Article

A Study on the Meaning Construction of Digital Narrative Discourse of Learners of English in Universities: Taking Holiday-Themed Multimodal Videos as an Example

Xuan Wang [1,*] and Mengmeng Ji [1]

[1] Ningxia University of Foreign Languages, Yinchuan, China
* Correspondence: Xuan Wang, Ningxia University of Foreign Languages, Yinchuan, China

Abstract: Learner digital storytelling transforms given narrative content into dynamic multimedia works, a pedagogical approach which has demonstrated a highly positive impact on contemporary English language teaching. Drawing extensively on the Visual Grammar theoretical framework proposed by Kress and van Leeuwen (2020), this empirical study systematically examines foreign language videos published on the official WeChat account of a university located in Northwest China. It conducts a rigorous multimodal discourse analysis across three primary dimensions—representational meaning, interactive meaning, and compositional meaning—to comprehensively explore how university English learners construct complex discourse meaning specifically at the visual level. The analytical findings reveal three key insights: 1) At the level of representational meaning, the utilized images frequently lack sufficient narrative strength; 2) At the level of interactive meaning, the overall meaning construction tends heavily toward the objective transmission of factual information; 3) In terms of compositional meaning construction, visual salience is notably insufficient across the sampled media. Therefore, while the visual modality in the short videos created by these learners generally serves an adequate communicative function, it functions primarily as a supplementary tool to the dominant textual modality, meaning the holistic construction of visual meaning remains incomplete. By analyzing bilingual videos focused on traditional Chinese festivals, this study explores the nuanced characteristics of learners' multimodal expression, offering valuable pedagogical insights for language teachers. Furthermore, it helps learners gain a deeper understanding of Chinese history, customs, and cultural heritage, ultimately laying a robust foundation for effective cross-cultural dissemination.

Keywords: digital storytelling; multimodal discourse; visual grammar; english teaching; discourse analysis

Received: 25 April 2026
Revised: 31 May 2026
Accepted: 12 June 2026
Published: 17 June 2026

1. Introduction

Multimodal discourse analysis refers to the analysis of multisign-based discourse that engages a single sense or multiple senses. The concept was initially explored in theoretical frameworks and has since been developed by scholars from various fields who have analyzed diverse materials to establish analytical methods and frameworks for multimodal discourse studies. Researchers have also examined the application of multimodal discourse analysis theory, conducting systematic reviews of its origins, developmental history, and future research directions [1]. This has sparked significant interest in multimodal discourse research both domestically and internationally.

In recent years, research on multimodal discourse can be broadly categorized into three types: the integration of multimodal discourse with translation, its combination with cultural communication or cross-cultural studies, and its application in teaching or second language acquisition. Within the teaching context, some researchers have explored the synergy among modalities in classroom settings through teachers' use of multimodal discourse. For example, studies have analyzed meta-discourse in online lectures to

identify recurring patterns and relationships between these elements and paralinguistic and kinesthetic elements. Other research has used teaching videos to explore the semantic synergy between scaffolding language in English instruction and gestures, body postures, and facial expressions. However, only a few studies have examined the construction of meaning in learners' multimodal discourse. For instance, some analyses have focused on how university learners create meaning through the interaction between images and texts in English as a Foreign Language contexts [1, 2]. Other research has identified intersemiotic synergy in students' linguistic, paralinguistic, and visual modalities, such as reading written texts aloud from slides, oral citations, animations, and word tracking through pointing gestures and pointer movements. In China, relevant research has primarily explored the characteristics and changes in students' narrative texts at a qualitative level, while studies on learners' current multimodal narrative practices remain scarce, particularly regarding the narrative abilities of university students.

In light of this, this study draws on visual grammar frameworks to conduct an in-depth exploration of the characteristics of learners' multimodal narrative discourse. The findings aim to provide insights for teachers seeking to reform multimodal narrative instruction, promote the development of learners' multimodal narrative skills, foster a deeper understanding of Chinese culture among learners, and lay the groundwork for future cultural dissemination.

