Indicating engagement in online workplace meetings: The role of backchannelling head nods

Abstract

Amid COVID-19 and the so-called 'digital pivot', online virtual communication has been placed at the heart of our daily lives, both professionally and privately. As we move into a post-COVID context, the affordances of the digital turn have shown that we can operate professionally online but there is a need for a better understanding of communication in the online workplace. This paper first contributes to our understanding of the dynamics of indicators of engagement in multi-party communication online, as evidenced by a small-scale but in-depth corpus-based multi-modal study. Second, it showcases the importance of building naturally-occurring spoken corpora that go beyond written transcription of the spoken word and include annotation of non-verbal behaviour. The work pays particular attention to the incidence, frequency, position and function of spoken and head nod backchannels, exploring the coordination/co-occurrence of these features in online talk. Findings point to a changing profile of how engagement is displayed in online workplace meetings and this appears to be linked to the functionality of the platforms.

Keywords: Backchannels, response tokens, virtual communication, online workplace meetings, multi-modal corpus linguistics.

1. Introduction

There is a need, particularly in a post-COVID world that has witnessed a 'digital pivot', to derive new protocols for examining virtual workplace interaction, to better understand what comprises effective communication online, and to better understand virtual communication itself. Research being carried out on the Interactional Variation Online (IVO[1] – http://ivohub.org) project aims to address this need by undertaking corpus-based multi-modal analyses of virtual workplace communication to gain depth of insight into the potential barriers and carriers of effective dyadic and multi-party talk in this context. This paper showcases the types of multi-modal research that the IVO corpus can facilitate.

Multi-modal research is broadly concerned with the ways in which different "production modalities" (e.g. acoustic, speech, and different types of gestural behaviour – see Paggio & Navarretta, 2017: 464) interact to generate meaning in talk, with the

[1] See appendix for a full list of acronyms used in this paper.

acknowledgement that communication comprises a "specialised, evolutionary manifestation of a multimodal gestural complex" (Wilcox, 2004: 525), possessing "two sides, only one of which is speech; the other is imagery, actional and visuo-spatial" (McNeill, 2000: 139). A multi-modal corpus is defined as "video-recorded collection in which contributions in two or more of these modalities are annotated" (Paggio & Navarretta, 2017: 464) and is represented via different forms of media. Multi-modal corpus linguistics enables the empirical analysis of language at both a macro and micro level, identifying and characterising patterns of language and gesture-in-use. Multi-modal corpora also provide the tools to examine interactions within and between different modalities in the generation of meaning in talk.

In this paper, we focus on an area that has received much attention in the study of spoken interaction, but which has yet to receive comprehensive treatment in the context of online workplace meetings (Fernández Polo, 2021), namely the use of backchannels in the context of communicative feedback. Work in this area has long been linked to technology-mediated communication: Fries (1952), for example, in an early study on listener feedback, looked at its role in telephone calls. However, it is Yngve (1970: 568) who is credited with the introduction of the widely used term 'backchannel' to refer to non-floor-grabbing responses. Yngve focused on vocal and spoken items such as *uh-huh, yes, okay* and some short phrasal items. Nowadays, with the affordances of technology, a participant's full communicative feedback repertoire from the spoken to the non-verbal can be captured and analysed.

Backchannels are associated with good listenership and engagement (Fernández Polo, 2021) and "grease the wheels of the conversation but constitute no claim to take over the turn" (Tottie, 1991: 255). While the term 'listenership' is often associated with backchannels in the study of discourse (McCarthy, 2002), in the context of computer-mediated communication, the term 'communicative feedback' is more appropriate as it encapsulates multi-modal communication. Communicative feedback, according to Allwood et al (2007b: 256) refers to "unobtrusive (usually short) vocal or bodily expressions" that allow a recipient of information to inform a contributor of information about whether they are able and willing to (i) communicate (have contact), (ii) perceive the information (perception), and (iii) understand the information (understanding) and it allows the possibility of (iv) feedback information on emotions and attitudes triggered by the information. They project an understanding between speaker and listener that the turn has not been yielded, but as Duncan and Niederehe (1974) note, it is often difficult to identify the boundary between brief utterances and proper turns where the *listener* becomes the *speaker*. As O'Keeffe and Adolphs (2008) note, this challenge is more for the analyst than the actual conversational participants, who, in real-time conversation, draw on clues, such as prosodic features, facial expressions and gestures, to interpret whether an interlocutor is trying to take the floor or display listenership in a given context. We also illustrate in this paper that with the aid of multi-modal data and tools, we are afforded greater precision in relation to the cues of backchannels across all meeting participants, and this can lead to greater insights into the role of *collaborative* and *convergent* communicative feedback in multi-party talk.

Our focus is on how participants display communicative feedback, investigating whether non-verbal forms of feedback, especially head nods, play a new role in minimal response behaviour in online multi-party workplace environments.

2. Backchannels in communication

Backchannels are known by many terms including listener responses (Dittman & Llewellyn, 1968), receipt tokens (Heritage, 1984), and minimal responses (Fellegy, 1995). Backchannels represent the antithesis of interruptions (Mott & Petrie, 1995) insofar as they are listener responses which are not intended to take the floor as a turn would, but are instead intended to offer some form of relevant feedback to the speaker. Whilst backchannels are typically 'minimal' and often non-verbal in form, their relevance to effective communication should not be underestimated.

There are various lexical forms of spoken backchannelling behaviour, ranging from the most minimal such as *mm* and *yeah*, to multiple and repeated forms, and 'non-minimal' forms (Tottie, 1991: 263) such as *oh really* and *I know what you mean*, respectively. Spoken backchannels adopt different discursive functions depending on their lexical form, position in talk, and their coordination with non-verbal phenomena (see Maynard, 1997; Gardner, 1998; O'Keeffe & Adolphs, 2008 and section 3.3).

O'Keeffe and Adolphs (2008), in their study of backchannels in British and Irish English conversations using spoken corpora, noted that distinguishing between types of backchannels was not clear cut when using data transcribed from audio recordings as these fail to capture non-verbal response tokens such as head nods and shoulder shrugs. There has long been an awareness that a wide variety of non-verbal phenomena function as backchannels in discourse. For example, Brunner (1979) examined the use of smiles as backchannels, Jokinen (2011) examined gaze, whilst Knight (2011a), and Malisz et al. (2016) examined the use and functions of head nods and their lexical coordinates (for discussions of other forms of backchannelling behaviours, see Heylen et al. 2007: 149).

Several studies of backchannelling behaviour have been carried out using multi-modal corpora. Knight (2011a), for example, analysed the Nottingham Multi-Modal Corpus, a 250,000-word corpus of academic (student-supervisor) talk in English, whilst Paggio and Navarretta (2017) examined patterns of language and gesture-in-use in dyadic first meeting encounters. In general, previous corpus-based research in this space has typically utilised video-recorded dyadic or multi-party talk that has taken place in situ (i.e. in face-to-face contexts) and/or in lab-type conditions. Though some work has been done in online contexts (e.g. Brunner, 2021; Fernández Polo, 2021), there remain some gaps in this literature on the complexities of the relationships between forms of backchannelling behaviours in virtual communicative contexts specifically. The present study starts to bridge this gap by drawing on eight extracts of multi-party online workplace meetings, sampled from the starts and ends of four meetings from the IVO corpus.

3. Methodology

3.1 Data: a sub-corpus of starts and ends

The IVO corpus is an English-language post-COVID dataset (i.e. post-2021) that includes workplace meetings from three main types of organisations (Table 1). The meetings captured in the IVO corpus occur either fully online, or in 'hybrid' format where clusters of individuals are co-present while simultaneously linking with the main online meeting. For the purposes of this paper, a sub-corpus was created comprising eight short extracts of multi-party online meetings collected in the UK and Ireland. For comparability, these comprised four starts and ends of meetings (i.e. the first and last five minutes), as outlined in Table 2.

