English Language Teaching and Linguistics Studies
ISSN 2640-9836 (Print) ISSN 2640-9844 (Online)
Vol. 7, No. 4, 2025
www.scholink.org/ojs/index.php/eltls

Original Paper

Visual Grammar and Cultural Meaning in *Festive China*: A Multimodal Discourse Analysis

Wenhan Pan[1]

[1] School of Foreign Languages, Guangzhou College of Commerce, Guangzhou, China

Received: May 29, 2025 Accepted: July 29, 2025 Online Published: August 13, 2025

doi:10.22158/eltls.v7n4p67 URL: http://dx.doi.org/10.22158/eltls.v7n4p67

Abstract

This study examines how the English-language documentary series Festive China constructs and communicates cultural meaning through visual semiotic resources by focusing on the role of visual modality in shaping cultural narratives. Drawing on Systemic Functional Linguistics and visual grammar, the study adopts a multimodal discourse analysis approach to investigate twelve episodes of the series. Using ELAN 6.0, over 1,800 segments were annotated across six visual categories: process types, symbolic meaning, perspective, social distance, framing, and color. The findings reveal that Festive China places strong emphasis on material and existential processes to represent tradition as dynamic and embodied. Symbolic elements are consistently embedded within narrative structures to express values like continuity, reverence, and harmony. Variations in gaze, angle, and shot distance shape audience engagement, while compositional features such as soft framing and culturally marked color schemes enhance visual coherence and meaning. The study suggests that such visual strategies not only reinforce national identity but also serve as effective resources for multimodal pedagogy. By highlighting how visual modality can mediate cultural experience, this research contributes to the growing field of visual discourse analysis and offers pedagogical implications for the use of documentary media in language education and intercultural learning.

Keywords

multimodal discourse analysis, visual grammar, visual semiotics, Festive China, documentary, intercultural communication

1. Introduction

President has clearly stated the need to improve how China communicates with the world. He believes that China must build a stronger voice in the global arena. To do this, he calls for messages that are more creative, more persuasive, and easier for people to trust. One key part of this effort is to tell good

stories about China. These stories should help people from different countries and cultures understand China better and feel closer to its traditions.

As the world becomes more connected, the way people share and receive information is also changing. The rise of digital media and global networks has made international communication faster and more complex. These changes create new chances for countries to share their ideas, but they also bring new challenges. Countries now need to find better ways to shape how they are seen. For China, this means finding ways to talk to the world that fit both its own culture and the interests of global audiences.

In this new media environment, storytelling tools must be chosen carefully. Among them, documentary films play a special role. They can show real life, explain ideas, and touch people's emotions. As Barsam (2013) explains, documentaries are more than just films—they can teach, persuade, and connect people. This makes them a strong tool for telling China's story to the world in a way that feels real and meaningful. They do so by constructing emotionally engaging and ideologically meaningful representations of reality. In contrast to fictional narratives, documentaries aim to depict curated yet authentic experiences. By combining multiple modes of expression—such as moving images, narration, background sound, and written text—documentaries serve as effective instruments for public diplomacy and international cultural engagement.

Given their inherently multimodal nature, documentary films are especially suitable for analysis through the framework of multimodal discourse analysis (MDA). MDA examines how meaning emerges from the interplay of different semiotic modes, including linguistic, visual, auditory, and spatial resources, thereby offering a more comprehensive understanding of mediated communication. In contemporary media texts, and especially in documentary films, the visual modality plays a central role in meaning construction. Elements such as camera angles, color palettes, composition, and framing critically shape the audience's interpretation and engagement.

In recent years, MDA has gained increasing scholarly attention, with applications spanning education, media studies, and sociolinguistics. Researchers have explored how multimodal texts communicate ideology, construct social identity, and reflect cultural values (Jewitt, 2009; O'Halloran, 2008). However, while MDA is widely applied in educational discourse and visual advertising, focused analysis of cultural documentaries from a visual semiotic perspective remains scarce, particularly in the context of Chinese cultural representation in English-language media.

One of the most influential theoretical tools in this domain is the visual grammar framework developed by Kress and van Leeuwen (2006). This model builds on Halliday's Systemic Functional Linguistics (SFL) and extends its core principles into the domain of visual communication. It proposes three metafunctions: ideational, interpersonal, and compositional. The ideational metafunction addresses what is depicted in the image, including participants, processes, and settings. The interpersonal metafunction examines how visual structures shape the viewer's relationship with the represented participants, through gaze, angle, and distance. The compositional metafunction explores how visual elements are arranged-focusing on information value, salience, and framing. The three metafunctions

together establish a structured methodology for interpreting how visual images communicate meaning. Each metafunction offers a different way to understand how images create meaning. These functions work together to form a clear method for studying visual content.

Based on this idea, Martinec and Salway (2005) introduced a model to show how images and words connect. They focused on the links between pictures and language elements like subtitles, spoken text, or captions. Their model puts these links into three groups: extension, elaboration, and enhancement. Each group explains a different way that words and images work together to build meaning. Such classification allows for detailed analysis of how different semiotic modes cooperate to construct cohesive and persuasive multimodal discourse.

However, despite the value of these models, empirical research on English-language documentaries portraying Chinese culture remains scarce. Existing studies primarily address translation practices, lexical features, or overarching thematic content. As a result, the independent contribution of visual semiotics in constructing cultural meaning has often been marginalized. Even when studies adopt a multimodal perspective, few offer a detailed, visual-focused analysis or apply visual grammar systematically to unpack cultural representation.

This study aims to address that gap by focusing on Festive China, an English-language documentary series co-produced by China Global Television Network (CGTN) and various domestic media institutions. The documentary series Festive China introduces a range of traditional Chinese festivals including the Spring Festival, Lantern Festival, Dragon Boat Festival, Mid-Autumn Festival, and Qixi Festival to viewers around the world. It makes use of a wide array of filmic techniques to convey cultural meaning. These include the use of symbolic color palettes, dynamic and expressive camera movements, and carefully composed visual frames. These elements work together to present Chinese traditions in ways that are both visually striking and emotionally engaging.

This research adopts an analytical approach that centers on visual meaning-making. It is grounded in the frameworks of visual grammar and image-text relation theory. This study looks at selected scenes from Festive China to understand how the documentary shows Chinese culture. It first focuses on what the images show, how the scenes speak to viewers, and how the visual space is used.

Next, the study looks at how the images work together with spoken words. It pays special attention to the voice-over. By doing this, it shows how pictures and language are put side by side to make the meaning clearer.

2. Related Work

Multimodal Discourse Analysis (MDA) is now widely used in language and communication studies. It gives researchers simple tools to study how different forms—like speech, writing, images, and sound—work together to make meaning. This method is based on Systemic Functional Linguistics (SFL), first developed by Michael Halliday. SFL sees language as something people use in real life to express meaning. Later, scholars like Halliday and Hasan (1989), Halliday and Matthiessen (2014), and

O'Halloran (2004) built on this idea. They used it to look at other types of meaning-making, not just language. Their work helped create a way to study meaning in different kinds of signs and symbols.

