



Castledown

 OPEN ACCESS

# Language Education & Assessment

ISSN 2209-0959

<https://www.castledown.com/journals/lea/>

*Language Education & Assessment*, 5(1), 1–17 (2022)

<https://doi.org/10.29140/lea.v5n1.685>

## Image-Text Relations and Interjections in Animated Language-Learning Materials



SHIN-YING HUANG

*Department of Foreign Languages and Literatures  
National Taiwan University, Taiwan  
syhg@ntu.edu.tw*

### Abstract

This study examined image-text relations in English-as-a-foreign-language textbooks focusing on moving-image animations that accompany verbal dialogues. It also explores how animations support verbal dialogues in the textbooks. Data sources include three major versions of junior high school textbooks locally-produced in Taiwan, each consisting of 6 books, one for each semester. Findings show that animations enter into one specific type of image-text relations with the textual dialogue, i.e., a relation of Complementarity: Augmentation, by providing information that is consistent with and additional to the dialogues in a number of ways. The animations provide the rationale for a particular part of the dialogue not otherwise apparent, they portray another participant in the dialogue in addition to the speaker of the conversation turn, and they reveal a speaker's emotional response that is not expressed verbally in the dialogue. In naturally-occurring conversations, these emotions would usually be expressed through interjections, but are mostly absent from the verbal dialogues in these textbooks. As such, the animations also serve to substitute the pragmatic markers that should have been present in the verbal text. Based on these findings, the article concludes by providing pedagogical recommendations for how animations could be used to guide language learners in exploring the possible communicative functions of appropriate interjections in the textbook dialogues in order to develop learners' pragmatic competence.

**Keywords:** image-text relations, interjections, pedagogical materials, English as a foreign language, textbooks

**Copyright:** © 2022 Shin-ying Huang. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within this paper.

## Introduction

Much effort has been made to determine the relations between image and text (e.g., Bateman, 2014; Hiippala, 2015; Liu & O'Halloran, 2009). However, most of these have been about texts other than language materials, such as picturebooks (e.g., de Oliveira *et al.*, 2018; Sipe, 1998; 2012), science textbooks (e.g., Bezemer & Kress, 2010), webpages (e.g., Djonov, 2007) and other media texts such as advertisements (Page, 2006) and news texts (Economou, 2014). While a few studies have examined visual and verbal modes in language learning materials (e.g., Chen, 2010a, 2010b; Liu & Qu, 2014; Teo & Zhu, 2018), they have focused on how the images that accompany writing serve pedagogical goals, such as the extent to which the images can contribute to learner comprehension of the written text. In other words, image-text relations in language learning materials have seldom been considered from a non-pedagogical perspective in the way that images in other texts such as picture books or news texts have. In addition, these studies have looked at images in learning materials that accompany expository or narrative passages and have seldom examined the images associated with dialogues in these materials. Finally, almost all attention has concentrated on still images and their relations to the verbal text, even though, increasingly, audiovisual materials such as animated cartoon clips of the verbal dialogues are often presented as supplementary teaching and learning materials to the printed textbooks.

Thus, this study aims to gain a better understanding of the ways that animations differ from still images in terms of how they relate to the accompanying verbal dialogues in English-as-a-foreign-language (EFL) textbooks. The research is guided by the following question: How do animations support the verbal dialogues in English-language textbooks? Specifically, rather than providing a general overview of the different types of image-text relations which exist in a particular series of EFL textbooks, such as Lai (2018) has already done, this article will highlight a particular type of animation-text relations that have particular implications for EFL conversations, i.e., animations providing interjections in which verbal dialogues do not. This study is important in that it highlights a regularly used but frequently overlooked language material, that is, animations which accompany textbooks. The article elucidates the relations between the animations and the dialogue texts, both of which have been mostly absent in prior research on image-text relations. The results of the study will also provide implications for classroom teachers on how they can further utilize the multimodality of recent language learning materials.

In the following review of the literature, I will discuss how multimodal materials relate to the dialogues (rather than expository or narrative passages) in language learning materials. I will also review the few studies that have examined how animated videos are more effective in supporting language learning than still images. What this review will also show is that there have not been studies about how animations of language learning materials relate to their verbal texts, which is the focus of this article.

## Literature Review

In the few studies that discuss image-text relations of English-language textbooks, two frameworks are commonly used. One is Kress and van Leeuwen's (1996, 2006) grammar of visual design (GVD), and the other is the system of image-texts relations proposed by Martinec and Salway (2005) and further expanded upon by Unsworth (2006, 2007).

### Grammar of Visual Design

As its name suggests, GVD is a framework that focuses on visuals, that is, the representational, interactive, and compositional meanings in visuals such as images. It is beyond the scope of this article to discuss the grammar of visual design in detail, but a brief description will be provided here. GVD is derived from systemic functional linguistics (Halliday, 1994; Halliday & Matthiessen, 2004), which

understands a text has having representational/ideational, interactive/interpersonal, and compositional/textual metafunctions. In terms of the representational meaning, that is to say, the idea represented in the image, images can be either narrative or conceptual. In terms of the interactive meaning, i.e., the relationship between the image and the viewer, issues such as contact, social distance, and modality need to be considered. The compositional meaning, that is, how the various parts of a text (an image or a multimodal passage) work together to achieve a holistic meaning, can be considered from the aspects of information value, salience, and framing. When analyzing the multimodality of textbooks from a GVD perspective, the starting point is the analysis of images, as the grammar of visual design is a framework of how visuals make sense. The results from the analysis of images are then compared to the related verbal text to determine if aspects of the same meaning can be found.

