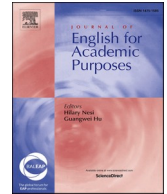




ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap

Modal density and coherence in science dissemination: Orchestrating multimodal ensembles in online TED talks and youtube science videos

Julia Valeiras-Jurado^a, Edgar Bernad-Mechó^{b,*}

^a Universitat de València (UV) / IULMA, Spain

^b Universitat Jaume I (UJI), Spain

ARTICLE INFO

Keywords:

Science dissemination
Multimodality
Modal density
Modal coherence
Digital genres

ABSTRACT

As science communication goes beyond the traditional borders of academia, new science dissemination genres emerge. This paper presents a multimodal analysis of two online science videos uploaded to YouTube: a recorded TEDx talk and a PBS Space Time video. We have determined the existence of two layers of modes in online science videos: embodied and filmic modes. The former are deployed by the speaker, while the latter are added to the ensemble during postproduction and editing. Our analysis sheds light on three main aspects of this multimodal orchestration: modal density, modal coherence, and genre definition. Our findings show that online science videos are more modally dense as a result of the incorporation of more filmic modes into the ensemble. In addition, we also find that the use of filmic modes requires some degree of expertise and know-how in order to keep the coherence of the multimodal ensemble. In short, the PBS Space Time video shows a higher degree of density and coherence. We conclude that the two genres are at different stages of development as digital genres.

1. Introduction

1.1. Science dissemination and recontextualization of scientific content

There seems to be an increasing trend for science to transcend the boundaries of academic institutions. As [Scotto di Carlo \(2014\)](#) points out, cultural organizations no longer hold the monopoly over knowledge dissemination. New communication channels have appeared in the picture that bring science closer to the public without the need of an academic institution as an intermediary. In this scenario, “the success of knowledge dissemination depends on how experts contribute to the way in which the audience approaches science not as something distant and separate, but as a heritage belonging to the whole community” ([Scotto di Carlo, 2015](#), p. 219).

The Internet has played a crucial role in this process, multiplying the possibilities for scientists to present their work to the general public. In fact, Rowley-Jolivet and Carter-Thomas claim that

* Corresponding author.

E-mail address: ebernad@uji.es (E. Bernad-Mechó).

<https://doi.org/10.1016/j.jeap.2022.101118>

Received 26 August 2021; Received in revised form 14 April 2022; Accepted 17 April 2022

Available online 8 May 2022

1475-1585/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

[t]he Internet has changed the game rules for scientific communication in several respects: the need to address a global, indeterminate audience, not an esoteric circle of peers, raises the problem of knowledge asymmetry between science researcher and audience, and requires suitable discursive strategies (2019, p. 82).

Online videos seem to be one of the preferred options to disseminate science (Erviti & Stengler, 2016; Kousha et al., 2012; León & Bourk, 2018) and the challenge to adapt to an audience that does not share the same academic background is apparent in the discursive strategies and semiotic modes used. This adaptation has been referred to in previous literature as *recontextualization*. From a multimodal perspective, that is, taking into account the variety of semiotic modes, i.e. ways of conveying meaning, involved in the communicative process, Bezemer and Kress have described recontextualization as:

moving meaning material from one context with its social organization of participants and its modal ensembles to another, with its different social organization and modal ensembles. Meaning material always has a semiotic realization, so recontextualization involves the re-presentation of the meaning materials in a manner apt for the new context in the light of the available modal resources (2008, p. 184).

In the case of science dissemination, recontextualization means moving scientific content from academia and its associated genres to a different context in which new genres and new multimodal ensembles, i.e. combinations of semiotic modes, become available, and more importantly, a context in which the audience has different needs and expectations. According to Bezemer and Kress (2008), recontextualization involves four rhetorical principles:

- a) *Selection* of both content that is relevant for the new audience and of semiotic modes which are available in the new context;
- b) *Arrangement* of the content in a way that is best for the audience and for the communicative purpose;
- c) *Foregrounding* of the elements that are particularly significant in the new context; and
- d) *Social repositioning* or reconstruction of social relations between the speaker and the audience.

These rhetorical principles can be realized by means of different strategies. For the case of scientific discourse, Luzón (2019) identifies four groups depending on their function: i) build credibility; ii) build persuasive arguments; iii) tailor information to the assumed knowledge of the audience; and iv) engage the audience. Although recontextualization has been extensively dealt with in previous research, to our knowledge, few studies approach this issue from a multimodal discourse analysis perspective.

1.2. From traditional to digital genres

As Maingueneau (2010) points out, the concept of genre evolves with history, and the creation of the Internet and new digital platforms has facilitated the emergence of new types of genres, which cannot always be analyzed with classic Discourse Analysis approaches. Of particular relevance for the present study is the case of digital genres, which are web-mediated genres that use the Internet as a medium. According to Shepherd and Watters (1998), these genres can be defined using three aspects: content, form and functionality (the capabilities afforded by the new online medium). As these authors explain, some digital genres (cybergenres) originate from a traditional genre and stay unchanged as they move to the online realm (replicated genres), while others evolve and acquire new traits using the affordances of their new medium (variant genres). There are also novel genres that are created directly in this new medium and are either substantially different from existing ones or created from scratch. In many cases the trend is as follows: a replicated genre starts adapting to the new medium and incorporating more affordances until it becomes a novel genre.

This paper wants to probe into the multimodal nature of online science videos, exploring the differences between two digital genres. In the case of the genres that concern this study, a TEDx Talk video and an episode from the YouTube channel PBS Space Time, we can identify a replicated genre in the former and a novel genre in the latter. We take this hypothesis of different stages of development of the two genres as a starting point for the analysis, as remarkable differences are noted in the exploitation of the affordances provided by the online medium. Replicated genres can only attain so much potential in comparison to the original genre they depart from, while novel genres typically take full advantage of their medium.

1.3. YouTube as a platform for online science dissemination

As mentioned in previous sections, the Internet has played and still plays a crucial role in the dissemination of science and in the creation of new genres. In particular, the video-sharing site YouTube has gained significant relevance and hosts an increasing amount of scientific content (Allgaier, 2020; Geipel, 2018). This growing popularity of YouTube has caught the attention of researchers in different fields. Snickers and Vonderau (2009) describe YouTube as a platform, an archive, a library, a laboratory, a database, and a medium at the same time. Burgess and Green (2009) see it as a spearhead of participatory culture. Soukup (2014), on the other hand, reflects on the complexity of the platform, claiming that it has grown into something bigger than a video-sharing site, a social media site, an educational resource or a communication channel, and therefore deserves more attention from communication research. The potential of YouTube for the communication and dissemination of science has been noted by several researchers who point out advantages like the rapid distribution, the amplification of content or the ability to have informal exchanges (Osterrieder, 2013).

