Original Research Article

A Multimodal Study on Discursive Strategies of Live Commerce Anchors

ABSTRACT

At present, discourse analysis is widely used in news reports and TV talk shows in China, but there is a lack of discourse analysis research on live streaming. Against this backdrop, this study adopts a multimodal discourse analysis approach, drawing upon the comprehensive theoretical framework for multimodal discourse analysis and Kress and Van Leeuwen's seminal visual grammar theory as the theoretical basis. Our research aims to investigate the intricate interplay between anchors and the amalgamation of diverse modalities in the process of constructing their image.Furthermore, we endeavor to analyze potential strategies that can be employed to enhance the efficacy of live broadcasting and augment viewers' engagement and consumption. In this paper, we decompose the multimodal discourse of the selected live video of "*Oriental selection*" into verbal and non-verbal modes, and analyze the process and role of how anchors use multimodality to form interactions and promote construction.

Keywords:Live streaming; multimodal discourse analysis;discourse strategy;Visual Grammar;Oriental selection; Elan.

1. INTRODUCTION

Since the so-called "first year of live e-commerce" in 2019, the live broadcast of goods has shown explosive growth [1]. During the "Douyin New Year Festival Promotion" period in 2023, the platform mall scene led to a year-on-year increase of 308%, and the search scene also led to a year-on-year increase of 124%. Payment of GMV（Gross Merchandise Volume）increased by 79% compared with the 22nd New Year Festival.Live streaming is essentially a kind of

economic behavior with multimodal language interaction as the core pivot for the purpose of reaching commodity transactions [2]. In the live streaming room, live-streamers promote the goods to their customers through the use of linguistic devices and with the assistance of their nonverbal strategies such as facial expression, gestures, and background, which directly affects the retention rate in the room and sales volume of certain goods. Both live streamers and audiences can act as message senders (sources) and message receivers (hosts), decoding and feeding back the messages, forming a double-loop interaction. This interactive process is presented or stored in the form of video, which is a typical dynamic multimodal discourse, containing linguistic features such as accent, voice tone, intonation, physical features such as gesture, eyes, expression, body posture, and environmental features such as goods, background, equipment, and network platform [3].

Nowadays, the live trade industry has become more and more standardized and mature, and live streamers have created a large number of “phenomenal” cases of selling goods.However, the multimodal language interaction of live commerce, as a new phenomenon of network language life, has rarely been analyzed from the perspective of multimodal discourse analysis [3].Consequently, we still lack an explicit understanding of the process of live streamers how to construct promotional discourse, and how to utilize multimodal resources to promote audience consumption.

Against this background, this study aims to investigate multimodal discourse strategies of live streamers to reveal their multifaceted identities from linguistic and non-linguistic perspectives. Specifically, we aim to analyze the linguistic and non-linguistic features of positive discourse used by broadcasters during live streaming sessions. Investigate the persuasive strategies employed to create a positive and engaging environment for viewers. Explore the relationship between positive discourse and the formation of social connections, emotional responses, and user engagement. Provide insights into the potential implications of positive discourse in live streaming for communication theory, online interaction, and social influence. Specifically, it tends to address the research question: How is the multimodal interaction of live streaming presented and functioned? The framework is able to incorporate both sides of the discourse

into the study and make the corpus analysis concrete and three-dimensional, so as to reveal the use and configuration of multimodality in live commerce and how the speaker uses a communicative strategy composed of multimodal elements to guide the listener into the practice of shopping.

2. MultimodalAnalysis of Videos

Although multimodal discourse analysis involves many language schools, the most suitable theoretical model is systemic functional linguistics, because it does not need to transform the theoretical framework itself to adapt to new purposes [4]. Therefore, this paper is mainly based on the theory of systemic functional linguistics to carry out multimodal discourse analysis.