2. Theoretical Framework

Based on the study of visual grammar, this research examines three dimensions: representational meaning, interactive meaning, and compositional meaning (see Figure 1).

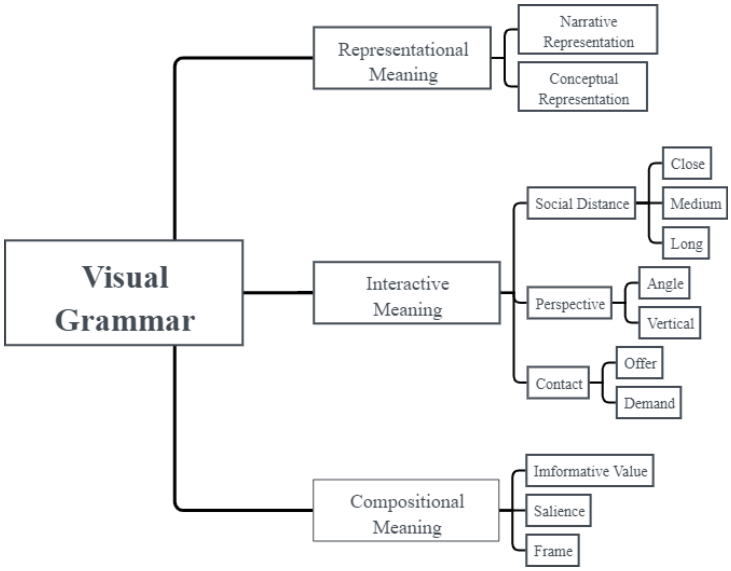

Figure 1. Visual Grammar

Representational meaning refers to the visual perception and understanding of the world, which can be distinguished into narrative representation (involving vectors) and conceptual representation (without vectors) based on the presence or absence of "vectors." The former encompasses behavioral and reactive processes, while the latter includes analytical, classification, and symbolic. The sources of meaning in symbolic can be further categorized into two types: symbolic suggestion and symbolic attribute. Classification involves hierarchical relationships, and visual grammar also highlights the importance of implicit classification [3]. Analytical representation emphasizes the relationship between parts and the whole. Narrative representation primarily involves action and reaction, where the vectors form diagonal lines connecting the actor or reactant to the target object.

Interactive meaning refers to the interaction between viewers and the subjects in an image, which is primarily achieved through three strategies: contact, social distance, and

perspective. Contact can be further categorized into offer and demand based on whether eye contact exists between the subjects and the audience. Social distance implies the closeness or distance between the subjects and the viewer through the framing scale; the greater the distance, the lower the viewer's sense of involvement [4, 5]. Perspective conveys subjective attitudes and is divided into horizontal and vertical categories.

Compositional meaning is constructed through information value, salience, and framing. Information value organizes elements and assigns them specific informational significance [6]. Salience endows elements with varying degrees of visual appeal, often achieved through techniques such as foregrounding and typography. Framing connects or separates elements within an image using solid or dashed lines.

3. Research Design

This section discusses the research questions, corpus selection, and research methods in sequence [6].

3.1. Research Question

This study primarily examines how learners construct discourse meaning at the visual level, addressing three specific questions [7].

What are the characteristics of multimodal discourse meaning construction at the representational level?

What are the characteristics of multimodal discourse meaning construction at the interactive level?

What are the characteristics of multimodal discourse meaning construction at the compositional level?

3.2. Corpus Selection

A certain school's video account has published 160 videos, primarily featuring stories related to idioms, Chinese solar terms, poetry, and legends. Among these, there are 27 videos on legends and historical anecdotes, 51 on poetry, 53 on solar terms, and 16 on festivals; each video averages about 5 minutes in length. Due to space limitations, this study selected seven videos related to festival-themed content as the corpus for analysis, specifically Dragon Boat Festival, Qixi Festival, Mid-Autumn Festival, Ghost Festival, Kitchen God Festival, Double Ninth Festival, and Spring Festival, to explore the characteristics exhibited by the school's learners in the construction of discourse meaning at different levels. The selected video corpus offers sufficient processability and analytical depth in terms of visual content, providing rich data support for the multimodal discourse analysis in this study.

