

# 蒙古语文语言政策文本树库的构建及其句法特征统计

安娜尔, 达胡白乙拉

内蒙古大学蒙古学学院, 内蒙古 呼和浩特

收稿日期: 2022年12月5日; 录用日期: 2023年1月12日; 发布日期: 2023年1月19日

## 摘要

语言政策作为国家政策的重要组成部分,在维护国家统一,促进社会和谐稳定发展中发挥着重要的作用。收集蒙古语文语言政策文本,进行句法计量研究属于语言政策研究和句法研究领域的新研究,也可以为蒙古语依存树库建设提供更多的数据,具有应用和研究价值。本文在阐述蒙古语文语言政策文本树库的加工处理及校对完善步骤的基础上对所构建的含有18,607个句子的树库语料句长、句类和句型进行了统计分析。

## 关键词

蒙古语文语言政策, 树库, 句长, 句型, 句类

# The Construction of Mongolian Language Policy Text Treebank and the Statistics of Its Syntactic Features

Anar, Dabhrbayar

School of Mongolian Studies, Inner Mongolia University, Hohhot Inner Mongolia

Received: Dec. 5<sup>th</sup>, 2022; accepted: Jan. 12<sup>th</sup>, 2023; published: Jan. 19<sup>th</sup>, 2023

## Abstract

As an important part of national policy, language policy plays an important role in maintaining national unity and promoting social harmony and stability. The collection of Mongolian language policy texts and the study of syntax metrology are a new research in the field of language policy research and syntax research, and can also provide more data for the construction of Mongolian

dependency tree, which is of application and research value. On the basis of elaborating the processing and proofreading work in the construction of the Mongolian language policy text tree, this paper makes a statistical analysis of the sentence length, sentence category and sentence pattern of the tree corpus containing 18,607 sentences.

## Keywords

Mongolian Language Policy, Treebank, Sentence Length, Sentence Pattern, Sentence Category

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近二十年来, 语言政策方面的研究活跃, 研究热情高涨, 研究成果颇多, 语言政策研究进入了一个全新的阶段、成为社会语言学的重要课题之一。可是, 缺乏对它的性质、理论基础、术语等作出的整体研究, 且尚未见到从语法或句法层面研究的研究成果。

依存树库中蕴藏着大量的词法和句法等信息, 对分析和理解句法结构提供丰富的信息。从蒙古语文语言政策树库抽取句法方面的数据, 并对其进行分析, 既可以检验前人的研究结论, 又可以发现内省法所观察不到的语言现象。目前构建的蒙古语树库多为随机抽取的多领域树库, 在语言分析上已经取得了较好的成果, 但是对特定的领域树库研究几乎空白。因此本文将阐述蒙古语文语言政策文本树库的加工处理及校对完善步骤并对语料句长、句类、句型进行统计分析。

## 2. 研究现状

### 2.1. 语言政策研究现状

21 世纪以来, 语言政策和语言规划研究逐渐在国内兴起, 相关成果陆续以专著或论文的形式发表[1]。不论是论文数量, 还是研究范围, 2005 年以来我国的语言政策研究发展速度堪称迅猛, 是本领域研究的主要发展时期[2]。国内语言政策研究的内容有关于我国语言政策的研究、关于其他国家的语言政策研究、中外语言政策对比研究、语言政策的理论研究及语言资源库的建设。语言资源库建设方面的研究体现了语言政策研究的新发展[2]。例如: 由上海外国语大学中国外语战略研究中心承担的教育部哲学社会科学重大课题攻关项目“世界语言政策综合资源库建设及比较研究”(15JZD047)于 2016 年获批。由上海外国语大学中国外语战略研究中心承担的国家语言文字工作委员会“十二五”科研规划重大项目“国际语言政策综合资源库建设”(ZDJ125-1)包括政策法规、规范标准、教材大纲、学术研究等四个子课题。张宝林的“‘世界各国语言政策数据库’建设与研究”[2]一文, 从研究目标、数据库内容、数据库建设、数据管理、基本思路、研究过程与方法等方面提出了为试图进行新的资源建设, 以满足学界的研究需求, 促进国内外语言政策研究的深入发展而建设“世界各国语言政策数据库”的设想。北京语言大学校级规划项目“世界各国语言政策数据库建设与研究”的核心成果之一“世界各国语言政策数据库系统 v1.1”是一个集成性的网络软件系统。国家社科基金重大项目“蒙古族语言生活调查”(17ZDA316)以社会、家庭、行政、司法、教育、传媒、文化艺术、宗教、商贸、政策等为研究领域; “蒙古语文语言政策数据库”是其阶段性成果之一。

