RiCL Research in Corpus Linguistics

"We've lost you Ian": Multi-modal corpus innovations in capturing, processing and analysing professional online spoken interactions

Anne O'Keeffe[a] – Dawn Knight[b] – Geraldine Mark[b] – Christopher Fitzgerald[a] – Justin McNamara[a]
Svenja Adolphs[c] – Benjamin Cowan[d] – Tania Fahey Palma[e] – Fiona Farr[f] – Sandrine Peraldi[d]

Mary Immaculate College, University of Limerick[a] / Ireland
Cardiff University[b] / United Kingdom
University of Nottingham[c] / United Kingdom
University College Dublin[d] / Ireland
University of Aberdeen[e] / United Kingdom
University of Limerick[f] / Ireland

Abstract – Online communication via video platforms has become a standard component of workplace interaction for many businesses and employees. The rapid uptake in the use of virtual meeting platforms due to COVID-19 restrictions meant that many people had to quickly adjust to communication via this medium without much (if any) training as to how workplace communication is successfully facilitated on these platforms. The *Interactional Variation Online* project aims to analyse a corpus of virtual meetings to gain a multi-modal understanding of this context of language use. This paper describes one component of the project, namely guidelines that can be replicated when constructing a corpus of multi-modal data derived from recordings of online meetings. A further aim is to determine typical features of virtual meetings in comparison to face-to-face meetings so as to inform good practice in virtual workplace interactions. By looking at how non-verbal behaviour, such as head movements, gaze, posture, and spoken discourse interact in this medium, we both undertake a holistic analysis of interaction in virtual meetings and produce a template for the development of multi-modal corpora for future analysis.

Keywords – Online workplace communication; corpus pragmatics; multi-modal corpus linguistics; corpus construction; transcription

1. Introduction[1]

The pandemic has acted as a catalyst for change and has impacted on the behaviours of producers and consumers of digital interactional content. Businesses have changed their

[1] This research/project was funded by the *Arts and Humanities Research Council* (UKRI-AHRC) and the *Irish Research Council* (IRC) under the *UK-Ireland Collaboration in the Digital Humanities Research Grants Call* (grant numbers AH/W001608/1 and IRC/W001608/1).

interaction with customers, relying more on social media to reinforce their brand image (Ibrahim and Aljarah 2023); cultural organisations have embraced different forms of digital delivery of content, often co-produced by their audiences, such as book readings broadcast online as well as in-person; education has seen the large-scale adoption of online modes of instruction and interaction through both synchronous and asynchronous means of tuition. The list goes on.

There is a need now to examine whether existing paradigms for analysing verbal and non-verbal discourse in both face-to-face and virtual contexts are fit-for-purpose and develop technical protocols for capturing and analysing online multi-modal interaction. The *Interactional Variation Online* project (IVO)[2] draws on the expertise of researchers and their collaborators in the UK and Ireland to evolve standardised ways of approaching questions about multi-modal communication that are accessible and (re)producible by other researchers and non-technical experts. These will inform research practice relating to the gathering, storage, processing and analysis of multi-modal data through community-building aspects of the project for multi-modal corpus linguistic research. This paper outlines the phases of corpus design and construction undertaken by the IVO project, including:

1. Surveying and partner engagement
2. Establishing a design frame
3. Data collection
4. Transcription
5. Coding and mark-up
6. Establishing an analytical framework.

Innovation in corpus linguistics is inextricably linked to technological developments that drive changes in corpus construction and analysis. Concurrently, analysis of human interaction in and with digital technologies is an ongoing concern of researchers in the digital humanities (Mackenzie 2020). Until recently, capturing video recordings of business meetings with each participant individually framed, for the specific purpose of looking at verbal and non-verbal behaviour, would have required vast amounts of hardware and intrusion on the meeting space (see Knight and Adolphs 2020). Now that virtual meetings have proliferated and become normalised, particularly during and since COVID-19, this

[2] https://ivohub.com/

type of data is easily captured using the users' own hardware. This, in turn, has provided a capacity to analyse non-verbal communication in meetings without the need for a laboratory-type set-up with multiple cameras to capture audio and video of all participants. The past two decades have seen early developments of multi-modal spoken corpora and analysis software, and many areas in applied linguistics have begun to investigate modes other than speech, e.g., in social semiotics (Harrison 2003), multi-modal (inter) action analysis (Cohn 2016), conversation analysis (Mondada 2019) and gesture studies (Cienki 2016). These studies are making advances in establishing a greater understanding of the interconnected network of modes that construct meaning (Levinson and Holler 2014; Holler and Levinson 2019).

Within Multi-Modal Corpus Linguistics (MMCL), a multi-modal corpus aligns multiple discursive modes (e.g., textual transcription, video and/or audio data), and provides the tools to examine interaction within and between different modalities in the generation of meaning. Allwood (2008: 210) provides the following rationale for collecting and analysing multi-modal corpora: "they provide material for more complete studies of 'interactive face-to-face sharing and construction of meaning and understanding' which is what language and communication are all about." Most current multi-modal corpora are 'specialised', so built to examine a particular discursive context, such as meeting rooms (Friedland *et al.* 2009), academic supervisions (Knight and Adolphs 2008), political interviews (Trotta *et al*. 2020), and/or to meet the requirements of a particular research area/project. There currently exist no 'general' large-scale multi-modal corpora, with data from a range of discursive contexts and/or socio-demographic groups, and few of the corpora that do exist are freely available to the research community. The CLARIN website[3] provides links to some of those multi-modal corpora that are accessible. Over a decade ago, when reflecting on the future for multi-modal corpora, Knight (2011) outlined a range of methodological and technical issues and challenges faced by researchers in MMCL, the majority of which remain pertinent today. Annotating and analysing multi-modal corpora remains an expensive, time-consuming and technically complex process. However, the proliferation of workplace communication via virtual post COVID-19 pandemic meetings gives rise to an opportunity for multi-modal corpus construction that we aim to illustrate in this paper.

