

# Dialogue Support Through the Identification of Utterances Crucial for the Listener's Interpretation if Missed

Kenshin Nakanishi<sup>1</sup><sup>a</sup>, Tomoyuki Maekawa<sup>2</sup><sup>b</sup> and Michita Imai<sup>1</sup><sup>c</sup>

<sup>1</sup>Keio University, Yokohama, Kanagawa, Japan

<sup>2</sup>Shizuoka University, Shizuoka, Japan

**Keywords:** Dialogue Context Understanding, Online Dialogue, Missed Utterances, Large Language Model, Natural Language Processing.

**Abstract:** When a listener becomes distracted and misses an important utterance, it can hinder their understanding of the conversation and their subsequent responses. In this study, we developed a chat system that simulates the impact of missed important utterances using an algorithm that identifies contextually significant dialogue, which we have been researching previously. The system assesses whether each user utterance contains important context and, if so, notifies the user to alert them of the possibility of misunderstanding by the other party. The results showed that when important utterances were missed, the listener often misunderstood the flow of the conversation. However, the effectiveness of the assistance that alerts users to potential misunderstandings varied depending on the case, and it became clear that the benefits of this feature in a chat system are limited.

## 1 INTRODUCTION

In conversations, the meaning of sentence can often be ambiguous on its own and depends heavily on the context (Sumita et al., 1988). Missing critical contextual information during a conversation can lead to misunderstandings, which can pose risks to the progress of the dialogue. This issue is particularly challenging in online meetings, where it is difficult to verify whether others have accurately understood one's statements. The purpose of this paper is to support smooth communication by identifying potential misunderstandings caused by missed contextual information and assisting in resolving such misunderstandings.


In dialogue, a concept called SCAINs (Statements Crucial for Awareness of Interpretive Nonsense) (Maekawa and Imai, 2023). represents utterances that are contextually significant for interpreting a statement (hereafter referred to as the "core statement"). To identify SCAINs, both omitted conversation histories (where parts of the dialogue are omitted) and full conversation histories are used as sources of contextual information. Using these histories, the


utterances are paraphrased using a large language model (LLM), and differences in the paraphrases arising from contextual variations are detected via similarity calculations. SCAINs are identified by comparing these differences with a predefined threshold.


Conventional SCAIN extraction algorithms consider only the immediately preceding utterance for SCAIN identification. In our previous study (Nakanishi et al., 2024), we expanded the scope of SCAIN candidates to include utterances located further away in the conversation history. Nonetheless, JPersonaChat (Sugiyama et al., 2023) was used as the dataset, and it has not been tested up to real-time interaction.

In this paper, we explore whether SCAINs can assist in real-time chat conversations by highlighting missed contextual information and helping to resolve misunderstandings. When SCAINs are detected for a user's utterance, the system presents these SCAINs to the user, making them aware of the context-dependent nature of their statement. Furthermore, the system informs users of the potential misinterpretation that may arise if the SCAINs are missed, thereby aiming to facilitate the resolution of misunderstandings.

A case study was conducted to evaluate the system's the quantity and quality of its assistance, and users' awareness of misunderstandings with their conversation partners. The results demonstrated the ef-

<sup>a</sup> <https://orcid.org/0009-0007-5754-0816>

<sup>b</sup> <https://orcid.org/0009-0003-0350-1011>

<sup>c</sup> <https://orcid.org/0000-0002-2825-1560>

fectiveness of the assistance varied depending on the case. While the system successfully helped users recognize misunderstandings, limitations were observed in the perceived benefits of the system's assistance in a chat-based format.

## 2 RELATED WORK

### 2.1 Dialogue Summarization

Summarization is a useful method for succinctly representing the key points of a dialogue (Feng et al., 2022). Nihei and Nakano (2020) proposed a model and browser that utilize multimodal and multi-participant features to estimate important utterances that should be included in meeting summaries. By incorporating a deep learning model that uses the linguistic information of the speaker, as well as the audio and head movement data of all participants, the browser makes it easier to understand the content of the meeting and the roles of the participants. However, this method does not consider factors such as missed or overlooked utterances, nor does it retrieve critical contextual information for specific utterances.

