Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 3150–3157
Marseille, 20-25 June 2022

A Speculative and Tentative Common Ground Handling for Efficient Composition of Uncertain Dialogue

Saki Sudo[1], Kyoshiro Asano[1], Koh Mitsuda[2], Ryuichiro Higashinaka[2], and Yugo Takeuchi[1]
[1]Shizuoka University: Hamamatsu, Shizuoka 4328011, Japan
[2]NTT Corporation: Yokosuka, Kanagawa 2390847, Japan
sudo.saki.16@shizuoka.ac.jp, asano.kyoshiro.16@shizuoka.ac.jp,
koh.mitsuda.td@hco.ntt.co.jp, ryuichiro.higashinaka.tp@hco.ntt.co.jp, takeuchi@inf.shizuoka.ac.jp

Abstract
This study investigates how the grounding process is composed and explores new interaction approaches that adapt to human cognitive processes that have not yet been significantly studied. The results of an experiment indicate that grounding through dialogue is mutually accepted among participants through holistic expressions and suggest that common ground among participants may not necessarily be formed in a bottom-up way through analytic expressions. These findings raise the possibility of a promising new approach to creating a human-like dialogue system that may be more suitable for natural human communication.

Keywords: common ground, analytic, holistic, data collection

1. Introduction

Common ground in dialogue is a set of information shared among participants that serves as a precondition for understanding individual utterances in dialogue with others. In other words, it is the basis of a common understanding in which the participants can grasp what is being said without needing excessively detailed explanation (Clark & Schaefer, 1989; Clark & Brennan, 1991). Moreover, Stalnaker (1978) roughly described common ground as follows: "the presuppositions of a speaker are the propositions whose truth he takes for granted as part of the background of the conversation… Presuppositions are what is taken by the speaker to be the COMMON GROUND of the participants in the conversation, what is treated as their *common knowledge* or *mutual knowledge*." When engaging in dialogue with others, people try to understand what the speaker is thinking and what they are saying or doing. In this way, it becomes easier to implicitly understand the meaning of the other's thoughts, words, and actions without having the other person explain the specific target of instructions, statements about ideas, or the intentions of actions taken.

Thus, for example, when two people who have worked together for a long time at a specific worksite discuss instructions about the work, we generally observe a dialogue with extremely simplified utterances that a third person could not understand. In such cases, indicative words such as "that," "this," and "it" are often used. Even if the specific object they refer to is not explicitly indicated, the work can be carried out smoothly without additional inquiry because they have established the belief that the object they refer to is the same thing/issue (Clark & Carlson, 1982). This belief arises from common ground between the participants that has been acquired through years of trial and error.

Clark and Schaefer consider dialogue a cooperative work process by participants toward a goal, and they explained the degree of establishing common ground through their "contribution model" (Clark & Schaefer, 1989). The contribution model has two stages, presentation and acceptance, and it is thought that the grounding is achieved through the joint action of the participants through these stages. However, although the contribution model shows a qualitative procedure of belief grounding among participants, it does not sufficiently explain its computational grounding procedure. Therefore, Traum (1994) proposed a network model for dynamic computation by reorganizing Clark and Schaefer's (1989) concept of "contribution" into seven types of base acts properly understood for joint belief formation, which were called "discourse units." This network model is based on a finite automaton network structure and assumes that the dialogue participants' internal states based on the acts of the current speech unit (Initiate, Continue, Ack, Repair, ReqRepair, ReqAck, Cancel) determine each grounding condition. Traum's model is similar to Clark's in that common ground in dialogue is achieved through the presentation and acceptance of information. Furthermore, Roque and Traum (2009) categorized the grounding process into nine stages and tried to comprehend the degree of grounding from each utterance.

However, such research still raises a question: Is it impossible to establish common ground in dialogue without empirically verifying that the participants have a common understanding of each other's beliefs by comparing their own beliefs with those of the other, as in the case described above? It is reasonable to adopt such a bottom-up approach to form the common ground in a dialogue between a dialogue system and a human, as in the work of Traum as well as related studies. However, it is also possible that such an approach is applied because the system designers themselves implicitly assume that the dialogue system is a first-time encounter with the human and that there is no shared experience or knowledge between them.

In this regard, Heller et al. noted that, when there is a possibility that the speaker shares knowledge with the listener in a dialogue, they communicate more efficiently and, by Grice's Maxims (Grice, 1975), use linguistic representations of that knowledge as symbols of common ground or provide an explanation of that knowledge (Heller, Gorman, & Tanenhaus, 2012). Furthermore, Keysar and colleagues experimentally demonstrated that dialogues produce utterances when using processing based on an egocentric perspective, since it is difficult to distinguish

between knowledge known only to oneself and knowledge shared with others during the dialogue process (Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003).

