

## **Multimodal Analysis of Image-Text Relations in the PEP Senior High English Textbooks**

**Jia Xu\***

*School of Foreign Languages, Liaocheng University, Liaocheng, Shandong, China*

*\*Corresponding Author*

**Abstract:** With the rise of multimodal teaching concepts, visual resources in English textbooks have become important teaching media. Illustrations are not only visual representations of textual content but also essential components of multimodal teaching resources. However, the phenomenon of "emphasizing text while neglecting images" remains prevalent in current senior high school English teaching, and the multimodal teaching potential of illustrations has yet to be fully exploited. Based on Martinee & Salway's image-text relation theory, this study conducts quantitative statistics and qualitative analysis of the image-text relationships in the People's Education Press (PEP) Senior High School English compulsory textbooks. The findings reveal that the proportion of co-occurring images and text is relatively high in the textbooks, with 74.1% demonstrating equal status between image and text, and 25.9% showing unequal status. Therefore, teachers should fully recognize the pedagogical value of illustrations, design teaching strategies according to different image-text relationship characteristics, and fully tap into the teaching potential of multimodal resources to promote the comprehensive development of students' language abilities and thinking qualities.

**Keywords:** Multimodal Analysis; Image-Text Relations; PEP Senior High English Textbooks; Textbook Illustrations; Multimodal Teaching

### **1. Introduction**

With the development of human social activities and the increasing abundance of material means, traditional mono-modal linguistic expression can no longer meet the demands of modern society. As Li & Lu point out, composite discourse involving multiple media has gradually become mainstream [1].

With multimedialization becoming the norm of social practice, multimodality constitutes an inherent characteristic of contemporary sociocultural systems. This contemporary context has profoundly influenced the field of education. The 2017 Edition of English Curriculum Criteria for Regular High School explicitly states that discourse types include not only traditional genres such as narratives and argumentative essays, but also multimodal texts [2]. This implies that in English textbooks, attention should be paid not only to the written language but also to the illustrations within the textbooks.

However, in current senior high school English teaching, the instructional value of textbook illustrations has not yet been fully recognized. Many teachers still focus on the written text during instruction, neglecting the important role of illustrations in multimodal teaching, which has led to the widespread phenomenon of "prioritizing text over images." This not only limits students' ability to comprehend multimodal texts but also fails to utilize the unique value of illustrations in deepening textual understanding, enriching aesthetic experiences, and enhancing educational functions. As Cheng & Cong note, visual resources such as images play a crucial role in various learning activities. Multiliteracy has become an essential component of communicative competence [3]. Students need to develop the ability to interpret images and communicate through visual means. As an important component of multimodal texts, illustrations in textbooks provide students with intuitive visual information, helping them better understand textual content, stimulate learning interest, and cultivate cross-cultural awareness and critical thinking skills. Therefore, conducting a study of the relationship between images and text in textbooks and uncovering the potential instructional value of illustrations is

of great significance for promoting the transformation of senior high school English teaching from a traditional mono-modal linguistic approach to a multimodal teaching model. Based on this, this paper takes the PEP Senior High School English compulsory textbooks as the research subject and employs image-text relationship theory to conduct an in-depth analysis of the relationship between images and text in the textbooks. The aim is to investigate the specific distribution and proportion of various image-text relationships within the textbook units, providing useful references and insights for the compilation and teaching of senior high school English textbooks.

## 2. Literature Review

"Mode" refers to the medium employed by humans in information transmission, encompassing linguistic text, gesture, image, etc [4]. Among these, a "monomodal" text relies on a single semiotic system, whereas a "multimodal" text integrates two or more distinct modal resources to achieve communication and meaning construction. In short, a multimodal text is a form of discourse that encodes and transmits meaning through multiple semiotic systems. The development of multimodal discourse has laid a theoretical foundation for image-text relationship analysis, making the interaction between images and text a prominent topic in linguistics that has garnered attention from numerous researchers.

