

DOI: 10.11992/tis.201706092

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180408.1131.012.html>

基于词缀的维吾尔谚语识别关键技术研究

穆妮热·穆合塔尔^{1,2,3}, 李晓^{1,2}, 杨雅婷^{1,2}, 艾孜尔古丽⁴, 周喜^{1,2}

(1. 中国科学院新疆理化技术研究所, 新疆 乌鲁木齐 830011; 2. 新疆民族语音语言信息处理实验室, 新疆 乌鲁木齐 830011; 3. 中国科学院大学, 北京 100049; 4. 新疆师范大学计算机科学技术学院, 新疆 乌鲁木齐 830054)

摘要: 在自然语言理解、机器翻译、舆情分析等自然语言处理领域中, 维吾尔谚语识别是整个文本实体识别的重要组成部分。为满足维吾尔谚语信息化的需求, 本文构建了比较完善的维吾尔谚语语料库。同时, 从传统语言学角度对维吾尔谚语的语法、语义结构进行分析, 构建了一个由维吾尔谚语功能语类(词缀)组成的、专属维吾尔谚语规则的知识库, 并将此知识库与自然语言处理技术相结合, 实现一个既能够从文本中识别出维吾尔谚语, 又能提供维汉互译等功能的信息软件系统。该系统也为开展计算机理解与处理维吾尔文字奠定了一个崭新的基础。

关键词: 维吾尔谚语; 谚语词缀; 谚语规则; 词缀覆盖率; 谚语规则库; 谚语语料库; 识别系统

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 1673-4785(2018)03-0452-06

中文引用格式: 穆妮热·穆合塔尔, 李晓, 杨雅婷, 等. 基于词缀的维吾尔谚语识别关键技术研究[J]. 智能系统学报, 2018, 13(3): 452-457.

英文引用格式: Munire·Muhetaer, LI Xiao, YANG Yating, et al. Affix-based key technology for Uyghur proverb recognition[J]. CAAI transactions on intelligent systems, 2018, 13(3): 452-457.

Affix-based key technology for Uyghur proverb recognition

Munire·Muhetaer^{1,2,3}, LI Xiao^{1,2}, YANG Yating^{1,2}, AZRAGUL⁴, ZHOU Xi^{1,2}

(1. Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Science, Urumqi 830011, China; 2. Xinjiang Key Laboratory of Minority Speech and Language Information Processing, Urumqi 830011, China; 3. University of Chinese Academy of Science, Beijing 100049, China; 4. School of Computer Science and Technology, Xinjiang Normal University, Urumqi 830054, China)

Abstract: In fields of natural language processing such as natural language understanding, machine translation, and public opinion analysis, Uyghur proverb recognition is an important part of the whole text entity recognition. To meet the need of Uyghur proverb informationization, this paper establishes a relatively complete corpus of Uyghur proverbs. The grammar and semantic structure of Uyghur proverbs were analyzed from the perspective of traditional linguistics, and a knowledge base that comprises functional genres (affixes) of Uyghur proverbs and obeys Uyghur proverb rules was constructed. In addition, the knowledge base was combined with natural language processing technologies to realize an information software system that can recognize Uyghur proverbs from text and mutually translate between Chinese and Uyghur language. The system also laid a new foundation for understanding and processing Uyghur language and characters by computer.

Keywords: Uyghur proverbs; proverbs affix; proverb rules; coverage rate of affix; proverb rule bases; proverb corpus; recognition system

收稿日期: 2017-06-30. 网络出版日期: 2018-04-08.

基金项目: 新疆维吾尔自治区重点实验室开放课题 (2015KL031); 新疆维吾尔自治区重大科技专项课题 (2016A03007-3); 新疆维吾尔自治区自然科学基金项目 (2015211B034); 中科院战略性先导科技专项项目 (XDA06030400); 新疆维吾尔自治区社会科学基金项目 (2016CYY067).

通信作者: 李晓. E-mail: xiaoli@ms.xjb.ac.cn.

