RiCL Research in Corpus Linguistics

A corpus-assisted genre analysis of the *Tunisian Lecture Corpus*: An exploratory study

Basma Bouziri
University of Gabés / Tunisia

Abstract –Multimodal, specialized corpora of academic lectures represent authentic classroom data that practitioners can draw on to design academic listening resources that would help students attend lectures. These corpora can also act as reflective practice corpora for teacher training or professional development programs with the objective of raising awareness of lecturing practices. Despite their contribution in shaping the type and quality of the learning that takes place in classrooms, multimodal lecture corpora are scarce, particularly in the Arab world. This paper addresses this research gap by designing and collecting a corpus of academic lectures delivered in English in Tunisia. The corpus was explored using a Systemic Functional Linguistics and English for Specific Purposes integrated genre analysis framework. A three-layered model of analysis was used to manually code various rhetorical functions as well as their realizations. Major findings include the pervasiveness of metadiscursive functions when compared to discourse functions, the identification of context-specific metadiscursive strategies, and the absence of verbal or non-verbal signaling of some rhetorical functions. Implications relate to the necessity of compiling and/or using lecture corpora that are multimodal, the value of adopting function-first approaches to explore these, particularly in non-native contexts, and the design of professional development programs and learning materials that would better account for local academic needs.

Keywords – Academic lectures; corpus; genre analysis; Systemic Functional Linguistics; English for Specific Purposes; Tunisia; exploratory studies

1. RATIONALE

In Tunisia, increasing attention and efforts are being devoted to quality pedagogy and teacher training in higher education. To contribute to these efforts, the study of the rhetorical features of lecturers' discourse and the way they are realized is a necessary step. Such research would lead to designing needs and context-specific courses and materials that would support students when attending lectures. It would also play a role in the design of teacher professional development programs with the aim of upgrading the quality of teaching in higher education. To study lectures, the use of specialized

corpora is pivotal. Nesi (2008: 1–2) maintains that "as teachers of languages for specific purposes, it is these small specialized corpora that interest us most." The main reason is that the analyst can integrate macro-contextual features that are essential for a sound interpretation of the corpus data (Flowerdew 2004; Camiciottoli 2008). In doing so, explanatory adequacy supplements the descriptive power that corpora already have (Bhatia 2002). The validity of the analysis and interpretation of the corpus data within its context of use is further accentuated when the compilers of specialized corpora are themselves the analysts.

In spite of the value that locally designed and specialized corpora have, they remain scarce. In the Tunisian context, some research has been conducted on classroom discourse (Abdesslem 1987; Touati 2004). However, to our knowledge, there is no study which has been carried out on academic lecture discourse in Tunisia except for Bouziri (2019). One major reason is that compiling spoken data is a daunting and expensive task, which is coupled with the sensitivity of the data under focus as well as with the current requirements for multimodal data. In fact, the use of muted spoken corpora (Ballier and Martin 2015), that is, transcripts of spoken language which are not distributed with audio and video files, does not account for the most basic and immediate context behind their production and their use fails to deliver a valid analysis and interpretation of the data (Deroey and Taverniers 2011). A second reason is the nature of the analyses that are conducted on academic lectures. Since pedagogical applications often represent their ultimate aim, genre analyses of lectures are particularly valuable. In focusing on macro discourse structures and functions, genre approaches draw one of the most significant links between corpora and contexts (Partington 2004) reflecting principled variations not often captured by large-scale corpus studies, since their interest is in the texts and contexts that generated the corpus data (Flowerdew 2013). Additionally, corpus-assisted genre analyses hold the potential for accounting for less marked discourse phenomena.

Although some macro approaches to lecture discourse analysis have been conducted in different contexts (cf. Young 1994; Alsop and Nesi 2014; Bouziri 2019), they are not as widespread as more form-based types of analyses. Along with data transcription, the manual annotation of lectures is, in fact, time-consuming and cognitively demanding because multiple viewings of the lectures and readings of the transcripts are necessary to identify and describe various rhetorical categories. These

difficulties are heightened due to the idiosyncratic nature of the lecture genre. In fact, when lecturing, "the lecturer is not under great pressure to exhibit control over conventionalized rhetorical structure" (Thompson 1994: 182). This is unlike the research article genre, for example, where researchers are pressured into adhering to international standards of writing for their work to be published in international journals (Abdesslem and Costello 2018). One reason for the idiosyncratic nature of lectures is that they are live events. This necessarily involves a certain degree of spontaneity and, hence, unpredictability. Another reason is the lecturers' individual lecturing styles which also account for the high variability that often characterizes this genre along with class size (Lee 2009; Cheng 2012), discipline (Young 1990; Thompson 1994; Deroey and Taverniers 2011), and culture (Alsop and Nesi 2014). All these variables sustain the rhetorical variation that academic lectures exhibit and thus makes their study challenging.

2. THE STUDY

To address the aforementioned research gaps and challenges, the *Tunisian Lecture Corpus* (TLC) project was started up. In this paper, I report on and discuss its collection, transcription, and coding with the objective of providing tools for the study of lectures in under-investigated contexts.[1] To this aim, I propose a theoretical framework and a model of genre analysis that are compatible with the specificity of the academic lecture genre. The framework and the model were used to approach the corpus and develop the coding scheme employed for the manual annotation of the various rhetorical functions in TLC. Accordingly, the subsequent sections are organized as follows. Section 3 presents and discusses the theoretical framework of the study: an integrated genre analysis framework that draws on both the Systemic Functional Linguistics (SFL) and English for Specific Purposes (ESP) traditions. In this section, the model of analysis is also presented. Section 4 provides a description of TLC with information about its collection, the participants, the corpus transcription, and its coding and analysis. The results of the corpus-driven study are then presented in Section 5. Section 6 summarizes the findings and discusses their implications.

[1]The transcription conventions and coding scheme developed and used in this study are available in the IRIS database following this link: https://www.iris-database.org/iris/app/home/detail?id=york:938327

3. THEORETICAL FRAMEWORK

Genre Analysis (GA) is a discourse approach that concentrates on both the linguistic and contextual aspects of texts in specific genres. The analysis is 'top-down' (Biber *et al.* 2007) starting from the rhetorical functions and moving down to their linguistic, non-linguistic, and/or multimodal realization. Within GA, three research traditions have been distinguished: New Rhetoric (NR), Systemic Functional Linguistics (SFL), and the English for Specific Purposes (ESP) approach. In this paper, I will concentrate on the SFL and ESP approaches for two reasons. The first is that NR adopts a social rather than a pedagogical orientation, considering the classroom as "an inauthentic environment lacking the conditions for complex negotiation and multiple audiences" (Hyland 2002: 114). The second is that New Rhetoricians adopt a non-linguistic approach to GA (Hyon 1996; Flowerdew 2002), investigating texts from an ethnographic rather than a discourse analytic perspective.

As opposed to New Rhetoricians, SFL and ESP genre analysts conduct both functional and linguistic analyses of genres with pedagogical motives in mind. Within these two approaches, functional categories are set as the starting point for the analysis. Subsequently, the linguistic features that characterize them are described. According to Callies (2015), such function-driven approaches are rarely implemented in linguistic research despite the fact that they are valuable particularly for research that is conducted in non-native contexts. One reason is that function-driven approaches enhance our understanding of the way forms correlate with the functions they realize in discourse (Callies 2015). They also seek to uncover 'non-canonical' strategies which non-native users may employ to convey meaning, and which can be easily neglected in form-driven approaches (Callies 2015: 54). For the aforementioned reasons, a GA approach has been adopted to analyze TLC.

