

# THE EMIBO CORPUS

## A resource for investigating lecture discourse across disciplines and lecture modes in an EMI context

JANE HELEN JOHNSON, MARIANGELA PICCIUOLO  
UNIVERSITÀ DEGLI STUDI DI BOLOGNA

**Abstract** – The aim of this paper is to introduce and describe the EmiBO corpus and present some initial data. EmiBO is a corpus of transcribed Master’s degree university lectures in English given by Italian lecturers, featuring different disciplines and lecture modes. The corpus is constantly being expanded as new recordings are acquired and their transcriptions added. At present it includes 21 complete lecture events by 14 different lecturers in Engineering and Economics subjects, corresponding to 36 lecture hours and just over 200,000 words. Lecturer and student participant turns are annotated. One part of the corpus includes transcripts of audio and video recordings of face-to-face (F2F) lectures, while the other features transcripts of online lectures, including written elements in the chat. The inclusion of audio and video recordings of different lecture modes make it possible to focus on the interplay between spoken and written input, image and body language, while variations in communicative practices may be tracked as new lectures by the same speaker are added. The different modes brought together in a single corpus constitute a unique opportunity to investigate and compare language and non-verbal elements across EMI lecture contexts. Insights are given into the hitherto under-investigated features of Online Distance Learning in EMI, thus being of interest to others besides EMI scholars. Also of note is that non-native English speaking lecturer discourse practices may be compared cross-sectionally across different modes from a truly ELF-oriented perspective. The paper presents and comments quantitative data resulting from corpus analysis as well as outlining some initial qualitative explorations with suggestions for further development.

**Keywords:** EMI; lecturer discourse; F2F lecture mode; online lecture mode; corpus analysis.

## 1. Introduction

Corpus linguistics tools are uniquely placed to provide insights into the English used by lecturers in the academic context of English as a Medium of Instruction (EMI) (Jablonkai 2021). This paper describes our purpose-built corpus of mainly spoken data from academic lectures in an EMI context, discusses challenges, presents some initial findings and outlines some ways forward.

The term EMI has become standard in reference to higher education programmes where the emphasis is conveying content rather than language, in English, and where a high level of English competence is required to guarantee success (Wilkinson 2017). In Italy, as in other countries, the number of courses taught in English has increased exponentially over the last decades (Campagna, Pulcini 2014; Dearden 2014). The degree of linguistic and pedagogical competence of many EMI lecturers, who frequently have to adapt quickly both the mode and the method of their teaching, has often come under scrutiny (see Lasagabaster 2022 for a useful summary). To this end, electronic collections of academic discourse can provide invaluable material, giving empirical evidence of language in use (Jablonkai 2021).

### 1.1. Literature review

Academic spoken discourse consists of a number of subgenres, including classroom management and office hours, as well as service encounters with administrative staff (Biber *et al.* 2002). However, it is the academic lecture which has been the focus of most research (Lynch 2011). Best described as an expository classroom genre (Fortanet-Gómez 2005), it involves both lecturer and student as active participants. Lecture functions include introducing theories and concepts, integrating ideas from previous lectures, and presenting information to the audience, as well as evaluating materials (Lee 2009).

Important contributions have been made to our understanding as regards aspects of the academic lecture (Fortanet-Gómez, Bellés-Fortuño 2005) and findings have been applied in a number of areas. Research into lecture discourse with particular reference to EMI and our own work has focused on the following:

- *metadiscourse*: Ädel (2010) devised a taxonomy of the discourse functions of spoken and written academic English; Molino (2018) uses this taxonomy to perform a corpus-based analysis of metadiscursive functions in EMI lecture discourse, relating such functions to both personal and impersonal metadiscourse and to the language forms used; Brogginì and Murphy (2017) investigate reflexive metadiscourse in a corpus of Italian EMI lecturers' speech, finding a predominance of personal markers; Mauranen (2012) focuses on metadiscourse in an English as Lingua Franca (ELF) context; Doiz and Lasagabaster (2022) also consider spoken interactive metadiscourse in EMI from an ELF perspective. In addition, Deroey and Johnson (2021) describe the use of lexico-grammatical importance markers in Engineering lectures, and found little difference in the frequency and range of the markers used by NS and NNS speakers;
- *interaction*: previous studies have investigated classroom interaction by means of (1) interviews focusing on lecturers' and students' perceptions (Morell 2007; Picciuolo, Johnson 2020; Revell, Wainwright 2009); and (2) corpus-based studies comparing the frequency, form and function of questions as linguistic devices through which lecturers and students perform their teaching/learning tasks. These latter have shown that it is the genre more than the disciplinary content which affects questioning in lecturers' discourse (Chang 2011; Crawford Camiciottoli 2004, 2008; Dafouz Milne, Sanchez García 2013) as well as lecturers' instructional style (Morell 2004; Northcott 2001). Other corpus-based investigations in this field have focused on lecturers' use of personal pronouns as a means for engaging students with lecture content and enhancing students' comprehension, comparing pronoun frequency and function across disciplines (Johnson, Picciuolo 2022; Yeo, Ting 2014);
- *pragmatics*: past studies (e.g. Alsop 2016; Alsop, Nesi 2013; Bellés-Fortuño, Fortanet-Gómez 2009; Björkman 2011) have emphasized the importance played by pragmatic strategies in promoting effective communication exchanges in the ELF classroom. In her work, Björkman (2011) focused on variations in the use of selected strategies in two different ELF speech events (students' work-groups and lectures), finding that such strategies were more frequently employed by students, allowing them to preserve mutual understanding "despite frequent non-standardness in the morphosyntax level" (p. 950). Alsop and Nesi (2013) investigated cross-cultural and cross-linguistic variations in Engineering lecturers' use of summary, showing that EMI Malaysian lecturers used summary for previewing and reviewing current talk more often than British and New Zealand lecturers, thus paying more attention to reinforcing content than NS lecturers. More recently, Picciuolo (2022) compared frequency and function of lexical spatial deictic references in the discourse of EMI

Engineering lecturers across different teaching modalities, showing that the virtual environment forces the lecturers to interact with the learning objects (both physical and virtual) more frequently than in the F2F classroom, thus anchoring students' attention and engaging them;

- *lexis*: the lexical features of both spoken and written academic discourse have been described in detail by Biber (2006) and Biber *et al.* (2002). These include the functions of lexical bundles (Biber, Barbieri 2007; Biber *et al.* 2004), their cohesive role (Nesi, Basturkmen 2006), and the frequency of formulaic expressions (Simpson 2004; Simpson-Vlach, Ellis 2010). Martinez *et al.* (2013) investigated lexical phrases used by lecturers to introduce key terms. Lexical elements also featured in Mazak and Herbas-Donoso's (2015) study of translanguaging in a Spanish EMI context. Research into lexis has also focused on discipline-specific vocabulary (e.g. Hyland, Tse 2007), particularly since this may affect the frequency, collocation, range, meanings and functions of the lexical features. Dang (2018a) found that both spoken and written texts from the hard sciences are more challenging for students as regards lexical coverage than those of the soft sciences, while a knowledge of a core vocabulary of general English was found to be more useful for soft than for hard sciences (Dang 2018b). Wordlists based on a particular subject (e.g. Engineering: Mudraya 2006; Gardner, Xu 2019) are therefore fundamental in planning ESP course material for both students and EMI lecturers (Dang 2018a, 2018b).

We may note that two variables in particular are used for comparison. One is the distinction between disciplines (e.g. Alsop, Nesi 2013; Crawford Camiciottoli 2004, 2007; Northcott 2001). Crawford Camiciottoli (2007), for example, offers an in-depth investigation into business lecturers' discourse, focusing not only on linguistic features, but also on paralinguistic and multimodal aspects. Instead, Alsop and Nesi (2013) investigated discourse variations in EMI Engineering lecturers from different L1. Another variable is the status of native (NS) or non-native speaker (NNS), as in Deroey and Johnson (2021). Other examples include Suviniitty (2012), who compared interaction in Finnish lecturers' speech in English and in their L1, and Hellekjaer (2010) who compared lecture comprehension in English and in their L1 by students in Norwegian HE. These two variables of non-native-speaker discourse and discipline-specific discourse were also of particular relevance in designing EmiBO.