[3] https://www.clarin.eu/resource-families/multimodal-corpora

2. User-driven design

For the IVO project, the team adopted a user-driven approach to the research and corpus design (i.e., one in which practitioners and end-users co-construct the design from the start to ensure that it has "relevance and application to real-world problems and uses beyond the academic context" (Knight *et al.* 2021: 44)). To achieve this, we looked beyond the immediate research team to gain a baseline understanding of the general population's working behaviours, and their perceptions of working online during the COVID-19 pandemic. This was undertaken via an online survey[4] which was circulated to the project partners and their networks and publicised via social media platforms. The survey attracted 371 responses from individuals working in a range of vocations including academic, pharmaceuticals, finance, real estate, IT, media, the creative arts, medicine and for charitable organisations, of whom 54 per cent were from Ireland, 20 per cent from the UK, 18 per cent from Malta and 8 per cent from other locations. Likewise, 54 per cent defined themselves as female and 46 per cent male. Age ranges of respondents are shown in Table 1.

Age range	**Percentage**
18–24	11
25–34	26
35–44	20
45–54	20
55–64	21
65+	2

Table 1: Age ranges of survey respondents

Results showed a substantial increase in online meetings during the pandemic, with 41 per cent of respondents saying that they never had online meetings prior to the pandemic. Just 3 per cent of respondents said that they never had online meetings at the time of the survey (January 2022, emerging from pandemic restrictions). When asked whether they preferred face-to-face or online meetings for specific types of meetings (e.g., whole organisation meetings, social events), most respondents (76%) were in favour of face-to-face social events but were happy to have other meetings virtually. Connected to this, there was a strong sense of loss of social interaction in online environments, something that is seen as more pervasive in face-to-face interaction. This is also noted in Milz *et al.*'s (2023) study of online public planning meetings during COVID-19.

[4] https://ivohub.com/wp-content/uploads/2022/10/IVO_Baseline_Infographic.pdf

Free-text responses to the question 'what doesn't work well in virtual meeting environments?' included: 1) 'less informal interaction'; 2) 'you can't pick up the mood music of the room'; 3) 'difficult to build team spirit'; 4) 'no face-to-face, presenter cannot see facial expressions'.

Whilst the survey gained a relatively small number of respondents, with most from just three English-speaking countries (Ireland, United Kingdom and Malta), some useful insights into the broad preferences towards certain platforms for different types of work interactions were gained. It also helped to highlight the perception that specific platforms are chosen depending on the relative formality of the event, for example, respondents showed a preference for *Zoom*[5] in social meetings and for *Microsoft Teams*[6] in team meetings. These results also show both positive and negative sentiment towards virtual meetings. They underscore the desire for an increase in the social interaction which is lost in this environment and a maintenance of the convenience that is gained through online meetings. Our results tally with early studies on the efficacy of virtual meetings (e.g., Panteli and Dawson 2001) and overlap with some of the findings in Milz *et al.* (2023), such as a preference for holding large team meetings online rather than face-to-face.

3. Challenges and considerations

3.1. Developing multi-modal corpora

Research in this space faces on-going challenges relating to the forms of data to be included, namely:

1. The modalities to be captured and represented, i.e., what hardware to use to track gaze direction.
2. Where to source the data (and how).
3. The format for storage, i.e., which encrypted shared platforms to use.
4. The method of transcription and coding and deciding on whether speech-to-text tools are preferable to manual transcription.

[5] https://zoom.us/
[6] https://www.microsoft.com/en-us/microsoft-teams/log-in

5. The best way to align annotations and different modalities in a meaningful way to map temporal and/or semiotic relationships between these, e.g., deciding on a tool like *ELAN*[7] to create a system of co-occurring tiers.

Some reflections on these elements are discussed below, with further guidance provided on our website.[8]

3.2. Data collection

3.2.1. Establishing a sampling frame

The starting point for any corpus project is to define 'what' data is to be included/recorded and, important in the case of multi-modal corpora, 'how' it is to be recorded. This scoping process is scaffolded using a corpus design frame, also known as a 'sampling frame', and defined by Knight *et al.* (2021) as a rubric that specifies which texts, from which genres, and in what proportions are to be sampled for use in a corpus. Design frames ensure that the data collection is principled so that the resultant corpus provides, as far as possible, an accurate representation of the communicative contexts it seeks to capture (and represent).

To construct a design frame for a corpus of online meetings, we needed to define both 'meeting' and 'agenda' in terms of how we intend to use these terms for our purposes. Schwartzman (1989: 7) defines a meeting as "a communicative event involving three or more people who agree to assemble for a purpose ostensibly related to the functioning of an organization or a group." We extend this definition to include meetings of two people, to account for dyadic interaction, those which are agenda-driven, and which take place in virtual environments. The purpose of an agenda, according to Svennevig (2012: 54), is to provide "the participants with a 'template' for the topics to be addressed and the activities to engage in during the meeting." For the purposes of this project, the design frame we adopted involves data from online meetings which we define as communicative events, involving two or more people who agree to assemble online for a purpose, with a predetermined formal or informal agenda, related to the functioning of an organisation or a group.

Decisions regarding what should be included in a design frame, and its associated design taxonomy (i.e., its explicit categorisation framework), are typically driven by the

[7] https://archive.mpi.nl/tla/elan
[8] https://ivohub.com/resources/

specific aims of the corpus/associated research project. Efforts have been made to establish frameworks for representativeness and corpus design (e.g., Egbert *et al.* 2022) and there is a general understanding that any taxonomy used "must be consistent and transparent so that corpus users can navigate the corpus with ease" (Knight *et al.* 2021: 28). The development of a design frame is often iterative and dynamic, undergoing changes as the context is understood more while data is being collected. While this leads to a design frame that is more tailored to the dataset as it is collected, this requires detailed documentation and justification throughout the process. The initial design frame provided criteria for data sources that were seen as essential variables in the construction of a corpus of virtual meetings for the IVO project. These encompassed three broad factors: 1) the meeting type (e.g., team meetings, one-to-one meetings), 2) the sector (public/private) and 3) the meeting context or goal (e.g., transactional, pedagogic, team updates).

Subsequently, as we collected recordings to be included in the IVO corpus, it became apparent that a focus on the sector categories was needed, as it would provide a design frame that would encompass a spread of team meetings of various sizes and configurations with enough variables in terms of goals (desired outcomes of a meeting) and context (organisational setting) to create a corpus that would suit the aims of the research project. From pursuing early versions of sampling frames that were focused on meeting type and context, we had inherently acquired data that fell into multiple categories but were biased towards the public sector. The adoption of a private/public sector-focused framework (as these categories are defined by Esteve and Ysa 2011) led us to pursue access to data aligning with these categories (see Table 2). To this end, following a review of industry categories based on data collected, we designed a sampling framework that would take, as its principle, the categorisation set out in Table 2.

