

# **A Corpus-informed Approach to Multimodal Pragmatics: Insights from an Exploration of Requests in B2 Spoken Language on Zoom**

Gerard O'Hanlon  
Mary Immaculate College, Ireland

*Embodied actions have a key role to play alongside spoken interaction in pragmatic development when learning a second language. Multimodal corpora have much to offer researchers, language educators and learners within pragmatics. This methodological study outlines a multimodal corpus pragmatic approach to analyse a video-mediated dataset, gathered via the videoconferencing platform Zoom. Two English language learners at B2 level participated in the study. Elicitation tasks were employed through open roleplays centred on request sequences. The data was transcribed and then manually annotated in tiers with xml coding, first for spoken request sequences, then for embodied features relating to those requests (gesture, head movement, gaze patterns and facial expression). The corpus software AntConc and the audiovisual tool ELAN were used for transcription, annotation and recall. A function-to-form corpus pragmatic approach was used to investigate the spoken and embodied resources of the participants' request sequence. Preliminary findings show that the participants augmented request sequences at key pragmatic junctures during video-mediated interaction with embodied resources. These include the use of paralinguistic (e.g., exhaling, pauses, loudness, rapid speech), gesture (literal, deictic, stress and non-literal movements), gaze patterns and gaze aversion connected to politeness, and the use of face and eyes to convey hesitation in response.*

## **Introduction**

This methodological study explores how speech act research can be broadened to include macropragmatic interactional perspectives and multimodal (spoken, visual and actional) resources using audiovisual data garnered from the videoconferencing platform Zoom. Firstly, multimodality involves the study of how different channels of meaning-making are combined into an integrated whole (Bezemer & Jewitt, 2010). Speech, paralinguistic, gesture, gaze patterns, facial expression, head movement and the use of space are core features of everyday human communication. Findings from this paper argue that these modes continue to be relevant in digital spaces. Secondly, pragmatics plays a key role in interactive digital technologies such as videoconferencing and the methods outlined below seek to explore how to carry out research in this area. Thirdly, the discourse of digital pragmatics is framed by the limitations and potentialities of videoconferencing technology, and these merit inclusion in this research. Limitations are evident with webcam framing: do the roles of gesture and gaze continue to be important in video-mediated interaction? Potentialities include the benefits of having the near simultaneous aural-actional-visual co-presence of one's interlocutors, across distance, for social interaction and meaning making. These

potentialities can be better harnessed for pragmatic teaching and language learning. In concluding, this paper proposes the extension of pragmatic research to include a multimodal corpus approach (Huang, 2021), analysing as an illustrative example a request sequence between two adult B2 learners of English in a multimodal dyadic interaction in videoconferencing contexts.

## **Background theoretical and methodological considerations**

### **Pragmatic competence**

Hymes (1972) situated the knowledge of language and the ability to use it within a construct of communicative competence. This guided later frameworks which sought to place pragmatic competence as a key component of second language acquisition. Canale and Swain (1980) and Bachman (1990) provided variations on the ability to use language in socially appropriate ways. Their models show that the acquisition of functional and sociolinguistic control of language is as essential to communicative competence as linguistic knowledge (Jones et al., 2018). Effective and appropriate communication in a target language requires a learner to develop sociopragmatic and pragmalinguistic competence (Leech, 1983). Sociopragmatic knowledge (referring to what to do, when and to whom) often proves difficult for language learners, given their lack of experience in the cultural context of the target language and a consequent lack of intuitive understanding of how people interact in target language cultures. Pragmalinguistic knowledge – knowing how language forms can be harnessed for pragmatic purposes - focuses on the forms and conventions of language use. Pragmatic competence necessitates that both types be mapped onto each other, and language users “need to be able to use the right kind of language tools given the social rules” (Roever, 2022, p. 10).

Félix-Brasdefer (2021) extends notions of pragmatic knowledge to include *macropragmatic competence*, “knowing how to behave in interaction in a variety of formal and informal communicative events” (p. 13). This perspective aligns with Roever’s description of *Interactional Competence*, “the language users’ ability to coordinate social actions with each other in mutually recognisable ways” (2022, p. 19). Given that illocutionary acts invite a response or a sequel, language learners must develop their interactional competence to coordinate their social actions with others.

Speech act sequences therefore concern the ability to produce and comprehend social actions in their sequential context—actions negotiated by two or more speakers across turns. Many features from the Conversation Analysis (CA) tradition (Schegloff, 2007) can inform interactional speech act research across sequences and inform pragmatic language pedagogy. These features include turn-taking, overlaps, conversational repair, topic management, preference organisation (how to organise conversational turns when social harmony is endangered), conversational openings and closings, and pre-speech act sequences (preambles to executing speech acts) (Schegloff, 2007). To these, Roever (2022) adds *recipient design* (how one adapts one’s talk to interlocutors, e.g., in terms of politeness). Learners demonstrate features of recipient design (Galaczi, 2014) for example, by producing topic shifts, transitions or backchannels (active listenership).

## Requests

The speech act of requesting is a directive, an illocutionary act, an attempt by the speaker to get the hearer to do something (Searle, 1976). Requests aim to engage the addressee in a future course of action, one that coincides with the requester's goals. Requests can be face-threatening, posing an imposition and intrusion into the addressee's territory. For this reason, requests tend towards indirectness and politeness (Leech, 2014), when English is the target language. Embodied actions (for example, our use of gaze, head movements or gesture) are also vital components during such events (Drew & Couper-Kuhlen, 2014).

Brown and Levinson propose strategies for carrying out face-threatening acts (1987). With directives, one can vary the level of directness utilising Brown & Levinson's superstrategies. These are presented as a three-part model, consisting of direct, indirect and non-conventional requests, in the Cross-Cultural Speech Act Realization Project (CCSARP) (Blum-Kulka et al., 1989). The level of (in)directness is further determined by the power and social distance between interlocutors and the (ranking) degree of imposition of a request (Brown & Levinson, 1987).

