

Constructing a Culinary Interview Dialogue Corpus with Video Conferencing Tool

Taro Okahisa, Ribeka Tanaka*, Takashi Kodama, Yin-Jou Huang, Sadao Kurohashi

Kyoto University

Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan

{okahisa, kodama, huang, kuro}@nlp.ist.i.kyoto-u.ac.jp

Abstract

Interview is an efficient way to elicit knowledge from experts of different domains. In this paper, we introduce CIDC, an interview dialogue corpus in the culinary domain in which interviewers play an active role to elicit culinary knowledge from the cooking expert. The corpus consists of 308 interview dialogues (each about 13 minutes in length), which add up to a total of 64,000 utterances. We use a video conferencing tool for data collection, which allows us to obtain the facial expressions of the interlocutors as well as the screen-sharing contents. To understand the impact of the interlocutors' skill level, we divide the experts into "professionals" and "enthusiasts" and the interviewers into "skilled interviewers" and "unskilled interviewers." For quantitative analysis, we report the statistics and the results of the post-interview questionnaire. We also conduct qualitative analysis on the collected interview dialogues and summarize the salient patterns of how interviewers elicit knowledge from the experts. The corpus serves the purpose to facilitate future research on the knowledge elicitation mechanism in interview dialogues.

Keywords: Interview Dialogue, Corpus Construction, Multi-modal Information, Knowledge Elicitation

1. Introduction

An interview is a special form of dialogue that occurs between an interviewer and an interviewee. Typically, the interviewer asks the interviewee some questions to elicit information of interest from the interviewee.

We are especially interested in the interview dialogues that aim to elicit technical knowledge from domain experts. Here, domain experts refer to people who possess technical knowledge about how to carry out specific tasks in a certain domain. Interview is an efficient way to elicit technical knowledge from such experts. Under the deliberate questions of the interviewers, the experts are prompted to reflect on and verbalize the technical knowledge that is critical to complete the task. The elicitation of technical knowledge can help the transmission of skills from domain experts to less-experienced workers, which is beneficial to many industrial fields.

Most of the existing interview dialogue corpora are based on news interviews which serve the purpose of information broadcasting or entertainment (Majumder et al., 2020; Sasayama and Matsumoto, 2020; Zhu et al., 2021). However, the interview dialogue data in which the interviewer plays an active role to guide the overall dialogue flow and elicit technical knowledge is still lacking.

To facilitate research on interview dialogues and technical knowledge elicitation, we construct an interview dialogue corpus in the culinary domain. Each interview dialogue takes place between two participants, who are given the role of interviewer and expert, respectively. The interviewer is asked to elicit cooking instructions

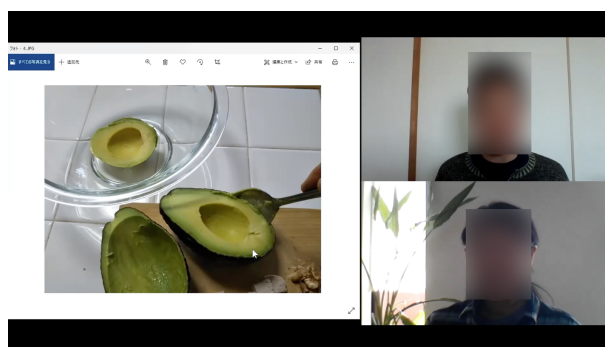
for a specific dish from a culinary expert in an online video dialogue. On the other hand, the expert is asked to verbalize the technical knowledge under the query of the interviewer.

Some existing corpora of spoken dialogues are based on telephone conversations (Den and Fry, 2000; Zhou et al., 2010). However, telephone conversations are very different in nature compared to normal dialogues. For example, the lack of visual context hinders to usage of demonstratives (e.g. *this*, *that*). Other spoken dialogues corpora consist of the recordings of face-to-face conversations (van Son et al., 2008; Brône and Oben, 2015; Koiso et al., 2018). However, recording actual face-to-face dialogues in a studio is very time-consuming and costly.

