Modern Linguistics 现代语言学, 2024, 12(6), 1-6
Published Online June 2024 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2024.126426

Hans汉斯

喀喇汗王朝时期维吾尔族诗歌文本的语言特征

艾克拜尔·司马依力江

北方民族大学外国语学院，宁夏 银川

收稿日期：2024年4月11日；录用日期：2024年6月3日；发布日期：2024年6月7日

摘　要

本文探究喀喇汗王朝时期维吾尔族诗歌文本的语言特征，研究发现：1) 喀喇汗王朝时期维吾尔族诗歌文本的词频分布符合齐普夫定律，其文本内部语言机制具有“自组织性”；2) 喀喇汗王朝时期的维吾尔族诗歌的最高频词与汉语的最高频词呈现相似性，因此维吾尔语与汉语这两种语言在齐普夫分布规律上是相似的。本文所得数据及标注方法为未来研究处理维吾尔语文本提供了参考，也进一步验证了前人的部分研究成果，有助于维吾尔族诗歌研究的系统性、客观性和科学性。

关键词

喀喇汗王朝时期，维吾尔族诗歌，语言特征，齐普夫定律

Language Features of Uyghur Poetry Texts during the Kara-Khanid Khanate Period

Ai Kebaier Si Mayilijiang

School of Foreign Studies, North Minzu University, Yinchuan Ningxia

Received: Apr. 11th, 2024; accepted: Jun. 3rd, 2024; published: Jun. 7th, 2024

Abstract

This paper explores the language features of Uyghur poetry texts during the Kara-Khanid Khanate period. The study finds that: 1) The word frequency distribution of Uyghur poetry texts during the Kara-Khanid Khanate period conforms to Zipf's law, indicating a self-organizing mechanism within the text; 2) The highest frequency words in Uyghur poetry during the Kara-Khanid Khanate period exhibit similarities with the highest frequency words in Chinese, suggesting that Uyghur and Chinese share similarities in Zipfian distribution patterns. The data and annotation methods obtained in this paper provide a reference for future research on Uyghur language texts, and further verify some of the previous research results, which is conducive to the systematicness, objectivity and scientificity of Uyghur poetry research.

文章引用: 艾克拜尔·司马依力江. 喀喇汗王朝时期维吾尔族诗歌文本的语言特征[J]. 现代语言学, 2024, 12(6): 1-6. DOI: 10.12677/ml.2024.126426

Keywords

Kara-Khanid Khanate Period, Uyghur Poetry, Language Features, Zipf's Law

1. 引言

维吾尔族语言是中国少数民族语言之一，从古至今先后使用了多种文字，包括回鹘文、汉文、阿拉伯文、拉丁文等多种文字[1]，因此为后人留下了许多珍贵的维吾尔族文学作品。其中，在回鹘文文献中，文学作品很多，以诗歌作品居多，可以说诗歌创作是古代维吾尔族文学最重要的成就之一[2]。以喀什噶尔和巴拉沙衮为首府的喀喇汗国(893~1212 年)在维吾尔族文学史上更是开创了一个崭新的时期，尤其以诗人兼思想家优素甫•哈斯哈吉甫和语言大师麻赫穆德•喀什噶里二人最为著名，其作品《福乐智慧》和《突厥语辞典》更是优异的代表作，堪称喀喇汗王朝的双璧，可以说维吾尔古代文学自有了这两块丰碑之后，才真正迈上了康庄大道，趋于成熟[3]。丰碑之一的《突厥语大词典》是由作者麻赫穆德•喀什噶里完成的[4]。每个古代维吾尔语词汇用阿拉伯文释意，其中有各部族的民歌、诗章等[3]。

1935 年 Zipf 有关语言统计的著作的出版标志着一个新的语言学分支学科(计量语言学，Quantitative linguistics)和一种新的语言研究方法的诞生[5]。计量语言学以真实语言活动中产生的言语材料为研究对象，力求通过定量方法来探索语言的结构模式与演化规律，是一个用精确的方法来研究人类语言的语言学分支学科[6]。

