



Re-contextualising ‘Greeting’: A multimodal analysis in an EFL Textbook

Dadan Jauhara*, Emi Emilia, and Iwa Lukmana

*Department of English Education, Faculty of Language and Literature Education,
Universitas Pendidikan Indonesia Jl. Dr. Setiabudhi No. 229 Bandung, West Java, Indonesia*

ABSTRACT

This paper provides an account of multimodal meanings represented in an English textbook as a teaching and learning resource. This study explores how ideational, interpersonal and textual meanings are represented across visual and verbal modes and their intermodal relationship. This study employed a qualitative case study design. The major theoretical basis for this research is Systemic Functional Linguistic Multimodal Discourse Analysis (SFL-MDA) which focuses on how meaning is made through the use of multiple modes of communication. The data were obtained from one chapter of the textbook for the seventh grade, which is about greeting. In general, the findings show that, ideationally, the images can be categorised as narrative representation and, verbally, the texts are dominated by relational processes. Moreover, a visual and verbal relationship can be categorised as projection-location. Interpersonally, it is shown that images can play an essential role in shaping readers’ feelings or attitudes through cartoon characters, color, natural coding orientation, and typeface. Together with the verbal mode, images play a role in positioning the readers’ interaction with the textbook. Textually, the book’s page layouts reflect the organisation of meaning across texts and images and affect the viewers’ reading path.

First received:

29 January 2020

Revised:

14 August 2020

Accepted:

2 November 2020

Published:

28 February 2021

Keywords:

intermodal relation;
multimodal; verbal;
visual

INTRODUCTION

The role of educational technologies has provided many opportunities to present multiple representations of content (text, images, audio, video) to cater more effectively to different learning style (Sankey et al., 2010). In the last two decades, English teachers should be able to teach a wide range of materials such as textbooks, diagrams, photographs and drawing to facilitate students’ learning style (Ajavi, 2012). It is in line with Gardner’s (1993) theory of multiple intelligences in which students possess a different kind of learning style, and one of them is visual intelligence.

A textbook is one of the learning materials in language teaching, and it typically incorporates pictures and other visuals (Weninger, 2018). The role and value of visual images in English textbooks have been recognised and reflected by some researchers and teachers. They have started to realise the importance of visual elements and graphic design in English textbooks. Based on their study, it is suggested that textbooks incorporating both aesthetic features and functional purposes can be used to trigger the art of teaching, conduct a better learning experience and promote students’ learning motivation and aesthetic experiences (Chan et al., 2012). Moreover, visual images presented in English textbooks can be used to develop students’ language skills and carry messages with them. Therefore, teachers and textbook writers must understand how language works together with other modes, such as images to create meaning.

* Corresponding author: danjauhara@gmail.com

Multimodal discourse analysis for English textbooks and other learning resources has been conducted by some researchers. These include an exploration of ideational, interpersonal, and textual meanings realised through the co-dependence of visual and linguistic semiotic modalities in page layouts (Low & Guo, 2004), an analysis of interpersonal meaning in textbooks for teaching English as a foreign language (Chen, 2009), a survey of teachers' perspective on a multimodal textbook for a secondary school in raising critical language awareness (Nguyen, 2011), and investigation on how multimodality in English textbooks may scaffold learning through visual text (Salbego et al., 2015). Moreover, in Indonesian EFL context, multimodal discourse analysis on English textbook has also been conducted by some researchers. These include an analysis of ideational meaning and the relationship between verbal and visual meanings in the textbook (Hermawan, 2013), the development of English materials for vocational school from the perspective of SFL social semiotic (Widodo, 2015), cultural context represented in English textbooks (Farida et al., 2018; Sumiyati & The, 2014).

In line with the basic competencies that should be learned by the students of junior secondary school in Indonesian EFL context, the expression of greeting is one of the topics that is provided in the textbook. Based on the English curriculum, the students are expected to be able to express greeting to the people around them. The greeting is one of the topics that should be learned because the greeting is an act of communication to show attention and to suggest a type of relationship or social status (formal or informal) between individuals or groups of people. This kind of conversation can be categorised as an interpersonal text because it is intended to maintain an interpersonal relationship between the speaker and interlocutor (Celce-Murcia & Olshtain, 2000).

