

コツを引き出す対話設定における オンライン料理インタビュー対話コーパスの構築

岡久 太郎^{†,††}・田中リベカ^{†,†††}・児玉 貴志[†]・Yin Jou Huang[†]・
村脇 有吾[†]・黒橋 禎夫[†]

インタビューは技能者からコツを引き出すための重要な対話形式の1つである。本研究では、料理ドメインにおける技能者からインタビュアーが料理のコツを積極的に引き出そうとしているインタビュー対話を収集したコーパス (CIDC) を構築した。CIDC は、308 のインタビュー対話 (1 対話あたり約 13 分)、約 6 万 4 千発話から構成される。対話収集には、ウェブ会議システムを用い、参加者の表情と共有されている料理工程を示す写真を発話音声とともに収録した。また、全ての発話を音声認識によって書き起こし、人手で修正した。なお、技能者とインタビュアーのそれぞれにおいて上級と一般の 2 つの熟達度の参加者を集めた。CIDC を活用することで、今後、インタビュー対話におけるコツの引き出し方に関する研究が進展することが予想される。

キーワード：インタビュー対話、コーパス構築、マルチモーダル情報

Domain Knowledge Elicitation in a Corpus of Online Interview Dialogues on Culinary Arts

TARO OKAHISA^{†,††}, RIBEKA TANAKA^{†,†††}, TAKASHI KODAMA[†], YIN JOU HUANG[†],
YUGO MURAWAKI[†] and SADAO KUHASHI[†]

Interviews are an efficient way to elicit knowledge from experts in different domains. In this paper, we introduce a corpus of interview dialogues on subjects in culinary arts called CIDC. In collecting this data, interviewers played an active role in eliciting culinary knowledge from cooking experts. The corpus consists of 308 interview dialogues, each about 13 minutes long, comprising a total of 64,000 utterances. We used a videoconferencing tool to collect the data, which included the content that the participants shared on their screens. This approach allowed us to classify the facial expressions of the interlocutors from the video data. We also categorized the experts as either professionals or enthusiasts and the interviewers as skilled or unskilled. We expect this corpus to facilitate future research on knowledge elicitation mechanisms in interview dialogues.

[†] 京都大学, Kyoto University

^{††} 現在, 静岡大学, Currently, Shizuoka University

^{†††} 現在, お茶の水女子大学, Currently, Ochanomizu University

Key Words: *Interview Dialogue, Corpus Construction, Multi-modal Information*

1 はじめに

対話において、話し手の発話に対して聞き手が質問や確認を行うことで、モノログにおいては表出しづらい情報を引き出すことが可能である。本研究では、対話のこのような機能に着目し、特定分野の技能者からその技能者が持つコツをインタビューによって引き出すという設定を考え、そうしたインタビュー対話のコーパス構築に取り組む。なお、本稿ではコツを以下のように定義する。

コツ 自発的には言語化しづらい、特定ドメインに関する深い知識、感覚

管見の限り、インタビュー対話によって技能者からコツを引き出すという目的で構築されたコーパスは存在しない。このようなコーパスの構築は、近年の産業界において課題となっている熟練者の技能伝承を支援する対話システムの開発に貢献する。しかし、産業界の技術者の対話を直接収集し、大規模なデータを作ることは難しい。そこで、比較的多くの人がコツを有する料理に着目する¹。

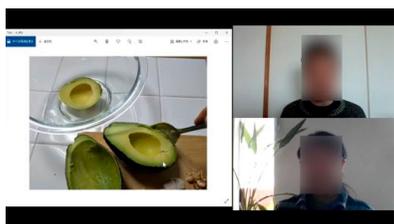
本研究では、オンラインビデオ対話において料理の技能者からインタビュアーが特定の料理の調理方法を聞き出すという設定で、料理インタビュー対話コーパス (**CIDC: Culinary Interview Dialogue Corpus**) を構築した。CIDCは、約6.4万発話の音声、その書き起こし、オンラインビデオ通話の画面映像のデータ(図1)と対話者の情報、技能者から収集した料理に関する情報、インタビュアーが事前に考えた質問内容をまとめたメタデータから構成される。

インタビュー対話の収集にオンラインビデオ通話を利用したのは、視覚的情報が共有できる対面对話に近い環境で、時間的、金銭的コストを抑えた対話の収集が可能となるためである。日本語を対象とした既存の話し言葉対話コーパスのほとんどは、電話会話を使用したものか(伝, フライ 2000)、参加者を実際に集め、その様子を録音・録画したもの(前川 2004; Fujimura et al. 2012; 伝, 榎本 2014; 小磯 他 2019)である。しかしながら、電話会話の場合は視覚的情報の共有ができないため、指示語の使用が減少する等、対面对話とは性質が大きく異なり、参加者をスタジオ等に集めて対話を収録する場合は時間的・金銭的コストがかかる。

現在、COVID-19の世界的感染拡大の中で、以前はオンラインビデオ通話を利用していなかった人々も様々なウェブ会議システムを使用するようになった。オンラインビデオ対話であれば、参加者はいつも使っている自身の機器を用いて自宅から対話に参加することが可能である。さらに、ウェブ会議システムを使用することで、参加者はお互いの表情を見つ、画面共有機能で

¹ 近年、スマートスピーカーが普及し、料理のレシピやコツをスマートスピーカーに対話的に質問する場面も増えてきている。そのため、料理ドメインのインタビュー対話を収集することは、そういったアプリケーションの開発にも貢献することが期待される。

| 発話者 | 開始時間 | 終了時間 | 発話内容 |
|-----|---------|---------|---|
| 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、 |
| 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。 |
| 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、 |
| 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。 |
| 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があります。 |
| 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で。はい。 |
| I | 05:12.3 | 05:16.2 | わかりました。少し(よーす) (よーす) (あの一) 余分を残して、 |
| E | 05:16.4 | 05:17.4 | (あ) そうですね。 |



a. 映像データ

b. 書き起こし

図 1 料理インタビューコーパスの例. 発話者の列の“E”は技能者を, “I”はインタビュアーを表す.

視覚的文脈を共有することが可能となり, 電話会話よりもより通常の対話に近いコミュニケーションを行うことができる. 通信の遅延などのずれは生じるものの, オンラインビデオ対話はこれらの視覚的情報を含めて対話の記録を行うことができる.

本稿では, 技能者からコツを引き出すインタビュー対話コーパスである CIDC の構築方法とその詳細について述べる. 2 節で関連研究について述べたのち, 3 節ではインタビュー対話の収集方法と書き起こしの方法について述べる. 4 節では CIDC の統計と特徴について述べる. そして, 5 節では全体をまとめ, CIDC の具体的な利用可能性について述べる. なお, CIDC は <https://nlp.ist.i.kyoto-u.ac.jp/?CIDC> にて公開中である².

2 関連研究

これまでのインタビュー対話システム研究においては, インタビュイーの意見や趣味嗜好, 日常生活に関する情報の取得を目指したインタビューを対象としたもの (Stent et al. 2006; Johnston et al. 2013; Kobori et al. 2016; Inoue et al. 2020; 曾, 中野 2021) やインタビュイーの精神状態や周辺環境について傾聴を行う医療目的のインタビューを対象としたもの (Gratch et al. 2014;

² タスク実施前に全ての参加者には, タスクの参加による身体的な危険が伴うことはないということを説明し, (i) インタビュー対話の映像データ, (ii) インタビュー対話の音声データ (文字化情報も含む), (iii) 参加者の役割・性別・年齢 (10 歳刻み)・使用するマイクの情報, (iv) インタビュー収録前に実施する事前調査の回答結果, (v) インタビュー収録後に実施するアンケートの回答結果を関連づけたデータベースを作成し, あらゆる研究者・研究機関が研究および研究成果の公表に用いることができるよう, 利用目的を学術研究に限定した上で公開することに関して同意を得ている. また, タスク終了後であっても, この同意を撤回する権利を有することに関しても事前に説明している.

