

Arash Ghazvineh\*

# An inter-semiotic analysis of ideational meaning in text-prompted AI-generated images

<https://doi.org/10.1515/lass-2023-0030>

Received August 28, 2023; accepted November 19, 2023

**Abstract:** This paper explores the inter-semiotic analysis of the ideational meaning in images generated by the text-to-image AI tool, Bing Image Creator. It adopts Kress and Van Leeuwen's Grammar of Visual Design as its theoretical framework as the original grounding of the framework in systemic functional grammar (SFG) ensures a solid theoretical basis for undertaking analyses that involve the incorporation of textual and visual components. The integration of an AI generative model within the analytical framework enables a systematic connection between language and visual representations. This incorporation offers the potential to generate well-regulated pictorial representations that are systematically grounded in controlled textual prompts. This approach introduces a novel avenue for re-examining inter-semiotic processes, leveraging the power of AI technology. The paper argues that visual representations possess unique structural devices that surpass the limitations of verbal or written communication as they readily accommodate larger amounts of information in contrast to the limitations of the linear nature of alphabetic writing. Moreover, this paper extends its contribution by critically evaluating specific aspects of the Grammar of Visual Design.



**Keywords:** inter-semiotic analysis; AI text-to-image generator; systemic functional linguistics; grammar of visual designs

## 1 Introduction

In recent years, artificial intelligence (AI) has emerged as a powerful, transformative force, reshaping every aspect of our lives. From healthcare and education to transportation and communication, AI has insinuated itself into the fabric of society, and its impact is only expected to grow in the coming years. It has become an integral part

---

**\*Corresponding author: Arash Ghazvineh**, Department of Art Research, Faculty of Arts, Tarbiat Modares University, Jalal-e-Al-e-Ahmad Highway, Nasr Bridge, 1411713116, Tehran, Iran, Phone: +98989109827095, E-mail: arash.e.ghazvineh@gmail.com. <https://orcid.org/0000-0003-2434-5786>

 Open Access. © 2023 the author(s), published by De Gruyter on behalf of Soochow University.  This work is licensed under the Creative Commons Attribution 4.0 International License.

of our daily lives, with products such as autocorrect features that prevent spelling errors, smart assistants, social media monitoring tools, healthcare management systems, maps and navigation tools, among numerous other instances.

As AI systems become more sophisticated, they are increasingly being harnessed for creative purposes, such as generating works of art, literature, and music. This rapid expansion of AI into the realm of human creativity necessitates a deeper understanding of the underlying processes and mechanisms that govern the translation of ideas and concepts, especially between humans and AI systems and across different representational modes.

Semiotics, as a discipline concerned with the investigation of signs, symbols, and their utilization and interpretation, has played a crucial role in comprehending the generation and dissemination of meaning across diverse modes of human expression. The advent of AI-powered creative tools, such as OpenAI's DALL-E, Midjourney, and Bing Image Creator, which generate visual representations based on input textual prompts, has heightened the necessity for semiotic examination of the transformative processes by which these tools reshape input data into pictorial depictions. These tools facilitate the inter-semiotic translation, a process that involves the conversion of meaning from one semiotic system (e.g., text) to another (e.g., visual imagery).

The aim of this article is to explore the AI-mediated inter-semiotic translation, focusing specifically on the way distinct images generated by such generative models represent the meanings already captured in the original textual prompts. By conducting a comparative analysis between the visual outputs produced by the AI model Bing Image Creator and their corresponding original textual prompts, our aim is to gain insights into the processes by which meanings are replicated and distributed within the composition of visual representations.

To accomplish this objective, we have adopted a social semiotic perspective and particularly drew on the systemic functional linguistic notion of ideational meta-function. The interdisciplinary nature of SFL has rendered it applicable to the examination of a wide spectrum of phenomena, spanning from films (e.g., Bateman and Schmidt 2012) to discourse and genre analysis (e.g., Rose 2023; Van Leeuwen and Han 2023), as well as visual imagery (e.g., O'Halloran 2008).

As AI continues to advance and permeate every aspect of our lives, it is crucial that semiotics engages with this new frontier to ensure a clear understanding of the implications and consequences of the AI-driven semiotic work. This research serves as a critical first step in that direction, paving the way for more nuanced and informed discussions about the role of semiotics in the age of AI. We hope this would contribute to the broader discourse on the role of semiotics in the rapidly evolving field of AI, and to shed light on the ways in which these advanced technologies are redefining the boundaries of human creativity.

## 2 Semiotic construal of inner/outer world

Following Halliday's SF linguistics, developed in the 1960s and later widely elaborated and extended to a whole body of meaning-making instances, language and other semiotic systems are conceived as systematic reservoirs that individuals utilize to effectively communicate and make meaning within various contexts of social life (see, for example, Halliday 1978; Halliday and Matthiessen 2014). The theory offers a comprehensive framework for understanding the underlying organization of such semiotic resources that enable the semiotic work. Central to this approach is an emphasis on the functional aspect of semiosis.

From a SF perspective, semiotic systems are characterized as socio-historically developed means of communication and as evolved (and, ever-evolving) toolkits of resources to create and express meaning within the community. Halliday recognizes three broad and all-inclusive strands of meaning semiotic resources have evolved to realize, namely *ideational*, *interpersonal* and *textual*. He argues that any specific use of language simultaneously construes these three fundamental domains of meaning, which he calls meta-functions (see, Halliday and Matthiessen 2014: 30–31).

Within this triadic model, the ideational meta-function pertains to the way language represents and construes the external world, including the expression of experiences, things, and processes as well as the internal world of consciousness, e.g., mental processes of thinking and the like (Halliday and Matthiessen 2014). Put differently, it looks at how the external world is represented within the representational system. It plays a crucial role in the way language users make sense of their experiences and effectively communicate their understanding to others (Butt et al. 2003).

The ideational meta-function is further subdivided into two distinct sub-types: the *experiential* and the *logical*. The experiential meta-function is concerned with the portrayal of events, processes, participants, and circumstances, whereas the logical meta-function pertains to the organization and interconnections among such constituents, as well as the expression of logical-semantic relations such as cause and effect, condition, or concession (Egins 2004).

