INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia

IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

Harmonizing Images, Texts, and Emoticons to Promote Indonesian Events on Social Media

Siti Aisyiyah[*, 1], Titik Ismailia [2]

[1, 2] *Language, Communication, and Tourism*
Politeknik Negeri Jember, Indonesia

**Corresponding email: aisyiyah@polije.ac.id*

Abstract

Social media is one of the most effective promotional channels in the digital era, as it can reach audiences widely. Social media facilitates features that enable its users to provide information clearly and attractively. One of the social media platforms, Instagram, is used to promote Indonesian events by the Indonesian Ministry of Tourism (Kementerian Pariwisata RI). This research applied multimodal discourse analysis to investigate the use of images, texts, and emoticons to promote Indonesian events on Instagram. Employing visual grammar theory, Appraisal framework, and intersemiotic complementarity, this paper analysed how the three modes collaborated to make meaning in English promotion. Data were images and English captions taken from two official Instagram accounts of the Indonesian Ministry of Tourism posted in April to June 2026. The images or visuals were the primary mode of Instagram posts. The posts were completed with captions which consisted of verbal texts and emoticons. Findings of the research highlighted that the images mostly contained cultural dances as unique events in Indonesia. Most images were delivered in narrative representation. The verbal texts of captions dominantly conveyed valuation and inclination to influence the audience's perception and evoke their feeling about the events. Additionally, emoticons were used to strengthen emotion and develop more personal communication with audiences. The harmonious collaboration of eye-catching images, persuasive verbal texts, and supportive emoticons developed effective messages for promoting Indonesian events on social media.

Keywords: *multimodal communication, images, texts, emoticons, Indonesian events, social media*

|| *Received: 25/05/2026* || *Accepted: 27/07/2026* || *Published: 29/07/2026*

1. Introduction

Advances in information and communication technology have transformed the way individuals and organizations interact with the public. Social media has become one of the most widely used platforms for marketing communication because it can reach a broad audience. Various social media platforms such as Instagram, Facebook, and

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia

IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

TikTok are not only used to share information but also serve as effective promotional tools for companies, business owners, and organizations.

The social media, particularly Instagram, is utilized by Ministry of Tourism of the Republic of Indonesia to promote one of its prime programs, that is Indonesia events. The Indonesia events were posted by official accounts of Ministry of Tourism, those are kemenpar.ri and @eventbyindonesia. The posts were completed with captions in Indonesian and English. BPS-Statistics Indonesia notes the number of foreign visitors in Indonesia from January to April 2026 was 4,68 millions which was the highest number of foreign visitors since 2020. It reflects the magnitude of social media influence on people's behaviour which makes the role of social media important as a media campaign.

Promotion through social media is widely used to attract visitors to come to Indonesia. The urgency of using social media as a media promotion particularly of Indonesia events is based on the large number of social media users in Indonesia. Based on data released by goodstats.id at the end of 2025, the number of Instagram users in Indonesia reach 82,4% of national internet users. It places Indonesia on the top ten of the highest number of Instagram users in the world. In this paper, Instagram was chosen as the source of data based on consideration that the characteristics of Instagram that can display visual and verbal text, in accordance with the promotional needs of Indonesia events that require good visualization supported with verbal information. Zappavigna (2016) investigates that Instagram is a social photography application designed on smartphones so that social media users can convey visual and textual messages, and readers can interact with these messages. She also states that Instagram provides users with a number of semiotic modes, such as visual and written modalities, and overcomes semiotic constraints on image dimensions and text length in captions.

The application, which was released in 2010, is still quite in demand today and has a lot of influence on the cyber culture of tourists. For example, in the decision-making process for traveling by the public, social media is a reference source for determining Indonesia events to visit. Ismarizal & Kusumah (2023) find that Instagram content can encourage travellers from the internet generation (iGen). They highlight that interesting Instagram content in terms of colour, captions, and community comments can be one of the factors that attract tourists to visit Indonesia events. Arefieva et al. (2021) states that with its visual characteristics and wide popularity, Instagram is the right platform to display images, attractions, and experiences offered by various tourism destinations. Gorji et al. (2023) add that when travellers do not have enough information about a particular destination, the photos shared on photo sharing platforms have an important role to play in influencing their decisions and choices.

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia

IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

Elements commonly used in Instagram include visuals such as images or photos, and captions containing text and emoticons. Visuals in the form of images, illustrations, or videos play a crucial role in capturing users' attention. In addition to visual elements, captions are also a key component of digital marketing communication. Captions are verbal texts that is often accompanied by emoticons. The texts serve as visual supplement that provides information, builds a narrative, and guides the audience to take specific actions. Meanwhile, emoticons function as visual symbols that can reinforce emotional expression, enhance the appeal of a message, and create more personal and easily understandable communication.

Although various studies have discussed the effectiveness of Instagram as a promotional tool, there remains a need to understand the specific roles of each element including images, verbal text, and emoticons. This study aims to analyse how images, verbal texts, and emoticons interplay to construct promotional language on Instagram. The results of this study are expected to contribute to the development of more effective digital promotion and serve as a reference for practitioners and academics in understanding the factors that influence promotional language on social media.

