

Making Meaning Across Modes: A Review of Theories and Multimodal Metaphor in Media Discourse

PENG Xiaoyu

South-Central Minzu University, Wuhan, China

Multimodal metaphor has become a key topic in cognitive and discourse studies, extending Conceptual Metaphor Theory from purely verbal data to coordinated uses of language, image, sound, and gesture. This article argues that rigorous multimodal metaphor analysis lies on three methodological commitments—explicit identification criteria that distinguish multimodal metaphor, cross-domain mapping operationalization that makes cues, and meaning construction through different modes. In this context, the review further argues that the central explanatory payoff lies at the discourse level, where mappings are orchestrated into coherent evaluative trajectories. By linking criteria, operationalization, and reliability to mapping dynamics and discourse effects, the article outlines a process-oriented framework for future multimodal metaphor research. Overall, the review shows a shift from theory-driven discussion to richer empirical and interdisciplinary work, and argues that multimodal metaphors are powerful semiotic resources for constructing stance, emotion, and ideology in contemporary communication.

Keywords: multimodal metaphor, Conceptual Metaphor Theory, cross-domain mapping; meaning construction

Introduction

Multimodal metaphor has emerged as a central topic in cognitive linguistics and discourse studies, as contemporary communication increasingly relies on complex constellations of language, image, sound, and gesture rather than on verbal texts alone. Building on Conceptual Metaphor Theory, which re-conceptualizes metaphor as systematic cross-domain mapping grounded in embodied experience, recent research has extended the analysis from written and spoken language to a wide range of media discourses, including print and TV advertising, film, news cartoons, emojis, and classroom interaction (Forceville & Urios-Aparisi, 2009; Peng & Yang, 2016; Suarez, 2019; Yang, 2023). This growing body of work demonstrates that multimodal metaphors do not merely ornament messages but play a constitutive role in structuring how abstract issues, social identities and emotions are framed, evaluated, and argued for. Against this background, this review offers a focused review of multimodal metaphor studies in media discourse. It first outlines the conceptual foundations of multimodal metaphor, and then examines how multimodal metaphors are identified in authentic data, how cross-domain mappings are constructed across modes, and how these mappings contribute to meaning construction and persuasion in specific media contexts.

Conceptual Metaphor Theory

While many regard metaphor as merely a linguistic device for rhetorical embellishment, its influence extends far beyond stylistic expression. Metaphors permeate everyday cognition and behavior, playing a pivotal

role in shaping human thought processes and actions. Their significance manifests not only in verbal communication, but also in fundamental cognitive frameworks and practical conduct. Thus, cognitive linguistics was then developed. Lakoff and Johnson (1980) established metaphors as the fundamental framework through which human conceptual systems are structured and comprehended. These cognitive mechanisms remain essential in contemporary society, continuing to shape our perception and interpretation of the world. Their groundbreaking publication, *Metaphors We Live by* (Lakoff & Johnson, 1980), systematically developed the theoretical foundations of conceptual metaphor, revolutionizing scholarly understanding of this phenomenon. This work marked a paradigm shift in metaphor studies, transitioning the field toward cognitive-oriented research and establishing new theoretical trajectories for subsequent investigation. The theory of conceptual metaphors suggests that metaphors form the basis of the human conceptual system, and that concepts are formed because of the experiences that people have gained through continuous concrete practice, which reflect the properties and characteristics of things. Conceptual metaphor is the expression of concepts in a metaphorical way (Lakoff & Johnson, 1980).

The theoretical framework incorporates two fundamental components: source domain and target domain. This cognitive process involves transferring characteristics from the source to the target domain, enabling comprehension of abstract concepts through more tangible representations. Metaphorical understanding occurs when we employ familiar concepts to interpret unfamiliar ones. Lakoff (1993) characterized the source domain as concrete knowledge derived from immediate physical experiences, while the target domain represents abstract concepts developed through mediated interactions with our environment.

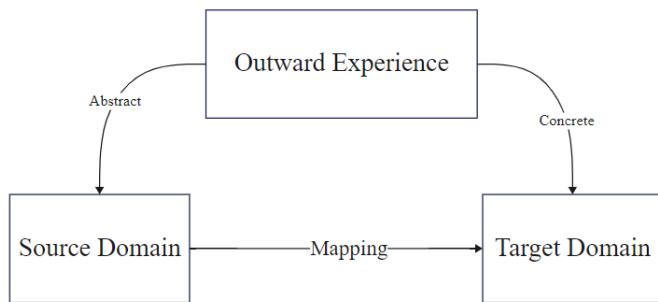


Figure 1. Mapping from source domain to target domain.

As illustrated in Figure 1, the metaphorical transformation occurs via cross-domain mapping where concrete elements from the source domain provide structure to abstract elements in the target domain, simultaneously making the abstract more tangible and the concrete more conceptually rich.

