

基于语料库的以“gɛp”构成的固定短语的多维研究

周雪晴, 阿孜古丽·夏力甫

新疆大学中国语言文学学院, 新疆 乌鲁木齐

收稿日期: 2026年6月22日; 录用日期: 2026年7月23日; 发布日期: 2026年8月5日

摘要

本文基于自建的以gɛp构成的固定短语语料库,从句法与语义两个维度对相关短语展开系统性分析。句法研究维度,引入生成语法理论中的X标杆理论,对gɛp固定短语的语法结构、层级构造规律系统进行梳理与分析;语义研究维度,结合言语行为动词相关研究对gɛp构成的言语行为短语进行分类与阐释。本研究通过多种研究方法,有效揭示此类短语在句法层面的结构规律与在交际层面的动态语义功能。不仅为维吾尔语固定短语的系统分析提供了新的视角,也为后续维吾尔语短语资源库的建设与完善提供了实践参考。

关键词

语料库, gɛp, 固定短语, X标杆理论, 言语行为动词, 语义特征

A Multi-Dimensional Study of Fixed Phrases Formed with “gɛp” Based on a Resource Database

Xueqing Zhou, Aziguli·Xialifu

School of Chinese Language and Literature, Xinjiang University, Urumqi Xinjiang

Received: June 22, 2026; accepted: July 23, 2026; published: August 5, 2026

Abstract

This paper carries out a systematic analysis of fixed phrases featuring gɛp based on a self-established corpus from syntactic and semantic aspects. Drawing on the X-bar theory of generative grammar, the

study explores the grammatical structure and hierarchical formation rules of these phrases in syntactic analysis. In semantic research, it classifies and explains relevant speech act phrases grounded on existing studies of speech act verbs. Employing multiple research methods, the paper identifies syntactic structural rules and dynamic semantic functions in interpersonal communication. It puts forward a new perspective for studying Uyghur fixed phrases, and offers practical guidance for the development and improvement of Uyghur phrase resource repositories.

Keywords

Corpus, gep, Fixed Phrases, X-Bar Theory, Speech Act Verb, Semantic Features

Copyright © 2026 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

语料库构建在机器翻译、词典编纂、第二语言学习等方面具有重要作用。随着人工智能与自然语言处理技术的快速发展,语言资源构建研究日益深入,其中语料库构建的研究尤为突出。现有成果中,胡传成(2015) [1]分析了语料库语言学方法在教学中的优势;祖力克尔江·司坎旦尔(2020) [2]则通过爬取“天山网”汉语新闻语料,经预处理构建了对应的词汇语料库,展现了完整的资源构建流程。不过目前,相关语料库研究多集中于两种语言的对齐技术以及语料库的实际运用,尚未系统延伸到固定短语等语言单位的深度挖掘与构建。

在维吾尔语句法与短语结构研究方面,学界已建立起较为成熟的分析框架。马德元(1993) [3]系统评述了维吾尔语短语的早期研究成果,针对短语定义、边界划界、类型标准及组合规则等争议问题展开梳理。厘清复合词与短语的边界,建立多维度的短语分类体系。张玉萍(1998) [4]进一步从句法平面切入,明确维吾尔语短语包含“实词 + 实词”“实词 + 虚词”两类核心组合形式,通过层次分析与结构模式归纳,系统阐释了短语的句法结构与内部句义、语法关系,重点揭示维吾尔语短语“同形异构”“同构异形”的特殊结构规律。

在固定短语研究层面,国内固定短语研究多集中于宏观特征归纳与两种语言对比研究,研究对象多为通用成语、惯用语,如刘珉《汉维共时对比语法》 [5]基于共时视角,对照汉维 13 类词类语法特征,重点对比两种语言短语的构成方式、结构类型与组合手段,揭示了词序、虚词在汉维短语建构中的差异化规律。针对固定短语细分范畴,申辉、柳慧(2022) [6]聚焦成语类固定短语,对比阐释汉维成语的定义、来源与基本特征,指出二者来源与界定存在差异,但成语结构的固定性是汉维两种成语所共有的特征。