To overcome the above shortcomings, we use video conferencing tools to collect our interview dialogue corpus. Following the outbreak of the global pandemic COVID-19, more and more people are using video conferencing tools to communicate with people in different physical locations. These tools not only allow the interlocutors to see each other's facial expressions in real time, but they can also utilize the screen-sharing feature to share visual context with each other. Utilizing video conferencing tools, we can collect spoken dialogues that are close in nature to face-to-face dialogues at a lower cost.

In this paper, we propose the Culinary Interview Dialogue Corpus (CIDC), which contains a total of 308 interview dialogues in Japanese. Utilizing the video conferencing tool, we collected approximately 64,000 utterances along with the corresponding video contents including the interlocutors' facial expressions and the visual context acquired through screen-sharing. See Figure 1 for example.

*Currently affiliated with Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo, 112-8610, Japan (E-mail: tanaka.ribeka@is.ocha.ac.jp).



Speaker	Start	End	Utterance
E	04:35.6	04:38.9	(で) 今回アボカドの皮を器にいたしますので、 As the avocado peel will be used as a bowl,
E	04:39.4	04:42.8	中の身をちょっとスプーンでくりぬいていきます。 hollow out the flesh with a spoon.
E	04:44.5	04:49.4	はい。(で) 皮がちょっと破れないように気をつけて、(えー) くりぬいていただい て、 Be careful not to break the peel when hollowing
E	04:50.0	05:01.8	(んで) (えーと) (ま) 身、結構ギリギリまで今回取ってるんですけど も、(ま) (いち) 1センチぐらいの厚さとか、(あの一) 深めに (えー) 残して いただいても全然大丈夫です。 This time, we took almost all the flesh, but it's also okay to leave a margin about 1cm.
I	05:02.0	05:07.4	そうですね。あんまりギリギリまでやろうとすると、私もよく皮を破ってしまうこ とあります。 I see. I often break the peel when trying to hollow it completely.
E	05:07.3	05:11.8	(あ) そうですね。(?) やっぱ器にするので気を付けていただいて。はい。 Yes. Please be careful because the peel will be used as a bowl for salad

Figure 1: An example of video image and transcription. The alphabets “E” and “I” in the first column stand for expert and interviewer, respectively.

The rest of the paper is structured as follows. In Section 2, we introduce the related works. In Section 3, we describe how to collect the interview dialogues. In Section 4, we describe the transcription method and statistics of the collected corpus. In Section 5, we conduct qualitative analysis on the collected corpus and include several dialogue examples to illustrate the characteristics of the knowledge-eliciting strategies. Finally, we summarize the whole paper and discuss the future prospects in Section 6.

2. Related Work

The collection of interview dialogues has attracted much attention in the past few years. Ishihara et al. (2018) collect dialogues between humans and robots for estimating the speaker’s willingness. Recently, INTERVIEW (Majumder et al., 2020) and MEDIASUM (Zhu et al., 2021) have been proposed, both of which are based on news interview transcripts. To understand the structure of interview dialogues, multimodal information (e.g., utterance audio data, participant’s facial expressions, hand gestures, visual contexts, etc.) as well as textual data are considered to be necessary. Most of the existing interview dialogue corpora do not contain such multimodal information, while our new corpus includes such multimodal infor-

mation, allowing for a deeper analysis of the interview structure.

The research of multimodal dialogue collection has also been well studied. In particular, there is a lot of research aimed at language learning (van Son et al., 2008; Saeki et al., 2021) instead of technical knowledge elicitation. On the other hand, DAIC (Gratch et al., 2014) is a large multimodal corpus of clinical interviews including the following settings: (i) Face-to-face, (ii) teleconference, (iii) Wizard-of-Oz, (iv) Automated. However, the interviewers play the role of a counselor, who does not actively elicit information from the interviewee, because DAIC is designed to contribute to the diagnosis of psychological distress conditions.

3. Method of Interview Dialogue Collection

3.1. Interview Dialogue Setting

In this work, we construct an interview dialogue corpus to facilitate the understanding of the knowledge-eliciting mechanism during an interview. We focus on the culinary domain and collect interview dialogues in which the interlocutors are asked to discuss the cooking procedures of specific recipes.

