面向“蒙古语语义信息词典”的名词语义分类体系*

海银花　那顺乌日图

内蒙古大学蒙古学学院 呼和浩特市 010021

E-mail: haiyinhua@imnu.edu.cn

摘　要：对于蒙古语名词进行语义分类是我们研发“蒙古语语义信息词典”时首要完成的基础工作。信息处理用蒙古语名词语义分类体系的构建，可为蒙古语语句自动处理提供语法和语义相结合的、全面的语言知识，它有助于提升蒙古文信息处理水平。本文简要介绍我们对名词语义分类的研究实践，着重说明分类的基础资源以及分类的依据、标准和具体分类体系。

关键词：蒙古语　名词语义分类体系　“蒙古语语义信息词典”

"Mongolian Semantic Information Dictionary"-Oriented Noun Thesaurus System

Hai.Yinhua　Nasun-urt

Academy of Mongolian studies, Inner Mongolia University, Huhhot 010021

E-mail: haiyinhua@imnu.edu.cn

Abstract: The classification of Mongolian noun in the light of thesaurus system is the fundamental work we have to accomplish first when we develop the “Mongolian Semantic Information Dictionary”. The construction of information processing with Mongolian noun thesaurus system have provide a combination of a comprehensive knowledge of syntax and semantic for Mongolian automatic processing, which will raise the level of Mongolian language information processing. Through the introduction of our work on the research of Mongolian noun thesaurus system, this paper explores foundation resources, standard, criteria, and its system of the Mongolian noun thesaurus.

Key words: Mongolian language, The thesaurus system of Mongolian noun, "Mongolian semantic information Dictionary"

一、基础资源

名词作为蒙古语三大“开放性”词类之一，对它进行语义分类描述是构建蒙古语语义知识库的前提。为了给计算机自动分析和自动生成提供更全面、更深入的语义知识，建立一项与语言工程应用相结合的、面向语义知识库的语义分类体系是我们研发“蒙古语语义信息词典”的基础。其中名词语义分类是该课题的重要组成部分，是我们首要完成的分支工作。

自2008年内蒙古大学蒙古学学院在国家自然科学基金的资助下着手研制“蒙古语语义信息词典”，其目的是在现有的语言资源—“蒙古语语法信息词典”的基础上构建“蒙古语语义信息词典”，为机器翻译、自动校对、文本检索等应用系统中蒙古语语句的自动处理提供语法和语义

*此项研究得到国家自然科学基金项目“《蒙古语语义信息词典》的设计与实现”（项目号：60873084）”的资助。

作者简介：海银花（1981—），女，蒙古族，内蒙古大学蒙古学学院博士生，研究方向为蒙古文信息处理。那顺乌日图（1959—），男，蒙古族，内蒙古大学蒙古学学院教授，博士生导师，研究方向为蒙古文信息处理。

相结合的、全面的语言知识。

自2000年内蒙古大学蒙古学学院在国家社会科学基金和国家自然科学基金，教育部、国家语委的项目资助下研发了“蒙古语语法信息词典”。经过近几年的努力，语法信息词典与其管理平台基本成形。“蒙古语语法信息词典”由“总库”和19个分库组成。“蒙古语语法信息词典名词分库”作为其有机组成部分以36个属性字段详细描述了蒙古语14105个名词的各种语法属性。该分库中通过设置“前面能否带数词定语”、“前面能否带数量名定语”、“前面能否带动词定语”等属性字段描述了名词的搭配特征方面的信息。但是本词典性质上不属于纯语义词典，所以它未能涵盖蒙古语名词更深入的语义属性。蒙古语名词语义分类工作是基于“蒙古语语法信息词典名词分库”而进行的探索性研究课题。“名词分库”作为基础资源，其14105个词条已成为语义分类的主要对象。我们参照现有各家语义分类体系的基础上，对蒙古语名词语义进行了更趋合理、较细的分类，旨在通过语义信息解决那些计算机在自动处理词语时所遇到的仅靠语法规则难以解决的问题。

