

# **Presentation Strategies of Anglophone Cultural Elements in Short-Form Videos: A Multimodal Discourse Perspective**

**Xiaoxia Zhang**

Changchun University of Architecture and Civil Engineering, Changchun, Jilin Province, China 130607

**Abstract:** In the era of rapid development of short-form videos and increasingly convenient intercultural communication, short-form videos have become an important medium for the dissemination of anglophone cultural elements. They employ various types of multimodal symbols such as images, audio, text, and actions to overcome the limitations of traditional single-mode dissemination. The multimodal discourse theory holds that a discourse is composed of multiple different symbols. This provides us with a new insight for understanding how to present anglophone cultural elements in short-form videos and how they influence the audience. Despite the diverse modes of presenting Anglophone cultural elements in short-form videos, persistent issues such as superficial presentation, ineffective integration, and inadequate conveyance of cultural signification continue to hinder effective cross-cultural communication. Based on the fundamental principles of multimodal discourse theory, this paper takes some common short-form videos encountered in daily living contexts as examples, and explores the specific manifestations of Anglophone cultural elements in short-form videos from four aspects: visual modality, auditory modality, textual modality, and multimodal synergy, while analyzing existing shortcomings and providing targeted recommendations to enhance the practical relevance and applicability of the research, ultimately elevating the quality and efficacy of Anglophone cultural dissemination through short-form videos.

**Keywords:** Multimodal Discourse; Short-Form Videos; Anglophone Cultural Elements; Presentation Strategies

**DOI:**10.12417/3029-2328.26.04.031

## **1. Introduction**

With the advancement of new media technologies, short-form videos have emerged as a key conduit for information acquisition and cross-cultural dialogues, leveraging their bite-sized format, engaging appeal, and viral spread characteristics. Anglophone culture is a component of multicultures. It has been well promoted through short-form videos. Characterized by multiple-modality, short-form videos can combine various symbols together to better present Anglophone culture [1]. The multimodal discourse theory holds that discourse is a symbolic system composed of different types of modes such as language, images, and sounds, and these various modes work together to accomplish the task of signification conveying. Currently, it is quite common to present Anglophone culture in short-form videos. However, most of them merely remain superficial and rarely delve into the cultural connotations behind it. Moreover, multimodal symbols do not cooperate with each other tacitly, making it difficult for the audience to truly grasp the essence of Anglophone culture. Given this issue, this paper takes a multi-modal discourse perspective and uses real-life examples to analyze how to present Anglophone cultural elements, and addresses the existing obstacles, promotes the effective implementation of cross-cultural communication, and eschews theoretical abstraction to ensure the concreteness and real-world utility of the research.

## **2. Core Principles of Multimodal Discourse Theory and the Inherent Affinities Between Short-Form Videos and Anglophone Cultural Dissemination**

### **2.1 Core Connotations of Multimodal Discourse Theory**

The multimodal discourse theory is based on system-functional linguistics and holds that “discourse signification is accomplished through various different modalities together”. Here, the modalities refer to various symbol systems that can be used to convey signification, integrating visual, auditory, textual, kinesthetic and gestural

---

Author Profile: Xiaoxia Zhang (b. 1994-), Female, Han Ethnicity, Native of Zhongwei, Ningxia; Master’s Candidate, Teaching Assistant; Research Focus: Discourse Analysis.

symbols. It points out that a single modality alone cannot fully convey discourse signification. Only when multiple modalities complement and collaborate with each other can information dissemination and cross-cultural dialogues be carried out more effectively and vividly. In the research process of multimodal discourse, visual modality (images, frames, colors), auditory modality (sound, music, sound effects), and textual modality (subtitles, titles, copy) are the three most important modalities. With the effective cooperation of these three modalities, the fundamental structure is constructed for short-form video discourse signification transmission. It is also an important channel for us to understand the presentation methodologies of Anglophone culture in short-form videos.

