蒙古语语义信息词典（SIKM）的研发与应用

海银花
（内蒙古大学 蒙古学学院，呼和浩特 010021）

摘要:“蒙古语语义信息词典”(SIKM)作为一部语言知识库已成为整个蒙古语语言资源的组成部分。自 2009 年通过国家自然科学基金和自治区相关部门的资助项目，该词典的研发取得了阶段性成果，也初步应用到了一些系统和研究中。SIKM 现已收录 57，000 多词条，涵盖 4 个词典库。其中包含全部词语的 1 个“总库”，名词、形容词、动词等 3 个词类各建一个数据库。每个数据库文件都详细刻画了词语及其语义属性的二维关系。目前已完成 57,000 多条词语的语义分类和属性描述，其初步应用获得了很好的效果。文中介绍词典规模及结构，语义分类体系，属性字段描述及其初步应用等内容。

关键词：蒙古语，语义信息词典，蒙古文信息处理，研发，应用
中图分类号： **文献标识码：**

Development and Application of Semantic Information Knowledge-base in Mongolian

HaiYinhua
(Inner Mongolia University, School of Mongolian Studies, Hohhot, 010021China)

Abstract: "the Semantic Information Knowledge-base in Mongolian" (SIKM) is an integral part of the overall Mongolian language resources as a knowledge base. The knowledge base achieved initial results with the project of National Natural Science Foundation and the relevant departments of Inner Mongolia Autonomous Region Since 2009, and had the initial application on some systems and researches. In SIKM there are 57,000 entries, covering four databases, which contains all the words in a "general base" and other sub-bases of noun, adjective and verb. Each database file has a detailed description on words and its semantic attributes with a two-dimensional relationship. At present, we completed the semantic classification and attribute descriptions on more than 57,000 words, and also had a satisfactory effect on its initial application. The paper will dissertate the scale and structure of the knowledge base, semantic classification, the description of attribute and its initial application in variety of system.

Keywords: Mongolian, Semantic Information Knowledge-base, Mongolian Information Processing, Development, Application

1 引言

众所周知，实现基于自然语言的智能信息处理，计算机需要理解大量的语义知识。随着 Internet 的迅猛发展，云计算(Cloud Computing)迈向大数据（Big Data），建设大规模的语义知识资源已成为一个热门课题。国内外语义资源建设的最新趋势表明信息搜索技术对语义理解要求越来越高，因此建设语义资源的重要性和必要性日益显著。

2009-2012 年间课题组依托单位在国家自然科学基金的资助下承担“蒙古语语义信息词典的设计与实现”（60873084）项目，着手研制“蒙古语语义信息词典”，目的是在语法分析的基础上，为计算机自动分析蒙古语句子和生成外语句子提供更翔实而深

基金项目：国家社科基金项目（12CYY062），国家自然科学基金重点项目（61032008）（与清华合作），国家自然科学基金项目（60873084），内蒙古自治区蒙古语言文字信息化专项扶持资金项目（2012399）
作者简介：海银花（1981—），女，蒙古族，内蒙古大学博士、讲师，主要研究方向为蒙古文信息处理。

入的语义知识。这是语义词典的初期开发阶段，期间已完成名词、形容词和动词等 3 大词类常用词语的语义分类和搭配信息描述，取得了阶段性成果。但是，传统蒙古语语义学研究底子很薄这一事实，使语义词典的研发变得一个长期性的语言工程。自 2013 年 12 月 SIKM 的二期研发受到了内蒙古自治区民族事务委员会的“蒙古语言文字信息化专项扶持项目”(2012399)的支持，并着手开发了词典总库。目前 SIKM 可以为短语结构辨别、词汇歧义消解、语料库标注等研究提供基础知识。

SIKM 现已收录 5.7 万多词条，涵盖 4 个词典库。其中包含全部词语的 1 个“总库”（图 1），名词、形容词、动词等 3 个词类各建一个数据库。每个数据库文件都详细刻画了词语及其语义属性的二维关系。各词典库囊括的词条数和信息量及其总信息量（信息量=记录数*属性字段数）计算如表 1 所示。“总库”与各类词典库可以通过“词语”、“音标”、“词类”等 3 个关键字段进行相互连接。

