

## Multimodality in discussion sessions: corpus compilation and pedagogical use<sup>1</sup>

Mercedes Querol-Julián

[querolm@ang.uji.es](mailto:querolm@ang.uji.es)

Universitat Jaume I, Spain

### ABSTRACT

Discussion sessions of conference paper presentations are spontaneous and unpredictable, in contrast to the prepared lecture that precedes them. These can be challenging, especially for novice presenters whose worst fear is to fail to understand the second meaning of a question or comment, and who know it is not only the quality of the research that is judged but also their prestige and worth. Additionally, spoken academic genres have traditionally been explored by focusing on the transcription of speech and disregarding the multimodal nature of spoken discourse. This study offers a comprehensive account of the design of a multimodal corpus of discussion sessions, where audio, video, transcriptions and annotations are time-synchronised. This multilayer analysis provides examples (not only of linguistic utterances of rhetorical moves and multimodal evaluation, but also of how they are actually expressed paralinguistically and kinetically), which can be used in the classroom and to design learning-teaching materials.

**Keywords:** *English for Academic Purposes, discussion sessions, multimodal corpora, multilayer annotation, research-based pedagogical materials*

### I. INTRODUCTION

The study of academic spoken research genres has received the attention of scholars in the last decade. They have focused primarily on conference paper presentations (Ventola et al. 2002) and particularly on lectures, where the outcomes of the research are presented. To date, however, discussion sessions (hereafter DSs) that follow lectures, and that round off conference paper presentations (CPs), have not received much attention. However, it is in this face-to-face forum that the scientific community can question, criticize and praise, or share knowledge and experience with presenters, who have to know how to respond and react to discussants' comments and questions in a clear and effective way. Therefore, DSs are inherently evaluative as proven by Wulff et al. (2009). These scholars identify considerable differences between the language used in the lecture and in the discussion session, which is characterised by patterns of evaluative language.

Discourse analysis of academic spoken research genres has in general adopted the traditional exploration of written genres, paying attention almost exclusively (Hood and Forey's (2005) work is one exception) to the transcription of speech. However, the complex multimodal nature of spoken discourse cannot be captured in a verbatim transcription of audio recordings; sometimes analysts also make prosodic or phonetic transcriptions and take notes of contextual aspects. Spoken discourse can roughly be described as the co-expression of verbal modes and non-verbal modes; hence, verbatim transcriptions and even transcriptions of paralinguistic (prosodic or phonetic) are only a partial representation of the original event (Thompson 2005). The process to register spoken data can be more problematic when we want to capture non-verbal features, such as the visual. Video recording of the events allows the analyst to explore verbal-visual (visible bodily motion, kinesics) or multimodal functions of linguistic patterns. Therefore, the analysis of speech events cannot be performed on the same basis as written discourse since they use different modes of expression. The difficulty arises because oral communication is multimodal, it is embodied and combines both verbal and non-verbal elements (Adolph and Carter 2007). In addition, most of the work on kinesics, and on paralinguistic, is done on conversation analysis, an area of interpersonal interaction widely explored by scholars who generally belong to multidisciplinary backgrounds such as anthropology, psychology, psychiatry, and sociolinguistics. Gesture is one of the kinesic features that has received most attention. The most influential approaches to the study of gesture are those by Efron (1941), Ekman and Friesen (1969), Kendon (2004) and McNeill (1992). These works see gesture as an activity of major importance to the understanding of the speaker's speech, which has a significant social meaning.

This paper is part of a study that aimed at making a cross-disciplinary analysis of the presenter's expression of evaluation in the DSs of two CPs in Linguistics and Chemistry. I set out to investigate evaluation in spoken academic discourse beyond the traditional linguistic approach. Thus, a multimodal approach, drawn mainly from conversation analysis studies, was followed to foreground **KINESICS** and **PARALANGUAGE** that CO-OCCUR with the **LINGUISTIC EXPRESSION OF EVALUATION**. The theoretical framework of the study, in which the design of the corpus was underpinned, was embedded in techniques of genre analysis (Swales 1990) and

discourse analysis, including the theoretical orientations of systemic functional linguistics (Halliday 1985), conversation analysis (Schegloff and Sack 1973), pragmatics (Brown and Levinson 1987), and multimodal discourse analysis (Kress and van Leeuwen 2001). Conversely, corpus linguistic techniques enabled me to make the application of the multimodal approach feasible. I used computer techniques for automated analytical procedures and qualitative techniques for the interpretation of the corpora. More precisely, I collected a video corpus, took part in the process of transcription, and annotated it. I used the multilayer annotation tool to time synchronise transcriptions (verbatim or orthographic, paralinguistic, and kinesic) and annotations (semantic evaluation and generic moves). Without this tool, it would not have been feasible to analyse evaluation on the comprehensive multimodal level as was done in the study. Nonetheless, a qualitative interpretation of the data was necessary to foreground the salient features that define evaluation in DSs.

The interpretation of findings and the multilayer annotation enabled me to see the potential of this material for pedagogical purposes. The multimodal annotated corpus that I introduce in this paper can provide real examples of the rhetorical moves in which the interaction is organised to express specific communicative purposes, and the linguistic and multimodal expression of evaluation that articulates the rhetoric of the interaction. These multimodal instances can be retrieved to be used in the classroom and in the design of learning-teaching materials. Students will be provided not only isolated linguistic utterances but also how these are expressed during the interaction enabling them to identify changes in paralinguistic features and kinesic features (gesture, head movement, facial expression, and gaze). This would be a significant contribution to the virtually non-existent pedagogical materials based on multimodal corpora research to learn-teach academic spoken genres. Currently, there is only one work (Ruiz-Madrid and Querol-Julián 2008) that devotes a few activities to discussion sessions, which design was based on the study of natural language from a multimodal approach.

The paper is structured in three sections. First, the design of the corpus is presented. I describe the data and give a detailed account of the steps followed to get the corpus ready for the analysis. Then, I suggest some pedagogical applications of the multimodal corpus in the design of activities and the use of the corpus in the classroom.

## II. CORPUS DESIGN

The corpus was designed and compiled within the framework of a major project, the compilation of the *Multimodal Academic and Spoken language Corpus* (MASC) (Fortanet-Gómez and Querol-Julián 2010). MASC is a multidisciplinary collection of Spanish and English spoken academic events at university (i.e. lectures, seminars, guest lectures, students' presentations, dissertation defences, plenary lectures, and conference paper presentations), collected by the research group GRAPE ([Group for Research on Academic and Professional English](#)) at the Universitat Jaume I. The multimodal nature of MASC is given by the five different types of data, gathered during the video recording of the events: slides, transcripts, handouts, and video and/or audio recordings

There are several aspects that need to be considered when designing a spoken corpus, such as the size, variety of language, level of proficiency, text types, and genre among others (Campoy and Luzón 2007). Prioritizing one aspect over another depends on the purpose of the research that is going to be conducted on the corpus. Hence, the aim of the analysis determines the compilation of the corpus, how the corpus is collected, transcribed, and annotated. The criteria followed in the design of the corpus used in the study were based on the main objective of MASC, the multimodal discourse analysis of academic spoken genres (the criteria will be described below). Additionally, a cross-disciplinary approach was adopted in the study which has also determined the design of the corpus.