The factors that make science videos successful have also been the object of recent research. For instance, Welbourne and Grant (2016) examined factors that affect channel and video popularity in science communication on YouTube. They found that user-generated content seems to be more popular among viewers than professionally generated content in the area of science

communication. They also found that credibility of the content does not necessarily depend on high production values but also on issues such as communicator experience, impartiality, affinity, accessibility, and ability to build meaningful connections with the audience. In addition, it must be noted that the line between professional and user- (amateur) created content is often blurry (Muñoz Morcillo et al., 2016). It is important to note that the number of views in YouTube is a metric that says more about the popularity of the video than of its scientific quality. As Thelwall et al. (2012) found out, the success of YouTube science videos depends on different variables (e.g., entertainment value, particular topic of research, visual images, etc.) that are not necessarily related to its scientific value. Saurabh and Gautam (2019) also point out that the number of views increases as the channel becomes established and more videos are uploaded. They also conclude that medium-length videos are preferable, and that the initial part of the video needs to be exciting enough to keep the viewers' interest until the end.

Concerning the presenter, Sugimoto and Thelwall (2013) found that science and technology TED videos presented by academics were "liked" more often than those by non-academics, which shows that scholars are not disadvantaged in this new media environment. González García et al. (2020) conclude that most successful YouTube videos feature a male, young scientific communicator who employs both animations and his public impression, i.e. the way he is perceived by the audience, as effective communication tools. Frobenius (2014) claims that successful YouTubers create the illusion of talking live to their audiences, and Pérez-Torres et al. (2018) suggest the key to success is being perceived as an equal.

Focusing now on one of the genres studied here, TED talks, Caliendo and Compagnone define this genre as "a series of short popularizing talks (of approximately 20 min), addressing a mass audience and delivered by top-level experts in a wide variety of domains" (2014, p. 105). TED talks have been often described as a blend of other genres. For example Ludewig (2017) finds elements of sales pitches, educational communication, and memoirs, and she claims that these talks have gone through a process of perform-ativization or theatricalization to become "infotainment" or "edutainment". Still, TED conferences have been criticized for elitism. However, a select group of videos is made available under a creative commons license through the TED website and YouTube (Sugimoto & Thelwall, 2013). As a result, there is a primary audience at the conferences, and a secondary audience for the video recordings of these conferences. These videos are carefully produced, sometimes even by professional filmmakers. Though smaller TEDx conferences (local and independently organized TED events) lack such elaboration, they still follow these general guidelines (Ludewig, 2017). It is precisely one of the video-recorded TEDx Talks that concerns this study.

The other YouTube videos we consider in this study are PBS Space Time Videos. PBS Space Time is a subsidiary of Public Broadcasting Service (PBS), a nonprofit organization which produces, broadcasts and distributes audiovisual products both on TV and online. As these types of videos are a more recent phenomenon, to our knowledge, they have not yet been defined as a genre. They could fit into what some scholars have called *scifotainment* or *edutainment* genres (Pérez-Llantada, 2021). They differ from TED talks in their length (they are significantly shorter), their level of formality (they are significantly more informal) and, above all, in the fact that they are not recorded live events but videos produced to be hosted online. In some ways they resemble other online scientific content like vlogs. However, we consider PBS Space Time videos to be different from vlogs, since the main characteristic of vlogs is their spontaneity, immediacy, looseness and amateur feel (Werner, 2012). PBS Space Time videos, on the other hand, are rehearsed, staged and professionally produced. They also have a sequentiality and continuity that makes them similar to a TV series (e.g. they are released as "episodes").

1.4. A multimodal approach to online science dissemination genres

A multimodal approach to genre is based on the assumption that different modes interact in communication to express meaning in the genre-creation process (Kress, 2003; Kress & van Leeuwen, 2006). The concepts of *mode* and *media* can help understand the production and interpretation processes involved in genres. Jewitt (2004) clarifies the difference between these concepts stating that *modes* enable representation of meaning, while *media* allow for the dissemination of meaning. When modes are combined to make meaning, they are orchestrated in *multimodal ensembles* (Kress, 2010). In addition, Norris (2004) claims that the number of modes in itself is less significant than their intensity (i.e. how much they are used) and the complexity of the interrelationships among them. She refers to this as *modal density*. Moreover, when we look further into these interrelationships, we can discern if these modes are orchestrated into coherent ensembles, that is, whether they are used in a consistent way and do not contradict each other. We refer to this as *modal coherence* (Valeiras-Jurado, 2019). The concept of modal coherence is not absolute, and to a great extent it is relative and perspectival. However, it is measurable according to some external, objective factors: a multimodal ensemble can be considered coherent if all modes contribute seamlessly to the communicative aim at hand (e.g. inform, educate, etc.), which normally implies following some genre conventions.

Online science dissemination genres are intrinsically multimodal, which calls for a multimodal discourse analysis approach. Boy et al. claim that

YouTube science videos are seen as well-organized multimodal arrangements consisting of a variety of visual and verbal modes like stills, moving images, text, spoken language, sounds, animations, graphics, etc. which is a much more complex system of communication than text only (2020, p. 4).

These authors have classified scientific online videos according to the number of modes that are deployed to compose audiovisual scientific content, and to the primary function of the multimodal orchestration. Four basic types of science videos are thus identified:

In *presentation films* the speaker talks directly to the camera in a medium closeup shot and answers a limited number of scientific questions. In addition to spoken language, it features other modes such as text over visuals, background images, animations, gestures and facial expressions. Their strength lies mainly in their personalization ability. In *expert films* the focus lies on an expert person who

discusses a topical field of research. These videos contain more moving image material than platform-specific presentation modes (animations, insertion of user comments, addressing users, etc.). *Animation films* show artificial moving images to illustrate a process, a problem, an issue, or a scientific theory. Finally, *narrative explanatory films* are based on a general question that is answered through arguments that follow logical reasoning. They are more complex and often contain elements that characterize the previous types, combining narrative and informative elements by telling an entertaining story while they give an explanation. They use mainly moving image material and have the highest number of cuts. Also, they can exercise strong attentional control. The two videos analyzed in this study might fit into this category.