In terms of English language learners, developmental studies show that language learners shift from brief formulaic utterances to longer, complex speech act structures at higher proficiency levels (Leech, 2014). Developmental trajectories result in more target-like norms related to directness and indirectness at higher proficiency levels. This is due to greater automatization of learner knowledge and a better understanding of the pragmalinguistic rules and sociopragmatic norms of target cultures (Taguchi & Roever, 2017). An increased use of modification devices in and around requests is observed alongside these developmental trajectories (Alcón Soler et al., 2005). Modification devices serve to vary the degree of politeness involved in performing a speech act (Trosborg, 1995). They may be internal (e.g., intensifiers and fillers) or external (e.g., grounders or expanders) (Alcón Soler et al., 2005). Jones et al. (2018) employed a corpus analysis of L2 requests at B2 and C1 and found that learners at higher levels complemented the core head acts through past or present modals *could* or *can*. Also frequent were pre-request external moves such as mitigators (*please*) and alerters (chunks such as *excuse me* or *sorry to bother you*) alongside an increasing sensitivity to the context and the interlocutor.

## Multimodality

Human communication is a multimodal process (Adami, 2016) employing a range of spoken and embodied channels to perform visible acts of meaning. Multimodality research considers modes other than language, enabled by video recordings of human communication. Some multimodal studies are based on assumptions that speech or writing is the dominant mode (i.e., carrying the essence of meaning) with other modes expanding or modifying these meanings. Other studies foreground the significance of a specific mode (e.g., gesture) and transcribe its functions in relation to other background modes, such as speech (Bezemer & Jewitt, 2010).

Multimodal digital resources impact and influence our social, educational and professional environments, notably through videoconferencing and social media. O'Halloran et al. (2014) place pragmatics within a multimodal approach, with language operating within

the functionalities of other resources (e.g., gesture or graphical interfaces). Meaning is created through the interplay of such modes. Multimodal pragmatic approaches can allow for a wider methodological scope and the further development of classic pragmatic theories (Huang, 2021). Given the vast array of modes in digital discourses and the potential of how they might intersect with pragmatic features, research in the field of multimodal pragmatics requires further investigation.

### **Mode**

A mode is a socially shaped and culturally given semiotic resource for meaning making. Examples include speech, writing, image, layout music and 3D objects (Kress 2010). Two sets of interactional modes can be distinguished: (1) *Embodied modes*, which include gaze (Bavelas et al., 2002), gesture (McNeill, 1992), facial expression and head (Norris, 2004) and body movements and posture (Harrigan, 2005). (2) *Disembodied modes*, which include still and moving image, layout, space, sound and music (Kress, 2010). Many modes are often referred to as *nonverbal communication*, which McArthur describes as ‘the ways that human communication operates alongside and beyond verbal messages’ (2016, p. 34).

### ***Gesture***

Gesture can be broadly analysed on two axes. The first involves the gesture phase (pre, peak and post) (Kendon, 2004). The examples highlighted in this study refer to gestures as they occur at their peak phase, when they demonstrate their maximal communicative potential. The second axis involves the forms and functions of the gestures themselves. Here the focus is on the following four features: iconicity (literalness) of the gestures, their metaphoric allusions, their deictic values and their beat functions (stress and emphasis) (McNeill, 1992). Gestures which occur alongside speech (thus augmenting the communicative value of that speech) are referred to as co-speech gestures (Gullberg, 2008).

### ***Gaze, face and head***

Gaze involves the organisation, direction and intensity of looking. Gaze is often sequentially structured, with the hearer gazing at the speaker more than the speaker gazes at the hearer (Norris, 2004). Furthermore, speakers typically look toward the hearer at the end of phrases, marking the potential for turn transition (Kendon, 1967). Therefore, gaze patterns revolve around holding the floor before periodically checking in for potential turn-taking. These check-ins have been termed “the gaze window” (Bavelas et al., 2002).

Facial displays refer to timed changes in eyebrow position, expressions of the mouth, movement of the head and of the eyes. Their functions can range from grammatical functions (raised eyebrow to accent words), semantic functions (nods for feedback) and social functions (a polite smile) (Allwood et al., 2004).

### ***The affordance and transformation of mode***

First, all modes have different affordances (i.e., potentialities and limitations) for meaning making. During video-mediated interactions, the potential for speech and writing are augmented while the potential for the employment of space is limited. Second, transformation describes the processes of meaning-change through the re-ordering of the elements in a

semiotic text (Kress, 2010). The mode stays, but different terms and relations appear in the transformed text. While speech and embodiment are present in video-mediated communication, these modes are necessarily transformed to adapt to the videoconferencing software and our devices.

### **Video-mediated communication**

In recent years, a social and cultural discourse has evolved where videoconferencing platforms have become online spaces where class and meetings occur, with interlocutors adapting to these environments by employing a range of multimodal resources to navigate the discourse of video-mediated interaction. This discourse is framed by the limitations and potentialities of the technology (Jewitt et al., 2016). These include the visual limitations of the webcam frame or the internet dropouts which cause disruption to the turn taking system. Potentialities include the benefits of having the near simultaneous aural-actional-visual co-presence of one's interlocutors, across distance, for the convenience of social interaction. Video-mediated communication requires further investigation and description given that it is part of the rapidly changing multimodal digital learning environment in which the pragmatics of spoken interaction is transformed for language learners and educators.