Such meaning capacities derive from the unique organization of the semiotic resources. Halliday distinguishes between language as a repository of abstract semiotic resources (i.e., language as system) and language as a concretized structure emerging from those resources (i.e., as structure), arguing that the meta-functional capacities can be related to specific abstract systems. He posits that the configuration of the realized structures should be traced back to specific systemic choices made from the available underlying resources (see Fawcett 1988; Halliday 1966a, 1966b). Notably, the realized structures are not seen merely as formal arrangements of

elements, but rather as representational structures formed to communicate meanings (Matthiessen 1995). From this vantage point, the capacity to realize meta-functional meaning is believed to be intrinsic to the entire architecture of the representational structures (cf. Martin 1991).

### 3 Ideational representation in semiotic systems

Within Halliday's comprehensive framework, the capacity to represent the world is captured in a number of options or terms, systematically arranged across networks of interrelated systems (Halliday 2013). Each individual system, in this regard, captures a specific facet of the meta-functional meaning and accumulates it within a representational structure (Matthiessen 1995).

One key system in the realization of the ideational meta-function is *transitivity* which plays a significant role in embodying experiential meaning by representing processes, participants, and circumstances (Halliday and Matthiessen 2014). Within this system, the "interaction between, or relations involving things" is reflected in a process-participant configuration (Davidse 2017: 80). Circumstances provide additional context to the configuration. Transitivity encompasses different types of processes in three primary domains. These processes, as relevant to our discussion, include *material*, *mental*, and *relational*. The material category involves observable actions and events in the physical world, while the mental category pertains to conscious cognitive processes like thinking. The relational category deals with the attribution and identification of entities in a given context (Davidse 2017). Additionally, Halliday identifies three secondary processes: *behavioral*, *verbal*, and *existential*. Behavioral processes, such as watching and listening, provide an external perspective on consciousness. Verbal processes occupy an intermediate position between mental and relational processes, involving the transmission of symbolic representations. Existential processes bridge the gap between relational and material processes, facilitating an external perspective on category realization (Davidse 2017).

Each process type is accompanied by specific participant roles and potential circumstances that provide additional context. In material processes, participant roles include *actor*, *goal*, *initiator*, and *range*. *Senser* and *phenomenon* are participant roles associated with mental processes. Relational processes involve *carriers*, *attributes*, *values*, *tokens*, *identifieds*, and *identifiers* (Halliday 1967a, 1967b, 1968, 1985). We will return to these roles in depth when drawing parallels between linguistic and visual representational structures.

The logical meta-function, on the other hand, is related to clause complex and various types of logical-semantic relationships between clauses in language (Fawcett

2000). It allows for the construction of complex meanings by linking various ideas or propositions together (Halliday and Matthiessen 2014). These relationships may be expressed through conjunctions, adverbial phrases, or other grammatical and lexical resources, and they primarily serve to organize and integrate ideational meaning at the level of discourse (Eggs 2004). The combination of experiential and logical resources as a repertoire of systemic choices enables the users of semiotic systems to conceptualize and make sense of the world around them.

## 4 Inter-semiosis: cross-modal meaning translation

Inter-semiotic translation or transmutation, put forth by Roman Jakobson (see Jakobson 1959), as the process of reinterpreting verbal signs through the use of non-verbal sign systems, has played a pivotal role in semiotics, particularly in the study of multimodal texts. This area of investigation has been extensively explored by scholars such as O'Halloran (1999, 2007, 2013), Royce (1998, 2002), and Thibault (2000), that, drawing on Halliday's SFL, have investigated how meaning converges or diverges in texts that utilize a range of semiotic resources, including but not limited to visual, verbal, and gestural elements.

A diverse range of approaches have been employed in the modelling of inter-semiotic dynamics, particularly focusing on the analysis of ideational signification transformations. One such example is the work by Yuen (2004), who frames the semantic expansion of ideational meaning in print advertisements through multi-semiotic interactions of visual and linguistic resources. Accentuating the contextual ties, Yuen (2004) introduces strategies for the generation of ideational meaning across modally diverse semiotic resources. She explores how contextual relations among various semiotic systems create an interpretive space which allows for a cross-investment of meaning. Through the lens of bi-directional investment of meaning across visual and linguistic components, this approach underscores the significance of contextualizing relations in the dynamics of inter-semiotic interactions, thereby enabling the enhancement of textual meaning in visual representational systems and vice versa.

In his work on page-based multimodal texts, Royce (1998) proposes the concept of inter-semiotic complementarity, which posits that in such multimodal entities the visual and verbal modes exhibit a semantic symbiosis that contributes to the creation of a cohesive textual phenomenon (Royce 1998: 26). He draws on the functionally-oriented frameworks of Halliday (Halliday and Matthiessen 2014) and Halliday and Hasan (1976, 1985) to demonstrate that within multi-semiotic texts, the inter-semiotic

relations are predominantly characterized by a type of complementarity which can be addressed using linguistic concepts and analytical techniques commonly utilized in cohesion analysis of language. He identifies a number of metafunctionally-based strategies, including synonymy, antonymy and repetition, through which language and images interact inter-semiotically across different modalities, thereby facilitating a cross-modal interface that fosters the exchange and integration of meta-functional meanings (Royce 1998).

Amongst such scholarly pursuits, *The Grammar of Visual Design* proposed by Kress and Van Leeuwen (2021) stands out as a comprehensive and highly pertinent framework for exploring inter-semiotic and trans-semiotic meaning-making. Initially developed as a social semiotic theory with a focus on visual representations, this framework demonstrates significant potential for the analysis of text-prompted AI-generated images through an inter-semiotic lens. Given its original grounding in the linguistic theory of systemic functional grammar, the framework offers a solid foundation for conducting such analyses that involve the incorporation of both textual and visual elements.

