Image-text Relations in Junior High School EFL Textbooks in China: A Mixed-methods Study

LAI Haiyan (Kelly)
Faculty of Education, The University of Hong Kong, Hong Kong

Abstract—**This study seeks to test the applicability of Martinec and Salway's (2005) framework of image-text relations in school textbook discourse, and adopts mixed methods by combining quantification of thirteen categories in Martinec and Salway's analytical framework with interpretations of the relations between visual images and verbal language in the six junior high school EFL textbooks to generalize multimodal trends in pedagogical discourse. The findings shed some light on cultivating students' multimodal literacy in the pedagogical context of Mainland China.**

Index Terms—**image-text relations, EFL textbook discourse, multimodal literacy, systemic-functional multimodal discourse analysis**

I. INTRODUCTION

Past decades have witnessed the trend of multimodality which involves multiple semiotic modes such as visual image, verbal language, and acoustic sound in human communication, as well as the mushroom development of SF-MDA (Multimodal Discourse Analysis) inspired by Halliday's (1994) Systemic Functional Linguistics. SF-MDA views diverse semiotic resources in terms of representational, interactive, and compositional tri-meta-functions, i.e., modeling visual representations of material world, constructing the relations between viewers and what is viewed, and creating visual coherence, according to Kress and van Leeuwen (2006). Pedagogical materials cannot escape from this multimodality trend. Nowadays, textbooks (especially primary and secondary school textbooks) are full of colorful pictures, cartoons, and highlighted texts. Some are even equipped with multimedia supplementary materials (such as DVDs), which draw students' attention more easily, due to the synchronization of multiple modes.

Inter-semiotic relations or inter-semiosis (i.e., the interactive communication power achieved by different semiotic resources in multimodal texts) are key aspects of MDA study (Jewitt, 2009; O'Halloran, 2011). Barthes (1977) first put forward three kinds of image-text relations, i.e., illustration, anchorage, and relay, which were applied to the analysis of newspaper photographs, moving images, and dialogue in film. Focusing on the ideational aspect of image-text relations, and drawing on the earlier version of the framework, Martinec and Salway (2005) integrated Barthes' (1977) categorization of image-text relations with Halliday's (1994) logico-semantic relations of expansion and projection to create a generalized framework of image-text relations. Most systemic-functional semioticians started to probe into intersemiotic relations in the late 1990s (Kress & van Leeuwen, 1996/2006; Lemke, 1998; Royce, 1998, 2002, 2006; O'Halloran, 2008; Stockl, 2004; Martinec & Salway, 2005; Kong, 2006; Unsworth, 2006, 2007; Matthiessen, 2007; Liu & O'Halloran, 2009; Painter & Martin, 2011). Previous multimodal studies on EFL textbooks in diverse contexts mainly adopted Kress & van Leeuwen's (2006) visual analysis framework such as Chen (2009) in China, Nordensvard (2010) in Sweden, Tahririan & Sadri (2013) in Iran, Weninger & Kiss (2013) in Hungrary, Salbego et al. (2015) in Brazil, and Torres (2015) in South Korea. Whilst other multimodal EFL textbook discourse studies have incorporated Kress & van Leeuwen's (2006) visual analysis model with other frameworks such as Royce's (1998) intersemiotic complementarity framework (e.g., Liu & Qu, 2014 in China), or Kong's taxonomy (2006) (e.g., Vorvilas, 2014), Kress and van Leeuwen's (2005) text-image relations framework (e.g., Yassine, 2014 in Algeria). However, the comprehensive framework of Martinec and Salway (2005) has not been adopted in these pedagogical materials studies on image-text relations to quantify categories in a series of textbook corpus to generalize trends and findings.

In light of the afore-mentioned, the research question is formulated below:

What are the multimodal relations between verbal language and visual images, as well as their appraisal meanings in junior high school EFL textbooks in Mainland China (with Guangzhou as a featured case)?

In order to address above issues, this study adopts mixed methods, in other words, for one thing, selects some representative examples for qualitative case study, and codes the data based on relevant theoretical or analytical frameworks to offer a snapshot of multimodal analysis of EFL pedagogical materials and their appraisal meanings; for another thing, quantifies kinds of image-text relations to generalize research findings in the six EFL textbooks. The data of the current study consist of six junior high school EFL textbooks in Mainland China, edited and published by the Shanghai Education Press in 2002 and 2012, which are currently used in junior high school of Guangzhou. These series of textbooks are officially recommended by Guangzhou Education Bureau, and about 90% of public junior high school students are using them. To be more specific, two volumes for Junior 1 (Year 7, Y7, hereafter) were newly published in

2012, two volumes for Junior 2 (Year 8, Y8, hereafter), and two volumes for Junior 3 (Year 9, Y9, hereafter) were published in 2002.

I first present the distribution of visual styles in these textbooks. It is followed by the distribution of verbiage-image relations in the six textbooks, and a discussion of the thirteen categories, namely, image & text independent, image & text complementary, image subordinate to text, text subordinate to image, exposition (i&t same generality), exemplification: text more general, exemplification: image more general, extension, enhancement (temporal), enhancement (spatial), enhancement (causal/reason/purpose), locution (wording) and idea (meaning) in Martinec and Salway's (2005) framework (see Figure 1). Finally, this paper is concluded with a summary and pedagogical implications for multimodal literacy.
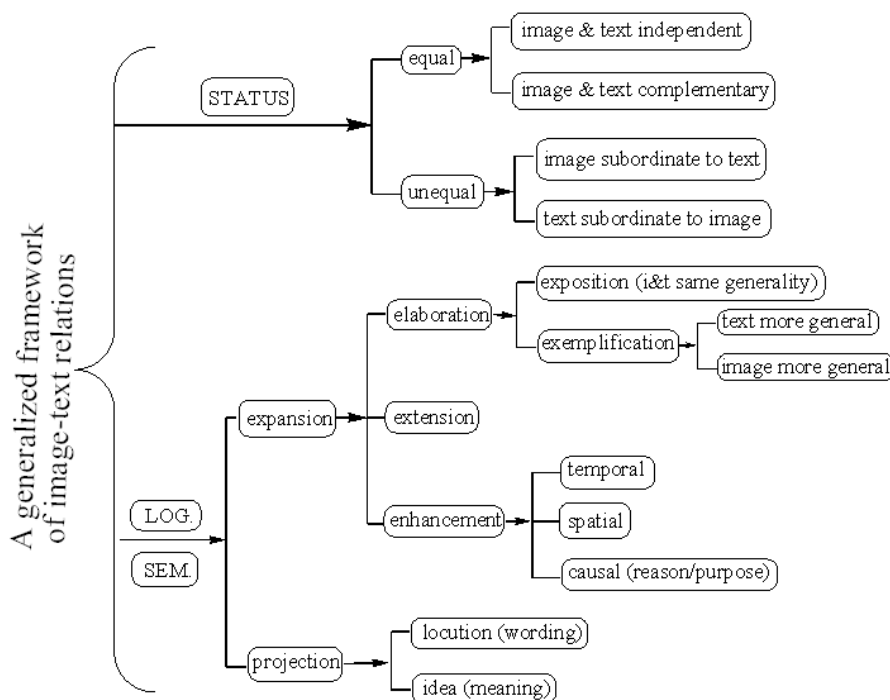

Figure 1: Network of combined status and logico–semantics (Martinec and Salway, 2005, p.358).

