

議論参加者の陳述評価に向けた複数人議論コーパスの構築

塩田 宰[†] 嶋田 和孝^{††}

[†]九州工業大学大学院 情報工学府
福岡県飯塚市川津 680-4

^{††}九州工業大学大学院 情報工学研究院
福岡県飯塚市川津 680-4

E-mail: †{t_shiota,shimada}@pluto.ai.kyutech.ac.jp

あらまし 教育において複数人議論を活用する機会が増加しており、物事を明確かつ論理的に伝える力が重要になってきている。教育という観点では、議論参加者の陳述を公正かつ定量的に評価、フィードバックすることが必要になってくるが、正確な評価を行うことはコストが高い。書き言葉を対象にした論述の質の自動推定について様々な研究が行われているが、書き言葉の論述と対面対話での陳述は性質が大きく異なる。そのため、対面対話における陳述を自動評価するためには対話形式の議論を収録したコーパスを基にした研究を進めていく必要がある。本研究は、議論参加者の陳述の自動評価に向けて複数人議論を収録し、説得力をはじめとする6種類の評価値を議論参加者に付与したコーパスの構築を目指す。

キーワード 対話理解, 複数人議論コーパス, 陳述の質, 説得力

Constructing a Multiparty Argument Corpus for Argumentation Quality Assessment in Discussions

Tsukasa SHIOTA[†] and Kazutaka SHIMADA^{††}

[†] Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN

^{††} Department of Artificial Intelligence, Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN

E-mail: †{t_shiota,shimada}@pluto.ai.kyutech.ac.jp

Abstract Since group discussion and debate have been employed to educate and even evaluate students in educational fields, it is important to develop a system that automatically assess participants' statement quality in a group discussion. However, there have not been enough research to assess argumentation quality in interaction, as compared with written language. In this paper, we construct a multiparty argument corpus for automatic argumentation quality assessment in discussions. We annotate 6 types of dimensions about effectiveness, which is one of the classes of argumentation quality, for the participants in our corpus and report the result of the annotation.

Key words Conversation Understanding, Multiparty Argument Corpus, Argumentation Quality, Persuasiveness

1. はじめに

教育において、複数人による議論を頻繁に用いる協調学習やPBLのような能動的学習が注目を集めて久しい。近年では複数人による議論は学習の場に止まらず、大学の入学試験においても受験者の知識以外の能力を測る指標として活用され、合否に影響を与える重要な役割を担ってきている。そのため、明確かつ論理的に自身の主張やアイデアを相手に理解、納得しても

らう能力が重要になってきている。

議論が上手くなるように教育を行うことやコミュニケーション能力を測る試験を行うことを考えた場合、議論参加者の陳述の質や内容を評価し、フィードバックすることは必須である。Wachsmuthら[1]のサーベイでは、論証の評価対象は「Cogency (適切性)」「Effectiveness (有効性)」「Reasonableness (合理性)」の大きな3つのカテゴリに大別でき、更にそれぞれのカテゴリに属する様々な評価軸が存在すると示されている。しかしなが

ら、評価者が複数人の言動からなるインタラクションをそのような細かい評価軸で評定することはとても労力が高く、客観性や公正さの信頼性が高いとは限らない。つまり、システムによる議論参加者の客観的な自動評価およびフィードバックを実現することは、評価者の負担軽減や被評価者の評価結果に対する納得感、コミュニケーション能力の向上など様々な点に貢献できると考える。したがって、議論中の話者の陳述について分析が可能なコーパスを構築することは有意義であると考えている。

