

CHAPTER II

REVIEW OF LITERATURE

A. Review Literature

1. Discourse Analysis in Media

Discourse Analysis (DA) is an interdisciplinary approach that examines language as a form of social practice. Discourse is not merely a collection of words or sentences, but a means through which meaning, identity, and social relations are constructed. Through discourse, social values, ideologies, and power relations are produced and circulated within society (Gee, 2025). Language, therefore, functions not only as a tool for communication but also as a medium through which social reality is shaped and negotiated. This perspective is strongly influenced by the work of Norman Fairclough, who emphasizes that discourse both reflects and constructs social structures, particularly in relation to power and ideology (Quyen, 2024).

In media studies, DA and MDA are widely applied to analyze media texts such as news, advertisements, films, and digital content. Media discourse is considered powerful because it does not simply reflect reality, but actively constructs it (Đorđević, 2022). Through processes such as framing, representation, and selection of linguistic features, media texts can influence how audiences perceive events, individuals, and social issues. For instance, the choice of words,

metaphors, and narrative structures can subtly shape public opinion and reinforce dominant ideologies.

However, traditional discourse analysis has often focused primarily on verbal language, which limits its ability to fully explain meaning-making processes in contemporary media that rely heavily on visual and auditory elements (Nielsen et al., 2022). In response to this limitation, scholars have developed multimodal approaches that consider multiple semiotic resources simultaneously. According to Kress & van Leeuwen (2006), meaning is constructed through the interaction of various modes such as image, sound, gesture, and spatial design, each of which contributes to the overall interpretation of a text.

Furthermore, media texts in the digital era are increasingly multimodal, integrating visual imagery, soundtracks, editing techniques, and written language into a unified communicative whole. This complexity requires an analytical framework that goes beyond linguistic analysis alone. Multimodal Discourse Analysis (MDA) enables researchers to examine how these different modes work together to produce meaning, construct identities, and position audiences within particular ideological frameworks. Therefore, combining DA and MDA provides a more comprehensive understanding of how media discourse operates, especially in audiovisual texts such as film trailers, music videos, and social media content.

2. Multimodal Discourse Analysis of Audiovisual Media

Multimodal Discourse Analysis (MDA) has developed as an extension of Discourse Analysis (DA), emerging in response to the increasing dominance of visual and audiovisual communication in contemporary society. MDA recognizes that meaning is not constructed solely through verbal language, but through the interaction of multiple semiotic modes such as language, images, color, layout, sound, music, movement, and spatial arrangement (Kress & van Leeuwen, 2006). In audiovisual media, these modes operate simultaneously and complement one another, making analyses that focus only on verbal text insufficient for understanding contemporary communication practices that are highly visual and digital. Kress & van Leeuwen (2006) argue that each mode has its own “grammar” or system of meaning-making, and that the combination of these modes is structured and strategic rather than random. Therefore, MDA examines how the selection, organization, and interaction of these modes are used to construct representation, social relations, and meaning within a multimodal text.

From a multimodal perspective, each semiotic mode possesses its own communicative potential and conventions. According to Kashaka (2025), visual elements such as framing, composition, and color interact directly with the way viewers interpret images, while auditory elements such as soundtracks, sound effects, and tempo

rhythm function as guides to viewers' emotions and strengthen the narrative flow of the story. This strategic selection aligns with the concept of “mode choice”, which emphasizes that communicators deliberately choose particular modes to achieve specific communicative goals. As argued by Samaras (2025), content creators explicitly design and integrate text, imagery, and audio to exploit platform affordances and successfully resonate with their target audience.

According to Milenković (2025), one of the key contributions of Kress & van Leeuwen (2006) is the concept of three visual metafunctions, adapted from Halliday's metafunctions in Systemic Functional Linguistics. First, the representational metafunction concerns how reality, events, and participants are depicted through visual elements, such as types of processes (action, reaction), objects, and settings. Second, the interactive metafunction explains how the relationship between the image and the viewer is constructed through gaze, social distance (close-up, medium shot, long shot), camera angle, and emotional expression. Third, the compositional metafunction relates to how elements within the frame are arranged to create coherence and emphasis of meaning, for example through information value (left/right, top/bottom), salience (color, size, contrast), and framing. This framework is highly relevant for analyzing audiovisual texts such as advertisements, films, music videos, and teasers, as these

forms rely on complex visual elements to guide audience interpretation (Jewitt et al., 2025).