In this respect, a contrastive study should compare items that are comparable; to put it in other words, the two corpora of Linguistics and Chemistry should have similarities to make the comparison possible. A close look to the factors that may influence the rhetoric and the performance (linguistically and non-linguistically) of the DSs of CPs might help to shed light on the *tertium comparationis* of the two corpora. I have identified six different aspects that may affect **INTERPERSONAL MEANING** in discussion and therefore might influence in the expression of evaluation: the purpose of the conference, the relationship among the participants, cultural and personal features, environmental factors, others' turns, and the discipline. These factors, however, do not operate individually but function as a whole. First, the **PURPOSE OF THE CONFERENCES** was to create a site for bringing together specialists in a field of research to share

investigation results and to open a forum for discussion. In the discussion sessions, as well as in the lectures, the major concerns of the speakers in both conferences were to present their views and to persuade the audience of the relevance and value of their research. Concerning the RELATIONSHIP AMONG THE PARTICIPANTS, both were small focused conferences, with no parallel sessions; thus, the audience size was similar in all the presentations, around 50 people. Small conferences may help presenters to establish a good rapport with the audience. Some participants in the conference in Linguistics, as well as the organisers of the conference in Chemistry were interviewed to find out the relationship between the participants and its possible influence on the discussion sessions. They maintained that most of the participants already knew each other before the conference, as they were international communities of experts with specific and common research interests. The use of first names to address them can linguistically confirm this affirmation. They also note that the DS in CPs could be considered the most stressful stage. The main reason they gave was that after presenting their research, presenters are fully exposed to an audience of experts (in these conferences most of them were senior researchers), who during approximately 20 minutes have been evaluating the presentation and comparing it with their previous knowledge and experience. Presenters should be ready to respond tricky questions and challenging comments, obviously easy questions and nice comments do not pose major problems; but the difficulty lies in the uncertainty of the audience reaction. In view of this, the relationship among the participants can play a crucial role to create a relaxed atmosphere for discussion. The main characters of the discussion are the presenter and the discussant; consequently, the relationship between them would be the most influential one to formulate their questions, comments, and responses. However, the discussion opened between them is not an isolated exchange. The relationship that the presenter and the discussant have with the rest of the participants may also constrain their performance. Of major interest to the contrastive study, however, is that the informants argued that the rhetoric and performance of the discussion did not differ from those adopted in other conferences on the same academic discipline.

So far, I have shown that the purpose of the meetings and the relationship among the participants of these specialised conferences seem to be the same. However, there are other factors that may influence these comparable corpora of DSs which are variables

rather than constants. In this respect, CULTURAL AND PERSONAL FEATURES may affect discussants' questions and comments, and presenters' responses. However, I am neither a biographer nor interested in adopting an ethnographic approach to go into what could be a fascinating analysis. My final objective in the study was to find out a new methodology of analysis from a multimodal perspective; that is the reason why I primarily focused on the linguistic and non-linguistic features of the speech, not putting much emphasis on the cultural and personal backgrounds of the speakers. On the other hand, DSs are organised around a dialogic exchange structure where discussant's and presenter's turns follow each other or overlap. Certainly, the OTHERS' TURN, its meaning and how it is performed, will constrain the response to the questions and comments. This is the way the discussion is constructed. Turns are central in the exchange structure, since it is by turn taking that participants take part in the discussion. Nonetheless, as stated above, the factors that may affect discussion do not do it individually but their spheres of influence overlap. How others' turns are performed depends on the rest of the factors already noted: the purpose of the conference, the relationship among the participants, cultural and personal features, ENVIRONMENTAL FACTORS (such as problems with microphones), and the discipline. Regarding the DISCIPLINE, cross-disciplinary differences have been a common topic of analysis from different perspectives in the studies of evaluation in academic written genres (Hyland 2000, 2004). As regards spoken academic genres, whereas a considerable number of studies have focused on the description and interpretation of a genre in a particular discipline (Flowerdew 1992, Olsen and Huckin 1991), not much work has been done to bring to the fore neither differences between two or more disciplines nor disciplinary differences concerning evaluation. An exception is the work of Poos and Simpson (2002) who explore the use of hedging in a corpus of academic spoken English. These scholars found disciplinary differences; however, neither attention has been paid yet to evaluation in discussion sessions of conference paper presentations, nor a multimodal approach has been adopted to the study of this interpersonal meaning in academic spoken genres.

The *tertium comparationis* of the two corpora is essential to conduct a scientific contrastive study. Nonetheless, although the factors discussed above might influence in the expression of evaluation, they are beyond the corpus designer's control, since they

are inherent to the event and the people that take part in it. There are other aspects, however, that can be controlled in the design of comparable corpora such as corpus size. The size of the present corpus has been determined by the approach adopted in the analysis, the multilayered exploration of evaluation. This type of analysis requires small corpora that enable to carry out a qualitative examination. The purpose of the study was to describe evaluation in both disciplines, rather than to make generalisations of linguistic and non-linguistic patterns, where a larger corpus would be required.

## **II.1. Corpus description**

As noted above, two corpora of CPs, lectures and discussion sessions, of two different academic disciplines were collected for the study. The Chemistry conference brought together leading scientists from all over the world, where a total of 36 papers were presented across a range of areas on the science of isotopes. Conversely, all contributions to the Linguistics conference, 24 in total, dealt with the topics of genre analysis and discourse analysis. Participants were international experts in the field of applied linguistics. For the investigation, however, only the discussion sessions were of interest, thus a subcorpus of ten DSs from each conference was selected. Two criteria were considered in the selection of these DSs. The first criterion was the number of presenters. Only one speaker should have presented the paper, and thus he or she should be the only one responsible for responding the audience's questions and comments. A preliminary analysis showed that when there is more than one presenter, speakers share responsibilities; in the sense that, presenters can give and seek for their colleague's support and even negotiate who is going to respond, using verbal and non-verbal language. Thus, turn-taking organization and rhetoric would be more complex. It is not only the interpersonal meaning between presenter and discussant/s that would come into play, but also the interpersonal meaning between presenters. The second criterion adopted in the selection was the number of turns. A turn is counted when a participant in the discussion (chair, presenter, or discussant) takes the floor. This criterion can give a tentative idea of the level of interaction in the discussion, which should be as similar as possible in both disciplines. Eventually, the Linguistics DSs corpus consists of nearly 12,000 words, 71 minutes, and 39 dialogic exchanges. Whereas, the Chemistry DSs

corpus amounts to nearly a total of 8,500 words, 59 minutes, and 34 dialogic exchanges. The analysis of the corpus of DSs was done at the macrostructure level. This analysis revealed the identification of patterns of dialogic exchanges in the two disciplines. Accordingly, two sub-corpora of dialogic exchanges were selected for the study of evaluation and the generic structure (moves). Sinclair et al. (1972) define *exchange* as the basic unit of the interaction, because it consists of the contribution of at least two participants. In the study, I have followed this definition and categorised what I have called **DIALOGIC EXCHANGES**. These types of exchanges refer to the dialogue held between discussant and presenter to make comments and questions, and to respond to them. The definition of this type of exchanges is necessary to distinguish them from other types of interaction where participants aim at organising the discussion rather than at engaging in a dialogue. Additionally, the concept of **DIALOGIC PATTERN** is used to go beyond the concept of adjacency pair postulated by Schegloff and Sacks (1973), where a question is followed by an answer, to embrace more complex structures; for example, discussant's comment is followed by a question which is responded by presenter, rather than the adjacency pair question – response.

