



Understanding Laughter in Dialog

Maria Koutsombogera¹ · Carl Vogel¹

Received: 19 July 2021 / Accepted: 31 March 2022 / Published online: 17 May 2022
© The Author(s) 2022

Abstract

This work explores laughter within a corpus of three-party, task-based dialogs with native and non-native speakers of English, each consisting of two players and a facilitator, in relation to whether the laughter is perceived as serving discourse functions or rather as genuinely mirthful according to a small number of annotators' (2) inspection of a substantial multimodal dialog corpus (18 interactions of approximately 10 min each). We test the hypothesis that those different types of laughter have occurrence patterns that relate in different ways to the topical structure of the conversations, with discourse laughter showing a stronger tendency to occur at topic termination points. All laughter events (569) are assigned to one of three values, discourse, mirthful or ambiguous, and are studied with respect to their distribution across the dialog topic sections. The analysis explores interactions among laughter type and section type, also with respect to other variables such as the facilitators' feedback and the speakers' conversational role and gender. Discourse laughter is more frequent at topic termination points than at topic beginnings, also in comparison to mirthful laughter. Discourse laughter is also highly associated with facilitators' feedback type, especially at topic ends. Finally, there are few distinctive effects of gender, and an interaction among speaker role and laughter type. The results strengthen the hypothesis of the discourse function of laughter, indicating a systematicity in discourse laughter, in that it is more predictable and highly associated with the dialog topic termination points, and, on the contrary, a less systematic distribution of mirthful laughter, which shows no particular pattern in relation to topic boundaries.

Keywords Conversational laughter · Topic boundaries · Laughter rate · Discourse laughter · Discourse marker

Introduction

Human interaction in dialog contains informative signals that go beyond the linguistic content of the communication, and it is important to basic knowledge of human cognition to identify how the signals within dialog are understood and used — concern with the interpretation of laughter, like many concerns about cognition, can be traced at least as far back as Socrates and Plato [1]. Laughter in spoken

interaction is an important form of paralinguistic expression, fulfilling a social function [2] such as displaying engagement and amusement, and, where laughter is shared, has been often described as a social cohesion mechanism [3]. In everyday interactions laughter can be very frequent and may take various forms in diverse contexts [4], having various effects or functions in the interaction when observed within its context [5]. In this work we argue that some forms of laughter are social signals which provide information about the discourse structure of dialog,¹ acting as a discourse connective does. We think that both “polite” laughter and “malicious” laughter fall into this category, but in this work we examine natural interactions in which malice does not arise (although malicious laughter appears to have been important to Socrates).

In linguistic theory, an idealization of language is that most uses of language are anchored in bi-directional signs, expressions produced by speakers in the same manner as

This research has benefited from support from the Science Foundation Ireland ADAPT Centre (Grant 13/RC/2106) and the EU H2020 Programme under the Marie Skłodowska-Curie grant agreement no. 701621 (MULTISIMO).

✉ Maria Koutsombogera
koutsomm@tcd.ie

¹ Computational Linguistics Group, Trinity Centre for Computing and Language Studies, School of Computer Science and Statistics, Trinity College, The University of Dublin, Dublin, Ireland

¹ We take discourse structure, in coarse grained terms, to be about the flow and relationships among successive topics.

understood by the same speakers when also acting as hearers, and that this is probably more efficient style of use than reasonable alternatives [6].² The analysis of laughter within linguistics and cognitive science may make use of any of a number of well-motivated taxonomies of laughter, including those that anchor the interpretation of laughter in laughter triggers, laughables. However, a methodological problem emerges in that only speakers know best what triggers their laughter, and it is not clear that speakers are reliable informants about what triggered any laugh, retrospectively. Laughter as a paralinguistic unit is fundamentally different from linguistic objects: a noun like “cat” may successfully denote spatio-temporal regions of the universe that contain cats, an adjective like “happy” may successfully denote spatio-temporal regions of the universe that contain happy individuals; any one laugh may be identified in relation to the space and time during which a person is laughing, but that appears to be the end of crispness in the denotation of laughter.

In as much as seemingly identical triggers in seemingly identical circumstances may not lead an individual to laugh in all such occasions, or even more than one such occasion, the fluidity of denotation of laughter creates a situation in which the linguistic expression with the closest correspondence might be “now”: outside fiction, “now” never points to the same spatio-temporal region of the universe twice. We think that speakers may be more reliable in evaluating whether they were amused or not for any instance of laughter than in specifying the exact amusement trigger and do not attempt to identify triggers. We also accept empirical evidence that the acoustic structure of mirthful laughter is different to that of the complement category of laughter. Our research explores the hypothesis that the laughter-complement of mirthful laughter is a kind of discourse connective. Familiar linguistic discourse connectives are “because”, “therefore”, “then” and so on. We are analyzing the extent to which instances of laughter are used or understood as a discourse connective, possibly as a paralinguistic counterpart of “I’m done with this topic”, “I don’t want to add to this topic”, or other alternatives. At present, we are interested in the dichotomy formed by “mirthful” laughter and “discourse” laughter, and hypotheses that relate these forms of laughter to the dynamics of dialog and properties of dialog participants that may be thought to interact with dialog dynamics (including gender, for example). Some of these dynamics may give a clue about the discourse relations underlying these forms of laughter. For example, if discourse laughter occurs more at the ends of topics than at the start of

topics, then it is probably not perceived or generated with the meaning, “I want to say more about this topic”.

We explore laughter in the MULTISIMO corpus of 3-party, task-based interactions carried out in English [7]. In each dialog, two players work together to provide answers to a quiz and are guided by a facilitator, who monitors their progress and provides feedback or hints when needed. The corpus was built to make available an English dataset that facilitates the analysis of interactional verbal and non-verbal qualities (e.g., gesture, gaze, linguistic content) in relation to participant qualities (e.g., personality traits measured with the big-five personality inventory) and perception by external raters (such as participant dominance and interlocutor collaboration). The corpus was thus not designed to exclusively address laughter studies. However, it is natural that conversational laughter occurs in the interaction among the group participants. Laughter events in the corpus exist throughout the conversation: they are used to ratify the speaker’s retention of the floor, to provide acknowledgments, to release tension and so on, or to spontaneously vocalize mirth. Laughter with the function “I’m listening; keep talking” is distinct from “I acknowledge that you just tried to be funny” and from “I agree to vocalize a smile at the same time that you are”, just as “next” is distinct from “because” and “moreover”. These and related forms are discourse connectives, that, in the case of laughter, provide social glue to segments of dialog contributed by participants.

For the present work, we stay very close to a binary distinction of laughter in the *Discourse* and *Mirthful* types, as we consider that this minimal distinction captures the predominant types of laughter in social interaction that can be easily perceived by external raters (and with reference to previous work from [8], who report on perceptual impressions of laughter). While this dichotomy has precedent in the literature, it is not without controversy. For example, a different labeling of the taxonomy (mirthful vs. polite) has been dismissed as conflating dimensions of analysis [9], noting that one may vocalize amusement while being polite or impolite. We do not dismiss more articulated taxonomies of laughter that have been developed for other studies to address different research questions, but continue to find theoretical and empirical interest in a binary distinction that captures observable features of the laugher’s demeanour and anchored in a spectrum with spontaneous unfettered mirth on the one side and contrived conscious control on the other. A natural binarization is between *Mirthful* and the complement, *non-Mirthful* laughter. We prefer a positive label for the complement and choose “*Discourse*”, as this corresponds, we think, to what is meant when referring to “polite” laughter — it is not about being polite, but using laughter in a manner that fits into dialog at certain moments just as do formulaic linguistic expressions, and the

² Acts of linguistic deception that depend on linguistic ambiguity are not precluded by this idealization, but are noteworthy for requiring more effort on the part of the speaker.

attitude behind those moments could as well be an intention to be rude as to be polite. Our primary annotations are of laughter of the individual with respect to this dichotomy.³ In contrast to other works, we do not annotate the laughter trigger [9], as we feel this often requires deeper inspection of speaker's consciousness than even speakers themselves have on reviewing recordings of their laughter and other dialog contributions. Undoubtedly, there is a subjective element to making annotations using this binary distinction since it does entail inferring whether a speaker has decided to laugh or succumbed to laughter. This appears to be delving into speakers' mental states. However, this is not far removed from the sort of reasoning that enters the annotation of speaker disfluency: a speaker may embark on the utterance of the "wrong word", and its annotation as disfluency depends upon it seeming not to have been intended. Objectivity may emerge from agreement in multiple subjective annotations.

In our work, laughter events in the dataset were manually annotated by two raters with either of the discourse and mirthful values on watching the dialog recordings, i.e., by having access to both audio and video signals. The disagreement among the two raters was then measured and it was used as a value of interest, i.e., the cases where the raters disagree were assigned with a new value, *Ambiguous* (cf. "[Laughter Classification](#)"). This is the extent to which we move beyond a binary classification of instances of laughter. We analyze them separately, but feel that those items thus given the *Ambiguous* label are best understood as having a discourse value — that is, we expect the instances of laughter that are deemed mirthful by one annotator and discourse by another are likely to pattern with discourse laughter in the independent measures we explore. We believe that the observed disagreement among raters is rooted in the nature of the task, i.e., the perception of something being funny, or not, and the fact that not all people perceive the same things as funny. We maintain that each of the laughter instances is likely to be independent of another, i.e., the annotation of one instance as *discourse* or *mirthful* does not entail that the next instance will be annotated as such. Therefore, when

³ It is necessary to consider problem cases for this dichotomy. An individual laugh that is a spontaneous outburst of mirth and simultaneously a signal of topic completion is not possible, on our view, inasmuch as acting on an *intent* to signal topic flow excludes the possibility that the laughter was spontaneous. Thus, in our view, laughter that seems spontaneous and mirthful and happens to occur at a topic boundary is recorded as *Mirthful*. Similarly, a laugh that appears to be impolite, a malicious laughing at someone in order to enhance the target's experience of humiliation, although it does not occur in our dataset, we would regard as an instance of *Discourse* laughter, not too far removed from the function that is typically transcribed with an exclamation mark, rather than as an episode of spontaneous mirth.

annotators come across a laughter event, they essentially have to acknowledge whether they perceive this as being amusing or not.⁴

All dialogs in the corpus analyzed here show an identical structure, consisting of three questions that are processed in a specified manner (i.e., by first identifying and then ranking the answers). To analyze properties of laughter, the dialogs were segmented in topics consisting of three sections, i.e., topic starts and ends (topic boundaries) and topic middle points; topic is thus defined as the part of the dialog related to either the identification or the ranking of answers (cf. "[Topic Segmentation](#)"). In addition, we look at laughter distribution during particular dialog moments, where facilitators provide feedback about the performance of the players, as feedback may trigger laughter responses by both players and facilitators.

We focus our analysis on the way laughter quantities, i.e., laughter frequency and duration, interact with dialog topic sections, speaker role, gender and feedback content, with the aim to address a set of hypotheses around the discourse function of laughter. Our background hypothesis related to the distinction between discourse and mirthful laughter is that we expect discourse laughter patterns to be more systematic than mirthful laughter, and mirthful laughter to be less associated with topic boundaries than discourse laughter. We believe that likely triggers for mirthful laughter are frequently possible to be uncovered in a post hoc manner or mirthful laughter may be related to a laugher-internal idea that has no direct visible counterpart in the dialog. In contrast, discourse laughter shows a more detectable structure, and we expect that more discourse laughter is found at topic termination points, implying that it functions similarly to a discourse connective.

⁴ One might ask whether one should obtain the annotations of many more than two raters in order to inform the judgements here. Of course, one could, but given that the annotation task, over the whole dataset, is arduous, while adding annotations might decrease the subjectivity of abstractions over all of them, it might also increase variability of judgements that make it necessary to study also factors associated with the annotators. This, too, would be valuable. Adding annotations would not guarantee higher agreement levels, because of the peculiarities of this subjective perception task and its dependence on the perception of the annotators of what is funny or amusing. Therefore, we also see value in starting with two annotations, treating annotations that agree as clear and treating annotations that disagree differently. In the same spirit as contemplating the impact of additional annotations of the laughter instances, one might ask whether one should study laughter in other datasets, as well. Of course, one should, but that has as pre-requisite determining if effects of interest appear in a subset of that data, in particular, the dataset that we have constructed. All of these angles merit exploration. We do not present our analysis as exhaustive, but rather as an important detailed study of a substantial corpus of interactions. Any one study is necessarily finite. Observational accounts [56] such as this one occupy a methodological space between anecdote [57] and large scale sampling [58].

Even with a binary distinction between discourse and mirthful laughter, a number of further dichotomies are relevant to the flow of dialog. A laughter event may be ratified by other dialog participants, or not. Laughter may interact with turn-taking, or not. For example, it is natural to imagine that shared laughter is likely at the mutually perceived end of a topic, a mutual acknowledgement that the topic has reached a conclusion. Possibly, the floor may be taken by a new speaker. On the other hand, one might imagine unrati-fied laughter to tend to accompany continuance of a topic and speaker. Just as by following the patterns of speaking or not in a dialog, one may have a sense of which interlocutor is dominating a conversation, one may inspect the patterns of laughter types in a dialog to make inferences about topic identity and speaker identity on either side of laughter events. Within the present work, we focus on the top-level binary classification, and we make observations in relation to this classification and other aspects of the dialog, such as gender effects.

By investigating the distribution of the *Ambiguous* label, we aim to find out more about those laughter events where annotators disagree about, i.e., the topic sections where they occur and their potential function. We expect cases in which annotators disagree to pattern more with discourse laughter, i.e., to function as a discourse connective and to be mainly located at topic termination points.

Our main hypothesis related to speaker role is that facilitators will produce more discourse laughter than players at the end of the topics, where their main task is to move to the next topic and keep on the dialog. We also test gender differences and their association with laughter and topic section types, without, however, having particular expectations about significant effects, as gender composition within groups was not among the controlled variables.

In relation to the interaction of laughter and facilitators' feedback, we expect that discourse laughter is more frequent than mirthful at the moments where feedback occurs, given that feedback is related to acknowledgements or used as politeness marker. We also expect feedback-related laughter to be mainly located at topic termination points (i.e., points where players' responses are evaluated and often trigger laughter reactions from participants) and that feedback at those points is mostly of a positive character (cf. "[Feedback Annotation](#)" for details about feedback types).

The above hypotheses are tested in "[Results](#)" and discussed in "[Discussion](#)"; in the next section we discuss the materials used for this study. The findings of this study highlight the discourse function of laughter, which shows systematic patterns, contrary to its mirthful counterpart, which is more arbitrarily distributed. Furthermore, discourse laughter was found to be prevalent at topic termination points, and highly associated with the moments where feedback is provided by the facilitators to the players. The conversational

role of facilitator presents higher discourse laughter rates than the role of the player; finally, females have in general more laughter instances, but shorter in duration than males.

Related Work

Originally, discourse markers have been considered as linguistic elements (conjunctions, interjections, adverbs, and lexicalized phrases) which help to build discourse structure, organize textual information, and construct conversation [10]. Pragmatically, discourse markers operate at the discourse level to signal discourse-related interaction, i.e., by initiating discourse, marking a topic shift or used as fillers [10, 11]. Also, they present discourse linking functions by marking coherence relations and indicating the structural organization of the discourse [12]. Discourse markers serve different communicative purposes and their meaning is context-dependent [13]. Furthermore, it has been reported that the term discourse marker is attributed to items that fulfill discourse marking functions, and besides linguistic expressions, such items may be speech formulas and non-lexicalized metalinguistic devices [14], and that word pronunciations, vocalizations such as filled pauses but also gestures provide markers that structure discourse and may fulfill conversational management functions [15]. In this paper, we explore the hypothesis that conversational laughter may serve as an informative signal regarding dialog topic segmentation.

Conversational laughter has been considered as a cooperative mechanism which can provide clues to dialog structure, i.e., it has quantifiable discourse functions, and, together with its social signaling capacity, it may be used in the signaling of topic changes and in enhancing topic boundary detection [16–19]. It has also been reported that in task-based interactions, laughter is significantly more frequent than expected when the subjects do not address the task at hand, i.e., when the interaction is more socially oriented than task-oriented [20]. Laughter has also been viewed as an event anaphor associated with two dimensions regarding the relation of the laughter to the laughable, i.e., the enjoyment of an event and the recognition of an incongruous event [21].

Several categorizations of laughter types have been reported in the literature, based on the physical properties of the laughter, its expressiveness and its functionalities. Four types of laughter were analyzed in [22], including the emotional categories of joyous, taunting and Schadenfreude laughter, and the physical type of tickling laughter. Poyatos [23] provides an emotion-based categorization, distinguishing among laughter types of affiliation, aggression, social anxiety, fear, joy, comicality and ludicrousness, amusement and social interaction, and self-directedness. Tanaka and Campbell [8] report on high accuracy in automatically classifying laughter in two types, polite and

mirthful, based on acoustic parameters; the selection of those types is grounded on their finding that, following a manual perceptual labeling of laughter, polite and mirthful were the two predominant types defined in the dataset.

Laughter annotation in the ILHAIRE Project [24] was performed with an annotation scheme that initially included two laughter types, hilarious and conversational, and was later modified to include the values of Hilarious, Social, Awkward, Fake and Not a Laugh [25]. From a social signal perspective, laughter may be additionally viewed with respect to social functions and to the intentionality of the speaker with positive or negative effect, such as cases of laughter elicitation from the audience through parody, or by discrediting or ridiculing another person or event [26, 27]. While identifying shortcomings of existing laughter classification schemes, [5] build upon the event anaphor account of [21] and introduce a multi-layered analysis of laughter based on different parameters, such as formal and contextual aspects, semantic meaning and functions. The need of multi-level studies of laughter has been also identified by [28], who provide an annotation scheme based on the physical features of laughter, and acknowledge the multiple dependencies among respiration, facial action, acoustics, and body movements in the expression of laughter.

Duration aspects of laughter have been addressed from the acoustics perspective in the literature, and laughter duration has been considered among features during laughter synthesis processes [29]. Duration differences have been also found for laughs expressed by males compared to those produced by females, with male laughter bouts being a bit longer than the female ones [30]. In their work about overlapping laughter, [31] found that the initiating laugh (the first one) is longer than the responding laugh (the second one). In examining the conversational phases of casual talk, [32] found that laughter is very frequent in phases of short contributions (chats) compared to the longer, monologic phases (chunks); however, there was no distinction in laughter types nor speaker roles, because of the nature of casual conversations, i.e., where there are no predefined roles.

Studies about the impact of gender on patterns of laughter have shown that female speakers laugh more than their audience [33] and have significantly higher spontaneous laughter than males, and slightly higher daily mean frequency of laughter than males, although the latter difference is not significant [34]. Bachorowski and Owren [35] report that in mixed-sex encounters females laugh more often than males, something that is not observed in same-sex encounters. Also, laughter duration aspects have been examined with respect to gender, i.e., longer laughter duration in males is related to dominance, while in females is linked to rewarding behavior [36]; female long laughter was also perceived as more spontaneous, an observation also supported by [37].

Investigation of laughter is related to natural human-computer interaction (HCI) applications, particularly in the development of intelligent user interfaces with social communication skills, i.e., the ability to represent and understand complex human social behavior in face-to-face communication, including laughter and smiling. Several studies have investigated laughter and smiling to help create more natural and efficient agents, in terms of understanding and predicting the type of laughter to use but also in synthesizing laughter. Becker-Asano and Ishiguro [38] have focused on the social effect of laughter produced by a robot, while [29] present an interactive system enabled to detect human laughs and respond appropriately, by integrating human behavior and context information. El Haddad et al. [39] report on a process for synthesizing laughter and smiling sequences, and a system predicting smile and laughter sequences in a dialog participant based on observations of the other participant's behavior.

As far as data collections are concerned, there are datasets designed exclusively for laughter investigation, such as the DUEL multilingual and multimodal corpus [40] and some have been used to train and predict laughter models using audiovisual information, including the AV-LASYN Database [41], the MMLI database [42] and the MAHNOB database [43]. Most of the databases of this kind are very helpful in modeling laughter and are usually limited to certain types of laughter and context.

In addition, discourse and conversational aspects of laughter have been studied in datasets of social or task-based interactions that were not created solely for that purpose (laughter investigation), still, they provide an important source for observations related to laughter as part of the behavior of the involved participants. Examples are the AMI [44] and TableTalk [45] corpora, which allow for comparisons of the laughter dimension in human interactions in different contexts (cf. [16]).

The contribution of this work is summarized in the following:

- verification of the discourse role of laughter in a new dataset and a new conversation setup, i.e., three-party, task-based interaction including two different roles, the players performing the task and a facilitator moderating the discussion;
- investigation of discourse aspects of laughter exploring annotations that distinguish among mirthful and discourse laughter;
- extending the analysis with respect to variables of conversational role and gender of the laughter owners, and the facilitators' feedback.

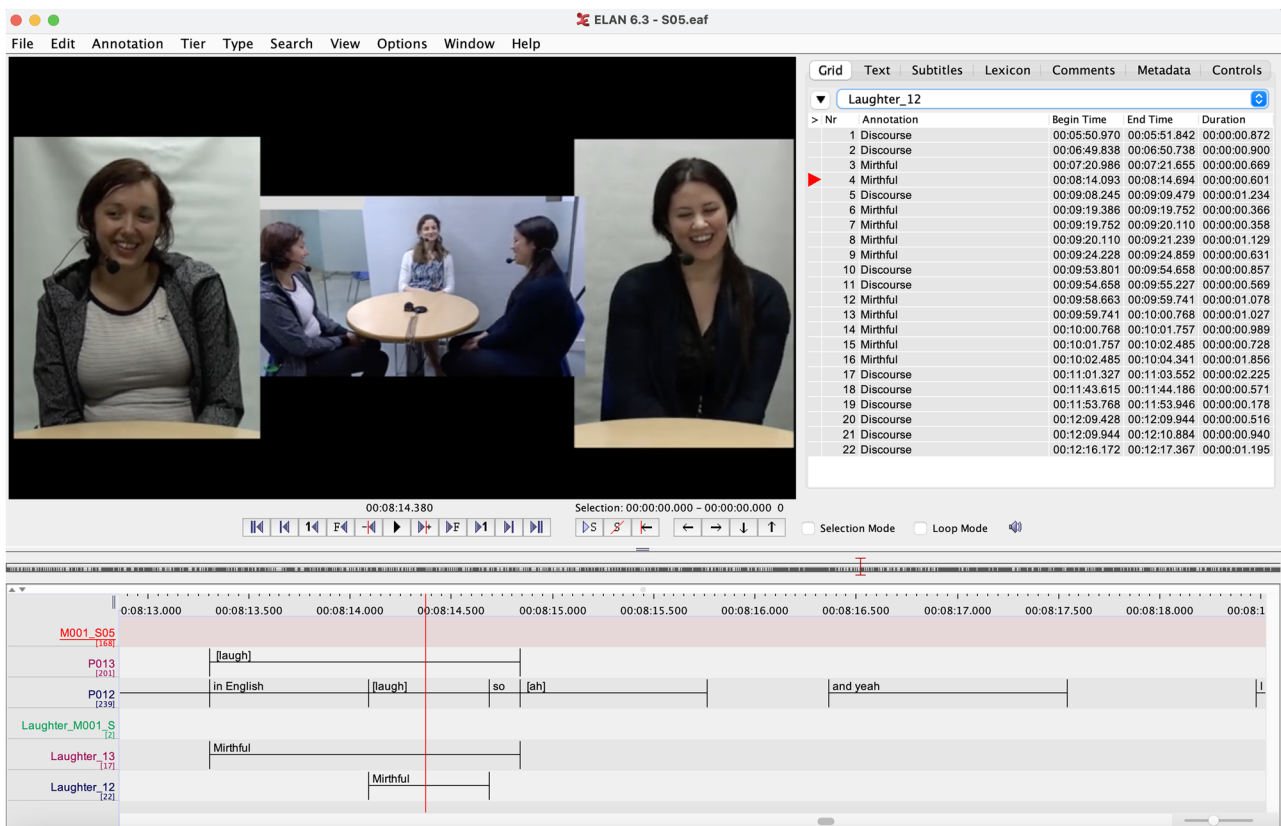


Fig. 1 Screenshot of a sample file in ELAN. There is one tier per speaker that includes the speaker turns, speech and laughter transcription (*M001_S05*, *P012* and *P013*). Tiers *Laughter_12* and *Laughter_13* include laughter instances of the 2 speakers that are laughing at that moment

Materials and Methods

Dataset

For this work we have used 18 dialogs from the MULTI-SIMO corpus of collaborative group interactions [7], where two players work together to provide answers to a quiz and are guided by a facilitator, who monitors their progress and provides feedback or hints when needed.⁵ The scenario was designed in a way that would encourage the participants' collaboration towards a goal. The dialog sessions were carried out in English and the task of the players was to discuss with each other, provide the three most popular answers to each of three questions posed (based on survey questions posed to a sample of 100 people),⁶ and rank their answers from the

most to the least popular. Basic knowledge questions were selected so that they would be easy to address and would trigger the discussion among the players.⁷

Players expressed and exchanged their personal opinions when discussing the answers, and they announced the facilitator the ranking once they reached a mutual decision. Task success corresponds to proposing a shared ranking of answers that correspond to the ranking aggregated from the 100 people surveyed independently. The task was designed to require participants to discuss with each other both candidate answers and estimates of the rankings provided by the independent group.

A dialog session has an average duration of 10 min. All sessions were recorded with high-definition cameras and head-mounted microphones, enabling high-quality audio capturing at distinct channels. The recordings took place in a dedicated room, and in each group the three participants were sitting around a table at an equal distance to each other (cf. Fig. 1). The pairing of players, who are in their majority Trinity College students and researchers (mean age = 30 years old), was randomly scheduled and was based on their availability to participate in the recordings. Thus, pairing

⁵ All subjects gave their informed consent to participate in the dialog recordings. The recordings and the study were conducted in accordance with the EU General Data Protection Regulation 2016/679 (GDPR), and the protocol was approved by the Trinity College Dublin, School of Computer Science and Statistics Research Ethics Committee. With complete compliance with the terms of consent provided by the participants, these 18 dialogs are represented in the version of the corpus available under a non-commercial license at <http://multisimo.eu/datasets.html>.

⁶ <http://familyfeudfriends.arjdesigns.com/>, last accessed 15.03.2022

⁷ The questions posed were: (1) name a public place where it's likely to catch a cold or a flu bug; (2) name 3 instruments you can find in a symphony orchestra; and (3) name something that people cut.

of participants was not designed to include experimental conditions based on familiarity and gender balance among players, even though in the resulting data there are three groups where players are familiar with each other. Also, there are eight groups where players are both female, seven where players are both male, and three mixed groups. One-third of the participants are native English speakers, while the rest of them span fifteen nationalities and mother tongues other than English. Three participants shared the role of the facilitator in the 18 dialogs. All three facilitators are female, non-native English speakers (of Greek nationality) and have the same professional background and competence in English (i.e., English teachers). The facilitators were briefed before the recordings about the quiz questions and answers and they were instructed to monitor the discussion and provide hints to the players when necessary. In this study we exploit laughter annotations and conversational features of all participants in the 18 corpus dialogs.

The audio signal of all corpus sessions was manually segmented in speaking turns and transcribed by two annotators using Transcriber.⁸ A rich transcription approach was followed so that the transcripts were as close as possible to the verbal content and the way it was uttered. In this respect, transcription includes the speakers' identification, the words they utter, as well as a set of labels marking speech disfluencies, overlapping talk, silence and laughter, all fully synchronized with the audio signal. Thus, laughter, as a voiced expression, was identified and marked in the transcripts every time it occurred. Transcripts were then imported into the ELAN annotation editor,⁹ so that all the information coming from the transcript was visible and further editable.

Topic Segmentation

In topic segmentation tasks, topics have been defined as lexically coherent segments of discourse or as a cohesive sequence of conversation turns about a particular subject [46, 47]. While topics and topic boundary identification have been shown to be reliably coded by human raters [48], automatic approaches for this task are reported to present challenges related to the nature and processing of the speech signal, such as speaker diarisation, speech recognition, separation of the speech signal from noise, etc. [49]. To identify topic boundaries, automatic approaches mostly use speech transcriptions [50–52] or vocalization events, such as pauses and speech overlaps [49].

In the MULTISIMO corpus each dialog was manually segmented into structural parts, as dialogs follow a specific structure, i.e., introduction, question 1, question 2, question 3 and closing. The data design defines the dialog topics and their order: by the nature of the game, each of the three quiz

questions is further structured into two parts, i.e., the identification of answers and the ranking of answers. Our focus of investigation are the three questions discussed in the dialogs. Thus, we define as *topic* each of the two parts of each of the three quiz questions; the entire dialog is therefore split into six topics. Following the methodology described in [53] and [16], we measure laughter events occurring in each dialog topic, by splitting each topic in the following three sections:

- section *wo1* represents the start of the topic (*wo*: segments without the topic core).
- section *wo2* represents the end of the topic and the transition to the next one.
- section *wi* is the middle part which represents the core of the topic (*wi*: segments having the topic core within).

The splitting is performed upon a temporally proportional method, i.e., the sum of the duration of the start and end sections (*wo1+wo2*) equals the duration of the core section (*wi*), for however long a topic is. The difference with respect to the methodology followed in [16] is that they use a single value for both start and end of a topic, indicating a transition between two topics, and that topic section types are defined in relation to the duration of a sequential pair of topics.

As regards the association of topics to the dialog content during the answer identification phase, in the *wo1* section the facilitator poses the question, requests acknowledgment from the players and explains, if necessary (e.g., *I think the third question should be very simple, so I would like you to name things that people cut.*). The *wi* section is the core of the topic where the players work with each other to find the possible answers to the questions and they check their responses with the facilitator (e.g., one of the utterances included in a given *wi* section would be: *Would a swimming pool be a weird answer or would it make sense?*). At the final, *wo2* section, the facilitator resumes the answers and briefly requests acknowledgment from players (e.g., *Great job, well done, you now have the three, ok?*).

During the answer ranking phase, in *wo1* section the facilitator asks players to proceed to the ranking of the answers, and provides further explanations, if necessary (e.g., *Now that you have the three most popular answers would you be able to discuss the ranking then?*). Players during the *wi* section collaborate and discuss among each other about the popularity of the responses and their possible ranking, they come to a decision and let the facilitator know about it (e.g., *So it might make sense to put that number one?*). In the *wo2* section, the facilitators provide their feedback to the ranking and the players react accordingly (e.g., *You were very fast when you had to find the three answers but I'm afraid you again don't have the right ranking.*).

⁸ <http://trans.sourceforge.net/>, last accessed 15.03.2022

⁹ <https://tla.mpi.nl/tools/tla-tools/elan/> last accessed 15.03.2022

Table 1 Laughter events examples from the corpus (from participants who also have the floor at that particular moments), their original annotation and their updated annotation for cases where the 2 raters disagree

	Example	Original Value		Final Value
		Rater1	Rater2	
1	You're getting there that's very close [laughter]	Discourse	Discourse	Discourse
2	I'm so sorry I don't have any more questions [laughter]	Discourse	Discourse	Discourse
3	She was just saying what they were [laughter]	Discourse	Discourse	Discourse
4	but I'm a vegetarian I'd never have thought of that [laughter]	Mirthful	Mirthful	Mirthful
5	know it's just a way of thinking of it [laughter]	Mirthful	Mirthful	Mirthful
6	oh well that's not gonna happen [laughter]	Mirthful	Mirthful	Mirthful
7	Well we're getting there [laughter]	Discourse	Mirthful	Ambiguous
8	ok, they have the same meaning [laughter] makes sense	Mirthful	Discourse	Ambiguous
9	but I don't think people generally get sick in restaurants [laughter]	Discourse	Mirthful	Ambiguous
10	You don't do a lot of cooking, do you? [laughter]	Mirthful	Discourse	Ambiguous
11	It's so hard [laughter] took us long enough	Discourse	Mirthful	Ambiguous

Participant Role and Gender

The dialogs involve two roles: facilitator and players. All three of the facilitators are female. The distribution of genders across players was described in "Dataset": the dialogs do not balance gender, given that three groups were mixed, eight involved females only and seven, males only. In analyzing gender, we consider the quantities associated with laughter (as described below) as observed among females and males, independently of dialog role. Thus, quantities associated with facilitators are among the quantities assessed for females. If effects associated with role are identical to those associated with gender, it may be concluded that there is insufficient information to separate effects of role from effects of gender.

Laughter Classification

Annotation related to laughter classification was carried out in ELAN employing the audio signal and the temporally aligned speech transcription and videos of the front view of the participants. As mentioned in "Dataset", there are transcripts for each speaker, and within those there exist annotations of audible laughter events that had been identified and marked on the time axis as such. Annotators were asked to label those laughter events that were already spotted from the transcription phase (but also any newly observed events that were missing from transcripts) with one of the 2 available values: either *Discourse* or *Mirthful*. Figure 1 presents an example of laughter annotation in a corpus dialog.

As outlined above, we use the *Discourse* label for laughter expressed as a politeness marker, an acknowledgment marker or a discourse filler by the conversation participants to maintain the conversation. Related cases in the dataset are, for example, greetings, acknowledgment

responses of players to facilitators, politeness responses among players or from the facilitator to the players, as well as fillers in cases where the speakers are not sure of what to answer. The *Mirthful* label is synonymous to enjoyment and hearty laughter, i.e., assigned to cases where laughter appeared to be a consequence of genuine amusement. Annotators were not informed of our hypotheses about the distribution of these laughter types within topics and near topic boundaries.

The annotation was performed by two raters (1 female, 1 male, both non-native English speakers and postgraduate TCD students), who applied either of the discourse and mirthful values to the identified laughter events. Annotations were based on the perception of laughter type as recorded in the audio and video signals. Thus, annotators were able to have access to the whole dialog, including the immediate context of utterances that included laughter. The inter-rater agreement was fair, with a percentage agreement of 65.7%, and a Cohen's Kappa coefficient of $\kappa = 0.28$.

Laughter events were then extracted from the annotations, including all related information, i.e., their timestamps and duration, the labels assigned by the two annotators and their mapping to topic sections. We further processed the annotated values and came up with a single value for each laughter segment; the *Discourse* value was set for cases that both raters considered as discourse laughter; the *Mirthful* value was decided for events that both annotators rated as mirthful; and a third value of *Ambiguous* was introduced for those cases where the two raters disagreed (cf. a set of examples on Table 1). We argue that the inter-rater disagreement is a source of information per se, and is directly related to the research in question, therefore we preserve the annotation of disagreement by introducing an additional (*Ambiguous*) value attributed to laughter annotation (see also [54] reporting on the value of studying correlations with variables that

Table 2 Distribution of laughter in the dataset. All durations are in seconds

	All	Discourse	Mirthful	Ambiguous	wo1	wi	wo2
Counts	569	245	129	195	92	294	183
Total Duration	707	254	200	253	115	375	217
Mean Duration (<i>SD</i>)	1.24 (0.82)	1.04 (0.55)	1.55 (1.03)	1.30 (0.88)	1.24 (0.73)	1.28 (0.93)	1.19 (0.68)

encode inter-annotator disagreement along with the annotators' ratings).

This process resulted in a set of 569 laughter events (cf. Table 2), each accompanied by one of the three values, which were analyzed with respect to their distribution and association with topic sections (cf. the results reported in "Results").

Feedback Annotation

The facilitator's contributions that take place during the discussion of the three questions in each dialog were additionally annotated with values that correspond to the type of the feedback they provide to the participants' responses, i.e., *positive*, *negative* and *neutral* feedback. All types of feedback are meant to help participants identify the correct answers and their ordering; the distinction in types serves to understand how feedback is lexicalized. Neutral feedback is about facilitator's interventions with hints and examples (*It is related to food, but think of a different category of food.*); positive feedback is given when the participants are doing well (*Great job, well done!*); negative feedback is given to responses that are not correct (*Unfortunately you didn't get this one.*).

Annotations related to the facilitator's feedback were originally developed for a study that detected the alignment between players by observing linguistic repetitions in the dialog transcripts and investigating the relation of the alignment to the type of the facilitator's feedback [55]. In that study, the facilitator's feedback was coded by one rater, and

annotations were edited for validity and consistency by a second rater, resulting in a set of 2576 feedback annotations.

In the current work, we explored the feedback annotations that are temporally aligned with laughter (i.e., 246 items), given that feedback from a facilitator may lead to both discourse and mirthful laughter responses; and that a feedback response may signify a topic transition, such as in cases where players elicit a reaction from the facilitator to move on to the next question. Feedback information consists of indicating the presence or absence of a feedback expression in relation to each of the laughter events identified, and of indicating the type of the feedback expressions. The analysis of feedback consists in identifying the distribution of feedback and feedback types in relation to laughter events, and the interaction of feedback with laughter type, role, gender and topic section types (cf. "Laughter and facilitator's feedback"). The hypotheses we want to test are that (a) laughter co-occurring with feedback is more of the discourse type than the mirthful one; (b) that this laughter is mainly located at topic termination points; and (c) that when feedback co-occurs with laughter, then feedback is of a positive type (rather than negative or neutral).

Results

Distribution of Laughter Across the Corpus

This section presents the observations about the distribution of laughter in the corpus, independently of the topic

Table 3 Distribution of laughter types per speaker role and gender (counts and duration; all durations are in seconds)

	Gender		Speaker Role	
	Female	Male	Facilitator	Player
All Counts	422	147	178	391
Mean Duration (<i>SD</i>)	1.21 (0.84)	1.33 (0.75)	1.09 (0.76)	1.31 (0.84)
Discourse Counts	186	59	76	169
Discourse Mean Duration (<i>SD</i>)	1.00 (0.53)	1.15 (0.61)	0.86 (0.41)	1.12 (0.59)
Mirthful Counts	84	45	40	89
Mirthful Mean Duration (<i>SD</i>)	1.47 (1.08)	1.70 (0.91)	1.30 (0.85)	1.67 (1.08)
Ambiguous Counts	152	43	62	133
Ambiguous Mean Duration (<i>SD</i>)	1.33 (0.94)	1.20 (0.60)	1.23 (0.94)	1.33 (1.08)

Fig. 2 Laughter counts in topic sections

sections, including distributions of laughter types per conversational role and per speaker gender.

There are 569 laughter events annotated in the corpus (cf. Table 2); the majority of those (245, 43%) are assigned with the discourse label, 129 (23%) with mirthful, and the remaining 195 (34%) with the ambiguous one. There are two speaker roles in the dataset, the facilitator and the participant (player). The laughter distribution per speaker role and per gender is shown in Table 3. Speaker role is not significantly associated with laughter type distribution ($\chi^2(2) = 0.04$, $p = 0.98$). On the contrary, there seems to be an effect of gender, with females having significantly more laughter quantities than males, in both discourse and mirthful laughter ($\chi^2(2) = 7.36$, $p < 0.03$).¹⁰

As regards the temporal length of laughter, the mean durations of laughter occurring per laughter type, role and gender are shown on Tables 2 and 3. A Kruskal-Wallis test revealed that discourse laughter is significantly shorter than mirthful ($H(2) = 28.50$, $p = 6.462e - 07$). Also, Wilcoxon tests showed that players' laughter is longer than facilitators' ($W = 28062$, $p < 0.001$), while females laugh for a shorter time than males ($W = 26924$, $p < 0.02$).

Laughter and Topic Sections

This section gives an overview of the observations related to the distribution of laughter and laughter types across the three topic section types of the dataset, and in relation to the conversational role and the gender of the speakers.

As mentioned in "Topic Segmentation", each of the 18 dialogs is segmented into 6 topics and each of the 6 topics is further split into 3 section types; *wo1* represents the starting section, and *wo2* the ending section, while the *wi* section is the middle section part which represents the core of the topic. In terms of content, section *wi* is usually the section where the players collaborate with each other to answer the question and rank their answers. Section *wo1* refers to dialog parts where the requests are posed and discussed, and section *wo2* is mostly related to the evaluation of responses and the feedback that the two players receive from the facilitator.

Out of the total 569 laughter occurrences, the majority of them is included in *wi* sections (52%), followed by section *wo2* (32%), while Section *wo1* presents the fewest counts (16%) (cf. Fig. 2).

Table 4 Mean (SD) laughter durations (in sec.) in topic sections

	Total	Laughter Type			Speaker Role		Gender	
		Discourse	Mirthful	Ambiguous	Facilitator	Player	Female	Male
wo1	1.24 (0.73)	1.08 (0.66)	1.66 (0.83)	1.08 (0.58)	1.04 (0.53)	1.32 (0.78)	1.15 (0.67)	1.56 (0.83)
wi	1.28 (0.93)	0.98 (0.53)	1.62 (1.17)	1.41 (1.01)	1.11 (0.88)	1.35 (0.94)	1.23 (0.95)	1.45 (0.83)
wo2	1.19 (0.68)	1.11 (0.54)	1.36 (0.88)	1.20 (0.70)	1.07 (0.63)	1.24 (0.69)	1.22 (0.73)	1.11 (0.54)

¹⁰ Data analysis was conducted in R. On applying Shapiro-Wilk normality test, variables involved in all tests are found to be non-normally distributed, therefore non-parametric tests were employed in the analysis (Pearson's chi-squared test (χ^2), Wilcoxon rank sum test with continuity correction (W), Kruskal-Wallis rank sum test (H)).

Table 5 Mean laughter-per-second rates (*SD*), per section type, role and gender

	Total	Speaker Role		Gender	
		Facilitator	Player	Female	Male
wo1	0.74 (0.70)	0.77 (0.83)	0.73 (0.63)	0.83 (0.76)	0.60 (0.57)
wi	0.82 (0.52)	0.93 (0.35)	0.77 (0.58)	0.86 (0.33)	0.75 (0.75)
wo2	0.88 (0.39)	1.01 (0.39)	0.81 (0.38)	0.89 (0.37)	0.85 (0.45)

By observing the distribution of laughter counts in topic sections, we found no interaction of counts with laughter type ($\chi^2(4) = 6.05, p = 0.2$), nor with speaker role ($\chi^2(2) = 0.49, p = 0.8$) or gender ($\chi^2(2) = 3.60, p = 0.2$).

As regards laughter mean duration, we note that there are slight differences among the three topic sections (cf. Table 4); however, those differences are not significant (Kruskal-Wallis $H(2) = 0.199, p = 0.9$).

Laughter durations are of varying length across topic segment types, with laughter in *wi* sections being the longest, followed by *wo1* sections, while laughter in *wo2* sections are the shortest (cf. column *Total* on Table 4), although for any individual topic the duration of *wo1* equals that of *wo2*, and the duration of *wo1* added to the duration of *wo2* is equal to the duration of *wi* (cf. "Topic Segmentation" for topic segmentation details). A direct comparison of laughter counts within topic sections would probably yield to misleading results, as, for example, laughter counts might be high for one section because they are measured over a longer period, and not because people are laughing more often. Therefore, to provide more accurate comparisons, we performed a set of tests involving the normalization of counts, i.e., by calculating the rate of *number of laughter counts per second* for each section type; and subsequently, for each of the laughter types along the three section types, and for each of the laughter types in each of the three section types. Tables 6 and 7 include variables that were constructed to represent those quantities (e.g., *wo1LpS* represents the mean laughter-per-second rate in *wo1* section), and to be tested with other variables.

Table 5 lists the laughter-per-second rates, per section type, role and gender. The first column shows that there are subtle differences of this quantity among topic

Table 6 Mean laughter-per-second rates (*SD*), per type, in topic sections (*w*: all topic sections)

	Discourse	Mirthful	Ambiguous
w	1.03 (0.55)	0.65 (0.69)	0.76 (0.40)
wo1	0.61 (1.18)	0.24 (0.43)	0.35 (0.56)
wi	0.93 (0.70)	0.41 (0.49)	0.63 (0.53)
wo2	0.89 (1.14)	0.44 (0.76)	0.69 (0.89)

Table 7 Mean laughter-per-second rates (*SD*), per type, role and gender, in topic sections (*w*: all topic sections)

	Gender		Speaker Role	
	Female	Male	Facilitator	Player
w-Discourse	1.15 (0.51)	0.84 (0.57)	1.30 (0.60)	0.90 (0.48)
w-Mirthful	0.69 (0.83)	0.59 (0.35)	0.67 (0.49)	0.64 (0.78)
w-Ambiguous	0.85 (0.34)	0.62 (0.44)	0.94 (0.39)	0.68 (0.37)
wo1-Discourse	0.80 (1.37)	0.29 (0.65)	0.50 (0.82)	0.66 (1.32)
wo1-Mirthful	0.17 (0.42)	0.35 (0.43)	0.12 (0.36)	0.29 (0.46)
wo1-Ambiguous	0.48 (0.62)	0.14 (0.34)	0.36 (0.58)	0.34 (0.55)
wi-Discourse	1.07 (0.53)	0.70 (0.89)	1.12 (0.59)	0.84 (0.74)
wi-Mirthful	0.42 (0.50)	0.39 (0.50)	0.50 (0.53)	0.36 (0.47)
wi-Ambiguous	0.76 (0.53)	0.41 (0.45)	0.79 (0.68)	0.55 (0.42)
wo2-Discourse	1.04 (1.36)	0.63 (0.51)	1.33 (1.79)	0.67 (0.51)
wo2-Mirthful	0.45 (0.86)	0.41 (0.58)	0.35 (0.49)	0.48 (0.87)
wo2-Ambiguous	0.70 (0.92)	0.68 (0.86)	0.82 (1.14)	0.63 (0.74)

sections; however, those differences are not significant ($W = 1296, p > 0.3, W = 1170, p > 0.07, W = 1222.5, p > 0.1$, for each of the comparisons among topic sections). We do note, however, that facilitators' laughter rate is significantly higher than the rate of players in *wi* sections ($W = 462, p = 0.01$); and that female laughter rate is significantly higher than male rate in *wi* sections ($W = 470, p = 0.02$).

A comparison of laughter type rates across the 3 sections shows that the rate of discourse laughter per second is significantly higher than the mirthful laughter rate ($W = 730.5, p = 7.615e - 06$) and the ambiguous laughter rate ($W = 1001, p = 0.005$) (cf. 1st column on Table 6). While there are no significant effects of laughter rate across sections, and gender, there are significant interactions with speaker role: mean discourse laughter rates are higher for facilitators than for players ($W = 467, p < 0.009$), and so are ambiguous laughter rates ($W = 431.5, p < 0.05$) (cf. 3 first rows on Table 7).

When we observe laughter rates in more detail, i.e., distribution of each of laughter type rates per each of the sections (cf. Table 6), we note that the rate of discourse laughter at topic termination points (*wo2*) is significantly higher than at topic beginnings (*wo1*, $W = 1037.5, p = 0.007$); a similar behavior is observed with ambiguous laughter rates ($W = 1110, p < 0.02$). Also, discourse laughter rate is significantly higher than mirthful at both topic termination points (*wo2*: $W = 916.5, p < 0.0006$) and at topic core points (*wi*: $W = 825, p = 6.857e - 05$). The rates of all laughter types are significantly higher in *wi* than in *wo1* ($W = 1998, p < 0.0006$ for discourse laughter; $W = 1770, p = 0.03$ for mirthful laughter; $W = 1943, p < 0.002$ for ambiguous laughter).

Table 8 Distribution of feedback and non-feedback laughter

	Feedback	Non-feedback
All Counts	246	323
Mean Duration (<i>SD</i>)	1.30 (0.83)	1.20 (0.81)
Female Counts	179	243
Male Counts	67	80
Facilitator Counts	59	119
Player Counts	187	204
wo1 Counts	36	56
wi Counts	101	193
wo2 Counts	109	74
Discourse Counts	127	118
Mirthful Counts	40	89
Ambiguous Counts	79	116
Positive Feedback Counts	127	-
Negative Feedback Counts	51	-
Neutral Feedback Counts	68	-

We also examined the rates of laughter types in the distinct topic sections conditioned by the speaker gender and the speaking role (cf. Table 7). We note that the rates of both discourse and ambiguous female laughter are higher than males in section *wi* ($W = 479.5, p = 0.01$, $W = 470, p < 0.02$ respectively), as well as in section *wol* ($W = 443.5, p < 0.04$, $W = 441.5, p = 0.03$ respectively). However, male mirthful laughter rate is observed to be higher than female in section *wol* ($W = 238, p < 0.03$). Finally, there are no significant effects of laughter type rates associated speaker role among the topic sections.

Laughter and Facilitator's Feedback

This section presents the distribution of laughter events that co-occur with dialog sections where feedback was provided to the players by the facilitator. First, we examine the quantity of laughter that is related to feedback, as opposed to feedback-unrelated laughter, as well as the topic sections where feedback is mostly present (cf. Table 8). Second, we focus on the feedback-related laughter and we look at the

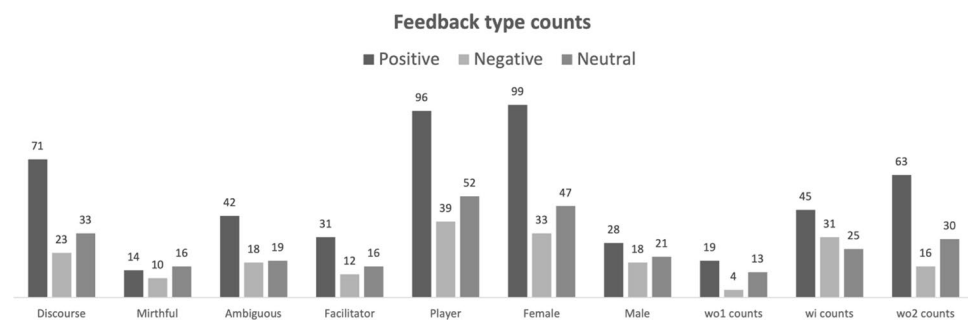
distribution of the three feedback types (i.e., positive, negative, neutral) in relation to laughter types and topic sections (cf. Fig. 3).

We formulate the hypothesis that since feedback is largely related to acknowledgements or politeness markers, it is very likely that discourse laughter value is more frequent than mirthful at feedback moments. Also, by the way the topic segmentation is designed, it is expected that feedback is mainly present at topic termination points (*wo2*), since this is the topic section where the evaluation of the players' responses takes place, involving laughter reactions from both speaker roles. Finally, because laughter (especially in the context of collaborative dialogs) happens mostly as an affirmative, amusing or encouraging reaction, we also expect that it will be mainly associated with positive feedback.

The distribution of feedback-related and feedback-unrelated laughter is shown on Table 8. There are more discourse laughter counts associated with feedback (52%) than mirthful (16%) or ambiguous ones (32%) also, there is more discourse laughter related to feedback than feedback-unrelated laughter ($\chi^2(2) = 15.833, p < 0.001$). Furthermore, there are more feedback-related laughter instances in players than in facilitators ($\chi^2(1) = 10.151, p < 0.002$), while there is not any significant interaction with gender ($\chi^2(1) = 0.32, p > 0.5$). When examining feedback in relation to topic sections, we note that laughter at feedback moments occurs more at topic termination points than topic beginnings or core topic parts. Furthermore, at topic termination points there is significantly more laughter co-occurring with feedback than feedback-unrelated laughter ($\chi^2(2) = 29.96, p = 3.122e - 07$). In terms of duration, feedback-related laughter is longer than feedback-unrelated laughter ($W = 35906, p < 0.05$).

As regards feedback types related to laughter, we observe that there is significantly more positive feedback than negative or neutral in all three topic sections, with *wo2* having the majority of positive feedback counts ($\chi^2(4) = 11.40, p = 0.022$) (cf. Fig.3). We also note no significant interaction of feedback types with either of the laughter type, speaker role, speaker gender, or laughter duration quantities.

Fig. 3 Profiling of feedback types in relation to laughter types, role, gender and topic sections



Discussion

Because of the nature of the empirical evidence that may be brought to analyze the theoretical perspectives, it is necessary to analyze appropriate data sets as they emerge. We used the MUTLTISIMO dataset, which, because of its design, it supports a structural notion of topic. Dialog topics were split using a temporally proportional method to generate sections representing the start, core and end of a topic. Our goal was to study the timing of laughter with respect to that topic individuation, as we maintain that topic transition and topic continuation signals are among the discourse functions of discourse laughter. Furthermore, our perspective is that the binary distinction in discourse and mirthful laughter types is viable, and our annotation results do not give us reason to think a more articulated taxonomy would yield less disagreement.

Our findings extend the main results reported in [16], who show that laughter is more likely to occur at topic terminations, rather than topic onsets, implying a discourse function of laughter which may serve as a feature to enhance detection of topic boundaries. Our observations are also in line with the findings in [8] that, in social communication, people typically express polite social laughter, while mirthful laughter is less frequent. This work adds to prior study of the timing of laughter with respect to topic boundaries (e.g., [16]) a demonstration of differences in these distributions according to the perceived type of laughter (discourse vs mirthful vs ambiguous). The earlier results, as they pertain to discourse laughter, are demonstrated to apply to the additional data provided by the MULTISIMO corpus. We therefore believe that our observations are consistent with findings of similar research work. In order to make a claim that this approach would yield the same results with other data, we would first need to test it. However, our findings are anchored in a dataset that is well-formed, and this is a reason to expect that testing those effects in other datasets would result in the same observations.

Prior quantitative work on discourse functions of laughter has not taken the distinction into account. We tested this theory further by quantifying the discourse functionality, this time by employing distinctive laughter types marked as discourse and mirthful, and by examining laughter subject to control variables such as role, gender and feedback content, and highlighted the evidence of discourse laughter exhibiting a non-random structure, acting as a discourse marker.

We analyzed annotated laughter data from a corpus of three-party dialogs and addressed the question of whether discourse laughter distributes distinctly from mirthful laughter. The results show that the mirthful laughter exhibits a more arbitrary distribution than its counterpart, while

discourse laughter is more predictable, more systematic, and highly associated with the dialog topic termination points. We see this as evidence that the instances of laughter given the label “discourse” in this dataset, have a genuine discourse function in the interpersonal management of the flow between topics in the dialogs.

Discourse laughter, when studied independently of other variables, such as topic sections or gender and speaker role, is attributed to the majority of laughter occurrences, is shorter than mirthful laughter (cf. Table 2), and has a higher rate of laughter per second (cf. Table 6). When examined within topic section types, the rate of discourse laughter is higher than its mirthful or ambiguous counterparts (cf. Table 6), and also higher for facilitators (cf. Table 7).

Mirthful laughter patterns with significant associations are rare: when measured in the whole corpus, independently of topic sections, mirthful laughter is longer in mean duration compared to discourse (cf. Table 2). Mirthful laughter rate is higher in topic middle sections than at topic beginnings, but this kind of distribution is observed also for the discourse type (cf. Table 6). Also, males tend to have higher mirthful laughter rates than females at topic beginnings (cf. Table 7).

Ambiguous laughter has been shown to have a similar behavior to discourse laughter in several cases, supporting our initial hypothesis that it is closer to discourse than to mirthful laughter. For example, ambiguous laughter is in general shorter than mirthful, but its rate is higher than the mirthful laughter rate in all three topic sections (cf. Table 6). Similarly to discourse laughter, the rate of ambiguous laughter at topic termination points and topic middle points is higher than at topic beginnings (cf. Table 7).

Our hypotheses related to speaker role were partly confirmed. Our expectation that there is an interaction between role and laughter type, i.e., that facilitators produce more discourse laughter than mirthful laughter is supported by the fact that mean discourse laughter rates are higher for facilitators than for players when examined across all section types (cf. Table 7). However, there is no significant evidence related to the hypothesis that facilitators are more prone to discourse laughing at topic termination points. Instead, we note that facilitators' laughter rate is significantly higher than players' at topic core sections (cf. Tables 5, 7). Finally, as an independent quantity, players' laughter is longer than facilitators' (cf. Table 3).

As regards laughter distribution with respect to gender, there was no particular hypothesis formulated; however, some significant differences emerge: females have in general more laughter instances, but shorter laughter than males, in both discourse and mirthful laughter (cf. Table 3). When examining laughter rates and gender, we note that female discourse and ambiguous laughter rates are higher

than males at topic beginnings and topic middle parts, while, on the contrary, males show higher mirthful rates at topic beginnings (cf. Table 7).

We also looked at cases where laughter co-occurs with moments where facilitators express their feedback about the performance of the players. We noted that laughter during feedback is longer than other moments and is mostly associated with players (cf. Table 8), presumably as a reaction to the facilitators' feedback responses. Most importantly, the results show that the discourse value is highly associated with laughter during feedback and that most of the feedback laughter, following our expectations, occurs at topic termination points (cf. Table 8). The players provide acknowledgment responses to the facilitator through laughter, also possibly indicating that since the validity of their responses has been checked, they can conclude a round and move on to the next question.

We have followed observational methods [56] in which we explore the corpus dialogs by making operational the theoretical construct of the discourse function of laughter, and then we measure relationships among those operationalized qualities: laughter events are a straightforward notion when viewed as identified occurrences in the corpus. However, laughter may be made operational in relation to the discourse structure. The concept of laughter functioning as a discourse marker is not directly measurable from the raw data, therefore our approach has been to construct a new way of making this concept operational and test the quantitative interactions among labeled laughter events and their durations on the one hand, and other qualities emerging from the dialog structure and participants, such as topic sections, speaker role, gender and feedback content on the other hand.

Conclusion

We consider this work as an incremental contribution to the study of laughter in relation to its function as a discourse marker, and not an exhaustive analysis of laughter functions and physiology within this corpus. In this work laughter is considered in its quantity (counts) and duration aspects, annotated by external raters. Other aspects, which are beyond the scope of this paper, would include qualitative and acoustic analysis of laughter, and examining how specific combinations of participants' sex or familiarity levels within the group composition may affect laughter expressions. Laughter may be approached under different perspectives, and the corpus data offer the opportunity to address those perspectives in future work. Such work would include yet another dichotomy relevant to the flow of the dialog, that of a laughter being shared or not among the participants: a laughter event may be ratified by other dialog participants, or not. And if laughter is shared, then it would be safe to

assume that this laughter is more likely to be mirthful, as dialog participants may recognize the mirth aspects in the dialog context, and ratify those by laughing.

In addition, we acknowledge that having the data annotated by additional raters might yield different disagreement levels among raters; however, in this study, we have chosen to treat disagreement as a value of interest (*Ambiguous*), instead of disregarding laughter events labeled as such, in an attempt to better understand and to inspect the distribution of laughter of those cases with respect to our primary research question, i.e., their function as a discourse marker.

Our results are rooted in a well-structured dataset and we argue that our approach can be replicated in other dialog data and bring similar findings. We can speculate on possible uses of the results. They are relevant to technology that is hoped to be imbued with believable social signals like laughter, where the distribution of such signals should be akin to what happens in natural dialog. They may also be relevant to the computationally assisted interpretation of multimodal recordings of dialog, where the distribution of laughter may support assessment of relative engagement of dialog participants. Plato's recording of the Socratic dialogs suggests that the timing of laughter has been important for millennia.

Acknowledgements The authors would like to thank the two raters, Siddhitha Sundari Bhoopathy and Thiago Sessegolo Marques de Almeida, who performed the laughter annotation of the dataset.

Author Contributions Conceptualization, M.K. and C.V.; methodology, M.K. and C.V.; formal analysis, M.K. and C.V.; investigation, M.K. and C.V.; data curation, M.K.; writing — original draft preparation, M.K.; writing-review and editing, M.K. and C.V.; visualization, M.K. and C.V.; supervision, M.K. and C.V.; project administration, M.K.; funding acquisition, M.K. and C.V.

Funding Open Access funding provided by the IReL Consortium.

Data Availability The data and annotations analyzed for this study are available from the MULTISIMO website (<http://multisimo.eu/datasets.html>).

Declarations

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Naas M. Plato and the spectacle of laughter. *Angelaki*. 2016;21(3):13–26. <https://doi.org/10.1080/0969725X.2016.1205253>.
- Trouvain J, Truong K. *Laughter*. Routledge handbooks in linguistics. Routledge, United Kingdom; 2017. p. 340–355.
- Glenn P. *Laughter in interaction*. Studies in Interactional Sociolinguistics. Cambridge University Press; 2003. <https://doi.org/10.1017/CBO9780511519888>.
- Vettin J, Todt D. Human laughter, social play, and play vocalizations of non-human primates: an evolutionary approach. *Behaviour*. 2005;142(2):217–40.
- Mazzocconi C, Tian Y, Ginzburg J. Multi-layered analysis of laughter. In: 20th Workshop on the Semantics and Pragmatics of Dialogue, Proceedings of the 20th Workshop on the Semantics and Pragmatics of Dialogue. New Brunswick, United States; 2016. p. 97–107. <https://hal-univ-diderot.archives-ouvertes.fr/hal-01371159>.
- Hurford J. Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*. 1989;77:187–222. [https://doi.org/10.1016/0024-3841\(89\)90015-6](https://doi.org/10.1016/0024-3841(89)90015-6).
- Koutsombogera M, Vogel C. Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus. In: N.C.C. chair, Choukri K, Cieri C, Declerck T, Goggi S, Hasida K, Isahara H, Maegaard B, Mariani J, Mazo H, Moreno A, Odijk J, Piperidis S, Tokunaga T, editors. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Miyazaki, Japan; 2018. p. 2946–2951.
- Tanaka H, Campbell N. Classification of social laughter in natural conversational speech. *Comput Speech Lang*. 2014;28(1):314–325. <https://doi.org/10.1016/j.csl.2013.07.004>. <http://www.sciencedirect.com/science/article/pii/S0885230813000545>.
- Mazzocconi C, Tian Y, Ginzburg J. What's your laughter doing there? A taxonomy of the pragmatic functions of laughter. *IEEE Transactions on Affective Computing*; 2020. p. 1–1. <https://doi.org/10.1109/TAFFC.2020.2994533>.
- Schiffrin D. *Discourse markers*. Studies in Interactional Sociolinguistics. Cambridge University Press; 1987. <https://doi.org/10.1017/CBO9780511611841>.
- Müller S. *Discourse markers in native and non-native english discourse*. John Benjamins; 2005.
- Fraser B. What are discourse markers? *J Pragmat*. 1999;31(7):931–952. [https://doi.org/10.1016/S0378-2166\(98\)00101-5](https://doi.org/10.1016/S0378-2166(98)00101-5). <https://www.sciencedirect.com/science/article/pii/S0378216698001015>.
- Fischer K. *The Loaded Discipline?*
- Fischer K. *Approaches to discourse particles*. Brill, Leiden, The Netherlands; 2006. <https://brill.com/view/title/23248>.
- Fischer K. *Towards an understanding of the spectrum of approaches to discourse particles: introduction to the volume*. Elsevier; 2006. p. 1–20.
- Schiffrin D. *Discourse Markers: Language, Meaning, and Context*, chap. 3. John Wiley & Sons, Ltd; 2005. p. 54–75. <https://doi.org/10.1002/9780470753460.ch4>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470753460.ch4>.
- Bonin F, Campbell N, Vogel C. Time for laughter. *Knowl-Based Syst*. 2014;71:15–24. <https://doi.org/10.1016/j.knosys.2014.04.031>.
- Bonin F, Campbell N, Vogel C. The discourse value of social signals at topic change moments. In: Proceedings of Interspeech 2015. 2015. p. 1967–1971. <https://doi.org/10.21437/Interspeech.2015-433>.
- Gilmartin E, Bonin F, Vogel C, Campbell N. Laughter and topic transition in multiparty conversation. In: Proceedings of the SIGDIAL 2013 Conference, The 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue, 22–24 August 2013, SUPELEC, Metz, France. 2013. p. 304–308. <http://aclweb.org/anthology/W/W13/W13-4045.pdf>.
- Holt E. The last laugh: shared laughter and topic termination. *J Pragmat* 2010;42(6):1513–1525. <https://doi.org/10.1016/j.pragma.2010.01.011>. <http://www.sciencedirect.com/science/article/pii/S037821661000024X>. *Laughter in Interaction*.
- Vinciarelli A, Chatziioannou P, Esposito A. When the words are not everything: the use of laughter, fillers, back-channel, silence, and overlapping speech in phone calls. *Frontiers in ICT* 2015;2:4. <https://doi.org/10.3389/fict.2015.00004>. <https://www.frontiersin.org/article/10.3389/fict.2015.00004>.
- Ginzburg J, Breitholtz E, Cooper R, Hough J, Tian Y. Understanding laughter. In: 20th Amsterdam Colloquium. Amsterdam, Netherlands; 2015. <https://hal-univ-diderot.archives-ouvertes.fr/hal-01371396>.
- Szameitat DP, Alter K, Szameitat AJ, Wildgruber D, Sterr A, Darwin C. Acoustic profiles of distinct emotional expressions in laughter. *J Acoust Soc Am*. 2009;126(1):354–66.
- Poyatos F. The many voices of laughter: a new audible-visual paralinguistic approach. *Semiotica*. 1993;93(1–2):61–82.
- Dupont S, Çakmak H, Curran W, Dutoit T, Hofmann J, McKeown G, Pietquin O, Platt T, Ruch W, Urbain J. *Laughter research: a review of the ILHAIRE Project*. Springer International Publishing, Cham; 2016. p. 147–181. https://doi.org/10.1007/978-3-319-31056-5_9.
- ILHAIRE FP7 Project. Laying the groundwork for database collection. Deliverable 1.1. 2012. http://www.ilhaire.eu/pdf/deliverable_fp7_ilhaire_270780_wp1_d11_version5.3.pdf. Version 5.3.
- D'Errico F, Poggi I. “The bitter laughter” when parody is a moral and affective priming in political persuasion. *Front Psychol*. 2016;7:1144. <https://doi.org/10.3389/fpsyg.2016.01144>. <https://www.frontiersin.org/article/10.3389/fpsyg.2016.01144>.
- D'Errico F, Poggi I, Vincze L. Discrediting signals, a model of social evaluation to study discrediting moves in political debates. *Journal on Multimodal User Interfaces*. 2012;6:163–78.
- Ruch W, Ekman P. The expressive pattern of laughter. In: A.W. Kaszniak (ed.) *Emotion, qualia, and consciousness*, pp. 426–443. Word Scientific Publisher, Tokyo (2001). <https://doi.org/10.5167/uzh-77825>
- Niewiadomski R, Hofmann J, Urbain J, Platt T, Wagner J, Piot B, Cakmak H, Pammi S, Baur T, Dupont S, Geist M, Lingenfeller F, McKeown G, Pietquin O, Ruch W. Laugh-aware virtual agent and its impact on user amusement. In: Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems, AAMAS '13. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC; 2013. p. 619–626. <http://dl.acm.org/citation.cfm?id=2484920.2485018>.
- Bachorowski JA, Smoski MJ, Owren MJ. The acoustic features of human laughter. *J Acoust Soc Am*. 2001;110(3):1581–97. <https://doi.org/10.1121/1.1391244>.
- Truong K, Trouvain J. Investigating prosodic relations between initiating and responding laughs. In: Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH 2014). International Speech Communication Association (ISCA); 2014. p. 1811–1815.
- Gilmartin E, Vogel C. Chat, chunk and topic in casual conversation. In: Proceedings 14th Joint ACL - ISO Workshop on Interoperable Semantic Annotation. Association for Computational Linguistics, Santa Fe, New Mexico, USA; 2018. p. 45–52. <https://www.aclweb.org/anthology/W18-4705>.

33. Provine RR. Laughter. *Am Sci* 1996;84(1):38–45. <http://www.jstor.org/stable/29775596>.
34. Martin RA, Kuiper NA. Daily occurrence of laughter: relationships with age, gender, and type a personality. *Humor*. 1999;12(4):355–84.
35. Bachorowski JA, Owren MJ. Not all laughs are alike: Voiced but not unvoiced laughter readily elicits positive affect. *Psychol Sci*. 2001;12(3):252–7. <https://doi.org/10.1111/1467-9280.00346> PMID: 11437310.
36. Wood A, Martin J, Niedenthal P. Towards a social functional account of laughter: acoustic features convey reward, affiliation, and dominance. *PLOS ONE*. 2017;12(8):1–19. <https://doi.org/10.1371/journal.pone.0183811>.
37. Lavan N, Scott SK, McGettigan C. Laugh like you mean it: authenticity modulates acoustic, physiological and perceptual properties of laughter. *J Nonverbal Behav*. 2016;40(2):133–49. <https://doi.org/10.1007/s10919-015-0222-8>.
38. Becker-Asano C, Ishiguro H. Laughter in social robotics - no laughing matter. In: *Intl. Workshop on Social Intelligence Design (SID2009)*. 2009. p. 287–300.
39. El Haddad K, Çakmak H, Gilmartin E, Dupont S, Dutoit T. Towards a listening agent: A system generating audiovisual laughs and smiles to show interest. In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction, ICMI 2016*. ACM, New York, NY, USA; 2016. p. 248–255. <https://doi.org/10.1145/2993148.2993182>.
40. Hough J, Tian Y, de Ruyter L, Betz S, Kousidis S, Schlangen D, Ginzburg J. DUEL: a multi-lingual multimodal dialogue corpus for disfluency, exclamations and laughter. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association (ELRA), Portorož, Slovenia; 2016. p. 1784–1788. <https://www.aclweb.org/anthology/L16-1281>.
41. Çakmak H, Urbain J, Dutoit T, Tilmanne J. The AV-LASYN database: A synchronous corpus of audio and 3D facial marker data for audio-visual laughter synthesis. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*. European Language Resources Association (ELRA). 2014. http://www.lrec-conf.org/proceedings/lrec2014/pdf/163_Paper.pdf.
42. Niewiadomski R, Mancini M, Baur T, Varni G, Griffin H, Aung MSH. Mmli: Multimodal multiperson corpus of laughter in interaction. In: Salah AA, Hung H, Aran O, Gunes H, editors. *Human Behavior Understanding*. Cham: Springer International Publishing; 2013. p. 184–95.
43. Petridis S, Martinez B, Pantic M. The MAHNOB laughter database. *Image Vis Comput*. 2013;31(2):186–202. <https://doi.org/10.1016/j.imavis.2012.08.014>. <http://www.sciencedirect.com/science/article/pii/S0262885612001461>. *Affect Analysis In Continuous Input*.
44. McCowan I, Lathoud G, Lincoln M, Lisowska A, Post W, Reidsma D, Wellner P. The AMI meeting Corpus. In: Noldus LPJJ, Grieco F, Loijens LWS, Zimmerman PH, editors. *Proceedings Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research*. Wageningen: Noldus Information Technology; 2005. p. 137–140.
45. Campbell N. An audio-visual approach to measuring discourse synchrony. In: *Multimodal Conversation Data*. Interspeech 2009, Brighton, UK. ISCA; 2009. p. 2159–2162.
46. Hearst MA. Texttiling: Segmenting text into multi-paragraph sub-topic passages. *Comput Linguist*. 1997;23(1):33–64.
47. Van Dijk TA. *Discourse, power and access*. Texts and Practices. Readings in Critical Discourse Analysis. Routledge; 1996. p. 84–104.
48. Gruenstein A, Niekrasz J, Purver M. Meeting structure annotation. Springer Netherlands, Dordrecht; 2008. p. 247–274. https://doi.org/10.1007/978-1-4020-6821-8_10
49. Luz S, Su J. The relevance of timing, pauses and overlaps in dialogues: detecting topic changes in scenario based meetings. In: Kobayashi T, Hirose K, Nakamura S, editors. *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010*. ISCA; 2010. p. 1369–1372. http://www.isca-speech.org/archive/interspeech_2010/i10_1369.html.
50. Galley M, McKeown K, Fosler-Lussier E, Jing H. Discourse segmentation of multi-party conversation. In: *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1, ACL '03*. Association for Computational Linguistics, USA; 2003. p. 562–569. <https://doi.org/10.3115/1075096.1075167>.
51. Hsueh P, Moore J, Renals S. Automatic segmentation of multi-party dialogue. In: *EACL*. 2006.
52. Sherman M, Liu Y. Using hidden Markov models for topic segmentation of meeting transcripts. In: *In SLT*. 2008. p. 185–188.
53. Bonin F. *Content and context in conversations : the role of social and situational signals in conversation structure*. Phd thesis, Trinity College Dublin; 2016.
54. Vogel C, Koutsombogera M, Costello R. *Analyzing likert scale inter-annotator disagreement* Springer Singapore, Singapore; 2020. p. 383–393. https://doi.org/10.1007/978-981-13-8950-4_34.
55. Reverdy J, Koutsombogera M, Vogel C. Linguistic repetition in three-party conversations. Springer Singapore, Singapore; 2020. p. 359–370. https://doi.org/10.1007/978-981-13-8950-4_32.
56. Vogel C, Koutsombogera M, Esposito A. Aspects of methodology for interaction analysis. In: *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. 2020. p. 141–146. <https://doi.org/10.1109/CogInfoCom50765.2020.9237908>.
57. Sacks O. *The President's Speech*, chap. *The man who mistook his wife for a hat*. Picador; 1985. p. 76–80.
58. Schmid H. Probabilistic part-of-speech tagging using decision trees. In: *Proceedings of the International Conference on New Methods in Language Processing*. Manchester, UK; 1994. p. 154–164.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.