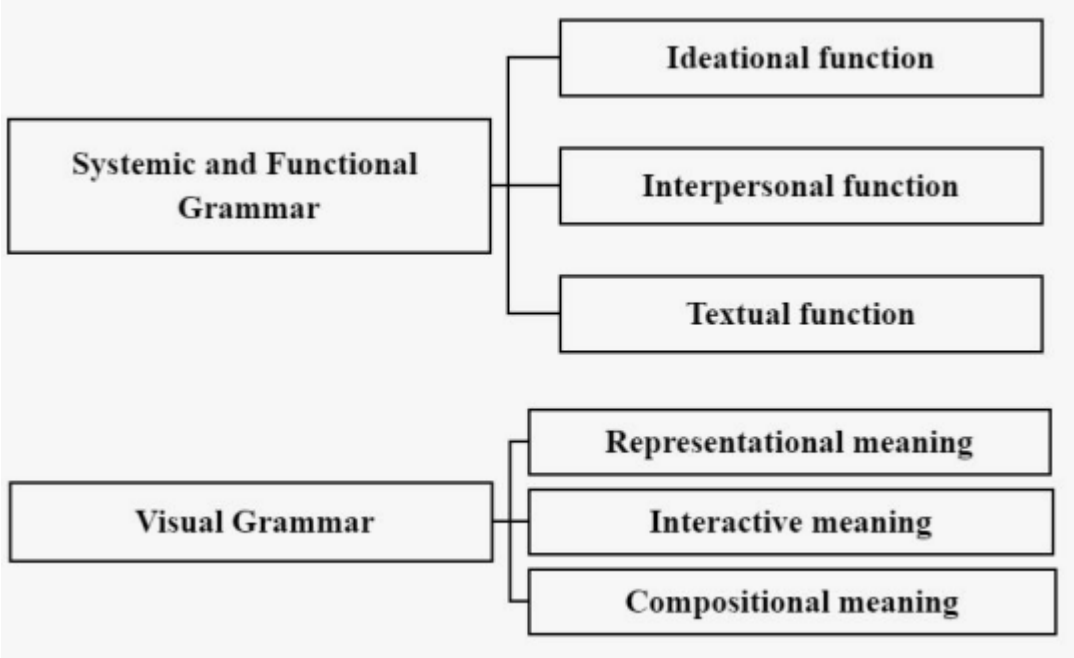
Halliday [5] regards language as a symbolic resource, and thinks that symbols are used to generate and express meaning and realize discourse communication. Martin [6] further developed this theory and applied it to discourse analysis to put forward an analytical framework, which mainly includes five levels: culture, context, meaning, form and media. Zhang Delu, a domestic scholar, further applied Martin's theory to foreign language teaching and put forward a comprehensive theoretical framework of multimodal discourse analysis. Since then, Zhang Delu, a domestic scholar, has further explained the theory and constructed a comprehensive framework of multimodal discourse analysis theory. This framework is divided into four levels: cultural level, context level, content level and expression level [4].

The Visual Grammar of Kress and Van Leeuwen presents how the described elements(such as people, places and things) are combined in a more complex and extensive wayinvisual statements. This is exactly the same as Halliday's Systemic Functional Grammar, which presents how language elements combine in clauses, sentences and discoursestoconstruct meanings. As Kress and Van Leeuwen [7] believe, Halliday's threemeta-functions are the starting point for describing images, not only this model is suitablefor language, but also it serves as a good source for thinking about all representationalmodels. They hold that all semiotic models encode three meanings at the sametime, namely ideational meaning, interpersonal meaning and textual meaning. Like all othersemiotic models, visual images meet

the requirements to the relationship of representingthe experience world, producers and recipients, interacting between participants andaudiences represented in visual design, and arranging the composition of visual resources [7].

Consequently, the interpersonal function, ideational function and text functioninHalliday's systemic functional grammar are correspondingly referred to as the interactivemeaning, representational meaning and compositional meaning in Kress andVanLeeuwen's Visual Grammar. (As Fig.1. shows)

***Figure 1.*Systemic functional grammar and Visual grammar**



3. DATA AND METHODS

In this study, the Douyin live account"东方甄选(Oriental Selection)"are selected as a representative of the industry live streaming, and select its live video (11 hours and 51 minutes) of the "New Year's Cargo Festival" on February 9, 2023 as the analytical material to establish a typical dataset. During the "Double Eleven" e-commerce promotion festival in 2022, Oriental Selection, which ranked first in the GMV ranking of Douyin live broadcasts for four consecutive months, generated average daily sales of 25 million to 50 million yuan.

At present, the mode of online live broadcast with goods tends to be mature, and the audience and consumers are tired of aesthetics. As one of the live streamers of Oriental Selection，Dong

Yuhui is famous for his bilingual delivery and his unique personal style. Since then, the Oriental Selection studio has received great attention. At present, theOriental Selection studio has attracted the attention of various domestic media, including China Youth Daily, Xinhua Daily and Nanfang Daily.The positioning of the live broadcast room is the live broadcast of knowledge-based agricultural assistance with goods. Based on the carefully controlled investment selection, combined with the anchor's extensive knowledge structure and humorous conversation, the live broadcast room confidently expresses Chinese culture and produces a video with a strong sense of network, which effectively attracts users to stop, and has achieved high web traffic and high income, becoming a model of knowledge-based live broadcast.