3.1. Key notions in genre analysis

In this study, three key notions in GA are used. They are drawn from the ESP and SFL approaches. In combination, they are viewed as complementary and relevant to the analysis of the academic lecture genre. The first two constructs identified in the ESP approach are 'moves' and 'steps'. According to Swales (2004: 228–229), a move is:

> a discoursal or rhetorical unit that performs a coherent communicative function in a written or spoken discourse. […] It is a functional, not a formal, unit. […] Sometimes, however, grammatical features can indicate the type or nature of move.

Thus, a move is defined in terms of the rhetorical goal that it seeks to achieve. It is usually "realized in stages […], all of which are more or less steps to the fulfillment of the function of the move" (Bhatia 2001: 86). A step, in turn, constitutes a specific function within a move and serves its higher purpose. Steps have been referred to elsewhere as 'strategies' (Kwan 2006; Yin 2016) or 'sub-functions' (Thompson 1994) to denote their non-obligatory, cyclical, and non-sequential nature (Lee 2009), in contrast to the way they have been interpreted in Swales' model. With respect to academic lectures, for example, several researchers have analyzed introductions as a sub-genre in terms of moves and steps (Lee 2009) or functions and sub-functions (Thompson 1994; Yaakob 2013). They have found that there are multifunctional units and point out that, at times, there is some difficulty to disentangle functions from one another. Another major finding is the lack of "robust preferred orders" (Swales 1990: 145). Thompson (1994), for instance, found that various functions in lecture introductions display a non-sequential structure. Duszak (1994), in her study of academic Polish texts, also reports that moves behave in a cyclical rather than a linear fashion and that various combinations of moves and steps are possible.

The third construct used in the study is that of 'phase' and is drawn from the SFL approach. A phase is defined as a "strand of discourse that recurs discontinuously throughout a particular language event" (Young 1994: 165). Two types of phases are identified: discourse phases and metadiscourse phases. A discourse phase embodies rhetorical functions such as defining, explaining, and exemplifying, whereas metadiscourse phases have structuring, evaluating, and closing functions. Phase boundaries are identified pragmatically rather than temporally using semantic, verbal, non-verbal, and contextual cues. The construct of phase is very useful because it enables the analysis to be conducted via a unit that does not imply any sequential ordering. Phasal analysis is indeed the most influential model proposed to analyze academic lectures within the SFL tradition (Young 1990; Gregory 2002; Wu 2013). As Young (1990: 45) points out, "it describes what actually happens in an instance of discourse where different strands recur and are interwoven to form the discourse plot of an instantiation of language." This kind of analysis is flexible because it does not enforce a linear or hierarchic structure to texts and captures the dynamic and non-sequential

nature of the academic lecture discourse. Most importantly, Young's work contributed to a systematic and functionally-oriented analysis of discourse segments where a function rather than a form-based analysis is adopted to identify phases. Another advantage of phasal analysis is that it provides a thick and comprehensive analysis of discourse along both macro-elements and micro-elements.

3.2. An SFL-ESP integrated framework

In combining ESP and SFL approaches to GA within a single theoretical framework, a function-driven model is developed. Phasal analysis is viewed as the analytical approach that best captures the dynamic nature of the academic lecture genre as the construct of phase is soft, non-codal, and non-predictive (Gregory 2002). It is soft, since one phase is likely to occur at any point in discourse and is not restricted to a particular temporal unit such as beginning, middle, or end. In fact, "there are many beginnings, many middles, and many ends." (Young 1994: 165). Similarly, phases are non-codal, as they are recognized and labelled in terms of their function (e.g., evaluation, content, and discourse structuring) and not in terms of an implied hierarchical structure (e.g., initiation, response, feedback). Finally, a phase is not predictive because it does not (necessarily) require a particular phase to occur next.

The terms moves and steps are interpreted in this study as functions and sub-functions, for this conceptualization does not suggest or, at least, impose any kind of a hierarchical order. The two constructs represent more fine-grained rhetorical distinctions that are not captured by phasal analysis. Along with phases, functions and sub-functions would be allowed to occur discontinuously and are characterized by recursiveness and a certain degree of unpredictability.

This looser view of genre is taken up in the SFL-ESP integrated framework that I am proposing in this paper. Within this framework, two interrelated theories and methodologies of genre, SFL and ESP, are assimilated within one single model that draws on their mutual strengths. The theoretical basis and the main components of this framework are summarized as follows:

1. Phases, moves, and steps are rhetorical spaces for the enactment of a set of rhetorical purposes.
2. Phases, moves, and steps are non-predictive and are organized along certain preferences or choices. This means that a significant degree of variability in the exploitation of those rhetorical spaces in the lecture genre should be acknowledged.
3. A fine-grained coding of rhetorical functions is compatible with a phasal analysis.

The constructs of phase, move, and step are used to design a corpus-assisted genre analysis model to analyze TLC. This model is described in the following section.

3.3. Model of analysis

The present corpus-assisted model aims to explore TLC in view of identifying a niche that would lead to develop a corpus-based study. This research process is depicted in Figure 1 where the first parse in the analysis is corpus-driven, involving "the inspection of corpus evidence" (Tognini-Bonelli 2001: 84).

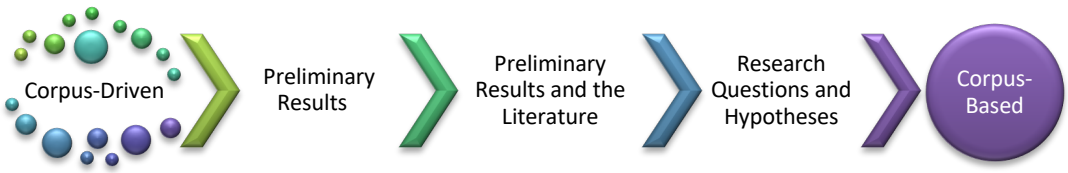


Figure 1: Corpus-driven to corpus-based analyses

The corpus-driven genre analysis set as its first objective the unveiling of rhetorical features that would not be captured by large scale and purely corpus-based techniques (Yaakob 2013). The relationship between phases, moves, and steps in the model is illustrated in Figure 2.

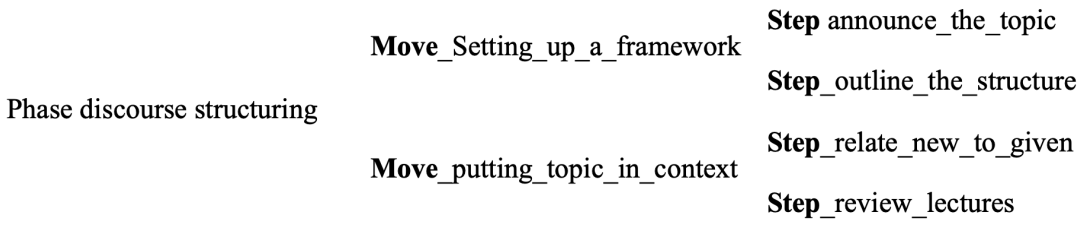


Figure 2: Rhetorical relations in the *Tunisian Lecture Corpus*

A phase is the upper level category with a general rhetorical function such as structuring, describing, and evaluating. It encompasses more specific moves which carry out its general purpose. Within these rhetorical functions, various steps are identified. This is illustrated in Figure 2 where the discourse structuring phase unfolds into two types of moves: setting up a framework and putting topic in context. In turn, these are divided into two steps each. It should be noted that a discourse structuring phase can be enacted at two different rhetorical planes which I shall call here domains: the content domain and the lecture domain. For example, setting up lecture framework refers to structuring the lecture as a whole, whereas setting up content framework refers to structuring a particular content unit or topic within the content phase.

