A Multi-party Dialogue Dataset for Dialogue Goal Tracking in a Hospital Setting and How It Can Be Used in LLM Prompt Engineering Experiments

Weronika Sieińska, Angus Addlessee, Daniel Hernández García, Nancie Gunson, Marta Romeo, Christian Dondrup, Oliver Lemon
Heriot-Watt University, Edinburgh, UK
{w.sieinska, a.addlesee, d.hernandez_garcia, n.gunson, m.romeo, c.dondrup, o.lemon}@hw.ac.uk

Abstract

We describe a multi-party dialogue dataset, which we collected, annotated, and released on GitHub for public use. The dataset is specifically designed for the task of dialogue goal tracking. It consists of transcriptions of 35 conversational interactions between 2 human speakers and a humanoid social robot called ARI in a hospital setting. The robot is there to alleviate the workload of medical staff by providing patients with information related to the hospital. In the dataset, each utterance that states a goal of a human speaker, e.g., to go to the reception or to find out where they can get a cup of coffee, is explicitly annotated with that goal. In this paper, we also describe a computational experiment we conducted with the use of the dataset to illustrate how it can be used. We prompt engineered 5 large language models for the task or dialogue goal tracking. While some of the models performed very poorly, others were able to grasp the task quite well and predicted most goal annotations correctly.

1 Introduction

Today's voice assistants are typically dyadic, with a single user interacting with a single system. However, as dialogue systems are getting deployed on social robots and placed in public spaces (Gunson et al., 2022; Moujahid et al., 2022), these systems are increasingly required to deal with challenges of multi-party dialogue. Importantly in this paper, they need to track user goals, which can be shared between multiple people or even answered not by the system but by other human speakers.

Regardless of the number of users, in order for a conversational system to work, it needs to contain a control mechanism for tracking the state of a dialogue, which is a separate, however similar, challenge. Researchers have been interested in tracking the state of a dialogue for years (Larsson and Traum, 2000; Williams and Young, 2007; Wang and Lemon, 2013; Ren et al., 2018; Balaraman et al., 2021). In 2013, Williams et al. (2013) started a series of scientific competitions called Dialogue State Tracking Challenge[1] (DSTC). In 2024, the dialogue research community can participate in the competition for the 12th time[2].

Dialogue goal tracking, on the other hand, is a form of dialogue system evaluation, especially in task-oriented (also called *goal*-oriented) dialogues, which creates the need for robust goal tracking strategies and suited datasets.

Researchers have been collecting multi-party dialogue data for years, some of which is even multimodal (Robinson et al., 2004; Djalali et al., 2012; Yamasaki et al., 2012; Mahajan and Shaikh, 2021; Reverdy et al., 2022). The existing variety of datasets also serve various purposes. Some datasets were constructed for the task of building common ground between different parties (Furuya et al., 2022), whereas others – for modeling social phenomena in discourse (Shaikh et al., 2010). Chen et al. (2020) built a multi-party dialogue dataset for the analysis of emotions and interpersonal relationships between speakers. To our knowledge, however, there are no available datasets built specifically for the task of speaker's goal tracking in human-robot interaction.

2 Multi-party Dialogue Dataset

In this paper, we describe a novel multi-party dialogue dataset consisting of transcriptions of 35 interactions between 2 human speakers and a humanoid robot called ARI (Cooper et al., 2020) in a hospital setting. The robot is there to alleviate the workload of medical staff by providing patients and their companions with information related to the hospital. We de-

[1]The competition is now known as Dialogue System Technology Challenge.
[2]https://dstc12.dstc.community/

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

signed our dataset specifically for the task of multi-party dialogue goal tracking and released it as a GitHub repository: `https://github.com/wsieinska/multi-party-dialogue-dataset`.

We annotated the data for speakers, addressees, and goals of speakers such as to get a cup of coffee, to find lifts, to go to the toilet, etc. We differentiate between *individual* goals – when only 1 speaker has the goal; and *shared* goals – when both speakers have the same goal (e.g., they both want to eat something). We think that, in multi-party interactions, the distinction between individual and shared goals may affect the way they are answered, and, hopefully, make the interactions feel more natural.