Table 1. Types of organisations in the IVO corpus

Organisation type	**Type of meeting**
Public sector	Routine city council meeting
Educational institution	University conference planning meeting
Non-governmental organisation	Arts NGO project team meeting

As Table 2 illustrates, word counts varied across the samples contained within the sub-corpus analysed.

Table 2. Breakdown of data in IVO sub-corpus

Reference	**Meeting type**	**Organisation type**	**Context**	**Word Count**	**No. of participants**	**No. of speakers**	**No. of turns**

DCC_2_start	Whole organisation	Public sector[2]	Hybrid	815	20	6	59
TaLC_1_start	Team	Educational institution	Online	1073	9	4	15
NCoL_1_start	Team	Non-governmental organisation	Online	1174	5	2	4
NCoL_4_start	Team	Non-governmental organisation	Hybrid	961	4	4	11
DCC_2_end	Whole organisation	Public sector	Hybrid	951	21	8	40
TaLC_1_end	Team	Educational institution	Online	1294	9	9	23
NCoL_1_end	Team	Non-governmental organisation	Online	1263	5	5	36
NCoL_4_end	Team	Non-governmental organisation	Hybrid	712	4	4	39
				8243	**77**	**42**	**286**

Note that in the case of DCC_2 (a city council meeting), whilst there are 20 and 21 visible participants on screen at the start and end of the meeting, respectively, it is likely that there is an additional number of participants engaging in the call; the total number of whom is unknown as cameras may have been turned off[3]. Note, also, that we have included two extracts from the same organisation to balance the number of online and hybrid meetings in the sample.

Of salience to this paper, we also note the disparity between the number of participants and the number of actual speakers and the number of turns per meeting. While not all participants spoke during the analysed meeting excerpts, they may have engaged in non-verbal feedback.

Whilst only short excerpts were included in the present study's sub-corpus, circa 40 minutes in total, these represent a sample of data types and contexts, arguably offering sufficient data for detailed analyses to be undertaken. We note the choice of the terms *starts* and *ends* as opposed to the norm associated with substantial work within the field of Conversation Analysis on *openings* and *closings.* Whereas within television, phone or radio data, there is a clear and canonical structure to openings and closings (see Clayman, 2002; O'Keeffe, 2006; Jucker & Landert, 2015), contemporary workplace meetings have staggered

[2] Public sector refers to government and government-controlled enterprises but for the purposes of the corpus, we have separated educational organisations.
[3] Note that none of the IVO team was present at the meetings and for this reason some of the details are unknown to us.

starts where participants join at different times, technical issues are dealt with, or phatic communication takes place and eventually someone changes *footing* (Goffman, 1981) to start the actual meeting[4]. At points associated with closings, again we find lack of alignment with the canonical order found in other mediated interactions on a television, phone or radio interview with endings often manifesting in collaborative waving. Hence, the term 'end' was seen as better representing this phase of discourse.

3.2 Transcription

To enable the analysis of multi-modal datasets, video recordings, as with spoken audio recordings, require orthographic transcription. While no agreed standard for transcription necessarily exists, shared practices are common across general spoken corpora (e.g. Cambridge and Nottingham Corpus of Discourse in English - CANCODE, Carter and McCarthy, 2004) and/or national corpora with spoken components (e.g. Spoken BNC2014, Love, 2020). 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, with an adapted version of the CANCODE conventions utilised (see https://ivohub.com/resources/).

Given the increasing number of speech-to-text tools, and the fact they are now integrated directly into the main videoconferencing software Teams (https://teams.microsoft.com/) and Zoom (https://zoom.us), it seemed appropriate to explore the potential for using a speech-to-text tool, Otter.ai, to generate "first-pass" transcripts here. Otter.ai (www.otter.ai) orients towards creating transcriptions that prioritise legibility and coherence rather than preserving all elements of the original speech. In the process, items such as backchannels, repetitions and hesitations (e.g. *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 required significant manual input/editing as a result. Despite these shortcomings, Otter.ai provided effective turn separation and time alignment, offered ease of editing, and generally increased the speed of transcription, so was deemed more of *benefit* than *cost* to use.

3.3 Coding

[4] It could be argued that this is not only due to the online context, but possibly due to the fact that there are more participants involved and it is not a premeditated beginning like on TV since bi-lateral video-mediated conversations do work similarly to telephone conversations (see Brunner & Diemer, 2018).

The next stage was to manually code the transcripts for the incidence, position and function of spoken and head nod backchannels with the aid of ELAN (https://archive.mpi.nl/tla/elan). ELAN enables the annotation and analysis of data across 'tiers' of information, supporting frame-based analyses of multiple modes of time series data, from audio and video data to sensor outputs.

The process for identifying and annotating backchannelling behaviours in the data is depicted in Figure 1. This involves playing at half-speed each recorded meeting and (i) identifying potential backchannels according to their lexical and head nod forms, (ii) checking if they function as backchannels, and (iii) annotating them accordingly. This is a particularly time-consuming process given that multiple participants were involved.

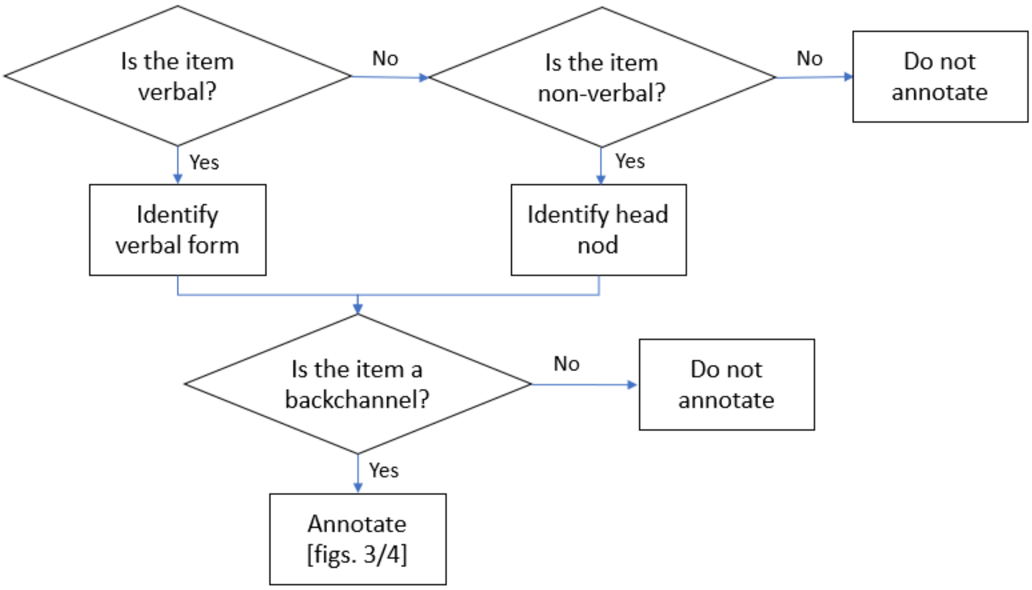

Figure 1. Approach to identifying and annotating spoken and head nod backchannels

Figure 2 illustrates the next stage of categorisation, which involves a two-stage process. Drawing on categorisation schemes in previous literature, spoken backchannels are first classified according to three common forms: 1) 'single' words/forms (see Oreström, 1983); 2) 'double' repeated words/forms (see Tottie, 1991: 263); or 3) 'multiple' (two or more) non-repeated words/forms (see Tottie, 1991: 263). Common examples of 1, 2 and 3 include *yeah*; *mm mm* and *that's right*, respectively. An additional category has been created for the purpose of this study: where a single word/form is seen to be used by a participant but, as their microphone was muted, they cannot be heard by other participants ('unheard' UH). This stage of form-based annotation is seen in yellow in Figure 2 (with codes in boldface).

