

# 《蒙古语语法信息词典叹词分库》的建设及难点分析\*

林八鸽 包艳花\*

(内蒙古大学 蒙古学学院蒙古语文研究所, 内蒙古 呼和浩特 010021)

(呼伦贝尔学院 蒙文系, 内蒙古 海拉尔 021008)

**摘要:** 为实现自动分析和自动生成蒙古语语句而研制的《蒙古语语法信息词典》由不同的层次构成。《蒙古语语法信息词典叹词分库》(以下简称《叹词分库》)是它第二层次的有机组成部分。本文详细介绍有关《叹词分库》的建设情况,并着重探讨在建设《叹词分库》时所遇到的一些难点问题,并且提出了相应的处理方法。

**关键词:** 蒙古语语法信息词典; 叹词分库; 属性字段; 难点; 处理方法

## The Exclamation Bank in the “Mongolian Grammatical information Dictionary”

Lin Bage; Bao Yanhua

(Academy of Mongolian studies, Inner Mongolia University, Huhhot, 010021, China)

(Mongolian department, Hulunbeier College, Hailar, 021008, China)

**Abstract:** The “Mongolian Grammatical Information Dictionary” is compiled for the automatic parsing and generating of the Mongolian language, which is composed of different levels. The exclamation bank in the “Mongolian Grammatical Information Dictionary” (for short “the exclamation bank”) is an important part of its second level. In this paper, authors will particularly introduce about the designing of the exclamation bank and mainly discuss some difficulties, and offer relevant measure.

**Keywords:** “Mongolian Grammatical Information Dictionary” “the exclamation bank”, attribute field, difficulty, measure

### 一、引言

《蒙古语语法信息词典》是面向计算机的、为实现自动分析和自动生成蒙古语语句而研制的电子词典。它囊括了词法形态、句法功能、搭配特征以及正字法等方面的知识。《蒙古语语法信息词典》由不同层次构成。第一层是包括该词典所有词条的总库,第二层是各词类分库。如,名词分库、动词分库、标点符号分库、副词分库等。

\* 获教育部、国家语委(MZ115 038)项目,国家自然科学基金项目(36963005)资助。

\* 林八鸽(1978-),内蒙古大学蒙古学学院博士生, e-mail: [xigoi2004@yahoo.com.cn](mailto:xigoi2004@yahoo.com.cn); 包艳花(1979-),内蒙古呼伦贝尔学院蒙文系,研究生, E-mail: [svrgvgceceg@126.com](mailto:svrgvgceceg@126.com)

《叹词分库》是《蒙古语语法信息词典》第二层次的有机组成部分。建设《叹词分库》时，在充分利用《蒙古语语法信息词典总库》、《100万词级现代蒙古语语料库》、传统语法研究成果、参考《现代汉语语法信息词典》中叹词属性字段及关于现代汉语语法信息词典方面的著作论文等知识资源的基础上，设计出易于机器处理的各种属性字段。由于叹词本身的一些语法特点，在设计该库的属性字段时，遵循以语义信息为主，语法、句法知识为辅的原则。

## 二、建设《叹词分库》时遇到的难点及相应的处理方法

在传统蒙古语语法研究中，《根据蒙古语词的语义、语法、句法特点，把蒙古语词划分为实词、虚词、叹词等三大类。》<sup>①</sup>，把叹词放在跟实词、虚词同一个层次上进行研究，足见叹词在蒙古语词类中的地位及它具有独特的特点。

从叹词的性质来讲，不像实词一样有明确的词汇意义，也没有虚词一样的语法意义，它以特定的语义表现形式，来表示人们对外界的感受、各种心理活动、愿望，对人畜的呼唤或答应等声音；叹词一般不充当句子成分，有时以《GE-》词根的连动词连接构成引语句；叹词能表达完整的意义，因此它能独立成句；叹词的结构比较松，并且没有任何语法形态变化，不表达任何语法范畴；多数叹词是词根词，有少数是派生词。其中有些呼唤牲畜的叹词是该牲畜的缩略名。叹词的这些特点是建设《叹词分库》的有利依据。