To illustrate this theoretical construct, Lakoff (1993) presented the case of “The only child as a little emperor”. In this metaphorical relationship, the notion of an only child (Target Domain) derives its meaning through association with the concept of a young monarch (Source Domain). Through cognitive projection, selected attributes from Source Domain are transferred to Target Domain, thereby enabling Target Domain to inherit and manifest particular characteristics originally associated with Source Domain.

Multimodal Metaphor

With the advent of social media, contemporary existence has become thoroughly saturated with multimodal communication, fundamentally transforming traditional metaphor theory. This paradigm shift extends beyond

purely verbal expression by incorporating visual, auditory, and symbolic elements into conceptual metaphor frameworks. Multimodal metaphor is a branch of study within Conceptual Metaphor Theory that examines how metaphors are constructed and understood across different modalities of expression, such as language, images, gestures, and sounds (Forceville, 1996). This theory recognizes that metaphors are not limited to verbal expressions, but can be expressed and perceived through a combination of multiple modalities.

The studies of multimodal metaphor not only expand the boundaries of conceptual metaphor theory, but also introduce new dimensions of meaning construction through integrated sensory channels. The concept of multimodal metaphor was developed to account for the interaction between verbal and non-verbal elements in communication. It suggests that metaphors can be created by mapping information from source domain to target domain across various modalities, enhancing the richness and complexity of meaning. For instance, a visual metaphor might use images to represent abstract concepts, while a gestural metaphor could involve expressing ideas through body movements. Similarly, an auditory metaphor might use sounds to convey a message that is typically associated with words. The study of multimodal metaphors is significant as it helps to understand how different communication modalities work together to create meaning and evoke emotions or reactions. It also provides insights into how people process and interpret information from multiple sensory inputs, which is crucial for fields like advertising, film-making, and cultural studies.

Lakoff (1996) provided the first comprehensive and systematic description of image metaphors within the framework of cognitive linguistics, laying a solid foundation for subsequent multimodal research. Since then, the research field has been expanded to the realm of nonverbal and multimodal metaphors. These researches have also broadened the discourse categories from print advertisements to include comics, television advertisements, and movies.

Identification of Multimodal Metaphors

Having clarified how multimodal metaphor is defined at the theoretical level, this section turns to the practical question of how such metaphors can be systematically identified in discourse. Compared with verbal metaphor, where established procedures, such as the MIP/MIPVU (Metaphor Identification Procedure/Metaphor Identification Procedure Vrije Universiteit) offer relatively clear guidelines, multimodal data pose additional challenges: Meanings are distributed across modes, boundaries between metaphor and metonymy are often blurred, and analysts' interpretations can be highly subjective. Recent studies therefore not only proposed concrete identification criteria, but also experimented with ways of operationalizing them in different media and ensuring inter-annotator reliability.

In terms of identification criteria, current methodological discussions converge on a small set of theoretically necessary conditions for claiming a multimodal metaphor. At minimum, analysts must: (a) establish the presence of two distinct conceptual domains; (b) justify a mapping that transfers structure from a relatively concrete source to a more abstract target; and (c) demonstrate that the inferred relation exceeds contiguity-based links typical of metonymy. Because meanings are distributed across modes, the "evidence" for domains and mappings rarely sits in a single sign: Verbal labels may cue the target, while image composition, movement, sound, or gesture may cue the source. Consequently, a defensible identification requires explicit argumentation for what counts as the target, what counts as the source, and which mapped correspondences are licensed by the co-deployed cues, rather than relying on impressionistic readings. In this context, Suarez (2019) worked in-depth with a single animated series, using a cognitive framework to uncover a specific multimodal metaphor "Fusion Is Relationship"

and to trace how it is anchored in culturally motivated conceptual metaphors, such as “Love Is a Whole” and “Love Is Economical Exchange”.

Regarding ways of operationalizing these criteria in different media, the key move is to translate “two domains + mapping” into observable annotation units and mode-specific cues (Ford & Paula, 2023). In static media (e.g., print ads and posters), operationalization often hinges on specifying the visual elements that cue source or target, and on describing how layout, salience, and juxtaposition trigger cross-domain inference. In time-based media (e.g., animation and film), operationalization additionally requires segmentation (frame/shot/scene), time-coding, and tracing how mappings unfold through sequential orchestration—recurring motifs, character actions, editing, soundtrack, and dialogue can each function as partial prompts that cumulatively stabilize a mapping. In other words, “operationalization” is not merely applying a definition, but constructing a procedure that makes cross-modal cues countable, locatable, and comparable across instances within the same medium.

