

Intersemiotic Relations Between Verbal and Visual in Children's Picture Books: A Multimodal Discourse Analysis

Margaretha Miki Tandipare
mkzkyy2308@gmail.com

Sahril Nur
sahrilfbsunm@unm.ac.id

Harnita Rosalia
harnita.rosalia@unm.ac.id

Universitas Negeri Makassar

ABSTRACT

This study examined how intersemiotic relations between verbal text and visual illustrations contribute to narrative meaning construction in selected children's picture books using the framework of Multimodal Discourse Analysis (MDA). The study focused on three types of intersemiotic relations proposed by Royce (2013), namely complementary, reinforcing, and expanding relations. A descriptive qualitative research design was employed. The data sources consisted of three children's picture books: *The Gruffalo* by Julia Donaldson and Axel Scheffler, *The Tale of Peter Rabbit* by Beatrix Potter, and *We're Going on a Bear Hunt* by Michael Rosen and Helen Oxenbury. Ten extracts were purposively selected through close reading as the most representative instances of verbal-visual interplay in narrative construction. Findings reveal that the extracts are distributed across three categories: reinforcing relations (5 extracts), where verbal and visual modes mutually strengthen meanings in scenes of movement, confrontation, and character description; complementary relations (3 extracts), where both modes contribute distinct but mutually necessary meanings across emotional and atmospheric dimensions; and expanding relations (2 extracts), where illustrations add emotional depth and symbolic significance beyond the verbal text. Reinforcing relations emerged as the dominant pattern. The study demonstrates that verbal-visual interplay is the fundamental mechanism through which narrative meaning is constructed in children's picture books, offering a methodological contribution to multimodal narrative analysis in children's literature.

Keywords: Intersemiotic Relations; Children's Picture Books; Multimodal Discourse Analysis

INTRODUCTION

In contemporary communication practices, meaning is no longer constructed solely through verbal language but emerges through the interaction of multiple semiotic modes such as images, color, layout, and visual composition. Bezemer and Kress (2008) argue that meaning in modern communication is produced multimodally

because various semiotic resources operate simultaneously in constructing messages. This multimodal nature of communication is particularly evident in children's picture books, which combine verbal text and visual illustrations within a single narrative structure to collaboratively produce meaning.

Children's picture books occupy an important position in both literary and educational contexts. In these texts, illustrations do not merely function as decorative elements but serve as active meaning-making resources that contribute significantly to narrative construction. Painter, Martin, and Unsworth (2013) explain that picture books are multimodal narrative texts in which verbal and visual elements collaboratively contribute to narrative progression, emotional atmosphere, and character representation. Serafini (2024) further states that words and images in picture books establish complex relationships that shape readers' narrative understanding, making it impossible to fully comprehend the story through verbal analysis alone. Despite the growing interest in multimodal studies of children's picture books, limited attention has been given specifically to how verbal and visual modes interact through intersemiotic relations in constructing narrative meaning as an integrated system. Previous studies, including Moya Guijarro (2014), Svensson (2020), Xie, Noor, and Hamed (2023), and Marmion (2025), have approached picture books from various multimodal perspectives but have not systematically examined the intersemiotic relations between verbal and visual modes using Royce's (2013) framework of intersemiotic complementarity. This gap motivates the present study.

Several multimodal frameworks could, in principle, be applied to this material. Painter et al. (2013) offer a systemic-functional model of image analysis attentive to narrative structure, character focalization, and page composition, while Marsh and Millard (2000) approach picture books mainly from a literacy-pedagogy perspective, concerned with how young readers are socialized into multimodal texts rather than with classifying verbal-visual relations themselves. Royce's (2013) intersemiotic complementarity was selected over these alternatives because it supplies an explicit, three-way taxonomy complementary, reinforcing, and expanding relations that names how meaning is distributed between the verbal and visual channels at the level of individual extracts. This granularity suits the study's aim of quantifying and characterizing verbal-visual interaction across extracts, whereas Painter et al.'s model is oriented more toward narrative image grammar and Marsh and Millard's toward literacy development; Royce's categories therefore give the clearest analytical fit for the research questions pursued here. Therefore, this study aims to examine how intersemiotic relations between verbal text and visual illustrations, specifically complementary, reinforcing, and expanding relations, contribute to narrative meaning construction in selected children's picture books. Understanding this interplay is important not only for advancing multimodal theory but also for providing practical insights for educators, writers, and illustrators regarding how verbal and visual resources work together in children's storytelling.

