

多模态视角下社交媒体视觉元话语的机制研究——以小红书平台为例

姜明汐

辽宁师范大学国际商学院 辽宁大连

【摘要】本文基于元话语理论及多模态话语分析视角，以小红书平台为例，探究社交媒体中视觉元话语的构成、功能与身份建构机制。社交媒体具备即时性、互动性与多模态性，为用户借助视觉资源实施语篇组织、意图调控与身份塑造提供了丰富场域。研究借鉴 De Groot 等（2016）的视觉元话语分类模型，结合平台特性，将视觉元话语划分为图片-文字类与视频-文字类两大类型，并进一步细分为布局类与信息类等子类，系统分析其模态特征及话语实现方式。研究发现，视觉元话语通过构图、注释、态度标记、自我提及、介入标记等策略，不仅优化信息传递与认知引导，更在互动中动态建构用户身份，适应多元社交情境。本研究为理解数字时代的多模态交际与身份实践提供参考，并为后续跨文化、跨平台视觉元话语比较研究奠定基础。

【关键词】社交媒体；视觉元话语；多模态话语分析；小红书

【收稿日期】2025年12月12日 **【出刊日期】**2026年1月7日 **【DOI】**10.12208/j.ssr.20260028

A study on the mechanism of visual meta-discourse on social media from a multimodal perspective: a case

study of Xiaohongshu platform

Mingxi Jiang

College of International Business, Liaoning Normal University, Dalian, Liaoning

【Abstract】From the perspective of metadiscourse theory and multimodal discourse analysis, this paper examines the composition, functions, and identity construction mechanisms of visual metadiscourse in social media, taking Xiaohongshu as a case study. Characterized by immediacy, interactivity, and multimodality, social media offers a fertile ground for users to employ visual resources in discourse organization, intentional regulation, and identity shaping. Drawing on De Groot et al.'s (2016) classification of visual metadiscourse and adapting it to platform-specific features, this study categorizes visual metadiscourse into image-text and video-text types, further subdivided into layout-oriented and information-oriented subcategories. It systematically analyzes their modal features and discursive realizations. The findings reveal that visual metadiscourse, through strategies such as composition, annotation, attitude markers, self-mention, and engagement markers, not only enhances information delivery and cognitive guidance but also dynamically constructs user identities in interaction, adapting to diverse social contexts. This study provides theoretical insights into multimodal communication and identity practices in the digital age and lays a foundation for future cross-cultural and cross-platform comparative research on visual metadiscourse.

【Keywords】Social media; Visual metadiscourse; Multimodal discourse analysis; Xiaohongshu

1 引言

元话语（metadiscourse）是构建语篇结构、调控人际互动的重要话语策略^[1]，其研究范式正经历由单模态向多模态的深刻转向。早期研究多集中于线性文本，尤其是学术语篇中的语言符号系统^[2]。随着数字媒介技术的发展，元话语的研究视野已扩展至多模态交际领域^[3]。在多模态语境下，文字与视觉元话语形成动态互补的

符号生态系统^[4]，共同构建网络空间的表意实践。社交媒体用户通过评论、话题标签、表情符号等多种视觉资源，不仅实现传统意义上的语篇组织功能，还执行复杂的人际定位与情感调节策略。因此，探究社交媒体中的视觉元话语，不仅有助于揭示数字时代人际互动的语用机制，更能深化对网络身份认同的理解，解析用户如何通过多元视觉策略在虚拟空间中构建差异化的自我

呈现（self-presentation）^[5]。本研究以小红书平台为分析对象，依据 De Groot 等（2016）提出的视觉元话语理论模型，结合平台特有的多模态表征形式，构建适用于社交媒体的视觉元话语分析框架，系统考察其模态构成、交互功能与身份建构机制所呈现的特征。

