

AN ANALYSIS OF COMMUNICATION MODES OF VISUAL IMAGES IN A MEDICAL TEXTBOOK

PHAM BICH NGOC*

ABSTRACT: In the teaching and learning process, a strong connection exists between visual literacy and the representation of images in textbooks, particularly in the context of English for Specific Purposes (ESP) textbooks. In order to perform a comprehensive analysis of communication modes, 30 images from a medical textbook were examined using Kress and Van Leeuwen's (2021) analytical strategies for visual images, alongside Halliday's (1985) transitivity system. This analysis was conducted with the premise that visual and verbal discourses interact and reflect various modes of communication. The study of multimodal discourse showed that visual images were designed with a different range of representational, interactive, and compositional modes. The findings may offer valuable theoretical and pedagogical insights for medical students and textbook designers, emphasizing the value of incorporating multimodal resources and effectively combining visual elements with verbal strategies to convey meaning.

KEY WORDS: Analysis, visual images, multimodal discourse, communication modes, medical textbooks.

RECEIVED: March, 25th, 2025.

EDITED: June, 26th, 2025

1. Introduction

In an era marked by an exponential increase in visual culture and digital communication, the concept of multimodality has gained critical importance in understanding how meaning is constructed across various platforms and disciplines. Multimodality refers to the strategic combination of different semiotic modes- such as visual images, verbal language, layout, gestures, and sound- to enhance communication and meaning-making. This paradigm shift has prompted a growing scholarly interest in how these modes co-exist and interact, particularly in educational and professional settings where clarity and engagement are essential. Among the key developments in this field is the emergence of Systemic Functional Multimodal Discourse Analysis (SF-MDA), a theoretical framework rooted in Halliday's (1994) Systemic Functional Linguistics (SFL). SF-MDA offers a powerful lens through which to examine how multimodal texts function in social contexts. Central to this framework are the three metafunctions - ideational, interpersonal, and textual - which have been extended by Kress and van Leeuwen (2021) to account for the visual mode. These metafunctions respectively correspond to how images represent the experiential world, how they construct relationships with viewers, and how elements are composed to create coherence and structure. As the boundaries of traditional linguistic analysis are expanded, the application of SF-MDA enables a richer, more nuanced interpretation of meaning beyond the verbal channel.

In this evolving communicative landscape, pedagogical materials- especially textbooks used in English for Specific Purposes (ESP)-have undergone a significant transformation. Contemporary ESP textbooks are no longer confined to dense blocks of text; instead, they are increasingly designed with a multimodal approach, integrating photographs, diagrams, charts, and typographic variation to cater to learners' diverse cognitive styles and to enhance comprehension. The visual elements in these materials serve not merely decorative purposes but act as integral carriers of disciplinary knowledge, functioning in tandem with the verbal text to scaffold learning, clarify concepts, and stimulate interest. Empirical research supports the pedagogical value of such multimodal designs. For instance, Bahari (2018) illustrates that visual images can stimulate learners' intrinsic motivation and emotional

*Hai Phong University of Medicine and Pharmacy; Email: ngocydhp@gmail.com

engagement, while Zhao et al. (2020) suggest that visual inputs enhance cognitive processing by creating more interactive and flexible connections between new information and learners' existing knowledge structures. Elmiana (2019) applied Kress and van Leeuwen's framework to Indonesian senior high school EFL textbooks, demonstrating that images not only complement but also critically shape the meaning of educational content. However, despite the evident role of images in facilitating learning, there remains a noticeable lack of research focusing specifically on how visual semiotics function in ESP materials, particularly in highly specialized fields like medicine.

To address this gap, the current study investigates the interplay of verbal and visual semiotic resources in medical ESP textbooks, with a particular focus on how these elements collaboratively function to convey specialized knowledge. The analysis draws upon a multimodal discourse framework that integrates both static (images, diagrams) and dynamic (interactional structure, gaze, salience) visual elements, as well as the linguistic structures embedded in accompanying texts. Following Baldry and Thibault (2006), this research conceptualizes semiotic resources as operating simultaneously at both the contextual (social-pragmatic) and code-deployed (formal-structural) levels, allowing for a comprehensive evaluation of how meaning is encoded and decoded within educational discourse. Furthermore, in line with Kress (2003), the study acknowledges that multimodal compositions differ significantly from purely verbal texts in terms of their spatial layout, information hierarchy, and semiotic density, necessitating new forms of analytical literacy.

The focal material for this research is the textbook *Oxford English for Careers: Medicine 1* (McCarter, 2009), a widely used resource in medical English instruction. Adopting a descriptive-analytic methodology, the study aims to address the following research questions:

1. *What are the most frequently occurring sub-types of Halliday's (1985) transitivity system in the verbal analysis of medical textbook images?*
2. *What are the most commonly observed sub-types of Kress and van Leeuwen's (2021) visual image interpretive strategies in the visual analysis of medical textbook images?*
3. *How do visual and verbal discourses interact in medical textbook images to reflect modes of communication?*

Through this inquiry, the study contributes not only to the field of multimodal discourse analysis but also to pedagogical design and curriculum development, offering insights into how ESP materials can be optimized to enhance learner engagement, comprehension, and retention in specialized domains such as medicine.

2. Literature review

2.1. Systemic Functional Linguistics (SFL) and Multimodal Discourse Analysis (MDA)

The foundation of this study rests upon Systemic Functional Linguistics, pioneered by Michael Halliday (1985, 1994). In contrast to formalist linguistics, SFL views language as a social semiotic system, where meaning is created through choices in context. Central to SFL are three metafunctions- ideational, interpersonal, and textual- which correspond to how language constructs experience, enacts social relations, and organizes discourse. These metafunctions have been expanded by multimodal theorists to analyze visual modes as legitimate meaning-making systems. Building on Halliday's theory, Kress and van Leeuwen (1996, 2006, 2021) developed a visual grammar that applies the SFL metafunctions to visual images. In their framework, the representational (ideational) metafunction explains how images depict actions, participants, and conceptual categories; the interactive (interpersonal) metafunction explores the relationship between image and viewer (e.g., gaze, social distance, and point of view); the compositional (textual) metafunction deals with how elements are arranged to create coherence and salience (e.g., information value, framing, and salience).

