

Multimodal Metadiscourse in Influencer Review Videos: A Case Study

Han Jikun^{1*}, Omer Hassan Ali Mahfoodh², Marlina Binti Jamal³

^{1,2,3} School of Languages, Literacies and Translation, Universiti Sains Malaysia (USM), 11800 Penang, Malaysia

*Corresponding author

DOI: <https://doi.org/10.47772/IJRISS.2026.100800871>

Received: 04 September 2026; Accepted: 09 September 2026; Published: 19 September 2026

ABSTRACT

Influencer review videos (IRVs) are a multimodal digital genre in which creators combine textual and visual resources to organize evaluations and address viewers. This case study examines the distribution and functions of textual and visual metadiscourse in 12 beauty-related IRVs from a YouTube influencer. The analysis combines quantitative and qualitative approaches, drawing on Hyland's textual metadiscourse model and De Groot et al.'s visual metadiscourse framework, with relevant concepts from Kress and van Leeuwen's visual grammar used to interpret visual realizations. Textual metadiscourse was identified with AntConc, whereas visual metadiscourse was manually annotated in ELAN by two coders. Both analyses were conducted by two coders, with Cohen's kappa indicating strong inter-coder agreement for textual ($\kappa = 0.97$) and visual coding ($\kappa = 0.92$). Textual frequencies were normalized to occurrences per 1,000 words, while visual frequencies were normalized to occurrences per minute of video footage. The qualitative analysis examined how the identified resources functioned in context. The findings show that textual metadiscourse is dominated by interactional resources, particularly self-mentions and engagement markers, while transitions are the most frequent interactive resource. Visual metadiscourse shows a more balanced distribution between interactive and interactional resources. Transitions, attitude markers, and boosters are particularly prominent, whereas visual evidentials and code glosses occur infrequently, and no visual hedges are identified. Functionally, textual and visual metadiscourse contribute to discourse organization, evaluation and stance, and viewer orientation. These functions underscore the significance of both textual and visual metadiscourse in IRVs. The study extends metadiscourse research into multimodal communication and the genre of IRVs, while also exploring the applicability of a refined visual metadiscourse framework to IRVs.

Keywords: multimodal metadiscourse, textual metadiscourse, visual metadiscourse, influencer review videos

INTRODUCTION

Early conceptualizations describe metadiscourse as writing about writing (Williams, 1981), discourse about discourse (Vande Kopple, 1985), and text about text (Mauranen, 1993). These perspectives underscore that discourse not only conveys information but also guides interpretation and evaluation. Building on these earlier views, Hyland (2005, p. 37) defined metadiscourse as "the cover term for the self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community".

Metadiscourse research has predominantly focused on academic genres (Pearson & Abdollahzadeh, 2023). However, non-academic and digitally mediated genres remain comparatively underexplored. Among emerging digital genres, influencer-generated content has become an important form of online communication, often combining personal experience and audience engagement through social media platforms (Abidin, 2016). In particular, influencer review videos (IRVs) provide a relevant context for

examining how creators present evaluations and engage viewers (Hallinan, 2023).

As a multimodal genre, IRVs combine textual and visual resources in the construction of meaning and viewer engagement (Indrawati et al., 2023). In such multimodal environments, visual resources play an important role in presenting information, expressing evaluations, and guiding viewer interpretation alongside textual resources (Kress & van Leeuwen, 2020). Extending this perspective to the metadiscursive functions of visuals, Kumpf (2000) introduced the concept of visual metadiscourse, highlighting how visual elements can organize information, guide interpretation, and shape readers' engagement with texts. Subsequent studies have increasingly examined textual and visual metadiscourse together across multimodal genres, including academic posters (D'Angelo, 2016), digital academic discourse (Carrió-Pastor, 2021), social-media advertising (Al-Subhi, 2022), and infographics (Li et al., 2023). These studies demonstrate the applicability of multimodal metadiscourse analysis across different communicative settings.

Despite this growing attention to textual and visual metadiscourse, little is known about how these resources are distributed and function in IRVs. Therefore, this study addresses the following research questions:

- (1) What is the distribution of textual metadiscourse in the selected influencer review videos?
- (2) What is the distribution of visual metadiscourse in the selected influencer review videos?
- (3) How do textual metadiscourse resources function in the selected influencer review videos?
- (4) How do visual metadiscourse resources function in the selected influencer review videos?

LITERATURE REVIEW

Influencer Review Videos

Influencer review videos (IRVs) are a form of social-media content produced by influencers and shared on platforms such as YouTube, TikTok, and Instagram. These videos combine product demonstration and evaluation with personal experience and narrative (Torjesen, 2021). Although influencer-generated content has been studied in media and marketing research, these review videos have only recently attracted attention in linguistics and discourse analysis (Hallinan, 2023). As video-based content, these reviews are inherently multimodal, as spoken commentary is coordinated with embodied and visual resources such as gestures, gaze, and facial expressions, enabling creators to assess products while addressing an absent but interactionally constructed audience (Gibson, 2020). These intertwined evaluative and multimodal features make IRVs a productive site for examining how textual and visual resources organize discourse and engage viewers.

Multimodal Metadiscourse

Metadiscourse research has traditionally focused on language. However, digital communication increasingly involves multiple semiotic modes, with meaning emerging through the interaction of language, image, gesture, layout, and other resources (O'Halloran, 2021). Existing multimodal metadiscourse studies have examined the interaction between textual and visual resources across a range of genres (e.g., Carrió-Pastor, 2021; Al-Subhi, 2022; Li et al., 2023). Across these studies, multimodal metadiscourse is operationalized primarily through textual and visual resources. In this study, multimodal metadiscourse refers to metadiscursive functions realized across textual and visual modes.

Multimodal Metadiscourse in Digital Contexts

Empirical studies show that meaning-making often relies on the interplay of textual and visual modes. Bernad-Mechó (2017) examined university lectures, showing that verbal metadiscursive resources are

coordinated with embodied resources such as gestures in the organization of discourse. Carrió-Pastor (2021) identified disciplinary variation in digital academic journals, with engineering articles relying more on visual metadiscourse and linguistics articles more on textual resources. Al-Subhi (2022) found that linguistic and visual metadiscourse work together to organize promotional messages and engage viewers. Li et al. (2023) showed that textual and visual resources jointly contribute to information guidance and reader engagement in student-produced infographics.

These studies indicate that multimodal metadiscourse varies across genres and contexts. However, multimodal metadiscourse remains a relatively limited area of research, with existing studies concentrated largely in academic discourse and comparatively less attention paid to other communicative contexts (Dłóciok et al., 2026). Little is known about how textual and visual metadiscourse are distributed and function in IRVs. Given the genre-specific patterns identified in previous multimodal metadiscourse research, examining this genre can extend the understanding of metadiscourse in dynamic, video-based digital communication.

