THE EXOTIC AND THE EXPLORER: A MULTIMODAL ANALYSIS OF IDENTITY CONSTRUCTION IN A VIETNAMESE STREET FOOD REVIEW VIDEO

Nguyen Hong Lien

Faculty of English, Hanoi National University of Education, Hanoi city, Vietnam

*Corresponding author: Nguyen Hong Lien, e-mail: nguyenhonglien@hnue.edu.vn

Abstract. This study investigates the multimodal construction of identity in food media, specifically analyzing how verbal, visual, and paralinguistic modes interact to represent identities. Focusing on the most viewed YouTube review video about Vietnamese street food, the research applies Systemic Functional Linguistics to uncover social, individual, role, and lifestyle identities based on van Leeuwen's framework of identity (2022). Findings reveal that these multimodal elements work synergistically to create an immersive experience where the food is constructed as an exotic culinary experience in relation to its perceived otherness from mainstream norms. Concurrently, reviewers leverage these resources to manifest their own identities as professional media personalities, adventurous food explorers, and expressive critics. This dual identity construction highlights how food media not only portrays diverse cuisines but also shapes public perception within digital discourse. This study contributes to the global appreciation for Vietnamese street food, thereby providing a new academic model for multimodal analysis of food media and offering practical insights for content creators to craft more convincing and authentic portrayals that challenge cultural stereotypes.

Keywords: identity construction, Vietnamese street food, Multimodal Discourse Analysis.

1. Introduction

Vietnamese street food is deemed to be favored around the world and has gained international recognition. The international food magazine TasteAtlas has announced that 26 dishes of Vietnamese street food have been listed in the top 100 street foods in Southeast Asia for 2025 (Tam Anh, 2025) [1]. Alongside pho and banh mi, crispy pancakes, chicken sticky rice and more are chosen to be in the list. Vietnamese street food has also been featured in the Netflix series about "Street Food: Asia" in 2019 as viewers are invited to embark on a global cultural journey into street food and discover the stories of the people who create the flavorful dishes (Nguyen, 2019) [2]. Hence, understanding how this vibrant cornerstone of Vietnamese cuisine is represented in the international sphere can not only spark interest in Vietnam as a travel destination to boost tourism but also shape Vietnamese culinary identity among global audiences.

Food, in general, and street food, in particular, can profoundly express identity because it serves as a powerful symbolic and social phenomenon, intrinsically linked to how individuals and groups perceive themselves (Cohen, 2024 [3]; Stano, 2023 [4]). Anthropologist Eugene Anderson notes in Everyone Eats: Understanding Food and Culture (2014) that food is a universal means of

communication in all communities [5]. Through food choices and practices, people actively convey who and what they are, to themselves and to others. Food can be an expressive dimension of various identity traits, including race, ethnicity, religion, national origin, gender, age, and socioeconomic status (Anderson, 2014 [5]). The act of consuming food can incorporate an eater into a culinary system and, by extension, into the group that practices it, while also serving to differentiate or separate groups (Fischler, 1988, as cited in Cohen, 2024 [3]). Ultimately, food culture provides social and cultural food rules that help define and characterize a group or nation to locate "what is familiar and ours from what is not" (Ranta & Ichijo, 2022, p. 11) [6]. However, investigation into Vietnamese street food is lacking in both the domestic sphere and the foreign one. Therefore, a study that examines Vietnamese street food in YouTube's review videos about Vietnamese street food from a multimodal lens would be of importance.

Thanks to the multimodal nature of videos, multimodal discourse analysis (MDA) is a powerful tool for investigating YouTube videos about street food. MDA allows researchers to analyze various modes of communication (e.g., visual, verbal, paralanguage, etc.) simultaneously to understand how meaning is constructed and communicated. Researchers can determine which semiotic principles are common to all modes of meaning production and how these principles are implemented differently in each mode by examining meaning-making across modes (Jewitt et al., 2025) [7].

With all the above-mentioned reasons, the study seeks answers to two research questions, with one focusing on the portrayal construction - how Vietnamese street food is shown to the audience, and the other elaborating on the identity reflection – what identity of Vietnamese street food and that of reviewers are construed in relation to others. The research questions are:

(1) How do verbal, visual, and paralanguage modes interact to construct the portrayal of Vietnamese street food in the most-watched review video about Vietnamese street food on YouTube?

(2) What identity of Vietnamese street food and the reviewers do the portrayal reflect?

2. Content

2.1. Literature review

2.1.1. The concept of identity
2.1.1.1. Multimodality and identity
The concept of identity is central to contemporary social and cultural life. Identity is broadly understood as how one sees oneself and is seen by others (Cohen, 2024) [3]. While it may appear to denote something "fixed and definite" (van Leeuwen, 2022, p. 5) [8], this study argues that it can be dynamic and fluid since it "can shift and recombine to meet new circumstances" and "is interactionally negotiated" (Bucholtz and Hall, 2004, p.375 [9]).

Adopting a social semiotics approach, van Leeuwen (2022) believes that identity and multimodality are strongly associated [8]. He claims that understanding identity requires drawing on two distinct types of resources: sociological and philosophical resources for comprehending what identity is, and semiotic resources for analyzing how identity is expressed. van Leeuwen (2022) highlights four primary concepts of identity [8]:

(1) Social identity is defined by an individual's place in a pre-existing social order and their relationships to others, such as kinship systems (p. 6) [8]. This form of identity is often expressed visibly on the body, like Maori face tattooings (mokos) that signify rank, position, marital status, kinship, work, prestige, sacred power, and birth status, while also conveying spiritual and moral significance.