そこで、本研究は議論参加者の陳述評価に向けた複数人議論を収録したマルチモーダルコーパスの構築を試みる。議論参加者の陳述評価のために、評価対象となる主張や討論をなるべく多様かつ数多く収録したい。そこで、4名の話者がディベートおよび合意形成を行う対話設定を構築し、「自身のスタンスがより良いことを相手に納得させる」「答えのない議論で全員が合意できるアイデアを提案する」という異なる性質の立論をマルチモーダルデータとして収録する。また、先行研究 [1] の分類と書き言葉における論証の強さを評価した先行研究 [2] を参考に、議論参加者の陳述の理解・納得感を示す評価指標である Effectiveness (有効性) に分類される 6 つの評価軸のアノテーションスキームを構築した。本研究ではそのスキームを用いたアノテーションを実施し、その結果の報告も行う。

2. 関連研究

近年、自然言語処理の分野で書き言葉を対象に論証の質を自動評価するためのデータセットが複数構築され [2] [3]、様々な推定モデルが提案されている [4] [5]。しかしながら、話し言葉による論証の評価は書き言葉よりも困難な点がある。例えば、エッセイは論述の内容を熟考できるため構成が綺麗に整っており、かつ他者からのインタラクションを考慮する必要がない。一方、対面形式の議論の場合、陳述は短時間であるため話し言葉や言い淀みによって構成が崩れることもしばしばあり、他の人からのリアルタイムな批判に対応することも必要となる。加えて、書き言葉は言語的情報のみを考慮する一方で、対人のコミュニケーションは人の動作や声など複数のモダリティを考慮する必要がある [6]。こういった点からも、議論における議論参加者の陳述自動評価に向けたマルチモーダルなコーパスの構築は必要である。

複数人によるインタラクションを理解したり分析することを試みる研究は古くから関心が持たれ、AMI Corpus [7] や ICSI Corpus [8] を代表とする様々なコーパスが構築されている [9] [10] [11] [12]。これらのコーパスは著者の知る限り、議論参加者の陳述について定量的なスコアリングを行うことを目的としておらず、スコアリングに関係する評価値が付与されていない。林ら [13] は就職試験を想定したグループディスカッションの参加者から取得できるマルチモーダル情報からコミュニケーションスキルを測ることを目的として、発話の書き起こしや Web カメラなどの機器から得られる動作・音声データを収録したコーパスを構築している。また、林らはディスカッションの形式を「自由討論」「インバケット」「ケーススタディ」「ディベート」の 4 つの型に分類できることを報告し、そのうち



図 1 収録環境の様子

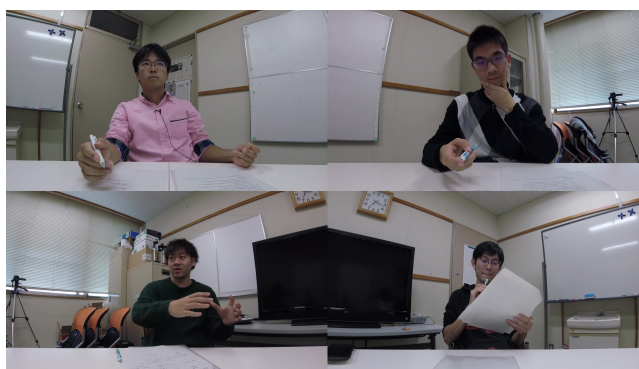


図 2 上半身撮影用カメラでの撮影例

採用試験で用いられることの多いインバケット型およびケーススタディ型のディスカッションを収録している。林らのコーパスでは「傾聴する姿勢」や「情報伝達力」などコミュニケーションスキルに関連する 5 つの項目について 5 件法で話者を評価した結果を付与している。一方、本研究は話者の陳述の理解・納得感に関して自然言語処理や議論学の知見を用い、より細かく普遍的な評価軸についてアノテーションスキームを構築している点が異なる。

