

# From Trending to Teaching: A Framework to Analyze TikTok Videos for Vocabulary Instruction

Alice Shanthi

Academy of Language Studies, Universiti Teknologi MARA (UiTM) Kampus Seremban 3, Malaysia

Nur Alyaa Fatinah Jumaat

Academy of Language Studies, Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia

Puspalata C Suppiah (Dr)

Academy of Language Studies, Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia

Elia Md Johar (Dr)

Academy of Language Studies, Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia

Nalini Arumugam (Associate Professor, Dr)

Faculty of Education, Languages, Psychology and Music, SEGI International University, Petaling  
Jaya Selangor, Malaysia

Received: March 29, 2024      Accepted: April 17, 2024      Published: May 14, 2024

doi:10.5296/ijssr.v12i2.21904      URL: <https://doi.org/10.5296/ijssr.v12i2.21904>

## Abstract

TikTok, a popular short-form video platform, has evolved into a versatile communication tool with potential applications in education. Examining the various multimodal components present in TikTok videos can offer valuable insights for vocabulary acquisition. TikTok combines multiple modes (multimodal), like visuals, text, audio, and movement, to engage users. Understanding the synergy between these modes on TikTok, in conveying meaning and maintaining engagement, could inform more effective vocabulary teaching methods. This paper delves into the examination of two frameworks for analysing multimodal elements in pedagogical materials. The first framework discussed is Serafini's (2010), while the second is Machin and Mayr's (2012). The findings suggest that a key distinction lies in Serafini's narrow focus on pedagogy compared to Machin and Mayr's broader approach, which scrutinises the construction of meaning. The study concludes that Serafini's framework is most appropriate for analysing multimodal elements in pedagogical materials as it directly addresses the learning objectives in student learning. The results of this study hold implications for students, educators, and educational establishments.

**Keywords:** TikTok, teaching and learning, Serafini (2010), Machin and Mayr (2012), multimodal

## 1. Introduction

### *1.1 TikTok Video for Vocabulary Enhancement*

Investigating the various multimodal aspects found within TikTok videos can provide valuable insights for vocabulary learning. Vocabulary forms the base for learning any language; a rich vocabulary correlates with greater academic performance across all language skills, including reading, writing, speaking, and listening, aiding in attaining quicker fluency in the language. In recent times, TikTok videos have emerged as a tool for enriching learners' vocabulary. TikTok combines multiple modes, such as visuals, text, audio, and movement, to engage users. Understanding how these modes work together in TikTok to convey meaning and sustaining attention could inform effective vocabulary instruction. Studying multimodal interplay in TikTok can reveal best practices for retaining students' focus, associating words with rich contextual cues, leveraging memory movement, and embedding definitions seamlessly into content. As a platform for short-form educational content, TikTok exemplifies multimodal techniques that resonate with learners. Examining these multimodal strategies using frameworks like Serafini's or Machin and Mayr's could uncover principles for effectively merging modes to teach vocabulary in engaging ways.

Vocabulary retention is a critical educational challenge, with students often encountering difficulty in fully internalising new words. At the same time, platforms like TikTok use multimodal techniques that seamlessly integrate audio, visuals, text, and movement to hold the viewer's attention and convey meaning effectively. While numerous studies have explored educational uses of social media platforms, TikTok remains relatively understudied in this regard. This gap presents an opportunity to investigate how TikTok's widespread multimodal content engages users and to consider how such approaches could enhance vocabulary instruction. Additionally, previous studies have also shown that there is insufficient exploration of TikTok's multimodal strategies from a pedagogical perspective (Sharma & Giannakos, 2020; Anderson & Kachorsky, 2019). This is problematic, as valuable insights into TikTok's engagement with learners are not being leveraged to address the recognised vocabulary retention issue. Further research is necessary to analyse TikTok's interplay of modes using frameworks like Serafini's, to derive evidence-based best practices for multimodal vocabulary in the context of teaching and learning.

