Multimodality as a framework for the study of verbal and non-verbal resources in web-based science communication. The case of science crowdfunding videos (SCVs)

Ana Cristina Vivas-Peraza
University of Zaragoza, Spain   ✉ ⓘ

ᴱᴺᴳ Abstract: In a multimodal approach to communication, language is just one of many communicative modes that can be used to deliver a message and the overall meaning of a semiotic product is conveyed by the orchestration of those modes into multimodal ensembles (i.e., combinations of semiotic modes). Multimodality studies are gaining recognition in ESP, particularly in the study of digital genres for science popularisation. These genres are web-based and rely on multiple semiotic resources afforded by digital technologies to disseminate scientific knowledge to lay audiences, such as speech, gestures, images, sound, and camera effects. This paper explores multimodal communication in science crowdfunding videos (SCVs), a digital genre intended to get the Internet audience involved in a research project. Drawing on a case study of 3 SCVs from *Experiment.com* and qualitative interviews to their author- scientists, the results show how the engagement of online audiences is achieved by combining language with a variety of non-verbal resources mediated through the scientist's body language and through video edition, such as the embedding of images and video footage, and the choice of camera shot and angle. Moreover, the study reveals eight multimodal ensembles that enhance audience engagement through the scientist's presence, text, and visual aids.
Key words: multimodality; science crowdfunding videos; digital genres; engagement.

Contents: 1. Introduction. 2. Multimodality and engagement. 3. The Science Crowdfunding Video (SCV). 4. Materials and Methods. 4.1. Selection of three SCVs for the study. 4.2. Multimodal analysis. 5. Results and discussion. 5.1. Constructing a competent and enthusiastic researcher identity through ME1, ME2 and ME3. 5.1.1. Researcher's self-introduction in SCV3. 5.1.2. Final call to action in SCV2. 5.2. Seeking common ground through ME4. 5.2.1. Opening and closing slides in SCV1. 5.3. Describing the research through ME5, ME6, ME7, and ME8. 5.3.1. Presenting the research project in SCV2. 5.3.2. Seeking compassion in SCV1. 5.3.3. Mixing popular culture images and specialised images in SCV3. 6. Conclusions. Acknowledgements. Informed consent statement. References. Appendix A.

1. Introduction

During the past decades, online science communication has become a major skill among scientists to share their scientific discoveries and gain society's trust and involvement in science, consistent with the ubiquitous Open Science movement (Luzón & Pérez-Llantada, 2019, 2022). The Internet has facilitated *science popularisation*, or the dissemination of specialised knowledge to lay audiences (Xia & Hafner, 2022); and as a result, new digital genres for science popularisation have emerged on the web, becoming a subject for academic discussion among ESP (English for Specific Purposes) researchers. Some examples of these genres are online science dissemination videos (Bernad-Mechó & Valeiras-Jurado, 2023; Luzón, 2019; Xia & Hafner, 2022), promotional online academic trailers (Maier & Engberg, 2023), and Twitter conference presentations (Villares, 2023). Science popularisation videos streamed through web platforms such as Youtube

(Bernad-Mechó & Valeiras-Jurado, 2023; Luzón, 2019) or TED talks[1] (Xia & Hafner, 2022) are intended to make scientific content accessible and comprehensible to all types of audiences through a range of aural, visual, and verbal modes of communication. In online academic trailers (Maier & Engberg, 2023), scientists also use the multimodal video medium to fulfil a double persuasive purpose: to popularise research work and to promote an academic course that is created on the basis of such work. And in the case of Twitter conference presentations (Villares, 2023), these consist in six-tweet threads shared by academics to disseminate their research projects with diversified audiences (i.e., expert and non-expert audiences, Luzón, 2013), for which they also orchestrate different semiotic modes to communicate and to establish a balance between academic and informal registers.

What these genres have in common is that they are intended to engage the public in scientific issues and that they are inherently multimodal, as they communicate meaning through a variety of semiotic modes ranging from the verbal modes of speech and writing, to the non-verbal modes of image, sound, gestures, and camera effects, to mention a few. Scientists producing these genres are thus challenged to make scientific content engaging to all types of audiences by making the best use of multimodal resources that digital media puts at their disposal (Bernad-Mechó & Valeiras-Jurado, 2023). Similarly, genre analysts, traditionally focused only on the linguistic properties of genres, have been challenged to consider modes of communication other than speech and writing to fully understand these new forms of digitally-mediated science communication, for which they are adopting multimodal approaches (Askehave & Nielsen, 2005; Bateman, 2014; Xia, 2020). The number of multimodal studies in ESP has increased with the advent of digital genres, and as Xia and Hafner (2022) recommend, they should continue to grow to understand new trends in science communication.

An emerging digital form for science popularisation that has received little attention is a genre that I will refer to as *science crowdfunding video* (henceforth, SCV). SCVs consist in a short video through which scientists promote a research project in need of funding, in order to get the public involved either by donating money or by sharing the proposal through social networks (Hui & Gerber, 2015). Similar to other science popularisation genres, such as science dissemination videos (Bernad-Mechó & Valeiras-Jurado, 2023; Luzón, 2019; Xia & Hafner, 2022) or academic trailers (Maier & Engberg, 2023), they are digitally mediated and exploit the multimodal affordances of the edited video to engage online audiences. That is, scientists producing SCVs also face the challenge to reach out to the public through this multimodal artefact and need to make semiotic decisions on the multimodal resources to be used and how these can be combined in the most effective fashion. Against this background, the present study investigates how different modes are orchestrated in this genre to engage online audiences and get them involved in a research project. This aim can be summarised in the following research questions:

i.   What semiotic modes are used in SCVs to engage the audience?
ii. How are these semiotic modes combined into multimodal ensembles?

These questions will be addressed by selecting a sample of three SCVs from the platform *Experiment. com*, on which multimodal analyses will be conducted (see Section 4 for more details). The results obtained could be of pedagogical application to ESP courses intended to improve scientists' communication skills, as will be stressed in the concluding section.

The paper is organised as follows. Section 2 focuses on multimodality and engagement, both core aspects in the study of science popularisation. Section 3 presents the genre of SCVs, which is the object of this study. In Section 4, the materials and methods used to carry out the analysis will be explained. The findings will be shown and discussed in Section 5, and conclusions about such findings will be drawn in Section 6.

2. Multimodality and engagement

Multimodality and engagement are key factors in the study of digital genres for science popularisation. Kress and van Leeuwen (2001) define *multimodality* as "the use of several semiotic modes in the design of a semiotic product or event, together with the particular way in which the modes are combined" (p. 10). In multimodal communication, a *mode* constitutes an "organised set of semiotic resources" that "are recognised within a community as realising meaning" (Jewitt et. al., 2016, p. 157). Some examples of modes are speech, writing, static and moving images, hand gestures, gaze, camera shot and angle, and music. Each of these modes convey meaning on their own and can contribute to the creation of meaning in an "integrated, multimodal whole" (Jewitt et.al, 2016, p. 2). In other words, a multimodal analysis of a communicative event involves not only the identification and interpretation of each individual meaning-making resource used, but also how and to what extent each of those resources contributes to the construction of meaning in a *multimodal ensemble* (Jewitt, 2009).

