

# 吴歌的计量特征研究

丁 婧

无锡科技职业学院文化旅游学院, 江苏 无锡

收稿日期: 2025年11月3日; 录用日期: 2025年11月24日; 发布日期: 2025年12月5日

## 摘 要

本文以吴歌为研究对象, 旨在揭示其语言系统中的统计规律与诗体生成机制。研究基于计量语言学的分析思路, 运用Zipf定律、Busemann系数与节奏建模等方法, 从词频分布、词类比例、节奏特征及语体聚类等维度, 对吴歌的语言结构进行系统考察。研究发现, 吴歌呈现开放而冗余的语言体系, 且有叙事主动性与抒情描写性并存的双重话语倾向, 其依附语气与声调的口语化节奏模式, 有别于古典诗与新诗的方言诗体系统。因此, 吴歌的诗性源于方言语音、语义与情感表达的自组织过程, 其语言特征为理解汉语诗体的地域分化与口头生成机制提供了新的实证依据。

## 关键词

吴歌, 计量语言学, Zipf定律, Busemann系数, 节奏结构, 方言诗体, 语体聚类

# A Study of Quantitative Properties of Wu Songs

Jing Ding

School of Culture and Tourism, Wuxi Vocational College of Science and Technology, Wuxi Jiangsu

Received: November 3, 2025; accepted: November 24, 2025; published: December 5, 2025

## Abstract

This paper is devoted to Wu Songs (Wuge) and aims to uncover the statistical regularities and generative mechanisms underlying their linguistic and poetic systems. Within the framework of quantitative linguistics, Zipf's law, Busemann's coefficient, and rhythm modeling were applied to systematically examine Wu Songs in terms of word-frequency distribution, part-of-speech proportion, rhythmic structure, and stylistic clustering. Wu Songs are found to embody an open and redundant linguistic system that exhibits a dual discourse tendency combining narrativity and lyricism. Their

rhythm, shaped by tonal intonation and affective cadence, forms a speech-like prosodic pattern distinct from both classical and modern poetic forms. The poetic nature of Wu Songs is interpreted as a self-organizing process emerging from the interaction among phonological, semantic, and emotional structures within the Wu dialect, providing new empirical evidence for the regional diversification and oral formation of Chinese poetic forms.

## Keywords

Wu Songs, Quantitative Linguistics, Zipf's Law, Busemann Coefficient, Rhythm Structure, Dialectal Poetics, Genre Clustering

Copyright © 2025 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

吴歌作为中国南方最具代表性的民间抒情歌体之一，自明清以来广泛流传于苏浙一带。其语言柔婉、节奏明快、叙事性强，以吴语方言为载体，兼具口语性与诗意性，既承续了古代民歌的“言志”传统，又孕育了江南文化中独特的“温柔敦厚”审美气质。吴歌通常包括“歌”与“谣”两部分，既有劳动号子、船歌、山歌，也有情歌、顺口溜等，体现出其题材的生活性与形式的多样性[1] [2]。

从历史源流看，吴歌的文化渊源可追溯至先秦至两汉时期的“吴歈”“越吟”等传统。《吴都赋》中“吴愉、越吟”之语，已可视为吴歌早期形态的文献显现[3] [4]。此后，南朝至唐宋时期的“吴声歌曲”在文献中频繁出现，吴歌在江南水乡的生产、生活与祭祀环境中逐渐形成独特的文化形态[5]。

近代以来，顾颉刚编纂《吴歌甲集》，以苏州一带歌谣为核心，系统搜集、校勘并分类整理吴歌文本，被普遍视为“独立的吴语文学的第一部”，在学术上确立了吴歌作为方言文学与民间诗体的重要地位[6]。刘半农在序文中强调吴歌在语言、风土与艺术上的整体性，指出其以方言为诗歌媒介的独特价值，为后来的语言学研究奠定了问题意识。此后各版影印与点校的出版，使吴歌文本由零散记录走向体系化，为基于语料的定量研究提供了重要条件。

围绕吴歌与“吴声歌曲”传统的关系，学界多从文学史与乐府学角度展开溯源。吴大顺在《魏晋南北朝乐府歌辞研究》中系统考察《乐府诗集》中的《吴声歌》类作品，指出东晋以来“吴声歌曲”在江东城市文化与民间情歌中的重要地位，为理解吴歌的历史结构提供了宏观背景[5]。王焰安则在《从〈吴声歌〉、〈西曲歌〉看南朝民间情歌》一文中揭示，南朝情歌以江南都市生活与恋爱情绪为核心，其叙事风格与后世吴歌在题材与情感结构上呈现显著的延续性[7]。可见，吴歌并非孤立文本，而是深嵌于自“吴声歌曲”以来的江南歌谣传统之中。

