

VIETNAM NATIONAL UNIVERSITY, HANOI
UNIVERSITY OF LANGUAGES AND INTERNATIONAL STUDIES

NGUYỄN HỒNG LIÊN

**MULTIMODAL REPRESENTATION
OF VIETNAMESE STREET FOOD
IN YOUTUBE FOOD REVIEW VIDEOS**

Major: English Linguistics

Code: 9220201.01

SUMMARY OF DOCTORAL THESIS

HANOI, 2026

**The thesis has been conducted at
University of Languages and International Studies, VNU**

Supervisors:

1. Prof. Nguyễn Hòa
2. Dr. Nguyễn Thị Hương Lan

Reviewer 1:

Reviewer 2:

Reviewer 3:

The thesis will be defended before the Doctoral Thesis Evaluation Committee
at University of Languages and International Studies, VNU
at, on,

The thesis can be accessed at:

- The National Library of Vietnam
- The Information and Library Center, Vietnam National University, Hanoi.

THESIS-RELATED PUBLICATIONS

1. Nguyen Hong Lien (2023). A case study of paralinguistic in videos about Vietnam's Tet holiday: An approach from Systemic Functional Linguistics theory, *Proceedings of 2023 International Graduate Research Symposium (IGRS 2023)*, pp.547-557.VNUP
2. Nguyen Hong Lien (2023). Book review: Modelling paralinguistic using systemic functional semiotics: theory and application, *VNU Journal of Foreign Studies*, 39(4), 148-152. <https://doi.org/10.63023/2525-2445/jfs.ulis.5148>
3. Nguyen Hong Lien (2024). Multimodal metaphors in videos about Vietnamese street food: A case study, *Proceedings of 2024 International Graduate Research Symposium (IGRS 2024)*, pp.326-338.VNUP
4. Nguyen Hong Lien & Nguyen Thi Hoai (2024). Attitude resources in videos about Vietnamese cuisine: An approach from appraisal theory, *Proceedings of International Conference on Trends of Modern Linguistics: Theoretical and Applied*, pp.673-680.VNUP
5. Nguyen Hong Lien (2025). From street to screen: Exploring ideational meaning in Vietnamese street food videos via Multimodal Discourse Analysis, *Proceedings of International Conference on "Functional Linguistics and Multimodality"*, pp.376-393. VNUP
6. Nguyen Hong Lien (2025). Street food as cuisine: A multimodal discourse analysis of Vietnamese street food, *Proceedings of International Conference on Language, Media and Culture*, pp.306-325.VNUP
7. Nguyen Hong Lien (2025). The exotic and the explorer: A multimodal analysis of identity construction in a Vietnamese street food review video, *HNUE Journal of Science*, 70(4), 3-15. <https://doi.org/10.18173/2354-1067.2025-0052>
8. Nguyen Hong Lien & Duong Thu Ha (2025). Multimodal mechanisms to engage viewers in appreciating Vietnamese street food in a food review video, *HUFLIS Journal of Inquiry into Languages and Cultures*, 9(3), 370-391. <https://doi.org/10.63506/jilc.0903.357>
9. Nguyen Hong Lien & Duong Thu Ha (2025). Applying the Multimodal Discourse Analysis approach to explore the representation of Vietnamese street food in extracts of food review videos on a social network platform (Vận dụng đường hướng phân tích diễn ngôn đa phương thức để tìm hiểu sự biểu hiện của ẩm thực đường phố Việt Nam trong các trích đoạn video đánh giá ẩm thực trên mạng xã hội); [University-level research project, Hanoi National University of Education, Hanoi].

CHAPTER 1: INTRODUCTION

1.1. Rationale

Vietnamese street food is deemed to be favored around the world and has gained international recognition. Beyond its gastronomic appeal, Vietnamese street food functions as a cultural text through which national identity, everyday life and social values are articulated. In the contemporary digital era, however, cultural meanings are no longer constructed solely through physical settings. YouTube emerge as powerful specialized tools within this broader digital landscape. Its affordances position YouTube as a key site for constructing and circulating meanings of Vietnamese street food. Recognizing this potential, the Vietnamese government has emphasized on developing tourism, especially cuisine tourism by harnessing the power of social networks as well as social media. Consequently, social media influencers, travel bloggers or content creators are increasingly considered as important contributors to the growth of Vietnam's tourism industry with their strong impact on their followers' travel choices. Furthermore, thanks to the multimodal nature of videos, there are advantages that make multimodality in general or multimodal discourse analysis (MDA) in particular a powerful tool for investigating YouTube videos about street food. MDA allows researchers to analyze various modes of communication (e.g., visual, verbal, paralinguistic, etc) simultaneously to understand how meaning is constructed and communicated. However, research on Vietnamese street food within YouTube review videos is scarce, both domestically and internationally. Addressing this gap is essential, as such an investigation can provide deeper insights into how Vietnamese street food is represented and perceived globally, while also informing tourism marketing strategies that align with national cultural promotion policies. To address these gaps, I conduct a study titled **“MULTIMODAL REPRESENTATION OF VIETNAMESE STREET FOOD IN YOUTUBE FOOD REVIEW VIDEOS”**.

1.2. Aim and objectives

The overarching aim of this research is to examine how Vietnamese street food is multimodally represented and constructed on social media, with a specific focus on English-language YouTube review videos.

To achieve this aim, the study pursues the following objectives:

- To analyze how verbal, visual and paralinguistic modes interact to construct the representation of Vietnamese street food in YouTube review videos;
- To examine how verbal, visual and paralinguistic modes orchestrate to engage viewers and foster appreciation of Vietnamese street food;
- To investigate how the multimodal representation of Vietnamese street food realizes different food identities in the food review genre.

