

CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter discusses about some theories related to the study. They are discourse analysis, systemic functional linguistics, multimodal discourse analysis, emotions in media and movies, the movie *Inside Out*2, and previous study.

A. Discourse Analysis

Discourse analysis is a field of linguistic study that investigates the use of language within social, cultural, and ideological frameworks, going beyond the limitations of grammatical structure or sentences alone. Gee (2015) highlights that discourse analysis allows for a deep understanding of how language shapes identity, values, and social interactions. Van Dijk (2018) complements this view by stating that discourse analysis encompasses written, spoken, and multimodal texts, as they are all interrelated in the process of meaning formation within a specific context.

Discourse analysis has developed and come to encompass various methods, such as Conversation Analysis, which describes everyday interactions to understand conversation patterns (Sidnell, 2016), and Critical Discourse Analysis, which emphasizes the relationship between language, social domination, ideology, and power (Wodak & Meyer, 2016). Additionally, the approach that is relevant to functional linguistics is Systemic Functional Linguistics (SFL), which focuses on the role of language in a social context. SFL views language as a system of meaning options that are applied to achieve social goals (Halliday, 2013), and it became the foundation for the development

of Multimodal Discourse Analysis (MDA) by Kress and van Leeuwen (2006). This SFL approach is crucial for explaining how verbal modes, such as dialogue in films, reflect and construct the emotional experiences of the characters.

B. Multimodal Discourse Analysis (MDA)

Multimodal Discourse Analysis (MDA), developed by Kress and van Leeuwen, is a logical extension of Systemic Functional Linguistics (SFL) by Halliday, specifically designed to examine how meaning is constructed through the interaction of various semiotic modes beyond spoken language. Kress and van Leeuwen (2006) built this framework by relying on the principles of social semiotics, in which each mode has its own grammar and is chosen by the discourse producer to convey a specific message. Their main work, *Reading Images: The Grammar of Visual Design*, often abbreviated as GVD (Grammar of Visual Design), serves as a specific model within MDA for examining visual meaning.

According to the MDA framework, the overall emotional message of a movie is formed by combining various semiotic elements. While traditional media studies often prioritize text-driven narratives, MDA recognizes that meaning is non-linear and multidimensional. Although both verbal and visual modes have a unique capacity for meaning formation, their real power lies in how they interact and support each other (Kress & van Leeuwen, 2006). This study breaks down the two main modes found in movies to systematically analyze how emotions are constructed in the movie *Inside Out 2* (2024). This is done using a specific theoretical framework, which is described below:

1. Verbal Mode

Verbal mode includes spoken and written language used to convey meaning, such as conversation, narration, and written text. This mode is one of the main mediums of communication because it allows ideas to be conveyed explicitly. Tan (2021) asserts that verbal mode plays an important role in constructing narratives and conveying clear emotions to the audience. In the context of multimodality, verbal mode often combines with other modes, so that the meaning of a message does not only depend on words, but also on accompanying visual, audio, and gestural support. Verbal mode is analyzed in this study to reveal how language linguistically reflects and constructs the emotional representations of the two main characters, Joy and Anxiety.

2. Visual Mode

Visual mode includes various aspects that can be seen, such as color, lighting, shot size, and camera angle. This mode is very significant because humans are essentially visual creatures who respond quickly to images. Song and Kwon (2024) found that in virtual reality settings, the appearance of virtual characters (including facial expressions and overall look) together with head movements significantly impact how emotions are perceived by viewers. The visual mode also allows for the implicit conveyance of meaning, for example through symbols or visual metaphors that are difficult to grasp through verbal language alone. This mode is analyzed to see how visual composition represents and builds emotional interaction.

In conclusion, the MDA approach is crucial in this study because emotions in animated media are never singular, but rather the result of systematic synergy between text and imagery. By adopting MDA as the primary framework, this study provides a comprehensive explanation of how the purposeful interaction between verbal mode and highly dominant visual mode provides the narrative meaning. Because the visual mode carries the primary semiotic weight in expressing emotion in this film, a more specialized and rigorous tool is required shifting the focus from the general multimodal umbrella to Kress and van Leeuwen's specific visual grammar (Grammar of Visual Design).

C. The Grammar of Visual Design (GVD)

While, MDA serves as the umbrella framework in this study, given that emotions in animated films arise from the simultaneous “orchestration” of verbal and non-verbal elements, the GVD functions as the primary operational instrument to analyze the visual mode. GVD provides a visual grammar to analyze how images communicate meaning through structural elements such as colour, framing, and vectors. This framework systematically adapts and applies Halliday's Three Metafunctions of SFL.