进入现代以来，吴歌研究逐渐被纳入民俗学与区域文化框架。陈书录指出，吴歌作为“诗化的江南民间文化”，以日常性与审美性相结合的方式展现江南民众“以歌抒情、以言成诗”的文化心理[2]。中国非物质文化遗产研究从曲牌类型、表演场域与传承谱系等方面系统梳理吴歌，确立其兼具历史深度与现实生命力的活态价值[8]；朱银华从区域互动角度揭示吴歌在更大空间范围内的流变与融合[9]。与此同时，学界也从“方言文学”与“口头传统”视角重估其语言价值。《吴歌甲集》及相关研究指出，吴歌在音系上充分利用吴语声调与入声，词汇上保留大量方言与生活化表达，句法上呈短句、重叠与衬词等口

语特征,使其成为观察吴语叙述与修辞的重要窗口[6][10]。此外,有学者从“吴歌与苏白”“吴歌与弹词”的互动入手,揭示其对苏州话书面化与戏曲化的影响,认为吴歌在一定程度上构成“吴语书写”的早期实验场[5][6]。总体而言,现有研究奠定了吴歌的文本与文化基础,但仍以质性描述为主,对整体语料的词频、句式与节奏规律缺乏系统的量化分析。

与吴歌研究的质性传统形成鲜明对照的是,计量语言学在汉语诗歌、民歌与语体研究中的迅速发展。刘海涛在《计量语言学导论》中系统介绍了 Zipf 定律、复杂网络与文本聚类等方法,为汉语文本的统计分析提供了理论框架[11]。黄伟、刘海涛通过语体聚类验证不同文本在词类分布上的显著差异[12];张晓瑾、刘海涛对“中国民歌‘花儿’”的研究证明“花儿”文本的词频分布符合 Zipf 定律,并借 Busemann 系数揭示其“主动性话语姿态”,为民歌计量研究提供了范式[13];刘海涛、潘夏星利用 Zipf-Alekseev 模型刻画汉语新诗的节奏分布,论证新诗在整体上呈现接近自然语言的概率结构[14]。此外, Pan 与 Liu 的研究表明,终韵结构与词类分布可作为区分古典诗与新诗的重要指标[15]。

国际研究亦提供了重要方法论启示。Popescu 与 Altmann 等人发展了 Zipf-Alekseev 与 Popescu-Altmann 模型,对多语种诗歌的词频、节奏及“主动性/描写性”特征进行函数拟合,展示了如何借助幂律分布与 Q 值等指标揭示诗歌风格差异[16]-[18]。Fang 对当代汉语计量语言学的综述指出,大规模语料与多模型拟合已成为国际趋势,且相关方法正广泛应用于诗歌与民歌研究[19]。艾克拜尔·司马依力江对喀喇汗王朝维吾尔族诗歌的语言特征分析,直接套用了“花儿”与新诗研究的技术路径,进一步验证了诗歌文本遵循 Zipf 定律的跨语言稳定性[20]。

然而,迄今为止,尚无将计量语言学方法系统应用于吴歌语料的研究。吴歌语言兼具方言性、口语性与强节奏性,其短句、重叠与助词构成使之在词频与句法分布上极具统计分析潜力。本文在前人研究基础上,综合运用 Zipf 定律、Busemann 系数与 Zipf-Alekseev 模型等方法,对吴歌语料的词频结构、韵脚词类与节奏分布进行系统分析,并结合聚类与共现网络揭示其语言自组织机制与方言诗体特征。

## 2. 吴歌的词频与 Zipf 规律

吴歌作为方言抒情歌体,其语言组织是否遵循自然语言的一般规律,是判断其口语性与诗体化程度的重要切入点。为此,本文以《吴歌甲集》为语料基础,对全书一百首吴歌进行分词与词频统计,总计词次  $N=5440$ ,词形数  $V=2917$ 。由此构建词频-序位分布,并在双对数坐标下进行 Zipf 定律的拟合分析。Zipf 定律认为,词频  $f(r)$  与其序位  $r$  之间存在幂律关系:

$$f(r) = cr^{-a},$$

其中指数  $a$  反映词汇分布的集中程度,常被用于衡量文本在“自然语言”与“文学语言”之间的分布差异。当  $a \approx 1$  时,文本表现出接近口语的自然分布;当  $a > 1$  时,则说明词汇使用更集中,具有较高的诗性密度或主题聚焦倾向。