表1　SIKM信息量计算表

库名	记录数	属性字段数	信息量	总信息量
总库	57495	16	919,920	1932,460
名词库	14105	26	416,000	
形容词库	11025	21	231,525	
动词库	32365	11	356,015	

2　SIKM 的语义分类体系

课题组通过前期所承担的国家社科基金“面向信息处理的蒙古语语义研究”项目(ZC105-008)对于蒙古语词语进行过大致的语义分类。但是，由于当时的理论和技术基础所限，该分类未能覆盖 SIKM 词典库所收录的蒙古语常用词语，未能满足蒙古文信息处理的深层需求。SIKM 语义分类是现有词典库所收录词条数量的基础上进行，并且其深度与广度完全取决于语义分析的需要。

SIKM 二期工程参照现有各家语义分类体系的基础上，针对语料库加工、汉蒙机器翻译中语言分析的实际要求，同时考虑到将来便于与蒙古文 Wordnet 兼容，或者与“多义词词典”、“同形词词典”等已有的各种语义知识库共享资源而实现的。 例如，蒙古语名词语义分类体系是通过前期课题-“面向信息处理的蒙古语语义研究”而研制出的名词语义分类成果而开发的，其大致框架如下所列。

1.HEREG(时间)(Nh)
　1.1 VLVS TORO(政治)(Nhv)
　1.2AJV AHVI(经济)(Nha)
　1.3HAVLI CAGAJA(法律)(Nhh)
　1.4CERIG DAYIN(军事)(Nhc)
　1.5UILECILEGE(科学)(Nhu)
　1.6JIGVLCILAL(旅游)(Nhj)
　1.7EMCILEGE(医疗)(Nhe)
　1.8$ASIN SVRTAHVN(宗教)(Nh$)
　1.9NAYIR NAGADVM(娱乐)(Nhn)
　1.10NER_E T0MIY_A(科学术语)(Nhs)
2.B0DAS(物)(Nb)
　2.1B0DATVB0DAS(具体物)(Nbb)
　　2.1.1AMIDVB0DAS(生物)(Nbb)
　　2.1.1.1HOMON(人)(Nbba1)
　　2.1.1.2AMITAN(动物)(Nbba2)
　　2.1.1.3VRGVMAL(植物) (Nbba3)
　　2.1.2AMIGUIB0DAS(有生物)(Nbbu)
　　2.1.2.1BAYIGALIB0DAS(自然物)(Nbbu1)
　　2.1.2.2J0HIYAMALB0DA(人工物)(Nbbu2)
　2.2HEYISBURIB0DAS(抽象物)(Nbh)
　　2.2.1 VCIR JUI(logic)(事理)
　　2.2.2 Y0S0 MVRAL(道德)(Nbhy)
　　2.2.3SVRGANHOMOJIL(教育)(Nbhh)
　　2.2.4SVRAGJANGGI(信息)(Nbhs)
3. 0YVN VHAGAN(智慧)(Nv)
　3.1 SEREL(感知)(Nvs)
　3.2 SEDHILGE(情感)(Nvh)
　3.3 VHAMSAR(意识)(Nvv)

3.4 TANIHVI(认识)(Nvt)

3.5 JANG CINAR(性格)(Nvj)

4. UILE HODELGEGEN(动作)(Nu)

4.1 JANG UILE(习俗)(Nu1)

4.2 HEREGUL JIGVRGAN(争吵)(Nu2)

4.3GUYUDELYABVDAL(走势)(Nu3)

4.4 HEMJIY_E CINAR(规格)(Nu4)

4.5FIZILAGI-YINUILE(生理活动)(Nu5)

4.6 DAGV CIMEGE(声响)(Nu6)

4.7HARICAG_AH0LB0G_A(关联)(Nu7)

4.8 HVBIRALTA(变化)(Nu8)

4.9 NEGUDEL SILJILTE(迁移)(Nu9)

4.A J0RILTA AJILLAG_A(主张)(NuA)

