Emoji in dialogue with images

A multimodal social semiotic approach

Cecilia Lazzeretti
Free University of Bolzano-Bozen

This paper explores the visual nature of emoji and their potential to co-create meaning when in dialogue with other visual semiotic resources, such as images, paintings, videos, and photographs. Drawing on a multimodal social semiotic framework and studies that emphasise the independent meaning-making capabilities of visual communication, it challenges the traditional view that emoji are merely paralinguistic elements dependent on verbal co-texts. The study analyses a corpus of art museum social media posts, highlighting instances of intermodal convergence between the shared images and emoji used in the captions. This analysis reveals formal and semantic dialogic relations, showcasing how emoji can function as independent semiotic resources, capable of establishing meaningful interactions with other visual elements.

Keywords: emoji, multimodal dialogue, intermodal convergence, multimodality, social semiotics

1. Introduction

Many of the current linguistic approaches (Sampietro 2016; Dürscheid et al. 2019; Zappavigna and Logi 2024) consider emoji as dependent on the verbal co-text and minimise their potential to make meaning *per se*, as a full means of representation. The ambiguity of emoji (Czestokowska et al. 2022), which are inherently polysemic and non-univocal in their interpretation, would require a co-text that acts as "a clear verbal anchorage" (Sampietro 2016, 110) to clarify the meaning made by the multimodal text as a whole. Likewise, Dürscheid et al. (2019, 175) treat emoji as a variable linguistic unit "determined by the context or the reader's interpretation of a given text in which it is used". Zappavigna and Logi (2024) go even further, defining emoji as a paralinguistic resource and claiming that even when emoji appear alone in a text, their meaning often depends on "co-occurring

https://doi.org/10.1075/ld.00211.laz | Published online: 29 September 2025
Language and Dialogue ISSN 2210-4119 | E-ISSN 2210-4127 © John Benjamins Publishing Company

language within the broader context of the situation", such as "a preceding linguistic move in an exchange" (Zappavigna and Logi 2024, 8).

The implication for this is that emoji would only make meaning when combined with a more or less explicit textual counterpart. In the absence of a co-text, this would have to be implied, imagined, or, paradoxically, visualised. Furthermore, the view that equates emoji with paralanguage, as understood by Abercrombie (1968), implies a hierarchical relationship between semiotic resources, with words dominating other forms of communication. Multimodalists, however, reject this hierarchy, asserting that visual communication should be taken as seriously as verbal communication (see, among others, Sindoni 2011; 2019).

Adopting a multimodal social semiotic approach that treats images on par with verbal language (Kress and Van Leuween 2006), and drawing on studies such as Maier (2023), who argues for a pictorial account of emoji, this paper aims to refocus on the essentially visual nature of emoji. It also highlights their potential to co-create meaning when in dialogue with other akin semiotic resources, such as images, videos, paintings and photographs. This affordance of emoji is probably less intuitive than their conventional ability to make meaning by paraphrasing verbal resources, as it involves unusual ways of visually expressing ideas, with creative and playful effects for both writers and readers. Therefore, it is worth exploring within the framework of multimodal discourse analysis and exploiting as an engagement strategy in digital communication.

To explore how emoji establish dialogue with other visual resources, the study draws on Linell's (1998, 9) broad definition of dialogue as an interaction that can be enabled both by verbal and non-verbal resources — "through language *or some other symbolic means* (my emphasis)". Not surprisingly, the idea of a dialogic relation instantiated between images in the absence of text has already been codified in the field of design and visual arts, namely in terms of formal or semantic association (Camillini et al. 2022). From a social semiotic perspective, the same association has been described in terms of intermodal convergence, to suggest that different modes can combine to make meaning, as observed in picture books (Painter et al. 2013; Moya-Guijarro 2024) and social media posts (Jovanovic and Van Leeuwen 2018).

The visual arts not only provide methodological inspiration for this study but also instantiate the dataset used. This dataset comprises a corpus of art museum social media posts, which I have analysed in a previous study (Lazzeretti 2023) and further enriched for the current research. The discussion will highlight instances of intermodal convergence between the images shared on social media and the emoji used in the captions in the absence of a co-text. These instances include formal dialogic relations, such as those based on the use of similar colours

or shapes, and semantic dialogic relations, such as those based on the representation of similar entities, concepts, or processes.

This paper is structured as follows. Section 2 discusses the visual nature of emoji, while Section 3 explores how they can establish dialogue with akin semiotic resources. Section 4 outlines the materials and methodology used on the study. Finally, Section 5 presents and analyses specific examples of dialogic relations between emoji and images in the multimodal context of museum social media posts.

2. Arguing for a pictorial account of emoji

Moving away from Barthes' essay *Rhetoric of the Image* (1977) on the meaning of images as always related to, and in a sense, dependent on, a verbal text, Kress and Van Leuween (2006) argue that language and visual communication can create meaning separately and through their own specific forms. According to Barthes, images by themselves are too polysemous, too open to a variety of possible meanings, so that the verbal text is needed to either extend or elaborate their meaning. What Barthes overlooked, according to Kress and Van Leuween (2006, 19), is that "the visual component of a text is an independently organised and structured message, connected with the verbal message, but in no way dependent on it — and similarly the other way round."

If we accept that emoji are equivalent to images and apply Kress and Van Leuween's (2006) tenet that images make meaning *per se*, we can argue for emoji as independent semiotic resources, capable of standing alone without needing a co-text to justify their presence.