DeVault et al. 2014) が多い。そのため、構築・公開されているコーパスに関しても、上記の目的を対象としたものが多い。

情報取得を目指したインタビューコーパスの多くは、マルチモーダル情報を含んでいない。杉山 他 (2016) は対話を通じた知識獲得の分析を目的として、「今頑張っていること」について学生に心理療法士が聞き取りを行った約 40 分間の対話音声を書き起こした傾聴対話コーパスを作成しているが、マルチモーダル情報は含まれていない。また、杉山らは今後の展望として対話数を増やすことを挙げているものの、収録対話数は 23 対話に留まっている。また、吉越他 (2021) では、話題の追従に着目した対話システム構築のため、聞き手が話し手から趣味を聞き出す雑談対話コーパスを構築している。この雑談対話コーパスはクラウドソーシングを用いたチャットによる対話コーパスであるため、5,114 対話 (1 対話あたり約 16 発話) からなる大規模なものであるが、画像のやりとり等は含まれていない。さらに、INTERVIEW (Majumder et al. 2020) や MEDIASUM (Zhu et al. 2021) といったニュースインタビューを対象とした大規模コーパスも構築されているが、これらも書き起こしデータのみを対象としている。

インタビュー対話の構造の分析には、テキストデータだけでなく、マルチモーダル情報 (発話の音声特徴、参加者の表情、ジェスチャー、視覚的情報等) の存在が重要である。特に本研究が目指しているコツを引き出すインタビューにおいては、技能者が説明に悩んでいる表情や実際の手つきを模したジェスチャーがコツを引き出している箇所に多く見られることが予想される。そのため、CIDC ではそれらの要素を含んだコーパスとなっており、コツの引き出し方を多角的に分析することが可能である。

一方、医療目的のインタビューを収集したコーパスとしては、花房、荒木 (2021) のカウンセリング対話コーパスや Gratch et al. (2014) の DAIC 等が挙げられる。花房、荒木 (2021) では、ナラティブ・セラピーのワークショップで行われた 4 つのカウンセリングにおけるカウンセラー役と相談者役の発話の文章化データを対象に、カウンセリングにおける対話の進行段階を 5 つに分けた対話段階を各発話に付与したコーパスを作成している。また、DAIC では鬱病や PTSD の症状についての医療インタビューを、(i) インタビュアーとの対面の対話、(ii) インタビュアーとのウェブ会議システム上での対話、(iii) Wizard-of-Oz 法を用いた対話、(iv) 自動対話システムとの対話の 4 つの設定でマルチモーダル情報を含めて収集している。これらのコーパスでは、インタビュアーはカウンセラーとして対話に参加しているため、傾聴対話としての特性が強く、特定分野の技能者からコツを引き出すインタビューとは大きく性質が異なる。

以上のように、インタビュアーとの対話の中で技能者からコツが引き出されるような対話設定でマルチモーダル情報を含んだコーパス構築を行っている研究はこれまでに見られない。

3 インタビュー対話の収集方法

3.1 対話設定

本研究では、技能者のコツがインタビューーとの対話によって引き出されるインタビューを実現するために、タスクとして料理の得意な技能者に対して特定の料理の調理工程について尋ねる2者対話を設定した。

参加者は各対話において、技能者かインタビューーのいずれかの役割で対話に参加した。

技能者 料理のコツを有し、インタビューーに対して、特定の料理の調理手順について説明する役割。事前に準備した料理の写真に対して大まかな説明を行った上で、インタビューーの質問に答える形で、具体的な作り方や注意すべき点を説明する。

インタビューー 技能者から料理に関するコツを引き出す役割。事前に技能者が紹介する料理について、技能者から提供された資料に目を通し、対話の準備を行う。

なお、対話前に技能者とインタビューーの双方に、技能者の役割は「自身の選んだある料理について写真をもとに手順などを説明」することであり、インタビューーの役割は、「技能者の説明を聞きながら、作業のコツなどのより詳細な情報を聞き出す」ことであるとインタビューの目的を明示している。

また、インタビューの収集コストとの兼ね合いで、プロレベルの技能者とインタビューーだけでなく、約半数は一般レベルの技能者とインタビューーを募集した。技能者については、料理に関する仕事に従事経験がある者を上級技能者³、料理を趣味としているが、職業とはしていない者を一般技能者に分類した。一方、インタビューーについては、人事面接など何らかのインタビューを実施した経験を持つ者を上級インタビューー、そのような経験がない者を一般インタビューーに分類した。なお、インタビューーは自炊経験があるなど最低限の料理の知識を持つことを参加の必須条件にしている。

対話データの収集はデータ収集の専門業者に依頼し、この業者が参加者の募集と事前事後のアンケート等を含めたインタビューの監督を行った。最初に暫定的な条件設定で20対話の予備収集を行い、その結果を受け、最終的な条件を定め、288対話を本収集した。以下の記述では、予備収集と本収集の両者共通の条件は断りをつけず記載し、本収集で新たに設けた条件についてはその都度記載する。

3.2 対話前の指示内容

CIDCが想定しているインタビューは、単なる調査インタビューやカウンセリングインタビューとは異なり、技能者とインタビューーの双方が技能者の持っているコツについて多く話すこと

³ 具体的な条件として、(i) 料理教室等で料理（種類は問わず）を教えた経験がある、(ii) 調理師免許を取得している、(iii) 料理家として活動している、という3条件のいずれかを満たす者を上級技能者とした。

ができるように事前に準備を行って取り組むものである。タスク実行前の事前準備として技能者とインタビュアーに以下のような指示を行った。

まず、技能者に以下の4つの料理の情報を準備し、提出するよう求めた。

- (1) a. **料理タイトル**: 紹介する料理のタイトル
- b. **料理概要**: 紹介する料理の概説 (50字程度)
- c. **手順写真**: 6-10枚の調理工程を説明するための写真
- d. **自慢したい点 (本収集のみ)**: 紹介する料理についての「自慢したい点 (こだわり、コツなど)」(少なくとも1つ)を記述した文章。対応する写真番号とともに報告を求めた。

実際に収集された料理の情報は図2のようなものである。

| | |
|---|--------------------|
| <p>料理タイトル 鰯の梅肉ロール天ぷら</p> <p>料理概要 梅肉にゴマ味噌風味をつけて、さっぱり味付け。ぱりぱりの骨せんべいと一緒につまみにぴったりです。</p> <p>自慢したい点 5枚目：魚の身の巻き終りは梅肉を塗らない。 9枚目：魚の骨を二度揚げする。</p> | <p>手順写真</p> |
|---|--------------------|

a. 「鰯の梅肉ロール天ぷら」の例

| | |
|---|--------------------|
| <p>料理タイトル 白玉あんみつ</p> <p>料理概要 なめらかもちもちな白玉団子と透明寒天、触感の楽しいおやつです。寒天はノンカロリーで食物繊維たっぷり、うれしい効果もおすすめです。</p> <p>自慢したい点 3枚目：沸騰してからかき混ぜながら2分。 6枚目：水は少しずつ入れる。 8枚目：浮き上がってきたら2分。</p> | <p>手順写真</p> |
|---|--------------------|

b. 「白玉あんみつ」の例

図2 実際に技能者から提出された料理の情報の例

さらに、技能者には以下のような指示を事前に行った。

- (2) a. 写真を順番に表示できるように準備をしておく（出来上がり写真→手順の写真→出来上がり写真）
- b. 手順の写真を表示しながら「まず野菜を切ります」「塩・胡椒をします」など、「何をしている場面か」を簡単に説明し、あとは質問を待つ。
- c. 次の写真に進むタイミングはインタビュアーが示す。ただし、技能者が説明し足りないことがあると感じた場合には、追加で説明を行って良い。
- d. 話している最中、写真に対象物が写っている場合には、マウスポインタで指し示しながら説明する。原則として、マウスポインタは見やすいように事前に拡大しておく。（本収集のみ）

(2a) から (2c) はインタビューの流れを統一するために設定した事項である。また、(2d) は予備収集の結果を受けて追加した事項であり、指差しに相当するものとしてポインタの使用を明示的に指示したものである。

一方、インタビュアーには対話タスクの前に上記 (1a), (1b), (1c) の 3 つの情報を与えた上で、以下の指示を事前に与えた。

- (3) a. 事前に料理タイトル、料理概要、手順写真を予習し、質問の内容とタイミングをイメージしておく。
- b. インタビュアーが進行役をつとめる。技能者は最低限の手順しか説明しないので、それ以上の説明を引き出すよう質問する。次の写真に進むタイミングもインタビュアーが示す。

(3a) の指示に関して、本収集では事前に「良い質問のタイプ」を教示し、1 つ以上の質問内容とその質問を何枚目の写真で行うつもりか事前に著者らに報告させ、その内容をインタビュー中に聞くように求めた。この指示は、実際のインタビューを想像しながらインタビュアーが予習を行えることを目的として設定したものである。