Several prominent frameworks, akin to TikTok, serve as useful tools for examining multimodal elements within social media. Firstly, Serafini's (2010) framework scrutinises multimodal aspects in pedagogical materials to understand their impact on learning, involving the analysis of modal interplay, strategies, objectives, and outcomes. Another notable framework is proposed by Machin and Mayr (2012), offering a flexible approach applicable to any discourse that considers contextual factors, modal types and combinations, information design, and meaning construction. Furthermore, Kress and Van Leeuwen (2001) have developed a social semiotic theory that explores how modes are combined to communicate meaning and power relations, serving as an important tool for critically examining multimodality in social media. O'Halloran et al. (2008) established Systemic Functional Multimodal Discourse Analysis (SF-MDA), drawing on systemic functional linguistics to

study communicative contexts such as textual, visual, and other modes in digital media. Baldry and Thibault's (2006) Multimodal Transcription and Text Analysis offer notation tools and analytical concepts for studying relationships between modes like gesture, speech, and images, applicable to interactions on social media platforms. Finally, Lemke's (2002) Typology of Intermodal Relations classifies types of connections between modal elements, such as redundancy, complementarity, or opposition, valuable for understanding mode interplay within social media contexts.

Given the abundance of frameworks available for studying multimodality in short videos, the selection of a framework relies on various factors including research goals, scope, platforms, and modes under investigation. It is also possible to integrate or adapt multiple frameworks to create a tailored analytical approach. Nonetheless, not all frameworks mentioned earlier are suitable for studying multimodality in TikTok videos, particularly for educational purposes. Therefore, this study opted to concentrate solely on the two frameworks for the following reasons. Firstly, both Serafini's (2010) and Machin and Mayr's (2012) frameworks are well-suited for analysing multimodal elements in TikTok videos, especially in comparison to other frameworks. Serafini's framework, specifically, is aligned with the pedagogical focus, examining multimodal elements in educational materials/contexts, which matches the goal of analysing TikTok videos for insights into vocabulary teaching.

Next, in terms of the socio-cultural aspect, Machin and Mayr's framework incorporates social and cultural contextual factors, which is essential for comprehending TikTok, deeply rooted within youth culture. Regarding the range of modes, both frameworks account for a broad spectrum of multimodal elements—textual, visual, auditory, and gestural—thus encompassing TikTok's varied mode utilisation. Moving on to the analysis of interplay, both frameworks facilitate the analysis of interconnections and relationships between modes, which is crucial as TikTok combines modes seamlessly. As for meaning construction, Machin and Mayr's model explores how modes construct meaning, allowing for an examination of how TikTok communicates messages multimodally. Next, in terms of systematic application, these frameworks provide clear guidelines and principles for systematically applying multimodal analysis. Finally, both frameworks are flexible yet rigorous, adaptable to different platforms and content types while maintaining theoretical rigour. Hence, this study chose to utilise and compare Serafini and Machin and Mayr's frameworks to determine which is more compatible in providing comprehensive, socially attuned, systematic frameworks for uncovering multimodal strategies, meaning-making, and educational insights, ensuring suitability for examining a platform like TikTok. Other frameworks may have narrower scopes or lack specific guidance.

### *1.2 Problem Statement*

While TikTok's combination of modes like visuals, text, audio, and movement appears effective for engagement and communication, comprehensively grasping the multimodal strategies at work necessitates a systematic analytical approach. Without a defined framework to evaluate and compare techniques, it is challenging to derive rigorous, structured insights on why particular multimodal interplays succeed. An exceptional framework provides

consistency in examining relevant modal categories, principles for probing the purpose and relationships behind multimodal choices, and an avenue for translating findings into pedagogical applications. Employing a framework like Serafini's ensures that TikTok's widely popular yet insufficiently studied multimodal strategies are thoroughly analysed and converted into evidence-based, practically applicable principles for vocabulary instruction. Ultimately, a research-grounded framework enables the systematic decoding of TikTok's engaging multimodality to effectively address the established challenge of vocabulary retention.

### *1.3 Literature Review*

Serafini et al. (2022) carried out a study entitled "Multimodal Analysis of First-grade Students Reading *We Are in a Book!* The Reading Teacher." Using Serafini's framework, the study investigated the interaction and comprehension of eight first-grade students from the picture book '*We Are in a Book!*'. The research categorised multimodal elements such as images and text according to learning objectives, examined student responses, and evaluated how these modes aided comprehension. The findings shed light on students' application of gestures, gazes, and actions during reading to understand the book's humour and interactivity.