METHOD

Corpus Collection

The present corpus (Table 2) consists of 12 influencer review videos, totaling 41,884 tokens and approximately 199.9 minutes of video footage. The selected videos primarily involve full-face makeup reviews covering multiple beauty-product categories, including foundation, concealer, powder, blush, eyeshadow, mascara, and lip products. As a case study of influencer review discourse, all videos were collected from James Charles' YouTube channel. YouTube was selected because it is the most visited video platform worldwide. James Charles was purposively selected because he has the highest YouTube subscriber count among the beauty influencers considered, a large international audience, and an active channel with a substantial body of beauty-related review content.

Table 1. Characteristics of the Selected Influencer Review Videos

Duration	Word count	Topic	URL
17'34"	3650	Full face of the most expensive makeup products	https://www.youtube.com/watch?v=gkuaRMZIOdk&t=145s
14'49"	3215	Full face of Selena Gomez's rare beauty	https://www.youtube.com/watch?v=e5Ob9b6Yza4&t=254s
17'00"	3591	Full face of beauty influencer makeup brands	https://www.youtube.com/watch?v=TG_RS6GvA9w&t=517s
20'23"	3871	I slept in my makeup and here's what happened	https://www.youtube.com/watch?v=xk6l_jn6gco&t=174s
15'10"	2827	Trying a full face of makeup from 5 below	https://www.youtube.com/watch?v=dvQ6jKvCRrs&t=426s
17'55"	3672	Trying a full	https://www.youtube.com/watch?v=ETwQxDx3hcY&t=351s

		face of makeup from forever 21	
12'54"	2567	The world's cheapest makeup store	https://www.youtube.com/watch?v=HbRo5HdOAE&t=545s
17'16"	3651	Trying a full face of Korean makeup	https://www.youtube.com/watch?v=VX2uRreCG5Y&t=117s
15'36"	3534	Full face of expired makeup	https://www.youtube.com/watch?v=GYCtRBE-gjw&t=157s
17'35"	3470	This brand says you can sleep in their makeup	https://www.youtube.com/watch?v=yHke1ZaCIRQ&t=452s
16'27"	3468	Full face of the world's most luxurious makeup	https://www.youtube.com/watch?v=8rzRFNk3Vaw
17'15"	4368	Basic but still cute ASF makeup looks	https://www.youtube.com/watch?v=XOKAL39mQeI&t=228s

The videos were selected according to the following criteria: (1) the videos focus on beauty-related products and include product evaluation or demonstration; (2) each video exceeds 10 minutes in length to provide sufficient discourse for analysis; (3) videos with available captions were included to facilitate accurate transcription; and (4) all videos had to be presented in monologue format to maintain consistency in discourse structure across the corpus. All selected videos were published in 2023, with one video selected from each month to maintain temporal consistency within the dataset. For transcription, Lark was used to generate the initial transcripts, providing automated speech recognition and text-editing functions. The generated transcripts were subsequently verified manually against the original videos to ensure the accuracy of the verbal data.

Corpus Analysis

This case study adopts a corpus-based discourse approach to analyze textual and visual metadiscourse in selected beauty-related IRVs. The analysis was conducted in two main phases: textual and visual metadiscourse analysis, each involving coding, reliability checking, and interpretation. Prior to coding, two coders participated in a training session to familiarize themselves with the categories and functions of textual and visual metadiscourse.

Textual Metadiscourse Analysis

In the first phase, textual metadiscourse markers were coded following Hyland's interpersonal model (Table 1). Hyland's model was selected for its wide use and systematic classification of interactive and interactional metadiscourse. AntConc's keyword list was first used to identify frequently occurring lexical items within the corpus. These items were subsequently examined in context, and those functioning as metadiscourse markers were assigned to the corresponding categories. For example, expressions such as "stunning," "beautiful," and "oh, my god" were categorized as attitude markers when they expressed the influencer's evaluation of products, while audience-oriented expressions such as "you guys" were identified as engagement markers when they directly addressed viewers.

This process allowed the marker list to incorporate genre-specific realizations. To refine the adapted marker list, two IRVs were independently coded by the two coders, following the 10%–25% range

recommended by O'Connor and Joffe (2020). Using the refined marker list, all potential instances were retrieved through AntConc's concordance tool for manual verification.

Table 2. Hyland's Interpersonal Model of Textual Metadiscourse

Category	Function	Example
Interactive	Help to guide the reader through the text	Resources
Transitions	express relations between main clauses	in addition; but; thus
Frame markers	refer to discourse acts, sequence or stages	finally; to conclude
Endophoric markers	refer to information in other parts of the text	noted above; see Fig.
Evidentials	refer to information from other texts	according to X; Z states
Code glosses	elaborate propositional meanings	namely; e.g.; in other words
Interactional	Involve the reader in the text	Resources
Hedges	withhold commitment and open dialogue	might; perhaps; possible
Boosters	emphasize certainty or close dialogue	in fact; it is clear that
Attitude markers	express writer's attitude to proposition	unfortunately; surprisingly
Self-mentions	explicit reference to author(s)	I; we; me; our
Engagement markers	explicitly build relationship with reader	consider; you can see that

The retrieved items were independently reviewed by the two coders, who annotated all instances separately. Their annotations were compared, and inter-coder reliability was assessed using Cohen's kappa ($\kappa = 0.97$), indicating strong agreement between coders. Any remaining disagreements were resolved through discussion until consensus was reached. The occurrences of textual metadiscourse markers were calculated, and all textual frequencies were normalized to occurrences per 1,000 words to facilitate comparison among textual metadiscourse categories.

Visual Metadiscourse Analysis

In the second phase, visual metadiscourse was initially coded based on De Groot et al.'s (2016) framework. As in the textual analysis, two IRVs were first coded independently by the two coders to examine the applicability of the initial framework to the present corpus.

All visual annotations were conducted using ELAN. Based on this preliminary coding, relevant concepts from Kress and van Leeuwen's (2020) visual grammar were incorporated where necessary. The framework was then further refined throughout the coding of the full dataset as additional visual patterns emerged. This iterative process continued until no further adjustments were required, resulting in the final framework presented in Table 3.