二、名词语义分类的必然性

名词语义分类信息在多义词义项判断、词语搭配、短语结构的正确分析和生成等蒙古语各种应用层面上均有重要作用。我们构建“蒙古语名词语义信息词典”时主要刻画蒙古语名词之间“语义”层面上的共性和差异，因此采用了“语义分类+属性描述”的方式来标记每一个名词跟其他名词之间的异同。换言之，以语义场理论为基础的语义分类法和基于复杂特征集的属性描述法相结合的语义描述方法：我们先进行语义分类，然后基于语义分类体系运用属性描述形式表示名词的语义属性。我们用属性字段形式描述其语义类信息，同时还设置其他语义特征的属性字段来刻画名词的全面的语义知识。

名词语义分类的当前目标是：

（1）将名词语义分类作为名词属性描述的基础，而不仅仅作为“蒙古语名词语义信息词典”的属性字段之一。在“蒙古语名词语义信息词典”中以“大语义类”和“子语义类”两个属性字段描述每个名词的语义类属，例如：“HONI”[①]（绵羊）的“大语义类”字段中填写“AMITAN”（动物类）的相关标记，其“子语义类”字段中填写“TABVN HVSIGV MAL”（五种牲畜类）的相关标记等。其中我们设置“大语义类”和“子语义类”两个属性字段的原因在于“大语义类”是针对刻画意义相同或相似的一些名词的共性，而“子语义类”是为了描述属于同一“大语义类”名词之间的差异。

（2）名词语义分类不仅揭示每一个名词的意义，而更为重要的是在语义分类和语义标记的基础上，对各类名词进行语义搭配研究，对语义搭配信息逐一进行描述。我们在“蒙古语名词语义信息词典”中通过描述每个名词的所属语义类和能够与其搭配的名词的所属语义类，来刻画名词与名词之间的搭配规律，即以语义属性字段的形式来描述每一个名词的搭配特征，刻画词语及其语义属性的二维关系。

（3）名词语义分类为名词的价量、价质的确定，价量、价质之间对应关系的判明等名词配价研究提供支持。在名词语义分类的基础上，确定名词的价量、价质等属性，也为名词的语义搭

① 文中的蒙古文以拉丁文转写形式提供。

配生成提供有效控制。

从信息处理的角度对蒙古语词语进行语义分类，制定相应的标记体系，并将其作为词语语义属性描述的基础规范是蒙古文信息处理中新近开展的工作，所以其基础还显薄弱。因此我们在充分利用蒙古语名词语义分类体系前人研究成果的同时，借鉴和参考了英语、汉语等其它语言的词语语义分类体系研究成果。

其中目前能够参考的蒙古语语义研究方面的成果有：

（1）由内蒙古大学承担完成的国家社会科学基金项目“面向信息处理的蒙古语语义研究”（项目号为 01BYY025）。该项目以蒙古语的高频词和真实语料为主要数据来源，对蒙古语名词、形容词和动词进行了较为粗浅的语义分类：将名词分为“HEREG”（事）、“BODAS”（物）、“HEM HEMJIGUR”（度量）、“UILE YABVDAL”（动作）、“VYVN VHAGAN”（智慧）、“CAG ORON”（时位）等 6 个大类，14 个中类以及诸多小类，最终分出 7 个层次 130 多个小类；

（2）姜迎春的“面向信息处理的蒙古语形容词语义研究”（2003 年）[②]中依据名词和形容词的语义搭配情况，将名词分为“HEREG”（事）、“BODAS”（物）、“CAG”（时间）、“ORON”（空间）等 4 个大类、8 个中类、23 个小类，继而将其分为更小的子类，最多的已达到 7 个层次；

（3）包金兰的“蒙古语复合形容词和名词的语义搭配研究”（2005 年）[③]立足于名词和复合形容词的搭配规律，将名词分为“BODAS”（物） 和“HEREG YABVDAL”（事件）等 2 个大类、13 个中类、117 个小类等等。

但由于各家语义分类体系的目的及应用范围不同，对同一个名词可能有不同的定义与归类，从而产生不同的语义分类体系。例如将同一个词“AMITAN”（动物）在第（1）语义分类体系中再分为“ARIYATAN”（兽类）、“JIGURTEN ”（禽类）两个子类，第（2）语义分类体系中分为“ARIYATAN”（兽类）、“GOROGESU”（鹿类）、“BAG_A AMITAN”（小动物类）、“JIGURTEN”（禽类）、“JIGASV”（鱼类）、“HOROHAI SIBAJI”（虫类）、“MOLHOGCI”（爬行类）、“GER-UN TEJIGEBURI”（家禽类）等 8 个子类，而在第（3）语义分类体系中则分为“TABVN HVSIGV MAL”（五种牲畜类）、“BVSVD GER-UN TEJIGEBURI”（其他家禽类）、“BVSVD AMITAN”（其它动物类）3 个子类等等。

