Journal of Pragmatics 228 (2024) 1–16

Contents lists available at ScienceDirect

Journal of Pragmatics

journal homepage: www.elsevier.com/locate/pragma

Backchannels are not always very short utterances. The case of Italian Multi-Unit Backchannels

Daniela Mereu [a, *], Francesco Cangemi [b], Martine Grice [c]

[a] University of Turin, Italy
[b] Tokyo University of Foreign Studies, Japan
[c] Universität zu Köln, Germany

ARTICLE INFO

Article history:
Received 18 October 2023
Received in revised form 11 May 2024
Accepted 14 May 2024

Keywords:
Backchannels
Feedback signals
Italian
Conversational speech
Duration

ABSTRACT

Backchannels (BCs) are signals produced by conversation partners to support the ongoing turn of the interlocutor. Audible backchannels can have a lexical (*yes*) or non-lexical form (*mhm*), consisting of one unit (*mhm*) or multiple units (*mhm yes*). In this paper, we analyse the form and function of Multi-Unit Backchannels (MUBs) in Italian conversational speech. Data are drawn from the DIA (Dialogic ItAlian) corpus, a collection of spontaneous and informal conversations between dyads of Italian speakers who know each other well. In this corpus, MUBs represent over 29% of total backchanneling signals. While most single-unit backchannels fulfil a single function, with *Continuer* being the most frequent, the majority of MUBs fulfil multiple functions simultaneously, usually involving *Agreement*. Although backchannels are regarded as very short utterances, a large proportion of MUBs are, in fact, not short in duration. Interestingly, in our corpus MUBs are produced more frequently by speakers who do not take a prominent role in the interaction.

1. Introduction

People involved in conversation tend to avoid talking all at once and to collaborate in managing the alternation of turns (Sacks et al., 1974). On the surface, at a given moment in the conversation only one interlocutor has the turn and engages mainly in speaking activities, while the other interlocutor is mainly listening. In truth, however, both interlocutors are simultaneously engaged in both speaking and listening, because the person playing a listener's role co-constructs the current speakers' talk. This is achieved via gestural and positional cues, such as gaze and head movements, and via vocal signals, such as *yes*, *mhm*, *okay*. These very short utterances represent crucial messages produced by listeners to support the ongoing turn of the interlocutor, by signalling attention, agreement or understanding (Yngve, 1970). In general, these signals play an essential role in managing conversation and, importantly, they contribute to the success of the communication and to the establishment of interpersonal rapport, based on mutual attentiveness, reciprocal exchange of positivity cues, and coordination of nonverbal behaviours (e.g. Tickle-Degnen and Rosenthal, 1990) over the course of a conversation. Recipients producing backchannels do not generally remain recipients for the duration of the entire conversation: even if they initially only produce feedback signals while the partner in the conversation is speaking, they may then use other signals to take the floor (Drummond and Hopper, 1993b).

* Corresponding author.
E-mail address: daniela.mereu@unito.it (D. Mereu).

https://doi.org/10.1016/j.pragma.2024.05.003

The terminology used to refer to these very short utterances varies across studies, resulting in a wide range of labels which overlap only partially, and which emphasize different specific functions of these signals. Among the different labels are 'signals of attention' (Fries, 1952), 'reactive tokens' (Clancy et al., 1996), 'response tokens' (Gardner, 2001), 'listener responses' (Maynard, 1990), 'acknowledgment tokens' (Jefferson, 1984), 'acknowledge moves' (Carletta et al., 1997), and 'backchannel response' (Yngve, 1970). In this paper we adopt the term 'backchannels' (BC), a label which remains neutral with respect to the conversational function (cf. Ward and Tsukahara, 2000). This term is also more inclusive, in the sense that it refers to items uttered in the back channel of the dialogue (i.e. speech produced by the person who does not have the floor) vs. the front channel (i.e. speech produced by the person who has the floor) (Bangerter and Clark, 2003; Peters and Wong, 2015). This neutral label can then be further specified with one or more discourse functions (see §1.1 and §2.2).

With the term 'backchannels' we refer to the short productions uttered by one participant in the conversation when the other participant occupies the floor. At their core, these verbal reactions show that the speaker has heard the contribution of the partner, but they also often show that this contribution has been understood and accepted. Even after explanations and instructions, BCs are not mandatory, because they can be provided non-verbally, especially in face-to-face contexts, and because the other participant may not expect them to occur (Carletta et al., 1997; Ward and Tsukahara, 2000).

According to our understanding (and following Ward and Tsukahara 2000), in order to be recognised as a BC, a specific utterance:

a) must be addressed to the content expressed in the utterance produced by the other speaker;
b) must be optional;
c) does not require a reaction from the other.

As pointed out by Ward and Tsukahara (2000), these three criteria are based on the perspective of the interlocutor producing the utterance. In our work, every utterance that did not meet these three conditions was not considered a BC. These parameters were strictly applied during the coding phase to distinguish between BCs and non-BCs. For example, all responses to a question do not fall under the category of BC, because they are not optional, but expected.

Based on this assumption, this paper focuses on Multi-Unit Backchannels (MUBs), i.e. BCs composed of multiple units, in a corpus of Italian conversational speech. Our aim is to better understand the functional and formal aspects of this phenomenon, and our findings suggest that our understanding of the form, use and function of backchanneling signals would be limited if MUBs were ignored.

1.1. Functional properties of backchannels

Backchannels can fulfil a variety of discourse and conversational functions. They can be used by the listener to express attention, understanding, agreement with the interlocutor, or a combination of these. At the same time, they can also serve as an invitation for the main speaker to continue to talk or, conversely, as a strategy to take the floor (Schegloff, 1982; Jefferson, 1984; Clancy et al., 1996; Jurafsky et al., 1998). Although many different function-based classifications of BCs have been proposed, we restrict our review of the literature to the principal functional categorizations.

Schegloff (1982) used the term 'continuers' for items that are used to signal attention and understanding. For example, *uh huh* may be used by listeners to signal that they understand that the talk produced by the current speaker is not finished. Another category of BCs contributing to discourse co-construction recognised by Schegloff (1982) is represented by 'assessments', e.g. *oh wow* and *really?* These utterances show an additional meaning related to a listener's reaction to the current turn, such as surprise (see also Goodwin, 1986).

With respect to the role played by BCs in mediating the exchange of turns in dialogue, Jefferson (1984) introduces the concept of 'passive recipiency' for those BCs used when the listener does not have the intention of starting a turn. In these cases, the listener recognizes that the co-participant is the current turn-holder and can continue to talk. By contrast, 'incipient speakership' refers to utterances which display the willingness of the listener to move from recipiency to speakership. In the study by Drummond and Hopper (1993a), both conversational functions ('passive recipiency' and 'incipient speakership') were related to the sequential location in which BCs occur in an ongoing stream of talk. The position of BCs in the speech sequence has been shown to be a factor that can predict the function of BCs in maintaining the turn or seeking turn change. Moreover, some BCs tend to occur in a specific position, e.g. *okay*, *oh* and *that's beautiful* occur near the end of speech sequences (Heritage, 1984; Merritt, 1984; Mandelbaum, 1987, 2013; Drummond and Hopper, 1993b).

Jurafsky et al. (1998) included four kinds of utterances in the category of BCs: continuers, assessments, incipient speakership, and agreements. The latter type refers to items that mark the degree of acceptance by the producer toward the content expressed by another speaker.

All the different functions mentioned above can overlap, and even be conveyed simultaneously by one single item (Kjellmer, 2009). Crucially, one of the properties of the most common BCs is polyfunctionality (Peters, 1997; Ward and Tsukahara, 2000; Gardner, 2001; McCarthy, 2003; Peters and Wong, 2015).

Taking a perspective in which dialogue is used by people to coordinate the common, everyday activities in which they are engaged, Bangerter and Clark (2003) analysed these signals as project markers. The role of project markers is to contribute within the dialogue to the coordination of two different kinds of project transitions, and namely vertical transitions (i.e.,

entering and exiting joint projects) and horizontal transitions (i.e., proceeding within joint projects). They found that signals such as *okay* and *all right* are used for vertical transitions, while items such as *uh-huh*, *m-hm* and *yeah* are specialized for horizontal transitions, thus allowing participants to continue with the current joint project. Focusing now only on the type of BC that will be analysed here from a functional perspective (MUBs), previous work has paid special attention to interactional aspects, in terms of both turn management and interactional function (Tao and Thompson, 1991; Tottie, 1991; Clancy et al., 1996; Wong, 2000; Stivers, 2004; Wong and Peters, 2007), whereas until now other aspects of the interaction, such as conversational dominance, have been overlooked. For example, Wong and Peters (2007), in their study based on data from telephone conversations in Australian and New Zealand English, point out that the relative complexity of BC structure appears to be correlated with its interactional function. While most single BCs have the function of supporting the speaker holding the floor, so that they can continue their turn, the increased complexity associated with BC clusters suggests a shift in importance from supporting the speaker holding the floor to the content of the talk itself.