2.1 Visual Grammar and Systemic Functional Foundations

A major step forward in multimodal discourse research is the visual grammar framework by Kress and van Leeuwen (2020). They built this model by drawing on Halliday's Systemic Functional Linguistics. They moved its three main functions—representational, interactive, and compositional—from language to images. This gives a clear way to study how pictures make meaning in different cultures and settings.

The representational function looks at what is shown in the image. It includes actions like movement and human contact. It also covers ideas like grouping and symbols. This part helps pictures show both real-life events and abstract ideas (Kress & van Leeuwen, 2020; Martinec, 2000). The interactive function focuses on how images connect with the viewer. It looks at things like where people in the image are looking, the camera angle, how close the shot is, and how real the image feels. These features shape how viewers feel—whether they are involved, distant, or simply watching.

The compositional function deals with how the image is built. It includes layout, focus, borders, and how information moves across the picture. These choices help the whole image stay clear and connected (Kress & van Leeuwen, 2001, 2020). Many researchers have used this model in different fields. They have looked at ads (Guo & Li, 2020), school materials (Jewitt, 2009), and tourism websites (Jia, 2021). In China, visual grammar has been localized and applied to textbooks, advertising, and media texts (Zhang & Jia, 2012; Liu, 2019; Wu, 2022), showcasing its flexibility and relevance across semiotic genres. Their work shows that this framework is useful across subjects and fits well with today's media.

2.2 Image–Text Interaction

Visual grammar gives a basic way to study how pictures show meaning. But many multimodal texts, like documentary films, do not use visuals alone. They often mix images with words, such as spoken narration or written text. Because of this, it is important to look at how these modes work together.

Martinec and Salway (2005) made a model to study how images and text relate to each other. They looked at two things: which mode stands out more, and how the image and text connect in meaning. These meaning links fall into three types—elaboration, extension, and enhancement. Many researchers have used this model to study websites (Djonov, 2007), school texts (Unsworth, 2006), and ads (Thibault, 2000). But people have not used it much to study documentary films.

Royce (2020) added another idea called intersemiotic complementarity. This means that different types of signs—like pictures and words—can work together to explain or support meaning. This idea works well for cultural documentaries, where images and voice-over often shape meaning together (Zhao & Djonov, 2020).

Later, Feng and Zhao (2017) brought in the idea of multimodal metonymy. It shows how images and language can link in symbolic ways to create meaning tied to a certain culture. Even though these ideas

are useful, few studies have used them to look at English-language documentaries about Chinese culture. Most research has focused more on subtitles, translation, or how audiences respond. So, how images and words really work together in these films is still not well understood.

2.3 Multimodal Construction in Documentaries

Recent studies have changed how researchers look at multimodal meaning in documentaries. Scholars no longer focus only on simple links between images and text. Now, they look at all the types of meaning-making used in film. Bateman and Schmidt (2021) point out that documentaries bring together visuals, sound, and language. These parts work side by side to build a clear story and create emotional impact. In the same way, Barsam (2013) and Nichols (2017) say that documentaries are not just neutral records. They use different modes to shape how viewers understand what they see.

New technology has made this kind of research easier. Tools like UAM CorpusTool (Donnell, 2008) and ELAN (Wittenburg et al., 2006) help researchers mark and study sound and images in detail. These tools make the process more accurate and easier to repeat.

On the theory side, some scholars offer ways to study meaning across time and space in digital media. O'Halloran's (2008) Systemic Functional Multimodal Discourse Analysis (SF-MDA) and Lemke's (2002) idea of hypermodality give useful models for this kind of work.

In China, more scholars are now using Multimodal Critical Discourse Analysis (MCDA). Researchers like Hu (2007), Zhu (2007), and Zhang (2018) say that social and cultural context should be part of multimodal studies. But most of this work still looks at political or business media. Documentaries made to reach global audiences have not been studied as much in this field.

2.4 Visual Grammar in Cultural Documentaries

In recent scholarship, visual grammar has been applied to the study of documentary media focusing on cultural identity and representation. For example, Hong and Duan (2020) explored Aerial China and noted how aerial imagery and landmark emphasis supported regional identity. Likewise, Bie (2022) examined Festive China and highlighted the role of color, composition, and close-up framing in conveying traditional values.

Nonetheless, many such studies adopt a predominantly descriptive approach, focusing on isolated visual features rather than applying all three metafunctions of visual grammar in a systematic manner. Moreover, comprehensive multimodal analyses of English-language documentaries produced for international dissemination remain scarce.

2.5 Research Gap and Orientation

A review of the literature reveals several critical gaps. Although the visual grammar framework has been widely used, few studies comprehensively apply all three metafunctions—representational, interactive, and compositional—in the analysis of documentary discourse. Most research isolates visual features without offering a holistic account of how meaning is visually constructed.

Moreover, despite theoretical progress in intersemiotic analysis, few empirical studies have explored how visual and verbal elements interact in intercultural documentary contexts. Research tends to

emphasize subtitles, translation, or audience reception, while the underlying semiotic mechanisms of image–text interplay remain insufficiently examined.

Finally, critical multimodal discourse analysis is rarely applied to English-language cultural documentaries. As a result, the ideological and communicative strategies embedded in these media texts are often overlooked.

To address these shortcomings, this study examines how Festive China, an English-language documentary series, uses the representational, interactive, and compositional resources of visual grammar to construct cultural meaning. It further investigates how image-text relations and intersemiotic complementarity contribute to cohesive multimodal narratives. In doing so, this research aims to enhance our understanding of intercultural communication strategies in documentary filmmaking and contribute to a more integrated multimodal analytical framework for visual-cultural discourse.

3. Materials and Methods

This study adopts a multimodal discourse analysis approach grounded in the theoretical framework of visual grammar. The methodological design integrates several interrelated components, including theoretical foundations, corpus construction, annotation tools, visual annotation schemes, and multimodal analysis procedures.

3.1 Theoretical Framework

This research builds on the visual grammar framework developed by Kress and van Leeuwen (2001), extending Halliday's Systemic Functional Linguistics (SFL) into the visual domain. Visuals are treated as meaning-making systems, capable of performing three metafunctions: representational, interactive, and compositional. These metafunctions serve as the analytical lens through which visual content in Festive China is examined.

Representational meaning concerns what is depicted in the image—such as social actors, cultural artifacts, or symbolic references—and whether the representation follows narrative processes (e.g., movement, interaction) or conceptual structures (e.g., classification or symbolism). Interactive meaning refers to how viewers engage with the represented subjects through gaze, camera perspective, and social distance. Compositional meaning relates to the organization of visual elements in the frame, including salience, information flow, and balance, which collectively influence how visual messages are interpreted.