Multimodality lies on the assumption that each semiotic mode has different affordances and limitations (Kress, 2010; Jewitt, 2013), and consequently, some meanings are expressed and represented in some modes better than in others (Bezemer & Jewitt, 2010). Images, for instance, can convey much information at a glance, whereas written language (also perceived through sight) requires harder and longer cognitive

---

[1]   TED (Technology, Entertainment and Design) talks consist in a recorded public-speaking presentation intended to share knowledge in a variety of topics and educate global audiences in an accessible way. They are also used for science dissemination (Xia & Hafner, 2022).

effort as the reader interprets the words and sentences that constitute a message (Rowley-Jolivet, 2004). At the same time, as Stöckl (2004) points out, images are stronger in connotation than words and can connect faster to the viewers' emotions. On the other hand, writing can represent ideas and processes that are more difficult for images to represent, such as the description of emotions or experiences that involve human senses, such as smell or touch. Since modes have different semiotic possibilities, they are often orchestrated into multimodal ensembles to make up for the limitations that each of them may present, in order to enhance communication and achieve a greater engaging effect on the audience. This way, the relation between modes within a multimodal ensemble can be *elaborative* or *extensive* (van Leeuwen, 2005).

A mode is considered elaborative if it summarises, details, and reformulates the meaning conveyed by another mode in order to restate or reinforce such meaning. On the other hand, an extensive mode adds new meaning to the one conveyed by another mode. By exploring these modal relations, the multimodal analyst can observe the semiotic potentialities of each mode used in an artefact and examine whether they have been arranged coherently in a multimodal ensemble. According to Valeiras-Jurado and Bernad-Mechó (2022), *modal coherence* is achieved in a multimodal ensemble if all modes "are used in a consistent way and do not contradict each other" and "contribute seamlessly to the communicative aim at hand" (p. 3). By observing modal coherence, an insight can be gained into modal affordances, multimodal design and the modal choices made by the *sign-maker* (i.e., the producer of semiotic material, Kress, 2010).

The study of how different semiotic modes communicate is nothing new. Modes conveyed through body language and voice have been studied as far back as Ancient Greece, when rhetoricians used to train politicians, priests, and lawyers on how to make their public speeches more engaging with the correct pronunciation, voice articulation, and body posture (Spang, 2005; Pullman, 2013). And since then, these modes, typically referred to as *non-verbal communication*, have attracted the attention of psychologists and anthropologists, and more recently, linguists (Cestero-Mancera, 2017, 2018, 2019; Cestero-Mancera & Díez-Prados, 2021, Poyatos, 1998, 2002). In the same manner, images (part of human culture longer than script) have been the object of study of disciplines such as Art History over centuries (Kress, 2010), with contemporary linguistic accounts on the interpretation of images in combination with writing, such as Kress & van Leeuwen's (1996, 2006, 2021) ground-breaking work on visual grammar. Body language and image, as well as other meaning-making resources such as music (van Leeuwen, 1999), have been a subject of interest in different academic and non-academic disciplines for ages. The term 'Multimodality', however, was only used for the first time in 1990 (Jewitt et al., 2016), when it emerged as "an attempt to bring all means of meaning making together under one theoretical roof, as part of a single field in a unified account, a unifying theory" (Kress, 2010, p. 5). Multimodality accounts not only for combinations of particular modes (image and writing, or language and gestures) but for all the verbal and non-verbal modes combined in a semiotic artefact.

With the explosion of Internet-mediated genres, interest in multimodal research has increased among ESP scholars to account for the great variety of semiotic modes that the web offers for the dissemination of science. In particular, attention has been paid on the multimodal strategies that scientists use to *engage* lay audiences in these genres, a practice that in academic discourse involves "recognising the presence of their readers, pulling them along their argument, focusing their attention, and (…) guiding them to interpretations" (Hyland, 2005, p. 176). As far as video genres are concerned, several recent studies on multimodal engagement are noteworthy. In her study on online popular science videos, Luzón (2019) examined the multimodal strategies used by scientists to recontextualise specialised knowledge for Internet audiences, noting the communicative potential of visual modes to popularise science. Xia and Hafner (2022) identified five strategic multimodal configurations to increase audience engagement in TED talks through the combination of different visual, linguistic, gestural, and spatial modes. In a more recent multimodal study on Youtube dissemination videos, Bernad-Mechó and Valeiras-Jurado (2023) showed the importance of non-verbal resources controlled during the production and editing process (i.e., what they call "filmic modes", p. 305) to attract the attention of viewers in the introductory section and highlighted the importance of well-orchestrated multimodal ensembles to engage online audiences. As for Maier and Engberg (2023), they explored the generic structure of academic trailers to observe the multimodal strategies used to achieve knowledge dissemination and promotion. These studies have served as an inspiration for the analysis of multimodal engagement strategies in SCVs.

3. The Science Crowdfunding Video (SCV)

Similar to online academic trailers (Maier & Engberg 2013), scientists producing SCVs exploit the multimodal affordances of the video to fulfil a twofold aim: to disseminate their research work online and to engage the Internet audience to finance a project derived from such work. The practice of science crowdfunding offers an alternative or additional funding strategy to the traditional application for national research grants. The objective of both procedures is similar in that scientists are seeking for research funding; however, they are different in terms of intended audiences (Mehlenbacher, 2019). Whereas grant applications are addressed to national granting institutions and evaluated by a panel of experts in the research field in question; science crowdfunding involves the engagement of online audiences with different levels of expertise. Crowdfunding thus challenges scientists to persuade the public to participate in the funding of their projects, and to make a wise choice of semiotic resources to achieve that goal.

SCVs are embedded within a science crowdfunding proposal, which scientists can create in web-based crowdfunding platforms. Such proposal presents the characteristics of the research project to be funded through two main elements: a textual description (which includes images), and the SCV itself. Previous genre

studies on science crowdfunding proposals have only focused on linguistic features of the textual description (Mehlenbacher, 2017, 2019; Pérez-Llantada, 2021; Vela-Rodrigo, 2023) to examine how the specialised language of science is imported into this genre to engage non-specialised audiences. As for the SCV, to the best of my knowledge, the only two previous genre studies have been conducted by Vivas-Peraza (2022a, 2022b) on videos taken from the platform *Experiment.com*. In my first study, I examined the representation of gender through verbal and visual resources in eight SCVs crafted by undergraduate students and found significant differences between the depiction of female and male scientists that contribute to the reinforcement of gender bias and the underrepresentation of women in STEM disciplines (Vivas-Peraza, 2022a). And in a later study, I collected a larger sample of SCVs made by novice and senior scientists to observe the visual and verbal engagement strategies used to call the audience to action (Vivas-Peraza, 2022b). This latter study only focuses on the final section of the videos (where that call to action is made) and interprets semiotic modes individually without observing how these are orchestrated into multimodal ensembles. This paper, therefore, serves as an extension of this second study and, by adopting multimodality as a methodological framework, examines the semiotic modes used in the whole video and on how they can be orchestrated into multimodal ensembles to make scientific content engaging to lay audiences.

In the science-specialised online crowdfunding platform *Experiment.com*, proposals can be submitted with their respective textual descriptions and SCVs. SCVs are optional; however, the very platform encourages scientists to include one with the textual description of the project, since "projects with videos are 60% more likely to be fully funded, and are shared twice as much" (Experiment, 2023, n.p.). This shows the persuasive potential that the videos present and the positive impact they can have on the success of the scientific campaign (Vachelard et al., 2016). Furthermore, the platform also establishes some guidelines for the production of SCVs. On the one hand, videos should be between two and three minutes long maximum. On the other hand, they should include a summary of the research project in question and a clear statement on why the project deserves funding (Experiment, 2023). According to Vivas-Peraza (2023a), SCVs typically start with an introduction of the researcher, followed by a presentation of the research project (background, goals, method, significance, etc.), and end with a final call to action (p. 11). And, since videos can be edited, this can be conveyed through a variety of semiotic possibilities, such as images, video footage, spoken narrations, sound effects, and written text. As Vivas-Peraza (2023b) observed, this flexibility in terms of audio-visual content leads to different types of SCVs, from voice-over slideshows presenting the science behind a project, to recordings in which the very scientist appears on screen to introduce themselves and 'sell' their research idea to the public, similar to what business entrepreneurs do to sell their products or services in popular elevator pitches (Daly & Davy, 2016; Díez-Prados, 2019).

