

蒙古语语义格的分析

德·萨日娜

(内蒙古大学 蒙古学学院, 内蒙古 呼和浩特 010021)

摘要: 我们认为, 作为信息处理的基础研究, 用格语法的理论、方法分析和概括蒙古语句义现象有一定的理论和应用价值。根据蒙古语语料中出现的各种语义成分的语义结构特征, 初步概括出了基本语义格 4 大种和它们下层的小类格共 40 种。这 4 大格是主体、客体、邻体和饰体。主体: 句子中充当动作的发出者或陈述、描写、说明的主题。客体: 句子中充当动作所涉及的对象。邻体: 句子中充当动作的情景、辅助材料及外围基因的成分。饰体: 句子中充当动作的修饰成分。

关键词: 蒙古语; 语义格; 语义格种类

中图分类号: H212 **文献标识码:** A

格语法是 20 世纪 60 年代初由美国语言学家菲尔默 (Fillmore) 提出的语法理论。它源于转换生成语法中的深层结构分析。该理论提出了简单句中谓语与名词短语之间的句法语义概念关系, 语义格即句子中由谓语所支配的各种语义成分。格语法理论在自然语言处理研究中被广泛运用。

菲尔默最初提出的英语的语义格有 6 种, 即施事格 Agentive、工具格 Instrumental、与格 Dative、使役格|使成格 Factive、处所格 Locative; Place、客体格|受事格 Object; Patient 等。格语法理论发展到今天, 其研究方法有了很大的改进。随之, 人们对语义格这一概念的认识也更加细致和完善。我们认为, 各种语言的语义格类型大体上相同, 不同的是它们的表现形式和由此引起的小类语义格的差异。对蒙古语这样形态发达的语言来说, 从形态特征上来说明语义格的类型有一定的特点和优势。

一、研究蒙古语语义格的意义

随着蒙古文信息处理工作的进一步发展, 基础研究和应用研究中都迫切需要蒙古语语义知识。根据这一需求, 我们正在进行面向信息处理的蒙古语语义学方面的探索性研究。我们认为, 作为信息处理的基础研究, 用格语法的理论、方法分析和概括蒙古语句义现象有一定的理论和应用价值。具体来讲, 我们目前所进行的汉蒙机器翻译、语料库加工和自动校对等应用研究都需要有关词语搭配方面的语义知识。整个蒙古文信息处理工作的发展也需要更多、更系统的蒙古语语义知识。

由于现有的蒙古语语义研究的基础非常薄弱, 格语法研究工作也刚刚开始, 我们现在的研究工作也只是初步的分析、统计和概括, 以作为下一步工作的基础。

二、分析方法的探索

在蒙古语语义格分析中有两方面的要求: 一是要认真观察语义格这一概念在蒙古语中的具体体现, 描述必须概念清楚、分类合理、特征鲜明; 二是分析得出的基本知识要符合蒙古语句法语义结构的同时, 也能够满足信息处理研究中形式化、属性化的要求。根据这两种要求, 我们在语料分析中采取了以下几个步骤: 首先, 从现代蒙古语语料库教科书部分中抽取不同体裁的语料 (民间故事、叙事文和议论文三种), 对其分别进行语义格分析; 其次, 对三种语料的分析结果, 分别从语义格的谓语中心词属性、格中心词属性、格标志等几个方面

逐一进行归纳和分类；最后，根据上述两项分析的结果，设计出蒙古语语义格一览表（包括格种类、格的含义解释、主要句型和例）和蒙古语语义格的总的层级表（包括语义格大类、中类、小类及它们的各属性），并列出了所用不同体裁语料中各语义格的分布状况和大概的数据。

对蒙古语来说，语义格分析涉及到句义、句法、词义和词法等多方面的知识。用这些知识的映射关系来解释或处理一种语义现象是一个崭新的课题，分析中所遇到的难题比较多，其中几个主要问题的初步认识和处理意见如下：

第一，语义格成分的结构问题：据分析，蒙古语的大部分语义格是 NP 短语结构，尤其句子的核心语义格主体和客体均属于 NP 短语结构，而邻体和饰体中极少数成分是 VP 结构的，如原因、动状两种语义格具有这样的结构特点。

