Higher Education and Practice Vol. 3 No. 5, 2026

An Exploration of the Application of Multimodal Discourse Analysis in English Reading Comprehension

Mao Mao
Anhui Vocational College of Art, Hefei, Anhui, China

Abstract: In contemporary English reading instruction, textual materials typically exhibit multimodal characteristics-visual elements such as images, charts, and layout designs collaborate with linguistic content to construct meaning. However, traditional reading teaching has long centered on linguistic modalities while neglecting the synergistic relationships between different modalities. Grounded in systemic functional linguistics and visual grammar theory, this study explores application pathways for multimodal discourse analysis in English reading comprehension education. The paper first outlines core concepts and theoretical frameworks of multimodal discourse analysis, then develops a comprehension model for multimodal English reading and proposes instructional strategies for three phases: pre-reading, during-reading, and post-reading. This paper introduces a case study of picture book reading for college students and a case analysis of public service advertisements. Research demonstrates that multimodal reading strategies enhance learners' ability to interpret text-image interactions and integrate cross-modal information, while fostering critical multimodal literacy. The paper further discusses practical challenges in implementation and potential solutions.

Keywords: Multimodal Discourse Analysis; English Reading Comprehension; Visual Grammar; Text-Image Relationship; Teaching Strategies

1. Introduction

1.1 Research Background

Contemporary English reading materials have evolved far beyond mere textual content. From illustrations and charts in textbooks to infographics in online articles and multimedia content on social media, multimodal information has become the norm. However, my teaching observations reveal that many students tend to focus solely on textual language while treating visual elements as optional decorations when interpreting such texts. During one lesson, I presented an English expository passage containing a bar chart; several students failed to grasp the data trends after reading the text and only realized upon inquiry that they had overlooked the accompanying chart. Students cited reasons such as "examination questions often lack visuals" or "being accustomed to reading only text." This phenomenon highlights a persistent issue in English reading instruction: the tendency to prioritize linguistic content over visual elements may hinder students' ability to effectively extract and integrate information from real-world multimodal texts.

Multimodal discourse analysis provides a systematic theoretical framework for understanding how non-linguistic symbolic systems contribute to meaning construction. Its application in English reading comprehension instruction enables a re-examination of reading's essence-not merely as text decoding, but as a process that integrates diverse modal resources to build holistic meaning. Nevertheless, research on applying this theoretical framework to the design of English reading teaching strategies remains limited.[1]

1.2 Research Objectives and Significance

This study aims to explore the application of multimodal discourse analysis theory in English reading comprehension instruction, specifically by clarifying the meaning construction mechanisms of multimodal reading, designing a practical three-stage teaching strategy, and evaluating its feasibility and effectiveness through teaching cases. At the theoretical level, the research seeks to expand the application scope of multimodal discourse analysis and enrich the theoretical perspectives in English reading instruction research; at the practical level, it provides frontline teachers with concrete

teaching strategies and case references to help learners adapt to contemporary multimodal reading environments and develop multimodal literacy skills.

1.3 Research Methods

This study employed a literature review approach to systematically examine the foundational theories of multimodal discourse analysis and relevant research findings both domestically and internationally. Additionally, using case studies, This program implemented teaching practices of picture book reading and public service advertisement reading for college students, collecting qualitative data on instructional implementation through classroom observations, analysis of student work, and teaching reflections.[2]

2. Theoretical Basis and Literature Review

2.1 Definition of Core Concepts

Multimodal discourse analysis refers to a research paradigm that extends the scope of discourse analysis from linguistic symbols to various symbolic resources such as images, colors, layouts, and sounds. The term "modality" encompasses different symbolic systems and their respective material carriers-including linguistic modalities (oral and written language), visual modalities (images, charts, layout design), and auditory modalities (sound and music). The core tenet of multimodal discourse analysis is that each modality possesses its own semiotic logic in meaning construction; they do not exist in a hierarchical relationship but rather interact synergistically.

