

A Review of Social Semiotics in the Interaction of Text and Image

Sai MA*

Department of English, School of Foreign Languages, North China University of Science and Technology, Tangshan 063000, China

**Corresponding author: 15830581048@163.com*

Abstract: *As the major form of anti-epidemic promotion in network media, anti-epidemic promotion picture applies diversified multi-modal symbol resources in the construction of meaning. By exploring the interactive relationship between visual grammar and conceptual metaphor/metonymy, a joint multimodal research framework is created to clarify the meaning construction and mechanism contained in anti-epidemic pictures. The research shows that conceptual metaphor and conceptual metonymy not only provide cognitive motivation for the realization of visual grammar, but also inspire the associations of images to the real world. Visual Grammar highlights the cognitive meaning expression of conceptual metaphor and metonymy, stimulates readers' interpretation of conceptual metaphor and metonymy. The research explores the combination of two mainstream multimodal research paths, and provides a new perspective for multimodal discourse analysis practice.*

Keywords: *Social Semiotics; Visual-verbal Interaction; Meta Functions*

Internet users are increasingly reading news on internet through screens, and this shift in reading habits has led the discourse analysis community to increasingly cover new media discourse under their researches. New media news is different from traditional paper news in that users no longer passively receive news information but actively select it in the process of reading. The data of users' selection is reflected in the number of news reads and subscriptions, which are the pointer to the popularity of online news. The change in turn produced new genre types on the new media platform through integration of different genres, in which the application of modes and the combination of symbols has been greatly diversified and innovated. The tendency has also brought earth shaking changes to people's reading habits. The design of discourses for our daily reading has shown an inclination toward a visual logic path rather than the original linear one, in which images and texts will be scrutinized in a circuitous routine and an all-around way repetitively. A multimodal shift thus took place in the realm of discourse analysis academia accordingly in order to adapt to this change and fully understand communication

1. Introduction

As the pioneers of modern semiotics, Ferdinand Saussure's and Charles Sanders Peirce's theories have far-reaching influence on the semiotic academia. The former pays attention to how people create meaning with symbols in social behavior. The latter regards symbols as a ternary relationship and divides them into three categories: icon, index and symbol. However, both of their theories have shortcomings. On one hand, they do not consider the relationship between language and paralinguistic; on the other hand, symbols are regarded as a static concept and ignore its specific use process, that is, how the symbol is produced and understood.

2. Systemic Functional Perspective

Barthes pioneered the exploration on relationship between different modes in his master piece *Rhetoric of the Image* in 1977. His discussion and analysis illustrated the interaction between text and image in press photographs and advertisements. Barthes defined relationships between image and text based on their dependency as anchorage (text supporting image), illustration (image supporting text) and relay (image-text co-operation), in which language serves to either elaborate or extend the meaning of the visual text. Barthes theory led to the consideration on relative relations between image and text.^[1]

With different concerns though, multimodal researches after 1990s have achieved an agreement that

both linguistic and the pictorial modalities shared the equal status in meaning construal. Researches regarded as Intrasemiosis focus on other modals rather than language and studied them as closed systems for selection in meaning making process. On the other hand, those deals with the interrelatedness between different semiotic systems and meaning arising from the relationship is deemed as intersemiosis. In the late 1990s, systemic functional semioticians began to inquire into intersemiotic relations. For instance, Kress and van Leeuwen discussed the interrelationship between different modes and proposed that the whole meaning of the multimodal discourse should be treated as interacting and affecting with one another (Kress & van Leeuwen 1996).^[2]

Based on the metafunctional perspective of SFL, Royce (2015) established an page-based intersemiotic complementarity framework covering the intersemiotic ideational, interpersonal and compositional relationships. Lim (2004) proposed two types of contextualizing relations in the meaning arising from intersemiosis: In cases where the meaning of one modality seems to reflect the meaning of the other through some type of convergence, the two resources share co-contextualizing relations. On the other hand, in cases where the meaning of one modality seems to be at odds with or unrelated to the other, their semantic relationship is one that creates divergence or dissonance. In the latter case, the resources share re-contextualizing relations. Also, Lim (2004) proposed an Integrative Multi-Semiotic Model (IMM) for the discussion on meaning-making process through both language and visual images as semiotic resources. As a meta-model, the IMM was established through the assimilation and incorporation of matrixes and frameworks available in the field of multimodal studies from the SFL approach. At a meta-level, the overarching research framework synthesized those advances and initiatives across multiple planes and dimensions for analyzing intersemiosis in a multimodal text.