We used ELAN[3] for annotation, which is a tool for annotating audio and video recordings (Brugman and Russel, 2004). We describe in detail how the data was collected in Appendix A and how it was annotated in Appendix B. Dataset statistics can be found in Appendix C. Appendix D contains an example dialogue from our dataset.

3 Computational Experiment

We conducted a computational experiment with the use of our dataset. We prompt engineered 5 large language models (LLMs) to perform goal tracking in multi-party conversations, namely: GPT-4o, GPT-4 Turbo, GPT-3.5 Turbo, Vicuna-13b-v1.5-16k, and Llama-2-13b-chat-hf-16k. The prompt we used can be found in Appendix E.

We took the few-shot learning approach and added 3 training examples to the prompt (dialogues 1, 11, and 21) leaving 32 dialogues for testing (3 was the highest possible number due to memory limitations). For each test dialogue file, we created a copy and replaced goal annotations with blanks. The task for the LLMs was to return these dialogues with blanks filled in with their predictions of goal annotations. It can be divided into two subtasks: (1) return the same text of the given test dialogue, (2) replace blanks with predictions of goal annotations.

We evaluated performance at subtask 1 by computing similarity scores between generated dialogues and dialogues from our dataset. We used `python3 difflib.SequenceMatcher` as our metric. Then, to evaluate performance at subtask 2, we extracted predicted goal annotations and compared them to gold annotations from our dataset with the use of the same metric. However, due to the fact that the LLMs did not perform very well at (it would seem straightforward) subtask 1 (especially Llama-2-13b-chat-hf-16k), some generated dialogues needed to be slightly altered to enable automatic extraction of predicted goal annotations. Both altered and unaltered dialogues are available for comparison in our GitHub repository.

[3] `https://archive.mpi.nl/tla/elan`

Table 1 presents our experimental results. Each result is a mean of results obtained for all 32 dialogues used for testing. GPT-4o obtained the best results at both subtasks reaching 84% at subtask 1 and almost 80% at subtask 2. Llama-2-13b-chat-hf-16k performed the worst and did not even reach 5% of goal annotations predicted correctly.

Model	Subtask 1	Subtask 2
Llama-2-13b-chat-hf-16k	31.04 ± 17.49	4.89 ± 11.85
Vicuna-13b-v1.5-16k	61.02 ± 21.22	36.99 ± 38.03
GPT-3.5 Turbo	73.37 ± 19.92	63.54 ± 34.37
GPT-4 Turbo	77.71 ± 23.32	66.09 ± 39.22
GPT-4o	**84.09 ± 20.25**	**79.33 ± 30.89**

Table 1: Experimental results for subtasks 1 and 2.

4 Conclusions and Future Work

Multi-party dialogue goal tracking is a complex and challenging task. In order to solve it, multi-party dialogue data must be collected and annotated for speakers' goals. Therefore, we hope that our dataset will be a valuable contribution.

In our experiment, we tested the ability of 5 state-of-the-art LLMs to track goals of speakers in multi-party interactions. Some of the models were able to grasp the task quite well, however, there is still a lot of room for improvement.

In the future, it would be interesting to repeat our experiment with other prompts, e.g., a more detailed prompt explaining the reasoning behind how goals are annotated, and more training examples.

In this work, we were solely interested in the task of tracking goals of speakers. However, our dataset could be annotated for split utterances, coreferences, anaphoras, ellipses, and clarification requests; and used for other tasks.

Lastly, we appreciate that the size of our dataset is rather small. Hence, another useful follow-up to our work would be further data collection.

Acknowledgements

This research was funded by the EU H2020 program under grant agreement no. 871245 (SPRING project `https://spring-h2020.eu/`).

References

Vevake Balaraman, Seyedmostafa Sheikhalishahi, and Bernardo Magnini. 2021. Recent Neural Methods on Dialogue State Tracking for Task-Oriented Dialogue Systems: A Survey. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 239–251, Singapore and Online. Association for Computational Linguistics.

