



中华人民共和国国家标准

GB/T 36452—2018

信息处理用藏文分词规范

Specification on Tibetan segmentation for information processing

2018-06-07 发布

2019-01-01 实施

国家市场监督管理总局 发布
中国国家标准化管理委员会

目 次

前言 I

引言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 分词规范 1

参考文献 15

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本标准起草单位：中国电子技术标准化研究院、西藏大学、西北民族大学、西藏自治区藏语文工作委员会办公室、青海师范大学、青海民族大学、中国科学院软件研究所、西藏自治区工业和信息化厅。

本标准主要起草人：扎西加、欧珠、尼玛扎西、熊涛、格桑多吉、多拉、拉巴泽仁、大罗桑朗杰、高定国、拉琼、仁青诺布、索南尖措、旺堆、小尼玛扎西、普次仁、顿珠次仁、赵栋材、边巴嘉措。

引 言

本标准以现代藏语的词类和分词研究成果为基础,根据藏文词汇特点与构词规律,并参考汉语分词及词类标记相关标准(见参考文献)的部分内容,规定了信息处理用藏文分词规范。

信息处理用藏文分词规范

1 范围

本标准规定了信息处理用藏文分词规范。
本标准适用于藏文信息处理各领域,其他行业和有关学科可参照使用。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。
GB/T 36337—2018 信息处理用藏语词类标记集

3 术语和定义

下列术语和定义适用于本文件。

3.1

词 word

区别事物意义的最小的语法单位。

3.2

词组 phrase

两个或更多词组合成的语言单位。

注:词组可以是实词与实词的组合,也可以是实词和虚词的组合。

3.3

藏文信息处理 Tibetan information processing; TIP

用计算机对藏文的音、形、义等信息进行处理。

3.4

分词单位 segment unit

在分词过程中出现的词。

注:分词单位不仅限于语法词,其中也包含了信息处理所需的一部分结合紧密、使用稳定的词组。

3.5

藏文分词 Tibetan segmentation

将连续的藏文音节序列按照一定的规范重新组合词序列的过程。

4 分词规范

4.1 藏文分词单位和词类的标记

本标准以“/”作为藏文分词单位的标记,藏语词类标记依据 GB/T 36337—2018 的规定。

4.2 一般名词 <མྱུར་བཏང་གི་མིང་།>(nn)

4.2.1 由单音节名词和单音节形容词组成的词为一个分词单位。

示例：

དབང་ཚུང་/ ལ་ཚུང་/ ར་ཚུང་/ ཞང་ཚུང་/ ཞང་ཚུང་/ ལམ་ཆེན་/ མི་ཆེན་/ དབང་ཆེན་/

4.2.2 由双音节名词和单音节形容词组成的三音节名词为一个分词单位。

示例：

སྒ་བོ་ཆེ་/ མི་བོ་ཆེ་/ མདུང་མོ་ཆེ་/ ར་བོ་ཆེ་/ དར་མོ་ཆེ་/ ལྷགས་པོ་ཆེ་/ ལ་བོ་ཆེ་/

4.2.3 由名词性分词单位和专职词“བ། བོ། བ། བོ། བ། བོ།”组成的名词为一个分词单位。

示例：

སྒོ་མ་ཆེན་བ། ཞང་བ། བདག་བོ། རོང་བ། རྩ་བོ། ར་མ། གནས་མོ། རྩམ་བ་བོ། རྩལ་མོ། དྲིགས་ལྷན་མ། རྟིང་མ་བ། རྩལ་མ་བ།
གངས་ཅན་བ།

4.2.4 名词性的构词后缀“ཅན། ལན། མཁན།”不单独切分。

示例：

ལག་ཤེས་ཅན་/ དཔྱད་ལན་/ ཡོན་ཏན་ཅན་/ ར་མཁན་/ ལྷ་བཙོང་མཁན་/

4.2.5 四音节构成的复合词为一个分词单位。

示例：

ཟག་པ་ཀླན་ཟད་/ རྩ་གར་ཏིག་ཏ་/ རྩ་ཡོན་ལྷན་ཁང་/ དབྱང་ཉིས་ཐན་འགྱུར་/ གྲང་ཁན་རིང་ལྷགས་/ ལ་ཆེ་གྲུར་གྲུམ་/

4.2.6 五音节构成的复合词为一个分词单位。

示例：

བྱ་སྔོན་ཕུང་པོ་རི་/ མི་མཐེད་འཛིག་རྟེན་ཁམས་/ པོ་རྟི་ལྷགས་ལུང་མ་/ ལྷང་ཀོང་ཕྱག་རྩ་བ་/

4.2.7 三音节构成的固定词组为一个分词单位。

示例：

རྩུ་ལོངས་སྤྱོད་/ ལྷག་དམར་ཡག་/ ར་ཁ་ཡག་/ འབྲང་བྲེ་མེར་/ ཏ་འདོ་བ་/ མཛེ་ཁྱུང་དཀར་/

4.2.8 四音节构成的固定词组为一个分词单位。

示例：

དབྱར་དགུན་སྟོན་དཔྱད་/ རྩེ་ག་ན་འཆི་/ ཞི་རྩུས་དབང་གྲགས་/ ཆགས་གནས་འཛིག་སྟོང་/ དཀར་མེར་དམར་ལྷང་/ འཆད་ཉན་སྟོན་སྤྱབ་/
ལེགས་ཉེས་ལྷང་དོར་/ ལྷ་སྤྱལ་ཡང་སྤྱལ་/ པ་མེས་ཡང་མེས་/

4.2.9 五音节及以上音节构成的固定词组为一个分词单位。

示例：

རུབ་སྤང་བ་མཐའ་ཡས་/ བསམ་པ་རྩུན་གྲུབ་མ་/ ཞིང་ཁམས་རབ་འབྱམས་རྩུ་མཆོ་/ ལྷེབ་སྦྱར་རིན་ཆེན་འབྱུང་གནས་/ དག་ཡིག་
མཁས་པ་དགའ་སྦྱོང་/

4.3 人名<མིའི་མིང་།>(nr)

