

Analyzing Visual Grammar In Language Pedagogy: A Case Of An Instagram ELT Account Kampung_Ingggris.Pare

Muntaha Muntaha¹, Umi Kartikawati², Muhammad Rashid Maulana Yusuf ³, and Solikha Nur Dzulfiani ⁴

^{1,3,4} UIN Raden Mas Said Surakarta, Indonesia

² Milligan Academia Indonesia, Indonesia

* E-mail: muntaha@staff.uinsaid.ac.id

Article history

Submitted: 2026/09/30; Revised: 2026/10/05; Accepted: 2026/10/07

Abstract

The rapid expansion of social media has transformed how English language teaching and learning are represented, promoted, and experienced beyond conventional classrooms. Instagram provides English language education providers with a multimodal space where linguistic, visual, spatial, and interpersonal resources construct pedagogical meanings. This study examines visual grammar in the Instagram content of Kampung Ingggris Mr. Mark, an English-language learning institution in Indonesia. Drawing on Kress and van Leeuwen's social semiotic theory, the study analyses representational, interactive, and compositional meanings through qualitative Multimodal Discourse Analysis of selected Instagram posts featuring English-learning activities, teachers, students, instructional messages, institutional identity, and promotional-pedagogical content. The analysis focuses on visual participants, narrative and conceptual processes, gaze, social distance, perspective, modality, and information value. Findings show that Instagram posts construct English learning as accessible, social, participatory, and visually engaging while simultaneously functioning as instructional and promotional communication. The study demonstrates that social media visuals can operate as pedagogical resources and that visual grammar provides a systematic framework for examining meaning-making in digital ELT environments.

Keywords

Instagram; Kampung Ingggris; language pedagogy; multimodality; visual grammar



© 2026 by the authors. This is an open-access publication under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY SA) license.

INTRODUCTION

The development of digital communication has substantially changed how language education is delivered, represented, and accessed (Jones & Hafner, 2012; Kress, 2010). English language learning is no longer restricted to textbooks, classrooms, and face-to-face interaction. Social media platforms have increasingly become spaces where learners encounter linguistic input, educational explanations, learning strategies, teacher identities, and representations of language-learning communities. Among these platforms, Instagram is particularly relevant because it

integrates photographs, illustrations, typography, written language, video, color, spatial arrangement, and interactive features within a single communicative environment (Leaver et al., 2020).

In English language teaching (ELT), Instagram offers opportunities to convey pedagogical messages through multimodal resources (Royce, 2002). An English-learning post may simultaneously contain a photograph of a teacher, an image of students interacting, English vocabulary, grammatical explanations, motivational quotations, institutional logos, typography, and visual symbols. Meaning is therefore not produced by written language alone. Rather, it emerges from the interaction of multiple semiotic modes (Jewitt, 2009; Kress & van Leeuwen, 2001).

This multimodal nature of digital communication makes Instagram particularly relevant to the study of visual grammar. Kress and van Leeuwen's (2006) social semiotic framework conceptualizes visual communication as a structured meaning-making system. As with language, visual resources can serve different communicative functions. Visual grammar generally examines three major dimensions: representational meaning, interactive meaning, and compositional meaning (Kress & van Leeuwen, 2006). These dimensions concern what is represented, how viewers are positioned relative to the represented participants, and how visual elements are organised to create hierarchy and coherence. Multimodality refers to the use and integration of multiple semiotic modes in communication (Kress, 2010). These modes may include written language, speech, images, gesture, color, typography, spatial arrangement, music, and movement (Jewitt, 2009; Kress & van Leeuwen, 2001). From a social semiotic perspective, meaning is not restricted to language, as different semiotic resources have distinct affordances for meaning-making (Kress, 2010).

Previous research has demonstrated the usefulness of visual grammar for examining multimodal texts in advertising, textbooks, tourism communication, and English language learning (Machin, 2007; Royce, 2002). Research on Indonesian EFL textbooks, for example, has shown that visual elements can construct representational, interpersonal, and compositional meanings and can support learners' comprehension and engagement. Research on Instagram has similarly demonstrated that visual and verbal resources operate together to construct meaning and communicative effects (Leaver et al., 2020). In contemporary ELT environments, multimodality is increasingly important because learners encounter English through websites, social media, videos, digital textbooks, learning applications, and other digitally mediated environments (Jones & Hafner, 2012). Consequently, language pedagogy needs to consider not only linguistic literacy but also learners' ability to interpret and produce

multimodal texts (New London Group, 1996; Royce, 2002).

Instagram provides a particularly relevant environment for multimodal language pedagogy because its communicative architecture encourages the integration of visual and verbal resources (Leaver et al., 2020). A single post can combine photographs, captions, typography, graphic design, emojis, symbols, and institutional branding. Kress and van Leeuwen's (2006) visual grammar provides a social semiotic framework for examining how images produce meaning. The framework is influenced by Halliday's (1978) systemic functional linguistics and proposes that visual communication can be analyzed through three broad metafunctions including: Representational meaning, Interactive meaning, Compositional meaning, Verbal mode, and Multimodal relationship (Halliday & Matthiessen, 2014; Kress & van Leeuwen, 2006).

Representational meaning concerns the participants, processes, and circumstances represented in an image (Kress & van Leeuwen, 2006, 2021). Two broad categories are particularly important: Narrative representation, which depicts participants engaged in actions, reactions, or processes; and Conceptual representation, which represents participants in terms of classification, structure, identity, or attributes (Kress & van Leeuwen, 2006). In an ELT Instagram account, narrative representations may include teachers explaining lessons, students participating in classroom activities, or learners interacting with one another (Jewitt, 2008; Lim, 2021). Conceptual representations may include portraits of teachers, institutional facilities, certificates, logos, or carefully arranged learning materials (Kress & van Leeuwen, 2006).