We adopt the dyadic interview setting in which two participants engage in each dialogue. The two participants

1	Ask numerical questions such as quantity of ingredients, how long a cooking procedure takes, etc.
2	Ask the expert to elaborate on subjective descriptions such as ‘thickening’, ‘browning.’, etc.
3	Ask about the possible ingredient and equipment substitutions.
4	Ask about preliminary preparation steps since they are often omitted.
5	If you notice some details in the picture that is not mentioned by the expert, ask for further explanation.
6	Once you have elicited an explanation from the expert, take it one step further and see if it’s a good thing.
7	Ask for the tips directly.
8	Ask about what should be avoided during the cooking procedure.
9	Express your concern and critics about the recipe.
10	Ask hypothetical questions based on your own culinary knowledge.
11	Instead of yes-no questions, try to ask “why” or “how” about the cooking procedure.

Table 1: Interviewer’s manual.

are assigned the role of **expert** and **interviewer**, respectively.

Expert The expert introduces the recipe of a specific dish to the interviewer. Whether spontaneously or prompted by the interviewer’s queries, the expert should try to verbalize the technical knowledge about how to make the dish.

Interviewer The interviewer actively elicits technical knowledge from the expert by asking questions based on the expert’s utterances.

We recruited Japanese native speakers to conduct the interview collection tasks.

3.2. Preparations and Instructions before Interview

In order to help the interviewers elicit culinary knowledge from the expert effectively, we ask the participants to do some preparation work in advance (Section 3.2.1) and give them some instructions regarding the interview process (Section 3.2.2).

3.2.1. Preparation in Advance

We ask the participants to do some preparation in advance so that the participants can be familiar with the contents that will be covered during the interview.

First, the expert is asked to submit the following information about the dish they want to talk about:

- The title of the dish
- A brief summary of the recipe of the dish (around 50 characters)
- 6–10 pictures of the cooking process
- Takeaway points (e.g. tips, special features) of the recipe and the corresponding photo number

The interviewer is given the title, the summary, and the pictures of the dish submitted by the expert. In addition, the interviewers are advised to read the “Interviewer’s manual” we provided. The manual consists of 11 classic patterns of interview questions in the culinary domain that we manually collected from a preliminary experiment (Table 1). Based on the above information, the interviewer is asked to prepare one or more

questions and when to ask these questions during the interview.

3.2.2. Instructions Regarding the Interview Process

Instructions are given to the participants regarding how to act in the interview process.

The following instructions are given to the experts:

- Display the pictures of the cooking procedures in order.
- Show each picture to the interviewer and introduce the corresponding cooking procedure (e.g. “First, cut the vegetables” or ‘Add salt and pepper’). After that, expect the interviewer’s questions about the cooking process.
- Wait for the interviewer’s prompt to proceed to the next picture. However, the expert can add extra explanations about the important points not elicited by the interviewer’s questions.
- The expert is advised to use the mouse pointer to indicate the reference object in the picture, when that given object is mentioned in the dialogue. (The mouse pointer should be enlarged beforehand so that it can be easily seen)

On the other hand, the following instructions were given to the interviewers:

- Review the title, summary, and pictures submitted by the experts before the interview begins, and go over the content and timing of the prepared questions.
- Act as the dialogue facilitator. Actively ask questions to elicit further explanations, since the expert will only give a minimum explanation of each cooking procedure. Also, the interviewer should prompt the interviewer to move on to the next picture.

	Experts	Interviewers	Total
# of words	370,771	218,751	589,522
# of utterances	35,478	28,677	64,155
Avg. # of utterances per dialogue	115.2	93.1	208.3
Avg. # of words per utterance	10.5	7.6	9.2
Avg. duration of utterance (sec.)	3.9	2.9	3.5
Avg. duration of dialogue video (min.)	—	—	12.6

Table 2: Statistics of the entire interview dialogue collected. Fillers are excluded.

3.3. Recording Method

We use the video conferencing tool Zoom¹ to carry out the interviews. The participants are asked to use external or built-in cameras and microphones. Also, the experts use the screen-sharing function to show pictures of the cooking process. We used the built-in recording function of Zoom to collect the audio and video contents of the interview dialogues. The audio contents of both speakers were separately saved in wav files. The video contents were recorded in mp4 files.