Despite acknowledging the visual component of emoji, linguists tend to consider them mostly as symbols: according to Zappavigna and Logi (2024, 6), emoji are "technically characters", working exactly like the other letters, digits, and symbols that the user can enter through the keyboard. They can be treated as an 'iconographetic mode', "encoded as Unicode characters and rendered as small icons" (Zappavigna and Logi 2024, 35–35). On the other hand, semioticians seem to give more weight to their visual appearance: Danesi (2016), for instance, argues for an emoji semantics based first on "denotative sense" — a cloud emoji means first and foremost a cloud — and then on further connotations, involving a creative interpretation of each picture, that are activated "at all levels latently" and by which the same cloud emoji can allude to dullness or boredom (Danesi 2018, 319). This echoes Sargeant's (2019, 152) observation on emoji as a "a random selection of disparate pictures" requiring us to "imbue them with meaning".

Perhaps not surprisingly, one of the most compelling perspectives on the inherently visual nature of emoji emerges in a language study rooted in a philosophical background: Maier (2023, 2) explicitly invites us to consider emoji as pictures, "a mode of signification loosely characterised as based on resemblance between form and content", emphasising that emoji, just "like photographs and drawings", are used "to depict what the world looks like". Maier's (2023) pictorial account allows us to explain any "creative, non-canonical emoji usage", based on the generic pictorial meaning. For example, it can effectively explain why certain emoji have acquired a sexual meaning (Thompson et al. 2018; Weissman 2019), considering their visual appearance, even though the denotative meaning of 🍆 and 🍑 has nothing to do with sex. A peach is just a peach, and an eggplant is just an eggplant. Yet, for the user, the visual recognition of shapes and colours starts earlier than the codification of meaning: these emoji do resemble something else.

In outlining a pictorial semantics for emoji, Maier (2023) distinguishes between "entity/event emoji", which depict the world, and "face/hand emoji", which depict the speaker. This aligns with Zappavigna and Logi's (2024) distinction between ideational emoji, representing physical entities, processes or states, and interpersonal emoji, expressing feelings, emotional states, or judgements. Maier also points out that "pictorial truth, unlike linguistic truth, is a matter of resemblance: a given picture is true of a world if it resembles part of that world" (2023, 6). In their resemblance to the world, emoji cannot be said to be completely true, as they are rather abstract and stylised. However, with some exceptions of emoji that are best analysed as symbols, such as for instance flags, which do not depict countries, but a conventional combination of colours and shapes associated with that particular country, Maier concludes that a substantial portion of emoji uses, such as those depicting animals, plants, objects, activities, hand gestures, people, and facial expressions, are primarily pictorial.

A salient consequence of equating emoji to pictures is that they can be viewed as elementary units of a pictorial discourse and can be combined to create meaning. Their combination, or the dialogue they instantiate when they are juxtaposed together, allows meaning to arise based on what Maier calls a "coherence-driven enrichment mechanism" (2023, 16). While the meaning of a single emoji is quite basic, a deliberate sequence of emoji can convey much more complex ideas, similar to how multiple utterances create meaning in a conversation. For emoji to make sense together, they need to be coherently related and organised into a pictorial narrative. This aligns with Seargeant's (2019) view of emoji as "a picture based writing system", which can be used for short, experimental narratives.

For example, in the mini-story above (Figure 1), emoji are able to create a simple narrative that has no pre-existing textual counterpart. Unlike other examples used by Seargeant (2019) to illustrate 'emojifiction', such as the famous incipit of

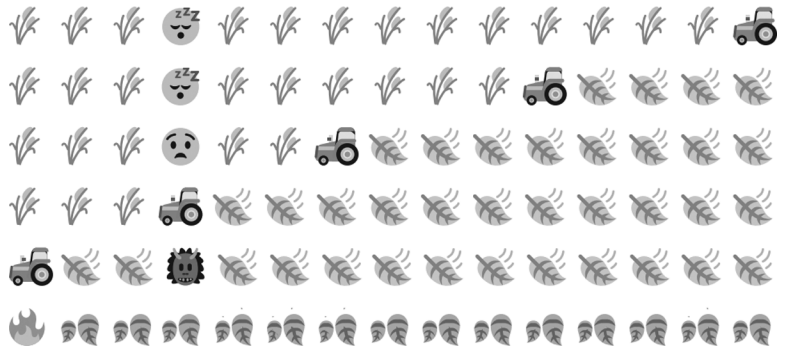


Figure 1. *'Deathly Harvest': A mini-story told in emoji* (Seargeant 2019, 166)

Kafka's Metamorphosis told in emoji, this mini-story has no script. It stands on its own, relying solely on its pictorial nature. Seargeant (2019) notes that the plot flows like a six-line comic strip, with each line describing a further step in the story. It emerges not only by reading each line from left to right, but also by comparing each line from top to bottom. All the basic components of a narrative are present: the setting (a field of corn), the characters (a tractor, a sleeping man) and the events (the tractor mowing the field, a man sleeping hidden in the plants, the tractor and the man colliding, the man dying and turning into a devil). The fire at the end leaves the finale hanging: will the tractor be burned by the devil?

As this mini-story shows, the coherence-driven enrichment mechanism proposed by Maier relies heavily on human interpretation. Users combine emoji based on their visual appearance and then on additional connotations that are activated at various levels, often influenced by subjective interpretation. Consequently, Maier's pictorial semantics implies that the semantics of emoji is user-dependent: emoji resemble their signified concepts but can be imbued with subjective meanings, ultimately determined by the user. This contrasts with the symbolic account, which posits that emoji depend on the linguistic co-text to extend or elaborate their meaning (Sampietro 2016; Dürscheid et al. 2019; Zappavigna and Logi 2024).

A final point to take away from Maier's work is that in considering the idea of deliberate sequences of images that co-create meaning, he also considers 'multimodal mixtures' (2023, 16) in which the communicative content of emoji is set in relation to other visual resources. This brings us to hypothesise that emoji can establish dialogue not only with the verbiage, but also with other images, creating a rich, multimodal tapestry of meaning. Such a perspective potentially broadens the scope of our understanding of emoji, viewing them as integral components in the larger context of visual communication.