维吾尔语是在新疆维吾尔自治区范围内使用人数较多的语言, 维吾尔语信息化是我国少数民族语言文字信息化的重要组成部分之一, 一直以来得到了党和国家的高度重视^[1]. 维吾尔谚语在维吾尔语

中广泛应用,通常由语素、词、词组、句子等基本单位逐级组建成^[2]。维吾尔谚语作为维吾尔语言的一个重要的组成部分,与其他的维吾尔语言成分结合在一起,构成了一种完整的语言网络系统^[3]。根据维吾尔谚语自身的语法、语义、结构特征将谚语从这一网络系统中提取出来是从计算机角度对维吾尔谚语语言结构的探讨与研究,提供了一个新的视角,同时该研究在舆情分析、语言理解以及机器翻译中将会被广泛使用^[4]。除此之外,考察维吾尔谚语的使用率、覆盖率,统计新谚语,在文本分析研究具有重要意义,并且在传承民族优秀文化、饮食文化、道德观念、哲学思想等方面起一定的作用^[5]。

1 维吾尔谚语规则知识库的建设

维吾尔族谚语是流传在维吾尔族人民群众口头上的定型的现成语,具有意义的完整性、结构的简短性、结构的稳定性、语言的通俗性、哲理及教育性等特点。在语法和结构上,维吾尔谚语也有其自成一格的结构系统^[6]。

根据语言学界对谚语的研究结论和语言学理论^[7],从维吾尔谚语的语义、语法结构两方面入手,提出辨别维吾尔谚语的基本原则,并根据基本的原则建立成规则库为计算机自动划界并识别维吾尔谚语提供基础^[8]。到目前为止,维吾尔谚语规则库包括词缀、关键词、关联词等几个形式。通过对语料库中的维吾尔谚语进行分析与研究归纳出谚语识别规则。

1.1 维吾尔谚语规则知识库

本文通过分析小学1年级到6年级的维吾尔语文教材后,收集了在文本中紧跟着维吾尔谚语前面或者是后面出现,用来引用当前谚语的词组或句子,用其作为搜索的对象从而缩小了文本中维吾尔谚语的搜索范围。其部分规则如下(总规则词组为12种):

- (1) 此比喻中这样说的(…) بۇ تەمسىلدە مۇنداق دېيىلگەن
- (2) 有这样一说(…) دېتەكەن
- (3) 有这样的俗话(…) كوناڭار مۇنداق دېگەن
- (4) 说过的话(…) دېگەن تەمسىل مانا مۇشۇنىڭدىن كەلگەن. شۇنىڭدىن كەلگەن

(从此留下这种话语/比喻)。其基于标点符号的规则:“<”和“:”等这些符号作为谚语识别候选规则,为下一步判断谚语作为参考。

1.1.1 基于维吾尔语附加成分特征的规则研究

根据维吾尔谚语的语法结构,对谚语进行分类判断^[9]。例如:

1) 肯定形式:谚语中的动词词干后附加词尾<-ر, -ار, -ىر>构成的谚语。

2) 将来时第三人称单数词尾<-ىدۇ, -ىدۇ>构成的谚语。

3) 第二人称祈使句分为肯定和否定两种形式的谚语。

4) 过去式时态接陈述式第三人称,即在副词、动词等词类后附加词尾<-تى>构成的谚语^[10]。

5) 否定形式的谚语:在动词词干后附加词尾<-مەس, -مەس>构成。

6) 谓语是以形容词结尾的谚语。等6种语法规则。例如:

سېخىنىڭ كۆلكىسى، بېخىلىنى يىغلىتار (慷慨者的笑能使小气鬼哭)

«مىۋ» قولى سۇنۇقنىڭ دەردىنى قولى سۇنۇق بىلىدۇ (断臂人的痛苦只有断臂人知道)

1.1.2 维吾尔谚语关系规则

根据维吾尔谚语提出了并列、假设、取舍、连锁等4种关系规则。

根据上述的维吾尔谚语的句型结构归纳的规则(词缀)总数有75种,词缀的含义是黏附在词根上构成新词的语素,它本身不能单独构成词。在维吾尔语言中黏附在词根前面的词缀称为前词缀(前缀),黏附在词根后面的词缀称为后词缀(简称后缀),插入词根中间的词缀称为中词缀(简称中缀)。另外,维吾尔谚语中有一种特殊词缀,即由两个或者两个以上中缀黏附在词根后面,本文中均可以双中缀和多中缀,例如表1所示。