Such frameworks have enabled scholars to analyze images as semiotic resources on par with language, thereby legitimizing the integration of visual analysis into linguistic inquiry. Van Leeuwen (2005) and Baldry and Thibault (2006) further contributed to the development of Systemic

Functional Multimodal Discourse Analysis (SF-MDA), stressing the need to consider intersemiotic cohesion, spatial configuration, and multimodal synergy in meaning construction.

2.2. Multimodal Discourse in Medical ESP: A SF-MDA Perspective

While multimodal analysis has gained increasing traction in fields such as English Language Teaching (ELT), English for Academic Purposes (EAP), and general education, its application to English for Specific Purposes (ESP), particularly in medical textbooks, remains limited and underdeveloped. Most existing studies either focus solely on verbal or visual modes, with little effort to examine how these modes interact in discipline-specific contexts. In medical ESP textbooks, visual elements- such as diagrams, illustrations, and clinical images-are often treated as supplementary or decorative, rather than as integral to meaning-making and knowledge construction. This underrepresentation neglects the epistemological weight that visuals carry in the medical field, where visual literacy is essential for both learning and professional practice.

At the same time, the complexity and density of medical discourse call for more robust analytical frameworks. Systemic Functional Linguistics (SFL), and particularly Halliday's transitivity system, offers valuable tools for analyzing how experience is constructed through language-highlighting who does what to whom, under what circumstances. However, the use of transitivity analysis in tandem with image analysis in medical ESP materials is still scarce. When combined with Kress and van Leeuwen's visual grammar-focusing on representational, interpersonal, and compositional meanings-this dual approach can reveal how textual and visual elements work together (or fail to) in communicating medical knowledge.

This study seeks to address these gaps by conducting a multimodal analysis of selected medical ESP textbook excerpts. Specifically, it applies transitivity analysis to the accompanying verbal texts to explore how medical actions, participants, and processes are linguistically represented. In parallel, it employs visual grammar analysis to examine the semiotic features of medical imagery. Crucially, it investigates how the verbal and visual modes interact-whether they complement, reinforce, or contradict one another in meaning. By integrating these perspectives, the study extends the application of SF-MDA to a high-stakes professional domain and contributes practical insights into multimodal pedagogy, medical discourse understanding, and the more effective design of ESP teaching materials.

2.3. Visual semiotics in ESP textbooks

Semiotics has gained considerable prominence in applied linguistics, drawing the attention of numerous scholars who explore this field from diverse perspectives. Among these, Zamani (2016) argues that semiotics plays a crucial role in language instruction as it helps students develop their cognitive abilities at all levels of comprehension. Semiotics holds immense potential in the field of language learning and teaching. Senel (2007) suggests that semiotics enhances the teaching and learning process through the use of non-verbal communication, images, visuals, films, videos and more. By leveraging semiotics, language learning becomes increasingly engaging and motivating. Furthermore, semiotics can significantly improve students' performance across all four language skills and aspects of language acquisition.

According to Hismanoglu (2005), semiotics is concerned with anything that can be used as a sign. A sign is defined as anything that can be interpreted as standing for something else. In this sense, semiotics can be understood as related to anything that can be perceived or interpreted as a sign. There are two primary models of semiotics: Saussure's (2000) dyadic model and Peirce's (1958) triadic model. The key difference between Saussure's and Peirce's perspectives lies in their focus. Saussure's semiotics emphasizes the semantic aspects, primarily focusing on establishing a connection between the vocal component of the sign, the signifier, and its referent, the signified. In contrast, Peirce approaches semiotics from a philosophical standpoint, concerned with the relationship between the physical world and nature (Masinambow and Rahayu, 2001).

Textbooks, especially in ESP contexts, are increasingly acknowledged as multimodal pedagogic

artefacts. They not only transmit knowledge but also mediate disciplinary identities and shape learners' epistemological access to specialized content (Hyland, 2009). In medical English textbooks, images such as diagrams, X-rays, procedural photos, and anatomical sketches are more than supplementary; they are core to conveying scientific concepts, clinical processes, and professional practice. A growing body of research investigates the semiotic roles of images in ELT and ESP materials. For example, Elmiana (2019) analyzed high school EFL textbooks using Kress and van Leeuwen's visual grammar, finding that most images served decorative rather than informational purposes. Zhao et al. (2020) highlighted the role of visuals in enhancing cognitive load management, particularly when learners encounter complex domain-specific vocabulary. Meanwhile, Bahari (2018) demonstrated that images can dynamically engage learner motivation, suggesting that visuals should not be seen merely as aesthetic additions, but as pedagogically active components in discourse.

The theoretical framework of this study is grounded in Kress and van Leeuwen's (2021) strategies for visual images in multimodal texts (Figure 1) and Halliday's (1985) transitivity system for the verbal analysis (Figure 2) of textbook images. This integration of multimodal theoretical approaches enables a comprehensive examination of both textual and visual elements in medical textbook images.