Salbego *et al.* (2015), for example, examined three images in three commercial English-language textbooks for how the images correspond to the related verbal text from the perspective of GVD's three metafunctions, although focusing mainly on the representational meaning in terms of visual-verbal relations. They identified representational correspondence between image and text in two cases. What is interesting in both examples is that each still image corresponds to two turns in the dialogue which the image accompanies. The image that shows Tom introducing Paulo to Sarah (i.e., Tom with his right palm facing up and simultaneously Sarah and Paulo shaking hands) portrays Tom's conversational turn ("Sarah, this is Paulo. He's from Brazil") as well as Sarah's turn ("Hello, Paulo. Are you on vacation?"). Similarly, in the second image discussed, Salbego *et al.* noticed that the woman's gesture corresponds to her turn ("Actually, Costa Rica isn't in South America. It's in Central America") while the man's gesture relates to his turn ("Oh, right. My geography isn't very good"). Although the two speaking turns occur sequentially one after the other, the accompanying image presents the two people's embodied interactions as simultaneously occurring. This reveals the contrast between the temporal/sequential nature of language and the spatial/simultaneous nature of image and is common to images relating to dialogues.

Jauhara *et al.* (2021) conducted a similar study using the GVD framework for the analysis of images in a locally-produced EFL textbook for seventh-grade students. They focused on content related to "greetings" and found correspondence between the verbal and the visual text. One instance given in the article shows two multimodal texts from the textbook. In both, the visual aspect presents two people in conversation. Emanating from them are speech bubbles that include all the conversational turns from the two participants. For example, the image at the top represents a three-turn greeting sequence between a student and the teacher. The student initiates the greeting (turn 1), the teacher replies (turn 2), and then the student responds (turn 3). Thus, the student has a total of two conversation turns while the teacher has one turn. In the visual, two speech bubbles emanate from the student (one for each of his conversation turn) and one speech bubble is attributed to the teacher (for his one conversation turn). Thus, the image represents two people in dialogue and the three speech bubbles are simultaneously shown in the picture to represent the three sequential turns-at-talk. Thus, the image represents the idea of two people having a conversation in general while the speech bubbles delineate all the conversation turns and the content. The representational and ideational meanings of image and text correspond, but the temporal and sequential nature of language cannot be exactly reproduced in still images.

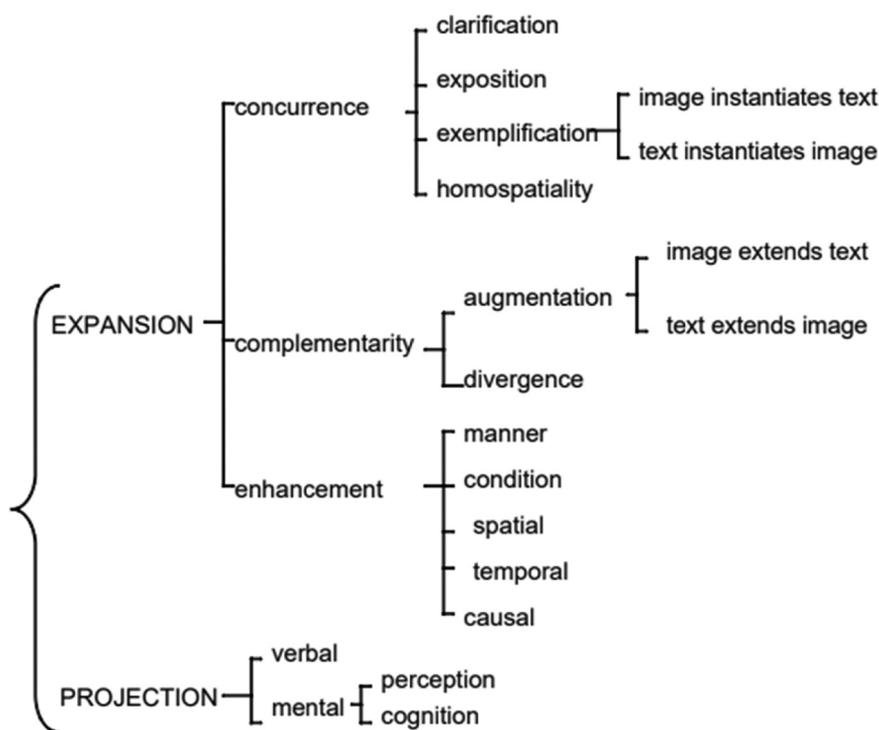
### Image-Text Relations

In Martinec and Salway's (2005) seminal study, they analyzed a large number of texts (although *not* including language learning materials) in their effort to establish a comprehensive typology of image-text relations. Their framework was examined and further expanded upon in Unsworth (2006, 2007) based on his own examination of additional texts (but still *not* including language learning materials).

It is beyond the scope of this paper to discuss the intricate details of Unsworth's (2007) framework. However, as *Expansion: Complementarity: Augmentation* is the image-text relations to be addressed in this paper, I will provide a brief explanation. Please see Figure 1 and also Unsworth (2006, 2007) for a detailed discussion.

As is shown in Figure 1, *Expansion: Complementarity* occurs either as Divergence or Augmentation, the latter of which is further classified as image extending text or text extending image. *Complementarity: Divergence* describes when "the ideational content of text and image are at variance" (p. 1176). This occurs most often in postmodern picture books, when the text and its related picture may be expressing two opposing things, an example of which can be found in John Burningham's *Time to Get Out of the Bath, Shirley*, wherein one might find a picture showing a mother giving her child a bath, while the text narrates the child's adventures across the seas, that is to say, the bath from the child's imaginative perspective. More relevant to this study is the relation of *Complementarity: Augmentation*, which occurs when "each of the modes provides meanings additional to and consistent with those provided in the other mode" (p. 1185). The additional meaning could occur in the image (i.e., image extends text) or in the text (i.e., text extends image).

A few scholars have looked at image-text relations from the system proposed by Martinec and Salway (2005) and Unsworth (2006, 2007). Lai (2018) examined the relations between text and image in junior high school EFL textbooks in China using the thirteen categories of image-text relations proposed in Martinec and Salway (2005). Examples of twelve types of relations were found in the six textbooks examined except for the relations of enhancement (temporal), for which there were none. Lai also found that the progression of proficiency level coincided with the decrease in the amount of images in the textbooks. As the textbooks' intended students moved up in grade level and thus English proficiency, the textbooks shifted from being image-based to text-based. This broadly coheres with Liu and Qu's (2014) study of the multimodality of Chinese college EFL textbooks from the perspective of page layout, which also found that textbooks directed at lower



**Figure 1** Unsworth's (2007, p. 1175) framework of ideational image-text relations.

proficiency learners were more picture-based while those for higher proficiency learners were more text-based, with much fewer visuals.