1.5. The complex interplay of modes in online TED talks and PBS Space Time videos

A characteristic of the TED and Space Time videos is that they feature multimodal ensembles that comprise a wide array of modes. Some of these modes can be called embodied, and are controlled by the speaker (Norris, 2004), while others are more related to the editing and production of the videos, and can be classified as filmic modes, e.g. cuts or shots (Tseng & Bateman, 2010). Some authors have analyzed this multimodality in videos by grouping modes into *layers* (Burn, 2014; Farias et al., 2021; Jimenez Hurtado & Soler Gallego, 2013; Kozloff, 2000; Vanoye, 1985). Vanoye (1985), for example, talks about a first layer in which communication takes place between the characters of a film, and a second layer in which communication happens between the filmmaker and the audience. This second layer is the one that hosts filmic modes. These modes should not be neglected in the multimodal analysis of science dissemination videos, since they convey crucial meaning. For example, shots do not only specify the actual distance between the camera and the character (or, in our case, the speaker) but also the virtual distance between the speaker and the audience, simulating the proxemics of interpersonal communication (Kress & van Leeuwen, 2006). Also, moving images can contribute to the narration and the interaction with the audience (Rowse, 2014). In fact, Muñoz Morcillo et al. claim that

[s]everal factors such as narrative strategies, video editing techniques, and design tendencies with regard to cinematography, the number of shots, the kind of montage used, and even the spread use of sound design and special FX point to an increasing professionalism among science communicators independent of institutional or personal commitments. In general, it can be said that “supposed” amateurs are creating the visual language of science video communication. (2016, p. 1).

As Mayer (2014) suggests, in this complex, multilayered modal orchestration it is of crucial importance that layers do not compete for attention, but rather complement each other, or, in other words, that they are coherent.

It must be pointed out that this division into layers should be understood as an analytical tool to facilitate the multimodal analysis. In reality, the two layers are intertwined and influence each other, especially if the video is conceived to be distributed online and filmic modes are included in the design from the onset. For example, a particular type of shot or angle can affect the use of gaze or gestures.

This study explores the differences between replicated and web-native genres in online science videos adopting a multimodal approach. The existence of a second layer of filmic modes has been discussed in multimodal analyses of films, and the increasing importance of professional production (which implies the use of filmic modes) has been noted in previous studies of YouTube science videos. However, we believe that the highly elaborate multimodal ensembles orchestrated in these videos pose some unanswered questions that we aim to answer in this study:

- How do filmic modes have an impact on modal density?
- How do the two modal layers (embodied and filmic) cohere with each other?
- In which ways does this inform the definition of the two genres? To what extent is the online medium an integral part of these definitions?

2. Methodology

2.1. The dataset

In order to answer the research questions presented above, two science dissemination videos were selected from YouTube for further analysis. Both videos feature Dr. Matthew O’Dowd, an astrophysics researcher at the City University of New York and science communicator, talking about Fermi’s Paradox and the probabilities of finding alien life in the universe based on the experience on Earth. The first video is entitled “Fermi’s Paradox and the Psychology of Galactic Empires¹” and it portrays a face-to-face TEDx talk given at the Technische Universität Wien in 2017 that was recorded and later on uploaded to YouTube after some editing on August 9, 2017. The video lasts 23 min and 44 s and, at the time this paper was written, it had over 1.1 million views and 12 thousand likes. The video was uploaded by the channel TEDx Talks, which had 34.4 million subscribers. The second video is entitled “Why Haven’t We Found Alien Life?²” and it consists of an edited recording made by the channel PBS Space Time. This video has a duration of 10 min and 33 s. It was uploaded to YouTube on Nov 5, 2015 and, at the time of writing, it had over 3 million views and about 35 thousand likes. Moreover, the channel PBS Space Time had 2.6 million subscribers. These videos were chosen as they entail two different

¹ <https://www.youtube.com/watch?v=3ITvVTnjU5U>.

² <https://www.youtube.com/watch?v=cJONS7sqi0o>.

representations of a rather similar content: a multi-camera recording of a live TED talk and an edited video made *ad hoc* for an online platform. In this sense, and for the purposes of this paper, both videos are seen as an online experience (as opposed to an *in situ* event in a TED talk), taking into account the affordances of the online medium. This means that we are not analyzing the TED talk as a live phenomenon, which from our point of view would represent a different communicative event and genre, and which has already attracted considerable attention from previous research (cf. introduction section in this paper).

Once the videos were selected, and after an initial viewing and analysis in terms of content, four short excerpts were extracted from each of the videos. These fragments constitute pairs which are similar in terms of content and therefore comparable. In this manner, a comparative analysis could be designed to explore the similarities and differences between the TEDx and the PBS Space Time videos. Table 1 below describes the dataset used in the analysis.

2.2. The multimodal annotation

The eight fragments were multimodally annotated using the software *Multimodal Analysis Video* (MAV) (O'Halloran et al., 2012). Following Kozloff (2000) and Vanoye (1985), two main layers of meaning were identified for the analysis: embodied modes, i.e. verbal and non-verbal body-performed modes, and filmic modes, i.e. the modes arising from the editing process of the YouTube videos. Whilst the former set of modes fully depends on the presenter, the latter is non-aligned to the speaker and is only perceived by the audience watching the videos.

Among the embodied modes, and in line with previous research on multimodal analyses of academic discourse (Bernad-Mechó, 2021), the following were taken into account: spoken language, paralinguistic, gestures, gaze, proxemics, head movement and facial expression. For spoken language, i.e. what is verbalized, we considered the verbal transcription of the speech (i.e. entirely corpus-driven). As for paralinguistic, the focus was put on the use of pauses, understood as silences that are longer than half a second; the presence of prominent strands of speech (e.g. a particular tone change, a salient stress, etc. that stands out from the rest); and the tempo in words per minute. For gestures, McNeill's (1992) classification was followed to identify *iconic*, *metaphoric*, *deictic* and *beats* gestures. We consider movements of the hands and arms as gestures, and they were classified in relation to the full ensemble in which they were integrated (i.e. not just in relation to spoken language). Our annotation system corresponds to Kendon's (2004) gesture units in the case of single gestures (that is, from relaxation moment to relaxation moment, including a preparation phase, a stroke and a recovery phase). In the case of repetitive gestures, however, we have annotated each iteration, which would be closer to what Kendon (2004) calls a gesture phrase (including the preparation and the stroke). Therefore, each iteration (or gesture phrase) may be classified differently. It is important to note that the unit used for annotation (gesture phrase or gesture unit) does not affect our quantitative analysis, since we do not count units, but the amount of time the speaker is gesturing relative to the total length of the video. In the case of gaze, the direction of the gaze was marked (whether aimed at the audience, the camera, the floor, etc.). In proxemics, a basic distinction was made between a standing position and the moments in which the presenter was pacing. This distinction was corpus-driven and is in line with previous studies (Bernad-Mechó & Fortanet-Gómez, 2019). As for head movements, nods, shakes and tilts were identified. Finally, eyebrow raisings, frowns and smiles were annotated to account for facial expression.