II. THE DISTRIBUTION OF VISUAL STYLES IN THE SIX TEXTBOOKS

In total, there are 43 teaching units covering verbal texts together with visual images in the six textbooks, and all units have visual images. To put it in another way, visual images are indispensable for junior high school students to scaffold to understand verbal texts. According to Chen's (2009) classification, visual styles can be categorized into three types: cartoon, photograph, and portrait. The distribution of visual styles in junior high school EFL textbooks is shown in Table 1, and Figure 2 displays the percentages in the form of a histogram.

TABLE 1:
THE DISTRIBUTION OF VISUAL STYLES IN THE SIX JUNIOR HIGH SCHOOL EFL TEXTBOOKS

English (Shanghai Education Press)			Cartoon	Photograph	Portrait	Total
Books		Codes				
Year 7, Book I	1st Volume	Y7-1	162 (59.1 %)	112 (40.9 %)	0 (0 %)	274 (24.8%)
	2nd Volume	Y7-2	154 (67.2 %)	75 (32.8 %)	0 (0 %)	229 (20.8%)
Year 8, Book II	1st Volume	Y8-1	160 (80.0 %)	40 (20.0 %)	0 (0 %)	200 (18.1%)
	2nd Volume	Y8-2	130 (75.1 %)	41 (23.7 %)	2 (1.2 %)	173 (15.7%)
Year 9, Book III	1st Volume	Y9-1	94 (70.7 %)	34 (25.6 %)	5 (3.7 %)	133 (12.1%)
	2nd Volume	Y9-2	65 (69.1%)	29 (30.9%)	0 (0%)	94 (8.5%)
Total			765 (69.4%)	331 (30.30%)	7 (0.6%)	1103 (100%)

Cartoon is the most popular visual style among the three, which accounts for 69.4% of the visual images used in the six textbooks, whereas photograph covers about one third (331/1103 = 30%) in the three visual styles of the six textbooks. Portrait is the least used visual style in the six textbooks, which only covers 0.6% of 1103 images. On the other hand, the distribution of visual styles decreases accordingly with the escalation of students' proficiency level and cognitive maturation, as indicated by the total number of the visual styles in each textbook. The higher the students' proficiency level, and the more cognitively mature the students, the fewer visual images and the more verbal texts in these textbooks.
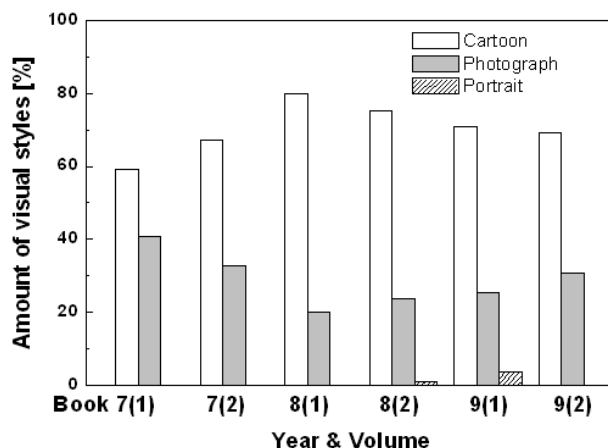
Figure 2: Distribution of visual styles in the six textbooks

This initial finding suggests that cartoon is the most acceptable and favoured form for junior high school students due to its low cognitive demand. To put it in another way, junior high school students like cartoon most because cartoon is not difficult to understand, and the cute pictures are easy for cognitive processing (Myer & Moreno, 2003). This is consistent with Paivio's (1986) limited capacity assumption. Paivio (1986) posits that visual and verbal channels are two sensory channels for humans to process information. The limited capacity assumption states that each sensory channel (either visual or verbal) has limited capacity to process cognitive information at any one time. Therefore, the juxtaposition of visual images and verbal texts in junior high school textbooks functions complementarily (Royce, 1998, 2002) to compensate for students' limited cognitive capacity of each sensory channel. Whereas, photograph covers second major proportions of visual styles in the six textbooks. As photograph captures the reality and is related to students' daily life, which may help students to understand situated contexts of verbal texts. Furthermore, with the advancement of students' grade and English proficiency, as well as cognitive maturation, the amount of cartoons and photographs reduces, while the number of portraits increases, and the portraits are about celebrities such as Charlie Chaplin and Albert Einstein, which promotes positive moral education for more cognitively developed students in later years.

There are also a large number of tables and graphical representations in the six textbooks which present teaching contents directly via the visual mode. Adapted from Vekiri's (2002) classification of graphical representations, the graphical representations in the present study consist of diagrams (e.g. schematic diagram such as water cycle), maps (e.g. geographical maps), graphs (e.g. spider-gram, pie chart, bar graphs, and line graphs), charts (e.g. tree diagrams, flow charts), and tables. But the total amount of these graphical representations does not show any remarkable curve in the six textbooks.

Based on the statistical evidence, visual images play a vital role in junior high school EFL textbooks in Mainland China, and understanding the inter-semiotic relations between visual images and verbal texts accords great significance to multimodal literacy (Chen, 2009). Students should understand intermodal relations to develop their multimodal literacy, and to understand teaching contents in a multimodal way.

III. THE DISTRIBUTION OF VERBIAGE-IMAGE RELATIONS IN THE SIX TEXTBOOKS

With reference to the aforementioned general picture of the distribution of visual styles in the six textbooks, the distribution of verbiage-image relations in these textbooks is presented with a more in-depth statistical analysis of the 13 categories of verbiage-image relations in Martinec and Salway's (2005) framework, as illustrated in Table 2 below.

