

# 藏语不规则动词的信息标注方法\*

江 荻

中国社会科学院 民族学与人类学所 语音学与计算语言学实验室, 北京 100081

E-mail: jiangdi@cass.org.cn

**摘 要:** 文章讨论了藏语单音动词的标注问题, 认为藏语单音动词分为有词形变化的不规则动词和无词形变化的不规则动词两类, 前者可利用词形变化表识别和标注, 后者需要建立句法识别规则加以标注。项目利用的分析工具是 Toolbox, 介绍了 Toolbox 词典中动词属性的内容, 通过词典技术以及与文本互动特点开展不规则动词的分析和标注。  
**关键词:** 藏语; 文本标注; 不规则动词; 信息内容; 电子词典

## The Tagging Methods for Tibetan Irregular Verbs in Texts

Jiang Di

Dept. of Phonetics and Computational Linguistics

Institute of Ethnology & Anthropology of Chinese Academy of Social Sciences, Beijing 100081

E-mail: jiangdi@cass.org.cn

**Abstract:** This paper discusses the issues of the annotation for Tibetan irregular verbs. All Tibetan monosyllabic verbs can be divided into two groups, one of which is called the irregular verbs with paradigms, another of which without form changes. For the first type, a tagging way is to make use of an irregular verbal table, which will help find infinitive forms. And for the second, some pre-built syntactic rules must be utilized. Furthermore, the paper introduces the constructions and contents of Toolbox dictionary the project uses, and discusses the analysis technology of mutual interactions between the dictionary and the text.

**Keywords:** Tibetan; text-annotation; irregular verbs; morphologic contents; dictionary

### 1 藏语单音动词形态变化现象

现代藏语单音节动词具有类似词形变化的特征, 有些学者称之为屈折变化<sup>[1,2,3]</sup>。传统文法认为藏语动词有时和式范畴, 每个动词都呈现现在时、未来时、过去时三种时形态, 自主动词还有命令式形式, 即传统文法所谓的“三时一式”<sup>[4]</sup>。参见表 1。

表 1 藏文三时一式变化形式

现在时	未来时	过去时	命令式	
词根	前缀+词根	前缀+词根+后缀	词根+后缀/元音变化	
སྤྲོད་srung	བསྤྲོད་b-srung	བསྤྲོད་སྤྲོད་b-srung-s	སྤྲོད་srung-s	遵守
འཇིག་v-jil	གཞིག་g-zhil	འཇིག་b-cil	ཇིག་chil	排除
ལྟོ་lta	བལྟོ་b-lta	བལྟོ་b-lta-s	ལྟོ་lto-s	看
འབད་vchad	བའབད་bshad	བའབད་bshad	འབད་shod	说
མཚོད་mchod	མཚོད་mchod	མཚོད་mchod	མཚོད་mchod	穿戴(敬)

按照形态分析, 也可以认为藏语动词受到时式范畴制约构成时式词形变化表。从词形上看, 藏语动词词形变化有 4 种基本类型: (1) 前缀与后缀变化, 例如“遵守”, (2) 词根声母变化 (可

\* 教育部—国家语委“民族语言文字规范标准建设及信息化科研资助项目”(编号: MZ115-020); 中国社会科学院重大课题(YZDA2011-18)资助。

能同时有前缀变化),例如“排除”,(3)前后缀变化或者词根声母变化同时命令式元音 a 变化为 o,例如“看”,“说”<sup>[5]</sup>。除此之外,还有相当部分动词本身并不发生形态变化,各时式形式一样(4)<sup>[3]</sup>,例如“穿戴”。

尽管现代藏语(拉萨话)已经没有典型时范畴,动词范畴的句法体现表现为体和示证,并且是以句法词方式呈现,但是动词自身仍然会根据体-示证范畴变换形式,这是所谓相对时概念<sup>[6]</sup>。例(1)的词根动词 gtong“给”在实现体(REA)条件下采用的是过去时形式 btang,动词 vdri“询问”也用了过去时形式 dris。

例(1)

- a) Ngas 'khyod kyis thugs spro ga dus btang ba yin nam' zhes dris pas .  
 b) nga gis 'khyod Gis thugs spro ga dus gtong-ed pa yin nam' zhes vdri-ed pas .  
 c) 1sg AG 2sg AG 筵席 何时 给与 -PST REA INT 说 询问-PST INTM .  
 d) 我问“你什么时候摆了筵席?”