Interactive meaning concerns the relationship established between the image and its viewers (Kress & van Leeuwen, 2006). Important resources include gaze, social distance, perspective, and modality (Kress & van Leeuwen, 2006). For example, gaze may indicate whether represented participants symbolically address viewers. A direct gaze creates a "demand" image, establishing an imaginary relationship of engagement, while an indirect gaze creates an "offer" image, positioning viewers as passive observers of an activity (Kress & van Leeuwen, 2006). Social distance is constructed through image framing: close-up images create intimacy, whereas long shots establish a more distant or contextual relationship (Hall, 1966; Kress & van Leeuwen, 2006). Perspective concerns the angle from which represented participants are viewed; horizontal angles convey involvement or detachment, while vertical angles shape perceptions of power and status (Kress & van Leeuwen, 2006). Modality concerns the degree of perceived reality, validity, or credibility that visual resources such as color,

contextualization, detail, and naturalistic representation construct (Kress & van Leeuwen, 2006).

Compositional meaning concerns how visual elements are organized into a coherent whole through three major resources: information value, salience, and framing (Kress & van Leeuwen, 2006). Information value concerns the placement of elements in positions such as left/right (Given/New), top/bottom (Ideal/Real), or center/margin (Kress & van Leeuwen, 2006). Salience concerns the visual prominence of particular elements achieved through size, color, contrast, sharpness, or placement (Kress & van Leeuwen, 2006). Framing concerns the degree to which elements are visually connected or separated through borders, lines, or white space (Kress & van Leeuwen, 2006). In an ELT Instagram post, these resources determine whether the teacher, learning message, students, institutional logo, or call-to-action receives the greatest visual prominence (Lim & Tan, 2018; Yi & Jiang, 2024).

Recent scholarship has also begun to investigate multimodal language pedagogy on social media platforms (O'Halloran et al., 2017; Yi & Jiang, 2024). A 2026 study of Instagram posts from BBC Learning English and Learn English with Emma found that linguistic content, visual elements, and interactional features collectively contribute to meaning-making and learner engagement (e.g., Chen & Zhang, 2026). Such findings indicate that social media content should not be treated merely as supplementary illustrations of linguistic information; instead, visual resources themselves constitute essential pedagogical and communicative resources (Jewitt, 2008; Lim, 2021).

Despite this growing interest, relatively little attention has been paid to Indonesian community-based English-learning institutions and the ways their social media content constructs pedagogical meanings (Kusuma, 2022). Kampung Inggris represents a unique and important Indonesian context because English learning is strongly associated with community-based, immersive, and informal learning environments (Abrar et al., 2018; Zein et al., 2020). In this context, the Instagram presence of Kampung Inggris Mr Mark offers an interesting case study for examining how English education is visually represented to prospective and existing learners.

The present study therefore investigates the visual grammar of selected Instagram posts associated with Kampung Inggris Mr Mark. Rather than examining Instagram solely as a promotional platform, this study considers its content as a multimodal pedagogical discourse in which educational and institutional meanings intersect (Kress & van Leeuwen, 2006; Lim & Tan, 2018). The study addresses the following research questions:

1. How are representational meanings constructed in the Instagram posts of Kampung Inggris Mr. Mark?
2. How do visual resources construct relationships between represented participants and Instagram audiences?
3. How are compositional resources used to organize pedagogical and institutional meanings?
4. How do visual and verbal resources work together to construct English language pedagogy on Instagram?

METHODOLOGY

This study adopts a qualitative case study design integrated with Multimodal Discourse Analysis (MDA) (Creswell & Poth, 2018; Kress & van Leeuwen, 2006; Yin, 2018). A qualitative multimodal approach is particularly suited for this inquiry, as it moves beyond simple content counting to interpret how visual, verbal, and spatial semiotic resources interact to construct pedagogical meanings within digital environments (Jewitt, 2009; Merriam & Tisdell, 2016). The unit of analysis for this case study is the official Instagram presence of Kampung Inggris Mr. Mark. This account was selected via purposive sampling because it exemplifies an English Language Teaching (ELT) institution that systematically leverages social media to blend pedagogical instruction, socio-cultural branding, and learner recruitment (Campbell et al., 2020; Miles et al., 2014).

The primary data consisted of digital artifacts published on the target Instagram account. Following Emmel (2013), explicit inclusion and exclusion criteria were established to construct a rigorously defined dataset. The inclusion criteria were that the posts were selected if they explicitly featured English language-learning content; depicted teachers or students in learning environments; offered pedagogical tips, grammar explanations, or motivational messages; or combined visual and verbal resources to construct the institution's identity and instructional philosophy (Lim, 2021; Yi & Jiang, 2024). Meanwhile, posts consisting exclusively of routine administrative notices or generic promotional material without substantive pedagogical content were excluded (Kusuma, 2022).

Following these parameters, a purposive sample of 20 multimodal posts was collected, establishing a manageable and rich dataset for deep qualitative analysis (Emmel, 2013; Merriam & Tisdell, 2016). Sampling proceeded until data saturation was reached, ensuring that further sampling yielded no new semiotic patterns or pedagogical themes. Data were gathered through non-reactive digital archiving techniques (Highfield & Leaver, 2016; Kozinets, 2020). Each artifact was systematically

logged with its high-resolution visual component (still image or reel video clip), caption copy, publication date, embedded typography, emojis, hashtags, and relevant user engagement features (Page et al., 2014). To ensure systematic data tracking and referencing throughout the analysis, each entry was assigned an alphanumeric code (e.g., MM01, MM02, MM03) (Miles et al., 2014).