The criterion followed for the selection of the dialogic exchanges that form the sub-corpora was to share similar dialogic patterns. Results show that only 4 and 3 dialogic exchange patterns were recurrent in Linguistics and Chemistry respectively, and only those performed in two turns were common in both disciplines: *Comment – Comment*, *Question – Response*, and *Comment + Question – Response*. On the other hand, it is worth noting that these three patterns are the most frequent “openers” of longer exchange patterns in the corpora with more than two turns. These data prove that participants in the discussion sessions in the small corpora analysed commonly follow these three dialogic exchange patterns (63% of the exchanges in Linguistics and 71% in Chemistry) to open discussion. The sub-corpora of dialogic exchanges were constituted by four exchanges of each pattern from each discipline. The sub-corpora of Linguistics was formed by a total of about 2300 words and 15 minutes, and around 2000 words and 14.30 minutes shaped the one of Chemistry.



## **II.2. Getting the corpus ready**

The corpora were compiled in three stages: data collection, transcription, and annotation. The several types of transcriptions and annotations were done in the following order: first, a verbatim transcription of the corpus of CPs (lectures and DSs); then the annotation of the generic structure (moves) and the semantic evaluation of the corpus of dialogic exchanges of DSs; and finally, the transcription of kinesic and paralinguistic features that co-express with the semantic evaluation already annotated. In following sections, I give an account of the process of collecting, transcribing, and annotating data; as well as of the multilayer annotation of the corpus. Figure 1, in the next page, gives a synoptic view of the design of the corpus that makes possible to carry out a multimodal approach for the exploration of evaluation in DSs, which is described throughout the section.

### ***II.2.1. Collecting the data***

The first stage in the compilation of a corpus is the collection of the data. However, there is a previous stage before collecting the data. We need presenters to give their permission to be video recorded. As commented, the corpus is part of a major project MASC. The procedure we follow to collect the data in MASC is first to contact the organisers of the events. In many cases, the organisers give us the go-ahead to email the speakers. But it can also happen that the organisers become mediators. In both cases, we write a formal email explaining the project they are going to be involved in. We only tape those speakers who give a positive reply to our request. In addition, the data are initially compiled for research purposes; however, participants also sign a consent form when part of the data is going to be published.

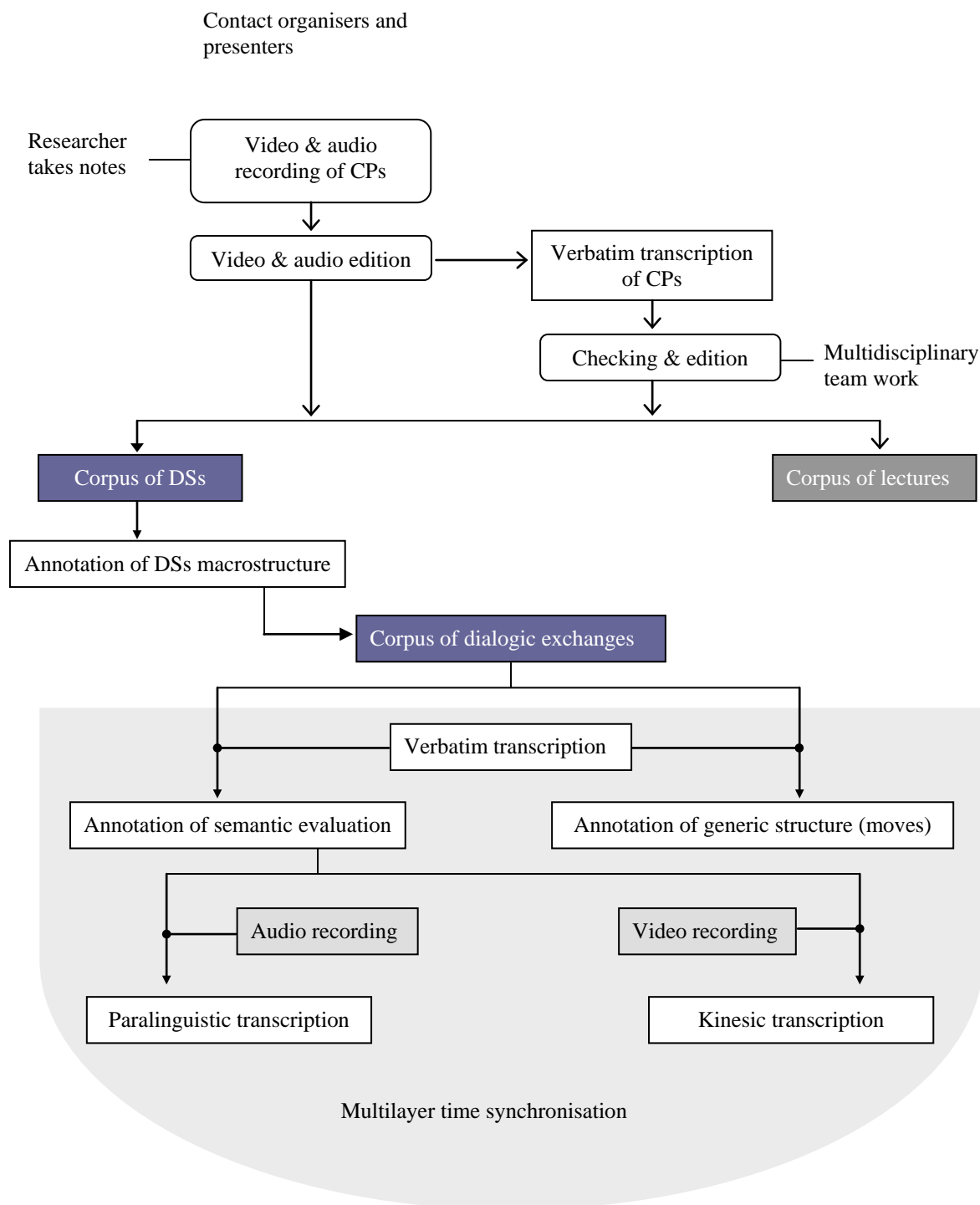


Figure 1. Design of a multimodal corpus of DSs.

For the present study, the original corpus (lectures and discussion sessions) was video recorded and the organisers of both conferences played the role of mediators. However, sometimes the use of go-betweens entails a risk. An example of the difficulties that may

appear when researchers do not contact directly with the speakers is what happened in the Chemistry conference. The organisers informed us that we only had permission to tape 11 out of the 36 presentations and discussion sessions; however, when the conference was over some of the speakers complained about not having been video recorded. A major obstacle to compile data for a multidisciplinary specialised academic spoken corpus is to have access to other areas of knowledge different from ours, since neither the organisers nor the participants in the event are familiar with the methodology we use. In those cases, it is essential that once the organisers green-light our project we try to personally contact speakers to avoid misunderstandings.