When the scientist is visible in SCVs, they appear either throughout the whole video or at different stages of the video while other stages display voice-over images and video sequences of, for example, the research object. In her latest study, Vivas-Peraza (2023b) focused on SCVs in which the scientist appears on screen at the end of the video to call the audience to support the project, in order to observe the visual and verbal strategies used to engage the audience in that final section. She found that in order to promote intimacy and dialogic involvement with the audience, scientists often use second person pronouns (Hyland, 2005), maintain eye-contact and smile, use a close camera shot (Kress & van Leeuwen, 2021), and choose their research workspace to film themselves in. In addition, she observed a frequent use of directives to instruct the audiences to perform the desired action (i.e., to donate money or diffuse the campaign) and a tendency to evoke positive feelings about the research project in question through the use of evaluative language. This study shows common semiotic modes used to engage the audience in the last section of the SCV. However, it overlooks other relevant modes used in other sections of the video that involve the use of static and moving images or text, as well as the interplay of modes in multimodal ensembles. The present study aims to shed light on these other multimodal phenomena.

4. Materials and methods

This paper examines different semiotic modes employed by scientists to engage the audience in different sections of the SCV, as well as to observe how those modes are combined in multimodal ensembles. In order to do that, multimodal analysis was undertaken on a small sample of three SCVs collected from the science-specialised crowdfunding platform *Experiment.com*, and qualitative interviews were conducted to the author-scientists of those videos to triangulate the results.

4.1. Selection of three SCVs for the study

The sample of SCVs is small because, as Jewitt (2009) states, "multimodal research can be applied to take a detailed look at 'big' issues and questions through specific instances" (27), meaning that the study of multimodal communication involves fine-grained analyses that are impossible to be carried out in a large number of artefacts. Table 1 provides details of the three selected SCVs, including the release date, project title, authorship (pseudonyms), and web link where they can be found online.

SCV1 deals with a Medicine project aiming to find a vaccine to protect new-born babies from Zika virus transmission from mother to child in countries with a high prevalence of this disease. Among the scientists in charge of this experiment, it is Sunan Wongyai, the interviewed scientist. SCV2 focuses on a Biology and Ecology project for the conservation of the Palo Santo tree species and is authored by Emily Crawford. As for SCV3, it also promotes a research project on Biology and Ecology that intends to use squid as an animal host to study innate immunity in space to understand astronauts' immune system dysregulations in long-term space travel better. This latter SCV was produced by two researchers, among whom is Anna Dawson.

Table 1. SCVs selected for multimodal analysis.

SCVx	Date	Duration	Project title	Author/s	Web link
SCV1	15/10/20	*2:58*	*Modeling Zika virus transmission from mother to child using uterine mini-organs.*	Sunan Wongyai (and colleagues)	https://experiment.com/projects/modeling-zika-virus-transmission-from-mother-to-child-using-uterine-mini-organs
SCV2	04/11/19	*1:33*	*Does the palo santo tree consist of multiple distinct species?*	Emily Crawford	https://experiment.com/projects/does-the-palo-santo-tree-consist-of-multiple-distinct-species
SCV3	18/01/18	*3:23*	*Squid in Space: Symbiosis and Innate Immunity.*	Anna Dawson (and colleagues)	https://experiment.com/projects/squid-in-space-symbiosis-and-innate-immunity

These three SCVs were picked from a larger research project that sampled 50 SCVs in English in the fields of Biology, Ecology, and Medicine from 2017 to 2021 (Vivas-Peraza, 2023b). In this larger sample, the semiotic modes used to engage the audience were identified and quantified. This was done with qualitative analysis software *Atlas.ti 8*, which allows the segmentation and codification of large amounts of multimedia data (Friese, 2019). With this software, the videos were first segmented into the different sections that conform them (i.e., introduction, presentation of research project and call to action), and then, codes referring to the observed semiotic manifestations of engagement were used to code those segments (see Figure 1). For instance, if in the opening section of the video (previously marked as a segment), the use of a close shot to engage the audience was observed, this opening section would be codified with the code "Close Shot", and with as many other codes as semiotic modes would be observed in it.

Figure 1. Interface of Atlas.ti 8 showing SCV2 segmented and coded.
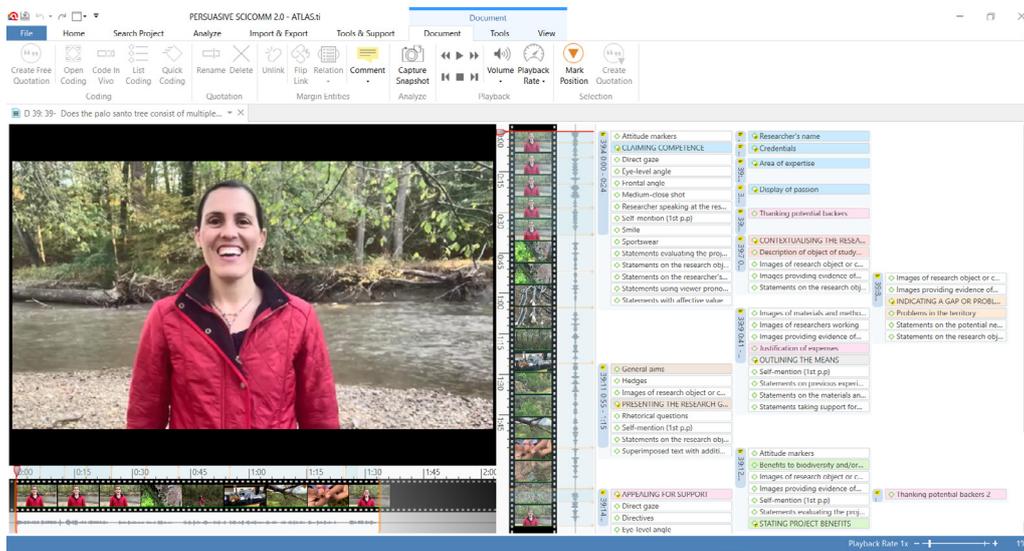
After coding the 50 SCVs, the *Atlas.ti* tool "Code-document table" was utilised to obtain a list of the SCVs with the total number of semiotic modes used in them. This list was ordered from the highest to the lowest number of semiotic modes and three of the top 10 were selected for in-depth multimodal analysis. These three were selected on the basis of two main criteria: (i) that they were semiotically rich (i.e., at the top of the list), (ii) and that their authors agreed to a personal interview to discuss their semiotic choices. The first criterion was based on the hypothesis that a deeper and more informative study of multimodal engagement would arise from videos that combined a larger number of semiotic modes. As for the second criterion, it was established to increase the reliability and validity of the results, since, by asking scientists about their semiotic choices in their videos, the interpretation of the semiotic modes in this study would detract from subjectivity. Thus, the multimodal analyses were carried out on three of the top-ranked SCVs whose authors would agree to do the interview, namely, SCV1, SCV2, and SCV3 (see Table 1).