4.B AHVI BAYIDAL(境况)(NuB)

4.C ARG_A B0L0LCAG_A(可能性)(NuC)

5. CAG(时间)(Nc)

5.1HARICANGGVICAG(相对时间)(Nch)

5.1.1 ONGGEREGSEN(过去)(Nch1)

5.1.2 0D0(现在)(Nch2)

5.1.3 IREGEDUI(将来)(Nch3)

5.2HARICALASIGUI CAG(绝对时间)(Ncg)

5.2.1CAG-VNHVBIYARI(分布)(Ncg1)

5.2.2 HVGVCAG_A(分配)(Ncg2)

5.2.3 VLARIL(季节)(Ncg3)

6. 0R0N(地点)(N0)

6.1 B0DATAI 0R0N(具体地点)(N0b)

6.1.1BAYIGALI-YIN 0R0N(自然地点)(N0b1)

6.1.2NEYIGEM-UN 0R0N(社会地点)(N0b2)

6.2 HEYISBURI 0R0N(抽象地点)(N0h)

7. HEM HEMJIGUR(度量)(Nq)

7.1VVGAL HEMJIGUR(固有度量)(Nqv)

7.2JIGELEGE HEMJIGUR(借用)(Nqj)

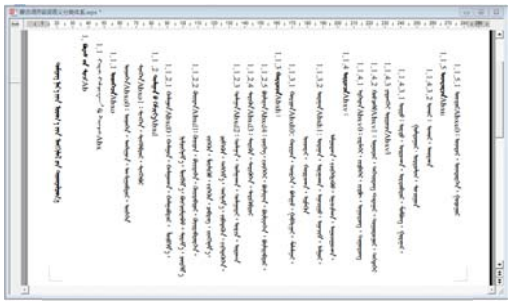

图1 形容词语义分类体系及标记集样本

目前，SIKM 语义分类涵盖名词、形容词和动词，根据每个词语的基本词汇意义进行归类，并制定其相关标记而获得的。它主要包括 7 大类、191 个子类、9 个层次的名词语义分类体系、6 个大类、217 个子类、2 个层次的形容词语义分类体系（图 1）和 5 个大类、121 个子类、4 个层次的动词语义分类体系（图 2），各自包括的大类、子类和层次数量统计如表 3 所示。

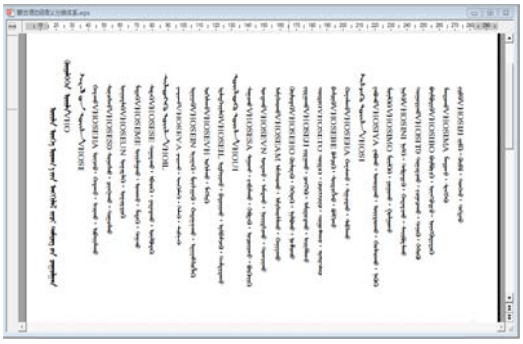

图2 动词语义分类体系及标记集样本

3 SIKM的属性描述

SIKM 是一部尽可能从多角度、多层次上描述现代蒙古语常用词语语义特征的知识库。SIKM 的初期开发阶段，由于传统蒙古语语义学研究基础特别薄弱和当时蒙古文信息处理的各种技术尚未成熟等理论和技术方面的限制，无法实现整体描述整个蒙古语词语的语义特征这一语言工程。所以只能从单独的词类切入描述其语义属性，从而 SIKM 的开发工作从名词、形容词和动词等 3 个分库开始的，未能涉足到“总库”的设计和研发。直到 2013 年，语义词典的二期工程在开发 3 个词典库的技术和经验基础上，着手研发了其“总库”（图 4）。

图 3 SIKM 总库样本

SIKM 所包含的属性描述大致可归纳为以下 5 种：

1）连接信息：这是从《蒙古语语法信息词典》（2012 版）中直接继承而来的“蒙古文、音标、词类”等 3 个关键字段。这些字段的链接不仅保证语义词典收词的规范

性，音标与词性标注的准确性，而且经过相互配合使用，使系统获得更加完备的语法和语义知识。

2）基本语义信息：词语本身的一些基本语义特征，如该词的义项、同形词、释义、同义词、反义词、熟语等。如表 2 对“AMA”这一词的 5 个义项描述，可以为词义消歧和词义研究提供丰富的知识；