Regarding their realization, the three rhetorical functions may take different forms: a phrase, a clause, or an utterance-(s). Their boundaries are set based on a mixed set of criteria (Swales 2004). The first and main criterion is linguistic where the meaning of words and expressions such as *I mentioned last week*, *today I'm going to* is used to assign a particular rhetorical function to a segment or to help identify boundaries between two units. A second criterion is the use of paralinguistic features (e.g., intonation, stress), and non-verbal features (e.g., gestures), which can help assign or confirm a particular rhetorical function to a segment. Contextual elements may also be employed to further check and resolve any ambiguity, thus representing the third criterion used during the coding process. Contextual clues can be found in the videos of the lectures, for example. Finally, the coder's knowledge of the way texts within a particular genre and discourse community tend to be structured has also an impact on the coding decisions as well (Dudley-Evans 1994).

4. CORPUS AND METHOD

In this section, details about the corpus and the method used for coding and analyzing it are provided.

4.1. Corpus design

TLC is a non-native, specialized, and multimodal corpus of academic lectures comprising over 106,000 words. It is made up of 12 video recordings and one audio recording. Details of the corpus are shown in Table 1 below.

Words in the corpus	106,200
Mean	8.169
Range	7.913
Class size	Average of 20 students
Hours of recording	20 hours and 50 minutes
Video files	12
Audio files	1
Course description	9 out of 12

Table 1: Overview of the *Tunisian Lecture Corpus*

The recorded lectures took place in two institutions of higher education in Tunisia, with lecturers teaching content courses in English in applied linguistics, cultural studies, and literature. Course descriptions were collected when available, as they can enhance the interpretation of the lectures. The names of participants were kept anonymous, and so were the institutions they were affiliated with. 12 participants provided written consent, and one gave oral consent, to make (some of) their data available. For instance, some participants agreed to make all their data publicly available whereas others agreed only for the transcripts to be shared publicly. All participants were Tunisian non-native speakers of English with Arabic as their native language. Their students were also non-native English speakers and were mostly Tunisians. English was used mainly to teach English language undergraduate and graduate degrees. English-mediated instruction is carried out in some public and private universities for subjects such as business and engineering.

The higher education landscape in Tunisia has also witnessed a shift towards more interactivity in academic lectures. In this regard, several courses now combine the lecture format with the seminar format rather than use the traditional monologic lecture format. The lecture format is adopted to teach theoretical content whereas the seminar format provides a space where students can apply that theory to tasks such as text analysis, oral presentations, or linguistics exercises. Some of the data collected in this study thus reflects a hybrid genre that might not be smilar to data collected in other contexts.

4.2. Data collection

The data was collected during the academic year 2014–2015. Lecturers of undergraduate content courses in four institutions of higher education were contacted. Clarifications and details concerning ambiguous points as well as technical aspects of the recording (e.g., the placement of the video camera and the setting up of the microphone) were discussed. Only 13 lecturers from two institutions granted consent. Recording sessions were then arranged with each one of them during their regular classroom hours. The equipment was tested in a real classroom situation in order to evaluate the quality of the image and sound, comfort in wearing the microphone, the degree of intrusiveness of the mounted camera in class, the students' reactions, and the actual battery and recording capacities. Participants were then recorded for two lectures in a row. One reason was to examine how two different lectures were connected. A second reason was that the first lecture of each participant could provide useful information for the interpretation of the second one which was under investigation. The video camera was mounted on a tripod at the back of the class and angled in order to capture the full front frame including the lecturer, the board, and the whole class with students sitting with their backs to the camera. A background information sheet on each lecture session was also filled. Variables gathered included gender, teaching experience, and language background.

4.3. Data transcription

The transcription system devised is described in Appendix 1. A low-level transcription requirement was opted for and as many relevant contextual elements as possible were

included. To balance quality and speed, *Soundscriber*, a transcription tool, was used to walk through the files. To operate this tool, audio files in the wave format were extracted from the video files. In addition to *Soundscriber*, two other windows were used during actual transcription. The first is the video file which was used to check pauses, contextual events, and the second is a plain text file used for transcription. The triangulation of data sources was effective for a smooth and reliable transcription. The transcription process also underwent three passes. Pass one concentrated on textual and specific spoken features like pauses, fillers, and backchannels. Pass two was carried out with the support of video files and was proven essential for a *bona fide* transcription. Indeed, features such as pauses, turn boundaries, and contextual events could be spotted and/or interpreted more appropriately thanks to visual cues. In the final pass, transcripts were edited for more consistency and transcription mistakes were corrected. It should be noted that grammatical mistakes that lecturers made were kept as they were with corrections added via the tag <error corr>. Disfluencies and mistakes pertaining to clausal and utterance structures were not corrected as this would have changed the data.

One last point pertaining to the transcription of students' turns is worth mentioning. Although the focus in this study is on lecturer discourse, students' turns were fully transcribed whenever possible. These were marked between square brackets in the transcripts because, most of the time, they could not be fully and/or clearly heard to be transcribed in any reliable way, and thus fully exploited for the purposes of the current study. In those cases, the speech act performed by the student was transcribed. For example, if a student responded to a question, the response was transcribed as <response>. If a student asked a question, it was transcribed as <question> and so on. Because of the gaps in students' contributions, these were not included in the total number of words in TLC.

4.4. Coding and analysis

To explore TLC, a preliminary coding scheme was devised based on the literature on lecture genre analysis, and thus included some rhetorical functions derived from it. The *UAM CorpusTool* (O'Donnell 2017) was used to draw the scheme and made it possible to manually code the lectures in terms of (pre-)designed features. As the coding proceeded, new categories emerged and were added to the original scheme. At the end of the process, an upgraded version was generated (see Appendix 2). The scheme makes

some slight, yet important distinctions between some rhetorical categories. For example, dividing a global rhetorical function such as the discourse structuring phase in terms of more specific rhetorical functions as illustrated in Figure 2 (see Section 3.3) highlights the distinction between two types of knowledge (respectively schematic knowledge and contextual knowledge) that the two moves, (*viz.*, setting up a framework and putting topic in context) activate.

The coding procedure was constructed on a small sub-set of the corpus comprising two lectures (<Civ-09-02-A> and <Ling-07-02-B>), which in turn constituted 10% of the corpus. The procedure involved various stages adapted from Biber *et al.*'s (2007: 12–13) model of 'top-down' approaches to corpus analyses. The first stage included a survey of rhetorical functions. Lecture introductions and lecture closings have been the subject of many research projects and, as such, they represented a major input for a number of rhetorical functions that could be initially included in the coding scheme. As for the rhetorical functions in the content phase, many were drawn from lecture research as well as research on other academic genres. The second stage is the warm-up stage in which the video of the lecture to be coded was viewed. This enabled the coder to get a general feel of the lecture, which in turn guided its interpretation. Notes were taken on details which were not fully captured during transcription, but which were thought to aid the coding process. The video of a previously recorded lecture and the course description were also consulted whenever the coder wanted to obtain further contextual information. The third stage of the coding procedure involved segmenting the text into phases, moves, and steps. Phase and move boundaries can sometimes be fuzzy and, therefore, needed to be constantly revised throughout the coding process in order to reach the finest delimitation of these different rhetorical categories. In the fourth stage, the coding scheme was upgraded and the coder then moved to the analysis of pervasive functions in the corpus in view of pinpointing interesting rhetorical and/or linguistic phenomena. Analysis of the coded categories was carried out using the statistics feature of the *UAM CorpusTool*, which enabled to calculate their frequencies. A qualitative analysis of the coded segments was also conducted in order to study the way the various rhetorical categories were realized verbally and non-verbally.

5. RESULTS AND DISCUSSION

The results of the corpus-driven study are organized into four parts. The first provides general findings about the academic lecture genre and the various functions identified in TLC as observed in the corpus-driven study. The remaining parts focus on the metadiscursive functions identified: their pervasiveness, context-specificity, and verbal and non-verbal realizations.