Before the dialogue begins, each interlocutor is given 5 minutes to check the operation status of the devices and rehearse the interview. We ask the interlocutors to complete the interview within 15 minutes. Shortly before the 15-minute time limit, the interlocutors are prompted to end the conversation as soon as possible.

3.4. Transcription of Dialogue Data

We use AmiVoice² to transcribe the dialogues automatically. Furthermore, the transcriptions are manually revised by professional transcribers (Advanced Media, Inc.), and the start/end time of each utterance is also annotated. The transcribers were the professionals in an annotation company. The unit of transcription is based on pauses of 500ms or longer, and if there are pauses between utterances of the same speaker, they are transcribed as different units. The purpose of the transcription method is to provide data for examining methods of eliciting knowledge from experts. For this reason, we focus on the content of the interlocutor’s speech itself, and kept the transcription to a minimum with regard to backchannel, laughter, and errors. Specifically, the following policies are adopted, and when issues arose, the transcription was carried out after consultation on a case-by-case basis:

- Fillers and restatements are enclosed in parentheses.
- Do not transcribe backchannel (“um”, “hmm”, etc.) while one interlocutor is speaking. However, if the speech is clearly audible, such as “I see” or “Okay,” it is transcribed.
- For laughter, it is transcribed as [laughter]. However, for laughter during speech, do not leave any information about it.

- If the speech was inaudible, it is transcribed as “(?)”
- If the utterance is longer than 10 seconds, punctuation can be added.
- If the speaker made a slip of tongue, transcribe it as the correct form.

Since it is difficult to distinguish between fillers (e.g. “(んで)(えーと)(ま)身、結構ギリギリまで今回取ってるんですけど”) (well) *this time, we took almost all the flesh*) and restatements (e.g. “(いち)1センチぐらいの厚さとか”) (*a margin about (o-) one cm*), we have adopted the policy of bracketing them with the same symbol parentheses in the transcription stage. Therefore, the following process was applied after the corpus was constructed to distinguish between fillers and other restatements:

- Throughout the corpus, items that occur less than 10 times and do not contain a double vowel are considered to be restatements and are bracketed with < >.
- All others are considered fillers and are bracketed by ().

3.5. Post-Interview Questionnaire

In order to check the quality of interviews, the participants are asked to answer the following post-interview questionnaire:

- Was the conversation going smoothly during the interview? (5-Likert scale)
- Was the interviewer able to elicit culinary knowledge from the expert skillfully? (5-Likert scale)
- Describe the overall impressions of the interview. (open question)
- How knowledgeable/skillful are you in the culinary domain? (5-Likert scale; only for the interviewers)

As for the questions using a 5-Likert scale, the participants choose among the following options: “5. Agree” “4. Somewhat agree,” “3. Neutral,” “2. Somewhat disagree” and “1. Disagree.”

¹<https://zoom.us>

²<https://www.advanced-media.co.jp>

4. Statistics of CIDC

4.1. Statistics of the Interview Dialogues

The corpus is constructed from 308 dialogues, and the statistics for the entire corpus are shown in Table 2³.

While the average number of words per utterance is 10.5 for the experts and 7.6 for the interviewers, the average number of utterances per dialogue is 115.2 for the experts and 93.1 for the interviewers, indicating that the experts deliver longer utterances than the interviewers. This trend indicates that the dialogue participants appropriately fulfilled the roles of the expert and interviewer.

4.2. Statistics of the Participants

In the dialogue tasks, each interlocutor was assigned the role of an expert or an interviewer. Among the 26 participants, 17 people participated as experts and 17 people participated as interviewers (8 people participated in both roles). See Table 3 for details.

For both experts and interviewers, we further divide them into subgroups by their skill levels. We divide the experts into a group of “professionals” and a group of “enthusiasts”, based on their skill levels in the culinary domain. The experts who match at least one of the following descriptions are classified as “professionals”: (i) have experience as a teacher in any type of cooking class, (ii) have a chef license, or (iii) are currently a professional cook. Enthusiasts are people who loves cooking but do not match the above conditions. On the other hand, the interviewers are classified based on their skill level to conduct an interview. The group of “skilled interviewers” consists of people who have experience in conducting any form of an interview (such as a job interview), while the others are classified as “unskilled interviewers”⁴.