Kress and Van Leeuwen's (2021) framework applies SFL notions to the analysis of images from a meta-functional perspective, aiming to explore their communicative potential in terms of representation, interaction and composition. The framework is based on the idea that Halliday's meta-functions and associated concepts serve as semantic-functional roles rather than rigid formal descriptions, thereby allowing for their extension to the analysis of other representational structures. Consequently, Kress and Van Leeuwen contend that these linguistic concepts can be effectively applied to explore various representational systems such as visual imagery. The primary objective of their research is to shed light on the grammatical structures that underpin the generation of meaning within images, as well as the cohesive integration of diverse elements within the visual compositions (Forceville 1999).

The integration of a generative adversarial network (GAN) model within the analytical framework enables a systematic connection between language and visual representations. This incorporation offers the potential to generate pictorial representations that are systematically grounded in controlled textual prompts. The textual prompts, along with the GAN model, serve as tools to facilitate the analysis as they provide a structured foundation of data. As the resulting images are influenced to a significant extent by the input texts, this analytical approach allows for the anticipation of intended visual meanings, thereby guiding the analysis process. Additionally, by implementing purposeful modifications to the textual prompts, it becomes feasible to examine the corresponding shifts manifested within the generated images. The approach introduces a novel avenue for re-examining the inter-semiotic processes.

Text-to-image GAN models operate through the utilization of generative adversarial networks (GANs) in conjunction with natural language processing (NLP) techniques to generate images based on textual descriptions or prompts. This process entails the integration of two primary components, namely a text encoder and an image generator. The text encoder undertakes the task of transforming textual inputs, such as sentences or captions, into a latent vector representation. This encoding procedure captures the semantic and contextual information embedded within the text, thereby functioning as a guiding mechanism for the subsequent image generator. The image generator, in turn, leverages the encoded text as input and proceeds to generate images that align with the provided textual descriptions. This transformation process involves the conversion of the latent vector into a visual representation, often employing convolutional neural networks (CNNs), to produce images that exhibit a high degree of realism (for a readily available elucidation of GANs, see Creswell et al. 2018; Wang et al. 2017).

## 5 Ideating: the visual versus the linguistic

As mentioned earlier, ideation encompasses the capacity of a representational system through which an external world, lying beyond the given system, is rendered and represented by the system. This capacity is contingent upon both the inherent potentialities and confines of that system as well as the impact of socio-cultural factors on the system's evolution through history (Kress 2010). In this sense, various semiotic systems have evolved, both semiotically and socio-culturally, as distinct representational systems, leading to their divergent manners in capturing the complexities of the external world. Viewed through an ideational lens, such external realities are modelled by the representational system in terms of a number of participants engaging in various processes. The system subsequently brings these participants together in its unique way, resulting in a coherent and meaningful depiction of the external world. Within the framework of SFG, participants are classified according to the roles they adopt in relation to the process in which they are involved. These roles, including actor, goal, senser, amongst others, are typically assumed by nouns and nominal groups which are associated with a verb that represents the process (Halliday and Matthiessen 2014).

Unlike the linguistic syntax, visual representations do not exhibit the same level of demarcation and ease of determination regarding participants. Images, particularly naturalistic ones, can incorporate a multitude of participants, with their identification dependent on the observer's focal point and the emphasis given to specific parts of the image. Kress and Van Leeuwen draw upon insights from art history to propose a reliable definition of visual participants. They argue that

participants in visual representations correspond to visual objects defined by formal attributes such as volumes or masses with distinct weight or gravitational pull (for further exploration of distinguishing visual participants, see Arnheim 1954, 1983). These identified participants subsequently engage in visually depicted processes that meaningfully relate them within the overall composition of the image. Kress and Van Leeuwen recognize two main categories of representational structures based on their process types.

## 5.1 Narrative representations

Kress and Van Leeuwen identify the primary processes within visual representations, beginning with narrative structures, wherein aspects of reality are depicted in terms of unfolding actions or events. Such structures construct the external world in a narrative manner, emphasizing doings and happenings. The key feature characterizing a narrative structure is the presence of a vector which is analogous to a linguistic action verb. Vectors are formed by depicted elements and appear as oblique, often diagonal lines that dynamically link the participants within the visual representation. Narrative structures stand in contrast to conceptual structures, wherein participants are interconnected based on classification, part-whole relationships, or symbolic attributions, facilitated by means of spatial configurations with or without non-vectorial connectors (Kress and Van Leeuwen 2021: 55).

In narrative structures, the participant from which an emanating vector can be discerned is designated the actor. Conversely, the goal refers to the participant to which the vector is directed. Based on the three roles of actor, vector, and goal, Kress and Van Leeuwen delineate the following possibilities (Kress and Van Leeuwen 2021: 58–61): a *transactional process* entails the presence of all three positions. It corresponds to a transitive structure in language, where the subject engages in an action on an object. On the other hand, for a *non-transactional structure*, there are two potential configurations. These include an *event* comprising only a vector and a goal, corresponding to a passive structure in language where only the deed and the object onto which the deed is performed are evident. For instance, the sentence “the man was killed” clearly shows the process and the goal, but there is no mention of the actor. The second possible configuration, simply termed as *non-transactional structure*, involves an actor and a vector, corresponding to linguistic intransitive structures where no object is present, as in the sentences “it rains” or “the wind blows.”

Within the visual system, one can envision a transactional structure when observing a player in a football match striking a ball with their foot. In this particular scenario, the player takes on the role of the actor, and the ball represents the goal,





**Figure 1:** A visual narrative structure generated based on the textual prompt “a football player shooting a ball with their foot”.

while the act of shooting serves as the process. Notably, the vector signifying the process is constituted by the outstretched leg of the player, distinctly directed towards the ball (Figure 1).

In an analogous scenario, the observation of a ball soaring through the sky unaccompanied by any indication of the shooter’s presence gives rise to a non-transactional event. Likewise, an image has the potential to depict a player in the act and posture of shooting, with the ball already discharged and out of the visual frame.

