Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 4051–4059
Marseille, 20-25 June 2022

Analysis of Dialogue in Human-Human Collaboration in Minecraft

Takuma Ichikawa, Ryuichiro Higashinaka
Graduate School of Informatics, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan
{ichikawa.takuma.w0@s.mail,higashinaka@i}.nagoya-u.ac.jp

Abstract

Recently, many studies have focused on developing dialogue systems that enable collaborative work; however, they rarely focus on creative tasks. Collaboration for creative work, in which humans and systems collaborate to create new value, will be essential for future dialogue systems. In this study, we collected 500 dialogues of human-human collaboration in Minecraft as a basis for developing a dialogue system that enables creative collaborative work. We conceived the Collaborative Garden Task, where two workers interact and collaborate in Minecraft to create a garden, and we collected dialogue, action logs, and subjective evaluations. We also collected third-person evaluations of the gardens and analyzed the relationship between dialogue and collaborative work that received high scores on the subjective and third-person evaluations in order to identify dialogic factors for high-quality collaborative work. We found that two essential aspects in creative collaborative work are performing more processes to ask for and agree on suggestions between workers and agreeing on a particular image of the final product in the early phase of work and then discussing changes and details.

Keywords: collaboration, creative task, dialogue system, minecraft, data collection

1. Introduction

Recently, as dialogue systems become part of our daily lives, towards more advanced dialogue systems, researchers have been working on dialogue systems that can collaborate with humans to complete tasks (Meena et al., 2014; He et al., 2017; Kim et al., 2019; Zhang et al., 2021).

Many studies related to collaborative work with dialogue systems use Minecraft[1] (Szlam et al., 2019). Minecraft is a sandbox-type 3D game in which the player controls an avatar and can move freely in the world. It also supports multiplayer, making it easy for multiple players to collaborate in the same environment.

Most studies using Minecraft have focused on problem-solving collaborative work (Narayan-Chen et al., 2019; Ogawa et al., 2020; Bara et al., 2021), and few studies have focused on creative collaborative work. To realize a society in which humans and systems can create new value, it is essential that systems and users can work together on creative collaborative work.

This study collected data on creative collaborative work between humans in Minecraft as a basis for developing a dialogue system that enables creative collaborative work. We devised the Collaborative Garden Task and collected text chats, action logs, and questionnaire results. We also collected third-person evaluations of the created gardens scored in terms of originality and beauty in order to objectively assess the quality of the collaboration. Furthermore, we analyzed the relationship between dialogue and collaborative work that received high scores on the subjective and third-person evaluations in order to identify dialogic factors for high-quality collaborative work. The contributions of the paper are as follows.

- We collected 500 human-human dialogues on creative collaborative work by using Minecraft; the data contain text chat, action logs, and subjective/objective questionnaire results.
- We found that, for collaborative work, performing more processes to ask for and agree on suggestions between workers is essential.
- We discovered that agreeing on a particular image of the final product in the early phase of work and then discussing changes and details is also essential.

2. Related Work

Regarding collaboration with dialogue systems, a system that manages network traffic flows through interaction with a user (Lochbaum et al., 1990), a system that collaboratively solves problems to plan evacuations (Ferguson and Allen, 1998), and a system that negotiates procedures for abstracted action through interaction to achieve a common goal (Sidner, 1994) have been proposed. In addition, COLLAGEN (Rich and Sidner, 1999) can collaborate with a user in manipulating software such as a video cassette recorder or a gas turbine simulator. PLOW (Allen et al., 2007) can collaborate with the user to search for information, book flights, and purchase books on a web browser by following the user's instructions on operating the web browser. Meena et al. (2014) developed a system that can perform the Map Task (Anderson et al., 1991) with users. In the Map Task, two workers play the roles of an information provider and an information follower. The information provider gives an explanation on the basis

[1] https://www.minecraft.net/

of a complete map, and the information follower follows the instructions and draws the correct route on the map. The system that these researchers developed acts as the information follower, and it acquires the necessary information by interacting with the user (i.e., the information provider) in order to complete the map.