3.3. Research Methodology

This study employs a mixed-methods approach combining quantitative and qualitative methods, utilizing the Elan 6.8 software to conduct multimodal discourse analysis of the videos. First, by annotating participant attributes in Elan, the video data is categorized into three parent layers: Representational Meaning, Interactive Meaning, and Compositional Meaning [8]. Under Representational Meaning, using Elan's participant attributes, two branches were established as Narrative Representation and Conceptual Representation. Narrative Representation was further subdivided into action and reaction by creating sub-levels; Conceptual Representation was similarly subdivided into three sub-levels: classification, analytical, and symbolic. Interactive Meaning and Compositional Meaning were directly subdivided into sub-levels. Finally, the videos were annotated, and data on annotation frequency and overall proportion were collected.

All annotation work was performed independently by a single annotator. To ensure the reliability of the annotation results, a second round of independent annotation was conducted one month after the first round was completed. Two months after the second round was completed, the same annotator performed a third round of re-annotation for verification. The consistency rate between the three rounds reached 90%.

4. Research Findings

This section presents descriptive statistics for the three dimensions, detailing the proportions of each category.

4.1. Distribution of Representational Meaning

After statistically analyzing the distribution of the representational meaning, the following table was generated:

As shown in Table 1, symbolic (31.92%) accounted for the highest proportion. This indicates that, in the videos, learners are adept at conveying abstract meanings through symbols, metaphors, or cultural imagery. Action (26.69%) ranks second after symbolic. Visual narratives centered on action emphasize what the subjects are doing, highlighting the daily habits and traditional customs of the people during the festival. Classification (18.99%) and analysis (12.43%) have similar proportions, both lower than action. Reaction (9.97%) focuses on the gaze, facial expressions, and emotional responses of characters or objects, but is used less frequently in the selected videos.

Table 1. Distribution of Representational Meaning

	Category	Frequency	Proportion (%)
Narrative Representation	Action	423	26.69%
	Reaction	158	9.97%
Conceptual Representation	Classification	197	12.43%
	analytical	301	18.99%
	symbolic	506	31.92%

4.2. Distribution of Interactive Meaning

The distribution of interactive meaning across different dimensions is presented in the table below, specifically covering three dimensions: contact, social distance, and perspective.

According to visual grammar, there is no inherent superiority between offer and demand. Learners flexibly utilize either type depending on the context to express needs or fulfill contextual requirements [9]. Offer aims to convey objective information, while demand focuses on establishing an emotional connection with the audience through vectors. As shown in Table 2, the number of offers in the selected videos is lower than that of demands, accounting for only 6.54% of the total interactive meanings, while demands account for 93.46% of the total. This distribution indicates that learners prioritize the description of objective facts in the process of multimodal description.

Table 2. Distribution of Interactive Meaning

	Category	Frequency	Proportion (%)
Contact	Offer	143	93.46%
	Demand	10	6.54%
Social Distance	Long	210	32.76
	Medium	139	21.68
	Close	292	45.55
Perspective	Frontal-angle	333	49.55
	Oblique-angle	339	50.45
	top-down	130	9.91
	Eye level	436	68.13
	bottom-up	74	5.64

In terms of social distance, close shots accounted for the highest proportion, reaching 45.55%. This was followed by long shots, accounting for 32.76%. Medium shots accounted for 21.68%. The concept of shot size in the visual system originates from interpersonal distance theory, enabling viewers to mentally close the distance between participants in the image. Consequently, as visual grammar suggests, long shots are used to convey public distance—a sense of remoteness—and to present a panoramic perspective, capturing the overall character of natural or urban landscapes. Close shots, by emphasizing a visual focal point at close range, magnify and highlight details that viewers should focus on, conveying emotional tones or more specific meanings [2, 10].