2 视觉元话语相关理论概述

元话语（metadiscourse）理论由 Zellig Harris（1959）在语用学研究中提出，为理解语言交际行为提供了新的分析视角^[6]。随后，Fowler（1987）从修辞学角度将其界定为一种重要的话语策略^[7]。在此基础上，Ken Hyland（2000）进一步系统阐述了元话语在构建语篇连贯性、传递作者立场等方面发挥的重要作用^[8]。Hyland&Tse（2004）提出的引导式元话语（interactive metadiscourse）和交互式元话语（interactional metadiscourse）二分法^[9]，为后续研究奠定了基础。随着多模态话语分析理论的兴起（Kress&Van Leeuwen, 2006），De Groot（2016）等学者将上述分类拓展至视觉领域，提出构图引导式视觉元话语（interactive visual metadiscourse）和人际交互式视觉元话语（interactional visual metadiscourse），强调通过符号元素创造受众身临其境的参与体验^[10,11]。然而，该框架主要基于静态图像（如企业年报），未能充分涵盖社交媒体中广泛使用的话题标签、表情符号等动态、交互性强的多模态资源，也未适应社交平台高度互动、实时反馈的语境特征。

相较于文字元话语，视觉元话语具有直观性强、跨文化认知门槛低的优势，更易于在全球化传播中被快速识别与接受。为适应社交媒体平台的多模态特性，本研究在 De Groot 等人分类的基础上，结合小红书平台的内容形态，将视觉元话语进一步区分为图片-文字类

视觉元话语与视频-文字类视觉元话语两大类，细化出布局类与信息类等子类，构建一个更贴合社交媒体语境的多模态元话语分析框架，通过目的性抽样的研究方法，系统探究其形式特征及话语实现方式。

3 图片-文字类视觉元话语

图片-文字类视觉元话语指创作者在语篇构建的过程中，通过整合文字、图像、表情符号、话题标签等多种符号资源的协同作用，实现对信息的可视化组织与意义构建。这类元话语不仅能直观呈现语篇信息与结构，更能有效引导受众的认知路径。基于 Kress&Van Leeuwen（2006）的视觉语法理论和 De Groot（2016）的多模态话语分类框架，本文将图片-文字类视觉元话语划分为布局类元话语和信息类元话语两大维度（见表 1）。二者在模态特征与语篇功能上既存在差异，又形成互补。

布局类元话语以图像模态为主导，其视觉表征包括：

- (1) 图像元素占据主导视觉空间；
- (2) 文字起辅助说明作用；
- (3) 强调通过视觉逻辑组织信息；
- (4) 依赖空间布局实现意义传达。

信息类元话语则以文字模态为核心，表现为：

- (1) 文字内容构成信息主体；
- (2) 图像仅作补充说明；
- (3) 注重语言文字的逻辑性；
- (4) 以线性叙述为主要表达方式。

这种模态差异反映了两者在信息编码策略上的本质区别：前者通过视觉语法实现意义建构，后者则依靠语言系统完成信息传递。

表 1 图片-文字类视觉元话语分类

分类	功能	表现形式
布局类	构图 composition	通过视觉设计引导受众视线路径 元素位置；字号尺寸；图片色彩；形态；对比度；清晰度
	视觉过渡标记 visual transition marker	运用符号或图像逻辑关系组织信息 表格；要点列表；流程图；树状图；思维导图
	视觉框架标记 visual frame marker	使用符号分割、连接语篇不同部分 线条；边框；方框
信息类	视觉内指标记 visual endophoric marker	引导受众定位语篇内其他信息 互动性引导语；数字化编码；符号引导
	视觉证源标记 visual evidential	标记外部信息来源 超链接；“@”符号
	视觉注释标记 visual code gloss	提供额外补充信息或解释 “#”等话题标签；图像解释文本信息
	视觉态度标记 visual attitude marker	传递创作者情感态度 表情；面部表情；手势
	视觉自我提及 visual self mention	通过视觉形象进行自我呈现 个人照片；企业标识；产品形象图
	视觉介入标记 visual engagement marker	将受众转化为积极参与者 开放式提问；引导型动作