In this study, we focus on the grounding process achieved through the mutual presentation, reference, and acceptance of beliefs among participants in a dialogue. Moreover, based on a quantitative and chronological analysis of the formation and use of common ground in human interaction, we explore the existence of new interaction approaches that adapt to human cognitive processes, which have not yet received much research attention.

2. Dialogue and Grounding Process

2.1 Human Natural Communication

In general, people find it easier to talk with close friends, family members, and other people who live in the same environment or belong to the same organization. One reason is that there is likely to be a rich underlayer of shared information in such relationships. Conversely, this idea implies that if there is a lack of underlying information between two participants in a dialogue, they are likely to find it difficult to communicate. Therefore, when the shared experience and knowledge between the two participants is unknown or scarce, as in the case of dialogue between a dialogue system and a human, the person is likely to find it challenging to communicate smoothly. The designers of dialogue systems most certainly recognize this issue. Therefore, to form common ground with people, conventional dialogue systems try to elicit information from people by reacting cooperatively to their utterances and actively performing interactions that contribute to forming common ground, at least in the initial stage of the dialogue. However, this type of interaction, in which we repeatedly confirm or question the other person's speech, is expensive and risks being perceived as impolite in person-to-person dialogues (Brown & Levinson, 1987).

Consequently, in the case of person-to-person interaction, even though the other person is unknown, i.e., not a close friend or acquaintance, and the information they share is uncertain, they may not necessarily ask questions as actively as in a dialogue system. Instead, people may choose a measure of encouraging the other person to speak spontaneously by nodding in response to the other person's speech and thus eliciting the information necessary for grounding. On the other hand, when talking to someone, the person can start talking with the speculative assumption that the other person has at least a certain shared level of knowledge and experience. In this way, only when the possibility of a discrepancy between one's own and the other's assumptions becomes apparent, it is natural for people to take the low-cost approach of forming common ground by confirming each other's assumptions through dialogue based on Traum's finite automaton network.

2.2 Analytic/Holistic Expressions

In the previous section, we mentioned that forming common ground for dialogue can be approached in two ways:

(I) A grounding approach, exemplified by Clark and Traum, that grounds the mutually assumed object on an analytical description of the object to be shared as a topic through dialogue.

(II) A grounding approach in which one speaker describes an abstract impression or representation of an object based on the speaker's subjectivity and the other speaker bases the object on the mutual assumption that the object is consistent with it, rather than on an analytical description of the object that involves comparative objectivity.

In this study, in our analysis of the results of dialogue experiments using tangrams, as well as the results of Schober and Clark (1989) described in the next chapter, the utterance types classified in (I) above are defined as **Analytic** expressions, and the speech types classified in (II) are defined as **Holistic** expressions. Table 1 shows the annotations and examples of each type of utterance.

In the dialogue experiment described in the next chapter, the two participants in the experiment refer to each other's set of tangram figures (six tangram figures) on their side and the other's side only in dialogue, and they are tasked with naming their respective tangram figures in a "tangram naming task."

3. Experiment

3.1 Tangram Naming Task (TNT)

3.1.1 Tangram

In this experiment, the participants mutually exchange their ideas, views, and articulations of a given tangram shape, which are used to perform the Tangram Naming Task. A tangram is a dissection puzzle consisting of seven flat polygons, called "tans," that are put together to form shapes

Table 1: Annotations and examples of utterances

Utterance Type	Annotations and Examples of Utterance
Analytic	Speech specifically describes the individual polygons that make up the tangram shape and where they are located in the overall shape. Ex1) There are two equilateral triangles, one on the right and one on the left... Ex2) You see, a triangle on the top right...
Holistic	An utterance that subjectively describes the whole or a part of a whole tangram shape without mentioning the polygons that make up the tangram (e.g., "it looks like," "it seems"). Ex3) It looks like an animal, like a horse. Ex4) Yes, this is the one who is sliding in soccer.

(Figure 1). The objective is to replicate a pattern (given only an outline) generally found in a puzzle book using all seven pieces without overlap (Wikipedia "Tangram," 2022).

Tangrams are exceptionally practical for observing how people explain abstract figures to others. Therefore, tangrams are used in human interactions and in experiments to verify shared understanding in human-human or human-robot interactions (Foster et al., 2008; Spanger et al., 2009; Tokunaga et al., 2012).

