

Title	PREP 話法のアノテーションに基づく就職面接における コミュニケーションスキルの分析
Author(s)	前田, 雄之介
Citation	
Issue Date	2024-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/18894">http://hdl.handle.net/10119/18894</a>
Rights	
Description	Supervisor: 岡田 将吾, 先端科学技術研究科, 修士(情報科学)

修士論文

PREP 話法のアノテーションに基づく就職面接における  
コミュニケーションスキルの分析

前田雄之介

主指導教員 岡田将吾

北陸先端科学技術大学院大学  
先端科学技術研究科  
(情報科学系)

令和6年3月

## Abstract

Interpersonal communication skills are required in various situations in the modern age, and these skills are especially important in job hunting. Previous studies have attempted to estimate the skills of interviewees based on their nonverbal behavior during interviews, but they have not taken into account the interviewer's point of view in interview evaluation. In this study, we focused on the four components of PREP-based discourse, which is considered to be an effective speech structure for obtaining a high evaluation in job hunting: "P: Point," "R: Reason," "E: Example," and "P: Summary Point. First, we collected data including interviewees' speech, language, and interview evaluations by skilled interviewers obtained during job interview training, and annotated the speech transcription data with each component of the PREP-based discourse to construct a data corpus. Next, we extracted the linguistic features of the responses to which each element of the PREP -based discourse was assigned, analyzed the relationship between the PREP-based discourse and the interview evaluation scores, and analyzed the relationship between the PREP-based discourse and the interview evaluation scores. We then analyzed the relationship between the PREP-based discourse and the interview score, and clarified the linguistic features related to the PREP speech that were related to the interview score. Finally, we report the results of our work on a classification problem for automatically annotating the four elements of the PREP-based discourse to the response data, with the aim of applying it to automatic skill estimation.

# 目次

<b>第1章</b>	<b>はじめに</b>	<b>1</b>
<b>第2章</b>	<b>関連研究</b>	<b>2</b>
2.1	自動面接評価	2
2.2	議論マイニングにおけるデータセットの構築	2
2.3	本研究の位置付け	3
<b>第3章</b>	<b>PREP 話法アノテーション</b>	<b>4</b>
3.1	PREP 話法の定義	4
3.2	データセット	5
3.2.1	就職面接システム	5
3.2.2	実験参加者	6
3.2.3	面接評価アノテーション	6
3.3	アノテーション	6
3.3.1	アノテーションスキーム	6
3.3.2	アノテーションの流れ	7
3.4	アノテーション結果	8
3.4.1	アノテーション結果の概要	8
3.4.2	アノテータの一致度	8
<b>第4章</b>	<b>アノテーション分析</b>	<b>10</b>
4.1	PREP の有無と面接評価の関係	10
4.2	PREP 話法で述べられた内容と面接評価の関係	11
4.2.1	J-LIWC によるカテゴリー付け	11
4.2.2	PREP の要素に含まれるカテゴリーと面接評価の関係	12
<b>第5章</b>	<b>実験と評価</b>	<b>20</b>
5.1	特徴量	20
5.2	使用した回帰モデルと実験設定	20
5.3	結果	21
<b>第6章</b>	<b>PREP アノテーションの自動化</b>	<b>25</b>
6.1	大規模言語モデルと議論マイニング	25

6.2	手法	25
6.2.1	プロンプト	26
6.2.2	評価	26
6.3	結果	26
<b>第7章</b>	<b>終わりに</b>	<b>28</b>
<b>付録A</b>	<b>プロンプト</b>	<b>30</b>
A.1	プロンプト1	30
A.2	プロンプト2	31
A.3	プロンプト3	33
A.4	プロンプト4	38

# 目次

6.1 人手によるアノテーションと GPT によるアノテーションの比較 . . .	27
---	----

# 表目次

3.1	質問文の内容	5
3.2	5つの評価項目	6
3.3	PREP 話法でよく使われるフレーズ	7
3.4	アノテーション結果の概要	8
3.5	PREP の各要素の有無	9
3.6	アノテータ間の一致度	9
4.1	PREP の有無と総合評価の検定	11
4.2	PREP の有無と論理性の検定	11
4.3	PREP の有無と言葉遣いの検定	12
4.4	回答文全体の J-LIWC の出現頻度と評価の相関係数	16
4.5	理由がある回答の J-LIWC の出現頻度と評価の相関係数	17
4.6	具体例がある回答の J-LIWC の出現頻度と評価の相関係数	18
4.7	まとめの結論がある回答の J-LIWC の出現頻度と評価の相関係数	19
5.1	ラッソ回帰の精度表	22
5.2	リッジ回帰の精度表	23
5.3	SVR の精度表	24
6.1	GPT の各プロンプトごとの評価	27

# 第1章 はじめに

現代ではコミュニケーション能力は社会の様々な場面で必要不可欠な能力とされており、特に就職活動で行われる面接官と求職者の面接ではこのスキルが重要視される。このため、求職者は面接訓練を必要としているが、効果的にトレーニングを行えるキャリアカウンセラーの数は限られており、面接訓練の自動化に向けた研究は様々に行われている [15, 20, 19, 23, 21]。これらの研究によって面接官の評価を自動化し、被面接者に対してフィードバックを行うことが可能になったが、面接官が具体的に被面接者のどのような点に着目しているかは明らかになっていない。

一方で、自然言語処理の議論マイニングは自然言語の論証の構造化に取り組んでいる。特に、小論文やディベートにおいてその論証の構造を分析することでスキルを評価、フィードバックを行う研究が進められている [5]。

就職面接においても回答の構造は重要であり、就職面接の訓練において PREP 話法と呼ばれる文章構造に沿って回答をすることがよく指導される。PREP 話法とは回答を P: 結論, R: 理由, E: 具体例, P: まとめの結論の4つの要素で構成することである。この構造に則って面接の回答を行うことでより説得力のある評価されやすい回答になるといわれている。

本研究では PREP 話法に基づいてされた回答が面接の評価にどのような影響をもたらすかを分析する。まず就職面接訓練で取得された被面接者の音声、言語、動画と面接官による評価を含むデータセットから発話の書きおこしを行い人手によって PREP 話法の各要素のアノテーションを行いデータコーパスを作成した。次に、PREP の各要素が付与された回答文の言語特徴量と面接官の評価の相関について分析した。そして、このようにして得られた言語特徴量がスキル予測に寄与するかについていくつかの線形回帰モデルによって調べた。最後に、将来的にスキルの自動推定と PREP を利用したフィードバックを行うために、回答の PREP の4要素への自動分類に取り組んだ。

## 第2章 関連研究

### 2.1 自動面接評価

プレゼンテーションスキルや就職面接のマルチモーダル情報を用いたスキル推定の先行研究について述べる。Ramanarayananら[13]は音声、言語、体の動き、感情などのマルチモーダルな特徴量からプレゼンテーションのスコアを予測し人間に近い推定を行った。sagaら[14]はソーシャルスキルトレーニングのスコア推定において言語、音声の特徴量が有効であることを示した。Rehmanら[12]は被験者のビデオ通話による就職面接の合否判定を音声、言語、Openfaceによって取得した被験者の表情などの特徴量から推定した。さらに、発話内容の単語の重要度の分析を行い合否に関係する単語の重要度を分析した。

本研究で用いたデータセットと同じデータセットを利用した先行研究はOhbaら[20, 19, 23, 21]がある。Ohbaら[8]では41名分の被験者データの音響、言語、生体信号、視線を収集し、それらの特徴量を線形回帰モデルと時系列モデル分析して面接スキルと被験者の面接に対する自信度とその2つのギャップについて推定した。大庭ら[23]ではそれまで収集していたデータの音響と言語特徴量を利用した線形回帰モデルによるスキル推定を行い、それに基づいて面接中の動画、スキルのレーダーチャート、フィードバックコメントを示すフィードバックシステムを開発した。さらに、新たに45名分の面接データを収集し就職面接に関する参考書とフィードバックシステムによって学習した被験者を比較してフィードバックシステムが被験者に与える影響について分析した。

これらの先行研究では被験者へのフィードバックはスキル推定によるルールベースのコメントにとどまっており、被験者の回答の構造に沿った教育は行われてこなかった。

### 2.2 議論マイニングにおけるデータセットの構築

議論マイニングは自然言語の主張と理由を抽出し論証の構造化を自動化する。この分野では文章の構造や要素に対するアノテーションが行われている。Parkら[9]はeRulemakingオンラインの投稿においてユーザーのコメントの命題を論証の種類によって4つにアノテーションしデータセットを構築した。Egawaら[2]はオン

ライン議論フォーラム ChangeMyView での投稿を5種の要素とその要素が支持関係にあるか反対関係にあるかの関係性をアノテーションした。

## 2.3 本研究の位置付け

これまでの研究では被面接者の音声や言語などの特徴量から面接スキルを推定してきたが、面接官の具体的な評価の着眼点については議論されてこなかった。議論マイニングにおいても Nguyen ら [7] の研究では小論文の論証性を考慮した自動採点が行われているが、就職面接の対話に対して回答の論証構造について分析したものはない。本研究ではまず就職面接の回答を構造的に分析するために PREP 話法に基づく4つの要素を付与するアノテーションを行う。次に、付与された要素の言語特徴量を抽出し面接評価点との関係を調べることで、どのような発話内容が評価に影響を与えるのかを明らかにする。

## 第3章 PREP 話法アノテーション

本研究では PREP 話法に基づく就職面接の回答へのアノテーションを行った。PREP 話法とは回答を P: 結論, R: 理由, E: 具体例, P: まとめの結論の 4 つの要素で構成することである。アノテーションでは就職面接対話データセット 50 人分に対し 5 人のアノテータによる PREP の 4 要素のアノテーションを行い、アノテータ間の一致率を求めた。

