

Multimodal Text Analysis of a Student's Book for Grade VII of Junior High School

Ratih Sri Pramestri, Joko Priyana

Universitas Negeri Yogyakarta, Indonesia

Email: ratihsri.2021@student.uny.ac.id, joko.priyana@uny.ac.id

ABSTRACT

This research investigates how multimodal texts in the English for Nusantara coursebook for Grade VII junior high school construct meaning through the integration of verbal and visual elements. The coursebook combines written language with diverse visual modes to support meaning-making in classroom contexts. Recognizing the importance of multimodality in English language education, the study applies Halliday's Systemic Functional Linguistics (SFL) to analyze how verbal texts realize ideational, interpersonal, and textual meanings. Kress and van Leeuwen's Visual Grammar is used to examine how visuals construct representational, interactional, and compositional meanings, while Martinec and Salway's framework explains how verbal and visual modes interact. A qualitative content analysis method was employed on three purposively selected chapters—About Me, Culinary and Me, and Home Sweet Home—for their rich multimodal content. Verbal data, including dialogues and reading passages, were segmented into clauses and analyzed according to SFL metafunctions. Visual data, such as illustrations and contextual images, were examined for narrative, conceptual, interactional, and compositional features. Findings indicate that verbal texts primarily realize ideational meanings through material, relational, and mental processes, while visual texts frequently employ actional, reactional, and analytical representations. Image-text relations show strong tendencies toward equal status, with elaboration and enhancement as the most common logico-semantic relations. Overall, meaning is collaboratively constructed across modes, highlighting the layered complexity of multimodal educational texts.

KEYWORDS *Image-Text Relations, Meaning-Making, Multimodal Text Analysis, SFL, Visual Grammar*



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International

INTRODUCTION

Multimodality refers to the integration of multiple communicative modes, such as visual, auditory, textual, imagery, video, and musical elements, which together convey meaning in communication (Hasyim & Arafah, 2023; Sandler, 2022). It plays a crucial role in language acquisition since nonverbal components enhance learners' comprehension and motivation (Dressman, 2019; Renandya, 2022). Coursebooks are no longer limited to written text but include layouts, icons, images, colors, and sometimes audio or video, all of which support learners in understanding and using language effectively (Fitria & Simbolon, 2024). These multimodal resources allow students, particularly beginners, to better interpret meaning through visual illustrations and contextual relations embedded in coursebook design (Lim & Tan-Chia, 2022; Weninger, 2021).

The integration of multimedia elements in coursebooks enhances both the aesthetic appeal and pedagogical effectiveness of language learning materials (Suman, 2023). Visuals stimulate imaginative thinking and improve comprehension by supporting the written content (Mohammed, 2024). Research shows that multimedia-rich coursebooks also address diverse learning styles (Jana, 2023). However, the effectiveness of multimodality depends on how well different modes are integrated (Boehm et al., 2022; Pawłowski et al., 2023). If visuals or sounds

are used merely for decoration, they may distract students from the intended message, lowering comprehension and reducing the learning experience (Renandya, 2022).

Meaning-making in multimodal discourse requires the complementary use of semiotic resources, such as text, images, and layout, to construct coherent messages (Ilonga & Mapunda, 2022). Poorly designed multimodal elements can hinder comprehension, as irrelevant or excessive visuals may distract learners from essential information (Alabi, 2024; Hoque, 2025; Rohi & Nurhayati, 2024). Therefore, careful multimodal design is essential to balance textual and non-textual resources and to foster effective learning (Holtlien, 2025; Vargas & Trevisan, 2025). Clear and coherent layouts engage students' cognitive skills and improve usability, motivation, and learning outcomes (Behnke, 2021; Ghai & Tandon, 2022; Permata et al., 2024; Wong & Hughes, 2023).

Several studies have analyzed multimodal discourse in coursebooks. International research highlights the interplay of images and text in constructing meaning (Adami, 2023; Cheng, 2021; Drajadi et al., 2025). In the EFL context, studies such as Elmiana & Shen (2025) and Stewart (2025) show that visuals support comprehension, particularly for beginners, by scaffolding learning through image-text relations. Similar findings in China (Haiyan, 2018) emphasize that diverse image-text relations promote multimodal literacy aligned with learners' proficiency. In Indonesia, Elmiana (2019), Fitriana (2021), and Lestari (2022) applied Kress and van Leeuwen's framework to EFL coursebooks, revealing both pedagogical benefits and limitations, such as issues of representation, design, and visual-textual integration.