As proposed by Lim, SFL is applied on multimodal analysis for its proficiency and versatility in the discovery of principles guiding the operation of semiosis other than language, based on that understanding systems and mechanisms in semiotic resources will be revealed. Consequentially, more specific text-types have been taken into account as the intersemiotic research target, such as scientific texts, magazines, TV advertisement, hypertexts, educational texts and newspapers^[3]

3. Tri-functional Perspectives

Halliday (2014) proposed a comprehensive understanding of language as a social semiotic system that fulfills specific purposes within a given social context. His perspective underscores the significance of language as a tool of communication, not just a system of sounds, words, and grammar, but as a medium that carries meaning and functions within a broader social framework.^[4]

Expanding this concept to other modes of communication, scholars have argued that visual mode, similarly, should be considered a completed semiotic system. Visual elements, such as images, colors, layouts, and typography, convey meanings and intentions just as language does. Therefore, they deserve equal attention and treatment in the process of meaning construction and semiotic analysis.

When it comes to multimodal texts, which incorporate both linguistic and visual elements, the application of Systemic Functional Linguistics (SFL) becomes particularly relevant. As a general principle for meaning making, SFL provides a framework to analyze not only the linguistic elements but also the visual mode and their cohesive links within the multimodal text as a unified whole. This approach recognizes that language and visual mode are complementary and interact with each other to create a more complex and nuanced meaning.

By applying SFL to the intersemiotic analysis of visual mode, a deeper understanding can be achieved on how visual elements contribute to the overall meaning of the multimodal text. Patterns, relationships, and functions of visual elements can be identified, which work together with linguistic elements to achieve the intended purpose of the text. This integrated analysis not only enhances our comprehension of the text but also highlights the significance of visual communication in today's multimodal world.

3.1 Ideational Perspectives

Halliday (2014) defines ideational function as a way of representing patterns of experience that enable human beings to build a mental picture of reality, to make sense of what goes on around them and inside them. Experience is a flow of chunked quanta of changes and each quanta is modelled as a figure of doing, happening, and being. Figures can be built up into sequences related by time, cause and the like. Based on the division, ideational metafunction has two components: experiential and logical. According to Royce (2015), the examination of the ways that the visual and verbal modes interact intersemiotically

in ideational terms. To illustrate cohesive links between visual and verbal modal, “reference points” should be derived on the visual side to connect its counterpart—lexical inventories in verbal text such as process, participants and circumstance etc. Those visual reference points are defined as Visual Message Elements (henceforth VMEs). With different semantic properties, VEMs involve four aspects, identification of the presented participants, activities or processes, circumstances and attributes.

Halliday (2014) defines cohesion as the system of textual resources that established cohesive links indicating the semantic relationships in the text. Based on Halliday’s cohesion theory, Royce came up with six ideational cohesive relations of two different modes in a multimodal discourse: Repetition(R) for the repetition of experiential meaning; Synonymy(S) for a similar experiential meaning; Antonymy(A) for an opposite experiential meaning; Hyponymy (H) for the classification of a general class of something and its sub classes; and Meronymy(M) for reference to the whole of something and its constituent parts . The general category of collocation(C) for words that tend to co-occur in various subject areas (Halliday, 2014; Royce, 2015).^[5]

