# 面向信息处理理的蒙古语代词语义的分类初探

白玲迪

（内蒙古呼和浩特市 010021）

摘要：蒙古 文信息处理理的很多 工作迫切需要蒙古语词语语义属性的描述体系②。蒙古语代词语义的形式化描述是蒙古 文信息处理理语义处理理中的 一项基础 工程③。本 文论述了了 面向信息处理理的代词语义分类的迫切性即代词语义分类的 目的、原则、 目标及具体语义分类体系与标记集。

关键字：信息处理理；蒙古 文；代词；语义分类

## 1. 引语

蒙古语信息处理理研究 工作从上世纪 八 十年年代着 手建立语料料库开始，基本完成了了字、词处理理阶段的 工作， 目前已全 面开展句句 子、篇章处理理阶段并取得了了 一定成效。然而，语义分析是句句 子处理理阶段的重要任务之 一。蒙古语信息处理理的实际需求，词性语义分类 目前已完成动词、名词、形容词并都付诸使 用。在标注蒙古语语料料库语义信息时，代词相关信息极少，遇到很多麻烦，因此完善语料料库语义标注中代词语义研究不不可或缺。代词是指代事物、现象，具有抽象概括性。它指代的事物、现象体现在实际上下 文指定的语 言单位。代词的指代范围宽，只有从语境中才能找到它的实际意义。蒙古语代词具有语法形式格和数的形态变化，且使 用率 高。从动词谓语简单句句的动词（含 一价、 二价、三价动词）与名词语义搭配研究来看，对句句 子进 行行语义加 工时，对动词标注语义分类和价数,对名词标注语义分类和介量量(语义 角 色), 而对代词只标了了语义 角色标注,所以不不确定代词在本句句中具体代替的内容。为此，着眼蒙古 文信息处理理的实际需求，迫切需要对代词语义进 行行分类。例例如：

① 1e3 TERE/Rj{UJ} SERI/Ve2{VSIH0VY,1}+L_E/Fs31 ./Wp1 含代词句句!

② 1e3 NARA/Ne2{Nbbu11,UJ} MANDV/Ve2{ VHOSIMA,1}+L_A/Fs31 ./Wp1 含名词句句!

## 2 代词语义分类的 目的与原则

### 2.1 目的

⑴从语 言学 方 面看，语义分类是有利利于描述代词的共性和特性。词汇的搭配是语 言词汇结合 而成的现象，与语义有紧密联系。

⑵蒙古语代词与汉语代词相 比语法变化较多。 比如汉语代词中 “他、她、它”等三种指定代词，但蒙古语中只有 一词“@]<╔”来表示上面汉语的三种形式，没有阴阳性、 人和物的区别，所指的语 言单位只能在它的语境中查明。语

① 白玲迪，蒙古族，毕业于内蒙古 大学，硕 士研究 生。

② 海 银花 《 面向信息处理理的蒙古语名词语义研究》[D]，内蒙古 大学，博 士论 文，2010 年年。

③ 额尔敦朝鲁 《 面向信息处理理的蒙古语动词语义研究》[D]，内蒙古 大学，博 士论 文，2005 年年。

1

料料库出现代词，必须分析其上下 文，并标注语义分类，最后归纳代词语义搭配的特征。

⑶ 代词在词汇体系词数虽少，但使 用频率 高且复杂。它包含实体（有 无 生命体）和抽象物（时间、地点、状态、数、范围、性质）等，使机器 识别代词时必须有指定的符号表示语义分类。

⑷ 通过语义分类能明确知道代词所指代的事物并理理解句句子语义关系。代词本身没有实际意义，必须从它所在的语境中了了解代词的实际意义。

⑸代词语义分类标记有利利于完善动词与名词语义搭配研究。

① 1e5 BI/Rb {UD} CU/Sq YEHE/Ac BAYARLA/Ve2{VSEVRBA,1}+L_A/Fs31 ./Wp1

⇩

② 1e5 BI/Rb{Rbe11}{UD} CU/Sq YEHE/Ac BAYARLA/Ve2{VSEVRBA,1}+L_A/Fs31./Wp1

### 2.2 原则

本 文基于语料料库代词指代的语境意义进 行行语义分类。以下为代词语义分类的原则：

⑴ 具有层次性。代词语义类分为 大语义类， 子语义类等清晰的层次性。原因在于“ 大语义类” 是针对刻画意义相同或相似的 一些代词的共性， 而“ 子语义类（包括中类， 小类）”是为了了描述属于同 一“ 大语义类” 代词之间的差异。