## 2.2. 句法研究现状

学者们认为蒙古语句法研究同蒙古语语音学和词法学比较起来, 是一个研究还不很深入的学科之一[3]。近 20 年来针对蒙古语句法诸多问题的新研究频繁, 在句子理论、句子结构、句子成分、句子类型、句子模式分析等诸多方面取得了重要成就, 并产生了用转换生成语法、统计语言学等新式理论研究蒙古语句子一些问题的著作。在蒙古语依存树库方面, 斯·劳格劳的《现代蒙古语依存句法自动分析研究》[4]一文中设计实现了基于依存语法的蒙古语自动句法分析系统。刘凯、乌日力嘎、斯钦图等的《基于双语约束的蒙古语无监督依存分析》[5]、斯·劳格劳、华沙宝、萨如拉等的《基于统计方法的蒙古语依存句法分析模型》[6]、苏向东、高光来、闫学亮等的《蒙古文依存句法分析》[7]、谢聪娇的《基于汉蒙依存句法树转换的蒙古文依存句法分析》[8]等论文中设计实现了一些基于依存语法的句法分析模型; 楠丁的《基于依存语法的蒙古语宾述关系描述与识别研究》[9]、白庆格勒图、斯·劳格劳的《基于依存语法的蒙古语间接宾述关系识别》[10]等论文中对现代蒙古语宾述关系及状述关系与间接宾语关系进行了描述; 包晓荣、华沙宝、达胡白乙拉等的《基于依存语法的蒙古语语义角色分类及其标记研究》[11]、包晓荣的《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》[12]等论文中进行了基于依存树库的语义研究; 除此之外还有, 骆凯、李淼、乌达巴拉等的《汉蒙翻译模型中的依存语法与形态信息应用研究》[13]、斯·劳格劳、萨如拉的《蒙古语依存树库的构建》[14]、齐力格尔的《蒙古语书面语和口语某些问题的计量研究》等关于依存树库的研究[15]。

## 3. 蒙古语文语言政策数据库

### 3.1. 蒙古语文语言政策的界定

国内外学者一直关注着语言政策研究, 其中语言政策的定义是一个重要课题。澳大利亚语言学家 Kalpan 和 Baldauf 认为: 语言政策是一个社会, 群体或体系为了实现规划的变化而制定和实行的语言观念, 法律, 规定, 规则和实践[16]。我国学者蔡永良(2002)认为: 语言政策是指人类社会群体在言语交际过程中根据对某种或某些语言所采取的立场, 观点而制定的相关法律, 条例, 规定, 措施等等[17]。

舍·那木吉拉(2007)认为: 发展语言, 改变语言的因素有很多……其中最重要的是语言环境, 语言态度和语言政策, 但是起主要作用的是政治原因, 也就是政治机构的语言政策[18]。色·贺其业勒图(1994)认为: 语言政策是国家、政党、阶级、阶层或社团对语言问题所持根本态度的反映[19]。他还说, 语言政策是党和国家或有关社会组织在一定的历史时期和社会环境下制定的指向语言应用及语言发展问题的措施系统[20]。乐·陶格敦白乙拉(2003)认为: 语言政策是指人们有意识、有目的地影响语言文字功能的各种行政方法[21]。具体来说语言政策就是一个国家、一个民族或地方规定的语言文字的方针、原则和与其有关的工作任务。国家明文规定的关于语言的规定是显性语言政策; 体现语言态度, 观点的是隐性语言政策。总的来说, 语言政策是以规定、命令、指示、决定、纲领、法律或公文、领导演讲、论文、批评等形式出现的。