Despite this growing body of research, gaps remain in the analysis of multimodal discourse in new Indonesian EFL coursebooks (Barus et al., 2024; Tafrijiyah & Andriani, 2025). Previous studies have mostly relied on Kress and van Leeuwen's framework, while Martinec and Salway's image-text relation model has not been widely applied. To address this, the present study investigates multimodal texts in the *English for Nusantara* coursebook for grade VII, published by the Ministry of Education and Culture under the *Kurikulum Merdeka*. Using Systemic Functional Linguistics, Kress and van Leeuwen's Visual Grammar, and Martinec and Salway's framework, this research examines how verbal and visual elements interact to create meaning. The study aims to evaluate how well multimodality in this coursebook supports learners' engagement, comprehension, and development of multimodal literacy.

This study identifies several issues related to multimodality in the English language textbook, particularly the latest *English for Nusantara* coursebook for junior high schools. Although various previous studies have analyzed multimodal elements in EFL textbooks in Indonesia, none have specifically focused on this coursebook, which is an important resource within the *Kurikulum Merdeka* framework. Most prior research has examined older textbooks or those aimed at different education levels, such as senior high school books. Additionally, many analyses have utilized the Kress and van Leeuwen framework which, while robust, does not fully address the interaction between images and text, creating a gap in the exploration of image-text relations.

This research aims to analyze multimodal texts in the *English for Nusantara* coursebook for Grade VII students, published by the Ministry of Education and Culture of Indonesia. The primary focus is on how these multimodal elements interact to create and convey meaning within the educational context. By employing a Multimodal Discourse

Analysis (MDA) approach and integrating Kress and van Leeuwen's Visual Grammar (VG) for visual elements, as well as Halliday's Systemic Functional Linguistics (SFL) for verbal elements, this research will explore the relationships between these two elements. Consequently, the findings of this study are expected to provide valuable insights into the utilization of multimodal resources in government-endorsed textbooks and contribute to future textbook development.

RESEARCH METHOD

This research employed a qualitative methodology to analyze multimodal elements in the English for Nusantara EFL coursebooks for junior high school, focusing on verbal and visual texts and their contributions to meaning making. Using content analysis, the study systematically examined communication content to uncover meanings, contexts, and intentions. The analysis applied Systemic Functional Linguistics (SFL) to verbal texts and Kress and van Leeuwen's Visual Grammar to visual texts, while Martinec and Salway's framework was used to assess the relationship between image and text. The unit of analysis comprised multimodal texts from three chapters—"About Me," "Culinary and Me," and "Home Sweet Home"—selected for their rich integration of verbal and visual elements. Data collection involved reading the coursebook and organizing findings using analytical tables. Trustworthiness was ensured through expert validation and peer review. Data analysis followed a structured approach, coding verbal texts for ideational, interpersonal, and textual meanings, while visual elements were examined for representation, interaction, and composition. This analysis provided insights into how these multimodal resources co-constructed meaning within the educational context.

RESULT AND DISCUSSION

Research Findings

1. Verbal Text Analysis

Ideational metafunctions

Transitivity analysis in the verbal text of the book "English for Nusantara" class VII revealed an interesting distribution pattern in terms of processes, participants, and circumstances. The data obtained from the analysis of 520 clauses shows that material processes dominate with the highest frequency.

Table 1. Distribution of Process Types, Participants, and Circumstances

Aspect	Type	Frequency	Percentage
Process	Material	215	41%
	Mental	83	16%
	Relational	133	25%
	Verbal	23	4%
	Behavioral	45	9%
	Existential	21	4%
Subtotal		520	100%
Participant	Actor	197	20%
	Goal	194	20%
	Senser	82	8%

Aspect	Type	Frequency	Percentage
	Phenomenon	85	8%
	Carrier	75	8%
	Attribute	75	8%
	Token	58	6%
	Value	58	6%
	Behaver	45	5%
	Verbiage	20	2%
	Exist	21	2%
	Receiver	8	1%
Subtotal		983	100%
Circumstance	Location	108	45%
	Cause	39	16%
	Matter	29	12%
	Manner	22	9%
	Accompaniment	19	8%
	Extent	16	7%
	Role	4	2%
	Angle	2	1%
Subtotal		239	100%

Source: Primary data analysis

Material processes dominated with 41% of the total, representing physical activity and concrete actions that emphasized textbook orientation to the student's practical experience, as in the clauses "I live on Jalan Sumatera" and "Write his identity in the empty boxes." Furthermore, mental processes occupy 16% and include cognition, perception, and emotions, such as "I like fishing" and "You do need what for fishing" (desiderative process), which show their important role in training the expression of internal states. The relational process of 25% is divided into attributions, such as "She has more than twenty novels," and identification, such as "Their favorite badminton player is Jonathan Christie," which supports the description and identification of vocabulary and basic structure. Other processes include verbal (4%) such as "Galang says Basmallah before eating," behavioral (9%) such as "You listen to Audio 1.3," and existential (4%) such as "There are two parts in the audio," which despite the small percentage still enriches the semantic diversity of the text in the textbook.