4.3.1 不易区分姓和名的笔名通常作为一个分词单位。

示例：

ཙང་ལ་བ་/ འཇུ་མི་པམ་/ བོད་མཁས་པ་/ རྩ་སྤྱིས་དགེ་བཤེས་/ འབྲོམ་སྟོན་པ་/

4.3.2 人名的译名按其习惯形式作为一个分词单位。

示例：

ཨ་དབྱིན་ས་ཐེན་/ མར་ཁི་སི་/ ལུ་ལུ་ལུ་/ པ་ཅེན་/ ཆའོ་ཞོ་ཆེན་/ ཏུ་ཏུ་ལི་/

4.3.3 人名(包括带格助词的人名)作为一个分词单位。

示例：

ཙང་ཁ་པ/ འབྲུ་མི་པམ/ བོད་མཁས་པ/ རྫོང་མི་དགེ་བཤེས/ འབྲོམ་སྟོན་པ/ གཟུགས་ཀྱི་ཉི་མ/ ཀྲན་ཏུ་བཟང་པོ/ ཆོས་ཀྱི་རྒྱལ་མཚན/ སྤང་ཅིའི་སློ་གྲོས/

4.4 地名<ས་རྒྱུའི་མིང་།>(ns)

4.4.1 山丘河流等名称作为一个分词单位。

示例：

དམར་པོ་ཟེ/ ལྷགས་པོ་ཟེ/ འབའ་ཆུ/ ཟུང་ཆུ/ མང་ཆུ/ རྟོ་རྫོག་གཙང་པོ/ ཡར་ལྷང་གཙང་པོ/ ལུ་ཅུ་རི་བོ/ གདང་ ལ་རི་བོ/ ཟེ་ཕུ་ལོ་ཟེ་སྤང་ཕྱན/

4.4.2 地名后有表示自然区划的普通名词，如“ལམ”等，作为一个分词单位。

示例：

རྫོ་སེང་གེ་ལམ/ སེ་ར་ལམ/ ཉང་བྲན་ལམ/ གཙང་བྲན་ལམ/ ལྷ་སྤྲིངས་ལམ/

4.4.3 地址叙述中的大小地名分别作为一个分词单位，如其中地名或路名是一个音节则不予切分。

示例：

སེ་ར་ལམ/ སློ་རྟགས་/ ཡུཾ/ པ/ ང་རྒྱག་པ

4.5 国名<རྒྱལ་ཁབ་ཀྱི་མིང་།>(ng)

国名不论长短，作为一个分词单位。

示例：

ཀྲང་གོ/ འཇར་པན/ ཨུ་རུ་སུ/

4.6 族群名<རིགས་རྒྱུད་ཚོ་བའི་མིང་།>(ne)

4.6.1 表示民族、人种的词作为一个分词单位。

示例：

བོད་རིགས་/ རྒྱ་རིགས་/ དབྱི་རིགས་/ ཉ་སག་རིགས་/ ཡུ་གུར་རིགས་/ བོད་མི/ རྒྱ་མི/ ཉན་མི/

4.6.2 表示部落、部族的词作为一个分词单位。

示例：

མགར་རྩེ་ཆོ་བ/ ཤར་པ/

4.6.3 表示文种、语种的词作为一个分词单位。

示例：

བོད་ཡིག/ དབྱིན་ཡིག/ རྒྱ་ཡིག/ སྐག་སྐད/ བོད་སྐད/ དབྱིན་སྐད/ ཨུ་རུ་སུའི་སྐད/ འབར་མའི་སྐད/ འབྲུག་པའི་སྐད/ བོད་ཚུ་/ རྒྱ་ཚུ་/

4.7 团体机构名<ལས་ཁུངས་ཀྱི་མིང་།>(na)

4.7.1 行政区划称谓中两个音节及以上做切分，地名或行政区划名如其中一个是单音节，则不切分，将地名与行政区划名合起来作为一个词。

示例：

བོད་རང་སྐྱོང་ལྗོངས་/ མཚོ་ལྷོ་རྫོང་/ མཚོ་སྤོ་ཁུལ་/ རྟག་ཆུ་ས་ཁུལ་/ རྒྱ་ཆ་རྫོང་/ སུམ་མདོ་ཤང་/ གོང་དཀར་རྫོང་/ བོད་མདོ་/ གོང་ཆོ་/ མང་རྫོང་/ སྐག་རྫོང་/ འབའ་རྫོང་/

4.7.2 团体机构名称作为一个分词单位。

示例：

མཉམ་འབྲེལ་རྒྱལ་ཚོགས་/ གྲུང་གུང་གྲུང་དབྱེང་/ མི་བཞི་ཤོག་ཁག་/ རྒྱ་གཞུང་དམར་པོའི་ཚོགས་པ་/ བོད་མྱོང་ས་སློབ་གྲྭ་ཆེན་མོ་
བོད་ མྱོང་ས་དཔེ་མཛོད་ཁང་/

4.8 专有名词〈ཆེད་བཏགས་ཀྱི་མིང་།〉(nz)

专有名词作为一个分词单位。

示例：

སྐྱལ་ཁོའི་ཞེ་གཟུར་/ འབྲི་གུང་སྤིང་ལོག་ ར་སྐྱལ་དོན་རྒྱུན་/ གོར་བོད་དམག་འབྲུག་/ ལྷ་བཞིའི་ལས་འགུལ་/ ལྷགས་པག་ལོའི་གསར་བཞེ་
དབྱིན་འགོག་དམག་འབྲུག་/ རྒྱ་རྒྱལ་གཏམ་རྒྱུད་/ མི་ལ་རས་པའི་རྣམ་ཐར་/ ཉིན་རེའི་ཆགས་པར་/

4.9 辞藻〈མདོན་བརྗོད།〉(no)

辞藻作为一个分词单位。

示例：

བརྗོད་པའི་འཁོར་ལོ་/ ཏྲ་བདུན་དབང་པོ་ / ཉིན་བྱེད་བདག་པོ་/ གཞི་སྒྲོན་སྐྱེས་/ རྒྱ་གཏེར་/ འདོད་ཡོན་/ གཟུགས་ཅན་/ བསེ་བྱ་
མི་མཁའ་རྒྱ་མཚོ་/

4.10 数称词〈གངས་འདྲེན།〉(nm)