3. Extending the notion of dialogue to visual modes

As Weigand (2010, 506) notes, "communication is always dialogic". However, dialogue has mainly been seen as an aspect of verbal, and especially spoken communication. Drawing on Bakhtin's theory of language, which identifies in an oral or written utterance between people "the real unit of speech communion" (Bakhtin 1986, 67), the idea of dialogue has traditionally been associated with human interaction, and in particular, with two people talking to each other. However, as Holquist says, although natural language is a very powerful carrier for dialogue, "it is only one of several ways in which dialogic relationships appear in the larger dialogue which is the event of existence" (Holquist 2003, 42). Similarly, Linell's (1998, 9) broad definition of dialogue as "any dyadic or polyadic interaction between individuals who are mutually co-present to each other and who interact through language *or some other symbolic means* (my emphasis)" seems to envisage a wider conceptualisation of dialogue, not necessarily limited to verbal resources. Both Holquist' and Linell's observations open up the possibility of considering dialogue as an interaction that can also be enabled by non-verbal resources. In particular, Linell's notion of 'symbolic means' evokes a wide range of semiotic resources, including representations, artefacts and images in general, which could be extended to emoji as well.

The idea of a dialogue between images is not new to visual arts and design. In their insightful essay on images in dialogue, Camillini et al. (2022) claim that an association of images is never causal and is both determined by the author's manipulation and the observer's perception. The visual dialogue can be either formal or semantic. The first is more intuitive, as it is realised when two images are placed side by side because of aesthetic similarities (colour, balance, or shape), while the second is based on 'nexuses of meaning' and 'analogy of contexts' between the images (Camillini et al. 2022, 71).

To illustrate the first type of visual dialogue, which is formal, Camillini et al. (2022) reference a work by artists Jason Fulford and Tamara Shopsin titled *This Equals That* (2014). This book features pairs of photographs placed together based on the similarity of their shapes, with no accompanying text. For example, one photograph shows a wooden set square, while the adjacent photograph depicts a large shark-shaped tunnel (Figure 2). As noted by Camillini et al. (2022, 64), "not only do the two photos repeat the same shape, but they are also balanced in terms of colours".

To exemplify the second type of visual dialogue, which is semantic, Camillini et al. (2022) cite Erik Kessels' work, *Attack of the Giant Fingers* (Kessels 2014). This collection features photographs where the photographer's finger accidentally obstructs the camera lens. Despite this common mistake, the association between

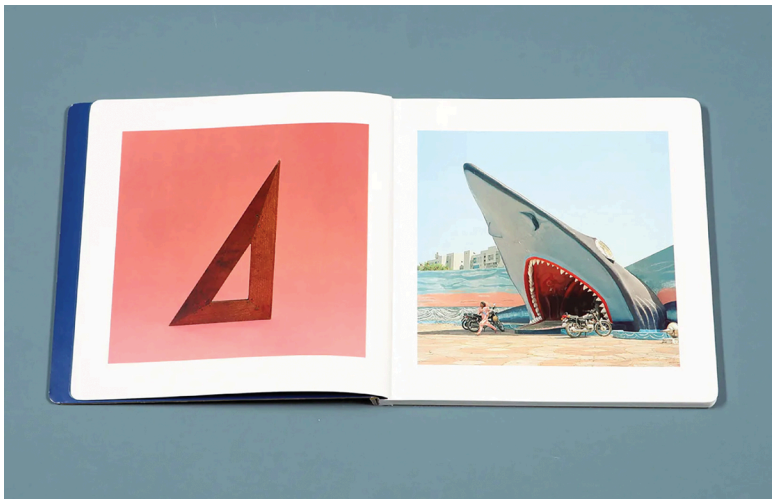


Figure 2. Jason Fulford and Tamara Shopsin (2014), photography, New York: Aperture. Courtesy of the authors and the publisher © Fulford, Shopsin and Aperture

these disparate, apparently unrelated images becomes creative. Page after page, one mistake after another, the finger highlights certain parts of the picture and simultaneously becomes "a graphic and narrative element of the sequence" (Camillini et al. 2022, 69–69).

Camillini et al.'s (2022) framework on dialogic relations established between images in the absence of text can be transposed to multimodal discourse analysis. This requires first acknowledging mode as "a socially shaped and culturally given semiotic resource for making meaning" (Kress 2009, 179) and, consequently, recognising the visual mode, i.e. images, "as a full means of representation", which can form "an alternative to writing" (Kress and Van Leeuwen 2006, 17).

Studies grounded in multimodal discourse analysis (Martin 2011) have shown that different modes work together to co-create meaning, establishing a convergence or divergence of meanings (Painter et al. 2013). A relation of intermodal convergence, in particular, implies an alignment between modes and is realised when modes synergistically complement each other, as observed in picture books (Painter et al. 2013; Moya-Guijarro 2024).

For Linell (1998, 14), complementarity is one of the most salient features of dialogue, as even parties communicating from different positions can achieve "some degree of shared understanding in and through their interaction." This kind of interaction is at the heart of Weigand's (2010, 513) dialogic speech act taxonomy, which distinguishes between initiative speech acts, where speakers make a claim, and reactive speech acts, where speakers are expected to respond to that claim.

If we adopt a social semiotic, multimodal approach to dialogue, in which images are treated as semiotic resources on par with verbal language and therefore able to co-create meaning dialogically, we need to analyse how the same scheme of initiative and reactive speech acts can be realised in a multimodal environment. This is what Jovanovic and Van Leeuwen (2018) have tried to do by observing different types of social media exchanges, which, by design, facilitate the use of visual modes (photographs, illustrations, drawings, infographics and other pictorial forms).

