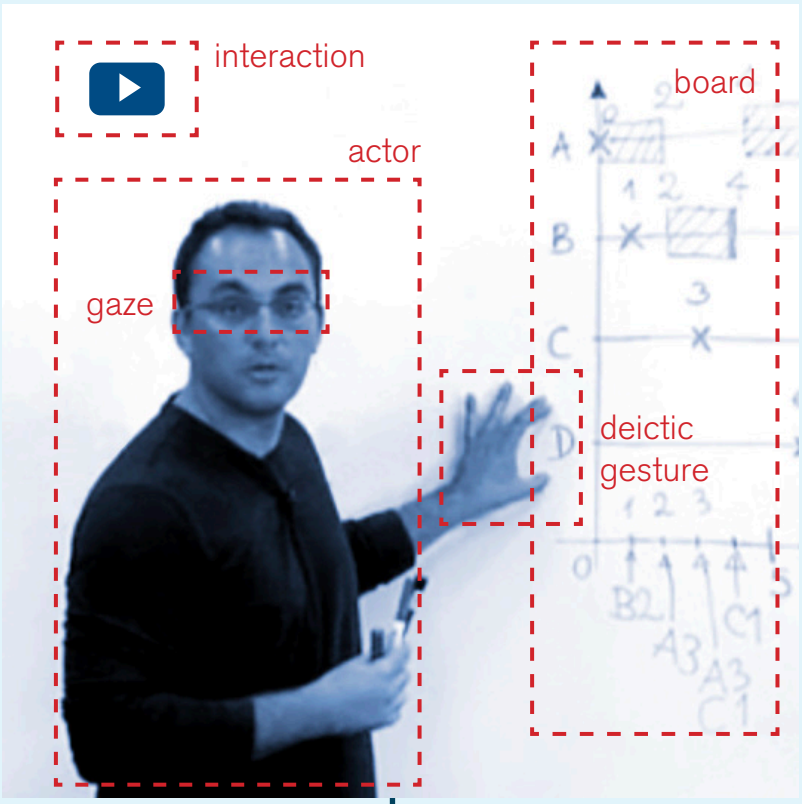
PhD Thesis

ANATOMY OF INSTRUCTIONAL VIDEOS

A SYSTEMATIC CHARACTERIZATION OF THE STRUCTURE OF ACADEMIC INSTRUCTIONAL VIDEOS

José Miguel Santos Espino



Supervisor
Francisco Mario Hernández Tejera

Doctoral program
Telecommunication Technologies and Computational Engineering

Universidad de Las Palmas de Gran Canaria
Las Palmas de Gran Canaria, June 2019

UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

D. GUSTAVO MARRERO CALLICÓ, COORDINADOR DEL PROGRAMA DE DOCTORADO «TECNOLOGÍAS DE TELECOMUNICACIÓN E INGENIERÍA COMPUTACIONAL» DE LA UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA,

INFORMA,

Que la Comisión Académica del Programa de Doctorado, en su sesión de fecha doce de julio de dos mil diecinueve tomó el acuerdo de dar el consentimiento para su tramitación a la tesis doctoral titulada “Anatomy of Instructional Videos: a Systematic Characterization of the Structure of Academic Instructional Videos” presentada por el doctorando D. José Miguel Santos Espino y dirigida por el Doctor D. Francisco Mario Hernández Tejera.

Y para que así conste, y a efectos de lo previsto en el Artículo 11 del Reglamento de Estudios de Doctorado (BOULPGC 7/10/2016) de la Universidad de Las Palmas de Gran Canaria, firmo la presente en Las Palmas de Gran Canaria, a quince de julio de dos mil diecinueve.

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

ESCUELA DE DOCTORADO

Programa de doctorado

Tecnologías de Telecomunicación e Ingeniería Computacional

Título de la tesis

Anatomy of Instructional Videos: A Systematic Characterization of the Structure of Academic Instructional Videos

Tesis Doctoral presentada por

José Miguel Santos Espino

Dirigida por

Dr. Francisco Mario Hernández Tejera

El Director, El Doctorando,

Las Palmas de Gran Canaria, a 27 de junio de 2019

anatomy *noun* [...] **3** : the art of separating the parts of an organism in order to ascertain their position, relations, structure and function.

— *Merriam-Webster Dictionary*

It is possible to teach every branch of human knowledge with the motion picture.

— *Thomas Alva Edison, 1913*

Abstract

Research on video-based learning has found several structural features in instructional videos with a potential influence in learning outcomes. The main goal of this thesis has been to build a systematic classification scheme for these characteristics.

This research has covered instructional videos from a broad perspective, considering three natures of instructional videos: as instructional films, as multimedia learning objects, and as multimodal texts. Thus, the classification scheme is grounded in a multidisciplinary theoretical framework, which includes Cognitive Multimedia Learning theories, Film Analysis, Multimodal Discourse Analysis and Systemic Functional Linguistics.

The process of building the classification starts with an extensive literature review and a field study on MOOC platforms. Features retrieved from the review and the field study led to a bottom-up conceptual clustering, ending with a full classification scheme. The architecture of the classification scheme is inspired on Multimodal Discourse Analysis models, particularly John Bateman's GeM framework.

The resulting classification scheme comprises eight taxonomical domains: Medium, Presentation, Interaction, Spatiotemporal, Speech, Social Appearance, Strategic and Generic (for video genres). These domains are organized in hierarchical layers, from the physical medium to more abstract levels. In addition, specific taxonomies have been developed for all domains except Strategic and Generic. These intradomain taxonomies have been elaborated by means of literature reviews that have compiled more than 200 authoritative references on the influence of audiovisual features on learning.

In summary, this research has delivered these products: 1) A classification scheme that systematically organizes the characteristics in instructional videos that researchers have found relevant in learning processes; 2) A survey of presentation styles and features currently used in instructional videos in online courses (MOOCs); and 3) A comprehensive literature review on the instructional video features that are related to learning effectiveness.

Agradecimientos / acknowledgements

Decidí escribir en inglés el contenido de esta memoria de tesis, salvo este apartado de agradecimientos. Por más que el inglés sea una lengua que aprecio y cultivo, mi corazón habla en español. Y desde el corazón están escritas estas líneas.

Quiero agradecer en primer lugar a Cayetano Guerra su invitación a embarcarme en este proyecto y haber pasado tantos buenos ratos juntos, que espero que continúen por muchos años. Igualmente agradezco a Marilola Afonso el haberme acompañado durante los primeros años de esta investigación. Y a Mario Hernández le agradezco el haber estado en los momentos oportunos para dar el consejo imprescindible y el apoyo necesario para la recta final del trabajo, como director de la tesis.

Quiero también dar mi agradecimiento a todos los profesionales que de algún modo han contribuido a este trabajo: a Soraya García, con la que además he tenido el placer de haber compartido tres publicaciones relacionadas con esta línea de investigación; a Michail Giannakos y su equipo en la NTNU en Trondheim, que gentilmente me acogieron en su campus y me dieron valiosos consejos; a Jon Baggaley por rebuscar entre sus viejos *papers* de los años setenta para enviarme algún original imposible de encontrar en Internet; a los profesores de la ULPGC que participaron en el Proyecto Prometeo y dedicaron su tiempo a relatarme su experiencia, que por cierto aún queda pendiente de publicar en un artículo no recogido en esta tesis; a José Pablo Suárez por ayudarme a aliviar cargas de trabajo; a mi compañero de despacho Agustín Trujillo por aguantar mis neuras en las etapas complicadas; y en general a todas las personas que por acción u omisión han tenido un papel en mi desarrollo como investigador.

A mis hijas, Nerea Nova, Carla y Silvia, les agradezco todo el tiempo que sin saberlo ellas, me han concedido. Espero poder devolverlo con creces.

A mi hermano Ulises, dondequiera que estés, que sepas que en algunos momentos en los que redactaba esta memoria te he recordado mientras hacías tus vídeos de *stop motion* con los Playmobil.

A mi madre le agradezco, primero que nada, y como ella misma diría, «haberme parido». Ella también ha cedido mucho tiempo en el que necesitaba de mí, para que yo pudiera sumergirme en este largo trabajo. Le doy las gracias por su aceptación y su generosidad, con todo el amor de un hijo.

Y a Eugenia le agradezco todo. El haber sido compañera en el sentido más originario de la palabra: compartir el pan. El haber tenido el lujo de contar con la mejor revisora de inglés que conozco. El haber sido mi soporte vital. Y su sacrificio para abrir esos huecos imposibles en nuestras frenéticas vidas para que yo pudiera dedicar algo de tiempo a la tesis. Eugenia, sin ti yo no estaría aquí escribiendo estas líneas. Te las dedico con todo mi amor.

Contents

Chapter 1. Introduction to study

1.1 The genesis of this research

The history of this research begins in the spring of 2013, in one of the walks that Cayetano Guerra and I took around the Computer Science Building at the ULPGC Campus while we were discussing what my incipient Ph.D. project would be. Cayetano was my doctoral supervisor at that time. We agreed that my doctoral work should be oriented towards the analysis of instructional videos, following our previous experience around the Project Prometeo at the ULPGC, an in-house multimedia production unit that during its serving period produced more than 500 learning objects, most of them video lectures (Afonso Suárez, M.D.; Guerra Artal, C.; Villalba Casas, A.; Elías Hernández, 2009).

As part of my initial training, we agreed that it would be interesting to build an inventory of the characteristics in instructional videos with potential to influence learning outcomes. After a first sketch built mainly from Project Prometeo findings (Santos Espino, Afonso Suárez, Guerra Artal, & García-Sánchez, 2013), I started to review the literature on audiovisual education and realized the many scientific contributions on that matter, which have revealed the effect on learning of certain properties of videos: an adequate combination of voice and pictures, the spatial and temporal coherence of information elements and using conversational rather than formal speech style, just to mention a few. Most of those findings took roots on the works of the multimedia learning research community, with landmarks like the Cognitive Theory of Multimedia Learning (CTML) and the collection of multimedia learning principles (Mayer, 2014a), each one associated to one or a small set of properties in the multimedia learning object.

I learnt from my initial research that there is certainly abundant experimental evidence of video properties with a measurable effect on learning. But at the same time, I realized that this overwhelming set of findings lacked *structuring*. Richard Mayer's multimedia learning principles shared a solid underlying theory of learning, but on the surface, they appeared to have loose structural connections between them. In fact, the last edition of the canonical book *The Cambridge Handbook of Multimedia Learning* (Mayer, 2014d) lists up to 23 different learning principles, without an explicit taxonomy or classification scheme. Modern reviews of research, such as Kay's (2012) on video podcasts, do not provide a comprehensive classification of characteristics.

I concluded that there was a need for a *map*: a cartography that would organize all these research findings around higher-order structural categories.

Besides, recent research on the learning effectiveness of video properties put the focus on a limited—yet important—set of features, such as the presentational items and their spatial and temporal relationships. Other characteristics, such as those

related to camera usage and the interpersonal features of the discourse, receive relatively scarce attention. The state of things was not always like today. Let us take the research in educational TV made in Britain by John Baggaley and collaborators (John Baggaley, Ferguson, & Brooks, 1980). They developed an extensive experimental research of how multiple filming techniques (camera shots, cuts, soundtracks and more) affected the audience's perceptions and attitudes. This is a line of research that by the end of 20th century was taken apart in favor of features related to computer-generated multimedia objects. Now, in the rise of online instruction and research on the efficiency of online video lectures, these features are starting to be studied again by younger researchers belonging to a community of research who in many cases are not aware of the past valuable knowledge they could take advantage of.

In contrast to the current trend, older reviews of research of educational films and educational television showed a broader perspective on the structure of the motion picture product. For instance, the remarkable work of Wetzel, Radtke and Stern (1993, 1994), perhaps the last comprehensive review on educational film and video before the digital revolution and the advent of multimedia learning theories, covered all kinds of attributes and properties of educational artefacts, not only basic representational items, but also aspects of the filming process (e.g. camera settings) and discourse (e.g. humor, pedagogic style). It is hard to find posterior reviews that have this wide coverage of feature domains.

Consequently, I arrived at a second conclusion: not only was there a need for a map of video characteristics, but in the last two decades *the territory explored by media learning researchers had been narrowed*. There was a need to make again explicit the full variety of relevant features in instructional videos, to regain that 'lost territory' of research.

Then my doctoral thesis had an objective.

1.2 Situation

Before elaborating the approach to this research, it will be helpful to introduce the key concepts and the context where this research lays. I will start with the definition of some key terms. Then I will introduce the historic context of videos in education and the related scientific research on the instructional effectiveness of video.

1.2.1 Key terms and concepts

What is 'video'?

The *Encyclopedia of the Sciences of Learning* (Seel, 2012) defines *video* as *"the simultaneous presentation of a continuous stream of visual and auditory information"*. This definition is independent of media technology with which video is recorded and delivered, in contrast with narrower definitions that differentiate *video* from *film*, as this one found in the same encyclopedia: "the word *video* refers to the technology of electronically capturing and broadcasting a sequence of still images representing scenes in motion".

→ I will use the term *video* throughout this thesis in the broader meaning of the first definition given above.

What is 'learning'?

The concern for learning "focuses on the way in which people acquire new knowledge and skills and the way in which existing knowledge and skills are modified" (Shuell, 1986, p. 412). Shuell stated three criteria to ascertain the existence of learning: there is a change in the individual; the change results from practice or experience; the change is enduring (Shuell, 1986). Shuell's criteria are brilliantly coalesced by Richard E. Mayer in this succinct definition of learning: *"the relatively permanent change in a person's knowledge or behavior due to experience"* (Mayer, 1982, p. 1040).

→ Though the above definitions insinuate some cognitivist bias, they actually fit into the other major epistemological viewpoints—behaviorist and constructivist. Hence, I will assume them as the overarching concept of 'learning' in this thesis.

What is 'instruction'?

In common language, 'instruction' is used as a synonym of 'teaching' (Oxford Dictionary) or, more precisely, as "the action, practice of profession of teaching" (Merriam-Webster Dictionary). With a more academic intent, 'instruction' is defined by Smith and Ragan, in their book *Instructional Design* (2005, p. 2), as "the intentional facilitation of learning toward identified learning goals", or, in other words, "the intentional arrangement of experiences, leading to learners acquiring particular capabilities". Similarly, Reigeluth and Carr-Chellman (2009) make a short definition of instruction as *"anything that is done purposely to facilitate learning"*.

What are instructional and educational videos?

→ According to the above definitions, the concept of 'instructional video' could be defined as *video made for instruction*, that is, video 'made purposely to facilitate learning' or, more precisely, *'video made purposely to facilitate a permanent change of knowledge or behavior in the viewer'*. This will be the meaning that I will adopt for 'instructional video' in this research.

Note that the intentional nature of the word 'instructional' excludes from this definition those films and videos originally made for non-instructional purposes and brought into the classroom as an instructional resource, as in the case of showing a Hollywood film to spark a student debate, as in (Bruti, 2015).

On the other hand, it is useful to clarify the traditional distinctions between the terms 'educational' and 'instructional' that have often been applied to films and videos. In the early age of audiovisual learning technologies, the word 'instructional' was applied mainly to short-length films which showed procedures or explained concepts. On the other hand, the word 'educational' was applied to documentary films and academic works like college lectures and interviews to experts. Educational films would often feature a longer duration. Also, the word 'educational' was associated to academic and inherently educational institutions (e.g. universities, educational television channels, educational government boards), while the word 'instructional' was generally associated to the industry and military. These semantic distinctions between 'educational' and 'instructional' can be observed in mid-century texts (e.g. Fairgrieve, 1941) and they have been maintained over the years to some extent. Nowadays, however, the terms 'educational video' and 'instructional video' are used almost interchangeably, and with the same frequency, as Figure 1-3 suggests.

What is video-based learning?

As *The Encyclopedia of the Sciences of Learning* says (Seel, 2012): "The term *video-based learning* is used in the sciences of learning and cognition to designate a knowledge or skills acquired by being taught via video". In practice, 'video-based learning' is also often used in the sense of 'video-based instruction', that is, the intentional facilitation of learning using the video as principal tool.

→ I will use 'video-based learning' with both meanings across this document.

What are instructional design strategies?

Instruction can follow methods, tools and techniques, some of them based in scientific evidence. Instructional professional activities can be developed in a variety of disciplines: instructional design, development, implementation, management and evaluation (Reigeluth, 1983).

As regards instructional design, the foundational work by David Jonassen and collaborators (Jonassen, Grabinger, & Harris, 1991) identified five classes of instructional strategies: a) contextualizing instruction, b) presenting and cueing

content, c) activating learner processing, d) activating and assessing learner outcomes and e) synthesizing and sequencing those processes into instructional lessons.

Instructional design strategies can be grouped in four methodological approaches: *hands-on*, *expository*, *interactive* and *collaborative* (Ormrod, 2017). In expository instruction, information is presented in the same form in which students are expected to learn it. In hands-on activities, students *do* rather than hear or read. In interactive and collaborative approaches, learning is constructed upon social interaction between students. Videos can be used as support material for all kinds of instructional strategies: for example, in expository designs, it may be a recorded lecture of some topic; in collaborative designs, it may be a set of videos made by learners to share their discoveries.

→ This research will be limited to those videos that are best suited for expository instruction.

1.2.2 Films and videos in education

Video enables the integration of multiple information channels (sound, pictures and text) into a single stream that allows for a captivating and immersive experience. A film projected on a large screen in the darkness of a classroom can catch the students' attention like no other learning resource can. These exceptional qualities of the motion picture have led to a century-long history of audiovisual education full of imaginative instructional practices.

This section makes a short resemblance of the historic development of educational films and videos, their uses and their benefits. I will highlight some aspects that will be relevant along this dissertation.

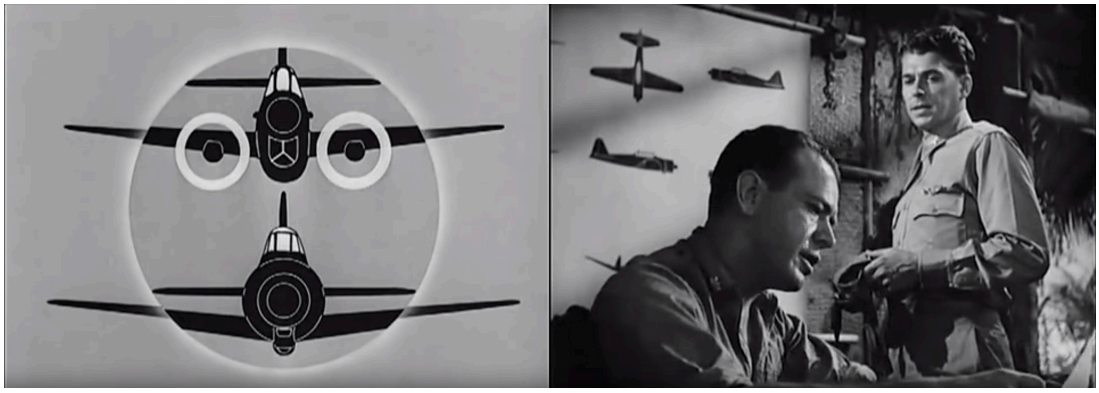
Brief history of educational films and videos

Cinematography served to education since its inception. The first known educational film dates from 1898; in 1908 Thomas Alva Edison started to produce educational films commercially ("Educational films," 1979). Thomas G. Smith concludes that "there had been educational films since movies were invented" (in Orgeron, Orgeron, & Streible, 2011). The needs of military mass training in Second World War spurred the film industry to produce training films which achieved high levels of quality, with remarkable examples as *Recognition of the Japanese Zero Fighter (1943)*[1], which relies on animated slides, diagrams, dialogic teaching and dramatization, a combination of features that even today seems sophisticated (see Figure 1-1).

[1] Available in YouTube: https://www.youtube.com/watch?v=uwo5uqOywFI

Figure 1-1. Two photograms from the instructional film *Recognition of the Japanese Zero Fighter (1943)*.

To the right, standing, a young Ronald Reagan in the role of a fighter pilot.



The post-WW2 period witnessed a flourishing era of production of 16mm educational films, which were regularly projected in European and American classrooms (Ruoff, 1992, p. 218). At the same time, television boosted as a new mass medium which broadcasted educational films. Some of the formats developed in educational television influenced in further developments of documentaries and other expository genres. Educational television has been traditionally focused on children audiences, but adult formats have been in use, often documentaries and *infotainment*. Distance education also benefitted from educational television and integrated this medium as a basic instructional resource.

The film technology was replaced in the eighties by compact videotapes, just to switch within few years to digital media: CD and DVD discs. Videotapes and digital discs enabled educational videos to be watched at home and at a user controlled pace, in contrast to the collective, pre-scheduled nature of educational television broadcasts and classroom sessions (Moreno, 2005; D. Zhang, Zhou, Briggs, & Nunamaker, 2006). At the same time, computers started to appear in schools, giving rise to a variety of educational software applications. Multimedia learning tools thrived, most of them including video clips as learning instructional resources.

At the beginning of the 21th century, streaming technology eliminated the need for physical media and allowed for real-time video delivery and user-controlled watching. Streaming video was received with enthusiasm by the academic community (Collis & Peters, 2000; Shephard, 2003; Thornhill et al., 2002; C. Young & Asensio, 2002). The web technology enabled a cheap and easy way to implement online instructional resources, including video materials.

Today, a vast amount of digital instructional videos is being produced every day in higher education institutions, with a dramatic increase in recent years. Massive Open

Online Courses (MOOCs) [2] make extensive use of instructional videos as a teaching resource. Meanwhile, millions of people turn to online courses and watch video lectures and tutorials in digital platforms as YouTube or Udemy. By 2013, 56% of online adults were watching how-to videos (Purcell, 2013). This boom in online video learning resources has been propelled by technical factors as the availability of cheap video streaming services, the steady growth in network bandwidth and low-latency for home users, and the spreading of smartphones and tablets, which are useful both as watching devices and for video recording and editing.

Figure 1-2. Frequency of terms related to film-based education, 1910-2008.
source: Google Books

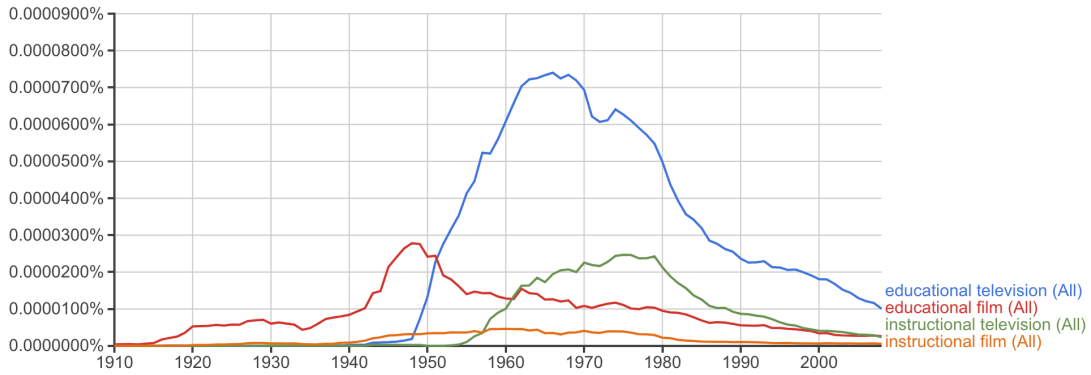


Figure 1-3. Frequency of terms related to educational videos, 1970-2008.
source: Google Books

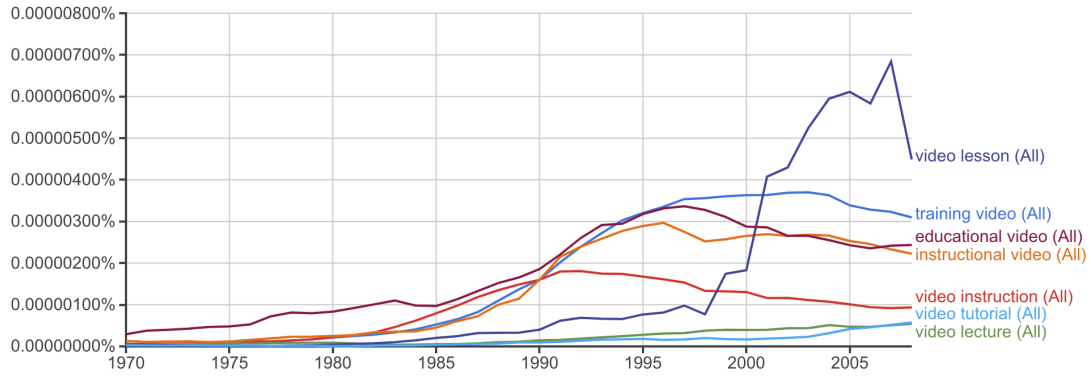


Figure 1-2 and Figure 1-3 show the usage of key terms about audiovisual educational media across the 20th century, retrieved from the Google Books corpus. The charts provide an insight to the historic evolution of moving pictures for instruction: we can see a rise in educational films in the Second World War and a boost of references to educational television starting in post-war times. Instruction with video media

[2] Chapter 3 includes a detailed description of the MOOC concept.

arises in the late sixties of 20th century. The data from Google Books corpus reach up to 2008, so they cannot attest the recent rise of online video delivery platforms.

Uses and benefits of video in instruction

The motion picture has the ability to depict realistic and immersive motion sequences, to capture and preserve real-world events and places that would be costly to see directly, and to change the size and speed of recorded natural phenomena that cannot be seen with the naked eye. These capabilities were recognized from early times as key advantages of the filmic medium as an educational device and today they can be considered "essential representational attributes" of motion pictures (Snelson & Perkins, 2009).

The technological changes at the end of the 20th century added new capabilities to video. First, new digital media enhanced interactivity features, particularly in playback control, navigation and search; and second, affordable technology enabled everyone to create and distribute video content, both instructors and students (Snelson & Perkins, 2009). As a result, new uses of video have emerged, such as video diaries, student-generated video tasks and videoconferences. Moreover, video clips can be integrated easily in wider learning environments, such as web pages and learning management systems. This historical development has been called by Young and Asensio (2002) as the "three 'I's": image, interactivity and integration.

Along the history of education, motion pictures have been used in a wide variety of formats: talking head lectures, recorded chalk-and-talk lectures, interviews, feature documentaries, dramatizations of historical events or professional scenarios, video diaries, demonstrations of procedures, videoconferences, trigger videos and many others. Some of the formats have been borrowed from other film and television fields, while others are more specific of education. Schwartz and Hartman (2007) show a comprehensive map of digital video formats in education, organized around four classes of intended outcomes: saying, seeing, doing and engaging. In a recent review of video usage in higher education, Winslett (2014) found these learning objective categories for video in instruction: show factual and procedural content, directly instruct/describe, provide exemplars, show real life practices and contexts, show complexity and trigger better practices, and democratize video production.

The communicative qualities of films and videos in education are summarized in the words that F. D. McClusky wrote in 1947:

> *The motion picture* [...] *is essentially a multiple method of communication. It is especially effective as a technique for telling a story. It presents facts realistically. It dramatizes human relations and events. It arouses emotions. It transmits attitudes. It records and reproduces phenomena for scientific study and analysis. It depicts the imaginative. And it can enable to see the unseen. (McClusky, 1947).*

The same ideas were revisited sixty years later by Koumi (2006), who states that video is well suited for three educational values: *cognitive value*, *nurturing value* and

experiential value. The cognitive value is concerned with demonstrations of processes, visual descriptions of concepts; the nurturing value is concerned with the affective connection with the learner; and the experiential value is concerned with sharing other's experiences and interactions.

Film and video technologies allow instructional content to be distributed to a large audience with a very low cost, compared to printed media and to direct classroom instruction from a teacher. This enabled the concept of educational television, and, later, video-intensive online learning as we can see in current MOOCs. Moreover, video instruction would serve as a replacement for direct face-to-face instruction when the latter is not feasible (e.g. a geographically disperse audience, or people with reduced mobility). Video instruction has also been claimed to be beneficial for students at risk or hard to reach (Passey, 2006). Video introductions have been reported to increase teacher's and student's social presence (Garrison, Anderson, & Archer, 1999) in distance education.

The inherent characteristics of the video medium in instruction have been studied and compared to other media attributes. Animated graphics have been found superior than static graphics in knowledge acquisition (see the meta-analyses by Berney & Bétrancourt, 2016; Höffler & Leutner, 2007). There is also a strong evidence that the simultaneous presentation of pictures and speech improves learning, which has driven to the enunciation of the *multimedia effect*: "people learn more deeply when they receive an explanation in words and pictures rather than words alone" (Mayer, 2002, p. 105). This effect is a strong empirical argument for video instruction over printed or audio-only media[3].

Finally, many studies have shown that different media types can lead to similar learning outcomes (Donkor, 2010; Means, Toyama, Murphy, Bakia, & Jones, 2009; Merkt, Weigand, Heier, & Schwan, 2011). These findings suggest that, even if one takes a cautious position, video-based instruction is not a harmful option in learning scenarios where other media are not possible (e.g. massive distance learning).

Disadvantages and criticism

On the other side, video as an instructional medium has some disadvantages. It enforces a sequential, constant-rate processing flow in the viewer (tempered by some navigational capabilities of modern streamed video). Compared to printed text, video requires much more real-time attention to grasp the whole contents: it is more sensitive to viewer's distraction (Große, Jungmann, & Drechsler, 2015; Kozma, 1991). Fortunately, this attentional disadvantage can be addressed by trimming the content in short-length segments, weeding out extraneous contents and using signals to reclaim attention to relevant content (an excellent case study is described in Ibrahim, Antonenko, Greenwood, & Wheeler, 2012). Nevertheless, it imposes a burden on instructional designers that printed books do not bear.

[3] I will elaborate more on the multimedia effect and multimedia learning principles in Chapter 2.

Stetz and Bauman (2013) warned about thoughtless use of video as a teaching resource. They identified up to thirteen factors to consider before integrating video in the classroom: video making students read less, lack of interactivity, and adherence to the "lecturing" model, among other factors. The risks of passive watching and superficial learning due to 'edutainment effect' have also been warned (Fill & Ottewill, 2006). There is also the risk of learner's overconfidence when learning from video (Salomon, 1984; Szpunar, Jing, & Schacter, 2014), and the related Pygmalion effect (Fries, Horz, & Haimerl, 2006). Some authors have noticed an exaggerate increase of video as a communication tool in educational and corporate contexts, which has been labeled as "video is the new PowerPoint" (Leshem, 2018), making a parallelism with the infamous role of PowerPoint as a presentation killer, e.g. "death by PowerPoint" (Garber, 2001). This statement should warn us about misuses of video in instruction.

Moreover, the argument of technology skeptic Richard Clark (1983, 1994) is worthy of consideration: it is not the instructional *medium* that produces benefits, it is the instructional *method*. Some meta-analyses on the effect of technology in learning (Schmid et al., 2009; Tamim, Bernard, Borokhovski, Abrami, & Schmid, 2011) point to a Clark's rebuttal: technology has shown proven positive effects in instruction. In spite of those evidences, I think that Clark's argument still contains a pertinent call to restrain excessive techno-enthusiasm. That is a reason to encourage research on video instruction that sheds light on the actual benefits of videos in instruction and how to introduce video media in instructional designs.

1.3 Goal

Once the matter of study has been put into context, I will return to the point at which I arrived when I decided to start this study. To recapitulate, I came to this position about the status of research on the learning effectiveness of instructional video characteristics:

- There is a need for a *map*: a cartography that organizes the research findings on video characteristics, using higher-level structural categories.
- There is a need to make explicit the full variety of relevant features in instructional videos, to regain areas of research that have been neglected in the last two decades.

To overcome this need, I propose this measure: to build a comprehensive *classification scheme* of instructional video characteristics. Now I will formulate this proposal as a main research goal and a main research question, and in the following section I will justify the proposal.

Main goal:

to build a classification scheme for instructional video characteristics

Main research question:

how can instructional video characteristics be systematically and usefully classified?

By **systematic**, I mean that this classification must be constructed in accordance with a planned method and must comply with sound scientific criteria.

By **useful**, I mean that it must provide meaningful, non-trivial information to the scientific community of interest. The beneficiaries of this study are the people who are involved in instructional video research, design and production.

The main goal explains why this thesis is entitled **"Anatomy of Instructional Videos"**. The New Oxford American Dictionary defines *anatomy* as "a study of the structure or internal workings of something". The Merriam Webster Dictionary defines it as "the art of separating the parts of an organism in order to ascertain their positions, relations, structure, and function". Certainly, this research aims to make an *anatomy* of instructional videos, to identify and classify those structural constituents that contribute to their instructional function.

1.4 Scope

1.4.1 Perspective: characteristics with influence in learning

A classification scheme or a taxonomy are always built from a given perspective. An object of study can be observed from different beliefs, purposes and epistemological views. That is, any taxonomy will be *arbitrary* and, consequently, taxonomies are not discovered, but crafted with a certain degree of choice. This ultimately means that one principal driving force when building a taxonomy is its *usefulness* (Reigeluth, 1983, p. 13).

In order to succeed in the usefulness of the classification to build in this study, an imperative delimitation of scope must be stated on the characteristics to be studied: this study shall be focused on characteristics in an instructional video that have—or may have—an effect in the learning outcomes derived from the use of the video under some instruction model and conditions. This delimitation includes both characteristics with observed effects (positive or negative) on learning outcomes, and those with an acknowledged *potential* to influence the learning—even if their effects have not been sufficiently studied so far.

I am aware that the relation between a single characteristic and the learning outcome of the whole instructional video may be weak, since multiple factors are in play: individual characteristics of learner, instructional design, etc. What is more, the very concept of 'learning effectiveness' of media attributes is contentious (R. E. Clark, 1983, 1994). I will dive into detail in Chapter 4 about this matter. The set of characteristics will be collected from the available research literature on video-based learning and related fields, as will be discussed in the 'Strategy' section of this chapter.

1.4.2 Typology of videos covered by this study

For the sake of feasibility, this study will be confined to a limited range of educational videos. More precisely, the study will include videos that meet all these five attributes: *academic*, *intrinsically instructional*, *instructor-authored*, *asynchronous* and *streamed*. These constraints make it possible an object of study that can be analyzed in the reasonable time span of a doctoral thesis, without losing the power of generalization of the findings from the study.

The **academic** attribute means that this study will be focused on videos that are produced and used in a post-compulsory, college learning setting. I will not consider videos specifically made for children, K-12 stages, non-academic adult education, or industrial training. Nevertheless, I will borrow findings obtained in other learning domains when they are reasonably extrapolatable to the academic context.

The **intrinsically instructional** attribute means that this study will take into account videos whose intrinsic intent is to provide instruction by themselves, with little or null need to rely on external resources. The most representative cases that meet this criterion are lectures, tutorials, demonstrations, worked examples,

documentaries and interviews to experts. I will therefore exclude from this study communicational videos (e.g. teacher-learner video chats, teacher feedback to assignments), trigger videos, welcome videos and advertisements. The study will also exclude non-instructional films that are commonly used as a teaching resource, for instance a scene of a feature film used to support a case-based session.

Only **instructor-authored** videos will be analyzed. Learner-produced instructional videos are completely out of the scope of this study. The "instructor-authored" term does not mean necessarily that it is an instructor who records or edits the video. Rather, this term refers to the *intellectual conception* of the video contents comes from an instructional designer.

This study will be limited to videos intended for **asynchronous** communication, which excludes synchronous video formats in which speakers and learners can interact in real time: videoconferences, live interviews and the like.

Finally, the study will focus on **streamed** videos, that is, videos that are delivered online to be watched asynchronously by learners at their demand. More specifically, this study is not directly concerned with the use of videos within a face-to-face classroom instructional design, though this study's findings could be easily extrapolated to those settings.

1.4.3 Included and excluded video formats

To recapitulate the limitations imposed in this research on the objects of study, I will list what video formats are explicitly included in the study and what formats are excluded.

Included formats:

- Documentaries
- Lectures (including recorded classroom lectures)
- Tutorials, demonstrations and worked examples
- Interviews to experts, testimonials and 'vox pops'

Excluded formats:

- Video diaries
- Trigger videos
- Recorded seminars
- Video conferences
- Mashup videos

1.5 Motivation and significance

The assumptions that have motivated this study need a justification. The relevance of the main research goal also needs a rationale. There are three main questions to address:

a) What are the benefits of building a classification scheme for instructional video characteristics?
b) Are classification schemes already available that make this research unnecessary?
c) Are there relevant features in instructional videos that have been overlooked in recent research in video-based learning?

This section will give answer to these three questions, thereby providing a justification for the proposed research.

1.5.1 What are the benefits of a classification scheme?

Scholars from many knowledge fields agree that classification schemes improve the understanding of the domain being classified (Nickerson, Varshney, & Muntermann, 2013; Pope, 1994; Reigeluth, 1983). Classification is particularly helpful in understanding complex domains (Nickerson et al., 2013). Bailey (1994) discusses extensively the advantages of classification in social science research. Among other advantages, Bailey mentions the reduction of complexity, the identification of similarities and differences among cases, easy comparison of types and easy study of relationships.

Conclusively, (Vegas, Juristo, & Basili, 2009) have accurately summarized the benefits of building a taxonomy for any knowledge field in these three properties: a) it provides a set of unifying constructs; b) it helps to understand interrelationships; and c) it helps to identify knowledge gaps.

Every emerging research subject demands some classification work. A case that is close to the current study is that of video games: recently, researchers have developed taxonomies of video game genres (Gunn, Craenen, & Hart, 2009) and video game structural properties (Wood, Griffiths, Chappell, & Davies, 2004). Even closer to instructional videos, the growing interest in multimedia design at the end of the 20th century prompted a long record of classifications and taxonomies over various aspects, such as representational elements (Bernsen, 1994; Heller & Martin, 1995) and interactivity (Aleem, 1998; Schwier, 1992). These examples show how communities of research try to achieve a better understanding of a new domain by means of classification schemes.

An example with instructional videos

Let us define an example to illustrate the convenience of a classification system. Table 1-1 shows the names and description of four multimedia learning principles (Mayer, 2014c): multimedia principle, coherence principle, personalization principle and voice principle. After reading the descriptions, it is clear that the first two (multimedia and coherence principles) have to do with the presence or absence of representational items (voice, text, pictures), while the other two (personalization and voice principles) have to do with the learner's response to the speaker, due to the perceived social distance.

There is a latent grouping in this set of principles: two representational principles and two social distance principles. These two relationships could be used as grouping criteria to classify learning principles and video characteristics in general. Therefore, we could create a 'social distance' grouping in which we include these two characteristics: 'discourse personalization' (with two possible values: impersonal and personalized) and 'speaker's accent' (whose values are combinations of human/robotic and native/foreign values).

The mere definition of the category 'social distance' has some value, because we are now able to group concrete video attributes around a higher-level concept. But there is more. One can speculate on more video characteristics not yet assessed that could belong to this class of social distance attributes. Does the degree of formality of the speaker's discourse affect to learning outcomes? What about the speaker's gaze: is learning influenced if the speaker makes frequent eye contact with the camera? What about the social status cues that may be portrayed by the speaker's garment? These characteristics are examples of candidates to belong to the 'social distance' category, whether they have been assessed in video-based learning research or not. This would be an example of the usefulness of classification in the discovery of unexplored areas in the study of video effectiveness on learning.

Table 1-1. A selection of four multimedia learning principles

principle name	people learn better when…
multimedia	words and pictures are used rather than words alone
coherence	extraneous material is avoided
personalization	words are presented in conversational rather than monologue style
voice	words are spoken in a standard-accented human voice

1.5.2 Are classifications of instructional video characteristics available?

Once the importance of having a classification system is justified, the next question is to find out as to whether valid classifications already exist for instructional video characteristics.

Probably the work that best represents the goal of providing a 'map' of the characteristics that influence the learning effectiveness of instructional videos is the report that Wetzel, Radtke and Stern made for the U.S. Navy: *Review of the Effectiveness of Video Media in Instruction* (Wetzel et al., 1993), later published as a book (Wetzel et al., 1994). This work gathered the findings on the effectiveness of all kinds of video features, covering both professional best practices and scientific

research on learning with film and video. In fact, if I were writing these lines before 2000, the main research goal of this study should be sufficiently satisfied by the work of Wetzel et al, except for minor reorganizations to achieve a systematic classification scheme. But now in the 2010s, Wetzel et al.'s work is considerably outdated, as it leaves out most developments of cognitive theories of multimedia learning, which are posterior to the cited report. Even the standard terminology has changed substantially.

It is hard to find a work later than Wetzel et al.'s that has a broad coverage of characteristics. There are several classification schemes on particular structural and functional domains, for example semiotic structures of educational television programs (De Vaney, 1991), gestures in lecture videos (J. R. Zhang, Guo, Herwana, & Kender, 2010), video annotation features (Aubert, Prié, & Canellas, 2014), interaction design patterns in video learning environments (Seidel, 2015) and video presentation formats in MOOCs (Hansch et al., 2015). But I have been unable to find a work that integrates all this research into a global picture. A couple of reviews of research on video-based learning have provided classification schemes for instructional videos. Kay's review on video podcasts (Kay, 2012) includes a typology based on four dimensions: purpose, pedagogy, segmentation and academic focus. Winslett (2014) categorizes educational videos around three dimensions: learning objective, educational topic and production type. But in both schemes, Kay's and Winslett's, structural categories of characteristics are only superficially covered. On the other side, Ploetzner and Lowe (2012) built a systematic characterization of instructional animations that matches the main goal of this study, but applied to a different type of learning object.

→ In summary, there exist many up-to-date categorizations of particular features, but *I have not found a modern top-level map that encompasses all the feature domains that are relevant to understand the structure of instructional videos.*

1.5.3 Are there overlooked characteristics in current research on video-based learning?

Computer killed the video star

To understand the gaps in current instructional video research, it is necessary to relate their historical roots. At the end of the 1970s, there was a drastic change in the interests of the educational technology sector: computers entered the scene. Computer Assisted Instruction took the lead in instructional research, initially with text-based applications and later with multimedia software. Much of Richard Mayer's group's research took computer-generated animations as the primary instructional material to develop their multimedia learning theories. Video was relegated to a secondary role, as another building block for multimedia objects.

A new scientific community was emerging around Multimedia Learning (MML), and in some way it grew at the expense of the pre-existing Educational Television (ETV)

community, as they competed for the same resources. Quoting Barford and Weston (1997):

> *As Educational Television Services within Universities fall prey to cutbacks, and investment in bespoke video gives way to Computer Based Learning and multimedia technologies, does video still have a place in Higher Education? (Barford & Weston, 1997).*

The newborn MML scientific community developed independently of the existing ETV community. There was little transfer of knowledge between them. ETV could only languish and contemplate its replacement by another scientific community. Gavriel Salomon, one of the leading members of the ETV community, narrated his own experience:

> *Research on the "old" media employed in education was giving way to the growing interest in a new and exciting technology—computers.* [...] *The old kind of research on media, particularly the visual media, its trials, many tribulations, and (alas, so few) successes were more or less forgotten. (Salomon, 1994, p. xvii).*

At the beginning of the 21st century, there was a renewed interest in instructional videos as a primary research object, fostered by technological changes (web, streaming, cheap recording devices). But by then the ETV community had been wiped out and its knowledge forgotten. The new researchers on instructional video were starting from scratch in many relevant aspects. The loss of the ETV community's legacy is particularly noticeable in two areas: film production technology and semiotic analysis of videos.

Film production technology

One of the main consequences of the disruptive replacement of communities is the lack of continuity of research in film production techniques, which flourished in the 1970s and 1980s (e.g. John Baggaley et al., 1980; Salomon, 1979a). To illustrate this point, let us take Wetzel, Radtke and Stern's review on learning efficiency of video (1993). It contains a brief section about professional best practices in editing techniques that may influence the viewer's response (pp. 96-98). These pages mention features such as frequency of shot change, appropriate moment to make the cut, differences and continuity between cuts (spatial, temporal and semantic continuity), use of cutaways, reverse-angle shots and camera viewpoint. Another section about cueing techniques for increasing viewer's attention (pp. 112-117) refers 40 selected scientific works[4] written between 1960 and 1991 that analyzed the effect of edition techniques such as zooming, dollying, shot length, panning, camera angle, cutting rate and continuity of shots.

The above excerpts show a diversity of topics that is unseen today in instructional video research. There are indeed some recent studies on camera viewpoint (Fiorella,

[4] I have discarded a couple of references prior to 1960.

van Gog, Hoogerheide, & Mayer, 2017), instructor's camera angle (Beege, Schneider, Nebel, & Rey, 2017) and cutting instructor's shots (Díaz, Ramírez, & Hernández-Leo, 2015; Rene F. Kizilcec, Bailenson, & Gomez, 2015), but they constitute a very recent trend, are still relatively scarce and are constricted to a very short range of film edition techniques.

As regards film production techniques, there is another prominent trait in many recent scientific papers on video-based learning: a disconnection with the discipline of film production. This is something that Greg Winslett has noticed in his review of research on educational video in higher education (Winslett, 2014). In his conclusions, he expressed that within the examined research "the vocabularies of film and TV production appear conspicuously absent". He follows on with this concern about neglecting production techniques when designing academic instructional videos (bold highlights are mine):

> *For example, a mise-en-scène understanding of educational video offers a shared vocabulary to consider how lighting, proxemics, framing, depth of field, camera angle, lens and film stock, composition and form may be used to support particular outcomes and signify specific learning activities.* ***Not considering these vocabularies or ways of thinking is to risk conveying a visual element that runs contrary to the desired outcome****. For example, poor consideration of these elements may result in a person who is considered an expert in their field being filmed as if they were a novice, unsure of their experience or position.*

Winslett's example does not only mean that poor knowledge of film production technology can lead to flawed instructional products. It also means that the results from some research experiments on instructional videos may be confounded by non-controlled effects originated in video production features. It is critical to (re)incorporate that knowledge on film production to improve the reliability, reproducibility and comparability of research.

Semiotic structure of instructional videos

Semiotics deals with how humans make *meaning* through communication systems: speech, text, signs and many other ways to build messages. A *symbol system* is a set of elements such as words, numbers, shapes, camera movements... that interrelate by syntactic rules (formal or informal) in order to create messages. Semiotics is strongly interlaced with instruction, as an instructional activity often takes place through an interchange of messages between participants. Those messages not only contain a concept or skill to be learnt, but also convey social and affective codes with relevance in the learning process.

Many educational researchers were interested in the semiotic structure of educational videos. The most prominent was Gavriel Salomon, who considered symbol systems as the "most essential characteristic" of a medium (Salomon, 1979b). Salomon observed that the level of understanding of the specific codes of a medium (e.g. television) influenced how children acquired knowledge from than medium

(Salomon, 1979a). Other researchers investigated the semiotic codes in educational television, as the grammar sketched in (De Vaney, 1991).

This structuralist perspective has been lost in the current research on instructional videos. Meanwhile, Semiotics and related linguistic theories such as Rhetorics have been applied to areas as the automatic generation of multimedia presentations (André, 2000; Taboada & Mann, 2006), the design of technical documentation (Farkas, 1999) and the analysis of television news (Morales Morante, 2012; Wojcieszak, 2009), to name a few. New developments in discourse analysis, genre theories and multimodality have fostered the semiotic analysis of academic discourse, for example the academic lecture (Deroey & Taverniers, 2011; Fortanet Gómez & Bellés Fortuño, 2005). Nevertheless, current research in video-based learning is not taking advantage of the knowledge that is being generated in these areas for a better understanding of instructional videos.

1.5.4 Conclusion: significance of this research

The preceding sections have shown that the proposed research makes sense in the current state of the art:

- First, it would be the first comprehensive systematic classification of instructional video features in a long time.
- Second, this classification scheme can contribute to improve the knowledge about video-based learning and instruction.
- Third, this classification scheme can help to bring a wider perspective about the features of instructional videos that have influence in learning outcomes, integrating knowledge findings from various communities of research into a single documental source.

1.6 Strategy and method

In order to achieve the research goals, I have implemented this strategy of action:

1. Identify the main scientific disciplines that will shape this research.
2. Find key scientific works in these disciplines that will provide a first set of relevant video characteristics.
3. Perform a field study on online courses, to obtain evidences of usage of video characteristics.
4. Perform a process of classification of characteristics, which will result in a classification scheme.
5. Refine the classification scheme with an extensive literature review, which will provide domain-specific taxonomies of characteristics.

The first step of this process is to identify the main scientific fields that will contribute to gain a comprehensive view of instructional video characteristics. Two fields that have been elicited in this first chapter are the classic research on educational films and television and the more recent community of Technology Enhanced Learning, particularly the theories on Multimedia Learning. This couple of fields has the drawback that it does not adequately cover certain communicative aspects of audiovisual products. In the case of the classical body of research, because it is outdated. In the case of recent research, as I have warned in this introduction, these aspects in instructional videos have been overlooked.

The field that will complete the gap stems from modern developments in Linguistics and Semiotics. Multimodal Discourse Analysis encompasses research on the process of meaning making in multimodal communication artefacts, such as pictures, web pages and, of course, videos.

Taking these three fields as root knowledge sources, the second step in this research has been to identify paramount research works within these fields that serve as references to relevant characteristics of instructional films and videos.

In addition, I have conducted a field study on current online platforms to assess the actual usage of instructional videos, and to obtain direct evidences on frequent structural features.

The following step in the strategy is the most decisive for this research. With the key scientific references identified and the insights from field evidence, I have performed a process of classification of characteristics, through a bottom-up method. This process starts with the identification of an inventory of characteristics that researchers have been considered relevant in learning with video. It continues with a conceptual clustering that results in a set of classification domains for the characteristics. This preliminary classification scheme has been confronted to conceptual frameworks found in Multimodal Discourse Analysis. I have found that John Bateman's GeM framework for multimodal printed documents (Bateman, 2008) suits very well to the classification scheme and, with minor modifications, can provide a sound theoretical support to the classification.

The classification step produces a classification scheme which is already a useful deliverable of this research. Nevertheless, I have taken a step further in order to enhance the classification and to provide a stronger evidence support. An extensive literature review has been made for each classification domain, resulting in a set of domain-specific taxonomies of characteristics. This refinement also helps to make fine adjustments on the previous classification scheme.

1.7 Detailed research questions and objectives

Once the goals and the method of this research have been described, we can reformulate the main research question into a set of more concrete research questions which correspond with some procedures of the strategy.

1.7.1 Research questions

Primary research question:

For academic instructional videos,
how can their characteristics that are influential in learning outcomes be classified around useful higher-level concepts?

Secondary research question 1:

How can Multimodal Discourse Analysis theories and findings be used to provide a classification scheme for instructional video characteristics?

Secondary research question 2:

What presentation formats are being used by the academic community for online instructional videos?

1.7.2 Research objectives

- Build an extensive inventory of academic instructional video characteristics:
  - Include properties that have a potential in fulfilling learning goals.
  - Include properties that are meaningful for video designers and instructors.
  - 'Regain territory': recover areas of interest that the 'multimedia learning' community has taken apart.
- Build a classification scheme of academic instructional video characteristics.
- Make an inventory of instructional video presentation styles.
- Make a review of research of the learning effectiveness of academic instructional video characteristics.

1.8 Structure of this dissertation

The remaining of this dissertation is structured as follows.

Chapter 2, "Characterization of Instructional Videos", contains a literature review about the topics that are closely related to this study. First, instructional videos are characterized as three types of entities: expository films, multimedia learning objects and multimodal texts. Next, two key scientific fields for this study, Cognitive Theory of Multimedia Learning and Multimodal Discourse Analysis, are introduced. A literature review follows that collects selected scholarly works that provide insights to the research questions posed in this study.

Chapter 3, "A Survey of Instructional Videos in MOOCs", presents a field study performed on Massive Open Online Courses (MOOCs) to assess the current and actual use of instructional video presentation formats and other key features. The results from this piece of research have already been published in a JCR-indexed journal (*Technical Communication*) in the year 2016 (Santos-Espino, Afonso-Suárez, & Guerra-Artal, 2016).

Chapter 4, "Building the Taxonomy", describes the process of building the classification scheme of video characteristics. Both the process design and the intermediate outcomes of the process execution are described.

Chapter 5, "A Taxonomy of Instructional Video Characteristics", summarizes one of the main research objectives of this thesis. It describes the full classification scheme as a result of the strategy described above in this Chapter 1. Both the main classification scheme and the domain-specific taxonomies are presented.

Chapter 6, "The Taxonomy in Detail", expands the description made in Chapter 5. All the components of the classification scheme are explained in detail, accompanied by annotated references to the scientific works that exemplify and support each of them.

Chapter 7, "Conclusions", summarizes the findings obtained across this research, discusses practical implications of the proposed classification scheme and identifies possible future research based upon the results of this thesis.

Chapter 2. Characterization of instructional videos

2.1 Chapter overview

I have conducted a preliminary research in order to situate instructional videos within appropriate analytical frameworks. With this goal in mind, I have performed an extensive literature review in which I have found various epistemological approaches to instructional video research. Cognitive learning sciences and discourse analysis are the two knowledge fields that may give an appropriate theoretical support to my thesis. This chapter presents the results of this research process.

This chapter goes beyond a mere literature review, as it introduces a characterization of instructional videos. This characterization takes three perspectives: videos as films, videos as learning objects and videos as texts[5] (linguistic-semiotic artefacts).

This chapter is structured as follows: first, I characterize instructional videos according to their three natures (film, learning object, text). Then I discuss the two main scientific sources that form the theoretical backbone of my thesis: cognitive theories of learning and multimodal discourse analysis. A final section shows a selected list of research works that are closely related to the objective of this thesis and will serve as the basis for building the classification scheme of video characteristics.

2.2 The three natures of instructional videos

In Chapter 1 I have explained how historically two perspectives on research in instructional films and videos have followed one another. In a first period, these products were examined as *educational films*. Later they were incorporated into the field of study of *multimedia learning objects*, which is the dominant perspective in current scientific research on video-based instruction. Now, we can consider a third approach to analyzing videos, rooted on the modern schools of Linguistics and Semiotics: Discourse Analysis theories and methods. From the perspective of Discourse Analysis, videos can be regarded as *texts* (productions of human language) which portray a discourse and build meaning.

In general, as communicative artefacts, all modern instructional videos share some key attributes: they are *expository*, *multimodal* and *digital*. The word 'expository' takes different meanings depending on the discipline. From a film analysis perspective, *expository* is opposed to terms as *narrative* and *participatory* (Brewer, 1980; Nichols,

[5] 'Text' is not used here with the common meaning of 'written language', but with the broader meaning of any language utterance, regardless of its mode.

2001). From an instructional design perspective, *expository* is opposed to terms as *interactive* and *collaborative* (Ormrod, 2017). The multimodal quality of videos makes them a type of *multimedia* learning objects (Churchill, 2007), and also a type of *multimodal* texts (Paltridge, 2012). Finally, the digital quality of modern instructional videos has added them some features over their analogue predecessors, in particular more affordances for interactivity (Merkt et al., 2011).

In brief, we conclude on these three natures of instructional videos:

- Instructional videos are **expository films**.
- Instructional videos are **multimedia learning objects**.
- Instructional videos are **multimodal texts**.

These three perspectives on the nature of instructional videos will shape the building of the classification schema of this research work.

During my research, I have explored what film analysis, multimedia learning theories and discourse analysis say about instructional videos and other related artefacts. As a result of my literature review, I have identified some properties that distinguish instructional videos from other communicative artefacts:

- As films, they are **expository** rather than narrative.
- As films, they usually use a **formal voice**.
- As films, they usually have a **short duration**.
- As films, they usually show **low visual complexity**.
- As learning objects, they are generally **presentational**.
- As learning objects, they are generally suited for **expository instruction**.
- As learning objects, they are **moderately interactive**.
- As learning objects, they show a **moderate level of multimodality**.
- As learning objects and as texts, they are usually **multimodal**.
- As texts, they usually show a **discourse of "truth and trust"**.

I will elaborate these statements along the remainder of this section.

2.2.1 Instructional videos are expository films

Instructional videos constitute a particular class of *expository films*. An expository film differs from a *narrative film* in that there is no story or plot to tell. Moreover, expository films usually portrait *nonfictional* facts, though it is not a required property. Brewer (1980) makes a more precise delimitation of narrative and expository texts. He defines narrative texts as "events that occur through time and are related through a causal or thematic chain. Narrative texts center on one or more protagonists who carry out different actions in order to satisfy a goal." According to Brewer, expository texts describe the *structure and processes involved in a system or event*.

In a line, expository films are *information-centric*, while narrative films are *story-centric* or *character-centric*.

Besides instructional or educational films and videos, other well-known classes of expository films are television broadcast news. Many documentaries and television commercials are also expository.

Film production techniques

Many scholars have studied how films are constructed, what components are involved and how they contributed to the film outcome in various dimensions, such as aesthetics and discourse. I will just mention Noel Burch (Burch, 1970) as a contributor to understanding the main building blocks of cinematographic visual language (e.g. the 'six spaces' model), and also Herbert Zettl and his monumental work *Sight, Sound, Motion: Applied Media Aesthetics* (Zettl, 2016), which covers all aspects of film production.

Zettl, Burch and many other film production experts have shown how various cinematographic techniques can be used as a means of artistic expression and expository communication: camera movements, frame layout, montage, mise-en-scène, etc. Historically, educational researchers have harnessed this knowledge on film production techniques to foster the effectiveness of film and video instruction, especially during the age of educational television (Wetzel et al., 1993).

Distinctive features of modern instructional videos

Recently, digital technologies and streaming delivery have favored the hybridization of educational videos with other media, such as computer animations, digital slides, interactive forms and hypertexts. This has given rise to so-called *multimedia learning objects*. For instance, it is now common that an instructional video is embedded as a learning resource into a web-based learning management (LMS) platform. This digital media hybridization has progressively distanced the structure of instructional video from that of the classic genres of expository and nonfiction films (such as documentaries) and it has accentuated its most distinctive features, such as a short duration and a low expository complexity.

Short duration

Most instructional videos have a short duration, with typical values of 2-15 minutes (see Chapters 3 and 5 for some field studies). Common exceptions to this short length are videos of unedited recorded classroom lectures. This short length of instructional videos is in line with TV commercials and TV broadcast news, which aim to convey a clear and short message, as instructional videos often do. Documentaries would be the only type of expository film that features long durations (20-200 minutes), similar to narrative films.

The short length of instructional videos is a property that has been with them practically from the beginning. In fact, by the half of the 20th century it was suggested that educational films should last no longer that 10-15 minutes (Sumner, 1950).

Some attributed the custom of short durations to mere technical reasons (e.g. to adjust to the size of film rolls), but the fact is that short time lengths have been pervasive over a century of educational films and instructional videos, and there is a trend to shorter durations in online streamed videos.

Low visual complexity

One of the most salient features of a typical instructional video when it is compared with other classes of expository films is its low visual complexity. This relative simplicity is really extreme when one compares a 15-minutes long, single-shot video lecture with an usual 20 second TV commercial having 15 or more shots. On the other hand, many video lectures consist of a single scene[6], perhaps alternating two or three different shots, such as a shot of the speaker and a shot of a board/slide. One study on actual video structure (Liu & Kender, 2002) reported that a typical one-hour class meeting video was composed of 50-100 scenes/shots, thus giving an average shot length of one minute. This is 15 times the typical shot length of a commercial video (4 seconds, according to the aforementioned study). The term *"illustrated radio"* has sometimes being used to refer video lecture recordings. This label illustrates well how certain instructional video modalities do not place much emphasis on visual editing.

Moreover, instructional videos are very frugal in terms of editing and shots. Some authors suggest that "such fast editing may not be needed as the video is shown in the context of other educational resources and activities" (Thornhill et al., 2002). To put it in other words, a television broadcast program must rely on itself to attract its audience, while a contemporary educational video is usually one piece in a wider educational resource, such as a web site or a multimedia-enhanced unit.

Film analysis and film semiotics

In the mid-twentieth century an interest arose in analyzing the semiotics of films, that is, exploring the internal structures of film products that are used to convey meaning. One of the most representative authors of the first wave of film semiotics was Christian Metz and his *Grande Syntagmatique* (Metz, 1966), in which he proposed a linguistic analysis of narrative films, organized in a taxonomy of syntagmatic structures. Another important contributor was Umberto Eco (Eco, 1976), who provided a more general view of film semiotics, not attached to linguistics like Metz, but closer to the Peircean semiotics. According to Eco, film images and sounds cannot be regarded as 'messages without codes': they have their own extralinguistic codes.

Eco's claim that films have their own semiotic codes was joined to the evidence that those codes were cultural and therefore had to be learned (Sekula, 1982). This intellectual climate fostered that educational researchers as Gavriel Salomon

[6] A *scene* is the recording of a continuous activity, as perceived by the viewer.

(Salomon, 1979a) advocated investigating film symbol systems as a key element in learning.

Voices and modes in expository films

Several scholars have dived into the structural analysis of nonfiction films (Nichols, 2001; Plantinga, 1997; Ruoff, 1992). One of the most interesting features of documentaries is the notion of *voice* or *mode* in which the film discourse is portrayed.

Plantinga found that documentaries may use these three voices: formal voice, open voice and poetic voice (Plantinga, 1997). The **formal voice** makes an "explanation [...] with a high degree of epistemic authority" (p. 107). The **open voice** is "epistemically hesitant [...], observes and explores rather than explains" (p. 108). The **poetic voice** is concerned not so much with explanation or observation as "art and/or as means of exploring representations itself" (p. 109). According to Plantinga, all three voices assert that what is presented in the film is actual and real. The differences lay in the epistemic authority, the hesitancy or the aestheticism.

The majority of instructional videos use the Plantingan formal voice, as it generates a tone of authority and reliability on the contents. Besides, the 'fly on the wall' format, in which the camera is filming real-life events with almost no intervention, is more related to an open voice.

Bill Nichols evolved the Plantingan notion of voices. In his *Introduction to Documentary* (Nichols, 2001), he introduced the six 'documentary modes' that characterize all varieties of this genre: *poetic*, *expository*, *participatory*, *observational*, *reflexive* and *performative*. Most educational films and videos show traits of the expository mode: the use of a narration voice over the footage (*voice of God*), and the dominance of exposition of facts over affective judgements, among other features. Though, traces of all six modes can be found in several instructional works.

2.2.2 Instructional videos are multimedia learning objects

As I described in Chapter 1, in the last quarter of the 20th century the interest in technology-based educational research turned radically to Computer Aided Instruction. This meant that videos and films came to be considered more as components of computer-based *learning objects* than as self-fulfilling products. Moreover, digital technologies have enabled the building of new kinds of instructional materials that combine several representations: text, voice, illustrations, video (moving images) and interactive animations, among others. Some popular classes of those *multimedia learning objects* are slide presentations, web pages, animations, videos, computer games and computer-based simulations.

A notable amount of research has been done to investigate how the use of multiple representations in learning materials influences the learning process. As I will discuss later in this chapter, a full-blown theory has been developed, the Cognitive Theory of Multimedia Learning, around the hypothesis that people learn better if content is provided in pictures and audio simultaneously.

Standard definition of learning object

The Learning Object Metadata standard defines a *learning object* as "any entity, digital or non-digital, that may be used for learning, education or training" ("IEEE Standard for Learning Object Metadata," 2002).

Cisco Systems corporation has refined the former definition (CISCO SYSTEMS, 2003) into what they call a *Reusable Learning Object (RLO)*. An RLO is characterized by these four features: it is based on a single learning or performance *objective*; it is built from a collection of *content* and *practice* activities; it can be tested through *assessments*; its contents are identified with *metadata* that enables referencing and searching.

Characterization of digital learning objects

Churchill (2007) aimed at overcoming the vagueness of these previous definitions and crafted his own characterization of learning objects: "a learning object is a representation designed to afford uses in different educational contexts". In this definition, the term *representation* is understood as a translation of some existing entity into some digital medium. Churchill's definition implies that this representation is an integration of multiple media modalities into a single meaningful learning unit.

Churchill also presented a classification of learning objects, comprising six types: presentation, practice, simulation, conceptual model, information and contextual representation (see Table 2-1). As stated by Churchill himself, instructional videos are canonical instances of the *Presentation* type, although they may contain embedded items belonging to other types (e.g. in-video interactive quizzes, conceptual maps).

Table 2-1. Characterization of digital learning objects (Churchill, 2007)

learning object type	description
presentation	direct instruction (e.g. lecture, PowerPoint)
practice	drill and practice, learn procedures (e.g. quiz)
simulation	representation of real-life systems
conceptual model	representation of key concepts
information	display of information with multiple modalities
contextual representation	data displayed as it emerges from a simulated scenario

Interactivity and multimodality

Instructional videos can be characterized against other learning objects according to two parameters: *interactivity* and *multimodality*. Interactivity is about the user's ability to manipulate the object's behaviour or appearance. Multimodality accounts the

variety of representational modes in the object (printed text, speech, sound, pictures, animations, video, 3-D virtual environments, etc.).