Serafini's framework was also employed in another study by Price and Clinton (2020), in which the researchers examined how an 11th-grade teacher adapted graphic novels for teaching purposes. The researchers identified learning goals, noted textual and visual components, evaluated navigation, and gauged effectiveness for literary analysis. The results revealed the teacher's implementation of multimodal scaffolding within the adapted graphic novels to support students' literacy development.

Studies have also applied Machin and Mayr's framework systematically to evaluate multimodal features in pedagogical materials and derive insights to inform teaching and learning. López (2018) conducted a study specifically aimed at analysing multimodal metaphors used in advertising. Using Machin and Mayr's framework, this research scrutinised multimodal metaphors present in print advertisements. The author identified and analysed semiotic modes such as images, text, layout, and interplay. The findings revealed how advertisers utilise multimodal metaphors to frame their products positively. Similarly, Vásquez (2014) in his study titled "Usually not one to complain but..." had applied Machin and Mayr's approach to explore identity construction in online consumer reviews. Multimodal elements such as usernames, profile pictures, and textual attributes were examined. The study uncovered how reviewers linguistically frame their identities to the companies and products being reviewed.

These instances demonstrate the versatility of Serafini's (2010) and Machin and Mayr's (2012) frameworks in analysing the potential for meaning-making within multimodal choices across genres. This approach equips researchers with tools to methodically investigate modes, interplay, contexts, and the resulting meanings, thereby offering constructive insights into communication strategies.

### *1.4 Objective of the Study*

The objective of this study is to evaluate the relative effectiveness of two analytical frameworks, Serafini's (2010) and Machin and Mayr's (2012), within the context of analysing TikTok videos for teaching vocabulary in English as a second or foreign language. The aim is to determine which framework provides a more inclusive and insightful approach for exploring the potential of TikTok videos in enhancing English learners' vocabulary acquisition.

A qualitative inquiry employing textual analysis methodologies is designed to address two key research questions that underpin this investigation: (1) to identify the multimodal elements embedded within TikTok videos, and (2) to analyse how these multimodal elements can enhance students' learning experiences. Multimodality encompasses a wide range of modes, spanning textual, auditory, and spatial elements. The effective adaptation and utilisation of these multimodal components are paramount for maximising the potential of TikTok videos in facilitating English vocabulary learning and enriching the effectiveness of conveying linguistic and cultural messages to learners.

## **2. Method**

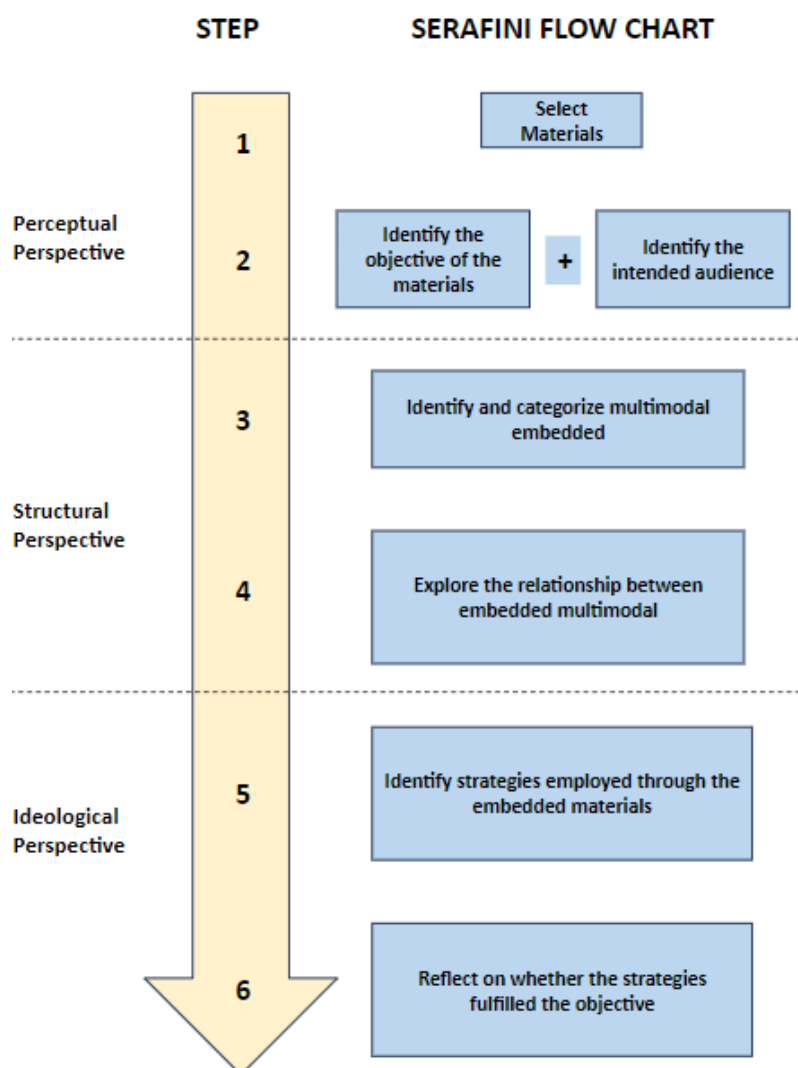
This study has chosen two frameworks to analyse the collected data to answer the critical questions of this study. Serafini's and Machin and Mayr's frameworks were used to analyse a TikTok video to address the research questions. These frameworks are employed to elucidate and further explore the analysed data by examining how these multimodal can affect learners' vocabulary enhancement.