In social media contexts, dialogue can be realised multimodally through initiating and responding moves that may be verbal, visual, or a combination of the two, such as visual-verbal, verbal-visual, or, importantly for our discussion, visual-visual. To illustrate the latter, Jovanovic and Van Leeuwen (2018) present an example of 'emologues', which are dialogic exchanges between users on Instagram, reacting solely through emoji to the image of a puppy: "no written comments, only 'off-the-shelves' emotive responses" (2018, 694). This type of visual-visual interaction exemplifies the focus of Section 5, where the potential of emoji to enter into dialogue with other images will be further explored.

4. Materials and methods

This research is inextricably linked to the disciplines of design and visual arts, which is perhaps unsurprising, considering that emoji have already earned recognition as artworks. In 2016, the Museum of Modern Art (MoMA) in New York acquired the original set of 176 emoji characters designed by Shigetaka Kurita and organised an exhibition to celebrate this new acquisition (Lee 2018).

Assuming that an environment imbued with the visual arts could reveal some untapped potential of emoji in meaning-making, I will use examples taken from a corpus of museum social media posts to illustrate my point about emoji as an essentially visual resource capable of establishing dialogue with other akin resources. The corpus consists of 100 social media posts collected from the Instagram accounts of 17 Anglo-American art museums over a six-month period (5 September 2021 to 18 February 2022). I randomly selected posts that contained at least one emoji and manually annotated them. The choice of museums was based on their active social media presence, as they are among the museums with the highest number of social media followers worldwide according to the Art Newspaper.[1] I have already used this corpus in a previous study (Lazzeretti

1. https://www.theartnewspaper.com/2025/04/01/museums-are-losing-social-media-followers-amid-users-mass-xodus

2023), to demonstrate that emoji have contributed to add an informal component to museum communication, according to the 'post-museum' style (Ravelli 2006), which is characterised by a more direct interactional attitude towards the audience. The original corpus has been further enriched for the purposes of the current research with ad hoc posts that are most likely to illustrate the dialogic relations instantiated between the images shared by museums in social media posts and the emoji used in the captions in the absence of a co-text acting as a verbal anchorage for them.

The study also draws on visual arts from a methodologically point of view. Camillini et al.'s (2022) framework on images in dialogue served as a springboard for the identification of relevant examples from the database and for the subsequent analysis. As outlined in Section 3, formal dialogic relations are those realised between two images sharing similar colours or shapes, while semantic dialogic relations are those realised between images depicting similar entities, concepts, or processes. Both types of relations, however, overlap the concept of intermodal convergence, which originates from studies on how text and images can align to create meaning (Painter and Martin 2012; Painter et al. 2013). This approach builds on Halliday's metafunctional perspective (1978) by exploring how various modes convey ideational, interpersonal, and textual meanings. The concept of intermodal convergence can also be extended to examine relationships between visual resources. In Kress and Van Leeuwen's (2006) visual grammar, Halliday's ideational, interpersonal, and textual metafunctions are respectively named representational meaning (concerned with how the world is represented in multimodal texts), interactive meaning (dealing with how the relationship between the visual, the producer, and the viewer is visually encoded), and compositional meaning (concerned with multimodal texts' inner coherence). In terms of visual grammar, formal dialogic relations pertain to an intermodal convergence of compositional meanings, while semantic dialogic relations pertain to an intermodal convergence of representational meanings. In the discussion that follows the focus will largely remain on these two meanings, touching only very briefly on convergent interactive meanings when these are realised in the observed examples.

5. Discussion

5.1 Convergence of compositional meanings between image and emoji

The first type of intermodal convergence to be examined involves compositional meanings. This refers to formal similarities between the images shared by

museum communicators in social media posts and the emoji used in the post captions. These dialogic relations may be established by analogy of colours and/or shape.



Figure 3. Convergence of compositional meanings

In the Instagram post above (Figure 3), the image chosen — a photograph by Jeffrey Czum — depicts a landscape with mountains in the background, dominated by multiple shades of pink. A pink van, surmounted by a neon sign, features at the bottom of the picture. The post caption acknowledges the artist's credentials and comments on the image content using a flamingo 🦩 emoji and the hashtags #digitalart and #backtowork.

Three different types of intermodal convergences are realised in this post. The first is one of interpersonal meanings and is established between the f* word on the neon sign and the hashtag #backtowork: both elements, visual and textual, converge to express an emotional state of frustration at having to resume work. The second convergence is one of compositional meanings and is enabled by formal similarities (colour) between the image and the emoji. The textual component of the post is not involved in this convergence, as the choice of the flamingo 🦩 emoji is driven by the dominant colour of the image. A third convergence of interactive meanings can be identified in the direction of the van in the picture and the posture of the flamingo 🦩 emoji, as they are facing each other, creating two colliding vectors from left to right and right to left.

The association between the pink colour in the image and the flamingo emoji is arbitrary; one could have chosen any other pink emoji available on the key-

board. This particular choice reflects the degree of creativity involved in selecting emoji according to the post image content and confirms our argument on emoji semantics as user-dependent. In this case, it is the museum communicator who has made the choice and infused their own creativity into the process.



Figure 4. Convergence of compositional meanings

Another case of dialogic formal relations based on colour is shown in Figure 4. While the use of the lemon 🍋 emoji is not surprising here, as lemons are explicitly mentioned in the artwork's title and in the post caption, the use of the yellow heart 💛 emoji at the end of the caption is not so straightforward to interpret. There is no clear counterpart for it in the co-text, neither at the ideational nor the interpersonal level. The idea of love, or friendship, does not seem relevant in this context, as the caption merely describes the painting and emphasises that it is being exhibited for the first time. Likewise, there is no expression of attitudinal meaning in the co-text, except for the adjective 'vibrant'. However, 'vibrant' does not couple at the interpersonal level with the yellow heart 💛 emoji. A more typical choice would be the sparkles ✨ emoji, often used to enhance the positive qualities of energy and light inherent in the adjective 'vibrant'. A sensible explanation is that the yellow heart 💛 emoji is image-driven, as it converges at a compositional level with a prominent colour in the painting.