「良い質問のタイプ」とは、予備収集の結果を受けて、インタビュアーが技能者からより深いコツを引き出せるよう作成した 11 個の質問方法とその例を載せた資料である (図 3)。本研究では、技能者が自発的には言語化しづらい知識、感覚を引き出すようなインタビュー対話を実現することを目的としている。そのため、技能者が自発的に言語化可能な内容（例えば、レシピにも書かれるような内容）のみを尋ねるのではなく、対話の中で技能者自身にも気づきを与える対話となることが望まれる。しかし、予備収集では、技能者がスムーズに説明を行い、インタビュアーはそれに相槌や反応を返す対話になってしまうことが多かった。そこで、本収集で追加したこの資料では、「相手（技能者）が答えにつまりそうな質問」や「相手自身も意識し

この資料は、相手（技能者）が答えにつまりそうな質問の仕方をまとめたものです。相手自身も意識していなかった料理のコツを聞き出せるよう、この資料を参考に質問を考えてみてください。

1. 調味料や材料の分量・重さ、煮込み時間などの数値的な情報を質問する
 - 「お醤油は小さじ何杯くらいですか」
 - 「何分ぐらいで火を止めますか」
2. 「とろみ」「こんがり」「まとまってきたら」などの抽象的な表現について、客観的に判断する手がかりを質問する
 - 技能者：「とろみが出てきたら火を止めます」
 - インタビューア：「とろみが出てきたのをどうやって判断したら良いでしょうか」
3. 同じ材料・器具がないときはどうするのかを質問する
 - 「黒酢がないときは、どうしたら良いでしょうか」
4. 食材の下準備の説明が省略される傾向にあるので、あえて質問する
 - 「お野菜はどうしてこの切り方をしたのですか」
 - 「どうして塩をかけるのですか」
5. 写真に映っているのに技能者が特に説明しなかったものを見つけ、それについて質問する
6. 技能者から説明を引き出したら、もう一歩踏み込んで、それが良いことなのかも質問する
 - インタビューア：「ネギは、青いネギじゃなくて白いネギですね」
 - 技能者：「はい、白いネギの青いところと白いところを両方入れました」
 - インタビューア：「その方が美味しいのですか」
7. ポイントを率直に尋ねる
 - 「何かコツはありますか？」
 - 「気をつけるべきことはありますか？」
 - 「難しいところはどこですか？」
8. 避けた方が良いことや、失敗経験を質問する
 - 技能者：「鯛の代わりに、他の白身魚やホタテでも良いですよ」
 - インタビューア：「良いですね。逆に、イマイチなものってありますか」
9. 反対の意見や、否定的な意見を投げかけてみる
 - 「ピーマンを入れると、子どもが食べないんじゃないでしょうか」
10. 自分なりの仮説を立てて質問する
 - 「砂糖の代わりにみりんを使うときも、同じ分量を入れたら良いでしょうか」
11. はい・いいえで答えられる質問を、「どうして」「どうやって」と相手に話させるような質問に言い換える
 - 「切り方はもっと小さくても良いですか」を「どうしてこの大きさに切ったんですか」と言い換える
 - 「これ、イタリアンパセリじゃなくても大丈夫ですか」を「イタリアンパセリがないときはどうしていますか」と言い換える

図 3 インタビューアに教示した「良い質問のタイプ」

ていなかった料理のコツを引き出すといった目的を明示し、インタビュアーと技能者の対話がレシピにも記載されるような内容を尋ねるのみに留まらないようにした。

「良い質問のタイプ」作成においては、予備収集の対話における以下の3つの場面に着目し、その際にインタビュアーがどのような質問をしていたのかを調査し、一般化した。

- 技能者の発話にコツが含まれていると思われる場面
- インタビュアーがもう少しでコツを引き出したのではないかと思われる場面
- 「うーん」という発話や視線を外すような仕草によって技能者が回答の内容を考えながら話している様子を見せた場面

技能者に提出を求めた(1)の料理の情報とインタビュアーに事前に提出を求めた質問内容は、料理の情報として本コーパスに含まれている。

3.3 対話の収録方法

インタビューにはオンライン会議システムのZoom⁴を使用した。参加者は外付け、またはコンピューターに内蔵されたカメラとマイクを使用し、画面共有機能によって技能者側の画面を共有した状態で対話を行った（共有画面についての詳細は後述）。録音・録画に関してはZoomの録画機能を利用し、音声、カメラの映像、共有された画面の全てを記録した。それぞれの参加者は、各対話タスクの前に最大5分程度の練習を個別に行った（具体的な手順については付録A）。1つの対話は15分程度で終了するように、時間超過が予想される場合はチャット機能によって合図が送られた。

3.4 対話終了後のアンケート

対話タスクの終了後、技能者とインタビュアーのそれぞれに以下の事項についてアンケートを実施した。

- (4)
- a. コミュニケーションはスムーズだったか
 - b. インタビュアーが技能者からコツをうまく聞き出したか
 - c. 自身が料理の知識や技能を持っているか⁵（インタビュアーのみ）
 - d. インタビューの感想

5件法で回答を求めた(4a), (4b), (4c)は「1. そう思う」「2. ややそう思う」「3. どちらでもない」「4. あまりそう思わない」「5. そう思わない」の選択肢から1つを選ばせた。なお、(4c)はインタビュアーが専門的な立場から質問を行っているのか、一般の立場からコツを引き出すこ

⁴ <https://zoom.us>

⁵ この問いは、紹介されている料理のジャンル（和食、洋食、洋菓子等）によって同一参加者であっても知識や技能が異なるため、タスクごとに回答を求めることとした。

とを試みているのかを判断するために実施した項目である。最後に、(4d)のインタビューの感想は、「その他、感想や印象に残ったこと、お気づきの点などがありましたら自由にお書きください。」と指示し、無回答での提出も許容した。

3.5 対話データの書き起こし

専門業者に依頼し、人手によって全ての対話データの発話内容を発話者、開始時間、終了時間とともに書き起こした。発話の単位は500ミリ秒以上のポーズを基準とし、同一話者の発話であっても間にポーズが入る場合は異なる発話として書き起こしを行った。書き起こし手順として、最初に音声認識システムのAmiVoice[®]⁶を用いて自動書き起こしを行い、それを人手で修正した。書き起こしは、技能者からコツを引き出す際に観察される発話の内容を検討するためのデータとすることを目的としている。そのため、参加者の発話内容そのものに焦点を当て、相槌や笑いは最小限の書き起こしに留めた。具体的には付録Bに記載の方針を採用し、疑義が生じた場合はその都度協議の上で書き起こしを実施した。

4 コーパスの構築結果と分析

4.1 コーパスの統計

料理インタビュー対話コーパス(CIDC)は308件の対話からなる。コーパス全体に関する統計情報は表1の通りである。総語数、発話数ともに技能者がインタビュアーよりも多いのは、事前に両者の役割を設定し、インタビュアーが技能者からコツを引き出すよう指示した効果であると言える。

| | 技能者 | インタビュアー | 合計 |
|----------------|---------|---------|---------|
| 総語数 | 370,771 | 218,751 | 589,522 |
| 異なり語数 | 7,266 | 5,079 | 8,226 |
| 発話数 | 35,478 | 28,677 | 64,155 |
| フィルター数 | 41,301 | 28,251 | 69,552 |
| アンケート(4a)の回答平均 | 4.64 | 4.25 | 4.44 |
| アンケート(4b)の回答平均 | 4.62 | 4.08 | 4.35 |
| 平均対話時間(分) | — | — | 12.63 |

表1 収集されたインタビュー対話全体の統計。単語数については、フィルターと言い誤りを取り除き、Juman++(Tolmachev et al. 2018)を用いて形態素解析を行った。発話数については、フィルターや言い誤りのみで構成される発話以外の数を示す。

⁶ <https://www.advanced-media.co.jp>

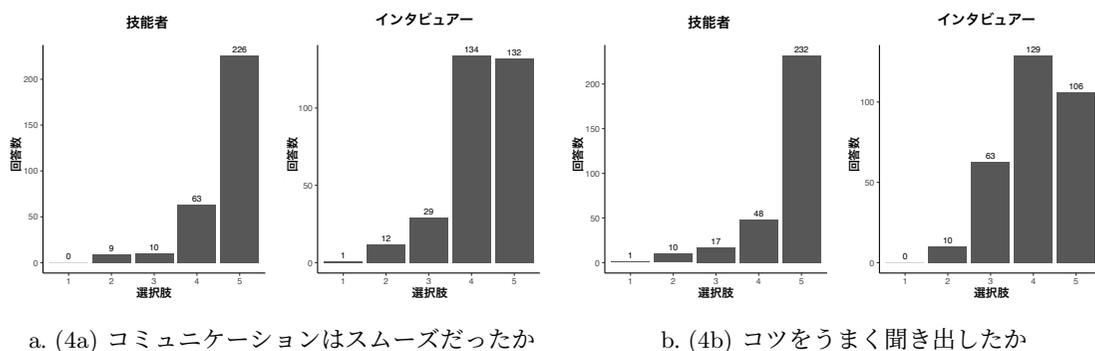


図 4 インタビュー後のアンケート (4a) と (4b) に対する回答の分布。

対話タスク実施後に行ったアンケートの結果を図 4 に示す。(4a)の「コミュニケーションはスムーズだったか」という問いでは、技能者の平均値が 4.64 であり、インタビュアーの平均値が 4.25 であった。回答の分布を見ても、技能者とインタビュアー共に 85%以上が「5. そう思う」と「4. ややそう思う」を選択している。