### 3.1 PREP 話法の定義

本研究では就職面接の回答を構造化するために PREP 話法を用いる。PREP 話法は就職面接の訓練においてよく指導されている文章構成方法で、相手にわかりやすく論理的に話を伝えることができるとされている。PREP とは Point: 結論, Reason: 理由, Exsample: 具体例, Point: まとめの結論であり、この 4 つの要素を以下のように定義する。

- **結論:** 全体で最も伝えたい重要な部分。
- **理由:** 結論を支持するもっとも端的な説明。
- **具体例:** 理由を裏付けるデータや体験談などの個別の例。
- **まとめの結論:** 最初に述べた結論を繰り返したもの。

例えば「あなたの強みは何ですか?」という質問に対して PREP 話法を用いて回答すると以下のように構成できる。

私の強みは忍耐力です（**結論**）。なぜなら私は苦手なことや難しいことであっても継続して取り組み成果を出すことができますからです（**理由**）。例えば高校生の時は英語が苦手だったのですが、大学生になってから毎日 3 時間継続して勉強することで TOEIC で 800 点を取ることができました（**具体例**）。こうした経験から何事にも私は粘り強く努力する力が身につきました。（**まとめの結論**）

PREP 話法は結論、理由、具体例、まとめの結論の順で構成することを推奨しているが、理由と具体例の順を入れ替えたり理由を最初に述べても同様な効果を生むことができる。

表 3.1: 質問文の内容

ID	質問文
1	あなたが大学時代最も力を入れたことについて具体的に教えてください
2	これまでに苦労した、または大変だったことについて教えてください。 また、それをどう乗り越えたかについて詳しく教えてください。
3	就職活動におけるあなたの仕事選びの軸を教えてください。 また、そう考えた理由やきっかけがあればそれを含めてお話しください。
4	あなたの志望業界についてお聞きします。あなたの志望業界を教えてください。 また、そう考えた理由やきっかけがあればお話しください。
5	第一志望の会社とその志望動機を教えてください。 もし、その第一志望の会社に入社したとして、入社後やってみたい仕事はありますか？
6	その理由を含めてお話しください。
7	あなたの強みは何ですか？その理由などを自身の経験をもとにお話しください。
8	あなたの弱みは何ですか？その理由などを自身の経験をもとにお話しください。 あなたは周囲の人からはどのような人だと言われますか？
9	また、そう考える理由や印象深いエピソードがあればそれも教えてください。
10	あなたはチームの中でどのような役割を果たすことが多いですか？ また、そう考える理由や印象深いエピソードがあればそれも教えてください。
11	あなたが対人関係で大切にしていることを教えてください。

## 3.2 データセット

本研究では Ohba らの研究 [8][23] で 2020 年から 2022 年に収集された就職面接対話実験データセットを使用した。このデータセットでは VR を用いて被験者がバーチャルエージェントと面接対話を行い音声、言語、動画、視線、生体信号の情報を収集した。実験参加者は日本での就職を予定している、もしくはすでに終えた大学生、大学院生である。このデータにキャリアカウンセラー 2 名による面接評価のアノテーションが付与されている。

### 3.2.1 就職面接システム

データセットで用いられた就職面接システムでは被験者は VR ヘッドマウントディスプレイ (FOVE01) を装着し VR 空間上でバーチャルエージェント 3 名と面接を行う。質問はキャリアカウンセラーの指導の下作成された一般的な一次面接で問われる 13 項目で行われたが、本研究のアノテーションでは以下の 11 質問に対する回答を対象とした。(表 3.1) データはマイクやビデオによって音声、言語、動画、視線、生体信号の情報を収集した。

表 3.2: 5 つの評価項目

評価項目	評価基準
総合評価	面接の総合点（どの程度その人を採用したいと思うか）
積極性	被面接者は質問に対してどの程度積極的な姿勢を見せたか.
声の印象	被面接者の声量や話すスピードは適切であったか.
言葉遣い	被面接者の言葉遣いは適切だったか.
論理性	論理性のある回答ができているか.

### 3.2.2 実験参加者

この実験に参加したのはこれから就職活動を始める人、または就職活動をしているかもう終えた大学院生で日本語が流ちょうに話すことができる。本研究ではこのデータセットの中から 50 名分のデータを対象とした。

### 3.2.3 面接評価アノテーション

このデータセットには熟練したキャリアカウンセラー 2 名が面接実験のビデオを見て、被験者の面接能力を積極性、声の印象、言葉遣い、論理性、総合評価の 5 項目（表 3.2）について 7 段階のリッカート尺度で評価を行った。先行研究 [23] によって 2 人の評価の一致度は十分確認されている。

## 3.3 アノテーション

PREP 話法アノテーションは被面接者の回答に対し PREP の各要素を付与する作業で 5 人のアノテータによって行った。手順書による教育と事前アノテーションによる疑問点やラベルが一致しない点についてのキャリブレーションを行い本アノテーションを行った。

### 3.3.1 アノテーションスキーム

アノテーションでは 3.2 節で収集された面接の回答の書きおこし文章について、原則一文に一つ PREP 話法の結論、理由、具体例、まとめの結論の 4 つのラベルをいずれかを付与する。ただし、一文に複数の要素がある場合には節を最小単位とし、どの要素も付与できない場合はその文や節に対し何もラベルを付与しないことを認めた。また、付与したラベルに対して自信がないときは「自信なし」ラベルを同時に付与した。アノテータにはガイドラインとして 3.1 節での定義に加えて、表 3.3 のその要素でよく使われるフレーズを与えた。

表 3.3: PREP 話法でよく使われるフレーズ

ラベル	よく使われるフレーズ
結論	私は～です 結論から申しますと～です
理由	それは～だからです なぜかという～だからです
具体例	例えば～ これは…の時ですが～
まとめの結論	以上のことから～ 繰り返しになりますが～

以下の例のように被面接者が具体例によって結論を支持する理由としようとしている場合や、理由と具体例がはっきり分けられないときには「理由または具体例」のラベルを付与することを認めた。例はチームで果たす役割に対する回答だが、エピソードによって結論を支持しようとしている場合にこのラベルを使用することができる。

私はキャプテンということではないんですけども、最終的な意思決定をするようなそういった部分がチームでの役割として多いことがあります。(結論) 例えばロボカップではチームのキャプテンは先生、教授だったんですけども、同期の中で私が誰がこれをやって次にこれをやろう、そういった部分を決めていたことが多々ありましたので意思決定役といったそういった部分が、私はチームの中では比較的多い役割となっています。(理由か具体例)

また、質問 ID8 の弱みを聞く質問に関してはまとめの結論を必ずしも結論の繰り返しとするのではなく、最初の結論で述べた弱みの改善の取り組みや補うための工夫も該当するものとして定義した。

### 3.3.2 アノテーションの流れ

アノテーション作業は3.2のデータセットの50人分(550回答)に対し、日本語話者5人のアノテータによってオープンソースのテキストアノテーションツールの doccano[6]を用いて行った。作業の前にアノテーションスキームを読むよう指示し、3人分33回答のデータに対し事前アノテーションを行った。次に、疑問点やアノテーションが一致しなかった部分についてチャット上でアノテータと議論を行いキャリブレーションを行いアノテーションスキームを修正した。そして、事前アノテーションとは異なる3人分のデータでアノテーションを行い、アノテータの疑問が解消されたことを確認して残りの44人分のデータで本アノテーション

表 3.4: アノテーション結果の概要

	結論	理由	具体例	まとめの結論	理由か具体例	自信なし
アノテーター 1	519	275	201	233	60	1
アノテーター 2	535	268	165	305	171	6
アノテーター 3	524	404	235	243	4	0
アノテーター 4	531	338	116	189	24	3
アノテーター 5	520	353	151	352	77	20

を行った。本アノテーション中も疑問点があればチャット上で議論しアノテータ全員と共有した。

## 3.4 アノテーション結果

### 3.4.1 アノテーション結果の概要

事前アノテーションで行った3人分のデータを除き、47人分517回答のデータに対し各アノテータが付与したラベルの数を表3.4に示す。この表では付与された回数を示しているため、結論が複数あるなどして一文に2回付与された場合もあるので回答の総数517個を上回る数が示されている。さらに、アノテータ3人以上がその回答に対してPREP話法の各ラベルを付与したものを要素有りとしたときの回答の数を数えると表3.5になる。

結論に関してはほぼすべてのアノテータが同じアノテーションをしていたが、理由、具体例についてはどちらを選ぶべきかたびたび議論が行われアノテータによって違いがみられた。自信なしのラベルを付与した数は少なくアノテータは自信をもって作業に取り組めたといえる。

PREPの有無についてはすべての文章で結論が存在していたが、結論しか要素を持たない回答は98回答であり、多くの回答は理由、具体例、まとめの結論のいずれかが欠けていた。複数の要素を持つ回答は少なく理由と具体例の両方を持つ回答は74回答で、PREPのすべての要素を含んだ回答は65回答(12%)と少数であった。

### 3.4.2 アノテータの一致度

回答の文に付与されたラベルについて5人のアノテータの一致度をFleiss's kappaによって求めると0.554であり、複雑なアノテーションであったが中程度の一致率だといえる。また、それぞれのアノテータ間の一致率をCohen's kappaによって求めると表3.6になる。0.526-0.624と一致率は安定していたが、「理由か具体例」ラ

ベルを多用したアノテータ2は他と比べて一致率が低くなっており、回答の理由と具体例の部分でアノテーションのずれが生じていたと考えられる。

表 3.5: PREP の各要素の有無

	有	無
結論	517 (100%)	0 (0%)
理由	310 (56%)	207 (44%)
具体例	169 (32%)	348 (68%)
まとめの結論	217 (42%)	300 (58%)
全ての要素	65 (12%)	452 (88%)