我们此次进行分来的目标是基于蒙古语已有的研究成果，制定一套能够满足“蒙古语语义信息词典”本身的体系结构和蒙古语语言模型需求的，既有科学性，又具操作性的语义分类体系。通过系统地划分名词语义并详细地描述各个子类的种种特征来编纂一部能够为蒙古文信息处理各种应用系统提供名词语义支持的知识库。所以此项分类与以往分类的相对而言其优点在于，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，将会提供更详细，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，应用语义知识解决只靠语法知识难以解决的那些问题，因此为名词提供语法和语义相结合的全面的语言知识。最后，我们建立蒙古语名词语义分类体系及语义关系描述体系是面向“蒙古语语义信息词典”的，因此此项分类体系以属性字段形式被容纳在语义知识库中，最终成为独立的词典数据库。

三、名词语义分类的依据和标准

② 内蒙古大学硕士学位论文，2003 年。

③ 内蒙古师范大学硕士学位论文，2005 年。

1. 依据

我们对于蒙古语名词语义进行具体分类时，基于前人研究成果，运用计算语义学的理论与方法，以名词的基本词汇意义作为主要依据初步完成了此项工作。这里以名词的基本词汇意义作为主要分类依据的原因在于名词的多义性。基本词汇意义是对客观事物、现象的性质、特征、功能等进行概括、抽象地描绘、解释等而形成的词的一种最基本的意义内容。词的多义性是由基本词汇意义通过词义的扩大、缩小、类比、转义等方式引申而产生的几个相互有联系的意义。对计算机处理自然语言来说多义性是义项划分和义项排列的核心，涉足到词义消歧、话语分析等问题。我们按照每个名词的基本词汇意义来划分或确定其语义类不仅是为了使分类简洁，避免冗余信息，而且也为名词的配价研究奠定了基础（因为名词的价量、价质是同样按照名词的基本词汇意义来确定）。蒙古语中绝大多数名词具有两个或两个以上的义项，不同的义项使同一个词可属于不同的语义类。例如名词“ONGGOCA”有（1）（喂饮牲口的）槽，如 GAHAI-YIN ONGGOCA（猪食槽）；（2）船，如 DARBAGVLTV ONGGOCA（帆船）；（3）畦，如 ONGGOCA TATAJV NOGOG_A GARGAHV（开畦种菜）；（4）（洗澡、洗衣等用的）的池子，如 VHIYALG_A-YIN ONGGOCA（洗脸池子）④等四个不同的义项。因此它以第（1）义项与“TOGOG_A、DEBUR_E、ABDAR_A、$UGUI、HORGO”等名词共同属于“SABA EdLEL”（器皿）类；以第（2）义项与“NISHEL、MASIN、HASAG、SEUHE、TERGE”等名词共同属于“JAM HARILCAGAH-V BAGAJI”（交通工具）类；以第（3）义项与“ALCAM、BAGLAG_A、JVRBVS、BVGVJI、HELHIY_E、”等名词共同属于“HEM HEMJIGUR”（度量）类；同时以第（4）义项与“ALCIGVR、TOLI、TONGPONG、SABVNG、SAM”等名词共同属于“VHIYALG_A-YIN HEREGSEL”（洗涤用具）类等。蒙古语名词的此种多义性使分类体系将会变得错综复杂甚至导致混乱。我们一方面为了避免此种可能性，比较科学地理顺名词的不同义项和名词语义类之间的复杂关系，将基本词汇意义作为对于名词进行语义分类的主要依据。

