

Multimodal Discourse Analysis of English-Language Instagram Posts: Caption, Image, and Platform Conventions

Aida Jalilzada

PhD Candidate, Lecturer, Nakhchivan State University, Nakhchivan, Azerbaijan

Email: aidajalilzada@ndu.edu.az

ORCID: <https://orcid.org/0009-0003-2463-2053>

© 2026 Aida Jalilzada. This is an open-access article published in *Global Spectrum of Research and Humanities (GSRH)*. Licensed under **Creative Commons Attribution 4.0 International (CC BY 4.0)** — free to share, adapt and use commercially, with attribution.

<https://creativecommons.org/licenses/by/4.0/>

DOI: <https://doi.org/10.69760/gsrh.0260304027>

Abstract

Instagram posts communicate through coordinated verbal, visual, and platform-specific resources. This conceptual article examines how captions, photographs, typography, colour, emojis, and hashtags contribute to meaning in English-language posts. Drawing on social semiotics, multimodal discourse analysis, and digital communication research, it proposes a practical framework for interpreting the relationship between caption and image, the organization of visual attention, the pragmatic work of concise language, and the role of platform conventions. The article treats these features as context-sensitive resources rather than fixed codes: the meaning of an emoji, colour, or hashtag depends on its placement, surrounding text, image, audience, and genre. It also distinguishes textual interpretation from claims about audience response, which require reception data. The framework is intended for close analysis and future corpus-based research; it does not report a newly collected Instagram dataset or statistical findings. By integrating linguistic and visual analysis, the article offers a way to describe how posts construct stance, identity, and audience alignment while remaining attentive to the limits of interpretation.

Keywords: *multimodality; Instagram; social semiotics; digital discourse; visual communication; emoji; hashtags*

1. Introduction

Instagram is a setting in which written language routinely appears alongside photographs, video, graphic design, and interface features. A post can include a caption, text embedded in an image, hashtags, emoji, tagged accounts, and audience-facing prompts. These elements do not simply accumulate: they are arranged to guide interpretation and invite particular forms of engagement. Studying the caption alone therefore gives an incomplete account of how a post communicates. Multimodality research provides concepts for examining how distinct semiotic resources contribute to a communicative event (Bateman et al., 2017; van Leeuwen, 2015).

This article focuses on English-language Instagram posts as multimodal discourse. It asks how verbal and visual resources relate, how linguistic choices contribute to stance and

audience address, and how platform conventions shape the interpretation of posts. The goal is a conceptual and analytical synthesis, not a report of findings from an empirical sample. Examples are illustrative constructions, and no frequency claims about Instagram users are made.

The discussion connects established multimodal scholarship with research on social media discourse, hashtags, and emoji. Jalilzada's (2022) analysis of abbreviations in media discourse is relevant to interpreting compressed language in online contexts. Ismayilli et al. (2025) examine gestures and facial expressions in EFL classroom interaction. Although their study does not concern Instagram, it supports a useful methodological point for this article: nonverbal resources should be interpreted in relation to verbal communication and the interactional setting, rather than assigned fixed meanings in isolation.

2. Conceptual and Methodological Approach

The article adopts a selective conceptual review. It brings together social-semiotic accounts of multimodality, frameworks for visual composition, and research on digital discourse. The analysis is organized around four relationships: caption and image; linguistic form and pragmatic purpose; visual composition and attention; and platform markers such as hashtags and emoji. This structure is intended to support future close readings and empirical studies.

Multimodal analysis considers how meaning is distributed across resources and how their arrangement creates a whole. It does not assume that each mode has a stable, independent meaning. A photograph, caption, or emoji can be interpreted differently depending on genre, sequence, audience knowledge, and the surrounding platform context (Jewitt et al., 2016; Kress & van Leeuwen, 2021). Accordingly, the interpretive claims below concern plausible functions in context. Claims about how actual viewers understood a post would require audience or reception evidence.