Nhat and Pha (2019) examined picture book comics using Unsworth's (2006) framework but focused exclusively on the representational meaning. They found that Expansion and Projection each accounted for around half of the image-text relations. For Expansion, Concurrence and Complementarity each accounted for two-fifths of the occurrences, suggesting that "the re-expression of meanings alternates with the extension of meanings throughout the story" (p. 135). Specifically, Complementarity relations accounted for 20.7% of the overall image-text relations, with Augmentation taking up 16% and Divergence taking up 4.7%. Complementarity in Unsworth's (2006) framework is similar to what Martinec and Salway (2005) termed extension, when there is additional but consistent meaning in either the image or the text. Compared to Lai's (2018) study of textbook image-text relations, which found that extension accounted for only 1.1% of the total amount, in this study of a picture book comic, complementarity between image and text was found in 20% of the cases. This is understandable, as picture books can include many types of image-text relations to creatively tell a story while language textbooks might need to emphasize reiteration of meaning between image and text for the images to support comprehension of the linguistic text, and by extension, the learners' language development.

Martinec and Salway's (2005) work has been widely used to explore image-text relations in a variety of text types except language learning materials, with Lai's (2018) study a rare example, as textbook analysis concerning visual-verbal relations are mostly conducted from the GVD framework. The framework of image-text relations proposed in Martinec and Salway (2005) and Unsworth (2006, 2007) provide a system which does not begin from visual analysis but specifically explores the types of relations between images and texts. Thus, this is a more suitable framework than GVD for the current research which examines image-text relations in the animations that accompany locally-produced English-language textbooks. Unsworth's (2007) framework will be drawn upon as it follows from Martinec and Salway's (2005) comprehensive study and further expands on it.

### **Animations**

In recent years, animations have begun to be explored for their impact on the learning of English-as-a-foreign-language, as they can better reflect the dynamics of the sequential and temporal nature of language. A few studies have explored the extent to which EFL learners' speaking ability can be enhanced as a result of using animated materials for teaching formulaic expressions. For example, Najafabadi *et al.* (2019) Ghaderi and Afshinfar (2014) similarly found that even though formulaic expressions can be learned with or without the use of animations, students improved more and better retained what they learned when animated materials were used.

Other scholars examined the potential of animation in supporting different aspects of learning. Niknejad and Rahbar (2015) examined animations and their potential for developing EFL learners' reading comprehension and found that animations were more effective than static images in enhancing learners' comprehension of texts. Bakla (2019) zeroed in on the outcome of animations on EFL students' learning of punctuation in writing, and found that animations provide more effective support for the learning of punctuations than text-based materials.

While more research is necessary to determine how animations support language learning, another fundamental but as yet unexplored aspect of research on animations is how they relate to the language learning materials they accompany, i.e., the verbal text and the associated images. The present study fills this gap by exploring image-text-animations relations using pedagogical language-learning materials for which all three aspects are available.

## Method

This study investigates image-text relations in EFL textbooks focusing on moving-image animations that accompany verbal dialogues. The study also explores how animations support verbal dialogues in the textbooks. Data sources include three major versions of junior high school textbooks locally-produced in Taiwan, each consisting of 6 books, one for each semester. Thus, a total of 18 junior high English-language textbooks were analyzed. In each textbook, there are five lessons or units, each composed of a dialogue section and a reading section. This study highlights the dialogue section of the EFL textbooks. The dialogues include consecutive conversational turns-at-talk, usually between two or three people. They are usually accompanied by one or two images that portray the dialogues. In recent years, publishers of these textbooks have also begun to provide animations, that is, short videos, of the dialogues. They are moving-image cartoons of the textbook dialogues. These accompanying videos are provided to teachers and can also sometimes be found online. The aforementioned images that are placed in the textbooks along with the dialogues are often screen captures from the animations. Examples of the dialogues and information on the animations can be found in the Findings section below.

In the analysis of the data, I began by reading the dialogues and their accompanying still images in the textbooks and watching the animations. After familiarizing with each, I read and viewed them alongside each other for coding purposes. As stated earlier, Unsworth's (2007) system of image-text relations served as the framework for this study. Specifically, for each lesson, I began with the still images that accompany each textbook dialogue. There are no more than two or three images for each dialogue. I identified the conversation turns the images are related to and coded them for their image-text relations. Then I moved on to the animations using the verbal dialogues as points of departure. That is, after reading each conversation turn in the dialogue, I looked at how this turn is constructed in the animation, made a screenshot, and coded for the intersemiotic relations. In cases when each conversation turn includes more than one sentence, each sentence functions as the unit of analysis, and I examined whether a different representational meaning is portrayed in the animation for the sentence, and if that were the case, a different screenshot was made and then coded. Furthermore, as animations are moving-image continuous portrayals, they allow for embodied interaction to occur *between* conversation turns. In the occurrence of such cases, screenshots were made and coded. This coding process was conducted for all the dialogues in each textbook across all 6 textbooks from the three publishers. In this first stage of data analysis, the first layer of image-text relations was examined, i.e., the image-text-animations were coded for whether they showed a relation of Concurrence, Complementarity, or Enhancement in terms of Expansion and Verbal or Mental in terms of Projection (see Figure 1 for categories of image-text relations).

Then, following an iterative process, in the second stage of data analysis, all of the data were examined again following the process described above, but this time, I examined the second-layer of relations, namely, each still image and screenshot from the animations were examined for whether they showed a relation of Clarification, Exposition, Exemplification, or Homospatiality in terms of Concurrence and whether they showed a relation of Augmentation and Divergence in terms of Complementarity with the verbal dialogue, as well as which of the 5 types of Enhancement they displayed. The same was also done for Projection: Mental looking for either Perception or Cognition. (see Figure 1 for details). Finally, in the third stage of data analysis, the constant comparative method was employed to ensure that the criteria for assigning the codes was consistent across all dialogues in the 18 textbooks.