In addition to visual elements, the aural dimension also plays a central role in multimodal meaning-making. Sound, including music, rhythm, tempo, timbre, and even silence, functions as a semiotic resource that conveys emotional, social, and ideological meanings (Forte, 2023). For instance, soft and melancholic music may create a sense of intimacy or sadness, while fast tempo and strong beats can evoke excitement, tension, or anticipation. According to van Leeuwen (2026), sound parameters like pitch, volume, and tempo are not merely physical attributes, but rather carry meanings shaped by socio-cultural contexts. In brief audiovisual texts, these auditory elements are designed strategically to build atmosphere, represent modal identity, and instantly trigger emotional engagement from the audience.

Gestural elements and bodily movements are also essential components in multimodal analysis, particularly in audiovisual texts featuring human participants such as advertisements, music videos, and K-Pop teasers. Facial expressions, gaze direction, posture, and movement can communicate emotions, interpersonal relationships, and power dynamics (Norris & Geenen, 2026). In the entertainment industry, elements such as choreography, body orientation, and performance style are not merely aesthetic choices, but also serve as communicative strategies that construct the artist's identity and image

(Belgrimet & Rabab'ah, 2021). These gestural elements may reinforce or even contradict verbal and visual messages, making it necessary to analyze the interaction between modes.

Furthermore, spatial arrangement and composition within the frame play a significant role in shaping meaning. The placement of elements, use of empty space, depth (foreground and background), and camera movement (pan, zoom, tracking shot) influence how viewers process and interpret visual information (Hanmakyugh, 2023). For example, a centrally positioned subject with strong lighting may indicate importance, while the use of empty space may evoke feelings of isolation or contemplation (Tseng et al., 2024). These spatial strategies contribute to the overall coherence and emotional impact of audiovisual texts.

A key concept in MDA is intermodality, which refers to the interaction and integration of multiple semiotic modes in the process of meaning-making. Meaning does not emerge from individual modes in isolation, but from the relationships between them, as modes can reinforce, complement, or even contradict one another. In audiovisual media, particularly in formats such as K-Pop teasers, meaning is often constructed through the dynamic interplay between visual symbolism, sound design, gesture, and editing techniques. Therefore, MDA provides a comprehensive and systematic framework for analyzing

how audiovisual texts orchestrate multiple semiotic resources to produce layered meanings and influence audience interpretation.

3. Teaser as Multimodal Promotional Texts

A teaser is a short form of promotional content designed to generate curiosity, anticipation, and speculation among audiences before the release of a complete product such as a film, television series, album, or music video (Revanthy et al., 2023). Unlike trailers, which typically provide more detailed previews of storyline or content, teasers are deliberately brief—often lasting only a few seconds to less than a minute—and are characterized by ambiguity and incompleteness (Trehan & Maan, 2012). This strategic incompleteness is not a limitation but a defining feature, as it encourages viewers to infer missing information, form interpretations, and actively seek further details, thereby increasing audience engagement.

From a multimodal perspective, teasers can be understood as compact, high-density texts that rely on the interaction of visual, aural, and sometimes linguistic elements to construct meaning. Visual components such as color palettes, lighting, camera angles, framing, and editing rhythms are carefully orchestrated to convey mood, identity, and thematic direction rather than explicit narrative (Kashaka, 2025). For instance, dark tones and low-key lighting may evoke mystery or tension, while bright and saturated colors may signal energy and optimism (Kakar & Muzafar, 2025). According to Sarriá

(2015), camera techniques play a crucial role in constructing meaning; for instance, low-angle shots can create a sense of power and dominance. On the other hand, the use of close-up shots can establish intimacy and allow the viewer to delve deeper into the character's inner emotional turmoil (Li, 2023).

