

# 汉语幽默风格多维语言特征计量研究——以相声文本为例

陈永胜

华侨大学, 中国

王雪琳\*

暨南大学, 中国

## 摘要

幽默作为一种重要的语言表现风格, 其生成机制与语言表达特征一直是自然语言处理领域关注的重要问题。基于此, 本文引入计量风格学研究范式, 以相声文本作为幽默风格载体、电视访谈文本作为非幽默风格对照, 对汉语幽默风格的语言特征开展实证分析。研究表明: (1) 与非幽默文本相比, 汉语幽默风格文本的介词与连词使用比例显著偏低, 而形容词与副词使用比例显著偏高; (2) 在语义层面, 汉语幽默风格文本表现出更高层次的语义不一致性, 反映出其依赖“语义违背”以实现幽默效果; (3) 在情感表达维度, 汉语幽默风格以中性情感与一定比例的消极情感相互交织为典型特征, 而非幽默风格文本整体呈现出更为鲜明的积极情感倾向。本研究的结果对揭示汉语幽默汉语文本的语言特征及其生成机制具有一定的参考价值, 同时为汉语幽默风格研究提供了一种可操作、可检验的量化研究路径。

## 关键词

汉语幽默, 计算幽默, 语言特征, 语言风格, 数据驱动

## 1 引言

语言风格是交际参与者运用语言表达手段综合表现出来的气氛和格调, 语言的表现风格是这种气氛格调所形成的审美形态 (黎运汉, 2002)。幽默风格是指诙谐而轻松的语言风格, 它通过语言互动在个人、群体或组织层面的交流中引发积极的情绪体验与认知反应 (胡范铸, 1987; Attardo, 2020)。近年来, 计算幽默学逐渐成为计算语言学中的一个重要研究方向。现有研究多以印欧语系语言为考察对象, 尤以英语语料最为集中, 并且已经在幽默识别、特征建模与机制探讨等方面取得了较为丰富的成果 (Ritchie, 2001; Mihalcea & Strapparava, 2006; Kao et al., 2016; Skalicky et al., 2020)。既有研究通常围绕数据集构建、语言特征提取与分类算法设计三大方向展开, 其中语言特征的选取被视为连接幽默风格学理论与计算模型的关键环节 (Kalloniatis & Adamidis, 2024)。研究者从语言学角度出发, 提出并验证了多种可用于刻画幽默风格文本的语言风格手段, 包括词性分布、句法结构、歧义性、语义不一致性、情感分布以及修辞特征等。

\* 通讯作者。联系电邮: [wangxuelin2022@gmail.com](mailto:wangxuelin2022@gmail.com)

然而,与英语计算幽默风格研究的快速发展相比,针对汉语幽默风格的计量研究仍十分有限(林鸿飞等,2024),相关研究在语料库建设、特征体系构建与理论探讨上均存在明显不足。这一缺口不仅限制了我们对汉语幽默风格语言机制的理解,也在一定程度上制约了计算幽默研究在跨语言层面的理论完整性。鉴于此,本文引入计量风格学研究范式,选取相声这一代表性的汉语幽默口语文本,并以电视访谈口语文本作为非幽默风格对照,采用词类分布、语义不一致性与情感分布下的语言风格计量指标,考察汉语幽默与非幽默风格文本的语言使用特征。

## 2 研究综述

### 2.1 语言风格的计量研究

计量风格学研究认为,语言风格可通过词汇、句法、语篇等层面的风格手段及其风格指标得到客观表征,从而为风格差异的比较、归类与解释提供实证基础(Mosteller & Wallace, 1964; 曾毅平、朱晓文, 2006; 胡显耀, 2021)。从发展脉络看,早期计量风格研究多依赖词频、词长、句长等浅层统计特征(Mosteller & Wallace, 1964)。此后,随着语料库资源日益丰富、多元统计方法日趋成熟,语言风格研究转向多维度特征分析,完成了从人工统计到自动化大规模提取、从单变量对比到多元统计与分类建模的范式升级(Stamatatos et al., 2001)。在研究主题上,计量风格学已广泛应用于多类语言对象,如作者身份判别与作家作品风格分析(Hoover, 2007; Hao et al., 2025)、功能性语体与文体风格的比较(Biber, 1988; McIntyre et al., 2019)、翻译风格与译者风格等研究(许明、蒋跃, 2020)。大量的计量风格学研究表明,不同语言风格在风格手段与风格指标上存在显著差异。本研究遵循这一基本假设,对汉语幽默与非幽默风格的口语文本展开计量分析。

### 2.2 幽默研究的理论脉络及其计算转向

幽默风格研究在西方学术传统中由来已久,其理论源头可追溯至古希腊哲学与修辞学传统。总体来看,从古希腊的亚里士多德时期一直到20世纪80年代前,主要形成了优越论、释放论和乖讹论三大经典理论(刘乃实、张韧弦, 2016: 25; 林鸿飞等, 2016)。这三种理论分别从社会学、神经分析与认知冲突的角度解释幽默现象,构成了幽默研究最为基础的理论框架(蔡辉、尹星, 2005; 刘乃实、张韧弦, 2016)。20世纪后半叶,幽默研究呈现出明显的多学科交叉特征。随着语义学、语用学及认知语言学的发展,语言学领域对幽默的探讨也进一步转向其认知机制与语义生成机制(Attardo, 1997)。

随着计算语言学与人工智能的发展,幽默研究逐渐由理论分析转向数据驱动的实证研究(Partington, 2006)。现有研究大体可归纳为以下几条路径。首先,幽默识别是计算幽默风格研究中最基础,也最早发展的研究路径。相关早期研究旨在区分“幽默”与“非幽默”文本,认为不协调性与预期违背是幽默风格生成的关键机制(Mihalcea & Strapparava, 2006)。幽默通常源于不同情境、意义或认知预期之间的不一致、不协调或强烈反差(Lefcourt, 2012)。因此,幽默识别需要充分利用幽默的表现形式,通过量化文本,捕获语义的差异程度,获取文本序列中的每个词语的词义特征,将词义特征与语义特征进行拼接,分析词语级别的语义不一致性(林鸿飞等, 2024)。随着研究任务与语料类型的扩展,幽默识别研究逐渐涵盖笑话、新闻标题及社交媒体文本等多种文本类型,并形成了较为规范的评测任务与数据集(Potash et al., 2017)。