Table 1. Visual grammar framework

Dimension	Analytical features
Representational meaning	Participants, actions, reactions, vectors, narrative processes, conceptual processes
Interactive meaning	Gaze, social distance, perspective, angle, modality
Compositional meaning	Information value, salience, framing
Verbal mode	Caption, headline, instructional text, slogan, hashtag
Multimodal relationship	Reinforcement, elaboration, extension, complementarity

The analytical procedure followed a structured four-stage workflow: first, the identification of Representational Meanings, in which each post was categorised by its visual participants, narrative vectors, and conceptual structures to establish what pedagogical activity was taking place. Second, the analysis of Interactive Resources coded social distance, gaze vectors, and camera angles to evaluate how the visual compositions positioned the viewer in relation to the instructors and students. Third was the deconstruction of compositional meaning. In this stage, the spatial layouts, framing boundaries, and visual salience hierarchies were mapped to identify which instructional elements were foregrounded. Fourth intersemiotic mapping: here, visual and verbal modes were analysed synchronously to determine how text overlays, captions, and audio tracks complemented, reinforced, extended, or elaborated the visual images. To ensure analytical trustworthiness and credibility, coding categories were consistently applied across the entire corpus. Peer debriefing was conducted, wherein a second researcher independently coded a subset (20%) of the dataset. Any discrepancies in semiotic interpretation were discussed and reconciled to ensure inter-coder reliability.

RESULTS AND DISCUSSION

4.1 Representational Meaning

Representational meaning focuses on how visual semiotic resources depict real-world objects, participants, and conceptual relationships. In social semiotics,

representational structures are broadly categorized into narrative processes, which depict dynamic actions, events, or transactional reactions through visual vectors, and conceptual processes, which visually organise participants in terms of their static class, structure, essence, or meaning.

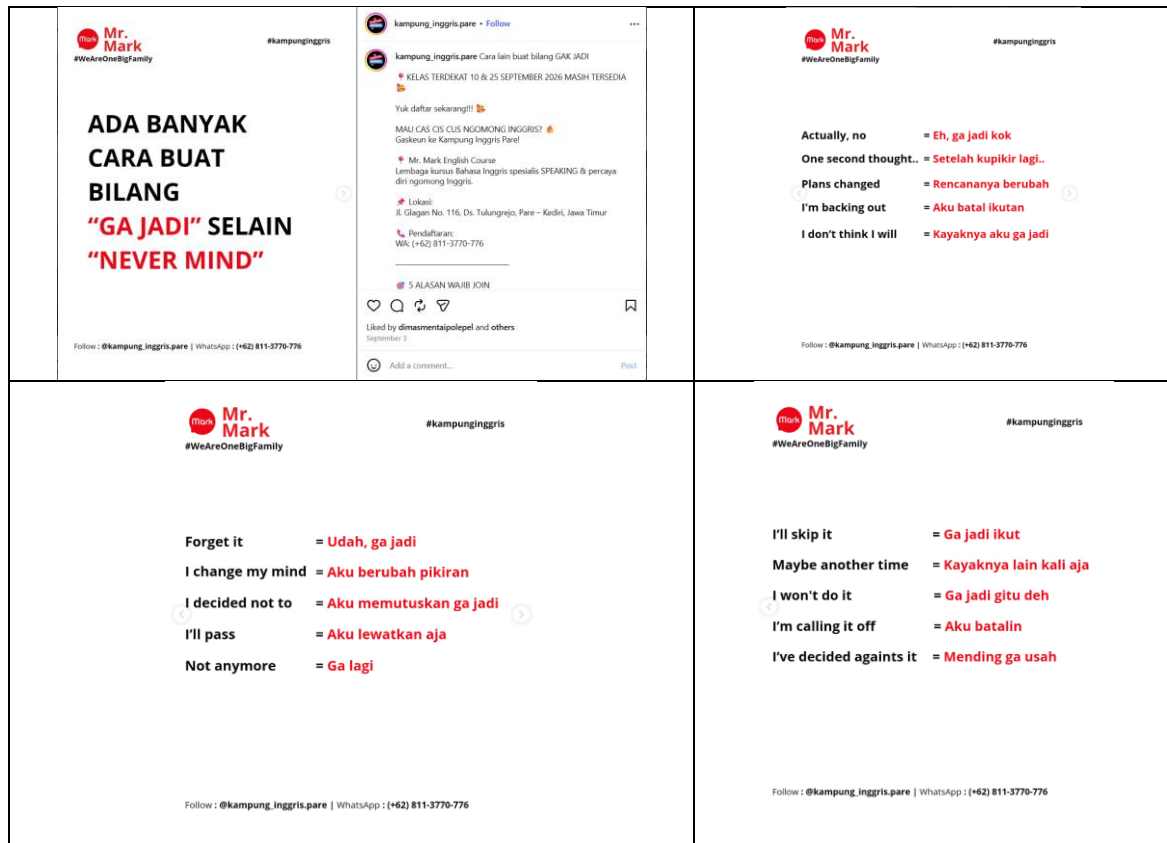


Figure 1. MMO 5 Ada banyak cara buat bilang gak jadi post

In this MMO 5 data, the images function as a visual-linguistic analytical structure, in which the language expressions themselves serve as the primary participants. Across both slides, the English phrases—such as "Actually, no", "On second thought..", "I'm backing out", "I'll skip it", and "I'm calling it off"—are positioned on the left side in bold black font. These English expressions act as the principal participants being analyzed. The Indonesian Equivalents as Symbolic Attributes was Positioned on the right, linked by an equals sign (=), are the informal Indonesian equivalents rendered in bold red font (e.g., "Eh, ga jadi kok", "Setelah kupikir lagi..", "Aku batalin", "Mending ga usah"). The red color serves as a symbolic marker that draws immediate visual attention, highlighting the familiar, casual Indonesian translations that anchor the meaning for the learner.

The use of the visual operator (=) explicitly establishes a relationship of

functional equivalence between formal/informal English idioms and everyday conversational Indonesian. This design breaks down complex pragmatic alternative phrases used when changing one's mind into bite-sized, paired units. The upper section of both slides integrates institutional identifiers, including the brand logo ("Mr. Mark"), the overarching slogan ("#WeAreOneBigFamily"), and the location hashtag ("#kampuninggris"). The bottom section provides contact details and social media handles. These elements visually frame the vocabulary lesson, transforming simple text into an institution-branded pedagogical product. By structuring the content as a text-dominant conceptual graphic rather than a human-centered narrative, the visual design directs the audience's full focus onto language as an object of study, allowing learners to easily read, digest, and save functional expressions in a fast-paced social media environment.

