Universität Paderborn

Fakultät für Kulturwissenschaften

Adaptive Explanations: The Involvement of Explainees

Dissertation zur Erlangung des akademischen Grades Doktor der Philosophie (Dr. phil.)
im Fach Psycholinguistik der Universität Paderborn

von

Josephine Beryl Fisher

Erstbetreuerin
Prof. Dr. Katharina J. Rohlfing

Zweitbetreuer
Prof. Dr. Anke Lenzing

Prüfungskommission
Prof. Dr. Ilka Mindt
Dr. Angela Grimminger

Paderborn
2025

Eidesstattliche Erklärung

Hiermit erkläre ich gemäß §12 der Promotionsordnung:

(a) dass die vorgelegte Arbeit selbstständig und ohne Benutzung anderer als der in der Arbeit angegebenen Hilfsmittel angefertigt wurde;

(b) dass die Arbeit bisher weder im In- noch Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt wurde;

(c) dass früher oder gleichzeitig kein Promotionsverfahren bei einer anderen Hochschule oder bei einer anderen Fakultät beantragt wurde, gegebenenfalls nebst vollständigen Angaben über dessen Ausgang.

Paderborn, 30.06.2025

J.B. Fisher

Contents

1 Introduction

"*Contributing to a discourse [...] appears to acquire more than just uttering the right words at the right time. It seems to consist of collective acts performed by the participants working together*" (Clark and Schaefer, 1989, p. 259). The question the authors pose is how the interaction partners contribute to discourse. Conversational partners can have different degrees of verbal involvement in interactions (Fisher et al., 2022). Various aspects that may be important are: the roles of the interaction partners, the topic under discussion, and their conversational partner. Whereas research concerned with the involvement of both partners considers how the speakers adjust their speech (e.g., Fischer (2016)), little is known about how both interaction partners are involved for a specific purpose such as explanations. The involvement for the purpose of an explanation can be described on the verbal level and in how far the interaction partners contribute to the goal of understanding. Explanations are important to come to an understanding (Quasthoff et al., 2017). Therefore, they are of great importance in various areas of life (e.g. educational and communicative contexts).

The importance of the explainee has been shown in various works qualitatively in Conversation Analysis (Quasthoff et al., 2017; Morek et al., 2017; Gosen et al., 2025). Especially in narrations, argumentations and explanations, the joint accomplishment of conversational tasks has been shown on the level of the discursive structure (Quasthoff et al., 2017). In Conversation Analysis the aim is to qualitatively examine interactions and how they are accomplished jointly (Gosen et al., 2025).

This work focuses on the interaction partners, especially the verbal behaviour of the explainee. The verbal involvement is examined on different levels: on a speaker move level, structural level, and semantic level. Thereby, the verbal involvement is linked to cognitive processes. Therefore, this analysis is classically placed within the field of Psycholinguistics as the examination is interaction driven and based on empirical data (Morek et al., 2017).

The role of the less knowledgeable interaction partner is becoming more important also in the linguistic fields of pragmatics and sociolinguistics (Hansen, 2025). So far, research has mainly focused on the explainer (Rohlfing et al., 2021; Sokol and Flach, 2020) and not on how the explainee becomes actively involved in an explanation.

Additionally, the need for the investigation of the explainee involvement arises as current approaches in Explainable Artificial Intelligence (XAI) attempt to satisfy multiple user needs by enabling their involvement, but for this, the empirical evidence is lacking (Fisher et al., 2022). The attempt is to make the algorithms, highly complex Machine Learning systems, explainable and to unpack their black-box systems (Sokol and Flach, 2020). As the authors note, a vast amount of tools to enable transparency is not interactively including the explainee, although these systems need to be understandable for their users and require bi-directional communication (Sokol and Flach, 2020). The shift towards the explainee in XAI is grounded in the growing need for explainability to heighten the trust in such systems (Miller, 2019). Consequently, the conversational partners and how they address certain topics is more important than what is being explained (Besold and Uckelman, 2018). Thereby, special attention is given to the needs, for example knowledge gaps, of the explainee. Therefore, interaction partners adapt their speech towards each other (Traxler, 2012). The aim of explaining is not to provide a complete explanation that covers all the details of a topic, but to adjust the content to their conversational partner (Rohlfing et al., 2021). In doing so, explanations can depict a social process (Miller, 2019). From an interactive perspective, research on human–machine interactions has underscored that successful interactions are driven not only by the actions of the explainer, but also by the active participation of the explainee (Sokol and Flach, 2020). In line with this from the co-constructive view (Rohlfing et al., 2021), we claim that explainees, the less knowledgeable person, also actively shape the explanation by various linguistic means. Therefore, we examine in human–human explanations to what extent the verbal behaviour of the explainee influences the explanation. In particular, we are looking at the concrete verbal behaviour of the explainees and what type of speaker moves they are using (Chi et al., 2008). These moves can be categorised as different speaking strategies by the interaction partners. With regard to this, we would like to investigate which moves are used by the interaction partners to adapt to their interaction partner and how the degree of adaptivity can be tested and measured. Beyond the moves, also their semantic content is analysed which allows for showing how the content is negotiated in the explanations. Consequently, the question arises: How is the explainee verbally involved in an explanation and what effect does it have on the explanation?

To examine the crucial role of the explainee in shaping explanations and fostering understanding, we explored multiple explanation datasets with different foci. These different datasets are the ECOLANG Corpus (Gu et al., 2025), ADEX game explanation corpus (Fisher et al., 2023; Fisher and Rohlfing, 2025) and a hospital dataset including consultations of children operations (Fisher et al., 2022). They offer complementary insights into the co-construction and verbal adaptation across varied topics. The ECOLANG Corpus (Gu et al., 2025) is crucial as it links the verbal behaviour of the explainees, specifically the questions they pose, with their subsequent learning outcomes. These outcomes were measured in form of a recall task, yielding a recall score. Therefore, the understanding in form of word learning is captured in relation to their different types of questions. For more details, see 4.1. The ADEX corpus of game explanations provides game explanations in form of three different studies with different conditions. Within the different studies, we vary the verbal behaviour of the explainee. In Study 1, explainees are actively engaged; in Study 2, they are trained to adopt a passive, non-substantive role (limited to neutral backchannels); and in Study 3, they are trained to engage substantively. This systematic variation enables a direct comparison of how the explainee's level of verbal involvement shapes the explainer's verbal behaviour. Consequently, we examine the mechanisms of interactive adaptivity in explanations. More information is provided in 4.5. The hospital dataset offers rare insight into naturally occurring high-stakes explanations in authentic clinical settings, unlike the lab-based environments of the other data sets. It captures the unfolding of explanations during paediatric surgical consultations. In particular, the hospital dataset offers results on how the explanations are taking place, regarding who is involved at which time point. For a detailed account, see 4.4.

The following chapters are structured as follows. In the theoretical background, key concepts and theories related to adaptivity and explainee involvement will be briefly introduced. Building upon this foundation, relevant published papers in this regard will be presented (see Research Questions and Hypotheses). The discussion of the papers encompasses their research questions, hypotheses, and the main results, as well as their relevance and connection with regard to the explainee involvement. Afterwards, in Chapter 5, the synthesis and the overarching discussion follows. Finally, Chapter 6, which presents the conclusion and outlook, concludes the work by reflecting the limitations and suggesting aspects for future research.

2 Theoretical Background

The following section establishes the framework for the investigation of interactive adaptivity. This encompasses previous research on the involvement of the explainee as such, and adaptation processes in explanations. Addressing the research gap of the lacking empirical evidence for how to involve the explainee in XAI research (Fisher et al., 2022), we set out to examine how an explainee verbally contributes in a human–human explanation. In order to do so first, the relevant theories and main mechanisms driving a dialogue are presented. This includes, some adaptation concepts such as alignment, synchronicity, and co-construction. Secondly, the structural organisation of explanations is presented in relation to conversational jobs. In addition, more monological and dialogical phases are discussed. These phases indicate how many people are involved in that specific phase of the explanation.

2.1 Interactive Adaptivity

As previously stated, we claim that explanations are co-constructed by both interaction partners (Rohlfing et al., 2021). The interactive view on dialogue drives this assumption. This research is placed in the paradigm of constructivism in the area of interactionism. According to research on constructivism (Kalina and Powell, 2009), explainees interactively construct knowledge in classroom settings. In line with interactionism, conversational partners aim to reach a shared understanding (Clark, 1996; Quasthoff et al., 2017) and continuously negotiate what is being explained (Rohlfing et al., 2021). Consequently, significant parts of the dialogue are established by "*interacting, cooperating, and collaborating on the content*" (Traxler, 2012, p. 308). Thus, dialogues are joint activities, which include task-oriented dialogues (Tenbrink, 2020), to which explanations also belong.

The overarching concept of interactive adaptivity frames the whole examination of the verbal behaviour of the interaction partners and especially the involvement of the explainee in our investigation. Interactive adaptivity captures the adaptation of conversational partners on a verbal level (Buhl et al., 2025). In order to interactively, on the verbal level, adapt towards each other, both partners presumably need to be involved in the dialogues, in our case explanations. To capture the adaptation of interaction partners, we will take a look at the concrete verbal behaviour (Buhl et al., 2025). Thereby, we are interested in how far the verbal behaviour of the interaction partners affects the other partner, which poses a typical question in discourse analyses (Ferreira, 2024).

In an explanation, interaction partners aim to reach a shared understanding (Quasthoff et al., 2017). Each explanation involves two interaction partners: an explainer, who possesses more knowledge, and an explainee, who has less knowledge. The subject of the explanation is referred to as the explanandum, which is made up of different kinds of explanans (Rohlfing et al., 2021). Following the co-constructive view (Quasthoff et al., 2017; Rohlfing et al., 2021) on explanations, what is being explained is not determined beforehand, but is continuously negotiated among conversational partners throughout the explanation. Therefore, the involvement of both interaction partners who jointly co-construct the explanation is important. In these types of interactions, a knowledge asymmetry exists between the speakers by the explainer knowing more about the explanandum than the explainee. The same applies to tutoring (Chi et al., 2004, 2001) and school settings (Bressler, 2023). In tutoring settings,the active involvement of both interaction partners has already been shown to lead to deep learning effects (see Chi, 2009, for a summary). The temporal dynamics and content structuring of explanations (e.g., nodes and moves) remain under-explored, justifying the research focus. Therefore, we attempt to address the research gap of how exactly an explainee is verbally involved in an explanation.

2.2 Common Ground

The aim of interaction partners is to come to a joint understanding in dialogues as such (Clark, 1996), and especially in explanations (Quasthoff et al., 2017). How they achieve this is described in the following. The basis for interactive adaptivity is formed by the phenomenon recognised as common ground, as it views interactions as joint activities (Clark, 1996). This is a prerequisite for successful understanding, which is especially important in the case of explanations (Rohlfing et al., 2021). According to Clark and Schaefer (1989), common ground includes the information that both speakers share. Throughout the conversation, the interaction partners jointly participate in the expansion of common ground (Clark, 1996). Discourse models from various fields such as linguistics, philosophy, psychology, and artificial intelligence mostly agree on three assumptions: (1) speakers assume a certain degree of common ground before a conversation, (2) this common ground increases over time; and (3) turns have to be placed at the right time to add to the common ground (Clark and Schaefer, 1989). Although "common ground" sounds very similar to "common knowledge" or "shared knowledge," it goes beyond mere shared information. Simply knowing the same thing does not constitute common ground. Interaction partners can both know something without realising that it is shared. However, common ground requires that both conversational partners are aware that the knowledge is mutually known. In essence, common ground is mutually recognised shared knowledge (Fischer, 2016; Traxler, 2012).

Depending on who we are talking to, we may or may not share a common ground, and which pieces of information are shared among interaction partners needs to be negotiated (Traxler, 2012). Therefore, speakers carefully monitor the knowledge construction of the other speaker in relation to their common ground. Consequently, they partake in *audience design* by adapting their speech in accordance to their conversational partner (Traxler, 2012). Thus, interaction partners design their language to be comprehensive, which is a regular part of speech production (Ferreira, 2024).

All in all, the common ground theory sheds light on the process by which interaction partners select specific expressions to denote particular concepts, as well as how these expressions develop and change throughout multiple interactions (Traxler, 2012). Therefore, one possible form of operationalisation of common ground can be made when focussing on the lexical items both interaction partners use. This will be further discussed later in section 2.4. I adopt the approach of Clark (1996) to see dialogue as a joint activity.

2.3 Turn-taking

One of the most essential mechanisms enabling interaction is turn-taking (Levinson and Torreira, 2015). A turn can consist of a single word, a sentence, or several sentences by a speaker (Sacks et al., 1974)

Turn-taking showcases the coordination of speaker transitions and the determination of who speaks at which point in time in a conversation (Sacks et al., 1974). The next speaker is either determined by explicitly addressing the other speaker or by self-selection. In the latter case, the original speaker can choose to continue speaking or the other speaker can choose to speak without being explicitly addressed (Sacks et al., 1974). As a result, the turn-taking mechanism facilitates effective interactions by managing the allocation of speakers and their turns. A smooth interaction is marked by speakers taking turns to discuss a given topic without many overlaps or interruptions (Sacks et al., 1974).

As Stukenbrock (2013) summarises, turns can be made up of a word, phrase, sentence, or even several sentences. Each of these can form a turn construction unit. A turn can consist of multiple turn construction units. Their length is not fixed beforehand. After each turn construction unit, a speaker change could possibly occur. This makes it a transition-relevant place. When a turn is being completed, there are indications that can appear at the syntactic, semantic , context, intonational or non-verbal level (Stukenbrock, 2013). During interaction, conversation partners also provide backchannels which do not function as turns. They, as well as turns, are used to display understanding (Traxler, 2012). Levinson (2016) highlights the universal nature of turn-taking across different languages and turn-taking within primates. Empirical evidence for the universal character of turn-taking is provided in the study by Stivers et al. (2009). Their results also show that it is applicable across languages and cultures.

In the following, the characteristics of turns, such as length and form, will be described in more detail. Turns are usually short, on average two seconds in length. Only 5% of speech constitutes simultaneous speech by two or more speakers. The average pause between speakers is only 200 ms long, though it takes 600 ms to produce a word. Thus, interaction partners often manage to have shorter pauses than the separate speech production takes (Levinson and Torreira, 2015). Consequently, the pause between speech comprehension and production should take longer if they take place separately and not parallel. This is the core of the so-called psycholinguistic puzzle that interaction partners often manage to have pauses that are shorter than the time it takes to produce speech (Levinson and Torreira, 2015). EEG analysis showed that as soon as speakers in their speech comprehension recognise the function of a turn, speech preparation for the response begins. In the first 400 ms after the start of a turn, speakers very quickly recognise the function of a speech contribution (Gisladottir et al., 2015). Levinson and Torreira (2015) purpose the rapid speech act detection in the turn as a key resource to overcome the psycholinguistic puzzle because it forms their own reply. This makes the reaction time faster, allowing for quick responses and leading to the suggestion of overlapping comprehension and speech production (Levinson and Torreira, 2015). The matter of speech acts will be further discussed in 3.2.

In general, turn-taking in interactions appears to be a challenging task for interaction partners because speech comprehension and production take place in parallel for quick response times (Levinson and Torreira, 2015). The mechanism as such appears to be universal. However, the concrete processes involved in the parallel language comprehension and production are not fully explained yet.

2.4 Alignment

Interaction partners can align their linguistic representations during a dialogue on different linguistic levels (Pickering and Garrod, 2004a). Pickering and Garrod (2004a) proposed the so-called interactive alignment account especially for dialogues, that explains and fosters successful communication. Like the turn-taking mechanism, the interactive alignment account describes an interactive mechanism that addresses language processing and production. It explains why it is possible to engage in such rapid turn-taking by coupling the representations of interaction partners in speech production and speech perception. Thus, it proposes a solution to the psycholinguistic puzzle, as, according to the authors, speakers exploit the benefits of interactive alignment (Garrod and Pickering, 2004). From a cognitive perspective, conversations should be extremely complex, since utterances can be elliptical, timing of turns needs to be socially acceptable, and the parallel comprehension and production of speech happens simultaneously (Garrod and Pickering, 2004). Garrod and Pickering (2004) follow the line of research of Clark and Schaefer (1989) and Clark (1996) that conversations are a joint activity in which both conversational partners aim to come to a joint understanding. When taking a closer look at common ground, the question arises of how it is established jointly among conversation partners. To address this matter, the interactive alignment account provides answers, encompassing priming and the automatic alignment across multiple linguistic levels. The reasons for interactive alignment are twofold. It appears because the linguistic representations of interaction partners are primed and there is a parity of representations in both processes of speech production and processing. Priming of representation leads to imitation and that in turn to alignment (Garrod and Pickering, 2004).

Pickering and Garrod (2004a) link alignment with common ground. According to the interactive alignment account, there are different linguistic levels of alignment which lead to an implicit common ground, previously termed as common knowledge. The common ground is implicit because it is automatically modelled. In contrast, common ground is achieved by the active modelling of shared understanding (Pickering and Garrod, 2004b). Interaction partners reuse their partners' lexical items, sounds, syntax, or meanings. The lexical items refer to the words which are being used. This form of alignment is being considered in the different types of verbal behaviour in the form of speaker moves in the category of *repeat* (3.2). This captures words being repeated verbatim. The phonetic representations relate to the different sounds, more particularly how the words are being articulated. On another level, syntactic alignment occurs when the syntax is copied from previous utterances. Semantically they align on the content they discuss. On an additional level, the situation models of the interaction partners become aligned. This includes the mental representation of what is being discussed.

In more detail, the interactive alignment account by Pickering and Garrod (2004a) proposes that alignment at one level, for example, the lexical level, automatically leads to alignment at higher levels (such as semantic or phonetic) and eventually to the alignment of the situational models. Therefore, interaction partners can better communicate as it ensures that both are talking about the same topic. If one uses the same words, the likelihood is higher that they have the same concepts in mind. The alignment at the different levels tends to increase throughout the interaction.

Following this line of research, Reitter and Moore (2014) investigated, in corpus-based studies, the syntactic alignment of speaking partners. Their findings support the claim that alignment at a lower level leads to higher-level alignment. The authors report that task success correlated with syntactic alignment of participants. Therefore, lower-level alignment enhances task success.

In contrast to this, Fusaroli et al. (2014) present a dynamic framework for analysing dialogue grounded in the concept of *interpersonal synergy*. At the core of this synergistic approach is the perspective of dialogue as a self-organising and dynamic, interpersonal system that enables coordinated and effective interactions. The researchers claim that explanation partners align with each other, not only by their mere alignment but also by their complementation. These complementary dynamics encompass that interaction partners are functioning in a dyad and not as individuals in dialogue. To illustrate how synergy extends beyond alignment, Fusaroli et al. (2014) examine cases involving speech impairments, drawing on research by Dressler et al. (2009); Goodwin (2003); Wilkinson et al. (2003) which deals with what happens when an interaction partner has a speech impairment. They claim that it does not lead to an overall impairment of the interaction. Instead, the system adapts because this is compensated by the unimpaired partner, while the impaired speaker develops strategies for compensation. Therefore, "*the dialog – as a synergetic system – can compensate in order to preserve its functionality*" (Fusaroli et al., 2014, p.154). Moving partially in the same direction, Dubuisson Duplessis et al. (2021) investigate the shared, as well as the individual lexicons of interaction partners. Building on this, they propose a measurement for lexical alignment. This goes in the same vein as Fusaroli et al. (2014) who point out that "*[a] joint task, for instance, could require the interlocutors to develop a shared vocabulary for jointly guiding attention to and talking about the particularities of the task*" (Fusaroli et al., 2014, p. 151). Consequently, both approaches mention the joint vocabulary of the interaction partners to investigate their alignment.

Alignment and common ground both depict that the interaction partners adapt towards each other but how they achieve this is not stated.

2.5 Structure of explanations

In the following, the focus shifts from mechanisms and phenomena in interactions towards their structure in order to examine more closely how conversational partners adapt towards each other. Dialogues follow specific structures, which can be operationalised as recurrent patterns in interactions that are constituted by a set of sequentially organised actions called pragmatic frames (Rohlfing et al., 2016).

In the work of Kobayashi (2022) the structure of explanations is empirically shown to be divided into initial and interactive phases. The initial explanation phase includes the tutor (explainer) providing some introductory information and the tutee (explainee) being rather passive by not providing verbal feedback. In the interactive phase, both interaction partners are involved in question-answer sequences in which problems in understanding are solved (Kobayashi, 2022).

A further approach to describe the sequential organisation of explanations, with regard to the discursive structure, is proposed by Quasthoff et al. (2017). There, the structure of a dialogue is described by different universal conversational tasks, so called conversational jobs, that are co-constructed by the interaction partners. According to the authors, interaction partners fulfil five different conversational jobs across narrations, argumentations, and explanations (Quasthoff et al., 2017). Fisher et al. (2022) adapted the job coding of explanations as follows for the consultations. The jobs remain the same as for other explanations, but what they include in detail is adjusted. In job 1 *Establishing topical relevance*, the following interaction is framed which starts with an extended monological phase. Job 2 *Constituting an explanandum* includes the interaction partners specifying the explanandum for the explanation. In the doctor–patient consultation this included the naming of the important body parts. In the core job 3 *Explication procedural, conceptual and/or causal relation*, the explainer provides all necessary information concerning the topic of the explanation, while the explainee signals attention or asks questions. This included, in particular, the relevant information of the upcoming operation with regard to the procedure and its justification. An example where the procedure is further discussed can be seen in the following, explainer (doctor) says "*The fixation usually takes four weeks*". In that example, the explainer specifies for how long the fixation of the broken bone will last. Job 4 *Closing* frames the end of the explanation by explicitly stating that this was the end or by opening the room for any kind of questions. An example for opening the room for questions, taken from the game explanation would be the explainer (doctor) asking "*Do you have any further questions?*". In Job 5 *Transition*, the interaction partners are getting out of the explanation and back to a turn-by-turn structure. In the data, this included signing the relevant documents by the legal guardians to allow for the operation. These conversational jobs are co-constructed by both the explainee and explainer (Fisher et al., 2022). For more examples for the conversational jobs see 9.

In the following, the conversational jobs will not be further examined because they describe on a rather broad level the discursive structure. I will focus more closely on the concrete verbal behaviour of the interaction partners. The jobs can be useful when examining the context in which a speaker move or semantic node is mentioned.

2.6 Phases in explanations

To account for the active involvement of the explainee in explanations, their degree of involvement at different time points can be analysed via their turns and resulting phases. Fisher et al. (2022) found that explanations consist of two alternating phases, namely monological and dialogical phases. These can be investigated by considering the speech. In a monological phase, there is only one speaker talking. The other speaker can make backchannels, but no turns. In the dialogical phases, both speakers have alternating turns. This shows that both speakers are actively involved in an explanation because there are several dialogical phases in which the explainee is involved, and sometimes they even initiate those themselves. Occasionally, explainees also have monological phases in which they explain certain aspects. Thereby, the coding of phases addresses the existing research gap concerning the lack of empirical evidence on how to involve the explainee in XAI research.

Taken together, this chapter has established the theoretical background for the investigation of interactive adaptivity. The concept is rooted in the idea that explanations are co-constructed by both interaction partners, and that successful communication relies on the establishment of common ground via turn-taking, and alignment. Therefore, the chapter has reviewed key theories and mechanisms driving dialogue. Additionally, the structure of explanations, the different phases, and conversational jobs that are co-constructed by the interaction partners have been discussed. The research gap in the field of XAI has been identified, highlighting the need for empirical evidence on how to involve the explainee in explanations. The next step is to address the methodological challenges of investigating interactive adaptivity in explanations, which will require the development of innovative methods and tools to analyse the complex dynamics of explanation dialogues.

3 Methodological challenges

In the following section, the focus lies on the analytical units employed to investigate the explainee's involvement and in how far they pose methodological challenges and demand methodological solutions. These units comprise speaker moves, assessed in terms of their substantiveness and semantic content (game nodes).

I will focus on the methodological procedure of how to examine the verbal involvement of the explainee. Deciding on which level of segmentation and detail to include in the transcript is difficult and should be motivated by the research questions one is investigating (Tenbrink, 2020). As we are interested in the involvement of the explainee and the quality of their speaker moves, these are the units to use for the analysis (for more details, see 3.3). The criteria to define a speaker move are semantic and prosodic. On the one hand, the single idea has to be transported semantically, and on the other hand the prosody confirms the unit. Similarly to this, Hu and Degand (2023) propose the conversational discourse unit. Beyond the previously mentioned features, also the syntax is included. We are following this approach by using the speaker moves as a unit and to have several nodes, pieces of information, included.

3.1 Nodes

In the following, the semantic analysis, the so-called nodes, will be further described. The nodes enable the analysis of adaptation on a semantic level. Consequently, the semantic coding scheme was developed to make the explanation content explicit, transferring the different semantic contents from the verbal data of the explanations into numerical data. In Fisher et al. (2023) and Fisher and Rohlfing (2025), explanation nodes were assigned to each utterance to support the speaker move analysis (content level).

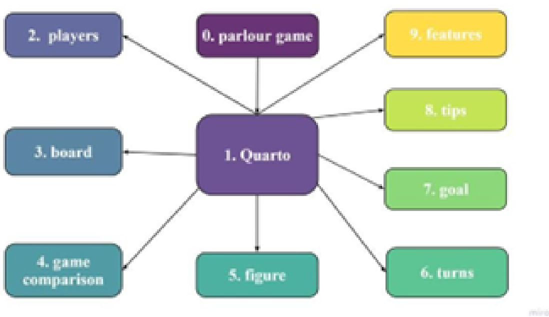


Figure 3.1: Blocknodes from the game explanations of Quarto (Fisher et al., 2023)

A game explanation encompasses various elements of a game, including its rules, pieces, and the game board. As noted by Kotthoff (2009), this type of explanation, similar to route and word explanations, can be quickly understood and immediately applied. The author also refers to game explanations as procedural explanations (Kotthoff, 2009). In order to divide the speech of these explanations into speaker moves, one has to first develop a coding scheme for the semantics (Chi et al., 2008). This was based on semantic maps developed for tutoring sessions (Chi et al., 2008; Miyake, 1986). Both maps were only based on the explainer explaining alone and not on the interaction with the explainee.

In contrast, Fisher et al. (2023) developed the node scheme based on the content both interaction partners addressed. In detail, the node scheme comprises nine blocks of nodes. Each block contains a certain topic and consists of different subnodes. For example, the block *Figures* has different subnodes such as *Amount*, *Individuality*, *Height*, *Form*, *Structure*, and *Colour* and each of those have further subnodes. The mentioning of multiple nodes in an utterance was possible. For example, when talking about the figures, two nodes are already addressed when saying they are tall and small. The nodes were also used to keep track of when the conversational partners talked about certain topics. This enabled detailed comparisons of the utterances. Thereby, one captures when certain topics are covered in the interaction. This is referred to as the interaction history in the following. Hereby, one could differentiate, for example, whether something was repeated or paraphrased (put into different words). Moreover, topic initiations and uptakes by the conversational partners could be investigated.

3.2 Speaker Moves

The following chapter proceeds to present the verbal behaviours captured and examined within the unit of speaker moves. A speaker move or explanation strategy is defined as an utterance that is related to the topic under discussion (Chi et al., 2008). In order to define the topic, the verbal behaviour (speech) requires a detailed semantic analysis (see 4.2). Taking the semantic analysis of the utterances as a basis, the speaker moves were assigned. Fisher and Rohlfing (2025) extended the coding scheme for speaker moves from tutoring research (Chi et al., 2001, 2004, 2008; Tare et al., 2011) to explanations containing a more diverse set of moves for the explainer and explainee. In more detail, the differentiation of certain moves being only available to the explainer or explainee has been resolved. In doing so, the verbal behaviour of the interaction partners can be examined more thoroughly. This includes different types of questions, information transporting moves, such as *providing information* or *additional information*, as well as moves that mark that information was already mentioned before, like *repeating* or *paraphrasing*. For a full list of speaker moves and examples, see 9.

In order to differentiate the coding of speaker moves from other coding schemes examining verbal behaviour of interaction partners, three related coding schemes are presented, namely the Human Communication Research Centre (HCRC) annotation scheme (Carletta et al., 1997), Dialogue Act Markup in Several Layers (DAMSL) scheme (Allen and Core, 1997), and the taxonomy of illocutionary acts (Searle, 2002). In doing so, the moves are put into context of other linguistic approaches examining the verbal behaviour of interaction partners, and the benefit of using our move coding scheme for our purpose will become clear.