其次，以语言学特征为导向的研究路径，强调从词汇、句法、语义与语篇层面，提取幽默可量化、可建模的语言特征。相关研究已系统考察了词类分布、语义不一致性、语篇连贯等语言特征在幽默生成与识别中的作用（Skalicky et al., 2016）。此研究路径不仅服务于幽默识别任务，也旨在揭示幽默风格是如何通过语言风格手段得以建构与呈现的，在可解释性方面具有明显优势。最后，随着研究目标的细化，面向幽默类型与功能的计算研究逐渐兴起，该路径关注幽默在交际中的具体功能，如讽刺、攻击、自嘲或娱乐，并进一步区分冒犯性幽默与非冒犯性幽默（Meaney et al., 2021）。相关研究结合语言特征与可解释模型，分析幽默在社会偏见、刻板印象和价值观表达中的作用，推动计算幽默研究与社会语言学及伦理研究的交叉融合。

### 2.3 汉语计算幽默研究的现状与不足

英语的计算幽默研究在范式、方法、语料资源上已日趋成熟。相比之下，汉语幽默的计算研究仍处于起步阶段，相关成果较为有限。幽默高度依赖具体语言的语言特征与文化约定，其语言实现机制具有显著的语言特异性（Vandaele, 2010），这使得在汉语语境下开展本土化的计算幽默风格研究具有重要意义。

近年来，学界针对汉语幽默风格已开展一定探索，但整体仍以个案分析与理论阐释为主。例如，戈玲玲与何元建（2012）以《围城》为语料，基于言语幽默的理论框架，指出汉语幽默的生成与理解不仅涉及修辞手段的运用，而且依赖于真诚性、相似性与和谐性等语用原则。李先进（2013）以《阿Q正传》为对象，结合语料库分析进一步提出，汉语幽默在表层主要通过各种修辞手法来实现，而在深层结构层面则呈现出典型的脚本对立机制。不同于以个案分析为主的研究路径，近年来亦有研究对汉语幽默文本开展大规模的计量分析。如林鸿飞等（2024）面向社交媒体语境构建了中文幽默计算的理论与实践框架，将情感特征、语义不一致性与模糊性等作为汉语幽默风格的核心语言风格手段纳入计算模型，为汉语幽默风格的计算研究提供了完整的研究思路。与此同时，汉语幽默研究也开始重视基础语料资源建设，例如面向微博文本构建的中文幽默语料库 HumorWB，成为首个融入社交语境信息的汉语幽默语料库（Zeng et al., 2024）。

尽管上述研究为理解汉语幽默的语言特征提供了有益探索，但总体来看仍存在若干明显局限。首先，现有汉语幽默风格研究对象多集中于单一作家、单部作品，结论的适用范围相对有限。其次，现有研究对汉语幽默风格文本语言特征的计量风格学分析仍相对不足。再次，相关研究在语料类型覆盖上仍有不足，现有研究主要关注文学文本和社交媒体文本，对以口语表演为核心载体的相声类幽默风格文本关注明显不足。

针对上述问题，本文选取相声这一极具代表性的汉语幽默口语体裁，以电视访谈类节目作为非幽默口语文本的对照语料，采用计量风格学范式，从词类分布、语义不一致性与情感分布三个维度，对汉语幽默与非幽默风格文本的语言特征展开系统对比分析。基于此，本文拟重点回答以下研究问题：

- （1）相较于电视访谈类节目文本，相声文本在词类分布上是否呈现显著差异？
- （2）相较于电视访谈类节目文本，相声文本在语义不一致性特征上是否呈现显著差异？
- （3）相较于电视访谈类节目文本，相声文本在情感分布上是否呈现显著差异？

### 3 语料与方法

#### 3.1 语料选取及处理

本文选取相声文本构建汉语幽默风格语料库，主要基于以下几方面的考虑。首先，相声作为以幽默为核心功能的传统曲艺形式，其语言表达具有高度的幽默密集性与持续性，幽默效果贯穿整体话语结构，因而为幽默语言特征的系统分析提供了相对稳定的研究对象。其次，相声语言以口语为基础，但并非自然发生的即兴口语，而是一种经过预先构思与程式化加工的表演语体，兼具口语表达的直观性与语言组织的规范性，能够在较大程度上集中呈现汉语口语幽默的典型语言特征。最后，相声在现代汉语语境中具有广泛的社会影响力与接受基础，是研究汉语幽默语言使用的重要代表性载体之一。

在此基础上，本文以《中国相声大全》为主要语料来源，构建汉语幽默风格语料库，并在保证文本完整性的前提下对语料进行筛选与整理。为开展对比分析，本文同时构建非幽默口语对照语料库。本文选取的“非幽默”语料并不是幽默风格的绝对缺失，而是不以制造幽默为主要交际目的的语料。访谈文本在话轮转换、即兴表达等方面与相声同属口语范畴，但交际目的为信息传递与叙事，而非制造幽默。因此，我们选取央视访谈节目《鲁豫有约》与《面对面》的转写文本作为数据来源，以在口语表达方式与话题多样性等方面与相声语料保持基本可比性。最终构建的语料库包含对口相声、群口相声以及两类访谈文本各 100 篇。选取语料后，我们首先清洗原始文本（删除节目标题、舞台提示等无关信息），随后，采用 LTP (<https://github.com/HITSCIR/ltp>) 对语料进行分词及词性标注处理，并进行人工校对。最后，利用自编 Python 程序提取相关语言学指标。

