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

樽谷洋希 [1]　Yin Jou Huang [1]　松田思鵬 [1]
村脇有吾 [1]　黒橋禎夫 [1]　近大志 [1]　岡久太郎 [2]
[1] 京都大学　[2] 静岡大学
{tarutani, huang, matta, murawaki, kuro}@nlp.ist.i.kyoto-u.ac.jp
chika.taishi.6p@kyoto-u.ac.jp　okahisa-taro@inf.shizuoka.ac.jp

概要

筆者らは技能者からその技能について聞き出すインタビューを収集した技能者インタビュー対話コーパスを構築し，このコーパスを用いてコツを含む発話 (コツ発話) を効果的に引き出すインタビューの特徴を解明することを目的として研究を行っている．本論文はこの技能者インタビュー対話コーパスにおける多面的なアノテーションを用いて，コツ発話の表出を予測するロジスティック回帰モデルを構築し，その予測精度とモデルの特徴，またそこから考察されるコツ発話の表出に関わるインタビューの特徴について報告する．

1　はじめに

特定のドメインの技能者は，その技能ドメインにおける一般に知られていないコツを持っていると考えられる．筆者らは技能者インタビューという，特定の技能ドメインの技能者からその技能について聞き出すインタビュー形式の対話に着目している．インタビュー形式の対話によってそのコツを引き出す方法，またはコツが引き出される状況とはどのようなものかが解明されれば，コツを明文化して後世に残しやすくなると筆者らは期待している．

技能者が非熟練者にコツを伝達する方法に着目した研究は相互行為研究 [1, 2, 3, 4] や教育心理学 [5, 6] の分野において行なわれてきたが，自然言語処理や言語学の観点からの研究は少数である．そこで筆者らは複数の技能者インタビュー動画を収録し，音声を書き起こし，それらにインタビューの特徴を示す多面的なアノテーションを施した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 [7, 8, 9]．EIDC は，料理，園芸の 2 つのドメインにおける技能者・インタビュアーの 2 者間のインタビュー対話を収録している．

本論文では，EIDC に付与された多面的アノテーションを使って，コツ発話の表出に関わるインタビューの特徴を分析する．具体的には，注目する発話と先行する文脈を見て，その発話がコツ発話かどうかを二値分類するタスクを設定する．分類のための手がかりとして多面的アノテーションから得られる各種特徴を用い，ロジスティック回帰モデルを訓練する．最後に，回帰モデルの予測精度やパラメータから，コツ発話の表出に関わる特徴を考察する．

実験の結果，主に以下のような特徴を持つ場合にコツ発話が表出しやすいことがわかった．(1) Purpose，Manner，Condition といった知識構造に関わるタグが当該発話に含まれる場合，(2) 質問に対する回答，(3) 直近でフィラーや頷きなどが多く表出している場合．

2　技能者インタビュー対話コーパス

データセット　技能者インタビュー対話コーパス (EIDC) には，料理ドメインの 313 回分のインタビュー，園芸ドメインの 100 回分のインタビューが，動画，書き起こしテキストの形で収録されている [7]．書き起こしテキストは発話単位 (以下，発話とよぶ) という単位で区切られている．

書き起こしテキスト上にコツ発話アノテーション，意味フレームアノテーション，発話意図アノテーションの 3 種類のアノテーションがなされており，動画上にパラ言語アノテーションがなされている．以下に各アノテーションの概要を示す．詳しくは文献 [8, 9] を参照されたい．

コツ発話アノテーション　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**コツ発話**とよび，当該発話に Key_Utterance というタグを付与している．

(1) インタビュアー: わかりました. これはかき混ぜたりとかしますか？
技能者: （えーと）最初はちょっとかき混ぜないですね.（あのー）（さい）（あのー）お箸を使ってかき混ぜるというよりは，フライパンをちょっと揺らす感じ ... (Key_Utterance)