4.2. Interactive Meaning

Interactive meaning examines the relationship established between the semiotic visual artifact and its audience/viewer. For example, in MMO 10, the interactive dimension constructs a nuanced relationship between the depicted participants and the Instagram audience, balancing objective observation with emotional intimacy. Across all three images, the participants display an indirect gaze (looking at the board, each other, or off-camera). This establishes an "offer" image, positioning the Instagram viewer as an unseen observer watching an authentic, active classroom session rather than being directly addressed. Image 1: A long/medium-long shot showcasing the entire physical environment (floor, whiteboard, roof structure, and student body). This creates an objective distance, allowing viewers to evaluate the study atmosphere. Image 2 & Image 3: Medium-close to medium shots focused on individual facial expressions and engagement. This shorter distance fosters emotional closeness, empathy, and personal connection with the teacher and students. The camera angles are primarily at eye-level, establishing an egalitarian relationship between the participants in the photo and the Instagram viewer. In Image 1, the high-angle perspective looking down at students seated on the floor emphasizes a grounded, humble, and informal learning environment. Natural lighting, unposed candidate expressions, realistic colors, and authentic classroom settings enhance the credibility and authenticity of the learning experience.



Figure 2. MMO 10 Progression over perfection post

Vertical camera angles primarily mediate the power dynamic between viewers and the represented participants. The images are overwhelmingly shot at eye level, which creates an egalitarian perspective that places the Instagram audience on equal footing with the learners and instructors. Where slight high-angle perspectives occur, such as looking down at students seated on the floor in Image 1, the angle does not diminish student agency; rather, it underscores the grounded, non-hierarchical, and informal pedagogical environment characteristic of the program. Regarding visual contact, all three images feature participants who display an indirect eye gaze, looking toward the whiteboard, conversing with peers, or focusing off-camera rather than making direct eye contact with the lens. In social semiotic terms, this establishes an "offer" visual structure. Rather than directly demanding the viewer's immediate attention or response, the images position the social media user as an unseen, non-intrusive observer invited to witness an authentic, ongoing learning session in real time.

This social connection is further calibrated through strategic variations in framing and spatial distance across the carousel posts. Image 1 employs a long-to-medium-long shot that captures the broader physical environment, including the traditional roof structure, floor seating arrangement, whiteboard, and student body. This broader framing maintains a degree of objective distance, enabling prospective students to evaluate the institution's overall communal atmosphere and physical infrastructure. Conversely, Image 2 and Image 3 transition into medium-close and medium shots that prioritize individual facial expressions, smiles, and focused eye

contact during class discussions. By closing the physical distance, these tighter frames foster personal empathy, social familiarity, and emotional resonance, drawing the viewer into the teachers' and students' learning experiences.

4.3. Compositional meaning

Compositional meaning analyzes how visual elements are arranged to form a coherent whole through Information Value, Saliency, and Framing. For example, in MMO 13, the compositional architecture of the post employs a dual-axis Information Value structure, organising meaning along both horizontal (Left-to-Right) and vertical (Top-to-Bottom) spatial axes. Along the horizontal plane, the composition follows a classic Given/New layout. The left panel serves as the Given, depicting an authentic classroom environment where seated students interact in front of a whiteboard covered with handwritten grammar notes ("Coulda Shoulda Woulda", "Must", "Could"). This visual space grounds the viewer in a familiar, real-world context representing the lived learning experience at Kampung Inggris. Conversely, the right panel serves as the New, providing structured, digital text that defines the core pedagogical distinction between "LIVE" (permanent dwelling) and "STAY" (temporary stay), accompanied by bilingual sample sentences. By positioning the explicit linguistic explanation on the right, the graphic introduces key takeaways as actionable knowledge that resolves the visual scenario presented on the left.

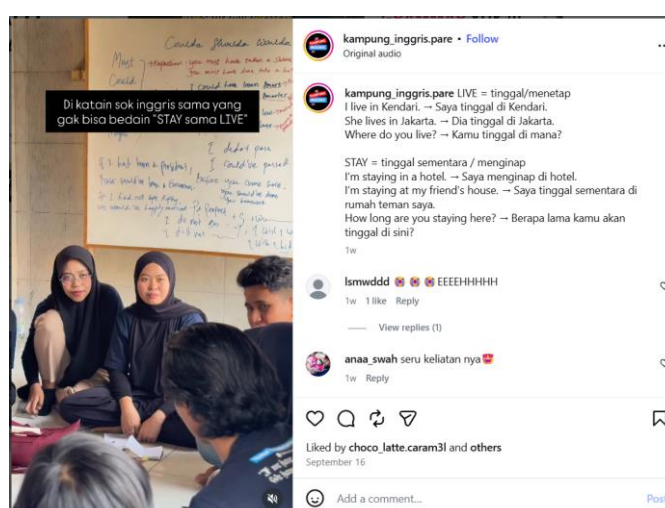


Figure 3. MMO 13 Dikatain sok Inggris post

Along the vertical axis within the left panel, the composition constructs an

Ideal/Real hierarchy. The Ideal region at the top left is dominated by a prominent black text box featuring the overlay: "Di katin sok inggris sama yang gak bisa bedain 'STAY sama LIVE'" (Mocked as being arrogant in English by someone who can't even tell the difference between 'STAY' and 'LIVE'). This top-aligned overlay presents the emotional hook and narrative ideal, a relatable social anxiety and humorous frustration common among language learners. The Real at the bottom panel grounds this narrative conflict in physical reality, depicting students sitting together on the floor in a relaxed, non-intimidating learning environment. This lower space visually presents the institution's supportive classroom culture as the tangible solution to overcoming social judgment. Visual salience regulates how viewer attention is directed across the composition through contrast, human figures, and typographical weight. Immediate focal entry is established by the stark black text box in the upper-left quadrant. Its bright white lettering contrasts sharply against the light wooden frame of the whiteboard, ensuring that the emotional and narrative hook is the first element processed by the viewer. Secondary salience is captured by the human subjects in the middle ground. The facial expressions of the seated students, particularly the central female student looking directly at the camera, draw strong viewer engagement through their central alignment, facial clarity, and welcoming affect. Within the right-hand textual interface, high typographic salience is assigned to the capitalized terms LIVE and STAY. Their distinctive font weight and capitalization immediately highlight the primary linguistic targets of the post.