需要说明的是，相声文本与电视访谈文本在文本长度与话轮结构上仍存在一定差异。相声作为表演性口语体裁，话轮一般较短，频繁切换，而电视访谈文本则呈现出更复杂的话轮特征。上述差异在一定程度上可能影响部分语言特征的绝对频次分布。为降低文本长度和话轮长度差异对结果的干扰，本文在统计分析过程中对各项指标进行了标准化处理，以尽可能减少文本规模差异对分析结果的干扰。尽管如此，文本长度和话轮带来的潜在影响仍需在后续研究中通过更精细的语段切分进一步加以考察。

表 1. 语料库规模

文本库	篇数	总词数
对口相声	100	258878
群口相声	100	247232
面对面	100	540134
鲁豫有约	100	623850

#### 3.2 指标计算方法

首先，本文计算各主要词类在文本中所占的相对比例，以消除文本长度差异对统计结果的潜在影响。词类划分依托于 LTP 系统提供的词性标注体系，在完成自动标注的基础上，通过人工校验进一步确保标注的准确性与可靠性。

其次，在语义不一致性 (SI) 测量方面，本文采用 sentence-transformers 模型对文本进行语义编码，并以句际语义一致性为基础，通过反向转换构建语义不一致性指标。具体而言，借

助预训练的 Sentence-BERT 类句向量模型，将文本中的每个句子表示为固定维度的语义向量，并据此计算相邻句子之间的余弦相似度，以表征句际语义一致性程度。一般而言，余弦相似度越高，说明相邻句子之间的语义衔接越紧密，文本整体的语义一致性越强；反之，则表明句子之间的语义偏离越明显。基于此，本文进一步采用“1-余弦相似度”的方式，将句际语义一致性反向转换为语义不一致性，并以所有相邻句子对该值的平均数作为文本语义不一致性的核心测量指标，计算公式如式（1）所示：

$$(1) SI = \frac{1}{M-1} \sum_{i=1}^{M-1} (1 - \cos(v_i, v_{i+1}))$$

其中， $M$  表示文本中的句子总数， $v_i$  和  $v_{i+1}$  分别表示第  $i$  个句子和第  $i+1$  个句子的语义向量， $\cos(v_i, v_{i+1})$  表示两个句子向量之间的余弦相似度。该指标值越大，说明相邻句子之间的语义跳跃越明显，文本整体的语义不一致性越高。

最后，在情感分析层面，本文采用大连理工大学情感词汇本体库（简称“大连理工情感词典”）进行情感词识别。大连理工情感词典包含 27466 个词条，每个词条标注为“积极”“中立”“消极”三类情感极性（徐琳宏等，2008）。我们将本研究的分词结果与情感词典进行匹配，统计各文本中不同情感极性词语的出现频次。为消除文本长度带来的影响，采用相对占比进行标准化处理，计算方法如式（2）所示。在此基础上，分别计算积极、消极与中性情感词的占比，以此作为衡量文本情感倾向的量化指标。

$$(2) \text{某一极性情感词占比} = \text{该极性情感词频次} / \text{情感词总频次}$$

## 4 结果

### 4.1 词类分布

表 2 和图 1 呈现了不同文本类型的词类占比分布特征。总体来看，相声文本与电视访谈节目文本在多类词项的使用比例上存在一定差异。统计检验结果进一步表明，显著差异主要集中在介词、连词、形容词和副词四类词项上。具体而言，介词占比（ $F=151.128$ ， $df=3$ ， $p<0.01$ ）、连词占比（ $F=406.223$ ， $df=3$ ， $p<0.01$ ）、副词占比（ $F=42.458$ ， $df=3$ ， $p<0.01$ ）以及形容词占比（ $F=7.112$ ， $df=3$ ， $p<0.01$ ）在不同文本类型之间均存在显著差异。

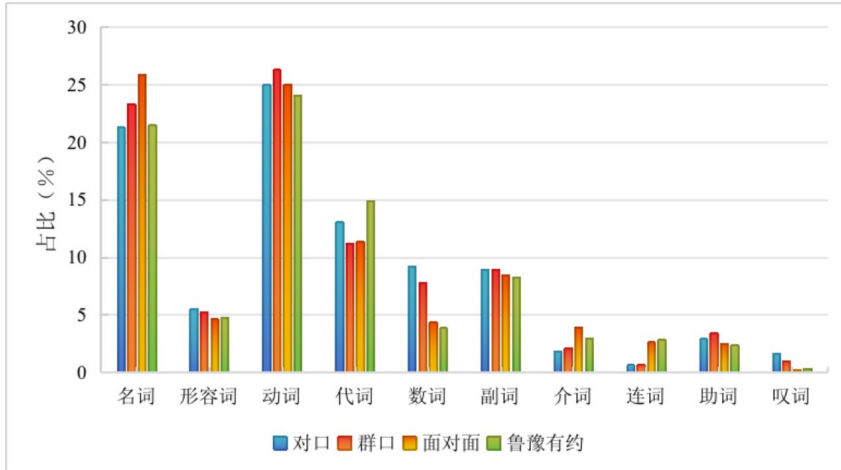
表 2. 幽默与非幽默口语文本词类占比结果（%）

类型	名词	形容词	动词	代词	数词	副词	介词	连词	助词	叹词
对口	21.3	5.48	24.98	13.06	9.23	8.93	1.84	0.62	2.9	1.63
群口	23.3	5.26	26.31	11.21	7.77	8.94	2.08	0.65	3.39	0.97
面对面	25.88	4.62	25.01	11.33	4.33	8.44	3.87	2.61	2.49	0.02
鲁豫有约	21.5	4.72	24.09	14.93	3.85	8.24	2.98	2.81	2.33	0.33

具体来说，相声文本中的介词和连词使用比例显著低于电视访谈文本，其中对口相声和群口相声的介词占比分别为 1.84% 和 2.08%，连词占比分别为 0.62% 和 0.65%；而电视访谈文本中上述两类词项的占比明显更高，《面对面》的介词和连词占比分别为 3.87% 和 2.61%，《鲁豫有约》分别为 2.98% 和 2.81%。与之相对，相声文本在形容词和副词的使用上显著高于电视访

