰难痛苦”, 有些与其它语素按照一定的组合方式组成双音节或多音节形容词, 如 ཇེ་ཅུང་ (che chung) “大小”, ཀ་མཉམ་པོ་ (kha yang ba) “口松; 说话随便”。

确定了上面的原则, 下面是提取归纳的具体步骤:

(1) 收录单音节形容词。这些语素在词典注为单音形容词, 如 རྩལ་ (kyang) “直的; 公正的”。可以把它们定义为成词语素, 它们常常出现在多音节形容词中, 是多音节复合形容词的构词语素, 如 རྩལ་པོ་ (kgyang po) “直的; 公正的”。在藏语中, 这样的单音节成词语素约 150 个。

(2) 带后缀的双音节形容词, 提取词根作为语素。如 དཀར་བ་ (dkar ba) “白的; 忠

心耿耿；清白/纯洁)、**གང་བོ་**(grang bo) “冷；寒冷；严寒”、**ཞིབ་མོ་**(zhib mo) “精细；精密；细致/吝啬” **སྐྱིད་པ་**(skyid pa) “快乐；幸福”。可以分别把**དཀར་གང་ཞིབ་སྐྱིད་**(dkar\grang\zhib/skyid)作为语素。在标注这些语素的意义时，对一个语素表示多种意义的，采取分条记录，比如：

དཀར་ (dkar)	ཞིབ་ (zhib)	སྐྱིད་ (skyid)
意义 1: 白的	意义 1: 精密；细致	意义 1: 快乐；幸福
意义 2: 忠心	意义 2: 吝啬	意义 2: 舒服的
意义 3: 纯洁		

这一步工作我们基本上参考现有的藏汉词典和已经加工整理过的语料。

(3) 对复合词的处理略微显得困难，首先得把复合词中的语素切分出来，然后提取形容词词性语素。如下面的表：

词	意义	类别	语素 1	意义	类别	语素 2	意义
དཀའ་བདེ་ (dkav bde)	平安；快乐	a	དཀའ་ (dkav)	快乐	a	བདེ་ (bde)	吉祥，安乐
བདེན་བྲལ་ (bden bral)	不公正；假的	a	བདེན་ (bden)	诚实，公正	a	བྲལ་ (bral)	分离的
བབ་བོདོ། (bab brdol)	莽撞；鲁莽	n	བབ་ (bab)	情状	a	བོདོ། (brdol)	狂乱的
ཐུགས་བྲེལ་ (thugs brel)	忙碌；繁忙	n	ཐུགས་ (thugs)	心	a	བྲེལ་ (brel)	忙

表一：双音复合形容词切分表

我们只收录了**དཀའ་བདེན་/བདེན་བྲལ་/བོདོ།/བྲེལ་**(dka/bden/bde/bral/brdol/brel)。因为其它的语素不是形容词性的。确定这些语素意义的方法是通过考察多个含有该语素的复合词，然后切分语素并归纳出它们的语素意义，如下面表中是含有**མངོན་**(mngon)这个语素的形容词，通过观察归纳出它的基本语素意义为“显著，显然”，把它确立为一个形容词性语素。

形容词	意义	语素	语素义
མངོན་ཅུང་ (mngon chung)	虚；拘谨/弱小	མངོན་ (mngon)	显著， 显然
མངོན་འཕགས་པ་ (mngon 'phags pa)	特别高贵		
མངོན་རྒྱལ་པ་ (mngon rlom pa)	自负；自大		
མངོན་སྲུང་ (mngon sum)	显现；显露		
མངོན་གསལ་ (mngon gsal)	明显；清楚		
མངོན་གསལ་དོད་པོ་ (mngon gsal dod po)	突出；显著		

表二：提取语素表

(4) 通过整理, 最后得到约 850 个形容词性的语素。比较发现, 从前两个步骤得到的语素一共有 660 个左右, 约占 75%, 而从 600 多个复合词的 1200 多个语素中提取出不同的形容词性语素 190 个左右, 三音节以上的复合形容词构成形式基本上是一个属性名词加上一个形容词。如表三。

从表中看出, 形容词的整个意义来源于名词词根语素义和形容词词根语素义的组合, 组合成的形容词多为最常见的形容词。可以把名词性语素和形容词性语素分开的意义和组合在一起的意义看成形象与抽象的关系, 比如: “价格大” 是形象的表示, 而“贵” 是在此意义上的抽象。

གཏོམ་རིང་པོ། (gtom ring po)	འཇམ་མཚན་པོ། (zheng kha chen po)	འཇམ་མཚན་ཅུང་ཅུང་། (zheng kha chung chung)	གླ་རིང་པོ། (thag ring po)
n (深度) + a (长)	n (宽度) + a (大)	n (宽度) + a (小)	n (距离) + a (长)
浅	宽	窄	远
རྩེ་རྩེ་པོ། (rtse rno po)	མཚན་མེད་པོ། (mdog nyes po)	གོང་ཅན་པོ། (gong chen po)	གོང་མེད་པོ། (gong khe po)
n(顶端)+a(尖)	n(面容)+a(坏)	n(价格)+a(大)	n(价格)+a(便宜)
尖	丑	贵	便宜