English textbooks for junior high school usually provide not only verbal text but also visual text. Therefore, the students should be able to interpret meanings of the texts which integrate visual and verbal modalities presented in the textbooks (Macken-Horarik, 2004). Moreover, the teachers also play an essential role in instructing their students to explore visual and verbal resources in the textbooks to support the teaching and learning process. Therefore, this paper aims to discuss the role of visual and verbal texts and its relation in promoting English teaching and learning process. Thus, image and text provided in an English textbook are analysed to look at the affordances of multimodal meaning in assisting language learning.

THEORETICAL FRAMEWORK

The study has been informed by the theory of systemic functional linguistics (SFL). It is a social theory of language that "...regards language as a meaning-making resource through which people interact with each other in given situational and cultural context" (Emilia, 2014, p. 63). The use of language is influenced by context, it "...extends along the cline of instantiation from the potential pole (context of culture) to the instance of pole (context of situation)" (Matthiessen et al., 2010, p. 77). In terms of context, the functional model of language concerns two strata of contexts which are derived from Malinowski, i.e. the context of culture and the context of the situation. Related to context, Malinowski claimed that "language only becomes intelligible when it is placed within its context situation" (Eggins, 2004, p. 88).

The study centred on uncovering ideational, interpersonal, and textual meanings of visual and verbal modes represented in an EFL textbook and intersemiotic relationship as well as its implication. The data were analysed under the framework of SFL - Multimodal Discourse Analysis (SFL-MDA), focusing on how meaning is made through the use of multiple modes of communication (O'Halloran, 2011). SFL-MDA is considered relevant to analysing the data because it concerns with theorising and analysing the multiple resources of different modes of text, image, sound, gesture and three-dimensional objects in visual, media or digital text which combine to create meaning in other contexts (Kress & van Leeuwen, 2001; O'Halloran, 2011).

Framework for Analysing Verbal Text

Based on the discussion above, it can be summarised that there are three different abstractions of meaning in SFL called metafunctions, namely ideational, interpersonal, and textual metafunctions. Ideational meaning construes ‘reality’; interpersonal meaning construes ‘social reality’; while textual meaning construes ‘semiotic reality’ (Martin, 1991 cited in Christie & Unsworth, 2000).

Systemic Functional Grammar (SFG) approach from Halliday (1985, 1994), Halliday and Matthiessen (2004, 2014), Eggins (2004), Martin et al. (1997), and Butt et al. (2000) is used in analysing ideational meaning of verbal mode, focusing on the types of processes and their participants and circumstances (TRANSITIVITY system); interpersonal meaning realised through what is called the MOOD and MODALITY systems; and textual meaning, realised through the THEME system.

Framework for Analysing Images

Visually, the data were analysed based on visual grammar (Kress & van Leeuwen, 1996, 2006) and, in turn, the theory has extended and adapted by other theorists such as Callow (2013), Humphrey et al. (2012), Joyce and Gaudin (2007), and Painter et al. (2013).

In terms of ideational (representation) meaning, the image can be narrative or conceptual representations. Narrative representation processes can be action, reaction, speech, mental, and conversion. Those actions are also represented circumstantially according to the represented context of the situation (Setting, Means, and Accompaniment), and also attributes or qualities and characteristics of the participants represented. Moreover, the categories of conceptual representations are Classification, Analytical, or Symbolic.

Interpersonal (interactive) meanings of visual mode, the interactive meanings are based on contact, social distance, attitude, gaze, proximity and angle between participants, and modality markers. Contact is realised by ‘gaze’ between the represented participant and the viewer (offer and demand). Social distance is realised through the size of frame based on the distance of shots (close shot = personal), (medium shot = social), and (long shot = impersonal). The realisation of attitude is based on the perspective or angle that creates power relation (vertical angle) and involvement (horizontal angle). Power relations consist of (represented power = low angle), (viewer power = high angle), and (neutral = eye level). Involvement consist of (oblique = detachment) and (frontal = involvement). Concerning the attitude in visual interaction, the present study primarily focuses on subjective images and ignores objective images. Finally, the textual (compositional) relates the representational and interactive meanings of the image to each other through three interrelated systems, namely INFORMATION VALUE, SALIENCE, and FRAMING.

