Turn-taking dynamics across different phases of explanatory dialogues

Petra Wagner[1,5], Marcin Włodarczak[2], Hendrik Buschmeier[3,5], Olcay Türk[1,5], Emer Gilmartin[4]

[1]Phonetics Workgroup, Faculty of Linguistics and Literary Studies, Bielefeld University
[2]Stockholm University, Stockholm, Sweden
[3]Digital Linguistics Lab, Faculty of Linguistics and Literary Studies, Bielefeld University
[4]INRIA Paris, Paris, France
[5]SFB/Transregio 318 Constructing Explainability, Paderborn and Bielefeld, Germany

Abstract

We examined the turn-taking dynamics across different phases of explanatory dialogues, in which 21 different explainers explained a board game to 2–3 explainees each. Turn-taking dynamics are investigated focusing on >19K floor transitions, i.e., the detailed patterns characterizing turn keeping or turn yielding events (Gilmartin et al., 2020). The explanations were characterized by three different phases (board game absent, board game present, interactive game play), for which we observed differences in turn-taking dynamics: explanations where the board game is absent are characterized by less complex floor transitions, while explanations with a concretely shared reference space are characterized by more complex floor transitions, as well as more floor transitions between interlocutors. Also, the speakers' dialogue role (explainer vs. explainee) appears to have a strong impact on turn-taking dynamics, as floor transitions that do not conform with the dialogue role tend to involve more effort, or floor management work.

1 Introduction

1.1 Floor transitions as indicator of different interactions and interaction styles

Floor management, the organization of the back and forth of the conversational floor between interlocutors, is no regular "ping pong game", during which the contributions of the conversation partners are neatly arranged in consecutive turns clearly delimited by minimally overlapping speech or very short pauses. Rather, periods when floor ownership can be clearly determined, with a single speaker producing solo (non-overlapping) speech, are often separated by a succession of shorter utterances, silences, and overlaps. Despite this, the bulk of the existing turn-taking literature has focused on strictly local descriptions of turn-taking centered around individual instances of silence or overlap, thus losing track of these extended patterns of floor negotiation (Sacks et al., 1974; Heldner and Edlund, 2010; Stivers et al., 2009). By not taking into account the diversity and complexity in how the floor is negotiated, research may easily overlook patterns that characterize more monological ("chunking") or more interactive ("chatting") phases of conversations, but also differences in language-specific interaction patterns such as the typical frequency of vocalized feedback or backchanneling (Dingemanse and Liesenfeld, 2022).

To overcome this apparent limitation, Gilmartin et al. (2020) proposed an alternative description of dialogue state in terms of *floor transitions*. Each floor transition consists of two longer intervals of solo speech exceeding some predefined duration (e.g., 1 second) separated by a series of *intervening intervals*: silences, overlaps, or shorter stretches of solo speech. Depending on whether or not they are associated with a speaker change, floor transitions can be furthered classified as between- or within-speaker (BST and WST, respectively). Previous work has demonstrated that both dyadic and multiparty conversations are greatly varied in terms of the floor transition patterns, with the majority of transitions involving more complex patterns of speech and silence than assumed by simple accounts of turn change and retention (Gilmartin et al., 2020; Włodarczak and Gilmartin, 2021; Gilmartin and Włodarczak, 2023).

One point to note is that across different corpora and interactions the vast bulk of floor state transitions between stretches of single party speech (BSTs and WSTs) have been found to involve odd numbers of intervening intervals. This is due to the very low probability of finding *exact* 'smooth switches' in the data – where one speaker starts speaking at exactly the same moment as another stops or where two or more speakers start and stop speaking at the same time.

Analysis of long multiparty casual conversations

Proceedings of the 28th Workshop on the Semantics and Pragmatics of Dialogue, September, 11–12, 2024, Trento, Italy.

has furthermore identified alternating phases differing in the length, composition in terms of speech, silence and overlap, the relative frequencies, and in the distribution of floor state transitions (Gilmartin et al., 2018). This is broadly in line with the findings of conversation analysis of multiparty casual talk, which has noted that conversations comprise a mixture of two different structural subgenres or phases – stretches of highly interactive chat with participation from several speakers, and longer almost monologic chunks (often narrative or expositionary – anecdotes, recounting of experience, . . .) where one speaker dominates and others mostly provide feedback (Eggins and Slade, 1997).

Between-speaker transitions in chat interaction were spread over more intervening intervals than in chunk, thus increasing the frequency of more complex transitions. This could reflect more turn competition, or indeed more backchannels and acknowledgment tokens being contributed by more participants. One-interval transitions comprised the largest class, with a higher proportion of one-interval transitions in chunk than chat, and higher proportions of within speaker than between-speaker one-interval transitions in both, but particularly in monologic chunk.