根据对《吴歌甲集》的统计结果,词频-序位分布在双对数坐标中呈现出良好的线性趋势,其拟合优度  $R^2=0.9839$ , Zipf 指数  $a=0.6524$ ,截距  $\log_{10}(c)=2.0840$ ,见图 1。这一结果表明,吴歌整体词汇分布符合 Zipf 定律的幂律特征,说明其语言系统具备自组织结构,且具有较强的自然语言特征。与传统汉语诗体(如五言、七言诗)中普遍  $a > 1$  的集中分布不同,吴歌的指数明显偏低,反映其词汇扩散度高、语汇丰富,体现了口头歌体的开放性与语言流动性。

需要注意的是,吴歌作为口头诗体,其实际演唱中可能存在停顿、拖长音与重复唱段,而这些特征在文本转写中往往被省略,因此词频结构所呈现的 Zipf 分布主要反映的是转写文本层面的语言组织,而非完整的口头生成机制。

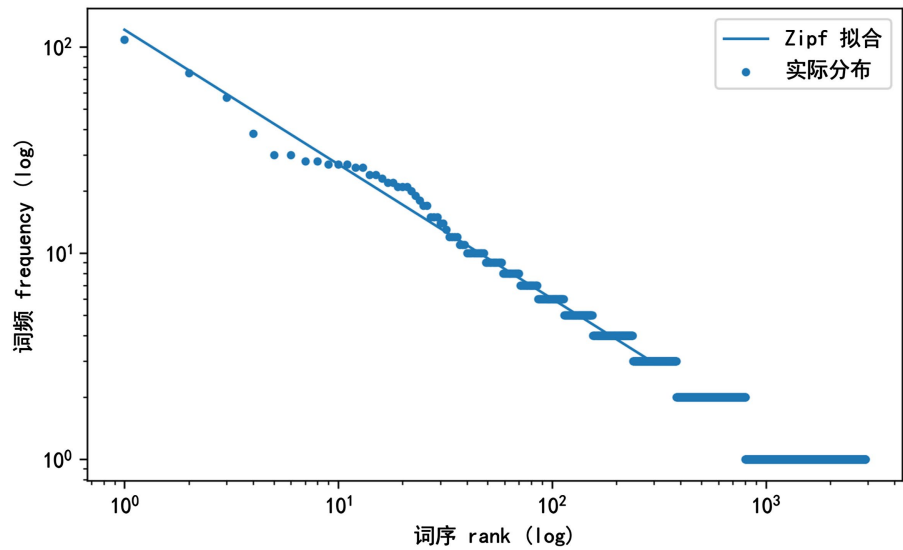


Figure 1. Log-log plot of word frequency distribution and Zipf fitting of Wu Songs  
图 1. 吴歌语料词频分布及 Zipf 拟合图

在高频词统计中，吴歌的前列词项以代词、助词与语气词为主，见表 1，如“你”（109 次）、“我”（75 次）、“啥”（57 次）、“个”（38 次）、“呀”（30 次）、“来”（30 次）等。此类高频词共同构成了吴歌的“口语框架”，表现出强烈的对话与互动特征，突显出其抒情主体与听话对象之间的互文关系。高频词的累计覆盖率达 15.7%，显示少量词汇在整体语料中承担了高比例的表达功能。这种高集中度并非书面诗歌中典型的韵律约束结果，而是源于口头交流语境下的程式性表达，是“方言诗体”的典型特征。尤其是“你 - 我”二元结构的反复出现，构成了吴歌抒情结构的基础，折射出其以人际互动为中心的审美取向。

Table 1. High-frequency words and Zipf model parameters in Wu Songs  
表 1. 吴歌高频词统计及 Zipf 模型拟合参数

词形	频次	相对频率	累计频率
你	109	0.020036765	0.020036765
我	75	0.013786765	0.033823529
啥	57	0.010477941	0.044301471
个	38	0.006985294	0.051286765
呀	30	0.005514706	0.056801471
来	30	0.005514706	0.062316176
吃	28	0.005147059	0.067463235
到	28	0.005147059	0.072610294
了	27	0.004963235	0.077573529
哉	27	0.004963235	0.082536765
做	27	0.004963235	0.0875
在	26	0.004779412	0.092279412
是	26	0.004779412	0.097058824

续表

哥哥	24	0.004411765	0.101470588
的	24	0.004411765	0.105882353
里	23	0.004227941	0.110110294
要	22	0.004044118	0.114154412
去	22	0.004044118	0.118198529
有	21	0.003860294	0.122058824
哭	21	0.003860294	0.125919118
郎	21	0.003860294	0.129779412
着	20	0.003676471	0.133455882
好	19	0.003492647	0.136948529
把	18	0.003308824	0.140257353
上	17	0.003125	0.143382353
人	17	0.003125	0.146507353
小	15	0.002757353	0.149264706
不	15	0.002757353	0.152022059
奴	15	0.002757353	0.154779412
叫	14	0.002573529	0.157352941