意味フレームアノテーション　意味フレームアノテーションのタグは特定のドメインに頻出の述語を表す**フレームタイプ**，特定のフレームタイプにおいて重要な要素である**フレーム要素**，他のタグの詳細を表す**指定要素**からなる．これらのタグは，発話内の該当するテキストスパンに付与される．タグのペアが動作/動作対象，動作/動作の詳細などの関係 (以下，Relation) を持つ場合，タグ同士を結び付ける．また，同じ動作を表しているフレームタイプ同士はイベント共参照の関係づけを行う．意味フレームアノテーションのタグの詳細は付録 A を参照されたい．

発話意図アノテーション　発話意図アノテーションは，情報を求める，確認するといったコミュニケーション上の意図をタグ付けしたものである．タグ付けは特定の発話意図を担う言語単位に対して行い，一般に発話は複数の発話意図単位からなる．ただし，ある発話が他の発話への反応である場合や，他の発話に付随する内容である場合はそれらを示す関係タグを付与する．発話意図アノテーションのタグの詳細は付録 B を参照されたい．

パラ言語アノテーション　パラ言語アノテーションのタグはインタビュー内での身体的動作を示すものである．各パラ言語アノテーションのタグに対応する時間軸上のスパンにアノテーションがなされている．パラ言語アノテーションのタグの詳細は付録 C を参照されたい．

3　提案手法

3.1　ロジスティック回帰モデル

ロジスティック回帰モデルは，説明変数 x_k とその係数 β_k，定数 α，目的変数 y を用いて以下の式で表される．

$$\log\left[\frac{P(y=1)}{1-P(y=1)}\right] = \alpha + \sum_{k=1}^{n} \beta_k x_k$$

本論文の場合，目的変数 y はコツ発話かどうか判定する対象の発話がコツ発話かどうかを示す二値変数（1 または 0）であり，$y=1$ はコツ発話であることを表す．

3.2　説明変数

本論文では 17 個の説明変数を設計する．これらはタグの出現頻度等に基づく事前調査で有効性が期待されたものである．以下では，コツ発話かどうかを判定する対象の発話を**判定対象**とよぶ．**二値変数**は 1 または 0 をとる説明変数を指す．**連鎖**とは発話意図アノテーションにおいて Relation がつけられているタグ同士を辿っていくとき，特定の順番のタグが見られることを指し，(タグ 1) → (タグ 2) → ··· のように示す．

連鎖に関係する説明変数　コツ発話が表出するとき，特定の発話意図の連鎖が頻繁に見られる．以下の説明変数は判定対象がそれらの連鎖における末尾に当たるかを表す．

- **req-ans**: 判定対象が Request_Info→ Answer における Answer であるかどうかの二値変数.
- **req-ans-sta**: 判定対象が Request_Info→ Answer→ Statement における Answer であるかどうかの二値変数.
- **req-ans-req-ans**: 判定対象が Request_Info→ Answer→Request_Info→Answer における 2 回目の Answer であるかどうかの二値変数.

判定対象内のタグに関係する説明変数　コツ発話内には特定のタグが出現しやすい傾向にあると考えられる．以下の説明変数は判定対象内に特定のタグが存在するかどうかを表す．

- **purpose**，**manner**，**condition**: 判定対象がそれぞれ意味フレームアノテーションの Purpose，Manner，Condition を含むかどうかの二値変数.
- **gesture**: Gesture を含むかどうかの二値変数.
- **long-filler**: 判定対象が 1 秒以上の Filler を含むかどうかの二値変数.
- **long-nod**: 判定対象が 0.8 秒以上の Nod を含むかどうかの二値変数.

直近発話列内のタグに関係する説明変数　コツ発話の直近の発話列には特定のタグが出現しやすいと考えられる．以下の説明変数は判定対象内，またはその直近の発話列の特徴を表す説明変数である．

- **recent-statement**: 直近の 5 発話（判定対象を

含む）内の Statement の数．3 以上の場合は 3 とする．

- **recent-prompt-act**: 直近の 5 発話（判定対象を含む）内に Prompt_Act が存在するかどうかの二値変数．
- **recent-trigger**: 直近の 5 発話（判定対象を含む）内に意味フレームアノテーションにおけるフレームタイプが存在するかどうかの二値変数．
- **successive-filler**: 直近の 5 発話（判定対象を含む）内の Filler の数．
- **successive-nod**: 直近の 5 発話（判定対象を含む）内の Nod の数．
- **smile**: 直近の 5 発話（判定対象を含む）内に S-laugh が存在するかどうかの二値変数．
- **kotsu-question**: 判定対象が「コツはなんですか？」や，「ポイントはなんですか？」など，明示的にコツを質問しているかどうかの二値変数．この説明変数のみアノテーションではなく書き起こし中の単語に着目する．具体的には，判定対象が Request-Info から連鎖したものであり，かつその Request-Info 内に「コツ」「ポイント」「注意」のいずれかの単語が存在するとき 1 をとる．

