

## *Original Paper*

# A Multimodal Discourse Analysis of Interpersonal Meaning in New-Type Farmers' Agricultural Promotion Short Videos on Douyin

Yixin Chen<sup>1</sup>

<sup>1</sup> Harbin Institute of Technology, Weihai, Shandong 264209, China

Received: February 12, 2026

Accepted: April 22, 2026

Online Published: May 7, 2026

doi:10.22158/eltls.v8n3p21

URL: <http://dx.doi.org/10.22158/eltls.v8n3p21>

### ***Abstract***

*This study selects 15 short videos published over the past year by new-type farmers (xin nong ren) on Douyin as research corpus. It adopts systemic functional grammar's interpersonal meaning theory, visual grammar's interactive meaning theory, and intersemiotic complementarity theory as analytical framework, using AntConc and ELAN multimodal video analysis software as research tools. This study employs both quantitative and qualitative methods to examine how agricultural promotion short videos construct interpersonal meaning through multimodal resources. The analysis focuses on three aspects, including verbal mode, visual mode, and their synergistic complementarity. According to the findings, in the verbal mode, declarative sentences account for the highest proportion in the mood system. Median-value and low-value modal expressions reach 92.8% in the modality system, which helps present information in a relative neutral and tentative manner and contributes to shaping a credible image of new-type farmers and maintaining audience trust. In the visual mode, demand image, medium close-up and medium shots, horizontal angle, and high-modality color expressions are commonly used. These visual choices effectively reduce the social distance between creators and audiences, support a more equal interpersonal relationship and encourage emotional engagement. In terms of intersemiotic complementarity between verbal and visual modes, attitudinal convergence, attitudinal divergence, and appellative reinforcement can all be observed. Among them, attitudinal convergence occurs most often, where verbal and visual elements work mutually to reinforce interpersonal meaning, improve coherence and optimize the audience's viewing experience. This study enriches empirical research on multimodal discourse analysis in the context of agricultural promotion short videos, and provides implications for the interactive design and content production of such videos.*

### ***Keywords***

*New-type farmers' agricultural promotion short videos, multimodal discourse analysis, interpersonal meaning, intersemiotic complementarity*

## 1. Introduction

Multimodal discourse refers to communication that makes use of several sensory channels such as hearing, vision, and touch, together with different semiotic resources like language, images, sound, and gesture (Zhang, 2009). This field develops from Halliday's systemic functional linguistics. Language is understood to realize three metafunctions, including ideational function, interpersonal function, and textual function (Halliday, 1978). On this basis, Kress and van Leeuwen (1996) extended these metafunctions to the study of images and put forward the idea of visual grammar (Kress, van Leeuwen, Kress, & van Leeuwen, 1996). Later, Li (2003) introduced multimodal discourse analysis into Chinese academia, which helped promote related research in China (Li, 2003). Zhang (2009) then worked on building a more complete analytical framework and examined how different modes relate to each other. The corpus of multimodal discourse analysis cover both static texts, such as textbooks and signs, and dynamic texts, including television programs and films. This broadens the scope of discourse analysis. It not only shows how the linguistic system functions in meaning exchange, but also explains how semiotic systems like images, music, and color take part in this process. In this way, discourse can be interpreted in a more comprehensive and accurate manner, and it becomes clearer how multiple modes are used in social communication (Zhu, 2007).

New-type farmers are individuals or groups who can use the internet to serve agriculture, rural areas, and farmers. They are equipped with new concepts, new ways of thinking, and new technologies. These people engage in the production, processing, and distribution of agricultural products, or provide services such as publicity, promotion, guidance, and consulting for agriculture (Ali Research Institute, 2015). The report of the 20th CPC National Congress points out the need to advance rural revitalization and speed up the building of an agricultural powerhouse. In this context, new-type farmers turn to short videos as a multimodal form to attract audiences, gather traffic, and reach consumers directly. Through this approach, agricultural products are connected with the market more efficiently, which helps increase farmers' income. During this process, the way interaction is designed in short videos affects how audiences receive and respond to the information, and it also influences the overall effect of agricultural promotion.

Based on this background and practical needs, this study focuses on agricultural promotion short videos created by new-type farmers on Douyin. From the perspective of multimodal discourse analysis, the study examines how these videos use different modes to construct interpersonal meaning and how they build emotional and cognitive connections with audiences. The study also aims to provide implications for the interactive design of such short videos so that they can better serve the development of the rural revitalization.

## 2. Literature Review

Drawing on relevant studies in multimodal discourse analysis, Xin (2008, p. 208) classifies the field into three major strands, including the social semiotics school, the interactional sociology school, and the cognitive linguistics school (Xin, 2008). The social semiotics school mainly focus on the generic features, intergeneric relations, and analytical methods of multimodal discourse, with applications extending to media texts, scientific discourse, mathematical discourse, textbooks, children's literature, advertising discourse, multimedia presentations, online communication, film and television works, video games, and other text types (Li & Lu, 2012). The interactional sociology school primarily concerns itself with social interaction in real-world situations, with corpora mainly drawn from naturally occurring interactive events such as classroom interaction, everyday conversations, and face-to-face communication in work contexts (Norris, 2004). Research within the cognitive linguistics school mainly concentrates on theoretical reviews and case analyses, with corpora encompassing political cartoons, television documentaries, print advertisements, video advertisements, posters, and speeches (Pan & Li, 2017).