## Findings

In presenting the findings, my aim is *not* to explicate the various types of animation-image-text relations found but to focus on *Complementarity: Augmentation* between the animations and verbal dialogues, as this aspect of image-text relations in language textbooks has not been researched at all and was an

unexpected finding. In what follows, each turn-at-talk in the textbook dialogues cited below has been numbered, and will henceforth be referred to as a turn (i.e., a conversation turn-at-talk) along with its corresponding number (such as turn 9).

### Complementarity: Augmentation of Emotions

As Su *et al.* (2007) pointed out, most of interpersonal communication is achieved through body language while relatively very little is accomplished linguistically. Animations, by nature of their moving-image dynamic aspects, better allow for the portrayal of body language, such as facial expression and gestures. In addition, animations also frequently incorporate comic symbols, which “are mainly used to visually represent emotions and the state of the characters” (Akai, Yamashita, & Matsushita, 2015, p. 1). Both comic symbols and gestures are often made use of in these textbook animations to show characters’ emotions.

The first example (see below) is taken from Lesson 2 Book 1 of Kang Hsuan publisher (Feng, 2019, pp. 22-23). The animation can be found on the publisher’s YouTube page (<https://www.youtube.com/watch?v=D2Qdz8HC9xA>). The Figures below are screenshots that I took from the animations found on YouTube page. The conversation takes place at Rita’s home, with Nick unable to correctly identify the animals in some of the pictures. What is particularly interesting about the animation is that it fills in the gaps between turns about how Nick feels about his failure to identify the animals, and this information is not present in the verbal dialogue in the textbook. In the dialogue, after Nick asks whether one animal is a lion and is told by Rita that it’s a dog (turns 3 and 4), he continues with a comment about the dog, i.e., that “it’s big and scary” (turn 5) before he goes on to begin another question about a different picture (turn 5). Thus, in the dialogue, Nick merely provides a commentary (turn 5) of the dog about it being “big and scary” after he fails to correctly identify the picture. The animation, however, reveals that Nick is very surprised about the animal being a dog rather than a lion through the use of comic symbols, i.e., two thick red flashing exclamation marks placed over his head (see Figure 2). This occurs in-between turns 4 and 5.

Then, after Nick learns that one of the pictures he thought to be a fox was actually a dog (turn 6), the animation portrays him with two flashing red hearts respectively placed over his eyes (see Figure 3). They appear during Nick’s turn 7 when he says “It’s cute.” The comic symbol of the two flashing hearts

(At Rita’s home)

- 1 Nick: Are these your pictures?
- 2 Rita: No, they’re not. They’re my parents’ pictures.
- 3 Nick: Their pictures are cool. What’s the animal in this picture? Is it a lion?
- 4 Rita: No, it’s a dog.
- 5 Nick: It’s big and scary. What about the animal in that picture? Is that a fox?
- 6 Rita: No, it isn’t. It’s a dog, too.
- 7 Nick: It’s cute.

(At Rita’s home; continued)

- 8 Nick: What are these? Are these sheep?
- 9 Rita: No, they’re dogs, too.
- 10 Nick: What are those animals in that picture? Are they dogs?
- 11 Rita: Wrong again. They’re black bears.
- 12 Nick: No way! These pictures are cool.

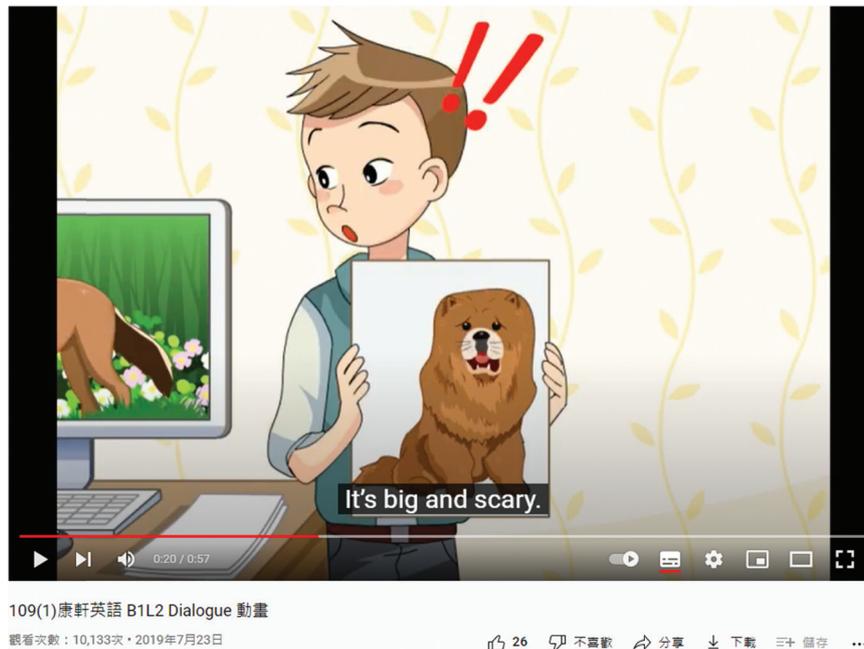


Figure 2 Thick red flashing exclamation marks.