LITERATURE REVIEW

Previous Related Studies

Several previous studies have explored multimodality in children's picture books from different perspectives. Moya Guijarro (2014) analyzed children's picture books using a Systemic Functional Linguistics approach and demonstrated that verbal and visual elements collaboratively contributed to meaning making. However, the study did not specifically examine intersemiotic relations as an integrated analytical focus. Xie, Noor, and Hamed (2023) explored visual narrative characteristics in Chinese classical poetry picture books and found that visual and textual elements contributed to narrative continuity, yet the focus remained primarily on visual narrative characteristics rather than the interaction between verbal and visual modes. Svensson (2020) examined how visual elements such as gaze, distance, and perspective influenced emotional engagement in picture books but did not systematically analyze how verbal-visual interaction through intersemiotic relations constructed narrative meaning. Marmion (2025) explored multimodal storytelling strategies in contemporary award-winning picture books and found that page layout and visual-verbal relationships contributed to narrative interpretation, yet the study focused on stylistic aspects rather than intersemiotic analysis. These gaps justify the present study's specific focus on intersemiotic relations as the analytical framework for examining narrative meaning construction in children's picture books.

Multimodal Discourse Analysis (MDA)

Multimodal Discourse Analysis (MDA) is an analytical framework used to examine how meaning is constructed through the interaction of multiple semiotic modes such as language, images, color, layout, and composition. O'Halloran (2008) explains that meaning in multimodal discourse emerges through the integration of linguistic and visual semiotic resources that function together in meaning-making processes. Kress and Bezemer (2023) argue that discourse analysis must consider the interaction among semiotic modes since meaning is fundamentally multimodally constructed. In this study, MDA serves as the overarching framework within which intersemiotic analysis is conducted to examine how verbal and visual modes interact in constructing narrative meaning in children's picture books.

Intersemiotic Complementarity

The central theoretical framework of this study is Royce's (2013) concept of intersemiotic complementarity, which explains how verbal text and visual illustrations interact collaboratively in constructing meaning within multimodal texts. Royce (2013) argues that meaning in multimodal texts emerges through complementary relationships between semiotic modes rather than through the dominance of one particular mode. Based on Royce (2013), intersemiotic relations in this study include three types. First, complementary relations occur when verbal

and visual elements complete each other in constructing meaning, with each mode contributing information that the other does not fully express. Second, reinforcing relations occur when verbal and visual elements emphasize similar meanings or information, strengthening the overall message through redundancy across modes. Third, expanding relations occur when one mode provides additional information beyond what is expressed in the other mode, enriching the narrative with meanings that extend beyond the verbal or visual text alone. Hu and Qiu (2020) explain that these intersemiotic relations enrich meaning-making processes because readers construct interpretation through the simultaneous interaction between words and images. In children's picture books, intersemiotic complementarity is particularly relevant because illustrations often clarify actions, strengthen emotional atmosphere, or provide narrative details not explicitly stated in verbal text. Svensson (2020) explains that visual elements such as gaze, perspective, and distance influence readers' interpretation and emotional engagement with narrative situations. Shi (2023) further states that visual-verbal interaction contributes significantly to multimodal meaning-making and visual literacy in children's picture books.