数称词作为一个分词单位。

示例：

ཚོགས་གཉིས་/ ལྷགས་གཉིས་/ ཆོས་སྤྱིད་གཉིས་/ ས་བརྒྱ་ དཔལ་ཡོན་གཉིས་/ བདེན་པ་བཞི་/ དཀོན་མཆོག་གསུམ་/ མཚོན་
བྱེད་གསུམ་/ རྒྱལ་བ་རིགས་ལྔ་/ རྣམ་ཤེས་ཚོགས་དྲུག་/ འཁོར་ལོ་རྩིབས་བརྒྱད་/ བད་མ་འདབ་བརྒྱད་/ དབུ་མ་བཞི་བརྒྱ་པ་/ རྩི་བརྗོད་
སུམ་ཅུ་མ་/ སྤྱོད་མངའ་རིས་སྒྲོར་གསུམ་/ བར་དབུས་གཙང་ཅུ་བཞི་/ དགུང་ཉི་ཟླ་དཀར་གསུམ་/ དཔལ་ལྷན་ས་གསུམ་མ་/ དགའ་
ལྷན་ལྔ་བརྒྱ་མ་/ རྣད་རིགས་བཞི་བརྒྱ་ཅུ་བཞི་/ དགག་རིགས་སྤོང་ཕྱག་བརྒྱད་ཅུ་/ ཆོས་རྒྱལ་མེས་དཔོན་རྣམ་གསུམ་/

4.11 处所方位词〈སྤྱོད་གས་གནས་ཀྱི་མིང་།〉(nf)

处所词方位词为一个分词单位。

示例：

ཕྱི་ལོག་/ རྣང་ལོག་/ ཕྱི་སོ་ རྣང་སོ་/ ཉེ་འགམ་/ མཐའ་འཁོར་/ ཤར་/ རྩོ་/ རྩུབ་/ བྱང་/ རྩེང་/ གཡས་/ གཡོན་/ རྒྱབ་/ མདུན་/ བཞུག་/ གཡས་གཡོན་/

4.12 时间词〈དུས་སྟོན་པའི་མིང་།〉(nt)

4.12.1 历史朝代的时间词为一个分词单位。

示例：

གུང་རྒྱལ་རབས་/ མན་རྒྱལ་རབས་/ ཡོན་རྒྱལ་རབས་/

4.12.2 季节、月份的时间词为一个分词单位。

示例：

དབྱར་ལ་/ དུས་ལ་/ སྟོན་ལ་/ དབྱིད་ཀ་/ ཟླ་བ་དང་པོ་/ ཟླ་བ་བརྒྱ་གཉིས་པ་/ གནམ་གང་/

4.12.3 民俗节庆的时间词为一个分词单位。

示例：

ས་ག་ཟླ་བ་/ ལ་མཚན་/ ཟླ་བ་མཚན་/ སྐར་མ་རེ་ཤིང་/ དཔལ་ལྷན་པ་མཚན་/

4.12.4 每周的七天、一天的时辰、一般的时期等为一个分词单位。

示例：

གཟའ་ཉི་མ་/ ས་ཟླ་བ་/ ལུ་མ་/ ཉིན་གུང་/ ཟྱ་ཏྲོ་/ ལ་སང་/ དེ་རིང་/ སང་ཉིན་/ གནང་ཉིན་/ ན་ནིང་/ གཞིས་ནིང་/ ད་ལོ་/ སང་ལོ་/ ད་ཉི་/ མ་འོངས་པ་/

4.12.5 “年、月、日、时、分、秒”分别为一个分词单位。

示例：

ཟྱི་ལོ་/ 2010/ལོ་/ ཆོས་/ ཆོས་པ་/ 10/ སྐར་མ་/ 11/ སྐར་ཆ་/ 11/ ཟྱི་ལོ་/ ཉིན་ལྟར་བཅུ་/

4.13 动名词<བྱ་ལས་གྱི་མིང་།>(nv)

动名词为一个分词单位。

示例：

ཉན་སྒྲོར་/ གཙོ་སྒྱུར་/ གྲོས་སྐྱུར་/ མཚུ་བ་སྒྱུར་/ རེ་སྒྲུག་/ ཟླ་ཁིད་/

4.14 职衔名<ཚ་ཁོང་མིང་།>(nx)

职衔、称谓名为一个分词单位。

示例：

གོ་ཤི་/ ཟླ་ཁིད་/ བཀའ་སྒྲོན་/ ཏུང་ཡིག་/ མདའ་དཔོན་/ ཟྱི་དཔོན་/ ཀྲུ་ཁི་/ ཏུང་ཅི་/ སྒོ་བ་ཟྱི་/ ཨ་ཁུ་/ ཨ་ ཁང་/ བ་པ་/

4.15 数词<གྲངས་ཀྱི་མིང་།>(m)

4.15.1 表示计数<བྲངས་གྲངས།>(mj)的词为一个分词单位。

示例：

གཅིག་/ བཅུ་/ བཅུ་ ཁྲི་/ འབུམ་/ ཏུང་ཕུར་/ ཏུག་ཅུ་རེ་ཏུག་/ བཅུ་ཅ་བཅུད་/ ཕྱེད་དང་བཅུད་/ ཆིག་སྟོང་བཞི་བཅུ་/ བཅུ་ ཆ་གཅིག་/ བཞི་ཆ་གསུམ་/ བཅུ་ཆ་ཏུག་ཅུ་/ སྟོང་ཆ་ལྔ་/

4.15.2 表示序数<རིམ་གྲངས།>(mx)的数词为一个分词单位。

示例：

ཨང་དང་པོ་/ དང་པོ་/ གཉིས་པ་/ གསུམ་པ་/ བཅུ་དང་ཏུག་ཅུ་པ་/

4.15.3 表示概数<ཚད་གྲངས་གྱི་མིང་།>(mg)的数词为一个分词单位。

示例：

འགའ་/ ལ་ཤས་/ བཅུ་ཡས་མས་/ བཅུ་ལྷག་/ གཅིག་ཙམ་/ ཏུ་མ་/

4.16 量词<འཇལ་བྱེད་གྱི་མིང་།>(q)

4.16.1 度量词<འཇལ་གཤོར་གྱི་མིང་།>(qd) 为一个分词单位。

示例：

སར་/ མཚན་/ འདྲིམ་/ རྒྱང་གྲགས་/ དཔག་ཚད་/ ལུ་/ ཚུན་/ ཟླ་/ རྒྱ་མ་/ སྐང་/ ཉེང་/ ཟེ་/ ལལ་/ མིད་/