Table 3. Refined Analytical Framework for Visual Metadiscourse in Influencer Review Videos

Category	Function	Resources
Interactive	help to guide the reader through the text	
Transitions	express semantic links between sections to create continuity in product presentation	overlap; integration; visual rhyme

Frame markers	distinguish sections among products or content	visual contrast; segregation; separation
Endophoric markers	interrelate images with verbal elements	referent video clip; referent photograph; referent image; arrows referent timestamp; referent location
Evidentials	refer to additional sources.	external video clip; external photograph;
Code glosses	elaborate or clarify meanings by visualizing discourse	illustrative video clip; illustrative photograph
Interactional	involve the reader in the text	
Hedges	no visually realized	—
Boosters	intensify certainty and foreground elements as especially important	emphatic close shot; emphatic text; foreground-background
Attitude markers	express the influencer's attitude toward the product or message	color; affective gesture; facial expression; affective emoji; affective video clip
Self-mentions	refer to the sender(s) of the discourse	self-directed gesture; self-display posture; influencer's photograph/video clip/name/brand/platform
Engagement markers	build relationship with viewers	viewer-directed gesture; attention-getting emoji; viewer address text; interactive video clip; viewer-directed close shot

To ensure consistent coding, the nine visually realized categories are distinguished according to their metadiscursive functions in context. Transitions are coded when visual resources link separate segments or elements to create continuity, whereas frame markers indicate boundaries between sections or content. Endophoric markers link verbal references to corresponding visual elements. Evidentials introduce external visual sources as supporting information, while code glosses visually clarify, exemplify, or elaborate verbal content. Boosters increase the visual salience or emphasis of a product or message, whereas attitude markers express evaluative or affective meanings. Self-mentions visually foreground the influencer's identity or presence, while engagement markers visually orient the discourse toward viewers through viewer-directed resources. This function-based approach also helped distinguish visually similar resources with different functions. For example, an emphatic close shot was coded as a booster when the camera moved closer to a product or the visible result of product application. In contrast, a viewer-directed close shot was coded as an engagement marker when the camera moved closer to the influencer or a product being directly presented to viewers. These distinctions show that coding decisions were based on the metadiscursive function of the visual resource in context rather than on visual form alone.

Once the framework was finalized, all annotations were reviewed to ensure consistency with the final

coding scheme. Inter-coder reliability for the full dataset was then assessed using Cohen's kappa ($\kappa = 0.92$), indicating strong agreement. For each subcategory of visual metadiscourse, frequencies were calculated and normalized to occurrences per minute of footage, together with percentages, to facilitate comparison among visual metadiscourse categories.

RESULT

Distribution of Textual Metadiscourse

Table 4 presents the frequency distribution of textual metadiscourse resources identified in the corpus. The findings show that interactional resources occur considerably more frequently than interactive resources. This distribution shows a clear quantitative predominance of interactional resources in the corpus. For interactive resources, transitions are the most frequently occurring category, followed by frame markers, while endophoric markers, evidentials, and code glosses occur relatively infrequently. Among interactional resources, self-mentions and engagement markers show the highest frequencies, whereas attitude markers, boosters, and hedges occurred less frequently.

Table 4. Frequencies of Textual Metadiscourse in a Beauty Influencer's Review Videos

Metadiscourse	Occurrence	F per 1000 words	Percentage
Interactive resources	1638	39.11	22.44%
Transitions	1160	27.70	15.89%
Frame markers	452	10.79	6.19%
Endophoric markers	6	0.14	0.08%
Evidentials	5	0.12	0.07%
Code glosses	15	0.36	0.2%
Interactional resources	5662	135.18	77.56%
Hedges	338	8.07	4.63%
Boosters	593	14.16	8.12%
Attitude markers	714	17.05	9.78%
Self-mentions	2328	55.58	31.89%
Engagement markers	1689	40.33	23.14%
Total	7300	174.29	100%

Distribution of Visual Metadiscourse

Table 5 presents the frequency distribution of visual metadiscourse resources identified in the corpus. Interactional resources occur slightly more frequently than interactive resources, accounting for 55.06% of all visual metadiscourse instances, compared with 44.94% for interactive resources. This distribution shows a relatively balanced quantitative pattern between the two broad categories. Among interactive resources, transitions represent the most frequently occurring category, followed by endophoric markers and frame markers. In contrast, code glosses and evidentials appeared relatively infrequently. Unlike the interactive resources, whose frequencies vary considerably across categories, interactional resources show a more balanced distribution: attitude markers lead, boosters follow closely, then engagement markers, with self-mentions least frequent. Hedges are not identified in the corpus.

Table 5. Frequencies of Visual Metadiscourse in a Beauty Influencer's Review Videos

Metadiscourse	Occurance	F per minute	Percentage
Interactive resources	923	4.62	44.94%
Transitions	382	1.91	18.60%
Frame markers	176	0.88	8.57%
Endophoric markers	280	1.40	13.63%
Evidentials	27	0.14	1.31%
Code glosses	58	0.29	2.82%
Interactional resources	1131	5.66	55.06%
Boosters	292	1.46	14.22%
Attitude markers	380	1.90	18.50%
Self-mentions	214	1.07	10.42%
Engagement markers	250	1.25	12.17%
Total	2054	10.27	100%

Functions of Textual Metadiscourse

Textual transitions in IRVs

Transitions organize the progression of product evaluation in the IRVs. These markers help establish logical relationships between product descriptions, comparisons, and evaluations, allowing the influencer to present product-related judgments in a coherent manner throughout the review. In the corpus, transitions are predominantly realized through common conjunctions, such as *and*, *but*, and *so*, which establish additive, contrastive, and consequential relationships between different stages of product assessment.

(1) We definitely have like the pinks in there, but it seems like these matte Brown, warm neutral colors are definitely more of the vibe. So, I think that we're gonna go with this palette. (Video 8)

Textual frame markers in IRVs

Frame markers signal shifts between different stages of the review process. In the corpus, markers such as *alright*, *now*, and *then* are frequently used to introduce new products, move between review activities, or indicate changes in the focus of discussion. These markers exhibit strong colloquial characteristics and are commonly found in spoken discourse (Macário Lopes & Amaral, 2006).

(2) So if you're excited for today's video, please go ahead and give it a big thumbs up down below. I would love to make my money back. Alright, you guys, for product No. 1, we're starting off with a bang. (Video 1)

Textual endophoric markers in IRVs.

Endophoric markers refer viewers to other parts of the review. These markers, such as *later on* and *by the end of the video*, help the influencer direct viewers' attention to other moments or stages within the ongoing video, thereby supporting the coherence of the review. An example is provided below:

(3) Or you could also use Shamps or Cloud 9 if you want a more pinky tone that's gonna match with the blush. I'm gonna use a little bit of Shamps for that pinky tone just because I'll show you guys Finnegan on

the eyes later on. (Video 12)

Textual evidentials in IRVs.

Evidentials mainly introduce external voices or sources to support the influencer's statements and evaluations. In the data, they are typically realized through source attribution structures such as “ ‘...’ [name]” or “.... — [name],” which are further categorized into expert and peer attributions.

Anyway, starting off, our first step is always going to be primer. So we're going to go in with the Nimya by Nikkie Tutorials License to Glow Serum because Nikkie always says 'not to prime is in fact a crime'. (Video 3)

Textual code glosses in IRVs.

