

# 蒙古语构形附加成分重叠使用特征及其模型

淑琴，艳花

(内蒙古大学 蒙古学学院蒙语所，内蒙古 呼和浩特 010021)

**摘要：**词法形态分析是蒙古文信息处理诸多应用系统的一个不可或缺模块。传统蒙古语语法学在这方面取得了一定的成果。然而，面向计算机的语言知识颗粒度极细、颗粒数量极大，在本质上不同于面向人的语言知识。本文根据“100万”语料查询统计、分析结果，归纳、总结了现代蒙古书面语构形附加成分重叠使用特征及其模型。它将为蒙古文信息处理实用系统提供关键的词法形态方面的知识。

**关键词：**蒙古语；构形附加成分；重叠使用；模型

**中图分类号：**H212

**文献标识码：**A

## 一、引言

蒙古语是属于黏着型语言，其单词在动态语料中的“词形式”<sup>①</sup>丰富多样。构形附加成分是蒙古语词的词法形态变化形式，本文沿用“不改变词汇意义，只增加语法意义的附加成分称为构形附加成分”<sup>②</sup>这个定义，但不排除构形附加成分的构词作用和构词附加成分的构形作用。

传统语法研究词法形态方面所取得的丰硕成果（如，构形附加成分的具体形式、类别、正字法、功能、所附加的语法意义、构成等）是归纳、总结构形附加成分重叠使用模型的重要理论依据。

现代蒙古书面语构形附加成分的重叠使用有几种类型、有什么具体形式、哪些形式合理、哪些形式不合理，每个形式所表达的功能、意义如何等方面不同学者有不同的观点，如，“GSAD|GSED, GSAN|GSEN, D”“重格”、复数附加成分的重叠使用、“零形式”与其它（领属附加成分）构形附加成分之间的关系等。我们先不考虑理论上的这些不同观点，从语言信息处理的角度，基于语料中所使用的外在形式，可以把构形附加成分的重叠使用分为同类构形附加成分的重叠使用和异类构形附加成分的重叠使用两种，每个构形附加成分的重叠使用都有它自己的特点，并且所缀接的词干的不同也会影响具体构形附加成分的重叠使用特征。本文归纳、总结构形附加成分重叠使用特征及其模型<sup>③</sup>时根据“100万词级现代蒙古语文数据库”<sup>④</sup>查询统计、分析结果，兼顾其它语料中的实际情况。并且在《蒙古语语法信息词典》构形附加成分分库中按构形附加成分类别设置了表示构形附加成分重叠使用特征的“Fp|<sup>⑤</sup>Fc|Fa|Fm|Fn|Ft|Fd|Fe|Fi|Fh”<sup>⑥</sup>十个属性字段，每个构形附加成分后面填写了其前面出现的具体构形附加成分标记。这些属性信息为蒙古语语料切分标注、机器翻译、文本自动校对及文字识别等应用系统提供关键的词法形态方面的知识。

## 二、同类构形附加成分的重叠使用

复数、格、级、数词和态附加成分可以重叠使用，领属、祈使式、陈述式、副动词、形动词、名动词、体附加成分不能重叠使用。同类构形附加成分的重叠使用情况如下：

### （一）复数附加成分

ᠳᠡᠭᠤᠭᠤᠨ (-NAR|-NER|NAR|NER) + ᠠᠨᠳᠤ (-VD|-UD): ᠳᠡᠭᠤᠭᠤᠨ ᠠᠨᠳᠤ DEGUU-NER-UD<sup>⑦</sup>