表三: 多音复合形容词切分表

总结上面藏语形容词以及形容词性语素在词汇与句法上的特点, 我们觉得构建形容词性语素知识库是藏语信息处理必不可少的。因此, 下面谈谈形容词性语素知识库的构建情况。

4. 形容词性语素库的设计

形容词性语素知识库设计的基本目的是可以为后续形容词研究服务, 长远目标是为藏语信息处理服务。因此, 首先需要对形容词作全面了解。如哪些语素可以加后缀, 可以加哪些后缀, 以及与其它语素组成复合词时的语序位置等等问题都需要涉及。

语素库中的一个语素可以有多个义项, 比如, ཇེ་ (che) 有两个义项, 一是“大”, 一是“老”。我们规定: 一个语素的一个义项为一条记录, 在这条记录下设置多个属性字段, 每一个字段标示这个语素的该义项的一个属性。各个属性字段如下:

藏文词形: 一个形式一般来说是一条记录, 但也存在一个形式有两条以上的记录, 比如 ཇུང་ (chung) 一个意义为“少”, 一个意义为“小”, 因此为两条记录。

词形转写: 转写形式采用本课题设计的转写方案。

意义: 指从形容词中提取出的语素意义, 使用汉语注释, 比如: མང་ (mang): 语素意义为“多”。

后缀: 有些词根语素不只一个后缀, 因此, 我们分别用了“后缀 1, 后缀 2, 后缀

3”三个字段属性。比如：ཀླུ་ལྷོ་ལྷོ་ (gsal) “亮”有三个比较常见的后缀，后缀 1: འ་ (ba)，后缀 2: མ་ (ma)，后缀 3: ཀ་ (kha)。

类别属性：我们计划对藏语语素全面研究，比如名词语素、动词语素等，考虑到以后整个语素库的一致性问题，规定形容词性语素类别为“As”。

语素组合位置：即一个形容词语素在与其它语素一起组成复合词时的位置。使用两个值表示，“自由”与“定位”。其中“定位”语素根据具体位置可分为：“前、中、后”，“前”表示该语素放在复合词的第一个位置，“后”表示该语素放在复合词的最后一个位置，“中”表示该语素放在复合词的非第一个和最后一个语素位置。但在藏语中，除了那些黏附性的后缀外，大多数语素都是自由语素。比如：ཆེ་可以组合成：བྱོལ་ཆེ་ (bros che) “膨大；体积大”，ཆེ་བུ་འཕུལ་ (che btsan) “强大；强狠；厉害”，སྐྱེད་ཆེ་བ་ (stobs che ba) “强大”三种位置均可，故为自由语素。

常用组合词：举例列出常见的该语素的组合词，可以加深对语素构词的了解。比如：མང་(mang)常见的组合词有：སྐྱེད་མང་ (sna mang) “繁多”，མང་ལྷོ་ལྷོ་ (mang nyung) “分量；数量/多与寡”，མང་པོ་ (mang po) “许多；众多”，མང་བ་ (mang ba) “许多；众多”等等。

反义语素：反义关系是形容词中最为常用的关系，在考虑形容词性语素时，应注意这种关系，有些形容词性语素有直接的反义语素，有些有间接的反义语素，比如：མཐོ་ (mtho) “高”与དམུག་ (dman) “低”，བློ་བཟང་ (bzang) “好”与དུག་ (ngan) “坏”，是两对直接反义语素。

反义构词：是词汇中比较特殊的一类现象，构词语素由两个意义相反的语素组合构成新词后表示一种新的意义。例如：མང་ལྷོ་ལྷོ་ (mang nyung) “多寡；多少”，ཆེ་ཆུང་ (che chung) “大小”，བློ་བཟང་ (bzang ngan) “善恶；好坏”。还有一种反义构词现象即一个形容词性语素加上一个表示否定的词，表示与该语素义相反的意义，否定词的位置不定。如果形容词性语素带一个后缀 [一般来说为ཅན་ (can)]，则要省去后缀再加否定词，比如：མཚོ་མཚོ་ (mdzes) “美”与མི་མཚོ་མཚོ་ (mi mdzes) “不美”，དབང་ཅན་ (dmang can) “有力”与དབང་ཅན་ཅན་ (dbang med) “无力”。

重叠式：重叠式是藏语构词的一种常用手段，形容词性语素往往可以重叠，而且重叠的形式也有多种形式，重叠后构成状态形容词。比如，AA 式：ཀུག་ཀུག་ (kug kug) “弯曲的；不直的”，ABB 式：སྐྱེད་འཕུལ་འཕུལ་ (skya ther ther) “光秃秃秃”，AAB 式：ལོ་ལོ་ལོ་ལོ་ (yom yom pa) “摇摆；颤动”，AABB 式：ལྷོ་ལྷོ་སྐྱེད་སྐྱེད་ (lkog lkog sud sud) “鬼鬼崇崇”，Aa^② 式：ཀླུ་ལྷོ་ (kya kyor) “衰弱；歪斜；摇摆”，ABaB 式：ཀ་རེ་ཀ་རེ་ (ka re ko re) “弯弯曲曲；磨蹭”。