From a horizontal perspective, the frequency of oblique and frontal shots is similar, though the former is slightly higher, accounting for 50.45% of horizontal shots. Shots taken directly from the front constitute the horizontal frontal perspective, accounting for 49.55%. The vertical perspective reflects power hierarchies, while the horizontal perspective signifies equality among participants, with no power differences. According to Table 2, horizontal shots account for 68.13% of the total, representing the highest number of shots. The top-down perspective often diminishes an individual's sense of presence, reducing them to a state of mediocrity on a moral level. In the videos, the top-down perspective accounted for 20.31% of annotations, significantly lower than the level perspective. The bottom-up perspective typically conveys a sense of superiority, grandeur, and triumph in the individuals depicted. However, it had the lowest number of annotations in the videos, accounting for only 11.56%.

4.3. Distribution of Compositional Meaning

The meaning of composition involves three key dimensions: contact, social distance, and perspective, which are distributed as follows:

As shown in Table 3, learners' videos exhibit a tendency toward center-periphery composition, which accounts for as much as 67.36% of the total. In this composition, the core element is placed at the center, while the peripheral areas contain secondary content. Meanwhile, the frequency of use for the Given-New composition and the ideal-reality composition is relatively similar, accounting for 16.47% and 16.17%, respectively. In the Given-New composition, information is distributed from left to right, with elements on the left typically conveying known information and those on the right presenting new content, guiding the audience's gaze to the right. In the ideal-reality composition, information is distributed from top to bottom, with elements at the top conveying idealized information and those at the bottom leaning more toward realistic information.

Table 3. Distribution of Compositional Meaning

	Category	Frequency	Proportion (%)
Information	Given-New	111	16.47
value	Ideal-Reality	109	16.17
	Center-Periphery	454	67.36
Salience	Size	57	5.81
	Color contrast	160	16.31
	Perspective	699	71.25
	Position	65	6.63
Frame	Connection	219	55.44
	Separation	176	44.56

Salience is determined based on visual cues. Viewers of a spatial composition can judge the relative importance of various elements. The more important an element is, the greater its salience. As shown in Table 3, perspective is the primary method used in the video to convey the salience of objects, accounting for as much as 71.25%. This technique highlights foreground objects by blurring the background. Color contrast was annotated

less frequently than perspective, accounting for 16.31%. Position and size accounted for 6.63% and 5.81%, respectively. In terms of position within visual images, elements become more significant when shifted toward the top or to the left.

The frame divides or connects elements within the image, thereby defining whether they belong to a single whole. As shown in Table 3, connections in the video are primarily achieved through vectors formed during the narrative representation process or through the arrangement of elements in the background, accounting for 55.44% of the total—a higher proportion than segmentation. This indicates that video imagery places greater emphasis on presenting an overall visual impression. Segmentation, on the other hand, accounts for 44.56% and is manifested through actual borders, blank spaces between elements, and discontinuous colors [11, 12].

5. Discussion

The videos selected for this study primarily explore three aspects in the construction of meaning: the origins of festivals, folk culture, and dietary habits. Learners consistently placed the origins of festivals at the beginning of the videos. This reflects the characteristics of production, daily life, and culture in ancient society, laying the groundwork for the subsequent sections on dietary habits and folk culture. Additionally, it serves as an introduction to Chinese culture. By gaining an understanding of history and culture, learners are better able to develop cultural awareness. Furthermore, by visually presenting relevant customs, the video serves as a concrete manifestation of the inheritance of traditional culture. Dietary habits are deeply rooted in everyday life; they not only reflect people's attitudes toward life but also highlight regional cultural differences. Through vivid demonstrations, the video enhances the effectiveness of cultural dissemination. In the process of visual meaning construction, learners exhibited the following characteristics of representational meaning.