Martinec and Salway (2005) proposed a generalized system of image-text relations for the analysis of different genres of multimodal discourse in which images and texts co-occur. Two relations concerned in the system are Image-text status and their logico-semantics relations, which can be chosen simultaneously in the analysis of multimodal text. The realization of each are also specified both for human analyst and machine. Image-text status can be subdivided into equal and unequal. In equal relation, a whole image is related to the whole text, which means that information conveyed exist in parallel. In unequal relations, image and text jointly play the role of participants of a process and form part of a larger syntagm. When image or text modifies their counterpart, the relation between is deemed as unequal and can be further divided into image subordinates to text or vice versa, where the modifier subordinates the modified. Logic-semantic relation includes expansion and projection, which applies Halliday’s grammar to image-text relations. Equal status was attached more attention by Martinec and Salway for its applicability to new media analysis. It is also pointed out that the system may function to distinguish image-text relations between new and old media.

From a systemic functional semiotic perspective, Unsworth (2014) outlined different meaning types in a range of texts and established a framework based on Halliday’s logico-semantics. The framework mainly concerns for the ideational relation between image and text, which involves expansion and projection. Expansion further involves concurrence, complementarity and enhancement, while projection includes verbal and mental. He stresses that a metalanguage system should be developed for describing the meaning-making resources of images and image/text interaction. In literacy pedagogy, meanings made through images or image-text interaction are crucial for comprehending and composing multimodal texts.

Chan (2011) further developed a tentative framework from Unsworth’s study to describe the contribution on ideational meaning through visual and verbal elements as well as the nature of relations between them. Following Unsworth’s concern on pedagogy, the frame work is developed for students’ reading of multimodal texts and description of inter-semiotic relations. Chan proposed two broad types of ideational image-text interaction as ideational concurrence and ideational complementarity, and their realization through projection and logic-semantic relations of expansion: elaboration, extension and enhancement.^[6]

3.2 Interpersonal Perspectives

Halliday (2014) proposed that the interpersonal metafunction is realized by the clause as exchange, which is organized as an interactive event involving the speaker, or writer and audience. Interpersonal meaning is meanly realized through mood and modality system which is also the focus of interpersonal intersemiotic researches.

Royce (2015) exemplified the visual interpersonal features which is deemed as the first step to the analysis on extant visuals. Interpersonal features consist of viewer address, levels of involvement, power relations and social distance in the visual Mood system, and in terms of the various modality markers in the visual Modality system .

In visual images, information is also exchanged in ways that the visuals address their viewers. However, in visual mode, verbal support is often required to make the nature of speech function clear. A verbal contact address may reinforce or support in offering of goods or service in an advertisement and a questioning facial expression may complemented by a printed verbal question (Kress & van Leeuwen, 1996). To determine which speech function is enacted, presence or absence of visual techniques that

directly address the viewer should be examined. The participant's gaze (and the gesture, if present) demands something from the viewer, demands that the viewer enter into some kind of imaginary relation with him or her (Halliday, 2014). In naturalistic coding orientation, a question is being asked through gaze or facial expressions toward the viewer in which a vector can be drawn from some point of origin to the viewer's face; A command can be made through the gesture of pointed finger which forms a direct vector to the viewer; Offering can also be made by gestures submitting or presenting some object); finally, if there are no vectors, gaze or gestures in the visual frame, then the image is offering information to the viewers.

The level of involvement is realized through the horizontal angle, which is related to the relationship between the two front planes: the front plane of the visual constructor and the front plane of the represented participant. These planes can be arranged in parallel or diverge by forming different angles to each other. Therefore, perspective of visual could be frontal or oblique, which forms a continuous incline according to the divergence angle. Even if the image can have either a frontal or an oblique point of view, there is not strictly an either/or distinction. Obliqueness is a matter of degree, for instance, so long as the vanishing point(s) still fall(s) within the vertical boundaries the image is still recognized as a frontal angle. The frontal angle represents the inclusion between the constructor and vision, while the oblique angle represents the degree of commitment to the subject or represented participants, and the degree of inclusion is expressed in different intensities. Right angle or perpendicular oblique angle, on the other hand, suggest the objectiveness and other than that, power in visuals.