3）语义类信息：根据 3 大词类语义分类体系填写的词语所属的语义类属性，包括大语义类和子语义类。

4）搭配信息：描述一个词语与其他词语发生语义联系的功能，主要包括前搭配、后搭配和并列搭配；

5）配价信息：刻画词语能够支配多少名词性成分的配价数量，例如，零价、一价、二价、三价和四价等；施事、客事、主体、客体、与事等配价质量；这是开发 SIKM 的重点内容，可直接服务于计算机语义自动分析（如同表 3 所示）。

表2 SIKM中不同义项的描述样本

词语	音标	词类	汉文	义项	同形	释义	义类	同义词	反义词	熟语
	AMA	Net1	嘴	1		进食、说话器官	器官			AMA ALDAHV
	AMA	Net2	边	2		敞开部分	构件		SEGUL	AMA SEGER
	AMA	Net2	闲谈	3		闲言	属性	HELE		AMA NIGETEI
	AMA	Net1	人口	4		统计人数单位	度量	HOMON		ERUHE AMA
	AMA	Net2	接近	5		接近的时期	时间	UY_E		AMAN DEGER_E

表3 “名词库”中的一价名词配价信息描述样本

名词	语义类	价量	价质	语义关系	实例
EJI（母亲）	亲属关系	1	领事	领事-属事	MINV EJI（我的母亲）
NABCI（叶子）	植物构件	1	整体	整体-部分	M0D0N NABCI（树叶）
ONOR（味儿）	属性	1	来源	来源-结果	AYIRAG-VN ONOR（酸奶味）
DABHILTA（奔跑）	动作	1	实体	实体-动作	M0RIN DABHILTA（马的奔跑）

除此之外，在语义词典库中还增添了一些必要的属性字段。例如，在“名词库”中增加了（1）西里尔文（kiril）、（2）英文（English）、（3）其他搭配、（4）语义关系（VDHARICAG_A）等属性。其中“语义关系”字段是针对有价名词的语义属性而设置，要填写当前词条能够构成的语义关系的代码。例如，“  ”(枝)的相关字段中填置“整体-部分”语义关系的代码“buh-bur”，表示该名词与其配价之间构成一种“整体—部分”关系（如同M0D0N MOCIR   ：树枝）（如同表3所示）。这种属性对于词语搭配、短语语义关系判断、歧义消解和语料库加工等自然语言处理各种领域的意义深重。

目前，SIKM的语义属性描述比较完整尤其名词库（图4）、形容词库（图5）、动词库（图6）的全部属性项目都已按照要求填满信息。“总库”包括16个字段，其属性字段说明如表4所示，各分库囊括的字段数量及其名称概括如表5所示。