Hennie Brugman and Albert Russel. 2004. Annotating Multi-media/Multi-modal Resources with ELAN. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Yi-Ting Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. MPDD: A Multi-Party Dialogue Dataset for Analysis of Emotions and Interpersonal Relationships. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 610–614, Marseille, France. European Language Resources Association.

Sara Cooper, Alessandro Di Fava, Carlos Vivas, Luca Marchionni, and Francesco Ferro. 2020. ARI: the Social Assistive Robot and Companion. In *Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 745–751, virtual. IEEE.

Alex Djalali, Sven Lauer, and Christopher Potts. 2012. Corpus Evidence for Preference-Driven Interpretation. In *Logic, Language and Meaning*, pages 150–159, Berlin, Heidelberg. Springer Berlin Heidelberg.

Yuki Furuya, Koki Saito, Kosuke Ogura, Koh Mitsuda, Ryuichiro Higashinaka, and Kazunori Takashio. 2022. Dialogue Corpus Construction Considering Modality and Social Relationships in Building Common Ground. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4088–4095, Marseille, France. European Language Resources Association.

Nancie Gunson, Daniel Hernández García, Weronika Sieińska, Christian Dondrup, and Oliver Lemon. 2022. Developing a Social Conversational Robot for the Hospital Waiting Room. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1352–1357.

Staffan Larsson and David R. Traum. 2000. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering*, 6(3–4):323–340.

Khyati Mahajan and Samira Shaikh. 2021. On the Need for Thoughtful Data Collection for Multi-Party Dialogue: A Survey of Available Corpora and Collection Methods. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 338–352, Singapore and Online. Association for Computational Linguistics.

Meriam Moujahid, Helen Hastie, and Oliver Lemon. 2022. Multi-party Interaction with a Robot Receptionist. In *17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 927–931. IEEE.

Liliang Ren, Kaige Xie, Lu Chen, and Kai Yu. 2018. Towards Universal Dialogue State Tracking. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2780–2786, Brussels, Belgium. Association for Computational Linguistics.

Justine Reverdy, Sam O'Connor Russell, Louise Duquenne, Diego Garaialde, Benjamin R. Cowan, and Naomi Harte. 2022. RoomReader: A Multimodal Corpus of Online Multiparty Conversational Interactions. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2517–2527, Marseille, France. European Language Resources Association.

Susan Robinson, Bilyana Martinovski, Saurabh Garg, Jens Stephan, and David Traum. 2004. Issues in Corpus Development for Multi-party Multi-modal Task-oriented Dialogue. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Samira Shaikh, Tomek Strzalkowski, Aaron Broadwell, Jennifer Stromer-Galley, Sarah Taylor, and Nick Webb. 2010. MPC: A Multi-Party Chat Corpus for Modeling Social Phenomena in Discourse. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Zhuoran Wang and Oliver Lemon. 2013. A Simple and Generic Belief Tracking Mechanism for the Dialog State Tracking Challenge: On the believability of observed information. In *Proceedings of the SIGDIAL 2013 Conference*, pages 423–432, Metz, France. Association for Computational Linguistics.

Jason Williams, Antoine Raux, Deepak Ramachandran, and Alan Black. 2013. The Dialog State Tracking Challenge. In *Proceedings of the SIGDIAL 2013 Conference*, pages 404–413, Metz, France. Association for Computational Linguistics.

Jason D. Williams and Steve Young. 2007. Partially observable Markov decision processes for spoken dialog systems. *Computer Speech and Language*, 21(2):393–422.

Shota Yamasaki, Hirohisa Furukawa, Masafumi Nishida, Kristiina Jokinen, and Seiichi Yamamoto. 2012. Multimodal Corpus of Multi-party Conversations in Second Language. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 416–421, Istanbul, Turkey. European Language Resources Association (ELRA).