The Human Communication Research Centre (HCRC) annotation scheme includes three different coding levels (Carletta et al., 1997). The dialogue acts capture individual dialogue moves and task relevant dialogue acts. As Tenbrink et al. (2013, p. 189) summarises, "*basic dialogue acts are coded as conversational moves (first level), which* *indicate the communicative purpose of an utterance. Each utterance is coded with a single move*". These are divided into initiation moves, such as instructions, explanations, questions, and requests to access understanding or the readiness to continue. An initiation move is followed by a response move, like acknowledgments, question replies, and clarifications. How different moves, in the first level of coding, collectively achieve the aim of an initiation move is described on the game level (second level). The term game is used here to describe the linguistic level and not a game which one plays as in the case of the game explanations. Therefore, the moves constitute the fundamental units for the game structure. These games can be grouped and further differentiated in managing or performing the task. This takes place at the transactional level (third level), the substructure of the interactions, which constitutes the third coding level (Carletta et al., 1997).

In the next section, the purposed Dialogue Act Markup in Several Layers (DAMSL) scheme by Allen and Core (1997) is presented. In total, it has four layers or so-called tags. This includes two different discursive functions. On the one hand, the forward-looking functions (1) shape what follows in the conversation, such as asking questions, giving instructions, making suggestions, or statements. On the other hand, backward-looking functions (2) relate to how an utterance responds to what came before, like answering a question, showing agreement, or acknowledging understanding. The information level (3) categorises utterances based on whether they involve performing tasks, managing task-related planning or coordination, or handling dialogue management (e.g., greetings, repairs, delays). Lastly, the communicative status (4), identifies whether an utterance is complete, incomplete, abandoned, or unclear (Allen and Core, 1997; Tenbrink et al., 2013).

The final coding scheme considered here regard the speech acts which are differentiated in Searle (2002) taxonomy of five types of illocutionary acts, which a speaker can choose from. These are assertives, directives, commissives, expressives, and declarations. In more detail, explanations, statements, and assertions are part of the first type, namely assertives in which the speaker commits, in different degrees, to the truthfulness of the uttered statement. They convey information about the state of affairs. Directives aim to instruct the interaction partners on what they should do (e.g. order, comands, requests) and commissives; on the other hand, the interaction partner themselves commits to a certain action (e.g. wows, pledges, contracts). In expessives as the name already reveals, the interaction partner expresses their emotional state and and their point of view like apologising or congratulating. The last act are declarations which bring about a change of affairs, such as declaring war or marriage (Searle, 2002). Searle (2002) claims that those speech acts capture all aspects that we can use language for.

In the next section, the different coding schemes of verbal behaviour are contrasted with our speaker moves. Comparing to what we have defined as speaker moves, the dialogue acts by Carletta et al. (1997) are on a higher linguistic level because they do not combine cognition with the interaction in such detail. They cluster the different sets of moves together and form subordinate categories (games) in regard to how those fulfil specific aims. Moreover, as stated, an explanation is one of the initiation moves and is not further analysed in the dialogue act scheme. In contrast, we take a more detailed look at how the explanation takes place. Compared to the DAMSL scheme by Allen and Core (1997), our speaker moves are more fine-grained. They do not only focus on the information level whether the task, in our case the game, is addressed, but via the detailed node scheme, we can also capture what is being said exactly when. Therefore, the utterances are seen in context with the previous utterances. Therefore, this categorisation is on a broader level than the speaker moves as they do not focus on the semantic content (nodes), but rather the form of the utterances. As in the HCRC coding scheme (Carletta et al., 1997), the category of explanations is not further differentiated in the taxonomy.

These coding schemes constitute different approaches to coding utterances. The HCRC (Carletta et al., 1997) differentiates between conversational move, game, and transactions. In contrast, the Dialogue Act Markup in Several Layers (DAMSL) coding scheme (Allen and Core, 1997) differentiates on forward- or back-looking function, information level, and communicative status. Searle (2002) taxonomy of speech acts concentrates on assertives, directives, commissives, expressives, and declarations. These approaches rather depict how those utterances are formulated and how the interaction partners position themselves towards what was being said. In contrast, the examination of speaker moves, as we analyse them, combines a thorough semantic analysis with an in depth analysis of the explanation strategies.

3.3 Substantive Moves

In the following, the analysis of substantiveness is added to the examination of speaker moves. In general, the substantiveness refers to the quality of the speaker move in the sense to what extent it contributes to the semantic content of the domain. According to tutoring literature (Chi et al., 2001, 2008), the explainee's speaker moves can be further differentiated into *substantive* and *non-substantive* moves. On the one hand, a move is substantive if it directly links to the topic under discussion and forms a meaningful contribution to an ongoing activity, e.g. problem solving or response to tutors' (in our case explainers) explanation. Examples of substantive moves are reading sentences aloud, giving unprompted self-explanations, asking questions, answering questions, responding to scaffolding prompt by tutor, and reflecting (comprehension monitoring) (Chi et al., 2008). On the other hand, a move is non-substantive when the information was previously mentioned in the same way or it is topic non-related. Examples are continuer, repetitions, agreement, and off-task remarks by tutee (in our case explainee) (Chi et al., 2001). Therefore, one first has to examine the explanation nodes and the speaker moves of the utterances. Subsequently, jointly considering the information enables the substantiveness analysis.

For example, in an explanation the explainee says "*Similar to Nine men's morris*" this is a substantive move because the game comparison is mentioned for the first time by the explainee, in form of the move *providing information.* A substantive question can be "*How big is the board?*" as it addresses relevant game content (later referred to as *factual questions*).

In order to compare the quantity of substantive moves across subjects in explanations, Fisher and Rohlfing (2025) developed a substantiveness score. This encompasses the total frequency of substantive moves by the explainee in an explanation. The total frequency is used because we argue that each substantive move has an effect on the explainer.

Overall, the examination of the substantiveness of the explainees' moves allowed us to investigate the impact of substantive moves on the number of nodes being addressed and also how they affect the verbal behaviour of the explainee in the sense of how many nodes they address and how diverse their verbal behaviour is, e.g. how many different speaker moves they utter (for more details see 4.3).

The previous chapter has discussed some methodological challenges regarding the coding of nodes, speaker moves, and substantiveness. Combining these different units of analysis is an advantage for the investigation of the verbal behaviour of the interaction partners as it enables a more fine-grained analysis, which in turn captures the adaptivity of interaction partners. Through the detailed examination of the verbal behaviour, the substantiveness of the contributions by the explainee and their driving impact on the explanation can be investigated. Within the method the innovation lies within 1) the semantic node coding scheme, which is based on the verbal behaviour of both interaction partners, 2) resolving the role-specific explainer and explainee moves, and 3) quantifying the quality of the speaker moves by the explainee in a substantiveness score.

4 Research Questions and Hypotheses

In the following, the different papers of this dissertation will be briefly summarised and put in relation to one another. This includes their research questions and hypotheses, methods, main results, and their overall contribution towards addressing the involvement by the explainee in explanations. This creates a multilayered analysis of the linguistic aspects of the explainee's verbal behaviour.

In order to accommodate a range of explanatory settings and to investigate differences between various types of explanation, multiple explanation types were considered in the different papers. The subsequent sections will examine three specific types of explanations: patient-doctor consultations, word learning, and game explanations. Notably, the latter category of game explanations will receive more extensive examination.

The following papers display various methods to shed light upon different aspects of the involvement by explainees. Empirical studies and subsequent discourse analyses of the recorded speech in ELAN (Wittenburg et al., 2006) formed the basis for the investigation. The analyses considered either the explainees' concrete verbal behaviour, ranging from different question types (see 4.1) to a broader range of speaker moves (see 4.3), the semantics (see 4.2) or the structural organisation of an explanation (see 4.4).

4.1 Explain with rather than explain to: How explainees shape their own learning (Fisher et al., 2024) published in Interaction Studies

In the article, we aimed to explore the explainees' different types of questions (Tare et al., 2011) in an explanation from a co-constructive standpoint (Rohlfing et al., 2021). Explanations as already noted are part of task-oriented dialogues, which are joint activities (Tenbrink et al., 2013) in which the interaction partners aim to come to a shared understanding (Clark, 1996; Quasthoff et al., 2017). One means to gain more information is question asking, and adults adapt their questions to be most effective (Ruggeri and Lombrozo, 2015). On an interactive level, when an interaction partner asks a question, the other is expected to answer. This is the case because a question belongs to the adjacency pair question and answer. If a question is posed, an answer is expected (Bieswanger and Becker, 2017).

In this paper, the involvement of explainees and how they actively guide the explanation process was captured by their question-asking. Afterwards, we examined in how far their type and number of questions relates to their learning. The participants in the presented study were 40 native English-speaking students (mean age 25 years) observed in dyadic explanation dialogues, i.e. 20 dyads were considered (21 women and 19 men). These dialogues are part of the ECOLANG Corpus (Gu et al., 2025). Each explanation included 24 stimuli objects (12 unknown, 12 known), which were chosen out of 36 objects in total. These objects were divided into four categories: tools, musical instruments, food, and animals. The study included three different stages. In the online training session, the explainer learnt about 12 unknown objects with video material prior to the dyadic interaction. Subsequently, in the explanation phase, the explainer provided a familiar partner (explainee) with information about the objects. The objects were presented in two conditions: present and absent. In advance, the explainee was informed about the upcoming recall test. Finally, during the recall task, the explainees' ability to label the unknown objects was assessed through a forced-choice test, where they had to determine whether a given picture and its accompanying label corresponded. Each correct choice was coded as 1, each incorrect choice as 0. The occurrence and type of the explainees' questions in the explanation session were explored and related to the explainees' subsequent recall score. The questions were categorised into six distinct groups and coded using ELAN (Wittenburg et al., 2006). The categories are mainly based on the ones presented in exhibition child-caregiver interactions (Tare et al., 2011) with some adjustments of the labelling and procedure questions. The number of different questions was correlated with the average recall task scores. We found that all explainees asked questions about the unknown objects. On average, explainees asked 29 questions (SD = 15.9, range 9–73). In the absent and present condition, factual and labelling questions were the most common question types. More function related questions were prompted in the object-present condition (M = 2.95, SD = 3.5) compared to the absent condition (M = 1.05, SD = 1.2). The other question types did not significantly differ between conditions. Concerning the results of the recall task, the explainees' individual performance in this task was calculated as the mean of their correct choices. The average score obtained was .8 (SD = .14), meaning that 80 percent of the labels were correctly identified. Only the question type of labelling in the present condition correlated positively with the recall score.

On a methodological level, we were able to apply and adapt a coding scheme developed for tutoring settings (Tare et al., 2011) to the context of explanations. This enabled a more fine-grained analysis of different question types by the explainee compared to (Chi et al., 2001, 2008), which in turn provided deeper insights into their verbal behaviour.

Therefore, the article provides first insights into the verbal involvement by the explainee in explanations by focussing on their question-asking behaviour. This constitutes a typical means to retrieve information (Ruggeri and Lombrozo, 2015) in an interaction and showcases how explainees can shape their own learning by involving themselves in the explanation.

We found that the explainees asked different types of questions, especially about the object's label and facts. The labelling questions were linked to better recall. In general, the presence of an object does not seem to trigger a better or different involvement of the explainee, except for function questions, which were asked more when objects were present. This indicates that the physical presence of the objects made their function-related properties more noticeable. To provide relevant and adaptive explanations, it is crucial to involve the explainee and to base the explanations on the users' questions. Therefore, the results contribute to the design of XAI, by suggesting that when explaining a system's functions, the presence of an object or referent might be helpful.

Overall, the task success in form of the recall task was related to the types and number of questions by the explainee. Thus, the empirical results that both interaction partners jointly co-construct the relevant information in the explanation support the co-constructive framework (Rohlfing et al., 2021). The results highlight the importance of actively involving the explainee in the explanation process, as they steer the interaction with their different questions.

4.2 Exploring the Semantic Dialogue Patterns of Explanations – a Case Study of Game Explanations (Fisher et al., 2023) published at SemDial in Proceedings of the 27th Workshop on the Semantics and Pragmatics of Dialogue

Not only is the question-asking behaviour of the explainee important, but also what they address semantically in an explanation. In the paper, we investigated several research questions related to the semantic patterns in explanations. A semantic pattern can be formed by the occurrence of different nodes in a particular order. Firstly, we expected sequential patterns comparable to the phases in direction giving (Psathas and Martin, 1976). To put it into other words, we expected the nodes to appear in a chronological order and, thereby, forming sequential patterns. Secondly, we expected the explainee to be an active participant (Rohlfing et al., 2021; Fisher et al., 2022), having the opportunity to introduce explanation nodes on their own. Thirdly, we expected the explanations to be co-constructive (Rohlfing et al., 2021; Clark and Schaefer, 1989).

The study included a subset of 51 dyadic game explanations of the game Quarto with a total of 102 participants. Of those, 60 were female, 38 male, and three diverse, with a mean age of 23. 96 participants were native German speakers and five were second language speakers. Lastly, 94 of them were students and seven had other occupations[1]. The explanations lasted on average 05:01 minutes. The semantic dialogue patterns were investigated using the annotation program ELAN (Wittenburg et al., 2006). In the node scheme development, we applied a data-driven adaptation from scientific explanations. Iteratively, the explanation nodes were added. This means that in the process, first the more general blocknodes (e.g. *Figures*) and then the more detailed subnodes (e.g. *Height*, *Form*, *Structure*) were established. We adjusted the level of detail to the topical occurrences in the data. Consequently, at the end of this process, it required the reanalysis of the data with the final version of the node coding scheme. For more details on nodes, see 3.1.

In the next section, the results regarding the order and sequentiality of the node are presented. The blocks *Game* and *Quarto* were rather mentioned at the beginning. *Board* and *Figures* were discussed subsequently, followed by the block *Goal* and *Turns*. The explanation typically ended with the blocks *Tips* and *Features* of the game. *Players* and *Comparisons* to other games occurred throughout the entire explanation. A cluster of certain blocks, explanation nodes connected to certain topics, such as *Figure* or *Goal*, tended to be explained closely together at about the same place in the explanation. Blocks such as *Players* or *Turns* were spread throughout the interaction. One reason might be that the node for *Turns* was mentioned several times and as a result the semantic connection to several other blocks is very strong. Another possibility is that the order of the blocks

[1]One data point each is missing due to technical problems.

differed in each dialogue, and consequently no strict clusters could be identified. We hypothesised that explanations, similarly to scientific explanations, also have sequential patterns. In scientific explanations, the nodes are ordered in a hierarchical and sequential manner (Chi et al., 2008; Miyake, 1986). In direction-giving certain sequential patterns occurred, which are phases that include specific information (Psathas and Martin, 1976). Game explanations are also sequential interactions, but the patterns are not as restrictive as in scientific explanations. Therefore, our hypothesis could only be partly verified for some explanation blocks, but not for all of them.

We also hypothesised the explainee to be actively involved in the co-construction of explanations. Our findings provide support, as we found that the explainee is active by demanding a more detailed explanation or pointing out knowledge gaps. While the explainer mentioned 49% on average (SD = 8.0) of the explanation nodes in their explanation, the explainee also addressed 20% of the explanation nodes on average (SD = 11). This results in a high variance in the explainee's verbal contributions. In particular, the explainees in Study 1 initiated 4.2 new nodes on average (min. 0, max. 10, SD = 2.6). This means that the explainee is the one addressing a node for the first time in an explanation, for example, by asking a question. In total, the explainees introduced 212 explanation nodes in the whole sample. 81% of these initiations were taken up by the explainer. This was done most often directly, 152 times, constituting 72%, indirectly only 19 times (9%) and without any uptake 41 times (19%). This shows that the explainee contributes meaningfully and actively to the content of the explanations, jointly constructing them, and our hypothesis was confirmed.

Evidence for the assumption that interaction partners are co-constructive is provided by the finding that the sequentiality and order of the nodes could not be clearly shown. Thereby, the content of an explanation depends on the interaction partners, and is therefore unique. This supports the assumption of the content of an interaction being a moving target (Rohlfing et al., 2021; Carletta et al., 1997; Klein, 2009), which in turn highlights the co-construction of explanations. Thus, our third hypothesis, expecting the explanations to be co-constructive, was also confirmed.

The establishment of the node scheme facilitated a nuanced semantic analysis of explanations. In more detail, the node coverage allowed for a precise mapping of the concrete verbal involvement of the explainee, the variety of nodes they addressed. Crucially, the scheme also enabled the assessment of the node uptake by the explainer of newly addressed nodes by the explainee. Consequently, it could be observed that the explainees actively take part in the explanation by addressing already mentioned nodes by the explainers or by addressing new nodes which the explainers take up. Thereby, both interaction partners co-construct the content of the explanation.

As such, the node scheme functions not merely as a descriptive tool, but as an analytical framework that elucidates: (1) the constellation of nodes covered during the interaction, (2) the temporal sequence of their emergence within the interaction history, and (3) the attribution of contributions to specific interaction partners.

4.3 Adaptive explanations: the role of explainees' substantive moves (Fisher and Rohlfing, 2025) submitted journal article in Text & Talk

As the involvement of the explainee could be evidenced in the previous paper, in the following their concrete verbal behaviour, speaker moves, was examined more closely. Little is known about in how far explainees are involved in an explanation. How do their substantive moves drive the uncovering of topics and influence the verbal behaviour of the explainer (Buhl et al., 2024)?

Thus, we intended to address the research gap concerning how the explainee, through their speaker moves and more specifically by their substantive input, plays an important role in shaping the explanation and contributes to its quality, which is defined as covered nodes. This contributes to exploring how the interactive adaptivity takes place. We explored whether the interaction partners adapt to one another and, thereby, jointly co-construct the explanation and its different elements. For this purpose, their types of speaker moves and coverage of semantic content were examined. We proposed that substantive moves of the explainee lead to more semantic content and adaptivity than non-substantive moves. Thus, the explanations gain a higher explanation quality, which is also mirrored in the verbal behaviour by the explainer in their diversity of moves and their nodes. In the analysis, we attempted to capture the interactive adaptivity of interaction partners, meaning which effect the speaker moves of the explainee have on the explainer. The verbal behaviour of interaction partners can be described by the total frequency of speaker moves, by the differentiation into the different types of moves, or by the range of different moves they use in an explanation. We operationalised adaptivity as the usage of different speaker moves as it captures how many speaker moves the interaction partners used to adapt towards their partner.

The ADEX game explanation corpus of Quarto includes video-recorded dyadic interactions. In the paper, two studies of the corpus were taken into account, the first being a semi-natural study and the second one being a control study. In the latter, Study 2, we regulated the explainees' behaviour to examine its impact on the explainer, investigating if the process of adaptation is impeded when one partner does not offer substantial verbal feedback. The studies had multiple sessions that included multiple questionnaires and assessment tools before and after the explanation phase. Before the explanation, the explainers were asked to familiarise themselves with the game. The first study consisted of 64 dyadic explanations, resulting in 128 participants. They had a mean age of 25 and consisted of 123 L1 and 5 L2 German speakers. 72 participants were female, 42 male, and 4 diverse. These interactions were audio and video recorded and lasted on average 5:30 minutes. In the second study, 41 game explanations were collected with 44 participants. The participants had a mean age of 24 years. Of the 44 total participants, 41 were explainers, and were paired with three pre-trained explainees. 41 participants were L1 German speakers, and three were L2 German speakers. 24 participants were female, 19 male, and one participant was diverse. The explanations lasted on average 2:30 minutes. The three trained explainees each participated in several explanations. In both studies, the explainers were instructed to explain the game in a manner that their explainee could win it. In Study 1, the explainees were encouraged to actively take part in the explanation, and in Study 2, they were instructed to only use neutral backchannels and to not make substantive contributions. A backchannel was being described as neutral when they only included continuers and no demonstration of changes in the knowledge construction, which would have been faked by the trained explainees because they knew the game already.

The coding of the verbal behaviour was also done in ELAN (Wittenburg et al., 2006). The analysis was carried out on multiple levels, namely: explanation nodes, speaker moves, and substantiveness (in the case of the explainee). The nodes were investigated with regard to their coverage and occurrence within the explanations for the speaker move analysis. The latter focused on the frequency and the different types of speaker moves, their diversity (different set of moves). The substantiveness analysis focused on the type of involvement by the explainee, if it was meaningful (substantive) or not (non-substantive).

The next section will briefly summarise the results of both studies. When focussing on the speaker moves by the explainee in Study 1, the most frequent moves (in order of their frequency) were: *factual questions*, *mentalising*, *paraphrasing partner*, and *additional info*. Notably, together these made up 84% of the substantive moves. Every explainee used substantive moves in the explanation and contributed meaningfully to the interaction.

Moreover, the substantive speaker moves led to a higher node coverage compared to non-substantive moves, meaning that more different types of nodes were addressed when more substantive moves were used. This is in line with our hypothesis. Turning to the results of the explainer, there was a significant positive correlation between move diversity and node coverage, indicating that a greater variety of moves contributes to covering a wider range of topics. This aligns with the observations made regarding the explainee. Furthermore, the primary factor influencing the joint node coverage was the number of speaker moves made by the explainer. This indicates that the more moves are used, the more nodes are being addressed.

In Study 2 the less engaged explainee led to significantly fewer jointly used speaker moves (the number of moves both interaction partners use), a lower joint node coverage (the number of nodes both interaction partners cover), and also a smaller move diversity. For the explainer alone, this was also the case, regarding their speaker moves (frequency and diversity) and covered nodes. These results taken together support the claim that the substantive involvement of the explainee is crucial in an explanation.

A special focus can be drawn towards the comparison of the node coverage in Study 1 and Study 2. Firstly, although the explainees only contributed to about 20.77% of the content in the explanations, their input has a significant impact on the explainer's presentation of various topics, suggesting that their role is more important than their relatively small contribution might suggest. In the second study, according to the design, the explainee barely covered any topics. With regard to the explainer, the study found that the explainer's coverage of explanation nodes decreased significantly from 48.69% to 36.97% when the explainee's verbal behaviour was controlled. This suggests that the explainer's ability to cover relevant topics was negatively impacted by the explainee's limited input. Furthermore, the joint node coverage, which measures the overall comprehensiveness of the explanation, also decreased significantly. These results indicate that the explainee plays a crucial role in shaping the explanation and that their involvement can lead to more comprehensive and detailed explanations. Overall, the study highlights the importance of the explainee's role for comprehensive explanations.

Regarding the implications and connections to the other papers in this work, we were able to combine the explanation nodes with the speaker move analysis and their substantiveness. Thus, the explanation node scheme serves as a tool to enable a thorough semantic analysis and lays the basis for determining the speaker moves which are used by the interaction partners. For example, the speaker moves providing *additional information* and *repeating* can only be differentiated when we can compare the utterances and their semantic content. If the information on a subnode is given for the first time, but there was already some information provided on the block, the move is *additional information.* If the information was already given and only a part is repeated in the same manner then in that case it is a *repeat.* Therefore, the nodes and speaker moves are tightly intertwined allowing for fine-grained differences between utterances and their explanation strategies to be examined. Additionally, we applied the explanation nodes to a larger data set of the ADEX corpus from 51 to 105 explanations. This included the entire first study and Study 2 consisting of 41 explanations. Therefore, we were able to apply the fine-grained analysis to a larger data set, proving the appropriate level of detail of our coding scheme.

Regarding the analysis of the verbal behaviour of the interaction partners, we enriched the speaker move analysis including different question types and going beyond those. Thereby, we advanced a detailed speaker move analysis. To describe the full picture of the verbal behaviour of the interaction partners, a coding scheme for the speaker moves was developed in a bottom-up and iterative manner (see 3.2 for more information). The most frequent move by the explainee in Study 1 was *factual questions.* This highlights that they are a common verbal means for the explainee to be involved in the explanation. These results support the approach proposed in Fisher et al. (2024) suggesting that a prototypical involvement of the explainee consists of asking questions. By advancing our coding scheme and adding several moves, we were also able to gain an even more fine-grained picture and examine the verbal behaviour of both interaction partners.

4.4 Exploring Monological and Dialogical Phases in Naturally Occurring Explanations (Fisher et al., 2022) in KI-Künstliche Intelligenz

To provide evidence for the active involvement of explainees in the explanation process, the following paper will be presented. It displays the role of monological and dialogical phases in explanatory processes and how the explainer and explainee negotiate the interaction dynamics in naturally occurring explanations. The explanandum of these consultations are the individual operations or treatments for the patients. This is a sensitive topic for the patient and legal guardians because it affects the health of the patient.

In this paper, we addressed the research questions of how the explainee is involved in an everyday explanation and how an explanation is structured. We argued in favour of two phases (more monological and more dialogical phase) in explanations. To investigate this, we audio- and videotaped eleven interactions between patient and physician taking place at a clinic for paediatric surgery in Germany. Therefore, we were able to collect naturally occurring explanations in a hospital setting. In total, eight doctors, ten patients, and 13 legal guardians took part in the study. The consultations lasted on average 9:30 minutes. In the presented study, we used conversation analysis (Selting and Kern, 2020) to code the phases into rather monological or dialogical phases based on the speech of the interaction partners (for more details see 2.6). In addition, the conversational jobs were examined in the paper. Our analysis showed that alternating monological and dialogical phases are central to the structure of naturally occurring explanations. We observed the following patterns in the initiation of phases: on the one hand, explainers tended to initiate monological phases more frequently than explainees. On the other hand, explainees tended to initiate dialogical phases more frequently. These results align with scientific and educational explanation patterns that consist of an initial monological phase in which the explainer provides information followed by a dialogical interactive phase in which both interaction partners actively participate (Kobayashi, 2022). Additionally, the conversational jobs occurred in both the monological and dialogical phases which implies that information given in the monological phase needs to be repeated in the dialogical phase to enable the explainee to connect it with their knowledge. This provides insights for the co-construction of explanations because both phases are necessary in explanatory sequences. With respect to the distribution of the phases, explanations appeared to predominantly include monological phases, on average 73.4 %. Dialogic phases occurred on average 26.6 % of the time and rather tended to be located towards the end of the explanations. The different phases in an explanation highlight the significance of explainees in the explanation process because they are mostly responsible for the dialogical phases. The most important implication for XAI is that the involvement of the explainee in an explanation needs to be possible throughout the interaction, especially more towards the end because explanations start with a monological phase followed by alternating dialogical and monological phases.

Therefore, it indicates that the explainees are actively involved in explanations through their phases and jobs. Describing the explanations and the explainee's involvement through the phases is done on a rather broad level, focussing on the overall structure of the explanation rather than on its specific content. The aim is to show that explainees are actively involved and may even initiate these phases, which we were able to do in this paper.

4.5 Complementary Analyses

Beyond the results in the previously discussed papers, some complementary analyses and their results concerning the explanation nodes and the speaker moves are presented in the following chapter. This includes more fine-grained analyses of the involvement of the explainee to examine more closely the adaptivity on a semantic and speaker move level. These include on the one hand the interaction partners semantic coverage of topics and how these topics are taken up by the explainee via their nodes and, on the other hand, their frequency and diversity of speaker moves. For this purpose, three different game explanation studies will be compared.

Regarding the results on the semantic level, the node initiations and which speaker moves are used for this by the explainee are reported. This is an extension of the investigation conducted in Fisher et al. (2023), expanding the analysis from 51 to 64 explanations and linking the speaker moves to the node initiation. In doing so, not only the frequency of the node initiations is considered, but also how those are introduced. This provides deeper insights into the interactive adaptivity by showcasing with which speaker move the explainee concretely introduces information, which is taken up in the following, thus steering the explanation content.

On average, explainees in Study 1 introduced 4.1 new nodes (SD = 2.5). In total, explainees brought up 256 new explanation nodes, and the explainer built upon 83% of these initiations. Most of the time, the explainer responded directly to the explainee's initiations (74%), while in some cases, the response was indirect (9% of the time), and in 17% of the cases, the explainer did not take up the initiation. The results do not differ greatly from those reported in Fisher et al. (2023), showing that the subset already provided a solid trend with respect to the node initiations. The uptake rate of the explanation nodes by the explainer of the explainee-introduced explanation nodes indicates that an explanation is a joint activity (Clark and Schaefer, 1989), in which both interaction partners are involved in the co-construction of the content (Rohlfing et al., 2021). The explainees can actively steer the explanation content via their initiation which are mostly taken up by the explainee. Thereby, this provides further information on how the interactive adaptivity takes place. The explainee introduces a node and the explainer takes it up.

In the following, the three moves that are most often used to initiate new nodes are listed. All moves used for a new node initiation are substantive. The majority of new nodes are addressed by asking *factual questions* that comprise 68% (173 occurrences) of the initiations. Another move that is used for the initiations is providing *additional information*, constituting 13% (34 occurrences) of the overall initiation moves. In the third place is *providing information* with 8% (21 occurrences), here the explanation block is mentioned for the first time.