3.1 布局类元话语包括构图、视觉过渡标记和视觉框架标记三个部分，其核心是通过视觉元素的有机组织来优化信息传递效率，实现传播效果的最大化。

(1) 构图作为视觉传达的核心机制，通过整合信息值、显著性和取景三大维度 (Kress&Van Leeuwen, 2006)，实现对图像元素的整体布局。在实践中，创作者依托图像空间的信息承载梯度，常将核心元素置于视觉中心以获取最大关注，次要信息依重要性层级呈放射状分布，形成从中心到边缘、自上而下、从左至右的信息衰减序列。同时，通过调控元素尺寸、色彩、形态、对比度等视觉变量，强化核心焦点，弱化次要干扰，从而引导受众注意力分配。例如在图 1 中，创作者将核心产品置于画面视觉锚点 (visual anchor) 位置，利用中心布局确保其获得首要关注；次要产品按优先级排列，形成以焦点为中心的信息同心圆结构^[12]，并对次要产品进行模糊处理。文本层面对关键话术放大字号、运用色彩，通过多模态协同实现产品信息与品牌叙事的有效结合。

(2) 视觉过渡标记指通过符号系统与视觉表征的

内在逻辑关联，对信息进行结构化重组与可视化呈现的元话语策略。其主要形式包括表格、要点列表、流程图等。这类标记通过空间布局与视觉逻辑的结合，实现文本信息的去冗存精，突出核心特征，优化认知路径，降低信息处理负荷，提升信息获取效率。

如图 2 所示，创作者采用对比表格直观呈现不同品牌洗衣液的产品特性，通过空间对应关系强化比较维度，使受众能够快速捕捉差异特征，显著提升信息的可读性与决策效率。

(3) 视觉框架标记指运用线条、边框等视觉分割元素对图文内容进行区块化处理，使不同信息单元保持相对独立性。这种标记通过统一的构图原则、版式布局及字体系统，实现分散区块的内容整合与视觉统一。

如图 3 案例所示，创作者采用直线分割将画面划分为 9 个规整单元，每个单元遵循相同视觉逻辑：左下象限集中呈现产品实物及功能，右上象限展示产品名称与说明。这种标准化网格布局设计通过重复的视觉节奏强化整体性，模块化设计提升信息获取效率，统一的框架引导视线流动，实现重点信息的快速识别。



图 1 构图示例



图 2 视觉过渡标记示例



图 3 视觉框架标记示例

3.2 信息类元话语包括视觉内指标记、证源标记、注释标记、态度标记、自我提及及介入标记六部分。与布局类元话语的空间组织功能不同，信息类元话语是通过信息流标记实现语篇结构的动态引导，其核心特征在于以文字内容为主导载体，视觉元素承担辅助说明功能。

(1) 视觉内指标记指创作者在同一语篇中使用特定标记引导受众注意力指向目标位置的策略。在小红书平台中，创作者倾向规避“如图所示”“如视频所示”等传统直白标记，转而采用“快看”“带你看”等互动性引导语，或在文本中标注图像对应页码（如“p2-7”），实现跨模态的精确信息定位（图4）。这种策略通过口语化表达增强亲和力，利用数字化编码实现精确定位。

(2) 视觉证源标记通常以超链接、“@”符号等形式呈现，主要功能是建立语篇与外部信息源的关联回路。在商业推广场景中，此类标记多用于单一产品或是特定商家的推介，通过“文字描述+超链接”形成营销闭环（图5）。在社交互动场景下，则承担

用户间的指向性交流，通过“@”提及或链接实现信息源定位与社交互动连接（图6）。

(3) 视觉注释标记通过图像符号系统对文字内容进行语义补充与信息重构。一是对文字信息进行具象化阐释，如图7，标题“开箱世界最贵垃圾袋”配以直观视觉呈现，弥补文字描述的局限；二是借助“#”等话题标签实现文本信息的符号化概括，如图8中，“清洁好物”“全屋清洁”等标签凝练主题，构建内容检索的语义节点。

(4) 视觉态度标记是创作者在语篇中表达态度的视觉化呈现方式，可分为显性与隐性两类。显性视觉态度标记通过直观视觉元素（如人物神态、动作表情）直接传达态度，以图9为例，戒指作为核心元素结合虚化背景，营造喜悦氛围，辅以表情符号强化情感传递；隐性视觉态度标记则需要受众结合视觉元素与文字信息进行推理判断，可能反语等修辞手法，如图10标题通过表面矛盾吸引注意，引导深入解读真实态度。