METHOD

This research was conducted to explore how ideational, interpersonal and textual meanings in visual and verbal modes are represented in a textbook and to explore the intermodal relationship of the visual-verbal modes and their effects in constructing the meaning, shaping readers’ interaction and feeling, and guiding reading path.

The study used a qualitative case study design especially text analysis by using SFL-MDA because it concerned with the theory and practice of analysing meaning arising from the use of multiple semiotic resources (O’Halloran, 2008). Multiple semiotic resources in this study refer to the use of verbiage and image in an EFL textbook. Therefore this study involves the analysis of linguistic and visual forms of semiosis based on SFL principles. The strength of SFL-MDA for this study is “metafunctional principle which provides an integrating platform for theorising how semiotic resources interact to create meaning” (O’Halloran, 2008, p. 444). For this study, the one-page layout was selected and analysed in the context of visual social semiotics (Callow, 2013; Joyce & Gaudin, 2007; Kress & van Leeuwen 2006; Painter et al., 2013). The data were taken from Chapter 1 of an EFL textbook for seventh-grade students of junior secondary school, entitled

“*Bahasa Inggris: When English Rings a Bell (Edisi Revisi 2014)*”. The selected page layout can be categorised as multimodal because this includes both verbal texts and images.

FINDINGS AND DISCUSSION

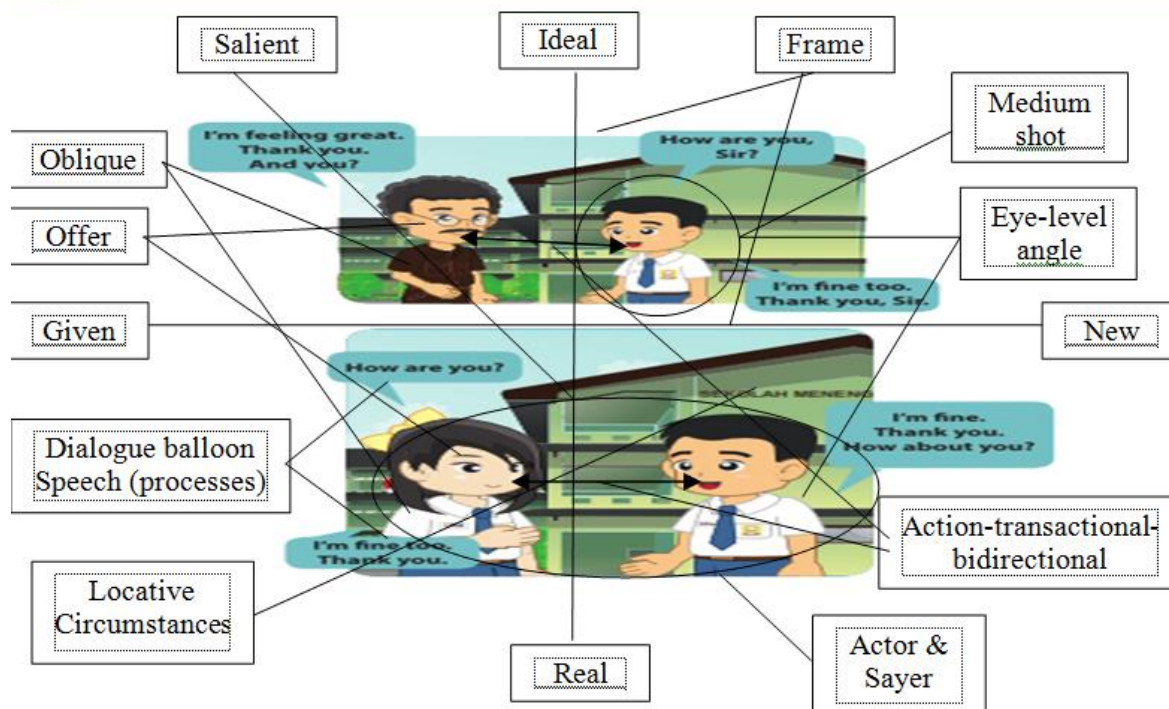
This section presents and discusses findings with regard to how ideational, interpersonal, and textual meanings in both visual and verbal modes are represented in an EFL textbook and how they work together to create meaning, shaping readers’ feeling and interaction, and guiding readers’ reading path. This section is divided into three subsections comprising ideational, interpersonal, and textual meanings.