4.16.2 个体量词<ཁེར་རྒྱུད་གྱི་མིང་།>(qg) 为一个分词单位。

示例：

ཉག་མ/ ཡ/ ཀང་/

4.16.3 集体量词<ཚོགས་སྒྲིལ་མིང་།>(qj) 为一个分词单位。

示例：

ཚོགས་/ ལྷ/ རོན་པོ་/ རུན་པོ་/ རྒྱུ་/ ཆ/

4.16.4 倍率量词<ཕྱབ་ཚད་གྱི་མིང་།>(ql) 为一个分词单位。

示例：

ཕྱབ་/ ཆ་ཤས་དྲུག་/

4.16.5 动量词<ལན་གྲངས་གྱི་མིང་།>(qc) 为一个分词单位。

示例：

ཐེངས་/ ཚར་/ དེ་བ་/ ལན་/

4.17 代词<ངེས་བཟུང་གི་མིང་།>(r)

4.17.1 人称代词<གང་ཟག་ངེས་བཟུང་གི་མིང་།>(rr) 为一个分词单位。

示例：

ང་/ ཁྱོད་/ ཁྱེད་/ ཁོ་/ ཁོ་པོ་/ ཁོ་མོ་/

4.17.2 单音节人称代词加“ཚོ་ཅག་”为一个分词单位。

示例：

ང་ཚོ་/ ཁྱེད་ཚོ་/ ཁོང་ཚོ་/ ཁྱེད་ཚོ་/ དེད་ཅག་/ བདག་ཅག་/ ཁྱེད་ཅག་/ ཁྱེད་ཅག་/

4.17.3 表示复数的代词为一个分词单位。

示例：

ཕྱི་མས་/ དག་/

4.17.4 疑问代词<དྲི་བ་སྟོན་པའི་མིང་།>(ry) 为一个分词单位。

示例：

ཅི་/ ཅི་/ སུ་/ གང་/ ལྟ་/ ཅི་ཙམ་/ དུ་/ ག་ཚོད་/

4.17.5 疑问代词与“སྟེད། སྟེད། ལྟར། བཞིན། སྐད། ཞིག་སྟེ། སྐད། འདྲ། ཕྱིར།”组合为一个分词单位。

示例：

ཅི་ཞིག་/ ཅི་འདྲ། གང་ཞིག་/ ཅིའི་ཕྱིར་/ ཅི་ལྟར་/ ཅི་སྐད་/ ཅི་འདྲ་/ ཅིའི་ཕྱིར་/ ཅི་སྟེད་/ ཅི་སྟེད་/ ཅི་ལྟར་/ ཅི་ བཞིན་/ ཅི་སྐད་/ ལྟ་ཞིག་/ གང་འདྲ་/ གང་སྟེད་/ གང་ལྟར་/

4.17.6 指示代词<བརྗོད་དོན་ངེས་བཟུང་གི་མིང་།>(rz) 为一个分词单位。

示例：

ལ་ཅིག་/ མ་མོ་/ འགའ་ཞིག་/ རེ་འགའ་/ ཅིག་ཤོས་/ ལ་ལ་/ གཞན་དག་/ འགའ་རེ་/ གང་ཅུང་/ དེ་/ དེ་ག་/ དེ་ ཙམ་/ དེ་ག་ཙམ་/ དེ་ག་/ དེ་ཚོ་/ འདི་ག་/ འདི་ཙམ་/ འདི་ག་ཙམ་/ འདི་/ དེ་འདྲ་/ འདི་འདྲ་/

4.17.7 反身代词<ཐུ་དོན་ངེས་བཟུང་གི་མིང་།>(rf) 为一个分词单位。

示例：

རང་/ མཚོག་/ ལགས་/ ཉིད་/

4.18 动词〈བྱ་བའི་མིང་།〉(v)

4.18.1 不及物动词〈བྱེད་མེད་ལས་ཚིག་།〉(vi) 为一个分词单位。

示例：

འཁོར་/ འཕྱར་/ བངས་/ ལངས་/ འདྲོགས་/ ཆད་པ་/ ལྷུང་བ་/

4.18.2 及物动词〈བྱེད་འབྲེལ་ལས་ཚིག་།〉(vt) 为一个分词单位。

示例：

བསྐྱབ་/ སྐོར་/ ལྷུར་/ བསྐྱངས་/ འཁྲོལ་བ་/ བསྐྱགས་པ་/

4.18.3 “AA”形式重叠的动词为一个分词单位。

示例：

འབབ་འབབ་/ འཁོར་འཁོར་/ འབྱུལ་འབྱུལ་/ རྒྱུང་རྒྱུང་/ འཁྲོལ་འཁྲོལ་/

4.18.4 “ABB”形式重叠的动词为一个分词单位。

示例：

འགྲོ་ཚག་ཚག་/ བསམ་ཚག་ཚག་/ ལེན་ཚག་ཚག་/ བསྐྱར་ཚག་ཚག་/

4.18.5 “AA+后缀”形式重叠的动词为一个分词单位。

示例：

དཔར་དཔར་པོ་/ བཅག་བཅག་པོ་/ བམ་པམ་པོ་/ སྤང་སྤང་པོ་/

4.18.6 “AABB”形式重叠的动词为一个分词单位。

示例：

གོ་གོ་རིག་རིག་/ བལྟས་བལྟས་རིག་རིག་/ བསམ་བསམ་འདོད་འདོད་/

4.18.7 “A+མི/མ+A”形式的动词为一个分词单位。

示例：

མཐོང་མ་མཐོང་/ ལྷུང་མི་ལྷུང་/

4.18.8 判断动词〈ནྟམ་གཅོད་ཀྱི་མིང་།〉(vj) 为一个分词单位。

示例：

ཡིན་/ རེད་/ མིན་/ མེད་/

4.18.9 存在动词〈གནས་པ་སྟོན་པའི་མིང་།〉(vc) 为一个分词单位。

示例：

ཡོད་/ མངའ་/ མཆིས་/ འདུག་/ གདའ་/ བདོག་/ ལུན་/ གནས་/ ལྷུང་/

4.18.10 助动词〈བྱ་མིང་རྗེས་མཐུན།〉(vu) 为一个分词单位。

示例：

ལྷུང་/ ལྷུང་/ ལྷུང་/ དགོས་/ ལྷུང་/ འཐད་/ འགྲིག་/ ལུས་/ དཔེ་/ འོས་/ ཚིག་/ མིས་/ བསམ་/ འདོད་/ རིགས་/ རྒྱུང་/ རེས་/ དགའ་/ སྐ་/ ཡང་/ འགོ་/ བཟད་/ འདུག་/ ལྷུང་/ སྤང་/