在言语行为理论应用研究层面,言语行为理论由奥斯汀提出,塞尔完善这套理论,核心观点是说话本身就是做事。国外先开展相关研究。首先奥斯汀区分两类话语,定下显性施为句基础规则。塞尔划分四层言语行为,提出五大施事行为类型。后续学者继续补充施为动词相关特点。国外学者 Wierzbicka [7]专门整理英语言语行为动词,给出分类思路。理论还用到外语教学、话语分析、人工智能对话多个领域。国内学者吴剑锋 [8] (pp. 39-42)专门整理汉语里的言语行为动词,分清言语行为动词和施为动词区别。钟守满 [9] (pp. 89-90)对比英汉两类动词,从信息传递、句式角度找出两者异同。

综上,国内学者已经搭建起维吾尔语句法、固定短语、言语行为三类基础研究框架。在句法研究层面,这些研究囿于语法单位层面的静态描写;在固定短语层面,现有研究很少搭建专属语料库,缺少短语内部搭配、使用场景的统计分析;言语行为理论相关研究,国内多用来分析汉语或汉英对比分析。

本文承接前人研究的基础内容，沿用已有句法分析形式，吸收言语行为理论的分析方法。依托自建的维吾尔语 *gep* 固定短语语料库，从句法和语义两个维度展开系统分析：句法层面，借助 X 标杆理论揭示其内部结构与层次构造规律；语义层面，则分析 *gep* 构成的言语行为短语。

2. *gep* 构成的固定短语语料库的构建与概览

现代维吾尔语的固定短语与现代汉语中的固定短语一样，其结构具有固定性、语义具有整体性。在把握固定短语特点的基础上，通过词典、国语助手、izdax 等主流翻译软件进行交叉验证，收集整理以 *gep* 构成的固定短语。采用 Excel 作为语料库的数据模型，设置国际音标、释义、结构类型等字段。语料库为每个短语条目设置的核心字段如表 1 所示。

Table 1. Schematic design of core corpus fields
表 1. 语料库核心字段设计示意

字段名称	说明与示例
固定短语	<i>gep tapmaq</i>
释义	无中生有
结构类型	动宾短语
句法结构	(NP + 主格) + VP
例句	<i>sen emdi gep tapma.</i>
例句译文	你不要无中生有了。

由上表 1 可知，语料库内的某个固定短语可以通过语义、结构等方面进行描述。如 *gep tapmaq* (无中生有)，*gep* 作为 *tapmaq* (找，求得)所涉及的对象，二者构成动宾结构。例句意为“无中生有”中 *gep tapmaq* 以否定形式出现，在句子充当谓语成分，陈述主语 *sen* (你)的行为。

从 *gep* 构成的固定短语的结构类型来看，*gep* 能与动词、名词、形容词等相结合构成短语，展现了其强大的组合能力。如图 1 所示。其中，以 *gep* 构成的动宾短语(67 个)和偏正短语(66 个)在数量上占据主导地位，如动宾短语 *gep ewetmek* (捎话)、*gep almaq* (探口风) *gep ifitmek* (挨骂)等；偏正短语还可分为定中结构如 *gep anisi* (话匣子)，状中结构如 *gepke qalmaq* (落不是)。此外，还有谚语，如 *gepni ojlap de, nanni fjajnap je.* (话要想着说，饔要嚼着吃)，是劳动人民生活经验的凝练体现，揭示道理。

