

基于小型语料库的维吾尔语 名词构形语素统计分析

拜佳宁, 王 静

新疆大学中国语言文学学院, 新疆 乌鲁木齐

收稿日期: 2024年7月24日; 录用日期: 2024年8月29日; 发布日期: 2024年9月5日

摘 要

维吾尔语作为一种形态结构复杂的语言, 其词与词之间的关系常常需要采用附加词尾的构形语法手段来表达, 名词是维吾尔语中的一种重要词类, 因此分析维吾尔语中的名词构形语素非常重要。文章主要通过切分统计基于YiCorpus平台所建立的法律法规小型语料库中出现的名词构形语素来分析不同构形语素出现频率不同的原因, 为维吾尔语教育教学提供一定的帮助, 对计算机语言学的进一步研究给予支持, 也为语言研究与语言教学提供更加可靠的统计数据。文章采用了文献研究法、统计分析法以及语料分析法对语料库内的语料进行收集、整理、分析, 总结出了维吾尔语名词构形语素的使用特点及规律, 通过此次研究发现, 针对法律 + 维吾尔语双学位学生的语言教学和面向司法干部的语言教学有一些不同于普通的第二语言教学的特点和规律, 教学人员应该根据教学对象的不同采取不同的教学方法和策略。

关键词

维吾尔语语料库, 构形语素, 词尾, 语料切分

Statistical Analysis of Uyghur Noun Morphological Morphemes Based on a Small Corpus

Jianing Bai, Jing Wang

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

Received: Jul. 24th, 2024; accepted: Aug. 29th, 2024; published: Sep. 5th, 2024

Abstract

Uyghur, as a language with a complex morphological structure, often uses morphological

文章引用: 拜佳宁, 王静. 基于小型语料库的维吾尔语名词构形语素统计分析[J]. 现代语言学, 2024, 12(9): 135-144.

DOI: 10.12677/ml.2024.129765

grammatical methods, such as suffixes, to express the relationships between words. Nouns are a significant part of the Uyghur language; hence, analyzing the noun morphological morphemes in Uyghur is crucial. This article primarily focuses on studying the noun morphological morphemes within a specialized legal corpus developed on the YiCorpus platform, investigating the varying frequencies of different noun morphological morphemes. This analysis aims to aid Uyghur language education and instruction, support further research in computational linguistics, and provide more reliable statistical data for both language research and teaching. The study employs methods such as literature review, statistical analysis, and corpus analysis to collect, organize, and scrutinize the corpus. It summarizes the usage characteristics and patterns of noun morphological morphemes in Uyghur, revealing specific distinctions and conventions in language teaching for students dual-majoring in law and Uyghur, as well as for judicial officials, compared to general second language instruction. Based on these findings, educators should adopt varied teaching methods and strategies tailored to the different needs of their students.

Keywords

Uyghur Language Corpus, Morphological Morphemes, Suffix, Corpus Segmentation

Copyright © 2024 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

语料库语言学是一门非常实用的科学,从诞生至今它不断伴随着普通语言学的发展而发展,并不断完善自身。如今语料库语言学不再仅局限于以往简单枯燥的语法、词汇和词典编纂,它涵盖了二语习得、外语教学与研究、翻译服务和实践、话语分析、语言处理等更为专业和全面的诸多语言学研究的领域,使其得到前所未有的发展[1]。语料库语言学的发展是以电子计算机技术的广泛应用为基础的,随着电子计算机技术的升级和相关软件的研发,语料库的构建和应用越发便捷,为我们研究语言现象和语言内部结构提供了诸多便利。