Complementing visual grammar, this study adopts Martinec and Salway's (2005) framework for analyzing image–text relations. Their model classifies the interactions between visual and verbal modes based on relative status (e.g., dominant or equal) and semantic relations (e.g., elaboration, extension, or enhancement). This model helps assess how visual and textual elements cooperate in building cohesive cultural narratives.

Additionally, the research draws on multimodal corpus analysis (MCA), which combines corpus

linguistics with multimodal theory to facilitate both qualitative and quantitative inquiry. As advocated by Baldry and Thibault (2006) and Bateman (2014), MCA enhances methodological transparency through systematic annotation. While its application in Chinese academic research remains limited, emerging studies (e.g., Liu, 2017; Huang, 2015) demonstrate its potential in addressing annotation standardization and cross-platform compatibility, which this study further explores through tailored implementation.

3.2 Corpus Construction and Data Collection

The empirical base of this study is a self-compiled multimodal corpus constructed from selected episodes of the English-language documentary Festive China, co-produced by CGTN and other domestic media institutions. Episodes were chosen based on thematic coverage of traditional Chinese festivals, visual richness, and cultural significance. The final corpus includes episodes focusing on the Spring Festival, Lantern Festival, Dragon Boat Festival, and Mid-Autumn Festival.

The corpus was developed through a structured three-stage process. First, episodes were segmented into analytical units based on shifts in visual and thematic content. Each segment was marked by scene transitions or changes in narrative focus. Second, key visual frames were extracted and matched with time-stamped transcripts of corresponding narration. These were manually cross-referenced to ensure accuracy. Third, the data were systematically catalogued with detailed metadata—such as festival type, geographic setting, symbolic imagery, and participant roles—to support later searchability and comparison.

Corpus construction followed methodological protocols outlined by Baldry and Thibault (2006). Manual segmentation ensured consistent identification of scenes, and metadata tagging enhanced analytical precision. Ethical clearance was obtained from the Ethics Review Committee at St. Paul University Philippines. Research integrity was maintained through precise timecode alignment, standardized file naming, and detailed documentation practices, ensuring reproducibility and analytical transparency.

3.3 Annotation Tools and Technical Implementation

To conduct multimodal annotation, this study employed three software tools that correspond to different analytical dimensions: visual, verbal, and temporal.

For visual annotation, UAM CorpusTool 6.0 was used to label and analyze visual elements according to the metafunctional categories of visual grammar. Its hierarchical coding system enabled consistent identification of recurrent patterns across visual segments.

Verbal analysis was conducted using AntConc 4.3.1 (Anthony, 2005), a corpus analysis tool that helped detect word frequencies, collocations, and rhetorical structures in narration. This allowed for systematic comparisons between verbal discourse and visual representation.

ELAN 6.8 was used to conduct time-aligned multimodal annotation. Customized annotation tiers were created to capture specific visual features—including camera angle, gesture, shot scale, framing, and symbolic imagery—as well as instances of image–text alignment. Through ELAN's interface, visual

annotations were directly linked to verbal data, facilitating synchronous cross-modal analysis.

The annotation protocols closely followed the standards set by O'Halloran (2004, 2008) and Baldry and Thibault (2006), ensuring theoretical coherence and methodological consistency across the software platforms.

3.4 ELAN-Based Visual Annotation Scheme

In this study, ELAN served as the primary tool for segmenting and annotating the visual modality of Festive China. Designed for time-aligned, multilayered annotation of audiovisual content, ELAN enabled the research team to systematically label visual elements based on the three metafunctions of visual grammar.

The annotation scheme was organized into six visual-semantic dimensions. First, process types [P-] were classified into material processes [PMT], mental processes [PMN], verbal processes [PVP], and existential processes [PEP]. Second, symbolic categories [S-] included seasonal symbols [SSS] and nature-related symbols [SNS], emphasizing culturally embedded iconography.

Perspective [P-] was coded by camera angle: low-angle [PLA], high-angle [PHA], and eye-level [PEL]. Social distance [S-] was captured through shot types: close-up [SCU], medium shot [SMS], and long shot [SLS]. Framing [F-] was labeled as clear boundary [FCB], blurry boundary [FBB], or contrastive framing [FCF], reflecting compositional strategies. Dominant color [C-] codes included red [CRD], green [CGN], white [CWT], pink [CPK], yellow [CYL], and blue [CBL].

Before the annotation began, a standard template was created to keep the work consistent across all twelve episodes. Each segment was marked by hand. Then, the team went through several rounds of checking and review. To make sure the coding was reliable, two trained researchers coded the same parts on their own. The agreement between them was high. The Cohen's Kappa scores were all above 0.85. These annotations became the base for later analysis.

3.5 Multimodal Analysis Procedures

The study used a five-step process to look at how images and words work together to make meaning. This process combined visual analysis, language patterns, and links between different modes. Each step helped build a full picture of how the documentary tells its story.

In the first step, selected frames were divided into two types: narrative and conceptual. Narrative images showed actions or people doing something. Conceptual images showed ideas, symbols, or groups. This sorting followed basic ideas from visual grammar. The second step looked at features like eye contact, camera angle, and how close the shot was. These helped show how the viewer connects with the image. The third step focused on layout. It looked at what stood out, how balanced the image was, and where things were placed on the screen. These choices helped make the image clear and organized.

At the same time, the study looked at the spoken words using AntConc. This helped find repeated words, common phrases, and main topics in the voice-over. These patterns were then compared with the images to see if they matched or said something different.

In the last step, the study used Martinec and Salway's (2005) model to look at how pictures and words were linked. It looked at three types of links: elaboration, extension, and enhancement. These links showed how the two parts worked together to shape meaning.

4. Results and Discussion

This analysis examines how Festive China conveys cultural meaning through visual elements. It is based on annotated data created using ELAN 6.0. These annotations were coded across six semiotic dimensions: type of process, symbolic category, visual perspective, perceived social distance, framing techniques, and use of color. Each category is examined in relation to the three metafunctions outlined in visual grammar—namely, representational, interactive, and compositional meanings. This combined framework supports a systematic interpretation of the documentary's visual discourse. To maintain clarity, the analysis is structured into five sections, each addressing a distinct aspect of visual meaning-making.

4.1 Overall Structure of the Discourse

Each episode of Festive China follows a general structural template. However, there are variations in the themes and festivals featured across the twelve episodes. By analyzing the annotated segments of each video, four primary structural parts were identified: Introduction, Origins, Tradition, and Transition. These divisions reflect the narrative rhythm of the documentary and guide viewers through the cultural themes presented in each episode.