### 2.2 Topic Shift Detection

Detecting topic changes is a crucial factor in understanding the flow of conversations. Kishinami et al. (2023) analyzed whether natural topic transitions could be detected using a corpus of casual conversations between humans. Additionally, a model was proposed to determine whether detected topics within a conversation are major topics (Konigari et al., 2021), demonstrating its usefulness for tasks such as dialogue summarization and information extraction. While topic shift detection contributes to understanding the general topics of a conversation, it does not identify contextually critical information for the current utterance. Moreover, it has not been examined whether topic shifts can alter the interpretation of subsequent utterances.

### 2.3 Identifying Important Missed Utterances: SCAINs

An existing system, SCAINs-Presenter (Tsuchiya et al., 2024), displays SCAINs during spoken dialogues. The system assumes a scenario where two participants engage in a dialogue while simultaneously performing chat tasks with chatbots, a setup designed to increase the likelihood of missed utter-

ances. During the chat subtasks, participants can refer to SCAINs displayed on the screen, enabling smoother progression in the primary task of spoken dialogue. However, the scope of SCAIN candidates is limited to the two utterances immediately preceding the core statement, and the potential for earlier utterances to be SCAINs is not considered. Additionally, while SCAINs are presented to individuals who may have missed them, there has been no investigation into whether the system can help resolve misunderstandings when the speaker's own utterances are ambiguous or unclear.

## 3 MISSED LISTENING SIMULATION SYSTEM

This section describes a simulation system designed to test whether it can help resolve discrepancies in interpretation caused by user speech that led to misunderstandings.

### 3.1 Overview

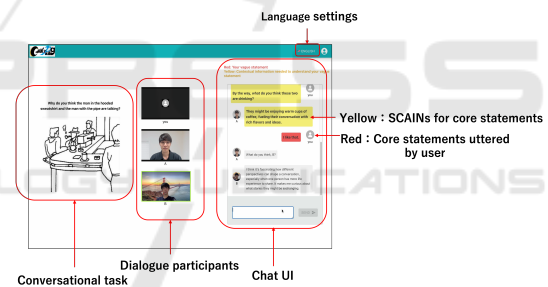


Figure 1: Simulation System UI – Interface with three participants (User, Agent A, Agent B) and SCAIN detection for the user's utterance. (Conversational task image is cited from Study 3 of Merged Minds (Rossignac-Milon et al., 2021)).

Figure 1 shows the UI, where discussions about the left-side image are conducted in a chat format. The central area represents three participants (User, Agent A, and Agent B), with the current speaker highlighted in green. The chat UI on the right treats the user's utterance as the core statement, performing SCAINs detection from the fourth utterance onward. If SCAINs are detected, the core statement is highlighted in red, and the SCAINs in yellow. Users can switch the interface language to Japanese via the header.

### 3.2 SCAINs Algorithm

Figure 2 illustrates the processing flow for extracting SCAINs, consisting of the following steps: prepar-

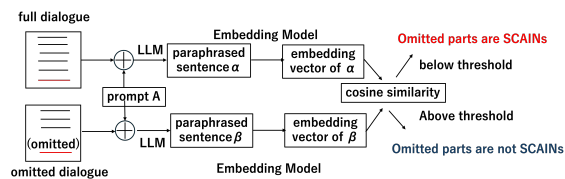


Figure 2: Process Flow for Extracting SCAINs.

ing dialogue texts, creating prompts for LLM input, vectorizing the LLM-generated text, and comparing similarities against a threshold. Detailed explanations of each step are provided in Sections 3.2.1 to 3.2.3.

3.2.1 Preparation of Conversation Histories

Two types of conversation histories are prepared for the current utterance: a full dialogue history with all utterances and an omitted dialogue history with selected omissions. As shown in Figure 3, utterances up to four positions preceding the core statement are treated as SCAIN candidates. For example, for the core statement #10, candidates include (#5, #6), (#6, #7), (#7, #8) and (#8, #9). The omitted dialogue history excludes these SCAIN candidates, and the processing in Figure 2 is executed four times for each core statement.

Previous research (Nakanishi et al., 2024) found that key contextual information is often within two preceding positions. However, this system considers up to four positions to evaluate the user’s impression of the quantity of extracted SCAINs.