### **Corpora, multimodality and pragmatics**

Corpora are large computerised collections of texts consisting of spoken and written language. Texts are principled, taken from real-life contexts and are often annotated with metadata (e.g., automated part-of-speech tagging). Corpora of various kinds have become fundamental to pedagogical researchers (Knight & Adolphs, 2020). The International Corpus of Learner English (Granger et al., 2020), the Trinity Lancaster Corpus (Gablasova et al., 2019) and The English Profile (Harrison & Barker, 2015) are just some of the English language learner corpora available today.

Corpus software facilitates the analysis of data through a variety of tools and methods. Many packages are open-access (e.g., *AntConc*) with built-in corpora for investigation. Software functions can display frequency wordlists, collocations and concordance lines instantaneously. Corpus linguistics research starts with a vertical reading of data, "instructing the software to plough through myriads of text samples in search of occurrences of a target item" (Rühlemann 2019, p. 7). Pragmatic research, on the other hand, involves a horizontal reading of data, "weighing and interpreting individual occurrences within their contextual environment" (Rühlemann 2019, p. 8). These two approaches are seen as complementary in certain research contexts and can be integrated within a corpus pragmatics approach (Rühlemann & Aijmer, 2015).

Corpora can be automatically marked up for part-of-speech (PoS) tagging, allowing for the recall of specific grammatical features such as verbs or nouns. Certain pragmatic features can be discerned from PoS tagged data, such as modality or discourse markers. However, there are complications in marking up speech acts (Weisser, 2018). Transcribed spoken data is often pragmatically bereft, which leaves the researcher to re-encode features pertinent to their research such as the illocutionary force or politeness strategy (O'Keeffe et al., 2020). This process can be approached manually by the researcher and is effective in smaller, more contextualised settings where the researcher has closer knowledge of the data.

This approach falls within a range of “function-to-form” approaches within corpus pragmatics (O’Keeffe, 2018). The most relevant of these to this paper is one-to-one searching (i.e., manually annotating requests to enable rapid later recall via corpus software for deeper contextualised investigation). As is discussed below, this aligns with social linguistic approaches to multimodality (i.e., qualitatively homing in on selected episodes for detailed analysis) (Bezemer & Jewitt, 2010).

Many corpora have traditionally been monomodal. Their datasets are text-based (written data and transcriptions of spoken data). Developments in technology now permit the capture, storage, annotation and exploration of large audiovisual datasets for multimodal study. This allows us to read, watch, hear and experience a dataset in greater detail. Multimodal corpora can be defined as physical repositories where records of behaviours are represented through multiple forms of media, typically aligned video, audio and textual representations of data within a single digital interface (Knight & Adolphs, 2020). Multimodal corpora permit a mixing of information of distinct modes from a given research perspective, namely digitized visual and acoustic data, textual transcriptions and annotations (Bateman, 2013). Automatic annotation of multimodal data is not yet possible, but PoS tagging and automated speech-to text software can help the researcher with the heavy lifting of the initial stages.

As noted, multimodal corpus research shares parallels with 1) multimodal social linguistics (i.e., the principled selection of episodes for further research) and 2) corpus pragmatics function-to-form traditions (i.e., manual annotation, one-to-one searching; sampling, searching and sifting) (O’Keeffe et al 2020). This is evident in Kendrick et al. (2023) who selected the opening five-minute segments of conversation initiations for the quantitative examination of the role of gesture and gaze in dyadic turn transitions. For their part, Knight et al., (in press) created a sub-corpus of extracts of starts and ends from multi-party online meetings to examine engagement through spoken and head-nod backchannels.

## **Methodology**

### **Research questions**

This paper seeks to outline a methodological approach (i.e., aligning multimodality and corpus pragmatics) to answer the following questions:

RQ1: How do B2-level English language learners perform request sequences in pairs during spoken video-mediated interactions?

RQ2: What modes do B2-level English language learners employ—and how do they employ them—during spoken video-mediated interactions?

### **Data, participants and roleplays**

This study involved the elicitation of requests from two adults at the B2 Level. Two female participants (one French L1; one Spanish L1) took part. The participants were recruited from a language school in Dublin, Ireland. The participants were resident in their

home countries at that time due to Covid 19 restrictions and were studying online. All participants signed permission forms agreeing to participate in this multimodal research project. These interlocutors are referred to as Participant A (performing the request) and Participant B (responding to that request).

Open roleplays were devised by this researcher to elicit request sequences from participants for this study (see example in Figure 1). Open roleplays are beneficial in that they result in conversational interactions (e.g., openings, closing, turn-taking), impose less researcher control and approximate authentic discourse (Kasper & Dahl, 1991). However, given that participants take on roles, there is always a risk they may not be able to properly inform that role, due to a lack of knowledge or language about the character they are supposed to be (Félix-Brasdefer, 2018). Participants A and B performed 9 roleplays in total. These were recorded via *Zoom*.

**Open roleplay (Student A)**

You need to get a specific book for your studies – *The Exam Guide*. It is expensive and you can't find a digital copy online.

You've seen a person in your class with a copy. You often say 'Hi' or 'goodbye' but don't know them very well.

You want to ask for a loan of the book for a few days to scan some chapters.

You see this person alone after class. Say hi and ask for the book!

**Open roleplay (Student B)**

You have just finished English class and a friendly classmate you don't really know comes over to say 'hello'. You are in a little bit of a rush because you want to go home and get studying with your new book (*The Exam Guide*) for an exam you have soon.

Continue the conversation.