For dialogue data collection, 20 dialogues are recorded as preliminary collection with tentative condition settings at first. Based on the analysis of the preliminary collection, the final conditions were set and the main collection was collected. In the following description, conditions common to both the preliminary and main collections are described without further explanatory notes.

We conduct a quantitative analysis on combinations of interview participants of different skill levels. The results are summarized in Table 4⁵.

We can observe that professionals (E_H) have a lower average number of utterances per dialogue compared to the enthusiasts (E_L). Also, we can observe that the

³After removing fillers and restatements, morphological analysis was performed using Juman++ (Tolmachev et al., 2018).

⁴The requirement for participating as an interviewer is to have a minimum level of culinary knowledge (e.g. having experiences of cooking).

⁵For experts, E_H stands for professionals and E_L for enthusiasts. For interviewers, I_H stands for skilled interviewer and I_L for unskilled interviewer.

I \ E	E		
	E_H	E_L	\times
I_H	2	1	5
I_L	3	2	4
\times	3	6	-

Table 3: Details of task participants. The symbol “ \times ” means “not participating as an expert/interviewer.”

average length per utterance of the professionals are longer than that of the enthusiasts. We speculate that the professionals have the ability to utter longer and more complex utterances compared to the enthusiasts, regardless of the skill level of the interviewer.

On the other hand, there is no significant difference in the average number of utterances per dialogue regarding the skill levels of the interviewers. However, we observe a larger number of interviewer’s utterances per dialogue when the interlocutors are “enthusiasts” ($E_L \& I_H$, $E_L \& I_L$), compared to the case when the interlocutors are “professionals” ($E_H \& I_H$, $E_H \& I_L$). Also, we observe that “skilled interviewers” (I_H) have a larger number of words per utterance compared to that of “unskilled interviewers” (I_L). This suggests the following two tendencies of the interviewers: (1) When the interviewee has a lower level of cooking skill, the interviewer is required to say more to elicit knowledge from the expert, regardless of his or her own interview skill, and (2) Interviewers with better interview skills can conjure up utterance with richer contents.

4.3. Results of the Post-Interview Questionnaire

The results of the post-interview questionnaire are shown in Figure 2 and Figure 3. For the question “Was the conversation going smoothly?”, the average scores are 4.64 and 4.25 for the the experts and the interviewers, respectively. Also, more than 85% of both the experts and interviewers chose “5. Agree” or “4. Somewhat agree.” These results indicate that both roles felt that they were able to conduct a smooth conversation in the task.

As for the question “Was the interviewer able to elicit culinary knowledge skillfully?”, the mean value for the experts is 4.62 and the mean value for the interviewers is 4.07. About 80% of the experts chose “5. Agree,” while only about 30% of the interviewers did. This difference is probably caused by the fact that the interviewers were instructed in advance to elicit culinary knowledge from the experts, and therefore, they judged their own performances in the interview more strictly than the experts.

The mean value for the interviewer’s culinary knowledge is 3.52. This result reflects the fact that the interviewers were required to have a minimum level of culinary knowledge as a condition for participation.

	$E_H&I_H$		$E_H&I_L$		$E_L&I_H$		$E_L&I_L$	
	E_H	I_H	E_H	I_L	E_L	I_H	E_L	I_L
Avg. # of utterances per dialogue	104.3	83.8	110.2	86.3	124.6	101.5	122.1	101.6
Avg. # of words per utterance	12.2	8.6	11.9	6.9	8.9	8.0	9.0	7.1

Table 4: Statistics for each dialogue setting. Fillers are excluded. For the notation, see the footnote 5.