同时维吾尔语作为一种通过附加语缀来表示语法意义的语言,分析其构形语素就显得十分重要,有些关于维吾尔语构形语素的研究并不是很多,在二十世纪六十年代由苏联从事维吾尔语研究的语言学家集体编写的《现代维吾尔语》中提到过“词素”这一概念。哈密提·铁木尔编写的《现代维吾尔语语法》一书中将语素分为了独立语素和词缀语素,又将其中的词缀语素分为构词语素和构形语素。邓浩在《论维吾尔语构形语素的历史发展》[2]中从功能-语义角度对名词的数、人称和格位形式的发展进行了探索研究,《维吾尔语的构形语素来源新探》[3]中对维吾尔语构形语素的来源做了新的阐释;李圃在《现代维吾尔语语素研究》[4]中对现代维吾尔语语素做了较为全面的考察,将维吾尔语语素进行了分类,在功能分类中的黏着语素分类中将语素分为了构词语素、转词语素、构形语素。亚热·艾拜都拉在《论维吾尔语构形语素的多义性——兼论粘着语和屈折语的共同点》[5]中通过实例证明在维吾尔语中一个构形语素可以同时表示几种语法意义,而同一个语法意义也可以用不同的构形语素来表示的语言现象。方晓华在《论维吾尔语的语素》[6]中对维吾尔语语素进行了分类,在黏着语素中分析了构形语素。霍盛在《关于维吾尔语附加成分的划分》[7]中深入研究了维吾尔语附加成分。

本文基于小型汉维法律法规平行语料库来展开研究,利用文献研究法、语料分析法以及统计分析法来分析名词构形语素在自然语言中出现频率背后的原因,为语言研究与语言教学提供更加可靠的统计数

据, 对学生的学习和老师的教学提供帮助, 也为计算语言学的进一步研究给予支持。

2. 语料库介绍及语料切分

2.1. 语料库介绍

语料库是指一个由大量的语言实际使用的信息组成的, 专供语言研究, 分析和描述的语言资料库。所谓平行语料库是指由一种语言的原始文本和它在另一种(或多种)语言中的翻译文本组成的双语(或多语)对照的语料库[8]。平行语料库主要用于语言研究、翻译研究、翻译培训、外语教学、词典编纂等; 它使翻译研究完成了从规定向描写的转变。本文主要围绕着大学生创新计划项目所建立的汉维法律法规平行语料库中的语料展开研究, 该语料库是基于 YiCorpus 平台所建立的, 共有 205 条法律法规, 维吾尔语语料总词数有 815,857 个, 语料内容均为国内法律法规, 库中语料均为项目小组成员收集对齐导入所获得的, 该语料库内的数据对于维语言专业进行汉维语语法、句法的对比研究, 对汉、维两种语言的法律法规语言的使用特点、翻译特点以及翻译时出现的问题的研究具有较高的参考、科研价值。

2.2. 语料切分的要求及步骤

维吾尔语是形态非常丰富的一种语言, 是一种黏着性语言, 丰富的构形词缀连接词干, 表示数、格、时态等语法功能[9], 在分析研究维吾尔语构形语素的过程中语料切分是十分重要的一个步骤, 我们利用 YiCorpus 平台对前期录入的 205 份文本文件中的词频进行统计汇总, 整合导出形成 excel 表格然后进行语料切分, 由于目前还没有较好的自动切分维吾尔语的平台所以我们采取了人工切分的方式, 完成初步的切分后为了确保准确再次进行了交换检查切分语料的结果, 最后由指导老师进行了最终的检查校对, 通过这些步骤也大大提升了切分结果的正确性。同时对于切分语料也有相应的要求, 切分过程中需将每一个词语后的构形语素全部切分出来, 不可漏切错切, 如维吾尔语词语“belgilengen”(规定的)需要切分成“belgile + n + gen”不可只切分成“belgilen + gen”。

2.3. 语料切分过程中遇到的困难

语料切分是完成论文过程中非常重要的一步, 同时它也是难度较大的一项工作, 需要切分语料的成员具有一定的维吾尔语基础, 同时切分的词语数量也较多, 需要花费大量的时间和精力, 对于准确性的要求也很高, 因为切分结果是论文统计分析构形词缀的依据, 在切分过程中也遇到了困难。维吾尔语存在语流音变的现象, 由于语音和谐的规律, 部分词语再加上词缀以后会出现弱化增音的现象[10], 例如“ifxana”加构形语素“niq”以后“ifxana”最后一个音节中的“a”弱化成了“i”, 该词语最后写为“ifxaniniq”对于这类词语都需要人工切分完成后还原该词原本的形式, 这也需要切分人员充分了解维吾尔语语流音变的变化规则并且运用到切分过程中去。