### 3.2. 语料的来源

在明确了语言政策的准确定位, 研究目的后, 最重要的就是收集大量的语料。收集蒙古语文语言政策语料的主要途径有:

1) 人工录入。这是本数据库最主要的来源。这部分包括蒙古文颁发的文件、论文以及演讲稿等。来源有: 由《蒙古语文》编辑部、巴彦淖尔盟蒙古语文工作委员会汇编的《民族语文政策文件汇编》(1980)、《实践》杂志、《民族语文》杂志、《内蒙古日报》等。

2) 网上下载。在上网收集之前必须对语言政策的概念有比较明确的界定, 根据界定去逐项选择, 而

不是盲目性地下载。然后依次搜索下载符合要求的语言政策文本。例如：内蒙古民族事务委员会网站(<http://mw.nmg.gov.cn/>)、呼和浩特市蒙古语文信息化综合平台(<http://www.mgl.gov.cn/>)、包头市民族事务委员会网站(<http://mw.baotou.gov.cn/>)等。

- 3) 使用现存的电子文本。
- 4) 使用 OCR 技术转换成文本。

根据语言政策的界定分为显性语言政策和隐性语言政策，在这基础上又依据公文分类，对文本进行了 12 大分类：指示、决定、通知、报告、条例、方法、规定、办法、总结、规划、演讲和论文；其中前十个为显性语言政策、后两个为隐性语言政策。

### 3.3. 文本的处理

因为是格式转换后无格式粘贴，所以存在字体不齐，段落不齐等问题。这些问题对后期查阅政策有着不可避免的影响，为解除这些不利因素我们对字体和格式进行了统一，再进行更深一步的处理。

1) 收集文本后用蒙古文自动校对系统(Mongolian Editor for Plain Text)进行校对。系统具有蒙古文词形非词检查、读音非词检查和读音非词自动纠错功能。

2) 按照传统蒙古文拉丁转写内蒙古大学方案，利用 Mongolian Editor，将蒙古文文本转写成内大拉丁字符文本。

### 3.4. 数据库管理平台

在数据库属性字段方面填入了政策标题，拉丁标题，发文字号，发文单位，发文日期，内容摘要，附件等信息。为实现蒙古语文语言政策的查询利用功能，为相关研究提供基础资料、节约研究时间、提高效率，在建立数据库之后，利用 Microsoft Office Access2010 研发蒙古语文语言政策数据库管理平台。蒙古语文语言政策数据库管理平台是用 Microsoft access2010 的宏和 VBA 完成的。

## 4. 蒙古语文语言政策文本树库的构建步骤

### 4.1. 蒙古语文语言政策文本树库的加工处理

在构建蒙古语文语言政策数据库的基础上选取文本构建蒙古语文语言政策文本树库时候需要一些标注体系。

#### 4.1.1. 固定短语识别

由于蒙古文单词间有空格隔开，因此不存在单词的切分问题，主要还是固定短语识别和标注问题。蒙古语固定短语是基本上由两个或两个以上的词组合而成的在形式和语义上结合紧密，表达一个词汇概念，构成一个词汇单位，充当一个句子成分或某一种附带成分的词组合[22]。根据“信息技术 - 信息处理用蒙古文词语标记集”(GB/T26235-2010)，固定短语可分为复合词、习用语、成语、固定词、名词术语等大类。词之间用“=”来连接表示。本研究用 MDPS (Mongolian Dependency Parsing System)自动识别复合词后根据大类定义和语言政策的特点，灵活地规定了几项标注规范。

#### 4.1.2. 词类标记代码

蒙古语文语言政策树库的词性标注采用自动标注的方式，辅以人工校对。蒙古语词汇可分为实词、虚词和感叹词。本研究依据“信息技术 - 信息处理用蒙古文词语标记集”(GB/T26235-2010)来划分，词类标记代码含有丰富的标记集，词性标记总数有 91 个。为了后续研究中的统计需求，我们选取了第一级中的 15 个代码。