Ideational Meanings

In Figure 1, the readers can identify the main characters, the activities, and the circumstances represented in the images. The participants’ attributes, such as their outfits can help the readers to understand who these participants are and what activities they are involved in. The participants can be identified as the students of Indonesian junior high school students due to the uniforms they are wearing. It is relevant to the point that “...colour clearly can be used to denote specific people, places and things as well as classes of people...” (Kress & van Leeuwen, 2002, p. 347). Meanwhile, a man who looks older than them with eyeglasses and his formal clothes seem to be a teacher. Figure 1 depicts two events of conversation. The upper image depicts a conversation between a boy (student) and his teacher, and the lower picture shows a conversation between a boy (a student) and his friend (girl).

Figure 1

Visual analysis (Bahasa Inggris: When English Rings a Bell Kelas VII 2014, p. 7)



The two successive images portray two moments of greeting and their responses. Related to this type of representation, the activities reflected in the image can be categorised as narrative representation. The participants are standing up and looking at each other, and their gestures, represented mainly by their facial expression and arms indicate that they are actively involved in a conversation. In terms of narrative representation, the images indicate that they represent

action:transactional:bidirectional because a vector emanates from the boy's eyes toward the girl or man and vice versa (Kress & van Leeuwen, 2006). Their arms also form vectors toward each other. Both images depict the process of conversation between represented participants. Therefore, both represented participants play the role as an Actor and Goal at the same time, in which both participants are speaker and listener (Kress & van Leeuwen, 2006). Further, the presence of the dialogue balloons indicates that the processes are speech processes. Six dialogue balloons project a topic discussed by the represented participants. Therefore, the represented participants in the image function as the Sayer, the speech balloons realise the verbal processes, and the texts play the role of projected wording.

With respect to circumstances, the images depict the locative circumstances. It is indicated by the presence of school building, plants, clouds and sky as a background. It can be seen in both images that the represented participants are foregrounded in front of the school building. The presence of locative circumstance can help the viewers to get additional information. They will get information that the conversations occurred at school in the day time. Therefore, the presence of image accompanying verbal text can lead the students' mental imagery that when he/she meets someone (friends or teachers) at school, they have to initiate to greet them in order to maintain the interpersonal relationship.

Related to verbal mode projected in the dialogue balloon, Figure 1 represents the expression of greeting and thanking. It can be seen from the following excerpt:

Student : *"How are you, Sir?"*
Teacher : *"I'm feeling great. Thank you. And you?"*
Student : *"I'm fine too. Thank you, Sir."*

The excerpt shows that the clause *"How are you"* indicates that the boy greets and asks his teacher about his condition, whether he is well or not and the teacher responds it by saying *"I am feeling great,"* giving information that he is well. The clause *"I am feeling great"* can be categorised as Mental processes because the word *"great"* indicates its quality of his feeling at that time (Derewianka, 2011). Similarly, the relational processes in the image are indicated by the clause *"I am fine too."* This type of relational process can be categorised as Attributive clauses because the clause is not reversible and it relates a participant to their attribute *"fine"* (Emilia, 2014).

After elaborating the ideational meanings of visual and verbal modes, the paper is now in a position to elaborate the relation between texts and images. There are two visual components: represented participants (students and teacher) and locative circumstance (school) accompanying verbal text (dialogue balloons). Each of visual and verbal modes is related to each other by logico-semantic relation of projection. It can be seen that the images and texts are interdependent, which is realised by each of them playing a role in a verbal projection; the image projects the text or vice versa. Therefore, the intersemiotic relation between text and image can be categorised as Projection–locution (Martinec & Salway, 2005; Unsworth, 2007) because the verbal mode is enclosed in dialogue balloons. The images and texts are independent, realised by the whole image being related to the whole texts. The represented participants in the images function as the Sayer, the speech balloons realise the verbal processes, and the texts play the role of projected wording.

