

技能者インタビュー対話コーパス (EIDC) v.2.0: コツ発話の同定に向けた相互行為アノテーション

近大志¹ 岡久太郎² Yin Jou Huang¹ 樽谷洋希¹
松田 思鵬¹ 村脇有吾¹ 黒橋禎夫¹
¹ 京都大学 ² 静岡大学

chika.taishi.6p@kyoto-u.ac.jp okahisa-taro@inf.shizuoka.ac.jp
{huang, tarutani, matta, murawaki, kuro}@nlp.ist.i.kyoto-u.ac.jp

概要

技能者インタビュー対話コーパス (Expert Interview Dialogue Corpus; EIDC) は、技能のコツを含む発話 (コツ発話) の性質や、効果的な引き出し方を解明する目的で構築されている。EIDC は園芸・料理ドメインの技能者インタビュー対話を対象とし、映像・音声、書き起こし、各種アノテーションを収録している。EIDC v.2.0 では、コーパスの一部を対象として、対話の相互行為的な特徴に着目した追加アノテーションを行った。本稿では、追加アノテーションの仕様と定量的特徴を報告し、コツ発話の検出に寄与する特徴を分析する。

1 はじめに

熟練した技能者はレシピや仕様書の指示を単に遂行できるだけでなく、動作で注意すべきポイントや次の動作に移行する基準等、技能の背後にあるコツを会得している。筆者らは、技能に関するインタビューの中で技能者が発する**コツ発話**の性質や、効果的な引き出し方を解明する目的で研究を行っている。コツ発話とは、初心者等が理解していない可能性のある作業における重要なポイントや発展的な事項を含む発話を指す。この目的のもと、**技能者インタビュー対話コーパス (Expert Interview Dialogue Corpus; EIDC)** を構築した。¹⁾

EIDC v.1.0 では、対話で示された技能の知識構造を意味フレームに基づいて記述し、コツ発話をアノテーションした。しかし、コツ発話は技能者が暗黙的に実践しているノウハウを含み、必ずしも言語化が容易ではないと考えられる。このことは、例えば、インタビュアーの鋭い質問に対して技能者が答

えに窮し、言い淀み (フィラー) を挟みつつ、ジェスチャーを交えながら返答する行動に反映される。コツ発話を検出するモデルを設計し、それを効果的に引き出す対話システムを開発するには、技能の知識構造やコツ発話単体だけではなく、話者がやりとりの中で発する多面的な情報が必要となる。

このような背景のもと、EIDC v.2.0 ではコーパスの一部に対して発話意図とパラ言語情報のアノテーションを行った。本稿の目的は、EIDC v.2.0 の仕様とその定量・定性的特徴を報告し、コツ発話の検出につながる特徴を分析することにある。分析から、発話意図の特定の連鎖や、ジェスチャーやフィラーといったパラ言語情報がコツ発話の予測にとって有用な特徴であることが示唆される [1]。

2 EIDC

2.1 データセット

EIDC は料理・園芸ドメインの技能者インタビュー対話を扱う。技能者は納豆餃子の作り方、ノウセンカズラの剪定といった特定の作業を段階ごとに解説し、インタビュアーは適切な納豆の量、茎を手でちぎってはならない理由等、技能の様々な側面に関して質問を投げかける。インタビューはオンライン通話システム (Zoom) 上で実施し、共有画面にある写真を参照しながら対話を進める (図 1)。EIDC では次のデータを収録している: (a) インタビュー対話の音声・動画ファイル, (b) インタビュー対話の書き起こしテキスト, (c) および各種アノテーションデータ。(a-b) については 1 対話あたり 10–15 分程度の対話を (i) 料理 308 対話, (ii) 園芸 100 対話分収録しており, (c) の対象はアノテーションごとに異なる。書き起こしに関する仕様については, [2] を参照されたい。