A Data Collection

We collected a dataset of 35 interactions between 2 human speakers and a humanoid robot called ARI in a hospital setting. We did that in the "Wizard of Oz" setup. Each interaction was recorded using cameras both on ARI itself and external ones.

In the videos, one can see two human speakers (the participants of the data collection) standing next to each other. Please note that the released dataset only contains written transcriptions of the interactions. We were not able to release videos due to privacy regulations (videos contained personally identifiable information of data collection participants – their faces).

Human speakers were assigned particular roles. One of them was a patient who came to the hospital to attend an appointment with a medical doctor, whereas the other was their companion.

Participants were also given tasks to complete in each interaction with ARI. They were supposed to: retrieve information about the location of lifts, room 17, and toilets; as well as find out where they can get something to eat, where they can get a cup of coffee, and what time they should expect their appointment to commence at. In the dataset, goal annotations often reflect the tasks participants were trying to complete. Figure 1 presents picture representations of the tasks given to the participants. The pictures allowed us to avoid suggesting the use of any particular words and fostered more diverse wording in the dataset.

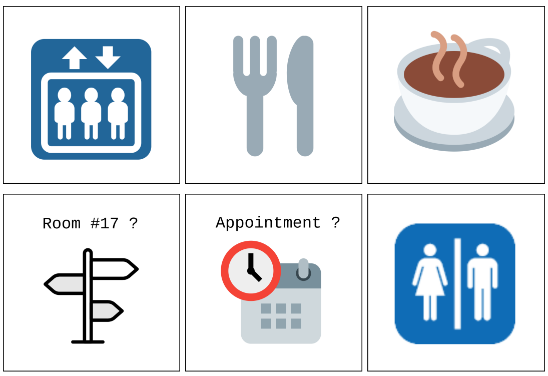

Figure 1: Picture representations of tasks given to participants during data collection.

The tasks were supposed to give participants an idea about what kind of information they can retrieve from ARI, however, participants were welcome to make other hospital-related requests, e.g., ARI was asked whether the hospital cafeteria serves cakes and whether consultations are covered by social security health insurance.

B Data Annotation

We annotated the data for speakers, addressees, and goals of speakers. All of the data was annotated by the first author of this paper, and 20% of the data was also annotated by the second author. Overall, the authors agreed with each other's annotations in 96.08%.

B.1 Speaker Annotation

Speaker is the participant who uttered the given utterance. It is either patient (`Pat`), companion (`Com`), or ARI (`ARI`). Speaker annotations were determined by the analysis of videos, in particular: head and body movements, and voice timbres. Unfortunately, it was not possible to determine who is speaking at the given moment by looking at participants' lips as they were covered by face masks. The inter-annotator agreement for speaker annotations is 100.00%.

B.2 Addressee Annotation

Addressee is the participant who the given utterance is addressed to. Similarly to speaker annotation, addressee annotation required the analysis of videos, head and body movements in particular. Sometimes, the speaker would address someone by their name, making the addressee annotation task trivial, e.g., "So, Mrs Companion, do you know what I'll be eating today?", "ARI, I've been waiting a long time, I'm tired.". Possible values of the addressee annotation are the following: ARI (`ARI`), patient (`Pat`), companion (`Com`) – one addressee; patient and companion (`Pat+Com`) – ARI addressing both human speakers; patient and ARI (`Pat+ARI`), companion and ARI (`Com+ARI`) – a human speaker addressing ARI and the other human speaker. The inter-annotator agreement for addressee annotations is 98.53%.

B.3 Goal Annotation

In each interaction, the patient and the companion have certain goals, which often reflect the tasks the participants were given during data collection (to get a cup of coffee, to find lifts, to go to the toilet, etc.). The inter-annotator agreement for goal annotations is 89.71%.