Framing mechanisms control the degree of visual connection or disconnection between the multimodal elements. A sharp, high-contrast vertical dividing border completely separates the left photographic panel from the right textual panel. This explicit visual severance creates a dual-mode layout that clearly delineates contextual, photographic evidence (left) from structured, formal instruction (right). The solid black background surrounding the overlay phrase ("Di katin sok inggris...") serves as a hard visual frame, separating the social commentary from the background whiteboard notes and the students below. This ensures optimal legibility without cluttering the photographic composition. In contrast to the rigid outer framing, the students within the photograph are depicted without internal borders, seated close together on the floor. This deliberate absence of internal dividers visually reinforces themes of community, peer support, and inclusive learning.

4.4. Verbal mode

In multimodal social semiotics, the verbal mode examines how written text,

spoken language, typographic choices, and linguistic structures interact with visual elements to convey educational and promotional messages. In these social media images (Reels/Carousel posts), the verbal mode operates across three distinct layers: On-Screen Text Overlays, Physical Whiteboard Text, and the Digital Caption Interface.

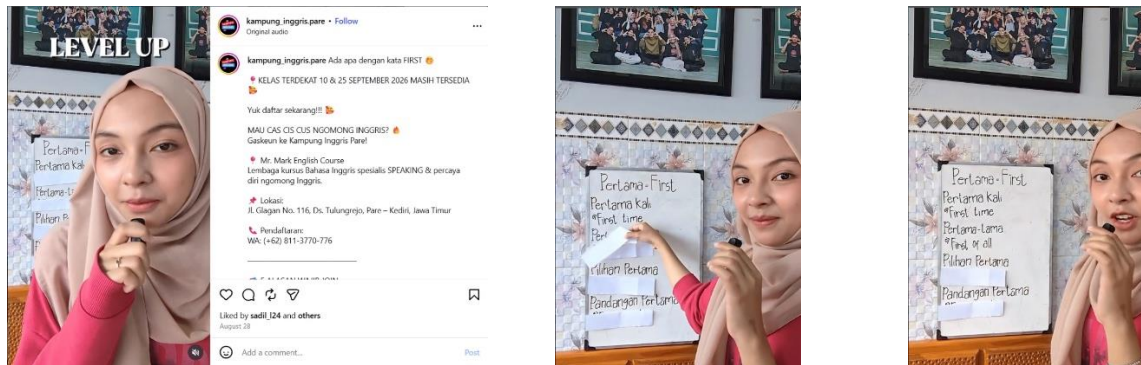


Figure 4. MMO 18 Level Up post

The visual media utilizes strategically placed text overlays to establish immediate thematic engagement and brand positioning. At the top-left corner of the video frame, a bold, sans-serif white overlay displaying the phrase "LEVEL UP" serves as a primary motivational anchor. By borrowing terminology from gaming culture and contemporary youth slang, this graphic overlay frames language learning not as a tedious academic chore, but as an empowering skill upgrade. Complementing this visual anchor, the initial line of the digital caption initiates interaction through a rhetorical query: "Ada apa dengan kata FIRST" (What's up with the word FIRST?). Paired with a inquisitive emoji, this open-ended teaser deliberately engineers a "curiosity gap," incentivizing viewers to sustain their attention through the video to uncover the linguistic resolution.

Within the photographic space, the physical whiteboard acts as the central instrument for explicit instruction, utilizing a structured bilingual comparative layout. The board presents Indonesian base expressions on the left—such as "Pertama", "Pertama kali", "Pertama-lama", "Pilihan Pertama", and "Pandangan Pertama"—paired side-by-side with their corresponding English target phrases (e.g., "First", "First time", "First of all"). To enhance engagement during short-form video consumption, the instructor incorporates a tactile "reveal" mechanic by physically covering the target English translations with paper flaps. This gamified visual-verbal technique transforms static vocabulary acquisition into an interactive guessing game, holding viewer focus through anticipation. Furthermore, the use of handwritten typography

on the whiteboard establishes an informal, authentic classroom modality, reducing learner anxiety and making complex grammatical nuances far more approachable than those in standardised textbook print.

The accompanying digital caption transitions the viewer from educational consumption to actionable institutional enrollment through three distinct linguistic techniques. The caption employs capitalized typography, location pin emojis, and exclamation marks to signal limited availability, stating: "📍 KELAS TERDEKAT 10 & 25 SEPTEMBER 2026 MASIH TERSEDIA" followed by an direct imperative call-to-action: "Yuk daftar sekarang!!!" (Come on, register now!). To align with the informal identity of Kampung Inggris and build cultural proximity with young learners, the copy integrates casual Indonesian expressions: "MAU CAS CIS CUS NGOMONG INGGRIS?" (Want to speak English effortlessly/fluenty?) "Gaskeun ke Kampung Inggris Pare!" (Full speed ahead to Kampung Inggris Pare!). The caption concludes with clear, itemized contact and administrative details—highlighting the program's primary value proposition ("spesialis SPEAKING & percaya diri") alongside precise physical location data and direct WhatsApp registration links—completing the functional arc from engagement to conversion.

