

蒙古语形容词短语语义角色特征的统计考察

乌云嘎

内蒙古大学/呼和浩特
2296192394@qq.com

包艳花

内蒙古大学/呼和浩特
svrgvgceceg@126.com

达胡白乙拉

内蒙古大学/呼和浩特
dabhvrbyar@163.com

摘要

蒙古语中形容词短语数量多且用途广，在句子中可以充当多种语义角色。统计考察蒙古语形容词短语语义角色的特征对语法、语义的研究和应用具有重要的理论意义和实用价值。本文基于《蒙古语形容词短语语义角色信息数据库》，从形容词短语语义角色类别及频率、长度、长度与类别、内部结构关系类型及其频率、内部结构关系与类别之间的关系、内部结构关系类型与长度之间的关系、与形容词短语语义角色相关联的句子成分类型及频率、与形容词短语语义角色类别关联的句子成分类别及频率等8个方面进行了统计分析。经研究发现，该数据库含有形容词短语语义角色640条，出现的类别有4个大类14个小类，出现最多的大类是饰体，出现最多的小类是饰体大类中的性状语义角色；形容词短语语义角色的长度分布在2-12个词之间，其中长度为2个词的出现频率最高；长度为2个词的形容词短语语义角色覆盖的语义角色类别和出现次数最多；形容词短语语义角色覆盖的内部结构关系类别有7种，其中定体关系出现频率最高，且所对应语义角色类别最多；形容词短语语义角色与4种句子成分相对应，其中与定语对应出现的最多；与形容词短语语义角色类别关联的句子成分类别及频率统计中，每个语义角色对应一种句子成分。

关键词： 蒙古语；形容词短语语义角色；特征；统计分析

A Statistical Study on the Semantic Role Characteristics of Mongolian Adjective Phrases

Abstract

Adjective phrases in Mongolian are numerous and widely used, and can play a variety of semantic roles in sentences. Statistically investigating the semantic roles of Mongolian adjective phrases has important theoretical significance and practical value for the study and application of grammar and semantics. Based on the Mongolian Adjective Phrase Semantic Role Information Database, this paper analyzes the semantic role category and frequency, length, length and category, internal structure relationship type and frequency, internal structure relationship and category, internal structure relationship type and length, sentence component type and frequency associated with the semantic role of adjective phrase. Eight aspects of sentence component categories and frequencies associated with the semantic role categories of adjective phrases are statistically analyzed. Through research, it is found that the database contains 640 semantic roles of adjective phrases, and there are 4 categories and 14 subcategories in the database. The largest category is adornment, and the largest subcategory is character semantic role in the adornment category; The length of the semantic role of the adjective phrase is between 2 and 12 words, of which 2 words have the highest frequency; The semantic roles of adjective phrases with a length of 2 words cover

本研究得到国家社科基金重点项目《基于标注语料库的蒙古语句法计量研究》(批准号19AYY018)、内蒙古自治区高等学校应用语言学创新团队项目资助。

the most semantic role categories and occurrence times; There are 7 types of internal structural relationships covered by the semantic roles of adjective phrases, of which the fixed relationship has the highest frequency, and the corresponding semantic role categories are the most; The semantic role of adjective phrase corresponds to four sentence components, of which the most corresponding to attribute; In the sentence component category and frequency statistics associated with the semantic role category of adjective phrase, each semantic role corresponds to one sentence component.