The results show that the explainee frequently asks questions as a means to address new topics or knowledge gaps. Thus, it becomes visible that this type of speaker move plays an important role in explanations. That they in fact steer the interaction with this move is highlighted by the high uptake rate of the explainer (Fisher et al., 2023). This supports the special focus on question-asking in the previous section (4.1). Moreover, the explainee not only elicits further information by question-asking, but also provides information themselves. This highlights that the explainees also possess, to a certain extent, game knowledge and contribute this to the explanation.

Investigating the node initiation closely thus shows that not only can the nodes be used to differentiate the speaker moves, but also to examine who initiates new information. This is important because it describes more deeply how explainees are involved in explanation processes in human–human interactions.

In the next section, the node coverage, the speaker moves, and the move diversity by the explainee in Study 3 will be reported and compared to the previously mentioned studies, Study 1 and Study 2. This will be done in order to investigate the different conditions of the studies and how they impacted the explainee involvement. Moreover, the results of Study 2 and Study 3 display how well we were able to script and train an explainee and their involvement, their speaker moves, in an explanation ourselves.

Before reporting the results of the explainee, the study design of Study 3 needs to be briefly described. The study was designed to highlight the importance of active explainees in comparison to Study 2 and to highlight that we were able to capture the verbal behaviour of explainees from Study 1, their substantive behaviour, and train explainees to verbally behave repeatedly in that manner.

Study 3 included 43 dyadic game explanations with 47 participants with mostly German L1 speakers of which 43 were explainers and 4 pre-trained explainees, mean age 23 years, 15 male and 32 female participants. The explanations lasted on average 7:44 minutes.

To play the role of an active and co-constructive explainee, student assistants were trained to follow a script with speech instructions during the explanation. These instructions included predefined speaker moves, based on correlational results with cognitive dimensions (Buhl et al., 2024), namely *factual questions*, *summarising information*, *paraphrasing*, and giving an *example*, as well as neutral backchannels.

After having described Study 3, the next section continues by presenting the results of Study 3 and comparing them to the other two studies.

On a semantic level, in Study 3 the explainees covered on average 22.1% of the nodes (min. 7.2, max. 47.8, *SD* = 0.1). According to a one-way between-groups ANOVA (F(2, 145)= 76.618, $p < 0.001$), followed by an independent t-test ($t(45.087) = 13.656$, $p < 0.001$, $\eta^2 = 0.806$) which we conducted considering 0.53 nodes (min. 0, max. 11.6, *SD* = 1.8) from Study 2, this is a significant difference. Comparing the results of Study 3 with Study 1 using the same statistical method ($t(105) = .653$, $p = .515$, $\eta^2 = 0.004$), in which we considered the 20.8% (min. 2.9, max. 46.4, *SD* = 10.9) node coverage from Study 1, there is no significant difference. Consequently, the node coverage of the explainee in Study 1 (20.8%) and Study 3 (22.1%) were very similar.

Turning to the move frequency of the explainees using the same statistical approach, the same trend described for the nodes applies to the speaker moves. The explainees in Study 3 had an average of 15.6 moves (min. 8, max. 32, *SD* = 5.7). A one-way ANOVA between-groups (F(2, 145)= 76.618, $p < 0.001$) followed by an independent samples t-test ($t(45.4) = 16.494$, $p < 0.001$, $\eta^2 = 0.857$) revealed that the average of 1 speaker move (min. 0, max. 5, *SD* = 1.1) from Study 2 was significantly lower compared to Study 3. In contrast, comparing Study 3 to Study 1 ($t(99.206) = 1.137$, $p = .258$, $\eta^2 = 0,013$), which included 17.25 speaker moves (min. 3, max. 54, *SD* = 11.2) from Study 1, again showed no significant difference. Thus, the move frequency in Study 1 and Study 3 was very similar.

In more detail, the frequency of the most common speaker moves in Study 3 on average were: *factual questions* 7.2, *paraphrasing partner* 1.7, and *mentalising* 1.5. This mirrors the results from the most frequent explainee moves for Study 1, except for the order of *paraphrasing* and *mentalising* (see 4.3). *Mentalising* is the second most frequently used move in Study 1 and *paraphrasing partner* the second most frequent move in Study 3. The results of Study 3 show that the explainees mostly made use of their moves in accordance with their pre-trained speaker moves. The only exception was the move of *mentalising* which was used slightly more often than *summarising information* on the fourth place regarding their frequency.

Consequently, one could propose that the moves *factual question*, *paraphrasing partner* and *mentalising* form a repertoire, core moves, that explainees use in game explanations of Quarto. The repeated occurrence of those moves in both studies makes them appear to form a pattern (set of speaker moves).

Finally, coming to the move diversity of the explainees in the different studies. As noted earlier, this refers to the number of different moves which were being used in the explanations by the explainees. In Study 3, the explainees used on average 5.3 different moves (min. 5, max. 9. *SD*= 1,58). Performing a one-way, between-groups ANOVA followed by an independent samples t-test ($t(75.916) = 15.416$, $p < 0.001$, $\eta^2 = 0.758$) revealed that the move diversity of 0.8 on average (min. 0, max. 2, *SD* = 0.7) from Study 2 was significantly lower compared to Study 3. In contrast to this result, comparing Study 3 with Study 1 (t $(104.259) = 1.788$, $p = .077$, $\eta^2 = 0.03$), which included a move diversity of 6 (min. 2, max. 12, *SD* = 2.6) from Study 1, no significant differences between the studies were found.

Therefore, as there are no significant differences between Study 1 and Study 3, the results show that we were able to very closely script the explainee's verbal behaviour and involvement in an explanation on a speaker move (frequency and diversity) and content level (nodes). The empirical data can serve as training material for adaptive XAI. This allows XAI systems to model an explainer that adapts to an "average explainee" which includes their verbal behaviour on a semantic and speaker move level. Thereby, it is important to consider empirical results of human–human explanations as a starting point for human–machine interactions in which the explainee is being considered more.

The differences between Study 1 and Study 2, as well as Study 3 and Study 2 highlight the similarities of the two studies (Study 1 and Study 3). The similarities of the studies regarding the verbal behaviour of the explainee can be accounted for by the speaker move frequency and their move diversity. We cannot state whether the temporal placement of the moves is the same. It is possible that the trained explainees in Study 3 get involved later in the explanation because they are more patient since they know the game and that the explainees in the semi-natural setting in Study 1 get involved earlier in the dialogue.

With regard to the most frequent explainee moves, some patterns were found which can be useful for the modelling of explainee involvement in XAI.

5 Synthesis and Overarching Discussion

In the following section, an extended discussion of the synthesis and results of the four main papers and the complementary analyses will be presented. The synthesis elaborates how individual studies contribute to addressing the overarching research question of how the explainee is involved in an explanation. To investigate this matter, we analysed the verbal behaviour of the explainees on different levels: on a speaker move level, structural level, and semantic level.

On the speaker move level, first considering only the question types, in the ECOLANG dataset, we found that word learning is connected to labelling questions an explainee asks (Fisher et al., 2024). Therefore, word learning relates to the number of questions asked by the explainee. This supports the claim by Ruggeri and Lombrozo (2015) that the questions are adapted to elicit specific information. Methodologically, we were able to adapt the coding scheme of question types from the tutoring context (Tare et al., 2011) to explanations and examine their effect on learning.

In the ADEX corpus, consisting of game explanations, we examined, on a semantic level, how both interaction partners contribute to the explanation (Fisher et al., 2023). Therefore, we developed an explanation node coding scheme which was based on our data. This allowed for a detailed semantic analysis, in which we found that not only the explainers initiate new topics but, the explainees as well. Consequently, we captured which topics were addressed by whom. Our results show that both interaction partners attempt to come to a common ground via their shared information (Clark and Schaefer, 1989). Furthermore, the content of the explanation and its temporal placement in form of the sequentiality and order of the nodes was not evident. We expected the nodes to occur in a certain order. This goes in line with the claim that the content is a "moving target" in an interaction (Rohlfing et al., 2021; Carletta et al., 1997; Klein, 2009).

Extending the analysis of the speaker moves beyond questions, we proceeded to examine in how far they are substantive (Fisher and Rohlfing, 2025). In other words, how do they contribute meaningfully to the explanation? To do so, we took tutoring as a starting point, as it also deals with asymmetric interactions. From this, the concept of substantiveness (Chi et al., 2008) was taken to differentiate more gradually between the different types of involvement by the explainee. In this paper (Fisher and Rohlfing, 2025), we could combine the methods of the first two papers (Fisher et al., 2024, 2023), coding schemes of different questions and nodes, as they lay the basis for the thorough speaker move analysis. Discussing these results in more detail, they reveal how the conversational partners, explainer and explainee, contribute to the co-construction (Rohlfing et al., 2021) of explanations through their verbal behaviour. Concretely, the speaker moves and their different types, especially the substantive ones, play an important role in eliciting a higher joint node coverage, number of joint speaker moves, and move diversity. The substantive moves heighten the number of covered nodes, thereby, the quality of the explanation (Fisher and Rohlfing, 2025). Study 2, where the explainee was only involved by making neutral backchannels and no substantive moves, showed that the explainer was influenced by less active explainees. This led to a significant decrease in the frequency and variety of the speaker moves. Thus, our results extend what was found for substantive moves in tutoring settings (Chi et al., 2001, 2008) to explanations. The fact that those substantive moves and their absence in Study 2 influenced the verbal behaviour of the explainer demonstrates their adaptation towards the interaction partner. Thereby, for the first time, it provides experimental evidence for the interactive adaptivity in explanations.

Focussing on the different types of phases within an explanation, the alternation of monological and dialogical phases, provide further support that the explainee is involved in an explanation (Fisher et al., 2022). The alternating monological and dialogical phases display the co-construction (Rohlfing et al., 2021) of the explainer and the explainee, which shapes the explanatory process. Additionally, the conversational jobs occurred in both the monological and dialogical phases. This shows that the active involvement by the explainees is crucial for the co-construction of the dialogue and not a single explanation works without it.

Within the complementary analyses, the node uptake by the explainer of newly initiated nodes by the explainee in Study 1 was also found when taking into account the whole sample (N= 64) compared to the subset (N=51) used in Fisher et al. (2023). The moves used for the initiation are asking *factual questions*, providing *additional information*, and *providing information*. This highlights that both interaction partners jointly take part (Clark and Schaefer, 1989) in the co-construction of the explanation content (Rohlfing et al., 2021). In Study 3, we scripted an explainee to use substantive speaker moves. When comparing the explainees from Study 3 with those from Study 1, no significant differences were observed in either the frequency or diversity of node usage or speaker moves. Thereby, we were able in Study 3 to script an explainee similar to Study 1. These findings provide an empirical foundation for modelling human explainees in Explainable AI (XAI) systems. In order to have a system that adapts towards the explainee, it is important to account for substantive contributions. Therefore, the feedback by the explainees need to be possible in the form of full utterances or sentences, which can be analysed regarding their substantiveness. Additionally, the system needs to be dynamic by allowing those contributions by the explainee throughout the explanation. Thus, empirical data of human–human explanations can be used to model XAI systems that adapt to an "average explainee" based on data.

The findings taken together indicate that the interaction partners adapt towards each other. The node scheme provides a quantitative framework to investigate the content of an explanation by analysing both when and by whom the nodes were addressed. This supported the detailed analysis of the speaker moves and their substantiveness. Having a substantiveness score provides a quantitive measure for the verbal involvement of the explainee. On a rather structural level, the different phases in an explanation offered further insights on the co-construction of explanations and the involvement of explainees. Overall, we highlighted the impact of the verbal behaviour of the explainee on the explainer. Consequently, the explainer tailors the explanation to the explainee. Therefore, we were able to prove our concept of interactive adaptivity because we were able to show how the interaction partners verbally adapt towards each other on a speaker move level, structural level, and semantic level. At the core of the interactive adaptivity, we see that the verbal behaviour of the explainee has an effect on the explainer, and thus this is one way for the explainee to steer the interaction.

The results taken together lead to a more detailed understanding of the explainee's involvement in co-constructing explanations for the presented datasets. The following points highlight how the different methodological developments allowed for such a detailed examination. Firstly, the different types of questions extended the speaker move analysis by Chi et al. (2001, 2004, 2008). Secondly, the node scheme established in an iterative and data-driven way focused on the semantic content the interaction partners addressed jointly and not only the explainer alone, unlike other node schemes from tutoring literature (Miyake, 1986; Chi et al., 2008). Thirdly, the speaker moves and their substantiveness were examined in different studies, demonstrating their impact. Moreover, the speaker move analysis by Chi et al. (2001, 2004, 2008) has been extended by viewing the individual speaker move categories as being available to both interaction partners. Finally, on a structural level, the involvement of the explainee was shown in connection to the accomplishment of certain conversational jobs.

While this dissertation provides valuable insights into the involvement of the explainee in explanations, there are several limitations. One key limitation in the new word learning study (Fisher et al., 2024) is that the explainees' prior knowledge of the recall task may have influenced their engagement and question-asking behaviour during the explanation, potentially impacting the validity of our findings. The explainees possibly pay special attention towards the labels of the objects and therefore ask more questions regarding their label to make sure they understand it correctly.

Additionally, our analysis of the semantic content of an explanation was limited to the explanation node level (Fisher et al., 2023), and we did not fully capture the complexity of the interlocutors' co-construction of meaning. One could have paid more attention towards the lexical alignment of the interaction partners beyond the speaker moves *repeating* and *paraphrasing*. Thereby, one could analyse all occurrences of lexical alignment and examine when and on what words the interaction partners align. To investigate more closely when interaction partners align lexically would provide further insights on where the adaptation takes place in the explanation.

Moreover, the lack of sequential patterns in the explanation on the explanation node level may be due to the size of the dataset or the nature of the explanations themselves. Future studies should aim to collect larger datasets and develop more sophisticated analytical techniques to uncover the nuances of explanatory dialogue.

Furthermore, our analysis did not account for the potential cognitive changes that may occur in the partner models of explainers as a result of the explainee's substantive contributions. Consequently, also the cognitive adaptivity could influence the number of substantive contributions (Buhl et al., 2024). Drawing on the research by Buhl et al. (2024) showing that certain speaker moves, such as *factual questions*, *summarising*, and *paraphrasing partner*, lead to changes within the partner models, this matter should be investigated in more detail. Future research should therefore explore the interplay between partner models and substantive speaker moves, addressing questions such as: How are the partner models connected to the verbal behaviour of the interaction partners?

Moreover, the papers presented here focus on specific everyday contexts, namely word learning, game explanations, and medical consultations. This may limit the generalisability of our results to other explanatory settings. The order and sequentiality known from scientific explanations (Psathas and Martin, 1976) might also be found in other everyday explanations. Additionally, their semantic analysis could reveal a hierarchical node scheme with a fixed order (Chi et al., 2008; Miyake, 1986). Therefore, one might not be able to apply the results from everyday explanations to scientific explanations.

Finally, it is essential to acknowledge some methodological limitations. Though we transcribed the backchannels provided by the explainees, we did not further analyse their function. However, they might contribute to the explanation and provide further insight into the way explanations unfold.

Despite these limitations, this dissertation provides a foundation for future research on explanations and highlights the importance of considering the complex interplay between the explainer, explainee, and explanation context.

6 Conclusion and Outlook

In the following, the key aspects of the involvement of the explainee in explanations are recapped. Subsequently, an outlook is provided with perspectives for future work.

The explanation nodes, the speaker moves and their substantiveness as well as the phases (monological and dialogical) are central for the operationalisation of the involvement by the explainee in explanations. Thus, we were able to demonstrate in how far the verbal behaviour of the explainee influences the verbal behaviour of the explainer. This provides answers to the general research question in discourse analysis of how the verbal behaviour of an interaction partner relates to another (Ferreira, 2024). This dissertation goes beyond the qualitative description of the involvement of the explainee, as previously done in Conversation analysis (Quasthoff et al., 2017; Morek et al., 2017; Gosen et al., 2025), by providing several quantitive measures for the explainee involvement in explanations.

Not only does the verbal behaviour of the explainee affect their interaction partner, but also their own learning. Through their questions, explainees can shape their own learning by asking for the information they require in the explanation. Taking into account the previously mentioned limitations of this work, in the future other forms of learning, other than word learning, should be tested in relation to different question types (Fisher et al., 2024) and other speaker moves, as they might also lead to learning. Thereby, the verbal behaviour would be better linked to different learning outcomes. In this investigation, the impact of the presence of the explanandum in explanations can also be examined. As it influences the success of word learning (Fisher et al., 2024), it needs to be extended to further learning assessments.

Additionally, the explanation nodes can also be of interest in relation to learning. Future studies could relate the frequency and coverage of the explanation nodes to the processes of understanding. Do the nodes need to be repeated several times or does a certain amount of coverage of nodes foster understanding? Also, the type of speaker move might demonstrate a different state and form of understanding. There is a difference whether one asks a question, paraphrases what has been previously said, or even provides information. This is not captured when solely considering the nodes. Within these different assessments on a verbal level, not only the speaker moves reveal understanding in form of nodes that they address, but also the backchannels could signal understanding. Therefore, backchannels should be included in future analyses. Consequently, understanding measurements would be needed to supplement the analysis.

Moreover, to deepen the investigation of lexical alignment, one could take into account all lexical alignments within the speech of the interaction partners. Thus going beyond the speaker moves of *repeat* and *paraphrase*. This would add a further measure of adaptation in form of lexical alignment. This can be connected to the measurement of shared lexicons by (Dubuisson Duplessis et al., 2021) in human–human and human–agent interactions which can be adjusted in the subsequent step for XAI. It would offer a more detailed analysis of lexical alignment across the speech of both interaction partners. This might offer valuable insights on which interaction partner aligns to whom and could be implemented in XAI systems.

To better understand the speaker moves, future research is needed. A question that remains open is: What effect does the temporal placement of speaker moves have? They might have a different effect depending on the order in which they appear. Therefore, the development of innovative methods to capture the sequentiality of moves is needed.

Several questions also remain unanswered in regard to the interplay of the different concepts related to the involvement of the explainee, namely explanation nodes, speaker moves, and phases. In how far do the number of covered nodes and the diversity and frequency of moves depend on the phases in which they occur. To investigate this matter more closely, the game explanations would need to be analysed according to their phases and related to the other levels of analysis.

Moreover, though the nodes were used as a basis to identify the speaker moves in the presented papers, it was not displayed which speaker moves are used to convey certain nodes. If one were to do so, one could, for example, make a statement about which nodes the explainees ask factual questions. In return, this would provide more detailed insights into which and how information is retrieved. This analysis was already performed for Study 1 in (Fisher and Terfloth, 2025), but could be extended to the other studies in order to examine more thoroughly the interaction dynamics in the different studies.

Additionally, one could also investigate whether the density of information is higher in monological phases than in dialogical phases. Do the dialogical phases serve as a repair for knowledge gaps? Are there maybe aspects that are specific to a certain phase or are there trends that certain moves rather occur in monological or dialogical phases? These questions will be further examined in the A01 project, *Adaptive explanation generation*, of the TRR 318 "Constructing Explainability". The next analyses we are currently working on concern the phases in game explanations, which we have so far only investigated in the medical domain. As previously mentioned, also the cognitive changes of interaction partners might have an effect on the substantive contribution of explainees. To examine this in the future, psychological data is needed, which is also one of the research aims of the project A01.

One of the main objectives of the project A01 is to develop adaptive XAI. Taken together, our results suggest practical applications for improving explanatory processes, and these can be applied to XAI. The results highlight the need in the development of XAI to make the involvement of the explainee flexibly possible and as important as the explainer's. Understanding and capturing the involvement of explainees in human–human interactions provides insights into how human–machine interactions should be interactively designed so that the explainee becomes actively involved in the explanation process. To increase their learning, they need to be able to be involved by introducing new topics and making substantive moves.

7 Declaration

In the following, the contributions of the authors for the publications displayed in Chapter 4 will be listed.

For Fisher et al. (2024) the authors were: Josephine B. Fisher (JF), Katharina J. Rohlfing (KR), ED Donnellan (ED), Angela Grimminger (AG), Yan Gu (YG) and Gabriella Vigliocco (GV). The contributions were the following: conceptualisation, AG, ED, GV, JF, KR; data acquisition: ED, GV and YG; method, JF and KR; data analysis, JF; original draft preparation: AG, JF, and KR; supervision, GV and KR; funding acquisition, GV and KR.

For Fisher et al. (2023) the authors were: Josephine B. Fisher (JF), Amelie S. Robrecht (AR), Stefan Kopp (SK) and Katharina J. Rohlfing (KR). The contributions were the following: conceptualisation, all authors; method, JF; data acquisition, JF; data analysis, AR and JF; original draft preparation, AR and JF; supervision, KR and SK; funding acquisition, KR and SK.

For Fisher and Rohlfing (2025) the authors were: Jospehine B. Fisher (JF) and Katharina J. Rohlfing (KR). The contributions were the following: conceptualisation, all authors; method, JF and KR; data acquisition, JF; data analysis, JF; original draft preparation, JF; supervision, KR; funding acquisition, KR.

In Fisher et al. (2022) the authors were: Josephine B. Fisher (JF), Vivien Lohmer (VL), Friederike Kern (FK), Winfried Barthlen (WB), Sebastian Gaus (SG) and Katharina J. Rohlfing (KR). The contributions were the following: conceptualisation, all authors; method, JF and KR; data acquisition, JF; data analysis, FK, JF, VL; original draft preparation, JF, VL; supervision, FK, KR; funding acquisition, FK, KR. SG and WB SG contributed to the patient and data acquisition and discussion.

8 Acknowledgements

The work in this dissertation was funded by the Deutsche Forschungsgemeinschaft (DFG): TRR 318/1 2021 438445824, project A01. I thank Viviane Gladow, Celina Nitschke, Marco Verwiebe, and Britta Meyer for supporting the data collection.

Special thanks go to my supervisors Prof. Dr. Katharina J. Rohlfing and Prof. Dr. Anke Lenzing, who made my dissertation possible. They supported me in my work in various ways by discussing my work, and developing methods and concepts together. Also, they guided me through the academic system and taught me how to write and publish papers. They helped me with any insecurities and worries throughout the years, being there for me whatever I needed. You are my scientific role models and are showing me how to become an inspiring researcher.

I am also very grateful to the other members of the disputation committee because Prof. Dr. Ilka Mindt and Dr. Angela Grimminger have taught me a lot during my studies and supported me in finishing my PhD. Prof. Dr. Ilka Mindt has supported my career in the department of English Linguistics by employing me in different areas and teaching me a lot during my studies. Dr. Angela Grimminger supported me in many lab-related questions and how to handle communication with other researchers.

In general, I would like to thank my lecturers during my course of studies in the department of English Linguistics and English Literary and Cultural Studies at the Paderborn University. They made me believe that I could become a researcher and, more importantly, that the topics that I am interested in actually matter. Just to name a few I would like to mention: Prof. Dr. Anke Lenzing, Prof. Dr. Ilka Mindt, Prof. Dr. Merle Tönnies, Prof. Christian Langstrof, and Dr. Alexandra Hartmann. I would also like to thank especially Dr. Emilila Nottbeck for sparking my passion for Psycholinguistics in my first bachelor semester!

I also would like to thank my colleagues and friends who supported me significantly. These were Vivien Lohmer and Marie Weise. They supported me through all stages of my PhD by reading my paper drafts, rehearsing for conferences, and supporting our data collections. Also personally, both of you were always there for me when I needed a friend. Thank you for always looking out for me and going out with me when I needed a distraction. The support of all these strong women made it possible for me to complete my dissertation. Of course, I am also grateful for all the colleagues in the TRR 318 "Constructing Explainability" who supported me in my work. Special thanks go to Dr. Ronja Hannebohm and Kai Biermeier. Ronja supported me with many writing sessions and writing days. Kai was always there for technical questions and study support. Thereby, they truly provided an interdisciplinary network!

Last but not least, I would like to thank my family. Especially my parents who made it possible that I could study what I love and always believed in me!

Bibliography

Allen, J. and Core, M. (1997). Draft of damsl: Dialog act markup in several layers. Working paper, University of Rochester, Rochester, NY.

Besold, T. R. and Uckelman, S. L. (2018). The what, the why, and the how of artificial explanations in automated decision-making. *CoRR*, abs/1808.07074.

Bieswanger, M. and Becker, A. (2017). *Introduction to English linguistics*, volume 2752. UTB.

Bressler, C. (2023). *Lehrende und Lernende: eine asymmetrische Beziehung: Eine rekonstruktive Studie zu Erfahrungen und habitualisierten Orientierungen von Lehrpersonen*. Verlag Julius Klinkhardt.

Buhl, H., Fisher, J. B., and Rohlfing, K. J. (2024). Changes in partner models effects of adaptivity in the course of explanations. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 46, pages 4976–4983.

Buhl, H. M., Wrede, B., Fisher, J. B., and Matarese, M. (2025). Adaptation. In Rohlfing, K. J., Främling, K., Thommes, K., Alpsancar, S., and Lim, B. Y., editors, *Handbook of social Explainable AI*. Springer. published soon.

Carletta, J., Isard, A., Isard, S., Kowtko, J. C., Doherty-Sneddon, G., and Anderson, A. H. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1):73–105.

Chi, M. T., Roy, M., and Hausmann, R. G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32(2):301–341.

Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., and Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25(4):471–533.

Chi, M. T. H., Siler, S. A., and Jeong, H. (2004). Can tutors monitor students' understanding accurately? *Cognition and Instruction*, 22(3):363–387.

Clark, H. H. (1996). Common ground. In *Using Language*, pages 92–121. Camebridge Usiversity Press, Cambridge UK.

Clark, H. H. and Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13(2):259–294.

Dressler, R. A., Buder, E. H., and Cannito, M. P. (2009). Rhythmic patterns during conversational repairs in speakers with aphasia. *Aphasiology*, 23(6):731–748.

Dubuisson Duplessis, G., Langlet, C., Clavel, C., and Landragin, F. (2021). Towards alignment strategies in human-agent interactions based on measures of lexical repetitions. *Language Resources & Evaluation*, 55(2):353–388.

Ferreira, F. (2024). *Psycholinguistics: A very short introduction*. Oxford University Press.

Fischer, K. (2016). *Designing speech for a recipient*. John Benjamins Publishing Company.

Fisher, J. B., Lohmer, V., Kern, F., Barthlen, W., Gaus, S., and Rohlfing, K. J. (2022). Exploring monological and dialogical phases in naturally occurring explanations. *KI-Künstliche Intelligenz*, 36:317–326.

Fisher, J. B., Robrecht, A. S., Kopp, S., and Rohlfing, K. J. (2023). Exploring the semantic dialogue patterns of explanations–a case study of game explanations. In *Proceedings of the 27th Workshop on the Semantics and Pragmatics of Dialogue*, pages 35–46.

Fisher, J. B. and Rohlfing, K. J. (2025). Adaptive explanations: the role of explainees' substantive moves. Manuscript submitted for publication.

Fisher, J. B., Rohlfing, K. J., Donellan, E., Grimminger, A., Gu, Y., and Vigliocco, G. (2024). Explain with, rather than how to: How explainees shape their own learning. *Interaction Studies*, 25(2):244–255.

Fisher, J. B. and Terfloth, L. (2025). The dual nature as a local context to explore verbal behaviour in game explanations. In *Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue*, pages 246–248.

Fusaroli, R., Raczaszek-Leonardi, J., Tylén, and Kristian (2014). Dialog as interpersonal synergy. *New Ideas Psychology*, 32:147–157.

Garrod, S. and Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, 8(1):8–11.

Gisladottir, R. S., Chwilla, D. J., and Levinson, S. C. (2015). Conversation electrified: Erp correlates of speech act recognition in underspecified utterances. *PloS one*, 10(3):e0120068.

Goodwin, C. (2003). Conversational frameworks for the accomplishment of meaning in aphasia. *Conversation and brain damage*, 1:90–116.

Gosen, M. N., Willemsen, A., and Hiddink, F. (2025). Applying conversation analysis to classroom interactions: students'‘oh'-prefaced utterances and the interactional management of explanations. *Classroom Discourse*, pages 1–19.

Gu, Y., Donnellan, E., Grzyb, B., Brekelmans, G., Murgiano, M., Brieke, R., Perniss, P., and Vigliocco, G. (2025). The ecolang multimodal corpus of adult-child and adult-adult language. *Scientific Data*, 12(1):89.

Hansen, M.-B. M. (2025). Social meaning as hearer's meaning: Integrating social meaning into a general theory of meaning in communication. *Journal of Pragmatics*, 241:81–91.

Hu, J. and Degand, L. (2023). The conversational discourse unit: Identification and its role in conversational turn-taking management. *Dialogue & Discourse*, 14(2):83–112.