In the field of short videos, studies from the perspective of multimodal discourse analysis can generally be divided into five areas, including public issues, cultural tourism promotion, science communication, media communication, and family life. In the area of public issues, Lucia Abbamonte (2021) examined the discourse strategies of short videos created by Greenpeace on YouTube and their official website (Abbamonte, 2021). Chandler (2023) analyzed popular TikTok videos under the hashtag #anxiety to explore how anxiety narratives are constructed through multiple modes (Mordecai, 2023). Raquel Martínez-Sanz and Amaia Arribas-Urrutia (2023) selected short videos under the #donasangre hashtag on TikTok as their corpus to examine the interpersonal meaning of blood donation-themed videos (Blood donors wanted: narrative innovation on TikTok to enable mobilization, 2023). From the perspective of cultural tourism promotion, visual grammar was combined with a multimodal construction model to analyze how tourism images are built in official Douyin videos from the Tibet Autonomous Region (Du & Cheong, 2025). A related study by Sang (2020) investigated the interactive meaning in Li Ziqi's Weibo short videos based on visual grammar (Sang, 2020). With regard to science communication, Ou et al. (2023) focused on the multimodal discourse features and visual persuasion strategies of short videos in science and technology journals, providing a reference for presenting research findings (Ou, Zhang, Liu, & Chen, 2023). Science communication videos by "Huazha Hua Xiaolao" on Bilibili were taken as the corpus in Zhou Shihan and Huang Wen's (2025) study, where multimodal narrative strategies were discussed in order to improve communication effectiveness (Zhou, & Huang, 2025). Media communication research is represented by Zhu Xiaoyu and Shi Yidan (2022), who analyzed discourse features of short videos produced by traditional media hosts within a multimodal framework (Zhu & Shi, 2022). Meaning construction in emotion-driven short video news was further examined through multimodal discourse analysis in Li Tingting's (2024) study. In the context of family life, Zhao (2025) applied multimodal theory to explore identity construction of

overseas Chinese parent influencers in Douyin videos (Zhao, 2025).

At present, multimodal discourse analysis research on short videos has paid relatively limited attention to agricultural promotion content. In addition, existing studies on agricultural promotion short videos tend to focus on verbal and visual modes, while quantitative analysis and discussions of multimodal synergy remain limited. Against this background, this study selects 15 agricultural promotion short videos posted on Douyin within the past year, each with thousands of likes. The analysis focuses on how interpersonal meaning is constructed through different semiotic resources, aiming to provide empirical support for multimodal discourse analysis and practical implications for the production of agricultural promotion short videos.

### 3. Theoretical Framework

Interpersonal metafunction mainly focuses on the realization of language in social interaction. In Systemic Functional Linguistics, Halliday (1985) believes that interpersonal meaning is mainly realized through the mood system and the modality system (Castro & Halliday, 1995). The mood system reflects the relationship between the speaker and the listener and the purpose of communication. In terms of communication function, discourse can be divided into giving and demanding. From the perspective of content exchange, it can be divided into information and goods and services. When these two dimensions are combined, four basic speech functions can be formed, including giving information, demanding information, giving goods and services, and demanding goods and services, which correspond to declarative, interrogative, offers and commands respectively. Modality is used to express the speaker's judgment and attitude towards propositions or behaviors. The modality system is generally divided into modalisation and modulation. Among them, modalisation involves probability and usuality, while modulation is related to inclination and obligation. At the same time, modality can also be divided into three levels, including high, medium and low according to the speaker's responsibility for the proposition, which reflect different degrees of commitment or obligation. The mood system and the modality system work together to construct social relations between speakers in the process of communication.

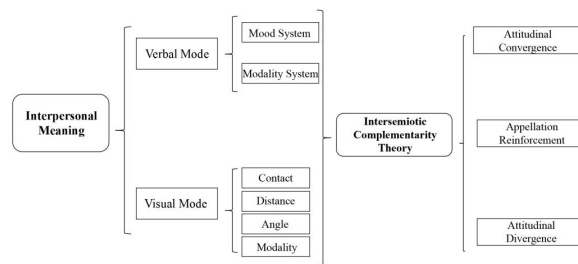
On this basis, Kress and van Leeuwen (1996) expanded the analytical framework at the visual level and proposed visual grammar, including representational meaning, interactive meaning and compositional meaning. The interactive meaning in visual grammar corresponds to the interpersonal meaning in language, mainly focusing on the relationship between image participants and viewers. This meaning is realized through the four resources of contact, distance, angle and modality. Among them, the contact depends on whether the character in the image forms eye contact with the viewer, which can be divided into demand and offer. The distance is related to the frame range of the picture, which is usually divided into close-up, medium close-up, medium shot and long shot, which are used to indicate different degrees of social distance. The perspective is analyzed from two dimensions, horizontal and vertical. The horizontal viewing angle includes frontal, oblique and back. The vertical perspective can

be divided into high angle, eye-level and low angle, which is used to present the power relationship between the viewer and the image participant. Visual modality is mainly realized through color presentation, including color saturation, differentiation, modulation and brightness, which will affect the realism and credibility of the image in perception. The framework of the interactive meaning is shown in Figure 1.

<b>Interactive Meaning</b>	<b>Contact</b>	Demand	Eye contact between image participant and viewer
		Offer	No eye contact between image participant and viewer
	<b>Distance</b>	Close-up shot	Only partial details of the subject appear
		Medium close-up shot	Only the subject's upper body (above chest) appears
		Medium shot	The subject appears from the knees up
		Long shot	The subject's full figure and surrounding scenery appear
	<b>Angle</b>	Horizontal angle	Camera shoots from the front, side, or back of the subject
		Vertical angle	Camera is positioned higher than, level with, or lower than the subject
	<b>Modality</b>	Color saturation	Ranges from fully saturated color to absence of color
		Color differentiation	Ranges from maximally diversified color range to monochrome
		Color modulation	Ranges from fully modulated color to simple, unmodulated color
		Contextualization	Ranges from no background to the most fully detailed background
		Representation	Ranges from maximum abstraction to finest pictorial detail
		Depth	Ranges from no depth to maximum depth perspective
		Illumination	Ranges from fully represented light and shadow to absence of light and shadow
Brightness		Ranges from maximum variation in brightness levels to only two brightness levels (black/white or dark/light grey), or a single color shifting from one brightness value to another	

**Figure 1. Theoretical Framework of Interactive Meaning (Kress & van Leeuwen, 1996)**

Royce analyzed the relationship between language and visual patterns and put forward the theory of intersemiotic complementarity. This theory is used to explain how different symbolic resources such as text and images interact in multimodal text and the complementary relationship between them. Different modes together act as meaning construction resources, forming a relatively coherent expression as a whole (Royce, 1998). Based on the three metafunctions of Systemic Functional Linguistics, Royce (2007) further divides intersemiotic complementarity into three categories: ideational, interpersonal and textual (Royce, 2013). Among them, interpersonal intersemiotic complementarity mainly focuses on how the mood and modality in language interact with visual contact, distance, angle and modality, and can be further divided into attitudinal convergence, attitudinal divergence and appellative reinforcement. The theoretical framework of this study is shown in Figure 2.