(2) Individual identity emphasizes the uniqueness and inwardness of the self. In this situation, the self is "no longer a social self, but an individual self which perceived itself as separate from society" (p. 11) [8]. Its expression is often subtle, regarded as unconscious symptoms of character rather than explicit social signs.

(3) Role identity, which views identity as multiple and changeable, based on the functional scripts individuals play. This understanding necessitates social recognition and is actively managed through "expressive equipment" (ibid., p. 20) such as setting, appearance, and manner to convey role-specific values like "cleanliness, modernity, competence, integrity" (Goffman, 1959, p. 35 as cited in van Leeuwen, 2022, p. 20) [8].

- Lifestyle identity centers on individually chosen leisure activities and consumption patterns that aim to connect individuals with like-minded communities. This identity is fluid, evolving with trends, and is more about "psychographics" (social trends, attitudes, lifestyles) than traditional demographics (van Leeuwen, 2022, p.20) [12]. Its expression involves styles of dress, consumer goods, and settings.

2.1.1.2. *Food and identity*

Food is profoundly connected to identity, serving as a powerful lens through which individuals and groups define themselves and are perceived by others (Reddy & van Dam, 2020 [10]; Cohen, 2024 [3]). This relationship extends beyond basic nourishment and encompasses intricate social, cultural, and political dimensions (Ranta and Ichijo, 2022) [6].

According to Stano (2023), food functions as a "sign" and a "system of signification and communication" that involves substances, practices, habits, and preparation techniques to convey meaning (p.308) [4]. For instance, showing how food habits are capable of manifesting cultures and historical traditions is the practice of avoiding waste and making clever use of leftovers in Ranta and Ichijo's study (2022) [6]. This practice is linked to the historical idea of being a "good wife and wise mother" in modern Japanese culture (p.35) [6].

Regarding self-present identity, individuals can define themselves through the consumption of the chosen food. Picky eaters in online discussions construct their identity by highlighting limited food choices and problematizing their eating habits (İkizoğlu & Gordon, 2022) [11]. They often describe a split self, where a rational adult self tries to force or trick a childlike picky-eating self into healthier eating, linguistically expressed through reflexive pronouns like "forcing myself to have plain Greek yogurt" or "teach myself not to be" (İkizoğlu & Gordon, p. 26) [11].

Food also serves as a boundary-setting mechanism, with specific food items, practices, or taboos used to define group and nation identity. In Singapore, Malay participants assert their cultural identity by preparing a greater variety of dishes for the fasting month and explicitly stating their food practices are "not like other races" (Reddy & van Dam, 2020, p. 4) [10]. Shared food habits provide a strong sense of belonging and cultural affirmation, as exemplified by the Burns Supper for celebrating Scottishness or Vietnamese New Year rice cakes, solidifying group identity and performance of the nation (Ranta & Ichijo, 2022) [6].

2.1.2. Food review discourse

2.1.2.1. *Product review videos*

According to Wang (2024), product review videos, as a type of discourse, introduce "the features, functions, advantages and disadvantages of the product to the audience through video" (p. 9) [12]. Often created by influencers, experts, or everyday consumers, they combine visual demonstrations with personal insights, making the review both informative and relatable.

Defining characteristics of product review videos are elaborated as follows. The initial element is their authenticity. By showing the product in action, they provide a level of transparency that written reviews or advertisements often lack. Viewers can see how a product performs in real-life scenarios, gaining a better understanding of its strengths and limitations.

Hence, video reviews are more trusted and persuasive to consumers than written reviews (Xu et al. 2015) [13]. Another aspect is multimodal communication. Product review videos rely not just on spoken language but also on visual and gestural cues, which complement verbal expressions (Jaakkola, 2020) [14]. For instance, a reviewer might emphasize a point by gesturing toward a product feature or using intonation to highlight its importance. The interplay between these modes enhances clarity and persuasiveness. Moreover, the use of evaluative language is pivotal thanks to its primary aim of expressing opinions, assessment, and stance (Cheng, 2024) [15]. Adjectives like "amazing," "disappointing," or "versatile" can be frequently employed to express opinions, while modal verbs such as "should" or "might" offer recommendations or convey uncertainty.

2.1.2.2. Online food reviews

The popularity of food reviews, recipe videos, and cooking tutorials has turned food into not just a necessity but a form of entertainment and exploration. According to Nielsen (2014) cited in Sahelices-Pinto et al. (2017) [16], 52% of internet users performed roles as opinion seekers, carrying out online processes of search for information about recipes, gastronomy, nutritional details of products, reviews of restaurants, or consumption experiences with brands before making a final decision about a purchase. In Bi's research (2018) [17], food reviews are popular among the seven types of products reviewed. In addition, the availability of online information displayed on YouTube, such as food vlogger reviews or recommendations, makes them easily accessible (Briliana et al., 2020) [18].

When consumers see YouTube shows, they look for information about vlogger reviews of the food and drinks offered as street food, including menu variations and recommendations, prices, how to consume the food, hours of operation, and the atmosphere. Food vloggers start by showing how the food is cooked and served, accompanied by spontaneous interviews between the food vloggers and the sellers or chefs. In the phase where the food vlogger is actually eating the food, sharing the experience of consuming becomes most important for inspiring purchasing decisions (Briliana et al., 2020) [18].