English reading comprehension is generally defined as the process by which readers extract and construct meaning from written texts. Classic reading models (such as the bottom-up model, top-down model, and interactive model) primarily focus on linguistic elements. In multimodal contexts, this process also involves processing cross-modal information associations and integration. In other words, multimodal reading competence encompasses not only text decoding and semantic reasoning but also the ability to recognize and interpret visual information such as images, charts, and layouts, and to coordinate it with textual content.

2.2 Theoretical Basis

The theoretical foundation of this study primarily draws from systemic functional linguistics and visual grammar theory. Halliday (1978) 's systemic functional linguistics categorizes language's metaprocesses into conceptual, interpersonal, and discursive functions, positing that language constitutes a system of meaning potentialities. Building upon this framework, Kress and van Leeuwen (2021) developed the visual grammar theory, which identifies three types of meaning conveyed by images: representational meaning (what content an image presents and its relationships), interactive meaning (the relationship between the image and the viewer manifested through visual elements such as distance, perspective, and proximity), and compositional meaning (the arrangement of elements within the image, their informational value, salience, and framing). This theory provides a systematic analytical framework for interpreting images in English reading materials.

Furthermore, Mayer's (2009) cognitive theory of multimedia learning posits that when text and images are presented simultaneously, learners must establish a mental connection between the verbal and visual channels; learning outcomes depend on how information is integrated rather than simply superimposed. Schema theory further suggests that multimodal information activates various types of background knowledge schemas, thereby enhancing text comprehension and retention.[3]

2.3 Current Research Status Domestically and Internationally

Multimodal discourse analysis research has a long history abroad, with scholars such as Kress, van Leeuwen, and Jewitt developing robust theoretical frameworks that have been applied to areas including textbook analysis, advertising discourse, and classroom interaction. In recent years, the integration of multimodality with reading instruction has garnered increasing attention, with research focusing on picture book comprehension, text-image relationships in scientific texts, and multimodal information processing in digital reading environments. Domestically, scholars like Zhang Delu (2009), Gu Yueguo (2007), and Li Zhanzi (2003) have introduced and advanced multimodal discourse analysis theory. However, systematic applications of this theory to English reading comprehension teaching remain limited: existing studies predominantly focus on content analysis

of multimodal features in textbooks or propose general instructional recommendations, rarely delving into specific teaching strategy design or effectiveness validation. This study aims to further explore this area.

3. The Multimodal English Reading Comprehension Model

3.1 Collaborative Processing Mechanism of Multimodal Information

In multimodal English reading, readers do not process linguistic and visual information sequentially or in isolation, but engage in dynamic cross-processing between the two sources. The dual-coding theory in cognitive psychology provides a compelling explanation: the human information processing system comprises both verbal and visual channels; when these two types of information are linked during encoding, cognitive load is distributed and processing efficiency improves. For instance, when reading an English science article on the greenhouse effect, the text states that "carbon dioxide concentrations have risen from 280 ppm before the Industrial Revolution to 420 ppm today," while a line graph visually illustrates this upward trend. Only by associating the numerical values in the text with the corresponding points and curve trajectories on the graph can readers fully comprehend the magnitude and significance of the change. This collaborative processing mechanism forms the cognitive foundation for enhanced multimodal reading efficiency.[4]

3.2 Basic Relationships Between Modalities

The relationship between images and text is not singular. The framework for image-text relationships proposed by Martinec and Salway (2005) distinguishes two dimensions: first, logical relationships-including elaboration (e.g., specification, exemplification) and projection (e.g., dialogue, ideas); second, positional relationships-including equality (both providing equally important information) and inequality (one being subordinate to the other). In English reading instruction, identifying these intermodal relationships helps readers allocate attention appropriately and accurately comprehend textual intent. For instance, when images and text complement each other, readers should give equal weight to both; when images merely serve as concrete extensions of text, readers may prioritize textual content while considering images as supplementary. However, my teaching observations reveal that many students lack this awareness-either completely ignoring images or overinterpreting decorative ones-which compromises reading accuracy and efficiency.