If a patient (`Pat`) has a goal to go to the hospital reception, the syntax of the goal annotation is the following: `G(Pat, go-to(reception))`. All goal annotations from this dataset are listed below (each of the annotations can represent a goal of any human speaker `Pat/Com`):

- `G(Pat, drink(<ARG>))` – the patient is thirsty and they specified that they would like to drink `<ARG>`, where `<ARG>` is, e.g., coffee, hot chocolate, tea, water, etc.;
- `G(Pat, eat(<ARG>))` – the patient is hungry and they specified that they would like to eat `<ARG>`, where `<ARG>` is, e.g., a piece of cake, croissant, sandwich, etc.;
- `G(Pat, get-info(<ARG>))` – the patient would like to get information about `<ARG>`, where `<ARG>` is, e.g., their appointment, day schedule in the hospital, cafeteria opening times, etc.;
- `G(Pat, go-to(<ARG>))` – the patient would like to go to `<ARG>`, were `<ARG>` is, e.g., the cafeteria, courtyard, lift, reception, toilet, vending machine, etc.;
- `G(Pat, sit-down())` – the patient is tired and would like to sit down.

If an argument is missing in the `G(Pat, drink())` or the `G(Pat, eat())` goal annotations, it means that the patient is thirsty/hungry but did not specify what they would like to drink/eat. In the dataset, the argument is always present for the `G(Pat, get-info(<ARG>))` and the `G(Pat, go-to(<ARG>))` goal annotations. `G(Pat, sit-down())` does not take an argument.

Other goal annotations, which are rare but also occur in the dataset, are: `G(Pat, request-escort(<ARG>))` – here `<ARG>` is a location and is always specified, `G(Pat, request-volume-up())`, and `G(Pat, get-help())` which do not take an argument.

B.4 Types of Goal Annotations

There are 5 types of goal annotations (each of the annotations can represent a goal of any human speaker `Pat/Com`):

- `G(Pat, get-info(cafeteria(location)))` – "open goal" – used when the patient asks for the location of the hospital's cafeteria by saying, e.g., "Where can I find the cafeteria?".
- `AGP(Pat, get-info(cafeteria(location)))` – "answer goal (positive)" – used when ARI or the companion gives the patient the information they requested by saying, e.g., "There's a cafeteria on the ground floor, near the courtyard.".
- `AGN(Pat, get-info(cafeteria(location)))` – "answer goal (negative)" – used when ARI or the companion expresses their inability to provide requested information by saying, e.g., "Sorry, I don't have this information.".
- `CGP(Pat, get-info(cafeteria(location)))` – "close goal (positive)" – used when the patient acknowledges they have received the requested information by saying, e.g., "Ok, great, thanks.".
- `CGN(Pat, get-info(cafeteria(location)))` – "close goal (negative)" – used when the patient acknowledges they will not receive the information they requested by saying, e.g., "Oh well, thanks anyway.".

Each utterance that states a goal is explicitly annotated with that goal – even if that particular goal has already occurred before and is still open. There is no need for more types of goal annotations: `RG(Pat, go-to(reception))` – "reopen goal" – is not necessary because it can be treated just like opening a new goal (it does not matter that the same goal has already occurred in the dialogue and that it is closed). We decided to take this approach for simplicity.

B.5 Shared Goal Annotation

All goal annotations described so far are examples of *individual* goal annotations – they describe goals of individual participants (the patient or the companion). Some goal annotations, however, describe goals, which are *shared* by the participants. We think that, in multi-party interactions, the distinction between individual and shared goals may affect the way they are answered, and, hopefully, make the interactions feel more natural, e.g., if a shared goal was opened, it could be more natural for ARI to address both participants while answering it, not just the one who was the speaker and opened it. Addressing both participants instead of just one of them could be reflected in the wording of the answer, ARI's head pose, ARI's gestures, etc.

Shared goals are built similarly to individual ones. Participants sharing a goal are joined by the "+" sign: `Pat+Com` (the order does not matter, however, it is always `Pat+Com` (not `Com+Pat`) in the dataset (for simplicity)), and if their goal is to eat a sandwich, the annotation is the following: `G(Pat+Com, eat(sandwich))`.