Several aspects should be taken into account before and during the recording to guarantee the quality of the data. Special mention should be made of those aspects related to the physical context and the speakers' performance. Before setting up the camera one should consider the size of the room, as well as the distribution of tables, computer/OHP, aisles, window/s and door/s. On the one hand, the intrusion of the camera should cause as little trouble as possible to the presenters in the sense that, they should not feel threatened by it, otherwise their behaviour could change. The smaller the room, the more difficult it is to create a comfortable environment and at the same time focus on the speaker. Moreover, the camera should neither prevent the audience from seeing the speaker, nor distract them from the presentation and discussion. On the other hand, a video recording can become a valuable source of data for the analysis, and for the design of pedagogical materials, if the quality of the image and the sound is good. Light conditions are essential for the quality of the image, an aspect that has to be negotiated with the organisers of the event beforehand. Regarding the sound, external microphones may help to improve it. The speakers' performance should also be taken into account when setting up the camera to be able to focus on them all the time. Presenters may be sitting or standing up, but they can also move around. Accordingly, it is a matter of extreme importance to be careful in this issue, otherwise we could lose relevant data for a multimodal approach.

The conference paper presentations that shape the data for the study were video recorded with a mini-DV digital video camera and an external unidirectional microphone plugged in the camera. One of the advantages of unidirectional microphones is that they seem to reduce ambient noise and to capture the sound of the

image that is in focus. In the corpus, presenters were in focus during the presentations and the discussion sessions. In the conference on Linguistics we were able to use two cameras which also allowed us to record the audience. This is an important difference in the data collection that has determined that only the presenters' performance should be the centre of the contrastive analysis. The external microphones helped to get an acceptable sound quality of the presenters' speech. However, the sound quality of the discussants was lower, which sometimes made the transcription hard. In the Chemistry conference, it was so because although the camera was set up in the middle of the room, among the discussants, the presenter was the one always in focus. In the Linguistics conference, the second camera was set up at the front of the room to focus on the audience; however, the quality of the audio recordings of those discussants sitting at the back was also reduced. Regarding the image, quality was good in the Linguistics conference, but in the conference on Chemistry it was a bit dark because, during the presentation and discussion session, lights were off on behalf of an excellent slide show and only light coming in from back windows illuminated the room. Light condition was a fruitless negotiation with the organisers of the conference. Unfortunately, this reduced the quality of the video recordings which will affect the analysis of kinesics, particularly of face expression and gaze. In addition, in Linguistics during few seconds in four exchanges the presenter was not on focus. These problems can be attributed principally to the inexperience of collecting a multimodal corpus at that time, that was the first contribution to the MASC, and therefore we were not so sensitive to those particular aspects of the recording and the consequences for this type of research.

The next step in the collection of data is the edition of the recordings. I used the video editing software Avid Liquid 7.0 to create .avi files. This format allowed me to manipulate the data creating audio files (.wav) to improve quality with the audio editor available in the program. In addition, after the analysis of the macrostructure of the DSs, I created the sub-corpora of dialogic exchanges making audio and video clips from the original recordings of the whole events. The format of these clips enabled me to export them to the multimodal annotation tool.

The collection of data involved the audio and video recording, but also the collection of contextual information. We observed how the paper presentations and the discussion sessions were performed and made a register in a form during the observation about

different aspects such as the type of event and communicative act (e.g. title, field of knowledge, duration), the speaker (academic status, nationality, mother tongue, age, and sex), the room (type of room and we sketch the distribution of participants, recording devices, furniture and props), the audience (type and number), the speaker resources (PPP, OHP, handouts, microphones, etc), the speaker/s' performance (mode of presentation (if explaining, reading or both) and posture adopted (if moving, sitting or standing up), the discussion (if there is discussion or not, when (during or after the presentation) and audience's turns (number, language, and sex), the recording (time and equipment), and any incident that occurs during the communicative act. The observation aims at fulfilling aspects that one cannot capture with the camera or the microphone and may help to understand the communicative act.

### ***II.2.2. Transcribing the data***

Once the audio and video recordings were edited, the next step was to transcribe what was said, that is, to create a verbatim transcription. The transcription was done for the corpus of CPs (lectures and discussion sessions) in a collaborative work between the GRAPE and the English Language Institute (ELI), at the University of Michigan. Transcriptions followed the established MICASE conventions, where some contextual data were also represented (i.e. XML tags and symbols were utilized to annotate potentially relevant features like speaker identity, speaker turns, speech overlap, laughter, backchannels and pauses<sup>2</sup>). Transcribers were native speakers of English who were previously trained. The process was implemented by checking and editing the transcriptions, a task that was accomplished by a multidisciplinary team since the help of an expert in the field was necessary to check the Chemistry transcripts. The transcripts of the conference in Linguistics were transferred to the ELI and gathered in a single corpus which was named *John Swales Conference Corpus (JSCC)*, a project that aims at complementing MICASE. As MICASE, transcripts of JSCC are also publicly available at the ELI corpora website<sup>3</sup>.

The other two types of transcriptions, kinesic and paralinguistic, were exclusively done for the analysis of evaluation in the corpora of dialogic exchanges when linguistic evaluation is expressed. Therefore, it was done after the orthographic transcription and

the annotation of semantic evaluation. Changes in kinesics and paralanguage that co-occur with semantic evaluation were identified and data were registered in the corpus with the help of the multimodal annotation tool ELAN (see detailed description in Section II.2.4.).

The scope of analysis of kinesics covered changes of: ARMS AND HANDS GESTURES, FACIAL EXPRESSION, GAZE DIRECTION, and HEAD MOVEMENT. Transcription of kinesics was a laborious job since the identification of the co-expression with linguistic evaluation was only possible by slowing down the video recording repeatedly to reveal any change, any micro expression (Ekman and Friesen 1969), not only of the face but of any of the kinesic aspects considered in the study, that are not observable in normal examinations. For example, in one of the exchanges in Linguistics the presenter used the expression “how it’s often taught” in her response to a discussant’s question, where the evaluative adverb “often” co-expressed with a kinesic feature of raising eyebrows that lasted 114 milliseconds. That would be difficult to capture without the annotator program. In Chemistry, it was not always possible to determine the exact direction of eye gaze. As a result, assumptions had to be made on body and head orientation. On the other hand, the transcription of gestures was made broadly, in the sense that in the study I was not interested in the gestures themselves, but in how they co-expressed with evaluative semantics. For this reason, I did not use an accurate identification of the three phases of prototypical gestures, i.e. preparation, stroke, and retraction<sup>4</sup> (Kendon 1980). Nonetheless, a preliminary study showed preparation and stroke commonly co-occur with linguistic evaluation.

Regarding paralanguage, as the starting point of the analysis was semantic evaluation, its examination was limited to changes in the pronunciation of discrete words. This approach narrowed the transcription to changes in the speaker’s VOICE QUALITY, i.e. LOUDNESS, and VOICE QUALIFIER, i.e. SYLLABIC DURATION (after Poyatos 2002). The identification of LOUDNESS was done by the comparison with the surroundings. Sound waveforms available in ELAN were essential at this stage, since waveforms reach the highest peaks when loudness goes up and the lowest peaks when it gets down. Figure 2 shows a sample of identification of loudness-up in ELAN of a fraction of clip in Chemistry, where the maximum amplitude of the waveform of the evaluative word *problems* corroborates the phonetic perception of the stressed noun.