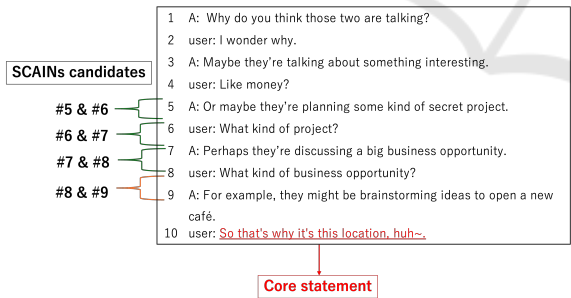


Figure 3: Example of SCAINs Candidate Selection (Core statement: 10th Line, Temporal Relative Position: 4).

3.2.2 Prompt for Paraphrasing Utterances Based on Conversation Context

Using both the full and omitted dialogue histories, the core statement is paraphrased with the help of a large language model (LLM). Table 1 presents the template for Prompt A, which is used to paraphrase utterances based on the conversation context. The conversation history and contextual information are provided to the LLM along with the core statement to guide the para-

Table 1: Prompt A: Template for paraphrasing utterances based on conversation context. {dialogue} represents the conversation history, and {core statement} is filled in with the core statement.

Rephrase a specific statement in a conversation to make it more concrete.

After a conversation between two people is input, the statement to be rephrased is specified. Rephrase the specified statement to be more concrete, avoiding the use of demonstratives, while using words from the conversation.

# Conversation  
{dialogue}

# Statement to Rephrase  
{core statement}

# Concrete Statement

phrasing process.

3.2.3 Embedding Paraphrased Utterances and Similarity Comparison

Table 2: Paraphrased utterances and cosine similarity in omitted dialogue for example in Figure 3.

Omitted Utterances	Paraphrased Utterance	Similarity
#5 & #6	Are they talking here because they're brainstorming ideas to open a new café?	0.97
#6 & #7	Are they talking here because they're brainstorming ideas to open a new café?	0.97
#7 & #8	Are they talking here because they're brainstorming ideas to open a new café?	0.97
#8 & #9	Perhaps they chose this location because they're discussing a significant business opportunity.	0.87

Paraphrased utterances based on full and omitted dialogues are embedded into vectors using an embedding model, with cosine similarity measuring vector similarity. High similarity suggests the omission of SCAINs candidates has minimal impact on the core statement, deeming them unimportant. Low similarity indicates significant impact, classifying the candidates as critical SCAINs. The similarity threshold, influenced by LLM performance and parameters, is

set through qualitative observation of paraphrased utterances.

Table 2 presents paraphrased utterances and cosine similarity scores in incomplete dialogues based on Figure 3. The paraphrased utterances were generated using gpt-4o-mini (OpenAI, 2023), and embeddings were created using text-embedding-ada-002 (Neelakantan et al., 2022). In the complete dialogue, the paraphrase was: "They chose this location because they're brainstorming ideas to open a new café."

As shown in Table 2, similarity scores drop significantly only when utterances #8 and #9 are omitted. In this case, the user's core statement, "Is that why it's this location?" connects the depicted image to a café, suggesting brainstorming about opening one. Omitting utterances #8 and #9 removes the café concept, shifting the dialogue to focus solely on business opportunities.

### 3.3 System Configuration

The system configures two AI agents (A and B) interacting with the user. Figure 4 shows the conversation history up to the first detection of SCAINs. The discussion starts with Agent A, and when the user makes a context-dependent ambiguous statement, the system highlights it in red and the SCAINs in yellow.

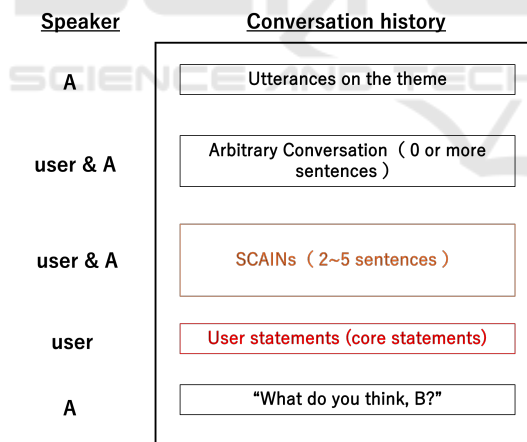


Figure 4: Conversation history when SCAINs occurred for the first time.