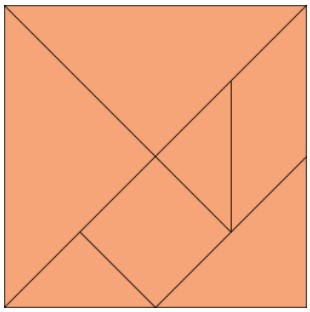

Figure 1: Seven tans of Tangram

3.1.2 Purpose of TNT

In the Tangram Naming Task (TNT), two participants in two different rooms must jointly name a set of six tangram figures displayed on a PC screen (one in each room) using only voice interaction. The six tangram figures are the same in each room; however, their placement and orientation on the screen are different (Figure 2). Therefore, the experiment participants must choose one of the six tangram figures and collaboratively decide on a name inspired by the shape. To accomplish this step, the participants need to mutually establish that the tangrams they choose are the same at the beginning of the TNT. After confirming that the selected tangram figure has the same shape, the next step is to name the selected tangram. These two-step tasks are repeated six times in one TNT for all six tangram figures.

In this experiment, we quantitatively elucidate the grounding process by measuring the time series of how a pair of participants working on TNT identify six mutually invisible tangrams and name them appropriately in each TNT session. Particularly, as an indicator of the grounding stage, we measure the frequency of the Analytic/Holistic expressions and the utterance ratio of these expressions from their verbal interaction described in Section 2.2 when the participants carry out each TNT.

3.2 Experimental Environment

3.2.1 Procedure

In this experiment, each participant works on the TNT twice. The first TNT is conducted with one partner (another participant); here, all participants interact with the others for the first time. In the second TNT, half of the participants work with the same partner as in the first TNT, while the other half work with a different partner than the one in the first TNT.

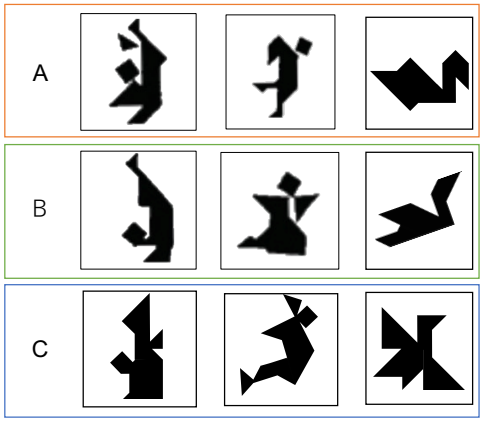

Figure 3: Three sets of tangram figures

A TNT employs nine types of tangram figures, shown in Figure 3, six of which were employed in the first TNT (A-B/B-C/C-A set) and six in different combinations in the second TNT (B-C/C-A/A-B set). Accordingly, at least three of the six tangram figures in the second TNT will be known figures that each participant has named in the first TNT, and the remaining three will be new to them. In this way, three of the six tangram figures presented in the first TNT session are also presented in the second TNT session, allowing us to examine the cognitive effects of the knowledge grounded by the first TNT session on the second TNT session.

Moreover, a TNT is limited to 30 minutes per session, during which time the six tangram figures must be mutually identified and named. In addition, the experimental participants are in different environments from each other and only they can see the set of tangram figures displayed on their respective PC screens. The dialogue of each TNT session between the two participants is recorded and analyzed as data, as described in the next section. Furthermore, the experiment participants are instructed on the following two points in advance:

1. The same tangram figures are displayed on the PC screens of both you and your partner.

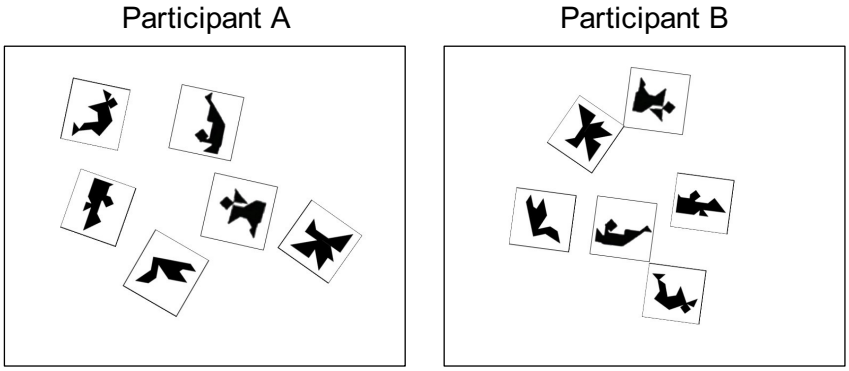

Figure 2: Six same-shape tangram figures presented to each participant

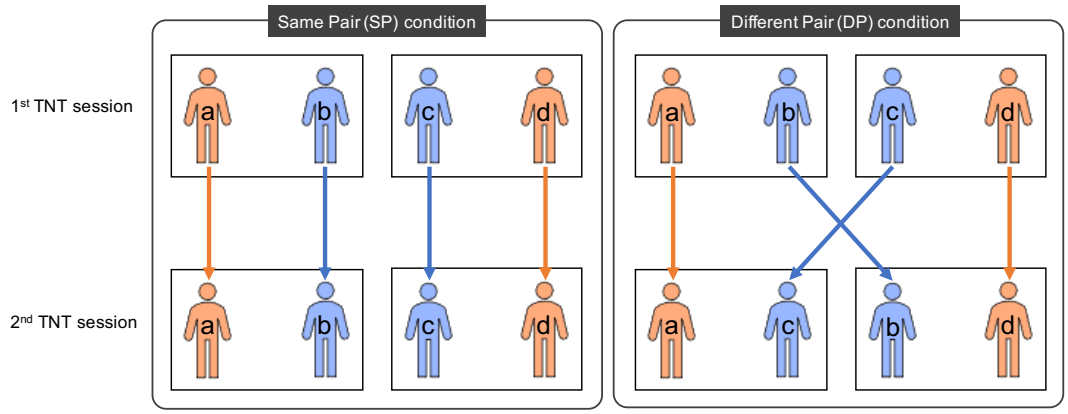

Figure 4: Arrangement design of participants for two experimental conditions

2. You can move or rotate the tangram figures on the screen.

3.2.2 Settings and Arrangement Design of Participants

Fifty-six native Japanese-speaking paid volunteers, recruited by a crowdsourcing company, participated in this experiment in an online environment.

We set up the following two experimental conditions in this experiment (Figure 4):

- **Same pair (SP) condition**: Two TNT sessions are performed with the same pair.
- **Different pair (DP) condition**: The second TNT is performed by a different pair than the first TNT pair.

In this experiment, the experimenter created 28 pairs of 56 participants at random. None of the participants in any of the pairs were acquainted with each other. Of the 28 pairs, 12 pairs were randomly assigned to the Same pair condition, while the remaining 16 pairs were assigned to the Different pair condition.

As an ethical consideration, the participants were informed through a written document in advance that they were free to discontinue their participation for any reason during the experiment. The participants then signed a consent form to participate in the experiment.

3.2.3 Hypothesis and Predictions of Results

Krauss and Glucksberg (1977) described a process in which two participants are asked to decide what to call an object by matching the cards of a novel figure, as in the present experiment. As a result, once common ground was established regarding what to call the object, the two parties used that name to indicate the object.

In this experiment, two unacquainted participants collaborating on TNT are expected to interact with each other through Analytic expressions, as described in Section 2.2, at the beginning of the task and then complete the grounding process through Holistic expressions. In other words, the participants exchange information for grounding through dialogues using Analytic expressions and then utter the grounded information and concepts using Holistic expressions. Accordingly, in the first TNT session, both the Same pair condition and Different pair condition participants interact mainly with Analytic expressions at the beginning of the session, with Holistic expressions gradually increasing from the middle of the session.

According to our hypothesis, noticeable differences should be observed in the dialogue of the second TNT session between the respective participants assigned to the Same pair and Different pair conditions described in the previous section, as follows:

A) The participants of the Same pair condition will continue to use the common ground formed in the first TNT session to start interacting with Holistic expressions from the beginning of the session.

B) The participants of the Different pair condition will interact with Analytic expressions at the beginning of the second TNT session because they are unacquainted with each other and do not share common ground such as that formed in the first TNT session.

These two predictions indicate that the number of Analytic and Holistic expressions in the participants' utterances may show a trade-off relation based on the condition.

3.3 Results and Analysis

3.3.1 Results

We collected 56 sets of dialogue from all TNT sessions, consisting of 10,639 utterances (Table 2). Using the classification of utterance expression in Section 2.2, the data means of counted utterances for Analytic and Holistic expressions in the first and second TNT sessions are shown in Table 3. In addition, the second TNT session shows each experimental condition described in Section 3.2.2.

