

基于格关系和配价的藏语动词再分类研究*

陈玉忠 李保利 俞士汶

北京大学计算机科学技术系 计算语言学研究所 北京 100871

E-mail: {degai, libl,yusw}@pku.edu.cn

摘 要: 本文着重介绍了我们根据研制藏汉机器翻译系统的实际需要, 在采用传统语法对藏语动词的已有分类基础上, 采纳格语法和配价理论的合理思想, 结合藏语的格接续特征, 通过每个动词的配价信息以及所能携带的相应的格助词类型(格标)及其数量来对藏语动词进行再分类, 从而形成一个集语法语义为一体的藏语动词再分类框架。

关键词: 藏语动词, 格关系, 配价理论

Re-Study on Tibetan Verb Categorization Based on Case-Relation and Valence Theory

Chen Yuzhong, Li Baoli, Yu Shiwen

Department of Computer Science and Technology, Peking University, 100871

Abstract: To develop a practical Tibetan-Chinese machine translation system, we propose a new scheme to classify Tibetan verbs. Valence theory and case grammar are combined in our approach. Each Tibetan verb is classified according to its valence information, type and number of possible case-auxiliary words accompanied, and Tibetan continuous features. In other words, not only syntactic feature but also semantic information is considered while classifying Tibetan verbs in this scheme.

Keywords: Tibetan Verb; Case-Relation; Valence Theory

一、引言

动词是句子的核心, 在整个句子中占据着支配的地位, 决定着全句的格局, 因此研究动词的语法语义范畴就自然成为了语法研究的一个重要方向。迄今为止, 研究动词语法语义范

* 本文研究工作得到国家自然科学基金项目(合同号: 69663001)和973项目(合同号: G1998030507-4)资助, 特此致谢。

畴的方法有各种各样并取得了不少成果,其中采用配价语法和格语法进行动词语法语义范畴研究取得的成果最为显著,而且这两种理论还在语言信息处理中得到了广泛的研究和应用[1][2][3][4]。但是,笔者还未曾见到利用配价理论和格语法对藏语动词进行再分类的相关报道,有鉴于此,我们在充分借鉴前辈、时贤的研究成果基础上,尝试着引入格语法和配价理论开展藏语动词的再分类研究,这项工作的开展,无论对汉藏机器翻译系统的研制还是对推动藏语信息处理的理论研究和应用研究都是非常有益的。

针对藏语信息处理的需要,具体在如何借鉴(如何取舍)已有成果和如何引入(如何结合)格语法和配价理论的问题上,我们所确定的基本研究思路如下:在传统藏语语法以格助词和动词为核心的理论框架指导下,尽可能地采用传统语法对藏语动词的已有分类;在此基础上采纳格语法和配价理论的合理思想,结合藏语的格接续特征,通过每个动词的配价信息以及所能携带的相应的格助词类型(格标)及其数量来对藏语动词进行再分类,从而形成一个集语法语义为一体的藏语动词再分类框架。

二、传统分类法简介

从传统语法对藏语动词的分类来看,主要包括:从动宾关系出发,分出及物动词与不及物动词;从动主关系出发,分出自动动词与不自动动词;从动词与主宾关系出发,分出使动动词与自动动词。对这6类动词采用二元属性描述法则在理论上可划分出8种类型,见表1。其中,(*)表示该项内容为空,下同。

表1 传统语法对藏语动词的分类

序号	及物	自主	使动	词例	参考译文
(1)	+	-	-	མཁྱེད་	(看见)
(2)	+	+	-	ཟ་	(吃)
(3)	+	-	+	(*)	(*)
(4)	+	+	+	བརྒྱུད་	(改变)
(5)	-	+	-	འཕྱུད་	(飞)
(6)	-	+	+	(*)	(*)
(7)	-	-	+	(*)	(*)
(8)	-	-	-	ཐུང་	(逃脱)
(8)'				ཞི་	(死)

由于表1中的几种动词范畴是从不同的角度划分的,它们之间的关系自然不是截然对立的,而是相互交错的。所以,我们实际上只得到了5种类型的藏语动词。即及物、不自主、自动动词(1),及物、自主、自动动词(2),及物、自主、使动动词(4),不及物、自主、自动动词(5),不及物、不自主、自动动词(8)。同时,我们从表1中也可以看出,所有动词均可以划入及物与不及物、自主与不自主这两个范畴,但使动与自动范畴则不可以。

三、 配价方法的引入

从配价理论来看，藏语动词应该有 3、2、1、0 价动词，但是如何确定一个藏语动词具体支配几个名词性成分却是一件比较困难的事情。譬如，藏语动词མཛད (交给) 应该是一个三价动词，不过，从藏语句子来看，在没有任何条件限制的情况下，我们用这个三价动词很自然地能够造出一个完全正确但有点啰嗦的有 4 价的句子。如：