(4b)の「コツをうまく引き出していたか」という問いでは、技能者の平均値が 4.62 であり、インタビュアーの平均値が 4.07 であった。回答の分布を見ると、技能者は約 90%が「5. そう思う」と「4. ややそう思う」を選んでいる。インタビュアーは、「5. そう思う」と「4. ややそう思う」を選んだのは約 75%であり、約 20%が「3. どちらでもない」を選んでいる。インタビュアーは、あらかじめ技能者からコツを聞き出すよう指示されていたことにより、厳しい自己評価をしていると考えられるが、全体としては参加者自身がコツを引き出すインタビューができたと感じていると言える。

さらに、インタビュアーのみに実施したアンケートである (4c)の「自身が料理の知識や技能を持っているか」については、平均値が 2.48 (SD = 1.09) であり、約 6 割が「1. そう思わない」か「2. あまりそう思わない」を回答していた。この結果から、インタビュアーは専門的な立場から質問するのではなく、あくまでも一般の立場からコツを引き出そうとすることが多かったことが分かる。

最後に、(4d)のインタビューの感想については、表 2 のような感想が集まった。技能者の回答にはインタビュアーに対する好意的な意見が多く見られ、インタビュアーの回答には自身がインタビュー中に聞けなかったポイント等の反省点が多く見られた。

インタビュー参加者数は 26 人 (男性: 5 名, 女性: 21 人; 20~29 歳: 1 人, 30~39 歳: 5 人, 40~49 歳: 12 人, 50~59 歳: 8 人) であり、その内訳は表 3 の通りである。3.1 節で述べたように、参加者それぞれに技能者とインタビュアーの 2 つの役割を設定した。技能者の役割でタスクに参加したのは 17 人、インタビュアーの役割でタスクに参加したのは 17 人であるが、そのうち

| 料理名 | 回答者の役割 | (4d)の回答 |
|---------------|---------|---|
| 芋煮 | 技能者 | こうしてもいいかなど代替案の提示があり、想像して考える機会を与えてもらった。料理の仕方がより良いものになっていくのではと思う。 |
| | インタビュアー | 音声が少し聞き取りづらかったです。時間配分がうまくいかず、聞きたいことがすべて聞けず残念でした。 |
| シュークリーム | 技能者 | ご質問のタイミングがつかみかね、こちらから写真の説明を先にするような流れになりました。それぞれの工程に興味を持って聞いていただきよかったです。 |
| | インタビュアー | 事前に写真は見せていただいています、各写真のご説明の途中で質問していいのかお話を終えられてからかとまどいました。お話の途中で質問をしてしまったりが何度かありました |
| 厚揚げと野菜の明太あんかけ | 技能者 | 同じお料理でも、質問される方によって視点が変わるんだなと感じました。「ひとつのお鍋でできるから便利」「小松菜だけ冷めている状態」など、お一人目のインタビュアーの方(対話ID:92)の質問とは少し違うので、また別の視点からもお答えできたかなと思いました。 |
| | インタビュアー | 明太子が味にも見た目にも良いアクセントになると思いました。胸肉が固くならない調理方法もわかりました。 |
| サーモンの押し寿司 | 技能者 | インタビュアーの方も料理をよく作られるとのことと料理のコツに関する質問が的を射ていたと思います。 |
| | インタビュアー | 先生が緊張されているかな?と感じながら進めました。前回のインタビューから少し間が空きましたが、自分が少し慣れてきた感がありました。 |
| ショコラロータルト | 技能者 | 材料を揃えたり下準備に時間はかかるけれど、工程そのものは非常に簡単であることを理解してもらえ、ご本人がそのようにまとめて言って下さりよかったです。 |
| | インタビュアー | ココナッツオイルに関して、無臭のココナッツオイルはない、とお話がありましたが、エキストラバージンココナッツオイルは新鮮なココナッツを非加熱処理しているの、全て匂いがある。一方、無臭のココナッツオイルは、加熱処理していて、エキストラバージンではない、という事だったのか、とインタビュー後にネットで確認して腑に落ちました。無臭のココナッツオイルがない、というコメントが気になりまして。質問をする際に、「どのようなココナッツオイルですか?」と広く質問をした方が良かった、と思いました。 |

表 2 アンケート (4d) の回答例

| インタビュアー | 技能者 | | |
|---------|-----|----|---|
| | 上級 | 一般 | × |
| 上級 | 2 | 1 | 5 |
| 一般 | 3 | 2 | 4 |
| × | 3 | 6 | — |

表 3 役割・熟達度ごとの参加者の内訳。×はその役割では不参加だったことを表す。

| 参加者の属性 | | 対話数 |
|--------|---------|-----|
| 技能者 | インタビュアー | |
| 上級 | 上級 | 79 |
| 上級 | 一般 | 80 |
| 一般 | 上級 | 74 |
| 一般 | 一般 | 75 |

表 4 参加者の熟達度ごとの対話数の内訳

8人は両方の役割で参加しているため、タスク参加者の合計は26人である。技能者のうち上級技能者は8人、一般技能者は9人であり、インタビュアーのうち上級インタビュアーは8人、一般インタビュアーは9人であった。また、参加者の属性ごとの対話数は表4の通りであった。

インタビューにおいて紹介された料理は152種類であった。紹介された料理を主食、主菜、副菜、デザート、その他に分類すると分布は図5のようになり、各分類の例は表5に挙げた。

また、本収集で集めた288対話から無作為に選んだ30対話において、「良い質問のタイプ」を読んだインタビュアーに事前報告させた質問内容が実際にインタビュー内で質問されているかを調査した。その結果、事前に報告された質問の数は1対話あたり平均3.07 (SD = 1.78)であった。その質問のうち、インタビュアーが実際に対話中で技能者に尋ねた質問の数は平均2.52 (SD = 1.57)であった。また、インタビュアー側から質問したかどうかに関わらず、事前に報告された質問の回答となる内容に技能者が言及していた数は平均2.93 (SD = 1.74)であった。これらの結果から、インタビュアーが事前に用意した質問に関してその回答となる内容をほぼ引き出すことが出来ていたことが分かる。

4.2 ウェブ会議システムを用いた影響

CIDC の特徴の一つは、ウェブ会議システムの Zoom を用いて収録を行ったという点である。ウェブ会議システムは、対話相手の表情が見えるだけでなく、画面の共有機能によって同一の視覚的コンテキストを参照することができる。その一方で、顔以外の対話相手の様子を観察できない点や、画面共有されている視覚的コンテキストに直接的な働きかけができない点は対面形式の対話には見られない制約である。このようなビデオ通話独自の問題について、参加者は柔軟に対処することができていた。

まず、対話相手の顔しか見えないという制約には、ジェスチャーが必要となる箇所ではカメラに映る位置に手を持ってきて、ジェスチャーを行うことで対処していた。(5)はその一例であり、技能者が発話の太字部分でケーキの型の底部分が持ち上がることを表すジェスチャー（図 6a）とケーキ自体が取り出せることを表すジェスチャー（図 6b）を行っている。

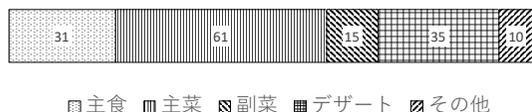
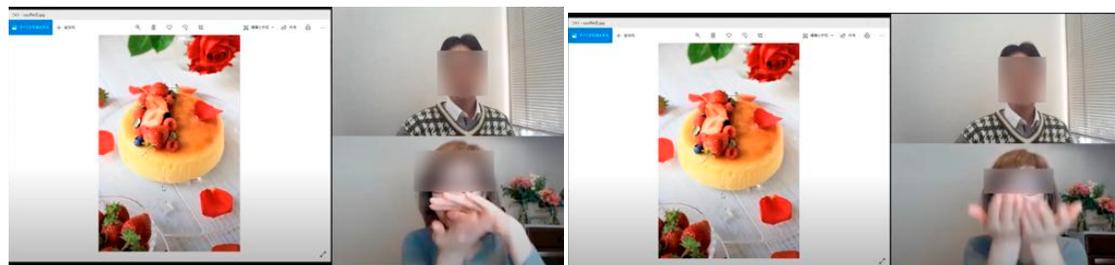