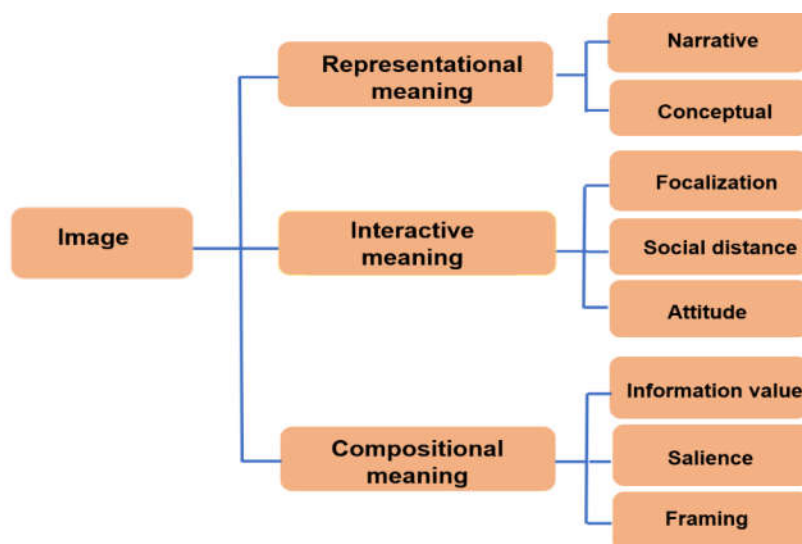


Figure 1. Kress and van Leeuwen's (2021) framework for visual images strategies

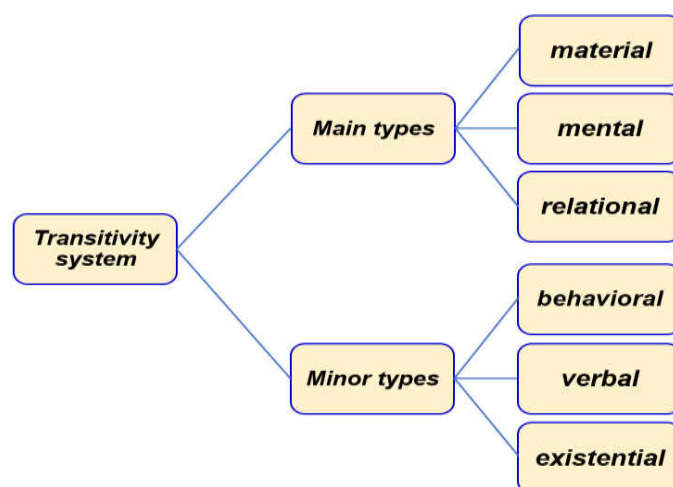


Figure 2. Halliday's (1985) framework for the verbal analysis

3. Methodology

For the purposes of the present investigation, the textbook *Oxford English for Careers: Medicine 1* by Sam McCarter (2009) was selected as the primary data source due to its pedagogical focus on English for Specific Purposes (ESP), particularly in the context of medical communication. From this text, a total of thirty medical-themed textbook images were randomly sampled to serve as the visual corpus. The study adopted a descriptive-analytic research design, allowing for the systematic examination of multimodal semiotic elements within the selected materials.

The multimodal analysis was conducted across two semiotic modes: the verbal mode and the visual mode. In the verbal mode, the unit of analysis was defined as either an independent or dependent clause, in alignment with systemic functional linguistic principles that consider the clause the fundamental unit of meaning (Halliday & Matthiessen, 2014). In the visual mode, analytical attention was given exclusively to those images accompanied by concise verbal annotations, thus ensuring a cohesive interplay between visual representation and linguistic content, and enabling a more nuanced interpretation of meaning within a social semiotic framework.

The overarching aim of the study was to investigate the representational strategies employed in the depiction of social actors- specifically agents- within textbook visuals used in medical English instruction. This objective was approached through the application of a dual theoretical framework. First, Kress and van Leeuwen's (2021) visual grammar model was employed to decode the compositional and representational structures embedded in the images, focusing on aspects such as narrative processes, salience, gaze, and vectorial relations. Second, Halliday's (1985, 1994) transitivity system, situated within the ideational metafunction of language, was utilized to analyze the linguistic descriptions associated with each image. The transitivity system provides a means to categorize processes (e.g., material, mental, relational) and participants (e.g., actors, goals, sensors), thereby revealing how experiential meanings are constructed and conveyed in pedagogical discourse.

Systemic Functional Grammar (SFG), as articulated by Halliday (1985) and later elaborated by Halliday and Matthiessen (2014), posits that language is organized around three metafunctions: the ideational, the interpersonal, and the textual. The ideational metafunction, which is of primary concern in this study, pertains to the construal of experience and the representation of the world through processes, participants, and circumstances. Accordingly, the verbal elements in the textbook images were analyzed through this lens to identify how experiential meanings align or diverge from the visual narratives.

To facilitate a coherent synthesis of findings, all data points- both visual and verbal- were coded and systematically presented in a comprehensive tabular format (Table 1). This methodological

choice allowed for cross-modal comparison and helped elucidate patterns of meaning-making within the multimodal representations of professional medical practice. Through this integrated approach, the study offers insights into the pedagogical strategies underlying multimodal representation in ESP textbooks and contributes to a broader understanding of meaning construction in instructional discourse.

Table 1. *Tabular Data Codification and Categorization*

<i>Category type</i>	<i>Codification options</i>
<i>Text-image relationship</i>	Direct; Indirect; <i>No association</i>
<i>Modes and realizations of visual meanings</i>	Representational; Interactive; <i>Compositional</i>
<i>Ideational functions</i>	Main type- Material; Main type-Mental Main type-Relational; Minor type-Behavioral Minor type-Verbal; <i>Minor type-Existential</i>

4. Results and discussions

The results of the study in relation to the first research question revealed that, at the verbal level, the most frequent communication mode representation in the discourse of medical textbook images was at the relational main-type level, while the least frequent was at the verbal minor-type level. Table 2 presents the distribution of transitivity process types within the verbal descriptions that accompany the selected visual images, analyzed according to Halliday's (1985, 1994) Systemic Functional Grammar. The table illustrates the frequency of each process type- material, mental, relational, behavioral, verbal, and existential- based on the clause-level analysis of 30 sampled utterances.