Keywords: Mongolian , Adjective phrase , Characteristic , Statistic analysis

1 引言

语义分析是自然语言处理中最关键的一环，语义角色标注是句子语义分析的重要内容。语义角色标注是在切词、词法标注的基础上对自然语言进行的浅层语义处理。

语义角色标注的研究最早开始于Dan Gildea和Dan Jurafsky，他们的实验所用语料是Berkeley大学开发的Framenet¹。从此语义角色标注研究引起了学界的关注，成为了研究热点。汉语语义角色研究方面最早开始的是SUN等人。他们采用人工标注的方法对含有动词的语料做了语义角色标注。此后语义角色标注方面，董静(2006)、于江德(2007)等使用概率论的方法进行了语义角色标注。汪红林(2009)、鞠九朋(2010)等基于依存关系进行了语义角色标注。王明轩等(2018)、买宇博(2020)、郝淼(2020)等利用神经网络对语义角色进行了分类标注。杨凤玲等(2018)结合短语句法进行语义角色标注，得出短语结构句法能有效提高语义角色识别率。除上述之外，汉语语义角色标注方法有融合多层次特征、深度神经网络、利用配价信息等。在语义角色识别方面，王鑫等(2011)分析依存树的特点，基于依存树的距离进行语义角色研究，识别率为98.5%。党帅兵(2015)基于词分布特征，利用神经网络模型，实现了多种特征相结合的神经网络测试和测试算法。以字和词为标注单位分别做了研究，以字为标注单位的结果为50.10%，以词为标注单位时达到72.89%。杨耀文(2016)基于山西大学汉语语义知识库(CFN)，利用神经网络进行汉语语义角色识别研究。以词为基本标注单位，使用神经网络模型，采用Dropout惩罚机制与词分布特征进行了对语义角色识别的显著作用的研究。王瑞波等(2017)基于汉语词、词类等特征分布，建立多特征相结合的神经网络结构进行汉语框架语义角色识别模型有关研究。曹雪飞(2020)基于BiLSTM模型建立汉语语义角色识别模型，重点研究了建模中的超参数调优和模型输入端特征的表示学习的改进方法。汉语语义角色研究取得了一定程度的成绩，促进了汉语语义信息处理研究。

蒙古文信息处理工作进入句处理阶段后，学者们从各自的研究目的出发，对蒙古语语义角色进行了分类研究。如宝·哈斯巴根(1995)将现代蒙古语词汇之间的语法—语义关系在动词句范围内分为施事格、当事格、与事格、共事格、时间处所格等10个分类。斯琴·德力格尔玛(1996)将常见的语义格与传统蒙古语语法格相结合提出了施事格、受事格、凭借格、工具格等9种语义格。德力格尔玛(2001)在前期研究的基础上提出了使役格，形成了由10个语义格组成的分类体系。从蒙古语语义格表现形式、语义格与语法格之间的关系等方面那顺乌日图(2002)和高莲花(2005)分别提到了蒙古语语义格。那顺乌日图(2002)提出了施事格、工具格、凭借格、使役格、处所格、客体格等9种语义格。高莲花(2005)提出了施事格、受事格、凭借格、工具格、处所格、时间格、使役格等14种语义格。德·萨日娜(2006)将语义角色分为主体、客体、邻体、饰体等4大类，施事、当事等28小类，并尝试对《SVYVL-I GAYIHABA》一文做了语义角色标注。特图克(2008)将语义角色分为主体、受体等6大类42小类。其木格(2008)对比研究了蒙汉语的常用语义格，总结出了共性和差异性，并将蒙古语常用语义格分为主题格、客体格、领体格、工具格、饰体格、环境格、根由格等8大类14小类。怖仁毕力格(2011)将蒙古语语义格分为施事、主题、处所、时间、来源、目标、方式等8大类12小类。富涛(2011)从标有词法标注的26万词现代蒙古语语料中选取了2701个动词，在分析2602个例句的基础上制定了蒙古语简单谓语句的语义角色分类体系，包括了主体、客体、邻体3大类，施事、当事等28小类。包晓荣等(2013)对5000句语料进行基于依存语法的语义角色人工标注，明确了依存关系和语义角

¹丁伟伟，常宝宝：《基于语义组块分析的汉语语义角色标注》，《中文信息学报》，2009年第23卷第25期，第54页

色之间的关系。将蒙古语语义角色分为主体、客体、邻体、饰体等4大类，施事、当事等24小类。额尔敦朝鲁等(2016)完成的内蒙古自治区民族事务委员会项目《蒙古语言文字数字资源建设与共享工程信息处理用蒙古文相关标准》的子项目《信息技术—传统蒙古文语义角色分类体系及其标记集》中将蒙古语语义角色分为主体、客体等4大类，施事、当事、主题等31小类。上述分类体系的共同点是每种分类都包含施事、受事、处所、时间等4种语义角色。相比较而言，有的分类体系将语义角色直接分为大类，而有的分为大类、小类。

在语义角色自动识别方面，伊好斯白音(2012)在德·萨日娜(2006)的语义角色分类体系及标记集的基础上增加了存在(Ors)语义角色。在此基础上分析了7646个名词短语充当语义角色的实例，归纳了813条识别规则。包晓荣(2013)从蒙古语依存树库中选取5087个简单动词谓语句，抽取施事和当事的特征，归纳出施事语义角色识别规则79条、当事语义角色识别规则62条，进行了自动标注研究。施事和当事语义角色自动标注的召回率达到了87.62%，准确率为93.0%。在蒙古语价自动识别方面，分别归纳了受事、客事、位事、同事、与事、表事、使事、构事、起事、止事、解事等价的自动识别规则。