### 1. 叹词子类划分问题

在传统蒙古语语法诸著作中，一般把叹词划分为表示心理活动的叹词和呼唤、驱赶牲畜的叹词等两大类。还有一些著作中，只针对叹词的定义，列出了一些例子，并没有对叹词进行具体的子类划分。《现代蒙古语》（1964年）<sup>②</sup>中对叹词进行了比较详细的划分，其分为表示某种情感或感觉的，表示呼唤答应的，〈A〉和〈Ai〉等感情词，诗歌、颂歌、歌谣、童谣等文学形式里用的称呼语。我们本着“以特定的语义表现形式，来表示人们对外界的感受、各种心理活动、愿望，表示对人畜的呼唤、答应、驱赶等声音的词为叹词”这一描述叹词的语义性定义，并兼顾叹词自身的词法功能、句法功能等特点，在制定面向信息处理的蒙古语标记集时把叹词分成表示心理活动的叹词（Is）和表示呼唤的叹词（Id）等两个子类。

### 2. 《叹词分库》收词问题

《蒙古语语法信息词典总库》是建设各《分库》的主要词语来源。在建设《叹词分库》时，从《蒙古语语法信息词典总库》中抽取117个叹词，作为《叹词分库》的词条。在建设《叹词分库》时：

1) 同一个词能兼类不同的叹词子类，则把它们作为不同词条收入《叹词分库》。

例如，“HA”一词可以表示高兴、惊奇、赞美等心理活动的同时，也可以表示对牛、骆驼等怒喝叫停的呼唤叹词，所以把“HA”这一词以两个词条收入《叹词分库》中。

2) 在现代蒙古语中，有一部分复合性叹词。

譬如，“AI JAYILAGVL、WA LAM\_A A、A HOGERUHUI、A YA YA、Ye HARALA、YO% YOHV……”等等。构成这些复合性叹词的词有的都可以独立使用；有的只能一个可以独立使用，而另一个不能独立使用。在建设《叹词分库》时，依据经济性原则，没有把这些复合性叹词作为独立的词条

来收入分库中，但在属性字段中专门设置了“前共现成分E-HAMTILR”和“后共现成分H-HAMTILR”字段来体现叹词的这一特点。

### 3) 关于叹词的其它一些问题

(I) 由于叹词有方言特点，义同形不同的词很多。如，“HAHAI、HALAHAI、HALAHADAI、HALAG、HALATAGAI、HALACAGAI”等叹词都表示嫌热的意思，“HO\$I、HOYI、HUI、HUII、HUIS”等叹词都表示呼叫的意思。在《叹词分库》中用“有没有同义词 ADLVD\_HA)”属性字段来描述了叹词的这一特点。

(II) 在叹词中，多义词占绝大部分。在该分库中把一个叹词所表示的意思——都给予了表述。如，《A》表示赞同，喜爱、赞美，后悔、觉醒，询问，接受，惊讶，责备、不满意，不称心等八种意思。

(III) 表示呼唤和答应牲畜的叹词和表示感觉的叹词主要以重叠形式出现。如，“GVRVI GVRVI (咕里)”，“TOYIG TOYIG (吹吹)”，“HALAHAI HALAHAI!YAGAHIGSAN HALAGVN YVM. (啊呀，真热呀!)”，“TI TI,YAGAHISAN HUITEN YVM! (啊呀，真冷!)”等。传统的蒙古语词典中，都是以严格意义上的“词”作为词条的。在建设《蒙古语语法信息词典》时，也以词为基本词条的，所以对这些词的处理也通过设置属性字段来完成的，即，“是否重叠使用 (DABTHE)”。