在切分统计的词语皆是语料库中法律法规的条文, 所以里面的部分词语是专业术语, 因此造成了部分词在单独看的前提下可以切分成词根+构形语素的形式, 但是在具体的条文中则是专门的法律词语, 这种情况下需要具体语境具体分析。

3. 维吾尔语名词构形语素的统计与分析

3.1. 领属性人称附加成分的统计与分析

维吾尔语名词的人称范畴是通过名词的形态变化表示名词所指事物的领属关系的语法范畴。表达的概念包括第一人称单数、复数, 第二人称单数、复数和第三人称(形式上不分单复数)[11]。

根据语料库中的语料的统计切分结果, 可以将库中领属性人称附加成分归纳如表 1。

Table 1. Statistical table of additional components for personal names of leading attributes**表 1.** 领属性人称附加成分统计表

| 领属性人称 | 构形语素 | 出现次数 |
|--------|------|--------|
| 第一人称单数 | im | 303 |
| | ym | 268 |
| | m | 11 |
| 第一人称复数 | imiz | 98 |
| | miz | 4 |
| 第三人称 | i | 38,346 |
| | si | 20,921 |

结果表明, 语料库中名词及名词性词语的领属人称有第一人称单数、复数和第三人称三种类型的构形语素, 没有出现第二人称的领属性人称, 其中出现数量最多的是第三人称领属性人称附加成分的两种形式。出现这个结果原因如下:

此次统计是基于法律法规语料库中的语料展开的, 库内语料均为法律法规的相关条款, 条款内容多以第三人称来书写的, 例如“*χελq høkymiti*”(人民政府); “*memurij mes’ul tarmiqi*”(行政主管部门)因此第三人称领属性人称词尾出现的频率极高, 明显多于其它领属性人称词尾, 同时虽然“i”和“si”均为第三人称附加词尾, 但附加在辅音后的“i”词尾数量是附加在元音后的“si”的 1.8 倍左右, 说明库中语料多为辅音结尾的词语; 而第一人称词尾则基本出现于“*dølitimiz*”(我们国家)这类词语中以及主体领有句式当中, 频率较低。

由此可见, 在库中语料中领属性人称附加成分的第三人称附加成分的出现频率总是远高于其他人称, 此次结果对老师教学和学生学习的参考价值, 老师可以在授课过程中优先讲授第三人称词尾, 学生在学习过程中也可以对第三人称词尾多加注意, 这样可以更好掌握维吾尔语语法, 提高教师的授课水平和学生的学习效率。

3.2. 格附加成分的统计与分析

维吾尔语名词的“格”指的是形态格。它是通过名词的形态变化表示名词与其它词之间的各种关系的语法范畴[11]。代词和名词化的其它类词也同名词一样可以按名词的格来变化。接下来将从领属格、向格、宾格等八个格位进行分析论述。

3.2.1. 领属格附加成分统计与分析

名词的领属格形式是由名词后附加词尾“niq”构成的, 例如: “*uniq kitabi*”(他的书)领属格表示名词与其他词之间的领属关系。库中语料领属格附加成分统计结果如表 2。

Table 2. Statistical table of additional components in the leading element**表 2.** 领属格附加成分统计表

| 构形语素 | 出现次数 |
|------|--------|
| niq | 33,997 |

结果显示, 领属格在语料库中出现的频率很高, 分析原因有以下几点:

- (1) 领属格是基础易于掌握和运用的格位, 在句中表示领有和限定的关系, 因此在维吾尔语中的使用频率很高。

例如: “«džunçua çelq džumhuriyitiniñ adwokatlar qanuni» niñ 28-maddisi: dewalqquñiniñ hawalisini qobul qilip, keliftıryf, kesim iñıwa qatniñiñ” (《中华人民共和国律师法》第二十八条: 接受当事人的委托, 参加调解、仲裁活动)。