5.1. Characteristics of Representational Meaning

In the process of constructing representational meaning, learners' video images exhibit a narrative strategy characterized by conceptual representation as the primary focus, supplemented by narrative representation. Figure 2 illustrates the typical characteristics of this symbolic implication process. Set against the backdrop of the Qixi Festival, the visuals feature a unified color palette, with details intentionally downplayed to create a profound atmosphere of traditional culture. Under this visual strategy, by anchoring the abstract concept of prosperity to tangible festive imagery, the images rapidly activate the audience's cultural consensus regarding national prosperity and people's well-being without the need for textual explanation.

Figure 2. Qixi Festival

Since symbolic meanings are shaped by specific sociocultural contexts, the abstract nature of the imagery in this narrative strategy may pose comprehension challenges for international audiences, resulting in weak visual storytelling [13, 14]. For example, in the videos Mid-Autumn Festival and Dragon Boat Festival (Figure 3), when the topics of "prayers, wishes, and homesickness" are raised, the creators use images of sky lanterns to convey visual meaning. However, the practice of lighting sky lanterns for blessings is a custom unique to China, and it is not easy for overseas audiences to associate sky lanterns

with making wishes. Furthermore, sky lanterns are not a cultural custom of the Mid-Autumn Festival but rather a traditional activity of the Lantern Festival. When learners misinterpret cultural traditions, it can also lead to misunderstandings among the audience.

Figure 3. Mid-Autumn Festival

However, some learners performed exceptionally well in this process. For example, in the case of the Kitchen God Festival, learners were able to use both action and reaction to express deep, abstract meanings through concrete, superficial behaviors, thereby reducing the difficulty of comprehension. When explaining the close relationship between the Kitchen God Festival and the Spring Festival, learners repeatedly linked Spring Festival-related scenes, such as paper cutouts, fireworks, lanterns, and preparing for the New Year (as shown in Figure 4) with action processes and reaction processes. The facial expressions of the people in the images served as reactions, externalizing the emotional state of joy and enhancing the festive atmosphere. This visual strategy transformed abstract symbolic processes into perceptible narratives.

Figure 4. Kitchen God Festival

5.2. Characteristics of Interactive Meaning

In the process of constructing interactive meaning, all seven sample videos primarily relied on three visual strategies, namely close-up shots, long shots, and oblique angles, with their frequency of use significantly higher than that of other options [15, 16]. These four visual strategies all emphasize the expression of objective meaning, indicating that the learners' videos tend to present information in a rational manner. As shown in Figure 2, the use of long shots and a bird's-eye view places the audience in a symbolic position of "elevated status," presenting characters, architecture, and space through a panoramic composition. The frame deliberately compresses the participants to the bottom of the screen, blurring individual identities and private interactions, and instead highlights the social and collective nature of the Qixi Festival as a public cultural event. At the same time, the shot is presented from a horizontal oblique angle, placing the audience in the position of a "bystander" and implying that they do not belong to the social world depicted within the footage, thereby further enhancing the objective presentation of the information.

Although learners place importance on conveying objective information, they rarely apply all three of the above visual strategies simultaneously. Instead, they combine them with techniques that bring the audience closer—such as close-ups, low-angle shots, and eye-level shots—to prevent the video from lacking interactivity. Their focus is primarily

on the objects involved in the food and cooking process, rather than on the people. Taking the video Mid-Autumn Festival as another example, numerous close-ups focus on mooncakes and osmanthus cakes. As shown in Figure 5, although the horizontal angle is an oblique perspective, the learner adopts a level perspective that places the audience and the objects in the frame on an equal footing. This choice of perspective helps create an interactive atmosphere, making the audience feel as though they are on the same level as the scenes or characters in the frame, while close-ups enhance the sense of immersion and engagement in the viewing experience. This aligns with findings that when the topic involves food, students tend to adopt visual strategies that bridge the gap between themselves and the audience (As shown in Figure 6).