TABLE 2:
THE DISTRIBUTION OF VERBIAGE-IMAGE RELATIONS IN JUNIOR HIGH SCHOOL EFL TEXTBOOKS

Books / Verbiage-image relations				Book I (Y7-1)	Book I (Y7-2)	Book II (Y8-1)	Book II (Y8-2)	Book III (Y9-1)	Book III (Y9-2)	Total
Status	equal	image & text independent		2 (1.2%)	3 (2.3%)	2 (1.6%)	5 (5.6%)	4 (5.4%)	2 (3.6%)	18 (2.8%)
		image & text complementary		48(28.7%)	30(23.0%)	43 (35.2%)	7(7.8%)	7(9.5%)	7 (12.5%)	142(22.2%)
	unequal	image subordinate to text		45(26.9%)	54(41.2%)	34 (27.9%)	42 (46.7%)	34 (45.9%)	33 (59.0%)	242(37.8%)
		text subordinate to image		3 (1.8%)	/	/	1 (1.1%)	2 (2.7%)	/	6 (0.9%)
LOG. SEM.	expansion	elaboration	Exposition (i&t same generality)	6(3.6%)	14(10.7%)	1(0.8%)	2(2.2%)	2(2.7%)	2(3.6%)	27(4.2%)
			exemplification text more general	23(13.8%)	9(6.9%)	14(11.5%)	3(3.3%)	3(4.1%)	4(7.1%)	56(8.8%)
			exemplification image more general	9(5.4%)	5(3.8%)	/	/	1(1.4%)	/	15(2.3%)
		extension		3 (1.8%)	/	/	1 (1.1%)	3 (4.1%)	/	7 (1.1%)
		enhancement	temporal	/	/	/	/	/	/	0 (0%)
			spatial	6(3.6%)	/	/	2 (2.2%)	/	1 (1.8%)	9 (1.4%)
			causal (reason/purpose)	1(0.6%)	2(1.5%)	2(1.6%)	/	/	/	5(0.8%)
	projection	locution (wording)		17(10.2%)	14(10.7%)	24 (19.7%)	26 (28.9%)	17 (23.0%)	7(12.5%)	105(16.4%)
		idea (meaning)		4 (2.4%)	/	2 (1.6%)	1 (1.1%)	1 (1.4%)	/	8 (1.3%)
Total				167 (26.1%)	131(20.5%)	122(19.1%)	90(14.1%)	74(11.6%)	56(8.8%)	640(100%)

According to the statistics demonstrated by Table 2, the total amount of instances which present image-text relations in each textbook is 167 (26.1%) in Y7-1, 131 (20.5%) in Y7-2, 122 (19.1%) in Y8-1, 90 (14.1%) in Y8-2, 74 (11.6%) in Y9-1, and 56 (8.8%) in Y9-2 respectively. In addition, Table 3.1 illustrates that percentage of "image & text complementary" decreases from 28.7% in Y7-1 to 12.5% in Y9-2, whilst percentage of "image subordinate to text" increases from 26.9% in Y7-1 to 59.0% in Y9-2. It suggests that textbook designers focus more on texts in textbooks of higher grade. The finding indicates that with the students' grade advancement and improved English proficiency, the amount of verbiage-image relations significantly decreases. These instances are then mapped onto Martinec and Salway's (2005) categories of visual-verbal relations. The frequency and ranking of each category is illustrated in Table 3.

TABLE 3:
RANKING THE TOTAL AMOUNT OF EACH CATEGORY OF IMAGE-TEXT RELATIONS IN THE SIX TEXTBOOKS

Ranking	Categories of image-text relations	Total amount in the six textbooks
1	Image subordinate to text	242(37.8%)
2	Image & text complementary	142(22.2%)
3	Locution (wording)	105(16.4%)
4	Exemplification: text more general	56(8.8%)
5	Exposition (i&t same generality)	27(4.2%)
6	Image & text independent	18(2.8%)
7	Exemplification: image more general	15(2.3%)
8	Enhancement (spatial)	9(1.4%)
9	Idea (meaning)	8(1.3%)
10	Extension	7 (1.1%)
11	Text subordinate to image	6 (0.9%)
12	Enhancement (causal/reason/purpose)	5(0.8%)
13	Enhancement (temporal)	0 (0%)

The following subsections illustrate each of the above 13 categories by drawing on representative examples from the six textbooks in my dataset.

A.   Unequal Status in Verbiage-image Relations

According to Martinec & Salway's (2005) classification, when image and text are in an unequal status in verbiage-image relation, it can be in either an "image subordinate to text" relation, or a "text subordinate to image" one.

1. Image Subordinate to Text

Table 3 indicates that "image subordinate to text" makes up a major proportion - more than one third (242/640=37.8%) in the image-text relations of the six textbooks. Image subordination can be defined as an inter-semiotic relationship between visual images and verbal texts "when the image only relates with a part of the text" (Martinec & Salway, 2005, p. 348), but not all of the text. Therefore, the relationship between the image and the text is 'unequal'. Figure 3 below is an example illustrating such an unequal relationship, which is extracted from *English* Year 7 (2nd Volume) Unit 1, p.3.
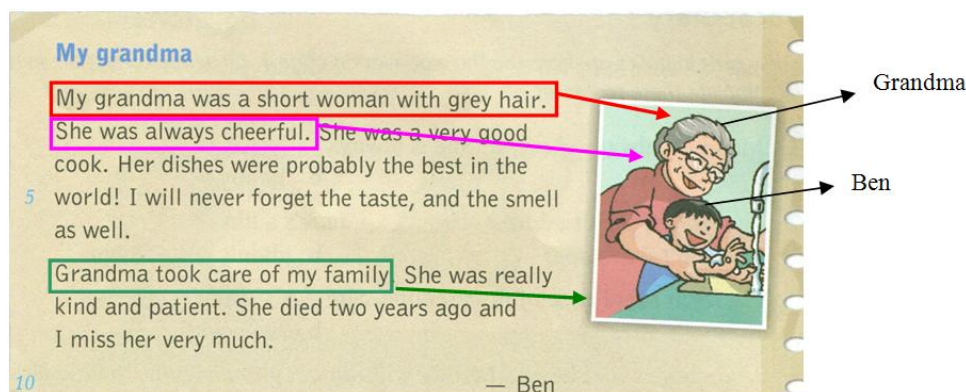

Figure 3: Example of "image subordinate to text" [from *English* Year 7 (2nd Volume) Unit 1, p.3]

The picture in Figure 3 depicts two cartoon characters Ben and his grandma, and his grandma assisted him in washing his hands at the wash basin. His grandma was not tall with grey hair. Both of them looked happy. What is depicted in the picture frame relates with three sentences in verbal texts "My grandma was a short woman with grey hair. She was always cheerful." and "Grandma took care of my family." For the remaining texts, there is no picture to illustrate the point. Image subordination plays a dominant role in the image-text relations of the six textbooks. Perhaps due to the limited space and junior high school students with certain-level cognitive ability, textbook editors cannot and do not necessarily print out all pictures of teaching contents, but just provide some vital relevant pictures in the visual-verbal relation type of "image subordinate to text" to sketch out the key theme of certain teaching unit and offer

some background information of teaching contents. In this way, readers can react emotionally to pictures for their low cognitive demand (Barry, 1997), and obtain an idea of what they are going to learn in certain unit whenever they process the pictures in the six textbooks. Thus, the visual images in the inter-semiotic relation of image subordination attract learners' attention, and achieve "immediacy" and "transparency" (Stockl, 2004) in the semiotic communication. They offer some hints or clues of teaching contents, and play a heuristic role in language learning. The facilitating effects of visual images on verbal textual comprehension have been empirically testified by many researchers (e.g., Pan & Pan, 2009).