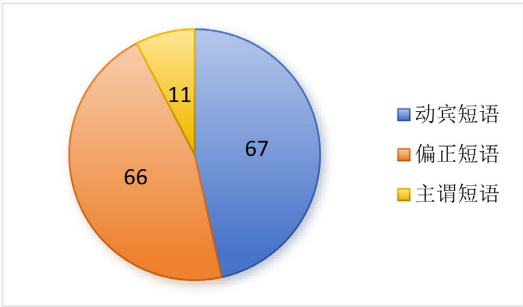


Figure 1. Distribution of structural types of fixed phrases with *gep*
图 1. 以 *gep* 构成的固定短语结构类型分布

综上所述，通过对自建的以 *gep* 构成的固定短语语料库从其核心字段设置、固定短语结构类型分布进行描述，为后续基于语料库对 *gep* 固定短语进行更深入的句法与语义研究奠定基础。

3. gep 构成的固定短语句法分析

力提甫·托乎提在其著作《现代维吾尔语参考语法》中指出：“维吾尔语是核心词在后的语言。核心词出现在补足语或附加语之后，或者被修饰语出现在修饰语之后。”[10]由 gep 构成的固定短语覆盖动宾短语、偏正短语、主谓短语三类核心类型。下面以力提甫·托乎提的短语界定与格省略规则为依据，参照《生成语法框架内的维吾尔语句法》中对维吾尔语句法描写规则——“将格附加成分确认为 K (格) 语类，把所谓的副动词、形动词和动名词词缀等确认为 SUBS (静词化) 语类，把时和人称附加成分确认为 INFL (屈折) 语类。”[11]对 gep 固定短语的内部结构、格标记功能、层级投射进行统一分析。

动宾短语由及物动词为中心语、名词性成分作受事宾语构成，典型语序为宾语(KP) + 动词(V)，gep 构成的动宾短语以动词为中心，gep 充当受事论元，按宾格标记有无分为两类。

无宾格标记的 gep 动宾短语中，gep 以零格形式直接充当动词补足语。

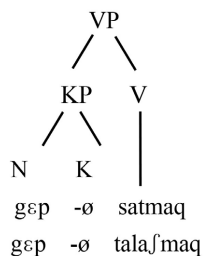
(1) ular ketfitfe gep setip tñiqti. (他们整晚高谈阔论。)

(2) hetqajsimiz gep talañmañli. (咱们谁也别争执。)

例 1, ular (他们)以主格形式出现，充当句子的施事；谓语部分“gep setip tñiqti”中 satmaq 以状态副动词的形式出现与助动词 tñiqmaq 进行连用，表达“一整晚持续交谈”的动作，整体在句子中充当谓语。

例 2, gep talañmaq (争执，抬杠)整体作为句子的谓语以祈使式出现，表达“不要争执”这一倡议。

上述例句中，短语 gep satmaq (高谈阔论)和 gep talañmaq (争执，抬杠)中 gep 作为 satmaq (出售)、talañmaq (争，抢)的受事论元，以“零”格词尾出现。可以用树状图表示，如下：



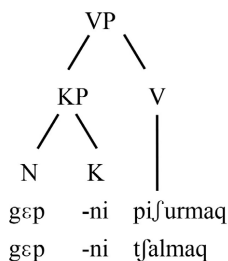
有宾格标记的 gep 动宾短语中，gep 以宾格(ni)的形式出现，与动词之间的支配关系明确。

(3) ular keliñtin burun biz gepni piñururwalajli, heli bolmisa tñenip qalidu. (他们来之前，我们统一口径，免得露马脚。)

(4) gepni tñalkınmañ,men gepimni qiliwalaj. (别打岔，让我说完。)

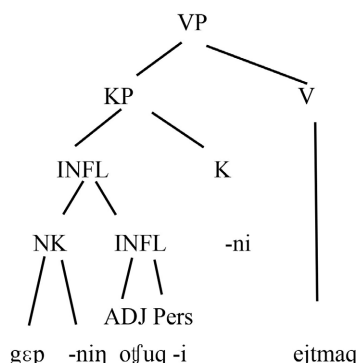
例 3, biz (我们)作为施事主语，gepni piñururmaq (统一口径)充当谓语描述主语的动作行为；例 4，复句中出现了两个由 gep 构成的固定短语 gepni tñalmaq (打岔)和 gepni qilmaq (说话)，两者均为分句的谓语成分。