ᠨᠣᠬᠣᠳᠤ (D) + ᠳᠡᠭᠤᠭᠤᠨ (-NAR|-NER): ᠨᠣᠬᠣᠳᠤ ᠳᠡᠭᠤᠭᠤᠨ NOHOD-NER

ᠨᠣᠶᠠᠳᠤ (D) + ᠠᠨᠳᠤ (-VD|-UD): ᠨᠣᠶᠠᠳᠤ ᠠᠨᠳᠤ NOYAD-VD

ᠬᠣᠮᠣᠰᠤ (S) + ᠠᠨᠳᠤ (-VD|-UD): ᠬᠣᠮᠣᠰᠤ ᠠᠨᠳᠤ HOMOS-UD

(二) 格附加成分

- നൾ ശ്ല (-YIN|-VN|-UN|-V|-U) + ശ്ല (-DV|-DU): നഹിറാജി നൾ ശ്ല JAHIRVGCI-YIN-DV
- നൾ ശ്ല (-YIN|-VN|-UN|-V|-U) + ശ്ല (-DVR|-DUR): നാഗാക്വ-നൾ ശ്ല NAGACV-YIN-DVR
- നൾ ശ്ല (-YIN|-VN|-UN|-V|-U) + നൾ (-ACA|-ECE): ന്ഗ-നൾ നൾ ENG-UN-ECE
- നൾ ശ്ല (-YIN|-VN|-UN|-V|-U) + നൾ (-TAI|-TEI): റുറ-നൾ നൾ EDUR-UN-TEI
- ശ്ല (-DV|-DU|-TV|-TU) + നൾ (-ACA|-ECE): റ്റ (നൾ ശ്ല നൾ GER-TU-ECE)
- നൾ (-TAI|-TEI) + നൾ (-YIN): ബാഗ്താഗാമജി-നൾ നൾ BAGTAGAMJI-TAI-YIN
- നൾ (-TAI|-TEI) + ശ്ല (-DV|-DU): നൾ നൾ ശ്ല JVD-TAI-DV
- നൾ (-TAI|-TEI) + ശ്ല (-DVR|-DUR): വന്തവ്ഗവ്ന-നൾ ശ്ല VNTVGVN-TAI-DVR
- നൾ (-TAI|-TEI) + നൾ (-YI): വെഹ്-നൾ നൾ VDH\_A-TAI-YI
- നൾ (-TAI|-TEI) + നൾ (-ACA|-ECE): റ്റ-നൾ നൾ ONGGE-TEI-ECE
- നൾ (-TAI|-TEI) + നൾ (-BAR|-BER): റ്റ-നൾ നൾ ITEGEL-TEI-BER
- ശ്ല ശ്ല (-LVG\_A|-LUGE) + നൾ (-YIN): സരങ്ഗവ്വ-നൾ ശ്ല SARANGGVW\_A-LVG\_A-YIN
- ശ്ല ശ്ല (-LVG\_A|-LUGE) + നൾ (-TAI|-TEI): ബോൾ-നൾ ശ്ല BOLHV-LVG\_A-TAI

(三) 级附加成分

- ശ്ല (-BTVR|-BTUR) + നൾ നൾ (-HAN|-HEN): ശ്ല-നൾ ശ്ല SVLABTVRHAN
- ശ്ല (-BVR|-BUR) + നൾ നൾ (-HAN|-HEN): (നൾ-നൾ ശ്ല CAYIBVRHAN)

(四) 数词变化形式

- നൾ നൾ (GAD|GED) + നൾ ശ്ല (DVGAR|DUGER): (നൾ-നൾ ശ്ല ARBAGADDVGAR)
- നൾ നൾ (GAD|GED) + നൾ ശ്ല (GVLA|GULE): (നൾ-നൾ ശ്ല HORIGADVGVLA)

(五) 态附加成分

- നൾ ശ്ല (GVL|GUL) + നൾ ശ്ല (GDA|GDE): (നൾ-നൾ ശ്ല TANILCAGVLVGDAJV)
- നൾ ശ്ല (GVL|GUL) + നൾ ശ്ല (JAG\_A|JEJE): ബെ-നൾ ശ്ല BEYELEGULJEJE
- നൾ ശ്ല (G\_A|GE) + നൾ ശ്ല (GVL|GUL): നൾ-നൾ ശ്ല OYILAGAGVL
- നൾ ശ്ല (G\_A|GE) + നൾ ശ്ല (LG\_A|LGE): നൾ-നൾ ശ്ല DAGAGALGAGSAN
- നൾ ശ്ല (G\_A|GE) + നൾ ശ്ല (GDA|GDE): നൾ-നൾ ശ്ല NICVGAGDA
- നൾ ശ്ല (G\_A|GE) + നൾ ശ്ല (LCA|LCE): നൾ-നൾ ശ്ല GARGALCA
- നൾ ശ്ല (G\_A|GE) + നൾ ശ്ല (JAG\_A|JEJE): നൾ-നൾ ശ്ല DUGURGEJEJE
- നൾ ശ്ല ശ്ല (GDA|GDE|DA|DE|TA|TE) + നൾ ശ്ല (GVL|GUL): നൾ-നൾ ശ്ല HAYAGDAGVL
- നൾ ശ്ല ശ്ല (GDA|GDE|DA|DE|TA|TE) + നൾ ശ്ല (JAG\_A|JEJE): നൾ-നൾ ശ്ല MORDAJAG\_A
- നൾ ശ്ല (DA|DE|TA|TE) + നൾ ശ്ല (GDA|GDE): നൾ-നൾ ശ്ല OLDAGDAJV
- നൾ ശ്ല (LDV|LDU) + നൾ ശ്ല (GVL|GUL): നൾ-നൾ ശ്ല MVSHILDVGVL
- നൾ ശ്ല (LDV|LDU) + നൾ ശ്ല (JAG\_A|JEJE): നൾ-നൾ ശ്ല UIMELDUJEJE
- നൾ ശ്ല (LCA|LCE) + നൾ ശ്ല (GVL|GUL): നൾ-നൾ ശ്ല TANILCAGVL
- നൾ ശ്ല (LCA|LCE) + നൾ ശ്ല (LDV|LDU): നൾ-നൾ ശ്ല BVLIALCALDV
- നൾ ശ്ല (LCA|LCE) + നൾ ശ്ല (JAG\_A|JEJE): നൾ-നൾ ശ്ല SVRVLCAJAG\_A