2. Multimodality in Digital Communication

Multimodality in digital communication deals with various modes of communication such as verbal language, images, symbols, and sounds. In social media promotions, messages are typically conveyed through a combination of visuals and captions that usually consists of verbal texts and emoticons that work simultaneously to produce a stronger meaning. Each element serves a distinct communicative function. Visuals serve to attract attention and convey information quickly, verbal texts provide context and explanation, while emoticons reinforce emotional expression. The interaction of these three elements creates a richer communication experience than text alone.

2.1 Visuals in social media promotional content

A visual is a primary feature in social media communication as it can attract users' attention in the first sight. From a multimodal perspective, visuals can function as standalone sources of meaning without being complemented by verbal texts. To understand the meaning of visuals, Kress & Leeuwen (2021) develop a visual grammar theory which is an extension of three meta functions in systemic functional linguistics (SFL). The SFL theory places language as part of social phenomena related to the social context in language use. In SFL, there are three main functions in language: the ideational function, the interpersonal function, and the textual function (Halliday & Matthiessen, 2014).

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia

IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

In the framework of visual grammar, Kress & Leeuwen (2021) adopt the three functions in systemic functional linguistics, namely ideational, interpersonal, and textual, into three visual meta functions consisting of representational, interactive, and compositional. Representational meaning explores the representation of objects displayed in the real world. The meaning can be understood through two processes: the narrative process and the conceptual process. The narrative process includes the action process, reaction process, verbal process, and mental process; while the conceptual process is related to the relational process and the existential process. Narrative process focuses on dynamic visual elements, unlike conceptual representations, which tend to be static and classificatory.

Interactive meaning is associated with the social relationships between actors and the evaluative orientations received by participants from one another and the displayed objects. Elements that can be used to understand interactive meaning include contact, social distance, and attitude. Contact relates to the direct or indirect gaze of participants in the image. Social distance is seen through the proximity or distance of the photo-taking, including: close shot, medium shot, and long shot. Attitude in visual grammar can be seen from the perspective of subjectivity and objectivity. Subjectivity is distinguished based on the horizontal and vertical angles of the photograph.

Compositional meaning refers to the layout of the multimodal discourse as a whole. The elements that constitute compositional meaning are information value, framing, and salience. Information value relates to the placement of certain information, such as new and old information. Framing answers the question of whether there are tools that connect these discourse elements or not. Salience refers to the different levels of elements for the reader, such as background and foreground.

2.2 Captions as promotional communication strategy

Captions are texts accompanying visuals in social media posts. The function of captions is not merely to provide additional information but also to build a narrative, convey a call to action, and create a connection with the audience. Besides verbal texts, captions are frequently completed with emoticons. Captions uploaded alongside images are important aspects of promotion. Iswandi et al., (2021) explain that language plays a crucial role in promotional media.

Captions on social media carry meaning related to social relationships, interactions, and expression of feelings. It means language of captions can be used to reflect author's attitude. An attitude is an expression that involves thoughts and feelings. The system of attitudes conveyed through language was proposed by Martin & White (2005) in Appraisal framework. They divide the attitude system into three dimensions: affect,

DOI: https://doi.org/10.25047/ijossh.v1i1.7365
IJOSSH is published by
Politeknik Negeri Jember, Indonesia
IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

judgment, and appreciation. Affect relates to a person's feelings or emotions; judgment relates to the behaviour of others; and appreciation relates to something, an object, or a phenomenon. Fang & Zou (2023) define appraisal as a term in systemic functional linguistics that describes a system of semiotic resources used to express opinions or attitudes. When expressing an attitude, a writer or speaker tends to use evaluative language. It relates to the writer's or speaker's attitude toward the material being conveyed. Appraisal framework explains how the expression of attitudes is used to make readers or listeners accept or agree with the attitudes and values being conveyed (Komninos, 2023).

The Appraisal framework is a development of the meta functions of the Systemic Functional Linguistics. Initially, the term appraisal was related to various applications of evaluative language. Subsequently, Martin & White (2005) develop the Appraisal framework that can be used for more comprehensive analysis related to evaluative language. The framework consists of three systems: Attitude, Engagement, and Graduation. Attitude is divided into three scopes, which include polarity, realization, and category. The polarity of the attitude system is divided into two poles, namely positive and negative. The realization of the attitude system can be seen explicitly (inscribed) and implicitly (invoked). The categories of the attitude system consist of three categories: Affect, Judgment, and Appreciation. Affect is related to the dimension of emotions. It is categorized into fear and desire (dis/inclination); unhappiness and happiness (un/happiness); insecurity and security (in/security); dissatisfaction and satisfaction (dis/satisfaction). Judgment refers to ethics, evaluating behaviour consisting of social esteem and social sanction. Social esteem is classified into normality (normality, how special), capacity (capacity, how capable), and tenacity (tenacity, how dependable). Social sanction is grouped into veracity (truthfulness, how honest) and propriety (decency, how ethical). Appreciation refers to the aesthetic aspect, evaluating texts/processes, natural phenomena. Appreciation is divided into three categories: reaction (reaction to something), composition, and valuation (valuation, how innovative, authentic, etc.). Next, the reaction category has two sub-categories: impact and quality. Similarly, the composition category includes two sub-categories: balance and complexity.