図 5 紹介された料理の内訳

| 分類 | 料理名 |
|------|---------------|
| 主食 | タイグリーンカレー |
| | こてっちゃん炒飯 |
| 主菜 | 厚揚げと野菜の明太あんかけ |
| | エビとレンコンの春巻き |
| 副菜 | ささみの胡麻和えサラダ |
| | おさしみの大根おろしあえ |
| デザート | チョコレートケーキ |
| | ポーロ |
| その他 | 米こうじ甘酒 |
| | 柚子ジャム |

表 5 料理の分類の例



a. ケーキ型の底部分が持ち上がるジェスチャー

b. ケーキが取り出されるジェスチャー

図 6 (5) のインタビュー中に見られたジェスチャー

(5) 「スフレチーズケーキ」(Interview_43: 00:00:28.9-00:00:49.6; 図6)

- 8 I : 底が、底が取れるっておっしゃられました？
- 9 E : はい。(あ) そうなんです。(あのー) スフレって(えーと) フランス語で泡っ
 という意味なんですけれども、とてもふわふわとして型から(えー) 取り出
 しにくいので、この(そそ) 底ですね、底が(こー) 持ち上がるようになって
 いて、ケーキがそのまま
- 10 I : (へー)。
- 11 E : 取り出せるような型を使っています。

次に、画面共有された視覚的コンテキストへの働きかけについては、マウスカーソルを動かすことで指差しに相当する指示を行っていた。(6)はその一例であり、太字部分でマウスカーソルを円を描くように動かしている。ここでは、インタビュアーが加熱用の生牡蠣を加熱済みの牡蠣と誤解しており、技能者はその説明を行っている。発話26で「これ生ガキなんですけど」と言いながら、写真のパッケージに書かれている「生かき」という文字の上でカーソルを動かし(図7)、その後、発話26から発話32にかけて「加熱調理用」という文字上にカーソルを移し、その説明を行っている。

(6) 「牡蠣と鮭のチャンチャン焼き」(Interview_177: 00:01:21.100-00:02:03.550; 図7)

- 24 E : (んでー) 基本的に火通す料理のときは加熱調理用の方、使ってもらった方が
 いいかなと思います。
- 25 I : (あ) なるほど。もう加熱してあるものの方が、

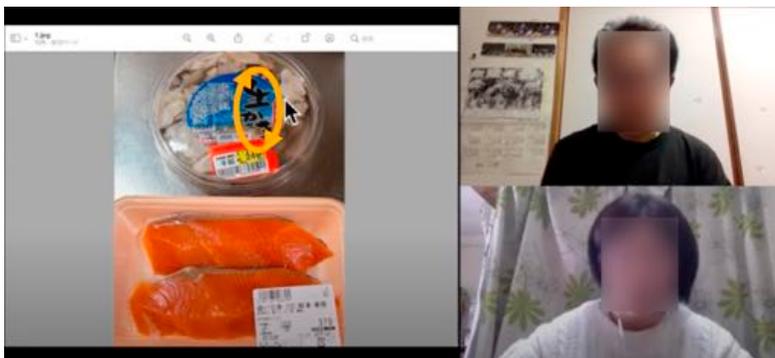


図7 (6)の発話26に見られたカーソルの動き。図中に示したように、「生かき」の文字上で円を描くようにマウスカーソルが動かされている。

- 26 E : いいえ (えーと) そうでは, そうではなくてですね, (えーと) これ生ガキな
んですけど,
- 27 I : おすすめということです.
- 28 E : (えと) ここにですね,
- 29 I : はい.
- 30 E : 加熱調理用ってかい.
- 31 I : はい.
- 32 E : (?) これは火を通して食べてくださいという意味なんですけど,

以上のように, ウェブ会議システムの利用において問題となりうるジェスチャー使用や視覚的コンテキストへの働きかけは柔軟に対応されている. これは, COVID-19 の感染拡大に伴うウェブ会議システムの普及によって, 対話参加者がビデオ通話という形式に慣れ親しんできていくことも影響していると考えられる.

4.3 参加者の熟達度による影響

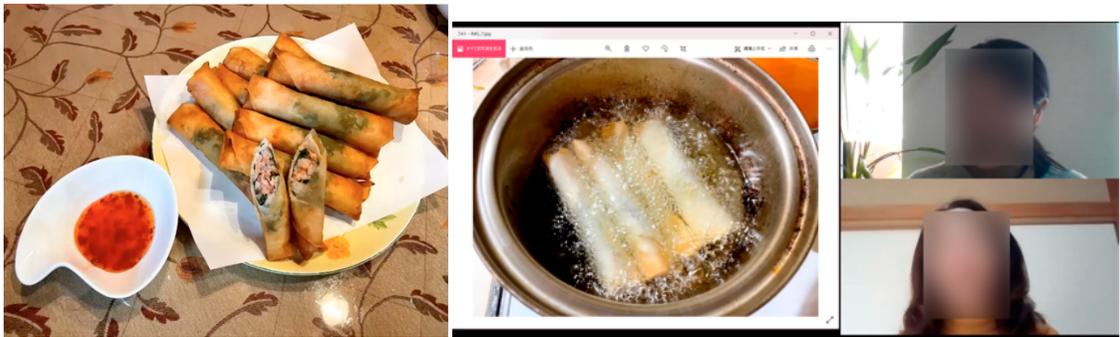
3.1 節で述べたように, 熟達度に応じて技能者を上級技能者と一般技能者に, インタビュアーを上級インタビュアーと一般インタビュアーに分けたが, インタビューの内容そのものからは熟達度に応じた大きな差異は見出せなかった.

技能者とインタビュアーの熟達度ごとの総語数と異なり語数を表 6 に示す. テキスト中に使用されている語彙の豊富さを表す指標である *Uber* 指標 (Dugast 1978, 算出方法は付録 C 参照) を見ると, 上級技能者と一般技能者, 上級インタビュアーと一般インタビュアーの間に大きな差は見られない. また, 語られている発話内容に関しても上級と一般との間に大きな質的差異は見られなかった.

なお, 発話の内容そのものではなく, 技能者の発話末尾に見られる語用論的特徴に熟達度による差が見られた. 具体的には, 調理手順の詳細や意図について質問された際, 上級技能者は断言する形式での回答が多かったが, 一般技能者の場合は「～と聞きました」「～だそうです」「多分～と思います」「～とされています」などと断言を避ける回答が多く見られた. (7) は一

| | 技能者 | | インタビュアー | |
|-------------|---------|---------|---------|---------|
| | 上級 | 一般 | 上級 | 一般 |
| 総語数 | 205,783 | 164,988 | 117,124 | 101,627 |
| 異なり語数 | 5,491 | 5,080 | 4,143 | 3,439 |
| <i>Uber</i> | 17.9 | 18.0 | 17.7 | 17.0 |

表 6 技能者とインタビュアーの熟達度ごとの総語数, 異なり語数, *Uber* 指標.



a. 料理完成写真

b. (7) の対話中の映像

図 8 「エビとレンコンの春巻き」に関する情報

一般技能者と一般インタビュアーの対話であるが、春巻きを一度に揚げる本数について発話 249 で「と聞きました」という形の伝聞体が用いられている。

(7) 「エビとレンコンの春巻き」(Interview_108: 00:10:02.200-00:10:21.050; 図 8)

244 E : これは (あの) 4 本が限界なので、

245 I : (あ) 入る分を順番に揚げていくような感じで、

246 E : (?) です。そう、はい、はい。

247 I : (あ)。

248 I : <まだ>。

249 E : 揚げました。はい、ただ (あのー) もしフライパンとか一気に入るのであれば、その方がいいと聞きました (あのー) 一気に揚げたほうが。

語用論において、このような発話内容の真偽を曖昧にする表現はヘッジ (Hedge) と呼ばれている。ヘッジがない発話は話者の観察や論理的思考を通して得た知識に基づくものであるが、ヘッジがある発話はそのような確固たる知識がなく、推測に基づく発話が多いと指摘されている (Prince et al. 1982)。上級技能者の発話にヘッジが少ないのは、彼らが料理を職業としていた経験があるため、一般技能者以上に様々な種類の料理を数多く作ってきた経験によるものであると考えられる。