Synthesizing Visual Grammar and Intersemiotic Complementarity

While Royce's (2013) framework identifies what kind of relation holds between verbal and visual modes, it does not by itself specify how the visual mode generates meaning. This study therefore draws on Kress and Van Leeuwen's (2020) systemic-functional visual grammar as the descriptive layer beneath Royce's classificatory layer. Kress and Van Leeuwen (2020) distinguish three simultaneous strands of visual meaning: representational meaning (who or what is depicted and the actions or attributes involved, comparable to transitivity in verbal grammar), interpersonal meaning (the relationship constructed between the depicted participants and the viewer, realized through gaze, social distance, and camera angle), and compositional meaning (how information value, salience, and framing are distributed across the page). In practice, representational choices (e.g., a character's posture or action) most often underpin reinforcing relations, since they mirror processes already named in the verbal text; interpersonal choices (e.g., gaze and distance) most often underpin complementary relations, since they add relational or emotional meaning the words do not state; and compositional choices (e.g., framing, salience, and what is included or omitted from the frame) most often underpin expanding relations, since they introduce symbolic or atmospheric meaning beyond the verbal text. This synthesis links the two frameworks explicitly, so that each extract's visual description in the Data Analysis section is grounded in a specific, named stratum of visual grammar rather than in impressionistic description.

METHOD

Design and Samples

This study employed a descriptive qualitative research design focusing on multimodal textual analysis. The primary data sources consisted of three selected children's picture books: *We're Going on a Bear Hunt* by Michael Rosen and Helen Oxenbury, *The Tale of Peter Rabbit* by Beatrix Potter, and *The Gruffalo* by Julia Donaldson and Axel Scheffler. These books were selected because they demonstrate strong verbal-visual interaction while employing different storytelling strategies. The data consisted of 10 selected extracts containing verbal texts and visual illustrations chosen based on their relevance to narrative progression, character representation, emotional development, and verbal-visual interaction. The 10 extracts were distributed across the three types of intersemiotic relations: five extracts illustrating reinforcing relations (Extracts 1, 2, 3, 4, 5), three extracts illustrating complementary relations (Extracts 6, 7, 8,), and two extracts illustrating expanding relations (Extracts 9, 10).

Representativeness of the 10 extracts relative to the three complete picture books was established using four purposive-sampling criteria. First, narrative-stage coverage: extracts were drawn from the orientation, complicating action, and resolution phases of each book so that the sample spans the full story arc rather than clustering at one point in the narrative. Second, salience of verbal-visual interaction: only pages where the illustration and the accompanying text visibly interacted (through shared, added, or diverging meaning) were eligible, excluding pages where images were purely decorative. Third, proportional coverage across sources: each of the three books contributed a comparable number of extracts (three to four each) so that no single book dominated the sample. Fourth, category saturation: extracts were added iteratively until each of Royce's (2013) three relation types was represented by at least two clear instances and additional candidate pages no longer surfaced new patterns of verbal-visual interaction, at which point selection was stopped. Together these criteria support treating the 10 extracts as analytically representative of verbal-visual interplay across the three books rather than as an arbitrary subset.

Instrument and Procedure

This study employed documentation and close reading as the primary data collection techniques. Documentation was used to collect verbal texts and visual illustrations from the selected pages of the picture books. Close reading was applied to examine verbal and visual elements in detail, focusing on linguistic choices, character representation, actions, visual composition, gaze, and verbal-visual interaction. The data collection procedure involved five stages: 1) reading the picture books repeatedly; 2) selecting 10 relevant extracts that demonstrate significant verbal-visual interaction; 3) documenting verbal and visual data from

the selected extracts; 4) categorizing the data based on intersemiotic relation types; and 5) rechecking all data to ensure consistency and accuracy.

Two verification steps were added to strengthen the trustworthiness of the classification. First, investigator triangulation was carried out: the first author produced an initial classification of all 10 extracts, after which the second and third authors independently re-coded the same extracts using Royce's (2013) three-way typology without seeing the first author's labels; classifications were then compared, and the few discrepancies were resolved through discussion until full agreement was reached. Second, peer debriefing was conducted with a colleague experienced in multimodal discourse analysis, who reviewed the coding rationale for each extract and the visual-grammar terminology used to describe it, flagging any classification or description that appeared inconsistent with Royce's (2013) or Kress and Van Leeuwen's (2020) definitions. Extracts flagged in this way were re-examined and, where necessary, re-coded before being finalized in the results reported below.