## 1.2. Review of existing corpora of academic English

While many corpora of written academic English are available, collections of spoken academic English are much less frequent, partly due to the time required for transcription as well as to issues of access and confidentiality, and the difficulties of data collection itself. Such corpora include the Michigan Corpus of Academic Spoken English (MICASE) with almost two million words (Simpson *et al.* 2002), the 2.7 million-word corpus compiled for the TOEFL 2000 Spoken and Written Academic Language Project (T2K-SWAL) (Biber *et al.* 2001), and the 1.6 million-word British Academic Spoken English (BASE)<sup>1</sup> corpus. These all contain mainly native speaker discourse from different academic sub-genres and across disciplines. Instead, the one million-word English as a Lingua Franca Academic corpus (ELFA) (Mauranen 2006) includes contributions from mainly non-native speakers in the context of a Finnish university, across different sub-

<sup>1</sup> The British Academic Spoken English (BASE) corpus was developed at the Universities of Warwick and Reading under the directorship of Hilary Nesi and Paul Thompson. Corpus development was assisted by funding from BALEAP, EURALEX, the British Academy and the Arts and Humanities Research Council.

genres and disciplines. Other specially compiled corpora include the Engineering Lecture Corpus (ELC)<sup>2</sup>, containing lectures in the same discipline in both NS and NNS contexts, and the Open Yale Course Lecture Corpus (Wingrove 2022), with lectures from three different disciplines.

The aforementioned corpora mainly contain transcripts of audio recordings. BASE also contains some annotation as regards prosodic and kinesic features. Indeed, since lecturers also depend on non-verbal resources such as gesture and visuals to transmit and get across meanings (Kress 2010; Kress *et al.* 2005), video recordings, especially those aligned with multimodal annotation, provide useful additional information (Fortanet-Gomez, Querol-Julián 2010; O’Keeffe *et al.* 2007). However, existing multimodal material in the academic field is scarce due to the amount of time needed to transcribe and annotate satisfactorily. The following are just a few examples of diverse multimodal academic and general corpora:

- *The Multimodal Academic Spoken Language Corpus* (MASC) (Fortanet-Gómez, Querol-Julián 2010) is a multidisciplinary corpus bringing together spoken academic events such as lectures in Spanish and English. Its multimodal nature is defined by the use of five different components: slides, transcripts, handouts, and video and audio recordings.
- *The Multimedia Adult English Learner Corpus* (MAELC) (Reder *et al.* 2003) consists of more than 3600 hours of adult ESL classroom interaction and associated written materials classes.
- *The Padova Multimedia English Corpus* (Padova MEC) (Ackerley, Coccetta 2007) includes lecture recordings and conference papers by English-speaking visiting professors as well as non-native speakers as a lingua franca, as well as student presentations.

A further development in academic lecturing is the use of online, network-based platforms that allow participants to synchronously interact with each other, sharing their voice, video and data, and giving and receiving instant feedback without being physically together in the same room (Massner 2021). In this way, students can “communicate with other students and the instructor through text-, audio-, and/or video-based communication of two-way media that facilitate[s] dialogue and interaction” (Martin *et al.* 2017, p. 3).

Studies of online teaching and learning have identified several problematic issues related to technology-mediated teaching and learning, including the professional development of teachers, knowledge construction, and patterns of interaction between the participants (Carrillo, Flores 2020). A perceptions study showed that Italian EMI lecturers teaching online following Covid-19 lockdown generally expressed dissatisfaction with their online courses, mainly due to the lack of interaction and direct feedback compared to their F2F classrooms (Cicillini, Giacosa 2020). However there is still little evidence-based research (Querol-Julián, Crawford Camiciottoli 2019) into the features of synchronous online teaching and learning in EMI.

To conclude, there is continued interest in research into EAP issues, with topics of interest continuing to be EMI and ELF, as well as an increasing focus on online education and enhancing EMI lecturers’ multimodal competence to engage students in different learning settings. The EmiBO corpus offers a unique opportunity to add to this research, since it contains lectures in English on Physical Sciences and Social Sciences from both F2F and online courses by Italian native speakers.

<sup>2</sup> <http://www.coventry.ac.uk/elc> (31.01.2022)

## 2. EmiBO: Background and Methodology

The development of the EmiBO corpus was begun as part of a research grant<sup>3</sup> investigation into EMI at the University of Bologna (UniBo). In the academic year 2018-19, UniBo offered more than 500 courses delivered through EMI and attended by both local Italian and international students, with about 30 full degree programs taught in English, involving 450 lecturers, most of whom taught at either the faculty of Engineering or Economics. These areas formed the most representative sample of EMI at UniBo (as at other Italian and European universities: Costa 2021; Costa, Coleman 2013; Dimova *et al.* 2015). Within these two areas, which we refer to respectively as Physical Sciences (PS) and Social Sciences (SS), lectures from four international Master's Degree courses<sup>4</sup> at UniBo were collected. The following sections describe corpus design, data collection and data transcription.

### 2.1. Corpus design

Our aim was originally only to focus on audio and video recordings of lectures taking place in a traditional classroom setting, where recordings were provided for research purposes as evidence of synchronous communication and not intended to be retained as asynchronous lectures. However, the advent of transition to online or blended synchronous learning<sup>5</sup> as a result of the Covid-19 pandemic provided an unexpected further source of material. Lecturers were suddenly obliged to deliver their lectures online, and videoconferencing platforms, such as Zoom and Microsoft Teams, abruptly became normalized so as not to interrupt teaching (Luporini 2020). A number of these lecturers were also willing to contribute recordings of their online lectures. This opened up new horizons for research, making it possible to envisage alignment of speech, gesture and visuals, as well as asynchronous mechanisms such as the chat.

In this way, we were able to store the transcripts from two distinct modes of lecture delivery: audio/visual in a face-to-face context, and online. The type of source data may thus be described as in Table 1:

Sub-corpus	Description
Part I/ F2F	audio or video recording of <i>face-to-face</i> (F2F) lectures given in the classroom
Part II / online	Online Distance Learning (ODL) either <i>completely online or blended</i> . This includes both audio and video, with lecturer and student interactions recorded both orally and written in the chat.

Table 1  
Sub-corpora and description.

Lecturers in both parts make a more or less spontaneous commentary, with or without the aid of a computer-based presentation, or using the whiteboard. The lecture style may be described as conversational (Dudley-Evans 1994).

<sup>3</sup> «Insegnare in inglese all'UniBo», funded by DIRI – International Relations Office of the University of Bologna. Research grant number ID-51465, Rep. 29/2018, Prot. 1084 del 17/7/2018.

<sup>4</sup> International Master's Degree courses in Civil Engineering, and Advanced Automotive Engineering (PS), Business and Economics, Health Economics and Management (SS).

<sup>5</sup> “Learning and teaching where remote students participate in face-to-face classes by means of rich-media synchronous technologies such as video conferencing, web conferencing, or virtual worlds” (Bower *et al.* 2015, p. 1).

Our aim was to sample as wide a selection of lectures as possible. We prioritised selection of complete lectures from start to end, rather than fragments of lectures in order to reduce variation in linguistic elements in different parts of the event (Biber 2006; Sinclair 1991). We used convenience sampling in order to collect lectures by a variety of different speakers as well as different lectures by the same speaker. It was not possible to identify the lecture position within the course, so there was no way of knowing how ‘representative’ the lecture was of the course. While these issues are important, the more lectures are added to the corpus, the less such variables affect the findings.

All the lecturers were NNS. The self-declared English language level of all contributors ranged from B1 to C1 on the Common European Framework of Reference (CEFR), with the majority stating C1. Our sample included lecturers with different years of experience teaching in English. Most lecturers had followed no specific teacher training. Such findings are in line with those from O’Dowd’s (2018) survey of EMI lecturers in European universities.