**Figure 1.** Example of an open roleplay used in this study (devised by author).

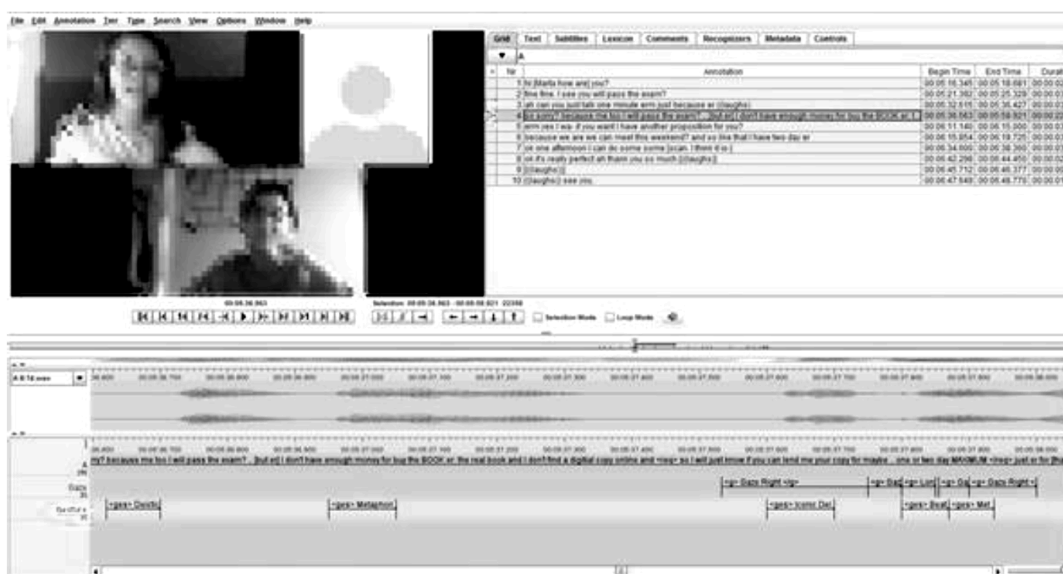
**Transcription, episode selection and annotation**

First, the spoken mode from all the roleplays was transcribed using the system devised by Diemer et al. (2016) for multimodal communication. This served as a base layer for initial searches within the corpus. These searches lead to the selection of the central request sequence (of the roleplay from Figure 1) as an episode that merited further multimodal investigation (see Results section). From a practical perspective this sequence was visible and audible, allowing for clear representation of the participants' multimodal pragmatic interaction.

*ELAN* was employed for transcription and annotation. Transcripts were exported (as .txt files) from *ELAN* to the corpus software *AntConc*. The (spoken) requests, gestures, facial

displays and gaze were manually annotated by the researcher as separate layers (tiers) using xml annotation (Kirk, 2016). The annotations were inserted using additional tiers, permitting the addition of various parallel modes (e.g., gesture, head movements, gaze patterns, facial expression) to the orthographic transcription.

Figure 2 shows the *ELAN* interface. The video player function is situated in the top left; the spoken turns of the transcriptions are visible in the top right. The acoustic waveforms run across the centre (useful for prosodic features) while the bottom segment contains tiers for the representation of speech, gaze and gesture. Note how these modes have been annotated in this sample using basic xml coding (e.g., <g> Gaze Right </g>). The transcriptions and annotations can be searched in *ELAN* for rapid indexing through the dataset.



**Figure 2.** Screenshot of *ELAN* Interface.

## Method

The multimodal methodology can therefore be summed up as involving four steps:

1. Transcription of the open role plays directly into *ELAN*
2. Exportation of the orthographic transcription from *ELAN* as a .txt file
3. Analysis of these texts with corpus software, in this case *AntConc*
4. Returning to *ELAN* to view segments of the dataset that have been identified for closer analysis through *AntConc* (i.e., exploring recurring and salient features from *AntConc* in their audiovisual contexts)

## Results

This section is taken from a roleplay dealing with the following context: Participant A wants to borrow a coursebook from Participant B for an exam. Participant B is reluctant to lend it as she needs the book herself.

The central request and its reply were selected and recalled (using *AntConc*) by the researcher as a multimodal episode. The audiovisual context was then analysed in detail (using *ELAN*) regarding which multimodal pragmatic features connected to that request sequence.

### Participant A's request

In terms of the spoken mode, Participant A modifies her speech act (Figure 3) with a range of devices, making it an indirect request. Modification occurs through two modal verbs. Firstly, “will” is commonly used with requests and in this case, it occurs with “just”, functioning to soften the periphery of the act. “Can,” present in the core of the act, is also a common feature of requests and permission. Note the participant’s use of an if-clause (“if you can”) with this modal, which functions to downtone the request further. After the request, the participant hedges her utterance (maybe) by asking for the book “for maybe one or two days MAXIMUM”. This use of “maximum” in her hedge is an attempt to minimise imposition, pointing to negative politeness.

<ONE>: but er I don't have enough money for buy the BOOK er: the real book and I don't find a digital copy online and so **I will just know if you can lend me your copy** for maybe .. one or two day MAXIMUM just er for that I can do some scan.

### Transcription Key

<ONE> Participant One

Core features of the request marked in **Bold**

Er: colon indicates elongated sound






CAPITALS – Stress/Loudness

**Figure 3.** Participant A's request.

### Gesture in Participant A's request

Gesture in Participant A's request (Figure 4) shows the utterances in Participant A's pre-request/request turn. This turn is broken down into six lexical segments, corresponding to the gestural units.

When the request in Figure 4 is viewed as a whole, the gestures become more frequent as the interlocutor closes in on the core of the request. This is evident in Segments 4, 5 and 6 where there are four gestures of differing types. While segments 1 and 2 employ longer spoken stretches, they each contain one gesture per segment, with segment 3 containing no gestures.