For ensuring inter-annotator reliability, agreement is regarded not as an afterthought but as part of the identification logic: If a criterion cannot be applied consistently, its analytical value is weakened. Reliability-oriented practice therefore emphasizes: (a) an explicit codebook that defines metaphor *vs.* metonymy decision rules, admissible evidence for domains, and minimum requirements for a mapping claim; (b) independent annotation followed by adjudication, with iterative refinement of categories and guidelines; and (c) reporting agreement metrics and documenting where disagreements concentrate (e.g., borderline metonymy-metaphor cases, ambiguous targets, or diffuse source cues) (Ford & Paula, 2023). Ford and Paula foregrounded the procedural dimension, designing multiple rounds of independent annotation to test how consistently different analysts can identify multimodal metaphors and metonymies. Methodologically, combining theoretically informed introspection with more data-driven checks (e.g., systematic cue inventories and repeated coding rounds) reduces individual bias and makes the resulting identifications more replicable across analysts.

Multimodal Metaphor in Media Discourses

Cross-Domain Mapping

Having addressed how multimodal metaphors can be identified in authentic data, their internal structure is needed to be clarified, that is, to the ways in which cross-domain mappings are actually constructed across different modes. Within Conceptual Metaphor Theory, metaphor is defined as a systematic mapping from a relatively concrete source domain onto a more abstract target domain; in multimodal discourse, this mapping is realized not only through verbal expressions, but also through visual design, layout, color, sound, and other semiotic resources.

Across different media, the mapping process is realized through medium-specific resources while preserving the same core mechanism of cue integration. From domain alignment to projection. In static visual artifacts, such as institutional logos, cross-domain mapping tends to be built through the semiotic affordances of graphic form, including shape, size, color, and layout, which provide concrete perceptual prompts that can be aligned with abstract institutional values (Forceville, 1996); the mapping is strengthened when multiple design parameters converge on a coherent source configuration that supports consistent inferences about the target (Sun & Yang, 2016). In lexicalized body metaphors, mapping is foregrounded through culturally sedimented source-domain vocabulary (e.g., “bone”, “blood”, and “tongue”), but the crucial point is that the same embodied source does not guarantee an identical target structure across languages: Mappings are constrained and reshaped by culturally

preferred existences and conventionalized extensions, yielding systematic cross-linguistic divergence in what is projected and what is backgrounded (Wang, 2020). In digitally mediated interaction, emoji-based emotional metaphors similarly rely on embodied source prompts, yet the mapping is materialized through stylized faces and symbols whose graphic schematization and valence patterns guide users toward particular affective interpretations. Here, mapping is stabilized through shared platform conventions that make certain source-to-target projections readily recoverable in context (Yang, 2023). Taken together, these strands indicate that multimodal cross-domain mapping is simultaneously grounded in embodiment and calibrated by culture: Different modes and media do not change the mapping logic itself, but they provide distinct semiotic routes for cueing domains, aligning correspondences, and licensing metaphorical inferences.

Meaning Construction

While cross-domain mapping specifies how source-target correspondences are licensed by multimodal cues, the next step is to explain how these correspondences are orchestrated across a text or interaction to yield coherent discourse-level meaning. What ultimately matters are not only which domains are mapped onto which, but how such mappings are sequenced, blended, and foregrounded to guide interpretation and achieve specific communicative purposes, such as persuasion, critique, or pedagogy. The following studies therefore investigate how multimodal metaphors function as meaning-making devices in documentaries, advertising, news cartoons, film, and language education.

A salient feature of this process is that multimodal metaphors frequently operate as compressed micro-narratives. Instead of mapping isolated attributes, multimodal cues can compress a causal or moral trajectory—problem emergence, intensification, responsibility assignment, and implied remedy—into a tightly packaged configuration (Vermenych, 2021). In time-based discourse, such as documentary or public communication genres, this micro-narrative is often realized through sequential orchestration: Visual evidence (images of scale, harm, or accumulation) may establish experiential “reality”, verbal commentary may stabilize the target and highlight causal relations, and sound design (music, rhythm, and silence) can calibrate affective stance and urgency. Meaning construction, in this sense, is inseparable from temporal accumulation: Mappings are reactivated across scenes and reinforced by recurring motifs, so that the metaphor becomes an interpretive backbone that connects local moments into an overall argument (Zhao & Su, 2010).

In more spatially compact genres, such as advertising and graphic design, meaning construction relies less on temporal buildup and more on compositional guidance (Wang & Fan, 2024). Here, the metaphorical inference is frequently triggered by deliberate juxtaposition, visual salience, and layout-based reading paths that steer attention toward a specific cross-domain alignment. Verbal elements (brand lines, slogans, and captions) often serve as anchors that constrain interpretation, while visual design supplies the inferential leap by presenting a source-domain scenario that is perceptually immediate but conceptually re-targeted (Vermenych, 2021). Meaning is thus constructed through a division of semiotic labor: One mode constrains “what the message is about”, and another mode enables “how the message should be understood”, with the audience’s inferential work completing the link.