5.1. Overview

The corpus-driven study confirmed the discontinuous and recurrent nature of phases and moves which were already reported in the literature as a key characteristic of the academic lecture genre. The evaluation phase typically illustrates this discontinuity and recursiveness. Additionally, not all functions weighed equally. A case in point is the difference between the functions: summarizing main points and indicating end in the lecture closing phase. A lecture which closes with a formal indication of an end and another which wraps up content bringing together the main and important points discussed do not have the same pedagogical value. Examining lecture closings, Cheng (2012) also found that reviewing key points has the lowest frequency when compared to the other functions in lecture closings. The finding above was possible thanks to the adoption of the SFL-ESP integrated framework which allowed for the coding of fine-grained rhetorical functions (*viz.*, moves and steps) in addition to more general ones (*viz.*, phases). The framework, particularly its integration of the construct of phase, was also useful as it distinguished between discourse phases and metadiscourse phases.

5.2. Pervasiveness of metadiscursive functions

The present corpus-assisted genre analysis led to identifying a number of rhetorical functions as displayed in Table 2. In Appendix 3, examples of each of the functions listed below are also provided.

RETHORICAL FUNCTION	<Civ-09-02-A>	<Ling-07-02-B>
STRUCTURING	**71**	**98**
FRAMING	**32**	**73**
SETTING_UP_LECTURE_FRAMEWORK		
Announce_the_topic	1	3
Outline_the_structure	0	0
Present_aims	0	0
Announce_start_of_lecture	0	0
Looking_ahead	0	0
SETTING_UP_CONTENT_FRAMEWORK		
Announce_content	19	35
Looking_ahead	2	1
REVIEW _CONTENT		
Indicate_end	1	1
Sum_up	5	32
REVIEW_ LECTURE		
Indicate_end	1	0
Housekeeping	1	0
Looking ahead	1	1
CONTEXTUALIZING	**39**	**25**
PUT_LECTURE TOPIC_IN_CONTEXT		
Refer_to_earlier_lectures	6	0
PUT_CONTENT_IN_CONTEXT		
Refer_to_earlier_content	14	6
Provide_rationale/context	19	19
ELABORATING	**44**	**47**
Explain_content	21	19
Exemplify_content	14	14
Specify_content	2	1
Draw_implication	7	13
DEFINING AND DESCRIBING	**51**	**66**
Define_content	16	14
Describe_content	35	52
EVALUATING_ LECTURE_ CONTENT	**1**	**7**
Show_importance_of_content	0	6
Indicate_Scope	1	1
EVALUATING_KNOWLEDGE	**51**	**19**
Enquire_about_ knowledge	10	5
Give_clues	1	1
Establish_knowledge	12	6
Indicate_attitude	5	0
Indicate certainty/uncertainty	3	1
Give_feedback	16	5
Enhance_understanding	3	1

Table 2: Significant rhetorical functions in two lectures in the *Tunisian Lecture Corpus*

A major observation is the frequency of metadiscursive functions when compared to discourse functions. The structuring function, where the lecturers both frame and contextualize their talk, is the most frequent one in the two lectures under study forming respectively 71 and 99 units. Framing and contextualizing enhance comprehension

through the organization of content and the presentation of contextual information that is critical to understand the content at hand. In <Civ-09-02-A>, 44 and 71 units were coded respectively as elaborating and structuring whereas 51 units were coded as defining and describing content. Similarly, <Ling-07-02-B> contains 98 units where the lecturer structures the talk and 47 units where he elaborates on the content introduced. In turn, defining content was coded 66 times. Based on the figures above, it seems that structuring is a key metadiscourse function in TLC. Looking closely at this finding, one can notice that there is a limited amount of lecture framing as reflected in the steps of setting up the lecture framework and reviewing lecture[2] when compared to content framing. This finding may indicate that the structure of lectures may not be obvious for students in order for them to recover the way content is organized in terms of macro and micro points. The pervasiveness of structuring denotes the lecturer's efforts in framing and contextualizing the talk with the aim of achieving coherence between different content units. However, lecture framing is mostly restricted to announcing the topic and indicating end rather than to presenting aims and/or outlining structure. The latter are typically realized in the lecture discourse structuring phases and lecture closing phases. Given the real time conditions under which lectures are delivered, they support and enhance comprehension. It is clear however that these were not fully taken advantage of as rhetorical spaces to carry out important pedagogical functions. In this regard, Palmer-Silveira (2004: 101) states:

> if the introduction is poorly prepared, the audience may lose interest and this can jeopardize the way our students will understand the topic. In the introduction, the audience will need to know the main topic, the purpose, the main concepts we will deal with.

While it is true that content framing contributes to the organization of various points within the lecture, it is nonetheless important for students to have global frames of reference they can resort to during the lecture in order to organize the various information presented. The low frequency of lecture framing has also been reported in the literature on global macro-markers (Thompson 2003; Palmer-Silveira 2004).

Besides structuring, evaluation is a pervasive metadiscursive function. One way in which evaluation unfolds in TLC is through the use of contextual comments as reflected in the function give feedback. Their use was particularly noted in the two lectures under investigation. Contextual comments broadly correspond to commentaries in Vande

[2] Four units each for <Civ-09-02-A> and <Ling-07-02-B>.

Kopple's (1985) and Crismore *et al*'s. (1993) models of metadiscourse, to contextual metadiscourse in Luukka's (1994) and Ädel's (2006) models, and to text parenthetical remarks in Goffman's (1981) descriptive account of the academic lecture. Their objective is "to recommend a mode of procedure or let the audience know what to expect" (Vande Kopple 1985: 85). Contextual comments characterize spoken rather than written discourse, which partly explains why they emerge as particularly interesting to reflect upon in this study.

A manual search of the whole corpus for contextual comments yielded 17 instances. Their realization in TLC is both different from the ones found in the literature on academic lectures in native settings and similar to some of the occurrences in data obtained in contexts similar to the present one. Differences and similarities with those various contexts can be noted when comparing extract (1) with extracts (2) and (3) below.

(1) <l_11> malcolm cowley he said the following words <lecturer dictates>indeed indeed <lecturer dictates> i quote here <lecturer dictates>indeed comma <lecturer dictates>these are one of the **rare** moments where I dictate in fact <lecturer smiles> <students laugh> I am **never** happy I feel **frustrated** all the time but er it is the strategy that we'll we'll opt for and it's for your er</l> <ss> <benefit></s>

<l_11>for your benefit yes er in order to you know try to make you yourselves help in the process of er building this so called course in fact er alright <lecturer dictates></l> <Lit-11-02-A>

(2) I'll move this slide a little bit so you can see better… (Luukka 1994: 80)
You might wish to read the last section first (Crismore *et al.* 1993: 46)

(3) <what I'm speaking is almost English more or less if you neglect the accent the rest should be more or less standard English> (Molino 2018: 946)

Extract (1) illustrates an instance of a contextual comment in TLC where the lecturer justifies his/her use of a dictation strategy. In this extract (as well as in all instances of contextual comments in TLC), contextual comments take the form of a relatively extended talk where the lecturer shares his/her own observations and evaluation of a particular event occurring during the lecture. The lecturer wants to prevent the students from conceiving the course as a mere imparting of information. Indeed, he/she implicitly advises the students against copying verbatim what he/she is dictating during exams. Dictation is negatively viewed in the Tunisian academic context because it is reminiscent of the traditional role of the teacher/lecturer as pouring information to the students' minds, of a view of lectures as information transfer only, and of a perception

of the lecturer as the one who holds knowledge. The adjectives *rare* and *frustrated* as well as the adverb *never* are employed by the lecturer to distance him/herself from this technique. The contextual comment also reflects the novice-expert relationship that typically characterizes the academic lecture genre. In the extracts under (2), the commentaries are relatively short and concern the practical management of the materials used during the lecture. They are in line with Vande Kopple's definition of commentaries cited above. In extract (3), however, a clear resemblance with extract (1) can be noted with respect to the nature and motivation of the contextual comment used. Indeed, the lecturer in this extract is referring to the quality of his/her English as somehow departing from the standard form. Molino (2018) argues that the comment acts as a kind of self-protective strategy and the same argument may also apply for the extract from TLC as the lecturer does not want to be negatively perceived as merely pouring information into students as empty vessels.