蒙古语语义角色分类、自动识别、自动标注等研究对蒙古语语义处理的深入和发展具有重要的学术意义和实用价值。

从国内外有关语义角色研究现状看，单词语义角色的分类、识别及标注有关研究较多，短语语义角色有关研究相对少，尤其是形容词短语语义角色的特征挖掘与统计分析方面的研究少之甚少，有必要基于真实语料统计分析的基础上开展深入探讨和研究。

2 关于形容词短语语义角色

形容词短语是指以形容词为主体，具有跟形容词相同的语法功能的短语，也有人称之为形容词性短语²。蒙古语学界根据短语的固化程度、中心词词类、词数、内部结构关系等的不同对蒙古语短语进行分类研究。其中，按中心词词类特征将蒙古语短语分为名词短语、动词短语、形容词短语、代词短语、数词短语、语气词短语、量词短语、方位词短语、时位词短语、副词短语、后置词短语、情态词短语等12类³。

一般来说，短语在句子中可以充当语义角色。按照短语的类别，对语义角色可以进一步分类研究。分类的结果可以有名词短语语义角色、形容词短语语义角色等。本文主要研究形容词短语语义角色有关特征。将对句子进行短语结构分析的基础上，对蒙古语形容词短语的语义角色特征进行统计分析。

3 语料加工与特征统计

3.1 语料加工

为挖掘蒙古语句法语义知识，从《现代蒙古语100万词级语料库》中选取了5000个句子，句长为3—20个词，对所选句子做了词法、句法、语义层面的加工，依次标注了固定短语标记、词法标记、短语标记、短语内部关系标记、语义类标记、句子成分标记、语义角色标记，构建了蒙古语句子多层次统一加工语料库。并基于多层次统一加工语料库构建了《蒙古语句子成分信息数据库》⁴。本研究利用《蒙古语句子成分信息数据库》的Phrase(短语类型)字段筛选形容词短语，构建了《蒙古语形容词短语语义角色信息数据库》。该数据库设有形容词短语(AP)、词类(Pos)、词法变化形式(Morph)、语义分类(Sem)、语义角色(Case)、短语内部关系(DH)、短语词数(UT)、句子成分(OGT)、来源句(IO)等9个属性字段。

语料加工中使用的固定短语分类及其标记依据来源于德·青格乐图(2005)《现代蒙古语固定短语语法信息词典详解》中使用的固定短语分类体系及其标记。词法分类及其标记来源于《信息技术——信息处理用蒙古文词语标记(GB/T26235-2010)》中制定的词类、构形附加成分、标点符号等的分类及其标记。短语分类和短语内部结构关系标记来源于达胡白乙拉(2014)《蒙古语短语结构知识库相关研究》中提出的短语分类和短语内部结构关系分类及其标记。动词语义分类及其标记来源于额尔敦朝鲁(2008)《面向信息处理的蒙古语动词语义研究》中提出的动词语义分类及其标记。海银花(2010)名词语义分类及其标记来源于《面向信息处理的蒙古语名词语义研究》中提出的名词语义分类体系及其标记。句子成分分类及其标记来源于阿日古

²李谱英：《形容词短语的结构和功能》，广西师范大学学报(哲学社会科学版)，1989年第1期，第39页。

³达胡白乙拉：《蒙古语短语结构知识库相关研究》，辽宁民族出版社，2014年，第28页。

⁴萨其荣贵：《蒙古状语信息数据库的构建》，内蒙古大学硕士学位论文，2013年，第35页。

娜(2012)《蒙古语名词短语句子成分的分析研究》中使用的句子成分分类及其标记。语义角色分类及其标记来源于伊好斯白音(2012)《蒙古语名词短语语义角色的统计分析研究》中使用的语义角色分类及其标记。

本研究利用该数据库中“形容词短语(AP)”、“语义角色(Case)”、“短语内部结构关系(DH)”、“短语词数(UT)”、“句子成分标记(OGT)”等属性字段及其值对蒙古语形容词短语语义角色特征进行了统计分析研究。