2. Text Subordinate to Image

Another kind of unequal status in image-text relations is "text subordinate to image". Table 3 displays that "text subordinate to image" ranks eleventh in all verbiage-image relations in the six textbooks, and nearly all the instances of "text subordinate to image" can be further categorized as "extension". Text subordination can be "realized by deixis from text to image, either by reference or present tense combined with material or behavioural process" (Martinec and Salway, 2005, p.348). For example, Figure 4 is extracted from *English* Year 8 (2nd Volume), Unit 3, p.46. The readers, junior high school students, are asked to read the notes below and write a paragraph about Charlie Chaplin.
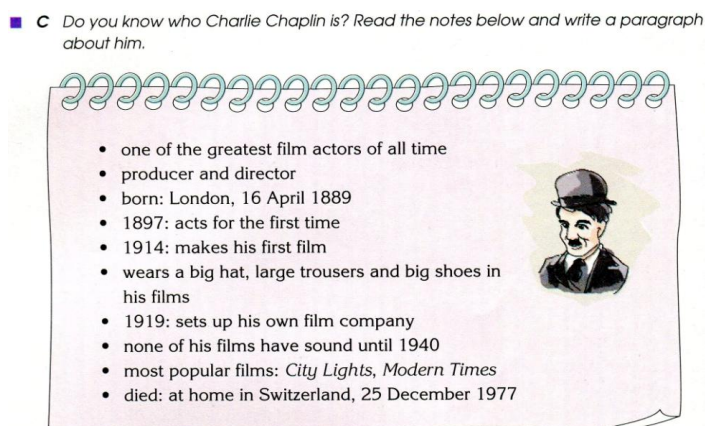

Figure 4: Example of "text subordinate to image: extension" [from *English* Year 8 (2nd Volume) Unit 3, p.46]

In Figure 4, the whole text relates with the portrait of Charlie Chaplin which visually presents his self-image in a comic style that complies with his humorous stereotype to enhance intimacy and cordiality with putative readers. On the other hand, verbal texts brief readers on Chaplin's contributions, occupation, year of birth, typical features, milestones in his life, representative works, and year of death and nationality. "Born: London, 16 April 1889; 1897: acts for the first time; 1914: makes his first film; 1919: sets up his own film company; died: at home in Switzerland, 25 December, 1977" add new, related background information to the content of the image of Charlie Chaplin. Furthermore, the verbal texts of "one of the greatest film actors of all time, producer and director; none of his films have sound until 1940; most popular films: *City Lights, Modern Times*" indicate the context of his life and representative works. What is mentioned above "goes beyond what is represented in the image, and beyond its participants, processes, and circumstances" (Martinec & Salway, 2005, p.350). What the image represented is "wears a big hat", while "wears large trousers and big shoes in his film" goes beyond what is represented in the image. "Text subordinate to image" relationship is "realized by deixis from text to image, either by reference or present tense combined with material or behavioural process" (Martinec & Salway, 2005, p.348), such as using present tense with material or behavioural processes like "acts, makes, and sets up". In the data, the verbiage-image relation of "text subordinate to image: extension" usually appears in the introduction of famous people, mostly in Book II or III, which is in the form of a portrait of celebrities on the right and the verbal text of their glorious history on the left. Although visual images are easier to access for sensory perception and cognition, they fail to offer a historical account for celebrities. The weaknesses of visual images are thus compensated with verbal texts. Both visual images and verbal texts simultaneously play a complementary (Royce, 1998, 2002) role in multimodal meaning-makings and create a semiotic harmonious landscape in junior high school EFL textbooks (Chen, 2009).

B. Equal Status in Verbiage-image Relations

While the previous subsection illustrates how words and images in textbooks may be unequal in status, there are also instances where they relate with each other in an equal manner, either independently or complementarily.

1. Image & Text Independent

From Table 3, we can see that there are not many "image and text independent" relations in the six textbooks, which ranks sixth in all verbiage-image relations of these textbooks. According to Martinec & Salway (2005), "image and text independent" is an inter-semiotic relation between an image and a text in which the whole image relates with the whole text in an equal and independent manner. For instance, Figure 5 was taken from *English* Year 7 (2nd Volume), Unit 5, p.58.

Figure 5: Example of "image & text independent" [from *English* Year 7 (2nd Volume), Unit 5, p.58]

The whole image in Figure 5 relates with the whole text "It runs into rivers. It rises from the sea to the sky. It runs into the sea. It falls as rain again. Water falls from the clouds as rain." in an equal and independent way. This kind of image-text relation usually appears in the units which elaborate on how a natural phenomenon happens such as the recycling of water. The juxtaposition of visual images and verbal texts co-deploys allows the readers to easily understand the abstract principle of the nature. Images in textbooks can shape students' sensory perceptions of the represented world with their more "topological meaning potentials" (Lemke, 1998), which is distinguished from the way languages work. Halliday (1978) states that "semiotic resources such as images and language are systems of meaning potentials" (p.39). However, images and language differ in their demanded cognitive operations, because "images are based on simultaneous and holistic gestalt-perception", while "language is a linear mode that calls for the successive integration of signs into phrases" (Stockl, 2004, p.17). Therefore, image is perceived as a quick mode than language which needs parsing.

2. Image & Text Complementary

Another kind of equal status in verbiage-image relations is called "image and text complementary". Table 3 shows that "image & text complementary" ranks second. According to Martinec & Salway (2005), the complementary image-text relation should be on the condition that the whole image relates with the whole text in an equal manner.

On the contrary, all kinds of expansion (see Table 2) can be categorized as "image and text complementary". In the multimodal analysis of verbiage-image relations in the six textbooks, nearly all image-text relations of "exposition (i&t same generality)", "exemplification: text more general", "exemplification: image more general", "extension", "enhancement: spatial", and "enhancement: causal (reason/purpose)" could be categorized as "image and text complementary" (Royce, 1998, 2002; Martinec & Salway, 2005). This is because simultaneous activation of readers' two or more sensory channels (i.e., visual and verbal) can be an effective way to reduce readers' cognitive load and to achieve pedagogical purposes, which is proved by some psychological experiments that "images are far more likely to be attention-getters in perception than language and can also be memorized much more easily and effectively" (Stockl, 2004, p.17). Because pictorial perception is considered to be much more easily accessible for its "simultaneity of gestalt formation" than language perception, without much parsing, "images directly tap into the emotions and provide immediate sensory input" (Stockl, 2004, p.17).

Martinec and Salway (2005) extend Halliday's (1994) functional account of logical-semantic relations in terms of experiential meaning into the analysis of visual-verbal inter-semiotic relations. In terms of more specific inter-dependency, the bi-modal logical-semantic relations at the grammar stratum perform the function of relating images with texts via relations of elaboration, extension, and enhancement, which will be discussed in the following sections.

C. Elaboration in Logical-semantic Image-text Relations

According to Martinec and Salway's (2005) framework, logical-semantic image-text relations consist of projection and expansion. The first kind of expansion in logical-semantic image-text relations is elaboration. In elaborating image-text relations, "some part of the main text, caption, or label may expand an aspect of an image or vice versa by restating or representing the same thing in a different semiotic resource, specifying an aspect in greater detail, refining it or making something more specific" (Zhao, 2011, p.146). Elaboration comprises exposition and exemplification. If images and texts have the same kind of generality, the image-text relation is exposition (i&t same generality). If image and text have different kinds of generality, the image-text relation is exemplification, either 'text more general' or 'image more general'. Table 3 demonstrates that "exposition (i&t same generality)" ranks fifth (27, 4.2%) in all image-text relations of the six textbooks. Exposition can be defined as a relationship between an image and a text in which they each plays a role in a relational identifying process (Halliday, 1994), and each has the same levels of

generality or abstraction. Martinec and Salway (2005) further state that the text has the tendency to be realized by a nominal group while the relational identifying process tends to be implicit. Whereas, "exemplification: text more general" accounts for 8.9% in all verbiage-image relations of the six textbooks. It suggests that textbook editors take junior high school students' cognitive features into consideration, and they provide concrete examples to illustrate abstract concepts or terminologies in order to reduce cognitive demand for targeted readers (Myer & Moreno, 2003).