谈文本,其中对口相声和群口相声的形容词占比分别为 5.48% 和 5.26%,副词占比分别为 8.93% 和 8.94%;而电视访谈文本中形容词和副词占比整体较低,《面对面》分别为 4.62% 和 8.44%,《鲁豫有约》分别为 4.72% 和 8.24%。

图 1. 幽默与非幽默口语文本词类柱状图 (%)



## 4.2 语义不一致性

表 3 展示了四类文本在语义不一致性指标上的统计结果。我们首先对该指标进行方差齐性检验,结果显示不满足方差齐性假设 ( $p < 0.01$ ),因此采用 Kruskal-Wallis 检验比较不同文本类型之间的差异。结果表明,四类文本的语义不一致性指标存在显著差异 ( $H(3) = 322.845, p < .001$ )。为明确差异的具体来源,我们进一步进行事后两两比较分析,结果显示幽默与非幽默风格文本之间均存在显著差异 ( $ps < 0.01$ )。

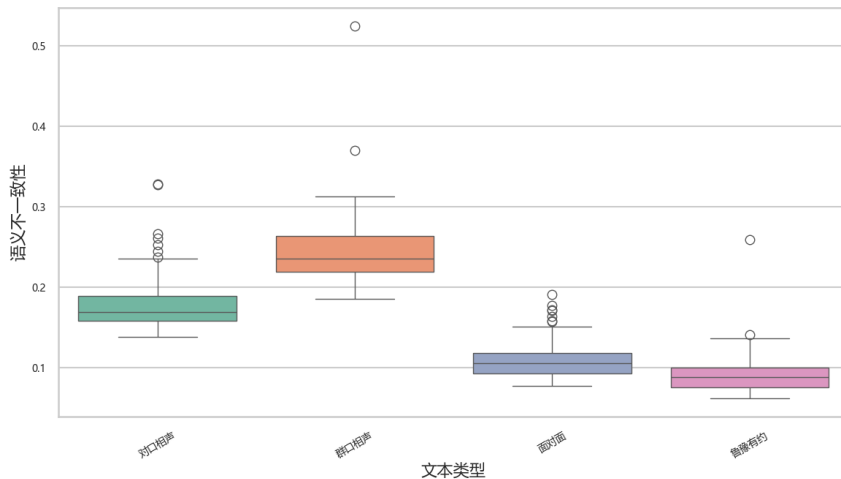
表 3. 幽默与非幽默口语文本语义不一致性统计检验结果

文本类型	N	平均值	标准差	标准误差	最小值	最大值
对口相声	100	0.179	0.034	0.003	0.138	0.328
群口相声	100	0.246	0.047	0.004	0.187	0.523
面对面	100	0.111	0.024	0.002	0.062	0.259
鲁豫有约	100	0.091	0.024	0.002	0.077	0.191

如图 2 所示,汉语幽默与非幽默风格文本在语义不一致性指标上的分布形态存在明显差异。群口相声的语义不一致性最高,对口相声次之,《面对面》再次,《鲁豫有约》最低。这一分布趋势与表 2 中的描述性统计结果一致,即群口相声平均值为 0.246,对口相声为 0.179,《面对面》为 0.111,《鲁豫有约》为 0.091,表明汉语幽默口语文本整体上具有更高层次的语义不一致性,而非幽默口语文本则表现出更强的语义连贯性。从中位数和箱体位置来看,群口相声的箱体整体最高,说明其语义不一致性在四类文本中最为突出;对口相声的箱体虽低于群口相声,但仍明显高于两类电视访谈文本,表明相声文本在句际层面普遍存在较强的语义跳跃。相比之下,《面对面》与《鲁豫有约》的箱体位置整体较低,且《鲁豫有约》最低,说明电视访谈文本的语义推进更趋平稳,句际衔接更为紧密,整体一致性较强。从离散程度看,群口相声的箱体范围和

须状线跨度最大，同时高值离群点也最为突出，说明该类文本内部的语义不一致性波动最为明显，不同文本之间差异较大。对口相声的分布相对集中一些，但仍存在多个高值离群点，表明其中部分语段同样具有较强的语义突转现象。相比之下，《面对面》与《鲁豫有约》的箱体更紧凑，四分位距较小，离群点数量也相对较少，反映出这两类文本内部的语义组织总体更加稳定。

图 2. 幽默与非幽默口语文本语义不一致性箱型图



### 4.3 情感色彩

如表 4 和图 3 所示，不同文本类型在情感色彩词使用分布上呈现出明显的差异。总体来看，电视访谈类文本表现出明显的积极情感倾向，积极情感词在两类访谈文本中均占据绝对优势。其中，《面对面》中积极情感词占比为 76.87%，《鲁豫有约》更高达 94.17%，而中性词和消极词的占比均明显较低，分别仅为 18.43%、4.70% 和 2.36%、3.47%。这表明电视访谈文本在情感表达上整体更偏向正向评价与积极建构，情感基调较为鲜明而集中。与此不同，对口相声和群口相声均以中性情感词占比最高，分别为 52.36% 和 62.33%，说明相声文本在情感表达上更倾向于以中性叙述为主。同时，两类相声中的消极情感词占比也相对较高，对口相声为 28.37%，群口相声为 21.03%，明显高于电视访谈文本；而积极情感词的占比则相对较低，分别仅为 19.27% 和 16.64%。

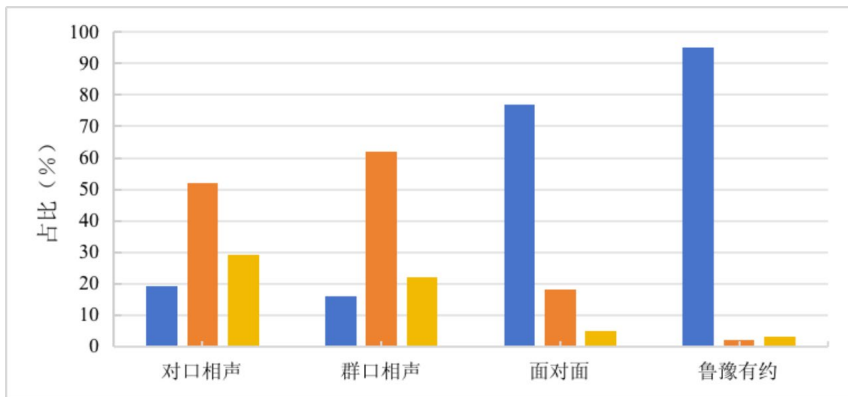
表 4. 幽默与非幽默口语文本情感占比结果 (%)