Several prominent theories in language research focus on multimodality analysis, each with distinct factors. Therefore, it is vital that accurate theory is adapted to ensure that the findings of these designed studies accurately reflect and address the research questions. For example, semiotic theory, advocated by key figures like Ferdinand de Saussure and Charles Sanders Peirce (Ibrahim & Sulaiman, 2022), is renowned for its examination of multimodality. Although this theory is used to analyse multimodality, it is mostly used to investigate how discourse influences society by interpreting the multimodality, with key concepts such as signified (modes) and signifier (embedded meanings) (Siregar & Yahaya, 2022). Nevertheless, this theory is not suited to the current study as it is unlikely to provide precise answers. Instead, the theories of Serafini and Machin and Mayr align more closely with the objectives of the present study. Thus, a pilot study was conducted to offer a comprehensive comparison of these frameworks to determine the most suitable one to be adapted for this research.

### *2.1 Serafini's (2010) Framework*

In 2010, Frank Serafini developed a framework for assessing multimodal pedagogical resources that was particularly designed to evaluate the efficiency of teaching materials such as books, instructional videos, and educational games. This framework looks at the integration of multimodal elements within these materials in facilitating a smooth teaching

and learning experience for educators. Serafini et al. (2022) conducted a study involving eight first-grade students in the United States. These students were instructed to read aloud from a storytelling book titled "We Are in A Book!", with specific pages marked using coloured sticky notes. Each student read individually, while a camera recorded their behaviours, including body language. The study's findings provided valuable insights for educators, which comprises the creation of a table to document students' reactions during class activities and to identify the effectiveness of the modes within the pedagogical materials.



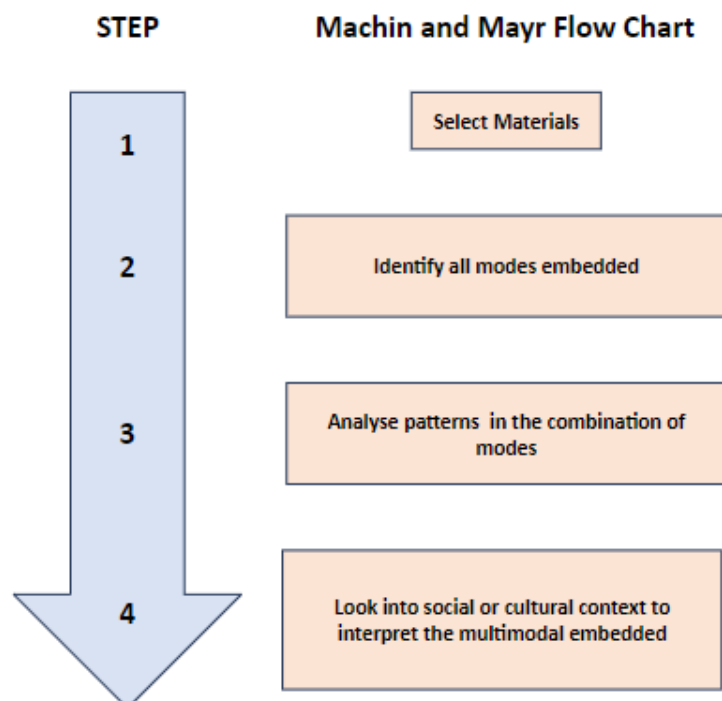
**Figure 1. Steps Taken in Serafini's (2010) Framework**