## **2.2 Inherent Affinities Between Short-Form Videos and Anglophone Cultural Dissemination**

The dissemination characteristics of short-form videos adapt to the dissemination needs of Anglophone culture, facilitating the display of Anglophone cultural elements in various forms. On the one hand, short-form videos are suitable for the fragmented reading habits of people today, leveraging their bite-sized format, and fast rhythm. They can quickly capture the audience's attention and reduce the dissemination difficulty of Anglophone culture, enabling the general public to access and understand Anglophone culture. On the other hand, short videos can accommodate various forms of information such as pictures, audio, and text. Within them, abstract Anglophone cultural elements (such as customs and habits, values, language usage rules, etc.) can be concretely and vividly presented, breaking through the limitations of words and language themselves. Meanwhile, short-form videos have strong interactivity. Viewers can leave comments in the comment sections or share the videos to participate in cross-cultural communication, thereby better promoting the dissemination of Anglophone cultural elements. This is precisely what the multi-modal discourse advocates, namely "Collaborative Synergy, Multimodal Meaning-Making" [2].

## **3. Concrete Presentation Strategies of Anglophone Cultural Elements in Short-Form Videos: A Multimodal Discourse Perspective**

### **3.1 Visual Modality: Directly Presenting Tangible Cultural Embodiments**

Visual modality serves as the primary and most immediate vehicle in short-form videos, conveying the surface manifestations of Anglophone culture through imagery, colors, and motion – enabling the audience to grasp cultural representations at a glance. In regular short-form videos, there are mainly two ways to showcase Anglophone cultural elements. One type is daily living contexts, such as Anglophone daily life and diet (hamburgers, French fries, afternoon tea), festival scenes (Christmas, Thanksgiving decorations and celebration activities), and street streetscape (London streets, New York subway), all of which are presented through authentic imagery to manifest the cultural life of Anglophone countries; The other type is the cultural symbols. For instance, in short videos, the national flags, emblems, landmark buildings (such as Big Ben and the Statue of Liberty), and traditional clothing from Anglophone countries are all symbols that embody Anglophone culture and enable to quickly convey cultural identity and cultural characteristics [3]. For instance, in food short-form videos, when a blogger records the production process of British afternoon tea with exquisite tea sets and Dim sum, it enables viewers to have a clear understanding of the emphasis and etiquette of British cuisine at a glance. Without the need for excessive textual explanations, viewers can grasp the main characteristics of British food culture.

### **3.2 Auditory Modality: Transmitting Culturally Distinctive Acoustic Signifiers**

Auditory modality mainly presents the auditory characteristics of Anglophone culture via speech, music, and sound effects, complementing the visual modality and making cultural elements more compelling. It manifests in three forms. The first type is spoken English demonstrations. In short-form videos, the hosts speak English or employ voice-over narration or play authentic audio excerpts (e.g., film dialogues, musical segments), enabling the audience to discern phonological patterns and prosodic features of English, thereby perceiving the aesthetic essence of the language. The second type is music related to Anglophone culture, such as adding English folk songs, pop songs, and holiday theme songs (such as Christmas songs) to videos. These are all indispensable parts of Anglophone culture, which can construct immersive sonic environment through audio-spatial orchestration, triggering embodied

cognitive resonance that makes audiences feel physically present in the narrative space. The third type is culture-related sound effects. In short-form videos, there are sounds such as the roasting of Thanksgiving turkeys, Christmas bells, and British street sirens, which enable people to experience the authentic living contexts in Anglophone countries. For instance, in holiday-themed short-form videos, adding the Christmas bell sound and the English song “Jingle Bells”, along with scenes of Christmas decorations, can quickly create a strong atmosphere of Christmas, enabling viewers to perceive the cultural significance of Western festivals [4].