4.19 形容词〈ཁྱད་ཚས་ཀྱི་མིང་།〉(a)

4.19.1 一般形容词〈ཐྱིར་བཏང་ཁྱད་ཚས་ཀྱི་མིང་།〉(as) 为一个分词单位。

示例：

མཐོས་པ་/ མཐོན་པོ་/

4.19.2 谓语形容词〈བྱ་དོད་ལྷན་མིང་།〉(ad)为一个分词单位。

示例：

ཡག་/ མཛེས་/ ལྷན་/ འཇམ་/

4.19.3 “AA”形式重叠的形容词为一个分词单位。

示例：

དཀར་དཀར་/ མོག་མོག་/ བཟང་བཟང་/ ངན་ངན་/ མཛེས་མཛེས་/

4.19.4 “AA+后缀”形式重叠的谓语形容词为一个分词单位。

示例：

ངན་ངན་པོ་/ ལེགས་ལེགས་པོ་/ འཁྱོག་འཁྱོག་པོ་/

4.19.5 “A+མི/མ+A”形式的形容词为一个分词单位。

示例：

ཡག་མི་ཡག་/ མཛེས་མི་མཛེས་/ དྲང་མི་དྲང་/ དགའ་མི་དགའ་/

4.19.6 两个及以上的多音节形容词为一个分词单位。

示例：

རིང་ལུང་/ ཡག་ཉེས་/ བཟང་ངན་/ བདེན་རླུན་/ དམར་སྒྲུག་/ སེར་སྒྲུག་/ ལྗང་ནག་/ ལྗང་ནག་/ ཆ་ཆ་ལུང་ལུང་/ དགའ་དགའ་སྒྲོ་སྒྲོ་/

4.20 状态词〈ངང་ཚུལ་སྟོན་པའི་མིང་།〉(zt)

状态词为一个分词单位。

示例：

ཤིག་སེ་ཤིག་/ ཐུ་ལི་ལི་/ ཐེབས་སེ་ཐེབས་/ ཤ་ར་ར་/ ལྷལ་ལྷལ་/ ཐེམ་སེ་ཐེམ་/ ལྷལ་ལྷལ་/ ལུང་ལུང་/ བན་བན་/ གྲུ་གྲུ་/ རྒྱུག་རྒྱུག་/ ངན་ངོན་/ དྲང་མ་དྲང་/ སྐན་ན་སྐན་ནེ་/ རྒྱབ་རྒྱག་སེ་/ བག་ལྷུག་ལྷུག་/ དམར་ལྷུག་ལྷུག་/ ལྗང་ལྷུག་ལྷུག་/ འཇམ་ཐིང་ཐིང་/ རྒྱུག་དིར་རེ་/ ཡོར་རོ་རོ་/ ཁྲོག་སེ་ཁྲོག་/ ཤར་སེ་ཤར་/ ཐ་ར་ར་/ འཇར་ར་ར་/ སེ་ལི་ལི་/ མེ་རེ་རེ་/ ལྷུ་ཅ་ཅ་/ འཇོལ་ལོ་ལོ་/

4.21 拟声词〈སྒྲ་ཚུལ་སྟོན་པའི་མིང་།〉(oo)

拟声词为一个分词单位。

示例：

དིར་དིར་/ ཐག་ཐག་/ བང་བང་/ སང་སང་/ ཆེལ་ཆེལ་/ ཟེར་ཟེར་/ ལུང་ལུང་/ ལྷང་ལྷང་/ རྒྱུག་རྒྱུག་/ ཐིག་ཐིག་/ ཁྲོག་ཁྲོག་/

4.22 叹词〈འཕོད་སྒྲ།〉(ee)

叹词为一个分词单位。

示例：

ཀྲེ་མ་/ ཀྲེ་ཏུད་/ ཞེ་མ་/ ཞེ་མ་ཏྲོ་/ ཏྲོ་ཏྲོ་ཏྲོ་/ ཀྲེ་ཀྲེ་/ ཏཱ་ཙཱ་/ ཏཱ་ཙཱ་ཙཱ་/ ཏཱ་ཏྲོ་/

4.23 副词〈མི་ཟད་པའི་མིང་།〉(d)

4.23.1 程度副词〈དངོས་ཚད་མི་ཟད་པ།〉(dc)为一个分词单位。

示例：

ཤིན་ཏུ་/ རྒྱུ་/ ཏྲ་ཙཱ་/ རྒྱུ་/ རྒྱུ་ཟད་/ རྒྱུ་ཟད་ཙཱ་/ ཆེས་/ ཆེས་ཆེར་/ རབ་ཏུ་/ རྣམ་པར་/ རྣམ་ཏུ་/ རྣམ་པར་/ རྣམ་པར་ཏུ་/ རྒྱུ་ཏུ་/ རྒྱུ་པར་/ རྒྱུ་པར་ཏུ་/ དག་ཏུ་/ མད་ཏུ་/ མཆོག་ཏུ་/ ཀྲུན་ཏུ་/

4.23.2 时频副词〈དུས་ཚད་མི་ཟད་པ།〉(dp)为一个分词单位。

示例：

སྒོ་བྱར་དུ/ སྒོ་བྱར/ ཟུར་དུ/ ཟུན་དུ/ ཅུ་མཐུད་དུ/ ཡུད་ཅམ/ བམ་ཡང་/ འཕྲལ་དུ/ སྐར་བཞིན/ བམ་རྒྱན/ བམ་རྒྱན་དུ/ སྐབས་རེར/ ད་ལན/ ཟུན་པར/ གཏན་གྱི/ གཏན་དུ/ སྐར/ སྐར་གྱི/ སྐར་ནས/ སྐར་བས/ སྐར་ལས/ སྐར་ལས་ཀྱང་/ ལམ་སང་/ ཅིག་ཅར/ ཅིག་ཅར་དུ/

