



P-ISSN 2355-2794
E-ISSN 2461-0275

Emerging Digital Factual Storytelling in English Language Learning: Investigating Multimodal Affordances

Wawan Gunawan^{1*}
Silmy Arizatul Humaira²
Tri Indri Hardini¹
Mahardhika Zifana¹

¹Department of English Language, Faculty of Language and Literature Education, Universitas Pendidikan Indonesia, Bandung 40154, INDONESIA

²Department of English Language, Faculty of Teacher Training and Education, Universitas Muhammadiyah Malang, Malang 65144, INDONESIA

Abstract

Attention has been given to multimodal texts to investigate their potential meaning affordances that facilitate learning and raise awareness of ideological meanings. However, how learners learn to make meaning by integrating intermodal relations involving language and visual images, especially in the context of learning English as a foreign language (EFL), has not been researched much. This study investigated the emerging process of meaning-making in digital factual storytelling practices in a senior high school in Indonesia. Fifty-six students participated in this study and each produced a digital storytelling video (DST). The DST videos were analyzed using an intersemiotic complementary framework to reveal the occurrences and typical intermodal meaning-making practices, as shown in the student digital storytelling (DST) videos on autobiography. The analysis focused on describing the emerging modes of making experiential meanings of the intermodal verbal-visual relations. It was found that the process of meaning-making was dominantly constructed in an exposition manner involving the verbs within clauses for identifying and describing. The emerging meanings resulting from the multimodal affordances allow storytellers to track with the potential meanings projected by the images. The selection of digital images helps students explore words when telling a story, shifting the conventional genre of autobiography. This study indicates the need to emphasize the purpose of text and to afford multimodal features to support the achievement of communication purposes.

* Corresponding author, email: wagoen@upi.edu

Citation in APA style: Gunawan, W., Humaira, S. A., Hardini, T. I., & Zifana, M. (2025). Emerging digital factual storytelling in English language learning: Investigating multimodal affordances. *Studies in English Language and Education*, 12(1), 312-327.

Received October 1, 2024; Revised November 15, 2024; Accepted December 21, 2024; Published Online January 31, 2025

<https://doi.org/10.24815/siele.v12i1.41596>

Copyright © 2025 by Authors, published by *Studies in English Language and Education*. This is an open-access article distributed under the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>)

Keywords: Digital storytelling, experiential meaning, intermodal text, intersemiotic meaning

1. INTRODUCTION

In the current digital era, meaning-making can be conveyed in many ways, more prominently through digital products permeated into more than one single mode, such as speech, writing, image, gesture, and sound (Hull & Nelson, 2005; Mestre-Mestre, 2015). In the current digital landscape, meaning-making processes have become increasingly complex, exceeding traditional modes of communication. The digital landscape is evolving and offers diverse opportunities for language learning. In the context of English learning, where English is a foreign language, the digital landscape presents intriguing possibilities to enhance learning, engagement, and productivity. The integration of visual, aural, and textual elements in the production of expressions has been appreciated as a way of learning about the use of authentic materials (Wong, 2019). When learning to speak English as a foreign language, such as in middle school contexts, learners use any possible means to accelerate the learning process. For instance, when learning to speak English through storytelling, learners create digital stories, providing them with meaningful experiences and enhancing their participation throughout the learning process (Wahyuni et al., 2018). The incorporation of multimodal elements in productive skill classes, such as speaking and writing, not only empowers students but also enables them to use language in more powerful ways (Shin & Cimasko, 2008). As such, the use of multimodal resources in written and spoken communication has become an urgent requirement in the educational curriculum. However, the use of multimodal sources to enhance the quality of spoken and written communication is not without challenges, as traditional academic norms may dominantly influence students' way of composing and expressing ideas (Alghamdi & Reed, 2019; Mills et al., 2020; Shin & Cimasko, 2008). The construction of meaning with multi-semiotic sources requires skills to operate a dynamic interplay among multiple systems: verbal and visual systems.

Researchers have highlighted the significance of multimodal composition and expressions in English as a foreign language context, emphasizing more empowering practices of English learning (Jiang, 2017; Kang, 2021; Salsabila et al., 2023; Wang, 2022). The significance of deploying multimodal sources in compositions and expressions of the English language in an EFL context allows learners to express ideas more dynamically involving diverse sources (Wang, 2022), embody movements and expressions of language (Salsabila et al., 2023), help EFL learners link print-based literacy and multimodal affordances in learning (Jiang, 2017), engage in creative processes of producing spoken and written texts across various genres (Kang, 2021). For instance, multimodal use fosters students' English learning in Taiwan with a deeper understanding of the content area that supports their composition and expression quality (Lai, 2024). It also demonstrates the significance of integrating multimodal sources as references for students to develop ideas in composing and expressing ideas (W. Yang et al., 2024) as a stimulus for more interpersonal interaction when composing and expressing ideas, as they entail attitudinal meanings (Unsworth & Mills, 2020). In a more specific context at the secondary school level, multimodal use helped the emergence of bilingual education, as it scaffolds learners' use of key ideas and expressions in communication (Smith et al., 2021). The significance of multimodal use in teaching highlights the fact that multimodality can shift the nature of literacy and writing, creating a more interactive and engaging learning environment (Alghamdi & Reed, 2019).

Other studies have investigated thinking processes (Jiang, 2018), representational meanings, meaning potentials, ideological representation (Jiang & Gao, 2020), and perspectives on using multimodal components in learning (Yeh, 2018). These studies have amplified the significant contributions of multimodal integration to EFL learning. The significance of multimodal integration into learning has indicated multimodal composing affordances in creating meanings to meet the demand for investigating the extent to which multimodal learning resources have contributed to learning quality and progression. However, the significance of learning by integrating multimodal sources and affordances overshadows emerging awareness and practices

of using multimodal for foreign language learning. In learning English as a Foreign Language (EFL), there is a paucity of research on how EFL learners make meanings of an ensemble of images as a design choice to help express ideas (Jiang & Gao, 2020; Olugbara et al., 2022; Unsworth & Herrington, 2022). Making sense of an ensemble of images as a design choice to help express ideas can be difficult for EFL (English as a Foreign Language) learners (Y. Kim & Kang, 2020). However, some studies have suggested a few strategies to optimize an ensemble of images for facilitating EFL learning, for example, identifying key elements or prominent recurring themes, colors, shapes, patterns, and other visual cues (D. Kim & Li, 2021; Lee et al., 2021).