4 実験

4.1 実験設定

目的変数 key_utterance (binary) は，コツ発話アノテーションの Key_Utterance タグに対応する．説明変数は，EIDC の書き起こしテキストと各種アノテーションから抽出した．訓練データを用いて訓練した回帰モデルをテストデータに適用し，その正誤を集計する．正誤の判定に使用する閾値は 0.01 刻みの 0–1 の値を使用し，各ドメインごとのテストデータにおける F1 スコアが最大となったときの閾値も記録する．ロジスティック回帰モデルの実装には Python の statsmodels ライブラリを用いた．

発話意図アノテーション，パラ言語アノテーションは EIDC の一部のみに付与されていることから，すべてのアノテーションが付与されたコーパスの一部のみを実験に用いた．

訓練データは料理ドメインの 13 回のインタビュー 4，15，20，38，48，65，88，113，120，155，174，212，281，テストに使用するデータは料理ドメインの 2 回のインタビュー 64，248，および園芸ドメインの 8 回のインタビュー 21，39，60，69，71，76，89，98 とした．後者は，訓練とテストのドメインを変更することで，ドメイン転移の影響を確認することを意図している．

表 1 各説明変数の係数とその標準誤差

説明変数	係数	標準誤差
purpose	1.9	0.3
req_ans	1.6	0.2
manner	1.4	0.2
condition	0.9	0.4
req_ans_sta	0.6	0.3
kotsu_question	0.5	0.5
recent_trigger	0.4	0.1
gesture	0.3	0.1
recent_prompt_act	0.1	0.2
recent_statement*	0.05	0.07
long_nod	0.03	0.2
successive_nod*	0.03	0.02
successive_filler*	0.02	0.02
req_ans_req_ans	-0.1	0.4
long_filler	-0.2	0.2
smile	-0.2	0.2
const	-3.6	0.2

表 2 各ドメインでの最良の F1 スコアとその時の閾値．

テストドメイン	閾値	F1
料理	0.31	0.691
園芸	0.11	0.461
園芸*	0.31	0.436

4.2 実験結果

訓練した回帰モデルのパラメータとその標準誤差を表 1 に示す．なお，説明変数名に後続する*は説明変数が二値変数ではないことを示す．

次に，各ドメインにおける最大の F1 スコアを表 2 に示す．なお，園芸*は料理ドメインにおける最良の閾値を用いて予測したときの F1 スコアである．

予測結果に対する F1 スコアは料理ドメインで 0.691，園芸ドメインで 0.461 (0.436) となった．

4.3 考察

説明変数の中で特に大きな係数を示したのは purpose, req_ans, manner, condition, req_ans_sta, kotsu_question, recent_trigger, gesture であった．

表 3　事前調査の BERT ベースのモデルとの比較

モデル	料理 F1	園芸 F1	F1 減少率
線形回帰	0.691	0.461	0.333
BERT ベース	0.501	0.330	0.341

特に発話意図に関係する req_ans，req_ans_sta の係数が大きいことから，EIDC においてコツ発話が表出する典型的なパターンは，質問への回答，またそれに付随する説明であるとみられる．

意味フレームアノテーションのタグに関係するものとしては，purpose，manner，condition，recent_trigger が挙げられる．purpose，manner，condition が言及されることがコツ発話の特徴として挙げられること，また直近でフレームタイプについて言及している時コツ発話が表出しやすいことが考えられる．

二値変数でない説明変数を見ると successive_filler，successive_nod が正でありかつ，共に少なくとも 25 になる発話が見られたため，例えば successive_filler，successive_nod がそれぞれ 25 だとすると，$\beta_{sf}x_{sf} = 0.5$，$\beta_{sn}x_{sn} = 0.75$ となるため，filler や nod が頻発している状況はコツの表出に大きく寄与している可能性があると考えられる．