表1 维吾尔谚语词缀(部分)

Table 1 Part of Uyghur proverbs suffixes display

后缀	例子	汉译(例子)
后缀	مەر،دەر	黄金的价值工匠知
中缀	بېرىپ يامان بولغۇچە ، بىرمەي يامان بول	与其给了得罪人，不如不给当恶人
双中缀&多中缀	پۇل ئالدىدا ئېزىقما، مەنسەپ دېسە قىزىقما	金钱面前莫动心，说起当官莫眼

根据6490条维吾尔谚语自身的特点,归纳了75种词缀,其中有35个后词缀(后加成分),6个中缀,34个双中缀和多中缀。

1.2 维吾尔谚语歧义现象

歧义是指语言中对一个词语或一种结构有两种

或多种解释,是语言研究中梳理的一个关键点^[11]。

1) 谚语成为句子成分

维吾尔谚语进行自动识别时在不同的规则层次中出现不同的歧义现象。不符合规则的现象分为以下两类:新谚语和谚语成为句子成分。对于基于关键词规则方法来说,有的谚语可以直接引用,前后不加任何词语,而且作为固定组合时也可以充当句子的其他成分,比如:主语、宾语等。但这种可能非常有限^[12]。如:

① 强盗收拾贼是免不了的。

ئوغرىنى قاراچى سوقۇپ كېتىشى ئېتىمالدىن يىراق ئىمەس ئىدى.

② 你可知瞧着被子伸腿,一个巴掌拍不响的道理。

سەن يوتقانغا قاراپ پۇت سۇنۇشنى، چاۋاك ئىككى قولدىن چىقىشىنى بىلمەسەن.

虽然固定语句是维吾尔谚语,但是具备了一个非维吾尔谚语该有的句子结构和语法结构^[13],比如上述例子①②。然而,这类谚语,往往需要借助人工处理后才能判断是否是谚语。

2) 谚语成为其他固定语句

维吾尔谚语作为语言中的一种固定的定型语句,谚语、成语和格言构成并列关系,三者同属于非生语范畴并且构成上下位关系。谚语、成语和格言的语义结构与功能存在某些交错。因此,计算机对维吾尔谚语进行自动处理时须有专属维吾尔谚语的特定规则,后再使用计算机进行处理^[14]。

谚语与成语。在结构上,虽然都大于词,但谚语一般为一个完整的句子或者复句,而成语则多为词组^[15]。系统自动识别时,部分维吾尔谚语除了句尾用句号以外,句中由逗号隔开,这就表明维吾尔谚语的形式是以句号、逗号相配合组成的谚语,而成语则只是以句号结束。

谚语和格言。在结构上,维吾尔谚语一般由一个或两个句子组成,这是由维吾尔谚语的语义简洁性所决定的。而格言有的是由一个或两个句子构成,有的则是由几个句子构成的语段。

谚语与歌谣。维吾尔谚语语言结构形式类似于维吾尔语歌谣,但是在结构、语义和句法功能上同样存在差别^[16]。通常维吾尔谚语最长不超过两三句,而维吾尔语歌谣则以两句以上为多见^[17]。

2 维吾尔谚语识别系统研究与设计

本研究中开发的维吾尔谚语自动识别系统分为维吾尔谚语判断模块、维吾尔谚语查询模块及维吾尔谚语统计模块三大模块,如图1所示。

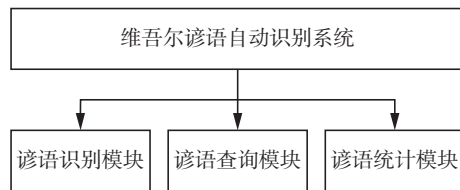


图1 维吾尔谚语识别系统功能图

Fig. 1 Functional diagram Uyghur proverb recognition system

1) 维吾尔谚语判断模块

维吾尔谚语判断模块从多种角度进行判别:

① 通过由关键词、标点符号、附加成分(词缀)组成的三层识别方法,从文本中识别出可能成为谚语的固定语句并判断该固定语句是否为维吾尔谚语;