1.2. Formal aspects of backchannels

Vocal BCs can be lexical (*yes*) or non-lexical items (*mhm*) and they can be composed of one unit (*yes*) or multiple units (*ah yes*, *no no*). Tottie (1991) distinguished BCs into three categories on the basis of their form: simple (*yes*, *mhm*), double (with the repetition of the same form, such as *no no*, *ah ah*), and complex forms (clusters composed of different items, such as *oh yes*, *that's right yeah*).

In addition, BCs may be produced as a single piece of feedback to the contribution made by the participant who is talking (*standalone* BC), or occurring in a string of different items (*string* BCs) (Peters and Wong, 2015).

While simple BCs have been extensively investigated from diverse perspectives and for different languages (e.g. Clancy et al., 1996; Couper-Kuhlen and Selting, 1996; Müller, 1996; Jurafsky et al., 1998; Ward and Tsukahara, 2000; Beňuš et al., 2007; Savino, 2014; Ha et al., 2016; Zellers, 2021; Sbranna et al., 2022; Wehrle, 2022), to the best of our knowledge, very few studies have analysed MUBs (i.e., BCs composed of multiple units).

Formal properties of MUBs have been researched mainly in terms of their internal structure, but not yet in terms of their phonetic properties. In the above-mentioned study by Wong and Peters (2007), MUBs are shown to be fairly frequent. Although in previous studies the majority of BCs produced were single forms (i.e. *mhm*, *yeah*, see White, 1989), in this study MUBs accounted for more than one fifth (22.57%) of BCs in the Australian corpus and more than one third (36.90%) in the New Zealand corpus. Among MUBs, only a small proportion were reduplicated BCs: 13% for the Australian corpus and 11% for the New Zealand corpus. The majority are composed of complex BCs, i.e. made up of non-repeated elements (over 85% of MUBs in each corpus).

As far as temporal properties are concerned, previous work on BCs has appropriately emphasised their “very short” duration (Koiso et al., 1998; Tanaka, 1999; Galley et al., 2004; Young and Lee, 2004; Edlund et al., 2010), because they generally consist of only very few syllables (Gardner, 2001). Young and Lee (2004) found that in face-to-face conversations the duration of BCs (continuers or assessments) produced by American English native speakers show a median value of 0.39 s (min = 0.24 s; max = 3.19 s). Similar median values were reported by Peters and Wong (2015) for the highest frequency short BCs in Australian English (*mhm* and *yeah*) when in final position in the speech string. Edlund et al. (2009, 2010), in the Swedish Map Task Corpus and the Columbia Games Corpus respectively, reported that the vast majority of BCs do not exceed a 0.5 s threshold.

With respect to factors that may influence the duration of high-frequency BCs such as *yeah* and *mhm*, findings from Peters and Wong (2015) suggested that the context of occurrence (*string* vs. *standalone*) clearly affects properties such as the duration of the BC and of the preceding interval. In addition, median durations of *yeah* and *mhm* turn out to be relatively longer in initial position in a speech sequence than in final position.

The very short duration of BCs emphasised by the previous studies can be explained by the type of utterance generally examined, i.e., consisting of one or at most two units. Apart from a few studies, in most research on BCs, multi-unit signals are generally grouped into a separate category (e.g., ‘other’) and do not constitute the focus of the analysis.

1.3. Rationale

In this study, we present an analysis of the functional and formal properties of Multi-Unit Backchannels. To this end, we explore a corpus of Italian conversational speech and ask the following questions:

a) How frequent are MUBs in the conversations examined?
b) What is the distribution of repeated (or complex) and lexical (or non-lexical) realizations?
c) Do MUBs have the same functions as simple BCs?

From a phonetic point of view, we also address the following questions:

d) To what extent can MUBs be characterised as “very short utterances”?
e) Is there a relationship between duration and conversational function?

The current paper is structured as follows: in §2 we will provide information on the methodology, data, participants and labelling procedures, including BC categorization; in §3 we will report on the results of the analysis conducted, with a focus on formal and interactional aspects of MUBs; in §4 we will discuss the results and their implications for future studies.

2. Method

In the current analysis we combine procedures from both Conversation Analysis and quantitative analysis. Principles and methods of the conversation analytic approach (Sacks et al., 1974; Goodwin, 1979; Schegloff, 1982; Goodwin and Heritage, 1990; Gardner, 2001) have been used for the identification of each BC token and the coding of its function. This coding was performed from the perspective of the participants in the conversation. In other words, each utterance was conceived as a form of action produced and situated in a concrete conversational context. From this perspective, functions were assigned on the basis of what participants do and say by means of their interactional actions. A detailed analysis of the organization of different conversational moves was then applied to several examples used to illustrate the main findings through specific concrete cases (see §2.2). The quantitative analysis, carried out using data visualization (*R tidyverse packages*, R Core Team 2023; Wickham et al., 2019), was conducted to identify general trends in the data.

2.1. Corpus

The material for our study comes from the DIA-Dialogic ItAlian corpus (Mereu and Vietti, 2021), a collection of 19 spontaneous and informal Italian conversations between dyads of speakers who know each other well (friends, relatives, couples, work colleagues). The corpus, collected in Bolzano (South Tyrol, Italy), has a total duration of 9 h and 26 min. It is currently archived at *The Language Archive*, and is available for research purposes.[1]

The speaker sample consists of 38 participants (age range 18–65; 13 M, 25 F): 22 native Italian speakers, 12 Tyrolean native speakers who learned Italian typically from primary school onwards (Italian sequential bilinguals), and 4 simultaneous bilinguals. The speakers represent a varied pool of occupational and educational profiles.

Dialogues took place in the presence of the researcher. Before the recording started, speakers were invited to either choose a topic from a list of suggestions (e.g., the social and linguistic situations in Bolzano, South Tyrolean festivals and traditions, the political situation in Bolzano, the environment) or to share a recent experience (e.g. a trip, film or book). The speakers were free to choose the topics that best suited their interests. After approximately 10 min of conversation, the researcher pretended to have to leave the room with an excuse, leaving the two interlocutors alone (cf. Ernestus et al., 2014). The speakers continued to talk for approximately 10 min, then the researcher returned to the room and silently observed the ongoing interaction for the last 10 min. Therefore, the approximate average duration of the interactions is 30 min. Since the researcher made no attempt to manage the conversation, the speaking times of each interlocutor were not necessarily balanced across the dialogues.

Data were recorded using two headset microphones (Shure SM35) and a digital audio recorder (Zoom H4 recorder with 44,100 Hz rate and 16-bit depth). Speakers sat face to face, approximately one meter apart, and were able to see each other. Video recordings are not available for these interactions.

Recordings were divided into interpausal units separated by a minimum silence length of 200 ms (Heldner and Edlund, 2010; Levinson and Torreira, 2015). Audible in-breaths, sighs and other similar human noises were considered as silent intervals. Filled pauses such as *mhm* and *eh* were annotated as part of non-silent intervals, that is as segments of speech. Laughter was labelled separately. All speech produced by speakers (including interrupted words, repetitions and speech laughs) was transcribed orthographically.

For this study, we analysed the dialogues between Italian native speakers (all born and living in Bolzano), comprising 10 dialogues of about half an hour each, consequently 20 speakers in total, with a duration of 5 h and 24 min.

2.2. Labelling

The labelling procedure comprised three steps. In the first step, we carried out a conversational analysis of the dialogues based on both orthographic transcripts and audio files. BC tokens were orthographically annotated in *Praat* (Boersma and Weenink, 2022) following the guidelines proposed by Ward and Tsukahara (2000) (see §1). From possible BC candidates we excluded cases of repetitions of speech portions used in previous turns. Out of 1212 candidate BC tokens, we excluded 30 cases which were produced during laughter.

In the second step, each token was labelled on the basis of its lexical form, distinguishing between simple and Multi-Unit Backchannels, i.e. cases composed by units separated by a silence of less than 200 ms. The category of MUBs was then further divided into repeated forms (*sì sì sì* 'yes yes yes') and combined forms (*sì esatto* 'yes right') (Tottie, 1991; Wong and Peters, 2007). Each unit was further categorised according to its lexical form, i.e. lexical (*certo* 'sure') vs. non-lexical (*eh*), and according to its length, i.e. short (one syllable, e.g. *no* 'no') vs. long (more than one syllable, e.g. *giusto* 'right'). As a result, MUBs

[1] https://hdl.handle.net/1839/cbb7632f-9143-494a-b363-948c0e6c9ae8.

can take different forms, including repeated long lexical units (*certo certo* 'sure sure'), combined short non-lexical units (*ah ok*) and other patterns.

