

蒙古语言文字数字资源建设与共享工程
信息处理用蒙古文相关标准
MGC/01-10

信息技术 传统蒙古文音节划分规范

Information technology — traditional Mongolian syllable classification

内蒙古自治区民族事务委员会

内蒙古大学

2016年11月

目 次

目 次.....	I
前 言.....	II
信息技术 传统蒙古文音节划分规范.....	- 1 -
1 范围.....	- 1 -
2 术语和定义.....	- 1 -
3 标记集的内容.....	- 2 -
4 标记代码的制定原则.....	- 2 -
5 标记集.....	- 2 -
6 音节划分原则方法.....	- 3 -

前 言

本标准是依据《现代蒙古语》（1964年）等主流语法著作，规定了信息处理用传统蒙古文蒙古文音节划分原则方法、音节分类与相应标记。

本标准由内蒙古自治区民族事务委员会提出。

本标准起草单位：内蒙古大学。

本标准起草人：达胡白乙拉、吴常福、艳丽

信息技术 传统蒙古文音节划分规范

1 范围

本标准规定了传统蒙古文音节划分规范，包括传统蒙古文书面语的音节分类及其标记、划分原则。

本标准适用于蒙古文的各种语料库、词汇集-词典、语文知识库等资源工程建设。

2 术语和定义

下列术语和定义适用于本标准。

2.1

蒙古文信息处理 Mongolian information processing

用计算机对蒙古文形、音、义等信息进行输入、排序、存储、输出、统计、提取等。

2.2

标记单位 tagging unit

蒙古文信息处理中使用的、音节、音位及其他单位，如元音、辅音等。

2.3

音节类 classification of syllable

词的语法分类，根据词法形态、句法功能、语义特征划分出来的类。

2.4

音节 syllable

人们在实际讲话当中，凭听觉或发音时肌肉的感觉，自然而然的分解出来的最小的语音片段。语音学上一般的说法是，人们在讲话当中每一次肌肉紧张度的增减交替是一个音节。一个音节一般有几个音素组成，有时一个因素也可以组成一个音节。这样的语音现实在相应的书面语中有确切稳定的体现，那就是有固定结构类型与特点而且可以确定原则方法来划分的书面语音节。

2.5

元音 vowel

发元音时，呼出的气流自由通过口腔，不遇到任何阻碍，发音器官的紧张状态是均衡的，声带振动，气流要比辅音弱一些。

2.6

辅音 consonant

发辅音时，呼出的气流，在通过口腔的时候需要冲破口腔中的发音器官所构成的障碍。因而气流要比元音强。发音时造成障碍的那一部分发音器官在气流冲出时也比较紧张一些。

2.7

元音音节 vowel syllable

一个元音可构成一个元音音节。

2.8

元音+辅音的节 vowel-consonant syllable

元音后面加辅音的音节。

2.9

元音+辅音+辅音的音节 vowel-consonant-consonant syllable

两个辅音前面有一个元音的音节。

2.10

辅音+元音的音节 consonant-vowel syllable

辅音后面加元音的音节。

2.11

辅音+元音+辅音的音节 consonant-vowel-consonant syllable

两个辅音中间有一个元音的音节。

2.12 辅音+元音+辅音+辅音的音节 consonant-vowel-consonant-consonant syllable

辅音加元音加复辅音的音节。

2.13 辅音+辅音+元音+辅音的音节 consonant-consonant-vowel-consonant syllable

复辅音后加元音在加辅音的音节。

2.14 辅音+辅音的音节 consonant-consonant syllable

口语中由双辅音构成的音节。

2.15 辅音+辅音+辅音的音节 consonant-consonant-consonant syllable

口语中连着由三个辅音构成的音节。

2.16 标记 tag

对文本中标记单位的类别进行标注的代码。

3 标记集的内容

本标准包括音节、元音、辅音、元音音节、元音+辅音的音节、元音+辅音+辅音的音节、辅音+元音的音节、辅音+元音+辅音的音节、辅音+辅音+元音+辅音的音节、复辅音节、三辅音等的标记符号。

4 标记代码的制定原则

本标准标记集包括二级标记。

4.1 《第一级标记》的命名方法

用大写英文字母标记第一级。如：

A (来自vowel)：元音

S (来自consonant)：辅音

4.2 《第二级标记》的命名方法

用大写英文字母后加数字标记第二级。如：

VC1 (来自一个consonant)：元音+辅音的音节

VC2 (来自两个consonant)：元音+辅音+辅音的音节

5 标记集

以下是标记集内容。右上角的数字表示该项所包含的构形附加成分或连接元音数。“-”表示与元音辅音的数字。“1,2”表示有几个元音或者是几个辅音构成。本标记代码分为音节分类标记代码和其他标记代码。

5.1 音节分类

音节分类标记代码见表1。

1 音节分类标记代码

序号	语文现象/单位	标记代码			音节类名称	代码说明
		第一级	第二级	第三级		
1	Oi ᠣᠢ	V			元音	<u>y</u> owel
2		VC			元音+辅音	<u>c</u> onsonant
3	ab ᠠᠪ		VC1		元音+辅音	
4	arslan ᠠᠷᠰᠯᠠᠨ		VC2		元音+辅音+辅音	
5	bi ᠪᠢ	CV			辅音+元音	
6		CVC			辅音+元音+辅音	
7	xeb ᠡᠬᠢ		CVC1		辅音+元音+辅音	
8	bars ᠪᠠᠷᠰ		CVC2		辅音+元音+辅音+辅音	
9	gxir ᠭᠢᠰᠢᠷ	CCVC			辅音+辅音+元音+辅音	

5.2 其他

其他标记代码见表2。

2 其他标记代码

序号	语文现象/单位	标记代码			类别名称	代码说明
		第一级	第二级	第三级		
1		CC			辅音	<u>c</u> onsonant
2	ǰaxidal		CC1		复辅音	
3	xotan-dv		CC2		三辅音	

6 音节划分原则方法

蒙古语中大多数词都是有多个音节构成的。因蒙古语词形变化而单词音节组合结构就有可能发生变化。正确划分书面语音节是一项重要工作。

现代蒙古语中虽然存在着上述复杂情况，书面语和口语中的音节划分有一种简单一致的方法。就是，不管是词的口语形式，书面语形式，有多个音节的词，经常在词的中间，末尾的元音前的辅音前面进行划分。例如：ün-düs-ten ᠤᠨ ᠳᠦᠰ ᠲᠡᠨ xa-rax ᠬᠠ ᠷᠠᠬ

man-dug-san ᠮᠠᠨ ᠳᠤᠭ ᠰᠠᠨ

上述划分原则与方法本质上遵循了肌肉紧缩度的增减交替原则。词根上加后缀之后原有的形式改变形成新的形式之后音节的组成规律也随着改变。这时同样用上述音节划分原则进行音节的划分。