精度についてはドメイン転移後に 0.230 (0.255) 低下する結果になった．理由としては料理ドメインのテストデータが小さいこと，またはドメインごとの説明変数の分布が異なることが考えられる．また，このコーパスで捉えきれていないインタビューの特徴があることが考えられ，そうした特徴を取り込めれば精度向上が見込めるかもしれない．

回帰モデルの比較対象として，BERT (DeBERTa V2) ベースモデルを fine-tuning した．BERT は入力に用いる発話列は回帰モデルと同じだが，アノテーションではなく発話列テキストを直接用いる．そのため意味内容に踏み込める一方で，ドメイン転移時の精度低下が予想される．比較結果を表 3 に示す．回帰モデルはドメイン転移前後ともに BERT を上回っており，十分に高い精度を達成していると判断できる．なお，回帰モデルのスコアは最良の F1 スコアとなる時の閾値を用いた時の数値を用いた．

4.4　Ablation 実験

考察において寄与度が高いと考えられた説明変数が実際に精度にどの程度寄与しているかを調べるために，上記の結果において係数の高かった順に説明変数を一つずつ除外してモデルの訓練，テストを行った．除外する説明変数は req-ans，condition，purpose，manner，successive_filler，successive_nod とした．テストにおける F1 スコアを以下の表 4 に示す．なお，どちらのドメインにおいてもベストスコアが出せる閾値における F1 スコアを示している．

表 4　Ablation 実験の F1 スコア

Ablation 対象	スコア（料理）	スコア（園芸）
なし	0.691	0.461
req_ans	0.714	0.456
condition	0.667	0.415
manner	0.674	0.446
purpose	0.607	0.438
successive_filler	0.713	0.438
successive_nod	0.691	0.451

園芸ドメインにおけるスコアを見ると，どの説明変数を除外してた場合もスコアは若干の減少に留まる．このことから，いずれかの説明変数が予測に必須というわけではないと推測できる．

5　結論

本論文では技能者インタビュー対話コーパスを用いてコツ発話の表出を予測するロジスティック回帰モデルを構築し，モデルの精度とパラメータを示し，コツ発話の表出に関わるインタビューの特徴を考察した．Purpose，Manner，Condition といったコツ発話内を伴いやすいタグや，インタビューという形式上質問に対する回答という形でコツ発話が表出しやすいこと，また直近でフィラーや頷きなどが多く表出している時にコツ発話が表出しやすいことが明らかになった．

本論文では，予測対象となる発話そのものも説明変数として取り込んだ．この設定は，コツ発話の表出を事後的に認識する目的には有効であるが，コツ発話の表出を促す目的には適さない．そこで，対象発話を参照せずに予測を行うという，より挑戦的な設定についても検討したい．さらに，説明変数の拡大や，現在のアノテーションではとらえきれていないインタビューの特徴のさらなる分析が今後の方向性として考えられる．

謝辞

この成果は, 国立研究開発法人新エネルギー・産業技能総合開発機構 (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] Charles Goodwin. Professional vision. **American Anthropologist**, Vol. 96, No. 3, pp. 606–633, 1994.

[2] 伝康晴. 身体的実演を伴う教授場面の相互行為分析—アドレス性に注目して. 田中廣明, 秦かおり, 吉田悦子, 山口征孝（編）, 動的語用論の構築へ向けて 第 3 巻, pp. 140–161. 開拓社, 2021.

[3] 林誠, 安井永子. サッカー指導場面での「身体的実演」に見られるコーチと選手の相互行為. 小宮友根, 黒嶋智美（編）, 実践の論理を描く—相互行為のなかの知識・身体・こころ, pp. 158–175. 勁草書房, 2023.

[4] Eiko Yasui. Japanese onomatopoeia in bodily demonstrations in a traditional dance instruction: A resource for synchronizing body movements. **Journal of Pragmatics**, Vol. 207, pp. 45–61, 2023.