Table 1 presents the duration of each segment across all episodes. This table illustrates how the time allocated to each section varies from one episode to the next. The Introduction segment accounts for around 6% to 8% of the total runtime in every episode. Typically, this part includes landscape shots, symbolic cultural motifs such as red lanterns or incense, and brief voiceovers that introduce the upcoming theme. The consistent inclusion of this segment serves as a framing device, helping the audience orient themselves within the festive theme of the episode, even before any historical or cultural practices are shown.

In comparison, the Origins and Tradition segments display more variation in both length and content. For example, in Episodes 1 and 4, over 40% of the runtime is dedicated to explaining the origins of the festivals being featured. On the other hand, Episodes 2 and 10 allocate much less time—around 15%—to this content, focusing instead on the enactment of cultural customs. This flexibility reveals that Festive China does not follow a rigid narrative structure.

Table 1. The Duration of Different Parts of The Episodes

Episode ID	Total Duration	Introduction Duration	Origins Duration	Tradition Duration	Transition Duration
1	04:19.1	00:17.1	01:45.3	01:36.8	00:39.9

2	03:55.1	00:17.2	00:28.2	02:57.7	00:12.0
3	03:30.8	00:16.8	00:51.5	01:56.1	00:26.3
4	03:29.7	00:17.2	01:24.9	01:26.1	00:21.4
5	04:10.0	00:17.3	01:29.6	02:03.1	00:20.0
6	03:59.5	00:17.6	01:06.4	02:05.6	00:29.9
7	03:52.0	00:17.4	00:38.1	02:47.1	00:09.4
8	03:45.5	00:18.3	01:51.2	01:16.1	00:19.9
9	04:16.7	00:18.2	01:03.5	01:59.9	00:55.1
10	03:53.1	00:18.0	00:22.5	02:49.0	00:23.6
11	04:13.5	00:17.3	00:54.2	02:56.5	00:22.8
12	03:51.6	00:17.3	00:13.8	02:56.5	00:24.0

Festive China does not rely on a fixed narrative formula. Instead, it modifies its episode structure according to the thematic content of each featured festival. This flexible framework enables the documentary to emphasize different cultural dimensions, depending on the nature and focus of the festival being portrayed.

The Tradition segment usually takes up the largest part of each episode. In Episodes 2, 5, and 11, for example, it covers more than 60 percent of the total time. This part focuses on how people express cultural heritage in daily life. It shows food preparation, ritual activities, seasonal customs, and events in local communities. By highlighting these everyday scenes, the segment presents tradition as something living and changing. It does not show tradition as something fixed or lost. Instead, it shows how it still matters today.

The Transition segment, on the other hand, is usually much shorter. It often makes up less than 10 percent of an episode. Although short, it plays an important role in the flow of the story. It often includes slow shots of nature, soft background sounds, and quiet voice-over. These parts help connect different themes or signal the end of the episode. Episode 9 is an exception to this pattern. In that episode, the Transition segment exceeds 15 percent of the total runtime. This extended duration highlights themes of cyclicality and seasonal transformation, aligning with the festival's symbolic meaning.

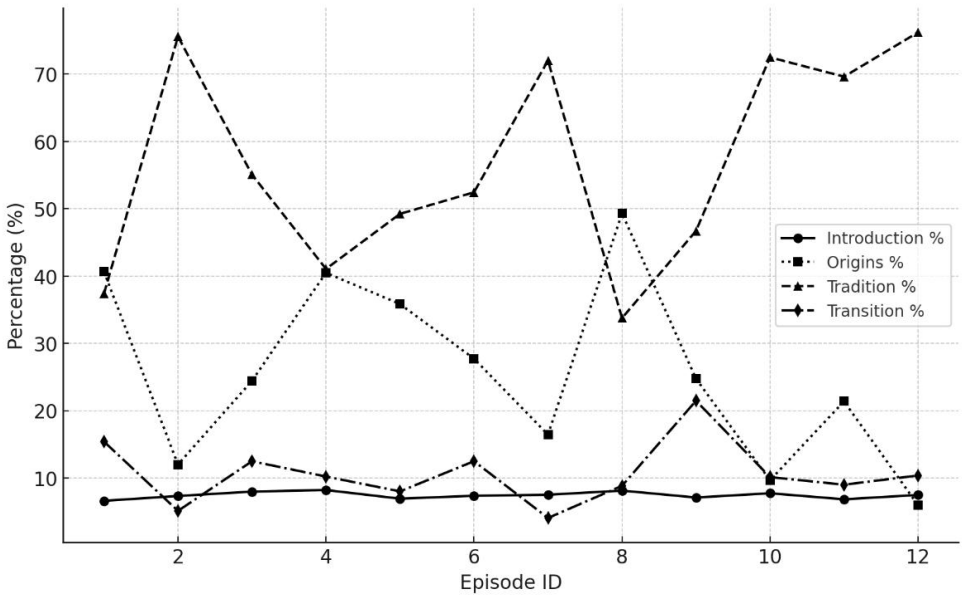

Figure 1. Percentage Of Each Section Relative To Total Duration (Per Episode)

Figure 1 visualizes the time allocation for each segment relative to the total runtime of the episodes. The data supports the argument that Festive China deliberately adjusts segment lengths to suit the narrative demands of each cultural topic. This structural variation reflects a balance between thematic focus and formal coherence.

The series operates through a dual narrative structure. The Introduction and Transition segments provide consistency and cohesion across episodes. Meanwhile, the Origins and Tradition segments offer interpretive flexibility. This combination of repetition and variation supports both educational engagement and aesthetic appeal. In doing so, the documentary creates a viewing experience that is culturally rich, structurally clear, and globally accessible.

4.2 Visual Modality of Festive China

The visual content in Festive China is studied using a structured, step-by-step annotation system. This system includes six visual categories: process type, symbolic meaning, point of view, social distance, framing, and color. The full set of data comes from all twelve episodes. It forms a multimodal corpus that supports both counting visual patterns and looking closely at what they mean.

Process type is the most common visual feature in the annotated data. Among its subtypes, material processes appear most often. A total of 142 instances were recorded, accounting for approximately 73.7% of the process-related visual duration. These sequences typically depict observable, physical actions, such as individuals preparing traditional dishes, organizing ritual spaces, or engaging in festive performances. The frequent recurrence of these images reflects the documentary's core representational strategy: to frame cultural traditions as dynamic, embodied, and enacted practices rather than static or symbolic references.

Existential processes form the second most prominent category, accounting for 65 instances and 25.6%

of the visual process duration. These shots usually present fixed or static cultural sites, including temples, riverbanks, ceremonial altars, or festive artifacts. Their function extends beyond scene-setting. These visuals contribute significantly to the construction of cultural continuity by asserting a sense of permanence, rootedness, and collective identity. In this way, the existential process acts as a symbolic anchor, embedding the narrative in specific spatial and cultural contexts.