3.3 Components of Multimodal Reading Competence

Based on the above analysis, multimodal reading competence can be broken down into three progressive levels. The first level is multimodal information decoding ability: the capacity to identify the fundamental meaning conveyed by images, charts, and layout elements, such as interpreting numerical comparisons in bar charts or understanding power dynamics implied by upward-looking camera angles in advertisements. The second level is cross-modal information association ability: the ability to establish corresponding, complementary, or contrasting relationships between text and visuals, for example, matching the phrase "increase steadily" with the rising trend line in a chart. The third level is multimodal critical reading ability: the capacity to evaluate the credibility of multimodal information, recognize logical consistency across modalities, and analyze designers 'rhetorical intentions. This advanced competency is particularly crucial in today's information-overloaded environment.

4. Application Strategies of Multimodal Discourse Analysis in English Reading Comprehension

4.1 Before Reading: Activating the Multimodal Schema

The preparatory phase before reading marks the starting point for applying multimodal discourse analysis. During this stage, teachers should guide students to systematically preview the text's overall layout, including headline font style and size, image types and placement, column-based information organization, color scheme usage, and even the presence of graphical elements such as icons, dividers, or background patterns. These visual elements convey crucial metadiscourse information: large bold headlines typically indicate core arguments or main ideas; color-boxed sections may summarize key points, provide definitions, or highlight warnings; central images usually represent thematic visuals that encapsulate the

text's central theme; and sidebars or marginalia often contain supplementary examples or background knowledge.

To activate students' multimodal schema effectively, teachers can employ a guided observation checklist:

What stands out most on the page (size, color, position)?

Are there recurring visual motifs (e.g., arrows, numbered steps, icons)?

How is white space used? Does it separate sections or create emphasis?

By answering such questions, students quickly identify the text type (news report, advertisement, instructional manual, storybook, or infographic) and form reasonable predictions about its content, purpose, and structural organization. For instance, before reading an English public service advertisement about environmental protection, a teacher might ask: "What perspective is employed in the ad's composition-bird's-eye, eye-level, or worm's-eye? How does this perspective suggest the relationship between the viewer and the subject matter (e.g., a polluted river seen from above may evoke a sense of helplessness or overview)?" This strategy of activating multimodal schemas not only builds reading expectations but also reduces cognitive load during subsequent detailed reading, as students already have a mental framework for where to locate different types of information.

4.2 Reading Process: Multimodal Information Processing

The core task during reading is to coordinate the processing of information from both linguistic and visual modalities. Unlike traditional reading that focuses solely on linear text, multimodal reading requires dynamic cross-referencing between modes. Teachers should guide students in employing a three-step text-image interpretation strategy:

Skim the image to grasp its representational content: identify the individuals, objects, scenes, actions, and relationships depicted. Pay attention to details such as gaze direction (do characters look at the viewer or at each other?), distance (close-up, medium shot, long shot), and angle (high, low, or eye-level).

Read adjacent textual passages (e.g., captions, the paragraph surrounding the image, or call-out boxes) to locate key phrases that correspond to elements in the image.

Cross-validate by mapping textual information onto specific areas of the image. Ask: "Does the text explicitly describe what I see? Are there inconsistencies (e.g., the text mentions a happy ending but the image shows a somber face)?"

The essence of these steps lies in establishing cross-modal referential relationships-recognizing when text and image reinforce each other (redundancy), when they provide complementary information (expansion), or when they create tension or contradiction (counterpoint).