In this paper, we take a quantitative and qualitative approach to analyze the live streaming corpus.We use the multimodal analysis software ELAN 6.6 to annotate the selected live video corpus and output the data so that the annotatable video corpus and the corresponding text corpus are created; then the modal coding is performed according to the verbal modality and non-verbal modality (including kinesthetic modality, visual modalityand auditory modality) (Table 2).Finally, we quantitatively and qualitatively analyze the kinds, densities and layouts of each modality in the live streaming video, and summarize their application rules.It should be noted that the video corpus in this paper belongs to the live replay video, which is difficult to annotate even if it is captured in the form of video recording due to the long duration of a live broadcast and the short time the audience's real-time interactive information stays on the interface. However, this “disappeared” information may not be fully reflected in the anchor's words, and sometimes the anchor may use the strategy of repeating audience questions or answering directly to provide feedback. Therefore, this paper can still continue to study from this perspective.

4. RESULTS

4.1 VerbalAnalysis

Verbal modality refers to the discourse used by communicators in the process of interaction.

The verbal modality of live video with goods contains phonological features, special terms, sentence class features and discourse strategies, among which discourse strategies have significant advantages, so this paper takes this as an example to do a specific analysis, see Table 1 for details.

Table1. Verbal Modality (Discourse Strategies) for Live Streaming Videos

Layers of content	**Number of labels**	**Average length oflabeling/s**	**Total labeling time/s**
Conversational history	49	13.2	646.8
Audience attraction	72	4.6	331.2
Repeatition	104	2.4	681.6
Immersive hawking	193	5.8	1119.4
Audience Interaction	336	6.8	2284.8
Layered narrative	389	58.4	22717.6
Performance	9	48.6	437.4
Total	1332	139.8	28218.8

As seen in Table 1, live broadcasting with goods builds an interactive relationship between “people-goods-field”, and the content of its live broadcasting discourse forms a fixed paradigm centered on “goods”: previewing goods → introducing goods (appearance, usage, etc.) → emphasizing key points (price, giveaways, etc.) → selling goods (launching links, urging purchase, etc.). (shape, usage, etc.) → emphasize the key points (price, gifts, etc.) → selling goods (launching links, urging to buy, etc.). The discourse expression effect of the live video with goods depends on the comprehensive use of the following six discourse strategies.

Conversational history means tracing back and responding to the topics that have been mentioned by the two sides, providing feedback on past information and triggering new interactions.Gavin [8] found that speakers would use retrospective narration to create new discourse segments, and that retrospective narration would only occur in the interactional context of the listening recipient. Live streaming interactions also rely on conversational history to take place.For example, the anchor replied to a comment posted by a viewer on the air: steak? When I sold steak before people said it was good, and just because you guys

mentioned steak, I made a special trip to find out by reviewing the new talk wheel that leads to the goods.

Audience attraction is to accumulate popularity for the live broadcast by inviting fans to join, purchase and share, such as guiding the audience to increase user stickiness and attention.For example, the anchor puts out this call to action: everyone helps me click on the followers and join our fan club. Double-tap the screen and give me some likes.

Repeat refers to the repetition of key information to stimulate the desire to buy, such as the example of the information on the gift of high-density output, and accompanied by high speed, high volume, to create a “buy or lose” consumer sentiment. For example, the anchor said: 249 yuan, today there is only one size. Yes, there is only one specification, as long as 249 yuan will be sent to you with the jar and so many Chenpi, the whole dried with half a catty, dried with half a catty ah!In live discourse, the role of repetition is not limited to the establishment of an initial connection, but is also highlighted in the “reading of comments” phase, where the anchor selectively reads and repeats the messages presented on the interface as a form of feedback to the audience, which is usually expressed as participatory listening or acknowledgement listening. When viewers ask for information about a certain product in the comment section, the anchor will repeat the question and answer.