Some transactional structures are *bidirectional*, wherein each participant assumes the role of both actor and goal interchangeably. As an example, in an image portraying a conversation between two individuals, wherein both are depicted as actively participating in the dialogue, each participant can alternate between being the actor and the goal throughout the exchange. Kress and Van Leeuwen employ the term *Interactors* to denote the participants’ dual role within such bidirectional transactional structures (Kress and Van Leeuwen 2021: 62). They note that the English language lacks specific mechanisms to adequately represent these cyclic interactions between participants, asserting the absence of any such a thing as *interactional process* in English. They, nonetheless, argue that clauses containing reflexive pronouns are able to partially establish a cyclic formation. However, in the subsequent analysis section (Section 6.1.3), we will demonstrate that certain English verbs, such as “to negotiate,” “to communicate,” and “to speak with,” can portray a cyclic interactional process in which the participants are actively depicted as transitioning between roles.

Kress and Van Leeuwen introduce a distinctive visual grammatical structure exclusive to participants with visible eyes. They label this structure as *reactional*,



**Figure 2:** Visual reactor and phenomenon generated based on the textual prompt “an image depicting an elderly man looking at a group of children playing”.

which occurs when the vector is defined by an eyeline, that is, the direction of the gaze of one or more participants (Figure 2). In reactional structures, the conventional terms actors and goals are replaced with *reacters* and *phenomena*, respectively, a nomenclature borrowed from Halliday (Halliday 1985). In such cases, the reactor designates the participant engaged in the act of looking, while the phenomenon, according to Kress and Van Leeuwen, can encompass either another participant or an embedded structure, such as a transactional structure that the reactor is directed towards (Kress and Van Leeuwen 2021: 62).

Reactions can also be classified into transactional and non-transaction categories. In the latter case, the phenomenon is conspicuously absent, and as a result, the viewer is prompted to engage in imaginative interpretation of the subject of the participant’s gaze and contemplation. This experiential gap can foster a potent sense of empathy and identification with the participants portrayed in the visual representation (Kress and Van Leeuwen 2021: 63). They can also be unidirectional or bidirectional.

The processes described so far typically occur within a contextual framework, known in SFG framework as *circumstances* (Halliday and Matthiessen 2014). Circumstances, as highlighted by Kress and Van Leeuwen, represent secondary participants that are linked to the primary participants not through vectors, but through other compositional means (Kress and Van Leeuwen 2021: 70). While the omission of participants within circumstances may not impact the proposition conveyed by the narrative structure, their exclusion results in the loss of pertinent information. Kress and Van Leeuwen specifically identify *Setting* as a particular type of circumstance. In Figure 1, the goal and the trees are all elements of the circumstance.

**Table 1:** Linguistic grammatical structures and their visual correspondents.

Linguistic structures	Visual representation	Roles
One-participant material process	Non-transactional actions	Process/actor or goal
Two-participant material process	Transactional actions	Process/actor/goal
Passive linguistic structures	Visual events	Process/goal
Reflexive structures/certain verbs like speaking, interacting, etc.	(Bidirectional) interactions	Process/interactors
Behavioral processes	Non-transactional reactions	Process/reacter/phenomenon
Projective processes (mental/verbal)	Reactions (only a small subset of projections)	Process/senser/phenomenon

Various linguistic representational structures and their corresponding visuals are illustrated in Table 1. These are structures relevant to our discussion. There are some other structures such as conversion that were not relevant to the current discussion.

## 5.2 Conceptual representations

Conceptual representations portray participants in a manner that emphasizes their generalized, stable, and timeless essence. They bear resemblance to the grammatical structures with stative verbs, which express states or conditions rather than actions. Kress and Van Leeuwen recognize three major kinds of conceptual structures: *Classification* structures establish hyponymical (‘kind of’) relations between participants, *analytical* structures establish meronymical (‘part of’) relations between participants, and *symbolic* structures visually define participants in terms of their symbolic identity, significance, or meaning (Kress and Van Leeuwen 2021: 76).

Within Classification structures, participants are linked to one another through a configuration characterized by symmetrical composition rather than a vector-based representation. This connection is facilitated by the specific size and orientation of the participants which are typically depicted in an objective, decontextualized manner. In such representations “[t]he visual order itself produces the relations” (Kress and Van Leeuwen 2021: 77). In these structures, there is “a tendency to equate all elements depicted on the same level in terms of one dimension” (Forceville 1999: 165). The drawing of a set of shoes in Figure 3 represents a visual classification structure.

In Figure 3, the arrangement of the shoes signifies their association as members of a shared class, inviting an interpretation in that light. In this case, the inherent visual order of the shoes gives rise to their relational connections rather than any possible vectorial connections.



**Figure 3:** A visual classification representation generated based on the textual prompt “a drawing of a set of shoes”.

Conversely, within analytical structures, participants are depicted as parts of a whole, with the whole assuming the role of the *carrier* that accommodates any number of *possessive attributes* (parts). Analytical structures lack vectorial representation of narrative representations, and they do not exhibit the compositional symmetry characteristic of classification visuals. In essence, they represent visuals as depictions of the current state of specific objects, capturing their essential features. They are “the default, ‘unmarked’ category” which do what visuals do best, presenting a visual ‘this is’ (Kress and Van Leeuwen 2021: 93). The image of a man donning a formal suit can exemplify an analytical structure, where the man himself acts as the carrier, and different pieces of his clothing, such as the coat, shoes, and trousers, serve as the possessive attributes.

Finally, Symbolic structures are visual representations that involve participants that are meant to be read symbolically due to their outstanding characteristics within the overall composition. They facilitate the communication of deeper meanings and invite viewers to engage in symbolic interpretation. Kress and Van Leeuwen (2021: 102–105) delineate two prominent categories of such structures, widely prevalent in contemporary visual culture and art history: *symbolic attributive* structures and *symbolic suggestive* structures.