Table 2. *Communication modes in verbal discourse of visual images*

<i>Clauses</i>		<i>Type of process</i>					
		<i>Material</i>	<i>Mental</i>	<i>Relational</i>	<i>Behavioral</i>	<i>Verbal</i>	<i>Existential</i>
No	30	7	6	11	2	1	3

As shown, relational processes were the most frequently used among all categories, appearing in 11 out of 30 clauses (36.7%). This high frequency of relational processes is particularly significant in the context of medical education, where definitions, identifications, and descriptions of conditions or anatomical features are prevalent. Relational clauses, which typically express states of being, attribution, or identity (e.g., "*The heart is a muscular organ*"), function to construct static, categorical knowledge, aligning with the declarative and factual orientation of medical discourse (Egins, 2004; Halliday & Matthiessen, 2014).

Material processes, which indicate actions or events involving actors and goals (e.g., "*The nurse injects the patient*"), accounted for 7 instances (23.3%). The presence of material clauses reflects an effort to encode dynamic actions within clinical procedures or physiological processes. Such clauses reinforce the procedural nature of medical knowledge, allowing learners to conceptualize medicine as a field grounded in practice and intervention (Martin & Rose, 2007). Mental processes, appearing in 6 clauses (20%), serve to encode processes of perception, cognition, and emotion (e.g., "*The patient feels pain*"). Their presence indicates an awareness of the subjective and experiential dimensions of healthcare, particularly when addressing diagnostic sensitivity or patient-centered care. The inclusion of these clauses in educational materials may signal an epistemological shift from a purely biomedical model to one that incorporates humanistic perspectives in clinical training (White, 1998). Behavioral processes were relatively scarce, with only 2 instances (6.7%). As processes that sit between material and mental types, behavioral clauses typically denote physiological or

psychological behaviors (e.g., “*The patient coughs*”). Their infrequent occurrence may reflect a disciplinary emphasis on either observable actions (material) or inner states (mental), rather than hybrid categories. Existential clauses (e.g., “*There is swelling in the abdomen*”) occurred 3 times (10%), typically serving to assert the presence or existence of medical phenomena. These clauses play an important role in establishing clinical conditions or symptoms without attributing agency, a rhetorical choice that aligns with the objectivity and diagnostic neutrality often found in scientific discourse (Gotti, 2005). Verbal processes, found in only one instance (3.3%), were markedly underrepresented. These typically report acts of saying or communication (e.g., “*The doctor explains the procedure*”). The rarity of verbal processes suggests that the communicative aspects of doctor-patient interaction are not a major focus in the selected textual extracts, reinforcing the argument that the textbook prioritizes technical knowledge over interpersonal communication in its linguistic design.

Taken together, the distribution of transitivity processes reveals a marked preference for clause types that construct factual, observational, and procedural knowledge, thereby reinforcing the epistemological stance of the medical discourse community. The dominance of relational and material processes over verbal and behavioral types reflects the genre-specific demands of instructional texts in medical education- where knowledge is primarily disseminated through classification and demonstration rather than dialogue or affect.

These findings also underscore the importance of multimodal curriculum design, wherein the verbal text is complemented by images that may, to some extent, fill the interpersonal or interactive gap left by the textual mode. Future instructional materials might benefit from a more balanced integration of process types to support a more holistic approach to medical language competence, as advocated by recent developments in ESP and medical humanities (Hyland, 2006; Candlin & Candlin, 2002).

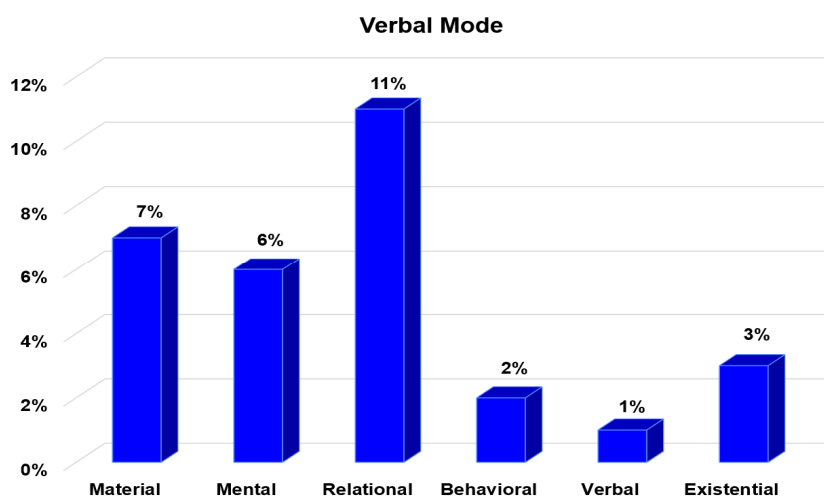


Figure 3. The frequency percentage of verbal levels based on Halliday's classification

The analysis further indicated that the most frequently observed sub-types of Halliday's (1985) transitivity system in the verbal analysis of textbook images were primarily relational (11%), functioning as a main-type ideational component. This finding suggests that visual images in the medical textbook predominantly serve to depict objects or concepts that exist without employing features to establish a direct connection between the signifier and the signified. In other words, visual images function as vehicles for conveying ideological concepts (i.e., extra-linguistic ideas) to learners rather than fulfilling linguistic functions (i.e., serving as tools for communication through language).

The analysis further revealed that among the sub-types of transitivity processes identified in the verbal descriptions accompanying the textbook images, relational processes emerged as the most frequently employed, accounting for approximately 11% of all identified clauses. As a central component of Halliday's (1985) ideational metafunction, relational processes construct meaning by establishing relationships of attribution, identification, or classification between participants, rather than encoding dynamic actions or mental states.