4.5. Multimodal relationship

In multimodal discourse analysis the interaction between different semiotic modes—visual (moving/still images), verbal (text overlays and captions), and auditory (audio/dubbing)—creates a unified communicative message. The relationship between on-screen text overlays and visual imagery functions predominantly through intersemiotic elaboration and extension. In the opening sequence of the video reel, an on-screen text overlay reading "Jangan panggil aku anak kecil paman 😊" (Don't call me a little kid, uncle) is super-imposed over a shot of two young child students standing on stage, holding microphones and cue cards. Semiotically, this text operates as a concurrence and framing mechanism. By projecting a popular pop-culture voice onto the child subjects, the overlay re-frames the young learners not merely as passive pupils, but as self-assertive, capable individuals performing in a mature public-speaking capacity.

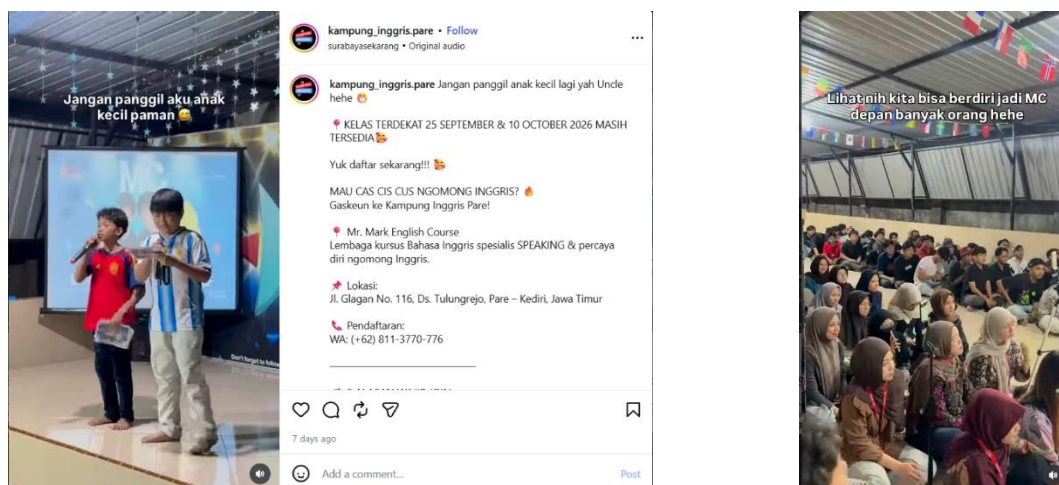


Figure 5. MMO 20 Jangan panggil aku anak kecil paman post

As the video camera pans from the stage to the surrounding environment, a second text overlay appears: "Lihat nih kita bisa berdiri jadi MC depan banyak orang hehe" (Look at us standing as MCs in front of many people hehe). Here, the visual-verbal relationship shifts to exposition and vectorial alignment. While the visual shot depicts a large, attentive adult audience seated on the floor, the accompanying text serves as an internal monologue from the children's perspective. This creates an intersemiotic action-result chain: the physical action of child bravery on stage is directly validated by the visual presence of a crowded audience room. Furthermore, the relationship between the video artifact and the digital caption exemplifies Barthesian anchorage and relay. While the video visually demonstrates public speaking competence and child empowerment, the digital caption ("Lembaga kursus Bahasa Inggris spesialis SPEAKING & percaya diri...") contextualizes this performance as commercial proof. The caption converts the entertaining visual demonstration into an institutional value proposition: if young children can overcome anxiety and master public speaking within this learning environment, adult prospective students can achieve similar fluency.

The auditory dimension introduces a layer of playfulness through audio-visual juxtaposition. Rather than using ambient classroom noise or standard instructional speech, the reel utilizes a viral, dramatic Indonesian voiceover featuring a child pleading ("Guru... Hentikan guru... Kami tidak ingin berdosa guru!"). This audio track stands in stark semiotic dissonance with the visual scene, which depicts two calm, disciplined children presenting on stage before a quiet audience. This intentional incongruity between the high-drama audio and the orderly physical environment generates humor, an essential pragmatic strategy for maximizing viewer

engagement and retention under short-form video algorithmic constraints.

Finally, spatial and temporal cohesion across modalities creates a complete narrative arc within a brief 10-second duration. The spatial transition from the stage performers to the wide hall audience is mirrored by a verbal shift from identity construction ("Don't call me a kid") to performance accomplishment ("MCing in front of many people"). By synchronizing spatial camera movement, narrative text progression, and comedic audio cues, the post effectively demonstrates the institution's core pedagogical outcome—building public speaking confidence (*percaya diri*), while maintaining high digital entertainment value.

Discussion

The empirical findings of this study demonstrate how an Indonesian English Language Teaching (ELT) institution (Kampung Inggris Mr. Mark) strategically orchestrates multimodal semiotic resources, visual, verbal, typographic, spatial, and auditory, to construct a hybrid identity that bridges informal social media culture with formal pedagogical authority. By deconstructing the representational, interactive, compositional, verbal, and intersemiotic dimensions of the Instagram dataset, this study uncovers three major thematic axes that define digital ELT discourse in the platform era: (1) The Pedagogical Re-semiotization of Language as a Digital Product, (2) Affective Engineering and Participatory Distance, and (3) Algorithm-Driven Pedagogical Humour and Intersemiotic Complementarity.

At the representational level, the findings reveal a distinct trend toward the re-semiotization (Iedema, 2003) of abstract linguistic knowledge into concrete, visually digestible artifacts designed for rapid social media consumption. Traditional language learning materials typically present grammar and pragmatic functions through continuous prose or formal textbook dialogues (Cook, 2013). However, in digital environments like Instagram, language is re-packaged as a conceptual object of study through structured visual-linguistic analytical layouts (as seen in MMO 05). By utilizing visual operators such as the equals sign (=) and color coding (e.g., bold red font for informal Indonesian equivalents), the institution constructs an explicit framework of pragmatic equivalence between target English idioms and everyday conversational Indonesian. This aligns with Halliday's (1978) ideational metafunction and Kress and van Leeuwen's (2006) concept of analytical processes, wherein complex socio-pragmatic concepts (e.g., changing one's mind using phrases like "On second thought" or "I'm backing out") are broken down into bite-sized, paired units.