1) <https://nlp.ist.i.kyoto-u.ac.jp/?EIDC>

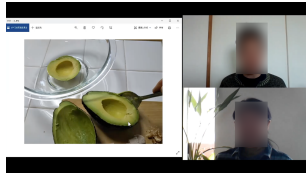


図1 インタビュー動画例

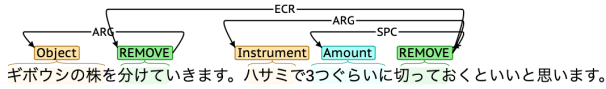


図2 Bratによるアノテーションの例

2.2 v.1.0のアノテーション

EIDC v.1.0では、(i) コツ発話アノテーション (特定の発話がコツ発話であるかどうか)、(ii) 意味フレームアノテーション (意味フレームに基づく技能の知識構造) を実施した。アノテーション対象は料理・園芸ドメインのすべての対話データであり、タグ付け作業には Brat²⁾ を用いた (図 2)。

コツ発話アノテーション コツ発話とは、初心者等は理解していないと思われる作業における重要なポイント・発展的な事項を含む発話を指す。5人のアノテーターが以下の手順で当該の発話行がコツ発話に相当するかどうかをタグ付けした。

1. コツを含むと思われる発話の発話番号に Key_Utterance タグを付ける
2. Key_Utterance タグを付与した発話に関して、当該のコツがインタビュアーとのやりとりを通して引き出されたと判断できる場合に属性タグ Interaction を付与する

- (1) インタビュアー: それを (こう) 違う方向に無理やりやっても大丈夫なんですか?
 技能者: (あ) はい。特に (あの一) (ま一) ちぎれないように気をつければ大丈夫です。朝顔の場合は (あの一) 反時計巻きに巻く習性があるので ... (Key_Utterance_Interaction)

表 1 に、ドメインごとにみたコツ発話の頻度を示す。なお、インタビュアー対話の冒頭 10-20 ターン前後は完成物の味や見た目等、作業には直接関わらないトピックを話し合うことが多い。

意味フレームアノテーション 対話内で明らかとなった特定技能の知識を構造化するため、フレーム意味論 [3, 4] の枠組みに準じたアノテーションを実施した。意味フレームは基本的に動詞によって喚起

2) <https://brat.nlplab.org/>

表 1 ドメインごとにみたコツ発話の頻度

	-Interaction	Interaction
料理 (69086 発話)	1514	5574
園芸 (16336 発話)	546	1210

される、個々のイベントに関する複合的知識を指し、その設定は各ドメインに依存する。フレームタイプと依存関係にある項や付加詞はフレーム要素・指定要素として設定しており、これらは基本的にドメイン間で共通している。他にも、同一のイベントを表す述語同士を結ぶ共参照タグや、技能者が避けるべき事項として述べている要素等に属性タグを付与している。アノテーションの仕様および分析については [5] を参照されたい。

3 EIDC ver. 2.0のアノテーション

3.1 発話意図アノテーション

相互行為や発話行為の観点からみて、対話に参加する話者は発話を用いて特定の談話行為 (Dialogue Act) を遂行しており、その背後には特定の発話意図 (Dialogue Intent) がある [6]。インタビューで頻繁に起きる質問と返答の対は単なる一問一答ではなく、質問に先行する内容確認や、返答に後続する情報の補足等、複雑な構造を有している。対話内で展開されやすい発話意図の連鎖を明らかにするため、発話意図をアノテーションした。

3.1.1 アノテーションの仕様

タグ付け対象 料理ドメイン 15 対話と園芸ドメイン 10 対話にアノテーションを実施した。2名のアノテーターが Brat を用いて各自の担当範囲をタグ付けした。なお、一致度を確認する目的で料理ドメインの 9 対話分は担当が共通している (Cohen's Kappa: $\kappa = 0.808$)。

タグセット ISO 24617-2 で規格化されている仕様 [6] を基に、主要な発話意図を (i) 情報交換, (ii) 共通認識の確認, (iii) 対話の管理, (iv) 社交的な行為の 4 つに大別し、下位分類として個別のタグを設定した。(2) はタグ付けの例である。

- (2) 技能者: ウィンナーを切っている工程になっています。(Statement)...