Aural elements are equally significant in shaping the meaning and emotional impact of teasers. Background music, sound effects, voice-overs, and even silence are strategically selected to align with the intended atmosphere and audience response (Kurki, 2025). Distinct musical hooks, rhythmic build-ups, and sudden silences can create tension, anticipation, or emotional release within a very limited duration. In many cases, teasers minimize or entirely omit spoken language, relying instead on non-verbal sounds, environmental audio, or musical cues as primary meaning-making resources. This emphasizes the role of sensory and affective engagement over explicit verbal explanation.

Teasers can also be conceptualized as narrative fragments. Rather than presenting a complete and linear storyline, they offer partial glimpses of characters, settings, and symbolic elements that function as interpretive clues. This aligns with the concept of “mode choice” by Kress & van Leeuwen (2006), in which producers deliberately select and combine semiotic modes to achieve specific communicative goals. In teaser production, decisions regarding what to

reveal visually (such as symbols, costumes, or settings) and what to suggest aurally (such as musical motifs or sound effects) are carefully balanced to construct a recognizable concept while maintaining ambiguity. Effective teasers therefore operate within a tension between clarity and mystery, providing enough information to attract attention while withholding enough to stimulate curiosity and speculation.

In the digital era, teasers are embedded within broader transmedia and social media ecosystems. Audiences no longer consume teasers passively; instead, they actively engage by replaying, analyzing, sharing, and discussing them across platforms (Yin, 2024). This participatory culture transforms teasers into objects of collective interpretation and significantly enhances their promotional impact. As a result, teasers function not only as advertisements but also as catalysts for audience interaction, speculation, and emotional investment.

B. Review of Relevant Research

This section reviews previous studies relevant to the present research in order to position the study within existing academic discussions and to identify research gaps that justify the current investigation. The reviewed studies are grouped into three main areas: (1) discourse analysis in media, (2) multimodal discourse analysis of audiovisual texts, and (3) studies related to teasers and K-Pop promotion.

First, several studies have applied Discourse Analysis (DA) to media texts to uncover how meaning, identity, and social values are constructed through language use. Yuliana et al. (2024), for example, analyzed K-Pop song lyrics to reveal semantic meanings and social values reflected in the texts, demonstrating how lyrics function as discursive tools to express identity and ideology. Similarly, Manninen (2021) examined identity narratives in K-Pop lyrics, focusing on verbal elements and thematic construction. While these studies provide valuable insights into discourse and identity formation, their analytical focus remains largely limited to linguistic aspects, with minimal attention to visual and aural modes that are central to contemporary media texts.

Second, a number of scholars have emphasized the importance of Multimodal Discourse Analysis (MDA) in understanding meaning-making beyond language. Foundational works by Kress & van Leeuwen (2006) provide the theoretical basis for analyzing visual grammar and its interaction with other semiotic modes. Subsequent studies have extended this framework to various audiovisual contexts, demonstrating how modes such as image, gesture, gaze, spatial arrangement, and sound interact to produce meaning (Norris, 2004). Hasyim & Arafah (2023) further argue that meaning in digital media emerges from the integration of multiple semiotic resources operating simultaneously. These studies support the view that audiovisual texts must be understood as complex multimodal ensembles rather than as simple carriers of verbal messages. However,

most of these studies focus on general audiovisual media and do not specifically address short-form promotional texts such as teasers.

Third, research on teasers and trailers has often been conducted from marketing, cinematic, or audience-response perspectives. One study examined how teaser trailers influenced audience expectations and consumer behavior, focusing on attention and recall. Johnston et al. (2016) discussed trailers as cinematic and aesthetic texts, while Revanthy et al. (2023) highlighted the role of teasers in building hype and shaping brand perception. Although these studies acknowledge the persuasive power of teasers, they tend to emphasize effectiveness and audience impact rather than providing detailed multimodal analysis of how meaning is constructed through visual, aural, and textual elements.