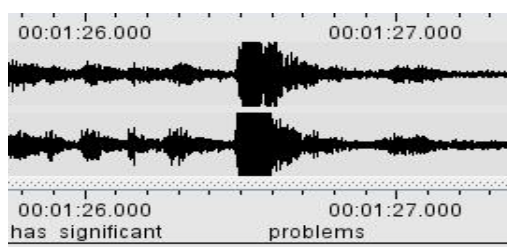


Figure 2. Sample view of identification of paralinguistic voice quality.

As for **VOICE QUALIFIER**, changes in the SYLLABIC DURATION refer to whether the word is pronounced faster or slower than expected in the discourse, that is, in comparison with the pronunciation of surrounding words. Figure 3 shows a sample of identification of long syllabic duration of a portion of a Linguistics exchange. By comparing duration of the evaluative utterance *tends to be more broad*, it can be observed that the adjective *broad* is attributed with the paralinguistic feature of long duration. Whereas the verb *tends to be* is pronounced in 582 ms and *more* in 222 ms; the adjective, despite being a monosyllabic word similar to *more*, lasts 594 ms, a duration even longer than the pronunciation of *tends to be*.

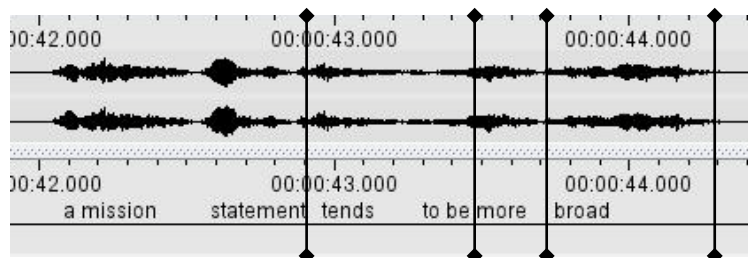


Figure 3. Sample view of identification of paralinguistic voice qualifier.

In addition, I have also included in the analysis the transcription of LAUGHTER, a type of differentiator or of **VOICE QUALIFICATOR**. I have considered the speakers' instances of individual laughter in contrast to episodes of general laughter, because I understand them as the expression of the speakers' attitude towards what they are saying. I cannot obviate the fact that this is a non-linguistic vocal effect which shows emotional reactions. Other paralinguistic aspects, such as intonation, would appear in holistic analysis rather than in the exploration of paralinguistic of discrete items, as done in the study.

### II.2.3. Annotating the data

Annotation differs from transcription in its content. Rather than capturing overtly observable aspects, annotation focuses on more abstract relationships. Annotation, as the collection and the transcription of the data, is determined by the purpose of the study. In view of that, a pragmatic or functional annotation was done on the verbal language to examine the structure of the discussion session and the linguistic evaluation.

Regarding the annotation of the structure, it is important to observe that the analysis conducted was corpus driven. Therefore, all the tags used in the annotation were not pre-selected before the analysis, but drawn from the findings. The macrostructure of the corpus of DSs was annotated to shed some light on the flow of the discussions, to see how turn-taking operates in DSs of specialised CPs. Three different types of tags were used for this aim: the identification of the **PARTICIPANTS** (speaker and addressee), the **TYPE OF TURN** and its **POSITION** in the discussion. All three were assembled in the following string which identifies each of the turns taken and overlapping:

speaker : type of turn \_ position of the turn ~ addressee

Regarding the identification of the **PARTICIPANTS**, even though it has been said that the identity of the speakers was already captured in the verbatim transcription, I have adapted MICASE conventions to identify the role the participants play in the interaction<sup>5</sup>. That is, rather than identifying the participant by the order they speak (S1, S2, etc.), I identified them by the primary role they play as: CHAIR (CH), PRESENTER (P), DISCUSSANT (D), or AUDIENCE (AUD). Besides, discussants were also assigned a number that shows the order in which they speak. I maintained unknown speaker/s (SU) and two or more speakers (SS) tags. Moreover, the name used for the tag was *participants* rather than *speakers* (as in the MICASE) since I aimed at identifying a further functional level, if they were speakers or addressees. As regards the **TYPE OF TURN**, the function that each turn had in the DS was tagged as: COMMENT (C), QUESTION (Q), and RESPONSE (R). The third tag identifies the **POSITION OF THE DISCOURSAL TURN** in the discussion. The dialogue between discussant and presenter can occur in two turns or in several turns. In order to trace the complexity of the sequence it has been annotated when the discussant's and presenter's turn **STARTS** the exchange (S), or when



it is a FOLLOW-UP turn (FU). Follow-up turns have also been numbered. When there is not follow-up, only start turns were tagged even though they started and finished the exchange.

The following example, taken from a Linguistics dialogic exchange, illustrates how the exchange between the first discussant in the DS and presenter was annotated in the corpus. The discussant formulates a question to the presenter to start her turn <D1:Q\_S~P> and the presenter responds <P:R\_S~D1>. However, the discussant does not consider the interaction is finished after the presenter's response and goes on with a follow-up question <D1:Q\_FU1~P> which is also responded by the presenter <P:R\_FU1~D1>, with first attempt in overlap and then in his turn.

D1:Q\_S~P: um, (were these others) that worked in these (fields) were guest editors or were they all the official editors

P:R\_S~D1: um, both both kinds. uh um and the\_ in in linguistic and in meds- in medical uh journals yes

D1:Q\_FU1~P: cuz i just wondered if they might get kind of a different, um, well different kind of type of editorial from a guest editor, who doesn't usually get the floor <P:R\_FU1~D1><OVERLAP> absolutely, mm </OVERLAP> and might use the opportunity to say things uh\_ you know, put forward their views and...

P:R\_FU1~D1: yep, yep. certainly, there's lot of variation from one journal to another, so that they seem to have their <SU-m><OVERLAP> in-house style </OVERLAP> in-house customs and perceptions of the genre, but also according to the the author. [...]

The annotation of the corpus of DSs allowed to identify, among other aspects, the sequence of the dialogues held in the exchanges (i.e. a question is followed by a response, a comment is followed by a comment and the like). This analysis has determined the selection of the recurrent patterns of the dialogues that make up the sub-corpora of dialogic exchanges to conduct the analysis of evaluation. The two sub-corpora (of Linguistics and Chemistry) were also functionally annotated in terms of the moves that shape the dialogic patterns and also in terms of linguistic evaluation. The generic structure of the exchanges was annotated to confirm the hypothesis that it is evaluation, both linguistic and non-linguistic, that articulates it. The tags used to mark the moves were also driven by the corpus. Conversely, the annotation of linguistic evaluation follows an abridged version of the appraisal model postulated by Martin and White (2005). I considered it interesting for the cross-disciplinary study to tag whether

the **SEMANTIC RESOURCES** expresses one or more than one of the three domains of evaluation in the model: ATTITUDE, ENGAGEMENT, and GRADUATION<sup>6</sup>.