A comparison of multi-party conversations and the dyadic phone conversations showed less silence and overlap in dyadic conversations (Gilmartin and Włodarczak, 2023). Also, dyadic interactions showed comparatively fewer occurrences of floor transitions with multiple intervals. However, it is unclear whether these results are mainly influenced by the number of speakers participating in the conversation, or whether the lack of the visual channel may have an independent influence: on the phone, speakers may wait for their interlocutor to finish before commencing to speak, and may not give as much verbal feedback in overlap.

1.2 Explanations as a special case of dialogues

Turn-taking has been investigated for dialogue generally (Sacks et al., 1974), for specific types of dialogue (free: Gilmartin et al. 2018, task-oriented: Gravano and Hirschberg 2011, chaired: Larrue and Trognon 1993), and for different types of interaction partners (e.g., children: Garvey and Berninger 1981, artificial conversational agents: Skantze 2021). In this paper, we examine the floor transitions in explanatory dialogues. These constitute a special case of dialogical interaction as they have interesting properties (they are task-oriented and goal-directed, but not too narrow and involve all participants) and are of practical interest and relevance to various fields such as health communication (Collins, 2005), education (Chi, 1996), explainable AI (Rohlfing et al., 2021), or human-robot interaction (Stange et al., 2022). In particular, we expect that successful explanations are not only shaped by an active explainer directed towards a passive explainee, but involve a high level of interaction, bidirectional monitoring and adaptation, or 'co-construction' of an ongoing explanation, with the collaborative goal of reaching understanding (Rohlfing et al., 2021). Fisher et al. (2022) could show for naturally occurring explanatory dialogues between doctors and patients, that explanations may contain both more monological and more dialogical phases, and such phases can be initiated independently of the conversational role. However, it is yet unclear whether and how these explanatory phases can be straightforwardly related to distinct floor transition patterns.

1.3 Research questions

First, we are aiming to discover whether the floor transition patterns we find for explanatory dialogues differ from those found for less constrained, free conversation such as in Switchboard (Godfrey et al., 1992). Second, we are interested in finding out whether floor transitions in explanations can reflect different phases (e.g., chatting vs. chunking) in an ongoing explanation, and how these interact with the different conversational roles (explainer vs. explainee).

2 Methodology

2.1 Dialogue setup

The analyzed data stems from a large corpus of dyadic interactions in German (Türk et al., 2023). The corpus consists of 87 explanatory dialogues, in which an explainer (ER) had the task to explain the board game 'Deep Sea Adventure' (Sasaki and Sasaki, 2014) to several (2–3) randomly chosen explainees (EE) consecutively. That is, each explainer is involved in 2–3 conversations each, thereby possibly adapting their explanation strategy, but also possibly adapting to different conversational partners. Prior to the study, the explainers had (a minimum of) two days to familiarize themselves with the board game rules. Explainers were entirely free in how they explained the board game. However, each explanation dialogue had to contain three phases: initially, there was a phase in which the physical

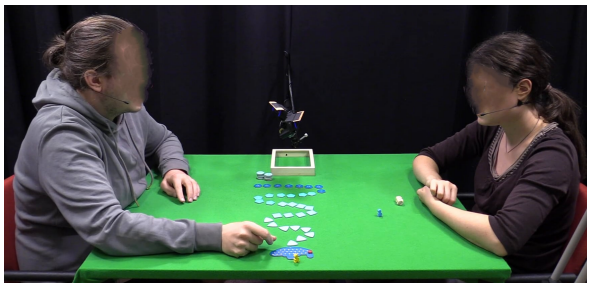

Figure 1: Explanation dialogue setup with the explainer (ER, left) and the explainee (EE, right). The figure shows a dialogue phase with the board game materials present.

board game was not present (gameAbsent). Next, explainers chose a moment at which the actual board game was introduced and the explanation was continued (gamePresent). Last, the explainers were asked to play the board game together with the explainees (gamePlay). This phase may or may not contain aspects of explanation. Explainers were free to choose when to end one explanation phase and begin the next. All interactions were video- and audio-recorded (see Figure 1) using individual head set microphones and multiple camera perspectives.

2.2 Annotations

The explanations were first transcribed with the help of the BAS Web Services (Kisler et al., 2017) or the automatic speech recognition software Whisper (Radford et al., 2022), and then corrected manually using Praat (Boersma and Weenink, 2022). In this annotation step, labels for disfluencies, backchannels, laughter, and audible breathing noises were added. Currently, the corpus is being annotated further for discourse functions, multimodal behaviors, acoustic-phonetic as well as symbolic prosody – but these were not analyzed further in the present study.

2.3 Participants

For the current analysis, we used dialogues from 21 explainers with 2–3 explainees each (75 explanatory dialogues in total). The mean duration of these explanations was M = 26min 47s (SD = 5min 55s). All explainers were German native speaking adults (age: $M = 23.33, SD = 2.58$; 6 male, 14 female, 1 diverse). Not all explainees chose to provide their socio-demographic information. However, they were all recruited based on their report of being a native German speaker. All participants signed a consent form, and the study had been approved by the university Ethics Board.