因以往对维吾尔族诗歌的研究大都采用的是定性的方法且缺乏科学语言数据的支持，而现代语言学的研究较注重真实语言数据的考察。因此，运用计量的方法，基于科学的语言数据，以语言结构特征为切入点来探究维吾尔族诗歌的文体特征是有意义的。

2. 研究综述

先前对维吾尔族诗歌的研究是从文化，音乐等方面进行的。如库尔班•买吐尔迪(2007)在《突厥语大辞典》论及至今仍存在于维吾尔民众中的撒树枝习俗及文化现象，阐释了其原因：在人们的心目中，自己的生命、生活和命运都与该棵树有关联[7]；此外，王明科(2016)从音乐的角度对《喀什葛尔之夜》的序曲到中间的十一章再到尾歌进行了详细地描述，指明阿布都热依木•吾提库尔的《喀什葛尔之夜》属于维吾尔说唱艺术中的第六种形式，即维吾尔的“达斯坦”，体现了一种诗歌叙事的古典美[8]。

目前鲜有研究是从语言学的角度分析维吾尔族诗歌，大多还停留于从某个词汇的分析回归到诗歌本身。如海热提江•乌斯曼(2015)对“迪旺”释意并将其与其它词之间的联系做了详细的论述，因此产生了编纂诗歌“迪旺”的传统[9]。哈司依提•艾迪艾木(2017)对维吾尔族口头诗歌的程式句法进行论述并提到其为一种缩略化的表达方式，是种达斯坦传统[10]。现有对维吾尔族诗歌，特别是喀喇汗王朝时期维吾尔族诗歌的研究成果较少，大多集中在某一词汇或句子的解释与理解，缺乏科学语言数据支持。所以，运用计量方法考察维吾尔族诗歌文本的内在规律是有意义的。

在 1938 年，Yule 的研究成为了使用现代统计学方法进行语体计量学研究的真正开端。黄伟和刘海涛(2009)通过两个 50 万词的语料样本发现在现代汉语口语体和书面语体中具有显著分布差异的 16 个语言结

构特征并以名词、代词等7个语言结构特征作为文本特征，将21个文本聚类为口语体和书面语体两类[11]。所以，以语言结构特征为切入点，运用计量方法探究维吾尔族诗歌是可行的。国内外已有研究从定量的角度对诗歌与民歌展开了探索。

国内方面，张晓瑾和刘海涛(2017)得出“花儿”文本的秩频分布符合齐普夫定律且呈现出了民歌文本的“语言自组织性”[12]。刘海涛和潘夏星(2015)在现代汉语新诗的计量特征研究中，新诗文本的“自然性”利用“齐普夫定律”得以验证[13]。

国外方面，在对罗马尼亚语民歌的研究中得出在数据不太复杂的情况下，可使用 Zipf 的幂律或Zipf-Alekseev 函数(Popescu *et al.*, 2010) [14]。另外，Popescu 和 Altmann 等研究者对 54 篇斯洛伐克诗人 Eva Bachletov 的诗展开计量研究并得出“基尼系数越大，文本的词汇丰富度越低”[15] [16]。

上述国内外研究证明了齐普夫定律作为指标可用来探究诗歌的语言特征，也证明运用计量的方法研究诗歌的可行性与科学性。然而以少数民族语言，对于诗歌为研究对象展开的计量研究数量匮乏。因此，本文对喀喇汗王朝时期维吾尔族诗歌文本的语言特征深入探究，旨在回答以下两个问题：

1) 喀喇汗王朝时期维吾尔族诗歌文本是否符合齐普夫定律？

2) 维吾尔语与汉语这两种语言在齐普夫分布规律上是否呈现相似性？

3. 语料与方法

本文使用的语料来源于《突厥语大词典(卷三)》[17]。它是以阿拉伯文诠释当时的突厥语(主要是维吾尔语)的辞典，成书于 1072~1074 年。作者麻赫穆德喀什噶里生于喀什噶尔乌帕尔乡阿兹克村(属今新疆喀什地区疏附县)，书中的题材具有多样性，因而为研究古代突厥语部族的历史地理、社会、文化和民俗等提供了丰富的资料，被称为 11 世纪各突厥部落社会生活的大百科全书[3]。因此，本文从《突厥语大词典(卷三)》中随机抽取了 15 首，总词数为 1144 词。随后，本文统计喀喇汗王朝时期维吾尔族诗歌文本的词频分布，并用阿尔特曼拟合器检验维吾尔族诗歌的文本是否符合齐普夫定律，并探究维吾尔语与汉语在齐普夫定律上是否具有相似性。