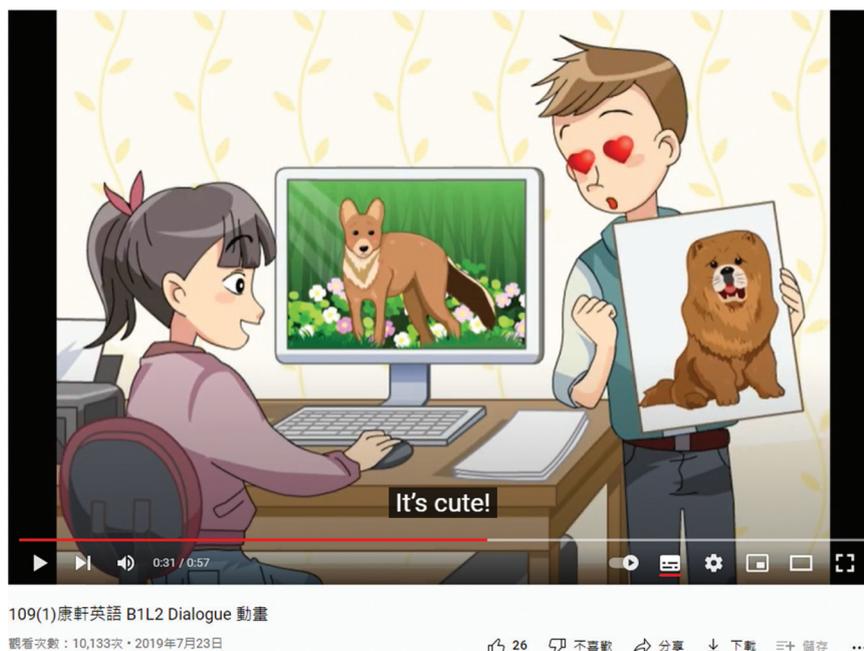


Figure 3 Flashing red hearts.

symbolizes that he really likes the dog. While the meaning of these flashing-heart symbols cohere with what Nick says in turn 7, it is nevertheless a relation of *Complementarity: Augmentation*. “It’s cute” is more a commentary of the picture, and does not necessarily reflect how the speaker feels about the picture. The comic symbols, however, show the emotion of love, or strong affection. They could be verbally glossed as “I really like it,” which would be an emotional expression rather than a commentary about the positive appearance of an animal such as “it’s cute.”

Finally, in the second part of the dialogue, Nick continues to inquire about other animal pictures. What he identifies as sheep turns out to be dogs (turns 8 and 9). In the dialogue, Nick does not show any

response or even make any commentary about the dogs. Instead, he immediately moves on to asking Rita about another picture. Thus, in the text, Nick is not shown as having any emotional response to his third failure at identifying an animal (turns 8 and 9). In the animation, however, he is portrayed as feeling embarrassed and/or puzzled, scratching the back of his head immediately after Rita gives her answer in turn 9 (see Figure 4) and continuing while he asks about the next picture (see Figure 5). The scratching head gesture can show “disagreement, puzzlement, shyness or anxiety” (Su *et al.*, 2007, p. 285), and in this case, expresses embarrassment on the part of Nick. The hand does not come away from the back of his head while he asks, “What are those animals in that picture?” in turn 10. This gesture of embarrassment/puzzlement is portrayed from turn 9 to turn 10. It is only until the start of the



**Figure 4** Nick scratching his head after Rita's conversation turn.



**Figure 5** Gesture continues into Nick's conversation turn.

second part of turn 10 (“Are they dogs?”) that he is portrayed as making another gestures, i.e., pointing to another picture.

Thus, there are three instances in this dialogue in which Nick wrongly identifies an animal in a picture. In one of the cases, Nick’s surprise (between turns 4 and 5) is shown in the animation using big red exclamation marks. In the next, his affection for the dog (in turn 7) is shown in the animation using two flashing red hearts over his eyes. In the third case, his embarrassment (between turns 9 and 10) is shown in the animation using the gesture of scratching the back of his head. None of these emotions is expressed in the verbal text but only displayed in the animation, providing a relation of *Complementarity: Augmentation* between the text and the moving-images.

The second example below is taken from Lesson 3 Book 1 of Nan I publisher (Chen, 2019, pp. 62-63). The animation can be found on the publisher’s YouTube page (<https://www.youtube.com/watch?v=Og-WxpLWscYM>). The conversation takes place at Jason’s potluck, attended by Rosa, Ming, and Marco.

(At Jason’s potluck)

- 1 Jason: Hi, Rosa! Come on in! Please put your dish on the table.
- 2 Rosa: Ok. Thanks.
- 3 Ming: Let’s try the beef noodles.
- 4 Marco: It’s really good.
- 5 Ming: Marco, let’s use chopsticks.
- 6 Marco: Oh, no. That’s not easy.
- 7 Ming: Just give it a shot!
- 8 Rosa: Let’s have tacos.
- 9 Jason: The tacos are delicious.
- 10 Ming: Yes, they are.
- 11 Rosa: Hey, guys. Don’t use chopsticks. Just use your hands.
- 12 Marco: That’s easy.
- 13 Jason: Hey, Marco, where is your dish?
- 14 Marco: I’m sorry. I...
- 15 Jason: It’s fine. Just bring TWO Italian dishes next time.

The second half of the conversation begins at turn 8. In turn 9, when Jason comments that “the tacos are delicious,” Ming replies, “Yes, they are” (turn 10). However, during Ming’s turn-at-talk (turn 10), not only is he depicted (see Figure 6) as holding a taco with his chopsticks and eating without difficulty, Marco, who does not have a speaking turn in the verbal text, is also shown (see Figure 6) as having trouble using chopsticks to pick up his tacos; the animation shows Marco dropping the tacos again and again. His frustration towards this is depicted in a series of black lines over the top half of his face. Colored lines over a character’s face is also a frequently used comic symbol (Kincaid, 2011) to show negative emotions such as frustration or annoyance. Marco’s difficulty eating tacos with chopsticks is not stated in the verbal text but only depicted in the animation, i.e., a relation of *Complementarity: Augmentation*. This is an important visual augmentation, as it provides the rationale for Rosa’s seemingly sudden and abrupt suggestion to not use chopsticks but to use hands (in turn 11), as no one in the previous turns (turns 8, 9, and 10) made any reference to how tacos should be eaten.