⑵ 有逻辑性。在同 一层次基于统 一标准分类。

⑶ 语义分类既不不能太过复杂也不不能太过简单。如果过于复杂，语义项多导致研究复杂。太过简单，语义项 又不不能明确描述代词之间细微的差异。

在上述原则基础上进 行行代词语义分类并系统地划分代词语义，此次分类与以往分类相 比优点在于， 首先，分类对象的规模较 大从 而其结果产 生的分类体系具有多层次、多类型、多关系的特征，将会提供更更详细，更更深 入的语义信息。其次，由于分类的深度和 广度取决于语法分析的需要。应 用语义知识解决那些只靠语法知识难以解决的问题，因此为代词提供语法和语义相结合的全 面的语 言知识④。

## 3 代词语义分类 目标

⑴ 将代词语义分类作为代词属性描述基础，同时语义分类的最终 目标是 自完善蒙古语语义知识库。代词语义是整个蒙古语语 言知识库的组成部分，为机器 翻译、 文本校对、 文 音转换、 文字识别、内容提取、编码转换等应 用开发提

④ 海 银花 那顺乌 日图 《 面向“蒙古语语义信息词典”的名词语义分类体系》[J]，http://www.cnki.net.

2

供语法和语义相结合的、全 面的语 言知识⑤。

⑵ 代词语义分类不不仅揭示每 一个代词的意义， 而更更为重要的是需要对各类代词进 行行语义搭配研究，对语义搭配信息逐 一进 行行描述的系统的语义分类和语义标记集。

⑶ 代词语义分类及确定标记集为研究代词的价量量、价质的确定，价量量、价质之间对应关系的判明等代词配价提供有效知识并在代词语义分类形式化基础上为代词与其它词性词的语义搭配 生成由此导出形式规则有着重要意义。

## 4 代词语义分类

代词语义分类是代词基本词汇意义为依据兼语法信息及结合语料料库代词使 用情况为基础来划分的。本次语义分类的重点是在于，因为蒙古语代词有格和数的形态变化，所以要结合这些代词特征来决定语义分类的细略略度。由于代词应 用 广泛复杂，同 一个代词可能有不不同的归类,从 而产 生不不同的语义分类体系。

划分的代词语义类为 人称代词、反身代词、指示代词、疑问代词、不不定代词、范围代词、代动词等七 大类。以下是七 大类

分别划分的 子类。 子类含中类和 小类等。代词语义总体分类为四层，第 一层为 7 个 大语义类、第 二层为 31 个中类，第三、四层为 30 个 小类。例例如：

人称代词：基于数的变化分为第 一，第 二，第三等三中类，再分中类为单数，复数等 子类。反身代词：基于数的变化分为单数、复数等 小类。

面向信息处理理的代词语义标记是相当于代词语义形式化，用相应的符号来表示语义，语义通过符号标记才能使机器 识别。设定代词语义标记时我们参考了了国家标准《信息处理理技术——面向信息处理理的蒙古语词性分类与标记集》与动词、名词，形容词的语义标记体系。代词符号标记第 一层采 用了了本层指代名的首 音节，第 二层分别在上 一层上同样加本层指代名的 首 音节，第三、第四层分别在上 一层上加阿拉伯数字来表示。所有符号标记前加在《信息处理理技术—— 面向信息处理理的蒙古语词性分类与标记集》 里里设定的《R》字 母。最后标记 长度为第 一层 3 位数，第 二层是 5 位数，第三层是 6 位数，第四层是 7 位数。

代词语义分类体系及标记集：

## 5. 结语

面向信息处理理的代词语义分类是研究蒙古语信息处理理语义发展的基本需求。

我们对于蒙古语代词语义进 行行具体分类时，基于前 人研究成果，运 用计算语义学的理理论与 方法为依据初步完成了了此项工作。由于时间缘故和研究尚未完全成熟,还有待完善。下 一步在代词与其他词性的搭配规律律、代词配价研究中将语义分类体系不不断充实、调整完善。

**参考文献：**

[1]《现代蒙古语》，内蒙古 人 民出版社，2005 年年。

[2]秦小锋《语义形式化研究的利利与弊初探》[J],语 言 文字探索， 2009 年年，第 5 期。

[3]林林杏光 《词汇语义和计算语 言学》[M]，语 文出版社， 1999 年年。

[4]《现代蒙古语 100 万词级语料料库》，内蒙古 大学蒙古学学院。