Sector	**Organisation types**
Private	Designated activity companies, sole proprietorships, partnerships, limited liability companies and considerations.
Public	Educational institutions, NGOs, government and health.

Table 2: Private/public sector-focused framework for data collection

To include a spread of organisational types, we aimed to collect data from each of the categories (in Table 2), although no set wordcount or prescriptions for balance were defined at the start, as these were likely to be somewhat driven by opportunism. We were essentially open to receiving whatever data was offered by those we contacted, and no prerequisites

were provided by the research team aside from the meetings taking place over video conferencing software. Thus, no prescriptions were established regarding the overall size of the corpus, the topics discussed within given meetings, nor the optimal lengths of recordings. Establishing targets is certainly advisable for corpus projects that intend to be large in scale, however. When building the *Spoken British National Corpus* 2014 (BNC2014; Love 2020), for example, it was vital that participants were asked "to make recordings of no less than 20 minutes in length" (Love 2020: 45) as it would have been near impossible to reach the *Spoken BNC2014*'s >11-million-word target in a timely way with shorter excerpts. By not imposing a time limit or restraint on the meetings recorded, the data is also more natural and authentic in style and that collection is therefore driven by participants, reflecting the actual process of online meetings.

3.2.2. Recording practicalities

When building spoken and multi-modal datasets in non-virtual environments, decisions regarding 'how' data is to be recorded also need close consideration. Access to equipment for recording and data storage can often be a challenge, and practical aspects such as 'where' to position the equipment, where participants will be in relation to this, and so on, need consideration. We posit that the digital pivot has certainly afforded a more streamlined approach here, whereby the decision making is somewhat not by the 'researcher' but the 'researched'.

In terms of camera settings, we were not prescriptive and essentially accepted all variants from those including recordings where some participants had their cameras off or had their audio muted. We also collected video-only and hybrid options as this reflects the reality of participant behaviour in virtual meetings. This variability, and the inclusion of multiple parties in the talk, increased the complexity and richness of the data, which needed to be factored into the time dedicated for annotation and analysis. This is because, as Goodwin (1994: 607) states

> like transcription, any camera position constitutes a theory about what is relevant within a scene - one that will have enormous consequences for what can be seen in it later - and what forms of subsequent analysis are possible.

The only requirement we did have was for data to be recorded by a representative from each of the meetings, using the built-in recording functionality of the given videoconferencing

software used (i.e., *Teams* or *Zoom*). Data could then be easily shared with members of the project team and was stored securely for subsequent analysis (guidelines on how to do this are available on the project website).[9] Such functionalities, and ever-increasing access to extensive cloud and desktop storage solutions, again, makes this stage of the corpus development process far quicker and easier than in face-to-face recording contexts which have resulted in other multimodal corpora such as the dyadic and triadic conversations that are the components of the *Freiburg Multimodal Interaction Corpus* (FreMIC; Rühlemann and Ptak 2023). However, a reliance on third-party software and the internet connectivity of participants can lead to technical issues, such as video and audio drop-out, resulting in participants relying on phrases such as that in the title of this paper to highlight such deficits to other participants.

3.2.3. Ethical considerations

Formal written consent was received from all participating organisations and, where possible, individuals as *a priori* for the development of the IVO corpus, and permission to re-use images/screenshots used in this paper were acquired from those participants who feature here. In cases where data were already in the public domain (for example, on company *YouTube* channels), explicit permission was granted from the organisations who made the data public and, where possible, the individuals participating in the recordings were contacted to request their consent. This is in line with best practice and in accordance with guidelines for best practice, such as that produced by the *British Association for Applied Linguistics* (BAAL 2021). Unfortunately, due to restrictions in copyright, publication and distribution, only excerpts of the dataset of the IVO corpus are publicly available for other users. The lack of availability and reusability of multi-modal corpora is an on-going issue within the field (see Knight and Adolphs (2020) and Knight (2011) for further discussions on this).

3.3. Orthographic transcription

Denham and Onwuegbuzie (2013) list four elements of spoken language as likely lost in transcriptions: 1) proxemics (the interpersonal space in the communication), 2) chronemics (the speed of the delivery and the length of silences), 3) kinesics (body language

[9] https://ivohub.com/resources/

and posture) and 4) paralinguistics (including volume, pitch and voice quality). Though the IVO corpus preserves much of the visual and auditory content of the original event, orthographic transcription is still required to enable searchability of spoken items after the corpus is constructed. While no agreed standard for transcription necessarily exists (i.e., the 'what' of transcription), shared practices are common across general spoken corpora as, for instance, the *Cambridge and Nottingham Corpus of Discourse in English* (CANCODE: Carter and McCarthy 2004) and/or national corpora with spoken components, like the *Spoken BNC2014*). In these cases, the "value" of spoken corpora is partly in revealing the "normal dysfluency" of speech (Biber *et al.* 1999: 1048), so there is an emphasis on transcribing *verbatim*, i.e., without standardising the content. This approach was also taken by the IVO team, using an adapted version of the CANCODE conventions (see website resource on transcribing multi-modal data).[10]

The actual process of transcription (i.e., the 'how' of transcription) has been noted as being a particularly time-consuming and arduous task (Knight and Adolphs 2022). As Lin and Chen (2020: 72) note, it can take "an hour to annotate the intonation and rhythm patterns in a single minute of speech," and "a further hour of video to conduct a detailed annotation for one minute of video." This is on top of the time taken to transcribe speech orthographically, whereby an hour of speech is estimated to take a trained researcher up to 14 hours (two working days) to transcribe (O'Keeffe *et al.* 2007). To speed up the process of transcription, the affordances of using speech-to-text and automatic captioning technologies have been explored by developers of spoken corpora. Love (2020: 104–107), for example, experimented with the use of a beta version of *Trint*,[11] an automatic speech-to-text transcription and editing tool, when constructing the *Spoken BNC2014*. He discovered that whilst the "time alignment and editing functionalities of the tool were very good, the accuracy of transcription appeared to be very low" (Love 2020: 107), with a "poor ability to separate turns according to the speakers who produced them" (*ibid*: 107). Love (2020) also tested other similar tools but concluded that, at the time of developing the *Spoken BNC2014*, they were all unfit for purpose as they did not produce fine-grained accurate outputs that are required for linguistic analysis.