A goal counts as shared when the speaker uses the word "we", e.g.:

- `Pat`: "How does it work here? We don't have any information. Is there any schedule for the day?" →`G(Pat+Com, get-info(day_schedule))`
- `Pat`: "And how do we get to the cafeteria?" →`G(Pat+Com, get-info(cafeteria( directions)))`
- `Pat`: "Could we have a little hot chocolate?" →`G(Pat+Com, drink(hot_chocolate))`

A goal also counts as shared when the speaker says "Me too." (or the like) following the specification of the other participant's goal, e.g.:

- `Com`: "I'm thirsty. I would like a glass of water." →`G(Com, drink(water))`
  `Pat`: "Oh yes, me too. Do you think there's a water fountain?" →`G(Pat+Com, drink(water))`
- `Com`: "Could you wait for me here? I need to go to the toilet." →`G(Com, go-to(toilet))`
  `Pat`: "I need to go too. I'll go with you." →`G(Pat+Com, go-to(toilet))`
- `Pat`: "I would grab a bite, I'm getting hungry." →`G(Pat, eat())`
  `Com`: "So am I. ARI, where can we get something to eat?" →`G(Pat+Com, eat())`

C Dataset Statistics

We analysed the data in terms of the number of turns, number of tokens (words), and the number of goal annotations. Table 2 presents statistics describing our dataset. On average, a dialogue from our dataset consists of 29.8 turns and includes 271.71 tokens and 8.17 individual `G`-type goal annotations.

D Example Dialogue

Table 3 presents an example dialogue from our dataset. In the dialogue, the patient and the companion want to eat a snack, go to the cafeteria, go to the toilet, and find out their appointment time. Their goals are opened, answered, and closed.

	Mean	St.Dev.	Min	Max
Turns	29.80	15.20	12	67
Tokens	271.71	162.19	86	766
Ind. G	8.17	6.89	1	30
Ind. AGP	3.74	3.32	0	11
Ind. AGN	1.71	2.30	0	9
Ind. CGP	1.69	1.43	0	4
Ind. CGN	0.60	0.81	0	2
Sh. G	2.63	2.28	0	8
Sh. AGP	1.51	1.63	0	6
Sh. AGN	0.60	0.91	0	3
Sh. CGP	0.74	0.92	0	3
Sh. CGN	0.23	0.60	0	3

Table 2: Dataset statistics (Ind. – Individual, Sh. – Shared, St.Dev. – Standard Deviation).

E The Prompt

I will give you a dialogue between two people, whose names are Pat and Com, and a robot, whose name is ARI. The dialogue will consist of multiple dialogue turns in the following format: "turn speaker->addressee: *utterance* `@goal$`". If the dialogue is "01 Pat->ARI: *I would like a cup of coffee, please.* `@G(Pat, drink(coffee))$`", then '01' is the turn number, 'Pat' is the speaker, 'ARI' is the addressee,"*I would like a cup of coffee, please.*" is the utterance, and "`@G(Pat, drink(coffee))$`" is an annotation of the goal of the speaker. However, each goal annotation will be replaced with the '`@[BLANK]$`' tag. I want you to guess missing goal annotations and return the dialogue with blanks filled in. You will find this dialogue between the '`<START>`' and '`<END>`' tags. Do not return any other text. I will also give you three example dialogues to learn from. Do not return the text of example dialogues or any other text. Remember, your task is to return the text between the '`<START>`' and '`<END>`' tags with the '`@[BLANK]$`' tags replaced by your guesses of goal annotations.
Example dialogue 1:
{example_dialogue_1}
Example dialogue 2:
{example_dialogue_2}
Example dialogue 3:
{example_dialogue_3}
Here is the dialogue, which I want you to return with blanks filled in:
<START>
{dialogue_with_blanks}
<END>