Many of the recent studies on collaboration with dialogue systems use Minecraft. Inspired by the Map Task, Ogawa et al. (2020) devised the Mansion Task, in which two workers receive different maps and try to reach the destination by exchanging the information on the maps. There are various obstacles, such as needing to press buttons simultaneously, along the way, and the workers need to work together to pass these obstacles. Bara et al. (2021) devised a task called MindCraft, in which two workers receive different craft recipes and create target items by exchanging information on item creation. They collected dialogues and workers' belief states by asking them explicitly about their belief states at periodic intervals in the dialogues, and they developed a model for predicting these states. Narayan-Chen et al. (2019) devised the Collaborative Building Task, in which two workers play the roles of an Architect and a Builder and create a target structure in accordance with their roles. The Architect has a blueprint of the target structure and instructs the Builder on how to manipulate the blocks on the basis of the blueprint. The Builder manipulates the blocks following the instructions given by the Architect and creates the structure. Köhn et al. (2020) and Jayannavar et al. (2020) collected such dialogue data and developed a model that generates block operations on the basis of instructions in the Collaborative Building Task. These studies focus on problem-solving, not creative collaborative work.

Few studies focus on creative collaborative work in environments that represent the real world, such as Minecraft. In this study, we focus on the Collaborative Garden Task and collect data such as text chats and action logs with the aim of developing a dialogue system that enables creative collaborative work.

A study that may resemble ours may be that by Mitsuda et al. (2022), who used a design task to lay objects out collaboratively; however, they used a simplistic interface with icons as objects, and the workers had individual layouts as their views in order to examine their mutual understanding through text chat, which is different from our setting in which people collaborate to create the same object in the same space.

3. Collaborative Garden Task Corpus

We collected data on human interactions during collaborative work in Minecraft. This section describes the Collaborative Garden Task, the data collection experiment we conducted, the dialogue data we collected, and the third-person evaluations. Note that we collected the data in Japanese. We obtained the approval of the ethics committee of our university to conduct the data collection experiment with regard to the usage of the data as well as the treatment of the personal information.

3.1. Collaborative Garden Task

We conceived the Collaborative Garden Task, in which two workers work together to create a unique and beautiful garden. By using a garden as a target, we considered it possible to balance the difficulty between creation and variation. As criteria, we used originality and beauty, which depend on individual values, to allow the workers to work creatively through discussions.

Workers interacted via text chat and manipulated blocks. We set the building area to ten squares in length and width and four squares in height. As shown in Figure 1, we placed "grass blocks," which correspond to lawn, in the bottom 10×10 squares in advance. The workers could use only 17 types of blocks such as "grass blocks," "oak leaves," and "bricks." There was no limit on the number of blocks. The time limit for the work was 20 minutes, and workers worked up to the time limit. From our observation, few pairs were unable to complete the garden within the time limit.

3.2. Data Collection Procedure

We used Minecraft Java Edition version 1.16.5 for data collection and SpigotMC[2] for the data collection platform. SpigotMC is an extension server for Minecraft, allowing us to customize the Minecraft environment with plug-ins. Some studies have used Malmo (Johnson et al., 2016) as a data collection platform (Narayan-Chen et al., 2019; Ogawa et al., 2020). However, it requires an additional step to introduce extensions to the client-side; therefore, we chose SpigotMC in this study.

The workers first joined the server and entered the Lobby World, which served as a waiting room where they waited until their partners arrived. Once the partner arrived, the workers created a new experiment room with a type-in command, and each pair moved to the experiment room. There was a separate text chat for each experiment room so that multiple pairs could work simultaneously. After moving to the experiment room, workers could start the Collaborative Garden Task. The timer displayed on the screen started running when the task started. Five minutes before the end of the timer, workers received an announcement of the remaining time. In addition, when they ran out of assigned time, they received a notification that the time was over. This was the process for one collaborative work session, which we refer to as one dialogue.

3.3. Data Specifications

We collected data on the text chats, block placement and removal, and player movement. All collected actions consisted of the following:

- timestamp when the action was performed
- player's position such as their coordinates (x, y, z) and viewpoint (yaw, pitch)