**Figure 2. Diagram of the Theoretical Framework**

#### 4. Research Design

Douyin has become an important platform for content production and the marketing of agricultural products. In 2025, Douyin e-commerce reported 10.2 billion orders of agricultural specialty products, and the total over the past three years exceeded 20 billion (Xinhua News Agency, 2025). Against this background, this article selects short videos created by new-type farmers on Douyin platform as research corpus, and adopts purposive sampling method to carry out research. Through the search bar, search for #newtypefarmers (新农人), #helpfarmers (助农), #newtypefarmersplan (新农人计划), #villageguardians (乡村守护人), #returntohometowntohelpfarmers (返乡助农) and other topic tags to screen relevant accounts, and then carry out manual screening. According to the basic screening criteria such as the number of fans and the amount of interaction of videos, 14 content creators are finally retained, whose accounts show stable audience engagement and continuous content production. By browsing the creator's homepage, select recent original agricultural product promotion short videos with relatively high audience engagement, and eliminate video works with incomplete data, non-original content and low relevance to agricultural product promotion.

This study takes the interpersonal meaning in Systemic Functional Grammar, the interactive meaning of visual grammar and the theory of intersemiotic complementarity as the analysis framework and discusses three issues. First, how verbal mode realizes interpersonal meaning in the short video of Douyin's new-type farmers' agricultural promotion. Second, how visual mode realizes interpersonal meaning in these videos. Third, how verbal and visual modes work together to construct interpersonal meaning. By addressing these questions, the paper examines how such videos connect with audiences, shorten social distance, improve audience recognition of agricultural products, and strengthen communication effects, in order to provide implications for improving farmers' income.

Both quantitative and qualitative methods are used in the study. For the verbal corpus, with the help of Wangyan OCR Extractor, the hard subtitles of the video were extracted and the text is then manually checked. This process produces a total of 660 clauses and AntConc was used to carry out text analysis. For the video corpus, ELAN 7.0 was used to set annotation tiers based on the research focus and mode features during playback. The software's statistical function then summarizes the annotation results and generates the data. At the level of text-image interaction, the analysis further integrates verbal and visual information in ELAN 7.0 to examine how they jointly construct interpersonal meaning.

#### 5. Verbal Mode Analysis

Systemic functional linguistics holds that the interpersonal metafunction of language refers to the use of language to express emotions and attitudes, construct interpersonal interaction, and establish interpersonal relationships. Mood and modality are the two main systems constituting the interpersonal meaning of the language system (Halliday, 1985). Accordingly, the following sections analyze the verbal mode of Douyin agricultural promotion short videos from these two systems.

### 5.1 Analysis of the Mood System

Within the systemic functional linguistics framework, Halliday (1985) identifies four primary speech functions in interpersonal communication, namely statement, question, offer, and command.

These functions reflect the different ways in which the speaker realizes the exchange of information and goods and services in communication. In terms of lexicogrammar, these functions are mainly realized through different types of clauses, including declarative clauses, interrogative clauses, imperative clauses and exclamative clauses. Declarative clauses are usually used to provide information, interrogative clauses are used to request information, imperative clauses are used to give instructions, and exclamative clauses are used to express strong emotions or subjective feelings. The statistical results of the mood system are shown in Table 1:

**Table 1. Statistical Results of the Mood System**

	Declarative	Imperative	Interrogative	Exclamative
Frequency	485	95	48	32
Proportion	73.5%	14.4%	7.3%	4.8%

In the analyzed short video corpus, a total of 660 clauses were identified, including 485 declarative clauses, accounting for 73.48%, 95 imperative clauses, accounting for 14.39%, 48 interrogative clauses, accounting for 7.27% and 32 exclamative clauses, accounting for 4.84%. The high-frequency use of these structures enables creators to present information such as planting conditions, production process and product characteristics objectively, which enhances the audience's sense of trust and plays a role in promoting the formation of purchase intention. The following selects some example sentences for interpersonal meaning analysis:

Declarative clauses:

(1) 我们最大的优势是使用水库水灌溉、使用有机肥料，而且没有使用有残留的农药。

Our greatest advantage lies in irrigation with reservoir water, the use of organic fertilizers, and the absence of pesticide residues.

(2) 像这样的一个冰淇淋甜瓜我们今天晚上就准备去直播了。

Sweet melons like this will be featured in our livestream tonight.

(3) 全国每五颗荔枝就有一颗来自我们茂名高州。

One out of every five lychees in China comes from Gaozhou, Maoming.

(4) 2025年10月底摘下的果子到现在还剩下20万斤没有卖出去。

The fruit picked at the end of October 2025 still has 200,000 jin left unsold.

(5) 这里的光照时间长，昼夜温差较大，所以口感也比较好。

The long hours of sunshine and large day-night temperature difference here contribute to a better

flavor.

The main function of the declarative clause is to provide information. Example (1) explains the planting advantages of Renjia Village sweet potatoes compared to other sweet potatoes on the market, introduces the growing environment by listing the production conditions, and gives a relatively objective response to the audience's questions about product quality. Example (2) presents information about the upcoming livestream and links the product display with the follow-up sales activities. Example (3) uses the declarative clause with numerical data to reflect the production scale of Gaozhou lychee, highlighting the prosperity of the Gaozhou lychee industry. Example (4) presents the pressure and urgency faced by Bapan pears' selling through specific time and quantity information, so as to guide the audience to pay attention to this problem. Example (5) explains the quality source of Minqin melon through causal explanation and explains the reason for the formation of its flavor. In general, declarative clauses in agricultural promotion short videos often convey information in combination with listing, numerical evidence, and causal explanations, which enhances the credibility of the content on the basis of a more direct tone and guides the audience to form a positive impression of the product.