The analysis of the gathered data involves several layered steps, as Figure 1 illustrates. Firstly, the researcher needs to identify the objectives of the materials and their intended audience. Questions regarding the learners' demographics and characteristics are asked to determine the suitability of the chosen materials for the target learners. Subsequently, once the educational

materials are chosen, the researcher proceeds to identify and categorise the embedded multimodal elements, including visual images, colours, and auditory components such as spoken words. Thirdly, the researcher explores the relationship between these embedded multimodal elements, examining how they deliver information and contribute to achieving the educational objectives. Fourthly, the researcher evaluates whether any strategies, such as colour coding, labeling, or cues, are employed in the materials to enhance the learning process. The researcher then analyses the roles played by these strategies in facilitating student learning. Finally, the researcher reflects on whether the embedded multimodal elements effectively fulfill the learning objectives.

## 2.2 Machin and Mayr's (2012) Framework

In contrast to Serafini's (2010) exclusive focus on studying multimodal elements within pedagogical materials, Machin and Mayr's (2012) theory provides a broader applicability, relevant to a larger range of discourses or materials. As described by Fairclough (2013, p. 9) in his book, critical discourse analysis involves "the process of analyzing linguistic elements to reveal connections between language, power and ideology that are hidden from the people". Consequently, this theory elucidates the meanings conveyed through various modes within a discourse, shedding light on how ideologies are constructed and represented. Machin and Mayr underscore the importance of considering social and cultural contexts, types, the integration of multimodal elements, and how the modes were designed to convey information.



**Figure 2. Steps Taken in Machin and Mayr's (2012) Framework**

Figure 2 demonstrates the steps involved in Machin and Mayr's framework. After completing the material selection process (which does not confine researchers to pedagogical materials,



unlike Serafini's theory), the researcher proceeds to identify all embedded modes within the materials, such as images, colours, and speech intonation. Next, the researcher analyses patterns in how these modes are combined to convey a message or information. Following this, the researcher takes into account how society understands a communication medium in order to contextualise the encoded messages within social or cultural settings. For instance, the researcher investigates consumers' responses to the colour red, which has distinct cultural connotations, for instances, American in USA usually perceive red as "danger", French in France perceive red as "aristocracy", and Chinese see red as "happiness" (Tursunovich, 2022). This stage allows the researcher to analyse potential interpretations of the discourse by the readers.

### *2.3 Sampling Procedures*

The samples for this study were selected using a purposive sampling technique. This technique relies on the researcher's discretion in choosing samples deemed appropriate (Sharma, 2017). One advantage of this method, according to Rai and Thapa (2015), is that it gives the researcher the ability to choose samples based on specific demographic characteristics of the population relevant to the research questions. In this study, the researchers established two criteria for selecting samples, and all chosen samples met the following criteria:

- a. The content creator's TikTok account must have followers exceeding 400K.
- b. The video posted focuses on teaching and learning English vocabulary.

The chosen TikTok content creators have indicated in their profiles that they were teaching English for communication purposes. The TikTok chosen for this study has a high number of followers, increasing the likelihood of its videos appearing on the For You Page (FYP) and, therefore, having higher possibilities of attracting a broader audience.