1.3. Research questions

The study addresses the overarching question: *How is Vietnamese street food multimodally represented and constructed in English-language YouTube review videos?*

This overarching question is operationalized into three specific research questions:

RQ 1: *How do verbal, visual and paralinguistic modes orchestrate to construct the representation of Vietnamese street food?*

RQ 2: *How do verbal, visual and paralinguistic modes orchestrate to engage viewers and foster appreciation of Vietnamese street food?*

RQ 3: *How are the identities of Vietnamese street food realized through its multimodal representation in YouTube food review videos?*

1.4. Methodology and scope of the study

This study is grounded in Multimodal Discourse Analysis (MDA), informed by both Social Semiotics and Systemic Functional Linguistics (SFL) as key theoretical foundations. The first objective is obtained through the analysis that focuses on the ideational metafunction which

unveils how the multifaceted characteristics of Vietnamese street food are constructed and represented. The second objective is obtained through the analysis that focuses on the interpersonal metafunction assesses how these multimodal resources are employed to engage viewers and cultivate appreciation for the culinary subject. This involves a detailed examination of how verbal descriptions, visual imagery and paralinguistic elements interact to construct meaning. To achieve the third objective, a systematic reflexive thematic analysis is conducted using an adapted form of van Leeuwen's (2022) identity framework. This qualitative method allows for the deductive and inductive identification and categorization of recurring descriptive patterns, thereby discovering the identities of Vietnamese street food through the interplay of verbal, visual and paralinguistic modes.

1.5. Significance of the thesis

Theoretically, this study makes a substantial contribution to the evolving landscape of MDA, particularly through its application of Social Semiotics to the analysis of online video content. Methodologically, the study proposes a systematic and replicable approach to qualitative research on online video discourse. Practically, the findings of this study offer considerable practical implications for a diverse range of stakeholders, ranging from content creators and digital culinary influencers to tourism boards and culinary enterprises.

1.6. Structure of the thesis

The thesis is comprised of eight chapters, including Introduction, Literature review, Research methodology, Multimodal orchestration for constructing the representations of Vietnamese street food, Multimodal orchestration for engaging viewers in appreciating Vietnamese street food, Identities of Vietnamese street food through its multimodal representation, Discussion and Conclusion.

CHAPTER 2: LITERATURE REVIEW

2.1. Online food reviews

Truong et al. (2025) suggest several key characteristics of food review videos. Food vlogs (video blogs) are a significant social media trend consisting of user-generated video content that narrates culinary experiences and aims to both entertain viewers and provide information about food and beverages. Food vlogs often provide authentic and personal perspectives, including detailed narratives, critiques and sometimes insider views to make them a preferred medium for an immersive food experience online. They utilize visual and auditory elements to engage viewers' senses and showcase food attractively. Therefore, the multimodal nature of videos has the ability to capture viewers' attention to the specific type of food review.

2.2. Representation and identity of street food in multimodal digital contexts

2.2.1. Representation and multimodality in digital media

In this study, I approach representation as a meaning-making practice in which semiotic resources such as language, image or gesture are deliberately organized to construct particular versions of reality. This perspective naturally leads to multimodality, since representation in digital media rarely relies on one mode alone but emerges through the interplay of multiple semiotic resources. In digital media contexts such as YouTube street food channels, representation thus becomes a site where dishes are multimodally staged, narrated, engaged with and circulated as cultural artifacts, which allows the food itself to perform identity and cultural meaning through its multimodal display.

2.2.2. The concept of street food identity

I posit that street food's identity is a socially constructed concept, emerging through an ongoing negotiation between the content creator's

performative representation of street and the viewer's engagement with that representation.

2.2.3. Food identity in digital multimodal food discourse

Food is deeply embedded in identity formation across personal, communal and national levels. Its identity emerges not simply from ingredients or taste but from preparation practices, cultural narratives, institutional promotion and mediated representation. Understanding food identity thus requires attention to the semiotic, discursive and multimodal processes through which food is represented, evaluated and made meaningful in contemporary media contexts.

2.2.4. Analytical framework for understanding street food identity

This study integrates an adapted version of Hall's (2013) Circuit of Culture with van Leeuwen's (2022) social semiotic identity framework to establish a unified analytical model for examining how Vietnamese street food identities are constructed in YouTube food review videos. Hall et al.'s model provides the macro-level cultural logic of meaning-making, which conceptualizes representation as a dynamic process embedded in creation, circulation and interpretation. However, its original formulation is insufficiently sensitive to the multimodal and platform-specific nature of digital food discourse. By adapting the circuit into five analytically relevant moments namely *creation*, *multimodal representation*, *engagement*, *norm* and *food/identity*, the study aligns Hall et al.'s cultural process with the affordances and practices of YouTube. Van Leeuwen's framework offers the micro-level semiotic pattern through which identity is realized. It reconceptualizes identity as a meaning enacted through semiotic choices. By adapting van Leeuwen's identity types from human subjects to food including *communal*, *particularized*, *functional* and *lifestyle identity*, the study connects identity formation to multimodal

representation and viewer engagement within the adapted circuit. Briefly, the two models operate in complementarity: Hall's framework explains where and why meanings circulate in digital culture, while van Leeuwen's explains how street food performs identity through the orchestration of multimodal resources.