4.4 コツが引き出されたか

3.2 節で述べたとおり、技能者が単に手順を説明するのではなく、インタビュアーが技能者から料理のコツを聞き出すというインタビューの目的を技能者とインタビュアーが互いに理解

し、その準備をした上で対話する設定を採用した。この対話設定により、CIDCではインタビュアーが効果的にコツを引き出している場面が多く見られた。その詳細な分析は今後の研究に譲り、ここでは「良い質問のタイプ」で事前に提示した以外の手法でインタビュアーがコツを引き出している場面の中で効果的であると思われる2つのパターンを紹介する。

パターン1は、直前に技能者が発言した内容をインタビュアーが繰り返したり、言い換えたりするというものである。繰り返しや言い換えは、対話でトラブルが生じた場合に修復の開始を示す機能 (Schegloff et al. 1977; Schegloff 1997) や、傾聴対話において相手への共感を示す手法として知られている (楡木 1989; 三島, 久保田 2003; ホールファミリーケア協会 2004)。一方、今回のインタビュー対話においてはトラブル修復や共感を示すのではなく、それによって技能者がコツに関する詳細な内容を語るという場面が見られた。

(8) 「セロリ多めのラグーソースの Pasta」 (Interview_56: 00:06:41.150-00:07:41.050; 図9)

88 E : はい。こちらは先ほど (あの一) 大きく (あの一) (ま) 切り分けたニンジン、セロリ、にんにく、玉ねぎを、もう一度に (ふ) (あの一) フードプロセッサーの容器に移しまして、かける準備をしているところです。

中略 (約 16.5 秒)

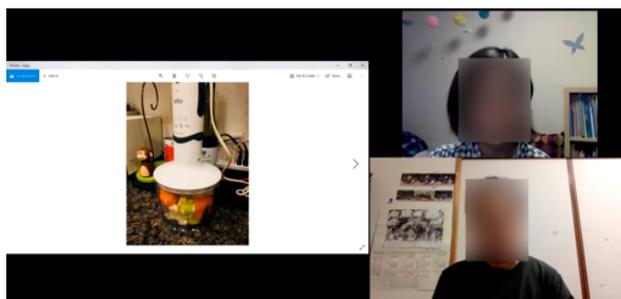
95 E : 3か4ぐらいに合わせていただいて、もう1分も回さないうちにみじん切りができますので、あまり細かくなり過ぎるとピューレになってしまいますので、それには気をつけてみじん切り程度

96 E : で終わるようにスイッチを切っていただく。

97 I : (あーあー)。(えっと)(ん)(み)みじん切り、みじん切り程度が目安っていう感じですかね？



a. 料理完成写真



b. (8) の対話中の映像

図9 「セロリ多めのラグーソースの Pasta」に関する情報

- 98 E : そうですね. ちょっと粒が(あの一) 2 ミリ角ぐらいで残る程度が, 野菜の(こ一) 食感も楽しめると思います.
- 99 I : (は一は一は一).
- 100 I : (あ一あ一) なるほどな. (な) なるほど. (えっと) かけすぎに注意っていうことですね. もう完全にドロドロになってしまうとまずいっていう.
- 101 E : そうですね. はい.

発話 95 で技能者が「みじん切り程度」と言ったのに対し, インタビュアーは頷きながらその話を聞いているため, この箇所に対話上のトラブルは見られない. しかしながら, 直後の発話 97 で「みじん切り, みじん切り程度が目安って感じですかね?」と繰り返しによる確認の発話を行っている. これにより, 発話 98 で技能者から具体的な野菜の大きさ(「2 ミリ角ぐらい」とその目的(「野菜の(こ一) 食感も楽しめると思います」)について引き出すことができている.

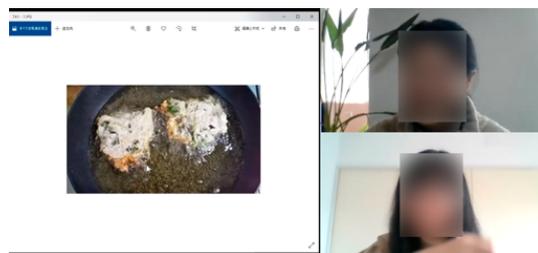
もう 1 つのパターンは, インタビュアーが自身の経験に言及し, 具体的な状況を設定, 提示することによって, 技能者から細部についての説明を引き出すというものである. 傾聴対話においては, 話し手の発言を受けて聞き手側が自己開示をする(e.g. 「私も〇〇なんです」)ことで, 話し手は自身の話に対して聞き手がより共感や理解をしてくれていると報告されている(石田 他 2018). 一方, 今回収集したインタビュー対話においては, 技能者の話を受けたインタビュアーの自己開示ではなく, むしろ話の起点として自身の経験について語るというパターンが見られた.

(9) 「生シラスのふっくらかき揚げ」(Interview_145: 00:10:34.250-00:11:28.450; 図 10)

- 177 I : (ふーん) (で) 何か (この) 揚げているときのコツとかってありますか?
- 178 E : (あ).



a. 料理完成写真



b. (9) の対話中の映像

図 10 「生シラスのふっくらかき揚げ」に関する情報

- 179 E : 揚げているときのコツですね.
- 180 E : (えーと)
- 181 E : 揚げているときの(こ).
- 182 I : 例えば, 私, よく(あのー)何か(こー)心配で, できているのかなって, (こー)菜箸でぐしゃぐしゃぐしゃ,
- 183 I : やったりしてしまうんですけど, (な)(な)(な).
- 184 E : (?).
- 185 E : そうですね, つまり, なるべく(あのー)なるべく触らないように.
- 186 I : (あー)やっぱりそうですか.
- 187 E : (?).
- 188 E : 片面を揚げましたら, (えーと)なるべく触らないようにそのまま(えー)じっくり火を通して, また裏返して, その後も(あのー)ほとんど触らないように, (えー)形がくずれないように揚げていって, (で)最終的に(あのー)
- 189 E : (えーと)バットに上げるときに, 油をしっかりと(こー)
- 190 E : 切るというのが, (あの)カリッとふわっと仕上がるコツかなと思いますので.

上記の例では, 発話 177 でインタビュアーが漠然と「揚げている時のコツとかってありますか?」と質問したため, 発話 178 から発話 181 にかけて, 約 6 秒間, 技能者が回答に窮している。これを見たインタビュアーは, 発話 182 で「例えば」と切り出し, 自身が揚げ物を行う際に失敗した事例を挙げることで, 技能者から発話 185 以降の揚げる時に気をつけるべき点を引き出している。

以上の 2 つのパターンは傾聴対話において見られる, 相手への共感や話を聞いていることを示す「繰り返し・言い換え」やインタビュアーの自己開示としての「自身の経験を語ること」とは性質を異にする。これらのパターンは, コツを引き出すことを目的とした対話設定によって生じたものであると考えられる。

5 おわりに

本稿では, 技能者から知識を引き出す対話システムの構築に向けた料理ドメインにおけるインタビューコーパス CIDC の構築について報告した。CIDC の特徴は以下の 2 点である。1 つ目の特徴は, 参加者が技能者とインタビュアーに分かれていることに加えて, コツを引き出すというという目的を事前に共有したことである。現在, 筆者らは CIDC に対して詳細なアノテ

ションを実施しており、そのアノテーションに基づき、4.4節で挙げたようなインタビュアーによってコツが引き出されている場面を定量的に分析することを目指している。2つ目の CIDC の特徴はウェブ会議システムを使ったことである。これにより、時間的に低コストでありながら、技能者からコツを引き出すインタビュー対話を大規模に収集するという目的を達成することができた。

なお、アンケートの (4d) で「インタビューの感想」を求めた自由記述の結果を見ると、インタビュアーが技能者よりも知識や経験がある場合、インタビュアーが技能者の説明に対して不満を持つ事例が少数見受けられた。今回の対話収集では、同一の参加者が技能者とインタビュアーの双方の役割で参加することが可能であった。そのため、技能者としては上級技能者と分類される者がインタビュアーとして一般技能者にインタビューをする対話が存在した。実際のインタビューにおいては、インタビュアーは技能者と同程度以下の知識を持った者であることが予想されるため、上記のような設定は避けるべきであった。

また、予備収集の結果を受けて、本収集ではインタビュアーに「良い質問のタイプ」を提示し、事前に質問内容を具体的に考えさせる手続きを取ったが、その結果、15分という制限時間が短く感じたという感想を提出した参加者が複数存在した。インタビューの内容に合わせた適切な時間設定についても今後のコーパス構築において重要な観点であることが明らかとなった。