4.23.3 范围副词〈ཁྱབ་ཚད་མི་ཟད་པ།〉(dw)为一个分词单位。

示例：

ཁྱོན་ནས/ མ་ལྟས་པ/ ཆ་ཆང་/ ཡོངས་རྒྱུས་/ ཐམས་ཅད་/ ཡོད་དོ་ཅག་/ འབའ་ཞིག་/ མཉམ་དུ/ མཉམ་གཅིག་དུ/ ལྟན་དུ/ ལྟན་ཅིག་དུ/ ཆབས་ཅིག་དུ/ ཡོངས་སུ/ མཐའ་གཅིག་དུ/ ཀྱན་ནས/

4.23.4 情态副词〈སེམས་འདུན་མི་ཟད་པ།〉(dx)为一个分词单位。

示例：

ཅི་ནས་ཀྱང་/ ཡང་དག་པར་/ རྩ་མ/ མདོན་སུམ་དུ/ མདོར་ན/ བལ་ཆེར་/ རོ་ཐོག་དུ/ བན་ཏན་དུ/ དོན་དམ་པར་/ ཇི་ཉར་ན/ དེ་ཉར་དུ/ རིམ་བཞིན་དུ/ རིམ་གྱིས་/ ཐད་ཀར་/ ཞིབ་དུ/ ཞིབ་པར་/ འད་ཀྱིས་/ གཏན་ནས་/ ཇེས་པར་/ ཇེས་པར་དུ/ ཡེ་ནས་/ ཁྱོན་ནས་/ ཅུ་བ་ནས་/ ཁོ་ཐག་

4.23.5 否定副词〈དགག་པ་མི་ཟད་པ།〉(df)为一个分词单位。

示例：

མ/ མི/ མིན/ མེད/

4.23.6 重复副词〈ལད་སྒྲོས་མི་ཟད་པ།〉(dl)为一个分词单位。

示例：

ཡང་ཡང་/ ཡང་ནས་ཡང་དུ/ ཡང་སྒྱར་/ ད་དུང་/ ད་ཅང་/ སྐར་/ སྐར་ཡང་/

4.24 助词〈གསལ་བྱེད་ཀྱི་མིང་།〉(u)

4.24.1 动词与时态助词〈དུས་གསལ་བྱེད་ཀྱི་མིང་།〉(us)组合为一个分词单位。

示例：

འགོ་བཞིན/ བསམ་གྱིན/ སྒོ་བ་གྱིན/ བྱི་རྒྱ/ འབྲི་གཤམ་/ འབྲི་རྩེས་/

4.24.2 语气助词〈རྗོད་དག་གསལ་བྱེད་ཀྱི་མིང་།〉(uu)为一个分词单位。

示例：

ཤས་/ ག་ལ་/ ཞེ/ གང་/ ཤ/ གག་/ ཟེར་/ ལེ/ རེ/

4.24.3 比喻助词〈མཚུངས་པ་གསལ་བྱེད་ཀྱི་མིང་།〉(ub)为一个分词单位。

示例：

འདྲ་/ བཞིན་/ མཚུངས་/

4.24.4 动词与祈愿助词〈སྐུལ་སྒོན་གསལ་བྱེད་ཀྱི་མིང་།〉(uq)组合为一个分词单位。

示例：

སྐུལ་ཅིག་/ ཤོངས་ཤིག་/ འཇུལ་ཞིག་/ ཁྱིར་ཤོག་/ ཤོད་དང་/ འབད་པར་གྱིས་/

4.24.5 终结助词〈སྐར་བསྐྱ་གསལ་བྱེད་ཀྱི་མིང་།〉(uz)为一个分词单位。

示例：

གོ/ ར/ ཇ/ ཉ/ ཐ/ ཕ/ ཏ/ ཨ/ ཀྵ/ ཪ/ ཫ/ ོ/

4.24.6 语尾助词〈མཐུག་སྐད་གསལ་བྱེད་ཀྱི་མིང་།〉(ud)为一个分词单位。

示例：

ཡིན/ རེད/

4.25 位格〈ལ་དོན་〉(l)

4.25.1 业格〈ལས་སྤྱ་བ།〉(ls)为一个分词单位。业格助词有：

སྤྱ/ ར/ ར/ ར/ བ/ ལ/ ར/

4.25.2 为格〈དགོས་ཆེད།〉(lg)为一个分词单位。为格助词有：

སྤྱ/ ར/ ར/ ར/ བ/ ལ/ ར/ རྩར/ སྐར/ ཇོན/

4.25.3 依格〈གནས་གཞི།〉(lt)为一个分词单位。依格助词有：

སྤྱ/ ར/ ར/ ར/ བ/ ལ/ ར/

4.25.4 同格〈དེ་ཉིད།〉(ld)为一个分词单位。同格助词有：

སྤྱ/ ར/ ར/ ར/ བ/ ལ/ ར/

4.25.5 时格〈ཆོས་ཀླས།〉(lh)为一个分词单位。时格助词有：

སྤྱ/ ར/ ར/ ར/ བ/ ལ/ ར/ ཆོ/ སྐབས་/ རྩས་/

4.25.6 黏着性格助词“ས། ར། ཏ། ཨ། ཀྵ། ཪ། ཫ། ོ།”应予切分，若黏着性“ར། ལ། ཏ། ཨ།”时用还原的方法来保留词的原形。

示例：

བ/ས་བཤད་/ གོ/འད་གོ/ ར་མ་མཁུ་འི་སྤྱན་/ ར་མ་མཁུ་ར་སྤྱན་/ཡོད་/ དག/ཏོ/时先切分黏着词然后其词补上一个འ，这时得出词的原形。

4.26 具格〈བྱེད་སྤྱ།〉(b)

4.26.1 施事格〈བྱེད་པོ།〉(bo)为一个分词单位。施事格助词有：

གིས་/ ཀྱིས་/ ཀྱིས་/ འིས་/ ཡིས་/

4.26.2 工具凭事格〈ལག་ཆའི་བྱེད་པ།〉(ba)为一个分词单位。工具凭事格助词有：

གིས་/ ཀྱིས་/ ཀྱིས་/ འིས་/ ཡིས་/

4.26.3 材料凭事格〈རྒྱ་ཆའི་བྱེད་པ།〉(bc)为一个分词单位。材料凭事格助词有：

གིས་/ ཀྱིས་/ ཀྱིས་/ འིས་/ ཡིས་/

4.26.4 方式事格〈ཐབས་ཚུལ་གྱི་བྱ་བ།〉(by)为一个分词单位。方式事格助词有：

གིས་/ ཀྱིས་/ ཀྱིས་/ འིས་/ ཡིས་/

4.26.5 缘由格〈བྱེད་སྤྱ་རྒྱ་མཚན།〉(bg)为一个分词单位。缘由格助词有：

འིས/ ཡིས/

4.27 属格<འབྲེལ་སྒྲི།>(gi)