Metafunksi Interpersonal

An analysis of interpersonal metafunctions reveals how textbooks construct relationships and social roles through mood systems and modalities. The mood distribution shows a pattern that reflects the pedagogical orientation of the textbook.

Table 2. Distribution of Mood and Modality Systems

Aspect	Type	Frequency	Percentage
Mood	Declarative	307	59%
	Imperative	164	32%
	Interrogative	47	9%
Subtotal		518	100%
Modality	Bond – High	10	53%

Aspect	Type	Frequency	Percentage
	Obligation – Medium	9	47%
	Inclination – High	4	9%
	Inclination – Medium	8	18%
	Inclination – Low	33	73%

Source: Primary data analysis

Declarative clauses dominate with 59% of the total clauses, confirming the main function of textbooks as conveyors of information, as seen in the simple structure of "I like mobile gaming" to the complex form of "One box has been completed for you" that utilizes passive construction; This dominance creates an informative and authoritative tone that fits the instructional purpose. The imperative clause occupies 32%, reflecting a strong directive orientation to instruct, both in complex forms such as "See the Wordbox at the end of the chapter to help you find the meanings of key wordings" or simple ones such as "Complete a mind map about Andre," both affirm an interactive approach to learning. Meanwhile, interrogative clauses are only 9%, but they play an important role in creating dialogical dynamics, for example in the yes/no questions "Are you enjoying the first day of school?" or "Can I come to your house?" which incorporate modulation of low inclination, making room for student responses. In terms of modality, bond modulation is divided relatively evenly between high (53%) and medium (47%) levels, as in "Plastic and glass bottles should be separated in the recycle bin" which indicates moderate liability, while inclination modulation is dominated by low levels (73%), e.g. "I want to go fishing," which emphasizes personal desires without external pressure. The overall distribution reflects a pedagogical strategy that balances information delivery, direction, interaction space, and respect for students' autonomy in the learning process.

Textual metafunctions

Textual metafunctions govern how information is organized and presented in a clause structure. Thematic and rema analysis reveals a consistent pattern of information organization in textbooks.

Table 3. Theme and Rema Distribution

Aspect	Type	Frequency	Percentage
Theme	Topical	480	80%
	Interpersonal	76	13%
	Textual	43	7%
Subtotal		599	100%
Markedness	Unmarked	428	83%
	Marked	88	17%
Subtotal		516	100%

Source: Primary data analysis

Topical themes dominate with 80% of the overall theme, reflecting the focus of experiential content and providing structural predictability that aids student understanding, as seen in the simple "It's great" or interrogative "Can you talk about yourself?". Interpersonal themes that amounted to 13% played a role in creating interactive nuances and emotional closeness, such as "What's your name?" and "You know I like school," which showed efforts

to build social engagement with readers. Meanwhile, textual themes with a frequency of 7% support cohesion and logical flow between clauses, as shown in "But I can't wait to go home" and "Or you can listen to both parts in one go," which mark contrast and alternative. Markedness analysis revealed the dominance of unmarked themes at 83% which gave structure ease of following, while 17% markedness themes were strategically used to provide emphasis and variety. These proportions show a balance between clarity of structure for ease of understanding and variations of form to keep reader engaged.

2. Visual Text Analysis

Representation Patterns

The analysis of visual representation patterns reveals how images in textbooks construct meaning through narrative and conceptual processes. The data obtained showed a distribution that reflected pedagogical priorities in visual representation.

Table 4. Distribution of Visual Representation Patterns

Aspect	Type	Frequency	Percentage
Narrative	Actional	54	43%
	Reactional	46	37%
	Projective	26	21%
Subtotal		126	100%
Conceptual	Analytical	31	69%
	Classificational	9	20%
	Symbolic	5	11%
Subtotal		45	100%

Source: Primary data analysis

The actional narrative process dominates 43% with movement visuals such as the Galang comic strip, helping students understand language through everyday activities. The reactionary process (37%) was evident from expressions and perceptions, for example students' reactions to foods they liked or rejected, which showed emotional nuances. The projective process (21%) is present through speech/thought balloons such as the greeting "Hi" which encourages active student participation. In the conceptual category, analytics dominated 69% with a section-whole chart, a 20% classification through the grouping of items such as "Ingredients" and "Tools," and a symbolic 11% such as a "Please Take Off Your Shoes" sticker that affirmed cultural norms. This distribution shows a balance between representations of actions, emotions, interactions, concepts, and cultural values in language learning.