The process of intermodal meaning-making is challenging because it offers semiotic opportunities in text production that are situated at the intersection of textual and social meaning (Kress, 2003). The social and textual layers of multimodal texts involving images, for example, shape speakers' and writers' ways of configuring meanings. Departing from this assumption, this study explores how meaning-making unfolded in students' multimodal language activities in relating past experiences. More specifically, this study identified students' experiential meanings realized through digital storytelling (henceforth: DST) as a digital multimodal artifact created by high school students when learning English as a foreign language. The concept of experiential meanings in this study refers to a topic being elaborated by verbs and referring to prior experiences and knowledge to achieve the purpose of communication (Gunawan, 2018). In this study, DST refers to personal stories of past events, in the genre of autobiography, by combining images in chronological order. The study of meaning-making in intermodal relations was guided by the following research questions:

1. What visual processes are involved in telling a factual story?
2. How do images afford intermodal relations when telling a factual story?

2. LITERATURE REVIEW

2.1 Experiential Meaning Resources in Digital Factual Storytelling

Experiential metafunctions are commonly discussed in Halliday's Systemic Functional Linguistics (SFL) theory (Halliday & Matthiessen, 2004). It is included as one of the three metafunctions referring to functional components that are always realized in making meaning with language and other semiotic systems, including images. The experiential metafunction is a meaning resource for expressing experiences, and constructing logical connections of meanings. The other two metafunctions are interpersonal and textual. These two metafunctions serve as meaning resources for respectively expressing meanings with regard to the audience, and with regard to modes of communication. The interpersonal metafunction involves interacting with other people. Meanwhile, textual metafunction deals with how the language is organized and managed (Halliday & Matthiessen, 2013).

Experiential metafunction is expressed in the transitivity system comprising the participants, processes which include material processes (verbs to express the acts of doing an action), behavioral processes (verbs to express physical or psychological aspects of a conscious being), mental processes (verbs to express the state of mind hearts, thoughts, feelings, and perceptions), verbal processes (verbs to refers to acts that involve talking, telling and saying), existential processes (verbs to express an existent), relational processes (verbs to express characteristics of the participants) and circumstances (Derewianka & Jones, 2016). The experiential metafunction is a resource for speakers and writers in creating meaning, including in storytelling.

In storytelling, meaning could be identified with reference to the genre conventions of either narrative or recount, in which common subjects and verbs usually unfold, as in other genres. The specific characteristic of storytelling in a digital manner is the integration of images and other semiotic systems, which transform into more multimodal types and demonstrate hybrid literacy skills in all contexts of learning language. This practice has now reached the context of learning

English as a second and foreign language (Towndrow & Pereira, 2018). Some practices are similar to a short film comprising various images, video, and music, including the voice recording of the narration (Bull & Kajder, 2005). The source for meaning-making is more complex and involves multi-semiotic resources in which the storytellers are faced with more ample choices of experiential resources for meaning-making, such as still pictures, moving pictures, personal experiences and other information to inform or retell (Robin, 2008). Therefore, Robin (2008) claimed that DST is a combination of various literacy skills to address digital literacy, global literacy, technology literacy, visual literacy, and information literacy (Hull & Nelson, 2005).

2.2 Multimodal Resources

Multimodal sources refer to diverse communication formats that use multiple semiotic systems, such as text, images, video, audio, and interactive elements. These sources can be found in academic publications, conference presentations, websites, social media posts, and other digital and non-digital media. The increasing use of multimodal sources has been driven by the increasing availability and accessibility of digital technologies, making the integration of the sources into learning seamless (Morell, 2015; Pan & Zhang, 2020). The integration of multimodal sources into learning has been defining the characteristics of the digital age, challenging the traditional notion of language as a monomodal tool for communication (Ravelli et al., 2018). Multimodal communication researchers examine not only the language used, but also the various modes that work together to convey a message, offering a more comprehensive understanding of communicative processes (Norris, 2012). These modes can include, but are not limited to, language, image, layout, color, music, and gesture, and they are often employed simultaneously to create cohesive and meaningful messages (Pan & Zhang, 2020).

From a social semiotic perspective, multimodality can be understood as the relationship of multiple modes of communication, where each mode contributes to the overall meaning-making process. Multimodal researchers have examined how semiotic resources are used, perceived, and interpreted in different communicative contexts with the goal of revealing the inherent complexity and richness of human expression (Norris, 2012; Ravelli et al., 2018). Multimodal complexity has been explored through visual grammar, using a social semiotic perspective to interpret intersemiotic meanings.

2.2.1 Visual grammar

Kress and van Leeuwen (2020) developed visual grammar based on Halliday's linguistics theory and Barthes' literary and semiotic theory. Images are also regarded as a means of communication because they construct a particular meaning constituting a Process, Participants, and Circumstances (Martinec & Salway, 2005). Martinec and Salway (2005) claimed that visual images are also structured by a certain grammar that can be observed from embedded elements such as color, perspective, framing, and composition. In this case, the language structure is expressed through the choices of word classes and clause structures, while the visual structure is related to the choice of color and compositional structures. In expressions of language, grammar allows people to share experiences by involving participants, verbs to describe the participants in terms of doing actions, sensing phenomena, defining or being defined, and showing their existence (Halliday & Matthiessen, 2013). Thus, the concept of experiential meanings from the systemic functional linguistic perspective is applicable not only to words but also to images. Theorists of visual grammar have rigorously developed a framework for understanding how images as semiotic systems function in making meaning (Kress, 2000).

An image has been conceptualized to possess a similar structure as a language in which three metafunctions: experiential, interpersonal, and textual, organize it (Kress & van Leeuwen, 2020). Like language, the three metafunctions in visual grammar consist of representational meaning, interactive meaning, and compositional meaning, which resemble the roles of the three metafunctions in language (J. Yang & Zhang, 2014). The interpersonal metafunction refers to the interaction between the creator, viewers, and objects in the image. When describing images as

visual resources, images represent experiences. In this case, the objects in the images refer to the represented participants and are related to other objects. The represented participants refer to any object that can be seen in an image, including living or inanimate objects. Meanwhile, the viewers who see the objects are called interactive participants. It is realized from the gaze, frame size, shot, and angle. These elements create social distances between creators and viewers. The textual metafunction refers to how the image is organized and presented. Similar to the language structure, the structure of an image also influences the meaning as a whole. For example, clause transitivity as conceptualized in SFL is associated with Kress and van Leeuwen's visual transitivity (Bateman, 2014).