A second aspect of evaluation is assessing students' knowledge. In <Ling-07-02-B>, 19 steps were coded as evaluating students' knowledge as compared to 51 in <Civ-09-02-A>. Conversely, there is a low proportion of lecture evaluation by the two lecturers, specifically one for <Civ-09-02-A> and seven for <Ling-07-02-B>. Evaluation of lecture points is reflected in sub-functions such as show importance of content and indicating scope. Extracts (4) and (5) illustrate the functions of evaluating students' knowledge and lecture evaluation, respectively.

(4) <l_07> okay classes are of four hours a day</l>
<presenter> <and six days a week></presenter>
<l_07> and six days a week [.] does this remind you of something in Tunisia?
<ss> <no></ss>
<l_07> primary school [.] primary school especially at the [.] o think <unintelligible token="1"/> the first and the second and third primary school this is what pupils do [.] it's five to six days four hours a day of teaching here okay? [.]</l> <Ling-07-02-B>

(5) <l_09> … so they asked for a bill of rights <lecturer writes on the white board> to be added to the constitution [.] what the federalist did is something else they wrote [.] </lecturer writes on the white board> documents commonly known as the federalist papers [.] those documents were <lecturer writes on the white board> a propaganda of federalism it means they talked about the great advantage advantages of federalism okay? so federalists wrote the federalist papers to support the American constitution the antifederalists asked for demanded it's more than asked they demanded <foreign>exiger</foreign> that a bill of rights to be added to the?</l>
<ss> <responses></ss>

> <l_09> constitution don't forget that okay federalist federalist papers antifederalists a bill of rights [..] clear so far? </l> <Civ-09-02-A>

In extract (4), the background knowledge marker *does this remind you of something in Tunisia?* enquires about the students' background knowledge. In doing so, the purpose is to relate the principles and concepts associated with suggestopedia to the students' own educational context. In extract (5), the use of the importance marker *don't forget that* focuses students' attention on the need to associate two important documents in American history, that is, *the Bill of Rights* and *the federalist papers*, to their respective political parties.

The discrepancy between the rhetorical functions of students' knowledge evaluation and lecture evaluation echoes those of Lee (2009) and Cheng (2012) who found that moves pertaining to lecture evaluation are optional. Similarly, Deroey (2017) found that importance markers are less frequent in authentic lectures than in EAP materials. Lecture evaluation refers to the lecturer's weighing of the points he/she makes in terms of importance or relevance. The highly infrequent units displaying evaluation of lecture points as compared to the pervasiveness of units reflecting evaluation of language may suggest that lecture comprehension is pursued at a more local, lexico-grammatical rather than at a global, discourse level of language. More examples and analysis of evaluation as a metadiscursive function are provided in the following section.

5.3. Context-specific metadiscursive functions

A closer inspection of the metadiscursive functions drawn from the present corpus-driven study led to identifying context-specific metadiscourse strategies that were adopted by the two lecturers. One such strategy is evaluating the students' linguistic knowledge through the use of metalinguistic comments (Ädel 2006). These comments are considered metadiscursive because the lecturer does not expand on the propositional meaning of the utterance. Their use expresses the desire to ensure that the students can process the incoming content. Although Ädel (2006: 109) discusses metalinguistic comments in relation to their use by Swedish students of English as "a filler strategy" and as a strategy that reflects non-native speakers' awareness of the situation of writing, the term 'metalinguistic comment' is adopted in this paper because it bears resemblance to those found in TLC as illustrated in extract (6).

(6) <l_09> yes they meet in a conference committee <lecturer writes on the white board> what's the role of a conference committee? </l>
<s><response></s>
<l_09> to iron out <.> differences <lecturer writes on the white board> <lecturer moves arm from right to left> to iron out to try to find a compromise a middle way solution </l> <Civ-09-02-A>

The lecturer in this extract uses both a gesture and a paraphrase in order to define the verb *to iron out*. In Flowerdew's (1992) taxonomy of definitions, the two strategies are referred to as substitution accompanied by visual support. The realization of the metalinguistic comment in extract (6) demonstrates the role that multimodal corpora play in appreciating the way the lecturer combines verbal and non-verbal strategies in order to assist students in following the lecture. In doing so, the lecturer is aware that vocabulary may hinder the comprehension of content. Hence, he/she strives to anticipate and address this issue before moving on to the delivery of academic content.

Metalinguistic comments have been identified under the labels 'low focus definitions' or 'embedded definitions', "which are not the focal point of the information" (Flowerdew 1992: 209) and are "incidental to the logical structure of the lecture" (Jackson and Bilton 1994: 73).[3] Interestingly, their frequency was reportedly greater in non-native contexts like Oman (Flowerdew 1992; Jackson and Bilton 1994), than in contexts where students' English proficiency is considered very high, like Canada (Lessard-Clouston 2009). In the metadiscourse literature, Vande Kopple's (1985) taxonomy of metadiscourse comprises the category 'glosses' that is similar to metalinguistic comments. However, this category, as realized in TLC, was rarely identified in current analyses of corpora, which may be explained by the relatively few studies conducted on academic lectures in an EFL context, a fact which may have made the marker go unnoticed. Moreover, most metadiscourse research was conducted in native and even Content and Language Integrated Learning contexts where the relationship between lecturers and students is considered to be native-to-native, despite the fact that one or both parties are non-native speakers of English. In those contexts, it is possible that metalinguistic comments as they were identified in TLC were simply not

[3] Please, note that definitions of technical and semi-technical words or phrases are included under the rhetorical function 'define content', which is a discourse rather than a metadiscourse function. For examples, see Appendix 3.

used and hence the analyst could not detect them. The aforementioned findings substantiate the need for a function and context-driven approach to corpus annotation.

A second context-specific strategy found is the use of lecturer evaluation markers. This is an evaluative metadiscursive device where the lecturers assess the students' performance to ensure that they appropriately applied theory and/or concepts introduced at an earlier phase of the lecture to case studies. These markers are metadiscursive, since they reflect the direct expression of stance by the lecturer towards the students' discourse. In extract (7) from <Ling-07-02-B>, the lecture evaluation marker takes the form of extended comments where the lecturer evaluates the way the micro-teaching session was carried out by the student.

(7) <l_07> … okay so you tried the maximum to apply things that are related to the natural approach at especially the first part in a natural approach of course there are other stages that [.] should be included if we have had more time of course but for the start especially for usually beginners we said this is a good strategy to make children of course there is usually a problem a remark that i usually repeat is that you do not adapt yourself to the level you are taught [.] normally you don't know things but you usually answer correct answers … of course you are not er [.] normally students at that level [.] they are learning okay? so they should usually expect the teacher to er help them find the names of things okay? to make the difference between healthy unhealthy [.] …you should adapt yourself to the level you are taught and try to help to the person who is actually teaching to teach okay? </l> <Ling-07-02-B>

A rather positive evaluation is highlighted at the beginning of the extract where the lecturer expressed satisfaction with the way the student applied the principles of the Natural Approach in class. This is conducted through the utterance *you tried the maximum to apply*. A negative evaluation was directed to the other students who were playing the role of pupils at an early stage of their English language learning. The negative evaluation is displayed in the use of the noun problem and the following series of negative grammatical constructions, *viz*. *do not adapt*, and *don't know*. As such, the negative connotations associated with these two language elements can also be exploited to realize evaluation.

The emergence of lecturer evaluation markers is closely connected to the context of the present study. In fact, the Licence-Master-Doctorate reform in Tunisia strongly urged a shift in focus from theory to practice and more classroom space for interactivity. Accordingly, more time was allotted to practicums in most courses within the undergraduate English curriculum at university. Evaluation of the students'

understanding of theory through the implementation of practicums coupled with the pressure for more interactivity in lectures has thus given rise to evaluative metadiscourse as reflected in the use of such markers.