2.4 Characterizing floor transitions

Using the methodology described in Gilmartin and Włodarczak (2023), the transitions of longer stretches of speech in our data set were characterized as being either examples of *within-speaker transitions* (WST) or *between-speaker transitions* (BST). This yielded a total amount of n = 19 458 floor transitions. For each dialogue, these were further split into the three explanation phases by partitioning the data into three equal parts of overall transitions, the first of which is assumed to roughly correspond to the dialogue phase gameAbsent, the second to the dialogue phase gamePresent, and the third to the dialogue phase gamePlay.

Additionally, each such transition was further characterized with respect to its structural detail: It is determined whether each transition contains stretches of solo speech, silences, or overlaps. Audible breaths, clicks, or laughter occurring on their own were excluded from the speech category and were not taken into account further. Based on the total number of events occurring in between two longer stretches of speech, each floor transition is then given a complexity score. That is, a floor transition that contains a single event in between longer stretches of speech, e.g., a silence, has the transition complexity of 1. With each further event, the complexity score increases.

Transitions types can also be represented with a shorthand notation using uppercase Latin letters to denote individual speakers (*A* and *B* for our dyadic explanations), combinations of letters to denote overlaps, and the letter *X* to denote global silence. Thus, for instance, *A_AB_B* is a between-speaker interval from speaker A to speaker B involving a single overlap, and *A_X_B_AB_A* is a within-speaker transition involving a silent interval, a shorter interval of solo speech by B and an overlap between A and B.

2.5 Analyses

In line with previous research (see Section 1), we expected most floor transitions to show an odd number of intervening intervals, and counted the most frequent patterns for transitions with one and three intervening events. As these counts revealed identical preferred transition patterns across dialogue phases (separately for BSTs and WSTs), we performed χ^2-tests to see whether the patterns distributed differently across the three different dialogue phases.

In order to test whether the transition complexity (measured as the number of individual events occurring between two longer stretches of speech) differed between dialogue phases, transition types, and the dialogue role of the speaker keeping or taking the turn, we calculated non-parametric Kruskal-Wallis tests, followed by post-hoc pairwise comparisons (Dunn tests, Bonferroni corrected). A non-parametric method was chosen, as regression models yielded non-normally distributed residuals.

In order to determine whether the odds for certain transition events (silences, solo speech, overlapping speech) differed between different dialogue phases, we calculated mixed logistic regression models with *silence*, *overlap* and *solo speech* as dependent variables, and dialogue phases (gameAbsent, gamePresent, gamePlay), direction of transition (EE, ER), as well as transition type (BST, WST) as fixed factors, and explainer as random intercepts. We also checked for significant interactions of the fixed factors, and carried out post-hoc pairwise comparisons where these occurred.

All statistical analyses were carried out using R version 4.1.3 (R Core Team, 2022), and the packages tidyverse (Wickham et al., 2019), lme4 (Bates et al., 2015), and rstatix (Kassambara, 2023). Post-hoc comparisons of factors involved in model interactions were performed using the package emmeans (Lenth, 2022).[1]

3 Results

3.1 Floor transitions across different transition types

In line with earlier research, the vast majority of floor transitions show an uneven number of intervening intervals (see Figure 2). Overall, there are fewer BSTs (n = 5284) than WSTs (n = 14 174). Simple transitions are more likely to be WST, while more complex transitions (>3 intervening intervals) are more likely to be BST (see Figure 2, right). That is, interlocutors invest more floor management work to yield or grab turns, and less to keep them. This tendency is statistically significant ($H(1, 19\,458) = 365.78, p < 0.001$), and post-hoc pairwise comparisons showed that this trend is stable for dialogues from 19 out of 21 explainers.

[1]The R-scripts and derived data sets (not the original recordings) can be obtained from the authors upon request.

Table 1: Frequencies of occurrence (raw counts) of floor transitions patterns for 1-interval transitions in BST and WST across the three different explanation phases.

	gameAbs		gamePres		gamePlay	
Pattern	**BST**	**WST**	**BST**	**WST**	**BST**	**WST**
A_X_B	247	2971	575	1899	720	1553
A_AB_B	56	200	171	198	157	118
total	303	3171	746	2097	877	1671

3.2 Floor transition complexities across different explanation phases

The dyadic explanations contain a higher proportion of simple (one-interval) floor transitions than what has been reported for the free dyadic conversations in the Switchboard corpus (Godfrey et al., 1992), especially for the first phase of the game explanations (see Figure 2, left). In later stages, the proportion of simple floor transitions drops strongly, more in line with less constrained conversational data. These differences in complexity across explanation dialogue are statistically significant ($H(2, 19\,458) = 482.53, p < 0.001$), and post-hoc pairwise comparisons showed that this trend is stable for dialogues from 19 out of 21 explainers.