1. Representational Metafunction

Kress and van Leeuwen theory views representational metafunction as how people, places, and objects are represented in images and text. In this study, it is used to describe how visual mode or images represent emotional expressions in the movie. In analyzing representational, it involves two process they are narrative process and conceptual process.

1.1 Narrative Process

Kress and van Leeuwen (2006) state that narrative process refers to an ongoing action or event where the elements within an image are connected by a vector. In short, narrative process is when participants are doing an action and connected by a vector. A vector is an imaginary or visible line indicating the direction of movement or interaction that create a visual story that places participants in an active role as agents or targets of an action. Furthermore, it can be formed by bodies, body parts, and objects appearing in the image. Narrative process is divided into two types called actional and reactional.

a. Actional Process

Actional process involves physical actions that occurs when a participant acts as an ‘actor’ who then emits a ‘vector’ that can be visualized by moving bodies, body parts, objects or a line of action directed toward specific target called a ‘goal’. So, the actional process involves an actor, vector, and goal. Actional process actions that have a ‘goal’ is a transactional. Actions in a transactional process have an ‘actor’ and a ‘goal’ directed at anyone or anything (Kress & van Leeuwen, 2006:63). For example, a little girl (the actor) gives a piece of bread (the vector) to a beggar (the goal). The girl acts as the actor performing the action, while the beggar serves as the goal receiving the action, then the vector is depicted by the direction of the girl’s hand movement, illustrating the connection between their action. Look at Figure 2.1.



Figure 2. 1 Transactional Actional Process

However, if the action in the actional process lacks a clear target but still demonstrate dynamic physical activity, it is called non-transactional. For example, a ballet dancer (the actor) is dancing (the vector) demonstrates physical action by moving her body, yet the action lacks a specific target or goal. Look at Figure 2.2.



Figure 2. 2 Non-transactional Actional Process

b. Reactional Process

Reactional process happens when a vector is formed by the eyeline or the participant's glance that creates a reaction rather than action. 'Reactor' is the participant who does the look, while 'phenomenon' is the object or participant who receive the gaze. The relationship between participants is not physical but rather mental or emotional, which the way the reactor gazes can take the form of anger, affection, or admiration, that is used to convey the type of

internal reaction the participants are currently experiencing. Just like the actional process, reactional process can be either transactional where the ‘reactor’ and ‘phenomenon’ exist, or non-transactional where only the ‘reactor’ appears (Kress & van Leeuwen, 2006:67). Look at figure 2.3 as the example of a transactional reactional process which a man (the reactor) is gazing at (the vector) a woman (the phenomenon). Afterwards, look at figure 2.4 as the example of non-transactional reactional process which a boy (the reactor) is daydreaming (the vector).



Figure 2. 3 Transactional Reactional Process

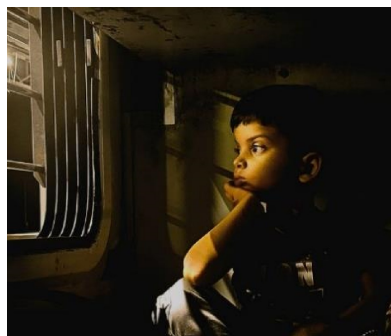


Figure 2. 4 Non-transactional Reactional Process

1.2 Conceptual Process

Unlike narrative process, conceptual process does not have a vector and serve to represent participants in terms of their class, structure, or more permanent meanings, making the images appear

more static and descriptive. Kress and van Leeuwen (2006:79) state that this process aims to define the fundamental ‘who’ or ‘what’ of the participant and it is often used in scientific illustrations, diagrams, or formal portraits intended to establish identity or classification rather than to narrate a specific event. Conceptual meaning can be categorized into classificational process, analytical process, and symbolic process.

a. Classificational Process

Classificational process shows the relationship between the participants in the images. It is a process that arranges participants symmetrically or in a specific order to indicate that they belong to the same class or category. This relationship offers one participant to be a superordinate and the others to be a subordinate that creates a taxonomy relationship among elements to provide an understanding of how members are grouped within a larger category (Kress & van Leeuwen 2006:79). For example, look at Figure 2.5 where ‘5 groups of vertebrates’ serve as the superordinate, it means the primary category or top-level in the classification. While, ‘fish, amphibians, reptiles, birds, and mammals’ serve as the subordinate, it means the members or the elements included within the superordinate. Then, taxonomy is this entire classification system or hierarchical structure of vertebrates that connects the superordinate with subordinate.