图 4 “名词库”样本

图5 “形容词库”样本

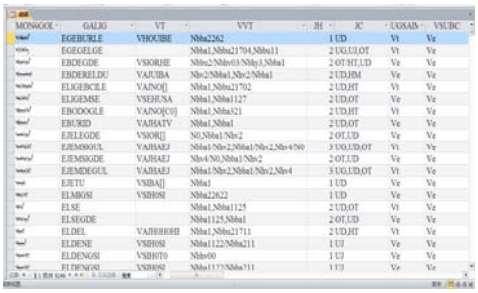

图 6　“动词库”样本

总体而言，通过 SIKM 可以得知一个词的全面语义信息，例如，名词“ᠤᠨᠳᠤᠰᠤ”（根）在词典中的属性描述有：【1】序号：[121]；【2】蒙古文：ᠤᠨᠳᠤᠰᠤ【3】音标：[UNDUSU]；【4】西里尔文：[үндэс]；【5】英文：[root]；【6】词类：[不可数名词]；【7】释义：[①高级植物器官之一。②事物根源][7]；【8】义项：1；【9】同形词：[ ]；【10】近义词：[VG]；【11】熟语：[UNDUSU-TEI-DEGEN VRGVDAG,UR_E-TEI-DEGEN DELGEREDEG]（有根有苗）；【12】大语义类：[生物]；【13】子语义类：[植物构件]；【14】前搭配语义类：[植物]；【15】前搭配形式：[名词属格、动态词尾、属格+动态词尾]；【16】前搭配实例：[EBUSUN UNDUSU、CECEG-UN UNDUSU、M0D0N-V UNDUSU]（ᠥᠪᠡᠰᠦᠨ ᠤᠨᠳᠤᠰᠤ、ᠴᠡᠴᠡᠭ ᠤᠨ ᠤᠨᠳᠤᠰᠤ、ᠮᠣᠳᠣᠨ ᠤ ᠤᠨᠳᠤᠰᠤ）（草根、花根、树根）；【17】后搭配语义类：[植物]；【18】后搭配形式：[形容词]；【19】后搭配实例：[UNDUSU MAGVTAI M0D0]（ᠤᠨᠳᠤᠰᠤ ᠮᠠᠭᠤᠲᠠᠢ ᠮᠣᠳᠣ）（烂根树）；【20】并列搭配语义类：[植物构件]；【21】并列搭配形式：[名词主格]；【22】并列搭配实例：[UNDUSU NAMAG_A]（ᠤᠨᠳᠤᠰᠤ ᠨᠠᠮᠠᠭ）（枝根）；【23】其他搭配：[NARIN UNDUSU]（细根）；【24】价量：[1]；【25】价质：[整体]；【26】语义关系：[整体-部分]。

表 4　SIKM“总库”属性字段说明表

序号	属性	标记	属性取值说明
1	序号	ID	自动生成的数字，表示 SIKM 包含的词条具体数目
2	蒙古文	MONGGOL	词语蒙古文书写形式。
3	拉丁转写	GALIG	蒙古语词条的拉丁转写形式。例如，“ᠮᠠᠭᠲᠠᠴᠢ”（马夫）的对应的 GALIG 字段中填置“AGTACI”。
4	词类	UGAYIMAG	词语的所属词类标记。例如，“AGTACI”（马夫）对应的“UGAYIMAG”字段中填置“Ne1”，表示可数名词。
5	汉文	HITAD	每个蒙古语词语的汉文翻译。例如，“AGTACI”对应的“汉文”字段中填置“马夫”。
6	西里尔文	KIRIL	传统蒙古文的西里尔文书写形式。例如，“AGTACI”对应的“KIRIL”字段中填置“аттач”。
7	英文	English	蒙古文词条的英文对应词。例如，“AGTACI”对应的“English”字段中填置“horse-herder”。
8	释义	TAYILBVRI	每个词条所包含的意义。例如，“AGTACI”对应的“TAYILBVRI”字段中填置“CERIG-UN AGTA-YI HARIGVLDAG HOMON.”。
9	义项	VDHALASV	多义词语的相关义项数字，使用“数字值”（1、2、3…等阿拉伯数字）表示其义项。例如，“AMVSV”（祭食）所对应的“VDHALASV”字段中填置“1”，表示第一义项。
10	语义分类	VT	当前词条所属的语义类的标记。例如，“AGTACI”所对应的“VT”字段中填置“Nbba1125”，表示身份。
11	同形词	DURSUIJL	若当前词条是同形词，填置“A、B、C、…”等大写字母表示同形词。例如，“AMA”(嘴、摔跤的另一方)对应的“DURSUIJL”字段中填置“A”。
12	近义	0YIRALCAG_A	填置当前词条的近义词的拉丁转写形式，例如，“AGTACI”（马夫）的相应“0YIRALCAG_A”字段中填置“AGTACIN”，均表示“马夫”。