Data Analysis

Data analysis focused specifically on intersemiotic analysis using Royce's (2013) concept of intersemiotic complementarity within the broader framework of Multimodal Discourse Analysis. Each of the 10 selected extracts was analyzed to identify whether the verbal-visual interaction demonstrated complementary, reinforcing, or expanding relations. The verbal text and visual illustration of each extract were examined simultaneously to determine how both modes worked together in constructing narrative meaning. All visual descriptions in this study follow Kress and Van Leeuwen's (2020) visual grammar terminology consistently: framing and salience for compositional meaning, gaze and social distance for interpersonal meaning, and vector/posture (action) for representational meaning. This terminology is applied uniformly across all ten extracts in the Results and Discussion section below so that lighting, framing, gaze, and posture are described with the same analytic vocabulary rather than ad hoc descriptive language.

RESULT AND DISCUSSION

Intersemiotic Relations in Narrative Meaning Construction

The analysis of 10 selected extracts from the three children's picture books revealed that intersemiotic relations between verbal text and visual illustrations play a fundamental role in constructing narrative meaning. As presented in Table 1 below, reinforcing relations emerged as the most dominant intersemiotic pattern, followed by complementary relations and expanding relations.

Table 1. Summary of Intersemiotic Relations

Intersemiotic Relations	Frequency	Percentage	Function in Narrative Construction
Reinforcing	5	50%	Strengthens and emphasizes meanings already expressed in the verbal text through visual support
Complementary	3	30%	Completes narrative meaning by integrating verbal and visual elements that mutually complete each other
Expanding	2	20%	Provides additional narrative and emotional information beyond what is expressed in the verbal text
Total	10	100%	

Reinforcing Relations

Reinforcing relations are the most dominant intersemiotic pattern in the 10 selected extracts, occurring when visual illustrations directly strengthen the meanings already expressed in the verbal text. This pattern consistently appeared in scenes involving physical movement, environmental challenges, character confrontations, and emotional atmosphere across all three picture books.

Extract 1 (Pages 1-2) – *We're Going on a Bear Hunt*

Verbal Text: We're going on a bear hunt. We're going to catch a big one. What a beautiful day! We're not scared. Oh-oh! Grass! Long, wavy grass. We can't go over it. We can't go under it. Oh, no! We've got to go through it!	Visual 
--	---

In Extract 1, material processes such as *going*, *catch*, and *go through* emphasize physical movement and the family's collective determination. The illustration visually reinforces these verbal meanings by depicting the family actively entering the long grass with energetic body movements and forward-directed postures. The bright and open composition reinforces the cheerful atmosphere established verbally through "What a beautiful day!", while the family's small size relative to

the wide grassy landscape visually reinforces the verbal description of the grass as a significant obstacle. Both modes work together in parallel to construct meanings of adventure, bravery, and determination, confirming reinforcing relations in this extract.

Extract 2 (Pages 3-4) – *We're Going on a Bear Hunt*

Verbal Text:	Visual
Swishy swashy! Swishy swashy! Swishy swashy!	

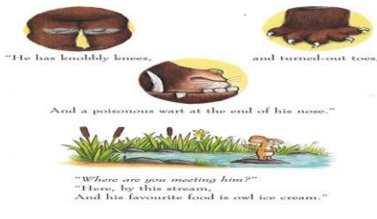
In Extract 2, the repeated onomatopoeic expression "Swishy swashy!" functions as a behavioral process representing the sound and physical sensation of moving through grass. The illustration visually reinforces this verbal meaning by showing the family running energetically across the grassy hill with bright colors and dynamic body movements. The open composition and scattered placement of the characters reinforce the sense of movement, freedom, and playfulness that the verbal sound pattern creates. Neither mode adds new information beyond what the other expresses; instead, both work in parallel to create an immersive and rhythmic reading experience, confirming strong reinforcing relations between the verbal and visual modes.

Extract 3 (Page 11) – *The Tale of Peter Rabbit*

Verbal Text:	Visual
Mr. McGregor was on his hands and knees planting out young cabbages, but he jumped up and ran after Peter, waving a rake and calling out, 'Stop thief!'	