[5] 城間祥子. 身体の技を伝え指導することば—能楽の稽古の事例をもとに. 日本語学, Vol. 23, No. 1, pp. 52–60, 2004.

[6] 生田久美子, 北村勝朗（編）. わざ言語: 感覚の共有を通しての「学び」へ. 慶應義塾大学出版会株式会社, 2011.

[7] 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.

[8] 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.

[9] 近大志, 岡久太郎, Yin Jou Huang, 樽谷洋希, 松田思鵬, 村脇有吾, 黒橋禎夫. 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. 言語処理学会第 31 回年次大会 発表論文集, 2025. (to appear).

A　意味フレームアノテーションのタグ

A.1　フレームタイプ

料理ドメイン

- **BAKE_FRY** 主に油を用いて火や熱源で調理する動作を表す.
- **DIVIDE**: 何かの全体ないしは部分を 2 つ以上に分ける動作を表す.
- **CHANGE**: 形状・向き・温度を変化させる動作を表す.
- **SIMMER**: 味付けのために液体で加熱する動作を表す.
- **HEAT**: 何かの温度を上げる動作を表す.
- **MIX**: 2 つ以上の物を合わせ，境界のない一つのものにする動作を表す.
- **PUT-ON**: 粉末状，粒状，液状，半固形状の物を別の物にかける・塗る動作を表す.
- **PLACE**: 調理のしやすさや食べやすさ，見栄えのために特定の方法で何かを配置する動作を表す.
- **WAIT**: 何かを特定の状態で意図的に放置する動作を表す.
- **COMPOUND**: 同程度の大きさの 2 つ以上の食材を接着させる動作を表す.
- **REMOVE**: 何かから何かを取り除く動作を表す.

園芸ドメイン

- **MIX**: 土と肥料，殺虫剤と水など，何かと何かを混ぜる動作を表す.
- **CHANGE**: 土に何らかの操作を加えて状態を変化させる動作を表す.
- **PLACE**: 特定の仕方で植物を配置する動作を表す.
- **SUPPLY**: 水や肥料，殺虫剤などを植物もしくは土にかける動作を表す.
- **SOW**: 植物の種を土や鉢などに蒔く動作を表す.
- **REMOVE**: 植物の全体，もしくは枝葉など植物の一部を取り除く動作を表す.
- **DIVIDE**: 植物を特定の大きさに分割する動作を表す.
- **TRANSPLANT**: 植物をある場所から別の場所へ移す動作を表す.
- **COVER**: 植物，土壌を覆うカバーなどを敷く動作を表す.
- **ELIMINATE**: 害虫を駆除，または追い払う動作を表す.
- **ARRANGE**: 植物の形状を意図的に変化させる動作を表す.
- **HARVEST**: 作物や植物の種子を採る動作を表す.

A.2　フレーム要素

- **Manner**: 当該動作（フレームタイプ）の実施様態を表す.

A.3　指定要素

- **Purpose**: 主に「〜のために」の形で表現され，当該のイベントを行うことでどのような利点があるかを表す.
- **Condition**:「〜の場合は」や「〜の人は」といった形で，通常のレシピ，工程から外れて，特定の条件を想定した説明をしている時，その条件部分に相当する.

B　発話意図アノテーションのタグ

- **Request-Info**: 情報を求める/引き出す発話を表す.
- **Answer**: 対話相手の質問に対する直接の応答を示す発話を表す.
- **Statement**: レシピや作業工程の情報を提供する発話を表す．Statement に該当するのは，技能者が手順を紹介する発話に加え，完成物の紹介，応答 (Answer) における前提や補足情報などを述べる発話である.
- **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を表す.

C　パラ言語情報のタグ

- **filler**:「えー」や「あのー」に代表される「言い淀み時などに出現する場繋ぎ的な表現」，または相槌を表す.
- **nod**: 頭の動き，主に頷きを表す.
- **laugh**: 笑いを表す．声を上げる笑いを L-laugh (Laugh)，声を上げない笑い，微笑みを S-laugh (Smile) とする.
- **gesture**: 意図的な手の動きを行っている部分を表す．インタビューは Zoom 上で行われているため，説明のためのマウスカーソルの動き等も gesture に含まれる.