The third stage consisted in the annotation of interactional functions. After listening to the recordings again for this purpose, one or more functions were determined for each BC. Functions were determined using evidence from the sequential context, i.e. from the perspective of the speakers, as detailed in the examples below. First, the authors jointly identified prototypical examples of the main functional categories (see Examples 1–4). On this basis, the first author performed the rest of the annotation, and unclear cases were discussed among the authors. The possibility of assigning more than one function to each item reduced the risk of arbitrariness and subjectivity in the coding. At the end of the third labelling step, each BC was assigned one or more of the following functions (Schegloff, 1982; Jefferson, 1984; Jurafsky et al., 1998; Norrick, 2010; Zellers, 2021):

- Continuer: following passively what the interlocutor is saying, without claiming the floor (also referred to as "Passive recipiency", cf. Jurafsky et al., 1998). The item serves as a phatic tool to signal attention and recipiency, and its production has no direct effect on the course of the conversation.
- Incipient speakership: attempting to take the floor. In this case the BC is followed by further talk from the same speaker.
- Agreement: expressing a degree of acceptance towards the statement made by the interlocutor in the previous turn.
- Assessment: providing an evaluative or emotional reaction to what the interlocutor has said.
- Evoking: evoking a reaction from the interlocutor, for example when the information conveyed in the preceding utterance is new for the speaker producing the BC (Allwood et al., 1992, Zellers, 2021).

For each of these 5 functions, below we provide an example drawn from the DIA corpus. In these transcripts we follow a simplified version of the GAT conventions for the transcription of conversational data (Couper-Kuhlen and Barth-Weingarten, 2011). Specifically, square brackets are used to indicate overlapping speech, while pauses are indicated with one or more hyphens in round brackets, depending on their length: (-) corresponds to a short pause, (–) represents an intermediate pause, (—) signals a longer pause. Real proper names have been replaced by fictional names.

(1) D1 (Continuer)

B is telling A about a business trip she took the previous week. In this excerpt B has just told A that she slept in a hotel room that wasn't very clean and she felt like she got lice.

1	A	ma va [ma sarà stat-] ma no
2	B	[è possibile]
3	A	ma sarà stato un po' (-) una sug[gestione]
4	B	[la suggestione]
5	A	ma anche io quando Anna aveva i pidocchi ho passato
6		una setti[ma]na a grattarmi furiosamente
7	B	[**eh**]
8	A	poi [la Eli]sa mi ha guardato mi detto gua' che non hai niente
9	B	[**eh**]

1	A	*come on [it must have bee-] no way*
2	B	*[is it possible]*
3	A	*it must have been a bit (-) an im[pression]*
4	B	*[an impression]*
5	A	*me too when Anna had lice I spent*
6		*one [week] furiously scratching*
7	B	*[**eh**]*
8	A	*then [Eli]sa looked at me as said listen there's nothing there*
9	B	*[**eh**]*

In Example (1), *eh* (in lines 7 and 9) are used by B to signal that the listener is following what the current speaker is saying, but without claiming the floor ('passive recipiency', cf. Jurafsky et al., 1998), as evidenced by the fact that A continues her narration.

(2) D15 (Incipient speakership)

A and B are talking about the sale of A's car. B has just told A that he cannot sell his car at the price he would like.

```
1  A  Luca devi abbassarlo sto pr[ezzo]
2  B                              [ma no]
3  A  all[ora in in minimo]
4  B     [l'ho pagata ventitr]è e ci ho speso
5     altri tremila euro di gan[cio traino]
6  A                          [eh hai fatto un bru]tto un brutto investimento

1  A  Luca, you have to lower this pr[ice]
2  B                                 [but no]
3  A  the[n at least]
4  B     [I paid twenty-three for it] and I spent
5     three more thousand euros for the tow[bar]
6  A                                    [eh you made a ba]d a bad investment
```

The utterance *eh* produced by A in Example (2), line 6, is followed by other speech, i.e., she took the floor to say that her boyfriend (B) made a bad investment in buying the car they are talking about. This produces an overlapping string of speech. Since in this case A attempts to take the floor, we code this BC as 'incipient speakership'.

(3) D4 (Agreement)

A and B are discussing how the children of immigrants arriving in Bolzano from other countries are more frequently enrolled in the German school than in the Italian school.

```
1  A  quasi quasi tutti ormai preferiscono la scuola tedesca
2     ris[petto alla scuola italiana]
3  B     [perché capiscono ma] perché capiscono che qui da noi è un'opportunità
4  A  eh sapere il tedesco [e di conseguen]za loro dicono
5  B                       [esatto]
6  A  già sanno l'ita[liano la loro lingua]
7  B                 [-liano la loro lingua d' origine]

1  A  almost everyone now prefers the German school
2     compa[red to the Italian school]
3  B       [cos they understand but] cos they realize that here it is an asset
4  A  eh the fact of knowing German [and consequent]ly they say
5  B                                [exactly]
6  A  they already know Ita[lian their language]
7  B                       [-lian their native language]
```

In Example (3) B used *esatto* ('exactly') in line 5 to express agreement towards the previous statement uttered by A, i.e. the fact that immigrants recognise the importance of German language in Bolzano. The agreement is further indexed by the joint sentence completion in line 7.

(4) D8 (Assessment)

A is telling B about her trip to Lampedusa and talks about the pollution on the island. B says that in Turin (the city in which he is living), despite being the most polluted city in Italy, he does not perceive the pollution.

1	A	mamma mia un inquinamento (-) terribile su quello bella però
2	B	mhmh [no Torino devo dire]
3	A	[son stata contenta] di respirare un po'
4		di aria (--) [bolzanina]
5	B	[cioè Torino in teoria è la cit]tà più inquinata d'Italia
6	A	non è Milano?
7	B	Torino
8	A	**dai**
9	B	però (-) ti dico io non stando lì non lo percepisco di p- sì infatti

1	*A*	*my goodness terrible pollution on that beautiful though*
2	*B*	*Mhmh [no Turin I have to say]*
3	*A*	*[I was glad] to breathe some*
4		*air from [Bolzano]*
5	*B*	*[I mean Turin is theoretically the] most polluted city in Italy*
6	*A*	*isnt'it Milan*
7	*B*	*Turin*
8	*A*	***no way***
9	*B*	*however I tell you not living there I don't feel it mo- yes indeed*

The BC *dai* ('no way!'), in Example (4), line 8, conveys an assessment value (that is, disbelief) towards the content expressed by speaker B, i.e. the fact that Turin is the most polluted city in Italy. This interpretation is corroborated by the fact that in line 9 speaker B feels the need to self-interrupt and to elaborate on the assessment of disbelief (*di p- sì infatti*, 'mo- yes, indeed').

(5) D8 (Evoking)

See Example (4) for context.

1	A	era mh io mh infatti ho un po' sofferto in motorino ti s- (---)
2		ovviamente se mi sei davanti mi sga[si addo]sso
3	B	[mhmh]
4	A	ma m- mi veniva mal di [tes]ta
5	B	[**sì**]
6	A	mamma mia un inquinamento (---) terribile su quello bella però

1	*A*	*me mh in fact I suffered a bit on the moped s- (---)*
2		*of course if you are in front of me you are crun[ching on] me*
3	*B*	*[mhmh]*
4	*A*	*but m- I got a [head]ache'*
5	*B*	*[**yes**]*
6	*A*	*Goodness, such a pollution, terrible about that, beautiful though*

In Example (5) *sì* ('yes') uttered in line 5 can be considered a continuer, since it signals that the listener is following what A is saying. However, it also has a supportive evoking function, since it displays surprise and evokes a further reaction from the interlocutor. This reaction is provided in the next turn, i.e. *mamma mia un inquinamento (—) terribile su quello bella però* (line 6). While the reaction to the Assessment BC in Example (4) was immediately relevant to the content of the previous turn (i.e. Turin being indeed the most polluted Italian city), the reaction to the Evoking BC in Example (5) does not address the content of the previous turn in a specific way. In the whole dataset analysed there were no cases of BCs with the sole function of Evoking, and only 5 cases where this function was combined with other functions. For these reasons, after the third stage of labelling we excluded these cases from the quantitative analysis.

3. Results

In the following section, we provide an overview of key results, offering counts and visualizations of frequency, form, function, and distribution of MUBs.