Imperative clauses:

(1) 大家可以一起帮西和县的老乡们把这个梨销出去。

Everyone can help the villagers of Xianhe County sell these pears.

(2) 大家看一下这个地上的地膜。

Everyone, take a look at the plastic mulch on the ground.

(3) 今天跟着我的脚步带你们一块去看看吧。

Today, come with me and let's go take a look.

(4) 上车跟我一起去收瓜。

Hop in and come with me to harvest the melons.

Imperative clauses are mainly used to issue instructions, invitations or calls for action. Example (1) is the action appeal issued by the new-type farmer creator to the audience, inviting the audience to participate in agricultural support initiatives and evokes their empathy. Example (2) guides the audience to pay attention to the plastic mulch, serving as a lead-in for the subsequent explanation of the advantages of highland eggplant planting. Example (3) is the blogger's invitation to the audience to visit the field, which helps to enhance the viewer's sense of on-site participation. Example (4) is an invitation to the audience before transitioning to the melon picking scene, which not only reduces the distance with the audience, but also hints the direct path of Minqin melon from the field to the market through the act of "picking", highlighting the product quality. Through instructions and invitations, imperative clauses pull the audience into the discourse context, reduce the social distance between the creator and the audience, and enhance interactional engagement.

Interrogative clauses:

(1) 楸水是什么呀?

What does "qiushui" mean?

(2) 这是老树对吧？

This is an old tree, right?

(3) 这个皮是不是后面要焮了水之后，他会更加的薄？

After the peel undergoes qiushui treatment, will it become even thinner?

(4) 来看我蒲城县这梨美不美？

Come and look, aren't these pears from Pucheng County beautiful?

Interrogative clauses are mainly used to demand information and advance discourse interaction. In examples (1), (2) and (3), the creator asked questions about planting technology, tree age and related technical effects in turn and gradually obtained more detailed information through continuous follow-up. This way of asking questions helps the audience to understand the production process and quality characteristics of the product more clearly, and breaks the flat rhythm of declarative clauses, making the expression more interactive and conversational. Example (4) poses a direct question to the audience to guide them to pay attention to the characteristics of pears, so as to enhance their sense of participation in the viewing process. Therefore, interrogative clauses not only serve the function of requesting information, but also encourage the audience to engage with the narrative process and enhance the interactivity of the video.

Exclamative clauses:

(1) 给大家看一下这个汁水太足啦！

Look at this—it's incredibly juicy!

(2) 沙漠种植出来的瓜就是香！

Melons grown in the desert are just so fragrant!

(3) 丰收咯！

What a harvest!

Exclamative clauses are mainly used to express strong emotions. Example (1) uses a vivid exclamation to highlight the abundant juiciness of Shenxian cantaloupe, using forceful language to enhance the product's appeal. Example (2) uses an exclamation to emphasize the origin advantages and unique flavor of Minqin melons, highlighting their scarcity and superior quality and deepening the audience's impression. Example (3) conveys the farmers' joy of a bountiful harvest through a concise and powerful exclamative clause, creating a celebratory atmosphere and stimulating audience emotions. The above exclamative examples, through strong emotions, create a positive atmosphere, highlight the advantages of agricultural products, and effectively stimulate audience emotions and purchasing intentions.

### 5.2 Analysis of the Modality System

Functional grammar regards modality as an excellent tool for realizing interpersonal function (Wang Hesi, Yin Pi'an, and Wang Furong, 2011). Within systemic functional grammar, Halliday (1985) divides modality into high, median, and low three categories. The higher the modal value, the greater the speaker's degree of certainty about a proposition. The lower the modal value, the more tentative and negotiable the expression becomes. High-value modality reflects the speaker's high degree of certainty

about a proposition and often conveys commands, requirements, obligations, or strong inclinations. Median-value modality reflects the speaker's median degree of certainty or inclination toward a proposition, balancing between affirmation and tentativeness. Low-value modality reflects the speaker's uncertainty and tentativeness and is typically used for speculation or polite expression. This study based on Peng Xuanwei's (2000) Chinese modal value table and the study's corpus, supplements and tabulates high-value, median-value, and low-value modal expressions, with statistical results as Figure 3 (Peng, 2000):

Modal Value	Expression Form	Frequency	Total	Percentage
High	(must)必须	1	6	7.2%
	(need to)要	3		
	(have to)得	2		
Median	(will)会	13	44	53.0%
	(would)要	13		
	(should)应该	3		
	(certainly)肯定	3		
	(would like to)想	11		
	(be willing to)愿意	1		
	(can)能	16		
Low	(may)可能	1	33	39.8%
	(could)可以	13		
	(might require)需要	3		

**Figure 3. Statistical Results of the Modality System**

According to the statistical results, high-value modal expressions account for 7.2% of the total, median-value modal expressions account for 53%, and low-value modal expressions account for 39.8% in the corpus of new-type farmers' agricultural promotion short videos. Median- and low-value modality are used more frequently.

By weakening the expression, the discourse can be closer to a more natural and approachable speaking style, making the information more interpersonally friendly. At the level of interpersonal relationship construction, this kind of expression takes the audience's feelings into consideration, establishing a certain sense of credibility and sincerity. Although high value modality is less frequent, it is still used in the text to emphasize key information, attract attention and guide the audience to take action. The following is an analysis of the selected example sentences:

High modality:

(1) 但是我们给消费者的必须精挑细选。

But what we give to consumers must be carefully selected.

(2) 这都得人工挑。

All of this has to be hand-sorted.