(1) 领属格具有多种属性意义, 此次分析的语料中不仅有领属格最基本的用法—物主领有, 还有其他用法, 比如主体领有和时空关系领有等。例如: “«džunçua çelq džumhuriyitiniñ adwokatlar qanuni» niñ 53-maddisi: 1980-jili 8-ajniñ 26-kyni 5-nöwetlik memliketlik çelq qurultiji daimij komitetiniñ 15-jikiniñ maqullanğan «džunçua çelq džumhuriyitiniñ adwokatlar waqitliq nizami» juniñ bilen teñ bikar qilinidu.” (《中华人民共和国律师法》第 53 条: 1980 年 8 月 26 日第五届全国人民代表大会常务委员会第十五次会议通过的《中华人民共和国律师暂行条例》同时废止)。

3.2.2. 向格附加成分的统计与分析

维吾尔语的向格主要表示行为动作的趋向[12], 还可以表示行为动作针对的对象, 事物达到或接近某一状态, 事物的价值等意义。名词的向格形式由名词后附加向格词尾构成。向格共有四个形式, 向格词尾的附加是根据词根与词缀之间的语音和谐规律来选择的, 库中向格附加成分的统计结果如表 3。

Table 3. Add component statistics table to the grid

表 3. 向格附加成分统计表

| 构形语素 | 出现次数 |
|------|--------|
| ka | 15,661 |
| ge | 10,506 |
| qa | 6163 |
| ke | 5446 |

结果显示, 在库中语料中向格的四种表现形式均有出现, 按照语音和谐的规律, 词干最后一个辅音若是清辅音则要与清辅音开头的词尾结合, 若是浊辅音则要与浊辅音开头的词尾结合, 元音结尾则要附加浊辅音开头的词尾, 同时还要注意元音的前后和谐规律。库中出现次数较多的是“ka”和“ge”形式的向格, 出现较少的则是“ke”和“qa”形式, “ka”“ge”两种形式的向格是附加在浊辅音和元音结尾的词语之后的, 如:

(2) «džunçua çelq džumhuriyitiniñ adwokatlar qanuni» niñ 14 - maddisi: muñu qanundiki belgilimilerge uñsun kelidigan adwokatlıri boluñ (《中华人民共和国律师法》第十四条: 有符合本法规定的律师)。

(3) «džunçua çelq džumhuriyitiniñ adwokatlar qanuni» niñ 28-maddisi: dewalqquñiniñ hawalisini qobul qilip, keliftıryf, kesim iñıwa qatniñiñ (《中华人民共和国律师法》第二十八条: 接受当事人的委托, 参加调解、仲裁活动)。

“ke”和“qa”形式则是附加在清辅音之后, 比如:

(1) «džunçua çelq džumhuriyitiniñ kepillik qanuni» niñ 76-maddisi: gory toxtami hoquq dekontlıri tapşurulğan kyndin bañlap kyñke ige bolidu (《中华人民共和国担保法》第七十六条: 质押合同自权利凭证交付之日起生效)。

(2) «džunçua çelq džumhuriyitiniñ kepillik qanuni» niñ 31 - maddisi: kepil kepillik džawabkarlıqini ystige albandin kejin, qerzdardin qerzni tölitiweliñqa hoquqluq (《中华人民共和国担保法》第三十一条: 保证人承担保证责任后, 有权向债务人追偿)。

根据此结果可以得出库中维吾尔语词语以浊辅音和元音结尾的数量会更多, 而且维吾尔语中的清辅音数量比浊辅音加元音的总数量少很多。所以各类附加成分中浊辅音开头的变体的出现频率比清辅音开头的变体的出现频率高很多。

“**ka**”形式的向格使用频率最高, 学生也可以最先学习该形式。

同时, 向格总体频率出现也很高, 这是由于向格具有多种意义, 不仅有其基本义表动作的趋向还有表示行为动作的目的、表示主体与客体的价值, 或与之等值的意义以及时间意义等多种含义, 因此使用范围广、次数多。