表 3.6: アノテータ間の一致度

	アノテータ1	アノテータ2	アノテータ3	アノテータ4	アノテータ5
アノテータ1	1.000	0.529	0.584	0.598	0.542
アノテータ2	0.529	1.000	0.534	0.539	0.549
アノテータ3	0.584	0.534	1.000	0.624	0.526
アノテータ4	0.598	0.539	0.624	1.000	0.548
アノテータ5	0.542	0.549	0.526	0.548	1.000

## 第4章 アノテーション分析

前章で収集したデータについて、まず回答の構造と面接官の評価の間にどのような関係があるのかを分析するために、回答の中の PREP の要素の有無によって面接官の評価に違いがあるかをマンホイットニーの U 検定によって調べた。そして、感情辞書の J-LIWC を用いて回答文でどのようなカテゴリーの単語を述べているか調べることで、どのカテゴリーの単語が評価に影響を及ぼすのかを分析した。

### 4.1 PREP の有無と面接評価の関係

PREP の要素の有無によって面接官の総合評価、論理性、言葉遣い (1~7 点) による評価に差が生じるかをマンホイットニーの U 検定で調べた。アノテーションにおいてその回答に 3 人以上のアノテータが PREP の各要素を付与したときその回答を PREP の要素有りとしそれ以外を要素なしとした。結論はすべての回答で要素有りだったので検定は行わず、理由、具体例、まとめの結論のすべての要素がありだった回答をすべての要素有りとした。また、要素の有無によってどの程度評価が変わるかを比較するために効果量  $r$  も求めた。それぞれの評価項目について各要素ごとに回答数、評価の平均、 $p$  値、効果量  $r$  を求めたものを表 4.1, 4.2, 4.3 に示す。

総合評価と論理性では理由、具体例、まとめの結論、すべての要素のすべてで  $p < 0.05$  を示し有意差はないといえる。また、効果量については理由が最も高く評価に影響を及ぼしていることが分かった。言葉遣いでは理由、まとめの結論、すべての要素が  $p < 0.05$  で有意差が認められたが、具体例では有意差は認められなかった。また、効果量についてはすべての要素が 0.10 で最も大きかったが、総合評価と論理性と比べると値は低くなっていた。

PREP の要素の有無は総合評価と論理性についてはすべての要素で有意差が認められ、言葉遣いでも具体例以外では有意差が認められたため、回答の構造と面接官の評価には関係があるといえる。さらに、効果量については理由が大きく、理由の有無によって評価が変化したといえる。しかし、回答の中には PREP 話法の構造に則って話しているにもかかわらず評価の低いものも存在しており、そういった回答はふさわしい内容を話すことができていないと考えられる。

表 4.1: PREP の有無と総合評価の検定

PREP の要素	要素有		要素無		p 値	効果量 r
	n	平均値	n	平均値		
理由	310	3.13	207	2.71	<b>0.000</b>	0.20
具体例	169	3.18	348	2.85	<b>0.003</b>	0.12
まとめの結論	217	3.20	300	2.79	<b>0.000</b>	0.11
全ての要素	65	2.89	452	3.46	<b>0.000</b>	0.16

表 4.2: PREP の有無と論理性の検定

PREP の要素	要素有		要素無		p 値	効果量 r
	n	平均値	n	平均値		
理由	310	3.16	207	2.79	<b>0.000</b>	0.18
具体例	169	3.21	348	2.91	<b>0.009</b>	0.11
まとめの結論	217	3.22	300	2.86	<b>0.000</b>	0.16
全ての要素	65	3.40	452	2.96	<b>0.000</b>	0.14

## 4.2 PREP 話法で述べられた内容と面接評価の関係

PREP 話法と面接評価の関係についてとくに総合評価と論理性においては PREP の要素の有無によって評価に差があることが分かった。しかし、PREP の要素があるにもかかわらず評価が低い回答があり、このような回答にはその質問に対する回答としてはふさわしくない内容を述べていると考えられる。そこで、心理学において広く用いられている感情辞書を日本語化した J-LIWC2015[4] を利用し、PREP の要素を持つ分の中で J-LIWC のどのカテゴリーが評価に関係しているのかを分析する。

### 4.2.1 J-LIWC によるカテゴリー付け

J-LIWC は心理学的なテキストデータ分析において英語で一般的に用いられる LIWC の日本語版であり、助動詞や動詞といった言語学的なカテゴリーと感情的プロセスや社会的プロセスの心的カテゴリーを日常でよく用いられる単語に割り当てた感情辞書である。ポジティブとネガティブの 2 カテゴリーのみの分類を行う感情辞書とは異なりネガティブ感情を上位のカテゴリーとして「怒り」や「悲しみ」のカテゴリーが存在する。J-LIWC では 11609 の単語と 69 のカテゴリーによって構成されておりその妥当性が示されている。

本研究ではそれぞれのアノテータがラベル付けした文章を Mecab を用いて形態素解析を行いそれぞれの単語を J-LIWC によるカテゴリー付けを行い文中での出現割合を求めた。これを各回答に行い、回答の PREP の要素ごとにカテゴリーの

表 4.3: PREP の有無と言葉遣いの検定

PREP の要素	要素有		要素無		p 値	効果量 r
	n	平均値	n	平均値		
理由	310	2.89	207	2.73	<b>0.023</b>	0.09
具体例	169	2.94	348	2.77	0.064	0.07
まとめの結論	217	2.93	300	2.86	<b>0.028</b>	0.09
全ての要素	65	3.06	452	2.79	<b>0.009</b>	0.10

出現割合を求めた。さらに、PREP の要素に分けない回答文全体のカテゴリの出現割合を求めた。そして、5 人のアノテータの平均をそのカテゴリの出現頻度とし面接評価との関係を調べた。

#### 4.2.2 PREP の要素に含まれるカテゴリと面接評価の関係

回答の内容と面接評価の関係を分析するために、回答文の中の J-LIWC の各カテゴリの出現頻度と総合評価、論理性、言葉遣いの評価（1～7 点）の相関を調べた。まず、質問の内容によって回答の内容は変わるため以下のように質問を 3 つのカテゴリに分け、質問の種類ごとに相関を分析した。

1. 自身がこれまで取り組んできたことに関する質問

質問 ID1：学生時代最も力を入れたこと

質問 ID2：これまでに苦労した経験

2. 志望する業界・企業についての質問

質問 ID3：仕事選びの軸

質問 ID4：志望業界について

質問 ID5：第一志望の会社について

質問 ID6：第一志望の会社に入ったとしてやってみたい仕事

3. 自分自身について質問 ID7：自分自身の長所について

質問 ID8：自分自身の短所について

質問 ID9：周囲の人からどう思われているか

質問 ID10：チームの中でどのような役割を果たすか

質問 ID11：対人関係で大切にしていること

そして、回答文全体での出現頻度と評価の相関を求め、評価項目ごとに絶対値の高い上位 5 つのカテゴリを調べた。(表 4.4) 表でのカテゴリ名の「all」は回

答全ての中のカテゴリーを示す. 同様に「r」は理由, 「e」は具体例, 「p 2」はまとめの結論の中でのカテゴリーを意味する. 例えば「all\_bio」は回答文すべての中の「bio」カテゴリーを意味する. 次に結論を除いてそれぞれの PREP の要素が含まれている回答と PREP の要素全てが含まれている回答の中で相関を求めた. この時, その要素内でのカテゴリーについて調べるため結論の要素とその要素での出現頻度についてのみ相関を調べた. このようにして PREP 話法にのっとなって話しているにもかかわらず評価に違いが生まれる要因を分析した. (表 4.5,4.6,4.7

### 回答文全体のカテゴリーの出現頻度と評価の相関

表 4.4 によると質問 ID1,2 についてはすべての評価項目で「all\_bio」の相関が高かった. 「bio」カテゴリーは「遺伝」や「生態」など生物学的プロセスを含むカテゴリーだが以下のような回答においてよく見られた.

私が一番力を入れたことはラグビーです。(結論) 私は元々**体**が小さくてですね、なかなか、ラグビーというのはできない状態だったんですが、実際に自分ができないことに挑戦したいと思ったので勝負いたしました。(理由) 実際に参加してみたら、やっぱり**怪我**も多くてですね、**骨折**とか**脳震盪**などをして、1年以上は試合に出れない状態があったんですがそこで自分が何で出来ないのかとか、どういったことでこれから試合に出れるように何をすればいいのかみたいなことを考えた時に、まずは**体重**の増加と**筋肉**の増加っていうところが一番主要な課題でしたので、1日の**食事**の量を、3回から5回に変えたりとか、毎朝起きて、タックル練習とかタイヤを押すって地道なトレーニングを重ねました。(具体例) その結果2年間で体重を20キロ以上増加させて、60キロぐらいだったものが、85キロとか80キロぐらいになることで、レギュラーに選ばれてチームに貢献することができました。(まとめの結論)

この例は大学時代力を入れたことに対する回答で評価は総合評価4.5, 論理性4, 言葉遣い4である. この回答では「体」, 「怪我」, 「骨折」, 「脳」, 「体重」, 「筋肉」, 「食事」が「bio」カテゴリーに含まれ運動部での経験を語る際によく使われる単語が当てはまる.