### 2.3.1 Research Design

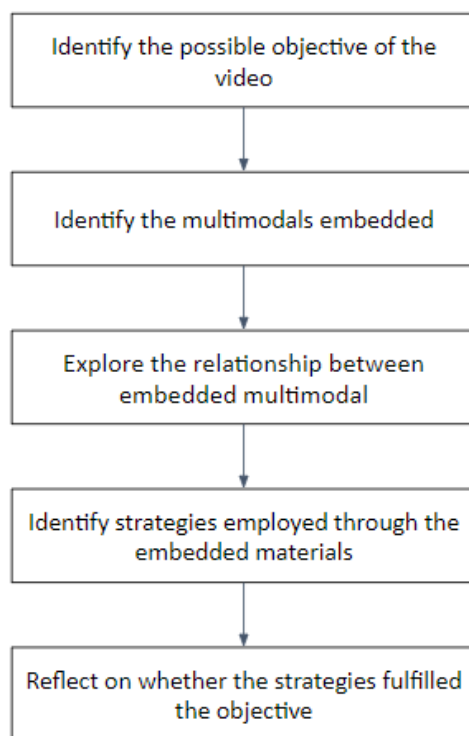




Figure 3. Data Analysis Procedures



Figure 3 illustrates the procedures undertaken to gather and analyse the collected data on multimodality from the TikTok video. The first step in this procedure is for the researchers to identify all embedded multimodal components, present in the videos such as pictures, text, and audio. The multimodal analyses conducted in this study encompass gesture, gaze, framing, fonts, colours, and spoken words. The researchers then examine the three theoretical analyses proposed by Serafini (2010); perceptual (identifying multimodal), structural (relationship between multimodal elements and the strategies employed), and ideological (whether the objective of the video is achieved by examining viewers' responses) as well as those suggested by Machin and Mayr (2012) to examine the multimodal components within the media. Finally, the researchers compare the data acquired using Serafini's and Machin and Mayr's frameworks.



### 3. Results



The two frameworks—Serafini and Machin and Mayr—are compared in this pilot project, which focuses on the collection and processing of multimodal data. To effectively compare the two presented theories, one of the samples was analysed using both Serafini's (2010) and Machin and Mayr's (2012) frameworks. The chosen sample is an international video posted by @antonioparlati, with over 6.3 million followers that lasts 34 seconds. The flow chart was meticulously followed during the data collecting and analysis procedures. Table 1 below displays the comparative findings.

Table 1. Comparison of Analysis Between Serafini (2010) and Machin and Mayr (2012)

Serafini's Theory		Marchin and Mayr's Theory
To provide students' vocabulary enhancement.	Objective	Not studied
<b>Gesture</b>		
<p>The character raised his left hand to attract and maintain audience focus on the words that he mentioned. In this picture, the character mentions “very important”, therefore, he lifted his left hand as the word mentioned is written in the left side of the video.</p>	<p>(Left hand lifted)</p>  <p>(Right hand lifted)</p> 	<p>The character raised his left hand to attract and maintain audience focus on the words that he mentioned. In this picture, the character mentions “very important”, therefore, he lifted his left hand as the word mentioned is written in the left side of the video.</p>
<b>Spoken Words</b>		
Just read the words appear on the screen without any explanation	No picture	Just read the words appear on the screen without any explanation

Fonts		
<p>1 type</p> <p>The creator only used 1 type of font named “Arial”.</p>		<p>1 type</p> <p>The creator only used 1 type of font named “Arial”.</p>
<p>The combination of the three modals (<b>text written, spoken words and gestures</b>) are used to direct the flow of the teaching.</p>	<p><b>Strategies Identified</b></p>	<p>Not studied</p>
Colours		
<p>White (question and answer), dark blue (colour in the background) and green (creator’s tops and highlighted answer)</p>	<p>White</p>  <p>Dark Green</p>	<p>White (question and answer), dark blue (colour in the background) and green (creator’s tops and highlighted answer)</p>

	 <p>Very beautiful - exquisite          Very happy - ecstatic          Very long - extensive          Very fast - swift          Very important -          Very clean -          Very tall -          Very upset -          Very lazy -          Very fancy -          Very painful -</p> <p>Dark Blue</p> 	
<p>The use of <b>three different colours</b> is used to shift students' attention from question to answer swiftly.</p>	<p><b>Strategies Identified</b></p>	<p>Not studied</p>
<p><b>Gaze</b></p>		