However, despite its immense popularity, research on online food reviews in video format has been extremely scarce. Therefore, a study that researches how Vietnamese street food is represented in online review videos is worth carrying out to bridge the existing gap.

2.2. Methodology

2.2.1. Research approach of the study

Grounded in a social semiotics theory, this case study adopts a qualitative approach. This study employs Systemic Functional Linguistics (SFL) as an analytical tool. In SFL, the three metafunctions - ideational, interpersonal, and textual (Halliday and Matthiessen, 2004) [19] - serve as powerful analytical tools for examining videos, particularly their linguistic and multimodal content. Due to the limit of the scope, this study is set to center around the construction of ideational and interpersonal meaning via selected modes. Ideational metafunction focuses on the representation of content and meaning (Halliday & Matthiessen, 2004) [19]. In video analysis, this metafunction can uncover how visuals, spoken language, and other semiotic elements represent ideas, experiences, or events. Interpersonal metafunction examines the relationship between speaker and hearer (Halliday & Matthiessen, 2004) [19], in particular the interaction between the creator of the video and its audience in the video format. It can investigate how language, tone, gestures, and gaze establish relationships and convey attitudes, such as authority, friendliness, or persuasion.

This case study investigates the most viewed YouTube video exploring Vietnamese street food, which belongs to an English-speaking channel. This channel distinguishes itself by pioneering a fresh, impactful approach to exploring global cuisines, which uniquely emphasizes the discovery of each country's distinctive foods through a dynamic presentation that fosters

cultural empathy and understanding. Based in Vietnam, the host created the channel on September 23, 2010, and uploaded his first video in December 2015. As of June 2025, the channel boasts approximately 11.3 million subscribers. The video tells about a journey in which the hosts discover local street food in Hanoi. The hosts are anonymized as Sobe and Tien while the screenshots are applied with filters not to reveal the identity of the content creator explicitly. The video's content is about reviewing street food in Hanoi, including water bug extract, sandworm cakes, roasted quails with Vietnamese curry leaves, pig penal and others. The video, lasting 13.07 minutes, was uploaded in 2018, and as of July 2024, it records the view count of 17,965,921 times.

The data processing consists of the following steps. Firstly, the video was chosen based on its topic, its popularity and its representativeness. Secondly, the video was transcribed and checked manually for accuracy. Thirdly, the data was analyzed according to the proposed analytical framework, which will be detailed in the next section. All of the analysis focused on the ideational and interpersonal meaning. Lastly, the representations of Vietnamese street food, in addition to the identities it uncovers, were synthesized across three modes, namely verbal, visual and paralanguage ones.

While the analyses of much research are detailed and systematic, they typically involve a focused and selective approach to the video data rather than a complete analysis of every single clause or image. In Liang and Lim's article (2024) [20], they identify and analyze clauses, images, or sequences that are particularly salient in realizing these metafunctions in relation to their research focus, rather than an exhaustive analysis of all video content related to the topic. This selective approach is necessary to maintain focus and depth in the analysis while managing the voluminous data that videos can present. Instead of coding every single frame or sentence, researchers identify critical moments that best illustrate how meaning is constructed through the interplay of multimodal elements. Therefore, this study focuses on the scenes when the hosts try the food to effectively showcase their appreciation of the dishes.

2.2.2. Analytical framework of the study

2.2.2.1. Constructing the portrayal of Vietnamese street food ideationally

For the analysis of verbal modes, this study utilized Halliday's Transitivity System (Halliday & Matthiessen, 2004) [19]. By identifying the different types of processes - Material (doing), Behavioral (behaving), Mental (sensing), Verbal (saying), Relational (being/attributing), and Existential (existing) – the results systematically uncovered how the verbal language in the YouTube videos constructed the ideational meaning of Vietnamese street food. This included how dishes were described in terms of their preparation, sensory attributes, ingredients, and the actions associated with consuming them.

The visual modes in the videos were analyzed through the lens of Kress and van Leeuwen's Visual Grammar (2021) [21], specifically focusing on their distinction between conceptual and narrative processes. Conceptual representations depict participants in terms of their stable or timeless essence, presenting them as a static concept (e.g., a close-up shot of an ingredient). In contrast, narrative representations present unfolding actions, events, and processes of change, typically identified by vectors that connect participants in an action (e.g., a shot showing a hand mixing ingredients). By applying this framework, the study determined how visual compositions contributed to representing the characteristics, preparation, and consumption of Vietnamese street food, highlighting either its static attributes or its dynamic processes.

The analysis of paralanguage modes drew upon the concept of semovergent paralanguage, as developed by Ngo et al. (2022) [22]. Semovergent paralanguage specifically investigates how non-verbal cues align with and reinforce the ideational meaning expressed verbally. This framework examines how hand shapes (e.g., mimicking an entity) or hand/arm movements (e.g., depicting an occurrence) concur with and enhance the verbal representations of entities,

occurrences, and qualities related to Vietnamese street food, thereby providing deeper insights into its multimodal construction.