Furthermore, framing these expressions within institutional branding (logos,

slogans like #WeAreOneBigFamily, and location hashtags) transforms raw linguistic data into a proprietary pedagogical product (Fairclough, 2006; Machin & Mayr, 2012). As Jewitt (2009) notes, digital pedagogy relies heavily on screen-based layout designs that allow learners to "scan, digest, and archive" information quickly. The present study extends this notion to ELT marketing: by treating language expressions themselves as the primary visual participants, the institution caters to the cognitive scrolling habits of modern digital natives (Prensky, 2001; Yi & Jiang, 2024), effectively turning social media feeds into supplementary mobile learning environments.

Analysis of the interactive and compositional dimensions (MMO 10 and MMO 13) demonstrates how spatial framing and camera angles are deployed to perform affective engineering—minimizing learner anxiety while maximizing social proximity (Krashen, 1982; Lim, 2021). In social semiotic terms, the dominant use of indirect gaze establishes an "offer" relation (Kress & van Leeuwen, 2006). Rather than confronting the viewer with a direct "demand" gaze that requires immediate response, the images position the social media user as an unseen, non-intrusive observer witnessing an authentic classroom setting in real time.

The prevailing eye-level framing creates a symmetrical, non-hierarchical power relationship between the viewer, the instructors, and the students (Bezemer & Kress, 2016). Even where slight high-angle shots occur (e.g., looking down at students seated on the floor), the perspective does not diminish learner agency; instead, it visually highlights the grounded, informal, and non-intimidating nature of the learning environment (Bourdieu, 1991). The movement from long shots (which display physical infrastructure and community scale) to medium-close shots (focusing on facial expressions and engagement) creates what Hall (1966) terms personal social distance. This transition fosters empathy and emotional closeness, allowing prospective students to project themselves into the physical space of the classroom.

Compositionally, the dual-axis layout identified in MMO 13 leverages Given/New and Ideal/Real structures (Kress & van Leeuwen, 2006) to resolve social anxieties associated with language learning in non-native contexts. By placing the narrative conflict—the fear of being mocked for speaking English ("Dikatain sok Inggris...")—in the top-left Ideal region, and grounding the physical solution (a supportive, floor-seated peer group) in the bottom Real region, the post visually operationalizes a problem-solution framework (Hoey, 2001). This confirms previous research in multimodal branding (Page, 2012; Vasquez, 2014), which suggests that social media educational posts succeed not merely by presenting information, but by addressing the socio-emotional barriers of their prospective clientele.

The analysis of the verbal mode and intersemiotic relationships (MMO 18 and MMO 20) highlights the crucial role of linguistic hybridity and cross-modal dissonance in driving engagement. Within the verbal captions and whiteboard text, there is a seamless integration of formal English instruction with informal Indonesian slangs (e.g., "Cas Cis Cus" and "Gaskeun"). In sociolinguistic terms, this code-mixing serves as an affective bridge (Holmes & Wilson, 2017; Zentz, 2015). By employing colloquial expressions popular among Indonesian youth, the institution signals cultural proximity and demystifies the perceived elitism of the English language in post-colonial contexts (Pennycook, 2017; Sulistiyo et al., 2016).

A critical theoretical finding emerges from the auditory-visual relationships in short-form video reels (MMO 20). Unlike traditional instructional videos where audio-visual harmony is maintained, digital social media pedagogy frequently employs intentional intersemiotic dissonance (Unsworth, 2006; Royce, 2007). In MMO 20, the extreme dramatic tension of a trending, crying voiceover ("Guru... Hentikan guru!") is juxtaposed against a serene visual scene of two young child students calmly presenting on stage as MCs.

This semiotic contradiction generates humor through incongruity theory (Attardo, 1994). Far from distracting from the educational message, this playful dissonance fulfills an algorithmic imperative: it hooks viewer attention within the first three seconds of scrolling, maximizing platform watch time and algorithmic reach (Gillespie, 2014; Zulli & Zulli, 2022). Once the viewer's attention is captured by the humorous audio-visual mismatch, the Barthesian anchorage and relay mechanisms (Barthes, 1977) in the caption redirect this focus toward the core institutional message: building public speaking confidence (percaya diri).

CONCLUSION

This study conceptualizes Instagram as an important multimodal environment for contemporary English language pedagogy. Through the framework of visual grammar, Instagram posts can be examined as structured semiotic texts in which representational, interactive, and compositional resources work together to construct educational meanings. The case of Kampung Inggris Mr. Mark provides an important Indonesian context for understanding how English learning can be represented through social media. The analysis suggests that visual representations of teachers, learners, learning activities, and educational messages can simultaneously construct pedagogical meanings and institutional identity. The study further highlights that digital ELT communication cannot be adequately understood by examining written captions alone. Visual resources contribute independently and interactively to the

construction of meaning. Consequently, multimodal literacy should become an increasingly important component of language education and digital pedagogy. Future studies may extend this research by comparing several Indonesian English-learning institutions, examining video-based Instagram content, investigating learners' interpretations of visual grammar, or combining multimodal discourse analysis with audience reception studies.