类型	积极词	中性词	消极词
对口相声	19.27	52.36	28.37
群口相声	16.64	62.33	21.03
面对面	76.87	18.43	4.70
鲁豫有约	94.17	2.36	3.47

由此可见，相声文本更多呈现出中性情感为主、消极情感并存的分布格局，其情感表达结构较电视访谈更为复杂，也更具反差性和张力。为检验上述分布差异的统计显著性，本文对四类文本的情感类别分布进行卡方独立性检验。结果显示，情感类别与文本类型之间存在显著关

联 ( $\chi^2(21, N = 400) = 85.3, p < 0.01$ ), 表明汉语幽默与非幽默风格文本在情感表达的分布结构上具有显著差异。

图 3. 幽默与非幽默口语文本情感词占比柱状图 (%)



## 5 讨论

### 5.1 词类分布

研究表明, 相声文本在介词与连词的使用占比上显著低于电视访谈文本, 而在形容词与副词的使用占比上显著高于电视访谈文本。该发现与既有计算幽默研究的结论一致, 即词类分布特征能够有效区分幽默文本与非幽默文本 (Skalicky et al., 2016)。在汉语中, 介词与连词主要承担句法连接与逻辑关系标记的功能 (黎锦熙, 1958), 其在相声文本中的低占比表明, 汉语幽默风格文本在语篇组织上较少依赖显性的语法标记来维系逻辑关系, 而更倾向于采用短句、并列结构或省略式表达这种非常规搭配来推进话语。如例句 (1) 中, 对包公的多种称谓进行连续罗列时未使用任何连接词, 句式简洁、节奏紧凑, 贴合口头表演的语言特点, 也借密集的称谓铺陈形成独特的表达趣味; 若加入“和”“还有”等连词, 反而会打破这份明快节奏, 削弱表达的趣味性。反观电视访谈节目, 其核心功能在于信息传递与叙事推进, 对话语逻辑的连贯性要求更高, 因此需更多借助连词、介词等显性连接手段, 保障表达的逻辑清晰, 这类词项的使用占比也随之更高。

(1) “他姓包名拯字希文, 世称包文正, 又叫包公、包大人、包先生、包老爷、密斯特包!” (《道法无边》)

此外, 相声文本还表现出对形容词与副词的显著偏好, 这与既有英语计算幽默研究中关于幽默文本“高描写性”的发现一致, 即幽默风格话语往往通过更为密集的修饰性成分来强化语义表现力 (Skalicky et al., 2016)。王金玲 (2002) 认为幽默风格文本中, 作者常运用丰富的想象着力把描述事物的本质特征极力地夸大或缩小, 使语句突出、鲜明, 从而使语产生幽默效果。而这些描写大都需要形容词和副词的参与, 在相声语境中, 形容词不仅承担对事物属性的描写功能, 更常用于即时评价与夸张刻画, 通过连续叠加的方式放大人物特征或情境张力, 从而为笑料的生成提供语义基础。如例 (2) 中连续叠加的“黑”“白”两个修饰词提供了更多细节, 使听众在心中构建形象时产生联想, 想到一个滑稽、古怪的丑角形象, 从而发笑。

(2) “哪儿呀! 就是那条黑狗, 黑狗白鼻梁儿。” (《醋点灯》)

副词主要修饰动作或状态, 可以调控语句的语气强度、表达程度与言语节奏。相声文本中副词的高频使用, 有助于在有限语篇内快速制造语义起伏与表达转折, 增强语言的画面感和表

现力。如例(3)中的“偏”标示出行为对既有认知前提的违背,不仅增强了语义上的不一致性,也进一步凸显了人物投机取巧的性格特征,从而强化了语言的戏谑效果与喜剧张力。

(3)“你想啊,人多,他离着门口远,卖票的挤得过去吗?明知道,可他偏喊。”(《白吃》)

这种通过修饰性成分强化叙述效果的策略,与幽默文本依赖夸张、评价和语义强化来建构幽默效果的机制相契合。相比之下,电视访谈文本以信息传递为主要目的,对语言准确性与客观性的要求更高,因而在形容词与副词的使用上相对克制。

## 5.2 语义不一致性

幽默风格话语的理解通常依赖语义与语用的动态交互。以语义脚本理论为代表的幽默理论认为,幽默风格效果的生成常通过激活并对立两个语义脚本,引导受众结合既有知识与当下的语言输入进行语义的重新解读(Raskin, 1985)。在相声中,这种机制常体现为“预期违背”。例如例(4)通过反复强调“住北房靠南头儿”营造“同住一屋理应互相认识”的强预期,最终却引出“仍不认识”的荒诞结果,形成明显的脚本冲突与认知落差,从而实现幽默效果。

(4)“甲:我住北房靠南头儿。乙:我也住北房靠南头儿……乙:我怎么不认识您哪?”(《拖妻献子》)

在计算幽默研究中,这类由预期违背产生的语义不一致性被认为是幽默风格的重要语言学特征。相关研究表明,语义不一致程度能够有效区分幽默与非幽默文本(Morales & Zhai, 2017)。本文结果显示,相声文本的语义不一致性特征较为突出。这可能与相声的表达特点有关。表演者往往先进行叙事铺垫,引导听众形成固定的理解预期;随后在包袱处突然转换语义,打破听众原有的判断,使前后话语出现明显的语义反差,进而产生幽默效果。正如黎运汉、盛永生(2010: 539)所指出的,超常搭配往往能够突破语义组合的常规,从而产生鲜明的幽默诙谐效果。