Within the editing layer, we drew from a series of studies incorporating filmic modes to their analyses (Desilla, 2012; Muñoz Morcillo et al., 2016; Tseng, 2008; Tseng & Bateman, 2010) to include type of shot, angle, *mise-en-scène*, cuts, music, visual prompts, sound effects and video effects into our multimodal analysis. For the analysis of shots, a classical taxonomy (Bowen, 2018) was adapted to fit the reality of the dataset, thus distinguishing between *close-ups*, *medium close-ups*, *medium shots*, *cowboy shots*, *medium full shots*, *full shots* and *long shots*. Furthermore, the angle of the camera was considered, whether it was a frontal or a lateral one. The analysis of *mise-en-scène* was limited to the description of the backgrounds used in the videos, regardless of the elements shown in this background (e.g. slides, curtains, TED logo, digital prompts etc.). Similarly, cuts in the editing and the use of music were also annotated whenever they were present. Visual prompts refer to images and texts (e.g. a full screen zoom in on the slides). Finally, sound and visual effects were also considered; these refer to any digital effects that were added during the editing. The selection of both embodied and filmic modes was corpus-driven.

Table 2 summarizes our framework for analysis. We argue that the selection of these specific modes allows for an in-depth

Table 1

Dataset of excerpts to be multimodally annotated.

CODE	CONTENT	EXTRACTED FROM	DURATION	NO. OF WORDS
TED_1	The Fermi Paradox. Possible civilizations and implicatures. Where are we on The Great Filter?	08:38 to 12:10	3' 32"	427
PBS_1	The Fermi Paradox. The Great Filter. Personal opinion.	00:46 to 01:22	0' 36"	119
TED_2	Humans as a proto-galactic civilization.	12:11 to 13:09 and 14:12 to 14:52	1' 38"	195
PBS_2	Humans might be first. Earth as the one instance to analyze.	01:23 to 02:01	0' 38"	122
TED_3	Early fossils. The origin of life.	05:17 to 07:16	2'	247
PBS_3	The origins of life. The late heavy bombardment. Early fossils.	03:22 to 06:10	2' 48"	489
TED_4	An answer to The Great Filter: it is behind us. "What if we are early?"	21:15 to 23:44	2' 30"	274
PBS_4	Answers to The Great Filter: "Multicellular organisms", "Emergence of intelligence", "What if we are early?"	08:09 to 08:46 and 09:00 to 10:24	2' 01"	359

3. Results and discussion

3.1. Modal density in double-layered multimodal ensembles

The first research question asked how multimodal ensembles were affected in terms of modal density with the inclusion of a second layer of analysis, i.e. the filmic modes. To answer this question, a quantitative analysis of the modes co-occurring in the multimodal ensembles was carried out using *State Machine* in MAV. Table 3 below provides average quantitative data on the use of each mode for both, the TEDx and PBS Space Time videos. The data are generally expressed in percentages over the total running time of the clips. In other words, percentages show how much time is devoted to each mode. As discussed in the methodological section above, three exceptions are to be considered: tempo is computed in average words/minute; *mise-en-scène* will only be explored from a qualitative point of view in the following section; and cuts are quantified in average number of instances/minute. Furthermore, it is important to remark that some modes may overlap. Thus, for example, a gesture could be interpreted both as a *beat* and as a *metaphoric* gesture at the same time. In such case, the specific percentages for each type of gesture will take this into account. However, in the total amount of gestures, only the moments in which the presenter is using gestures will be computed, as opposed to the moments in which he is not using any gestures, i.e. a gesture carrying two functions will only be counted as one. Finally, the averages are calculated in proportion to the length of each of the clips. In this sense, the values produced by longer clips will have a higher weight on the average than those arising from shorter ones.

Within the embodied modes, the quantitative data shows obvious differences between the two formats, as the PBS Space Time video appears to be much more modally dense. Within the paralinguistic features, for instance, pauses are common in the TEDx video, accounting for 17.68% of the time, while they are almost non-existent in the PBS Space Time video (0.54%). As for the tempo, the average of words/minute is approximately 50% higher in the PBS Space Time video (180.49 vs. 119.89 words/minute). Both of these data show a tendency to mark a more intense rhythm in the PBS Space Time video, in line with what is expected from YouTubers to

Table 3
Average account of modal density for embodied and filmic modes.

			TED Average	PBS Average
EMBODIED MODES	Paralanguage	pauses (%)	17.68	0.54
		prominence (%)	20.36	42.58
		tempo (average words/minute)	119.89	180.49
	Gestures	iconic (%)	0.33	0.82
		metaphoric (%)	25.03	6.87
		deictics (%)	1.23	1.93
		beats (%)	44.35	85.97
		TOTAL (%)	62.72	95.58
	Gaze	floor (%)	1.54	1.64
		clicker (%)	4.33	0.00
		back (%)	3.64	0.00
		prompter (%)	0.00	0.00
		audience (%)	70.50	0.00
		camera (%)	0.00	92.35
		other (%)	14.31	2.43
	Proxemics	standing (%)	75.47	96.41
		pacing (%)	18.87	0.00
	Head movements	nod (%)	5.35	28.86
		shake (%)	1.04	14.01
		tilt (%)	0.50	16.22
	Facial expression	TOTAL (%)	6.89	55.79
		eyebrow raising (%)	2.93	15.34
		frown (%)	1.20	20.84
smile (%)		0.17	0.00	
FILMIC MODES	Type of shot	TOTAL (%)	4.30	36.18
		close-up (%)	0.00	0.00
		medium close-up (%)	0.00	0.00
		medium shot (%)	27.30	36.23
		cowboy shot (%)	0.00	29.97
		medium-full shot (%)	14.14	11.47
		full shot (%)	12.85	3.04
	Angle	long shot (%)	39.70	15.70
		frontal (%)	69.96	96.41
		lateral (%)	24.03	0.00
		Cuts (average instances/minute)	0.00	2.98
Music	(%)	0.83	97.53	
	Visual prompts	image (%)	5.66	76.01
Sound effects	text (%)	1.54	66.93	
	TOTAL (%)	5.66	90.47	
	(%)	0.00	30.98	
	Visual effects (%)	0.83	42.37	