In contrast, mental and verbal processes are notably scarce. Mental processes appear only four times, while verbal processes are represented in a single instance. Together, they constitute less than 1% of the total visual time. This marginal presence highlights a deliberate narrative choice to minimize introspective or discursive modes of meaning-making. Instead, the documentary privileges externalized, collective action and nonverbal semiotic resources. The visual design of Festive China reflects a carefully constructed approach that emphasizes shared experience over individual interpretation. Festive China does not rely on direct explanations or detailed voice-over. Instead, it encourages viewers to watch, feel, and connect with what they see. This choice matches a broader visual style that values immersion. The film invites the audience to take part in cultural scenes, rather than just observe them from a distance.

Table 2. Annotation of Visual Modality in Festive China

Annotation Tier	Annotation Feature	Occurrence	Duration	Percentage
Process Type [P-]	Material Process [PMT]	142	1420.43	73.70
	Mental Process [PMN]	4	11.11	0.58
	Verbal Process [PVP]	1	2.02	0.10
	Existential Process [PEP]	65	493.86	25.62
Symbolic Category[S-]	Seasonal Symbol [SSS]	69	708.39	50.27
	Nature Symbol [SNS]	73	700.72	49.73
Perspective [P-]	Low Angle [PLA]	168	841.68	32.81
	High Angle [PHA]	214	1156.41	45.08
	Eye Level [PEL]	113	566.94	22.10
Social Distance[S-]	Close-up [SCU]	128	893.7	34.32
	Medium Shot [SMS]	177	1088.64	41.81
	Long Shot [SLS]	103	621.41	23.87
Framing[F-]	Clear Boundary [FCB]	60	389.29	14.52
	Blurry Boundary[FBB]	178	1812.87	67.62
	Contrasting Framing[FCF]	60	478.93	17.86
Color[C-]	Red[CRD]	36	220.49	18.39
	Green[CGN]	58	407.84	34.02

White[CWT]	19	98.77	8.24
Pink[CPK]	17	93.54	7.80
Yellow[CYL]	58	369.33	30.81
Blue[CBL]	2	8.88	0.74

One key part of the analysis is the use of symbols. The data shows that seasonal and natural symbols appear in nearly equal amounts. Seasonal symbols—like red lanterns, mooncakes, and banners—often appear with natural signs such as snow, rivers, flowers, and seasonal plants. This pairing suggests a link between human activities and nature. It reflects the idea that Chinese festivals follow the rhythm of the seasons and are closely tied to the environment.

Perspective also plays an important role. High-angle shots appear most often, making up 45.08% of the data. These views, often taken by drones, show landscapes and group events from above. They create a calm and distant feeling. Low-angle shots follow at 32.81%. These usually look up at temples, statues, or rituals, adding a sense of power or respect. Eye-level shots are less common (22.10%), but they help build emotional closeness. These shots often show faces or personal moments, helping viewers feel connected to the people on screen.

The way people are shown in terms of distance adds more detail. Medium shots are the most used, taking up 41.81% of the footage. These show people doing things in their space, like preparing food or taking part in a festival. Close-up shots come next at 34.32%. These focus on small details—such as a hand gesture or a special object. Long shots make up 23.87%. These help set the scene and show the wider setting.

Framing also shapes how the story moves. Most of the shots have soft or unclear edges (67.62%). These make transitions smooth and help create a poetic mood. Clear edges (14.52%) are used when focus is needed. Strong contrasts in framing (17.86%) appear when the film shows clear differences, like city versus countryside, or old versus new.

Color is another key tool. Green (34.02%) and yellow (30.81%) are the most common. They suggest life, energy, and celebration. These colors match scenes of farming and nature. Red is less frequent (18.39%), but it stands out. It often shows up in important or emotional scenes. Other colors—like white, blue, and pink—add small differences in mood and meaning.

4.3 Representational Meaning in Cultural Discourse

The representational meaning in Festive China is constructed through the intentional use of key visual elements. These elements include material processes, existential processes, and culturally embedded symbols. Combined, they form a cohesive visual system that presents cultural traditions as both physical practice and symbolic expression. The documentary frames tradition as something not only performed in the present but also embedded in specific locations, collective memory, and shared cosmological worldviews.

Material processes are the most frequently used among the various types of visual representation. They account for more than 70 percent of all identified process types in the annotated dataset. These scenes depict individuals engaged in concrete, observable actions. Common examples include preparing festival foods, performing traditional dances, lighting incense, or taking part in communal rituals. The camera often highlights motion and bodily activity. It repeatedly focuses on hands kneading dough, bodies moving in coordinated dances, or people participating in festive gatherings.

These images serve more than one purpose. On the one hand, they document cultural heritage through detailed observation. On the other, they act as narrative tools that express how tradition is physically embodied and performed in everyday life. This emphasis on agency aligns with the theoretical insights of Kress and van Leeuwen, who propose that narrative meaning in visual media often emerges through action and involvement.

In contrast, existential processes appear less frequently but are still meaningful. These scenes typically depict static visuals, such as sacred spaces, ritual altars, or symbolic objects. Though they lack movement, they convey atmosphere, spirituality, and cultural significance. Instead, they assert the continued presence of tradition by highlighting settings that function as anchors of cultural permanence. In doing so, they lend authority to the material processes that follow. The spatial environment becomes a legitimizing frame. Consequently, the relationship between material and existential processes is both sequential and complementary. The former illustrates how tradition is performed; the latter affirms the cultural and spatial context in which such enactment takes place.

Figure 2. Mid-Autumn Festival – Material and Existential Processes in Visual Harmony

A representative instance of visual layering appears in the Mid-Autumn Festival episode. The scene features a family spanning multiple generations, gathered to share mooncakes under the full moon. The gathering takes place beneath a circular moon gate and is softly lit by red lanterns. The act of eating together is presented as a material process. It expresses familial closeness, cultural continuity, and the

symbolic notion of reunion.

At the same time, the moon, lanterns, and stillness of the architectural setting function as existential elements. These components evoke a sense of timelessness and emotional depth. They frame the material activity within a space rich in cultural symbolism. This combination demonstrates how the documentary constructs representational meaning by merging performative acts with symbolic environments.

Mental and verbal processes are almost entirely absent, appearing in fewer than one percent of scenes. This scarcity reflects a deliberate narrative strategy. Festive China avoids direct introspection or explicit spoken commentary. Instead, it relies on visual cues—gesture, expression, spatial layout—to communicate meaning. The documentary invites viewers to watch and reflect, rather than follow a fixed explanation. It encourages meaning-making through what people see and feel. Symbols play a key role in adding depth to what the images show. Many of these come from seasonal or natural scenes. Red lanterns, snowflakes, and mooncakes often appear with plum blossoms, rivers, and fire.