<p>Directly to the camera</p> <p>The character's gaze is seen to look straight to the camera throughout the video.</p> <p>Here we can assume that he is trying to create a connection between him and the audience by creating a scenario of a one-to-one conversation.</p>		<p>Directly to the camera</p> <p>The character's gaze is seen to look straight to the camera throughout the video.</p> <p>Here we can assume that he is trying to create a connection between him and the audience by creating a scenario of a one-to-one conversation.</p>
<p><b>Framing</b></p>		
<p>Half body and close-up</p>		<p>Half body and close-up</p>
<p>The combination of <b>gaze and framing</b>. These two modes play an important role in not only in creating a bond between the audience and the speaker and at the same, urging the audience to answer the questions asked in the video.</p>	<p><b>Strategies Identified</b></p>	<p>Not studied</p>
<p>Yes</p>	<p>Objective fulfilled</p>	<p>Not studied</p>

#### 4. Discussion

Table 1 provides a detailed comparison of Serafini's and Machin and Mayr's theories, contrasting them for enhanced understanding. According to the findings derived from Serafini's theory, three strategies were found to be most effective in maximising the use of multimodal elements. These strategies are thought to be crucial for creating a more productive and favourable learning environment. The constant introduction of new terminology may cause student distraction, which is not something many educators desire. Therefore, educators need to use a variety of strategies to guide lessons and keep students' attention over the whole class. The first strategy employed in the video involves the integration of written text, spoken words, and gestures. The teaching process was led by this combination of modalities. The content producer in the video reads the phrases from left to right in a sequential manner, which can confuse viewers. To mitigate confusion, the speaker in the video uses spoken words and gestures to clarify and draw students' attention to the terms being taught. As a result, the content producer skillfully employed these three modes to navigate the learning session.

The second strategy identified is the combination of gaze and framing. These two modes are pivotal not just for connecting learners and speakers, but also in prompting the learners to respond to the questions posed in the video. Throughout the 36-second duration of the video, the content creator maintained his gaze in one direction, making direct eye contact with the camera and creating a sense of connection. Furthermore, the close-up shot framing in this example intensifies the connection. The combination of these two modes can be interpreted as a way to foster a sense of rapport between the educators and learners involved while instilling a sense of urgency and prompting the audience to stay attentive and participate in the discussion.

The third strategy used in this example was the use of three different hues (dark blue, white, and light green). The background lighting was dark blue, which had little effect on the background of the video. White was chosen as the font colour for both questions and answers. The visibility element is increased by the contrast between dark blue and white, which helps students stay focused on the lessons being taught. Finally, light green was used to accentuate the responses. It was observed that the content creator used this hue to set the questions and responses apart.

Finally, the researchers examined the ideological perspective, determining if the content creator succeeded in achieving the objective by integrating multimodal elements. As mentioned earlier, the content creator effectively employed multimodal elements to facilitate a smoother and easily accessible learning experience. However, the researchers believe that the content creators should give careful consideration to how they employ colour. While choosing green to highlight the answers is a good strategy, the green colour did not stand out since it was similar to the colour of the content creator's shirt. Despite slight differences in hue, the poor contrast between them reduced the visibility of the text. As a result, the response became less salient, which made it harder for students to understand the text. Nonetheless, by leveraging the use of multimodal elements to expand learners' vocabulary,

the content creator successfully managed in developing valuable pedagogical material. Machin and Mayr's framework, on the other hand, lays more of a focus on examining multimodal elements to comprehend the effects they have on culture and society. As a result, it is clear from the previously stated sample, that the video's content creator tried to create a less formal classroom setting by taking into account a variety of multimodal elements including background and lighting tone.

Typically, when we think of a learning environment, we image a well-lit classroom with conventional educational props like tables, instructional materials displayed on the walls, and whiteboards or blackboards. But even with the video's educational content, the background is devoid of these standard classroom components. The dim lighting in the background also differs from the atmosphere of a conventional classroom. The absence of familiar educational props and the unconventional lighting suggests that the content creator in the video aims to establish a less formal learning environment for the learners. By making the learning atmosphere more relaxed and pleasant, this method aims to help students enjoy the process of learning and improve their vocabulary at the same time.

## 5. Conclusion

In conclusion, this analysis compared and contrasted the theoretical frameworks of Serafini and Machin and Mayr in examining the use of multimodal elements in an educational video. Serafini's social semiotic perspective highlighted three key strategies the video creator employed to maximize the pedagogical benefits of multimodality: 1) integrating written text, spoken words, and gestures, 2) using gaze and framing to connect with learners, and 3) utilizing contrasting colors to increase visibility and salience.