We have pointed out that dialogic formal relations between images and emoji may also be established by analogy of shapes. In the following two examples, Figure 5 and 6, emoji have been chosen for their formal resemblance to the exhibits, or details of the exhibits, shown in the post images.

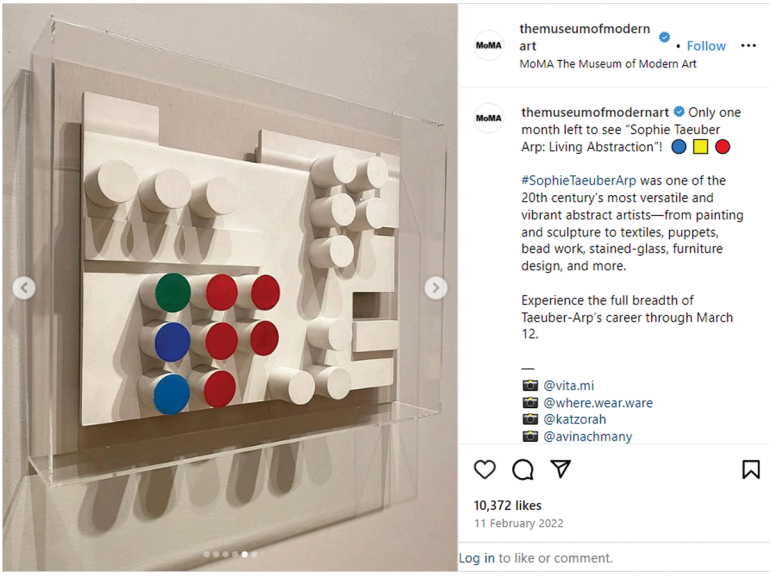


Figure 5. Convergence of compositional meanings

The first post (Figure 5) shows an abstract sculpture, characterised by a blend of vertical and horizontal geometry, with elements of various shapes and colours. The caption identifies the artist as Sophie Taeuber Arp. The emoji chosen to enrich the caption (the blue circle 🔵, the yellow square 🟨, and the red circle 🔴) are inspired not only by some of the primary colours used in the artwork but also by its geometric composition. The convergence of compositional meanings is therefore rather clear.

The analogy becomes more subtle and personal in the second example (Figure 6), where a sculpture showing a human hand seemingly crushing a face, as if it were made of plasticine, is paired with the 'woman facepalming' 🤦‍♀️ emoji. The convergence of compositional meanings arises from a free interpretation of the two visual resources — the sculpture of the crumbling head and the woman pressing her hand against her head look roughly the same — and is based on formal similarities, since nothing in the caption can explain their convergence in ideational or interpersonal terms. As in the case of the pink van associated with a flamingo emoji, creativity is very much part of the picture: it is the creator of the social media post who determines the dialogic relations and manipulates visual resources, according to their own interpretation. Given that the association here involves formal similarities, but also gestures, which in Kress and Van Leuween's visual grammar (2006, 118) are typically associated with interactive meanings, the intermodal convergence achieved in this particular case can be interpreted as one that both involves compositional and interactive meanings.



Figure 6. Convergence of compositional meanings

5.2 Convergence of representational meanings between image and emoji

The second type of intermodal convergence between the social media post image and the emoji used in the post caption to be discussed concerns representational meanings. This semantic dialogic relation is enabled by similarities at the representational level, i.e. images and emoji represent similar entities, concepts or processes. The relation can be straightforward, when there is a plain correspondence between the object depicted in the image and the corresponding emoji e.g., a red book appearing in the image and the red book 📕 emoji in the caption, or based on figurative constructions such as metonymy or metaphor, when the emoji represent a part of what is depicted in the image or symbolically evoke its referent. This latter strategy has already been highlighted by Wicke and Bolognesi (2020) as one of the most frequently used to represent conceptual content through emoji and is seen as a consequence of the fact that sometimes there is no emoji that "literally represents the referent designated by the concept" (2020, 631).

The example below (Figure 7) shows a first example of intermodal convergence of representational meanings between emoji and details of the post image, where the semantic dialogic relation is plain and straightforward. The post contains a video showing details of a large landscape drawing by the artist Joseph E. Yoakum. It is therefore a dynamic rendering of an originally static image, meant to allow Instagram users to zoom in on the work and slowly dwell on different details, one at a time, rather than offering an overall view. The opening frame of the video lingers for a few seconds on the detail of a rising sun, and it is this detail that the author of the post has chosen to call out in the caption of the post, using

the sun ☀️ emoji. The co-text is excluded from the convergence, which takes place only at the visual level, since nothing else in the caption refers to the detail of the sun.