3.1. Frequency and lexical aspects

The analysis carried out on the subsection of DIA corpus under investigation has shown that the percentage of MUBs is considerable, accounting for 29.1% of the total BCs (344/1182). The most frequent lexical types are: *sì sì* ('yes yes'), *eh sì* ('uh yes'), *sì sì sì* ('yes yes yes'), *ah okay* ('oh okay'), and *no no* ('no no').

Among the 344 MUBs we analysed, those starting with a non-lexical item (such as in *eh sì* 'uh yes', *ah okay* 'oh okay') correspond to 47.4%, while 49.4% begin with a short (monosyllabic) lexical item (such as in *sì sì* 'yes yes'). Only 3.2 % start with a long lexical item (such as in *infatti infatti* 'indeed indeed').

Almost 90% of all MUBs are composed of two (65.4%) or three units (23.8%). The remaining tokens consist of four-unit BCs (4.9%) and five-unit BCs (2.9%). Utterances with more than five units represent a very small percentage (2.9%) – one exceptional case features eight units, *no no no no no no no certo* 'no no no no no no no of course'.

In terms of the internal relationships within the units, 50% include a repetition of one or more items (*eh sì certo certo* 'uh yes of course of course'; *no no ovvio* 'no no clear'), while the other 50% consist of complex forms with no repetition (*eh beh sì* 'uh well yes'; *no infatti* 'no indeed'). The distribution between these two types is perfectly balanced, whereas, in the case of the corpora examined in Wong and Peters (2007), the majority consisted of complex BCs and only a small percentage (11–13%) of repeated BCs.

Out of 344 MUBs, only 7 (2%) have a short silence separating different units. MUBs with a silence of more than 200 ms between each unit were treated as separate BCs, whereas if the silence between two units was shorter than 200 ms, we considered it a multi-unit BC.

3.2. Temporal aspects

Overall, the duration of BCs ranges between 0.06 and 2.37 s (M = 0.45; SD = 0.32). Single BCs show a median value of 0.34 s (SD = 0.17; min = 0.06; max = 1.8), while the median duration of MUBs is 0.7 s (SD = 0.4; min = 0.18; max = 2.37). The median duration of single BCs is similar to measures already reported in other studies for other languages (Young and Lee, 2004; Peters and Wong, 2015), while the duration values of MUBs are higher.

Interestingly, almost 30% (28.6%) of total BCs show a duration greater than 500 ms (Fig. 1), and 68% of all MUBs have durations greater than 500 ms. The Figure shows that the duration increases with the number of syllables, but also that MUBs tend to be longer than single BCs with the same number of syllables.

As shown in Fig. 1, and in line with the available literature, most simple BCs with one or two syllables fall below the 500 ms threshold. However, this threshold would cut off most MUBs and also most BCs with more than two syllables.

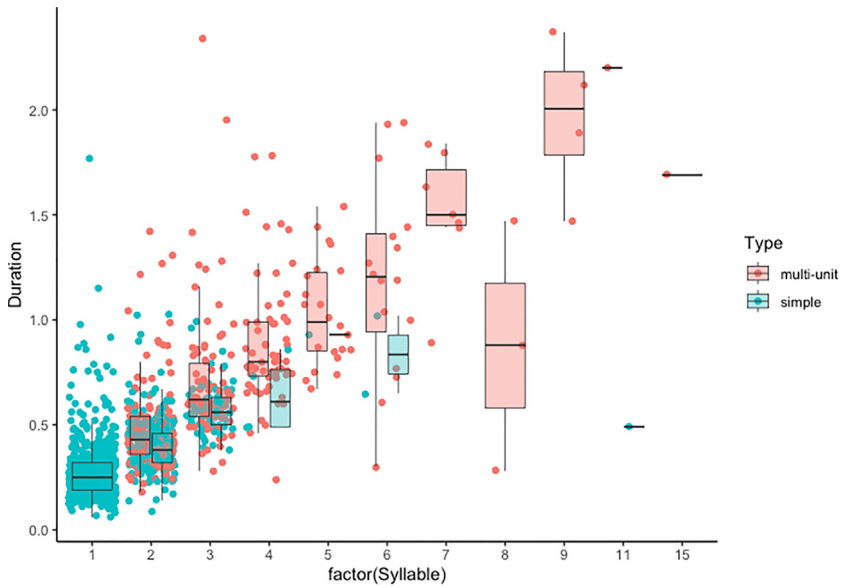

Fig. 1. Duration (in seconds) of multi-unit and simple backchannels, by number of syllables.

3.3. Functional aspects

Fig. 2 visualises the results for the functional properties of backchannel signals, separately for the 344 cases of MUBs (top row) and the 838 cases of simple BCs (bottom row).

Simple BCs were found to be predominantly monofunctional (65.6%), with Continuer being the most frequent function (see red cell in bottom row).

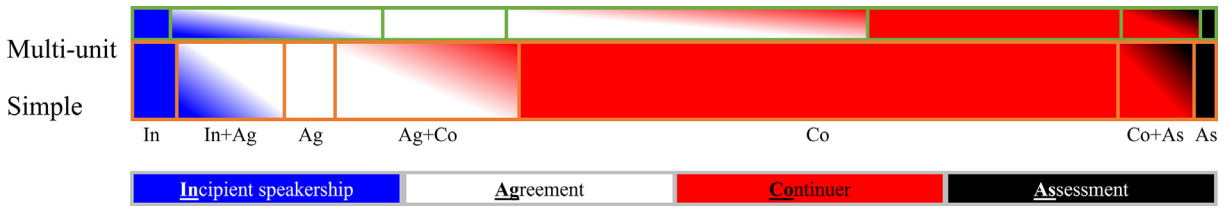

Fig. 2. Function of multi-unit and simple backchannels.

MUBs, on the other hand, were found to have more than one function (60.5%), with Agreement being the most frequent, either on its own, or in combination with Continuer or Incipient speakership (see cells containing white in the top row).

Focussing on MUBs (Fig. 3), it is worth noting that tokens with the shortest duration are used for incipient speakership (In), while utterances conveying an assessment function, although few and variable, tend to be longer (As). These results are not surprising, given that the role of incipient speakership BCs is usually merely to take the floor, while assessment BCs signal an evaluation in the content of the talk at hand (e.g. *ma va ma va* 'you don't say you don't say').

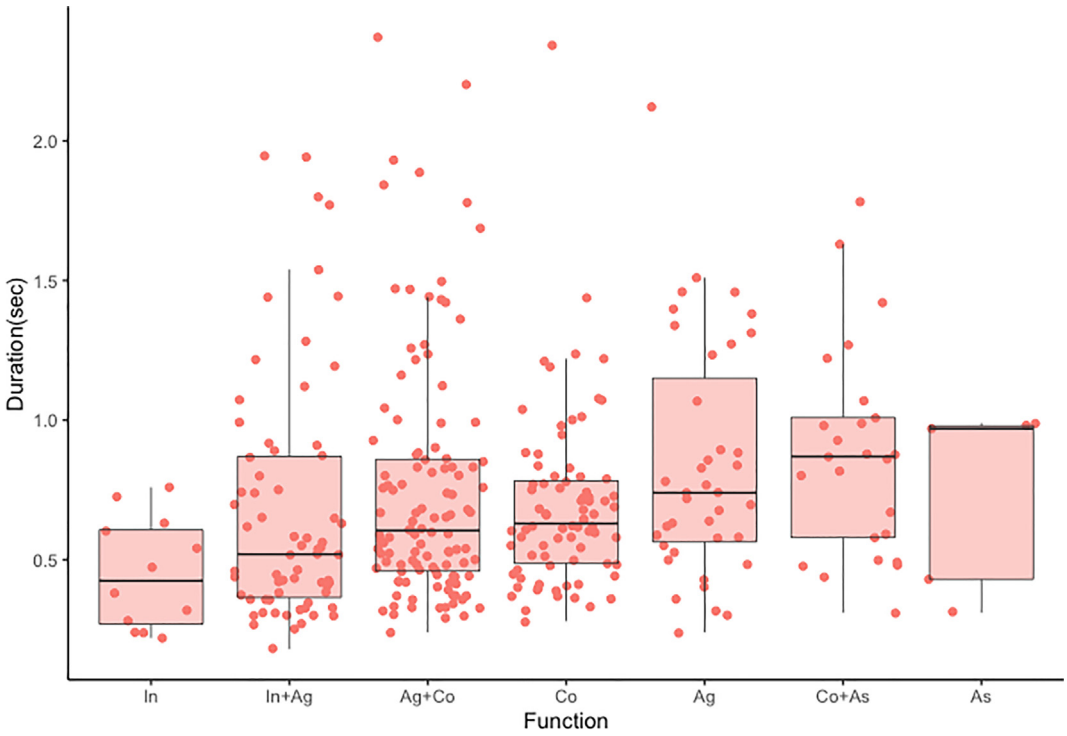

Fig. 3. Duration of multi-unit backchannels, by function.