三、异类构形附加成分的重叠使用

异类构形附加成分的重叠使用内容繁多，不宜一一列举，按不同类别归纳如下：

(一) 复数附加成分

1. 复数+格: നൾ-നൾ ശ്ല BAYISING-VD-TV
2. 复数+领属: നൾ-നൾ ശ്ല DARVG\_A-NAR-IYAN
3. 复数+附属: നൾ-നൾ ശ്ല JARVCA-NAR-TVNI
4. 复数+格+领属: നൾ-നൾ ശ്ല BAGATVR-VD-VN-IYAN



10. 态+体+形动词:  $\text{ᠲᠠᠭᠠᠷᠠᠭᠪᠯᠴᠢᠬᠠᠭᠰᠠᠨ}$  TAGARAGVLCIHAGSAN
11. 态+体+形动词+格:  $\text{ᠣᠷᠠᠭᠪᠯᠴᠢᠬᠠᠬᠤᠨ}$  OROGVLCIHAHV-DV
12. 态+体+形动词+领属:  $\text{ᠶᠠᠪᠪᠠᠭᠪᠯᠴᠢᠬᠠᠭᠰᠠᠨ}$  YABGVLCIHAGSAN-IYAN
13. 态+副动词+领属:  $\text{ᠣᠷᠠᠭᠪᠯᠠᠩᠭᠭᠢᠲᠠ}$  OROGVLVNGGVTA-BAN
14. 态+形动词+格:  $\text{ᠪᠡᠶᠡᠯᠡᠭᠤᠯᠬᠤ}$  BEYELEGULHU-YI
15. 态+形动词+领属:  $\text{ᠰᠣᠩᠭᠣᠭᠳᠠᠭᠰᠠᠨ}$  SONGGOGDAGSAN-IYAN
16. 态+形动词+附属:  $\text{ᠪᠠᠷᠢᠭᠳᠠᠭᠰᠠᠨ}$  BARIGDAGSAN-DVNI
17. 态+形动词+格+领属:  $\text{ᠪᠠᠷᠢᠭᠳᠠᠬᠤ}$  BARIGDAHV-ACA-BAN

### (九) 体附加成分

1. 体+祈使式:  $\text{ᠬᠠᠷᠢᠴᠢᠬᠠᠢ}$  HARICIHAY\_A
2. 体+陈述式:  $\text{ᠳᠠᠭᠠᠷᠢᠴᠢᠬᠠᠵᠠᠢ}$  DAGARICIHAJAI
3. 体+副动词:  $\text{ᠪᠷᠢᠴᠢᠬᠠᠭᠰᠠᠨ}$  VRICIHAGSAN
4. 体+形动词:  $\text{ᠵᠠᠶᠢᠯᠠᠰᠢᠬᠢᠬᠤᠨ}$  JAYILASHIHV
5. 体+名动词:  $\text{ᠫᠪᠸᠳᠠᠶᠢᠬᠠᠯᠲᠠᠢ}$  PVVDACIHALTAI
6. 体+形动词+格:  $\text{ᠬᠣᠷᠢᠴᠢᠬᠠᠭᠰᠠᠨ}$  HORICIHAGSAN-I
7. 体+形动词+领属:  $\text{ᠵᠠᠶᠢᠯᠠᠰᠢᠬᠢᠬᠤᠨ}$  JAYILASHIHV-BAN
8. 体+形动词+格+领属:  $\text{ᠵᠠᠨᠴᠢᠭᠳᠠᠬᠤ}$  JANCIGDAHV-ACA-BAN