Table 2: Collected data from all TNT sessions

1st TNT session			
28 sets of dialogue		*6,282 utterances*	
2nd TNT session			
Same pair condition		Different pair condition	
12 sets of dialogue	*1,622 utterances*	*16 sets of dialogue*	*2,735 utterances*

Table 3 : Mean numbers of utterances and percentages in each experiment condition

Mean numbers (percentages)	1st TNT session	2nd TNT session	
		SP condition	DP condition
All Utterances	224.36	135.17	170.94
Analytic Expression	24.79 (**10.90%**)	9.34 (**6.90%**)	16.69 (**9.76%**)
Holistic Expression	46.86 (**20.89%**)	28.67 (**21.21%**)	29.38 (**17.19%**)

Table 4: Example dialogue (Same pair condition)

	Utterances (Translated from Japanese)	Type
A	Do you know what this is, something like *a question mark*?	Holistic
B	Question mark?	
A	It looks like a question mark, or maybe *just a curve*.	Analytic
B	Mmm…	
A	You know, the ones that look like *they're bent*.	Analytic
B	Is that it? There are *two small triangles*, aren't there?	Analytic
A	Yes, that's right!	
B	Moreover, it also contains *a square*.	Analytic
A	And *square*, right?	Analytic
B	That is it!	

Tables 4 and 5 show examples of dialogues obtained in this experiment for each experimental condition.

A commercial annotation service annotated the utterances collected in this experiment. We confirmed the match rate of the three annotators based on how they had annotated the utterances in a sample of 11 out of all 56 dialogues. As a result, the Kappa coefficient was 0.94, which is a nearly perfect match.

Figure 5 (box plot[1]) shows the ratios of Analytic and Holistic expressions to total utterances in the naming of six tangram figures in order for the first TNT session. Figures 6 and 7 similarly show these ratios for the Same and Different pair conditions of the second TNT session, respectively.

The dots in Figures 5 through 7 indicate the ratios of Analytic/Holistic expressions to speech for each of the six tangram shapes listed in order from one to six, starting with the one named first. Therefore, in the case of tangrams where the number of utterances during the task is small, the number of dots is small and, in the case of tangrams where the number of utterances is large, the number of dots is large. Accordingly, for example, the TNT for the first three tangram figures in Figure 6 shows that the number of utterances between the participants required to accomplish the task was lower than the TNT for the subsequent three tangram figures.

Table 5: Example dialogue (Different pair condition)

	Utterances (Translated from Japanese)	Type
B	*A sea otter.*	Holistic
A	*A sea otter, trying to break the shell*, isn't it?	Holistic
B	Does *that sea otter have a flag on its head*?	Holistic
A	No, there's *no flag on that otter's head*… Rather than, well…, yes, a shell, *a square shell it has*.	Holistic
B	OK.	
A	*The feet are raised a little higher than the head*. If the *boat's bottom* is down, it looks like that to me.	Holistic
B	Oh…yes.	
A	It only has one of those... So there is *only one square*.	Analytic

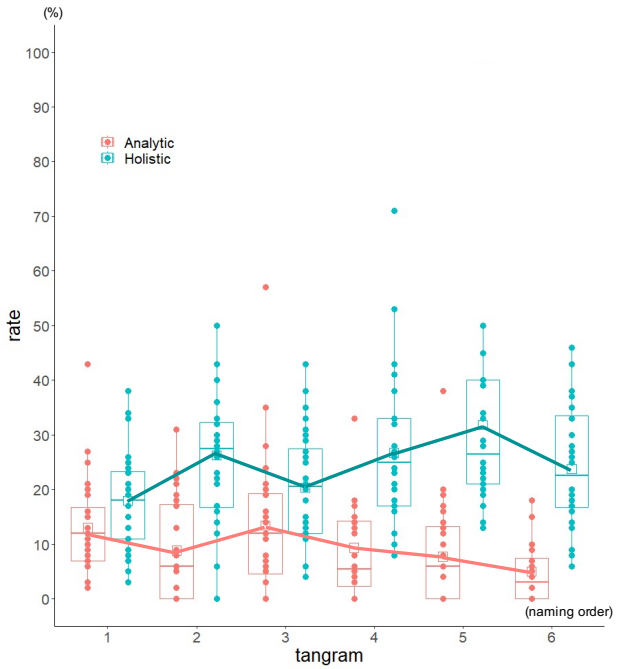

Figure 5: Ratio of Analytic and Holistic expressions of total utterances in first TNT session

[1] The box's upper end indicates the upper quartile, and the lower end indicates the lower quartile. The top end of the bar indicates the maximum value, and the bottom end of the bar indicates the minimum value. The horizontal lines in the boxes indicate the median values. Points not included at both ends of the bar are outliers.