Figure 5. Mid-Autumn Festival

Figure 6. Double Ninth Festival

5.3. Characteristics of Compositional Meaning

Regarding the construction of compositional meaning, findings indicate that learners adopt a central composition strategy when the topic involves food, whereas they employ a peripheral composition strategy when the topic involves tourism. In contrast to these findings, learners in this study tended to place key information at the center of the frame and used perspective to blur background details, thereby enhancing the salience of the objects. As shown in Figure 6, the two elderly individuals are positioned at the center of the frame, while the background is blurred to reduce visual distractions, further highlighting the subject's characteristics. This approach not only enhances the image's sense of depth but also makes the core content more distinct. At the same time, the figures' movements in the image form vectors that organize the various elements into a cohesive whole, achieving the goal of creating connections within the composition and guiding the audience to understand and perceive the deeper meaning the image intends to convey. This compositional approach not only enriches the image's content and adds depth but also enhances the viewer's experience.

However, in the learners' videos, some scenes lack sufficient visual salience, a finding consistent with research on visual composition [17, 18]. Take Figure 7 as an example, which features an overall dark color tone. Although the learner attempted to use lighting to direct the audience's attention toward the central participants and their actions, the participants occupy too small a portion of the frame, making it difficult to achieve the desired visual salience.

Figure 7. Mid-Autumn Festival

5.4. Comprehensive Analysis Across Three Dimensions

During the video production process, the learners actively gathered festival-related materials, demonstrating a keen interest in the cultural theme. Considering all three dimensions, although learners' performance was somewhat weak in certain sub-levels, overall, the parent levels were able to coordinate effectively and complement one another, thereby compensating for deficiencies across the various levels. Learners were able to select different strategies based on the video content, deciding whether to focus on objectively conveying information or enhancing interactive effects, demonstrating their potential and ability in overall conceptualization and expression. However, the connection between the video content and the festival theme was not sufficiently strong, and there was a lack of cross-cultural awareness. Specifically, in the presentation of images symbolizing the process, some representations were based solely on Chinese cultural contexts, which may be difficult for audiences from other cultural backgrounds to understand (as shown in Figure 3).

Furthermore, while learners have made significant progress in terms of cultural awareness and relevant knowledge, they still need to move toward the goal of embracing cultural differences and striving to engage and influence their audience, ultimately applying the skills they have learned to innovatively achieve a balance between meeting audience needs and expressing their own values [2].

6. Conclusion

From a Visual Grammar perspective, this study examines the characteristics exhibited by learners in the process of constructing discourse meaning at the visual level across three dimensions: representational meaning, interactive meaning, and compositional meaning. The findings reveal that: 1) In the seven videos created by the learners, the images are largely rooted in a local cultural context, making them difficult for audiences from other cultural backgrounds to understand. Additionally, the videos rely heavily on symbolic processes, resulting in weak narrative coherence. 2) In terms of interactive meaning construction, the videos tended to lean toward objective statements, reflecting an equal relationship between the images and the audience; however, the interactivity of the videos could be enhanced through other visual strategies. 3) In terms of compositional meaning construction, the sense of prominence is notably weak. The focal point of the frame is not sufficiently highlighted, making it difficult to guide the audience's attention, which results in an overall suboptimal visual effect. Therefore, while the videos produced by learners are often capable of conveying meaning, their overall meaning construction remains incomplete and can only serve as a supplementary tool to the textual level.

Based on the above findings, this study offers the following recommendations for multimodal narrative instruction: 1) Since many of the same scenes appear repeatedly across the videos in the learners' submissions, instruction should focus on improving students' efficiency in retrieving information, synthesizing data, and producing content. 2) Multimodal narrative instruction should be grounded in authentic contexts and serve genuine purposes, with narrative themes closely tied to students' lives. For instance, this study focuses on traditional festivals. On the one hand, learners possess a certain

familiarity with this topic; on the other hand, the process of gathering information allows them to gain a deeper understanding of traditional history and culture, thereby enhancing their interest. 3) Teachers should encourage students to post short videos on video-sharing platforms, allowing them to develop multimodal narrative skills through practice, thereby significantly boosting their enthusiasm and ability to effectively tell stories.