[2] https://www.spigotmc.org/

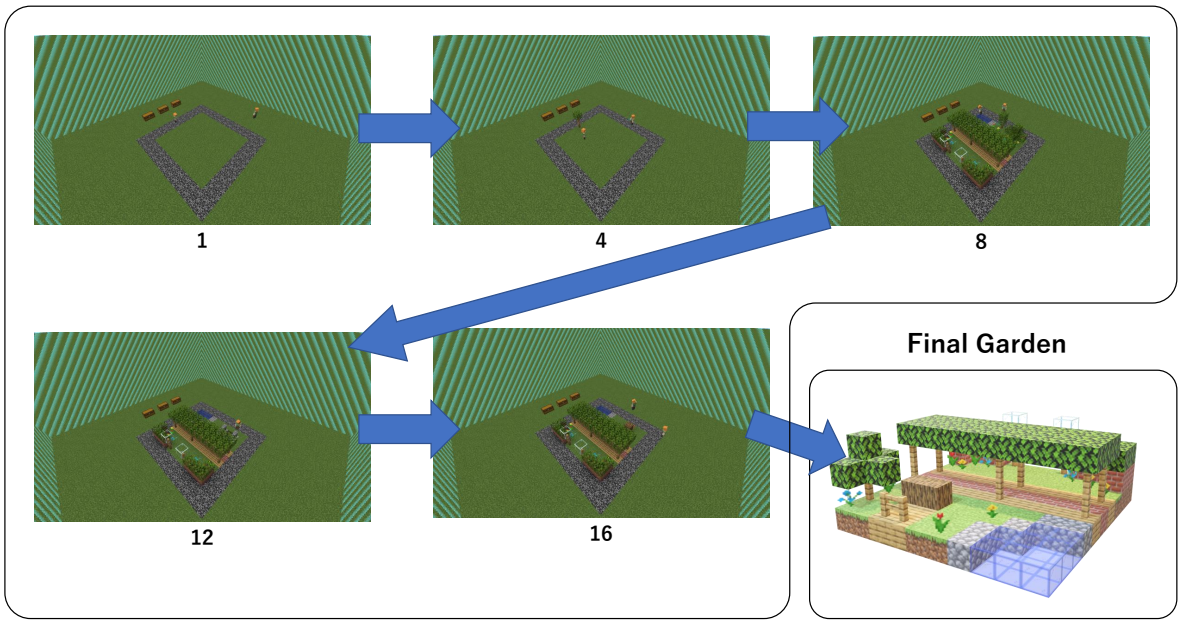


Figure 1: Process of Collaborative Garden Task and picture of final garden (excerpted from collected data). Two workers control their avatars in Minecraft and create garden by placing and removing blocks. Number below pictures is utterance ID in Table 2, and each picture corresponds to time of that utterance.

Number of workers	79
Number of pairs	500
Number of dialogues	500
Number of utterances	16,221
Mean number of utterances	32.4
Number of placements	182,355
Number of removals	131,552
Mean number of placements	182.4
Mean number of removals	131.6
Mean of moving distance	1,899.5
Mean number of blocks	201.8
Mean number of block types	11.6

Table 1: Statistics on collected data

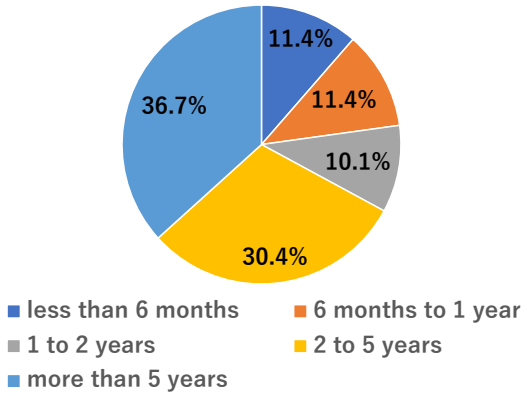


Figure 2: Workers' experience with Minecraft

- action-specific information such as chat messages, block types, etc.

- block information in the building area ($10 \times 10 \times 4$)

We collected these data for every time step when an action was performed by either of the workers, making it possible to reproduce the situation in which the actions were made.

Before participating in the experiment, workers answered a questionnaire about their age, gender, and experience in Minecraft. In addition, they answered a questionnaire about their satisfaction with the dialogue and the created garden after each work session. Here, we used a five-point Likert scale.

3.4. Collected Data

Table 1 shows the statistics of the 500 dialogues collected in this study. Seventy-nine people recruited by crowd-sourcing[3] participated in the experiment, and each worker worked an average of 12.7 times and a maximum of 24 times. Different pairs worked together in each work session. The working time was about 25 minutes, and the workers were paid 585 yen (approx. 5 dollars) per session.