create successful videos (Welbourne & Grant, 2016). In fact, a faster tempo might improve audience attention and persuasiveness (Chambers, 2001; Miller et al., 1976; Smith & Shaffer, 1995). Finally, PBS Space Time clips contain twice as many prominent uses of language as TEDx's (20.36% in TEDx vs. 42.58% in PBS Space Time). This higher density is also observed in other embodied modes such as the total use of gestures (62.72% in TEDx vs. 95.58% in PBS Space Time), head movements (6.89% vs. 55.79%) and facial expression (4.30% vs. 36.18% respectively). These modes seem to work in combination and serve as frequent emphatic rhythm markers. However, these ample differences, as we will discuss below, might be the result of choices in the filmic modes: in the TEDx video, even if the presenter is performing complex embodied ensembles, they are almost imperceptible to the online audience if a long shot is chosen for the camera. Finally, relevant differences are also found in the direction of gaze (almost fully focused on the camera in the PBS Space Time video -92.35%, while mostly focused on the audience in the TEDx video -70.50%). It is important to remark at this point that both videos are seen as YouTube videos. Indeed, the viewer is not attending a face-to-face TEDx talk, but they are watching an online recording. From this point of view, it can be argued that the presenter is only directly addressing the real audience (the online viewers) in the case of the PBS Space Time clip.

When considering the filmic layer of modes, several results stand out. The type of shot is one of such: while medium shots or cowboy shots (a slightly wider medium shot) are preferred in the PBS Space Time video, totaling 66.2% of the total duration of the clips, full shots and long shots are more common in the TEDx video, adding up to 52.55% of the duration of the clips. When looking at the use of angles, only frontal angles fully showing the presenter are chosen in PBS Space Time while the TEDx video combines both frontal shots (69.96%) and lateral shots (24.03%). Note that the remaining 6.01% refers to moments in which the presenter is not portrayed (for instance when showing a slide in full-screen). In general, frontal medium shots contribute to better conveying meaning through the embodied modes. On the contrary, longer shots offer a wider view of the setting (for instance, used in the TEDx video to include part of the background presentation within the frame). Cuts are only used in the PBS Space Time videos; at an average rate of 2.98 instances per minute. These contribute to a swifter montage without pauses, emulating operating modern television conventions such as the MTV-style editing (Dancyger, 2011). This kind of editing is usually also accompanied by music (present in 97.53% of the PBS Space Time clips) while this mode is only present in the outro of the TEDx video (0.83%). Similarly, visual prompts like the use of text and image are prevalent in the PBS Space Time videos (90.47%) while they are almost negligible in the TEDx video (5.66). As for the remaining filmic modes, they are quasi-exclusive of the PBS Space Time format. Sound and visual effects are almost non-existent in the TEDx talk, while they are very common in PBS Space Time (30.98% for sound and 42.37% for visual effects). Again, this kind of edition follows the standards of successful popular science videos on YouTube (Muñoz Morcillo et al., 2016)

In conclusion, PBS Space Time clips are denser, not only in terms of quantity and frequency of modes, but also in terms of


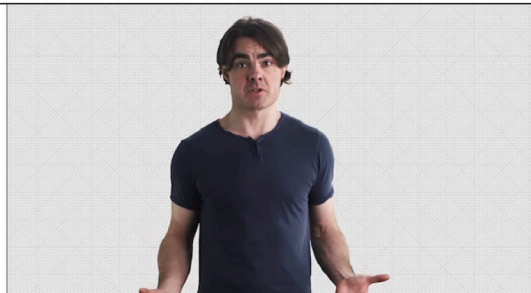
TED_3 – 01:01	PBS_3 – 00:25
	
<i>[...] is countless generations for a civilization that managed to come ahead of us.</i>	<i>I just don't think there's an inevitable great filter still ahead of us.</i>
Paralanguage: prominence	Paralanguage: prominence
Gestures: metaphoric	Gestures: beats
Gaze: audience	Gaze: Camera
Proxemics: standing	Proxemics: standing
Head movements: nod	Head movement: shake
Facial expression: --	Facial expression: eyebrow raising
Type of shot: long shot	Type of shot: medium shot
Angle: frontal	Angle: frontal
Mise-en-scène: TED logo and slides	Mise-en-scène: light gray background
Cuts: --	Cuts: --
Music: --	Music: mysterious
Visual prompts: --	Visual prompts: --
Sound effects: --	Sound effects: --
Visual effects: --	Visual effects: --

Fig. 1. Use of shots and camera angles to capture embodied modes.

complexity and the interactions among them, which contribute to building up a more coherent format. The level of professionalism in the production as well as the complexity of montage are reflected through a higher density in the filmic layer of modes, exploiting the affordances of the medium. Nevertheless, further qualitative analyses need to be carried out to account for the implicatures of specific multimodal ensembles and their conveyed meanings. These analyses are provided in the section below.

3.2. Modal coherence in online science videos

Our second research question inquires into modal coherence in the TEDx and PBS Space Time videos. In particular, we focus on coherence across the two layers of modes, i.e. embodied and filmic (Mayer, 2014). The analysis revealed that, in general, PBS Space Time shows greater coherence across the embodied and the editing layer, while in the TEDx video the two layers frequently contradict each other and do not always work jointly towards the same communicative aim. Some examples are offered in the following paragraphs.

The type of shot and camera angle in the TEDx video is very often not coherent with the rest of the embodied modes. There are examples of long shots that prevent the correct visualization of gestures, face expression and head movements for the audience watching the video, which in turn jeopardizes the expression of attitudinal meaning. It is true that the speaker might be orchestrating coherent ensembles that convey this meaning to people in the conference room; however, these cannot be identified by the online audience. In contrast, PBS Space Time shots are carefully chosen to exploit the affordances of embodied modes, for example shifting from a long to a medium shot and focusing on the speaker's gestures and face expression when he expresses his personal opinion. Unlike in the TEDx video, the speaker in PBS Space Time addresses the camera directly. In this sense, the interaction between embodied multimodal ensembles and the actual -online- audience in the TEDx video is constrained by the choices in filmic modes. Fig. 1 compares the use of shot and camera angle in the two videos.