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	ONCA VRTV	Ac+Ac	Morph	seg	UT	Case	DH	OGT	IO													
2	BVRVNGSHI BALAMHD-IYAR	Ac+Ac	Pc51	#+#	2	yan	d	adb	e4	{ { {NCA/Ac VRTV/Ac} AP2s/adb/yan BAI/Ve2/DAG/Pc21 (HAWOR) /prf: VP3b . /Wp1} S4												
3	BLDES SUNIN-IYAR	Ac+Ac	Pc51	#+#	2	yan	h	adb	e5	{ { {BVRVNGSHI/Ac BALAMHD/Ac-IYAR/Pc51} AP2h/adb/yan {SESGUCE/Vel+GSEGER/Frc (VAJHASO) DAT/Ve2+EA/Ps14} VP2s/prd: VP4b . /#												
4	DAG_A UKEI	Ac+Su		#+#	2	yan	s	adb	e6	{ { {BLDES/SUNIN-IYAR/Pc51} AP2s/adb/yan VNT												
5	YEHLE_S TOMOGATAI	Ac+Ac		#+#	2	yan	d	adb	e6	{ { {DAG_A UKEI/VEI} AP2s/adb/yan CVLA/Ve2+L_A/Ps14} VP2s/prd: VP5u												
6	DOOHEH CIJ	Ac+Sq		#+#	2	yan	s	adb	e7	{ { {OLON HOOSON} AP2s/adb/yan {HARU/Vel+DAG/Pc21 (HOLLYA) UGE/Suj} VP2s/prd: VP4b												
7	YEHLE_S TOMOGATAI	Ac+Ac		#+#	2	yan	d	adb	e7	{ { {YEHLE_S TOMOGATAI} AP2s/adb/yan {YEHLE_S TOMOGATAI/Ac} AP2s/adb/yan {YEHLE/Vel+JU/Pnl} V4C												
8	DOOHEH CIJ	Ac+Sq		#+#	2	yan	s	adb	e7	{ { {DOOHEH CIJ} AP2s/adb/yan {DOOHEH/Ac} AP2s/adb/yan {DOOHEH/Vel+GSAN/Pc11 (VAJHASE) UC												
9	AMTATAIHAN BOGED NAMJIGVN	Ac+Cj+Ac		#+#	2	yan	h	adb	e8	{ { {AMTATAI/Ax+HAN/Pc11} AP2s/adb/yan {BOGED/CJ} AP2s NAMJIGVN/Ac} AP2h/adb/yan {VNTA/Ve2												

图 1: 《蒙古语形容词短语语义角色信息数据库》样本

3.2 特征统计

基于《蒙古语形容词短语语义角色信息数据库》对形容词短语语义角色的类别及频率、长度、长度与类别、内部结构关系类型及其频率、内部结构关系与类别之间的关系、内部结构关系类型与长度之间的关系、与形容词短语语义角色相关联的句子成分类型及频率、与形容词短语语义角色类别关联的句子成分类别及频率等方面进行了统计。形容词短语语义角色的类别及频率统计结果如下表所示:

表 1: 形容词短语语义角色的类别及频率统计

大类	主体			客体			
	主题	施事	当事	受事	涉事	共事	与事
频次	33	25	6	12	8	2	1
频率(%)	5.16	3.91	0.94	1.88	1.25	0.31	0.16
大类	邻体			饰体			
	处所	时间	方向	原因	性状	情态	数量
频次	13	8	2	1	382	143	4
频率(%)	2.03	1.25	0.31	0.16	59.69	22.34	0.63

《蒙古语形容词短语语义角色信息数据库》中含有640个形容词短语, 640个形容词短语充当语义角色, 即该数据库中包括640个形容词短语语义角色, 共出现714次。经统计发现形容词短语语义角色类别有4个大类14个小类; 根据4个大类出现频次从高低排列的顺序为饰体、主体、邻体、客体; 其中出现频次最高的是饰体语义角色, 共出现529次, 占82.66%; 出现频次最低的是客体语义角色, 共出现23次, 占3.59%。根据形容词短语语义角色14个小类语义角色出现频次从高低排列的顺序为性状、情态、主题、施事、处所、受事、时间、涉事、当事、数量、共事、方向、与事、原因; 其中性状语义角色出现频率最高, 共出现382次, 占59.69%; 与事、转事、原因等语义角色出现频率最低, 分别出现了1次, 分别占0.16%。

下面从主体、客体、邻体、饰体大类语义角色各举一例说明形容词短语语义角色, 如图2所示。

1. 包含主题形容词短语语义角色的例句:

S2472e9 {{{GANGGANG/Ac(Aag)/spc\cin SARAL/Ac(Aaj)}AP2d/sbj\sed
B0L/Vz2\AP3s/sbj\sed {UNEHHER/H\yan {H0B0R/Ac SAYIHAN/Ac}AP2h/0Ca\0ns
 ADVGV/Ne1(Nbba22622)/GPenn}NP3d}NP4b}NP7u GE/Vx+N_E/Fs21/IP}NP8s ./Wp1}S9
 (骏马是稀有的良驹) 该句中“骏马是”充当着主题形容词短语语义角色。