Figure 2 shows the workers' experience with Minecraft. About three-fourths of the workers had been playing for more than one year. As for the experience of working together to create buildings in

[3] https://www.lancers.jp/

ID	S	Utterance	
1	A	なにか作りたいものありますか？	*(Do you want to make something?)*
2	B	藤？みたいな屋根みたいなのつくってみたいです	*(I want to make a roof like a wisteria trellis.)*
3	A	いいですね！	*(Sounds good!)*
4	A	真ん中にどーんと作ってみてください！	*(Try making it in the middle!)*
5	B	道を真ん中に作ってみます	*(I will make a path down the middle.)*
6	B	両脇にもなにかつくりたいです	*(I also want to make something on both sides.)*
7	A	横、こんな感じで大丈夫ですか？	*(This side, is this okay?)*
8	B	いいとおもいます！	*(It's good!)*
9	B	2本の木のところに、試しに椅子置いてみてもいいですか？	*(Can I try to put a chair at the two trees?)*
10	A	どうぞ！邪魔だったら壊してください！	*(Go ahead! If it's in the way, please remove it!)*
11	B	ありがとうございます	*(Thank you.)*
12	B	いいかんじの椅子と机の作り方わかりますか？	*(Do you know how to make a nice chair and desk?)*
13	A	思いつかないですね・・・ハーフブロックとかあればいいのに・・・	*(I cannot think of any... I wish I had a half block or something...)*
14	B	はい．．．	*(Yeah...)*
15	A	でっかいですね！	*(So big!)*
16	B	これ一個でテーブルは無理がありますかね．．．	*(I wonder if it is too much to ask for a table with just one of these...)*
17	A	いや・・・いけます！	*(No... that's okay!)*
18	B	いいです！	*(Great!)*

Table 2: Dialogue (collected in Japanese and translated to English by authors) during collaborative work shown in Figure 1

Questionnaire item	Mean	SD
Q1. Did you communicate your opinions and ideas?	4.61	0.76
Q2. Did your partner share his/her opinions and ideas with you?	4.52	0.87
Q3. When you disagreed, did you reach an agreement through discussion?	4.40	0.86
Q4. Did you successfully do building work such as placing and removing blocks?	4.65	0.67
Q5. Did you come up with any ideas that you could not imagine on your own?	4.40	1.02
Q6. Were you satisfied with the interaction with the partner?	4.34	0.99
Q7. Were you satisfied with what you created?	4.60	0.74

Table 3: Mean scores with standard deviations for questionnaire items averaged over all participants (N=1,000)

Minecraft multiplayer, 72.2% of the workers answered yes, and 27.8% answered no. Many of the workers had experience in multiplayer, experience in text chatting, and experience building things in Minecraft.

The mean number of utterances in the collected dialogues was 32.4, about 1.6 utterances per minute. The mean number of placements in the collected collaborative work per worker was 182.4, and that of removals was 131.6. The mean number of blocks in the gardens was 201.8, about half of the block placements for two workers. This shows that the workers repeatedly tried to place and remove the blocks.

Figure 1 shows the process of the collected collaborative work, and Table 2 shows an excerpt of a dialogue.

Table 3 shows the questionnaire results. The post-work questionnaire results show that all items had high scores, indicating that the quality of the collected collaborative work was high.

Evaluation item	Mean	SD
Originality	4.84	0.73
Beauty	4.32	0.75

Table 4: Mean scores with standard deviations for evaluation items averaged over all gardens (N=500)

3.5. Third-Person Evaluations

Even if the workers subjectively evaluated the gardens they created as good, they may not be objectively so. We thus conducted third-person evaluations of the 500 gardens using crowd-sourcing[4]. A total of 254 evaluators participated. Ten evaluators rated each garden. The average working time was about 42 minutes, and the evaluators were paid 700 yen (approx. 6 dollars). The evaluators first looked at the 30 gardens collected in the preliminary experiment conducted separately and

[4] https://crowdworks.jp/

Classification criteria	Expression (4-gram)	
Satisfaction	ありがとうございます！*,	*thank you!**,
	がいいですか**,	*would you like?***,
	いいですかね*, にしてみ*,	*let's try***,
	になりました*, みましょうか**,	*it looks**,
	なんですよ**, でどうでしょう**,	*that's right***, *how about?***,
	そうですよね*, いいですよ！*,	*great!**,
	感じでどうでしょ*, してみて*,	*I think we could try**,
	もよさそう*, よかったです！*,	*good!**,
	あったりします**, 感ありますね*,	*do you have***,
	こんな感じでどう**,	*it has a feel**,
	ありがとうございました**, てもよさ*	*thank you***
Originality	のはどうです*, っぽくなりました*	*how about?**, *it looks like**
Beauty	しますか？*, てみても*,	*would you like?**, *let's try**,
	みてもいい*, してみたいです*,	*I would like to**,
	いいかんじですね*, もいいと思い**	*looks nice**