3.3 Floor transitions patterns across different explanation phases

The frequencies of occurrence for different floor transition patterns are presented in Tables 1 and 2, separate for BSTs and WSTs. For transitions with one intervening interval (Table 1), the preferred floor transition patterns remain similar across the different dialogue phases for BSTs ($\chi^2(2, 1876) = 0.184$, n.s.), but change for WSTs, with a slightly higher proportion of overlapping transitions in the later dialogue phases, gamePresent and gamePlay ($\chi^2(2, 6953) = 24.62, p < 0.001$). For transitions with three intervening intervals (Table 2), the relative distribution of preferred floor transition patterns change significantly, both within BSTs ($\chi^2(2, 1024) = 13.83, p < 0.05$) and WSTs ($\chi^2(2, 3325) = 28.36, p < 0.05$), but it is difficult to identify a clear-cut pattern in these changes.

Generally, it can be observed that the occurrences of WSTs decrease in course of the dialogue, while the numbers of BSTs increase, indicating a higher level of floor transition related 'work' in the later stages of the explanation.

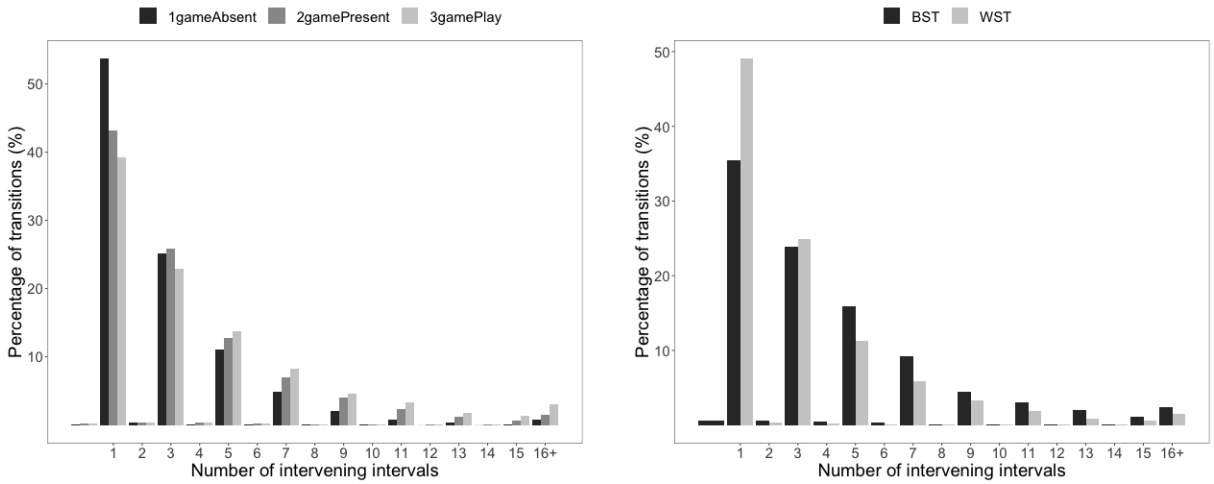

Figure 2: Frequencies of occurrence (%) of transition complexities across the three different dialogue phases (left) and transition types BST and WST (right).

Table 2: Raw counts of the most frequent floor transitions patterns for 3-interval transitions in BST and WST across the three different explanation phases.

	Pattern	gameAbs	gamePres	gamePlay
BST	A_X_B_X_B	43	181	254
	A_X_A_X_B	39	103	126
	A_AB_B_X_B	25	75	68
	A_AB_A_X_B	17	45	48
	total	124	404	496
WST	A_X_A_X_A	733	537	390
	A_X_B_X_A	358	285	234
	A_X_B_AB_A	120	108	101
	A_AB_A_X_A	77	68	54
	A_AB_B_X_A	76	101	83
	total	1367	1099	862

3.4 Floor transitions across different dialogue roles

BST transitions are equally often concerned with transfer of the floor to EEs ($n = 2643$) as to ERs ($n = 2641$), but BSTs to ERs are more complex ($M = 4.62, SD = 4.88$) than those to EEs ($M = 4.1, SD = 4.44$). That is, less complex BSTs tend to correspond to floor transitions to EEs, and more complex BSTs tend to correspond to floor transitions to ERs (see Figure 3, left). These complexity differences are statistically significant ($H(1, 5284) = 17.0, p < 0.001$). In WSTs, this pattern is almost reversed (see Figure 3, right): a lot more WST floor transitions are targeted to ERs ($n = 12\,270$) than to EEs ($n = 1904$), and transitions to ERs have fewer intervening intervals ($M = 3.2, SD = 4.16$) than those to EEs ($M = 4.12, SD = 4.74$). These differences are statistically significant ($H(1, 14\,174) = 128.37, p < 0.001$). Taken together, this indicates that more floor management work is necessary when the floor transitions are not aligned with the assigned dialogue roles, where the explainer's task is to keep the floor (and continue with the explanation), and the explainee's main task is to react and signal understanding, non-understanding, or ask for clarification.

