

ウェブ会議システムを利用した料理インタビュー対話コーパス

岡久 太郎 田中 リベカ* 児玉 貴志 Yin Jou Huang 黒橋 禎夫

京都大学大学院情報学研究科

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

概要

我々はウェブ会議システムを用いて、料理の技能者に対するインタビュー対話のマルチモーダルコーパスを構築した。技能者として料理のセミプロと料理好きを、インタビュアーとしてインタビューの経験を持つ者と持たない者を参加者として集めた。本稿では、コーパスの構築方法と仕様について報告し、技能者からコツを引き出すためのインタビュー手法について質的に分析する。

1 はじめに

対話システムの研究においては、気の効いた設定をデザインし、良質の対話コーパスを作ることが重要であるが、インタビュー対話についてはこれまであまり取り扱われることがなかった。そこで、本研究では、熟練者の技能伝承が産業界において課題となっていることを踏まえ、技術のコツなどを聞き出すことを目的としたインタビューに着目する。既存のインタビュー対話コーパスは、テレビ番組等で第三者に対話の内容を聞かせることを目的とした対話を対象とするものが多い [1, 2, 3]。そのため、技術や知識を引き出すことができる対話システムの研究には、その基盤としてインタビュアーが積極的に対話者から情報を引き出すことを試みているインタビューデータが不可欠である。

本研究では、ウェブ会議システムを用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タスクを設定し、**料理インタビュー対話コーパス**を構築した。料理インタビュー対話コーパスは、308 対話における約 6.9 万発話の音声とその書き起こしに加えて、対話者双方の表情と対話者間で共有されているコンピューター画面を含んだ映像から構成される (図 1)。

日本語を対象とした既存の話し言葉対話コーパス

* 現在の所属は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 (連絡先:tanaka.ribeka@is.ocha.ac.jp)



図 1 料理インタビューコーパスの動画の一部とそれに対応する部分の書き起こしの例。“E”は技能者，“I”はインタビュアーを表す。

のほとんどは、電話会話を使用したものか [4]、対話者を実際に集め、その様子を録音・録画したもの [5, 6, 7, 8] 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、通常の対話とは性質が大きく異なる。また、対話者をスタジオ等に集めて対話を収録する場合は時間的、費用的コストがかかる。

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。オンラインビデオ通話であれば、対話者は自宅からいつも用いている自身の機器を用いて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、対話者はお互いの表情を見ながら、画面共有機能で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。この点は、対話研究にとっても、表情や視覚的文脈を発話と同期させて記録することが容易になるという利点となる。

本稿では、料理インタビュー対話コーパスの構築方法とその詳細について述べる。

2 インタビュー対話の収集方法

2.1 収録方法と参加者

本コーパスは、ウェブ会議システムを用いて、料理の技能者とインタビュアーの間で行われた約15分の対話を308件録画したデータから構成される。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があらかじめ用意した写真(詳細は後述)を互いに見ながら対話を行った。録音・録画に関してはZoom¹⁾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。

各対話収集の際には対話者に技能者とインタビュアーの役割をそれぞれ与えた(Appendix A 参照)。一部の参加者は、技能者とインタビュアーの両方の役割で参加している。また、技能者とインタビュアーには事前にそれぞれ2つの属性を設定した。技能者は、一定の条件(Appendix B 参照)を満たす者を「セミプロ」、それ以外の者を「料理好き」として分類した。インタビュアーは、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場合は「技能あり」、それ以外の者を「技能なし」と分類した²⁾。

なお、対話データの収集については最初に20対話、暫定的な条件設定で予備収集を行い、その結果を受け、最終的な条件を定め、本収集を行った。

2.2 参加者への教示及びアンケート

今回の対話タスクでは、技能者から料理の技能を効果的に聞き出すような質問をインタビュアーにさせることが目的であるため、タスクの事前準備として技能者とインタビュアーに以下の指示を行った。

まず、技能者として参加する者には事前に以下の4つについて準備し、提出するよう求めた。

- (1) a. 紹介する料理のタイトル
- b. 紹介する料理の概要

1) <https://zoom.us>

2) インタビュアーは自炊経験があるような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。

- c. 6–10枚の手順写真
- d. 自慢したい点とそれに対応する写真番号(本収集のみ)