2.3 Emoticons in social media communication and promotion

Social media platforms are increasingly providing users with new visual resources for realizing both initiating moves and responses, so making it possible to replace words and sentences with emoticons or emoji, stickers, and so on (Jovanovic & Van Leeuwen, 2018). Emoticons are visual symbols used to express emotions, attitudes, or specific meanings in digital communication. In the current era, the employment of emoticons on social media is frequently regarded as a component that bridges the gaps in spoken mode

IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365
IJOSSH is published by
Politeknik Negeri Jember, Indonesia
IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

and facilitates the effective and engaging expression of emotions and ideas—a phenomenon known as popular language (Danesi, 2017).

On Instagram, emoticons are usually used alongside verbal modes in captions. Most emoticons are placed at the end of sentences as a substitute for punctuation (Surbakti et al., 2025). Although the number of emoticons in captions is not as many as the verbal mode, the symbol or emoticon is one of the unique characteristics in communication on social media. The expression of attitudes in verbal and emoticon modes is used to create a positive impression or image, arouse the reader's emotions, and encourage action to visit. A positive image is built by expressing the aesthetics of the destination, while emotions are touched by conveying positive feelings. Emoticons or emoji are symbols, icons, or characters that depict facial expressions, attitudes, or emotions.

2.4 Interplay of images, verbal texts, and emoticons

In multimodal discourse, various semiotic modes are used to construct meaning both individually and collectively. These semiotic modes interact in various forms. Barthes (1977) finds that verbal can add/extend the meaning of images and vice versa. It can also be said that verbal elaborates on images and vice versa. For Barthes, the meaning of images in particular, and other semiotic modes, are always connected to and dependent on the meaning of verbal texts. Meanwhile, for Kress & Leeuwen (2021) the visual component of the text, the image, is a message that is independently composed and arranged, related to the verbal text but not dependent on it.

Chen (2010) investigates the message conveyed visually can contradict (disclaim) the message conveyed verbally. However, these verbal and visual texts appear in the same context simultaneously. This relationship is also found by Bednarek & Caple (2015) who studied photojournalism in print and online newspapers. Martinec & Salway (2005) offer a system to examine the relationship between visual and verbal elements in a multimodal text. Their approach is based on the status and logical-semantics relationship between visual and verbal elements.

The approach differs from the system developed by Liu & O'Halloran (2009), which is more focused on the discourse relations between verbal and visual texts in their analysis of cohesive devices between verbal and visual texts. It includes logical relations that may arise in visual and linguistic messages, such as comparative, additive, consequence, contingency, and time. Comparative occurs when the components of language and images have the same experiential meaning. Additive emerges when one semiotic component adds new information to another, and then both messages are connected with verbal and visual messages. Additive conveys related but different messages. Consequential occurs when one semiotic message causes or defines another

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia

IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

message. Conversely, contingency arises when the cause has the potential to determine a possibility, but there is no definite effect. Time occurs when different procedural stages are represented both verbally and visually, so that the relationship between text and image can be described with the logic of time.

Royce (2013) defines intersemiotic complementarity in visual and verbal modes based on three meta functions: ideational, interpersonal, and compositional. He elaborated on the ideational meaning, where the relationship between the two modes includes repetition, synonymy, antonymy, hyponymy, meronymy, and collocation. Repetition occurs if there is a repetition of experiential meaning in the visual and verbal modes. Synonymy arises if the visual and verbal modes have the same experiential meaning. Antonymy occurs when the visual and verbal modes have opposing experiential meanings. Hyponymy arises when there is a classification of a general class and its subclasses. Meronymy occurs in relationships that are part-whole relations. Collocation occurs when a general category of collocation appears in words that tend to emerge in various subject areas.

In interpersonal meaning, Royce (2013) distinguishes relations between visual mode and verbal mode into three relations, which include: reinforcement of address, attitudinal congruence, and attitudinal dissonance. Reinforcement of address occurs when identical meanings are found in both visual and verbal modes (an identical form of address). Attitudinal congruence, which describes harmony or suitability, appears when the meanings in visual and verbal modes are similar (a similar kind of attitude). Attitudinal dissonance, which involves dissonance, incongruity, or mismatch, arises when the meanings in visual and verbal modes are ironic or contrasting. The third relation refers to compositional meanings. In compositional meaning, the relations of visual and verbal modes are analysed based on the coherence of layout or composition. It consists of information value in the image, salience in the image, degree of framing of elements in the image, inter-visual synonymy, and reading path.

3. Methods

This study employs a qualitative approach that allows researchers to collect data in a natural setting (Creswell & Creswell, 2018). In this case, the researcher utilized Instagram posts containing images and captions without any manipulation. The images and captions used as data in this study align with the characteristics of qualitative research, which relies on data in the form of text and images (Creswell & Creswell, 2018). The data sources were the @kemenpar.ri and @eventbyindonesia accounts, which were the official accounts of the Ministry of Tourism of the Republic of Indonesia. Both accounts have blue verification badges. The data was Instagram posts in the form of

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia
IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

images completed with captions comprising verbal text and emoticons. The data was selected based on English-language captions intended to provide information and attract the interest of foreign tourists to visit Indonesia. Each post was saved as a screenshot of the photo and caption.