4.2. Multimodal analysis

As mentioned before, multimodal methodology involves thorough analysis of a reduced number of cases (Jewitt, 2009), so the results of this study are not intended to provide quantitative data but rather an insight

into scientists' semiotic choices and combinations to engage the audience in different sections of the SCV. Once the three videos were selected for analysis, they were watched multiple times to observe the semiotic interactions occurring in each of the three main sections, i.e., introduction of the researcher, description of the research project, and call to action (Vivas-Peraza, 2022a). For each section of each video, I produced a *multimodal transcript*, which consists of a written detailed description of a selected range of semiotic modes co-occurring at different stages of a multimodal artefact (Jewitt, 2009). Following similar analytical procedures to those followed by Baldry & Thibault (2006), Hampel & Hauck (2006) and Lancaster & Roberts (2006), I presented each multimodal transcription with a matrix representing temporal and spatial sequences of modal interactions, with speech (or writing, in the absence of speech) used as the "anchor mode", or the mode against which the other co-occurring modes are observed (Jewitt, 2009, p. 48).

As can be seen in Appendix A, a table (i.e., a grid of cells) was used to create a matrix for each video section. Rows corresponded to the various shots or "continuously filmed stretches of video, without cuts" (Hafner, 2014, p. 666) that make up the section's content, while columns represented the semiotic modes that were employed in those shots. Given that speech was used as a starting point to study the modes that were combined with it, the transcript for each shot was further split into more rows that corresponded to the various utterances used in that shot in order to notice the modes co-occurring with those utterances. Consequently, a shift in utterance or row would occur when a noteworthy mode different that speech would occur. For instance, in sections where the scientist speaks in front of the camera, a new noteworthy body gesture would lead to a change of utterance, which would result in the creation of a new row in the table. And in sections with voice-over images, changes of utterance would be determined by a change of image. Shots with no voice-over would consist of writing and image, and, in this case, writing would be used as the anchor mode to observe other modes converging with it.

The wide array of semiotic resources that multimodal artefacts such as videos present is impossible to be analysed in a single study (Paltridge, 2012; Tseronis & Forceville, 2017), so I decided to examine the following verbal and non-verbal modes (see Table 2): speech, writing, static and moving images, clothing, setting, camera angle and shot, gaze, smile, eyebrow raising, head movements, and hand and arm gestures. These modes were selected because they played a significant role in audience engagement in the three analysed SCVs. The meaning conveyed by the verbal modes were interpreted according to Hyland's (2005) model of interaction in academic discourse. As for the non-verbal modes, Kress & van Leeuwen's (2021) visual grammar was useful to interpret the use of camera shot and angle, images, gaze, and smile; and Cestero-Mancera's (2018) taxonomy of non-verbal resources in persuasive communication was considered to examine eyebrow raising, head movements, and gestures of the scientists. Other semiotic modes involving the use of music and paralinguistic features (i.e., rhythm of speech, intonation, tone, and stress, Poyatos, 1998) were not considered, as their interpretation involves disciplines such as musicology, sound acoustics, phonetics, and phonology (van Leeuwen, 1999), all of which fall outside the theoretical scope of this study.

Table 2. Semiotic modes observed for the multimodal analysis of SCVs.

Modes		Description
Verbal	Speech	Spoken narration made by scientists.
	Writing	Written text appearing on screen.
Non-verbal	Static images	Standstill pictures, maps, diagrams, and logos.
	Moving images	Video footage.
	Clothing	Dress code followed by the scientists while speaking on camera.
	Setting	Background from which the scientist speaks.
	Camera angle	Scientist's standpoint view while speaking (e.g., frontal, oblique).
	Camera shot	Distance between the represented subject and the camera.
	Gaze	Scientists' eye orientation while speaking.
	Smile	Presence or absence of scientist's smile while speaking.
	Eyebrow raising	Scientist's up-and-down eyebrow movement while speaking.
	Head movement	Scientist's head tilts and nods while speaking.
	Hand/ arm gestures	Scientist's hand and arm shakes while speaking.

Once the multimodal transcripts were produced, I identified several multimodal ensembles through which scientists engage the audience in SCVs (see Section 3 in this paper). For each multimodal ensemble, I observed the modes arranged and the relations between those modes to examine modal coherence (i.e., the orchestration of modes in a consistent and non-contradictory manner) and the predominance of certain modes over others for engagement purposes. In order to contrast the results obtained from the analysis

and further interpret scientists' semiotic choices, I interviewed the author scientists of the three videos (i.e., Wongyai, Crawford, and Dawson). During those interviews, the participants and I watched their respective videos together by means of a *playback session* (Norris, 2011, p. 59), during which I would inquire in their communicative intentions when choosing the semiotic modes arranged in each video section. The three interviews were audio-recorded and transcribed orthographically with web-based computer software *Transcribe by Wreally*.

5. Results and discussion

The multimodal transcripts obtained from the analysis of each SCV reveal eight different types of multimodal ensembles, as illustrated in Table 3.

Table 3. Multimodal ensembles found in the three analysed SCVs.

ME* x	Combination of semiotic modes
ME1	Speech, gestural modes, camera shot and angle, setting, and clothing
ME2	Speech, gestural modes, camera shot and angle, setting, clothing, and writing
ME3	Speech, gestural modes, camera shot and angle, setting, clothing, and static images
ME4	Writing and static images
ME5	Speech (voice-over narration) and static images
ME6	Speech (voice-over narration), static images, and writing
ME7	Speech (voice-over narration) and moving images (video footage)
ME8	Speech (voice-over narration), moving images (video footage), and writing

* Multimodal Ensemble.

The multimodal ensembles with the most semiotic diversity are ME1, ME2, and ME3. They are all comprised of speech as well as the non-verbal modes that entail the presence of the scientist speaking in front of the camera: the gestural modes of gaze, smile, eyebrow raising, head movement and hand and arm gestures; and the choice of clothes. ME2 and ME3 include the additional modes of writing and superimposed static images, respectively. In the three videos analysed, these ensembles are used during the researcher's self- introduction and the final call to action. ME4 combines writing with static images and it was only found in SCV1 as a way of opening and closing the video content. As for ME5, ME6, ME7, and ME8, the primary combination of modes is speech and static/moving images; ME6 and ME8 also employing written language. These multimodal ensembles are common in the three chosen SCVs when describing the research project in the middle section.

Constraints of space do not allow for a full multimodal description of each video. Instead, in what follows I provide examples taken from the three videos to illustrate how those multimodal ensembles have been used to engage the audience in different sections. The multimodal transcripts produced for these examples can be found in Appendix A.

5.1. Constructing a competent and enthusiastic researcher identity through ME1, ME2 and ME3

5.1.1. Researcher's self-introduction in SCV3

In SCV3, the two researchers introduce themselves with an ME2, an ME3, and an ME1 (in this order) produced in one single shot. Figure 2 includes the utterances produced in each ensemble, as well as a screenshot worth discussing in terms of semiotic relations.

One the one hand, several semiotic modes are present throughout the entire shot as the scientists talk. These modes are common to the three multimodal ensembles used and include camera shot and angle, the setting from where the scientists are filmed, and the attire worn by them. In order to simulate an intimate and dialogic environment and engage the viewer, both scientists are captured in medium-close shots as well as frontal and eye-level angles (Kress & van Leeuwen, 2021). It is confirmed that this semiotic decision is intentionally designed for engagement purposes because Dawson stated during the playback session that she and her colleague did not want to be filmed too far away from the camera and, consequently, from the audience. Even if scientists look at each other at some specific moments to show team cohesion and affinity (see ME3 in Figure 2), they keep eye-contact with the audience along the shot by means of a direct gaze which, together with the type of camera shot and angle chosen, contribute to build rapport with the audience (Kress & van Leeuwen, 2021). Besides, the researchers decided to film themselves in their laboratory as a way of demonstrating their affiliation to a research institution, which is also demonstrated by the lanyard with employee ID worn by Dawson (Interview).