インタビュアー: ウィンナーは結構細かく切ってるんですね。(Request-Info)

インタビュアー: やはり、細かめのほうがよ

ろしいでしょうか? (Opinion)

技能者: はい、そうですね。 (Answer)

上記の例において、 Weiner を切る工程に関する技能者の情報提供 (Statement) を受けたインタビュアーは画面上の写真を参照し、細かく切っているかどうかを尋ね (Request-Info)、細かい方がよいという意見 (Opinion) を表明したのち、技能者が「そうですね」と肯定的に返答 (Answer) している。

単一の発話意図が複数の発話から構成される場合もあれば、複数の発話意図が単一の発話内に生起する場合もある。そのため、ラベリング対象となるテキストの範囲は1つの発話意図を担う箇所として柔軟に設定しており、これは、長い発話単位 [7] に対応する。また、発話意図間の関係もタグ付けしており、意味内容の連続 (Discourse) や、ある行為に対する応答 (Reactive) 等を記録している。主要な発話意図タグの定義については、付録 A を参照されたい。

3.1.2 分析

定量的特徴 タグの分布をドメインごとにまとめたものを図 4 に示す。どちらのドメインでも、おおむね同じ傾向が見られる。ただし、料理ドメインでは1つのインタビューで取り上げる調理工程が園芸ドメインより多く、相手の行動を促す発話 (Prompt-Act) の頻度が園芸ドメインに比べて高い。

コツ発話との関係 インタビュアー対話においては、技能者の返答 (Answer) に対してインタビュアーが当該トピックについて質問 (Request-Info) を深掘りし、技能者が自らの作業を振り返りつつ、再度返答を与える場面がある。

(3) インタビュアー: これはボールに入っているのは (あ) 原材料、何になりますでしょうか? (Request-Info)

技能者: ボールの中には、(えー) バター、粉糖、塩、(えー) 卵、そして (えー) 粉類ですね。 (Answer)

インタビュアー: これは (えー) その材料を混ぜる順番っていいのは、どういった順番で混ぜ合わせておられますでしょうか? (Request-Info)

技能者: はい。(え) まずバター、クリーム状にさせていただいて、粉糖入れて、(えー) よく攪拌していただきます、... (Answer,

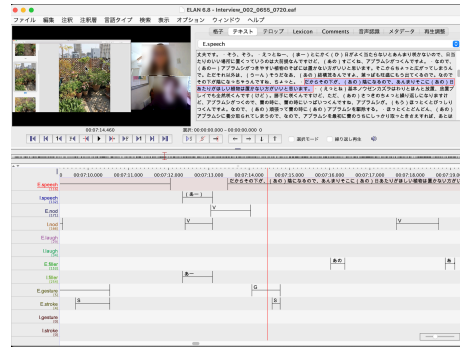


図 3 ELAN によるパラ言語アノテーションの例

(Key_Utterance_Interaction)

(3) の例において、「(ケーキの生地をつくるために) 何を混ぜるのか」という質問に対する返答を受け、インタビュアーは混ぜ合わせの順番を再度尋ねている。コツ発話は明言が容易ではない技能の側面を指すことから、(3) にみられる質問-返答からなる連鎖の反復にコツ発話が生じる可能性がある。

3.2 パラ言語アノテーション

パラ言語は伝達に伴う周辺の要素ではなく、円滑な意図伝達にとって欠かせない情報の1つである。回答が難しい質問に対して、技能者は答えに悩みつつ、フィラーやジェスチャーなどの様々なパラ言語情報 [8, 9, 10] を示しながら、その回答を構築することがある。コツ発話、およびコツ発話周辺の発話にみられる特徴を明らかにするため、パラ言語情報をアノテーションした。

3.2.1 アノテーションの仕様

タグ付け対象 アノテーション対象は料理ドメイン 20 対話と料理ドメイン 14 対話であり、作業においては、対話映像と書き起こしテキストのタイムスパンを同期させるために ELAN³⁾ を用いた (図 3)。