此外，吴歌的词频分布在中低频区间出现较长“尾部”，即大量低频词的存在。这一“长尾”现象意味着吴歌在固定表达之外保留了相当的创造性空间，使方言叙述与生活语汇能够灵活进入诗歌表达体系，从而兼具稳定与开放的语言特征。这与江南口头传统中“随口成歌”的创作机制密切相关，也与吴语方言自身的语义黏着性、助词丰富性相呼应。因而，从统计视角看，吴歌语言既符合自然语言的 Zipf 分布，又通过局部高频词的反复实现情感凝聚，体现出一种“自然与程式共存”的复合结构。

不难看出，吴歌语料的 Zipf 分布结果揭示其语言系统在统计层面既具有自然口语的自组织性，又通过核心词汇的高频聚焦，形成抒情诗体的程式化倾向。由此，吴歌在语言结构上介于自然口语与文学诗体之间，体现出民间抒情歌体的过渡性特征，为下文从词类分布与节奏层面探讨其“主动性话语姿态”提供了定量基础。

3. 词类分布与主动性分析

在诗歌与民歌的计量语言学研究中，词类分布不仅反映语体风格特征，也揭示了文本的“话语姿态”——即文本是更偏向动作、行为与交流的“主动性”，还是更偏向描写、客体与静态意象的“描写性” [21]。Busemann 系数(Busemann’s coefficient,  $Q$ )正是衡量这一倾向的重要参数，其定义为：

$$Q = \frac{N_{verb} + N_{pron}}{N_{noun} + N_{adj}}$$

其中，分子表示与主观、动态表达相关的词类(动词与代词)，分母代表与客观描写与名物指称相关的词类(名词与形容词)。当  $Q > 1$  时，文本具有较高的主动性，体现行为、情感与互动导向；当  $Q < 1$  时，则偏向描写性，强调物象、环境与静态场景[22]。

对《吴歌甲集》语料的词性统计显示：在总计 5868 个词次中，动词 1493 个，占 25.4%；代词 351 个，占 6.0%；名词 2172 个，占 37.0%；形容词 264 个，占 4.5%；其他词类 1588 个，占 27.1%。由此计算得 Busemann 系数  $Q = 0.7570$ ，见表 2。该系数低于 1，表明吴歌整体上更偏向“描写性话语”，即以具象名物和静态意象为主要构成。然而，与汉语古典诗歌(一般  $Q \approx 0.5$ )相比，吴歌的  $Q$  值明显更高，反映出其口语歌体在保留物象描写的同时，强化了交互与叙事成分，体现出“抒情中的动作化倾向”。

**Table 2.** Distribution of major part-of-speech categories and Busemann coefficient in Wu Songs  
**表 2.** 吴歌主要词类分布与 Busemann 系数

词类	词数	比例
代词(PRON)	351	6.0%
动词(VERB)	1493	25.4%
名词(NOUN)	2172	37.0%
形容词(ADJ)	264	4.5%
其他(OTH)	1588	27.1%
Busemann 系数 $Q$	-	0.7570

从词类比例上看，吴歌的动词与代词总占比达 31.4%，其中动词种类繁多，覆盖“来、去、做、吃、哭、唱、笑”等日常生活行为，显示其语言根植于生活经验的动作场景；代词以“你”“我”为核心，构成了频繁的对话结构，与第二节高频词分布结果一致，体现出吴歌强烈的言语指向性与人际互动特征。这种以“你-我”互动为核心的表达机制，使吴歌在语义上呈现出典型的“对话化抒情”格局，而非单向的叙述与描写。

相对而言，名词与形容词在吴歌中占 41.5%，主要承担描写自然物象、身体部位与日常事物的功能，如“花”“水”“郎”“奴”“门口”“河边”等，形成浓厚的地域生活氛围。这一描写性成分的比重，使吴歌在语言层面保持了与物质世界的紧密联系，强化了江南民歌的“生活诗性”特征。由此可见，吴歌的 Busemann 系数虽不高，却体现出“生活化的主动性”——它并非抽象思辨式的自我言说，而是通过身体动作与人际互动实现情感传递的“在场性”表达。

概言之，吴歌的词类分布揭示出一种兼具叙事性与描写性的平衡状态，其主动性成分使语言充满动势与交流意图，而描写性成分则维系了地方物象的感性氛围。正是在这种平衡之中，吴歌的语言系统实现了方言歌体的诗化转化，使口语话语在生活语境中自然生成诗性。