2.2.2 Intersemiotic meaning

Text–image relations work in a way that verbal text extends the meanings of an image or vice versa, or verbal text elaborates the meaning of an image or vice versa (Kress & van Leeuwen 2020). Kress and van Leeuwen (2020) emphasized more dynamic relations between verbal texts and images, signifying that visual image, verbal text, and other semiotic modes are related to one another but do not depend on one another. Chan (2011) developed Kress and van Leeuwen's (2020) concept of text-image relations. One relationship between meaning of visual images and verbal texts is conceptualized in logic semantic meaning relations to communicate a topic, idea or field (see Figure 1) (Chan, 2011).

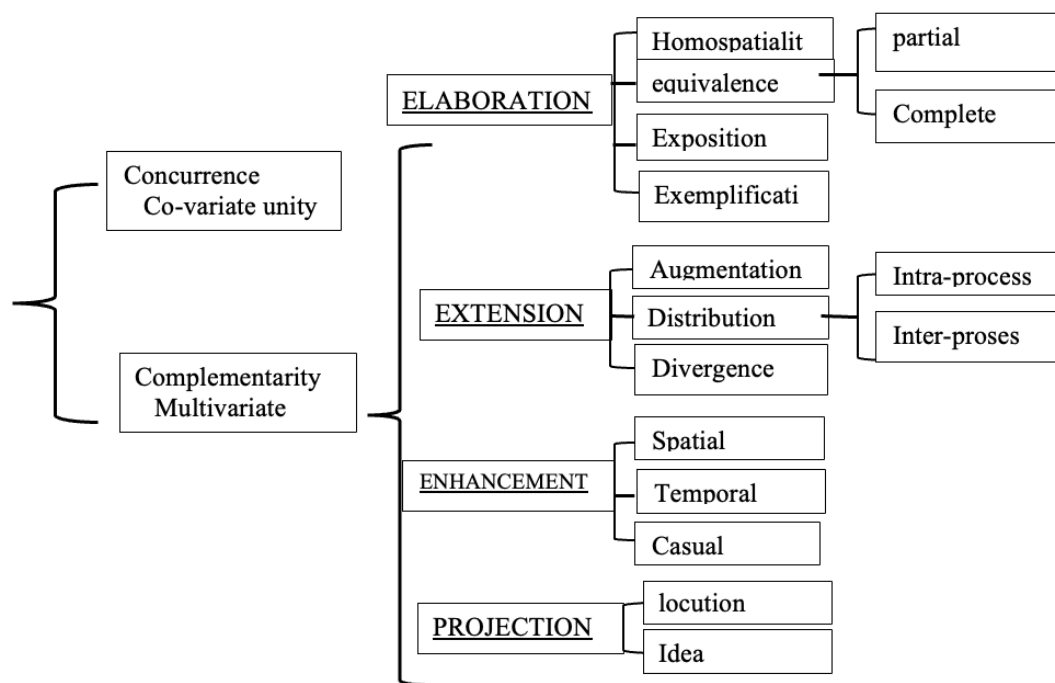


Figure 1. Intermodal relations (Chan, 2011, p. 15)

Chan (2011) developed the intersemiotic relations of verbal texts and images by dividing them into two major types: concurrence and complimentary. Concurrence is related to the equal relations between language and image when the configurations of participants, processes, and phenomena are equally structured. In discussing this concept, Unsworth and Herrington (2022) described that concurrence occurs when the experiential meaning between images and texts is equal. It is categorized into four classifications: homospatiality, equivalence, exposition, and exemplification. The first term refers to the most salient category of concurrence, indicating that language and image occur together in a similar form as a bounded entity. Equivalence refers to an image that has a function to clarify or explain the text in detail. Then, exposition means that the image represents a similar meaning to the text in an alternative version. Meanwhile,

exemplification is distinguished into two ways related to whether the image instantiates the text or vice versa (see [Chan, 2011](#); [Lim, 2004](#)).

The other big category is experiential complementary, which refers to joint intermodal interaction, in which the representation of meaning between language and image seems different, but they complement one another. This category includes extension, enhancement, and projection. The extension means that the text extends its meaning by using the image and does not provide examples of the opposite situation. This category includes augmentation (adding new meanings), distribution (constructing sequences of events or activities), and divergences (adding different aspects between image and the text). Enhancement is subdivided into manner, condition, spatial, temporal, and causal categories. According to [Martinec and Salway \(2005\)](#), enhancement deals with time, place, and reason or purpose. The other category is projection, which consists of a 'locution' that refers to the words, ideas, or thoughts being quoted or reported ([Chan, 2011](#)). The projection includes quotations or report of speech, such as those appearing in the speech bubbles of comic cartoons. According to [Coffin \(2006\)](#), projection is described as a locution or verbal projection if it is related to speech, but as an idea or mental projection if it refers to thoughts and ideas.

2.3 Autobiography Recounts for Factual Information Telling

An autobiography is a type of recount text that records the history of a specific participant ([Derewianka & Jones, 2016](#)). Similar to the general purpose of recount text, the social purpose of an autobiographical recount is to entertain and inform readers ([Coffin, 2006](#)). [Coffin \(2006\)](#) also noted that the story is not always told by famous figures like politicians or celebrities, but sometimes it is written by an unknown person and contains sequences of events written in chronological order. As it is categorized into a recount genre, it consists of orientation, events, and reorientation ([Knapp & Watkins, 2005](#)).

According to [Coffin \(2006\)](#), orientation involves the introduction of a person, time and space. Additionally, [Wood and Stubbs \(2000\)](#) explained that the function of orientation is to provide background information that is necessary for readers to understand a story. Orientation may consist of information such as what happened, who was involved, what was involved when the events occurred, where they occurred, and why they occurred. Meanwhile, [Wood and Stubbs \(2000\)](#) and [Coffin \(2006\)](#) pointed out that the record of events reveals the series of past events. The events are sequenced in chronological order. The last component is re-orientation, which composed of additional comments or expressions of attitude ([Coffin, 2006](#)). [Wood and Stubbs \(2000\)](#) confirmed that this re-orientation refers to the summary of the story, such as evaluating the importance of the topic, giving comments, and expressing opinions.