In ME2, ME3 and ME1, the semiotic modes above mentioned are combined with speech and gestural modes, as well as writing in ME2 and static images in ME3 (see Figure 2). In ME2, the two researchers state

their names while conveying a friendly relaxed smile and superimposing a written text that repeats their names and adds the title of their research project, which is not conveyed through speech. Writing and speech are thus involved in both an elaborative and an extensive modal relation, with the two modes restating and adding new information on each other (van Leeuwen, 2005). Extensive relations, in particular, are useful for scientists to convey multiple information simultaneously in this time-constrained video genre.

Figure 2. ME2, ME3, and ME1 while introducing the researcher in SCV3.[2]

ME2	ME3	ME1
(Researcher 1) Hi, I'm [name omitted]. (Dawson) And I'm [name omitted].	(Researcher 1) And we are graduate students with University [institution name omitted] working out of the space Life Sciences lab, which is right next to Kennedy Space Center.	(Dawson) And we're most interested in trying to figure out how to keep astronauts healthy during long-term space travel.

In ME3, Dawson's colleague establishes the credentials of both researchers (i.e., their qualification and institutional affiliations). As she speaks, two semiotic expressions can be highlighted. One of them is the speaker clasping her hands with palms facing upwards when uttering the key content words "graduate", "students", and "Florida". According to Cestero-Mancera (2018), this hand gesture is used to emphasise information and facilitate understanding of key content, and is considered a positive, audience-oriented gesture to engage the audience in the communicative activity (p. 86). Another salient semiotic expression is when she points out that their laboratory is very close to a remarkable site, the Kennedy Space Centre, where NASA space shuttles are launched from. By foregrounding this remarkable site, they add prestige to the laboratory they belong to, and therefore, to their work as researchers, since, as Dawson stated, "the idea of working at a location by Kennedy Space Center might be really exciting to some people" (Interview). As she speaks, a static image of the laboratory building pops up in the lower right corner of the screen, which illustrates that building. This would be an example of semiotic elaboration, where the pop-up image elaborates on the utterance conveyed by the speech mode. As Dawson confirmed, emphasis on their workplace is made on purpose to show their research competence. Both scientists were only graduate students with little experience to share at the time of launching this crowdfunding campaign (Interview), so that their laboratory was probably their best endorsement to persuade about their credentials and preparation.

As for ME1, Dawson indicates their research interest while producing engaging gestural modes. When she utters "we're most interested in…", she nods her head gently, which creates emphasis and seeks agreement from the public (Cestero-Mancera, 2018). Then, she closes her eyes slightly as a way of creating suspense of what she is about to announce and keeping the attention of the audience (see Figure 2). And then, when she reveals that they are interested in "trying to figure out how to keep astronauts healthy", she raises her eyebrows substantially and nods her head more vigorously to highlight the keywords "astronauts" and "healthy", at the time that her colleague smiles to the audience to also show her enthusiasm about the research topic.

The combinations of semiotic resources produced in this shot are thus intended to engage the audience and persuade them to participate in the campaign by constructing a competent and enthusiastic researcher identity that deserves funding. In order to show competence, speech, setting and clothing are combined along the whole shot and elaborate meaning on each other. On the other hand, enthusiasm is shown by combining speech with gestural modes that convey the researcher's emotional side. As for the setting, camera shot and angle, the clothes, and the written text and superimposed image used in ME2 and ME3 respectively, they are strategically orchestrated with the other modes to produce a more engaging effect.

5.1.2. Final call to action in SCV2

SCV2 ends with an ME1, where the scientist makes the final call to action in a single shot. As can be seen in Figure 3, she thanks the audience for watching and invites the audience to participate in the campaign, and these utterances are made more engaging with the semiotic choices made for the setting, the clothes, the type of camera shot and angle, and the gestures conveyed by the researcher while speaking.

---

2   The image of the scientist on the right has been distorted for ethical reasons, unlike the image of the other scientist, who was the interviewee and who gave her consent to use her image in the work resulting from my research.

Figure 3. ME1 while calling the audience to action in SCV2.

ME1	
Thanks again for watching my video.	I hope that you'll join me in conserving the amazing Palo Santo tree.

Unlike researchers in SCV3, Crawford chose an outdoors location to film herself in. Given that her research revolves around the study of the Palo Santo tree species, the non-specialised viewer might assume the background to be a Palo Santo forest. However, during the interview, the researcher explained that the chosen forest is not where this species can be found, but a park in Washington D.C., her home city. Although she would have liked to stand in a Palo Santo forest during the filming, these are located in South America and she was not there at the moment of launching the crowdfunding campaign. Hence, it is noteworthy that, in spite of not having access to the research location, Crawford gave the location enough attention to purposefully go to a natural setting so that viewers would at least identify her with "some sort of forest or some pretty place with plants" (Interview).

The setting choice, together with the smile she maintains almost during the whole shot, constitute a semiotic combination to show enthusiasm about her field of study and her research project on trees, but also to show competence in a friendly and approachable way in combination with her clothes choice (i.e., hiking jacket), which is also coherent with the setting. As for the camera shot and angle, Crawford also selected a medium-close shot and a frontal angle to further engage the audience. As she commented, "I remember trying many different standing-furthers and this one felt like 'I'm just talking to you'" (Interview). This camera shot and angle choice, and the permanent direct gaze towards the audience (Kress & van Leeuwen, 2021), in combination with statements using the second person pronoun "you" (Hyland, 2005), establish rapport with the audience and therefore are consciously arranged to engage the audience, as also discussed in Vivas-Peraza's (2022b) study.

In terms of gestural modes, smile, eyebrow raising, and head movements are worth highlighting since these, in combination with speech and camera shot and angle chosen before the recording, contribute to audience engagement. When thanking the audience, the researcher smiles at the audience and keeps her eyebrows raised and her head tilted slightly backwards as a way of creating emphasis (Cestero-Mancera, 2018). These gestures, together with the direct gaze maintained with the audience and the type of camera shot and angle chosen, increase the rapport with the audience. At the same time, the combination of smile with the evaluative adjective "amazing" (Hyland, 2005), used in the latter utterance, contribute to a positive evaluation of the research object and demonstrates the researcher's self-confidence, which as Cestero-Mancera (2018) points out, also predisposes the audience to persuasion (pp. 79-80).

5.2. Seeking common ground through ME4

5.2.1. Opening and closing slides in SCV1

The combination of writing and image in ME4 was only observed in SCV1, where this ensemble is materialised in an opening and closing slide that frames the video content; that is, right before the researcher's self-introduction and right after the call to action. In both cases, a written message is conveyed in English and in Spanish in the form of subtitles, with a static image in the background (see Figure 4.).