When looking at embodied modes, we note that they are more coherent among themselves in both videos. For example, nods very frequently coincide with gaze directed at the (physically present) audience, in what seems an attempt to seek agreement (Fig. 1). Similarly, frowning and eyebrow raising also coincide to a great extent with gaze directed at the audience. This can be interpreted as an attempt to show the audience what reaction the speaker is expecting from them. A further example is that metaphoric gestures are coherent with spoken language. Very frequently they offer a visual representation of what is conveyed with words and mostly coincide with gazing at the audience, as if checking that the message is getting across (see Fig. 2).

Visual prompts in the TEDx video have a delay with respect to embodied modes, as they entail editing decisions taken during the recording and not during postproduction. A similar type of incoherence is found in the *mise-en-scène* when the TEDx video uses a long shot in which a slide can be seen in the background which does not necessarily complement or support what the speaker is saying at that moment. No examples of these mismatches are found in the PBS Space Time clips. In fact, visual prompts seamlessly cohere with embodied modes such as spoken language. Fig. 3, for example, shows moments in which the speaker explains the origin of multi-cellular life. While in TEDx the slide on the background just shows a question mark, the visual prompts in PBS clearly support, and even duplicate, the information conveyed through spoken language.

What these results suggest is that editing and production are more planned and professionalized in the PBS Space Time video. The production of the TEDx video is limited in terms of capturing the simultaneity of modes and the complexity of the full modal ensemble. To mention two illustrative examples, capturing the slides in the background and speaker at the same time requires a very long shot which makes embodied modes less effective, and a close up on slides means leaving out the presenter. PBS Space Time clips can overcome these limitations. For example, gestures are intentionally performed around the upper area of the torso to keep them under camera frame. In addition, the production and editing of PBS Space Time are significantly more complex. This has been referred to by Muñoz Morcillo et al. (2016) as moderate complexity of production and high complexity of montage. As indicated in the previous section, the PBS Space Time video includes sound effects, visual effects and cuts to ensure a lively rhythm, none of which are present in the TEDx video.



Fig. 2. Coherent use of metaphoric gestures.

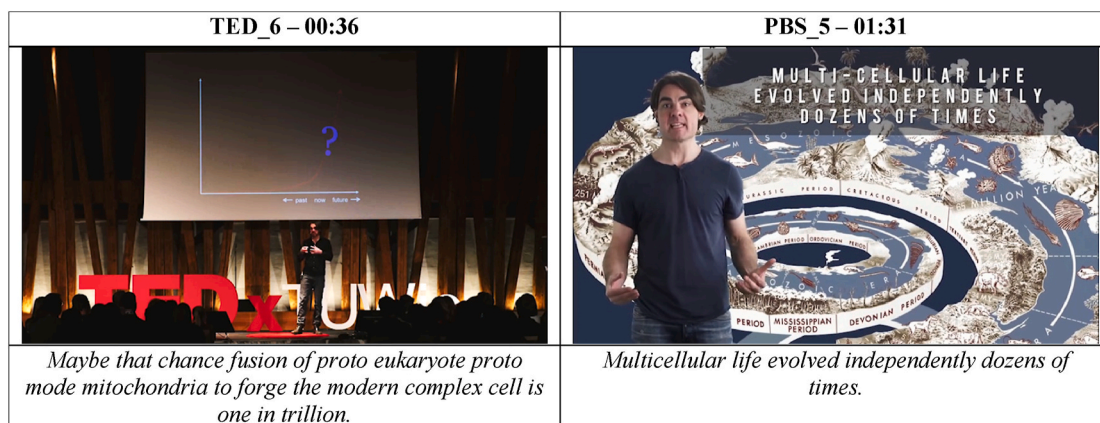


Fig. 3. (In-)coherent use of visual prompts.

3.3. Implications for genre analysis

Lastly, the third research question aims at deriving implications for genre analysis and the definitions of the genres at stake. The findings discussed in the previous sections can help us inform and refine the definitions of TED and PBS science videos as genres. Taking into account [Shepherd and Watters's \(1998\)](#) classification of digital genres and the results of our multimodal analysis, we can describe these genres as follows:

The PBS Space Time video, as hypothesized in the introduction, is a novel genre that originates directly in the online environment facilitated by YouTube. As such, this video makes good use of all the affordances of its medium and exploits both embodied and filmic modes to the fullest. The video is not only denser in terms of number and intensity of modes used, but also more coherent in its orchestration of modes.

The TEDx video, on the other hand, seems to be a good example of a replicated genre evolving from the live TED talk that is using this new medium to reach a wider audience. However, this adaptation to the online setting is not complete and it entails several drawbacks. Thus, although the video incorporates filmic modes such as different camera shots and angles, their use is significantly more modest than in the PBS video. This evidently shows that filmic modes are not an integral part of the genre since they are not considered from the onset, and, as a consequence, their use is less coherent and contributes less to the communicative aim of the genre.

Understanding the use of filmic modes becomes essential for the producer of research dissemination videos to ensure a coherence and effective use of modes in multimodal ensembles. All in all, the TEDx video may be seen as being at an earlier stage of development as a digital genre. It can be argued that one reason for this is that they are “recordings of live events”. It can further be argued that, to become fully developed, digital genres should be conceived as digital content from the start and oriented to an online audience. In short, the two genres are affected by the online medium and its affordances. The online medium (YouTube) serves just as a mere vehicle for the dissemination of science to maximize its impact. It can also be argued that the PBS video is actively creating a certain type of viewer, which will become a trait of the genre as it develops into a well-established digital genre. Both are topics that deserve further research.

4. Conclusion

In this paper we have analyzed two examples of online science videos in which scientific content is recontextualized and adapted to a new audience and medium: a recorded TEDx talk and a PBS Space Time video. We have approached these videos from a multimodal discourse analysis perspective to unveil the complex orchestration of modes that takes place in them. Drawing from multimodal studies of films, we have determined the existence of two layers of modes in online science videos, which we have called embodied and filmic modes, respectively. The former are deployed by the speaker, while the latter are added to the ensemble during production and editing. Our analysis sheds light on three main aspects of this multimodal orchestration: i) the way it affects modal density, ii) the way it affects modal coherence, iii) what this can tell us about TED and PBS videos as genres.