Additionally, [Wood and Stubbs \(2000\)](#) stated that recount text may consist of ideas for speculating future events. Regarding this view, although simple past tense is mostly used in recount text, it is possible to insert simple present tense to indicate immediacy and future tense to conclude or predict what might happen in the future, especially in autobiographical recounts. Despite being presented in a multimodal manner, storytelling cannot effectively achieve the purpose of communication unless it follows the rhetorical convention of telling.

3. METHODS

3.1 Design of Intermodal Analysis

This study identified patterns of meaning constructed through verbal and visual images in students' digital storytelling (DST) videos. The design of this study followed the concept of Systemic Functional Linguistics (SFL) in interpreting experiential meanings by specifically following the transitivity system developed by [Halliday and Matthiessen \(2004\)](#) and [Chan's \(2011\)](#) concept of intermodal relation (see Figure 1). To support the analysis, this study used the concept of an intermodal SFL relation to interpret meanings from visual elements descriptively.

3.2 Participants

The participants of this study were third-grade high school students in Malang, Indonesia. A female teacher taught the students how to create digital storytelling (DST) videos. The teacher is a single participant, so the necessary data are available for this study. The teacher had had years of experience teaching DST in several schools, attended DST training in America, and written some journals, a thesis, and books related to DST. The teacher prepared and designed instructions for teaching storytelling through images. After a semester of teaching, the students were assigned to produce DST videos. There were fifty-six students submitting their works on DST through an online platform called Edmodo. From the collected videos, three videos were selected purposively, especially to enable a deep description of the meaning-making process. The teacher selected one high-quality, one moderate-quality, and one low-quality DST. We selected three videos to make our analysis manageable. Manageability in qualitative data analysis is a complex element that should be considered in a qualitative study (Hull & Nelson, 2005).

3.3 Data Analysis

The analysis focused on visual image affordances as emerging practices in using digital technology to learn English as a foreign Language. The focus on images was driven by the limited resources of technology, which allowed the students to compile and arrange static images into a story. To align with the data characteristics, the analysis employed two main procedures.

First, the procedural analysis followed Chan's (2011) intermodal relations to reveal the number of frequencies of types of meanings projected by images' choices. The number of occurrences of meanings projected in the images was calculated to serve as complementary bases for describing the students' DSTs in general. As a complementary work for such data processing developments, a simple calculation provides the meaning of this investigation for descriptive purposes. The downloaded students' DST videos from Edmodo were screenshots to make some images static for the analysis. Of the fifty-six DST videos created by the students, three videos were selected for qualitative descriptive analysis. The three DST videos were selected based on their DST scores. DST 1 was created by AM (i.e., one of the participants of this study, and they are coded), representing the highest quality DST. DST 2 was created by DK, representing the moderate-quality DST. DST 3 was created by HS representing the lowest quality DST. All student-created intermodal products were reviewed to assess visual and verbal elements in factual storytelling.

Second, the verbal data were transcribed and placed next to the captured images. The analytical processes of the verbal and visual data focused on seeking the relation between the presented images and the verbal spoken languages. The analysis applied a model of image-text relation in presenting experiential meanings, that is, what is presented and how the presented topic is attributed by adapting Chan's (2011) points projected by experiential meanings. Several stages were taken to reveal the intermodal relations. First, the videos were watched several times to capture the overall components of the videos. Second, images from the DST videos were identified and divided into two parts: verbal (narration) and nonverbal (images) data. Third, the verbal data were transcribed into words. Then, the text was divided into several clauses.

4. RESULTS

4.1 Visual Processes in Factual Storytelling

The analysis indicated that the narrative notions reflected in the students' images included action, reaction, speech and mental, classificational, analytical, and symbolic notions. The most frequent notion the students used in their factual storytelling was symbolic attributes (see Table 1). Referring to Kress and van Leeuwen (2006), the analysis found that the students dominantly employed symbolic attributes presenting more than one participant in the selected images. Some

images consist of a single represented participant, which is called symbolic suggestive (Kress & van Leeuwen, 2006). Symbolic attributes were chosen to complement verbal communication when telling stories. The images serve as a backdrop of the context of the students' factual storytelling. The selected images also lead the topics of factual stories. In the students' factual storytelling, the images were structured in a way that flowed the language of the story. The narrative flow was assisted by orchestrating the images.

Table 1. Distribution of visual processes.

No.	Process/Text	Frequency	Ideational meaning indicator
1	ACTION	163	Vector
2	REACTION	53	The two-vector interaction
3	SPEECH AND MENTAL	13	Locution and thoughts
5	CLASSIFICATIONAL	239	Classification of items in images
6	ANALYTICAL	159	Explanation of the process
7	SYMBOLIC ATTRIBUTIVE	419	Identifying characters
8	SYMBOLIC SUGGESTIVE	264	Modulation of characters

As shown in Table 1, the story was primarily conveyed through symbolic attributes to express ideas and identify the character represented in the visual images accompanying the factual storytelling. The symbolic attribute played a dominant role in using images to illustrate the connection between the participant mentioned in the story and the assigned attributes. An example of a symbolic attribute is presented in Figure 1. Hence, for instance, in Figure 2, a young boy was depicted in a specific location holding binoculars. The represented actors included the boy and the binoculars. Based on the presence of the binoculars and the setting, it was evident that he was taking a picture at a tourist spot. This image contained no action, allowing the storyteller to describe the scene. The image was then further elaborated. It primarily served to present the subject to viewers without additional contextual information beyond what was explicitly depicted.



Figure 2. Example of a symbolic attribute in storytelling in Text 2.

Another example of a symbolic attribute was presented to elaborate on what the storyteller did and achieved. The interaction of a hand with a trophy symbolized an event and indicated possession, as shown in Figure 3.

Verbal realization	Image relations
In several times, I've won some test times like fourth winner in four pillar MPR Jawa Timur and coming finalist or semi-finalist in a paper contest(s). I have won some test times.	