Immersive hawking isthe continuation of the traditional yelling style of offline selling, accompanied by close display of the goods or ringing a bell to attract high attention. The essence of live broadcast with goods is a marketing to achieve the purpose of “selling goods” through “bringing goods”, which still continues the traditional shouting practice in the real society. The roar of the anchor group not only created a warm live broadcast atmosphere, but also strengthened the interaction with the audience. Immersive hawking highlights the appeal. The anchor appeals to the audience to buy by displaying or knocking on the goods at close range, holding up the brand of goods quota, and accompanying the uniform slogan of “link up” or “start snapping up”. Call words are usually loud, fast and information-intensive, creating a tense shopping atmosphere and trying to persuade people to buy.

Audience interaction means through the social discourse in the live broadcast to build a sense of intimacy and trust with the audience who are not present, to realize the leap from commodity consumption to emotional consumption, for example, the use of the terms “babies” and the buzzwords “family members” and “sisters” strengthens the sense of community through linguistic symbols.live streaming with goods is obviously different from other multi-modal videos. The general vlog is similar to an asynchronous communication, while the live broadcast is a real-time interaction, and the interaction between the speaker and the listener is stronger[3].

The Layered narrative is to reintroduce the sense of community through storytelling. For instance, before the anchor introduces Chenpi(Orange peel), he quotes the poem “One tael of Chenpi is one tael of gold, and one hundred years of Chenpi is better than gold”. Poetic discourse are used to create a real situation, to make up for the virtual live field due to the absence of the body brought about by the sense of indirectness to achieve the commodity symbols of emotional value of the leap. The highly infectious language description is the key to attract the audience to pay attention to the goods, which is mainly manifested as follows: a. A large number of adverbs of degree+adjectives are used, such as “super beautiful”, “very good”, “very good reputation in the industry”, “extremely beautiful” and “absolutely beautiful”. These words which are forbidden in the advertising law are presented here with exaggerated effects, and the word “very” has become the highest frequency adverb. B high-frequency use of online buzzwords or unique adjectives, such as “pure desire”, “atmosphere”, “rich” and “retro”, to describe a certain lipstick as “really fairy!” Describing a coat as “a retro Hong Kong girl” is used to emphasize the effect of product use.

Performance are tactics used by anchors to increase the number of people in the studio and to attract the interest of viewers, and the Oriental Selection Studio has become popular with the distinctive performances of the anchors. They use singing, reciting and storytelling to popularize knowledge, and the audience will buy and consume after enjoying the performance and increasing their goodwill towards the live streamers and the live brand.

4.2 Nonverbal Analysis

4.2.1 KinestheticAnalysis

Kinesthetic modality can be divided into head, gesture and posture language and other symbolic resources[9]. Due to the limitations of the “vertical screen” dialog box of smartphones, most of the live streaming videos involve facial expression and gesture language, this paper selects the more frequently used gesture language for specific analysis, see Table 2.

Table 2. Kinesthetic Modality (Gesture language) for Live Streaming Video

Gestures	Number	Average length /s	Total /s
Metaphorical gestures	106	1.9	201.4
Numeric gestures	208	2.3	478.4
Iconic gestures	226	3.6	813.6
Beat gestures	322	3.9	1255.8
Deictic gestures	405	2.8	1134
Total	1267	14.5	3883.2

As seen in Table 2, the gesture language in the live video includes the following five types: “metaphorical gestures”、 “Numeric gesture”、“Iconic gesture” 、 “Beat gesture” and “Deictic gesture”[10].

Metaphorical gesture refers togesture was used to describe abstract ideas or categories.For instance, when explaining “bad fruit is guaranteed, buy with confidence”, the anchor spread the palm of the hand and hold it firmly in the air, which indicating that the audience should “feel at ease”. The gesture of stability refers to the stability of the mind.

Numeric gesture means using gesture to represent quantities or price.For example anchor use gestures to suggest prices when talking about the price of 1RMB, using a finger to represent the number 1.In addition, the anchor will also use numeric gestures to indicate the number of days, for example, using a gesture of 2 fingers to indicate that the good will only take 2 days to be delivered.

Iconic gesture is similar to numeric gesture,iconic gestures use gestures to represent shapes. For example, when the anchor describes the outer packaging of a mooncake box, he puts his hands together and then opens them, indicating the "double-door" style of the box.

Beat gesturerefers to using gestures to reinforce Rhythmic gestures. For instance, the live streamer rhythmically tapsher fingers to follow the accents as she explaining the flavor of oranges in order to highlighting product information.

Deictic gestures are used to refer to specific things, highlighting the product information with gestures.For example, when introducing the item steak, the anchor points a finger at the steak being fried on the pan.