Then, at turn 13, Jason questions Marco about not bringing any dishes to the potluck. Marco replies by saying at turn 14, “I’m sorry. I...” The animation shows Marco in a posture of embarrassment

during turn 14, with his right hand scratching the back of his head (see Figure 7). This gesture means either “disagree, puzzled, shy, anxious,” according to Su *et al.*'s (2007, p. 285) observations. Marco is most likely anxious and embarrassed to be found out. The animation also portrays what happens to Marco between turns 13 and 14, i.e., before he states his apology in turn 14. Marco is depicted as having gone completely gray, accompanied by an impact mark above his head (see Figure 8). These



**Figure 6** Black lines over Marco's face as he has trouble using chopsticks.



**Figure 7** Marco's posture of embarrassment.



**Figure 8** Comic symbols portraying Marco between turns 13 and 14.

two comic symbols cohere with each other and repeats the message about how Marco must feel when his lack of contribution to the potluck was pointed out by Jason in turn 13. Darkness is understood to show disappointment or anxiety (Shinohara & Yoshihiro, 2009). In this case, based on the contents of the dialogue, it shows anxiety. Impact marks are also often found in manga, and is “a crown-like mark banging away from the character head...used when they notice something, realize something, or are surprised by something” (Realization Mark, 2020). Marco is most likely shocked and embarrassed that he was accused in front of everyone. This visual is not an accompaniment of any verbiage but is shown between Jason’s (turn 13) and Marco’s (turn 14) speaking turns. It has a relation of *Complementarity: Augmentation* as it shows Marco’s emotional reaction immediately after being questioned by Jason in turn 13, which was not expressed in the dialogue.

To sum up, animations share a relation of *Complementarity: Augmentation* with the verbal dialogue they depict in a number of ways. They can reveal a speaker’s emotional response that is not expressed verbally in the dialogue (such as in Nick’s and Marco’s cases), they can provide the rationale for someone’s suggestion (such as in Rosa’s case), and they can portray another participant in the dialogue in addition to the speaker of the conversation turn (such as in Marco’s case during Ming’s speaking turn).

### Animation as Interjection

What can be gleaned from the above discussion is that many interjections are missing from the textbook dialogues. Following Wierzbicka (1992) and Goffman (1981), Wharton (2003) explained that form-wise, an interjection “is capable of constituting an utterance by itself in a unique, non-elliptical manner” (Wharton, 2003, p. 42) and function-wise, it “expresses a mental or emotional attitude or state” (p. 43). Similarly, Ameka (1992) categorized interjections as either primary or secondary based on their form. Primary interjections are words that “are not used otherwise” (p. 105) while secondary ones “belong to other word classes based on their semantics and are interjections only because they can occur by themselves non-elliptically as one-word utterances” (p. 105). Ameka (1992) also classified interjections based on their communicative function, including, most relevant to the current discussion, the expressive function, which can be subdivided into the emotive and the cognitive. Emotive

interjections are “those that express the speaker’s state with respect to emotions and sensations they have at the time” while cognitive ones “are those that pertain to the state of knowledge and thoughts at the time of utterance” (p. 113). As such, interjections show emotional or mental responses in verbal dialogues in the way comic symbols do for manga characters. While the animations of the textbooks examined “insert” these responses of the characters using comic conventions, students also need to learn verbal interjections. Without their inclusion, the textbook dialogues seem rather stiff, unnatural, and stilted.

In the dialogue between Nick and Rita when Nick fails to correctly identify animals in Rita’s parents’ pictures, the conversation seems contrived because it lacks many interjections that “encode speaker attitudes and communicative intentions” (Ameka, 1992, p. 107). In turn 5, as the comic symbol of two red exclamation marks show, Nick is astonished or even shocked by Rita’s answer. In this case, a possible interjection might be “what?” in the turn-initial position to show Nick’s emotional reaction towards the picture being that of a dog rather than a lion. An interjection should also have been present in turn 7 and turn 10 in the same dialogue between Nick and Rita. In turn 7, a common interjection might be “aww” before the assessment “it’s cute.” Instead, there is no interjection in the dialogue; Nick’s emotional response is shown only through the comic symbol of two hearts over his eyes. In turn 10, which is the most peculiar of all, Nick does not make any verbal comment or give any reaction to getting the animal picture wrong (in turns 8 and 9) for the third time but immediately moves on to talking about another picture. However, in the animation, he is shown as being embarrassed, as expressed in the gesture of scratching his head. In addition, finding out that the picture is of a bear rather than a dog certainly is new information to Nick. Such a case would certainly necessitate an expression that “indicates a shift in speaker orientation to information presented by another interlocuter” (Fuller, 2003, p. 28), such as an “oh” before moving on to asking another question. That is why “oh” is often described as a change-of-state token (Bolden, 2006) used to show that the listener has received new knowledge.

In the other dialogue of Jason’s potluck, an interjection is also missing in the turn-initial position in turns 12 and 14. Because the animation shows that Marco has difficulty eating taco with chopsticks, he feels relieved when Rosa suggests to use their hands (turn 11). In the animation, Marco is shown to be happy and relieved, which necessitates an interjection, possibly “phew,” before announcing that it’s easy to eat with hands (turn 12). Also, after Jason questions Marco about not contributing a dish to the potluck (turn 13), Marco replies in turn 14 “I’m sorry. I...” A common interjection in this case that prefaces an apology might be “Oops,” which is an interjection that is frequently used with apologies (Jucker, 2019). Instead of a verbal interjection, the scratching head gesture that shows embarrassment is portrayed in the animation.

Wong (2002, p. 37) argued that “textbook conversations fail to match findings from empirical studies” (p. 37). This article shows that the lack of interjections is one big reason. While on many occasions the verbal dialogues do not show the speakers having any emotional reactions, the animations do portray them clearly, either through gestures or comic symbols. Thus, in these EFL pedagogical materials, rather than using verbal pragmatic markers such as interjections (Norricks, 2007) in the conversations, they are replaced by comic symbols or portrayal of gestures in the animation. This is an interesting aspect of how animations support EFL textbook dialogues that warrant further research.