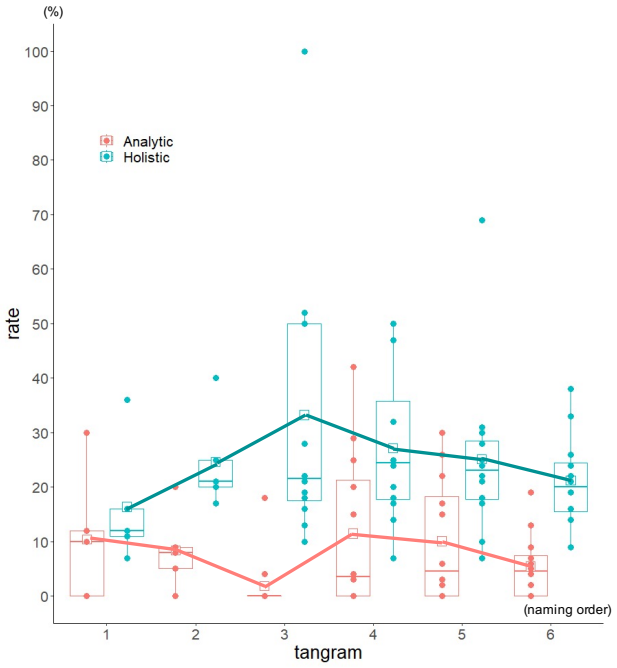

Figure 6: Ratios of Analytic and Holistic expressions to total utterances in Same pair condition in second TNT session

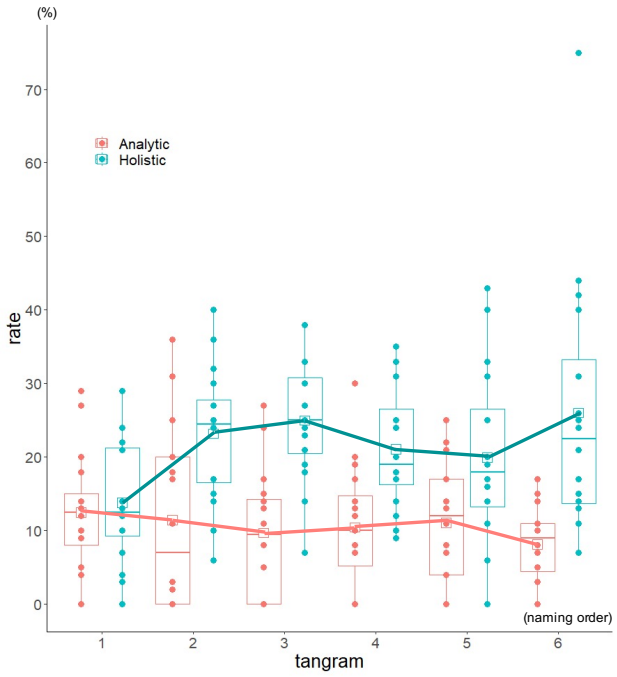

Figure 7: Ratio of Analytic and Holistic expressions to total utterances in Different pair condition in second TNT session

3.3.2 Considerations

First, we consider and analyze the experimental results shown in Table 3.

(a) Since the participants assigned to the Same pair condition work in the Same pair for two consecutive TNT sessions, the second TNT session can utilize the knowledge and concepts built upon in the first TNT session. Therefore, the percentage of utterances with Analytic expressions was lower in the second TNT session than in the first TNT session.

(b) The second TNT session of the participants assigned to the Different pair condition is essentially the same as the first TNT session, and no common ground has been established among the participants. Therefore, the percentage of utterances with Analytic expressions is about the same in the first and second sessions.

These results provide support for the hypothesis discussed in Section 3.2.3. However, we predicted that there would be an increase in Analytic expressions in the early stages of dialogue between unacquainted people with insufficient common ground. As the grounding progresses, there would be a decrease in Analytic expressions and an increase in Holistic expressions, but this could not be confirmed.

On the other hand, Figures 5, 6, and 7 illustrate the ratios of Analytic and Holistic expressions to total utterances in the first TNT session, in the Same pair condition in the second TNT session, and in the Different pair condition in the second TNT session, respectively. These results indicate that the ratios of the Analytic and Holistic expressions are comparable at the beginning of each TNT session. Furthermore, in the middle to the latter half of the interaction, there is a tendency for the percentage of utterances with Analytic expressions to decrease slightly as the naming of tangram figures progresses. On the contrary, the number of utterances with Holistic expressions was generally higher than that with Analytic expressions. This indicates that there is not necessarily a trade-off between the frequency of production of the two types of utterances described in Section 3.2.3. From the interaction partly illustrated in Table 5, even though common ground had not yet been established between the participants, one unilaterally spoke to the other in Holistic expressions. In addition, the other participant rarely spoke in a way that would allow the other participant to inquire about what the participant was indicating and, moreover, behaved as if they were sharing knowledge and concepts that had already been established.