Agent A then prompts Agent B, who was initially distracted. Agent B uses an incomplete dialogue history, possibly missing critical context and misunderstanding the user's statement. The system helps the user identify the ambiguity in their own statement and resolve misunderstandings if Agent B's response is unrelated. The system continuously evaluates and highlights SCAINs to support smoother communication.

Table 3: Prompt B: A template for A's utterances. Input the conversation history in {dialogue} and the speaker's name in {agent}.

---

Please discuss, continuing from the following conversation:  
 # Dialogue:  
 {dialogue}

---

As {agent}, please make a concise statement in one sentence.

---

Prompt B for conversing with Agent A is shown in Table 3. Since the task is discussion-based, the instruction explicitly states, "Please discuss." Additionally, as the conversation progresses while referring to an image, both the prompt and the image are input into the LLM request to enable multimodal interaction.

Table 4: Prompt C: A template for B's utterances. Input the conversation history into {dialogue} and the speaker's name into {agent}.

---

As {agent}, please continue the following conversation with 1-2 sentences.

---

# Dialogue:  
 {dialogue}

---

Additionally, Prompt C for conversing with Agent B is shown in Table 4. Since B assumes the possibility of missing parts of the conversation, the instruction does not explicitly direct them to "discuss," but rather to respond naturally following the flow of the conversation. Furthermore, to check B's level of understanding, the prompt specifies that B's responses should be slightly longer—by about one additional sentence—compared to Prompt B.

## 4 CASE STUDY

A case study was conducted to investigate whether the system's presentation of SCAINs to participants during real-time chats with three parties helps resolve misunderstandings caused by ambiguous utterances. SCAINs were used to highlight the possibility of missed context by the other party. The research questions (RQs) are as follows:

- RQ1: Is the quality of dialogue and response time of the AI agents appropriate?
- RQ2: Is the system necessary for conversations involving missed utterances?
- RQ3: Is the frequency of assistance provided by the system appropriate?



- RQ4: Do the SCAINs outputs align with the users' intuition?

#### 4.1 Setup

Table 5: Parameters of a simulation system for case study.

Parameters	Values
Model	"gpt-4o-mini"
Embedding Model	"text-embedding-ada-002"
Temperature	0
Cosine Similarity Threshold	0.88
Time Relative Position	4

Table 5 presents the parameter values used in the system. The large language model employed was gpt-4o-mini (OpenAI, 2023), with the prompts in Tables 2, 3, and 4 input as the content for the user role. The embedding model used was text-embedding-ada-002 (Neelakantan et al., 2022). The cosine similarity threshold was set to 0.88, based on qualitative observations of the paraphrased sentences. The temporal relative position was set to 4, consistent with previous research (Nakanishi et al., 2024).

To initiate the conversation, the "topic-related utterance" shown in Figure 4 was used, framed as a question about the characters in the illustration: "What do you think, B?".

#### 4.2 Experimental Setup

The experiment involved six participants (five males and one female) with an average age of 23.17 years. Participants engaged in an 8-minute chat session using the system. As shown in Figure 1, participants were instructed to discuss the topic, "Why do you think the man in the hooded sweatshirt and the man with the pipe are talking?" based on the illustration displayed on the left side of the system interface. The discussion task was adapted from Study 3 of Merged Minds (Rossignac-Milon et al., 2021).

The conversation involved three participants: the experiment participant, Agent A, and Agent B. Initially, the participant conversed with Agent A. If the participant made a context-dependent ambiguous statement, the system highlighted the ambiguous utterance in red and the necessary contextual information in yellow. Later, when Agent B was prompted by Agent A, they joined the conversation. Participants were instructed to use the information provided by the system to facilitate the discussion and resolve any misunderstandings that might arise.

#### 4.3 Questionnaire

Table 6: Details of Questionnaire A (Q5 is descriptive; others are rated on a 7-point scale).