Kalina, C. and Powell, K. (2009). Cognitive and social constructivism: Developing tools for an effective classroom. *Education*, 130(2):241–250.

Klein, J. (2009). Erklären-was, erklären-wie, erklären-warum. typologie und komplexität zentraler akte der welterschließung. *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven*, 2:25–36.

Kobayashi, K. (2022). Learning by teaching face-to-face: the contributions of preparing-to-teach, initial-explanation, and interaction phases. *European Journal of Psychology of Education*, 37(2):551–566.

Kotthoff, H. (2009). Erklärende aktivitätstypen in alltags-und unterrichtskontexten. *Erklären im Kontext. Neue Perspektiven aus der Gesprächs-und Unterrichtsforschung*, pages 120–146.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences*, 20(1):6–14.

Levinson, S. C. and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6:731.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38.

Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science*, 10(2):151–177.

Morek, M., Heller, V., and Quasthoff, U. (2017). Erklären und argumentieren. In Meißner, I. and Wyss, E. L., editors, *Begründen–Erklären–Argumentieren: Konzepte und Modellierungen in der Angewandten Linguistik*. Tübingen: Stauffenburg.

Pickering, M. J. and Garrod, S. (2004a). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Science*, 27(2):212–225.

Pickering, M. J. and Garrod, S. (2004b). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(2):169–190.

Psathas, G. and Martin, K. (1976). The structure of directions. *Semiotica*, 17(2):111–130.

Quasthoff, U., Heller, V., and Morek, M. (2017). On the sequential organization and genre-orientation of discourse units in interaction: An analytic framework. *Discourse Studies*, 19(1):84–110.

Reitter, D. and Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, 76:29–46.

Rohlfing, K. J., Cimiano, P., Scharlau, I., Matzner, T., Buhl, H. M., Buschmeier, H., Esposito, E., Grimminger, A., Hammer, B., Häb-Umbach, R., et al. (2021). Explanation as a social practice: Toward a conceptual framework for the social design of ai systems. *IEEE Transactions on Cognitive and Developmental Systems*, 13(3):717–728.

Rohlfing, K. J., Wrede, B., Vollmer, A.-L., and Oudeyer, P.-Y. (2016). An alternative to mapping a word onto a concept in language acquisition: Pragmatic frames. *Frontiers in Psychology*, 7:470.

Ruggeri, A. and Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, 143:203–216.

Sacks, H., Scheglof, E., and Jeferson Gail (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

Searle, J. R. (2002). Speech acts, mind, and social reality. In *Speech acts, mind, and social reality: Discussions with John R. Searle*, pages 3–16. Springer.

Selting, M. and Kern, F. (2020). Conversation analysis and interactional linguistics. *The Encyclopedia of Applied Linguistics*, pages 270 – 275.

Sokol, K. and Flach, P. (2020). Explainability fact sheets: a framework for systematic assessment of explainable approaches. *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 56–67.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., Yoon, K.-E., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26):10587–10592.

Stukenbrock, A. (2013). Sprachliche interaktion. In *Sprachwissenschaft: Grammatik—Interaktion—Kognition*, pages 217–259. Springer.

Tare, M., French, J., Frazier, B. N., Diamond, J., and Evans, E. M. (2011). Explanatory parent–child conversation predominates at an evolution exhibit. *Science Education*, 95(4):720–744.

Tenbrink, T. (2020). *Cognitive Discourse Analysis: An Introduction.* Cambridge University Press.

Tenbrink, T., Eberhard, K., Shi, H., Scheutz, M., et al. (2013). Annotation of negotiation processes in joint-action dialogues. *Dialogue & Discourse*, 4(2):185–214.

Traxler, M. J. (2012). *Introduction to psycholinguistics: Understanding language science.* John Wiley & Sons.

Wilkinson, R., Beeke, S., and Maxim, J. (2003). Adapting to conversation. *Conversation and brain damage*, pages 59–89.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). Elan: A professional framework for multimodality research. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1556–1559.

9 Appendix

The following appendix includes the coding scheme of speaker moves 3.2 and the original papers in Chapter 4.

Preliminary ADEX Coding Scheme for Speaker Moves 01.01.2024

Speaker Moves

Following the work by Chi [1] moves are defined as statements including a single idea presented by a single speaker within one turn. In a data driven approach the speaker moves are coded on three different levels. On the content level, on the rhetorical level and on the discourse management level. Each move takes place on the content and rhetorical level but not necessarily on the discourse management level. In the table, the various moves are provided with definitions and examples. In order to better classify the content of the moves, the conversation history (topics covered) is also collected via the game nodes.

Speaker Move	***Definition***	***Examples***
Content level		
providing info [2]	Introduction of topic at the start of each explanation element. Game node is mentioned for the first time. When a node is introduced for the first time in a question by the EE, then the answer to that question is considered as providing info. The information is game specific.	*EX: So you have sixteen pieces. (01)** *EX: Eh you have to put four pieces with the same feature in a row. (50)* *EX: There is eh a game surface that is like square. (28)*
providing personal info [5]	Includes individual preferences and experiences of the explainer and explainee.	*EE: Often Pictionary or Trivial Pursuit. (01)*
providing gesture	Gesture transports the information without verbal supplementation.	*EE shows on the table the possible formations of a row. (17)*
summarising info [3]**	Previous information is summarised. At least two nodes need to be addressed. It is still considered to be a summary even if a new blocknode is introduced, as long as the subnodes of this node are familiar.	*EE: Categories were light brown dark brown with hole without hole tall and small pieces (Nodes: 5.6.2., 5.6.1., 5.5.1., 5.5.2., 5.3.1. & 5.3.2.) (16)*
additional info [2]	Reasoning of actions, numbers, definitions, names, facts and contextualization. Node is already mentioned. The subnode might be mentioned for the first time within the explanation. Can also include additional info on the subnode. When a subnode is introduced for the first time in a question by the EE, then the answer to that question is considered as additional info.	*EX: And ehm there are two shapes one square and one round. (07)* *EX: And they are round or eh square the game pieces. (32)* *EX: You can place it wherever you want as long as that square is not occupied yet. (45)*
example [2]**	Providing an example. In games often game situations are provided. Subnode has been previously mentioned.	*EX: You can for example count for small pieces. (23)*
mentalising	Formulating a wish, goal or emotional state. Can also include agreement or an update within the knowledge state.	*EE: No it sounds like I would be able to do it now. (57)*
repeating [2] ***repeating_self*** ***repeating_partner*** ***repeating_both***	Information is repeated in the same manner as previously mentioned by the speaker themselves, their conversational partner or both. This also includes utterances with a different word order as long as the choice of words remains the same.	*EE: Yes so a row of four.* *EX: exactly a row of four.* *EE: A row of four. (09)*
paraphrasing [2] ***paraphrasing_self*** ***paraphrasing_partner*** ***paraphrasing_both***	Information is put into different words than previously mentioned by the speaker themselves, their conversational partner or both. A paraphrase has to be connected to a previous part of the conversation. Answers to questions with familiar nodes are also considered paraphrases.	*EX: you do not have only two colors* *EX: so it is not that you just need to have four of the same color in a row there are more categories.* *EX: there a two different colors that is one of the categories (02)*

focus monitoring [4]	The speaker repeats an aspect from the previous utterance of their speaking partner. The focus of the explanations is drawn to it.	*EX: That each have three attributes either they are tall small they are round square or they have a hole in the middle or they are just solid* *EE: Solid that eh. (07)*
Question Type		
Content level		
labeling [5]	Target the name of an explanandum in an oral or orthographic form.	*EE: And it is called Quarto? (22)*
factual [5]	Seek information of a game node. Include general demands for more information and examples.	*EE: So everybody has there are sixteen figures for both players? (49)*
reassurance [4]	Make sure the conversational partner is serious, a sign of disbelief or surprise or auditory difficulties.	*EE: Yeah? (19)* *EE: Really? (51)* *EX: Really not? (01)*
personal [5]	Includes individual preferences and experiences of the explainer and explainee.	*EE: I don't know do you know Nine Men's Morris? (16)*
procedure [5]	Is an off-topic remark on. Also includes information on the pre-session of the explainer.	*EE: Am I allowed to ask you a question? (49)*
comprehension [2]**	how to continue with the dialogue	*EX: exactly eh do you have any questions about that or did I explain it clearly? (26)*

* English translation of the German transcripts.

** according to literature tutor move, for us a speaker moves, can be used by both EX and EE

References

[1] Chi, M. T., Roy, M., & Hausmann, R. G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32(2), 301-341.

[2] Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive science*, 25(4), 471-533.

[3] Chi, M. T. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied cognitive psychology*, 10(7), 33-49.

[4] Data-driven categories.

[5] Tare, M., French, J., Frazier, B. N., Diamond, J., & Evans, E. M. (2011). Explanatory parent–child conversation predominates at an evolution exhibit. *Science Education*, 95(4), 720-744.

John Benjamins Publishing Company

Explain *with*, rather than explain *to*

How explainees shape their own learning

Josephine B. Fisher,[1] Katharina J. Rohlfing,[1] Ed Donnellan,[2,4]
Angela Grimminger,[1] Yan Gu,[3] and Gabriella Vigliocco[2]
[1] Paderborn University | [2] University College London | [3] University of Essex | [4] University of Warwick

Research about explanation processes is gaining relevance because of the increased popularity of artificial systems required to explain their function or outcome. Following an interactive approach, not only explainers, but also explainees contribute to successful interactions. However, little is known about how explainees actively guide explanation processes and how their involvement relates to learning. We explored the occurrence and type of explainees' questions in 20 adult — adult explanation dialogues about unknown present and absent objects. Crucially, we related the question types to the explainees' subsequent recall of the unknown object labels. We found that explainees asked different types of questions, especially about the object's label and facts. Questions about the object's function were asked more when objects were present. In addition, requests for labelling were linked to better recall. The results contribute to designing explainable AI that aims to provide relevant and adaptive explanations and to further experimental approaches to study explanations.

Keywords: explanations, explainee, interactive learning, types of questions, presence of referents

Introduction

Whereas research on explanations has traditionally focused on how a cause of an event can be revealed, recently, the focus has shifted toward explanations of everyday situations. These studies are motivated by the fact that technology surrounding us requires explanations to be accessible, adjustable, and thus useful. To make technology accessible is the goal of Explainable Artificial Intelligence (XAI) that develops models capable of explaining their functions. Scholars within the field increasingly recognize that not only is it important to develop contents to

https://doi.org/10.1075/is.23019.fis
Interaction Studies 25:2 (2024), pp. 244–255. ISSN 1572-0373 | E-ISSN 1572-0381

explain (the *what*), but because the goal of explaining is to create understanding in an explainee, *how* to explain is also critical. Surprisingly, it seems difficult to achieve a good understanding in an explainee, because people differ in what they are interested in and thus require diverse solutions. In XAI, it is argued that the answer to "what" is not as important as information about "how" (Besold & Uckelman, 2018) as it takes the users' needs, interests, and experiences into account when formulating a successful explanation. Miller (2019) claims therefore to recognize explanations as a social process to better understand how the required information emerges in the social interaction. Following this line of research, current approaches allow the user to steer the explanations by asking questions (Sokol & Flach, 2020), thus, allowing personalization of the interaction. However, although novel approaches in XAI build on the users' ability to ask questions (Sokol & Flach, 2020) and thus allow for a user's contribution in the explanation process, they currently lack an empirical basis to allow an interactive and thereby adaptive explanation. Previous studies on voice assistance systems revealed that users' familiarity with them influences their question-asking behavior (Le Bigot et al., 2008). However, the familiarity effect does not inform about how explainees might give cues and prompts to their needs in understanding.

In fact, little is known about explainees' question-asking behavior. When people are asked to explain spontaneously, they often provide long and monological statements which contribute only a little to learning (Chi et al., 2008). In contrast, the active involvement of both participants contributed considerably to deep learning effects (see Chi, 2009 for a summary). Question asking is used as a means to retrieve specific information (Rugerri & Lombrozo, 2015). Drawing from tutoring research, it is important to understand the type of questions that explainees ask (as a measure of their active involvement) and how these questions can foster their knowledge acquisition to provide a better empirical basis for the design of XAI.

The present study addresses this gap by analyzing the verbal interaction in dyads conversing about objects (provided by the experimenters). However, the experimental design is limiting explainers' behaviors to providing labels, functions, and facts of novel objects. Thus, one objection is that information that is typical in explanations (relations, causes) is not elicited in our setting. According to Klein (2009), however, contents such as labels, functions, and facts are essential and occur in more complex explanations as well. Thus, even though limited, the design of this study is providing ground for future experiments on interactive explanations and the enhancement of the design of adaptive systems.

For the analysis of question types, we took inspiration from work carried out on caregiver – child interactions, by Tare and colleagues (2011) that investigated how parents support children's learning at an exhibit. Because little is known about questions during explanations, we aimed to explore what types of questions will be raised.

In addition, in this study, two conditions were included: one in which the object explained was present during the interaction, and one in which it was absent. Reasoning that the presence of an object is a resource for communication, and non-present referents may elicit different forms of verbal and nonverbal behavior (e.g., Stukenbrock, 2014; Vigliocco et al., 2019), we hypothesized that the two conditions will give rise to different question types.

The question types observed were then used to relate performance by the explainee in a recognition task.

Method

ECOLANG Corpus

The audio data used in the analyses is taken from the dyadic adult-adult interactions of the ECOLANG Corpus[1] (Gu et al., in press). This is a corpus of dyadic interaction between two adults, or an adult and a child.

Participants

Participants in the dyadic explanation dialogues were 40 native English speakers (mean age of 25 years). 20 dyads were considered (5 female-female, 4 male-male, 11 female-male).

Stimuli

For each dyad, 24 stimuli objects (12 unknown, 12 known) out of a total of 37 objects were chosen. These objects belonged to four categories: tools, musical instruments, food, and animals (see Figure 1).

1. Detailed information on the ECOLANG Corpus: https://ucl.primo.exlibrisgroup.com/discovery/collectionDiscovery?vid=44UCL_INST:UCL_VU2&collectionId=81414722010004761

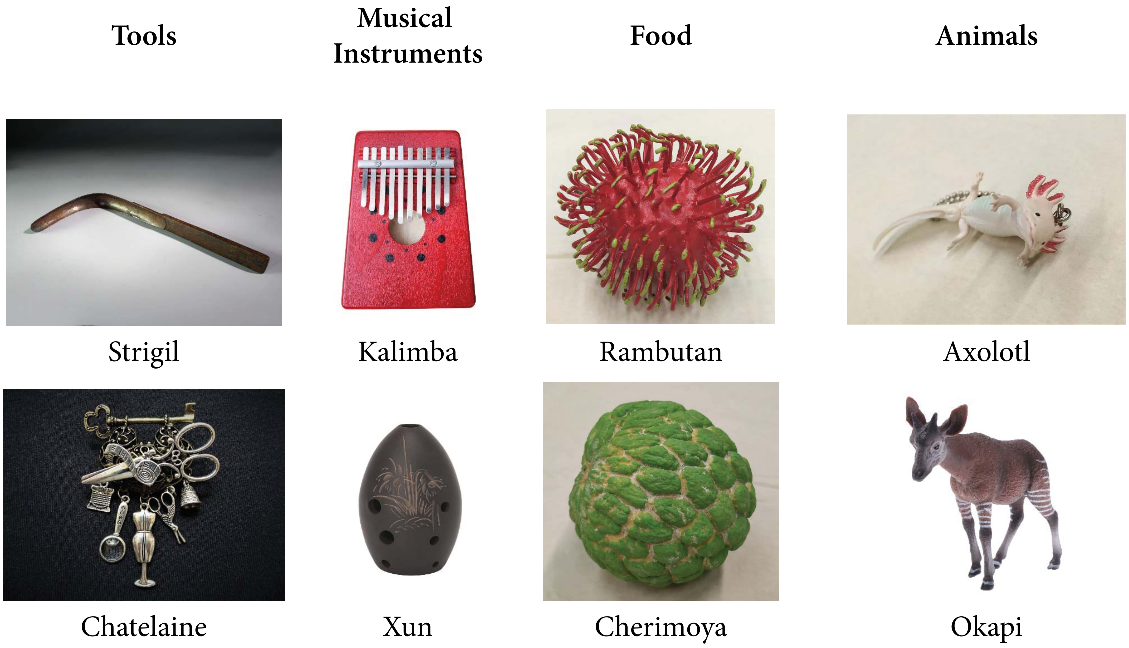


Figure 1. Four object categories as a pool

Procedure

The study was conducted in three stages: training session, explanation session, and recall task. Only the explanation session was recorded. For each dyad, one of the participants (explainer) was more knowledgeable than the other (explainee). Each of them was taking part in two sessions.

The online training session took place one or two days before the explanation session. In this session the explainer received materials informing them about different objects. They were presented with videos of 12 unknown objects which contained information about the labels, functions, and facts of these objects. For the 12 known objects, bulletin summaries of the object characteristics listed next to a photo were provided to the explainer. Additionally, in this session the explainer was instructed to talk about the objects in the explanation session. Right before the explanation session, the explainer went over a PowerPoint presentation listing all the objects including their pronunciation. Before the session, the explainee was informed about the subsequent recall task, in which labels were asked. In the explanation session, the dialogical partners were asked to engage about objects provided by experimenters, while they were audio- and video-recorded. The explainer spontaneously provided the explainee with knowledge about the unknown and known objects. The possible 12 unknown objects were each consecutively presented in two conditions: present and absent. When the first presentation was absent, the second was present.

If the objects were not present initially, the experimenter informed the explainee that the explainer was already aware of what the items were, and the explainer would provide information about the items while the experimenter retrieved them.

The order was counterbalanced, and half of the dyads started with the object-present condition and the other half of dyads with the object-absent condition. They were given four to five minutes to discuss the objects in each condition.

The recall task was administered directly after the explanation session and only with the explainee. During the task, explainees were tested on remembering the labels of the unknown objects in a forced choice test: The participants had to decide whether a presented picture and sound of a label matched.

Afterwards, explainees were asked to fill in a questionnaire in which they had to indicate their prior knowledge of the unknown objects to check whether these were unknown to them before the explanation.[2]

Coding

The question type analysis focuses on the unknown objects. For this, the explainee's questions were coded into six categories and transcribed using ELAN (2021). The categories are based on the work by Tare and colleagues (2011) with some modifications that concern the category of labelling and procedure questions (see Table 1). In the coding scheme by Tare and colleagues (2011), labelling questions were counted as factual questions. However, we decided to further differentiate them to examine more closely which particular types of questions lead to higher word recognition scores. With procedure questions, speakers in our data aimed to negotiate with what object to proceed in their conversation. In the work of Tare et al. (2011) this category was used with reference to space because the parents and children talked about where to go next within the exhibition. Based on observations in the data set, the category of reassurance was added. For the definition and examples of the question types, see Table 1.

In addition to the question types, we also coded the explainees' performance in the recall task: Each correct choice in the recall test was coded as 1 and each incorrect choice as 0. Their average performance was correlated to the question types.

---

2. According to a post-experiment survey, if explainees knew the objects before the interaction, the items were excluded from the analysis. Therefore, the number of unknown objects varies for each explainee between eight and eleven objects (mean: 9.5).

Table 1. Coding scheme for the question types. All examples are taken from the recordings

Question type	Definition	Examples
labeling	Target the name of a stimulus in an oral or orthographic form.	*What is that?* *How do you spell that?* *What's it called okapi?*
function	Different functions of a music instrument, tool or body parts of an animal and what they do. How to eat a fruit.	*What does that do?* *And you eat it, like, with a spoon or something?*
factual	Seek information of a profile or product description. Include general demands for more information and examples.	*What's a gourd?* *Where do they live?* *When was it invented? And then?*
reassurance	Make sure the speaker is serious, a sign of disbelief or surprise or auditory difficulties.	*Really? Shells right? Is it?*
procedure	Is an off-topic remark on how to continue with the discussion. Also includes information on the pre-session of the explainer.	*Can I try?* *Did you watch videos?* *Well they not told you?*
personal	Includes individual preferences and experiences of the explainer and explainee.	*Have you played it before?* *Did you know that yourself?*

Results

All explainees engaged in asking questions. Overall, explainees asked 29 questions on average across the 12 unknown stimuli (*SD* = 15.9, range 9–73).

In the following, we will present the results concerning the questions types and their relation to the learning measure.

What type of questions do explainees ask?

For this analysis, we will consider the mean number of question types in the present and absent condition (see Figure 2).

As can be seen from Figure 2., all question types were present in both conditions, although there were only very few procedure questions overall. Factual and labelling questions were the most frequent ones in both conditions.

Thus, participants varied highly in their repertoire of questions, with the exception of factual questions.

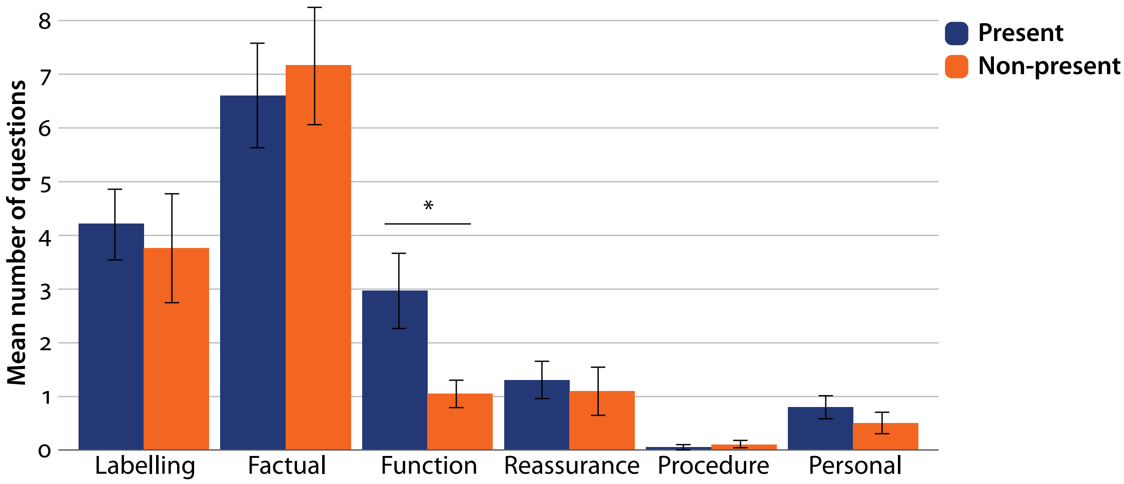


Figure 2. Mean number of question types

With respect to the present and absent condition, we found a significant difference in the occurrence of the function questions: According to a Wilcoxon test, the object-present condition ($M = 2.95$, $SD = 3.5$) elicited significantly more function questions than the absent condition ($M = 1.05$, $SD = 1.2$), ($Z = -2.46$, $p < .05$). This effect persisted also when we controlled for the overall number of questions that an explainee raised and considered percentages of the different question types.[3] However, this was the only significant difference. In contrast to our hypothesis, thus, our results suggest remarkable similarity between the conditions to a large extent.

Do questions relate to learning measures?

In the next step, we considered how the specific types of questions related to the recall score, and thus to the learning of the participants. On average, the participants scored .80 ($SD = .14$, Range .37–1.00) in the recall task. We performed a Spearman correlation analysis of the number of questions in the different question types and the recall score. The obtained r-values can be seen in Table 2. Bonferroni corrected results revealed that only the question type of labelling in the present condition correlated positively with the recall score ($r_s = .60$). Interestingly, although the occurrence of this question type did not differ between the condi-

3. In the study the number of questions were not put into proportion over total number of utterances of the speaker because they had a similar time limit.

tions, labelling questions in the absent condition were not significantly related to participants' learning.

Table 2. Spearman-Rho coefficients as results from the correlation analysis relating the question types to the recall score in the present and absent conditions (Bonferroni-corrected significance level, $p < .05$, indicated by *)

Question type	*r*-value in condition, *N* = 20	
	Present	Absent
Labeling	.60*	.35
Factual	.26	−.01
Function	.16	−.12
Reassurance	.27	.02
Procedure	.30	.22
Personal	−.04	.12

Discussion

In this study, we analyzed dialogues in which unknown objects were explained. Complementing current research focusing on the explainer, our analysis provides important insights into how an explanatory dialogue is shaped by an explainee: Instead of being only a passive receiver, the explainee asks various kinds of questions seeking specific information from the explainer. Following an approach created to describe child – caregiver interactions (Tare et al., 2011), we classified the questions asked by the explainee and found that questions concerning the label, facts, and functions were the most frequent. In addition, we investigated whether the type of questions asked differed depending upon whether the object was present or absent. Given that objects can be considered as resources for communication (e.g., Stukenbrock, 2014; Vigliocco et al., 2019), we hypothesized that they would elicit different question types. We found however, that the frequency of the questions occurring was similar in these two conditions. Only questions concerning the object's function were more common when the objects were present suggesting that the physical presence of the objects rendered their function-affording properties more saliently. For the other types of questions, the present study does not have evidence for the hypothesis that the two conditions result in different set of explainee's behavior. For the development of XAI, this might be of particular importance as XAI targets to explain the system's function. Thus, XAI research needs to consider that the presence of objects or referents might be crucial to elicit questions about systems' functions. Further research is necessary to investigate technical contents in explanatory dialogues, how they are talked about, and how they result in a better understanding of the presented functions.

A crucial point of interest in our study was whether different questions were especially likely to relate to the mapping between the object and their label (as assessed in the recall task). In this task, the explainee had to decide whether the novel label and a picture of the object matched. Our results suggest that the more labelling questions the participants asked in the present condition the higher they scored in recalling the new labels and vice versa. One likely explanation for the correlation is that learning to map labels to objects is easier when the label is provided at the same time as the referent. The co-occurrence of label and referent is considered to be critical for early word learning in infants (Smith et al., 2002), our results indicate that this advantage exists in adulthood as well (see also Derks & Dunman, 1974). Without knowing the objects' appearance, participants' knowledge about the labels was little helpful in this task, as it was demonstrated by the lack of effectiveness of labelling questions in the absent condition. However, in our study, learning was limited to the mapping task. Future research needs to regard this limitation by creating comprehension or learning tasks that are assessing not only the mapping but also broader knowledge about the unknown object, its function, and facts. In this respect, Rohlfing and colleagues (2021) suggest different forms of comprehension that everyday explanations might yield ranging from remembering some facts to actual performance or ability to decide.

As our study was designed with a clear focus on explanation of novel objects' labels, facts, and their function, two critical points follow from it: Firstly, that the explainees were informed about the recall task prior to the explanation might have motivated them to ask comprehension questions that directly secured the acquired knowledge. We, thus, cannot rule out the possibility that our results about the engagement of the explainees are due to the prior announcement of the recall task. Secondly, in everyday settings, explanations extensively connect to emerging interests and various knowledge gaps requiring different forms of comprehension and follow-up actions. For this purpose, further studies have to take a rich behavior of the explainer into account in which different information is connected and combined.

Importantly, the results speak to a crucial involvement of the explainee, as all participants engaged in asking questions. Our results line up with the interactive learning account (Chi, 2009) proposing that episodes of tutoring that are initiated by the tutee foster their learning. Our results extend, however, this approach to explanatory dialogues showing that it is not only the sole initiative of the explainee that is crucial for recall of the labels. Even though they were as frequent as in the present condition, labelling questions in the absent condition were not significantly related to the learning scores suggesting that the learning advantage cannot result from hearing the label more frequently and on own initiative. Instead, our results suggest that for the recall task, the visual information about the object was crucial. We see great potential to extend the analysis presented in this paper focusing on how the explainees' shape the dialogue to interaction-based measures relating the behaviors of the explainee and explainer. Such an extension would provide novel insights into how explainers react to the explainees' questions and whether there are patterns of responses.

For the development of XAI, the results of this study are informative as they show the necessity to involve the explainee as much as the explainer in the process of explaining: By raising quite a lot and a variety of different types of questions, the explainee is significantly steering the explanatory dialogue and, thus, contributing to the construction of the relevant information.

Funding

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR 318/1 2021 – 438445824 and the European Research Council Advanced Grant (ECOLANG, 743035).

References

Besold, T.R., & Uckelman, S.L. (2018). The what, the why, and the how of artificial explanations in automated decision-making. arXiv:1808.07074 [cs.AI].

doi Chi, M.T., Roy, M., & Hausmann, R.G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive science*, *32*(2), 301–341.

doi Chi, M. (2009). Active-Constructive-Interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*, 73–105.

doi Derks, P.L., & Dunman, J.E. (1974). The effect of repetition of objects and object names on free recall. *Bulletin of the Psychonomic Society*, *3*(4), 289–292.