The data was analysed using multimodal discourse analysis combining visual analysis on images, caption analysis on verbal texts and emoticons, and the interplay of those modes. Images were analysed based on representational meaning of visual grammar theory by Kress & Leeuwen (2021). Captions including verbal texts and emoticons were analysed based on the attitude system of the appraisal framework by Martin & White (2005), encompassing the dimensions of affect, judgement, and appreciation. Then, the visuals, verbal texts, and emoticons were analysed to find the interplay of the different modes in constructing the promotion of Indonesia events considering intersemiotic complementarity by Royce (2013).

4. Findings and discussion

There were 33 selected posts in the form of images completed with English captions. They were captured from the official Instagram account of Ministry of Tourism of The Republic of Indonesia @kemenpar.ri and @eventbyindonesia. The images were posted from April to June 2026. Before that period, there was no specific post about Indonesia events accompanied with English captions. The posts provided multiple images featuring events and their schedule. The first image displayed the upcoming event, and next images provided the schedules of related events. The images which were analysed were the images of the events regarding their function to attract viewers at the first time. The images were classified into six categories including cultural dances, music performances, sports, hiking, artworks, and crowd/spectacle.

***Table 1**. Image Categories*

No	Category	Number
1	Cultural dances	8
2	Music performances	5
3	Sports	5
4	Hiking	4
5	Artworks	3
6	Crowd/Spectacle	7

Narrative representation images

Narrative representation focuses on dynamic interactions, identifying vectors (lines of action or sight) that tell a story rather than presenting a static state. Conceptual

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia
IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

representation deals with the essence, structure, or meaning of the subjects, rather than their actions. Most images used to promote Indonesia events were dynamic narrative images.



Image 1. Cultural Dances

Image 1 presents the masked traditional Balinese dancer positioned in the centre. The action lacks an explicit internal goal, so the narrative focus shifts entirely to the quality of the movement itself. The image narrates energy, expression, and the essence of dance, projecting this vitality directly outward to invite the viewer into the performance. There is no specific physical object or person inside the frame being targeted by this action. The eyes of the mask are directed straight ahead, breaking the frame to look directly toward the viewer. This creates a non-transactional reaction that transforms into a direct address. By looking straight at the audience, the character creates an immediate psychological connection. The narrative is no longer just a spectacle to be observed passively; it becomes a personal invitation, reinforcing the festival's tagline, "Where The World Comes To Move."

The text "World Dance Festival Indonesia 2026" and institutional logos "Wonderful Indonesia" serve as tools that clarify the purpose of the action, transforming a cultural performance into a concrete global event. Narratively, the strong physical vectors and direct eyeline engagement create an active, high-energy invitation that pulls the viewer into the experience. The traditional attire and architectural backdrop define the event's authentic identity, while the structural text establishes its grand, international scale. Ultimately, the image communicates that the World Dance Festival Indonesia 2026 is not just an event to watch, but a dynamic global celebration of culture that the viewer is are personally invited to experience.

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia

IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License



Image 2. Music Performance

Unlike the previous image featuring traditional dances, image 2 captures a contemporary musical performance, utilizing lighting, technology, and performance gestures to establish narrative processes. The narrative is centred around the experience of live performance. The combination of intense digital backdrops, sharp laser-like spotlights, and active singing postures narrates a high-energy, modern entertainment event meant to be projected outward toward an audience. The four modern musical performers on stage (singers/artists) are standing in a line under the spotlight. The performers are not directing their actions toward an object or person inside the scene. This creates a transactional reaction.

The silhouette of the crowd watching the performers bridges the gap between the viewer of the image and the stage. It narrates a collective experience of admiration, focus, and immersion, inviting the image's viewer to imagine being part of that captivated audience. Narratively, the sharp vectors of the spotlights and the transactional gaze of the audience silhouettes create an immersive story of a vibrant, live concert experience. The futuristic visuals, modern styling, and clean event details define the weekend picks as trendy, metropolitan, and globally influenced youth entertainment.

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia

IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

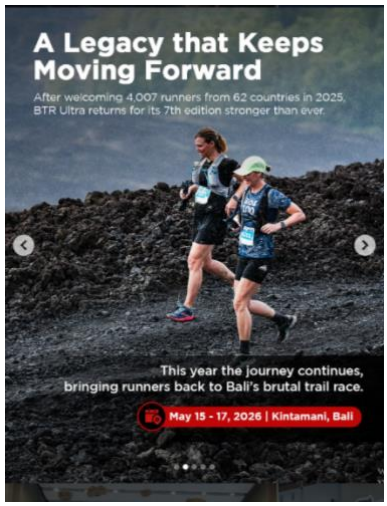


Image 3. Sports

Image 3 shifts from the previous stage performances to an outdoor athletic feat, using the real-world movement of human bodies across a harsh landscape to build its narrative. The narrative center completely on the process of endurance and determination. Because there is no visible finish line or final destination within the frame, the story emphasizes the act of running itself. The left-to-right movement directly mirrors the headline, "A Legacy that Keeps Moving Forward," literalizing the concept of progress.

Two female trail runners are running side by side. The runners are moving toward a space outside the right side of the frame. This represents a transactional reaction. Neither runner looks at the camera or the viewer; they are completely absorbed in their immediate environment. Their focused downward and forward gazes tell a narrative of intense concentration, awareness, and respect for the challenging path they are conquering.

The smiles on their faces, however, add a narrative layer of joy and camaraderie despite the physical strain. Narratively, the left-to-right directional vectors and the runners' focused gazes construct a story of active progression, intense focus, and shared endurance through a demanding landscape. The stark, jagged volcanic setting and the textual legacy details define *BTR Ultra* as a highly prestigious, uniquely challenging, and elite international trail race that tests human limits.