Code glosses mainly function to clarify and elaborate the influencer's descriptions and evaluations, making product-related information more accessible to viewers. They are mainly realized through markers such as I mean and for example, as well as occasional definitional cues such as called, that is, and known as. An example is provided below:

The packaging does look gorgeous. I mean, it's a nice, like very heavy glass component. (Video 1)

Textual hedges in IRVs.

Hedges function to soften evaluations and express claims with caution in IRVs. They allow the influencer to present judgments as personal observations or possibilities rather than absolute claims. This cautious use of language may also reflect the interactive nature of digital platforms, where audience-oriented communication can encourage more careful expressions (Zou & Hyland, 2022). Here is an example.

(6) I love getting a gorgeous full beat face of makeup for not a lot of money, but with the cheaper prices sometimes comes lower quality products. (Video 7)

Textual boosters in IRVs.

Boosters function to strengthen the force of the influencer's evaluations and present judgments with a higher degree of certainty. Through intensifying expressions such as really, definitely, never, obviously, and actually, the influencer emphasizes personal assessments and conveys confidence in product-related evaluations.

Look at that. I mean, such a simple eye. Obviously, this isn't anything crazy or evolutionary, but it is really impactful. (video 11)

Textual attitude markers in IRVs.

Attitude markers function to express the influencer's evaluations and emotional responses toward products and experiences. In the corpus, attitude markers are mainly realized through evaluative adjectives, attitude verbs, and sentence adverbs. Positive valuations are frequently expressed through items such as love, good, beautiful, and amazing, while expressions such as oh my god and honestly further emphasize the influencer's affective stance. These resources help construct a personal and emotionally engaging evaluation of products.

(8) Oh, it smells amazing. This smells so good. It's like minty. Oh my God. I literally love that. Wait, this smells so good. (Video 5)

Textual self-mentions in IRVs.

Self-mentions foreground the influencer's personal presence and perspective. Singular forms (I, my, me) occur much more frequently than plural forms (we, our). This pattern reflects the persona-centered nature of the IRVs, where product evaluations are often grounded in the influencer's personal experiences and viewpoints. An exception is observed in Video 12, where plural self-mentions (we, our, us) are used when referring to the influencer's own brand. These forms construct a collective voice representing the brand and its associated team (Li, 2022). (9) Now this product I was actually very shocked but also impressed to find on my luxury makeup list. (Video 11)

Textual engagement markers in IRVs.

Engagement markers involve viewers directly and orient the discourse toward the audience. Direct address forms such as you and you guys are particularly prominent, positioning viewers as immediate addressees and co-participants. Directives, including inclusive invitations such as let's and guidance expressions such as look at, make sure, have to, and should, further invite viewers to follow the influencer's actions and evaluations. Questions mainly function rhetorically rather than seeking actual responses, positioning viewers toward alignment with the influencer's perspective (Janoschka, 2004). Personal asides, such as by the way, occur less frequently and contribute only minimally to viewer engagement. Here is an example: (10) Do you see how we kind of have this vibe going on out here? This is what we're looking for. (Video 12)

Functions of Visual Metadiscourse

Visual transitions in IRVs.

Visual transitions maintain visual continuity and connect separated elements across the review. Following Kress and van Leeuwen's (2020) concept of visual connection, transitions are operationalized through three strategies: overlap, integration, and visual rhyme. Overlap occurs when an element crosses or extends beyond a boundary. Integration refers to the combination of multiple elements within a shared visual space. Visual rhyme involves the repetition of similar visual features across separated elements. In the dataset, these strategies reflect the dynamic and sequential nature of IRVs. Overlap is operationalized as the reappearance of the same product, for example in both display and application, which visually links distinct moments and creates cohesion across scenes (Figure 1(a)). Integration is realized through visual consistency across shots, such as matched opening and closing frames or unified scenes (Figure 1(b)). Visual rhyme involves repeating the formatting of contextual information, including product names, time, and location, with consistent styles across segments (Figure 1(c)).

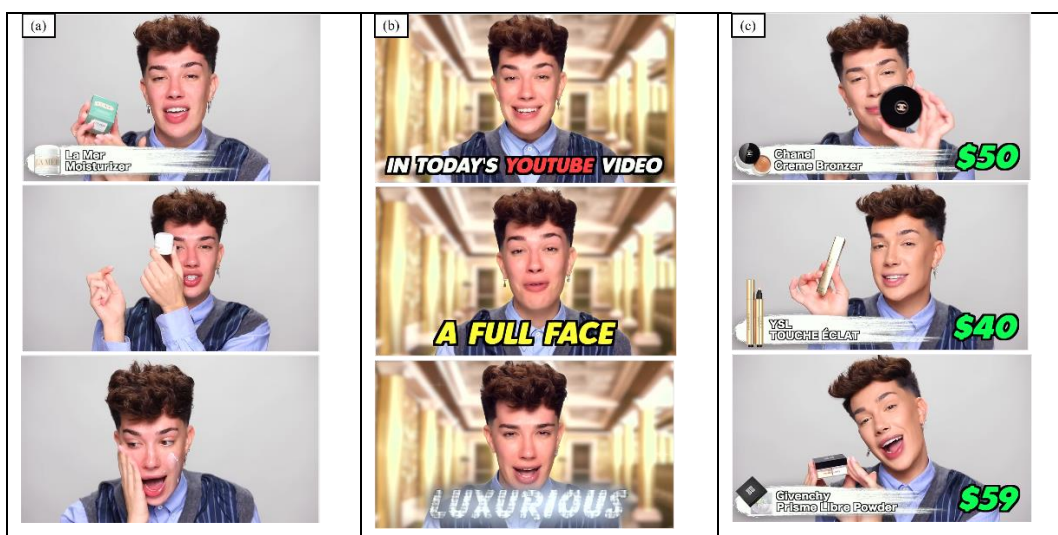


Figure 1. Examples of Visual Transitions: (a)-(c) Video 1

Visual frame markers in IRVs.

Visual frame markers establish boundaries and segment visually distinct parts of the review. Drawing on Kress and van Leeuwen's (2020) concept of visual disconnection, they are operationalized through three compositional strategies: segregation, separation, and visual contrast. Segregation uses visible frame lines or depicted structures to delimit units. Separation relies on empty space between components to signal functional distinctness. Visual contrast highlights difference through variations in visual or typographic features.

In the dataset, these strategies reflect the dynamic organization of IRVs. Segregation is realized through inserting short segments that function as visual dividers between sections (Figure 2(a)). Separation is realized through disappear-and-reappear motions that create brief pauses between sections (Figure 2(b)). Visual contrast is realized through changes of scene, including shifts in background or location. These contrasts often function rhetorically to mark different stages of the video, such as introduction, demonstration, and conclusion (Figure 2(c)).