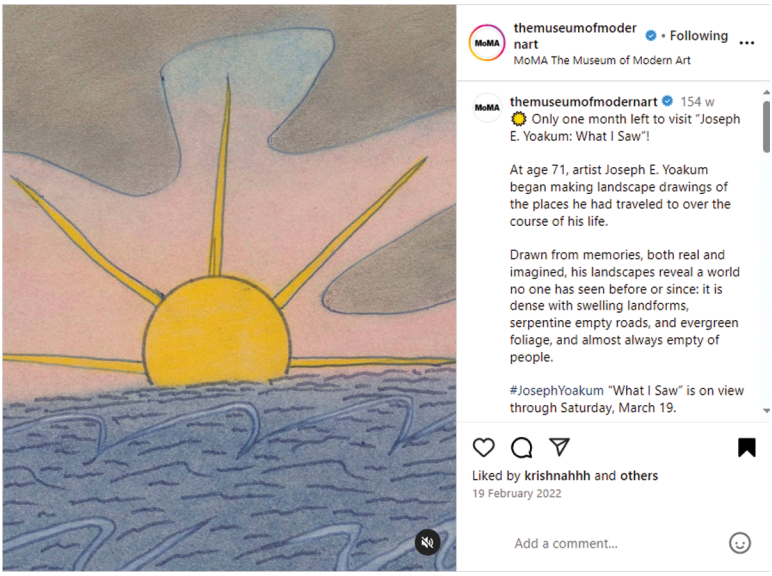


Figure 7. Convergence of representational meanings

Similarly, in the following example (Figure 8), there is a clear convergence of representational meanings between the post image and the emoji in the caption. The entities appearing in the image — a curious combination of a shoe and a cactus resembling a female leg — are reproduced almost verbatim by the cactus 🌵 emoji and the high-heeled shoe 👠 emoji. Furthermore, if we look at the arrangement of the subjects in the image (shoe-cactus) and the reverse order of the emoji (cactus 🌵 — shoe 👠), the image and the emoji seem to mirror each other, potentially creating a convergence of interactive meanings.

The two examples above, which show a clear convergence of representational meanings between images and emoji, appear to be an attempt on the part of the post creator to provide an accurate and faithful translation of the artwork into emoji language. However, the more literal the translation, the less creative the meaning-making process becomes and, consequently, the less engaging for users who enjoy the cognitive effort of detecting the convergence of representational meanings. It becomes more enjoyable and challenging for social media users when post creators provide a personal, non-literal translation of the artwork into emoji language and build creative dialogic semantic relations between images and emoji based on figurative constructions, i.e. metonymy or metaphor. The effec-

tiveness of this cognitive strategy is particularly evident in the following example to be analysed (Figure 9).



Figure 8. Convergence of representational meanings



Figure 9. Convergence of representational meanings

The painting reproduced in the post shows a view of Claverton Street in Pimlico after a snowfall and the caption explains that this is the place where the painter, Charles Ginner, lived for a time. Three emoji are used in the caption: the love letter 💌, the snowflake ❄️, and the admission ticket 🎟️. While the snowflake emoji ❄️ and the ticket emoji 🎟️ are coupled with textual elements in the caption ("snow", "ticket"), the choice of the love letter 💌 emoji would be difficult to justify by simply reading the post, as no textual element is able to create a convergence of meaning with it. Its presence can only be understood by looking more closely at the painting and identifying the detail of an old red mailbox at the bottom. This detail has driven the use of the love letter 💌 emoji, which is metonymically related to the mailbox. There are many details in the painting that could have been chosen to create a salient convergence with emoji: the woman who is sweeping the street, for example, could have been associated with the broom 🧹 emoji, while the columns of the building could have been paired with the classical building 🏛️ emoji. However, the creator of the post chose to draw attention to the mailbox and used the love letter 💌 emoji to create a meaningful association with it, thereby inviting the readers to look closer at the image and enjoy its details.

The same strategy of drawing the viewer's attention to salient details of an image seems to be employed in the Courtauld's post above (Figure 10), which presents a well-known painting by Manet. Unlike the other examples we have presented so far, which were characterised by a minimal presence of text, this is a case of an informative caption that aims to engage users on a deeper level, telling a story rather than simply eliciting an emotional response. However, this may require a fairly long, descriptive text. The museum communicators have therefore chosen three image-driven emoji to use as bullet points to divide the text into three main sections and format the caption's layout in a playful way. Interestingly, three details were extracted from the paintings and translated into emoji: the bowl of oranges on the right, the flowers adorning the barmaid's chest (or, alternatively, the reference could be to those in the glass in front of her), and the bottles of beer, liquor, and champagne to the left and right of the bar. The bowl of oranges is summarised in the tangerine 🍊 emoji, the flowers in the bouquet 💐 emoji, while the various alcoholic beverages on the bar are encapsulated in the beer mug 🍺 emoji. The convergence of representational meanings is almost straightforward in the case of the flowers, while it relies on metonymical relations in the case of the bowl of oranges and the bottles of drinks.

There is no mention of these details in the text, which might have been expected, given that the emoji feature at the beginning of each section, as if to introduce some related content. In terms of layout, the emoji have a purely decorative function here. Their semantic purpose, perhaps less explicit, is a pragmatic

one: they play the role of an attention-directing device, as if they were telling the reader "look at those oranges, look at those flowers, look at those drinks" — but they perform it visually, without saying a word.