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia
IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

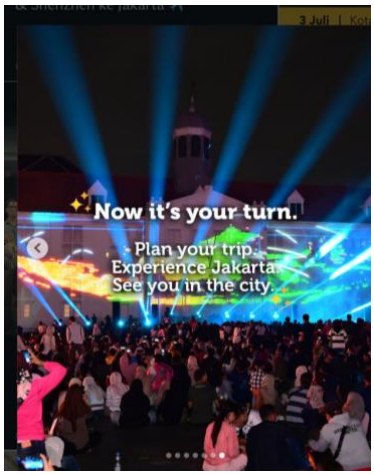


Image 4. Crowd/Spectacle

Image 4 shifts from individual human performers to a civic spectacle, focusing on the dynamic interplay between technology, architecture, and a collective audience. The narrative tells a story of an urban celebration and modern technological immersion. The radiating light beams and shifting colours transform a static historical space into an active, breathing event environment that projects its vibrant energy directly into the night. There are two actors including the stage lighting equipment / projection system (implicit) and the crowd in the foreground. The vectors of light radiate outwards into the sky and onto the building rather than targeting a specific person or object within the frame, making this a non-transactional action.

The large crowd is seated and standing in the foreground. The individuals in the crowd are positioned with their backs or sides to the viewer, all pointing their heads, gazes, and camera lenses forward and upward toward the illuminated building. This creates a clear transactional reaction. The collective gaze of the audience establishes the light show as a highly engaging phenomenon. Because the viewer of the image occupies a space right behind this crowd, we are narratively positioned as participants in the event, looking at the city's attraction alongside the local spectators.

The text overlay "Now it's your turn. Plan your trip. Experience Jakarta. See you in the city." acts as a direct narrative prompt, using the visual energy of the festival as a persuasive tool to inspire travel planning. The image functions as an analytical process that defines the identity of Jakarta as a tourism destination. Narratively, the radiating vectors of the searchlights combined with the transactional gaze of the captivated crowd construct an immersive story of city pride, shared entertainment, and public celebration. By overlaying futuristic light technology onto a historic monument, the image defines

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia
IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

Jakarta's identity as a vibrant, multifaceted metropolis where history and modern excitement intersect, ending with a direct personal call to action: "Now it's your turn."

Attitudinal message of captions

Attitude delivered through verbal language can be classified into three categories: *affect*, *judgement*, and *appreciation* (Martin & White, 2005). In this research, two categories were found in the captions including affect and appreciation. Affect consists of *inclination*, *happiness, security*, and *satisfaction.* Appreciation includes *reaction*, *complexity,* and *valuation.* Meanwhile, judgement was not found in the captions. It can be understood because the posts were aimed to promote events, not persons. Judgement is used to evaluate person's behaviours, not events. In the captions, attitude was used to evaluate Indonesia events including appreciation and affect. Appreciation refers to aesthetic aspects, while affect relates to feeling and emotion. Mostly, appreciation was delivered through *valuation,* and affect was conveyed through *inclination.*

Valuation is related to mental process, perspective to view something. It refers to the evaluation of aesthetics, texts, processes, or phenomena. In this research, valuation was used to give value to Indonesia events. The valuation was conveyed through verbal language such as: *where stories come to life, it's not just about distance; Bali's wildest trails, Jakarta just got a global shoutout; or From heritage streets to skyline views, from island sunsets to family-friendly fun.* Besides, the valuation was expressed throuh emoticons like *sparkles* (✨), *OK hand* (👌), or *heart* (♡)

Inclination is one of affect categories dealing with feeling or emotion. It refers to tendency, desire, or wish to do something. In the research, inclination deals with desire to watch or join the Indonesia events as the primary preferences. It was used to evoke feeling or willing to join or watch the events. The inclination were found in some expressions including: *Want to know more?; Swipe, save it, and don't leave the weekend on read;* or *Plan your trip. Feel the vibe. Enjoy Jakarta.*

Caption 1

From stage performances to street parades, this is where stories come to life ✨
World Dance Festival Indonesia 2026 brings together dancers from across the world, from traditional performances to modern expressions, all on one stage.
Want to know more? Head over to @wdfindonesia for details and how to experience it live 👀

Caption 1 was written to support image 1 which presents the masked traditional Balinese dancer. The first sentence of the caption provides valuation in expression **"where stories come to life ✨".** The text assigns high social and cultural value to the

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia

IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

event by characterizing it as a transformative space. The phrase implies that the festival is not just an ordinary exhibition of movement, but an impactful medium where narrative and culture gain vitality. The use of the sparkle emoticon (✨) intensifies this positive evaluation. Next expression *"brings together dancers from across the world... all on one stage."* also gives positive valuation. This highlights the scale, diversity, and unifying power of the event. By emphasizing the integration of global talent ("across the world") and contrasting genres ("traditional performances to modern expressions") onto a single platform ("all on one stage"), the text establishes the festival's high institutional worth and global prestige.