从文本层面看,相声中的语义不一致性可表现为多个层级。第一,在人物塑造与情节推进上,相声常采用高度程式化的喜剧角色结构与互动模式,例如甲乙角色的显著反差(甲夸张吹嘘、乙较真追问),使得同一主题在连续追问中不断发生偏转与失误累积(马季, 1980)。第二,在语篇组织上,相声为了维持听觉注意力与笑料密度,往往允许话题在较短跨度内频繁跳转,通过插入式情节、语用包袱或语言游戏实现局部连贯、整体跳跃的结构,如例(5)从民间买卖膏药突然转入打把式的展示性描写,情节与语义框架发生明显切换,降低了跨片段的语义连贯性。

(5)“一请人,先不说卖膏药。先说打把式……要值好,你给我叫声好……我临离‘镖局子’的时候,带点儿宝贝东西来。”(《生意经》)

这种语义上的跳跃与突转,在语用层面常常表现为对合作原则中“关系准则”的刻意违反。格莱斯(Grice)(1989)在合作原则中对关系准则做出界定,即所说的话要有关联。这就要求会话中双方所提供的信息需要与会话主题相关。但是在幽默文本中,言者常会采用“回避话题”“插科打诨”等方式有意违反关系准则制造幽默效果(黄影, 2023)。

电视访谈类文本通常以信息传递与主题探讨为核心目标,其话轮组织与提问策略更注重话题的有序展开与逻辑推进。主持人往往通过追问、复述、转承等主持技巧将与嘉宾的谈话维持在相对稳定的主题范围内,从而提升语义聚焦与整体连贯性。因此,在本研究的计算结果中,电视访谈文本呈现出更高的语义一致水平,而相声文本则表现出更显著的语义跳跃,与相声依赖“预期—突转”结构实现幽默的表达机制相一致。相声的“预期违背”特性契合了幽默风格的语义不一致性特征,而其他类别的幽默文本具体如何,是我们将来要关注的问题。

### 5.3 情感色彩

Cela et al.(2024)研究发现,负面情绪会提升单口喜剧的收视率。晏青、万旭婷(2024)指出,脱口秀将敌对的、攻击的幽默风格转化为常态化的语言策略,成为公然表达不满、回击和自我保护的方式。这说明,幽默风格往往伴随着一定程度的负面情绪,这些情绪因素在幽默加工过程中起到重要触发作用(Reyes al., 2012)。本研究结果显示,相声文本中的消极感情词和中性情感词使用占比较高。这可能是由于相声的叙事往往建立在角色冲突、包袱铺垫、夸张扮演与讽刺表达之上,因而更容易出现中性与消极情感交织的复杂情况。具体来看,相声文本的中性情感词使用比例高于电视节目,这可能也与其表演性质相关。相声文本的许多台词承担着情节推进、背景交代、包袱铺垫及角色串场等功能,因而不直接承载情感信息,会被标注为中性。相对而言,电视节目强调温情与真实表达,《面对面》与《鲁豫有约》中嘉宾倾诉个人故事时,主持人引导话题保持连贯,都会主动营造正向积极的氛围,因此积极情感词使用较高。

相声中的消极情感包括戏谑、讽刺、愤怒与悲伤等。例如戏谑情感可能是由于相声中不乏对人物性格缺陷的调侃、对社会伦理道德的讨论,这种表演往往围绕着日常生活中常见的矛盾和小冲突展开,营造出一种既令人发笑又引人深思的氛围(黄乐平等,2022)。正如肖复兴(2009)指出,反映生活的幽默是相声流淌的血液,对现实的讽刺是相声的灵魂,如果两者都缺失,相声就不再是相声。相声讲究“台上无大小,台下立规矩”,即相声演员在台上可以没大没小,在台下需要守规矩。如例(6)中对师父的调侃,将“名师”说为食物(“粽子”)起到戏谑效果。

(6) “甲:合着你师父就是粽子;乙:还是粽子!”(《傻瓜镖》)

相声自清末在北京的市井社会中形成以来,始终与讽刺性表达紧密相连。作为一种扎根民间的口头表演艺术,相声在揭示社会不公与权力失衡的同时,也为中下层民众提供了情绪宣泄与象征性发声的渠道(祝鹏程,2024)。如例(7)以轻描淡写、插科打诨的方式,讽刺了皇权婚姻的随意性及女性命运的被动性。

(7) “皇上喜欢谁,谁就能当上皇后啦……只要皇上不死,三年娶一拨儿。”(《慈禧入宫》)

## 6 结语

本研究采用计量风格学范式,从词类分布、语义不一致性与情感表达三个维度,对汉语幽默风格文本的语言特征进行了量化考察。研究表明,相声的幽默效果背后蕴含着相对稳定的语言规律。在词类层面,它通过减少显性逻辑连接、强化修饰性成分来提升语言节奏与表现张力;在语义层面,它借助更高层次的语义不一致性建构“预期—突转”的幽默机制;在情感层面,它则呈现出以中性情感为基础、消极情感交织其中的复杂表达格局。由此可见,汉语幽默风格并非不可捉摸的审美直觉,而是一种能够通过计量方法加以揭示和验证的语言风格现象。这些发现为理解汉语幽默风格的生成机制与认知模式提供了量化证据,验证了计量风格学方法在揭示汉语幽默语言风格的有效性。

本文亦存在一定局限。首先,研究对象仅聚焦于相声这一单一幽默体裁,相关结论在其他汉语幽默体裁中的适用性仍有待检验;其二,本文所选取的语言指标有限。鉴于此,未来研究可进一步拓展语料类型,如构建脱口秀、小品、笑话、网络幽默文本等不同幽默类型语料库,同时丰富和挖掘更多幽默风格的风格手段和风格计量指标,并结合跨语言、跨文化对比视角,深入探究汉语幽默风格的独特性。