### 4.1.3. 依存句法标注体系

构建依存树库需要遵循标注体系, 标注体系的差异首先体现在标注标签的设置上。各种依存关系采用的依存标注在数量和标注内涵上也各有不同。本研究所用的依存树库都遵循蒙古语依存关系标记集。该标记集包含 17 种依存关系。

## 4.2. 蒙古语文语言政策文本树库的校对与完善

树库的规模和质量与句法分析精度呈正相关关系, 对一定规模树库进行两遍校对所得句法分析器性能优于对两倍规模树库进行单遍校对, 因此在质量和规模间进行平衡时应优先考虑质量[23]。分析依存树库错误, 对提升树库质量有重要作用。

### 4.2.1. 固定短语标注错误

由于固定短语作为一个词汇单位、句子中充当一个句子成分或辅助成分, 所以固定短语标注是有关于句长统计和依存距离计算的工作。本文用 MDPS, 复合词自动识别后, 参考德·青格乐图的《现代蒙古语固定短语语法信息词典详解》进行一校。该词典中共选录了 7417 个固定短语, 其中复合词、习用词的结构成分限定为两个词, 名词术语限定为三个词, 成语为三个或三个以上的词组成[22]。二校由一人完成。针对语言政策的特殊性, 校对之前制定使用了如下规范:

- 1) 将机关名称视为一个单位。如“OBOR MONGGOL-VN OBERTEGEN JASAHV ORON-V ARAD-VN JOBLEL”。
- 2) 当下的政治术语视为一个单位。如“HAR\_A BULUGTEN HARGIS HUCUN”。
- 3) 书名、题名、文件名为一个单位。如“《EB HAMTV NAM-VN TVNGHAG》”。
- 4) 地区名称视为一个单位。如“HORINGER SIYAN”。
- 5) 语言文字类视为一个单位。如“MONGGOL HELE BICIG”。

### 4.2.2. 词性标注错误

词性标注是给定句子中判定每个词的语法范畴, 确定其词性并加以标注的过程[24]。蒙古语现有 AYIMAG, DAGA, DARHAN、Mglex 等切分、词性标注系统及基于深度学习的蒙古文词性标注方法。Mglex 是 2010 年内蒙古大学和中国社会科学院联合开发的词法标注系统, 它可以分词、词性标注, 能以“词干/词类标记 + 词缀/词类标记”为格式标注蒙古语词干词类信息和构形附加成分的信息。对 20 万词做处理的话准确率在 90% 左右, 平均准确率是 97.8%。分析器还具有以下特点: 用“/”表示词语变形形式、人名前用“[]”、地名前用“[]”。本研究是基于蒙古语词法分析系统 Mglex 进行了词性标注工作。在机器自动词性标注以后, 要进行人工校对, 人工校对分两步进行。

### 4.2.3. 依存关系标注错误

在标注过程中利用到以树库计算为目的的蒙古语依存句法分析软件 MDPS (Mongolian Dependency Parsing System)。标注后进行人工校对, 形成了包含 361,512 个词, 18,607 个句子的蒙古语文语言政策依存句法树库。

## 5. 蒙古语文语言政策文本句法特征统计分析

### 5.1. 蒙古语文语言政策文本句长分析

句子的测量单位有两种: 直接单位(小句)和间接单位(短语或词) [25]。大多数研究的句长测量单位为词。19 世纪以来, 以词数计算句长成为了定量研究的重点对象之一。首次用统计方法对句长进行系统研

究的人是谢尔曼(Sherman)。其中,用于句末的标点符号表示句末停顿和句子的语气,包括句号、问号、感叹号。这些符号是判断句子的标准。总字数除以总句数得平均句长,蒙古语文语言政策文本的平均句长为:  $361,512/18,607 = 19.43$ 。

## 5.2. 蒙古语文语言政策文本句型分析

句型,分为单句研究和复句研究。划分的主要依据是参考内蒙古大学蒙古学学院蒙古语文研究所(2005)编写的《现代蒙古语》[26]中的定义:句子可以按结构中所包含的谓语属性的多少分为简单句和复合句。如果句子结构只有一个谓语性质,这种句子就叫单句;句子结构中包含两个或两个以上谓语属性的称为复句。因此复合句可以看作是由数个简单句句型句模组成的句子。蒙古语文语言政策从句子句型和句模上看是简单句和复合句混用的。