Interpersonal Meanings

With respect to interpersonal meaning, the analysis focused on interactive and evaluative meanings. In terms of interactive meanings, the images can be categorised as 'Offer' since none of the participants depicted is looking at the readers/viewers. It means that none of the vector relates the eye-line of the represented participants to the readers/viewers. The absence of gaze or facial expressions directly towards the viewer suggests that it is simply offering information (Kress & van Leeuwen, 2006). It indicates that the represented participants invite the viewers to observe what is going on in the images, and they cannot involve directly with the participants. The images portray on which the viewer can look with really no requirement to react other than to agree with

it. Therefore, from the perspective of Orientation, the images do not invite direct engagement from the viewers. It just affords them to observe what is going on in the images.

In terms of social distance between the participants and the viewer, both images offer social relationship (less intimate) since the participants are taken from the medium shot (the waist up of human figure). With regard to the involvement and power relations, it can be approached from the viewpoint of the horizontal axis (involvement) and the vertical axis (power relation). Power relation between represented participants and the viewers in a textbook is encoded by the vertical angle formed between them. Figure 1 indicates that the authors and the textbook designers construct equal power relation because the participants are depicted in eye-level angle. Therefore, there is no power gap between the viewers and the participants, and thus it reflects an equal position in power (Callow, 2013; Chen, 2009; Kress & van Leeuwen, 2006). Horizontally, the participants in Figure 1 are depicted in oblique angle, which suggests detachment toward the readers/viewers. Therefore, what is depicted in the image is not part of the viewer's world. Thus, it does not need viewer involvement.

Further, gaze, proximity, and angle between the characters allow the readers to interpret various relationship and emotional connection between the represented participants in images (Callow, 2013). Figure 1 shows that the represented participants (boy and girl) gaze at each other, and the proximity between them is quite close in terms of social relationship. It can be identified by way of a boy stares directly at the girl's eyes in quite intense demand. It indicates a strong emotional connection between them and suggests deepening friendship between them.

In terms of modality or ambience (Painter et al., 2013), the choice of colours contributes to the enactment of interpersonal meanings in the image. The colour can be used to impress or to intimidate and reduce or raise stress levels of the viewers. Both images use natural colour for indicating the represented participants and setting. Natural colour helps the students in understanding meaning in the real context of the situation. The bright, warm, full realistic colour of the participants and locative background can make the readers amplify their emotion. Therefore the reader emotionally engaged with the participants in the textbook (Callow, 2013; Kress & van Leeuwen, 2006).

In terms of the degree of representation as modality markers, Figure 1 illustrates the cartoons as represented participants. The participants perform different expressive faces. The boy's eyes show happiness, which can be observed from his enlarged pupils that imply great interest. His facial expression with mouth open demonstrates quite enthusiastic gestures (Chen, 2009). However, the facial expressions of the teacher and the girl seem unenthusiastic because their mouths are closed with the lips down. These different facial expressions shown by represented participants can influence viewers/readers' feeling.

The conversation excerpt in Figure 1 indicates that the boy initiates the conversation by saying "*How are you?*" to his teacher. In terms of speech function, the boy asks a question to his teacher to 'demand' information (interrogative Mood) and the teacher responds it by answering the question and take turn to asks his students by saying "*And you?*".

The verbal text indicates that there are no modality markers presented in the text. The excerpt shows that participants use positive polarity clause. It means that the participants provide factual information about their condition. Moreover, the absence of modality markers in the verbal text indicates that the authors provide the reader with the polarity clauses that indicate absolute values of truth (Halliday & Matthiessen, 2014). However, the words '*great*' and '*fine*' indicate a positive feeling of the participants.