These symbols are not just for decoration. They belong to a shared cultural system. For example, red lanterns often stand for happiness and good fortune during the Spring Festival. But when they appear with things like ancestral offerings or winter landscapes, they suggest respect and the cycle of life and renewal.

The Qingming Festival episode offers another example. Willow branches show up many times. People place green willow twigs on family graves. The green color stands for life and new beginnings. This small act connects the present with the past. It shows how seasonal images carry deeper values across time, which is one of the main ideas behind the documentary's visual choices.

4.4 Symbolic and Cultural Signification in the Visual Mode

The symbolic mode plays a central role in how Festive China builds its visual message. Throughout the series, symbols from seasonal customs and nature appear again and again. These images are not added by chance. They are placed carefully in the scenes to suggest shared cultural meanings. Each symbol acts like a short visual message. It helps show values such as new beginnings, harmony, memory, and joy. Together, these symbols form an important part of how the documentary tells its story.

One of the most common and powerful symbols is the red lantern. It appears in episodes about major festivals, such as the Spring Festival, Lantern Festival, and Mid-Autumn Festival. The red lantern is more than decoration. It stands for wealth, happiness, and family ties. In many scenes, lanterns hang above doors, in narrow alleys, or across public squares. These places are often shown at night. The soft red light makes the lanterns stand out, both visually and emotionally. The glow gives off a sense of warmth, guidance, and lasting memory.

A strong example comes from the Lantern Festival episode. In one scene, people walk through a city street at night. Over their heads hangs a dense layer of red lanterns. Even with bright store signs and LED lights nearby, the lanterns dominate the image. This mix of modern and traditional creates a special mood. The repeating red lights turn a normal street into a festive space. The visual focus on the

lanterns gives viewers a feeling of shared joy and connection across time. Here, the red lantern is not just a symbol. It becomes a key sign that links everyday life with cultural memory (see Figure 3).

Figure 3. Lantern Festival – Symbolic Saturation in a Commercial-Cultural Space

This mix of old and new is seen in many parts of the series. The film often shows traditional signs next to modern places—like neon lights, city markets, or digital screens. But this is not shown as a clash. Instead, it shows how tradition fits into the present. Symbols are flexible. They change and move with time. They stay meaningful by becoming part of daily spaces and modern life. In this way, the film shows tradition as something alive and changing.

Besides man-made symbols, the documentary also uses natural signs. Things like rivers, plum blossoms, snow, and fire show up again and again. Each natural image has its own meaning. Fire stands for life, change, and spirit. Water means cleaning, reflection, and connection between generations. Plum blossoms show strength and fresh starts. These images do more than decorate the scene. They connect culture with nature and give meaning without words.

In many scenes, symbols are layered in one frame. A temple scene, for example, may show red lanterns, burning incense, and blooming flowers all at once. Each object adds a piece of meaning. Together, they form a full picture. These layers tell the viewer about beliefs, values, and ideas about the world. The film does not explain these ideas out loud. It lets viewers understand them by watching, feeling, and drawing on what they already know.

4.5 Interactive Strategies and Compositional Cohesion

Festive China uses a set of visual strategies to guide how viewers see and feel. These choices help build an emotional connection while showing cultural meaning. The approach follows the visual grammar model by Kress and van Leeuwen. In this model, images are treated as ways of communication. The documentary focuses on four main features: gaze, social distance, camera angle, and composition. Together, they shape a viewing experience that is both engaging and thoughtful.

Gaze is one of the clearest ways to connect with the viewer. Sometimes, people look straight into the camera. This is called a “demand” gaze. It creates a direct and personal feeling. But in most cases, the documentary uses the “offer” gaze. Here, people look at objects or at others within the scene—not at the viewer. This makes the viewer an observer. It adds a sense of distance, but still keeps them involved in the story.

Social distance also affects how viewers relate to what they see. The film moves between close-ups, medium shots, and long shots. Close-ups show details like faces, hands, or ritual items. These moments bring viewers closer and help them feel more connected. Medium and long shots, on the other hand, place people within a group or setting. These wider views highlight the shared nature of cultural life and community action.

A good example comes from the Mid-Autumn Festival episode. In one scene, a young girl reaches for a mooncake. The camera captures her hand in close-up, showing her excitement. She looks at the mooncake, not at the camera. This is a clear case of the “offer” gaze. The viewer watches quietly, not as part of the moment but as a witness. The warm light and soft colors add to the emotional tone. The girl’s action also becomes a symbol of family bonds and tradition passed down over time (see Figure 4).

Figure 4. Visual Interaction and Emotional Proximity in Cultural Context

Camera angle adds another layer to how scenes feel. Most shots are at eye level. This keeps the relationship between viewer and subject equal and calm. Sometimes, the film uses low angles to show rituals or important acts. These shots add weight and meaning. High angles appear less often. When they do, they open up space or bring a quiet, thoughtful mood.

Composition helps organize what the viewer sees. Following visual grammar rules, things on the left often stand for what is known. Things on the right bring in new meaning. Objects placed in the center

are usually the most important. In the mooncake scene, the pastry sits right in the middle of the frame. It draws the girl's gaze—and the viewer's. It becomes more than a snack. It becomes a cultural sign that links everyday moments to long-standing traditions.

The film also uses framing devices like moon gates, round tables, and ceremonial arches. These features help shape space and hold the viewer's attention. They mark where ritual happens and reflect traditional design. Often, the film shows these symbols next to modern buildings or city streets. But it does not show this contrast as a problem. Instead, it presents tradition as something that continues inside today's world. It moves forward while staying rooted.

5. Conclusions

5.1 Summary and Interpretation of Key Findings

This study shows that Festive China uses a clear and planned visual style to share cultural meaning. At the level of representation, the film often shows actions. These include ritual events, food preparation, and public gatherings. These scenes present tradition as something people do in real life—not just an idea. Along with these, the film uses scenes that show meaningful places and objects. Sacred spaces, familiar symbols, and traditional settings help build a strong sense of place and history.

Symbolic images add more depth to the story. The film often shows seasonal and natural signs. Red lanterns, snowflakes, plum blossoms, and fire appear in many episodes. These symbols point to a deep link between nature and culture. They suggest that tradition grows with the seasons and reflects a repeating cycle of life.

The film also guides how viewers connect with what they see. It uses changes in camera angle and distance to shape the viewer's feelings. Low-angle shots give a sense of power or importance. High angles create space and show reflection. Close-ups bring the viewer close to people or objects. Long shots or drone views pull the viewer back and create distance. These changes help move between watching and feeling.

The way each shot is built also matters. Framing, focus, and color are used to guide attention and build meaning. Together, these choices shape how the story unfolds and how tradition is seen on screen. The use of blurred transitions and layered visuals creates a poetic and flowing aesthetic. This style draws inspiration from traditional Chinese visual art. The color palette—dominated by red, green, and yellow—enhances both emotional impact and cultural resonance. Each color carries symbolic meaning, reinforcing values such as joy, renewal, and vitality.