こちらも留学中のエピソードと重なるのですが、留学中に感じた言語の壁と文化の壁を乗り越えることがとても大変でした。(結論) この点において私は初め自分の英語のスキルに自信がなく、自分からクラスメイトやお店の人などに話しかけることができなかつたのですが、これではこのまま1年を過ごしてしまつたら、何も変わらないで終わってしまうという思いから、1日1時間以上クラスメイトかお店の人など、地元の英語を話す方と喋る機会を設けるという約束を自分として、それを毎日意識して、心がけるようにしました。そうすると自然と英語の言い回しなども出てくるようになり、言語の壁を乗り越えることができました。(具体例) また、それに伴って、相手がどういった**生活**であったり、文化を基にして**生きて**いるのかということもコミュニケーションの中から知ることができるようになって、最後には、言語と文化の壁両方を乗り越えることができた1年間になったと感じています。(まとめの結論)

この例は学生自体苦勞したことに関する回答で評価は総合評価5, 論理性4, 言葉遣い4.5である。この回答では「生活」, 「生きて」が「bio」カテゴリーに含まれるが、いずれも留学先での体験について述べている。体育会系の部活や留学先での体験は就職活動でアピールされる定番だが、これらの単語はより自身の体験を具体的に語るのに用いる言葉であり表面的な体験ではなく具体的に自分自身がどのように取り組む必要性を示している。

### 理由を含む回答のカテゴリーの出現頻度と評価の相関

表4.5では質問ID7~11においてすべての評価項目で「r\_cogproe」が高い相関を示した。「cogproe」はCognitive Processes (認知的プロセス) で「心がけ」や「図る」の単語が当てはまる。

周囲の人間からは、真面目な人であると言われることが多い**と思います**。(結論) **理由**としましては、私自身が**決して**その頭や効率がいい人間ではないという自負がありまして、それに**基づいて**人の見えていないところであったりだとか、或いは**本来**自分がやらなくてもいいようなタスクっていうものもあの影で実行しているような所があるからだ**と思います**。(理由)

この例は周囲の人からどう思われているかへの回答で評価は総合評価5, 論理性4.5, 言葉遣い5である。理由の部分で「理由」, 「決して」, 「基づいて」, 「本来」が「cogproc」カテゴリーである。このように結論と理由しか述べられていない場合でも、自らを洞察できていることをアピールできればよい評価を得られる場合もある。

### 具体例を含む回答のカテゴリーの出現頻度と評価の相関

表 4.6 では質問 ID 3～6 において「p\_bio」がすべての評価項目で負の値となっているが、これは「bio」カテゴリーの「生活」や「生きる」という単語によって起きていると考えられる。仕事選びの軸ややりたい仕事において自分がただ生活できれば良いといった回答や生活に支障がないよう働きたいといった回答は、消極的な回答とみなされ低い評価を受ける。

### まとめの結論を含む回答のカテゴリーの出現頻度と評価の相関

表 4.7 によると質問 ID7～11 で「p\_negate」が負の相関であるにもかかわらず、「p2\_negate」は正の相関を示している。「negate」は否定語で「ない」や「ず」が該当する単語であるが、自分の弱みを聞く質問においてこの傾向がよくみられる。最初の結論において否定語が評価が低い回答は「朝起きられない」や「計画性がない」など社会人として問題があるような短所を述べてしまうものが多かった。反対にまとめの結論で否定語が用いられるのは最初の結論で述べた短所を打ち消すための工夫について述べられていることが多いため高い評価につながると考えられる。

表 4.4: 回答文全体の J-LIWC の出現頻度と評価の相関係数

質問 ID1,2					
総合評価		論理性		言葉遣い	
カテゴリー	出現頻度	カテゴリー	出現頻度	カテゴリー	出現頻度
all_bio	0.393	all_tentat	0.352	all_negemo	0.325
all_body	0.365	all_bio	0.333	all_bio	0.288
p2_casepart	0.337	all_leisure	0.324	all_discrep	0.286
r_function	0.320	all_body	0.307	r_function	0.277
all_leisure	0.317	all_health	0.250	r_casepart	0.274
質問 ID3,4,5,6					
p2_function	0.277	all_anger	0.258	r_casepart	0.206
all_anger	0.272	all_percept	0.233	all_feel	0.201
r_casepart	0.261	p2_function	0.227	r_function	0.196
p2_casepart	0.258	r_casepart	0.225	r_verb	0.196
r_function	0.253	all_verb	-0.221	r_informal	0.192
質問 ID7,8,9,10,11					
r_verb	0.200	all_relativ	0.170	r_verb	0.203
r_casepart	0.196	all_social	-0.169	r_casepart	0.181
all_social	-0.196	all_time	0.167	r_function	0.172
r_function	0.188	all_negemo	-0.156	all_negemo	-0.148
all_work	0.176	r_verb	0.143	r_cogproc	0.148

表 4.5: 理由がある回答の J-LIWC の出現頻度と評価の相関係数

質問 ID1,2					
総合評価		論理性		言葉遣い	
カテゴリー	出現頻度	カテゴリー	出現頻度	カテゴリー	出現頻度
r_affect	0.244	p_cause	-0.241	p_relig	0.288
p_casepart	0.219	r_discrep	0.210	p_discrep	0.273
p_discrep	0.217	p_relig	0.209	r_discrep	0.266
r_conj	0.211	r_ipron	-0.195	p_negemo	0.227
r_filler	0.194	r_affect	0.179	r_social	0.218
質問 ID3,4,5,6					
r_verb	0.209	r_verb	0.180	r_verb	0.162
r_casepart	0.209	r_casepart	0.178	r_casepart	0.152
r_function	0.195	r_function	0.158	r_function	0.137
r_informal	0.193	r_time	0.142	r_informal	0.135
r_filler	0.159	r_informal	0.138	p_insight	0.125
質問 ID7,8,9,10,11					
r_cogproc	0.225	r_cogproc	0.201	r_cogproc	0.229
r_verb	0.205	r_verb	0.162	r_verb	0.204
r_casepart	0.203	p_conj	0.159	r_casepart	0.189
r_function	0.193	r_cause	0.136	r_function	0.169
p_conj	0.162	r_casepart	0.133	p_conj	0.162

表 4.6: 具体例がある回答の J-LIWC の出現頻度と評価の相関係数

質問 ID1,2					
総合評価		論理性		言葉遣い	
カテゴリー	出現頻度	カテゴリー	出現頻度	カテゴリー	出現頻度
p_money	0.187	p_money	0.201	p_relig	0.193
e_relativ	0.180	p_percept	-0.151	p_negate	-0.171
e_time	0.174	e_work	0.148	p_money	0.168
e_space	0.164	p_negate	-0.131	p_discrep	0.160
e_casepart	0.146	e_ingest	-0.127	p_percept	-0.156
質問 ID3,4,5,6					
p_bio	-0.256	p_drives	0.308	p_bio	-0.275
p_reward	0.205	p_male	-0.242	p_affiliation	0.220
e_i	0.201	p_bio	-0.240	p_reward	0.202
p_negate	0.190	p_affiliation	0.237	e_verb	0.187
p_informal	0.172	p_ingest	-0.227	e_i	0.181
質問 ID7,8,9,10,11					
p_negate	-0.143	e_function	-0.198	p_negate	-0.166
e_negemo	0.104	e_auxverb	-0.194	p_certain	0.104
p_number	-0.098	e_casepart	-0.188	e_feel	-0.101
e_see	0.096	p_negate	-0.188	e_auxverb	-0.099
p_auxverb	0.094	e_conj	-0.179	p_pronoun	-0.099

表 4.7: まとめの結論がある回答の J-LIWC の出現頻度と評価の相関係数

質問 ID1,2					
総合評価		論理性		言葉遣い	
カテゴリー	出現頻度	カテゴリー	出現頻度	カテゴリー	出現頻度
p2_casepart	0.365	p2_casepart	0.244	p2_casepart	0.293
p2_function	0.340	p2_drives	0.228	p2_function	0.263
p2_verb	0.290	p2_function	0.224	p2_quant	0.248
p2_relativ	0.281	p2_social	0.220	p2_cause	0.248
p2_informal	0.277	p2_relativ	0.210	p2_verb	0.240
質問 ID3,4,5,6					
p_home	0.207	p_see	0.197	p_insight	0.173
p_see	0.163	p_home	0.183	p2_filler	0.159
p_time	-0.160	p_male	-0.156	p2_nonflu	0.159
p_number	0.151	p_insight	0.143	p2_function	0.157
p_insight	0.147	p_filler	0.142	p2_casepart	0.145
質問 ID7,8,9,10,11					
p2_negate	0.137	p_negate	-0.150	p2_negate	0.146
p_negate	-0.130	p_work	0.134	p_negate	-0.145
p2_relig	0.126	p_ipron	-0.131	p2_feel	0.116
p_ipron	-0.124	p_social	-0.113	p_differ	0.115
p_interrog	-0.122	p_interrog	-0.108	p2_relig	0.111

## 第5章 実験と評価

本節ではPREP 話法をもとに作成した特徴量が評価の予測にどのように寄与するかを分析する，BERTによる特徴量とPREP 話法によって得られるJ-LIWCの特徴量を線形回帰モデルによってスコア予測を行い特徴量の組み合わせによって精度にどのような影響を与えるか調べる。

### 5.1 特徴量

3.2 節の面接の回答からBERTと4章で用いたJ-LIWCのカテゴリによって特徴量を作成する。

BERT[1]はTransformer[16]アーキテクチャを基盤として双方向の言語表現学習を行うことができるため文章の文脈を考慮することができ，自然言語処理の様々なタスクにおいて高い精度を達成している。まず，3.2 節の面接の回答の書きおこし文をMeCabでトークン化しサブワードによって分割した。次に，事前学習された日本語のBERTモデル<sup>1</sup>によって出力された最終層のひとつ前の隠れ層を取得し768次元の言語特徴量を得た。

4章で取得したそれぞれの回答のJ-LIWCの各カテゴリの出現頻度を特徴量とした。PREP 話法の攻勢を反映するため，ある回答文全体の出現頻度を取得し69次元で特徴量を作り，一方でその回答文をPREPに分割したときそれぞれの要素の文で同様に出現頻度を特徴量とし69times4=276次元の特徴量を得た。特徴量はこの3種類を組み合わせる。