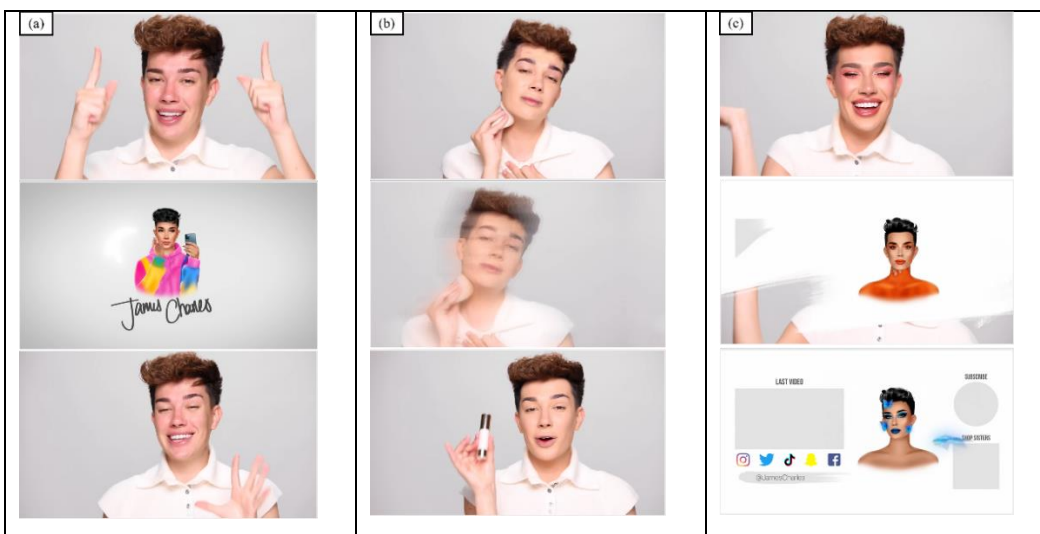


Figure 2. Examples of Visual Frame Markers: (a)-(c) Video 2

Visual endophoric markers in IRVs.

Visual endophoric markers connect verbal references with corresponding visual elements to construct cross-modal cohesion (De Groot et al., 2016). In the corpus, they take six forms, illustrated respectively in Figures 3(a)–(f): referent images, the product displayed when it is named; referent photographs, a photograph of the person shown when mentioned; referent video clips, an embedded clip from a previously posted video referenced by the influencer; timestamps, a time label shown when the time is mentioned; arrows, a graphic arrow indicating the region being referenced; and location references, an on-screen label naming the place when mentioned.

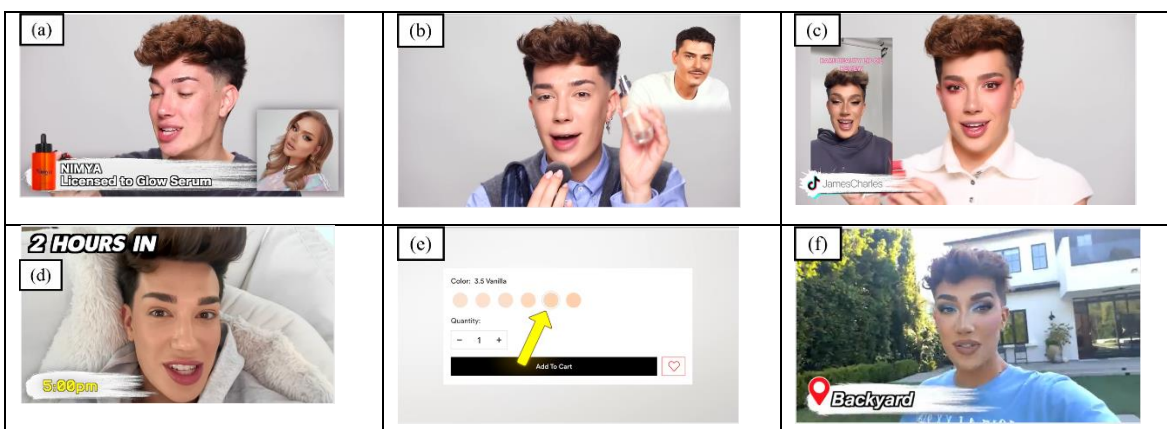


Figure 3. Examples of Visual Endophoric Markers: (a) Video 3; (b) Video 1; (c) Video 2; (d) Video 10; (e) Video 8; (f) Video 4

Visual evidentials in IRVs.

Visual evidentials introduce external visual material to support the influencer's statements and evaluations (Cruz García, 2017). In the corpus, they are mainly realized through external photographs and video clips. Figure 4(a) shows an embedded externally sourced clip that the influencer references when describing the viral "powder technique" video by Alex Earl, attributing the method to an outside source. Figure 4(b) overlays manufacturer images of Too Faced palettes to support the claim that "they're all the same."



Figure 4. Examples of Visual Evidentials: (a) Video 4; (b) Video 7

Visual code glosses in IRVs.

Visual code glosses clarify or elaborate verbal information through corresponding visual content (Ma & Jiang, 2025). In the corpus, they are mainly realized through illustrative photographs and video clips. Figure 5(a) shows a photograph accompanying the description of "Korean-style brows," providing a concrete visual illustration of the verbal description; Figure 5(b) presents a short clip following the comment "This is very millennial core...", visually exemplifying the influencer's characterization.

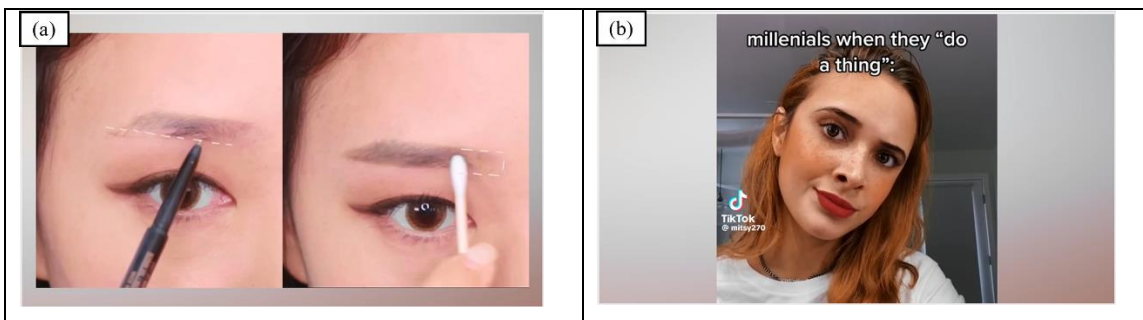


Figure 5. Examples of Visual Code Glosses, (a) Video 8; (b) Video 10

Visual boosters in IRVs.