Overall, the analysis confirms that Kress and van Leeuwen's visual grammar effectively explains how Festive China communicates meaning through both visual content and formal design.

5.2 Contributions and Implications

This study makes several key contributions to the field of multimodal discourse analysis. On a theoretical level, it expands the application of visual grammar from static images to moving-image narratives. While much prior research has focused on visuals in textbooks or advertisements, this study

demonstrates that the three metafunctions—representational, interactive, and compositional—are equally applicable to documentary film. Through motion, sequencing, and symbolic layering, film enhances the expressive potential beyond what still images can offer.

From a pedagogical standpoint, the documentary Festive China provides valuable content for teaching visual literacy and cultural understanding. Annotated segments can be integrated into classroom settings in disciplines such as language education, media studies, or intercultural communication. Students can learn how to interpret visual narratives, recognize symbolic meanings, and reflect on image-based communication. This approach supports the development of both multimodal awareness and cross-cultural competence.

In terms of communication, the documentary acts as a cultural mediator. It presents Chinese traditions in a visually engaging way that does not depend heavily on verbal explanation. For local audiences, it affirms cultural identity. For international viewers, it provides accessible insight into Chinese festivals.

Methodologically, the study introduces a structured framework that combines ELAN-based time-aligned annotation with a theory-driven coding scheme. This model allows for both qualitative analysis and basic quantitative validation, supporting future comparative research across media and cultures.

5.3 Limitations and Directions for Future Research

While the study provides meaningful insights, it has several limitations. First, the focus is limited to visual modality. The role of sound, narration, and music—each critical to meaning-making in documentary—was not analyzed. Future research should adopt a broader multimodal framework that considers how these modes interact to shape interpretation and emotional tone.

Second, this study does not include data on audience reception. It does not address how viewers interpret, respond to, or engage with the documentary's visual strategies. Audience studies using interviews, surveys, or eye-tracking could help explore how meaning is actually received and negotiated by different viewer groups.

Third, the annotation process, while systematic, involved subjective judgments. Classifying symbols or evaluating shot composition inevitably reflects the researcher's perspective. Although inter-coder reliability was tested and found high, future research could benefit from collaborative annotation models or semi-automated tools to reduce bias and increase replicability.

Finally, the study focuses on a single documentary series rooted in Chinese culture. Its findings may not fully apply to other genres or cultural settings. Comparative studies across national traditions or themes—such as diaspora, urbanization, or globalization—could help broaden our understanding of how documentary form and visual rhetoric mediate cultural meaning globally.

In sum, Festive China illustrates how visual modality serves not just aesthetic purposes but also functions as a powerful tool for cultural storytelling. Through representational richness, symbolic depth, and compositional elegance, the series communicates cultural values to diverse audiences. This study reaffirms the importance of visual discourse in cross-cultural communication and demonstrates the

analytical potential of visual grammar in understanding complex, multimodal texts.

Acknowledgement

The author sincerely thanks Guangzhou College of Commerce for its support. The school's helpful environment and useful resources made this research possible.

References

Alyousef, H. S. (2016). A multimodal discourse analysis of international postgraduate business students' finance texts: An investigation of theme and information value. *Social Semiotics*, *26*(5), 486-504.

Anthony, L. (2005). AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom. In *IPCC 2005* (pp. 729-737). IEEE.

Aufderheide, P. (2007). *Documentary film: A very short introduction*. Oxford University Press.

Baldry, A., & Thibault, P. J. (2006). *Multimodal transcription and text analysis: A multimedia toolkit and coursebook*. Equinox Publishing.

Barsam, R. (2013). *Looking at movies: An introduction to film* (4th ed.). W. W. Norton & Company.

Bassam, R. (2013). *Records and reality: Criticism of non-fiction films in the world* (W. Yawei, Trans.). Yuanliu Publishing House.

Bateman, J. A. (2014). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. Palgrave Macmillan.

Bateman, J. A. (2014). *Text and image: A critical introduction to the visual/verbal divide*. Routledge.

Bateman, J. A. (2022). Visual semiotics in cinematic storytelling: Exploring the role of color and composition. *Journal of Film Semiotics*, *17*(4), 78-96.

Bateman, J. A. (2023). *Multimodal discourse analysis: Theories and methods*. Cambridge University Press.

Bateman, J. A., & Schmidt, K. H. (2021). *Multimodal film analysis: How films mean*. Routledge.

Bednarek, M. (2022). *Text, discourse, and corpora: Theory and analysis*. Bloomsbury Academic.

Bie, J. Q. (2022). Multi-modal construction of new media discourse—Taking *Festival China–Spring Festival* as an example. *Journal of Hanjiang Normal University*, *42*, 1-5.

Bie, X. (2022). Multimodal discourse analysis of cultural documentaries: A study of *Exploring China: Culinary Adventure*. *Journal of Language and Culture*, *39*(2), 56-72.

Bill, N. (2017). *Introduction to documentary*. Indiana University Press.

Bordwell, D., & Thompson, K. (2022). *Film art: An introduction* (12th ed.). McGraw-Hill Education.

Chahine, I. C. (2022). Semiotic systems in contemporary documentary filmmaking: An analysis of narrative and visual modes. *Semiotica*, *234*, 45-67.

Chen, L., & Wang, J. (2023). Balancing tradition and modernity in *Festive China*: A focus on the Spring Festival and Lantern Festival. *Cultural Studies Quarterly*, *21*(2), 98-110.

Chen, L., Zhao, Q., & Liu, S. (2023). AI-enhanced translations and adaptive storytelling in *Festive China*: Expanding global accessibility and fostering cross-cultural engagement. *Journal of Multilingual Communication*, *19*(4), 91-107.

Chen, X. (2021). Reconsidering the visual dynamics of Chinese festivals: The case of composition. *Journal of Chinese Visual Culture*, *14*(2), 145-158.

Chen, Y., & Huang, G. (2009). Multimodal construal of heteroglossia: Evidence from language textbook. *Technology Enhanced Foreign Language Education*, *11*, 35-41.

Deng, J. J. (2023). A multimodal discourse analysis of posters based on visual grammar: The 19th Asian Games Hangzhou 2022. *International Journal of Linguistics*, *15*(6), 165-173.

Ding, J. S. (2020). *1.4 Billion We Are China*: Multimodal discourse analysis of CCTV public service advertisements. *Popular Literature and Art*, *23*, 118-119.

Djonov, E. (2007). Website Hierarchy and the Interaction between Content Organization, Webpage and Navigation Design: A Systemic Functional Hypermedia Discourse Analysis Perspective. *Information Design Journal*, *15*(2), 144-62.