### 5.2 使用した回帰モデルと実験設定

回帰モデルはラッソ回帰，リッジ回帰，サポートベクトル回帰（SVR）を用いて回答1つを1サンプルとして質問の種類ごとに学習した。

実験設定はラッソ回帰モデルにおけるパラメータは，K-fold（K=3）クロスバリデーションで最適化し， $\alpha$ パラメータ値は[0.001, 0.01, 0.1, 1, 10, 100]から選択した。リッジ回帰モデルのパラメータは，K-fold（K=3）クロスバリデーションで最適化し， $\alpha$ パラメータ値は[0.001, 0.01, 0.1, 1, 10, 100]から選択した。線形SVRの

---

<sup>1</sup><https://github.com/cl-tohoku/bert-japanese>

パラメータも同様に、K-fold (K=3) クロスバリデーションで最適化し  $C$  と  $\epsilon$  のパラメータ値は [0.001, 0.01, 0.1, 1, 10, 100] の中から選択した。

回帰性能の評価としてピアソンの相関係数 ( $r$ ) と二乗平均平方根誤差 (RMSE) を使用した。また、この実験では一個抜き交差検証を行っている。

## 5.3 結果

表 5.1, 5.2, 5.3 に回帰の精度を示す。相関係数は負の値となっている場合があり、これは回帰に失敗したためであるのでこのような失敗例では  $r = 0$  とした。特徴量は BERT のみ, J-LIWC のみ, PREP の要素を含んだ J-LIWC (J-LIWC (PREP)), BERT と J-LIWC, BERT と PREP を含んだ J-LIWC (J-LIWC (PREP)) の 5 種の組み合わせで行った, BERT と J-LIWC, BERT と PREP を含んだ J-LIWC の二つの特徴量による精度では, BERT と PREP を含んだ J-LIWC のほうがどのモデルにおいても向上した。すべての質問の中で最も高い精度を出したのは総合評価ではラッソ回帰の J-LIWC (PREP) と BERT + J-LIWC (PREP) の質問 ID4 で  $r=0.818$ , 論理性ではラッソ回帰の J-LIWC (PREP) と BERT + J-LIWC (PREP) の質問 ID4 で  $r=0.833$ , 言葉遣いではラッソ回帰の J-LIWC (PREP) の質問 ID8 の  $r=0.748$  だった。これらの結果から PREP に基づいた特徴量は評価の予測において多く寄与するといえる。

表 5.1: ラッソ回帰の精度表

質問 ID	BERT						J-LIWC						J-LIWC(PREP)						BERT+JLIWC						BERT+JLIWC(PREP)					
	総合評価		論理性		言葉遣い		総合評価		論理性		言葉遣い		総合評価		論理性		言葉遣い		総合評価		論理性		言葉遣い		総合評価		論理性		言葉遣い	
	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE	r	RMSE
1	0	0.657	0	0.942	0.087	0.664	0	0.7	0	0.949	0	0.68	0.056	0.685	0	0.96	0	0.698	0	0.731	0	0.698	0	0.731	0.056	0.685	0	0.96	0	0.698
2	0	0.827	0	0.917	0	0.582	0	0.849	0	0.981	0.057	0.587	0.087	0.855	0.37	0.883	0	0.603	0	0.887	0	1.04	0.057	0.587	0.11	0.85	0.369	0.885	0	0.603
3	0	0.907	0	0.987	0	0.878	0	0.791	0	0.972	0	0.843	0.572	0.628	0.298	0.908	0.605	0.649	0	0.791	0	0.972	0	0.843	0.572	0.628	0.298	0.908	0.605	0.649
4	0.123	1.294	0	1.38	0.028	1.015	0.016	1.331	0.018	1.464	0.251	0.934	0.818	0.699	0.833	0.746	0.671	0.7	0.005	1.337	0.011	1.466	0.251	0.934	0.818	0.699	0.833	0.746	0.671	0.7
5	0	1.471	0	1.307	0	1.083	0	1.471	0	1.307	0	1.096	0	1.471	0	1.307	0	1.083	0	1.471	0	1.307	0	1.083	0	1.471	0	1.307	0	1.083
6	0	0.862	0.061	0.722	0	0.918	0	0.905	0	0.737	0	0.799	0	0.848	0.06	0.751	0.731	0.536	0	0.885	0	0.737	0	0.799	0	0.848	0.06	0.751	0.731	0.536
7	0	0.748	0	0.632	0	0.651	0	0.777	0	0.652	0	0.632	0	0.698	0	0.632	0	0.641	0	0.777	0	0.652	0	0.632	0	0.698	0	0.632	0	0.641
8	0	1.347	0	1.182	0	0.842	0	1.637	0	1.181	0	0.85	0.079	1.391	0.369	1.104	0.748	0.555	0	1.551	0	1.181	0	0.85	0.102	1.375	0.369	1.104	0.738	0.57
9	0	0.762	0	0.692	0	0.554	0	0.762	0	0.655	0	0.554	0	0.762	0	0.655	0	0.554	0	0.762	0	0.655	0	0.554	0	0.762	0	0.655	0	0.554
10	0	0.592	0	0.632	0	0.628	0	0.595	0	0.595	0	0.572	0	0.603	0.394	0.559	0	0.572	0	0.595	0	0.595	0	0.572	0	0.603	0.358	0.586	0	0.572
11	0.074	0.895	0	0.943	0	0.779	0	0.906	0	0.932	0.025	0.81	0.33	0.838	0.502	0.782	0.476	0.671	0	0.906	0	0.928	0.045	0.808	0.33	0.838	0.502	0.782	0.476	0.671



表 5.3: SVR の精度表

質問 ID	BERT						J-LIWC						J-LIWC(PREP)						BERT+J-LIWC						BERT+J-LIWC(PREP)											
	総合評価			論理性			言葉遣い			総合評価			論理性			言葉遣い			総合評価			論理性			言葉遣い			総合評価			論理性			言葉遣い		
	r	RMSE	r	r	RMSE	r	r	RMSE	r	r	RMSE	r	r	RMSE	r	r	RMSE	r	r	RMSE	r	r	RMSE	r	r	RMSE	r	r	RMSE	r	r	RMSE	r	r	RMSE	r
1	0	0.768	0	1.051	0	0.68	0.09	0.667	0.017	1.087	0	0.742	0.276	0.635	0.281	0.905	0.036	0.672	0.108	0.657	0.167	0.961	0	0.729	0.26	0.635	0.259	0.909	0	0.689						
2	0.091	0.798	0	0.915	0.099	0.567	0	0.845	0	2.15	0	1.326	0.266	0.784	0	0.928	0.379	0.529	0.105	0.806	0	1.185	0.346	0.573	0.33	0.765	0	0.929	0.41	0.521						
3	0	0.765	0.092	0.957	0.064	0.834	0	0.766	0.081	0.986	0.056	0.83	0.114	0.961	0.088	1.035	0.477	0.91	0	0.766	0.075	1.099	0.209	0.813	0.24	0.835	0.142	1.011	0.576	0.732						
4	0.093	1.276	0	1.353	0.24	0.929	0.153	1.662	0	1.353	0.24	0.941	0.56	1.253	0.63	1.246	0.581	0.772	0.127	1.305	0	1.378	0.24	0.942	0.626	1.063	0.689	1.063	0.59	0.768						
5	0	1.449	0	1.309	0.055	1.065	0	1.449	0	1.309	0	1.069	0	1.449	0	1.309	0.393	0.985	0	1.449	0	1.309	0	1.068	0	1.449	0	1.309	0.394	0.983						
6	0	0.973	0	0.845	0	0.997	0.185	0.868	0.181	0.754	0	1.167	0.111	1.301	0.114	1.182	0.479	0.884	0	0.923	0.02	0.812	0.115	0.939	0.18	1.09	0.169	0.986	0.531	0.764						
7	0	0.84	0	0.636	0.166	0.638	0	0.814	0	0.636	0.107	0.611	0.202	0.908	0	0.636	0.315	0.621	0	0.82	0	0.636	0	0.654	0.256	0.805	0	0.636	0.312	0.62						
8	0.196	1.322	0	1.344	0.105	0.864	0.214	2.12	0.08	2.132	0	0.876	0.303	1.302	0.536	1.037	0.616	0.657	0.14	1.371	0.13	1.376	0.114	0.857	0.301	1.303	0.554	1.018	0.565	0.706						
9	0	1.035	0	0.666	0	0.61	0	0.782	0	0.661	0	0.592	0.131	0.905	0.223	0.724	0	0.569	0	0.791	0	0.661	0	0.642	0.115	0.86	0.286	0.648	0	0.566						
10	0	0.213	0.601	0	0.701	0.091	0.609	0	1.002	0	0.595	0	1.499	0.401	0.562	0.381	0.585	0.392	0.558	0	0.758	0	0.595	0	0.81	0.311	0.583	0.305	0.577	0.265	0.595					
11	0	1.198	0	1.058	0	0.791	0	0.957	0.161	0.904	0.309	0.738	0.072	1.05	0.605	0.723	0	0.775	0	0.906	0.15	0.91	0	0.774	0.159	0.987	0.6	0.726	0	0.78						

# 第6章 PREP アノテーションの自動化

## 6.1 大規模言語モデルと議論マイニング

大規模言語モデルによって議論マイニングを行う研究は現在進展しており, Pojoni ら [10] では Podcasts の書きおこしに対し ChatGPT を用いて主張や反論といった議論の構造を取り出そうとした. GPT モデル [11] は OPENAI の大規模言語モデルで GPT-4 はその前身の GPT-3.5 よりも大幅に改善されており特にさまざまな言語においても優れた性能を示している [3].