3. コーパス収録設定

本節ではコーパス収録を実施する環境および収録タスクについて述べる。

3.1 収録環境

対話の収録環境を図 1 に示す。部屋に円形テーブルを設置し、計 4 名の議論参加者を均等な間隔で配置する。対話の様子を俯瞰して記録するためにテーブルの中心に 360 度カメラ (Ricoh Theta V) と頭上カメラ (GoPro HERO5 Session)、360 度カメラの予備として部屋の壁際に予備のビデオカメラを設置する。加えて、参加者の表情や視線、上半身の動作情報を記録するために各参加者の正面にも上半身撮影用カメラ (GoPro HERO5 Session) を設置する。撮影された映像の一例として、各議論参加者の正面に設置されたカメラによる映像のキャプチャを図 2 に示す。また、議論参加者の発言をノイズを減らした状態で録音するために、各議論参加者にピンマイクレコーダー (TASCAM DR-10L) を装着する。

3.2 議論収録の手続き

議論収録は以下の (1) から (3) の要素で構成される。本コーパスは (2) および (3) を収録する。以下に、各要素の具体的な手続きを示す。

表 1 議論コーパス収録命題一覧

トピック ID	命題
T0	(小中高の) 生徒は制服を着用すべきである
T1	成人の拳銃所持・携帯の権利を認めるべきである
T2	小中高の教材はタブレットに置き換えるべきである
T3	飲酒可能年齢は 20 歳から下げられるべきである
T4	未成年の暴力的ゲームのプレイを禁止すべきである

(1) 収録に関する説明およびサーベイ (20 分)

集まった議論参加者 4 人は 2 人 1 チームに分かれた後、実験遂行者から収録の趣旨や今回議論する命題、自身が演じる立脚点について 5 分程度説明を受ける。その後、15 分間実験遂行者から与えられた資料や実験条件の再確認およびオンライン上で議論用のサーベイを行い、論点を整理する。

(2) ディベート (20 分)

議論参加者は前フェーズで整理した内容を用いて 20 分間を目安にディベートを行う。本研究ではディベートに命題単一型の立脚点混合タイプを採用する。例えば「小中校の教材はタブレットに置き換えるべきである」という命題に対して「置き換えるべき」という肯定的立脚点と「置き換えるべきでない」という否定的立脚点がそれぞれ主張を展開するディベートが命題単一型の立脚点混合タイプである。

(3) 合意形成 (20 分)

ディベート終了後から更に 20 分間を目安に、話し合った内容を基にして各立脚点が納得する合意形成を行う。例えば「小中校の教材はタブレットに置き換えるべきである」という命題に対して肯定的/否定的立脚点の両方が納得のいく折衷案として「導入コストや学習者の集中力欠如を考慮して理科などタブレット導入によって学習効果の向上が期待できる科目については導入する」などを導く議論を行う。

3.3 議論内容

本研究では procon.org^(注1)に掲載されている issues を参考に 5 つの命題を作成した。命題のリストを表 1 に示す。本研究はディベートポータルで行われるような匿名ユーザによるチャット形式のディベートとは異なり、議論参加者に対面形式で行う議論を依頼するため、コーパスとしての質の担保や倫理的問題を考慮する必要がある。そこで、命題選定の際には「命題は個人の主義/信条による心的負担が少なく、かつ知識差によって一方の立脚点が有利または不利になりにくいものにする」という点を考慮した。例えば、「死刑制度は秩序の担保のために必要である」といった命題は個人の背景や主義/信条によっては議論参加者の心的負担がとても大きい可能性がある。また、「イランとの核合意は破棄されるべきである」といった命題は特定の個人が専門的知識に長けていて残りの参加者が発言できなくなる可能性があると考えられる。

4. データ収録

本節では実施した収録実験およびそこから獲得したデータについて述べる。

話者 ID 開始時間 終了時間 発話

B(Pro2)	00:09:19.475	00:09:21.272	その始めるタイミング
B(Pro2)	00:09:21.492	00:09:24.866	っていうのは必ずしも人は物を扱えるわけではないので
B(Pro2)	00:09:25.068	00:09:27.458	どのタイミングから始めても変わらないんじゃないかなと思います
C(Con1)	00:09:28.407	00:09:30.990	小学校 1 年生のときにももうそれをする
B(Pro2)	00:09:31.418	00:09:32.508	それをしてしまっても
C(Con1)	00:09:33.003	00:09:35.759	まだ字も漢字も書けないときにそれをする
B(Pro2)	00:09:33.253	00:09:33.863	えっ