4. 结果与讨论

齐普夫定律是关于文本中词频分布的定律，即按词语出现次数的多少从大到小进行排序的分布[6]。齐普夫定律要求词出现的频数与其频数秩(序号)之间具有反比例关系。1928 年，Paul Menzerath 在研究词和音节的长度关系后得出一个词所含音节数的增加，这些音节的平均长度会减小并将这一定律称为 Menzerath Altmann 定律，公式经过推导后为 $y = ax^{-b}$，变量 y 为词语的频率，x 为词语序列，a 和 b 为两个参数[18]。

表 1 为 15 首维吾尔族诗歌样本的频次和频率，图 1 为喀喇汗王朝时期维吾尔族诗歌的齐普夫曲线。表 1 及图 1 显示：按照递减顺序排列的喀喇汗王朝时期维吾尔族诗歌中出现的词的频次和频率符合齐普夫定律。数据与某一个模型或者定律相吻合的程度可用拟合优度系数 R^2 来衡量。图 1 显示的拟合优度系数 $R^2 = 0.9501$，说明其文本的词频分布符合齐普夫定律；同时，其频率最高的 15 个词中有 10 个是单音节词，5 个为双音节词，累计频率为 0.1923，占比近 20%，可知其符合齐普夫定律的省力原则。刘海涛(2017a: 49-50)对兰卡斯特汉语语料库的词频进行统计发现词频最高的“的”的频率为 0.0615，小于 0.1;英语中词频最高的 the 的频率为 0.071；而本文中词频最高的“دى，تى”(了)的频率约为 0.06，占比为 6%，都小于 0.1。因此，在词频分布规律上，喀喇汗王朝时期维吾尔族语言与其他语言可能并无太大差异。

此外，根据《维汉大词典》和《现代维吾尔语参考语法》，دىو تى，为过去式语缀，其意义相当于汉语的“了”字，如قوشۇلدى (同意了)，بۇزۇلدى (坏了)；نى为宾格语缀，缀加在名词类及其短语末尾，表示该名词性成分是动作的客体或行为动作所涉及的对象，因此要求句中出现宾格短语的动词一般都是及

物动词，如كۆرسەن كىنونى (你去看电影吧)；ئۇ指向离说话者相对远的人或事物，ئۇ又可作第三人称代词“他/她/它”，可与格和后置合并，当代词词干与格或后置词合并时两者之间出现领属格نىڭ，如ئۇ+دا (位格) =ئۇنىڭدا (在他/她/它那里)，ئۇچۈن+ئۇ (为了) =ئۇچۈن ئۇنىڭ (为了他/她/它等)，它的指示代词功能更多地体现在修饰其他成分上，如ئۇمەسىلە (那个问题) 与ئۇ ئادەم (那个人)等；بىلەن在词与词，词组与词组，句子与句子之间充当连接成分，在汉语中相当于“跟、和、同、与等”，如ئۇ بىلەن مەن (他和我)；ئادەم为名词，相当于汉语中的“人”，如ئۇ ئادەم (那个人)；تۇتۇپ为动词，相当与汉语中的“抓”；لار为复数语缀，它不仅可与名词相结合表示复数，如ئىشچىلار (工人们)，还能跟形容词，代词，代词，量词，名词化，形容词化短语其他名词性语类合并，所以有时表达的意义不仅是复数，还要依具体情况而定，如سولار (各种渠道的水)，而不是水的复数；نىڭ为领属格，表示人或事物的领属关系以及各种关系，用做定语或修饰语。它与领属者合并后要求后面的从属者成分要带上与领属者的人称和数相一致的从属语缀，从而两者构成严紧的领属一从属结构，如منىڭ قولوم (我的手)，سىنىڭ قولوڭ (你的手)，ئۇنىڭ قولى (他的手)；دۈشمەن为名词，意为敌人；تۆۋەن为方位名词，意为下面；گە为向格，缀加在名词类及其短语末尾，表示动词所表达的动作的目的、去向、指向等，因此有些趋向性动词要求句中与向格短语一起出现，如گەكىرۈى (进屋子)，有时表示行为动作的媒介或价值如ادەمگەتولدى (充满了人)；راسا为程度副词，意为“不一般地，狠狠地”；ئىدى，ئى该系词来自古代突厥语系词ر“是”经过历时音变，目前以ئى的形式固定下来，其意义保持不变。在现代维吾尔语里ى与过去时成分دى和相应的人称成分合并，表示对过去的判断，如ئىدىمئالدىراشمەن (我当时很忙)，ئىدىڭلىشچىسەن (你曾经是工人)等；بىر为数词，意为“一”，如بىر پىيالە چاي (一杯茶)。本文的发现与刘海涛(2017a:49-50)统计的汉语词频前 15 个中的 7 个是一致的，它们分别是“的”“了”“是”“一”“和”“人”“他”，这在一定程度上说明维吾尔语与汉语在某方面是相似的。