タグセット パラ言語情報として、(i) 頭の動き (nod), (ii) 笑い (laugh), (iii) フィラー・相槌 (filler), (iv) ジェスチャー (gesture) をアノテーションした。なお、ジェスチャーを行う一連の身体動作のうち、核となる要素を指してストロークと呼ぶ [10]。そのため、ジェスチャーについては、ストローク部分 (stroke) とそれ以外を区別している。各タグの詳細については付録 B を参照されたい。

3) <https://archive.mpi.nl/tla/elan>

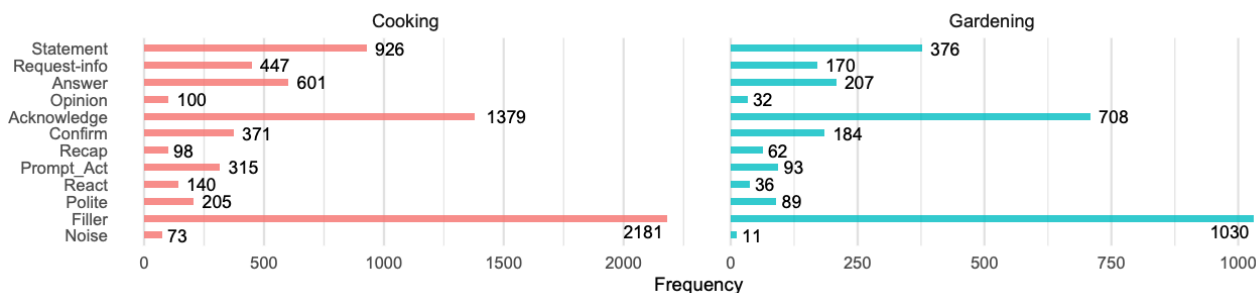


図4 発話意図言語タグの分布

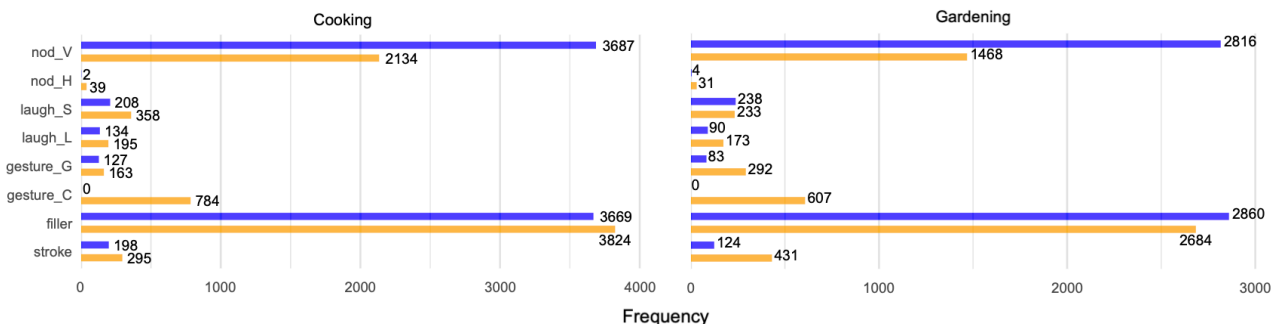


図5 パラ言語タグの分布. 青色がインタビュアーを指し、オレンジ色が技能者を表す.