In the next section I describe how these annotations and the transcriptions were incorporated to the corpus to carry out the analysis. Before moving to the description of how the multimodal annotated corpus was created, I would like to note the importance of tagging not only by the examination of the verbatim transcription but, even at this stage, by the consideration of the whole performance, that is, audio and video recordings. The multimodal approach might help the analyst to make a more accurate interpretation of the original event, closer to reality. It is important to bear in mind that, in the interaction, participants interpret their interlocutors' speech on the basis of what they hear, the content and the way it is said (that is, linguistics and paralanguage), and what they see (kinesics, visual aids, and any physical interaction with the surroundings). I consider thus, that the study of certain aspects of interpersonal meaning in spoken discourse (like those examined in the study), which were based exclusively on the analysis of verbatim transcripts could cause analysis inaccuracy, because a significant part of the modes of expression that speakers use are disregarded.

#### ***II.2.4. Creating a multimodal annotated corpus***

As described in previous sections, the study conducted with the corpus analysed the data from two approaches. First, I focused on the macrostructure of DSs from a top-down approach. At this level, the analysis was conducted on the corpus of DSs. Then, I explored moves and multimodal evaluation in the subcorpora of exchanges. The examination of moves similarly followed a top-down approach, but the exploration of multimodal evaluation followed a bottom-up approach. At this level of analysis the use of a multimodal annotation tool made the work easier, since it was necessary to time-synchronise the different levels of transcriptions (verbatim or orthographic, kinesic, and paralinguistic), annotations (moves and evaluative semantics), and the audio and video data. I used the multimodal annotation tool ELAN<sup>7</sup> (EUDICO Linguistic Annotator) (Wittenburg et al. 2006) to accomplish this task. This tool enabled me to create as many layers or tiers (as the program calls them) as needed for the different types of transcriptions and annotations. I use ten tiers in this corpus: two for verbatim

transcriptions (discussant's and presenter's), two for linguistic evaluation (discussant's and presenter's), one for moves, one for paralinguistic, and four for kinesics (gesture, head movement, gaze, and facial expression).

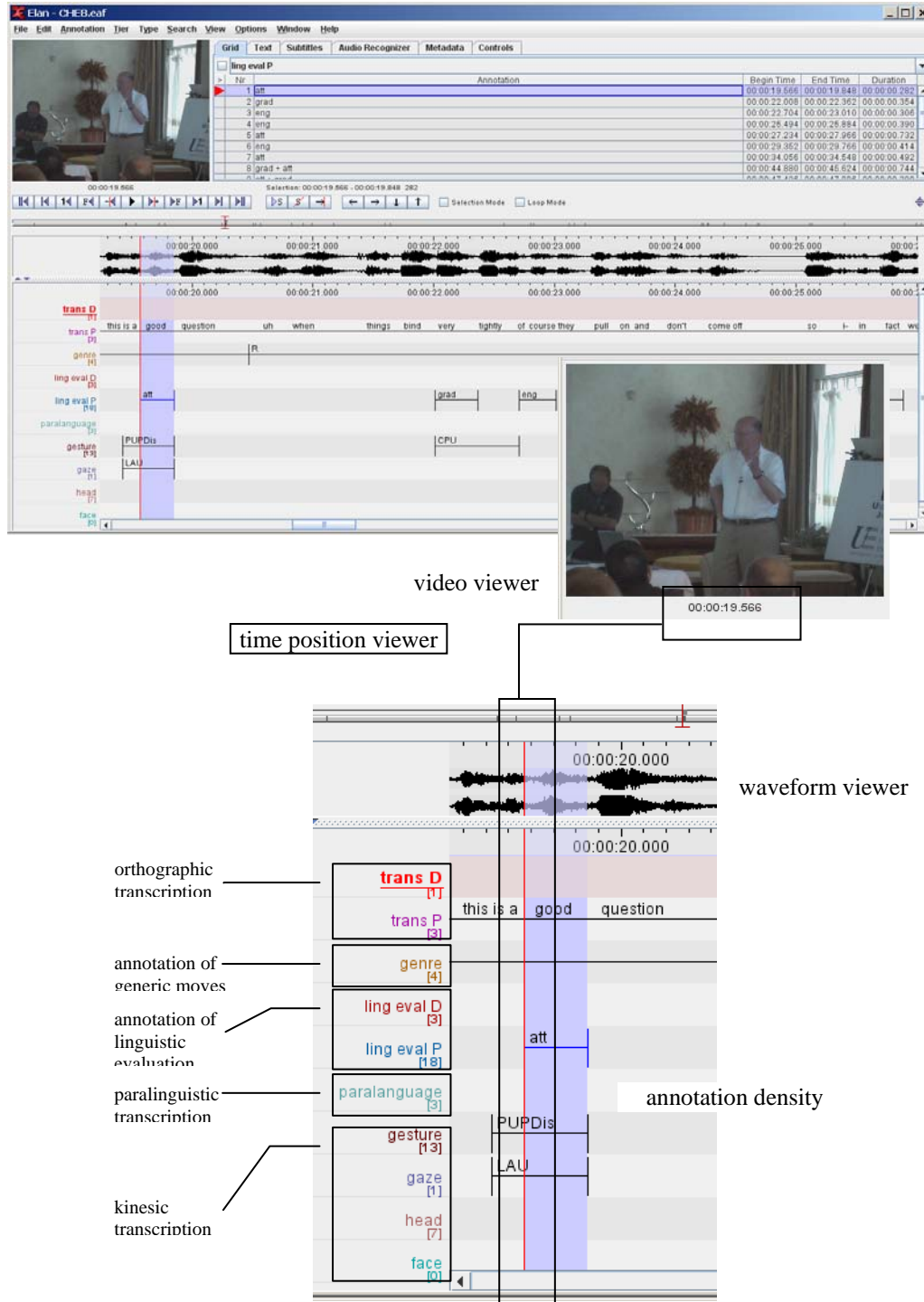


Figure 4. Sample view of multimodal annotation in ELAN.

Figure 4 shows a sample of multimodal annotation view in ELAN of a portion of a

Chemistry exchange. I have enlarged in the figure the four viewers that work in ELAN: video, waveform, annotation density, and time position. All viewers are synchronised and thus displayed at the same point(s) in time. The first stage was to introduce the plain verbatim transcriptions and synchronise them with audio and video data. Sound waveforms were a useful aid at this point. Then, I annotated moves and linguistic evaluation of presenter and discussant. Finally, the transcriptions of kinesics and paralinguistics were done on the grounds of the semantic evaluation. Once all the data were introduced, I could start the analysis with the aid of a search tool also available in the program. Manual extraction of data was necessary in the qualitative approach of the study.

### III. PEDAGOGICAL APPLICATIONS

As noted, the compilation of the corpus described in the previous section was done to study presenters' multimodal expression of evaluation in DSs of two academic disciplines. However, although the results of the study can find applications in English for Academic Purposes courses that focus on communicative skills, the multimodal annotated corpus itself can also be used as a pedagogical tool in the classroom, and as a valuable source of instances to create teaching and learning material to understand this academic research genre and the interpersonal feature that characterises it. In this section, I make some suggestions of the pedagogical potential of the annotated corpus, which due to the newness of the research I have not yet had the opportunity to put it in practice.