The two messages include linguistic features that favour audience engagement with the research project. One of them is the use of inclusive pronouns, as in "*we* are living in a new era", "shaping *our* social life", "*we* work together", and "together *we* can ensure". These pronouns establish rapport with the audience, by bringing them to the discourse and by indicating they share similar views with the scientist as a sign of membership (Hyland, 2005). Another one is the use of evaluative statements such as "to improve global public health and mitigate future re-emerging infectious diseases" or "ensure the quality of life for pregnant mothers and newborns". These statements evaluate the project positively by emphasising its potential benefits (Hyland, 2005), in order to prompt the participation of the public. Furthermore, it is apparent that in the initial message, the researcher seeks common ground by making a subtle allusion to COVID-19. The second wave of this unprecedented pandemic had already started when this campaign was released, so Wongyai used the opportunity to find connections between his research on Zika virus and what was going on in the world at that time to demonstrate how everyone could benefit from his project and urge them to participate.

Figure 4. ME4 as an opening and closing slide in SCV1.

With regards to the static image chosen for the background to accompany these two written messages, it displays the facilities of Thailand Science Park in Bangkok (see Figure 4), the setting from which the research is taking place and thus where Wongyai was conducting his experiment on Zika virus. The bird view is not casual and aims to impress viewers with its size and fanciness, as well as to add credibility to the scientist and his project, as Wongyai explained during the interview. The image is thus chosen to show expertise of the researcher and confer prestige, credibility, and reliability to him. In the meantime, it extends the meaning of inclusive pronouns "we", "our", and "us" conveyed in writing, since it specifies who (apart from the viewers) is involved: those behind those facilities, in this case, Wongyai, the scientist protagonist. The information about the project benefits is thus conveyed through writing, and even though the background image does not elaborate or extend on those benefits, it is strategically chosen as a symbolic way to show the prestigious institution that is behind those potential benefits.

5.3. Describing the research through ME5, ME6, ME7, and ME8

5.3.1. Presenting the research project in SCV2

In SCV2, Crawford, the author scientist, draws on ME5s to describe her object of study (the Palo Santo tree). As she commented during her interview, the images used were taken by herself, which shows she has been into the research field and is involved with the research object, thus adding credibility to her scientific persona and her project.

Figure 5. ME5 to present the research project in SCV2.

ME5	ME5
Palo Santo trees have been used by people for thousands of years for their aromatic wood and for their essential oil.	Due to deforestation and overharvesting, Palo Santo populations are decreasing in many areas of its range.

As can be seen in Figure 5, the first shot includes an oral statement that provides some general information about the Palo Santo tree while an image about its trunk shows up. In this image, attention is drawn to the oil spilling out of the log by providing a very close shot of the tree and by using a blurring background that makes the trunk and the oil stand out (Kress & van Leeuwen, 2021). Here, there is an elaborative and extensive relationship between speech and image. On the one hand, both semiotic modes inform that the wood and the oil are essential parts of the tree, with words in the case of speech and with camera focus and shot that highlight these parts of the three in the case of the image. On the other hand, without speech it cannot be determined that the wood is aromatic (smell cannot be seen) or that both elements have been used by people for a long time. Furthermore, without the image, it cannot be observed the specific and close physical characteristics of the tree. This demonstrates how language and image offer different

affordances (Stöckl, 2004; van Leeuwen, 2017) and proves how one might be used to expand the meaning of the other. This multimodal ensemble also shows the researcher's expertise, as she is both informing the audience about the characteristics of the tree species she investigates (thus demonstrating she knows well her object of study), and providing an image taken by herself in the research field (showing her involvement with her object of study).

In the second shot (see Figure 5), an oral statement is used to discuss the decline of the species of Palo Santo while the image displays some felled tree branches on a forest floor. Speech and image here also elaborate and extend meaning on each other. The image elaborates on the tree deforestation announced in speech, but also extends speech as it illustrates the physical characteristics of the felled Palo Santo tree branches. In addition, the meaning conveyed through both modes attempts to arouse the audience's sympathy with the problem. The present continuous form chosen for the verb phrase "Palo Santo populations are decreasing" implies that the number of trees will keep decreasing if the research project is not carried out, suggesting a possible total extinction of the species. Similarly, the intention behind the image of the tree branches on the ground is to increase public awareness of the situation. The statement on the research object, together with the image taken by the very researcher in the field (Crawford, Interview), demonstrates as in the first shot that she is knowledgeable of the current status of the tree, thus showing her expertise in the field of study.

5.3.2. Seeking compassion in SCV1

In SCV1, Wongyai informs about the sanitary, social and economic consequences of Zika virus disease on the population. To do that, he includes three different shots consisting of ME6s where he combined static images, speech, and writing (see Figure 6). Static images are accompanied by voice-over narration and overlay text used to include the image source, to provide Spanish subtitles, and to emphasise key information that is conveyed in speech.
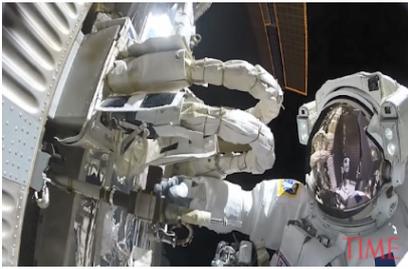
Figure 6. ME6 to seek compassion in SCV1.

ME6	ME6	ME6
Zika epidemic causes significant socio-economic burdens on the population because it is transmitted from pregnant mother to her unborn child.	Foetuses exposed to Zika virus through this transmission have a 20 percent chance of developing severe complications, including lifelong developmental defects and intellectual disability.	This burden causes a significant financial and emotional cost to the family.

As can be seen, as the author deals with the consequences of Zika virus on the population through spoken utterances, three images show children suffering from the effects of the Zika virus contracted from birth. As Wongyai commented during the interview, he consciously selected those images to prompt an emotional reaction in the audience (Interview). Static images in these ME6s, therefore, are used to appeal to the emotions of the audience in order to persuade them to donate money for the project, a strategy that is also used frequently in advertising to convince people to buy a product (van Leeuwen, 2017). As far as writing is concerned, it extends the meaning of the other modes by indicating the sources of the borrowed pictures, and it also elaborates on speech by emphasising key words of the three oral statements and by providing subtitles that translate the full statements into Spanish. Therefore, the three modes—speech, writing, and image—engage in an elaboration-extension semiotic relationship in which they both add new information and restate that which has been conveyed by other modes. The three communicate that pregnant women infected with Zika put their foetuses at risk. Meanwhile, the three modes extend information on their own. Images quickly convey the physical and behavioral abnormalities of a new-born afflicted with Zika; written and spoken statements explain that the virus spreads during pregnancy and offer statistics on the likelihood of transmission—information that is challenging to convey through images; and writing acknowledges the authors of the selected photos.

5.3.3. Mixing popular culture images and specialised images in SCV3

In SCV3, the two scientists introduce their research project topic with three shots composed by two ME7 and one ME8, as illustrated in Figure 7.

Figure 7. ME7 and ME8 to present the research project in SCV3.

ME7	ME7	ME8
Astronauts will continue to go to space...	...and for longer periods of time. But how do we make sure that they stay healthy during space travel?	Research on astronauts has found that their immune systems are highly dysregulated in space. And this may be in part due to the disruption of their normal healthy microbiome or the microorganisms that are associated with us.

In the first shot, the utterance is combined with a footage of seven NASA astronauts who look like they are about to go on a space mission based on the way they are raising their fists in triumph before getting on the NASA bus. And in the second shot, the utterance is accompanied by another footage that features an astronaut floating in space while repairing some space shuttle. Dawson revealed that the images were taken from IMAX Hubble 3D (2010), a well-known spatial documentary concerning the final Hubble repair mission (Interview), which demonstrates that the researchers have used imagery from popular culture into their SCV to enhance its attractiveness and thus engage the audience in their project. These images are combined with the use of inclusive "we" (confirmed by Dawson during the interview) and the rhetorical question asked in the second shot, which are clear linguistic strategies for audience engagement (Hyland, 2005). Speech constitutes the semiotic mode that conveys information about the objectives of the research project. Conversely, the images serve a more decorative purpose, serving to evoke memories of a significant previous space mission—akin to potential future space journeys for which astronauts may preserve their health better. Images here thus do not provide information about the project but make the SCV more appealing and engaging.