3.2.2 分析

定量的特徴 定量的特徴として、ドメインごとのタグの分布を図5に示す。料理・園芸どちらのドメインでもパラ言語の使用傾向がおおむね共通していることが分かる。肯定的なうなずきを表す頭の縦の動き (nod_V) がインタビュアーの方が高頻度であるのは、インタビュー対話の性質上、インタビュアーが聞き役に徹する機会が多いためである。

コツ発話との関連 コツ発話とパラ言語情報の関係を探る上で、コツ発話内に用いるパラ言語情報を調べるアプローチと、コツ発話周辺の発話に伴うパラ言語情報を調べるアプローチが考えられる。前者のアプローチからは、例えば、コツに相当する箇所を強調するために行われている、以下のようなジェスチャーが観察される (S 付きの括弧はストローク部分)。

- (4) 技能者: この後はもう s_1 [このまま 10] s_2 [分間放置] になります。

上記の例では、S1で右手のひらを前方に向け静止させるジェスチャーを行い、続くS2でその手を前方に2回振るジェスチャーを行っている。

後者のアプローチに対しては、技能者によるコツ発話に対して納得したことを示すために、インタビュアーが複数回のフィラーを用いる例が挙げら

れる。

- (5) 技能者: 水にさらず時間をもうぐーっと少なくすることができるので、この点はちょっとお勧めですね。

インタビュアー: はーはー(あー)なるほど。

ただし、コツ発話に限らずジェスチャー・フィラーを頻繁に用いる話者もいれば、選択的に用いる話者もいる。そのため、パラ言語の頻度とコツ発話をより厳密に関連付ける際には、話者ごとの個人差を考慮する必要があるだろう。

4 結論

技能者インタビュー対話コーパス (Expert Interview Dialogue Corpus; EIDC) v.1.0では、料理・園芸ドメインの知識構造を意味フレームの観点から明示し、コツ発話をアノテーションした。コツ発話は明言が難しく、暗黙的なノウハウが含まれる。コツ発話を検出するモデルを設計し、コツ発話を効果的に引き出すようなインタビュー手法を構築するには、発話意図やパラ言語情報等の多面的な情報を基に、やりとりの中で知識構造が提示される方略を明らかにする必要が生じる。EIDC v.2.0では、パラ言語情報や発話意図の追加アノテーションを実施した。新たに加えた対話上の特徴を活用することで、コツ発話の検出につながることを期待される [1]。

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。ここに感謝いたします。

参考文献

- [1] 樽谷洋希, 近大志, 岡久太郎, Yin Jou Huang, 松田思鵬, 村脇有吾, 黒橋禎夫. 技能者インタビュー対話におけるコツ発話の表出に至った発話列の特徴の分析. 言語処理学会第 31 回年次大会 発表論文集, 2025. (to appear).
- [2] 岡久太郎, 田中リベカ, 児玉貴志, Yin Jou Huang, 村脇有吾, 黒橋禎夫. 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. 自然言語処理, Vol. 30, No. 2, pp. 773–799, 2023.
- [3] Charles J. Fillmore. Pragmatically controlled zero anaphora. In **Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society**, pp. 95–107, 1986.
- [4] Josef Ruppenhofer, Michael Ellsworth, Myriam Schwarzer-Petruck, Christopher R Johnson, and Jan Scheffczyk. FrameNet II: Extended theory and practice. Technical report, International Computer Science Institute, 2016.
- [5] Taishi Chika, Taro Okahisa, Takashi Kodama, Yin Jou Huang, Yugo Murawaki, and Sadao Kurohashi. Domain transferable semantic frames for expert interview dialogues. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, pp. 5299–5308, Torino, Italia, May 2024. ELRA and ICCL.
- [6] Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum. Towards an ISO standard for dialogue act annotation. In **Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)**, Valletta, Malta, May 2010. European Language Resources Association (ELRA).
- [7] Japanese Discourse Research Initiative. 発話単位ラベリングマニュアル version 2.0.
- [8] 小磯花絵, 西川賢哉, 間淵洋子. 転記テキスト. 日本語話し言葉コーパスの構築法, pp. 23–132. 国立国語研究所, 2006.
- [9] 泉子 K メイナード. 会話分析. くろしお出版, 1993.
- [10] 細馬宏通. 非言語コミュニケーション研究のための分析単位: ジェスチャー単位. 人工知能, Vol. 23, No. 3, pp. 390–396, 2008. https://www.jstage.jst.go.jp/article/jjsai/23/3/23_390/_article/-char/ja/.