In Figure 6, which shows the results of the Same pair condition in the second TNT session, the reason why the number of utterances while performing TNT on the first three tangram figures is significantly lower than the others is that these three tangram figures reappeared as the three figures named in the first TNT session. Therefore, in the second TNT session, the number of utterances was lower than the others because the first step was to confirm the existence of the tangram figures named in the first TNT session and then to name them again with their names from the first TNT session. On the other hand, in the Different pair condition, each participant's screen showed three tangram figures they had named in the first TNT session, but it was unclear whether they were the exact same figures as their partner's. As a result, the participants in this condition followed the same procedure as in the first TNT session, but their utterances were dominantly produced in Holistic expressions based on the general experience of naming and building common ground with the different partner in the first TNT session.

In summary, our experiment shows that grounding through dialogue is mutually accepted among participants through Holistic expressions and suggests that common ground among participants may not necessarily be formed in a bottom-up way through Analytic expressions. This implies a new dimension to the traditional approach to dialogue composition. In other words, through this experiment, we found that people do not necessarily ground themselves in mutually cooperative dialogue, as Clark's contribution model claims, and that the cognitive approach, in which we speculate on what the other person is indicating and talking about based on our tentative beliefs, may become an essential method for future dialogue research.

4. Discussion

The challenge of establishing common ground with others is not an easy one, and it is disconcerting to assume that one has been able to form mutually shared beliefs with others only from the linguistically expressed information of dialogue. Therefore, people attempt to infer the beliefs of others not only from linguistic information but also from the other's behavior and the context of their actions. However, such an approach of passively inferring others' beliefs based on what they say and how they behave is only possible if it is guaranteed that we will receive accurate information from them. Therefore, if such a situation is not guaranteed, we need to apply a different method. Approaches that infer the beliefs of others based on passively obtained information and then respond accordingly are expensive. On the other hand, the approach of naively hypothesizing the beliefs of others and evaluating such hypotheses through their dialogue is efficient and low-cost if the hypothesized beliefs generally reflect those of others.

The following properties can be considered for the high-cost/low-cost interactions in building common ground in dialogue as pointed out in Section 3.2.3.

High-cost interaction

An interaction that directs the other person's attention to the object through an analytical expression of what the directed object is.

1. It mentions the attributes and features of the object.
2. By sharing one attribute or feature of the object with others, the scope of the shared information is expanded. Finally, an understanding of the object is shared with others (grounding).
3. Questions, confirmations, and responses to partners are frequent.

Low-cost interaction

An interaction that directs the other person's attention to the object through words that express a subjective image or overarching concept.

1. The person directing attention to an object engages in dialogue under the speculative assumption that he or she shares knowledge of that object with the other person.
2. The other party tentatively responds that they share knowledge of the indicated object and continues to search for it as the dialogue progresses.
3. The dialogue progresses with iterations of step 1 (speculative assumption) and step 2 (tentative acceptance).
4. When both participants are convinced that they have succeeded in grounding, the grounding is treated as successful.
5. In cases where step 4 does not succeed, the "High-cost interaction" described above is performed. In that case, it is not a low-cost interaction.

The issue of costs associated with building common ground through such dialogue has been the subject of several discussions regarding efficient ways of providing information (Gegg-Harrison & Tanenhaus, 2016; Wu & Keysar, 2007). Recent related research suggests that the cognitive science perspective on the estimation of other people's minds and the discussion of statistical prediction of other people's minds inferred from utterance sequences are essential in examining the grounding process in dialogue (Hupet & Chantraine, 1992; Clark & Wilkes-Gibbs, 1986). Therefore, it is probably necessary to discuss grounding through dialogue based on models of human memory, mental lexicon and knowledge, and such a cognitive topic. In other words, a dataset without semantic information about the object itself, such as the identification of spatially arranged meaningless visual objects discussed in Udagawa and Aizawa (2019), cannot address issues related to privileged, shared, or novel knowledge in the process of groundedness or issues related to the amount of information that is available to the participants. Moreover, it is also inadequate for problems such as those in our study, where the participants are on an equal level and have equal information.

The experimental procedure and the dialogue data used in this study make contributions to the discussion of human mental activity from the perspective of cognitive science by observing how the differences between the reuse of knowledge and information at the individual level and such reuse at the interpersonal level manifest themselves in utterances in multiple dialogue sessions.

This study investigates how the grounding process is formed and explores novel interaction approaches that adapt to human cognitive processes. The results of this study's experiment indicate that grounding through dialogue is mutually accepted among participants through Holistic expressions and suggest that common ground among participants may not necessarily be formed in a bottom-up way through Analytic expressions. These findings contribute to a promising new approach to achieving a human-like dialogue system that is suitable for natural human communication.