## 参考文献

- Attardo, S. (1997). The semantic foundations of cognitive theories of humor. *Humor*, 10(4), 395–420.
- Attardo, S. (2020). *The linguistics of humor: An introduction*. Oxford University Press.
- Barbieri, F., & Saggion, H. (2014). Automatic detection of irony and humour in Twitter. In Proceedings of the *International Conference on Computational Creativity (ICCC)* (pp. 155–162).
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Boase-Beier, J. (2023). Stylistics and translation. In *The Routledge handbook of stylistics* (pp. 420–435). Routledge.
- Cai, Hui (蔡辉), & Yin, Xing (尹星). (2005). 西方幽默理论研究综述 [A review of Western humor theories]. *外语研究 [Foreign Language Research]*, (1), 5–8.
- Cela, H., Veit, S. V., & Wood, G. (2024). Breaking down the laughter: An exploration into the linguistic dimensions in stand-up comedy ratings. *BMC Psychology*, 12(1), 679.
- Fan, Xiaochao (樊小超), Yang, Liang (杨亮), Lin, Hongfei (林鸿飞), Diao, Yufeng (刁宇峰), Shen, Chen (申晨), Chu, Yonghe (楚永贺), & Zhang, Tongxuan (张桐瑄). (2021). 基于多维潜在语义特征的幽默识别 [Humor recognition based on multidimensional latent semantic features]. *中文信息学报 [Journal of Chinese Information Processing]*, 35(8), 38–46.
- Ge, Lingling (戈玲玲), & He, Yuanjian (何元建). (2012). 基于言语幽默概论的汉语幽默文本特征研究 [A study on features of Chinese humorous texts based on verbal humor theory]. *外国语 (上海外国语大学学报) [Journal of Foreign Languages]*, 35(4), 35–44.
- H. P. Grice (1989). *Studies in the way of words*. Harvard University Press.
- Hao, Y., Chen, Y., Wang, X., & Liu, H. (2025). Metricizing diaspora dualities: a comparative quantitative analysis of poetry and prose in Southeast Asian Chinese literature. *Digital Scholarship in the Humanities*, fqaf094.
- Hoover, David L. (2007). Corpus stylistics, stylometry, and the styles of Henry James. *Style*, 41(2), 174–203.
- Hu, Fanzhu (胡范铸). (1987). 幽默语言学 [Humor linguistics]. 上海社会科学院出版社 [Shanghai Academy of Social Sciences Press].
- Hu, Xianyao (胡显耀). (2021). 语料库文体统计学方法与应用 [Corpus stylistic statistics: Methods and applications]. 外语教学与研究出版社 [Foreign Language Teaching and Research Press].
- Huang, Leping (黄乐平), Wang, Junchao (王俊超), & Liu, Yang (刘洋). (2022). 中国相声幽默英译理念探赜: 传统译论“信达切”原则拓展新视角 [On the translation of Chinese crosstalk humor: Expanding the traditional principle of “faithfulness, expressiveness, and elegance”]. *外语学刊 [Foreign Language Journal]*, (5), 67–72. <https://doi.org/10.16263/j.cnki.23-1071/h.2022.05.010>
- Huang, Ying (黄影). (2023). 基于汉英对比的言语幽默的多视角研究 [A multi-perspective study of verbal humor based on Chinese-English comparison], 河北大学博士学位论文 [Unpublished Doctoral Dissertation, Hebei University].
- Kalloniatis, A., & Adamidis, P. (2024). Computational humor recognition: A systematic literature review. *Artificial Intelligence Review*, 58(2), 43.
- Kao, J. T., Levy, R., & Goodman, N. D. (2016). A computational model of linguistic humor in puns. *Cognitive Science*, 40(5), 1270–1285.
- Lefcourt, H. M. (2012). *Humor: The psychology of living buoyantly*. Springer.
- Li, Xianjin (李先进). (2013). 基于语料库的汉语幽默文本特征研究 [A corpus-based study of Chinese humorous text features]. *外语学刊 [Foreign Language Journal]*, (6), 47–50.
- Li, Yunhan (黎运汉). (2002). 1949年以来语言风格定义研究述评 [A review of studies on language style definitions since 1949]. *语言文字应用 [Applied Linguistics]*, (1), 100–106.

- Li, Yunhan (黎运汉), & Sheng, Yongsheng (盛永生). (2010). 汉语修辞学 (修订版) [Chinese rhetoric (revised edition)]. 广东教育出版社 [Guangdong Education Press].
- Li, Jinxi (黎锦熙). (1958). 新著国语文法 [New Chinese grammar]. 商务印书馆 [Commercial Press].
- Lin, Hongfei (林鸿飞), Tuonike Tuersun (吐妮可·吐尔逊), & Zhang, Dongyu (张冬瑜). (2024). 面向社交媒体的幽默计算理论与实践 [Computational humor theory and practice for social media]. 外国语文 [Foreign Languages and Literature], 40(4), 45–59.
- Lin, Hongfei (林鸿飞), Zhang, Dongyu (张冬瑜), Yang, Liang (杨亮), & Xu, Bo (徐博). (2016). 幽默计算及其应用研究 [Computational humor and its applications]. 山东大学学报 (理学版) [Journal of Shandong University (Natural Science)], 51(7), 1–10.
- Liu, Naishi (刘乃实), & Zhang, Renxian (张韧弦). (2016). 现代幽默语言学 [Modern humor linguistics]. 外语教学与研究出版社 [Foreign Language Teaching and Research Press].
- Luo, Qun (罗群). (2021). 脱口秀会取代相声吗? [Will stand-up comedy replace crosstalk?]. 中国文化报 [China Culture Daily], (004).
- Ma, Ji (马季). (1980). 相声艺术漫谈 [On the art of crosstalk]. 广东人民出版社 [Guangdong People's Publishing House].
- McIntyre, D., & Walker, B. (2019). *Corpus stylistics*. Edinburgh University Press.
- Meaney, J. A., Wilson, S., Chiruzzo, L., Lopez, A., & Magdy, W. (2021). SemEval 2021 task 7: HaHackathon, detecting and rating humor and offense. In *Proceedings of the 15th International Workshop on Semantic Evaluation* (pp. 105–119).
- Mihalcea, R., & Strapparava, C. (2006). Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence*, 22(2), 126–142.
- Miller, T., Hempelmann, C. F., & Gurevych, I. (2017). SemEval-2017 task 7: Detection and interpretation of English puns. In *Proceedings of the 11th International Workshop on Semantic Evaluation* (pp. 58–68).
- Morales, Michelle A., & Zhai, Chengxiang. (2017). Identifying humor in reviews using background text sources. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 492–501).
- Mosteller, F., & Wallace, D. L. (1964). *Applied Bayesian and classical inference: The case of the Federalist Papers*. Addison-Wesley.
- Partington, A. (2006). *The linguistics of laughter: A corpus-assisted study of laughter-talk*. Routledge.
- Potash, P., Romanov, A., & Rumshisky, A. (2017). SemEval-2017 task 6: #HashtagWars: Learning a sense of humor. In *Proceedings of the 11th International Workshop on Semantic Evaluation* (pp. 49–57).
- Raskin, V. (1985). *Semantic mechanisms of humor*. Springer Netherlands.
- Reyes, A., Rosso, P., & Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74, 1–12.
- Ritchie, G. (2001). Current directions in computational humour. *Artificial Intelligence Review*, 16(2), 119–135.
- Skalicky, S., Berger, C. M., Crossley, S. A., & McNamara, D. S. (2016). Linguistic features of humor in academic writing. *Advances in Language and Literary Studies*, 7(3), 248–259.
- Skalicky, S., Duran, N., & Crossley, S. A. (2020). Please, please, just tell me: The linguistic features of humorous deception. *Dialogue & Discourse*, 11(2), 128–149.
- Stamatatos, E., Fakotakis, N., & Kokkinakis, G. (2001). Computer-based authorship attribution without lexical measures. *Computers and the Humanities*, 35(2), 193–214.
- Vandaele, J. (2010). Humor in translation. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies* (Vol. 1, pp. 147–152). John Benjamins.