本研究では人手で行っていた PREP アノテーションを GPT-4 を用いることで自動的にを行い, 将来的にリアルタイムで PREP 話法に則った構造の回答かを分析しより詳細なフィードバックにつなげることを目指す.

## 6.2 手法

PREP の自動分類を行うために 3 章で用いたデータセット 47 人分 517 回答に対しアノテーションスキームやアノテーションの例によって作成したプロンプトをもとに GPT-3.5 と GPT-4 によって PREP 話法の 4 要素結論, 理由, 具体例, まどめの結論の分類を行う. Python の ChatGPT ラッパー<sup>1</sup>によって GPT を使用し以下のような出力を得ること目指した.

私の強みは忍耐力です.: 結論  
なぜなら私は苦手なことや難しいことであっても継続して取り組み成果を出す  
ことができるからです. : 理由  
例えば高校生の時は英語が苦手だったのですが, 大学生になってから毎日 3 時  
間継続して勉強することで TOEIC で 800 点を取ることができました.: 具体例  
こうした経験から何事にも私は粘り強く努力する力が身につきました.: まどめ  
の結論

この分類結果を人手によりアノテーションとの類似度により評価し, GPT-3.5 と GPT4 の性能比較とプロンプトの検討を行う.

<sup>1</sup><https://github.com/llm-workflow-engine/llm-workflow-engine>

## 6.2.1 プロンプト

GPT-4をはじめとする大規模言語モデルで正確な答えを得るには効果的なプロンプトを作成することが重要である。例えばChain of Thought (思考の連鎖) 推論にステップを踏ませることで推論精度を向上させた [18].

本研究では以下の4つのプロンプトを使用した。

1. PREP 話法の定義と分類例 1 つ
2. PREP 話法の定義と分類例 5 つ
3. PREP 話法の定義と分類例 10 つ
4. PREP 話法の定義と推論のステップを踏ませるプロンプト

それぞれの詳細なプロンプトは付録に掲載する。プロンプト 1 から 3 では 3.1 節で定義した PREP 話法の各要素の定義に加え、複数の PREP の分類例をプロンプトにした。プロンプト 4 では PREP 話法の定義と段階的に結論、理由を探すようなプロンプトで構成した。これは PREP 話法は各要素の間に支持関係があるため、ステップを踏んでそれぞれの要素を推論させることが重要だと考えられるためである。

## 6.2.2 評価

GPT が出力した結果を評価するためにレーベンシュタイン距離を用いて人手による分類との類似度を求める。レーベンシュタイン距離はある文字列をもう一方の文字列に変換するときの挿入、削除、置換を行う最小回数で定義される。このレーベンシュタイン距離を文字列の長さで割り 1 から引くことで 2 つの文字列の類似度を 0 から 1 であらわすことができる。まず、GPT が出力した各要素の文とアノテータがラベル付けした分の類似度を計算する、そして、5 人のアノテータでそれぞれ類似度を求め平均をとり、各要素の人手によるアノテーションとの類似度を求める。

## 6.3 結果

このようにして求めた各プロンプトの評価を表 6.1 に示す。まず、GPT-4 と GPT-3.5 を比較するとまとめの結論以外の要素では評価が改善している。一方でまとめの結論については両方のモデルで分類に失敗することが多かった。

図 6.1 は人手によるアノテーションと GPT による分類を比較した図でオレンジが結論、黄緑が具体例、黄色がまとめの結論を表しており、上部が人手で下部が GPT である。このように最後の文章を結論とすることによる誤りが多くみられた。

プロンプト間においても大きく変化することはなかったが、プロンプト 1~3 で事例を増やしていったにもかかわらず、例の数が少ないプロンプトのほうが類似度が高い場合もあった。

人手  
GPT

私が第一志望にしている会社は農機具メーカーの最大手であるクボタです。ライバル会社であるヤンマーと比較してクボタは農機具のスマート化やスマート農業に積極的に注力しており海外シェアも高く特に北米事業では、大規模農業のスマート化としてセンサーや衛星データ、衛星から取得した位置情報を元に農機具の無人化といった最先端の農業研究にも注力しています。そのため私はクボタを志望しております。

図 6.1: 人手によるアノテーションと GPT によるアノテーションの比較

表 6.1: GPT の各プロンプトごとの評価

	GPT-4				GPT-3.5			
	prompt1	prompt2	prompt3	prompt4	prompt1	prompt2	prompt3	prompt4
結論	0.839	0.714	0.719	0.712	0.492	0.509	0.500	0.511
理由	0.648	0.651	0.644	0.650	0.518	0.476	0.468	0.458
具体例	0.559	0.654	0.563	0.567	0.421	0.356	0.365	0.352
まとめの結論	0.249	0.250	0.252	0.251	0.285	0.265	0.261	0.270

## 第7章 終わりに

本研究ではPREP話法に基づく就職面接の回答が面接評価にどのような影響をもたらすかについて分析を行った。就職面接訓練の発話の書きおこしデータに対し、アノテータ5名によるPREPの各要素を付与しデータコーパスを作成した。そして、面接官による評価がPREPの要素の有無によって差があるかを検定し、ほとんどの評価項目において有意差が認められた。また、PREPの各要素に含まれるJ-LIWCの各カテゴリーの出現頻度を取得し、評価との相関について調べどのようなカテゴリーの単語が評価にかかわっているかを明らかにした。次に線形回帰モデルによって評価の推定を行い、PREP話法を利用した特徴量が予測に寄与したことを示した。最後に、GPTによるPREPアノテーションの自動化に取り組み、4つのプロンプトについてどれほど人手に近づいた分類が行えるかを検討した。

本研究の今後の課題としては、実際にPREP話法を取り入れるように指導したときどの程度面接スキルが向上するかを調べることにある。そして、PREP話法による指導が有効であるとしたとき、どのようなフィードバックが有効であるかを検討する必要がある。被験者実験を通じた研究が必要である。また、発話の内容についてもより詳細な分析ができ、就職面接でよく使われる単語に特化した辞書を制作すれば面接評価の差について詳しく調べることができるだろう。

# 謝辞

主指導教員の岡田将吾准教授には日頃より多大なご助言，ご指導をいただきました。深く感謝を申し上げます。

また，副指導教員の井之上直也准教授にもとくにアノテーションに関して多くのご助言をいただきました。お礼を申し上げます。

キャリアカウンセラーの松木様，四柳様におかれましても研究において様々なご相談をさせていただきました。本当にありがとうございました。

# 付録A プロンプト

## A.1 プロンプト1

次のテキストを「結論」「理由」「具体例」「まとめの結論」の4つのラベルに分類してください。該当する箇所がない場合はなしとします。一つの文章に複数のラベルが付く場合もあります。

分類のラベルは以下の定義に従ってください。

結論: 全体で最も伝えたい重要な部分。

理由: 結論を支持するもっとも端的な説明。

具体例: 理由を裏付けるデータや体験談などの個別の例。

まとめの結論: 最初に述べた結論を繰り返したもの。

以下は答え方の例です  
例に従って回答してください

テキスト

私の強みは忍耐力です。なぜなら私は苦手なことや難しいことであっても継続して取り組み成果を出すことができるからです。例えば高校生の時は英語が苦手だったのですが、大学生になってから毎日3時間継続して勉強することでTOEICで800点を取ることができました。こうした経験から何事にも私は粘り強く努力する力が身につきました。

私の強みは忍耐力です: 結論

なぜなら私は苦手なことや難しいことであっても継続して取り組み成果を出すことができるからです.: 理由

例えば高校生の時は英語が苦手だったのですが、大学生になってから毎日3時間継続して勉強することでTOEICで800点を取ることができました.: 具体例

こうした経験から何事にも私は粘り強く努力する力が身につきました.: まとめ  
の結論

では次のテキストについて4つのラベルに分類してください

テキスト

回答は原則として各センテンスごとにラベルを示してください

## A.2 プロンプト2

次のテキストを「結論」「理由」「具体例」「まとめの結論」の4つのラベルに分類してください。該当する箇所がない場合はなしとします。一つの文章に複数のラベルが付く場合もあります。

分類のラベルは以下の定義に従ってください。

結論: 全体で最も伝えたい重要な部分。

理由: 結論を支持するもっとも端的な説明。

具体例: 理由を裏付けるデータや体験談などの個別の例。

まとめの結論: 最初に述べた結論を繰り返したもの。

以下は答え方の例です  
例に従って回答してください

テキスト

私の強みは忍耐力です。なぜなら私は苦手なことや難しいことであっても継続して取り組み成果を出すことができるからです。例えば高校生の時は英語が苦手だったのですが、大学生になってから毎日3時間継続して勉強することでTOEICで800点を取ることができました。こうした経験から何事にも私は粘り強く努力する力が身につきました。

私の強みは忍耐力です：結論

なぜなら私は苦手なことや難しいことであっても継続して取り組み成果を出すことができるからです。：理由

例えば高校生の時は英語が苦手だったのですが、大学生になってから毎日3時間継続して勉強することでTOEICで800点を取ることができました。：具体例

こうした経験から何事にも私は粘り強く努力する力が身につきました。：まとめの結論

テキスト

私に対人関係で大事にしていることは相手の気持ちを尊重しつつ自分の価値観と決して合わなかったとしてもそれをあいつらはそれで相手価値観であり、自分とは違うんだという風に一旦受け入れるということを大事にしています。

私に対人関係で大事にしていることは相手の気持ちを尊重しつつ自分の価値観と決して合わなかったとしてもそれをあいつらはそれで相手価値観であり、自分とは違うんだという風に一旦受け入れるということを大事にしています。：結論

テキスト

”周りの友人や、アルバイト先のお世話になった方々からは明るくて真面目だという風におっしゃっていただけることが多いと思います。心がけていることは、どんな方にもどのような人に対しても笑顔で明るく接することを心がけております。