② 对话框中输入固定语句,系统通过匹配当前固定语句的附加成分(词缀)后判断该输入语句是维吾尔谚语或者是成语或者是其他类固定语句。

2) 维吾尔谚语统计模块

此模块主要功能分为两部分:一是对已判别好的维吾尔谚语成分进行统计;二是按照规则知识库对语料库中的谚语逐条自动进行判断其是否为维吾尔谚语。

3) 维吾尔谚语查询模块

查询模块的主要功能是根据使用者的需求,输出维吾尔谚语其他相关信息。

2.1 系统功能的实现

本系统识别功能主要有以下几个:1) 识别关键词;2) 识别标点符号后自动地确定目标,固定语句的界定范围,以便后续进行判断;3) 识别功能语类(功能语类是指用来表达语法意义的成分,本文中的词缀);4) 将维吾尔谚语语料库中的所有谚语作为查询对象验证查询功能,即当输入维吾尔谚语上一句或下一句时,系统就将自动查询维吾尔谚语的上一句(下一句),并且输出维吾尔文解释、汉译、汉译解释、拼音和类型等一系列信息供用户使用;5) 统计功能,此功能包括维吾尔谚语成分统计和词缀覆盖率统计。前者根据谚语语料库里每一个完整的维吾尔谚语行数 and 词数进行统计。后者则根据已有的词缀对“纯谚语”语料库中的所有维吾尔谚语进行统计与计算,即检测维吾尔谚语关键技术研究系统的精准度。维吾尔谚语自动识别系统总流程图如图2所示。

本文中设计的维吾尔谚语自动识别系统在文本中充分利用3层条件识别出维吾尔谚语。识别系统判断率已超过97%。

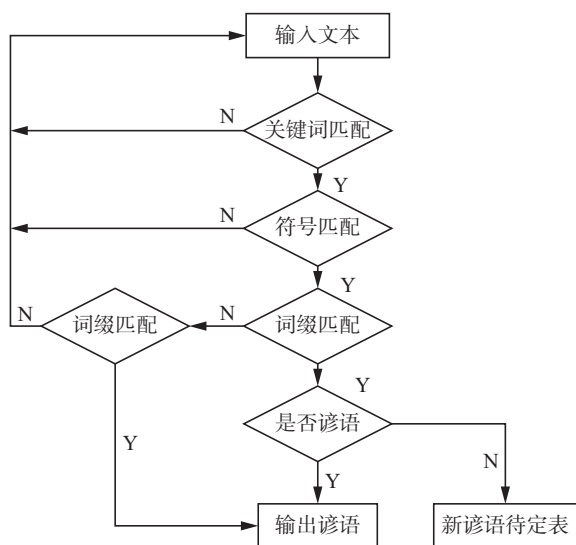


图2 维吾尔谚语识别系统流程图

Fig. 2 Uyghur proverb recognition system flow chart

2.2 维吾尔谚语语料库的构建

本研究所使用的文本为新疆师范大学“维吾尔语文研究基地”提供的以《语文》为名的小学维吾尔文语文教材、新疆人民出版社出版的《维汉对照维吾尔谚语》(总共3 000条)、《维吾尔谚语释解词典》(总共6 800条^[18])。

以《维吾尔谚语释解词典》和《维汉对照维吾尔谚语词典》中的6 490条维吾尔谚语(其中2 956条包含维汉解释、读法(拼音)、类型等一系列信息)组成的语料库为研究对象。用户根据需求进行查询与匹配、统计与分类数据库,从而能方便快捷查找到所需信息。

本文中收集的维吾尔谚语数据如表2所示。

表2 维吾尔谚语收集率

Table 2 Uyghur proverb collection rates

语料库	总数	收集总数	差别	收集率/%
纯语料	6 800	6 490	310	95.44
平行语料	3 000	2 956	44	98.53

根据表2可知,本文中维吾尔谚语的收集率达到95.44%,维汉平行谚语语料库的收集率达到98.53%。无论是维吾尔谚语纯语料(单语言),还是维汉谚语平行语料,其收集率都较完整。