Roland Barthes was the first to systematically explore the interactive relationship between images and language in meaning generation from a semiotic perspective, proposing the influential "image-text status theory"[5]. Barthes identified three core modes of interaction in meaning generation between images and text: first, "anchorage"; second, "illustration"; and third, "relay". Kress & van Leeuwen extended Halliday's social semiotic theory and core concepts of Systemic Functional Linguistics (ideational, interpersonal, and textual metafunctions) into multimodal research, establishing the theoretical framework of "Visual Grammar"[6]. They focused on the organizational principles governing image-text spatial composition on the page, positing that these are governed by three core

dimensions: information value, salience, and framing. Building upon image-text status theory and Visual Grammar, Martinec & Salway proposed a more systematic framework for image-text relations [7]. This theoretical framework clearly categorizes image-text relationships into two main dimensions: status relations and logico-semantic relations. Among these, status relations include: an "equal relation" where images and text are interdependent and jointly construct meaning (comprising independent and complementary subtypes), and an "unequal relation" where one is subordinate to the other (including image subordinate to text, and text subordinate to image). Logico-semantic relations are further subdivided into: "expansion" (covering "elaboration," "extension," and "enhancement" of the primary information) and "projection" (including "locution" and "idea")

Currently, many domestic scholars have applied Martinec & Salway's theoretical framework of image-text relations to various fields, including art design, advertising creativity, journalism and communication, and children's picture books, among other academic domains. However, multimodal research focusing on textbooks remains relatively limited, with most studies concentrated on textbooks for subjects such as English and Chinese. Yang applied the image-text relationship analysis theory to the study of illustrations in learners' English dictionaries, confirming that illustrations can significantly enhance the learning-assistance function of dictionaries by forming semantic synergy with definition texts [8]. Wang & Jiang applied the image-text relationship theory to a multimodal discourse analysis of the story sections in the PEP Primary School English textbooks, revealing an evolving pattern across grades: lower grades primarily exhibit image-text complementarity, middle grades show a transition toward image-text independence, while higher grades demonstrate increased image-text subordination, with semantic relations shifting from projection to elaboration and extension [9]. Zheng & Yang further applied the analytical framework to the Welcome Unit section of the PEP Senior High School English compulsory textbooks [10]. Their

study indicated that the proportion of equal image-text relationships is significantly higher than that of unequal relationships, and the simultaneous occurrence of images and text is frequent in the textbooks. This finding suggests that teachers should fully utilize multimodal resources in textbooks to foster the development of students' core competencies.

In summary, the image-text relationship theory proposed by Martinec & Salway, by defining status relations and logico-semantic relations, provides a structured analytical framework for examining image-text interactions in multimodal discourse. Its successful application across numerous fields demonstrates its significant academic value and practical relevance. Therefore, based on the dimension of status relations within Martinec & Salway's image-text relationship theory under the framework of multimodal discourse analysis, this study will explore the image-text relationships in the PEP Senior High School English compulsory textbooks. The aim is to investigate the overall characteristics and underlying patterns of these image-text relationships, thereby providing empirical evidence for textbook compilation and instructional application.

### 3. Research Design

#### 3.1 Research Questions

With the development of the times, the integration of various image resources into textbooks has gained widespread acceptance. The application of multimodal resources reflects the pedagogical philosophy and instructional functions advocated by the new curriculum standards, rendering the investigation of image-text relationships in textbooks of considerable research value. Therefore, this study proposes the following two research questions:

1. What distributional characteristics do image-text relationships exhibit in the PEP Senior High School English (2019 Edition) compulsory textbooks?
2. What pedagogical implications do these image-text relationships hold for senior high

school English teaching practice?

#### 3.2 Research Objects

The research materials of this study consist of three volumes, namely the three compulsory volumes of the PEP Senior High School English textbooks (2019 edition), totaling 15 units. These textbooks were selected primarily due to their wide adoption nationwide and high representativeness. This textbook series includes Compulsory Volumes 1–3 and Selective Compulsory Volumes 1–4, totaling seven volumes. Among these, Compulsory Volumes 1–3 constitute the mandatory curriculum for senior high school students, holding universal reference significance for both teachers and students. Therefore, this study focuses on the illustrations from PEP Senior High School English Compulsory Volumes 1–3 (2019 Edition) as research objects.