Figure 2-1 locates some common learning object types on a coordinate map in which the X-axis represents the degree of multimodality and the Y-axis represents the degree of interactivity. Digital videos are placed in a relatively low position for both parameters. They only surpass printed texts and analog films. As regards interactivity, digital video enables a certain degree of interactivity, especially the playback control and the possibility of content-based navigation. Moreover, a video could have embedded items that increase its level of interactivity (e.g. in-video quizzes). I will discuss these elements of interactivity in subsequent chapters of this dissertation.

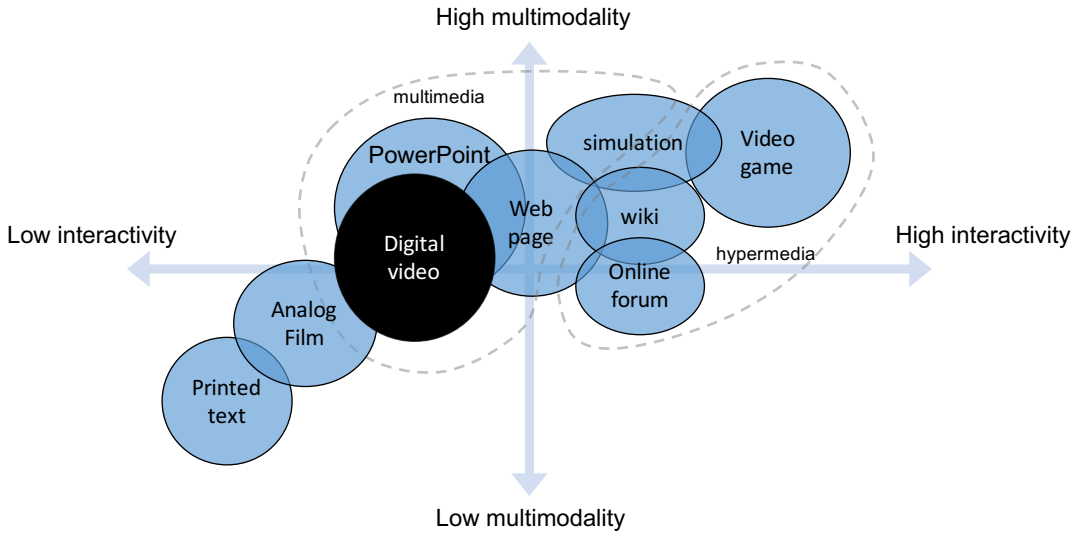
Expository instruction

Learning objects can be used in a variety of instructional designs. In the case of the instructional videos that are part of my research, their expository nature, their limited interactivity and the fact that they are pre-elaborated material make them more suitable for expository instructional designs than to other approaches, such as constructive or interactive instruction (Ormrod, 2017). In fact, this has been a characteristic established *a priori* for the research object of this thesis.

2.2.3 Instructional videos are multimodal texts

The vast majority of instructional videos consist primarily of spoken or written language, sometimes supported by other complementary representations. Thus, linguistic analyses would contribute to describing the structures of instructional videos, and semiotic analyses would contribute to understanding how these structures are used to construct meaning. This semiotic/linguistic way was already explored in the past (e.g. Metz, 1966). More recent approaches to linguistics and semiotics allow for richer and more powerful insights into instructional video. This is where Discourse Analysis and related disciplines can help.

Figure 2-1. Learning object types in the multimodality/interactivity space



Discourse and texts

What is discourse analysis? According to the classic Crystal's definition, *discourse* is "a continuous stretch of (especially spoken) language larger than a sentence, often constituting a coherent unit such a sermon, argument, joke, or narrative" (Crystal, 1992, p. 25). The object of study of Discourse Analysis are *texts*. Here 'text' means any verbal language utterance, in any modality: written, spoken or signed. What discourse analysis does is to analyze language in its social and cultural context, discovering usage patterns beyond classic grammar. There are several approaches to discourse analysis, from textual analysis (a linguistic approach) to 'discourse as social construction of reality' (an ethnographic approach). For a general overview of Discourse Analysis approaches and methods, see (Paltridge, 2012).

Multimodality

As with the perspective of multimedia learning objects, *multimodality* appears as a shaping dimension of instructional videos when they are considered as texts. Instructional videos as texts are inherently multimodal: written text is combined with spoken language and other modes, such as gestures and pictures.

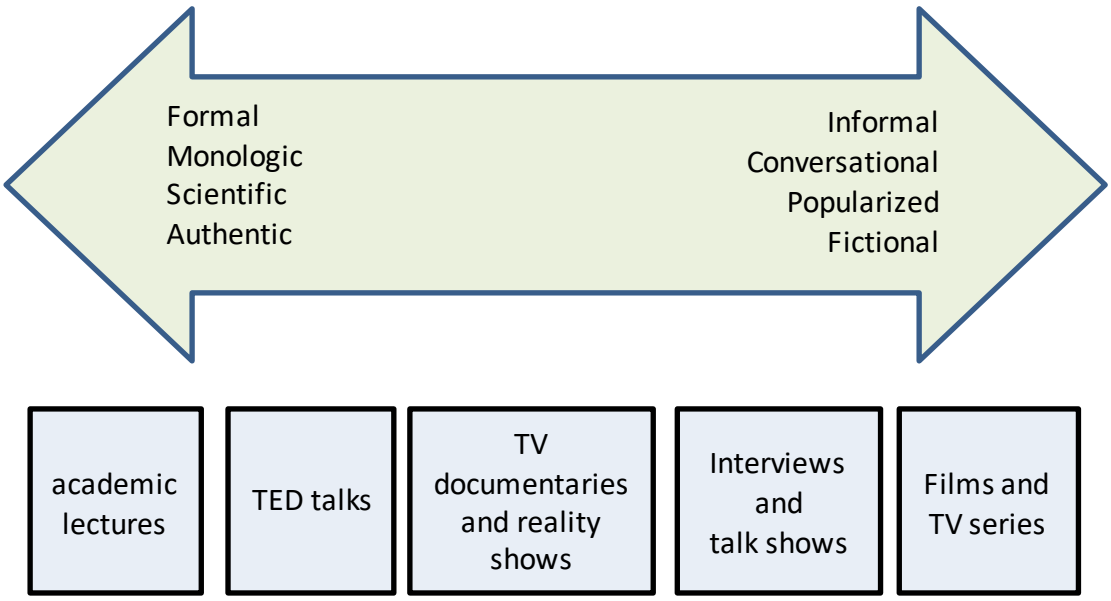
Beyond instructional videos, multimodality has been considered a requirement for the study of the academic discourse in general (Myers, 2003), given that many teaching and learning activities use multiple modes of representation. For example, Roth (2001) pointed out the importance of deictic gestures in the classroom lectures. And of course, the multimedia learning principle encourages the simultaneous use of auditory and visual modalities in learning materials.

Analyzing academic discourse

Academic genres —forms of communication in academic contexts— started to be studied with modern approaches thanks to authors as Swales and his analysis of research articles (Swales, 1981). Swales showed how certain textual patterns were used recurrently, orchestrated in characteristic sequences of *moves*. Those patterns or 'communicative events' were linked to functional goals. Following Swales, a cohort of researchers have been studying various formats of academic discourse through the lens of modern Discourse Analysis. In particular, academic lectures have been profusely studied (see Yaakob, 2013 for a literature review on this matter). A landmark in the understanding of the rhetoric structure of academic lectures was made by Lynne Young (1994). Young found that the macro-structure of a lecture is composed of several interleaved discourse *strands* which in turn are composed by *phases* such as discourse structuring, conclusion, evaluation, interaction, theory/content and examples. Young's phasal model has been validated and extended by other corpus-based studies (e.g. Deroey & Taverniers, 2011).

In recent years, related forms of spoken academic discourse have begun to be analyzed, for instance, academic oral presentations (Barrett & Liu, 2016).

Figure 2-2. The spectrum of video genres discourse
(Crawford Camiciottoli & Bonsignori, 2015)



Multimodality is gaining interest, e.g. the combined use of slides and talk in lectures (Degano, 2012) and the joint expression or speech, gestures and printed text in conference paper presentations (Morell, 2015). Nevertheless, researchers have paid little attention to instructional videos as a distinct type of academic communication. Only recently some research has been published regarding the discourse analysis of MOOC video lectures (Atapattu & Falkner, 2017; Bernad-Mechó, 2015).

In absence of specific discursive analyses of studio-recorded instructional videos, it is reasonable to presume that many discursive features of classroom lectures are shared by video lectures and tutorials. My personal review of literature in academic lectures reveals some outstanding features: extensive use of meta-discourse markers, use of interactive features even in monologic lectures, conversational language, significant participation of nonverbal language and extensive use or humor[7].

Discourse of 'truth and trust'

In Wetzel, Radtke and Stern's (1993) review of educational video, three psycho-social properties of video for instruction are found throughout the text: the *verisimilitude* of the setting, the *credibility* of the narrators and the discourse, and the *realism* of the representations. These properties agree with the thoughts of documentary theorists, such as Nichols, who wrote that "the documentary tradition relies heavily on being able to convey to us the impression of *authenticity*" (Nichols, 2001, p. xiii), or Plantinga's formal voice that conveys an "epistemic authority" that grants credibility to the audience (Plantinga, 1997, p. 107).

[7] Chapter 6 provides a detailed account of these findings on spoken academic discourse.

Also, when modern academic lectures (and instructional videos) are compared to other audiovisuals, such as TED talks and documentaries, or to less related television formats, such as talk shows and fiction series, we can identify some distinguishing discursive features. Figure 2-2 is taken from Crawford Camiciottoli and Bonsignori (2015) and shows how different formats are placed in a discourse spectrum with dimensions such as formality, number of participants and realism. Instructional videos lay in one extreme side of the spectrum, being particularly formal, monologic, scientific and authentic.

If we have to define this aggregate of properties with two words, they would be **'truth and trust'**. What is told in the instructional video is *true*, and the audience is impelled to *trust* in the facts and in the speaker.

2.3 Theoretical foundations for instructional video characterization

In the previous section I have described how instructional videos can be characterized according to three perspectives: as films, as learning objects and as texts (linguistic-semiotic artefacts). As an operational conclusion, this literature review has allowed me to identify two major theoretical sources which are capable of give support to the research goals of this thesis: first, cognitive theories of learning; and second, modern discourse analysis theories, founded on Semiotics and Functional Linguistics.

Cognitive theories of multimedia learning have been successfully used as a theoretical foundation within the Technology Enhanced Learning (TEL) research community. These theories are backed by strong experimental support. As far as this thesis is concerned, these theories help to identify many interesting features of the instructional videos that have a clear effect on learning. Thus, these theories can provide a raw collection of relevant characteristics to be classified. On the other hand, Discourse Analysis can provide a high-level picture about how linguistic and semiotic resources work to make meaning in an instructional video. Thus, this field can provide an insight to the architecture for the classification scheme. In addition, Discourse Analysis would fill the gap in linguistic and semiotic features that are usually missing in multimedia learning research.

Therefore, both theories combined can result in a powerful tool to achieve my intended *anatomy of instructional videos*.

In the following, I will make a brief exposition of the components of both research fields that can contribute substantially to my research.

2.3.1 The Cognitivist perspective: multimedia learning theories

The multimedia learning hypothesis states that "people can learn more deeply from words and pictures than from words alone" (Mayer, 2014d, p. 1). This idea was suggested several times in the past, for instance by Carpenter (1953) and Schmidt

(1972). The difference of Mayer's theory is that it is based on a sound theoretical framework from cognitive sciences. That is, modern multimedia theories of learning are rooted on these two hypotheses about the architecture of the human learning system:

- Dual Coding Theory (two separate processing channels for pictorial and verbal information).
- Cognitive Load Theory (the processing channels have a limited capacity of working memory to process the inflow of information).

These cognitive theories explain several multimedia learning effects that have been confirmed by experiments (split-attention effect, redundance effect...). It has been hypothesized that there are brain modules that constitute a system for information processing and learning, in which there exist separate channels for auditory and visual processing, and perhaps for language processing. Despite the experimental support for the theory, for the moment there is little evidence at the neurological or physiological level.

Dual Coding and Cognitive Load Theory

The "dual-channel hypothesis" was first formulated by Allan Paivio (Paivio, 1990), and states that the brain has separate channels for processing visual and verbal information. The verbal information may have a written or spoken representation.

The Cognitive Load Theory was initially developed by John Sweller (Sweller, 1988). Sweller considers that learning consists in the acquisition of mental *schemas* in the learner. The process of schema acquisition requires that working memory resources process the incoming information, which demands an amount of *cognitive load*. Working memory has a limited capacity, therefore it may overload if the learning material is too complex or poorly organized.

The Cognitive Load Theory distinguish three components of cognitive load: *intrinsic*, *extraneous* and *germane* cognitive load (Sweller, van Merrienboer, & Paas, 1998). The intrinsic load is that inherent to the learning subject (in terms of learner's current capabilities). The extraneous load comes from processing elements that do not contribute to the learning goal. The germane load is an extra cognitive effort to integrate diverse information sources in order to build mental schemas. According to the theory, these three components are additive, and the acquisition of new schemas can only succeed if the total demand of cognitive load does not exceed the capacity of the learner's working memory resources.

From an instructional perspective, the aim is to reduce extraneous load and contribute to germane processing, while keeping the global amount of cognitive load at a level achievable by the learner.

Cognitive Theory of Multimedia Learning

An implication of the above theories (Dual Channel hypothesis and Cognitive Load Theory) is that learning may be more effective when the information is shown

simultaneously in verbal and pictorial representations. Another implication is that if cognitive load is adequately distributed in both verbal and pictorial channels, the risk of overloading one channel is diminished and therefore learning may be better than using a single mode of representation (verbal or pictorial). These hypotheses conform the basis for the now called Cognitive Theory of Multimedia Learning (CTML): under certain circumstances, people learn better from a combination of words and pictures than from words alone. "understanding occurs when learners are able to build meaningful connections between pictorial and verbal representations" (Mayer, 2014d).

CTML has been developed by Richard Mayer and collaborators in the last quarter of 20th century. The most recent 'canonical' text about CTML is the *Cambridge Handbook of Multimedia Learning* (Mayer, 2014a).

CTML is based on these three assumptions:

- There are two separate channels for processing visual and auditory information (dual channel hypothesis).
- A channel can process a limited amount of information at a time (cognitive load theory).
- Learning consists of *active processing*.

As regards active processing, CTML theory assumes that, rather than being passive receptors of information, humans actively engage in cognitive processing. The active learning processes involve selecting what material seems to be relevant for transfer to working memory; organizing the selected information and eventually integrating this organized information into a knowledge structure (as in Sweller's schemas). All instructional material that favor these activities would enhance the learning outcomes: signaling cues, relations to prior knowledge, avoiding irrelevant material, etc.

Multimedia learning principles

CTML comprises several *instructional principles*, as the multimedia principle, the coherence principle, or the segmenting principle, to name a few. Each principle expresses a hypothesis of instructional design that fosters learning outcomes. For example, the redundancy principle says that people learn better when the same information is not presented in more than one format. Each principle is supported by empirical evidence.

Table 2-2 lists the most relevant Mayer's multimedia instructional principles, as they are specified in the 2014 Second Edition of the *Handbook of Multimedia Learning*. Earlier formulations included some other principles, as the guidance, interactivity and reflection principles (Moreno, 2005), that have been dropped from the later compilations of the theory.

The multimedia instructional principles can be applied straightforwardly to the production of instructional material. For instance, the book by Clark & Mayer (2008) shows concrete recommendations on how to design multimedia learning

resources, informed by CTML principles. Table 2-2 lists the guidance recommendations from that book, each one linked to its corresponding instructional principle.

Since its inception, CTML has been constantly adding principles and learning dimensions, as research has been producing new findings that cannot be fully explained with the classic set of principles. For instance, in consonance to alternative models to Paivio's dual coding, such as Bucci's Multiple Coding Theory (Bucci, 1997), motivational, affective and emotional factors in learning were eventually added to Cognitive Multimedia Learning Theory. Mayer (2014b) admitted that motivational elements in instruction contribute to germane processing and germane load.

Table 2-2. Principles of multimedia learning (Mayer, 2014)
As stated in the *Cambridge Handbook of Multimedia Learning*, 2nd Edition.

instructional principle	people learn better when…[1]	guidance[2]
multimedia	words and pictures are used rather than words alone	use words and graphics rather than words alone
modality	pictures and speech are used rather than pictures and printed text	present words as audio narration, rather than on-screen text
redundancy	the same information is not presented in more than one format	explain visuals with words in audio or text, not both
signaling	cues are added that highlight the key information and its organization	*(not discussed in the book)*
split-attention	multiple sources of information are physically and temporally integrated (spatial contiguity; temporal contiguity)	align words to corresponding graphics (place printed words near corresponding graphics; synchronize spoken words with corresponding graphics)
coherence	extraneous material is avoided	adding interesting material can hurt learning avoid lessons with extraneous audio, graphics, or words
segmenting	the message is presented in learner-paced segments rather than as a continuous unit	break a continuous lesson into parts
pre-training	people know the names and characteristics of the main concepts to be presented	ensure that learners know in advance the names and characteristics of key concepts
personalization	words are presented in conversational rather than monologue style	use conversational rather than formal style
embodiment	on-screen agents display humanlike gestures and movements	use effective on-screen coaches
voice	words are spoken in a standard-accented human voice	*(not discussed in the book)*
image	(on-screen speaker's image may not help to better learning)	*(not discussed in the book)*

Bold typeface is applied on principles that were enunciated in earlier formulations of the theory.

[1] According to the *Cambridge Handbook of Multimedia Learning*, 2nd edition (2014).
[2] According to *e-Learning and the Science of Instruction* (2011).

2.3.2 The Semiotic perspective: multimodal discourse analysis

Discourse Analysis is an area of Linguistics which is tightly linked to Semiotics. It is concerned with how people interact through language, and how semiotic resources (speech, text, pictures, body movements) are used to make meaningful communication. Discourse Analysis thrived throughout the second half of 20th century in a wide variety of fields, from Linguistic to Cognitive sciences and Sociology.

The concept of *meaning* in Semiotics has a broad scope. Meaning is not only about the textual topic that is being informed or taught. Meaning is also social, cultural and emotional (show relations of power in the conversation, affect, trust). Both dimensions, textual and social-cultural-emotional, have a strong impact in learning processes. Therefore, the analysis techniques and the research findings made in the communities of Discourse Analysis research will help us to discover patterns, structures and processes that are relevant when designing an effective instructional film or video.

As I wrote above, Discourse Analysis studies *texts*, defined as any language utterance in any modality (written, spoken or signed). Modern discourse analysis deals both with grammatical features of texts and higher-level rhetorical functions. Besides, Semiotics has long time ago realized that meaning is not only constructed upon verbal language, but many kinds of human artefacts convey coded messages and symbols, including films (Eco, 1979).

Genre analysis

Genres are a key concept in contemporary discourse analysis. Swales, a founder of the English for Specific Purposes school, defines a genre as "a class of communicative events, the members of which share some set of communicative purposes" (Swales, 1990, p. 58). The Sydney School characterizes a genre as "a staged, goal-oriented, purposeful activity in which speakers engage as members of a given culture" (James Robert Martin, 1994). Examples of genres are lectures, speeches, poems, jokes, advertisements, shopping lists or even casual conversations. Each genre has its own characterizing features, which can be linguistic, paralinguistic (e.g. print size, gesture) and contextual. Genres are used and recognized by members of a *discourse community* (Swales, 1990).

Genre analysis is being used actively in educational research to investigate how language is used and improve foreign language learning and technical communication (Paltridge, n.d.). From an educational perspective, an advantage of genres awareness is that producers of texts can assume that readers/listeners will already have knowledges and expectation about the presentation of the message, hence can serve to increase the efficiency of communication (Chandler, 1997).

Genres have similarities with *design patterns*, as used in Architecture and Software Engineering (Alexander, Ishikawa, & Silverstein, 1977; Lea, 1994). However, genres

are not necessarily intended and related to a design goal, nor they can be formally specified for reuse, as design patterns are.

Multimodal discourse analysis (MDA)

Traditionally, the main object of study of Discourse Analysis has been human language (spoken or written). Over time, it has expanded to other modalities of human symbolic communication, as photographs and illustrations. One pioneering example is the work led by Theo van Leeuwen (van Leeuwen & Kress, 1995), who applied discourse analysis to printed posters and newspaper pages. Van Leeuwen and Kress demonstrated how the *spatial layout* of text blocks and pictures was orchestrated to build meaning, through relations like contiguity, signaling or placement on certain page areas.

Van Leeuwen and other researchers gave rise to the new field of *Multimodal Discourse Analysis*, which encompasses the analysis of *multimodal texts*. Multimodal artefacts combine speech (written or spoken) with other modes of expression, such as diagrams, drawings, photographs, video and music. This approach has been applied to multimodal printed documents, such as newspapers, printed advertisements and illustrated nonfiction books (Baldry & Thibault, 2006; Paltridge, 2012, Chapter 8; van Leeuwen, 2008). More recently, researchers have focused in multimodal film analysis, for example Bateman (Bateman & Schmidt, 2012) and O'Halloran (O'Halloran, 2009).

The availability of computer technology has hybridized classic text-only genres, such as the educational manual, with graphics, animations, audio streams and interactive software applications. This trend in digital texts has spurred the need for a multimodal approach to discourse analysis. As J. L. Lemke remarked:

> *Genres are not what they used to be* [...] *Many genres of interest are increasingly multimodal, making their meanings through the codeployment of resources from both language and other semiotic systems." (Lemke, 2005).*

One example of this generalization of the multimodal approach that is closely related to this dissertation is the growing interest in the analysis of audiovisual modalities of conference presentations, e.g. (Morell, 2015; Valeiras-Jurado, 2017).

Systemic Functional approach to MDA (SF-MDA)

The Systemic Functional Linguistics theory (SFL) provides a foundational framework for the analysis of multimodal documents. Complex audiovisual artefacts, such as video lectures, can be described in terms of objects and patterns of interactions among them. Those interactions may be spatial-temporal relations, semantic or syntactic relations, phases and transitions, among others (O'Halloran, 2009; van Leeuwen, 2008).

The study of meaning. According to Halliday, semiotics goes beyond the Saussurian "general study of signs", to a more ambitious "study of meaning in its most

general sense" (Michael A. K. Halliday & Hasan, 1985). Hallidayan semiotics embraces the study of other communication forms apart from traditional linguistic modes: "[there are] other ways of meaning, in any culture, which are outside the realm of language". That vision enabled several studies on nonverbal forms of communication, such as photographs, comics, illustrated textbooks, and films.

The SFL theory has been applied beyond verbal language. As a landmark study, Kress and van Leeuwen (2006) explained how SFL can be expanded to the analysis of visual designs. The Systemic Functional approach to Multimodal Discourse Analysis is concerned with "the way people use semiotic resources to produce communicative artefacts and events and to interpret them [...] in the context of specific social situations and practices." (O'Halloran, 2009).

Some features of Systemic Functional Linguistics should that will be relevant in this study are *stratification*, *constituency* and *metafunctions*. I will write a short introduction to these concepts.

Stratification and constituency. Semiotic resources are organized in patterns that follow some properties. The organization takes place in space, time, grammar, semantics, or whatever physical or abstract dimension. The organization of semiotic resources usually is organized in hierarchic *strata*. The components in a given stratum are in turn constituents of higher strata. Thus, meaning is built up as a hierarchical series of functional elements, as in phoneme, word, phrase, clause, clause complex genres (Lim-Fei & O'Halloran, 2014).

Metafunctions and register. Halliday (M. A. K. Halliday, 1978) describes language as a 'social semiotic', that is, a system of signs that encodes social meaning. Halliday views language as *functional*: it is an instrument to fulfill goals. Language functions in a text can be grouped in three broad categories or *metafunctions*: **ideational** (what is the text about), **interpersonal** (who are the participants and their relationships), and **textual** (how elements of the text relate to each other) (Michael A. K. Halliday & Matthiessen, 2014, pp. 30–31). A speaker makes choices in the language systems. Choices are constrained by social identities and situations. A particular configuration of language choices is called a 'register'. In oral speech, Halliday defines three main *register variables* that build ideational meaning: **field**, **tenor** and **mode** (2014, pp. 33–34). The utilization of field, tenor and mode influences how the speakers make meaning in their language productions.

Semiotic resources are involved to develop one or more metafunctions. The analysis of multimodal documents involves to discover how people make semiotic choices to fulfill metafunctional goals (Lim-Fei & O'Halloran, 2014). For example, in an instructional video we can identify how a lecture statement constructs trust in the audience or how the instructor queries the viewers to ensure they remember the key ideas of the lecture.

2.4 Literature review: selected works

This section contains the result of a first survey of academic research that contains references that will serve as a first seed to the development of the instructional video structural taxonomy. I have collected works that refer to the learning efficiency of instructional videos, as well as works that try to describe the internal structure of films and multimedia artefacts. I will take these works as the starting point in developing the taxonomy, which I will describe in Chapter 4.

2.4.1 Wetzel, Radtke & Stern's review of educational video

Weztel, Radtke and Stern performed a research funded by the United States Navy consisting of an extensive review of the research literature of "dynamic video media" in instruction. The result was delivered in 1993 as a 220-page report containing an considerable collection of findings on the effectiveness of several features and production techniques in instructional films and videos (Wetzel et al., 1993). The report was later published as a book (Wetzel et al., 1994). For my dissertation, I have used the original U.S. Navy report.

The findings collected by Wetzel et al are organized around two classes, according to the source from which they were retrieved: *professional tradecraft rules* and *research findings* from scientific literature. Professional tradecraft rules are common conventions accepted by the film production industry and are exposed in the first part of Wetzel and collaborators' study. For this first part, the authors used as a guiding text the second edition of Herbert Zettl's *Sight, sound, motion: Applied media aesthetics* (Zettl, 1991) and other works by Zettl. The second part of the study is the result of an extensive review of research on educational films and educational television. It describes a list of audiovisual features that have been investigated as regards its potential influence in learning. Professional tradecraft rules and research findings are in turn organized in non-systematic categories. I have summarized Wetzel et al's findings in Table 2-3.

As I have mentioned in Chapter 1, Weztel, Radtke and Stern's work was probably the last comprehensive review of the internal structure of instructional videos from the age of Educational Television. Unfortunately, this work was prior to the development of Cognitive Theory of Multimedia Theory and therefore the authors could not take advantage of a theoretical framework on which to weave their results. There are no posterior works with the amplitude of this work.

2.4.2 Guidelines for instructional video design

Practitioners have developed guidelines to instructional video production that describe design features that are related to better learning outcomes. Some features are tightly related to multimedia learning principles, while others suggest rhetorical patterns considered to be effective. I have surveyed some of the post-Wetzel academic research on video production, resulting in some remarkable references that

Table 2-3. Characteristics of educational films (Wetzel, Radtke, & Stern, 1993)

professional tradecraft rules	
camera technique	shot length: long, medium, close-up, extreme close-up
	camera position and movement: direction, angle, distance / panning, tilting, dollying, crane, trucking
	camera angle
	zoom and focus lens effects: zoom, overlapping planes, size and distance, linear perspective, depth of field
shot composition	picture complexity, balance and proportion, movement, framing, lines, lighting, color
editing	cutting, shot order, continuity
special effects	animation, sound, music, text, captioning
research findings	
presentation format	dramatization; expository format
presentation pace and length	presentation pacing; program length
strategies, techniques and devices	(common instructional strategies and techniques)
	encouraging student participation, inserting questions
attention-getting devices	general viewer arousal, setting verisimilitude, rapid cutting, sound effects, music, humor, color
cuing	camera effects: zooming, dollying, shot length, panning, camera angle
	cutting and editing; composition
verbal material	captions, commentary/narration

I summarize in the following paragraphs. I will elaborate more on these works later in Chapter 5, when I discuss the rhetorical structure of instructional videos.

Jack Koumi developed one of the most comprehensive frameworks for educational video design, based on his long-term experience in the BBC Open University. A first public version of the framework was published in 1991 (Koumi, 1991), which was later expanded into a book (Koumi, 2006). A final refinement was published in 2015 (Koumi, 2015). Koumi's framework consists of 'pedagogic design principles' that comprise a catalogue of 25 discursive structures. Those principles are grouped around eight categories: *hook, signpost, facilitate cognitive engagement, enable construction*

of knowledge, sensitize, elucidate, reinforce, and *conclude/consolidate*. This ordering roughly corresponds to the actual sequence of these categories appearing in the video.

Loch and Mcloughlin (2011) sketched instructional design guidelines for self-regulated learning by using screencasts. These guidelines comprise three areas: provide an overview, activate prior knowledge, ask students to set learning goals, present questions and tasks, encourage students to reflect, ask students to self-assess their performance.

Swarts and Morain investigated the features that correlated with quality of online video tutorials (Morain & Swarts, 2012; Swarts, 2012). After a content analysis of 46 YouTube tutorials, they developed a rubric for assessing the quality of online video tutorials. This rubric comprises three design perspectives: physical design, cognitive design and affective design. The rubric applies some multimedia learning principles, as well as some expository patterns similar to Koumi's pedagogic design principles. A similar work was made by van der Meij and van der Meij (2013), who proposed eight guidelines for designing procedural video tutorials, based on prior research.

Kay (2014) developed a framework for creating worked-example video tutorials. This framework contains sixteen design characteristics, organized around four design goals: establishing the context, creating effective explanations, minimizing cognitive load and student engagement. As in the previous references, Kay suggests some rhetoric organization patterns as well as physical representation advices, closely related to multimedia learning principles.

2.4.3 Characterization of multimedia learning objects

Ploetzner and Lowe: expository animations

An interesting work because of its close relationship with instructional videos is Ploetzner and Lowe's systematic characterization of expository animations (Ploetzner & Lowe, 2012). Expository animations share many similarities with instructional videos: they have an instructional purpose, they have limited interactivity, and they rely strongly in audiovisual information—in fact, some instructional videos *are* expository animations.

The authors performed a systematic search in scientific literature by using a method similar to the 'snowball' technique. They started with a set of initial dimensions from secondary research: overviews, reviews and meta-analyses. Then, a couple of iterations were applied to search for those initial dimensions in literature databases. Ploetzner and Lowe eventually established four dimensions for organizing expository animation features: *presentation*, *user control*, *scaffolding* and *configuration*. Table 2-4 shows a slightly abridged depiction of Ploetzner and Lowe's characterization.

Ploetzner and Lowe's characterization cannot be immediately generalized to all kinds of instructional videos, since there are several characteristics that are not considered

Table 2-4. Ploetzner and Lowe (2012) characterization of expository animations

dimension	subdimension	values
presentation	1.1. representation	
	1.1.1. visual	iconic pictures, analytic pictures, symbols, formal notations, labels, text
	1.1.2. auditory	sound, speech, narration
	1.2. abstractions	iconic, abstract
	1.3. explanatory focus	behavior, structure, function
	1.4. viewer perspective	single, multiple
	1.5. spatio-temporal arrangement	
	1.5.1. spatial resolution	constant, variable
	1.5.2. spatial structure	dimensionality: two, three organization: flat, hierarchical
	1.5.3. temporal resolution	discrete, continuous with pauses, continuous with cuts, continuous
	1.5.4. temporal structure	representation: persistent, implicit, singular; chronology: linear, cyclic concurrency: sequential, simultaneous; organization: flat, hierarchical
	1.6. duration	presentation time
user control	2.1. time line	
	2.1.1. temporal navigation	(re-)start, stop, pause…
	2.1.2. temporal scaling	change speed
	2.2. presentation	
	2.2.1. appearance	magnify, change perspective
	2.2.2. information content	zoom, show/hide entities or layers, narration on/off
scaffolding	3.1. visual	visual cues; written prompts
	3.2. auditory	spoken prompts
configuration	4.1. execution	single, repeated
	4.2. setting	stand-alone, embedded
	4.2.1. (embedded) surroundings	(11 possible items)
	4.2.2. (embedded) concurrency	sequential, simultaneous

in the study, such as the pedagogical agent, voice attributes or discourse structure, that are essential in video lectures or tutorials.

What is remarkable about Ploetzner and Howe's proposal is the method to obtain the characterization: a systematic review of the scientific literature. By using this approach, each feature in the characterization corresponds to items that have been received attention by the scientific community.

2.4.4 Multimodal genre analysis

Some researchers on multimodality have defined conceptual frameworks that could be useful to understand the semiotic structure of multimedia learning objects. I have selected two relevant works: Vorvilas et al's framework for multimedia genres (Vorvilas, Karalis, & Ravanis, 2011) and John Bateman's 'GeM' framework for multimodal texts (Bateman, 2008).

Vorvilas et al: genre analysis for learning object design

One of the scarce crossovers between genre analysis and multimedia learning is the work by Vorvilas et al. (2011). The authors developed a conceptual framework for interpreting and designing content for learning objects, from a Genre Theory perspective. The model assumes that a learning object is a composition of elementary *content objects* that aggregate through *generic* patterns, in order to achieve communication goals, as in the Swalesian formulation of genres. The authors adopt Martin's microgenre concept (James Robert Martin, 1994) and apply it to multimedia learning objects: content objects are "digital microgenres which serve particular communicative goals inside a learning object". The resulting model (see Table 2-5) is composed of three layers of complexity: items, content objects and learning objects. Content objects in Level 2 relate to each other through rhetorical relationships, as defined in the Rhetorical Structure Theory (RST) (Mann & Thompson, 1988).

Table 2-5. Layered framework for multimedia genres (Vorvilas et al., 2011)

level 3	learning objects	digital macrogenres	tutorials, simulations, drills and practices, lessons...
level 2	content objects	digital microgenres	objectives, assessments, reports, explanations, summaries, representations...
level 1	items	communicative acts	buttons, symbols, captions, sounds, boxes, texts, lines...

Bateman's multimodal genres and the GeM framework

John A. Bateman, in his book *Multimodality and Genre: a Foundation for the Systematic Analysis of Multimodal Documents* (Bateman, 2008), introduces a comprehensive model for analyzing multimodal text. Bateman takes former descriptive frameworks for written text and extends them to properly describe multimodal printed documents (i.e. documents composed of text and graphics), taking into account physical features such as page layout, typography and color.

Bateman develops a descriptive framework (*GeM: Genre and Multimodality*) to facilitate the analysis of multimodal documents. The GeM framework considers that a multimodal document is constructed on a physical or virtual *canvas* on which several constraints operate, in order to render a *virtual artefact*. Those constraints may originate on physical properties of the canvas, on the production process, or on the intended use of the artefact. In addition, social conventions select arbitrary configurations of those artefacts, which are *genres* of multimodal documents.

According to the GeM framework, a multimodal document can be analyzed by using six different descriptive layers: *content, genre, rhetorical, linguistic, layout* and *navigation*. Each layer puts its lens through one particular meaning-making dimension, and through one particular mode of relationships between the constituent components of the document. To characterize the types of elements in the document, Bateman proposes a simplified layered structure, as shown in Table 2-6. Each layer in Table 2-6 defines its own basic set of units, and relations and structures defined on these units. The units in GeM base are the basic constituents: every layout, rhetorical, navigational or generic unit consists of a collection of GeM base units. All this means that the GeM model shows the stratification and constituency properties of systemic functional languages (Lim-Fei & O'Halloran, 2014).

The GeM framework states that multimodal documents are produced on a *canvas* (a canvas is more or less a synonymous of *medium*). The canvas holds a set of constraints grouped in three types: constraints of the physical medium (in printed text, it may be the minimum size of typography, or a limited color palette, etc.), restrictions from production technology, and restrictions from intended uses. All these restrictions conform a type of *virtual artefact*.

Though Bateman's model was originally intended for multimodal printed texts, it can be applied to other static media, such as web pages. Bateman continued his work on multimodal text towards narrative film analysis (Bateman, 2013; Bateman & Schmidt, 2012). Dynamic images in films can be supported in GeM by extending the concept of 'layout' so that is accounts for both spatial and temporal segmentation. Music and sound can be regarded as basic semiotic resources in the lowest physical layer (nonetheless, Bateman's work in applying GeM to films is not as deep as it was to printed texts). The application to narrative films shows that GeM framework is solid enough to describe many modes of audiovisual communication, such as

Table 2-6. Bateman's GeM layers and elements (Bateman, 2008)

Layer	Description
GeM base	The basic elements physically present on a page
Layout base	The layout properties and structure: nature, appearance and position of communicative elements on the page, and their hierarchical interrelationships
Rhetorical base	The rhetorical relationships between content elements: how the content is 'argued' and structured rhetorically
Navigation base	The elements that contribute explicitly to direct or assist the reader's consumption of the document, supporting 'movement' around the document in various ways
Genre base	A representation of the grouping of elements from other layers into generically recognizable configurations distinctive for particular genres or document types

instructional videos. In fact, I will use GeM as the foundation for the architecture of my taxonomy, as I will explain in Chapter 4.

2.4.5 Multimodal transcription of films

The community of Discourse Analysis has developed frameworks for the transcription of multimodal artefacts such as films. These frameworks include inventories of structural items that may occur during the film: e.g. shot transitions, camera movements, or character interactions. These sets of characteristics lack experimental validation of their actual significance in instructional effectiveness, but at least they indicate that they are relevant to a scientific community interested in understanding meaning-making.

Baldry and Thibault (2006)

Baldry and Thibault (Baldry & Thibault, 2006) developed a system for the analysis of multimodal discourse, grounded in Systemic Functional Linguistics. Their approach included methods for annotating films, which is summarized in Table 2-7.

Since Baldry and Thibault's proposals, other authors such as O'Halloran, Crawford Camiciottoli and their collaborators (Crawford Camiciottoli, 2007; Crawford Camiciottoli & Bonsignori, 2015; Lim Fei, O'Halloran, Tan, & E, 2015; O'Halloran, 2009) have been analyzing video-recorded academic lectures using multimodal transcriptions, where visual images extracted from the video are paired with the corresponding verbal text, together with other nonverbal acts such as lecturer's hand and body gestures.

Table 2-7. A schema for film transcription (Bauldry & Thibault, 2006)

Domain	Description items
Segmentation	Phase, subphase, transition
Visual frame	Frame, shot; given/new; sequencing
Visual information	Camera position, perspective, distance, visual collocation, visual salience, color, coding orientation, visual focus, gaze
Kinesic action	Movement: spatiotemporal arrangements of the agents
Soundtrack	Sound event/act, rhythm, loudness, tempo, continuity, pause Auditory voices: sequentiality, overlap, turntaking, vocal register
Metafunctional interpretation	SFL metafunctions: experiential, interpersonal, textual, logical

Multimodal Analysis Video tool

O'Halloran has applied the systemic functional linguistics theory to multimodal discourse analysis (Lim-Fei & O'Halloran, 2014). Her work includes the development of film annotation techniques. O'Halloran's team has implemented their framework into a software application (Lim Fei et al., 2015). This application is currently offered as a commercial product under the name of *Multimodal Analysis Video*[8]. The application provides a flexible schema for annotating the discourse events in all kinds of audiovisuals (*visual texts*): TV commercials, narrative films and recorded lectures, among others. The software comes with a predefined base of categories and labels, inspired by the theoretical developments of the O'Halloran's team. Users can add custom categories and label to build their own taxonomies.

As part of my review, I have examined the *Multimodal Analysis Video* computer application to obtain its internal scheme for video analysis. The result is shown in Table 2-8 and Table 2-9. The first chart is the default inventory of annotation items that are available for users. Language and visual items are more numerous, so these are listed in a separate chart in Table 2-9. As it can be seen, the scheme of this tool is an evolution of that of Baldry and Thibault (2006).

Table 2-8. Taxonomy of annotations in *Multimodal Analysis Video* tool

1) Composition
   a) Phase: a set of coherent meaning selections that are associated with specific theme
   b) Sequence: the camera moves with specific sub-topic across time-spaces
   c) Scene: a set of shots in the same time-space
   d) Shot
2) Connotation
   a) Myths: dominant ideology that is culture-specific
   b) Values: idealized social construct (e.g. happiness)
   c) Ideas: meaning associated with a value (e.g. curiosity, technology)
   d) Gender
3) Language
   a) General: Experiential / interpersonal / textual
   b) Spoken: intonation / pitch / pace / timbre
   c) Written: text organization / typography
4) Visual elements
   a) Experiential meaning
   b) Interpersonal meaning
   c) Textual meaning
5) Auditory elements
   a) General: volume, pitch movement, sound setting (mono, stereo), pitch, sound prominence (background, foreground)
6) Inter-relations
   a) Similarity: co-contextualization
   b) Difference: re-contextualization

[8] Available in http://multimodal-analysis.com/

Table 2-9. Predefined language and visual items in *Multimodal Analysis Video* tool

Metafunction	Language items	Visual items
Experiential meaning	Processes: material, mental, relational, verbal, behavioral, existential Participants: actor, target, senser, concept, sayer, behaver, existent Circumstances: extent, location, manner, cause, condition, accompaniment, role, matter, angle	Processes: action, reaction, interaction, state, conceptual Participants Participant roles: actor, reactor, target, concept
Interpersonal meaning	(role of each actor) speech function: statement, question, offer, command	Gaze: direct, indirect Social distance: long shot, medium shot, close shot Zoom: zoom in/out Camera movement: stationary, pan, tilt, pedestal, dolly Horizontal viewing perspective Vertical viewing perspective Visual prominence: sharpness of focus, colour contrast, lighting, foreground, background
Textual meaning	Topic / theme	Comparative relations: similarity, contrast Spatial relations: overview, detail Temporal relations: simultaneity, sequentiality Shot transitions: cut, dissolve, fade-in/out

2.4.6 Classifications of video presentation formats

Extensive research effort has been taken to build general classifications of online educational videos, resulting in several proposals. Table 2-10 and Table 2-11 show a summary of some selected classification schemas.

Educational video communication styles

An early work by Goodyear & Steeples, C. (1998) identified six main communication styles used in videoclips shared within communities of practice. The JISC organization (Thornhill et al., 2002) adapted the Goodyear & Steeples' classification

to the then-emerging field of streaming video in education, suggesting seven frequent usage patterns: *Talking Head*, *Events*, *Instructional*, *Simulation*, *Think Aloud*, *Fly on the Wall*, and *Real Life*.

Schwartz and Hartman (2007) developed a conceptual model that classifies educational video around four classes of learner outcomes: *seeing*, *engaging*, *doing* and *saying*. For each outcome, the authors established four or five video genres, as shown in Table 2-10.

Kay's (2012) literature review of research on video podcasts includes a classification that uses four dimensions: *purpose*, *segmentation*, *pedagogy,* and *academic focus*. Among the video types, Kay identifies *Lecture-based*, *Enhanced* (slides with added voiceover), *Supplementary* (administrative support, real-world demonstrations, summaries), and *Worked Examples* video podcasts.

The EU-funded REC:all project defined a model for lecture capture technologies (Moes, 2012), framed in Bloom's Taxonomy. Their model involves six types of lectures, such as the *Knowledge Clip*, a genre that REC:all project investigated in depth.

Winslett (2014) made a literature review of online video in higher education. As a result, he drafted various typologies for online video, according to expected learning outcomes, educational topics and video production formats. Winslett's typology of video production is shown in Table 2-11.

Video styles in MOOCs

Researchers have tried to characterize and classify the video styles that are prominent in MOOCs. There are two recent works that provide remarkable results on the classification of MOOC videos. First, the study by Guo, Kim and Rubin (2014) measured the influence of video production style over student engagement in MOOCs. The authors labeled videos using six production styles: *Slides*, *Code* (screencast), *Khan-style*, *Classroom*, *Studio* and *Office Desk*. The study was limited to a small set of courses of Scientific and Technology disciplines in the edX platform. In a second study, Hansch et al (2015) assessed current MOOCs and identified a set of 18 video style typologies (see Table 2-11), based on their qualitative reviews and interviews with MOOC designers and instructors.

Discussion of classification schemas

At present, there is no standardized taxonomy of educational video styles. In addition, the terminology is still maturing, as new video techniques emerge, and neologisms are coined. Some classifications tend to be organized around *functional features*, such as teaching purposes or communication patterns (Goodyear & Steeples, 1998; Kay, 2012; Schwartz & Hartman, 2007; Thornhill et al., 2002; Winslett, 2014), while other classifications tend to be more aware of *representational formats*, evidenced by terms as 'screencast', 'webinar', 'Khan-style' and the like (Guo et al., 2014; Hansch et al., 2015; Moes, 2012). Finally, the two MOOC classification schemas share a concern of researchers on the scene background or scenario setting,

which drives the building of the typologies. Half of the Guo, Kim and Rubin's categories are actually scenario settings (classroom, studio, office desk).

Table 2-10. Summary of classifications of instructional videos

Reference	Video typologies
Goodyear & Steeples (1998) *Subject: video clips* *Context: sharing of working practice*	Fly on the Wall Think Aloud Action with Commentary Talking Head Prepared Script Professionally Acted
Thornhill, Asensio & Young (2002) *Subject: streaming videos* *Context: best practices in education*	Talking Head Events Instructional Simulation Think Aloud Fly on the Wall Real Life
Schwartz & Kartman (2007) *Subject: designed videos* *Context: instructional design*	- *Seeing:* Tour, Portrayal, Point of View, Simulation, Highlighting - *Engaging:* Ad, Trailer, Trigger, Narrative, Anchor - *Doing:* Modeling, Identification, Demonstration, Step-by-step - *Saying:* Association, Chronicle, Analogy, Commentary, Expository
Kay (2012) *Subject: video podcasts* *Context: review of research*	- *Purpose:* Lecture-based, Enhanced, Supplementary, Worked examples - *Segmentation:* Non-segmented, Segmented - *Pedagogy:* Receptive viewing of podcasts, Problem-solving podcasts, Learner creation of podcasts - *Academic focus:* Practical, Conceptual
S. Moes et al (2012) *Subject: video lectures* *Context: best practices in education*	Weblecture Slidecast Knowledge Clip Screencast Tutorial Pencast Webinar Virtual Classroom

Table 2-11. Summary of classifications of instructional videos (continued)

Reference	Video typologies
Winslett (2014) *Subject: educational videos* *Context: review of literature*	Fly on the wall Mashing up Presenting to the camera Dramatic works Interviews, testimonials and vox pops Producing video games Recording and/or transmitting a teaching event Simulating/modelling/representing/capturing hard to see processes and contexts Video diaries Video enabled communication and collaboration
Guo, Kim & Rubin (2014) *Subject: MOOC videos* *Context: video usage patterns*	Slides Code *(screencast)* Khan-style Classroom Studio Office Desk
Hansch et al (2015) *Subject: MOOC videos* *Context: review of MOOC designs*	Talking Head Text-Overlay Actual Paper/Whiteboard Webcam Capture Green Screen Classroom Lecture Live Video Recorded Seminar Interview Conversation Presentation Slides with Voiceover Picture-in-picture Screencast Khan-Style Tablet Capture Udacity-Style Tablet Capture On-Location Animation Demonstration

Chapter 3. A survey of instructional video features in MOOCs

3.1 Chapter overview

I decided to conduct a field study to obtain first-hand evidence about the current usage of instructional video characteristics. This evidence will complement and extend the literature review of Chapter 2. This field study is focused on the instructional videos used in MOOC platforms.

Five global generalist MOOC platforms were selected for this study, which was conducted in two phases: first, a qualitative survey was made to identify frequently used video styles and build a classification scheme. Second, a sample of 115 courses in the selected MOOC platforms was used to account for video feature and style frequency. Various statistical tests were performed to discover associations between course characteristics and video style usage.

As a result, seven video presentation styles have been identified as the most frequent in MOOC courses. They fully describe the video stock of 85% of the sampled courses. A typical course uses two different styles. The study reveals two broad competing approaches to display instructional contents in MOOC videos: *speaker-centric* (a visible person speaks the contents) and *board-centric* (a large rectangular surface displays the contents). The actual usage of each approach is significantly related with the course subject area: Arts and Humanities courses exhibit a preference for speaker-centric styles, while Engineering and "Hard Science" courses favor board-centric videos. Social Sciences and Health courses lay in a neutral position.

3.2 Introduction to study

3.2.1 Background on MOOCs

Massive Open Online Courses (MOOCs) have arisen in recent years as a new model of large-scale online learning service in the context of Higher Education. At present, there are several MOOC platforms that provide thousands of online courses in a wide range of disciplines (see Karsenti, 2013 for a critical review of the MOOC history and characteristics). As of 2015, the media coverage on MOOCs has cooled down compared to the initial hype in 2012, but that does not mean a business decline: MOOC market keeps growing at a fast pace in course offer, enrolment and revenues (Shah, 2015).

MOOCs raised early attention in the scientific community (Liyanagunawardena, Adams, Williams, & Rekha Liyanagunawardena, 2013), including the research in

video-based learning (Giannakos, Jaccheri, & Krogstie, 2014). MOOC providers, teachers and scholars have a growing interest in the research of instructional video production techniques and how they relate to factors such as production cost, learning efficiency and student engagement.

It is important to point out that there are two pedagogical models of MOOCs: the cooperative "cMOOCs" and the more conventional "xMOOCs" (Rodriguez, 2012). cMOOC emphasizes collaboration between learners. A cMOOC is a platform that facilitates knowledge sharing and construction. On the other side, xMOOCs rely on a more traditional pedagogy, based on the delivery of learning contents from instructors to learners. xMOOCs have an instructional design heavily based on audiovisuals, most of them short video units provided by the course instructors. Many xMOOC videos have the format of recorded lectures or talks, screencasts or Powerpoint-like slideshows, all of them presenting descriptive content about the course topic (Karsenti, 2013). cMOOCs follow a *connectivist* pedagogy, while xMOOCs adhere to a *cognitive-behaviorist* model (T. Anderson & Dron, 2011).

xMOOCs platforms appeared later than cMOOCs and were usually associated to commercial ventures. Most large and well known MOOC platforms, like Coursera, edX, Khan Academy and Udacity are all xMOOCs. This study will focus exclusively on this dominant xMOOC course model and how videos are used to provide learning contents. In this text, I will use the term "MOOC" as a synonym of "xMOOC", a common custom in press and research works.

3.2.2 Objectives of the study

In this study I want to explore the usage of instructional videos in current xMOOC platforms. My purpose is to obtain a reliable account of the actual usage of communication styles in MOOC videos and how these styles are associated with MOOC characteristics, such as the platform, language and subject.

The study also proposes a categorization of MOOC video styles, based on their relative frequency and communicative approach. In addition, this article includes a final discussion on the possible causes of the observed usage of video styles.

I believe that this work will contribute to a better understanding of instructional videos in MOOCs and will help researchers in the characterization of video features, production techniques and communication patterns.

3.2.3 Typologies of online educational videos

To accomplish the aforementioned objectives, it is necessary to define a conceptual framework for categorizing MOOC videos. In Chapter 2 I have enumerated some works on classification of educational video presentation formats, with an emphasis in MOOC instructional videos. At present, there is no standardized taxonomy of educational video styles. In addition, the terminology is still maturing, as new video techniques emerge, and neologisms are coined. In this changing scenario, I will try

to adjust to common terms, but I will have to define precisely what I understand for each term used in the typology that I will provide.

3.3 Study design

3.3.1 Research questions

This research will classify videos according to their *communication style*: how the instructional contents are expressed in the video, by arranging visual items and sounds in space and time. It is important to remark that *communication style* deals with the *representational* features of videos, and not with the video *production technique*. Therefore, I will categorize the product outcome, not the way it has been produced.

With this definition given, this study has two main research questions to investigate:

- What are the communication styles and features used in instructional videos in current MOOCs?
- Are there significant differences between MOOC platforms or course subjects, regarding the usage of representational styles in their instructional videos?

3.3.2 Method overview

MOOC platforms

Five MOOC platforms were chosen for this study: *Coursera* and *edX* from United States; *FutureLearn* from UK; *MiriadaX* from Spain and *FUN* from France. These all are generalist MOOCs serving worldwide communities: a global audience for Coursera and edX; British and English-speaking people for FutureLearn; Spain and Latin America for MiriadaX; the Francophone community for FUN. I have discarded popular MOOC sites like Khan Academy and Udacity because they cover very narrow fields of knowledge, like Mathematics or Computer Programming.

Phases of the study

The study was developed in two stages. In a first phase, a qualitative survey was made to account what types of video communication styles are most frequently used in the selected MOOC platforms. This study resulted in the identification of seven video styles and the emergence of a broad categorization of "board-centric" and "speaker-centric" styles, as I will discuss below. The second phase consisted of a quantitative survey of video style usage in a sample of 116 MOOCs. With these data, a statistical analysis (descriptive and inferential) was made to answer the research questions and to extract more general conclusions.

3.3.3 Method for Phase 1

The main goal of the first stage was to set up a useful classification schema for the video communication styles. A small sample of courses from the evaluated MOOC platforms was examined by the authors and was contrasted with preceding classification proposals, in order to build a taxonomy of styles whose elements should exhibit a good balance of these criteria: a) meaningful for non-scholars; b) non-ambiguous; c) non-overlapping.

3.3.4 Method for Phase 2

Sampling period

All the selected platforms offer courses in time windows, so the course population varies over time. In order to reduce sampling biases related to season, five sampling periods were defined in a time window covering the first semester of 2015: February 11-14; March 24-31; April 20-May 11; May 20-June 8; June 18-July 8.

Course selection

For FUN, FutureLearn and MiriadaX platforms, we enrolled in all the courses that were available in each sampling window. Coursera and edX have a large number of available courses (more than 300 courses during the sampling window), so we enrolled in a subset of courses, following the same sequence shown by the platform's course search page. Some courses had not published the full course material when the sample was made: those courses having available less than one third of the full syllabus were discarded from the sample.

Video selection

For each course, we examined the full stock of instructional videos they contain. Some courses included links to external videos as part of reference material: this kind of video was not considered for the study. Other videos that have been discarded are course presentation videos, welcomes, promotional clips and descriptions on how to use the course material. Finally, we have not considered recorded hangouts and other additional videos that sometimes can be found in course discussion forums.

Video styles and features

For each course, a number of data were collected: subject area (see below); video styles used and other video features: freehand writing or marking; background/setting (neutral, office, on location); in-video quizzes; use of cartoons; use of actors instead of instructors. The video styles were coded using the definitions set in Phase 1, by a team of three evaluators. Each course was evaluated independently by two persons. When a discrepancy in the coding occurred, the evaluator pair had to reach consensus on a unanimous decision. Some videos are composed of segments that may use different styles: if this was the case, all styles were accounted for. If a video could

not be clearly assigned to any of the seven styles, the evaluator would assign it to the "others" category, as well as write an explanation.

Subject areas

Every course was assigned to one of these seven subject areas: Arts and Humanities; Business and Management; Social Sciences; Natural Sciences and Mathematics; Health and Medicine; Engineering and Computing; Everyday Life. This last category has been conceived to assign non-academic, practical courses about common life matters such as training for finding a job.

3.4 Phase 1: identifying video styles in MOOCs

Among all the reviewed classification schemas, I have considered Hansch et al (2015) catalog as a good starting point. However, I consider that this typology intermingles different dimensions in a single list, making it unnecessarily long and concealing the existence of some category groupings. For example, the "Talking Head", "Webcam Capture" and "Green Screen" types share much in common than they do with other types: in fact, these three only differ in the scene setting or background. That led me to simplify the Hansch et al's scheme, as well as adding the "Documentary" style that was observed in a significant amount of Phase 1 sample courses. I decided to suppress the "Animation" and "Demonstration" types because of the lack of relative frequency in the previewed courses.

The resulting seven styles are named "Talking Head", "Live Lecture", "Interview", "Slides", "Screencast", "Virtual Whiteboard" and "Documentary". Some of these terms have well established meanings, while others need a precise definition for the context of this research, which I will provide below. Sometimes a MOOC video uses a combination of these basic styles, but usually a single style is the dominant one. Table 3-1 sketches the mapping between these seven styles and the list by Hansch et al's report on MOOC video formats (Hansch et al., 2015). It can be seen that this taxonomy works roughly as a grouping of Hansch et al's typologies in more general classes.

Table 3-1. Mapping with Hansch et al (2015) video typologies

Styles in this study	Correspondences with Hansch et al (2015)
Talking Head	Talking Head Text-Overlay Actual Paper/Whiteboard Webcam Capture Green Screen On-Location (*)
Live Lecture	Classroom Lecture Live Video (*)
Interview	Recorded Seminar Interview Conversation Live Video (*)
Slides	Presentation Slides with Voiceover Picture-in-picture (**)
Screencast	Screencast
Virtual Whiteboard	Khan-Style Tablet Capture Udacity-Style Tablet Capture
Documentary	On-Location (*)
Other styles	Animation Demonstration

(*) "On-Location" and "Live Video" types match more than one style.

(**) "Picture-in-picture" is equivalent to my "Head and slides" substyle.

3.4.1 Style definitions

This section contains definitions for the seven main video styles, such as they will be considered along this paper. Figure 3-1 shows screenshots of all these styles. I made an effort to construct definitions that ease the following phase of style coding in the course sample, therefore avoiding ambiguity.

Talking Head. It is a video lecture whose most frequent shot is a talking human speaker who covers a large frame area (+30%) and is *not* surrounded by slides or other text-rich elements. The speaker addresses the audience: she or he looks at the camera

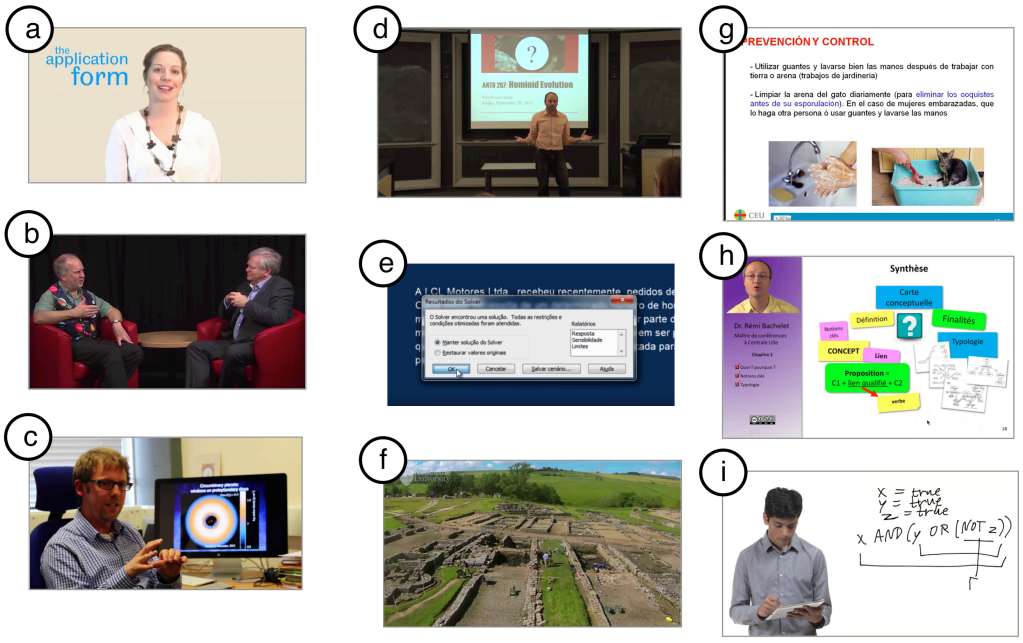
most of the time in a pretended eye-to-eye contact (the learner is addressed to using grammar second person). Sometimes overprint text is shown to enforce key ideas of the narration, or the scene switches to show another kind of material (still images, short video clips, etc.). Those insertions represent a relatively small amount of video time.

Live Lecture. It is the live recording of a classroom lecture or conference talk. An in-classroom audience is visible or implied. The learner's role is third person. The video should show some degree of edition (i.e. switching shots and cameras), but always keeping the overall perception of being recorded in a single take.

Interview. One person or more answer questions or discuss about a topic. An interviewer may or may not be present. There are two main approaches for the interviews: the *dialogic* (several people are involved in a conversation) and the *declarative* (each speaker answers a tacit question, but there is no explicit conversation). The key feature that differentiates an "Interview" from a "Talking Head" video is that in the first case, speakers do not address the audience and do not show direct eye contact (learner is third person).

Slides. In its most basic form, it is an animated sequence of PowerPoint-like slides with a voiceover talk (*slideshow* or *slidecast*). Most frequent versions of this style display the speaker as a small "talking head" placed in a marginal area of the frame (most commonly at the right bottom). Sometimes this substyle has been referred as

Figure 3-1. Screenshots of most frequent video styles in MOOCs



a) Talking Head. b) Interview (dialogic). c) Interview (declarative). d) Live Lecture. e) Screencast. f) Documentary. g) Slides. h) Head and Slides. i) Virtual Whiteboard.

"picture-in-picture" (Hansch et al., 2015; René F. Kizilcec, Papadopoulos, & Sritanyaratana, 2014), but I will call it with a more specific term: "**Head and slides"**, to avoid confusion with other picture-in-picture layouts.

Screencast. This is the visual recording of a computer session screen output, as defined by Udell (2005). It will usually include a voice narration with a description of the actions being taken.

Virtual whiteboard. This style has been popularized by Khan Academy videos. A virtual whiteboard is shown where an instructor draws content, for example mathematical formulas, diagrams or short text. The whiteboard is often blank at the video start. The instructor's face is usually not displayed, though some variants of this style show human hands and/or a pen doing the drawings.

Documentary. This is the standard cinematographic genre whose typical structure consists of a narration and filmed segments of stock material about a topic. The narrator may or may not be displayed; in this latter case, their presence represents a minimal fraction of the video length.

3.4.2 Characterization of styles

Every style can be characterized by a set of features:

- The *content displayer* is the main representational item that provides instructional information within the video frame. All styles but one use a human speaker or a rectangular board for this purpose. I will discuss this in a next section.
- The *learner's role* is the position of the viewer in the video narrative: second person or third person.
- The *text density* is the average amount of written text that is displayed in the video frame. Some styles use substantially more text than others.
- The *setting* or scenario may be *natural* (a classroom, an office, etc.) or *artificial* (chroma display, computer screenshot, etc.).

The characterization for all the styles is shown in Table 3-2.

3.4.3 Speaker-Centric versus Board-Centric videos

Table 3-2. Characterization of the video styles

style name	content displayer	learner's role	text density	setting
talking head	speaker	2nd person	low/medium	variable
live lecture	speaker	3rd person	low/medium	natural
interview	speaker	3rd person	low	natural
slides	board	2nd/3rd person	high	artificial
screencast	board	2nd/3rd person	medium/high	artificial
virtual whiteboard	board	2nd/3rd person	medium	artificial
documentary	n/a	3rd person	low	variable

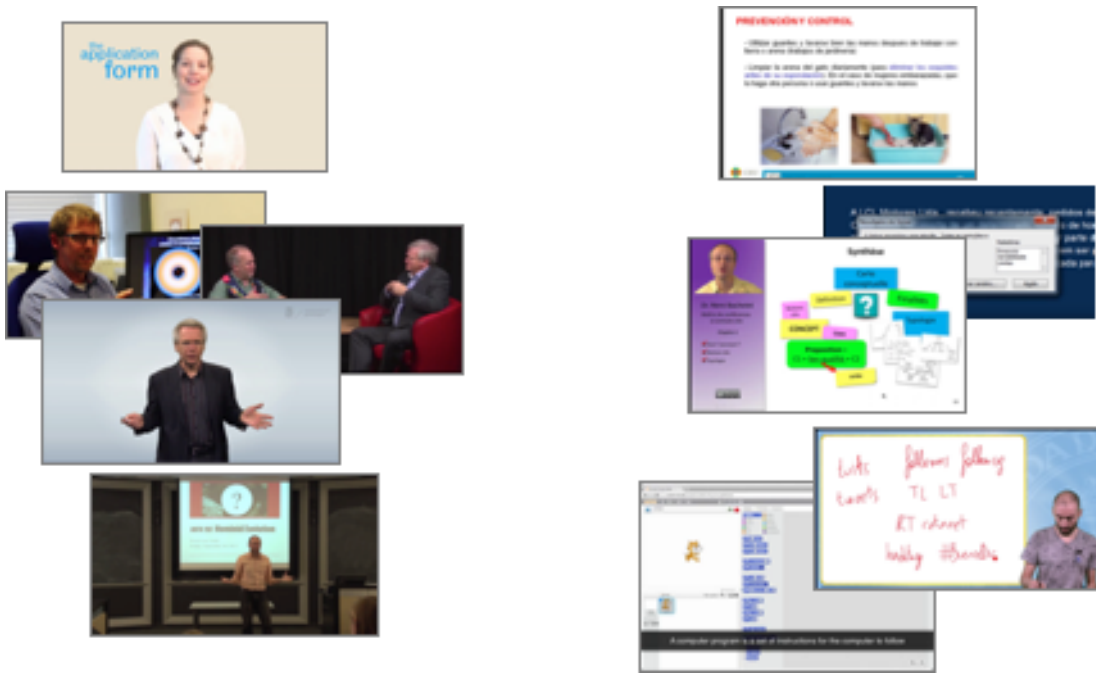
During my qualitative study, I have noticed that most MOOC videos take a single structural item as its main provider of instructional content. Depending on what type of item plays that role, I find two classes of videos: "board-centric" and "speaker-centric".