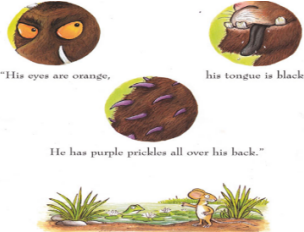
In Extract 3, multiple material processes such as *planting*, *jumped up*, *ran after*, and *waving* create a dynamic and tense chase sequence. These verbal actions are directly reinforced by the illustration depicting Mr. McGregor aggressively chasing Peter with a rake. Mr. McGregor's threatening posture and Peter's stretched, panicked body posture visually reinforce the urgency and danger already established verbally. The eye-level perspective strengthens emotional involvement by placing readers directly within the action. Both modes consistently reinforce each other in constructing a coherent and powerful representation of conflict, danger, and pursuit.

Extract 4 (Page 7) – *The Gruffalo*

Verbal Text: He has knobbly knees, and turned-out toes, And a poisonous wart at the end of his nose. Where are you meeting him? Here, by this stream, And his favourite food is owl ice cream.	Visual  "He has knobbly knees," and turned-out toes, And a poisonous wart at the end of his nose," "Where are you meeting him?" "Here, by this stream, And his favourite food is owl ice cream."
---	---

In Extract 4, relational processes such as *has* and *are* are used to verbally construct the gruffalo's frightening physical attributes, including "knobbly knees", "turned-out toes", and "a poisonous wart at the end of his nose". These verbal descriptions are directly reinforced by the illustrations presenting the gruffalo's body parts in prominent and vivid visual detail at the top of the page. The composition directs readers' attention to the same grotesque features described verbally, making the creature's threatening image more vivid and memorable. The visual and verbal modes operate in parallel without adding new information, confirming reinforcing relations in this extract.

Extract 5 (Page 11) – *The Gruffalo*

Verbal Text: His eyes are orange, his tongue is black; He has purple prickles all over his back. Where are you meeting him? Here, by this lake, And his favourite food is scrambled snake.	 "His eyes are orange," his tongue is black; He has purple prickles all over his back." "Where are you meeting him?" "Here, by this lake, And his favourite food is scrambled snake."
--	--

In Extract 5, the mouse continues describing the gruffalo's frightening features to the snake using relational processes *are* and *has*. The verbal descriptions of orange eyes, a black tongue, and purple prickles are directly reinforced by the illustrations visually presenting these specific features in striking detail. The sharp eyes, black tongue, and purple prickles create a frightening visual impression that parallels the verbal descriptions precisely. The mouse appears calm and confident near the lake while the gruffalo's body parts dominate the composition, reinforcing the contrasting emotional positions of the two characters established verbally. Both modes work together to construct a consistent and intensified image of the gruffalo's dangerous appearance, confirming reinforcing relations.

Complementary Relations

Complementary relations account for 30% of the 10 selected extracts and occur when verbal and visual elements complete each other in constructing meaning, with each mode contributing information that the other does not fully express on its own. These relations are particularly evident in extracts where emotional, atmospheric, and characterization meanings are distributed across both modes.

Extract 6 (Page 2) – *The Tale of Peter Rabbit*

Verbal Text:	Visual
"Now, my dears," said old Mrs. Rabbit one morning, "you may go into the fields or down the lane, but don't go into Mr. McGregor's garden."	 "Now, my dears," said old Mrs. Rabbit one morning, "you may go into the fields or down the lane, but don't go into Mr. McGregor's garden."

In Extract 6, the verbal text constructs Mrs. Rabbit's warning through contrasting material processes *may go* and *don't go*, establishing the central narrative tension between permission and prohibition. However, the verbal text alone does not fully express the emotional dimension of maternal authority and care. The illustration completes this meaning by visually positioning Mrs. Rabbit as the dominant figure through her central placement and larger body size, with her body oriented attentively toward her children. Furthermore, the illustration subtly shows Peter Rabbit positioned slightly apart from his siblings, visually foreshadowing his disobedience in a way that the verbal text has not yet expressed. The verbal text provides the narrative content while the illustration contributes the emotional and characterization layers, confirming complementary relations.

Extract 7 (Pages 17-18) – *We're Going on a Bear Hunt*

Verbal Text:	Visual
Tiptoe! Tiptoe! Tiptoe! WHAT'S THAT?	 