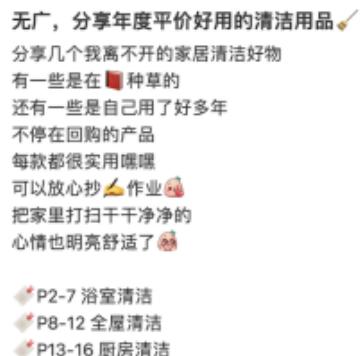


图4 视觉内指标示例

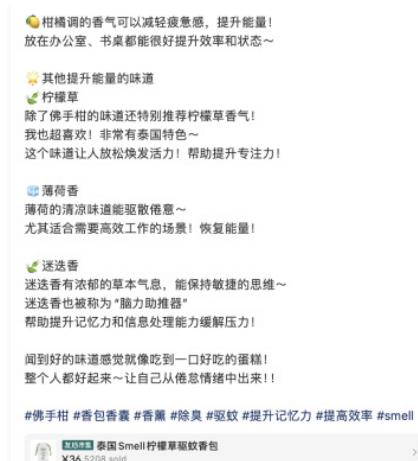


图5 视觉证源标记示例 1

复刻简简厨房纽约大曲奇【空气炸锅版】

一次就成功啦 感谢这个配方 @简简厨房
当然我有根据自己的口味做了细微的调整~

糖：白砂糖 40g (我用了 30g 区别不大) 红糖 20g

油：黄油 100g 室温软化

粉：低筋面粉 150g 泡打粉 3g 盐 2g

鸡蛋：一颗打散

图 6 视觉证源标记示例 2

▣ 开箱世界最贵垃圾袋

如果你和我一样是在苏黎世生活，读书，旅游，就会发现一个非常奇怪的情况：

所有的垃圾桶开口都极小，稍微大一点的垃圾都扔不进去

原因是政府不希望你把大垃圾直接扔在路边垃圾桶

比如我的生活垃圾，必须购买这种政府认可包含了垃圾清理费的垃圾袋，

当然也要扔在宿舍指定的垃圾回收处

10个140人民币，每天心都在滴血

来这里长期生活的要慎重啊 😱😱😱

图 7 视觉注释标记示例 1

洁癖星人无限回购的 9 件清洁神器 🍋 好物盘点

拒绝清洁焦虑、精简清洁 ✅ 只留最需要的，家里一样可以做到长久如新

+

整理了一些一直跟着我的清洁好物 🌸

很多都是可以一物多用，真正做到精简清洁

希望大家都可以保持家里干净的风水 😊

迎接美好的生活

#清洁好物 #全屋清洁 家清好物正当时 家清好物正当时 #无限回购好物 #
清洁神器 #宅家大扫除 #买了不后悔的家居神器 #浴室清洁 #清洁小妙招

图 8 视觉注释标记示例 2



图 9 显性视觉态度标记示例

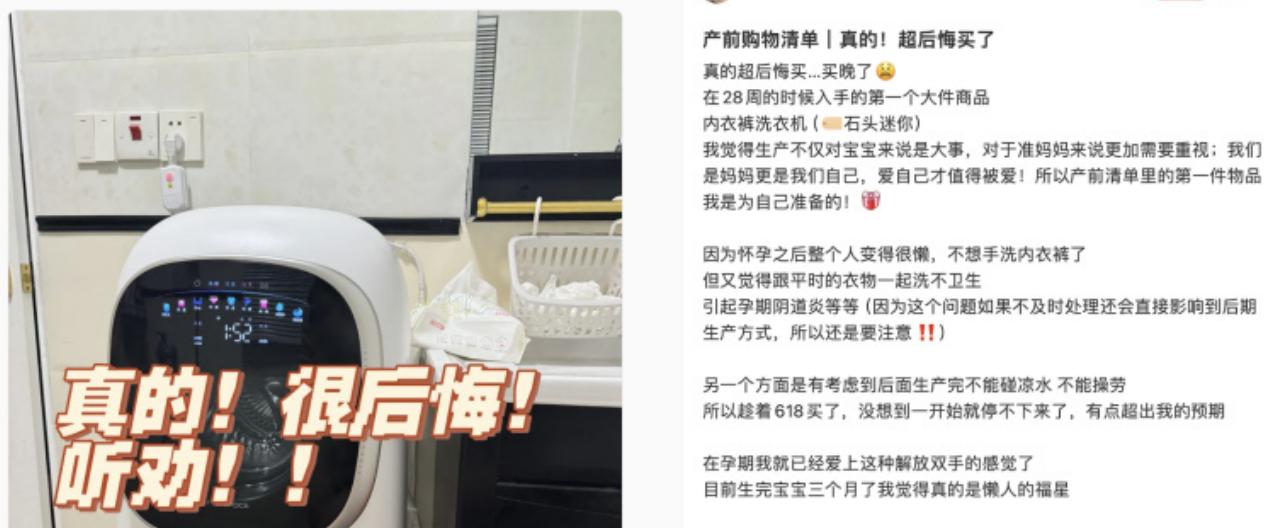


图 10 隐性视觉态度标记示例



图 11 视觉自我提及示例



图 12 显性视觉介入标记示例



图 13 隐性视觉介入标记示例

(5) 视觉自我提及指创作者通过视觉形象（如个人照片、企业标识）在语篇中进行自我呈现的策略，具有明确身份、强化认知记忆的双重功能。从指代对象来看，视觉自我提及既适用于个人创作者，也可用于企业或机构等集体主体。如图 11 中，演员刘奕君发布个人形象照，直观强化公众人物形象，避免冗长文字描述。类似的，企业账号也常使用品牌 logo 或产品形象作为视觉自我提及标记，以增强品牌辨识度。