These resources strengthen the visual emphasis of particular products, features, or information in IRVs. Drawing on Kress and van Leeuwen's (2020) notions of salience and distance, which explain how visual achieve prominence, three realizations are identified in the dataset: emphatic close shot, foreground-background composition, and emphatic on-screen text.

Emphatic close shots bring products or product features closer to viewers, foreground-background composition directs attention to selected details, and emphatic on-screen text increases the prominence of key information. Figure 6(a) shows a close shot foregrounding color; Figure 6(b) highlights the brow-pencil tip through foreground-background composition; and Figure 6(c) uses prominent on-screen text to emphasize the brand name.

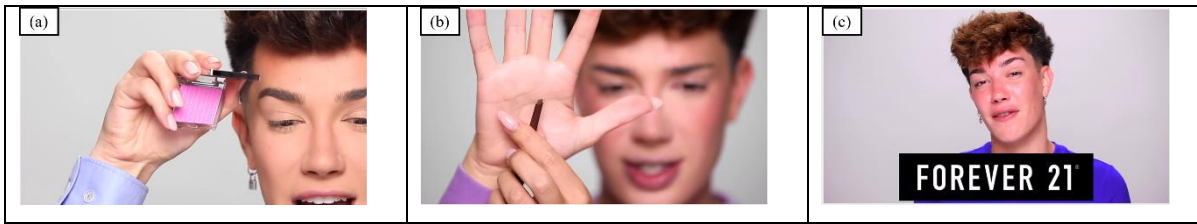


Figure 6. Examples of Visual Boosters: (a) Video 1; (b) Video 8; (c) Video 6

Visual attitude markers in IRVs.

Drawing on Kress and van Leeuwen's (2020) notions of modality and contact, visual attitude markers function to express the influencer's evaluations and affective responses through visual means. In the corpus, they are realized through facial expressions, affective gestures (e.g. thumbs-up), color treatments, affective emoji, and short affective clips. Figure 7 illustrates typical instances: (a) an exclamation emoji marking surprise as the influencer says "Oh, wow! ... I've never seen anything like this before"; (b) an exuberant facial expression accompanying "so much better"; (c) a double thumbs-up supporting the positive evaluation "...a big success"; (d) an embedded reaction clip expressing criticism ("The audacity, the gumption"); and (e) near-desaturation accompanying "That's crazy," visually conveying astonishment.

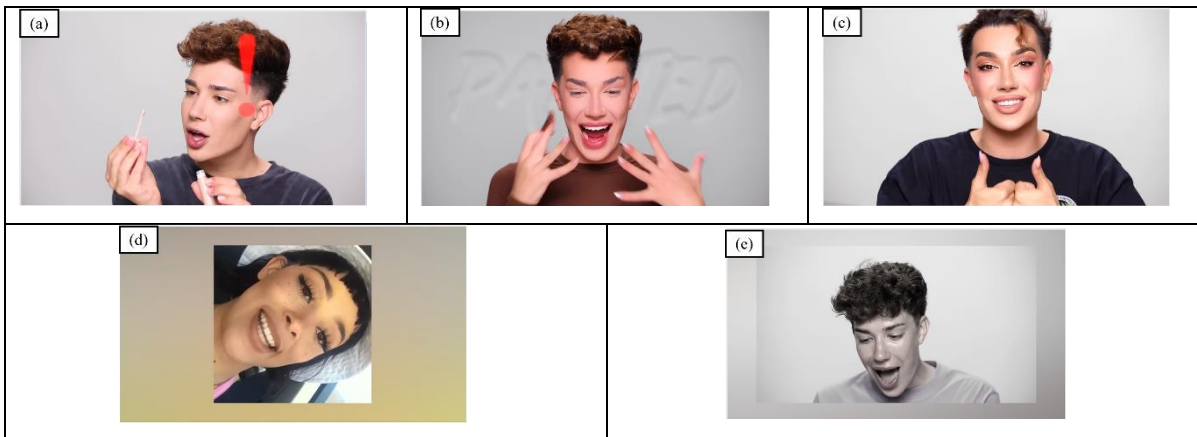
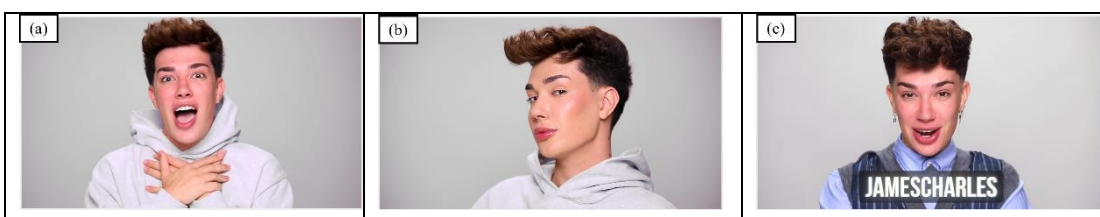


Figure 7. Examples of Visual Attitude Markers: (a) Video 3; (b) Video 12; (c) Video 7; (d) Video 1; (e) Video 8

Visual self-mentions in IRVs.

In IRVs, visual self-mentions foreground the influencer's presence and reinforce self-representation within the review. They are realized through self-directed gestures, photographs, video clips, name displays, platform identifiers, postures, and references to the influencer's own brand. These resources make the creator's identity visually salient. Figure 8 illustrates these realizations: (a) a self-directed gesture, where the influencer points to himself while speaking; (b) a self-display posture after completing the makeup look; (c) the influencer's name appearing on screen during self-introduction; (d) a reference to earlier photographs of the influencer's makeup styles; (e) platform identifiers that visually reinforce the influencer's self-identification; (f) an embedded clip featuring him alongside another creator as he refers to their friendship; and (g) a reference to his own brand, Painted.



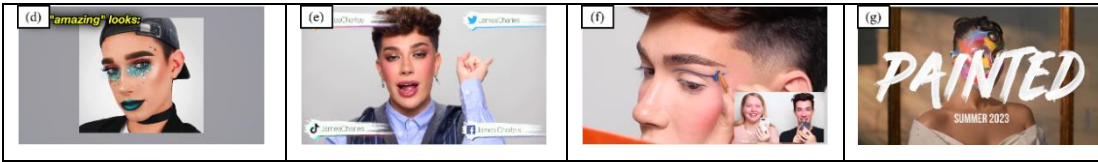


Figure 8. Examples of Visual Self-mentions: (a) Video 10; (b) Video 10; (c) Video 1; (d) Video 3; (e) Video 1; (f) Video 3; (g) Video 3

Visual engagement markers in IRVs.