### 5.2.1. 蒙古语文语言政策文本中的简单句

蒙古语文语言政策虽然以复合句为主,简单句使用较为少见。经统计,共有1204个简单句。简单句,从结构看,这些单句,以主谓句为主,非主谓句相对少见。如:“HELE=BICIG-UN TAL\_A-BAR SIN\_E NER\_E=TOMIY\_A-YI NIGEDHEHU AJIL HIN\_E。”。

### 5.2.2. 蒙古语文语言政策文本中的复合句

大多数研究的句长测量单位为词。蒙古语文语言政策文本是由大量复合句构成的,有16764个复句。这些复合句中有复合从属句、并列复合句和多重复合句。

并列复合句,也叫联合复句,是由几个意义上没有明显主次关系的分句组成的复句。并列复句没有主句从句之分,前后分句分别描述几个行为、时间或事件的几个方面。如,“NEYIGEM JIRVM-DV 0R0JV, DELEHEI dAHIN NIGEDUN\_E。”。

复合从属句。两个单句之间以支配关系为句型来表示完整的意思。如,“OBOR MONGGOL-VN OBERTEGEN JASAHV 0R0N BAYIGVLVDAMAGCA MAN-V NAM MONGGOL HELE BICIG-I MONGGOL UNDUSUTEN-U ARAD TUMEN-U OBERTEGEN JASAHV BAYIGVLVLG\_A-YIN ERHE TVSIYAL-IYAN YABVGVLHV CIHVLA HEREGLEGUR B0LGAHV TVHAI 0LAN TAL\_A-BAR HUCUTEI ARG\_A HEMJIY\_E ABCV CINGGADHAGSAN YVM。”。

多重复句,包含三个或三个以上分句的复句。蒙古语文语言政策以复句为主,复句中多重复句也是常见的。其中有三重、四重、五重复句等。“ARBAN TABVN JIL-UN TVRSI OBERTEGEN JASAHV 0R0N-V NAM-VN H0RIY\_A B0LVN JASAG-VN GAJAR-ACA MONGGOL UNDUSUTEN-U OBER-UN-IYEN HELE BICIG-I HEREGLEHU BA HOGJIGULHU ERHE CILOGE-YI BATVLAGSAN-IYAR BARAHV UGEI, OBOR MONGGOL-VN OBERTEGEN JASAHV 0R0N-V UNDUSUTEN-NUGUD-UN ARAD TUMEN-I VDVRIDCV 0R0N-V OBERTEGEN JASAHV-YI YABVGVLVN, NEYIGEM-UN OGERECILELTE YABVGVLJV AJV AHVI S0Y0L-VN BUTUGEN BAYIGVLVLTA-YI HIGSEN-U HAMTV MONGGOL HELE BICIG-UN AJIL-I HUCUTEI B0LGAHV ARG\_A HEMJIY\_E-NUGUD-I ABCV, MONGGOL HELE BICIG-UN HEREGLEHU HEBCIY\_E-YI YEHEBER ORGEJIGULUGSEN BOGED VRIDV GARVG\_A UGEI HOGJIGULUN BAYALIGJIGVLJV TEGUN-I ORGEN 0LAN TUMEN-U AMIDVRAL, TEMECCEL-UN JINGHINI HEREGLEGUR BA MONGGOL UNDUSUTEN-U AJV AHVI S0Y0L-I HOGJIGULHU CIHVLA HEREGLEGUR B0LGAJAI。”

### 5.2.3. 蒙古语文语言政策文本中的引用内嵌句

在句子结构中插入自己或别人的言论,称为引用内嵌句,也就是插入句。用连接动词“GE-”、“HEME-”

与当句进行连接。该库中有 639 处。如, “NOHOR StALIN: 《HELE B0L HEREGLEGUR, BAGAJI MON, EGUN-IYER HOMUS H0G0R0ND0-BAN HARILCAG\_A ABCV, UJEL SANAG\_A-BAN S0LILCAJV, HARILCAN 0YILAGALCADAG YVM》 GEJU HELEGSSEN BAYIN\_A.”。