2.3 维吾尔谚语识别对比实验

为了衡量上述所构建维吾尔谚语语料库和维汉平行谚语语料库的实用性以及谚语识别方法的有效性,本文中分别对维吾尔谚语语料库、维汉平行谚语语料库的开放以及关闭状态进行识别。实验结果表明,系统通过规则词组、标点符号、谚语词缀等3层识别步骤筛选后的维吾尔谚语识别率达到96.9%,如果将此结果再与《维吾尔谚语语料库》

比较后其识别准确率提高到99%。这说明,从文中搜索到的维吾尔谚语是在维吾尔谚语语料库里出现的,因此可以确切地判断当前的固定语句为维吾尔谚语。因为“维汉平行谚语语料库”中的维吾尔谚语语料总数少于“维吾尔谚语语料库”中的维吾尔谚语总数,因此,基于“维汉平行谚语语料库”(其他语料库关闭)从文本中识别谚语的准确率为43.47%。“维汉平行谚语语料库”关闭状态与前面所述的“维吾尔谚语语料库”关闭的状态是一个样的,就是说系统只能通过上述的3层识别方法识别维吾尔谚语,其数据分析如表3所示。

表3 谚语识别率对比性实验结果

Table 3 Proverb recognition rate comparative experimental results

语料库状态	识别率/%
维吾尔谚语语料库(开放)	99
维吾尔谚语语料库(关闭)	96.9
维汉平行谚语语料库(开放)	43.47
维汉平行谚语语料库(关闭)	96.9

通过对比实验结果表3可知,维吾尔谚语语料库开放状态下文本中识别谚语效率较高。当然用户也可以在没有谚语语料库状态下根据专属谚语的特定规则判别当前固定语句是否为谚语,但是其判别率并不很理想。因此建议通过规则库进行筛选之后再使用维吾尔谚语语料判断当前固定语句是否为维吾尔谚语。

2.4 添加新谚语

维吾尔谚语是广大维吾尔族人民口头流传的固定语句^[19]。到目前为止很多研究者已经收集了大量维吾尔谚语,但是随着维吾尔族文化的发展,可能会生成新的维吾尔谚语,并且在文本中依然能与关键词连用,或者出现在括号里。这些固定语句不仅满足关键词规则和标点符号规则,也符合谚语语法、词缀规则。另外,这些固定语句可直接存放在待定谚语库里供使用者人工处理,并进一步判定该固定语句是否为新的维吾尔谚语。表4是小学维吾尔文语文教材中的新谚语数据分析。

表4 新谚语比例

Table 4 Proportion of new proverbs

语料库	谚语总数	新谚语	新谚语比例/%
语文教材	170	28	16.47

从表4可知,小学维吾尔文教材中识别的维吾尔谚语的总数共170条,其中新谚语数目为28条,占总谚语数目的16.47%。

2.5 词缀覆盖率

计算词缀覆盖率时,对收集的6490条维吾尔谚语进行分析。其中包含后缀的维吾尔谚语共3786条,由中缀相连接的维吾尔谚语共1555条,由双中缀和多中缀相连接的维吾尔谚语共2356条。因为谚语本身句法结构多样,附加成分(词缀)连接特征极其不规律,包含上述词缀的维吾尔谚语共4934条,覆盖率百分比为75.81%。其余1556条维吾尔谚语是没有特殊的语法结构或者没有包含特定的词缀,未覆盖概率百分比为24.19%。其数据分析如表5所示。

表5 谚语词缀覆盖率
Table 5 Proverbs suffix coverage

词缀类型	个数	匹配谚 语个数	匹配覆 盖率/%	未匹配 谚语	未匹配 覆盖率/%
双中缀和多中缀	34	2356	36.30	4134	63.9
中缀	6	1555	23.96	4935	76.04
后缀	35	3786	58.33	2704	41.66

表5中分别显示每一种功能语类(词缀)在维吾尔谚语语料库中占的百分比,其中被双中缀和多中缀覆盖的维吾尔谚语占总谚语的36.30%,唯有中缀和后缀覆盖的维吾尔谚语分别占为23.96%和58.33%。值得一提的是,维吾尔谚语语料库里的6490条维吾尔谚语中的某些维吾尔谚语有同时与后缀和中缀以及双中缀相连接的可能性,即系统从文本中识别出维吾尔谚语时,有可能会在同一条维吾尔谚语本身匹配到上述的3种词缀。比如:

غورا بولماي مېغىز بولماس : 维吾尔文

汉译:(没有青杏哪来杏仁)^[11]。

此谚语中包含后缀和中缀,假设关系...مى...格式,还有否定形式词缀:在动词词干后附加词尾مىس,مىس。

本文中的覆盖率是指所有谚语词缀的频率由高到底降序排列时其每一条谚语词缀与其整个谚语的频率之和在全部语料中所占的比重,用来度量测试完整性和测试有效性^[20],如公式(1)所示。

$$F_i = \sum \frac{n_i}{N} \times 100\% \quad (1)$$

式中: F_i 为识别对象*i*的覆盖率, n_i 为识别对象*i*的出现次数, N 为所有语料中调查对象出现的总量。

3 结束语

本文从计算语言学理论角度归纳出维吾尔谚语的语法结构模式,为文本中自动识别维吾尔谚语系统的实现构建了较完整的维吾尔谚语语料库与只属于维吾尔谚语语法的维吾尔谚语规则知识库。

维吾尔谚语系统的分析、设计与实现详细地说明系统的设计路线和功能模块,并进行小规模实验,即测试系统。本文阐述了维吾尔谚语规则库、源数据库和设计系统的模型与设计中的细节,并且对维吾尔谚语语料库的各项语料统计进行分析,为自动识别系统提供了良好的基础。系统依据用户所提交的含有维吾尔谚语的任何一种格式的文本,通过关键词和相关符号,与语料库、规则库进行判断,获得所需的维吾尔谚语,根据需要进行查询与匹配,进行统计与分类,可快速地找到所需数据。本研究除了在语言理解、维汉/汉维机器翻译中起重要作用之外,还能对中考、高考中考维吾尔语文的全体学生提供一个良好的学习平台,对教学研究及推广维吾尔语言也会起到很大作用。

参考文献:

- [1] 杨艳,何天宇. 基于短语的关系数据库关键词查询方法[J]. 计算机研究与发展, 2012, 49(S1): 276-282.
YANG Yan, HE Tianyu. Phrase-based keyword search over relational databases[J]. Computer research and development, 2012, 49(S1): 276-282.
- [2] 张红雨. 汉维谚语对比[D]. 重庆: 西南大学, 2010: 15-20.
ZHANG Hongyu. The development of chemical curriculum comparision of Chinese Uighur proverbs[D]. Chongqing: Southwest University, 2010: 15-20.
- [3] 杨德明. 再论维吾尔谚语的翻译[J]. 民族翻译, 2014(4): 81-86.
YANG Deming. On the translation of Uyghur proverbs[J]. Minority translators journal, 2014(4): 81-86.
- [4] 马俊民. 论维语谚语的汉译[J]. 语言与翻译, 1985(3): 47-52.
MA Junmin. On the translation of Uyghur language into Chinese[J]. Language and translation, 1985(3): 47-52.
- [5] 芦凤娟. 基于语料库的英语谚语及其变体使用研究[D]. 临汾: 山西师范大学, 2013: 12-15.
LU Fengjuan. A corpus-based study on the Usage of English proverbs and their variations[D]. Linfen: Shanxi Normal University, 2013: 12-15.
- [6] 马俊民, 廖泽余. 维汉对照维吾尔谚语[M]. 乌鲁木齐: 新疆人民出版社, 2007: 1-300.
- [7] 陈雪慧. 汉维谚语中比喻修辞对比研究[D]. 乌鲁木齐: 新疆师范大学, 2013: 8-12.
CHEN Xuehui. Chinese proverbs and Uyghur proverbs comparative study of figures of speech[D]. Urumchi: Xinjiang Normal University, 2013: 8-12.
- [8] 力提甫·托乎提. 从短语结构到最简方案: 阿尔泰语言的句法结构[M]. 北京: 中央民族大学出版社, 2004: 80-140.
- [9] 刘正江. 维吾尔语成语研究概述[J]. 喀什师范学院学报,