Table 1. Frequency and frequency of 15 Uighur poetry samples
表 1. 15 首维吾尔族诗歌样本的频次和频率

频序	字词	频次	频率
1	دى	36	0.0315
2	نى	32	0.0280
3	تى	28	0.0245
4	ئادەم	16	0.0140
5	بىلەن	16	0.0140
6	تۇتۇپ	15	0.0131
7	نىڭ	12	0.0105
8	دۈشمەن	12	0.0105
9	تۆۋەن	10	0.0087
10	لار	9	0.0079
11	گە	8	0.0070
12	راسا	8	0.0070
13	ئۇ	6	0.0052
14	ئىدى	6	0.0052
15	بىر	6	0.0052

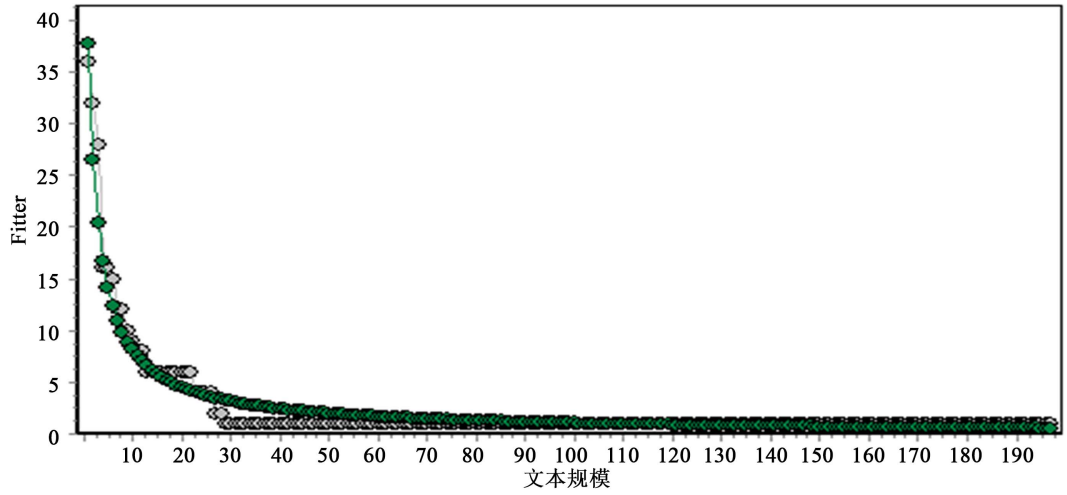

Figure 1. Zipf curve of Uyghur poetry during Kara-Khanid Khanate period
图 1. 喀喇汗王朝时期维吾尔族诗歌的齐普夫曲线