The power relations between the viewers and the represented participants in a visual are encoded in the vertical angle formed between them (Royce, 2015). Kress (2010) analyzed the power relation between viewers and participants (also the image producer) in visuals realized by different camera height as high angle, low angle and eye-level angle. High angle position viewers to look down to the presented participants in the image and suggest that viewers have power or superiority over participants or producer of the image. On the other hand, low angle may force viewers to look up to participant's image and suggest image participant's or producer's power over them. Finally, if the image is at eye level, then equality relation is achieved and there is no power difference involved (Royce, 2015).

The degree of social distance encoded between the represented participant and the viewer is realized by the size of frame. The size of the frame decides how much human body displayed in the image, through different shots, such as close-up, medium and long shot. It is analogical to the distance between people talking face to face and similarly could be intimate, friendly or unknown (Royce, 2015).

Based on Halliday's definition on interpersonal meaning and Visual Grammar, Royce (2015) proposed three types of intersemiotic interpersonal relations as Reinforcement of address, Attitudinal congruence and Attitudinal dissonance. While other scholars tried to discuss the evaluation in interpersonal meaning under the existed ideational framework (Martinec & Salway, 2005), either from the perspective of image-text status or logical semantics.

Attitudes are rarely explored in the intersemiotic researches. Several scholars made the initiative steps to discuss how attitudes interact each other across verbal and visual modes. Based on Martin's text-image relation framework, Mao and Sang (2021) proposed a framework to illustrate the interpersonal image-text relation framework, Appraisal Theory and Visual Grammar. For Mao and Sang, there are four types of image-text status according to the density of interpersonal meaning, as image-text independent (high interpersonal meaning in both modes), image-text complementarity (low interpersonal meaning in both modes), image subordinates to text (only image with high interpersonal meaning) and text subordinates to image (only text with high interpersonal meaning). Mao and Sang's research identified the interpersonal meaning in images based on Kress and van Leeuwen's Visual Grammar, as eye contact, camera angles and distance. On the other hand, evaluative resources in text are also scrutinized so as to decide the status between text and image. However, it is of question to compare the density of interpersonal meaning identified by visual-grammar in image to those identified by appraisal-theory in text.

3.3 Compositional Perspectives

Information value refers to the value that endowed by the placement of visual elements in different zone of the image: left and right, top and bottom, center and margin. In left and right structure, what is placed on the left usually suggested as established or given information, thus possess low informational value. On the contrary, the right is for New and more valuable information. Top and bottom distinguish the quality of ideal or real of the entities or information presented in the visuals. What is presented on

the top is regarded of high value, while the bottom is that of less highly valued. Saliency deals with the assigning of visual weight to visual elements for attracting viewers through factors as placement, size, contrasts and sharpness. Element placed foregrounded is visually more salient than that placed in background. Also, related to the information value, entities placed on right or top of the images also visually “heavier” than those on left or bottom. Larger in Size also give objects saliency for it is easier to attract viewer’s attentions than smaller ones. Contrast in terms of color or shape as well as differences in sharpness can also add visual saliency to objects.

Framing refers to the device that represent the connection or disconnection of elements of image through the usage of framing line or dividing line which signifying whether they belong together or not. Royce (2015) proposed that the compositional features shared similar principles as in visuals and those orthographic compositional or layout techniques are generally referred to as principles of typography.

As for Information value in verbal text, left-right and Top-Bottom orientation also operate in the whole page space and realizes the Given-New and Ideal-Real ordering composition. Verbal saliency can be realized through size, boldness, underling as well as contrast on type and color. The first and most obvious is the use of frame is realized by frame lines under, above and between the typography, and borders around the typography to set it apart from the margins and any visual elements. Empty space also serves the function so as to around mark typographical elements off from both the space of the margins and the gutters (Royce, 2015).^[7]

To analyze the compositional intersemiotic complementarity, the layout of multimodal text should be taken into account since visual and verbal ingredients are not placed randomly on the page but for certain principles and purpose. Compositional elements from both visual and verbal mode are deployed to produce a coherent multimodal text as an integrity according via as Information valuation on the page, Saliency on the page, Degrees of framing of elements on the page, Inter-visual synonymy and the importance of potential reading paths (Royce, 2015).