The corpus is constantly being expanded as new recordings are collected, transcribed and added. We are thus following a ‘monitor corpus’ approach (Sinclair 1991, pp. 24-26). In this way a number of different discourse events become available for analysis, thus providing more representative sampling. It is also possible to track developments as lecturers changed mode and/or became more experienced in teaching and in their language competence.

## **2.2. Data collection**

Data collection was differentiated according to the source. Audio recordings were done by means of portable mini-digital voice recorders or mobile phones. While most lecturers recorded their own session and provided the material in electronic format, on occasions the lectures were recorded by the researchers in person. One advantage of this was that the audio recordings could be supplemented by field notes as aids to subsequent transcription. Instead, video recordings delivered online through platforms such as Microsoft Stream and YouTube were shared by lecturers with our research team.

## **2.3. Lecture transcription**

Recordings were transcribed manually and proof-read by both researchers. Participant turns were marked and the participants identified only by their roles as ‘lecturer’, ‘student’ and ‘researcher’. Annotation also made it possible to distinguish between lecturer’s and students’ spoken turns and written turns in the chat in Part II / online.

A simplified annotation system based on Jefferson (2004) was used to note micropauses (.) and overlapping talk []. Discourse disfluencies such as false starts and hesitations were lexicalised (e.g. *uhm*, *erm*). Similarly, non-verbal actions like laughing or coughing were signalled in angle brackets (e.g. <laugh>). Standard punctuation marks (comma, full stop, question mark) were used.

Transcription decisions involved how to deal with non-standard English usage as well as domain-specific words. For example non-standard pronunciation is often found among non-native English speakers. Such items in our corpus were transcribed in the standard form. An example is one lecturer’s non-standard pronunciation /retʃ:ikləbʊlz/ which we transcribed as ‘recyclables’. Instead, non-standard grammatical structures, such as “the water look blue”, were transcribed as such. What is more, a number of domain-specific words were unfamiliar to both the English-native and the Italian-native

transcribers. Here an internet search for possible ‘word candidates’ proved invaluable. It is worth noting that this situation is similar to that experienced by students in the international EMI class, though these would have the disadvantage of having to deal in real-time with the problem.

### 2.3.1. Part I / F2F - issues with audio and video transcription

Audio recording, particularly in settings such as university lecture halls, requires reliable technical tools. However, recording quality sometimes impeded comprehension to such an extent that some lectures could not be included in the EmiBO corpus. This was particularly unfortunate considering that it was not easy to gain the trust of the lecturers in order to audio-record their lectures. There were several reasons for poor recording quality, related to the physical distance between lecturer, students and microphone; classroom acoustics; sound reverberation; size of lecture hall and ambient noise. In some cases, spoken output was very difficult to distinguish, and double-checking by both researchers was invaluable. Parts which were still impossible to decipher were marked [inaudible] in the transcript. Though the tendency was for teacher-fronted delivery, when students did intervene, two main problems arose for the transcriber. First of all, in a large lecture hall, a digital voice recorder will not pick up sound properly if the source is too far away. What is more, when students’ interventions overlap, it is almost impossible to distinguish individual speakers clearly. Video recordings of lectures delivered in person present similar difficulties to audio recording considering that the camera is generally directed towards the lecturers while students are usually off-camera and their voice is often barely audible. In such cases, the researcher’s field notes where present proved invaluable.

### 2.3.2. Part II / online - issues with recordings and transcription

In Microsoft Teams, as in most videoconferencing platforms, users are provided with a set of communication tools (e.g. messaging and chat, file sharing, videoconferencing, document editing, polling and survey) that, however effective and engaging these tools may be, are mainly restricted to three communication modes – textual, visual, and oral. These pose objective difficulties when collecting and transcribing data.

Firstly, collecting *textual* data in online communication addresses the issue of representing “sequentiality” (Helm, Dooly 2017) – i.e. turn-taking – considering that participants can communicate not only by speaking into the microphone, but also by writing in the chat, such that conversations may thus *silently* overlap. In our corpus, this is a major problem considering that while lecturers mainly speak into the microphone, students largely communicate through the chat. Although timestamps usually allow us to establish the conversation flow, sequentiality is nonetheless hard to establish when (1) students simultaneously write comments in the chat, considering that researchers can see the chat only when the lecturer shifted back from presentation mode to desktop view; (2) the lecturer writes in the chat and speaks at the same time; (3) the lecturer replies to students asking questions in the chat but without showing the chat. This problem was solved in part by using separate ‘tiers’ in ELAN<sup>6</sup>, a software that allows for extensive annotation of multimodal data, to keep track of overlapping turns and the modes through which participants communicate (see Section 2.4).

<sup>6</sup> ELAN (Version 6.2) [Computer software]. (2021). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan> (31.01.2022).

Secondly, collecting *visual* data from online recordings poses considerable challenges in terms of access to visual cues such as participants' positioning and gaze (Adolphs, Carter 2007; Bedenlier *et al.* 2021). Researchers have neither control of webcam positioning, nor of the size and quality of the image (Helm, Dooly 2017) which mainly depends on infrastructural constraints as well as on the affordances provided by the videoconferencing platform. In this respect, lecturers usually give salience primarily to the textual material shared in the presentation mode. While this made transcription of lecturers' commentary easier, it made transcription of the 'invisible' chat more difficult.

Thirdly, as regards the *oral* mode, good audio quality in online classrooms enables further investigation into paralinguistic cues such as tone of voice, intonation, speech rate, accent and hesitation. However, when the lecturer moves around the classroom and speaks away from the microphone, it is difficult to distinguish the speaker's voice clearly. Similarly, students are generally far from the microphone and their interventions are often inaudible.

Finally, there are two further issues when *transcribing* multimodal data, related to *selection* and *taxonomical classification* of non-textual modes. The transcribing process inevitably involves selection according to the interests of the researcher (Helm, Dooly 2017) particularly when dealing with multimodal data. Selection at present involved verbal resources – speech, and written material in the chat – and non-verbal modes, including visuals – such as emojis and 'graphicons' (Al-Zou'bi, Shamma 2021; Herring, Dainas 2017), but also tables, graphs, images, gesture – including mouse pointing and toggling between windows – gaze, and spatial positioning (Morell *et al.* 2020). As regards taxonomy, at present, there is no consistent taxonomy used for tagging non-textual cues in multimodal classroom communication in online settings, such that scholars have used a variety of taxonomies when referring to gestures, gaze and positioning (Lim *et al.* 2012; Morell *et al.* 2020; O'Halloran *et al.* 2014; Querol-Julián 2021). We make use of Morell *et al.*'s (2020) taxonomy for the description of non-verbal modes in our study as regards positioning, gaze and gesture, while we refer to O'Halloran *et al.* (2014) for the analysis of non-verbal resources using technological tools.

#### 2.4. Metadata

Contextual information such as lecturer identification, course subject, date, time and mode of recording was added to each file. Details of individual student origin were unavailable. Transcriptions were lemmatised and part-of-speech tagged by SketchEngine<sup>7</sup>. Transcription and annotation of online discourse was done with ELAN, which offers a timed multimodal transcription, where different modalities can be separately transcribed on different parallel and synchronous horizontal tiers such that duration and coincidence of each mode are easily recognisable. Furthermore, ELAN allows for multimodal data playback, and audio representation as waveform or pitch contour. Finally, ELAN projects can be exported in CSV-files for later analysis, allowing the researcher to look for statistical significance in the dataset, e.g. the frequency of co-occurring semiotic modes of interest. ELAN has proved to best suit the needs of our monitor corpus since it enables new tiers to be added as other aspects of interest emerge (Helm, Dooly 2017). However, ELAN does not allow for automatic transcription, such that textual and non-textual cues need to be transcribed manually.

<sup>7</sup> Kilgarrif *et al.* 2004.



### 3. Quantitative data

This section presents and comments on some quantitative data in relation to the transcription of the EmiBO corpus so far. Throughout, we shall distinguish between mode (Part I / F2F and Part II / online), and macro discipline (PS and SS). At present, the transcription includes all audio and video commentary as well as all written contributions in the chat.