This predominance of relational processes suggests that the verbal discourse embedded within the visual images primarily serves a taxonomic and descriptive function. That is, it tends to define, categorize, or characterize entities, rather than narrate events or represent cognitive or interactive experiences. This pattern is indicative of a discourse orientation that prioritizes knowledge representation over communicative interaction, aligning with the epistemological priorities of scientific and medical texts, which frequently privilege objectivity, classification, and decontextualized knowledge (Halliday & Matthiessen, 2014; Martin & Rose, 2007).

Regarding the second research question, the researcher applied Kress and van Leeuwen's (2021) visual image strategies across three levels: representational, interactive and compositional to assess the frequency of each level. The frequency of each level and its sub-levels was presented in Table 3.

Table 3. Visual images framework based on Kress and van Leeuwen's (2021) strategies

Mode	Category		Sub-category
Representational	Participant	Human	Age
			Gender
			Sociocultural Portrayal
	Distance	Non-human	Close-up
			Medium shot
			Long shot
Interactive	Perspective	Horizontal	Frontal / Oblique
		Vertical	High /Low/ Eye-levelled
	Modality	Color	
		Contextualization	
Compositional	Information value	Left-Right	
		Top-Bottom	
		Centre/ Margin	
	Text-Image	Information	
		Illustration	
		Decoration	

Drawing on Kress and van Leeuwen's (2021) social semiotic framework, the representational metafunction of visual communication is understood as encompassing the ways in which images depict participants- whether human, non-human, animate, or inanimate- along with their actions, experiences, relationships, and associated attributes. This metafunction addresses the ideational dimension of meaning-making in visual discourse, particularly how depicted entities are constructed

to convey culturally embedded knowledge and values.

In applying this theoretical lens to the visual data, the present study first categorized the participants in each image according to ontological type- human or non-human- and systematically recorded their frequency of occurrence across the sample. This initial classification served as the basis for a deeper representational analysis, focusing on how different types of participants are visually constructed and positioned.

Subsequent analysis examined the relational configurations and semiotic properties of these participants, including age, gender, attire, physical positioning, and associated visual cues. Particular attention was paid to the implicit sociocultural meanings encoded in these representations, such as normative gender roles, hierarchical professional identities, and ethnocultural invisibilities. The analysis revealed that visual images in the medical textbook are not neutral depictions of reality; rather, they function as ideologically saturated semiotic resources that contribute to the reproduction of dominant discourses- such as biomedical objectivity, gendered labor divisions, and Western clinical norms- within the pedagogical space.

By interrogating these visual constructions, the study uncovered latent patterns of representation and exclusion, illuminating how certain identities are foregrounded while others are marginalized or rendered invisible. These findings underscore the importance of critical multimodal literacy in the design and evaluation of educational materials, particularly in fields like medicine where visual representations carry epistemological and sociocultural weight.

The frequency and percentage of the representational mode were illustrated in Figure 4.

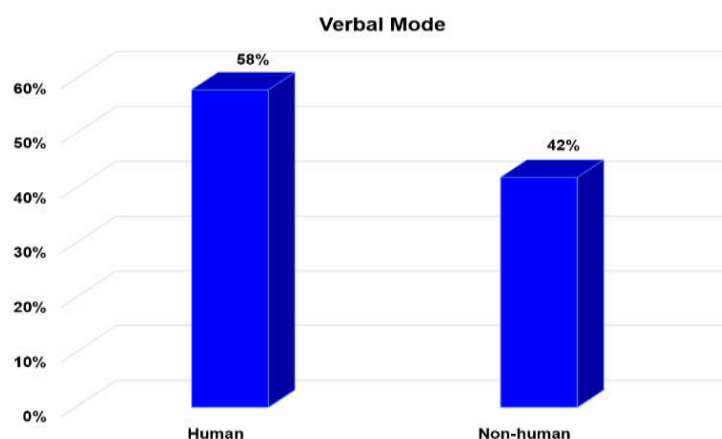


Figure 4. *The frequency percentage of representational mode*

As shown in Figure 4, the verbal analysis of the representational metafunction revealed that human participants constituted 58%, while non-human participants accounted for 42%. This distribution suggests that the textbook gives slight priority to human-centered medical discourse, reflecting the pedagogical aim of emphasizing clinical roles, professional identities, and patient interaction.

According to Kress and van Leeuwen (2021), representational choices in multimodal texts construct not only what is shown, but also how knowledge and values are encoded. The frequent presence of human participants aligns with the need to represent experiential and interpersonal aspects of medical practice, while the substantial representation of non-human elements-such as organs, tools, or diseases- highlights the discipline's commitment to technical precision and scientific objectivity (Lemke, 1998). These findings confirm that even in the verbal mode, the textbook serves a dual function: conveying factual knowledge and shaping learners' professional worldview through ideologically informed representational choices.

The analysis of the interactive metafunction- as theorized by Kress and van Leeuwen (2021)- focused on the semiotic strategies through which visual elements position viewers and establish communicative relationships within the multimodal discourse of medical textbook images. This metafunction addresses how images serve not merely to represent but also to interact with their audience, shaping the viewer's alignment, engagement, and epistemic stance. In the current study, three primary parameters were examined: social distance, perspective, and modality. The frequency and percentage distribution of these features are presented in Figure 5.