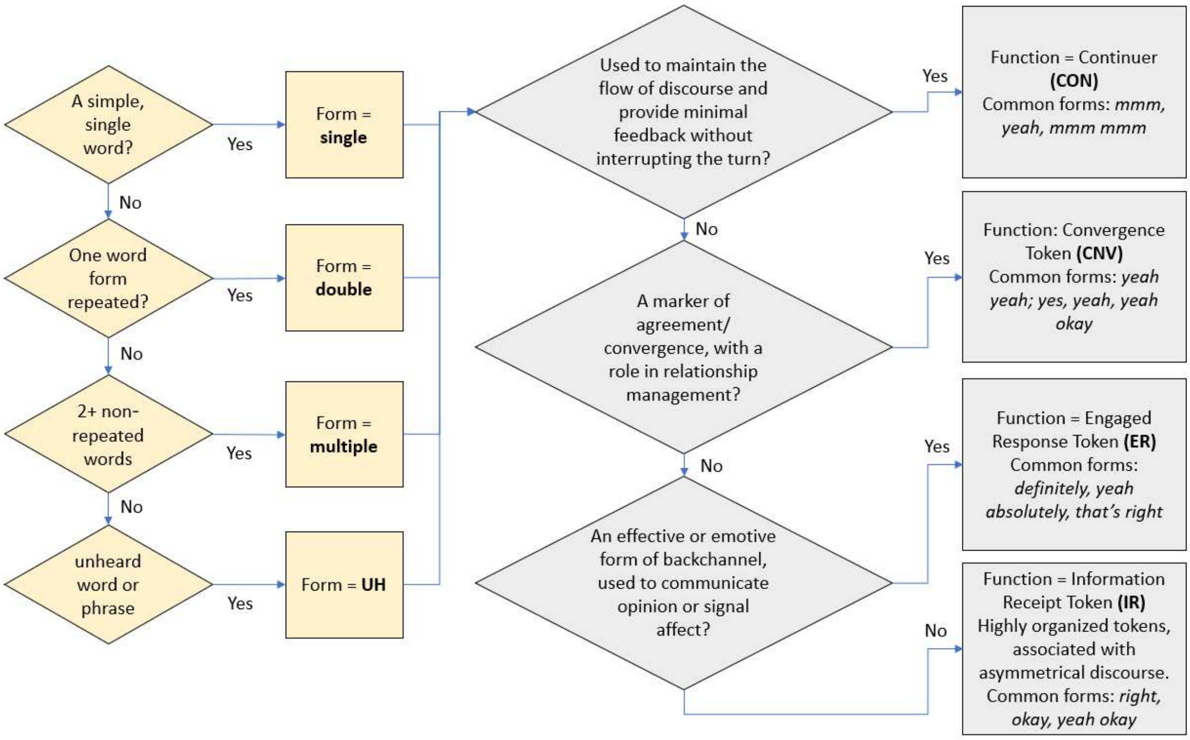

Figure 2. Approach to the annotation of spoken backchannel forms and functions

During stage three (indicated in grey), discursive functions are classified according to O'Keeffe and Adolphs' paradigm (2008 – adapted by Knight, 2011a), which was developed following the micro-analysis of the functions of backchannels in 60,000 words of data from the Limerick Corpus of Irish English and CANCODE. The principal functions of spoken backchannels in this paradigm are *continuers* (CON), *convergence tokens* (CNV), *engaged response tokens* (ER) and *information receipt tokens* (IR), with short codes indicated in boldface in Figure 2. These four functions exist on a cline. At one end are the most "facilitative" forms of spoken backchannels (continuers, CON, and convergence tokens, CNV). These typically have a relatively low/empty lexical content (e.g. *mm, yeah*) and are associated with discourse management and structuring between speaker and listener(s). At the other end of the cline are two distinct types of more engaged lexically content-heavy forms: engaged response tokens (ER) associated with relational/affective feedback and information receipt tokens (IR) which O'Keeffe and Adolphs found in asymmetrical (institutional) interactions where one speaker had organisational control of the flow of discourse.

Regarding backchannelling head nods, short codes are used to mark the incidence of all nods and to mark structural features of the movement (Figure 3). These codes were arrived at through initial observation of the data and the correlation of these observations with previous coding schemes used for categorising nod forms. Figure 3 illustrates the approach to annotating backchannelling head nods.

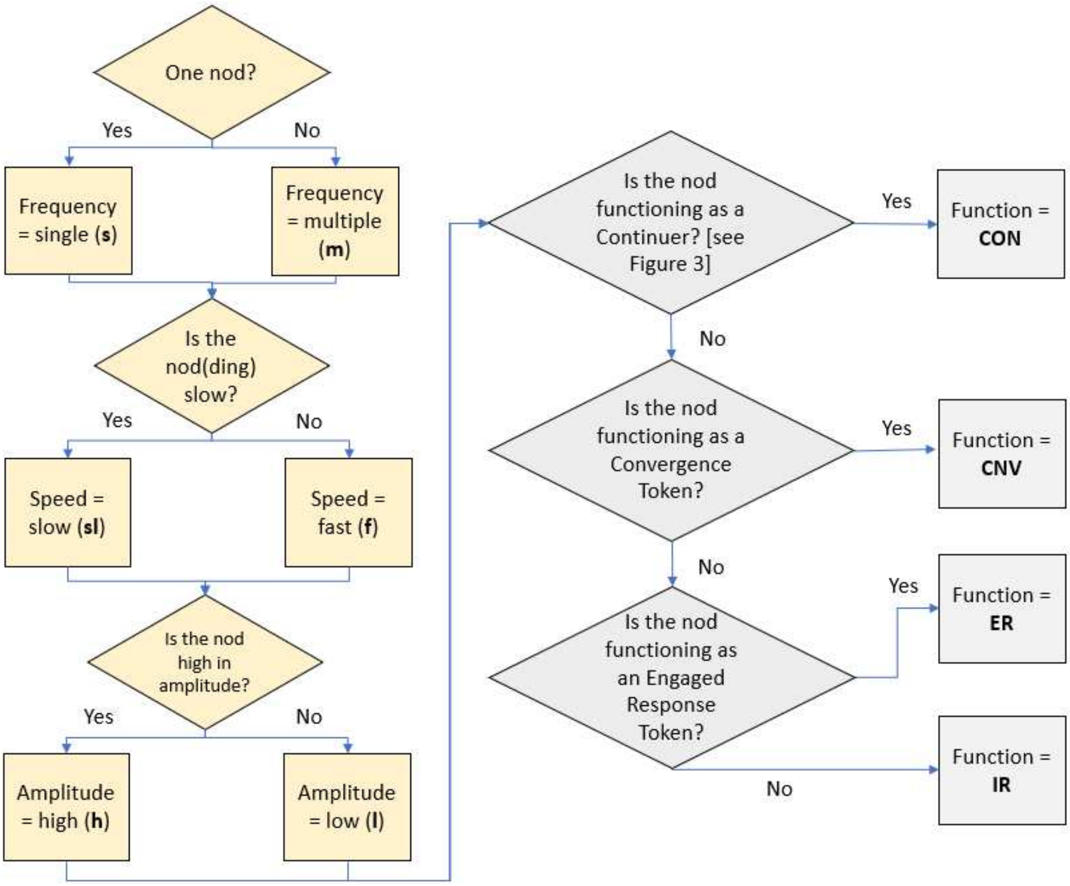

Figure 3. Approach to the annotation of backchannelling head nod forms and functions

The annotation scheme for *forms* of nods (in yellow,) was inspired in part by the MUMIN (Allwood et al., 2007a) and subsequent NOMCO schemes (based on MUMIN – see Paggio and Navarretta, 2017), as well as the work of Knight (2011a). The three categories are:

(i) **Frequency**: Ishi et al. (2010) and Knight (2011a) outline the functional difference between single and multiple backchannel head nods. Thus, we categorised frequency as s(single) and m(multiple).

(ii) **Speed:** the speed of head movements is found by Hadar et al. (1985) to contribute to their alignment with agreement and so was determined to be an important variable to categorise. The codes sl(slow) and f(fast) were used to code for speed of nods.

(iii) **Amplitude**: defined by Wagner et al. (2014) as the degree of displacement of a gesture. In relation to nods, this refers to a high or low vertical displacement associated with a forward and back motion. As such, we used the codes h(high) and l(low) for this category.