In light of the 'digital pivot' and the increasing number and ubiquity of speech-to-text tools, and the fact they are now integrated directly into the main online meeting software

[10] https://ivohub.com/transcribing-mm-data/
[11] https://trint.com/

Microsoft Teams and *Zoom*, for example, it seemed appropriate to revisit the potential for using speech-to-text tools here. To this end, *Otter.ai*[12] was used in the first instance to generate a 'first-pass' of the collected data. This was then carefully checked and edited through close viewing and listening. *Otter.ai* is oriented towards creating transcriptions that prioritise legibility and coherence rather than preserving all elements of the original speech. In the process of automated transcription, items such as backchannels, repetitions and hesitations (for example, *uh, um, ah*) are omitted and so require manual addition by the analyst during the checking phase. In addition, for some recordings featuring strong regional or national accents, the accuracy of the transcription was low and, as a result, required a lot more manual input/editing. Despite these shortcomings, *Otter.ai* proved effective for turn separation and time alignment, offered ease of editing in its user interface, and generally increased the speed of transcription, so it was deemed more of benefit than *cost* to use.

The speed of transcription checking per minute is strongly determined by the transcript that is being reviewed and the elements of the transcript that are determined as necessary elements of the review process. For the IVO project, these elements of review were necessary but time-consuming components of the process, and they include 1) checking for accuracy and editing accordingly; 2) inserting fillers (such as *uh* and *um*) which *Otter.ai* is programmed to ignore; 3) inserting symbols and codes for items like interruptions, coughs and non-verbal sounds; and 4) anonymising any content that might reveal the identity of the participants or organisations involved.

3.4. Coding and mark-up

As non-verbal behaviours (such as gesture, gaze, posture, head nodding) are not readily analysable units, annotation is required as a precursor to the analysis phase. As noted by Allwood *et al.* (2007b: 274), "annotation schemes often reflect the specific requirements that drive the creation of such a [multi-modal] corpus" and these different needs and requirements often result in the use of bespoke coding schemes for marking-up non-verbal behaviours. Despite this lack of universality, there are broadly two types of schemes: 1) those which focus on form and 2) those which focus primarily on communicative function.

Form-based schemes typically concentrate on marking up non-verbal behaviour purely in kinesic terms, capturing, for example, the size, shape and relative position of sequenc-

[12] https://otter.ai/

es of movements that form non-verbal behaviours. Examples of these include McNeill's (2000) *Gesture Phase Coding Scheme*, which allows the modelling of a range of bodily movements, but predominantly concentrates on defining sequences of hand movement, and Ekman and Friesen's (1978) *Facial Action Coding Scheme* (FACS), which focuses on classifying facial expressions through the movement of specific facial muscles, known as *Action Units* (AUs). Other schemes include the *Bielefeld Speech and Gesture Alignment Corpus* (SaGA; Lücking *et al.* 2010), the *REmote COL-laborative and Affective Interactions Corpus* (RECOLA; Ringeval *et al.* 2013) and *Video-mediated English as a Lingua Franca Conversations Corpus* (ViMELF; Brunner and Diemer 2021), which are designed to code gestures and signs which co-occur with talk.

Function-based schemes, which are more relevant to this current research, enable the mark-up of the semiotic and/or pragmatic relationship between verbalisations and non-verbal behaviour, that is, the communicative function of multi-modal interaction. These are schemes which annotate, for example, where non-verbal behaviours co-occur (or not) with speech, and the basic discursive function of the non-verbal behaviour and speech within and across such patterning. Examples of these types of coding schemes include Holler and Beattie's (2002) *binary coding scheme for iconic gestures*, and Allwood *et al.*'s (2007a) *MUMIN coding scheme*. To account for the simultaneous annotation of the form, pragmatic meaning and prosodic profile of gestures, the *M3D System* (Rohrer *et al.* 2020), instead, proposes a multidimensional approach to labelling gestures that goes beyond traditional systems, such as McNeill's (2000), which are solely descriptive. In addition, the *Database of Speech and Gesture* corpus (DoSaGE; Pak-Hin Kong *et al.* 2015), annotated via three independent tiers: a tier for linguistic information of the transcript, a tier for forms of gestures, and a tier with functions for each gesture used.

The IVO project, as described in Knight *et al.* (in press), used an annotation scheme that considered both form and function in an approach to analysing head nods in virtual meetings. This entailed creating two tiers for each speaker and annotating form on one tier and function on the other. The form categories were frequency, speed and range, while the functions were the categories of backchannels described by O'Keeffe and Adolphs (2008).

3.5. Establishing an analytical framework

Close qualitative viewing of the data collected led us to construct a suitable descriptive framework for online meeting stages and practices as outlined in Figure 1, below, and Table 2 (cf. Section 3.2.1). This is loosely based on Handford's (2010) business meeting stages and discursive practices. Handford's model consists of six meeting stages, including three pre- and post-meeting stages, representing access to participants before, during and after the data collection, and which accommodate the intertextual nature of the meeting (i.e., references made to emails, agendas and other communications outside of the meeting). The data collected for the IVO project contains only data recorded and collected during a scheduled meeting time, with no access to participants, before and after the meeting recording span. The simplified structural model we designed for our corpus (informed by Handford's model) is composed of four main stages (1–4) identified in our data, illustrated in Figure 1 and exemplified in Table 2 (Section 3.2.1), respectively. Figure 1 represents the broader context within which the meetings occur. It shows additional exchanges and activities (e.g., email communications), both virtual and face-to-face, which typically take place between participants whilst preparing for meetings (e.g., agenda creation, slides for item presentation) and responding to activities after the event (completing action points, writing up meeting notes) in spatio-temporal contexts, other than the meeting itself. These are labelled 'meeting preparation' and 'post meeting follow-up' and are illustrated in grey before and after the four main stages of the meeting. While we do not have corpus data from these two peripheral stages, we recognise their relevance to the recorded meeting data; for example, we have evidence, either through direct reference to an agenda or from the way meetings are progressed, that the preparation of an agenda before the meeting is central to the management of each of the meetings analysed. In summary, in our model, we include the pre-meeting and post-meeting phases as a component to understand inter-textuality, e.g., participants may reference emails sent or discussions between individuals before meetings.