Dong, M., & Yuan, X. L. (2021). Constructing a framework for multimodal aesthetic criticism discourse analysis. *Foreign Language Teaching*, *42*(1), 77-82.

Donnell, M. 2008. Demonstration of the UAM Corpus Tool for Text and Image Annotation. Pp. 13-16 in *Proceedings of the ACL-08 (Ed.): HLT Demo Session*.

Duan, Y. (2019). Multimodal meaning construction in documentaries: An analysis based on systemic functional grammar and visual grammar. *Discourse & Communication*, *14*(3), 265-283.

Eggins, S. (2004). *An introduction to systemic functional linguistics* (2nd ed.). Continuum.

Feng, D. Z., & Zhao, X. F. (2017). Multimodal metonymy and image textual meaning construction. *Journal of Foreign Languages*, *6*, 8-13.

Feng, D., & O'Halloran, K. L. (2021). Systemic functional multimodal discourse analysis of Chinese television advertisements. *Discourse & Communication*, *15*(2), 189-213.

Geng, J., & Chen, Z. (2014). The dynamic multimodal discourse analysis of documentaries. *Journal of Xi'an International Studies University*, *4*, 24-28.

Guo, Y. Q., & Li, R. (2020). Visual grammar in practice: Analyzing discourses with words and images of COVID-19 pandemic in China. *International Journal of Social Science and Education Research*, *3*(9), 10-18.

Halliday, M. A. K., & Hasan, R. (1989). *Cohesion in English*. Longman.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge.

Hong, X., & Duan, C. (2020). The construction of Sichuan image under multimodal visual grammar: Taking the documentary *Aerial China (Sichuan)* as an example. *Open Journal of Social Sciences*, *8*, 108-120.

Hu, Z. (2007). Multimodalization in social semiotics. *Language Teaching and Linguistic Studies*, *1*, 1-10.

Huang, L. H. (2015). Corpus 4.0: Multimodal corpus construction and its applications. *Foreign Languages Bimonthly*, *38*(3), 1-7.

Iedema, R. (2003). Multimodality, resemiotization: Extending the analysis of discourse as multi-semiotic practice. *Visual Communication*, *2*(1), 29-57.

Jewitt, C. (2009). *The Routledge handbook of multimodal analysis*. Routledge.

Ji, W. K. (2021). Multimodal collaborative reconstruction of foreign translation and communication of cultural promotional films. *Foreign Language Education*, *42*(5), 82-86.

Jia, Y. Y. (2021). Multimodal discourse analysis of urban image construction—Taking the "Best Jinan" city image promotion video as an example. *Modern Communication*, *19*, 94-96.

Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.

Kress, G., & van Leeuwen, T. (2002). Colour as a semiotic mode: Notes for a grammar of colour. *Visual Communication*, *1*(3), 343-368.

Kress, G., & van Leeuwen, T. (2020). *Reading images: The grammar of visual design* (3rd ed.). Routledge.

Lemke, J. (2002). Travels in hypermodality. *Visual Communication*, *1*(3), 299-325.

Li, X., & Lu, Y. (2012). The most successful analytical framework for extending functional grammar to visual morphology. *Journal of Applied Linguistics and Language Research*, *5*(2), 89-105.

Liu, J. (2017). A review of foreign multimodal corpus construction and related research. *Foreign Language Education*, *38*(4), 40-45.

Liu, Y. (2019). Visual grammar framework under multimodal discourse analysis: A study of English listening and speaking textbooks. *Journal of Shanxi Datong University (Social Sciences Edition)*, *33*(3), 99-101.

Martin, J. R. (2002). Fair trade: Negotiating meaning in multimodal texts. In C. Patrick (Ed.), *The semiotics of writing: Transdisciplinary perspectives on the technology of writing* (pp. 311-337). Brepols.

Martinec, R. (2000). Types of process in action. *Semiotica*, *130*(3-4), 243-268.

Martinec, R., & Salway, A. (2005). The image–text relation: A functional analysis of semiotic resources. *Visual Communication*, *4*(3), 337-363.

Nichols, B. (2017). *Introduction to documentary* (3rd ed.). Indiana University Press.

O'Halloran, K. L. (2004). Visual semiosis in film. In K. L. O'Halloran (Ed.), *Multimodal discourse analysis: Systemic functional perspectives* (pp. 109-131). Continuum.

O'Halloran, K. L. (2008). Systemic functional-multimodal discourse analysis (SF-MDA): Constructing ideational meaning using language and visual imagery. *Visual Communication*, *7*(4), 443-475.

Qin, Z. M. (2025). Constructing multimodal discourse meaning in Chinese cultural videos based on large model annotation: A case study of “Chopsticks: it’s more than just a utensil!” *Journal of Humanities and Social Sciences*, *1*(5), 57-65.

Royce, T. D. (2020). Intersemiotic complementarity: A framework for multimodal discourse analysis. *Text & Talk*, *40*(5), 585-602.

Thibault, P. J. (2000). The multimodal transcription of a television advertisement: Theory and practice. In A. Baldry (Ed.), *Multimodality and multimediality in the distance learning age* (pp. 311-385). Palladino.

Tian, H. L., & Pan, Y. Y. (2018). From meaning to intention: New development of multimodal discourse analysis to multimodal critical discourse analysis. *Shandong Foreign Language Teaching*, *39*(1), 23-33.

Unsworth, L. (2006). Towards a metalanguage for multiliteracies education: Describing the meaning-making resources of language-image interaction. *English Teaching: Practice and Critique*, *5*(1), 55-76.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556-1559). European Language Resources Association.

Wu, J. G., Li, D. Q., & Zhang, H. S. (2021). A study on multimodal translation of the explanatory notes of *Beautiful China* and the construction of national image. *Shandong Foreign Language Teaching*, *42*(5), 31-41.

Wu, S. (2022). Multimodal construction grammar research: Theoretical motivations, research framework, and developmental prospects. *Journal of Beijing Second Foreign Languages Institute*, *44*(2), 96-108.

Yang, Y. J., & Chen, L. (2024). A multimodal discourse analysis of the film *Lighting Up The Stars* from the perspective of visual grammar. *Modern Linguistics*, *2024*(124), 52-60.

Zhang, J. Y., & Jia, P. P. (2012). A few thoughts on visual grammar. *Contemporary Foreign Language Studies*, *(3)*, 38-42.

Zhang, L. 2018. The Interplay between Sound and Visuals in Documentaries: A Multimodal Discourse Analysis Perspective. *Discourse Studies Quarterly*, 9(3), 111-29.

Zhao, S., & Djonov, E. (2020). A social semiotic approach to multimodal discourse: Intersemiotic relations in multisemiotic texts. *Journal of Pragmatics*, *155*, 65-78.

Zhu, Y. S. (2007). Theoretical foundations and research methods of multimodal discourse analysis. *Foreign Language Research*, *5*, 82-86.