ངས་ལྷོ་ལ་མ་དཔེ་ཆ་དམར་པ་ཞིག་ལག་ལམ་གྱིས་མཛད། (我用右手交给卓玛一本红书)。理论上我们是不是还可以造的更长一些呢？似乎不难办到。因此，如何确定藏语动词的配价、进而如何将传统语法的动词分类与配价理论结合起来是进行藏语动词再分类要解决的关键问题，而制订一个可操作的确定动词配价指数的标准则是首要问题。针对藏语句子的特点，我们提出确定藏语动词配价指数的标准如下：

- 1) 在藏语基本句[†]中提取配价指数；
- 2) 单句中若包含格助词，应在由单个格助词或虚词所引导的最多的名词性成分中确定配价指数；
- 3) 藏语动词的配价指数最高为 3。

由这一确定配价指数的标准，可以得到藏语动词有 3、2、1、0 价四类动词。那么，这四类动词具体与上述传统语法中的 5 类动词之间有着什么样的关系呢？为了说明这一问题，我们依据以上标准，就对前面 5 个动词的具体配价指数做一番实际考察。以下例句中，下划线部分为名词性成分。

及物、不自主、自动动词 (1) “མཛོང” 在基本句中由单个格助词所引导的最多的名词性成分为 2，其配价指数为 2。如：ངས་གསེར་གྱི་བུ་མ་པ་ཞིག་མཛོང། (我看见了一只金碗)[‡]。

及物、自主、自动动词 (2) “བ” 在基本句中由单个格助词所引导的最多的名词性成分为 3，其配价指数为 3。如：ངས་བུར་མས་བ་བ། (我用筷子吃饭)。

及物、自主、使动词 (4) “བསྐྱར” 在基本句中由单个格助词所引导的最多的名词性成分为 3，其配价指数为 3。如：ངས་ཤིང་གྱིས་འཁོར་མའི་ཁུའྱག་བསྐྱར། (我用力改变轮子的方向)。

不及物、自主、自动动词 (5) “འཕུར” 在基本句中由单个格助词所引导的最多的名词性成分为 2，其配价指数为 2。如：གན་མ་ནས་བྱ་འཕུར། (鸟在空中飞翔)。

[†] 基本句是指只能包含一类显性的格助词或提示助词的单句。

[‡] 为便于排版用图形格式表示的藏文字母，以下我们尽可能地采用同一类藏文词和句子来举例说明。

不及物、不自主、自动动词(8)“ཐར་”在基本句中由单个格助词所引导的最多的名词性成分为2,其配价指数为2。如:ཇི་བོད་ཀྱི་ལས་ཐར་། (兔子从网中逃脱)。

不及物、不自主、自动动词(8)’,“གི་”在基本句中由单个格助词所引导的最多的名词性成分为1,其配价指数为1。如:མི་དམངས་ཀྱི་དོན་དུ་གི། (为了人民的事业而死)。

由此可见,按配价理论划分的3价、2价和1价动词与传统语法分类中的5类动词有对应关系,而且是单射关系。但0价动词在传统语法中没有对应成分。也就是说,传统语法的分类细致但分类所涵盖的动词类别不全。而按配价理论的分类,涵盖了所有的动词类别但分类相对要粗一些。

那么,传统语法中未分类(未讨论)的动词有那几类?按配价理论划分出来的0价动词具体对应藏语中的那一类或哪几类动词呢?实际上,藏语动词从有无形态变化的角度可以划分为两大类,传统藏语语法的分类主要针对有形态变化的一类,而对无形态变化的动词类未作详细的分类研究。由现代藏语词语分类体系中所划分的基本词类可知,判断动词和助动词就是没有形态变化的两大类动词。我们可以利用配价标准来确定这两类动词以及0价动词的配价指数,归入到相应的类中去,进而形成一个有配价信息的完整的藏语动词分类框架。如:

判断动词“ཡོད་”在基本句中由单个格助词所引导的最多的名词性成分为2,所以其配价指数为2。实例如:ལྷ་ས་ན་མངའ་ལ་དུ་ཡོད། (拉萨有哈达)。

判断动词“ཡིན་”在基本句中由单个虚词所引导的最多的名词性成分为2,所以其配价指数为2。实例如:ང་ནི་སྐྱབས་ཡིན། (我是学生)。

判断动词“དཔེར་ན།”在一个基本句中不能带名词性成分,所以其配价指数为0。实例为:དཔེར་ན། (例如)。

综上所述,在传统语法对动词分类的基础上,引入配价理论来对无形态变化的藏语动词再分类是对传统语法的补充和完善,也是非常有实用价值的。

四、配价与格助词的有机结合

最后,再来看看如何将各类格助词与藏语动词上述的语法语义范畴直接挂起钩来进行再分类。瞿霭堂教授曾经根据藏语动词与主语形式上的一致关系,分出三类动词:主语加施动格助词的及物动词,主语加存在助词的存在动词,主语不加助词的不及物动词和判断动词,