3.4.1 Distributions of overlaps, solo speech, and silences across different game phases

The analysis of preferred floor transition patterns already indicated shifting patterns across different dialogue phases (see Section 2.4). In the following, these tendencies are examined in more detail using mixed logistic regression models.

The regression model for overlaps shows that both game phases and transition types influence the likelihood of overlapping speech (see Figure 4, left). In particular, gamePlay makes overlapping speech less likely ($est = -0.34, se = 0.08, z = -4.1, p < 0.001$) and WST transitions make overlapping speech less likely ($est = -0.14, se = 0.08, z = -17.0, p < 0.001$). Also, there is a significant interaction between dialogue phase and floor transition type, leading to opposite effects for BSTs and WSTs in course of the dialogue: BSTs are losing their stronger likelihood tendency to show overlap in course of the game, showing least overlapping speech during gamePlay, while WSTs are increasing their likelihood to show overlap in course of the game, and are least likely to show overlap during gameAbsent (see Figure 4, left). For BSTs, a pairwise post-hoc comparison showed significant differ-

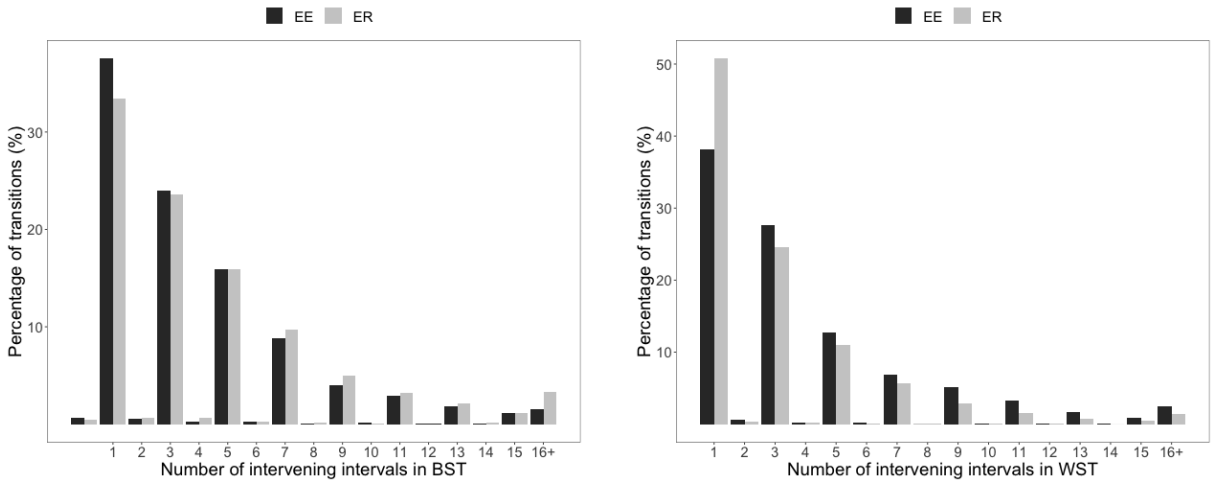

Figure 3: Frequencies of occurrence (%) of transition complexities to ER and EE in BST (left) and WST (right).

ences between gamePlay and the earlier gameAbsent ($est = 0.33, se = 0.08, z = 4.03, p < 0.001$) and gamePresent ($est = 0.21, se = 0.06, z = 3.37, p < 0.01$). For WSTs, a pairwise post-hoc comparison showed significant differences between gameAbsent and the later gamePresent ($est = 0.58, se = 0.05, z = 12.1, p < 0.001$) and gamePlay ($est = 0.73, se = 0.05, z = 2.96, p < 0.001$).

Furthermore (see Figure 4, right), we found that overall, solo speech is less likely to occur in WSTs than in BSTs ($est = -0.6, se = 0.07, z = -7.67, p < 0.001$), and later phases of the dialogue increased the likelihood for solo speech to occur in floor transitions (gamePresent: ($est = 0.19, se = 0.09, z = 2.14, p < 0.05$); gamePlay: ($est = 0.20, se = 0.08, z = 2.38, p < 0.05$). A post-hoc test revealed that the tendency of solo speech to increase in the later stages of the dialogue is largely due to WSTs, which show a significant increase in solo speech between gameAbsent and gamePresent ($est = 0.37, se = 0.04, z = 8.89, p < 0.001$) as well as between gamePresent and gamePlay ($est = 0.19, se = 0.04, z = 4.28, p < 0.001$). For BSTs, this tendency can only be found when contrasting the early gameAbsent and the late gamePlay phases ($est = 2.0, se = 0.08, z = 2.34, p < 0.05$).