4.2.2 Visual Analysis

Visual modality is the symbolic resources obtained by stimulating the visual nerve, specifically by videos, objects, special effects and display boards to create intuitive consumer scenes, see Table 3.

Table 3. Visual Modality for Live Streaming Video

Layers of content	Number	Average length/s	Total/s
Vidoe	15	14.7	220.5
Material object	289	10.6	3063.4
Special effects	9	323.2	2908.8
Display board	41	42.0	1722
Total	354	390.5	7914.7

Short videos have made “vertical screen” a new form of social interaction. Although the 9:16 ratio of the “screen” as a framework for dialogue limits the scope and space of communication to a certain extent, verticalization is more in line with the visual habits of the contemporary generation. “It has become the most convenient, comfortable, and customary short video viewing option for users, and a new form of audiovisual that is in line with the language, thinking, and behavioral patterns of the mobile Internet generation”[11].With narrower bezels

and detailed structural settings that are more conducive to focusing attention. The following is an example of the interface of the Oriental Selection live broadcasting room, which draws on the analytical framework of visual grammar to analyze how the screen composition induces interactions between people “inside” and “outside” the screen in terms of the meanings of reproduction, interaction and composition.

Representational meaning: Representational meaning means that the image represents or reproduces the relationship between people, places and events [12]. The categorical structure of live public screens usually presents a modal configuration of ‘image + text’. The image mode is the main mode, and the text mode is the auxiliary mode. The former includes the image of the anchor, the background of the live broadcast room and the interactive icon, while the latter includes the comment part in the lower left corner of the screen and the text embedded in the background, which describes the possible environment of real interaction. The analyzed structure can be divided into bearers and several features.The bearers, i.e., all the building blocks of the live broadcast room, include the live broadcast background, sharing icons, liking icons, shopping cart icons, and dialogue boxes; the features can be manifested in the form of content, i.e., different anchors, different speech contents, etc.

Interactive meaning: Interactive meaning refers to the relationship between the image viewer and the image world, namely “the listener” and “the speaker”. Here, “the speaker” is the direct performer of the image, namely the anchor, and everything he says and does belongs to the dynamic elements of constructing a long video.Interactive meaning consists of distance, point of view and contact. In terms of distance, the vertical screen intercepts only the upper half of the anchor's body, and this type of close-up excludes superfluous information and highlights the anchor itself. It is a visual representation of the speaker's initiative to get close to the listener.The point of view reflects the power relationship between the speaker and the hearer, and the hearer views the image at a level, which is a dialogue gesture of equality. Contact refers to the construction of an imagined relationship in which the speaker looks directly at the listener, an image known as a solicited class of images, which corresponds to solicited speech acts in functional grammar [13], implying that the anchor wants to get feedback from the

listener from the comment section and passes on this signal to the listener, encouraging them to speak interactively in the comment section.

Compositional meaning: Information value, framing and salience are important compositional resources.Kress and van Leeuwen [14] argue that the elements on the left and right of an image correspond to the structure of “known-new information” and the elements on the top and bottom correspond to the structure of “ideal reality”, but that certain contexts lead to different results.On the live public screen, the elements on the left and right are not symmetrically arranged. This does not fit the structure of “known-new information”, but the upper and lower elements present an “ideal-real” structure, and the upper part is a generalized substance and the lower part is real, more practical information. Salience refers to the degree of attracting the audience's attention. In the eyes of previous researchers, the information above is the most significant part. However, on the public screen, the information below is more significant, the actual information such as goods on the shelves and dialogues that the listener pays attention to following the speaker's guidance.

5. DISCUSSION

Head anchors are considered authorities in the live streaming industry, and they use verbal strategies such as hyperbole and emotionally provocative expressions when presenting their products to successfully engage viewers. David Myers [15] points out that persuasion has a “peripheral path”, when the person being persuaded is distracted by the message and does not pay attention to whether the argument is convincing or not, familiar and easy-to-understand expressions are more persuasive. For example, when the anchor says “all the girls”, “let's go”, “listen to me, 3-2-1”and other words, these call to order for the audience has the infectious force of collective carnival, the audience clicked on the link to implement the purchase, the persuasion has been completed [16]. Therefore, the interactive activity of live streaming with goods is essentially a persuasive behavior, and its core intention is to achieve audience purchase. The vertical screen of the live broadcast viewing experience and unique camera framing brings the communication distance between the anchor and the audience closer. The inducing prompts in the live broadcast, such as “someone is buying” and “like”,

give the audience an immersive shopping experience. Anchor's humorous words can create a warm dialogue atmosphere, and in the close social situation of live broadcasting, the anchor's words have a significant impact on the audience's purchasing.