13	反义	ESERGU	填置当前词条的反义词拉丁转写形式，例如，“DOGOM”（简单）的相应“ESERGU”字段中填置“BVDVLIYANTAI”,表示“复杂”。
14	搭配	VYALDVG_A	填置当前词条搭配实例的拉丁转写形式，例如，“TOBCI”（简单）的相应“VYALDVG_A”字段中填置“TOBCI AGVLG_A\TOBCI ABSARHAN”，表示“内容简介\简单明了”。
15	价量	JH	当前词条的配价数量，用“0,1,2,3,4”等数字值表示其价量。例如，“NABCI”(叶子)的相应“JH”字段中填置“1”，表示一价名词。
16	价质	JC	当前词条的配价性质的相关标记。例如，“NABCI”(叶子)的相应“JC”字段中填置“buh”，表示“整体”（表示整体-部分关系）。
17	实例	JI$IY_E	每个词条在语料库和纸质辞书中的实例。例如，在“AGVR”（生气）的“实例”字段中填置“AGVR BEY_E-YI JVBAGAN_A,AGVLA MORI-YI JVBAGAN_A”。

表 5 语义词典 3 个分库属性字段简介

库名	字段数	字段名称及其标记
名词库	26	序号（NO）；蒙古文(MONGGOL)；音标（GALIG）；词类（UGSAYIMAG）；汉译（HITAD）；西里尔文(KIRIL)；英文(English)；语义关系(VDHARICAG_A)；释义(TAYILBVRI)；义项（VDHALASV）；同形词（DURSIJ）；近义词（OYIRCG）；熟语（HELELCE）；大语义类（TOVT）；子语义类（JIVT）；前搭配语义类(EVVT)；前搭配形式(EVH)；前搭配实例（EVJ）；后搭配语义类（HVVT）；后搭配形式（HVH）；后搭配实例（HVJ）；并列搭配语义类（JVVT）；并列配形式（JVH）；并列搭配实例（JVJ）；价量（JH）；价值(JC)；
形容词库	16	序号(NO)；蒙古文写法(MONGGOL)；音标(GALIG)；词类(UGSAYIAG)；汉译(HITAD)；释义(TAYILBVRI)；义项(VDH_A-DOTOM)；西里尔文(KIRIL)；近义词(OYIRALCAG_A)；反义词（ESERHU）；主体（AGENT）；客体（OBJECT）；语义类（Semantic class）；配价数(valence)；共现的名词例子(NER_E_VYAL)；共现的动词例子(UILE_VYAL)；
动词库	11	序号(NO)；蒙古文写法(MONGGOL)；音标(GALIG)；汉译(HITAD)；词类(UGSAYIAG)；读音（DAGVDALG_A）；异形词（ONDOO）；语义类（VT）；搭配的名词语义类(VVT)；价量（JH）；价质（JC）；

4 语义词典的初步应用

SIKM的语义属性在多义词消歧、同形异义词的辨别、短语结构关系判定以及语义角色的标注等层面均提供知识支撑。目前，SIKM初步被应用到了短语结构判断和语义角色研究、语义网络的开发以及语料库加工等科研工作中。

4.1 短语研究中的作用

1)《蒙古语名词短语语义角色的统计分析研究》[1]中应用语义词典SIKM的“词语”和“语义类”字段，通过对5107个蒙古语简单句进行语义角色标注，统计分析7646条名词短语充当语义角色情况，归纳出了813条名词短语的语义角色识别规则。其中构建“名词语义角色分析库”、统计分析名词短语语义角色中心词的语义分类等环节中充分利用了上述语义属性。例如，文中对于“存在”(0rs)这一语义角色已归纳出10条规则，其第一条规则为：

R1NP0rs-Ne2|Ne1+Ve1+Ne2

< Nsubcat > =Ne

< Morph > =0

< Nsem > =Nvt|Nbba1112

< Vsubcat > =Ve2

[1] 伊好斯巴雅尔. 蒙古语从比格形式名词的语义角色辨析研究，内蒙古大学硕士学位论文，2012 年；

< Vsem > =VHA0R

< Valent > =0rs

…

2）《蒙古语从比格形式名词的语义角色辨析研究》[2] 基于语料分析结果探索名词和谓词语义分类的相互匹配和约束条件时应用语义词典 SIKM 的“语义类”字段，归纳出从比格形式名词的294个语义角色辨析模型。例如，“MINV JIRUHE TVNG HUCU/TEI TUGSI/JU MANGNAI-ACA HOLOSO CIHIGLEBE.”例句的语义角色辨析模型为ERHETEN+ACA+UILEDHU UILE=EGUSBURI，表示“人的器官+ACA+主动动词=来源”，即“Nbba12+Fc40+VHOUI=egs”。