Regarding the third shot, an ME8 is used to orchestrate speech, writing, and video footage. In this instance, a timelapse video footage of microscope images are displayed together with written text acknowledging the firm that generated those images (Timelapse Vision, Inc.). Written language extends the meaning conveyed by image and speech, whereas these two latter modes are involved in an elaborative relationship. According to Dawson, these images are illustrating the immune system dysregulation that they mention in their spoken narration, so that they elaborate meaning on each other. However, non-experts cannot really interpret what is going on in those microscope images, and since the SCV was created for all publics, they may have been included for credibility purposes rather than to help them understand what is being said. From this point of view, images show that the researchers have access to laboratory equipment, thus showing their status as scientists and their expertise in the field.

6. Conclusions

The Open Science movement has emphasised the importance of public communication of science for scientists to disseminate specialised knowledge among lay audiences, gain the trust of society, and advance their research careers. The Internet and Web 2.0 have enabled scientists to reach diverse audiences through new digital genres for science popularisation, which are inherently multimodal. The traditional notion of academic literacy has thus shifted to include multimodal artefacts, requiring scientists to develop digital composing and multimodal communication skills. In light of this background, the present study has proposed multimodality as a framework to the analysis of science crowdfunding videos, with the aim of advancing knowledge in current trends of science communication and offer pedagogical applications to train future scientists in the production of multimodal artefacts.

The SCV is an emerging digital genre produced to disseminate scientific content and engage the Internet audience in a research project in need of funding, and the findings derived from this study have shown that scientists can fulfil those two purposes with eight different multimodal ensembles. The three scientists interviewed seem to be aware of the need of gaining people's trust and invest their efforts in combining semiotic modes that add credibility to their scientific capacity and enthusiasm. In this regard, it has been observed how the statements about their credentials (in speech or writing) become more powerful in combination with a choice of clothing and setting typical of a scientific person. On the other hand, scientists orchestrate linguistic, gestural, and visual modes to promote a climate of dialogue, closeness and familiarity with the audience. Gestural semiotic modes, such as gaze, smile, arm and hand gestures, eyebrow raising, and head

movements have, since ancient times, constituted non-verbal resources used by speakers in all kinds of persuasive communication (Cestero-Mancera, 2018), and when scientists are recorded in front of the camera, they are also observed to make use of them. The modal affordances of video also allow the researcher to simulate closeness and intimacy when addressing the audience through camera shot and angle effects. In addition, the possibility of editing the videos also allows the researcher to include static and moving images that perform different functions. One of them is, as shown in SCV1, to move the audience and urge them to act in the best interest of science. As (Stöckl, 2004) points out, images are more intense in connotation than the linguistic modes and tap more directly into human's emotions, which are fundamental for engagement. Another function is to serve a pedagogical purpose to aid in the understanding of scientific issues. Images make the video content engaging and inviting for all kinds of audiences, especially non-disciplinary audiences, and since they are more efficient cognitively than words (Rowley-Jolivet, 2004), they play a fundamental role in science popularisation. One further function is more pragmatical and has to do with time-saving, which is essential in the production of short videos such as the SCV. In this sense, the combination of image and language has shown to be very effective in the analysed videos.

All in all, the results of this study show how the combination of different semiotic resources enhance audience engagement as they allow showing expertise, seeking common ground and intimacy, and educating people in science in a same time-space slot. Even if the small size of the sample prevents the generalisation of results and the obtention of quantitative data on the most recurrent use of semiotic resources, the results of this study can nonetheless be potentially used for the elaboration of ESP genre-based courses that would train science practitioners in developing multimodal literacy to make scientific content engaging to all publics. Future ESP studies should continue this trend, by examining other emerging digital genres for science popularisation and by including the analysis of aural semiotic modes not considered in this study, such as the use of paralinguistic features or music. In the same manner, further multimodal research on SCVs should analyse a somewhat larger sample to provide quantitative data that would also lead to interesting results on multimodal patterns on this genre.

Acknowledgements

Informed Consent Statement

Informed consent was obtained from the three interviewed subjects involved in the study. The study was conducted in accordance with the Data Protection Department at the University of Zaragoza, and the protocol was approved by the Ethics Committee for Clinical Research of Aragon *(*CEICA*)* with the ethical approved code PI22/519. Informed consent was also obtained from the subjects to use their images in this and future academic papers. This documentation is not publicly available due to privacy and ethical restrictions.

References

Askehave, Inger, & Nielsen, Anne Ellerup. (2005). Digital genres: a challenge to traditional genre theory. *Information technology & people*, *18*(2), 120-141. https://doi.org/10.1108/09593840510601504

Baldry, Anthony Peter & Thibault, Paul J. (2006). *Multimodal Transcription and Text Analysis: A Multimodal Toolkit and Coursebook with Associated On-line Course*. Equinox.

Bateman, John A. (2014). Genre in the age of multimodality: Some conceptual refinements for practical analysis. In Paola Evangelisti-Allori, Vijay K. Bhatia & John A. Bateman (Eds.), *Evolution in Genre: Emergence, Variation, Multimodality* (pp. 237-269). Peter Lang.

Bernad-Mechó, Edgar, & Valeiras-Jurado, Julia. (2023). Multimodal engagement strategies in science dissemination: A case study of TED talks and YouTube science videos. *Discourse Studies 25*(6), 1-22. https://doi.org/10.1177/14614456231161755

Bezemer, Jeff & Jewitt, Carey. (2010). Multimodal analysis: Key issues. In Lia Litosseliti (Ed.), *Research Methods in Linguistics* (pp. 180-197). Continuum.

Cestero-Mancera, Ana María. (2016). La comunicación no verbal: propuestas metodológicas para su estudio. *Lingüística en la Red, Monográfico XIII* (2), 1-36. http://hdl.handle.net/10017/25226

Cestero-Mancera, Ana María. (2017). La comunicación no verbal en discurso persuasivo empresarial. *Pragmalingüística, 25*, 124-145. http://hdl.handle.net/10498/20146

Cestero-Mancera, Ana María. (2018). Recursos no verbales en comunicación persuasiva: gestos. *ZER. Revista de Estudios de Comunicación 23*(44), 69-92. https://doi.org/10.1387/zer.18130

Cestero-Mancera, Ana María & Díez-Prados, Mercedes (2021). Paralinguistic resources in persuasive business communication in English and Spanish. In Eliecer Crespo-Fernández (Ed.), *Discourse Studies in Public Communication* (pp. 245-272). John Benjamins.

Daly, Peter & Davy, Dennis (2016). Structural, Linguistic and Rhetorical Features of the Entrepreneurial Pitch: Lessons from the Dragon's Den. *The Journal of Management Development 35*(1),120-132. https://doi.org/10.1108/JMD-05-2014-0049

Díez-Prados, Mercedes (2019). Verbal and nonverbal engagement devices in business persuasive discourse. The elevator pitch. In Carmen Sancho-Guinda (Ed.), *Engagement in professional genres* (pp. 217-242). John Benjamins.

Experiment. (2023). Experiment.com. https://www.experiment.com

Friese, Susanne (2019). Qualitative data analysis with ATLAS. ti. *Qualitative data analysis with ATLAS. ti*, (3rd. ed). Sage Publications.