For a future dataset, researchers should specify the account and post types sampled, the collection period, inclusion criteria, and treatment of deleted, private, or ephemeral content. They should preserve relevant context while minimizing unnecessary personal data, especially when posts involve private individuals. A coding scheme could record caption structure, image–text relation, salient compositional features, emoji and hashtag placement, and interactional prompts. A second analyst or an explicit recoding procedure could help identify ambiguous cases. These are methodological recommendations, not procedures already carried out in this conceptual article.

3. English-Language Features in Instagram Discourse

3.1. Lexical choice and compressed syntax

Instagram captions range from extended narratives to short phrases, sentence fragments, and single-word evaluations. Concision can make a caption visually easy to process and can create an informal tone, but brevity is not a universal platform requirement. Users may also write long reflections, promotional descriptions, or explanatory texts. The form should therefore be interpreted in relation to post genre and communicative purpose rather than explained as a simple consequence of character limits.

Colloquial vocabulary, evaluative adjectives, idioms, contractions, and direct address can make a caption sound conversational. A phrase such as “Such a good day” may frame an image as personal recollection; paired with a product photograph, the same expression could

function as an endorsement or promotional stance. These examples demonstrate interpretive possibilities, not observed data. Jalilzada's (2022) discussion of abbreviations in media discourse is relevant to the broader point that compressed forms need contextual interpretation: an abbreviation can be ambiguous when detached from its communicative setting.

Ellipsis and fragments can also rely on the image to supply omitted information. A caption such as “Finally here” may be intelligible when paired with a recognizable travel scene, while without the image it remains underspecified. The image does not simply replace missing words; it supplies contextual cues that readers combine with linguistic knowledge and assumptions. Analysis should identify what information is explicit, what is inferred, and what remains uncertain.

3.2. Pragmatic purposes and audience address

Posts may inform, narrate, entertain, promote, express affiliation, or solicit a response. A question in a caption can invite comments, while an imperative may direct viewers toward a link, event, or action. First-person forms foreground personal experience; second-person forms create direct address. These grammatical choices can help construct an imagined relationship with viewers, though the presence of a prompt does not establish that viewers responded or felt addressed in the same way.

Digital discourse research examines how language contributes to relationships and searchable forms of participation online (Page, 2012; Zappavigna, 2012). In Instagram posts, shared tags and recurring expressions may signal affiliation, but a hashtag alone cannot prove membership in a community. It may also categorize a post, increase its discoverability, mark a campaign, or be used ironically. Researchers should interpret a marker together with the post and account context rather than infer stable identity from one token.

4. Verbal and Visual Relations

4.1. Caption–image relations

A caption may complement, anchor, extend, or contrast with an image. A descriptive caption can identify a place or person; an evaluative caption can direct attention toward an emotional reading; a narrative caption can supply events that are not visible. In other posts, the caption and image may create irony or productive tension. Multimodal research on Instagram demonstrates the value of studying document intent through the integration of image and text rather than treating either as sufficient by itself (Kruk et al., 2019). Studies of brand communication make a similar point: Tenca (2022) applies multimodal discourse analysis to the construction of brand personality and user engagement on Instagram, showing how verbal and visual choices work together in promotional posts.

The relation is not always one of simple illustration. A photograph may invite several readings, while the caption narrows or redirects them. Conversely, an image may complicate a caption that appears straightforward. A useful close-reading procedure is to describe each mode separately first, then explain how their juxtaposition changes the interpretation. Researchers should avoid claiming a single intended meaning unless supported by contextual evidence such as the creator's explanation or recurring practice.

4.2. Composition, colour, and visual attention

Visual composition can shape the path through which viewers encounter a post. Salience, framing, angle, distance, gaze, and placement influence the relation between depicted participants and viewers. Kress and van Leeuwen's (2021) visual grammar offers analytical

categories for describing such choices, while social-semiotic accounts emphasize that visual conventions are culturally and historically situated. A compositional feature should be described before assigning it a psychological or ideological effect.