3.2.3. 宾格附加成分的统计与分析

名词的宾格表示主体与客体的关系, 通过在名词后附加“**ni**”来表示, 宾格并无其他变体形式, 附加了宾格词尾的词在句子中充当宾语, 宾格词尾后不再接受其他词尾[13]。库中也有许多附加宾格词尾的词语, 库中语料附加的宾格词尾统计结果如表 4。

Table 4. Statistical table of bingge additional components

表 4. 宾格附加成分统计表

| 构形语素 | 出现次数 |
|------|--------|
| ni | 45,218 |

由此可知, 宾格“**ni**”在库中一共出现了 45,218 次, 主要以客体宾格的形式出现如:

(1) «**džunxua xelq džumhuriyitiniñ kepillik qanuni**» niñ 36-maddisi: jeza (bazar), kent karxaniliriniñ jer ijlitiñ hoquqini ajrim renige qojufqa bolmajdu. (《中华人民共和国担保法》第三十六条: 乡(镇)、村企业的土地使用权不得单独抵押。)

(2) «**džunxua xelq džumhuriyitiniñ kepillik qanuni**» niñ 96-maddisi: alaqidar terepler zakalet toxtamida zakaletni tapjurf mudditini pytyjfyfi lazim. (《中华人民共和国担保法》第九十六条: 当事人在定金合同中应当约定交付定金的期限。)

宾格出现次数之所以这么高有以下几个原因:

(1) 法律条文中的宾语多是带有限定成分的, 这种情况下宾格无法省略, 因此最终的出现频率很高。例如:

«**džunxua xelq džumhuriyitiniñ kepillik qanuni**» niñ 75-maddisi: tōwendiki hoquqlarni goryge qojufqa bolidu” (《中华人民共和国担保法》第七十五条: 下列权利可以质押)此时的“**hoquqlar**”前面有限定词“**tōwendiki**”因此宾格不可省略。

(2) 法律条文中的维吾尔语动词多为及物动词可以支配宾格宾语。例如: “«**džunxua xelq džumhuriyitiniñ qanuni**» niñ 93 maddisi: fu deridžilik xelq hōkymitiniñ testiqi bilēn, xejr-saxawet tejkilatiniñ pul muamile isfotini syryfityryf” (《中华人民共和国慈善法》第九十三条: 经本级人民政府批准, 可以查询慈善组织的金融账户;)此时“**syryfityryf**”为及物动词, 支配“**xejr-saxawet tejkilatiniñ pul muamile isfotini**”。

3.2.4. 从格附加成分的统计与分析

名词的从格由名词附加从格词尾 **din/tin** 来表示, 附加了从格词尾的名词表示动作的起点, 经过的处所以及被比较的事物等, 在句中作状语。对于库中从格词尾的统计结果如表 5。

Table 5. Statistical table of additional components from grid

表 5. 从格附加成分统计表

| 构形语素 | 出现次数 |
|------|------|
| din | 7739 |
| tin | 3987 |

结果显示, 库中对于从格的两种形式均有出现, 附加在浊辅音和元音后面的“din”形式约是附加在清辅音后面的“tin”形式的将近两倍, 说明库中语料更多是以浊辅音或元音结尾的, 产生这个结果的原因还有一个, 即在维吾尔语名词后附加词缀是按照数+人称+格的形式书写的, 附加在人称后面的从格都是“din”形式, 例如: “«*džunçua xelq džumhuriyitiniñ statistika qanuni*» niñ 37-maddisi: statistika xadim-liridin aldinqi tarmaqniñ (3) tarmaqfisisidin (5) tarmaqfisisitfe bolkan tarmaqfilarida kersitilgen qilmiflarniñ birini sadir qilkanliri tyzitişke bujruvidu.” (《中华人民共和国统计法》第三十七条: 统计人员有前款第(三)项至第(五)项所列行为之一的, 责令改正。)