2.2.2.2. Constructing the portrayal of Vietnamese street food interpersonally

Verbal modes were analyzed using Halliday and Matthiessen's Mood system (2004) [19] to identify interrogative, declarative, and imperative structures, alongside with Martin and White's Appraisal framework (2005) [23] to explore attitudes, including affect (emotions of the speaker/writer), judgment (assessment of human behaviors based on normative principles), and appreciation (evaluation of things) in addition to engagement and graduation. Verbal modes in this study are spoken language, which includes voiceover and the spoken discourse presented by the characters in the videos.

Additionally, Kress and van Leeuwen's (2021) [21] visual grammar framework provides insight into visual meaning-making, analyzing contact (demand/offer), social distance (personal, social, impersonal), and attitude (involvement, detachment, viewer power, equality, participant power).

Concerning paralanguage, according to Ngo et al (2022) [22], paralanguage includes gestural resources (including gesticulation, pantomime and emblems, but not sign language); facial expression, posture and body movement; and the vocalizations. In this study, the interpersonal meaning in paralanguage was expressed via expression, voice quality, body gestures, expressing feelings and enacting social relations in cooperation with spoken language. Paralinguistic affect modeled expressions of emotion in facial affect with features realized through muscle movement of the face, and in voice affect with features realized through qualities of the voice.

2.2.2.3. Applying the Intermodal relation

The multimodal interaction to construct meanings of three selected modes were realized using Zhang's model (2009) [24], which categorizes intermodal relationships into complementary relationships or non-complementary relationships. Complementary relationships include Reinforcement (Highlighting, Primary and secondary, and Extension) and Non-reinforcement (Intersection, Coalition, Coordination). Non-complementary consists of three main relations, namely Overlapping (Redundancy, Exclusion, Counteraction), Inclusion (Whole and part, Abstract and concrete) and Context Interaction (Context-independence, Context-dependence) (Zhang, 2022 [25]).
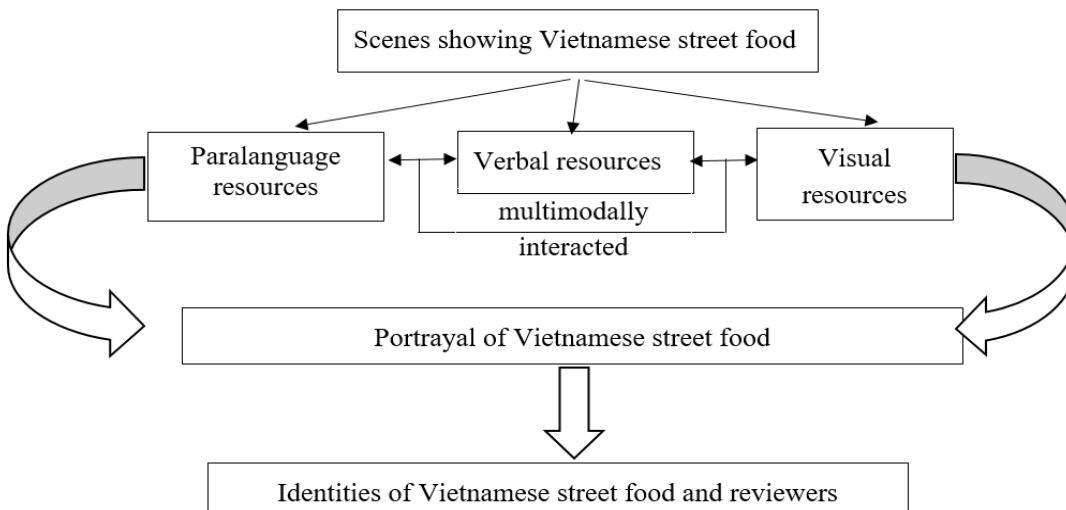
2.2.2.4. Uncovering the identities

Figure 1. The analytical framework of the study

Van Leeuwen's framework of identity (2022) [8] offers a powerful lens for understanding how identity is constructed through semiotic resources, making it equally applicable to humans and objects. According to the framework, identity is a dynamic, socially mediated concept that can be expressed through a variety of communication channels. For humans, this involves how individuals present themselves through language, appearance, and social interactions to project social, individual, role, or lifestyle identities. Similarly, objects are meticulously designed, marketed, and consumed in ways that instill them with comparable layers of identity. By analyzing the semiotic choices embedded in an object's creation, representation, and use, one can uncover its intended social connections, functional roles, and alignment with consumer lifestyles, therefore demonstrating the framework's applicability beyond human subjects.

The analytical framework of the study is visualized in Figure 1.

2.3. Findings and discussion of the study

2.3.1. The multimodal portrayal of Vietnamese street food

"Bizarre and visually shocking" have been found prominent in the data. Due to limited space, two examples (the data is embedded in the QR code at the end of the article) that best illustrate the portrayal will be displayed subsequently.

Extract 1 shows a scene where the hosts try the "water bug extract," an add-on for the classic Hanoian street food, banh cuon.

Ideationally, the water bug is primarily represented as a peculiar entity with surprising sensory attributes and challenging consumption processes. Verbally, the initial Relational processes like "it is like a perfume" and "smells like an appletini" immediately establish its unexpected aromatic qualities, while the later Behavioral process "the body feels so disgusting" highlights an undesirable physical sensation. These verbal descriptions are visually underscored by the "close-up" shot, which, as a Conceptual representation, allows for a detailed, almost scientific examination of the bug's static form, emphasizing its unusual appearance. Simultaneously, the visual "two people eating" acts as a Narrative process, depicting the unfolding action of consumption. This is further enhanced by paralanguage, where Sobe's "finger on neck to show it's like a perfume" semovergently mimics the verbal attribute, and the slow, deliberate movement of "slowly put the water bug inside their mouth" depicts the challenging occurrence, collectively constructing the water bug's identity as a strange, yet tangible, object of consumption.