## 三、《叹词分库》属性字段描述及取值

对《100万词级现代蒙古语语料库》中叹词的实际使用情况进行分析并参考传统语法研究成果的基础上，在《叹词分库》中设计出了易于机器处理的16个属性字段。其中“序号”，“蒙古语”，“标记”，“拉丁转写”，“例子”等五个属性字段是从《蒙古语语法信息词典总库》中继承下来的。

《叹词分库》中所设置的属性字段的值有字符型、逻辑型两种类型。属性字段名称的设计，尽量使用表示该特征的蒙古语词，以使用户很容易从字面上能够意会，采取直接用自然语言形式表示如，(MONGGOL)、(GALIG)、(DABTAHV)、(SEREL)……等。但出于在系统中易于描述和节省字节考虑，有的没有直接用自然语言形式表示，而采用了一些简略代码，如，(UGESAIMAG)、(E-HAMTILR)和(H-HAMTILR)。《叹词分库》中设置的16个属性字段的含义及其取值规格如下：

1. 序号(NO)：指该叹词在《叹词分库》中的排列次序。
2. 蒙古语(MONGGOL)：指该叹词的蒙古文书写形式，填写该叹词的蒙古文书写形式。
3. 子类(UGESAIMAG)：填写叹词的子类标记，取值为字符型。  
如，TETEGEI(Is)、HOHO(Id)等。
4. 拉丁转写(GALIG)：填写该词的拉丁转写形式，取值为字符型。  
如，COHO(COHO),PAH\_A(PAH\_A)等。
5. 能否重叠使用(DABTAHV)：是，填Yes；不是，填No。  
如，TOYIG(Y),PAH\_A(N)等。
6. 是否派生词(EGUSUMEL)：是，填Yes；不是，填No。  
如，HA(N),HALAHAI(Y)等。
7. 是否多义词(OLANVDH\_A)：是，填Yes；不是，填No。

- 如, A(Y), GVRVI(N)等。
8. 是否有同义词 ( ADLVD\_HA ): 有, 填相应词的拉丁转写; 没有, 不填。  
如, AYILVV( AYILVG ), AYILVG( AYILVV )等。
  9. 前面是否有共现成分 ( E-HAMTILR ): 有, 填相应共现成分的拉丁转写; 没有, 不填。  
如, JAYILAGVL( AI ), HOGERUHUI( A )等。
  10. 后面是否有共现成分 ( H-HAMTILR ): 有, 填相应共现成分的拉丁转写; 没有, 不填。  
如, LAM\_A(A), WASI(DA)等。
  11. 是否表示心理活动 ( SEDHIL ): 是, 把该词表示的意思用拉丁转写的形式填写; 不是, 不填。  
如, PAH\_A( JIGSIHU )。
  12. 是否表示感觉 ( SEREL ): 是, 把该词表示的意思用拉丁转写的形式填写; 不是, 不填。  
如, HALAHAI( HALAHV )。
  13. 是否表示呼唤 ( DAGVDAHV ): 是, 把被呼唤的牲畜名字用拉丁转写的形式填写; 不是, 不填。  
如, TOYIG( HONI )。
  14. 是否表示驱赶 ( HOGEGEHU ): 是, 把被驱赶的牲畜名字用拉丁转写的形式填写; 不是, 不填。  
如, COHO( TEMEGE )。
  15. 是否表示愿望、要求 ( HUSEL ): 是, 把该词表示的意思用拉丁转写的形式填写; 不是, 不填。  
如, BAYIJA( HORIGLAHV ), AYILVV( ICIGEHU )等。
  16. 例子 ( JI\$IY\_E ): 填写该词在真实语料中出现的典型例子。  
如, HENG( HENG! BITEGEI TEGEJU SINVHAI AGASILA! ), COHO( COHO! COHO! GEGSEGER HOLADABA )等。