3.4. Interactional aspects

MUBs are not produced with the same frequency in all dialogues. Fig. 4 shows the distribution of MUBs across dyads (symbols) and speakers (individual points). Two speakers from dyads D2 and D4 stand out as producing a large number of MUBs (i.e. high values on the y-axis). These speakers are also those who talk considerably less than their interlocutors, both in terms of word tokens and of speech time (i.e. low values on the x-axis).

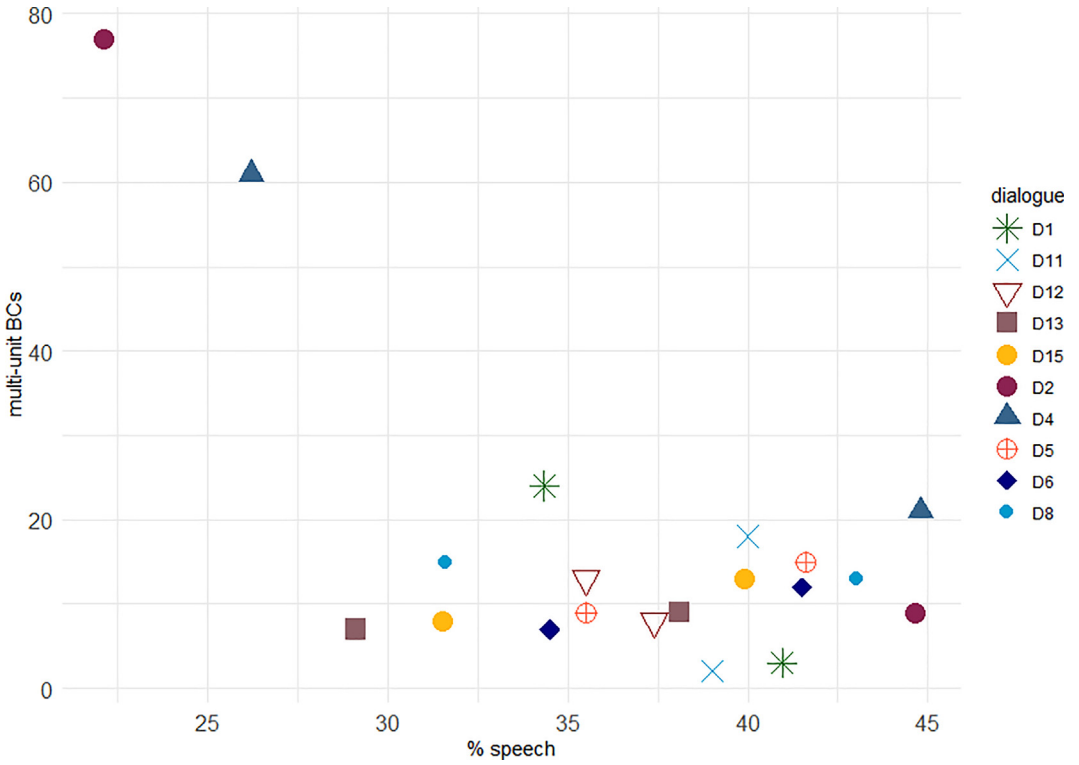

Fig. 4. Distribution of multi-unit BCs across dialogues and speakers, according to speaker activity. Speakers of the same dyad are represented by the same symbol.

The remaining dyads do not seem to support the claim of an inverse relationship between the frequency of produced MUBs and the amount of speech activity. However, given the exceptionally high number of MUBs produced by the less talkative speakers in dyads D2 and D4, in the following we provide a closer look at one of these interactions. Example 6 provides the verbal transcript of about one and a half minutes of the conversation produced by D2.

(6) D2 (00:00:58 – 00:02:32)

Speakers A and B talk about pollution in Bolzano.

A e per quanto ri- per esempio le fabbriche che dici però a me è successo
di dovergli mandare (-) a sensorcivico che mi sono iscritto che è quel
(--) un un un sito del comune dove puoi (--) co[municare segn]alare
B [mh segnalare]
A le cose e del del fumo che usciva delle acciaierie
B **ah** [**sì sì**]
A [e non ero l'unico] perché ce [n'erano tanti che]
B [**sì sì**]
A l'hanno segnalato [però quel fumo]
B [c'era anche sul giornale]
mi [sembra]
A [sì però quel fumo] continua ad uscire (--) [regolarmente]
B [tranquillamente]
A e una cosa secondo me se fanno se c'è abbastanza attenzione
non è neanche vero perché (---) io avevo fermato due vigili urbani(--)
chiedendo se era normale (--) che quel fumo uscisse da quella fabbrica
e hanno detto che è una cosa che non riguarda loro che devo rivolgermi
rivolgermi io [all'ufficio all'ufficio am]bientale
B [all'ufficio preposto]
A e mi ha detto da chi andare quindi anche da quel lato lì

```
21  B  sì [sì]
22  A     [mentre] magari se c'è una macchina che fa un (--) pelo di fumo
23     in più (--) allora ecco che viene controllata così come ci sono anche
24     altre (--) regole secondo me che non sono giuste (--) eh (--) ad esempio
25     sempre nell'ambito della macchina perché io che ho una macchina vecchia
26     un euro diesel due sono bloccato non posso girare anche se la uso
27     pochissimo (--) e poi ci sono le macchine che hanno il filtro
28     antiparticolato come (--) eh il gommista sotto casa mia che l'altro
29     giorno si stava facendo la pulizia del filtro quindi è una macchina che
30     va per un'ora fumando come un
31  B  sì perché [ci vuole tempo prima che entri] in [vigore in fun]zione
32  A           [una ciminiera]                    [esatto]
33  B  [no]
34  A  [e però] quella è un filtro antiparticolato che secondo loro inquina
35     di meno quando invece è una delle cose più dannose che ci sia
36  B  sì sì [no infatti]
37  A        [lì ci sono] veramente pochi pochi controlli

 1  A  and regarding for example the factories you say but it happened to me
 2     I had to send (-) to sensorcivico that I signed up for which is that
 3     a a a website of the municipality where you can com[municate re]port
 4  B                                                      [mh report]
 5  A  things and about the smoke that was coming out of the steel mills
 6  B  ah [yes yes]
 7  A     [and I wasn't the only one] because there [were many who]
 8  B                                              [yes yes]
 9  A  Reported it [but that smoke]
10  B              [it was in the newspapers too]
11     I [think]
12  A    [yes but that smoke] keeps coming out (--) [regularly]
13  B                                              [without issues]
14  A  and one thing in my opinion if they do if there is enough attention
15     it's not even true because I had stopped two traffic cops
16     asking if it was normal for that smoke to be coming out of that factory
17     and they said that it's not their business, that I have to go
18     I should go [to the office the office for en]vironmental
19  B              [to the appropriate office]
20  A  and he told me whom to talk to so even on from that point of view
21  B  yes [yes]
22  A     [while] maybe if there is a car that puts out a little bit more smoke
23     then here it is checked as well as there are also
24     other rules in my opinion that are not fair eh for example
25     always speaking about cars why if I have an old car
26     a euro diesel two I'm stuck I cannot drive it around even if I use it
27     very little and then there are the cars that have the particulate filter
28     like eh the guy from the tire dealer shop under my house who the other
29     day was getting the filter cleaned so it's a car that
30     goes for an hour smoking like a
31  B  yes because [it takes time for it to] take [effect in ope]ration
32  A              [a smokestack]                [exactly]
33  B  [no]
34  A  [and yet] that is a particulate filter which they say pollutes
35     less when in fact it is one of the most harmful things there is
36  B  yes yes [no indeed]
37  A          [in that sense there are] very few controls
```

The extract shows that the contribution of two participants to the conversation is asymmetrical. Itakura (2001) suggests to evaluate symmetry in conversation along three dimensions: the number of word tokens produced (quantitative dimension), the efforts to set the direction of the interaction (sequential dimension), and the restriction of speaking rights, particularly through interruptions and overlaps (participatory dimension). The asymmetry along the quantitative and sequential dimensions is demonstrated by the fact that speaker A starts the topic and develops it through long, informative turns. On the other hand, speaker B only gives very brief responses and provides feedback to the content produced by A (e.g. lines 6, 10–11, 13, 19, 21, 33, 36), using different strategies such as simple backchannels, MUBs, repetitions, and sentence completion. With respect to the participatory dimension, speaker A appears to achieve conversational control, since his turns are followed by

positive responses. When speaker B tries to elaborate on the topic (e.g. lines 10–11, 31), his controlling actions are not successful, since speaker A restricts B's speaking rights through interruptions and overlap. Interestingly, in this conversational segment the interactionally prominent role of speaker A co-occurs with a high rate of MUBs production by speaker B (*ah yes yes*, *yes yes* (twice), *yes yes no indeed*).