T.	Sp.→Add.	Utterance	Goal
01	ARI→Pat+Com	Hello, how can I help you?	–
02	Com→ARI	Hello.	–
03	Pat→Com	Well, it's my first time here, I don't know if this is the case for you. I'd really like to um... to be able, to be able to eat a little bit before going to my appointment, do you know where that is?	`G(Pat, get-info(food(location)))`
04	Com→Pat	Oh, I don't know. We'll ask the question.	`AGN(Pat, get-info(food(location)))`
05	Com→ARI	I don't know what your name is. Can you give us information and tell us what your name is?	`G(Com, get-info(ari(name)))`
06	ARI→Com	Hello, my name is ARI. How can I help you?	`AGP(Com, get-info(ari(name)))`
07	Com→ARI	Where's the cafeteria?	`G(Com, get-info(cafeteria(location)))`
08	ARI→Com	There are a few options available as part of your visit, and there's also a cafeteria on the ground floor.	`AGP(Com, get-info(cafeteria(location)))`
09	Com→Pat	Ok, on the ground floor.	`CGP(Com, get-info(cafeteria(location)))`
10	Pat→ARI	And how do we get to the cafeteria?	`G(Pat+Com, get-info(cafeteria(directions)))`
11	ARI→Pat+Com	You have to enter the second building behind you. Then it's the second door on the left.	`AGP(Pat+Com, get-info(cafeteria(directions)))`
12	Pat→ARI	Right, the second building.	`CGP(Pat+Com, get-info(cafeteria(directions)))`
13	Com→ARI	I'd really like to know what's on the menu.	`G(Com, get-info(menu))`
14	ARI→Com	Today we have an endive salad to start, followed by a ham and cheese escalope with a side of mashed potato and a fruit compote for dessert.	`AGP(Com, get-info(menu))`
15	Pat→ARI	Yes, but that's for lunch. We would like a snack, wouldn't we? Grab a coffee or breakfast, since we're early for our appointment.	`G(Pat+Com, eat(snack))`
16	ARI→Pat+Com	Of course. There are a few options available as part of your visit, and there's also a cafeteria on the ground floor.	`AGP(Pat+Com, eat(snack))`
17	Com→Pat	And I'd really like to go to the toilet too.	`G(Com, go-to(toilet))`
18	Com→ARI	Can you tell us where the toilets are?	`G(Com, get-info(toilet(location)))`
19	ARI→Com	Very easy! You have to go to the corridor at the end of the room. It's the second door on the right.	`AGP(Com, get-info(toilet(location)))`
20	Com→ARI	Second door on the right. Oh well, we'll find it. Thank you.	`CGP(Com, get-info(toilet(location)))`
21	Pat→ARI	Next to... So it's next to the cafeteria?	`G(Pat, get-info(toilet(location)))`
22	ARI→Pat	Absolutely.	`AGP(Pat, get-info(toilet(location)))`
23	Com→Pat	I feel like it's further away. That, that's not on the ground floor...	–
24	Pat→ARI	I'm having a difficulty understanding what you're saying. I can't hear very well. Could you repeat where the toilets are, please?	`G(Pat, get-info(toilet(location)))`
25	ARI→Pat	Very easy! You have to go to the corridor at the end of the room. It's the second door on the right.	`AGP(Pat, get-info(toilet(location)))`
26	Pat→ARI	In the corridor at the end of the room, ok...	`CGP(Pat, get-info(toilet(location)))`
27	Pat→ARI	Um, how do I know what time my appointment is at?	`G(Pat, get-info(appointment(time)))`
28	ARI→Pat	Sorry, I don't have this information.	`AGN(Pat, get-info(appointment(time)))`
29	Pat→ARI	That's a shame. Um...	`CGN(Pat, get-info(appointment(time)))`
30	Com→Pat	Maybe it's at the reception. Maybe there's a service, a receptionist who will let you know, who will tell you. Let's go there.	`G(Com, go-to(reception))`
31	Pat→Com	Ok, I'll go and ask at the reception then.	`G(Pat, go-to(reception))`
32	ARI→Pat+Com	Thank you. See you around. Have a nice day.	–
33	Com→ARI	Thank you.	–
34	Pat→ARI	Have a nice day.	–

Table 3: Example dialogue from our dataset where individual and shared goals are opened, answered, and closed (T. – Turn, Sp. – Speaker, Add. – Addressee).