To ensure the accuracy of quantitative analysis of image-text relationships, the scope of illustration statistics is strictly delimited to those appearing within the main body content of textbook units. Cover illustrations, background illustrations in unit introductions, and solid-color background pages are excluded from statistical calculation.

#### 3.3 Coding Scheme

Within Martinec & Salway's theoretical framework of image-text relations, image-text relations can be classified according to two subsystems: status relations and logico-semantic relations. The former comprises equal status relations, including image-text independence and image-text complementarity, as well as unequal status relations, encompassing image subordinate to text and text subordinate to image. Accordingly, this study classifies the illustrations in PEP Senior High School English compulsory textbooks based on the dimension of status relations within the image-text relation theoretical framework. This classification process constitutes the coding, with the specific coding rules presented in Table 1.

**Table 1. Coding Scheme for Image-Text Status Relations**

Category		Indicators	
Status Relations	Equal Status	Image & text independent	Both illustration and text possess complete representational functions; the information expressed is independent and equal
		Image & text complementary	Illustration and text require mutual support to form a greater community of meaning

Unequal Status	Image subordinate to text	Illustration is drawn based on partial information from the text or serves as a display of partial information
	Text subordinate to image	Text connects with partial or complete information in the illustration, and functions as an explanation or elaboration of the illustration's content

**4. Analysis of Research Results**

According to the classification theory of image-text relations, status relations encompass equal relations and unequal relations. This section presents a quantitative analysis of the image-text relationships in the three compulsory volumes of the PEP Senior High School English textbooks (2019 edition). It further explores the relationships between illustrations and their corresponding texts qualitatively through case studies of typical examples from the textbooks. This combined approach aims to provide textbook users with a comprehensive understanding of the overall characteristics of image-text relations within the instructional materials.

**4.1 Analysis of Equal Image-Text Status Relations**

Equal relations emphasize that image and text possess equal status and function in conveying information, with both capable of expressing messages independently and completely. Equal relations can be further categorized into image & text independent and image & text complementary. The following section will present the quantitative data on equal image-text status relations in the PEP Senior High School English textbooks (2019 edition) and analyze them in conjunction with specific examples. (as shown in Table 2).

**Table 2. Quantitative Statistics of Image-Text Equal Status Relations in PEP Senior High School English Compulsory Textbooks**

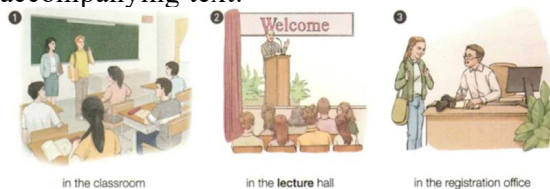
Grade	Status	Equal		Total
		Image & text independent	Image & text complementary	
Compulsory Volume 1		38	57	95
Compulsory Volume 2		22	35	57
Compulsory Volume 3		25	50	75
Total		85	142	227
Percentage		27.8%	46.3%	74.1%

**4.1.1 Image & text independent**

In image-text independent relations, images and text are connected on an equal footing. Although centered on the same theme, they independently and completely convey information, with no direct dependency or explanatory relationship between them.

For example, in the Listening and Speaking 1 section of the Welcome Unit in Compulsory Volume 1 (as shown in Figure 1), Illustration 1 depicts a classroom setting (featuring visual signs such as students, a teacher, a blackboard, desks, chairs, and books). Through the teacher's body language, facing the student group and the presence of learning materials, it clearly constructs the event scenario of "a teacher introducing a new student to the class," and can independently convey complete information without relying on the caption below. Illustration 2 presents a typical conference scene (a teacher speaking at a podium with a microphone, student audience below, and a welcome banner), indicating the event of "a teacher delivering a welcome speech to

students" without textual assistance. Illustration 3 shows a teacher and a student engaged in close conversation, with notetaking actions, and details such as the student wearing a backpack and smiling, supplemented by items commonly found in an office environment like notebooks, pens, and computers. From this, it can be inferred that the location is a registration office. All three illustrations are capable of conveying meaning independently without relying on the accompanying text.