In order to quantify these insights, Fig. 5 provides a content-free speech activity record (see Cangemi et al., 2023) for the first 10 min of dialogue D2. In this kind of visualization, time flows from left to right, and each line corresponds to 1 min of the interaction (i.e., 10 lines correspond to 10 min). The figure uses colour to display active states (speech), and blanks to display inactive states (no speech). Speech from speakers A and B are represented by blue and red segments, respectively. Their backchannel signals are represented with lighter hues, i.e. cyan and pink for speakers A and B respectively.

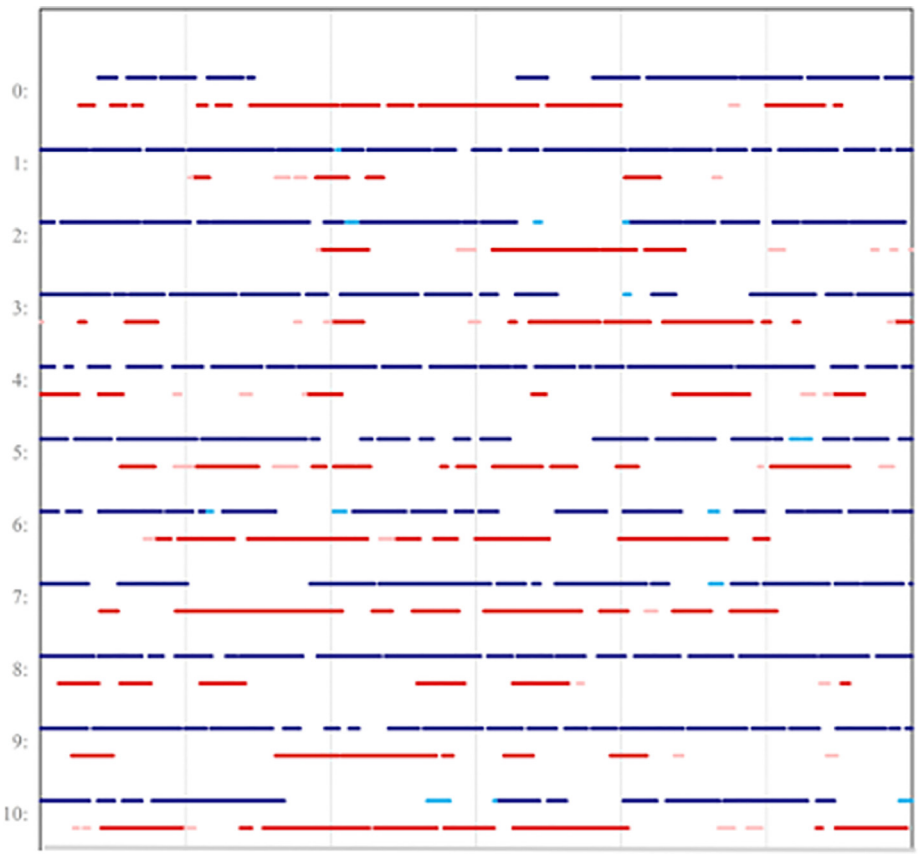

Fig. 5. Content-free speech activity record for the first 10 min of the Dialogue 2. Blue lines represent speech activity of A, while red lines represent activity of B. Lighter hues represent backchannels. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

The number and length of blue segments in Fig. 5 show that speaker A is extremely active, and produces comparatively few backchannel signals (cyan segments). Speaker B, on the other hand, remains frequently silent, produces many short speech turns (red segments) and a comparatively high number of backchannels (pink segments).

Crucially, as mentioned above, speaker B from dyad D2 is also the speaker who produces the highest amount of MUBs in our dataset. Fig. 6 illustrates the relationship between the number of produced MUBs and the degree of activity in the conversation. Data from dyad D2 is represented in the left panel. The plot shows speaker B in red and speaker A in blue. The top dimension shows that speaker B produces a large amount of MUBs, while the other four dimensions show that speaker A takes an interactionally prominent role. These dimensions are based on sequences of speech activities (Cangemi, 2023). In these sequences, each activity is coded as *S* (silence), *V* (overlap) or *x* (speech, either by speaker A or B, depending on the colour of the profile). Specifically:

- *xSx* is a sequence where one speaker speaks (*x*), goes silent (*S*), and resumes speaking (*x*). In other words, in this sequence the speaker's interlocutor does not produce overlapping speech and does not try to take the conversation floor.
- *SxS* is a sequence where a speaker's speech (*x*) is preceded and followed by a silence (*S*), i.e. without overlapping speech from the interlocutor.

- *duration of SxS* is the average duration of *SxS* sequences.
- *VxS/VxV* is the ratio between counts for *VxS* and *VxV* sequences. In *VxS* sequences, the conversation transitions from overlap (*V*) to a single speaker's speech (*x*) and then to silence (*S*). In *VxV* sequences, the conversation transitions from overlap (*V*) to a single speaker's speech (*x*) and then to yet another overlap (*V*), suggesting that the interlocutor is attempting to take an active role.

According to the definition of participatory dominance by Itakura (2001), interactionally prominent speakers can be expected to have higher values for *xSx*, *SxS*, *duration of SxS* and *VxS/VxV*.

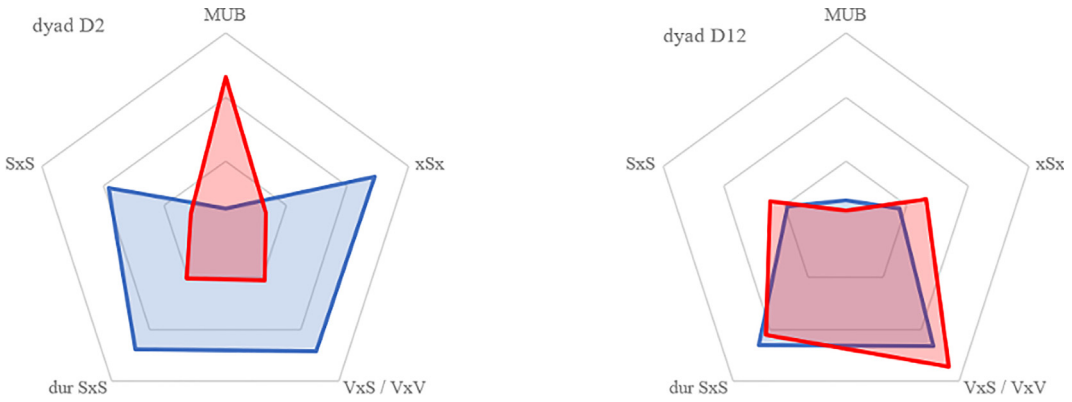

Fig. 6. Interactional profiles for an imbalanced (D2, left) and a balanced (D12, right) conversation.

Fig. 6 shows a spider plot with data extracted from dyad D2, in which speaker A (in blue) takes an interactionally prominent role, with high values for *SxS*, *duration of SxS*, *xSx* and *VxS/VxV*. On the other hand, speaker B (in red) replies with a high count of *MUB*s. The plot of data extracted from conversation D12, by contrast, represents a more balanced conversation, as shown by the large overlap between the profiles for speaker A (in blue) and speaker B (in red), dyad D12. Interestingly, neither of the two speakers from this dyad produced a noticeably large number of MUBs.

4. Discussion

In the current study we explored temporal and functional properties of Multi-Unit Backchannels, a relatively understudied type of backchannel signal.

Our finding that 30% of the overall BCs consist of Multi-Unit (repeated and complex) realizations suggests that these kinds of utterances represent a non-negligible part of conversational speech. As for the formal properties, most MUB tokens start with a non-lexical item (47.4%) or with a short lexical item (49.4%), while the remaining tokens begin with a long lexical item. Moreover, almost 90% of all MUBs consist of two or three units. With regard to their temporal properties, contrary to a common perception, a sizeable proportion of BCs are not short in duration, as we found that more than 28% of the total BCs and 68% of MUBs show a duration longer than 500 ms.

Crucially, this suggests that backchannelling behaviour can only be thoroughly assessed by considering the relatively understudied category of MUBs. Like simple backchannels, MUBs can also be expected to further interact with gestural, postural and other visual information.

The high percentage of MUBs with a relatively long duration may be related to the type of interaction we analysed. Perhaps unsurprisingly, previous studies featuring task-oriented dialogues (e.g. data from the Swedish Map Task Corpus in Edlund et al., 2009) highlighted the frequency and the importance of short, simple backchannels. On the other hand, MUBs seem to be more easily observable when interaction is more spontaneous, less task-oriented, and deals with familiar topics, as in our corpus (and in telephone conversation data in Wong and Peters, 2007). Before reaching conclusions about MUBs in communication in general, it is necessary to perform targeted explorations across different types of interaction.

In terms of conversational functions, the analysis revealed that MUBs tend to convey several functions simultaneously (Ward and Tsukahara, 2000; Gardner, 2001; Peters and Wong, 2015). MUBs often occur with an Agreement function, optionally supported by Continuer or Incipient speakership functions. For a thorough understanding of how MUBs are used to seek turn change, it would be useful to carry out a closer investigation of the sequential placement of MUBs, along the lines of Drummond and Hopper (1993b).

Moreover, the observation of how MUBs are distributed across the conversation and across speakers revealed that the type of BCs could be connected not only to specific conversational functions, but also to the amount of speech produced by the interlocutor. Specifically, in two cases we observed that a large number of MUBs are produced by speakers with dominant interlocutors (in the sense of Itakura, 2001), perhaps as a strategy to partially compensate for the asymmetry in the

interaction. Despite the exploratory and case-based nature of this analysis, the results highlighted how MUBs can shed light on both the management of conversation and the dominance relationships among participants.

Taken together, the findings of this paper suggest that MUBs represent an important part of backchanneling strategies in conversational speech. As such, they should receive appropriate attention in future studies on spoken interaction.

CRediT authorship contribution statement

Daniela Mereu: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Francesco Cangemi:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Martine Grice:** Writing – review & editing, Supervision, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 281511265, SFB 1252 Prominence in Language at the University of Cologne: “Individual behaviour in encoding and decoding prosodic prominence” (Principal Investigators: Martine Grice and Kai Vogeley). Open Access funding was enabled and organized by the University of Turin.

We would like to thank Michelina Savino and two anonymous reviewers for their valuable comments.

References

Allwood, Jens, Nivre, Joakim, Ahlsén, Elisabeth, 1992. On the semantics and pragmatics of linguistic feedback. J. Semant. 9 (1), 1–26. https://doi.org/10.1093/jos/9.1.1.

Bangerter, Adrian, Clark, Herbert H., 2003. Navigating joint projects with dialogue. Cognit. Sci. 27 (2), 195–225. https://doi.org/10.1207/s15516709cog2702_3.

Beňuš, Štefan, Gravano, Agustin, Hirschberg, Julia, 2007. The prosody of backchannels in American English. In: Proceedings of the 16th International Conference of Phonetic Sciences, pp. 1065–1068. https://doi.org/10.7916/D8WW7S1P.

Boersma, Paul, Weenink, David, 2022. Praat: doing Phonetics by Computer. https://www.praat.org.

Cangemi, Francesco, 2023. Visualising fluency: English interviews of Japanese students. In: Proceedings of the 37th General Meeting of the Phonetic Society of Japan, Sapporo, Japan, 2023.

Cangemi, Francesco, Grice, Martine, Janz, Alicia, Lucarini, Valeria, Spaniol, Malin, Vogeley, Kai, 2023. Content-free speech activity records: interviews with people with schizophrenia. Lang. Resour. Eval. https://doi.org/10.1007/s10579-023-09666-z.

Carletta, Jean, Isard, Amy, Isard, Stephen, Kowtko, Jacqueline C., Doherty-Sneddon, Gwyneth, Anderson, Anne H., 1997. The reliability of a dialogue structure coding scheme. Comput. Ling. 23 (1), 13–32.

Clancy, Patricia M., Thompson, Sandra A., Suzuki, Ryoko, Tao, Hongyin, 1996. The conversational use of reactive tokens in English, Japanese, and Mandarin. J. Pragmat. 26, 355–387. https://doi.org/10.1016/0378-2166(95)00036-4.

Couper-Kuhlen, Elizabeth, Barth-Weingarten, Dagmar, 2011. A system for transcribing talk-in interaction: GAT 2: English translation and adaptation of Selting, Margret et al. Gesprächsanalytisches Transkriptionssystem 2. Gesprächsforschung 12, 1–51.

Couper-Kuhlen, Elizabeth, Selting, Margret, 1996. Prosody in Conversation: Interactional Studies. Cambridge University Press, Cambridge.

Drummond, Kent, Hopper, Robert, 1993a. Back channels revisited: acknowledgment tokens and speakership incipiency. Res. Lang. Soc. Interact. 26 (2), 157–177. https://doi.org/10.1207/s15327973rlsi2602_3.

Drummond, Kent, Hopper, Robert, 1993b. Acknowledgment tokens in series. Commun. Rep. 6 (1), 47–53. https://doi.org/10.1080/08934219309367561.

Edlund, Jens, Heldner, Mattias, Pelcé, Antoine, 2009. Prosodic features of very short utterances in dialogue. In: Vainio, Martti, Aulanko, Reijo, Aaltonen, Olli (Eds.), Nordic Prosody - Proceedings of the Xth Conference. Peter Lang, Frankfurt am Main, pp. 57–68.

Edlund, Jens, Heldner, Mattias, Moubayed, Samer A., Gravano, Agustìn, Hirschberg, Julia, 2010. Very short utterances in conversation. Work. Pap./Lund University, Department of Linguistics and Phonetics 54, 11–16. https://doi.org/10.7916/D83B67MN.

Ernestus, Miriam, Kočková-Amortová, Lucie, Pollak, Petr, 2014. The Nijmegen corpus of casual Czech. In: Calzolari, Nicoletta, Choukri, Khalid, Declerck, Thierry, Loftsson, Hrafn, Maegaard, Bente, Mariani, Joseph, Moreno, Asuncion, Odijk, Jan, Piperidis, Stelios (Eds.), Proceedings of LREC 2014: 9th International Conference on Language Resources and Evaluation, pp. 365–370.

Fries, Charles C., 1952. The Structure of English. Longmans, Green and Company, London.

Galley, Michel, McKeown, Kathleen, Hirschberg, Julia, Shriberg, Elizabeth, 2004. Identifying agreement and disagreement in conversational speech: use of Bayesian networks to model pragmatic dependencies. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL 2004), pp. 669–676. Barcelona.

Gardner, Rod, 2001. When Listeners Talk: Response Tokens and Listener Stance. John Benjamins, Amsterdam.

Goodwin, Charles, 1979. The interactive construction of a sentence in natural conversation. In: Psathas, George (Ed.), Everyday Language: Studies in Ethnomethodology. Irvington Publishers, New York, pp. 97–121.

Goodwin, Charles, 1986. Between and within: alternative sequential treatments of continuers and assessments. Hum. Stud. 9, 205–217. https://doi.org/10.1007/BF00148127.

Goodwin, Charles, Heritage, John, 1990. Conversation analysis. Annu. Rev. Anthropol. 19, 283–307.

Ha, Kieu-Phuong, Ebner, Samuel, Grice, Martine, 2016. Speech prosody and possible misunderstandings in intercultural talk: a study of listener behaviour in standard Vietnamese and German dialogues. Proc. Speech Prosody 8, 801–805. https://doi.org/10.21437/SpeechProsody.2016-164. Boston.
Heldner, Mattias, Edlund, Jens, 2010. Pauses, gaps and overlaps in conversations. J. Phonetics 38 (4), 555–568. https://doi.org/10.1016/j.wocn.2010.08.002.
Heritage, John, 1984. A change of state token and aspects of its sequential placement. In: Atkinson, Maxwell, J., Heritage, John (Eds.), Structures of Social Action. Cambridge University Press, Cambridge, pp. 299–345.
Itakura, Hiroko, 2001. Describing conversational dominance. J. Pragmat. 33 (12), 1859–1880. https://doi.org/10.1016/S0378-2166(00)00082-5.
Jefferson, Gail, 1984. Notes on a systematic deployment of the acknowledgement tokens 'Yeah'; and 'Mm Hm'. Pap. Linguist. 17 (2), 197–216. https://doi.org/10.1080/08351818409389201.
Jurafsky, Daniel, Shriberg, Elizabeth, Fox, Barbara, Curl, Traci, 1998. Lexical, prosodic, and syntactic cues for dialog acts. In: Proc. ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers, pp. 114–120.
Kjellmer, Göran, 2009. Where do we backchannel? On the use of mm, mhm, uh huh and such like. Int. J. Corpus Linguist. 14 (1), 81–112. https://doi.org/10.1075/ijcl.14.1.05kje.
Koiso, Hanae, Horiuchi, Yasuo, Tutiya, Syun, Ichikawa, Akira, Den, Yasuharu, 1998. An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. Lang. Speech 41, 295–321. https://doi.org/10.1177/0023830998041004.
Levinson, Stephen C., Torreira, Francisco, 2015. Timing in turn-taking and its implications for processing models of language. Front. Psychol. 6, 731. https://doi.org/10.3389/fpsyg.2015.00731.
Mandelbaum, Jenny, 1987. Speakership Recipiency in the Accomplishment of a Telling. Unpublished doctoral dissertation. University of Texas, Austin, TX.
Mandelbaum, Jenny, 2013. Storytelling in conversation. In: Sidnell, Jack, Stivers, Tanya (Eds.), The Handbook of Conversation Analysis. Blackwell Publishing Ltd, Oxford, pp. 492–507.
Maynard, Senko K., 1990. Conversation management in contrast: listener response in Japanese and American English. J. Pragmat. 14 (3), 397–412. https://doi.org/10.1016/0378-2166(90)90097-W.
McCarthy, Michael, 2003. Talking back: small, interactional response tokens in everyday conversation. Res. Lang. Soc. Interact. 36, 33–63. https://doi.org/10.1207/S15327973RLSI3601_3.
Mereu, Daniela, Vietti, Alessandro, 2021. Dialogic Italian (DIA): the creation of a corpus of Italian spontaneous speech. Speech Commun. 130, 1–14. https://doi.org/10.1016/j.specom.2021.03.002.
Merritt, Marilyn, 1984. On the use of "OK" in service encounters. In: Baugh, Joel, Sherzer, John (Eds.), Language in Use. Prentice-Hall, Englewood Cliffs, NJ, pp. 167–196.
Müller, Frank Ernst, 1996. Affiliating and disaffiliating with continuers: prosodic aspects of recipiency. In: Couper-Kuhlen, Elizabeth, Selting, Margret (Eds.), Prosody in Conversation. Cambridge University Press, Cambridge, pp. 131–176.
Norrick, Neal R., 2010. Listening practices in television celebrity interviews. J. Pragmat. 42 (2), 525–543.
Peters, Pamela, 1997. Micro- and macrolinguistics for natural language processing. In: Wichmann, Anne, Fligelstone, Steven, McEnery, Tony, Knowles, Gerry (Eds.), Teaching and Language Corpora. Longman, London, pp. 175–185.
Peters, Pamela, Wong, Deanna, 2015. Turn management and backchannels. In: Aijmer, Karin, Rühlemann, Cristoph (Eds.), Corpus Pragmatics: A Handbook. Cambridge University Press, Cambridge, pp. 408–429. https://doi.org/10.1017/CBO9781139057493.022.
R Core Team, 2023. R: the R Project for Statistical Computing [computer programme], Version 4.2.2.
Sacks, Harvey, Schegloff, Emanuel A., Jefferson, Gail, 1974. A simplest systematics for the organization of turn-taking in conversation. Language 50, 696–735. https://doi.org/10.2307/412243.
Savino, Michelina, 2014. The intonation of backchannel tokens in Italian collaborative dialogues. In: Vetulani, Zygmunt, Mariani, Joseph (Eds.), Human Language Technology Challenges for Computer Science and Linguistics. LTC 2011. Lecture Notes in Computer Science. Springer, Cham, pp. 17–28.
Sbranna, Simona, Möking, Eduardo, Wehrle, Simon, Grice, Martine, 2022. Backchannelling across languages: rate, lexical choice and intonation in L1 Italian, L1 German and L2 German. In: 11th International Conference on Speech Prosody, 2022. https://doi.org/10.21437/SpeechProsody.2022-149.
Schegloff, Emanuel A., 1982. Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences. In: Tannen, Deborah (Ed.), Analyzing Discourse: Text and Talk. Georgetown University Press.
Stivers, Tanya, 2004. "No no no" and other types of multiple sayings in social interaction. Hum. Commun. Res. 30, 260–293. https://doi.org/10.1111/j.1468-2958.2004.tb00733.x.
Tanaka, H., 1999. Turn-taking in Japanese Conversation: A Study in Grammar and Interaction. John Benjamins, Amsterdam.
Tao, Hongyin, Thompson, Sandra A., 1991. English backchannels in Mandarin conversations: a case study of superstratum pragmatic 'interference'. J. Pragmat. 16 (3), 209–223. https://doi.org/10.1016/0378-2166(91)90093-D.
Tickle-Degnen, Linda, Rosenthal, Robert, 1990. The nature of rapport and its nonverbal correlates. Psychol. Inq. 1 (4), 285–293.
Tottie, Gunnel, 1991. Conversational style in British and American English: the case of backchannels. In: Aijmer, Karin, Altenberg, Bengt (Eds.), English Corpus Linguistics: Studies in Honour of Jan Svartvik. Longman, pp. 254–271.
Ward, Nigel, Tsukahara, Wataru, 2000. Prosodic features which cue back-channel responses in English and Japanese. J. Pragmat. 32 (8), 1177–1207. https://doi.org/10.1016/S0378-2166(99)00109-5.
Wehrle, Simon, 2022. A Multi-Dimensional Analysis of Conversation and Intonation in Autism Spectrum Disorder. Ph.D. dissertation. University of Cologne.
White, Sheida, 1989. Backchannels across cultures: a study of Americans and Japanese. Lang. Soc. 18 (1), 59–76. https://doi.org/10.1017/S0047404500013270.
Wickham, Hadley, Averick, Mara, Bryan, Jennifer, Chang, Winston, McGowan, Lucy D'Agostino, François, Romain, Grolemund, Garrett, Hayes, Alex, Henry, Lionel, Hester, Jim, Kuhn, Max, Pedersen, Thomas Lin, Miller, Evan, Bache, Stephan Milton, Müller, Kirill, Ooms, Jeroen, Robinson, David, Seidel, Dana Paige, Spinu, Vitalie, Takahashi, Kohske, Vaughan, Davis, Wilke, Claus, Woo, Kara, Yutani, Hiroaki, 2019. Welcome to the tidyverse. J. Open Source Softw. 4 (43), 1686. https://doi.org/10.21105/joss.01686.
Wong, Jean, 2000. Repetition in conversation: a look at "First and second sayings". Res. Lang. Soc. Interact. 33 (4), 407–424. https://doi.org/10.1207/S15327973RLSI3304_03.
Wong, Deanna, Peters, Pamela, 2007. A study of backchannels in regional varieties of English, using corpus markup as the means of identification. Int. J. Corpus Linguist. 12 (4), 479–509. https://doi.org/10.1075/ijcl.12.4.03won.
Yngve, Victor H., 1970. On getting a word in edgewise. In: Papers from the Sixth Regional Meeting. Chicago Linguistic Society, pp. 567–578.
Young, Richard, Lee, Jina, 2004. Identifying units in interaction: reactive tokens in Korean and English conversations. J. Sociolinguistics 8 (3), 380–407. https://doi.org/10.1111/J.1467-9841.2004.00266.X.
Zellers, Margaret, 2021. An overview of forms, functions, and configurations of backchannels in Ruruuli/Lunyala. J. Pragmat. 175, 38–52. https://doi.org/10.1016/j.pragma.2021.01.012.

Daniela Mereu is a Researcher in Linguistics at the University of Turin. Her research interests include sociophonetic variation, speech reduction, phonetic encoding of conversational structures in spontaneous speech and development of speech resources. Her work mainly focuses on Sardinian and Italian. After receiving her PhD in Linguistics from the Universities of Bergamo and Pavia, she worked as a post-doctoral researcher at the Free University of Bozen/Bolzano.

Francesco Cangemi is Adjunct lecturer and Research fellow at the Tokyo University of Foreign Studies. After defending a dissertation on prosody in Aix-en-Provence, he worked as a post-doctoral researcher in Cologne, focussing on the phonetics and phonology of linguistic prominence. He specializes in the analysis of individual differences in the production and perception of language and in communicative styles. He currently explores how conversational behaviour can be explored, visualized and quantified, with a focus on interactions between adults with Autism Spectrum Disorder, with a diagnosis of schizophrenia, and between language learners.

Martine Grice is Professor of Phonetics at the University of Cologne. She has served as President of the Association for Laboratory Phonology and edits Studies in Laboratory Phonology. Her work on intonation theory includes the analysis of complex tonal structures and the negotiation between tune and text. She has worked extensively on the intonation of Italian, English and German and has addressed specific challenges in Vietnamese, Tashlhiyt Berber and Maltese. More recently she has been working on Italian and German and investigating how intonation is used in attention orienting and how perspective taking abilities (especially in autism and schizophrenia) affect conversation management.