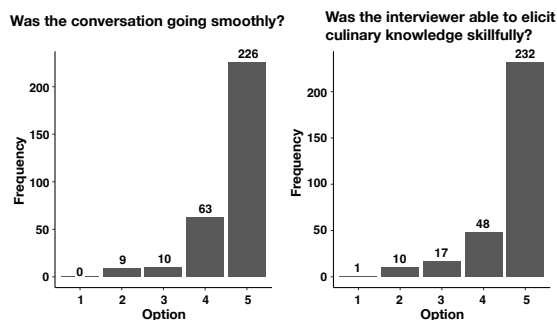


Figure 2: Results of questionnaire to the experts.

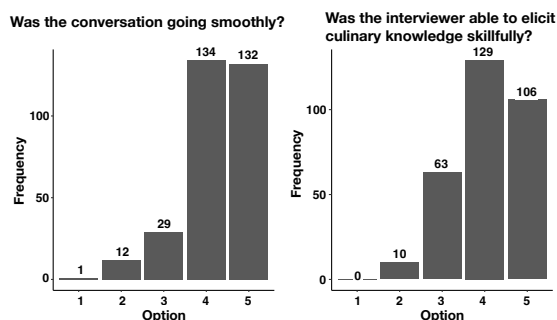


Figure 3: Results of questionnaire to the interviewers.

5. Qualitative Analysis

5.1. Skill levels of Interlocutors and Characteristics of the Interviews

As mentioned earlier, we categorize the experts into “professionals” and “enthusiasts” and the interviewers into “skilled interviewers” and “unskilled interviewers.” In the following, we conduct qualitative analysis to understand the characteristics of the interview participants of different skill levels, as well as the characteristics of interview dialogue under different settings (i.e. $E_H&I_H$, $E_H&I_L$, $E_L&I_H$, $E_L&I_L$).

We first analyze the characteristics of experts of different culinary skill levels. When asked about detailed information such as the quantities of ingredients, the professionals were able to answer smoothly while looking at the camera. On the other hand, the enthusiasts often were at a loss for an answer and referred to the recipes in such a situation. In addition, when asked about the purpose and intentions of a particular cooking procedure, the professionals are often more assertive, while the enthusiasts used the form of euphemism (e.g. *I*

heard that ..., I think that ..., it is said that ...).

We also analyze the characteristics of interviewers of different interview skill levels. When it comes to back-channeling, the skilled interviewers tend to use a combination of verbal and body languages to show their understanding, while the unskilled interviewers often just nod imperceptibly. In addition, the skilled interviewers actively used gestures when asking questions. In addition, the unskilled interviewers often need to look at the prepared questions when speaking, while the skilled interviewers look at the camera to show their attention to the experts.

Content-wise, the unskilled interviewers often respond to the experts’ answers with “I see” or “I understand”, showing their difficulties to dig into the experts’ answers. On the other hand, the skilled interviewers often integrate their own cooking experiences with the experts’ answers and ask further follow-up questions. Next, we analyze the characteristics of interview dialogues under different dialogue settings. In the setting of “ $E_H&I_H$,” both participants looked at the camera and proceeded with the dialogue at an appropriate pace. In the setting of “ $E_L&I_L$,” the pace of the conversation was sometimes too fast or too slow. In addition, in the setting of “ $E_H&I_L$,” the interviewers sometimes struggle to understand the experts’ intentions or act confused when the experts give an answer that he/she did not expect. On the other hand, in the setting of “ $E_L&I_H$,” while the conversation seems to be going smoothly, the interviewers often have to make more effort to elicit answers from the experts.

5.2. Characteristics of a Knowledge-eliciting Interview

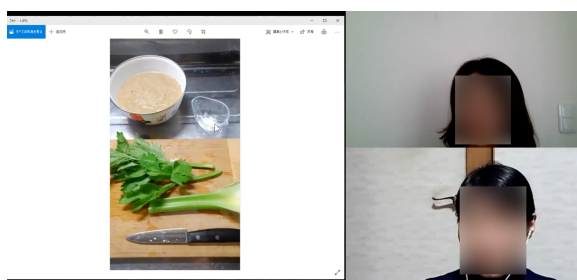
During the dialogue collection, the interviewers are given the instruction to actively elicit culinary knowledge from the experts. Thus, there are situations in which the interviewer’s question prompted the experts to verbalize the implicit technical knowledge that he/she was not initially aware of. In the following, we analyze the characteristics of the interviewer’s utterances and summarize two salient patterns of how interviewers elicit implicit technical knowledge from the experts.