No.	Question
Q1	The frequency of assistance (highlighting text) was appropriate.
Q2	The responses from the AI agents (A and B) were natural.
Q3	The response speed of the chat was appropriate.
Q4	You could not understand what B said when they spoke for the first time.
Q5	Why do you think B made that statement when they spoke for the first time? Please explain based on the conversation content.
Q6	The system was useful for the dialogue.

Details of Questionnaire A are shown in Table 6. Except for Q5, responses were rated on a 7-point Likert scale from "Strongly Disagree" (1) to "Strongly Agree" (7), followed by optional free text. Q5 was an open-ended question and, along with Q4 and Q6, was designed to indirectly confirm whether users recognized the possibility that their statements led to misunderstandings by their conversation partner.

Since B missed the SCAINs and made a statement, it was expected that users might respond in Q4 by stating they could not understand what B said. Ideally, in Q5, users would refer to the possibility that B missed the SCAINs, and in Q6, they would feel that SCAINs helped resolve misunderstandings caused by B.

Q1 assessed the appropriateness of the frequency of assistance, such as whether too much highlighting caused discomfort or too little made the system's effect negligible, thus serving as an evaluation of the relevance of extracted SCAINs. Q2 and Q3 were included to evaluate the AI's response quality and the system's latency issues, respectively.

Table 7: Details of Questionnaire B on Conversation History (7-point scale).

No.	Question	Targeted Responses
Q7	Was the amount of highlighting applied to the utterances appropriate?	Utterances where SCAINs were extracted
Q8	Did the utterances contain important information?	Utterances where SCAINs were extracted
Q9	Was the presence or absence of assistance for each utterance appropriate?	All participant utterances

In Questionnaire B (Table 7), participants evaluated whether the assistance provided by the system for each of their utterances was appropriate using a 7-point Likert scale ranging from "Strongly Disagree" (1) to "Strongly Agree" (7). Q7 and Q8 focused on assessing the quantity and quality of SCAINs when they occurred. Q9 evaluated the quality of SCAINs detection by considering cases where assistance was provided for unnecessary sections or absent in sections where it was desired.

It should be noted that these questions reflect the participants' subjective impressions. The value of the system lies in its ability to assist in areas that participants might not otherwise recognize. Therefore, these evaluations are primarily used as a reference to identify specific examples for further analysis.

#### 4.4 Results

Figure 5 visualizes Questionnaire A results in a boxplot with mean values. The horizontal axis represents questions, and the vertical axis shows 7-point Likert scale values.

Q3 received high median and mean values, with responses concentrated in the 6–7 range, indicating consensus and positive evaluations. Conversely, Q4 had low median and mean values near 2, reflecting predominantly negative evaluations and significant variation. Some neutral or slightly positive ratings were observed, but most were low.

For Q1, Q2, and Q6, medians ranged between 4 and 5, indicating neutral to slightly positive evaluations, with notable variation. Q2 showed responses ranging from 3 to 6, highlighting divergence in opinions. Additionally, an outlier in Q1 suggests one extreme rating significantly differed from the others.

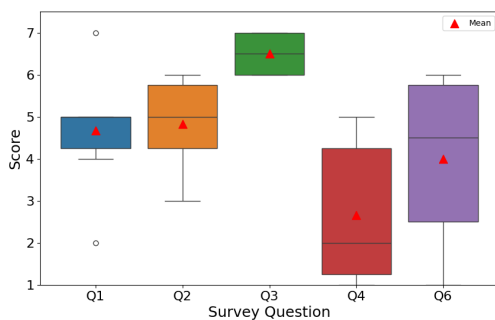


Figure 5: Boxplot of Questionnaire A. Red triangles are average values.

## 5 DISCUSSION

### 5.1 RQ1: Is the Quality of Dialogue and Response Time of the AI Agents Appropriate?

Based on Figure 5, Q3 results indicate that the response speed of the chat was deemed appropriate. The issues of execution time identified in previous SCAINs studies (Maekawa and Imai, 2023) appear to have been resolved through the implementation of parallel processing and caching techniques.

Regarding the naturalness of AI responses (Q2), free-text feedback highlighted inflexibility, such as repeated similar responses. This suggests that improvements are needed for the prompts shown in Tables 3 and 4.