Moreover, related to the typography of the text, it is not deniable that recognising words with some fonts is easier than others (van Rossum, 1998 cited in Tarasov et al., 2015). The *Helvetica 11* typeface of the text presented in the textbook is perceived to enhance the legibility of the texts by helping the readers to distinguish the letters and words more easily in reading the text, the font type and size have a significant effect on reading speed (Chandler, 2001). Moreover, the black

colour for the typeface makes it contrast with the colour of the paper. Therefore, students can read the text clearly.

After discussing visual and verbal interpersonal meanings, it is time to elaborate on the relation between images and texts with regard to the multimodal construal of greeting expression. In terms of the relationship between images and verbiage, Figure 1 suggests that the absence of modality markers and evaluative words presented in the text do not provide evaluative meanings. Therefore, verbally the text cannot provoke the evaluation of readers to co-articulate Affect, Judgment, and Appreciation. However, visually, the authors position the readers and represented participants in an equal position. It can be evidenced by the depiction of the social distance of the represented participant and eye level angle of viewing. The texts and visual images try to invite readers to share feelings regarding the expression of greeting. The images utilise evaluative meanings to align viewers with the feeling they intended to share through participants' facial expression and colour. The enthusiastic face expression of the represented participant, especially the boy, can affect the reader's positive feeling on greeting expression.

Textual Meanings

The textual meaning of the book involved compositional relations, focusing on three main points; (1) Information Value realised through the placement of images and texts (left/right, vertical/horizontal and centre/ margin) on the page, (2) Salience, and (3) Frame. In this study, textually, the layout of the book can be categorised as top-down composition because the reader is led by the image and verbal text of instruction. At the top position of the page (Ideal), the image of the teacher with the dialogue balloon instructs the reader to practice greeting with the people around them, followed by the image of the girl saying "How are you" as an example of greeting expression. Meanwhile, on the downside (Real), there are two images performing dialogue of greeting. Therefore, it indicates that the instruction (Ideal) is a more prominent element in the page layout; it tends to contain big ideas, the promise of what is to follow. Meanwhile, the example of conversation (Real) is considered less prominent, featuring more detailed information (Kress & van Leeuwen, 2006).

Regarding the images of conversation, it can be seen that the teacher is positioned on the left side (Given), the student (boy) is positioned on the right side (New), while the dialogue balloons indicate that the student initiates conversation and the teacher responds to it. This image indicates that the placement of the participant is not determined by the rule of a reading path, from left to the right because the initiator of the conversation is positioned on the right side while the responder is positioned on the left side. Therefore, it requires readers' ability to understand the meaning of verbal text projected in the dialogue balloons. This type of position can confuse the readers whose reading path are from left to right and thus, the layout of the textbook can be considered irrelevant with one of the characteristics of a good textbook as offered by Tomlinson (1998) in which a textbook should have an attractive layout, make the students feel at ease, and feel relaxed when seeing or reading a textbook. Further, the lower image has a different pattern in terms of positioning the role of speaker and responder in which the speaker asks the girl her condition and the girl responds to it. The second image places the initiator of conversation (addresser) on the left side (Given) while the responder on the right side (New). It is relevant with the left to the right reading path, and thus, it can be considered good as it fulfils one of the criteria of a good books as discussed by Tomlinson (1998).

In terms of salience, the images in the textbook are more salient than the text by virtue of their colour intensity, Given and New position and its size. Figure 1 demonstrates that the most salient image is represented by participants because they display the greatest tonal contrast in colour and their sizes. They are also positioned as foreground. Meanwhile, the school building, blue sky, and trees are placed behind as the background of the represented participants. The text "How are you" is printed in larger font size than others. Therefore, it can attract the reader's attention to read the text. It may give a cue to readers on what they are going to learn.

Further, the framing of both images is separated by the white space; there is no line or actual frame lines to indicate that both images are in the frame. Moreover, both images can be considered related because one of the represented participants is presented in both images. Therefore, both images indicate the process of greeting between students and between the student and the teacher.

In terms of verbal textual meaning, both conversations indicate that the topical Theme (unmarked) through nominal groups may lead students to look through the rest of the paragraphs and clause to discover the point of the text; that is, where the text is heading (Butt et al., 2000). Moreover, it can assist the reader in accessing the information and can make it easier for readers because they connect new information to information already presented in the text (Joyce & Feez, 2012).