#### 四、《分库》的应用

《蒙古语语法信息词典》囊括了词法形态、句法功能、搭配特征以及正字法等方面的知识。由于叹词本身的语法特点, 该分库的属性字段以语义信息为主, 语法、句法知识为辅。它在蒙古语自动分析和生成、机器翻译、自动校对等各种信息处理系统中起到重要作用。

##### 1. 在蒙古语自动分析中的应用

上面所述的《叹词分库》字段信息对于蒙古语语料库词性标注是有利的。例如: 我们可以通过“UGESAIMAG”属性字段信息对句子《HOYI! CINGGI/JU YAGAHI/JV BOL/DAG YVM? (喂! 这样怎能行呢?)》中的《HOYI》标注词性。

《100万词级现代蒙古语语料库》是以拉丁转写蒙古文字母的方法建立的, 此分库的“GALIG”字段会给查询、统计语料库带来方便。再用句子《HOYI! CINGGI/JU YAGAHI/JV BOL/DAG YVM? (喂! 这样怎能行呢?)》为例, 可以通过《HOYI》从语料库中查找并统计该词及其上下文信息。

## 2. 在蒙古语自动生成中的应用

《分库》为汉-蒙机器翻译系统正确生成蒙古语“词形式”提供叹词正字、正音、语义搭配方面的属性信息。汉-蒙机器翻译系统中汉语句子“哎！好像一只鹅受伤了”翻译成蒙古文后“AI HOGERUHUI ! NIGE GALAGV SIRHATV/GSAN BOLOLTAI ,”，其中“AI”的正确生成可依据《分库》中的“H-HAMTILR”、“SEDHIL”等属性字段值自动生成。

## 五、结语

《蒙古语语法信息词典》总库及其各“分库”是蒙古语语言知识库的有机组成部分。《叹词分库》中虽然包含了叹词的读音、结构、语义等属性信息，但由于语料所限，叹词本身也有很多问题尚未解决，所以《叹词分库》不可能反映现代蒙古语叹词的全部特点。但随着语法研究的深入和语言信息处理系统的不断发展，会及时补充《叹词分库》的有关信息。

### 注释：

①②内蒙古大学蒙古学学院蒙古语研究所：《现代蒙古语》（蒙文）（P. 367）（P. 570-572），内蒙古人民出版社，2005年。

## 参考文献

- (1)内蒙古大学蒙古学学院蒙古语研究所：《现代蒙古语》（蒙文），内蒙古人民出版社，2005年。
- (2)清格尔泰：《现代蒙古语语法》（蒙文），内蒙古人民出版社，1999年。
- (3)纳·格日勒图：《蒙古书面语语法研究》（蒙文），内蒙古大学出版社，1998年。
- (4)嘎日迪：《现代蒙古语》（蒙文），内蒙古教育出版社，2001年。
- (5)德·青格勒图：《面向信息处理的蒙古语固定词组研究》（蒙文），内蒙古教育出版社，2001年。
- (6)内蒙古大学蒙古学学院蒙古语研究所：《蒙汉词典》，内蒙古大学出版社，1999年。
- (7)巴图格日勒：《关于叹词》（蒙文），《西北民族学院学报》，1985年第2期。
- (8)艳花：《蒙古语语法信息词典副词分库》的建设，内蒙古大学学报（哲学社会科学版），2007年第3期。
- (9)俞士汶等 《现代汉语语法信息词典详解》（第二版），北京大学出版社，2003年。
- (10)那顺乌日图 《蒙古语语法信息词典》框架设计，内蒙古大学博士学位论文，2000年。
- (11)巴 萨日娜：《蒙古语语法信息词典》动词语法属性字段设置，内蒙古大学学报（人文社会科学版），2005年11月。
- (12)刘云，俞士汶，朱学锋，段慧明：《现代汉语虚词知识库的建设》，语言文字应用，2005年2月，第一期。
- (13)淑琴 《蒙古语语法信息词典构形附加成分分库》的设计与实现，内蒙古大学硕士学位论文，2005年。