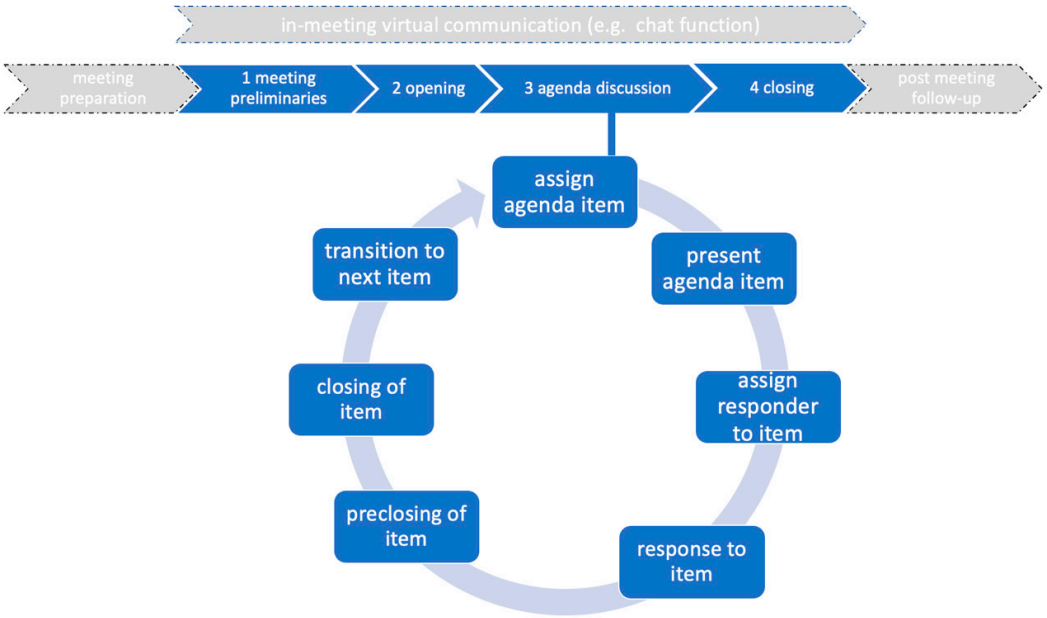

Figure 1: Linear and cyclical meeting stages

As Figure 1 illustrates, there are four core meeting stages which are fixed in order, and clearly identifiable in the data. Stage 1 represents a preamble to the meeting and may include work-related talk and small talk (Mirivel and Tracey 2005) as well as, crucial to this context, technology-related content specifically related to the setting up of the virtual meeting or the visibility of participants (particularly in meetings of large attendance, i.e., 20 participants or more). Stage 2 represents the start of the meeting proper. Stage 3 is typically a cyclical iterative stage, driven by multiple agenda items. Stage 4 can also contain more phatic communication once the 'business' of the meeting is over. In our corpus, stages 1 and 4 are regulated by the data contributors and the intactness of what takes place in these stages depends on when the recording is started and stopped.

As Handford (2010) notes, participants do more than simply progress through each of the structural stages of the meeting; they engage in a variety of goal-oriented dynamic discursive practices (e.g., setting the agenda, bringing the discussion back on track, checking shared understanding, moving from one agenda item to the next, bringing the meeting to a close). To scaffold and investigate these practices, Handford adopts McCarthy's (1998) four strands of linguistic behaviour found in spoken discourse (expectations, formulations, recollections and instantiations) and exemplifies them in terms of discursive practices, e.g., setting the agenda, bringing discussion 'on track' and bringing meeting to a close (Handford 2010: 77). Through our structural framework, we adopt a similar ap-

proach, isolating sub-sections within each of these stages, to examine specific discursive practices, at specific meeting stages, (e.g., managing turn-taking in the agenda discussion stage, closing of agenda items, timekeeping in the opening stage and agenda discussion stage). This gives us a means to systematically explore how these stages and practices are managed and executed both verbally and non-verbally, investigating whether certain verbal and non-verbal routines are realised at varying levels of granularity, at a stage, practice or linguistic feature level of analysis. It also allows for further cross-categorisation from different perspectives (e.g., identifying whether certain practices are characteristic of certain linguistic strands, examining the relational and transactional verbal and non-verbal behaviours and their co-occurrence at particular stages or with specific practices, or within the different discourse communities within our sampling frame). The stages and examples of practices and the verbal and non-verbal linguistic items used to enact them are set out in Table 3. We note that, while the meeting stages are relatively fixed, the practices listed and exemplified in Table 3 are not confined to these stages of the meeting. They provide us with targeted text external and internal to meetings (i.e., via structural/contextual factors or by linguistic components of the texts). Though not annotated as a component of the corpus or part of the corpus construction process, this framework was established from engagement with corpus after construction and provides us with a means of targeting language via meeting stages. These stages and discursive practices are essentially ways into our data for analytical purposes. In addition to the discursive practices enacted verbally and nonverbally, online meetings facilitate virtual means of enacting these via the use of the chat box, virtual reactions and emojis such as hand raises. Though our recordings do not include these, we have evidence that they are being used in verbal responses, e.g., “did you have your hand up?”

Stages and Discursive Practices	Exemplars
1. Premeeting (participants are present/joining)	
Setting up technology and hosting administration	*so yeah sorry we were just getting the live stream sorted there*
Introducing members	*we welcome X who is my director shadow*
Greetings	*hi everyone;* waves (physical/virtual); *hello, hi* comments in chat box
Engaging in small talk	*it's a beautiful sunny day here*
Transition move to opening	*okay, good; right, this meeting won't take too long*
2. Opening	
Reference to previous meeting	*as we said in the previous meeting*
Time keeping	*we're going to keep the presentations to 10 minutes*
Housekeeping	*please keep yourself on mute if you're not talking*
Previewing meeting	*this meeting is really going to be just about*
Acknowledging absentees and late arrivals	*X can't make it today*
3. Agenda discussion	
Assigning agenda item with nomination	*first up we have X over to you*
Contributing agenda item	*thanks everyone I'll just give you an update on ...*
Assigning responders to agenda item	X did you want to come in there; go ahead X
Request to contribute	hand up (physical/virtual); *can I just pop/jump in here*
Responding to agenda item (e.g. expressing gratitude, praise, encouragement; requesting more information/clarification; adding commentary; summarising; acknowledging contribution / endorsement of update or work done; displaying support)	*thank you and all your staff for the hard work you do; good; great; fantastic;* thumbs up; hand clap (physical/virtual); *we've covered a lot there*
Preclosing of agenda item	*so a massive thanks for the presentation um I think the questions have shown there's lots of interest in all the work you're doing...we'll move on to our next presentation; can I take that motion as adopted?*
Closing with upshot/gist	*we'll follow up on that again*
Transition move to next agenda item	*up next is item eight*
4. Closing	
Opening up closings	*any other business; does anyone have anything else to say?*
Concluding meeting	*alright. lovely to see you all and hopefully see many of you in person next week and we'll be in touch over over email over the coming days*
Goodbyes	*see you next time*; *bye*; waves, comments in chat box