### Conclusion and Implications

This study investigated image-text relations in EFL textbooks focusing on moving-image animations that accompany verbal dialogues and explored how animations support verbal dialogues in the textbooks. Findings revealed that animations enter into a relation of *Complementarity: Augmentation* by providing information that is consistent with and additional to the dialogues in a number of ways.

In the examples discussed above, the animations provide information regarding a person's emotional state not made apparent in the dialogues. The animations show the emotional reactions (including astonishment and embarrassment) of a boy who repeatedly failed to correctly identify pictures of animals. Comic symbols were used to portray another character's frustration with chopsticks and delight at using his hands in addition to his shock and embarrassment at having being accused of not contributing a dish to a potluck. Interestingly, in naturally-occurring conversations, these emotions would usually be expressed through interjections, but are mostly absent from the verbal dialogues in these textbooks. As such, the animations also serve to substitute the pragmatic markers that should have been present in the verbal text.

Wong (2002, p. 37) observed that the "mismatch between textbooks and naturally occurring language" (p. 37) has serious consequences for language learners because textbooks are often "marketed as offering authentic, natural language, or language which is true to life" (p. 37). Similar critiques of language textbooks have been abundant (e.g., Bardovi-Harlig, 2001; Boxer & Pickering, 1995; Tatsuki, Kite, & Maeda, 2007). As Belz (2007) summarized, language textbooks "include little information on L2 pragmatics, lack explicit discussions of conversational norms and practices, and contain inauthentic language samples that are based on introspection rather than genuine language use" (p. 48). In the case of this study, interjections, often missing in the textbook dialogues, are an important part of verbal communication (Hismanglu, 2010). Padilla (2010) argued that usage of interjections is an integral aspect of communicative competence for foreign language learners, and pedagogical materials need to incorporate them as a necessary component of the language curriculum.

Multimodal pedagogical materials can be a good place to begin, such as in the examples discussed above. If focusing only on the verbal dialogues in the textbooks, it might be difficult for language learners to consider what might be missing from them. However, using the animations that share a relation of *Complementarity: Augmentation* with the dialogues, teachers can have students examine the comic symbols and gestures that occur along with the characters' speech in the animations. When students have identified them, teachers can then guide students to explore what might be said verbally that cohere with or re-present what is in the visual and gesture modes. As such, the communicative power of interjections to express an emotive or cognitive response to the interlocutor's prior talk can be better exemplified and understood by students. These pragmatic markers might often be missing in the textbook dialogues because they do not alter the syntax of a sentence but makes a difference in whether a speaker's language shows idiomaticity.

Despite increasing interest in the multimodality of pedagogical language materials from the perspective of image-text relations, a large area that has been neglected to date is the examination of animations that accompany language materials such as textbooks. This study has shown that the ways in which animations relate to and support textbooks are different from how still images relate to texts. Thus, more research is necessary that explores textbook-related animations from other locales and also commercially-produced textbooks that are distributed globally. In addition, research that addresses how teachers and learners make use of these animations is also required in order to come to a more comprehensive understanding of how different modes of pedagogical materials can work together to support language learning and development.

### Acknowledgment

This research received funding support from the Ministry of Science and Technology of Taiwan (MOST 108-2628-H-002-003-MY2).

## References

- Akai, Y., Yamashita, R., & Matsushita, M. (2015). Giving emotions to characters using comic symbols. *Proceedings of the 12<sup>th</sup> International Conference on Advances in Computer Entertainment Technology*. 1–4.
- Ameka, F. (1992). Interjections: The universal yet neglected part of speech. *Journal of Pragmatics*, 18, 101–118.
- Bakla, A. (2019). A mixed-methods study of tailor-made animated cartoons in teaching punctuation in EFL writing. *ReCALL*, 31(1), 75–91.
- Bardovi-Harlig, K. (2001). Empirical evidence of the need for instruction in pragmatics. In K. Rose & G. Kasper (Eds.), *Pragmatics and language teaching* (pp. 13–22). Cambridge, UK: Cambridge University Press.
- Bateman, J. (2014). *Text and image: A critical introduction to the visual/verbal divide*. London: Routledge.
- Belz, J. A. (2007). The role of computer mediation in the instruction and development of L2 pragmatic competence. *Annual Review of Applied Linguistics*, 27, 45–75.
- Bezemer, J., & Kress, G. (2010). Changing text: A social semiotic analysis of textbooks. *Designs for Learning*, 3(1-2), 10–29.
- Bolden, G. B. (2006). Little words that matter: Discourse markers “so” and “oh” and the doing of other-attentiveness in social interaction. *Journal of Communication*, 56, 661–688.
- Boxer, D., & Pickering, L. (1995). Problems in the presentation of speech acts in ELT materials: The case of complaints. *ELT Journal*, 49, 44–58.
- Chen, H. R. (2019). *Junior High School English. Book 1. 國民中學英語 (第一冊)*. Taipei: Nan I.
- Chen, Y. (2010a). Exploring dialogic engagement with readers in multimodal EFL textbooks in China. *Visual Communication*, 9(4), 485–506.
- Chen, Y. (2010b). The semiotic construal of attitudinal curriculum goals: Evidence from EFL textbooks in China. *Linguistics and Education*, 21, 60–74.
- de Oliveira, L. C., Smith, S. L., Jones, L., & de Almeida, C. R. (2018). Strategies for working with image-text relations in picture books. In N. Guler (Ed.), *Optimizing elementary education for English language learners* (pp. 177–195). Hershey, PA: IGI Global.
- Djonov, E. (2007). Website hierarchy and the interaction between content organization, webpage and navigation design: A systemic functional hypermedia discourse analysis perspective. *Information Design Journal*, 15(2), 144–162.
- Economou, D. (2014). Telling a different story. In E. Djonov & S. Zhao (Eds.), *Critical Multimodal Studies of Popular Discourse* (181–201). New York: Routledge.
- Feng, H. P. (2019). *Junior High School English. Book 1. 國民中學英語 (第一冊)*. Taipei: Kang Hsuan.
- Fuller, J. M. (2003). The influence of speaker roles on discourse marker use. *Journal of Pragmatics*, 35, 23–45.
- Ghaderi, V., & Afshinfar, J. (2014). A comparative study of the effects of animated versus static funny pictures on Iranian intermediate EFL students’ intake and retention of idiom. *Procedia—Social and Behavioral Science*, 98, 522–531.
- Goffman, E. (1981). *Forms of talk*. Oxford: Blackwell.
- Halliday, M. A. K. (1994). *Functional grammar* (2<sup>nd</sup> ed.). London: Arnold.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar*. London: Arnold.
- Hismanglu, M. (2010). Interjections in English: Neglected but important aspect of foreign language learning and teaching. *Journal of Theory and Practice in Education*, 6(1), 17–35.
- Hiippala, T. (2015). *The structure of multimodal documents: An empirical approach*. New York: Routledge.