Paraphrasing and Repetition One of the characteristic patterns of interviewers’ utterances when eliciting implicit technical knowledge is to repeat or paraphrase what the expert had said. See the following example:

- (1) “Marinated celery” (Figure 4)



(a) The picture of the dish.



(b) The screenshot of the dialogue example (1).

Figure 4: Marinated celery.

- 16 E: This is coarse salt, unlike the smooth table salt, it is quite rough and lumpy, with an apparent sweetness.
- 17 I: Coarse salt.
- 18 I: does it have a strong flavor?
- 19 E: Yes. The salt tastes good even if you eat it alone. **You can use any kind of salt you like, but since there are only a few ingredients, it is better to use flavored salt.**

In the dialogue example (1), the expert described the coarse salt as having “apparent sweetness” (utterance 16), while the interviewer rephrased it as “it have a strong flavor” (utterance 18). Stimulated by the interviewer’s utterance, the expert was able to verbalize the implicit technical knowledge (highlighted in bold font, utterance 19) regarding the purpose of using coarse salt.

The technique of paraphrasing and repetition is a simple but effective way of knowledge elicitation. The interviewer can ask this type of questions even if he/she doesn’t not have much experience in cooking. It has been pointed out that hearer’s repetition of the speaker’s utterance often functions as a trigger to repair conversational troubles (Schegloff et al., 1977; Schegloff, 1997).

Situation-oriented questions based on interviewer’s own experience On the other hand, there are some characteristics that can only be observed from the knowledge elicitation process of interviewers with sufficient cooking experience. Based on the interviewer’s own experience in cooking, he/she can describe a specific situation during the cooking procedure and elicit a more in-depth explanation from the expert. See the following example:

- (2) “Whitebait tempura” (Figure 5)

- 177 I: Do you have any tips for frying?
- 178 E: (Um)
- 179 E: Tips for frying.
- 180 E: (Well)
- 181 E: Frying...

- 182 I: For example, I am often worried and feel the urge to check whether the ingredients are ready,
- 183 I: I end up stirring it with chopsticks to check
- 184 E: (?)
- 185 E: It’s better not to touch it.
- 186 I: (Oh) I see.
- 187 E: (?)
- 188 E: **After frying one side or it, let it cook slowly without touching at all. After that, flip it over and deep-fry it without touching, so that it does not lose its shape. At the end of the frying, drain off all the oil when you lift it onto the bat,**
- 189 E: **this is the key to have it crispy and fluffy.**
- 190 E: **this is the key to have it crispy and fluffy.**

In the dialogue example (2), in the utterance 177, the interviewer vaguely asked, “Do you have any tips for frying?” In response to the question, the expert was at a loss for an answer for about 6 seconds (utterance 178 to 181). In utterance 182, the interviewer starts describing her own experience of frying, which successfully elicit the knowledge from the expert about frying. The interviewer’s experience in touching with chopsticks not only prompted the experts to answer “Try not to touch it” (utterance 185), but it also leads to a series of detailed explanation such as “let it cook slowly,” “flip it over” (the utterance 188), “drain off all the oil when you lift it onto the bat” (the utterance 189).

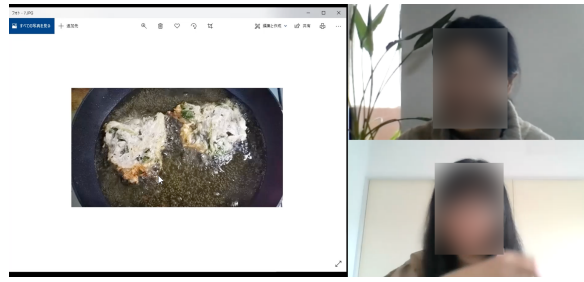
As can be observed from the above example, the interviewer’s reference to a specific situation based on his/her own experience can prompt the expert to recall more detailed procedures and add explanations for implicit points. In this way, interviewers can use his or her own knowledge to set up a specific situation, allowing experts to speak smoothly about their knowledge and skills that they are not aware of beforehand.