In the context of K-Pop, a growing body of research has explored marketing strategies, globalization, and audience engagement. Tanudjaja (2024) analyzed K-Pop promotional strategies and their effectiveness in building global fandoms, while Lee (2023) highlighted the systematic use of digital platforms and pre-release content in shaping fan interaction. Hargitai et al. (2023) also emphasized the role of social media in disseminating Korean Wave content. These studies demonstrate how K-Pop agencies strategically manage promotional materials, including teasers, to maximize visibility and engagement. However, they generally approach K-Pop promotion from a strategic or macro perspective, rather than from a detailed multimodal analytical lens.

Despite these contributions, there remains a significant gap in research that integrates Multimodal Discourse Analysis with a specific focus on K-Pop teasers. Most existing studies either analyze lyrics using discourse approaches or examine promotional strategies without exploring how meaning is constructed at the semiotic level. The multimodal construction of meaning in K-Pop teasers—through visual symbolism, sound design, spatial composition, and minimal textual cues—remains underexplored. In particular, short-form promotional content such as “Coming Soon” teasers has received little scholarly attention, despite its strategic importance in shaping audience anticipation and engagement. Furthermore, the interaction between visual and aural modes in teaser formats, including the role of music, sound effects, and silence, has not been extensively examined. Studies that explicitly analyze ultra-short promotional videos in K-Pop are especially scarce, indicating the need for more focused multimodal investigations.

Within this research landscape, the “Coming Soon” teaser of Rosé of Blackpink remains largely unstudied, despite her global influence and the significant public attention surrounding her releases. Existing research tends to focus more on music videos, fandom, or representation, rather than on teaser content as distinct multimodal texts. This indicates a clear research gap and provides a strong rationale for the present study. Therefore, this research aims to address this gap by applying a Multimodal Discourse Analysis (MDA) framework, primarily based on the visual

grammar theory proposed by Gunther Kress and Theo van Leeuwen (2006) which is further supported by the social semiotic approach to sound by Theo van Leeuwen (2026) and the Multimodal Interaction Analysis of gesture by Sirgid Norris (2004). Through this integrated framework, the study seeks to contribute both to the theoretical development of MDA in short promotional formats and to the empirical understanding of how K-Pop teasers construct meaning, shape audience perception, and generate anticipation in contemporary digital culture.

C. Theoretical Framework

This study is grounded in Multimodal Discourse Analysis (MDA) as its primary theoretical framework. MDA is employed to examine how meaning is systematically constructed through the interaction of multiple semiotic modes, including visual imagery, color schemes, sound, music, written text, gestures, facial expressions, and other audiovisual elements Kress & van Leeuwen (2006). This framework is particularly relevant in contemporary media contexts, where communication is no longer dominated by verbal language but is increasingly shaped by complex combinations of sensory modes.

At the core of MDA is the assumption that communication is inherently multimodal. Meaning is produced through the co-deployment of different semiotic resources, each with its own communicative affordances and constraints Kress & van Leeuwen (2006). These modes do not function independently but interact dynamically to form coherent and

layered meanings. As a result, meaning is not neutral or accidental but is strategically constructed through the selection and orchestration of modes to guide audience interpretation and emotional response (Kassa, 2024).

A central component of this framework is Kress and van Leeuwen's theory of visual grammar, which adapts Halliday's Systemic Functional Linguistics into three metafunctions: representational, interactive, and compositional meanings Kress & van Leeuwen (2006). The representational metafunction focuses on how people, objects, and actions are depicted, including the distinction between actional processes (doing, happening) and conceptual processes (being, symbolizing). The interactive metafunction concerns the relationship between the image and the viewer, realized through gaze (demand vs offer), social distance (close-up, medium shot, long shot), and camera angle (high, low, eye-level). The compositional metafunction examines how elements are arranged within the frame, including information value (given/new, ideal/real), salience (size, color contrast, focus), and framing (Umair et al., 2022). These metafunctions provide a systematic approach to analyzing how visual elements construct meaning and guide audience perception.