The scene powerfully engages the viewer and conveys the reviewer's immediate attitudes towards the water bug, shaping its interpersonal representation as a shocking experience. Verbally, the exclamative "Whoa" and the affect appraisal type of "disgusting" instantly signal strong, negative emotional reactions. Visually, the "close-up" shot creates a sense of intimacy and low social distance, forcing the viewer into a direct, unmediated encounter with the unusual food. The "eye-level angle" positions the viewer as an equal, sharing the raw experience rather than observing from a detached perspective. Paralinguistically, the initial vocal "whoa" (Voice affect: surprise) and the facial expression that "shows disgust" (Facial affect: disdain) are crucial. These non-verbal cues directly convey the reviewer's direct, unfiltered shock, positioning the water bug as an object that elicits strong, challenging emotional responses from both the reviewer and, by extension, the engaged audience.

The representation of the water bug as bizarre and visually shocking is achieved through a dynamic and complex intermodal interplay, particularly through strong complementarity and moments of subtle counteraction. The verbal declaration "Whoa, it is like a perfume" is immediately complemented and reinforced by Sobe's paralanguage (finger on neck gesture), creating a strong, multi-sensory impression of its unusual aroma. However, the most prominent interplay for "bizarre and shocking" occurs when the verbal "body feels so disgusting" finds powerful reinforcement in Sobe's paralanguage (facial expression showing disgust), while the

visual close-up on the act of "slowly put[ting] the water bug inside their mouth" elaborates on the physical challenge, drawing the viewer into the confronting experience. This initial, overwhelming convergence of modes on the negative and unusual aspects of the bug creates its primary portrayal as bizarre and visually shocking, setting up a dramatic contrast with the later, surprising positive appraisal of its flavor, which then shifts the intermodal relations towards a different kind of complementarity.

Briefly, the hosts' experience with the water bug extract is portrayed as a shocking but authentic culinary challenge. Verbally, they describe the bug as both a peculiar "perfume" and a "disgusting" physical sensation, while an eye-level camera angle and close-up shots display a direct, intimate encounter with the food. This is powerfully reinforced by paralanguage, such as facial expressions of disgust and a surprised vocal "whoa," which collectively construct the bug's identity as a bizarre and confronting food. The strong intermodal convergence of these modes highlights the surprising and challenging nature of the experience.

Extract 2 shows a scene where the hosts try a treat found in Hanoi's old quarter, described as sandworm pancakes.

Ideationally, the scene meticulously represents the process of preparing the sandworm pancake, highlighting its unusual and dynamic characteristics. Verbally, Material processes dominate, detailing the actions: "She's frying it up," "putting it right into heavy oil," and crucially, "putting the worms on top." The unexpected nature of these participants is emphasized by the observation, "The worms are still moving," and their subsequent action, "They are moving, writhing, in there," describing an unusual self-mixing. Visually, the "close up of putting the pancake in heavy oil and putting live sandworms on top" acts as a powerful Narrative representation, explicitly showing the unfolding, dynamic process with vectors connecting the vendor's hand to the ingredients. The "high angle" further positions the viewer to observe this peculiar creation process in detail. While paralanguage primarily serves interpersonal functions in this scene, the "hands cover mouth" gesture, in a semiverbal sense, physically depicts the initial disbelief at the occurrence of live worms being added, thereby enhancing the ideational representation of an unusual preparation method.

Interpersonally, the scene is rich with expressions of surprise, disdain, and a carefully curated open-mindedness that shapes the reviewer's relationship with both the dish and the audience. Verbally, initial shock is conveyed through exclamatives like "Oh" and "What the?", while "The worms are still moving" carries an undertone of disbelief. The appraisal of "disgusting" is initially hinted at through Sobe's rhetorical question, "people don't feel bad for worms, right?", and explicitly shown via paralanguage. Visually, the "high angle" on the cooking process, while ideationally showing the action, also positions the viewer in a somewhat detached, observing stance, possibly emphasizing the otherness of the dish. The "medium shot covering two hosts looking at the vendor preparing the pancake" maintains an interpersonal distance that allows the audience to observe the reviewers' reactions. Paralinguistically, the "surprise voice 'oh'" directly conveys immediate effect. Crucially, the "disdain facial affect" and "hands covering mouth" are powerful non-verbal appraisals of disgust or aversion. Yet, this initial negative affect is subtly countered by a "spirit down voice saying 'She'd mmm, good'," indicating a complex internal struggle where positive verbal appraisal ("good," "cool," "excites me") clashes with the suppressed vocal affect. This crafts the reviewer's persona as both shocked and "culinarily woke" for the audience.