ELAN (Version 6.2) [Computer software]. (2021). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

doi Gu, Y., Donnellan, E., Grzyb, B., Brekelmans, G., Murgiano, M., Brieke, R., Perniss, P. & Vigliocco, G. (2025). The ECOLANG Multimodal Corpus of adult-child and adult-adult Language. *Scientific Data*.

Klein, J. (2009). Erklären-Was, Erklären-Wie, Erklären-Warum. Typologie und Komplexität zentraler Akte der Welterschließung. In R. Vogt (Ed.), *Erklären. Gesprächsanalytische und fachdidaktische Perspektiven*. Stauffenburg.

Le Bigot, L., Bretier, P., & Terrier, P. (2008). Detecting and exploiting user familiarity in natural language human-computer dialogue. In K. Asai (Ed.), *Human computer interaction: New developments* (pp. 369–382). Intech Open Limited.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38.

Rohlfing, K. J., Cimiano, P., Scharlau, I., Matzner, T., Buhl, H. M., … (2021). Explanation as a social practice: Toward a conceptual framework for the social design of AI systems. *IEEE Transactions on Cognitive and Developmental Systems*, *13*(3), 717–728.

Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, *143*, 203–216.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological science*, *13*(1), 13–19.

Sokol, K., & Flach, P. (2020). One explanation does not fit all. *KI-Künstliche Intelligenz*, *34*(2), 235–250.

Stukenbrock, A. (2014). Pointing to an 'empty' space: Deixis am Phantasma in face-to-face interaction. *Journal of Pragmatics*, *74*, 70–93.

Tare, M., French, J., Frazier, B. N., Diamond, J., & Evans, E. M. (2011). Explanatory parent–child conversation predominates at an evolution exhibit. *Science Education*, *95*(4), 720–744.

Vigliocco, G., Motamedi, Y., Murgiano, M., Wonnacott, E., Marshall, C. R., Milan Maillo, I., et al. (2019). Onomatopoeias, gestures, actions and words in the input to children: How do caregivers use multimodal cues in their communication to children? Paper presented at the 41st Annual Conference of the Cognitive Science Society, Montreal, QB.

Address for correspondence

Josephine B. Fisher
Faculty of Arts and Humanities
SFB/TRR 318 Research Center "Constructing Explainability"
Paderborn University
Warburger Str. 100
33098 Paderborn
Germany
j.b.fisher@uni-paderborn.de

Biographical notes

Josephine B. Fisher is a PhD student at the TRR 318 Research Center "Constructing Explainability" in the Psycholinguistics Group at Paderborn University in the project A01 "Adaptive Explanation Generation". Her research interests are interactive adaptivity in explaining processes.

https://orcid.org/0000-0002-9997-9241

Katharina J. Rohlfing received the Master's degree in linguistics, philosophy and media studies from the University of Paderborn in 1997. As a member of the graduate program "Task oriented communication", she received the Ph.D. degree in linguistics from the Bielefeld University in

2002. Since 2015, she is full professor of psycholinguistics at the Paderborn University heading the SprachSpielLabor. Her research focusses on learning processes that drive fundamental communication abilities. Working across disciplines, her interest is to apply insights from research on language acquisition to develop artificial learning systems that interact with people.

https://orcid.org/0000-0002-5676-8233

katharina.rohlfing@uni-paderborn.de

Ed Donnellan is a Research Fellow in the Department of Psychology at the University of Warwick and an Honorary Research Fellow in the Department of Experimental Psychology at UCL. His research interests are language learning, curiosity and multimodal communication.

https://orcid.org/0000-0002-2739-7322

ed.donnellan@ucl.ac.uk

Angela Grimminger is a senior researcher at the Psycholinguistics research group at Paderborn University and Head of the SprachSpielLabor. Her research focuses on language and gesture development, child-robot interactions and multimodal interaction. She is a principal investigator in Project A02 "Monitoring the Understanding of Explanations" at the TRR 318.

https://orcid.org/0000-0002-5749-9362

angela.grimminger@uni-paderborn.de

Yan Gu is a lecturer (assistant professor) at the University of Essex and an honorary research fellow at UCL. His research interests include multimodal communication.

https://orcid.org/0000-0001-6093-3919

yan.gu@essex.ac.uk

Gabriella Vigliocco is a Professor of Psychology at the University College London in the department of Experimental Psychology. She earned her Doctor of Philosophy from the University of Trieste in 1995, following her Bachelor of Science from the same institution in 1990. Her research focuses on the psychological and neural mechanisms underlying human language use, with particular emphasis on the integration of conceptual and linguistic information.

https://orcid.org/0000-0002-7190-3659

g.vigliocco@ucl.ac.uk

Publication history

Date received: 2 May 2023
Date accepted: 5 February 2024

Exploring the Semantic Dialogue Patterns of Explanations – a Case Study of Game Explanations

Josephine B. Fisher
Psycholinguistic
Paderborn University
TRR 318 Constructing Explainability
j.b.fisher@upb.de

Amelie S. Robrecht
Social Cognitive Systems
Bielefeld University
TRR 318 Constructing Explainability
arobrecht@techfak.de

Stefan Kopp
Social Cognitive Systems
Bielefeld University
TRR 318 Constructing Explainability
skopp@techfak.de

Katharina J. Rohlfing
Psycholinguistic
Paderborn University
TRR 318 Constructing Explainability
katharina.rohlfing@upb.de

Abstract

Contributing to the research on social design of explainable AI, we studied 51 German dyadic explanations to reveal how an explanation process is unfolding and to what extend both, the explainer (EX) and the explainee (EE) are contributing to the content. In this paper, we exploratively examine semantic dialogue patterns of semi-naturally and spontaneously occurring explanations of the game Quarto, which are – compared to an expert explanation – less restrictive. We apply the notion of explanation nodes to identify explanation blocks as well as their order that constitute the internal structure of these explanations. In particular, we analyse which information is covered by an explanation dialogue in terms of both, coverage and frequency. Our results reveal the engagement of both interlocutors and provide a basis for the study of adaptivity in explanations and its realisation in dialogue systems.

1 Introduction

Explanations provide an interesting case for the study of how semantic structure is built up during a dialogue: As explanations have the goal to result in understanding, it is reasonable to assume that both partners need to contribute to the structure (Rohlfing et al., 2021). However, little is known about how this joint co-construction unfolds. At the same time, there is a growing need for understanding how explanations succeed. In the last years, Explainable AI (XAI) is driven by the General Data Protection Regulation (GDPR) (Carey, 2018) and the right to have an algorithm explained. This is particularly the case for "blackbox" machine learning algorithms. While many approaches to how to make the blackboxes inspectable exist, the process of explaining, i.e., the way of how to present the relevant content (the *explanandum*) and how to ensure sufficient understanding of it, receives little attention (Anjomshoae et al., 2019). The research area of XAI seems to be unbalanced, prioritising what aspects and features to explain instead of how to explain (Baniecki et al., 2023). Thus, empirically-driven studies are demanded to address the research gap from the perspective of a more user-centred and social interaction (Madumal et al., 2019). To create systems that are adaptive and provide an explanation that addresses the users' knowledge gap, it is crucial to explore how humans achieve an adaptive process when interacting with each other in explanatory dialogues.

In our investigation we address this gap by focusing on the co-constructive character of explanations, subscribing to the view that explanation is a social and co-constructive process (Rohlfing et al., 2021; Miller, 2018). How this co-constructive process is reflected in the dialogues can be addressed by contrasting the distributions of semantic contributions of the interlocutors. Thereby, we take into account the influence the EE and EX can take within a dyad. This allows both interaction partners to shape the content of the discourse. Who is planning and structuring and who is confirming the explanation?

While there is a well established research focus on modelling the structures of direction-giving (guiding a person to a specific place via verbal instruction) by extracting different phases out of human-human interactions (Psathas and Martin, 1976; Ewald, 2010), there is little done on spontaneous explanation dialogues. Due to this research gap, this paper will describe the semantic dialogue patterns of human-human everyday explanations to point out reoccurring structures. By introduc-

Proceedings of the 27th Workshop on the Semantics and Pragmatics of Dialogue, August, 16–17, 2023, Maribor.

ing an explanation node scheme, we also allow a more fine-grained semantic analysis. To reach this goal, we combine linguistic analyses with methods from computer science to work towards an implementation of humans' adaptive capabilities. The linguistic analyses focus on the explanation structure by introducing an explanation node scheme where each explanation node – which is the smallest unit in the system (see Sec. 3.3) – captures a semantic dialogue pattern which can be observed in the interaction. We employ this explanation node scheme to study the semantic dialogue structure of explanations between two interlocutors engaged in explaining a board game. The structures and relations that are represented by the explanation node scheme can be transferred to an ontology and, e.g., serve as a knowledge base in an adaptive explanation dialogue system.

Based on current research, we expect a game explanation to be sequential and co-constructive. (1) Concerning sequentiality, we expect sequential patterns comparable to the phases in direction-giving introduced by Psathas and Martin (1976). In addition, because the setting is eliciting everyday explanations, (2) we expect the EE to be an active participant (Rohlfing et al., 2021; Fisher et al., 2022) having the opportunity to introduce explanation nodes on their own. (3) Based on Rohlfing et al. (2021) and Miller (2018), we further expect the explanations to be co-constructive. For that, we will investigate the EE's contributions and how they are addressed jointly. If the EE is the first to mention an explanation node, we expect the EX to take it up.

2 Background and Related Work

Much work on how information is established during an interaction was characterised by Clark (1996) as introduced by his theory of *common ground*. It displays how conversational partners agree on their shared information, during the course of an interaction. Any type of discourse is a joined activity in which the common ground between interlocutors increases, and in which "sections and subsections [are]n't fixed beforehand, but [are] negotiated as [they go] along" (Clark, 1996, p.36). This includes "the knowledge, beliefs, and suppositions they believe they share about the activity" (Clark, 1996, p.38).

2.1 Structures of Explanations and Tutoring

Taking a broader perspective towards human explanatory dialogue, each explanation involves two conversational partners with an asymmetric knowledge distribution: an EX, who is more knowledgeable, and an EE, who is less knowledgeable. The subject of the explanation is the so-called *explanandum* which is constituted by different types of *explanans* (Rohlfing et al., 2021). Looking at an explanation as a process, it unfolds because the EX and the EE work together on specifying what information is needed for the EE to understand (Klein, 2009) as well as what is or should be the object of explaining (*explanandum*). Klein (2009) claims that there are several types of explanations; they relate to the How, the Why and the What. Scientific explanations rather focus on the Why, whereas everyday explanations reveal a variety in their types. The subtype of everyday explanations we are focusing on, are game explanations which cover different aspects, such as rules, figures and the game board. Kotthoff (2009) classifies game explanations in more detail as procedural explanations. This goes in line with the categorisation of Klein's (2009) definition of How explanations.

One can define an explanation process as a sequence of phases that contain explanation and verification blocks (El-Assady et al., 2019). How to find the optimal order of these blocks and which explanation strategy to choose depends on the level of detail, the EE, and the desired amount of interactivity. In this paper, we investigate such explanation blocks in human-human explanations and study how to extract their internal structure from explanation dialogues. An explanation involves two processes, the cognitive process, which can be described as the planning and construction of the explanans, and the social process, which focuses on the interaction between the EX and the EE (Miller, 2018). This paper will put the spotlight on the explanation as a conversation, by focusing on the content structure.

Similar to explanation, in the context of tutoring, a knowledge asymmetry exists. However, the addressee is supposed to learn, which is not necessarily the case in explaining, where the focus is on understanding or enabling (Rohlfing et al., 2021). Research on tutoring (Chi et al., 2008; Miyake, 1986) has established mind maps, in which the different elements that are part of a topic are listed and numbered in individual nodes, to account for the contents which were already discussed and understood in a conversation. These mind maps differ from what is known in linguistics as semantic maps that sound similar. Haspelmath (2003) has proposed the semantic map method that displays the lexical relatedness of words. It uses graphs to present relatedness of co-expressed meanings, connecting nodes by edges to describe which concepts can be expressed by the same words. However, it does not focus on the semantic relatedness of the explanation elements and is thus little relevant for the idea of the mind map. Explicitly in the work by Chi et al. (2008) on scientific explanations, the problem solving nodes were based on the verbal explanations of the tutors when they explained the steps alone. There, the individual nodes relate to a problem solving step. Miyake (1986) similarly listed the different elements in a hierarchical fashion which belong to a problem regarding the stitches of a sewing machine. For this purpose, the framework was called "the function-mechanism hierarchy". In this, the contents are differentiated in two ways. They address the function – what is taking place – and the mechanism – how it is performed. They are in such a way connected that the mechanism at a lower level is needed to describe the function of the next higher level. Here, the categorisation of the elements and its level of detail is justified as being appropriate to examine the ongoing process of understanding (Miyake, 1986).

2.2 Models in Computer Science and XAI

In contrast to the previously introduced node system, using an ontology or a knowledge graph (KG) to store information is a common method in dialogue systems (Axelsson and Skantze, 2023; Robrecht and Kopp, 2023; Axelsson and Skantze, 2020; Ma et al., 2015; Lin et al., 2015). The KG concept was first introduced by Minsky (1968), who called them *semantic networks*. Today it is used in approaches such as *the semantic web* (W3C, 2012) or *Wikidata* (Vrandečić and Krötzsch, 2014). The domain that is stored using a KG varies from scientific publications or E-commerce to social networks and geopolitical events (Kejriwal et al., 2021). While most ontologies are defined by the Resource Description Framework (RDF), other approaches or variations – such as Resource Description Framework Schema (RDFS) – are used. We will focus on RDF, as popular languages, such as the *Ontology Writing Language* (OWL) (W3C, 2012) derives from it. An RDF graph consists of a set of triples, each consisting of a *subject*, an *object* and the connection *predicate*. In other words: Two entities (subject and object) are connected via a relationship (predicate). Further information and restrictions on designing an RDF ontology can be found in Kejriwal et al. (2021). The subject, object or predicate – the smallest unit in an ontology – captures only one single entity or relationship. Therefore, a node might be, but not necessarily has to be, broken down into multiple triples, when transforming an explanation node scheme into an ontology.

In human robot interaction (HRI), the majority of research aims to create *Explainable Agency* or *Goal-Driven XAI*. As the agent explains behaviour and decisions, the interaction becomes predictable to the user (Anjomshoae et al., 2019). Next to predictability, understandability is a key goal when thinking about explainable agency. Both can be increased by improving the agents human-likeness. By looking at the processuality of human-human explanations, we aim to find patterns that can be transferred to HRI settings at a later state. Currently, effects on communication and explanation structure are usually measured using interaction studies (Stange and Kopp, 2020, 2021). Subsequently, the explanation is adapted to the best condition. There is research that uses a bottom-up approach by analysing multiple explanation interactions for their model (Madumal et al., 2019), but none of the considered dialogue types, the model is based on, are verbal everyday human-human explanations. Nevertheless, in the final study the agent performs an explanation on the board game "ticket to ride", which can be considered an everyday explanation in an agent-human setting. Most of the current papers define the communication of the explanation as their most important future work project (Anjomshoae et al., 2019).

3 Method

3.1 Participants

A subset of 51 game explanation dyads with a total of 102 participants from the *ADEX* (Adaptive Dialogical Explanations) corpus, which we collected in the project A01 *Adaptive Explanation Generation* in the *TRR 318 Constructing Explainability*[1], was considered. In the recorded (video and audio)

[1] https://trr318.uni-paderborn.de/en/projects/a01

Figure 1: Study design of ADEX corpus[2]

dyadic explanation dialogues 60 female, 38 male and three diverse subjects took part (age range: 18-55 years). 96 participants were native German speakers and five were second language speakers. Lastly, 94 of them were students and seven had other occupations [3]. The study was conducted in six phases (Fig. 1). Phase 1, 3, 5 and 6 were different questionnaires, which included psychological and understanding instruments. In Phase 2, the participants took part in the explanation without the game being present. Before the study, the EX was asked to learn the game Quarto. Quarto is a strategic board game that includes game figures with four different characteristics. The goal for each player is to place four figures in a row that have one of those characteristics in common. They were free to use any resources they liked for their preparation. We provided them with some exemplary sources and the possibility to take a look at the physical game before the study. After the first phase, the EX was instructed to spontaneously explain the game in such a manner that the EE would have the chance to win the game. The EE was told to actively take part in the explanation. The participants had no time restrictions for the explanation phase. Consequently, the explanations can be considered diverse because the subjects were free in their preparation of the game and their speech. In Phase 4, the dyads were instructed to play a couple of games of Quarto and to continue explaining. This phase was excluded in the current analysis.

3.2 Linguistic Coding

To explore the semantic dialogue structure of the explanations, we coded the speech according to their content with the program *ELAN* (Wittenburg et al., 2006)[4]. Therefore, we adapted the node scheme from scientific explanations to game explanations of Quarto [5]. The speech is divided into moves by the conversational partners. Following the work of Chi et al. (2008) moves are defined as statements including a single idea presented by a single speaker within one turn. Thus, the explanation nodes serve as a foundation for the speaker move analysis. Backchannels (such as *mh, yeah and okay*) are not considered in the analysis because they do not function as separate turns that attempt to take the conversational floor (Dideriksen et al., 2019). For the reliability check, six explanations (about 12% of the data) were coded concerning the blocknodes by two researchers. Thereby, an unweighted Cohen's kappa yielded an inter-coder reliability that can be considered almost-perfect (Landis and Koch, 1977) (k=0.90). The majority of mismatches related to the count of the parent - when one of their childnodes was discussed. Henceforth, deviations between the two coders were smoothed via post-hoc agreement. Based on this, the analysis of the whole data set was adjusted.

3.3 Explanation Structures

In contrast to the hierarchical and sequential order of scientific explanations, game explanations occur in a more flexible manner. In our approach, the semantic dialogue structure is captured in an explanation node scheme. Each *explanation node* captures specific semantic information. The explanation nodes are connected via arrows, whose direction represents an increase of detail. A parentnode is an explanation node at an upper level, while the next more detailed explanation node connected by an arrow is referred to as a childnode. A group of explanation nodes referring to the same semantic category form an *explanation block*, the highest node in a block is called *blocknode*. Together, the explanation nodes form a map that can be revisited

[2]©Paderborn University, Patrick Pollmeier

[3]One data point each is missing due to technical problems.

[4]Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands https://archive.mpi.nl/tla/elan

[5]The preliminary ADEX Codingscheme for Explanation Nodes can be found at https://go.upb.de/ADEX_Explanation_Nodes.

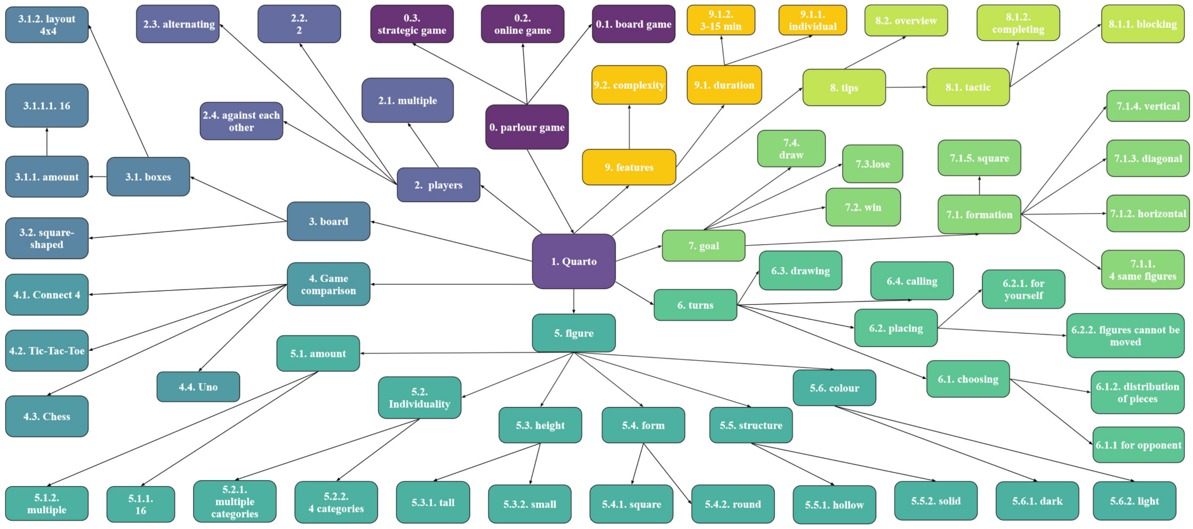


Figure 2: Node scheme: Each block is represented by a specific colour. The colour coding is consistent within all presented figures.

by the interlocutors (Fig. 2).[6] In contrast to (Chi et al., 2008), in which the tutor's explanation served as the only source for the node system construction, we instead used the explanation nodes in interaction. First, the blocknodes were established and in an iterative process the subnodes were added. We adjusted the level of detail to the topical occurrences in the data. As one can see, the game explanations cover ten blocknodes divided further in several subnodes. Taken together, 69 explanation nodes were identified.

The *Quarto* block only contains one node, its name, all the other nodes are placed around this central node. In the *Parlour Game* block, the game is put into the broader game context. All information related to the players, how many there are and in which mode they play, are grouped in *Players*. The third block captures the different characteristics of the *Board*. A special block is the game *Comparison* which contains the games that are frequently compared to Quarto. *Figures* is the largest block describing the characteristics of the game pieces. In *Turns*, the required game turns are listed and *Formations* names the possible formations of the figures and their impact on the goal of the game. Tactical tips are depicted in block *Tips* and the final block, *Features*, includes general features of the game, such as duration and difficulty. The block dependency is expressed through the colours, while each node has its own reference number.

[6]In the empirically developed explanation node scheme nodes were divided in subnodes if mentioned separately.

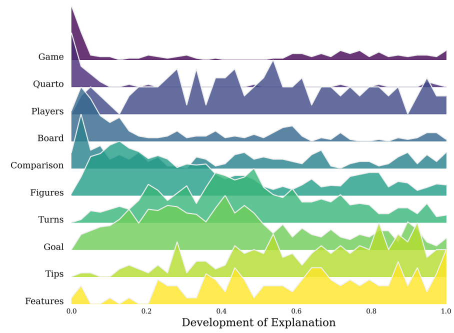


Figure 3: Reference to blocks by EX in relation to the time in all dialogues

4 Results

4.1 Order and Sequentiality

As previously introduced, we hypothesise the explanation blocks occur in certain patterns. These patterns will be described by focusing on the order the explanation blocks and nodes are either introduced or mentioned in. The order in which the blocks are mentioned by the EX can be seen in Figure 3[7].

It becomes apparent that the blocks *Game* and *Quarto* – if mentioned at all – are discussed in the very beginning of the explanation. The blocks *Board* and *Figures* are discussed subsequently, followed by the blocks *Goal* and *Turns*. The expla-

[7]The length of the interaction is normalised and the frequency of appearance is normalised for each block independently

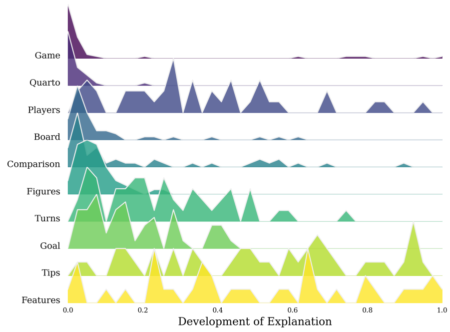


Figure 4: Introduction of the blocks by EX in relation to the time in all dialogues

nation is typically closed by referring to *Tips* and *Features* of the game. The block *Players* is not as explicitly connected to a specific part of the explanation; it can be addressed in the very beginning or at the end of the explanation or at both times. Apart from the discussed blocks, there is one block, the *Game Comparison*, that can be relevant at each state of the interaction. This shows how comparisons differ from the other blocks, as a *Game Comparison* unites nodes by their function and not primarily by their semantic meaning. Figure 4 displays the occurrence of a block being mentioned by the EX for the first time. Especially the blocks that are discussed in the beginning, such as *Parlour game*, *Quarto*, *Board* and *Figures*, are typically introduced in the beginning as well. The moment the block *Players* is mentioned first, shows a higher variance. Some explanations refer to the block *Players* at an early stage, while others first mention the block only in the second half of the explanation. Blocks that are discussed in the second half of the explanation, such as *Turns*, *Goal*, *Tips* and *Features*, are nevertheless often already introduced in the first half.

When distributing the explanation nodes separately (App. A Fig. 6), it becomes clear, that explanation nodes connected to certain blocks, such as *Figure* or *Goal*, tend to be explained close together at more or less the same place in the discourse, while blocks such as *Players* or *Turns* are spread over the whole interaction. This can be explained by either the fact that they are mentioned several times, as their semantic connection to other blocks is very strong, or that the order of the blocks differs in each dialogue. The explanation blocknode *2.0. Players* is not mentioned by any EX. A reason for this might be that the EX prefers the other – more detailed – explanation nodes of the block. Considering the individual explanation nodes, helps to understand, why the block *Game Comparison* is spread over the whole explanation. There are explanation nodes in the comparison block, that appear close to others due to their semantic relation (e.g. *4.4. Uno* and *6.4. Calling*): Similar to Uno, one also has to verbally indicate in the game that one has won. In Example 1 the EE notices the upcoming game comparison and brings in the name. Thereby, they co-construct the explanation and the EE displays their active participation.

Example 1 from D02

EX: Äh und dann ja hat man das Spiel gewonnen also es ist nen bissel dieser Ausruf kennt man ja so [von] genau von Uno letzte Karte.
Uh, and then yes, you won the game, so it's a bit like this exclamation that you know [from] exactly from Uno last card.[a]

EE: [Uno]
[Uno][a]

[a]English translation of the German transcripts.

Other comparisons, such as *4.2. Tic-Tac-Toe* or *4.1. Connect Four* can be used to compare multiple aspects of the game, as they have several semantic relations to Quarto.

4.2 Coverage and Frequency

In the following, the coverage and frequency analysis of the explanation nodes will be presented[8]. This includes answering the questions: (a) How many explanation nodes are addressed in the explanations and (b) How often is an explanation node addressed in an explanation (and by whom)? Turning to the coverage of the explanation nodes by the conversational partners. On average, the EX mentions 49% (min. 33% and max. 67%, SD = 8.0) of the explanation nodes in their explanations. In other words, about half of all explanation nodes are covered by the EX in the explanations. In contrast, the EE addresses on average 20% of the explanation nodes (min. 4% and max. 48% , SD = 11.0). Therefore, the EE relates to the explanation nodes less frequently and contributes less to the overall map coverage. We will now take a look at how the individual explanation nodes are covered in coverage and frequency. There are ex-

[8]For the analysis ELAN Annotation Frequency and Coverage (Biermeier, 2023) was used.

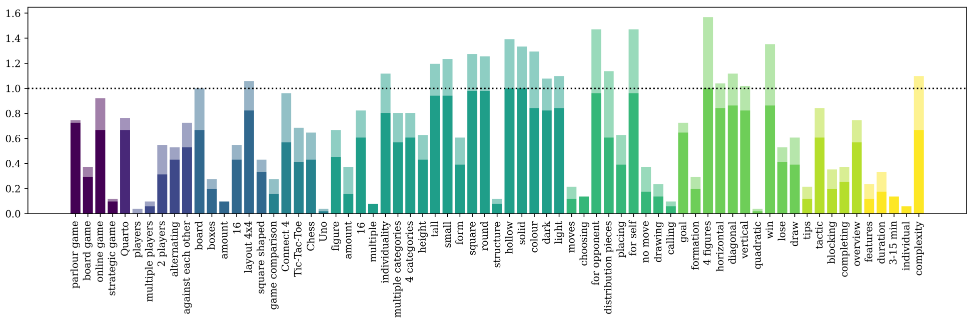


Figure 5: Explanation node coverage by EX (bottom bar) and EE (top bar) - each bar displays in how many of the dialogues the explanation node is mentioned. It shows the proportionate occurrence of the explanation nodes in the entire data set.

planation nodes in each block that are covered in almost every explanation, while others are discussed rather sparsely. When looking at the frequency of the explanation nodes (Fig. 5), it becomes clear that neither the more general blocknode [9] nor the more specific childnodes have a higher frequency of being mentioned. No explanation node specific patterns can be found, but block specific tendencies are observable. When describing the categories of figures, in three of four cases the contrastive characteristics are used more often (*5.3.1. tall - 5.3.2. small* (94.12%) – *5.3. height* (43.14%), *5.4.1. square -5.4.2. round* (98.04%) – *5.4. form* (41.18%), *5.5.2. solid - 5.5.1. hollow* (100%) – *5.5. structure* (9.8%), while in one case the category is used slightly more often (*5.6. colour* (88.24%) *5.6.1. dark - 5.6.2. light* (82.35%-84.31%). In general, the more detailed contrasting information is preferred. The comparison that is used in most explanations is *4.1. Connect 4*, which is mentioned in 60.78% of the explanations, while *4.4. Uno* is only used as a comparison in 1.96%. The explanation nodes that are mentioned in every explanation by the EX (Fig. 5 EX-darker bar) are *5.5.1. hollow, 5.5.2. solid* and *4 Figures*. There is no explanation node that is addressed in every explanation by the EE. The explanation node with the highest frequency is *7.1.1. 4 Figures* (Fig. 5 EE-lighter bar). The explanation nodes that are mentioned in less than 10% of the explanations are: *2. Players* (3.9%), *2.1. multiple (players)* (7.84%), *4.4. Uno* (1.96%), *5.1.2. multiple (figures)* (7.84%), *5.5. structure* (9.8%), *6.4. calling* (7.84%), *7.1.5. square* (1.96%) The option to arrange the figures in a quadratic shape is an optional rule and is not

[9]There is only one block where the blocknode has the highest frequency (*0. parlour game*).