Figure 3. Example of a symbolic attribute in storytelling in Text 1.

Figure 3 depicts a hand holding a trophy. The image was chosen to convey the idea of winning a competition. The storyteller narrated that she had won the paper contest as a semi-

finalist. There was no justification for selecting this trophy other than symbolizing the category of the competition and the achievement she had obtained. The trophy symbolized her success.

Symbolic attributives were not only related to certain material entities but could also be represented in the form of text written in an image. As exemplified in Figure 4, the word ‘LEADER’ symbolized what the storyteller was trying to convey. The event she related to was chosen based on her story of gaining leadership experiences. Simply put, she was a leader. The topic of leadership guided her to continue telling factual stories.

The image in Figure 4 provided symbolic meaning, particularly through the red color and capital letters embellishing the word “LEADER.” The color and capitalization symbolized certain meanings associated with the word. Red and yellow represented vibrant colors, close to life, as [Kress \(2000\)](#) suggested. The image served as a trigger for building ideas in the story, which aimed to express keywords extending the meaning of the unforgettable event, her life achievement, and her future orientation. Colors played a significant role in the overall meaning of the story discourse. In this case, red symbolized bravery, and the capital letters represented a symbolic meaning related to the characteristics of a leader, firmness and clarity of character.

Verbal realization	Image relations
“... and unforgettable event in my life was joining a boarding organization. That I was pointed as a leader. Achievement in my life is that when I (went to) Mecca with my family. In the next ten years, I want to be a success manager in a company. The favorite quote in my life is that “Don’t kill the dream, execute it”.	

Figure 4. Example of symbolic attribute in Text 3.

An example of a symbolic suggestive was the way in which the storyteller included an image of a bird, as shown in Figure 5. This image brought symbolic meaning to the storyteller’s awareness of a bird. It was a single eagle representing herself, as illustrated in the verbal language.

Verbal realization	Image relations
... I was born (on) 11 22 2000 in Malang. I would compare myself to a bird. Because I can fly everywhere. I want to be as free as a bird. I feel myself is fun, active, and fast thinker. Because I like doing activities which people tend (to think) I am an open minded person, which people tend to think I am an open minded person ...	

Figure 5. Example of a symbolic attribute in Text 3.

The storyteller compared herself to a bird to suggest how free she was in her life. The image of the bird inspired the storyteller to associate all of the words in her poem with bird characters, more specifically, the characteristics of an eagle. The choice of the image served as a symbolic suggestive entity because the reason for choosing the image could be anything that fit the storyteller’s imagination associated with the topic, which, in this case, was the introduction. There was no justification for why the bird was chosen to symbolize her ideas. When selecting images, justification was not essential; rather, the meanings should be represented ([Kress & van Leeuwen, 2020](#)).

The findings of multimodal resources predominantly used by the storyteller aligned with the data depicted by [J. Yang and Zhang \(2014\)](#), which displayed a map as an example of things and the colors of the states as possessive attributes. In their work, color choices marked the ideas the storyteller elaborated and extended. The findings of this study addressed the need for

developing students' multimodal competencies and language learning engagement (García-Pinar, 2024).

4.2 Intermodal Relations through Images in Factual Storytelling

Multimodal affordances were identified in terms of how much the students explored the meanings projected by the images. Following the concepts of meaning-making in intermodal relations, the analysis revealed how logico-semantic relations were created. The interrelationship between verbal and visual representation was concurrent, involving processes categorized as dominant exposition. In general, 45 of the 50 videos were accessible. Of the accessible videos, the students showed the exposition category of logico-semantic relations (405 occurrences) as the most frequent category of meaning-making. Exposition referred to expressing meanings of an image or text through other modes in similar ways (Unsworth & Herrington, 2022). The participants relied much more on this exposition to express the primary topic (experiential meaning) by employing more elaboration in an exposition manner. As seen in Figure 1, exposition was a way of elaboration. The fact that the students' DST videos consisted of dominant exposition meant that the students were likely to integrate images and verbal text by re-expressing the corresponding meanings in two types of modes: verbal and visual modes with an equal level of generality (see Figure 6 as an example).

Verbal realization	Image relations
Besides write I can draw sketch. A memorable event in my life is when I was in primary school. In every test, I always get a highest score. And in middle school I joined an acceleration program because my IQ was 130.	

Figure 6. Example of exposition in Text 1.

An example of exposition, as shown in Figure 6, in AM's video instantiated the same level of generality between the verbal text and the image. The meaning conveyed in the verbal text was on an equal level by employing verbs of action (material processes such as 'can draw,' 'joined') and verbs of attribution (relational processes such as 'was,' 'is'). The meanings were reciprocated equally, with the word 'draw' represented by the image of a pen and paper. Similarly, in DK's video (see Figure 7), the main idea of the verbal text, which was related to 'who is the author', was conveyed through the photograph to expose what had been verbalized and visualized. The common process of verbs to convey this kind of exposition was the relational process of attributing and providing attributes to the participant, which was 'I' in this case. In this exposition mode, the employed verbs were relational verbs of identifying. The storyteller showed his own photograph to introduce himself. The meaning realized in the verbal text corresponded to the visual image.

Similarly, in HS's video in Figure 7, the verbal text re-expressed the visual image by exposing an activity, as shown in the image. The image was elaborated by employing a variety of processes (verbs), including verbs describing the storyteller's perception of the presented visual images, relational verbs describing the storyteller's providing labels to the images, and material verbs describing actions the storyteller had done in the past.

The verbal text depicted an elaboration, which was a strategy for exposing ideas as an alternative to narrative storytelling. It was a more favorable means of describing past experiences. For example, storytellers tended to provide instantiations and extend relations between images and verbal texts. In providing examples, their storytelling reflected a class-member relationship because the example was taken from the class set and did not have the same level of generality

(Chan, 2011), employing verbs of attribution (such as ‘am,’ ‘was,’ and ‘work’) to provide attributes to the represented participants.

Verbal realization	Image relations
You can call me DK. I am 18th years old. I was born on 3 August 2000 in Tuban city. I am the last child from three brother. My father work as a civil servant and my mother work as a housewife.	

Figure 7. Example of exposition in Text 2.