4. 节奏与行长分布

节奏结构是诗歌形式组织的重要维度，也是判断其语言节律与口语化程度的关键指标。对于民歌这类口头抒情文本而言，节奏不仅受语言韵律的约束，更与表演、吟唱方式密切相关。通过统计诗行长度(即每行所含汉字数)并进行 Zipf-Alekseev 模型拟合，可揭示吴歌在节奏层面的组织规律与自组织特征。该模型以幂律形式描述行长  $x$  与其出现频率  $f(x)$  的关系：

$$f(x) = k(x+b)^{-a},$$

其中，指数  $a$  反映节奏集中度， $b$  为平移参数， $k$  为比例常数。当  $a$  值较大时，诗行长度分布趋于集中，节奏呈规则化特征；当  $a$  较小或接近零时，则表示行长分布更为分散，体现口语歌体的自由韵律特征。

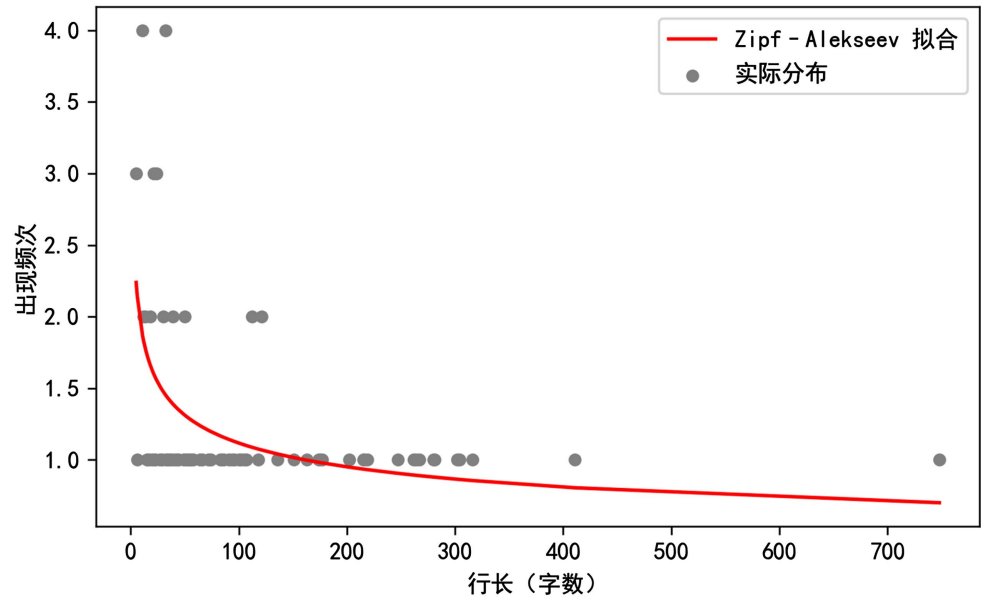
对《吴歌甲集》的 95 首吴歌进行逐行统计后，共获得 95 个诗行，平均行长为 93.91 字，标准差高达



111.57 字，见表 3。行长分布的波动极大，表明吴歌在结构上不存在严格的行句定型，而是随表演语气与叙事节奏自由展开。Zipf-Alekseev 模型拟合结果为  $a = 0.2321$ ， $b = 0.0000$ ， $k = 3.25$ ，其指数  $a$  明显低于 1，显示出吴歌行长分布的高度平缓性和自由性。这种“近似随机”的节奏分布意味着吴歌在节奏构型上几乎完全脱离书面诗体的定格模式，而保持了民间口语吟唱的自然波动，见图 2。事实上，由于行长统计基于现存文本形式，而吴歌在口头演唱中的节奏组织高度依赖声调、停顿与旋律性拖长，其实际节奏复杂度可能高于文本分析所呈现的结果。

**Table 3.** Rhythm statistics of Wu Songs (mean, standard deviation, and model parameters)  
**表 3.** 吴歌节奏统计参数(行长均值、标准差与模型系数)

指标	数值	说明
总诗行数	95	语料总行数
平均行长	93.91	每行平均字数
行长标准差	111.57	行长波动性指标
幂律指数 $a$	0.2321	节奏集中度系数
平移参数 $b$	0.0000	节奏分布平移项
比例常数 $k$	3.25	模型常数