Interaction Patterns

The pattern of visual interaction determines how the reader is positioned in relation to the participants represented in the image. The analysis reveals a consistent strategy in building visual interpersonal relationships.

Table 5. Distribution of Visual Interaction Patterns

Aspect	Type	Frequency	Percentage	Naturalistic	Abstract
Contact	Demand	15	16%		
	Offer	76	84%		
Subtotal		91	100%		
Social Distance	Close shot	10	11%		
	Medium shot	63	67%		
	Long shot	21	22%		
Subtotal		94	100%		
Attitude	Subjective	40	44%		
	Objective	51	56%		
Subtotal		91	100%		
Visual Modality	Color Saturation			89 (100%)	0 (0%)
	Color Differentiation			89 (100%)	0 (0%)
	Color Modulation			3 (3%)	86 (97%)
	Contextualization			56 (63%)	33 (37%)
	Detail			2 (2%)	87 (98%)
	Depth			8 (9%)	81 (91%)
	Lighting			4 (4%)	85 (96%)
	Brightness			89 (100%)	0 (0%)

Source: Primary data analysis

Visual analysis showed that the contact pattern was dominated by the offer type (84%), where participants did not look at the reader directly, thus placing the reader as an objective observer, as in the image of Made playing wheelchair basketball. In contrast, the demand type (16%) was used strategically, for example in the comic strip of Ibu Ida staring at the reader while teaching pronouns, thus creating direct engagement. In terms of social distancing, medium shots dominated (67%) with body framing from the waist up, such as the image of Monita with two plates of food, which balanced emotional closeness and visual context; while close shots (11%) were used for intimate details such as Monita's shopping list, and long shots (22%) were used to give a broad context such as Galang's house. Visual attitudes are relatively evenly divided between a neutral objective perspective (56%), such as a diagram of Galang's house, and a subjective perspective (44%) that presents the reader in a situation, such as Pak Rahmansyah's conversation with Galang. The visual modality shows a balance between realism and abstraction: 100% naturalistic color saturation and differentiation favor recognition, while other aspects such as color modulation (97%), detail (98%), depth (91%), and lighting (96%) are more abstract, thus simplifying visual representation without losing familiarity, while reducing the cognitive load for beginners.

2.3 Composition Patterns

Composition patterns govern how visual elements are organized in space to create a hierarchy of meaning and guide the reader's attention. The analysis reveals the consistent and purposeful use of compositional principles.

Table 7. Composition Pattern Distribution

Aspect	Type	Frequency	Percentage
Information Value	Left–Right	56	50%
	Top–Bottom	38	34%
	Center–Margin	17	15%
Subtotal		111	100%
Salience	Size – Large	87	98%
	Size – Small	2	2%
	Color Contrast – Strong	89	100%
	Sharpness – Sharp	89	100%
Framing	Strong	6	6%
	Weak	89	94%
Subtotal		95	100%

Source: Primary data analysis

The structure of the information value is dominated by a left–right pattern (50%) that follows the reading convention, placing the illustration as Given and the empty information box as New, thus creating a logical flow from example to practice; followed by an up–down pattern (34%) that distinguishes the Ideal realm of symbolic objects at the top from the Real realm of concrete tasks below, as well as the center–edge pattern (15%) that emphasizes the main focus on the central element while presenting a supporting visual context. Salience is achieved through dominant size (98%), strong color contrast (100%), and consistent focus sharpness (100%) that ensure important elements are clearly visible, with strategic placement in high-traction areas. In terms of framing, the majority of representations used weak framing (94%) to create cohesive visual integration, while strong framing (6%) was used strategically to separate conceptual information and tasks to support a more focused processing and learning sequence.

3. Image-Text Relationship

3.1 Relative Status

Analysis of image-text relationships reveals a sophisticated pattern of multimodal integration in the construction of meaning. The relative distribution of status shows a clear preference for collaborative meaning-making between the visual and verbal modes.