Figure 2. 5 Classficcational Process

b. Analytical Process

Analytical process presents a relationship between the ‘parts’ and the ‘whole’. It involves one participant as the main focus of the image (the carrier) and the other participant as the part of the carrier (possessive attributes). In other words, the carrier is the whole and the possessive attributes are the part in the images. The examples are human anatomy, machine diagrams, and maps. They are aim to break down the components of a subject so that the viewer can understand how these small parts work together to form a functional whole (Kress & van Leeuwen, 2006:87). Look at Figure 2.6.

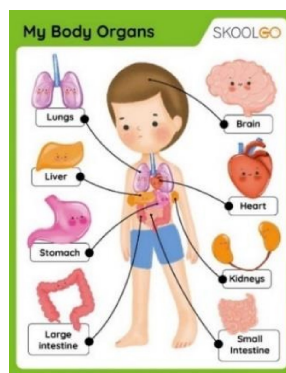


Figure 2. 6 Analytical Process

c. Symbolic Process

Symbolic process focuses on finding out the symbolic attributes in the images and then reveal the meaning behind it. Symbolic process can be attributive if the participant's identity is conferred by a specific symbolic object placed nearby creating a particular identity or a value. For example, Look at Figure 2.7 a woman wearing a white robe and a crown which symbolize purity, holiness and honor. These attributes function as symbols that impart meaning to the participant.



Figure 2. 7 Attributive Symbolic Process

Symbolic process can be suggestive if the meaning emerges from the overall atmosphere. Look at Figure 2.8, a man standing atop of a mountain with sunlight behind him creating an atmosphere that may imply victory, hope, or freedom. Thus, the symbolic meaning emerges implicitly from the overall visual composition, not from specific attributes attached to the participant. According to Kress and van Leeuwen (2006:105), in the symbolic process, the physical details of the participants are often set aside in order to highlight an

abstract quality or deep identity that one wishes to convey, such as power, sanctity, or dignity.



Figure 2. 8 Suggestive Symbolic Process

2. Interactive Metafunction

Interactive metafunction determines how an image establishes social and emotional connections between the participants and the viewers. Producers visually encode social meanings into an image through the participant's expressions, participant's distance from the camera, and the viewer's point of view (Kress & van Leeuwen, 2006).

2.1 The Gaze

The gaze happens when a participant looks towards the viewer creating a connection established between the participant and the viewer. Not only through the gaze, but it can be accompanied by gesture at the same time. The gaze is the most powerful tool for visual interaction, where the direction of the participant's gaze determines the type of social interaction taking place and the role the viewer must assume in that visual communication exchange.

a. Demand

Demand happens when the participant in the image looks directly at the viewer, compelling the viewer to enter into the active social relationship with the participant. This action creates an imaginary relationship where the viewer is compelled to acknowledge the participant's presence and respond to the emotions or authority emanating from that gaze, as if the participant were speaking directly and demanding direct engagement from the audience (Kress & van Leeuwen, 2006:118).

Look at Figure 2.9 as the example, where the woman stares directly at the viewer and seeks validation regarding her opinion that the man next to her is annoying. The woman's facial expression also helps create emotional closeness as well as humour, as if the viewer is being invited to agree with her assessment of the other character.



Figure 2. 9 Demand

b. Offer

Offer happens when participants look away and do not make eye contact with the viewer, thereby positioning the viewer as an objective outside observer who is 'offered' the opportunity to view the information or subject within the image without having to become emotionally or

socially involved. It is called an offer because it presents participants to the viewer impersonally as items of information, objects of contemplation, as if they were samples in a display case (Kress & van Leeuwen, 2006:119). For instance, in an image of a student reading a book intently without looking at the camera, the viewer is merely offered information about the activity without any expectation of involvement. Look at Figure 2.10.



Figure 2. 10 Offer

2.2 Social Distance

Social distance is determined by how close or far the participant is positioned within the frame. Social distance conveys the closeness of the relationship between the participant and the viewer through the shooting distance. For example, ‘a close-up shot’ stimulates a close or intimate personal relationship. Typically, only the face or the upper body is visible causing the distance creates a personal impression, as if the viewer has a close relationship with the participant. Meanwhile, ‘a medium shot’ represents a more neutral or general social relationship. Medium shot shows the participant’s body from the waist up, creating the impression of ordinary social interaction. In the other hand, ‘a long

shot' places the participant at a public distance, creating an impression of formality and objectivity. It usually shows the entire body or even the surroundings.