As for silences (see Figure 5), the model reveals they have a high likelihood to occur in all floor transitions – in line with the results displayed in Tables 1 and 2. Also, silences are more likely to occur in WST transitions ($est = 2.0, se = 0.08, z = 2.34, p < 0.05$). Due to interactions between the transition types and dialogue phases, we performed pairwise post-hoc comparisons, which revealed that silences are distributed differently for BSTs and WSTs across the dialogue: For WSTs, silences are most likely in the initial gameAbsent and the final gamePlay phase, and differing significantly from gamePresent (gameAbsent-gamePresent: $est = 0.25, se = 0.09, z = 2.59, p < 0.05$; gamePlay-gamePresent: $est = 0.4, se = 0.11, z = 3.73, p < 0.001$). For BSTs, silences are least likely in gameAbsent, and do not differ in their probability to occur in the later phases in the dialogue (gameAbsent–gamePresent: $est = 0.38, se = 0.14, z = 2.69, p < 0.05$; gameAbsent–gamePlay: $est = 0.38, se = 0.14, z = 2.77, p < 0.05$).

4 Discussion

Overall, our results show that explanatory dialogues differ from free dyadic phone conversations in various ways. In particular, they have a higher likelihood to have less complex floor transitions, especially in the first phase of the ongoing explanations, where the physical board game was not yet present and the explanations were made in an 'abstract' fashion. This indicates that floor transition patterns can differentiate between different types of dyadic interactions (phone conversations on a given topic vs. explanations). However, at first glance, this result is not in line with our expectation about explanations being characterized by a high degree of co-construction (Rohlfing et al., 2021). Rather, the floor transitions appear to reveal a strong degree of monologic chunking rather than dialogic chatting. This impression is strengthened by the general prevalence of WSTs (rather than BSTs), and the fact that WSTs rarely coincide with overlaps, but almost always with silences. Also, WSTs to explainers tend

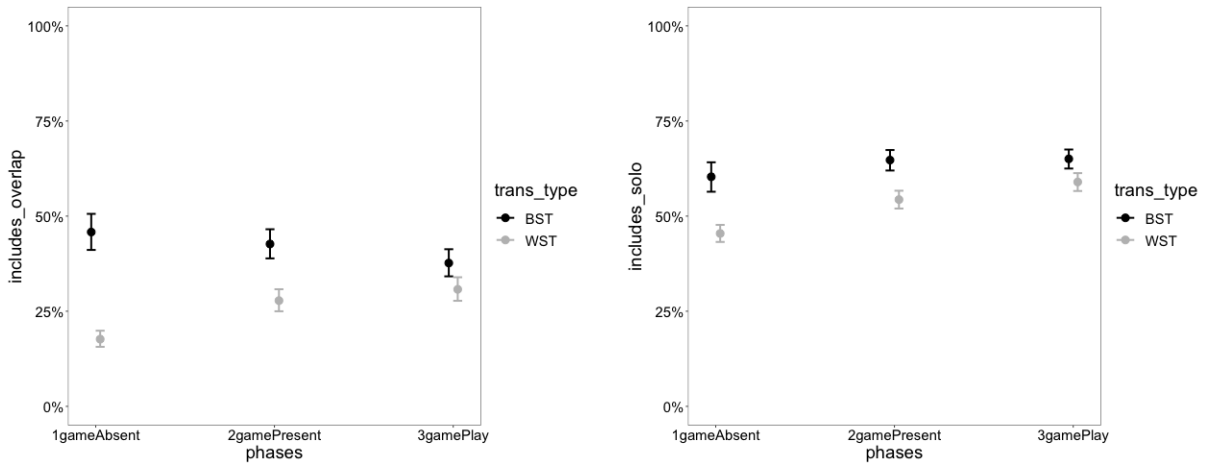

Figure 4: Predicted model probabilities (with 95% CIs) for occurrence of overlapping speech (left) and solo speech (right) across game phases.

to be least complex. Taken together, this gives the impression of an explainer mainly speaking and holding the floor (and not being challenged), and the explainee mostly being in a listening role.

However, we also clearly see that both the floor transition complexities as well as the proportion of BSTs increase in the later phase of the explanation, where the physical board game is introduced as a shared space that interlocutors can refer to, both verbally and multimodally (e.g., by deictic gestures). This is in line with findings by Fisher et al. (2022), who showed that explanations can take more monologic or more dialogic forms, but in our data, this change coincides with a change in situation (visible board game), which probably caused a higher degree of interaction by our interlocutors. A higher degree of co-construction during the gamePresent phase is also indicated by the drop in silences for WSTs, together with a higher proportion of overlaps and complex floor transitions. Currently, we cannot say whether this impact on co-construction can be generalized to other types of explanatory interactions (e.g., doctor-patient, teacher-student), both of which may come with and without a shared physical frame of reference, but our results ask for further analyses across different contextual settings.