**Figure 1. Example of Image & Text Independent**

**4.1.2 Image & text complementary**

Unlike image-text independent, image-text complementary emphasizes the interdependence and synergy between images and text. The two must be combined to construct the complete discursal meaning,

particularly when the text alone cannot fully express the semantics; the illustration provides key supplementary information, enabling the joint construction of a coherent textual message.

For example, in the Video Time section of Unit 2 in Compulsory Volume 3 (as shown in Figure 2), this part adopts a table format, with illustrations of visual symbols representing Confucian core concepts on the left and corresponding English definitions on the right. The term Ren (benevolence) is inherently abstract; relying solely on the textual definition (e.g., "Whatever you don't like done to yourself, don't do to others") makes it difficult for students to fully grasp its connotation as a code of conduct. The accompanying illustration uses classical bamboo slips as the background, depicting a Confucian scholar performing a ceremonial bow with hands clasped. The jade pendant at the waist symbolizes the virtue of a gentleman, while the serene expression and posture convey a scene of "respect, humility, and courtesy" in practice, translating Ren into a visualized behavioral paradigm. When students combine the textual definition—"loving people, loving everything in the world"—with the inclusive gesture in the image (arms outstretched toward nature), they construct a complete understanding of "the benevolent love of others": the text clarifies

the ethical essence, while the image concretizes the spatial dimension of "universal love." Without the image, the concept risks being reduced to didacticism; without the text, the image would remain merely a generic ancient-style illustration.



Figure 2. Example of Image & text complementary

#### 4.2 Analysis of Unequal Image-Text Status Relations

Unequal relations manifest as one mode being dependent on the other, with the dominant mode presenting the complete information while the subordinate mode conveys only partial information. Unequal relations include two subtypes: image subordinate to text and text subordinate to image. The following section will present a quantitative overview of unequal image-text status relations in the PEP senior high school English compulsory textbooks (2019 edition) and then analyze them in depth through illustrative examples of image subordinate to text and text subordinate to image (as shown in Table 3).

Table 3. Quantitative Statistics of Image-Text Unequal Status Relations in PEP Senior High School English Compulsory Textbooks

Grade	Status	Unequal		Total
		Image subordinate to text	Text subordinate to image	
Compulsory Volume 1		34	7	41
Compulsory Volume 2		13	3	16
Compulsory Volume 3		15	7	22
Total		62	17	84
Percentage		20.3%	5.6%	25.9%

##### 4.2.1 Image subordinate to text

Image subordinate to text refers to illustrations serving as auxiliary explanations of the written content. Their primary function is to support and supplement the information conveyed by the text, but the illustrations themselves lack independent meaning, with the text acting as the main carrier of information.

For instance, in the text "A Day in the Clouds" from Unit 2 of Compulsory Volume 2 (as shown in Figure 3), the illustration visually depicts antelopes on a plateau

meadow. However, its pedagogical function is entirely subservient to the text: First, as a concrete supplement to the verbal description, the image can only display the animals' appearance and habitat environment; it cannot independently convey historical contexts such as "the antelope population plummeted by 50% in the 1980s" or the poaching crisis. Second, the text constructs an ecological ethics dimension through the volunteer Tashi's dialogue: "Actually, we're trying to save ourselves"—a depth far beyond the representational capacity of the image.

Crucially, the specific conservation measures detailed in the text (such as adding passage bridges and establishing the Qiangtang Nature Reserve) are not represented in the illustration. This case confirms the core characteristic of "unequal image-text relations" in Martinec & Salway's theory: the image, as an adjunct to the text, serves to anchor the specific objects described, while the text carries the task of conveying the core information.