4.2 语义网络开发中的价值

名词语义网络是蒙古文WordNet的一项基础工程。其中名词同义词集合的建立是其核心内容。该研究基于语义词典“名词库”的名词概念和语义分类属性确定了各自的synset ID，参照WordNet概念之间关系以半自动方式创建一个适用于蒙古文信息处理的蒙古文名词同义词集合（图7）。尤其确定其概念之间关系或synset ID的词汇以及以synset为组织单位开始标注语义关系时充分利用了名词语义分类。但是，蒙古文wordnet的构建是长期的、动态的、工程量极大的项目。目前名词同义词集合信息在完善当中。

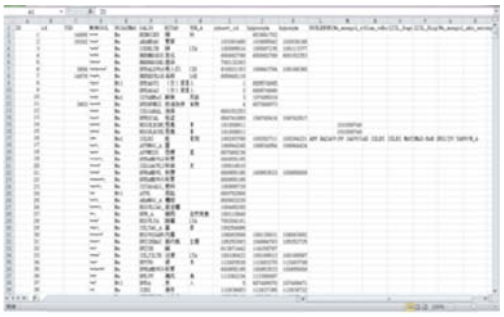

图7 “蒙古语名词同义词表”样本

4.3 语料库标注中的意义

语义词典 SIKM 对于语料库的词法标注和语义标注提供基础资源，为构建多级精加工语料库奠定基础。目前，课题组已完成语料库中的词法标注工作，构建了“100 万词现代蒙古语词法标注语料库”（如图 8）。

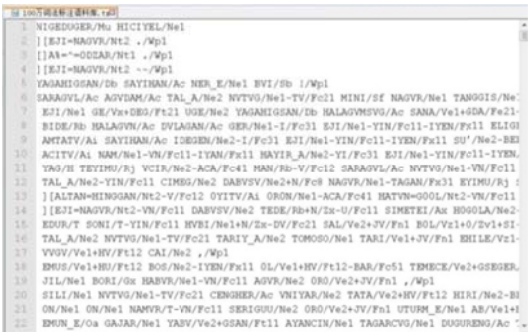

图 8 100 万词现代蒙古语词法标注语料库

下一步，我们将这份语料加工成词法-句法-语义信息标注的多级精加工语料库。其中，SIKM 作为一项基础资源投入应用。充分利用其“词语”、“语义类”、“语义关系”、“价量”、“价值”等诸多属性信息为多级精加工语料库的开发提供语义知识。例如，在语料库中标注名词语义分类信息如下所列（带下划线的表示语义分类信息标注）：

[]GANGGAM_A/Nbba1123-YIN EJI/Nbba1127 NI HEUHEN/Nbba1127 -IYEN HARI/Noh-DV M0RDA/JV , H0LA YABV/HV-ACA LA AYV/GAD SAyI MINGGA/N LANG/Nqj4-IYAR HOMON/Nbba1-U AMA/Nbba1111-YI TAGLA/JV BAYI/G_A HEREG/Nhy·…》（9/pageE190）[3]。

除此之外，在汉蒙机器翻译系统中语义词典判断多义词时，可以通过“义项”、“同形词”、“语义类”等字段中的选一可以说明当前的词条是否一个多义词。当同一个名词的多个义项属于不同语义类时，它们在句子中所受到的搭配限制也有所不同。其中可以利用“语义类”、“释义”、“价量”和“价质”等属性在目标语生成过程中对当前多义词进行消歧，从多义词的不同译法中挑选最合适的一个译词来提高译文质量。