Furthermore, identifying visual metadiscourse markers is a crucial strategy for navigating complex layouts. While text employs metadiscourse markers such as "first," "however," "in conclusion," or "for example" to indicate rhetorical structure, layouts incorporate visual equivalents: numerical sequences (1, 2, 3), arrows pointing from one block to another, color-coded sections (red for warnings, green for positive outcomes), icon-based navigation (home buttons, next arrows), and spatial hierarchies (top-to-bottom, left-to-right, or center-to-periphery reading paths). Recognizing these visual metadiscourse markers helps students grasp the text's structural logic-for instance, a flowchart with arrows explicitly signals a cause-effect or process sequence. Without such awareness, students may jump randomly between columns or miss important connections. Teachers can design visual metadiscourse hunts where students circle arrows, number sequences, or color-coded labels before reading the main content, thereby internalizing the text's architecture.

4.3 After Reading: Multimodal Meaning Reconstruction

The post-reading phase should not merely involve recounting the textual content; instead, students should be encouraged to engage in cross-modal information transfer and meaning reconstruction. This transforms reading from passive absorption into active synthesis. Two powerful strategies are particularly effective:

Strategy 1: Modal conversion

This requires students to convert information from one mode to another. For example:

Text → Diagram: After reading a descriptive paragraph about the water cycle, students create a labeled flowchart with arrows.

Diagram → Text: Given an infographic on global warming trends, students write a short

explanatory paragraph that verbalizes the data patterns.

Text → Multimodal presentation: Students summarize a news article using a combination of bullet points, a sketched cartoon, and key quoted phrases.

This process forces students to repeatedly compare and integrate information across different modalities, thereby deepening their understanding of the text. It also reveals gaps or ambiguities: if a student cannot convert a textual argument into a mind map, they likely have not fully grasped the logical relationships.

Strategy 2: Multimodal critical evaluation

Teachers guide students to examine potential ideological biases or rhetorical manipulations in the interplay between text and image. Critical questions include:

Does the image embellish the facts described in the text? (E.g., a product review text mentions "durable material," while the image shows a pristine, never-used item-does this avoid showing wear and tear?)

Does the layout marginalize certain information by placing it in the lower right corner, using small font sizes, employing dark colors over black backgrounds, or grouping it under a vague subheading? (Such positioning can signal "less important" or "optional" even if the content is crucial.)

Who is represented in the images, and who is absent? For instance, a textbook chapter on "modern family life" might show only nuclear families with parents and two children, excluding single-parent or extended families, thus subtly normalizing one model.

Such critical training helps develop students' multimodal literacy-the ability to not only decode but also question and resist manipulative design. Over time, learners transform from passive information recipients into active analysts who can articulate how visual choices shape meaning and persuade audiences.

4.4 Differentiation Strategies for Various Text Types

Different types of reading materials require distinct approaches to multimodal processing. Teachers should adapt their instructional strategies based on genre conventions and the dominant mode of communication. Below are detailed, genre-specific guidelines.

4.4.1 English picture books (e.g., Where the Wild Things Are, The Very Hungry Caterpillar)

In picture books, text is often concise and visuals dominate the narrative function. Students should be guided to "read the images" with the same rigor as reading words. Key observation points include:

Character expressions and postures: A furrowed brow, slumped shoulders, or open arms convey emotions not always stated in the text.

Background changes: Shifts in color palette (e.g., from warm yellows to cool blues) or scenery (e.g., from a tidy room to a wild forest) indicate mood transitions or plot progression.

Layout and framing: A full-page spread may signal a climax, while small vignettes suggest quick successions of events.

Teachers can prompt: "Why does the illustrator use a close-up of the character's face here? What do the jagged lines around the monster suggest?" By practicing visual narrative analysis, students infer plot developments and emotional tones independently.

4.4.2 Advertising texts (print or digital ads)

In advertisements, the synergy between visuals and text often carries clear rhetorical purposes. Students need to identify:

Appeal strategies: Rational appeals (facts, statistics, product features) versus emotional appeals (happiness, fear, nostalgia, social belonging).