Table 8. Relative Status of Image-Text

Aspects	Jenis	Frequency	Percentage
Equal Status	Independent	0	0%
	More complementary	76	85%
Unequal Status	Subordinate image	12	13%
	Subordinate text	1	1%
Total		89	100%

Source: Primary data analysis

Complementary relationships dominate 85% of multimodal units, demonstrating advanced integration in which the visual and verbal complement each other to convey meaning, as in *Comic Strip 1.4* that combines body expressions and word balloons in student introductions or house diagrams with labels that connect spatial understanding with vocabulary. This relationship provides a dual path of understanding that is very beneficial for beginners. In contrast, the relationship of subordinate images to text accounts for 13%, as in *Picture 2.4* of Nasi Goreng, where text becomes the center of instruction while images only reinforce understanding. The relationship of subordinate text to image is very limited (1%), for example *Picture 2.1* about Monita and her favorite food, where visuals dominate the meaning while text only serves as a general topic label. The absence of independent relationships (0%) confirms that all multimodal units are designed to be integrated to reinforce learning objectives.

3.2 Logical-semantic relationships

Logical-semantic relationships reveal how meaning is constructed and developed across visual and verbal modes through different types of semantic connections.

Table 9. Logical-Semantic Relationships

Aspects	Jenis	Frequency	Percentage
Expansion	Elaboration	84	47%
	Extension	19	11%
	Enhancement	53	30%
Projection	Loci	23	13%
	Goes	0	0%
Total		179	100%

Source: Primary data analysis (2023)

Elaboration relationships dominate 47% of logico-semantic connections with a pattern of repetition or re-explanation between text and images, such as in Text 1.3 about hobbies that display students' visual activities and then reinforce descriptive texts. Enhancement covers 30% by adding details of time, place, cause, or culture, for example in Comic Strip 2.1 about a favorite food that links the visual of fried bananas with verbal explanations and religious context. Extension (11%) expands the scope of information by adding instrument details or personal preferences that aren't visualized, while Projection (13%) appears through word balloons that project direct speech in interactions, such as a home tour by Galang. The absence of projection ideas (0%) indicates the development space to represent the character's thoughts or feelings. This combination forms a multi-layered multimodal approach, from strengthening basic understanding through elaboration to complex synthesis through enhancement and extension, as well as authentic representation of interaction through projection.

1. Construction of Meaning in Verbal Texts

The analysis of ideational metafunctions in the book "English for Nusantara" reveals a pattern of strategic meaning construction for beginner-level English language learning. The dominance of material processes (41%) is in line with the findings of Bondar et al. (2018) in the context of EFL Indonesian textbooks, but shows a more balanced distribution with other

processes. These findings reflect a pedagogical orientation that emphasizes grounding language learning in concrete experiences and observable actions, in accordance with the principle of Halliday & Matthiessen (2014) that material processes are fundamental building blocks for representing human experience.

The presence of substantial mental processes (16%) distinguishes this book from the research of Niman & Darong which found limited mental process representation in other EFL Indonesia textbooks. This more generous distribution indicates a conscious effort to develop students' capacity in expressing internal states, emotions, and cognitive processes. This is critical for developing holistic communicative competence, where students not only describe external actions but also articulate personal feelings, preferences, and thoughts that are essential for meaningful interaction.

The relational process (25%) functions as a bridge between concrete actions and abstract concepts, providing linguistic resources for describing, identifying, and categorizing. This distribution supports vocabulary development and conceptual organization which is fundamental for beginner learners. Integration with behavioral, verbal, and existential processes, albeit with a more limited frequency, shows an attempt to provide semantic diversity without overwhelming cognitive capacity students.

From the perspective of interpersonal metafunctions, the dominance of declarative clauses (59%) reflects the information-delivery model characteristic of instructional materials, where coursebooks are positioned as authoritative knowledge sources. However, substantial presence imperative clauses (32%) indicate a shift toward a more interactive pedagogical approach that encourages active student participation. This aligns with the contemporary language teaching philosophy that emphasizes learner agency and active engagement, consistent with the Merdeka Belajar curriculum that promotes student-centered learning (Kemendikbudristek, 2022).

The use of modality systems, particularly inclination modulation with the dominance of low level (73%), reflects cultural sensitivity in the Indonesian educational context which values politeness and indirect communication. This finding resonates with Cortazzi & Jin's research on cultural values in language learning materials, where softness in expressing suggestions or preferences aligns with Indonesian social norms that avoid direct imposition.

Textual metafunctions showed a preference for unmarked topical themes (83%) which provides structural predictability essential for beginner-level comprehension. Consistency in thematic organization creates scaffolding that helps students process new information efficiently. These findings support the research of Rosi & Ningrum (2019) which emphasizes the importance of thematic regularity in EFL materials for facilitating information processing and building reading fluency.