High-value modality can highlight and emphasize information. Example (1) uses the high-value modal verb "must" (必须) when introducing the sales guarantee for Bapan pears, reflecting their high regard for product quality and their sense of responsibility toward consumers, to some extent deepening audience trust. Example (2) uses the high-value modal expression 'has to' (得) to convey a strong sense of necessity, explaining the reliance of highland eggplant on manual sorting during transport down the mountain, strongly emphasizing the constraints that transport conditions impose on eggplant sales, and

capable of evoking audience empathy.

Median modality:

(1) 这里应该有个 1 万多斤吧。

There should be about 10,000-plus jin here.

(2) 如果您愿意参与这样一场助农公益活动，也可以跟我们书记联系。

If you are willing to participate in such an agricultural charity event, you may also contact our village secretary.

(1) 在这里要特别感谢当地部门帮我解决了很多问题。

I would especially like to thank the local authorities for helping me solve many problems.

(2) 这个已经熟了，再放个几天口感会更好。

This one is already ripe. If it is left for a few more days, the flavor will be even better.

The median-value modality can present information in a relatively objective tone. In Example (1), the blogger uses median-value modal expression “should” (应该) when analyzing the scale of stalled melons and fruits, providing information while avoiding being overly absolute and making the presentation of lagging sales more objective and enhancing the credibility of the description. Example (2) uses a relatively gentle tone to invite the audience to participate in the sales of Bapan pears to help farmers and reserve a certain choice for the audience so it is easier to be accepted. In example (3), the blogger uses median-value modality to express “want to” (要) when expressing gratitude, with a certain necessity, but the overall tone is not too positive or hesitant, showing a more natural and sincere expression. Example (4) uses the softer median value modal verb 'will' (会) to speculate on the possible changes in the flavor of pomelo, so that the tone is not absolute and reflects the handling of the complexity of the actual flavor change.

Low modality:

(1) 如果流量有价值，我希望能成为农户增收的桥梁。

If traffic has value, I hope it can become a bridge for increasing farmers' income.

(2) 让农户这些茄子尽量不要烂在地里，能卖出去一点是一点。

Let these farmers' eggplants not rot in the fields as much as possible. Every bit sold counts.

(3) 我们可以随便打开一个看一下吗？

Can we just randomly open one to take a look?

Low-value modal expressions use a tentative tone, leaving flexible space in the discourse. Examples (1) and (2) both use the low-value modal verb “can” (能), reducing the certainty of the utterance. In example (1), the blogger expresses aspirations for the agricultural promotion cause, using the low-modality form “hope” (希望) to humbly express the wish of helping farmers, shaping a credible new-type farmer image that reduces the distance with the audience. In example (2), the speaker uses a low-modality expression to predict the results of the agricultural promotion sales of highland eggplant, using a sincere tone to express the uncertainty of eggplant sales, evoking audience empathy. Example (3) uses the low-value modal expression “can” (可以) to politely inquire whether an agricultural

product parcel can be opened for display, and the consultative tone facilitates a natural transition in the video to the product display, enhancing the audience's sense of immersion.

## 6. Visual Mode Analysis

According to the visual grammar proposed by Kress and van Leeuwen (2006, pp. 114-148), interactive meaning is mainly realized through contact, social distance, angle, and modality. Accordingly, this chapter examines how interpersonal meaning is constructed in the visual mode of Douyin new-type farmers' agricultural promotion short videos from these four perspectives.

### 6.1 Analysis of the Contact System

Contact refers to the imaginary relationship established between image participants and image viewers through gaze. It is divided into demand and offer according to whether eye contact is established. When represented participants look directly at viewers and demand attention or involvement, a demand image is formed. When no eye contact is made and participants are presented as objects of observation, an offer image is formed (Kress & van Leeuwen, 2006, pp. 116-122). The statistical results of the contact system are shown in Table 2:

**Table 2. Statistical Results of the Contact System**

	Demand	Offer
Frequency	125	113
Percentage	52.5%	47.5%

According to Table 2, a total of 125 demand images were identified, accounting for 52.5%, and 113 offer images, accounting for 47.5%. The proportion of demand images is slightly higher, which shows that new-type farmer creators attach more importance to the establishment and maintenance of interaction with the audience. Through the use of direct gaze, the audience is introduced into the communicative context, thus reducing the social distance, strengthening emotional identification with farmers and enhancing audience engagement. At the same time, offer images also account for a certain proportion. Creators do not make eye contact with the audience, but present agricultural product information or the real difficulties faced by farmers in a relatively objective way.



Figure 4 Figure 5

Figure 4 shows a demand image in which a new-type farmer creator smiles and looks directly into the camera while holding half of a melon against a desert background. The direct gaze conveys confidence in the product and commitment to supporting agriculture, while inviting viewers to engage with the promotional theme of Minqin melons. Figure 5 is an 'offer' image in which the blogger interviews a plum farmer beside the orchard. Neither participants make eye contact with the audience and objectively present information about drought, water shortages, and reduced yields, reflecting the difficulties faced by the farmers.

### 6.2 Analysis of the Distance System

Distance mainly constructs the social and emotional relationship between the image participant and the viewer through framing size (Kress & van Leeuwen, 2006, pp. 124-129). The closer the camera is, the more intimate the relationship. The farther the distance, the easier it is to form a relatively objective and alienated viewing relationship. Close-up shots usually focus on the character's face or a certain detail to strengthen emotional expression. The medium close-up shot generally presents the part above the chest of the character, reflecting the closer interactive relationship. The medium shot shows the area above the knees of the characters, which corresponds to a daily social distance. The long shots present the characters and the surrounding environment, showing a relatively detached public relationship. The results are depicted in Table 3:

Table 3. Statistical Results of the Distance System

	Close-up	Medium Close-up	Medium Shot	Long Shot
Frequency	53	91	151	28
Percentage	16.4%	28.2%	46.7%	8.7%

It can be seen from Table 3 that among the 15 agricultural promotion videos, close-up shots account for 16.4%, which are mainly used to magnify the details of agricultural products. The medium close-up shot accounts for 28.2%, mostly used to show the facial expressions of bloggers or present the external

details of agricultural products. This kind of lens can bring the audience closer to the blogger and further enhance the audience's sense of trust. The medium shot accounts for the highest proportion, reaching 46.7%, used to convey product information and sales content to the audience at a relatively natural social distance. The long shot accounts for 8.7%, which is usually used to show the growth environment of agricultural products and provide support for the overall narrative.