由此可见, 蒙古语文语言政策以复句为主, 其中有并列复句、复合从属句、多重复句等, 多重复句中也有三重复句、四重复句和五重复句等。同时蒙古语文语言政策文本中也有一些单句和插入句。

### 5.3. 蒙古语文语言政策文本句类分析

根据句子的语气倾向或所叙述的东西与现实关系的区别, 可分为陈述句, 疑问句, 祈使句, 感叹句。陈述句是用来叙述或说明情况、带有陈述语调的句子, 句末标点符号一般用句号; 疑问句是用来表示疑问、带有疑问语调的句子, 句末标点符号一般用问号; 祈使句是要求、命令、禁止或建议、请求对方做某事或不做某事的句子, 句末标点符号用句号或感叹号; 感叹句是用来表示强烈感情、带有感叹语气的句子, 句末标点符号用感叹号。本库中共有 18607 个句子。

其中陈述句有 18,003 个。例: “NER\_E=T0MIY\_A-YI SVDVLVN\_A.”。

疑问句有 188 个。疑问句的例子有: “ENE HESEG TEUHE-YI T0BCIHAN YARIGSAN NI BIDEN-DU YAMAR ACI TVSA TAI BVI ?”。

祈使句有 331 个。例: “DVMDADV VLVS-VN ARAD TUMEN-DU J0L JIRGAL 0LG0JV, DVMDADV-YIN UNDUSUTEN-U SERGUN MANDVLTA-YI BEYELEGULUY\_E!”。

感叹句有 85 个。例: “ENE CINI SAYI M0NGG0L GAJAR-VN MAyIG-TAI B0LVGSAN!”。

可见, 在句子语气上以陈述句子为主, 然后是祈使句较多, 感叹句最少。这体现了语言政策文本的政治性、客观性和中立性特征。

## 6. 结论

构建蒙古语文语言政策文本树库并对其进行句法特征研究, 对蒙古语文语言政策而言是在研究思路和研究方法上的拓展。经过统计研究, 主要取得了以下结论:

首先, 构建的蒙古语文语言政策数据库包括了与蒙古语言文字规范化、标准化、应用化、发展化、信息化有关的蒙古语文指示、决定、通知、报告、条例、方法、规定、办法、总结、规划、演讲和论文。

其次是对蒙古语文语言政策树库进行了固定短语识别、词性标注、依存句法标注等处理。树库的质量是影响句法分析精确度的重要因素。处理之前说明了一些标注体系: 固定短语识别要遵循“信息技术-信息处理用蒙古文词语标记集”(GB/T26235-2010)中的定义和《现代蒙古语固定短语语法信息词典详解》中的示例、以“=”连接表示; 词性标注以“信息技术-信息处理用蒙古文词语标记集”(GB/T26235-2010)标记为主; 依存句法标注体系以蒙古语依存关系标记集中的 17 个关系为主。分别使用了 Mgllex 和 MDPS 软件进行自动分析, 再辅以人工校对。

再次, 对其基础句法特征进行了统计研究。统计得出, 蒙古语文语言政策文本的平均句长为 19.43; 句型上主要以复合句为主, 复合句中有三重、四重、五重复句等; 句类上以陈述句为主。

这个研究还可以深入扩展, 这里只是做了初步的工作, 最终获得的成果将对语言政策颁布和研究、依存句法专属领域的研究和扩充规格有着重要的意义。本文的研究正是融合了语言政策研究和计量研究, 为语言政策和计量语言学的发展尽一点绵薄之力。