根据前面的统计结果, 领属性人称大约出现了将近六万次, 这也是造成“din”出现次数高于“tin”的原因。从格整体出现次数超过一万次, 也是出现次数很高的格位, 出现这个结果是因为从格具有多种意义, 不仅可以表示事物的起源, 还可以表示被比较事物、原因等意义。

3.2.5. 时位格附加成分统计与分析

时位格主要表示行为动作的发生, 存在的时间或空间, 也可以表示工具[12]。名词的时位格词尾共有4个变体, 它们是 da/te/ta/dε, 附加规则同向格一样遵循语音和谐规律。库中时位格统计结果如表6。

Table 6. Statistical table of additional components in time lattice
表 6. 时位格附加成分统计表

| 构形语素 | 出现次数 |
|------|------|
| da | 9224 |
| dε | 8831 |
| ta | 2836 |
| te | 1432 |

结果显示, 按照清浊和谐规律来看, 附加在浊辅音和元音后的“dε”和“da”数量远高于附加在清辅音后的“ta”和“te”数量, 按照前后元音的和谐规律来看, 附加在最后一个音节的元音是后元音上的“ta”和“da”比附加在最后一个音节的元音是前元音的“dε”和“te”多, 由此可见库中语料更多的是以后元音结尾或是最后一个音节的元音是后元音同时以浊辅音结尾的词语, 这个统计结果也与向格统计结果得出的结论一致。时位格具有多种意义也是形成它出现次数如此高的原因之一。

时位格可以表示时间范围, 例如: “«*džunçua xelq džumhuriyitiniñ adwokatlar qanuni*» niñ 12-maddisi: adwokatlar kesip bilen juwullanwanda rajon feklimisige ufrimajdu.” (《中华人民共和国律师法》第十二条: 律师执业不受地域限制。)此时“juwullanwanda”表示在从业这个时间范围内。

可以表示地点, 例如: “(1) «*džunçua xelq džumhuriyitiniñ adwokatlar qanuni*» niñ 44-maddisi: birla waqitta ikkidin artuq adwokatliq ornida kesip bilen juwullanwankanlar;” (《中华人民共和国律师法》第四十四条: (一)同时在两个以上律师事务所执业的;)此时“ornida”表示在这个单位。

还可以表示手段, 起到“工具格”的功能, 现代维吾尔语没有独立的“工具格”附加成分。例如: “«*şindzañ ujkur arptom rajoniniñ medžburijet maaripini jolka qojuf farisi*» niñ 4-maddisi: qof tilda oqutuf az sanliq milletler tili we xenzu tilidiki oquf-oqutuf endizisini kersitidu.” (《新疆维吾尔自治区义务教育实施办法》第四条: 双语教学是指使用少数民族语言和汉语言文字组织教育教学的教育教学模式。)

3.2.6. 其他次要格位统计与分析

维吾尔语共有十个格位, 分为六个主要格位和四个次要格位, 在前面的统计分析中我们列举了除主

格之外的其他五个主要格位, 接下来将分析语料库出现的范围格、界限格、形似格这三个次要格位。

名词结合词尾-diki/-tiki 就构成名词的范围格形式。范围格表示人或事物存在的范围或领域[12]。统计结果如表 7。

Table 7. Statistical table of additional components in scope grid

表 7. 范围格附加成分统计表

| 构形语素 | 出现次数 |
|------|------|
| diki | 8245 |
| tiki | 1129 |

名词的界限格由名词附加界限格词尾-giŋe/-qiŋe/-ɳiŋe/-kiŋe 来表示的。界限格表示行为状态所持续的时间界限和空间界限[12], 在句中作状语, 统计结果如表 8。

Table 8. Statistical table of additional components in boundary grid

表 8. 界限格附加成分统计表

| 构形语素 | 出现次数 |
|------|------|
| giŋe | 91 |
| qiŋe | 42 |
| ɳiŋe | 40 |
| kiŋe | 27 |