It comes as no surprise that the last gamePlay phase in our explanations turned out to be most 'chatty', with a more equal distribution of WSTs and BSTs, and and almost equal proportion of overlap and solo speech in WSTs and BSTs. In earlier, more explanatory phases, BSTs are characterized by considerably more overlap, indicating that more turn 'grabbing' effort is necessary in the explanatory phases than during gamePlay.

In our view, the most interesting finding concerns the interaction between the interlocutors' role (explainee/EE vs. explainer/ER) and floor work necessary in BSTs and WSTs: WSTs to EE were more complex than those to ER, while BSTs to ER were more complex than those to EE. This shows that speakers had to invest more conversational effort whenever they were not conforming to their assigned roles of a predominantly 'speaking explainer' (who tends to have the turn, and may yield it when feedback is needed), or a predominantly 'listening explainee' (who might react with feedback to an explanation, but does not typically keep the turn). We therefore see that dialogue roles influence our floor transition behaviors, and in more equal interactions such as the gamePlay phase, these role-specific behaviors are adapted.

Obviously, our study has several limitations. As conversational data differs across many dimensions, it is difficult to compare results across different settings. Here, we not only compared dyadic free conversations (in American English) to dyadic explanations (in German), but we also compared phone conversations to conversations where interlocutors could see each other, and interact both verbally and non-verbally. We know from prior work that visibility alone has an effect on floor management in instructional dialogues, as visibility decreases overlaps and turn durations, but increases verbal backchanneling (Boyle et al., 1994). It is yet unclear, whether our result for a predominance of monologic interaction in the explanations were not exaggerated, as it currently ignores a large amount

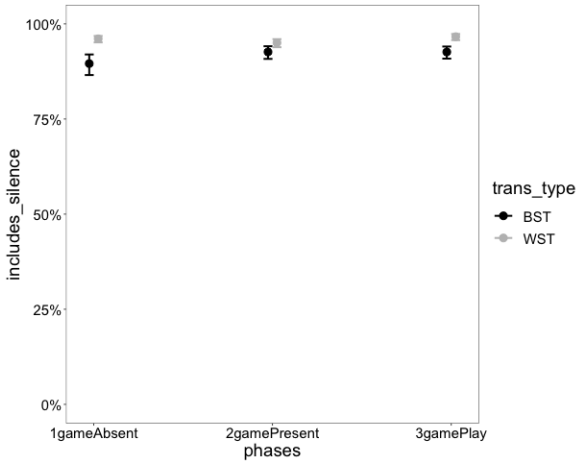

Figure 5: Predicted model probabilities (with 95% CIs) for occurrence of silences across game phases.

of non-verbal feedback behaviors as well as non-verbal cues related to floor management such as gaze, blinking, or head gestures (Malisz et al., 2016; Hömke et al., 2017; Kendrick et al., 2023). It is indeed possible, that interlocutors reduce their usage of gestural floor management cues once the board game is present during the explanation, as they need their hands to carry out the actual movements of the game, need to look at the board game, or use their hands to perform deictic gestures. Because of this, they may switch to a higher proportion of verbalized floor management cues, which we interpreted as more co-constructive interaction. In future work, we will therefore investigate whether non-verbal, gestural floor management follows a similar pattern throughout the various phases of the explanations, or whether verbal floor management compensates if the non-verbal cues cannot be expressed.

Another possibly confounding factor in our data relates to the way that the explanatory dialogue would have evolved without asking our participants to go through various explanatory phases. It is possible, that some of the findings presented here are the result of interlocutors 'warming up' to one another, and becoming more chatty in course of an interaction after a somewhat awkward initial phase. While this cannot be ruled out, we are still confident that this does not explain all our findings, as we see very stable tendencies across many speakers, who also displayed a wide variation in their individual interactive behaviors, or readiness to chat. Also, for silences, we found similarities for the initial and late stages of the conversation, which are difficult to explain if the explanatory phases, as implemented by our design, did not play a role at all.

Lastly, it has to be critically mentioned that our coarse split into three different phases does not properly reflect the three explanation phases. However, as our analysis yielded interesting differences between those three phases, we believe that this approach was successful as a first approximation.

5 Conclusions

Overall, our findings show that explanatory interactions follow turn-taking dynamics that differ from other types of conversational interactions, and shed light on the special turn-taking dynamics in different phases of explanatory interactions. Also, our analysis corroborates the usefulness of floor transitions as a measure for characterizing conversational dynamics and involvement of conversational partners.

Acknowledgments

This research was funded by the German Research Foundation (DFG): TRR 318/1 2021–438445824. We are grateful to two anonymous reviewers who made very helpful comments, and pointed out highly relevant literature we would have missed.

