



# 中华人民共和国国家标准

GB/T 36472—2018

---

## 信息处理用藏语短语分类与标记规范

Specification on Tibetan phrase classification and tagging  
for information processing

2018-06-07 发布

2019-01-01 实施

---

国家市场监督管理总局  
中国国家标准化管理委员会 发布

## 目 次

前言 .....	I
引言 .....	II
1 范围 .....	1
2 术语和定义 .....	1
3 总则 .....	1
4 短语类别的划分及标记代码 .....	1
附录 A (规范性附录) 信息处理用藏语短语分类及标记代码表 .....	7
参考文献 .....	11

## 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、青海师范大学、西藏自治区藏语文工作委员会办公室、中国藏学研究中心、青海民族大学、西藏大学、中国科学院软件研究所、西北民族大学。

本标准主要起草人:才让加、吉太加、熊涛、华却才让、扎洛、才藏太、才智杰、头旦才让、周毛先。

## 引 言

本标准依据藏语传统语法分类体系,根据藏语信息处理的特点和要求,按照语法功能和便于计算机自动分析和处理的原则划分藏语短语词类,并规定了藏语短语标记代码。

# 信息处理用藏语短语分类与标记规范

## 1 范围

本标准规定了信息处理中藏语短语的类别划分和标记代码。  
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 术语和定义

下列术语和定义适用于本文件。

### 2.1

**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。

### 2.2

**短语类别 phrase categories**

短语的语法分类,主要是根据语法功能划分出来的类。

### 2.3

**标记 tag**

对文本中短语类别进行标注的代码。

## 3 总则

### 3.1 短语的基本范围

本标准的类别特指信息处理用藏语短语类别,如名词性短语(NP)、动词性短语(VP)、形容词性短语(AP)等。

### 3.2 短语类别划分的原则

本标准针对藏语信息处理的特点和要求,依据藏语传统语法分类体系,并参考汉语短语分类研究成果(见参考文献),确定了藏文短语类别划分原则。短语大类之间不应出现交叉,包括标注形式层和语义层;小类间允许出现标注形式层交叉,但不应出现语义层的交叉。

### 3.3 标记代码的制定原则

依据国际通常做法,标记代码主要采用英文术语的字母。例如,“名词短语”,采用 NP;“数词”,采用 MP 作为标记代码。藏语独有的,或使用英文术语字母标记不便的,依据国内通常做法,标记代码采用汉语拼音字母。如,“判断性短语”,采用汉语拼音的首字母作为标记代码 PP。

## 4 短语类别的划分及标记代码

### 4.1 短语类别的划分

本标准将藏语短语划分为名词性短语、动词性短语、形容词性短语、数量词性短语、判断性短语、固