一方、インタビュアーとして参加する者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。インタビュアーには、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せる質問をするようあらかじめ指示した。

さらに、予備収集の結果を受け、本収集では事前に「良い質問のタイプ」をインタビュアーに教示し、1つ以上の質問内容とその質問を何枚目の写真でするつもりかについて事前に報告を求めた。「良い質問のタイプ」は、「技能者が答えにつまりそうな質問」について11個の項目を例と共に示した資料である(Appendix C 参照)。

また、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。

- (2) a. コミュニケーションがスムーズだったか
- b. コツをうまく引き出していたか
- c. 自身がどの程度料理の知識や技能を持っているか(インタビュアーのみ)
- d. インタビューの感想

(2a), (2b), (2c)の回答は「1. そう思わない」「2. あまりそう思わない」「3. どちらでもない」「4. ややそう思う」「5. そう思う」のいずれかを選択させた。また、(2d)については自由記述での回答を求めた。

3 収集したインタビュー対話の概要

3.1 発話内容に関する統計情報

構築した料理インタビュー対話コーパスの統計情報は表1の通りである。

1発話あたりの平均単語数は技能者が10.13語、インタビュアーが6.98語であり、1対話当たりの平均発話数は技能者が241.23、インタビュアーが207.38であることから、技能者はインタビュアーよりも長い発話をより多く行っていることが分かる。この傾向は、対話参加者が技能者とインタビュアーの役割を適切に果たしていることを示している。

また、技能者とインタビュアーの属性ごとに統計情報を詳細化した結果は表2のようになる。技能者は、対話相手の属性に関わらず、セミプロ(E_H)の

表 1 料理インタビュー対話コーパスの統計情報. フィラーや言い誤りを取り除き, Juman++ [9] を用いて形態素解析を行った.

	技能者	インタビュアー
1 対話あたりの平均発話数	241.23	207.38
1 発話あたりの平均単語数	10.13	6.98
1 発話あたりの平均継続時間	3.79 s	2.66 s

表 2 技能者とインタビュアーの属性ごとの統計情報. 技能者については E_H がセミプロ, E_L が料理好きを表し, インタビュアーについては I_H がインタビュー技能のあるインタビュアーを, I_L がインタビュー技能のないインタビュアーを表している.

	$E_H&I_H$		$E_H&I_L$		$E_L&I_H$		$E_L&I_L$	
	E_H	I_H	E_H	I_L	E_L	I_H	E_L	I_L
1 分あたりの平均発話数	8.8	7.7	9.0	7.4	10.3	9.1	10.1	8.8
1 発話あたりの平均単語数	11.7	7.6	11.5	6.4	8.7	7.3	8.8	6.7

方が料理好き (E_L) よりも 1 分あたりの発話数が少ない. また, 技能者の 1 発話あたりの単語数を見ると, インタビュアーの属性に関わらず, セミプロの方が料理好きよりも長い発話を行っている. これらを踏まえると, 技能者はインタビュアーの属性に関わらず, セミプロが料理好きよりも多くの内容を語っているために, 時間あたりの発話数が少なくなっていることが分かる.

一方, インタビュアーについては, その属性に関わらず, 対話相手が料理好きの場合 ($E_L&I_H$, $E_L&I_L$) の方が, 対話相手がセミプロの場合 ($E_H&I_H$, $E_H&I_L$) よりも 1 分あたりの発話数が多くなっている. しかしながら, それぞれの対話設定における 1 発話あたりの単語数については, 対話相手の属性に関わらず, インタビュー技能があるインタビュアー (I_H) の方が, 技能がないインタビュアー (I_L) よりも 1 発話あたりの単語数は多い. 以上から, 一般的な傾向として, インタビュー技能を有するインタビュアーの方が, インタビュー技能がないインタビュアーよりも内容の多い発話をしているが, 技能者の料理技能が低い場合は, インタビューの技能に関わらず, 技能者から多くの話を引き出すために多くの発話を行う傾向があることが分かる.

3.2 対話後のアンケート結果

図 2, 3 より, 技能者もインタビュアーも参加したインタビューについて概ね満足していることが分かる. 「コミュニケーションはスムーズだったか」という問いについては, 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.