## 基金项目

国家社会科学基金重大项目《蒙古族语言生活调查》(17ZDA316)。

国家社会科学基金重点项目《基于标注语料库的蒙古语句法计量研究》(19AYY018)。

## 参考文献

- [1] 王福美. 我国语言政策与语言规划研究概述[J]. 民族翻译, 2014(4): 25-30.
- [2] 张宝林. 世界各国语言政策数据库建设与研究总体设计[J]. 语言规划学研究, 2015(4): 38-44.
- [3] 巴·达瓦达格巴. 蒙古语文研究丛书-蒙古句法研究[M]. 呼和浩特: 内蒙古人民出版社, 2008: 1.
- [4] 斯·劳格劳. 现代蒙古语依存句法自动分析研究[D]: [博士学位论文]. 呼和浩特: 内蒙古大学, 2011.
- [5] 刘凯, 乌日力嘎, 斯钦图, 等. 基于双语约束的蒙古语无监督依存分析[J]. 中文信息学报, 2011, 25(4): 24-29.
- [6] 斯·劳格劳, 华沙宝, 萨如拉. 基于统计方法的蒙古语依存句法分析模型[J]. 中文信息学报, 2012, 26(3): 27-32.
- [7] 苏向东, 高光来, 闫学亮. 蒙古文依存句法分析[J]. 计算机科学, 2014, 41(8): 97-100.
- [8] 谢聪娇. 基于汉蒙依存句法树转换的蒙古文依存句法分析[D]: [硕士学位论文]. 呼和浩特: 内蒙古大学, 2016.
- [9] 楠丁. 基于依存语法的蒙古语宾述关系描述与识别研究[D]: [硕士学位论文]. 呼和浩特: 内蒙古大学, 2017.
- [10] 白庆格勒图, 斯·劳格劳. 基于依存语法的蒙古语间接宾述关系识别[J]. 内蒙古民族大学学报(社会科学版), 2018, 44(3): 30-34.
- [11] 包晓荣, 华沙宝, 达胡白乙拉. 基于依存语法的蒙古语语义角色分类及其标记研究[J]. 中文信息学报, 2013, 27(4): 70-73.
- [12] 包晓荣. 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[D]: [博士学位论文]. 呼和浩特: 内蒙古大学, 2013.
- [13] 骆凯, 李淼, 乌达巴拉, 等. 汉蒙翻译模型中的依存语法与形态信息应用研究[J]. 中文信息学报, 2009, 23(6): 98-104.
- [14] 斯·劳格劳, 萨如拉. 蒙古语依存树库的构建[C]//澳门大学人文学院, 澳门语言学会, 中国中文信息协会. 第十五届汉语词汇语义学国际研讨会论文集. 澳门: 中国中文信息协会, 2014: 328-337.
- [15] 齐力格尔. 蒙古语书面语和口语某些问题的计量研究[D]: [硕士学位论文]. 呼和浩特: 内蒙古大学, 2017.
- [16] 罗伯特·卡普尔, 小理查德·巴尔道夫. 太平洋地区的语言规划和语言教育规划[M]. 北京: 外语教育与研究出版社, 2014.
- [17] 蔡永良. 论美国的语言政策[J]. 江苏社会科学, 2002(5): 194-202.
- [18] 舍·那木吉拉. 语言政策与语言发展变化——同时论述蒙古语的兴衰史[J]. 中国蒙古学, 2007, 35(2): 1-6.
- [19] 色·贺其业勒图. 论语言政策[J]. 内蒙古师范大学学报, 1994(2): 45-48.
- [20] 《蒙古学百科全书》编辑委员会《语言文字卷》委员会. 蒙古学百科全书(语言文字卷) [M]. 呼和浩特: 内蒙古人民出版社, 2004.
- [21] 乐·陶格敦巴雅尔. 蒙古文语言规划与语言管理[J]. 蒙古学研究, 2003(1): 60-69.
- [22] 德·青格乐图. 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.
- [23] 丘立坤, 史林林, 王厚峰. 多领域中文依存树库构建与影响统计句法分析因素之分析[J]. 中文信息学报, 2015, 29(5): 69-75.
- [24] 宗成庆. 统计自然语言处理[M]. 北京: 清华大学出版社, 2019.
- [25] Kohler (2012) Quantitative Syntax Analysis. De Gruyter, Berlin.
- [26] 内蒙古大学蒙古学学院蒙古文研究所. 现代蒙古语[M]. 呼和浩特: 内蒙古人民出版社, 2005.