Figure 6



Figure 7 Figure 8 Figure 9

Figure 6 adopts a close-up shot, and highlights the golden yellow and fibrous flesh details of Renjia Village sweet potato with a blurred background, making its texture more intuitive and easier to stimulate the audience's appetite. By drawing viewers into close proximity with the product, the shot reduces social distance and elicits consumer desire. Figure 7 uses a medium close-up shot. The camera focuses on the blogger's gentle expressions and movements, making it easier for the audience to enter the introduction of Pinghe pomelos, and at the same time weakens the sense of distance between the two sides. Figure 8 adopts the medium shot to bring the audience into the on-site interview scene about the sales dilemma of highland eggplant, and the audience's attention is also guided to the agricultural support theme. Figure 9 presents the overall growth environment of Dangtu through the long shot. The dense river network and large area of water in the picture illustrate the good natural hydrological conditions of Dangtu water chestnuts, and also provide a visual basis for the expression in the video.

### 6.3 Analysis of the Angle System

The angle reflects the attitude of the image viewer towards the image participant, which is divided into horizontal angle and vertical angle. The horizontal angle reflects the involvement or detachment relationship between the viewer and the participant. The frontal angle fosters empathy and forms a common viewing feeling, while the oblique angle allows the audience to engage with the image as a

spectator, thus forming a relatively objective narrative atmosphere. The vertical angle reflects the status relationship between the viewer and the participant. A bird's-eye view angle means that the viewer is in a higher position, the level angle reflects the equal interpersonal relationship between the two sides, and a worm's-eye view angle means that the image participant is in a higher position (Kress & van Leeuwen, 2006, pp. 139-146). The statistical results are shown in Table 4.

**Table 4. Statistical Results of the Angle System**

	Horizontal Angle		Vertical Angle		
	Frontal	Oblique	Level	Worm's-eye	Bird's-eye
Frequency	131	157	288	13	29
Percentage	45.5%	54.5%	87.3%	3.9%	8.8%

Statistics show that there are a total of 288 instances of horizontal angles, of which 131 are frontal angles, accounting for 45.5%, and 157 oblique angles, accounting for 54.5%. The proportion of oblique angles is slightly higher, so that the audience can engage in the video more as observers, facilitating the objective delivery of promotional information. Although the proportion of frontal angles is relatively low, it can still reduce the social distance and enhance the audience's sense of trust.



**Figure 10 Figure 11**

Figure 10 adopts a frontal angle, placing the camera on the same plane as the viewer. In the picture, the blogger holds a large kale in the live stream setting and shows it directly to the audience. This perspective strengthens the real and simple atmosphere of the video and enhances the audience's sense of participation, making the viewing process more immersive. Figure 11 uses an oblique angle to present the scene of lychee workers sorting and processing fruits. The audience watches the whole process from the perspective of bystanders. This composition not only shows the busy atmosphere in the production process of Gaozhou lychee, but strengthens the sense of harvest, while enhancing the authenticity of the narrative.

In the vertical angle statistics, the level angle accounts for the highest proportion, and the interpersonal relationship with the audience is relatively equal in the agricultural promotion short video narrative. The proportion of bird's-eye view and worm's-eye view is relatively lower. The bird's-eye view is mainly used to show the overall view of the growing environment of agricultural products, while the worm's-eye-view is more used to highlight the details of agricultural products such as fruits on trees, emphasizing their fresh and plump characteristics.



Figure 12 Figure 13 Figure 14

Figure 12 adopts a level angle, allowing the audience to participate in the conversation of the medicinal value of Bapan pears in a relatively equal capacity, so as to enhance the audience's sense of experience. Figure 13 uses a bird's-eye view to survey the natural environment and village layout of the plum planting land as a whole, so that the audience can observe the planting environment from a higher perspective. This broad perspective further strengthens the connection between the excellent growing environment and quality of plums, and allows the audience to form a more intuitive overall understanding of the production environment. Figure 14 uses a worm's-eye view to frame the picture of fruit farmers picking lychee, highlighting the rich yield of lychee, and conveying the joyful atmosphere in the harvest scene.

#### 6.4 Analysis of the Modality System (Visual)

In visual grammar, the modality system mainly involves the authenticity, credibility and sense of reality presented by images. Visual modality is usually judged by multiple indicators, including color saturation, color differentiation, color modulation, contextualization, representation, depth, illumination and brightness. According to the degrees of these indicators, images can be divided into high modality, median modality and low modality (Kress & van Leeuwen, 2006). Agricultural promotion short videos mainly adopt high modality visual strategies, such as bright colors, clear details, rich contextual information and bright lighting. These visual features can enhance the audience's perception of the authenticity and quality of agricultural products.



Figure 15 Figure 16 Figure 17

Figure 15, Figure 16 and Figure 17 all reflect the high visual modality. Figure 15 adopts high saturation warm colors and strong color differentiation to highlight the bright and full appearance of Minqin melon, thus enhancing the audience's positive impression of the product. In Figure 16, there is a sharp contrast between the green water chestnut field and the workers' red clothing, which not only enhances the visual prominence of the picture, but also conveys the pleasant atmosphere in the harvest scene. Figure 17 uses bright lighting and high saturation green to highlight the fresh and tender characteristics of kale, enhancing the realism of the picking scene and increasing audience's interest in fresh and high quality agricultural products.