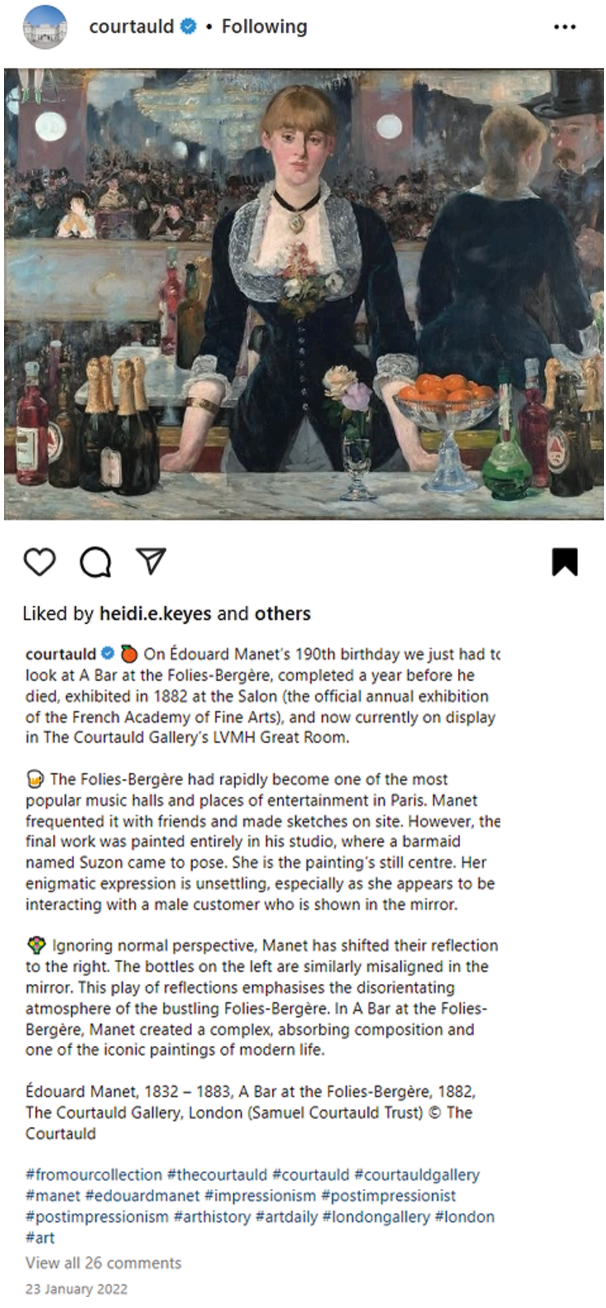


Figure 10. Convergence of representational meanings

6. Conclusions

Most linguistic studies have so far considered emoji as a paralinguistic resource that relies on the co-occurring verbiage to clarify or extend its meaning (Sampietro 2016; Dürscheid et al. 2019; Zappavigna and Logi 2024). However, the examples discussed in this paper show that emoji do not necessarily require a co-text to make meaning. In a multimodal environment such as that of social media posts, they can resonate with other akin visual modes — photographs, illustrations, drawings, paintings, moving images, and all other pictorial forms —, and activate an implicit, reciprocal semiosis, without the support of a textual anchorage to amplify it.

This affordance can be seen as a consequence of the essentially visual nature of emoji. Emoji look roughly like their generic pictorial referent, and, as such, they resemble something to the user. This resemblance to the world, as Maier (2023) points out, is far from perfect, as emoji are rather abstract and stylised. However, it is precisely this approximation that opens up for emoji to a variety of significations, which are ultimately negotiated by the user. As Seargeant (2019) puts it, emoji invite us to imbue them with meaning, drawing on our knowledge, our experience of the world and our interpretation of its images, engaging us in "a contest of forging sense out of the symbols available" (2019, 152). This strengthens the argument of a user-dependent, rather than text-dependent semantics for emoji.

A further observation that follows from the above is that emoji, as pictorial resources, resemble or evoke other images, and are, therefore, inherently dialogic: they call out for other images. By their very nature, then, emoji can enter into dialogue with other images, based on formal or semantic affinities. This idea has already been theorised in the field of visual arts and design (Camillini et al. 2022). From a social semiotic perspective, the same idea can be codified in terms of intermodal convergence (Painter et al. 2013), a concept that involves the alignment of semiotic modes to co-create meaning — typically a visual mode converging with a verbal mode, or vice versa, but also visual modes converging with other visual modes, as in the case of images converging with emoji (Jovanovic and Van Leeuwen 2018).

Examples taken from a dataset comprised of museum social media posts have highlighted instances of intermodal convergence between images and emoji involving both compositional and representational meanings, realised in the absence of a verbal anchorage supporting these convergences or making them more explicit. For example, we have highlighted examples of emoji driven by similarities at the compositional level, such as dominant colours or relevant shapes in the post image: the flamingo 🦩 emoji driven by a pink-dominated post image; a

sequence of circle and square emoji — 🔵 🟨 🔴 — driven by similar geometric elements in the post image. Intermodal convergence between images and emoji can also be realised at the representational level: for example, in the case of a sun ☀️ emoji driven by a rising sun appearing at the beginning of a video embedded in the social media post, or a love letter 💌 emoji metonymically driven by the detail of a mailbox in a painting.

Sometimes the convergence can involve interactive meanings, concerning the viewpoint and position of the viewer, as well as those of the participants represented in the images. This finding was unexpected and somewhat against the odds, given that emoji are stylised and the gamut of gazes, gestures and directions they can embody is very limited. Nevertheless, there were cases where images and emoji seemed to face each other with a mirroring effect, or to replicate similar gestures directed at the viewer e.g., the sculpture of a head crushed by a hand converging with the woman facepalming 🤦‍♀️emoji.

The effects of the making meaning potential of emoji that is dialogically activated by akin visual resources are twofold. First, there is a display of creativity on the part of the creators, resulting in unusual 'translations' of image content into emoji language. This creative process invites the audience to engage with the social media content in a novel and imaginative way. Secondly, there is a cognitive effort on the part of the users who work out these 'translations', resulting in a playful challenge that increases their engagement.

These findings have implications for both multimodal discourse studies and professional communication. For multimodal discourse studies, they broaden the scope of our understanding of emoji, inviting us to consider them as a comprehensive means of representation that extends beyond the boundaries of paralanguage. This perspective acknowledges their potential as a primary resource rather than an ancillary one. For professional communication, these findings suggest novel engagement strategies that exploit the use of emoji in dialogue with images. Such strategies can be employed across various contexts, not limited to the visual arts, to engage users in a playful and meaningful way.

Finally, it is important to note that this study is based on a small and specialised dataset. This limitation calls for an extension of the research to other areas of multimodal communication that involve the use of emoji and images, such as, for instance, instant messaging applications. This extension would provide a more comprehensive understanding of the role and impact of emoji in different communicative contexts.

References

Abercrombie, David. 1968. "Paralanguage." *British Journal of Disorders of Communication* 3(1): 55–59.