Table 3: Meeting stages and discursive practices

3.6. (Re)presentation for analysis

While there is a dearth in the existence of freely available and widely used multi-modal corpora (Knight and Adolphs 2020), a range of digital tools and resources that exist to support the analysis of bespoke multi-modal datasets do exist. Examples of such software include *Transana*,[13] which enables users to integrate, transcribe, categorise and code their data then search and explore it in more detail (Halverson *et al.* 2012) and *ExMARaLDA* (Schmidt and Wörner 2014), which facilitates time-aligned transcription and annotation of multi-modal data. Perhaps the most popular, free and frequently used tool is Max Planck Institute's *ELAN* (Wittenburg *et al.* 2006). As seen in the screenshot in Figure 2, *ELAN* enables the annotation and analysis of data across multiple 'tiers' of information.

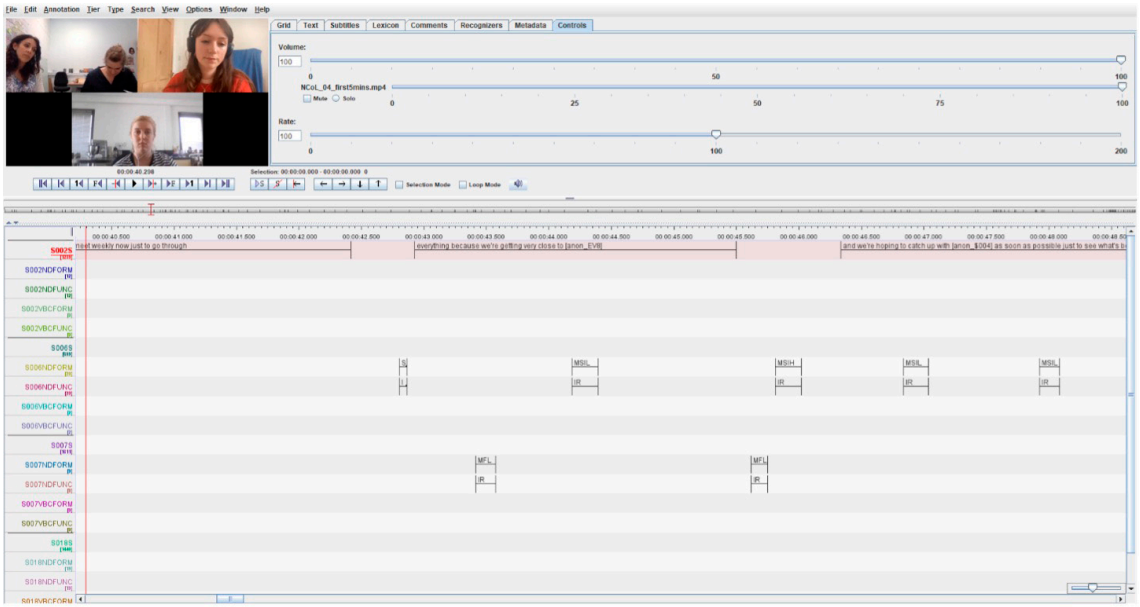

Figure 2: Screenshot of transcribed and coded data in *ELAN*

ELAN's tiers support frame-based analyses of multiple modes of time series data, from audio and video data to sensor outputs, which allows for synchronisation of tiers of data through the media timeline. Another feature of *ELAN* is its interoperability and flexibility in accommodating a range of file types, both for importing and exporting purposes. For example, different types of files can be imported into different tiers, e.g., .txt files or segmented .srt and Audacity files can be imported for transcription of speech, *Praat* files for analysis of sound, .mp4 files for video. Due to the ease of accessibility and use of *ELAN*, this software was the preferred option for the present study. Other studies which have used *ELAN* include the *Human-Computer Interaction Technologies* corpus (HuComTech; Pápay *et al.* 2011), DoSaGE (Pak-Hin Kong *et al.* 2015) and the *NEUROpsychological GESture*

[13] https://www.transana.com

corpus (NEUROGES; Lausberg 2019). Following the process of cleaning-up transcripts in otter, to enable the transcripts obtained from *Otter.ai* to be used in *ELAN*, two steps needed to be taken. These can be usefully repeated in other studies/projects of this nature:

1. Exporting transcription as SubRip (.srt) file. This file format is predominantly used for the creation of subtitles for integration into video files. To be used as subtitles, these files are timestamped to align with video to allow for future re-alignment in video annotation and analysis software. *Otter* has various options for line and character breaks, which result in different segmentation parameters in the subsequent transcription. We set the max number of lines to 1 and max characters per line to 2, which segments transcripts to approximate inter-pausal units when imported subsequently into tiers in *ELAN*.
2. Isolating individual speakers for use as single-speaker tiers in *ELAN*. To separate individual speakers to be treated as single-speaker tiers in *ELAN*, the .srt files are processed by a *python* code which isolates individual speakers. These files are then ready to be imported into *ELAN* as individual speaker tiers.

The above process results in an *ELAN* project that has the speech of individual speakers separated onto individual tiers. Additional tiers are then added for the annotation of non-verbal behaviour. Once these tiers have been defined, an *ELAN* template with tiers and a controlled vocabulary with set descriptions for annotations on tiers is created that can be used in other projects with the same analytical focus. For example, for an analysis of head nods as backchannels (see Knight *et al.* in press), we set a controlled vocabulary that was used on tiers for both form and function of backchannels. Having annotated data in *ELAN*, projects can be exported in a range of formats such as tab-delimited text, inter-linear text and subtitles text file. For our purposes, tab-delimited text files were used to observe incidence, frequency and co-occurrence of backchannels in excel spreadsheets.