D. Extension in Logical-semantic Image-text Relations

The third kind of expansion in logical-semantic image-text relations is extension. Table 3 illustrates that there are only a few examples of "extension" (7, 1.1%) among all verbiage-image relations of the six textbooks. Extension can be defined as a relationship between an image and a text in which "one mode extends or adds new information to the other mode, because it goes beyond what is represented in the image, and beyond its participants, processes, and circumstances" (Martinec & Salway, 2005, p.350). For example, Figure 4 illustrates the verbal text of Chaplin's historical events and drawing of his image are semantic and semiotic complementarity, which fall into the category of "semiotic division of labor" (Matthiessen 2007, p.46).

E. Enhancement in Logical-semantic Image-text Relations

Another kind of expansion in logical-semantic image-text relations is enhancement which is further distinguished by temporal, spatial, and causal (reason/purpose).

1. Enhancement: spatial

Table 3 indicates that "enhancement: spatial" ranks eighth in all image-text relations of the six textbooks. Enhancement can be defined as a relationship between an image and a text in which "one mode enhances the other mode by referencing it with circumstantial information like a time, a place, a reason, and a purpose" (Martinec & Salway, 2005, pp.350-351). A text enhancing an image or an image enhancing a text should be associated with its ideational content. To illustrate, Figure 6 was taken from *English* Year 7 (1st Volume) Unit 5, p.65.
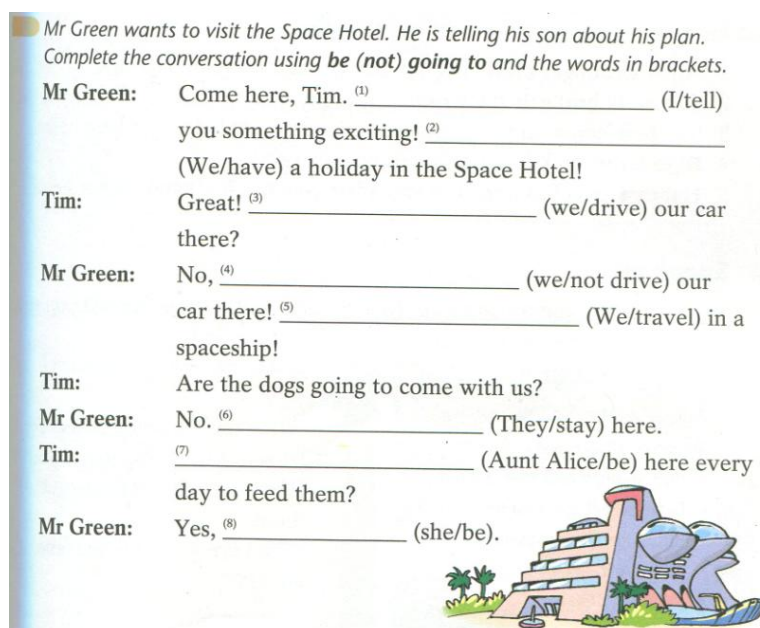

Figure 6: Example of "enhancement: spatial" [from *English* Year 7 (1st Volume) Unit 5, p.65]

In Figure 6, the image of "space hotel" relates with the text by enhancement, and the image of "the Space Hotel" qualifies the text circumstantially. To be more specific, the image of "the Space Hotel" enhances the text by space, and the image specifies the place where Mr. Green wants to visit. There are a few other examples of "enhancement: spatial" in all of image-text relationship of the six textbooks, and they usually appear in the 'speaking' section of some units.

2. Enhancement: causal (reason/purpose)

Only a small proportion (5, 0.8%) of "enhancement: causal (reason/purpose)" image-text relations have been identified in the six textbooks. Figure 7 demonstrates the image-text relation of "enhancement: causal (reason/purpose)" which is extracted from *English* Year 7 (2nd Volume) Unit 6, p.84.

Figure 7: Example of "enhancement: causal (reason/purpose)" [from *English* Year 7 (2nd Volume) Unit 6, p.84]

In Figure 7, the image enhances the text by way of 'causal (reason/purpose)', as the visual image depicts that people turned their lights off for one hour but burned the candle in order to show that they cared about climate change. This enhances the meaning denoted by the verbal text of 'Earth Hour'. This is one example to counter-prove Stockl's (2004) statement on some semantic limitations of visual images that "some meaning relations like causality cannot be expressed by images" (p.18).

3. Enhancement: temporal

From Table 3, there is no "enhancement (temporal)" image-text relation in the six textbooks, perhaps because time is an abstract concept and it is less than easy to display physically.

F.   Locution (Wording) vs Idea (Meaning)

"Locution (wording)" and "idea (meaning)" are two kinds of projection in logical-semantic image-text relations. Table 3 shows that "locution (wording)" ranks third (105, 16.4%) among 13 categories of verbiage-image relations in the six textbooks. In the image-text relations in Martinec and Salway's (2005) framework, "locution (wording)" is one of two kinds of projection which repeats or projects the information in one mode from the other mode via "talking bubble" linking with a sayer, and another kind of projection is information transmission inter-modally via "thinking cloud" linking with a thinker. Two respective examples were extracted from *English* Year 7 (1st Volume) Unit 2, p.15, and *English* Year 8 (1st Volume) Unit 5, p.71.
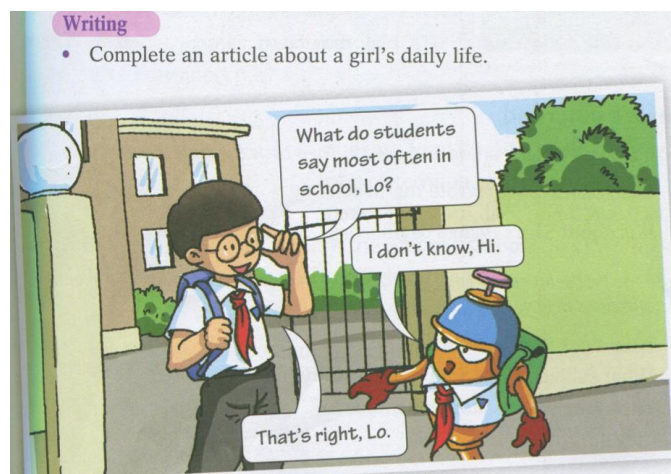