图 2: 主题形容词短语语义角色示例

2. 包含受事形容词短语语义角色的例句:

S4412e9 {{{ENE/Rj/spc VDAG_A/Qn}QP2/adb\cag {{{YAGV/Ra
 B0L/Vz2+0/Zv1+GSAN/Ft11}RP2u CV/Sq}RP3s/adb\cin UNEN/Ac CAGAN/Ac-
I/Fc31}AP2h/obj1\hur
 ILGARA/Ve2+GVL/Fe11+V/Zv1+N_A/Fs21(Vsb6)/prd}VP3t}VP6b}VP8b./Wp1}S9 (这次怎么
 也得证明清白) 该句中的“清白”是受有意发起行为影响的承受者, 因此在句中充当受事形
 容词短语语义角色。

图 3: 受事形容词短语语义角色示例

3. 包含处所形容词短语语义角色的例句:

S2522e9 {{{HABI/On TOSORTU/Ai}AP2d/adb\0r0 {{{[HILAR=DAMBI/Nt1-
 ACA/Fc41(Nbba1123)\ort OGER_E/Ri/GPenr}RP2d/spc\cin JIRVGCIN/Ne1(Nbba1126)}NP3d
 GE/Vx+JU/Fn1}NP4s/sbj {BAYI/Ve2+HV/Ft12(Vr0) UGEI/Su}VP2s/prd}VP6u}VP8b . /Wp1}S9
 (周围除了丹毕没有别的画家) 该句中的“周围”是表示动作行为发生的环境, 因此在句中
 充当处所形容词短语语义角色。

图 4: 处所形容词短语语义角色示例

4. 包含性状形容词短语语义角色的例句:

S4373e9 {[SARANGGVW_A/Nt1-DV/Fc21(Nbba1123)/obj1\hur {TERE/Rj/sbj\uid
 {{{HEDUI/Ra ONDOR/Ac}AP2b/spc\cin UNELELTE/Ne2(NuA)}NP3d/obj1\hur
 {OG/Ve1+CU/Fn1(Vas2) BAYI/Vz2+HV/Ft12}VP2t/prd}VP5t
 BVI/Sb}VP6s}VP7u}VP8t ?/Wp1}S9 (她给萨仁高娃多高的评价呢?) 该句中的“多高”这
 个形容词短语修饰名词“评价”充当性状形容词短语语义角色。

图 5: 性状形容词短语语义角色示例

形容词短语语义角色的长度分布如何?为回答这个问题以词数为单位统计了形容词短语语义角色长度,并进一步对其出现频次进行了统计。统计结果如下表所示:

表 2: 形容词短语语义角色长度和出现频次统计

词数	2	3	4	5	6	7	8	9	12
频次	460	94	44	26	6	7	1	1	1
频率%	71.88	14.69	6.88	4.06	0.94	1.09	0.16	0.16	0.16

从表2可以看到形容词短语语义角色的长度为2-12个词,根据出现频率从高低进行排列的顺序为2个词、3个词、4个词、5个词、7个词、6个词、8个词、9个词、12个词;其中长度为2个词的形容词短语语义角色出现频次最高,共出现460次,占71.88%;长度为8个词、9个词、12个词的形容词短语语义角色出现频次最低,分别出现1次,分别占0.16%。长度为7个词的形容词短语语义角色的出现频次比长度为6个词的形容词短语语义角色多,这个现象随着语料规模的扩大、语料体裁的增加发生变化是后期研究中需要关注的问题。

不同长度的形容词短语语义角色分布如何?为更好的回答这个问题,对不同长度的形容词短语语义角色进行了统计。结果如下表所示:

表 3: 形容词短语语义角色长度与类别关系统计

长度 类别	2个词	3个词	4个词	5个词	6个词	7个词	8个词	9个词	12个词
施事	11	4	4	3	1	2			
当事	4	2							
主题	11	9	3	4	3		1	1	1
受事	3	4	3		1	1			
涉事	5	3							
与事	1								
共事	1		1						
时间	6		2						
处所	6	3	2	1		1			
方向	2								
原因		1							
性状	305	41	20	13	1	2			
情态	103	25	9	5		1			
数量	2	2							

根据不同长度的形容词短语语义角色与14个小类语义角色类别相对应的频率高低进行排列的顺序为:2个词、3个词、4个词、5个词、7个词、6个词、8个词、9个词、12个词;其中长度为2个词的形容词短语语义角色分别与13个语义角色类别相对应,对应的类别数量最多;长度为8个词、9个词、12个词的形容词短语语义角色分别只与1个语义角色相对应。