名词结合词尾-dek/-tek 就构成名词的形似格形式, 名词的形似格表示事物和事物之间在性质、形状、特征等方面具有某些共性[12]。统计结果如表 9。

Table 9. Statistical table of additional components similar to grids

表 9. 形似格附加成分统计表

| 构形语素 | 出现次数 |
|------|------|
| dek | 86 |
| tek | 69 |

结果显示, 无论是范围格还是界限格或者形似格, 均为附加在浊辅音和元音之后的格位形式出现次数多, 这也跟它们能够组合的浊辅音和元音的数量比清辅音多有关。

3.2.7. 格位统计与分析总结

如表 10 所示, 通过统计分析库中语料中出现的五个主要格位和三个次要格位, 不难看出维吾尔语主要格位的使用频率很高, 次要格位中范围格的使用频率也很高, 界限格和形似格相对较低, 无论是向格还是时位格范围格, 只要是存在变体形式的格位均是以浊辅音开头的构形语素使用频率高于以清辅音开头的构形语素, 这也再次说明了维吾尔语是一种黏着语, 许多词语的发音都是偏向于浊音。

并且“格”附加成分除了可以出现在名词之外, 也可以附加在形容词(形容词名词化)、代词、模拟词以及动名词、形动词等静词化的动词后面。所以, 名词的“格”语法范畴不只是属于名词的语法范畴, 其他词类也可以附加这些“格”附加成分。这也是“格”附加成分出现次数多的原因。

Table 10. Statistical table of additional components in grid
表 10. 格附加成分统计表

| 构形语素 | 格位名称 | 出现次数 |
|---------------------|------|--------|
| niŋ | 领属格 | 33,997 |
| gɛ/ke/qa/ka | 向格 | 37,776 |
| ni | 宾格 | 45,218 |
| din/tin | 从格 | 11,726 |
| da/ta/de/te | 时位格 | 22,323 |
| diki/tiki | 范围格 | 9374 |
| giŋɛ/qiŋɛ/kiŋɛ/kiŋɛ | 界限格 | 200 |

同时由于浊辅音开头的构形语素是附加在以元音结尾或是以浊辅音结尾的词根之后, 而元音与浊辅音的数量加起来高于清辅音的数量, 这也可以说明为什么附加浊辅音开头的构形语素数量更多。而像是宾格(ni)领属格(niŋ)虽然是只有一个形式的构形语素, 但是却是使用非常频繁的构形语素, 因此也需要重视和掌握它们的不同用法。

3.3. 数附加成分的统计与分析

维吾尔语的名词有数的变化, 叫做名词的数。名词有单数和复数两种不同的形式。名词后没有附加任何词尾的形式是名词的单数形式。名词的单数形式表示单个的人或事物。例如: “kitab” (表示一本书), 名词后附加 lar 或 ler 词尾是名词的复数形式, 表示两个或两个以上的事物, 例如: “kitablar” (表示两本及两本以上的书), 由于维吾尔语单数是零形式的原因, 此次语料中只统计了维吾尔语的复数形式, 统计结果如表 11。

Table 11. Statistical table of additional components
表 11. 数附加成分统计表

| 构形语素 | 出现次数 |
|------|--------|
| lar | 31,984 |
| ler | 16,612 |

结果显示, lar 的出现次数是 ler 的将近两倍, 由此可以分析出语料库中的词语更多的是最后一个音节是后元音的词语, 同时数附加成分的整体出现次数有将近五万, 说明使用频率极高, 我认为原因如下:

(1) 名词的数不仅可以附加在名词之后, 还可以附加在名词化的其他词类之后。例如: “«ŋindɜaŋ uŋɜur apton«dɜuŋɜua ɜelq dɜumhurijitiniŋ beqiweliŋ qanuni» niŋ 31-maddisi:balilarni beqiweliŋ nami bilen aldap satqanlar qanun bojiŋɛ dɜinaŋi dɜawabkarliqqa tartilidu.» (《中华人民共和国收养法》第三十一条: 以收养名义拐卖儿童的, 依法追究刑事责任。)