第二，复合句的语义格切分问题：单句的语义格切分容易，而语料中往往是单句结构少，复合句结构多。在复合结构句子的处理上，我们采取的办法是：若在复合句中的两个以上述体结构（包括无主、客体的述体结构）之间存在语义上的并列关系（分句和分句之间的并列、转折、选择等关系）、制约关系（分句和分句之间用形动词的 ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ 或 ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ 形式及副动词形式连接），将每个述体结构与单句同样进行语义格分析；若在复合句中的任何一个述体结构整体上或充当句中主体、客体等格的定语成分、或充当某种语义格的，那么该述体结构内部的语义格分析无法和单句同一层次上进行。换句话说，包含大量复合句结构语料的语义格分析实际上是分层次进行的。我们初步认为充当某一种语义格定语成分的述体结构和整体上充当某一种语义格的述体结构内部的格分析不属于第一层分析。作为初步的探索性研究，本文所描述的内容都是在第一层分析中所获得的知识。

第三，语义格的属性问题：我们认为在蒙古语中决定语义格性质的主要因素，因语义格的种类不同而有所不同。对核心格（主体、客体）而言，谓语中心词的语义属性决定着语义格的种类。谓语中心词的语义属性主要包括词类、动词的及物和不及物、自发和非自发及语义类等特征。对非核心格（邻体、客体）而言，从谓语中心词的属性上很难说明它们的类型，而从它们格中心词的属性和格标类型的异同来区分语义格不同类型比较有把握。格中心词的属性主要指词类、语义类的区别。蒙古语的格标有语法格附加成分、领属附加成分或语气词、虚词、构词词缀、构形词缀等 5 种。

第四，在蒙古语语义格分析中不属于格分析范围的成分有位于谓语中心词后面的所有辅助成分、复合句的连接成分、插入语、感叹词等。

鉴于以上分析，我们对蒙古语语义格的描述主要围绕着语义格与谓语中心词之间的语义关系、谓语中心词属性、格中心词属性和格标类型等几方面来进行的。

另外，在语义格分析中，除了理应处理的复合句结构的二层、三层的语义结构未能得到解决以外，分析中未处理的成分还有：

第一，蒙古语语料中频繁出现的“ ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ ”词干连接动词不仅形式多样，而且在句子中所承担的语义功能也比较复杂。对于分析中出现的《 ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ 》动词的语义问题还没有做出分析和概括，关于这方面的研究需要在更多的语料分析基础上进行。

第二，在语料中出现的个别语义成分，如“ $\text{ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ}$ ”、“ $\text{ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ ᠠᠨᠠᠭᠠᠨᠠᠨᠠᠭ}$ ”等语义成分的语义格类型未能予以确定。

三、蒙古语语义格的种类和特征

我们根据蒙古语语料中出现的各种语义成分的语义结构特征，初步概括出了基本语义格 4 大种和它们下层的小类格共 40 种。这 4 大格是主体、客体、邻体和饰体。

主体：句子中充当动作的发出者或陈述、描写、说明的主题。它的下层包括施事、当事、系事、领事、描述、存在和说明等共 7 种。

客体：句子中充当动作所涉及的对象。它的下层包括受事、结果、客事、与事、受益、转变、同事、触事、疏远、属事、乘事、向跟、包含、遭遇和关联等共 15 种。

邻体：句子中充当动作的情景、辅助材料及外围基因的成分。它的下层包括时间、地点、源点、终点、路径、原因、目的、方向、工具、材料、基准、依据和范围等共 13 种。

饰体：句子中充当动作的修饰成分。它的下层包括性状、动状、方式、数量和频次等共 5 种。

蒙古语的这 4 种大格及其下属各小类的表现形式和例子请见表 1 蒙古语语义格一览表。从这一表上看出, 每个大格的下层各小类在语义上或形式上又存在着互相分和的特征。根据这一特征, 我们可以对各大格内部再进行一次归类, 这样会得出大格和小类中间的又一个类。请看表 2 蒙古语语义格总的层级表。该表中还包含了每个大、中、小类语义格的谓语中心词、格中心词和格标等的属性信息。

表中所用标记说明:

Nx/体词, C/零主格, Vt/及物动词, Ve/不及物动词, Vb/存在动词, T/时间词, O/处所词, JU/方向词, D/副词, A/形容词, U/摹拟词, M/数词, Q/量词, S/语气词, -/格附加成分标记, +/词干和词缀的连接符号。