上述例句中，gep 作为被强调的对象，与动词结合构成动宾关系的短语在句子中作谓语。短语结构树状图如下：



(5) gepnij ofuqini ejtmaqligimni ketfyrj. (请恕我直言。)

gepnij ofuqini ejtmaq (打开天窗说亮话)的句法功能相当于一个不及物动词, ofuq (开放)经名词化后接受 gep 的领有和限定与之搭配组成的名词短语同中心语 ejtmaq (说)构成动宾短语, 充当句子的谓语。短语结构树状图如下:

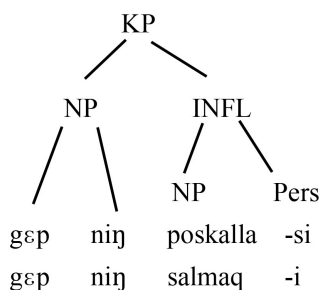


遵循修饰语前置、中心语后置的语序, 偏正短语其结构为修饰语 + 中心语, 分为定中偏正(名词性)与状中偏正(动词性)两类。gep 在偏正短语中可充当修饰语或中心语, 分别构成定中短语、状中短语。gep 在定中结构中可作(领属)修饰语或被描写中心语, 因此可分为 gep (领属格) + 名词、gep + 名词和形容词 + gep 三种结构类型。

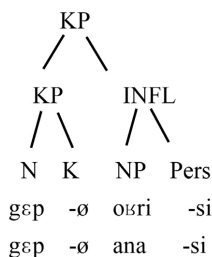
gep 作领属修饰语的定中短语中, gep 以领属格的形式出现, 对其他名词进行限定。

(6) gepnij poskallisi, unij iqtidari nahajiti jaxʃi. (简而言之, 他的能力很出色。)(7) unij ejide bowa gepinij salmiqi bek exir. (在他家爷爷话的分量很重。)

例 6, gepnij poskallisi (简而言之)在句中位于句首作为插入性状语, 对后续内容进行总括、归纳; 例 7, gepinij salmiqi (话的分量)在句子中充当主语, 谓语 bek exir 对其性质进行描述; poskalla (正题)和 salmaq (分量)受 gep 的限定, 需要附加对应的人称词尾(si/i)短语结构树状图如下:

(8) gep oxrisi neq mejdanda tutuldi. (窃听者被当场抓了现行。)(9) bu gep anisi gep qilsila toxtimajdu. (这个话匣子说起话来没完没了。)

例 8, gep oxrisi (窃听者)是整个句子的主语, tutmaq (抓)以被动语态的形式出现, 陈述主语的遭遇, neq mejdanda (在现场)是动作发生的地点状语, 补充说明动作的发生场景; 例 9, gep anisi (话匣子)作为整个句子的施事主语。gep 修饰 oxri (贼)和 ana (母亲), 对其的范围、性质进行缩小, 需要附加对应的人称词尾(si/i)。短语结构树状图如下:

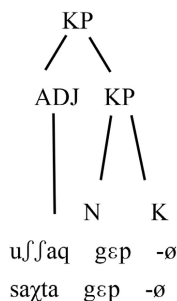


геп 作为中心名词，形容词作为前置修饰语，对其进行描写或限定，构成形容词 + геп 型定中短语。

(10) uffaq gep keipijatinizka tesir jetkyzmisun. (别让烦言碎语扰了心境。)

(11) bu retsip sayta gəpkə ox[ʃɑ]. (这药方像鬼画符一样。)