A controlled vocabulary was created in ELAN with the codes and integrated into tiers for the annotation of backchanneling head nod forms. This resulted in eight possible combinations as annotations on head nod form tiers, as seen in boldface (in the process/square boxes in Figure 3). Each nod was then categorised (and checked, by two researchers from the team) according to its discursive function, again using the categorisation scheme proposed by O'Keeffe and Adolphs (2008) and adapted by Knight (2011a) for the categorisation of backchannels in a multi-modal context.

3.4 Analytical approach

To determine frequencies of backchannels, the annotated data were exported from ELAN as 'tab-delimited text' in .csv files for use in Microsoft Excel. Results were cross-checked with manual counts using ELAN's annotation spreadsheet function, which provides sortable lists of annotations corresponding to individual annotation tiers. Overviews of co-occurrences and patterns of annotations were viewed using ELAN's 'annotation density plot'. This depicts the relative position of where individual annotations occur. It can, for example, identify clusters of backchannels with certain forms and/or functions. To obtain a more fine-grained view of co-occurrence, clusters of annotations on multiple tiers were viewed to identify and classify patterns of co-occurrence such as the exact alignment of nod with spoken backchannels.

This analysis process constitutes a function-to-form approach (O'Keeffe, 2018) often used in corpus pragmatics whereby corpus data is coded manually for a function and then qualitatively analysed. This approach contrasts with the more prevalent form-to-function approach (Aijmer, 2018) whose starting point is a frequency or key word list generated using corpus tools from which a qualitative functional analysis is then conducted. The function-to-form approach involves manual micro analysis, but it leads to a fuller functional description of forms. Essentially, a function-to-form approach allows for depth of analysis while form-to-function approach allows for breadth.

4. Analysis

4.1 Form and frequencies of backchannelling

4.1.1 Overall occurrences across the sub-corpus

Table 3 shows a breakdown of the frequency of occurrences of backchannelling head nods (NBCs) and spoken backchannels (SBCs) across the starts and ends of the four meetings in our sub-corpus. Blank entries here, and for subsequent tables, indicate a frequency of zero.

Table 3. Frequency of backchannelling head nods (NBCs) and spoken backchannels (SBCs)

	NBC		SBC	
	starts	**ends**	**starts**	**ends**
TaLC_1	84	110	5	11
DCC_2	9	3	7	
NCoL_1	56	149	4	11
NCoL_4	53	52	3	13
TOTAL	**202**	**314**	**19**	**35**

Initial observations indicate that:

(i) across the four meetings, nods are more frequent than spoken backchannels, at both the starts and ends of meetings.

(ii) with the exception of the DCC_2 data, backchannelling behaviour (spoken and nodding) tends to occur more frequently at the ends of meetings than at meeting starts.

(iii) where spoken (including those 'unheard', where participants were muted) backchannels occur, they are more frequent at the ends of meetings than at the starts of the meetings (DCC_2 excepted – as there are no spoken backchannels at the end of this meeting).

(iv) there is a noticeable increase in backchannelling head nods (NBCs) in the TaLC_1 and the NCoL_1 data at the ends of the meetings. This may be because ends tend to include discussion phases, while the starts tend to be dominated and managed by the chair in phases of information sharing.

(v) head nods are not equally distributed across the meeting samples.

Strikingly, we observe that backchannelling head nods (NBCs) are ten times more frequent than spoken backchannels (SBCs) in this sub-corpus. This observation contrast with previous findings that found that there were no marked differences in the rate of backchanneling head nod (and spoken backchannel) use at the start, middle and end of each face-to-face conversation (Knight, 2011a): "spoken and backchannelling head nod behaviours are used at a near constant rate" (Knight, 2011a: 118) (see also Oreström, 1983; Gardner, 1998). This may be an indication that the online medium prompts more backchannelling head nods than on site interaction as a compensatory strategy.

4.1.2 Forms

We now turn to the specific *types* of backchannels used in the sub-corpus. Tables 4 and 5 detail the frequencies of specific forms of backchannelling head nods and spoken backchannels, respectively.

Table 4. Backchannelling head nod (NBC) forms in the sub-corpus

	NCoL 1 start	NCoL 1 end	NCoL 4 start	NCoL 4 end	DCC 2 start	DCC 2 end	TaLC 1 start	TaLC 1 end	Total	%
SFL		7	4		1		2	7	21	4.1
SFH	14	4	2	1	1		1		23	4.5
SSlL		1	1				3	1	6	1.2
SSlH		1					1		2	0.4
MFL	27	122	29	47	4	1	69	84	383	74.2
MFH	2	13	3	1	1			2	22	4.3
MSlL	11	1	6	2	2	2	7	12	43	8.3
MSlH	2		8	1			1	4	16	3.1
Total	**56**	**149**	**53**	**52**	**9**	**3**	**84**	**110**	**516**	
%	**10.9**	**28.9**	**10.3**	**10.1**	**1.7**	**0.6**	**16.3**	**21.3**		

The most frequent form of nod backchannel is multiple, fast, low (MFL), a series of quick nods with low amplitude. These constitute 74.2% of all nod backchannels and are found across all other data sources, apart from the public sector DCC_2 data (in which nod backchannels are relatively scarce). The MFL nod occurs most frequently in the meeting ends, with one third found in one data sample alone (NCoL_1_end). The second most frequently occurring nod type, multiple, slow, low (MSlL), constitutes only 8.3% of nod backchannels, with 25% of these occurring in NCoL_1_start, 27% in TaLC_1_end and 16% in TaLC_1_start. The remaining nod types fluctuate in their distribution across the data samples. For example, 60% of single, fast, high (SFH) nods occur at the start of NCoL_1 and the remainder are thinly distributed across five of the other seven samples. Whilst 59% of multiple, fast, high (MFH) backchanneling nods occur in the end of NCoL_1, they are used by only two of the five meeting participants. These results diverge with those seen by Knight (2011a), where the most frequent forms of backchanneling nods found in academic dyadic supervision data were small nods of a short duration (corresponding to single, slow, low 'SSlL' here), comprising 47% of all instances. Nods corresponding to the multiple, fast, low (MFL) forms occurred in 28% of instances in Knight (2011a) (although this form was found to be the second most frequently used).

Results indicate that there is an uneven distribution of backchanneling nods across the sub-corpus. Almost 40% occur in the non-governmental organisation NCoL_1 data (10.9% in the start, 28.9% in the end), 37.6% in the educational institution TaLC_1 data (16.3% in the start, 21.3% in the end), and 20.3% in NCoL_4 (10.3% in the start, 10.1% in the end). Only 2.3% occur in the public sector DCC_2 data (1.7% in DCC_2_start, 0.6% in DCC_2_end).

Table 5. Spoken backchannel (SBC) forms in the sub-corpus

Form	**NCoL1 start**	**NCoL1 end**	**NCoL4 start**	**NCoL4 end**	**DCC start**	**TaLC1 start**	**TaLC1 end**	**Total**	**%**
mm	1	1						2	**4%**
okay	1	1		2			1	5	**9%**
yeah		4	2	7	1	1	4	19	**35%**
yes		1				1		2	**4%**
mmhmm			1					1	**2%**
no				1				1	**2%**
nice				1				1	**2%**
cool		1						1	**2%**
yes yeah					1			1	**2%**
very good					1			1	**2%**
that's cool				1				1	**2%**
yes fabulous						1		1	**2%**
oh okay							1	1	**2%**
yeah absolutely							1	1	**2%**
yeah we've got that		1						1	**2%**
yes yeah					2			2	**4%**
yes yeah fire away					1			1	**2%**
yes please thank you					1			1	**2%**
yes yeah fabulous						1	1	2	**4%**
yeah I think she is						1		1	**2%**
unheard (UH)	2	2		1			3	8	**15%**
Total	**4**	**11**	**3**	**13**	**7**	**5**	**11**	**54**	

As Table 5 indicates, *yeah* is the most frequently spoken backchannel, accounting for 35% of forms. This tallies with O'Keeffe and Adolphs' (2008) finding the *yeah* was the most frequently occurring single-word response token in their face-to-face conversation data. The occurrences of the unheard (UH) category of items also appear to be vocalisations of *yeah*. Of note here is the frequency of *yeah* in ends of meetings compared to starts (18 vs. 4 occurrences). As mentioned in relation to Table 3, the ends tend to involve more open discussions, while the starts tend to be dominated and managed by the chair in phases of information sharing. All 22 occurrences of *yeah* in the data are convergence tokens. While Knight and Adolphs (2008) found strong convergence usage of *yeah*, they also found frequent usage as a continuer which is absent here (discussed below).