Example 2 from D36

EE: Wie lange dauert das?
How long does it take?[a]
EX: Ne Runde höchstens zehn Minuten.
One round, ten minutes at the most.[a]

[a]English translation of the German transcripts.

captured in every external explanation of the game. As each participant was supposed to learn Quarto in advance with a source of their choice, this might be the reason for the low coverage. and *9.1.1. individual* (5.88%).

When looking at the frequency of an explanation node in a dialogue, it is considered to be discussed in depth, if it is mentioned more than five times by either the EX or the EE. There are only three explanations where no explanation nodes are discussed in depth and each block has at least one explanation node that is deeply discussed in either of the dialogues. Especially the fact, that a line needs four figures and that the figures are picked for the opponent are deeply discussed in more than half of the explanations (Tab. 1). The other explanation nodes that are deeply discussed occur in fewer explanations. They occur in a range of six till eighteen explanations. Overall, 221 times an explanation node is discussed in depth. In 95.48% of these, the EX is referring to the explanation node more often, than the EE. In these cases, the EE is rather passive. Nevertheless, there are explanations with a highly active EE. On the one hand, the EE can contribute nearly as many moves as the EX. On the other hand, the EE can introduce new explanation nodes (see example 2). Table 2 displays all explanation nodes that were referred to in more depth

#Dialogues	Label	Node
28	7.1.1.	4 Figures
28	6.1.1.	For Opponent
18	6.2.1.	For Self
16	5.5.1.	Hollow
12	7.2.	Win
11	5.4.2.	Round
11	5.6.	Colour
10	5.4.1.	Square
9	5.3.2.	Small
8	5.3.1	Tall
8	5.5.2	Solid
7	5.6.2	Light
6	8.1.	Tactic

Table 1: Number of explanations an explanation node is discussed in depth (>5) by either of the interlocutors

D-Number	Label	Node	EX	EE
D16	5.2.	Individuality	5	7
	5.5.1.	Hollow	5	5
	5.5.2.	Solid	5	11
D17	7.2.	Win	3	5
	8.1.	Tactic	3	6
D23	4.1.	Connect 4	2	6
	8.1.1.	Blocking	2	5
D42	5.2.	Individuality	3	7
D49	4.3.	Chess	2	5
	5.1.1.	16	5	5

Table 2: Dialogues in which explanation nodes are mentioned more frequently by the EE

by the EE than by the EX. Both, visualisations and examples show, how much the explanations differ from each other concerning their coverage and frequency. Finally, in our analysis, we addressed the question how explanations are co-constructed. For this purpose, the explanation nodes by the EE were analysed in detail to investigate which explanation nodes were introduced by the EE and whether the EX addressed these and when. In almost all of the dyads (50/51), the EE introduced a new explanation node. On average, the EEs initiated 4.2 new explanation nodes in a conversation (min. 0 and max. 10, SD = 2.6). Out of 212 explanation nodes that were introduced by the EE in the whole data set, the EX took up the explanation node directly 152 times (72%); 19 times (9%) they did not directly address the explanation node, but later on in the conversation. In 41 cases (19%), the EX did not take up the explanation node at all. Out of the 69 explanation nodes in total, the EEs introduced 48 (69%) throughout the different dyads. The results taken together show that an explanation is a unique interaction and highly depends on both conversational partners.

5 Discussion and Conclusion

In this paper, we introduced an explanation node scheme as a tool to model and explore the semantic dialogue structures of explanations. This tool allowed us to investigate the contributions of both dialogue partners to the domain knowledge. Concerning our research question (1), we were able to show that a game explanation is a sequential interaction. Nevertheless the patterns are not as restrictive as in a scientific explanation. A reason for this is likely to be the active participation of the EE, which we addressed in research question (2). In contrast to this, in an everyday explanation, the EE can be more active by demanding a more detailed explanation or pointing out knowledge gaps. In more naturally occurring explanations Fisher et al. (2022) also found a lot of variance in interaction patterns.

Further and with respect to research question (2), we expected the EE to be actively involved. We found support for this in our data showing that in each of the dyads in the corpus, up to 48% of the explanation nodes were covered by the EE showing also a high variance in the EEs' verbal contributions (Fisher et al., 2022). For future work, we hypothesise that the more active the EE is in the explanation, the less predictable it is to the EX, who has to adapt their explanation accordingly. This might account for why the sequentiality of the explanation nodes varies, even in our semi-natural game explanations. For naturally occurring explanations, we expect a higher variance in the contribution of the EE. This highlights the need for adaptive dialogue systems.

In research question (3), we set out to examine the relationship between the explanation nodes introduction by the EE and their uptake by the EX. The findings regarding the explanation node introduction by the EE being uptaken by the EX indicate that an explanation is a joined activity (Clark and Schaefer, 1989) in which the conversational partners co-construct their content (Rohlfing et al., 2021). To what extent the mentioned explanation nodes correlate with the EX's speaker moves, should be investigated in the future to provide more foundations for adaptive dialogue systems.

To conclude, based on first exploratory empirical results, we were able to display the content of explanations via explanation nodes. Thereby, we highlighted the active involvement of the EE by their explanation node introduction. The co-

construction of explanations is demonstrated by the take up of the explanation nodes by the EX.

The next steps in the linguistic analysis are: (1) to connect the explanation nodes with the verbal behaviour (speaker moves) of the conversational partners, as it was done in the works of Miyake (1986) and Chi et al. (2008). By making use of the nodes, one can keep track of the interaction history, i.e., the progress of the dialogue. Hereby, the explanation nodes can serve as a tool to support the future speaker move analysis because one is capable of telling whether information was already discussed and compare whether it has been modified. Chi et al. (2008) adds the concept of substantiveness to the contributions of the conversational partners. We hypothesise that the explanation nodes will correspond to this concept. This can be considered in future analyses. We only considered how the EX takes up the the explanation nodes the EE brings into the explanation and not all of their contributions. This could be an additional step for further analyses. When taking the modelling of human-agent explanation into account, the results will also be beneficial to the enhancement. The observed semantic dialogue patterns will be implemented into the dialogue system *SNAPE* (Robrecht and Kopp, 2023) . The order of the blocks will be used to define transition probabilities for a high level semantic decision process.

6 Limitations

We have to stress that because little is known about semantic structure being built by both partners during explanations, we followed an explorative approach. In our current analysis we excluded backchannels because they do not attempt to take the conversational floor. Nevertheless, backchannels might contribute to the dialogue. We attempted

Current Node *t*	**Next Node** *t+1*	**Frequency**
For Opponent	For Yourself	55
Square	Round	40
Vertical	Horizontal	35
For Opponent	For Opponent	32
Light	Dark	29
Hollow	Round	25
Tall	Small	23
For Yourself	For Yourself	23
Quarto	Parlour Game	21
Individuality	Individuality	21

Table 3: Explanation nodes with a cooccurence > 20

to find clusters in the explanation nodes by seeking high frequent bigrams, to not only see patterns in the block, but also in the explanation node sequences. The combinations in Table 3 were the ones that appeared more than 20 times. Some are repetitions of the same explanation node which can be interpreted as a deeper discussion of a particular explanation node. The others with a strong semantic connection are the (contrastive) characteristics for the figures and the categorisation that Quarto is a parlour game. With the exception of these bigrams, we were unable to find sequential patterns on the explanation node level. This can be either due to the interlocutors' co-construction or due to the size of the dataset. Following the first assumption, it could be that in expert explanations that occur without the involvement of the EE and in a more monological form (Klein, 2009), more patterns on the explanation node level can be found. With the current data size and method, we cannot provide clear indications. It might be possible to find patterns on the explanation node level within the explanation of the EX, if one controls the behaviour of the EE. Thereby, the influence of the EE on the explanation dialogue can be minimised. As we analyse only a subset in this paper, a next step is to expand the analysis to the whole study and with similar data from other projects.

Ethics Statement

The study with adult participants was approved by the Paderborn University Ethics Committee.

References

Sule Anjomshoae, Amro Najjar, Davide Calvaresi, and Kary Främling. 2019. Explainable Agents and Robots: Results from a Systematic Literature Review. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems*, volume 19, Montréal, Canada.

Agnes Axelsson and Gabriel Skantze. 2023. Do You Follow?: A Fully Automated System for Adaptive Robot Presenters. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 102–111, Stockholm Sweden. ACM.

Nils Axelsson and Gabriel Skantze. 2020. Using knowledge graphs and behaviour trees for feedback-aware presentation agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, pages 1–8, Virtual Event Scotland UK. ACM.

Hubert Baniecki, Dariusz Parzych, and Przemyslaw Biecek. 2023. The grammar of interactive explanatory model analysis. *Data Mining and Knowledge Discovery*.

Kai Biermeier. 2023. ELAN Annotation Frequency and Coverage: Commented Release.

Peter Carey. 2018. *Data protection: a practical guide to UK and EU law*, fifth edition edition. Oxford University Press, Oxford, United Kingdom.

Michelene T. H. Chi, Marguerite Roy, and Robert G. M. Hausmann. 2008. Observing Tutorial Dialogues Collaboratively: Insights About Human Tutoring Effectiveness From Vicarious Learning. *Cognitive Science*, 32(2):301–341.

Herbert H. Clark. 1996. Common Ground. In *Using Language*, pages 92–121. Camebridge Usiversity Press, Cambridge UK.

Herbert H. Clark and Edward F. Schaefer. 1989. Contributing to Discourse. *Cognitive Science*, 13(2):259–294.

Christina Dideriksen, Riccardo Fusaroli, Kristian Tylén, Mark Dingemanse, and Morten H. Christiansen. 2019. Contextualizing Conversational Strategies: Backchannel, Repair and Linguistic Alignment in Spontaneous and Task-Oriented Conversations.

Mennatallah El-Assady, Wolfgang Jentner, Rebecca Kehlbeck, Udo Schlegel, Rita Sevastjanova, Fabian Sperrle, Thilo Spinner, and Daniel Keim. 2019. Towards XAI: Structuring the Processes of Explanations. *Proceedings of the ACM Workshop on Human-Centered Machine Learning, Glasgow, UK*, 4:13.

Jennifer D. Ewald. 2010. "Do you know where X is?": Direction-giving and male/female direction-givers. *Journal of Pragmatics*, 42(9):2549–2561.

Josephine B. Fisher, Vivien Lohmer, Friederike Kern, Winfried Barthlen, Sebastian Gaus, and Katharina J. Rohlfing. 2022. Exploring Monological and Dialogical Phases in Naturally Occurring Explanations. *KI - Künstliche Intelligenz*, 36(3-4):317–326.

Martin Haspelmath. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. *The new psychology of language*, pages 217–248.

Mayank Kejriwal, Craig A. Knoblock, and Pedro Szekely. 2021. *Knowledge Graphs: Fundamentals, Techniques, and Applications*. MIT Press.

Josef Klein. 2009. Erlären-Was, Erklären-Wie, Erklären-Warum: Typologie und Komplexität zentralier Akte der Welterschließung. In R. Vogt, editor, *Erklären: Gesprächsanalytische und fachdidaktische Perspektiven*, pages 25–36. Stauffenburg Verlag, Tübingen.

H. Kotthoff. 2009. Erklärende Aktivitätstypen in Alltags- und Unterrichtskontexten. In J. Spreckels, editor, *Erklären im Kontext. Neue Perspektiven aus der Gesprächsund Unterrichtsforschung*, pages 120–146. Schneider, Germany.

Richard J. Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–74.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1).

Yi Ma, Paul A. Crook, Ruhi Sarikaya, and Eric Fosler-Lussier. 2015. Knowledge Graph Inference for spoken dialog systems. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5346–5350, South Brisbane, Queensland, Australia. IEEE.

Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. 2019. A Grounded Interaction Protocol for Explainable Artificial Intelligence.

Tim Miller. 2018. Explanation in Artificial Intelligence: Insights from the Social Sciences.

Marvin Minsky. 1968. *Semantic Information Processing*. MIT Press.

Naomi Miyake. 1986. Constructive Interaction and the Iterative Process of Understanding. *Cognitive Science*, 10(2):151–177.

George Psathas and Kozloff Martin. 1976. The Structure of Directions. *Semiotica*, 17(2).

Amelie Sophie Robrecht and Stefan Kopp. 2023. SNAPE: A Sequential Non-Stationary Decision Process Model for Adaptive Explanation Generation. In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence*, ICAART '23. SCITEPRESS.

Katharina J. Rohlfing, Philipp Cimiano, Ingrid Scharlau, Tobias Matzner, Heike M. Buhl, Hendrik Buschmeier, Elena Esposito, Angela Grimminger, Barbara Hammer, Reinhold Hab-Umbach, Ilona Horwath, Eyke Hullermeier, Friederike Kern, Stefan Kopp, Kirsten Thommes, Axel-Cyrille Ngonga Ngomo, Carsten Schulte, Henning Wachsmuth, Petra Wagner, and Britta Wrede. 2021. Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI Systems. *IEEE Transactions on Cognitive and Developmental Systems*, 13(3):717–728.

Sonja Stange and Stefan Kopp. 2020. Effects of a Social Robot's Self-Explanations on How Humans Understand and Evaluate Its Behavior. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '20, pages 619–627, Cambridge United Kingdom. ACM.

Sonja Stange and Stefan Kopp. 2021. Effects of Referring to Robot vs. User Needs in Self-Explanations of Undesirable Robot Behavior. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '21, pages 271–275, Boulder CO USA. ACM.

Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

W3C. 2012. OWL 2 Web Ontology Language Document Overview (Second Edition). *OWL 2 Working Group*.

Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a Professional Framework for Multimodality Research. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*.

A Appendix

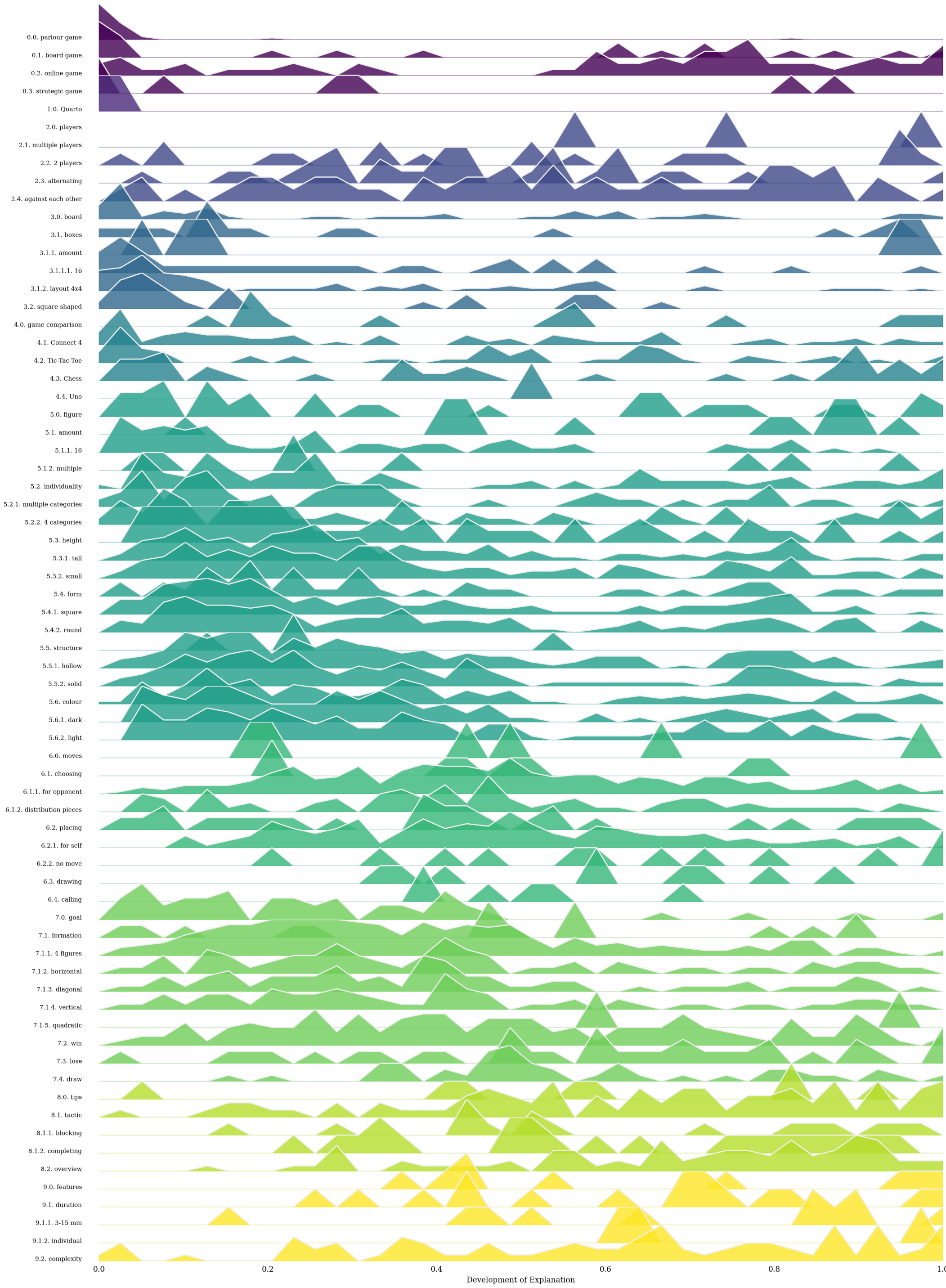


Figure 6: Reference to explanation nodes by EX in relation to the time over all dialogues

Adaptive explanations: the role of explainees' substantive moves

*Josephine B. Fisher and Katharina J. Rohlfing in Text & Talk: An Interdisciplinary Journal of Language, Discourse & Communication Studies

1 Abstract

Dialogues require both interaction partners to adapt to jointly contribute to an interactional goal, often referred to as common ground or alignment. However, we still know little about the substantive role individual interaction partners play in it. Explanations constitute an interesting case of asymmetric interactions with the goal of filling knowledge gaps. Assuming that both interaction partners contribute to it and adopting the methodology of micro-analysing individual contributions in our explanations, we hypothesised that the explainees' substantive moves enhance the semantic content of the explanation and guide the explainer to adaptive behaviour. To test this, we conducted a game explanation study with two conditions: a semi-naturalistic setting and one with a controlled, passive explainee. To reveal the effects of the explainee, study 1 (N = 128) was contrasted with study 2 (N = 44). We performed a discourse analysis of the verbal behaviours (moves) and the semantic content (explanation nodes). Comparing both conditions, the results show that less semantic content and fewer moves occurred without substantive involvement of the explainees. We conclude that with substantive moves by the explainee, explainers adapt their content and behaviours better. Hereby, the explainee actively shapes and thus co-constructs the explanation with the explainer.

Keywords: explanations, adaptation, co-construction, explainee, substantiveness

*Added to the anonymous version which is currently under review

2 Introduction

Adaptation during a dialogue has been characterised as a grounding process occurring between two interaction partners to achieve a shared goal (Clark 1991). In this vein, Clark and Schaefer (1989) have suggested that both interaction partners have to establish their common ground continuously in a dialogue. This means that in the conversation, they have to agree on their shared information. Continuing on this, all types of discourse constitute a joint activity in which the common ground expands (Clark 1996). The matter is complicated when facing an asymmetric interaction, in which there is one person knowing more than the other. It occurs that in such a situation, the more knowledgeable partner provides knowledge to the less knowledgeable partner. This is the case for explanations: Interaction partners have a knowledge asymmetry (Quasthoff et al. 2017). Therefore, explanations aim to arrive at understanding (Quasthoff et al. 2017). Many approaches acknowledge that an explanation is important in achieving knowledge symmetry (Clark and Schaefer 1989; Chi et al. 2001). Taking a co-constructed view, it has also been suggested that not only the explainer, the person explaining something, is vital for a successful interaction. In fact, the explainee, the person who is addressed, also takes an active role (Fisher et al. 2022). However, so far, it has barely been investigated to what extent the role of the explainee is significant. Both interaction partners negotiate and adjust the knowledge gap throughout the explanation and it is not fixed beforehand (Rohlfing et al. 2021). In this paper, we aim to investigate the verbal adaptation and take a closer look at the explanadum, the semantic content that is explained. Whether the explainer and explainee jointly co-construct the different elements, can be examined by the investigation of how the semantic content (i.e., the coverage of topics) and speaker moves develop.

Taking a look at the relevant literature addressing how speakers negotiate and adapt to each other when talking, both common ground and alignment are important. Following the interactive alignment account by Pickering and Garrod (2004), alignment at the lexical or syntactic level, so-called low-level alignment, leads to alignment at higher levels, such as semantics, phonology and eventually to the highest alignment of situational models. It therefore seems that lexical alignment is a driving force to jointly establish implicit common ground (Pickering and Garrod 2004) that in turn is the basis for interactive adaptation. Central to this synergistic approach is the view of dialogue as a self-organising and emergent interpersonal system that facilitates coordinated and functional interaction.

From this standpoint, task-oriented interactions, aimed at a shared goal consist of two phases: (1) presentation, where one interaction partner proposes ideas, and (2) acceptance, where the listener provides feedback on their understanding. The phases can include multiple turns. Through this iterative process, the two individuals gradually establish a shared knowledge base (common ground), which both participants recognise as mutual understanding (Clark and Schaefer 1989).

To shed light on how a common ground is achieved, we take the literature from tutoring dialogues as a starting point but extend it to include not only speaking strategies (Chi et al. 2001) but also strategies in providing feedback. More specifically, we investigate the concept of "speaker moves" (Chi et al. 2008: 323) among both interaction partners. Hereby, we especially focus on the explainee and whether their speaker moves are *substantive* contributions with respect of how they contribute to the dialogue and push the joint knowledge construction forward. A detailed description of the substantiveness and speaker moves follows in the next section.

2.1 Substantiveness

The concept of substantiveness is taken from the field of tutoring studies by Chi et al. (2001, 2008). Tutoring and explaining have in common that often one person has more domain knowledge than the other person. A knowledge asymmetry exists, which from a co-constructive perspective, fosters joint knowledge construction (Quasthoff et al. 2017). Explanations aim to identify and close knowledge gaps (Rohlfing et al. 2021). In the literature, mostly scientific explanations were considered which typically focus on answering why-questions to convey causal knowledge (e.g., Chi et al. 1994). Everyday explanations, in contrast, address a broader range of questions (what, where, how) to meet diverse informational needs (Klein 2009). Tutoring aims at knowledge acquisition, often to prepare for tests, whereas everyday explanations, like explaining a game, prioritise simpler and domain-specific understanding. This difference in goals shapes the nature of the interaction.

It is beneficial to see how different types of engagement by the tutee are characterised in these settings to describe joint knowledge construction – a co-constructive way of tutoring. “Substantive moves” (Chi et al. 2008) capture in how far the tutee contributes meaningful content verbally to the topic under discussion. For example, they can ask a content-related question. Explanations are comparable to tutoring. In particular, the explainee asks whether the figures are available for both players (Explainee: “There are sixteen figures for both players?”). In doing so, the explainee clearly addresses the matter of the distribution of figures among the players. Thereby, more concrete information is given by the explainer regarding the distribution of the figures, which in turn enriches the covered domain knowledge. The substantive vs. non-substantive moves are indicative of different modes of engagement and distinct processes of knowledge construction – this directly connects to the ICAP framework behind the investigation of the substantive contributions (Chi and Wiley 2014). They propose that in interaction, partners exchange ideas. In doing so, both partners jointly take part in it by their substantive contributions. Several assumptions underlie the ICAP framework, such as that verbal contributions need to be content relevant. Therefore, the verbal contributions need to be analysed in detail. Because of crucial learning and engagement effects, Chi et al. (2001) make a binary differentiation in the knowledge construction of the verbal contributions by the tutee: substantive moves are meaningful utterances to an ongoing activity. On the one hand, examples are: reading sentences aloud, giving unprompted self-explanations, asking questions, answering questions, responding to scaffolding prompt by tutor and reflecting (Chi et al. 2008). On the other hand, non-substantive moves with lower knowledge construction which do not contribute substantially to an explanation are: continuer, repetitions, agreement, and off-task remarks by tutee (Chi et al. 2001).

To transfer this approach to explanations, some adjustments needed to be made. We will discuss them in the method section of study 1. We propose that substantive moves are important to understand the co-construction of specific dialogues, and their analysis will provide novel insights into how joint knowledge is established.

In a general sense, the substantiveness, thus, depicts the visible (audible) knowledge construction of the tutee or explainee. It not only measures whether the tutee or explainee is verbally involved, but also the quality of the engagement. Chi et al. (2001) suggest that the effectiveness of tutoring sessions not only derived from the tutor's pedagogical skills, but also from the tutee and their substantive construction. The substantive contributions thereby play an important role in the interaction to achieve learning – in the case of tutoring – or understanding, in the case of explaining.

2.2 Semantically related content

In order to investigate the establishment of common ground in interactions more thoroughly, in tutoring settings, Miyake (1986) and Chi et al. (2008), categorised the emerging topics that the interaction partners addressed, in the so-called problem-solving steps, as mind maps with different nodes depicting the different topics. The limitation of these categorisations of the contents is that it was not based on the contents addressed in the tutoring sessions by the partners in a bottom-up manner. Instead, it was determined top-down by the tutor explaining alone (Chi et al. 2008) or the hierarchical functioning of a device (Miyake 1989). Nevertheless, the top-down approach laid the basis for investigating the knowledge construction of the tutee. Additionally, in Chi et al. (2008), it was analysed to what extent the partners jointly 'covered' the semantic content (in the form of nodes). The first node coverage includes the nodes which were addressed only by each speaker, and the latter the nodes which were jointly addressed by the speakers. Both types of "node coverages" (Chi et al. 2008: 333) were found to correlate with learning of the tutee. Therefore, the node coverage refers to the breadth of semantic content.

In the context of game explanations, Fisher et al. (2023) analysed the semantic content via a node system. In comparison to Chi et al. (2008), they have a more fine-grained and bottom-up analysis of the semantics which was possible because of the detailed coding schema of explanation nodes and the rather closed domain of the game Quarto (Fisher et al. 2023). The system was developed to make the explanation content explicit. Labelling the verbal data of the explanations with the different semantic contents.

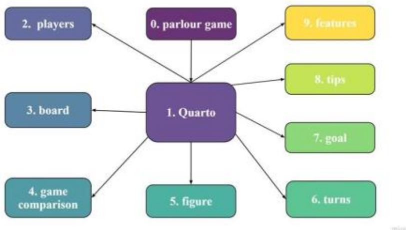


Figure 1: Blocknodes from the game explanations of Quarto (Fisher et al., 2023)

As argued in Fisher et al. (2023), one needs to connect the explanation nodes with the speaker moves, of the interaction partners to provide a more detailed basis for adaptive systems. In this paper, we aim to do so. We propose that speakers negotiate their verbal adaptation directed towards each other via their speaker moves.

The approach to assess the semantic content provides an opportunity to better evaluate the substantive moves of the explainee, and thus, the explainees' role in the explanation process. Following the above-mentioned co-constructive view on explanations (Quasthoff et al. 2017; Rohlfing et al. 2021), we hypothesised that the substantive moves of the explainee enhance the semantic content of the explanation and guide the explainer to adaptive behaviours in form of diverse speaker moves.

3 Method ADEX Corpus

The above-mentioned hypothesis was tested on verbal behaviours gained from the ADEX corpus. Below, we first describe the corpus before then moving to the study design.