Figure 2. 11 a Close-Up Shot



Figure 2. 12 a Medium Shot



Figure 2. 13 a Long Shot

2.3 Angle

Angle is used to distribute power and emotional engagement between the viewer and the participant, giving the viewer control over how they should feel and assess the participant's social position within

the available visual space. It can also be referred as the point of view, which is divided into horizontal and vertical.

a. Horizontal

Horizontal angle relates to the degree of involvement. The frontal angle indicates that what is in the image is part of our world, as if we were involved in the story, while the oblique angle indicates detachment or disconnection, signifying that the participant is part of the other world from us (Kress & van Leeuwen, 2006:134). Look at Figure 2.14, where all participants face the camera directly. Their faces and bodies are mostly visible from the front, and some of them even make eye contact with the viewer. This indicates the use of frontal angle. In other words, it feels like the participants invite the viewer involved to be part of their group.



Figure 2. 14 Frontal Angle

Now, Look at Figure 2.15, the participants are seen from the side and not facing at the camera directly. Their bodies are positioned diagonally, and their focus is on the activity lifting the teacher not at the viewer. This is known as an oblique angle. The

viewer is positioned as an outside observer, simply witnessing this moment of togetherness without being directly involved.



Figure 2. 15 Oblique Angle

b. Vertical

Vertical angle represents the distribution of power between the participants and the viewer, where a high angle gives the viewer the position of superiority and power over the participants in the image. In contrast, a low angle makes the participant appear more powerful and superior. Finally, an eye-level angle indicates a relationship of equality that is not dominated by either party (Kress & van Leeuwen, 2006:140). Look at these figures as the examples.



Figure 2. 16 High Angle



Figure 2. 17 Low Angle



Figure 2. 18 Eye-level Angle

2.4 Modality

Modality refers to the degree of truth or credibility of an image, which is determined by the extent to which the image represents reality accordance to a certain standard. Kress and van Leeuwen (2006:155) explain that the highest level of naturalistic modality is achieved when the use of colour, texture, and lighting closely resembles what we see with naked eye. However, in scientific context simpler images such as diagrams and tables are actually considered to possess a higher modality because it focuses on function rather than appearance.

Table 2. 1 Modality Marker

Modality marker	Maximum scale value	Minimum scale value
Colour saturation	Full of colour saturation	Black and white
Colour differentiation	Fully diverse range of colours	Monochrome
Colour modulation	Fully modulated colour (many different shades of colour)	plain, unmodulated colour
Contextualization	Detailed background	Absence of background
Representation	Maximum representation of every detail of participant	Minimal detail of participant (e.g., soft focus)

Depth	Deep perspective, strong convergence of vertical line is shown (e.g., fish-eye)	Absence of depth
Illumination	Full representation of light and shade	Absence of light and shade
Brightness	Maximum number degrees of brightness	Two brightness values of the same colour (e.g., dark grey & lighter grey)

3. Compositional Metafunction

Compositional metafunction describes how the representational and interactive metafunctions are combined into a whole coherent. Composition determines how the viewer's attention is directed and how different elements are related to one another to convey a coherent message (Kress & van Leeuwen, 2006:175). The three principles of composition are information value, salience, and framing.

3.1 Information Value

Information value is determined by the placement of elements within specific zones in a composition such as left and right, top and bottom, centre and margin. The left and right zones distinguish between established or known information (given) and information that is new or provocative (new). Then, the top placement often represents the ideal, abstract, or a promise (ideal), while the bottom placement is filled with information that is more practical, factual (real). Finally, the centre zone indicates placement of the main information in the centre of the image while the margin zone is where the supporting information at (Kress & van Leeuwen, 2006:179).



Figure 2. 19 Information Value

3.2 Salience

Salience is a technique used to make certain elements in an image more visually striking to the viewer than others. Kress and van Leeuwen (2006:201) state that salience does not depend on a single factor, but rather on a complex interplay of size, focus, contrast, and placement which collectively draw the viewer's eye to those elements first.

a. Foreground and Background

Foreground consists of elements placed in the front image, which naturally have a higher level of prominence because they appear larger and clearer detail making them the primary focus of the message. Meanwhile, background is placed behind the foreground and serves as the context that provides spatial depth (Kress & van Leeuwen, 2006: 202). Look at Figure 2.20 as an example.



Figure 2. 20 Foreground and Background

b. Font

Font is a set of letters in particular type, size, and colour. The choice of font, size, and weight plays a crucial role in providing visual emphasis to specific text messages where larger or bolder text has a higher level of prominence.

3.3 Framing

Framing refers to the presence or absence of borders or empty space that connect or separate elements within an image, where a strong frame indicates that these elements constitute distinct units of information, while the absence of a frame implies that they form an integrated whole.