### **3.3 Textual Modality: Supplementing and Elucidating Contextual Cultural Connotations**

Textual modality serves as a supplement to the visual modality and auditory modality. In short-form videos, it mainly appears in the form of titles, subtitles, and copy, aiming to explain the signification of Anglophone cultural elements and prevent viewers from misunderstanding due to lack of relevant background knowledge. It is manifested in two forms. The first type is explanatory text, such as the original English subtitles in videos and their Chinese translations, as well as text annotations, enabling to further explain the content of the imagery and sounds, thereby spelling out the meanings of Anglophone cultural elements. For example, when presenting the imagery about an English festival, a line of small characters will be added above the imagery: “Thanksgiving is a traditional festival in the United States, used to express gratitude for others’ help and gifts”, so that the audience can understand the cultural connotations behind this festival. The second type is the heuristic text, such as titles or copywriting, which uses questioning or summarizing methods to guide the audience to discover the essential attributes of Anglophone cultural elements. For example, there is a title called “Why Do Britons Like to Talk About the Weather When They Meet?”, which will arouse the audience’s interest and enable them to understand the interpersonal association in Anglophone countries. The application of the textual modality can remedy the shortcomings of audio-visual modalities, making the dissemination of Anglophone cultural elements more diverse and richer, and facilitating viewers to deeply understand their cultural connotations. It is necessary to keep up with the times, update the course content, improve the organization and presentation forms of the course content, and emphasize intra-disciplinary knowledge integration and interdisciplinary integration. Based on the course content and according to the development level of core competencies, academic quality standards are proposed to guide and assist teachers in grasping the depth and breadth of instructional delivery. Guidance efficacy is enhanced through scaffolded academic requirements, pedagogical cues, and exemplar assessment cases.

### **3.4 Multimodal Synergy: Achieving Profound Conveyance of Cultural Signification**

Multimodal collaboration is the primary way to present Anglophone cultural elements in short-form videos, and the basic requirement of the multimodal discourse theory. In short-form videos, it manifests as a semiotic trinity of visual imagery, auditory narration, and linguistic annotation – orchestrating complementary meaning-making pathways. Visual imagery present concrete scenes of Anglophone culture, auditory speech convey the language and cultural climate, while text further elaborates and explains the culture. These three elements complement each other, making the presentation of Anglophone cultural elements more vivid and rich. For example, in videos on Anglophone etiquette culture, viewers can observe the scene where British people shake hands and smile when they meet, concurrently, accompanied by a English greeting (“How do you do?”), and text subtitles: “In initial encounters, the British prioritize ritualized cordiality – manifested through modulated handshake pressure (tactile modality) and a protocol-bound smile (visual modality) – constituting the basic requirement of British etiquette.” In this way, the audience can clearly see the etiquette scenes, hear the relevant words, and understand the cultural signification contained therein, thereby achieving the goal of cultural dissemination.

## **4. Current Challenges in the Presentation of Anglophone Cultural Elements in Short-Form Videos**

### **4.1 Superficial Cultural Presentations with Inadequate Excavation of Intrinsic Connotations**

Currently, the presentation of Anglophone cultural elements in most short-form videos mostly remains

superficial, and rarely involves in-depth analysis of the culture. Short video creators prioritize visual sensationalism, showcasing detached cultural signifiers (e.g., food, festive decorations) while neglecting their semiotic depth – reducing cultural transmission to hollow iconography that obscures value systems and historical consciousness. For instance, many short-form videos merely focus on shooting Christmas decorations, gifts, etc., without delving into the historical roots and symbolic meanings of Christmas, resulting in that viewers only know that Christmas is a traditional western festival, but fail to understand the religious beliefs and humanistic spirit it embodies, impeding their cross-cultural communication efficacy.

#### ***4.2 Insufficient Multimodal Coordination Leading to Suboptimal Presentation Efficacy***

Although multi-modality presentation is used in some short-form videos, insufficient cooperation and even disconnection among modalities hinder cultural dissemination. For instance, in some short-form videos, the video imagery depicts English festival view, but video music is English songs that are unrelated to the festivals, and the subtitles equally do not match the imagery and audio, failing to effectively convey the information. There are also some short videos that overly emphasize the role of the visual modality, neglecting the functions of other modalities such as audio and text, resulting in ambiguous cultural dissemination and making it challenging for the audience to grasp the main meanings of the cultural elements.

#### ***4.3 Homogeneous Presentation Patterns Reflecting Deficient Innovation Capabilities***

The presentation forms of Anglophone cultural elements are relatively monotonous and highly homogeneous in short-form videos. Most video creators focus on cultural content such as food, festivals, and street streetscape. Moreover, they mostly adopt an “imagery + brief introduction” method, lacking novelty and personalized expression techniques, which fails to capture the audience’s attention. Additionally, some videos blindly follow others’ content without considering their own characteristics and the needs of their target audience, making the presentation of Anglophone cultural elements purposeless and ineffective in cross-cultural communication.