例 7 和例 8 中, 两个短语均属于“形容词 + *геп*”构成的定中结构(偏正短语), *уффаг геп* (烦言碎语) 在句中充当主语。它是后面谓语部分 *тесир жеткызмисун* 所描述的动作的发出者; *сахта геп* (鬼画符) 附加向格, 表示相似的目标, 在句中充当状语, 修饰谓语形容词 *оџаф* (像……一样) 对主语 *ретсеп* (药方) 进行描述。短语结构树状图如下:

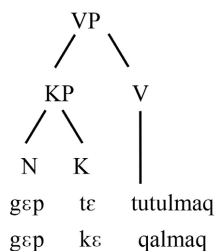


gep 构成的状中结构偏正短语, 核心特征是“gep”充当状语, 修饰后续的动词性中心语, 说明中心语的方式、状态、范围或条件。与定中结构不同, 状中结构的核心是动作的修饰与限定, 凸显动作的实施方式或状态。

(12) fewket özünün gepte tutulup qalxanlıqını hes qildi. (肖开提感觉到自己说漏嘴了。)

(13) *seni dəp gəpkə qaldim.* (因为你我才落不是。)

例 10, 由 *gep* 构成的状中短语 *gepte tutulmaq* (漏嘴) 以状态副动词出现与助动词 *qalmaq* 结合, 整体附加动名词 (*kanliq*) 词尾, 作为动词 *hes qilmaq* (感觉) 的受事, 充当句子的宾语成分; 例 11, 句子中隐含 “*mən*” (我) 这一主语, *seni dəp* (因为你) 修饰整个谓语成分 “*gepkə qaldım*” (落不是)。在以上两个短语中动词 *tergəsmək* (追问)、*qalmaq* (落, 陷于) 为中心语, *gep* 附加时位格 (*te*) 和向格 (*kə*) 作为动作发生的地点或趋向。短语结构树状图如下:



力提甫·托乎提指出:“在主谓式固定词组中主语不带主格。”[12] *gep* 构成的主谓短语中,其作为被陈述的对象,以“*gepi*”的形式出现且不带主格。有 *gep* + 形容词和 *gep* + 名词两种结构类型。

(14) bu *gepi tyz* qiz iken. (这是个心直口快的姑娘。)

(15) unin *gepi dzajida*. (他说得有道理。)

例 14, *gepi tyz* (心直口快)作为主语 *qiz* (女孩)的定语;例 15, *gepi dzajida* (言语得体)对主语 *u* (他)的行为进行阐述,充当句子的谓语成分。短语结构树状图如下:



综上所述,探究以“*gep*”为核心的各类短语的句法功能,参考 X 标杆理论对 *gep* 固定短语内部结构进行划分,并构建不同短语类型的结构树状图,为后续维吾尔语短语句法研究提供实践支撑。

4. *gep* 构成的固定短语的语义分析

言语行为动词研究源于奥斯汀的言语行为理论,在日常交际中人们离不开这类词。在国内关于言语行为动词的研究中,学者们普遍认同动词在句子结构中的核心地位,并从语义认知视角展开了深入探讨。钟守满(2004) [9] (p. 89)认为英汉言语行为动词从信息传递动作方向、客体认知等方面进行分类。王熙华(2016) [13]从言语行为理论中的行事行为视角分析维吾尔族情景剧,表明言语行为理论能有效阐释维吾尔语日常交际中含“指令式”“断定式”“表情式”的言语行为。

吴剑锋在《现代汉语言言语行为动词研究》中对言语行为动词进行阐释:“言语行为在语言词汇层面的投射形成言语行为动词,如‘告诉’‘命令’‘问’‘宣布’‘批评’‘骂’等”[8] (p. 34)。依据吴剑锋提出的三大界定原则,维吾尔语言言语行为动词短语需满足两项核心要求:一是基本功能为阐释言语行为本身,二是语义具有整体性。由 *gep* 构成的固定短语,正是维吾尔语言言语行为动词体系中的重要组成部分。