6. Conclusion

We introduced the construction of CIDC, the culinary interview dialogue corpus which aims to facilitate the



(a) The picture of dish.



(b) The screenshot of the dialogue example (2).

Figure 5: Whitebait tempura.

research of knowledge elicitation from domain experts. We collected dialogues using the video conferencing tool Zoom, which allowed us to record not only the participant’s speech but also their facial expressions and the visual contexts shared between the interlocutors through the screen-sharing feature.

We reported the basic statistics of the corpus and the quantitative analysis results regarding the impact of interlocutors’ skill level in cooking/interview. Also, the qualitative analysis of the collected data reveals the following patterns of how interviewers elicit knowledge from the experts: (1) paraphrasing and repetition and (2) situation-oriented questions based on the interviewer’s own experience. The CIDC will be publicly available in near future. We believe that the corpus can contribute to the future research of interview dialogue systems and the elicitation of domain knowledge.

7. Acknowledgements

This paper is based on results obtained from a project, JPNP20006, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

8. Bibliographical References

- Brône, G. and Oben, B. (2015). Insight interaction: a multimodal and multifocal dialogue corpus. *Language resources and evaluation*, 49(1):195–214.
- Den, Y. and Fry, J. (2000). Callhome japanese corpus. *Journal of the Phonetic Society (in Japanese)*, 4(2):24–30.
- Gratch, J., Artstein, R., Lucas, G., Stratou, G., Scherer, S., Nazarian, A., Wood, R., Boberg, J., DeVault, D., Marsella, S., Traum, D., Rizzo, S., and Morency, L.-P. (2014). The distress analysis interview corpus of human and computer interviews. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, pages 3123–3128, Reykjavik, Iceland, May. European Language Resources Association (ELRA).
- Ishihara, T., Nitta, K., Nagasawa, F., and Okada, S. (2018). Estimating interviewee’s willingness in multimodal human robot interview interaction. In *Proceedings of the 20th International Conference on Multimodal Interaction: Adjunct*, ICMI ’18, New York, NY, USA. Association for Computing Machinery.
- Koiso, H., Den, Y., Iseki, Y., Kashino, W., Kawabata, Y., Nishikawa, K., Tanaka, Y., and Usuda, Y. (2018). Construction of the corpus of everyday Japanese conversation: An interim report. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC’18)*, pages 4259–4264, Miyazaki, Japan, May. European Language Resources Association (ELRA).
- Majumder, B. P., Li, S., Ni, J., and McAuley, J. (2020). Interview: Large-scale modeling of media dialog with discourse patterns and knowledge grounding. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8129–8141, Online, November. Association for Computational Linguistics.
- Saeki, M., Matsuyama, Y., Kobashikawa, S., Ogawa, T., and Kobayashi, T. (2021). Analysis of multimodal features for speaking proficiency scoring in an interview dialogue. In *2021 IEEE Spoken Language Technology Workshop (SLT)*, pages 629–635.
- Sasayama, M. and Matsumoto, K. (2020). Construction of interview dialogue corpus and design of dialogue intention tags. In *Proceedings of The 82th International Convention of IPSJ (in Japanese)*, volume 1, pages 51–52, Feb.
- Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382.
- Schegloff, E. A. (1997). Practices and actions: Boundary cases of other-initiated repair. *Discourse processes*, 23(3):499–545.
- van Son, R., Wesseling, W., Sanders, E., and van den Heuvel, H. (2008). The IFADV corpus: a free dialog video corpus. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC’08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA).
- Zhou, K., Li, A., Yin, Z., and Zong, C. (2010). CASIA-CASSIL: a Chinese telephone conversation corpus in real scenarios with multi-levelled annota-

tion. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May. European Language Resources Association (ELRA).

Zhu, C., Liu, Y., Mei, J., and Zeng, M. (2021). Mediasum: A large-scale media interview dataset for dialogue summarization. *CoRR*, abs/2103.06410.

9. Language Resource References

Tolmachev, Arseny and Kawahara, Daisuke and Kurohashi, Sadao. (2018). *Juman++: A Morphological Analysis Toolkit for Scriptio Continua*. Association for Computational Linguistics.