2. Construction of Meaning in Visual Text

The visual representation patterns in this book demonstrate a sophisticated understanding of visual grammar principles to support language learning. The dominance of narrative processes (74% of total representation) with actional processes as the most prominent (43%) aligns with the Qi research (2024) which found the effectiveness of action-based visuals in concretizing abstract language concepts. Vector-based representations that show clear Actor-Goal relationships create immediately recognizable narratives that are accessible to beginner learners.

Substantial presence reactional processes (37%) indicate conscious effort to incorporate emotional engagement dimensions in visual design. This contrasts with Serafini's criticism of limited emotional representation in educational materials. "English for Nusantara" appears to prioritize affective dimension learning through expressive visuals that enable students to recognize and interpret emotional states, crucial for developing interpersonal communication competence.

Conceptual processes, meskipun less frequent (26%), show balanced distribution antara analytical (69%) dan classificational (20%) representations. Analytical representations particularly effective dalam supporting thematic vocabulary learning melalui part-whole relationships yang visually organize information. Namun, limited symbolic processes (11%) indicates potential area untuk development dalam representing abstract cultural concepts yang important untuk intercultural communicative competence.

Visual interaction patterns reveal preference for offering images (84%) that position students as objective observers. While this supports clarity and focus on content, excessive distancing can limit emotional engagement, as warned by Painter, Martin & Unsworth. Strategic use demand gazes (16%) create moments of interpersonal connection, but greater frequency can enhance student involvement and motivation.

Medium shot dominance (67%) creates comfortable social distance yang simulates conversational interaction, appropriate untuk building familiarity tanpa overwhelming intimacy. Modality findings menunjukkan sophisticated strategy dalam balancing naturalism (color saturation 100%) dengan abstraction (detail 98%, depth 91%) yang reduces cognitive load sambil maintaining recognizability, consistent dengan cognitive load theory principles (Sweller, 2011).

3. Integrasi Multimodal dan Pedagogical Effectiveness

High level multimodal integration (85% complementary relationships) exceeds findings dari previous studies seperti Zhu (2022) yang menemukan predominantly simple elaborative relationships dalam Chinese EFL textbooks. Strong integration ini indicates sophisticated multimodal design dimana both modes contribute equally untuk meaning construction, creating redundancy yang pedagogically beneficial untuk reinforcing comprehension.

Dominasi elaboration relationships (47%) provides essential scaffolding untuk beginner learners melalui meaning reinforcement across modes. Enhancement relationships (30%) add contextual richness yang helps students connect language use dengan real-world situations dan cultural practices. Combination ini aligns dengan Vygotskian principles zone proximal development, dimana multimodal support helps students access complex meanings yang initially beyond their independent capability.

Locution projection (13%) melalui speech bubbles creates authentic language use representation yang important untuk developing pragmatic competence. However, absence idea projection indicates missed opportunity untuk representing internal mental processes yang could deepen students' understanding tentang thoughts, feelings, dan perspectives dalam communication.

4. Implications for Education Practices and Policies

4.1 Implications of Pedagogical Practice

The findings of the dominance of material processes and narrative visuals show alignment with the communicative language teaching approach that emphasizes language use in meaningful contexts. Teachers can leverage this strength by integrating more task-based activities that utilize visual representations for scaffolding language production. Professional development programs need to incorporate multimodal literacy training to help teachers optimize image-text relationships in classroom instruction.

High multimodal integration demands teachers develop competence dalam interpreting dan utilizing visual-verbal connections effectively. Ini requires shift dari traditional text-focused teaching toward more holistic multimodal approach yang recognizes equal importance visual dan verbal modes dalam meaning construction.

4.2 Implications of Curriculum Policy

Strong ideational metafunction representation dengan concrete experience focus supports Kurikulum Merdeka implementation yang emphasizes contextual learning. Policy guidelines untuk textbook development perlu maintain multimodal integration approach sambil encouraging greater variation dalam cultural representation dan abstract concept visualization.

Limited symbolic processes dan idea projection indicate need untuk more specific guidelines regarding abstract concept representation dan higher-order thinking skill development dalam textbooks. Ini important untuk preparing students untuk linguistic dan cognitive complexity di higher levels.

4.3 Implications of Material Development

Research findings suggest need untuk increased variation dalam visual interaction patterns, particularly greater use demand gazes dan close shots untuk enhancing emotional engagement. Material developers perlu create balance antara beginner-appropriate clarity dengan sufficient challenge untuk cognitive growth.