### 7. Analysis of Text-Image Intersemiotic Complementarity

The above has analyzed the interpersonal meaning construction of verbal mode and visual mode in agricultural promotion short videos based on systemic functional grammar and visual grammar respectively, but the interaction between the two semiotic resources has not been discussed. The theory of intersemiotic complementarity proposed by Royce (1998) mainly focuses on the complementary relationship between language and images in multimodal discourse. Therefore, this study adopts the theory of intersemiotic complementarity to analyze the interaction between verbal mode and visual mode. The theory of intersemiotic complementarity includes ideational intersemiotic complementarity, interpersonal intersemiotic complementarity and textual intersemiotic complementarity. Among them, interpersonal intersemiotic complementarity can be further divided into attitudinal convergence, attitudinal divergence and appellation reinforcement. The following will analyze how verbal mode and visual mode jointly participate in the construction of interpersonal meaning in combination with specific cases.

#### 7.1 Attitudinal Convergence

Attitudinal convergence means that in multimodal texts, linguistic and visual semiotic resources cooperate with each other. The two modes jointly express a consistent interpersonal orientation and construct a unified attitudinal stance (Royce, 2013, pp. 93-96). This method appears frequently in agricultural promotion short videos, which can make the content more compelling and help enhance the audience's sense of participation.



Figure 18 Figure 19

In Figure 18, the blogger wears a straw hat and a farming attire, showing the melon to the audience and raising questions about the reason why the melon is sweet. Verbal mode guides the audience to participate through interrogative clauses, stimulates the audience's curiosity, and prepare for the follow up content. At the same time, the visual mode also constructs a similar interpersonal meaning. In the contact system, the creator looks directly at the audience and forms a demand image to attract the audience to participate in the interaction. At the distance level, the medium shot reduces the social distance between the creator and the audience, making the overall atmosphere approachable. In the angle system, the video adopts a frontal and level perspective, placing the creator and the melon in the center of the picture at the same time, so that the audience can engage with the interaction process in a more equal position. Together, the verbal and visual modes reinforce one another. Interrogative clauses, direct gaze, medium shot and level angle jointly build a closer interpersonal relationship between the creator and the audience, thus stimulating the audience's interest in Minqin melons.

In Figure 19, the video introduces the high altitude planting environment of Bapan pears through declarative clauses so that the audience can understand the advantages of the product in a more objective way. Correspondingly, the visual mode adopts a long shot and a high-angle perspective, which further enhances the sense of objectivity of the image and creates greater distance. At the same time, viewers are positioned above the orchard, enabling them to perceive the geographical characteristic of high-altitude cultivation more directly. In this case, verbal mode and visual mode jointly build a consistent interpersonal meaning through shared attitudinal orientation, and both present product information in a more credible way.

### 7.2 Attitudinal Divergence

Although the frequency of attitudinal divergence in agricultural promotion short videos is relatively low, it still plays an important role in the construction of interpersonal meaning. Attitudinal divergence refers to the inconsistency or even opposition between linguistic and visual semiotic resources in the expression of interpersonal meaning, which is usually used to highlight the key points, form contrasts or convey implicit evaluation meaning (Royce, 2013, pp. 96-97).



**Figure 20**

In Figure 20, the blogger uses the third person to objectively tell the prediction that “Minqin will be swallowed by the desert”, distancing himself from the audience. However, the visual modality constructs the opposite interpersonal meaning. The video adopts demand images, with Minqin melon farmers form direct eye contact with the audience to attract the attention of the audience. At the same time, the video uses a level angle to invite the audience to participate in the narrative in a more equal way. In addition, it adopts relatively high-modality colors, creating a positive emotional atmosphere through bright colors such as the red of the field tractor and the yellow of the melon. Verbal mode and visual mode have formed different interpersonal meaning expressions through attitudinal divergence. This contrast further highlights the resilience of the Minqin people who endured hardships to grow melons and combated the desert and finally turned the desert into farmland, defying the prophecy of their disappearance.

### *7.3 Appellation Reinforcement*

In addition to attitudinal convergence and attitudinal divergence, appellation reinforcement is also an important mechanism in interpersonal intersemiotic complementarity. It refers to the consistent realization of two or more semiotic systems in the sense of appellative meaning. By addressing the audience directly, creators can strengthen the connection with the audience and improve the audience’s sense of participation (Royce, 2013, pp. 92-93).



**Figure 21 Figure 22**

In Figures 21 and 22, the creator invited the audience to the livestream to buy products at the end of the

video and assured the quality and price of melon. The repeated use of the address “Everyone (大家)” directly talks to the audience, which reduces the social distance between the two sides. The visual mode further reinforces this appellative meaning. At the contact level, the creator looks directly at the camera and establishes a demand relation with the audience. At the distance level, the creator kept approaching the camera with a melon in her hand, and the framing gradually turned from the medium shot to the medium close-up, progressively reducing distance. At the angle level, the video adopts a frontal and level perspective to further enhance the interactive relationship. The consistent realization of verbal mode and visual mode in the sense of appellative meaning makes two semiotic resources jointly participate in the construction of interpersonal meaning, and finally achieve appellation reinforcement.

## 8. Conclusion

Based on the theory of interpersonal meaning of systemic functional linguistics, the theory of interactive meaning of visual grammar and the theory of interpersonal intersemiotic complementarity, this study analyzes the verbal mode and visual mode in 15 new-type farmers’ agricultural promotion short videos on the Douyin, focusing on how the two modes jointly construct interpersonal meaning and the interaction and complementarity between them. The study found that in the verbal mode, the frequency of declarative clauses and medium- and low- value modal expressions is relatively high, which makes the information transmission relatively objective. In terms of visual mode, demand images, medium close-up and medium shots, level angles and high modality expressions appear more frequently. This kind of visual expression helps reduce social distance with viewers and present the actual situation of farmers and information about agricultural products, thus enhancing the audience’s sense of identity with the video content. In the interaction between verbal mode and visual mode, attitudinal convergence is the most common complementary way. The consistent cooperation of the two modes can further highlight the narrative focus of the video and deepen the audience’s impression of the content. Appellation reinforcement mainly incorporates the audience into the specific context through the combination of direct address and relevant visual strategies, so as to strengthen the interaction between the creator and the audience. In contrast, the frequency of attitudinal divergence is relatively low, and it is mainly used to form a contrast effect, highlight key information, enrich the video narrative layers, and improve viewing engagement.