Our findings show that online science videos are more modally dense as a result of filmic modes being incorporated into the ensemble. In addition, we also find that the use of filmic modes requires some degree of expertise and know-how in order to keep the coherence of the multimodal ensemble. The PBS Space Time video that is part of our data reveals itself as denser and more coherent than the TED video. In the case of the TED video, we note a modest use of production and editing, and a lack of coherence between embodied and filmic modes. In addition, the multimodal ensembles deploy a use of embodied modes that is perceived as addressed to the audience that is physically present rather than to the online audience, which might negatively affect how meaning reaches the viewer. In fact, a clear acknowledgement of the recording devices in the conference room can increase engagement in the viewers as suggested by [Xia and Hafner \(2021\)](#). Having this idea in mind, we believe that research on the multimodal deployment of engagement strategies in online videos would yield enlightening results in further research. Likewise, we suggest that a higher modal density does

not necessarily imply a higher cognitive load if there is modal coherence, because all modes contribute seamlessly to the same communicative effect. In fact, there are results from previous literature that point in this direction (Chambers, 2001; Muñoz Morcillo et al., 2016; Welbourne & Grant, 2016). These findings confirm our hypothesis that the two genres are at a different stage of development in Shepherd and Watters's (1998) classification of cybergenres. The PBS Space Time video is a novel cybergenre fully adapted to the YouTube medium and exploiting all its affordances. The TED video, on the other hand, is a replicated genre that is still adapting to the new medium and makes a less efficient use of the filmic modes afforded by YouTube.

One of the limitations of the study is the absence of input from online viewers, which might complement our interpretations. Furthermore, given the limited size of the dataset, the results of this study cannot be generalized. Still, they can pave the way for larger scale studies, which could include, among others, a more rigorous linguistic analysis as part of the mode *spoken language*, or the analysis of full videos in order to fully grasp the macrostructure of the genre. These can provide a deeper understanding of the fast changes originated by the ever-growing trend towards science dissemination and popularization. These changes, as shown in this paper, go beyond the need to recontextualize scientific content. They also involve the use of new media with new semiotic modes and affordances, as well as new genres. In this scenario, multimodal literacy becomes a crucial skill among science communicators, who now more than ever need to skillfully orchestrate increasingly complex and multi-layered multimodal ensembles.

Authors statement

Dr. Valeiras-Jurado was in charge of formal analysis, investigation, methodology, writing, reviewing and editing. Dr. Bernad-Mechó was in charge of conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, writing, reviewing and editing.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Acknowledgements

This study is encompassed within a research project funded by the Ministerio de Ciencia, Innovación y Universidades, Spain (Ref: PGC2018-094823-B-I00) and entitled *Análisis de los géneros académicos en la docencia en inglés de grados internacionales desde una perspectiva multimodal y multicanal*. The analysis conducted in the study was also possible thanks to grant E-2020-31 awarded by Universitat Jaume I to Dr. Bernad-Mechó. Funding for open access charge: CRUE-Universitat Jaume I.

References

- Allgaier, J. (2020). Science and medicine on YouTube. In J. Husinger, M. M. Allen, & L. Klustrup (Eds.), *Second international handbook of Internet research* (pp. 7–27). Springer.
- Bernad-Mechó, E. (2021). Combining multimodal techniques to approach the study of academic lectures: A methodological reflection. *Atlantis. Journal of the Spanish Association of Anglo-American Studies*, 43(1), 178–198. <https://doi.org/10.28914/Atlantis-2021-43.1.10>
- Bernad-Mechó, E., & Fortanet-Gómez, I. (2019). Organizational metadiscourse across lecturing styles: Engagement beyond language. In C. Sancho-Guinda (Ed.), *Engagement in professional genres: Deference and disclosure* (pp. 321–340). John Benjamins.
- Bezemer, J., & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication*, 25(2), 166–195.
- Bowen, C. J. (2018). *Grammar of the edit* (4th ed.). Routledge.
- Boy, B., Bucher, H.-J., & Christ, K. (2020). Audiovisual science communication on TV and YouTube. How recipients understand and evaluate science videos. *Frontiers in Communication*, 5, 1–18.
- Burgess, J., & Green, J. (2009). *YouTube: Online video and participatory culture*. Wiley.
- Burn, A. (2014). The kineikonic mode: Towards a multimodal approach to moving image media. In C. Jewitt (Ed.), *The routledge handbook of multimodal analysis* (2nd ed., pp. 375–385). Routledge.
- Caliendo, G., & Compagnone, A. (2014). Expressing epistemic stance in university lectures and TED talks: A *Contrastive Corpus-Based Analysis*, 11, 105–122.
- Chambers, H. E. (2001). *Effective communication skills for scientific and technical professionals*. Perseus Publishing.
- Dancyger, K. (2011). *The technique of film and video editing: History, theory, and practice*. Elsevier.
- Desilla, L. (2012). Implicatures in film: Construal and functions in Bridget Jones romantic comedies. *Journal of Pragmatics*, 44, 30–53.
- Erviti, M. C., & Stengler, E. (2016). Online science videos: An exploratory study with major professional content providers in the United Kingdom. *Journal of Science Communication*, 15(6), 1–15.
- Farias, C., Seremani, T., & Fernández, P. D. (2021). Popular culture, moral narratives and organizational portrayals: A multimodal reflexive analysis of a reality television show. *Journal of Business Ethics*, 171, 211–226.
- Frobenius, M. (2014). Audience design in monologues: How vloggers involve their viewers. *Journal of Pragmatics*, 72, 59–72.
- Geipel, A. (2018). Wissenschaft@YouTube. Plattformspezifische Formen von Wissenschaftskommunikation. In E. Lettkemann, R. Wilke, & H. Knoblauch (Eds.), *Knowledge in action. Neue Formen der Kommunikation in der Wissensgesellschaft* (pp. 137–163). Springer.
- González García, R., Serrano Pascual, L., & Roca Martín, D. (2020). Perfil de las youtubers españolas dedicadas a la divulgación científica. In A. Castro Higuera, A. Sandulescu Budea, & R. García Orellán (Eds.), *Fórmulas comunicativas de vanguardia* (pp. 403–418). Ediciones Pirámide.
- Jewitt, C. (2004). Multimodality and new communication technology. In R. Scollon, & P. LeVine (Eds.), *Discourse and technology: Multimodal discourse analysis* (pp. 184–195). Georgetown University Press.
- Jimenez Hurtado, C., & Soler Gallego, S. (2013). Multimodality, translation and accessibility: A corpus-based study of audio description. *Perspectives*, 21(4), 577–594.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Kousha, K., Thelwall, M., & Abdoli, M. (2012). The role of online videos in research communication: A content analysis of YouTube videos cited in academic publications. *Journal of the American Society for Information Science and Technology*, 63(9), 1710–1727.
- Kozloff, S. (2000). *Overhearing film dialogue*. California: University of California Press.
- Kress, G. (2003). *Literacy in the new media age*. Routledge.
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

- Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- León, B., & Bourk, M. (2018). Investigating science-related online video. In B. León, & M. Bourk (Eds.), *Communicating science and technology through online video* (pp. 1–14). Routledge.
- Ludewig, J. (2017). TED talks as an emergent genre. *CLCWeb: Comparative Literature and Culture*, 19(1), 1–9.
- Luzón, M. J. (2019). Bridging the gap between experts and publics: The role of multimodality in disseminating research in online videos. *Iberica*, 37, 167–192.
- Maingueneau, D. (2010). Types of genres, hypergenre and Internet. In M. J. Luzón, M. N. Ruiz-Madrid, & M. L. Villanueva (Eds.), *Digital genres, new literacies and autonomy in language learning* (pp. 23–42). Cambridge Scholars Publishing.
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge University Press.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. The University of Chicago Press.
- Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology*, 34, 615–624.
- Muñoz Morcillo, J., Czurda, K., & Robertson-von Trotha, C. Y. (2016). Typologies of the popular science web video. *Journal of Science Communication*, 15(4), 1–32.
- Norris, S. (2004). *Analyzing multimodal interaction. A methodological framework*. Routledge.
- O'Halloran, K., Podlasov, A., Chua, A., & E, M. K. L. (2012). Interactive software for multimodal analysis. *Visual Communication*, 11(3), 363–381.
- Osterrieder, A. (2013). The value and use of social media as communication tool in the plant sciences. *Plant Methods*, 9(26).
- Pérez-Llantada, C. (2021). Genres and languages in science communication: The multiple dimensions of the science-policy interface. *Language & Communication*, 78, 65–76.
- Pérez-Torres, V., Pastor-Ruiz, Y., & Abarrou-Ben-Boubaker, S. (2018). YouTuber videos and the construction of adolescent identity. *Comunicar. Revista Científica de Comunicación y Educación*, 55, 61–70.
- Rowley-Jolivet, E., & Carter-Thomas, S. (2019). Scholarly soundbites. Audiovisual innovations in digital science and their implications for genre evolution. In M. J. Luzón, & C. Pérez-Llantada (Eds.), *Science Communication on the Internet: Old genres meet new genres* (pp. 81–106). John Benjamins.
- Rowsell, J. (2014). The mood is in the shot": The challenge of moving-image texts to multimodality. *Text & Talk*, 34(3), 307–324.
- Saurabh, S., & Gautam, S. (2019). Modelling and statistical analysis of YouTube's educational videos: A channel owner's perspective. *Computers & Education*, 128, 145–158.
- Scotto di Carlo, G. (2014). Ethos in TED talks: The role of credibility in popularised texts. *Facta Universitatis – Series: Linguistics and Literature*, 12(2), 81–91.
- Scotto di Carlo, G. (2015). Stance in TED talks: Strategic use of subjective adjectives in online popularisation. *Iberica*, 29, 201–221.
- Shepherd, M., & Watters, C. (1998). The evolution of cybergenres. In , Vols. 6–9. *Proceeding HICSS '98. Proceedings of the thirty-first annual Hawaii international conference on system sciences - Volume 2. January* (pp. 97–109).
- Smith, S. M., & Shaffer, D. R. (1995). Speed of speech persuasion: Evidence for multiple effects. *Personality and Social Psychology Bulletin*, 21, 1051–1060.
- Snickars, P., & Vonderau, P. (2009). *The YouTube reader*. Logotipas.
- Soukup, P. A. (2014). Looking at, with, and through YouTube. *Communication Research Trends*, 33(3), 3–34.
- Sugimoto, C. R., & Thelwall, M. (2013). Scholars on soap boxes: Science communication and dissemination in TED videos. *Journal of the American Society for Information Science and Technology*, 64(4), 663–674.
- Thelwall, M., Kousha, K., Weller, K., & Puschmann, C. (2012). Assessing the impact of online academic videos. In G. Wildén, & K. Holmberg (Eds.), *Social information research* (pp. 195–213). Emerald Group Publishing.
- Tseng, C. (2008). Coherence and cohesive harmony in filmic text. In L. Unsworth (Ed.), *Multimodal Semiotics: Functional analysis in contexts of education* (pp. 87–104). Continuum.
- Tseng, C., & Bateman, J. A. (2010). Chain and choice in filmic narrative: An analysis of multimodal narrative construction in "The Fountain. In C. R. Hoffman (Ed.), *Narrative revisited: Telling a story in the age of new media* (pp. 213–244). John Benjamins.
- Valeiras-Jurado, J. (2019). Modal coherence in specialised discourse: A case study of persuasive oral presentations in business and academia. *Iberica*, 37, 87–114.
- Vanoye, F. (1985). Conversations publiques. *Iris*, 3(1), 99–188.
- Welbourne, D. J., & Grant, W. J. (2016). Science communication on YouTube: Factors that affect channel and video popularity. *Public Understanding of Science*, 25(6), 706–718.
- Werner, E. A. (2012). *Rants, reactions, and other rhetorics: Genres of the YouTube vlog*. Unpublished PhD Thesis. University of North Carolina at Chapel Hill.
- Xia, S. A., & Hafner, C. A. (2021). Engaging the online audience in the digital era: A multimodal analysis of engagement strategies in TED talk videos. *Iberica: Revista de la Asociación Europea de Lenguas para Fines Específicos*, 42, 33–58. <https://doi.org/10.17398/2340-2784.42.33>

Dr. Julia Valeiras-Jurado is Assistant Professor at Universitat de València, Spain. Her research interests are multimodality, academic and business discourse, oral genres and persuasive language and ESP. Her most recent publications have appeared in *Text&Talk*, *International Journal of English Studies* and *Language and Communication*.

Dr. Edgar Bernad-Mechó is Assistant Professor at Universitat Jaume I, Spain. He specializes in the use of metadiscourse and multimodality in academic contexts. His latest publications include a paper in *ESP Today*, a methodological article approaching the multimodal analysis of lectures in *Atlantis*, and the co-edition of a special issue on multimodality for *Iberica*.