5 结论

总体而言，语义词典 SIKM 的研发不但对于机器翻译、语料库加工、树库构建以及

[2] 陈红霞. 蒙古语从比格形式名词的语义角色辨析研究，内蒙古大学硕士学位论文，2012 年；

[3] 括号（）中的编号表示当前实例在“100 万词现代蒙古语语料库”中的位置。

文献检索和文本校对等NLP系统提供知识资源，而且对于词义定量分析和词汇语义学研究都有很大的价值。目前课题组已实际完成了5.7万个词语的语义分类与属性描述。不久的将来，我们对于蒙古语代词、时位词等其他词类进行语义分类，并将其归入到语义词典 SIKM 中，扩充其规模，为自然语言处理提供更广范围的、更多语义知识。但是，语义知识库的开发毕竟是一项长期性语言工程。我们应该根据实际应用结果和反馈意见，不断地完善和调整词典库，并将其投入到更大面积的应用领域，实现其实用价值。

参考文献

[1] Nasun-urt. Mongolian knowledge base and Mongolian information processing, *the 17th International Conference KOREA and MONGOLIA,* Seoul.2004, p51-58;

[2] Nasun-urt. Exploitation and application of the Mongolian linguistic knowledge resource,*Proceeding of the International Conference of Chinese Computing(ICCC2005)* COLIPS Publication, Singapore. 2005, p213-218;

[3] HaiYinhua,Nasun-urt,WangSirguleng. New progress of Mongolian Grammatical Information Dictionary [C], *Proceedings of Mongolian Academy of Sciences*4. Ulaanbaatar , 2008.p75-84;

[4] Mongolian Academy of Science, Mongolian Dictionary with Detailed Explanation, 1st ed. Ulaanbaatar, 2008.

[5] Nasun-urt, “The Presumption of Mongolian Language Resource Platform Framework,” in *Collection of Chinese Language Resources Essays*, Zhangpu,WangTiekunEd. the Commercial Press,2009,P236-248.

[6] HaiYinhua,Nasun-urt, and Wuyungaowa, “The Initial Framework of Developing Semantic Knowledge Base of Mongolian Idioms,” Altaic Hakpo, vol. 22, June2012. p121-139;

[7] 王慧，詹卫东，刘群.现代汉语语义词典的设计与概要，《1998中文信息处理国际会议论文集》，清华大学出版社，1998.pp361-367；

[8] 俞士汶，朱雪峰，王慧.《现代汉语语法信息词典》的新进展.《中文信息学报》，2001年第1期；

[9] 于江生，俞士汶.CCD的结构与设计思想，《中文信息学报》，2002年第4期；

[10] 俞士汶，朱雪峰，王慧，张华瑞等.《现代汉语语法信息词典详解（第2版）》，清华大学出版社.2002；

[11] 何英玉编，迈向21世纪的语言学《语义学》，上海外语教育出版社，2004年；

[12] 德.青格乐图等..现代蒙古语固定短语语法信息词典详解〔M〕.呼和浩特：内蒙古教育出版社，2005；

[13] 那顺乌日图.蒙古语语言资源平台架构设想.中国语言资源论丛(一). 张普，王铁琨主编，商务印书馆，2009年，p236；

[14] 呼日乐吐什.蒙古语语言知识库管理平台的设计与实现.呼和浩特市内蒙古大学硕士学位论文[D].2010.5。

[15] 陈章太. 论语言资源.中国语言资源论丛(一)..张普、王铁琨主编.北京：商务印书馆，2009，p13；

[16] 包志红.蒙古语语义信息词典形容词分库的构建.内蒙古大学硕士学位论文[D].2010，p23-24；

[17] 那顺乌日图.蒙古语语言知识库的建立与应用.中文信息学报[J].2011年第6期，p163-165；

[18] 海银花.面向信息处理的蒙古语名词语义分类体系.内蒙古大学学报(哲社蒙文版)·2012年第4期 ·p79-88；

[19] 伊好斯巴雅尔. 蒙古语从比格形式名词的语义角色辨析研究，内蒙古大学硕士学位论文，2012年；

[20] 陈红霞. 蒙古语从比格形式名词的语义角色辨析研究，内蒙古大学硕士学位论文，2012年。