Figure 8: Example of "locution (wording)" [from *English* Year 7 (1st Volume) Unit 2, p.15]
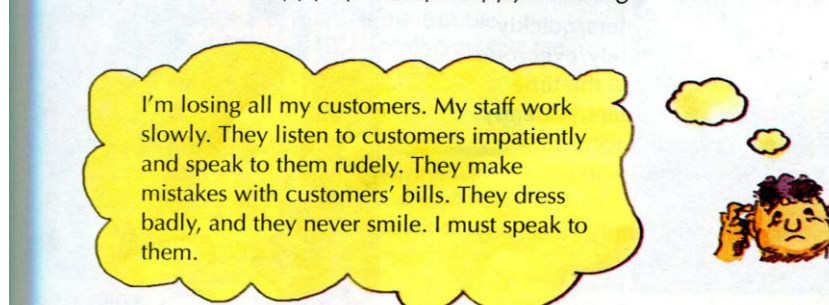

Figure 9: Example of "idea (meaning)" [from *English* Year 8 (1st Volume) Unit 5, p.71]

The image-text relation in Figure 8 exemplifies locution (wording), because information transmission from the visual mode (speaker) to the verbal mode (utterance) is achieved through "speech bubble", while Figure 9 performs the function of information projection from the visual mode (thinker, usually a person using his/her finger to scratch his/her head) to the verbal mode (what the thinker is thinking) via "thought cloud". From Table 3, we can see that "locution (wording)" is the third major type in the verbiage-image relations of the six textbooks, and they are especially evident on the 'herald' page of each unit to predict what is going to be learned in this unit. Textbook editors use cartoons to elicit the theme of a unit and to present the locution (wording) via "speech bubble" in order to catch readers' attention on the theme and make it easier for student readers to process verbal texts through visual images. On the contrary, there are a total of only eight "idea (meaning)" image-text relations, which ranks ninth in all verbiage-image relations of the six textbooks. This is perhaps due to the fact that for junior high school students, their cognitive ability is neither too low as primary pupils nor too high as senior high school students. Thus, textbook editors would be in no position to impose too many "thought cloud" on junior high school EFL textbooks.

IV. SUMMARY

This study has attempted to answer the research question, "What are the multimodal relations between verbal language and visual images in junior high school EFL textbooks in Mainland China (with Guangzhou as a featured case)?" by referring to Martinec and Salway's (2005) framework of 13 categories of verbiage-image relations, namely, image & text independent, image & text complementary, image subordinate to text, text subordinate to image, exposition (i&t same generality), exemplification: text more general, exemplification: image more general, extension, enhancement (temporal), enhancement (spatial), enhancement (causal/reason/purpose), locution (wording) and idea (meaning).

After presenting the distribution of visual styles and verbiage-image relations in the six textbooks, and illustrating the 13 categories of image-text relations one by one, the findings demonstrate that the total amount of instances which present image-text relations in each textbook is 167 (26.1%), 131 (20.5%), 122 (19.1%), 90 (14.1%), 74 (11.6%), and 56 (8.8%) respectively. These figures suggest that with students' grade advancement and improved English proficiency, the amount of verbiage-image relations significantly decreased, and textbook design changes gradually from image-based to text-based. "Image subordinate to text" (242, 37.8%), "image & text complementary" (142, 22.2%), and "locution (wording)" (105, 16.4%) are the three major types among the 13 categories of image-text relations in the six junior high school EFL textbooks. The visual-verbal inter-semiotic resources not only have distinctive "functional specification" (Bezemer & Kress, 2008), but complement and co-instantiate (Chen, 2009) with each other, which maximize this bi-modal "epistemological potential" (Bezemer & Kress, 2008). Dialogue relations are visually presented via dialogue balloons or locution (wording), which is a more acceptable form for junior high school students.

There is, however, no "enhancement (temporal)" image-text relation in the six textbooks, perhaps because time is an abstract concept that is not easy to display physically. There are a small number of remaining types of image-text relations such as exemplification: text more general (56, 8.8%), exposition (i&t same generality, 27, 4.2%), image & text independent (18, 2.8%), exemplification: image more general (15, 2.3%), enhancement (spatial) (9, 1.4%), idea (meaning) (8, 1.3%), extension (7, 1.1%), text subordinate to image (6, 0.9%), and enhancement (causal/reason/purpose) (5, 0.8%). These figures suggest that EFL textbooks for junior high school students need visual images to assist them in understanding verbal texts due to their immature cognitive ability (Myer & Moreno, 2003), relatively low English proficiency, and interest in visual images. The current textbook design of image-text relations facilitates students' learning and development of multimodal literacy based on students' English proficiency and cognitive maturation. Visual images play a vital role in junior high school EFL textbooks. Better understanding of image-text relations is not only beneficial for foreign language learners, but helpful for EFL professionals and textbook designers to cultivate students' multimodal literacy and design user-friendly multimodal textbooks. Comparatively speaking, textbook designers focus too much on the intermodal relations of "image subordinate to text", while other types of image-text relations such as "exemplification: text more general", "exemplification: image more general", "enhancement: causal (reason/purpose)", and "idea (meaning)" should be used more, as junior high school students' immature cognitive ability, they need more examples and visual images to scaffold to understand abstract concepts, logical reasoning, and implicitly articulated meanings.

The present study has made the first attempt to test the applicability of Martinec and Salway's (2005) framework to detail visual-verbal relations into 13 categories in a mini-corpus of the six junior high school EFL textbooks in Mainland China, and adopted quantitative measurement of various types of visual-verbal relations to evaluate findings in a theoretical way. It is a general trend that with students' grade advancement, improved English proficiency, and their cognitive maturity, the total amount of various types of visual styles, image-text relations, decreased accordingly, and junior high school EFL textbooks changed gradually from image-based to text-based. It can be inferred that students' grade, English proficiency, their interest in visual images, and their cognitive maturity play decisive roles in the design of EFL multimodal textbooks.

V. PEDAGOGICAL IMPLICATIONS FOR MULTIMODAL LITERACY

In light of the above discussion of major findings, we can infer that visual images play subsidiary and complementary roles in students' verbal text understanding, and provide some background information or direct sensory perception for students to understand verbal texts. Visual images are conducive for easy sensory processing and representing physical objects and space. Visual images act as emotional motivator and "attention-getter" (Stockl, 2004, p.17) to engage readers in targeted texts before verbal texts can be cognitively understood. The effects of visual images are empirically measured by many scholars. Pena and Quilez (2001) found that the use of visual images improved students' memorizing verbal texts. For second language learners, even though they do not know the meanings of some new words in their reading, the images offer some hints for them to guess the gist of the texts. The visual images facilitate or enhance textual comprehension for EFL learners (Pan & Pan, 2009). This is also consistent with Johnson-Laird's (1983) theory of mental models which states that visual images facilitate the construction of mental structure, and they "present basic information more concisely than equivalent textual statement and reduce the cognitive load in complex tasks" (p.66).

On the other hand, verbal texts have the advantages of depicting states-of-affairs and events. Verbal texts usually further specify "what is conveyed via images" in the verbal mode. Some verbal texts uttered by interlocutors are presented via speech bubbles, and these more acceptable cute pictures function to arouse young viewers' attention (Stockl, 2004), engage readers in targeted texts, and achieve "immediacy" and "transparency" (Stockl, 2004). In addition, with the upgrading of students' grade and English proficiency, in general, the amount of verbiage-image relations generally decreased. With this statistical evidence, assumption can be put forward that with the maturation of students' cognitive ability, the upgrading of their English proficiency, and their interest in visual images, students' understanding of verbal texts can be achieved without the help of visual images, they less rely on visual images to understand verbal texts.