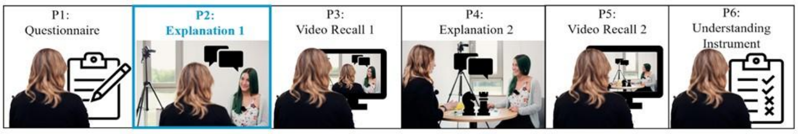


Figure 2: Study Design of ADEX Corpus study 1

The ADEX (Adaptive Dialogical Explanations) corpus (Buhl et al. 2024) includes dyadic video-recorded game explanations of Quarto. The aim of the corpus is to examine the adaptation processes of the interaction partners on a verbal and psychological level. To investigate this matter, the first two studies of the ADEX corpus will be compared. Whereas the first one elicits semi-naturalistic dialogues, in the second study, we controlled for the explainees' behaviours to see the impact on the explainer, testing whether the adaptation is hindered when one of the partners does not provide meaningful verbal feedback. The studies included three (study 2) to six phases (study 1). In both studies prior to the dialogue, the explainer was instructed to familiarise themselves with the game Quarto. Phase 1 and 3 consisted of various questionnaires that incorporated a sociodemographic questionnaire as well as psychological and comprehension assessment tools. Phase 2 was the explanation phase, on which the following two studies are based. In Phase 4, an explanation with the game being present took place, in which the interaction partners were instructed to continue explaining while playing the game. This was followed again by assessment tools (Phase 5) and video-recalls (Phase 6).

3.1 Study 1

3.1.1 Participants

In our study 1, we considered 64 game explanation dyads, total of 128 participants, from the ADEX corpus (Buhl et al. 2024). This included 123 L1 and 5 L2 German speakers (mean age: 25 years)[1]. The participants were either paid in money or in credit points for a university course. These interactions were administered, audio- and video-recorded, and lasted 5.5 minutes (min. 02:16 minutes, max. 12:24 minutes, *SD* = 0.1) and were therefore semi-naturalistic.

3.1.2 Procedure

In study 1, both participants were free in their verbal behaviour and there were no time restrictions for the task. However, in Phase 2, for the explanation without the game being present, the explainee was instructed to actively take part in the explanation. In addition, the explainer, was instructed to provide the explainee with an explanation by which they would be able to win the game after the explanation.

3.1.3 Verbal coding

The verbal behaviour of the interaction partners was coded with ELAN (Wittenburg et al. 2006). The analysis is fourfold and takes place on several levels. We wanted to analyse the explanation strategies the so-called "speaker moves" (Chi et al. 2008: 323). These consist of an utterance which includes a single idea by a speaker within a turn (Chi et al. 2008). In order to better classify the

[1]One data point is missing due to technical problems.

single idea, first, the semantics of the utterances were examined. For this, the explanation nodes serve as a basis for the analysis of the speaker moves. Further, when focusing on the explainee and the categorisation of their contributions, their substantiveness is based on the speaker moves.

We developed a speaker move coding scheme[2] in order to classify the verbal behaviour of the interaction partners. We mainly based our coding scheme for the speaker moves on tutoring literature (Chi et al. 2001, 2004, 2008; Tare et al. 2011) and extended the repertoire for the explainee (Fisher et al. 2024). In the following, a selection of relevant speaker moves is presented.

When information is provided for the first time, the speakers are *providing information* (Chi et al. 2001). Extraneous information to one of the previously introduced contents is then *additional information* (Chi et al. 2001). These moves are directly connected to the topic and are hence substantive. With respect to other speaker moves, we deviate from Chi et al. (2001, 2008) to better account for the explainee's behaviours. For example, Chi et al. (2001) do not differentiate between different types of questions by the explainee. The differentiation of the question types is taken from Tare et al. (2011) who consider child parent interaction in exhibitions. Focussing on the most prominent, namely *factual question*, addresses any topic related question regarding the game.

Information cannot only be questioned, but also previous information can be taken up again. One of those substantive moves is *paraphrasing*. Here, previous information is put into other words. As Chi et al. (2001) refer to repeats as non-substantive and note that one of the reasons that tutees learn while observing a tutoring session is that the explanation content was paraphrased or additional information was given, we added both categories to our coding scheme of substantive speaker moves. More specifically, repeats only remain in our coding as non-substantive when the

[2] The complete coding scheme with examples for each move can be found on: https://osf.io/twza4

previous information is literally repeated. By differentiating between *repeating* and *paraphrasing,* the different degrees of alignment become visible: From the perspective of interactive alignment (Pickering and Garrod 2004), repeats display lexical alignment (lower level) and paraphrases semantic alignment (higher level). Additionally, we also differentiated the reference of the repeats or paraphrases. They could either refer to the topics the speakers addressed themselves *(self)*, their partner had brought up previously *(partner)* or which contents were already covered by both interaction partners *(both)*. Taking the reference into account is inspired by the literature on alignment, regarding the establishment of shared lexicons (Dubuisson Duplessis et al. 2021). In doing so, one is aware of the direction of the alignments and thereby adaptations.

On the more personal level, two further moves are important for the substantiveness, namely *mentalising* and *providing personal information*. In the direction of knowledge reflection, *mentalising* captures this as the explainees give insights into their perceived knowledge state. In addition, it can also address a wish, goal or emotional state. *Providing personal information* refers to any preferences and experiences of the individual speakers. With regard to the game, for example, it can be the information about what other similar games the partners know.

Beyond classifying the different types of moves, they can also be described quantitatively. This includes, the total frequency of moves as well as the number of different moves. Later referred to as *move diversity*. This can be essential as it reflects the repertoire of behaviours with which a speaker can adapt.

3.1.4 Explanation nodes

In order to enable the analysis of the speaker moves, the first step was to examine the explanation nodes or so-called semantic content as developed in Fisher et al. (2023). The analysis of the nodes provides insights regarding the node frequency and the node coverage. In this paper, we focus on the node coverage that displays the variety of nodes addressed in the explanation (Chi et al. 2008). This can also be referred to as the breadth of semantic content.

Explanation nodes, from here referred to as nodes, were assigned to each utterance to support the speaker move (content level) analysis. The mentioning of multiple nodes in an utterance was possible. For example, when talking about the figures, already two nodes are addressed when saying they are tall and small. The nodes were also used to keep track of when the interaction partners talked about certain topics. This enabled detailed comparisons of the utterances. Hereby, one could differentiate, for example, whether something was repeated (same words) or paraphrased (put into different words). Moreover, also topic initiations and uptakes by the interaction partners could be investigated. For more details including the development of the coding scheme, see Fisher et al. (2023).

3.1.5 Substantiveness score

In order to apply the concept of substantiveness, the nodes and speaker moves need to be analysed beforehand. Based on this analysis, the utterances are classified as being substantive or non-substantive. Determining the substantiveness of a contribution, is two directional. In the case, in which the utterance is not content-related, the utterance is non-substantive. In the case, in which the utterance is content-related, the utterance needs to be analysed in further detail. For example, if the content was repeated literally, the utterance is non-substantive. However, if the content is put into different words (paraphrased), they display a higher level of alignment and it is substantive. The intercoder reliability of the substantiveness was measured by considering 10% of the explanations of study 1 by two independent coders. It resulted in a Cohens Kappa: 0.87, which displays a near perfect coding agreement.

To measure the substantive involvement by the explainee, we transformed the level of substantiveness into a score. The substantiveness score includes the total frequency of substantive speaker moves by the explainee. It shows in how many instances the explainee makes substantial contributions within the explanation. The use of the total frequency rather than the relative frequency is justified because it can be argued that each substantive move has an influence on the explanation. Additionally, if one were to put it in ratio with the non-substantive moves, speakers who make use of both types of moves would end up with a lower score in comparison to speakers that only contribute substantively to the explanation once. For example, Speaker 1 produces one substantive and one non-substantive move resulting in a score of 0.5, while Speaker 2 with one substantive move and no non-substantive move achieves a score of 1. However, this approach fails to accurately capture the dynamics of the conversation. By focusing solely on the total number of substantive contributions, the analysis emphasises the explainee's frequency of meaningful contributions, without confounding the results by including non-substantive moves.

3.2 Results

In the following section, the results of the explainees verbal behaviour in the semi-naturalistic study 1 will be reported. We begin with the semantic content that is covered by the explainee, the so-called, node coverage. The analysis of the semantic content is followed by the results of the substantive moves and the resulting substantiveness score. Finally, we examine the relationship of the semantic content and move diversity. In doing so, we aim to achieve a better insight into the adaptivity of the verbal behaviours.

3.2.1 Joint moves and nodes

As mentioned, our nodes were coded according to a bottom-up schema. Overall, there was no single explanation that covered all nodes. Instead, the interaction partners jointly covered 50.36 % of nodes (min. 33.33%, max. 69.57%, *SD* = 8.58%) by jointly using 54.55 (min. 23, max. 129, *SD* = 23.95) speaker moves on average.

3.2.2 Explainee's and explainer's contributions

In this section, the individual contributions to the node coverage will be examined. More specifically, explainees covered 20.77% (min. 2.9%, max. 46.38%, *SD* = 10.94%) of the nodes in the explanations. Whereas this might appear to be a small part, it could have a crucial effect on the explainer presenting the different topics. The nodes formed the basis for the move analysis. Accordingly, we found that all explainees had substantive moves in their verbal behaviour. Focusing on the different types of substantiveness, the explainees used 4.16 different kinds of substantive moves (min. 1, max. 9, *SD* = 1.84) on average in each explanation. In Table 1, the substantive moves are listed in order of frequency. Interestingly, the four most frequent moves

Move	**Average (*N* = 64)**	**Min.**	**Max.**	**SD**
Providing info	0.36	0	2	0.63
Providing personal info	0.45	0	2	0.64
Summarising info	0.59	0	5	0.99
Additional info	1.58	0	12	2.16
Example	0.11	0	2	0.44
Mentalising	2.36	0	10	2.16
Focus monitoring	0.03	0	1	0.18
Paraphrasing self	0.17	0	2	0.46
Paraphrasing partner	1.75	0	6	1.71
Paraphrasing both	0.38	0	5	0.92
Label question	0.05	0	1	0.21
Factual question	5.92	0	20	4.46
Comprehension question	0	0	0	0

constitute 84% of the total substantive moves.

Table 1: Substantive explainee moves study 1 average frequency

In addition, analysing also the non-substantive contributions of the explainees, we can report that only 3.42 (min. 0, max. 15, *SD* = 3.14) of their speaker moves were non-substantive with a move diversity of 1.89 (min. 0, max. 4, *SD* = 1.16). We can summarise that the majority of moves that the explainee provided in our setting were substantive.

Based on our analyses, we applied a substantiveness score to account for individual differences in the way how the explainees were engaged in the explaining process. The individual scores of the explainees in each data set are visualised in Figure 3. It becomes clear that a high variance is displayed in the scores. On average, the explainees scored 13.83 (min. 2, max. 43, *SD* = 8.97) in their substantiveness.

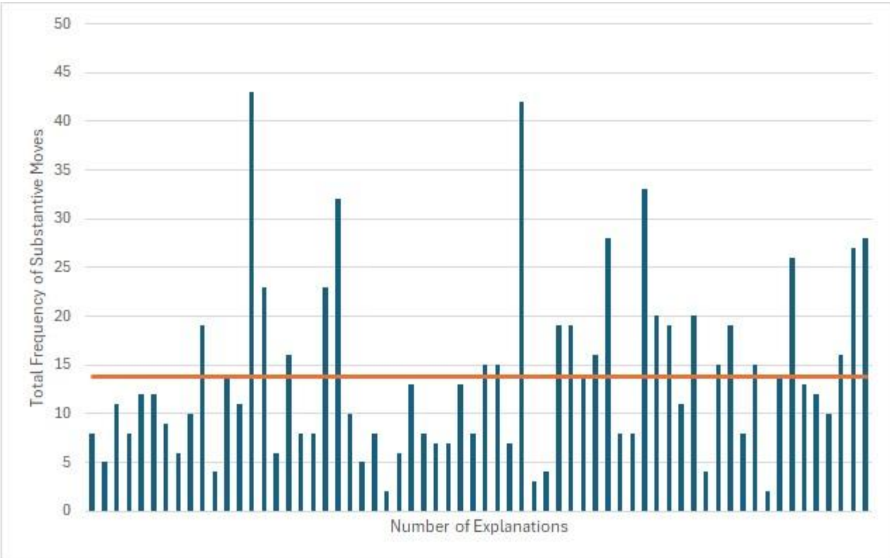


Figure 3: Explainee substantiveness score

Furthermore, the score is not normalised according to the length of the explanations because there is no evidence to believe that the longer the explanation is, the more the explainee gets involved. It rather shows the diverse range of degrees of involvement by the explainees.

To measure the knowledge construction by explainees, we examined their speaker moves and their level of substantiveness in relation to the node coverage. As previously introduced, we hypothesised that the substantive moves will address a larger number of nodes than non-substantive moves. When comparing the difference between substantive and non-substantive moves, we found that substantive moves correlate with the coverage of nodes (*r* = .86) to a greater degree than non-substantive moves (*r* = .609). Simple linear regression was used to test whether the substantiveness of moves predicts the node coverage. It was found that substantive moves significantly predicted the coverage of nodes (β = .80, p = <.001). Importantly, non-substantive moves did not predict the coverage of nodes (β = .13, p = <.081) in a significant manner. The overall regression was statistically significant (R^2 = .79, *F*(2, 61) = 111.630, p = <.001). The results thus confirm our hypothesis.

Another important aspect contributing to the coverage of overall nodes, and thus, to the quality of explanation, is the diversity of speaker moves. Therefore, the relationship between the move diversity and node coverage by the explainee was investigated in more detail. In assessing adaptivity of the partners with respect to the joint goal, we have proposed that the more diverse the moves are, the higher the node coverage by the explainee. Regarding the move diversity, explainees disposed of a repertoire of on average 6.05 different moves (min. 2, max. 12, *SD* = 2.58). A high correlation between the move diversity and node coverage was found ($r_s = .76, p < .001$). Therefore, the hypothesis can be verified.

For the explainers, we found that they covered 48.69% (min. 33.3%, max. 68.12% *SD* = 8.05%) of the nodes on average. More specifically, Table 2 presents the average frequencies of the explainers' speaker moves. The explainer had on average 37.3 speaker moves (min. 17, max. 75, *SD* = 13.79).

Table 2: Explainer moves study 1 average frequency

Move	**Average (*N* = 64)**	**Min.**	**Max.**	**SD**
Providing info	6.86	4	10	1.48
Providing personal info	0.59	0	4	1.05
Summarising info	0.97	0	4	1.13

Additional info	13	5	25	4.65
Example	1.98	0	7	1.86
Mentalising	2.23	0	9	2.38
Focus monitoring	0.02	0	1	0.13
Repeating self	0.19	0	2	0.47
Repeating partner	0.17	0	2	0.42
Repeating both	0.08	0	1	0.27
Paraphrasing self	5.77	0	20	3.97
Paraphrasing partner	0.56	0	5	0.89
Paraphrasing both	2.8	0	12	2.58
Label question	0.02	0	1	0.13
Factual question	0.23	0	3	0.66
Comprehension question	1	0	7	1.26
Reassurance question	0.03	0	1	0.18
Personal question	0.41	0	3	0.71
Procedure question	0.02	0	1	0.13

Turning to the diversity in the speaker moves of the explainer, their repertoire consisted of a set of 8.22 moves on average (min. 4, max. 14, *SD* = 2.47). In regard to the connection between the move diversity and the node coverage of the explainer, a significant and strong correlation exists ($r_s = .581$, $p <.001$). The results support our hypothesis.

Finally, we considered all results together. We follow up on the question which contributions pushed forward the quality of the explanation in terms of its comprehensiveness. To determine the key predictors of the joint node coverage, a stepwise regression analysis was conducted. The dependent variable was the joint node coverage, while the independent variables included the explainer speaker moves (frequency) and the substantiveness score of the explainee. The final regression model only considered the explainer move frequency ($\beta = 0.42$, $p < 0.001$) and joint node coverage ($\beta = .776$, $p < 0.001$). The model accounted for 60.2 % of the variance in joint node coverage ($R^2 = 0.602$, Adjusted $R^2 = 0.59$). The substantiveness score did not meet the significance threshold ($p > .10$) for inclusion and was excluded from the model.

To conclude, we found that the explainer covered more nodes than the explainee, which confirms the knowledge asymmetry in explanations. Furthermore, a strong positive correlation between move diversity and node coverage suggests that more diverse moves help to address a broader range of topics. This mirrors the findings found for the explainee. Additionally, the key predictor for the joint node coverage was the number of explainer speaker moves.

4 Study 2

Above, our analyses yield correlational results, from which we learn little about the effect of partners' contributions. In order to examine whether the involvement by the explainee has an effect on the explanation, we conducted study 2. Specifically, the study controls the explainee's behaviours and examines how this changes the explainer's speaker moves, their frequency, and diversity, as well as the number of explanation nodes addressed.

4.1 Method

4.1.1 Participants

In study 2, we conducted 41 dyadic game explanations as part of the ADEX no-response corpus. The participants had a mean age of 24 years. Of the 44 total participants, 41 were explainers, and were paired with three pre-trained explainees. Most of them were students (42), with two participants with other occupations. 41 participants were L1 German speakers, and three were L2 German speakers. The explanations lasted 2.5 minutes (min. 00:50 min, max. 04:49 min, *SD* = 0.04).

4.1.2 Procedure

The study consisted of the first three phases of study 1. For the control condition, explainees were restricted using backchannels in a neutral manner. As defined by Dideriksen et al. (2019), backchannels include head nods or brief one-word utterances such as "uh-huh," "yes," or "okay." These backchannels do not function as independent turns. The explainees were jointly trained, using a coding scheme for backchannel behaviours. On average, the trained explainees provided 1 (min. 0, max. 5, *SD* = 1.11) substantive move in each explanation and covered 0.53% (min. 0%, max. 11.59%, *SD* = 1,89%) of the explanation nodes. Therefore, the manipulation of the explainee worked well as they barely addressed any nodes. This becomes even clearer when taking a look at the node coverage in study 1 (see Table 4). The moves that were used were *providing personal information, additional information* and *mentalising* (see Table 3). Only once a non-substantive move was used by an explainee (average: 0.02, min. 0, max. 1, *SD* = 0.16). In contrast to study 1 with an average of 3.4 non-substantive moves, there are also fewer non-substantive moves in study 2. The majority of verbal behaviour included backchannels which are not defined as moves and thus are not coded according to their substantiveness.

Table 3: Substantive explainee moves study 2 average frequency

Move	**Average (*N* = 41)**	**Min.**	**Max.**	**SD**
Providing personal info	0.24	0	2	0.58
Additional info	0.02	0	1	0.16

Mentalising	0.73	0	4	0.81

Based on the substantive moves, the explainee had a substantiveness score of 1 (min. 0, max. 5, *SD* = 1.11).

4.1.3 Verbal coding

The coding of the data included the same steps as in the previous study 1. It included the analysis of the semantic content in forms of nodes, of the speaker moves, and the substantiveness.

4.2 Comparison results study 1 with study 2 verbal behaviour explainee

In the following, the results of study 2 will be compared directly with the results from study 1, starting with a general description of the dyads.

In study 2, the interaction partners jointly used 19.12 speaker moves (min. 5, max. 35, *SD* = 5.53). According to an independent t-test ($t(73.38)$[3] = 11.36, $p < .001$) that we conducted considering 54.55 (min. 23, max. 129, *SD* = 23.95) moves from study 1, this is significantly less. It shows that overall, the dialogues in study 2 were shorter. The difference pertained also to the topics that the dyads jointly covered. As can be seen in Table 4, the joint node coverage was significantly less in study 2. Therefore, the general verbal behaviour within the dyads is less.

Table 4: Comparison study 1 and study 2

		Study 1	**Study 2**	**t-test**	**η2**

[3] The variances are not equal and this results in a smaller comparable group size.

Explainer	Node coverage Min. Max. SD	48.69% 33.33% 68.12% 8.05%	36.97% 20.29% 52.17% 7.36%	7.544*	0.356
	Average Moves Min. Max. SD	37.3 17 75 13.79	18.12 5 34 5.33	10.02*	0.532
Explainee	Node coverage Min. Max. SD	20.77% 2.9% 46.38 10.94%	0.53% 0% 11.59% 1.89%	-	-
	Average Moves Min. Max. SD	17.25 3 54 11.17	1 0 5 1.12	-	-
Joint	Node coverage Min. Max. SD	50.36% 33.33% 69.57% 8.58%	37.01% 20.29% 52.17% 7.08%	8.237*	0.397
	Average Moves Min. Max. SD	54.55 23 129 23.95	19.12 5 35 5.53	11.36*	0.638

Significant at a level of $p < .001$

Our analysis of study 2 will focus on the explainer. It begins with an examination of the explainers' node coverage, followed by the joint node coverage of both speakers. Next, the speaker moves of the explainer are explored in terms of their diversity and frequency. Finally, the connection of the node coverage and move diversity is examined.

The explainer mean node coverage was 36.97% of the explanation nodes in study 2 (see Table 4). The maximum of 52.17% is only a little higher than the average in study 1. According to an independent *t*-test, the decrease from 48.69 % to 36.97 % covered nodes ($t(103) = 7.544$, $p < .001$) is significant. In comparison, the results, thus, indicate that the explainer node coverage was negatively affected by the controlled verbal behaviour of the explainee. To reveal whether this effect impacted the overall comprehensiveness of the explanation, we analysed the joint node coverage and compared the results from study 1 with 50.36% with study 2, which is 37.01% (see Table 4). An independent t-test revealed that the difference was significant ($t(103) = 8.237$, $p < .001$). Consequently, this indicates that the joint node coverage was also negatively affected. This supports that explainees play a crucial role in addressing different topics within an explanation. Together, the results demonstrate that the involvement of the explainee led to a higher coverage of nodes, and thus to more comprehensive explanations.

We now turn to a more detailed account of the explainer verbal behaviour regarding their frequency of speaker moves. On average, the explainer used 18.12 speaker moves (min. 5, max. 34, $SD = 5.33$). Comparing both studies with each other, the average speaker moves are significantly less in study 2 ($t(88.26)$[4] $= 10.02$, $p < .001$). This indicates that the amount of speaker moves by the explainer depends on the verbal behaviour of the explainee. Interestingly, the top three moves, *additional information*, *providing information* and *paraphrasing self* remain the same as in study 1 (see Table 2 and Table 5). Furthermore, when taking a closer look at the standard

[4] The variances are not equal and this results in a smaller comparable group size.

deviation of the moves, these are smaller in study 2, suggesting that the verbal behaviour of the explainer is more consistent or patterned.

Table 5: Explainer Moves Study 2 Average Frequency

Move	**Average (*N* = 41)**	**Min.**	**Max.**	**SD**
Providing info	5.78	2	9	1.42
Providing personal info	0	0	0	0
Summarising info	0.29	0	2	0.51
Additional info	7.41	1	14	2.97
Example	0.93	0	4	1.19
Mentalising	0.56	0	5	0.95
Focus monitoring	0	0	0	0
Repeating self	0.1	0	1	0.3
Repeating partner	0	0	0	0
Repeating both	0	0	0	0
Paraphrasing self	1.32	0	7	1.51
Paraphrasing partner	0.02	0	1	0.16
Paraphrasing both	0	0	0	0
Label question	0.02	0	1	0.16
Factual question	0.07	0	2	0.35
Comprehension question	1.02	0	3	0.67
Reassurance question	0	0	0	0
Personal question	0.41	0	3	0.81
Procedure question	0.15	0	1	0.36

This observation is also reflected in the analysis of the move diversity. We found that in study 2, explainers disposed of a repertoire of only 5.2 different speaker moves (min. 2, max. 8, *SD* = 1.38). In study 1 the moves the explainer uses consist of a set of 8.22 moves on average (min. 4, max. 14, *SD* = 2.47). The difference between the two studies is highly significant according to a t-test ($t(101.447) = 8.022$, $p < 0.001$).

For the examination of the closer relationship between the move diversity and node coverage, we hypothesised earlier that a more diverse set of moves relates to more covered nodes. This relationship was not found statistically significant ($r_s = .318$). Therefore, in study 2, the relationship seems to be different.

Taken together, the minimal verbal engagement of the explainee in study 2 negatively affected the explainer's verbal behaviour, resulting in a reduced node coverage, fewer speaker moves, and lower move diversity.

5 Discussion

Whereas the notion of common ground (Clark 1996) points to the fact *that* during a dialogue, partners establish joint knowledge, little is known about *how* the partners contribute to it. Adopting a micro-analysis methodology originating from tutoring literature (Chi et al. 2001, 2008; Miyake 1986), we conducted a discourse analysis of explanations and investigated the speakers' verbal moves and how they were related to the emerging semantic content.

Our investigation followed (co-)constructive approaches (Rohlfing et al., 2021; Chi and Wiley 2014) to asymmetric interactions. Accordingly, for the first study, we hypothesised that not only the explainer but also the explainee will contribute more to the semantic content when performing substantive moves in comparison to non-substantive moves. Our analyses supported this hypothesis, revealing that there was no explanation that covered all semantic nodes confirming what is known from research on explaining (Sokol and Flach 2020) suggesting that explanations require adaptation. Our analyses further revealed that both partners contribute to the adaptation, not only the more knowledgeable explainer but also the explainee is actively involved. In fact, the substantive moves of the explainee significantly predicted the coverage of nodes (i.e., the breadth of semantic content), while non-substantive moves did not show a significant relationship with the coverage of nodes.

To better control for the impact of the explainee, in the second study, the explainee's behaviours were reduced to mostly backchanneling. We hypothesised that a lack of substantive moves by the explainee will impact the explainer's speaker moves and reduce the explanation quality in terms of covered nodes. The results confirmed that the limited involvement influenced the explainer, leading to fewer moves, their reduced diversity, and decreased the node coverage, thus, the explanation quality. When comparing the results across studies, we found more pronounced diversity in moves in the first study speaking to the possibility that explainers adapt to individual explainees' interests and progress in understanding, when they can contribute substantive content to the explanation process. This interpretation is supported by a strong positive correlation between move diversity and node coverage suggesting that more diverse moves led to a broader range of addressed topics.

A closer analysis of the explainees' substantive moves revealed, *factual questions* guide interactions toward the explainee's knowledge gaps, making them effective for achieving shared understanding. Fisher et al. (2024) highlight that question asking, particularly *label questions* related to the overall task and its goal. In contrast to earlier work, our findings reveal that *factual questions* were most common, while *label questions* were rare. This suggests that predefined knowledge assessment, such as previously announced tests, influences question types, aligning question-asking behaviour with the format of testing.

Another common move by the explainees was *mentalising*, which reveals various mental processes. By *mentalising*, speakers likely make their process of explaining or knowledge construction accessible to their partner. Research across disciplines highlights that metacommunicative behavior allows participants to convey what they know, do not know, or are interested in, serving as a way to monitor understanding (Näykki et al. 2017). This monitoring likely facilitates adaptivity.

Focussing on the explainers' moves in both studies, *additional information* and *providing information* were the most common. Indicating that the essence of explaining largely involves providing information within asymmetric interactions. Further, the third and fourth most common moves were *paraphrasing self* and *paraphrasing both*, which result from the explainer addressing many nodes. In this process, explainers are more likely to paraphrase themselves, while *paraphrasing both* indicates they also incorporate nodes previously addressed by the explainee. Thus, these two latter types of moves clearly add to the adaptive aspects of the explanation process.

Whereas with our results, we argue that the verbal contribution of the explainee provides important information for the explainer to adapt, there are other mechanisms for adaptation possible. Buhl et al. (2024) have proposed that substantive contributions of the explainee can stimulate cognitive changes of the partner models in explainers, according to which the latter tailor their explanations. Our analyses cannot rule out the possibility of cognitive adaptation. Thus, our results can only underline the importance of substantive contributions that was also established in tutoring (Chi et al. 2001, 2008) and extend it to the explanatory dialogues. Hereby, the joint nodes were useful in the operationalisation of common ground to detect the shared information. In this respect, we proved the usefulness of a developed quality measure for explanations – that in XAI literature (e.g. Sokol and Flach 2020) is considered as comprehensiveness.

Critically reflecting our method, we were crucially guided by the assumption that a higher node coverage (i.e., the breadth of the semantic content) is reflecting a higher explanation quality. This is justified in the comprehensiveness that explanations seek to achieve – as shown by research on explanations (e.g. Garfinkel 1981). However, Miller (2019) pointed out that explanations are inherently selective. Thus, we need to get better insights into how explanations advance the understanding in explainees and whether in everyday explanations, it is done by the breadth of semantic content or rather selected content (e.g., specific nodes) that relates tightly to the goal underlying the explanation. In future research, we will investigate whether some crucial nodes require certain moves.

Our approach has further limitations. First, we considered the frequency of moves in a quantitative manner. However, the temporal placement of the speaker moves within the explanation would also be interesting to consider. Certain moves may occur early in monological phases, while others may characterise dialogical phases (Fisher et al. 2023). The temporal placement could explain why the substantiveness score did not predict the joint node coverage, suggesting that both the frequency and timing of substantive moves are important. This aligns with Fusaroli et al. (2014) who claim that methods need to be established to capture the changing linguistic patterns in a dialogue. Consequently, less frequent moves might also contribute to the establishment of individual patterns or to larger patterns including several moves.

Second, we introduced the phenomenon of interactive alignment (Pickering and Garrod 2004) to highlight partner adaptation. Our current analysis focuses on *paraphrasing* and *repeating*, limiting a comprehensive examination of alignment. Other moves may also involve lexical, semantic, or syntactic alignment, which would require closer investigation. Despite this, considering alignment improved the analysis of substantiveness.

Third, the study 2 design may have been overly constrained. Limiting the verbal behaviour of the explainee to a minimum, might have resulted in interactions that felt rather unnatural. It might have been more effective to encourage greater use of non-verbal behavior or non-substantive moves.

Whereas we are aware of all these limitations, we have presented the first steps toward assessing the explainee's contributions in an explanatory dialogue that – so far – has been considered as explainer-driven. Our analysis revealed that explainees are contributing to the dialogue to a considerable degree by their nodes and substantive moves. It helps us to understand how a dialogue is a joint endeavour with both participants fulfilling dialogical roles that are relevant for the speaker moves and the semantic content.

6 Acknowledgements

The work in this paper was funded by the Deutsche Forschungsgemeinschaft (DFG): TRR 318/1 2021 438445824, project A01. We thank Viviane Gladow, Marie Weise, Celina Nitschke, Marco Verwiebe and Britta Meyer for supporting the data collection.

7 References

Buhl, Heike, Fisher, Josephine B. & Katharina J. Rohlfing. 2024. Changes in Partner Models Effects of Adaptivity in the Course of Explanations. In Samuelson, Larissa K., Frank, Stefan, Toneva, Mariya, Mackey, Allyson & Eliot Hazeltine. (eds.): *Proceedings of the Annual Meeting of the Cognitive Science Society* 46. 4976–4983.

Chi, Michelene T.H. 1996. Constructing Self-Explanations and Scaffolded Explanations in Tutoring. *Applied Cognitive Psychology* 10(7). 33–49.

Chi, Michelene T.H., De Leeuw, Nicholas, Chiu, Mei-Hung & Christian Lavancher. 1994. Eliciting self-explanations improves understanding. *Cognitive Science* 18(39). 439–477.

Chi, Michelene T. H., Roy, Marguerite & Robert G. Hausmann. 2008. Observing Tutorial Dialogues Collaboratively: Insights About Human Tutoring Effectiveness from Vicarious Learning. *Cognitive Science* 32(2). 301–341.

Chi, Michelene T. H., Siler, Stephanie A. & Heisawn Jeong. 2004. Can Tutors Monitor Students' Understanding Accurately? *Cognition and Instruction* 22(3). 363-387.

Chi, Michelene T. H., Siler, Stephanie A., Jeong, Heisawn, Yamauchi, Takashi & Robert G. Hausmann. 2001. Learning from human tutoring. *Cognitive Science* 25(4). 471–533.

Chi, Michelene T. H & Ruth Wylie. 2014. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist* 49(4). 219–243.

Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Clark, Herbert H. & Edward F. Schaefer. 1989. Contributing to Discourse. *Cognitive Science* 13(2). 259–294.

Dideriksen, Christina R., Fusaroli, Riccardo, Tylén, Kristian, Dingemanse, Mark & Morten H. Christiansen. 2019. Contextualizing Conversational Strategies: Backchannel, Repair and Linguistics Alignment in Spontaneous and Task-Oriented Conversations. *Proceedings of the 41st Annual Conference of the Cognitive Science Society: Creativity + Cognition + Computation.* 261–267.

Dubuisson Duplessis, Guillaume, Langlet, Caroline, Clavel, Chloé & Frédéric Landragin. 2021. Towards alignment strategies in human-agent interactions based on measures of lexical repetitions. *Language Resources & Evaluation* 55(2). 353–388.

Fisher, Josephine B., Lohmer, Vivien, Kern, Frederike, Barthlen, Winfried, Gaus, Sebastian & Katharina J. Rohlfing. 2022. Exploring Monological and Dialogical Phases in Naturally Occurring Explanations. *KI-Künstliche Intelligenz* 36. 317–326.

Fisher, Josephine B., Robrecht, Amelie S., Kopp, Stefan & Katharina J. Rohlfing. 2023. Exploring the Semantic Dialogue Patterns of Explanations-a Case Study of Game Explanations. *Proceedings of the 27th Workshop on the Semantics and Pragmatics of Dialogue*. 35–46.

Fisher, Josephine B., Rohlfing, Katharina J., Donellan, Ed, Grimminger, Angela, Gu, Yan & Gabriella Vigliocco. 2024. Explain with, rather than how to: How explainees shape their own learning. *Interaction Studies* 25(2). 244–255.

Fusaroli, Riccardo, Raczaszek-Leonardi, Joanna & Kristian Tylén. 2014. Dialog as Interpersonal Synergy. *New Ideas Psychology* 32. 147–157.

Garfinkel, Alan. 1981. *Forms of Explanation: Rethinking the Questions in Social Theory.* New Haven: Yale University Press.

Klein, Josef. 2009. Erlären-Was, Erklären-Wie, Erklären-Warum: Typologie und Komplexität zentraler Akte der Welterschließung. In R. Vogt (ed.) *Erklären: Gesprächsanalytische und fachdidaktische Perspektiven*, 25–36. Tübingen: Stauffenburg Verlag.

Miller, Tim. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267. 1–38.

Miyake, Naomi. 1986. Constructive Interaction and the Iterative Process of Understanding. *Cognitive Science* 10(2). 151–177.

Näykki, Piia, Järvenoja, Hanna, Järvelä, Sanna & Paul Kirschner. 2017. Monitoring makes a difference: Quality and temporal variation in teacher education students' collaborative learning. *Scandinavian Journal of Educational Research* 61(1). 31–46.

Pickering, Martin J. & Simon Garrod. 2004. The Interactive-Alignment Model: Developments and Refinements. *Behavioral and Brain Science* 27(2). 212–225.

Quasthoff, Uta, Heller, Vivien & Miriam Morek. 2017. On the sequential organization and genre-orientation of discourse units in interaction: An analytic framework. *Discourse Studies* 19(1). 84–110.

Reitter, David & Johanna D. Moore. 2014. Alignment and task success in spoken dialogue. *Journal of Memory and Language* 76. 29–46.

Rohlfing, Katharina J., Cimiano, Philipp, Scharlau, Ingrid, Matzner, Tobias, Buhl, Heike M., Buschmeier, Hendrik, Esposito, Elena, Grimminger, Angela, Hammer, Barbara, Häb-Umbach, Reinhold, Horwath, Ilona, Hüllermeier, Eyke, Kern, Friederike, Kopp, Stefan, Thommes, Kirsten, Ngomo Axel-Cyrille N., Schulte, Carsten, Wachsmuth, Henning, Wagner, Petra & Britta Wrede. 2021. Explanation as a Social Practice: Toward a Conceptual Framework for the Social Design of AI systems. *IEEE Transactions on Cognitive and Developmental Systems* 13(3). 717–728.

Sokol, Kacper & Peter Flach. 2020. Explainability fact sheets: a framework for systematic assessment of explainable approaches. *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency.* 56–67.

Tare, Medha, French, Jason, Frazier, Brandy N., Diamond, Judy, & E. Margaret Evans. 2011. Explanatory parent–child conversation predominates at an evolution exhibit. *Science Education* 95(4). 720–744.

Wittenburg, Peter, Brugman, Hennie, Russel, Albert, Klassmann, Alex, & Han Sloetjes. 2006. ELAN: A professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006).* 1556–1559.

Two Phases of an Explanation

Josephine B. Fisher[1*], Vivien Lohmer[2], Friederike Kern[2], Winfried Barthlen[3], Sebastian Gaus[3] and Katharina J. Rohlfing[1]

[1*]Faculty of Arts and Humanities, Paderborn University, Germany.
[2]Faculty of Linguistics and Literature, Bielefeld University, Germany.
[3]Clinic of Pediatric Surgery, Evangelisches Klinikum Bethel, Bielefeld, Germany.

*Corresponding author(s). E-mail(s): jbfisher@mail.uni-paderborn.de; Contributing authors: vivien.ebben@uni-bielefeld.de; friederike.kern@uni-bielefeld.de; winfried.barthlen@evkb.de; sebastian.gaus@evkb.de, katharina.rohlfing@uni-paderborn.de

Abstract

Recent approaches to Explainable AI promise to satisfy diverse expectations of the users by allowing them to steer the interaction in order to elicit the content that is relevant. However, little is known about how the explainee takes part actively in the process of explaining. To tackle the empirical gap, we examined N = 11 naturally occurring explanations in doctor-patient-interactions. We analyzed the verbal behavior of both the explainer and explainee, in the sequential context and could identify phases that are either predominantly monological or dialogical. Further, by applying a structure of conversational jobs, we analyzed whether jobs proposed in the literature can be verified for our setting and what jobs were fulfilled in what kind of phases. The results display that naturally occurring explanations feature monological and dialogical phases. These findings offer important implications regarding the structure of an explanatory dialogue for future designs of artificial explainable systems.

Keywords: explanation, co-construction, conversational jobs, doctor–patient communication

1 Introduction

Algorithms that guide and determine our everyday life take over information processing that is getting more and more complex. They learn and represent data based on hierarchical architectures that are difficult to assess yielding the models to be black boxes. To make these boxes assessable is the goal of recent Explainable Artificial Intelligence (XAI) approaches. One such recent approach proposed by Sokol and Flach (2020) identifies the challenge of satisfying diverse expectations and competing objectives. Derived from the theoretical work by Miller (2019), new interactive approaches to this challenge allow both partners, the explainer (EX) and the explainee (EE), to be involved. However, the empirical basis for how to involve the EE is scarce. Addressing this gap, we investigated when and to what extent both partners are involved in everyday explanations. Following research suggesting that when asked spontaneously, EXs provide long-winded and monological statements without involving the EE visibly (Chi, 1996; Graesser & Person, 1994), we hypothesized that the EE will be involved later in the interaction, as the dialogue unfolds. This hypothesis is motivated by linguistic research suggesting that conversational partners do not only exchange ideas/content in a conversational interaction, but they also gradually co-construct conversational tasks, necessary for the accomplishment of a common goal (Dausendschön-Gay et al., 2015).

For naturally occurring explanations that are usually embedded in a larger conversational context, Quasthoff and colleagues (2017) identified parts referred to as conversational jobs. Because the first of the jobs paves the way for a subsequent explanation, and the last job organizes the transition back to the turn-by-turn talk, we prefer to speak of explanatory sequences rather than explanations. While all jobs are co-constructed by both EX and EE, the EX takes the principal speaker part and thus the main effort in the core job of explicating procedural, conceptual and/or causal relations (Quasthoff et al., 2017). Other jobs require the EE to be more engaged. This is usually the case at the beginning and ending of an explanatory sequence. First, by establishing topical relevance, a transition is jointly organized from turn-by-turn talk, and a principal speaker (Wald, 1978) is selected. Subsequently, both EX and EE must constitute an explanandum. Likewise, the final jobs of closing and transition getting back to turn-by-turn talk, must be jointly accomplished to finish the explanatory sequence (Quasthoff et al., 2017; Sacks et al., 1978).

Research on scientific explanations so far has mainly concentrated on the core job, i.e., *explicating procedural, conceptual and/or causal relations*. However, as naturally occurring explanations are usually embedded in an ongoing interaction with specific common goals, we studied all conversational jobs in our investigation, to find out more about their occurrences in monological and dialogic phases during explanatory sequences.

2 Method

2.1 Participants

For the purpose of this study, 11 consultations have been audio- and videotaped at the Clinic for Pediatric Surgery, Bethel hospital in Bielefeld. The majority (9/11) of the consultations were held in the children's outdoor department (NoKi); the other two in the pediatric surgery department. In total, 8 doctors, 10 patients and 13 legal guardians took part in the study. Because the type of surgery often differed among the consultations, the length of the explanation differed as certain surgeries took more time to explain than others. The duration of the conversations varied between 5 and 26 minutes (mean: 9.5 minutes, SD = 6.5). In order to keep the situation as natural as possible, the conversational partners did not receive any instructions.

2.2 Coding

To explore the engagement of the EE, we coded the content, the different phases of the explanatory sequences, speech of the EE and EX and the jobs performed by the participants. The coding of the jobs and phases was done simultaneously by two researchers. After the completion of the coding process, initial the two coders switched their focus and re-coded two explanations to check the reliability of their coding schemes, by calculating an unweighted Cohens Cappa (Cohen, 1960, 1968). The explanations correspond in their temporal length to 20% of the entire data set. The number of turns of the EX forms the sum of the sample (N). The coding of the two phases, monological and dialogic, results in Cohens Kappa of k = 0.7. The Cohens Kappa for the coding of the jobs was k = 0.86. Both coders thus have moderate to near-perfect coding agreement (Landis & Koch, 1977). Following the reliability test, the deviations between the two coders were smoothed.

Content-related segmentation

Additionally, the explanatory sequences were coded according to the content they provided. The consultations in our dataset included the following three content-related elements: organizational issues (1), health check of the patient (2), and information on the surgery (3). Only the latter element was considered in the analysis because this constituted the core explanandum.

Phase segmentation

The different phases of an explanation were segmented into more monological and more dialogical phases. To back up the analysis, the speech of the participants was considered. The monological phases could include backchannels but no turns of other speakers besides the current speaker. In the dialogical phase, multiple speakers can have turns and backchannels. Pauses between the phases were excluded because it is rather difficult to assign to which phase they belong.

Speech segmentation

The speech of the participants was segmented into turns and backchannels. A turn can have the length of at least one word, or it can take up to several sentences (Sacks et al., 1974). Backchannels are defined by Dideriksen et al. (2019, p. 262) as ”head nods or short utterances consisting of a word (e.g., ’uh-huh, ‘yes’, ’okay’), or short sentences, often repeating the previous turn (e.g., A: ’let’s meet Monday at 10’, B: ‘Monday at 10’)“.

Job segmentation

For coding, the content of the turns of EX or EE was considered and processed according to (Quasthoff et al., 2017). Methodologically, the analysis and coding follow conversational analysis (CA). However, only the verbal level was considered. Other modalities such as gaze, gesture, or positioning of the participants were excluded in the analysis.

Table 1: Coding scheme for the conversational jobs[1]

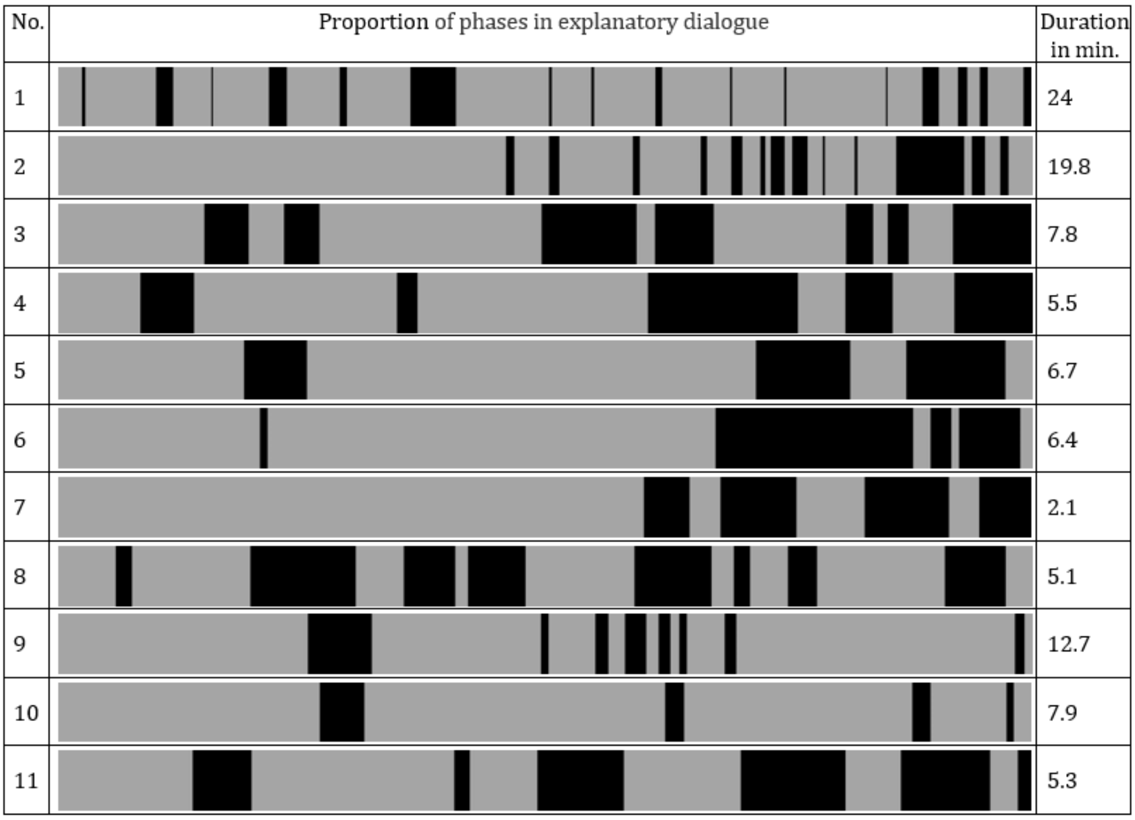
Job	Definition	Examples
Establishing topical relevance	Framing the upcoming sequence as a longer monological phase.	EX: What we also need to talk about… (04) EX: Planned is quasi for the fracture that we already talked about… (03)
Constituting an explanandum	Specifying the explanandum by localization of the crucial body areas or parts.	EX: As my colleague already said there are two types of a circumcision. (11) EX: So, he has up here a groove at his penis. (09)
Explication procedural, conceptual and/or causal relation	Providing information on the processes, concepts and/or reasons underlying the operation as a whole or for individual steps.	EE: I don't know if it is relevant but he is a premature baby, he had, he came five weeks too early and that is the risk he had in the past but he outgrew them. (04) EX: The procedure is nevertheless the same. So, we will make for one, a cut at the groin. (06)
Closing	Framing the upcoming sequence as a completion/ending of the explanation.	EX: But one cannot exclude it for 100 percent, that is basically it. (04) EX: Further questions? EE: No. (04)
Transition	Transfer back to a turn-by-turn structure in a small talk character or by signing the document.	EE: Thank you very much. (04) EX: Then we will do it that way, yes? (04) EX: Good I have shortly written down everything. (05) EX: Then I need your signature. (05)

3 Results

The results confirm our hypothesis: Explanations, even though accomplished co-constructively throughout, appear to have two phases: predominantly monological phases in which the EX acts as the principal speaker, and dialogic phases in which the EE is more involved. In more detail, as can be seen in Figure 1, the average distribution of the phases was 73.4 % for the monological (minimum 57.7 % and maximum 90.8 %, *SD* = 11) and 26.6 % for the dialogic phases (minimum 9.2 % and maximum 42.3 %, *SD* = 11). Regarding the type and length of the first phase, the explanation typically starts with a longer monological phase. The dialogic phases tend to be located at the end of an explanation. Consequently, the monological phases occur rather at the beginning.

[1] The consultations were conducted in German in 10/11 cases (02-11). Sentence structure and semantics were adapted to the English language and thus correspond broadly to the German transcripts.

Figure 1: Phases in the explanatory dialogue (N = 11) are indicated in grey (monological) and black (dialogical). Whereas the phases are displayed in proportions, the duration time of the dialogue is presented in the right column.



Moreover, our job analysis has shown that all jobs occur in both the monological and dialogical phase. In more detail, it shows that all jobs not only occur in the monologic phases, but they are repeated in the dialogical phases. Having the goal not only to explore the extent to which an explainee is involved but also how an EE contributes to the conversational jobs of an explanation, we analyzed which jobs are accomplished in monological vs. dialogical phases. We found that the jobs were present in all dialogues.

When taking a closer look at the first monological phase of each explanation in Figure 1, the tendency that this monological phase is the longest becomes visible. To examine this more thoroughly the proportion of the first monologue in relation to the average proportion of each monological phase was calculated. On average, the first monologue made up 30 % of the monological phases and on average the others made up 22 % each. In more detail, in eight out of eleven consultations (73 %), the first monological phase was the longest. Consequently, in the remaining three (27 %) the first monological phase was shorter. If one considers the first occurrence of the main job, it is crucial to see that it appears in 90.91 % (10/11) of the cases in the first monological phase. This highlights the importance of this job in the unfolding explanation process.

We also found that the jobs of closing and transition were co-constructed in the monological and dialogical phases. In more detail, 49 % of closings and 46.7 % of transitions occurred in the monological phases. Consequently, 51 % of closings and

53.3 % of transitions took place in the dialogical phases. This finding draws a more fine-grained picture on the work of Quasthoff et al. (2017).

The following example from explanatory sequence 04 demonstrates the co-construction of the final two jobs (closing and transition) and is located at the transfer into the second dialogic phase (see Fig. 1). The EEs in this interaction are two family members, the mother (EE 2), who must give her consent to the surgery and the patient (EE 1), who´s bone fracture must be treated surgically. EX addresses the two EEs indeterminately (EX: "Any other questions (?)"). After a pause (1.2 sec.) the son answers (EE 1: "Not from me"). After a short pause (0.6 sec.) the mother replies (EE 2 "no I understood it all too"). Both indicate to EX they have understood the explained procedure of the surgery without further questions. It can be clearly seen that this job is fulfilled dialogically and co-constructively. Question-answer sequences are an examples of pair sequences with strong conditional relevance – meaning that recipients are strongly expected to answer – (Levinson, 1983; Schegloff, 2007; Schegloff & Sacks, 1973). Therefore, the EX depends on the EE to fulfill this job.

The next example occurs subsequently to the transition. In the same vein, the next fulfillment of a job is also performed in a co-constructive manner. EX and EE 2 are talking at the same time (EE 2: "Thank you very much” EX: "fine") and thus fulfill the job closing.

In sum, our results suggest that:

(a) Explanations have two phases, monological and dialogic. The majority of explanatory sequences are monological hence the minority are dialogical.

(b) When averaged across the explanatory sequences, the jobs occurred within the monological as well as dialogic phase thus, suggesting that both phases are omnipresent.

(c) Within the explanatory sequences, particular jobs (the core job of explicating procedural, conceptual and/or causal relations) occurred in longer monological parts. These were repeated in the dialogical part indicating that even though the information was presented, it requires the EEs to link it to their background to make it relevant.

4 Discussion

In this study, we investigated naturally occurring explanations with the aim to provide an empirical basis for the involvement of the explainee (EE). For this purpose, we expanded the model proposed by Quasthoff et al. (2017), and further distinguished between monological and dialogical phases within an explanatory sequence. In line with previous research, we found that all explanatory sequences were initiated and performed predominantly by the explainer (EX). However, we also found the monological phases to be followed by dialogic phases involving the EE to a high degree. This result supports the study by Kobayashi (2021) who investigated similar phases (initial and interactive phases) in an experiment by demonstrating that the two phases occur naturally. In further analyses, the data set revealed a considerable variability concerning the succession and placement of dialogic phases: the EE can be involved at any time during the unfolding interaction. Nevertheless, our data suggests that to a large extent, longer monological phases occurred at the beginning of an explanatory sequence and especially featuring the core conversational job (*explication procedural, conceptual and/or causal relation*). Consequently, the dialogical phases appeared more frequently at the end. These findings expand previous research (Quasthoff et al., 2017) by shedding light onto the co-constructiveness of an explanatory sequence. In more detail, monological phases alternate with dialogic phases systematically in co-occurrence with specific conversational jobs. Furthermore, especially towards the end of an explanatory sequence, the dialogic phases appear to increase, and their relation to the explainees' understanding need to be investigated in future works. We must critically remark, however, that in our setting, explainees are not only informed about the surgery but also need to give their consent at the end of the consultation. The conversation follows thus rules of a medical consultation. As a consequence, our data set was acquired in a rather specific everyday context. To what extent the results can be generalized to other explanatory settings, is a topic for further research.

It is of vital interest to gain a deeper insight into the involvement of the EE in an explanation. Therefore, following Kobayashi (2021) systematically eliciting the two phases, one could run a more detailed analysis focusing on the specific verbal moves that the EE and EX make use of to reveal what kind of moves occur at what phases. The selection of specific speaker moves could serve as conversational techniques to guide an explanation and to adapt to the EE. For Explainable Artificial Systems, adaptation to the EE is of high interest to ensure the relevance of the explanation. The aspect of speaker moves is therefore, included in the A01 project, *Adaptive explanation generation*, of the TRR 318 "Constructing Explainability". In connection to the speaker moves, one could also investigate whether the particular moves are related to specific jobs.

In a further project of the TRR 318, Integrating the technical model into the *partner model in explanations of digital artifacts* (A04) the scope of interest will be extended from verbal to multimodal resources (i.e., gaze, deixis). Thereby, possible (nonverbal) involvements of the EE and also the overall co-construction of understanding will be analyzed in more depth. Additionally, by building on the results presented here, the intertwinement of monological and dialogical phases with conversational jobs will be studied further, with respect to their situative purposes.

5 Conclusion

Whereas first approaches to interactive explainable systems are on their way (Sokol & Flach, 2020), the empirical basis for how to involve the explainee in an explanatory interaction is scarce. Our investigation of naturally occurring explanations suggests that explainees can be involved at any time during the unfolding explanation and XAI systems need to implement both, monological and dialogical phases. Even though an explainer often starts with a monological phase providing crucial information (core conversational job), further information is then negotiated in the upcoming dialogical phases. XAI thus needs to consider that to provide an explanation can only serve as a starting point of a process in which the explainee is to be involved.

6 Declarations

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the Paderborn University on January 11, 2022. Informed consent was obtained from all subjects involved in the study. We specify the following contribution: Conceptualization, all authors; methodology, JF and KR; data acquisition, JF; data analysis, JF, VL, FK; original draft preparation, JF, VL; supervision, FK, KR; funding acquisition, FK, KR. All authors have read and agreed to the published version of the manuscript. WB and SG contributed to the patient and data acquisition and discussion. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR 318/1 2021 – 438445824.

References

Chi, M. T. H. (1996). Constructing Self-Explanations and Scaffolded Explanations in Tutoring. *Applied Cognitive Psychology*, *10*(7), 33–49. https://doi.org/10.1002/(SICI)1099-0720(199611)10:7<33::AID-ACP436>3.0.CO;2-E

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, *25*(4), 471–533. https://doi.org/10.1207/s15516709cog2504_1

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. https://doi.org/10.1037/h0026256

Dausendschön-Gay, U., Gülich, E., & Krafft, U. (Eds.). (2015). *Ko-Konstruktionen in der Interaktion: Die gemeinsame Arbeit an Äußerungen und anderen sozialen Ereignissen.* transcript Verlag. https://doi.org/10.1515/9783839432952

Dideriksen, C., Fusaroli, R., Tylén, K., Dingemanse, M., & Christiansen, M. H. (2019). *Contextualizing Conversational Strategies: Backchannel, Repair and Linguistic Alignment in Spontaneous and Task-Oriented Conversations* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/fd8y9

Graesser, A. C., & Person, N. K. (1994). Question Asking During Tutoring. *American Educational Research Journal*, *31*(1), 104–137. https://doi.org/10.3102/00028312031001104

Kobayashi, K. (2021). Learning by teaching face-to-face: The contributions of preparing-to-teach, initial-explanation, and interaction phases. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-021-00547-z

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Quasthoff, U., Heller, V., & Morek, M. (2017). On the sequential organization and genre-orientation of discourse units in interaction: An analytic framework. *Discourse Studies*, *19*(1), 84–110.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). *A simplest systematics for the organization of turn-taking for conversation. Language (pp. 696–735).*

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A Simplest Systematics for the Organization of Turn Taking for Conversation. In *Studies in the Organization of Conversational Interaction* (pp. 7–55). Elsevier. https://doi.org/10.1016/B978-0-12-623550-0.50008-2

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press.

Schegloff, E. A., & Sacks, H. (1973). Opening up Closings. *Semiotica*, *8*(4). https://doi.org/10.1515/semi.1973.8.4.289

Sokol, K., & Flach, P. (2020). One Explanation Does Not Fit All: The Promise of Interactive Explanations for Machine Learning Transparency. *KI - Künstliche Intelligenz*, *34*(2), 235–250. https://doi.org/10.1007/s13218-020-00637-y

Wald, B. (1978). Zur Einheitlichkeit und Einleitung von Diskurseinheiten. In *Sprachstruktur – Sozialstruktur*.