Hafner, Christoph A. (2014). Embedding digital literacies in English language teaching: Students' digital video projects as multimodal ensembles. *TESOL quarterly*, *48*(4), 655-685. https://doi.org/10.1002/tesq.138

Hampel, Regine & Hauck, Mirjam (2006). Computer-mediated language learning: Making meaning in multi-modal virtual learning spaces. *JALT CALL Journal*, *2*(2), 3-18.

Hui, Julie S. & Gerber, Elizabeth M. (2015). Crowdfunding science: Sharing research with an extended audience. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing* (pp. 31-43). http://dx.doi.org/10.1145/2675133.2675188

Hyland, Ken. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse studies*, *7*(2), 173-192. https://doi.org/10.1177/1461445605050365

Jewitt, Carey. (Ed.). (2009). *The Routledge handbook of multimodal analysis*. Routledge.

Jewitt, Carey. (2013). Multimodal methods for researching digital technologies. In Sara Price, Carey Jewitt & Barry Brown (Eds.), *The SAGE handbook of digital technology research* (pp. 250-265). Sage Publications.

Jewitt, Carey, Bezemer, Jeff, & O'Halloran, Kay. (2016). *Introducing multimodality*. Routledge.

Kress, Gunther. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, Gunther & van Leeuwen, Theodoor. (1996). *Reading Images: The Grammar of Visual Design*. Routledge.

Kress, Gunther & van Leeuwen, Theodoor. (2001). *Multimodal Discourse, The Modes and Media of Contemporary Communication*. Oxford University Press.

Kress, Gunther & van Leeuwen, Theodoor. (2006). *Reading Images: The Grammar of Visual Design*. (2nd edition). Routledge.

Kress, Gunther & van Leeuwen, Theodoor. (2021). *Reading Images: The Grammar of Visual Design*. (3rd edition). Routledge.

Lancaster, Lesley & Roberts, Mark (2006) *Grammaticisation in Early Mark Making: A Multimodal Investigation*. End of Award Report: RES-000-22-0599.

Luzón, María José. (2019). Bridging the gap between experts and publics: The role of multimodality in disseminating research in online videos. *Ibérica 37* (2019), 167-192.

Luzón, María José & Pérez-Llantada, Carmen. (Eds.). (2019). *Science communication on the internet: old genres meet new genres.* John Benjamins.

Luzón, María José & Pérez-Llantada, Carmen. (2022). *Digital genres in academic knowledge production and communication: Perspectives and practices*. Multilingual Matters.

Maier, Carmen Daniela & Engberg, Jan (2023). Multimodal Knowledge Communication in a Recontextualized Genre: An Analysis of Expertise Dissemination and Promotion Strategies in Online Academic Trailers. *ESP Today 11*(2), 261-279. https://doi.org/10.18485/esptoday.2023.11.2.4

Mehlenbacher, Ashley Rose. (2017). Crowdfunding science: Exigencies and strategies in an emerging genre of science communication. *Technical Communication Quarterly*, *26*(2), 127-144. https://doi.org/10.1080/10572252.2017.1287361

Mehlenbacher, Ashley Rose. (2019). *Science communication online: Engaging experts and publics on the internet*. The Ohio State University Press.

Norris, Sigrid. (2011). *Identity in (inter) action: Introducing multimodal (inter) action analysis*. Walter de Gruyter.

Paltridge, Brian. (2012). *Discourse Analysis* (2nd edition). Bloomsbury.

Pérez-Llantada, Carmen. (2021). Genres and languages in science communication: The multiple dimensions of the science-policy interface. *Language & Communication*, *78*, 65-76. https://doi.org/10.1016/j.langcom.2021.02.004

Poyatos, Fernando. (1983). Language and nonverbal systems in the structure of face-to-face interaction. *Language & Communication*, *3*(2), 129-140.

Poyatos, Fernando. (Ed.). (2002). *Nonverbal communication across disciplines*. John Benjamins.

Pullman, George. (2013). *Persuasion. History, theory, practice*. Hackett Publishing Company.

Rowley-Jolivet, Elizabeth. (2004). Different visions, different visuals: A social semiotic analysis of field-specific visual composition in scientific conference presentations. *Visual communication*, *3*(2), 145-175. https://doi.org/10.1177/147035704043038

Spang, Kurt. (2005). *Persuasión: fundamentos de retórica*. Eunsa.

Stöckl, Hartmut. (2004). In between modes. In Eija Ventola, Martin Kaltenbacher & Charles Cassily (Eds.), *Perspectives on Multimodality* (pp. 9-29). Routledge.

Tseronis, Assimakis & Forceville, Charles. (Eds.) 2017. *Multimodal Argumentation and Rhetoric in Media Genres*. John Benjamins.

Vachelard, Julien, Gambarra-Soares, Thaise, Augustini, Gabriela, Riul, Pablo & Maracaja-Coutinho, Vinicius (2016). A guide to scientific crowdfunding. *PLoS biology*, *14*(2), e1002373. https://doi.org/10.1371/journal.pbio.1002373

Valeiras-Jurado, Julia, & Bernad-Mechó, Edgar. (2022). Modal density and coherence in science dissemination: Orchestrating multimodal ensembles in online TED talks and youtube science videos. *Journal of English for Academic Purposes, 58*, 101118. https://doi.org/10.1016/j.jeap.2022.101118

van Leeuwen, Theodoor. (1999). *Speech, Music, Sound*. Palgrave Macmillan.

van Leeuwen, Theodoor. (2005). Multimodality, genre and design. In Sigrid Norris & Rodney H. Jones (Eds.), *Discourse in Action* (pp. 73-93). Routledge.

van Leeuwen, Theodoor. (2017). Rhetoric and Semiotics. In Michael J. MacDonald (Ed.), *The Oxford Handbook of Rhetorical Studies* (pp. 673-682). Oxford University Press.

Vela-Rodrigo, Alberto Ángel. (2023). A lexical bundle analysis of art-related crowdfunding projects. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*, (46), 321-349. https://dialnet.unirioja.es/descarga/articulo/9223213.pdf

Villares, Rosana. (2023). Exploring the rhetorical strategies of stance and engagement in Twitter conference presentations. *ESP Today 11*(2), 280-301. https://doi.org/10.18485/esptoday.2023.11.2.5

Vivas-Peraza, Ana Cristina. (2022a). The role of women in STEM disciplines. A multimodal analysis of crowdfunding project videos for the iGEM science competition. *Sociología y tecnociencia*, *12*(2), 228-251. https://doi.org/10.24197/st.2.2022.228-251

Vivas-Peraza, Ana Cristina. (2022b). Engaging the public in science crowdfunding: Scientists calling to action through visual and verbal strategies. *VISUAL REVIEW*, *9*(3), 1-15. https://doi.org/10.37467/revvisual.v9.3534

Xia, Sichen Ada. (2020). Genre analysis in the digital era: Developments and challenges. *ESP Today*, *8*(1), 141-159. https://doi.org/10.18485/esptoday.2020.8.1.7

Xia, Sichen Ada & Hafner, Christoph A. (2022). Engaging the online audience in the digital era: A multimodal analysis of engagement strategies in TED talk videos. *Ibérica*, *2021*(42), 33-58. https://doi.org/10.17398/2340-2784.42.33

Appendix A

The multimodal transcripts generated from the analysis of the three SCVs can be found at: https://drive.google.com/file/d/1Zz1tiKAiFEZlqR2cxuYMvajH1CJApcSw/view?usp=sharing