Segment	Speech	Gesture	Visual Unit
1	because <b>me</b> too I will pass the Cambridge exam	Points to self Both hands ( <b>Deictic</b> )	
2	but er: I don't have <b>enough money</b> for buy the BOOK	Both hands out, Palms up ( <b>Metaphoric</b> )	
3	er: the real book and I don't find a digital copy online  and so I will just know	---	---
4	if you can <b>lend me</b> your copy	Brushes hands to herself on 'lend' - ( <b>Iconic</b> )  Point to self 'me' ( <b>Deictic</b> )	
5	for maybe .. <b>one or two day</b>	Parallel hands, move up and down vertically ( <b>Beat</b> )	
6	<b>MAXIMUM</b>	Hands up, open palms out ( <b>Metaphoric</b> )	







**Figure 4.** Gestural units in Participant A's request.

Segment 2 employs a metaphoric gesture – hands out, denoting supplication as she emphasises that she has no money to purchase the book. The gestures in segments 4 and 5 are clear emphatic gestures, these being iconic (literal), deictic (reference) and beat (stress). Their function is to emphasise the spoken discourse they accompany. Segment 6 (*MAXIMUM*) is the gesture that most intertwines with politeness. It may be denoted as a metaphoric gesture

(hands up, palms out towards interlocutor) whose function is to minimise imposition and reduce the face-threat of the preceding request (“if you can lend me your copy for maybe... one or two day *MAXIMUM*”).

When mapped upon the lexical segments of her discourse, Participant A’s gaze repeatedly shifts from her device away to her right and back again. These gaze shifts occur at moments of potential turn-transition. The turn-taking concept of the gaze window (Bavelas et al., 2002) is important in this section. In video-mediated interaction, the speaker will gaze at the screen/webcam, with such devices serving as a proxy for their interlocutor.

Nonetheless, two of the segments display gaze aversion that may be explicable because of the constraints of linguistic politeness (Figure 5). In Segment 4, Participant A displays a longer than normal blink. This coincides with the beginning of how long she wants to borrow the book for. This might therefore be linked to an awareness of imposition. Furthermore, in Segment 6, we see the impact of the word “maximum” on her embodied actions. Having prompted a metaphoric gesture, this word now causes the participant to gaze away. It could be argued that this rightward gaze is part of her ‘gaze window’ patterning. Yet the grouping of spoken and embodied resources at this juncture points to its multimodal importance: the modes of paralanguage (loudness), gesture (metaphoric hands up, open palms out) and gaze (momentary rightward gaze) all combine at one crucial spoken moment (*MAXIMUM*) immediately following the key pragmatic feature (the core request) in this conversation.

Segment	Speech	Gaze	Visual Unit
1	and so I will just know if you can lend me your	Avert Gaze (her) right	
2	copy	Fix gaze interlocutor/camera	
(Potential for turn transition here)			
3	For maybe	Averts gaze (her) right	
4	One or	Eyes closed (longer than blink)	
5	Two day	Fix gaze interlocutor/camera	
(Potential for turn transition here)			
6	MAXIMUM	Averts gaze (her) right	

**Figure 5.** Gaze patterns in Participant A's request.

### **Participant B – response**

This section explores Participant B's embodied reactions and spoken response to Participant A's request. First, Participant B's use of backchannels during Participant A's request are examined. Next, Participant B's immediate, non-verbal initial reaction to the

request is detailed. Finally, Participant B's reply is examined in terms of her use of speech, paralanguage and gaze.

### **Participant B – backchannel**

Embodied backchannels, specifically continuers, pertaining to the function of maintaining the flow of discourse (O'Keeffe & Adolphs, 2008), are present as Participant B hears out Participant A's request sequence (see Figure 6). These are evidenced as the verbal continuer *yeah* (occurring once during Participant A's request) and head nods (three occurrences) during that request.

<b>Segment</b>	<b>Participant A</b>	<b>Participant B Backchannel</b>
1	so sorry? because me too I will pass the exam? .. [but er]	[Yeah] Overlap
2	I don't have enough money for buy the BOOK [er:]	[Nod] Overlap
3	the real book and I don't	
4		Nod
5	find a digital copy online	
6		Nod
7	and so I will just know if you can lend me your copy for maybe .. one or two day MAXIMUM	

**Figure 6.** Embodied backchannels and overlap.

Of the four backchannels detailed above, the first two overlap with features of Participant A's request (Segments 1 and 2). Overlap is denoted through square brackets [ ]. The verbal continuer "yeah" overlaps with Participant A's hesitator (but er) while the first of the embodied backchannel (continuer nod) overlaps with the second hesitation in the request (er). This indicates a connection between utterance-final hesitation and overlapping, embodied and verbal backchannels. The requester is reaching the end of her turn/utterance, and the listener takes advantage to nod or issue a response token without taking the floor.



Backchannels three and four occur without overlap, as nods, during brief gaps in Participant A's utterances (Segments 4 and 6). These nods occur in extremely short spaces between spoken segments and denote listenership. The three instances of nodding are examples of *beat* style nods, of minimal intrusion, a firm up-and-down movement consisting of two or three vertical strokes. They point to interactional competence, to Participant B reading and anticipating A's spoken and embodied request.

### ***Participant B's – facial display***

We have seen above how a grouping of embodied resources (speech, paralanguage, gesture and gaze) formed around Participant A's use of the word *MAXIMUM* to emphasise

hedging and negative politeness. Noteworthy here is Participant B's initial embodied reaction to that request, immediately after the word *MAXIMUM*. Participant B, surprised at the request, uses several embodied resources in rapid succession before formulating a spoken reply.

Firstly, she communicates surprise through her facial expression (Figure 7, Segment 1). In doing this she widens her eyes and raises her eyebrows. She then employs the use of space, to lean back and away from the camera in apparent exasperation. Then, this is compounded with exhalation (Segment 2). This conjunction of embodied actions is carried out rapidly, communicating an answer to the request before words are uttered; a signal that there will be a dispreferred response.