This study contributes to research on agricultural promotion short videos from a linguistic perspective and enriches empirical studies of interpersonal meaning construction in multimodal discourse analysis. At the practical level, the research results also provide practical implications for the interactive design of new-type farmers’ agricultural promotion short videos. However, there are still some limitations in this study. First of all, the research only selected 15 short videos on Douyin as its corpus. The sample size is relatively limited, so the representativeness of the research results is still limited to a certain extent. Secondly, the analysis of the complementary relationship between text and image is mainly based on typical cases and lacks more systematic quantitative statistics. Follow up research may further

expand the sample size and combine quantitative methods to carry out a more in depth study on the interpersonal intersemiotic complementarity in agricultural promotion short videos.

## References

- Abbamonte, L. (2021). The “sustainable” video-narratives of Greenpeace—an ecolinguistic investigation. *Forum for Modern Language Studies*, (57), 145-185.
- Ali Research Institute. (2015). *2014 Research Report on China's New-Type Farmers*. [http://www.nongshijie.com/a/article\\_9610.html](http://www.nongshijie.com/a/article_9610.html)
- Blood donors wanted: Narrative innovation on TikTok to enable mobilization. (2023). *Profesional De La información*, 32(3).
- Castro, C. D., & Halliday, M. A. K. (1995). An introduction to functional grammar. *Language*, 71(4), 831.
- Du, shuangyan, & Cheong, C. Y. M. (2025). Unveiling ethnic tourism imaginaries: A multimodal evaluation of Xizang Autonomous Region's Douyin promotions. *Ibérica*, (49), 77-106.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*.
- Kress, G., van Leeuwen, T. Kress, G., & van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design* (p. 288). Londres: Routledge.
- Li, T. (2024). Multimodal Discourse Analysis of Emotion-driven Short-video News. *Journalism Lover*, (12), 28-30. <https://doi.org/10.16017/j.cnki.xwahz.2024.12.009>
- Li, Z. (2003). Social Semiotic Approach to Multimodal Discourse. *Foreign Languages Research*, (05), 1-8+80. <https://doi.org/10.13978/j.cnki.wyyj.2003.05.001>
- Li, Z., & Lu, D. (2012). Multimodal Semiotics: Theoretical Basis, Research Paths and Future Prospects. *Foreign Languages Research*, (02), 1-8. <https://doi.org/10.13978/j.cnki.wyyj.2012.02.006>
- Mordecai, C. (2023). #anxiety: A multimodal discourse analysis of narrations of anxiety on TikTok. *Computers and Composition*, (67), 102763.
- Norris, S. (2004). *Analyzing Multimodal Interaction: A Methodological Framework*. Routledge.
- Ou, L., Zhang, T., Liu, Y., & Chen, P. (2023). Multimodal discourse analysis and visual persuasion strategy of short videos of scientific journals: A case study on journals listed in Excellence Action Plan for China's STM Journals. *Chinese Journal of Scientific and Technical Periodicals*, 34(11), 1494-1501.
- Pan, Y., & Li, Z. (2017). Multimodal Discourse Analysis in China's Academia: An Overview of the Publications with CSSCI Journals in Latest Decade. *Journal of Fujian Normal University(Philosophy and Social Sciences Edition)*, (05), 49-59+168-169.
- Peng, X. (2000). *A Comprehensive Contrastive Study of English and Chinese Discourse*. Shanghai Foreign Language Education Press.
- Royce, T. (1999). *Visual-verbal intersemiotic complementarity in the economist magazine*. University

- of Reading.
- Royce, T. (2013). *Intersemiotic complementarity: A framework for multimodal discourse analysis*. Lawrence Erlbaum & Assoc.
- Sang, X. (2020). A study of the construction of interactive meaning in multimodal discourses: A case study of Li Ziqi's videos. *Foreign Studies*, 8(04), 26-32+104.
- Wang, H., Yin, P., & Wang, F. (2011). A Comparative Study of Modality Between English and Chinese Political Speeches. *Journal of Xi'an International Studies University*, 19(02), 38-42. <https://doi.org/10.16362/j.cnki.cn61-1457/h.2011.02.016>
- Xin, Z. (2008). New Developments in Discourse Analysis: Multimodal Discourse Analysis. *Social Science Journal*, 2008(5), 4. <http://dspace.xmu.edu.cn:8080/dspace/handle/2288/21137>
- Xinhua News Agency. (2025). ByteDance Releases the 2025 Rural Assistance Data Report. *Xinhua Net*. <https://www.xinhuanet.com/tech/20250930/bd87a324922f46fe81050c8211d9815b/c.html>
- Zhang, D. (2009). On A Synthetic Theoretical Framework for Multimodal Discourse Analysis. *Foreign Languages in China*, 6(01), 24-30. <https://doi.org/10.13564/j.cnki.issn.1672-9382.2009.01.004>
- Zhao, A. (2025). Overseas Chinese parent wanghong on Douyin: Discursive constructions of diasporic parenthood. *Deakin University. Journal contribution*.
- Zhou, S., & Huang, W. (2025). Optimization Strategies for Multimodal Narratives in Science Popularization Short Videos: A Case Study of Bilibili Account "Huazha Hua Xiaolao". *Studies on Science Popularization*, 20(02), 81-90+104+109-110. <https://doi.org/10.19293/j.cnki.1673-8357.2025.02.009>
- Zhu, X., & Shi, Y. (2022). Short-video Production Practices of Traditional Media Hosts: An Analysis from the Perspective of Multimodal Discourse. *Media*, (20), 94-96.
- Zhu, Y. (2007). Theory and Methodology of Multimodal Discourse Analysis. *Foreign Language Research*, (05), 82-86. <https://doi.org/10.16263/j.cnki.23-1071/h.2007.05.034>