Table 5: Frequent expressions (4-gram of words in Japanese and English translation by authors) that appear in dialogues with high ratings in satisfaction, originality, and beauty (p-values are based on Fisher's exact test; *: $p < 0.05$, **: $p < 0.01$).

established their own evaluation criteria. Then, they evaluated 20 gardens in terms of originality and beauty. Originality and beauty were evaluated on a seven-point Likert scale. In addition, they were requested to write a comment (over ten characters in Japanese) stating the basis for their evaluations. The purpose of collecting the comments was to make the evaluator look at the gardens carefully before providing ratings.

We calculated the average of the ten evaluators' evaluation scores. Table 4 shows the mean and standard deviations of the obtained scores. The mean score for originality was 4.84 with a standard deviation of 0.73, and that for beauty was 4.32 with a standard deviation of 0.75. In the third-person evaluations, a rating of four means a middle-rated garden, and five means a relatively good one. Although the scores were not as high as those of the subjective evaluations, the third-person evaluations also confirmed the good quality of the collaborative work.

4. Analysis of Dialogue in Collaboration

As an analysis to develop a dialogue system that enables the Collaborative Garden Task to be done, we analyzed the dialogic factors that enable high-quality collaboration.

4.1. Expressions Useful for Collaboration

We first performed an analysis on the basis of word frequency. We classified all 500 dialogues into the top 20% and the bottom 80% on the basis of three classification criteria: one was the sum of the scores of the seven subjective evaluations (satisfaction), and two were the third-person evaluations (originality and beauty) of the gardens.

Focusing on word 4-grams that contain at least one content word, we first listed the top 300 frequent such 4-grams in the 16,221 collected utterances from our data. Then, we mined frequent 4-grams that appeared in the high-quality dialogues. More specifically, we used Fisher's exact test to verify whether the expressions frequently appeared in the top 20% of dialogues. We used Janome[5] as a Japanese morphological analyzer to extract words from utterances. We created a 2×2 cross table of the 500 dialogues with one axis indicating whether a dialogue included a specific word 4-gram and the other indicating whether a dialogue was among the high-quality dialogues, and we conducted a Fisher's exact test on them. Table 5 shows the expressions identified by test.

Two main types of expressions frequently appeared in dialogues for which satisfaction was high. The first was expressions that ask for a partner's agreement, such as "would you like?", "let's try," and "how about?" These expressions appear when suggesting an idea to a partner or deciding on a work plan. The dialogue shown in Table 2 was in the top 20% of satisfaction, and we can see expressions such as "This side, is this okay?" and "Can I try to put a chair at the two trees?"

The second was expressions showing positive appreciation toward the partner, such as "that's right" and "great!" These expressions show positive appreciation and agreement with a partner's opinion or what the partner created. In Table 2, we can see expressions such as "Sounds good!" "It's good!", and "No... that's okay!"

As for the dialogues with high third-person evaluations, we observed similar kinds of expressions. Regarding originality, we obtained two expressions, "how about?" and "it looks like." "How about?" is an expression that asks for a partner's agreement on one's own suggestion. We obtained six expressions regarding beauty. These expressions, such as "would you like?" and "let's try,"