Both visual images and verbal language "co-contextualize" (Liu & O'Halloran, 2009) to achieve multimodal meaning-makings in their distinctive and complementary (Royce, 1998, 2002) manners in the six textbooks. Visual and verbal modes have their relatively distinctive role in meaning-makings, meanwhile, they are also inseparably intertwined with each other in harmonious forms of visual communication -"mode mixing" (Stockl 2004, p.18). The appropriate arrangements of visual images and verbal texts are conducive to reducing cognitive demand (Myer & Moreno, 2003), which is especially suitable in the design of EFL textbooks. Better understanding of the inter-semiotic relations between visual images and verbal texts helps textbook designers to design more suitable textbooks, helps teachers better elicit teaching contents, and helps to scaffold learners' acquisition of multimodal literacy (Chen, 2009).

To summarize, the study of visual-verbal inter-semiotic construal in EFL textbook discourse has some pedagogical implications.

First, EFL teachers should provide tier guidance or affordance for students based on their English proficiency. For instance, for Junior 1 students, teachers can elicit the teaching content with more pictures, while for Junior 3 students, due to their higher English proficiency, more mature cognitive ability, and less interest in visual images, teachers can introduce more abstract terms without the help of visual images.

Second, textbook designers should be aware of multimodal literacy and different image-text relations, and take students' cognitive maturation, English proficiency, and their interest in visual-verbal modes into consideration while arranging visual images and verbal language of the textbook in an appropriate way. In this way, they won't add images just for the sake of adding images. They should optimize the inter-semiotic affordances of visual-verbal "functional specification" and "epistemological potential" (Bezemer & Kress, 2008).

APPENDIX I.   MAIN CARTOON CHARACTERS IN *ENGLISH* YEAR 7, 1ST VOLUME
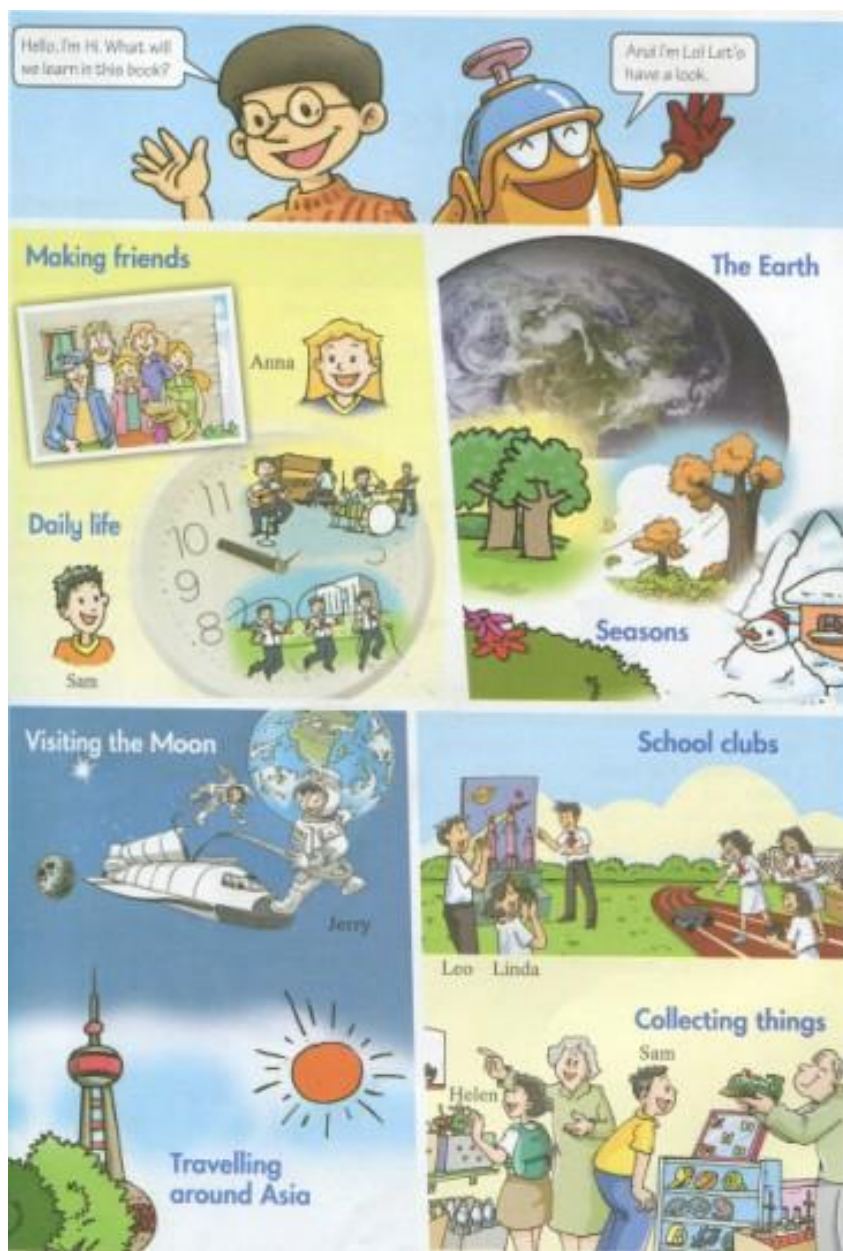
APPENDIX II.   MAIN CARTOON CHARACTERS IN *ENGLISH* YEAR 7, 2ND VOLUME

APPENDIX III.   THE OUTLINE OF JUNIOR HIGH SCHOOL EFL TEXTBOOK (*ENGLISH* YEAR *7*, 1ST VOLUME)

Module	Unit
1 My life	1 Making friends
	2 Daily life
Project 1 Our new classmates	
2 The natural world	3 The earth
	4 Seasons
Project 2 Our home town	
3 Travels	5 Visiting the Moon
	6 Travelling around Asia
Project 3 A travel plan	
4 Fun time	7 School clubs
	8 Collecting things
Project 4 A survey about free time activities	

APPENDIX IV.   THE OUTLINE OF JUNIOR HIGH SCHOOL EFL TEXTBOOK (*ENGLISH* YEAR *7*, 2ND VOLUME)

Module	Unit
1 People and place	1 People around us
	2 Travelling around the world
Project 1 Cities of the world	
2 Man's best friends	3 Our animal friends
	4 Save the trees
Project 2 Planning green activites	
3 Natural elements	5 Water
	6 Electricity
Project 3 Saving water and electricity	
4 Colorful life	7 Poems
	8 From hobby to career
Project 4 Who has the same hobby as me?	