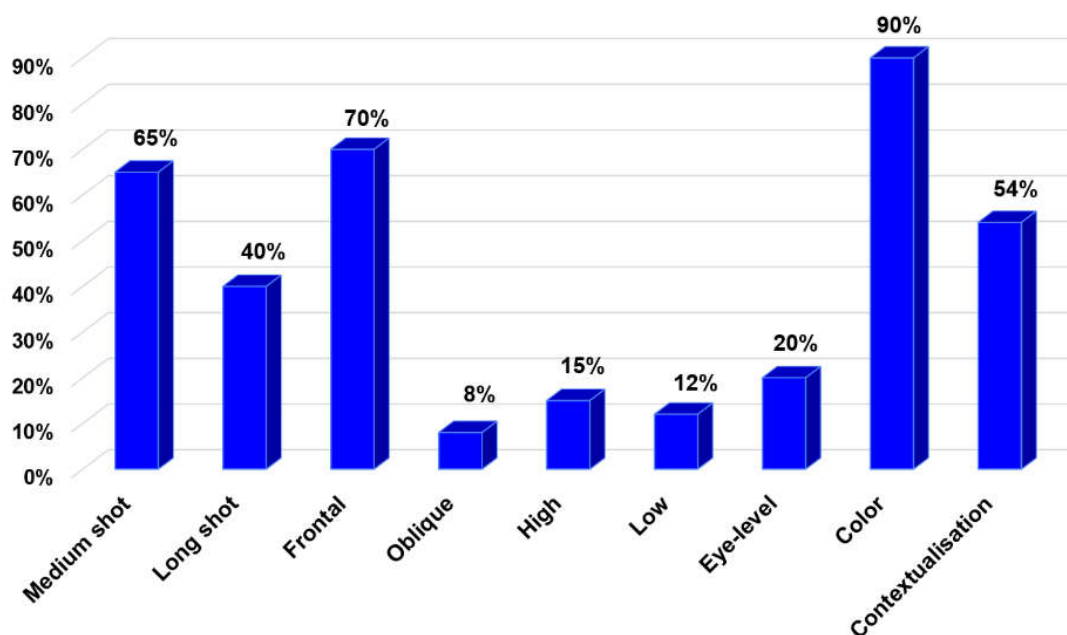


Figure 5. The frequency percentage of interactive mode

In terms of social distance, the majority of visuals employed medium shots (65%), followed by long shots (40%), suggesting a deliberate positioning of the viewer at a mid-range interpersonal distance. This framing likely serves to balance clinical detachment with sufficient engagement for instructional clarity. The use of medium shots is consistent with pedagogical goals in medical education, which prioritize both procedural overview and anatomical detail without excessive intimacy (Machin, 2016).

Regarding perspective, the horizontal axis revealed a dominant preference for frontal angles (70%), with a relatively small proportion using oblique angles (8%). Frontal positioning tends to construct alignment and interpersonal involvement, placing the viewer as an active participant in the scene (Kress & van Leeuwen, 2021). On the vertical axis, eye-level perspectives (20%) were employed to establish equality between viewer and represented participant, while high (15%) and low (12%) angles signaled subtle power asymmetries- typically elevating the professional authority of healthcare providers or diminishing the agency of patient figures (van Leeuwen, 2005).

The modality analysis focused on visual realism and semiotic intensity. A significant 90% of images utilized color, a high- modality feature that enhances representational immediacy and instructional salience (Lemke, 1998). Moreover, 54% of visuals were contextually embedded, meaning that participants were situated within recognizable clinical or anatomical settings. This use of contextualization supports knowledge transfer by anchoring abstract content in concrete, discipline-specific environments, thereby reinforcing epistemic authenticity.

Overall, the data demonstrate that the interactive metafunction is strategically employed to construct a balanced pedagogical stance, allowing for both engagement and professional distance.

These visual strategies work not only to instruct, but also to socialize learners into the discursive norms and interpersonal dynamics of the medical field.

The analysis of the third research question revealed that the text– image relationship in the selected medical textbook images is predominantly indirect, indicating a nuanced and non- linear form of multimodal communication. The frequency of text-image relationship as communication modes is presented in Figure 6.

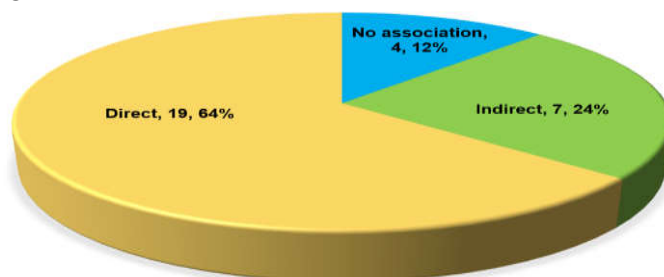


Figure 6. Frequency of text-image relationship as communication modes

As presented in Figure 6 and supported by image samples in the Appendix, only 19,64 % of the visuals were found to have a direct correspondence with the accompanying text- where visual and verbal modes reinforced the same semantic content. In contrast, 7,24 % exhibited an indirect relationship, in which the image and text shared thematic alignment but conveyed distinct or complementary meanings. Notably, 4,12 % of the visuals appeared entirely unconnected to the associated verbal content.

These results align with Barthes' (1977) framework of anchorage and relay, suggesting that visuals in medical textbooks often act not merely as illustrations but as independent semiotic resources contributing to meaning-making in more abstract or inferential ways. Such an approach implies that the image is not simply reinforcing the written word but is engaging the learner in a more cognitively demanding process of interpretive integration. This design reflects broader shifts toward constructivist multimodal pedagogy, where learners are expected to synthesize information across modes (Kress & van Leeuwen, 2021; Jewitt, 2008). However, it also highlights the importance of fostering multimodal literacy in learners, particularly in specialized academic fields like medicine.

According to Kress and van Leeuwen (2021), compositional meaning in multimodal texts is shaped by spatial arrangements that govern how information is organized and interpreted by the viewer. In a left- right structure, the left-side element typically encodes what is already known or contextually assumed- the Given- while the right-side element presents what is new, unfamiliar, or informationally prominent- the New. Similarly, in a top- bottom configuration, the top portion of the image is interpreted as the Ideal, conveying generalized, abstract, or value-laden content, whereas the bottom portion reflects the Real, emphasizing practical, concrete, or observable aspects of the message.