However, this study also has its limitations, such as a relatively small corpus and a lack of in-depth exploration of the overall characteristics of learners' narrative discourse meaning-making at the linguistic level and in terms of intermodal relationships. Therefore, future research could conduct an in-depth analysis of learners' multimodal narrative abilities from other dimensions and examine their developmental patterns from a diachronic perspective.

Funding: This paper was supported by the Department of Education of Ningxia Hui Autonomous Region Higher Education Scientific Research Project (NYG-2024-052).

References

1. C. Jewitt, Ed., *The Routledge handbook of multimodal analysis*, vol. 1. London: Routledge, 2009.
2. K. O'Halloran, "Mathematical and scientific forms of knowledge: A systemic functional multimodal grammatical approach," in *Language, Knowledge and Pedagogy: Functional Linguistic and Sociological Perspectives*, London: Continuum, 2007, pp. 205–236.
3. S. Norris, "Multiparty interaction: A multimodal perspective on relevance," *Discourse Studies*, vol. 8, no. 3, pp. 401–421, 2006.
4. J. Lemke, "Multimodal genres and transmedia traversals: Social semiotics and the political economy of the sign," *Semiotica*, vol. 173, no. 1, pp. 283–297, 2009.
5. D. Lindenberg, "Modes and intersemiotic cohesion in student presentations performed online: An SF-informed multimodal discourse analysis," *English for Specific Purposes*, vol. 69, pp. 67–79, 2023.
6. R. Barthes, *Elements of semiology*. Macmillan, 1977.
7. E. M. Mestre-Mestre, "The construction of meaning in the Second Language Classroom. A multimodal discourse analysis," *Procedia-Social and Behavioral Sciences*, vol. 173, pp. 228–233, 2015.
8. E. Bernad-Mechó, "A multimodal discourse analysis of linking metadiscursive elements in two opencourseware lectures (MOOCs)," *Procedia-Social and Behavioral Sciences*, vol. 212, pp. 61–66, 2015.
9. G. Kress and T. Van Leeuwen, *Reading images: The grammar of visual design*. Routledge, 2020.
10. A. S. Urbieta and E. A. Peñalver, "Multimodal discourse in digital storytelling: An assessment tool proposal," *Computer-Assisted Language Learning Electronic Journal*, vol. 22, no. 2, pp. 14–25, 2021.
11. Y. F. D. Yang, "Multimodal composing in digital storytelling," *Computers and Composition*, vol. 29, no. 3, pp. 221–238, 2012.
12. S. Meier, "Digital storytelling: A didactic approach to multimodal coherence," *Frontiers in Communication*, vol. 7, Art. no. 906268, 2022.
13. C. G. Signes, "Digital storytelling and multimodal literacy in education," *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, no. 22, pp. 237–250, 2014.
14. M. Y. Liang, "Beyond elocution: Multimodal narrative discourse analysis of L2 storytelling," *ReCALL*, vol. 31, no. 1, pp. 56–74, 2019.
15. Z. M. Chemodurova, "The art of storytelling in the digital age: A multimodal perspective," *Vestnik Volgogradskogo Gosudarstvennogo Universiteta. Seriya 2. Yazykoznanie*, vol. 21, no. 6, p. 110, 2022.
16. I. Alonso, S. Molina, and M. D. Porto, "Multimodal digital storytelling," *Multimodality and Cognitive Linguistics*, vol. 78, p. 147, 2015.
17. P. Bou-Franch, "Multimodal discourse strategies of factuality and subjectivity in educational digital storytelling," *Digital Education Review*, vol. 22, pp. 80–91, 2012.
18. I. Alonso, S. Molina, and M. D. P. Requejo, "Multimodal digital storytelling: Integrating information, emotion and social cognition," *Review of Cognitive Linguistics. Published under the auspices of the Spanish Cognitive Linguistics Association*, vol. 11, no. 2, pp. 369–387, 2013.