An explanation was also provided in the form of exemplification, realized through image-verbal text or verbal text-image instantiation. AM’s and HS’ DST videos in Figure 8 provided exemplifications. Their pictures instantiated an example of what was conveyed in the text. For instance, as identified in Figure 8, she presented the logo of UGM, a university she knew in Indonesia, to exemplify the best universities.

Verbal realization	Image relations
Hopefully, 1 day, I will join one of the best economic faculties and the best universities like UGM, UI, and ITB. Because I want to become an expert in the ministry of finance. My favorite quotes is “I am whatever the future holds want to be, because I can make many plans for my life but besides that God’s plan is always better than ours.”	

Figure 8. Example of exemplification in Text 1.

The verbal text, as seen in Figure 8, depicts an exemplification, which is another strategy for exposing ideas as an alternative way to be more descriptive than narrative. The text tends to be more descriptive, focusing on the best economic faculty, as exemplified by the logo of one of the preferred universities. This text did not discuss past experiences despite the teacher’s prompting to the students to write an autobiography. Instead, it was a more detailed description of future plans with some possibilities as indicated by the use of modalities such as ‘can’. The possibility is elaborated with verbs to express one’s mental state, such as ‘want’ and attributes directed to the author telling who he/she is. Like in the previous texts, the storyteller’s personal information was presented in a descriptive mode to elaborate on who she is rather than what she did, as commonly written in autobiography.

5. DISCUSSION

This study identified the inter-semiotic features of students’ DST videos in the process of learning to tell a factual story, an autobiography, in English as a foreign language. DST videos are digital artifacts that were analyzed to explore the topics being dominantly spoken in verbal modes and enhanced in visual images. The selection of topics and other linguistic and visual features concurrently realized meanings. The analysis of the verbal and visual features of the students’ DST videos revealed that images facilitated the students’ expression of their ideas in sharing their experiences. Language and images make meanings correspondingly. The findings of this study rely their explanation on the concepts of visual grammar (Kress & van Leeuwen, 2020; Martinec & Salway, 2005; Unsworth & Herrington, 2022). This explanation indicates that

the meaning of the integration of image and language that occurs in a single discourse is extended more than the meaning of the mode that occurs separately. This study found that by composing intermodal texts, the students learned to elaborate meanings through different multi-semiotic modes and logico-semantic relations. Deploying more relations of exposition in their DST videos was intended to clarify and elaborate the ideas represented from the visual images. Despite less significant contribution to building a factual story, the use of images to relate to verbal realizations indicates the significant attribution of visual representations in factual storytelling. The storytellers' understanding of visual elements, such as, size, color, lighting and arrangement is the basis for constructing the content of the story and promoting audience engagement (Cohn, 2020; Hoecker, 2014; Meier, 2022).

Additionally, this practice offers opportunities for students to develop their multimodal competencies. Multimodal competence encompasses the ability to communicate effectively using various modes of communication, such as visual, auditory, spatial, and linguistic modes. Multimodal activities in learning languages such as writing enhance students' linguistic skills and motivation (Kang, 2022; Arslan, 2000). The students' practices of using multimodal sources allowed them to elaborate on ideas in telling a factual story. This activity induces a habituation of learning and communicating with available sources. Meanwhile, the findings of this study support the findings of other studies (such as Jiang, 2017; Peng et al., 2017; Zhang et al., 2024), indicating the positive impact of using multimodal sources on the students' writing competence and their willingness to communicate by utilizing the available choices of verbal and visual modes of communication.

However, this study showed how images as multimodal components for multi-semiotic language practices offered students many expression choices. The images served as meaning resources from which the students could select the purpose of communication. In this study, the students showed less control of the recount genre, whose purpose was to tell or share factual information. In this genre, conventionally, communication is delivered through stages of activities. Instead, the story was more dominantly realized through relational processes (verbs to identify and define the subject), as it was almost always realized in every visual image. Meanwhile, material processes (verbs to express actions), and mental processes (verbs to express feelings and emotions) are the major language features of recounting (Derewianka & Jones, 2016). Derewianka and Jones (2016) stated, "an autobiography relates significant events from a person's own life... the writer needs to deal with long stretches of time when recounting a recent incident" (p. 130). An autobiography, according to Derewianka and Jones (2016), should show the major language resources commonly used for recounting such as by including specific, familiar participants (e.g. 'my family', 'Class 2Z', 'our school'), verbs to express actions, thoughts, and feelings, primarily in the past simple tense (e.g. 'went', 'was', 'were', 'had', 'watched') and adverb to express circumstances of time to locate actions in the past and to sequence events (e.g. 'yesterday', 'for two weeks', 'at first').

The findings of this study indicate that learners of English as a foreign language (EFL) focused too much on the elaboration of visual images without paying more attention to the purpose of the text. Thus, their autobiographies did not record a longer period of time in recounting the selected incidents in their life. Instead, the learners used exposition to clarify ideas by providing examples, paraphrasing, and summarizing the language through images. Their autobiographies tended to deploy similar major characteristics of the information report genre, which is characterized by the large number of relational verbs of defining, classifying, and exemplifying. The findings suggest that pedagogically, the lesson should begin with the learning of an autobiography model text to have the learners understand what language features to use to achieve the purpose of communication and to avoid being distracted by the salience of visual images in raising more ideas to tell. However, images or other emerging semiotic tools such as videos, graphics, and flowcharts help students develop ideas (Purwaningtyas, 2020) and process information and experiences, which could lead to improved literacy learning experiences by dynamically exposing their experiences and reflections with appropriate verbal systems (Mestres & Pellicer-Sánchez, 2019).

6. CONCLUSION

Digital technology has enabled the integration of images and videos to help students convey experiential meaning. The data analysis reveals that multimodal sources incorporate verbal elements to express ideas in various forms. These features, embedded within multimodal elements, provide meaning-making resources that students can utilize in both written and spoken communication. The potential of images as a multimodal unit allows for a more elaborate construction of storytelling. This study found that students predominantly employed an exposition strategy by describing, explaining, exemplifying, and adding new meanings to selected images in their digital videos. The inclusion of images enabled learners to articulate, describe, and explain concepts more effectively. Analysis of verbal data revealed that students used more verbs to express, define, and describe actions, an approach that diverges from conventional storytelling practices. Instead of adhering to traditional narrative conventions, students tended to provide information or attribute meaning through verbs that identify, describe, and signify. Their storytelling language was primarily structured through intermodal exposition. The process types used to express meaning, whether by describing, taking action, or sensing, play a key role in achieving communicative purposes. These process types in factual storytelling reflect register variables and contribute to the construction of meaning. The role of major linguistic features in intermodal relations should not be solely determined by the presence of still or moving images, which, while visually engaging, do not solely define meaning-making in storytelling.