(6) 视觉介入标记作为实现受众互动最直接的视觉修辞手段，旨在将受众转化为积极参与者，可分为显性与隐性两类。显性视觉介入标记采用直接互动形式，如图 12 中，通过开放式提问“哪款更百搭”并嵌入投票链接，构建对话情境，将虚拟互动转化为具象参与行为；隐性视觉介入标记则依托非语言符号实现互动引导，如图 13 通过构图与肢体语言构建“邀请者-受邀者”关系，以视觉隐喻唤起受众参与意识，实现从观看到参与的认知转换。

4 视频-文字类视觉元话语

视频-文字类视觉元话语是数字媒介时代视频内容的核心叙事策略。通过视频、文字、声音、界面等多模态形式的有机融合，在创作者与受众之间搭建起认知桥梁。其特征体现为第一人称视角带来的沉浸式体验与文字元素辅助的认知引导，帮助受众在碎片化信息环境中快速把握视频核心。

4.1 视频内容的叙事策略

(1) 主观镜头的沉浸式体验：创作者运用手持拍

摄和 POV（第一人称视角）镜头语言，结合视线引导（包括眼神交流、手势指示）和画面重点标注（箭头指示、高亮标记），营造强烈参与感。面部微表情、语音语调等非语言元素进一步强化叙事的亲和力与真实感，在 vlog 等互动型视频中效果显著。

(2) 人称转换的修辞艺术：成熟创作者善于灵活转化人称以优化叙事效果。采用“我-你”直接对话形式（如教程类视频中的即时指导），建立强烈互动联结；使用“我们”等包容性称谓，培养受众群体归属感，构建想象的观看共同体；适时穿插第三人称客观叙述（如画外音解说），则增强内容的专业性与可信度。

4.2 文字元素的元话语功能

在视频创作中，文字元素已从传统信息辅助工具演变为构建受众认知体系的核心媒介。通过系统化设计调控注意力流向，强化信息传递效率。

(1) 标题与封面的认知引导作用：视频标题和封面承担认知预设功能，常采用悬念式标题（如“99%的人都不知道这个技巧”）激发探索欲望。视觉呈现上使用大字号醒目标语配合高对比色彩，确保短时间内传递核心信息，优化内容吸引力，提升在信息流中的曝光权重。