形容词短语语义角色的内部结构关系有哪几种?为回答这个问题对形容词短语语义角色内部结构关系进行了统计,结果如表4所示:

经统计发现,形容词短语语义角色内部结构关系有宾述、体述、辅助、状述、联合、定体、复指等7种。根据出现频率从高低进行排列的顺序为定体、联合、辅助、宾述、体述、状述、复指;其中出现频率最高的是定体关系,共出现270次,占42.19%;出现频率最低的是复指关系,共出现3次,占0.47%。

那么,形容词短语语义角色内部结构关系相对于其类别的出现情况如何?为回答这个问题,又进一步考察了这些结构关系与语义角色类别之间的对应出现情况。统计考察结果如表5所示:

根据形容词短语语义角色内部结构关系与14个小类语义角色类别相对应的频次高低进行排

表 4: 形容词短语语义角色内部结构关系的类型及频次统计

形容词短语语义角色内部结构关系	定体	联合	辅助	宾述	体述	状述	复指
频次	270	203	75	32	29	28	3
频率(%)	42.19	31.72	11.72	5.00	4.53	4.38	0.47

表 5: 形容词短语语义角色内部结构关系与类别之间的关系统计

语义角色 \ 结构关系	定体	联合	辅助	宾述	体述	状述	复指
性状	145	155	28	23	20	11	
情态	59	38	17	7	8	14	
主题	10	4	19				
施事	18	1	4				2
处所	9	1		1		2	
受事	9	2	1				
时间	5		2	1			
涉事	6	2					
数量	2				1	1	
当事	4		2				
方向	1		1				
共事	1						1
与事	1						
原因			1				

列的顺序为：定体、辅助、联合、宾述、状述、体述、复指；其中定体关系对应语义角色类别的数量最多，与13种语义角色类别相对应；复指关系对应语义角色类别的数量最少，与2种语义角色类别相对应。

形容词短语语义角色内部结构关系类型与长度之间的关系如何？为回答这个问题，对形容词短语语义角色内部结构关系类型与长度之间的关系进行了统计分析。结果如下表所示：

根据形容词短语语义角色内部结构关系与不同长度的形容词短语语义角色类别之间对应出现的频次高低进行排列的顺序为定体关系、辅助关系、联合关系、宾述关系、状述关系、体述关系、复指；其中定体关系与长度为2-8个词的形容词短语语义角色对应；辅助关系与长度为2-6个词和9-12个词的形容词短语语义角色对应；联合关系与长度为2-7个词的形容词短语语义角色对应；宾述和状语关系与长度为2-5个词的形容词短语语义角色相对应；体述关系与长度为2-4个词的形容词短语语义角色相对应；复指关系与长度为2个词的形容词短语语义角色对应。

语义角色和句子成分是不同的概念。语义角色与句法角色相互独立又相互关联⁵。那么与形容词短语语义角色相关联的句子成分类型都有哪几类？经统计发现有主语、定语、状语和宾语。具体统计结果如表7所示：

根据与形容词短语语义角色类别关联的句子成分类别出现次数的高低进行的排列顺序为定语、状语、主语、宾语；其中定语出现次数最高，共出现385次，占总数的60.16%；出现次数最低的是宾语，共出现26次，占总数的4.06%。在此数据库中出现的宾语均是直接宾语，间接宾语未出现。

从形容词短语语义角色分类角度看，与形容词短语语义角色每个类别相关联的句子成分类别又有哪几类呢？经统计发现的结果如表8所示：

根据与形容词短语语义角色类别相关联的句子成分类别出现次数的高低进行的排列顺序为

⁵张蓉：《语义角色与句法角色》，沈阳师范大学学报(社会科学版)，2006年第30卷第4期，第80页

表 6: 形容词短语语义角色内部结构关系类型与长度之间的关系统计

长度	内部结构关系						
	定语	联合	辅助	宾述	体述	状述	辅助
2个词	199	172	38	9	26	13	3
3个词	34	12	24	11	1	12	
4个词	24	6	4	7	2	1	
5个词	5	8	6	5		2	
6个词	3	2	1				
7个词	4	3					
8个词	1						
9个词			1				
12个词			1				