CONCLUSION

This paper has elaborated how ideational, interpersonal, and textual meanings contribute to construing the topic of “greeting” in language learning. Utilising SFL-MDA, this paper exemplified how those resources create meanings, shaping feeling and interaction, and guiding the readers to read the textbook. As discussed in this paper, the analysis of ideational, interpersonal, and textual meanings and their intermodal relations suggest that these visual and verbal resources play an important role in assisting students in understanding the expression of greeting when learning English.

Ideational meanings presented in a textbook verbally and visually can construct the nature of the event, the object and the participants involved, and the circumstances in which they occur. These representations are used to construct the meaning to represent or make a statement about the reality of the expression of greeting. The interpersonal resources across images and verbal modes can create an interactive relationship between a textbook and the readers. The interpersonal analysis of visual and verbal modes has revealed the way of a textbook author in creating an imaginary relationship between the textbook and the viewers/readers and also shaping readers’ feeling. The various techniques of the depiction of the images and speech functions are intended to create interaction between the represented participants in the image with the readers. Moreover, the textual meaning in visual and verbal modes contributed to the organisation of meanings across semiotic modes in the textbook.

REFERENCES

- Ajavi, L. (2012). How teachers deploy multimodal textbooks to enhance English language learning. *TESOL Journal*, 6, 16-35.
- Butt, D., Fahey, R., Feez, S., Spinks, S., & Yallop, C. (2000). *Using functional grammar: an explorer’s guide*. National centre for English language teaching and research.
- Callow, J. (2013). *The shape of text to come: How Image and text work*. PETAA
- Celce-Murcia, M. & Olshtain, E. (2000). *Discourse and context in language teaching: A guide for language teachers*. Cambridge University Press.
- Chan, B. J., Yeh, L. W., & Chen H. L. (2012). Applying illustration and layout design for textbook to enhance the art of teaching: A case of social studies textbook. *Journal of Textbook Research*, 5(1), 47-84. <https://ej.naer.edu.tw/JTR/v05.1/2012-04-jtr-v5n1-047.pdf>
- Chandler, S.B., (2001). *Running head: Legibility and comprehension on screen type* (Doctoral dissertation). Retrieved from, <http://scholar.lib.vt.edu/theses/available/etd-11172001-152449/unrestricted/chandler.pdf>
- Chen, Y. (2009). *Interpersonal meaning in textbooks for teaching English as a foreign language in China: A multimodal approach* (Unpublished doctoral thesis). University of Sydney, Australia.
- Christie, F. & Unsworth, L. (2000). Developing socially responsible language research. In. L. Unsworth (Ed), *Researching language in school and community* (pp. 1-26). Cassel.