- 2004, 25(2): 59-61.
- LIU Zhengjiang. The general research situation about idiom of Uyghur language[J]. Journal of Kashgar teachers college, 2004, 25(2): 59-61.
- [10] 力提甫·托乎提. 维吾尔语及其他阿尔泰语言的生成句法研究[M]. 北京: 民族出版社, 2001: 60-120.
- [11] 张勇. 维吾尔谚语研究[D]. 乌鲁木齐: 新疆大学, 2005: 10-97.
- ZHANG Yong. Study on Uyghur proverbs[D]. Urumchi: Xinjiang University, 2005: 10-97.
- [12] 张勇. 维吾尔谚语研究视角与理论方法琐谈[J]. 新疆大学学报: 社会科学版, 2004, 32(1): 130-134.
- ZHANG Yong. On the perspective and theoretical methods of Uyghur proverb research[J]. Journal of Xinjiang university: social science edition, 2004, 32(1): 130-134.
- [13] 阿比达·吾买尔, 吐尔根·依布拉音. 维吾尔语句子边界识别算法的设计与实现[J]. 新疆大学学报: 自然科学版, 2008(3): 360-363.
- UMAR A, IBRAYIM T. Research and implementation of Uyghur sentence boundary detection[J]. Journal of Xinjiang university: natural science edition, 2008(3): 360-363.
- [14] 艾斯卡尔·亚克甫. 维吾尔谚语检索系统的研究与实现[D]. 青岛: 中国海洋大学, 2012: 34-70.
- YAKUP A. Research and realization of Uyghur ideom retrieval system[D]. Qingdao: Ocean University of China, 2012: 34-70.
- [15] 刘正江. 谈维吾尔语成语汉译的可译性限度[J]. 西北民族大学学报: 哲学社会科学版, 2005(3): 47-50.
- LIU Zhengjiang. On the translatability limits of Uyghur idioms in Chinese translation[J]. Journal of northwest university for nationalities: philosophy and social science, 2005(3): 47-50.
- [16] 艾孜尔古丽·玉素甫. 现代维吾尔语常用词计量研究[D]. 乌鲁木齐: 新疆师范大学, 2013: 40-48.
- AZRAGUL. The research on modern Uyghur language common words[D]. Urumchi: Xinjiang Normal University, 2013: 40-48.
- [17] 艾孜尔古丽, 努尔艾合买提, 玉素甫·艾白都拉. 现代维吾尔语常用词统计关键技术研究[J]. 中文信息学报, 2014, 28(5): 192-197.
- AZRAGUL, NURAHMAT, ABAYDULA Y. Research on key technology of modern Uyghur language common[J]. Journal of Chinese information processing, 2014, 28(5): 192-197.
- [18] 玉素甫·艾白都拉, 艾孜尔古丽, 祖丽皮亚. 基于网站用词调查的现代维吾尔语词长研究[J]. 计算机应用与软件, 2012, 29(5): 32-34.
- ABAYDULA Y, AZRAGUL, ZULPIYA. Research on word length of modern Uyghur based on surveying website wording[J]. Computer applications and software, 2012, 29(5): 32-34.
- [19] 李燕萍. 从维吾尔语谚语中的复句看现代维吾尔语复句的演变[J]. 喀什师范学院学报, 2006, 27(5): 57-60.
- LI Yanping. Observe and study the development orbit of the sentence of two clauses in the modern Uighur language from those are retained in the Uighur proverbs[J]. Journal of Kashgar teachers college, 2006, 27(5): 57-60.
- [20] 付东明, 陈得军. 维吾尔谚语研究趋势与反思[J]. 语言与翻译, 2014(1): 29-33.
- FU Dongming, CHEN Dejun. Uyghur proverb research trends and reflection[J]. Language and translation, 2014(1): 29-33.

作者简介:



穆妮热·穆合塔尔, 女, 1989年生, 博士, 主要研究方向为多种语言信息处理、机器翻译、自然语言处理。



李晓, 男, 1957年生, 研究员, 主要研究方向为多语种信息处理、人工智能。主持或承担过多项国家 863、中科院战略先导项目。发表学术论文 60 余篇。



杨雅婷, 女, 1985年生, 副研究员, 博士, 主要研究方向为机器翻译和自然语言处理。承担过多项国家 863、中科院战略先导项目。发表学术论文 30 余篇。