### **5. Pathways for Optimizing the Presentation Strategies of Anglophone Cultural Elements in Short-Form Videos**

#### ***5.1 Excavating the Connotative Essence of Culture to Achieve Presentation Transcending from Signifier to Signified***

Short-form video creators should avoid merely adopting superficial presentation techniques. Instead, they should delve into the essential attributes of Anglophone cultural elements, integrating cultural symbols with cultural signification and values, so that viewers can not only see the detached cultural signifiers but also understand their connotations. For instance, when it comes to presenting the culture of Anglophone festivals, apart from filming the festival scenes, video creators can also use subtitles or voice-over to explain the origin, customs, and significance of festivals. When it comes to presenting the English language, not only should the spoken English expression modalities be presented, but also the social rules and value orientations contained therein should be analyzed. Only in this way can it be considered as deep cultural dissemination and truly cross-cultural communication.

#### ***5.2 Optimizing Multimodal Synergy to Holistically Enhance Presentation Efficacy***

From the perspective of multimodal discourse analysis, it is important to effectively integrate visual, auditory and textual modalities. During the process of creating videos, it is necessary to make reasonable combinations based on specific situation of Anglophone cultural elements: visually constructing cultural scenes through authentic verisimilitude and semiotic playfulness; In terms of auditory modality, emphasis is placed on matching with the imagery, leveraging sound, music, and sound effects to create targeted cultural climate. The textual modality should concisely and accurately expound the cultural connotations. The three elements develop a good interactive relationship, thereby better conveying the cultural connotations.

### **5.3 Innovating Presentational Paradigms to Engineer Personalized Content Architectures with Cross-Cultural Resonance**

It is necessary to transcend the homogenized presentation paradigms to carry out innovative expressions of Anglophone cultural content based on audience needs and their own characteristics. For instance, by adopting scenario reenactment, video creators can present daily living contexts and social occasions in Anglophone countries through role-playing, thereby making cultural elements more interesting. By adopting “comparative analysis”, video creators can compare the differences between Chinese and Western cultures, enabling the audience to viscerally decode Anglophone cultural DNA. By integrating with current hot events and infusing Anglophone cultural elements into them, the dissemination power and influence of short-form videos can be enhanced, making the spread of Anglophone culture align more closely with daily living contexts and more appealing.

### **6. Conclusion**

From the perspective of multimodal discourse, short-form videos leverage their unique audio-visual combinations and multiple other forms (e.g. text) to present Anglophone cultural content, and open up a new channel for cross-cultural communication with their intuitive, vivid, and convenient features. Currently, although the Anglophone cultural content in short-form videos has diverse expression modalities, there are still problems such as superficial presentation, lack of effective integration and synergy, and severe homogenization, which impede the efficacy of cultural dissemination. Therefore, when creating short-form videos, video creators should, based on multimodal discourse philosophy, explore the profound connotations of Anglophone culture, enhance the cooperation among different types of multi-media tools, and explore new presentation techniques to make the content of Anglophone culture more deeply integrated into people’s daily living contexts, thereby better meeting the audience’s cultural pursuit while effectively conveying Anglophone cultural information to the whole world to promote cross-cultural communication and comprehension.

### **References:**

- [1] Le Zhang.(2026)*International Communication Strategies of Traditional Chinese Medicine Culture in English within the New Media Context:A Case Study of Traditional Chinese Medicine Short-Form Videos on TikTok*[J]. *Overseas English*,3,37-39.
- [2] Miao Yan.(2025)*Cultural Adaptability Characteristics and Optimization Strategies of English Short-Form Video Communication of Maoming Intangible Cultural Heritage from the Perspective of International Audience*[J]. *Insight-Shaanxi*,12,91-93.
- [3] Yan Zhang.(2024)*Strategies for TikTok Short Videos Enabling Local Culture to Integrate into College English Teaching:A Case Study of Shanxi Culture*[J].*Journal of News Research*,15(21),173-177.
- [4] Sulan Yang.(2022)*Exploration and Practice of Enhancing College Students’ Capabilities to Spread Excellent Traditional Chinese Culture in English:Taking the “Mastering Fluent English·Showcasing Chinese Culture” Short Video Competition as an Exemplary Case*[J].*Journal of Suzhou Education Institute*,25(05),26-29.