拟合优度系数和高频词都表明喀喇汗王朝时期维吾尔族诗歌文本的词频分布符合齐普夫定律，说明其文本的内部语言机制具有“自组织性”；且喀喇汗王朝时期的维吾尔族诗歌的最高频词与汉语的最高频词呈现相似性，这在一定程度上说明了维吾尔语与汉语这两种语言在齐普夫分布规律上是相似的。维吾尔族诗歌是维吾尔族人民口口相传的文化作品，传唱者大多为普通的大众，维吾尔族诗歌之所以受到普通劳动人民的喜爱，代代相传，其原因就在于维吾尔族诗歌文本简单易懂。因此，不追求锤炼造字的维吾尔族诗歌的文本自然就呈现出一种“自然”状态。本文结果已显示喀喇汗王朝时期维吾尔族诗歌的文本分布符合齐普夫定律，可为后续探究维吾尔语文本的规律性语言特征提供参考。

5. 结语

本文在计量语言学理论框架下引入计量研究方法对喀喇汗王朝时期的维吾尔族诗歌的文本计量特征进行了系统探索，可以回答本文提出的问题。本文从词频分布特征角度验证了喀喇汗王朝时期维吾尔族诗歌的文本词频分布高度符合齐普夫定律，其最高频词与汉语的最高频词及与汉语在齐普夫定律上呈现相似性。可见，同其他少数民族语言一样，用维吾尔语书写的维吾尔族诗歌文本符合人类语言的内部机制规律。

参考文献

[1] 阿不都热扎克·沙依木. 关于加强维吾尔文字专门研究的思考[J]. 西域研究, 2006(4): 105-107.
[2] 张巧云. 回鹘诗歌对回鹘文佛经偈颂的诗化影响[J]. 民族文学研究, 2016, 34(3): 115-123.
[3] 苗文军. 简论喀喇汗王朝时期的维吾尔族诗歌[J]. 兰州大学学报, 1998(4): 133-139.
[4] 牛汝极. 《突厥语大辞典》写本的流传[J]. 北方民族大学学报(哲学社会科学版), 2009(3): 27-30.
[5] 刘海涛. 依存语法的理论与实践[M]. 北京: 科学出版社, 2009.
[6] 刘海涛. 计量语言学导论[M]. 北京: 商务印书馆, 2017.
[7] 库尔班·买吐尔迪. 从《突厥语大辞典》中的“擀树枝”看维吾尔族的擀树枝习俗[J]. 民族文学研究, 2007(1): 52-54.
[8] 王明科. 《喀什葛尔之夜》的古典叙事诗美[J]. 中华文化论坛, 2016(12): 78-83.
[9] 海热提江·乌斯曼. 维吾尔诗歌“迪旺”述论[J]. 民族文学研究, 2015(6): 115-121.
[10] 哈司依提·艾迪艾木. 维吾尔口头爱情达斯坦的程式句法——以《艾力甫与赛乃姆》为例[J]. 中国韵文学刊, 2017,

31(2): 71-75+104.

[11] 黄伟, 刘海涛. 汉语语体的计量特征在文本聚类中的应用[J]. 计算机工程与应用, 2009, 45(29): 25-27+33.

[12] 张晓瑾, 刘海涛. 中国民歌“花儿”的计量特征[J]. 宁夏大学学报(人文社会科学版), 2017, 39(5): 76-80+91.

[13] 刘海涛, 潘夏星. 汉语新诗的计量特征[J]. 山西大学学报(哲学社会科学版), 2015, 38(2): 40-47.

[14] Popescu, I.-I., Mačutek, J. and Altmann, G. (2010) Word Forms, Style and Typology. *Glottotheory*, **3**, 89-96. https://doi.org/10.1515/glot-2010-0006

[15] Popescu, I.I., Čech, R. and Altmann, G. (2011) Vocabulary Richness in Slovak Poetry. *Glottometrics*, **22**, 62-72.

[16] Popescu, I. and Altmann, G. (2013) Descriptivity in Slovak lYrics. *Glottotheory*, **4**, 92-104. https://doi.org/10.1524/glot.2013.0007

[17] 新疆社科院. 突厥语词典[M]. 乌鲁木齐: 新疆人民出版社, 1984.

[18] 刘海涛, 黄伟. 计量语言学的现状、理论与方法[J]. 浙江大学学报(人文社会科学版), 2012, 42(2): 178-192.