判断动词又根据能带表语和无体的变化两个特点同不及物动词区别开来。关于动词与格助词的关系问题上，我们在此基础上作进一步的延伸：直接考察动词与格助词的类型、数量之间的关系，其基本立足点是：每一类（或每一个）动词都可能有用格助词和禁用格助词（与动词有关的格助词中不包括属格助词），格助词的类型因动词而已，就特定的动词而言格助词的类型和数量是固定的。根据这一思想，我们仍以上述的几个动词为例来进行考察。例句中的下划线部分为格助词或助词。

2 价的及物、不自主、自动动词 (1) “མཐོང་” 在一个基本句中最多能带 2 个格助词，而能带的格助词分别是作格助词和 LA 格助词，禁用格助词则是从格助词。如：

མཇུ་ལ་ངེས་གསེར་གྱི་བུ་མ་ལ་ཞིག་མཐོང་། (上午我看见了一只金碗)。

3 价的及物、自主、自动动词 (2) “བ” 在一个基本句中最多能带 2 个格助词，分别是作格助词和 LA 格助词，禁用格助词为从格助词。如：
བཟུང་བྱེད་ཀྱི་ལུ་ལྟོ་ལྟོ་ལྟོ་བཟུང་བཟུང་། (我用筷子吃饭)。

3 价的及物、自主、使动词 (4) “བསྐྱེད་” 在一个基本句中最多能带 3 个格助词，分别是作格助词、LA 格助词和从格助词，没有禁用格助词。如：
ངེས་མེད་གྱི་ལུ་ལྟོ་ལྟོ་ལྟོ་ལྟོ་ལྟོ་ལྟོ་བསྐྱེད་བསྐྱེད་བསྐྱེད་། (我用力改变轮子的方向)。

2 价的不及物、自主、自动动词 (5) “འབྱུང་” 在一个基本句中最多能带 2 个格助词，分别是 LA 格助词和从格助词，禁用格助词是作格助词。如：
བྱ་ལྟོ་ལྟོ་ལྟོ་ལྟོ་ལྟོ་ལྟོ་འབྱུང་། (飞机从北京飞向拉萨)。

2 价的不及物、不自主、自动动词 (8) “ཐར་” 在一个基本句中最多能带 2 个格助词，分别是 LA 格助词和从格助词，禁用格助词也是作格助词。如：
ནགས་ཚལ་དུ་རི་བོ་ལྟོ་ལྟོ་ལྟོ་ཐར་། (森林里一只兔子从网中逃脱)。

1 价的不及物、不自主、自动动词 (8) “གླིང་” 在一个基本句中最多能带 1 个格助词，是 LA 格助词，而作格助词和从格助词是禁用格助词。如：
མི་དམངས་ཀྱི་དོན་དུ་གླིང་། (为了人民的事业而死)。

2 价判断词 “ཡོད་” 在一个基本句中最多能带 1 个格助词，即 LA 格助词，禁用格助词是作格助词和从格助词。如：
ལྷ་ས་ནུ་མཇུ་ལ་དུ་ཡོད་། (拉萨有哈达)。

2 价判断词 “ཡིན” 在一个基本句中不能带格助词，禁用格助词也就是作格助词、LA 格助词和从格助词，但可以带一个提示助词。如：
ང་ལྟེ་སྲིབ་པ་ཡིན། (我是学生)。

0 价判断词 “དཔེར་ན།” 在一个基本句中不能带格助词，禁用格助词也就是作格助词、LA 格助词和从格助词，也不能带任何其他助词。实例如：
དཔེར་ན། (例如)。

五、 分类结果和讨论

通过总结和整理以上三种分类方法所得到的结果，我们可以得到包括 9 个子类的藏语动词再分类框架，如下表 2 所示。

表 2 基于格关系和配价的藏语动词分类框架

配价分类	传统分类	格助词分类	分类框架	词例	参考译文
3 价动词	及物、自主、使自动动词(4)	作格助词 LA 格助词 从格助词	I	བརྒྱུར	(改变)
	及物、自主、自动动词(2)	作格助词 LA 格助词	II	ཟ	(吃)
2 价动词	及物、不自主、自动动词(1)	作格助词 LA 格助词	III	མཐོང	(看见)
	不及物、自主、自动动词(5)	LA 格助词 从格助词	IV	འཕུར	(飞)
	不及物、不自主、自动动词(8)	LA 格助词 从格助词	V	ཐར	(逃脱)
	(*)	LA 格助词	VI	ཡོད	(有)
	(*)	提示助词	VII	ཡིན	(是)
1 价动词	不及物、不自主、自动动词(8)'	LA 格助词	VIII	ཤི	(死)
0 价动词	(*)	(*)	IX	དཔེར་ན	(例如)