(说明: a 是文本原句, b 是分析行, c 是标注行, d 是自由翻译行)

藏语动词时式词形变化可以类比英语动词称为不规则变化动词。根据统计,囊括以上 3 种词形变化的藏语不规则变化动词约有 759 条左右<sup>[7]</sup>,比英语不规则动词数量可能更多。

## 2 有词形变化不规则动词的标注

通常情况下,计算处理英文文本时对于不规则动词都采用形态还原方法,即建立不规则动词变化表,通过查词表获取词类信息或语义信息。藏语的不规则动词虽然有自身特点,在处理操作上基本可以仿照英语方法。

本文项目收集了藏语单音动词 2203 条,同形异义词分别单列,其中有词形变化不规则动词 1099 条。实验目的是抽取不规则单音动词词类信息、时式信息和语义信息。表中的相关信息是:词条、原形形式,词类,属性(自主/不自主,使动/自动,及物/不及物),时式(现在时,未来时,过去时,命令式),释义(汉语,英语)。

不规则动词的信息抽取主要依赖词形变化表。例如,例(2)的文本表层出现 spyad“使用”,通过查单音动词表,获取原形形式 spyod、时式信息(未来时),因此分析行表示为(原型) spyod-ing,标注行是“使用-FUT”。

例(2)

- a) gal te rgyal khab btsugs pa nas lo brgyavi ring gi dus tshod  
 b) ci ste rgyal khab vdzugs-ed pa nas lo brgya gi ring gi dus tshod  
 c) 假如 国家 建立-PST NOM ABL 年百 GEN 期间 GEN 时间  
 a) spyad de rang rgyal vdi nyid dar rgyas vbring tsam gyi rgyal khab  
 b) spyod-ing te rang rgyal vdi nyid dar rgyas vbring tsam gi rgyal khab  
 c) 使用 -FUT COR 我国 本身 繁荣 中等 稍 GEN 国家  
 a) Cig Tu vdzugs skrun byed thub na ,de ni ha las dgos pa zhis yin .  
 b) gcig La vdzugs skrun byed thub na ,de ni ha las dgos pa gcig yin .  
 c) 一 ALA 建设 能 如果,那 PAP 了不起 一 是 .  
 d) 如果从建国起,用一百年时间把我国建设成中等水平的发达国家,那就很了不起!

小句(名词化短语)中 btsugs 的原形是 vdzugs“建立;竖立;刺入;指定”,也是通过查词表获取的释义和语法信息(释义的多义项选择涉及其他问题,此处不论)。

### 3 无词形变化不规则动词的标注

如果承认藏语动词有时式变化,按照屈折形态原理,词形不发生变化的动词实际上也是不规则动词,标注的时候仍然要标注词类时式属性信息。不过,由于词表中词形同形,这类信息无法通过词形变化表获取,需要采用句法规则方法判断。

无词形变化的动词之所以也看作不规则动词有两方面原因,其一,人们总是认为词形变化是藏语动词常态,无变化动词是由于语音变化造成了所谓“三时一式”同形或合并<sup>[8]</sup>;其二,现代藏语(拉萨话)的谓语通过动词后的体貌标记或其他词语表示事件时间,多数情况下谓语所表达时间与动词内在时间(三时)词形一致,但也有不一致的时候<sup>[9]</sup>,让人们感觉动词词形与所表达时间相互矛盾,故产生不规则观念。

谓语事件时间与动词时式不一致的现象往往发生在特定情况,多数能够找到形式上的特征。以下简单讨论,举例说明谓语时间与动词时式形式不一致的现象。

(1) 过去时动词表示现在的事件。格式是“V-PST+Nom+Gen+sgang+Be-Verb”,其中Nom是名词化标记,Gen是属格标记,sgang是名词,原义为“上面”。实际上这个句子格式是判断句,其中的过去时动词只是结构上的名词化短语,但整个句子表示正在进行的事件。比较例句(3)和(4)例(3)

- |                                     |  |
|-------------------------------------|--|
| a\ nga mgo Na pavi sgang yin .      | a\ nga dpe cha bltas pavi sgang yin .    |
| b\ nga mgo na -ed pa gi sgang yin . | b\ nga dpe cha lta -ed pa gi sgang yin . |
| c\ 1sg 头痛 Nom Gen 上面 是              | c\ 1sg 书 看 -Pst Nom Gen 上面 是             |
| d\ 我头正在痛。                           | d\ 我正在看书。                                |

例(3)动词na“痛”并没有词形变化,因为na“痛”形式上已经三时同形,与相同结构的例(4)相比可以发现该形式应该是过去时动词。

(2) 过去时动词用于祈使句。祈使句式通常仅用原形动词(或现在时),但也有一些特殊结构。例如表示恭敬的请求句或劝阻句格式是:敬语动词+语气词,或者非敬语动词过去时+敬语标记nang(+语气词),劝阻句另加否定词。比较例(5),注意例(6)用的是无词形变化不规则动词。

- |                                       |                                 |
|---------------------------------------|---------------------------------|
| 例(5)                                  | 例(6)                            |
| a\ Gzhas jig btang gngang dang .      | a\ zhal lag brjed ma gngang .   |
| b\ Gzhas gcig gtong -ed gngang dang . | b\ zhal lag brjed -ed ma gngang |
| c\ 歌 一 唱 -Pst Hon Imp                 | c\ 饭 忘记 -Pst Neg Hon            |
| d\ 请唱支歌吧!                             | d\ 请不要忘了!                       |