5.4. Signaling of metadiscursive functions

When turning to the linguistic signaling of rhetorical functions, and particularly those which are metadiscursive, a few observations can be made. The lecturer in <Civ-09-02-A>, for instance, did not signal the transition between the lecture structuring phase and the content phase by any means. In some other instances, the relationship between two units was not explicitly signaled, as in extract (8), below, where the cause-effect relationship between the two propositions is implied rather than expounded.

(8) <l_09>so the articles of confederation failed they needed a document</l><Civ-09-02-A>

An explicit realization of the cause-effect relationship would be *so **because** the articles of confederation failed they needed a document*. The absence of signaling, as far as this metadiscursive function is concerned, adds to the value of function-driven analyses of discourse, especially when they are conducted with a pedagogical purpose in mind. Absence of signaling means that students may not notice the relationship between the two propositions, a fact which may affect their comprehension of the ongoing discourse.

In addition to absence of signaling, other issues were found. In <Civ-09-02-A>, the lecture topic was announced non-verbally as the lecturer wrote the lecture title on the board. No verbal iteration accompanied the visual marking of the topic. This finding further highlights the role that multimodal corpora play in data analysis and interpretation. In her study of academic lectures in universities in Spain, Martín del Pozo (2017: 26) made a similar observation commenting that non-native lecturers "need **more overt** [emphasis added] signaling of lecture phases and a wider stylistic variety enabling them to do so." This is important, since part of the lecturers' role is to produce language that develops the language competence of the students not only at the lexico-grammatical level, but also at the discourse and pragmatic levels. Students should indeed be exposed to models that would encourage them to structure their language productions and develop their academic literacy.

6. CONCLUSION

This study reports on the compilation of the *Tunisian Lecture Corpus* and provides tools for the analysis of similar corpora for researchers desiring to explore lectures in their own contexts. One major limitation concerns data collection and the absence of follow-up interviews with the lecturers in this study. This kind of data could have provided useful insights into the participants' own perceptions of their pedagogical and linguistic behaviors. Notwithstanding this limitation, the present research contributed to the provision of theoretical, analytical, and methodological tools that can be used to design and approach a corpus of academic lectures. These are the SFL-ESP integrated framework of genre analysis, a corpus-assisted model of genre analysis, and a coding scheme which can serve as a diagnostic tool to approach a corpus when no research questions or hypotheses have been pre-set.

The study yielded four important findings. The first is the discontinuous and cyclical nature of the rhetorical functions in TLC, which is in line with other findings in the literature. The second is the dominance of the metadiscursive functions in the present corpus, particularly structuring when compared to the discourse functions. A third set of findings relates to the new meanings that some metadiscursive categories acquired when compared to other data. This is the case of contextual comments and metalinguistic comments. Both findings reflect the significant role that context plays in shaping the different meanings and linguistic realizations that rhetorical functions have, as well as the value of researching genres in under-investigated contexts. Finally, there were some issues relating to the linguistic signaling of some metadiscursive functions as reflected in the absence of (verbal) signaling for the topic in one of the lectures investigated (Bouziri forthcoming). Given that metadiscursive functions and their realizations are particularly highlighted in the present corpus-driven study, a corpus-based study has been set up to further investigate the use of metadiscourse in TLC (Bouziri 2019).

Important implications of this study are drawn. Firstly, there is a necessity for designing multimodal corpora of lectures which would account for the use of non-verbal strategies. Indeed, the findings of this study highlighted the way some rhetorical functions were realized non-verbally. Along with verbal strategies, these contribute to fulfill the lecturers' pedagogical goals. Secondly, exploring a corpus using a genre analysis framework means that the corpus needs to be human readable. In this respect,

the present study provides a coding scheme that could be used as a diagnostic tool. It may be tempting to embark in corpus alignment or the preparation of XML files as one way to make the corpus ready for automatic extractions of some linguistic forms via corpus tools. However, this approach may not be effective for two reasons. The first is that such files are difficult to read for the analyst who is interested in the macro-level and manual discourse coding of lectures. As it was the case for this study, it is important to first start by exploring a sub-set of the corpus using a basic transcription system (cf. Appendix 1) and a manual analysis before deciding to invest time in such endeavors. A second reason is that the search for frequent words or multiword expressions may not yield useful results which would capture special features of the lectures. Again, this stems from the highly idiosyncratic nature of the lecture genre.

Thirdly, a corpus-assisted genre analysis is relevant when corpus studies are conducted with pedagogical objectives in mind. This type of analysis becomes even more significant when conducted in non-native contexts and on genres exhibiting a high degree of variability as it is the case of the academic lecture genre. Such a function-driven approach holds the potential of unveiling rhetorical functions worthy of further research and occasionally their non-canonical realizations. Furthermore, adopting a top-down approach in this study led to uncovering functions which were not signaled or were context-specific. Pedagogically, this is important when it comes to designing professional development programs for lecturers whose aim is to raise their awareness of the potential difficulties that their students may encounter when attending their lectures. It is also significant for the design of local academic materials that would integrate such strategies in order to better reflect the type of discourse that students in Tunisia are exposed to. Despite the difficulty of their implementation and their use of small corpora, macro-level discourse approaches to the lecture genre are rewarding when pedagogical applications are the ultimate objective.

REFERENCES

Abdesslem, Habib. 1987. *An Analysis of Foreign Language Lesson Discourse: With special Reference to the Teaching of English in Tunisian Secondary Schools*. Sheffield: The University of Sheffield thesis.

Abdesslem, Habib and Hassan Costello. 2018. Introductions in locally published research articles in linguistics: Towards a syntagmatics of moves. *Arab Journal of Applied Linguistics* 3/1: 5–46.

Ädel, Annelie. 2006. *Metadiscourse in L1 and L2 English.* Amsterdam: John Benjamins.

Alsop, Siân and Hilary Nesi. 2014. The pragmatic annotation of a corpus of academic lectures. *LREC*: 1560–1563.

Ballier, Nicolas and Philippe Martin. 2015. Speech annotation of learner corpora. In Sylviane Granger, Gaëtanelle Gilquin and Fanny Meunier eds., 107–134.

Bhatia, Vijay K. 2001. Analyzing genre: Some conceptual issues. In Marting Hewings ed. *Academic Writing in Context: Implications and Application*. Birmingham: University of Birmingham Press, 79–92.

Bhatia, Vijay K. 2002. A generic view of academic discourse. In John Flowerdew ed., 21–39.

Biber, Douglas, Ulla Connor and Thomas A. Upton. 2007. *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins.

Bouziri, Basma. 2019. *A Corpus-assisted Genre Analysis of the Tunisian Lecture Corpus: Focus on Metadiscourse*. Louvain La Neuve: Université Catholique de Louvain thesis.

Bouziri, Basma (Forthcoming). Topic signaling in the *Tunisian Lecture Corpus*. *ESP Today* 8/2: 2–24.

Callies, Marcus. 2015. Learner corpus methodology. In Sylviane Granger, Gaëtanelle Gilquin and Fanny Meunier eds., 35–56.

Camiciottoli, Belinda. 2008. Interaction in academic lectures vs. written text materials: The case of questions. *Journal of Pragmatics* 40/7: 1216–1231.

Cheng, Stephanie W. 2012. "That's it for today:" Academic lecture closings and the impact of class size. *English for Specific Purposes* 31/4: 234–248.

Crismore, Avon, Raija Markkanen and Margaret S. Steffensen. 1993. Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication* 10/1: 39–71.

Deroey, Katrin L. B. and Miriam Taverniers. 2011. A corpus-based study of lecture functions. *Moderna Språk* 2: 1–22.