In the current study, the analysis of compositional metafunctions in the selected textbook images revealed a strong preference for the left - right structure, which accounted for 80% of all cases (see Figure 7). This dominant orientation suggests that image designers systematically positioned familiar or context-setting elements on the left, followed by new or focal information on the right - a spatial arrangement that guides the learner's reading path and supports cognitive processing of scientific content (Kress & van Leeuwen, 2021). By contrast, only 11% of the images employed a top- bottom structure, indicating less frequent use of Ideal - Real contrast in organizing visual meaning.

This pattern reflects a broader tendency in educational materials to prioritize linear, sequential modes of information delivery, especially in disciplines such as medicine where clarity, categorization, and hierarchical organization are essential to knowledge construction. The preference

for Given- New structures may also facilitate learning by leveraging familiar schemas before introducing novel or technical information- thus enhancing comprehension in line with information design principles (Tufte, 2006) and multimodal learning theory (Mayer, 2009).

The frequency and percentage distribution of these compositional modes are depicted in Figure 7.

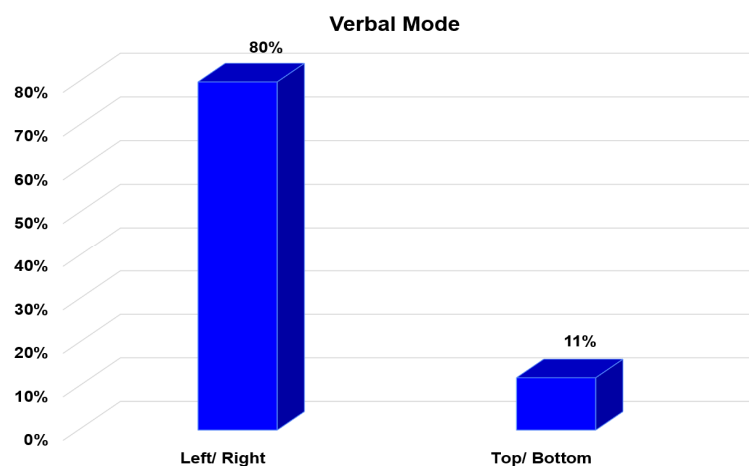


Figure 7. The frequency percentage of compositional mode

The findings show partial alignment with those reported by Svensson (2020) who examined the emotional dimension of text and image in Swedish textbooks. This study highlighted both similarities and differences in visual and verbal meaning-making constructs, emphasizing the importance of drawing styles and their potential effects on various age groups. The results also corroborate the findings of Andra et al. (2020) who affirmed the importance of integrating images into pedagogy to enhance learning development.

Several limitations, however, influenced the outcomes of this study. The multimodal discourse analysis conducted was based on two specific frameworks and interpreted the data according to their guidelines, meaning that the same textbooks could potentially yield different insights when analyzed through alternative frameworks. Additionally, the study lacked direct and convenient access to the textbook author who could have provided further insights into the rationale behind the selection of certain images over others.

5. Conclusions

Based on the findings from the multimodal discourse analysis of verbal– visual interactions in *Oxford English for Careers: Medicine 1* (McCarter, 2009), it can be concluded that the visual images are systematically constructed through a variety of representational, interactive, and compositional metafunctions, as outlined in the social semiotic framework of Kress and van Leeuwen (2021). These multimodal resources do not merely illustrate content; rather, they function as semiotic instruments that contribute to meaning-making within the specialized discourse of medical English.

The study underscores the need for a comparative and integrative theoretical framework that can account for how meaning is negotiated across modes- particularly in the context of cognitive reception and visual literacy among medical learners. This entails further investigation into how the sequencing, alignment, and interplay of multimodal elements support or hinder knowledge construction and shape learners' perceptions of content validity, clarity, and relevance.

Moreover, the findings support the claim that visual images- when purposefully designed- can enhance learners' ability to connect ideational meanings embedded in visuals with the verbal content of the curriculum, thereby facilitating deeper understanding (Fernández- Fontecha et al., 2020). Such integration not only improves conceptual clarity but also aligns with pedagogical principles of

multimodal learning (Mayer, 2009). Therefore, future research is encouraged to expand the analytical corpus to include a broader range of image types and genres, enabling a more nuanced understanding of the epistemic and instructional functions of visual representations in ESP and medical education.