D. Emotions in Media and Movies

Emotion is central to audience engagement with movie. Movies are designed to elicit emotional responses that determine narrative meaning. This study adopts the Basic Emotions framework proposed by Paul Ekman (1992). Ekman asserts that several basic emotions such as Happiness, Sadness, Anger, Fear, Disgust, and Surprise are biologically innate, universally recognized, and characterized by distinctive facial expressions and physiological changes. In the context of animated film analysis, this framework is crucial because it provides a clear and scientifically supported basis for identifying core emotional states, which animation artists then amplify through stylized visual features, such as distinct character shapes, lightning, and predominantly bright or symbolic colour palettes to convey emotions.

This study also incorporates the framework of Fredrickson (2001) to further classify these emotions. While Ekman identifies what basic emotions

exist, Fredrickson (2001) categorizes them into positive emotions and negative emotions. The positive emotions such as happiness, optimism, and cheerfulness manifest visually through open, expansive expressions. While, negative emotions such as anxiety, fear, panic, and worry producing constricting and focused behavioral patterns. Both frameworks work complimentary, Ekman's framework guides the identification of specific emotions in each scene, while Fredrickson's framework classifies whether those emotions belong to positive or negative states.

E. The Movie Inside Out 2

Inside Out 2 belongs to the animated coming-of-age genre. This genre is chosen because it directly reflects the real-life emotional growth and social struggles of teenagers. According to an ELT study by Valencia (2025), animated movies like Inside Out 2 use visual signs, colors, and facial expressions that effectively help students decode complex and abstract emotional concepts. Inside Out 2 (2024) continues the story of Riley, who is now entering adolescence, a phase of life full of changes and emotional challenges. The movie introduces new emotions such as Anxiety, Envy, Embarrassment, and Ennui, which are present in the "control centre," replacing the dominance of the five main emotions, namely Joy, Sadness, Anger, Fear, and Disgust. The presence of these new emotions depicts a more complex psychological dynamic, in which Riley must navigate the anxiety, envy, and social vulnerability typical of puberty. The addition of these emotions not only enriches the narrative, but also makes Riley's psychological representation more

realistic and relevant to the audience. With the expansion from five to nine emotions, the movie presents a deeper and more diverse multimodal study.

The movie consciously uses character design, color, music, dialogue, and gestures to explicitly display emotions. O'Halloran (2021) states that multimodal analysis can reveal the interaction between modes in conveying meaning. Therefore, *Inside Out 2* is very suitable as an object of study to understand the representation of emotions in animated movies.

F. Previous Study

The application of Multimodal Discourse Analysis (MDA) in film studies has received strong support from various empirical studies. For example, Julianingsi's (2023) study examined the film “*Cinderella* (2021)” and found that text (dialogue) and images interacted to form interpersonal and interactive meanings, using the frameworks of Kress and van Leeuwen and Halliday. Similar findings also emerged in Asis et al.'s (2021) analysis of the film “*Maleficent: Mistress of Evil*”, where verbal and visual elements such as facial expressions, body movements, and costume colours were found to be interrelated in conveying ideological messages about female power. Furthermore, the most directly relevant research is the study by Siddique et al. (2022) in their thesis “*Sadness and Joy in Inside Out: A Multimodal Analysis*”. This study investigates how emotions are represented in the film *Inside Out* (2015) through the multimodal framework of Kress and van Leeuwen, and finds that emotions are conveyed through verbal dialogue as well as visual representations such as color and character design.

Although these studies reinforce the validity of the MDA framework and demonstrate the multimodal nature of emotions, most previous studies still have limitations that can be addressed in this study. Specifically, the study by Siddique et al. (2022), although related to the same film object, it is focused on two characters (Joy and Sadness) and focused generally on the equal status of the text and visuals regarding primary emotions, without deeply exploring how an animated narrative can heavily shift its semiotic weight onto the visual mode to depict more complex, conflicting adolescent emotions. In general, these existing studies tend to evaluate text and image with identical parameters, failing to provide a detail, rigorous analysis of colour semiotics using Grammar of Visual Design (GVD) to explain situations where visual presentation clearly takes precedence especially in the movie *Inside Out* where the use of colour is very strong to express emotions.

Therefore, this study aims to fill this gap through a comprehensive analysis of the film *Inside Out 2* (2024), which introduces new emotional characters with greater complexity. By emphasizing the two dominant main characters, Joy and Anxiety, this study applies a theoretical framework of Kress & van Leeuwen to provide more integrated insights into the construction of emotional expression through the interaction of these two semiotic modes.