In Extract 7, the repeated material process *Tiptoe!* emphasizes cautious movement while the sudden exclamation "WHAT'S THAT?" introduces fear and uncertainty. However, the verbal text alone does not convey the dark, claustrophobic atmosphere of the cave that intensifies the danger. The illustration completes this meaning by visually depicting the family moving carefully through an enclosed cave environment using darker tones, tight framing, and tense body postures. The cave's visual darkness contributes an atmosphere of mystery and psychological

tension that the verbal text only implies. The verbal text provides the action and emotional shift while the illustration supplies the spatial and atmospheric context, confirming complementary relations.

Extract 8 (Page 2) – *The Gruffalo*

Verbal Text:	Visual
A mouse took a stroll through the deep dark wood. A fox saw the mouse and the mouse looked good. "Where are you going to, little brown mouse? Come and have lunch in my underground house." "It's terribly kind of you, Fox, but no— I'm going to have lunch with a gruffalo."	A mouse took a stroll through the deep dark wood. A fox saw the mouse and the mouse looked good. "Where are you going to, little brown mouse? Come and have lunch in my underground house." "It's terribly kind of you, Fox, but no— I'm going to have lunch with a gruffalo."  "A gruffalo? What's a gruffalo?" "A gruffalo! Why didn't you know?"

In Extract 8, the verbal text presents the first encounter between the mouse and the fox. The process *saw* signals the fox's predatory interest while the mouse's polite refusal introduces the gruffalo as a strategy for survival. However, the verbal text alone does not fully convey the visual power imbalance between the two characters. The illustration completes this meaning by depicting the fox as physically much larger than the mouse, visually communicating the predator-prey dynamic that underlies the seemingly polite social exchange. The fox's gaze toward the mouse suggests predatory intent while the mouse stands confidently despite its smaller size, visually contributing the emotional dimension of the mouse's cleverness and composure. The verbal text provides the narrative dialogue while the illustration completes the power dynamics and emotional tension, confirming complementary relations.

Expanding Relations

Expanding relations account for 20% of the 10 selected extracts and occur when one mode provides additional information beyond what is expressed in the other mode, enriching the narrative with emotional depth, symbolic significance, and interpretive layers that extend the verbal or visual text. Although the least frequent, these relations contribute the most distinctive dimension of meaning to the narrative.

Extract 9 (Pages 25-26) – *We're Going on a Bear Hunt*

Verbal Text:	Visual
We're not going on a bear hunt again.	 We're not going on a bear hunt again.

The verbal statement "We're not going on a bear hunt again" functions as the narrative resolution, expressing the family's fear and decision not to repeat the

dangerous adventure. However, this single sentence does not fully convey the emotional dimensions of comfort, warmth, and family closeness that characterize this moment. The illustration expands the verbal meaning significantly by showing the entire family hiding closely together under soft blankets inside the house, with warm colors and gentle lighting creating a powerful sense of safety and togetherness. The illustration also symbolically contrasts the warm indoor setting with the dangerous outdoor environments encountered throughout the journey, adding a layer of emotional resolution and symbolic meaning that extends well beyond what the verbal text expresses. These expanded meanings reward readers with a deeper emotional understanding of the story's conclusion, confirming expanding relations.

Extract 10 (Page 12) – *The Tale of Peter Rabbit*

Verbal Text:	Visual
Peter was most dreadfully frightened; he rushed all over the garden, for he had forgotten the way back to the gate. He lost one of his shoes among the cabbages,	 PETER was most dreadfully frightened; he rushed all over the garden, for he had forgotten the way back to the gate. He lost one of his shoes among the cabbages.

In Extract 10, the verbal text explains Peter's fear and confusion through processes such as *frightened*, *rushed*, *forgotten*, and *lost*. While these processes convey Peter's panic and physical loss, the illustration expands the emotional meaning significantly by focusing not on Peter himself but on his abandoned shoe lying alone among the cabbages. This visual choice symbolically communicates vulnerability, helplessness, and psychological disorientation in ways that the verbal description alone cannot fully achieve. The absence of Peter's body from the illustration further expands the emotional impact by emphasizing loneliness and the sense of being completely overwhelmed by the situation. The illustration therefore adds a rich layer of symbolic emotional meaning that extends significantly beyond the narrative information provided in the verbal text, confirming expanding relations.