通过对抽样语料的分析,发现不同体裁语料中出现的语义格种类及其频率都有所不同。因此,我们对抽样的不同体裁语料的语义格的分布状况也进行了大概的统计,请见表3 语义格在不同体裁语料中的分布。

语料分析例:

[illegible]

ገጽ/ACT ገጽ/GOAL ገጽ/PATH ገጽ/QIM ገጽ/P
 ገጽ/GOAL ገጽ/P ገጽ/PATH ገጽ/QIM ገጽ/P ገጽ/QIM
 ገጽ/QIM ገጽ/P ገጽ/ACT ገጽ/P ገጽ/LOK ገጽ/LOK
 ገጽ/P ገጽ/OBJ ገጽ/LOK ገጽ/P ገጽ/LOK ገጽ/LOK
 ገጽ/LOK ገጽ/OBJ ገጽ/P ገጽ/P ገጽ/LOK ገጽ/LOK
 ገጽ/TIM ገጽ/JU ገጽ/OIM ገጽ/P

[illegible]

表 1 蒙古语语义格一览表

[illegible]

表 2 蒙古语语义格总的层级表

格	中类	小类	谓语中心词属性	格标	格标与谓语成分共现对	格中心词属性
主 体	动作类	施事	自发动词	᠘、'、ᠰ、ᠨ、ᠭᠠ	᠘→一般动词词缀 ᠨ→不及物使役态词缀 ᠭᠠ→及物使役态词缀 ᠭᠡ→及物被动态词缀 ᠰᠢ→特殊动词	体词（名词、代词、形动词、数词）
		当事	非自发动词	、ᠭᠠ、ᠰᠢ		
	非动作类	系事	体词+语气词	᠘、ᠭᠠ、ᠰ、ᠨ		
		领事	体词+ ᠭᠠ 体词+ ᠭᠠ	᠘		
		描述	性状词类	᠘		
		存在	存在动词	᠘		
		说明	动词	᠘、'		
客 体	宾、主格类	受事	及物动词	ᠨ、ᠭᠠ、᠘、	（无）	体词
		结果	创造类动词	᠘、ᠨ	（无）	体词
		客事	非自发动词	᠘、ᠨ	（无）	体词
	与格类	与事	及物动词	ᠭᠡ、ᠭᠠᠭᠠ、᠘	（无）	非时位体词
		受益	给予动词	ᠭᠡ、ᠭᠠᠭᠠ	（无）	非时位体词
		乘事	骑乘类动词	ᠭᠡ、ᠭᠠᠭᠠ、᠘	（无）	交通工具
		包含	包含类动词	ᠭᠡ	（无）	体词
	出格类	触事	手类动作动词	ᠰᠢ	（无）	构件类名词
		疏远	疏远类动词	ᠰᠢ	（无）	名词、代词
		向跟	特殊动词	ᠰᠢ	（无）	名词、代词
	共同格类	同事	和同类动词	ᠭᠠ	（无）	名词、代词
		关联	关系动词	ᠭᠠ（ᠰᠢᠨᠭᠡ）	（无）	体词
	主格类	属事	ᠭᠠ 存在动词	᠘	（无）	名词
		转变	ᠭᠠ ᠮ 词干动词	᠘	（无）	名词
		遭遇	遇见类动词	᠘	（无）	名词、代词
邻 体	时间类	时间	（不确定）	ᠭᠡ、᠘	（无）	时间词
	空间类	地点	（不确定）	ᠭᠡ、᠘、ᠰᠢ	（无）	处所词、方位词
		源点	（不确定）	ᠰᠢ	（无）	处所词、方位词
		终点	ᠭᠠ ᠮ 词干动词	ᠭᠡ、᠘	（无）	处所词、方位词
		路径	（不确定）	ᠭᠡ、ᠮ	（无）	方位词、处所词
		方向	（不确定）	ᠮ、ᠮ	（无）	方向词
	器材类	工具	（不确定）	ᠮ	（无）	名词
		材料	（不确定）	ᠮ	（无）	名词
	关涉类	原因	（不确定）	ᠭᠡ、ᠭᠡ	（无）	体词
		目的	（不确定）	ᠭᠡᠭᠡ、ᠮ ᠮ	（无）	体词
		基准	（不确定）	ᠰᠢ、ᠭᠡᠭᠡ、ᠭᠡᠭᠡ	（无）	体词
		依据	（不确定）	ᠭᠡᠭᠡᠭᠡ	（无）	体词
		范围	（不确定）	ᠭᠡᠭᠡ ᠮ、ᠭᠡᠭᠡ ᠮ	（无）	体词
饰 体	性状类	性状	（不确定）	᠘	（无）	副、形容、摹拟、情态词
		动状	（不确定）	ᠭᠡ、ᠰᠢ ᠮᠮ	（无）	动词
		方式	（不确定）	ᠰᠢ、ᠮ	（无）	体词
	度量类	数量	（不确定）	᠘	（无）	数、量词
		频次	（不确定）	᠘	（无）	语气词