**Figure 2.** Line-length distribution and Zipf-Alekseev fitting of Wu Songs  
**图 2.** 吴歌行长分布及 Zipf-Alekseev 拟合图

从统计图形上看，吴歌行长分布在短句区间集中度较高，部分诗行呈现极长的扩展结构，形成显著的“长尾”现象。这种分布形态揭示了吴歌在节奏建构上的两层特征：其一，在宏观层面，节奏分布符合自然语言的幂律衰减规律，即行句长度越长，出现概率越低；其二，在微观层面，个别长行的存在强化了叙事与情感表达的张力，使得吴歌在节奏上兼具“口语性”与“表现性”。这种以短句为基调、长行为张力点的结构模式，与吴语方言中丰富的助词体系及句法黏着特征密切相关，也体现出吴歌的语言节奏与语调系统之间的内在一致性。

进一步地，从诗行长度的统计特征看，吴歌平均行长显著高于传统五言、七言诗的定型结构，而标准差的巨大波动(111.57 字)则显示出节奏组织的高度变异性。这一变异性在口头创作语境中具有功能意义。它并非形式松散的表现，而是与吴歌在歌唱与表演中“即兴句法”的语义节奏对应。换言之，吴歌的节奏并非音步式的格律，而是一种“言语节奏”(speech rhythm)，通过自然语流的断续、重复与延宕形成情感节拍。这种节奏形式恰与江南方言的音调系统相契合，使得吴歌在听觉上具备强烈的语气感与互动感。

因此，从 Zipf-Alekseev 模型的参数分布可见，吴歌在节奏维度上呈现出典型的非线性自组织特征：其幂律指数的低值说明节奏自由度高、系统复杂度强，而诗行长度的极端波动则揭示了口头诗体在情感爆发与叙事拉伸中的结构机制。总体来看，吴歌的节奏既非古典诗体的格律均衡，也非现代自由诗的完全随机，而是一种介于吟唱与言说之间的“中间节奏结构”。这种结构体现了民歌作为口头诗体的生成逻辑，即通过语言的流动性与方言语调的韵律特征，建构出独特的诗性秩序。

5. 与新诗及民歌的聚类比较

为进一步揭示吴歌在汉语诗体系统中的类型归属与语言结构特征，本节综合前文所得的 Zipf 指数、Busemann 系数及节奏参数等特征数据，采用多维聚类分析方法，将《吴歌甲集》与民歌“花儿” [13]、汉语新诗 [14]、古典五言诗 [15] 与七言诗 [15] 进行整体比较。其中“花儿”、新诗及古典诗体的统计指标分别来自 [13]-[15] 的计量研究成果。研究旨在考察吴歌在语言组织、主动性话语特征及节奏构型等方面与其他诗体的统计相似性，从而确定其在汉语诗歌谱系中的位置 [11] [12]。

根据标准化特征矩阵，见表 4，吴歌的 Zipf 指数  $a_z = 0.6524$  显著低于其他诗体，显示词频分布的平缓性和开放性；Busemann 系数  $Q = 0.7570$  则位于民歌“花儿”(0.81) 与新诗(0.63) 之间，表明吴歌在话语倾向上兼具“主动叙述”与“抒情描写”的双重特性；节奏参数  $a_r = 0.2321$  与极高的行长标准差(111.57) 共同揭示了其在节奏结构上的高度自由性与口语性。这一结果与刘海涛、张晓瑾等对“花儿”与新诗的统计研究相互印证，表明吴歌虽与民歌共享口头性特征，但在句法与节奏层面表现出更强的非均质性 [13]-[15]。不过，吴歌语料的口头性与方言特征在转写过程中往往被标准化，而新诗与古典诗多以书面形式留存，这种文本形态的不一致可能对特征矩阵的距离结构产生一定影响。

Table 4. Feature matrix of poetic genres  
表 4. 各诗体特征参数矩阵

语体	Zipf 指数 $a_z$	Busemann 系数 $Q$	节奏指数 $a_r$	节奏标准差	平均行长
吴歌	0.6524	0.7570	0.2321	111.57	93.91
民歌“花儿”	0.6900	0.8100	0.4200	48.60	19.30
汉语新诗	0.7100	0.6300	0.5500	35.10	21.40
古典五言诗	0.7800	0.4900	0.6700	8.40	5.00
古典七言诗	0.7500	0.5300	0.7100	9.70	7.00

层次聚类分析采用 Ward 最小方差法，以欧氏距离为度量，所得树状图如图 3 所示。聚类结果显示，古典五言诗与七言诗首先聚为一支，反映出格律诗体在语言结构上的高度一致性；其次，民歌“花儿”与汉语新诗形成一类，显示两者在自由句法与主动话语特征上的统计接近性；而吴歌则在最高层次上单独分出一枝，与其他所有诗体保持显著距离。这一结构表明，吴歌在统计空间中构成独立聚类单元，其语言模式无法由任何单一诗体代表性地归纳，体现出鲜明的区域诗体特征与语言系统独立性 [16]-[18]。