Unexplored audio-visual integration indicates important development area untuk digital age. Future material development perlu consider more comprehensive multimodal integration termasuk interactive elements dan multimedia resources yang support 21st-century learning competencies.

Overall, "English for Nusantara" demonstrates thoughtful multimodal design yang pedagogically appropriate untuk beginner learners. However, continuous improvement dalam cultural diversity representation, emotional engagement, dan higher-order thinking skill development akan enhance learning effectiveness dan prepare students untuk broader communicative competence dalam global contexts.

CONCLUSION

This study explores how verbal and visual elements in the English for Nusantara coursebook for SMP/MTs Grade VII collaboratively construct meaning, utilizing frameworks such as Systemic Functional Linguistics (SFL), Kress and van Leeuwen's Visual Grammar, and Martinec and Salway's image-text relations. It highlights the significance of interrelated systems, demonstrating how verbal texts convey ideational, interpersonal, and textual meanings, while visual texts enhance comprehension through narrative and conceptual processes. The findings indicate that the interaction between text and images shapes learners'

understanding and engagement, emphasizing the importance of equal status and logico-semantic relations in multimodal texts. The implications for future research and classroom practices encourage a multimodal perspective in language education, suggesting that educators should actively integrate visual analysis into teaching strategies to enhance student comprehension and promote multimodal literacy. Recommendations for coursebook authors stress the need for well-designed multimodal content that balances verbal and visual elements, ensuring cultural relevance and clarity. Future research could further investigate learner interactions with multimodal resources to deepen understanding of their practical application in educational contexts.

REFERENCES

- Adami, E. (2023). Multimodality and the issue of culture: A social semiotic perspective onto the interculturality of communication. In *Multimodal communication in intercultural interaction* (pp. 17–40). Routledge.
- Alabi, M. (2024). Visual learning: The power of visual aids and multimedia. *Journal of Educational Technology*, 15(4), 123–135.
- Barus, F. L., Harahap, S. H., Naelofaria, S., & Simanjuntak, E. E. (2024). Discourse Analysis in Developing Multimodal E-Book Teaching Materials using Book Creators in Psycholinguistics Lessons. *Journal of Pragmatics and Discourse Research*, 4(1), 101–110.
- Behnke, Y. (2021). Usability qualities of ‘well-designed’ geography textbook visuals. *Journal of Visual Literacy*, 40(1), 15–33.
- Boehm, K. M., Khosravi, P., Vanguri, R., Gao, J., & Shah, S. P. (2022). Harnessing multimodal data integration to advance precision oncology. *Nature Reviews Cancer*, 22(2), 114–126.
- Behnke, Y. (2021). Usability qualities of ‘well-designed’ geography coursebook visuals. *Journal of Visual Literacy*. <https://doi.org/10.1080/1051144X.2021.1902042>
- Cheng, L. (2021). A Multimodal Discourse Analysis on the Reports of China's First Gold: A Comparative Study of BBC and VOA. *2021 3rd International Conference on Economic Management and Cultural Industry (ICEMCI 2021)*, 2081–2084.
- Drajati, N. A., Ilmi, M., Rakerda, H., Mahmudah, K., Mardiana, R., & Supraptiningsih, N. (2025). Designing E-Multimodal Books for EFL Teaching Materials: A Call for Teacher Changes and Development. *Designing Learning with Multimodality in English Medium of Education (EME) Classrooms Across Asia*, 217.
- Dressman, M., & Sadler, R. W. (Eds.). (2020). *The handbook of informal language learning*. John Wiley & Sons Ltd.
- Elmiana, D. S., & Shen, S. (2025). Let Pictures Explain the World: Text–Image Narration in Picturebooks. *Journal of Research in Childhood Education*, 39(3), 458–471.
- Elmiana, D. S. (2019). Pedagogical representation of visual images in EFL coursebooks: A multimodal perspective. *Pedagogy, Culture & Society*, 27(4), 613–628. <https://doi.org/10.1080/14681366.2019.1569550>
- Fitria, T. N., & Simbolon, N. E. (2024). English language teaching (ELT) for students with visual-spatial learning style. *JELLE: Journal Of English Literature, Linguistic, and Education*, 5(02).
- Fitriana, W., & Wirza, Y. (2021). An analysis of multimodal text in EFL coursebook of secondary school in Indonesia in assisting students' text understanding. *Advances in Social Science, Education and Humanities Research*, 546, 627–636. <https://doi.org/10.2991/assehr.k.210427.120>