In symbolic attributive structures, two participants interact to confer a deeper connotative significance on the representation. In this case, the carrier assumes the role of the participant bearing a symbolic attribute that holds the symbolic meaning. Consequently, the carrier is imbued with the symbolic significance associated with the symbolic attribute. On the other hand, symbolic suggestive compositions involve a single participant subject to symbolic interpretation. In such structures, the atmosphere created within the visual composition plays a crucial role in inviting specific symbolic interpretations. In Figure 4, the apple serves as a symbolic attribute with the potential to evoke the symbolic narrative of the fall of humankind in the Garden of Eden. Consequently, the figure depicted holding the apple gains an



**Figure 4:** A symbolic visual representation generated based on the textual prompt “Jesus with an apple in his hand, standing within a heavenly silhouette. Surround his head with a radiant hue of light, evoking a divine aura”.

association with the persona of Christ. The presence of the apple imbues the composition with deeper symbolic meaning. Even in the absence of the apple, the distinct hue, color, and atmospheric qualities it imparts to the scene persist, thereby potentially alluding to the figure’s sacred identity as a holy individual, most probably Jesus Christ.

In SFG, the fundamental conceptual representations are realized through *relational* and *existential* processes, as identified by Halliday (Halliday and Matthiessen 2014). These processes depict the world in terms of relatively enduring states of affairs, rather than focusing on actions or mental processes. Halliday recognizes two primary categories of relational processes: *attributive* processes and *identifying* processes (see Halliday and Matthiessen 2014: 259–299).

Attributive processes involve designating something as an attribute of a particular carrier, such as the word “sad” in the sentence “she is sad.” These processes can be *intensive*, indicating the nature or quality of the carrier as in the previous example, *circumstantial*, denoting when, where, or with what the carrier is (e.g., “Jack is here” where “here” is a circumstantial attribute of Jack), or *possessive*, signifying what the carrier possesses, for example, “She owns a bag.”

Identifying relational processes are structured as “A is the identity of B,” with “A” being the *value* and “B” the *token*. In “She is the president,” “she” is the token with the value of being a president. The value serves to identify the status or function or meaning of the token.

Existential clauses, on the other hand, simply “represent that something exists or happens” (Halliday and Matthiessen 2014: 307). Such structures consist of only one participant termed *existent* whose existence the clause reaffirms. Existents can be

**Table 2:** Relative correspondence between conceptual structures in language and visuals.

Linguistic structures	Visual representations	Participant roles
Intensive/possessive clauses	Classificational/analytical	Carrier/attributive or possessive process/attribute
Identifying clauses	Symbolic attributive	Token/identifying process/value
Existential clauses	Symbolic suggestive	Existent (events/entities)

*events*, as in “It is a good day” or *entities*, as in “There is a car in the street.” As Halliday notes, the presence of a dummy subject like “there” or referenceless “it” is indicative of an existential clause (Halliday and Matthiessen 2014: 308).

According to Kress and Van Leeuwen, visual classificational and analytical structures perform functions similar to intensive and possessive attributive clauses, and there may also be some resemblance between symbolic attributive structures and identifying clauses, as well as between symbolic suggestive structures and existential clauses (Kress and Van Leeuwen 2021: 108). However, the dissimilarities between these linguistic and visual structures are substantial. Visual representations offer a range of structural devices that lack equivalence in verbal or written communication, allowing for the integration of numerous participants within a single structure, e.g., through tree structures, timelines, networks, etc. This is due to the linear nature of alphabetic writing, which limits the capacity to process linear information chunks, while visual representation can accommodate significantly larger amounts of information that can be readily perceived at a glance. Table 2 displays the relative correspondence between visual conceptual representations and their linguistic counterparts.

## 6 Analysis

Having established a robust theoretical framework, we now turn our attention to its practical application. Specifically, we delve into the analysis of a select set of AI-generated images, which are created in response to thoroughly-prepared textual prompts. We first start with the narrative structures representing material processes.

### 6.1 Narrative structures

#### 6.1.1 Non-transactional/transactional actions

An instance of one-participant material process is exemplified by the sentence “The man jumped up,” wherein the phrasal verb “jumped up” entails a Process of change



**Figure 5:** Visual rendition of the sentence “the man jumped up.”

in the outside world and designates “the man” as the actor who brought about the change. In Figure 5, this process is signified by the established sense of directionality inherent in the specific bodily depiction of the participant, wherein his limbs, both legs and arms, are extended in an upward orientation, establishing a vector. This figure represents a non-transactional visual representation.

The verb “jump” possesses an additional capacity to portray an amount of transformation within the external world that is specifically directed at or extended towards an entity distinct from the primary actor and thus represent a *doing* rather than a *happening*. Consequently, when employed in a transitive form, as in the sentence “the man jumped on the cat” the verb demonstrates the potential to capture an additional aspect of the world beyond the process and the actor. In this case, the final structure will be, in Kress and Van Leeuwen’s terms, a transactional configuration which also entails a goal as the participant that undergoes the process. Figure 6 shows an AI-generated visual representation of the second sentence.

Figure 6 encompasses two primary participants, i.e., the man and the cat, whose prominence within the image’s composition is accentuated by their relative sizes and placement in the foreground. An explicit connection between the two can be readily established, facilitated by a discernible vector originating from the outstretched hand of the man, assuming the role of the actor, and extending toward the cat, representing the goal. This configuration exemplifies a transactional representation. However, within the depicted image, a second narrative structure becomes apparent, distinct from the transactional framework. This second structure is characterized by the presence of an additional vector, formed by the direction of the cat’s paws, indicating its attempts to evade the man. It portrays a non-transactional





**Figure 6:** Visual representation of the sentence “the man jumped on the cat.”

representation in which the cat takes on the role of an actor involving a process of evasion. The visual representation, therefore, would be closer in meaning to the linguistic structure “the man jumped on the cat and the cat tried to evade the man.” This underscores the intrinsic potential of images to convey a diverse range of meanings in a single glance, emphasizing that the interpretation of the image depends on the observer’s intentional direction of visual attention.