2. 标准

我们建立蒙古语名词语义分类体系时不是用单一标准分类到底，而是组合使用了多个标准。首先按辩证唯物主义联系观把语义类第一层为“HEREG”（事）、“BODAS”（物）、“VYVN VHAGAN”（智慧）、“CAG”（时间）、“ORON”（空间）、“UILE HODELGEGEN”（动作）、“HEM HEMJIGUR”（度量）七大类，即顶层（一分为七）：从第二层起由于分类的标准不同而各类的层次就变得不一致。“HEREG”（事）根据其包含的领域分为“VLVS TORO”（政治）、“AJV AHVI”（经济）、“HAVLI CAGAJA”（法律）、“CERIG DAYIN”（军事）、“UILECILEGE”（服务）、“JIGVLCILAL”（旅游）、“EMCILEGE”（医疗）、“$ASIN SVRTAHVN”（宗教）、“NAYIR NAGADVM”（娱乐）、“NER_E TOMIY_A”（名词术语）等 10 个子类；“BODAS”（物）根据其“有无形状”、“有无颜色”、“有无质量”等基本义素分为“BODATV BODAS”（具体物）和“HEYISBURI BODAS”（抽象物）两类，继而把“BODATV BODAS”（具体物）以它的“有无生命”作为标准分为“AMIDV BODAS”（生物）和“AMIGUI BODAS”（非生物）两类；依据“CAG”（时间）的存在形式作为标准把它分为“HARICANGGVI CAG”（相对时间）和“HARICALASIGUI CAG”（绝对时间）；“ORON”（空间）是否具体存在作为其标准把它分成“BODATV ORON”（具体空间）和“HEYISBURI ORON”（抽象空间）；将“VYVN VHAGAN”（智慧）的心理学中的命名作为标准分了“SEREL”（感性）、“TANIHVI”（认识）、“VHAMSAR”（意识）、“SEDHILGE”（情

④ 内蒙古大学蒙古学研究院蒙古语文研究所：《蒙汉词典》（修订本），内蒙古大学出版社，1999 年，第 192 页

感)、“JANG CINAR”(性格)等子类;根据“UILE HODELGEGEN”(动作)的性质分为“AHVI BAYIDAL”(形势)、“JORILTA UILE”(目的)、“HARICAGATV UILE”(关系)、“SILJILTE UILE”(移动)、“HEREGUL JIGVRAGAN”(纠纷)、“AYANDAGAN-V UILE”(自然动作)、“JANG UILE”(习俗)、“FIJIVLVGI”(生理)、“HVBIRALTA”(变化)等;把“HEM HEMJIGUR”(度量)以它的来源作为主要标准分为“VVGAL HEMJIGUR”(固有的)和“JIGELEGE HEMJIGUR”(借用的)两类等等。我们明确了上述依据和标准的基础上把名词语义分类的深度和广度都控制在一定范围之内,分类避免过于复杂,分类体系倾向于简单;分类的逻辑性较强等原则指导了该分类体系。

四、名词语义分类体系⑤

NIGE、HEREG(一、事)

1. VLVS TURU(政治):VLVS, GURUN, TURU, HVRLADVGAN
2. AJV AHVI(经济):ARALJIY_A, HORONGGE, TURIYESU, CALING
3. HAVLI CAGAJA(法律):EREGUU, JARGV, SIDHEL, SIGULTE
4. CERIG DAYIN(军事):DAYIN, BAYILDVGAN, TVLVLDVGAN, HITVGAN
5. UILECILEGE(服务):ASARAMJI, TEDHUBURI, DAGADHAL
6. JIGVLCILAL(旅游):JIGVLCILAL, AYAN, AYALAL, TEGEGEBURI
7. EMCILEGE(医疗):MESE, EMCILEGE, TARILG_A, HANALG_A
8. $ASIN SVRTAHVN(宗教):SOSOG, SONESU, TARNI, ISLAM
9. NAYIR NAGADVM(娱乐):BUHE, VRVLDVGAN, TOGLAGAM, NAYIR
10. SINJILEHU VHAGAN-V NER_E TOMIY_A(科学术语):ABIY_A, SILUG, ALGeBRA, PROgRAM

HOYAR, BODAS(二、物)