言语交流具有交互性的特征,其集中体现在“信息在听说双方的相互传递”[8] (p. 54)。因此,言语行为存在“方向”这一属性,存在两种方向:一种是单向的,从信息发出者向信息接受者传递,体现“给出”或“获得”,如“告诉”“建议”等言语行为;另一种是互向的,信息发出者和信息接收者同时向对方发送信息,双方协同动作,既体现“给出”又体现“获得”,如“讨论”“商量”等言语行为。依据言语交际的“方向性”这一属性,可以将维吾尔语以 *gep* 构成的固定短语分为单向言语行为短语、互向言语行为短语。

以 *gep* 构成的单向言语行为短语具有[+单向]的特征,在此基础上对这类短语根据[±给出/获得]和[±认知/情感/意志]这两组区别性语义特征,可分为以下四类:

告知类言语行为短语核心表现为说话人输出观点、信息,听话人被动接受,没有明显的情感和意志倾向,体现[+给出][+认知-情感-意志]的语义特征。

(16) gep qilgili qojsaŋ bolmamdu? (让我说话可以吗?)

(17) ezizizni besiwiliŋ burader, gepni jorinatsiniz tŋataqniŋ tŋoni tŋiqidu. (冷静点兄弟, 你把话说大了可就麻烦大了。)

(18) u jirinka jygryeemeslikidin endiŋe qilip, sawaqdifiŋi muellimge gep ewetken. (他担心赶不上会议, 就让同学给老师捎了话。)

例 16, 说话人表达想要“说话”这一行为, 句中的反问语气让 gep qilmaq (说话) 具有了情感性的表现, 但从核心语义看, 仍是请求获得“输出信息”的祈使行为; 例 17 是说话人对听话人的 gepni juyinatmaq (夸大其词) 这一言语行为进行客观的评述与提醒; 例 18 是说话人需要通过同学 gep ewetmek (捎话) 给老师的事实陈述。

询问类言语行为短语核心表现为说话人获得信息、发起疑问, 核心意图在于从听话人处获得反馈信息, 体现[+获得][+认知-情感-意志]的语义特征。

(19) uniŋ tujuqsiz alimge gep taŋliŋusi kelip qaldi. (他突然间想试探一下阿里木的口风。)

(20) men qajta - qajta gep qoqlap, andin sewebini ejtti. (我反复追问, 他才说出原因。)

(21) zerikiŋin tutuluŋ ytfyn uni gepke salqum keldi. (为了消除寂寞, 我想把他的话套出来。)

例 19, 说话人主动 gep taŋlimaq (试探) 对方态度与想法, 目的是从对方那里获取隐含信息; 例 20, 说话人反复向听话人 gep qoqlamaq (追问), 核心意图是让对方提供具体原因; 例 21, gepke salmaq (套话) 不是直接的提问, 具有一定的策略性, 但本质仍是以获取信息为目的。

表态类言语行为短语核心表现为说话人输出情感态度, 核心目的是引发听话人情感回应, 体现[+获得][-认知+情感-意志]的语义特征。

(22) seniŋ destiŋdin anamdin dŋiziq gep aŋlidim. (因为你挨妈妈的骂了。)

(23) u xataliq otkyrgendin kejin, rehberlik uni qattiq gep qojdi. (他犯错后, 领导严厉地责备了他。)

上述两个例子中, 均表达说话人被 gep aŋlimaq (训斥) 和 gepke qojmaq (责备) 后传递出委屈、不满的情绪, 想要引发听话人的情感回应。

承诺类言语行为短语, 这类短语中说话人单向传递“承诺”“保证”等意志, 体现[+给出][-认知-情感+意志]的语义特征。

(24) biz adem boluŋta, iŋ qiliŋta gepimizde turuŋimiz kerek. (我们做人做事都得说话算数。)

(25) li muellim gepi gep adem. (李老师说说话算数。)