### 6.1.2 Visual events

Within visual events that bear resemblance to passive linguistic structures, two primary participants, namely the goal and the process, function as the capsules of information about the outside world. These structural formations enable ideation and model the experience of change by presenting a framework that exclusively captures the unfolding of an event and the participant upon which the event transpires. The sentence “the basketball was dunked into a hoop” exemplifies the representational capacity of such structures. This sentence conceptualizes the event in which a basketball, as the goal, undergoes the process of being dunked into a hoop. Figure 7 visually translates such ideational meanings using the representational capacities of the visual system.

In this depiction, the process is once again indicated by a theoretical vector aligning with the trajectory of the basketball’s movement. This vector passes through the hoop, signifying a vertical movement from top to bottom. The presence of the hoop as a circumstantial participant allows for the vertical motion to be transducted





**Figure 7:** Visual representation of “the basketball was dunked into a hoop”.

into the linguistic verb “to dunk.” In this structure, similar to passive sentences, there is no mention of what Halliday (Halliday and Matthiessen 2014: 224) characterizes as the “input energy,” i.e., the actor, that brings about the process.

### 6.1.3 Interactions

Within bidirectional transactional actions, or interactions, the participants involved alternately switch roles within a single process of interaction, such as a conversation. As mentioned earlier, Kress and Van Leeuwen (2021) assert that the English language lacks the linguistic resources to effectively represent these cyclic formations. They argue that English lacks verbs that accurately capture the essence of this aspect of reality, which can be visually conveyed through an interactional process. Furthermore, they contend that sentences such as “A and B communicate with each other,” which may seem to depict interaction similar to visual images, actually result in the participants losing their distinct identities. This transformation, in their opinion, leads to a jointly authored non-transactional action rather than a reciprocal, bidirectional transaction (Kress and Van Leeuwen 2021: 74).

However, we present an alternative perspective, positing that certain verbs in English do capture the essence of an interaction, implying a continual exchange of roles between the participants. For instance, in sentences like “A and B communicated with each other” or “A and B talked to each other,” it can be observed that the verb necessitates a shift of roles to establish the meaning of communication or conversation between the participants. If persistence in roles is maintained, the verb’s intended meaning would be altered, resulting in “talked to” rather than “talk with.”



**Figure 8:** AI-generated visual representation of “Jack and Mary communicate together”.

To explore this further, we used the AI image generating model to analyze the sentence “A and B communicate with each other,” which is cited in the Grammar of Visual Design to determine if this linguistic structure effectively captures the essence of an interaction. We replaced A and B with the proper nouns Jack and Mary. Figure 8 illustrates a visual representation of the sentence.

The presented image illustrates the active participation of both individuals in a shared process, i.e., in a conversational situation. Their active presence in the conversation is visually represented by their facing each other and the hypothetical vectors originating from their hands, directed towards one another, connecting and linking each participant to the other. This conversational engagement is further reinforced through a bidirectional reaction, which emerges from the eyelines of both participants. Within such structures, as described by Kress and Van Leeuwen (2021: 71), “[t]he eyelines connect two interactors who look at each other so that both combine the role of reactor and phenomenon.” It is important to distinguish between the two narrative structures evident in this context. Firstly, there exists a bidirectional transactional process, as indicated by the depiction of both individuals communicating with each other. Secondly, a bidirectional reaction is also formed through the eyelines of the participants. It should be noted that bidirectional reactions, by themselves, do not imply active engagement in a conversation, but rather solely focus on the act of visual interaction, encompassing the act of looking and being looked at.

#### 6.1.4 Reactions

The category of projective processes, which constitutes the final category of narrative structures pertinent to our current discussion, encompasses mental and verbal

processes. Mental processes encompass a range of cognitive activities, including perception processes such as seeing and hearing, affective experiences and emotions, and cognitive activities such as thinking and knowing. Evidently, these non-visual phenomena elude direct representation within visual systems of depiction, due to the inherent limitations in the material and semiotic capacities of these systems, which have not yet sufficiently evolved to encapsulate this facet of the inner cognitive domain. However, transactional reactions possess the capacity to partially depict a restricted subset of perceptual processes. According to Kress and Van Leeuwen (2021: 62), these visual configurations are commonly associated with the process of thinking. Such representations gain deeper ideational meaning regarding the inner world of the thinker through the semantic potential of particular facial expressions and gestures. In Figure 9, the sentence “the man is thinking” has been



**Figure 9:** The sentence “the man is thinking” rendered visually.

visually rendered. To further explore the visual representation of thinking, we employed analogous prompts centered on this cognitive process, each instance of which portrayed thinking through a gaze directed at an unspecified and unrepresented object, resembling the conventions observed in non-transactional reactions. Moreover, this portrayal of thinking consistently included a hand supporting the chin of the depicted individual, accompanied by a facial expression characterized by a pensive look.

Visual semiotic systems have also developed certain indirect ways of depicting verbal processes, such as the utilization of *dialogue balloons* as a means of conveying linguistic expressions in the form of direct quotations. Additionally, compositional components can be used to enable direct depiction of written pieces, thereby providing a visual incorporation of textual content.

## 6.2 Conceptual representations

### 6.2.1 Classificational/analytical structures

Let us now proceed to the representation of conceptual processes using the inherent capacities of visual systems. Analogous to linguistic structures, visual modes of representation have evolved to capture aspects of the world that pertain more to existence and state rather than action. In other words, they can conceptualize the experience in terms of *being* rather than *doing*. In fact, representations by visual systems essentially frame corners of reality, first and foremost, in terms of being, wherein a measure of action may or may not be present. While the intensive sentence “Mary is happy” merely states Mary’s state of happiness without providing further details, a visual representation of Mary being happy inherently involves depicting her within a specific context of other visual elements. Moreover, the visual representation employs a diverse array of semiotic resources, such as composition, texture, color, etc., at its disposal to establish the notion of happiness, as exemplified in Figure 10.