- Jauhara, D., Emilia, E., & Lukmana, I. (2021). Re-contextualizing “greeting”: A multimodal analysis of an EFL textbook. *Indonesian Journal of Functional Linguistics*, 1(1), 15–24.
- Jucker, A. H. (2019). Oops, I forgot, sorry: The spill cries oops and whoops in the history of American English. *Lingue Linguaggi*, 31, 1–33.
- Kincaid, C. (2011). Anime’s visual language. Retrieved from <https://www.japanpowered.com/japan-culture/animes-visual-language>
- Kress, G., & van Leeuwen, T. (1996/2006). *Reading images: The grammar of visual design*. New York: Routledge.
- Lai, H. (2018). Image-text relations in junior high school EFL textbooks in China: A mix-methods study. *Journal of Language Teaching and Research*, 9(6), 1177–1190.
- Liu, Y., & O’Halloran, K. L. (2009). Inter-semiotic texture: Analyzing cohesive devices between language and images. *Social Semiotics*, 19(4), 367–288.
- Liu, X., & Qu, D. (2014). Exploring the multimodality of EFL textbooks for Chinese college students: A comparative study. *RELC Journal*, 45(2), 135–150.
- Martinec, R., & Salway, A. (2005). A system for image-text relations in new (and old) media. *Visual Communication*, 4(3), 337–371.
- Najafabadi, M. R., Bakhshizadeh, Y., & Mirzaei, A. (2019). Developing EFL learners’ oral proficiency through animation-based instruction of English formulaic sequences. *Journal of Teaching Language Skills*, 38(1), 157–190.
- Nhat, T. N. M., & Pha, N. T. M. (2019). Exploring text-image relations in English comics for children: The case of “Little Red Riding Hood”. *VNU Journal of Foreign Studies*, 35(3), 128–138.
- Niknejad, S., & Rahbar, B. (2015). Enhancing EFL learners’ reading comprehension ability through multimedia-based visualization. *Journal of Applied Linguistics and Language Research*, 2(6), 119–127.
- Norrick, N. R. (2007). Interjections as pragmatic markers. *Journal of Pragmatics*, 4, 866–891.
- Padilla C. M. (2010). Teaching interjections in the ESL/EFL class: A pragmatic approach. In L. Pérez Ruiz, I. Parrado Román & P. Tabarés Pérez (Eds.), *Estudios de metodología de la lengua inglesa, vol. V* (pp. 23–33). Valladolid: Universidad de Valladolid.
- Page, J. T. (2006). Myth and photography in advertising: A semiotic analysis. *Visual Communication Quarterly*, 13(2), 90–109.
- Realization Mark. (2020, March 22). *Japanese with anime and kanji with manga*. Retrieved from <https://www.japanesewithanime.com/2020/03/realization-mark.html>
- Salbego, N., Heberle, V. M., & Balen, M. G. S. D. S. (2015). A visual analysis of English textbooks: Multimodal scaffolded learning. *Calidoscopio*, 13(1), 5–13.
- Shinohara, K., & Yoshihiro, M. (2009). Pictorial metaphors of emotion in Japanese comics. In C. Forceville & E. Urios-Aparisi (Eds.), *Multimodal Metaphor* (pp. 265–293). Berlin: Mouton de Gruyter.
- Sipe, L. R. (1998). How picture books work: A semiotically framed theory of text-picture relationships. *Children’s Literature in Education*, 29, 97–108.
- Sipe, L. R. (2012). Revisiting the relationships between text and pictures. *Children’s Literature in Education*, 43, 4–21.
- Su, W. P., Pham, B., & Wardhani, A. (2007). Personality and emotion-based high-level control of affective story characters. *IEEE Transactions on visualization and computer graphics*, 13(2), 281–293.
- Tatsuki, D., Kiate, Y., & Maeda, J. (2007). *What kinds of pragmatic information are included in language textbooks?* A poster presented at the 17<sup>th</sup> International Conference on Pragmatics and Language Learning, March 27m 2007, Honolulu, Hawaii.
- Teo, P., & Zhu, J. (2018). A multimodal analysis of affect and attitude education in China’s English textbooks. *Multimodal Communication*, 7(1), 1–15.

- Unsworth, L. (2006). Toward a metalanguage for multiliteracies education: Describing the meaning-making resources of language-image interaction. *English Teaching: Practice and Critique*, 5(1), 55–76.
- Unsworth, L. (2007). Image/text relations and intersemiosis: Towards multimodal text description for multiliteracies education. *Proceedings of the 33<sup>rd</sup> International Systemic Functional Congress (33<sup>rd</sup> ISFC)* (pp. 1165–1205). Pontificia Universidade Catolica De Sao Paulo, Brazil.
- Wharton, T. (2003). Interjections, language, and the “showing/saying” continuum. *Pragmatics & Cognition*, 11(1), 39–91.
- Wierzbicka, A. (1992). The semantics of interjection. *Journal of Pragmatics*, 18, 159–192.
- Wong, J. (2002). “Applying” conversation analysis in applied linguistics: Evaluating dialogues in English as a second language textbook. *IRAL*, 40, 37–60.