また, 「コツを引き出していたか」という問いについては, 技能者の平均値が 4.62 であり, インタ

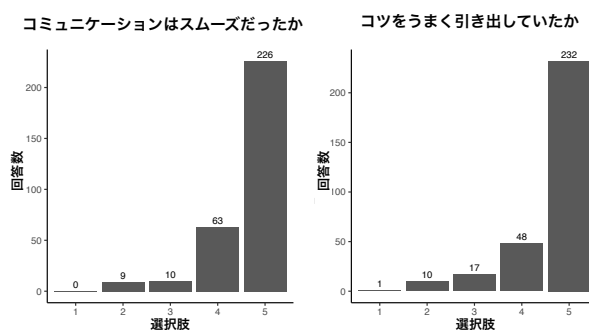


図 2 技能者に対するアンケート結果

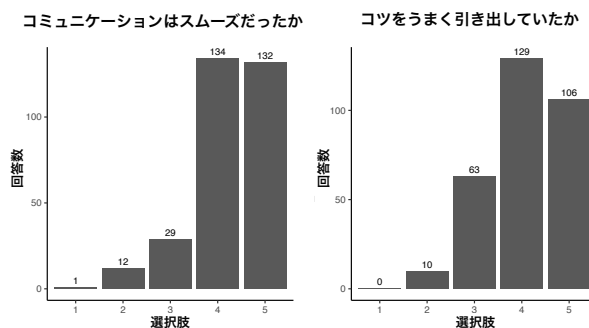


図 3 技能者に対するアンケート結果

ビュアーの平均値が 4.07 であった. この差は, インタビュアーが技能者からコツを聞き出すようあらかじめ指示されていたことにより, インタビューの出来について技能者以上に厳しい判定をしたことにより生じた差異であると考えられる.

なお, インタビュアーの料理の知識については平均値が 3.52 であった. これは, 最低限の料理の知識を持つことをインタビュアーの参加条件としたことにより, 知識があると回答する参加者が比較的多かったためであると考えられる.

4 コツの引き出し方

本研究で収集した対話データは、インタビュアーが技能者から料理のコツを聞き出すようにタスクを設定した。そのため、インタビュアーの発話が契機となり、技能者が当初は話すことを予定していなかった知識が引き出される場合が少なくない。本節ではコツの引き出し方の典型的パターンを事例ベースで説明する。

技能者からコツを引き出す際によく見られるインタビュアーの発話として、直前に技能者が発言した内容を繰り返したり、言い換えたりするというものが挙げられる。

(3) 「セロリのぬか漬け」

- E: (えーと) 塩と。これは(あの一)粗塩というか、(あの一)サラサラした食卓塩っていうよりは、(あの一)結構(あの一)ザラザラゴツゴツとした結構甘みのある塩です。
- I: 粗塩、はい。
- I: (あ)味が(じゃ)しっかりとしてるような感じですか？
- E: そうですね。もう<これなめ>これだけ舐めても美味しい感じの塩です。はい。(まーね)(あの一)なんでも塩好きなの使っていたらいいんですけど、(ま)塩だけで美味しいっていうのは結構、(あの一)材料少ないんでね、(あの一)美味しい塩でされたらいいかなと思います。

例3では、技能者の発言にある「甘みがある」をインタビュアーが「味がしっかりとしている」と言い換えたことにより、粗塩を使う理由を技能者が追加で説明している。

このような質問の仕方は、インタビュアーに料理の経験が十分でない場合も可能なものである。一方、インタビュアー自身にも料理の経験があるからこそ可能となるコツを引き出す質問の特徴として、インタビュアーが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パターンが挙げられる。

(4) 「生シラスのふっくらかき揚げ」

- I: (ふーん)(で)何か(この)揚げているときのコツとかがありますか？
- E: (あ)。
- E: 揚げているときのコツですね。
- E: (えーと)
- E: 揚げているときの(こ)。