APPENDIX V.   THE OUTLINE OF REMAINING JUNIOR HIGH SCHOOL EFL TEXTBOOKS (SHANGHAI EDUCATION PRESS)

English	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6	Unit 7
Year 8, 1st Volume	Newspaper	Detectives and crimes	Modern machines	Historical stories	Strange creatures	Jobs	Environment
Year 8, 2nd Volume	Success stories	Cartoons and comics	All about films	Educational exchanges	Family life	The Adventures of Tom Sawyer	Mind and memory
Year 9, 1st Volume	Body language	Hair care	Health and food	What should I do?	Going to the theatre	Great minds	Plants
Year 9, 2nd Volume	Consumer society	Sports	Fashion and style	All about learning	Friends or enemies	Entertainment	_____

REFERENCES

[1]   Barthes, R. (1977). Rhetoric of the image. In R. Barthes (Ed.), *Image-Music-Text.* Heath, S. (trans.) (pp.32-51). London: Fontana.
[2]   Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication, 25*, 166-195.
[3]   Chen, Y.M. (2009). Interpersonal meaning in textbooks for teaching English as a foreign language in China: a multimodal approach. Unpublished PhD thesis. University of Sydney.
[4]   Halliday, M.A.K. (1994/2000). An Introduction to Functional Grammar (2nd Ed.). London: Edward Arnold (Publishers) Ltd/Beijing: Foreign Language Teaching and Research Press.
[5]   Jewitt, C. (2009). Different approaches to multimodality. In C. Jewitt (Ed.), (2009). *The Routledge handbook of multimodal analysis* (pp. 28-39). London: Routledge.
[6]   Johnson-Laird, P.N. (1983). Mental models: Towards a cognitive science of language, inference and consciousness. Cambrige, MA: Harvard University.
[7]   Kong, K. (2006). A taxonomy of the discourse relations between words and visual. *Information Design Journal, 14*(3), 207-230.
[8]   Kress, G., & van Leeuwen, T. (1996/2006). Reading Images: The Grammar of Visual Design. London/New York: Routledge.
[9]   Lemke, J.L. (1998). Multiplying meaning: visual and verbal semiotics in scientific text. In J.R. Martin, & R. Veel (Eds.), *Reading Science: Critical and Functional Perspectives on Discourses of Science (pp.87-113).* London: Routledge.
[10]   Liu, Y., & O'Halloran, K. (2009). Intersemiotic texture: Analyzing cohesive devices between language and images. *Social Semiotics, 19*(4), 367-388.

[11] Liu, X., & Qu, D. (2014). Exploring the multimodality of EFL textbooks for Chinese college students: A comparative study. *RELC Journal, 45*(2), 135-150.

[12] Martinec, R., & Salway, A. (2005). A system for image-text relations in new (and old) media. *Visual Communication.* 4(3), 337-371.

[13] Matthiessen, C.M.I.M. (2007). The multimodal page: A systemic functional exploration. In Royce, T. and Bowcher, W.L. (Eds). *New Directions in the Analysis of Multimodal Discourse (pp.1-62).* Mahwah, NJ: Lawrence Erlbaum Association.

[14] Myer, R.E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38 (1), 43-52.

[15] Nordensvard, E. (2010). Multimodality: An EFL textbook comparison using multimodal discourse analysis. Sodertorn University College, Sweden.

[16] 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.

[17] O'Halloran, K. L. (2011). Multimodal discourse analysis. In K. Hyland, & B. Paltridge (Eds.), *Companion to discourse analysis* (pp. 120-137). London: Continuum.

[18] Pan, Y.C., & Pan, Y.C. (2009). The effects of pictures on the reading comprehension of low-proficiency Taiwanese English foreign language college students: An action research study. *VNU Journal of Science, Foreign Languages.* 25, 186-198.

[19] Painter, C., & Martin, J.R. (2011). Intermodal complementarity: Modelling affordances across image and verbiage in children's picture books. In Fang Yan (Ed.) *Studies in Functional Linguistics and Discourse Analysis (pp.132-158).* Beijing: Education Press of China.

[20] Paivio, A. (1986). Mental Representations: A Dual Coding Approach. Oxford, England: Oxford University Press.

[21] Pena, B.M., & Quilez, M.J.G. (2001). The importance of images in astronomy education. *International Journal of Science Education, 23*(11), 1125-1135.

[22] Royce, T. (1998). Synergy on the page: Exploring inter-semiotic complementarity in page-based multimodal text. In JASFL Occasional Papers No. 1 (pp. 25-49). Tokyo: Japan Association of Systemic Functional Linguistics (JASFL).

[23] Royce, T. (2002). Multimodality in the TESOL classroom: Exploring vsual-verbal synergy. *TESOL Quarterly, 36*(2), 191-205.

[24] Royce, T. (2006). Inter-semiotic complementarity: a framework for multimodal discourse analysis. In T. D. Royce, & W. L. Bowcher. (Eds.) *New Directions in the Analysis of Multimodal Discourse (pp.63-109).* Mahwah, New Jersey, London: Lawrence Erlbaum.

[25] Salbego, N., Heberle, V.M., & Maria Gabriela Soares da Silva Balen. (2015). A visual analysis of English textbooks: Multimodal scaffolded learning. *Calidoscopio, 13*(1), 5-13.

[26] Stöckl H. (2004). In between modes: Language and image in printed media. In E. Ventola, C. Charles, & M. Kaltenbacher, (Eds.) *Perspectives on Multimodality* (pp. 9-29). Amsterdam: John Benjamins.

[27] Tahririan, M.H. & Sadri, E. (2013). Analysis of images in Iranian high school EFL course books. *Iranian Journal of Applied Linguistics (IJAL),* 16(2), 137-160.

[28] Torres, G. (2015). "Reading" *World Link*: A visual social semiotic analysis of an EFL textbook. *International Journal of English Language Education, 3*(1), 239-253.

[29] Unsworth, L. (2006). Towards a meta-language for multi-literacies education: Describing the meaning-making resources of language-image interaction. *English Teaching: Practice and Critique, 5* (1), 55-76.

[30] Unsworth, L. (2007). Image-text relations and intersemiosis: Towards multimodal text description for multi-literacies education. In B. Leila, & T. Berber-Sarinha (Eds.) *Proceedings 33rd International Systemic Functional Congress (pp.1165-1205).* PUCSP, Sao Paulo, Brazil.

[31] van Leeuwen, T. (2005). Introducing Social Semiotics. London & New York: Routledge.

[32] Vekiri, I. (2002). What is the value of graphical displays in learning? *Educational Psychology Review*, *14* (3), 261-312.

[33] Vorvilas, G. (2014). Logical meanings in multimedia learning materials: A multimodal discourse analysis. *Acta Didactica Napocensia, 7*(2), 25-40.

[34] Weninger, C., & Kiss, T. (2013). Culture in English as a foreign language (EFL) textbooks: A semiotic approach. *TESOL Quarterly,* 47, 694-716.

[35] Yassine, S. (2014). Multimodal design of EFL textbooks: A social semiotic multimodal approach. *Anglisticum Journal (IJLLIS), 3*(12), 84-90.

LAI Haiyan (Kelly) is a final-year PhD candidate major in English Language Education at the Faculty of Education, The University of Hong Kong. Her research interest includes Critical/Multimodal Discourse Analysis; Academic Writing; Content and Language Integrated Learning. Her recent work appeared in Cognitive Linguistics studies.