

Original Research Article

A study on the visual communication mechanism of Chinese literary works on short video platforms

Xiaolin Sun

Guangxi University, Nanning, Guangxi, 530004, China

Abstract: On short-video platforms, literary characters, plots, images, and emotions are recast through footage, on-screen text, sound, and interactive prompts. Drawing on publicly available empirical studies of BookTok, digital reading, and short-video cultural communication in China published during the past five years, this article develops a five-stage framework: Literary semantic retention, visual re-narration, multimodal coordination, platform visibility, and audience participation. Rather than pooling incompatible datasets, it compares findings across studies. The evidence suggests that brief formats, music, and recognizable visual motifs help a video secure initial attention. References to a work's title, characters, relationships, key images, and context help viewers identify and interpret the source text, while creator credibility and community interaction are more closely associated with discussion, saving, sharing, and intention to read. Semantic retention and coherence across modes therefore matter if fleeting visual attention is to develop into sustained literary engagement. The framework separates reach metrics from indicators of substantive engagement and offers a basis for assessing the quality, rather than only the scale, of literary communication through short video.

Keywords: short-video platforms; literary works; visual re-narration; multimodal communication; BookTok

1. Introduction

Mobile short-video platforms have moved literary communication into a media environment shaped by images, captions, sound, performance, and comment threads. In Merga's analysis of 116 BookTok videos, recommendations, accounts of reading experiences, emotional responses, and reader community were the principal themes^[1]. Dezuanni et al. showed how selfies, shelf displays, and individualized forms of address helped mediate young people's reading culture^[2]. Interviews conducted by Guíñez-Cabrera and Mansilla-Obando indicate that rapid content cycles, algorithmic recommendation, and shared reading interests sustain both production and community ties^[3]. Audience research suggests that BookTok appeals especially to frequent leisure readers, while also reaching less frequent readers who already hold positive attitudes toward reading^[4]. Maddox and Gill likewise argue that hashtags, trending audio, search prompts, and users' practices jointly define what becomes visible as a literary community^[5]. Young readers use BookTok to support voluntary reading and construct reader identities^[6], and its language, camera work, and interaction patterns can support peer learning^[7]. Reviewing this literature, Dera identifies three main strands - review formats, community formation, and reading promotion - and calls for stronger empirical designs^[8]. Yet these studies seldom explain, within a single account, how literary meaning, visual organization, multimodal relations, and platform distribution interact. This article addresses that gap by asking how a literary text becomes a visual point of entry and how that encounter may lead to interpretation, discussion, or reading. Chinese literature makes the problem particularly visible: The concentrated imagery of classical poetry, the complex character networks of traditional chapter-based novels, and the historically situated meanings of modern writing all resist simple visual compression.

2. A visual communication model for literary works

2.1. Defining the process and the overall framework

In this study, visualizing a literary work does not mean merely illustrating it. It refers to the creator's

selection of people, events, settings, motifs, and emotional cues from the source text and their rearrangement through footage, captions, recitation, music, performance, or drawn images. The proposed process moves from retaining recognizable literary meaning, through visual re-narration and coordination across modes, to platform visibility and, finally, audience response.

2.2. Modeling semantic retention and visual re-narration

To assess semantic retention, we ask whether a video keeps five kinds of information legible: Relations among characters, pivotal events, recurring or representative images, emotional direction, and thematic language. We then classify the dominant method of visual re-narration as text-led, illustrative, scene-based, commentary-led, or mixed. The relative weight of these elements varies by genre. A poem may become recognizable through motifs such as the moon, a boat, or a pavilion, whereas a novel usually requires characters and a plot connection. A title, a named character, a short quotation, or a brief note on context can prevent the image from becoming detached from its literary source. In coding the material, claims that can be checked against the original work serve as the baseline. This makes it possible to distinguish a direct representation from visual metonymy and, in turn, from decorative imagery that has only a loose relation to the text.

2.3. Multimodal coupling and the reinforcement of visual motifs

Multimodal coupling concerns whether captions or narration, images, and sound tell a compatible story. Words carry plot or commentary; images give visible form to characters, objects, and settings; sound establishes pace and mood. Even in a very short video, these modes can leave viewers with a coherent line of interpretation when they converge on the same theme. By contrast, striking imagery or trending music may attract attention while pulling the video away from the source text. Visual motifs are therefore most useful when they grow out of recurring images, characters, or places in the work itself. Music and color count as reinforcement only when their mood supports, rather than overwrites, the emotional direction of the text. A pleasing surface is not, by itself, evidence of literary content.