How images reinforce or undermine textual claims: A luxury watch ad may show an elegant celebrity (emotional appeal) while the text lists "Swiss precision" (rational appeal). The image creates desire; the text justifies purchase.

Gaze and address: Does the model look directly at the viewer (demanding engagement) or look away (offering contemplation)? Direct gaze often creates a sense of personal address, common in charity or political ads.

A useful activity is ad deconstruction: students cover the text and infer the intended message from images alone, then read the text and compare their inference with the actual claims. This reveals how advertisers strategically combine modes.

4.4.3 Graphical expository texts (e.g., Science Charts, Infographics, Maps)

In these texts, visuals convey dense data while text provides interpretive context. A three-step approach is recommended:

Examine titles and legends first to understand what the graphic measures (e.g., "Annual CO_2 Emissions by Country (2000–2020)").

Analyze axes and trends: Identify independent

vs. dependent variables, units of measurement, and patterns (increase, decrease, fluctuation, plateau).

Review textual explanations to verify interpretations. Often, the text highlights key findings (e.g., "Emissions peaked in 2015") that may not be immediately obvious from a cluttered graph.

Teachers should also warn against common pitfalls: misreading scale (e.g., a truncated y-axis exaggerating differences), confusing colors in a legend, or ignoring error bars and footnotes.

4.4.4 Digital multimodal texts (e.g., English Learning Websites, Interactive E-books, News Portals)

Digital environments add dynamic elements that static texts lack: animations, hyperlinks, pop-ups, hover effects, embedded videos, and audio narration. These can enhance engagement but also increase cognitive load. Teachers should guide students in attention allocation strategies:

Preview interactive elements before starting: note which parts are clickable, which animate on scroll, and which play sound.[5]

Set a reading path: For example, "Read the main article first, then watch the 30-second video summary, then click the interactive timeline for details." Without a plan, students may click randomly and lose the narrative thread.

Distinguish core information from decorative or supplementary material: Animated icons that dance on every page are often purely aesthetic; a highlighted text box that expands when clicked may contain essential definitions.

Teachers can model think-aloud protocols while navigating a digital text: "I see a pop-up ad on the right-I will ignore it. The main text begins here. A hyperlinked word, 'photosynthesis', probably leads to a glossary. I'll finish this paragraph first, then click it." Such explicit metacognitive instruction helps students manage split attention and avoid missing core information due to distractions.

By differentiating strategies according to text type, teachers ensure that multimodal discourse analysis is not a one-size-fits-all method but a flexible toolkit. Over time, students internalize the habit of asking, "What kind of text is this, and what multimodal strategies does it demand?"-a skill transferable to academic, professional, and everyday reading contexts.[6]

5. Real Challenges and Approaches to Address Them

5.1 Main Challenges Faced

The application of multimodal discourse analysis in English reading comprehension instruction faces multiple practical challenges.[7] At the teacher level, the majority of English educators have received traditional linguistic and literary training, lacking systematic visual analysis literacy. While concepts such as "information value," "significance," and "framing" from Kress and van Leeuwen's visual grammar theory are theoretically understandable, teachers often struggle when applying them to actual classroom analyses of texts. Regarding teaching materials, the quality of illustrations in existing textbooks varies significantly: some complement textual content effectively, while a considerable number serve merely as decorative elements, and a few even contradict textual content-a situation that adds unnecessary complexity to multimodal instruction. For learners, there are individual differences in modal preferences: visual learners tend to overlook textual details, whereas verbal learners frequently skip images. Additionally, students may experience cognitive overload when encountering information-dense multimodal texts, where abundant colors, graphics, and text interfere with focusing on key information.[8]