Inclination is delivered through expression *"Want to know more?"*. The writer directly interrogates the reader's state of curiosity and desire for information. It acts as a behavioral trigger, prompting the audience to acknowledge their own interest in the event. Then, expression *"Head over to @wdfindonesia for details..."* initiates call to action. This imperative clause channels the reader's sparked curiosity into a specific, intended direction. The text strategically shifts the reader from a passive consumer of information to an active seeker. Expression *"...how to experience it live* 👀" indicates desire for immersion. The phrase appeals to the reader's desire for real-life experiential engagement ("experience it live"). The addition of the eyes emoticon (👀) serves as a visual marker of anticipation, alertness, and keen interest, heightening the collective sense of excitement and fear of missing out (FOMO).

Caption 2
The weekend's calling… you picking up? 😁 📞
Because suddenly staying in doesn't sound as tempting anymore. So if your weekend needed a little push, we've got a few ideas waiting for you soo consider this a sign. ☺️↕
Swipe, save it, and don't leave the weekend on read. ☝

Caption 2 was written to support image 2 featuring the four modern musical performers on stage. The positive valuation can be found in expression *"So if your weekend needed a little push, we've got a few ideas waiting for you"*. The text assigns a high functional utility and worth to its own content ("a few ideas"). By presenting these ideas as the perfect "push" to rescue a boring weekend, the brand positions its recommendations as highly valuable, useful solutions for the reader's leisure time.

The expression of inclination is delivered through expression *"soo consider this a sign.* 🙂↕". It is a validation of desire. By framing the post as a "sign" accompanied by a nodding emoticon (🙂↕), the text seeks to validate and nudge the reader's latent inclination to go out. It gives the audience a playful justification to stop hesitating and make plans. Next expression is invited inclination *"Swipe, save it, and don't leave the*

INTERNATIONAL JOURNAL OF
STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia
IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

weekend on read. ☝". The text utilizes imperative verbs ("Swipe, save it") to channel the reader's impulse into digital actions. The clever modern slang "don't leave the weekend on read"—borrowing from messaging apps where ignoring someone is seen as a negative choice—directly targets the reader's sense of social obligation and inclination. It acts as an urgent, highly relatable prompt to not ignore the opportunities of the weekend.

Caption 3

What does it look like to run from one volcano to another? 👀
BTR Ultra is coming back, and it's not just about distance, it's about elevation, terrain, and the kind of determination that starts before sunrise.
See how runners take on Bali's wildest trails this year.
Running next year? You might want to save this!👌📌

Caption 3 was written to support image 3 showing two female trail runners are running side by side. The first valuation can be found in expression *"it's not just about distance, it's about elevation, terrain, and the kind of determination that starts before sunrise."*. The text assigns intense value to the event by breaking it down into rigorous, meaningful components. By stating it is "not just about distance" and emphasizing "elevation, terrain, and... determination," the writer evaluates the race as something far more profound than a standard marathon. It frames the event as a monumental test of human spirit and physical endurance, raising its institutional prestige and worth. Next valuation can be read in *"Bali's wildest trails"*. The use of the superlative "wildest" evaluates the natural setting of the race as untamed, challenging, and pristine. This adds high experiential value to the event, positioning it as an extraordinary, elite adventure rather than a conventional road run.

The expression od inclination is written in an invitation *"See how runners take on..."*. The use of the imperative verb "See" serves as a direct invitation, guiding the reader's triggered curiosity toward a specific action (watching or observing the event's media content). The inclination is continued by *"Running next year? You might want to save this!*👌📌". It is a **c**all to intention. The final conditional query ("Running next year?") directly targets the reader's personal athletic goals and future intentions. The recommendation "You might want to save this!" coupled with the OK hand and pushpin emoticons (👌📌) converts that potential desire into an immediate, practical action (saving the post), actively fostering a readiness to commit to the event in the future.

Caption 4

Jakarta just got a global shoutout 🌏✨

INTERNATIONAL JOURNAL OF STUDIES IN SOCIAL SCIENCES AND HUMANITIES
IJOSSH, Vol 3(1) 2026
DOI: https://doi.org/10.25047/ijossh.v1i1.7365

IJOSSH is published by
Politeknik Negeri Jember, Indonesia
IJOSSH is licenced under a Creative Commons Attribution-ShareAlike 4.0 International License

Ranked #53 in the World's Best Cities 2026 by Resonance Consultancy, Jakarta is officially one of the world's cities to watch. From heritage streets to skyline views, from island sunsets to family-friendly fun. This city has a rhythm all its own.
Already trending on Instagram and TikTok, now it's your turn to experience it in real life. Plan your trip. Feel the vibe. Enjoy Jakarta ♡

Caption 4 was written to support image 4 displaying the large crowd is seated and standing in the foreground. The valuation is expressed in *"Jakarta just got a global shoutout 🌏✨ / Ranked #53 in the World's Best Cities 2026 by Resonance Consultancy, Jakarta is officially one of the world's cities to watch."*. It is a high positive valuation. The text establishes a very high social worth for Jakarta by grounding its praise in external, authoritative validation ("Resonance Consultancy"). The phrases "global shoutout," "World's Best Cities," and "cities to watch" elevate the city's institutional prestige from a local destination to a globally recognized metropolis. The emojis (🌏✨) serve to intensify this celebratory valuation. Next valuation is *"From heritage streets to skyline views, from island sunsets to family-friendly fun. This city has a rhythm all its own."*. The writer values the composition of the city by highlighting its multifaceted appeal (combining history, modernity, nature, and family leisure). By declaring that the city "has a rhythm all its own," the text attributes a unique, authentic, and highly desirable aesthetic value to Jakarta. Another valuation is conveyed in *"Already trending on Instagram and TikTok..."*. This evaluates the city's current social currency. By framing Jakarta as "trending," the text assigns a high contemporary value to the city, making it culturally relevant and desirable to a digitally active audience.