Table 5 shows that *yes* is a component of all multi-word backchannels (while occurring twice as a single-word BC). This corresponds with what McCarthy (2003: 35) refers to as *yes-plus* response tokens. While *yes* in isolation performs a backchannelling function, the addition of words (or repetitions of *yes*) provide emphasis or clarity (e.g. *yes yeah*). However, this contrasts with Farr's (2003: 78) analysis of tutor-student meetings in which repetitions of *yes* are noted as rare possibly because "its strength as a single token suffices". With this pragmatic force of *yes* in mind, we posit that *yes* (and multi-word units with *yes*) appears to be playing some kind of (online) context-specific role in backchannelling that may be highly procedural in these institutional contexts and may play an expediency role in moving proceedings on quickly. This is borne out by the fact that five of the seven occurrences of *yes-plus* tokens occur at the start of the most formal public sector dataset, the DCC city council meeting. Four of the five DCC tokens are by a single participant who is not the chair. Clearly, they are keen to progress the meeting and draw on the pragmatic force of a *yes-plus* item (See S054S *yes yes yeah fire away* in Figure 4).

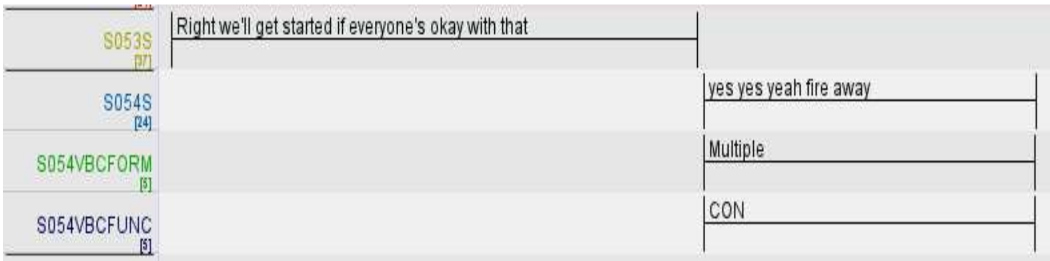

Figure 4. Example of *yes-plus* spoken backchannels (SBC) in public sector DCC_start

4.1.3 Form co-occurrences across sub-corpus

There were five common patterns of backchannelling head nods (NBCs) and spoken backchannels (SBC) the sub-corpus:

(i) backchannelling head nod (NBC) only, typically in the form of multiple times, fast, low range of movement (i.e. multiple, fast, low - MFL),
(ii) spoken backchannel (SBC) only, typically single *mm, yeah*,
(iii) backchannelling head nod (NBC) and spoken backchannel (SBC) co-occurring, typically an MFL nod co-occurring with *yeah,*
(iv) NBC preceding SBC (typically MFL nods), and

(v) SBC preceding NBC for example *yeah* occurring before single, fast, high (SFH) nods.

Table 6 represents the usage patterns of backchannelling head nods (NBCs) and spoken backchannels (SBCs). It illustrates that pattern (i) NBCs accounts for 90%.

Table 6. Patterns of SBC and NBC use in the sub-corpus

	Pattern	**Frequency**	**Percentage**
i	NBC only	491	90%
ii	SBC only	28	5%
iii	NBC and SBC co-occurring	15	3%
iv	NBC preceding SBC	6	1%
v	SBC preceding NBC	4	>1%

These results differ from Knight's (2011a) findings: for more than 70% of the time when spoken or nod backchannels were used, it was not in isolation. Here, we see that the inverse pertains: 95% of spoken or nod backchannel forms do not co-occur. This may relate to the spotlighting and mute functions of meeting platforms which we discuss in Section 4.3 (i.e. whereby even minimal SBCs can impact on turn allocation).

In the next section we take a detailed look at the frequency and distribution of specific backchannelling functions in the sub-corpus.

4.2 Backchannel functions

4.2.1 Frequency and distribution across the sub-corpus

Figures 5 and 6 show the frequency and distribution of NBCs and SBCs across O'Keeffe and Adolphs' (2008) four functional categories: *continuers* (CON), *convergence tokens* (CNV), *engaged response tokens* (ER) and *information receipt tokens* (IR) (see Section 3).