6. CONCLUSION

In this study, we combine a comprehensive theoretical framework for multimodal discourse analysis with visual grammar theory to analyze the configuration and use of multimodal resources in a live streaming scene in order to better illustrate that modality, which is traditionally used as a “paralanguage”, may also play a more important role in communication than verbal communication, and to understand aspects such as on-screen and off-screen human interactions from a multimodal point of view [1]. The study also aims to understand inter-personal interactions from a multimodal perspective, going beyond verbal discourse to understand modality's role in conveying meaning in discourse. In terms of methodology, this study is not limited to the traditional analysis of social interaction, but is based on the dynamic discourse analysis of multimodal annotated corpus; in terms of analytical framework and mode, a new multimodal discourse analysis framework is proposed, which not only can include both sides of the discourse in the scope of the study, but also can concretize and three-dimensionally analyze the corpus; in terms of the object, in the past, focusing on the analysis of the effect of the live streaming with the audience, and seldom explored the process of the interaction. In terms of the object, the previous focus on the effect of live broadcasting to bring goods, rarely exploring the interaction process, and how anchors interact with the audience in a multimodal way to facilitate the shopping behavior is the focus of this paper; in the scope of the study, it is not confined to a single linguistic dimension, but rather, it will include the three types of multimodal features of the language, the body, and the environment.

This paper establishes a dynamic corpus of Oriental Selection live streaming videos and examines the basic situation of their multimodal application. Through quantitative and qualitative analyses, it illustrates the use and configuration of multimodal symbolic resources in the live streaming video, and provides innovative strategies to enhance the effect of multimodal discourse expression. However, due to the non-singular linear relationship

presented by multimodal discourse expression, the complex hierarchical structure of verbal modality and the annotation analysis of non-verbal modality at the sensory level are the difficulties of data processing in the early stage of this paper, and other data analysis techniques can be applied to further deepen the research in the future.

REFENCES

1. Huang Chuxin, Wu Mengyao. The development status, existing problems and optimization paths of live streaming in my country. Media. 2020;17: 11-14. Chinese.

2. Li Yanping. The establishment and impact of trust relationships in live broadcast delivery. Young Journalists, 2021;8: 44-45. Chinese.

3. Wang Yubo, Pan Danting. Multimodal language interaction in live streaming. Language Strategy Research, 2022;7(3):34-46. Chinese.

4. Zhang Delu. Exploration of the comprehensive theoretical framework of multimodal discourse analysis. Chinese Foreign Languages, 2009;6(1): 24-30. Chinese.

5. Halliday M A K, R Hasan. Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective. Australia: Deakin University Press; 1985.

6. Martin J R. English Text: System and Structure. Amsterdam: JohnBenjamins; 1992.

7. Kress G, T van Leeuwen. Reading Images: The Grammar of Visual Design (2nd edn.). London: Routledge; 2006.

8. Gavin L. Towards a green applied linguistics: Human-Sea turtle semiotic assemblages in Hawai'i. Applied Linguistics. 2019;41(6), 922–946.

9. Shi Lei, Chen Shi. Understanding new consumption: the connotation, causes and practical rules of media consumption. Journal of Southwest University for Nationalities (Humanities and Social Sciences Edition), 2022;43(9): 133-140. Chinese.

10. Norris S. Analyzing Multimodal Interaction. London: Routledge; 2004.

11. Tang Miao, David. Research on the interactive behavior of e-commerce Internet celebrity live broadcasts from the perspective of interpersonal communication. New Media Research, 2020;6(20):1-3+7. Chinese.

12. Kress G, T van Leeuwen. Reading Images: The Grammar of Visual Design. London: Routledge; 1996.

13. Li Zhanzi. Social semiotic analysis of multimodal discourse. Foreign Language Studies, 2003;(5):1-8+80. Chinese.

14. Kress G, T van Leeuwen. Front pages: (The critical) analysis of newspaper layout. In A. Bell & P. Garrett (Eds.), Approaches to Media Discourse, 186–291. New Jersey: Blackwell; 2000.

15. Myers David G. Psychology with Updates on DSM-5. Macmillan Higher Education; 2014.

16. Wang Zheng. Research on the analysis model of multi-modal video discourse. Journal of Northeast Normal University (Philosophy and Social Sciences Edition), 2013;(01):105-108. Chinese.