[5] https://github.com/mocobeta/janome

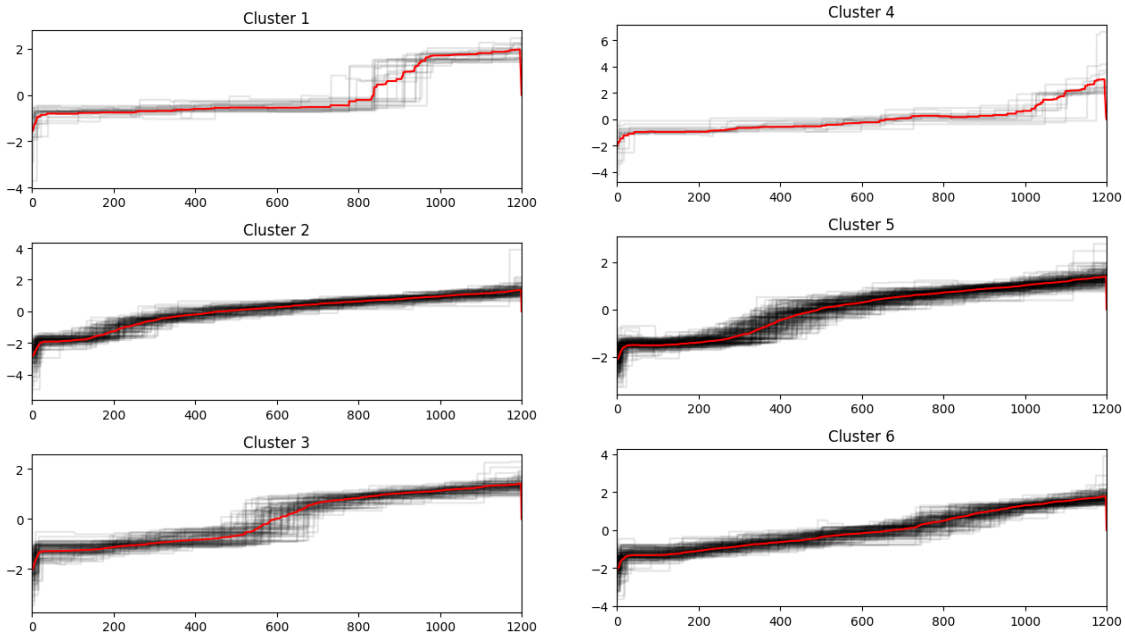


Figure 3: Results of time-series clustering of completion rate

are those that ask for a partner's agreement. In addition, "looks nice" is an expression of positive appreciation toward the partner.

In summary, we can see quite a few expressions of agreement, indicating the importance of performing more processes to ask for and agree on suggestions between workers in collaborative work.

4.2. Analysis of Collaboration Process

We investigated which types of dialogue lead to high-quality collaboration. To analyze the progress being made in the collaboration, we plotted the gardens' progress on the basis of the completion rate (see Eq. 1) and conducted a cluster analysis to find patterns in the collaboration process.

Considering that the final layout of the blocks was the result of the agreement between the workers, we could use the existence of these blocks at time t as a measure to calculate the completion rate.

$$\text{Completion Rate} = \frac{|Blocks_t \cap Blocks_{last}|}{|Blocks_{last}|} \quad (1)$$

$Blocks_t$ represents the set of blocks installed at time t, and $Blocks_{last}$ represents the set of blocks used in the final garden. We used the time-series data of the completion rate at the time of an utterance and conducted a cluster analysis on it. We used a k-means based method, k-Shape (Paparrizos and Gravano, 2015), for clustering, and following Mitsuda et al. (2022), we normalized the completion rate to follow a normal distribution with mean 0 and variance 1 in advance. We used the Elbow method to determine the number of clusters. Since there were several clusters in which the sum of squares of intra-cluster errors stopped decreasing, one of the authors manually checked the number of clusters and determined it to be six.

Figure 3 shows the clusters we obtained. Cluster 1 accounted for 4.4%, Cluster 2 for 19.6%, Cluster 3 for 16.4%, Cluster 4 for 3.0%, Cluster 5 for 32.2%, and Cluster 6 for 24.4% of the total.

In Clusters 1 and 4, the completion rate increased rapidly in the last phase. This happened because there were few utterances in the middle phase. In dialogues in these clusters, the workers seemed to have continued to work silently after deciding what to make through discussion in the early phase.

In Clusters 3 and 5, the completion rate increased significantly in the middle phase. In particular, in Cluster 3, we can see a sharp increase in the completion rate in the middle phase. In contrast, in Cluster 5, the completion rate increased only slightly before the middle phase. We assume that in these clusters, the workers proceeded with the outline discussed in the early phase and discussed the details in the middle and later phases.

The increase in the completion rate for Clusters 2 and 6 was almost constant from the early to late phases. In these dialogues, the workers constantly interacted with each other, discussing and agreeing on the construction of the garden step by step.

4.3. Relationship between Collaboration Process and Evaluations

To identify which cluster was associated with good collaboration, we investigated the frequency of high-quality dialogues within the clusters. To classify the dialogues by their quality, we used 11 evaluation criteria: the seven subjective evaluations, the sum of all subjec-

tive evaluations, two third-person evaluations, and the sum of all third-person evaluations, and we classified the top 20% as high-quality dialogues and the bottom 80% as poor-quality dialogues in descending order by score. We created a 2×2 cross table of the 500 dialogues with one axis indicating whether a dialogue was in a specific cluster and the other indicating whether a dialogue was among the high-quality dialogues, and we conducted a Fisher's exact test to find the collaboration process that led to high-quality collaborative work.