上述两个例句中 gepide turmaq 和 gepi gep 都表达“说话算数”这一承诺性意志, 说话人向听话人传递可靠、履约的态度。

以 gep 构成的互向言语行为短语具有[+互向]的特征, 信息双向传递, 说话人与听话人互为信息发出者与接收者, 体现[+给出][+获得]的语义特征。

(26) ular sewebsizla gep tegiŋip qaldi. (他们无缘无故抬起了杠。)

(27) tŋonlar gep qiliwatqanda, kitŋik balilar gepke qofuq salmanlar. (长辈说话的时候小孩们别插话。)

例 26, gep tergeŋmek (抬杠) 核心表达“交际双方互相反驳、争执不下”这一互向言语行为; 例 27, 听话人在对方说话过程中, gepke qofuq salmaq (插话) 既接收对方信息(获得), 又主动输出新内容(给出), 形成双向互动。

综上所述, 参照吴剑锋《现代汉语言行为动词研究》中的界定标准[8], 结合自建语料库, 对以 gep 构成的言语行为动词短语进行界定和分类, 并对每类短语的语义特征进行描述。该类短语功能上主要用

于阐释言语行为本身, 符合言语行为动词的基本特征。

5. 结语

本文对维吾尔语中以 *gep* 构成的固定短语进行语法语义分析, 揭示以 *gep* 构成的固定短语的内部结构规律; 从言语行为理论的视角出发, 分析该类短语在语言交际中的功能并对其进行分类。总言之, 以 *gep* 构成的固定短语是维吾尔语表达言语行为的重要手段。本研究既印证了言语行为理论、句法理论在维吾尔语研究中的适用性, 也为维吾尔语短语结构、语义分类和言语行为研究提供了新的实证材料。尽管本研究力求在多维分析框架下实现对以 *gep* 构成的固定短语的系统描写与解释, 但自建语料库的规模相对有限, 在一定程度上影响了研究结论的代表性与外推性。后续研究将着力于语料资源的扩充与完善, 进一步细化 *gep* 及其相关短语的语义特征分类, 力求更全面、深入地揭示维吾尔固定短语的系统规律。

参考文献

- [1] 胡传成. 语料库在维吾尔语词汇教学中的应用研究[J]. 语言与翻译, 2015(2): 86-90.
- [2] 祖力克尔江, 艾孜海尔江, 艾孜尔古丽. 汉维双语平行词汇语料库构建技术研究[J]. 电脑知识与技术, 2020, 16(6): 202-204.
- [3] 马德元. 维吾尔语词组研究述评[J]. 新疆大学学报(哲学社会科学版), 1993(1): 100-104.
- [4] 张玉萍. 维吾尔语短语结构分析[J]. 民族语文, 1998(5): 76-80.
- [5] 刘珉. 汉维共时对比语法[M]. 乌鲁木齐: 新疆人民出版社, 1991.
- [6] 申辉, 柳慧. 汉维成语对比分析[J]. 文化创新比较研究, 2022, 6(7): 58-61.
- [7] Wierzbicka, A. (1987) *English Speech Act Verbs*. Academic Press.
- [8] 吴剑锋. 现代汉语言行为动词研究[M]. 北京: 北京大学出版社, 2016.
- [9] 钟守满, 张伟华. 英汉言语行为动词分类及其语义认知解释[J]. 上饶师范学院学报(社会科学版), 2004(5): 88-91.
- [10] 力提甫·托乎提. 现代维吾尔语参考语法[M]. 北京: 中国社会科学院出版社, 2012.
- [11] 力提甫·托乎提. 生成语法框架内的维吾尔语句法[J]. 民族语文, 2005(6): 35-43.
- [12] 力提甫·托乎提. 维吾尔语格的省略和 X 标杆理论[J]. 民族语文, 1999(2): 65-71.
- [13] 王熙华, 史倡铭. 从言语行为理论中的行事行为视角看维吾尔族情景剧——以《幸福家庭的烦恼》为例[J]. 开封教育学院学报, 2016, 36(1): 56-57.