2.3. Social Semiotics as the selected approach of this study

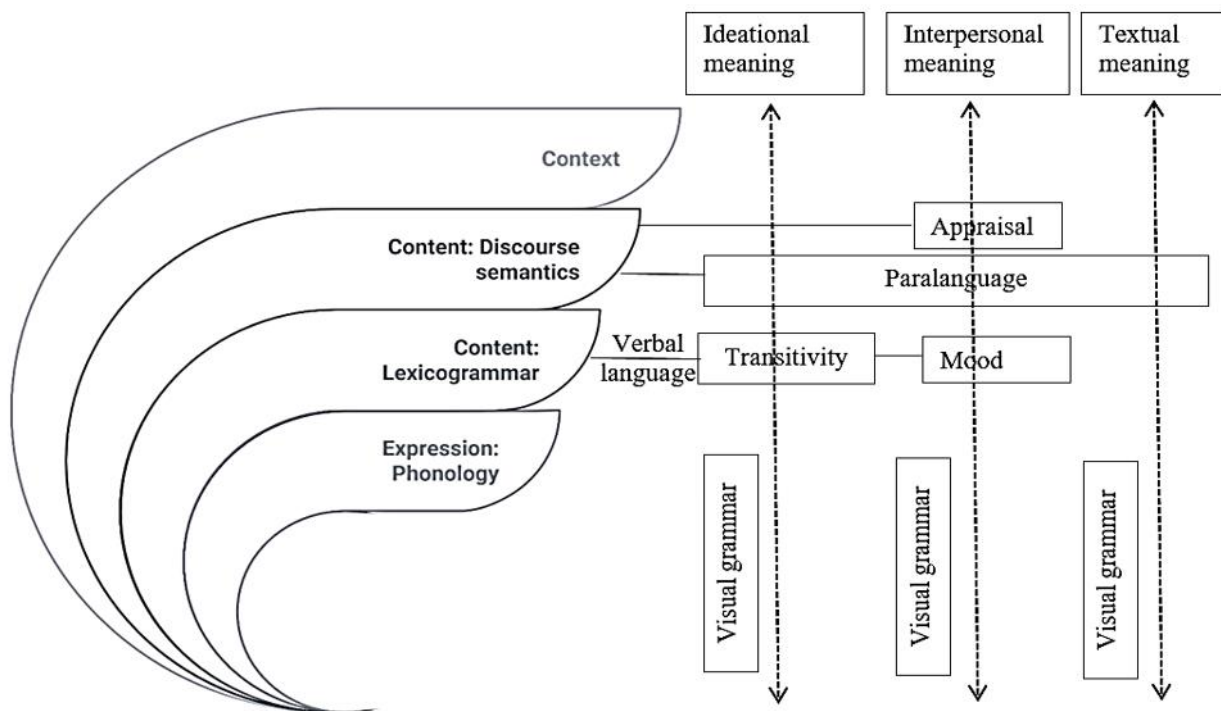
This study adopts Social Semiotics as its primary analytical approach because it offers a theoretically robust framework for analyzing meaning-making in multimodal texts. Moreover, Social Semiotics conceptualizes all communication as a social act, highlighting the relationship between representation, identity and value.

2.4. Meaning-making resources to be investigated in the study

The study investigates three modes in YouTube food review videos: the verbal mode, the visual mode and the paralinguistic mode.

Figure 2.3

Relationships of verbal, visual and paralinguistic resources in the study



2.5. Summary of selected aspects of SFL-based models to be investigated in this study

Table 2.8

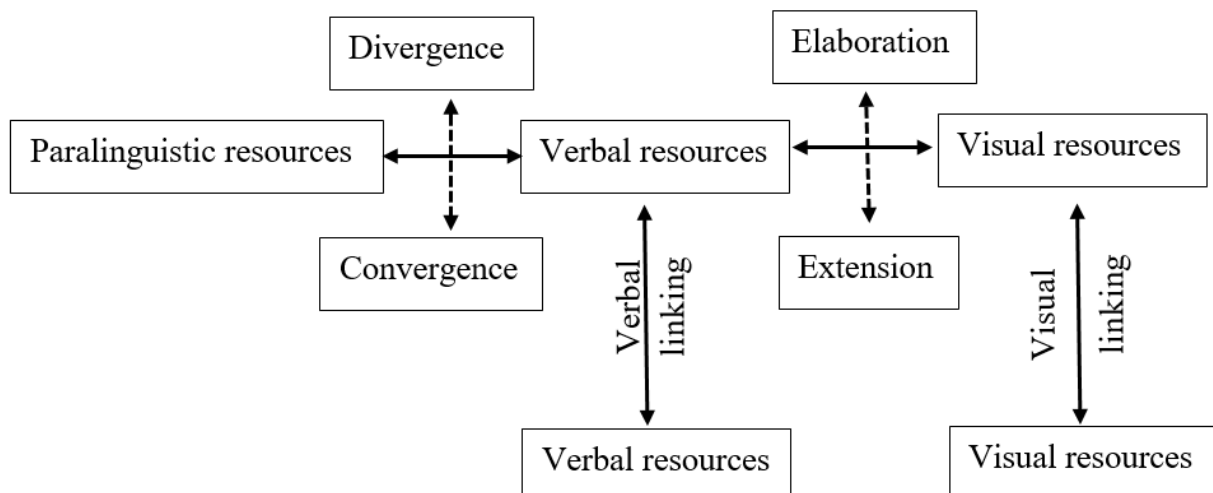
Summary of key theories in meaning-making resources to be inspected in the study

Metafunctions	Verbal resources	Visual resources	Paralinguistic resources
Ideational/ Representational function	The Transitivity system (Halliday & Matthiessen, 2004)	Conceptual/ Narrative processes (Kress & van Leeuwen, 2021)	Hand shapes Hand or arm motions (Ngo et al., 2022)
Interpersonal/ Interactive function	Mood system: (Halliday & Matthiessen, 2004) Appraisal framework – Attitude (Martin & White, 2005)	Participants Distance Attitude Gaze (Kress & van Leeuwen, 2021)	Facial affect Voice affect (Ngo et al., 2022)

2.5.6. An adapted multimodal orchestration model

Figure 2.4

An adapted multimodal orchestration model



CHAPTER 3: METHODOLOGY

3.1. Research design

This study adopts a qualitative interpretivist-constructivist paradigm as its philosophical foundation, recognizing that meaning in street food representation arises through social interpretation and semiotic construction (Creswell & Creswell, 2018; Miyahara, 2020).