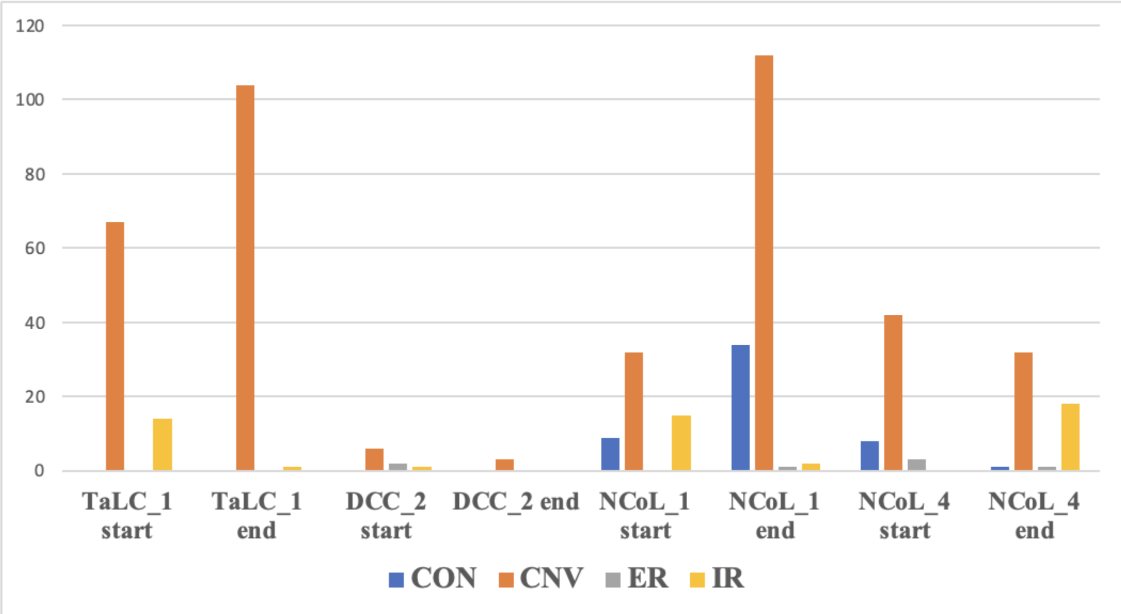

Figure 5. Nod backchannel functions across the sub-corpus

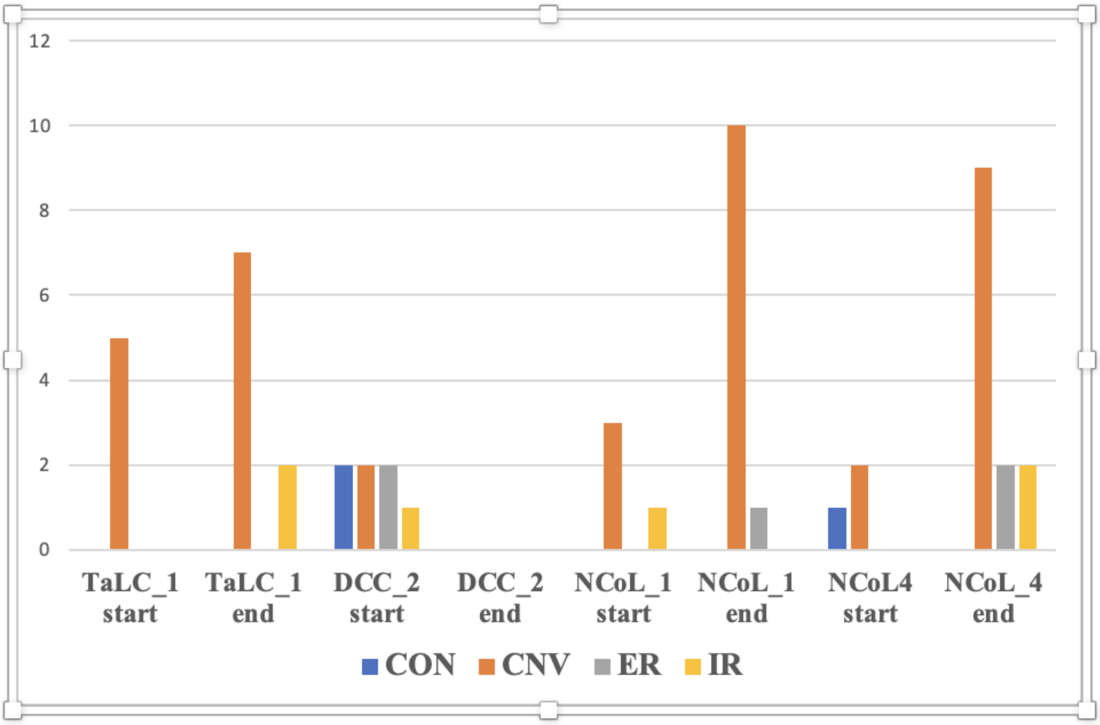

Figure 6. Spoken backchannel functions across the sub-corpus

Figures 5 and 6 indicate that the CNV function is the most dominant for both NBCs and SBCs, in both starts and ends, while CONs are relatively infrequent.

4.2.2 Form by function

Tables 7 and 8 illustrate the distribution of form types across the four discursive functions.

Table 7. NBCs by type and their corresponding functions in the sub-corpus

NBC form	Function			
	CNV	CON	IR	ER
MFL	304	44	32	3
MFH	17	2	1	2
MSlL	32		9	2
MSlH	15		1	
SFL	17	2	2	
SFH	14	4	5	
SSlL	6			
SSlH	1		1	
Total	**405 (77%)**	**52 (10%)**	**51 (10%)**	**7 (1%)**

Table 8. SBCs by form and their corresponding functions in the sub-corpus

SBC Form	Function			
	CNV	CON	IR	ER
mm	1		1	
okay	2		2	
yeah	19			
yes	2			
mmhmm		1		
no	1			
nice			1	
cool	1			
that's cool	1			1
yes yeah	1			
very good				1
yes fabulous	1			
yeah absolutely	1			
oh okay			1	
yeah we've got that	1			
yes yeah	1			1
yes yeah fire away		1		
yes please thank you		1		
yes yeah fabulous	2			
yeah I think she is	1			
UH	7			1

Total	**41 (77%)**	**3 (7%)**	**5 (9%)**	**4 (8%)**

These tables indicate that 75% of backchanneling nods with a CNV function are performed with a multiple, fast form with low amplitude (MFL). This form dominates all the functions (85% of nods which function as a CON are also multiple fast and low in form, as are 63% of IR nods). 46% of CNV SBCs are single form *yeah*. The unheard form (in each case that these occurred in our data, they were single unheard *yeah*) constitutes an additional 7% of the CNV function.

4.3 Continuers (CON)

The CON function is noticeable by its absence or relatively low occurrence, constituting just three of the 53 SBCs (5.66%). Spoken backchannels with a CON function occur in two of the eight samples, DCC_2_start (twice) and NCoL_4_start (once). Backchanneling nods with a CON function only occur in the non-governmental organisation data: NCoL_1_start (nine times), NCoL_4_start (eight times) and NCoL_4_end (once), and NCoL_1_end (34 times).

This low occurrence contrasts with previous findings where both CON and CNV backchannels are most frequent, with the CON function often occurring most frequently in institutional contexts (Knight, 2011a). We note here that the spoken CON backchannel, characteristically minimal forms which are unobtrusive in a face-to-face interaction, appear to have changed in terms of valency in online communication. Technology plays a role in this: firstly, when a microphone picks up noise from a participant, they become the focus of the interaction, either through their screen image lighting up (e.g. Zoom) or through a movement of their screen image to the top left-hand corner (e.g. Teams). With this type of 'spotlighting' comes an expectation that the participant has something to contribute even if they are merely backchannelling. Secondly, and added to this, is an element of displacement in the precision of the positioning of the backchannel because utterances can be perceived milliseconds after they are uttered. This creates an exchange which is minimally out of sync. When microphones are switched on, even a minimal vocalisation can appear to be a potential turn grab. While a spoken backchannel in a face-to-face setting with a CON (or CNV) function does not disturb the flow of the discourse, in online communication, the valency (or pragmatic force) of the 'minimal response' appears to have changed. We acknowledge that nod backchannels do not prompt the same highlighting effect, but the mere act of nodding could also draw unwanted attention. We propose that participants are aware of this change in valency and, as a result, there is a shift of behaviour away from the CON function in both spoken and nod backchannels as reflected in our results.

The few occasions where nod backchannelling is seen with a CON function, it appears to be a demonstration of "on-recordness" whilst engaged in another task related or not to the meeting discourse. We observe regular nodding from participants who are multi-tasking, e.g. reading something onscreen or looking down and reading other materials, or attending to messages on a mobile phone. The multi-tasking facilitated by the online environment appears to generate compensatory head nodding.

4.4 Convergence tokens (CNV)

CNV is the most frequently occurring backchannelling function in this sub-corpus (78.66% of NBCs, 77.35% of SBCs), performing a relational function and marking agreement or convergence in the discourse, thus playing an active role in reinforcing commonality. This finding is in contrast to O'Keeffe (2006); Knight (2011a); Knight and Adolphs (2008), where continuer (CON) rather than CNV is the most frequent function in institutional settings, and where CNV backchannelling is associated with casual conversation between friends (O'Keeffe and Adolphs, 2008). This is possibly because convergence, as a relational marker, is important for the smooth running of meetings especially given the reduction in use of continuers as discussed in Section 4.3.

This convergence coincides with participants accommodating towards each other's responses. Figure 7 illustrates this pattern in a screenshot of the TaLC_1_end data with nod form and function tiers isolated. Of relevance here is the pattern of convergence in both function (CNV) and form (multiple, fast, low). This is the most frequently occurring pattern of co-occurrence of form and function in the sub-corpus.

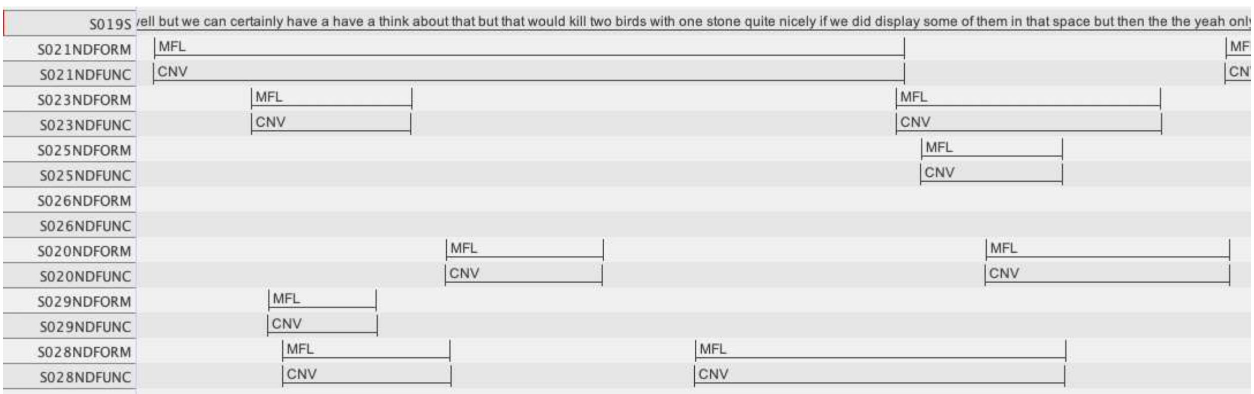

Figure 7. Screenshot of ELAN showing clustering of patterns in TaLC_1_end

At this point in the meeting, the chair is responding to a suggestion offered by another participant. We note here the alignment of convergence between a willingness to consider the suggestion (through *we can, certainly, quite nicely*) and the pattern of agreement among the other participants through convergent head nodding. This extract is representative of the kind of convergent behaviour that is likely to occur in settings where participants may desire the approval of others. This could be viewed as a type of accommodation where participants are converging in their spoken and non-verbal behaviours.