The expression of provoked inclination is *"...now it's your turn to experience it in real life."* The phrase "now it's your turn" operates as an interpersonal bridge. It transforms the objective global trend into a subjective, personal invitation. It provokes a sense of readiness and longing in the reader to step out of the digital space ("trending on Instagram and TikTok") and into active participation ("experience it in real life"). Expression *"Plan your trip. Feel the vibe. Enjoy Jakarta ❤️"* is a call to action. The text closes with a succession of short, high-impact imperative clauses ("Plan," "Feel," "Enjoy"). These function as direct commands designed to channel the reader's sparked desire into immediate physical actions. The concluding heart emoticon (❤️) infuses the directive with warmth, cultivating a positive emotional inclination toward the city.

Interplay of images and captions

Interplay of images and captions was analysed based on intersemiotic complementarity by Royce (2013), particularly on interpersonal meaning. It categorizes visual-verbal intercomplementarity into three relations. Those are reinforcement of address, attitudinal congruence, and attitudinal dissonance. Reinforcement of address happens when there is

DOI: https://doi.org/10.25047/ijossh.v1i1.7365
IJOSSH is published by
Politeknik Negeri Jember, Indonesia
IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

identical meaning of visual and verbal modes. It is also known as an identical form of address. Attitudinal congruence reflects harmony, suitability, or conformity on the meaning of visual and verbal modes. It takes place when visual and verbal modes have similarity on attitude or a similar kind of attitude. Attitudinal dissonance contains discrepancy, mismatch, or non-conformity on the meaning of visual and verbal modes. It appears due to an opposite or ironic attitude of visual and verbal modes' meaning.

Image 1 presents the Balinese dancer for promoting an event the World Dance Festival Indonesia 2026. The image is reinforced by caption 1 functioning as a persuasive promotional tool by combining high valuation with targeted inclination. It first establishes the cultural and global significance of the *World Dance Festival Indonesia 2026* to build social value. Once the event's prestige is recognized, it immediately shifts to interpersonal engagement, utilizing direct queries, visual emojis, and imperative prompts to cultivate and direct the reader's inclination toward active participation and attendance.

Image 2 features music performance of a live concert. It is completed with caption 2 operating as a highly persuasive, casual lifestyle prompt by shifting the reader's valuation of their weekend plans. It systematically reduces the perceived value of staying home ("not tempting") while inflating the utility of the brand's suggestions ("a little push"). Once the value is established, it leverages conversational text metaphors ("picking up the call," "leaving on read") to strongly provoke the reader's inclination, nudging them from a state of hesitation into active social participation.

Image 3 promotes running events by showing two female trail runners. It is strengthened by caption 3 which functions as a highly compelling promotional prompt by seamlessly bridging extreme valuation with strategic inclination. It first validates the immense worth and physical prestige of the *BTR Ultra* by framing it around volcanic landscapes and raw determination. Once the event is established as an epic challenge, the text shifts to an interpersonal level, utilizing curiosity prompts ("What does it look like?") and a direct call to future intention ("Running next year?") to nudge the reader from an amazed observer into a committed future participant.

Image 4 shows the large crowd in the foreground. The image is assisted by caption 4 which functions as a highly effective tourism promotion by balancing institutional valuation with persuasive inclination. It first validates Jakarta's global prestige using high-profile indexing ("World's Best Cities 2026") and modern social media metrics ("trending") to maximize the city's perceived worth. Once the value is secured, it directly targets the reader's inclination, using conversational shifts ("now it's your turn") and active travel directives to push the audience from passive admiration to booking a trip.

DOI: https://doi.org/10.25047/ijossh.v1i1.7365
IJOSSH is published by
Politeknik Negeri Jember, Indonesia
IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

Among the posts of Indonesia events, images and captions are completed each other. There is no discrepancy on the meaning of images and the attitude of captions. The images functions as visualisation of the events which can attract the viewers' attention. Meanwhile, the captions do not only provide information about images, but also strengthen the message by expressing valuation to the events and evoking inclination of the viewers. In other words, there is harmonious meaning of images and captions containing verbal texts and emoticons. Adopting Royce's intersemiotic complementarity of interpersonal meaning, the interplay of images, verbal texts, and emoticons on the posts of Indonesia events is attitudinal congruence.

5. Conclusion

Images, verbal texts, and emoticons play a crucial role in conveying the meaning of promotional messages on social media. These three elements complement one another in creating message that is informative and attractive to encourage viewers' desire to watch or join the events. The images mostly feature cultural dances as the unique event and identity of Indonesia. They are presented in dynamic narrative representation. The images are the most influential element in capturing the attention of Instagram users. They also play a role in conveying information more effectively.

Captions, containing verbal texts and emoticons, are written to support the existence of images. The captions provide additional information while guiding the audience to take specific actions. The verbal texts of captions dominantly conveyed valuation and inclination to influence audience's perception and evoke their feeling about the events. Additionally, emoticons were used to strengthen emotion and develop more personal communication with audiences. It is crucial to tailor the use of the emoticons to the context and viewers' characteristics to avoid diminishing the effectiveness of the promotional message.