2.4. Platform circulation and audience response

For analytical purposes, platform circulation is separated from audience engagement. Plays, likes, and short dwell times mainly record initial exposure. Comments, saves, shares, plans to read, and derivative creations point more clearly to continuing involvement. Strong visuals may secure visibility, but semantic retention and agreement among modes make the video easier to interpret. Trust in the creator and interaction within the community may then help move viewers from watching to discussing the work or reading it.

3. Comparative results

3.1. Data sources and comparative procedure

We compare findings from published studies rather than pool samples that differ in design and population. Rovira-Collado et al. analyzed 299 Spanish-language BookTok videos^[9]. Da Silva et al. surveyed 663 followers of literary content to examine how BookTokers influence reading intentions^[10]. Xu et al. coded 1,642 short videos about traditional Chinese painting on Douyin and Xiaohongshu and compared their visual framing and circulation^[11]. **Table 1** summarizes these evidence blocks. Each finding is interpreted only within the limits of its own study and then related to one of five shared dimensions: Semantic Retention, Visual Re-narration, Multimodal Coupling, Platform Visibility, or Audience Participation. Differences in sample size and method remain part of the comparison rather than being treated as interchangeable evidence.

Table 1. Empirical studies used to examine the proposed mechanism.

Study	Platform or corpus	Sample	Method	Evidence focus
Merga (2021) ^[1]	TikTok / BookTok	116 videos	Content analysis	Recommendations and community
Maddox & Gill (2023) ^[5]	TikTok / #BookTok	150 videos	Textual analysis	Interface and visibility
Rovira-Collado et al. (2025) ^[9]	TikTok / BookTok	299 videos	Content analysis	Visual and audio features
da Silva et al. (2025) ^[10]	Cross-platform	663 respondents	Survey model	Reading intention
Xu et al. (2026) ^[11]	Douyin / Red Note	1,642 videos	Comparative analysis	Framing and circulation

3.2. Effects of different forms of visual re-narration

Among the 299 videos examined by Rovira-Collado et al., 81.94% placed a printed book at the center of the image and 67.82% lasted less than 39 seconds^[9]. The combination is telling: Severe time compression did not prevent creators from making the subject immediately identifiable. Recommendation videos were also prominent in Merga's sample^[11] and in Maddox and Gill's study of popular #BookTok content^[5]. Commentary-led and mixed formats can therefore identify a work while still providing some context. Clips that rely almost entirely on images, by contrast, depend more heavily on viewers' prior familiarity with a symbol or story.

3.3. Interaction between semantic retention and visual appeal

Brief videos fit the rhythms of mobile viewing, but leave little room for sustained literary analysis^[9]. On Chinese platforms, educational and cultural framing has been associated with stronger interaction and circulation than performance-only content, although the size of this difference varies by platform^[11]. A recognizable image can introduce a work quickly. The title, a quotation, a character relationship, or a line of context must then keep that image connected to the source text. Without such anchors, a high play or like count says little about whether viewers have understood the literature.

3.4. Contribution of multimodal features to communication outcomes

Table 2 brings together the most relevant observations. In the 299-video BookTok sample, 58.53% of videos used music, while 41.47% used the creator's voice as a significant audio source^[9]. Semiotic analysis further suggests that choices of language, camera angle, and platform convention position viewers differently as learners^[7]. The modes do different work: Visible books and other objects support identification; captions and speech aid interpretation; music shapes affect and rhythm; and hashtags or interface prompts influence discoverability.

Table 2. Published signals and their interpretation within the mechanism.

Observed signal	Published evidence	Mechanism interpretation
Very brief format	67.82% < 39 s ^[9]	Rapid initial exposure
Printed book as visual focus	81.94% ^[9]	Semantic anchor
Use of music	58.53% ^[9]	Affective coordination
Creator narration	41.47% ^[9]	Interpretive mediation
Influence of BookTokers	55.9% variance explained ^[10]	Deeper engagement
Cultural or educational framing	Higher engagement ^[11]	Semantic depth and platform fit

3.5. Comparing surface-level interaction and deeper participation

Lightweight interactions often register an immediate response to imagery and rhythm. Deeper engagement depends more on interpretation, trust, and relationships within the community. In da Silva et al.'s structural model, attitudes toward BookToker recommendations and perceptions of creators together explained 55.9% of the variance in book-consumption intention^[10]. Interviews with young readers likewise connect sharing, emotional exchange, and reader identity with voluntary reading practices^[6]. Exposure alone is therefore insufficient. Creators still need a consistent interpretive voice, accurate information about the work, and room for discussion if viewers are to form an intention to read.