Segment	Action	Visual Unit
1	Facial reaction (Eyes, eyebrows)  → ( <i>MAXIMUM</i> )	
2	Moving back from camera (Proxemics)  Exhales	

**Figure 7.** Participant B's non-verbal reaction to a request.

#### ***Participant B's verbal reply***





Participant B starts her reply to the request with features typically indicating a dispreferred response (Figure 8). It starts off with a hesitation (*er*) followed by a paralinguistic feature associated with exhaling (*uff*) (Segment 1). Also noteworthy is the long pause (1.2 seconds) near the beginning of the response (Segment 3) and the rapid execution of *I don't know* later in her reply (Segment 5). We also see features associated with positive politeness, namely acknowledging the requester's wants (*I understand you*) (Segment 2) which is followed by an indirect offer of an alternative arrangement (*if I let you...one morning...*) (Segment 4 and 6). Overall, while Participant B doesn't immediately accede to the request, she also doesn't explicitly refuse it outright, offering instead a solution suitable to all parties.

Segment	Participant B Response	Feature
1	er uff	Hesitation follow by exhalation
2	two days it's a lot er I understand you but	Positive Politeness: acknowledging the requester's wants
3	(1.2)	Silence 1.2 seconds
4	you think you can do it a little bit faster if I let you	
5	>I don't know<	Rapidly spoken speech
6	one morning for example	Positive Politeness: indirect offer of an alternative arrangement

**Figure 8.** Participant B's spoken response.

When Participant B takes the floor, she uses gaze to regulate the interaction (Figure 9). This allows her response to be broken down into 4 segments. If we map these four gaze-reaction segments onto the range of pragmatic features detailed in her spoken reply, we can make some hypotheses about possible links in her pragmatic gaze behaviour.

The left gaze aversion in Segment 1 is longer than the gaze aversion Participant A employed during her request gaze segments. This may be due to idiosyncrasies in their gaze behaviour. Nonetheless, Participant B's long opening aversion also coincides with the beginning of her dispreferred response (i.e., hesitancy to accede). It also co-occurs with several salient spoken features, including hesitation, exhalation and a pause. She returns her gaze in Segment 2 to offer an alternative solution, averts her gaze again in Segment 3 (possibly allowing herself cognitive space to formulate the details of such an offer) before finishing by returning her gaze with the details of the offer (Segment 4).

Segment	Participant B Response	Gaze	Visual
1	er uff two days it's a lot er I understand you but (1.2)	Averts Gaze (her) left	
2	you think you can do it a little bit faster	Fixes gaze interlocut or Screen/Camera	
3	if I let you <I don't know>	Averts Gaze (her) left	
4	one morning for example	Fixes gaze interlocut or Screen/Camera	

**Figure 9.** Gaze in Participant B's reply.

## Discussion

This paper has employed a multimodal corpus pragmatic approach to analyse a sample dyadic video-mediated request sequence annotated for multimodal features including gesture, gaze, head nods, facial expression and paralinguistic. This process allows for the sequence to be described in terms of its constituent multimodal parts.

While videoconferencing software may have its limitations (limited camera framing, connection failure), it is far from restrictive to communication. Many potentialities are communicated alongside the spoken mode during videoconferencing sequences. Speech acts such as requests go beyond verbal communication alone, with the illocutionary force being expressed in this study through a myriad of modes. Participant A's multimodal ensemble shows a wide range of spoken and paralinguistic devices used to execute her request (modes and modification devices), complemented by a range of gesture and gaze strategies effected

at key points during the core head act, presenting important questions for further research about how these features align with and augment learner language and communication in relation to second language acquisition (see Gullberg, 2008). Furthermore, Participant B's response demonstrates active listenership alongside other nonverbal reactions including facial expression and the use of space within the confines of the videoconferencing lens. This again asks for more research on how learners employ nonverbal communication strategies in video-mediated spaces and about how effective this communication might be.

Moving forward, researchers face many challenges with multimodal data. The transcription and annotation of a multimodal corpus is a laborious process (Chanier & Lamy, 2017). Also, with the possibility of multiple participants, multiple viewpoints and multiple modes, multimodal datasets can become large and unwieldy (Bateman, 2013). Therefore, a focused guiding research question should always steer the research process from the early stages of the project (Jewitt et al., 2016). This also highlights the importance of finding a principled "way in" to the data, as evidenced in corpus pragmatics (function-to-form) (O'Keefe et al., 2020), social linguistic approaches to multimodality (e.g., episode selection) (Bezemer & Jewitt, 2010) and the rationalisation of this selection of episodes (see Knight et al., in press; Kendrick et al., 2023).

Eliciting data using roleplays still represents a certain degree of researcher control, resulting in contrived learner spoken production (Jucker, 2018). More natural, authentic conversational data collected from language learners, and in a variety of face-to-face and video-mediated contexts (e.g., classroom, social, professional), would permit greater description of how we behave in modern educational spaces, from a multimodal perspective (see Kress et al., 2005). Finally, multimodal research projects tend to work with smaller datasets. Larger, open-access multimodal corpora would facilitate opportunities to explore how generalised the features examined in this paper might be across learner corpora.

### **Pedagogical implications**

Advances in digital technology provide new communicative contexts and genres. This supports the case for fostering multimodal literacy within second language pedagogy. Interest in multimodal research has increased in line with the broad range of modalities present in modern learning environments (O'Halloran et al., 2016). For example, the multiliteracies approach (The New London Group, 1996) first introduced a pedagogy which promoted the idea that students develop a broader set of literacies to better engage with multiple modes in meaning making (the visual, the audio, the behavioural, etc.).