Bakhtin, Mikhail M. 1986. *Speech Genres and Other Late Essays*. Austin: University of Texas Press.

Camillini, Gianluca, Marcello Barison, and Roberto Gigliotti. 2022. "Images in Dialogue: How They Talk and What They Say." *Img Journal* 7: 56–73.

Częstochowska, Justyna, Kristina Gligorić, Maxime Peyrard, Yann Mentha, Michał Bień, Andrea Grütter, Anita Auer, Aris Xanthos, and Robert West. 2022. "On the Context-free Ambiguity of Emoji." *Proceedings of the International AAAI Conference on Web and Social Media* 16: 1388–1392.

Danesi, Marcel. 2016. *The Semiotics of Emoji*. London: Bloomsbury Publishing.

Danesi, Marcel. 2018. "Emojis: New Language or Technology-based Trend?" *ETC: A Review of General Semantics* 75(3/4): 317–328.

Dürscheid, Christa, Dimitrios Meletis, and Yannis Haralambous. 2019. "Emojis: A grapholinguistic approach." *Grapholinguistics and Its Applications* 1: 167–183.

Fulford, Jason and Tamara Shopsin. 2014. *This Equals That*. New York: Aperture.

Halliday, Michael Alexander Kirkwood. 1978. *Language as Social Semiotic*. London: Hodder Arnold.

Holquist, Michael. 2003. *Dialogism: Bakhtin and His World*. London: Routledge.

Jovanovic, Danica and Theo Van Leeuwen. 2018. "Multimodal dialogue on social media." *Social Semiotics* 28(5): 683–699.

Kessels, Erik. 2014. *In Almost Every Picture 13 — Attack Of The Giant Fingers*. Amsterdam: Kesselskramer.

Kress, Gunther and Theo Van Leeuwen. 2006. *Reading Images. The Grammar of Visual Design*. Abingdon: Routledge.

Kress, Gunther. 2009. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.

Lazzeretti, Cecilia. 2023. "Exploring the Use of Emoji in Museum Social Network Sites." *Discourse, Context & Media* 53, 100697.

Lee, SooJin. 2018. "Emoji at MoMA: Considering the 'original emoji' as art." *First Monday* 23(9).

Linell, Per. 1998. *Approaching dialogue: Talk, interaction and contexts in dialogical perspectives*. Amsterdam: John Benjamins.

Maier, Emar. 2023. "Emojis as Pictures." *Ergo: An Open Access Journal of Philosophy* 10(12): 317–355.

Martin, James R. 2011. "Multimodal semiotics: Theoretical challenges." In *Semiotic margins: Meaning in multimodalities*, ed. by Shoshana Dreyfus, Maree Stenglin, and Susan Hood, 243–270. London: Bloomsbury Publishing.

Moya-Guijarro, Arsenio Jesús. 2024. "A Multimodal Analysis of Character-character Interaction in LGTB Picture Books and its Educational Implications." *Linguistics and Education* 82: 101312.

Painter, Clare, Jim Martin, and Len Unsworth. 2013. *Reading Visual Narratives: Image Analysis of Children's Picture Books*. Sheffield: Equinox Publishing Ltd.

Ravelli, Louise J. 2006. "Genre and the Museum Exhibition." *Linguistics and the Human Sciences* 2(2): 299–317.

Sampietro, Agnese. 2016. "Exploring the Punctuating Effect of Emoji in Spanish Whatsapp Chats." *Lenguas Modernas* 47: 91–113.

Seargeant, Philip. 2019. *The Emoji Revolution: How Technology is Shaping the Future of Communication*. Cambridge: Cambridge University Press.

Sindoni, Maria Grazia. 2011. *Systemic-functional grammar and multimodal studies: An introduction with text analysis*. Pavia: Ibis.

Sindoni, Maria Grazia. 2019. "Mode-switching in Video-mediated Interaction: Integrating Linguistic Phenomena into Multimodal Transcription Tasks." *Linguistics and Education* 62: 1–11.

Thomson, Samantha, Emily Kluftinger, and Jocelyn Wentland. 2018. "Are You Fluent in Sexual Emoji? 😉: Exploring the Use of Emoji in Romantic and Sexual Contexts." *The Canadian Journal of Human Sexuality* 27(3): 226–234.

Weigand, Edda. 2010. "Language as Dialogue." *Intercultural Pragmatics* 7(3): 505–515.

Weissman, Benjamin. 2019. "Peaches and Eggplants or… Something Else? The Role of Context in Emoji Interpretations." *Proceedings of the Linguistic Society of America* 4(29): 1–6.

Wicke, Philipp and Marianna Bolognesi. 2020. "Emoji-based Semantic Representations for Abstract and Concrete Concepts." *Cognitive Processing* 21(4): 615–635.

Zappavigna, Michele and Lorenzo Logi. 2024. *Emoji and Social Media Paralanguage*. Cambridge: Cambridge University Press.

Address for correspondence

Cecilia Lazzeretti
Faculty of Education
University of Bozen-Bolzano
Regensburger Allee 16 — viale Ratisbona, 16
39042 Brixen — Bressanone, BZ
Italy

Cecilia.Lazzeretti@unibz.it

Biographical notes

Cecilia Lazzeretti, PhD, is a researcher at the University of Bolzano and a former museum press officer and freelance journalist. Her research focuses on the communicative demands of specialised language contexts, particularly in the fields of tourism and museums. She combines different methodologies in her work, including discourse analysis, corpus linguistics and multimodal discourse analysis. Cecilia has contributed extensively to the field of museum discourse, including the 2016 monograph *The Language of Museum Communication* and various articles in international journals.

Publication history

Date received: 21 February 2025
Date accepted: 9 June 2025
Published online: 29 September 2025