An important observation is that within these structures, the compositional configuration represents happiness within a broader context that encompasses a range of other possible meanings. These additional meanings emerge as a result of the intrinsic structural aspects of the image, such as the very nature of the frame, which have evolved over time and acquired cultural meanings. Additionally, the compositional elements employed in these visual structures do not function in isolation to exclusively realize specific concepts. Instead, it is their cumulative effect that engenders the potential for such interpretations. In essence, a distinguishing feature of visual representations lies in their ability to incorporate a collection of



**Figure 10:** Visual rendition of “Mary is happy”.

elements within a singular frame, wherein these elements interact as interconnected networks of signs to convey complex and layered meanings.

The material on which visual representational systems realize their semiotic work both supports and, at the same time, constrains certain interpretations. That arises, in part, from the fact that different material substrates afford distinct manipulations, yielding diverse possibilities for carving out potential signs. Additionally, the semiotic efforts invested by a community in a particular mode bestow it with certain capacity for signification. For instance, visual systems exhibit the ability to spatially disperse their components throughout their composition, enabling the formation of diverse spatial relations across various directions. Importantly, communities, driven by cultural factors, attribute semantic significance to these spatial relations. In contrast, language system organizes its constituents along a sequential structure and thus solely permits linear relations.

### 6.2.2 Symbolic structures

Within the scope of this analysis, we will now turn our attention to the symbolic attributive and suggestive structures, which bear resemblance to the identifying and existential structures as outlined by Kress and Van Leeuwen (2021: 109–110). Within these structures, the representational systems employ their internal resources to portray the participant’s existence or status in the external world. Symbolic attributive structures employ discernible visual cues or symbols to depict the value associated with the token. This is exemplified in Figure 11, where it visually renders



**Figure 11:** Visual representation of “Jack is the manager of the company”.

the identifying clause “Jack is the manager of the company.” In this figure, the semantic features inherent to the value “manager” are visually represented as symbolic attributes belonging to the token. These attributes collectively function to bestow upon the token the attribute of managerial status.

In this figure, an array of circumstantial elements, such as papers, desks, and the whiteboard in the background, collaborate with the individual’s attributes – his suit, tie, wristwatch, tablet, and the like – that typify a managerial role. It is noteworthy to highlight that the symbolic attributive structure is not the sole configuration semiotically operating within the composition of this image. An alternative analytical configuration is discernible, wherein the individual undertakes the role of a carrier of attributes. This particular configuration exclusively portrays the attributes as possessions of the individual, and can be transduced into a possessive construction, for instance: “Jack possesses a suit, a tie, a wristwatch ...” Additionally, the eyeline of the individual forms a non-transactional reaction. Each of these configurations captures a distinct facet of reality and translates it into a visual format.

In addition to the aforementioned structures, the image portrays an alternative structure that imparts a symbolic significance. Within this image, the man assumes a prominent position in the foreground and occupies the central area, while the surrounding background elements are blurred. Furthermore, the portrayal of the man from a low angle accentuates his authoritative stature, complemented by his confident body language and firm stance. These compositional elements collectively evoke a sense of power and authority associated with the individual. This configuration corresponds to a symbolic suggestive structure, which signifies the inherent power and influence





**Figure 12:** Visual rendition of “a tired manager of a company with a suit and tie and wristwatch worn casually with a tired face in his work place”.

embodied by the character. It is important to note that while the circumstantial elements and other structures also contribute to the establishment of such meaning, as they are integral components of the same composition and mutually influence one another, the specific meaning attributed to the symbolic suggestive structure remains exclusive to its framework. To illustrate, one can consider the subsequent image (Figure 12) as a counterpoint to the aforementioned structure. In this image, while many details bear resemblance to the previous one and thus maintain several structures, the absence of the symbolic suggestive structure is notable, as it fails to convey a sense of authority.

The aforementioned example, along with previous instances, demonstrates that visual representational systems inherently differ in their capacities for representation from language, leading to their distinct mechanisms for capturing specific nuances of ideational meaning. Consequently, when transitioning to the visual domain, visual semiotic resources present diverse spatial and compositional possibilities for conceptualizing the world outside the representational system. Furthermore, the examples highlight the applicability of the proposed generative AI framework in effectively addressing these differences and enhancing the modeling of the inter-semiosis process.

## 7 Conclusions

In conclusion, this paper has provided a comprehensive exploration of the inter-semiotic analysis of ideational meaning in images generated by Bing Image Creator, a

text-to-image AI tool. By adopting Kress and Van Leeuwen's Grammar of Visual Design as the theoretical framework, the study has established a solid foundation rooted in systemic functional grammar, ensuring a robust basis for analyzing the integration of textual and visual elements. The integration of an AI generative model within the analytical framework has facilitated a systematic connection between language and visual representations. This integration has opened up new possibilities for generating well-regulated pictorial representations that are grounded in controlled textual prompts. The utilization of AI technology has proven to be a powerful tool in this process, offering regulated and controlled visual data for analysis.

The study has illuminated the distinctive structural mechanisms inherent in visual representations, surpassing the constraints of linguistic or written modes of communication. Visual representations possess a remarkable capacity to accommodate substantial volumes of information, in contrast to the linear nature of representation in language. The study has also provided insights into the relationship between language and visuals, expanding our understanding of inter-semiotic processes and offering avenues for future research in this field.

## References

- Arnheim, Rudolf. 1954. *Art and visual perception: A psychology of the creative eye*. Berkeley: University of California Press.
- Arnheim, Rudolf. 1983. *The power of the center: A study of composition in the visual arts*. Berkeley: University of California Press.
- Bateman, John & Karl-Heinrich Schmidt. 2012. *Multimodal film analysis: How films mean*. London: Routledge.
- Butt, David, Rhondda Fahey, Susan Feez, Sue Spinks & Colin Yallop. 2003. *Using functional grammar: An explorer's guide*. Sydney: National Centre for English Language Teaching and Research, Macquarie University.
- Creswell, Antonia, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta & Bharath, Anil A. 2018. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine* 35(1). 53–65.
- Davidse, Kristin. 2017. Systemic functional linguistics and the clause: The experiential metafunction. In Tom Bartlett & Gerard O'Grady (ed.), *The Routledge handbook of systemic functional linguistics*, 79–95. London: Routledge.
- Eggs, Suzanne. 2004. *An introduction to systemic functional linguistics*, 2nd edn. London: Continuum.
- Fawcett, Robin P. 1988. What makes a 'good' system network good? In James D. Benson & William S. Greaves (eds.), *Systemic functional approaches to discourse*, vol. XXVI, 1–28. Norwood: Ablex.
- Fawcett, Robin P. 2000. *A theory of syntax for systemic functional linguistics*. Amsterdam: John Benjamins.
- Forceville, Charles. 1999. Educating the eye? Kress and Van Leeuwen's reading images: The grammar of visual design (1996). *Language and Literature* 8(2). 163–178.