3.6. Cross-platform applicability and robustness

Platform context changes how the same cultural material is framed. Xu et al. report a stronger entertainment orientation on Douyin, whereas Xiaohongshu gives greater weight to educational and aesthetic framing; on Xiaohongshu, cultural symbols and educational frames also carried a clearer circulation advantage^[11]. Research comparing reading-related content on Instagram and TikTok similarly shows that display conventions affect how reader identity is performed and mediated^[2]. Across these settings, semantic retention, coherence among modes, and community participation remain useful points of comparison. More practical choices, such as shot density, the length of commentary, and the wording of prompts, need to follow the pace and norms of each platform. Cross-platform adaptation should therefore change the presentation without loosening the video's connection to the literary source.

4. Conclusion

Taken together, the reviewed evidence supports a five-part account of literary communication on short-

video platforms: Semantic retention, visual re-narration, multimodal coupling, platform visibility, and audience engagement. Recognizable images, brevity, and music help secure attention. Titles, characters, relationships, quotations, and contextual explanation keep the literary work identifiable, while trusted creators and active communities are more closely tied to discussion, saving, sharing, and intentions to read. Evaluation should therefore consider both reach and the quality of subsequent engagement. Because the studies use different samples, measures, and research designs, their results cannot be reduced to a single effect size or used to rank literary genres. A useful next step would be to build a longitudinal dataset of short videos about Chinese literature that links comment content and saving behavior with evidence of reading the source works. Such a design could compare poetry, fiction, and essays across platforms and examine whether generative visuals introduce semantic drift or aesthetic homogenization.

References

- [1] Merga, M. K. (2021). How can Booktok on TikTok inform readers' advisory services for young people? *Library & Information Science Research*, 43(2), 101091. <https://doi.org/10.1016/j.lisr.2021.101091>.
- [2] Dezuanni, M., Reddan, B., Rutherford, L., & Schoonens, A. (2022). Selfies and shelfies on #bookstagram and #booktok – social media and the mediation of Australian teen reading. *Learning, Media and Technology*, 47(3), 355–372. <https://doi.org/10.1080/17439884.2022.2068575>.
- [3] Guíñez-Cabrera, N., & Mansilla-Obando, K. (2022). Booktokers: Generating and sharing book content on TikTok. *Comunicar*, 30(71), 119–130. <https://doi.org/10.3916/C71-2022-09>.
- [4] Dera, J., Brouwer, S., & Welling, A. (2023). #BookTok's appeal on ninth-grade students: An inquiry into students' responses on a social media revelation. *Journal of Adolescent & Adult Literacy*, 67(2), 99–110. <https://doi.org/10.1002/jaal.1303>.
- [5] Maddox, J., & Gill, F. (2023). Assembling “Sides” of TikTok: Examining community, culture, and interface through a BookTok case study. *Social Media + Society*, 9(4), 20563051231213565. <https://doi.org/10.1177/20563051231213565>.
- [6] Asplund, S.-B., Ljung Egeland, B., & Olin-Scheller, C. (2024). Sharing is caring: Young people's narratives about BookTok and volitional reading. *Language and Education*, 38(4), 635–651. <https://doi.org/10.1080/09500782.2024.2324947>.
- [7] Dezuanni, M., & Schoonens, A. (2024). #BookTok's peer pedagogies: Invitations to learn about books and reading on TikTok. *Social Media + Society*, 10(4), 20563051241309499. <https://doi.org/10.1177/20563051241309499>.
- [8] Dera, J. (2024). BookTok: A narrative review of current literature and directions for future research. *Literature Compass*, 21(10–12), e70012. <https://doi.org/10.1111/lic3.70012>.
- [9] Rovira-Collado, J., Martínez-Carratalá, F. A., & Miras, S. (2025). BookTok time: Characterization of new videos for mobile reading promotion. *Pixel-Bit. Revista de Medios y Educación*, 72, 180–197. <https://doi.org/10.12795/pixelbit.107744>.
- [10] da Silva, T. A. A., Bezerra, E. L. C., de Brito Silva, M. J., & de Oliveira Campos, P. (2025). Reading in the age of influence: An analysis of the impact of BookTokers on book consumption intentions. *Social Network Analysis and Mining*, 15, 100. <https://doi.org/10.1007/s13278-025-01542-8>.
- [11] Xu, A., Li, M., & Lyu, Y. (2026). Framing traditional Chinese painting on short-form video platforms: A comparative study of Douyin and Red Note. *Digital Scholarship in the Humanities*, fqag089. <https://doi.org/10.1093/llic/fqag089>.