Deroey, Katrin L. B. 2017. How representative are EAP listening books of real lectures? In Jenny Kemp ed. *2015 BALEAP Conference Proceedings: EAP in a rapidly Changing Landscape: Issues, Challenges and Solutions*. Reading: Garnet Publishing.

Dudley-Evans, Tony. 1994. Variations in the discourse patterns favored by different disciplines and their pedagogical implications. In John Flowerdew ed., 146–158.

Duszak, Anna. 1994. Academic discourse and intellectual styles. *Journal of Pragmatics* 21/3: 291–313.

Flowerdew, John. 1992. Definitions in science lectures. *Applied Linguistics* 13/2: 202–221.

Flowerdew, John ed. 1994. *Academic Listening: Research Perspective.* Cambridge: Cambridge University Press

Flowerdew, John ed. 2002. *Academic Discouse*. London: Longman.

Flowerdew, John. 2002. Genre in the classroom: A linguistic approach. In Anne Johns ed. *Genre in the Classroom: Multiple Perspectives.* London: Lawrence Erlbaum Associates, 91–120.

Flowerdew, Lynn. 2004. The argument for using English specialized corpora to understand academic and professional language. In Ulla Connor and Thomas A. Upton eds. *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins, 11–33.

Flowerdew, John. 2013. *Discourse in English Language Education*. London: Routledge.
Goffman, Erving. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
Granger, Sylviane, Gaëtanelle Gilquin and Fanny Meunier eds. 2015. *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.
Gregory, Michael. 2002. Phasal analysis within communication linguistics. In Michael Cummings, Peter Fries, David Lockwood and William Spruiell eds. *Relations and Functions within and around Language*. London: Continuum, 316–345.
Hyland, Ken. 2002. Activity and evaluation: Reporting practices in academic writing. In John Flowerdew ed., 115–130.
Hyon, Sunny. 1996. Genre in three traditions: Implications for ESL. *TESOL Quarterly* 30/4: 693–722.
Jackson, Jane and Linda Bilton. 1994. Stylistic variations in science lectures: Teaching vocabulary. *English for Specific Purposes* 13/1: 61–80.
Kwan, Becky S. C. 2006. The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes* 25/1: 30–55.
Lee, Joseph. 2009. Size matters: An exploratory comparison of small and large-class university lecture introductions. *English for Specific Purposes* 28/1: 42–57.
Lessard-Clouston, Michael. 2009. Definitions in theology lectures: Implications for vocabulary learning. *The Asian ESP Journal* 5/1: 7–22.
Luukka, Miina. Riitta. 1994. Metadiscourse in academic texts. In Britt-Louise Gunnarson, Per Linell and Bengt Nordberg eds. *Text and Talk in Professional Contexts*. Uppsala: The Swedish Association of Applied Linguistics.
Martín del Pozo, María Ángeles. 2017. Training teachers for English Medium Instruction: Lessons from research on second language listening comprehension. *Revista de Lingüística y Lenguas Aplicadas* 12/1: 55–63.
Molino, Alessandra. 2018. 'What I'm speaking is almost English…': A corpus-based study of metadiscourse in English Medium Lectures at an Italian university. *Educational Sciences: Theory and Practice* 18/4: 935–956.
Nesi, Hilary. 2008. Corpora and EAP. In *LSP: Interfacing Language with other Realms: Proceedings of the 6th Languages for Specific Purposes International Seminar*. Johor Bahru: Universiti Teknologi Malaysia. https://warwick.ac.uk/fac/soc/al/research/collections/bawe/papers/corpora_and_eap.pdf
O'Donnell, Mike. 2017. *UAM CorpusTool*. Madrid: Universidad Autónoma de Madrid.
Palmer-Silveira, Juan Carlos. 2004. Delivery strategies in classroom lectures: Organising the message. In Pilar Garcés, Reyes Gómez, Lucía Fernández and Manuel Padilla eds. *Current Trends in Intercultural, Cognitive and Social Pragmatics*. Sevilla: Editorial Kronos, 97–115.
Partington, Alan. 2004. Corpus-assisted discourse studies. In Alan Partington, John Morley and Louann Haarman eds. *Corpora and Discourse*. Berlin: Peter Lang, 9–18.
Swales, John. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
Swales, John. 2004. *Research Genres: Explorations and Applications*. Cambridge: Cambridge University Press.
Thompson, Susan. 1994. Frameworks and contexts: A genre-based approach to analyzing lecture introductions. *English for Specific Purposes* 13/2: 171–186.
Thompson, Susan. 2003. Text-structuring metadiscourse, intonation and the signaling of organization in academic lectures. *Journal of English for Academic Purposes* 2/1: 5–20.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
Touati, Walid. 2004. *Native vs. Non-native Treatment of Oral Errors in an EFL Context.* Tunisia*:* Institut Supérieur des Langues de Tunis dissertation.
Vande Kopple, William J. 1985. Some exploratory discourse on metadiscourse. *College Composition and Communication* 36/1: 82–93.
Wu, Shuxuan. 2013. Discourse structure and listening comprehension of English academic lectures. *Theory and Practice in Language Studies* 3/9: 1705–1709.
Yaakob, Salmah. 2013. *A Genre Analysis and Corpus-based Study of University Lecture Introductions*. Birmingham: The University of Birmingham dissertation.
Yin, Bin. 2016. An exploratory genre analysis of three graduate degree research proposals in applied linguistics. *Functional Linguistics* 3/7: 1–28
Young, Lynn. 1990. *Language as Behavior, Language as Code: A Study of Academic English*. Amsterdam: John Benjamins.
Young, Lynn. 1994. University lectures-macro-structure and micro-features. In John Flowerdew ed., 169–178.

Corresponding author
Basma Bouziri
University of Gabés
Institute of Arts and Crafts of Tataouine
Department of Education
Cité Mahrajène
3200 Tataouine
Tunisia
Email: basma.bouziri@isamt.rnu.tn