This study has some limitations that can be addressed in future research. This study should have involved more multimodal resources, such as gestures, pictures, and space. This study should have included more data, such as data videos showing more multimodal storytelling interactions. Future research directions could address more multimodal elements and form pedagogical practices of exploiting the potential meaning-making from multimodal learning sources and how students' genre awareness leads them to develop their communicative competence in composing and expressing meanings.

REFERENCES

- Alghamdi, M., & Reed, M. C. (2019). The value students and instructors place on multimodal composition within academic life. *Arab World English Journal (AWEJ)*, 10(1), 283-297. <https://doi.org/10.24093/awej/vol10no1.24>
- Arslan, S. (2020). Multimodal writing to promote global competence for EFL learners. *Sakarya University Journal of Education*, 10(3), 589-608. <https://doi.org/10.19126/suje.777878>
- Bateman, J. (2014). *Text and image: A critical introduction to the visual/verbal divide*. Routledge.
- Bull, G., & Kajder, S. (2005). Digital storytelling in the language arts classroom. *Learning & Leading with Technology*, 32(4), 46-49.
- Chan, E. (2011). Integrating visual and verbal meaning in multimodal text comprehension: Towards a model of intermodal relations. In S. Dreyfus, S. Hood, & M. Stenglin (Eds.), *Semiotic margins: Meaning in multimodalities* (pp. 145-147). Bloomsbury Academic.
- Coffin, C. (2006). *Historical discourse: The language of time, cause, and evaluation*. Continuum.
- Cohn, N. (2020). Visual narrative comprehension: Universal or not? *Psychonomic Bulletin & Review*, 27(2), 266-285. <https://doi.org/10.3758/s13423-019-01670-1>
- Derewianka, B., & Jones, P. (2016). *Teaching language in context* (2nd ed.). Oxford University Press.
- García-Pinar, A. (2024). Initiating language engagement with multimodal learning tasks. *Studies in English Language and Education*, 11(3), 1316-1332. <http://dx.doi.org/10.24815/siele.v11i3.35450>
- Gunawan, W. (2018). Re-appropriation of ideational meanings through drawings: A case of 5-year-old child experience in learning to mean. *Indonesian Journal of Applied Linguistics*, 8(2), 244-253. <https://doi.org/10.17509/ijal.v8i2.13300>

- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). Arnold.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2013). *Halliday's introduction to functional grammar*. Routledge.
- Hoecker, R. (2014). Visual narrative and trauma recovery. *Narrative Inquiry*, 24(2), 259-280. <https://doi.org/10.1075/ni.24.2.05hoe>
- Hull, G. A., & Nelson, M. E. (2005). Locating the semiotic power of multimodality. *Written Communication*, 22(2), 224-261. <https://doi.org/10.1177/0741088304274170>
- Jiang, L. (2017). The affordances of digital multimodal composing for EFL learning. *ELT Journal*, 71(4), 413-422. <https://doi.org/10.1093/elt/ccw098>
- Jiang, L. (2018). Digital multimodal composing and investment change in learners' writing in English as a foreign language. *Journal of Second Language Writing*, 40, 60-72. <https://doi.org/10.1016/j.jslw.2018.03.002>
- Jiang, L., & Gao, J. (2020). Fostering EFL learners' digital empathy through multimodal composing. *RELC Journal*, 51(1), 70-85. <https://doi.org/10.1177/0033688219898565>
- Kang, J. (2021). Adolescent English language learners' digitally mediated multimodal compositions: Multimodal enactment across different genres of writing in the EFL context. *English Teaching: Practice & Critique*, 20(2), 196-209. <https://doi.org/10.1108/etpc-07-2020-0069>
- Kang, J. (2022). Developing multimodal communicative competence: Adolescent English learners' multimodal composition in an after-school programme. *Literacy*, 56(4), 355-370. <https://doi.org/10.1111/lit.12294>
- Kim, D., & Li, M. (2021). Digital storytelling: Facilitating learning and identity development. *Journal of Computers in Education*, 8, 33-61. <https://doi.org/10.1007/s40692-020-00170-9>
- Kim, Y., & Kang, S. (2020). Writing to make meaning through collaborative multimodal composing among Korean EFL learners: Writing processes, writing quality, and student perception. *Computers and Composition*, 58, 102609. <https://doi.org/10.1016/j.compcom.2020.102609>
- Knapp, P., & Watkins, M. (2005). *Genre, text, grammar*. University of New South Wales Press, Ltd.
- Kress, G. (2000). Multimodality: Challenges to thinking about language. *TESOL Quarterly*, 34(2), 337-340. <https://doi.org/10.2307/3587959>
- Kress, G. (2003). *Literacy in the new media age*. Routledge.
- Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- Kress, G., & van Leeuwen, T. (2020). *Reading images: The grammar of visual design*. Routledge.
- Lai, C. J. (2024). Enhancing multimodal output in CLIL education: The impact of VR games on fourth-grade students' English poster designs and presentations in Taiwan. *Humanities and Social Sciences Communications*, 11(1), Article number 1465. <https://doi.org/10.1057/s41599-024-03999-y>
- Lee, S. Y., Lo, Y. H. G., & Chin, T. C. (2021). Practicing multiliteracies to enhance EFL learners' meaning-making process and language development: A multimodal problem-based approach. *Computer Assisted Language Learning*, 34(1-2), 66-91. <https://doi.org/10.1080/09588221.2019.1614959>
- Lim, V. F. (2004). Developing and integrative multi-semiotic model. In K. O'Halloran (Ed.), *Multimodal discourse analysis: Systemic functional perspectives* (pp. 220-246). Continuum.
- Martinec, R., & Salway, A. (2005). A system for image-text relations in new and old media. *Visual Communication*, 4(3), 337-371. <https://doi.org/10.1177/1470357205055928>
- Meier, S. (2022). Digital storytelling: A didactic approach to multimodal coherence. *Frontiers in Communication*, 7, 906268. <https://doi.org/10.3389/fcomm.2022.906268>