- **Board-centric** videos use a rectangle-shaped surface (a *board*) where instructional contents are presented. This board fills a large frame area or the full frame.
- **Speaker-centric** videos use a visible human speaker as the main vehicle to provide content. The speaker is visible most of the time. Sometimes more than one speaker may be present.

Figure 3-2 shows examples of both style families. Board centric styles include Virtual whiteboard (Khan-style tablet drawings); Slides (including "Head and Slides") and Screencasts. Speaker centric styles include Talking Head videos, Live Lecture recordings and Interviews.

Speaker centric videos tend to provide oral information, while in board centric videos the visual information (text or figures) is principal. Visual content items like charts

Figure 3-2. Speaker centric videos (left) and board centric videos (right)



and pictures may be also present in speaker centric videos, but in that context they often work as a complement to the primary spoken contents. On the other hand, board centric videos may display a speaker (as a small picture-in-picture, or occasional interleaved full shots), but this representation often works as a stylistic complement whose main role is not to provide contents by itself, but to help in other purposes such as enhancing attention or engagement, as suggested by Kizilcec, Papadopoulos and Sritanyaratana (2014). Engagement and attraction originated in speaker presence may also be related to the development of an intimate tutorial relationship observed by Adams, Yin, Vargas Madriz and Scott Mullen (2014) in their exploration of the learning experience of MOOC students.

The board- and speaker-centric divide has links with video production techniques. Board centric videos tend to be rooted on screen capturing procedures, while typical speaker centric videos are based on real-life video recordings.

Finally, I have to note that this classification works actually as a *spectrum*. One can find "pure" board centric styles, such as a screencast or a slideshow, and "pure" speaker centric videos like an unedited lecture capture; in the middle, there are styles with a combination of degrees of "board" and "speaker" centricity. This occurs with the "Head and Slides" substyle, which is still board-centric but shows some degree of speaker presence.

3.5 Phase 2: quantitative survey

3.5.1 Course demographics

A total of 116 courses have been evaluated. One of them did not use videos at all, thus it was removed from the study, leaving a total of 115 courses. Table 3-3 shows the distribution of courses by MOOC platform and course subject area.

Surveyed courses have been made by 84 institutions from 15 countries worldwide: United States (44 courses), Spain (20), United Kingdom (19), France (15), Latin America (7), Eastern Asia (4), Australia (3) and other European countries (3).

English is by large the most used language (74 courses), followed by Spanish (25), French (15) and Portuguese (1). All courses made in French language were provided by FUN, while Spanish-speaking courses came from MiriadaX (20), Coursera (4) and edX (1). The course in Portuguese was hosted in MiriadaX.

Table 3-3. Number of sampled courses, grouped by platform and course subject

	Coursera	edX	FUN	FutureLearn	MiriadaX	**Total**
Arts and Humanities	3	5	1	7	3	**19**
Business and Management	7	3	1	1	2	**14**
Engineering and Technology	5	8	4	1	5	**23**
Everyday Life	2	0	0	2	2	**6**
Health and Medicine	4	1	2	2	5	**14**
Science	3	6	2	3	0	**14**
Social sciences	4	6	5	6	4	**25**
Total	**28**	**29**	**15**	**22**	**21**	**115**

3.5.2 Video styles

Table 3-4 and Table 3-5 show the usage of video styles across MOOC platforms and course subjects, respectively. “Talking Head” and “Slides” are the most used styles in general, but there are some differences across platforms: “Slides” is the most used style in Coursera and MiriadaX, while “Talking Head” is the most frequent in edX and FutureLearn.

Table 3-4. Video style usage by MOOC platform

	Coursera	edX	FUN	FL	MX	All platforms
Talking Head	15 (54%)	21 (72%)	6 (40%)	21 (95%)	10 (48%)	73 (63%)
Live lecture	2 (7%)	9 (31%)	1 (7%)	2 (9%)	1 (5%)	15 (13%)
Interview	7 (25%)	12 (41%)	6 (40%)	13 (59%)	6 (29%)	44 (38%)
Slides	21 (75%)	9 (31%)	11 (73%)	10 (45%)	15 (71%)	66 (57%)
Screencast	6 (21%)	9 (31%)	1 (7%)	2 (9%)	5 (24%)	23 (20%)
Virtual whiteboard	5 (18%)	7 (24%)	1 (7%)	0	2 (10%)	15 (13%)
Documentary	2 (7%)	1 (3%)	2 (13%)	7 (32%)	4 (19%)	16 (14%)
Other styles	3 (11%)	4 (14%)	3 (20%)	3 (14%)	4 (19%)	17 (15%)

Notes. Each cell shows the count of courses that use videos with the corresponding style, and the percentage from total number of courses in the column's MOOC platform. Notice that one course may use several video styles, thus the sum of percentages in each column may be higher than 100%. FL = FutureLearn. MX = MiriadaX.

Table 3-5. Video style usage by course subject area

	Hum	BMg	Soc	HMed	Sci	Tech	EL
Talking Head	18 (95%)	8 (57%)	18 (72%)	6 (43%)	7 (50%)	13 (57%)	3 (50%)
Live lecture	3 (16%)	2 (14%)	3 (12%)	1 (7%)	3 (21%)	2 (9%)	1 (17%)
Interview	10 (53%)	6 (43%)	10 (40%)	5 (36%)	4 (29%)	8 (35%)	1 (17%)
Slides	4 (21%)	8 (57%)	15 (60%)	8 (57%)	8 (57%)	18 (78%)	5 (83%)
Screencast	2 (11%)	2 (14%)	1 (4%)	0	4 (29%)	14 (61%)	0
Virtual whiteboard	0	2 (14%)	1 (4%)	1 (7%)	5 (36%)	6 (26%)	0
Documentary	7 (37%)	0	3 (12%)	3 (21%)	1 (7%)	2 (9%)	0
Other styles	2 (11%)	0	3 (12%)	6 (43%)	3 (21%)	2 (9%)	1 (17%)
TOTAL	**19**	**14**	**25**	**14**	**14**	**23**	**6**

Notes. Each cell shows the count of courses that use videos with the corresponding style, and the percentage from total number of courses in the column's subject. Notice that one course may use several video styles, thus the sum of percentages in each column may be higher than 100%. Hum = Arts and Humanities. BMg = Business and Management. Soc = Social Sciences. HMed = Health and Medicine. Sci = Natural Sciences and Mathematics. Tech = Engineering and Technology. EL = Everyday Life.

3.5.3 Slides

Slide based videos are the second most frequent style, nearly tied with the "Talking Head" style. There are several variants of the slideshow, whose usage is shown in Table 3-6. The most frequent format is the "Head & Slides", found in 44% of courses with slide-based videos (25% of all courses). It is remarkable that 23 courses (20% of total) use slideshows with no visible speaker.

Table 3-6. Frequency of Slide substyles

slideshow types	courses	% of slideshows	% of total
all slideshow classes	66	100%	57%
head & slides	29	44%	25%
no visible speaker	23	35%	20%

3.5.4 Freehand writing and marking

Only nine courses contain videos that can be considered pure Virtual Whiteboard (Khan style or similar). This scarcity contrasts with the high reputation that this style holds. I have also accounted the use of handwriting in any form: occasional writing, marking text areas on screen, etc. A total of 30 courses (26%) exhibit some form of handwriting in their videos. This feature happens mostly over slides (13 courses) and during screencasts (8 courses).

3.5.5 Background and setting

Guo, Kim and Rubin (2014) speculate on the influence of the background and setting where the speaker is placed. They suggest using natural and simple settings such as an office. This hypothesis applies mostly to my "Talking Head" style, since the other ones are bound to a natural or artificial setting. I have explored the kind of background that is used in the "Talking Head" videos: rows in Table 3-7 show how many courses use "Talking Head" videos with a given scenario: a working place or office; a neuter background (i.e. chroma); a TV studio; on location; a classroom, conference room or theater.

Table 3-7. Background setting usage in Talking Head videos

background/setting	courses
working place	29 (40%)
neuter background	17 (23%)
TV studio	8 (11%)
on location	6 (8%)
classroom / theater	6 (8%)

3.5.6 Other features

Other minor features were also annotated:

- 12 courses make use of **actors** instead of instructors, mostly in documentary fragments, demonstrations and storytelling.
- 8 courses make use of **cartoons** and animated movies.
- 4 videos make explicit use of **humor** to communicate ideas.
- 8 courses use **in-video quizzes**, 7 of them in Coursera and the other one in edX.
- 4 courses make use of **voiceless videos**: 3 in MiriadaX and 1 in FutureLearn.

3.6 Statistical analysis

3.6.1 Style diversity

A typical course uses two different video styles (mean=2.36, SD=1.16, mode=2). Individual diversity measures for each platform and course subject are shown in Table 3-8. FutureLearn and edX host slightly more diversity than the other platforms. On the side of subjects, there is a group of lower diversity areas (Business & Management, Social Science and Health) and another one with higher diversity (Arts & Humanities, Science and Technology).

Sixty-three courses (55%) can be entirely described with just three styles: Slides, Talking Head and Interview. On the other side, only 17 courses (15%) use videos that cannot be labeled with the seven styles identified in this study. These additional styles include roleplays, storytelling, cartoons, *howtos* and others. These data reveal a low diversity of styles in MOOC courses.

Table 3-8. Style diversity, by platform and subject

Number of styles used in a course. Average and standard deviation.

	Coursera	edX	FUN	FutureLearn	MiriadaX
diversity	2.18 (1.09)	2.52 (1.27)	2.13 (1.30)	2.64 (1.18)	2.24 (0.94)

	Humanities	Business	Social Sc.	Health	Science	Tech & Eng.
diversity	2.47 (1.17)	2. (1.04)	2.2 (1.0)	2.14 (0.77)	2.5 (1.09)	2.83 (1.5)

3.6.2 Style co-occurrence

Given that a typical course uses two video styles, I have analyzed the co-occurrences of styles within a single course. The most common style pairings are listed in Table 3-9.

The most frequent pairing (Talking Head with Slides) has a lower frequency than expected if styles were chosen at random. To obtain deeper conclusions, I have performed chi-square tests of independence for all the 21 style pairings (N=115 for all tests). A statistically significant relation ($p<0.05$) was found in five pairs. For those cases, the Yules' Q statistic has been calculated to state the sign of the association (positive or negative).

Table 3-9. Top five style pairings

pairing	number of courses using that pairing
Talking Head with Slides	34 (30%)
Talking Head with Interview	30 (26%)
Slides with Interview	22 (19%)
Slides with Screencast	13 (11%)
Talking Head with Screencast	13 (11%)

These are the three positive associations:

- Virtual Whiteboard with Screencast (p=0.04, Q=+0.53)
- Virtual Whiteboard with Lecture (p=0.01, Q=+0.64)
- Interview with Documentary (p=0.03, Q=+0.53)

And these are the two negative associations:

- Slides with Lecture (p=0.01, Q=-0.63)
- Slides with Talking Head (p=0.003, Q=-0.54)

Furthermore, I have performed a similar test for independence for "Slides" against all other styles, and against the speaker centric family as a whole. The results are: a) the usage of slides and all other styles are independent (χ^2=0.236, p=0.627); b) a significant negative association is found between slides and speaker centric videos in general (χ^2=8.67, p=0.003, Yule's Q = -0.64).

This set of associations suggests a trend in course design to avoid combinations of slide-based videos with speaker centric modalities. The data also suggest that when a speaker-centric course needs to benefit from board-centric instructional material, there is a slight preference for Screencast and Virtual Whiteboard video instead of Slides.

3.6.3 Associations between style and course attributes

I performed chi-squared tests of independence to find statistically significant relations between the style usage in courses and other variables: platform, language and style. An alpha level of 0.05 has been used for all tests. A Fisher's Exact Test was made when data were not adequate for chi-squared tests. When considering subjects, I have excluded "Everyday Life" courses, since this subject category involves only 6 courses.

Given those considerations, the only significant relation found was between style usage and MOOC platform: $\chi^2(24, N=115) = 43.41$, $p<0.01$.

3.6.4 Board-centric versus speaker-centric styles

All reviewed courses make use of board- or speaker-centric videos. Only 30 courses (26%) harness styles outside the board-speaker centric spectrum. I have searched for associations between board-speaker centricity and two aspects: the MOOC platform and the course subject area. For that purpose, every course has been labeled as serving speaker centric videos, board centric videos, or both.

Table 3-10 shows the distribution of the three course classes across MOOC platforms. Tests for independence reject significant differences between platforms, neither chi-squared: $\chi^2 (8, N=115) = 13.07$, $p=0.109$, nor Fisher's Exact test: $p=0.088$.

Table 3-10. Distribution of style classes across MOOC platforms

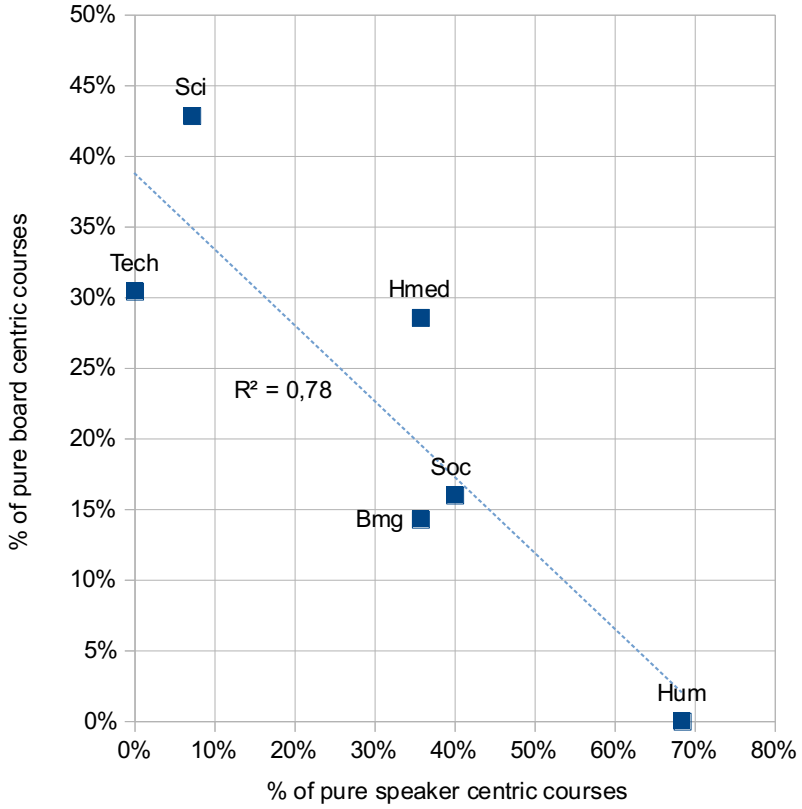
	Coursera	edX	FUN	FutureLearn	MiriadaX	All platforms
pure speaker centric	4	11	3	11	6	35
pure board centric	7	6	5	1	7	26
mixed	17	12	7	10	8	54

Table 3-11 shows how the three course classes are represented in each course subject area. The category "Everyday Life" was omitted, since it provides too few courses (6) and it has no representation in all platforms. A statistically significant association was found between course style approach and course subject area, by using Fisher's Exact Test: $p<.0001$.

Table 3-11. Distribution of style classes across subject areas

	Humanities	Business	Social Sc.	Health	Science	Tech. & Eng.
pure speaker centric	13	5	10	5	1	0
pure board centric	0	2	4	4	6	7
mixed	6	7	11	5	7	16

Figure 3-3. Proportions between pure speaker centric and pure board centric courses for each subject area



The linear regression function is plotted as a dashed blue line.

Hum = Arts and Humanities. Bmg = Business and Management. Soc = Social Sciences. Hmed = Health and Medicine. Sci = Natural Sciences and Mathematics. Tech = Engineering and Technology.

Figure 3-3 plots the ratio between speaker centric and board centric style preferences for each subject area, excluding "Everyday Life". Each point (x,y) represents the x proportion of courses having exclusively speaker centric videos and the y proportion of courses having exclusively board centric videos. When a linear regression is applied, this speaker/centric ratio shows a high correlation ($R^2 = 0.78$). Figure 3-3 also reveals a certain "Art-to-Science cline" that starts from "Art & Humanities", passes through "Soft Science" and ends in "Hard Science". In this subject spectrum, Humanities clearly favor speakers, Hard Science and Technology prefer boards, while "Soft Sciences" keep an intermediate position.

3.6.5 Discussion on survey results

This survey evidences a low diversity of communication styles in MOOC videos, both internal (a typical course employs two styles) and as a population: seven styles describe the majority of courses, and three of them (Talking Head, Slides and Interview) can label more than a half of the whole sample. Some well-known communication styles, as storytellings, live demonstrations and cartoons, have a testimonial presence in the examined MOOCs. Much variation of MOOC video styles can be explained with the speaker-centric -to– board-centric spectrum.

There are observable differences between Humanistic and Technological disciplines in the frequency of use of board centric or speaker centric videos. This relationship between speaker/board centricity and course subject may be linked to cultural factors, or there may be some relation to intrinsic properties of the contents: for example, teaching about mathematical formulas or complex engineering structures may be hard without displaying equations or diagrams. That would explain part of the Science and Technology preference for board centric videos. The actual source of these differences in communication styles is something to be explored in further research. Whatever is the case, the advice is not to immediately generalize a research finding on the learning efficiency of instructional video styles when the research has been limited to one particular subject area.

3.7 Interpretation of MOOC video style diversity

I want to finish this chapter with a reflection on the systemic causes of the observed MOOC focus on "Speakers and Boards" videos and the observed lack of style diversity.

Many MOOCs found in generalist platforms are built from existing face-to-face courses. Usually the original course is adapted to the MOOC format, changing the interfaces and course materials, but keeping the same instructional design. Most redesign effort is pushed into course assignments and discussion forums management due to the course attendance change of scale (Fredette, 2013; Kellogg, 2013).

The results from this survey are consistent with the observation that the underlying didactic technique in most MOOCs is the classic instructional lecture (Karsenti, 2013). Lectures are often adapted from the classroom to the MOOC platform with no fundamental changes, except for the audience decoupling. There are two simple approaches to accomplish this direct adaptation: one is to record your lecture talk (and edit it) and the other one is to print out your lecture notes. These two approaches lead respectively to the "Talking Head" and "Slides" communication styles. These kinds of videos not only are cheaper to record than other modalities, but easier to design out of current conventional course material. Other communication styles and techniques, like Khan-style tutorials, storytellings or animated demonstrations, would require more elaboration and therefore will tend to be relatively scarce in the current MOOC ecosystem. This reasoning may partially explain the distribution of video communication styles that this study reveals.

The attachment of current MOOCs to the video lecture has received criticism, as it has been acidly quoted by Ian Brogost (cited by Adams et al., 2014): "MOOCs [...] still rely on the lecture as their principal building block [...] The lecture is alive and well, it's just been turned into a sitcom". Despite the critics, the fact is that MOOCs have been successful, at least in terms of course volume and attendance. Indeed, their pedagogical conservative strategy may be one of the factors behind that success: it may have facilitated a fast transfer of educational content to the new online platforms. More disruptive or innovative approaches would have increased course production costs and consequently would have slowed the MOOC ecosystem growth. The production costs of MOOCs have been considerable: Hollands & Tirthali (2014) estimated a production cost ranging from USD 38,980 to USD 325,330 per MOOC. Being too much innovative should lead to poor sustainability. Moreover, some tools have been created to significantly reduce productions costs. For example, "Head and Slides" short-length videos can be recorded and edited with the Polimedia framework (Turro, Cañero, & Busquets, 2010) in almost real-time with a very affordable investment.

All things considered, the conservative scenario depicted by this study may be just a sign of the early stages of a MOOC ecosystem that has been focused on the growth of course offer, thus favoring pre-existing course reuse. It is feasible that in near future, MOOCs would add more diversity and more innovation in their audiovisual styles, as more courses will be developed free from the ties of legacy contents. Today, some may consider ironic that one Coursera's course on Digital Storytelling production (Coursera, 2015) is made up of slideshow videos and talking head lectures. Instead we can view it as a natural step in MOOC evolution, in which the current system leverages its modest resources to help teachers learn new communication styles, so they can build more innovative next generation MOOCs.

Chapter 4. Building the taxonomy

4.1 Chapter overview

This chapter describes the process of building the taxonomy of instructional video characteristics. The process starts from the theoretical bases and the literature review described in Chapter 2, as well as the finding of the field study of Chapter 3.

The taxonomy has been constructed through an iterative method which starts with the extraction of a raw inventory of characteristics taken from the selected works cited in Chapter 2. From this inventory, a process of clustering and categorization has been performed, giving as a result a first draft of the classification scheme. This scheme has been integrated into the theoretical frameworks of Systemic Functional Linguistics and Multimodal Discourse Analysis. John Bateman's GeM framework (Bateman, 2008) has been chosen as the basis to adapt the classification scheme, since it is a multimodal model for printed texts that is easily adaptable to other media, such as instructional videos.

This chapter will describe the method of the process and how the process was developed for the current study, including the intermediate results of the iterations. It is important to note that the final outcome of the process, that is, the final taxonomy, will be described in Chapter 5 (next chapter). With this organization I avoid mingling intermediate details about the development with the description of the final framework.

4.2 Building the taxonomy: goals and method

4.2.1 Conclusions from the literature review

The literature review made in Chapter 2 has provided a broad and diverse of interesting features of instructional videos. These references help to understand how various structural elements in videos contribute to learning and communication goals. The literature review comprises both Multimedia Learning and Discourse Analysis theories, which offer multiple perspectives on the video structures and their functions. Multimedia Learning theories provide strong empirical support for their claims, but lack insights on certain areas such as the structure of spoken discourse. On the other side, Discourse Analysis lacks empirical support for the effects of discourse in learning, but it provides a sound descriptive model for discourse structures.

The literature review from Wetzler et al. (Wetzel et al., 1993) has a broad coverage of video characteristics that is the closest to the scope that this study aims, but it lacks a theoretical framework and it is outdated. We need to update this landmark

study with more recent findings and a theoretical support. Multimedia Learning and Discourse Analysis will help in this endeavor.

In particular, **Systemic Functional Linguistics (SFL) will provide a useful descriptive framework to build the taxonomy**. The concepts of semiotic resources, stratification, constituency and rhetoric relations are powerful and suitable to organize the structural description of an instructional video. Every video structural item can be regarded as a semiotic resource. Semiotic resources define meaningful relations between them, of any kind: spatial, temporal, conceptual... Those *rhetoric* relations make meaning and have influence on learning. For example, the signaling principle may be expressed in terms of spatial, temporal and logical relations between one semiotic resource (a cueing device, e.g. an arrow) and another (a signaled item, e.g. a displayed text). The stratification means that semiotic resources are organized into layers, each one with their own physical or logical properties. All these concepts from Systemic Functional Linguistics will be of great help in a taxonomy of video structures, in particular if we organize the structures in a layered hierarchy of domains or *strata*.

4.2.2 Design goals

These will be the design criteria for the taxonomy:

- The taxonomy will be built starting with a bottom-up process from the literature sources found in Chapter 2.
- Systemic Functional Linguistics and Multimodal Discourse Analysis (SFL-MDA) will be the theoretical framework for the architecture of the taxonomy.
- The taxonomy will be validated and refined with further evidence from scientific literature reviews.

4.2.3 Outline of the method

To satisfy the design goals, this method has been defined for the taxonomy construction:

1. Collect raw characteristics from literature review sources.
2. Cluster and categorize the collected characteristics.
3. Add the findings from typologies of video styles.
4. Adapt Bateman's GeM (Bateman, 2008) as a framework to organize the categories.
5. Refine the classification scheme with domain-specific literature reviews.

4.2.4 The method in detail

Extract raw characteristics

The process starts with a bottom-up discovery of characteristics. We will start with a selected collection of research works from the literature review performed in Chapter 2: Wetzel at al.'s review on instructional video, the Multimedia Learning Principles, guidelines for instructional video design and film annotation methods and tools. From each literature source, a set of characteristics in videos/films will be extracted. The sources are diverse and relevant enough to trust on the significance of the set of collected characteristics.

The outcome of this iteration is a list of instructional video characteristics.

Categorization and clustering

A process of categorization and clustering will be performed on the raw inventory of characteristics. The outcome of this iteration will be a categorization scheme. The clustering will be manual, made by the author of this study. The aim is to define a set of categories that reflects the concepts used in instructional video design and analysis, and that at the same time keeps intragroup homogeneity.

The outcome of this iteration is a classification scheme of characteristics.

Typologies of instructional videos

There is another set of research works in Chapter 2 that are qualitatively different to those selected in the first iteration of this process. These are the works on typologies of instructional video styles and formats. A process of categorization will be performed on these sources, to be blended with the classification scheme of the previous iteration.

The outcome of this iteration is a refined classification scheme of characteristics.

Enhancing Bateman's GeM

In this iteration, the classification scheme will be integrated into the theoretical framework of SFL-MDA, by adapting Bateman's GeM framework (Bateman, 2008) to the specificities of instructional videos and the categories found in the former iterations. The new scheme would reflect the desirable property of stratification (layered architecture).

The outcome of this iteration is a classification scheme inspired in Bateman's GeM.

Refining the classification scheme

The last stage of the taxonomy construction is to perform specific literature searchs on each one of the categories identified in the classification scheme. The goals of this review process are: a) validate the classification scheme through evidences in the literature; b) discover characteristics not identified in previous iterations and

incorporate them in the classification scheme; and c) create lower-level taxonomies for each of the classification categories.

The outcome of this iteration is a refined classification scheme, plus a collection of second-level taxonomies, one for each category of the classification.

4.3 Collecting raw characteristics

The first stage of the taxonomy building is the extraction of raw characteristics from the research sources found in the literature review that is described in Chapter 2. For each source, I have extracted those characteristics in videos/films that have been pointed as relevant by the authors. In my opinion, these sources constitute an adequate basis to start the process, since many of these research works purposefully seek features that are related to learning or to meaning making. In addition, these works hold enough authority and relevancy in their respective fields to have confidence on the significance of the set of extracted characteristics.

4.3.1 Selected sources

The selected sources are organized in these four groups:

- Wetzel et al's review on instructional video (Wetzel et al., 1993).
- Multimedia Learning Principles, as found in *The Cambridge Handook of Multimedia Learning* (Mayer, 2014d).
- Guidelines for instructional video design (Kay, 2014; Koumi, 2006; Loch & Mcloughlin, 2011; Morain & Swarts, 2012; Swarts, 2012; van der Meij & van der Meij, 2013).
- Film analysis annotation methods and tools (Baldry & Thibault, 2006; Lim Fei et al., 2015).

Wetzel et al.'s review has been chosen as a singular source because, despite its age (1993), it stills contains an unpaired comprehensive compilation of findings in theory and practice of educational film/video production.

As regards Multimedia Learning Principles, I have collected every characteristic that is linked to some learning principle, both for the set of 'classic' principles and the so-called 'advanced' principles from the latest edition of the *Cambridge Handbook*.

4.3.3 Results

The collection process in this stage resulted in a list of 55 characteristics, which are enumerated in alphabetical order in Table 4-1. It can be noticed that there is some overlap between sources. For instance, all sources agree in the video segment length as a noteworthy property. Other pervasive features among sources are the cueing devices, the information complexity, the speech style (e.g. conversational versus formal) and the temporal contiguity of information sources.

sources	characteristic
-mf-	agent's gestures
wm--	animation
---p	audio and video quality
---p	building confidence about the speaker
w-f-	camera position and angle, perspective, viewpoint
w---	caption/subtitle
---p	cognitive complexity
-m-p	connect with pre-existing knowledge
w-f-	continuity
---p	create and fulfill learner's expectations
wm-p	cueing device (auditory, visual)
w-f-	cutting and editing (cuts, shots)
-m-p	decorative or irrelevant material
---p	discourse signposts
-m-p	embodied on-screen agent
w---	encouraging student participation
---p	explicit rhetoric organization
-m--	expository on-screen text
w---	focus length effects
w-f-	framing
w-f-	humor
-m--	in-video interpolated test
wm-p	information density, picture complexity
w---	questions
-m--	interactive controls
---p	introduction of new concept
---p	introduction, step, summary, conclusion
w-f-	lighting and color
w-f-	music (background)
--f-	myths, values, ideas, gender
-m--	navigation controls
---p	non-essential details of the task
---p	overview (tutorial)
---p	persuasive appeals
w---	presentation format: dramatization, expository
w-f-	presentation pacing
---p	real-life recording vs CGI
---p	repetitions of content
-m-p	segmentation of video contents
-m--	self-explanatory prompt
w-f-	shot length
---p	show user action, show system reaction
-m-p	simultaneity of multiple modalities (e.g. text and voice)
w---	sound effect
-mf-	spatial contiguity of information sources
-mfp	speech style (conversational/formal)
-m-p	speech style (native/foreign, human/robotic)
---p	tempo variations
-mfp	temporal contiguity of information sources
---p	title (video title or section title)
-m--	user-generated annotation
w-f-	verisimilitude
---p	video resolution
wmfp	video segment length
w-f-	zooming

Table 4-1. Raw list of instructional video characteristics.

Sources: **w:** (Wetzel et al., 1993). **m:** multimedia learning principles. **f:** film discourse analysis. **p:** guidelines for instructional video design.

4.4 Categorization and clustering

4.4.1 Categorization: first stage

From the raw list of characteristics that were collected in the first stage, I performed a process of categorization and clustering. At a first sight, three structural dimensions emerged from the inventory of characteristics:

- **Identity.** Entities versus properties.
- **Composition.** Simple versus compound features.
- **Functional Domain.** Audiovisual representation versus spoken discourse.

Table 4-2 shows the list of 55 characteristics with their associated dimensions.

Identity

On the one hand, we find characteristics that refer to identifiable entities, such as an on-screen agent, a block of text, a gesture or a spoken phrase. On the other hand, we find characteristics that are attributes of some entity which cannot be constituted as identifiable objects. These are the cases of the duration of a video segment, the mood of a voice or the ethnicity of a speaker.

To accommodate this difference into the taxonomy, we can create a basic binary dimension: the *identity*. This dimension has only two instances, which we will call *entities* and *properties*. Entities are objects having an identity and properties are attributes belonging to some entity. This dichotomy between entities and properties is the same that is practiced in data modeling methods, such as the Entity-Relationship model[9].

Composition

Some features in the list are defined or built as a composition of other listed features or they result from relationships between multiple items. Many of these compound features are related to physical representations. For example, the 'spatial contiguity of information sources' emerges from the spatial relation between some physical items. The 'explicit rhetoric organization' results from the logical relations between the rhetorical segments that compose the spoken discourse.

Functional domain

A salient grouping in this list of features is between *audiovisual representational* features and *spoken discourse* features. Examples of audiovisual features are the audio and video quality and the on-screen agent. Examples of spoken discourse features are persuasive appeals, self-explanatory prompts and rhetoric segments such as the introduction and the conclusion.

[9] See https://en.wikipedia.org/wiki/Entity%E2%80%93relationship_model

Most features from the list match in one of those two categories, though there are two kinds of exceptions. First, there are some features that can be labeled in both domains: segmentation of video contents, continuity, cutting/editing, presentation pacing and tempo variations. Second, there is a group of features that are neither spoken discourse nor plain representational items. Instead, they are devices intended for user *interaction*: in-video interpolated tests, interactive controls, navigation controls and user-generated annotations.

categories	characteristic
En·Si·Re	agent's gestures
En·Si·Re	animation
Pr·Si·Re	audio and video quality
Pr·Co·Sp	building confidence about the speaker
Pr·Si·Re	camera position and angle, perspective, viewpoint
En·Si·Re	caption/subtitle
Pr·Co·Re	cognitive complexity
Pr·Co·Sp	connect with pre-existing knowledge
Pr·Co·--	continuity
Pr·Co·Sp	create and fulfill learner's expectations
En·Si·Re	cueing device (auditory, visual)
En·Si·--	cutting and editing (cuts, shots)
En·Si·Re	decorative or irrelevant material
En·Si·Sp	discourse signposts
En·Si·Re	embodied on-screen agent
En·Co·Sp	encouraging student participation
En·Co·Sp	explicit rhetoric organization
En·Si·Re	expository on-screen text
Pr·Co·Re	focus length effects
Pr·Co·Re	framing
Pr·Co·Sp	humor
En·Si·In	in-video interpolated test
Pr·Co·Re	information density, picture complexity
En·Si·Sp	inserting questions
En·Si·In	interactive controls
En·Si·Sp	introduction of new concept
En·Si·Sp	introduction, step, summary, conclusion
Pr·Si·Re	lighting and color
En·Si·Re	music (background)
Pr·Co·Sp	myths, values, ideas, gender
En·Si·In	navigation controls
En·Si·Sp	non-essential details of the task
En·Si·Sp	overview (tutorial)
En·Si·Sp	persuasive appeals
Pr·Co·Sp	presentation format: dramatization, expository
Pr·Co·--	presentation pacing
Pr·Si·Re	real-life recording vs CGI
En·Co·Sp	repetitions of content
En·Co·--	segmentation of video contents
En·Si·Sp	self-explanatory prompt
Pr·Si·Re	shot length
En·Co·Sp	show user action, show system reaction
Pr·Co·Re	simultaneity of multiple modalities (e.g. text and voice)
En·Si·Re	sound effect
Pr·Co·Re	spatial contiguity of information sources
Pr·Si·Sp	speech style (conversational/formal)
Pr·Si·Sp	speech style (native/foreign, human/robotic)
Pr·Co·--	tempo variations
Pr·Co·Re	temporal contiguity of information sources
En·Si·Re	title (video title or section title)
En·Si·In	user-generated annotation
Pr·Co·Sp	verisimilitude
Pr·Si·Re	video resolution
Pr·Si·Re	video segment length
En·Si·Re	zooming

Table 4-2. Categorization of video characteristics: first stage.
Categories: **En**: Entity. **Pr**: Property. **Si**: Simple. **Co**: Compound.
Re: Representational. **Sp**: Spoken discourse. **In**: Interaction device.

4.4.2 Categorization: second stage

I have harnessed the preliminary three-dimensional categorization (identity, composition, function) to craft a clustering on the characteristics that makes more sense in the conceptual world of video analysis and production, and at the same time exhibits a reasonable intragroup homogeneity. As a result, I have come to these eight categories:

- **Medium properties.** These are low-level properties provided directly by the video medium or the recording devices.
- **Basic representational entities.** A group of entities in the raw list can be regarded as basic 'building blocks' with which the video is built: on-screen agent, text, voice, to name a few. These entities are arranged in space and time at the video designer's convenience.
- **Interaction devices.** This category was already identified in the first categorization stage, as one of the functional domains.
- **Spatial arrangements.** These are manipulations over the video frame or over the representational entities in order to obtain a particular visualization of the recorded scene or subject. They are often related to cinematographic techniques of camera settings.
- **Temporal arrangements.** These are entities and properties dealing with the temporal sequencing, such as the chosen video segment duration.
- **Complexity properties.** The list of characteristics contains some properties that measure the complexity of the depicted scene or of the content flow. Some authors relate the complexity to the learning effort.
- **Spoken discourse structures and properties.** This is by far the cluster with the most items. Most of them are rhetorical structures which with a spoken discourse is built.
- **Communicative goals.** Some discourse properties are objectives that the video designer establishes about verbal and non-verbal communication.

Table 4-3 shows the properties and entities that compose every category.

Table 4-3. Categorization of video characteristics: second stage

category	properties and entities
medium properties	audio and video quality, video resolution, lighting and color, image recording (real-life, CGI)
basic representational entities	agent's gestures, animation, caption/subtitle, cueing device, decorative or irrelevant material, embodied on-screen agent, expository on-screen text, background music, sound effect, title
interaction devices	in-video interpolated test, interactive controls, navigation controls, user-generated annotation
spatial arrangements	camera position and angle, perspective, viewpoint, focus length effects, framing, shot length, spatial contiguity of information sources
temporal arrangements	continuity, cuts/shots, presentation pacing, segmentation, simultaneity of multiple modalities, tempo variations, temporal contiguity of information sources, video segment length, zooming
complexity properties	cognitive complexity, information density, picture complexity
spoken discourse structures and properties	discourse signposts, explicit rhetoric organization, humor, inserting questions, introduction of new concept, introduction/step/summary/conclusion, myths/values/ideas/gender, non-essential details of the task, overview, persuasive appeals, presentation format, repetitions of content, self-explanatory prompt, show action/reaction, speech style (conversational/formal), speech style (native/foreign, human/robotic)
communicative goals	building confidence about the speaker, connect with pre-existing knowledge, create and fulfill learner's expectations, encouraging student participation, verisimilitude

4.5 Typologies of instructional videos: video genres

The second stage of categorization completed the process of extracting characteristics from the main set of research sources from Chapter 2. After that, I have incorporated other set of sources from the literature review made in Chapter 2, which are the works on typologies of instructional video styles and formats. Eight studies were selected, covering various aspects of video instruction and communication (Goodyear & Steeples, 1998; Guo et al., 2014; Hansch et al., 2015; Kay, 2012; Moes, 2012; Schwartz & Hartman, 2007; Thornhill et al., 2002; Winslett, 2014).

4.5.1 Features extracted from typologies

The selected studies are heterogeneous in goals, methods and outcomes, but there are some commonalities around video features of interest. As I pointed in Chapter 2 conclusions, some studies make emphasis on *functional features* such as teaching/learning purposes and communication patterns (Goodyear & Steeples, 1998; Kay, 2012; Schwartz & Hartman, 2007; Thornhill et al., 2002; Winslett, 2014), while other classifications tend to be more concerned with *representational formats*, evidenced by terms as 'screencast', 'webinar', 'Khan-style' and the like (Guo et al., 2014; Hansch et al., 2015; Moes, 2012).

I have made a synthesis of this material in order to elicit the aspects around which those classifications are elaborated. The result is outlined in Table 4-4 and discussed in the following. The synthesis consists of five dimensions that have been used to classify video styles:

- **Purpose/goal.** Every instructional video has a learning goal.
- **Type of recorded action.** Most videos can be distinguished by the kind of action being shown. In many cases, videos record some type of real-life event, while others are staged recordings.
- **Communication format.** Videos that record the same kind of action may have different approaches to the presentation of that action. The approach may be 'neutral', as in the 'fly on the wall' style, or it may be 'involved', as in an interview.
- **Frame format (layout and dynamics).** Several studies from the Multimedia Learning community are concerned with representational aspects of the frame, for example whether it is helpful to set an on-screen speaker in a small box overlaid on the frame (picture-in-picture). My own field study on MOOCs unveils that the relative sizes of the displaying board and the on-screen speaker is corelated with the course subject field. Other studies, e.g. (Hansch et al., 2015; Moes, 2012) distinguish the dynamics of information presentation: pen-based, typed or overprinted.
- **Scenario/background.** Various studies pay attention to the variations in the appearance of the scene setting. These variations may be originated by

the recording technique (green screen, screencast, webcam) or by the selected recording stage (on location, studio, office desk, classroom).

Table 4-4. Dimensions used to classify video styles

dimension	examples of video styles and formats
purpose/goal	seeing, engaging, doing, saying (Schwartz & Hartman, 2007); pedagogy, academic focus (Kay, 2012)
type of recorded action	lecture, conversation, testimonial, demonstration, simulation, real-life event
communication format	fly on the wall, think aloud, action with commentary, talking head, prepared script, professionally acted, dramatization, interview, testimonial, vox pop (Goodyear & Steeples, 1998; Thornhill et al., 2002; Winslett, 2014)
frame format	text-overlay, whiteboard, slides with voiceover, picture-in-picture, screencast, Khan-style, Udacity-style, animation (Guo et al., 2014; Hansch et al., 2015; Moes, 2012)
scenario/background	real classroom, virtual classroom, green screen, on location, studio, office desk (Guo et al., 2014; Hansch et al., 2015)

4.5.2 Adapting the classification scheme: video genres

Once the feature extraction on video typologies has been completed, their outcomes have to be contrasted with the first clustering that was obtained in the former section of this chapter (see Table 4-3).

Most of the five dimensions match well with the preliminary clustering: the cluster already identified a 'communicative goals' category; the communication format dimension fits in the 'spoken discourse structures and properties'; the frame format dimension is a kind of 'spatial arrangement', while the scenario/background dimension can be considered a subset of 'basic representational entities' cluster. If purpose and goals are to be added in the 'communicative goals' category, the title should change to 'goals and purposes', so that this category is not limited to communicational dimension.

The only unmatchable dimension is the 'type of recorded action', which was not explicitly identified in the previous stages. It should therefore constitute a new cluster in an enhanced classification scheme.

Before acting on the classification, it will be convenient to draw attention to the fact that the characterization of a given video style frequently does not derive from a single characteristic, but from a combination of several features. For example, a Khan-style video is recognizable as a combination of type of recorded action (lecture), a particular frame arrangement (virtual whiteboard session) and a particular

background (neutral, often dark). Testimonial 'vox pop' videos often follow an on-location recording scenario and use medium shots with the speaker in angle view.

Many configurations regularly found in certain video styles seem more cultural conventions than functional requirements. This is the point where the concept of *genre* may help in the classification scheme. Genres where introduced in Chapter 2 as a key concept used in Discourse Analysis sciences. Genres are defined as "a class of communicative events, the members of which share some set of communicative purposes" (Swales, 1990). According to V. K. Bhatia, linguistic genres are characterized by conventionalized communicative settings, such as the "recurrence of rhetorical situations, shared communicative purposes and regularities of structural organization" (Bhatia, 1997). In the context of this research, we can borrow the extended concept of *multimodal genres*, also discussed in Chapter 2 (Bateman, 2008; Lemke, 2005; Vorvilas, Karalis, et al., 2011).

A (multimodal) genre can be described as a particular configuration of semiotic resources which is generically recognizable by a social community. The involved semiotic resources may be acts of speech, visual elements, and even scene settings (which often portray semiotic value).

For example, a 'Khan-style' video is identifiable both by its visual appearance and by its purpose (usually an expository tutorial). A 'picture-in-picture' video belongs to a variant of video lecture, characterized by a particular spatial configuration which is recognizable by the community (video designers and learners who watch those videos).

Consequently, I will add a new cluster to the classification scheme, called **"video genres"**. This cluster will incorporate any configuration that could be considered a multimodal genre: a type of recorded action, but also formats as the screencast, the picture-in-picture, the talking head tutorial and many other popular styles.

4.5.3 Resulting categorization (third stage)

In summary, this iteration has resulted in a taxonomy of nine categories of video characteristics:

1. medium properties.
2. basic representational entities.
3. interaction devices.
4. spatial arrangements.
5. temporal arrangements.
6. complexity properties.
7. spoken discourse structures and properties.
8. goals and purposes.
9. video genres.

4.6 Adapting Bateman's GeM framework

The classification obtained in previous iterations could be enough to fulfill the goal of building a taxonomy of instructional video characteristics. Nevertheless, it is highly desirable that it be linked to a theoretical framework that endows it with a solid justification and enables further generalization and evolution. With this aim in mind, I will adapt the classification scheme to the theoretical field of Multimodal Discourse Analysis, more precisely to the GeM (Genre and Multimodality) framework proposed by Bateman (Bateman, 2008), which I referred in Chapter 2.

4.6.1 Why GeM

In the first place, I was clear that the framework for the taxonomy would be based in Systemic Functional Linguistics and Multimodal Discourse Analysis (SFL-MDA). I could have decided to derive the taxonomy from existing characterizations made on multimedia presentations (e.g. M. Williams, 2003) or on expository animations (Ploetzner & Lowe, 2012), but both approaches lack the support for discourse structures that are an essential requisite for a complete description of instructional videos.

Among SFL-MDA models, I have found Bateman's GeM framework as the best starting point to integrate the taxonomy of instructional video features. Bateman's proposal was originally intended to describe printed documents. It generalizes the rhetorical relationships that are usually found in classic analysis of spoken discourse, extending them to spatial and conceptual relationships than can be found in page layout blocks and graphic signs. In GeM model any semiotic resource, whatever its nature (text, picture, drawing, graphic sign), may establish rhetoric relations with other resources. Those generalized relationships take place also in other *semiotic modes* that include dynamic visualizations, such as narrative films.

Some SFL-MDA key principles that are manifest in Bateman's framework are the *stratification* and *constituency*: concepts are organized in hierarchical layers; higher level concepts are built on lower level constituents. Bateman's GeM defines five layers: GeM base (basic physical elements), layout, rhetorical, navigation and genre.

Bateman's GeM is not the only framework intended for multimodal documents. In fact, other authors (Vorvilas, Karalis, et al., 2011) have proposals that include general multimedia objects, thus covering more modalities than GeM. Nonetheless, Bateman's model is the closest that I have found to the taxonomic dimensions obtained in the initial stages of my study. GeM only needs minor adaptations to adequately describe instructional videos.

One virtue of Bateman's GeM is that it generalizes rhetorical relationships between semiotic resources, so that discursive, logical and spatial relationships can be accommodated in the framework. The model can be seamlessly extended to express other kinds of relationships, such as the temporal relationships that occur among representational items in a video (e.g. time sequencing, temporal contiguity,

continuity), as Bateman himself developed in later works (Bateman & Schmidt, 2012). The 'Layout' layer would describe both spatial and temporal relationships between semiotic resources. What is more, social and affective relationships could be also expressed in the model with minor adjustments.

As in SFL-MDA models, the stratification will allow us to clearly separate different levels of analysis of the instructional video: the physical medium, basic representational entities, complex structures, and reoccurring patterns (genres).

4.6.2 Evolving GeM to support the classification

In order to evolve GeM to support instructional videos in their contemporary form, we have to tackle this main limitation: current streamed video medium enables modalities of navigation and interaction that are far different to those in the printed medium. This limitation will be solved by adding an 'Interaction' domain in the base layer that will contain the interaction devices used in current video media.

Other changes that have been made to the Bateman's original framework are a more detailed layered architecture that is closer to the design process of instructional videos; and a new 'strategic' layer to include learning goals and purposes as characteristics supported by the model.

Interaction domain

Bateman's GeM considered a 'Navigation' layer which describes certain kind of semantic relations between physical entities that set navigational cues to the reader, e.g. arrows that link text blocks and paragraph numbering. On the other hand, the video medium provides other kind of navigational artefacts, such as a playback panel or an interactive table of contents. Some videos also include in-video interactive artefacts like interpolated quizzes or forced pauses. Moreover, research has found that those interactive and navigational devices influence the learning outcomes. Those interactive features are distinct enough to merit a separate domain in the taxonomy.

Spatiotemporal domain (renaming of GeM's layout)

The most obvious enhancement to be made on the original GeM model is to add a temporal domain to support properties and relationships that deal with time, such as cutting and editing (film montage), camera setting and movements, and temporal contiguity of elements. Bateman proposed that the concept of 'Layout' will include this kind of relationships. The first stages of my classification have identified two domains: spatial and temporal.

Following Bateman's rationale, I will merge spatial and temporal domains into a single one. In fact, many characteristics in these two domains are actually interleaved in space and time (e.g. camera manipulations, contiguity of items). Also, frame spatial layouts need not be static: they may switch over time, with a possible effect on

learner's attention. What is more, complexity measures may involve both space and time arrangements.

The only difference between this classification scheme and GeM is nominal: I will call this domain 'Spatiotemporal' instead of 'Layout'. I think this term captures better the underlying concept than a word -layout- that is often associated to spatial arrangements[10].

Layered architecture

Our taxonomy claims a more detailed stratification architecture than GeM's. From Vorvilas et al.'s models (Vorvilas, Karalis, et al., 2011; Vorvilas, Vergidis, & Ravanis, 2011), I will borrow a more nuanced layered architecture. This model will be closer to the design process of instructional videos: a physical medium (the video frame), elementary building blocks (e.g. on-screen agent, whiteboard), and upon them scenes and shots.

Strategic layer: goals and purposes

The clustering stages of the taxonomy building have revealed some features that deal with high-level purposes and goals in the video design. Bateman's GeM does not support explicitly the design goals. It does include the design decisions based on cultural factors and physical constraints, but the intentions or purposes do not form part of the model.

I have decided to add a new layer in this classification scheme to support those characteristics dealing with high-level purposes. It is called the 'Strategic' layer. This way, the enhanced classification scheme draws a full path from the designer's mind to the physical representation medium.

4.6.3 Result: a 'GeM-ified' classification scheme

Once all the modifications have been applied, we obtain the classification scheme depicted in Table 4-5. This scheme contains five hierarchical layers, from the physical medium to higher level constructs: physical medium, basic entities, compositional layer, strategic layer and generic layer. Each layer contains one or more domains that organize the video characteristics (entities and properties). Domains in the same layer have equal hierarchy. The seven domains are: medium, presentation, interaction, spatiotemporal, speech, goals and generic.

The seven domains in the scheme correspond to their respective categories in the preliminary clustering, with the exceptions of the 'spatial' and 'temporal' categories which have been merged, and the 'complexity properties' category, that has no corresponding domain. I have decided to suppress the 'complexity' from the domain inventory, since it is a small set of properties that can be predicated from any

[10] Actually, I considered also the term 'Filmic' for this domain, to highlight that many features in this domain are related to cinematographic techniques. I finally dropped this term as being somewhat obscure.

combination of compositional domain (space, time, speech). Complexity properties can be integrated into the remaining domains without losing expressive power in the scheme.

This is the (almost) final step in the conceptual building of the taxonomy. A final refinement has been made by validating the scheme with extensive literature reviews. The next chapter will describe this final classification scheme in full detail.

Table 4-5. Layered classification scheme, before refinement phase

layer	domain	description
Layer 4: generic	generic	genres: common configurations of resources in underlying layers
Layer 3: strategic	goals and strategies	communicative and instructional goals and strategies involved in the product design
Layer 2: compositional	speech	articulation of spoken/written discourse
	spatiotemporal	item composition in the video frame and compositions of frames across time
Layer 1: basic entities	presentation	entities that carry informational contents
	interaction	entities that enable direct viewer interaction
Layer 0: physical	medium	properties of the physical video medium

4.7 Refining the classification scheme

The last stage of the taxonomy construction is the one that requires more work. An extensive literature research will be performed for each one of the domains in the classification scheme.

The goals of this review process are as follows:

- Validate the classification scheme through evidences in the literature.
- Discover characteristics not identified in the previous phase and incorporate them in the classification scheme.
- Create lower-level taxonomies for each of the classification domains.

In order to keep the research effort within reasonable bounds for a doctoral thesis, only the first three layers have been addressed: physical layer, basic entities layer (presentation and interaction) and compositional layer (spatiotemporal, speech). 'Goals' and 'generic' domains have been left out of this process of refinement.

4.7.1 Method

The starting point is the set of characteristics and categories that were identified in the previous stages of this process. Over each concept, a search process is performed to find academic references and to expand the inventory of characteristics. When enough characteristics are found for a domain, a classification process is performed to provide specific sub-taxonomies for that domain.

Search process

A search on scientific databases is performed. I have used Scopus and Google Scholar as the databases. Searches return a first batch of references, which is examined to find relevant sources that support the concept as a useful characteristic in the taxonomy. A relevant source is an academic work that satisfies one of these conditions: a) it is a paper published after a peer review process in a journal or conference indexed in JCR (Journal Citation Reports) or SJR (Scimago) databases; b) it is a book or technical report with more than 20 citations in Google Scholar. The time span of the searched works reaches up to the year 2017, when this review phase occurred.

To narrow the search results, search queries usually have been constrained to include terms such as 'instructional video', 'educational video', 'video lecture' and 'video tutorial' in their abstracts, titles or keywords.

In some cases, relevant sources have been complemented with other references to peer-reviewed or authoritative works that offer additional information about the topic being reviewed.

A search round on a characteristic usually delivers a number of relevant sources that contain further references to other characteristics that were not in the current inventory. Each new characteristic is searched with the method described above. If relevant sources are found that support the characteristic, it is added to the inventory.

The process finishes when no new relevant characteristics are found. It is often reached in two or three depth levels after the first search on an original concept.

Classification process

The search process generates a set of video features and characteristics for each domain that extends the raw list obtained in the previous phases. Thanks to this wider inventory, I can perform a classification process for each domain that results in a hierarchical set of taxonomies.

Adjustments in the classification scheme

Sometimes the search and classification processes may surface new characteristics that suggest adjustments in the original classification scheme. In that case, the adjustments are made, and the process goes on.

4.7.2 Changes after the refinement process

As a result of the refinement, one major change has been applied to the classification scheme. **A 'social appearance' domain has been added.** Several properties of various items are able to trigger a social response in the learner, for example, the speaker's social group and the scene setting. These properties can be manipulated by the video designer and they can be manipulated together, giving rise to a mise-en-scène or social atmosphere. Several authors state that the social response is relevant in instructional video design, therefore I have promoted this dimension as a separate domain.

With these letters I finish the description of the whole process of building the classification scheme for the instructional video characteristics. The final outcome can be read in Chapter 5 (overall description of the taxonomy) and Chapter 6 (the taxonomy in detail with accompanying literature references).

I invite the reader to follow on the next pages.

Chapter 5. A taxonomy of instructional video characteristics

5.1 Chapter overview

This chapter describes the final result of the process of building the classification scheme for instructional video characteristics. As explained in Chapter 1, the characteristics being classified are those which have been found (or suspected) to have relevance in the learning outcomes of video watching. In Chapter 4 I have explained how the classification scheme has been built through a bottom-up conceptual clustering, departing from a list of characteristics taken from key scientific sources. The draft classification scheme presented in Chapter 4 has been refined by an extensive literature review from various fields, such as Educational Technology, Educational Psychology, Applied Linguistics and Discourse Analysis. This review has provided a broad inventory of instructional video characteristics. This inventory and the corresponding references are shown in Chapter 6.

The resulting classification scheme is organized around structural domains, which in turn are conformed into a hierarchy of layers. This hierarchy ranges from the most physical domains to the most abstract: physical medium, basic entities, compositional domains, a strategic domain and video genres. For the lower domain layers (basic entities and compositional domains), specific taxonomies have been built. These taxonomies are introduced in this chapter and will be discussed in detail in Chapter 6.

5.2 Meta-model specification

This framework uses several terms whose meaning must be clarified for better understanding. These terms are: *entity*, *property*, *class*, *domain* and *layer*. These concepts and their relations altogether conform a meta-model for this framework.

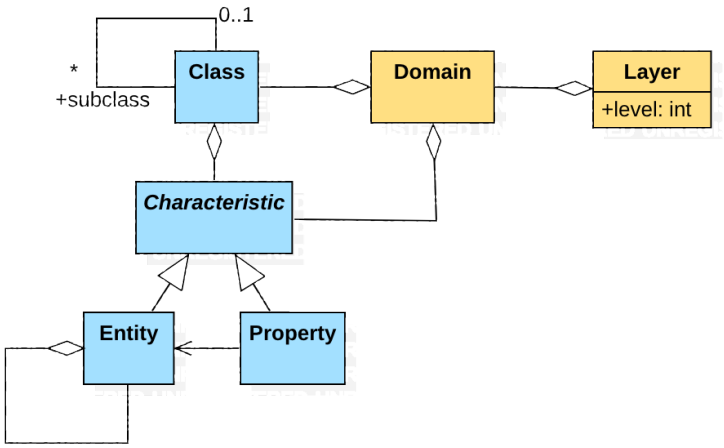
Figure 5-1 displays a UML class diagram of the meta-model.

- A **characteristic** is any of the objects for which this taxonomy has been defined. A characteristic can be either an *entity* or a *property*.
- An **entity** is any object belonging to the video that is identifiable. An entity may be a composite of many lower-level entities. Some examples of entities are: actor, audio narration, scene, rhetoric stage.
- A **property** is a value that can be measured from an entity, from a set of entities, or from the whole video object. Examples of properties are: duration, speed, word count, color, size. Some properties, such as duration, may be

applicable to both single entities and to the entire object. Other properties only have sense for the whole object, such as the video genre.

- For convenience, characteristics (entities and properties) can be grouped in **classes** that share common traits. Classes provide conceptual generalization and descriptive economy. For example, actor's gender, age and ethnicity are properties that can be grouped in a class of 'social appearance' properties. Classes can be hierarchized by using superclass-subclass relationships.
- A **domain** is a semantic space that provides an 'observational lens' to analyse the video object. Two examples of domains are the presentation domain and the spatio-temporal domain. This framework classifies characteristics across a set of relevant domains. In general, every characteristic in this framework will be assigned to one single domain.
- The domains are grouped in a hierarchy of **layers**. Layers are stacked according to their position in a physical-to-abstract spectrum, ranging from the lowest physical medium to the higher-level mindset of design goals and cultural conventions.

Figure 5-1. Taxonomy meta-model (UML class diagram)



5.3 Domains and layers

The classification scheme defines eight structural domains: *medium*, *presentation*, *interaction*, *spatiotemporal*, *speech*, *social appearance*, *goals and strategies*, and *generic*. These domains are grouped in five hierarchical layers, as shown in Table 5-1.

Table 5-1. Taxonomy domains and layers

layer	domain	description
Layer 4: generic	generic	genres: patterns of use of the basic and compositional entities, recognizable by the community of video creators and viewers
Layer 3: strategic	goals and strategies	communicative and instructional goals and strategies involved in the product design
Layer 2: compositional	social appearance	social and cultural traits that influence viewer's response to contents
	speech	articulation of discourse in its textual modality (written or spoken)
	spatiotemporal	articulation of discourse in space and time, by following film production methods
Layer 1: basic entities	presentation	entities that carry informational contents
	interaction	entities that enable direct viewer interaction
Layer 0: physical	medium	the physical substrate of the video medium, as it is available to video creators

5.3.1 Layers: from mind to frame

The structural domains can be organized in a hierarchy of layers that starts from a lowest physical level (the video frame and its surroundings) to increasingly abstract levels. Intermediate layers deal with structural building blocks and discourse making. The top layers in this hierarchy are related with the general design of the product.

This is the complete list of layers:

- **Layer 0: physical.** This is the physical substrate that serves to place the content and at the same time restricts the modes of expression of the content.
- **Layer 1: basic entities.** User-recognizable entities that carry content or provide basic mechanisms for user interaction. Video creators place these basic entities in the video as part of the production process. Most prominent

entities in instructional videos are *actors* (e.g. narrators, models) and *boards* (e.g. text boxes, slides, computer screens).

- **Layer 2: compositional.** The basic entities from Level 1 are articulated in different domains to elaborate the instructional discourse. Entities are arranged in space and time, following film edition techniques. The speech is articulated in rhetoric structures, using linguistic functions. Social appearance qualities are composed to trigger social responses on viewers. An adequate coordination of basic entities in space, time, speech, and social domains will help to reinforce the meaning of the discourse, as well as its learning efficiency. This layer contains higher level entities, such as scenes, video segments and rhetoric phases.
- **Layer 3: strategic.** How authors plan the elaboration of the instructional video. Video creators aim at fulfilling goals and purposes, most often related to learning outcomes. They also apply pedagogic principles, strategies and designs.
- **Layer 4: generic.** Basic and compositional entities are usually organized in recurring patterns that are recognized by a community of practice. These patterns are called *genres*. Particular genres can be identified in instructional videos.

5.3.2 Spaces of choices, spaces of constraints

Each domain offers its own set of entities to video creators. At the same time, it imposes some constraints. For example, the Spatiotemporal domain offers temporal segmentation resources such as scenes, sequences and clips. These resources cannot be used at will, because some constraints operate on them, such as the segmentation principle: video segments should be short, for better learning outcomes. The Social Appearance domain provides choices to set the instructor's gender or ethnicity, but these choices will influence the social response of the learner. Finally, the preferences of the social community would favor one presentation format over others, not only due to instructional efficiency or economic utility, but also due to social customs. This leads to apply the notion of 'genre' to instructional video.

In short: each domain is both a space of choices and a space of constraints.

5.3.3 The Physical layer: video medium

This lowest-level domain involves how the audiovisual data is recorded, stored and delivered. The Medium offers video creators a substrate of audiovisual devices in which basic semantic entities can be inserted. In its current form, the Medium provides a rectangular frame consisting of a two-dimensional array of pixels where moving images can be displayed, as well as a soundtrack. This medium offers some resources over which the video designer has a limited choice of manipulation: color palette, typography, sound volume, etc. The Medium will also provide a user interface where interaction entities can be placed.

5.3.4 The Basic entities layer

The Presentation domain

This domain offers structural items that convey meaningful contents: slides, screen captures, audio narration, one or more *actors* (visible agents), subtitles, sound, music and many other entities.

The Interaction domain

Interactivity is the ability of receiving external feedback from the user to alter the information flow. Nowadays, most digital videos show some degree of interactivity, since users can control the video playback with basic controls. In addition, there are higher levels of interactivity that can be enabled. In instructional videos, inserting forced pauses at some selected points is becoming widely used, so that the user stops watching and performs some task (reflects on the previously shown concepts, answers a given question, etc.). In some cases, the video can only resume playing if the user solves a simple in-video quiz that provides feedback about her understanding.

5.3.5 The Compositional layer

The compositional layer is the semiotic level in which the instructional message is articulated. The message is built by handling communicative features of space, time, language and social responses. Basic entities from the lower layer interrelate through all kinds of relations: spatial layout, temporal sequencing, rhetoric functions, and parasocial responses. This model will take into account two perspectives: a semiotic perspective (meaning making) and an instructional perspective (learning). The semiotic perspective tries to "identify the particular functional contributions made by the elements of a document to the intended communicative purposes of that document as a whole" (Bateman, 2008). The instructional perspective tries to identify learning principles that guide the design of an effective multimedia learning object, as described in the Cognitive Theory of Multimedia Learning (Mayer, 2014d).

This compositional layer contains three domains: the *Spatiotemporal* domain, the *Speech* domain and the *Social Appearance* domain.

Spatiotemporal domain

This domain describes how content pieces are arranged over space and time. This domain is mostly managed with methods from film production, such as the montage and the camera setting (Burch, 1970), resulting in the articulation of a complex spatio-temporal language (Metz, 1974). The effects in learning of spatial and temporal relations have been extensively researched, resulting in widely demonstrated learning principles such as the segmentation principle, the spatial and temporal contiguity principles and the redundancy principle (Mayer, 2014d).

Speech domain

The Speech domain describes how written and spoken language are articulated across the instructional video. Rhetoric is a fundamental feature of the instructional discourse, as the rhetoric has the goal of persuading the listener. The message should be articulated in a way that is understandable, and at the same time attractive. Systemic Functional Linguistics (Michael A. K. Halliday & Matthiessen, 2014) is used as a theoretical framework to describe the linguistic features that have been found relevant to achieve the learning and rhetoric goals of instructional communication.

Social Appearance domain

The social appearance characterization of an instructional video goes beyond a mere collection of noticeable attributes of the narrator/instructor. The social appearance is a complex construct in which the video designer projects a social image of the visible agent, a social image that in turn causes a response on the viewer. This response may affect processes such as learner's motivation (Baylor, 2011), the speaker's perceived credibility (John Baggaley et al., 1980) and the learner's involvement in the video discourse (Mayer, Sobko, & Mautone, 2003).

5.3.6 The Strategic layer

Video creators aim at fulfilling goals and purposes, most often related to learning outcomes. They also apply pedagogic principles, strategies and designs. The Strategic layer refers to those properties found in instructional video that reflect those goals, purposes and pedagogies. This layer contains (in this current unelaborated version) a single domain, which I call the Goals and Strategies domain.

The direct instructional purposes of videos may be to introduce concepts, demonstrate a procedure, show the operation of a process or a system, and many other purposes (Kay, 2012; Winslett, 2014). In addition, other communicative goals may be set, such as engage the audience (Schwartz & Hartman, 2007) and build reputation on the institution sponsoring the video.

To fulfill the communicative and instructional goals, several authors have proposed design strategies to make effective instructional videos: building confidence about the speaker, connect with pre-existing knowledge, create and fulfill learner's expectations, encouraging student participation, and building verisimilitude (see references in Chapter 2.4 and further discussion in Chapter 4). These strategies are usually deployed as spoken rhetoric structures, with the collaboration of other domains, such as the social appearance. Thus, these strategies condition how the entities in lower level are selected and combined. Strategies relate tightly to instructional design, as in the instructional strategies identified by Jonassen, Grabinger and Harris (1991). Communicative strategies that can be identified in instructional videos are segmenting content, cueing content, connect with previous knowledge, and activate the social presence, among others.