In Cluster 3, we verified that high-quality dialogues appeared significantly ($p < 0.05$) more frequently in those of high evaluation on the basis of the total score of the two third-person evaluations. In Cluster 5, we verified that high-quality dialogues appeared significantly ($p < 0.05$) more frequently in those of high evaluation on the basis of the total score of the seven subjective evaluations. Clusters 3 and 5 are clusters in which the completion rate increased significantly in the middle. In these dialogues, there were few utterances until the middle of the dialogue when the dialogue resumed. This happened because the workers agreed on the central concept of the garden in the early phase and worked on it smoothly.

The dialogue in Table 2 had two such phases, an early phase and a later phase, with no conversation for about 10 minutes in between. In the early phase, the workers discussed the garden concept of a wisteria trellis and agreed to work on it. On the other hand, in the last phase, they discussed chairs and desks by looking at what they created. This dialogue was included in Cluster 3 and was in the top 20% of all subjective and third-person evaluations.

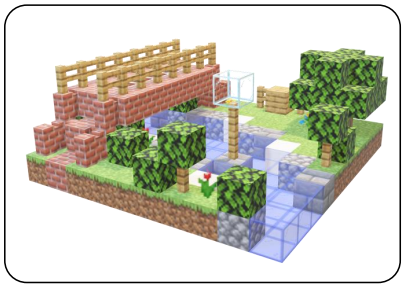
Figure 4 shows the gardens created from dialogues in Cluster 3 and Cluster 5, respectively. The following is an excerpt from an early phase of a dialogue included in Cluster 3. This dialogue is denoted as Dialogue A.

(1) A: 小さくていいので橋みたいなのをつくりたくて・・どうでしょう？
(I would like to create a small bridge or something like that... how about it?)
B: いいですね！じゃあ小川もですね
(That's great! Then let's create a creek.)

The following is an excerpt of a dialogue included in Cluster 5, and this dialogue is denoted as Dialogue B.

(2) A: じゃあ花壇やってみますか．．．？
(Well, let's create a flower bed...?)
B: そうしましょう！私、屋根付きルーフみたいなの作ってみたいんですけど
(Yeah, let's do that! I would like to create something like a covered roof.)

Dialogue A was in the top 20% for all evaluation criteria, and Dialogue B was in the top 20% for all evaluation criteria except for originality and the total score



Garden from Dialogue A

Garden from Dialogue B

Figure 4: Examples of collected gardens

of the two third-person evaluations. These utterances, described above, appeared in the early phase of the dialogues when the workers discussed the garden concept and came up with ideas followed by agreement. In the later phase, we can see the following exchange in Dialogue A.

(3) A: 橋の真ん中をガラスにしてみたんですがどうでしょう？
(I tried to turn the middle of the bridge into glass blocks, how about that?)
B: おぉ～魚が見えそう、可愛いと思います！
(Wow, I can almost see the fish, it is pretty!)

In Dialogue B, the following exchange was observed in the later phase.

(4) A: 白だと目立ちすぎるかなあ
(I guess white is too conspicuous.)
B: んーちょっと高級感が出るかもです
(Hmmm... maybe it will look a little more luxurious.)

These utterances appeared in the later phase, when the workers discussed the details of the garden and worked on it while looking at the created objects.

In contrast to Cluster 3 and Cluster 5, Cluster 2 and Cluster 6 did not include significantly more high-

quality dialogues. Although we expected constant discussion and consensus to lead to better collaboration, it was interesting to note that this was not the case.

From the analysis, we discovered that agreeing on a particular image of the final product in the early phase and then discussing the details to complete the garden is essential in both subjective and third-person evaluations. We consider it essential to develop skills for dialogue systems to be able to discuss basic concepts before creation and discuss the details while looking at the created objects.

5. Conclusion and Future Work

We collected data on human interactions in the Collaborative Garden Task in Minecraft with the aim of developing a dialogue system that enables creative collaborative work. We collected subjective and third-person evaluations of dialogues and gardens and analyzed dialogic factors that were evaluated highly from several aspects. As a result, we discovered that, in creative collaborative work, it is essential to perform more processes to ask for and agree on suggestions between workers and to agree on a particular image of the final product in the early phase of work and then discuss changes and details.