Visual engagement markers orient visual communication toward viewers and invite their involvement in the discourse (De Groot et al., 2016; Ma & Jiang, 2025). Drawing on Kress and van Leeuwen's (2020) framework, these resources are realized in the dataset through viewer-directed gestures, viewer address text, interactive video clips, viewer-directed close shots, and attention-getting emojis. Figure 9 illustrates these strategies. (a) shows a viewer-directed gesture when the influencer says "you guys"; (b) presents on-screen text at the end of a video inviting viewers to leave comments; (c) incorporates a subscription prompt through a short video clip at the beginning or end; (d) uses a viewer-directed close shot as he says "you guys can see we got tons of things in the bag"; and (e) opens with a greeting "Hi Sisters!" accompanied by a waving-hand emoji.

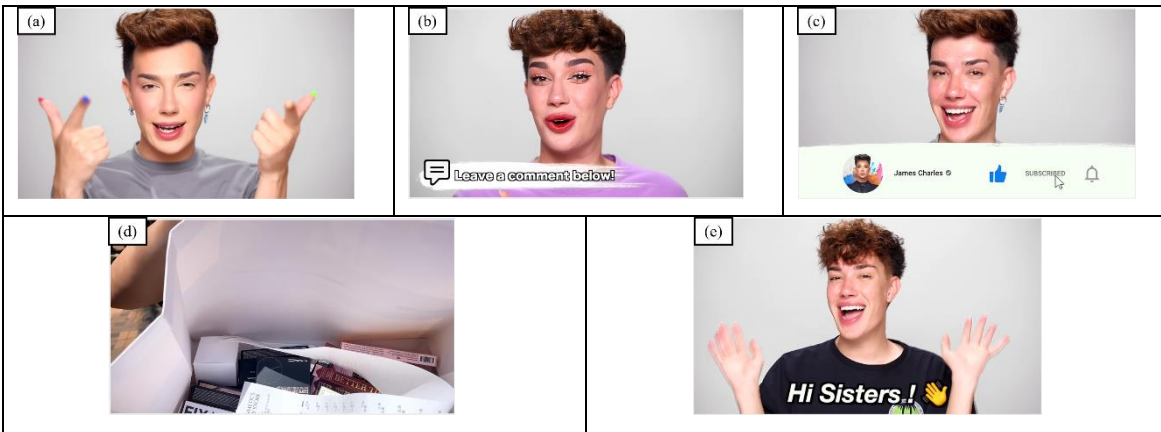


Figure 9. Examples of Visual Engagement Markers: (a) Video 5; (b) Video 8; (c) Video 5; (d) Video 7; (e) Video 7

DISCUSSION

This case study examined the distribution and functions of textual and visual metadiscourse in beauty-related YouTube review videos produced by a single influencer. Regarding textual metadiscourse, the present corpus shows a clear quantitative predominance of interactional over interactive resources, suggesting a discourse profile that combines explicit organization with a marked orientation toward the influencer and viewers. Among interactive resources, transitions are the most frequent, consistent with their function in establishing discourse relations and with Al-Subhi's (2022) findings in social-media advertising. Frame markers are also relatively prominent, which may reflect the extended and sequential nature of review videos, where product presentation, testing, and evaluation unfold across different stages. By contrast, endophoric markers, evidentials, and code glosses are less frequent, distinguishing the present corpus from academic writing, where explicit cross-referencing, source attribution, and elaboration play a more substantial role (Vasheghani Farahani, 2020). The low frequency of evidentials may reflect greater reliance on the influencer's own evaluations than on external sources for support. Unlike academic writing, where code glosses often function to define specialized concepts (Guziurová, 2022), code glosses in IRVs primarily clarify and elaborate the influencer's descriptions and evaluations. At the interactional level, self-mentions and engagement markers dominate, reflecting an emphasis on personal experience and direct viewer orientation. The prominence of self-mentions is consistent with research on YouTube reviews showing that reviewers foreground personal opinion and reviewer identity in the evaluation process

(Gibson, 2022; Hallinan, 2023). Engagement markers further orient the discourse toward absent viewers, resembling audience-design practices in vlog communication (Frobenius, 2014). Boosters also outnumber hedges, while attitude markers remain relatively frequent, suggesting a stronger tendency toward explicit and confident evaluation than toward cautious qualification, a pattern that contrasts with the substantial use of hedging reported in research article discussions (Liu & Tseng, 2021).

The videos show visual interactive and interactional resources used at roughly equal rates, indicating a relatively balanced quantitative distribution between the two categories. Functionally, these resources contribute to both discourse structuring and interpersonal positioning. This balanced profile differs from some static multimodal genres. For example, academic posters and conference slides tend to rely heavily on visual interactive devices to organize information (D'Angelo, 2016; Ma & Jiang, 2025). In the IRVs, visual transitions, endophoric markers, and frame markers occur at relatively high frequencies. This frequency pattern may reflect the sequential nature of product demonstrations, where transitions maintain continuity, endophoric markers link verbal and visual elements, and frame markers segment stages of the review. It also parallels the way images guide readers in digital academic media (Carrió-Pastor, 2021). Likewise, the relatively frequent use of visual attitude markers, boosters, and engagement markers suggests that visual resources are repeatedly drawn on to express evaluation, heighten salience, and orient the review toward viewers, consistent with De Groot et al.'s (2016) account of visual metadiscursive functions. By contrast, no visual hedges are identified, indicating that mitigation is not visually realized in the present corpus. This pattern suggests a stronger visual tendency toward emphasis and evaluation than toward mitigation.

CONCLUSION

This case study has examined textual and visual metadiscourse in beauty-related review videos produced by a single influencer, focusing on their distribution and functions across textual and visual modes. In the IRVs examined here, a broad range of metadiscursive resources contributes to discourse organization and viewer orientation, highlighting how these functions are realized across textual and visual modes. The findings extend metadiscourse research to dynamic, video-based influencer review discourse. The study further refines De Groot et al.'s (2016) visual metadiscourse model at both the category and resource levels. At the category level, frame markers and endophoric markers are incorporated because they are visually realized in the dynamic review videos examined here. At the resource level, concepts from Kress and van Leeuwen's (2020) visual grammar, such as overlap, frontal angle, and close shot, are adapted to the review-video context, while gestures are further differentiated into self-directed, viewer-directed, and affective forms according to their metadiscursive functions. These refinements reveal expanded category coverage and genre-specific visual realizations that are not fully captured by the original model.

Nonetheless, the study has several limitations. The corpus consists of only 12 videos from a single influencer within the beauty genre, which limits the generalizability of the findings. Future research could therefore include larger and more diverse samples involving different influencers, genres, and platforms. The analysis also focuses primarily on textual and visual modes, leaving aside other semiotic resources such as sound and music that may contribute to meaning-making in review videos. Future studies could incorporate these modes to provide a more comprehensive multimodal analysis.