The representation of the dish as initially shocking and bizarre, followed by a nuanced acceptance, is constructed through dynamic intermodal interplay. The immediate "shock" is powerfully created through complementary reinforcement: Sobe's verbal exclamations "What the?" and "The worms are still moving" are reinforced and highlighted by his paralanguage (surprise voice "oh," facial affect showing disdain, hands covering mouth). Simultaneously, the

visual "close up of putting the pancake... and putting live sandworms on top from high angle" provides vivid elaboration, showing the very action that triggers this multimodal shock. This initial strong alignment across all three modes immerses the viewer in the reviewers' disbelief. As the scene progresses, a more complex non-complementary counteraction emerges. Sobe's paralanguage with a spirit-down voice saying 'She'd mmm, good' (indicating the adding of live sandworms is good) subtly counteracts the full enthusiasm of his verbal appraisal. This disjunction between positive verbal framing and hesitant vocal affect highlights the internal struggle and ultimately constructs a more complex identity for the dish – one that is bizarre enough to elicit genuine shock, yet appealing enough to be embraced by an adventurous food explorer.

Accordingly, a complex culinary adventure is represented in the experience of the hosts with the sandworm pancake. A high-angle, close-up shot of the cooking process, combined with verbal descriptions of the worms "writhing," creates a sense of surprise and confrontation. This is further amplified by paralanguage like a surprised vocal "oh" and facial expressions of disdain. However, a crucial moment of counteraction emerges as the hosts' positive verbal appraisals clash with their hesitant vocal tones, thus revealing a genuine internal struggle. This interplay of modes ultimately constructs a portrayal of the dish as both bizarre and appealing and reinforces the hosts' identity as adventurous food explorers.

2.3.2. Identities of Vietnamese street food and reviewers

The encounters with the water bug extract and sandworm pancakes provide compelling case studies for understanding how van Leeuwen's four types of identity (2022) [8] are simultaneously represented in multimodal food media.

The manifestation of *social identity* is central to both extracts, which emerges from the interaction between the food's perceived status and the reviewers' cultural positioning. Both the water bug extract, described as peculiar with challenging consumption processes, and the sandworm pancakes, highlighted by "live sandworms" that are "still moving" and "writhing," acquire a social identity as culturally othered or unconventional foods. This identity is established through their stark contrast with an implied, more mainstream culinary norm. The reviewers' immediate paralinguistic responses, such as Sobe's facial affect showing disgust and hands covering mouth, function as powerful semiotic signals of their own social identity as individuals whose cultural foodways diverge from those presented dishes. These reactions position the foods as outside the familiar boundaries of their social group, thereby reinforcing the food's unusual social identity and highlighting the reviewers' own cultural affiliation. Furthermore, Sobe's rhetorical interrogative in the sandworm cake scene, "So a lot of foreigners eat this, huh?" directly invokes a social dimension approving the acceptance and typical consumption patterns of such food. This highlights such a particular social group as adventurous foreigners who engage with these extreme aspects of Vietnamese street food. The initial disdain shown in paralanguage across both scenes also suggests a universal human reaction to such unusual foods. Afterwards, this feeling is overridden by a performance of acceptance, creating a social space for those who dare to transcend common culinary boundaries.

The *individual identity* of both the food and the reviewers is co-constructed through a shared experience of unique and often trembling sensory attributes. The representation of both the water bug and sandworm cake as "bizarre and visually shocking" infuses them with a distinct identity as unique entities with challenging qualities. The water bug extract develops its distinct individual identity through its surprising aromatic qualities—verbally described by the Relational processes as "like a perfume" and smelling "like an appletini"—and its undesirable physical appearance by the Behavioral process as "the body feels so disgusting." Similarly, the sandworm cake's individual identity as bizarre is solidified verbally through exclamations like "What the? I thought that's premix. The worms are still moving," emphasizing its unusual living components. Visually, the close-up, eye-level angle of two people consuming the water bug, and the close-up shot of

putting the pancake in heavy oil and putting live sandworms on top from a high angle, both confront the viewer directly with the shocking visual reality of these dishes. Paralinguistically, immediate "whoa", surprise voice "oh," and faces showing disdain across both scenes vividly capture the shock these foods uniquely provoke, thus establishing their individual identity as remarkably unusual and challenging. These unusual sensory profiles of the foods directly elicit and manifest the reviewer's own individual identity as an expressive reviewer. Their spontaneous vocalizations ("Whoa"), genuine expressions of aversion (disgusting verbal appraisal, disdainful facial affect), and the visible internal struggle (the spirit down voice saying "She'd mmm, good" counteracting initial shock) reveal their identity to challenging experiences.

In terms of *role identity*, both extracts showcase a clear manifestation of this type, with the food and reviewers collaboratively participating in a mediated performance of culinary discovery. Both dishes, through its close-up visuals, allow detailed and almost scientific examination, take on the role of specimens for analysis within a structured review format. This positions the dishes as objects through which the reviewers fulfill their role identity as professional food explorers. In the sandworm pancake scene, despite paralinguistic cues of disdain and hands covering mouth - authentic initial reactions to the bizarre preparation - Sobe's verbal reframing ("This is cool," "It excites me") transforms the shocking experience into an appealing one, consistent with his asserted "culinarily woke" role. Similarly, in the water bug scene, Sobe's journey from initial verbal "disgusting" to a final "amazing" through paralanguage, such as eyes closing to taste, showcases his individual capacity to overcome challenges. The reviewers strategically employ visual techniques like the "eye-level angle" to establish intimacy and shared experience, thereby performing their role as engaging facilitators of the audience's encounter with the unusual. This presentation enables the reviewers to effectively perform their role identity as professional media personalities who not only show their willingness to confront culinary extremes but also make it even desirable to an audience who values authenticity and bold experiences.