5.2 Preliminary Considerations on Response Strategies

To address these challenges, several approaches can be adopted. In teacher training, workshops on visual grammar could be incorporated into both pre-service and in-service programs for English teachers, guiding them to conduct practical analyses of text-image interactions in teaching materials. For resource development, a multimodal English reading teaching case library could be established by selecting high-quality textual-materials (e.g., National Geographic Children's Editions, BBC Infographics) accompanied by analytical guidelines. Regarding differentiated instruction, visual learners could benefit from text-focused scaffolding (e.g., highlighting key sentences), while verbal learners might use image observation cue cards listing essential visual elements. To enhance metacognitive skills, students should be encouraged to reflect on their reading processes, gradually developing habits like self-assessment questions such as "Did I miss any visual information?" or "Is there consistency between

text and images?" The feasibility of these strategies requires further empirical validation.[9]

6. Conclusion and Prospects

6.1 Main Conclusions

This study re-examines English reading comprehension instruction from the perspective of multimodal discourse analysis. The findings demonstrate that multimodal discourse analysis provides an effective theoretical framework for understanding the essence of contemporary English reading-reading is no longer a passive reception of linear text but rather an active integration and negotiation of multiple symbolic resources. Building on these insights, this study develops a multimodal comprehension model for English reading, proposes application strategies applicable across all three stages (pre-, during, and post-reading), and validates its feasibility through teaching cases. Specifically: First, the text-image mutual interpretation strategy significantly enhances students 'comprehension of both emotional and informational texts; Second, training in recognizing visual metadiscourse markers improves students' awareness of textual structure; Third, multimodal critical reading practice fosters higher-order thinking skills, enabling learners to transcend mere literal comprehension.

6.2 Research Limitations

This study has the following limitations: the teaching practice scale was small and lacked systematic quantitative data support; it primarily focused on the combination of visual and linguistic modalities, with limited discussion of dynamic modalities such as audio and video; it did not adequately examine adaptive differences among learners with varying English proficiency levels; and the study period was short, failing to evaluate the long-term effects of multimodal strategies.

6.3 Future Research Directions

Future research can be advanced in the following directions: conducting larger-scale experimental studies that employ tools such as eye-tracking to quantitatively evaluate the effectiveness of multimodal teaching; exploring AI-assisted multimodal reading tools, including those that utilize natural language processing and computer vision technologies for automatic annotation of text-image correspondences; extending research to mobile reading contexts by examining the processing mechanisms of novel multimodal texts such as short videos, interactive infographics, and scrolling pages; and undertaking cross-cultural comparative studies to investigate how learners from different cultural backgrounds interpret cultural symbols within images. Multimodal discourse analysis has opened up new research frontiers for English reading instruction, and we anticipate that further empirical research and practical exploration will enrich achievements in this field.

Acknowledgments

This research was supported by Scientific Research Projects of Philosophy and Social Sciences at Universities in Anhui Province (Grant No. 2025AHGXSK30604)

References

[1] Kress, G., & van Leeuwen, T. (2021). *Reading Images: The Grammar of Visual Design* (3rd ed.). Routledge.

[2] Halliday, M. A. K. (1978). *Language as Social Semiotic*. Edward Arnold.

[3] Jewitt, C. (2014). *The Routledge Handbook of Multimodal Analysis* (2nd ed.). Routledge.

[4] Mayer, R. E. (2009). *Multimedia Learning* (2nd ed.). Cambridge University Press.

[5] Zhang Delu. (2009). Exploration of a comprehensive theoretical framework for multimodal discourse analysis. China Foreign Languages, (1),24–30.

[6] Gu Yeguo. (2007). Analysis of Multimedia and Multimodal Learning. Foreign Language Audiovisual Teaching, (2),3–12.

[7] Li Zhanzi. (2003). A sociolinguistic analysis of multimodal discourse. Foreign Language Studies, (5),1–8.

[8] Wang Lu. (2018). Research on the relationship between text and images in English reading instruction from a multimodal perspective. Foreign Language Teaching in Primary and Secondary Schools, (10),23–28.

[9] Liu Dan. (2020). Application of multimodal discourse analysis in high school English reading instruction. Basic Foreign Language Education, (3),45–51.