A 発話意図のタグ

情報交換 発話によって情報を交換し、自らの意見を表明する行為を指す。

- **Statement** 作業工程の情報を提供する発話、および、質問の応答に対する前提や補足情報に相当する発話断片に付与する。
- **Request-Info** 情報を求める、または引き出す発話に付与する。典型的には疑問文が多いが、他の言語形式によって表現される場合もある。
- **Answer** 相手の質問に対する応答で中心的な返答内容となる発話断片に付与する。
- **Opinion** 相手の発言内容に対する、自分の意見や解釈を述べる発話に対して付与する。典型的には発話者が提示した技能に関する感想が相当する。インタビュー対話はオンライン通話システム (Zoom) で行っているため、技能者が画面に共有した写真に対するコメント等 (e.g. おいしそうですね) も Opinion とみなしている。

共通認識の確認 対話の内容について、相手との共通認識を高めたり確認したりする行為を指す。

- **Acknowledge** 発言内容に対する理解を示す発話、および発言への傾聴を示す相槌に対して付与する。
- **Confirm** 相手の発言内容を確認する発話に対して付与する。
- **Recap** 相手の理解を促進する目的で自身の発言内容を繰り返す発話部分、または言い換えた部分に付与する。

対話の管理 対話の流れをコントロールする行為を指す。

- **Prompt-Act** 相手の行動を促す発話、または自分が何かを行うように提案する発話に対して付与する。
- **React** Prompt-Act 相当の発話が表す提案や申し出を受け入れる、あるいは拒否するような発話に対して付与する。

社交的な行為 社交性を示す行為を指す。

- **Polite** 社交性を示す行為を指す。例えば、挨拶や感謝を述べるといった儀礼的な発話に付与する。

B パラ言語情報のタグ

頭の動き (nod) 相手の発言に対する相槌や肯定・否定の態度を示すために発話者が行っている頭の動きを指す (E/I.nod)。反応を区別するため、頭の動きの方向を縦 (E/I.nod.V) と横 (E/I.nod.H) に区別している。本アノテーションでは、複数回繰り返される頭の動きであっても、一連の動きがひとまとまりの機能を担うと解釈できる動作については単一のアノテーションを施している。

笑い (laugh) 技能者・インタビュアーの笑い (E/I.laugh) をタグ付けする際、笑い声が伴った笑い (E/I.laugh.L) と、笑い声を伴わないが瞬時に表情が笑顔に変化した笑い (E/I.laugh.S) に区別している。なお、笑い声を伴う笑いの一部は書き起こしのテキストに含まれていないケースがあり、その場合はアノテーターがテキストを編集している。

フィラー・相槌 (filler) フィラーとは「えー」や「あー」に代表される「言い淀み時等に出現する場繋ぎ的な表現」を指し [8]、相槌は「話し手が発話権を行使している間に聞き手が送る発話権の移動を伴わない発話」と定義される [9]。「えー」「はい」等、言語化されるフィラー・相槌の情報を E/I.filler の注釈層に記録している。相槌については、「はい」や「なるほど」のような発話だけでなく、相手の発話中に見られる「本当に」のような内容語を含む言語情報もアノテーション対象とし、これらを E/I.filler とタグ付けしている。

ジェスチャー (gesture) 意図的な手の動きを行っている部分に対して E/I.gesture をタグ付けしている。インタビューでは Zoom を用いており、技能者は言葉だけでなく、互いが共有している画面に対してマウスカーソルで指示をしたり、カメラに映るように身振りを交えたりすることで意図の伝達を実現するケースがある。そのため、意図的なジェスチャー (E/I.gesture_G) と、カーソルを用いた意図伝達 (E/I.gesture_C) を区別している。

また、先行研究においては、ジェスチャーを準備・ストローク・復帰という一連の段階に分割した上で、核となる動作をストロークと呼んでいる [10]。そのため、意図的なジェスチャーが生じた部分のうち、ストロークの開始と終了までを (E/I.stroke) としてタグ付けしている。