- Mestre-Mestre, E. M. (2015). The construction of meaning in the second language classroom: A multimodal discourse analysis. *Procedia - Social and Behavioral Sciences*, 173, 228-233. <https://doi.org/10.1016/j.sbspro.2015.02.057>
- Mestres, E. T., & Pellicer-Sánchez, A. (2019). Young EFL learners' processing of multimodal input: Examining learners' eye movements. *System*, 80, 212-223. <https://doi.org/10.1016/j.system.2018.12.002>
- Mills, K. A., Stone, B. G., Unsworth, L., & Friend, L. (2020). Multimodal language of attitude in digital composition. *Written Communication*, 37(2), 135-166. <https://doi.org/10.1177/0741088319897978>
- Morell, T. (2015). International conference paper presentations: A multimodal analysis to determine effectiveness. *English for Specific Purposes*, 37, 137-150. <https://doi.org/10.1016/j.esp.2014.10.002>
- Norris, S. (2012). Multimodality in practice: Investigating theory-in-practice-through-methodology. In S. Norris (Ed.), *Multimodality in practice* (pp. 222-226). Routledge. <https://doi.org/10.4324/9780203801246-18>
- Olugbara, C. T., Letseka, M., & Olugbara, O. O. (2022). A systematic review of digital storytelling as an educational tool for teaching and learning in Southern Africa. In J. Olivier, A. Oojorah, & W. Udhin (Eds.), *Multimodal learning environments in Southern Africa* (pp. 165-195). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-97656-9_9
- Pan, X., & Zhang, Z. (2020). An empirical study of application of multimodal approach to teaching reading in EFL in senior high school. *International Journal of Emerging Technologies in Learning*, 15(2), 98-111. <https://doi.org/10.3991/ijet.v15i02.11267>
- Peng, J. E., Zhang, L., & Chen, Y. (2017). The mediation of multimodal affordances on willingness to communicate in the English as a foreign language classroom. *TESOL Quarterly*, 51(2), 302-331. <https://doi.org/10.1002/tesq.298>
- Purwaningtyas, T. (2020). Didactic symbol of visual images in EFL textbook: Multimodal critical discourse analysis. *Pedagogy: Journal of English Language Teaching*, 8(1), 51-63. <https://doi.org/10.32332/pedagogy.v8i1.1959>
- Ravelli, L., Adami, E., Boeriis, M., Veloso, F. O., & Wildfeuer, J. (2018). Visual communication: Mobilizing perspectives. *Visual Communication*, 17(4), 397-405. <https://doi.org/10.1177/1470357218794838>
- Robin, B. R. (2008). Digital storytelling: A powerful technology tool for the 21st-century classroom. *Theory into Practice*, 47(3), 220-228. <https://doi.org/10.1080/00405840802153916>
- Salsabila, S., Nurami, N., Oktaviana, R. T., Muljanto, S., & Kurnia, A. (2023). Seeing how multimodal sources are used in EFL classrooms. *English Education and Applied Linguistics Journal*, 6(2), 102-106. <https://doi.org/10.31980/eeal.v6i2.50>
- Shin, D. S., & Cimasko, T. (2008). Multimodal composition in a college ESL class: New tools, traditional norms. *Computers and Composition*, 25(4), 376-395. <https://doi.org/10.1016/j.compcom.2008.07.001>
- Smith, B. E., Pacheco, M. B., & Khorosheva, M. (2021). Emergent bilingual students and digital multimodal composition: A systematic review of research in secondary classrooms. *Reading Research Quarterly*, 56(1), 33-52. <https://doi.org/10.1002/rrq.298>
- Towndrow, P. A., & Pereira, A. J. (2018). Reconsidering literacy in the 21st century: Exploring the role of digital stories in teaching English to speakers of other languages. *RELC Journal*, 49(2), 179-194. <https://doi.org/10.1177/0033688218754943>
- Unsworth, L., & Herrington, M. (2022). Visualization type and frequency in final-year high school science examinations. *Research in Science Education*, 53, 707-725. <https://doi.org/10.1007/s11165-022-10094-2>
- Unsworth, L., & Mills, K. A. (2020). English language teaching of attitude and emotion in digital multimodal composition. *Journal of Second Language Writing*, 47, 100712. <https://doi.org/10.1016/j.jslw.2020.100712>

- Wahyuni, W., Sujoko, S., & Sarosa, T. (2018). Improving students' speaking skill through project-based learning (digital storytelling). *English Education*, 6(2), 161-161. <https://doi.org/10.20961/eed.v6i2.35943>
- Wang, H. (2022). International English learners' perspectives on multimodal composing and identity representation via multimodal texts. *SAGE Open*, 12(2), 21582440221103526. <https://doi.org/10.1177/21582440221103526>
- Wong, D. (2019). Digital storytelling using both keyboard and pen. In D. Wong (Ed.), *Digital humanities and new ways of teaching* (pp. 195-218). Springer. https://doi.org/10.1007/978-981-13-1277-9_11
- Wood, S., & Stubbs, K. (2000). *Targeting text*. Blake Education.
- Yang, J., & Zhang, Y. (2014). Representation meaning of multimodal discourse: A case study of English editorials in *The Economist*. *Theory and Practice in Language Studies*, 4(12), 2564-2575. <https://doi.org/10.4304/tpls.4.12.2564-2575>
- Yang, W., Chan, A., & Gagarina, N. (2024). Remote online language assessment: Eliciting discourse from children and adults. *Frontiers in Communication*, 9, 1463182. <https://doi.org/10.3389/978-2-8325-5142-4>
- Yeh, H.-C. (2018). Exploring the perceived benefits of the process of multimodal video making in developing multiliteracies. *Language Learning & Technology*, 22(2), 28-37. <https://doi.org/10.125/44642>
- Zhang, Y., Peng, J., & Zheng, Y. (2024). Teachers' perceptions of implementing digital multimodal composing in tertiary classrooms: Voices from Chinese EFL teachers. *International Journal of Applied Linguistics*, 34(4), 1265-1282. <https://doi.org/10.1111/ijal.12560>