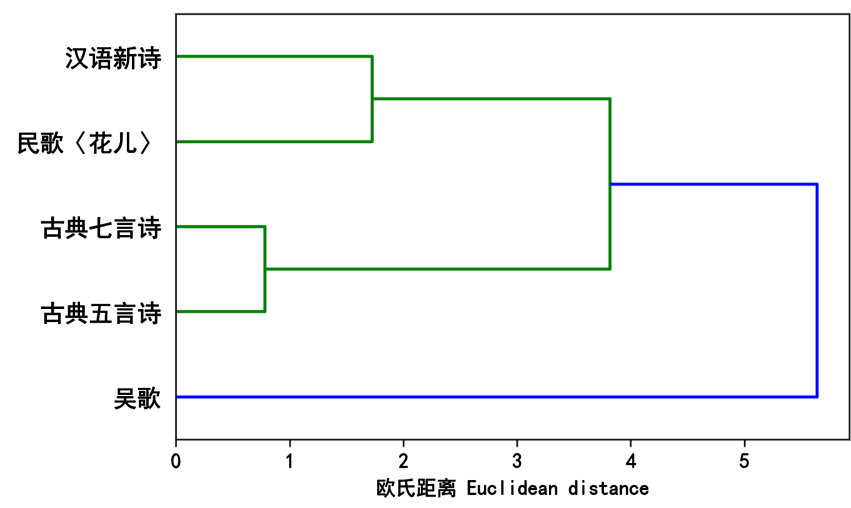


Figure 3. Hierarchical clustering of poetic genres  
图 3. 诗体特征层次聚类图

为了进一步验证聚类结构的可视化表现，本研究采用主成分分析(PCA)进行二维投影，见图 4。投影结果同样显示，吴歌样本点位于坐标系的左上象限，与其余诗体群体明显分离。民歌“花儿”与汉语新诗在图中相互靠近，说明两者在语言活性与句法灵活度上具有统计同构性；古典诗体两点则集中于右侧区域，显示其节奏紧密与词汇控制度高的特征。吴歌与所有诗体的空间距离均较大，印证其在语体谱系中的“非对称”地位：既承续民歌的口语传统，又超越新诗的现代书写范式，在统计意义上表现为一种“方言自由诗”的类型形态[14] [17] [19]。

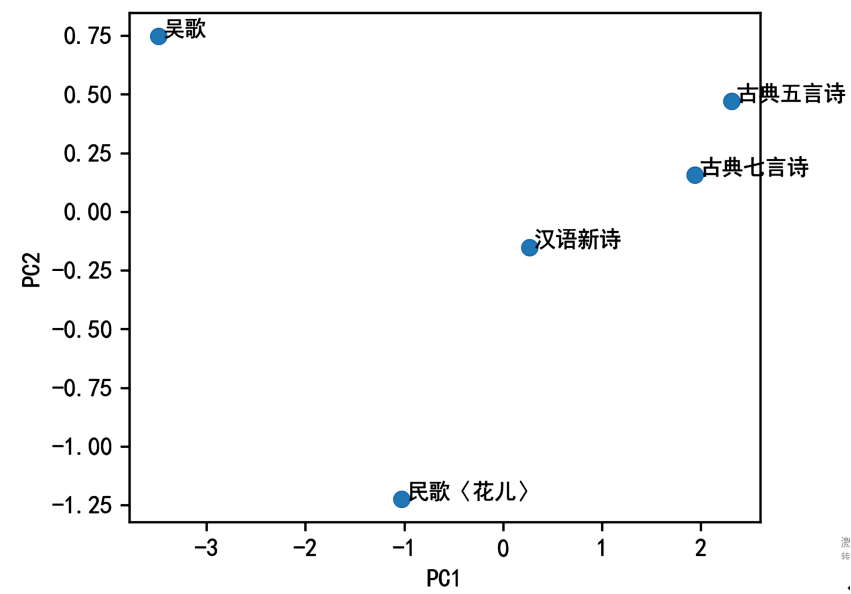


Figure 4. Similarity of poetic genres (PCA projection)  
图 4. 诗体特征相似性分布(PCA 投影)

从语言计量学角度看，吴歌的这种“异质性聚类”揭示了方言诗体在语言系统中的特殊演化机制。一方面，吴歌的高 Busemann 系数与低 Zipf 指数共同表明，其语言组织同时具备主动叙述性与高自由度分布；另一方面，节奏参数的低集中度与高标准差反映出其音步结构的自然生成性。换言之，吴歌在统