- Derewianka, B. M. (2011). *A new grammar companion for teachers* (2nd ed.). Primary English Teaching Association.
- Eggs, S. (2004). *An introduction to systemic functional linguistics* (2nd ed.). Continuum.
- Emilia, E. (2014). *Introducing Functional Grammar*. Pustaka Jaya.
- Farida, A. R., Warsono, & Hartono, R. (2018). The cultural relation between verbal and visual image in English textbooks for junior high school. *English Education Journal*, 8(2), 162-168.
- Gardner, H. (1993). *Multiple intelligences: The theory in practice*. Basic Books.
- Halliday, M.A. K. (1994). *Functional grammar* (2nd ed.). Arnold.
- Halliday, M. A. K. & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar*. Arnold.
- Halliday, M. A. K. & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th Ed.). Routledge.
- Hermawan, B. (2013). *Status-logicosemantic relation of texts and images in English textbook in Indonesia: How messages are constructed and understood* (Unpublished research report). Universitas Pendidikan Indonesia, Indonesia.
- Humphrey, S., Droga, L., & Feez, S. (2012). *Grammar and meaning*. PETAA.
- Joyce, H. & Gaudin, J. (2007). *Interpreting the visual*. Phoenix Education Pty. Ltd.
- Joyce, H. & Feez, S., (2012). *Text-based language and literacy education: Programing and methodology*. Phoenix Education Pty. Ltd.
- Kress, G. & van Leeuwen, T. (1996). *Reading images: the grammar of visual design*. Routledge.
- Kress, G & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Bloomsbury.
- Kress, G. & Van Leeuwen, T. (2002). Colour as a semiotic mode: notes for a grammar of colour. *Visual Communication*, 1(3), 343-369
- Kress, G. & van Leeuwen, T. (2006). *Reading images: the grammar of visual design* (2nd ed.). Routledge.
- Low, F., & Guo, S. (2009). Analysing English language textbooks: Case studies from a multimodal perspective. *Proceeding International Systemic Functional Linguistic Congress* (pp. 196-203). International Systemic Functional Linguistics Association.
- Macken-Horarik, M. (2004). Interacting with the multimodal text: Reflections on image and verbiage in *Art Express*. *Visual Communication*, 3(1), 5-26.
<https://doi.org/10.1177/1470357204039596>
- Matthiessen, C. M. I. M., Teruya, K., & Lam, M. (2010). *Key terms in systemic functional linguistics*. Continuum.
- Martin, J. R., Mathiessen C. M. I. M., & Painter C. (1997). *Working with functional grammar*. Arnold.
- Martinec, R. & Salway, A., (2005). A system for image—text relations in new (and old) media. *Visual Communication*, 4(3), 337-371.
- Nguyen, T.T. (2011). *An analysis of multimodal texts in the EFL textbook for secondary school in Vietnam from teachers' perspective on raising critical language awareness: A Multimodal Approach* (Unpublished Dissertation). London Metropolitan University.
- O'Halloran, K. 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, K. L. (2011). Multimodal analysis and digital technology. In A. Baldry & E. Montagna (Eds.), *Interdisciplinary perspectives on multimodality: Theory and practice* (pp. 21-34). Palladino.
- Painter, C., Martin J. R., & Unsworth, L. (2013). *Reading visual narratives: Image analysis of Children's picture books*. Equinox
- Salbego, N., Heberl M. V., & Balen, M. G. S. d. S. (2015). A visual analysis of English textbooks: Multimodal scaffolded learning. *Calidoscopio*, 13(1), 5-13.
- Sankey, M., Birch, D., & Gardiner, M. (2010). Engaging students through multimodal learning environments: The journey continuous. *Proceeding ASCILITE Sydney 2010* (pp. 852-863). Australasian Society for Computers in Learning in Tertiary Education.

- Sumiyati & The, Y.H (2014). Multimodal analysis and the cultural context of Indonesian English textbooks. *Working Papers*. Retrieved from:
https://www.researchgate.net/publication/319643671_Multimodal_analysis_and_the_cultural_context_of_Indonesian_English_textbooks/link/59b77c96458515c212b3fb31/download
- Tarasov, A. D., Sergeev, P.A. & Filimonov, V.V., (2015). Legibility of textbook: a literature review. *Procedia - Social and Behavioral Sciences*, 174, 1300-1308.
<https://doi.org/10.1016/j.sbspro.2015.01.751>
- Tomlinson, B. (1998). *Developing materials for language teaching*. Cambridge University Press.
- Unsworth, L. (2006). Toward a metalanguage for multiliteracies education: Describing the meaning-making resources of language-image interaction. *English Teaching: Practice and Critique*, 5(1), 55-76.
- Unsworth, L. (2007). Image/text relations and intersemiosis: Towards multimodal text description for multiliteracies education. *Proceedings 33rd International Systemic Fuctional Congress (33rd ISFC)* (pp. 1165-1205). *Pontificia Universidade Catolica De Sao Paulo, Brazil*. Retrieved from:
http://www.pucsp.br/isfc/proceedings/Artigos%20pdf/59pl_unsworth_1165a1205.pdf
- Weninger, C. (2018). Textbook analysis. In C.A. Chappelle, (Ed.) *The encyclopedia of applied linguistics*. Willey and Sons.
- Widodo, P.H. (2015). *The development of English materials from a social semiotic perspective: Parcipatory action research* [Unpublished doctoral dissertation]. Department of Linguistic, University of Adelaide.