### (十) 附属附加成分

1. 附属+领属:  $\text{ᠭᠠᠷᠠᠲᠠᠬᠢ}$  GAR-TAHI-BAN
2. 附属+格:  $\text{ᠲᠠᠰᠪᠯᠭᠠᠨᠠᠨᠭᠢᠭᠳᠠᠨᠠᠨᠢᠬᠢ}$  TASVLG\_A-NVGVD-VN-HI-YI
3. 附属+格+领属:  $\text{ᠶᠠᠭᠤᠨᠠᠨᠢᠬᠢᠲᠠᠢ}$  VG-VN-HI-TAI-BAN

## 四、构形附加成分重叠使用模型

基于上述同类和异类构形附加成分的重叠使用特征归纳出构形附加成分的重叠使用模型，具体内容如下：

- 复数（重叠）+附属+格（重叠）+附属+格（重叠）+领属
- 数词（重叠）+级（重叠）+附属+格（重叠）+领属
- 态（重叠）+体+祈使式
- 态（重叠）+体+陈述式
- 态（重叠）+体+副动词+领属
- 态（重叠）+体+副动词+级
- 态（重叠）+体+形动词+复数（重叠）+格（重叠）+领属
- 态（重叠）+体+形动词+复数（重叠）+格（重叠）+附属
- 态（重叠）+体+名动词+格（重叠）

## 五、应用

构形附加成分重叠使用特征及其模型是正确识别、分析、生成每一个具体“词形式”的重要依据。

“蒙古语语料切分标注软件”<sup>⑩</sup>正确切分、还原、标注蒙古语语料，离不开构形附加成分重叠使用特征及其模型所提供的信息，并且该软件可以直接运用《蒙古语语法信息词典》构形附加成分分库中前接构形附加成分类别属性字段“Fp|Fc|Fa|Fm|Fn|Ft|Fd|Fe|Fi|Fh”的值。在汉-蒙机器翻译系统中，蒙古语“词形式”的自动分析与自动生成也可以从中获取词法形态方面的信息。

此项研究将提高蒙古文文本自动校对质量和文字识别准确率。从前接构形附加成分类别



- [6] 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
- [7] 冯志伟,《机器翻译研究》,中国对外翻译出版公司,2004年
- [8] 黄昌宁,李涓子,《语料库语言学》,商务印书馆,2002年
- [9] 俞士汶,《计算语言学概论》,商务印书馆,2004年
- [10] 吴立德等,《大规模中文文本处理》,复旦大学出版社,1997年
- [11] Kenneth R. Beesley, Lauri Karttunen, Finite State Morphology, CSLI publications, 2002.
- [12] Kennedy. G., Introduction to Corpus Linguistics, Addison Wesley Longman Limited, 1998.

## The Sequence of Mongolian Morphological Form and it's Model

Shuqin, Yanhua

(The Institute of Mongolian Studies, Inner Mongolia University, Hohhot 010021, China)

**Abstract:** Morphological analysis is an indispensable component of many Mongolian information processing systems. Traditional Mongolian grammar has made large amounts of efforts in this area and achieved, in some cases, promising results. However, the linguistic knowledge which is applied in computer systems is essentially different from that used by human being, because in the former the granularity of knowledge is much more fine, and the amount is much larger. This paper sums up the sequence of Mongolian morphological form and it's model based on the statistical and analytical results of One Million Word Mongolian Corpus. It will provide key morphological knowledge for Mongolian information processing applied systems.

**Key words:** Mongolian; morphological form; sequence; model

**收稿日期:** 2005-10-31;

**基金项目:** 国家 863 计划项目 (2003AA115510); 国家自然科学基金项目 (36963005)

**作者简介:** 淑琴(1979-),女,内蒙古科左后旗人,博士研究生,主要研究方向为蒙古文信息处理;艳花(1979-),女,内蒙古科左中旗人,硕士研究生,主要研究方向为蒙古文信息处理。