The findings from this study provide valuable insights into the cognitive processing of the composition and use of common ground in dialogue. However, these findings must still be interpreted with some limitations. The first concern is whether the phenomena shown by the experiment's results are general enough to be observed under conditions other than those of the TNT in this study. As mentioned

previously, it has been reported in studies that Holistic utterances are observed more frequently than Analytic utterances, as also observed in this experiment, because this contributes to improving the efficiency of dialogue, which is a phenomenon frequently observed in general dialogue activities as well (Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003). Therefore, it is not possible to strongly assert from the results of the experiments in this study alone whether this is a generality that can be applied to dialogue universally. However, this possibility should not be rejected out of hand. As another issue, the findings of this experiment may depend on cultural differences in communication, including differences in the languages used. Since no experiments have been conducted in languages other than Japanese, and no experiments have been conducted in other cultures, the possible dependence on language and culture cannot be neglected. Suppose that language or cultural dependence exists in building common ground in dialogue. Such a mechanism would be interesting in itself. Moreover, this suggests the possibility that dialogue using speech translation systems, currently rapidly becoming popular, will need to be designed not only for language translation but also for the building and use of common ground in dialogue.

More detailed analysis and exploration are required in the future, including the above limitations to the conclusions drawn from this study. For example, in the current study, we did not analyze whether a unique dialogue occurred for all of the individual tangram shapes used in the experiment. Furthermore, the ontological aspects of the Holistic and Analytic expressions have not been considered. Consequently, unresolved issues such as these should be considered in future works.

References

Brown, P. & Levinson, S. C. (1987). "Politeness: some Universals in Language Usage." Cambridge: Cambridge University Press.

Clark, H. H. & Brennan, S. E. (1991). "Grounding in communication." In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition*. 127-149, American Psychological Association.

Clark, H. H. & Carlson, T. B. (1982). Hearers and Speech Acts, Language, 58:332-373.

Clark, H. H. & Schaefer, E. (1989). Contributing to discourse. Cognitive Science, 13(2): 259-294.

Clark, H. H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, 22(1), 1-39.

Foster, M. E., Bard, E. G., Hill, R. L., & Guhe, M. (2008). The roles of haptic-ostensive referring expression in cooperative, task-based human-robot dialogue. The 3rd ACM/IEEE International Conference on Human Robot Interaction (HRI2008), 295-302.

Grice, H. (1975). "Logic and Conversation." In P. Cole & J. Morgan (Eds.), *Syntax and Semantics*, Vol.3, Speech Acts, 41-58, New York: Academic Press.

Heller, D., Gorman, K. S., & Tanenhaus, M. K. (2012). To name or to describe: shared knowledge affects referential form. Topics in Cognitive Science, 4, 290-305.

Hupet, M. & Chantraine, Y. (1992). Change in repeated references: Collaboration or repetition effects? Journal of Psycholinguistic Research, 21(6), 485-496.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. Psychological Science, 11, 32-37.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. Cognition, 89, 25-41.

Krauss, R. M., & Glucksberg, S. (1977). Social and Nonsocial Speech. Scientific American, 236(2), 100-105.

Roque, A. & Traum D. R. (2009). Improving a Virtual Human Using a model of Degrees of Grounding. Proceedings of 21st IJCAI (IJCAI2009), 1537-1542.

Schober, M. F. & Clark, H. H. (1989). Understanding by addressees and overhearers. Cognitive Psychology, 21, 211-232.

Spanger, P., Yasuhara, M., Iida, R., & Tokunaga T. (2009). Using extra linguistic information for generating demonstrative pronouns in a situated collaboration task. Proceedings of PRE-CogSci2009, 1-8.

Stalnaker, R. (1978). "Syntax and Semantics." New York Academic Press, 9:315-332.

Tokunaga, T., Iida, R., Terai, A., & Kuriyama, N. (2012). The REX corpora: A collection of multimodal corpora of referring expression in collaborative problem solving dialogue. Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC2012), 422-429.

Traum, D. R. (1994). A Computational Theory of Grounding in Natural Language Conversation. Technical report, Rochester University (NY), Department of Computer Science.

Tangram: https://en.wikipedia.org/wiki/Tangram (last referred: 1st Jan, 2022).

Udagawa, T. & Aizawa, A. (2019). A Natural Language Corpus of Common Grounding under Continuous and Partially-Observable Context. Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI19), 7120-7127.

Wu, S. & Keysar, B. (2007). The Effect of Information Overlap on Communication Effectiveness. Cognitive Science, 31, 1-13.