The findings from the 10 selected extracts demonstrate that narrative meaning in children's picture books is fundamentally constructed through the continuous interplay between verbal text and visual illustrations via intersemiotic relations. The dominance of reinforcing relations (50%) across the selected extracts reflects the collaborative nature of verbal-visual storytelling in children's picture books, in which both modes consistently work together to strengthen and amplify core narrative meanings. As demonstrated in Extracts 1, 2, 3, 4, and 5, reinforcing relations ensure that readers receive consistent and strengthened meaning across both verbal and visual channels, which is particularly important for young readers who are still developing their narrative comprehension skills. This finding aligns

with Royce's (2013) argument that meaning in multimodal texts is constructed through the interaction between semiotic modes.

The presence of complementary relations (30%) in Extracts 6, 7, and 8 indicates that verbal and visual modes frequently take on distinct but mutually supportive roles in meaning construction. Rather than simply repeating the same information, verbal text and illustrations contribute different types of meaning that together form a complete narrative understanding. This finding supports Hu and Qiu's (2020) explanation that intersemiotic relations enrich meaning-making processes because readers construct interpretation through the interaction between words and images. Complementary relations are particularly significant in constructing emotional dimensions, characterization, and atmospheric context that neither mode could fully achieve independently.

The expanding relations (20%) in Extracts 9 and 10, while the least frequent, contribute the most distinctive layer of meaning by adding emotional depth, symbolic significance, and narrative richness beyond the explicit verbal content. These relations are especially important in moments of emotional resolution and psychological vulnerability, where the illustration elevates the narrative through symbolism and visual metaphor. This finding supports Shi's (2023) argument that visual-verbal interaction contributes significantly to multimodal meaning-making and visual literacy in children's picture books.

Beyond emotional enrichment, expanding relations play a distinct role in fostering critical visual literacy among young readers. Because expanding relations require the reader to notice information that exists only in the image the discarded shoe in Extract 10, the sheltering blankets in Extract 9 they demand an interpretive move beyond straightforward decoding: the child must infer meaning, connect symbol to emotion, and tolerate a degree of ambiguity that reinforcing relations, by design, do not require. Serafini (2014) frames this kind of interpretive work as central to visual literacy, in which readers learn to treat images as sites of meaning-making rather than as illustrations of a pre-established verbal meaning. Repeated exposure to expanding relations, therefore, plausibly trains exactly the inferential and symbolic reading skills that later support critical engagement with more complex multimodal texts; because reinforcing relations dominate the sampled extracts (50%) while expanding relations are comparatively rare (20%), this also suggests that picture books deliberately balance redundancy which supports comprehension and confidence for emergent readers against sparser, higher-demand expanding relations that stretch interpretive skill, rather than maximizing either strategy alone. Overall, the analysis confirms that verbal text and visual illustrations cannot be treated as independent meaning-making resources in children's picture books. Narrative meaning emerges precisely through their continuous interaction via reinforcing, complementary, and expanding relations, demonstrating that picture books must be understood as integrated multimodal literary texts in which both language and image are equally significant contributors to narrative construction.

CONCLUSION

This study examined how intersemiotic relations between verbal text and visual illustrations contribute to narrative meaning construction in selected children's picture books. The analysis of 10 selected extracts from *The Gruffalo*, *The Tale of Peter Rabbit*, and *We're Going on a Bear Hunt* confirmed that reinforcing relations are the most dominant intersemiotic pattern (50%), followed by complementary relations (30%) and expanding relations (20%). These findings demonstrate that narrative meaning in children's picture books is fundamentally multimodal because it emerges through the continuous interplay between verbal and visual semiotic modes rather than through either mode alone. The analysis of the 10 selected extracts further demonstrates that each type of intersemiotic relation contributes differently to narrative construction. Reinforcing relations amplify core narrative meanings through parallel verbal-visual expression. Complementary relations distribute meaning across both modes so that each contributes what the other cannot fully express independently. Expanding relations enrich the narrative by adding emotional depth, symbolic meaning, and interpretive layers that extend beyond the explicit verbal content. Together, these three types of relations enable children's picture books to construct rich, multi-layered narratives that engage readers both linguistically and visually.