- Ghai, A., & Tandon, U. (2022). Analyzing impact of aesthetic visual design on usability of e-learning: An emerging economy perspective. *Higher Learning Research Communications*, 12(2), 1.
- Haiyan, L. (2018). Image-text relations in junior high school EFL coursebooks in China: A mixed-methods study. *Journal of Language Teaching and Research*, 9(6), 1177–1185. <https://doi.org/10.17507/jltr.0906.07>
- Hasyim, M., & Arafah, B. (2023). Semiotic multimodality communication in the age of new media. *Studies in Media and Communication*, 11(1), 96–103.
- Holtlien, A. C. (2025). *The impact of Multimodal texts on language proficiency and engagement in early English education*. Inland Norway University.
- Hoque, M. (2025). *Multimodal literacy in higher education: university students' perspective on digital resources in English language learning*. BRAC University.
- Ilonga, E., & Mapunda, G. (2022). Complementarity of communicative modes on meaning making in Tanzania's digital telecom marketing: A social semiotic multimodal perspective. *Southern African Linguistics and Applied Language Studies*, 40(1), 87–99.
- Jana, S. (2023). Interactive E-Books And Multimedia Learning: Enhancing Engagement And Retention In Secondary English. *Journal of Namibian Studies*, 33.
- Lim, F. V., & Tan-Chia, L. (2022). *Designing learning for multimodal literacy: Teaching viewing and representing*. Routledge.
- Lestari, A. R. (2022). *Multimodal analysis: Pedagogical representation of visual images in EFL coursebooks* [Master's thesis, Universitas Pendidikan Indonesia].
- Mohammed, R. J. (2024). Visual Aids As a Tool to Stimulate Creative Thinking in The Language Teaching Process. *JOURNAL OF LANGUAGE STUDIES*, 8(9), 278–289.
- Pawłowski, M., Wróblewska, A., & Sysko-Romańczuk, S. (2023). Effective techniques for multimodal data fusion: A comparative analysis. *Sensors*, 23(5), 2381.
- Permata, N. A., Lubis, M., Perdana, I., Kurniawan, A., & Sembiring, A. (2024). Enhancing educational games: Addressing design challenges and improving effectiveness. *2024 12th International Conference on Cyber and IT Service Management (CITSM)*, 1–5.
- Qi, J. (2024). Analyzing image meanings in Chinese EFL coursebooks: A multimodal perspective. *Theory and Practice in Language Studies*, 14(8), 2498–2509. <https://doi.org/10.17507/tpls.1408.23>
- Renandya, W. A. (2022). Learning from multimodal texts. *Willy's ELT Corner*. Retrieved June 2024, from <https://willyrenandya.com/learning-from-multimodal-texts/>
- Rohi, M. P., & Nurhayati, L. (2024). Multimodal learning strategies in secondary EFL education: Insights from teachers. *Voices of English Language Education Society*, 8(2).
- Rosi, I. K., & Ningrum, Y. S. (2019). Textual meaning of English coursebook instruction (A case of Buku Bahasa Inggris Kelas X by Indonesian Ministry of Education and Culture). *ELT Forum: Journal of English Language Teaching*, 8(2), 184–191. <https://doi.org/10.15294/elt.v8i2.32956>
- Sandler, W. (2022). Redefining multimodality. *Frontiers in Communication*, 6, 758993.
- Stewart, O. G., González, J., & Midgett, E. (2025). A framework for critically analyzing multimodal texts for English Language Learners through intermodal complementarity. *Qualitative Research Journal*.
- Suman, C. (2023). The effective integration of multimedia in foreign language teaching. *A Peer Review Research Journal of Language and Human Development*, 1(1), 83–93.
- Tafrijayah, K., & Andriani, M. (2025). Multimodal Discourse Analysis Of An English Language Learning Textbook In Indonesia. *Lire Journal (Journal of Linguistics and Literature)*, 9(1), 61–79.
- Vargas, P. N., & Trevisan, D. G. (2025). Seeing What We Can't See in the Non-textual Representation from the Educational Context. *IFIP Conference on Human-Computer*

Interaction, 333–355.

Weninger, C. (2021). Multimodality in critical language textbook analysis. *Language, Culture and Curriculum*, 34(2), 133–146.

Wong, J. T., & Hughes, B. S. (2023). Leveraging learning experience design: digital media approaches to influence motivational traits that support student learning behaviors in undergraduate online courses. *Journal of Computing in Higher Education*, 35(3), 595–632.