- Halliday, Michael A. K. 1966a. Lexis as a linguistic level. In Charles Ernest Bazell, John Cunnison Catford, Michael A. K. Halliday & Robert Henry Robins (eds.), *In memory of J.R. Firth*, 148–162. London: Longman.
- Halliday, Michael A. K. 1966b. The concept of rank: a reply. *Journal of Linguistics* 2(1). 110–118.
- Halliday, Michael A. K. 1967a. Notes on transitivity and theme in English 1. *Journal of Linguistics* 3(1). 37–81.
- Halliday, Michael A. K. 1967b. Notes on transitivity and theme in English 2. *Journal of Linguistics* 3(2). 199–244.
- Halliday, Michael A. K. 1968. Notes on transitivity and theme in English 3. *Journal of Linguistics* 4(2). 179–215.
- Halliday, Michael A. K. 1978. *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.
- Halliday, Michael A. K. 1985. *An introduction to functional grammar*. London: Arnold.
- Halliday, Michael A. K. 2013. Meaning as choice. In Lise Fontaine, Tom Bartlett & Gerard O'Grady (eds.), *Systemic functional linguistics: Exploring choice*, 15–36. Cambridge: Cambridge University Press.
- Halliday, Michael A. K. & Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.
- Halliday, Michael A. K. & Ruqaiya Hasan. 1985. *Language, context, and text: Aspects of language in a social-semiotic perspective*. Victoria: Deakin University Press.
- Halliday, Michael A. K. & Christian M. I. M. Matthiessen. 2014. *Halliday's introduction to functional grammar*, 4th edn. London: Routledge.
- Jakobson, Roman. 1959. On linguistic aspects of translation. In Robin Arthur Brower (ed.), *On translation*, 232–239. Cambridge: Harvard University Press.
- Kress, Gunther. 2010. *Multimodality: A social semiotic approach to contemporary communication*. New York: Routledge.
- Kress, Gunther, Theo Van Leeuwen. 2021. *Reading images: The grammar of visual design*, 3rd edn. Abingdon: Routledge.
- Martin, James R. 1991. Intrinsic functionality: implications for contextual theory. *Social Semiotics* 1(1). 99–162.
- Matthiessen, Christian M. I. M. 1995. *Lexicogrammatical cartography: English systems*. Tokyo: International Language Sciences Publishers.
- O'Halloran, Kay L. 1999. Interdependence, interaction and metaphor in multisemiotic texts. *Social Semiotics* 9(3). 317–354.
- O'Halloran, Kay L. 2007. Systemic functional multimodal discourse analysis (SF-MDA) approach to mathematics, grammar and literacy. In Rachel Whittaker, Mick O'Donnel & Anne McCabe (eds.), *Advances in language and education*, 75–100. London: Continuum.
- O'Halloran, Kay L. 2008. Systemic functional-multimodal discourse analysis (SF-MDA): Constructing ideational meaning using language and visual imagery. *Visual Communication* 7(4). 443–475.
- O'Halloran, Kay L. 2013. Multimodal analysis and digital technology. In Elena Montagna (ed.), *Readings in intersemiosis and multimedia*, 35–53. Jerusalem: IBIS Editions.
- Rose, David. 2023. Genre, register and discourse in systemic functional linguistics. In Michael Handford & James Paul Gee (eds.), *The Routledge handbook of discourse analysis*, 2nd edn., 328–345. London: Routledge.
- Royce, Terry. 1998. Synergy on the page: Exploring intersemiotic complementarity in page-based multimodal text. *JASFL Occasional Papers* 1(1). 25–49.
- Royce, Terry. 2002. Multimodality in the TESOL classroom: Exploring visual-verbal synergy. *Tesol Quarterly* 36(2). 191–205.
- Thibault, Paul J. 2000. The multimodal transcription of a television advertisement: Theory and practice. In Anthony Baldry (ed.), *Multimodality and multimodality in the distance learning age*, vol. 31, pp. 311–385. Campobasso: Palladino Editore.

- Van Leeuwen, Theo, Joshua Han. 2023. Evaluation and discourse analysis. In Michael Handford & James Paul Gee (eds.), *The Routledge handbook of discourse analysis*, 2nd edn., 23–38. London: Routledge.
- Wang, Kunfeng, Chao Gou, Yanjie Duan, Yilun Lin, Xinhua Zheng & Fei-Yue Wang. 2017. Generative adversarial networks: Introduction and outlook. *IEEE/CAA Journal of Automatica Sinica* 4(4), 588–598.
- Yuen, Cheong Yin. 2004. The construal of ideational meaning in print advertisements. In Kay L. O'Halloran (ed.), *Multimodal discourse analysis: Systemic functional perspectives*, 163–195. London: Continuum.

## Bionote

### Arash Ghazvineh

Department of Art Research, Faculty of Arts, Tarbiat Modares University, Tehran, Iran

[arash.e.ghazvineh@gmail.com](mailto:arash.e.ghazvineh@gmail.com)

<https://orcid.org/0000-0003-2434-5786>

Arash Ghazvineh is a PhD student in Arts Research at Tarbiat Modares University, Tehran, Iran. He completed his master's in Dramatic Literature at the same university with a thesis on the inter-semiotic translation from dramatic texts to films. His research centers on multimodality, sign-processes and semiosis and cultural semiotics.