The study contributes to Multimodal Discourse Analysis by providing empirical evidence that intersemiotic relations play a critical role in narrative construction in children's picture books. Building on this evidence, four concrete pedagogical recommendations are offered for early childhood educators. First, conduct a guided "picture walk" before reading aloud, pausing on pages with strong reinforcing relations (e.g., Extract 1, 4) so children learn to match action words to depicted actions, building foundational word-image correspondence. Second, use explicit visual-inference questions such as "What does the picture tell us that the words don't?" on pages carrying complementary or expanding relations (e.g., Extract 6, 9, 10) to scaffold symbolic and emotional interpretation rather than literal decoding. Third, design classroom activities that have children compare a page's words and image side by side and sort the relationship as "same," "both add something different," or "picture tells more," giving young readers accessible, age-appropriate language for the three relation types identified in this study. Fourth, provide targeted training for early childhood educators in basic visual-grammar concepts (framing, gaze, salience) so they can more confidently guide picture-book discussions that draw on both verbal and visual evidence. Future researchers are encouraged to examine intersemiotic relations in picture books from different cultural backgrounds and genres, and to test these classroom recommendations empirically, to obtain broader comparative and pedagogical findings.

REFERENCES

- Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication*, 25(2), 166–195. <https://doi.org/10.1177/0741088307313177>
- Callow, J. (2020). Visual and verbal intersections in picture books: Multimodal assessment for middle years students. *Language and Education*, 34(2), 115–134.
- Gao, Y. (2023). Complementarity between language and image in multimodal texts of classical Chinese poems. *TASMANIA*, 100.
- Hu, Y., & Qiu, Q. (2020). A study on verbal and image relations in multimodal texts from the perspective of intersemiotic complementarity. *Canadian Social Science*, 16(10), 50–56. <https://doi.org/10.3968/11938>
- Kress, G., & Bezemer, J. (2023). Multimodal discourse analysis. In *The Routledge handbook of discourse analysis* (pp. 139–155). Routledge.
- Kress, G., & van Leeuwen, T. (2020). *Reading images: The grammar of visual design*. Routledge.
- Marmion, P. I. (2025). *Exploring multimodal storytelling in contemporary award-winning picture books: A stylistic study* [Doctoral dissertation, Vilniaus universitetas].
- Moya Guijarro, A. J. (2014). *A multimodal analysis of picture books for children: A systemic functional approach*. University of Toronto Press.
- 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.
- Painter, C. N., Martin, J. N., & Unsworth, L. N. (2013). *Reading visual narratives: Image analysis of children's picture books*. University of Toronto Press.
- Royce, T. D. (2013). Intersemiotic complementarity: A framework for multimodal discourse analysis. In *New directions in the analysis of multimodal discourse* (pp. 63–109). Routledge.
- Serafini, F. (2014). *Reading the visual: An introduction to teaching multimodal literacy*. Teachers College Press.
- Serafini, F. (2024). The complex relationship of words and images in picturebooks. *Journal of Visual Literacy*, 43(3), 233–249. <https://doi.org/10.1080/1051144X.2024.2394338>
- Shi, D. (2023). Intermodality and visual literacy: Exploring visual-verbal instantiation in children's picture books on coronavirus. *Journal of Visual Literacy*, 42(3), 183–209.
- Svensson, B. (2020). The enacting of emotions in two picture books: Interpretations through a multimodal discourse analysis. *International Journal of Language Studies*, 14(2).
- Tsapiv, A. (2022). Multimodal imagery in picture books for children. *Cognition, Communication, Discourse*, 25, 80–88.
- van Leeuwen, T. (2023). Social semiotics and multimodality. In *The Routledge handbook of applied linguistics* (pp. 320–336). Routledge.

Xie, J., Noor, M., & Hamed, P. K. (2023). Exploring the visual narrative characteristics of Chinese classical poetry picture books in multimodal discourse analysis. *Eurasian Journal of Educational Research*, 108.