For a dialogue system to demonstrate such behaviors, we need to develop technologies for the system to discuss ideas rather deeply at the beginning of work, having its own ideas and exchanging them with users. The system also needs the capability to agree on a concept. In addition, at the end of the collaboration, the system needs to understand the current product of the work and needs to discuss possible improvements with users, putting the final touches on the work. One straightforward experiment we can consider is to train an encoder-decoder model that can output user utterances on the basis of the state of the garden and the dialogue history. Eventually, we will work on developing a system that can creatively collaborate with users.

6. Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 19H05692.

7. Bibliographical References

Allen, J., Chambers, N., Ferguson, G., Galescu, L., Jung, H., Swift, M., and Taysom, W. (2007). PLOW: A Collaborative Task Learning Agent. In *Proceedings of the 22nd National Conference on Artificial Intelligence - Volume 2*, pages 1514–1519.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., and Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4):351–366.

Bara, C.-P., CH-Wang, S., and Chai, J. (2021). MindCraft: Theory of Mind Modeling for Situated Dialogue in Collaborative Tasks. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1112–1125.

Ferguson, G. and Allen, J. F. (1998). TRIPS: An Integrated Intelligent Problem-Solving Assistant. In *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence*, pages 567–572.

He, H., Balakrishnan, A., Eric, M., and Liang, P. (2017). Learning Symmetric Collaborative Dialogue Agents with Dynamic Knowledge Graph Embeddings. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1766–1776.

Jayannavar, P., Narayan-Chen, A., and Hockenmaier, J. (2020). Learning to execute instructions in a Minecraft dialogue. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2589–2602.

Johnson, M., Hofmann, K., Hutton, T., and Bignell, D. (2016). The Malmo Platform for Artificial Intelligence Experimentation. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 4246–4247.

Kim, J.-H., Kitaev, N., Chen, X., Rohrbach, M., Zhang, B.-T., Tian, Y., Batra, D., and Parikh, D. (2019). CoDraw: Collaborative Drawing as a Testbed for Grounded Goal-driven Communication. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6495–6513.

Köhn, A., Wichlacz, J., Torralba, Á., Höller, D., Hoffmann, J., and Koller, A. (2020). Generating Instructions at Different Levels of Abstraction. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2802–2813.

Lochbaum, K. E., Grosz, B. J., and Sidner, C. L. (1990). Models of Plans to Support Communication: An Initial Report. In *Proceedings of the Eighth National Conference on Artificial Intelligence - Volume 1*, pages 485–490.

Meena, R., Skantze, G., and Gustafson, J. (2014). Data-driven models for timing feedback responses in a Map Task dialogue system. *Computer Speech & Language*, 28(4):903–922.

Mitsuda, K., Higashinaka, R., Oga, Y., and Yoshida, S. (2022). Dialogue Collection for Recording the Process of Building Common Ground in a Collaborative Task. In *Proceedings of the 13th Language Resources and Evaluation Conference*.

Narayan-Chen, A., Jayannavar, P., and Hockenmaier, J. (2019). Collaborative Dialogue in Minecraft. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5405–5415.

Ogawa, H., Nishikawa, H., Tokunaga, T., and Yokono, H. (2020). Gamification Platform for Collecting Task-oriented Dialogue Data. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 7084–7093.

Paparrizos, J. and Gravano, L. (2015). k-Shape: Efficient and Accurate Clustering of Time Series. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1855–1870.

Rich, C. and Sidner, C. L. (1999). COLLAGEN: A Collaboration Manager for Software Interface Agents. *Computational Models of Mixed-Initiative Interaction*, pages 149–184.

Sidner, C. L. (1994). An Artificial Discourse Language for Collaborative Negotiation. In *Proceedings of the Twelfth AAAI National Conference on Artificial Intelligence*, pages 814–819.

Szlam, A., Gray, J., Srinet, K., Jernite, Y., Joulin, A., Synnaeve, G., Kiela, D., Yu, H., Chen, Z., Goyal, S., et al. (2019). Why Build an Assistant in Minecraft? *Computing Research Repository*, abs/1907.09273.

Zhang, Y., Xia, G., Levy, M., and Dixon, S. (2021). COSMIC: A Conversational Interface for Human-AI Music Co-Creation. In *International Conference on New Interfaces for Musical Expression*.