To consider information valuation from the perspective of intersemiotic complementarity requires the analysis on relative placement of verbal and visual mode as vertical or top/bottom, horizontal or left/right, and the center/margin relation. Saliency is realized through relative sizes of visual and verbal mode. The proportion between both mode indicates the degree of significance of each element in the layout of multimodal text. Also, variations or contrasts in tonal values, type and colors are all to be took into account in the same way they are analyzed in respective mode. The examination of degrees of framing involves the way visual and verbal mode divide them and the empty space remaining on the page.^[8]

Summary

This review delves into the intricate integration of multimodal analysis within the robust framework of systemic functional linguistics, placing a pivotal emphasis on the three core metafunctions: the ideational, interpersonal, and textual. Under the umbrella of the ideational metafunction, multimodal research scrutinizes how a diverse array of semiotic resources, ranging from vivid images to resonating sounds and expressive gestures, come together to convey profound meanings and conceptualize intricate experiences. These modalities offer unique perspectives into the world, adding layers of complexity and nuance to the messages being communicated. The interpersonal metafunction is explored through a lens that highlights how various modalities facilitate social interactions and the formation of meaningful relationships. Here, the role of nonverbal cues such as facial expressions, body language, and tone of voice is emphasized, as they play a crucial part in interpersonal communication, often conveying emotions and intentions that are difficult to articulate through words alone.

The textual metafunction unravels how modalities collaborate seamlessly to create a cohesive textual organization. It examines how different modalities complement and enhance each other, working in harmony to convey a unified and compelling message. This metafunction underscores the importance of coherence and cohesion in multimodal texts, ensuring that the various elements work together effectively to achieve the intended purpose.

The multimodal approach within systemic functional linguistics reveals the intricate complexity of meaning construction in real-world communication. It underscores the interdependence of various modalities and their indispensable contributions to the overall meaning of a text or discourse. By incorporating multimodal elements into their analyses, researchers gain a more comprehensive and nuanced understanding of how meaning is created, shaped, and communicated in both verbal and

nonverbal contexts, providing valuable insights into the dynamics of human communication.

Reference

- [1] Kress, G., & van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. Routledge.
- [2] Royce, T. D. (2015). *Intersemiotic complementarity in legal cartoons: An ideational multimodal analysis*. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, 28, 719–744.
- [3] Lim, F. V. (2004). *Developing an integrative multisemiotic model*. In K. L. O'Halloran (Ed.), *Multimodal discourse analysis* (pp. 220–246). London & New York: Continuum.
- [4] Halliday, M.A.K., & Matthiessen, C.M.I.M. (2014). *Halliday's introduction to functional grammar* (4th ed). Routledge.
- [5] Martinec, R., & Salway, A. (2005). *A system for image-text relations in new (and old) media*. *Visual Communication*, 4, 337-371.
- [6] Unsworth, L., & Cléirigh, C. (2014). *Multimodality and reading: The construction of meaning through image-text interaction*. Routledge.
- [7] Chan, E. (2011). Integrating visual and verbal meaning in multimodal text comprehension: Towards a model of inter-modal relations. In S. Dreyfus, M. Stenglin & S. Hood (Eds.), *Semiotic Margins: Meaning in Multimodalities* (pp. 144-167). New York, NY: Continuum.
- [8] Mao, M.Y. & Sang, H.Y. (2021). A study on image-text relations of new-media discourses from the perspective of interpersonal meaning: taking articles of China Daily's WeChat official account as an example. *Journal of University of Science and Technology Beijing (Social Sciences Edition)*, 37, 147-153.