ELAN offers many possibilities to retrieve multimodal data, which can be used in the classroom or in the design of activities. There are two ways to access the annotated corpus. The focus could be on the analysis of a single dialogic exchange and all the aspects transcribed and annotated in it. That is, it could be interesting to show students instances of:

- semantic evaluation
- semantic evaluation + audio
- semantic evaluation + audio + video
- semantic evaluation and co-expression with kinesic and/or paralinguistic features

- generic moves

Figure 5 illustrates the exploration of a dialogic exchange from Chemistry. You can select from the list of the ten tiers the feature that you are interested in. In the example, I have selected “gesture” as one of the kinesic features. Once the selection is done, you access to a list of all the instances of gestures that co-express with semantic evaluation. In the dialogic exchanges below there are 13 instances. For the annotation, I have used different tags to simplify the reference to the gestures. In the example, I have selected “CPU” that stands for “closing palms up”. A click on it, gesture Nr 2, and one has access to the video, audio, and annotation density where that gesture is performed.

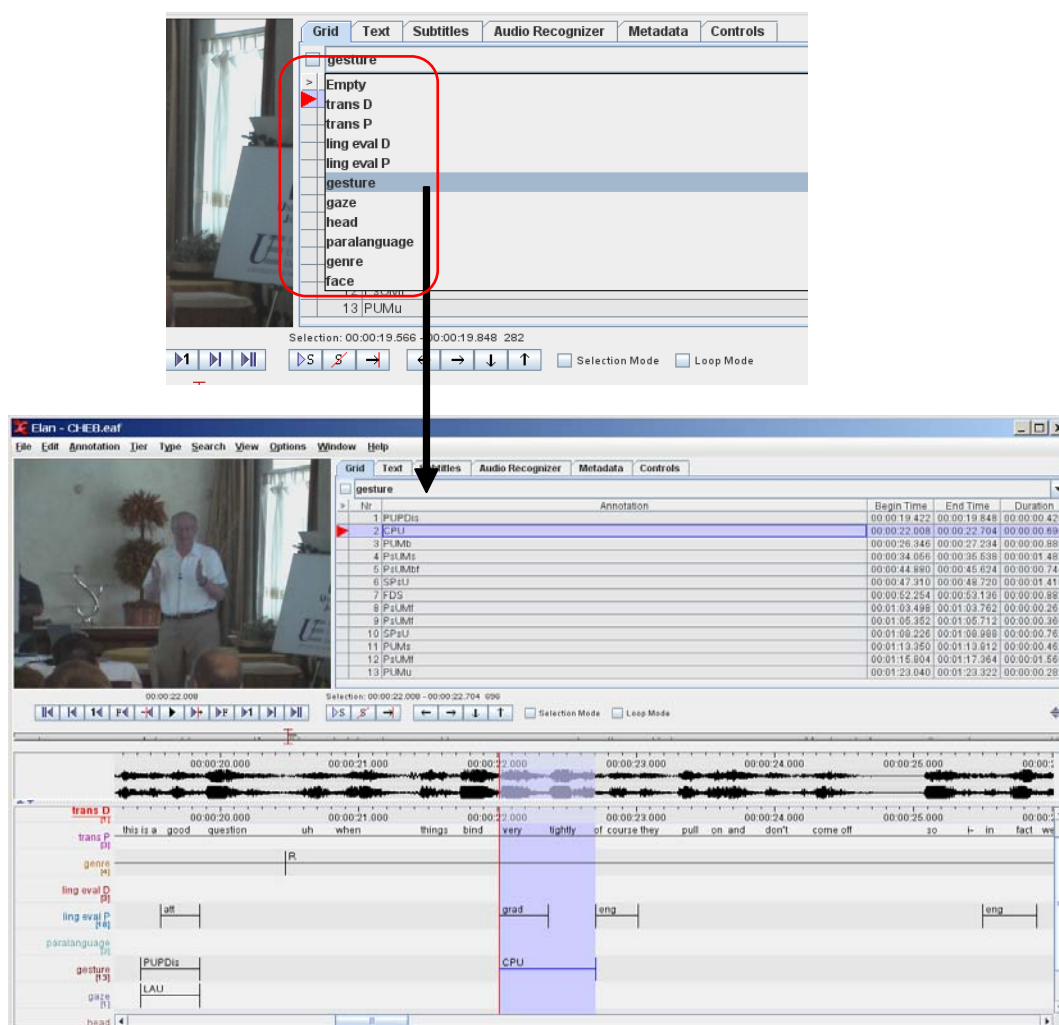


Figure 5. Sample view of the exploration of a dialogic exchange in ELAN.

The other way to retrieve data is using the searching tool. This allows me to focus on

one annotation (this is the general term used in the program, but it embraces both annotations and transcriptions) to find all the instances of it that appear in the corpus. In Figure 6, I illustrate the example of the move “OPT”, “opening the turn” that is used in the two corpora 14 times (6 in Chemistry and 8 in Linguistics). If I click on instance Nr 6, ELAN opens a new window to display the video, audio, and annotation density viewer where this move is expressed in the exchange.

**Search eaf files**

Annotations  
 Search  Search Define search domain Search complete  
 regular expression  case sensitive Export

14 occurrences in 14 annotations in 24 files (0.27 seconds)

Nr	File	Tier	Before	Annotation	After	Begin Time	End Time	Duration
1	CHE1	move	TB	OPT	CC	00:00:16.908	00:00:18.402	00:00:01.494
2	CHE10	genre	TA	OPT	TB	00:00:02.976	00:00:04.056	00:00:01.080
3	CHE10	genre	TA	OPT	CQ	00:00:10.560	00:00:13.080	00:00:02.520
4	CHE12	genre		OPT	CQ	00:00:02.424	00:00:05.958	00:00:03.534
5	CHE8	genre	FQ	OPT	R	00:00:18.200	00:00:20.468	00:00:02.268
6	CHE9	genre		OPT	CQ	00:00:00.520	00:00:02.448	00:00:01.928
7	LE1	genre		OPT	CC	00:00:03.040	00:00:10.340	00:00:07.300
8	LE10	genre	FQ	OPT	R	00:00:38.124	00:00:49.140	00:00:11.016
9	LE11	genre		OPT	CQ	00:00:00.492	00:00:02.256	00:00:01.764
10	LE4	genre		OPT	CC	00:00:01.207	00:00:02.413	00:00:01.206
11	LE7	genre	TA	OPT	FQ	00:00:03.480	00:00:04.242	00:00:00.762
12	LE8	genre	RFQ	OPT	R	00:00:22.326	00:00:24.102	00:00:01.776
13	LE9	genre	CQ	OPT	CQ	00:00:03.114	00:00:03.984	00:00:00.870
14	LE9	genre	FQ	OPT	R	00:00:20.358	00:00:23.484	00:00:03.126

**Elan - CHE8.eaf**

File Edit Annotation Tier Type Search View Options Window Help

Grid Text Subtitles Audio Recognizer Metadata Controls

Annotation

Nr	Annotation	Begin Time	End Time	Duration
1	FQ	00:00:02.976	00:00:18.096	00:00:15.120
2	OPT	00:00:18.200	00:00:20.468	00:00:02.268
3	R	00:00:20.468	00:01:22.992	00:01:02.524
4	CT	00:01:22.992	00:01:24.312	00:00:01.320

Selection: 00:00:18.200 - 00:00:20.468 2268

0:18.000 00:00:19.000 00:00:20.000 00:00:21.000 00:00:22.000 00:00:23.000 00:00:24.000

trans D (1) uh that so this is a good question uh when things bind very tightly of course they pull on and don't come off

trans P (3)

genre (4) OPT R

ling eval D (3)

ling eval P (3)

paralanguage (3)

gesture (3)

gaze (1)

head (1)

face (3)

att grad eng CPU LAU

Figure 6. Sample view of searching an annotation in ELAN.