### 5.2 RQ2: Is the System Necessary for Conversations Involving Missed Utterances?

In Q4, Mr. B's statements reflect content prior to the omitted SCAINs, so they are not entirely unrelated to the conversation. Users might have understood the content, and the visible conversation history in this chat format likely reduced the perceived inconsistency. A more accurate approach might have been asking whether Mr. B's statements aligned with the conversational flow.

In Q5, we asked a question to confirm the sense of discomfort of Mr. B's statement, but many participants felt that "he made a statement including a keyword for the time being". In other words, they did not feel that they could not understand the content, but at least they felt some discomfort.

For Q6, opinions on color-coded highlighting were mixed. Some appreciated its clarity, while others found it unnecessary, believing their views were already understood. Misunderstandings about the system's purpose or benefits in real-time dialogues suggest clearer explanations might improve engagement and effectiveness.

### 5.3 RQ3: Is the Frequency of Assistance Provided by the System Appropriate?

For Q1, opinions on assistance frequency varied, with some finding it too infrequent to notice and others deeming it appropriate, depending on the conversation content. The frequency is influenced by the co-

sine similarity threshold, suggesting the need for user-adjustable thresholds for personalization.

Figure 6 presents a heatmap showing the appropriateness of assistance for user utterances across participants. Assistance starts from the fourth utterance, with warm-colored areas indicating appropriateness. Participants 5 and 6 had relatively suitable frequencies, while Participant 1 scored lower for early utterances, highlighting insufficient assistance at the start.

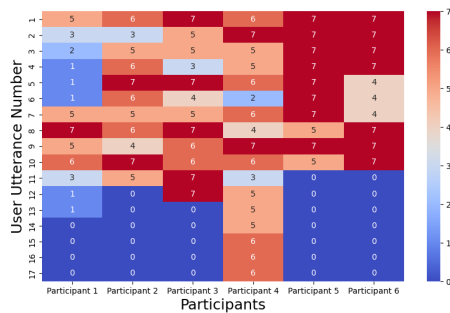


Figure 6: Heatmap of Q9. Adequacy of assisted or unassisted for each user statement (7-point Likert scale) for each participant.

#### 5.4 RQ4: Do the SCAINs Outputs Align with the Users' Intuition?

Specific qualitative evaluations are shown in the Appendix. Agreement with user intuition varied from case to case. Confirming the conversations in the case studies, SCAINs were more likely to occur when short utterances or directives were used. For example, "Let's do that" and "Indeed". While the system was able to detect utterances that could be interpreted differently depending on the context alone, there were differences among users in terms of the effectiveness and frequency of the assistance in resolving misunderstandings.

## 6 CONCLUSIONS

This paper introduced the SCAINs algorithm into a chat system to identify utterances influencing interpretation and highlight potential misunderstandings caused by missing contextual utterances.

Case study results showed SCAINs occurrences for all participants in an AI-driven, multimodal discussion task, with the system providing a seamless experience. Feedback on SCAINs frequency was positive despite case-specific variations. Future work will focus on enhancing prompt and threshold settings and assessing the system's effectiveness in human-to-human interactions.

## ACKNOWLEDGEMENTS

This work was supported by JST CREST Grant Number JPMJCR19A1, Japan.