Figure 1 gives details of the number of lectures and different lecturers across the two modes and disciplines in the EmiBO corpus.

	PHYSICAL SCIENCES	SOCIAL SCIENCES	Total
Lecturers	14	3	17
Part I / F2F Lectures	12	3	15
Part II / online Lectures	6	0	6

Figure 1  
Number of lecturers and lectures according to mode and discipline.

#### 3.1. Duration and word counts

Figure 2 gives details of the latest EmiBO corpus release (2021) as regards minutes and number of words.

		PHYSICAL SCIENCES	SOCIAL SCIENCES	Total
Duration in minutes	Part I / F2F	1078	402	1480
	Part II / online	655	0	655
	<b>Total</b>	<b>1733</b>	<b>402</b>	<b>2135</b>
Number of words	Part I / F2F	107,729	22,400	130,129
	Part II / online	73,029	0	73,029
	<b>Total</b>	<b>180,758</b>	<b>22,400</b>	<b>203,158</b>

Figure 2  
Duration in minutes and number of words in EmiBO.

#### 3.2. Word ratio

Lecturer discourse makes up the majority of words in EmiBO (figures 3 and 4), though students also contributed and on one occasion the researcher answered a question directly addressed to her by the lecturer. All this was transcribed.

	<b>Participant</b>	<b>Part I / F2F</b>	<b>Part II / online</b>	<b>Total</b>
<b>Spoken</b>	Lecturer words	121,643	72,821	194,464
	Student words	7,584	85	7,669
	Lecturer: student ratio	16:1	857:1	
	Researcher words	137	0	137
<b>Written</b>	Lecturer words (chat)	-	50	50
	Student words (chat)	-	43	43
<b>Total</b>		130,129	73,029	<b>203,158</b>

Figure 3  
Number of words per participant per *mode*.

	<b>Participant</b>	<b>PHYSICAL SCIENCES</b>	<b>SOCIAL SCIENCES</b>	<b>Total</b>
<b>Spoken</b>	Lecturer words	176,978	17,486	194,464
	Student words	3172	4497	7669
	Lecturer: student ratio	56:1	4:1	25:1
	Researcher words	137	-	137
<b>Written</b>	Lecturer words (chat)	50	-	50
	Student words (chat)	43	-	43
<b>Total</b>		180,758	22,400	<b>203,158</b>

Figure 4  
Number of words per participant per *discipline*.

The ratio of lecturer words to student words in EmiBO is 25:1 as a whole. In Part II / online we found an extremely unbalanced lecturer/student talk ratio (857:1), with students never speaking in most classes. This could mean that students feel intimidated by the medium or are unaccustomed to speaking out and merits further investigation.

A difference, though less striking, is also seen if the two macro disciplines are considered separately (figure 4), with lecturer-to-student word ratio in Physical Sciences being 56:1 but only 4:1 in Social Sciences. This indicates that students are contributing much more during Social Sciences lectures than during Physical Sciences lectures.

### 3.3. Participant turn ratio

The higher the number of speaker turns, the more likelihood that this indicates greater interaction. The average number of participant turns in Part I / F2F lectures is 99, while

the average number of spoken participant turns in Part II / online is 5. This lack of audience input in Part II / online suggests that our online lecturers have a non-interactive style (Morell 2004), though we need to bear in mind that some contributions were not spoken but written in the chat.

The average number of participant turns in Physical Sciences lectures is 50, while the average number of speaker turns in Social Sciences lectures is 203. Social Sciences lecturers would thus seem to have a much more interactive style (Northcott 2001) than Physical Sciences lecturers.

We are not suggesting that balanced speaker turns alone create satisfactory interaction. There may indeed be other resources at play, since the combination of various semiotic resources in online lectures in co-occurrence with lecturer's speech has been found to foster interaction with students as well as make meaning more understandable (Querol-Julián 2021).

### **3.4. Lexical variation**

The ratio of word types to word tokens gives an idea of lexical variation. The closer the type-token ratio (TTR) is to 1, the greater the lexical richness. The TTR is the same (0.05) across the two different lecture modes (F2F and online), while Social Science lectures have greater lexical variety (0.10) than the Physical Science lectures (0.04). TTR thus appears to be affected by the discipline and not the lecture mode.

### **3.5. Speech rates**

Paralinguistic features such as speech rates may be compared in the two parts. The mean speech rate in Part I / F2F is 91 words per minute (wpm) which is slightly slower than for native speakers (Tauroza, Allison 1990). The difference might be due to EmiBO lecturers being NNS, or they might be deliberately speaking more slowly to make allowances for their student audience. There might also be another explanation, if we look closer at the two sub-corpora. In our Social Sciences F2F lectures, for example the mean speech rate is 56 wpm. This slow rate might be due to the higher number of student interventions in the Social Sciences lectures, since interaction has also been found to slow the pace (Morell 2020, p. 65). Instead, in Part II / online the mean speech rate is 113 wpm, suggesting that lecturers tend to talk at a slightly faster rate online than in F2F classrooms. However as mentioned in section 2.3.2, sequentiality is a problematic issue in online lectures and speaking often overlaps with writing, thus making speech rates difficult to compare with precision.

### **3.6. Lecturer gender**

Past research has found differences between male and female spoken discourse in the academic context (Schleef 2008). While four out of ten of our Physical Sciences lecturers are female, at present EmiBO contains no lectures by female Social Sciences lecturers. Two of the top three lectures in Part I / F2F as regards number of participant turns were from female lecturers, suggesting greatest interaction. Given that female lecturers made up only 13% of the F2F lecturers, we might conclude that female lecturers tend to be more interactive than male (also suggested by Schleef 2008), at least in a F2F context. In Part II /online, speaker turns were very few on average, suggesting little interaction overall. Since, as mentioned above, there appears to be little interaction in online lectures in

general, this suggests that gender alone does not determine the degree of classroom interaction in online EMI lectures, but that other aspects are involved.

### 3.7. *Idiosyncratic lecture style*

Contributions by the same lecturer across different lecture modes reveal possible evidence of idiosyncratic lecture style and adaptation to the medium. For example, there are four lectures by the same lecturer and none have many participant turns. This is evidence of little or no interaction and a purely monologic lecture style, which could in part be due to the medium used (three of them are online) and/or to the lecturer's idiosyncratic teaching style.

To conclude, the quantitative data shows there are notable differences throughout the EmiBO lectures as regards both the ratio of lecturer/student participant turns, and the number of words within the turns. There are also some major differences in speaking rate. Some of these differences might be due to idiosyncratic lecturing style, whereas others may be due to the amount and quality of lecturer/student interaction, or to the teaching mode itself. While Part II / online also includes contributions written in the chat, which may also involve icons and emojis, in the current state we see that this facility is little used by both lecturer and students. Female lecturers, though not represented in equal numbers, tend to be more interactive than male, particularly in F2F lectures. Online lectures appear to be less interactive than audio/video lectures, though further research is necessary to rule out the effect of personal lecturing style and class size, since it is easier to stimulate interaction in small classes (Crawford Camiciottoli 2005).

## 4. Research using EmiBO

The EmiBO corpus has been used to focus on the linguistic features of lecture discourse in an EMI setting *tout court*, as well as to compare lexical realisations across different learning settings in different macro disciplines. The use of participant turn annotation allows us not only to compare the amount and function of interaction between speakers but also to distinguish between lecturer and student discourse, both spoken and written. Part-of-speech annotation and lemmatization has allowed a comprehensive lexicographic analysis of the data, making it possible to investigate patterns and stylistic features. In this section we describe research which has already been carried out using EmiBO, and highlight areas for further exploration.