3.2. Data selection and collection

3.2.1. *Data selection criteria*

The following criteria guided the initial search and subsequent filtering of YouTube videos: Relevance, Popularity, Language, Diversity and Public domain availability.

3.2.2. *Description of the final dataset*

The final dataset for this study comprises three YouTube videos. All selected videos are the most viewed Vietnamese street food content from their respective channels to ensure their significant reach and impact within the online culinary sphere. The three channels are (all names are pseudonyms):

- *Video 1* was published on January 12, 2018; this video from the Sobe channel takes viewers on an adventure through Hanoi, Vietnam, seeking out incredibly rare and exotic street food dishes. With nearly 18 million views and a runtime of 13 minutes and 7 seconds, it offers a widespread culinary journey.

- *Video 2* is from the Fora channel; this video was published on August 3, 2017. It promises an insane street food tour in Saigon, Vietnam, uncovering hidden Vietnamese street food spots, including options for street breakfast and eggs. It boasts approximately 7 million views and lasts 22 minutes and 44 seconds.

- *Video 3* is from the Mawi channel; this video was released on May 9, 2023 and zeroes in on five “must-eat” and extreme” street food experiences in Hanoi, Vietnam. It has nearly 5 million views and is substantially 40 minutes and 19 seconds long.

3.2.3. Data collection process

3.2.3.1. The unit of analysis

In this study, a scene is adopted as the unit of analysis, which is primarily identified by its space-time continuity. Each video is segmented into scenes, excluding the introduction and the ending. In total, there are 91 scenes.

3.2.3.2. Multimodal transcription and annotation

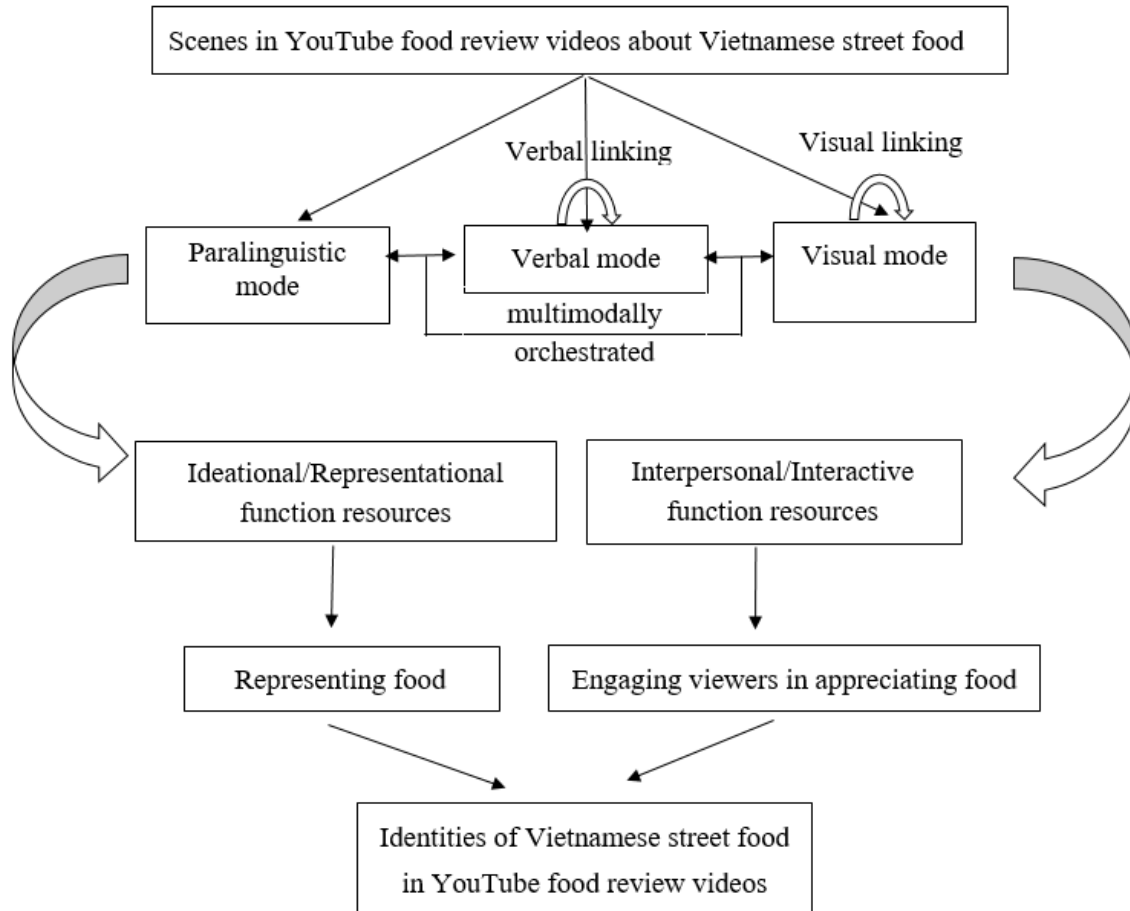
This study takes on Norris's (2019) viewpoint that a selected number of key dimensions should be analyzed and discussed in-depth. Only the aspects that are most pertinent to the phenomenon being studied would be the focus of the intricate multimodal annotation (Zappavigna, 2019).