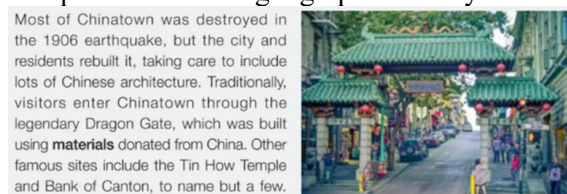
**Figure 3. Example of Image Subordinate to Text**

#### 4.2.2 Text subordinate to image

Text subordinate to image refers to the text serving as an auxiliary explanation of the illustration. Its primary function is to support and supplement the information conveyed by the image, but the text itself lacks independent meaning, with the image acting as the primary information carrier.

For example, in the text "Welcome to Chinatown" from Unit 3 of Compulsory Volume 3 (as shown in Figure 4), the illustration vividly portrays the street view of San Francisco's Chinatown (such as Dragon Gate, shops, and crowds), capturing its lively atmosphere and cultural characteristics. The text introduces the history, landmarks, cuisine, and cultural activities of Chinatown, providing contextual information for the image. For instance, the text notes that this Chinatown is the largest and oldest in the United States, attracting more visitors than the Golden Gate Bridge. It mentions that Dragon Gate was constructed using materials donated from China, highlighting the strong connection between Chinatown and Chinese heritage. Furthermore, the text introduces Tin How Temple, the Bank of Canton, and the

historic Portsmouth Square, enabling students not only to observe the external features of Chinatown but also to gain deeper insight into its cultural value. In teaching, this "text subordinate to image" relationship can help students better comprehend the information conveyed by the illustration while stimulating their interest in and desire to explore multiculturalism. Teachers can guide students to observe details in the image and then engage in in-depth discussions incorporating the textual content, thereby enhancing students' cross-cultural communicative competence and language proficiency.



**Figure 4. Example of Text Subordinate to Image**

### 5. Conclusion

Based on image-text relation theory, this study conducts quantitative analysis and qualitative discussion of image-text relations in PEP Senior High School English compulsory textbooks (2019 Edition) (see Table 4). From the quantitative findings, equal status relations between image and text dominate the textbooks, accounting for 74.1%, comprising 46.4% image-text complementarity and 27.8% image-text independence. Unequal status relations constitute 25.9%, with image subordinate to text at 20.3% and text subordinate to image at 5.6%. This distribution indicates that textbook design emphasizes collaborative interaction between image and text, prioritizing equal status relations while maintaining a proportion of unequal relations to provide diversified resources for varied instructional contexts.

**Table 4. Quantitative Statistics of Image-Text Status Relations in PEP Senior High School Compulsory English Textbook**

Garde	Status	Equal		Unequal		Total
		Image & text independent	Image & text complementary	Image subordinate to text	Text subordinate to image	
Compulsory Volume 1		38	57	34	7	136
Compulsory Volume 2		22	35	13	3	73
Compulsory Volume 3		25	50	15	7	85
Total		85	142	62	17	306
Percentage		27.8%	46.4%	20.3%	5.6%	100%

### **5.1 Pedagogical Implications of Image-Text Independence**

Image-text independence accounts for a certain proportion (27.8%) in the textbooks, indicating that teachers should not overlook its instructional value during teaching. For content featuring this relationship, teachers should guide students to extract information independently from both the image and the text, thereby cultivating their ability to process multimodal information. For example, teachers can first ask students to observe illustrations, describing details such as scenes, characters, and events, then have students read the text to extract key information, including background knowledge, viewpoints, or attitudes. Finally, teachers can organize students to compare and analyze the similarities and differences between the image and the text to deepen comprehension. Teachers can also utilize this relationship for extended activities, such as guiding students to make inferences, engage in discussions, or brainstorm based on clues from either the image or the text, thereby enriching their knowledge base and honing their thinking quality.