One fruitful point of departure is the *wordlist*. Indeed, the two macro disciplines may be distinguished semi-automatically by means of a keyword list to reveal lexis which is more or less frequent than would be expected when compared with another word set (Scott 1997; Scott, Tribble 2006). When comparing Physical Sciences with Social Sciences lectures, the keyword list features a number of specialised items, giving an idea of 'aboutness' (Scott, Tribble 2006, p. 59) and found in many lectures in the macro discipline. We would expect domain-specific vocabulary to emerge from such a comparison when comparing lists of the most frequent words and lexical bundles. In this regard, Picciuolo (2022) found that in online settings, where EMI lecturers tend to interact with visuals – like graphs, tables, pictures – more often than in F2F and blended settings, even domain-specific sentences and word-related vocabulary become more visually salient. This is because the lecturer often interacts with the visuals by highlighting and zooming in. In so doing, the lecturer tends to guide students' attention to some of the most

critical lecture contents, particularly for hard sciences, which are characterised by a high lexical coverage.

Our Physical Sciences corpus features a number of sub-disciplines of Engineering with many lectures on the subject of hydrology and water resource management, thus determining a predominance in our word lists of ‘water-related’ lexis. The spoken PS wordlist included verbs (e.g. *let, see, say, know, mean, go, get, need, look, want, think, make*) which in context might be typical of the spoken language of the worked example: a “step-by-step demonstration of how to perform a task or solve a problem” (Clark, Mayer 2011, p. 224). We note also that both PS and SS lecturers use a relatively limited range of common verbs (e.g. *be, do, go, have, know, let, make, say, see, use*), perhaps because academic lecturers are under pressure to transmit a large amount of complex content in a comprehensible way in a short space of time (Biber 2006; Deroey, Taverniers 2011).

The discipline has been found to affect the lexical behaviour of an item (Hyland, Tse 2007). Examples from EmiBO are the lemmas *time* and *number*, whose collocates in PS are different from those in SS, as is their relative frequency. Such information has practical applications as regards material development for coaching both lecturers and students.

The above studies refer to single word usage. However *multiword units* (MWUs) or lexical bundles - more common in spoken than written genres (Biber *et al.* 2004) - also provide interesting areas of research using the EmiBO corpus, particularly because they are a useful focus for language learning. Indeed, learning phrases rather than single words is important for developing fluency in students (Simpson-Vlach, Ellis 2010). In addition, it has been suggested that frequently recurring phraseology is more important than recurrent words or lemmas in denoting the content of a text (Sinclair 2004, p. 148), a fact which may be exploited when presenting new terminology. Lecturer training courses may also include a focus on MWUs in context. In EmiBO, for example the most frequent 5-word unit in SS lecturer discourse is “*so the idea is that*”, whose function in context appears quite different from that of the most frequent 5-word unit in PS lecturer discourse (“*if you look at the*”), giving an idea of the different rhetorical underpinnings of the two disciplines.

Initial findings from a comparison between Part I / F2F and Part II / online (Picciuolo 2022) show notable variation in the use of *lexical place deixis* as a consequence of the reconceptualization of space in the online environment. For example, we found a higher occurrence of the proximal locative adverb *here* in F2F classes (in line with Bamford 2004, and Friginal *et al.* 2017) signalling the lecturer’s need to anchor students in the physical space of the classroom. In contrast, the proximal demonstrative deictic *this* occurs more frequently in online classes, suggesting the lecturer’s greater need to anchor students to the objects displayed on the screen.

Furthermore, Picciuolo (2022) compared the frequency of lexical spatial deictics co-occurring with gestures and actional resources (O’Halloran *et al.* 2014) – e.g. mouse pointing, zooming in – in the discourse of an Italian EMI Engineering lecturer across different lecture modalities. She found that, although in online and blended modalities stimuli are more visually salient, such that we might expect a constant higher co-occurrence of lexical spatial deixis with actional resources, the occurrence of multimodal spatial deictics – i.e. gesturally and/or actionally anchored lexical spatial deictics – seems not to be dependent upon the lecture mode. Instead, it would depend on the way the lecturer exploits the affordances of the tool, whether a projector in F2F or the mouse pointer in online lectures, with different degrees of awareness. However, in online and blended settings, spoken spatial referring and gestural positioning – through the means of

mouse pointing – often go together, allowing immediate clarification in cases of misunderstanding due to non-standard pronunciation, or the use of proper names or specific terminology.

Deroey and Johnson (2021) compared the use of lexico-grammatical *importance markers* in Engineering lectures in English in an EMI and L1 context. Overall, EMI and L1 lectures were similar in importance marker frequency, (sub)types, and lexemes. Both EMI and L1 lecturers most frequently used verb markers, predominantly in the directive V/N clause (e.g. *remember they don't know each other*). The study suggests that EMI and L1 English speakers highlight key lecture points in a similar way and that EMI lecturers are not limited in their expression of importance markers.

Following on from the above-mentioned study, Johnson (2022) compared the use of *assessment-related expressions* as importance markers in an EMI and L1 context, with particular focus on modality, phraseology, positioning and clustering. Though she found few clear distinctions in terms of NS and NNS, there were differences in the positioning and modality of assessment-related expressions which might affect student reception of such cues (Miller, Parlett 1974). She also found examples of L1 interference in phraseology among the NNS lecturers as well as a tendency among native speakers to use certain idiomatic phrases – noted also by Seidlhofer (2009). Both of these features might hinder comprehension for students of different language backgrounds. As a follow up to this, investigation of student awareness of such usage is planned.

Picciuolo and Johnson (2020) conducted a preliminary investigation into the *perceptions and attitudes* of SS and PS lecturers teaching through English, and compared these findings with classroom observation. PS lecturers were found to be particularly concerned about their own speaking and listening abilities, and appeared to be strongly influenced by the ideology of the Native Speaker model (Jenkins 2014). Conversely, SS lecturers were less worried about their English language competence and more responsive towards aspects of classroom management. Classroom observation through querying the EmiBO corpus confirmed that there are differences according to the disciplinary areas, particularly as regards lecturer-to-student talk ratio, with students in PS classes generally being asked to provide short answers to closed confirmation questions, while SS lecturers adopted a more inclusive pedagogy, allowing students to actively participate through classroom discussions and presentations.

Johnson and Picciuolo (2020) looked at EMI *lecturer questioning* as evidence of increased interaction in the classroom in order to co-construct meaning (Dafouz Milne, Sanchez Garcia 2013; Morell 2004). To do this, they triangulated EmiBO corpus data with lecturers' perceptions of their own classroom practice, as well as taking into consideration student questionnaires. They found that, while most lecturers claimed to make use of strategies to promote interaction, including questions, few questions were actually asked, particularly in Physical Science lectures, and these tended to be content-oriented (Thompson 1998) rather than audience-oriented questions, which made them less useful for the purposes of encouraging interaction (Crawford Camiciottoli 2008). Social Sciences lecturers showed greater awareness about different ways to promote interaction than their colleagues from Physical Sciences. Students admitted that they preferred asking questions at the end of class. While this was less face-threatening than asking during the lecture, we posit that this represented a missed opportunity for all students to learn from the answer.

Johnson and Picciuolo (2022) investigated the frequency and function of *personal pronouns* in EMI lecturers' classroom talk in order to focus on inclusion, and compared them with lecturers' own perception of the degree of inclusion in their lecture. They found *you* to be the most frequent pronoun used by lecturers from both PS and SS, possibly

suggesting direct reference and enabling the lecturer to interact with the students (Dafouz *et al.* 2007), and thus providing evidence of a certain degree of co-operation and inclusion between lecturer and students. However, they found that *you* tended to function instead as an impersonal indexical which is not directly deictic to the hearer (Dafouz *et al.* 2007, p. 653), suggesting that lecturers are reluctant to promote bidirectional speech exchanges and have an unclear view of their own teaching practices.

Exploration of the corpus enables us to highlight different patterns of usage across the two macro disciplines, across different speaker types, and across modes, with reference to other variables such as teaching experience and language competence perhaps revealing an emerging use of English as a lingua franca in the university setting. Further research projects based on EmiBO include an investigation of lecturer translanguaging in an Italian context, while other areas of research focus on student participation in EMI lectures.