CIDC の具体的な利用方法の一例として、CIDC に対して述語項構造や意味フレームに基づくアノテーションを実施し、それを学習データとした知識構造解析システムを開発することが考えられる。また、マルチモーダルな情報に関しては、ポイントの指示物と言語化されている内容に対応させるアノテーションを実施することで、上述の知識構造化をより精緻に行うことが可能になるだけでなく、技能者の表情やジェスチャーを利用することでコツについて語る場面を視覚的にも検討することが可能となる。このような知識構造解析システムが開発できれば、技能伝承の場面において、技能者が言及した知識をリアルタイムで構造化し、まだ言及されていない知識をインタビュアーに提示するようなインタビュー支援システムを実現することもできるだろう。

謝 辞

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務 (JPNP20006) の結果得られたものです。ここに感謝致します。

本資料は Okahisa et al. (2022) に基づくものであり、当該論文は ELRA が著作権を有して CC BY-NC で頒布しています。本資料を CC BY で頒布することについては ELRA の許可を得ました。

参考文献

- 伝康晴, 榎本美香 (2014). 千葉大学 3 人会話コーパス (Chiba3Party). 国立情報学研究所 音声資源コンソーシアム. (データセット). <https://doi.org/10.32130/src.Chiba3Party>, [Y. Den and M. Enomoto (2014). Chiba3Party. National Institute for Japanese Language and Linguistics Speech Resources Consortium. Dataset. <https://doi.org/10.32130/src.Chiba3Party>].
- 伝康晴, フライジョン (2000). 日本語 CallHome コーパス. 音声研究, 4 (2), pp. 24–30. [Y. Den and J. Fry (2000). CallHome Japanese Corpus. *Onsei Kenkyu*, 4 (2), pp. 24–30.].
- DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Xu, Y., Rizzo, A., and Morency, L.-P. (2014). “SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support.” In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS '14, pp. 1061–1068, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.
- Dugast, D. (1978). “Sur Quoi se Fonde la Notion D’etendue Theoratique du Vocabulaire?” *Le Francais Modern*, 46 (1), pp. 25–32.
- Fujimura, I., Chiba, S., and Ohso, M. (2012). “Lexical and Grammatical Features of Spoken and Written Japanese in Contrast: Exploring a Lexical Profiling Approach to Comparing Spoken and Written Corpora.” In *Proceedings of the VIIth GSCP International Conference*, pp. 393–398. Speech and Corpora.
- 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 9th International Conference on Language Resources and Evaluation (LREC'14)*, pp. 3123–3128, Reykjavik, Iceland. European Language Resources Association (ELRA).
- 花房竜馬, 荒木健治 (2021). ナラティブ・セラピーにおける対話段階予測のための素性設計. 日本知能情報ファジィ学会 ファジィ システム シンポジウム講演論文集, 37, pp. 425–430. [R. Hanabusa and K. Araki (2021). Feature Design for Dialogue Stage Prediction in Narrative Therapy. *Proceedings of the Fuzzy System Symposium*, 37, pp. 425–430.].
- Inoue, K., Hara, K., Lala, D., Yamamoto, K., Nakamura, S., Takashi, K., and Kawahara, T. (2020). “Job Interviewer Android with Elaborate Follow-up Question Generation.” In *Proceedings of the 2020 International Conference on Multimodal Interaction*, pp. 324–332, New York, NY, USA. Association for Computing Machinery.

- 石田真也, 井上昂治, 中村静, 高梨克也, 河原達也 (2018). 共感表出と発話促進のための聞き手応答を生成する傾聴対話システム. 人工知能学会研究会資料 言語・音声理解と対話処理研究会, **82**, pp. 7–12. [M. Ishida et al. (2018). An Attentive Listening System Generating Empathetic Responses and Promoting User Utterances. *Zinko Chino Gakkai Kenkyu-kai Shiryo Gengo Onsei Rikai-to Taiwa Shori Kenkyukai*, 82, pp. 7–12.].
- Johnston, M., Ehlen, P., Conrad, F. G., Schober, M. F., Antoun, C., Fail, S., Hupp, A., Vickers, L., Yan, H., and Zhang, C. (2013). “Spoken Dialog Systems for Automated Survey Interviewing.” In *Proceedings of the SIGDIAL 2013 Conference*, pp. 329–333, Metz, France. Association for Computational Linguistics.
- Kobori, T., Nakano, M., and Nakamura, T. (2016). “Small Talk Improves User Impressions of Interview Dialogue Systems.” In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 370–380, Los Angeles. Association for Computational Linguistics.
- 小磯花絵, 天谷晴香, 石本祐一, 居關友里子, 白田泰如, 柏野和佳子, 川端良子, 田中弥生, 伝康晴, 西川賢哉 (2019). 『日本語日常会話コーパス』 モニター公開版の設計と特徴. 国語研究所日常会話コーパスプロジェクト報告書 3. [H. Koiso et al. (2019). *Nihongo Nichijo Kaiwa Kopasu Monita Kokai-ban-no Sekkei-to Tokucho*. Kokugo Kenkyu-jo Nichijo-kaiwa Kopasu Projekuto Hokokusho, 3.].
- 前川喜久雄 (2004). 『日本語話し言葉コーパス』の概要. 日本語科学, **15**, pp. 111–133. [K. Maekawa (2004). *Nihongo Hanashi-kotoba Kopasu-no Gaiyoo*. *Nihongo Kagaku*, 15, pp. 111–133.].
- 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)*, pp. 8129–8141, Online. Association for Computational Linguistics.
- 三島徳雄, 久保田進也 (2003). 積極的傾聴を学ぶ—発見的体験学習法の実際. 中央労働災害防止協会. [N. Mishima and S. Kubota (2003). *Sekkyoku-teki Keicho-o Manabu—Hakken-teki Taiken Gakushu-no Zissai*. Chuo Rodo Saigai Boshi Kyokai.].
- 楡木満生 (1989). 積極的傾聴法. 医学教育, **20** (5), pp. 341–346. [M. Niregi (1989). *Sekkyoku-teki Keicho-ho*. *Medical Education*, 20 (5), pp. 341–346.].
- Okahisa, T., Tanaka, R., Kodama, T., Huang, Y. J., and Kurohashi, S. (2022). “Constructing a Culinary Interview Dialogue Corpus with Video Conferencing Tool.” In *Proceedings of the 13th Language Resources and Evaluation Conference*, pp. 3131–3139.
- Prince, E., Frader, J., and Bosk, C. (1982). “On Hedging in Physician-Physician Discourse.” In *Linguistics and the Professions. Proceedings of the 2nd Annual Delaware Symposium on*

Language Studies., pp. 83–97.

- Schegloff, E. A. (1997). “Practices and Actions: Boundary Cases of Other-initiated Repair.” *Discourse Processes*, **23** (3), pp. 499–545.
- Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). “The Preference for Self-correction in the Organization of Repair in Conversation.” *Language*, **53** (2), pp. 361–382.
- Stent, A., Stenchikova, S., and Marge, M. (2006). “Dialog Systems for Surveys: The Rate-a-course System.” In *2006 IEEE Spoken Language Technology Workshop*, pp. 210–213.
- 杉山享志朗, 吉野幸一郎, 田中宏季, 近藤真人, ニュービッググラム, 中村哲 (2016). 傾聴対話コーパスの作成と知識獲得行為の分析. 人工知能学会全国大会論文集, **JSAI2016**, p. 3. [K. Sugiyama et al. (2016). Making Listening Dialogue Corpus and Analysis of Knowledge Acquisition. Proceedings of the Annual Conference of JSAI, p. 3].
- Tolmachev, A., Kawahara, D., and Kurohashi, S. (2018). “Juman++: A Morphological Analysis Toolkit for Scriptio Continua.” In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 54–59, Brussels, Belgium. Association for Computational Linguistics.
- ホールファミリーケア協会 (2004). 傾聴ボランティアのすすめ. 三省堂. [Horu Famiri Kea Kyokai (2004). Keicho Borantia-no Susume. Sanseido.].
- 吉越卓見, 児玉貴志, 新隼人, 坂田亘, 田中リベカ, 黒橋禎夫 (2021). 主題性にもとづく雑談対話システムの構築. テクニカル・レポート 2021-NL-250-7, 研究報告自然言語処理 (NL) . [T. Yoshikoshi et al. (2021). Shudai-sei-ni Motozoku Zatudan-Taiwa Shisutem-no Kochiku. Kenkyu Hokoku Shizen-gengo Shori (NL), 2021-NL-250-7.].
- 曾傑, 中野有紀子 (2021). 知識と話題の埋め込み表現に基づく質問生成と対話システムへの適用—料理嗜好インタビューシステムに向けて—. 自然言語処理, **28** (2), pp. 598–631. [J. Zeng and Y. Nakano (2021). Question Generation in Dialogue Systems based on Embedding Representation of Knowledge and Topics: Toward Food Preference Interview System. Journal of Natural Language Processing, 28 (2), pp. 598–631.].
- Zhu, C., Liu, Y., Mei, J., and Zeng, M. (2021). “MediaSum: A Large-scale Media Interview Dataset for Dialogue Summarization.” *CoRR*, **abs/2103.06410**.