属格为一个分词单位。属格助词有：

གི/ གྱི/ གྱི/ འི/ ཡི/

4.28 源格<འབྱུང་ཁུངས།>(jg)

源格为一个分词单位。源格助词有：

ནས/ ལས/ ལ/

4.29 单纯连词<ཚིག་པ་རྒྱུ་བ།>(c)

4.29.1 据引连词<གཞི་འཛིན་དོན་འདྲེན།>(cv)为一个分词单位。据引连词有：

ན/

4.29.2 引指连词<ངོས་འཛིན།>(cl)为一个分词单位。引指连词有：

ལ/ ར/

4.29.3 衔接连词<མཚམས་སྒྲླེ།>(cn)为一个分词单位。衔接连词有：

ཏི/ དེ/ ཟླ/ ཅིང/ ཞིང/ ཤིང/ ནས/

4.29.4 比较连词<བསྟར་བའི་སྒྲ།>(cb)为一个分词单位。比较连词有：

ལས/ བས/

4.29.5 所指连词<ངོས་བཟུང་།>(ci)为一个分词单位。所指连词有：

ནི/ ལ/

4.29.6 添加连词<ཁ་སྐོང་།>(ck)为一个分词单位。添加连词有：

ནི/

4.29.7 统分连词<དང་སྒྲ་འབྲེད་སྒྲ།>(cd)为一个分词单位。统分连词有：

དང་/

4.29.8 缘由连词<དང་སྒྲ་རྒྱ་མཚན།>(ca)为一个分词单位。缘由连词有：

དང་/

4.29.9 时间连词<དང་སྒྲ་ཆོས་བས།>(ct)为一个分词单位。时间连词有：

དང་/

4.29.10 随同关系连词<བྱ་བའི་སྐར་གཞི་ལྟན་པའི་དང་སྒྲ།>(cc)为一个分词单位。随同关系连词有：

དང་/

4.29.11 复加桥接连词<མཐུན་པའི་ཚིག་རྒྱ།>(cf)为一个分词单位。复加桥接连词：

ཁྱང་/ ཡང་/ འང་/

4.29.12 转折连词<མི་མཐུན་པའི་ཚིག་རྒྱ།>(cz)为一个分词单位。转折连词有：

ཀྱང་/ ཡང་/ འང་/

4.29.13 复指连词<འོག་མས་གོང་མ་སྟེང་བ།>(cu)为一个分词单位。复指连词有：

ཀྱང་/ ཡང་/ འང་/

4.29.14 关系引连连词<ལྟག་མ་འདྲེན་བ།>(cg)为一个分词单位。关系引连连词：

ཏེ/ དེ/ རྟེ/

4.29.15 具指连词<འདྲེན་སྟེ།>(ce)为一个分词单位。具指连词有：

ཅས་/ ཞས་/ ཤས་/

4.29.16 释疑连词<འདྲི་གཞི་སྟོན་བ།>(cs)为一个分词单位。释疑连词有：

ཅེས་/ ཞེས་/ ཤེས་/ ཅེན་/ ཞེན་/

4.29.17 分义连词<འབྲེད་སྟེ།>(cm)为一个分词单位。分义连词有：

གམ་/ ངམ་/ དམ་/ རམ་/ བམ་/ མམ་/ འམ་/ རམ་/ ལམ་/ སམ་/ ཏམ་/

4.29.18 疑问连词<དྲི་བ།>(cy)为一个分词单位。

示例：

གམ་/ ངམ་/ དམ་/ རམ་/ བམ་/ མམ་/ འམ་/ རམ་/ ལམ་/ སམ་/ ཏམ་/

4.29.19 数起止连词<གངས་སྟེང་བ།>(cr)为一个分词单位。数起止连词：

ནས་/

4.29.20 境起止连词<ཡུལ་སྟེང་བ།>(cj)为一个分词单位。境起止连词有：

ནས་/

4.29.21 时起止连词<དུས་སྟེང་བ།>(cp),时起止连词为一个分词单位。时起止连词有：

ནས་/་་་་བར་

4.29.22 区别连词<དགར་བ།>(cq),区别连词为一个分词单位。区别连词有：

ནས་/ ལས་/ ཅུ་/

4.30 复合连词<ཕན་འདུས་མ།>(f)

4.30.1 语义转折连词<དོན་གཞན་འདྲེན་བ།>(fh)为一个分词单位。

示例：

ཡིན་ན་ཡང་/ དེ་ལྟ་མོད་ཀྱི་ ཡིན་མོད་ འོན་ཀྱང་/

4.30.2 语义关联连词<དོན་གྱི་སྐབས་དབྱེ་དང་དམིགས་འཛིན་མཚམས་སྟེང་།>(fg)为一个分词单位。

示例：

དེ་ལ་/ དེ་ཡང་/ དེ་ནི་/ འོ་ན་/ དཔེར་ན་/ མདོར་བསྟུས་ན་/ མདོར་ན་/ མདོར་དེལ་ན་/ ཟུར་ན་/ ཟུར་བཤད་ན་/ གཞན་ཡང་/ ད་དུང་/ ད་ཅུང་/

4.30.3 泛指关联连词<དམིགས་ཀྱིས་མི་འཛིན་པའི་མཚམས་སྟེང་།>(fz)为一个分词单位。

示例：

གང་དུ་ན/ གང་གི་ཕྱིར/ གང་ཉར་ཡང་/ ཇི་ཉར་ཡང་/ ཅིར་ཡང་/ ཅང་/ ཅི་ཡང་/

4.30.4 假设关联连词〈བརྟག་པ་མཐའ་བཟུང་གི་མཚམས་སྒྲིག〉(fj)为一个分词单位。

示例：

གལ་ཏེ/ གལ་སྲིད/ གཅིག་བྱས་ན/ ཆ་བཞག་ན/ དབང་དུ་བཏང་ན/

4.31 词缀〈ཕྱི་རྒྱུན།〉(h)