5.3.7 The Generic layer

I introduced in Chapter 2 the concept of *genre* in modern discourse analysis. Although there are several definitions for the *genre* concept, we can state that they are communicative events that pursue some communication goal and that exhibit structural features that are recognizable and used within a certain social community (James Robert Martin, 1994; Swales, 1990). Each genre has its own characterizing features, which can be linguistic, paralinguistic and contextual.

Genres can be viewed as "selected points" in the infinite space of possible configurations of semiotic resources. Not all technically feasible configurations are used in practice. In instructional videos, not every conceivable layout of presentation items is used. Instead, a few presentation formats are used recurrently, such as Powerpoint-like slideshows, short talking head presentations or handwritten tablet session captures, to name a few. This limited use of the expressive potential of videos may be driven by multiple factors: aesthetic, functional and cultural.

The use of recognizable genres in instructional videos has the advantage of creating an expectation in the audience about what is shown in the video, the steps that the presentation will follow and where the informational contents should be placed. These expectations, if fulfilled, can increase the efficiency of the communication (Chandler, 1997).

Genres in instructional videos

Several categories of genres can be identified in instructional videos. First, we have communication approaches such as the lecture, the demonstration and the interview, which can be regarded as proper instructional video genres. Second, we have presentation formats such as the screencast, the slideshow and the talking head, and many other formats listed in Chapter 2 (see Table 4-4). As suggested by the literature review made in Chapter 2, video genres may be characterized by five kinds of features:

a) **Communicative goals** (e.g. seeing, engaging, doing, saying).
b) **Type of recorded action** (e.g. lecture, conversation, demonstration, simulation...).
c) **Communication format** (e.g. fly on the wall, talking head...).
d) **Frame format** (e.g. text-overlay, slides with voiceover, picture-in-picture...)
e) **Mise-en-scène** (e.g. scenario, background).

It is important to note that the same content can be presented in different ways, with varying effects on learning outcomes. For example, in Chapter 6 I discuss a line of research that has identified benefits in a dialogic exposition rather than monologues (Chi, Kang, & Yaghmourian, 2017; Muldner, Lam, & Chi, 2014). The choice of using a monologic lecture, a tutorial dialogue or an interview will surely not be irrelevant in terms of learning.

It could be argued that the different genres are no more than the consequence of solving communicative needs raised by the learning content: lectures are well suited

to present conceptual knowledge, while demonstrations are better for practical and procedural knowledge. But there is something else: the same content, the same learning goal can be presented in different ways, but the choice of presentation format cannot be reduced to a utilitarian decision. Aesthetics and cultural traditions also play a role. For example, in Chapter 3 a correlation was found between the course scientific field and the choice of video layout. This correlation cannot be fully explained on a utilitarian basis: cultural factors may apply.

Other argument on video genres is that video lectures and tutorials are derivatives of classroom and face-to-face formats. It could be argued that most video styles are not proper genres, but mere recordings of offline events, or a reformatting of conventional offline genres. Nonetheless, the field evidence shows that video genres are something more. Crook and Schoefield (2017) show how video lectures have their own distinctive and characterizing features. Nordkvelle, Fritze and Haygsbakk (2010) point how "bringing the lecturer into the studio implies that most of those elements that make a lecture interesting or fun to follow are weakened" and how teachers "become actors performing in contexts viewers would most likely compare with media" (p. 65). What is more, rhetoric patterns in instructional videos differ from classroom events, as Chapter 6 will show.

Towards a taxonomy of video genres

As I have warned, this thesis will not develop a taxonomy of video genres. This Generic domain is introduced here as a blank canvas in which a specific research is needed. As a starting point, we have the inventories of video presentation formats and communication styles in section 2.4.6, and some findings such as the Speaker-Board centric spectrum described in Chapter 3. A first draft towards a taxonomy is Table 4-4, which summarizes the findings of Chapter 2. Among all the references, the fourfold classification of Schwartz and Kartman: *seeing*, *engaging*, *doing* and *saying*, may be a promising base for a taxonomy, since it links video presentation formats with high level communication goals.

One issue that hinders the development of a taxonomy of video genres is that there is no standardized nomenclature for video formats. New buzzwords are constantly emerging and going out of fashion. Consolidating a vocabulary on video genres would be a great help.

5.4 Domain-specific taxonomies

This section presents the domain-specific taxonomies that I have built for the lower level layers of the main taxonomy. In order to dimension a reasonable workload for my thesis, this dissertation has developed extensively layers 0, 1 and 2 (medium, presentation, interaction, filmic, speech and social appearance domains). The topmost layers (Strategic and Generic) have been characterized in this thesis, but the development of domain-specific taxonomies have been left apart of this study to keep a reasonable workload on the research project.

This section shows some tables that summarize the specific taxonomies elaborated for each domain. Chapter 6 will describe deeply the concepts under each taxonomy and the corresponding evidence from scientific literature.

5.4.1 Medium domain

The Medium domain is characterized by a small set of structural features: the video frame, the soundtrack, the overlays and user controls. The specific literature review has unveiled only three physical medium properties that have been considered to have influence in learning: the frame size, video quality and audio quality. All these entities and properties are summarized in Table 5-2, organized around the four structures that the modern video medium provides.

Table 5-2. Taxonomy of the Medium domain

structural class	definition	properties
video frame	visual representation of video	frame size, video quality
soundtrack	auditory representation of video	audio quality
overlay	item superimposed to video frame	---
user control	item that is actionable by the viewer	---

5.4.2 Presentation domain

The Presentation domain exhibits a rich variety of entities that are available to video creators to build content. The specific review of literature has found that presentation entities can be grouped around the axial concepts of **actors and boards**. Both concepts were revealed in the field study of Chapter 3. An actor is an agent with human qualities, real or virtual, visible or not, that actively provides content in the instructional video. A board is a surface on which instructional contents are presented. Common examples of boards are PowerPoint-like slides, captures of computer screen sessions and physical whiteboards.

The proposed taxonomy is built around the couple of actor and board. Many video characteristics studied by researchers can be labeled as one of these three classes: actor entities, board entities and interaction devices between an actor and a board. The rest of presentation entities are peripheral to the actor and the board, and are classified according to their instructional function: auxiliary instructional entities and non-instructional entities. This last class has been often considered to have a *negative* effect on learning, though this conclusion needs to be nuanced, as Chapter 6 will discuss.

Table 5-3. Taxonomy of the Presentation domain

class	entities
board entities	instructional text, diagram, picture, map, graph, sound, animation
actor entities	voice, face, gestures
actor-on-board interaction	handwriting/drawing, virtual pointer, deictic gesture
auxiliary instructional entities	acoustic signal, subtitle, closed caption
non-instructional entities	non-instructional text, visual decoration, background music

5.4.3 Interaction domain

The Interaction domain hosts a number of entities that enable a dialog between the viewer and the medium. After the literature review, four functional categories have been identified for the interaction devices: playback, navigation, system-user dialog and user commentary (see Table 5-4). The first two functions (playback and navigation) allow the user to *control* how the video is watched, while the other two functions (system-user dialog and user commentary) are vehicles to convey user *feedback* about the video.

Table 5-4. Taxonomy of the Interaction domain

class	subclass	entities
control	playback	basic playback panel playback speed control presentation control
	navigation	timeline (simple, enhanced) table of contents visual summary in-video hyperlink
feedback	system-user dialog	interpolated test system-generated pause
	user commentary	user-generated annotation

5.4.4 Spatiotemporal domain

The Spatiotemporal domain holds elaborate physical relationships between basic components of the video, which conform higher level entities. Many of these

constructs can be named using the conventional cinematography vocabulary. Four classes of characteristics have been identified: *spatial layout*, *temporal segmentation*, *linearity* and *informational complexity*. The experimental support for the effect of the collected characteristics in learning is variable. Some features such as the video segment length or the spatial contiguity have strong support, while others such as the cutting rate have not been explored in depth.

Table 5-5. Taxonomy of the Spatiotemporal domain

class	type	characteristics (entities and properties)
spatial layout	property	use of frame areas with semiotic relevance
	property	camera setting: angle, shot, perspective, zooming
temporal segmentation	entity	film segmentation hierarchy: shot, slide, scene, sequence, clip, hypervideo
	entity	segment transitions: pauses and temporal cues
	property	video segment length (duration)
linearity	property	linear vs. nonlinear
	entity	navigation graph (only for nonlinear videos)
informational complexity	property	presentation speed: words per minute, items per minute
	property	filmic complexity: cutting rate (shots per minute), continuity
	property	timing between informational events (temporal contiguity, redundancy)
	property	spatial contiguity of informational items

5.4.5 Speech domain

The Speech domain describes how written and spoken language are articulated across the instructional video. Two taxonomies describe this domain. The first taxonomy is shown in Table 5-6 and describes the entities that compose the discourse in the video. These entities are inspired in the catalogue of discursive structures made by Koumi (2006, 2015) and enriched by the field study described in Chapter 3 and the further literature review performed in Chapter 6.

The second taxonomy describes the discourse properties that are found (or suspected) to influence the learning outcome of the video. The list is shown in Table 5-7. This taxonomy is based on concepts of Systemic Functional Linguistic (language metafunctions) (Michael A. K. Halliday & Matthiessen, 2014), which I have found to fit perfectly as conceptual clusters of the research findings on instructional video discourse.

Table 5-6. Taxonomy of the Speech domain (entities)

class (rhetoric goal)	entities (rhetoric phases)
organize discourse	opening / closing shot overview of the contents explain pre-requisites and context relate to other contents announce following section rhetoric pause summarize contents
communicate content	theory / content demonstration / task execution example reformulation evaluation: indicate attitude evaluation: indicate commitment
query the learner	ask to recall/repeat exposed content ask to perform tasks ask for reflection and transfer
engage the learner	hook (capture attention) justify/motivate content build confidence/authority in speaker create and fulfill learner's expectations

Table 5-7. Taxonomy of the Speech domain (properties)

class (SFL metafunction)	subclass	properties
textual (mode)	spoken/written	spoken vs. written text
	action/reflection	spontaneous vs. acted speech
	interactivity	monologic vs. dialogic questions and prompts
interpersonal (tenor)	speech function	statement, question, offer, command
	social distance	conversational vs. formal style politeness humor
	personalization	personalization (addressing to 2nd person)
	standing	authority claims
	appraisal	attitude, engagement, graduation
	stance	modality: epistemic vs. deontic uncertain vs. confident narrator

5.4.6 Social Appearance domain

All the characteristics that have been identified in the Social Appearance domain are properties. Most of them are directly linked to the 'actor' entity. A small subset of properties refers to the global setting or atmosphere expressed in the video: these have been grouped using the cinematographic term *mise-en-scène*. The actor's social appearance properties are grouped in four classes: *realism*, *fluency*, *social distance* and *social group*.

Table 5-8. Taxonomy of the Social Appearance domain

class	properties
realism	voice: robotic vs. human picture: computer-generated, cartoon, natural
fluency	native vs. foreign accent, speech rate, speech fluency, visual addressing, gesture-speech synchronization
social distance	display: shot length language: personalization, formality, politeness
social group	gender, age, ethnicity, social affiliation, language register/dialect
mise-en-scène	social-cultural setting, spatiotemporal setting, scene atmosphere

Chapter 6. The taxonomy in detail

6.1 Chapter overview

In the previous chapter I have outlined the classification scheme for instructional video characteristics. This chapter develops the components of the classification scheme and discusses the evidences from the scientific literature that support the components of the taxonomies. This chapter will traverse the classification scheme across each one of the analyzed domains: Medium, Presentation, Interaction, Spatiotemporal, Speech and Social Appearance. For each domain, the domain-specific taxonomies are explained.

This chapter ends with a discussion on the findings, with an emphasis on the 'missing areas' of the scientific literature: features of instructional videos that are underexplored by researchers.

6.2 The Medium domain

The medium is the physical substrate that is used to build the informational contents of the instructional video. The medium is both an instrument to express the design ideas of the author, and a constraining force that limits the expressive modalities. In this section I will introduce the structure of the streamed video medium that is used for instructional purposes.

From a semiotic perspective, "a semiotic mode is developed by virtue of the work that a group of users puts into using some material substrate as a tool for constructing meaning" (Bateman, 2011). Bateman's definition of semiotic mode makes clear that each mode is born on community of practice rather than in the sensorial qualities of the representation: photography and painting are both visual artefacts, but they are socially and technically acknowledged as different modes. In this sense, films and, more specifically, instructional videos, are a distinct semiotic mode. A semiotic mode requires a *medium*, that is, a material substrate to carry the semiotic resources (Hiippala, 2014). Each type of medium defines its own set of semiotic resources (pages, links, paragraphs, audio narration, video sequences, etc.) that can be selected by authors to satisfy their communicative goals, and thereby build an instance of a particular semiotic mode (in our case, a particular genre of instructional videos).

From this perspective, we can decompose this semiotic material substrate in two layers: one physical layer which a) provides the representational infrastructure where meaningful entities can be displayed; b) provides devices for user interaction on the artefact; and one higher lever layer consisting of the multiple representational entities that can be allocated and placed in the underlying physical layer. In this taxonomy, this lower-level, physical layer is what we name "the medium" in the present study.

6.2.1 Structure of the digital video medium

In this section I list some characteristics of the video medium that according to the reviewed literature have been studied regarding their possible influence on instruction and learning.

Basic entities

The video medium provides audiovisual information streams. The visual stream is characterized as *framed* and *two-dimensional*. The core entity in the Medium dimension is the **video frame**, which is a continuous surface, usually rectangular, that contains the visual information flow, and where the course of action is shown. A companion of the video frame is the **soundtrack**, which being an optional item, it is seldom omitted in instructional videos.

The frame and the soundtrack both host a collection of presentational entities: written text, animations, speech, sound and music, and filmic footage. The most relevant presentational entities will be discussed in Section 0(the Presentation domain).

In addition to the video frame and the soundtrack, a video may show **overlays** that are optional visual items added on top of the video frame or in surrounding areas. A common type of overlay is a text box showing additional commentary. Overlay visualization is often user-controlled (user can switch off overlay visualization). The summary of basic medium entities is shown in Table 6-1.

The structure of streaming video medium cannot be limited to a mere representational device. Streamed video supports some forms of direct user interaction, for example basic playback buttons, or more advanced artefacts such as tables of contents or in-video quizzes. Those **user controls** are also basic structures of the video medium and are crucial to understand how instructional videos contribute to the learning process.

Table 6-1. Entities of the Medium domain

entity	definition
video frame	visual representation of video
soundtrack	auditory representation of video
overlay	item optionally superimposed to video frame
user control	item that is actionable by the viewer

The extended video medium

The medium basic entities are spatially organized in a higher-level structure. Figure 6-1 shows the sketch of a common layout for an instructional video user interface.

We can distinguish three areas: a *video frame* area, a *video interaction* area, and a *video learning environment* area. The video interaction area usually covers the entire video frame and contains all the basic user interface controls, such as playback, timeline, etc. The video learning environment contains the frame and the interaction areas, as well as the rest of artifacts that compose the full product with which the learner interacts. A video learning environment may be a web page, an LMS, or any other type of digital platform.

This classification scheme will cover the inner structures of this macro-organization: the video frame and the video interaction area.

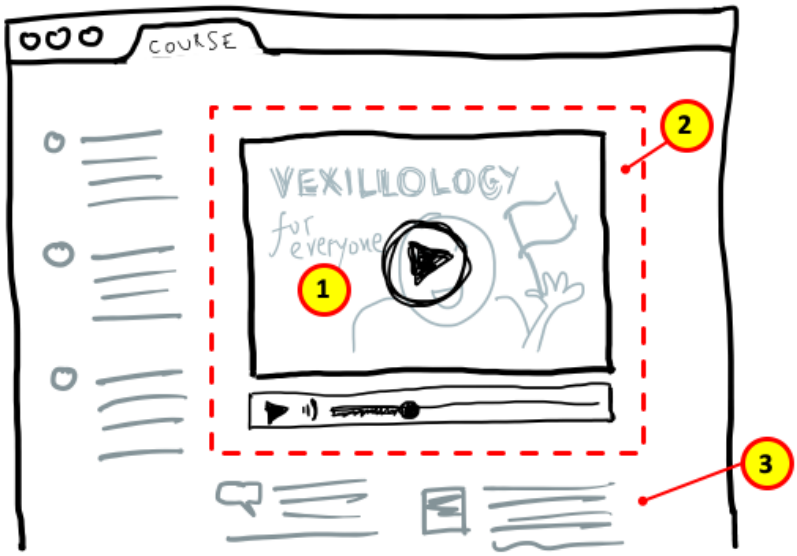


Figure 6-1. The enhanced Medium structure
1: Video frame. 2: Video interaction area. 3: Video learning environment area.

A note on video annotations

Many modern MOOC and video distribution platforms support video annotations. For that reason, annotations should be included as a basic medium constituent. Video annotations are "information pieces that can be anchored in the temporality of the video so as to sustain various processes ranging from active reading to rich media editing" (Aubert et al., 2014). Annotations can take many modalities: text, pictures, animations, or other types of multimedia objects (Seidel, 2015). Annotations are commonly attached to temporal points. There exists technology to bind annotations to moving objects (Goldman, Gonterman, Curless, Salesin, & Seitz, 2008), but this study has not found reported applications of this kind of annotation in instructional video. (Aubert et al., 2014) specify four use cases for video annotations in learning: active reading, live lecture annotation, performance annotation, and annotation for assignment.

Annotations can be rendered in many ways. A common representation is as overlays: in that case, annotation display can be switched off by the user. They may be

displayed on top of the video frame or in other locations in the extended video medium structure.

For the purpose of this study, I will restrict Aubert's et al. definition of video annotation to a piece of information that is: a) anchored to a temporality of the video; b) distinguishable and separable from the video frame contents. This classification scheme will assign annotations to three structural areas: presentation, navigation and user feedback (see Table 6-2). Navigational and feedback annotations belong to the Interaction domain (section 0).

Table 6-2. Types of video annotations

class	entities
presentation	subtitles, overlaid enrichment to basic contents
navigation	annotated timeline, hyperlinks
feedback	interpolated tests, user comments

6.2.2 Medium properties

Instruction is influenced by the medium. Characteristics such as the physical qualities of the representation (e.g. printed, displayed on a digital screen) and the delivery mode (streamed vs. broadcast) shape different affordances and expected responses on learners (Fahy, 2004; Shephard, Ottewill, Phillips, & Collier, 2003).

Moreover, video has some particularities over other media: the information shown in a video is more transient than that in printed text, thus it demands more cognitive resources to the learner, such as attention and short-term memory (Kozma, 1991). The degree of interactivity of the streamed video is higher than that of a broadcast television, but lower than that of an interactive computer simulation. De Vaney (1991, p. 253), when discussing semiotic codes in educational television, observed that grammar codes in certain Hollywood films had adapted various codes from large screens to home video screens: shot ranges and increase of physical action.

Although all streamed videos share the same general attributes, there are still differences at the medium level that are secondary, but they can still condition the instruction. The clearest and most documented example in literature is screen size.

The iconic visual representation of films is strongly *framed*: it adopts a particular fixed size and shape, usually rectangular (Bateman & Schmidt, 2012, p. 137). In addition to, due to digital video technology, video designers are constrained to a highly limited set of frame aspect ratios. I have not found studies that explore the geometry of the video frame (shape or aspect ratio), but there is some literature on the effects of the frame size in the viewer, as I will discuss below.

Global medium properties may be manipulated by film author for expressive reasons: for example, a segment may be shown in black and white to mean a flashback in the

story; a lower quality in audio or video may be deliberately used to mean that it is unedited, spontaneous footage, etc. This expressive device can be used in instruction to provide social or emotional cues that influence learning. I will discuss some of these manipulations when introducing higher level domains of this classification scheme.

Screen size

The physical characteristics of the visualization device will produce different experiences in the viewer. Compare, for example, watching a movie in a small smartphone screen with the immersive experience of watching the same movie with a virtual reality headset. In educational settings, the recent boom of mobility and handheld devices has caused that many learners watch instructional videos with small screen sizes. Also, handheld devices have user interfaces with ergonomic qualities that differ of those of computer monitors or classroom projectors. In particular, those small screens challenge the experience of learners and, consequently, the criteria for designing streamed instructional videos.

Research provides evidence that larger screen sizes amplify the psychological viewer responses. A remarkable experiment on this matter was conducted by Reeves et al. (Reeves, Lang, Kim, & Tatar, 1999). They compared three screen heights: 2, 13, and 56 inches. The largest screen stimulated more valence and arousal in viewers. Additionally, there is some support for the hypothesis that larger screen size also raises attention and memory (Grabe, Lombard, Reich, Bracken, & Ditton, 1999).

The effect of screen size in learning has been investigated for a long time. In 1992, Bruijn et al. (Bruijn, Mul, & Oostendorp, 1992) tested the effect of screen sizes in CRT computer monitors. They compared learning retention from 12-inch and 15-inch displays, resulting in no significant differences in retention, but some size effect in learning time. Maniar et al. (Maniar, Bennett, Hand, & Allan, 2008) performed a set of experiments on various modalities of tutorials (video, audio, text) with three screen sizes. Video instruction was superior to audio-only and text; the smallest screen size (1.65 inches) produced lower learning retention.

Audio and video quality

The consensus of scholars and practitioners is that audio and video quality affects learning, particularly when the noise-to-signal ratio is high enough to penalize perception. Above some undetermined level, the learning experience should not be significantly affected. It is difficult to delimit what are the satisfying levels of medium quality, since there is a high individual variability in learners, depending of factors like age, physiology, and even cultural background (Jaimes & Dimitrova, 2006). On the other hand, a study about conceptual modeling learning objects found low correlation between multimedia quality and perceived quality of the product (Churchill, 2014), but found high correlation between perceived quality and other basic features, such as variety of fonts and colors. An assessment of YouTube video tutorials (Morain & Swarts, 2012) found that poorly user-rated videos usually had problems in audio track quality. Morain and Swarts' assessment rubric for video

tutorials includes a 'viewability' objective that is achieved when audio and video quality are noise-free and high quality.

There are some experiments that have explored the relation between audio/video quality and efficiency in instructional videos. In a first study, higher screen resolutions have been found to have positive effects in vocabulary learning (D. Kim & Kim, 2012). P. ten Hove and H. van der Meij have studied the correlation between popularity and video characteristics of YouTube instructional videos (P. E. ten Hove, 2014; P. ten Hove & van der Meij, 2015). They have found a strong correlation between image quality (screen resolution) and popularity: 56% of unpopular videos use a very low resolution (less than 480 pixels height), while 64% of popular videos use a high resolution (HD, 1080 pixels high). HD resolution is used by 8% of unpopular and average videos. What is more, 33% of unpopular videos and 34% of average popularity videos suffer from noise, which is present in only 4% of popular videos. This observation supports the notion that audio and video quality affects viewer experience and preferences. Similar conclusions were obtained by Morain and Swart's assessment of YouTube tutorials (Morain & Swarts, 2012).

A multitasking medium

Another important issue is that many students watch videos in devices where multitasking is omnipresent: multiple concurrent applications, social networking notifications, etc. Exposure to multitasking can negatively influence the learning process, or at least influence how this process develops (Subrahmanyam et al., 2013). This has been a concern, especially in the learning of the younger, who are more accustomed to multitasking (Van Der Schuur, Baumgartner, Sumter, & Valkenburg, 2015). As far as this study has found, this is an open question that is still being investigated.

6.2.3 Practical implications for instructional design

The distinctive characteristics of video medium have a strong influence in the selection of presentational entities and their internal attributes. The transient nature of displayed video information (as compared with printed text) compel to a moderation in information density of texts, and to minimize distracting or redundant visuals, as stated by multimedia principles (redundancy principle).

The limitations in screen size of handheld devices has led to Some instructional recommendations have been proposed to overcome the limitations of handheld devices. I will mention two sound examples on this topic. First, Churchill and Hedberg (2008) study suggested: full screen presentation, landscape presentation, minimize scrolling, design for short contacts and tasks centeredness, one step interaction, zooming facility. Second, Churchill's study on presentation design in learning objects (Churchill, 2014) came up with this a set of recommendations that include: design for a single screen, design for small space, use color in moderation, avoid unnecessary decorative elements, design with a single font, and use frames to logically divide the screen area. These two studies refer to other authors that have

come to similar conclusions. It seems that the academic consensus is that multimedia learning objects, and instructional videos in particular, must follow simplicity at all levels: multiplicity of entities, informational density and physical attributes of entities such as text and graphics.

6.3 The Presentation domain

6.3.1 Presentation domain structure

The video frame is built up on some basic components, such as a box where text or pictures are displayed, an area where a narrator speaks, onscreen subtitles, etc. There may also exist audio components, like a voice narration or background music. Most of these items provide information directly related to the learning goal, while other function as navigational items, signals to interesting information, and other purposes.

The semiotic division of labor between word and image is contextually specific (van Leeuwen, 2008, p. 136). Some videos will use images as a supplement or support for the main spoken exposition, while other videos will use images as the principal vehicle of the exposition, and words become the supplement.

Actors and boards

Based on my field study on MOOC videos (see Chapter 3), two representational entities stand out in instructional videos, particularly on video lectures: the *actor* and the *board*. Both are the principal content displayers in many video lectures and tutorials.

The **actor** is defined as an agent with human qualities, real or virtual, visible or not, that actively provides content. Sometimes more than one actor may be present in the frame, or multiple actors may appear during the video presentation.

The **board** is defined as a rectangle-shaped surface on which instructional contents are presented. This surface, when present, usually covers a large portion of the video frame. Common examples of boards are slides, captures of computer screen sessions, and physical whiteboards.

The physical properties of the actor and the board in the video layout, their mutual relationships, and their presence or absence, are key features to characterize every genre of instructional video. In fact, this study has found that many video presentation styles can be defined in terms of *board-centric* and *actor-centric* classes (Santos-Espino et al., 2016), as I have discussed in Chapter 3.

In their most basic presentation formats, boards are visual entities (slides, computer screens, etc.), while actors are auditory entities (a voice narration). Of course, a board can show sounds, and an actor can be visible in the frame, but these attributes are usually considered as *additions* to the default setting. Their essential, basic nature is summarized as 'the board is seen, the actor is heard', a property that many video

styles satisfy. The only exception is the model-based demonstration, where the model may be shown as a video recording without speech.

Signals and cues

Signaling provide support for the learner's cognitive process. Auditory and visual cues constitute a vehicle for guidance to the learner, by directing the learner's attention to relevant information (Moreno & Mayer, 2005). Signals can be shown as visual items (pointers, arrows, highlighted areas) that focus attention on some specific area of the display. Signals can be oral or written prompts (e.g. self-explanation prompts) aimed at triggering a reaction on the learner. Almost any type of shape or action can be used as a signaling device in an instructional video. There are some cultural conventions in visual signals, as it is the case of the various 'mouse pointers', often arrow-shaped. Even the instructor's picture has been proposed as a signaling entity: in (René F. Kizilcec et al., 2014), slide-based lectures are enhanced with the presence of a picture-in-picture instructor, which appears intermittently only in specific moments when the learner has to pay direct attention to the slides (instead of the speech).

The effect of signaling or cueing in learning has been studied for a long time in narrated animations. Typically, a narrated animation consists of an animated illustration and a narration. Mautone and Mayer (2001) used a mix of signals in their experiments on signaling effects: colored arrows pointing to relevant areas of the illustration, color coding of illustration components and relationships, summary icons, and speech cues (changing in intonation). Ozcelik et al. (2010) verified that learning performance was increased if the terminological labels in the illustration were signaled (changing color to red) in synchrony when the narration mentioned each term. Visual cues can be used as an artifact for procedural scaffolding (P. Sharma & Hannafin, 2007), providing guidance to use correctly the instructional video: point to interesting areas in the video frame, raise attention when something important is shown, etc. However, visual cues may not affect learners' performance in animations with low complexity (de Koning, Tabbers, Rikers, & Paas, 2011). Visual and textual cues are well studied by multimedia learning scholars, and their conclusions are valid for instructional video in general.

This classification scheme considers signaling as a potential *function* of each and every presentational entity (actors, boards and the rest). Any entity may work as a signal or cue. Notwithstanding, there are a few items that are used primarily for signaling. One class is that of actor-on-board interaction entities such as real-time handwriting, visual pointers and deictic gestures. The other class are the acoustic signals that some videos use. I will introduce both classes later in this chapter.

Taxonomy of presentational entities

Once that actors and boards have been introduced, the taxonomy of entities in the Presentation Domain can be described. Actors and boards are also containers of other lower-level entities, which are shown in Table 6-3. Apart from actor- and board-specific entities, there is a number of entities shared by the actor and the

board. These are representations of actions that the actor performs and are visualized in the board, such as the mouse movements shown as a moving pointer on a computer screen, or actor's drawings on the board. Other entities provide auxiliary instructional content that is not tied to actors and boards (acoustic signals, subtitles and closed captions). Finally, there is a class of non-instructional entities which serve other purposes than instruction, such as to identify authors or participants, declare intellectual rights, date the video, or just for decoration.

Next sections will describe each relevant entity of this taxonomy, together with their respective research evidences found in my review.

Table 6-3. Taxonomy of entities in the Presentation domain

class	entities
board entities	instructional text, diagrams, pictures, maps, graphs, sound, animations
actor entities	voice, face, body language (gestures, pose)
actor-on-board interaction	handwriting/drawing, virtual pointer, deictic gesture
auxiliary instructional entities	acoustic signals, subtitles, closed captions
non-instructional entities	non-instructional text, visual decoration, background music

Characterization of actors

From the perspective of a characterization of the learning effects, the actor is a rather complex entity. The actor does not just convey direct instructional, verbal speech. The actor's representation also transports a lot of nonverbal speech and social cues: gestures, gaze, intonation, accent and garment, just to mention some. This diversity of nonverbal inputs reflects the essential multimodal nature of human language, as described in Poyatos' basic triple structure: language, paralanguage and kinesics (Poyatos, 1983). Those actor's nonverbal features have effects in learner's cognitive load, attention, engagement, and many other learner's internal processes.

An actor may operate with different *roles* in the video. It may be an instructor, a tutor, a model, an expert being interviewed, or a combination of those. If the actor is not visible, it is usually called a *narrator*. Table 6-4 shows common terms used in different instructional video genres for these roles.

Table 6-4. Common terms used to designate actors

terms	video genres	roles
instructor, teacher	lecture, tutorial	a person that shows and narrates the contents
tutor, tutee	tutorial	in a dialogic tutorial, a person who explains and guides (tutor) or learns (tutee)
interviewer, interviewee, expert	interview	a person that is interviewed (interviewee, expert) or asks questions (interviewer)
model	demonstration, interview	a person that shows how a procedure is executed or explains a testimonial
narrator, presenter	documentary, demonstration	a non-diegetic voice that describes the contents

In the literature review performed for this study, a large amount of research has been found that is focused on actor's attributes, their relationships with other entities in the video structure and their influence on learning. Covered topics are the following:

- Speech quality
- Visibility
- Relative size in video frame
- Face
- Gestures
- Gaze (showing eye movements onscreen)
- Handwriting / drawing on the board
- Social appearance: humanity, gender, age, ethnicity
- Multiplicity (e.g. in dialogic lecturing)

Instructional video designers are able to decide on which concrete actor's attributes will be shown in the video. In accordance with this composite internal structure, this classification scheme will regard the actor as a composite entity that contains several semi-autonomous sub-entities. These sub-entities are: voice, face, gestures, handwriting/drawing and gaze. Actor sub-entities are discussed in the following sections of this chapter.

Other actor's attributes will be considered in this classification scheme as properties of the actor entity: the *social appearance* and the *multiplicity*.

Social appearance. Attributes such as the humanity, gender, age and ethnicity form a construct that provides a social appearance of the actor, which in turn raises a response in the viewer (Gunawardena & Zittle, 1997). This construct deserves a full domain in this taxonomy (the Social Appearance Domain), which is discussed later in this chapter (see Section 6.7).

Multiplicity. An educational video can show multiple actors, either simultaneous or separated in time. We can find *concurrent* actors in dialogic interviews and dialogued lectures (Chi et al., 2017; Santos-Espino et al., 2016). A special case of concurrent actors occurs in demonstrations in which a visible model performs a procedure, while an audio narration explains it. The model and the narrator may be different persons. This case will be called 'modal concurrency'. A common case of *sequential* multiple actors is a video lecture consisting in several segments in which each segment is presented by a different instructor. The topic of multiple actors will be discussed in detail in section 6.6 (The Speech domain), where the rhetoric features of instructional videos are characterized.

Table 6-5. Actor entities and properties

type	class	entities/properties
entities	actor structure	voice face gestures
	actor on board	pointer movement handwriting/drawing gaze
properties	social appearance	humanity gender age ethnicity social affiliation
	multiplicity	single multiple sequential multiple concurrent modal concurrent

Characterization of boards

A board can contain whatever visual representation that can be conceived: diagrams, images, maps, compositional diagrams, graphs, and conceptual diagrams (Bernsen, 1997). The examination of hundreds of videos during this investigation has revealed that boards can be characterized by means of three dimensions: a *writing/non-writing* dimension, and a *physical/virtual* dimension and a *static/dynamic* dimension. Table 6-6. Characterization of boardsshows some examples of the two first dimensions. A *writing* board shows drawings or text being written or drawn by hand, usually by the actor. A *non-writing* board shows information that is not perceived as drawn by hand. Examples of writing boards are chalkboards, whiteboards or paper sheets. Examples of non-writing boards are PowerPoint slides or computer screen captures. *Physical* writing boards as chalkboards, paper sheets or lightboards contrast with *virtual* boards as Khan-style blackboards, the canvas of a paint application, etc.

Many video lectures make use of *static* boards that show still pictures, screenshots and text. The static representation changes from time to time, as in a slideshow. *Dynamic* boards show animated content, such as an inline video clip, an animation or a computer screen session. The use of static or dynamic boards shapes the filmic montage of the instructional video, as I will discuss in Section 6.5 (The Spatiotemporal Domain).

Table 6-6. Characterization of boards

	writing boards	**non-writing boards**
physical boards	vertical writing board paper sheet lightboard	physical printed card PowerPoint-like slide
virtual boards	virtual canvas drawable slide	computer screen capture

6.3.2 Actor entities

Voice

The voice is an essential component of most instructional videos. The cognitive theories of multimedia learning claim that learning efficiency increases as content is presented using multiple modalities, for example, visual diagrams and audio narration (multimedia principle). Furthermore, it has been found that it is better to show images with audio narration, rather than images with printed text, or simultaneous audio and text (modality and redundancy principles). Those evidences from multimedia learning research grant the voice a paramount position in digital instructional products.

For the present study, it is important to stress the difference between 'voice' and 'speech'. The voice is the auditory form of speech. And the latter may be shown in different modalities: as printed text, as auditory utterances, or even as visual gestures. The speech as an organized informational flow will be discussed later in Section 6.5 .

The voice that is shown in an instructional video may be **diegetic** or **non-diegetic**. A diegetic voice is that which is perceived as happening in the course of filmed action. This is the case of the instructor's voice in a recorded lecture. A non-diegetic voice is that perceived as occurring externally to the filmed action, thus, not linked to any visible actor in the video. This is the case of a narration in a documentary or in a narrated animation.

The voice as an auditory expression of instructional content and its influence in learning have been explored. Several properties have been tested for influences in the learner's experience:

- Presence or abscense of voice
- Redundancy of voice and text
- Speech rate
- Social attributes: humanity, accent, etc.

Some studies have compared the relative learning efficiency of adding voice to video presentations. For example, (Mohamad Ali, Samsudin, Hassan, & Sidek, 2011) found that narrated screencast was more effective than silent screencast in learning performance. The modality principle of CTML says that people learn better from graphics and narration than from graphics and printed text. The practical application of modality principle is to present words as audio narration, rather than on-screen text (R. C. Clark & Mayer, 2008).

Voice effect. As regards the humanity, the voice principle of CTML declares that people learn better when the words are spoken in a human voice rather in a machine-generated voice (Atkinson, Mayer, & Merrill, 2005). The effect size is reportedly strong (d = .74) (Mayer, 2014d, p. 345), though state-of-the-art synthetic voices are very close to human speech and the voice effect would be reformulated. I will discuss the voice effect in the Social Appearance Domain section (page 188).

Speech rate. There are many studies on the effect of speech rate in educational settings. This characteristic is discussed in detail in Section 6.5.7 of this thesis (see page 169). Time-Scale Modifications and playback speed controls allow learners to adjust the speech rate at will, therefore adjusting to their listening preferences and capabilities. The ability for learner adjustment of video speech rate lowers the relevance of this feature in a modern characterization of instructional video. Playback speed controls are discussed in Section 6.4.3 of this dissertation (see page 148).

Showing the actor: helping or harmful?

The basic, simplest arrangement for a video lecture or tutorial is to show the speaker's voice together with a dynamic visualization (an animation, a live recording, a slideshow, a computer session capture, or the like). What would be the effect of adding a picture of the instructor's face or body? Is it neutral, beneficial, or harmful for learning?

Before discussing possible answers to these questions, a key fact must be taken into account: instructor's presence in video entices learner's attention. Eye-tracking heat maps made with alternating instructor absence and presence clearly demonstrate that learners pay visual attention to instructor's depiction (René F. Kizilcec et al., 2014) (see Figure 6-2). That study reported that spent 41% of the time looking at the instructor's face. Other eye-tracking studies also report high visual attention to instructor's face (Zhongling Pi & Hong, 2016; J. Wang & Antonenko, 2017).

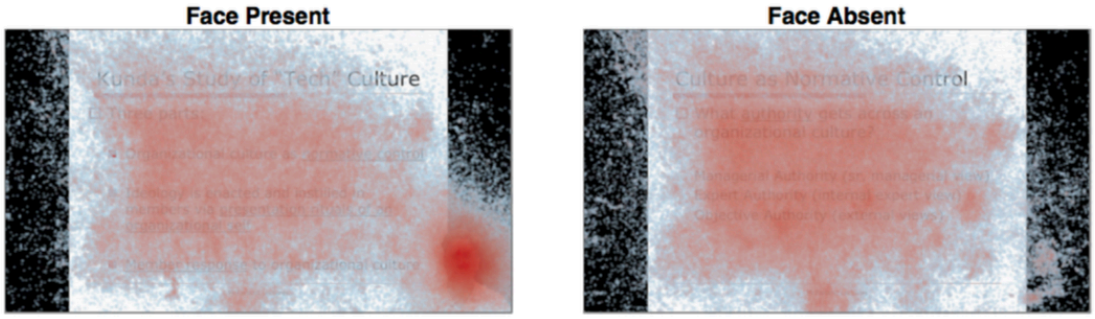


Figure 6-2. Heat map of visual attention to instructor's face (from Kizilcec et al., 2014)

What has to be ascertained is the instructor effect in learning. There are contending arguments about adding the actor's picture in the video frame, rooted on social, psycholinguistic and cognitive theories. The intuition that showing human presence in mediated communication fosters user engagement has evolved to the formulation of the Social Presence Theory (Gunawardena & Zittle, 1997). According to this theory, the presence of a pedagogical agent in a multimedia learning object will promote learning. What is more, the asynchronous nature of online learning triggers "the issue of how teacher immediacy behaviors and social presence are affected by the environmental transformation" (Richardson & Swan, 2003). In that context, showing the instructor's face in lecture videos may contribute to a teacher's higher social presence.

In the beginnings of the era of multimedia learning environments, there was an intuition that adding an animated instructional agent would result in better learning (Rist, André, & Müller, 1997). The research on multimedia learning has found strong evidence of an **embodiment principle**: people learn better when on-screen agents display humanlike gestures, movements, eye contact and facial expression (Mayer, 2014d, p. 345). Together with the voice principle and the image principle, experimental evidence from CTML suggests that adding a socially-balanced video recording of the human instructor may enhance learning. More generally, a meta-analysis (Schroeder, Adesope, & Gilbert, 2013) concluded that the presence of a pedagogical agent in multimedia learning objects has a significant, but small, positive effect on learning. Analogously, Amy L. Baylor, a prominent researcher on computer-generated agents, states: "while the addition of an anthropomorphic interface agent to a learning system generally has little direct impact on learning, it potentially has a huge impact on learner motivation" (Baylor, 2011).

Finally, from a psycholinguistic perspective, the instructor's face may contain valuable information for enhancing the understanding of the narration. The instructor's face may help in lip reading, and instructor's body language may reinforce some meanings of the speech (importance, judgement, etc.). The mutual reinforcement between visual and auditory streams is well known in psychology, with notorious examples such as the McGurk effect (McGurk & MacDonald, 1976).

Notwithstanding, the instructor's image is an additional visual stimulus that has to be processed by the learner, and much of the information portrayed by the instructor may be irrelevant or redundant. The extraneous cognitive load produced by the instructor's image may overload the visual processing channel, and as a consequence deter learning. This would be particularly harmful because of the learner's natural tendency to be attracted to the human picture. This question has been examined by multimedia learning researchers. Frechette and Moreno (2010) examined how animated pedagogical agents affect students' perception and learning. Students were exposed to one of four possible conditions: static agent; agent with deictic movements; agent with facial expressions; agent with both deictic movements and facial expressions. They found that the presence of the agent did not have an effect on learning. Moreover, the only significant difference was that the static agent was preferred over that with facial expressions.

There is evidence that simple screen layouts provide enough learning performance. For example, Lents and Cifuentes (2009) examined the introduction of simple PowerPoint with voice over lectures for undergraduate students. The authors found that video-based instruction (without visible instructors) produced similar outcomes than face-to-face instruction in student preparation. If videos without the instructor work, why take the risk of adding it?

The arguments and evidences that have been introduced in this section lead to the need of experimental research on the effect of the instructor's presence in video lectures and tutorials, that takes an account of concrete learners' behavior and learning outcomes, and how the contending principles (social presence vs. redundancy) balance together.

Actor's face

One of the first experimental studies on instructor presence in lecture videos was (Homer, Plass, & Blake, 2008). The researchers made two laboratory experiments where two versions of a recorded lecture were compared: one included a video of the lecture plus synchronized slides; other included the slides and the audio narration (video was omitted). Learning, cognitive load and social presence were assessed. The results found no significant differences for learning and social presence, and some effect in cognitive load. One of the experiments revealed that the effect on cognitive load varied on visual/verbal learning preference: the full-video demanded more cognitive load for low visual-preference students, and vice versa.

In other study (Lyons, Reysen, & Pierce, 2012), participants were split in two groups. Each group viewed a slightly different version of the same video lectures, one including the instructor's face and other without it. Self-reports were gathered on perceived effects of video: in learning, usefulness, interactivity, and comfort. Technological efficacy of students was also assessed. The results indicated that adding the instructor's image affected negatively to perceived learning and interactivity, and this effect was higher in students with low technological efficacy. This finding is in consonance with the redundancy effect: for students with lower

competence, the instructor's image may drain cognitive resources with extraneous load, preventing potential positive effects from social presence.

A set of experiments designed by Kizilcec and colleagues (2015; 2014) found that learning outcomes were not different across instructor or instructorless video lectures. They tried the approach of strategically presenting the instructor only when direct attention would be demanded on the visual slides, and they found again that it did not result in an increase of learning outcome, or variations over social presence effects compared to a constant visualization of the instructor. More recently, Wang & Antonenko (2017) compared instructor's presence in mathematics lectures. Two lectures were designed with varying topic difficulty. They found no instructor effect in learning transfer. Learners paid higher visual attention to the instructor in the easy topic lecture.

The presence of the instructor's face has been investigated in procedural knowledge videos, also with mixed results. One study (van Gog, Verveer, & Verveer, 2014) found that the presence of the instructor's face raised learning performance of a puzzle-solving task, but a subsequent replication (van Wermeskerken & van Gog, 2017) measured no effect in learning performance, despite of the face capturing learners' attention. Hong, Pi, & Yang (2016) compared the instructor effect in videos showing different types of knowledge, and found that adding a visible instructor increased learner's cognitive load only when learning procedural knowledge, but not declarative knowledge.

Other important finding is that a great proportion of learners seem to be satisfied with instructorless videos. In the study by Kizilcec and collaborators (2015) on a MOOC, 35% of learners preferred to watch videos *without* the instructor face. This fact must be considered to balance the importance of social presence effect with respect to redundancy effect.

An additional attribute to be pondered is the **instructor's eye orientation**. The perceived position of the instructor (e.g. frontal vs. lateral) may induce different parasocial behaviors in the learner. A frontal orientation, with instructor's eyes looking front-to-front to the learner, work as an engaging social stimulus that, in turn, can fosters learning. This hypothesis has been tested by Beege et al. (Beege, Schneider, Nebel, & Rey, 2017). They manipulated two instructor attributes: proximity (near vs. far) and body-eye orientation (frontal vs. lateral). Results found a strong orientation effect in learning, with higher outcomes for frontal orientation. No significant effect was found for proximity. The experiment used real-life recordings from human instructors.

Actor's gestures

Gestures are a key element in human intercommunication. They provide informational contents, add emotional qualities to speech, and they even can constitute a full-blown communication system, as signed languages are. Consequently, gestures play an important role in teaching and learning. In words of Pozzer-Ardhengi & Roth (2007), teaching "involves not only the words and

sentences a teacher utters and writes on the board during a lesson, but also all the hands/arms gestures, body movements, and facial expressions a teacher *performs* in the classroom". The teacher can use gestures to point to important elements in a graphical presentation, to emphasize some key point of the speech, or even to make a visual representation of a concept (e.g. mimicking the direction of a moving force, or hand drawing the shape of the concept being explained).

McNeil's taxonomy of gestures (McNeill, 1992) classify gesture movements into four major categories: *beats* or 'batons' provide no topical content and emphasize communication utterances; *deictic* gestures are concrete or abstract pointing; *iconic* gestures represent a concrete entity or event; *metaphoric* gestures reference to an abstract content. A fifth additional category is that of *cohesive* gestures, composite gestures that signal continuities in thematically related but temporally separated discourse.

According to Pozzer-Ardhengi & Roth (2007), gestures contribute to the integration of all the other resources that build the meaning unit being taught in a lecture. Neurophysiological studies evidence how gestures help to integrate the multiple sources of meaningful information such as words, pictures and concepts (Kelly, Creigh, & Bartolotti, 2010; Wu & Coulson, 2007). Gestures also provide their own contents, particularly the representational gestures (iconic and metaphoric): they represent diagrammatic elements, physical objects, events, etc. (Coskun & Acartürk, 2015). The review from de Koning and Tabbers (2011) show copious evidence of the improvements in learning afforded by involving human actions, gestures included. Among other findings, they suggest that "observing gestures performed by another agent aids understanding and facilitates learning".

According to Roth (2001), in lectures and other teacher presentations, iconic gestures may be more easily suited than oral language, since they are encoded as images and they do not require a translation. Roth also argues that gestures "can orient students to aspects of a visual representation that the lecture can point to and highlight by tracing", therefore acting as a learning scaffolding device. All these considerations are in line with the signaling principle of CTML.

Embodiment principle. The effect of gestures is closely linked to the embodiment principle of CTML. This principle remarks that human-like gestures activate a social response in the learner that may foster learning (Mayer, 2014d, Chapter 14). This means that instructor's gestures have an additional benefit over mouse pointer and other virtual signaling cues. The effect of gestures and body language has deep roots in human biology. It is argued that observing an animation showing a human performing physical action triggers the mirror neuron system that is involved in executing the same action (van Gog, Paas, Marcus, Ayres, & Sweller, 2009).

De Koning and Tabbers (2011) propose four strategies related to human movements and learning from animations: let the learner follow the movements using gestures, make the learner manipulate the movements through interaction, embody the movements, and stimulate learners to reconstruct the perceptual processing of the movements.

Showing hands. (Castro-Alonso, Ayres, & Paas, 2014) found that the presence of hands improved the effectiveness of static presentations, but decreased effectiveness in animations. (D. S. Cheng et al., 2014) extracted several actor characteristics from video lectures in the *Videolectures*[11] repository and correlated them with video popularity. Among other results, they found that amplitude of gestures tends to be higher for less valued lectures.

Gesture annotation in video lectures. Zhang (2012) developed a specific taxonomy of gestures in video recorded lectures. Nine semantic categories were identified: put, spread, swipe, close & open, flip & swing, touch, pint, hold, and others. He also applied Principal Component Analysis to reduce the dimensionality of registered gesture data, subsequently finding that two gestures accounted for more than half of the variance of poses: a point-vs-rest gesture (mostly deictic), and a spread-vs-rest gesture (mostly iconic or metaphoric). Furthermore, Tian and Bourget (2016) annotated gestures from five video lectures and used student surveys to determine relations between gestures and content significance. They found that three types of gestures were strongly correlated to pedagogical significance: pointing (stretch arm, fingers pointing), ball (form a ball shape using two palms or fingers), circle (draw one or more circles using a finger or a palm).

6.3.3 Actor-on-board interaction

Some presentational entities are devices with which the actor interacts with the board. The actor uses their own body to direct attention to points of interest in the board, using deictic gestures and their gaze. A stronger modality of actor-board interaction happens when the actor physically writes on the board.

Handwriting and drawing

The positive effect of observing how the instructor draws diagrams on the board may be related to fundamental multimedia learning principles, as discussed by Fiorella and Mayer in (Fiorella & Mayer, 2016). First, the moving hand and drawing may act as visual cues that catch learner's attention (signaling principle). Second, the drawing is often performed simultaneously with the oral narration, which enforces the integration of information and knowledge acquisition (temporal contiguity principle). Third, the progressive elaboration of drawing diagrams produces a segmentation of the contents in smaller, easier to learn, pieces, compared to a sudden visualization of the full diagrams (segmentation principle). Fourth, a more general cognitive principle states that humans have evolved to learn by observation of other's movements (embodiment principle).

The recording of drawing and handwriting does not require a physical writing surface. Today, instructors are able to write and draw directly on tablets and other drawing devices. The so-called *digital ink* yields technological support for handwriting on screen. Digital ink is used in classrooms to enhance lectures by adding handwriting

[11] http://www.videolectures.net

on top of visual content created in advance (Venema & Lodge, 2013). Lectures with recorded digital ink and audio can be stored and distributed as 'pencasts' (Herold, Stahovich, Lin, & Calfee, 2011). Early studies on digital ink showed that teachers use this resource for signaling relevant visual content (R. J. Anderson, Hoyer, Wolfman, & Anderson, 2004): this suggests that digital-ink based presentations may contain visual cues that benefit learning.

If the instructor handwriting is displayed instead of typeface text, legibility issues may arise in learners. To overcome this risk, techniques have been developed for automatic transformation of handwriting into typeface script, as in (Cross, Bayyapunedi, Cutrell, Agarwal, & Thies, 2013).

The effect of handwriting in learning is beginning to be specifically tested. Fiorella and Mayer (2016) also conducted a set of four experiments to test the combined effect of drawing, as well as showing the instructor's hand and the instructor's body. The combined results show that observing the instructor drawing and showing the instructor's hand has small, positive effects in learning, though showing the instructor's body did not help in learning. Türkay (2016) assessed overall differences in retention and engagement between whiteboard animations and other formats: narrated slides, audio only, and text only. She found significant differences in favor of whiteboard animations.

Actor's deictic gestures and gaze

Sharma et al. have been investigating hoy learners follow teacher's gestures in video lectures. They defined a "with-me-ness" scale (K. Sharma, Jermann, & Dillenbourg, 2014) that uses two measures: the *perceptual* measures the degree of following instructor's deictic actions; and the *conceptual* measures the degree of following teacher discourse, that is, eye attention to on-screen points of interest. They measured both dimensions in an experiment and found great correlation between the scale and learning outcomes. This supports the idea that showing deictic gestures positively guides learner's attention, but that correlation can be explained by other causes.

Another source of signaling in classroom lectures is the instructor's *gaze*. As with hand or body gestures, instructor's gaze may point to areas of interest in visual presentation contents (eg. slides or diagrams) that can guide learners. Some videos implement the method of enhancing instructional videos by adding the actor's eye movements overlaid on the video, to show relevant visual parts and guide learner's attention. The effect of explicitly adding instructor's gaze to educational videos has been assessed in several studies (see the review of van Gog & Jarodzka, 2013; and the introduction section of van Marlen, van Wermeskerken, Jarodzka, & van Gog, 2016).

A set of recent studies by Sharma et al. explored gaze-augmentation in conceptual learning on adult education settings. In a first study (K. Sharma, Jermann, & Dillenbourg, 2015), they incorporated teacher's gaze to a MOOC video lecture with complex visual contents (urban landscapes). Teacher's gaze was captured with an eye

tracking device as he recorded the video and incorporated in the video edition as a visual cue. Authors compared user behavior of gaze-augmented videos with gaze-absent videos, finding that there were fewer video replays and less video pauses in the gaze-augmented videos, therefore implying that gaze enhances understanding. Later, they set other experiment (K. Sharma, D'Angelo, Gergle, & Dillenbourg, 2016) in which they compared three conditions of visual signaling of relevant content: no signaling, using a pen-like pointer and showing teacher's gaze. The content was a slide-based tutorial on cloud type recognition. They found that visual signaling enhanced learning, but no difference between pen-like pointer and gaze augmentation conditions.

Gaze-augmentation is a promising technique, but it must be taken with caution, since gaze patterns may exhibit substantial differences according to expertise and cultural roots (McIntyre, Mainhard, & Klassen, 2017).

6.3.4 Auxiliary instructional entities

Subtitles and closed captions

Subtitles are a common feature in video media. They can be placed as a fixed box in the frame, as optional overlaid text, or as an out-of-frame text. Its contents may be a full transcription of the narration, or a closed caption suitable for deaf or hard of hearing (DHH) people viewers. Closed captions include a non-verbatim transcription, as well as annotations of relevant sounds. Subtitles and transcriptions may be edited in advance or may be created in real time by using automatic speech recognition. Subtitles may be presented in a language other than the original narration, thus allowing non-native learners to access the narrated contents. The potential role of subtitles is to increase user accessibility: non-native, DHH, or low-skilled learners.

What does the research say about the use of subtitles when learning with video? Groundbreaking eye tracking experiments by d'Ydewalle and colleagues (d'Ydewalle, Van Rensbergen, & Pollet, 1987) showed that when presenting concurrent audio and subtitles in the same language, people spend considerable time reading the subtitles even if they are native speakers and they understand the audio. Reading subtitles seemed to be an automatic task. Further research has confirmed these findings for both native and non-native speakers: for example, two eye tracking studies (Kruger, Hefer, & Matthew, 2014; Winke, Gass, & Sydorenko, 2013) report learners spending between 43% and 75% of the time watching subtitles. Whether this visual attraction to subtitles is harmful or beneficial for learning has been an intense research topic, with contradictory results. To mention an example, (Lavaur & Bairstow, 2011) found a negative subtitle effect in English native speakers watching video with French subtitles, while (Hinkin, Harris, & Miranda, 2014) reported just the opposite. Research on cognitive multimedia learning provides evidence of coherence and redundancy effects when adding printed text to voice and visuals (Mayer, Heiser, & Lonn, 2001). Moreno & Mayer (2002) showed that the

combination of audio narration and subtitles can enhance learning, provided that no other concurrent visual material is shown. A practical application of those findings is that audio and subtitles should be placed before presenting visuals, but never simultaneously. Based on those evidence, the orthodoxy of CTML advices to avoid simultaneous voice and text.

Nevertheless, there is little doubt that subtitles are a need for the hearing-impaired learner, and useful in language learning. For example, an experiment with captioned instructional videos for English as a Foreign Language (EFL) (BavaHarji, Alavi, & Letchumanan, 2014) showed that learners who watched captioned lectures outperformed those who watched videos without captions in vocabulary acquisition and language proficiency development.

A study from van der Zee et al. (van der Zee, Admiraal, Paas, Saab, & Giesbers, 2017) offers an explanation for the disparity of results on subtitle effects. They hypothesize that the effect of subtitles is tightly related to the visual-textual information complexity (VTIC) of the video. For videos with low VTIC, subtitles can help learning (or at least, do not damage learning), because learner's cognitive resources are sufficient to process the extra load demanded by subtitles. On the other hand, adding subtitles to a video with high VTIC would lead to cognitive overload and resulting inconvenient effects as split-attention. Zee and colleagues tested successfully this hypothesis in a L2 context, and in addition they found no main effect of subtitles once video complexity and learner's language proficiency were accounted for. This VTIC hypothesis can explain why subtitles have positive memory effects in documentaries (Lång, 2016) and narrative films (Hinkin et al., 2014) that have little visual complexity, but consistently show detrimental effects in complex multimedia animations, as in (Moreno & Mayer, 2002).

The question of closed captions has also been raised, e.g. in (Tisdell & Loch, 2017). This study measured student preferences about closed captions in mathematics worked examples. Students declared a very high agreement on the usefulness of closed captions, with higher values for L2 speakers.

In summary, there is not still a conclusive evidence of the effect of subtitles in instructional videos, but some advices could be done: a) subtitles are very useful for non-native learners, as well as the deaf and hard of hearing; b) subtitles may deter learning if they concur with simultaneous visual information.

Acoustic signals

The synchrony of image and sound is routine practice in film making. "Hollywood filmmakers use cinematic techniques of image and sound to focus the attention of the spectator on the salient elements that further the narrative action" (Ruoff, 1992). Convenient use of background music in narrative films helps to remember highlighted scenes (M. Boltz, Schulkind, & Kantra, 1991) and affects the cognitive processing of visual scenes (M. G. Boltz, 2001). Many studies have noticed positive effects of sound in educational television (see references in Wetzel et al., 1993, p. 109). In multimedia learning environments, sounds may attract learner attention on

relevant information, reduce distracting stimuli, and increase engagement. Recurring sounds can be associated with semantic references shown in the video scenes. (Bishop & Sonnenschein, 2012) describe nine qualities of sound and their common associations in audiovisuals: intensity, pitch, timbre, speed, rhythm, shape, reverb/echo, directionality, harmony.

In spite of the powerful potential of sound and music, the signaling use of sound is rarely observed in instructional videos. Sound is often limited to a narration, and occasionally to music in start and end credits. In my survey of older Project Prometeo footage I have observed some usage of music cues for topic transitions and end-of-clip shots, but it's a scarce feature in current MOOC videos. One of the few cases of sound cue is the mouse click. In many screencast tutorials and lectures, clicking sounds are heard as the instructor manipulates her/his computer when showing the example (tutorial) or the presentation slides (lecture). In live lectures, these sounds are raw recording and probably they were not intentionally integrated as signaling cues in the video. In fact, they may be regarded as environmental noise. On the other hand, recorded screencasts may use mouse clicks explicitly as signals. Video capturing software tools often offer clicking sounds as a configurable setting for the output soundtrack.

The sound scarcity in instructional videos contrasts with other expository genres, as broadcasting news, in which sounds are frequently used to signal relevant events such as a section ending; or documentaries, that adopt all the sound conventions of narrative films. This absence also has been pointed out in instructional software (Bishop, Amankwatia, & Cates, 2008): "while sound is being incorporated into many learning environments, many instructional designers are using sound only for literal, information conveyance and not yet exploring how to exploit the associative potential of music, sound effects, and narration to help learners process the material under study more deeply". This same study suggested advanced uses of sound beyond 'signaling bells and whistles', for example, creating a systematic auditory syntax for categorizing main ideas, or using sound to tie into previous knowledge.

I have not found direct research on signaling sound cues in instructional video. The closest works are the aforementioned by Bishop and colleagues (Bishop et al., 2008; Bishop & Sonnenschein, 2012). This lack of research is another sign of the underexploited potential of sound as an instructional resource. I advocate for more research and more experimentation on this matter.

6.3.5 Non-instructional entities

Most videos include a number of visual and auditory objects which do not provide direct instructional content. Some of them contain valuable non-instructional information, as credits, titles, overlaid logos, and copyright notices. Others are purely decorative and serve to esthetic purposes. In the middle of these extremes, we can find non-instructional entities that provide subtle informational, social or emotional cues, such as background music, or the physical objects belonging to the scene setting (furniture, wall painting, etc.).

Multimedia learning theories tend to discourage non-instructional entities, since they add extraneous load into the learner. Moreover, adding more entities to the video may increase production costs. These two arguments (minimize extraneous load and save production costs) are a rationale for a minimalist design of instructional videos: focus on the instructional content and avoid ornaments. These principles result in videos with few editions, few shot transitions, and scarce decoration. The study of Morain and Swarts on YouTube tutorials (Morain & Swarts, 2012) backed that conclusions: good videos focused on the essential explanation and avoided extraneous discourse, visuals and sounds. Nevertheless, some non-instructional entities in an instructional video will bring valuable contents or have some qualities that are directly linked to the learning process.

Non-instructional text

Opening credits showing an institution name situate the social source of the product and provide authority. An overlaid box under the instructor that shows her/his name and position attaches a formal label of authority, which in turn raises the instructor's credibility with the learner. The location of recording may show a luxurious office or laboratory that manifests power. All these elements contribute to a rhetoric of power that frames learners into a favorable attitude to accept the message. A very different recording setting may show a busy office desk filled with piled papers and family photos that transmit warmness and intimacy, that seeks empathy and openness in learner's attitude. Later in this dissertation, I will cover in more detail the social and cultural aspects of instructional videos (Section 6.7).

Decorative pictures

Decorative illustrations may distract learners with extraneous load, especially if they are too salient. The so-called *seductive details* may threaten learning by drawing the learner's attention away from relevant contents (Harp & Mayer, 1998). A seductive detail effect has been found, by which learning performance decreases in presence of this kind of material. This effect is considerable in retention and transfer performance (Rey, 2012). Nevertheless, some research has found that learners can benefit from decorative entities, when they are designed to induce positive emotional states or metacognitive support in the learner (Schneider, Nebel, & Rey, 2016; Sitzmann & Johnson, 2014). Specifically, Sitzmann and Johnson found experimental evidence that in video-based instruction, seductive details indirectly improve learning by reducing negative affect, and hinder learning by decreasing time on task. Also, seductive details moderate attentional focus and the effect on learning on attrition. All these findings are in consonance with the *emotional design hypothesis*, which proposes that designing features with a potential emotional response in learners will influence learning performance (Park, Knörzer, Plass, & Brünken, 2015).

Background music

From the perspective of cognitive load theory, background music should be considered a source of extraneous cognitive load and should be avoided, just like any other noise or sound not belonging to the instructional information stream, as it is stated in the coherence principle (Moreno & Mayer, 2000a). However, music has been proven to affect cognition. For example, (Ilie & Thompson, 2011) showed that listening to music before cognitive tasks changed subjects' mood in a way that increased creativity and speed of processing. In addition, music has a great potential in higher education online instruction for motivation, learning enhancement, and even as a primary instructional resource (Dunlap & Lowenthal, 2010). In the specific case of documentaries, the type of background music has been proven to influence the perceived reliability of the narrator (John Baggaley et al., 1980).

Many YouTube procedural tutorials use background music. The study of ten Hove on YouTube video tutorials (P. E. ten Hove, 2014) revealed that many poorly rated videos use background music and sound as *the only* auditory resource. Average videos usually do not use background music, and finally, best valued videos do use background music, but only in selected segments of the clip. These results suggest that background music may have some positive effect in learner's engagement when it is shown in carefully selected times of the video rather than as a continuous flow, thus using background music as a signaling cue.

6.4 The Interaction domain

Multimedia learning theories embrace interactivity as one pillar for effective learning. The experiments made by Richard Mayer and collaborators (Mayer & Chandler, 2001) showed that even the simple ability to control the pace of an instructional animation enhances knowledge acquisition. The Interaction domain covers several types of entities that are added to the video medium in order to enable direct interaction from the viewer. This section will describe a taxonomy of interaction entities, grounded on previous characterizations of interactivity in multimedia learning objects. Then a list of distinctive interaction entities will be described and discussed.

6.4.1 Interactive video

Interactive video was once defined as "a video program in which the sequence and selection of content is determined by the learner's response" (Floyd, B. & Floyd, S., as cited in McNeil & Nelson, 1991). This definition intended to identify a new class of learning video use, apart from in-classroom video exhibitions where the learner could only watch passively. Affordable media such as VCR and DVD allowed students to watch instructional video at home and acquire more control on the time of watching and the playback pace (Hofmeister, Engelmann, & Carnine, 1986). Personal computers and online systems increased the ability to control and also to dialog with the multimedia learning materials.

It is important to notice that there is no single definition of 'interactivity'. There are multiple proposals for a definition, which vary depending on the field of research and application (Domagk, Schwartz, & Plass, 2010; Johnson, Bruner II, & Kumar, 2006). Interactivity can be characterized from multiple perspectives: technological, psychological or sociological. From a technological perspective, we can describe the features that certain device offers, or the affordances that have potential to engage the learner in activities. From a psychological perspective, we can focus on the cognitive and behavioral activities that take place in the learner when interacting. Sociologically, we can center on the dialogs that emerge from the interaction.

In this study, I will use the term 'interactivity' from a technological perspective. I will examine the features and affordances of video technology that enable the learner to exert control over the content presentation and to dialog with the system.

Effects of interactivity (and its absence) in instructional video

The drawbacks of traditional, non-interactive video were well known decades ago. Quoting Schwan & Riempp (Schwan & Riempp, 2004):

> *Viewers of traditional, non-interactive educational videos typically face problems similar to those of the Tetris players. They must rapidly organize information that is presented at a rate they cannot change. While texts*

> *allow their readers to tune their reading behavior to their cognitive needs, this is not the case for traditional videos or films.*

Under this premise, it was expected that adding more user control over video playback should be beneficial for the learner. In fact, several studies have found that adding interactivity to videos enhances the learning process (Cherrett, Wills, Price, Maynard, & Dror, 2009; D. Zhang et al., 2006).

The addition of interactive features to digital video may enhance user's watching experience. This hypothesis was tested in (F. C. Li, Gupta, Sanocki, He, & Rui, 2000) in six video categories: classroom lectures, conference presentations, entertainment shows, news, sports, and travel. A prototype video browser was built with interactive features as table of contents, jump to video shot boundaries, playback speedup, and user annotations. For the classroom lectures, users made extensive use of table of contents and time speedup controls. Users reported very high satisfaction rates of table of contents, but low rates for personal annotations.

An early meta-analysis on interactive video effects (McNeil & Nelson, 1991) found that program controlled lessons where learners have little or no control over the instructional path offered better learning outcomes compared to interactive videos where learners could control the content path. In addition, the meta-analysis found a higher achievement effect in interactive video with guided control of review and practice. Therefore, program-controlled interactive video appeared to be more effective than learner-controlled interactive video. The authors suggested that interactive video "is best accomplished when it is guided and structured as opposed to being entirely under the control of the learner". Later developments in cognitive theory have shown that "the very nature of an interactive learning environment implies an increased cognitive load on the learner because of the number of activities required and decisions needed" (Schwan & Riempp, 2004).

A recorded video lecture lacks the interactivity of a live lecture, where teachers and learners can dialog, formulate questions, ask for repetition or elaboration, and in general modulate the flow of the activity. Even formal, academic lectures allow for some person-to-person interaction that a bare video will not provide. That absence of interactivity of instructional video has been signaled as a risk of the integration of video as a learning resource (Stetz & Bauman, 2013).

Other external resources may provide interactivity to the learning experience. A video learning platform may include forums, instant messaging, etc. In fact, there is a video-based technique that is inherently interactive: the video conference. However, this study is focused on pre-recorded educational video.

In summary, the evidence says that interactivity is most beneficial for learners with prior knowledge and having guidance to use the learning environment features, as stated in Mayer's *learner control principle* (Scheiter, K., in Mayer, 2014d, pp. 487–512). This assumption leads to design the content with a clear navigation system that follow instructor-defined learning paths, for example by using tables of contents. Furthermore, this assumption advocates to use interactive devices that obtain

information about the learner's performance and give feedback, for example by using interpolated tests.

Apart from the effect on learning, learner interactions with digital video can be analyzed in order to extract conclusions about learners' behavior that may enhance the teaching and learning process. The click stream retrieved by MOOC platforms is a fertile source of information about learner behavior that is generating findings with relevance in instructional design and video production (Guo et al., 2014; N. Li, Kidzinski, Jermann, & Dillenbourg, 2015; N. Li, Kidziński, Jermann, & Dillenbourg, 2015; Seaton, Nesterko, Mullaney, Reich, & Ho, 2014).

Support for interactivity in current video learning environments

Video digital storage allowed for instant random search to any time position. Increasing computing power facilitates the implementation of complex algorithms for navigation and searching. Today, instructional video still shows a limited interactivity, often provided by basic video play controls (play/pause/stop). As user interfaces evolve, more user controls are provided, such as playback speed, automatic captioning, and user-generated annotations, but the degree of user control has kept limited compared to other audiovisual products, as video simulations and video games.

A review of 2010 (Schoeffmann, Hopfgartner, Marques, Boeszoermenyi, & Jose, 2010) over more than 40 video browsing and retrieval interfaces and shows a state-of-the-art panorama of playback and navigation interfaces used in commercial systems. Most video browsing applications offered a VCR-player-like interface with enhancements. Niels Seidel developed an extensive survey of interactive features in video learning platforms (Seidel, 2015). Seidel performed a content analysis of 118 video-based environments, covering learning platforms and MOOCs, and major video repositories. Based on his analysis, Seidel identified 40 *interaction design patterns* in use in video-based learning, organized in two layers: macro-level interactivity (13 patterns) and micro-level interactivity (27 patterns).

6.4.2 Taxonomy of Interaction entities

Other works on characterization of interactivity

Moreno and Mayer (Moreno & Mayer, 2007) describe five types of interactivity in multimedia learning objects: *dialoguing* (learner receives questions and answers or feedback); *controlling* (learner determines pace or order of presentation); *manipulating* (learner alters the content of presentation); *searching* (learner finds new content material by querying); and *navigating* (learner moves to different content by selection).

Merkt and collaborators (Merkt et al., 2011) group interaction operations with learning objects in two layers: *micro-level* and *macro-level* activities. Micro-level activities are performed inside the learning object: re-reading, loopback, etc. Macro-level activities require the use of external objects, as top-down organizers (table of

contents, index). Merkt's work only considers how the learner moves across the learning object contents; micro-level and macro-level activities could be mapped respectively to the 'controlling' and 'navigating' types in Moreno and Mayer's classification.

One of the four dimensions in Ploetzner and Lowe's characterization of expository animations (Ploetzner & Lowe, 2012) is *User control*, which comprises the interactions that the learner may apply over the animation. The forms of user control in this characterization are: *time line* (play, stop, change speed...) and *presentation* (change appearance or information).

The proposed taxonomy

Table 6-7 shows the proposed taxonomy of interaction entities for this classification scheme, derived from the three references above and on-the-field observation. The list of entities is not exhaustive, since user interfaces are continuously emerging from engineer ingenuity. Nevertheless, the list is highly representative of the current state of instructional video interaction features. Basically, interaction entities can be grouped in two categories: *control* and *feedback*. Control entities enable learners to establish the pace, order and settings of the video presentation. Feedback entities serve to more complex learner-system dialogue, whereby the learner sends information that is directly related to the instructional content.

Additionally, four functions have been proposed to group interaction entities: *playback*, *navigation*, *dialog*, and *comment*. As regards the control entities, we distinguish fine-grained user control (playback) from an overall control over the presentation (navigation). Playback entities provide actionable items to change the visualization pace or appearance. Navigation entities allow users to have a global view of the contents, for example with a timeline, a content schema or a graph. Within the class of feedback entities, we consider two functional types: dialog entities are those which enable a question-reply interaction, usually to get some feedback on learner understanding or satisfaction. The comment entities allow learners to send annotations that will be permanently stored in the video, and optionally could be watched by other users.

The following sections will discuss in more detail these four functional types of interaction entities, showing concrete examples and research findings of their instructional effectiveness.

Table 6-7. Taxonomy of interaction entities

class	subclass (function)	entities
control	playback	basic playback panel playback speed control presentation control
	navigation	timeline (simple, enhanced) table of contents visual summary in-video hyperlink
feedback	system-user dialog	interpolated test system-generated pause
	user commentary	user-generated annotation

6.4.3 Playback entities

As they were introduced above, playback entities are the most basic case of control entities. They have been present since the early times of interactive video, as in the classic VCR panel: play, pause, stop, rewind, fast-forward. More advanced controls have been added to current interfaces, such as setting the speed of playback. Other control entities allow for changing presentation settings: image size, zooming an area, sound volume, hide/show overlays (subtitles, transcripts, annotations), etc.

Figure 6-3. Example of basic playback controls (YouTube).

This limited form of *controlling* interaction enables the learner to control the pace of presentation and repeat watching difficult passages, among other possibilities. The increased control over the video reproduction may foster learning. In addition, the playback panel is widespread and intuitive, thereby its usage while watching video does not represent a relevant cognitive load on the learner's side.

This hypothesis has been addressed and tested by some researchers. Schwan and Riempp (Schwan & Riempp, 2004) designed an experiment where students watched an instructional video to learn how to tie nautical knots. One group (interactive condition) could control video playback (pause, rewind...) while another group (non-interactive condition) only was allowed to watch the video from start to finish. Results showed that the overall viewing times were similar in both groups, but interactive users controlled the difficult knots video segments with acceleration, pausing, reversing and replaying, resulting in substantially less watching time for that interactive group and better understanding of the topic.

Playback speed control

Playback speed control was one of the interaction features included in the experiments of (Schwan & Riempp, 2004). They found that users able to change the playback speed obtained better learning outcomes that subjects with the non-interactive interface. Evidence shows that trained learners can adapt and benefit from time-compressed speech (Simhony, Grinberg, Lavie, & Banai, 2014), with no loss of learning performance. On the other side, playing in slow motion allows for paying more attention in complex parts of the video. Moreover, a fast playback can be used as a searching mechanism: the user can quickly traverse the video receiving both auditory and visual input, in search for a point of interest. Non-native learners may benefit from slowing down the playback speed, as it was agreed by Coursera students in a recent survey (Mamgain, Sharma, & Goyal, 2015).

A study found that approximately 2-4% of students in two Coursera courses changed the playback speed while watching video courses (N. Li, Kidzinski, et al., 2015). The same study found that decreasing playback speed was related to perceived difficulty of contents.

As regards the implementation of playback speed controls, Time Scale Modification (TSM) technology allows for stretching and compressing the time scale of the speech without affecting the perceived qualities of the voice, such as pitch and tonal inflections (Smith III, 2011). Efforts to introduce TSM-based playback speed controls were made to enhance fast video browsing, with noticeable results by the turn of the 20th century (Amir et al., 2000). Current computing power makes TSM implementation straightforward, so many video playback interfaces currently support changing playback speed within a range, usually from 50% to 200% of the original rate. Currently, there is not a universally adopted interface for playing speed control. In YouTube and some MOOC platforms, playback speed is controlled with a pull-down menu in the video settings icon bar. Other controls have been used, such as a slider.

Presentation control

Many video play interfaces allow users to manipulate certain presentation attributes: visibility of overlays and subtitles, zooming, etc. In particular, the effect of subtitles can be positive or negative depending on learner characteristics, such as language proficiency or perceptual impairments (see page 136), therefore it is very convenient that subtitles and captions visibility can be user selectable.

Advanced playback controls

Some advanced playback controls are being developed specifically for instructional video. A prototype of automated speed control (Song, Hong, Oakley, Cho, & Bianchi, 2015) applies heuristics about learner's body behavior. The system automatically adjusts playback status according to learner's head position: if the head is down, assume that the learner is taking notes, then reduce playback speed; if learner's head is not pointing to video the system reduces speed, or pauses or rewinds,

depending on the timespan of the distraction; etc. This contribution is valuable not only for the engineering achievement, but also because it helps to explore learner's micro-level activities.

6.4.4 Navigation entities

A higher level of learner control on the presentation of an instructional video is enabled by artefacts that show an overall representation of the video contents, and allow users to select one particular segment to be reproduced. These are the navigational entities. The most extended artefact for navigation is the *timeline*, a clickable horizontal bar, as seen in the top side of Figure 6-3. The timeline bar is a universal feature in modern video playback interfaces. Other navigational interfaces are tables of contents, which show the semantic organization of the video so that the user can directly jump to any of them.

Basic and enhanced timeline

The classic timeline bar is a horizontal bar that shows the relative time point where the video is playing. This basic arrangement can be enhanced in many ways. The timeline may show points of interest along the video: chapters, tagged events, in-video quizzes, among others. Typically, points of interest are rendered as small highlighted segments in the timeline.

One recent innovation is to show a histogram or a heat map of the display frequency of each time point, as logged from user video watching. By observing this representation, the viewer can have an idea of the 'most interesting' points in the video. The viewer can even make deductions from the usage patterns, to decide if the video is interesting or not. The timeline with added histogram is called *rollercoaster timeline*. In (J. Kim, Guo, et al., 2014) a full-fledged version of this kind of enhanced timeline is demonstrated. It provides three visual features: a *rollercoaster timeline* (histogram), *interaction peaks* (highlights most visited areas in the timeline), and a *personal watching trace* (a visual footprint of segments already watched by the user). This timeline offers novel navigation methods: when the user slides the pointer over an interaction peak, the cursor decelerates, to attract user's attention.

Figure 6-4. Rollercoaster timeline

Playback user interfaces have been redesigned in order to enhance engagement. For instance, Lee and Doh (2012) built a user interface for gamified e-learning K-12 courses where the progress bar displays an avatar running on a track. The track may show intermediate in-video learning goals. Besides, the avatar will change its shape as the student achieves more course goals.

During this research, no studies have been found that assess the effect in learning of any form of timeline enhancement.

Object-based navigation

Object-based navigation tools are suitable for board-centric video formats, such as screencasts, Khan-style and whiteboard presentations, where multiple graphical objects are drawn during the exposition. NoteVideo (Monserrat & Zhao, 2013) enhances browsing of Khan-style presentations with a novel interface that allows pointing and clicking on visual objects. The user can jump to the time where the selected object was first drawn or drag on the object to watch its history. Tool developers claim that user response time is significantly lower than other navigation mechanisms (text-based navigation and object scrubbing).

Another example of object-based interaction enhancements is a very sophisticated system for browsing software video tutorials (Nguyen & Liu, 2015). With this method, video watchers can interact directly on the video frame as if it is the original software tool. When the user triggers an action on the screen, the video playback is moved to the time when that action was recorded. This system requires capturing author's events during video recording.

In-video hyperlink

A hyperlink may connect to other object, or to a given time point in current video. Temporal hyperlink implementation was first documented in the *Elastic Charles* hypermedia journal project from MIT Media Lab (Brøndmo & Davenport, 1991). Video-to-video linking introduces challenges not present in hypertext links: limited lifetime within video playback, the meaning of the 'back' action, etc.

The hyperlink is seldom used in isolation. It is usually the building block of higher-order navigational structures, as tables of contents, indexes, user-annotated timelines, or a graph of supplemental material. Nevertheless, enhanced video interfaces such as YouTube allow to insert small clickable overlays with an associated URL to another video.

Table of contents, indexes and summaries

Many video-playing interfaces offer tables of contents and indexes. Instead of navigating over a time scale, the user is enabled to navigate over a 'semantic' map of the contents, therefore helping in the comprehension of the material. The table of contents may be shown as a separate graph, or may be embedded on an enhanced timeline. The *video summary* or *video abstract* is an (usually) automated table of contents that shows relevant events in the video, with links to their respective time points. There are three common methods to display a video summary: keyframes, video skims, and table of contents (Biswas, Gandhi, & Deshmukh, 2015).

As a general principle, we can contrast the classic time-based navigation with this *tag-based* navigation offered by tables of contents, indexes and summaries. Tag-based

navigation consists in adding marks to specific points in video timeline to facilitate user localization of segments and events.

In the user review of several interactive features made by (F. C. Li et al., 2000), viewers of video lectures gave the highest rating to the table of contents, well above other navigational features as *jump to the next section* and *fast-forward*. Many students watch recorded lectures to review contents while preparing for their examinations. For this use case, students will want to skip irrelevant footage and search for the relevant parts for review. A navigational feature that helps to locate the interesting time points of the captured video would be very useful. Video tagging has proven to be useful in these scenarios (Gorissen, Van Bruggen, & Jochems, 2015).

In one of the most referenced experiments about interactive video, Zhang, Zhou, and Briggs (2006) tested the differences in learning effectiveness of interactive video by designing an experiment under which students were exposed to four learning conditions: e-learning with interactive video; e-learning with non-interactive video; e-learning without video; traditional classroom. The interaction was provided by a web interface with basic playback panels and a table of contents that allowed to choose whatever unit to watch at any moment. The authors found that students in the e-learning condition with interactive video performed better and declared higher satisfaction than those in the other experimental conditions.

Merkt, Weigand, Heier, and Schwan (2011) explored the use of learner's macro-level activities. Two experiments involved three kinds of media: "common videos" with a simple interaction panel, enhanced videos with a table of contents, and illustrated textbooks. They found that the table of contents was less frequently used than micro-level actions as stop-rewind-forward, and also that this micro-level strategy was superior than macro-level browsing for processing the information in the instructional video. The studies were conducted with K-12 students. Later, Merkt and Schwan (2014) confirmed in an experiment that enhancing video with a table of contents may be beneficial, but it requires that the learner has previous training of search strategies.

Regardless of the non-conclusiveness of the above experiments, there are circumstances where a top-level navigational instrument is essential. For example, non-linear video design requires a table of contents or some type of navigational map (Meixner, Matusik, Grill, & Kosch, 2014), so that the user can have a global overview of the contents and choose the view path.

Tables of contents can be generated automatically from the video contents. The MMToC system (Biswas et al., 2015) takes slide-based video lectures and obtains salient words from the text in the slides and the speech transcript. Then a segmentation process produces a table of contents. Another interesting approach is crowdsourcing the generation of the table of contents: ToolScape (J. Kim, Nguyen, et al., 2014) process a how-to video and generates an interactive navigation toolbar with text and thumbnails of intermediate steps (see Figure 6-5). The steps are identified and verified by crowd workers recruited in Amazon's Mechanical Turk.

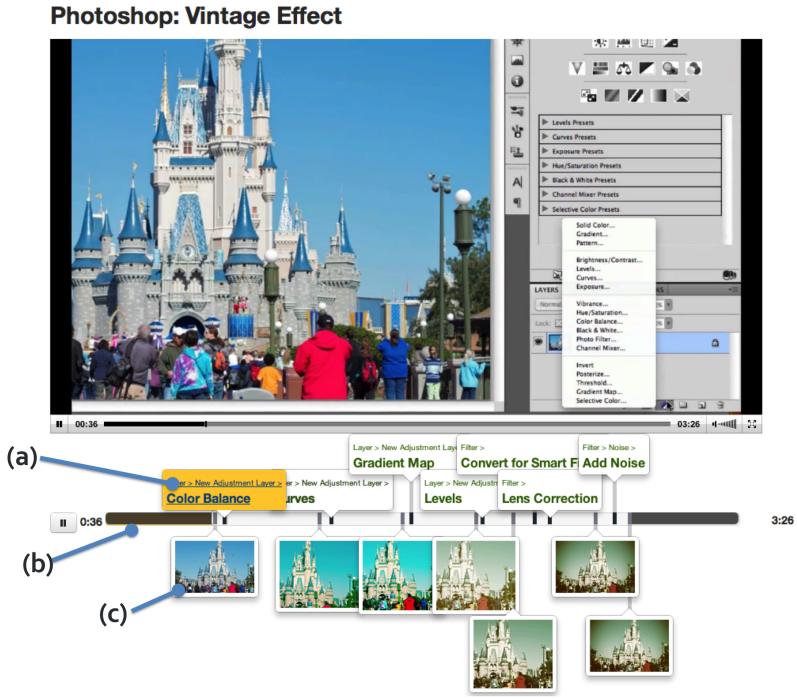


Figure 6-5. (Kim, Nguyen, et a., 2014) example of enhanced timeline interface

6.4.5 Feedback entities

Feedback entities get complex queries from the learner that are directly related to the higher-level processing of the instructional contents. This study has identified two types of feedback entities: interpolated tests (in-video quizzes) and user-generated annotations.

Interpolated tests

As early as 1991, the meta-analysis of (McNeil & Nelson, 1991) on interactive video features showed a significant effect of *guidance* features (e.g. embedded questions and subsequent suggestions). As the authors remarked: "it appears that learners benefitted more from guidance or suggestions on what to review after incorrectly responding to embedded questions". That observation is not surprising: learning sciences have shown that making practice tests during learning enhances the long-term retention of the information being learnt. The so-called *testing effect* has been demonstrated in many empirical studies (van den Broek et al., 2016).

Interpolated testing has been used in instructional video for decades. Today, many video platforms (YouTube, Coursera, etc.) support the addition of interpolated tests inside learning videos, usually in the form of multiple-choice quizzes. HTML5 standard has facilitated the implementation of those in-video quizzes. The usual mechanism starts showing the quiz when some predefined time in the video is reached. The video stops and the system waits for user input. When the user submits her answer, the system validates it and shows some feedback. Then the user is allowed to resume video playback.

The effectiveness of interpolated tests in instructional video has been studied by researchers. An experiment designed by (Cherrett et al., 2009) showed that interactive video with embedded tests was an effective tool. College students were assessed about risk identification. The evaluation tool used videos that incorporated embedded tests, which involved textual questions as well as locating hazardous points in scenes. The results were positive, both in learning outcome and in student experience satisfaction. (Vural, 2013) also evaluated the learning effectiveness of embedded tests by comparing video lectures with added in-video quizzes with simple video lectures. The experiment found positive effect of the embedded test feature. Similar results were obtained in an experiment with embedded tests on screencasts (Woodruff, Jensen, Loeffler, & Avery, 2014)

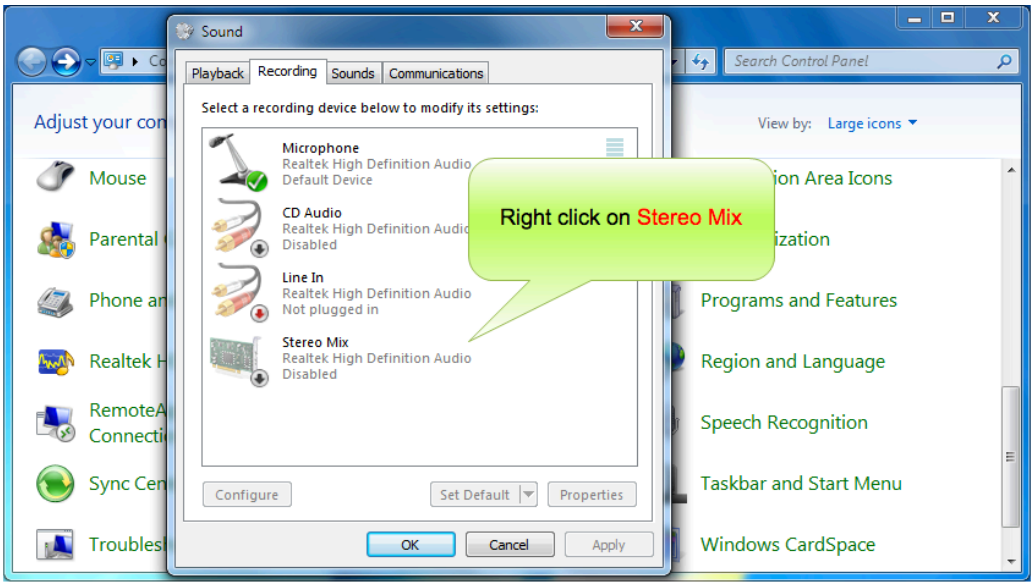
An issue to be warned is that placing the test in-video is prone to accessibility problems for users with a low-speed network connection (Mamgain et al., 2015). In that scenario, post-video or separate tests would work better.

There is a tendency for the learner to be overconfident in their prediction of learning, when the teaching design is based on video lectures. (Szpunar et al., 2014) analyzed the effect of interpolated tests in predicted and actual performance in video lecture-based training. The experiment involved 54 high school students in a statistics introductory course. Results showed that interpolated testing helped students to adjust their predicted performance to the actual outcome. Therefore, interpolated testing would be helpful to calibrate learning. In addition, the years-long work of Szpunar and his collaborators has revealed that interpolated tests within video lectures reduce mind wandering, increase task-relevant behaviors and favors learning (Schacter & Szpunar, 2015).

System-generated pauses

Some video interfaces allow for inserting system-generated pauses, that force the user to click for resuming video playback. This feature can be used to remind the user that she/he has to fulfill some external task, as in the screenshot shown in Figure 6-6, which was captured from a video tutorial explaining how to configure a software tool. The screenshot corresponds to a moment where the tutorial has required the user to perform some action in their computer. The video pauses itself and waits for user feedback to resume.

Figure 6-6. Screenshot of a system-generated pause in a video tutorial



User-generated annotations

Learners can use annotations to mark points of interest in the video. This feature is a support for self-learning and self-reflection processes. It can work in conceptual learning (lectures) as well as in procedural learning (Rich & Hannafin, 2009). Researchers have tested enhanced video interfaces which allow learner note-taking. The experiment of (Delen, Liew, & Willson, 2014) found better results in students that learned a topic using an enhanced video interface over the control group.

Some tools have been available for user video annotation (Hosack, 2010). YouTube also provides a simple system for author annotation. Based on (Hosack, 2010), there are two desirable properties of a video annotation system: a) *readability*: accommodate large amounts of text in annotations without obscuring video contents; b) *synchronous playback*: the playback of video contents must be synchronized with the display of annotations.

One interesting application of annotation technologies is to enable users to share their video annotations so that small improvements and corrections can be applied to the online video in a collaborative way (Cross, Bayyapunedi, Ravindran, Cutrell, & Thies, 2014). This method can even be used to make embedded translations of text items in video.

Manual annotation during video watching may be difficult for the student. To overcome this dificultades, prototypes of gaze-based annotation tools have been developed (Nguyen & Liu, 2016).

6.5 The Spatiotemporal domain

As the film critic Noel Burch observed, film production is about articulating space and time: "from a formal viewpoint, a film is a succession of fragments of time and space"[12] (Burch, 1970). Film production includes the *decoupage* or planning, which is "the operation consisting of decomposing an action (story) in shots and sequences, in a more or less precise way, before the film shooting". The *montage* of films involves the articulation of space and time by using several techniques. Montage constructs the message of the film and amplifies its qualities, enhancing the viewer's response. Cuts and transition shots mark the ending of a narrative segment. Camera angles and frame layout establish the roles and relations of characters. Sound effects and background music serve as cues to signal changes in mood, space and time. Applied to instructional films and videos, this filmic language can be harnessed to clarify the structure of the exposition, signal the relevant pieces of content and arise affective responses from the viewer that would foster learning (Wetzel et al., 1993).

6.5.1 Handling space and time

The spatiotemporal articulation comprises two dimensions: space and time. The spatial articulation deals with the layout of scenes and frames. This layout is built from the physical arrangement of objects and the camera setting. The camera is an essential element in film. It is the primary source of filming footage. Several camera parameters affect the way the recorded scene is perceived: camera position, angle, focal length, aperture and zoom will change the scene point of view, the perceived sizes and salience of the subjects, and many other qualities that affect to the cognitive and affective responses of the viewer. On the other hand, the temporal organization includes structuring the film by using a hierarchy of segmentation levels: shots, scenes and sequences. Furthermore, for instructional videos it is appropriate to add some levels to this basic temporal hierarchy, such as the partition in separate video clips and the non-linear structures of hypervideos.

The spatiotemporal codes are so complex and elaborate that they constitute a language in itself. Indeed, this language can be taught and harnessed by instructors to facilitate cognitive activities, as Gavriel Salomon demonstrated in his pioneering research on educational television (Salomon, 1979a).

Spatiotemporal articulation and multimedia learning principles

Several multimedia learning principles are tightly related to the spatio-temporal manipulation of the instructional film. The most relevant are the *segmenting principle* and the *contiguity principle* (Mayer, 2014d). The segmenting principle states that people learn better when the message is presented in short, learner-paced segments rather than as a continuous unit. Applying the segmenting principle to instructional video design leads to splitting the video into separate clips, or, less radically, to

[12] Burch's quotes have been translated to English from the Spanish book edition.

delimit the exposition segments into clearly bounded filmic units (sequences). The contiguity principle states that people learn better when words and pictures are physically and temporally integrated. This principle commands the film editor to pay special attention to scene layout and camera settings, in order to shoot an integrated frame in which related items are close enough. If meaningful objects are too distant from each other, a *split-attention effect* may arise: viewer's attention is divided in different areas of the screen, resulting in extra cognitive load and distraction. Moreover, the film edition should ensure a proper synchronization between the imagery and the speech, and an adequate orchestration of item entries and exits.

Filmic segmentation hierarchy

The film production industry has established a conventional film montage hierarchy: *frame*, *shot*, *scene*, *sequence*. This is precisely the hierarchy of analysis levels that has been proposed for the discourse analysis of film and television (De Vaney, 1991; Iedema, 2001; Paltridge, 2012, p. 177). Discourse analysis expands this hierarchy to account for higher levels of discourse: the *generic stage* and the *work as a whole* in Iedema (2001), or the *phase* in Baldry and Thibault's (2006) proposal for multimodal transcription. Similarly, Video Content Analysis researchers have proposed models in order to categorize the internal structure of video: the classic shot/scene/sequence hierarchy is common for characterizing broadcast videos or feature films (Truong & Venkatesh, 2007).

Truong and Venkatesh (2005) describe a model in which a video can be segmented in *logical units*, which can be shots, scenes, sequences or 'topics'. A *topic* is a segment where a single narrative or expository item is presented, equivalent to the *phase* in Baldry and Thibault's coding (2006). Figure 6-7 shows a UML diagram based on that model. The 'topic' logical unit is well suited to analyze expository formats, such as TV news broadcasts and educational videos.

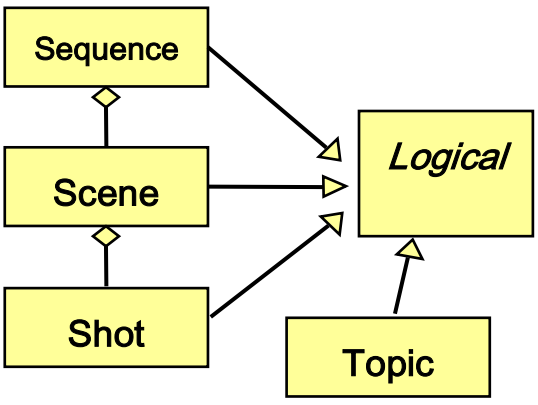


Figure 6-7. Conceptual model of temporal structures for video content analysis (based on Truong & Venkatesh, 2005).

6.5.2 Entities and properties of the Spatiotemporal domain

My review of research on the spatio-temporal and filmic aspects of instructional videos resulted in many video characteristics from these domains that have been considered relevant by the scientific community. In summary, the evidence that emerges from the review is that instructional videos should be designed with particular attention to these manipulations:

- Segment the contents into short, understandable pieces.
- Optionally, use a non-linear arrangement of the video segments.
- Ensure an adequate spatial and temporal coordination between spatiotemporal events and visible objects.
- Manage the information complexity in space and time: presentation speed, scene layout.

Therefore, spatial and temporal characteristics can be grouped in these four structural classes: *spatial layout*, *temporal segmentation*, *linearity* and *informational complexity*. These classes and some examples of characteristics are shown in Table 6-8.

Some properties can be used as quantitative measures of certain aspects in the video. For example, the speech rate (words per minute) and the cutting rate (shots per minute) are both measures of informational complexity. Unfortunately, I have not found proposals of standardized measures of informational complexity, neither partial (for some aspect of complexity) nor global. The only outstanding global property in the spatio-temporal domain is the video length (duration), which has been extensively studied.

The following sections will discuss in detail every group of characteristics and their respective supporting research evidence.

Table 6-8. Taxonomy of properties and entities in the Spatiotemporal domain

class	properties and entities
spatial layout	use of frame areas with semiotic relevance
	camera setting: angle, shot, perspective
temporal segmentation	film segmentation hierarchy: shot, slide, scene, sequence, clip, hypervideo
	segment transitions: pauses and temporal cues
	clip length (duration)
linearity	linear vs. nonlinear
	navigation graph (for nonlinear video)
informational complexity	presentation speed: words per minute, items per minute
	spatiotemporal complexity: cutting rate (shots per minute)
	timing between informational events (temporal contiguity, redundancy)
	spatial contiguity of informational items

6.5.3 Spatial layout

The spatial organization of an instructional video deals with three major design factors: the frame/scene layout and the camera setting. The scene layout defines the relative sizes and the locations of the scene items (actors and boards, principally). The camera setting not only contributes to the desired scene layout, but it also allows for multiple adjustments to enhance both static and dynamic attributes of the film: zooming, panning, focal length, and field of depth, among others. The foremost characteristic of the camera is that it provides a *viewpoint* for the scene: what is the viewer's (virtual) physical location in the scene. The viewer may be a member of a theater audience, may be sitting close to the model, or even she/he may feel as *being* the model (in a subjective camera recording).

In the following paragraphs, I will describe some findings that I have collected during my review that are related to the spatial manipulation of the instructional video.

Frame areas

Current instructional videos often exhibit little variations in layout across their time span. Many procedural tutorials are recorded using a single, fixed point camera. Analogously, talking head lectures and screencasts rely on one main layout that is used across the entire video length. This fixed layout may be occasionally interspersed by transition shots, cutaways and other spatiotemporal events, but the same frame/scene layout usually keeps consistent over time. That fixed arrangement

of objects in the screen leads to question what are the most convenient layouts for an instructional video.

A first relevant fact is that different areas in the frame receive different attention from viewers. It is believed that the center of the screen is the most important area; the right side is more important that the left side (at least, in left-to-right reading cultures); and the top side is more important than the bottom. Wetzel, Radtke and Stern (Wetzel et al., 1993, p. 92) and Bateman (Bateman, 2008, p. 51) mention abundant experimental evidence of the distinct relevance of certain screen areas. The 'up is good, bad is down' metaphor is well known in psychology research (Meier & Robinson, 2004).

Spatial contiguity

The spatial proximity of scene items is another important factor in learning effectiveness (the spatial contiguity effect in the CTML). There is an association between spatial proximity of displayed pairs of words and pictures and the type of judgments people make about them (Casasanto, 2008). On the contrary, if semantically salient items are presented in distant parts of the screen or they are not properly integrated, the viewer's attention may split between the multiple sources of content (Ayres & Sweller, 2014). In order to address the split-attention effect, arranging spatial proximity of relevant items in the frame is crucial. This arrangement may involve a careful edition of texts, pictures and instructor recording. For instance, text should be segmented in pieces that are small enough to be displayed near the relevant imagery. Moreover, text should be omitted sometimes and replaced by labels attached to the pictures. Florax and Ploetzner (2010) tested the effects of text segmentation and picture labeling and found that the absence of text segmentation and picture labeling produced the split-attention effect with respect to retention, but not for comprehension. The effect was stronger in text segmentation.

Research on item layouts

In my review, I have gathered some investigations on the relative placement of specific scene items. Most research has been dedicated to the size and position of the instructor and the board in video lectures. Additionally, I have found some research on the placement of subtitles.

Layout of instructor and board. Most research has to do with the relative size and location of the instructor and the board. (Bhat, Perry, & Chinprutthiwong, 2015) students of a MOOC course were offered two versions of the same lectures: one with a picture-in-picture instructor, other with a larger instructor blended with the slides. Students clearly preferred the blended format. (Korving, Hernández, & De Groot, 2016) compared relative sizes of instructor and board (slides). No significant effect was found on learning or attention between large or small relative sizes. (Z. Pi, Hong, & Yang, 2017) compared relative sizes of instructor and board (slides). No significant effect was found on social presence and cognitive load, but there was a significant effect in learning: the small instructor size produced better outcomes. One experiment (Korving et al., 2016) tested three layout configurations:

small board and large instructor, large board and small instructor, and board only (with no visible instructor). The authors measured learners' attention from self-reports. Participants reported more attention when the instructor image was present (whether large or small), but only after having a first video watching session (15 minutes long). Instructor's appeal was controlled for moderation: no significant effect was found. Similarly, another study performed with Chinese students (Z. Pi et al., 2017) tested the effects of instructor's image size in social presence, cognitive load and learning. Three layouts were considered: small, medium and large instructor's size. The experiment found that instructor's image size did not influence social presence and cognitive load, but negatively affected learning performance.

Subtitles are commonly placed in the bottom area of the screen. Some methods have been proposed for subtitle placement that attempt to reduce the spatial distance between the subtitles and the relevant items in the display. One method places each on-screen subtitle block next to the associated speaker picture (Hu, Kautz, Yu, & Wang, 2015). A further evolution of this technique was presented by (Akahori, Hirai, Kawamura, & Morishima, 2016): it is a 'region-of-interest' subtitle placement system which collects eye-tracking information from multiple viewers to infer what is the region of interest in each time frame. Subtitles are placed just below the calculated region of interest. These advanced methods should promote coherence between textual and pictorial information, compared to a conventional fixed-position subtitling, but this hypothesis is still to be properly tested. The ongoing research has shown mixed results at the moment (Chen, Yan, Liu, & Jiang, 2017; Ouzts, Snell, Maini, & Duchowski, 2013).

As a whole, there are no clear effects from the relative sizes of the frame items. There is clear evidence of the effect of placing items in some frame areas, and the effect of the spatial contiguity of meaningful items. The relative sizes of the instructor and board seem not to be relevant in learning effects. The spatial arrangement of these two items has more to do with customs than with effectiveness, therefore the diversity of layouts may be a matter of communication genres.

Camera setting

A cinematographic camera allows to manipulate multiple shot parameters (Wetzel et al., 1994), such as shot length, camera movement, camera angle, zoom, and focus lens effects. Camera movement include actions such as panning, tilting, dollying, crane movement, and trucking. Camera movement should have made purposedly and controlled, in a way that it does not call attention to itself (Zettl, 2016). Lens effects may affect several parameters, such as depth of field, linear perspective, and perceived size and distance of subjects. The review of Wetzel, Radtke and Stern (Wetzel et al., 1993) found some experimental evidences that camera and lens manipulations produced weak cueing effects (signal or guide to a relevant point in the film) that should be useful to enhance learning effectiveness, or at least learner's motivation. Furthermore, Salomon (Salomon, 1979a) demonstrated that zooming and other camera techniques can be used to support cognitive processes and facilitate learning.

Actor's social construction. Research has shown that the perceived physical distance to observed objects influence the observer's emotional response to objects (L. E. Williams & Bargh, 2008). Moreover, the perceived size and distance of the onscreen actor may modulate the actor's social presence. Multimodal discourse theorists such as van Leeuwen (van Leeuwen, 2008, p. 138) consider that shot length (close, long) gives a measure of social distance, and camera angle (frontal, oblique, high, low) induces a social relation. What has to be explored is the actual influence of this camera-induced social priming in the learning process. It is worth mentioning the experiments on educational TV by Baggaley (John Baggaley et al., 1980), in which he demonstrated that several camera settings produced different social responses on viewers. I will elaborate these social aspects in section 6.7 , where I discuss the social appearance domain of this classification scheme.

Subjective camera. One particular camera setting is the *subjective camera*, also called *first-person* shooting. In this configuration, the action is filmed from the actor's point of view. Subjective videos are potentially useful in procedural videos that manage physical objects, because the viewer can become more easily involved in the observed action. In the post-war US Army experiments on educational films, Roshal (1949) found that in a procedural instructional video (tying knots), a 'subjective camera' from the demonstrator viewpoint produced better learning than a viewpoint of a trainee watching the demonstrator. Wetzel et al agree that procedural demonstrations should use a subjective viewpoint (1993, p. 98). The current boom of video tutorials has stirred new academic interest in the subjective camera: multimedia learning researchers Fiorella, van Gogh, Hoogerheide and Richard Mayer (2017) have presented a set of experiments that evidence some *perspective effect*: students performed better after viewing a first-person instructional video, compared with students who watched a third-person version.

6.5.4 Sequencing hierarchy in instructional video

This taxonomy will enhance the classic hierarchy of film sequencing with some additions that will allow for a better characterization of instructional videos. Table 6-9 depicts the expanded hierarchy used in this taxonomy. Two levels have been added to the classic hierarchy: the *clip* and the *hypervideo*. These levels are proposed to account actual instructional video production techniques: when one instructional material is segmented into separate small video files (clips), following the segmentation principle of CTML; and when several video clips are arranged in a way that multiple navigation paths are allowed, resulting in a non-linear video or hypervideo. These two levels, clip and hypervideo, are discussed in detail later in this section.

Many recorded lectures, especially in the 1990s-2000s, consist of a set of slides plus a voiceover. Many modern online videos keep using the slide as a content organizational unit. In those 'slideshow' video styles, the slide replaces the shot as the basic segmentation entity.

Table 6-9. Film segmentation hierarchy for instructional video

entity	description
shot	a continuous recording of a scene or object, perceived as recorded in a single camera take
slide	a board with a static depiction of text or graphics
scene	a sequential arrangement of shots that share the same spatial and temporal setting
sequence	a sequential arrangement of scenes and shots that fulfill one or more expository goals
clip	a distinct and identifiable video document
hypervideo	a collection of video sequences or video clips that can be watched in different orderings by navigating links

The difference between 'scene' and 'sequence' can be unclear, particularly in video tutorials and recorded lectures, which hold a very simple montage structure. It would be helpful to recall Bateman and Schmidt's criterion for scene identification (Bateman & Schmidt, 2012, sec. 7.3.1): "the scene is considered here to be recognizable by an observer if it is possible for that observer to carry out a substitution test of the following sort. If the observer can perform a thought experiment in which the putative scene can be designed as a single audiovisually reliable shot, i.e., can play the scene out as a shot in a spatially-unified context, then there is good evidence to accept that a scene is on hand.". On the other hand, a *sequence* is a sequential composition of scenes and/or shots that do not necessarily share the same spatial and temporal coordinates and are grouped to develop some expository or narrative step. In many instructional videos, the distinction between scene and sequence is blurred and can be collapsed into one single level of analysis, as will be discussed later in this section.

Spatiotemporal vs. rhetoric segments

An instructional video can be structured in both spatiotemporal and rhetoric dimensions: there is a spatiotemporal segmentation (scenes, sequences, clips) and there is a rhetoric segmentation (introductions, topic expositions, evaluations, conclusions, etc.). Spatiotemporal and rhetoric segmentations are closely intertwined. Usually the boundaries of scenes and sequences coincide with the boundaries of rhetoric phases. A frequent arrangement in instructional videos is a sequence consisting of one or more rhetoric phases. Nevertheless, one cannot make an exact equivalence between spatiotemporal and rhetoric units.

Physical vs. conceptual structures

The shot and the clip are physical structures (more related to the physical attributes of the video), while the scene and the sequence are conceptual structures that are more coupled with the semantics of the contents. Automatic detection of shot boundaries can be straightforwardly implemented by signal processing algorithms, but scene/sequence detection needs some degree of content analysis. This also applies to the learner: for a human watcher, it will be easy to delimit the shots in the video, but delimiting scenes and sequences could require more effort. This is a rationale for including special shots to better delimit sequence boundaries, for example with special cut shots, fadeouts, etc.

The shot

The basic unit of film production is the shot. A shot is the continuous recording of a scene or object from the time the camera starts until it stops (Beaver, 2006). A shot is perceived as uninterrupted and recorded by a single camera in the same take. The frame layout may change during the shot, but it will be due to changes in camera setting, or due to changes in the objects themselves, but never by editing cuts. Many instructional videos are synthetic instead of filmed, for example computer-generated animations. To properly cover those cases, the concept of 'shot' has to be detached of a physical camera take. The criterion for considering that a consecutive set of frames qualifies as a 'shot' is that it is *perceived* as if taken in an uninterrupted recording, even when that's not the actual way it was generated.

The shot has been considered the minimum 'building block' for film semantic. Metz (1974) considers the shot as the basic unit of semiotic construction in films, linguistically closer to a sentence rather than to a word. As Bateman and Schmidt (2012) point out, film analysis scholars agree with Metz's statement, perhaps due to strong semiotic properties of the shot concept, that is, *commutability*, "the possibility of changing the order of elements in order to produce distinctive variations in significance".

Feedback events. Some advanced instructional videos may contain user feedback events that stop the video and force a user response to resume: interpolated tests and system-generated pauses (see "The Interaction Domain" in this chapter). These interaction events should be considered as part of the montage, because they introduce cuts in the video sequence, thus generating a segmentation.

Cutaway shots. Baggaley and Duck (1975) showed that cutaway shots in interviews, showing a short shot of the interviewer at certain times, increased the perceived interviewee's tension, sincerity and understandability.

Scenes and sequences: a shallow hierarchy

Narrative films require a complex organization of shots. Segments of edited action are organized into sequences and scenes that evolve around the narration, which selects episodes over multiple points in time and space. Sometimes there are segments that are external to the narrated story (*non-diegetic* units), or multiple

subplots are intertwined across the montage. On the contrary, there is no comparable complexity in conventional instructional videos. Many video lectures and tutorials develop their exposition in a single act, during one single spatial and temporal setting, thus resulting in a single scene video. Single-shot instructional videos are not uncommon, for instance a recorded live lecture shot with a single camera. When a tutorial describes a procedure that last for a long time, the resulting film is indeed decomposed in sequences, one per temporal episode.

In summary, instructional videos usually display a shallow sequencing hierarchy. The difference between scene and sequence is somewhat blurred in instructional videos. The only exception to this structural shallowness occurs in the documentary genre, in which full-fledged cinematographic techniques are commonly deployed. Therefore, in many cases it will be appropriate to collapse the levels of 'sequence' and 'scene' into a single level. A proper name for this combined level would be 'shot cluster', as in (Messina, Montagnuolo, & Sapino, 2006), who used this name to designate a set of consecutive shots with similar properties.

The clip as a segmentation entity

The 'clip' has been added to the classic hierarchy. Many instructional video materials are edited as a set of independent clips which can be watched separately. The clips in the same set remain semantically related and they usually must be watched in a predetermined sequence. In those cases, the film editor has designed to split the contents in separate video files to properly segment the contents in small chunks. This is indeed a decision of edition (or montage), thus the clip deserves a recognition as a level in the edition hierarchy. This feature was already identified in Kay's literature review on video podcasts (Kay, 2012), which differentiates *non-segmented* and *segmented* video podcasts, being the last those "broken up into smaller chunks that can be searched and viewed according to the needs of the user".

The video clip is an organizational level that matters in film industry: for example, TV series are organized into *episodes* and *seasons*. Analogously, instructional videos are frequently decomposed in separate video file for each 'chapter', 'unit' or 'lesson'. The reasons for decomposition are different between instructional videos and entertainment TV. The media industry imposes strong time length restrictions on the format of TV series. Furthermore, inter-episode organization is driven by narrative. On the other hand, instructional videos are decomposed in different clips due to time length constraints based on learning effectiveness, according to learning goals and, more decisively, according to the content topics. Some courses will add separate clips for discourse organization: an introduction clip, a summarization clip, an end-of-course clip, and the like.

6.5.5 Non-linear video

We can consider the instructional video as a set of informational segments to be watched. There should be a plan or method for ordering those segments over time. The method should be as simple as a fixed, pre-planned sequence, which is the design

that most videos follow. Beyond this basic sequencing, there are more ways to define a content flow: for example, the video may contain multiple sequence paths so that a different path can be taken in each watching session, depending on user input or some external feedback. This approach leads naturally to *non-linear video*.

A non-linear video is a multimedia object, based on video segments, that can be navigated by using links embedded in the videos. A non-linear video is characterized by a non-linear structure, alternative playback paths, choice elements and the influence on the order of scenes (Meixner et al., 2014).

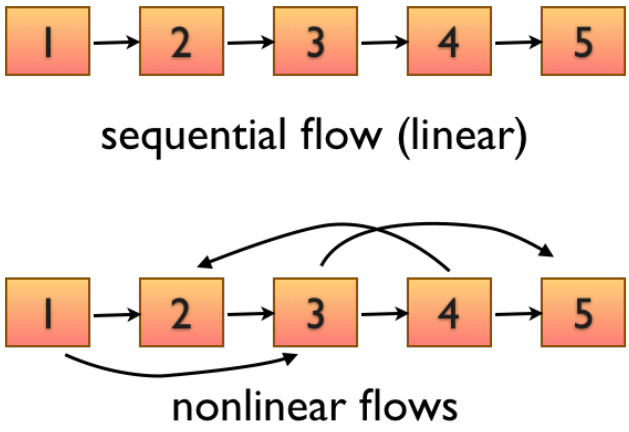


Figure 6-8. Flows of linear vs. non-linear videos

Figure 6-9 shows an example of a non-linear video, retrieved from YouTube platform. This is a tutorial about how to play the twelve major chords with a guitar. The lower area of the video frame shows an array of chord codes, all of which are clickable. When the user clicks on a chord code, a new clip is played demonstrating how to play that chord. This new clip also contains the array of chord codes.

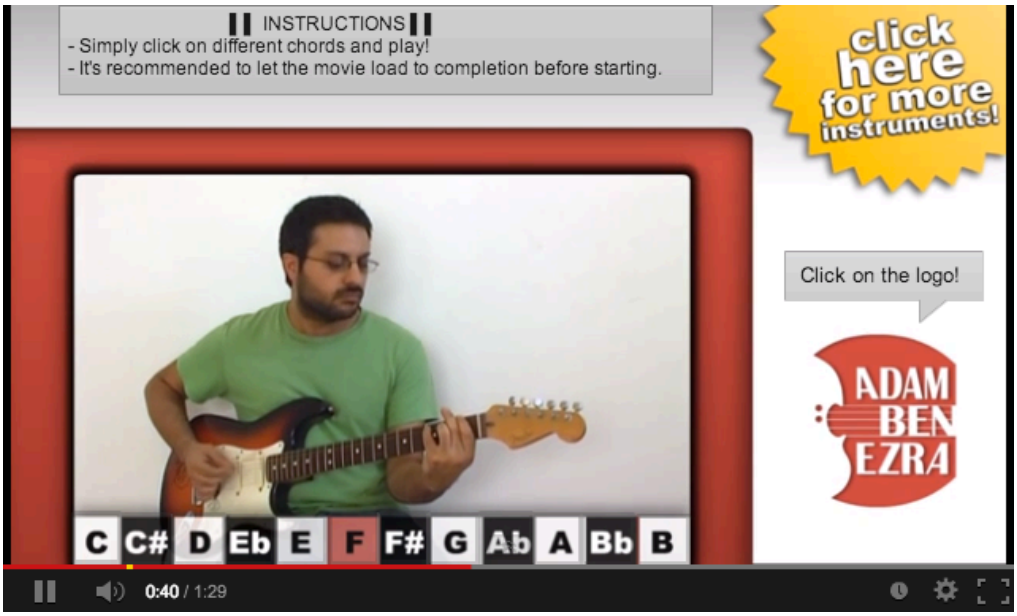


Figure 6-9. Screenshot of a YouTube non-linear video

According to Mujacic and collaborators (Mujacic, Debevc, Kosec, Bloice, & Holzinger, 2012), a hypervideo document is built out of these five types of components: scene, narrative sequence, temporal link, space-temporal link, and navigation. The *scene* is a linear sequence of video frames. The *narrative sequence* is a possible viewing path for a group of scenes. *Links* are references from one source video to a destination video. A *temporal link* is activated when the source video reaches some time point. A *space-temporal link* is activated when user points to some area in the frame. Finally, the *navigation* is a collection of static elements that are always available to the user. The navigation element can be implemented in various forms: a scene graph, a table of contents, a keyword search interface, or others (Meixner et al., 2014).

Although a hypervideo would allow for a huge number of narrative sequences (viewing paths), the reality is that instructional designers of a hypervideo propose a bounded set of navigation paths. Some common navigation graphs are the linear, the 'fishbone', and the 'branched tree' (Fernando Caro & Romero Moreno, 2012).

Hypervideo allows for some degree of *adaptability*, that is, show different content, or allow/restrict different viewing paths according to user behavior (Mujacic et al., 2012). For example, an in-video quiz appears at some point in the video. Depending to the user answer, the system may take different story paths. More complex adaptations may take into account user's profile (geolocation, history of past watched videos, academic record, etc.).

Nonlinearity enable users to choose the ordering of the content presentation, for example to discard introductory segments that are not necessary if they have prior knowledge of the course subject. Nevertheless, more navigational freedom carries its disadvantages: some learners may get disoriented or may skip essential contents. Moreover, exploring non-linear content may demand an extra cognitive load.

The actual benefit of hyperlinked video instruction is not clear. Three examples of research exemplify what can we expect from hypervideo. First, Reiss (2008) tested the effect of different enhancements to a simple video lesson: adding embedded triggers that activate HTML supplemental material; or adding user activated links that show various additional material. Reiss found that the simple video interface was more effective in information acquisition, thus suggesting that a single stream of information should be preferred, at least for novice learners. Second, Pusic, LeBlanc and Miller (2007) compared with novice college students two versions of a computer tutorial: one linear version and other non-linear, with hyperlink-based navigation. Both versions worked well, with no significant differences in ability gain, though the linear version required less time to complete. Third, Tonndorf et al. (2015) investigated the usability of non-linear video in elderly learners. Two versions of a video tutorial on physical training were used: one was a linear video, and the other was a hypervideo. The authors found slightly more usability problems in the hypervideo version, but users of hypervideo showed more active learning behavior.

6.5.6 Temporal segmentation

The temporal arrangement of the instructional video is driven by two main principles: to split the contents into small temporal chunks (segmentation principle), and to follow an articulated, understandable sequence. Both principles should reinforce each other: ideally, each filmic segment contains a full expository phase (or a small group of phases); and, inversely, every expository phase is designed to be short.

The montage techniques should contribute to a better segmentation. Segments can be implemented by using any level of the filmic hierarchy, for example at the sequence level or at the clip level. Also, editing methods should be used to make the segmentation explicit and easy to comprehend: for example, by using transition shots, pauses, and temporal cues.

Evidence of the segmentation principle

There is abundant evidence of the segmentation effect. Schittek Janda et al. (2005) compared two versions of instructional content: one sequential video, and another video fragmented into eight short clips. The fragmented version caused better learning results and more learner satisfaction. In the field of expository animations, it has been observed that segmentation of dynamic visualization into smaller units and providing pauses between segments reduces the cognitive load imposed in the learner by the transience of the video stream (Spanjers, van Gog, & van Merriënboer, 2010). A practical test on segmentation was performed by Ibrahim et al (Ibrahim et al., 2012), who took an educational video and applied over it some treatments: segmentation into smaller units, weeding extraneous material and adding signaling. Learners who used the modified video obtained better results than learners who used the original video. Furthermore, (Spanjers, Wouters, van Gog, & van Merriënboer, 2011) showed that, with animated worked examples, learners with low level of prior knowledge learned more efficiently if content was segmented instead of continuous,

but segmentation did not help learners with high level of prior knowledge, thus suggesting a boundary effect for the segmentation principle.

What is the optimal duration?

The duration is the most obvious and global temporal property of a video. There is a wide consensus between researchers and practitioners that the length of an instructional video clip should be kept as short as possible. When surveyed about this matter, students also demand shorter clips rather than long ones (Mitra, Lewin - Jones, Barrett, & Williamson, 2010). Mayer's segmentation principle advocates to split the learning contents in short duration clips. Many guides for producing videos suggest durations of few minutes. The scientific rationale for this statement is that learner's attention and assimilation capacity tend to decrease over time. Research evidence from psychology supports this hypothesis: for instance, (Medina, 2008) provides evidence that suggests that learner's attention decreases after about ten minutes of passively listening to instructional material.

The problem with the short video duration is to give a measure of what can be considered 'short'. Many studies establish boundaries for what should be an adequate length for an instructional video, with ranges from 1 to 30 minutes (Kay, 2014; Morris & Chikwa, 2014; Thomson, Bridgstock, & Willems, 2014; Wetzel et al., 1993); but many of these studies support their claim in anecdotal evidence. An old advice for teaching films in schools appealed for ten to fifteen minutes in length (Sumner, 1950), a measure that would take roots on the length of a school lesson, or even the length of a standard 16mm reel. A more grounded conclusion comes from the study by Guo, Kim and Rubin (2014) on MOOC videos found that video length was strongly related to relative watching time. They estimated that the 'median engagement time' for all viewer was at most 6 minutes, regardless of total video length. The absolute watching time in long videos was lower than that of shorter videos. The authors suggested that 6 minutes could be an upper bound for an engaging video clip. More recent research (Ozan & Ozarslan, 2016) has confirmed this finding: short videos (less than 10 minutes) are more likely to be watched entirely than long length videos (more than 30 minutes).

Two recent technologies are shedding more light on this question: eye-tracking and video usage analytics. (Zhongling Pi & Hong, 2016) monitored undergraduate students when watching video lectures in various formats, finding that fatigue started in average at 11 minutes, and reached to a peak level at 22 minutes (fatigue level was measured with eye blink duration).

Apart from those results and suggestions about video duration, it should be noted that the optimal duration, whatever it is, will be different across presentation formats: a documentary should (and must) have a longer duration than an instructional video showing how to cook an omelette, just to mention one trivial example.

Coordination between expository content and film segmentation

The need for coherence between narration and visualization has been stressed by several authors (Morain & Swarts, 2012; Sugar, Brown, & Luterbach, 2010; van der Meij & van der Meij, 2013). Experimental evidence suggests that the video content should not only be segmented, but also the segmentation structure should be highlighted with pauses between segments and other temporal cues (transition shots, discourse markers, sound cues, etc.). (Spanjers et al., 2011) found that even short pauses of two seconds are beneficial to learners. Sounds can be used to enhance temporal composition, for example by using recurring sounds when a particular type of video segment starts or ends, or when a topic is introduced (Bishop & Sonnenschein, 2012).

Spanjers, van Gog and collaborators (Spanjers, Van Gog, Wouters, & Van Merriënboer, 2012) further investigated on the explanation of the segmentation effect in learning from animations. They found evidence for two plausible explanations: pausing and temporal cueing. By inserting pauses between segments, learners are given additional time to process the contents. Additionally, the segmentation makes explicit the underlying structure of the presentation, thus it provides a cue to better understanding the contents. In the practicioner's arena, Morain and Swarts, in their quality assessment rubric for instructional videos (Morain & Swarts, 2012), include these two items: 'natural breaks are included to allow viewer to pause' and 'audio announces step just slightly before the step is executed'.

6.5.7 Informational complexity

The presentation speed and the amount of information in the presentation flow may determine the learning effectiveness. A flow that is too fast will be less understandable, while an excessively slow pace can be boring and therefore distracting. In any case, the potential drawbacks of a fast pace can be partially moderated by the learner, thanks to the playback video controls, such as pause, rewind and speed control (Kay, 2014). See also Section 6.4.3 (page 148) for a discussion of playback video controls. In the following lines I describe the relevant studies from my review that inform about the presentation speed and informational density.

Presentation speed and density

Animation speed. It is not clear that showing a fast-paced animation is always harmful for learning. (de Koning et al., 2011) presented an animation at low, normal and high speed. Students showed equal performances on comprehension and transfer for both speeds (also for cued and not-cued versions). The authors hypothesized that a low speed may make it easier for learners to extract relevant parts of the animation, and, on the contrary, higher speeds may force learners to quickly decide which information requires attention, thus increasing cognitive load and hindering learning. By contrast, the experiment by Fischer, Lowe and Schwan (2008) using an animation

of a clockwork mechanism with two versions (normal and fast pace) revealed that students who watched the fast presentation produced more correct and less false concepts about the mechanism.

Speech rate. The research on effect of voice speed shows that most listeners can cope with very fast speech paces, compared to natural speed. A classical experiment on learning retention in audio lectures (Aiken, Thomas, & Shennum, 1975) found that learners performed better with a doubled speech rate of 240 words per minute (wpm) than with a normal speech rate of 120 wpm. Subsequent research has confirmed that untrained listeners can comfortably understand compressed speech up to double the natural rate (see references in Gade & Mills, 1989), being 240-275 words per minute an upper level of comprehensible speech rate for adult listeners. The ability to adapt to unusually high speed rates decreases with age (Simhony et al., 2014). Non-native learners may found challenging to listen to high speech rates, but they can be trained to overcome their limitations (Banai & Lavner, 2012). Apart from the direct effect on contents understanding, speech rate modulates the learner's affective perception of the speaker. Research has found that speakers with a faster speech rate receive higher perceived credibility and immediacy (Simonds, Meyer, Quinlan, & Hunt, 2006), which in turn fosters student motivation.

Information density in slides. In slide-based presentations, one design question is how much content to include and how much time is spent on each slide. The informational flow will be bounded by the narrated exposition, but the choice still remains as to how many slides will be used for a given content: many slides with very little content, or few slides with dense content? In each case, the presentation pace will be faster or slower to stay in synch with the narration. This issue was tackled by an experiment (Fish, Mun, & A'Jontue, 2016), in which two versions of the same instructional video lessons were compared: one version with 611 slides and a transition of 40 seconds between slides, and another with 1993 slides and 10 seconds transitions. The effectiveness was evaluated with student surveys. The overall rating was higher for the version with fewer slides. As the authors suggest, "too many visuals may negatively affect learning".

Cutting rate. The *spatiotemporal complexity* may be a relevant feature in instructional videos. Older studies suggest that expository formats should avoid excessive edition. Drew (1985) found that viewers of highly edited television news tended to focus on audio and ignored the discontinuity in the video. Wetzel et al (1993, p. 116) warn that "cutting a complex scene too rapidly prevents the learner from interpret the contents. On the contrary, too lengthy cuts may foster mind wandering". Unfortunately, I have not found recent studies on spatiotemporal complexity that explore Wetzel et al.'s postulate.

Proper timing and avoiding redundancy

The temporal order of appearance of the presentation items is paramount for optimal learning. As stated by cognitive theories of multimedia learning, "visual-textual information can be segmented without affecting the video duration by only

showing new information from the moment it is mentioned in the narration and becomes relevant" (Moreno & Mayer, 2002). This is called the *redundancy principle*. The delay between narration and visuals affects learning. The effect of narration-visual contiguity was demonstrated by Patricia Baggett (1984). She performed an experiment with different temporal offsets between narration and visuals. Seven versions were tested, with different positive and negative offsets between visuals and narration (7, 14, 21 seconds, and a version in synchrony). Presenting narration before visuals produced worse results, while the versions with full synchrony and with visuals presented 7 seconds before narration resulted in the best learning performance.

The redundancy effect has been noticed from early educational film studies. Northrop (1952) compared three versions of a film: one with no inherent organization, other with a logical development, and a third with a chronological story. Each version was edited with and without explanatory titles. Results showed that titles improved the learning for the film without inherent organization, but decreased learning in the other versions: redundancy hindered learning.

The redundancy principle recommends that multimedia learning objects present words as a narration and avoid using simultaneous text and narration. Nevertheless, for learners with perceptual hindrances it is useful that visual text accompanies spoken words. Evidence is showing that non-native learners in MOOCs tend to skip clips with narration only, and pause significantly more in portions that show simultaneous text and narration (Uchidiuno, Hammer, Yarzebinski, Koedinger, & Ogan, 2017). In the case of subtitles, Moreno and Mayer concluded that it is better to show verbal and written explanations *before* the visual information is shown (Moreno & Mayer, 2002).

Timing of instructor presence. Some researchers have been particularly interested in the timing of instructor presence. One reason is the intuition that continuous instructor's presence may be redundant and therefore a source of distraction, but at the same time it is convenient some human presence to enact positive affective responses in the viewer. In this way, self-presentation of the narrator/instructor has been proposed as a good practice for video tutorials (Morain & Swarts, 2012). This rhetoric pattern is considered an adequate balance between the personalization principle and the redundancy principle. Kizilcec et al. (Rene F. Kizilcec et al., 2015) hypothesized that a 'strategic' presence of the instructor, appearing occasionally in low-demanding moments, should be better than a continuous display. Their experiments found no significant effect in that method. Finally, Díaz et al. (Díaz et al., 2015) used EEG to measure the viewer's response to different instructor presentation modes in instructional videos. They tested three versions: instructor's picture always present, instructor absent (voice only), and instructor presented only at video start. Results showed similar values in emotional states, and a significant higher cognitive load in the third version (instructor presented at video start). One preliminary suggestion from these evidences is to use the instructor's image only if it is clearly justified, and in moments where it cannot be harmful in terms of cognitive load.

6.6 The Speech domain

Instructional video content is basically *text*. As I explained in Chapter 2, the term *text* is used here in the same sense as Semiotics and Discourse Analysis: any realization of human language, whether auditory or written. Text in instructional videos is highly structured. Instructional videos show common regular patterns that are identifiable across multiple genres. For instance, all meaningful instructional expository formats follow a regular rhetoric structure of introduction, development and closing. The regularity and formality of discourse in instructional video makes them closer to formal genres such as academic writing or corporate communication, though they still share some characteristics of spontaneous spoken language.

The regular patterns found in video discourse may foster learning by making the content more predictable. Some linguistic patterns in lectures, such as discourse markers, serve to guide the learner. Evaluation of the contents by the lecturer may increase learner's motivation. What is more, guidelines for designing instructional videos encourage to make the rhetoric structure explicit (Koumi, 2006; Swarts, 2012; van der Meij & van der Meij, 2013). Definitely, linguistic structures are relevant in instructional video design and therefore they deserve a deep analysis and characterization.

Discourse analysis is a convenient tool to characterize the structure of the text in instructional videos. Moreover, Systemic Functional Linguistics (SFL) provides a powerful theoretical framework to ground discourse analysis methods, as in O'Halloran's Systemic functional-multimodal discourse analysis (SF-MDA) (Lim-Fei & O'Halloran, 2014; O'Halloran, 2008). In fact, SFL and discourse analysis has been used to analyze classroom lectures (Lim, 2011). I will utilize the Discourse Analysis and the SFL framework to characterize the linguistic functions that have been found by multimedia learning researchers. This characterization is described in Section 6.6.2 . Before this, I will start in Section 6.6.1 with a summary of the literature review that I have performed in order to develop and support the characterization. Sections 0and following will describe in detail the main linguistic features that have been found to be relevant in learning with instructional video.

6.6.1 Review of research on instructional discourse

Classes of instructional videos

A top-level classification of discourse styles in instructional videos is between interactive and non-interactive discourses. Some instructional videos show a non-interactive discourse: monologic lectures, how-to demonstrations by a speaker, documentaries, etc. Other instructional videos show an interactive language: interviews, tutorial dialogues, and panel discussions, among others. The characteristics of the speech are fairly different between interactive and non-interactive formats.

As far as discourse structure is considered, we can distinguish two families of videos: conceptual-factual knowledge videos (lectures) and procedural knowledge videos (demonstrations, how-to videos). At present, there is more specific research on procedural video discourse than in conceptual-factual videos. There are numerous articles on the rhetoric structure of YouTube tutorials, demonstration and how-to videos (Morain & Swarts, 2012; Sugar et al., 2010; P. ten Hove & van der Meij, 2015; van der Meij & van der Meij, 2013). Quite the reverse, the analysis of discourse of online lectures is limited to pre-recorded classroom lectures. The 'recorded for video delivery' lecture has not been analyzed in depth. Therefore, for the case of conceptual knowledge videos, we have to rely on extrapolation of research made on classroom lectures, or analogous research on other online formats, such as TED Talks (Compagnone, 2015) and video blogs (Frobenius, 2014).

Discourse in classroom lectures

Classroom lectures have been studied for long with the lens of Discourse Analysis. Most lectures fit in one of three lecturing styles: reading style, conversational style, and rhetorical style (Fortanet Gómez & Bellés Fortuño, 2005). The reading style is the traditional in academic settings, with a lecturer speaking in a strongly monologic style. Conversational and rhetorical styles have gained ground over the last few decades: academic lectures have become more interactive and the audience is more involved in the realization of the lecture.

A landmark in the understanding of the rhetoric structure of academic lectures was made by Lynne Young (1994). Young found that the macro-structure of a lecture is composed of several discourse *strands*. Each strand goes through different *phases* such as 'discourse structuring', 'conclusion' and 'evaluation'. Strands may overlap over time. Young identified six classes of phases in the corpora she analyzed. Three of them are metadiscoursal (they talk about the discourse itself): *discourse structuring* (the lecturer announces new directions of the lecture), *conclusion* (the lecturer summarizes points previously developed), and *evaluation* (the lecturer evaluates information that has already been transmitted or is about to be transmitted). The other three phases in Young's model are: *interaction*, *theory/content*, and *examples*. *Interaction* is the phase where lecturer and their audience enter a dialogue, usually in a question-and-answer sequence. *Theory* or *content* phase is the presentation of the content of the lecture. Finally, the *examples* phase occurs when the lecturer provides concrete examples that give support to their content. These three strands are often interspersed as the lecture progresses.

Young's phasal model has been validated or modulated by other corpus-based studies. For instance, Deroey and Taverniers (2011) analyzed a sample of lecture corpus and identified these six functions in lecture discourse: *informing*, *elaborating*, *evaluating*, *organizing discourse*, *interacting*, and *managing the class*. Moreover, they decomposed these functions into 18 subfunctions.

Characteristics of academic lectures

I have performed a review of recent research of discursive patterns in academic lectures that may have influence in the translation of discourse to the video modality. The outstanding features are summarized below:

- Extensive use of meta-discourse markers (discourse about the ongoing discourse, for example to announce what topic is about to be introduced): (Ädel, 2010; Alharbi, Ng, & Hain, 2015; Bernad-Mechó, 2015).
- Use of interaction features such as questions, even when the lecture style is highly monologic (Crawford Camiciottoli, 2008). Lecturer's discourse displays typical features of conversational language: incomplete clauses, pauses, false starts, redundancies, and discourse markers (Flowerdew & Miller, 1997).
- A significant participation of nonverbal language (gestures and gaze) in discourse meaning-making (Crawford Camiciottoli, 2007; Lim, 2011).
- The lecturer changes their discourse depending on audience and classroom size, e.g. (S. W. Cheng, 2012; J. J. Lee, 2009).
- Extensive use of humor (irony, self-deprecation, refer to salient events, etc.) (D. Lee, 2006; Nesi, 2012).
- The lecturer reflects her/his tenor with the students (Molina Plaza & Argüelles Álvarez, 2013).

From face-to-face lectures to studio-recorded video

There are marked differences between face-to-face lectures and studio-recorded video lectures. Canned video lectures lack the dialogic units of face-to-face lectures. Hence, the act-move-exchange model (Sinclair & Coulthard, 1975) is almost useless in monologic expositions, and neither is the common Initiation-Response-Feedback (IRF) exchange. Other discursive difference is a lower spontaneity. As J. J. Lee points out (2009), "while [classroom] lectures may plan and most likely utilize their notes in the delivery of lectures, these communicative events are nevertheless performed in real time". This is not the case of studio-recorded lectures, in which several takes can be recorded for a given shot, until an error-free take is obtained.

One principal difference between a classroom lecture and a studio-recorded lecture is the absence of a live *audience*. This change of setting causes changes in the speaker's rhetoric. For example, Frobenius (2014) has found that video bloggers (*vloggers*) are aware of their future audience and make choices of semiotic resources that are suited to the new medium: questions, personalization, gestures, gaze, and pointing gestures to locate areas in the viewing interface shared with the audience. Also, as the lecture speakers do not know in advance what will be the exact configuration of their audience, they will tend to make worst-case assumptions to broaden the efficiency of the contents. For example, if they assume a large, nonnative audience, they would tend to speak at a slower pace and use fewer redundancies (Barrett & Liu, 2016).

Despite the differences between face-to-face and studio-recorded modalities, some theoretical models can still be applicable. There is evidence to postulate that Young's phasal model (L. Young, 1994) is valid for canned video academic lectures, although this claim has not been properly assessed. Youngian phases are present in video lectures, though the relative weight of certain phases changes if compared with the face-to-face modality. The 'Interaction' phase is usually substituted by self-evaluation prompts or task proposals.

Rhetoric structure of instructional videos

Jack Koumi's framework for educational video design (Koumi, 2006, 2015) consists of 'pedagogic design principles' that comprise a catalogue of 25 discursive structures. Those principles are grouped around eight categories:

1. Hook. Make learner want to know. Capture attention. Sustain interest.
2. Signpost. Introduce scene, what's next, what to look out for...
3. Facilitate cognitive engagement. Pose questions, encourage prediction...
4. Enable construction of knowledge. Pause commentary for contemplation, invent visual metaphors...
5. Sensitize. Personalize the teacher, signal change of mood/topic...
6. Elucidate. Vary tempo to indicate syntax, enhance legibility/audibility...
7. Reinforce. Repetition, re-exemplify, compare/contrast...
8. Conclude/consolidate. Recapitulate, summarize key features, chapter ending...

Examples of rhetoric patterns within Koumi's organization are 'chapter heading: what's next?', 'pause commentary for contemplation', and 'summarize key features', to name a few.

Much research has been done in the rhetoric structure of procedural videos. Researchers on online video tutorials works have grounded their work on principles for technical documentation writing (Farkas, 1999) and for information design (Carliner, 2000). According to Farkas, procedural discourse is structured in three phases: *explanation*, *demonstration* and *doing*.

The screencast format has received specific attention from video design researchers. Sugar and others (2010) examined the presentation structure of a set of screencast tutorials. They found some common structural patterns in discourse organization: bumpers (opening and closing messages), overview, describe procedure, present concept, focus attention, and elaborate content. Loch and Mcloughlin (2011) proposed design guidelines for self-regulated learning by using screencasts. These guidelines include to provide an overview, activate prior knowledge, ask students to set learning goals, present questions and tasks, encourage students to reflect, ask students to self-assess their performance.

Swarts and Morain developed an assessment rubric for the quality of online video tutorials (Morain & Swarts, 2012; Swarts, 2012). They found that good quality videos exhibit certain exposition patterns: begin with an overview of the task ahead; add non-essential details to clarify and add perspective, but do not abuse of them;

make statements to build confidence about the speaker; make persuasive appeals to the learner; create and fulfill learner's expectations. In addition, Swarts points out that the rhetoric structure of video should be visible and persistent, to facilitate learner's navigation (Swarts, 2012).

Van der Meij and van der Meij (2013) proposed eight guidelines for designing procedural video tutorials. Some of these guidelines make advice on rhetoric patterns. The first guideline, 'craft the title carefully', advocates for a title that guides the user and provides a succinct description of the goal that is demonstrated. The second guideline, 'promote the goal', includes a preview of the task to orient the user.

Kay (2014) developed a framework for creating worked-example video tutorials. Kay's framework contains a series of principles to follow when designing an effective tutorial. Among these principles, there are some rhetorical patterns:

a) Establish the context: at the beginning of the clip: clear problem label, background information; key elements are clearly articulated before trying to solve them; highlight key features that learners should attend to
b) Create effective explanations: problem is broken down into meaningful chunks; the reason for conducting each step is explained.
c) Minimize cognitive load: the important elements are written down as needed (not all at once).

An interesting point of view on the "establish context"/ "signpost" rhetoric stage was formulated by Thomson, Bridgstock and Willems (2014): consider not placing the signposting in the video itself, but in other surrounding object (e.g. in the container web page). They reason that "the less video-based verbal or written commentary/explanation around the core learning message, the greater the engagement; and also, removing 'administrative' elements from the video such as explanations of associated assessment items means that the footage may be reusable in other learning contexts and/or for other subjects.".

The lack of previous social interactions between lecturer and audience to show and negotiate their respective social status make it convenient to insert opening and ending sequences in all video lecture clips. This opening and closing phases have been identified as 'bumpers' in Sugar et al. analysis of screencasts (Sugar et al., 2010).

Supplying interaction

One-on-one tutoring is perhaps the best instructional design for effective learning, far away from lectures (Bloom, 1984). One of the virtues of one-on-one tutoring is the rich interaction that flows between the tutor and the learner, and the ability to adapt the tutorial contents to the individual learner's characteristics. Studio-recorded videos lack most of the interactivity and adaptability that is needed to implement tutorial dialogue. Nevertheless, developers of instructional videos often try to overcome this limitation by adding some interactive, dialogic features to the instruction discourse. The interactive features that I have identified in the present study are:

- Record dialogic speech.
- Add questions to monologue.
- Add self-explanation prompts.
- Add interpolated tests and tasks.

I will discuss these quasi-interactive features in a following section.

6.6.2 A taxonomy for the Speech domain

The analysis of instructional discourse shows two levels of spoken/written language that are relevant in the learning effectiveness of instructional videos. First, we have the overall organization of discourse, the discourse *rhetoric*, which can be roughly defined as the logical concatenation of small pieces of verbal information (rhetoric *stages*). Second, we have the language *register*, which is the concrete configuration of multiple language functions for a given situation. According to Halliday's Systemic Functional Linguistics, the language register is organized around three variables, named *field,* mode and *tenor*. My review of research on instructional video finds that the SFL concept of language register (field, mode, tenor) is a useful device to organize the linguistic features that have been found relevant in learning effectiveness.

The following sections describe two taxonomies, the first for classifying rhetoric phases and the second for classifying language features around the SFL theory.

Rhetoric structure: strands and phases

Instructional videos exhibit a very regular rhetoric structure. In fact, all kinds of instructional delivery usually follows a sequence of introduction, body and closure (Smith & Ragan, 2005). Both conceptual-factual (lectures) and procedural (demonstrations, how-to videos) are often organized in a sequence of episodes or topic expositions (discourse strands). In turn, a typical structure for an episode or topic is an introduction, an elaboration, some kind of discussion/evaluation/task assignment, and a conclusion. This internal structure is composed of several *phases*, that is, one or more sentences with a main communicative goal and with a characteristic linguistic profile.

In summary, the macro-level organization of discourse in instructional videos can be described in terms of two layers, as in (L. Young, 1994): one first layer of 'episodes' or 'discourse strands', and a second layer of rhetoric 'phases'. In the following, I will propose a taxonomy for classifying the rhetoric phases.

In the light of the reviewed literature, a synthesis proposal can be made for classifying instructional video phases, which is outlined in Table 6-10. Rhetoric phases of instructional videos can be grouped in four broad categories, according to their respective function in the overall rhetoric design: *organize discourse*, *communicate content*, *query the learner* and *engage the learner*. The *organize discourse* function is, as in Young's classification (L. Young, 1994) and as the *signposting* in Koumi and others (see above sections), for making explicit the organization of the discourse phases,

signpost topic transitions, or initiating/closing a discourse strand. The *communicate content* function satisfies the essential goal of the discourse, which is to expose some kind of knowledge. This function in turn contains a number of subfunctions, such as informing, elaboration and evaluation functions in (Deroey & Taverniers, 2011). I prefer to group these three areas in a single category, to hold the model simplicity. The *query the learner* function is a surrogate of the interaction functions in other models: it collects all rhetoric phases where the speaker raises questions or tasks addressed to the audience. Finally, the *engage the learner* function collects all rhetoric phases whose purpose is to activate the learner interest in the content, for example by asserting expectations on its usefulness.

Table 6-10 shows representative examples of rhetoric phases associated to each rhetoric function. Those examples are collected from the reviewed literature or have been found in my own study on MOOC videos. They cover a wide range or actual realizations of instructional video discourse, but they cannot be considered an exhaustive list, since the variety of rhetoric phases is only limited by the creativity of the instructional designer.

Table 6-10. Taxonomy of rhetoric phases in instructional videos

function in discourse	common rhetoric phases
organize discourse	opening / closing shot overview of the contents explain pre-requisites and context relate to other contents announce following section rhetoric pause summarize contents
communicate content	theory / content demonstration / task execution example reformulation evaluation: indicate attitude evaluation: indicate commitment
query the learner	ask to recall/repeat exposed content ask to perform tasks ask for reflection and transfer
engage the learner	hook (capture attention) justify/motivate content build confidence/authority in speaker create and fulfill learner's expectations

Language register: mode and tenor

Table 6-11 associates the findings from multimedia learning with a subset of SFL metafunctions and functions. The first two columns respectively show SFL

metafunctions and functions. The third column shows features that research on multimedia learning has been found relevant in learning effectiveness. These association, thus, can be considered a mapping between Systemic Functional Linguistics and Cognitive Theory of Multimedia Learning.

Some remarks must be done about Table 6-11. First, the ideational/experiential metafunction (field) of SFL has been omitted in this table, because it does not provide value for this classification scheme: it just describes the subject matter of the instructional video, which is not a choice in discourse design or discourse realization. Second, apart from the classic SFL functions, I have added the *appraisal* function, borrowed from the Appraisal Theory (J. R. Martin & White, 2005), to account for the large amount of evaluation that instructional discourse contains.

Table 6-11. Properties of instructional videos as Systemic Functional Linguistics functions

metafunction	function	instructional video properties
textual (mode)	spoken/written	spoken vs. written text
	action/reflection	spontaneous vs. acted speech
	interactivity	monologic vs. dialogic questions and prompts
interpersonal (tenor)	speech function	statement, question, offer, command
	social distance	conversational vs. formal style politeness humor
	personalization	personalization (addressing to 2nd person)
	standing	authority claims
	appraisal	attitude, engagement, graduation
	stance	modality: epistemic vs. deontic uncertain vs. confident narrator

6.6.3 Dialogic videos

Beyond the common monologic lecture video, we can find several types of educational videos that record dialogues. The most frequent dialogic video styles are the following:

- Tutorial dialogue (tutor-tutee).
- Team lecturing (dyadic lecture).
- Interview to an expert (interviewer-interviewee).

- Panel discussion (similar to an interview, with more participants).

The following paragraphs introduce those dialogic modalities and the alleged benefits in the learning process.

Observational learning and tutorial dialogues

It has been proposed that people learn from watching other people behaviors. For example, Bandura coined the term *vicarious learning* (Bandura, 1962) to refer how people learned some behavior from watching videos of that behavior. The vicarious learning theory has broadened into an instructional method by which the student learns from watching a learning situation where the learner cannot interact. This learning situation is usually recorded in video. A novel approach for instruction is 'learning by observing others learn': record tutorial dialogue and show to other learners to watch the recorded video. This approach has been found superior to watching monologue tutorials.

Observational learning has been extensively studied in tutorial dialogues. In a common setting, a student watches a recorded session of tutor-student dialogue. (Muldner et al., 2014) in two studies, they compared collaborative observation of recorded dialogue with two other treatments: one-on-one tutoring and collaborative observation of monologue. In both studies, the observation of dialogue outperformed the observation of monologue in learning outcomes. (Chi et al., 2017) reported a series of experiments that explore the differences between watching dialogue and monologue videos. The authors have proposed a novel approach for instruction: record tutorial dialogue and show to other learners to watch in dyads, so they will be able to solve the activities collaboratively. They have found that students learned more when they collaboratively watched tutorial dialogue-videos than lecture-style monologue-videos. The authors also have found that the results are similar with expert and non-expert recorded tutors, therefore validating the scalability of this approach.

The discussion of misconceptions in the classroom has been shown to enhance learning. This method has also been tested in non-interactive, video instruction. (Muller, Bewes, Sharma, & Reimann, 2008) compared four presentation formats for an instruction on Physics: expository lecture, expository lecture with additional interesting information, exposition with refutation of misconceptions, and student-tutor dialogue with refutation of misconceptions. Results showed that the two refutation modes outperformed the non-refutation modes as regards learning retention. Also, (Muller et al., 2008) found that watching a video of student-tutor dialogue with discussion of misconceptions was superior than watching an expositive lecture, with a strong effect size in learning retention (d=0.83). Low-skilled learners were most benefited by the dialogic format.

Team lecturing

Dyadic lecturing, though uncommon, has been used in college education for a long time (Dey & Low, 1968). Dey and Low described a lecture with two instructors,

where the instructors exchange roles as the lecture progresses (lecturer, observer, expand upon the other lecturer, etc.). Several modalities of collective lecture teaching have been documented, including guest-host pair lecturing (Yanamandram & Noble, 2006), 'driver and navigator' roles (Burden, Heldal, & Adawi, 2012), and team lecturing with a student assuming the role of a teacher (student teacher) (Gray & Halbert, 1998). Researchers and practitioners report several advantages in team lecturing. It remains an open question whether watching a dyadic lecture video provides some benefit.

Despite the possibilities of dyadic lecturing, it is scarcely used in MOOC videos. My study accounted for only three courses exhibiting two simultaneous instructors (*Introduction to Public Speaking* in Coursera, *Cosmology* in edX, and *The Science of Medicines* in FutureLearn).

Figure 6-10. Screenshot of a dyadic lecture
From *Cosmology* MOOC (ANUx). Reproduced with permission (CC BY).

6.6.4 Asking the learner

In general, instructors strive to keep certain level of interactivity in their discourse, even when there is no visible audience. In her contrastive analysis of face-to-face academic lectures versus written text materials, Crawford Camiciottoli (2008) found that the frequency of questions was 'strikingly similar' in both modalities. It is expected that instructional videos are following this trend, although it is something to be confirmed.

It is well known that inserting questions in an exposition enhances learning. The effect has been found even in simple interventions, such as inserting multiple-choice questions slides in an offline PowerPoint lecture (Campbell & Mayer, 2009). Moreover, the introduction of deep-level-reasoning questions in the exposition enhances learning (Craig, Sullins, Witherspoon, & Gholson, 2006), when compared to monologue or shallow questions.

Loch and Mcloughlin (2011) proposed an instructional design model for screencast videos in which questions take an essential role. According to this model, questions

may be used in three stages of self-regulated learning: planning (for activating prior knowledge), monitoring processes (for understanding ongoing concepts and engaging in the problem-solving process), and reflection on self-knowledge and task achievement.

Self-explanation prompts

A special case of questions are the *self-explanation prompts*. In this type of question, the instructor asks the audience some question to be explained by the learner on her own. There are two principal classes of self-explanation prompts: *prediction* prompts, e.g. 'what will happen next?', 'will x be greater than zero?'; and *reflection* prompts, e.g. 'why this has happened?', 'could you explain that with your own words?' Self-explanation prompts foster reflection on learners, as verified by Moreno and Mayer (2005). Self-explanation prompts can be considered instances of conversational style, therefore they are linked to the personalization principle.

Eliciting learner's self-explanations of contents has been demonstrated to improve understanding of the instructional contents (Chi, De Leeuw, Chiu, & Lavancher, 1994). Lin et al. (2016) studied the combined effect of visual cues and self-explanation prompts in a multimedia lesson. Combined cues increased learning outcome (retention) and had positive effect on cognitive load and learner's motivation.

6.6.5 Interpersonal function of language

Mood in instructional discourse

In the successive stages of a common instructional video, the speaker uses all four Hallidayian moods: offer, statement, question, command (Michael A. K. Halliday & Matthiessen, 2014). For example, at the start of the lecture, the speaker is in the 'offer' mode (giving goods and services); in lecture development, we can find the speaker in 'statement' mood (communicating declarative knowledge) or 'command' mood (giving instructions). When closing a topic, it is frequent to enter 'question' mood (interpolated questions to the audience). An appropriate balance of moods would serve to guide the learner across the lecture/tutorial, and to establish what phase is being realized.

Standing: authority claims

Morain and Swarts' assessment rubric for video tutorials (Morain & Swarts, 2012) includes an objective called 'confidence', defined as "Narrator inspires confidence by presenting self as knowledgeable and skilled and also emphasizes association with reputable organizations". Their analysis of YouTube tutorials found that several good quality videos show the narrator making overt claims to credibility by referencing a company sponsor or by noting credentials.

Appraisal

The appraisal function (J. R. Martin & White, 2005) is extensively used in instructional discourse. Frequently, a lecturer utters judgments and expectations about the topic she/he is explaining, about the task being performed, about the audience, and many other forms or appraisal. Some researchers have paid attention to the effect of appraisal in instructional video. For instance, Morain and Swarts' assessment rubric for video tutorials (Morain & Swarts, 2012) includes two objectives called 'self-efficacy' and 'engagement', which include evaluative statements about the contents and expectations about the viewer/learner involvement in the task. The personalization principle, which advocates for a conversational style (Moreno & Mayer, 2000b), also promotes to make (positive) judgement about the learner.

Personalization

The personalization principle states that people learn better when words are presented in conversational style rather than formal style (Mayer, Fennell, Farmer, & Campbell, 2004). In the context of multimedia learning theories, *conversational style* means that first and second persons are used, and that comments are directed to the learner. *Formal style* means that third person forms are used, and no comments are directed to the learner. Moreno and collaborators demonstrated the personalization principle in multiple experiments (Atkinson et al., 2005; Moreno & Mayer, 2000b). In Moreno and Mayer's experiments, the personalization condition featured first person sentences (self-referencing), second person sentences, as well as a conversational style (e.g. 'what do you think if...?'). The impersonal condition used third-person sentences and monologic discourse. A meta-analysis on instructional text design (Ginns, Martin, & Marsh, 2013) showed moderate or large effects of personalization in learning outcomes.

Kartal (2010) went deeper into the elements of personalization principle, and tested the separate effects of personalized grammar and text formality, in an experiment with Turkish students. He found that both personalized and informal interventions produced a large effect in learning retention.

The personalization effect can be moderated by learner's social and cultural framework. For instance, an experiment tested the combined effect of audiovisual modality and personalization between young and old learners (Bol, Van Weert, De Haes, Loos, & Smets, 2015). The study found no significant personalization effect alone, but personalization enhanced the effect of audiovisual modality. This combined effect was higher in older learners. In the case of aversive content, personalization may lead to an inverted effect in learning (Kühl & Zander, 2017).

Politeness

The level of politeness of language expresses the interpersonal situation between the speaker and the viewer. The difference between 'now you have to learn about photosynthesis!' and 'let's find out something about photosynthesis'. Learners are

aware of the degree and direction of politeness in instructional sentences in a multimedia learning environment (Mayer, Johnson, Shaw, & Sandhu, 2006). It has been proposed that polite speech would contribute to better learning than direct instructions. Such politeness effect was observed in a series of studies, that showed that learners obtained better outcomes in a polite version of a computer tutorial (N. Wang et al., 2008). The politeness effect has been also demonstrated by Schneider, Mebel and Pradel (Schneider, Nebel, Pradel, & Rey, 2015), who found large effect sizes of polite instruction texts in both learning retention and learning transfer.

Spontaneous vs. acted speech

Wetzel et al. (Wetzel et al., 1993), citing the work of Baggaley and Duck (1975), stated that "Baggaley and Duck [...] further suggest that a speaker is best presented as an independent source of information rather than a mere conduit of another person's message. Consequently, an obviously rehearsed, scripted or edited presentation is likely to be perceived as being less reliable and credible than a less polished, but more natural presentation.". Evidence suggests that there must be a balance between spontaneity and formality: the speech has to be rehearsed previously, in order to avoid visible hesitations and digressions that would hinder credibility. But if the speech is too rehearsed and too formal, it would appear to be socially distant and consequently less engaging. Swarts came to this same conclusion after analyzing the effectiveness of YouTube tutorials (Swarts, 2012).

Humor

Several theories have been proposed to justify humor in teaching. For example, the instructional humor processing theory (IHPT) proposes an explanation of how humor may facilitate learning: "students need to perceive and then resolve the incongruity in a humorous instructional message. Further, the IHPT proposes that the recognition of humor will increase students' attention" (Banas, Dunbar, Rodriguez, & Liu, 2011). One social theory that favors the use of humor in the classroom derives from Sigmund Freud's Relief Theory[13], according to which humor is a human device for relieving tension and disguise aggression.

A meta-analysis of the effect of humorous lectures in learning (D. M. Martin, Preiss, Gayle, & Allen, 2006) found a positive, moderate effect of humor in learning, but also found that this effect may vary depending on factors such as the type of knowledge being learnt and the teacher's gender. The meta-analysis of Banas et al. (Banas et al., 2011) identified 22 types of humor, of which 7 were considered appropriate for classroom use, and 13 were considered context-dependent.

In the academic realm, humor in academic lectures may serve to "illustrate an argument, state a point more memorably, establish rapport, create a sense of informality or enliven the atmosphere" (D. Lee, 2006). Hence, humor may release social stress in the classroom. Whatever is the cause, the fact is that humor and laughter are very frequent in academic discourse: 144 of 152 analyzed speech in

[13] See https://plato.stanford.edu/entries/humor/#RelThe

MIBASE corpus contained at least one "laugh" event (D. Lee, 2006). Lectures contained in average 0.21 laughs per minute. In the British BASE corpus, 136 out of 160 university lectures contained at least one laughter event (Nesi, 2012).

Some studies have addressed humor specifically in educational video. Wetzel, Radtke and Stern's review on educational television (Wetzel et al., 1993) reports claims from authors of the second half of 20th century, who considered that humor in instructional media should be discouraged because it creates distraction, a "playful" state not able to process instruction, and dissatisfaction when learner later watches non-humorous presentations. Other authors argued that humor reduces anxiety and increases attention (1993, p. 110). Similarly, Aagard's doctoral dissertation on humor and instructional video (Aagard, 2014) contains several references to studies on the humor effect on learning within video lectures, also with mixed results. Related to studies on seductive details, (Sitzmann & Johnson, 2014) found evidence that in instructional videos, jokes and other irrelevant but seductive details may foster learning by increasing engagement, but hinder learning by increase of time dedicated to the task.

A remarkable study is Kaplan and Pascoe (1977), who found that adding humorous examples to a videotaped undergraduate lecture did not improve short-term recall, but increased retention after several weeks. They found that the retention occurred in test items that were related to the humorous examples.

Aagard (2014) conducted experiments in which two versions of a video-recorded university lecture were produced: one neutral and other humorous. An experiment was made to find out whether humor had an effect on students learning and motivation. The results found no significant differences in learning and motivation between humorous and neutral conditions.

In contrast to face-to-face academic costumes, my study of MOOC instructional videos (see Chapter 3) revealed that humor was almost completely absent: only four courses contained videos that were humorous or contained salient humorous moves. The humoresque moves were also scarce. It seems as if the lecturers, who routinely use humor in classroom, refrain to record humor for their online materials. It is possible that video authors perceive that the social benefits of the Relief Theory do not apply to online video, since there is no live audience that shares the laughter. Other explanation is that authors prefer not taking the risk of making faux pas with humor, given that the recorded modality does not allow to fix misunderstandings.

The conclusion of this discussion is that humor should be considered as a relevant feature of this taxonomy, as it has a potential effect on learning. And, of course, a thesis containing some humor is a nicer thesis (Gromenauer & de la Calzada, 2017).

6.7 The Social Appearance domain

The social appearance in an instructional video goes beyond a mere collection of actor's attributes. The social appearance is a complex construct in which the video designer projects a social image of the visible agent, a social image that in turn causes a response on the viewer. This response may affect processes such as learner's motivation (Baylor, 2011), the speaker's perceived credibility (John Baggaley et al., 1980) and the learner's involvement in the video discourse (Mayer et al., 2003).

6.7.1 Social response theories

The social effect in digital media as well as in instruction has been observed for a long time. Several theories have tried to explain this behavior and to measure its influence on learning. Some outstanding theories are the parasocial interaction theories, the vicarious learning theory, the social presence theory. I will give a brief review on these theories before proposing a taxonomy of social appearance features.

Parasocial behavior

When a pedagogical agent is included in a multimedia object, learners tend to attribute social qualities to the device and develop a social interaction. The *parasocial interaction* was first defined by Horton and Wohl (Horton & Wohl, 1956), as a one-sided, non-directional, conversational 'give and take' between observers and media figures, characters and entities. In their bestseller book *The Media Equation* (Reeves & Nass, 1996), Byron Reeves and Clifford Nass described multiple situations where people, when interacting with electronic devices, attribute them human characteristics and treat them as real people. Nass later contributed to spread the term *Computers as Social Actors* (CASA) to define this unconscious social response of people when interacting with computer devices (Nass & Moon, 2000).

Social presence theory

Gunawardena and Zittle (Gunawardena & Zittle, 1997) characterize *social presence* as a construct that comprises a number of dimensions relating to the interpersonal contact. They mention the concepts of *intimacy* and *immediacy*, which are achieved by features as physical distance, eye contact, facial expression (e.g. smiling), dress, and voice inflection. Immediacy is the psychological distance that the speaker puts between her/him and the listeners. The social presence effect has been demonstrated in several experiments. For instance, the Presentation Domain section in this chapter discusses some evidences of the effects of displaying the instructor in the video frame.

Observational learning

Bandura coined the term *vicarious learning* (Bandura, 1962) to designate how people learn a behavior from watching videos showing that behavior. The way that vicarious learning operates is tightly related to social responses in learners. A visible, human-

like pedagogical agent can be seen as a **virtual social model** to be imitated by the learner. As Baylor points out: "one can acquire the behaviors or expertise mediated thorough a social model using processes such as observation, vicarious experience and social interaction. Research indicates that the most effective social model is similar to the observer while representing someone whom the observer aspires to be like. Consequently, one of the most important attributes for a social model is *appearance:* how s/he looks with respect to age, status, attractiveness, and credibility". (Baylor, 2011). Peer agents, or agents perceived as similar to the learner, may capture more attention: "according to the **similarity-attraction hypothesis**, learners may be more attracted to animated pedagogical agents that are similar to them, especially with regard to gender" (Kervellec et al., 2016).

Social agency theory

In the context of multimedia learning, Richard Mayer and collaborators have proposed that this parasocial behavior can enhance learning, in what is called **social agency theory**: "The main thesis in social agency theory is that social cues in a multimedia message can prime the social conversation schema in learners. Once the social conversation schema is activated, learners are more likely to act as if they are in a conversation with another person." (Mayer et al., 2003).

Limitations of research findings

There are some factors that limit the generalization of experimental findings of viewer responses to social appearance in learning environments.

First, the social response is modulated by the viewer's cultural framework. Cultural differences affect all production levels, from low-level features such as color, to more subtle characteristics such as icon semantics (Jaimes & Dimitrova, 2006; Jaimes, Sebe, & Gatica-Perez, 2006). Nevertheless, some general inferences about learners' social response can be extracted from experimental research in Western countries that are potentially extensive to other cultural environments.

Second, most recent research on social appearance has been made on computer-generated agents, in many cases with a low presentation quality. The extensibility of these research findings to human agents, or to highly realistic computer renderings should be checked. Some experiments are being replicated with more modern technologies, for example with voice (Craig & Schroeder, 2017).

6.7.2 Taxonomy of social appearance properties

Based on the theoretical framework and the experimental evidence that I have collected in my review, I have developed a taxonomy of social appearance characteristics, as shown in Table 6-12. The characteristics are grouped in five classes: *role, realism, fluency, social distance,* and *social group.* I have tried to ensure that the classes correspond to features that are not too much dependent on subjective judgment: hence, I have avoided properties such as 'attractiveness' or 'intimacy'. In

the following I will enumerate the references to studies that have explored each of these five classes.

Table 6-12. Taxonomy of social appearance properties

class	description	examples of properties
role	the actor may play different roles in the exposition	narrator, instructor, model
realism	the actor/scene may be rendered in a realistic or synthetic way	voice: robotic vs. human picture: computer-generated, cartoon, natural
fluency	the actor/scene can be perceived as easy to understand and self-confident	native vs. foreign accent, speech rate, speech fluency, visual addressing, gesture-speech synchronization
social distance	the actor can be perceived as socially 'close' or 'far' with respect to the viewer	display: shot length language: personalization, formality, politeness
social group	a set of observable attributes that situate the actor in a social group	gender, age, ethnicity, social affiliation, language register/dialect

6.7.3 Evidences from the research

Social role

An actor may play different roles in the instructional video: it may be an instructor (usually, an expert who explains a content at which she/he is highly capable), a narrator (a neutral speaker who communicates a content as given by others), or a model (a person that performs some activity or declares a personal experience). The actor's role frames the social appearance and may influence the learner's attitude. This matter has been explored by Nugent, Tipton and Brooks (1980), who studied what presentation format was best for promoting affective learning via educational television. Four presentation formats were tested: dramatization, on-camera host/narrator, authority/model testimonial, and visuals with narration. The same content was produced in the four formats. Results showed that the 'authority/model testimonial' was superior in both viewer's appeal and in affective learning. Similarly, (Töpper, Glaser, & Schwan, 2014) tested two versions of a multimedia learning environment about prostheses, in a within-subjects design. One version was preceded with videos with testimonials of users of prostheses. A social cue effect was found in the knowledge acquisition.

Realism

The humanity of pedagogical agents has sparked interest between multimedia learning researchers. Features that have been assessed include gesturing, voice quality and picture appearance.

Human-like appearance. Frechette and Moreno (2010) compared various agent presentation formats: static, with deictic movements, with facial expressions, and with both deictic movements and facial expressions. The static agent was superior than the agent with only facial expressions, and no other significant differences were found. A meta-analysis (Yee, Bailenson, & Rickertsen, 2007) found that the effect of just adding an agent is larger than the effect of animating the agent to behave more realistically. The effect of non-human agents in learning has been tested. (J. Li, Kizilcec, Bailenson, & Ju, 2015) found that a robotic appearance is less appealing than that of a human-like agent.

Robotic voice. Several experiments have confirmed that people learn better from instruction given by a human-like voice rather than a robotic voice (Atkinson et al., 2005; Mayer et al., 2003). Today, robotic voices are seldom used in educational videos. Advances in text-to-speech technologies enable today to synthesize voices with near-human quality, hence there is no need to resort to robotic voices. The voice effect has been reanalyzed recently, showing that state-of-art synthetic voices no longer produce a difference in learning (Craig & Schroeder, 2017).

Actor's fluency

The *fluency* comprises several properties that contributed to the perceived easiness of the content flow: seamless speech, appropriate gesturing and pose, direct addressing, native accent, etc.

Addressing. Includes eye contact and camera angle. (Beege, Schneider, Nebel, & Rey, 2017) studied the addressing effect in video instruction. They controlled addressing (near vs. far) and orientation (frontal vs. lateral) of an onscreen lecturer. They found a large orientation effect for retention performance, with higher learning outcomes for frontal orientation. Proximity did not show a significant effect.

Standard accent versus foreign accent. (Mayer et al., 2003) showed that learners performed better with native voices, instead of voices with a foreign accent. They proposed a *voice effect* that included also the robotic vs. human voice. (Sanchez & Khan, 2016) showed that a speaker with a foreign accent changes students' appraisal about learning difficulty, but does not impact effective learning.

Speech rate. Speakers with a faster speech rate receive higher perceived credibility and immediacy (Simonds et al., 2006).

The influence of instructor's fluency in actual learning has been questioned by some researchers. Several experiments performed by Carpenter and collaborators (S. K. Carpenter, Wilford, Kornell, & Mullaney, 2013; Toftness et al., 2017) have compared 'fluent' instructors with 'disfluent' instructors in educational videos. For example, in the first experiment (S. K. Carpenter et al., 2013), the fluent instructor

stood upright, made direct eye contact, used balanced gestures, and produced seamless speech, while the disfluent instructor hunched on a chair, broke eye contact, read from notes, and spoke in a halting manner. Participants who watched the fluent instructor self-rated their learning outcome higher than the non-fluent group, but no significant difference in learning retention was found between both groups.

Physical social distance

The graphic representation of a displayed actor is considered to generate different social responses on the viewer, according to the perceived spatial relation between the viewer and the actor. According to scholars, parameters such as the shot distance (close-up, long shot), the camera angle and the actor's address would result in a perceived social distance, power relation and interaction, respectively (van Leeuwen, 2008, pp. 138–142). Parameters such as the shot length and the camera angle may have an impact on speaker's perceived credibility and social distance (John Baggaley et al., 1980; Landström, 2008). However, I have not found recent and conclusive experiments on the effect of physical social distance in learning.

Actor's social group

A pedagogical agent, or a model, will often show visible traits of belonging to a social group, particularly through properties as gender, age and ethnicity. This group of appearance features has been studied in several studies.

Studies on instructor social group start as early as 1950 with the experiment by Kishler (1950), that revealed how the religious affiliation has an effect in the perceived prestige attached to the speaker in instructional films. In the context of computer-generated agents, Baylor (Baylor, 2009) observed that learners tend to be more influenced by an agent of the same gender, age and ethnicity.

As regards gender and age, (Hoogerheide, van Wermeskerken, Loyens, & van Gog, 2016) studied the influence of gender in instructional video tutorials. They tried a 2×2 design with female and male models viewed by female and male observers. Results showed no effects of gender on learning and near transfer, but there were some gender differences in self-reported results: male models enhanced perceived competence, and male students reports showed that learning from a male model was less effortful and more enjoyable. Studies from newscasts (Weibel, Wissmath, & Groner, 2008) show that newscasters' gender and age affect to their credibility, depending on audience gender and age. Also, (Meseguer-Martinez, Ros-Galvez, & Rosa-Garcia, 2017) found that Microeconomics YouTube video tutorials with a female speaker receive more 'likes' than those with a male speaker. (Beege, Schneider, Nebel, Mittangk, & Rey, 2017) tested the coherence between the onscreen agent's age and the age primed with the text. They made a between-subjects 2×2 factorial design (young vs. old female agent; young vs. old text priming) and tested learning retention and transfer, as well as cognitive load and motivational data. Results showed that learning transfer was higher when the agent's age was coherent with the text. Other parameters showed no effect of age coherence. (Schroeder & Adesope, 2015) found no significant gender effect in learning and learner's perception.

Regarding ethnicity of real actors, (Aronson, Plass, & Bania, 2012) presented short videos (2 minutes) showing a dialogue between a health expert and a patient about HIV tests. Four versions of the video were recorded, varying in emotional content (positive, negative) and speaker's ethnicity (African American, White). They showed the video to people belonging to different ethnic groups. Results showed weak effects of ethnicity in user responses.

6.7.4 The Mise-en-scène

A side aspect of social appearance lays outside the actor, in the surrounding displayed environment. Some instructional videos are situated in a neutral space (e.g. recorded with chroma), while others are situated in a real-life setting, such as a television studio, an office, a classroom, or on location. The use of a certain setting is not neutral. It would influence the construction of the social appearance of the actor, hence the viewer's response in terms of credibility and motivation. For example, a video podcast filmed in the instructor's office room, showing scholar books will enhance the status of the instructor as a qualified expert. Analogously, a videocast that has been recorded with a laptop's webcam in the instructor's office, with a long-angle shot, implies an informal setting that may predispose the viewer to a more intimate social response.

Studies on MOOC video lectures have noticed such differences in scene setting. Guo, Kim and Rubin (2014) and Hansch et al (2015) used the mise-en-scène as a characterizing feature in their respective classifications of video styles. Guo et al distinguishes *classroom*, *studio* and *office desk* videos. Hansch et al.'s *webcam capture*, *on location*, *green screen* and *classroom lecture* types are all characterized by scenario and camera shot. Both studies suggested that these differences in setting may influence viewer engagement. Unfortunately, I have not found experimental studies that measure potential effects on scene setting in learning.

6.8 Final discussion

The work described in this chapter has developed the most basic levels of the classification scheme of this thesis. More than 200 references to scientific works have been collected, which have contributed to round out the catalog of the instructional video characteristics. These references have also helped to validate the classification scheme obtained in Chapter 4.

The features that have been retrieved from the literature research fit naturally into the classification domains identified in Chapter 4. Only minor adjustments have been applied to the final draft of Chapter 4. It was considered convenient to add a Social Appearance domain, due to the abundance of references dealing with the social presence and the learner's social response to certain features. After this change, the first domains of the classification scheme have been rearranged to six: physical/medium, presentation, interaction, spatiotemporal, speech and social appearance. The strategic and generic domains remain undeveloped.

6.8.1 Findings

Each of the feature domains being reviewed shows a wealth of relevant characteristics, though the amount and depth of research evidence is uneven. There are some 'shaded areas' that are suggestive but require further study. The next section will account for some of these understudied areas.

Among the inventory of video characteristics, the **actor** stands out as a complex construct around which the video content is organized. The actor influences both the visual and auditory presentation of content. The actor also shapes subtler aspects of discourse rhetoric and the social and affective connection with the audience. Coordination mechanisms between the actor and the displaying **board** are key to understand the dynamics of many instructional video presentation styles.

The literature review has evidenced the orchestration of cinematographic language and written/spoken language to provide an integrated, effective instructional discourse. Different techniques have been identified, such as the filmic montage for sequencing video segments with continuity, or an adequate rhetoric structure that provides clarity and engagement.

As for the spoken discourse, this research has shown that expository video formats cannot be regarded as a mere reformatting of conventional classroom lectures. Research evidence shows other modalities of discourse having interesting learning effects, such as the dialogic exposition and the use of subjective camera (point of view).

Another remarkable observation is how the intrinsic lack of interactivity of the video medium is tackled by video designers with multiple methods: in-video quizzes, navigational devices and interpersonal features of speech.

One of the most relevant findings of this research is that **Systemic Functional Linguistics** (SFL) has proven useful for characterizing multimedia learning principles that are related to speech, as is the case with the Voice principle and the Personalization principle. Most speech features of instructional videos can be classified around Halliday's SFL concepts of language metafunction and function.

6.8.2 What is missing in current research

The literature review has revealed some areas with inconclusive results, or insufficient research effort. In this section, I develop my personal judgment about some areas of research that, after my review, I consider to be underexplored and worth investigating.

What is missing in the Presentation domain

There are some entities belonging to the Presentation domain which need particular research efforts:

Subtitles. There are evidences that subtitles may be beneficial or harmful for learning, depending on learner characteristics. More research is needed to delimit what are the boundary conditions for the learning effects of subtitling videos, as well as testing the modern machine-generated close captions and translations.

Deictic gestures vs. other cues. I have noticed a lack of studies assessing the effect of instructor deictic gestures in video lectures, compared to face-to-face gestures, or compared to other virtual pointing cues.

Gaze augmentation. Gaze augmentation has been assessed on procedural videos, modeling examples and classifying tasks. Research should be expanded to expository lectures.

Signaling sounds. They are rarely used in common instructional videos, despite their potential benefits as cues. They have been frequently used in multimedia learning applications and in other expository genres, such as broadcast news. Signaling sounds in video lectures and tutorials deserve more research and practical use.

What is missing in the Interaction domain

This domain appears to be fully covered by all kind of devices and supporting research. I will only mention that most recent papers tend to focus on integrating learners' collaboration in the video.

What is missing in the Spatiotemporal domain

The Spatiotemporal domain has strong support from research. Several Multimedia Learning principles are linked to spatial and temporal relations between video entities. Nevertheless, many features have been taken apart since the big shift in research from educational films to computer applications. These features deal with

processes of cinematography, such as **camera settings** and **cutting** (cutting rate, shot alternance).

In addition, there are few works of **film analysis** that focus on instructional video, which contrasts with the relative abundance of film analysis of other expository genres, such as television commercials and broadcast news. Film analysis on instructional videos would clarify their peculiarities and would pave the way to a better understanding of instructional video design principles.

I have struggled to find research on **measures of spatial and temporal complexity** suitable for instructional videos, with almost no success. Basic measures such as words per minute or slides per minute are used by some authors, but I have not found standardized indicators for scene complexity or cutting rate. I consider necessary that the research community produces synthetic indicators of complexity for the visual, textual and spoken information provided in the video.

What is missing in the Speech domain

At present, discourse analysis of studio-recorded video lectures has been barely addressed. The Speech domain needs extensive work that delves into the specificities of the discourse of instructional videos. Those analyses would query if existing face-to-face rhetoric models are also valid for instructional video formats and genres. They would also study the structural and functional differences between dialogic and monologic expository formats. In general, more quantitative data is needed to measure the actual use of linguistic and rhetoric features in instructional video.

What is missing in the Social Appearance domain

As I conclude from my review, the balance between the social presence effect and principles related to cognitive efficiency is not clear. The presence of an on-screen actor and seductive details often add irrelevant information that potentially hinders learning, but at the same time social presence increases learner engagement. This is an open question at the time of this writing. In any case, social appearance properties must be included in an up-to-date inventory of instructional video characteristics, as they continue to be the subject of active and intense research.

On the other hand, the mise-en-scène is suspected to contribute to the social response of the video watcher, and therefore it may be non-neutral in the learning outcomes. It is advisable to make some experimental research that sheds light on this hypothesis.

Chapter 7. Conclusions

7.1 Results of this research

The proposed goal for this doctoral thesis was *to build a classification scheme for instructional video characteristics*, with this accompanying research question: *how can instructional video characteristics be systematically and usefully classified?* This dissertation has described the process of building such classification scheme, starting from an extensive literature review which led to a bottom-up classification work, ending with the full scheme that has been presented in Chapter 5. This classification scheme is based on Multimodal Discourse Analysis theories and tools, particularly Bateman's GeM framework (Bateman, 2008).

The classification scheme contains eight descriptive domains: medium, presentation, interaction, spatiotemporal, speech, social appearance, strategic and generic (for video genres). These domains are organized in hierarchical layers, from the physical medium to more abstract levels.

In addition, specific taxonomies have been developed for each of these domains: medium, presentation, interaction, spatiotemporal, speech and social appearance. All of these taxonomies are introduced in Chapter 5 and described in detail in Chapter 6.

In summary, this research has delivered these products:

- A classification scheme that systematically organizes the characteristics in instructional videos that researchers have found relevant in learning processes.
- A survey of presentation styles and features currently used in instructional videos in online courses (MOOCs).
- A comprehensive literature review on the instructional video features that are related to learning effectiveness.

7.2 Discussion and contributions

7.2.1 A wide and comprehensive scope

This research on instructional videos has been conducted from a broad perspective. It was the will of this research to overcome the restricted viewpoint of the Technology Enhanced Learning research community, which has traditionally considered video as one more component of computer-based learning objects. Instead, this research has considered three natures of instructional videos: as instructional films, as multimedia learning objects, and as multimodal texts.

This threefold perspective has led to a classification scheme that addresses a wide variety of video features, expanding those routinely discussed in Multimedia Learning research. The resulting scope is as wide as that of Wetzel et al's review on instructional films and video (Wetzel et al., 1993). The filmic aspects of instructional videos (montage, camera settings, mise en scène) have been integrated into the taxonomies, as well as elements of discourse: rhetoric, language functions and genres.

In a certain way, this research could be considered as an update of the dated Wetzel et al's review. In order to elaborate the classification scheme, I had to carry out an exhaustive review of the scientific literature on instructional videos. This review constitutes a substantial part of this dissertation and I consider it valuable in itself. The research works discussed in Chapter 2 and Chapter 6 may serve as an up-to-date reference to the state of the art of instructional video research.

7.2.2 A scheme grounded in Linguistic and Semiotic theories

One of the differences between this dissertation and other scholarly reviews on instructional video is that my research offers a strong taxonomical framework that organizes instructional video characteristics around meaningful descriptive domains.

The architecture of the classification system is grounded on recent developments of Multimodal Discourse Analysis. In particular, John Bateman's GeM framework (Bateman, 2008, 2013) has been used as the overarching foundation for organizing the classification domains. The layered architecture of the classification scheme proposed in this dissertation has been inspired by the concept of stratification from Systemic Functional Linguistics (Lim-Fei & O'Halloran, 2014), as well as by related design frameworks (Vorvilas, Karalis, et al., 2011).

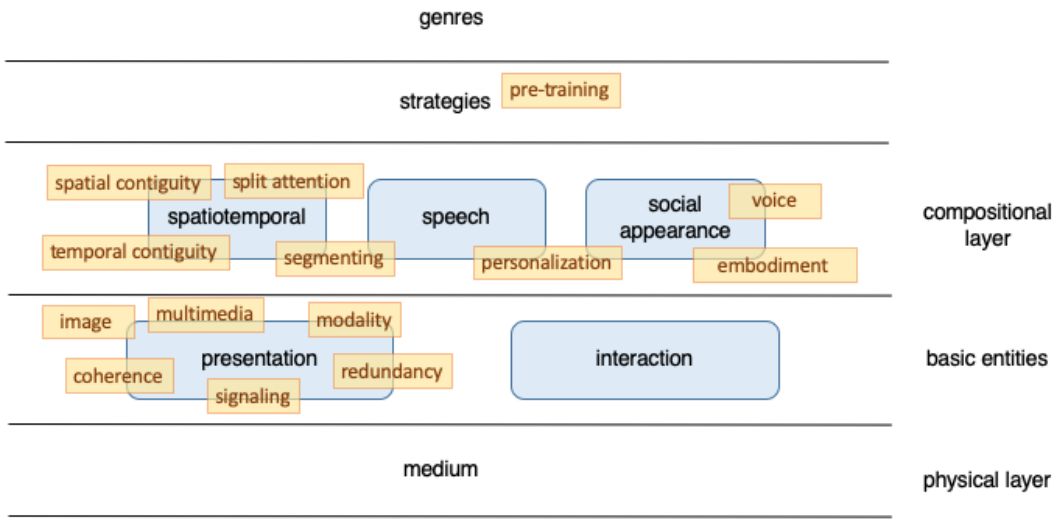
In addition, Systemic Functional Linguistics has proven useful for characterizing multimedia learning principles that are related to speech, as is the case with the Voice principle and the Personalization principle. Other speech features of instructional videos can be classified around SFL concepts of language function, as it occurs with the self-explanation prompts, humor, and dialogic versus monologic discourse.

7.2.3 A meaningful scheme

The descriptive domains have a tight correspondence with functional areas in learning research: modalities of representation, learner interaction, spatial layout, temporal segmentation, language and social presence.

Figure 7-1 shows how most relevant Mayer's Multimedia Learning Principles fit into the proposed classification scheme. We can see two main clusters, one around the Presentation domain and another around the Spatiotemporal domain. Three principles (personalization, voice, embodiment) are linked to the Social Appearance domain. The Speech domain is directly related to two principles (segmenting and

Figure 7-1. Multimedia Learning Principles and the classification scheme



personalization), though these principles are also related to other domains. This representation may help to realize that Multimedia Learning Principles can be grouped around structural domains.

7.2.4 Other findings

Apart from the classification scheme, this research has found some facts and insights that provide more knowledge on instructional videos.

Actors and boards

Among the inventory of video characteristics, the **actor** stands out as a complex construct around which the video content is organized. The actor influences both the visual and auditory presentation of content. The actor also shapes subtler aspects of discourse rhetoric and the social and affective connection with the audience. Coordination mechanisms between the actor and the displaying **board** are key to understand the dynamics of many instructional video presentation styles.

The field study on MOOC videos (see Chapter 3) has revealed a spectrum of design patterns dealing with the role of actors and boards in the frame layout. The spectrum goes from **speaker-centric** styles (a visible person talks about the content) to **board-centric** styles (a large rectangular surface shows the content). One of the findings of the study is that the adoption of each of these patterns is significantly related to course subject: Arts and Humanities courses show a preference for speaker-centric styles, while Engineering and "Hard Science" courses prefer board-centric formats. The Social and Health Sciences courses are in a neutral position.

Missing areas in current research

The extensive literature review taken as the final stage of this research has evidenced some video features which need more research effort to ascertain their contribution to learning processes.

Within the Presentation domain, I suggest more exploration of the effect of subtitles, the relative benefit of deictic gestures compared to other signaling cues, the effect of gaze augmentation in expository lectures, and the effect and use of signaling sounds.

The Spatiotemporal domain needs to recover the research on cinematographic manipulations, such as the effect of camera setting, cutting rate, continuity, and many other filmic features. Also, I perceive the need to elaborate integrated measures of spatial and temporal complexity, and their relations to cognitive load and learning outcomes.

The review of the Speech domain has made it more obvious that, within the field of Discourse Analysis, there is little specific research of instructional videos, with very few exceptions (Atapattu & Falkner, 2017; Bernad-Mechó, 2015). There is very little work on the rhetoric patterns used in video lectures and tutorials, compared to the relative abundance of studies on classroom lectures. Do rhetoric principles and guidelines for classroom videos apply also to video lectures? Does the new medium promote a shift in discourse patterns?

In the age of educational television, De Vaney (1991) warned that the transition from theater film to television changed semiotic codes in use (e.g. shot lengths, dialogue codes). The same may be happening now as the physical medium has changed from desktop devices to handheld, or from broadcast to interactive, on-demand watching. This is something that needs further research.

Finally, in the Social Appearance domain, the evidence of research does not conclude what is the benefit of the social presence effect in relation with learning outcomes. The mise en scène should also be investigated as a feature that influences the learner's response to the contents.

7.2.5 Contribution to the scientific community

As a conclusion of this discussion, these are the three main contributions of this research project to the scientific community:

- An up-to-date and broad review of the state of the art of instructional video research.
- A classification scheme that comprehensively lists and conceptually organizes the characteristics of instructional videos that are related to learning processes.
- New evidences about current usage patterns of instructional video in online courses.

I also believe that a major contribution of the classification scheme is that it contributes to the integration of two disciplines: Multimodal Discourse Analysis and Technology Enhanced Learning. Both scientific communities can benefit from this integrated result:

- For the Technology Enhanced Learning research community, the new classification scheme is a tool that organizes the findings of the Cognitive Theory of Multimedia Learning and related disciplines in a structured and meaningful way. The holistic perspective taken by the classification will help to identify non-obvious relationships among features, as well as to (re)discover areas of research that deserve more attention.
- For the Multimodal Discourse Analysis research community, the classification scheme is a demonstration that multimodal analysis frameworks and theories can be applied to instructional video, and also an invitation to integrate the findings from Multimedia Learning theories into their research activity.

7.3 Limitations of this work

This work has not yet explored all the dimensions identified in the classification scheme. In particular, there is a whole area that has remained undeveloped, which are video genres. The Strategic domain has not been developed either.

With regard to the methodology, it must be said that the process of collecting evidence has been rather exploratory. In the final phase of the research, a systematic process of reference collection has been followed, but the judgement on when a feature should be explored has been ultimately driven by subjective criteria. This may have introduced some researcher bias into the final result of the taxonomies. It would therefore be desirable to carry out a further iteration of the process, in which a new review of the literature is undertaken using a much more systematic and objective method, especially for the exclusion or inclusion criteria of references.

In order to construct a very broad map of characteristics, the selection criteria in the final literature review have been unrestrictive. The taxonomies collect characteristics that have a strong experimental support along with others that are only mentioned as worthy of interest and with potential in learning, but not backed by conclusive experimental results. As a consequence, the taxonomies are very broad in scope, but at the same time they have the risk of confusing the categories of high and low quality.

Related to the above is that this study does not attempt to differentiate the effectiveness of the characteristics according to the context in which they are applied. There are characteristics whose usefulness and effectiveness may depend on factors such as the individual characteristics of the learner, the pedagogical context, and instructional design. This classification scheme does not fall into these distinctions.

Finally, there are several research fields related to video-based instruction that have not been explored. For instance, emotional design (Heidig, Müller, & Reichelt, 2015; Schneider et al., 2016) and other modalities of video design, such as storytelling and gamification with narrative style (Zhu, Pei, & Shang, 2017).

7.4 Proposals for future work

The result of the work presented in this dissertation should be considered the end of an iteration within a research process that can continue in many directions. The most immediate move is to develop the levels of taxonomy to which the literature review process has not been applied: Strategic Layer (goals and purposes) and Generic Layer (instructional video genres). This last layer is the one that can be developed more easily in the present state of this research.

In the following paragraphs I will mention these and other lines of work that can be derived immediately from this thesis.

7.4.1 Video genres

To develop the Generic domain of the classification scheme, it would be necessary to build an inventory of video genres (based on field evidence), and then characterize the genres according to a set of dimensions or traits. The literature review presented in Chapter 2 has provided a number of catalogs of video presentation formats that is an excellent starting point to build the inventory of video genres, along with the survey made in Chapter 3 on MOOC videos. According to the preliminary research described in Chapter 4, video genres could be classified by dimensions such as purpose, type of recorded action, communication format, frame format and scenario/background. All these materials and findings can be developed, resulting in a structured catalog of video genres.

7.4.2 A refined literature review

I have mentioned that the outcome of the final categorization process may have some weaknesses related to the selection criteria for characteristics used in the final literature review. These weaknesses can be overcome by carrying out a new iteration of the process. From the current classification scheme and its taxonomies, a new revision of the literature can be made, this time using a more systematic and objective method. The process would start from a state similar to the end of Chapter 4, with the general classification scheme, but without the domain-specific taxonomies.

To collect relevant evidence, a much more restrictive inclusion criterion can be used: for example, including only characteristics that have an experimental support above a certain threshold. An additional criterion may be to limit the literature search only to higher order research: reviews or meta-analyses.

This method would result in a more limited inventory of characteristics, but at the same time more homogeneous and with a higher quality content. In that case, this

restricted review must be done cautiously, as there are several video features that are still poorly researched: there is a risk of producing an overly limited map of characteristics, which is precisely something that this research project has tried to avoid.

A workaround is to admit characteristics with different levels of support and label each characteristic with its associated level. In this way users of the classification scheme could have information about the quality of each characteristic or taxonomic group.

7.4.3 Other works

Expand and refine the field study. Extend the survey made on MOOCs to other public video repositories, such as YouTube. The new survey can be made using the new classification scheme as the conceptual framework for the surveyed features.

Build a corpus of videos and characteristics. For a better understanding of this classification scheme, a corpus of real-life examples of instructional videos can be elaborated. The corpus specimens would show the characteristics collected in this classification. In fact, my research has resulted in a large inventory of sample video clips, which is ready to be curated and structured to become a usable corpus.

Apply the framework in video analysis. Discourse Analysis tasks include the annotation of communication events that occur in audiovisual recordings. One immediate application of this classification scheme is to use one software annotation tool, such as *Multimodal Analyis Video* by O'Halloran team (Lim Fei et al., 2015). This tool allows to easily define custom categories. The classification scheme can be transferred to the tool and some analysis can be performed on sample instructional videos. This action would help validate the classification scheme in a practical setting.

Bibliography

Aagard, H. P. (2014). *The effects of a humorous instructional video on motivation and learning* (Doctoral dissertation, Purdue University).

Adams, C., Yin, Y., Madriz, L. F. V., & Mullen, C. S. (2014). A phenomenology of learning large: the tutorial sphere of xMOOC video lectures. *Distance Education*, *35*(2), 202–216. http://doi.org/10.1080/01587919.2014.917701

Ädel, A. (2010). Just to Give You Kind of a Map of Where We Are Going: A Taxonomy of Metadiscourse in Spoken and Written Academic English. *Nordic Journal of English Studies*, *9*(2), 69–97.

Afonso Suárez, M.D.; Guerra Artal, C.; Villalba Casas, A.; Elías Hernández, A. (2009). Prometeo and IESCampus, Multimedia Learning Management Systems, towards Intelligent Agents. In *14th E-Learn World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education*. Vancouver, Canada.

Aiken, E. G., Thomas, G. S., & Shennum, W. A. (1975). Memory for a lecture: Effects of notes, lecture rate, and informational density. *Journal of Educational Psychology*, *67*(3), 439–444. http://doi.org/10.1037/h0076613

Akahori, W., Hirai, T., Kawamura, S., & Morishima, S. (2016). Region-of-Interest-Based Subtitle Placement Using Eye-Tracking Data of Multiple Viewers. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX'16* (pp. 123–128). New York, New York, USA: ACM Press. http://doi.org/10.1145/2932206.2933558

Aleem, T. A. (1998). *A Taxonomy of Multimedia Interactions* (Doctoral Dissertation, The Union Institute Cincinnati).

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). *A Pattern Language*. Oxford University Press.

Alharbi, G., Ng, R. W. M., & Hain, T. (2015). Annotating Meta-discourse in Academic Lectures from Different Disciplines. In *ISCA Workshop on Speech and Language Technology in Education (SLaTE)* (pp. 161–166). Leipzig.

Amir, A., Ponceleon, D., Blanchard, B., Petkovic, D., Srinivasan, S., & Cohen, G. (2000). Using audio time scale modification for video browsing. In *Proceedings of the 33rd Hawaii International Conference on System Sciences. IEEE.*

Anderson, R. J., Hoyer, C., Wolfman, S. a., & Anderson, R. (2004). A study of digital ink in lecture presentation. *Proceedings of the 2004 Conference on Human Factors in Computing Systems - CHI '04*, *6*(1), 567–574. http://doi.org/10.1145/985692.985764

Anderson, T., & Dron, J. (2011). Three generations of distance education pedagogy. *International Review of Research in Open and Distance Learning*, *12*(3), 80–97.

André, E. (2000). The generation of multimedia presentations. In *Handbook of natural language processing* (pp. 305–327).

Aronson, I. D., Plass, J. L., & Bania, T. C. (2012). Optimizing educational video through comparative trials in clinical environments. *Educational Technology Research and Development : ETR & D*, *60*(3), 469–482. http://doi.org/10.1007/s11423-011-9231-4

Atapattu, T., & Falkner, K. (2017). Discourse Analysis to Improve the Effective Engagement of MOOC Videos. In *Learning Analytics and Knowledge Conference* (pp. 580–581). http://doi.org/10.1145/3027385.3029470

Atkinson, R. K., Mayer, R. E., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, *30*(1), 117–139. http://doi.org/10.1016/j.cedpsych.2004.07.001

Aubert, O., Prié, Y., & Canellas, C. (2014). Leveraging video annotations in video-based e-learning. In *7th International Conference on Computer Supported Education (CSEDU)*. Barcelona, Spain.

Ayres, P., & Sweller, J. (2014). The Split-Attention Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 206–227). Cambridge University Press.

Baggaley, John, Ferguson, M., & Brooks, P. (1980). *Psychology of the TV Image*. Gower.

Baggaley, Jon, & Duck, S. W. (1975). Experiments in ETV - effects of edited cutaways. *Educational Broadcasting International*, *8*(1), 36–37.

Baggett, P. (1984). Role of temporal overlap of visual and auditory material in forming dual media associations. *Journal of Educational Psychology*, *76*(3), 408–417. http://doi.org/10.1037//0022-0663.76.3.408

Bailey, K. D. (1994). *Typologies and Taxonomies: An Introduction to Classification Techniques* (Vol. Sage Unive). Thousand Oaks, CA: SAGE Publications Inc.

Baldry, A., & Thibault, P. J. (2006). *Multimodal Transcription and Text Analysis*. Equinox. http://doi.org/10.1558/equinox.24010.

Banai, K., & Lavner, Y. (2012). Perceptual Learning of Time-Compressed Speech: More than Rapid Adaptation. *PLoS ONE*, *7*(10), e47099. http://doi.org/10.1371/journal.pone.0047099

Banas, J. A., Dunbar, N., Rodriguez, D., & Liu, S. J. (2011). A review of humor in educational settings: Four decades of research. *Communication Education*, *60*(1), 115–144. http://doi.org/10.1080/03634523.2010.496867

Bandura, A. (1962). Social learning through imitation. In M. R. Jones (Ed.), *Nebraska symposium of motivation* (pp. 211–269). Lincoln: University of Nebraska Press.

Barford, J., & Weston, C. (1997). The use of video as a teaching resource in a new university. *British Journal of Educational Technology*, *28*(1), 40–50. http://doi.org/10.1111/1467-8535.00005

Barrett, N. E., & Liu, G.-Z. (2016). Global Trends and Research Aims for English Academic Oral Presentations: Changes, Challenges, and Opportunities for Learning Technology. *Review of Educational Research*, *86*(4), 1227–1271. http://doi.org/10.3102/0034654316628296

Bateman, J. A. (2008). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. Palgrave Macmillan. http://doi.org/10.1057/9780230582323

Bateman, J. A. (2011). The Decomposability of Semiotic Modes. In K. L. O'Halloran & B. A. Smith (Eds.), *Multimodal Studies: Exploring Issues and Domains* (pp. 17–38). Routledge.

Bateman, J. A. (2013). Multimodal analysis of film within the GeM framework. *Ilha Do Desterro*, (64), 49–84. http://doi.org/10.5007/2175-8026.2013n64p49

Bateman, J. A., & Schmidt, K.-H. (2012). *Multimodal Film Analysis: How Films Mean*. (K. O'Halloran, Ed.). Routledge.

BavaHarji, M., Alavi, Z. K., & Letchumanan, K. (2014). Captioned instructional video: Effects on content comprehension, vocabulary acquisition and language proficiency. *English Language Teaching*, *7*(5), 1–16. http://doi.org/10.5539/elt.v7n5p1

Baylor, A. L. (2009). Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3559–3565. http://doi.org/10.1098/rstb.2009.0148

Baylor, A. L. (2011). The design of motivational agents and avatars. *Educational Technology Research and Development*, *59*(2), 291–300. http://doi.org/10.1007/s11423-011-9196-3

Beaver, F. E. (2006). *Dictionary of film terms: the aesthetic companion to film art*. Peter Lang Publishing.

Beege, M., Schneider, S., Nebel, S., Mittangk, J., & Rey, G. D. (2017). Ageism – Age coherence within learning material fosters learning. *Computers in Human Behavior*, *75*, 510–519. http://doi.org/10.1016/j.chb.2017.05.042

Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2017). Look into my eyes! Exploring the effect of addressing in educational videos. *Learning and Instruction*, *49*, 113–120. http://doi.org/10.1016/j.learninstruc.2017.01.004

Bernad-Mechó, E. (2015). A Multimodal Discourse Analysis of Linking Metadiscursive Elements in Two Opencourseware Lectures (MOOCs). *Procedia - Social and Behavioral Sciences*, *212*, 61–66. http://doi.org/10.1016/j.sbspro.2015.11.299

Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education*, *101*, 150–167. http://doi.org/10.1016/J.COMPEDU.2016.06.005

Bernsen, N. O. (1994). Foundations of multimodal representations: a taxonomy of representational modalities. *Interacting with Computers*, *6*(4), 347–371. http://doi.org/10.1016/0953-5438(94)90008-6

Bernsen, N. O. (1997). Defining a taxonomy of output modalities from an HCI perspective. *Computer Standards & Interfaces*, *18*(6), 537–553. http://doi.org/10.1016/S0920-5489(97)00018-4

Bhat, S., Perry, M., & Chinprutthiwong, P. (2015). Seeing the Instructor in Two Video Styles: Preferences and Patterns. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 305–312).

Bhatia, V. K. (1997). Genre analysis today. *Revue Belge de Philologie et d'histoire*, *75*(3), 629–652. http://doi.org/10.3406/rbph.1997.4186

Bishop, M. J., Amankwatia, T. B., & Cates, W. M. (2008). Sound's use in instructional software to enhance learning: A theory-to-practice content analysis. *Educational Technology Research and Development*, *56*(4), 467–486. http://doi.org/10.1007/s11423-006-9032-3

Bishop, M. J., & Sonnenschein, D. (2012). Designing with sound to enhance learning: Four recommendations from the film industry. *Journal of Applied Instructional Design*, *2*(1), 5–15.

Biswas, A., Gandhi, A., & Deshmukh, O. (2015). MMToC: A Multimodal Method for Table of Content Creation in Educational Videos. In *Proceedings of the 23rd ACM international conference on Multimedia - MM '15* (pp. 621–630). New York, New York, USA: ACM Press. http://doi.org/10.1145/2733373.2806253

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, *13*(6), 4–16. http://doi.org/10.3102/0013189X013006004

Bol, N., Van Weert, J. C. M., De Haes, H. C. J. M., Loos, E. F., & Smets, E. M. A. (2015). The effect of modality and narration style on recall of online health information: Results from a web-based experiment. *Journal of Medical Internet*

Research, *17*(4), e104. http://doi.org/10.2196/jmir.4164

Boltz, M. G. (2001). Musical Soundtracks as a Schematic Influence on the Cognitive Processing of Filmed Events. *Music Perception*, *18*(4), 427–454. http://doi.org/10.1525/mp.2001.18.4.427

Boltz, M., Schulkind, M., & Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Memory & Cognition*, *19*(6), 593–606. http://doi.org/10.3758/BF03197154

Brewer, W. F. (1980). Literary theory, rhetoric, and stylistics: Implications for psychology. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 221–239). Hillsdale, NJ: Erlbaum.

Brøndmo, H. P., & Davenport, G. (1991). Creating and viewing the Elastic Charles - a hypermedia journal. In R. MacAleese (Ed.), *Hypertext: State of the Art* (pp. 43–51). Intellect Books.

Bruijn, D. D., Mul, S. D., & Oostendorp, H. V. (1992). The influence of screen size and text layout on the study of text. *Behaviour & Information Technology*, *11*(2), 71–78.

Bruti, S. (2015). Teaching Learners How to Use Pragmatic Routines Through Audiovisual Material. In *Multimodal Analysis in Academic Settings: From Research to Teaching* (p. 213).

Bucci, W. (1997). *Psychoanalysis and cognitive science: A multiple code theory*. Guilford Press.

Burch, N. (1970). *Praxis del Cine* (1st ed.). Madrid, ES: Fundamentos.

Burden, H., Heldal, R., & Adawi, T. (2012). Pair Lecturing to Enhance Reflective Practice and Teacher Development. In *LTHs 7:e Pedagogiska Inspirationskonferens*.

Campbell, J., & Mayer, R. E. (2009). Questioning as an instructional method: Does it affect learning from lectures? *Applied Cognitive Psychology*, *23*(6), 747–759. http://doi.org/10.1002/acp.1513

Carliner, S. (2000). A Three-part Framework for Information Design. *Technical Communication*, (Fourth Quarter 2000), 561–576.

Carpenter, C. R. (1953). A theoretical orientation for instructional film research. *Audiovisual Communication Review*, *1*(1), 38–52. http://doi.org/10.1007/BF02713169

Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, *20*(6), 1350–1356. http://doi.org/10.3758/s13423-013-0442-z

Casasanto, D. (2008). Similarity and proximity: When does close in space mean close in mind? *Memory & Cognition*, *36*(6), 1047–1056. http://doi.org/10.3758/MC.36.6.1047

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2014). Learning from observing hands in static and animated versions of non-manipulative tasks. *Learning and Instruction*, *34*, 11–21. http://doi.org/10.1016/j.learninstruc.2014.07.005

Chandler, D. (1997). An Introduction to Genre Theory. Retrieved from http://visual-memory.co.uk/daniel/Documents/intgenre/chandler_genre_theory.pdf

Chen, H., Yan, M., Liu, S., & Jiang, B. (2017). Gaze inspired subtitle position evaluation for MOOCs videos (Vol. 10443, pp. 1044315–1044318).

Cheng, D. S., Salamin, H., Salvagnini, P., Cristani, M., Vinciarelli, A., & Murino, V. (2014). Predicting online lecture ratings based on gesturing and vocal behavior.

Journal on Multimodal User Interfaces, *8*(2), 151–160. http://doi.org/10.1007/s12193-013-0142-z

Cheng, S. W. (2012). "That's it for today": Academic lecture closings and the impact of class size. *English for Specific Purposes*, *31*(4), 234–248. http://doi.org/10.1016/j.esp.2012.05.004

Cherrett, T., Wills, G., Price, J., Maynard, S., & Dror, I. E. (2009). Making training more cognitively effective: Making videos interactive. *British Journal of Educational Technology*, *40*(6), 1124–1134. http://doi.org/10.1111/j.1467-8535.2009.00985.x

Chi, M. T. H., De Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477. http://doi.org/10.1016/0364-0213(94)90016-7

Chi, M. T. H., Kang, S., & Yaghmourian, D. L. (2017). Why Students Learn More From Dialogue- Than Monologue-Videos: Analyses of Peer Interactions. *Journal of the Learning Sciences*, *26*(1), 10–50. http://doi.org/10.1080/10508406.2016.1204546

Churchill, D. (2007). Towards a useful classification of learning objects. *Educational Technology Research and Development*, *55*(5), 479–497. http://doi.org/10.1007/s11423-006-9000-y

Churchill, D. (2014). Presentation design for "conceptual model" learning objects. *British Journal of Educational Technology*, *45*(1), 136–148. http://doi.org/10.1111/bjet.12005

Churchill, D., & Hedberg, J. (2008). Learning object design considerations for small-screen handheld devices. *Computers and Education*, *50*(3), 881–893. http://doi.org/10.1016/j.compedu.2006.09.004

CISCO SYSTEMS. (2003). *Reusable Learning Object Strategy: Designing and Developing Learning Objects for Multiple Learning Approaches*.

Clark, R. C., & Mayer, R. E. (2008). *e-Learning and the Science of Instruction* (2nd ed.). Pfeiffer.

Clark, R. E. (1983). Reconsidering Research on Learning from Media. *Review of Educational Research*, *53*(4), 445–459. http://doi.org/10.3102/00346543053004445

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development*, *42*(2), 21–29. http://doi.org/10.1007/BF02299088

Collis, B., & Peters, O. (2000). Educational Applications of WWW-Based Asynchronous Video. In N. Correia (Ed.), *Multimedia '89. Springer Computer Sciences Series* (pp. 177–186). Springer Vienna. http://doi.org/10.1007/978-3-7091-6771-7_19

Compagnone, A. (2015). The Reconceptualization of Academic Discourse as a Professional Practice in the Digital Age: A Critical Genre Analysis of TED Talks. *HERMES-Journal of Language and Communication in Business*, *27*(54), 49–69. http://doi.org/10.7146/hjlcb.v27i54.22947

Coskun, M., & Acartürk, C. (2015). Gesture Production under Instructional Context : The Role of Mode of Instruction. In P. P. Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio (Ed.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society (CogSci 2015)* (pp. 459–464). Pasadena, California: Cognitive Science Society.

Coursera. (2015). Powerful Tools for Teaching and Learning: Digital Storytelling. Retrieved October 1, 2015, from https://www.coursera.org/course/digitalstorytelling

Craig, S. D., & Schroeder, N. L. (2017). Reconsidering the voice effect when learning from a virtual human. *Computers & Education*, *114*, 193–205. http://doi.org/10.1016/J.COMPEDU.2017.07.003

Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The Deep-Level-Reasoning-Question Effect: The Role of Dialogue and Deep-Level-Reasoning Questions During Vicarious Learning. *Cognition and Instruction*, *24*(4), 565–591. http://doi.org/10.1207/s1532690xci2404

Crawford Camiciottoli, B. (2007). *The language of business studies lectures: A corpus-assisted analysis*. John Benjamins Publishing Company.

Crawford Camiciottoli, B. (2008). Interaction in academic lectures vs. written text materials: The case of questions. *Journal of Pragmatics*, *40*(7), 1216–1231. http://doi.org/10.1016/j.pragma.2007.08.007

Crawford Camiciottoli, B., & Bonsignori, V. (2015). The Pisa Audiovisual Corpus Project: a Multimodal Approach To ESP Research. *Journal of English for Specific Purposes at Tertiary Level*, *3*(2), 139–159.

Crawford Camiciottoli, B., & Fortanet-Gómez, I. (Eds.). (2015). *Multimodal Analysis in Academic Settings*. Routledge.

Crook, C., & Schofield, L. (2017). The video lecture. *The Internet and Higher Education*, *34*, 56–64. http://doi.org/10.1016/j.iheduc.2017.05.003

Cross, A., Bayyapunedi, M., Cutrell, E., Agarwal, A., & Thies, W. (2013). TypeRighting: Combining the Benefits of Handwriting and Typeface in Online Educational Videos. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 793–796. http://doi.org/10.1145/2470654.2470766

Cross, A., Bayyapunedi, M., Ravindran, D., Cutrell, E., & Thies, W. (2014). VIdwiki: Enabling the Crowd to Improve the Legibility of Online Educational Videos. *Proc. CSCW '14*, 1167–1175. http://doi.org/10.1145/2531602.2531670

Crystal, D. (1992). *Introducing Linguistics*. London: Penguin.

d'Ydewalle, G., Van Rensbergen, J., & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: The case of subtitling. In J. K. O'Regan & A. Lévy-Schoen (Eds.), *Eye Movements: From Physiology to Cognition* (pp. 313–321). Elsevier.

de Koning, B. B., & Tabbers, H. K. (2011, December 26). Facilitating Understanding of Movements in Dynamic Visualizations: An Embodied Perspective. *Educational Psychology Review*. Springer US. http://doi.org/10.1007/s10648-011-9173-8

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2011). Attention cueing in an instructional animation: The role of presentation speed. *Computers in Human Behavior*, *27*(1), 41–45. http://doi.org/10.1016/j.chb.2010.05.010

De Vaney, A. (1991). A grammar of educational television. In D. Hlynka & J. C. Belland (Eds.), *Paradigms Regained: The Uses of Illuminative, Semiotic, and Post-Modern Criticism As Modes of Inquiry in Educational Technology* (pp. 241–280). Educational Technology Publications.

Degano, C. (2012). Texture Beyond the Text: Slides and Talk in Conference Presentations. In S. M. Maci & M. Sala (Eds.), *Genre Variation in Academic Communication. Emerging Disciplinary Trends* (pp. 135–152). Bergamo: Centro di Ricerca sui Linguaggi Specialistici.

Delen, E., Liew, J., & Willson, V. (2014). Effects of interactivity and instructional scaffolding on learning: Self-regulation in online video-based environments. *Computers and Education*, *78*, 312–320. http://doi.org/10.1016/j.compedu.2014.06.018

Deroey, K. L. B., & Taverniers, M. (2011). A corpus-based study of lecture functions. *Moderna Sprak*, *105*(2), 1–22.

Dey, G., & Low, H. (1968). Team Teaching: A Dyadic Approach. *Improving College and University Teaching*, *16*(1), 23–25. http://doi.org/10.1080/00193089.1968.10532690

Díaz, D., Ramírez, R., & Hernández-Leo, D. (2015). The effect of using a talking head in academic videos: An EEG study. In *Proceedings - IEEE 15th International Conference on Advanced Learning Technologies: Advanced Technologies for Supporting Open Access to Formal and Informal Learning, ICALT 2015* (pp. 367–369). IEEE. http://doi.org/10.1109/ICALT.2015.89

Domagk, S., Schwartz, R. N., & Plass, J. L. (2010). Interactivity in multimedia learning: An integrated model. *Computers in Human Behavior*, *26*(5), 1024–1033. http://doi.org/10.1016/j.chb.2010.03.003

Donkor, F. (2010). The Comparative Instructional Effectiveness of Print-Based and Video-Based Instructional Materials for Teaching Practical Skills at a Distance. *International Review of Research in Open and Distance Learning*, *1*, 96–116.

Drew, D. G., & Cadwell, R. (1985). Some Effects of Video Editing on Perceptions of Television News. *Journalism Quarterly*, *62*(4), 828–849. http://doi.org/10.1177/107769908506200417

Dunlap, B. J. C., & Lowenthal, P. R. (2010). Using Digital Music to Enhance Students ' Experience in Online Courses. *TechTrends*, *54*(4), 58–73.

Eco, U. (1979). *A Theory of Semiotics* (1st ed.). Bloomington; Indiana University Press.

Educational films. (1979). In *The Great Soviet Encyclopedia* (3rd ed.). Советская Энциклопедия.

Fahy, P. J. (2004). Media Characteristics and Online Learning Technology. In T. Anderson & F. Elloumi (Eds.), *Theory and Practice of Online Learning* (pp. 137–172).

Fairgrieve, J. (1941). Instructional Films and Educational Films. *Geography*, *26*, 77–80. http://doi.org/10.2307/40562080

Farkas, D. K. (1999). The Logical and Rhetorical Construction of Procedural Discourse. *Technical Communication*, *46*(1), 42–54.

Fernando Caro, M., & Romero Moreno, M. C. (2012). Exploración de la narrativa audiovisual para el diseño y producción de hipervideos educativos. *Escenarios*, *10*(2), 45–56.

Fill, K., & Ottewill, R. (2006). Sink or swim: taking advantage of developments in video streaming. *Innovations in Education & Teaching International*, *43*(4), 397–408. http://doi.org/10.1080/14703290600974008

Fiorella, L., & Mayer, R. E. (2016). Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology*, *108*(4), 528–546. http://doi.org/10.1037/edu0000065

Fiorella, L., van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology*, *109*(5), 653–665. http://doi.org/10.1037/edu0000161

Fischer, S., Lowe, R. K., & Schwan, S. (2008). Effects of presentation speed of a dynamic visualization on the understanding of a mechanical system. *Applied Cognitive Psychology*, *22*(8), 1126–1141. http://doi.org/10.1002/acp.1426

Fish, K., Mun, J., & A'Jontue, R. A. (2016). Do visual aids really matter? A comparison of student evaluations before and after embedding visuals into video

lectures. *Journal of Educators Online*, *13*(1), 194–217.

Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction*, *20*(3), 216–224. http://doi.org/10.1016/J.LEARNINSTRUC.2009.02.021

Flowerdew, J., & Miller, L. (1997). The teaching of academic listening comprehension and the question of authenticity. *English for Specific Purposes*, *16*(1), 27–46. http://doi.org/10.1016/S0889-4906(96)00030-0

Fortanet Gómez, I., & Bellés Fortuño, B. (2005). Spoken academic discourse : An approach to research on lectures. *Revista Española de Lingüistica Aplicada*, (25), 161–178.

Frechette, C., & Moreno, R. (2010). The roles of animated pedagogical agents' presence and nonverbal communication in multimedia learning environments. *Journal of Media Psychology*, *22*(2), 61–72. http://doi.org/10.1027/1864-1105/a000009

Fredette, M. (2013). How to Convert a Classroom Course into a MOOC. *Campus Technology*.

Fries, S., Horz, H., & Haimerl, C. (2006). Pygmalion in media-based learning: Effects of quality expectancies on learning outcomes. *Learning and Instruction*, *16*(4), 339–349. http://doi.org/10.1016/J.LEARNINSTRUC.2006.07.005

Frobenius, M. (2014). *The pragmatics of monologue: interaction in video blogs* (Doctoral thesis, Universität des Saarlandes, Germany).

Gade, P. A., & Mills, C. B. (1989). Listening Rate and Comprehension as a Function of Preference for and Exposure to Time-Altered Speech. *Perceptual and Motor Skills*, *68*(2), 531–538.

Garber, A. R. (2001). Death By Powerpoint. Retrieved November 18, 2018, from https://www.smallbusinesscomputing.com/biztools/article.php/684871/Death-By-Powerpoint.htm

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education*, *2*(2–3), 87–105. http://doi.org/10.1016/S1096-7516(00)00016-6

Giannakos, M. N., Jaccheri, L., & Krogstie, J. (2014). Looking at MOOCs Rapid Growth Through the Lens of Video-Based Learning Research. *International Journal of Emerging Technologies in Learning*, *9*(1), 35–38. http://doi.org/10.3991/ijet.v9i1.3349

Ginns, P., Martin, A. J., & Marsh, H. W. (2013). Designing Instructional Text in a Conversational Style: A Meta-analysis. *Educational Psychology Review*, *25*(4), 445–472. http://doi.org/10.1007/s10648-013-9228-0

Goldman, D. B., Gonterman, C., Curless, B., Salesin, D., & Seitz, S. M. (2008). Video object annotation, navigation, and composition. In *Proceedings of the 21st annual ACM symposium on User interface software and technology - UIST '08* (pp. 3–12). ACM. http://doi.org/10.1145/1449715.1449719

Goodyear, P. M., & Steeples, C. (1998). Creating shareable representations of practice. *Research in Learning Technology*, *6*(3), 16–23. http://doi.org/10.1080/0968776980060303

Gorissen, P., Van Bruggen, J., & Jochems, W. (2015). Does tagging improve the navigation of online recorded lectures by students? *British Journal of Educational Technology*, *46*(1), 45–57. http://doi.org/10.1111/bjet.12121

Grabe, M. E., Lombard, M., Reich, R. D., Bracken, C. C., & Ditton, T. B. (1999). The role of screen size in viewer experiences of media content. *Visual Communication Quarterly*, *6*(2), 4–9. http://doi.org/10.1080/15551399909363403

Gray, T., & Halbert, S. (1998). Team teach with a student: New approach to collaborative teaching. *College Teaching*, *46*(4), 150–153.

Gromenauer, A. A., & de la Calzada, C. (2017). Hilarious citations increase morale of doctoral students. In *The World of Today* (Vol. 1, pp. iv–vi). https://bit.ly/2I2DUKD.

Große, C. S., Jungmann, L., & Drechsler, R. (2015). Benefits of illustrations and videos for technical documentations. *Computers in Human Behavior*, *45*, 109–120. http://doi.org/10.1016/j.chb.2014.11.095

Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education*, *11*(3), 8–26. http://doi.org/10.1080/08923649709526970

Gunn, E. a a, Craenen, B. G. W., & Hart, E. (2009). A Taxonomy of Video Games and AI. *Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, AISB 2009*, 4–14.

Guo, P. J., Kim, J., & Rubin, R. (2014). How Video Production Affects Student Engagement : An Empirical Study of MOOC Videos. *L@S 2014 - Proceedings of the 1st ACM Conference on Learning at Scale*, 41–50. http://doi.org/10.1145/2556325.2566239

Halliday, M. A. K. (1978). *Language as social semiotic : the social interpretation of language and meaning*. Edward Arnold.

Halliday, Michael A. K., & Hasan, R. (1985). *Language, Context and Text: Aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press.

Halliday, Michael A. K., & Matthiessen, C. M. I. M. (2014). *Introduction to Functional Grammar*. http://doi.org/10.4324/9780203431269

Hansch, A., Hillers, L., McConachie, K., Newman, C., Schildhauer, T., & Schmidt, P. (2015). *Video and Online Learning: Critical Reflections and Findings from the Field*. *HIIG Discussion Paper Series No. 2015-02.*

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, *90*(3), 414–434. http://doi.org/10.1037/0022-0663.90.3.414

Heidig, S., Müller, J., & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in Human Behavior*, *44*, 81–95. http://doi.org/10.1016/j.chb.2014.11.009

Heller, R. S., & Martin, C. D. (1995). A Media Taxonomy. *IEEE Multimedia*, *2*(4), 36–45.

Herold, J., Stahovich, T., Lin, H., & Calfee, R. C. (2011). The Effectiveness of "Pencasts" as an Instructional Medium. In *Proceedings of the American Society for Engineering Education*.

Hiippala, T. (2014). Multimodal genre analysis. In S. Norris & C. D. Maier (Eds.), *Interactions, Images and Texts. A Reader in Multimodality* (pp. 111–124). Berlin, Boston: De Gruyter. http://doi.org/10.1515/9781614511175.111

Hinkin, M. P., Harris, R. J., & Miranda, A. T. (2014). Verbal redundancy aids memory for filmed entertainment dialogue. *Journal of Psychology: Interdisciplinary*

and Applied, *148*(2), 161–176. http://doi.org/10.1080/00223980.2013.767774

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, *17*(6), 722–738. http://doi.org/10.1016/j.learninstruc.2007.09.013

Hofmeister, A. M., Engelmann, S., & Carnine, D. (1986). Videodisc Technology: Providing Instructional Alternatives. *Journal of Special Education Technology*, *7*(3), 35–41. http://doi.org/10.1177/016264348600700304

Hollands, F. M., & Tirthali, D. (2014). Resource requirements and costs of developing and delivering MOOCs. *International Review of Research in Open and Distance Learning*, *15*(5), 113–133. http://doi.org/10.3102/0162373714557519

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior*, *24*(3), 786–797. http://doi.org/10.1016/j.chb.2007.02.009

Hong, J., Pi, Z., & Yang, J. (2016, September 20). Learning declarative and procedural knowledge via video lectures: cognitive load and learning effectiveness. *Innovations in Education and Teaching International*, pp. 1–8. http://doi.org/10.1080/14703297.2016.1237371

Hoogerheide, V., van Wermeskerken, M., Loyens, S. M. M., & van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction*, *44*, 22–30. http://doi.org/10.1016/j.learninstruc.2016.02.004

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observation on intimacy at a distance. *Psychiatry*, *19*(3), 185–206.

Hosack, B. (2010). VideoANT: Extending online video annotation beyond content delivery. *TechTrends*, *54*(3), 45–49. http://doi.org/10.1007/s11528-010-0402-7

Hu, Y., Kautz, J., Yu, Y., & Wang, W. (2015). Speaker-Following Video Subtitles. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *11*(2), 1–17. http://doi.org/10.1145/2632111

Ibrahim, M., Antonenko, P. D., Greenwood, C. M., & Wheeler, D. (2012). Effects of segmenting, signalling, and weeding on learning from educational video. *Learning, Media and Technology*, *37*(3), 220–235. http://doi.org/10.1080/17439884.2011.585993

Iedema, R. (2001). Analysing film and television: A social semiotic account of hospital: An unhealthy business. In T. van Leeuwen & C. Jewitt (Eds.), *The Handbook of Visual Analysis* (pp. 181–204). Los Angeles: Sage.

IEEE Standard for Learning Object Metadata. (2002). IEEE Computer Society. http://doi.org/10.1109/IEEESTD.2002.94128

Ilie, G., & Thompson, W. F. (2011). Experiential and cognitive changes following seven minutes exposure to music and speech. *Music Perception: An Interdisciplinary Journal*, *28*(3), 247–264.

Jaimes, A., & Dimitrova, N. (2006). Human-centered multimedia: Culture, deployment, and access. *IEEE Multimedia*, *13*(1), 12–19. http://doi.org/10.1109/MMUL.2006.8

Jaimes, A., Sebe, N., & Gatica-Perez, D. (2006). Human-Centered Computing : A Multimedia Perspective. In *MM '06 Proceedings of the 14th ACM international conference on Multimedia* (pp. 855–864). Santa Barbara, US. http://doi.org/10.1145/1180639.1180829

Johnson, G. J., Bruner II, G. C., & Kumar, A. (2006). Interactivity and its Facets Revisited: Theory and Empirical Test. *Journal of Advertising*, *35*(4), 35–52.

http://doi.org/10.2753/JOA0091-3367350403

Jonassen, D. H., Grabinger, R. S., & Harris, N. D. C. (1991). Analyzing and Selecting Instructional Strategies and Tactics. *Performance Improvement Quarterly*, *4*(2), 77–97. http://doi.org/10.1111/j.1937-8327.1991.tb00505.x

Kaplan, R. M., & Pascoe, G. C. (1977). Humorous lectures and humorous examples: Some effects upon comprehension and retention. *Journal of Educational Psychology*, *69*(1), 61–65. http://doi.org/10.1037/0022-0663.69.1.61

Karsenti, T. (2013). MOOC: What the research says. International Review of Technology in Higher Education. *International Journal of Technologies in Higher Education*, *10*(2), 23–37.

Kartal, G. (2010). Does language matter in multimedia learning? Personalization principle revisited. *Journal of Educational Psychology*, *102*(3), 615–624. http://doi.org/10.1037/a0019345

Kay, R. H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior*, *28*(3), 820–831. http://doi.org/10.1016/j.chb.2012.01.011

Kay, R. H. (2014). Developing a Framework for Creating Effective Instructional Video Podcasts. *International Journal of Emerging Technologies in Learning (IJET)*, *9*(1), 22–30.

Kellogg, S. (2013). How to make a MOOC. *Nature2*, *499*, 369–371. http://doi.org/10.1126/science.7939709

Kelly, S. D., Creigh, P., & Bartolotti, J. (2010). Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. *Journal of Cognitive Neuroscience*, *22*(4), 683–694. http://doi.org/10.1162/jocn.2009.21254

Kervellec, A.-L., Jamet, E., Dardier, V., Erhel, S., Le Maner-Idrissi, G., & Michinov, E. (2016). A Study of Gender Similarity Between Animated Pedagogical Agents and Young Learners (pp. 510–517). Springer, Cham. http://doi.org/10.1007/978-3-319-39483-1_46

Kim, D., & Kim, D.-J. (2012). Effect of screen size on multimedia vocabulary learning. *British Journal of Educational Technology*, *43*(1), 62–70. http://doi.org/10.1111/j.1467-8535.2010.01145.x

Kim, J., Guo, P. J., Cai, C. J., Li, S.-W., Gajos, K. Z., & Miller, R. C. (2014). Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14* (pp. 563–572). New York, New York, USA: ACM Press. http://doi.org/10.1145/2642918.2647389

Kim, J., Nguyen, P. T., Weir, S., Guo, P. J., Miller, R. C., & Gajos, K. Z. (2014). Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 4017–4026). http://doi.org/10.1145/2556288.2556986

Kishler, J. P. (1950). *The effects of prestige and identification factors in attitude restructuring and learning from sound films. Technical Report SDC 268-7-10. The Instructional Film Research Program*. Port Washington, US.

Kizilcec, Rene F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: evidence from two large-scale field studies. *Journal of Educational Psychology*, *107*(3), 724–739. http://doi.org/10.1037/edu0000013

Kizilcec, René F., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction: Effects on Information Retention, Visual Attention, and Affect. In *Proceedings of the 32nd annual ACM conference on Human factors in*

computing systems - CHI '14 (pp. 2095–2102). http://doi.org/10.1145/2556288.2557207

Korving, H., Hernández, M., & De Groot, E. (2016). Look at me and pay attention! A study on the relation between visibility and attention in weblectures. *Computers & Education*, *94*, 151–161. http://doi.org/10.1016/j.compedu.2015.11.011

Koumi, J. (1991). Narrative Screenwriting for Educational Television: A framework. *Journal of Educational Television*, *17*(3), 131–148. http://doi.org/10.1080/1358165910170302

Koumi, J. (2006). *Designing Video and Multimedia for Open and Flexible Learning*. Routledge.

Koumi, J. (2015). Learning outcomes afforded by self-assessed, segmented video–print combinations. *Cogent Education*.

Kozma, R. B. (1991). Learning with Media. *Review of Educational Research*, *61*(2), 179–211. http://doi.org/10.3102/00346543061002179

Kress, G., & van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design*. (Routled, Ed.) (2nd ed.). Routledge.

Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs . L2. *Journal of Eye Movement Research*, *7*(5), 1–15. http://doi.org/10.16910/jemr.7.5.4

Kühl, T., & Zander, S. (2017). An inverted personalization effect when learning with multimedia: The case of aversive content. *Computers and Education*, *108*, 71–84. http://doi.org/10.1016/j.compedu.2017.01.013

Landström, S. (2008). *CCTV, Live and Videotapes: How Presentation Mode Affects the Evaluation of Witnesses.* (Doctoral thesis, University of Gothenburg, Sweden).

Lång, J. (2016). Subtitles vs. narration: The acquisition of information from visual-verbal and audio-verbal channels when watching a television documentary. In S. Hansen-Schirra & S. Grucza (Eds.), *Eyetracking and Applied Linguistics* (pp. 59–82). http://doi.org/10.17169/langsci.b108.230

Lavaur, J.-M., & Bairstow, D. (2011). Languages on the screen: Is film comprehension related to the viewers' fluency level and to the language in the subtitles? *International Journal of Psychology*, *46*(6), 455–462. http://doi.org/10.1080/00207594.2011.565343

Lea, D. (1994). Christopher Alexander: An Introduction for Object-Oriented Designers. *Software Engineering Notes*, *19*(1), 39–46.

Lee, D. (2006). Humor in spoken academic discourse. *NUCB Journal of Language Culture and Communication*, *8*(1), 49–68.

Lee, H., & Doh, Y. Y. (2012). A Study on the Relationship between Educational Achievement and Emotional Engagement in a Gameful Interface for Video Lecture Systems. In *2012 International Symposium on Ubiquitous Virtual Reality* (pp. 34–37). IEEE. http://doi.org/10.1109/ISUVR.2012.21

Lee, J. J. (2009). Size matters: an exploratory comparison of small- and large-class university lecture introductions. *English for Specific Purposes*, *28*(1), 42–57. http://doi.org/10.1016/j.esp.2008.11.001

Lemke, J. L. (2005). Multimedia Genres and Traversals. *Folia Linguistica*, *39*(1–2), 45–56. http://doi.org/10.1515/flin.2005.39.1-2.45

Lents, N. H., & Cifuentes, O. E. (2009). Web-Based Learning Enhancements: Video Lectures Through Voice-Over PowerPoint in a Majors-Level Biology Course. *Journal of College Science Teaching*, *39*(2), 38–46.

Leshem, Y. (2018). The Hidden Reason Behind The Increase of Video Usage In Enterprises. Retrieved February 23, 2018, from https://www.entrepreneur.com/article/307092

Li, F. C., Gupta, A., Sanocki, E., He, L., & Rui, Y. (2000). Browsing digital video. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00* (pp. 169–176). New York, New York, USA: ACM Press. http://doi.org/10.1145/332040.332425

Li, J., Kizilcec, R., Bailenson, J., & Ju, W. (2015). Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior*, *55*, 1222–1230. http://doi.org/10.1016/j.chb.2015.04.005

Li, N., Kidzinski, L., Jermann, P., & Dillenbourg, P. (2015). How Do In-video Interactions Reflect Perceived Video Difficulty? In-Video Interaction Analysis. *Proceedings of the Third European MOOCs Stakeholder Summit*, *1*(17), 112–121.

Li, N., Kidziński, L., Jermann, P., & Dillenbourg, P. (2015). MOOC video interaction patterns: What do they tell us? *Lecture Notes in Computer Science*, *9307*(1), 197–210. http://doi.org/10.1007/978-3-319-24258-3_15

Lim-Fei, V., & O'Halloran, K. L. (2014). Systemic functional multimodal discourse analysis. In S. Norris & C. D. Maier (Eds.), *Interactions, Images and Texts. A Reader in Multimodality* (pp. 137–154). De Gruyter.

Lim Fei, V., O'Halloran, K. L., Tan, S., & E, M. K. L. (2015). Teaching visual texts with the multimodal analysis software. *Educational Technology Research and Development*, *63*(6), 915–935. http://doi.org/10.1007/s11423-015-9395-4

Lim, V. F. (2011). *A Systemic Functional Multimodal Discourse Analysis Approach to Pedagogic Discourse* (Doctoral thesis, National University of Singapore).

Lin, L., Atkinson, R. K., Savenye, W. C., & Nelson, B. C. (2016). Effects of visual cues and self-explanation prompts: Empirical evidence in a multimedia environment. *Interactive Learning Environments*, *24*(4), 799–813. http://doi.org/10.1080/10494820.2014.924531

Liu, T., & Kender, J. R. (2002). Rule-based semantic summarization of instructional videos. *Proceedings. 2002 International Conference on Image Processing.*, *1*, I-601-I–604 vol.1. http://doi.org/10.1109/ICIP.2002.1038095

Liyanagunawardena, T. R., Adams, A. A., Williams, S. A., & Rekha Liyanagunawardena, T. (2013). MOOCs: a systematic study of the published literature 2008- 2012. *The International Review of Research in Open and Distance Learning*, *14*(3), 202–227. http://doi.org/10.3329/bjms.v12i4.16658

Loch, B., & Mcloughlin, C. (2011). An instructional design model for screencasting: Engaging students in self-regulated learning. *ASCILITE 2011 Changing Demands, Changing Directions.*, 816–821.

Lyons, A., Reysen, S., & Pierce, L. (2012). Video lecture format, student technological efficacy, and social presence in online courses. *Computers in Human Behavior*, *28*(1), 181–186. http://doi.org/10.1016/j.chb.2011.08.025

Mamgain, N., Sharma, A., & Goyal, P. (2015). Learner's perspective on video-viewing features offered by MOOC providers: Coursera and edX. In *Proceedings of the 2014 IEEE International Conference on MOOCs, Innovation and Technology in Education, IEEE MITE 2014* (pp. 331–336). IEEE. http://doi.org/10.1109/MITE.2014.7020298

Maniar, N., Bennett, E., Hand, S., & Allan, G. (2008). The effect of mobile phone screen size on video based learning. *Journal of Software*, *3*(4), 51–61. http://doi.org/10.4304/jsw.3.4.51-61

Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*. http://doi.org/10.1515/text.1.1988.8.3.243

Martin, D. M., Preiss, R. W., Gayle, B. M., & Allen, M. (2006). A meta-analytic assessment of the effect of humorous lectures on learning. In *Classroom Communication and Instructional Processes: Advances through meta-analysis* (pp. 295–313).

Martin, J. R., & White, P. R. R. (2005). *The Language of Evaluation. Appraisal in English*. Palgrave Macmillan. http://doi.org/10.1057/9780230511910

Martin, James Robert. (1994). Modelling big texts: a systemic funcional approach to multi-genericity. *Network 21*, 29–52.

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, *93*(2), 377–389. http://doi.org/10.1037/0022-0663.93.2.377

Mayer, R. E. (1982). Learning. In H. E. Mitzel, J. H. Best, W. Rabinowitz, & A. E. R. Association (Eds.), *Encyclopedia of educational research* (5th ed., pp. 1040–1058). Free Press.

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation*, *41*, 85–139. http://doi.org/10.1016/S0079-7421(02)80005-6

Mayer, R. E. (2014a). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Second Edi, pp. 43–71). Cambridge University Press.

Mayer, R. E. (2014b). Incorporating motivation into multimedia learning. *Learning and Instruction*, *29*, 171–173. http://doi.org/10.1016/j.learninstruc.2013.04.003

Mayer, R. E. (2014c). Research--Based Principles for Designing Multimedia Instruction Overview of Multimedia Instruction.

Mayer, R. E. (Ed.). (2014d). *The Cambridge Handbook of Multimedia Learning* (Second Edi). Cambridge University Press.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology*, *93*(2), 390–397. http://doi.org/10.1037/0022-0663.93.2.390

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A Personalization Effect in Multimedia Learning: Students Learn Better When Words Are in Conversational Style Rather Than Formal Style. *Journal of Educational Psychology*, *96*(2), 389–395. http://doi.org/10.1037/0022-0663.96.2.389

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*(1), 187–198.

Mayer, R. E., Johnson, W. L., Shaw, E., & Sandhu, S. (2006). Constructing computer-based tutors that are socially sensitive: Politeness in educational software. *International Journal of Human-Computer Studies*, *64*(1), 36–42. http://doi.org/10.1016/J.IJHCS.2005.07.001

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social Cues in Multimedia Learning : Role of Speaker's Voice, *95*(2), 419–425. http://doi.org/10.1037/0022-0663.95.2.419

McClusky, F. D. (1947). The Nature of the Educational Film. *Hollywood Quarterly*, *2*(4), 371–380. http://doi.org/10.2307/1209533

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*,

264(5588), 746–748. http://doi.org/10.1038/264746a0

McIntyre, N. A., Mainhard, M. T., & Klassen, R. M. (2017). Are you looking to teach? Cultural, temporal and dynamic insights into expert teacher gaze. *Learning and Instruction*, *49*, 41–53. http://doi.org/10.1016/j.learninstruc.2016.12.005

McNeil, B. J., & Nelson, K. R. (1991). Meta-Analysis of Interactive Video Instruction: A 10 Year Review of Achievement Effects. *Journal of Computer-Based Instruction*, *18*(1), 1–6.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*.

Medina, J. (2008). *Brain Rules*. Seattle, US: Pear Press.

Meier, B. P., & Robinson, M. D. (2004). Why the Sunny Side Is Up: Associations Between Affect and Vertical Position. *Psychological Science*, *15*(4), 243–247. http://doi.org/10.1111/j.0956-7976.2004.00659.x

Meixner, B., Matusik, K., Grill, C., & Kosch, H. (2014). Towards an easy to use authoring tool for interactive non-linear video. In *Multimedia Tools and Applications* (Vol. 70, pp. 1251–1276). http://doi.org/10.1007/s11042-012-1218-6

Merkt, M., & Schwan, S. (2014). Training the use of interactive videos: Effects on mastering different tasks. *Instructional Science*, *42*(3), 421–441. http://doi.org/10.1007/s11251-013-9287-0

Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction*, *21*(6), 687–704. http://doi.org/10.1016/j.learninstruc.2011.03.004

Meseguer-Martinez, A., Ros-Galvez, A., & Rosa-Garcia, A. (2017). Satisfaction with online teaching videos: A quantitative approach. *Innovations in Education and Teaching International*, *54*(1), 62–67. http://doi.org/10.1080/14703297.2016.1143859

Messina, A., Montagnuolo, M., & Sapino, M. L. (2006). Characterizing Multimedia Objects Through Multimodal Content Analysis and Fuzzy Finerprints. *Sitis 2006*, 154–165.

Metz, C. (1966). La grande syntagmatique du film narratif. *Communications*, *8*(1), 120–124.

Metz, C. (1974). *Film language: a semiotics of the cinema*. Oxford University Press and Chicago Universtity Press.

Minnesota State University. (1998). Social Presence Theory.

Mitra, B., Lewin-Jones, J., Barrett, H., & Williamson, S. (2010). The use of video to enable deep learning. *Research in Post-Compulsory Education*, *15*(4), 405–414. http://doi.org/10.1080/13596748.2010.526802

Moes, S. (2012). REC:all Framework. Retrieved from http://www.rec-all.info/profiles/blogs/definitions-of-various-formats-of-captured-lectures-in-relation

Mohamad Ali, A. Z., Samsudin, K., Hassan, M., & Sidek, S. F. (2011). Does Screencast Teaching Software Application Needs Narration for Effective Learning?. *Turkish Online Journal of Educational Technology - TOJET*, *10*(3), 76–82.

Molina Plaza, S., & Argüelles Álvarez, I. (2013). University large lectures in MICASE: A systemic functional analysis. *Revista Española de Lingüística Aplicada*, (Extraordinary 1), 183–207.

Monserrat, T., & Zhao, S. (2013). NoteVideo: facilitating navigation of blackboard-style lecture videos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1139–1148). ACM. http://doi.org/doi: 10.1145/2466110.2466147

Morain, M., & Swarts, J. (2012). YouTutorial: A Framework for Assessing Instructional Online Video. *Technical Communication Quarterly*, *21*(1), 6–24. http://doi.org/10.1080/10572252.2012.626690

Morales Morante, L. F. (2012). Structure and meaning of television news: Parameters for the construction and analysis of the message in the audiovisual environment. *Estudios Sobre El Mensaje Periodistico*, *18*(2), 805–821.

Morell, T. (2015). International conference paper presentations: A multimodal analysis to determine effectiveness. *English for Specific Purposes*, *37*, 137–150. http://doi.org/10.1016/j.esp.2014.10.002

Moreno, R. (2005). Instructional Technology: Promise and Pitfalls. In L. Pytlik-Zillig, M. Bodvarsson, & R. Bruning (Eds.), *Technology-based education: Bringing researchers and practitioners together* (pp. 1–19). Information Age Publishing.

Moreno, R., & Mayer, R. E. (2000a). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, *92*(1), 117–125. http://doi.org/10.1037//0022

Moreno, R., & Mayer, R. E. (2000b). Engaging Students in Active Learning: The Case for Personalized Multimedia Messages. *Journal of Educational Psychology*, *92*(4), 724–733. http://doi.org/10.1037//00224)663.92.4.724

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, *94*(1), 156–163. http://doi.org/10.1037//0022-0663.94.1.156

Moreno, R., & Mayer, R. E. (2005). Role of Guidance, Reflection, and Interactivity in an Agent-Based Multimedia Game. *Journal of Educational Psychology*, *97*(1), 117–128. http://doi.org/10.1037/0022-0663.97.1.117

Moreno, R., & Mayer, R. E. (2007). Interactive Multimodal Learning Environments. *Educational Psychology Review*, *19*(3), 309–326. http://doi.org/10.1007/s10648-007-9047-2

Morris, C., & Chikwa, G. (2014). Screencasts: How effective are they and how do students engage with them? *Active Learning in Higher Education*, *15*(1), 25–37. http://doi.org/10.1177/1469787413514654

Mujacic, S., Debevc, M., Kosec, P., Bloice, M., & Holzinger, A. (2012). Modeling, design, development and evaluation of a hypervideo presentation for digital systems teaching and learning. *Multimedia Tools and Applications*, *58*(2), 435–452. http://doi.org/10.1007/s11042-010-0665-1

Muldner, K., Lam, R., & Chi, M. T. H. (2014). Comparing learning from observing and from human tutoring. *Journal of Educational Psychology*, *106*(1), 69–85. http://doi.org/10.1037/a0034448

Muller, D. A., Bewes, J., Sharma, M. D., & Reimann, P. (2008). Saying the wrong thing: Improving learning with multimedia by including misconceptions. *Journal of Computer Assisted Learning*, *24*(2), 144–155. http://doi.org/10.1111/j.1365-2729.2007.00248.x

Myers, G. (2003). Words, Pictures, and Facts in Academic Discourse. *Ibérica: Revista*

de La Asociación Europea de Lenguas Para Fines Específicos (AELFE), (6), 3–13.

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. http://doi.org/10.1111/0022-4537.00153

Nesi, H. (2012). Laughter in university lectures. *Journal of English for Academic Purposes*, *11*(2), 79–89. http://doi.org/10.1016/j.jeap.2011.12.003

Nguyen, C., & Liu, F. (2015). Making Software Tutorial Video Responsive. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 1565–1568. http://doi.org/10.1145/2702123.2702209

Nguyen, C., & Liu, F. (2016). Gaze-based Notetaking for Learning from Lecture Videos. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2093–2097). ACM. http://doi.org/10.1145/2858036.2858137

Nichols, B. (2001). *Introduction to Documentary* (1st ed.). Indiana University Press.

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, *22*(3), 336–359. http://doi.org/10.1057/ejis.2012.26

Nordkvelle, Y. T., Fritze, Y., & Haugsbakk, G. (2010). The visual in teaching – from Bologna to YouTubiversity Precursor : The lecture as learning object. In R. Maier & T. Hug (Eds.), *Medien - Wissen - Bildung: Explorationen visualisierter und kollaborativer Wissensräume* (pp. 59–71). Innsbruck: Innsbruck University Press.

Northrop, D. S. (1952). *Effects on learning of the prominence of organizational outline in instructional films. Human Engineering Report, SDC 269-7-33*. Santa Monica, US.

Nugent, G. C., Tipton, T. J., & Brooks, D. W. (1980). Task, learner, and presentation interactions in television production. *Educational Communication and Technology Journal*, *28*(1), 30–38. http://doi.org/10.1007/bf02791378

O'Halloran, K. L. (2008). Systemic functional-multimodal discourse analysis (SF-MDA): constructing ideational meaning using language and visual imagery. *Visual Communication*, *7*(4), 443–475. http://doi.org/10.1177/1470357208096210

O'Halloran, K. L. (2009). Multimodal Analysis and Digital Technology. In A. Baldry & E. Montagna (Eds.), *Interdisciplinary Perspectives on Multimodality: Theory and Practice. Proceedings of the Third International Conference on Multimodality* (p. 26). Palladino, Campobasso.

Orgeron, D., Orgeron, M., & Streible, D. (Eds.). (2011). *Learning with the Lights Off: Educational Film in the United States*. Oxford University Press.

Ormrod, J. (2017). *Educational psychology: Developing learners*. *Educational psychology: Developing learners*. Pearson/Merrill Prentice Hall. http://doi.org/10.1016/0257-8972(86)90040-X

Ouzts, A. D., Snell, N. E., Maini, P., & Duchowski, A. T. (2013). Determining Optimal Caption Placement Using Eye Tracking. In *Proceedings of the 31st ACM international conference on Design of communication* (pp. 189–190). http://doi.org/10.1145/2507065.2507100

Ozan, O., & Ozarslan, Y. (2016). Video lecture watching behaviors of learners in online courses. *Educational Media International*, *53*(1), 27–41. http://doi.org/10.1080/09523987.2016.1189255

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior*, *26*(1), 110–117. http://doi.org/10.1016/j.chb.2009.09.001

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University

Press. http://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Paltridge, B. (n.d.). Genre and English for Specific Purposes. Retrieved January 21, 2018, from https://genreacrossborders.org/genre-and-english-specific-purposes

Paltridge, B. (2012). *Discourse Analysis* (2nd ed.). Bloomsbury Academic.

Park, B., Knörzer, L., Plass, J. L., & Brünken, R. (2015). Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms. *Computers & Education*, *86*, 30–42. http://doi.org/10.1016/J.COMPEDU.2015.02.016

Passey, D. (2006). Digital video technologies enhancing learning for pupils at risk and those who are hard to reach. In M. Childs, M. Cuttle, & K. Riley (Eds.), *DIVERSE proceedings : 2005 & 2006 : 5th International DIVERSE Conference, 5th-7th July 2005, Vanderbilt University, Nashville, USA, 6th International DIVERSE Conference, 5th-7th July 2006, Glasgow Caledonian University, Glasgow, UK* (pp. 156–168). Glasgow Caledonian University Press.

Pi, Z., Hong, J., & Yang, J. (2017). Does instructor's image size in video lectures affect learning outcomes? *Journal of Computer Assisted Learning*, *33*(4), 347–354. http://doi.org/10.1111/jcal.12183

Pi, Zhongling, & Hong, J. (2016). Learning process and learning outcomes of video podcasts including the instructor and PPT slides: a Chinese case. *Innovations in Education and Teaching International*, *53*(2), 135–144. http://doi.org/10.1080/14703297.2015.1060133

Plantinga, C. B. (1997). *Rhetoric and Representation in Nonfiction Film* (1st ed.). Cambridge University Press.

Ploetzner, R., & Lowe, R. (2012). A systematic characterisation of expository animations. *Computers in Human Behavior*, *28*(3), 781–794. http://doi.org/http://dx.doi.org/10.1016/j.chb.2011.12.001

Pope, S. T. (1994). A Taxonomy of Computer Music. *Computer Music Journal*, *18*(1), 5–7.

Poyatos, F. (1983). Language and nonverbal systems in the structure of face-to-face interaction. *Language and Communication*, *3*(2), 129–140. http://doi.org/10.1016/0271-5309(83)90010-1

Pozzer-Ardenghi, L., & Roth, W.-M. (2007). On Performing Concepts During Science Lectures. *Science Education*, *91*(1), 96–114. http://doi.org/10.1002/sce.20172

Purcell, K. (2013). *Online Video 2013*.

Pusic, M. V., LeBlanc, V. R., & Miller, S. Z. (2007). Linear Versus Web-Style Layout of Computer Tutorials for Medical Student Learning of Radiograph Interpretation. *Academic Radiology*, *14*(7), 877–889. http://doi.org/10.1016/j.acra.2007.04.013

Reeves, B., Lang, A., Kim, E. Y., & Tatar, D. (1999). The Effects of Screen Size and Message Content on Attention and Arousal. *Media Psychology*, *1*(1), 49–67. http://doi.org/10.1207/s1532785xmep0101

Reeves, B., & Nass, C. (1996). *The Media Equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press.

Reigeluth, C. M. (1983). Instructional Design: What is it and why is it? In C. M. Reigeluth (Ed.), *Instructional theories in action* (p. 5). Hillsdale, NJ: Lawrence Erlbaum Associates.

Reigeluth, C. M., & Carr-Chellman, A. A. (2009). Understanding Instructional Theory. In C. M. Reigeluth & A. A. Carr-Chellman (Eds.), *Instructional-Design*

Theories and Models. Volume III. Building a Common Knowledge Base. Routledge.

Reiss, D. (2008). Video-based multimedia designs: A research study testing learning effectiveness. *Canadian Journal of Learning and Technology / La Revue Canadienne de l'apprentissage et de La Technologie*, *33*(3). http://doi.org/10.21432/T2FG64

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, *7*(3), 216–237. http://doi.org/10.1016/j.edurev.2012.05.003

Rich, P. J., & Hannafin, M. (2009). Video Annotation Tools: Technologies to Scaffold, Structure, and Transform Teacher Reflection. *Journal of Teacher Education*, *60*(1), 52–67. http://doi.org/10.1177/0022487108328486

Richardson, J. C., & Swan, K. (2003). Examining social presence in online courses in relation to students' perceived learning and satisfaction. *Journal of Asynchronous Learning Network*, *7*(1), 68–88. http://doi.org/10.1016/j.pec.2009.03.021

Rist, T., André, E., & Müller, J. (1997). Adding animated presentation agents to the interface. In *Proceedings of the 2nd international conference on Intelligent User Interfaces* (pp. 79–86). http://doi.org/10.1145/238218.238298

Rodriguez, O. C. (2012). MOOCs and the AI-Stanford Like Courses: Two Successful and Distinct Course Formats for Massive Open Online Courses. *European Journal of Open, Distance and E-Learning*, *2*, 13.

Roshal, S. M. (1949). *Effects of Learner Representation in Film-Mediated Perceptual-Motor Learning. Technical Report SDC 269-7-5*. Port Washington, US.

Roth, W.-M. (2001). Gestures: Their Role in Teaching and Learning. *Review of Educational Research*, *71*(3), 365–392. http://doi.org/10.3102/00346543071003365

Ruoff, J. K. (1992). Conventions of Sound in Documentary. In R. Altman (Ed.), *Sound Theory, Sound Practice* (pp. 217–234). Routledge.

Salomon, G. (1979a). *Interaction of Media, Cognition, and Learning*. L. Erlbaum Associates.

Salomon, G. (1979b). Media and symbol systems as related to cognition and learning. *Journal of Educational Psychology*, *71*(2), 131–148. http://doi.org/10.1037/0022-0663.71.2.131

Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, *76*(4), 647–658. http://doi.org/10.1037/0022-0663.76.4.647

Salomon, G. (1994). *Interaction of Media, Cognition and Learning: An Exploration of How Symbolic Forms Cultivate Mental Skills and Affect Knowledge Acquisition*. Routledge.

Sanchez, C. A., & Khan, S. (2016). Instructor accents in online education and their effect on learning and attitudes. *Journal of Computer Assisted Learning*, *32*(5), 494–502. http://doi.org/10.1111/jcal.12149

Santos-Espino, J. M., Afonso-Suárez, M. D., & Guerra-Artal, C. (2016). Speakers and Boards : A Survey of Instructional Video Styles in MOOCs. *Technical Communication*, *63*(2), 101–115.

Santos Espino, J. M., Afonso Suárez, M. D., Guerra Artal, C., & García-Sánchez, S. (2013). Measuring the Quality of Instructional Videos for Higher Education. In *6th International Conference of Education, Research and Innovation. ICERI 2013.* (pp. 1053–1060). Seville: IATED.

Schacter, D. L., & Szpunar, K. K. (2015). Enhancing Attention and Memory During Video-Recorded Lectures. *Scholarship of Teaching and Learning in Psychology*, *1*(1),

60–71. http://doi.org/10.1037/stl0000011

Schittek Janda, M., Tani Botticelli, A., Mattheos, N., Nebel, D., Wagner, A., Nattestad, A., & Attstrom, R. (2005). Computer-mediated instructional video: a randomised controlled trial comparing a sequential and a segmented instructional video in surgical hand wash. *European Journal of Dental Education*, *9*(2), 53–58. http://doi.org/10.1111/j.1600-0579.2004.00366.x

Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R., Abrami, P. C., Wade, C. A., ... Lowerison, G. (2009). Technology's effect on achievement in higher education: a Stage I meta-analysis of classroom applications. *Journal of Computing in Higher Education*, *21*(2), 95–109. http://doi.org/10.1007/s12528-009-9021-8

Schmidt, W. D. (1972). *Design Elements in instructional films: an attempt to derive some operational generalizations based on research and on producer opinion* (Doctoral thesis, The Ohio State University).

Schneider, S., Nebel, S., Pradel, S., & Rey, G. D. (2015). Mind your Ps and Qs! How polite instructions affect learning with multimedia. *Computers in Human Behavior*, *51*(PA), 546–555. http://doi.org/10.1016/j.chb.2015.05.025

Schneider, S., Nebel, S., & Rey, G. D. (2016). Decorative pictures and emotional design in multimedia learning. *Learning and Instruction*, *44*, 65–73. http://doi.org/10.1016/j.learninstruc.2016.03.002

Schoeffmann, K., Hopfgartner, F., Marques, O., Boeszoermenyi, L., & Jose, J. M. (2010). Video browsing interfaces and applications: a review. *SPIE Reviews*, *1*(1), 18004–18035. http://doi.org/10.1117/6.0000005

Schroeder, N. L., & Adesope, O. O. (2015). Impacts of pedagogical agent gender in an accessible learning environment. *Educational Technology and Society*, *18*(4), 401–411.

Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How Effective are Pedagogical Agents for Learning? A Meta-Analytic Review. *Journal of Educational Computing Research*, *49*(1), 1–39. http://doi.org/10.2190/EC.49.1.a

Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction*, *14*(3), 293–305. http://doi.org/10.1016/j.learninstruc.2004.06.005

Schwartz, D. L., & Hartman, K. (2007). It is not television anymore: Designing digital video for learning and assessment. In R. Goldman, R. D. Pea, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 335–348). Routledge.

Schwier, R. A. (1992). A Taxonomy of Interaction for Instructional Multimedia. In *Annual Conference of the Association for Media and Technology in Education in Canada*. Vancouver, Canada.

Seaton, D. T., Nesterko, S., Mullaney, T., Reich, J., & Ho, A. D. (2014). Characterizing Video Use in the Catalogue of MITx MOOCs. *Elearning Papers*, *37*(March), 33–41.

Seel, N. M. (Ed.). (2012). *Encyclopedia of the Sciences of Learning*. Springer US. http://doi.org/10.1007/978-1-4419-1428-6

Seidel, N. (2015). Interaction design patterns for spatio-temporal annotations in video learning environments. In *Proceedings of the 20th European Conference on Pattern Languages of Programs - EuroPLoP '15* (pp. 1–21). Kaufbeuren, Germany: ACM. http://doi.org/10.1145/2855321.2855338

Sekula, A. (1982). On the Invention of Photographic Meaning. In *Thinking Photography* (pp. 84–109). London: Macmillan Education UK. http://doi.org/10.1007/978-1-349-16716-6_5

Shah, D. (2015). By the numbers: MOOCs in 2015. Retrieved January 6, 2016, from http://www.class-central.com/report/moocs-2015-stats/

Sharma, K., D'Angelo, S., Gergle, D., & Dillenbourg, P. (2016). Visual augmentation of deictic gestures in MOOC videos. *Proceedings of International Conference of the Learning Sciences, ICLS*, *1*(2012), 202–209.

Sharma, K., Jermann, P., & Dillenbourg, P. (2014). "With-Me-Ness": A Gaze-Measure for Students' Attention in MOOCs With-Me-Ness. *Proceedings of the International Conference of the Learning Sciences*, (2010), 1017–1021.

Sharma, K., Jermann, P., & Dillenbourg, P. (2015). Displaying teacher's gaze in a MOOC: Effects on students' video navigation patterns. In *10th European Conference on Technology Enhanced Learning, EC-TEL 2015, Toledo, Spain, September 15-18, 2015, Proceedings* (Vol. 9307, pp. 325–338). Springer International Publishing. http://doi.org/10.1007/978-3-319-24258-3_24

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. *Interactive Learning Environments*, *15*(1), 27–46. http://doi.org/10.1080/10494820600996972

Shephard, K. (2003). Streaming Video To Support Student Learning. *British Journal of Educational Technology*, *34*(3), 295–308. http://doi.org/10.1111/1467-8535.00328

Shephard, K., Ottewill, R., Phillips, P., & Collier, R. (2003). From videocassette to video stream: Issues involved in re-purposing an existing educational video. *Alt-J*, *11*(2), 14–22. http://doi.org/10.1080/0968776030110203

Shuell, T. J. (1986). Cognitive Conceptions of Learning. *Review of Educational Research*, *56*(4), 411–436. http://doi.org/10.2307/1170340

Simhony, M., Grinberg, M., Lavie, L., & Banai, K. (2014). Rapid adaptation to time-compressed speech in young and older adults. *Journal of Basic and Clinical Physiology and Pharmacology*, *25*(3), 285–288. http://doi.org/10.1515/jbcpp-2014-0023

Simonds, B. K., Meyer, K. R., Quinlan, M. M., & Hunt, S. K. (2006). Effects of Instructor Speech Rate on Student Affective Learning, Recall, and Perceptions of Nonverbal Immediacy, Credibility, and Clarity. *Communication Research Reports*, *23*(3), 187–197. http://doi.org/10.1080/08824090600796401

Sinclair, J., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press.

Sitzmann, T., & Johnson, S. (2014). The paradox of seduction by irrelevant details: How irrelevant information helps and hinders self-regulated learning. *Learning and Individual Differences*, *34*, 1–11. http://doi.org/10.1016/j.lindif.2014.05.009

Smith III, J. O. (2011). Time Scale Modification. In *Spectral Audio Signal Processing* (pp. 375–385). W3K Publishing.

Smith, P. L., & Ragan, T. J. (2005). *Instructional Design* (3rd ed.). Wiley.

Snelson, C., & Perkins, R. A. (2009). From Silent Film to YouTube ™ : Tracing the Historical Roots of Motion Picture Technologies in Education. *Journal of Visual Literacy*, *28*(1), 1–27. http://doi.org/10.1080/23796529.2009.11674657

Song, S., Hong, J., Oakley, I., Cho, J. D., & Bianchi, A. (2015). Automatically Adjusting the Speed of E-Learning Videos. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15* (pp. 1451–1456). New York, New York, USA: ACM Press. http://doi.org/10.1145/2702613.2732711

Spanjers, I. A. E., van Gog, T., & van Merriënboer, J. J. G. (2010). A Theoretical

Analysis of How Segmentation of Dynamic Visualizations Optimizes Students' Learning. *Educational Psychology Review*, *22*(4), 411–423. http://doi.org/10.1007/s10648-010-9135-6

Spanjers, I. A. E., Van Gog, T., Wouters, P., & Van Merriënboer, J. J. G. (2012). Explaining the segmentation effect in learning from animations: The role of pausing and temporal cueing. *Computers and Education*, *59*(2), 274–280. http://doi.org/10.1016/j.compedu.2011.12.024

Spanjers, I. A. E., Wouters, P., van Gog, T., & van Merriënboer, J. J. G. (2011). An Expertise Reversal Effect of Segmentation in Learning from Animated Worked-out Examples. *Computers in Human Behavior*, *27*(1), 46–52.

Stetz, T. A., & Bauman, A. A. (2013). Reasons to Rethink the Use of Audio and Video Lectures in Online Courses. *Higher Learning Research Communications*, *3*(1), 38–45. http://doi.org/10.18870/hlrc.v3i4.168

Subrahmanyam, K., Michikyan, M., Clemmons, C., Carrillo, R., Uhls, Y. T., & Greenfield, P. M. (2013). Learning from Paper, Learning from Screens. *International Journal of Cyber Behavior, Psychology and Learning*, *3*(4), 1–27. http://doi.org/10.4018/ijcbpl.2013100101

Sugar, W., Brown, A., & Luterbach, K. (2010). Examining the anatomy of a screencast: Uncovering common elements and instructional strategies. *The International Review of Research in Open and Distributed Learning*, *11*(3), 1–20.

Sumner, W. L. (1950). *Visual Methods in Education*. Oxford: Basil Blackwell.

Swales, J. M. (1981). *Aspects of Article Introductions. Aston ESP Research Report #1*. Birmingham, UK.

Swales, J. M. (1990). *Genre analysis : English in academic and research settings*. Cambridge University Press.

Swarts, J. (2012). New Modes of Help: Best Practices for Instructional Video. *Technical Communication*, *59*(3), 195–206.

Sweller. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. http://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296. http://doi.org/10.1023/A:1022193728205

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, *3*(3), 161–164. http://doi.org/10.1016/j.jarmac.2014.02.001

Taboada, M., & Mann, W. C. (2006). Applications of Rhetorical Structure Theory. *Discourse Studies*, *8*(4), 567–588. http://doi.org/10.1177/1461445606064836

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What Forty Years of Research Says About the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study. *Review of Educational Research*, *81*(1), 4–28. http://doi.org/10.3102/0034654310393361

ten Hove, P. E. (2014). *Characteristics of Instructional Videos for Conceptual Knowledge Development* (Ms.C. dissertation, University of Twente, The Netherlands).

ten Hove, P., & van der Meij, H. (2015). Like It or Not. What Characterizes YouTube's More Popular Instructional Videos? *Technical Communication*, *62*(1), 48–62.

Thomson, A., Bridgstock, R., & Willems, C. (2014). "Teachers flipping out" Beyond the online lecture: Maximizing the educational potential of video. *Journal of*

Learning Design, *7*(3), 67–78. http://doi.org/10.5204/jld.v7i3.209

Thornhill, S., Asensio, M., Young, C., Hodgson, V., Jackson, P., Strom, J., ... Zenios, M. (2002). *Video Streaming: A guide for educational development*. (S. Thornhill, M. Asensio, & C. Young, Eds.). The JISC Click and Go Video Project, ISD, UMIST.

Tian, Y., & Bourguet, M.-L. (2016). Lecturers' Hand Gestures as Clues to Detect Pedagogical Significance in Video Lectures. In *Proceedings of the European Conference on Cognitive Ergonomics - ECCE '16* (pp. 1–3). New York, New York, USA: ACM Press. http://doi.org/10.1145/2970930.2970933

Tisdell, C., & Loch, B. (2017). How useful are closed captions for learning mathematics via online video? *International Journal of Mathematical Education in Science and Technology*, *48*(2), 229–243. http://doi.org/10.1080/0020739X.2016.1238518

Toftness, A. R., Carpenter, S. K., Geller, J., Lauber, S., Johnson, M., & Armstrong, P. I. (2017). Instructor fluency leads to higher confidence in learning, but not better learning. *Metacognition and Learning*, pp. 1–14. http://doi.org/10.1007/s11409-017-9175-0

Tonndorf, K., Handschigl, C., Windscheid, J., Kosch, H., & Granitzer, M. (2015). The effect of non-linear structures on the usage of hypervideo for physical training. In *2015 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. http://doi.org/10.1109/ICME.2015.7177378

Töpper, J., Glaser, M., & Schwan, S. (2014). Extending social cue based principles of multimedia learning beyond their immediate effects. *Learning and Instruction*, *29*, 10–20. http://doi.org/10.1016/J.LEARNINSTRUC.2013.07.002

Truong, B. T., & Venkatesh, S. (2005). *Segmentation and Annotation of Video Data: A Survey*. Institute for Multi-Sensor Processing & Content Analysis, Curtin University of Technology, Western Australia. Perth (Australia).

Truong, B. T., & Venkatesh, S. (2007). Video abstraction: A systematic review and classification. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *3*(1), 3-es. http://doi.org/10.1145/1198302.1198305

Türkay, S. (2016). The effects of whiteboard animations on retention and subjective experiences when learning advanced physics topics. *Computers and Education*, *98*, 102–114. http://doi.org/10.1016/j.compedu.2016.03.004

Turro, C., Cañero, A., & Busquets, J. (2010). Video learning objects creation with Polimedia. *Proceedings - 2010 IEEE International Symposium on Multimedia, ISM 2010*, 371–376. http://doi.org/10.1109/ISM.2010.69

Uchidiuno, J., Hammer, J., Yarzebinski, E., Koedinger, K. R., & Ogan, A. (2017). Characterizing ELL Students' Behavior During MOOC Videos Using Content Type. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale - L@S '17* (pp. 185–188). New York, New York, USA: ACM Press. http://doi.org/10.1145/3051457.3053981

Udell, J. (2005). What Is Screencasting. Retrieved December 7, 2015, from http://archive.oreilly.com/pub/a/oreilly/digitalmedia/2005/11/16/what-is-screencasting.html

Valeiras-Jurado, J. (2017). *A multimodal approach to persuasion in conference presentations* (Doctoral thesis, Universitat Jaume I, Spain).

van den Broek, G., Takashima, A., Wiklund-Hörnqvist, C., Karlsson Wirebring, L., Segers, E., Verhoeven, L., & Nyberg, L. (2016, June 1). Neurocognitive mechanisms of the "testing effect": A review. *Trends in Neuroscience and Education*. Urban & Fischer. http://doi.org/10.1016/j.tine.2016.05.001

van der Meij, H., & van der Meij, J. (2013). Eight Guidelines for the Design of Instructional Videos for Software Training. *Technical Communication*, *60*(3), 205–228.

Van Der Schuur, W. A., Baumgartner, S. E., Sumter, S. R., & Valkenburg, P. M. (2015). The consequences of media multitasking for youth: A review. *Computers in Human Behavior*, *53*, 204–215. http://doi.org/10.1016/j.chb.2015.06.035

van der Zee, T., Admiraal, W., Paas, F., Saab, N., & Giesbers, B. (2017). Effects of subtitles, complexity, and language proficiency on learning from online education videos. *Journal of Media Psychology*, *29*(1), 18–30. http://doi.org/10.1027/1864-1105/a000208

van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance (meta-) cognitive processes in computer-based learning environments. In *International Handbook of Metacognition and Learning Technologies* (pp. 143–156). Springer, New York, NY. http://doi.org/10.1007/978-1-4419-5546-3_10

van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The mirror neuron system and observational learning: Implications for the effectiveness of dynamic visualizations. *Educational Psychology Review*, *21*(1), 21–30. http://doi.org/10.1007/s10648-008-9094-3

van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers and Education*, *72*, 323–327. http://doi.org/10.1016/j.compedu.2013.12.004

van Leeuwen, T. (2008). *Discourse and Practice: New Tools for Critical Discourse Analysis*. Oxford University Press. http://doi.org/10.1093/acprof

van Leeuwen, T., & Kress, G. (1995). Critical layout analysis. *Internationale Schulbuchforschung*, *17*(3), 25–43. http://doi.org/10.2307/43056999

van Marlen, T., van Wermeskerken, M., Jarodzka, H., & van Gog, T. (2016). Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computers in Human Behavior*, *65*, 448–459. http://doi.org/10.1016/j.chb.2016.08.041

van Wermeskerken, M., & van Gog, T. (2017). Seeing the instructor's face and gaze in demonstration video examples affects attention allocation but not learning. *Computers and Education*, *113*, 98–107. http://doi.org/10.1016/j.compedu.2017.05.013

Vegas, S., Juristo, N., & Basili, V. R. (2009). Maturing software engineering knowledge through classifications: A case study on unit testing techniques. *IEEE Transactions on Software Engineering*, *35*(4), 551–565. http://doi.org/10.1109/TSE.2009.13

Venema, S., & Lodge, J. M. (2013). Capturing dynamic presentation: Using technology to enhance the chalk and the talk. *Australasian Journal of Educational Technology*, *29*(1), 20–31. http://doi.org/10.1234/ajet.v29i1.62

Vorvilas, G., Karalis, T., & Ravanis, K. (2011). Designing Learning Objects: A genre-based approach. *Journal of Baltic Science Education*, *10*(2), 114–126.

Vorvilas, G., Vergidis, D., & Ravanis, K. (2011). Multimodal microgenres for designing Learning Objects. *Themes in Science & Technology Education*, *4*(2), 89–102.

Vural, Ö. F. (2013). The Impact of a Question-Embedded Video-based Learning Tool on E-learning. *Educational Sciences: Theory and Practice*, *13*(2), 1315–1323.

Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior*, *71*, 79–89. http://doi.org/10.1016/j.chb.2017.01.049

Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., & Collins, H. (2008). The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies*, *66*(2), 98–112. http://doi.org/10.1016/J.IJHCS.2007.09.003

Weibel, D., Wissmath, B., & Groner, R. (2008). How gender and age affect newscasters' credibility - An investigation in Switzerland. *Journal of Broadcasting and Electronic Media*, *52*(3), 466–484. http://doi.org/10.1080/08838150802205801

Wetzel, C. D., Radtke, P. H., & Stern, H. W. (1993). *Review of the Effectiveness of Video Media in Instruction*. San Diego, CA: Navy Personnel Research and Development Center.

Wetzel, C. D., Radtke, P. H., & Stern, H. W. (1994). *Instructional Effectiveness of Video Media*. Lawrence Erlbaum Associates.

Williams, L. E., & Bargh, J. A. (2008). Keeping one's distance: the influence of spatial distance cues on affect and evaluation. *Psychological Science*, *19*(3), 302–8. http://doi.org/10.1111/j.1467-9280.2008.02084.x

Williams, M. (2003). *A Taxonomy of Media Usage in Multimedia (T-MUM)* (Doctoral dissertation, Nova Southeastern University).

Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *Modern Language Journal*, *97*(1), 254–275. http://doi.org/10.1111/j.1540-4781.2013.01432.x

Winslett, G. (2014). What counts as educational video?: Working toward best practice alignment between video production approaches and outcomes. *Australasian Journal of Educational Technology*. http://doi.org/10.14742/ajet.v30i5.458

Wojcieszak, M. (2009). Three Dimensionality: Taxonomy of Iconic, Linguistic, and Audio Messages in Television News. *Television and New Media*, *10*(6), 459–481.

Wood, R. T. A., Griffiths, M. D., Chappell, D., & Davies, M. N. O. (2004). The Structural Characteristics of Video Games: A Psycho-Structural Analysis. *CyberPsychology & Behavior*, *7*(1), 1–10. http://doi.org/10.1089/109493104322820057

Woodruff, A. E., Jensen, M., Loeffler, W., & Avery, L. (2014). Advanced screencasting with embedded assessments in pathophysiology and therapeutics course modules. *American Journal of Pharmaceutical Education*, *78*(6). http://doi.org/10.5688/ajpe786128

Wu, Y. C., & Coulson, S. (2007). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin & Review*, *14*(1), 57–63.

Yaakob, S. (2013). *A Genre Analysis and Corpus Based Study of University Lecture Introductions* (Doctoral dissertation, University of Birmingham).

Yanamandram, V., & Noble, G. (2006). Student experiences and perceptions of team-teaching in a large undergraduate class. *Journal of University Teaching & Learning Practice*, *3*(1), 49–66.

Yee, N., Bailenson, J. N., & Rickertsen, K. (2007). A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1–10). New York, New York, USA: ACM Press. http://doi.org/10.1145/1240624.1240626

Young, C., & Asensio, M. (2002). Looking through three 'I's: the pedagogic use of streaming video. In *Proceedings of the 2002 Networked Learning Conference*. Sheffield, UK.

Young, L. (1994). University lectures - macro-structure and micro-features. In J. Flowerdew (Ed.), *Academic Listening: Research Perspectives* (pp. 159–176). Cambridge University Press.

Zettl, H. (1991). *Sight, Sound, Motion: Applied Media Aesthetics* (2nd ed.). Belmont, CA: Wadsworth Pub Co.

Zettl, H. (2016). *Sight, Sound, Motion: Applied Media Aesthetics* (8th ed.). Wadsworth Pub Co.

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker, J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management*, *43*(1), 15–27. http://doi.org/10.1016/j.im.2005.01.004

Zhang, J. R. (2012). Upper body gestures in lecture videos. *Proceedings of the 20th ACM International Conference on Multimedia - MM '12*, 1389. http://doi.org/10.1145/2393347.2396499

Zhang, J. R., Guo, K., Herwana, C., & Kender, J. R. (2010). Annotation and taxonomy of gestures in lecture videos. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*, 1–8. http://doi.org/10.1109/CVPRW.2010.5543253

Zhu, Y., Pei, L., & Shang, J. (2017). Improving video engagement by gamification: A proposed design of MOOC videos. In *Lecture Notes in Computer Science* (Vol. 10309 LNCS, pp. 433–444). Springer, Cham. http://doi.org/10.1007/978-3-319-59360-9_38

List of tables

List of figures

Anexos

Resumen en español

Glosario español/inglés

Anexo A. Resumen en español / Abridged version in Spanish

A.1 Antecedentes

El vídeo es un medio de comunicación que permite la integración de múltiples canales de información (sonido, imágenes y texto) en un único flujo que permite una experiencia cautivadora y envolvente. Una película proyectada en una gran pantalla en la oscuridad de un aula puede captar la atención de los estudiantes de una manera difícil de igualar por otro recurso de aprendizaje.

Las cualidades excepcionales de la imagen en movimiento han dado lugar a una historia centenaria de educación audiovisual que arranca en los orígenes del cine educativo en 1898, florece con la Segunda Guerra Mundial y la explosión de vídeos «instruccionales» para las tropas y llega a su cénit con la televisión educativa. Tras un cierto declive en la última década del siglo XX, en favor de la instrucción basada en computador, el vídeo recupera su protagonismo a principios del presente siglo XXI gracias a las tecnologías de *streaming*, que facilitan su distribución y reproducción a la carta.

Actualmente, las instituciones académicas producen diariamente miles de vídeos didácticos. Los cursos masivos en línea (MOOC) utilizan extensamente lecciones en vídeo como un recurso educativo esencial (Santos-Espino et al., 2016). Al mismo tiempo, millones de personas siguen cursos en línea y consumen tutoriales y lecciones en vídeo en plataformas digitales tales como YouTube o Udemy (Purcell, 2013).

El vídeo (o cine) se ha demostrado como un excelente medio educativo, por su capacidad para mostrar secuencias en movimiento realistas, para capturar y preservar eventos y lugares del mundo real que serían difíciles de presenciar directamente, o para cambiar el tamaño y la velocidad de fenómenos naturales que no pueden apreciarse a simple vista (Snelson & Perkins, 2009). El vídeo digital añade algunos atributos sobre el cine convencional, tales como una moderada capacidad interactiva y la posibilidad de integración con otros medios (C. Young & Asensio, 2002).

Los beneficios directos del formato de vídeo en la instrucción han sido estudiados desde una perspectiva científica. Los gráficos animados han demostrado su ventaja sobre las imágenes estáticas (véanse los metaanálisis de Berney & Bétrancourt, 2016; Höffler & Leutner, 2007). Hay una fuerte evidencia de que la presentación simultánea de ilustraciones y voz mejora el aprendizaje (Mayer, 2002, p. 105). Además, la utilización de vídeo despierta la presencia social de instructores y alumnos en la educación a distancia (Garrison et al., 1999). En definitiva, el vídeo es un recurso didáctico de primera línea en la actualidad.

A.2 Necesidad: cartografiar el conocimiento

Casi un siglo de investigación científica ha acumulado una abundante evidencia acerca de determinadas características de los vídeos que tienen efectos medibles en el aprendizaje. Una primera observación sobre este amplio conjunto de hallazgos es que carece de estructuración sistemática. Los principios de aprendizaje multimedia de Richard Mayer tienen en común el apoyo de sólidas teorías sobre el aprendizaje humano, pero tal y como están expuestos en las obras científicas, muestran entre ellos una débil conexión estructural. La última edición de la obra canónica *The Cambridge Book of Multimedia Learning* (Mayer, 2014d) enumera hasta 23 principios de aprendizaje multimedia que carecen de una taxonomía explícita o un esquema de clasificación. Otras revisiones actuales sobre el aprendizaje basado en vídeo, como la de Kay (2012) o la de Winslett (2014), tampoco proporcionan un esquema completo de clasificación de características.

La conclusión es que **se hace necesario construir un *mapa* estructurado de las características del vídeo didáctico**: una cartografía que organice todos los hallazgos científicos alrededor de categorías estructurales útiles y entendibles.

Por otro lado, es notorio que la investigación reciente sobre la eficacia del aprendizaje mediante vídeo ha estado focalizada en un conjunto restringido de características (si bien importantes), como los elementos de representación y sus relaciones espaciales y temporales. Sin embargo, otras características, tales como las relacionadas con la manipulación de la cámara y las características del discurso hablado, especialmente la retórica y la función interpersonal, reciben una atención relativamente escasa y podríamos decir que inconexa, a pesar de que en general se reconoce su potencial para el proceso de aprendizaje.

Este estado de las cosas no siempre fue así. Como muestra, tomemos la investigación en televisión educativa realizada en Gran Bretaña por John Baggaley y sus colaboradores (John Baggaley et al., 1980), quienes desarrollaron una extensa investigación experimental sobre cómo las múltiples técnicas de filmación (p.ej. tipos de plano, montaje y banda sonora) afectaban a las percepciones y actitudes del público. Se trata de una línea de investigación que a finales del siglo XX se desmanteló en favor del estudio de características relacionadas con los objetos multimedia, desde una perspectiva del aprendizaje asistido por computador.

En el momento actual, gracias al auge de la enseñanza en línea basada en vídeo en *streaming*, las propiedades cinematográficas de los vídeos están volviendo a ser estudiadas por una nueva generación de investigadores. Sin embargo, en muchos casos estos investigadores no son conscientes de los valiosos conocimientos del pasado de los que podrían sacar provecho. Es más, el desconocimiento sobre la técnica cinematográfica es notorio. Winslett (2014), tras una revisión de la literatura científica sobre el vídeo didáctico en la enseñanza superior, alerta de que «el vocabulario sobre producción cinematográfica y televisiva brilla por su ausencia».

A partir de estas observaciones, he llegado a una segunda conclusión: no solo hace falta un mapa de las características de los vídeos, sino que **en las últimas dos décadas el territorio explorado por los investigadores se ha estrechado**. Es necesario volver a poner de manifiesto toda la variedad y riqueza de las características de los vídeos que son relevantes en el aprendizaje. En definitiva, se trata de recuperar ese *territorio perdido* de la investigación.

A.3 Objetivos de la investigación

Según he expuesto en los antecedentes, mi investigación preliminar llegó a esta posición acerca del estado de la investigación sobre la efectividad del aprendizaje de las características de los vídeos didácticos:

- Se necesita un *mapa*: una cartografía que organice los resultados de la investigación sobre las características de los vídeos, utilizando categorías estructurales de nivel superior.
- Es necesario sacar a la luz toda la variedad de características relevantes en los videos didácticos, para recuperar áreas de investigación que han sido descuidadas en las últimas dos décadas.

Para superar estas necesidades, propongo construir un esquema de clasificación completo de las características de los vídeos didácticos. Esta propuesta se puede formular como un objetivo de investigación y una cuestión de investigación:

Objetivo principal:

Elaborar un esquema de clasificación para las características de los vídeos didácticos

Cuestión de investigación principal:

¿Cómo pueden clasificarse las características de los vídeos didácticos de una forma sistemática y útil?

Por **sistemática**, quiero decir que esta clasificación debe construirse de acuerdo con un método planificado y debe cumplir con criterios científicos sólidos.

Por **útil**, quiero decir que debe proporcionar información significativa y no trivial a la comunidad científica interesada. Los beneficiarios de este estudio son las personas que están involucradas en la investigación, diseño y producción de vídeos didácticos.

A.4 Alcance

A.4.1 Definiciones

El objeto de esta investigación son los **vídeos didácticos**, también llamados *vídeos instruccionales*, como calco del inglés *instructional videos*.

Por *vídeo* entenderemos «la presentación simultánea de un flujo continuo de información visual y auditiva» (Seel, 2012). Esta definición es independiente del soporte de visualización y distribución, así que no diferencia entre películas de cine y vídeo digital.

En el contexto de esta tesis, un *vídeo didáctico* se define como un vídeo elaborado con el objetivo de instruir. En este caso, la palabra *instruir* significa «cualquier acción intencionada con el objetivo de facilitar el aprendizaje» (Reigeluth & Carr-Chellman, 2009). A su vez, para el término *aprendizaje* se utilizará la definición cognitivista de Richard Mayer: «el cambio relativamente permanente en el conocimiento o la conducta de una persona causado por la experiencia» (Mayer, 1982, p. 1040).

A.4.2 Delimitación

Es preciso delimitar cuál es el alcance de la investigación, para asegurar un trabajo de investigación viable y un resultado útil.

- Se clasificarán las características que tengan influencia en el aprendizaje, bien sea basada en evidencias experimentales, bien sea porque se les reconoce un potencial en el aprendizaje, aunque actualmente no haya resultados concluyentes.
- En cuanto a las tipologías de vídeos estudiados: académicos, intrínsecamente didácticos, intencionadamente didácticos, asíncronos y en *streaming*.
- Formatos incluidos: documentales, clases magistrales (incluyendo clases en aula grabadas), tutoriales, demostraciones, ejercicios, entrevistas y testimonios.
- Formatos excluidos: diarios, vídeos promocionales, seminarios grabados, videoconferencias, vídeos tipo *mashup*.

A.5 Motivación

A.5.1 Beneficios de elaborar una clasificación

Los estudiosos de numerosos campos científicos coinciden en que los esquemas de clasificación mejoran la comprensión del dominio objeto de clasificación (Nickerson et al., 2013; Pope, 1994; Reigeluth, 1983). La clasificación es especialmente beneficiosa para entender dominios complejos (Nickerson et al., 2013). Entre otras ventajas, Bailey (1994) menciona la reducción de complejidad, la identificación de similitudes y diferencias entre casos, la fácil comparación de tipos y el fácil estudio de relaciones internas.

De forma concluyente, Vegas, Juristo y Basili (2009) resumieron los beneficios de construir una taxonomía para cualquier campo de conocimiento: a) proporciona un conjunto de construcciones unificadoras; b) ayuda a entender las interrelaciones; c) ayuda a identificar vacíos en el conocimiento.

Otras tecnologías relacionadas con los vídeos didácticos han tenido experiencias recientes de clasificación de conceptos: géneros de videojuegos (Gunn et al., 2009), propiedades estructurales de videojuegos (Wood et al., 2004), o bien diversas características de los objetos multimedia, tales como los elementos representacionales (Bernsen, 1994; Heller & Martin, 1995) o los elementos de interactividad (Aleem, 1998; Schwier, 1992).

A.5.2 ¿Existen ya esquemas de clasificación?

Probablemente el trabajo que mejor representa el objetivo de proporcionar un «mapa» de las características que influyen en la efectividad del aprendizaje de los videos didácticos es el informe que Wetzel, Radtke y Stern hicieron para la Marina de los Estados Unidos: *Review of the Effectiveness of Video Media in Instruction* (Wetzel et al., 1993), publicado posteriormente como libro (Wetzel et al., 1994). Este trabajo es un compendio de los hallazgos hasta esa fecha sobre la eficacia de todo tipo de características de vídeo, abarcando tanto las mejores prácticas profesionales como la investigación científica sobre el aprendizaje con el cine y el vídeo. El problema de esta obra es su antigüedad, pues se deja atrás muchos hallazgos posteriores sobre las teorías cognitivas del aprendizaje multimedia.

Resulta difícil encontrar un trabajo posterior al de Wetzel et al con una cobertura amplia de las características de los vídeos. Podemos encontrar esquemas de clasificación sobre dominios muy concretos, como el uso de gestos en lecciones en vídeo (J. R. Zhang et al., 2010), anotaciones sobre vídeos (Aubert et al., 2014), patrones de diseño de interacción en entornos de aprendizaje basado en vídeo (Seidel, 2015) y formatos de presentación de vídeos en cursos en línea (Hansch et al., 2015). Pero he sido incapaz de encontrar un trabajo que dé una panorámica amplia sobre todos los ámbitos estructurales del vídeo didáctico.

A.5.3 ¿Hay características que se han dejado de lado?

Hay que tener en cuenta que a finales de los años setenta del siglo XX hubo un cambio drástico en los intereses del sector de la tecnología educativa, debido a la entrada de los ordenadores. Hasta entonces, la investigación en Televisión Educativa (ETV) era floreciente, pero decayó rápidamente a favor de la investigación en Instrucción Asistida por Ordenador (CAI) y más tarde, la investigación en Aprendizaje Multimedia (MML) (Barford & Weston, 1997; Salomon, 1994, p. xvii). Estas nuevas comunidades científicas barrieron a la ETV sin que hubiera transferencia de conocimiento entre ellas. Este suceso provocó varias pérdidas de conocimiento científico. El primer campo olvidado fueron las tecnologías cinematográficas aplicadas a la educación, aunque actualmente están recobrando interés gracias al vídeo en *streaming*. Otro campo perdido fue el de la semiótica aplicada al vídeo educativo desde una perspectiva estructuralista (De Vaney, 1991; Salomon, 1979a). La semiótica y la retórica permiten analizar la estructura del discurso en los vídeos, sus sistemas de símbolos y cómo se articulan para construir un mensaje eficaz. Otras modalidades de comunicación han seguido recibiendo atención desde estas disciplinas, no así el audiovisual educativo.

Como conclusión, el marco actual de investigación sobre el vídeo en la enseñanza adolece de integrar las características fílmicas del vídeo (Winslett, 2014) y la estructura lingüística y semiótica del discurso. Es importante recuperar el terreno perdido y ofrecer un marco amplio de análisis de los vídeos didácticos que abarque no solo los elementos «multimedia» convencionales, sino también esos otros aspectos.

A.6 Método de la investigación

El plan de actuación para alcanzar los objetivos de esta investigación ha sido el siguiente:

1. Identificar las principales disciplinas científicas que modelarán esta investigación.
2. Hallar trabajos científicos claves dentro de esas disciplinas, que proporcionen un primer conjunto de características relevantes de los vídeos didácticos.
3. Realizar un estudio de campo sobre cursos en línea, para obtener evidencias de uso real de características de vídeos didácticos.
4. Realizar un proceso de clasificación sobre las características identificadas en los pasos anteriores, lo cual resultará en un esquema de clasificación basado en dominios estructurales.
5. Refinar el esquema de clasificación mediante una revisión profunda de la literatura científica. Este paso dará también una colección de taxonomías para cada dominio de características.

A.7 Una caracterización de los vídeos didácticos

Como primera etapa del trabajo de tesis, he tratado de situar a los vídeos didácticos en un marco de análisis apropiado y amplio. He realizado una exploración de las fuentes científicas y he encontrado varias aproximaciones epistemológicas hacia la investigación en vídeos educativos. En particular, las ciencias cognitivas y el análisis del discurso son los dos campos de conocimiento que pueden dar un soporte teórico óptimo para mi investigación.

Como síntesis de esta primera investigación, puede afirmarse que los videos didácticos tienen estas tres naturalezas:

- Los vídeos didácticos son **filmes expositivos**.
- Los vídeos didácticos son **objetos de aprendizaje multimedia**.
- Los vídeos didácticos son **textos multimodales**.

Estas tres perspectivas hacia la naturaleza de los vídeos didácticos son las que modelarán el esquema de clasificación propuesto en esta tesis.

A.7.1 Los vídeos didácticos como filmes expositivos

Los vídeos didácticos constituyen una clase particular de *filmes expositivos* (*expository films*, en inglés). Un filme expositivo difiere de un *filme narrativo* en que no tiene una historia o una trama. Los filmes expositivos normalmente representan hechos no ficticios y describen la «estructura y procesos involucrados en un sistema o evento» (Brewer, 1980). Dentro de la familia de géneros expositivos, los vídeos didácticos modernos suelen caracterizarse por una **duración corta** y una **baja complejidad visual**. La inmensa mayoría de los vídeos didácticos utilizan la «voz formal» de Plantinga (1997). Dentro de los seis *modos* del marco de Nichols (2001), muchos vídeos didácticos encajan con el «modo expositivo».

A.7.2 Los vídeos didácticos como textos multimodales

El *discurso* se define como «una producción continua de lenguaje (típicamete oral) más larga que una frase y que normalmente conforma una unidad coherente tal como una conferencia, una discusión, un chiste o una narración» (Crystal, 1992, p. 25). El objeto de estudio del Análisis del Discurso son los **textos**. Un «texto» es cualquier emisión de lenguaje verbal, en cualquier modalidad: escrita, hablada o signada. El Análisis del Discurso estudia el lenguaje en su contexto social y cultural, y descubre patrones de uso que van más allá de la gramática convencional.

El **Análisis del Discurso Multimodal** (ADM) amplía el estudio de los textos a cualquier combinación de modos, no solamente verbales, sino también icónicas, musicales, táctiles o de cualquier otro tipo. El ADM ha estudiado toda clase de producciones multimodales, desde fotografías hasta películas (Baldry & Thibault, 2006; Bateman, 2008; Kress & van Leeuwen, 2006; O'Halloran, 2009), o acciones

que implican lenguaje no verbal tales como las conferencias o lecciones en clase (Barrett & Liu, 2016; Morell, 2015).

La **Lingüística Sistémico-Funcional** (LSF) (Michael A. K. Halliday & Hasan, 1985; Michael A. K. Halliday & Matthiessen, 2014) proporciona un marco teórico básico para el análisis multimodal. Esta teoría enfoca el análisis del lenguaje y la gramática desde un punto de vista funcional, con énfasis en las interacciones entre los hablantes, sus objetivos funcionales y sus roles en la comunicación. La LSF ha sido aplicada para definir relaciones complejas en objetos multimodales, tales como películas de cine o carteles publicitarios (O'Halloran, 2009; van Leeuwen, 2008).

Los vídeos didácticos han sido apenas estudiados de forma específica desde la perspectiva del ADM. Sin embargo, se han realizado análisis sobre vídeos expositivos en general (ej. documentales) y también sobre el discurso académico (ej. conferencias y clases magistrales). De estas investigaciones, por analogía, puede concluirse que los vídeos didácticos tienen un discurso caracterizado por la **verdad y confianza** (*truth and trust*), con una tendencia a la **formalidad**, el **monólogo**, a la argumentación científica y la no ficción (Crawford Camiciottoli & Bonsignori, 2015; Crawford Camiciottoli & Fortanet-Gómez, 2015).

A.7.3 Los vídeos didácticos como objetos multimedia

Los vídeos didácticos pueden considerarse un tipo de «objetos de aprendizaje» (Churchill, 2007) caracterizados por un nivel de interactividad moderado y un grado relativamente alto de multimodalidad, comparados con otros tipos de objetos.

La hipótesis del aprendizaje multimedia establece que «las personas aprenden más profundamente mediante palabras e imágenes juntas que solamente con palabras» (Mayer, 2014d, p. 1). Las teorías modernas del aprendizaje multimedia se basan a su vez en dos hipótesis sobre la arquitectura del sistema de aprendizaje humano. Por un lado, la **Teoría de la Codificación Dual** (*Dual Coding Theory*), según la cual existen dos canales separados para el procesamiento de la información visual y verbal (Paivio, 1990). Y por otro lado, la **Teoría de la Carga Cognitiva** (*Cognitive Load Theory*), que sugiere que los canales de procesamiento tienen una capacidad limitada de memoria de trabajo para procesar el flujo entrante de información (Sweller, 1988).

Sobre la base de las dos teorías mencionadas se ha ido desarrollando la llamada **Teoría Cognitiva del Aprendizaje Multimedia** (*CTML, Cognitive Theory of Multimedia Learning*), cuyo principal valedor es Richard Mayer. La CTML ha evidenciado un conjunto de *principios didácticos*, tales como el principio multimedia, el principio de coherencia y el principio de segmentación, por mencionar algunos. Cada principio establece una hipótesis sobre el diseño de la instrucción que influye positivamente sobre los procesos de aprendizaje. Cada principio está soportado por evidencias experimentales. La Tabla 2-2 del Capítulo 2 de esta tesis muestra los principios de aprendizaje multimedia más relevantes, tal y como están recogidos en la obra *The Cambridge Handbook of Multimedia Learning* (Mayer, 2014d).

A.8 Estudio de campo sobre cursos online (MOOC)

Para complementar la revisión de la literatura, decidí realizar un estudio de campo para obtener evidencias de primera mano acerca de la utilización actual de las características de los vídeos didácticos. Este estudio de campo se ha centrado en los vídeos que se emplean en las plataformas MOOC (Massive Online Open Courses).

Para este estudio se seleccionaron cinco plataformas MOOC generalistas de alcance internacional: Coursera, edX (Estados Unidos), MiriadaX (España), FutureLearn (Reino Unido) y FUN (Francia). El estudio se desarrolló en dos fases. En primer lugar, se realizó un estudio cualitativo para identificar los estilos de vídeo más utilizados y con ello crear un esquema de clasificación. En segundo lugar, se utilizó una muestra de 115 cursos en las plataformas MOOC seleccionadas para hacer un recuento de las características de vídeo y la frecuencia de uso de cada estilo. Se realizaron varias pruebas estadísticas (fundamentalmente, estadística no paramétrica) para descubrir asociaciones significativas entre las características del curso y el tipo de vídeo utilizado.

Como resultado, se han identificado siete estilos de presentación en vídeo como los más frecuentes en los cursos de MOOC: **«busto parlante» (*Talking Head*)**, **«clase en vivo»**, **«entrevista»**, **«presentación con diapositivas»**, **«screencast»**, **«pizarra virtual»** y **«documental»**. Esta clasificación es una simplificación de las tipologías identificadas en el informe de Hansch et al sobre formatos de vídeo en cursos MOOC (Hansch et al., 2015). Los siete estilos identificados describen la totalidad del stock de vídeos del 85% de los cursos muestreados. Un curso típico utiliza dos estilos diferentes.

El estudio de campo pone de manifiesto una escasa diversidad en la utilización de recursos de representación. Muy pocos cursos utilizan actores o modelos en lugar de instructores (12/115), dibujos animados (8), humor explícito (4), tests empotrados en el vídeo (8) y vídeos sin voz (4).

El estudio revela dos patrones para mostrar los contenidos educativos en los vídeos de MOOC: por un lado se encuentran los estilos **centrados en el orador** (una persona visible habla de los contenidos) y por otro, los estilos **centrados en el tablero** (una gran superficie rectangular muestra los contenidos). Uno de los hallazgos del estudio es que la adopción de cada uno de estos patrones está significativamente relacionada con el tema del curso: los cursos de Artes y Humanidades muestran una preferencia por los estilos centrados en el actor, mientras que los cursos de Ingeniería y «ciencias duras» prefieren los videos centrados en el tablero. Los cursos de Ciencias Sociales y de la Salud se encuentran en una posición neutral. Así pues, se observa una cierta distinción entre «ciencias y letras» en las preferencias de formatos de los vídeos.

A.9 El proceso de clasificación

La taxonomía de esta tesis ha sido construida a través de un método iterativo que comienza con la extracción de un inventario crudo de características tomadas de un conjunto de trabajos seleccionados, que he considerado fuentes principales para mi estudio: el informe de Wetzel, Radtke y Stern (1993) sobre cine y vídeo educativos; los Principios de Aprendizaje Multimedia de Mayer (Mayer, 2014d); varias guías para el diseño de videotutoriales basadas en evidencia experimental (Kay, 2014; Koumi, 2006, 2015; Loch & Mcloughlin, 2011; Morain & Swarts, 2012; Swarts, 2012; van der Meij & van der Meij, 2013) y varios métodos y herramientas de anotación del discurso multimodal en cine (Baldry & Thibault, 2006; Lim Fei et al., 2015; O'Halloran, 2009).

Este conjunto seminal de fuentes ha dado lugar a un inventario de 55 características relevantes en los vídeos didácticos. A partir de este inventario, se ha realizado un proceso de agrupación (*clustering*) y categorización, dando como resultado un primer borrador del esquema de clasificación. A este borrador se le ha incorporado un análisis sobre las tipologías de vídeos didácticos, que ha incorporado el concepto de género de vídeo al esquema original. Finalmente se han propuesto nueve agrupamientos clasificatorios: propiedades del medio, entidades básicas de representación, dispositivos de interacción, arreglos espaciales, arreglos temporales, propiedades de complejidad, estructuras del discurso hablado, metas comunicativas y géneros de vídeos.

El resultado de las fases anteriores ha sido integrado en los marcos teóricos de la Lingüística Funcional Sistémica y del Análisis del Discurso Multimodal. En particular, el modelo GeM de John Bateman (Bateman, 2008) ha sido elegido como la base para adaptar el esquema de clasificación. En primer lugar, GeM es un modelo multimodal fácilmente adaptable a muchos soportes de comunicación, incluyendo el audiovisual. En segundo lugar, las categorías de GeM encajan casi a la perfección con las resultantes del esquema inicial de esta investigación. Al modelo GeM se le ha incorporado una estructura por capas jerárquicas inspirada en los trabajos de Vorvilas y colaboradores (Vorvilas, Karalis, et al., 2011; Vorvilas, Vergidis, et al., 2011) y que traslada al modelo el aprovechable concepto de *estratificación* de la Lingüística Sistémico-Funcional.

Como fase final del proceso, se ha realizado una serie iterativa de búsquedas de literatura científica específicas para cada una de las categorías de clasificación. El objetivo de este proceso de búsqueda es triple. En primer lugar, se pretende validar el esquema de clasificación mediante evidencias generales en la literatura. En segundo lugar, se quieren descubrir características que no estaban presentes en las primeras fuentes seleccionadas. Finalmente, se elaboran taxonomías de más bajo nivel para cada una de las categorías de clasificación. En esta fase se han seleccionado más de 200 referencias relevantes. El resultado detallado puede leerse en el Capítulo 6 de la versión completa en inglés.

A.10 Resultado final: el esquema de clasificación

A.10.1 Especificación del metamodelo

El marco de clasificación propuesto utiliza varios conceptos: *entidad*, *propiedad*, *clase*, *dominio* y *capa*. Estos conceptos y sus relaciones conforman un metamodelo para este esquema de clasificación.

- Una **característica** es cualquiera de los objetos para los que se ha definido esta taxonomía. Una característica puede ser una *entidad* o una *propiedad*.
- Una **entidad** es cualquier objeto identificable dentro del vídeo. Una entidad puede ser una combinación de muchas entidades de nivel inferior. Algunos ejemplos de entidades son: actor, narración de audio, escena, fase retórica.
- Una **propiedad** es un valor que se puede medir de una entidad, de un conjunto de entidades o del objeto de vídeo completo. Ejemplos de propiedades son: duración, velocidad, número de palabras, color, tamaño.
- Por conveniencia, las características (entidades y propiedades) pueden agruparse en **clases** que comparten rasgos comunes. Las clases proporcionan generalización conceptual y economía descriptiva. Por ejemplo, el género, la edad y el grupo étnico de los actores son propiedades que pueden agruparse en una clase de propiedades de «apariencia social». Las clases se pueden jerarquizar utilizando relaciones superclase-subclase.
- Un **dominio** es un espacio semántico que ofrece una perspectiva de análisis sobre el vídeo. Dos ejemplos de dominios son el dominio de presentación y el dominio espaciotemporal. En general, cada característica de este marco se asignará a un solo dominio.
- Los dominios se agrupan en una jerarquía de **capas**. Las capas están apiladas según su posición en un espectro que va del medio físico hasta lo más abstracto.

A.10.2 El esquema de clasificación: capas y dominios

El esquema de clasificación propuesto define ocho dominios estructurales: medio físico, presentación, interacción, espaciotemporal, habla, apariencia social, metas/estrategias y el dominio genérico. La Tabla 1 muestra los ocho dominios organizados en capas jerárquicas.

Tabla 1. Esquema de clasificación propuesto

capa	dominio	descripción
Capa 4: genérica	genérico	géneros: patrones de utilización de las entidades básicas y compuestas, reconocibles por la comunidad de autores de vídeos y su audiencia
Capa 3: estratégica	metas y estrategias	metas y estrategias comunicativas y educativas involucradas en el diseño del producto
Capa 2: composicional	apariencia social	rasgos sociales y culturales que influyen en la respuesta del usuario a los contenidos
	habla	articulación del discurso en su modalidad textual (escrita o hablada)
	espaciotemporal	articulación del discurso en el espacio y el tiempo, mediante métodos de producción cinematográfica
Capa 1: entidades básicas	presentación	entidades que contienen información
	interacción	entidades que posibilitan la interacción con el usuario
Capa 0: física	medio	el soporte físico que sirve de sustrato a los niveles superiores

A.10.3 Capas: de la mente al fotograma

Los dominios estructurales pueden organizarse mediante una jerarquía de capas que va desde el nivel físico más bajo (el fotograma de vídeo y su entorno) hasta niveles cada vez más abstractos. Las capas intermedias se ocupan de los elementos constructivos estructurales. Las capas superiores de esta jerarquía están relacionadas con el diseño general del producto.

Capa 0: física. Es el sustrato físico que sirve para ubicar el contenido. Al mismo tiempo restringe qué modos de expresión pueden darse al contenido.

Capa 1: entidades básicas. Entidades reconocibles por el usuario que o bien portan contenido, o bien proporcionan mecanismos básicos para la interacción con el vídeo. Los autores de vídeos colocan estas entidades básicas en el vídeo como parte del proceso de producción. Las entidades más prominentes en los videos instructivos son los *actores* (por ejemplo, narradores y modelos) y los *tableros* (por ejemplo, cajas de texto, diapositivas y pantallas de ordenador).

Capa 2: composicional. Las entidades básicas del la Capa 1 se articulan en diferentes dominios para elaborar el discurso didáctico. Las entidades están dispuestas en el espacio y en el tiempo, haciendo uso de las técnicas habituales de producción cinematográfica. A su vez el discurso se articula en estructuras retóricas, utilizando funciones lingüísticas. Las propiedades de apariencia social se componen para desencadenar respuestas sociales en los espectadores. Una adecuada coordinación de las entidades básicas en el espacio, el tiempo, el habla y los aspectos sociales ayuda a reforzar el significado del discurso, así como su eficiencia de aprendizaje. Esta capa contiene entidades de nivel superior, tales como escenas, segmentos de vídeo y fases retóricas del discurso.

Capa 3: estratégica. Esta capa contiene propiedades acerca de cómo los autores plantean el diseño del vídeo. Los creadores se proponen con el vídeo distintos objetivos y metas, la mayoría de las veces relacionados con los resultados del aprendizaje. También aplican principios, estrategias y diseños pedagógicos.

Capa 4: genérica. Las entidades básicas y las composicionales suelen organizarse en patrones recurrentes que son reconocibles por una comunidad de práctica (*community of practice*). Estos patrones se denominan *géneros*. En los vídeos didácticos pueden identificarse ciertos géneros, tales como los *screencasts*, los *talking heads*, etc.

A.10.4 Los dominios de la clasificación

El dominio del Medio. El dominio más básico viene dado por las características que el medio del vídeo ofrece a los creadores: un cuadro (*frame*) para ubicar el contenido visual, una banda sonora y cierta capacidad de interacción y anotación.

El dominio de Presentación. Este dominio ofrece elementos estructurales básicos que portan contenido significativo: diapositivas, capturas de pantalla, narración en audio, uno o más *actores* visibles, subtítulos, sonidos y música, entre otras entidades.

El dominio de la Interacción. La interactividad en un sistema multimedia se refiere a su capacidad de recibir retroalimentación externa del usuario para alterar el flujo de información mostrada. Actualmente, muchos vídeos didácticos presentan cierto grado de interactividad, como mínimo a través del control de reproducción del vídeo. Además, muchas veces se añaden elementos más sofisticados de interacción, tales como pausas forzadas en puntos seleccionados y preguntas empotradas en el vídeo.

El dominio Espaciotemporal. Este dominio describe cómo las piezas básicas del contenido son organizadas en el espacio y el tiempo, casi siempre utilizando las técnicas convencionales de la cinematografía, tales como el montaje y los ajustes de cámara (Burch, 1970). La organización espacial y temporal de los elementos influye en la respuesta del espectador, según demuestran los principios de aprendizaje multimedia, tales como el principio de segmentación, los principios de contigüidad espacial y temporal, y el principio de redundancia (Mayer, 2014d).

El dominio del Habla. Este dominio agrupa las características de alto nivel que describen la articulación del lenguaje hablado y escrito a lo largo de la duración del

vídeo. La retórica es un factor fundamental en el discurso didáctico, ya que el objetivo de la retórica es justamente persuadir al oyente. El mensaje debe ser articulado de manera que sea al mismo tiempo comprensible y atractivo.

El dominio de la Apariencia Social. La caracterización de la apariencia social en un vídeo didáctico va más allá de una mera colección de atributos del narrador o el instructor visible en el vídeo. La apariencia social es un constructo complejo mediante el cual el autor del vídeo proyecta una imagen social que puede disparar una respuesta en el espectador, que a su vez puede afectar a procesos como la motivación (Baylor, 2011), la credibilidad (John Baggaley et al., 1980) y la implicación del espectador en el discurso del vídeo (Mayer et al., 2003).

El dominio Estratégico. Los creadores de vídeos quieren cumplir metas y propósitos en los futuros espectadores, normalmente relacionados con resultados de aprendizaje. Este dominio incluye todas aquellas propiedades de los vídeos que de alguna manera reflejan esas metas, propósitos y pedagogías.

El dominio Genérico. En Semiótica y Lingüística, los *géneros* son eventos de comunicación que persiguen alguna meta comunicativa y que muestran características estructurales reconocibles y utilizables dentro de una comunidad social (James Robert Martin, 1994; Swales, 1990). Cada género tiene sus características distintivas, que pueden ser lingüísticas, paralingüísticas y contextuales. El uso de géneros reconocibles en los vídeos didácticos tiene la virtud de crear una expectativa en la audiencia sobre qué se va a mostrar en el vídeo, qué pasos va a seguir la exposición y dónde se ubicarán los contenidos relevantes. Estas expectativas, si se cumplen, pueden incrementar la eficacia comunicativa del vídeo (Chandler, 1997).

A.10.5 Taxonomías específicas de los dominios

Para cada uno de los dominios de más bajo nivel, se han elaborado unas taxonomías específicas. Las siguientes tablas muestran esas taxonomías. La presente tesis no ha elaborado taxonomías para los dominios Estratégico y Genérico.

Taxonomía para el dominio del Medio

Tabla 2. Taxonomía del dominio de Medio

clase	Definición	propiedades
fotograma	cuadro en el que se muestra el vídeo	tamaño, resolución
banda sonora	representación sonora del contenido	calidad de audio
superposición	elemento superpuesto al fotograma	---
control de usuario	elemento accionable por el usuario	---

Taxonomía para el dominio de Presentación

El dominio de Presentación es la «caja de herramientas» básica que utiliza el autor del vídeo para construir el contenido. El trabajo de campo en los cursos en línea y la posterior revisión de la literatura muestran que las entidades del dominio de Presentación se pueden agrupar en torno a la pareja de conceptos **actor y tablero**. Un «actor» es un agente con cualidades humanas, real o virtual, visible o no, que suministra contenido de forma activa. Un «tablero» (*board*) es una superficie en la cual se muestran contenidos didácticos. Ejemplos habituales de tableros son las diapositivas tipo PowerPoint, grabaciones de una sesión de ordenador o pizarras físicas.

Las entidades de la taxonomía se agrupan en torno a los conceptos de actor y tablero de esta forma: hay entidades del actor, entidades del tablero y otras entidades que sirven de dispositivos de interacción entre un actor y un tablero. Las restantes entidades identificadas en esta investigación actúan de forma periférica al actor y el tablero, bien como entidades auxiliares (ej. subtítulos), bien como entidades no didácticas (ej. decoraciones).

Tabla 3. Taxonomía para el dominio de Presentación

clase	entidades
entidades del tablero	texto didáctico, diagrama, ilustración, mapa, gráfico, sonido, animación
entidades del actor	voz, rostro, gestos
interacción actor-tablero	trazo o dibujo a mano alzada, puntero virtual, gesto deíctico
entidades auxiliares didácticas	señal acústica, subtítulo, transcripción
entidades no didácticas	texto no didáctico, decoración visual, música de fondo

Taxonomía para el dominio de Interacción

El dominio de interacción incluye una variedad de dispositivos de interacción más o menos avanzados entre el usuario y el propio vídeo. Se han identificado cuatro clases de categorías funcionales (Tabla 4): las dos primeras (reproducción y navegación) permiten al usuario *controlar* cómo se visualiza el vídeo, mientras que las otras dos (diálogo usuario-sistema y comentario de usuario) sirven de vehículos para que el usuario aporte una retroalimentación de información al vídeo y este altere su comportamiento.

Taxonomía para el dominio Espaciotemporal

La Tabla 5 muestra la taxonomía de este dominio. La mayoría de las características pertenecen al ámbito de las técnicas cinematográficas. El soporte experimental sobre el efecto en el aprendizaje de estas características es variable: algunas características

como la duración del segmento y la contigüidad espacial tienen un efecto bastante alto, mientras que otras como el ritmo de planos o la continuidad no han sido exploradas en profundidad.

Tabla 4. Taxonomía para el dominio de Interacción

clase	subclase	entidades
control	reproducción	panel de reproducción básico control de velocidad de reproducción control de presentación
	navegación	línea de tiempo (simple o mejorada) tabla de contenidos resumen visual hiperenlace incrustado
retroalimentación	diálogo usuario-sistema	test interpolado pausa forzada por el sistema
	comentario de usuario	anotación generada por el usuario

Tabla 5. Taxonomía para el dominio Espaciotemporal

clase	tipo	características (entidades y propiedades)
composición espacial	propiedad	áreas del fotograma con relevancia semiótica
	propiedad	ajustes de cámara: ángulo, plano, perspectiva, zoom
segmentación temporal	entidad	jerarquía de segmentación fílmica: plano, diapositiva, escena, secuencia, clip, hipervídeo
	entidad	transiciones de segmentos: pausas y pistas temporales
	propiedad	duración del segmento de vídeo
linealidad	propiedad	lineal vs. no lineal
	entidad	grafo de navegación (solo en vídeos no lineales)
complejidad informacional	propiedad	velocidad de presentación: palabras por minuto, elementos por minuto
	propiedad	complejidad fílmica: ritmo de planos (planos por minuto), continuidad
	propiedad	ritmo entre eventos informativos (contigüidad temporal, redundancia)
	propiedad	contigüidad espacial de los elementos

Tabla 6. Taxonomía para las entidades retóricas del dominio del Habla

clase (meta retórica)	**entidades (fases retóricas)**
organizar el discurso	plano de apertura o cierre del vídeo, visión general de los contenidos, explicación de los prerrequisitos y el contexto, relacionar con otros contenidos, anuncio de la siguiente sección, pausa retórica, resumen del contenido
comunicar el contenido	teoría / contenido, demostración / ejecución de tarea, ejemplo, reformulación, evaluación: indicar actitud, evaluación: indicar cumplimiento
interpelar al espectador	solicitar recordar o repetir el contenido expuesto, solicitar ejecución de tarea, solicitar reflexión y transferencia
atraer al espectador	enganchar (capturar la atención), justificar/motivar contenido, generar confianza/autoridad en el locutor, crear y cumplir las expectativas del espectador

Tabla 7. Taxonomía para las propiedades del dominio del Habla

clase (metafunción)	**subclase**	**propiedades**
textual (modo)	hablado/escrito	texto hablado vs. escrito
	acción/reflexión	habla espontánea vs. ensayada
	interactividad	monólogo vs. diálogo preguntas e interpelaciones
interpersonal (tenor)	función del habla	enunciación, cuestión, oferta, orden
	distancia social	estilo conversacional vs. formal amabilidad humor
	personalización	personalización (hablar en 2ª persona)
	posición (standing)	afirmación de autoridad
	evaluación (appraisal)	actitud, involucración, graduación
	postura (stance)	modalidad: epistémica vs. deóntica narrador inseguro vs. confiado

Taxonomías para el dominio del Habla

Para este dominio se proponen dos taxonomías. La primera (Tabla 6) sirve para describir las *entidades* que componen el discurso del vídeo. Está inspirada en el

catálogo de estructuras retóricas de Koumi (2006, 2015) y enriquecido por el estudio de campo realizado en esta tesis. La segunda taxonomía (Tabla 7) sirve para describir las *propiedades* del discurso que, de acuerdo con la revisión de la literatura, parecen tener influencia en la eficacia del mensaje. Esta segunda taxonomía utiliza los conceptos de metafunción y función lingüística de la teoría de la Lingüística Sistémico-Funcional. Estos conceptos encajan a la perfección como agrupaciones conceptuales para describir los hallazgos de investigación sobre el discurso en los vídeos didácticos.

Taxonomía para el dominio de la Apariencia Social

Todas las características de este dominio son propiedades. La mayoría de ellas tienen que ver con el actor del vídeo, excepto una clase de propiedades que tienen que ver con la puesta en escena general (*mise en scène*).

Tabla 8. Taxonomía para el dominio de la Apariencia Social

clase	propiedades
realismo	voz: robótica vs. humana imagen: generada por ordenador, dibujo animado, natural
fluidez	acento nativo vs. extranjero, velocidad del habla, fluidez del habla, mirada directa, sincronización entre gestos y habla
distancia social	tamaño del plano lenguaje: personalización, formalidad, amabilidad
grupo social	género, edad, grupo étnico, afiliación social, dialecto
puesta en escena	ambiente sociocultural, ambientación espaciotemporal, atmósfera

7.4.5 Géneros de vídeos didácticos

Varias clases de géneros son reconocibles en los vídeos didácticos. Primero, tenemos distintos enfoques de comunicación, tales como la clase magistral, la demostración y la entrevista, que pueden considerarse géneros genuinos. En segundo lugar, nos encontramos variaciones en los formatos de presentación tales como el *screencast*, la presentación de diapositivas y el «busto parlante» (*talking head*).

Tras la revisión de la literatura desarrollada en este trabajo, puede establecerse que los géneros de vídeos didácticos pueden caracterizarse según cinco tipos de rasgos:

a) Metas comunicativas (ej. ver, atraer, hacer, decir).
b) Tipo de acción grabada (ej. clase, conversación, demostración, simulación).
c) Formato de comunicación (ej. «fly on the wall», «busto parlante»).
d) Organización de la pantalla (ej. diapositivas tipo PowerPoint, imagen empotrada o *picture in picture*...).
e) Puesta en escena (ej. escenario, fondo de pantalla).

A.11 Conclusiones

A.11.1 Resultados de la investigación

El objetivo principal propuesto para esta tesis ha sido «elaborar un esquema de clasificación para las características de los vídeos didácticos», de forma sistemática y con un resultado útil. Esta tesis ha descrito el proceso de elaboración de esta clasificación, que comenzó con una extensa revisión de la literatura científica que dio lugar a un proceso de clasificación ascendente, cuyo resultado final es el esquema de clasificación expuesto en el apartado A.10 de este resumen y en el Capítulo 5 de la versión completa en inglés. El esquema de clasificación está basado en las teorías y herramientas del Análisis del Discurso Multimodal, en concreto el marco de trabajo GeM propuesto por John Bateman (Bateman, 2008).

Otro resultado colateral de la investigación es el estudio sobre los estilos de presentación y características empleados actualmente en los vídeos didácticos de cursos en línea MOOC.

Por último, un resultado valioso de la investigación es la propia revisión de la literatura sobre características de los vídeos didácticos. En sí misma ofrece una visión muy amplia y actualizada sobre el conocimiento sobre los componentes estructurales de los vídeos y su efecto en el aprendizaje.

A.11.2 Hallazgos destacables

El estudio sobre los cursos MOOC pone de manifiesto que hay una correlación entre el tipo de materia del curso y las preferencias en el formato de presentación del vídeo. Esta correlación no parece explicarse del todo por el contenido y todo apunta a que influyen factores culturales. Esto es algo que necesita más investigación.

El estudio sobre los MOOC y la revisión de la literatura revelan dos componentes cruciales en la estructura de los vídeos didácticos: el **actor** y el **tablero**. Sus atributos y su interacción condicionan bastantes propiedades generales del producto.

Una observación digna de mención es cómo los diseñadores de vídeos superan la falta de interactividad de este medio en comparación con otros formatos (incluida la instrucción presencial): test empotrados dentro del vídeo, herramientas de navegación y la inclusión de funciones interpersonales del lenguaje, tales como las interpelaciones a la audiencia.

Un resultado que considero muy interesante es que la Lingüística Sistémico-Funcional (LSF) de Halliday se muestra como un instrumento útil para caracterizar los principios de aprendizaje multimedia relacionados con el lenguaje, como ocurre con el Principio de Voz (*voice principle*) o el Principio de Personalización (*personalization principle*). La mayoría de las propiedades del discurso hablado y escrito de los vídeos pueden clasificarse de forma útil mediante conceptos de la LSF.

A.11.3 Contribuciones a la comunidad científica

A modo de resumen, las contribuciones principales de esta investigación a la comunidad científica son:

- Una revisión actualizada y amplia del estado actual de la investigación sobre el vídeo educativo.
- Un esquema de clasificación que enumera de forma exhaustiva y organiza conceptualmente las características del vídeo relacionadas con los procesos de aprendizaje.
- Nuevas evidencias empíricas acerca de los patrones de uso de los vídeos didácticos en los cursos en línea.

Hay que remarcar que la revisión de literatura y el esquema de clasificación se ofrecen desde una perspectiva amplia que además de los desarrollos de la comunidad del Aprendizaje Mejorado por Tecnología (TEL) acoge el conocimiento que nos pueden proporcionar el Análisis Cinematográfico y la Lingüística Sistémico-Funcional, así como los resultados previos de la investigación en Televisión y Cine Educativos.

A.11.4 Trabajos futuros

Como continuación directa del trabajo de investigación aquí expuesto, pueden desarrollarse estas líneas:

- **Géneros de vídeos**. A partir del estado actual, se puede completar la investigación descrita en los capítulos 2 y 3 y crear un catálogo de distintas configuraciones de presentación de los vídeos, sustentado en las taxonomías presentadas en esta tesis.
- **Crear un corpus de vídeos y de características.** Para un mejor entendimiento del esquema de clasificación, se puede elaborar un corpus de muestras reales de vídeos didácticos que muestren las diferentes características identificadas en esta clasificación.
- **Refinar la revisión de la literatura.** La fase final del proceso de clasificación (construcción de las taxonomías específicas) presenta algunas debilidades metodológicas que podrían superarse realizando una nueva iteración en la revisión de la literatura, con unos criterios de inclusión más estrictos.
- **Ampliar el estudio de campo.** El estudio sobre cursos MOOC se puede aplicar en otros repositorios públicos como YouTube, o incluso replicarlo en los mismos MOOC, aunque en esta ocasión utilizando el nuevo esquema de clasificación elaborado en esta tesis.
- **Aplicar el esquema de clasificación en el análisis de vídeos.** La clasificación se puede usar como un sistema de codificación en la anotación de vídeos didácticos, aplicado a una muestra de vídeos reales. Esto serviría para validar el sistema de clasificación en un contexto práctico.

Anexo B. Glosario bilingüe inglés/español

This is an English/Spanish glossary of the most relevant terms used in this dissertation.

Este es un glosario bilingüe inglés/español con los términos más destacados que aparecen en esta tesis. Su misión es ayudar a un lector castellanohablante en la delimitación del significado de los términos usados en el original inglés de esta investigación. En algunos casos se hacen aclaraciones sobre el contexto en el que se usa el término en este trabajo.

Actor. Actor.

Audio podcast. Podcast de audio.

Blended learning. Aprendizaje híbrido.

Board. Tablero.

Board-centric. Centrado en el tablero *(estilo de presentación de vídeo)*.

Classification scheme. Esquema de clasificación.

Clip (video). Videoclip.

Closed caption. Subtítulos.

Cognitive Load Theory. Teoría de la Carga Cognitiva (Sweller, 1988).

Cognitive Theory of Multimedia Learning (CTML). Teoría Cognitiva del Aprendizaje Multimedia.

Compositional. Composicional *(lingüística)*.

Computer Assisted Instruction (CAI). Instrucción Asistida por Computador.

Computer Based Learning (CBL). Aprendizaje Basado en Computadores.

Deictic. Deíctico (ej. «gesto deíctico»).

Diegetic / non-diegetic. Diegético / no diegético *(música o sonido en una película)*.

Discourse Analysis. Análisis del Discurso (Paltridge, 2012).

Documentary. Documental *(género cinematográfico)*.

Dual Coding Theory. Teoría de la Codificación Dual (Paivio, 1990).

Educational film. Filme educativo. Cine educativo. Película educativa.

Educational television. Televisión educativa.

Expository. Expositivo.

Film Analysis. Análisis cinematográfico.

Flipped class. Clases invertidas. Clases inversas.

Fluency. Fluidez *(de un orador, tanto oral como gestual)*.

Fly on the wall. Literalmente "Mosca [posada] en la pared". No se suele traducir.

Frame (video). Fotograma, cuadro (de la pantalla).

Genre. Género *(según el concepto manejado en Semiótica)* (Bhatia, 1997; Swales, 1990).

Gesture. Gesto.

Hypermedia. Hipermedia, hipermedios.

Hypervideo. Hipervídeo (vídeo no lineal).

Instructional design. Diseño de la instrucción.

Instructional video. Vídeo didáctico. Vídeo educativo.

Interactive, interactivity. Interactivo, interactividad.

Interpolated test. Test empotrado (en el vídeo).

Interview. Entrevista *(estilo de vídeo)*.

Layout. Composición, disposición *(de objetos físicos)*.

Learning object. Objeto de aprendizaje.

Lecture. Lección, conferencia, clase magistral *(depende del contexto)*.

Medium. Medio, soporte.

Microlecture. Microlección.

Mise-en-scène. Puesta en escena *(de una película)*.

MOOC (Massive Open Online Course). MOOC (curso en línea abierto y masivo).

Multimedia learning (MML). Aprendizaje multimedia.

Multimedia Learning Principles. Principios de aprendizaje multimedia (Mayer, 2014d).

Multimodal, multimodality. Multimodal, multimodalidad.

Multimodal Discourse Analysis. Análisis del discurso multimodal.

Narrative. Narrativo.

Overlay. Superposición.

Pedagogical agent. Agente pedagógico.

Podcast. Podcast *(no se traduce)*.

Rhetoric. Retórica.

Scene. Escena *(cine)*.

Screencast. Screencast *(no se traduce)*.

Segment. Segmento.

Semiotics. Semiótica (ciencia). *No confundir con «semiología»*.

Sequence. Secuencia.

Shot. Plano *(cine)*.

Slide. Diapositiva (elemento de presentación).

Social Appearance. Apariencia social.

Social Distance. Distancia social.

Social Presence Theory. Teoría de la Presencia Social (Minnesota State University, 1998).

Speaker-centric. Centrado en el orador *(estilo de presentación de vídeo)*.

Streamed video. Vídeo en *streaming*.

Systemic Functional Linguistics (SFL). Lingüística Sistémico-Funcional (Michael A. K. Halliday & Matthiessen, 2014).

Talking Head. "Busto parlante" (formato de presentación de video).

Taxonomy. Taxonomía.

Technology Enhanced Learning (TEL). Aprendizaje mejorado por la tecnología.

Video annotation. Anotación en vídeo.

Video-based learning (VBL). Aprendizaje basado en vídeo.

Video podcast. Podcast de vídeo.

Video tutorial. Videotutorial, tutorial de vídeo.

Videocast. Videocast.

Voiceover / voice over. Voz superpuesta.

José Miguel Santos Espino

ANATOMY OF INSTRUCTIONAL VIDEOS

A SYSTEMATIC CHARACTERIZATION OF THE STRUCTURE OF ACADEMIC INSTRUCTIONAL VIDEOS

Research on video-based learning has found several structural features in instructional videos with a potential influence in learning outcomes. The main goal of this thesis has been to build a systematic classification scheme for these characteristics. An inventory of characteristics has been collected through an extensive literature review and a field study on MOOC platforms. The development of the classification scheme is grounded in a multidisciplinary theoretical framework, which includes Cognitive Multimedia Learning theories, Film Analysis, Multimodal Discourse Analysis and Systemic Functional Linguistics. The resulting classification scheme comprises eight taxonomical domains: Medium, Presentation, Interaction, Spatiotemporal, Speech, Social Appearance, Strategic and Generic (for video genres). This dissertation also includes domain-specific taxonomies with a complete catalog of video characteristics.

ANATOMÍA DE LOS VÍDEOS DIDÁCTICOS

UNA CARACTERIZACIÓN SISTEMÁTICA DE LA ESTRUCTURA DE LOS VÍDEOS DIDÁCTICOS ACADÉMICOS

La investigación en el aprendizaje mediante vídeo ha identificado multitud de características estructurales de los vídeos con potencial para influir en el aprendizaje. El objetivo principal de esta tesis ha sido construir un esquema sistemático de clasificación para esas características. Se ha recopilado un inventario de características mediante una extensa revisión de la literatura y un estudio de campo sobre varias plataformas MOOC. La elaboración del esquema de clasificación ha recurrido a un marco teórico multidisciplinar que se nutre de las teorías cognitivas del aprendizaje multimedia, el análisis cinematográfico, el análisis del discurso multimodal y la Lingüística Sistémico-Funcional. El esquema de clasificación resultante comprende ocho dominios taxonómicos: Medio, Presentación, Interacción, Espaciotemporal, Habla, Apariencia Social, Estratégico y Genérico (géneros de vídeos). Esta tesis también incluye taxonomías específicas para cada uno de los dominios.

* contiene resumen en español