3.3. Data analysis

3.3.1. The analytical framework of the study

Figure 3.2

The analytical framework of the study



3.3.2. Data analysis procedure

Firstly, the data corpus for this study consisted of 91 scenes in three YouTube videos about Vietnamese street food. Secondly, all verbal, visual and paralinguistic content was transcribed and stored in Excel documents. Next, the analysis proceeded to address the research questions. At the initial stage of analysis, scenes were organized sequentially within each video to preserve narrative coherence and genre-specific progression. This chronological organization made it possible to observe how meaning unfolds over time, through the common structuring of introductory framing scenes, core tasting and evaluation sequences and concluding assessment. At the subsequent analytical stage, scenes were reorganized according to their dominant semiotic functions. Each scene was examined in terms of its primary contribution to representation (RQ1), viewer engagement (RQ2) and identity construction (RQ3). Representational scenes foregrounded the material, sensory or procedural properties of street food such as ingredients, preparation techniques, and texture/flavor. Engagement-oriented scenes prioritized interpersonal alignment including direct address, embodied evaluation, affective display and strategies designed to establish credibility and rapport. Identity-oriented scenes were those in which street food was explicitly or implicitly positioned within broader cultural, social and lifestyle discourses such as community belonging, tradition, craftsmanship and tourism. Finally, the findings were synthesized and integrated. For detailed discussion in the thesis, a subset of scenes was selected based on their representativeness of recurrent multimodal patterns, analytical richness, functional clarity and distribution across the dataset.

CHAPTER 4

MULTIMODAL ORCHESTRATION FOR CONSTRUCTING THE REPRESENTATIONS OF VIETNAMESE STREET FOOD

4.1. Representing Vietnamese street food as a sensory experience

4.1.1. Representing Vietnamese street food as a sensory experience through the concurrence of the verbal, visual and paralinguistic mode

The analysis reveals that Vietnamese street food is multimodally represented as a deeply sensory experience, where meaning is constructed through the interaction of verbal, visual and paralinguistic resources. Across scenes, the five senses - sight, aroma, taste, touch and sound - are not simply mentioned but semiotically performed to transform food into an embodied and affective encounter. When all three modes converge, sensory meanings are vividly enacted through verbal articulation, visual evidence and gestural embodiment. In contrast, when only the verbal and visual modes are active, meaning is achieved through seeing and saying, with visual cues and linguistic precision compensating for the absence of gesture. Through such sensory orchestration, the representation of Vietnamese street food moves beyond pure description to embody a living sensory culture, inviting viewers to taste, see, smell, hear and feel its features.

4.2. Representing Vietnamese street food as tradition-rooted and everyday-life cuisine

The representations of Vietnamese street food as tradition-rooted and embedded in everyday life emerge through a dynamic interplay of intrasemiosis within verbal and visual modes and intersemiosis between them. Verbal descriptions establish coherent ideational chains that specify where the food is eaten, the tools involved, longstanding traditions, the times of day associated with particular dishes and the manner of consumption. These verbal sequences are complemented by visual linking through narrative or conceptual shots of alleys, utensils, cooking setups, vendors and typical dining environment, which independently construct cohesive spatial, temporal and cultural meanings. When combined, verbal and visual modes mutually elaborate one another in order to produce specification alignments that anchor each cultural practice within concrete scenes of everyday life. Notably, paralanguage in the form of hand gestures is largely absent, demonstrating that for this category of ideational meaning, contextual cues from speech and setting are sufficient and embodied gesture plays little role in representing routine cultural practices.

4.3. Representing Vietnamese street food as the product of skilled labor and artisanal mastery

The analysis demonstrates that Vietnamese street food is as the product of skilled labor and joyful creator. Through the concurrence of verbal, visual and paralinguistic modes, the scenes focus on the dexterity and emotional investment of vendors who transform ordinary ingredients into artful dishes. Whereas the verbal mode highlights expertise through descriptions of speed, balance and control, the visual mode showcases rhythmic movements and procedural fluency. Additionally, where present, paralanguage reinforces these qualities through expressive gestures. These modes work together to construct a narrative of craftsmanship and pride in manual skill. Regarding intrasemiosis, by employing visual linking and verbal linking in the preparation scenes, the videos focuses on the vendor's meticulous technique and dedicated attitude. This multimodal representation elevates Vietnamese street food beyond mere sustenance, ultimately framing it as a living art form that celebrates the dignity of labor.

4.4. Representing Vietnamese street food as a culinary challenge

The analysis reveals that Vietnamese street food is represented as a culinary challenge that simultaneously evokes fear and openness. Across the scenes, the interplay of verbal, visual and paralinguistic resources constructs eating as both a test of endurance and a cross-cultural encounter. Verbal expressions of hesitation or disgust (“disgusting,” “frightening”) evolve into reflective and empathetic comments (“open mind,” “another person’s culture”). This forms a coherent intrasemiotic shift from aversion to acceptance. Visually, close-ups of the raw food magnify the sensory intensity of the challenge, while narrative shots situate reactions within a social and dialogic space. Paralinguistic gestures - whether absent or expressive - reinforce emotional tension and negotiation. These modes orchestrate a representation of Vietnamese street food as both a test of personal and cultural boundaries and a means of cultivating understanding through encounter.

CHAPTER 5

MULTIMODAL ORCHESTRATION FOR ENGAGING VIEWERS IN APPRECIATING VIETNAMESE STREET FOOD

5.1. Engaging viewers through eliciting involvement

Viewers' involvement can be constructed through two multimodal patterns. The first pattern relies on meaning congruence across all three modes, where each mode reinforces the same interpersonal stance. When verbal appraisals, expressive facial cues and medium shots align intersemiotically, they collectively draw viewers into the tasting moment to make them feel as though they are co-experiencing the food with the host. The second pattern achieves involvement through the absence of verbal language, where interpersonal meaning emerges instead from the parallel enactment of visual and paralinguistic cues. Together, they demonstrate that involvement can be powerfully achieved even without verbal explanation.