4.5 Engagement response tokens (ER)

ER tokens are infrequent in the data, constituting 7.54% of SBCs and 1.35% of nods, making ER the least frequent nod type and third most frequent SBC type. This is in contrast to O'Keeffe and Adolphs (2008) who found ER tokens to be the second most frequent response token type in their casual conversation data. This is not unexpected given the institutional nature of the data. We notice, however, that though ER backchannels are infrequent, there are phases of meetings especially where participants engage in a personal narrative that they do occur especially: for example, in DCC_start.

We posit that the low frequency of ER tokens is due to the institutional nature of the data, which is focused largely on agenda items and so does not involve personal discourse that could involve emotional responses in the form of SBCs or nods.

4.6 Information receipt tokens (IR)

Similar to both CON and ER functions, IR responses are notably infrequent. Just 9.4% of SBCs are IR, while 9.9% of nods function as IR. This contrasts with Knight and Adolphs' (2008) who found that 23.5% (16 of 68) of spoken BCs in their dyadic institutional data were IR. O'Keeffe and Adolphs (2008) note that IR tokens are associated with institutional discourse (e.g. radio phone in) and especially within power asymmetries where they play a role in managing discourse. We propose that in online workplace discourse IR tokens, though less frequent, may play an important role not just for the power-role holder (the chair) because in the online medium there is an onus on the listeners/participants to clearly mark the receipt of information or instruction as a compensation for lack of physical co-presence. However, more online meeting data across more workplace contexts is needed to fully explore this. Gardner (2007) found the frequency of occurrences of IR tokens to be heavily context-dependent (e.g. *right* in dietetic consultations). Such a necessity is not present in the current data. Also, as has been noted, these tokens may be avoided so as not to cause interruption due to spotlighting.

We also observe that four of the five spoken IRs co-occur with IR nods. Figure 8 shows this type of co-occurrence in the TaLC_1_end sample with a speaker responding to the meeting chair by saying *oh okay* while engaging in a multiple, slow, high (MSlH) type IR nod.

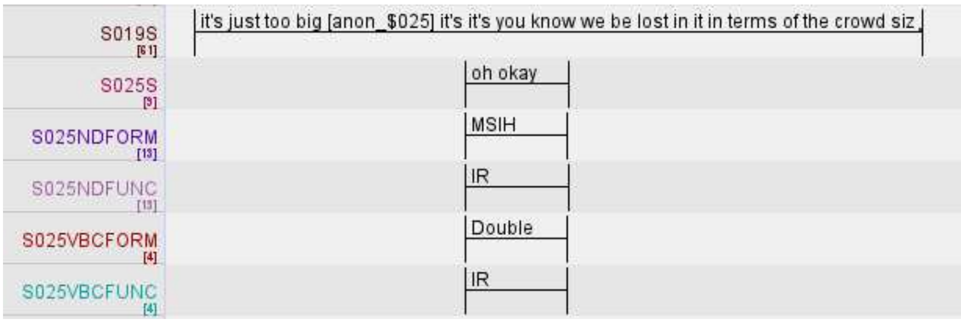

Figure 8. Screenshot of ELAN showing co-occurrence of IR SBC with IR nod in TaLC_1_end

5. Conclusion

In this multi-modal corpus study of backchannels in starts and ends across four online workplace meetings, we note various findings that merit further investigation:

(i) head nods backchannels are found to be more frequent than spoken backchannels. However, in contrast to previous work on face-to-face institutional discourse settings, the type of nod that is most frequent and its discursive function has shifted. In terms of form, we note a move from SSlL (single, slow, low) to MFL (multiple, fast, low). In terms of the function of these nods, we note a move from continuing to non-floor grabbing convergence.

(ii) participants appear to be adapting to the challenges and affordances of the digital platform, which presents participants in a two-dimensional plane. This creates a *gallery effect* whereby participants see both themselves and their interlocutors, as well as their spoken and non-verbal behaviour. The spotlighting function in particular appears to have an impact on turn-taking and listenership because even when the participant is making a minimal response it can appear as a turn grab. Returning to Mott and Petrie (1995), backchannels are the antithesis of interruptions because they are not intended to take the floor as a turn. We note that participants appear to be mitigating against backchannels being confused with turn grabs by:

a. using head nods over spoken backchannels to indicate listenership,
b. minimising the cooccurrence of head nods and spoken backchannels in general,
c. minimising the use of continuer head nods with spoken backchannels,

d. favouring displays of convergence (over continuers) using head nods and spoken backchannels, and
e. accommodating towards each other's convergent response behaviour in clusters, especially at ends of meetings.

Put simply, in an online setting, participants appear to be avoiding displays of continuer backchannelling because of the spotlighting function and because it is not necessary to assure the floor-holder that they can continue in this medium. Additionally, we note that because of the spotlighting function the pragmatic force of a continuer backchannel appears to be stronger in this medium and this also contributes to avoidance. For example, the turns of the chairing participants are long. Participants appear to have developed an understanding that, with an effective chair, their turn will come, and that it is through convergence that they can offer listenership.

The city council meeting (DCC_2) appears to be an outlier. Despite having the most participants, it contains fewest nod and spoken backchannels (Table 3). Less than a third of participants speak during the start of the meeting and just over a third of participants at the end. When backchannels occur in these data, they tend to be in dyadic or small group discussions. It is possible that the "on-recordness" of these political data (which are publicly available) contributes to the lack of backchannels as well as a highly structured and managed agenda with items given set timeframes. Also, the number of participants in the DCC_2 meeting is likely to effect opportunities for, degrees of and types of participation and engagement. In line with Knight (2011a) and O'Keeffe (2006) backchannels can link to roles and so occurrences across participants will not be consistent. For example, the meeting chair is more likely to be holding/managing the floor and less likely to have a high occurrence of backchannelling (e.g. Figure 7 extract).

Another major contextual factor is the effect of the global pandemic on the nature of the interaction and the relationships between the participants involved in the meetings. In addition, layers of workplace hierarchical structures are strong determiners of communicative behaviour in both face-to-face and online workplace settings (Marlow & Wilson, 2012). At the time of writing, online meetings still appear to be widely used but their nature appears to include more hybrid formats (as evidenced in the DCC_2 city council data). Future work will be needed on the impact of hybrid workplace meetings.

We are conscious that this paper examines (in detail) a small dataset and that the quantitative analysis and interpretation is sometimes based on very low frequencies and we acknowledge that much more data needs to be analysed if generalisations are to be made. Indeed, some patterns may be idiolectal. However, we feel that our paper makes an important contribution to showcasing the affordances of taking a multi-modal approach in the use of corpus linguistics to study spoken discourse.

As a result of COVID-19, the use of online virtual communication became a norm where it had been an exception. Post-COVID, it is a residual and normal practice, with the addition of hybridity. We argue that this major digital turn in workplace meeting practices requires in-depth research and here we hope to have contributed to an enhanced understanding of the dynamics of indicators of engagement in multi-party communication, through a small-scale but in-depth corpus-based multi-modal study. We hope that in the process, we have showcased both the potential and the importance of building naturally-occurring spoken corpora that go beyond written transcription of the spoken word and include annotation of non-verbal behaviour. Our methodology pays meticulous attention to the incidence, frequency, position and function of spoken and head nod backchannels and this has facilitated an exploration of the coordination/co-occurrence of these features in online meetings. Our findings point to a changing profile of how engagement is displayed in online workplace meetings, and this appears to be linked to the functionality of the platforms. Without a multi-modal corpus methodology, our results would have serious misrepresented how engagement through backchannelling and would not have identified the evolving role of head nods in this.