The harmonious collaboration of eye-catching images, persuasive verbal texts, and supportive emoticons develop effective messages for promotion. There is no doubt that the three elements play crucial role in the promotional language on social media. However, this research was limited to investigate representation meaning of visuals, attitudinal message of captions containing verbal texts and emoticons, as well as their interplay. There are more aspects of multimodality that are needed to be investigated further. Due to the limitation of this research, it is suggested to have more comprehensive study on multimodality to find the best feature of promotion on social media.

References

Arefieva, V., Egger, R., & Yu, J. (2021). A machine learning approach to cluster

IJOSSH is published by
Politeknik Negeri Jember, Indonesia

IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

destination image on Instagram. *Tourism Management*, *85*(March), 104318. https://doi.org/10.1016/j.tourman.2021.104318

Barthes, R. (1977). *Image Music Text* (S. Heath (ed.)). Fontana Press.

Bednarek, M., & Caple, H. (2015). News Discourse. *Brno Studies in English*, *41*(1), 213–216. https://doi.org/10.5817/bse2015-1-13

Beni Ismarizal, & Ahmad Hudaiby Galih Kusumah. (2023). The Instagram Effect on Tourist Destination Choices: Unveiling Key Attraction Elements. *Journal of Consumer Sciences*, *8*(2), 124–137. https://doi.org/10.29244/jcs.8.2.124-137

Chen, Y. (2010). Exploring dialogic engagement with readers in multimodal EFL textbooks in China. *Visual Communication*, *9*(4), 485–506. https://doi.org/10.1177/1470357210382186

Creswell, J. W., & Creswell, D. J. (2018). Research Design. Qualitative, Quantitative, and Mixed Methods Approaches. In *Sage Publications, Inc.* (Fifth). SAGE Publications, Inc. https://doi.org/10.4324/9780429469237-3

Danesi, M. (2017). *The Semiotics of Emoji: The rise of visual language in the age of the internet*. Bloomsbury Academic.

Fang, S., & Zou, L. (2023). Narratives of Women and Gender Relations in Chinese COVID-19 Frontline Reports in 2020. *International Journal of Environmental Research and Public Health*, *20*(5). https://doi.org/10.3390/ijerph20054359

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday ' s Introduction to Functional Grammar* (4th ed.). Routledge.

Iswandi, F., Sururia, L., & Fahmia, M. (2021). Pembentukan Sikap Penulis Melalui Teori Appraisal Dalam Situs Pariwisata 'Www.Indonesia.Travel' (Wonderful Indonesia) Versi Bahasa Indonesia. *Nusantara: Jurnal Ilmu Pengetahuan Sosial*, *8*(2), 2373–2389.

Jovanovic, D., & Van Leeuwen, T. (2018). Multimodal dialogue on social media. *Social Semiotics*, *28*(5), 683–699. https://doi.org/10.1080/10350330.2018.1504732

Komninos, N. (2023). Discourse Analysis of the 2022 Australian Tennis Open: A Multimodal Appraisal Perspective. *Hermes - Journal of Language and Communication in Business*, *63*, 83–98. https://doi.org/10.7146/hjlcb.vi63.140134

Kress, G., & Leeuwen, T. Van. (2021). *Reading Images; The Grammar of Visual Design* (3rd ed.). Routledge.

Liu, Y., & O'Halloran, K. L. (2009). Intersemiotic Texture: Analyzing cohesive devices between language and images. *Social Semiotics*, *19*(4), 367–388. https://doi.org/10.1080/10350330903361059

Martin, J. R., & White, P. R. R. (2005). The Language of Evaluation: The Appraisal Framework. *Lecture Notes in Computer Science*, 256.

Martinec, R., & Salway, A. (2005). A system for image–text relations in new (and old) media. *Visual Communication*, *4*(3), 337–371. https://doi.org/10.1177/1470357205055928

Royce, T. D. (2013). Intersemiotic complementarity: A framework for multimodal discourse analysis. In T. D. Royce & Bowcher (Eds.), *New Directions in the Analysis of Multimodal Discourse* (First, pp. 63–109). Routledge.

DOI: https://doi.org/10.25047/ijossh.v1i1.7365
IJOSSH is published by
Politeknik Negeri Jember, Indonesia
IJOSSH is licenced under a Creative Commons
Attribution-ShareAlike 4.0 International License

https://doi.org/10.4324/9780203357774

Siyamiyan Gorji, A., Almeida-García, F., & Mercadé Melé, P. (2023). Analysis of the projected image of tourism destinations on photographs: the case of Iran on Instagram. *Anatolia*, *34*(2), 144–162. https://doi.org/10.1080/13032917.2021.2001665

Surbakti, A. N., Lubis, A. D., Israra, R. F., & Wardani, R. (2025). Emojis as Modern Punctuation and Their Effect on Digital Communication. *Jurnal Pendidikan Tambusai*, *9*(1), 4583–4588.

Zappavigna, M. (2016). Social Media Photography: Construing subjectivity in Instagram images. *Visual Communication*, *15*(3), 271–292. https://doi.org/DOI 10.1177/1470357216643220