Colour may contribute to mood, contrast, brand identity, or thematic association, but colour meanings are not universal. The interpretation of a hue depends on its combinations, cultural setting, design genre, and relation to depicted objects. Likewise, a close-up may foreground detail without necessarily implying intimacy in every context. Multimodal analysis is strongest when it distinguishes observable features from interpretations and supports interpretations through the broader design of the post (Bateman et al., 2017; van Leeuwen, 2015).

This contextual principle also applies to nonverbal communication. Ismayilli et al. (2025), studying EFL classroom interaction, report on the role of gesture and facial expression alongside speech. The classroom setting differs from asynchronous Instagram communication, so their findings cannot be transferred as evidence about social-media users. They nevertheless reinforce the need to analyze visible resources in conjunction with verbal content and communicative context. For Instagram research, this means describing the image or gesture, its relation to the caption, and the post genre before inferring stance or emotion.

Video and carousel posts add sequence and temporal organization. The order of frames, transitions, and on-screen text can structure attention differently from a single image. A static screenshot may not preserve these relations. Future research should record the format and sequence of the original post where ethically and technically possible, while noting that platform interfaces and affordances can change over time.

4.3. *Emoji and hashtag functions*

Emoji can contribute affect, stance, emphasis, humour, or disambiguation. Their function depends on where they appear and what accompanies them. An emoji at the end of a caption may intensify its tone; the same symbol used repeatedly may serve as a visual motif or account convention. Logi and Zappavigna (2023) approach emoji from a social-semiotic perspective, emphasizing the interaction between emoji and language. Highfield (2018) likewise examines emoji and hashtag practices as flexible resources shaped by platform cultures. These accounts caution against assigning a fixed dictionary meaning to each icon.

Hashtags can categorize content, connect a post to a topic, participate in campaigns, or add evaluative commentary. Their visibility can also make them part of the post's visual design. The discursive relations among tags may connect posts across topics and communities, and the use of tagging to align around shared values has been described as communing affiliation (Zappavigna & Martin, 2018); nonetheless, actual patterns of circulation need platform-specific evidence (Rathnayake & Suthers, 2025; Zappavigna, 2018). Researchers should record placement, spelling, capitalization, and relation to the caption, since these details may affect readability and interpretation.

5. Digital Identity, Interaction, and Research Limits

A profile and its posts can contribute to a performed digital identity through recurring choices of language, image, colour, and genre. This identity is neither fully contained in a single post nor directly accessible from visual appearance. Posts are curated representations produced within platform conditions, and audience interpretation may differ from creator intention. Research on social media therefore benefits from treating identity claims as situated and

provisional rather than as transparent descriptions of the person behind an account (Page, 2012; Zappavigna, 2012).

Interaction features, including comments, likes, and replies, add a participatory dimension. They may provide evidence of engagement, but visible metrics do not reveal the full meanings or motivations of viewers. Access to platform data is also selective and subject to change. Claims about reach, persuasion, or community formation should rely on appropriate observational or reception methods rather than be inferred from post design alone.

The present framework has a deliberate limitation: it does not analyze a defined corpus of public Instagram posts. It therefore identifies analytical categories and hypotheses, not empirical trends. Future work could compare public-facing genres — such as personal, educational, and promotional posts — using a transparent sample and a consistent coding protocol. Qualitative close reading could be combined with corpus methods or interviews, provided privacy, consent, and platform terms are considered. This would make it possible to test how often particular relations occur and how audiences interpret them.

6. Conclusion

English-language Instagram posts create meaning through interaction among language, image, visual design, and platform conventions. Captions can describe, evaluate, narrate, or address viewers; images can contextualize or complicate those captions; composition can organize visual attention; and emoji and hashtags can contribute stance, categorization, and affiliation. These functions are context-sensitive and should be demonstrated through analysis of the post as a whole.