5.2. Engaging viewers through building trust

Reviewers cultivate trustworthiness through two complementary multimodal strategies that reinforce their credibility as narrators of Vietnamese street food. The hosts firstly establish authenticity through congruent multimodal appraisal. This pattern relies on semantic amplification, where the hosts' appreciative language, facial expressions, tone and visual evidence, whether positive or negative, are collectively aligned. This high degree of modal congruence eliminates doubt and is capable of persuading the viewers that the hosts' feelings are genuine and unmediated, thus setting the emotional foundation for credibility. Secondly, the host achieves expert authority through objective attitude. This is realized by strategically employing multimodal divergence - for instance, suppressing intense Paralinguistic Affect while delivering strong Verbal Appraisal. This act of emotional control signals that the appraisal is the result of reflective and critical evaluation rather than mere enthusiasm.

This establishes the host as a meticulous and authoritative guide whose final verdict is earned and trustworthy. Collectively, this orchestration of sincerity and control provides the compelling validation that is necessary for viewers to accept the culinary appreciation of the street food.

5.3. Engaging viewers by providing rational justifications of appreciation

This section shows another strategy that reviewers engage viewers is by offering rational explanations that legitimize their appreciation of Vietnamese street food. Through logical reasoning, reviewers construct a verbally coherent argument of highlighting the nutritional benefits of sandworms which create intrasemiotic cohesion within the verbal mode and position positive evaluations as grounded and credible rather than purely emotional. These verbal justifications interact with visual evidence and paralinguistic cues with a view to creating intersemiosis that reinforces the rational claims. Likewise, reviewers deepen viewer engagement by situating appreciation within geographical and environmental contexts to connecting the dish to its specific location, ecological condition and seasonal cycle. Visual sequences of ponds and raw materials anchor these explanations spatially, while steady paralinguistic delivery of neutral tone signals informational reliability. Across both patterns, meaning arises from the coordinated orchestration of verbal reasoning, visual evidence and stabilising paralinguistic cues, which allows reviewers to engage viewers by presenting Vietnamese street food as not only enjoyable but also nutritionally and environmentally meaningful.

5.4. Engaging viewers through engineering conversion

This section illustrates that reviewers actively guide viewers toward appreciating unfamiliar or potentially intimidating Vietnamese street foods by engineering conversion, a multimodal process that transforms hesitation into openness. Through emotionally balanced reactions, reviewers acknowledge apprehension verbally yet immediately counter it with positive evaluations, creating an intrasemiotic verbal pattern that moves from uncertainty to reassurance. These shifts are reinforced by

paralinguistic markers such as expressive intonation or softened tones and by visual close-ups that demystify the dish in order to generate intersemiosis in which all three modes align with each other. In parallel, the reviewers employ familiar comparisons, where unexpected textures or ingredients are verbally reframed through analogies to known foods, which enables viewers to map the unfamiliar onto recognizable categories. These comparisons are supported visually through shots of the dish being cooked or eaten, which stabilize the analogy, while paralinguistic cues like enthusiastic tone or fast tempo signal confidence in the re-framing. Across both patterns, meaning is constructed through intermodal cooperation, where verbal explanation, visual substantiation and paralinguistic affect jointly reduce perceived strangeness. In this way, engineering conversion functions as a strategic interpersonal tool that makes challenging foods accessible, relatable and ultimately appreciable to viewers.

5.5. Engaging viewers through constructing narration

This section demonstrates that viewers are engaged through the strategic construction of narrative coherence across modes. Through temporal sequencing, reviewers use verbal descriptions, paralinguistic cues and visually ordered shots to guide viewers step-by-step through unfolding actions, thus generating anticipation and shared experiential progression. In parallel, spatial grounding situates the narrative within clearly defined places, where verbal references to specific locations align with visual contextualization and paralinguistic shifts in tone to anchor the viewer in a sense of physical and cultural context. Across both patterns, intersemiosis, particularly between verbal narration and visual sequencing, is central to creating narrative flow. Meanwhile intrasemiosis within each mode ensures internal coherence that stabilizes the storyline. Briefly, these multimodal resources transform simple demonstrations of food into immersive narrative experiences, which enables viewers to follow the journey through both time and space and deepens their engagement with Vietnamese street food.

CHAPTER 6

IDENTITIES OF VIETNAMESE STREET FOOD THROUGH ITS MULTIMODAL REPRESENTATION

6.1. Communal identity of Vietnamese street food

The analysis shows that Vietnamese street food is constructed as a distinctly communal identity through its multimodal representation, emerging as an everyday, publicly embedded and culturally legitimized practice. Across scenes, the verbal narration, visual documentation and paralinguistic cues consistently position street food within shared social rhythms, specifically prepared in open, mobile spaces, consumed in close proximity and embedded in interactions among vendors, diners and passers-by. This communal identity is further reinforced through signs of cultural ratification, including recognizable cooking methods and the visible reputation of long-standing vendors. At the same time, references to seasonal traditions, inherited techniques and iconic ingredients anchor street food in broader cultural continuities.