REFERENCE

1. Abidin, C. (2016). Visibility labour: Engaging with influencers' fashion brands and #OOTD advertorial campaigns on Instagram. *Media International Australia*, 161(1), 86–100. <https://doi.org/10.1177/1329878X16665177>
2. Al-Subhi, A. S. (2022). Metadiscourse in online advertising: Exploring linguistic and visual metadiscourse in social media advertisements. *Journal of Pragmatics*, 187, 24–40. <https://doi.org/10.1016/j.pragma.2021.10.027>
3. Bernad-Mechó, E. (2017). Metadiscourse and topic introductions in an academic lecture: A multimodal insight. *Multimodal Communication*, 6(1), 39–60. <https://doi.org/10.1515/mc-2016->

0030

4. Carrió-Pastor, M. L. (2021). Multimodal metadiscourse in digital academic journals on linguistics, engineering and medicine. *European Journal of English Studies*, 25(3), 259–277. <https://doi.org/10.1080/13825577.2021.1988254>
5. Cruz García, L. (2017). Evidentials in advertising: A sample study. *Revista de Lingüística y Lenguas Aplicadas*, 12, 1–12. <https://doi.org/10.4995/rlyla.2017.7074>
6. D'Angelo, L. (2016). Academic posters: A textual and visual metadiscourse analysis. Peter Lang.
7. De Groot, E., Nickerson, C., Korzilius, H., & Gerritsen, M. (2016). Picture this: Developing a model for the analysis of visual metadiscourse. *Journal of Business and Technical Communication*, 30(2), 165–201. <https://doi.org/10.1177/1050651915620235>
8. Dłóciok, J. R., Oben, B., & Meex, B. (2026). Visual metadiscourse in mediated communication: A systematic scoping review. *Frontiers in Communication*, 11, 1786155. <https://doi.org/10.3389/fcomm.2026.1786155>
9. Frobenius, M. (2014). Audience design in monologues: How vloggers involve their viewers. *Journal of Pragmatics*, 72, 59–72. <https://doi.org/10.1016/j.pragma.2014.02.008>
10. Gibson, W. (2020). Sensory communication in YouTube reviews: The interactional construction of products. *Discourse & Communication*, 14(4), 383–403. <https://doi.org/10.1177/1750481320910523>
11. Gibson, W. (2022). All you need is guitar pedals: The communicative construction of material culture in YouTube product reviews. *Discourse, Context & Media*, 49, 100626. <https://doi.org/10.1016/j.dcm.2022.100626>
12. Guziurová, T. (2022). Glossing an argument: Reformulation and exemplification in L2 Master's theses. *Topics in Linguistics*, 23(2), 18–35. <https://doi.org/10.2478/topling-2022-0009>
13. Hallinan, B. (2023). No judgment: Value optimization and the reinvention of reviewing on YouTube. *Journal of Computer-Mediated Communication*, 28(5), zmad034. <https://doi.org/10.1093/jcmc/zmad034>
14. Hyland, K. (2005). Metadiscourse: Exploring interaction in writing. Continuum.
15. Hyland, K. (2007). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied Linguistics*, 28(2), 266–285. <https://doi.org/10.1093/applin/amm011>
16. Indrawati, D., Luckieta, M., & Nadelia, F. (2023). Critical multimodal discourse analysis of YouTube influencers' impact on purchase intention. *AIP Conference Proceedings*, 2594(1), 150015. <https://doi.org/10.1063/5.0110894>
17. Janoschka, A. (2004). Web advertising: New forms of communication on the Internet. John Benjamins Publishing Company.
18. Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
19. Kress, G., & van Leeuwen, T. (2020). *Reading images: The grammar of visual design* (3rd ed.). Routledge.
20. Kumpf, E. P. (2000). Visual metadiscourse: Designing the considerate text. *Technical Communication Quarterly*, 9(4), 401–424. <https://doi.org/10.1080/10572250009364707>
21. Li, M., Gibbons, J., & Pham, Q. N. (2023). Re-exploring writer-reader interaction: Analyzing metadiscourse in EAP students' infographics. *Journal of English for Academic Purposes*, 66, 101303. <https://doi.org/10.1016/j.jeap.2023.101303>
22. Li, X. (2022). The discursive construction of corporate identity in the corporate social responsibility reports: A case study of Starbucks. *Frontiers in Psychology*, 13, 940541. <https://doi.org/10.3389/fpsyg.2022.940541>
23. Liu, C., & Tseng, M.-Y. (2021). Paradigmatic variation in hedging and boosting: A comparative study of discussions in narrative inquiry and grounded theory research. *English for Specific Purposes*, 61, 1–16. <https://doi.org/10.1016/j.esp.2020.08.002>
24. Ma, Y., & Jiang, F. K. (2025). Guiding and engaging the audience: Visual metadiscourse in PowerPoint slides of Three Minute Thesis presentations. *English for Specific Purposes*, 77, 56–70. <https://doi.org/10.1016/j.esp.2024.10.003>
25. Macário Lopes, A. C., & Amaral, P. (2006). From time to discourse monitoring: Agora and então in European Portuguese. *Belgian Journal of Linguistics*, 20, 3–18.

- <https://doi.org/10.1075/bjl.20.03mac>
26. Mauranen, A. (1993). Contrastive ESP rhetoric: Metatext in Finnish-English economics texts. *English for Specific Purposes*, 12(1), 3–22. [https://doi.org/10.1016/0889-4906\(93\)90024-I](https://doi.org/10.1016/0889-4906(93)90024-I)
 27. O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 19, 1–13. <https://doi.org/10.1177/1609406919899220>
 28. O'Halloran, K. L. (2021). Multimodal discourse analysis. In K. Hyland, B. Paltridge, & L. Wong (Eds.), *The Bloomsbury handbook of discourse analysis* (2nd ed., pp. 249–266). Bloomsbury Academic. <https://doi.org/10.5040/9781350156111.ch-018>
 29. Pearson, W. S., & Abdollahzadeh, E. (2023). Metadiscourse in academic writing: A systematic review. *Lingua*, 293, 103561. <https://doi.org/10.1016/j.lingua.2023.103561>
 30. Torjesen, A. (2021). The genre repertoires of Norwegian beauty and lifestyle influencers on YouTube. *Nordicom Review*, 42(2), 168–184. <https://doi.org/10.2478/nor-2021-0036>
 31. Vande Kopple, W. J. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication*, 36(1), 82–93. <https://doi.org/10.58680/cc198511781>
 32. Vasheghani Farahani, M. (2020). Metadiscourse in academic written and spoken English: A comparative corpus-based inquiry. *Research in Language*, 18(3), 319–341. <https://doi.org/10.18778/1731-7533.18.3.05>
 33. Williams, J. M. (1981). *Style: Ten lessons in clarity and grace*. Scott, Foresman.
 34. Zou, H. (J.), & Hyland, K. (2022). Stance in academic blogs and three-minute theses. *International Journal of Applied Linguistics*, 32(2), 225–240. <https://doi.org/10.1111/ijal.12411>