4.31.1 中接成分〈བར་གྱི་གསལ་བྱེད།〉(hz){~མ~}组合的词为一个分词单位。

示例：

ར་མ་ལུག/ ས་མ་འབྲོག/ རྩ་མ་འབྲོག/ རྩ་མ་རྟོལ/ གངས་མ་ཆར/ རྩ་མ་བོད/

4.31.2 后接成分〈མིང་མཐའ།〉(hh)组合的词为一个分词单位。

后接成分主要有：

a) 带元音{ཨ}的后接成分有{~པ}{~བ}{~མ}{~ཀ}{~ལ}{~ག}{~ཆ}{~ཏ}{~སྟེང}{~བྱེད}{~ཆས}等形式。

示例：

ཟམ་པ/ མིང་པ/ གང་པ/ རྒྱལ་པ/ སེར་པ/ ལྷ་པ/ ར་མ/ ལྷ་མ/ ཇ་མ/ གངས་ཀ/ ལས་ཀ/ འཁོས་ཀ/ དགུན་ལ/ གཡེར་ལ/ རྒྱལ་ལ/ དང་ག/ སྤང་ག/ ཐང་ག/ ལག་ཆ/ རས་ཆ/ རྒྱང་ཆ/ ལལ་ཏ/ ལཔ་ཏ/ ལ་ཏ/ འབྲེག་སྟེང/ ཟ་སྟེང/ འབྲེག་བྱེད/ ཉན་བྱེད/ རྩ་ཆས/

b) 带元音{ཨ}的后接成分有{~ཅ}{~ཆ}等形式。

示例：

དྲོང་ཅ/ ཅག་ཅ/ ལཔ་ཅ/ མེལ་ཆ/ ལཔ་ཆ/

c) 带元音{ཨ}的后接成分有{~པོ}{~བོ}{~མོ}{~སོ}{~ཏོ}{~པ་པོ}等形式。

示例：

དངོས་པོ/ སོག་པོ/ ཆོན་པོ/ རྩ་པོ/ སྟེང་པོ/ ཇོལ་པོ/ ལང་པོ/ བང་པོ/ ནང་པོ/ འབྲེག་པོ/ དགུག་ཏོ/ མཐུག་ཏོ/ ལྷ་པོ/ རྩ་པ་པོ/

d) 带元音{ཨ}的后接成分有{~འུ}{~བུ}{~གུ}{~ུ}等形式。

示例：

རྟེའུ/ རྟེའུ/ མེའུ/ བེའུ/ རེའུ/ རྟེའུ/ རྩལ་བུ/ གང་བུ/ རྩལ་བུ/ འཛེར་བུ/ ལ་གུ/ སུ་གུ/ རྩལ་གུ/ མཐུག་གུ/ རྩང་ུ/ རྩང་ུ/ རྩང་ུ/

e) 带{ཅིག}{ཞིག}后接成分的词为一个分词单位。

示例：

ཆབས་ཅིག/ ལ་ཅིག/ ལྷན་ཅིག/ འབའ་ཞིག/ རེ་ཞིག/

f) 带{རྟེངས}{རྟེན}{ལྟགས}{རང་བཞིན}后接成分的词为一个分词单位。

示例：

འགྲོ་རྟེངས/ བཤད་རྟེངས/ ལྟ་རྟེན/ བསམ་རྟེན/ བདེ་འཇགས་རང་བཞིན/

4.32 成语(习语)〈དཔེ་ཆོས།〉(xx)

4.32.1 四字格成语为一个分词单位。

示例：

སྐྱེ་ནང་སྤྱང་བསྐྱུས་/ གྲིབ་ནག་འདྲེར་མཐོང་/ ངན་པ་སྤྱག་འདྲན་/ ལུ་སྤྱ་གར་ཁྲིང་/ ར་མ་མཁའ་མེ་ཏོག་/ ཅུས་སྤུལ་གྱི་སྤྱ/ གང་གྲུའི་བྱེ་མ་/ ལྷང་ཁའི་མར་མེ་/ མོ་གཤམ་གྱི་བྱ/ གཅིག་ཤེས་ཀྱན་གྲོལ་/ བརྒྱ་ཁ་གཅིག་གཙོང་/ ལུ་ངན་ལམ་ཤོར་/ ཅུར་བལྟས་ལེབ་མཐོང་/ ལྷག་མཐོང་ཞེ་མེར་/

4.32.2 约定俗成的习语为一个分词单位。

示例：

ཅི་བསམ་སྤྱན་འབྲུབ་/ བཀྲ་ཤིས་བདེ་ལེགས་/ སྤྱན་སྤུམ་ཚོགས་པ་/ སྤྱེལ་ལད་འཁྲབ་པ་/ ལྷ་མོ་གཡོག་པ་/

4.32.3 能产性强的四字格词为一个分词单位。

示例：

སྐྱེ་མེན་མེར་མེན་/ དེས་མངོན་དེས་ཡིབ་/ དེས་མོག་དེས་གསལ་/ ཅུས་བདེ་གནས་བདེ་/ ལྷ་མེན་འདྲེ་མེན་/ མི་ཡག་མི་བཙོག་/

4.33 语素<བཞག་ཅུང་གི་སྒྲ།>(ii)

语素为一个分词单位

示例：

ད/ བ/ ས/ གྱ/ ག/ ང/ གི/

4.34 标点符号<ཚོག་ཏྱགས།>(ww)

藏文标点符号与其他语种的标点符号为一个分词单位。

示例：

༄༅ / ལྟོ / ལ།:། / ལེེ ཡ།

4.35 非藏文字符<བོད་ཡིག་མ་ཡིན་པ།>(ot)

藏语中出现的非藏文字符,如其他语言的符号串、数理化符号、阿拉伯数字等作为一个分词单位。

示例：

H₂O/ 汉文/ =/ cm/ 12/ 3.14/ 10%/ H₂O/ kg/ 1m/ TB/

4.36 梵文<ལེགས་སྤྱར།>(sk)

梵文为一个分词单位。

示例：

བུནྟ་/ ལྷམྟ་/

参 考 文 献

- [1] GB/T 13715—1992 信息处理用现代汉语分词规范
 - [2] GB/T 20532—2006 信息处理用现代汉语词类标记规范
-