Acknowledgement

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

References

Aijmer, K. (2018). Corpus Pragmatics: From form to function. In A. H. Jucker, K. P. Schneider, & W. Bublitz (Eds.), *Methods in Pragmatics* (pp. 555-585). Berlin: Mouton de Gruyter.

Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C., & Paggio, P. (2007a). The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation, 41*(3), 273-287.

Allwood, J., Kopp, S., Grammer, K., Ahlsén, E., Oberzaucher, E., & Koppensteiner, M. (2007b). The Analysis of Embodied Communicative Feedback in Multimodal Corpora: A Prerequisite for Behavior Simulation. *Language Resources and Evaluation, 41*(3/4), 255–272.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman, London.

Brunner, L. J. (1979). Smiles can be back channels. *Journal of Personality and Social Psychology, 37*(5), 728–734

Brunner, M-L. (2021). *Understanding Intercultural Communication - Negotiating Meaning and Identities in English as a Lingua Franca Skype Conversations*. Berlin: Springer.

Brunner, M-L., & Diemer, S. (2018). "Okay ... so ... nice to meet you? {smiles}" – Openings in ELF Skype conversations. In M. de los Ángeles Gómez González & J. L. Mackenzie (Eds.), *The Construction of Discourse as Verbal Interaction* (pp. 171-197). Amsterdam: Benjamins.

Carter, R., & McCarthy, M. J. (2004). Talking, creating: Interactional language, creativity and context. *Applied Linguistics, 25*(1), 62-88.

Clayman, S. (2002). *The News Interview: Journalists and Public Figures on the Air.* Cambridge: Cambridge University Press.

Dittman, A., & Llewellyn, L. (1968). Relationships between vocalizations and head nods as listener responses. *Journal of Personality and Social Psychology, 9*: 79–84.

Duncan, S., & Niederehe, G. (1974). On signalling that it's your turn to speak. *Journal of Experimental Social Psychology, 10*(3), 234–47.

Farr, F. (2003). Engaged listenership in spoken academic discourse: the case of student–tutor meetings. *Journal of English for Academic Purposes*, 2(1), 67-85.

Fellegy, A. M. (1995). Patterns and functions of minimal response. *American Speech, 70*(2), 186–99.

Fernández Polo, F. J. (2021). Backchannels in video-mediated ELF conversations: a case study. *Journal of English as a Lingua Franca, 10*(1), 113-140.

Fries, C. C. (1952). *The Structure of English*. New York: Harcourt, Brace and Company.

Gardner, R. (1998). Between speaking and listening: the vocalisation of understanding. *Applied Linguistics, 19*(2), 204–24.

Gardner, R. (2007). The Right connections: Acknowledging epistemic progression in talk. *Language in Society*, 36(3), 319-341.

Goffman, E. (1981). *Forms of talk*. Philadelphia, PA: University of Pennsylvania Press.

Hadar, U., Steiner, T. J., & Clifford Rose, F. (1985). Head movement during listening turns in conversation. *Journal of Nonverbal Behavior, 9*(4), 214-228.

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson, & J. Heritage (Eds.), *Structures of Social Interaction: Studies in Conversation Analysis* (pp. 299-345). Cambridge: Cambridge University Press.

Heylen, D., Bevacqua, E., Tellier, M., & Pelachaud, C. (2007). Searching for Prototypical Facial Feedback Signals. *Proceedings of IVA: International Virtual Agents* (pp. 147-153), Paris, France.

Ishi, C.T., Liu, C., Ishiguro, H., & Hagita, N. (2010). Head motion during dialogue speech and nod timing control in humanoid robots. Proceedings of the *2010 5th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 293-300), Osaka, Japan.

Jokinen, K. (2011). Turn taking, Utterance Density, and Gaze Patterns as Cues to Conversational Activity. *Proceedings of ICMI 2011 Workshop Multimodal Corpora for Machine Learning* (pp. 31-36), Alicante, Spain.

Jucker, A.H., & Landert, D. (2015). Historical pragmatics and early speech recordings: Diachronic developments in turn-taking and narrative structure in radio talk shows. *Journal of Pragmatics, 79*, 22-39.

Knight, D. (2011a). *Multimodality and active listenership: a corpus approach*. London: Bloomsbury.

Knight, D. (2011b). The future of multimodal corpora. *Brazilian Journal of Applied Linguistics (BJAS), 11*(2), 391-416

Knight, D., & Adolphs, S. (2008). Multi-modal corpus pragmatics: The case of active listenership. In J. Romero-Trillo (Ed.), *Pragmatics and corpus linguistics*, 175-190.

Love, R. (2020). *Overcoming Challenges in Corpus Construction.* London: Routledge.

Malisz, Z., Włodarczak, M., Buschmeier, H., Skubisz, J., Kopp, S., & Wagner, P. (2016). The ALICO corpus: analysing the active listener. *Language Resources and Evaluation, 50*(2), 411–442.

Marlow, E., & Wilson, P. O. C. (2012). *The breakdown of hierarchy: communicating in the evolving workplace.* London: Routledge.

Maynard, S. K. (1997). Analyzing interactional management in native/non-native English conversation: a case of listener response. *International Review of Applied Linguistics, 35*(1), 37–60.

McCarthy, M. J. (2002). Good listenership made plain: non-minimal response tokens in British and American spoken English. In R. Reppen, S. Fitzmaurice, & D. Biber (Eds.), *Using Corpora to Explore Linguistic Variation* (pp. 49-72). Amsterdam: John Benjamins.

McCarthy, M. (2003). Talking back: "Small" interactional response tokens in everyday conversation. *Research on language and social interaction*, 36(1), 33-63.

McNeill, D. (2000). Action and Thought. In D. McNeill (Ed.), *Language and Gesture* (pp. 139-140). Cambridge: Cambridge University Press.

Mott, H., & Petrie, H. (1995). Workplace interactions: women's linguistic behaviour. *Journal of Language and Social Psychology, 14*(3), 324–36.

O'Keeffe, A. (2006). *Investigating media discourse*. London: Routledge.

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). Berlin: Mouton de Gruyter.

O'Keeffe, A., & Adolphs, S. (2008). Using a corpus to look at variational pragmatics: response tokens in British and Irish discourse. In K. P. Schneider, & A. Barron (Eds.), *Variational Pragmatics* (pp. 69-98). Amsterdam, Netherlands: John Benjamins.

Oreström, B. (1983). *Turn-Taking in English Conversation.* Lund, Sweden: LiberFörlag Ltd.

Paggio, P., & Navarretta, C. (2017). The Danish NOMCO corpus: Multimodal interaction in first acquaintance conversations. *Language Resources and Evaluation, 51*(2), 463–494.

Tottie, G. (1991). Conversational style in British and American English: the case of backchannels. In K. Aijmer, & B. Altenberg (Eds.), *English Corpus Linguistics* (pp. 254-271). London: Longman.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication, 57*, 209-232.

Wilcox, S. (2004). Language from gesture. *Behavioral and Brain Sciences,* 27(4), 524–5.

Yngve, V.H. (1970). On getting a word in edgewise. *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society,* 567–77.

Appendix – guide to acronyms

BNC – British National Corpus

CANCODE – Cambridge and Nottingham Corpus of Discourse in English

CNV – convergence token

CON – continuer

DCC – Dublin City Council

ER – engaged response token
IR – information receipt Token
IVO – Interactional Variation Online [project name]
MFH – multiple, fast, high
MFL – multiple, fast, low
MSlH – multiple, slow, high
MSlL – multiple, slow, low
NBC – nod backchannels
NCoL – Nottingham Cities of Literature
NGO – Non-Governmental Organisation
SBC – spoken backchannels
SFH – single, fast, high
SFL – single, fast, low
SSlH – single, slow, high
SSlL – single, slow, low
TaLC – Teaching and Language Corpora [committee]
UH – unheard