REFERENCES

- Abbas, N. F., Jasim, M. M., & Hussein, K. A. (2026). A social semiotic analysis of Instagram as a multimodal platform for supporting EFL learners. *Arab World English Journal*, 17(1), 49–65. <https://doi.org/10.24093/awej/vol17no1.4>
- Abrar, M., Mukminin, A., Habibi, A., Asyraf, F. A., Makmur, M., & Marzulina, L. (2018). "If our English is bad, our future is gloomy": English language learning motivation in a community-based learning center in Indonesia. *The Qualitative Report*, 23(1), 129–145.
- Atmawijaya, T. D., Ningsih, H. K., & Kushartanti, B. (2026). Visual grammar and intersemiotic complementarity in BIPA textbooks: A multimodal analysis of *Sahabatku Indonesia*. *Alinea: Jurnal Bahasa, Sastra dan Pengajaran*, 6(2), 1010–1024.
- Attardo, S. (1994). *Linguistic theories of humor*. Walter de Gruyter.
- Barthes, R. (1977). *Image, music, text*. Fontana Press.
- Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication*, 25(2), 166–195.
- Bezemer, J., & Kress, G. (2016). *Multimodality in learning and knowing: The assignment of meaning*. Routledge.
- Bourdieu, P. (1991). *Language and symbolic power*. Harvard University Press.
- Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Murdoch, N. (2020). Purposive sampling: Complex or simple? Research case examples. *Journal of Research in Nursing*, 25(8), 652–661. <https://doi.org/10.1177/1744987120927206>
- Cook, G. (2013). *Discourse of advertising*. Routledge.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Emmel, N. (2013). *Sampling and choosing cases in qualitative research: A realist approach*. SAGE Publications.
- Fairclough, N. (2006). *Language and globalization*. Routledge.
- Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). MIT Press.

- Hall, E. T. (1966). *The hidden dimension*. Doubleday & Company.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge.
- Highfield, T., & Leaver, T. (2016). Instagrammatics and digital methods: Studying visual social media, Instagram and contemporary visual culture. *Communication Research and Practice*, 2(1), 47–62. <https://doi.org/10.1080/22041451.2016.1155332>
- Hoey, M. (2001). *Textual interaction: An introduction to written discourse analysis*. Routledge.
- Holmes, J., & Wilson, N. (2017). *An introduction to sociolinguistics* (5th ed.). Routledge.
- Iedema, R. (2003). Multimodality, resemiotization: Extending the analysis of discourse as multi-semiotic practice. *Visual Communication*, 2(1), 29–51.
- Jewitt, C. (2008). Multimodality and literacy in school classrooms. *Review of Research in Education*, 32(1), 241–267. <https://doi.org/10.3102/0091732X07310586>
- Jewitt, C. (Ed.). (2009). *The Routledge handbook of multimodal analysis*. Routledge.
- Jones, R. H., & Hafner, C. A. (2012). *Understanding digital literacies: A practical introduction*. Routledge.
- Kozinets, R. V. (2020). *Netnography: The essential guide to qualitative social media research* (3rd ed.). SAGE Publications.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.
- Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- Kusuma, I. P. I. (2022). EFL teachers' instructional design and multimodal practices in social media-assisted language learning. *Journal of Language and Education*, 8(2), 85–98.
- Leaver, T., Highfield, T., & Abidin, C. (2020). *Instagram: Visual social media cultures*. Polity Press.
- Lim, F. V. (2021). *Designing learning for multimodal literacy: Investigating the pedagogy of digital texts*. Routledge.
- Lim, F. V., & Tan, S. K. Y. (2018). Developing multimodal literacy in the English classroom. *English in Education*, 52(2), 163–180. <https://doi.org/10.1080/04250494.2018.1472504>
- Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, 116(1), 75–98.
- Machin, D. (2007). *Introduction to multimodal analysis*. Bloomsbury.
- Machin, D., & Mayr, A. (2012). *How to do critical discourse analysis: A multimodal approach*. SAGE Publications.

- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.
- Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60–92.
- O'Halloran, K. L., Tan, S., & E, M. K. (2017). Multimodal analysis for critical thinking. *Learning, Media and Technology*, 42(2), 147–170.
- Page, R. (2012). *Stories and social media: Identities and interaction*. Routledge.
- Page, R., Harper, R., & Unsworth, L. (2014). *Researching language and social media: A student guide*. Routledge.
- Pennycook, A. (2017). *The cultural politics of English as an international language*. Routledge.
- Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the Horizon*, 9(5), 1–6.
- Royce, T. (2002). Multimodality in the TESOL classroom: Exploring visual-verbal synergy. *TESOL Quarterly*, 36(2), 191–205.
- Royce, T. D. (2007). Intersemiotic complementarity: A framework for multimodal discourse analysis. In T. D. Royce & W. Bowcher (Eds.), *New directions in the analysis of multimodal discourse* (pp. 63–109). Lawrence Erlbaum Associates.
- Sakdiyah, R., Perangin-angin, A., Zein, T. T., Sinar, T. S., & Rangkuti, R. (2026). Multimodal literacy: A visual grammar analysis of Indonesian EFL textbooks in senior high schools. *Journal of General Education and Humanities*, 5(2).
- Sulistiyo, U., Mukminin, A., Abdurrahman, A., & Edy, S. (2016). Learning to teach English: Indonesian student teachers' views on English language teaching practices. *The Qualitative Report*, 21(11), 2073–2089.
- Unsworth, L. (2006). Towards a metalanguage for multiliteracies education: Describing the meaning-making resources of language-image interaction. *English Teaching: Practice and Critique*, 5(1), 55–76.
- Vasquez, C. (2014). *The discourse of online reviews*. Bloomsbury Publishing.
- Yi, Y., & Jiang, L. (2024). Multimodal social media practices in language teaching and learning: A systematic review. *Computers & Education*, 208, 104940.
- Zein, S., Sukyadi, D., Hamied, F. A., & Lengkanawati, N. S. (2020). English language education in Indonesia: Policy, research, and practice. *Asian Englishes*, 22(1), 85–101. <https://doi.org/10.1080/13488678.2019.1629402>
- Zentz, L. (2015). "Is English also the language of Islam?" Ideologies of 'English' and 'globalization' in Indonesian Islamic higher education. *Journal of Language, Identity & Education*, 14(2), 70–85.
- Zhu, J. (2021). A multimodal discourse analysis of second language instruction videos. *Open Access Library Journal*, 8, 1–11.
- Zulli, D., & Zulli, D. J. (2022). Extending the internet meme: Conceptualizing technological mimicry and viral engagement on TikTok. *New Media & Society*, 24(8),

1872–1890.