## REFERENCES

- Feng, X. et al. (2022). A survey on dialogue summarization: Recent advances and new frontiers. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 5453–5460. International Joint Conferences on Artificial Intelligence Organization.
- Kishinami, Y. et al. (2023). Topic transition modelling in human-to-human conversations. In *Proceedings of the 29th Annual Meeting of the Association for Natural Language Processing*, pages 408–413.
- Konigari, R. et al. (2021). Topic shift detection for mixed initiative response. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 161–166. Association for Computational Linguistics.
- Maekawa, T. and Imai, M. (2023). Identifying statements crucial for awareness of interpretive nonsense to prevent communication breakdowns. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12550–12566. Association for Computational Linguistics.
- Nakanishi, K. et al. (2024). Extending contextual understanding techniques to account for missed listenings. In *Proceedings of the 38th Annual Conference of JSAL*.
- Neelakantan, A. et al. (2022). Text and code embeddings by contrastive pre-training. *ArXiv*, abs/2201.10005.
- Nihei, F. and Nakano, Y. (2020). Evaluating a multimodal meeting summary browser that equipped an important utterance detection model based on multimodal information. In *Proceedings of the 34th Annual Conference of JSAL*.
- OpenAI (2023). Gpt-4 technical report. *ArXiv*, abs/2303.08774.
- Rossignac-Milon, M. et al. (2021). Merged minds: generalized shared reality in dyadic relationships. *Journal of Personality and Social Psychology*, 120:882–911.
- Sugiyama, H. et al. (2023). Empirical analysis of training strategies of transformer-based japanese chat systems. In *2022 IEEE Spoken Language Technology Workshop (SLT)*, pages 685–691.
- Sumita, K. et al. (1988). Information of the interpretation for context understanding. In *Proceedings of the 37th National Convention of IPSJ*, pages 1129–1130.
- Tsuchiya, A. et al. (2024). Scains presenter: Preventing miscommunication by detecting context-dependent utterances in spoken dialogue. In *Proceedings of the 29th International Conference on Intelligent User Interfaces, IUI '24*, page 549–565. Association for Computing Machinery.

## APPENDIX

This is a qualitative evaluation based on actual cases. Examples that match the user's intuition are shown in Table 8, and examples that do not match are shown in Table 9.

Table 8: Example of a Dialogue Where Both the Quality and Quantity of Assistance Were Appropriate. The "No." corresponds to the vertical axis values in Figure 6. (Participant 2). The red line represents the core statement, while the yellow lines indicate the SCAINs.

No.	Speaker	Utterance
	A	Why do you think those two are talking?
	User	Maybe they're making some shady deal?
	A	That might be true. Their behavior is suspicious.
1	User	Doesn't the outfit of the guy in the front stand out?
	A	True, that outfit is quite conspicuous. He might be hiding something.
2	User	Maybe in his sleeves?
	A	Something hidden in his sleeves.? That's intriguing.
3	User	Could it involve those two as well?
	A	Yeah, those two might be involved in something.
4	User	Feels like a drama.
	A	It's like a scene from a suspense movie. Makes you curious!
5	User	Totally!
	A	What do you think, B?
	B	Hmm, that outfit is definitely unusual. There might be a special reason for it.

Table 8 presents an example where the results for Q7, Q8, and Q9 were all rated 7, indicating that SCAINs were appropriately extracted when desired, and both the quality and quantity were sufficient. The core utterance, No.5, included all five preceding statements as SCAINs.

The user's response that the quantity of assistance was appropriate suggests that they considered two aspects, "what might be hidden in the sleeves" and "what's happening," as critical points of interest. The utterance "Totally!" is difficult to interpret on its own, making it unclear what the user is referring to, which could hinder B from engaging in the conversation. This led B to misinterpret "Totally!" and shift the discussion back to clothing, disregarding the SCAINs.

Table 9 shows an example of Participant 1's conversation where the assistance was inappropriate due to the absence of detected SCAINs, despite the likely

Table 9: Example of a conversation with Participant 1 who was assisted inappropriately, where No. corresponds to the vertical axis value in Figure 6.

No.	Speaker	Utterance
	A	Why do you think those two are talking?
	User	Maybe they just want to enjoy a conversation at the café.
	A	That's possible. They might want to relax and have a chat in a cozy atmosphere.
1	User	What do you think, A?
	A	Perhaps they gathered to discuss something important.
2	User	What kind of important matter?
	A	For example, they might be discussing a work project.
3	User	That's true. With their detective-like outfits, maybe they're talking about a case.
	A	If so, they might be looking for some clues.
4	User	What do you think the person in the hoodie does for a living?
	A	The person in the hoodie might be a detective or a journalist.
5	User	What other topics do you think they might be discussing?

need for them. For instance, in No. 4, the user's inquiry about "the person in the hoodie" references the image without prior mentions in the history, possibly explaining the lack of SCAINs extraction. However, earlier mentions of a "detective-like outfit" could have been extracted to provide context about professions, reducing the risk of misunderstanding. Notably, a SCAINs candidate with a cosine similarity score of 0.89 existed, emphasizing the importance of carefully setting thresholds.