(2) 数附加成分不仅可以表示复数的概念, 还可以表示类的概念。

因此, 数附加成分的出现频率很高, 是学生学习必须要掌握的一个知识点。

4. 结语

本文统计了语料库中出现的领属性人称、格、数这三个附加在名词性词语后的构形语素的出现次数,

分析了这几种构形语素出现频率高低的原因, 根据结果可以看出, 附加在名词性词语后的构形语素中宾格(ni)领属格(nij)虽然只有一种形式但是出现频率极高, 同时向格、时位格、从格这些主要格位具有四种形式, 整体出现次数也很多, 领属性人称词尾中第三人称词尾(si/i)出现次数最多, 次要四格出现次数相较于主要六格明显降低, 这些结果为学生学习维吾尔语及老师教学提供了帮助, 同时也为计算机语言学的研究提供了支持。

维吾尔语是具有形态变化的语言, 这对于母语为汉语的第二语言学习者来说具有一定的困难, 因此教师在教学过程中可以优先讲授这些出现频率高的名词构形语素, 学生也要优先学习这些语缀, 因为他们日常中运用广泛, 需要重点理解和掌握以便于更好的深入学习维吾尔语语法。通过本研究发现, 虽然面向第二语言学习者的语言教学都有一些共同特点, 但是面向法律 + 维吾尔语双学位学生的语言教学和面向司法干部的语言教学也有一些不同于普通的第二语言教学的特点和规律, 比如: 领属性人称的第三人称的使用频率比日常维吾尔语和文学语言高很多。这也说明面向普通第二语言学习者和面向法律专业学生的第二语言教学还是有一定的差异, 教学人员应该根据教学对象的不同采取不同的教学方法和策略。

参考文献

- [1] 于淼. 国内语料库语言学发展研究[J]. 英语广场, 2018(9): 46-47.
- [2] 邓浩. 论维吾尔语构形语素的历史发展[J]. 喀什师范学院学报, 1988(1): 47-59.
- [3] 邓浩. 维吾尔语构形语素来源新探[J]. 中央民族学院学报, 1989(4): 86-91.
- [4] 李圃. 现代维吾尔语语素研究[D]: [硕士学位论文]. 乌鲁木齐: 新疆大学, 2002.
- [5] 亚热·艾拜都拉. 论维吾尔语构形语素的多义性——兼论粘着语和屈折语的共同点[J]. 乌鲁木齐成人教育学院学报, 1999(2): 25-27.
- [6] 方晓华. 论维吾尔语的语素[J]. 喀什师范学院学报, 1990(1): 87-92+60.
- [7] 霍盛. 关于维吾尔语附加成分的划分[J]. 新疆社科论坛, 1991(2): 88-95.
- [8] 热西旦. 汉文-维吾尔文双语语料库构建的实验性研究[D]: [硕士学位论文]. 乌鲁木齐: 新疆大学, 2007.
- [9] 吐尔洪·吾司曼, 杨雅婷, 艾孜孜·吐尔逊, 程力. 字符级的维吾尔语形态协同分析方法[J]. 北京大学学报(自然科学版), 2019, 55(1): 47-54.
- [10] 哈里旦木·阿布都克里木, 孙茂松, 刘洋, 阿布都克力木·阿布力孜. THUyMorph: 维吾尔语形态切分语料库[J]. 中文信息学报, 2018, 32(2): 81-86.
- [11] 阿依克孜·卡德尔, 开沙尔·卡德尔, 吐尔根·依布拉音. 面向自然语言信息处理的维吾尔语名词形态分析研究[J]. 中文信息学报, 2006(3): 43-48+98.
- [12] 早克热·卡德尔, 艾山·吾买尔, 吐尔根·依布拉音, 艾斯卡尔·艾木都拉. 维吾尔语名词构形词缀有限状态自动机的构造[J]. 中文信息学报, 2009, 23(6): 116-121.
- [13] 马维和, 李玉梅. 现代维吾尔语格的分类[J]. 石河子大学学报(哲学社会科学版), 2004(2): 8-10.