周りの友人や、アルバイト先のお世話になった方々からは明るくて真面目だという風におっしゃっていただけることが多いと思います。:結論  
心がけていることは、どんな方にもどのような人に対しても笑顔で明るく接することを心がけております。:理由

#### テキスト

チームの中では、どちらかと言うと、みんなの意見を纏めて、あと必要な情報を集めたりするというのが好きで、そういう立場についています。ただ、リーダーではなく、リーダーを補佐するまたは、そのリーダーや他のメンバーとの間を取り持つというようなポジションに好んでついていきます。情報を集めたりみんなの意見を取りまとめるとか、重大な意思決定をする大きく方向を決めるということは苦手なんですけれども、必要な情報を集めたり、その時間や場所、全員の意見を取りまとめた上で、ここがベストなんじゃないかっていう選択肢を探すっていうのが得意なので、そうですね、リーダーを補佐する、または人と人との仲介をするというポジションにつくのが向いていると、自分では思っております。

チームの中では、どちらかと言うと、みんなの意見を纏めて、あと必要な情報を集めたりするというのが好きで、そういう立場についています。:結論  
ただ、リーダーではなく、リーダーを補佐するまたは、そのリーダーや他のメンバーの間を取り持つというようなポジションに好んでついていきます。:結論  
情報を集めたりみんなの意見を取りまとめるとか、重大な意思決定をする大きく方向を決めるということは苦手なんですけれども、必要な情報を集めたり、その時間や場所、全員の意見を取りまとめた上で、ここがベストなんじゃないかっていう選択肢を探すっていうのが得意なので、:理由  
そうですね、リーダーを補佐する、または人と人との仲介をするというポジションにつくのが向いていると、自分では思っております。:まとめの結論

テキスト私はチームの中では、サポーターに回ることが多いです。理由としては、やはりリーダーとして周りを引っ張っていくというものも大事だと思うんですけど、チームの一員であるというポジションが、逆に自由に動けてよりそのチームを支えられることがあると、私自身考えているため、そのようなチームの中でのポジションは取ることが多いです。例としては高校の時の学園祭の活動などでは、敢えてリーダーではなく副リーダーになることで、リーダーではできないような仕事などに、関わってチームを支えたりはしました。

私はチームの中では、サポーターに回ることが多いです。:結論

理由としては、やはりリーダーとして周りを引っ張っていくというものも大事だと思うんですけど、チームの一員であるというポジションが、逆に自由に動けてよりそのチームを支えられることがあると、私自身考えているため、そのようなチームの中でのポジションは取ることが多いです。:理由

例としては高校の時の学園祭の活動などでは、敢えてリーダーではなく副リーダーになることで、リーダーではできないような仕事などに、関わってチームを支えたりはしました。:具体例

では次のテキストについて4つのラベルに分類してください

テキスト

回答は原則として各センテンスごとにラベルを示してください

### A.3 プロンプト3

次のテキストを「結論」「理由」「具体例」「まとめの結論」の4つのラベルに分類してください。該当する箇所がない場合はなしとします。一つの文章に複数のラベルが付く場合もあります。

分類のラベルは以下の定義に従ってください。

結論: 全体で最も伝えたい重要な部分。

理由: 結論を支持するもっとも端的な説明。

具体例: 理由を裏付けるデータや体験談などの個別の例。

まとめの結論: 最初に述べた結論を繰り返したもの。

以下は答え方の例です

例に従って回答してください

テキスト

私の強みは忍耐力です。なぜなら私は苦手なことや難しいことであっても継続して取り組み成果を出すことができるからです。例えば高校生の時は英語が苦手だったのですが、大学生になってから毎日3時間継続して勉強することでTOEICで800点を取ることができました。こうした経験から何事にも私は粘り強く努力する力が身につきました。

私の強みは忍耐力です: 結論

なぜなら私は苦手なことや難しいことであっても継続して取り組み成果を出すことができるからです。:理由

例えば高校生の時は英語が苦手だったのですが、大学生になってから毎日3時間継続して勉強することでTOEICで800点を取ることができました。:具体例

こうした経験から何事にも私は粘り強く努力する力が身につきました。:まとめの

## 結論

### テキスト

私に対人関係で大事にしてることは相手の気持ちを尊重しつつ自分の価値観と決して合わなかったとしてもそれをあいつらはそれで相手価値観であり、自分とは違うんだという風に一旦受け入れるということを大事にしています。

私に対人関係で大事にしてることは相手の気持ちを尊重しつつ自分の価値観と決して合わなかったとしてもそれをあいつらはそれで相手価値観であり、自分とは違うんだという風に一旦受け入れるということを大事にしています。:結論

### テキスト

”周りの友人や、アルバイト先のお世話になった方々からは明るくて真面目だという風におっしゃっていただけることが多いと思います。心がけていることは、どんな方にもどのような人に対しても笑顔で明るく接することを心がけております。

周りの友人や、アルバイト先のお世話になった方々からは明るくて真面目だという風におっしゃっていただけることが多いと思います。:結論  
心がけていることは、どんな方にもどのような人に対しても笑顔で明るく接することを心がけております。:理由

### テキスト

チームの中では、どちらかと言うと、みんなの意見を纏めて、あと必要な情報を集めたりするというのが好きで、そういう立場についています。ただ、リーダーではなく、リーダーを補佐するまたは、そのリーダーや他のメンバーとの間を取り持つというようなポジションに好んでついていきます。情報を集めたりみんなの意見を取りまとめるとか、重大な意思決定をする大きく方向を決めるということは苦手なんですけれども、必要な情報を集めたり、その時間や場所、全員の意見を取りまとめた上で、ここがベストなんじゃないかっていう選択肢を探すっていうのが得意なので、そうですね、リーダーを補佐する、または人と人との仲介をするというポジションにつくのが向いていると、自分では思っております。

チームの中では、どちらかと言うと、みんなの意見を纏めて、あと必要な情報を集めたりするというのが好きで、そういう立場についています。:結論  
ただ、リーダーではなく、リーダーを補佐するまたは、そのリーダーや他のメンバーとの間を取り持つというようなポジションに好んでついていきます。:結論  
情報を集めたりみんなの意見を取りまとめるとか、重大な意思決定をする大きく方向を決めるということは苦手なんですけれども、必要な情報を集めたり、その時間や場所、全員の意見を取りまとめた上で、ここがベストなんじゃないかって

いう選択肢を探すっていうのが得意なので、：理由

そうですね、リーダーを補佐する、または人と人との仲介をするというポジションにつくのが向いていると、自分では思っております。：まとめの結論

テキスト 私はチームの中では、サポーターに回ることが多いです。理由としては、やはりリーダーとして周りを引っ張っていくというものも大事だと思うんですけど、チームの一員であるというポジションが、逆に自由に動けてよりそのチームを支えることがあると、私自身考えているため、そのようなチームの中でのポジションは取ることが多いです。例としては高校の時の学園祭の活動などでは、敢えてリーダーではなく副リーダーになることで、リーダーではできないような仕事などに、関わってチームを支えたりはしました。

私はチームの中では、サポーターに回ることが多いです。：結論

理由としては、やはりリーダーとして周りを引っ張っていくというものも大事だと思うんですけど、チームの一員であるというポジションが、逆に自由に動けてよりそのチームを支えることがあると、私自身考えているため、そのようなチームの中でのポジションは取ることが多いです。：理由

例としては高校の時の学園祭の活動などでは、敢えてリーダーではなく副リーダーになることで、リーダーではできないような仕事などに、関わってチームを支えたりはしました。：具体例

テキスト

対人関係で一番意識していることは、相手へのリスペクト尊敬です。相手がどのような体験をしてきたのか、経験を積んできたのか、また何を学んで、どういう価値観を持っているのかというのを理解することで相手のことをより、客観的にも見ることが出来ると考えております。その客観的な視点から相手を尊重しながら話すことによって、お互い信頼関係を生むこともできると考えております

対人関係で一番意識していることは、相手へのリスペクト尊敬です。：結論

相手がどのような体験をしてきたのか、経験を積んできたのか、また何を学んで、どういう価値観を持っているのかというのを理解することで相手のことをより、客観的にも見ることが出来ると考えております。：理由

その客観的な視点から相手を尊重しながら話すことによって、お互い信頼関係を生むこともできると考えております。：具体例

なし：まとめの結論

テキスト

私の就活の軸は、やはりそのやりたい仕事ができることが一番最初ですね。やりたいことは半導体なんですけど、元々私は半導体が好きで、自作PCの組み立てがきっかけで半導体に興味を持ったんですけど、やりたいこと社会人経験ってすごい、一生長いものなので、やはりやりたいことができるということは長く続けられるモチベーションを保つことができるってことなんで、そういった意味では、やっぱり自分自身も、人生充実できるし、その相手側、会社側も、そうして長く

続けていただければ嬉しいと思うんで、お互い winwin の関係になると思うので、私の就活の軸はやりたいことができるってことにおいて、フォーカスしといて就職活動をしています。

私の就活の軸は、やはりそのやりたい仕事ができることが一番最初ですね。：結論

やりたいことは半導体なんですけど、元々私は半導体が好きで、自作 PC の組み立てがきっかけで半導体に興味を持ったんですけど、：具体例

やりたいこと社会人経験ってすごい、一生長いものですので、やはりやりたいことができるということは長く続けられるモチベーションを保つことができるってことなんで、そういった意味では、やっぱり自分自身も、人生充実できるし、その相手側、会社側も、そうして長く続けていただければ嬉しいと思うんで、お互い winwin の関係になると思うので、：理由私の就活の軸はやりたいことができるってことにおいて、フォーカスしといて就職活動をしています。：まとめの結論