## 5. Concluding remarks and future prospects

The EmiBO corpus was conceived within the scope of a two-year research project and data presented so far refer to data transcribed in those two years. However, lecturers have shown they are increasingly willing to participate in this study, and new material is still being collected. We are currently adding new transcriptions of both audio/visual and online material, thus enabling further analysis of non-verbal semiotic features in EMI classroom interactions.

To sum up, the EmiBO corpus makes it possible to focus on lecture discourse in different modes (online and F2F), lecture discourse in different macro disciplines (Physical Sciences and Social Sciences), and lecture discourse with different speaker status (lecturer vs student). Similarities and differences across the sub-sections of the corpus may be highlighted for further analysis. Other corpora (e.g. BASE) may be used as reference to compare L1 lectures with the EMI lectures of EmiBO. The latter may also be used to investigate EMI delivered by speakers of a different L1 (e.g. in comparison with ELFA, with mainly Finnish L1 speakers). We expect analysis of these elements to bring new insights to the understanding of the nature of interactions and pedagogical practices in the EMI classroom.

The lectures in our corpus provide evidence of episodes of interaction which might be able to shed some light on how interactional multimodal strategies might differ from F2F to the online classroom; how multimodal strategies are transferred from one environment to the other; and to what extent the lecture mode or the discipline determines the way language is used and its purposes. All of this further contributes to an awareness of “EMI lecturers’ multimodal interactional competence” (Querol-Julián 2021, p. 311).

With the EmiBO corpus, we aim to make the most of the great potential of corpus linguistics in analysing different aspects of discourse in an EMI context (Jablonkai 2021). The outcomes of this research may be useful in designing teacher training tools and student support, particularly in these times of change, when online learning models are likely to predominate in the future. In this respect, further research on EMI classroom communication across different lecture modes might even pave the way for the design of a desktop videoconferencing platform for synchronous language teaching in EMI settings.

**Bionotes:** Jane Helen Johnson is Research Assistant at the Department of Modern Languages, Literatures and Cultures of the University of Bologna. A qualified teacher of English as a Foreign Language, she currently teaches English for Specific Purposes on the Master's degree course in Language, Society and Communication. Her research interests include English-Medium Instruction (EMI) and Corpus-Assisted Discourse Studies (CADS), as well as Corpus Stylistics and Translation. Her current research projects concern the teaching and learning possibilities offered by online distance learning, combining discourse analysis with EMI and ESP.

Mariangela Picciuolo is currently a PhD student in English Linguistics at the University of Bologna. Her research project focuses on English-Medium Instruction (EMI) and the analysis of discourse and interaction strategies used by lecturers in different teaching modalities – face-to-face, blended and online teaching – from a multimodal perspective. Her research interests also include: Corpus Linguistics, Computer-Mediated Communication, and Multimodal Discourse Analysis. Since 2015 she has been teaching English Language at the Department of Political and Social Sciences at the same university. She is actively engaged in several projects related to teaching and learning of new contents (multimodal literacy) and methods (MOOC).

**Authors' addresses:** [janehelen.johnson@unibo.it](mailto:janehelen.johnson@unibo.it) ; [mariangela.picciuol2@unibo.it](mailto:mariangela.picciuol2@unibo.it)



## References

- Ackerley K. and Coccetta F. 2007, *Enriching language learning through a multimedia corpus*, in “ReCALL” 19, pp. 351-370.
- Ädel A. 2010, *Just to give you kind of a map of where we are going. A Taxonomy of Metadiscourse in Spoken and Written Academic English*, in “Nordic Journal of English studies. Special issue on metadiscourse” 9 [2], pp. 69-97.
- Adolphs S. and Carter R. 2007, *Beyond the word. New challenges in analysing corpora of spoken English*, in “European Journal of English Studies” 11 [2], pp. 133-146.
- Alsop S. 2016, *The 'Humour' element in Engineering lectures across cultures: An approach to pragmatic annotation*, in López-Couso M.-J., Méndez-Naya B., Núñez-Pertejo P. and Palacios-Martínez I.M. (eds.), *Corpus Linguistics on the Move: Exploring and Understanding English through Corpora*, Brill Rodopi, Leiden, pp. 337-361.
- Alsop S. and Nesi H. 2013, *The summarising function of university Engineering lectures: a cross-cultural perspective*, in Archibald A.N. (ed.) *Multilingual Theory and Practice in Applied Linguistics: Proceedings of the 45th Annual Meeting of the British Association for Applied Linguistics, University of Southampton, 6-8 September 2012*, BAAL/ Scitsiugnil Press, pp. 11-14.
- Al-Zou'bi R. and Shamma F. 2021, *Assessing instructors' usage of emojis in distance education during the COVID-19 pandemic*, in “Cypriot Journal of Educational Sciences” 16 [1], pp. 201-220.
- Bamford J. 2004, *Gestural and Symbolic Uses of the Deictic “here” in Academic Lectures*, in Aijmer K. and Stenström A.-B. (eds.), *Discourse Patterns in Spoken and Written Corpora*, John Benjamins: Amsterdam, pp. 113-38.
- Bedenlier S., Wunder I., Gläser-Zikuda M., Kammerl R., Kopp B., Ziegler A. and Händel M. 2021, *Generation invisible? Higher Education Students' (Non)Use of Webcams in Synchronous Online Learning*, in “International Journal of Educational Research Open”, 2, pp. 1-8.
- Bellés-Fortuño B. and Fortanet-Gómez I. 2009, *Pragmatic Markers in Academic Discourse: The Cases of well and the Spanish Counterparts bien and bueno*, in Gómez Morón M., Padilla Cruz M., Fernández Amaya L. and de la O Hernández López M. (eds.) *Pragmatics Applied to Language Teaching and Learning*, Cambridge Scholars Publishing, Newcastle, UK, pp. 280-304.
- Biber D. 2006, *University language: A corpus-based study of spoken and written registers*. John Benjamins Publishing, Amsterdam.
- Biber D. and Barbieri F. 2007, *Lexical bundles in university spoken and written registers*, in “English for Specific Purposes”, 26 [3], pp. 263-286.
- Biber D., Conrad S. and Cortes V. 2004, *If you look at...: Lexical bundles in university teaching and textbooks*, in “Applied Linguistics” 25 [3], pp. 371-405.
- Biber D., Conrad S., Reppen R., Byrd P. and Helt, M. 2002, *Speaking and writing in the university: A multidimensional comparison*, in “TESOL Quarterly” 36 [1], pp. 9-48.
- Biber D., Reppen R., Clark V. and Walter J. 2001, *Representing spoken language in university settings: the design and construction of the spoken component of the T2K-SWAL Corpus*, in Simpson R.C. and Swales J.M. (eds.), *Corpus Linguistics in North America*, University of Michigan Press, Ann Arbor, pp. 48-57.
- Björkman B. 2011, *Pragmatic strategies in English as an academic lingua franca: Ways of achieving communicative effectiveness?* in “Journal of Pragmatics” 43, pp. 950-964.
- Bower M., Dalgarno B., Kennedy G. E., Lee M. J. and Kenney, J. 2015, *Design and implementation factors in blended synchronous learning environments: Outcomes from a cross-case analysis*, in “Computers & Education”, 86, pp. 1-17.
- Broggini S. and Murphy A. C. 2017, *Metadiscourse in EMI lectures: Reflections on a Small Corpus of Spoken Academic Discourse*, in Valcke J., Murphy A. C. and Costa F. (eds.), *Critical Issues in English –Medium Instruction at University, L'analisi linguistica e letteraria*, 25, pp. 325-340.
- Campagna S. and Pulcini V. 2014, *English as a Medium of Instruction in Italian Universities*, in “Textus”, 1, pp. 173-190.
- Carrillo C. and Flores M. A. 2020, *COVID-19 and teacher education: a literature review of online teaching and learning practices*, in “European Journal of Teacher Education”, 43 [4], pp. 466-487.
- Chang Y. 2011, *The use of questions by professors in lectures given in English: Influences of disciplinary cultures*, in “English for Specific Purposes”, 31, pp. 103-116.
- Cicillini S. and Giacosa A. 2020, *English-Medium Instruction Lecturers' and Students' Perceptions about the Transition from in-Person to Emergency Remote Education*, in “European Scientific Journal, ESJ”, 16 [38], pp. 46-60.
- Clark R.C. and Mayer R.E. 2011, *e-Learning and the Science of Instruction: Proven Guidelines for Consumers*