关于表 2，需要作以下三点补充说明：

(1) 这一分类的目的很明确，概括起来主要有二：一是希望得到一个藏语动词的再分类框架，二是要考察配价与格助词之间的内在联系或者说在二者之间建立一种直接联系。对于第一点而言，我们得到的动词细类 I、II、III、IV、V、VI、VII、VIII、IX 就构成了一个动词再分类框架。其含义其实已蕴含在三种具体分类方法之中，具体的名称我们暂时可以结合

三种分类方法来命名。如二价及物、不自主、自动且可携带作格和位格助词的动词，简称为Ⅲ类动词等等。关于第二点，根据目前的研究，我们可得到如下初步的认识：由价（单个动词）来考察格助词，在基本句内，3价动词可携带作格助词，0、1、2价动词则基本不能携带作格助词（Ⅲ类动词例外）；在一个完整句内考察格助词，3价动词可携带三个格助词，2价动词可携带两个格助词（Ⅳ、Ⅴ类动词例外），1价动词可携带1个格助词，0价动词不能携带格助词。反过来，由格助词来考察（单个动词）价，同样可得到上述结论。

（2）从纯分类的角度来讲，分类的基本要求是不杂（标准统一）、不重（互相排斥）和不漏（对象的全集）。由以上讨论可知，事实上，传统的分类方法本质上可以看作是一种采用格关系分类方法。就这种分类方法本身而言，不杂是做到了，但在不重和不漏方面还不能说完全达到了要求，特别是在不漏方面确实没有做到。对这一点，虽然我们一方面通过只采用及物和不及物、自主和不自主两个范畴，另一方面增加一个“其他动词”的方法来满足这一要求，但这样的分类标准毕竟不是传统意义上的分类标准了。而且，在没有可靠的理论依据和扎实的实践检验的基础上，草率地做这样的修改反而不一定适合于藏语的实际。所以，在我们的再分类框架中，基础部分的分类还是沿用了传统的分类标准。从严格的意义上来说，格助词的分类方法其实不是一个独立的分类标准，他只是配价和传统分类方法的一种统一的形式表现，或者说是把上述两种分类标准有机地联系起来的一种形式标记系统。一个动词所能带的格的数目是配价语法着力研究的主题，而一个动词所能带的从属名词的语义角色正是格语法着力研究的主题[5][6]。通过格助词把动词的配价属性同格关系显性地联系起来，而这有助于计算机进行自动句法语义分析。配价的分类方法应该说是完全满足了分类标准的基本要求，但实事求是地说，虽然我们提出了一个可操作的提取动词配价指数的标准，由于标注几万条动词的配价信息是一项不小的工程，具体实施起来必须要慎重，在理论上作进一步的完善是很有必要的。因此，我们已有的藏语电子词典也好，汉藏英电子词典也好，目前只描述了按传统分类法所划分出来的子类信息和主要格助词信息，对动词的配价信息还未作详细描述。而本项研究可以说是为进一步开展这项工作所做的理论探讨。

（3）坦率地讲，以上藏语动词的分类还是个尝试性的分类框架，特别是在传统语法对动词分类的基础上，结合格助词和配价语法的理论对藏语动词的再分类更是一个探索性的研究工作。但我们坚信，这一分类对藏语信息处理是非常有价值的，而且根据这一基本思路还可以依据格助词更详细的语义特征对藏语动词作更为细致的分类研究。同时，在此基础上仔细地探究动词、格助词和动词配价三者之间深层的语义关系，有助于建立面向信息处理的更为细致的藏语动词的语义分类模式，而这又是一项非常有实用价值的课题。

参 考 文 献

- [1] 林杏光：“基于格关系的现代汉语述语动词分类系统”，《计算语言学进展与应用》，陈力为、袁琦主编，清华大学出版社，1995年10月。
- [2] 冯志伟：“自然语言机器翻译新论”，语文出版社，1994年10月。
- [3] 沈阳，郑定欧：“现代汉语配价语法研究”，北京大学出版社，1995年8月。
- [4] 袁毓林，郭锐：“现代汉语配价语法研究”，北京大学出版社，1998年7月。
- [5] 袁毓林：“汉语动词的配价研究”，江西教育出版社，1998年8月。
- [6] 瞿霭堂：“论藏语的语法体系”，载《藏族学术讨论会论文集》，西藏人民出版社，1984年9月。