テキスト自分の長所は、現状の問題を的確に分析して、それを解決することができるような能力であるという風に考えています。自分が東芝という企業で四週間の研究開発のインターンシップに参加した際には、水処理施設の異常診断システムの改善というようなテーマを与えられました。その時担当の方と話し合い、初めはアルゴリズムの改善などによって異常診断の精度を改善しようということで、研究を進めていたのですが、なかなか精度を上げるということができませんでした。そこで、改めてその自らデーターの分析を行い測定しているそのセンサーデーターの中から異常のその識別にその寄与する測定データーと、寄与しないような測定データーっていうものがあるということ、自らその分析したことによってわかりました。また分析した結果から主体的に、自らそのような有効なデーターのみを活用できるようなそのアルゴリズムというものを導入することによって、最終的に現行のその異常診断システムよりも、高精度で、予測が出来るモデルというものを実現することができました。このような経験を通して、自分で主体的に問題というものを把握して、それに対して解決を行うということができるとい能力は、企業に入社後のその研究開発だったりする業務の中でも、活かすことができるのではないかと考えています。

自分の長所は、現状の問題を的確に分析して、それを解決することができるような能力であるという風に考えています。自分が東芝という企業で四週間の研究開発のインターンシップに参加した際には、水処理施設の異常診断システムの改善というようなテーマを与えられました。：結論

なし：理由

自分が東芝という企業で四週間の研究開発のインターンシップに参加した際には、水処理施設の異常診断システムの改善というようなテーマを与えられました。その時担当の方と話し合い、初めはアルゴリズムの改善などによって異常診断の精度を改善しようということで、研究を進めていたのですが、なかなか精度を上げるということができませんでした。そこで、改めてその自らデーターの分析を行

い測定しているそのセンサーデータの中から異常のその識別にその寄与する測定データと、寄与しないような測定データっていうものがあるということ、自らその分析したことによってわかりました。また分析した結果から主体的に、自らそのような有効なデータのみを活用できるようなそのアルゴリズムというものを導入することによって、最終的に現行のその異常診断システムよりも、高精度で、予測が出来るモデルというものを実現することができました。：具体例  
このような経験を通して、自分で主体的に問題というものを把握して、それに対して解決を行うということができるという能力は、企業に入社後のその研究開発だったりする業務の中でも、活かすことができるのではないかと考えています。：まよめの結論

テキスト

私は、自分の短所は一つのこと集中しすぎて、その周りが見えなくなってしまうというふうなことであるという風に考えています。大学における研究であったり課題においても一つの作業にその没頭しすぎてしまい、最終的に、締め切りのそのギリギリになって、その慌ててしまうというふうなことが、今までに数多くありました。’,’そこで、このような短所を自覚して、自分の中では、複数のタスクについて、あらかじめ、計画し、スケジュールリングを行うことによって余裕を持って物事に集中して取り組めるような環境というものを自分で作るように意識しています。

私は、自分の短所は一つのこと集中しすぎて、その周りが見えなくなってしまうというふうなことであるという風に考えています。：結論

なし：理由

大学における研究であったり課題においても一つの作業にその没頭しすぎてしまい、最終的に、締め切りのそのギリギリになって、その慌ててしまうというふうなことが、今までに数多くありました。：具体例

そこで、このような短所を自覚して、自分の中では、複数のタスクについて、あらかじめ、計画し、スケジュールリングを行うことによって余裕を持って物事に集中して取り組めるような環境というものを自分で作るように意識しています。：まよめの結論

テキスト

私はIT業界を志望しています。理由としては大学時代、大学院含めて、ITに関する技術を学んでいて実際学んでもこの技術を一生続けていきたいと考えたためです。

私はIT業界を志望しています。：結論

理由としては大学時代、大学院含めて、ITに関する技術を学んでいて実際学んでもこの技術を一生続けていきたいと考えたためです。：理由

では次のテキストについて4つのラベルに分類してくださいテキスト

回答は原則として各センテンスごとにラベルを示してください

## A.4 プロンプト4

次のテキストを結論、理由、具体例、まとめの結論の4つのラベルに分類してください。該当する箇所がない場合はなしとします。一つの文章に複数のラベルが付く場合もあります。

分類のラベルは以下の定義に従ってください。

結論: 全体で最も伝えたい重要な部分。

理由: 結論を支持するもっとも端的な説明。

具体例: 理由を裏付けるデータや体験談などの個別の例。

まとめの結論: 最初に述べた結論を繰り返したもの。

まず結論となる箇所を探してください

次にその結論を説明するものとしてもっとも端的なものを理由としてください

そして、その理由を裏付ける具体的なエピソードや根拠を示したものを具体例としてください

最後に結論で述べたことが述べられていればそれをまとめの結論としてください

回答は原則として各センテンスごとにラベルを示してください

回答は以下のように出力してください

テキスト: 結論

テキスト: 理由

テキスト: 具体例

テキスト: まとめ

## 参考文献

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.
- [2] Ryo Egawa, Gaku Morio, and Katsuhide Fujita. Annotating and analyzing semantic role of elementary units and relations in online persuasive arguments. In Fernando Alva-Manchego, Eunsol Choi, and Daniel Khashabi, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pp. 422–428, Florence, Italy, July 2019. Association for Computational Linguistics.
- [3] Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. Measuring massive multitask language understanding. *CoRR*, Vol. abs/2009.03300, , 2020.
- [4] Tasuku Igarashi, Shimpei Okuda, and Kazutoshi Sasahara. Development of the japanese version of the linguistic inquiry and word count dictionary 2015. *Frontiers in Psychology*, Vol. 13, , 3 2022.
- [5] John Lawrence and Chris Reed. Argument mining: A survey. *Computational Linguistics*, Vol. 45, No. 4, pp. 765–818, December 2019.
- [6] Hiroki Nakayama, Takahiro Kubo, Junya Kamura, Yasufumi Taniguchi, and Xu Liang. doccano: Text annotation tool for human, 2018. Software available from <https://github.com/doccano/doccano>.
- [7] Huy V. Nguyen and Diane J. Litman. Argument mining for improving the automated scoring of persuasive essays. In *AAAI Conference on Artificial Intelligence*, 2018.
- [8] Tomoya Ohba, Candy Mawalim, Shun Katada, Haruki Kuroki, and Shogo Okada. Multimodal analysis for communication skill and self-efficacy level estimation in job interview scenario. pp. 110–120, 12 2022.
- [9] Joonsuk Park and Claire Cardie. Identifying appropriate support for propositions in online user comments. In Nancy Green, Kevin Ashley, Diane Litman, Chris

- Reed, and Vern Walker, editors, *Proceedings of the First Workshop on Argumentation Mining*, pp. 29–38, Baltimore, Maryland, June 2014. Association for Computational Linguistics.
- [10] Mircea-Luchian Pojoni, Lorik Dumani, and Ralf Schenkel. Argument-mining from podcasts using chatgpt. In Lukas Malburg and Deepika Verma, editors, *Proceedings of the Workshops at the 31st International Conference on Case-Based Reasoning (ICCBR-WS 2023) co-located with the 31st International Conference on Case-Based Reasoning (ICCBR 2023), Aberdeen, Scotland, UK, July 17, 2023*, Vol. 3438 of *CEUR Workshop Proceedings*, pp. 129–144. CEUR-WS.org, 2023.
- [11] Alec Radford and Karthik Narasimhan. Improving language understanding by generative pre-training. 2018.
- [12] Wasifur Rahman, Sazan Mahbub, Asif Salekin, Md Kamrul Hasan, and Ehsan Hoque. Hirepreter: A framework for providing fine-grained interpretation for automated job interview analysis. In *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 1–5, 2021.
- [13] Vikram Ramanarayanan, Chee Wee Leong, Lei Chen, Gary Feng, and David Suendermann-Oeft. Evaluating speech, face, emotion and body movement time-series features for automated multimodal presentation scoring. pp. 23–30, 11 2015.
- [14] Takeshi Saga, Hiroki Tanaka, Hidemi Iwasaka, and Satoshi Nakamura. Objective prediction of social skills level for automated social skills training using audio and text information. pp. 467–471, 10 2020.
- [15] Fuhui Tian, Shogo Okada, and Katsumi Nitta. Analyzing eye movements in interview communication with virtual reality agents. In *Proceedings of the 7th International Conference on Human-Agent Interaction, HAI '19*, p. 3–10, New York, NY, USA, 2019. Association for Computing Machinery.
- [16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023.
- [17] Henning Wachsmuth, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. Computational argumentation quality assessment in natural language. In Mirella Lapata, Phil Blunsom, and Alexander Koller, editors, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume*

*l, Long Papers*, pp. 176–187, Valencia, Spain, April 2017. Association for Computational Linguistics.

- [18] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023.
- [19] 黒木春樹. スキル改善フィードバックに基づく就職面接訓練システムの構築と評価, 2022. 北陸先端科学技術大学院大学 先端科学技術研究科 修士論文 (未公刊) .
- [20] 森泰. 面接訓練対話エージェントシステムを用いた面接スキルの分析, 2021. 北陸先端科学技術大学院大学 先端科学技術研究科 修士論文 (未公刊) .
- [21] 大場知也. マルチモーダル面接対話コーパスを用いた被面接者の面接スキルと自信度の分析, 2023. 北陸先端科学技術大学院大学 先端科学技術研究科 修士論文 (未公刊) .
- [22] 友秀大嶋. 話すスキルUP すぐできる! 論理的な話し方 話の組み立て方が上手になる PREP 法の使い方. 日本能率協会マネジメントセンター, 単行本, 3 2013.
- [23] 大庭知也, 黒木春樹, マワリム ケンディオリビア, 岡田将吾. マルチモーダル機械学習モデルに基づく就職活動面接訓練フィードバックシステム効果の分析. 人工知能学会全国大会論文集, Vol. JSAI2023, pp. 3Q1OS19a05–3Q1OS19a05, 2023.