- and Designers of Multimedia Learning* (3rd Edition), Pfeiffer, San Francisco.
- Costa F. 2021, *Introduction. EMI Stakeholders and Research in the Italian Context. Moving Towards ICLHE?* in Mastellotto L. and Zanin R. (eds.), *EMI and beyond: Internationalising higher education curricula in Italy*, bu,press, Bozen/Bolzano, pp. 1-16.
- Costa F. and Coleman J.A. 2013, *A Survey of English-Medium Instruction in Italian Higher Education*, in "International Journal of Bilingual Education and Bilingualism", 16 [1], pp. 3-19.
- Crawford Camiciottoli B. 2004, *Interactive discourse structuring in L2 guest lectures: Some insights from a comparative corpus-based study*, in "Journal of English for Academic Purposes", 3 [1], pp. 39-54.
- Crawford Camiciottoli B. 2005, *Adjusting a business lecture for an international audience: a case study*, in "English for Specific Purposes", 24, pp. 183-199.
- Crawford Camiciottoli B. 2007, *The Language of Business Studies Lectures*, John Benjamins, Amsterdam/Philadelphia.
- Crawford Camiciottoli B. 2008, *Interaction in academic lectures vs. written text materials: The case of questions*, in "Journal of Pragmatics", 40 [7], pp. 1216-1231.
- Dafouz Milne E., Nunez B., Sancho C. and Foran D. 2007, *Integrating CLIL at the tertiary level: teachers' and students' reactions*, in Marsh D. and Wolff D. (eds.), *Diverse contexts - converging goals. CLIL in Europe*, Peter Lang, Frankfurt, pp. 91-102.
- Dafouz Milne E. and Sanchez García M. D. 2013, *'Does everybody understand?' Teacher questions across disciplines in English-mediated university lectures: An exploratory study*, in "Language Value", 5 [1], pp. 129-151.
- Dang T.N.Y. 2018a, *A Hard Science Spoken Word List*, in "International Journal of Applied Linguistics", 169 [1], pp. 44-71.
- Dang T.N.Y. 2018b, *The nature of vocabulary in academic speech of hard and soft-sciences*, in "English for Specific Purposes", 51, pp. 69-83.
- Dearden J. 2014, *English as a medium of instruction – A growing global phenomenon*, The British Council, London.
- Deroey K. L. B. and Johnson J.H. 2021, *Metadiscourse by 'native' and 'non-native' English speakers: importance marking in lectures*, paper presented at "3rd metadiscourse across genres conference", 27-28 May 2021, Universitat Jaume I de Castelló, Spain.
- Deroey, K. L. B. and Taverniers, M. 2011, *A corpus-based study of lecture functions*, in "Moderna Språk", 2, pp. 2-22.
- Dimova S., Hultgren A.K. and Jensen C. (eds.) 2015, *English-Medium Instruction in European Higher Education*, De Gruyter Mouton, Berlin.
- Doiz A. and Lasagabaster D. 2022, *Looking into English-medium instruction teachers' metadiscourse: An ELF perspective*, in "System", 105, pp. 1-12.
- Dudley-Evans T. 1994, *Variations in the Discourse Patterns Favoured by Different Disciplines and Their Pedagogical Implications*, in Flowerdew J. (ed.), *Academic listening: research perspectives*, Cambridge University Press, Cambridge, pp. 146-158.
- Fortanet-Gómez I. 2005, *Honoris causa speeches: An approach to structure*, in "Discourse studies", 7 [1], pp. 31-51.
- Fortanet-Gómez I. and Bellés-Fortuño B. 2005, *Spoken academic discourse: an approach to research on lectures*, in "Revista española de lingüística aplicada", 1, pp. 161-178.
- Fortanet-Gómez I. and Querol-Julián M. 2010, *The video corpus as a multimodal tool for teaching*, in Campoy M.C., Bellés Fortuño B. and Gea-Valor L. (eds.), *Corpus-based approaches to English language teaching corpus and discourse*, Continuum, London/New York, pp. 261-270.
- Friginal E., Lee J.J., Polat B. and Roberson, A. 2017, *Exploring Spoken English Learner Language Using Corpora: Learner Talk*, Springer, New York.
- Gardner S. and Xu X. 2019, *Engineering registers in the 21st century. SFL perspectives on online publications*, in "Language, Context and Text", 1 [1], pp. 65-101.
- Hellekjær G. O. 2010, *Lecture Comprehension in English-Medium Higher Education*, in "Hermes-Journal of Language and Communication Studies", 45, pp. 11-34
- Helm F., and Dooly M. 2017, *Challenges in transcribing multimodal data: A case study*, in "Language Learning & Technology", 2 [1], pp. 166-185.
- Herring S. C. and Dainas A. R. 2017, *"Nice picture comment!" Graphicons in Facebook comment threads*, in Proceedings of the Fiftieth Hawai'i International Conference on System Sciences (HICSS-50), IEEE, Los Alamitos, CA.
- Hyland K. and Tse P. 2007, *Is there an "academic vocabulary"?*, in "TESOL Quarterly", 41 [2], pp. 235-253.
- Jablonkai R. R. 2021, *Corpus linguistic methods in EMI research: A missed opportunity?*, in Pun J. and Curle S. (eds.), *Research methods in EMI*, Routledge, London, pp. 92-106.