In digitally mediated interaction, multimodal metaphor meaning is further shaped by contextual dependence and interpersonal positioning (Yu & Xu, 2023). Emojis and icon-based metaphors rarely stand alone; their metaphorical force is typically activated by co-text, shared platform conventions, and turn-by-turn interactional goals. Meaning construction therefore involves not only ideational content (“What is being conceptualized”), but

also interpersonal functions (how stance, affect, and alignment are performed). The same graphic source prompt can support different target constructions depending on local context, while repeated community use stabilizes certain projections into relatively conventional readings (Awier, 2021). In this setting, multimodal metaphors contribute to meaning by enhancing narrative expressivity, managing evaluation, and coordinating social relations—often compensating for the reduced availability of embodied cues in text-based communication.

Taken together, the meaning construction process of multimodal metaphor can be understood as a discourse-level orchestration of mappings: Modes anchor, distribute, and reinforce interpretive cues; genres regulate whether metaphors unfold sequentially; and audiences assemble these cues into coherent messages that typically include evaluation, causality, and implied action.

Conclusion

Overall, this article supports a process view in which multimodal metaphor operates through a chain of identification, mapping, and discourse-level meaning construction. At the level of mapping and interpretation, the evidence suggests that embodiment provides recurrent source structures, while media affordances and cultural conventions calibrate how projections are cued, aligned, and stabilized. Crucially, multimodal metaphors build meaning not only by aligning domains but by orchestrating mappings across modes and across time or layout to produce evaluative trajectories—whether via sequential “micro-narratives” in time-based discourse, compositional prompting in static designs, or context-dependent stance work in digital interaction. Future research would benefit from cross-linguistic and cross-platform comparisons, and mixed-method designs that combine fine-grained qualitative analysis with replicable annotation and corpus-driven or experimental validation.

References

Awier, M. (2021). Multimodal metaphor in ELT: Combining theoretical knowledge and skills development. *Linguodidactic, XXV*, 7-20.

Forceville, C., & Urios-Aparisi, E. (2009). *Multimodal metaphor*. Berlin: Mouton de Gruyter.

Forceville, C. (1996). *Pictorial metaphor in advertising*. London: Routledge.

Ford, S., & Paula, P. (2023). What counts as a multimodal metaphor and metonymy? Evolution of inter-rater reliability across rounds of annotation. *Language and Cognition, 15*(4), 786-814.

Hong, G., & Zhang, Z. (2010). An analysis of the compositional meaning of the multimodal video discourse “You and Me”. *Technology Enhanced Foreign Language Education, 32*(6), 20-24.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 202-251). Cambridge: Cambridge University Press.

Peng, Y., & Yang, H. Y. (2010). The construal of the integrative meaning of multimodal commercial advertising discourse. *Journal of University of Science and Technology Beijing (Social Sciences Edition), 26*(4), 1-6.

Suarez, G. (2019). Metaphors made live: Multimodal metaphor analysis in animation. In *Perspectives on visual learning* (Vol. 3, pp. 93-102). Budapest: Hungarian Academy of Sciences.

Sun, Y., & Yang, W. T. (2016). A panoramic survey of award-winning LOGO works of GDUFS Center for teacher development from the perspective of multimodal metaphor theory. *Technology Enhanced Foreign Language Education, 38*(6), 22-28+89.

Vermenych, Y. (2021). Multimodal metaphor patterns in documentaries about plastic pollution. *Theory and Practice in Language Studies, 11*(8), 878-883.

Wang, T. M. (2020). A contrastive analysis on the construction modes of English and Chinese body metaphor’s target domain. *Journal of Changchun University of Science and Technology (Social Sciences Edition), 33*(1), 139-142.

Wang, Y. Y., & Fan, Z. Y. (2024). Collaborative construction of meaning by modes: An interpretation of three Episodes of Red Cliff. *Foreign Language and Literature Studies, 41*(4), 41-53+70+134.

Yang, Y. (2023) A study on multimodal metaphor of emotions of emojis. *International Journal of Multilingualism*, 22(3), 1683-1720. doi:10.1080/14790718.2025.2451696

Yu, Q. F., & Xu, D. Z. (2023). Meaning construction of multimodal metaphors in anti-corruption news cartoons from the perspective of conceptual blending theory. *Journal of Mudanjiang University*, 32(10), 65-74.

Zhao, X. F., & Su, H. Y. (2010). The cognitive construction of meaning in multimodal metaphorical discourse: Integration under the interaction of multimodal metonymy and metaphor. *Journal of University of Science and Technology Beijing (Social Sciences Edition)*, 26(4), 18-24+30.