The article has offered a conceptual framework rather than empirical findings. It connects multimodal theory with digital-discourse research and incorporates Jalilzade's (2022) work on abbreviations in media discourse and Ismayilli et al.'s (2025) study of nonverbal communication in EFL classrooms. The latter is used as a contextual comparison, not as evidence about Instagram. Further research should use transparent sampling, careful ethical practice, and audience evidence when making claims about reception. Such work can clarify how multimodal choices operate across Instagram genres and communities.

Declarations

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Ethics and Data Availability: This conceptual article analyses no newly collected dataset. All examples are illustrative constructions, and no personal or participant data were used.

Use of Generative AI: Generative AI assisted with structural organization, language editing, and reference formatting; all content was reviewed and verified by the author.

Author Contributions: Conceptualization, analysis, drafting, and revision: A.J.

How to Cite:

Jalilzada, A. (2026). Multimodal discourse analysis of English-language Instagram posts: Caption, image, and platform conventions. *Global Spectrum of Research and Humanities*, 3(4), 272-278. <https://doi.org/10.69760/gsrh.0260304027>

References

- Bateman, J. A., Wildfeuer, J., & Hiippala, T. (2017). *Multimodality: Foundations, research and analysis — A problem-oriented introduction*. De Gruyter. <https://doi.org/10.1515/9783110479898>
- Highfield, T. (2018). Emoji hashtags // hashtag emoji: Of platforms, visual affect, and discursive flexibility. *First Monday*, 23(9). <https://doi.org/10.5210/fm.v23i9.9398>
- Ismayilli, T. M., Mammadova, K. M., Qurban, I. N., & Heyderova, N. A. (2025). Confirmation gestures, mimics and their effects on foreign language learning: Examples of EFL students. *Pegem Journal of Education and Instruction*, 15(2), 198–206. <https://doi.org/10.47750/pegegog.15.02.20>
- Jalilzada, A. (2022). *The place of abbreviation in media discourse*. <https://doi.org/10.5281/zenodo.6560297>
- Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing multimodality*. Routledge. <https://doi.org/10.4324/9781315638027>
- Kress, G., & van Leeuwen, T. (2021). *Reading images: The grammar of visual design* (3rd ed.). Routledge.
- Kruk, J., Lubin, J., Sikka, K., Lin, X., Jurafsky, D., & Divakaran, A. (2019). Integrating text and image: Determining multimodal document intent in Instagram posts. In *Proceedings of EMNLP-IJCNLP 2019* (pp. 4622–4632). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1469>
- Logi, L., & Zappavigna, M. (2023). A social semiotic perspective on emoji: How emoji and language interact to make meaning in digital messages. *New Media & Society*, 25(12). <https://doi.org/10.1177/14614448211032965>
- Page, R. (2012). *Stories and social media: Identities and interaction*. Routledge. <https://doi.org/10.4324/9780203120345>
- Rathnayake, C., & Suthers, D. (2025). Towards a “pluralist” approach for examining structures of interwoven multimodal discourse on social media. *New Media & Society*, 27(2), 1172–1192. <https://doi.org/10.1177/14614448231189800>
- Tenca, E. (2022). Creating brand personality and engaging users in Instagram for marketing purposes: A multimodal discourse analysis. *Iperstoria*, 20. <https://doi.org/10.13136/2281-4582/2022.i20.1208>
- van Leeuwen, T. (2015). Multimodality. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis* (2nd ed.). Wiley Blackwell.
- Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web*. Continuum.

- Zappavigna, M. (2018). *Searchable talk: Hashtags and social media metadiscourse*. Bloomsbury Academic.
- Zappavigna, M., & Martin, J. R. (2018). #Communing affiliation: Social tagging as a resource for aligning around values in social media. *Discourse, Context & Media*, 22, 4–12. <https://doi.org/10.1016/j.dcm.2017.08.001>

Received: 10 September 2026

Accepted: 27 October 2026

Published: 28 October 2026