Camiciottoli & Campoy-Cubillo (2018) emphasise that work remains to be done regarding the description of what communicative modes contribute to communication, which multimodal elements need to be observed in receptive, productive and interactive contexts and how modes can be valued and assessed. A better understanding of how multimodality operates in communicative environments will provide educational stakeholders with better tools to identify, analyse, interpret, evaluate and apply multimodal communication effectively across proficiency levels. The Common Framework of Reference for Intercultural Digital Literacies (CFRIDiL) (Sindoni et al., 2019) provides a noteworthy example of multimodal metalanguage descriptors for assessment (with video-mediated interaction very much at its

core), with the aim of promoting change in teaching, learning and assessment within intercultural digital literacies.

Corpora are most strongly connected to language education through data-driven learning (DDL) (Johns, 2002), an inductive approach designed to help learners explore concordance lines and form generalisations about the target language. DDL involves a range of approaches and resources which have kept pace with methods and technological developments in teaching (Viana, 2022). O’Keeffe et al. (2020) point to how corpora can offer pragmatic insights into areas such as negative politeness or vague language, insights which can be extended by utilising audiovisual content and contexts. Multimodal corpora can also inform classroom materials and lesson design, offering potential for implicit and explicit pragmatic instruction. Ruiz-Madrid & Valeiras-Jurado (2020) contend that a multimodal pedagogy that goes beyond language has the potential to promote alternative ways of reading, interpreting and composing semiotic texts.

## Conclusion

Findings from this paper show that it is possible to combine pragmatics, corpus linguistics and multimodality when exploring verbal and non-verbal features connected to spoken request sequences in an online video-mediated dataset of English language learners. More specifically, analysis through corpus and audiovisual software permitted detailed observation of how learners interact during request sequences and how they use paralinguistic, gaze, gesture, face and head movement as they do so. A multimodal corpus approach showed how these embodied actions interconnect with a range of pragmatic features connected to requests, e.g., gaze aversion, heightened gesture usage and features of politeness. Larger datasets using a multimodal corpus pragmatics approach will help to generalise and further test the validity of these findings.

## References

- Adami, E. (2016). Multimodality. In O. Garcia, N. Flores, & M. Spotti (Eds.), *Oxford handbook of language and society* (pp. 451–472). Oxford University Press.
- Alcón Soler, E., Safont Jordà, P., & Flor, A. (2005). Towards a typology of modifiers for the speech act of requesting: A socio-pragmatic approach. *RAEL: Revista Electrónica de Lingüística Aplicada*, 4, 1–35.
- Allwood, J., Cerrato, L., Dybkjaer, L., Jokinen, K., Navaretta, C., & Paggio, P. (2004). The MUMIN Multimodal Coding Scheme. *NorFA Yearbook 2005*, 129–157.
- Anthony, L. (2022). *AntConc* (Version 4.1.4). Waseda University.  
<https://www.laurenceanthony.net/software>
- Bachman, L. F. (1990). *Fundamental considerations in language testing* (5th ed.). Oxford University Press.
- Bateman, J. A. (2013). Multimodal corpus-based approaches. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 3983–3991). Wiley Blackwell.

- Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, 52, 566–580. <https://doi.org/10.1111/j.1460-2466.2002.tb02562.x>
- Bezemer, J., & Jewitt, C. (2010). Multimodal analysis: Key issues. In L. Litosseliti (Ed.), *Research methods in linguistics* (pp. 180–197). Continuum.
- Blum-Kulka, S., House, J., & Kasper, G. (Eds.). (1989). *Cross-cultural pragmatics: Requests and apologies*. Ablex.
- Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.
- Camiciottoli, B. C., & Campoy-Cubillo, M. C. (2018). Introduction: The nexus of multimodality, multimodal literacy, and English language teaching in research and practice in higher education settings. *System*, 77, 1–9. <http://dx.doi.org/10.1016/j.system.2018.03.005>
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <https://doi.org/10.1093/applin/I.1.1>
- Chanier, T., & Lamy, M. (2017). Researching technology-mediated multimodal interaction. In C.A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 428–443). Wiley Blackwell.
- Diemer, S., Brunner, L., & Schmidt, S. (2016). Compiling computer-mediated spoken language corpora: Key issues and recommendations. *International Journal of Corpus Linguistics*, 21(3), 349–371. <https://doi.org/10.1075/ijcl.21.3.03die>
- Drew, P., & Couper-Kuhlen, E. (2014). Requesting – from speech act to recruitment. In P. Drew & E. Couper-Kuhlen (Eds.), *Requesting in social interaction* (pp. 1–35). John Benjamins.
- ELAN (Version 6.4) (2022). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. <https://archive.mpi.nl/tla/elan>
- Félix-Brasdefer, J. C. (2018). Role plays. In A. H. Jucker, K. P. Schneider, & W. Bublitz (Eds.), *Methods in pragmatics* (pp. 305–331). Mouton de Gruyter.
- Félix-Brasdefer, J. C. (2021). Pragmatic competence and speech-act research in second language pragmatics. In J. C. Félix-Brasdefer & R. L. Shively (Eds.), *New directions in second language pragmatics* (pp. 11–26). Mouton de Gruyter.
- Gablasova, D., Brezina, V., & McEnery, T. (2019). The Trinity Lancaster Corpus: Development, description, and application. *International Journal of Learner Corpus Research*, 5(2), 126–158. <https://doi.org/10.1075/ijlcr.19001.gab>
- Galaczi, E. D. (2014). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics*, 35(5), 553–574. <https://doi.org/10.1093/applin/amt017>
- Granger, S., Dupont, M., Meunier, F., Naets, H., & Paquot, M. (2020). *The International Corpus of Learner English* (Version 3). Presses Universitaires de Louvain.
- Gullberg, M. (2008). Gestures and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 286–315). Routledge.
- Harrigan, J. (2005). Proxemics, kinesics and gaze. In J. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research* (pp. 137–198). Oxford University Press.
- Harrison, J., & Barker, F. (Eds.) (2015). *English profile in practice* (Vol. 5). Cambridge University Press.
- Huang, L. (2021). Toward multimodal corpus pragmatics: Rationale, case, and agenda. *Digital Scholarship in the Humanities*, 36(1), 101–114. <https://doi.org/10.1093/dsch/hgaa080>