Figure 6.1

Communal identity of Vietnamese street food in YouTube food review videos



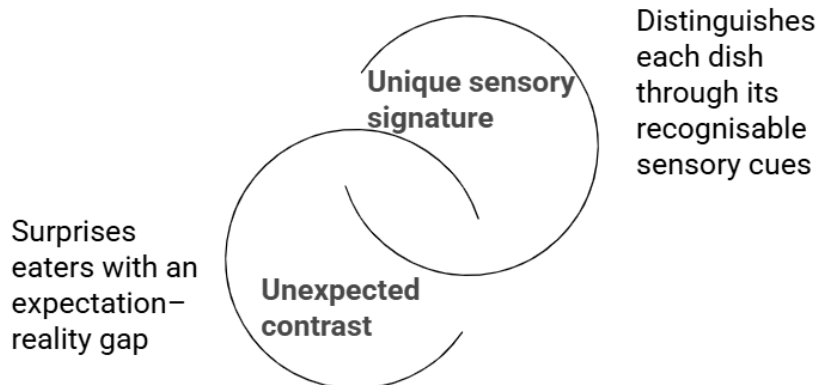
6.2. Particularized identity of Vietnamese street food

The particularized identity of Vietnamese street food is constructed through multimodal cues that foreground each dish's distinctive sensory personality. Verbal descriptors such as detailed references to sound, aroma, texture, sight and taste combine with close-up visuals and occasional paralinguistic cues (hand gestures, facial reactions) to mark the food's unique sensory profile. Unexpected sensory contrasts, such as a confronting appearance paired with a surprisingly delicious flavor, are also multimodally staged to produce a narrative of surprise.

Figure 6.2

Particularized identity of Vietnamese street food in YouTube food review videos

Particularized identity of Vietnamese street food in YouTube food review videos



6.3. Functional identity of Vietnamese street food

The functional identity of Vietnamese street food is shaped by how the dishes are positioned within the social and experiential scripts enacted in the videos. Some foods take on the role of a daring experience framed through verbal anticipation, visual close-ups and paralinguistic reactions that dramatize fear, courage or shock. Others become objects of expertise, where technical vocabulary, shots of skilled preparation and evaluative tone construct the food as professional craft. Dishes function as social binders, facilitating conversation, shared reactions and communal eating

and as cultural ambassadors, used by hosts to explain Vietnamese culinary norms and express local pride.

Figure 6.3

Functional identity of Vietnamese street food in YouTube food review videos



6.4. Lifestyle identity of Vietnamese street food

Lifestyle identity emerges as Vietnamese street food is framed not only as nourishment but as a curated cultural experience aligned with contemporary leisure practices. The multimodal portrayal constructs adventurous eating as a desirable activity with voluntary risk-taking and excitement communicated through enthusiastic tone, expressive gestures and dynamic visuals. The videos also present street food as a means of cultural exploration, offering authentic immersion and satisfying tourist curiosity through contextual commentary and location-rich visuals. Finally, lifestyle identity extends to online participation, where direct address, inclusive camera angles and shared reactions create a sense of viewer involvement and communal enjoyment. Collectively, these strategies situate Vietnamese street food within global lifestyle cultures that center on cultural exploration, digital participation and culinary adventure.

Figure 6.4

Lifestyle identity of Vietnamese street food in YouTube food review videos

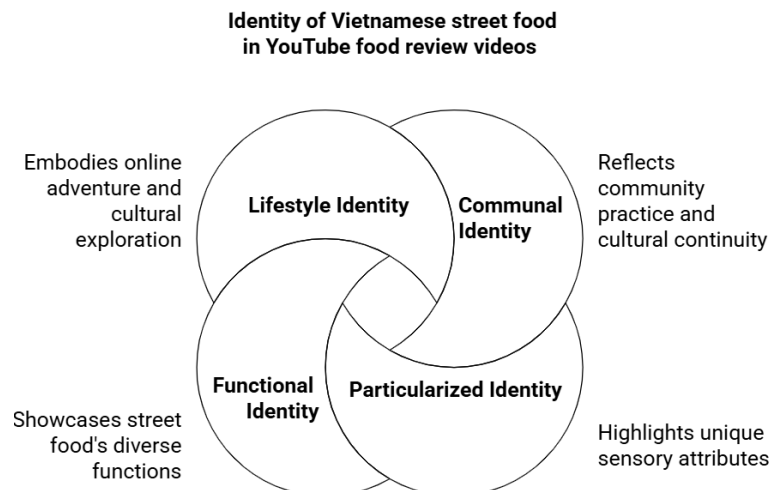


6.5. Identity interactions in multimodal representation

The analysis reveals that the identities of Vietnamese street food are not realised as isolated categories but emerge through layered multimodal cues that operate both within individual scenes and across the dataset as a whole. This intersectional pattern reflects van Leeuwen's (2022) view that identities "often co-exist in complex combinations" (p. 6), formed through the interaction of the four above-mentioned concepts of identity.

Figure 6.5

Multilayered identity of Vietnamese street food in YouTube food review videos



CHAPTER 7

DISCUSSION

7.1. The logical pathway across the findings of three research questions

The three research questions form a collective analytical route that moves from meaning construction, to viewer engagement and ultimately to identity formation.

- RQ1 – What it means: multimodal representation constructs the discursive substance of Vietnamese street food.
- RQ2 – Why it matters: multimodal engagement patterns draw viewers into these meanings.
- RQ3 – What it becomes: identities formed from the accumulation of representational and interactional meanings.

7.2. Vietnamese street food as tourist appeal via multimodal discourse of YouTube food review videos

The findings of this study suggest that YouTube food review videos do more than simply describe Vietnamese street food; they actively participate in shaping the “tourist imagination” of Vietnam’s food culture. Vietnamese street food, represented in this study through its sensory intensity, artisanal expertise and cultural embeddedness, fits directly into this pattern of gastronomic appeal. Digital media plays an especially crucial role in mediating tourist desire towards food. The YouTube creators present Vietnamese street food as accessible, authentic and aesthetically captivating, thereby contributing to Vietnam’s soft-power visibility. The findings of this study significantly boost this touristic appeal. Portrayals of Vietnamese street food correspond to the key quality that food tourists seek – authenticity (Field, 2004; González-Sánchez et al., 2024). Meanwhile, the engagement patterns translate these representations into an accessible narrative for global audiences. Moreover, identity construction further strengthens the connection between multimodal

discourse and tourism. Importantly, these processes are expressed via semiotic means. The multimodal orchestration documented in this study (verbal-paralanguage interaction, verbal–visual intersemiosis, paralanguage–visual parallel realizations of the same stance and instances of intrasemiosis) demonstrates how media discourse creates the conditions for food tourism. It is by rendering taste imaginable, risk manageable and authenticity linguistically/visually/paralinguistically accessible that tourist imagination is shaped. These findings illustrate that semiotic choices are not outer but central patterns to tourism promotion.