The potential of these small corpora is significant. To mention a few data, 521 evaluative utterances have been annotated (373 expressed by presenters and 188 by

discussants) where the identification of the three appraisal categories has been done (attitude, graduation, and engagement). In addition, 276 kinesic features and 56 paralinguistic features were co-expressed with presenters' semantic evaluation and transcribed. Regarding the generic structure, 90 moves were annotated.

In this paper, I have described the aspects that need to be considered when compiling an interactive spoken academic genre for the study of evaluation. As proven, the use of multimodal corpora represents a major breakthrough in the field of corpus linguistics and academic spoken discourse analysis; since, taking into account the multimodal nature of oral communication provides a more comprehensive picture of the events. The corpus linguistics techniques used here open a new line of research to explore academic spoken discourse and to provide multimodal material for teaching and learning English for Academic Purposes.

## Notes

<sup>1</sup> The work described in this paper was supported by Universitat Jaume I (Grant CONT/2010/08).

<sup>2</sup> For a detailed documentation of the MICASE transcription conventions, cf. the MICASE manual at <http://micase.elicorpora.info/micase-statistics-and-transcription-conventions/micase-transcription-and-mark-up-convent> > 6 November 2010.

<sup>3</sup> <<http://www.elicorpora.info/>> 6 November 2010.

<sup>4</sup> The phase of the movement that is closer to the apex, the main part of the gesture, is called *stroke*. The phase of movement leading to the stroke is named the *preparation*. And the phase of movement that follows the stroke is referred to as the *recovery* or *retraction*.

<sup>5</sup> MICASE transcription conventions identify speakers as: speaker IDs assigned in the order they first speak (S1, S2, etc); unknown speaker, without and with gender identified (SU); probable but not definite identity of speaker (SU-1); two or more speakers, in unison (SS).

<sup>6</sup> The attitudinal system has to do with 'evaluating'. Engagement has to do with the negotiation of other voices in the text apart from the authorial voice. The third dimension in the appraisal model is graduation. A distinctive feature of attitudes is that they can be gradable.

<sup>7</sup> <<http://www.lat-mpi.eu/tools/elan/>> 6 November 2010.

## REFERENCES

- Adolph, S. and Carter, R.** 2007. "Beyond the word. New challenges in analysing corpora of spoken English". *European Journal of English Studies*, 11 (2), 133-146.
- Brown, P. and Levinson, S.** 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Campoy, M. C. and Luzón, M. J.** (Eds.). 2007. *Spoken Corpora in Applied Linguistics*. Bern: Peter Lang.
- Efron, D.** 1941. *Gesture and Environment*. Morningside Heights: King's Crow Press.
- Ekman, P. and Friesen, W.V.** 1969. "The repertoire of nonverbal behavioral categories: Origins, usage, and coding". *Semiotica*, 1, 49-98.
- Flowerdew, J.** 1992. "The language of definitions in science lectures". *Applied Linguistics*, 13, 202-221.
- Fortanet-Gómez, I. and Querol-Julián, M.** 2010. "The video corpus as a multimodal tool for teaching". In Campoy, M. C., B. Bellés and Ll. Gea (Eds.) *Corpus-based Approaches to English Language Teaching Corpus and Discourse*. London & New York: Continuum, 261-270.
- Halliday, M.A.K.** 1985. *An Introduction to Functional Grammar*. London: Arnold.
- Hood, S. and Forey, G.** 2005. "Introducing a conference paper: Getting interpersonal with your audience". *Journal of English for Academic Purposes*, 4, 291-306.
- Hyland, K.** 2000. *Disciplinary Discourses: Social Interactions in Academic Writing*. London: Longman.
- Hyland, K.** 2004. "Engagements and disciplinarity: The other side of evaluation". In Del Lungo Camiciotty, G. and E. Tognini Bonelli (Eds.). *Academic Discourse. New Insights into Evaluation*. Bern: Peter Lang, 13-30.
- Kendon, A.** 1980. Gesticulation and speech: Two aspects of the process of utterance. In Key, M. (Ed.). *The Relationship of Verbal and Non-verbal Communication*. The Hague: Mouton, 207-227.
- Kendon, A.** 2004. *Gesture. Visible Action as Utterance*. Cambridge: Cambridge



University Press.

**Kress, G. and van Leeuwen, T.** 2001. *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Edward Arnold.

**Martin, J.R. and White, P.** 2005. *The Language of Evaluation: Appraisal in English*. London: Palgrave Macmillan.

**McNeill, D.** 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago & London: The University of Chicago Press.

**Olsen, L. and Huckin, T.** 1991. "Pint-driven understanding in engineering lecture comprehension". *English for Specific Purposes*, 9, 33-47.

**Poos, D. and Simpson, R.C.** 2002. "Cross-disciplinary comparisons of hedging: some findings from the Michigan Corpus of Academic Spoken English". In Reppen, R., S. Fitzmaurice and D. Biber (Eds.). *Using Corpora to Explore Linguistic Variation*. Philadelphia: John Benjamins, 3–21.

**Poyatos, F.** 2002. *Nonverbal Communication across Disciplines. Volume II. Paralanguage, Kinesics, Silence, Personal and Environmental Interaction*. Amsterdam: John Benjamins.

**Ruiz-Madrid, N. and Querol-Julián, M.** 2008. GRAPE Online Activities for Academic English. 6 November 2010 <<http://www.grape.uji.es/activities/pagina%201/index.html>>

**Schegloff, E.A. and Sacks, H.** 1973. "Opening up closings". *Semiotica*, 8, 289-327.

**Sinclair, J., Forsyth, I.M., Coulhard, R.M. and Ashby, M.** 1972. *The English Use of Teachers and Pupils*. Final report to SSRC. University of Birmingham.

**Swales, J.M.** 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

**Thompson, P.** 2005. "Spoken language corpora". In Wynne, M. (Ed.). *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, 59-70. 6 November 2010 <<http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm>>

**Ventola, E., Shalom, C. and Thomson, S.** 2002. (Eds.) *The Language of*

*Conferencing*. Frankfurt: Peter Lang.

**Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. and Sloetjes, H.** 2006.

“ELAN: A professional framework for multimodality research”. *Proceedings of Language Resources and Evaluation Conference*. 6 November 2010  
<<http://www.mpi.nl/publications/escidoc-60436/@@popup>>

**Wulff, S., Swales, J.M. and Keller, K.** 2009. “‘We have seven minutes for questions’:

The discussion sessions from a specialized conference”. *English for Specific Purposes*, 28, 79-92.

*Received September 2010*

*Cite this article as:*

**Querol-Julián, M.** 2010. “Multimodality in discussion sessions: corpus compilation and pedagogical use”. *Language Value*, 2 (1), 1-26. Jaume I University ePress: Castelló, Spain. <http://www.e-revistas.uji.es/languagevalue>.

**ISSN 1989-7103**

**Articles are copyrighted by their respective authors**