属性名词：名词和形容词之间的关系是属性和属性值的关系，属性词语是表示事物的某种属性，由名词来充当；表值词语表示事物在某种属性上的取值，由形容词来

^② 小写字母表示元音有变化。

充当。我们知道一个实体概念能够具有多种属性，而每一个属性如果都对应一个属性值的话，就可以形成一个概念的属性与属性值之间的关系列表，比如：མི(mi)“人”这个概念，可以从各个方面来界定：ལོ(lo)“年龄”的属性值可以用རྒན་པོ་(rgan po)“老”和ཅང་(chung)“小”来赋值，ལྗང་པོ་(ljid pa)“体重”可以用ལྗང་པོ་(ljid po)“重”和ཡང་པོ་(yang po)“轻”来赋值等。确定属性名词比较困难，也有许多方法可以选用。但对藏语来说，可以找到一种比较简便的方法，即从复合形容词中提取属性名词，前面已经谈过复合形容词中有一类特殊现象，即复合形容词的基本结构由名词(属性)+形容词(属性值)构成，比如：འཇམ་མཉམ་པོ་(zheng kha chen po)“宽”，其中འཇམ་པོ་(zheng kha)为属性名词，མཉམ་པོ་(chen po)为属性值。因此，可以直接提取出一个属性名词。

例句：例句是从真实语料中抽取出来对词项加以描述的必要支撑，以མཐོ་(mtho)“高”为例：

ཕོ་བྲང་རྩེ་པོ་ཏུ་ལ་མགོ་ལོས་ཡོད་རེད་།

pho-brang-rtse-po-ta-la mtho-los-yovo-red

布达拉宫殿 高度多有

布达拉宫殿很高。

上面描写了每一个形容词性语素的属性字段，下面看一个具体的例子：

藏文 词形	词形 转写	意义	后缀 1	后缀 2	类别 属性	语素组 合位置	反义 构词	重叠 式	属性 名词	组合词	例句
ཇེ	che	大	བ	ཀ	A s	自由	ཇེ་ཅང་	ཇེ་ཇེ	ཇེ་ལྗང་པོ་	略	略 [®]

表四：属性字段表

5. 结语

藏语形容词性语素库是藏语语法信息词典研发中不可缺少的一部分，由于受到各种条件的限制，我们的研究规模比较小，目前，已经从形容词中提取 800 多个形容词性语素。完成了对形容词性语素库的设计，填写了约 200 个最常用的形容词性语素，剩下的语素正在完成之中。

参考文献：

- [1] 胡坦等，1999年，拉萨口语读本，民族出版社
 [2] 霍凯特：《普通语言学教程》，北京大学出版社
 [3] 金鹏，《藏语简志》，民族出版社

[®] 为了便于排版而省略。

-
- [4] 江荻, 2005 年, 面向机器处理的现代藏语句法规则和词类、组块标注集, 《中国民族语言工程研究新进展》, 社会科学出版社
- [5] 罗秉芬等, 2002 年, 《拉萨口语会话手册》, 西藏人民出版社
- [6] 俞士汶等, 1999 年, 现代汉语语素库的开发及应用, *世界汉语教学*, 第 2 期
- [7] 苑春法, 1998 年, 基于语素数据库的汉语语素及构词研究, *世界汉语教学*, 第 2 期
- [8] 张怡荪主编, 1993 年, 《藏汉大词典》, 北京民族出版社
- [9] 周季文、谢后芳, 2003 年《藏语拉萨话语法》, 民族出版社
- [10] Jiang Di: The Construction and Identification Approaches of Adjectival Predicate in Modern Tibetan.

Research on adjectival morpheme in Tibetan

Congjun Long

Institute of Ethnology & Anthropology Academy of Social Sciences

Department of Computational Linguistics, Beijing 100081

longcongjun@163.com

Abstract:

Most of adjectives are Disyllabic in Tibetan; and they consist of the root and the affix. However, when acting as the predicate and complement in the sentences, the affix is omitted. So, we cannot find the form of the adjectival root in dictionaries existed, In terms of the phenomena, we raise to construct an adjectival morpheme database, we abstract the adjectival morphemes from 150 monosyllabic adjectives, disyllabic adjectives and polysyllabic adjective, in the three section, we discuss the principle and method to abstract the adjectival morpheme, And then, we gain about 800 adjectival morphemes. In order to organize those adjectival morphemes, we will construct the adjectival morpheme database, in the fourth section; we discuss the device of the adjectival morpheme database. According to the different attribute of the adjectival morpheme, we will device the different fields and then fill those fields with the attribute of adjectival morphemes. At present, we had finished the adjectival morpheme database.

Keyword:

Modern Tibet, Adjective affix, adjectival morpheme, adjectival morpheme database morpheme abstraction