

Belinda Crawford Camiciottoli*

Department of Philology, Literature and Linguistics
University of Pisa, Italy
belinda.crawford@unipi.it

Veronica Bonsignori**

Language Centre
University of Pisa, Italy
bonsignori@cli.unipi.it

THE PISA AUDIOVISUAL CORPUS PROJECT: A MULTIMODAL APPROACH TO ESP RESEARCH AND TEACHING

Abstract

This paper presents an ongoing project sponsored by the University of Pisa Language Centre to compile an audiovisual corpus of specialized types of discourse of particular relevance to ESP learners in higher education. The first phase of the project focuses on collecting digitally available video clips that encode specialized language in a range of genres along an 'authentic' to 'fictional' continuum. The video clips will be analyzed from a multimodal perspective to determine how various semiotic resources work together to construct meaning. They will then be utilized in the ESP classroom to increase learners' awareness of the key contribution of different modes in specialized communication. We present some exploratory multimodal analyses performed on video clips that encode instances of political discourse across two different genres on the extreme poles of the continuum: a fictional political drama film and an authentic political science lecture.

139

Key words

multimodality, political discourse, ELAN, gestures, academic lectures, films.

* Corresponding address: Belinda Crawford Camiciottoli, Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa, Via Santa Maria, 67, 56126 Pisa – PI (Italy).

** The research was carried out by both authors together. Belinda Crawford Camiciottoli wrote sections 1., 2., 2.1., 2.2., 3. and 3.2. Veronica Bonsignori wrote the abstract and sections 3.1 and 4.

Sažetak

U radu se predstavlja projekat u toku finansiran od strane Centra za jezike Univerziteta u Pizi, čiji je cilj prikupljanje audio-vizuelnog korpusa specijalizovanih vrsta diskursa od posebnog značaja za studente engleskog jezika nauke i struke na visokom nivou obrazovanja. Prva faza projekta sastoji se iz prikupljanja digitalnih video klipova koji sadrže specijalizovani jezik iz čitavog niza žanrova u rasponu od onih 'autentičnih' do 'fikcionih'. Video klipovi se analiziraju sa stanovišta multimodalnosti kako bi se utvrdilo kako različiti semiotički resursi funkcionišu zajedno u konstrukciji značenja. Video klipovi će se koristiti u nastavi engleskog jezika struke u cilju povećanja svesti studenata o značaju različitih modaliteta u specijalizovanoj komunikaciji. U radu predstavljamo nekoliko preliminarnih multimodalnih analiza video klipova koji sadrže primere političkog diskursa iz dva različita žanra na krajnjim tačkama kontinuuma: filmske političke drame i autentičnog predavanja iz oblasti političke nauke.

Ključne reči

multimodalnost, politički diskurs, ELAN, gestikulacija, akademska predavanja, filmovi.

1. INTRODUCTION

Since Kress and van Leeuwen's (1996) and Lemke's (1998) pioneering work on the contribution of visual modalities to the construction of meaning, over the past two decades multimodality has become a 'hot topic' in discourse studies. It is now widely recognized that other semiotic resources (e.g. visual, aural, kinesic and spatial) beyond verbal language are crucial components of human communication (Jewitt, 2014). As O'Halloran (2011: 123) aptly commented, "communication is inherently multimodal". Reflecting this view are numerous highly influential works in the area of *multimodal discourse analysis* (O'Halloran, 2004; Scollon & Levine, 2004; Norris, 2004; Baldry & Thibault, 2006; Bateman & Schmidt, 2012, among others), which aim to shed light on the meanings and functions of multiple semiotic modes (and how they interact) in situated interaction.¹

The growing interest in multimodality is clearly linked to the rapid and ongoing development of digital interactive media (Hyland, 2009), which now

¹ Other research approaches that have drawn extensively on semiotics include conversation analysis (e.g. Heath, 1984; Kendon, 1990), mediated discourse analysis (e.g. Scollon, 2001), semiotic remediation (e.g. Prior, Hengst, Roozen, & Shipka, 2006), in addition to studies in the area of linguistic anthropology (e.g. Silverstein, 2004).

permeates our social practices across all spheres of activity. This trend is also evident in educational settings, which now strongly promote the acquisition of *multimodal literacy*, i.e. the ability to construct meanings through “reading, viewing, understanding, responding to and producing and interacting with multimedia and digital texts” (Walsh, 2010: 213). The need to foster multimodal literacy and practices in education is widely recognized (Royce, 2002; Jewitt & Kress, 2003), also to take advantage of today’s learners’ increasing propensity to use multi-semiotic digital resources in their daily lives (Street, Pahl, & Rowsell, 2011). To help learners develop skills for comprehending and producing texts that combine several modes, new pedagogic practices are needed to enhance the awareness of multimodality not only on the part of learners, but also of teachers. This dual necessity is seen in two different approaches to multimodal research that target instructional contexts. The first focuses on the communication process between the teacher and the students mediated through both face-to-face classroom interaction and learning materials, e.g. texts, images, websites, audiovisual resources. The second approach emphasizes how to teach multimodal discourse to students, i.e. how different semiotic resources combine to construct meaning in educational settings. Much of the research in this area stems from the New Literacy Studies paradigm, which rejects the idea of literacy based only on the ability to read and write (Gee, 1996), as well as the work of the New London Group that addresses issues relating to the teaching of multimodal discourse associated with new technologies (New London Group, 1996).²

In the language classroom, the multimodal approach has numerous advantages. In addition to helping learners develop skills to understand and produce texts in the target language which integrate several modes (O’Halloran, Tan, & Smith, in press), it also raises awareness of cultural differences in how people communicate non-verbally. With particular reference to English language teaching, some studies have demonstrated the benefits of exposing learners to input that combines several modes. Busà (2010) found that this approach enabled students to improve their ability to structure different types of discourse and to communicate orally using language, intonation patterns and gestures. In the context of listening comprehension, gestures can reinforce a verbal message, which has the effect of reducing ambiguity and therefore enhancing understanding (Antes, 1996). Gestures can also effectively integrate with speech when used by teachers to assist learners during unplanned explanations of vocabulary (Lazarton, 2004). Sueyoshi and Hardison (2005) showed that learners of English at lower proficiency levels benefitted from gestures and facial cues to improve understanding. However, gestures were also found to help higher proficiency learners, not only by replicating verbal meanings, but also by extending them in

² The original members of the New London Group were Courtney B. Cazden, Bill Cope, Norman Fairclough, James Paul Gee, Mary Kalantzis, Gunther Kress, Allan Luke, Carmen Luke, Sarah Michaels and Martin Nakata. They met in New London, New Hampshire (USA) to discuss new prospects for teaching and learning literacy, also introducing the term *multiliteracies*.

more complex ways (Harris, 2003). Because non-verbal signals often replicate verbal messages, they reduce ambiguity and therefore enhance understanding, particularly among lower proficiency L2 learners who may not completely grasp verbal meaning (Kellerman, 1992; Antes, 1996).

As the research reviewed above shows, the ability to exploit different semiotic resources plays a key role in language learning. It thus stands to reason that input beyond the verbal mode is particularly important in situated communicative contexts where domain-specific linguistic, discursive, pragmatic and cultural features can create significant obstacles for language learners. More specifically, in ESP contexts, a multimodal approach to teaching and learning can provide students with a new strategy to cope with the challenges of specialized language. In the following section, we describe an ongoing project sponsored by the Language Centre of the University of Pisa that aims to respond to the special needs of our ESP students from a multimodal perspective.

2. THE PISA AUDIOVISUAL CORPUS PROJECT

At the University of Pisa, a considerable amount of English language teaching takes place in non-humanistic instructional contexts, including business, tourism, political science, law and medicine. In these degree programmes, large numbers of students are required to achieve adequate levels of English language competence in order to graduate and successfully enter the professional world. To address these needs, a project was recently launched by the Corpus Research Unit of the University of Pisa Language Centre to compile an audiovisual corpus which represents specialized types of discourse in the above-mentioned discourse domains that are of particular relevance both to our students and to ESP learners in higher education in general. The underlying rationale of the project is to exploit audiovisual resources not only as a way to increase exposure to the target language, but also to better connect with learners who make massive use of multi-semiotic digital resources in their daily lives. Referring to such resources, Rose (1997: 283) pointed out that “in foreign language contexts, exposure to film is generally the closest that language learners will ever get to witnessing or participating in native speaker interaction”. In addition, audiovisual materials allow learners to understand how language is used in the domain-specific contexts that are represented in the digital sources.

2.1. Corpus design and preparation

The Pisa Audiovisual Corpus project was inspired by the Library of Foreign Language Film Clips developed at the Berkeley Language Centre of the University

of California,³ a large database with currently over 15,000 short clips (max 4 minutes) from approximately 400 films in 24 different languages. Of particular interest is the fact that the clips are annotated with searchable descriptors of genre, linguistic features (e.g. idioms, irony, metaphor), speech acts (e.g. greetings, apologies, compliments) and cultural-related content linked to contexts of usage (e.g. politics, sports, employment), all of which renders them highly effective tools for teaching foreign languages. Similarly, the aim of our project is to collect a corpus of video clips, but with content that is carefully selected to represent specialized discourse domains (rather than general English), also including a variety of different genres beyond filmic language. We plan to annotate linguistic elements that often create difficulties for ESP learners (e.g. idioms, use of humour, figurative language, phrasal verbs, rhetorical devices, culture-specific references), as well as the non-verbal features (e.g. gesturing, gaze direction, prosodic patterns) that may contribute to their meanings in situated communication. In this way, we can gain insights into how non-verbal signals may contribute meanings in various ways, including replication, reinforcement and integration (cf. Antes, 1996; Lazarton, 2004; Harris, 2003), as well as modification and contradiction (Argyle, 2007), and thus render such meanings more accessible to ESP learners. The multimodally annotated corpus will then be used for both research and teaching purposes. More specifically, the annotated video clips can be utilized in the ESP classroom to increase learners' awareness of the contribution of different semiotic modes to meaning in specialized communicative contexts.

The corpus has been designed to represent specialized language from the domains of business and economics, political science, law, science (e.g. biology, environmental issues), medicine and health, and humanities (e.g. art, tourism). The video clips are currently being collected from a variety of different genres to which ESP students may be exposed, reflecting a continuum in terms of formality, communicative mode, scientificity and authenticity, as illustrated in Figure 1.

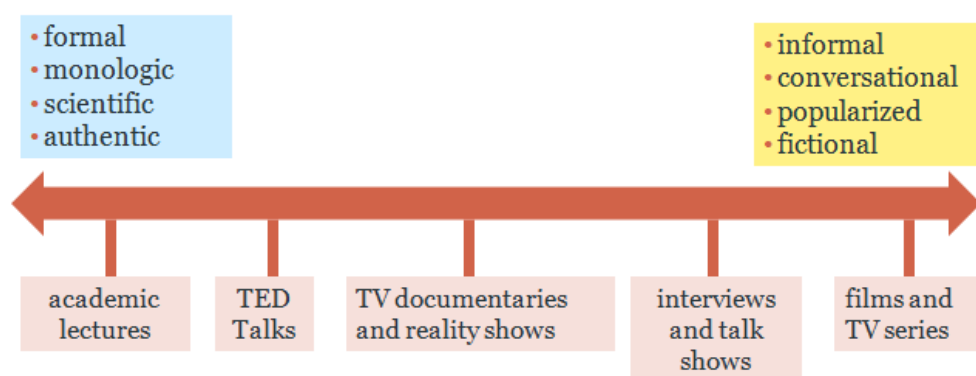


Figure 1. The genre continuum of the Pisa Audiovisual Corpus

³ <http://blcvideoclips.berkeley.edu/>

As can be seen, on the far left end of the continuum we have discipline-specific academic lectures and TED Talks that are typically instances of formal, monologic, scientific and authentic discourse. On the far right end, films and TV series can be prototypically described as instances of informal, conversational, popularized and fictional discourse. Other genres represent various stages along the continuum. For example, TED Talks are short speeches (18 minutes or less) given by experts on a wide range of topics from science to business to global issues, but in a popularized format.⁴ TED Talks are now recognized as an important resource in instructional contexts and particularly in language teaching (Takaesu, 2013; Crawford Camiciottoli & Querol-Julián, forthcoming 2015). Moving towards conversational and relatively informal discourse are genres such as interviews and talk shows. All of these genres constitute important resources for ESP teaching, especially when learners have limited opportunity to experience the target language in face-to-face interactions outside the classroom.

The first phase of the project entails selecting and preparing short video clips that are available from accessible digital resources. In particular, digital materials were first viewed according to precise criteria in order to identify specific scenes to be cut into clips, i.e. those that contained specialized vocabulary, speech-acts typical of specialized discourse, rhetorical devices, prominent gestures, etc., and thus particularly relevant to our study. Depending on the particular genre, the preparation process may also involve the incorporation of other instruments that can facilitate teaching, for example a transcription of the speech if not already available, as well as translation, dubbing and subtitling options.

2.2. The analytical approach

The multimodal analysis of video clips focuses on the interplay between non-verbal features (e.g. facial expressions, gaze direction, hand and arm gestures, body posturing, prosodic elements) and verbal features that may characterize the particular genre and communicative context. Towards this aim, we decided to adopt both *corpus-based* and *corpus-driven* approaches. In the first approach, digital sequences are selected according to pre-established verbal elements that are characteristic of the discourse in question, and then co-occurring non-verbal features are analyzed to determine their contribution to meaning. In the second approach, features of interest are allowed to emerge from the data, without referring to pre-established sets of verbal or non-verbal features to be investigated. In other words, we first view the video resources to identify any prominent non-verbal signals (i.e. those that we perceive as particularly visible or intense with respect to others used by a given speaker), and then analyze the co-occurring verbal elements.

⁴ <https://www.ted.com/talks>

To analyze the video clips, we broadly refer to the principles of multimodal discourse analysis (cf. O'Halloran, 2004; Scollon & Levine, 2004; Norris, 2004), which place particular emphasis on the interaction of multiple semiotic resources in situated communicative contexts, and the meanings that this interaction generates. This approach can make use of various systems for the multimodal transcription, all of which are capable of incorporating both verbal and non-verbal elements and highlighting the interaction between them (cf. Baldry & Thibault, 2006; Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006; Wildfeuer, 2013). In the following section, we describe an application of multimodal discourse analysis to some exploratory work with video resources from two genres in the domain of political discourse: a political drama film and a political science lecture. We chose to focus on these two genres that correspond to the extreme poles of the continuum illustrated in Figure 1 in an effort to offer the most diversified illustration of the corpus as possible.

3. POLITICAL DISCOURSE: TWO CASE STUDIES

Both of the analyses presented below utilized ELAN software (Wittenburg et al., 2006) as the analytical instrument for the multimodal transcription and annotation of the digital resources.⁵ With this application, it is possible to create and insert information that is synchronized with streaming audio and video input. For example, the transcript can be time aligned with speech production and user-defined annotations that mark specific verbal and non-verbal features of interest may also be incorporated. All annotations are then displayed in multiple levels or tiers under the streaming video (Wittenburg et al., 2006).

Because hand and arm gestures were frequent in the video resources of both cases, we decided to adopt common frameworks for the physical description of the gestures and for their functional interpretation. For gesture description, we referred to Querol-Julián's (2011) system of abbreviations, e.g. PalmUMUp (palm up and moving upwards), HandsRotOut (hands rotating outward) and FingRing (fingers forming a ring). For gesture function, we followed 1) Kendon's (2004) pragmatic functions of gestures, i.e. *modal* (to express certainty/uncertainty), *performative* (to illustrate a speech act) and *parsing* (to separate units within a stretch of discourse), and 2) Weinberg, Fukawa-Connelly and Wiesner's (2013) functional classification of gestures as *indexical* (to indicate a referent), *representational* (to represent an object or idea) and *social* (to emphasize the message or increase the speaker's connection to the audience).

⁵ ELAN was developed at the Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. It is freely available at <http://tla.mpi.nl/tools/tla-tools/elan/>.

3.1. The political drama film

Film language has been defined as a hybrid genre along the continuum between spoken and written language. More specifically, referring to the film script, Gregory and Carroll (1978: 42) described it as “written to be spoken as if not written”, because, as Taylor (1999: 262) explained, in this particular text type “the original dialogue is not real, merely purporting to be real”. The fictional character of film language is supported by other scholars who also define it as “prefabricated orality” (Chaume, 2004) or as “written discourse imitating the oral” (Gambier & Soumela-Salme, 1994). This is mainly due to the fact that film language is subject to various constraints in terms of length, immediacy, linguistic relevance of the utterances and, above all, it is the result of the decision of the scriptwriter. However, recent studies have also demonstrated the similarities between film language and everyday conversation in terms of spontaneity and authenticity (cf. Kozloff, 2000; Quaglio, 2009; Forchini, 2012; Bonsignori, 2013). In addition, it is worth noticing that certain films do not even have a written script because of the choice of the director to let actors be free and sound as spontaneous as possible – examples are Mike Leigh and Ken Loach who also often recruits non-professional actors (Taylor, 2006; Bonsignori, 2013). Finally, in teaching settings film discourse is seen as “an authentic source material (that is, created for native speakers and not learners of the language)” (Kaiser, 2011: 233), and there are several studies showing how to exploit the multi-semiotic character of films in language learning (among the most recent, see Kaiser & Shibahara, 2014; Bruti, 2015), by integrating the various elements such as language, image, sound, gestures, that altogether contribute to creating meaning.

For the purposes of the present paper, the political drama film *The Ides of March*⁶ (2011, George Clooney) was chosen. It is an adaptation of the play by Beau Willimon *Farragut North* (2008), based on his work as a former campaign staffer. In fact, it tells the story of a junior campaign manager, Stephen Meyers (played by Ryan Gosling), for Governor Mike Morris (George Clooney), a Democratic presidential candidate who is competing against Senator Ted Pullman in the Democratic primary. Therefore, the film is full of specialized vocabulary and various instances of political discourse, ranging from political speeches, political interviews and the like, which makes it a perfect case study for research purposes and ideal material for teaching.

The film was analyzed using a corpus-based approach (cf. section 2.2.). More specifically, clips were created by selecting the film sequences where relevant verbal rhetorical strategies were used as a characterizing feature of political discourse.⁷ Subsequently, any co-occurring non-verbal elements were investigated.

⁶ For further information on this film, see http://www.imdb.com/title/tt1124035/?ref_=nv_sr_1.

⁷ For example, the use of parallel structures, metaphors, irony (Beard, 2000; Partington & Taylor, 2010).

In the following paragraphs, we present the analysis of the two clips extracted from *The Ides of March* using the multimodal annotation software ELAN.

Tiers	Controlled vocabulary	
	Abbreviation	Description
Transcription		
Gesture_description	OPsU	opening palms up
	CPsD	closing palms down
	OPUtl	opening palm up towards interlocutor
	PUMf	palm up moving forward
	JHsC	joined hands to the chin
	OPDhT	open palm down hitting the table
Gesture_function	indexical	to indicate position
	modal	to express certainty
	parsing	to mark different units within an utterance
	performative	to indicate the kind of speech act
	representational	to represent an object/idea
Gaze	social	to emphasize/ highlight importance
	Up	up
	Down	down
	Left	left
	Right	right
Face	Out	out
	frowning	frowning
	smiling	smiling
Prosody	Stress	paralinguistic stress
Notes	description of camera angles	

Table 1. Annotation framework for the political drama film (from clip 2)

Table 1, which refers to one of the clips (namely, clip 2), exemplifies the multi-level or tiered analytical structure created in the software: (1) the left-hand column lists some verbal and non-verbal elements, such as gesture, gaze, face, etc., and (2) the controlled vocabulary column contains the labels used to annotate them, abbreviated when necessary along with the explanation or description of what they refer to.

Starting with clip 1, it lasts only 32 seconds and shows a part of a debate for the primary elections with Senator Pullman and Governor Morris at the auditorium of Miami University, Ohio. In the full sequence, apart from the two candidates who stand on the podium, there is also a moderator and, of course, the audience. However, in this clip, Governor Morris exploits the provocative question asked by his opponent, Senator Pullman, to deliver a speech. This is why this extract has been classified as monologic political speech. Clip 1 is characterized by some examples of the use of parallelism, an important rhetorical strategy in political discourse used for evaluation and persuasion (Partington & Taylor, 2010). A parallel structure can take on various forms: a bicolon, namely expressions

containing two parallel phrases, and a tricolon which consists of expressions containing three parallel phrases or even the repetition of three words.⁸ Following is the transcription of the speech delivered by Governor Morris, answering his opponent's question: "Would you call yourself a Christian?", which contains two bicolons at the beginning and a tricolon at the end:

I am not a Christian or an atheist. I'm not Jewish or a Muslim. What I believe, my religion, is written on a piece of paper called the Constitution. Meaning that I will defend until my dying breath your right to worship whatever god you believe in, as long as it doesn't hurt others. I believe we should be judged as a country by how we take care of people who cannot take care of themselves. That's my religion. If you think I'm not religious enough, don't vote for me. If you think I'm not experienced enough or tall enough, then don't vote for me, 'cause I can't change that to get elected. (clip 1)

The screenshot in Figure 2 shows the multimodal analysis of the two bicolons at the beginning of his speech. As can be seen in the Transcription tier, in correspondence with the parallel structure some words are stressed – as pointed out by the annotation "stress" in the Prosody tier below – that is, words corresponding to new content, e.g. *Christian*, *Jewish* and *Muslim*, and some gestures are performed. For example, focussing on the first bicolon, a gesture labelled as PU-TFf in the Gesture_description tier, corresponding to "palm up - thumb and forefinger forward", is employed with the function of parsing to mark different units within the utterance. Finally, gaze is directed down and then to the right. In the case of the second bicolon, corresponding to the words *Jewish* and *Muslim* the gestures "PUMbf", namely "palm up moving back and forth" and "PUMud", that is "palm up moving up and down", are used respectively.

⁸ The tricolon may even entail the same exact word repeated three times, as in the famous slogan against Margaret Thatcher: "Maggie Maggie Maggie! Out out out!".

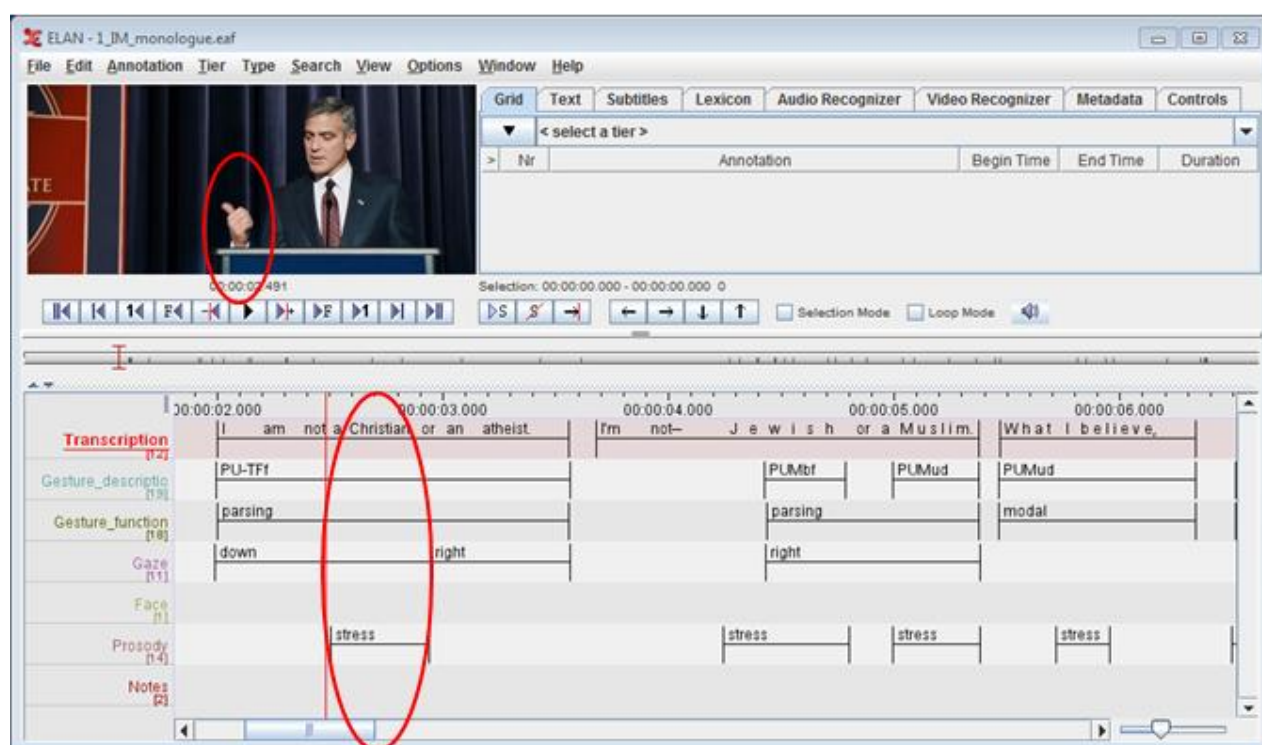


Figure 2. Multimodal analysis of parallelism from clip 1 – political speech

Clip 2 is 1 minute long and shows a part of an interview for the primary elections between journalist Charlie Rose and Governor Morris, surrounded by members of Governor Morris' staff and TV crew. The setting is a hotel conference room transformed into an interview space. The communicative exchange in this clip corresponds to the political interview, a type of spoken institutional language that takes the form of various types of questions and responses (Partington & Taylor, 2010). Figure 3 below shows a screenshot of the multimodal analysis of clip 2.

As Figure 3 below shows, there are two Transcription tiers: one for Governor Morris and the other for journalist Charlie Rose, which also allows us to point out cases of overlapping. The labels that appear in the following tiers refer to the person who is speaking. For instance, in this case and highlighted in green, the gesture labelled as "PUMF", namely "palm up moving forward", refers to Charlie Rose, who employs it with an indexical function when asking the Governor a question. Moreover, the Notes tier is also important because it describes camera angles and close-ups of one speaker, thus leaving out the other interlocutor, or as in this case, it tells us when it focuses on the audience – more specifically, on Governor Morris' staff – so that both interlocutors are actually off-screen and we can only hear but not see them. This is an important feature of the film genre, which has to be taken into account when analyzing audiovisual material.

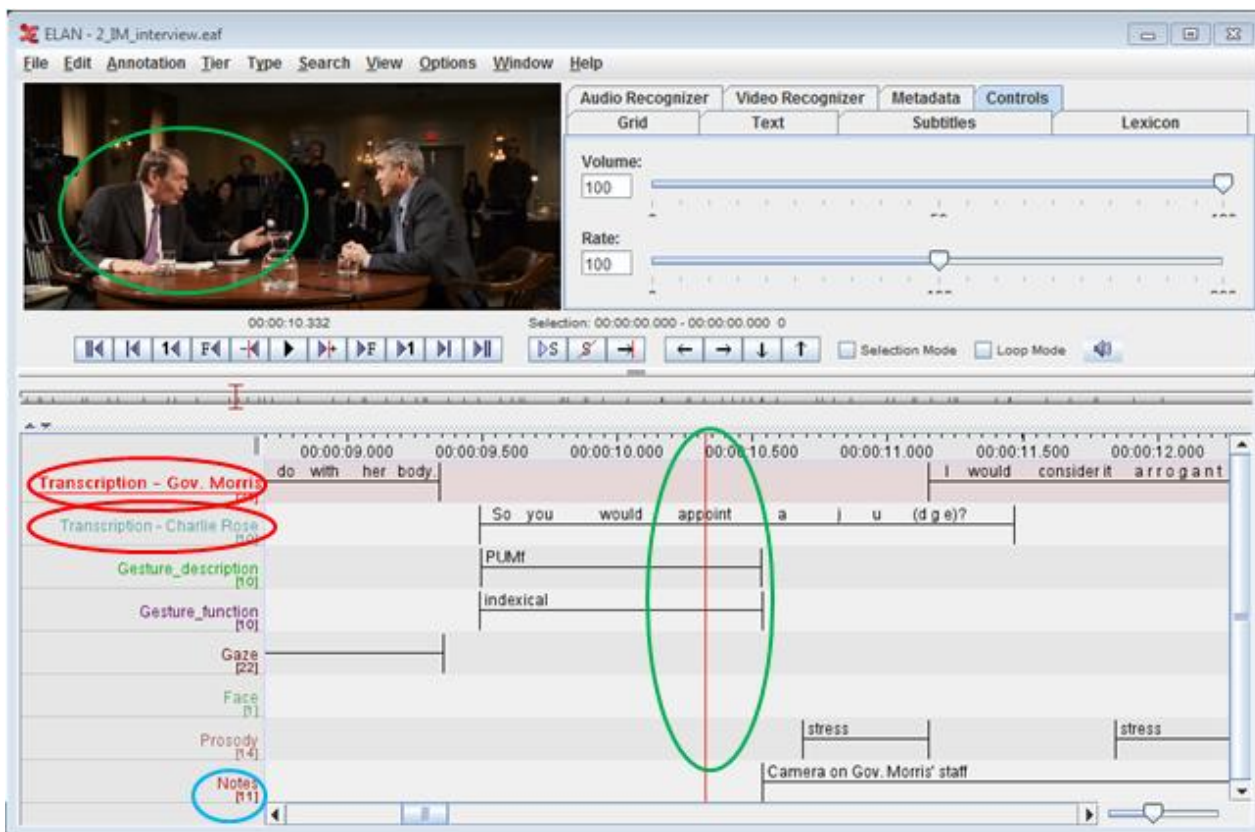


Figure 3. Multimodal analysis of Q&R from clip 2 – political interview

Finally, Table 2 below summarizes the types of gestures occurring in the political interview portrayed in clip 2, which is useful to draw some preliminary conclusions on this type of political discourse.

Gesture	Transcription	Abbreviation	Description	Function
1 (1)	(GM) Eh... Would I do it?	OPsU	opening palms up	social
2 (1)	(GM) No!	CPsD	closing palms down	modal
3 (1)	(GM) certainly not a government ,	OPUtl	opening palm up towards interlocutor	modal
4 (1)	(CR) So you would appoint a ju (dge)?/	PUMf	palm up moving forward	indexical
5 (1)	(CR) It gets more complicated when it's personal.	JHsC	joined hands to the chin	modal
6 (5)	(CR) So you, you , Governor, would impose a death penalty.	OPDhT	open palm down hitting the table	social

Table 2. Summary of gestures from clip 2

In the first column, we can see the six types of gestures found in the whole clip, along with their frequencies in parentheses. In the transcription column, lexical

items in bold indicate the presence of prosodic stress. Regarding the total number of gestures, fewer gestures appear in the political interview format than in the monologic political speech, i.e. 10 vs. 18 respectively, despite the longer length of clip 2. Concerning their function, most of the gestures have a social function (in the case of Charlie Rose) and modal function (in the case of Governor Morris). This can be explained by the fact that the journalist uses them to give the floor to the Governor to get an answer, while in his responses, the Governor needs to show some kind of self-confidence in order to be as convincing and persuasive as possible. One last remark refers to the need to point out and take into account that in this clip, neither of the two speakers appears in many sequences, which highlights once again the paramount importance of the role of camera angles and the types of shots in the film genre.

3.2. The political science lecture

Academic lectures have always formed the backbone of higher education. As a type of pedagogic discourse, academic lectures consist of a set of specialized practices that are designed to transmit knowledge and skills from experts to novices (Bernstein, 1986). In the context of language teaching, the ability to understand academic lectures is often a challenging endeavour for L2/FL listeners. In fact, they are faced with a combination of potentially unfamiliar phonological, lexicosyntactic, structural, pragmatic and cultural features that must be decoded during the listening process (Flowerdew, 1994). In particular, ESP learners must also deal with discipline-specific vocabulary and discursive patterns. The need to acquire solid academic listening skills has become even more pressing, given the growing trend in student mobility in which increasing numbers of L2/FL learners are undertaking study programmes in English-medium universities. It is thus important to prepare ESP listeners for content lectures before study abroad experiences to avoid persistent difficulties linked to lack of comprehension (Crawford Camiciottoli, 2010). For these reasons, we have included academic lectures among the genres represented in the Pisa Audiovisual Corpus.

Thanks to ongoing developments in educational technology, it is now possible to access authentic academic lectures in digital format from Internet sources. Since their launch in 2002 by the Massachusetts Institute of Technology, OpenCourseWare lectures have seen an enormous expansion, giving vast numbers of people all over the world the opportunity to experience high-quality instruction.⁹

This case study is based on a political science lecture entitled “Introduction: What is Political Philosophy”. It was accessed from Yale University’s Open Courses

⁹ For example, thousands of lectures across disciplines are freely available from the Open Education Consortium Internet portal and iTunes University.

website, specifically from the course *Introduction to Political Philosophy*, and delivered by a highly esteemed professor at Yale University.¹⁰ The lecture was recorded in the fall of 2006 and is approximately 37 minutes long. The lecture took place in what appeared to be a large lecture hall, with the lecturer positioned behind a podium. No supporting materials were used during the lecture and there was no verbal interaction with students, thus reflecting a monologic, frontal and non-interactive style (Morell, 2004). The rationale for this traditional format is perhaps linked to the aims of the Open Courses website. In fact, the lectures available on the website are mostly monologic and frontal. Other more participatory formats would likely be too dispersive in terms of content and also problematic to film clearly for a remote Internet audience.

The lecture was analyzed using a corpus-driven approach (cf. section 2.2.), i.e. without referring to any pre-established features. The entire lecture was first viewed to identify potentially prominent non-verbal features that could be systematically observed. Hand/arm gestures and direction of the lecturer's gaze (i.e. outward towards the audience or downwards toward the podium) were clearly visible, while facial expressions could not be analyzed due to lack of close-ups during filming. During this process, we determined that the audio quality was also sufficiently clear to distinguish instances of prosodic stress. The video was then viewed a second time to isolate the types of verbal elements that co-occurred with prominent non-verbal features. On the basis of this information, we created a structural framework of verbal and non-verbal features to be annotated in lecture video, again using ELAN. Table 3 provides an overview of the analytical framework.

Analytical tiers	Annotations in ELAN	Description/explanation
Transcript		Speech inserted under sound wave
Humour	<ul style="list-style-type: none"> • Pun • Joke • Irony 	<ul style="list-style-type: none"> • pun • joke • irony
Interaction	<ul style="list-style-type: none"> • Persuasion • Idiom • Imperative • Compcheck 	<ul style="list-style-type: none"> • modal verbs/personal pronouns used as persuasive features • idioms • imperatives • question to verify comprehension
Prosody	<ul style="list-style-type: none"> • Stress 	<ul style="list-style-type: none"> • presence of prosodic stress
Gaze	<ul style="list-style-type: none"> • Out • Down 	<ul style="list-style-type: none"> • gaze directed outwards • gaze directed downwards
Gesture_description	<ul style="list-style-type: none"> • FingTouch • FingBunch • FingerPoint • PalmsOpApUD 	<ul style="list-style-type: none"> • fingers touching together • fingers bunched, moving up and down • finger pointing towards audience • palms open and apart, moving up and down

¹⁰ Steven B. Smith, *Introduction to Political Philosophy* (Yale University: Open Yale Courses), <http://oyc.yale.edu/> (Accessed December 14, 2013). License: Creative Commons BY-NC-SA. (<http://oyc.yale.edu/terms>): <http://creativecommons.org/licenses/by-nc-sa/3.0/us/>.

	<ul style="list-style-type: none"> • PalmOpUD • HandVertUD • HandsSwSD • HandsPray • HandsRotCtr • HandsWApD • HandsClasp 	<ul style="list-style-type: none"> • palm open, moving up and down • hand vertical moving up and down • hands sweeping sideways • hands in praying position • hands rotating at centre of body • hands wide apart moving down • hands clasped together in front of body
Gesture_function	<ul style="list-style-type: none"> • Social • Repres • Index • Parsing • Perform • Modal 	<ul style="list-style-type: none"> • social (emphasizes a message) • representational (represents object/idea) • indexical (indicates a referent) • parsing (distinguishes units of speech) • performative (illustrates speech act) • modal (expresses certainty/uncertainty)

Table 3. Annotation framework for the political science lecture

In the following paragraphs, we illustrate the multimodal analysis of some co-occurrences of verbal and non-verbal elements. For reasons of space, these will be limited to four episodes involving an idiom (humour), a pun (humour), persuasion (interaction) and a comprehension check (interaction). Figure 4 reproduces two screenshots from ELAN that analyze the contribution of non-verbal cues to the idiom “cart before the horse” followed by the pun “the cart before the course”. In the left screenshot, we see the analysis of the idiom with the corresponding lexical items positioned under the sound wave of the speech in the Transcript tier. Underneath this is the annotation “Idiom” in the Interaction tier, “Out” in the Gaze tier, “PalmsOPApUD” in the Gesture_description tier, and “Social” in the Gesture_function tier. The still image shows how the lecturer uses gesturing and gaze to call attention to the idiom. In the right screenshot, the pun (in the Humour tier) has been annotated in the same way for Gesture tiers. However, prosodic stress is present in this case, also evident from the sound wave above the word “course”. Gaze instead is downward, which could be a reflection of the lecturer’s overall understated style of communication.

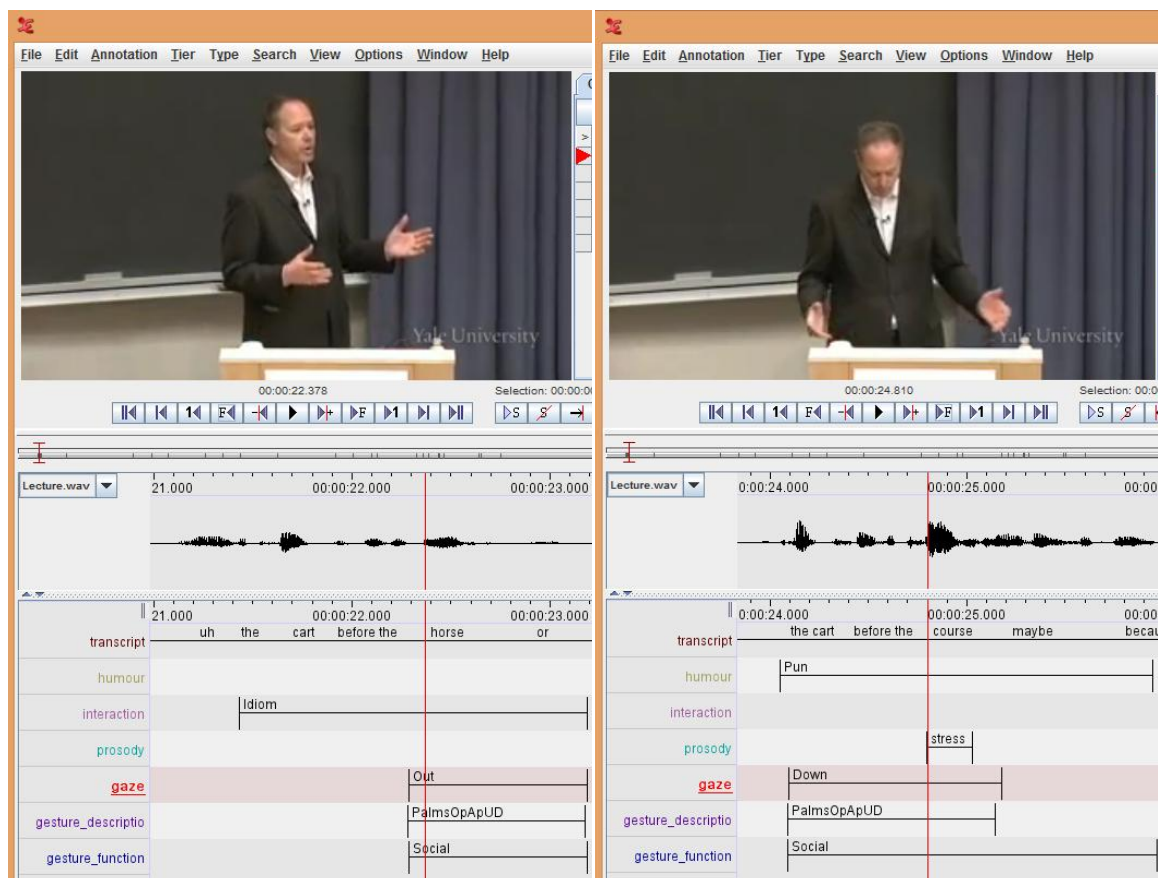


Figure 4. Multimodal analysis of an idiom followed by a pun

Figure 5 reproduces a screenshot that illustrates the multimodal annotation of an episode of persuasive language containing the modal *should* and the inclusive pronoun *we* (“one thing we should not do”) followed by a comprehension check (“right?”). As can be seen, the verbal elements have been both annotated in the Interaction tier. They are accompanied by gaze outwards, as well as a praying-like gesture that has a social function to reinforce the message. The comprehension check that follows is instead not accompanied by any prominent non-verbal cues. The lecturer then continued by repeating once again the same verbal input, thereby marking its rhetorical force even more.

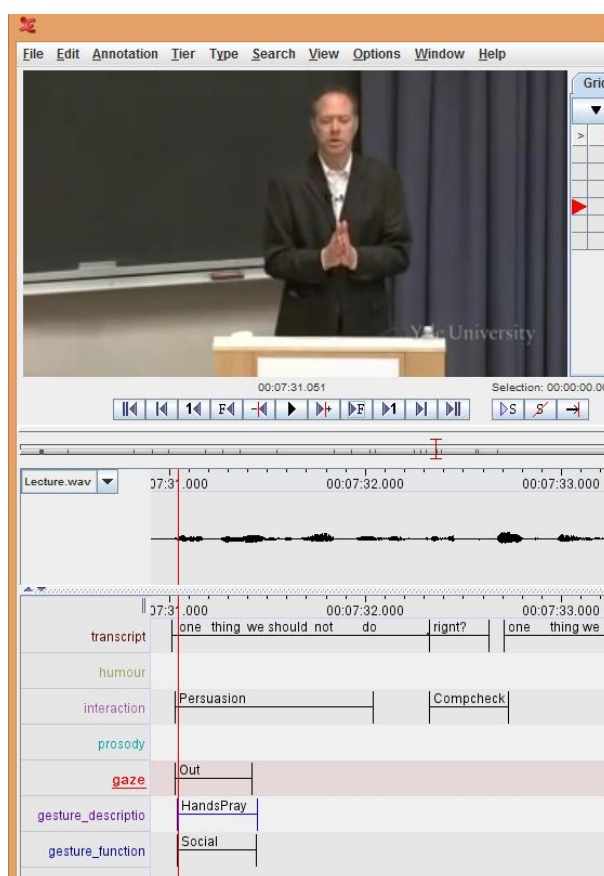


Figure 5. Multimodal analysis of persuasion followed by a comprehension check

The above examples of multimodal analysis of the interplay between verbal and non-verbal elements during this academic lecture suggest that the latter makes an important contribution to instructional discourse. We interpret the synergistic interaction of the two modes as a means to enhance understanding and promote an atmosphere that facilitates learning. Once this type of analysis has been performed on the entire lecture video, the next step would be to prepare short clips of about five minutes that contain particularly rich combinations of different semiotic modes for use in the classroom to help ESP learners to approach lecture comprehension from a multimodal perspective.

4. CONCLUDING REMARKS

Across the two different genres analyzed, i.e. the academic lecture and the film (which comprises both monologic speech and the interview format), non-verbal signals often co-occurred with verbal rhetorical elements. This leads us to hypothesize that rhetorical meaning may be a characterizing feature of political

discourse in general, regardless of genre and communicative mode, as suggested in many studies reporting the importance of bodily behavior in persuasive discourse and political communication (cf. among many, Atkinson, 1984; Poggi & Pelachaud, 2008). Moreover, interestingly, gestures co-occurred with specific intonational patterns, such as prosodic stress. Clearly, to corroborate our findings, it would be necessary to carry out multimodal analyses of political discourse in other genres, beyond the two dealt with in this exploratory study.

With particular reference to the Pisa project, the next step entails continuing with data collection across different genres indicated in the continuum illustrated in Figure 1, including TV documentaries, TED Talks, authentic interviews, and across other domains such as medicine and health, economics, law, tourism, etc. In addition, it will also be necessary to define a finer-grained set of annotations to enhance the description of audiovisual resources from a multimodal perspective across such genres and domains, and finally, to develop a database of annotated video clips useful for ESP teaching and research.

[Paper submitted 2 Sep 2015]

[Revised version accepted for publication 27 Nov 2015]

References

- Antes, T. A. (1996). Kinesics: The value of gesture in language and in the language classroom. *Foreign Language Annals*, 29(3), 439-448.
- Argyle, M. (2007). *Bodily communication* (2nd ed.). London: Routledge.
- Atkinson, M. (1984). *Our masters's voices. The language and the body language in politics*. London: Mathuen.
- Baldry, A., & Thibault, P. J. (2006). *Multimodal transcription and text analysis*. London: Equinox.
- Bateman, J., & Schmidt, K.-H. (2012). *Multimodal film analysis: How films mean*. London: Routledge.
- Beard, A. (2000). *The language of politics*. London: Routledge.
- Bernstein, B. (1986). On pedagogic discourse. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 205-240). New York: Greenwood Press.
- Bonsignori, V. (2013). *English tags. A close-up on film language, dubbing and conversation*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Bruti, S. (2015). Teaching learners how to use pragmatic routines through audiovisual material. In B. Crawford Camiciottoli, & I. Fortanet-Gómez (Eds.), *Multimodal analysis in academic settings. From research to teaching* (pp. 213-236). London: Routledge.
- Busà, M. G. (2010). Sounding natural: Improving oral presentation skills. *Language Value*, 2(1), 51-67.
- Chaume, F. (2004). *Cine y traducción* [Cinema and translation]. Catedra: Signo e Imagen.

- Crawford Camiciottoli, B. (2010). Meeting the challenges of European student mobility: Preparing Italian Erasmus students for business lectures in English. *English for Specific Purposes*, 29(4), 268-280.
- Crawford Camiciottoli, B., & Querol-Julián, M. (in press). Chapter 23. Lectures. In K. Hyland, & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes*. London: Routledge.
- Flowerdew, J. (Ed.) (1994). *Academic listening. Research perspectives*. Cambridge: Cambridge University Press.
- Forchini, P. (2012). *Movie language revisited. Evidence from multidimensional analysis and corpora*. Bern: Peter Lang.
- Gambier, Y., & Soumela-Salme, E. (1994). Subtitling: A type of transfer. In F. Eguíluz, M. R. Merino, V. Olsen, E. Pajares., & J. M. Santamaría (Eds.), *Transvases culturales: Literatura, cine, traducción* [Cultural transfer: Literature, cinema, translation] (pp. 243-252). Vitoria-Gasteiz: Universidad del País Vasco.
- Gee, J. P. (1996). *Sociolinguistics and literacies: Ideology in discourses* (2nd ed.). London: Taylor & Francis.
- Gregory, M., & Carroll, S. (1978). *Language and situation: Language varieties and their social contexts*. London: Routledge & Kegan Paul.
- Harris, T. (2003). Listening with your eyes: The importance of speech-related gestures in the language classroom. *Foreign Language Annals*, 36, 180-187.
- Heath, C. (1984). Talk and reciprocity: Sequential organization in speech and body movement. In J. M. Atkinson, & J. Heritage (Eds.), *Structures of social action. Studies in conversation analysis* (pp. 247-265). Cambridge: Cambridge University Press.
- Hyland, K. (2009). *Academic discourse: English in a global context*. London: Continuum.
- Kozloff, S. (2000). *Overhearing film dialogue*. Berkeley: University of California Press.
- Jewitt, C. (Ed.) (2014). *The Routledge handbook of multimodal analysis*. London: Routledge.
- Jewitt, C., & Kress, G. (Eds.) (2003). *Multimodal literacy*. New York: Peter Lang.
- Kaiser, M. (2011). New approaches to exploiting film in the foreign language classroom. *L2 Journal*, 3(2), 232-249. Retrieved from <http://escholarship.org/uc/item/6568p4f4>
- Kaiser, M., & Shibahara, C. (2014). Film as a source material in advanced foreign language classes. *L2 Journal*, 6(1), 1-13. Retrieved from <http://escholarship.org/uc/item/3qv811wv>
- Kellerman, S. (1992). 'I see what you mean': The role of kinesic behaviour in listening and implications for foreign and second language learning. *Applied Linguistics*, 13(3), 239-258.
- Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kress, G., & van Leeuwen, T. (1996). *Reading images. The grammar of visual design*. London/New York: Routledge.
- Lazaraton, A. (2004). Gesture and speech in the vocabulary explanations of one ESL teacher: A microanalytic inquiry. *Language Learning*, 54(1), 79-117.
- Lemke, J. L. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin, & R. Veel (Eds.), *Reading science: Critical and functional perspectives on discourses of science* (pp. 87-113). London: Routledge.
- Morell, T. (2004). Interactive lecture discourse for university EFL students. *English for Specific Purposes*, 23(3), 325-338.

- New London Group (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60-92.
- Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. London: Routledge.
- O'Halloran, K. L. (2011). Multimodal discourse analysis. In K. Hyland, & B. Paltridge (Eds.), *Companion to discourse analysis* (pp. 120-137). London: Continuum.
- O'Halloran, K. L. (Ed.) (2004). *Multimodal discourse analysis: Systemic functional perspectives*. London: Continuum.
- O'Halloran, K. L., Tan, S., & Smith, B. A. (in press). Multimodal approaches to English for academic purposes. In K. Hyland, & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes*. London: Routledge.
- Partington, A., & Taylor, C. (2010). *Persuasion in politics. A textbook*. Milano: LED.
- Poggi, I., & Pelachaud, C. (2008). Persuasive gestures and the expressivity of ECAs. In I. Wachsmuth, M. Lenzen, & G. Knoblich (Eds.), *Embodied communication in humans and machines* (pp. 391-424). Oxford: Oxford University Press.
- Prior, P., Hengst, J., Roozen, K., & Shipka, J. (2006). 'I'll be the sun': From reported speech to semiotic remediation practices. *Text & Talk*, 26(6), 733-766.
- Quaglio, P. (2009). *Television dialogue. The sit-com Friends vs. natural conversation*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Querol-Julián, M. (2011). *Evaluation in discussion sessions of conference paper presentation: A multimodal approach*. Saarbrücken: LAP Lambert Academic Publishing.
- Rose, K. (1997). Pragmatics in the classroom: Theoretical concerns and practical possibilities. In L. F. Bouton, & Y. Kachru (Eds.), *Pragmatics and language learning*, Vol. 8 (pp. 267-295). Urbana, IL: University of Illinois at Urbana-Champaign.
- Royce, T. (2002). Multimodality in the TESOL classroom: Exploring visual-verbal synergy. *TESOL Quarterly*, 36, 191-205.
- Scollon, R. (2001). *Mediated discourse: The nexus of practice*. London: Routledge.
- Scollon, R., & Levine, P. (Eds.) (2004). *Multimodal discourse analysis as the confluence of discourse and technology*. Washington, DC: Georgetown University Press.
- Silverstein, M. (2004). "Cultural" concepts and the language-culture nexus. *Current Anthropology*, 45(5), 621-652.
- Street, B., Pahl, K., & Rowsell, J. (2011). Multimodality and new literacy studies. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 191-200). London: Routledge.
- Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55, 661-699.
- Takaesu, A. (2013). TED Talks as an extensive listening resource for EAP students. *Language Education in Asia*, 4, 150-162.
- Taylor, C. (1999). Look who's talking. An analysis of film dialogue as a variety of spoken discourse. In L. Lombardo, L. Haarman, J. Morley, & C. Taylor (Eds.), *Massed Medias. Linguistic tools for interpreting media discourse* (pp. 247-278). Milano: Led.
- Taylor, C. (2006). The translation of regional variety in the films of Ken Loach. In N. Armstrong, & F. M. Federici (Eds.), *Translating Voices, Translating Regions* (pp. 47-52). Roma: Aracne.
- Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *Australian Journal of Language and Literacy*, 33(3), 211-223.

- Weinberg, A., Fukawa-Connelly, T., & Wiesner, E. (2013). Instructor gestures in proof-based mathematics lectures. In M. Martinez, & A. Castro Superfine (Eds.), *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (p. 1119). Chicago: University of Illinois at Chicago.
- Wildfeuer, J. (2013). *Film discourse interpretation: Towards a new paradigm for multimodal film analysis*. New York: Routledge.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*. Retrieved from http://www.lrec-conf.org/proceedings/lrec2006/pdf/153_pdf

BELINDA CRAWFORD CAMICIOTTOLI is Associate Professor of English Language and Linguistics at the University of Pisa. Her research focuses mainly on interpersonal, pragmatic and multimodal features of discourse found in both academic and professional settings, with particular reference to corpus-assisted methodologies. She has published in leading international journals, including *Intercultural Pragmatics*, *Discourse & Communication*, *English for Specific Purposes* and *Business and Professional Communication Quarterly*. She recently published the book *Rhetoric in financial discourse. A linguistic analysis of ICT-mediated disclosure genres* (Rodopi) and co-edited with Inmaculada Fortanet-Gómez the volume *Multimodal analysis in academic settings: From research to teaching* (Routledge).

VERONICA BONSIGNORI holds a PhD in English Linguistics (2007). She teaches English Language and Linguistics at the University of Pisa where she has a scholarship as a temporary researcher. Her interests are in the fields of pragmatics, sociolinguistics and audiovisual translation. She has published several articles, focusing on the transposition of linguistic varieties in Italian dubbing (2012, Peter Lang), linguistic phenomena of orality in English filmic speech in comparison to Italian dubbing, as in the paper co-authored with Silvia Bruti and Silvia Masi (2012, Rodopi). She authored the monograph *English tags: A close-up on film language, dubbing and conversation* (2013, Cambridge Scholars Publishing).