(2) 字幕的动态表达功能：字幕超越单纯文字说明，通过动态设计（字号缩放、色彩变换或动态效果等）帮助受众在信息过载环境中建立内容优先级别。自动字幕生成过程中可能出现的识别错误，有时会意外形成幽默效果或赋予视频新的解读空间。这种非刻意为

之的误差反而可能成为视频的传播助力。

(3) 界面文字的交互设计：视频平台界面文字系统构建完整的交互引导体系：进度条标记（如“00: 30 重点来了！”）实现观众注意力精准调度；行为召唤按钮（如“解锁下一集”），运用视觉证源标记提升转化率；话题标签（如“#夏日运动挑战”“#简单生活技巧”）则是通过视觉注释标记实现内容概括。

4.3 视频与文字的协同叙事

视频中的视觉内容与文字通过跨模态语义整合实现更丰富的表达效果，主要表现为两种形式：

(1) 互补关系：视觉展示过程，文字提炼结论。在教程类视频中，动态画面展示操作流程的同时，文字标注同步完成信息抽象化处理，二者形成认知闭环。

(2) 冲突关系：创作者通过反讽表达制造批判性张力，或运用悬念设置构建叙事期待，刻意制造的图文差异能产生独特的修辞效果。

5 结论

本研究通过分析小红书平台的视觉元话语的表征体系，基于 De Groot 等学者提出的多模态话语分类框架，构建图片-文字类视觉元话语与视频-文字类视觉元话语的分析框架。研究发现，相较于受制于语言壁垒的纯文字元话语，视觉元话语凭借其直观性与跨文化性，已成为数字时代叙事的基础语法架构，显著提升信息传递效率。创作者通过巧妙的视觉构图策略、文字辅助注解和多模态协同机制，不断优化内容表达，建构具有文化适应性的传播路径。

具体而言，构图、过渡标记、框架标记等布局类策略与内指标记、证源标记、注释标记、态度标记、自我提及和介入标记等信息类资源，共同构成完整的视觉元话语体系，实现语篇组织、人机互动的多重功能。随着 AI 技术持续发展，视觉元话语表达形态正经历从静态表达向动态交互的演变升级，但其核心价值始终聚焦于提升传播精度——这为数字时代的跨文化传播研究提供了新视角。

未来研究可进一步关注视觉元话语在不同文化背景下的适应性，及其在新兴技术环境中的应用特征与

发展趋势，以深化对多模态交际机制的理解。

参考文献

- [1] HYLAND K. *Metadiscourse: Exploring Interaction in Writing*[M]. Continuum, 2005.
- [2] 付晓丽, 徐赳赳. 国际元话语研究新进展[J]. 当代语言学, 2012.
- [3] 姜峰, HYLAND K. 互动元话语：学术语境变迁中的论辩与修辞[J]. 外语教学, 2020.
- [4] 冯德正. 多模态语篇分析的基本问题探讨[J]. 北京第二外国语学院学报, 2017.
- [5] GOFFMAN E. *The presentation of Self in Everyday Life*[J]. New York: Doubleday, 1959.
- [6] HARRIS Z. *An Introduction: Metadiscourse*[M]. London Press, 1959.
- [7] FOWLER R. *Linguistic Criticism*[M]. Oxford University Press, 1987.
- [8] HYLAND K. *Disciplinary Discourses: Social Interactions in Academic Writing*[M]. London: Longman, 2000.
- [9] HYLAND K, TSE P. Metadiscourse in Academic Writing: A Reappraisal[J]. *Applied Linguistics*, 2004.
- [10] KRESS G, VAN LEEUWEN T. *Reading Images: The Grammar of Visual Design*[M]. New York: Routledge, 2006.
- [11] DE GROOT E, NICKERSON C, KORZILIUS H, GERRITSEN M. Picture This: Developing a Model for the Analysis of Visual Metadiscourse[J]. *Journal of Business and Technical*, 2016.
- [12] BERTIN J. *Semiology of Graphics*[M]. Conference on Computer Networks, 1983

版权声明：©2026 作者与开放获取期刊研究中心（OAJRC）所有。本文章按照知识共享署名许可条款发表。

<http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS