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 introducing 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.

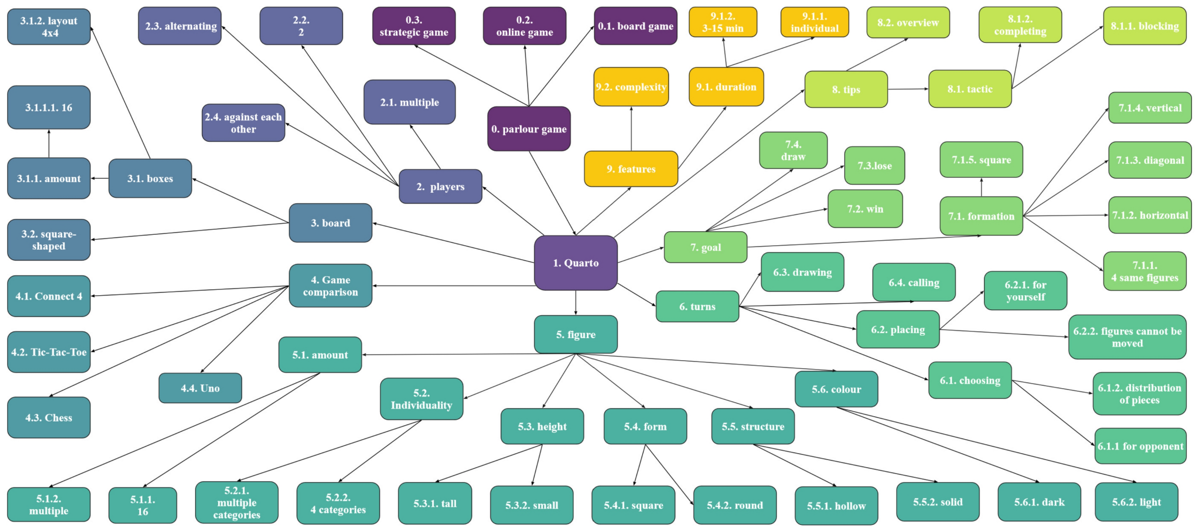

Figure 2: Node scheme: Each block is represented by a specific colour. The colour coding is consistent within all presented figures.

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.

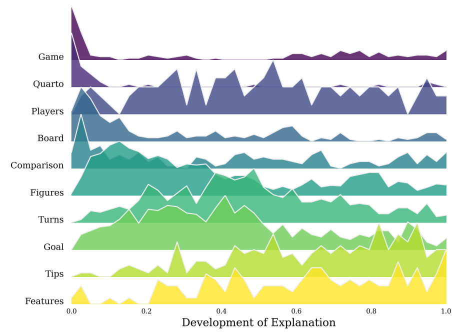

Figure 3: Reference to blocks by EX in relation to the time in all dialogues

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

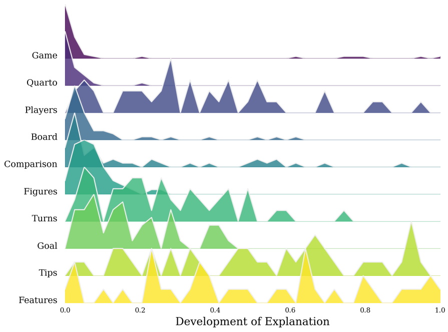

Figure 4: Introduction of the blocks by EX in relation to the time in all dialogues

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.

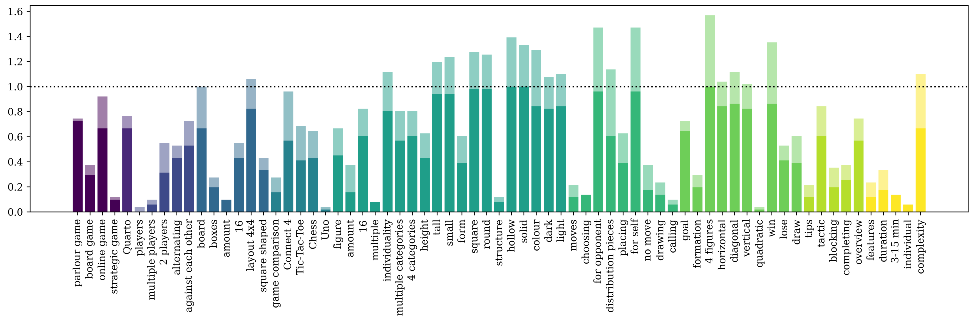

Figure 5: Explanation node coverage by EX (bottom bar) and EE (top bar) - each bar displays in how many of the dialogues the explanation node is mentioned. It shows the proportionate occurrence of the explanation nodes in the entire data set.

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

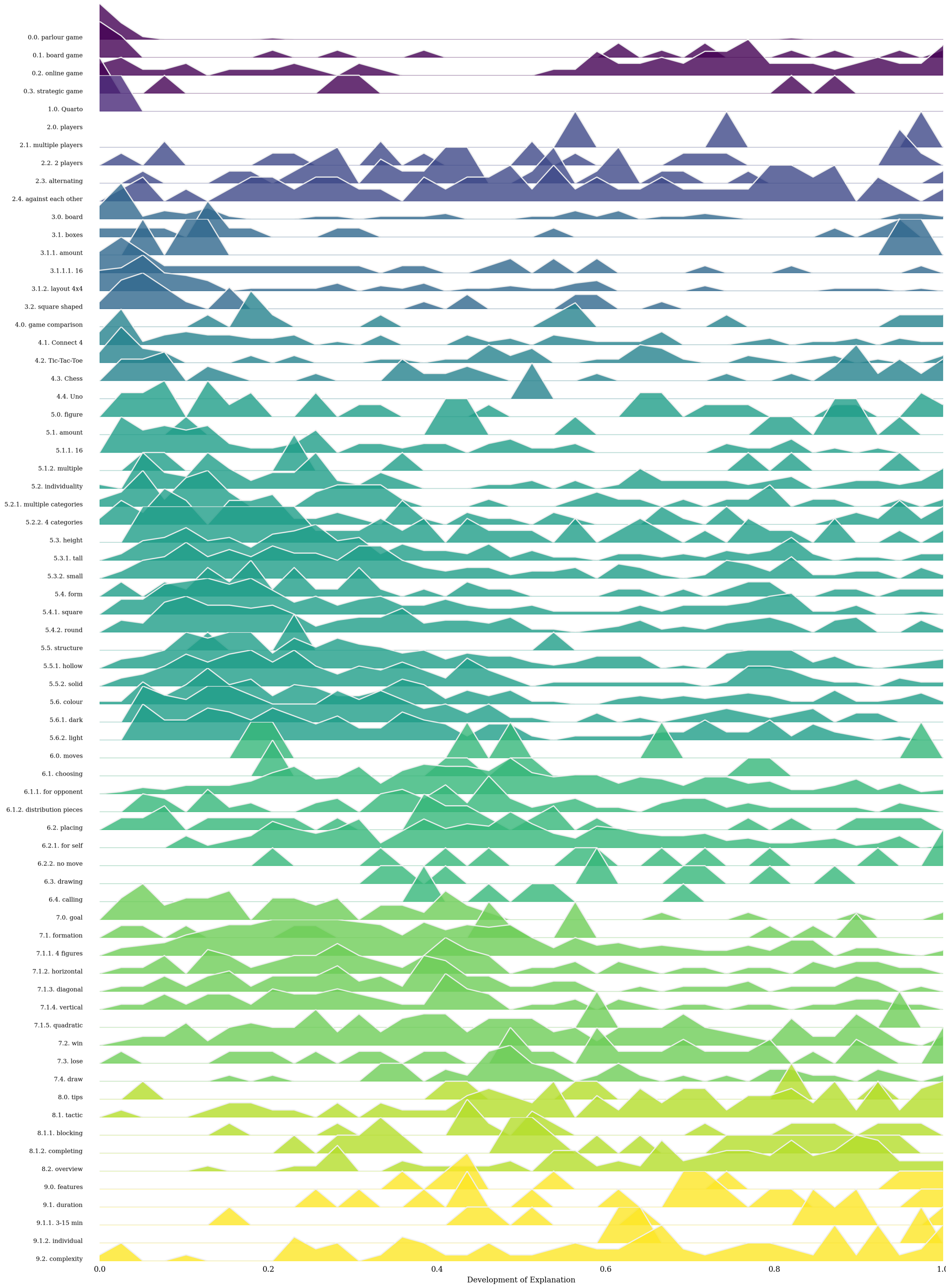

Figure 6: Reference to explanation nodes by EX in relation to the time over all dialogues