采用过去时形式可能是避免生硬的命令式语气,这里不深究。但是,为了获得正确的标注,应该为这些特定结构建立标注规则。目前,我们已建立9项无词形变化不规则动词句法特征,更完备的特征库尚在建设之中。

### 4 动词词典库的信息内容

将2203条藏语单音动词按照时式不同形式构建词形变化库,共计词形5095条,同形异义词分别单列。现简单介绍库中动词的相关属性。

每一词条都标注了动词时式属性(tense: Fut, Pres, Pst, IMF)。动词词条可能与原形相同,例2、7、10,也可能不同,例1、3...;较为复杂的情况有,例如gcag是两个不同动词的同形未来时形式,vchag(gcag, bcags, chogs)“踩,踏”,gcog(gcag, bcag, chog)“打破,砸碎”;其他信息分别是动词原形形式(Ori);及物与否(vt/-);使动与否(Cas/-);自主与否(vl/-);中文释义(Chn);句例(Sent)。<sup>[10]</sup>

No	Form	Tense	Ori	vt/-	Cas/-	VI/-	Chn	Sent.
1	dgang	Fut	vgengs	+	+	+	装满	bang mdzod vbrus bkang ba 粮食装满仓
2	dgongs	Pres	dgongs	+	HON	+	考虑	thugs la de ltar dgongs te gsungs pa red 心里那样想了才说的
3	dgug	Fut	vgugs	+		+	召引	dmag mi vgugs pa 调兵 mgron vgugs pa 延宾
4	gcag	Fut	vchag	+		+	踩	phyi rgyal gyi sa chavi stef du vchag pa 踏上外国的土地
5	bcags	Pst	vchag	+		+	踩	zhabs kyis bcags pavi gnas 曾经到过的地方
6	gcag	Fut	gcog	+	+	+	打破	rdo gcog pa 打碎石头
7	gtug	Pres	gtug	+		+	会见	nga sman khang la em rje btug par phyin pa yin. 我去了医院见大夫。
8	gtugs	Pst	gtug	+		+	会见	dge ngan gyi zhabs la gtugs pa 亲近老师
9	brngams	Pst	mgam	-		-	贪念	khrag la brngams pavi dmag pung 嗜血部队
10	mched	Pst	mched	-		-	蔓延	rims tshad mched pa 瘟疫流行

动词原形的作用特殊，如果文本表层的动词词形与之不一致，则从时式属性提取表层形式的信息，分析或还原为“原形+时式属性 (-ing, -ed, -!<折返> )”，并标注为“释义+时式标记 (-Fut, -Pst, -IMF)”。

## 5 结语

本项目是利用 Toolbox 隔行对照化工具开展的藏语文本句法规则研究，该工具的重要特点是：文本与词典互动，利用词典对文本进行分析和标注。本文正是利用这项特点，构建了适合藏语分词和形态分析的词典技术，包括深层与表层技术来解决不规则变化动词的分析。该工具的不足之处是未提供识别率或召回率的统计分析，不能实现量化研究。

这篇文章涉及的只是藏语动词一个细微方面，有关藏语动词其他研究内容，例如句法语义分类的处理、Toolbox 词典深层与表层形式设置等，作者曾经另文介绍<sup>[11]</sup>。希望本文对研究藏语动词文本处理的同行有所助益。

## 参考文献

- [1] Csoma de Kőrös, Alexander. 1834. A Grammar of the Tibetan Language. Calcutta: the Baptist Mission Press.
- [2] Wolfenden, Stuart N. 1929. *Outlines of Tibeto-Burman linguistic morphology*. London: Royal Asiatic Society.
- [3] Li, Fang-Kuei 1933. Certain Phonetic Influences of the Tibetan Prefixes upon the Root Initials. *Bulletin of the Institute of History and Philology* 6.2: 135-157.
- [4] 图弥三菩札等著，黄明信编：《西藏文法四种合编》（藏文），北京，民族出版社，1956年
- [5] 江获. 1991. 藏语动词的历史形态研究. 《中国藏学》1991年第1期
- [6] Zeisler, B. 2004 *Relative Tense and Aspectual Values in Tibetan Languages: A Comparative Study*. Berlin: Mouton de Gruyter.
- [7] 江获. 1992. 藏语动词屈折现象的统计分析. 《民族语文》1992年第4期
- [8] 胡书津. 2002. 藏语动词类型释要. 成都：四川民族出版社
- [9] 周季文，谢后芳. 2003. 藏语拉萨话语法. 北京：民族出版社
- [10] Jiang Di. et. 2005. The Verbal Entries and Their Description in a Grammatical Information- Dictionary of Contemporary Tibetan. In Robert Dale, Kam-Fai Wong, Jian Su, Oi Yee Kwong (Eds), *Natural Language Processing – IJCNLP 2005*, 874-884. Springer.
- [11] 江获. 2006. 藏语语法词典的构建方法：词条与文本互动，载：曹右琦，孙茂松（主编），中文信息处理前沿进展—中国中文信息学会二十五周年学术会议，第107-113页。北京：清华大学出版社