Finally, both the food items and the reviewers actively contribute to constructing a shared *lifestyle identity*, centered on adventurous culinary exploration. These dishes are framed as to-be-conquered experiences, aligning with a thrill-seeking and open-minded approach to food. Both dishes, represented as bizarre and visually shocking, acquire a lifestyle identity as an exotic curiosity. In parallel, the reviewers consistently manifest their own lifestyle identity as culinary adventurers through their engagement with such items. Verbally, Sobe's explicit self-description in the sandworm pancake scene as "someone who's open-minded and might I say, culinarily woke. It excites me," directly links the consumption of such unconventional food to a particular adventurous lifestyle. While not as explicitly stated in the water bug scene, the reviewer's journey from initial "disgusting" to "amazing" (verbal appraisal) through paralinguistic "eyes close to taste" and "hand shape showing the flavor permeating out" enacts this same adventurous journey. The visuals across both scenes, from the intimate close-up of eating the water bug to the confronting process of seeing how live worms are added to the pancake, invite the audience to either indirectly participate in or admire this boundary-pushing culinary lifestyle. The reviewers' managed reactions, particularly the subtle interplay between initial aversion and later acceptance, reinforce this lifestyle identity as culinarily woke.

2.3.3. Discussion

Upon analyzing how verbal language interacts with paralanguage and visual elements to construe meanings, it becomes evident that these modes do not operate in isolation but interact to create a compelling and immersive experience for the audience. The results of this study illuminate how various multimodal resources complement or counteract to portray a cultural phenomenon, which aligns with prior research (Lu, 2024 [26]; Zhang, 2022 [25]).

The findings of the study demonstrate how unusual foods, like the water bug extract and sandworm pancakes, acquire a social identity rooted in their perceived otherness from a presumed

culinary norm. This aligns with prior research that identifies certain Vietnamese food items, particularly insects, as exotic and even repulsive to non-Vietnamese audiences, since they tap into Western taboos against insect-eating (Avieli, 2011) [27]. By strategically highlighting the strange visual and physical attributes of these dishes, the video leverages this cultural divide to create a shocking yet compelling narrative. Furthermore, this act of consuming and portraying such foods can be seen as a form of cultural assertion, similar to how the consumption of specific foods in Japan (whale meat) or South Korea (dog meat) justifies their traditions and acts as a form of resistance to perceived Western cultural imperialism (Ranta & Ichijo, 2022) [6]. Consequently, when a food's identity is shaped by being seen as different, it not only sets one group apart from another but also becomes a powerful way for cultures to assert themselves in a globalized world. This, thereby, turns mealtime choices into statements about identity and power.

Additionally, the study's analysis supports other research regarding reviewers' roles and identity in online contexts. Reviewers perform their role primarily by offering a critical evaluation of various subjects with the fundamental goal of persuading and influencing viewers' choices (Kathpalia, 2021) [28]. This involves actively expressing their attitude, stance, viewpoint, or feelings towards the item being reviewed. In this study, by employing multimodal resources, the reviewers have effectively constructed the representation of Vietnamese street food as an exotic culinary experience. Beyond evaluation, reviewers also use their platform for identity construction, thus potentially building "a personal brand" (Jaakkola, 2020, p. 386) [14]. A professional media personality, an adventurous food explorer and an expressive reviewer are explicit identities of reviewers that have been discovered in the most-watched review video about Vietnamese street food on YouTube. Ultimately, reviewers serve critical functions in informing, persuading, and shaping public opinion about various products and cultural phenomena.

3. Conclusion

The study contributes to a deeper understanding of how multimodal resources work together to represent cultural identity and traditions through food media. By focusing on the identity concept, the analysis shows how Vietnamese street food is constructed as an exotic culinary experience, largely based on the portrayal of dishes like the water bug extract and sandworm pancakes as in this particular food review video. This portrayal not only highlights unique aspects of Vietnamese food culture but also allows reviewers to build their personal brand as adventurous and influential figures in the digital food landscape. This interplay eventually shapes audience perceptions of both the cuisine and the content creators themselves.

Furthermore, the findings of the study underscore its academic and practical contribution. This study provides a novel multimodal analytical framework that integrates established SFL models with verbal, visual and paralinguistic theories to analyze food media. By demonstrating how these diverse semiotic resources orchestrate to construct meaning, it offers a more comprehensive approach than studies that focus on single modes. The findings also offer insights for content creators, particularly food reviewers and travel influencers. By understanding how the orchestration of visuals, verbal language, and paralanguage can build an authentic narrative, creators can design content that is more engaging.

While this study provides valuable insights, certain limitations should be acknowledged. The analysis focuses on specific modes (verbal, paralanguage, and visual) while omitting others, such as textual metafunctions or soundscapes, which could offer additional perspectives on multimodal meaning construction. Furthermore, while the dataset is limited to one video, broader analyses could explore more diverse representations of Vietnamese street food. Investigating more data also helps to avoid lengthy descriptions; instead, comparisons across videos can be made. Future research could expand on this study by examining other food cultures or incorporating omitted modes to provide a more holistic view of multimodal interaction in food studies.