表 7: 与形容词短语语义角色相关联的句子成分类型及出现次数统计

句子成分	定语	状语	主语	宾语
频次	385	165	64	26
频率(%)	60.16	25.78	10.00	4.06

宾语、状语、主语、定语；其中宾语与6种形容词短语语义角色对应；状语与4种形容词短语语义角色对应；主语与3种形容词短语语义角色对应；定语与2种形容词短语语义角色对应；并且14种形容词短语语义角色只与一种句子成分对应。

4 结果与讨论

本文基于多层面统一加工语料库从形容词短语语义角色类别及频率、长度、长度与类别、内部结构关系类型及频率、内部结构关系与类别之间的关系、内部结构关系类型与长度之间的关系、与形容词短语语义角色相关联的句子成分类型及其频率、与形容词短语语义角色类别关联的句子成分类别及频率等8个方面统计分析形容词短语语义角色特征。

经统计发现，形容词短语在句中主要充当性状语义角色；长度分布在2-12个词，长度为2个词的形容词短语语义角色出现频次最高的同时覆盖语义角色类别最多；形容词短语语义角色内部结构关系有7种，其中定语关系出现的最多并且覆盖的语义角色类别也最多；形容词短语语义角色可与4种句子成分对应，其中与定语对应最多。

综上，形容词短语语义角色类别统计中未出现的是存在、客事、受益、乘事、触事、转事、成事、路径、范围、境况、过程、工具、目的、依据、频度等15个语义角色；形容词短语语义角色内部结构关系统计中未出现的是总括关系；与形容词短语语义角色对应的句子成分中间接宾语未出现；长度统计中未出现的是长度为10个词的形容词短语语义角色。本研究是基于《现代蒙古语100万词级语料库》中选取的5000个简单句得出上述结果。后期研究中扩大语料规模、增加语料体裁、改变标注分类体系等的基础上对蒙古语形容词短语语义角色特征进一步深入研究。不同的语料规模、语料体裁、分类标注体系等是否对上述统计结果产生影响，以及未出现的形容词短语语义角色类别和内部结构关系是否在后续研究中出现是值得进一步探讨的问题。随着语言量化研究的发展，计量语言学者们相继发现了齐普夫定律等具有普适性的计量语言学定律。那么基于短语结构语法的蒙古语形容词短语语义角色类别及出现频率、长度及出现频率、内部结构关系及出现频率等的关系能否与齐普夫定律或计量语言学其他定律符合也是值得探究的问题。

5 结语

本文在5000个简单句的基础上对形容词短语语义角色特征进行了统计分析。后期会在扩大语料规模、增加语料体裁、不同标注体系等的基础上进一步对蒙古语形容词短语语义角色特征

表 8: 与形容词短语语义角色类别关联的句子成分类别及出现次数统计

句子成分 语义角色	主语	定语	状语	宾语
施事	25			
当事	6			
主题	33			
受事				12
涉事				8
共事				2
与事				1
时间			8	
处所			13	
原因			1	
方向				2
数量		4		
性状		381		1
情态			143	

进行对比统计分析。语义角色标注已在机器翻译、问答系统、信息抽取、信息检索等领域得到广泛应用，统计分析形容词短语语义角色功能对以蒙古语为一方的机器翻译、信息检索等系统的开发提供有效的参考数据。蒙古语形容词短语语义角色特征的统计分析为今后归纳形容词短语语义角色的识别规则、自动识别、自动标注等工作奠定基础。

参考文献

- 丁伟伟, 常宝宝. 2009. 基于语义组块分析的汉语语义角色标注. 中文信息学报, 23(25): 53-61+74. 董静. 2007. 基于CRF模型的语义角色标注研究. 中国科学院软件研究所, 北京, 中国.
- 于江德, 樊孝忠, 庞文博, 余正涛. 2007. 基于条件随机场的语义角色标注, 东南大学学报, 23(3): 361-364.
- 汪红林, 王红玲, 周国栋. 2009. 基于依存关系的语义角色标注. 计算机工程, 35(15): 82-84.
- 鞠九朋, 王红玲, 周国栋. 2010. 依存关系语义角色标注研究. 计算机工程与应用46(14), 158-161+175.
- 王明轩, 刘群. 2018. 基于深度神经网络的语义角色标注. 中文信息学报, (2): 50-57.
- 买宇博. 2020. 基于深度神经网络的语义角色标注研究. 河南大学, 郑州.
- 郝淼. 2020. 基于深度注意神经网络的语义角色标注研究. 中北大学, 太原.
- 杨凤玲, 周俏丽, 蔡东风, 季铎. 2018. 结合短语结构句法的语义角色标注. 中文信息学报, 32(6): 1-11.
- 王鑫, 穗志方, 李芸. 2011. 基于依存树距离的语义角色识别方法. 中国计算语言学研究前沿进展(2009-2011).
- 党帅兵. 2015. 基于词分布表征的汉语框架语义角色识别研究. 太原: 山西大学.
- 扬耀文. 2016. 基于神经网络模型的汉语框架语义角色识别. 太原: 山西大学.
- 王瑞波, 李济洪, 李国臣, 扬耀文. 2017. 基于Dropout正则化的汉语框架语义角色识别. 中文信息学报, (1): 147-154.
- 曹学飞. 2020. 基于BiLSTM的汉语框架语义角色识别方法研究. 太原: 山西大学.
- 宝.哈斯巴根. 1995. 现代蒙古语动词句研究. 北京: 民族出版社.
- 斯琴, 德力格尔玛. 1996. 语义学. 呼和浩特: 内蒙古人民出版社.