- Jefferson G. 2004, *Glossary of transcript symbols with an introduction*, in Lerner G. H. (ed.), *Conversation Analysis: Studies from the First Generation*, John Benjamins, Amsterdam, pp.13-31.
- Jenkins J. 2014. *English as a lingua franca in the international university: The politics of academic English language policy*, Routledge, London.
- Johnson J.H. 2022. "Don't worry: when it comes to Examsville, I'm not going to ask you this" *Assessment-related expressions in Engineering lectures in an EMI and L1 context*, paper presented at "TaLC– Teaching and Language Corpora Conference", 12-15 July 2022, University of Limerick, Ireland.
- Johnson J.H. and Picciuolo M. 2020, *Interaction in spoken academic discourse in an EMI context: the use of questions*, in Proceedings of 6th International Conference on Higher Education Advances (HEAd'20), June 2-5 2020, Universitat Politècnica de València, València, pp. 211-219.
- Johnson J. H. and Picciuolo M. 2022, *Inclusività e performatività nel parlato del docente EMI: un'indagine sull'uso della deissi personale*, in Fusari, S., Ivancic, B. and Mauri, C. (eds.), *Diversità e inclusione. Quando le parole sono importanti*, Meltemi Press srl, Milan, pp. 109-129.
- Kilgariff A., Rychly P., Smrz P. and Tugwell D. 2004, *The Sketch Engine*, in Proceedings EURALEX 2004, Lorient, France, pp. 105-116.
- Kress G. 2010, *Multimodality: A social semiotic approach to contemporary communication*, Routledge, London.
- Kress G., Jewitt C., Bourne J., Franks A., Hardcastle J., Jones K. and Reid E. 2005, *English in Urban Classrooms. A multimodal perspective on teaching and learning*, Routledge, London.
- Lasagabaster D. 2022, *Teacher preparedness for English-medium instruction*, in "Journal of English-Medium Instruction", 1[1], pp. 48-64.
- Lee J. J. 2009, *Size matters: an exploratory comparison of small- and large-class university lecture introductions*, in "English for Specific Purposes", 28, pp. 42-57.
- Lim. F. V., O'Halloran K. L. and Podlasov A. 2012, *Spatial Pedagogy: Mapping Meanings in the Use of Classroom Space*, in "Cambridge Journal of Education", 42 [2], pp. 235-251.
- Luporini A. 2020, *Implementing an online English linguistics course during the Covid-19 emergency in Italy: Teacher's and students' perspectives*, in "ASp", 78, pp. 75-88.
- Lynch T. 2011, *Academic listening in the 21st century: Reviewing a decade of research*, in "Journal of English for Academic Purposes", 10, pp. 79-88.
- Martin F., Ahlgrim-Delzell L. and Budhrani K. 2017, *Systematic review of two decades (1995 to 2014) of research on synchronous online learning*, in "American Journal of Distance Education", 31 [1], pp. 3-19.
- Martinez R., Adolphs, S. and Carter R. 2013, *Listening for needles in haystacks: how lecturers introduce key terms*, in "ELT Journal", 67 [3], pp. 313-323.
- Massner C. 2021, *The Use of Videoconferencing in Higher Education*, in Pollák F., Soviar J. and Vavrek R. (eds.), *Communication Management*, IntechOpen, London, DOI: 10.5772/intechopen.99308.
- Mauranen A. 2006, *A rich domain of ELF: the ELFA Corpus of academic discourse*, in "Nordic Journal of English Studies, Special Issue: English as a Lingua Franca", 5 [2], pp.145-159.
- Mauranen A. 2012, *Exploring ELF: academic English shaped by non-native speakers*, Cambridge University Press, Cambridge.
- Mazak C. and Herbas-Donoso C. 2015, *Translanguaging Practices at a Bilingual University: a Case Study of a Science Classroom*, in "International Journal of Bilingual Education and Bilingualism", 18 [6], pp. 698-714.
- Miller C. and Parlett M. 1974, *Up to the Mark: A Study of the Examination Game*, Society for Research into Higher Education, London.
- Molino A. 2018, *'What I'm Speaking is almost English...': A Corpus-based Study of Metadiscourse in English medium Lectures at an Italian University*, in "Educational sciences: theory and practice", 18[ 4], pp. 935-956.
- Morell T. 2004, *Interactive lecture discourse for university ELF students*, in "English for Specific Purposes", 23, pp. 325-338.
- Morell T. 2007, *What enhances ELF students' participation in lecture discourse? Student, lecturer and discourse perspectives*, in "Journal of English for Academic Purpose", 6, pp. 222-237.
- Morell T. 2020, *EMI teacher training with a multimodal and interactive approach: A new horizon for LSP specialists*, in "Language Value", 12 [1], pp. 56-87.
- Morell T., Norte, N. and Beltran-Palanquez V. 2020, *How do trained English-medium instruction (EMI) lecturers combine multimodal ensembles to engage their students?*, in Roig-Vila R. (ed.) *La docencia en la Enseñanza Superior. Nuevas aportaciones desde la investigación e innovación educativas*, Ediciones OCTAEDRO, S.L., Barcelona, pp. 308-321.
- Mudraya O. 2006, *Engineering English: A lexical frequency instructional model*, in "English for Specific Purposes", 25 [2], pp. 235-256.
- Nesi H. and Basturkmen H. 2006, *Lexical bundles and discourse signaling in academic lectures*, in "International

- Journal of Corpus Linguistics”, 11, pp. 283-304.
- Northcott J. 2001, *Towards an ethnography of the MBA classroom: a consideration of the role of interactive lecturing styles within the context of one MBA programme*, in “English for Specific Purposes”, 20, pp. 15-37.
- O’Dowd R. 2018, *The training and accreditation of teachers for English medium instruction: an overview of practice in European universities*, in “International Journal of Bilingual Education and Bilingualism”, 21 [5], pp. 553-563.
- O’Halloran K. L., Tan S. and Marissa K. L. E. 2014, *Multimodal pragmatics*, in Schneider K. P. and Barron A. (eds.), *Pragmatics of Discourse*, De Gruyter Mouton, The Hague, pp. 239-68.
- O’Keefe A., McCarthy M. and Carter R. 2007, *From corpus to classroom*, Cambridge University Press, Cambridge.
- Picciuolo M. 2022, *Reconceptualising space in academic lectures: online and face-to-face lecturer discourse in the context of English-Medium Instruction*, paper presented at “Digital Genres and Open Science” International Conference, 26–27 May 2022, University of Zaragoza, Spain.
- Picciuolo M. and Johnson J.H. 2020, *Contrasting EMI lecturers’ perceptions with practices at the University of Bologna*, in Miller D. R. (ed.), “Quaderni del CeSLiC. Occasional papers AlmaDL” Centro di Studi Linguistico-Culturali (CeSLiC) e Alma Mater Studiorum, Università di Bologna, Bologna, pp. 1-23.
- Querol-Julián M. 2021, *How does digital context influence interaction in large live online lectures? The case of English-medium instruction*, in “European Journal of English Studies”, 25 [3], pp. 297-315.
- Querol-Julián M. and Crawford Camiciottoli B. 2019, *The Impact of Online Technologies and English Medium Instruction on University Lectures in International Learning Contexts: A Systematic Review*, in “ESP TODAY”, 7 [1], pp. 2-23.
- Reder S., Harris K. and Setzler K. 2003, *The Multimedia Adult ESL Learner Corpus*, in “TESOL Quarterly”, 37 [3], pp. 546-557.
- Revell A. and Wainwright, E. 2009, *What makes lectures ‘unmissable’? Insights into teaching excellence and active learning*, in “Journal of Geography in Higher Education”, 33 [2], pp. 209-23.
- Schleef E. 2008, *Gender and Academic Discourse: Global Restrictions and Local Possibilities*, in “Language in Society”, 37 [4], pp. 515-538.
- Scott M. 1997, *PC analysis of key words — And key key words*, in “System”, 25 [2], pp. 233-245.
- Scott M. and Tribble C. 2006, *Textual patterns. Keywords and corpus analysis in language education*, John Benjamins, Philadelphia.
- Seidlhofer B. 2009, *Accommodation and the Idiom Principle in English as a Lingua Franca*, in “Intercultural Pragmatics”, 6 [2], pp. 195-215.
- Simpson R. 2004, *Stylistic features of academic speech: The role of formulaic expressions*, in Connor U. and Upton T. A. (eds.), *Discourse in the Professions: Perspective on Corpus Linguistics*, John Benjamins, Amsterdam, pp. 37-64.
- Simpson R. C., Briggs, S. L., Ovens, J. and Swales, J. M. 2002, *The Michigan Corpus of Academic Spoken English*, The Regents of the University of Michigan, Ann Arbor.
- Simpson-Vlach R. and Ellis N. C. 2010, *An academic formulas list: New methods in phraseology research*, in “Applied Linguistics”, 31 [4], pp. 487-512.
- Sinclair J. 1991, *Corpus, concordance, collocation*, Oxford University Press, Oxford.
- Sinclair J. 2004, *Trust the text*, Routledge, London.
- Suviniitty J. 2012, *Lectures in English as a Lingua Franca - Interactional Features*, University of Helsinki Doctoral Dissertation, Helsinki.
- Tauroza S. and Allison D. 1990, *Speech rates in British English*, in “Applied Linguistics”, 11 [1], pp. 90-105.
- Thompson S. 1998, *Why ask questions in a monologue? Language choice at work in scientific and linguistic talk*, in Hunston S. (ed.), *Language at Work. Selected papers from the Annual Meeting of the British Association of Applied Linguistics*, University of Birmingham, September, 1997, Multilingual Matters Ltd, Clevedon, pp. 137-150.
- Wilkinson R. 2017, *Trends and issues in English-medium instruction in Europe*, in Ackerley K., Helm F. and Guarda M. (eds.), *Sharing perspectives in English-medium instruction*, Peter Lang, Frankfurt/Berne, pp. 35-75.
- Wingrove P. 2022, *Academic lexical coverage in TED talks and academic lectures*, in “English for Specific Purposes”, 65, pp. 79-94.
- Yeo J.Y. and Ting S.H. 2014, *Personal pronouns for student engagement in arts and science lecture introductions*, in “English for Specific Purposes” 34, pp. 26-37.