In addition to visual grammar, this study incorporates perspectives on sound and music as essential semiotic modes. Sound, including music, rhythm, tempo, timbre, and silence, functions as a powerful resource for constructing emotion, atmosphere, and identity (van Leeuwen, 2026). In short-form media such as teasers, sound plays a crucial role in shaping

audience engagement by creating tension, anticipation, or emotional resonance within a limited duration (Dashkewytch, 2021). The interaction between sound and visual elements, including synchronization between musical cues and visual transitions, is therefore analyzed to understand how meaning is reinforced and intensified (Pauwels, 2012).

Gestural elements are also central to this framework. Facial expressions, body posture, gaze, and movement are treated as meaningful semiotic resources that convey emotions, interpersonal relations, and identity (Norris, 2004). In performance-oriented media such as K-Pop teasers, gestures are highly stylized and contribute significantly to the construction of the artist's persona (Necula, 2017). These embodied expressions interact with visual framing and sound to produce nuanced and layered meanings.

Another important concept guiding this study is the deliberate selection and combination of semiotic modes to achieve specific communicative purposes, as emphasized by Constantinou (2005), which refers to the deliberate selection and combination of semiotic modes to achieve specific communicative purposes. In the context of a "Coming Soon" teaser, mode choice is reflected in decisions regarding what information to reveal or conceal, how visual and aural elements are prioritized, and how minimal textual cues are used to guide interpretation. These choices contribute to the creation of an intentionally incomplete yet evocative text that stimulates audience curiosity and anticipation.

Building upon this theoretical foundation, the analysis in this study is organized around five key semiotic modes, which correspond directly to the research problems:

a. Linguistic Mode

Linguistic mode encompasses spoken and written language, while written language also employs typography, layout, spacing, punctuation, and other graphic resources to construct meaning (Wildfeuer et al., 2019). Language functions as a framing device that guides interpretation and highlights particular aspects of a message, thereby shaping how audiences understand an issue (Flusberg et al., 2024).

b. Visual Mode

This encompasses imagery, color usage, lighting, framing, and symbolic representation Kress & van Leeuwen (2006). Visual elements play a dominant role in conveying mood, atmosphere, and conceptual meaning.

c. Spatial Mode

This refers to the layout and composition of elements within the frame, including positioning, salience, and visual hierarchy (Kress & van Leeuwen (2006). Spatial organization directs viewer attention and structures meaning.

d. Aural Mode

This includes music, sound effects, rhythm, volume, and silence (van Leeuwen, 2026). Aural elements function as affective resources that shape emotional tone, pacing, and anticipation.

e. Gestural Mode

This involves facial expressions, body movement, posture, and gaze (Belgrimet & Rabab'ah, 2021). Gestural elements communicate emotion, attitude, and interpersonal meaning.

These five modes do not operate in isolation but through intermodal interaction, where meaning emerges from their combined and coordinated use. Visual symbolism may be reinforced by sound, gestural expressions may gain intensity through framing and lighting, and linguistic minimalism may be supported by strong audiovisual storytelling. Through this interaction, the teaser functions as a unified multimodal discourse rather than a collection of separate elements.

Finally, this framework is situated within the broader context of K-Pop promotional practices. In this study, K-Pop teasers are treated as multimodal promotional texts that combine various semiotic resources to communicate artistic concepts, shape the artist's public image, and attract audience interest before the official release of new content. However, rather than focusing on audience reception, this study emphasizes the internal semiotic mechanisms through which meaning is constructed. In summary, this theoretical framework integrates Halliday's Systemic

Functional Linguistics (SFL) for the linguistic analysis, Kress & van Leeuwen's (2006) visual grammar, van Leeuwen's (2026) theory of sound, and Norris's (2004) framework on gesture as a semiotic resource to provide a comprehensive multimodal analysis of the data. This integrated approach enables a comprehensive analysis of how Rosé's "Coming Soon" teaser constructs layered meanings, shapes artistic identity, and evokes emotional responses through the interaction of linguistic, visual, spatial, aural, and gestural modes.