7.3. Contributions of the study

This study offers significant theoretical, methodological and empirical contributions to multimodal discourse studies, digital food communication and the understanding of Vietnamese street food in media environments. Theoretically, the research extends social-semiotic multimodality, particularly the notion of multimodal orchestration, by demonstrating that meaning-making in YouTube food reviews is not a fixed or predictable configuration of modes but a flexible and context-responsive semiotic process. Methodologically, the study contributes an adapted multimodal orchestration framework that integrates SFL-based inter/intrasemiosis with Hall's (2013) circuit of culture and van Leeuwen's (2022) multimodal identity theory. Empirically, the study provides one of the original systematic linguistic accounts of Vietnamese street food as represented through global digital platforms. By identifying four identity types - communal, particularized, functional and lifestyle - the findings contribute to the literature on Vietnamese street food. Overall, the study enriches current knowledge by showing that Vietnamese street food, when mediated through YouTube review videos, becomes not only a culinary object but a multimodally constructed cultural identity, shaped by representation, engagement and the communicative logics of digital media.

CHAPTER 8: CONCLUSION

8.1. Recap of key findings

Findings of this study show that YouTube representations of Vietnamese street food do not merely document food practices but actively construct cultural meaning, value, and identity for global audiences. By demonstrating how verbal, visual, and paralinguistic resources are flexibly orchestrated, the study suggests that appreciation and authenticity are produced through multimodal choices rather than residing inherently in the food itself. The identification of layered and accumulating identities further indicates that food identity in digital media is dynamic, relational, context-dependent and discursively constructed, rather than fixed or singular. These insights advance multimodal discourse analysis by illustrating how identity construction unfolds across scenes and interactions, not isolated moments. The study also sheds light on how digital food reviews function as cultural mediators, transforming everyday street food into a sensory, social, and symbolic experience.

8.2. Implications of the study

This study offers a number of important implications. Theoretically, it extends SFL-informed multimodal frameworks beyond traditional written and spoken texts to dynamic, user-generated video discourse, demonstrating how ideational and interpersonal meanings are jointly realized through flexible multimodal orchestration rather than fixed modal hierarchies. Applying van Leeuwen's identity framework to food rather than human actors, the study advances identity research by showing that food can function as a semiotic actor with layered, cumulative, and interactionally activated identities (communal, particularized, functional and lifestyle). Methodologically, the study offers a replicable qualitative model for analyzing online video discourse. The integration of Social Semiotic analysis of ideational and interpersonal metafunctions with reflexive thematic analysis for identity construction provides a robust

multi-layered analytical procedure. This approach is particularly valuable for researchers studying YouTube, TikTok, Instagram Reels, or other video-centric platforms. Practically, the findings have clear relevance for content creators, tourism marketers and cultural policymakers. Pedagogically, the findings also hold implications for language, media and communication studies.

8.3. Limitations of the study

This study is not without limitations which are acknowledged below. Firstly, the dataset is intentionally small and purposive, consisting of three English-language YouTube food review videos. This limits the generalizability of the findings across platforms, languages and genres of food discourse. Secondly, the study focuses exclusively on English-language content, which privileges representations oriented toward international audiences. As a result, the findings may not fully capture how Vietnamese street food is represented and negotiated in Vietnamese-language videos or in locally oriented digital contexts. Thirdly, although viewer engagement is analytically inferred from multimodal strategies, the study does not include audience reception data such as comment analysis or interviews. Consequently, claims about viewer appreciation and interpretation remain analytically grounded rather than empirically validated from the audience's perspective. Finally, the analysis centers on representational and interpersonal metafunctions, leaving the textual/metafunctional dimension (e.g., cohesion, rhythm, editing logics across entire videos) relatively underexplored.

8.4. Suggestions for future research

Building on the above mentioned limitations, several directions for future research are recommended. Future studies could adopt a comparative design to examine Vietnamese street food alongside other national cuisines or compare representations across platforms such as YouTube, TikTok and Instagram. This would illuminate how platform affordances and

constraints shape multimodal identity construction. Another possibility is to expand the linguistic scope to include Vietnamese-language or bilingual food review videos would allow researchers to compare inward-facing and outward-facing culinary representations and explore how national identity is negotiated differently for domestic versus international audiences. In addition, research could incorporate audience reception methods such as comment discourse analysis, surveys, or interviews, to triangulate how viewers interpret, negotiate or resist the identities constructed in food review videos. Finally, longitudinal research could examine how multimodal representations of Vietnamese street food evolve over time, particularly in relation to tourism development, globalization or platform algorithm changes.

8.5. Reflections on the research

From the start of my PhD, my motivation was largely personal. As a Vietnamese person born and raised in Hanoi, street food was never merely food. To me, it was part of the rhythm of everyday life, embedded in memories, social interactions and a sense of belonging. This doctoral research marked a turning point, in which it required me to confront the challenge of theorizing the everyday, for example the ordinary plastic stool, the street-side vendor, the fleeting act of eating as meaningful semiotic and cultural phenomena. Completing this doctoral project marks a form of intellectual maturation. I have learned to balance personal involvement with analytical precision and to situate my own voice within scholarly debates. Looking back, the journey from the opening quotation to the final chapter mirrors my own development as a researcher: from experiencing street food as art to understanding how that art is semiotically constructed, mediated and made meaningful in contemporary digital culture.