表 3 语义格在不同体裁语料中的分布

格	序	种类	民间故事	叙事文	议论文
主体	1	施事	√	√	√
	2	当事	√	√	√
	3	系事	√	√	√
	4	领事	√		
	5	描述	√	√	
	6	存在	√	√	√
	7	说明	√		√
		总数	187	121	46
		比率	约 30%	约 36%	约 27%
格	序	种类	民间故事	叙事文	议论文
邻体	1	时间	√	√	√
	2	地点	√	√	√
	3	源点	√	√	
	4	终点	√	√	
	5	路径	√	√	
	6	原因	√	√	
	7	目的	√	√	√
	8	方向	√	√	
	9	工具	√	√	√
体	10	材料	√		
	11	基准	√		
	12	依据	√		
	13	范围			√
		总数	164	84	19
		比率	约 26%	约 25%	约 11%

格	序	种类	民间故事	叙事文	议论文
客	1	受事	√	√	√
	2	结果	√		√
	3	客事	√	√	√
	4	与事	√		
	5	受益	√		
	6	转变	√	√	√
	7	同事	√	√	
	8	触事	√		
	9	疏远	√		
体	10	属事	√	√	√
	11	乘事	√	√	
	12	向跟			√
	13	包含			√
	14	遭受	√		√
	15	关联			√
		总数	201	60	73
		比率	约 32%	约 18%	约 43%
格	序	种类	民间故事	叙事文	议论文
饰体	1	性状	√	√	√
	2	动状	√	√	
	3	方式	√	√	√
	4	数量	√	√	
	5	频次	√	√	√
		总数	79	70	31
		比率	约 12%	约 21%	约 18%

从自然语言处理研究的发展来看, 格语法的应用从语言学的语义分析拓展到认知领域, 并已经达到能够支持多种应用系统开发的程度。目前我们所进行的蒙古语语义格研究, 其目的是为解决现阶段的蒙古语词语、语句生成中语法知识所无法解决的语义知识问题, 为蒙古文信息处理语义研究打下基础。

参考文献

- [1] 清格尔泰. 蒙古语语法 [M]. 呼和浩特: 内蒙古人民出版社, 1991.
- [2] 林杏光. 词汇语义和计算语言学 [M]. 北京: 语文出版社, 1999.
- [3] 徐昌火. 归纳、演绎与句子的语义成分研究 [J/OL]. www.js-sk1.cn/xuehuil/cxh/xchh.htm
- [4] 冯志伟. 现代语言学流派 [M]. 西安: 陕西人民出版社, 1999.
- [5] 黒橋禎夫. 自然言語による知識の表現 [J/OL] www.kc.t.u-tokyo.ac.jp/kuro/paper/sakigake.doc

The analysis of Mongolian Semantic case

De·Sarina

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

Abstract: We think, as a basic research of information processing, analysis and conclusion of Mongolian sentence meaning with theory and method of case grammar are valuable work for theoretical and applied research. According to characteristics of semantic structures of different semantic components in the Mongolian language material, 4 big classes of basic semantic case including 40 sub-classes are concluded tentatively. These 4 big cases are subject, object, neighbor and decoration. Subject: a body of making movements or a theme stated, described and explained in a sentence. Object: a body involved in movements in a sentence. Neighbor: component expressing scene, subsidiary material and components of peripheral factors in a sentence. Decoration: decorative component of movements in a sentence.

Key word: Mongolian language; Semantic case; Classes of semantic case

收稿日期: 2004-9-9;

基金项目: 国家社会科学基金 (01BYY02);

作者简介: 萨日娜 (1966-), 女, 内蒙古科左后旗人, 博士生, 内蒙古社会科学院蒙古语言文字研究所助理研究员, 主要研究领域为蒙古语语义学、蒙古文信息处理。