付録

A 個別練習の内容

対話タスクに初めて参加する者に対し、インタビュー収録の前に監督者が以下の点について確認した。

- 名前などの個人情報を発話しないこと
- ビデオ映像に表情が良く映るよう、カメラとの距離は適度に保たれていること
- 対話時間が15分を超えそうな場合に監督者が送信するチャットが見られること
- 対話中にトラブルが起きた場合、チャットで監督者に知らせること
- 技能者が写真を表示し、次の写真に移動することができること
- インタビュアーが進行役を務め、写真を元に最低限の作業手順を聞いて、質問を交えながら話の流れでインタビューを進めてもらうこと

また、リハーサルとして技能者に最初の写真を表示させ、冒頭部分のみを説明をさせた上で、インタビュアーが進行役として次の写真に進める手順を実際に行わせた。

B 書き起こしの方針

書き起こしの方針は以下の通りである。書き起こしの引用においては、各発話の左に付している番号は当該インタビュー全体通しての発話番号を表し、“E”は技能者の発話であることを、“I”はインタビュアーの発話であることを表す。

- (10) a. フィラーや、発話が尻つぼみになったり相手の発話と重複し途中で聞こえなくなったりした場合は、丸括弧を用いてその箇所を明示する。
- 「ブリの照り焼き」 (Interview_170: 00:01:06.150–00:01:17.500)

27 I : (んー)はい。これは(せ)背側とか腹側とか、どっちが美味しいとかありますか？

28 E : (うーん)どちらでも美味しく食べられると思うんですけど。
- b. 一方の話者が発話中の、もう一方の話者の相槌（「うん」「ふん」等）については書き起こさない。ただし、「なるほど」や「そうなんですね」のような発話がはっきりと聞き取れる場合は書き起こしの対象とする。
- 「茄子のぬか漬け」 (Interview_15: 00:02:49.533–00:3:00.243)

46 E : そうでしょうね, 多分. (あの一) 多分冷蔵庫の中で, 水分的にも温度的にもいい状態が(あの) 茄子のまわりに作られると思うんですね.

47 I : なるほど. はい. わかりました.

c. 単発の笑いについては【笑い】と書き起こし, 発話中の笑いについては笑いの情報は残さない.

- 「薔薇のアップルパイ」 (Interview_96: 00:03:12.150-00:03:21.750)

60 I : こちらのパイシートは手作りのものになりますでしょうか?

61 E : (あ) もう

62 E : スーパーに売ってる冷凍食品にある冷凍パイシートです.

63 E : 【笑い】

d. 発話内容が聞き取れない場合は「(?)」と表記する.

- 「黒豆」 (Interview_159: 00:03:40.250-00:03:54.200)

84 I : (ゆ) 丹波の黒豆とか(?)やつですか?

85 E : (あ) そうですね, それはもちろんこの丹波の黒豆が最上級ですけれども, 丹波の黒豆(しゆ)を使ったものが, (あの一) 他に北海道とか

e. 10秒を超える程度の長い発話の場合は, 書き起こしの読みやすさを考慮し, ポーズとは無関係に句読点で区切っても良い.

- 「紅茶のムースケーキ」 (Interview_165: 00:05:12.500-00:05:46.450)

54 I : ゼラチンは(こ)粉や板や何かお勧めというか

55 I : ありますか?

56 E : そうですね. (ま) 私はもう板が慣れているので(え一)板をお勧めするんですけども, 板は(ま)計りやすい, 大体1枚もう4グラムって決まっていますので計りやすいですし, お水も(え一)粉でしたら計らないといけないんですが, もう板の場合は, その板がなんていうんでしょ,

57 E : (え一と一)全部(こ一)(お)お水が

58 E : 浸ってればいいようなんです, なので(え一)計量も楽ですし扱いも楽かなと思います.

f. 言い誤りなどで正確に話せていないような場合は正表記に直して書き起こす.

なお, (10a)については, フィラーと文頭の言い直しや末尾のはっきりとは聞き取れない部分の区別が難しいため, 書き起こしの段階では同一の()で括るという方針を取っている. そのため, コーパス構築後に次のような処理を施し, 意図的なフィラーとそれ以外の言い誤りを

区別した.

- (11) a. コーパス全体を通して, 出現頻度が 10 回未満かつ長音記号を含まないものを文頭の
言い直しや末尾のはっきりとは聞き取れない部分とみなし, 〈 〉で括った
b. それ以外はフィラーとみなし, ()で括った
- 「海老の旨煮」(Interview171.:00:08:00.000– 00:08:16.250)
 - 191 E : そうですね. でも 2 分ぐらいは置いといてもらった方がいいと思
います. (あの)ちっちゃい, 大きいあると思うんですけど.
 - 192 I : (あー)はい.
 - 193 E : (で)さっと<あがっ>
 - 194 E : 上げて, その後も余熱で火が中にも入るので.

C Uber 指標の算出方法

テキスト中の総語数を N , 異なり語数を $V(N)$ とした時, Uber 指標は以下のように求める.

$$Uber = \frac{(\log N)^2}{\log N - \log V(N)}$$

略歴

岡久 太郎: 2014 年東京学芸大学教育学部中等教育教員養成課程国語専攻卒業.
2016 年京都大学大学院人間・環境学研究科修士課程修了. 2020 年同博士後期
課程研究指導認定退学. 2021 年京都大学より博士(人間・環境学)を取得.
博士(人間・環境学). 2020 年京都大学大学院人間・環境学研究科特定研究
員, 2021 年京都大学大学院情報学研究科特定研究員, 2021 年静岡大学情報学
部助教, 現在に至る. 理論言語学, コミュニケーション研究に従事. 言語処
理学会会員.

田中リベカ: 2013 年お茶の水女子大学理学部情報科学科卒業. 2015 年同大学院
博士前期課程修了. 2018 年同博士後期課程単位修得退学. 2021 年お茶の水女
子大学より博士(理学)を取得. 博士(理学). 2018 年から 2021 年まで京大
大学院情報学研究科特定研究員. 2021 年よりお茶の水女子大学文理融合
AI・データサイエンスセンター特任講師.

児玉 貴志: 2019 年京都大学工学部電気電子工学科卒業. 2020 年京都大学大学
院情報学研究科博士前期課程修了. 同年より同大学院情報学研究科博士後期
課程に進学. 現在に至る. 修士(情報学). 自然言語処理の研究に従事. 言語

処理学会会員.

Yin Jou Huang : 2015 年台湾大学卒業. 2017 年京都大学大学院情報学研究科博士前期課程修了. 博士 (情報学). 2021 年より京都大学大学院情報学研究科特定研究員. 2023 年より京都大学大学院情報学研究科特定助教. 自然言語処理の研究に従事. 言語処理学会会員.

村脇 有吾 : 2011 年京都大学大学院情報学研究科博士後期課程修了, 博士 (情報学). 同年京都大学学術情報メディアセンター特定助教, 2013 年九州大学大学院システム情報科学研究院助教, 2016 年京都大学大学院情報学研究科助教, 2020 年同講師, 現在にいたる. テキスト解析および計算言語学に関する研究に従事. 言語処理学会, 情報処理学会各会員.

黒橋 禎夫 : 1994 年京都大学大学院工学研究科博士課程修了. 博士 (工学). 2006 年 4 月京都大学大学院情報学研究科教授. 2023 年 4 月より同特定教授および国立情報学研究所長. 自然言語処理, 知識情報処理の研究に従事. 言語処理学会 10 周年記念論文賞, 同 20 周年記念論文賞, 文部科学大臣表彰科学技術賞等を受賞.

(2022 年 9 月 30 日 受付)

(2023 年 1 月 13 日 再受付)

(2023 年 2 月 20 日 採録)