- I: 例えば、私、よく(あの一)何か(こ一)心配で、できているのかなって、(こ一)菜箸でぐしゃぐしゃぐしゃ、
- I: やっぱりしてしまうんですけど、(な)(な)(な)。
- E: (？)³⁾。
- E: そうですね。つまり、なるべく(あの一)なるべく触らないように。
- I: (あ一)やっぱりそうですか。
- E: (？)。
- 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(え一)じっくり火を通して、また裏返して、その後も(あの一)ほとんど触らないように、(え一)形がくずれないように揚げていって、(で)最終的に(あの一)
- E: (えーと)バットに上げるときに、油をしっかり(こ一)
- E: 切るというのが、(あの一)カリッとふわっと仕上がるコツかなと思いますので。

例4では、最初に漠然とコツを尋ねられた技能者その後約6秒間回答に詰まっている。それを受けたインタビュアーは下線部のように自身の経験談を語ることで、天ぷらを揚げる時のコツを引き出している。このように、技能者が質問の回答に窮している時に、インタビュアーが具体的な状況を提示することで、技能者は自身では事前に意識していなかった知識や技能についてスムーズに話すことが可能となる例が多く見られた。

これらの質問の仕方については、事前にインタビュアーに提示した「良い質問のタイプ」(Appendix C 参照)には含まれていない方略であり、インタビュアー役を務める対話システムの研究において、重要な要素となりうる。

5 おわりに

本稿では、ウェブ会議システムを用いた料理インタビュー対話コーパスの構築方法とその詳細について報告した。

コーパスの質的分析の結果、技能者のコツを引き出す場面においては、言い換えや繰り返しを行うパターンとインタビュアー自身の経験をもとに具体的な状況を想定した質問を行うパターンの2つが見出された。これらのパターンをどのように使い分けるか、どのように対話システムに組み込むのかについては、今後の課題である。

3) 対話書き起こし中の(？)は聞き取れなかった箇所を表す。

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。

参考文献

- [1] Bodhisattwa Prasad Majumder, Shuyang Li, Jianmo Ni, and Julian McAuley. 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, November 2020. Association for Computational Linguistics.
- [2] Chenguang Zhu, Yang Liu, Jie Mei, and Michael Zeng. Mediasum: A large-scale media interview dataset for dialogue summarization. **CoRR**, Vol. abs/2103.06410, , 2021.
- [3] 篠山学, 松本和幸. インタビュー対話コーパスの構築と発話意図タグの設計. 第 82 回全国大会講演論文集, 第 2020 巻, pp. 51–52, feb 2020.
- [4] 伝康晴, フライジョン. 日本語 callhome コーパス. 音声研究, Vol. 4, No. 2, pp. 24–30, 2000.
- [5] 前川喜久雄. 『日本語話し言葉コーパス』の概要. 日本語科学, Vol. 15, pp. 111–133, 2004.
- [6] Yasuharu Den and Mika Enomoto. A scientific approach to conversational informatics: Description, analysis, and modeling of human conversation. pp. 307–330. John Wiley & Sons, Hoboken, NJ, 2007.
- [7] Itsuko Fujimura, Shoju Chiba, and Mieko Ohso. 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**, p. 393. Speech and Corpora, 2012.
- [8] 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. 『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, 2019.
- [9] Arseny Tolmachev, Daisuke Kawahara, and Sadao Kurohashi. 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, November 2018. Association for Computational Linguistics.

A タスク参加者数の詳細

表6 参加者の役割・属性ごとの内訳. I, Eはそれぞれインタビュアー, 技能者を指す. また×はその役割では不参加だったことを表す.

	E	セミプロ	料理好き	×
I				
技能あり		2	1	5
技能なし		3	2	4
×		3	6	-

B 「料理のセミプロ」の条件

以下の3条件のいずれかを満たす技能者を「料理のセミプロ」として分類した.

- 料理教室等で料理（種類は問わず）を教えた経験がある
- 調理師免許を取得している
- 料理家として活動している

C 良い質問のタイプ

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
 - 「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
 - 「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」
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, もう一步踏み込んで, それが良いことなのかを質問する
 - I: 「ネギは, 青いネギじゃなくて白いネギですね」
E: 「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」

I: 「その方が美味しいのですか」

7. ポイントを率直に尋ねる
 - 「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. 避けた方が良くいことや, 失敗経験を質問する
 - E: 「鯛の代わりに, 他の白身魚やホタテでも良いですよ」
I: 「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