REFERENCES

[1] Tam Anh, (2025). *TasteAtlas ranks Vietnamese street foods among the best in Southeast Asia*. Retrieved from https://e.vnexpress.net/photo/food-recipes/tasteatlas-ranks-vietnamese-street-foods-among-the-best-in-southeast-asia-4856205.html

[2] Nguyen, Q. (2019). *Netflix series on Asian street food focuses on Saigon*. Retrieved from https://e.vnexpress.net/news/travel/food-recipes/netflix-series-on-asian-street-food-focuses-on-saigon-3908602.html

[3] Cohen M, (2024). Food Identity. *Journal of Food Law & Policy, 20*(2). https://doi.org/10.54119/jflp.vmbu3563

[4] Stano, S. (2023). Food, Ideology and Critical Semiotics. *Lexia*, 41-42: 307-322.

[5] Anderson EN, (2014). *Everyone Eats*: *Understanding Food and Culture*. New York: NYU Press. https://muse.jhu.edu/book/28733.

[6] Ranta R & Ichijo A, (2022). *Food, National Identity, and Nationalism:* From Everyday to Global Politics. Springer.

[7] Jewitt C, Bezemer J & O'Halloran K, (2025). *Introducing multimodality, 2nd edition*. London and New York: Routledge.

[8] Van Leeuwen T, (2022). *Multimodality and Identity* (1st ed.). Routledge. https://doi.org/10.4324/9781003186625

[9] Bucholtz M & Hall K, (2004). Language and identity. In A. Duranti (Ed.), *A Companion to Linguistic Anthropology* (pp. 369-394). Blackwell.

[10] Reddy G & van Dam RM, (2020). Food, culture, and identity in multicultural societies: Insights from Singapore. *Appetite*, 149, p.104633

[11] İkizoğlu D & Gordon C, (2022). Vegetables as a Chore": Constructing and Problematizing a "Picky Eater" Identity Online. In *Identity and Ideology in Digital Food Discourse: Social Media Interactions across Cultural Contexts*, edited by Alla Tovares and Cynthia Gordon, pp. 13-32. London: Bloomsbury Academic

[12] Wang Y, (2024). The Impact of Product Review Videos on Electronics Consumers. Published thesis.

[13] Xu P, Chen L & Santhanam R, (2015). Will video be the next generation of e-commerce product reviews? Presentation format and the role of product type. *Decision Support Systems*, 73, 85–96. doi:10.1016/j.dss.2015.03.001

[14] Jaakkola M, (2020). Useful Creativity: Vernacular Reviewing on the Video-Sharing Platform Vimeo, *Culture Unbound*, Volume 12, Issue 2, 2020: 373–392. Published by Linköping University Electronic Press: http://www.cultureunbound.ep.liu.se

[15] Cheng S, (2024). A review of interpersonal metafunction studies in systemic functional linguistics (2012–2022). Journal of World Languages, 10(3), 623-667. https://doi.org/10.1515/jwl-2023-0026

[16] Sahelices-Pinto C, Lanero-Carrizo A, Vázquez-Burguete JL & Gutierrez-Rodriguez P, (2017). Ewom and 2.0 Opinion Leaders in the Food Context: A Study with a Sample of Spanish Food-Related Weblogs. Journal of Food Products Marketing, 24(3), 328–347. doi:10.1080/10454446.2017.1266561

[17] Bi NC, (2018). Product review videos on YouTube as eWOM. In L. Ha (Ed.), *The audience and business of YouTube and online videos* (pp. 59–72). Rowman & Littlefield

[18] Briliana V, Ruswidiono W & Deitiana T, (2020). Do Millennials Believe in Food Vlogger Reviews? A Study of Food Vlogs as a Source of Information. *Journal of Management and Marketing Review*, 5(3) 170 – 178

[19] Halliday MAK & Matthiessen CMI, (2004). An Introduction to Functional Grammar. London: Routledge.

[20] Liang WJ & Lim FV, (2024). Representing youth as vulnerable social media users: a social semiotic analysis of the promotional materials from The Social Dilemma. *Semiotica* 2024 (256):153-174.

[21] Kress G & van Leeuwen T, (2021). Reading images: The grammar of visual design, 3rd edition. London: Routledge.

[22] Ngo T, Hood S, Martin JR, Painter C, Smith B & Zappavigna M, (2022). Modelling paralanguage from the perspective of systemic functional semiotics: Theory and application. Bloomsbury.

[23] Martin JR & White PR, (2005). *The language of evaluation: Appraisal in English*. New York: Palgrave Macmillan.

[24] Zhang D, (2009). On A Synthetic Theoretical Framework for Multimodal Discourse Analysis [J]. *Foreign Languages in China*, 6 (01): 24-30. DOI:CNKI:SUN: ZGWE.0.2009-01-010.

[25] Zhang F, (2022). Meaning Construction of Multimodal Synergy in Documentary Discourse: Taking The Lockdown: One Month in Wuhan as an Example. *International Journal of Linguistics, Literature and Translation*, 5(6): 52-60, doi: 10.32996/ijllt.2022.5.6.7

[26] Lu Y, (2024). Multimodal Discourse Analysis of the Promotional Film Countdown: Beginning of Spring for the Opening Ceremony of Beijing Winter Olympics. *International Journal of Language, Literature and Culture,* 4(4), 17–29. https://doi.org/10.22161/ijllc.4.4.3

[27] Avieli N, (2011). Making Sense of Vietnamese Cuisine. *Education About Asia*, 16(3), 42-45.

[28] Kathpalia SS, (2021). *Persuasive Genres: Old and New Media (1st ed*.). Routledge. https://doi.org/10.4324/9780429243721

Appendix