计意义上展现出一种“非格律－高流动”的诗体类型，其语言节奏更接近自然语流的动态平衡，而非古典诗歌的韵律约束[18][21][22]。这种结构特征表明，吴歌在语言演化上处于从民间口头性向现代诗性表达过渡的关键位置，构成连接方言民歌与现代汉语诗体的重要桥梁。

综合来看，聚类与投影分析均证明吴歌在统计空间上独立于传统格律诗与现代书面诗之间，形成一个具有显著方言特征与高复杂度的诗体系统。这一发现不仅为吴歌研究提供了量化依据，也从语言计量学的角度揭示了区域文化与诗歌形式之间的深层对应关系。

## 6. 结语

吴歌作为江南方言文化的产物，在语言结构上展现出高度的复杂性与独立性。本文通过词频、词类、节奏与聚类四个维度的计量分析发现，吴歌不仅在词汇分布上呈现低集中度的开放体系，在话语构型上亦体现出主动与抒情并行的双重倾向，其节奏组织则脱离格律约束而趋向口语自发的语流结构。这一整体特征揭示出吴歌诗体的生成并非格律模仿或书面转写，而是方言语音、语义与情感系统在长期民间演化中自组织形成的诗性样态。吴歌的语言规律表明，方言诗体同样具有可统计化的结构特征，其诗性来源于语流内部的动态平衡与文化记忆的音韵化表达。因而，吴歌不仅是民间口语的延伸，更是汉语诗体在地域化、口头化与现代化三重张力中形成的一种独立范式。

## 基金项目

本研究得到了江苏高校哲学社会科学研究一般项目(编号：2022SJYB1069)、无锡市哲学社会科学招标课题(编号：WXS22-C-83、WXS25-C-02)的资助和支持。

## 参考文献

- [1] 中国社会科学网. 吴歌千年传唱赓续中华文脉[EB/OL]. [https://www.cssn.cn/skgz/bwyc/202309/t20230912\\_5685076.shtml](https://www.cssn.cn/skgz/bwyc/202309/t20230912_5685076.shtml), 2023-09-12.
- [2] 陈书录. 吴歌越吟：诗化的江南民间文化[EB/OL]. <https://www.chinesefolklore.org.cn/web/index.php?NewsID=9568>, 2011-12-14.
- [3] 左思. 吴都赋[M]. 上海：上海古籍出版社，1986.
- [4] 周勋初. 吴越文化史论[M]. 杭州：浙江人民出版社，2004.
- [5] 吴大顺. 魏晋南北朝乐府歌辞研究[M]. 上海：上海古籍出版社，2009.
- [6] 顾颉刚. 吴歌甲集[M]. 北京：北京大学出版部，1926.
- [7] 王焰安. 从《吴声歌》《西曲歌》看南朝民间情歌[J]. 深圳大学学报(人文社会科学版), 2016, 33(5): 120-126.
- [8] 国家文化和旅游部. 国家级非物质文化遗产代表性项目名录[EB/OL]. <https://www.ihchina.cn/project.html>, 2022-06-01.
- [9] 朱银华. 论扬州民歌对不同时期吴声歌曲的影响与贡献[J]. 扬州大学学报, 2018(4): 45-52.
- [10] 杨俊光. 吴歌私情母题及女性称谓原型探佚[J]. 民间文学, 2019(4): 45-56.
- [11] 刘海涛. 计量语言学导论[M]. 北京：商务印书馆，2017.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] Pan, X. and Liu, H. (2019) Quantitative Stylistic Evolution of Chinese Poetry. *Journal of Quantitative Linguistics*, **26**, 258-275.
- [16] Popescu, I., Mačutek, J. and Altmann, G. (2010) Word Forms, Style and Typology. *Glottology*, **3**, 89-96. <https://doi.org/10.1515/glot-2010-0006>
- [17] Popescu, I., Čech, R. and Altmann, G. (2013) Descriptivity in Slovak Lyrics. *Glottology*, **4**, 92-106. <https://doi.org/10.1524/glot.2013.0007>

- [18] Altmann, G. (2014) Unified Modeling of Diversification in Language. RAM-Verlag.
- [19] Fang, Y. and Wang, Y. (2017) Quantitative Linguistic Research of Contemporary Chinese. *Journal of Quantitative Linguistics*, **25**, 107-121. <https://doi.org/10.1080/09296174.2017.1352478>
- [20] 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.
- [21] Busemann, A. (1923) Das Verhältnis zwischen Substantiv und Verb in der deutschen Sprache. *Indogermanische Forschungen*, **41**, 195-228.
- [22] Liu, H. and Liang, J. (2013) The Quantitative Study of Chinese Poetic Language: Busemann's Coefficient Revisited. *Glottometrics*, No. 25, 45-59.