1. BODATV BODAS(具体物)
   1.1 AMIDV BODAS(生物)
      1.11 HOMON(人)
         1.111HVBI HOMON(个人)
            1.1111YERUNGHEI NEREYIDUL(总称):HOMON
            1.1112TVSHAI NEREYIDUL(专称):ER_E, OGEDEI, SEGEGETEN, EJI
         1.112 HAMTV NEYITE(集体):TATAR, TOBED, NIGECE,,HAMTVLIG
      1.12AMITAN(动物)
         1.121 YERUNGHEI NEREYIDUL(总称):HOHOTEN, SEGERTEN, NVGVSVTAN, EBERTEN
         1.122 TVSHAI NEREYIDUL(专称):ARSLAN, ELIY_E, TVVJA, MOGAI
      1.13 VRGVMAL(植物)
         1.131BURILDUHUN(构成):NABCI, UNDUSU, ESI, UR_E
         1.132TOROL JUIL(类型):NARASV, SIRALJI, SARAN_A, HEMHE
   1.2 AMIGUI BODAS(无生物)
      1.21 BAYIGALI-YIN BODAS(自然物):NARA, ELESU, GOROH_A, ALTA

⑤ 由于篇幅限制,文中未能列出更详细的子类。

1.22 JOHIYAMAL BODAS（人工物）：BAYISING, JIDA，GABA, DEBEL

2. HEYISBURI BODAS（抽象物）

2.1VCIR JUI（事理）：HEBCIY_E, VDH_A, JORILG_A, UILEDUL

2.2YOSO MVRAL（道德）：HARICAGVLG_A, ALBA，ALDAG_A, EY_E

2.3 SVRGAN HOMOJIL（教育）：HICIYEL, HOMOJIL, SVRVLG_A, SVRGAL

2.4 SVRAG JANGGI（信息）：MEDEGE, SVRAG, CIMEGE, MEDEGELEL

GVRBA, VYVN VHAGAN（三、智慧）

1. SEREL（感性）：MEDEREMJI, TAGALAMJI, HALAGV, DAGARADASV

2. SEDHILGE（情感）：GVNIG, BAYASVL, HILING, EMIYEL

3. VHAMSAR（意识）：ABIYAS, VR_A, VHAGAN, JORIG

4. TANIHVI（认识）：UJELTE, HUSELENG, OYILAGALTA, BODOLG_A

5. JANG CINAR（性格）：JANG, AGASI, ABVRI, OBOR

DORBE, CAG（四、时间）

1. HARICANGGVI CAG（相对时间）：VRJIDUR, ODO, NOGOGEDUR, IREGEDUI

2 .ARICALASIGUI CAG（绝对时间）：ORLOGE, HVGVCAG_A, EREN，HABVR

TABV, ORON（五、空间）

1. BODATV ORON（具体空间）：MVJI, MAYIHAN, CAHAR, HABVRJIY_A

2. HEYISBURI ORON（抽象空间）：HOBEGE, BAYIRISIL, TALABVR, SAGVCA

JIRGVG_A, UILE HODELGEGEN（六、动作）：VDVMSIL, JAMNAL，HAMJILCAG_A, OROSIL

DOLOG_A, HEM HEMJIGUR（七、度量）

1 VVGAL（固有的）：ALCAM, GODOLI, BAGCA, DVSVL

2 JIGELELGE（借用的）：ROBLI, JING, dVLLAR, $ANG

结束语

我们通过查词典、分析语料等手段“蒙古语语法信息词典名词分库”大部分常用名词进行语义归类，但还有相当一部分词语尚无法确定其类属。由于时间缘故和研究尚未完全成熟，这里还有很多有待完善的地方。所以我们在下一步工作—名词与名词的搭配规律、名词配价研究中将语义分类体系不断充实、调整。

参 考 文 献

[1] 陈小荷. 一个面向工程的语义分类体系. 语言文字应用, 1998, 2.

[2] 林杏光. 词汇语义和计算语言学. 语文出版社, 1999.

[3] 陈群秀，等. 信息处理用现代汉语语义分类词典的设计与实现. 中国中文信息学会成立二十周年学术会论文集. 清华大学出版社，2001.

[4] 那顺乌日图. 关于面向信息处理的蒙古语语义研究. 内蒙古大学学报，2002, 5.

[5] 马勇腾, 亢世勇. 新编同义词词林语义分类体系. 第三届学生计算语言学研讨会论文集，2006.

[6] 王惠, 等. 现代汉语语义词典”的结构及应用. 语言文字应用, 2006, 1.

[7] 海银花, 等. “蒙古语语法信息词典”的新进展. 民族语文国际学术研讨会，北京:2007.