4. Conclusion

Innovation is often based on how previous approaches are integrated in new ways. In this paper, we have outlined how, through the design, construction and analytical framing of the IVO corpus (see Knight *et al.* in press), we have engaged in practices that are innovative in how they integrate approaches in the following four areas:

1. Design frame: taking an approach to a design frame that is both user-based and focused on sectors.
2. Data collection: integrating data sourced from project partners that was recorded for this project with pre-existing recordings in the public domain.
3. Corpus construction: integrating speech to text transcription with a tiered system of multi-modal corpus analysis.
4. Analytical frame: integrating and adapting frameworks from previous workplace discourse to establish a framing that facilitates approaches to the data based on meeting stages, discursive practices and discourse features.

The challenges outlined in this paper regarding components of the corpus construction process are preceded by obstacles faced when attempting to acquire data. According to a survey carried by *KPMG International Limited* in 2022,[14] both businesses and consumers are growing evermore concerned about privacy and data security. In this environment, it is challenging for organisations to submit recordings of meetings for research purposes. To ensure the acquisition of data from sectors that fit into the IVO sampling frame, we have drawn upon a network of partners and investigators who have taken interest in and trusted the project from the outset. This trust-building has been essential to both the acquisition and sharing of data that constitutes the IVO corpus. Thus, having several project members with connections to various industries has been integral to the IVO corpus construction.

The construction of multi-modal corpora is still a relatively new endeavour. While recordings of virtual meetings promise a representation of an event that is close to what the participants in that event experienced, it remains challenging to ensure a process of corpus construction that is both efficient and reusable. The temptation to annotate everything in fine-grained detail is superseded by the understanding that this is extremely time-consuming, laborious and challenging in a context where you may be presented with thirty panels of speakers on a screen at one time. As with any project of this nature, we have found that clearly defined analytical goals (with a framework such as that outlined in this paper), and research questions aid the process of determining what to annotate and how. The decisions we have made in approaching data collection and analysis have ensured that we can both collect data that represents online meetings in a reasonably representative way and gain insights into this data in a manner that is achievable within the limited scope of a project such as this.

[14] https://advisory.kpmg.us/articles/2021/bridging-the-trust-chasm.html

References

Allwood, Jens. 2008. Multimodal corpora. In Anke Lüdeling and Merja Kytö eds. *Corpus Linguistics. An International Handbook*. Berlin: Mouton de Gruyter, 207–225.

Allwood, Jens, Loredana Cerrato, Kristiina Jokinen, Costanza Navarretta and Patrizia Paggio. 2007a. The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation* 41/3: 273–287.

Allwood, Jens, Stefan Kopp, Karl Grammer, Elisabeth Ahlsén, Elisabeth Oberzaucher and Markus Koppensteiner. 2007b. The analysis of embodied communicative feedback in multimodal corpora: A prerequisite for behavior simulation. *Language Resources and Evaluation* 4/3: 255–272.

BAAL = The British Association for Applied Linguistics. 2021. *Recommendations on Good Practice in Applied Linguistics* (fourth edition). Available at www.baal.org.uk

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Brunner, Marie-Louise and Stefan Diemer. 2021. Multimodal meaning making: The annotation of nonverbal elements in multimodal corpus transcription. *Research in Corpus Linguistics* 9/1: 63–88.

Carter, Ronald and Michael McCarthy. 2004. Talking, creating: Interactional language, creativity and context. *Applied Linguistics* 25/1: 62–88.

Cienki, Alan. 2016. Cognitive Linguistics, gesture studies, and multimodal communication. *Cognitive Linguistics* 27/4: 603–618.

Cohn, Neil. 2016. A multimodal parallel architecture: A cognitive framework for multimodal interactions. *Cognition* 146: 304–323.

Denham, Magdalena A. and Anthony John Onwuegbuzie. 2013. Beyond words: Using nonverbal communication data in research to enhance thick description and interpretation. *International Journal of Qualitative Methods* 12/1: 670–696.

Egbert, Jesse, Douglas Biber and Bethany Gray. 2022. *Designing and Evaluating Language Corpora: A Practical Framework for Corpus Representativeness*. Cambridge: Cambridge University Press.

Ekman, Paul and Wallace V. Friesen. 1968. Nonverbal behavior in psychotherapy research. In John M. Shlien ed. *Research in Psychotherapy Volume III*. Massachusetts: American Psychological Association, 179–206.

Esteve, Marc and Tamyko Ysa. 2011. Differences between the public and the private sectors? Reviewing the myth. *ESADEgov e-bulletin* https://esadepublic.esade.edu/posts/post/differences-between-the-public-and-the-private-sectors-reviewing-the-myth

Friedland, Gerald, Hayley Hung and Chuohao Yeo. 2009. Multi-modal speaker diarization of real-world meetings using compressed-domain video features. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. Tapei: IEE, 4069–4072.

Goodwin, Charles. 1994. Professional Vision. *American Anthropologist* 96/3: 606–633.

Halverson, Erica Rosenfeld, Michelle Bass and David Woods. 2012. The process of creation: A novel methodology for analysing multimodal data. *The Qualitative Report* 17/11: 1–27.

Handford, Michael. 2010. *The Language of Business Meetings.* Cambridge: Cambridge University Press.

Harrison, Claire. 2003. Visual social semiotics: Understanding how still images make meaning. *Technical Communication* 50/1: 46–60.