REFERENCES

1. Bahari, Akbar. (2018), Sacred Text Motivation for General L2 Learners: A Mixed Methods Study. *Journal of Academic Ethics* 16 (4): 377-407. <https://doi.org/10.1007/s10805-018-9316-3>.
2. Baldry, Anthony, and Paul J. Thibault. (2006), Multimodal Transcription and Text Analysis (London: Equinox). *Developing Multimodal Texture* 55.
3. Barthes, R. (1977). *Image, Music, Text* (S. Heath, Trans.) (pp. 38-39). London: Fontana Press.
4. Candlin, C. N., & Candlin, S. (2002). *Introduction: Discourse, Expertise and the Management of Risk in Health Care Settings*. In *Research on Language and Social Interaction*, 35(2), pp. 118-123.
5. Eggins, S. (2004). *An Introduction to Systemic Functional Linguistics* (2nd eds., p. 206). London: Continuum.
6. Elmiana, Dewi Satria. (2019), Pedagogical Representation of Visual Images in EFL Textbooks: A Multimodal Perspective. *Pedagogy, Culture & Society* 27 (4): 613-628. <https://doi.org/11.1080/100481366.2019.1569550>.
7. Fernandez-Fontecha, Almudena, Kay L. O'Halloran, Peter Wignell, and Sabine Tan. (2020), Scaffolding CLIL in the Science Classroom via Visual Thinking: A Systemic Functional Multimodal Approach. *Linguistics and Education* 55: 100788. <https://doi.org/10.1016/j.linged.2019.100788>.
8. Gotti, M. (2005). *Investigating Specialized Discourse* (pp. 90-95). Bern: Peter Lang.
9. Halliday, M.A.K. (1985), *An Introduction to Functional Grammar* (1st eds.). London: Edward Arnold.
10. Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. Second Edition. London: Edward Arnold.
11. Halliday, M. A. K. and C. M. I. M. Matthiessen (2004). *An Introduction to Functional Grammar*. Third Edition. Great Britain: Hodder Arnold.
12. Hişmanoğlu, Murat. (2005), Semiotic elements and Difficulties in Teaching Vocabulary Items. *Dil Dergisi* 128: 51-68.
13. Hyland, K. (2006). *English for Academic Purposes: An Advanced Resource Book* (pp. 48-54). London: Routledge.
14. Hyland, K. (2009). *Academic Discourse: English in a Global Context*. London: Continuum/Bloomsbury Academic, p.112.
15. Jewitt, C. (2008). *Multimodality and Literacy in School Classrooms* (pp. 15-28). Routledge.
16. Kress, G. R., & van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design* (1st eds.). London: Routledge.
17. Kress, G., & Van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design* (2nd eds.). London: Routledge.
18. Kress, Gunther. (2003), *Literacy in the New Media Age*. Routledge.
19. Kress, G. R., & van Leeuwen, T. (2021), *Reading Images: The Grammar of Visual Design* (3rd eds.). Routledge. <https://doi.org/10.4324/9781003099857>.
20. Lemke, J. L. (1998). Multimodality. *Linguistics and Education*, 10(3), 359-386.
21. Machin, D. (2016). *The Need for a Social and Affordance-driven Multimodal Critical Discourse Studies*. *Discourse & Society*, 27(3), 322-334.

22. Martin, J. R., & Rose, D. (2007). *Working with Discourse: Meaning Beyond the Clause* (2nd eds., pp. 100-105). London: Continuum.
23. Masinambow, E. K. M., and Rahayu S. Hidayat. (2001), *Semiotik: Mengkaji tanda dalam artifak*. Jakarta: Balai Pustaka.
24. Mayer, R. E. (2009). *Multimedia Learning* (2nd eds., pp. 34-42). Cambridge University Press.
25. Sam McCarter. (2009), *Oxford English for Careers, Medicine 1*. Oxford University Press.
26. Senel, Mufit. (2007), The Semiotic Approach and Language Teaching and Learning. *Journal of Language and Linguistic Studies* 3 (1): 117-132.
27. Tufte, E. R. (2006). *The Cognitive Style of PowerPoint: Pitching Out Corrupts Within* (pp. 51-60). Cheshire, CT: Graphics Press.
28. Van Leeuwen, Theo. (2005), *Introducing Social Semiotics*. Psychology Press.
29. White, P. R. R. (1998). Extended Reality: Pronouns and the Vernacular. In J. R. Martin & R. Veel (Eds.), *Reading Science: Critical and Functional Perspectives on Discourses of Science* (pp. 266-296). London & New York: Routledge.
30. Zamani, Sirvan. (2016), Semiotics and Language Teaching (Teaching English to Kurd Students). *World Scientific News* 42: 293-305. <https://doi.org/10.1111/j.1540-4781.1996.tb01613.x>.
31. Zhao, Fang, Schnotz Wolfgang, Wagner Inga, Gaschler Robert. (2020), Texts and Pictures Serve Different Functions in Conjoint Mental Model Construction and Adaptation. *Memory & Cognition, New York*, 48 (1): 69-82. <https://doi.org/10.3758/s13421-019-0096>.

Phân tích các phương thức giao tiếp của hình ảnh trực quan trong một giáo trình y khoa

Tóm tắt: Trong lĩnh vực giảng dạy và học tập, đặc biệt là đối với các giáo trình tiếng Anh chuyên ngành hình ảnh trực quan đóng vai trò quan trọng trong việc hỗ trợ năng lực đọc hiểu hình ảnh (visual literacy) và truyền đạt nội dung chuyên môn. Nghiên cứu này nhằm phân tích toàn diện các phương thức giao tiếp được thể hiện qua 30 hình ảnh trực quan trích từ giáo trình *Oxford English for Careers: Medicine 1* (McCarter, 2009). Phương pháp phân tích diễn ngôn đa phương thức được áp dụng, kết hợp giữa các chiến lược phân tích hình ảnh của Kress và van Leeuwen (2021) và hệ thống chuyên tác của Halliday (1985), nhằm khám phá cách thức hình ảnh và ngôn ngữ tương tác để kiến tạo ý nghĩa. Kết quả nghiên cứu cho thấy các hình ảnh trực quan được thiết kế trên cơ sở ba phương diện chính: biểu hiện (representational), tương tác (interactive), và bố cục (compositional), phản ánh một cấu trúc truyền thông đa dạng và có chủ đích. Những phát hiện này cung cấp những đóng góp đáng kể cả về mặt lý thuyết và thực tiễn, đặc biệt trong việc nâng cao hiệu quả thiết kế giáo trình và phát triển kỹ năng diễn giải thông tin trực quan của người học chuyên ngành. Nghiên cứu cũng đề xuất mở rộng phạm vi khảo sát trong tương lai để hiểu sâu hơn về vai trò của hình ảnh trong việc hỗ trợ tiếp nhận và chuyển giao tri thức chuyên môn trong bối cảnh đào tạo y khoa.

Từ khóa: phân tích; hình ảnh trực quan; diễn ngôn đa phương thức; phương thức giao tiếp; giáo trình y khoa.