- Wang, Jinling (王金玲). (2002). 论幽默语言的特征与技巧 [On the features and techniques of humorous language]. 外语学刊 [Foreign Language Journal], (3), 58–63.
- Xiao, Fuxing (肖复兴). (2009). 相声的讽刺与幽默 [Satire and humor in crosstalk]. 艺术评论 [Art Review], (11), 15–17.
- Xu, Linhong (徐琳宏), Lin, Hongfei (林鸿飞), & Zhao, Jing (赵晶). (2008). 情感语料库的构建和分析 [Construction and analysis of an emotional corpus]. 中文信息学报 [Journal of Chinese Information Processing], (01), 116–122.
- Xu, Ming (许明), & Jiang, Yue (蒋跃). (2020). 《阿Q正传》译入译出文本的风格计量学对比 [A stylometric comparison of translated and back-translated texts of *The True Story of Ah Q*]. 外语研究 [Foreign Language Research], 37(3), 86–92+103. <https://doi.org/10.13978/j.cnki.wyj.2020.03.014>
- Yan, Qing (晏青), & Wan, Xuting (万旭婷). (2024). 异项美感狂欢：脱口秀节目的幽默实践及青年情感结构反思 [Carnival of heterogeneous aesthetics: Humor practices in stand-up comedy and reflections on youth emotional structure]. 新闻界 [Journalism Review], (7), 49–62.
- Zeng, Yiping (曾毅平), & Zhu, Xiaowen (朱晓文). (2006). 计算方法在汉语风格学研究中的应用 [Application of computational methods in Chinese stylistics]. 福建师范大学学报（哲学社会科学版） [Journal of Fujian Normal University (Philosophy and Social Sciences)], (1), 14–17.
- Zeng, Z., Li, Z., Yang, L., & Lin, H. (2024). Leveraging social context for humor recognition and sense of humor evaluation in social media with a new Chinese humor corpus—HumorWB. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 10393–10402).
- Zhu, Pengcheng (祝鹏程). (2024). 谁之传统，传统何为：对当代北京相声传统化实践的考察及反思 [Whose tradition and what tradition: An examination and reflection on the traditionalization practices of contemporary Beijing crosstalk]. 民俗研究 [Folklore Studies], (3), 136–147+160.

投稿：2025年12月31日；接受：2026年3月8日；出版：2026年4月15日

## 作者简介

陈永胜，华侨大学华文学院研究生。研究方向：海外华语和计量语言学。

王雪琳，暨南大学华文学院在读博士研究生。研究方向：语言风格、计量语言学、二语习得。

## **A Multidimensional Quantitative Study of the Linguistic Features of Chinese Humorous Style: A Case Study of *Xiangsheng* Texts**

**Yongsheng Chen**

Huaqiao University, China

**Xuelin Wang**

Jinan University, China

### **Abstract**

Humor, as an important form of linguistic style, has long been a major topic in natural language processing, especially in relation to its generative mechanisms and linguistic features. In view of this, the present study adopts a quantitative stylistic approach to examine the linguistic features of Chinese humorous style, using *xiangsheng* texts as representatives of humorous discourse and television interview texts as a non-humorous reference group. The results show that: (1) compared with non-humorous spoken texts, spoken texts with a humorous style in Chinese use significantly lower proportions of prepositions and conjunctions, but significantly higher proportions of adjectives and adverbs; (2) at the semantic level, humorous texts display a higher degree of semantic incongruity, suggesting that they rely on an “expectation-shift” structure to produce humorous effects; and (3) in terms of emotional expression, Chinese humorous style is mainly characterized by an interplay of neutral emotion and a certain proportion of negative emotion, whereas non-humorous texts show a clearer tendency toward positive emotion overall. These findings help reveal the linguistic features and generative mechanisms of Chinese humorous texts, and also provide an operational and testable quantitative approach to the study of Chinese humorous style.

### **Keywords**

Chinese humor, computational humor, linguistic features, language style, data-driven

*Yongsheng Chen* is a graduate student at the College of Chinese Language and Culture, Huaqiao University. His research interests include overseas Chinese language style and quantitative linguistics.

*Xuelin Wang* is the corresponding author and a doctoral candidate at the College of Chinese Language and Culture, Jinan University. Her research interests include language style, quantitative linguistics, and second language acquisition.