- Hymes, D. (1972). Models of interaction of language and social life. In J. J. Gumperz & D. H. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 35–71). Holt, Rinehart and Winston.
- Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing multimodality*. Routledge.
- Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus linguistics* (pp. 105–117). Rodopi.
- Jones, C., Byrne, S., & Halenko, N. (2018). *Successful spoken English: Findings from learner corpora*. Routledge.
- Jucker, A. H. (2018). Data in pragmatic research. In A. H. Jucker, K. P. Schneider, & W. Bublitz (Eds.), *Methods in pragmatics* (pp. 3–36). Mouton de Gruyter.
- Kasper, G., & Dahl, M. (1991). *Research methods in interlanguage pragmatics*. University of Hawai'i Press.
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26, 22–63. [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4)
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Kendrick, K. H., Holler, J., & Levinson, S. C. (2023). Turn-taking in human face-to-face interaction is multimodal: Gaze direction and manual gestures aid the coordination of turn transitions. *Philosophical Transactions of the Royal Society*. <https://doi.org/10.1098/rstb.2021.0473>
- Kirk, J. M. (2016). The pragmatic annotation scheme of the SPICE-Ireland Corpus. *International Journal of Pragmatics*, 21(3), 299–322. <https://doi.org/10.1075/ijcl.21.3.01kir>
- Knight, D., & Adolphs, S. (2020). Multimodal corpora. In M. Paquot & S. Th. Gries (Eds.), *A practical handbook of corpus linguistics* (pp. 353–371). Springer International Publishing.
- Knight, D., O'Keeffe, A., Fitzgerald, C., Mark, G., & McNamara, J. (In press). Indicating engagement in online workplace meetings: The role of backchannelling head nods. *International Journal of Corpus Linguistics*.
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Kress, G., Jewitt, C., Bourne, J., Franks, A., Hardcastle, J., Jones, K., & Reid, E. (2005). *English in urban classrooms: A multimodal perspective on teaching and learning*. Routledge.
- Leech, G. (1983). *Principles of pragmatics*. Longman.
- Leech, G. (2014). *The pragmatics of politeness*. Oxford University Press.
- McArthur, J. A. (2016). *Digital proxemics: How technology shapes the ways we move*. Peter Lang.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66, 60–92. <https://doi.org/10.17763/haer.66.1.17370n67v22j160u>
- Norris, S. (2004). *Analysing multimodal interaction: A methodological framework*. Routledge.
- O'Halloran, K. L., Tan, S., & Marissa, K. L. E. (2014). Multimodal pragmatics. In K. P. Schneider & A. Barron (Eds.), *Pragmatics of discourse* (pp. 239–270). Mouton de Gruyter.
- O'Halloran, K. L., Tan, S., & Smith, B. A. (2016). Multimodal approaches to English for academic purposes. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 256–269). Taylor & Francis Group.
- O'Keeffe, A. (2018). Corpus-based function-to-form approaches. In A. H. Jucker, K. P. Schneider, & W. Bublitz (Eds.), *Methods in pragmatics* (pp. 587–618). Mouton de Gruyter.
- O'Keeffe, A., & Adolphs, S. (2008). Response tokens in British and Irish discourse: Corpus, context, and variational pragmatics. In K. P. Schneider & A. Barron (Eds.), *Variational pragmatics: A focus on regional varieties in pluricentric languages* (pp. 69–98). John Benjamins.
- O'Keeffe, A., Clancy, B., & Adolphs, S. (2020). *Introducing pragmatics in use*. Routledge.

- Roever, C. (2022). *Teaching and testing second language pragmatics and interaction: A practical guide*. Routledge.
- Rühlemann, C. (2019). *Corpus linguistics for pragmatics*. Routledge.
- Rühlemann, C., & Aijmer, K. (2015). Corpus pragmatics: Laying the foundations. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics: A handbook* (pp. 1–26). Cambridge University Press.
- Ruiz-Madrid, N., & Valeiras-Jurado, J. (2020). Developing multimodal communicative competence in emerging academic and professional genres. *International Journal of English Studies*, 20(1), 27–50. <https://doi.org/10.6018/ijes.401481>
- Schegloff, E. A. (2007). *Sequence organization in interaction* (Vol. 1). Cambridge University Press.
- Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, 5, 1–23.
- Sindoni, M. G., Adami, E., Karatza, S., Marenzi, I., Moschini, I., Petroni, S., & Rocca, M. (2019). *The common framework of reference for intercultural digital literacies*. <https://www.eumade4ll.eu/common-framework-of-reference-for-intercultural-digital-literacy/>
- Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford University Press.
- Trosborg, A. (1995). *Interlanguage pragmatics: Requests, complaints and apologies*. Mouton de Gruyter.
- Viana, V. (Ed.) (2022). *Teaching English with corpora: A resource book*. Routledge.
- Weisser, M. (2018). *How to do corpus pragmatics on pragmatically annotated data: Speech acts and beyond*. John Benjamins.

### **Recommended APA citation**

- O'Hanlon, G. (2025). A corpus-informed approach to multimodal pragmatics: Insights from an exploration of requests in B2 spoken language on Zoom. In M. González-Lloret, J. M. Sykes, & J. K. Yoshioka (Eds.), *Pragmatics & Language Learning* (Vol. 15, pp. 1–21). National Foreign Language Resource Center, University of Hawai'i. <http://hdl.handle.net/10125/75347>