### **5.2 Pedagogical Implications of Image-Text Complementarity**

Image-text complementarity accounts for the highest proportion (46.4%), underscoring its significance within the textbooks. In instruction, teachers should fully exploit its advantages by guiding students to focus on the mutual reinforcement between images and text, fostering their ability to grasp multimodal texts holistically. Specifically, teachers can first guide students to interpret the image independently to grasp its intuitive information (e.g., setting, actions, and emotions); then have them read the text to obtain detailed information (e.g., causes, processes, inner thoughts); finally, instruct students to integrate the information from both sources to construct a complete discursive meaning. Furthermore, teachers can design interactive activities based on this relationship, such as group discussions or role-playing, enabling students to appreciate the dynamics of image-text complementarity through collaboration, thereby enhancing their comprehensive language competence

and cognitive abilities.

### **5.3 Pedagogical Implications of Image Subordinate to Text**

Accounting for 20.3%, the relationship of image subordinate to text indicates that the illustration primarily serves an auxiliary function. In teaching, teachers should guide students to focus on how the image enhances textual expressive effects, facilitating their comprehension of written content. For example, teachers can first have students read the text to extract key information, then instruct students to observe the image to explore how visual elements (such as color, line, and composition) reinforce emotions, atmosphere, or key points conveyed verbally. Meanwhile, this relationship can be utilized to train students' ability to convert between visual and verbal information—for example, by having them describe the image's message in their own words, which deepens their understanding and retention of the text.

### **5.4 Pedagogical Implications of Text Subordinate to Image**

Although constituting a smaller proportion (5.6%), the relationship of text subordinate to image should not be neglected. Under this relationship, text mainly serves to explain and elaborate on illustrations. Teachers can guide students to first observe the image and extract its key information, then read the text to explore how it supplements the visual content. Additionally, teachers can design writing or speaking activities based on this relationship, prompting students to expand on the image's content using the textual cues, thereby enhancing their language output competence.

In summary, the image-text relations in PEP Senior High School English compulsory textbooks (2019 Edition) demonstrate rich diversity, with each relationship possessing distinctive pedagogical value. Teachers should flexibly employ teaching strategies according to the characteristics of different relations, fully exploiting multimodal resources in the textbooks to promote the holistic development of students' language competence, thinking qualities, and cultural awareness. Meanwhile, textbook developers should continue to focus on the rational coordination of image and text in future

editions, further optimizing the multimodal design of textbooks to provide richer and more effective instructional resources for English language teaching.

### References

- [1] Li Zhanzi and Lu Danyun. "Multimodal Semiotics: Theoretical Foundations, Research Approaches, and Development Prospects." *Foreign Language Research*, 2012(2): 1-8.
- [2] Ministry of Education of the People's Republic of China. *The 2017 Edition of English Curriculum Criteria for Regular High School*. Beijing: People's Education Press, 2020.
- [3] Chen Xiaotang and Cong Lin. "The Design and Use of Imagery Resources in the Writing of English Textbooks." *Curriculum, Teaching material and Method*, 2020, 40(8): 78-85.
- [4] Zhu Yongsheng. "Theoretical Foundations and Research Methods of Multimodal Discourse Analysis." *Foreign Language Research*, 2007(5): 82-86.
- [5] Barthes, Roland. *Image-Music-Text*. London: Fontana Press, 1977.
- [6] Kress, G.R. and Van Leeuwen, T. *Reading images: The grammar of visual design*. London: Psychology Press, 1996.
- [7] Martinec R. "A System for Image-Text Relations in New (and Old) Media." *Visual Communication*, 2005, 4(3): 337-371.
- [8] Yang Xinzhang. "Text-image Relations in Learner's English-Chinese Dictionaries." *Journal of University of Science and Technology Beijing (Social Sciences Edition)*, 2012, 28(4): 45-51.
- [9] Wang Ronghua and Jiang Guiying. "Multimodal Foreign Language Teaching: Blending of Image and Language Resource." *Journal of Schooling Studies*, 2015, 12(3): 84-90.
- [10] Zheng Zihao and Yang Lijuan. "Image-Text Relations in Textbooks from a Multimodal Perspective: A Case Study of Welcome Unit in PEP Senior High School English (Compulsory Volume 1, 2019 Edition)." *Overseas English*, 2023(20): 207-210