received: May 2020
accepted: September 2020

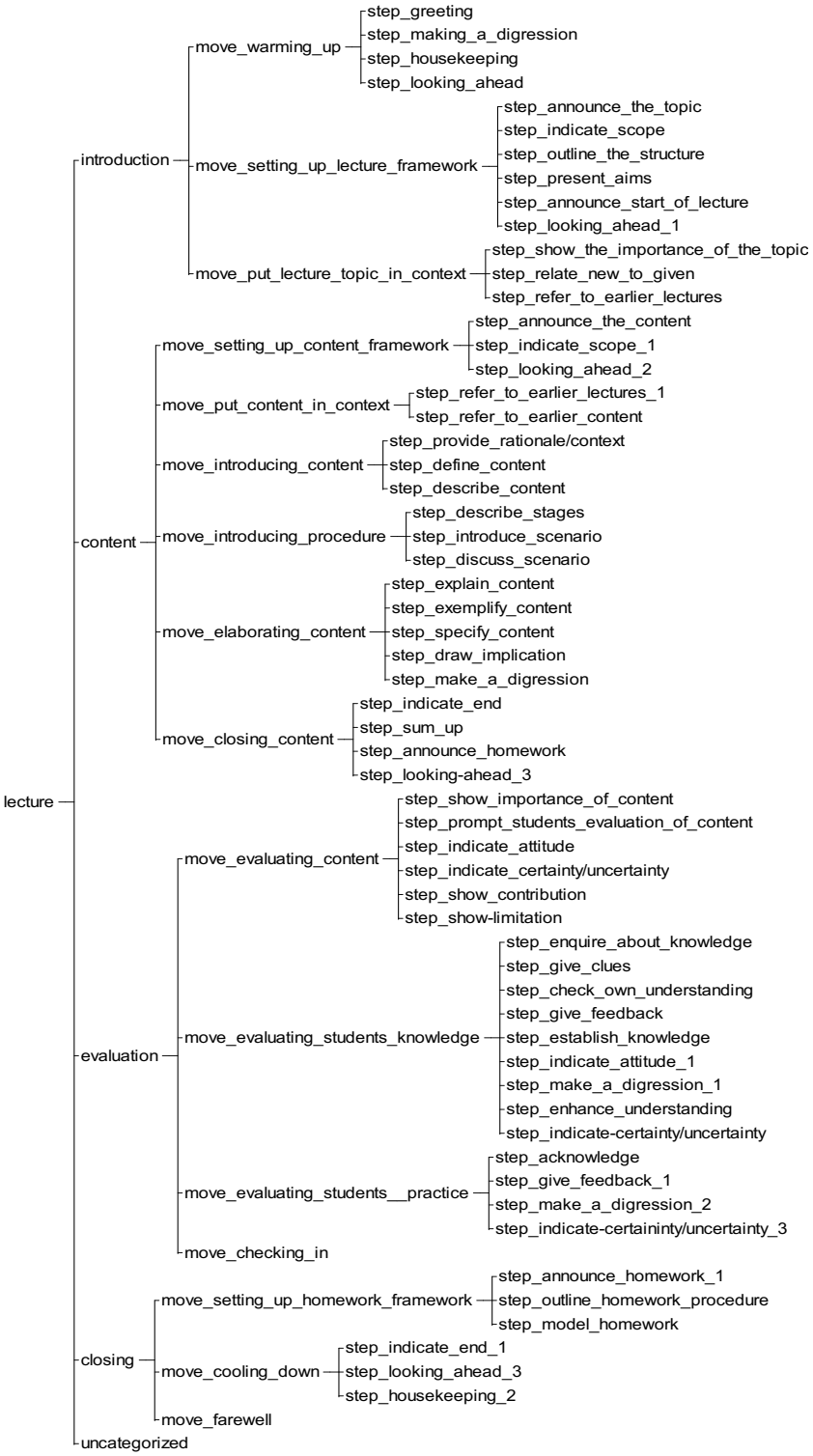
APPENDICES

Appendix 1: Transcription guidelines[4]

ELEMENT	TRANSCRIPTION CONVENTION
Backchannels	Examples include: *oh, oops, really, okay, yeah, uhm, no, yes, right*
Contextual events	-Teacher does not speak while he writes <the lecturer writes on the white board dur= 5 secs> -Teacher speaks while he writes <the lecturer writes on the white board>text</the lecturer writes on the white board>
Error correction	<error corr=this>these</error corr=this>
Fillers	Fillers like er, err, erm, mm are transcribed
Participants	Lecturer is transcribed as 'l' followed by the code attributed to him/her. Example: <l_01> <s> refers to one student. <ss> refers to a group of students. <presenter> refers to a student making a presentation. <presenters> refer to two or more students making a presentation.
Pauses	[.] [..] [pause dur=7 secs]
Sounds	<sound= 'tch'/>
Translation	Translation is included between a double parenthesis <foreign>Oay</foreign> ((yes))
Unintelligible tokens	<unintelligible token='1'/>

[4] A more detailed version of the transcription guidelines is available at https://www.iris-database.org/iris/app/home/detail?id=york:938327

Appendix 2: Coding scheme for rhetorical functions



Appendix 3: Examples of rhetorical functions

RETHORICAL FUNCTION	EXAMPLES
STRUCTURING	
FRAMING	
SETTING_UP_LECTURE_FRAMEWORK	
Announce_the_topic	<l_7>today we're going to deal with a method which is completely different from the others [.]</l>
Outline_the_structure	None
Present_aims	None
Announce_start_of_lecture	None
Looking_ahead	None
SETTING_UP_CONTENT_FRAMEWORK	
Announce_content	<l_07>so we start first with the er [.] background yes</l>
Looking_ahead	<l_07>a lot of arts and drama [..] in the teaching **and we will see later on how can this happen during the process**<l>
REVIEW _CONTENT	
Indicate_end	<l_07> **so this is the first thing** the second this is</l>
Sum_up	<l_07>so this is the first thng you write [.] the first observation of lozanov is [.] that learners do not use more than ten percent of their brain capacity under traditional methods of teaching [.] </l>
REVIEW_ LECTURE	
Indicate_end	<l_09>okay so let's stop here</l>
Housekeeping	<l_09>so did I tell you about the test? <s> <responses> <l_09>it will be the first week after the holidays which means the thirtieth of march</l>
Looking ahead	<l_07>I see you next time [.] for multiple intelligence lessons</l>
CONTEXTUALIZING	
PUT_LECTURE TOPIC_IN_CONTEXT	
Refer_to_earlier_lectures	<l_09>so we already focused in the first sessions on different periods of american history we talked about the colonial period [.]
PUT_CONTENT_IN_CONTEXT	
Refer_to_earlier_content	<l_07>remember that pupils said that the system is not fair because you teach us in the same way even if we have different? levels</l>
Provide_rationale/context	<l_07>teachers do not give homework because he discovered also through a questionnaire and the studies that he did that one of the problems with school is? Homework</l>
ELABORATING	
Explain_content	<l_07>it's a method of teaching that has an objective to? [.] lower or eliminate all the psychological barriers that pupils develop during their learning process **and hence what can we do? We can maximize what now? their use of their brain capacities [.] if we lower these psychological barriers students are going to use more their cognitive skills**</l>
Exemplify_content	<l_07>there are classrooms which are special for teaching english okay? and teachers make an effort to decorate them..with different things that are related to the e,glish culture maps I don't know the maps of Britain of the US some poems famous famous famous sayings different things pictures in which you find the parts of the body the different flags</l>
Specify_content	<l_07>this is his diagnosis of the state of learners while they are taught according to the conventional teaching methods **of course here we talk about grammar translation method audiolingual err method of teaching the cll the different types of methods that existed before when pupils are taught using the methods**</l>
Draw_implication	<l_07>remember that we said that he discovered that this brain capacity is limited to? [.] ten percent **so we try to? maximize that through lowering these psychological barriers** okay?</l>

RETHORICAL FUNCTION	EXAMPLES
DEFINING AND DESCRIBING	
Define_content	<l_07> so can you tell us what is desuggestion?</l> <s><student resumes presentation dur=12 secs></s> <l-07>okay meaning offering options and proposals</l>
Describe_content	<l_07> a suggestopedic class should be like this the first thing as you said comfortable [.] what? <s>comments</s> <l_07> okay so comfortable class so we should have a comfortable environment yes?so this is important the chairs are arranged in a semi-circle </l>
EVALUATING_ LECTURE_ CONTENT	
Show_importance_of_content	<l_07>and it is very important [.] to highlight that this person is a psychotherapist because all of this is going to influence</l>
Indicate_Scope	<l_09>okay so let's move to the most important part which is the constitution itself</l>
EVALUATING_KNOWLEDGE	
Enquire_about_ knowledge	<_07> georgi lozanov [.] who is actually what? <s>response</s> <l_07> a psychotherapist basically and educator [.] what is a psychotherapist?</l>
Give_clues	<l_09>what is the meaning of a census? **sta? [.]** census? statistics which are made every ten years…<:l>
Establish_knowledge	<l_07>lowering means of course diminishing the psychological barriers</l>
Indicate_attitude	<l_09>no i personally i <foreign>belEaks</foreign> ((on the contrary)) he is the least charismatic of all the russian presidents er</l>
Indicate certainty/uncertainty	<l_07>**I don't know whether you have ever gone to a psychotherapist** but in a cabinet of a psychotherapist there are two key features</l>
Give_feedback	<l_07>so you tried the maximum to apply things that are related to the natural approach of course there are other stages that [.] should be included</l>
Enhance_understanding	<l_07>mozart bethoven vangelis bach all of these are type of baroque music<:l>