Holler, Judith and Geoffrey Beattie. 2002. A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk. *Semiotica* 142/1: 31–69.
Holler, Judith and Stephen C. Levinson. 2019. Multimodal language processing in human communication. *Trends in Cognitive Sciences* 23/8: 639–652.
Ibrahim, Blend and Ahmad Aljarah. 2023. The era of Instagram expansion: Matching social media marketing activities and brand loyalty through customer relationship quality. *Journal of Marketing Communications* 29/1: 1–25.
Knight, Dawn. 2011. The future of multimodal corpora. *Brazilian Journal of Applied Linguistics* 11/2: 391–416.
Knight, Dawn and Svenja Adolphs. 2008. Multi-modal corpus pragmatics: The case of active listenership. In Jesús Romero-Trillo ed. *Pragmatics and Corpus Linguistics: A Mutualistic Entente*. New York: Mouton De Gruyter,175–190.
Knight, Dawn and Svenja Adolphs. 2020. Multimodal corpora. In Stefan Gries and Magali Paquot eds. *A Practical Handbook of Corpus Linguistics*. Paris: Springer, 353–371.
Knight, Dawn and Svenja Adolphs. 2022. Building a spoken corpus. In Anne O'Keeffe and Michael McCarthy eds. *The Routledge Handbook of Corpus Linguistics.* London: Routledge, 21–34.
Knight, Dawn, Steve Morris, Laura Arman, Jennifer Needs and Mair Rees. 2021. *Building a National Corpus: A Welsh Language Case Study.* London: Palgrave.
Knight, Dawn, Anne O'Keeffe, Geraldine Mark, Chris Fitzgerald, Justin McNamara, Svenja Adolphs, Benjamin Cowan, Tania Fahey-Palma, Fiona Farr and Sandrine Peraldi. In press. *Interactional Variation Online* (IVO): Corpus approaches to analysing multi-modality in virtual meetings. *International Journal of Corpus Linguistics*.
Lausberg, Hedda. 2019. *The NEUROGES® Analysis System for Nonverbal Behavior and Gesture*. *The Complete Research Coding Manual including an Interactive Video Learning Tool and Coding Template*. Berlin: Peter Lang.
Levinson, Stephen C. and Judith Holler. 2014. The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2013.0302
Lin, Phoebe and Yaoyao Chen. 2020. Multimodality I: Speech, prosody and gestures. In Svenja Adolphs and Dawn Knight eds. *The Routledge Handbook of English Language and Digital Humanities*. London: Routledge, 66–84.
Love, Robbie. 2020. *Overcoming Challenges in Corpus Construction.* London: Routledge.
Lücking, Andy, Kirsten Bergmann, Florian Hahn, Stefan Kopp and Hannes Rieser. 2010. The *Bielefeld Speech and Gesture Alignment* corpus (SaGA). In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner and Daniel Tapias eds. *Proceedings of the 7th International Conference on Language Resources and Evaluation.* Valletta: English Language Resource Association, 92–98.
Mackenzie, Jai. 2020. Digital interaction. In Svenja Adolphs and Dawn Knight eds. *The Routledge Handbook of English Language and Digital Humanities*. London: Routledge, 49–65.
McCarthy, Michael J. 1998. *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
McNeill, David. 2000. Action and Thought. In David McNeill ed. *Language and Gesture*. Cambridge: Cambridge University Press, 139–140.

Milz, Dan, Atul Pokharel and Curt D. Gervich. 2023. Facilitating online participatory planning during the COVID-19 pandemic. *Journal of the American Planning Association*: 1–14.

Mirivel, Julien C. and Karen Tracy. 2005. Premeeting talk: An organizationally crucial form of talk. *Research on Language and Social Interaction* 38/1: 1–34.

Mondada, Lorenza. 2019. Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics* 145: 47–62.

O'Keeffe, Anne and Svenja Adolphs. 2008. Using a corpus to look at variational pragmatics: Response tokens in British and Irish discourse. In Anne Barron and Klaus P. Schneider eds. *Variational Pragmatics*. Amsterdam: John Benjamins, 69–98.

O'Keeffe, Anne, Michael J. McCarthy and Ron Carter. 2007. *From Corpus to Classroom – Language Use and Language Teaching*. Cambridge: Cambridge University Press.

Pak-Hin Kong, Anthony, Law Sam-Po, Connie Ching-Yin Kwan, Cristy Lai and Vivian Lam. 2015. A coding system with independent annotations of gesture forms and functions during verbal communication: Development of a *Database of Speech and GEsture* (DoSaGE). *Journal of Nonverbal Behavior* 39/1: 93–111.

Panteli, Niki and Patrick Dawson. 2001. Video conferencing meetings: Changing patterns of business communication. *New Technology, Work and Employment* 16/2: 88–99.

Pápay, Kinga, Szilvia Szeghalmy and István Szekrényes. 2011. HuComTech Multimodal Corpus Annotation. *Argumentum* 7: 330–347.

Ringeval, Fabien, Andreas Sonderegger, Juergen Sauer and Denis Lalanne. 2013. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In Rama Chellappa, Xilin Chen, Qiang Ji, Maja Pantic, Stan Sclaroff and Lijun Yin eds. *Proceedings of the 10th IEEE International Conference on Automatic Face and Gesture Recognition*. Shanghai: Curran Associates, 1–8.

Rohrer, Patrick Louis, Ingrid Vilà-Giménez, Júlia Florit-Pons, Núria Esteve-Gibert, Ada Ren, Stefanie Shattuck-Hufnagel and Pilar Prieto. 2020. *The Multimodal Multidimensional (M3D) Labelling Scheme for the Annotation of Audiovisual Corpora*. *Gesture and Speech in Interaction Conference.* Stockholm: University of Stockholm.

Rühlemann, Christoph and Alexander Ptak. 2023. Reaching beneath the tip of the iceberg: A guide to the Freiburg Multimodal Interaction Corpus. *Open Linguistics* 9/1: 20220245. https://doi.org/10.1515/opli-2022-0245

Schwartzman, Helen B. 1989. *The Meeting: Gatherings in Organizations and Communities*. New York: Plenum Press.

Schmidt, Thomas and Kai Wörner. 2014. EXMARaLDA. In Jacques Durand, Ulrike Gut and Gjert Kristoffersen eds. *Handbook on Corpus Phonology*. Oxford: Oxford University Press, 402–419.

Svennevig, Jan. 2012. The agenda as resource for topic introduction in workplace meetings. *Discourse Studies* 14/1: 53–66.

Trotta, Daniela, Alessio Palmero Aprosio, Sara Tonelli and Elia Annibale. 2020. Adding gesture, posture and facial displays to the polimodal corpus of political interviews. In Nicoletta Calzolari (Conference chair), Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk and Stelios Piperidiseds eds. *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille: European Language Resources Association, 4320–4326.

Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann and Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. In Nicoletta Calzolari, Khalid Choukri, Aldo Gangemi, Bente Maegaard, Josheph Mariani, Jan Odijk and Daniel Tapias eds. *Proceedings of the 5th International Conference on Language Resources and Evaluation*. Genoa: European Language Resources Association, 1556–1559.

Corresponding author
Anne O'Keeffe
Department of English Language and Literature
Mary Immaculate College
University of Limerick
South Circular Rd
V94 VN26
Limerick
Ireland
Email: anne.okeeffe@mic.ul.ie

received: June 2023
accepted: January 2024