Serafini's approach focused more on identifying specific strategies teachers can use to maximize the pedagogical effectiveness of multimodal elements in instructional videos. While Machin and Mayr concentrate more on analysing how these modalities are intended to convey embedded information and their relationship to social and cultural contexts, whereas Serafini emphasises maximising the utilisation of multimodal elements to enhance the effectiveness of learning materials. As a result, compared to Machin and Mayr's findings, Serafini's paradigm provides a more thorough examination of the TikTok videos' effectiveness of teaching materials.

## References

- Anderson, K., & Kachorsky, D. (2019). Assessing students' multimodal compositions: an analysis of the literature. *English Teaching*, 18(3), 312–334. <https://doi.org/10.1108/etpc-11-2018-0092>
- Baldry, A. P., & Thibault, P. J. (2006). *Multimodal transcription and text analysis: A multimedia toolkit and coursebook*, 1, 1-288. Equinox.
- Fairclough, N. (2013). *Critical discourse analysis: The critical study of language*. Routledge.
- Ibrahim, I., & Sulaiman, S. (2020). Semiotic communication: An approach of understanding a meaning in communication. *International Journal of Media and Communication Research*



(IJMCR), 1(1), 22-31.

Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.

Lemke, J. L. (2002). Travels in hypermodality. *Visual communication*, 1(3), 299-325.

López, X.A. (2018). Multimodal metaphor in advertising: A multimodal discourse analysis. *Review of Cognitive Linguistics*, 16(1), 207-226.

Machin, D., & Mayr, A. (2012). *How to do critical discourse analysis: A multimodal introduction*. SAGE.

O'Halloran, K. L. (2008). Systemic functional-multimodal discourse analysis (SF-MDA): Constructing ideational meaning using language and visual imagery. *Visual communication*, 7(4), 443-475.

Price, C. & Clinton, V. (2020). Implementing and adapting the Serafini framework. *Journal of Adolescent and Adult Literacy*, 41(2), 196-225. <https://doi.org/10.7560/LAMRA41203>.

Rai, N., & Thapa, B. (2015). *A study on purposive sampling method in research*. Kathmandu School of Law, 5.

Serafini, F. (2010). Reading multimodal texts: Perceptual, structural and ideological perspectives. *Children's Literature in Education*, 41(2), 85–104. <https://doi.org/10.1007/s10583-010-9100-5>.

Serafini, F., Gee, J., & Murphey, D. (2022). Multimodal analysis of first-grade students reading *We Are in a Book!* *The Reading Teacher*. 76(3), 384-395. <https://doi.org/10.1002/trtr.2268>

Sharma, K., & Giannakos, M. N. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, 51(5), 1450–1484. <https://doi.org/10.1111/bjet.12993>

Sharma, G. (2017). Pros and cons of different sampling techniques. *International Journal of Applied Research*, 3(7), 749-752.

Siregar, I., & Yahaya, S. R. (2022). Semiotic exploration of *Roti Buaya* as a cultural ornament. *British Journal of Applied Linguistics*, 2(1), 06-13. <https://doi.org/10.32996/bjal.2022.2.1.2>

Tursunovich, R. I. (2022). Linguistic and cultural aspects of literary translation and translation skills. *British Journal of Global Ecology and Sustainable Development*, 10, 168-173.

Vásquez, C. (2014). "Usually not one to complain but...": Constructing identities in user-generated online reviews. In C. Lassen, J. Strunck & T. Vestergaard (Eds.) *Mediating Ideology in Text and Image: Ten Critical Studies*, 65-88. John Benjamins.

## Acknowledgments

We want to thank Universiti Teknologi MARA for the assistance provided for this research.

**Funding**

Not Applicable.

**Competing interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Informed consent**

Obtained.

**Ethics approval**

The Publication Ethics Committee of the Macrothink Institute.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

**Provenance and peer review**

Not commissioned; externally double-blind peer reviewed.

**Data availability statement**

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

**Data sharing statement**

No additional data are available.

**Open access**

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.