References

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67:1–48.

Paul Boersma and David Weenink. 2022. Praat: Doing phonetics by computer.

Elizabeth A. Boyle, Anne H. Anderson, and Alison Newlands. 1994. The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech*, 37(1):1–20.

Michelene T. H. Chi. 1996. Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology*, 10:33–49.

Sarah Collins. 2005. Explanations in consultations: the combined effectiveness of doctors' and nurses' communication with patients. *Medical Education*, 39:785–796.

Mark Dingemanse and Andreas Liesenfeld. 2022. From text to talk: Harnessing conversational corpora for humane and diversity-aware language technology. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pages 5614–5633, Dublin, Ireland.

Suzanne Eggins and Diana Slade. 1997. *Analysing Casual Conversation*. Cassell, London, UK.

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*, 26:317–326.

Catherine Garvey and Ginger Berninger. 1981. Timing and turn taking in children's conversations 1. *Discourse Processes*, 4:27–57.

Emer Gilmartin, Kätlin Aare, Maria O'Reilly, and Marcin Wlodarczak. 2020. Between and within speaker transitions in multiparty conversation. In *Proceedings of Speech Prosody 2020*, pages 799–803, Tokyo, Japan.

Emer Gilmartin, Carl Vogel, and Nick Campbell. 2018. Chats and chunks: Annotation and analysis of multiparty long casual conversations. In *Proceedings of the 11th International Conference on Language Resources and Evaluation*, Miyazaki, Japan.

Emer Gilmartin and Marcin Włodarczak. 2023. Getting from A to B: Complexities of turn change and retention in conversation. In *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, pages 3457–3461, Prague, Czech Republic.

John J. Godfrey, Edward C. Holliman, and Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 517–520, San Francisco, CA, USA.

Augustín Gravano and Julia Hirschberg. 2011. Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, 25:601–634.

Mattias Heldner and Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38:555–568.

Paul Hömke, Judith Holler, and Stephen C. Levinson. 2017. Eye blinking as addressee feedback in face-to-face conversation. *Research on Language and Social Interaction*, 50:54–70.

Alboukadel Kassambara. 2023. *rstatix: Pipe-friendly framework for basic statistical tests*. R package version 0.7.2.

Kobin H. Kendrick, Judith Holler, and Stephen C. Levinson. 2023. Turn-taking in human face-to-face interaction is multimodal: Gaze direction and manual gestures aid the coordination of turn transitions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378:20210473.

Thomas Kisler, Uwe Reichel, and Florian Schiel. 2017. Multilingual processing of speech via web services. *Computer Speech & Language*, 45:326–347.

Janine Larrue and Alain Trognon. 1993. Organization of turn-taking and mechanisms for turn-taking repairs in a chaired meeting. *Journal of Pragmatics*, 19:177–196.

Russell V. Lenth. 2022. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.7.3.

Zofia Malisz, Marcin Włodarczak, Hendrik Buschmeier, Joanna Skubisz, Stefan Kopp, and Petra Wagner. 2016. The ALICO corpus: Analysing the active listener. *Language Resources and Evaluation*, 50:411–442.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust speech recognition via large-scale weak supervision. *Preprint*, arxiv:2212.04356.

Katharina Rohlfing, Philipp Cimiano, Ingrid Scharlau, Tobias Matzner, Heike Buhl, Hendrik Buschmeier, Angela Grimminger, Barbara Hammer, Reinhold Häb-Umbach, Ilona Horwath, Eyke Hüllermeier, 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*, 13:717–728.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50:696–735.

Jun Sasaki and Goro Sasaki. 2014. *Deep Sea Adventure (Tabletop Game)*. Oink Games, Tokyo, Japan.

Gabriel Skantze. 2021. Turn-taking in conversational systems and human-robot interaction: A review. *Computer Speech & Language*, 67:101178.

Sonja Stange, Teena Hassan, Florian Schröder, Jacqueline Konkol, and Stefan Kopp. 2022. Self-explaining social robots: An explainable behavior generation architecture for human-robot interaction. *Frontiers in Artificial Intelligence*, 5:866920.

Tanya Stivers, Nick J. Enfield, Penelope Brown, et al. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, 106:10587–10592.

Olcay Türk, Petra Wagner, Hendrik Buschmeier, Angela Grimminger, Yu Wang, and Stefan Lazarov. 2023. MUNDEX: A multimodal corpus for the study of the understanding of explanations. In *Proceedings of the 1st International Multimodal Communication Symposium*, pages 63–64, Barcelona, Spain.

Hadley Wickham, Mara Averick, Jennifer Bryan, et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

Marcin Włodarczak and Emer Gilmartin. 2021. Speaker transition patterns in three-party conversation: Evidence from English, Estonian and Swedish. In *Proceedings of Interspeech 2021*, pages 801–805.