- 德力格尔玛. 2001. 蒙古语语义学. 沈阳: 辽宁民族出版社.
- 那顺乌日图. 2002. 关于面向信息处理的蒙古语语义研究. 内蒙古大学学报(汉文版), 30(5): 43-48.
- 高莲花. 2005. 试论蒙古语语义格. 内蒙古大学学报, (4): 14-21.
- 德·萨日娜. 2006. 蒙古语格语法框架研究. 呼和浩特: 内蒙古大学.
- 特图克. 2008(1). 蒙古语句子语义学结构. 中国蒙古学, 36(1): 1-7.
- 其木格. 蒙汉语常用语义格对比研究, 北京: 中央民族大学, 2008年.
- 怖仁毕力格. 2011. 蒙古语语义格分类分析. 中国蒙古学(蒙古文), 39(6):159-163.
- 富涛. 2011. 面向信息处理的蒙古语简单谓句句模研究. 呼和浩特: 内蒙古大学.
- 包晓荣, 华沙宝, 达胡白乙拉. 2013. 基于依存语法的蒙古语语义角色分类及其标记研究. 中文信息学报, 27(4):70-73.
- 伊好斯白音. 2012. 蒙古语名词短语语义角色的统计分析研究. 呼和浩特: 内蒙古大学.
- 包晓荣. 2013. 基于语料库的蒙古语简单句施事和当事语义角色特征研究. 呼和浩特: 内蒙古大学.
- 德·青格乐图. 2005. 现代蒙古语固定短语语法信息词典详解. 呼和浩特: 内蒙古教育出版社.
- 额尔敦朝鲁. 2008. 面向信息处理的蒙古语动词语义研究. 呼和浩特: 内蒙古人民出版社.
- 郭嘉伟. 2010. 汉语语义角色研究述评. 牡丹江教育学院学报, (1): 21-22.
- 海银花. 2010. 面向信息处理的蒙古语名词语义研究. 呼和浩特: 内蒙古人民出版社.
- 路青, 崔新春, 胡艳波. 2012. 基于文献计量的国内语义角色标注研究现状分析. 情报杂志, 31(4): 43-47+53.
- 陈菜芳. 2012. 中文语义角色标注研究概述. 文教资料, (27): 139-140.
- 阿日古娜. 2012. 蒙古语名词短语句子成分的统计分析. 呼和浩特: 内蒙古大学.
- 白萨其日拉图. 2013. 蒙古语谓语句信息数据库的构建. 呼和浩特: 内蒙古大学.
- 达胡白乙拉. 2014. 蒙古语短语结构知识库相关研究. 沈阳: 辽宁民族出版社.
- 达胡白乙拉, 萨其日拉图, 萨其荣贵, 其其格, 王旭芳, 阿丽玛. 2014. 蒙古语文本语料多层次统一加工相关的一些问题. 内蒙古大学学报, (3):1-18.
- 达胡白乙拉, 萨其日拉图, 萨其荣贵, 其其格, 王旭芳, 阿丽玛. 2014. 蒙古语文本语料多层次统一加工相关的一些问题. 内蒙古大学学报, (4):15-31.
- 楠丁. 2014. 蒙古文语义角色研究概述. 内蒙古师范大学学报, 43(6):132-134+138.
- 乌优坛. 2022. 基于蒙古语学术语体语料库的句法计量研究. 呼和浩特: 内蒙古大学.
- 宋衡, 曹存根, 王亚, 王石. 2022. 一种改进的汉语语义角色分类体系与标注实践. 中文信息学报, 36(12):140-151.
- 冯国丽, 于秀金. 2023. 格标记和语义角色互动的类型学研究. 南开语言学刊, 37(1): 16-32.