図 3 発話の書き起こしテキスト例 (トピック T2)

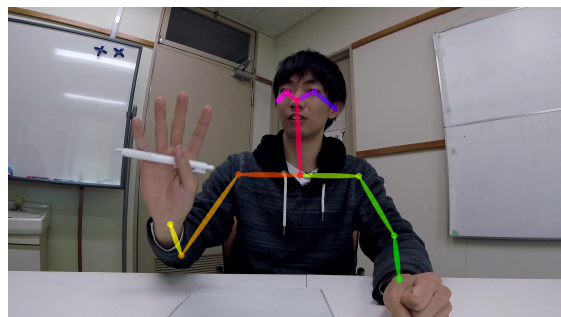


図 4 OpenPose による骨格特徴点解析例

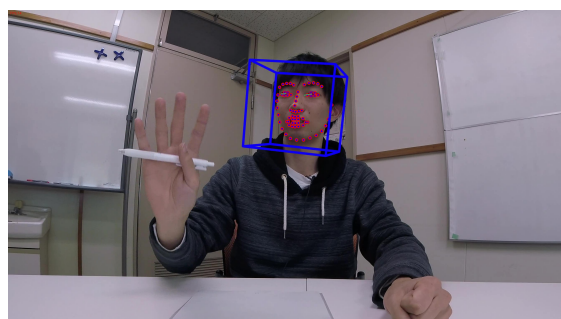


図 5 OpenFace による顔特徴点解析例

4.1 収録実験概要

九州工業大学に所属する計 13 名の大学生・大学院生に対して収録の協力を依頼した。4 名の被験者を 1 つのグループとして、合計 5 グループ 10 対話の収録を行った (一部の学生は 2 つのグループに属している)。

収録開始前に被験者は「研究室の研究用途に限定したデータの利用」「個人を特定できる情報を匿名化した対話データの Web 上への無償公開」「研究室の利用許可を得た外部機関へのデータ配布/公開」という 3 つのデータ取り扱いの同意/不同意について意思表示と署名を行った。更に、被験者は収録終了後に「決定した折衷案は何ですか？」など議論についての簡単なアンケートに回答した。

4.2 獲得データと整形

上記手続きで 5 対話のディベートと 5 対話の合意形成を収録した映像データを獲得した。ここから、コーパス構築に向けたデータ整形について述べる。

はじめに、全体を俯瞰して撮影した 360 度カメラの映像からディベートと合意形成議論の開始から終了までの映像を手動で切り出した。次に、PluralEyes4^(注2)を用いて、切り出した 360 度カメラの映像と各議論参加者の動作を撮影した上半身撮影用カメラおよび頭上カメラの映像を同期し、議論の開始から終了までの区間の全ての映像を獲得した。また、PluralEyes4 は音

(注1) : <https://www.procon.org/>

(注2) : <https://www.redgiant.com/products/shooter-pluraleyes/>

表2 評価軸一覧

評価軸	スコアレンジ	概要
説得力	1-6	陳述にどの程度説得力があったか
発想力	1-6	陳述がどの程度具体的に独創的だったか
雄弁さ	1-6	陳述がどの程度雄弁だったか
一貫性	1-6	陳述にどの程度一貫性があったか
批判的思考力	1-6	陳述がどの程度批判的であったか
説得戦略	Yes/No	Ethos/Logos/Pathosを有していたか

声データを基に同期を行うため、360度カメラの映像音声とピンマイクで収録した音声を同期することで各議論参加者の音声データを獲得した。続いて、上半身撮影用カメラの撮影データを結合した動画(図2)と音声動画アノテーションツールであるELAN[14]を用いて発話を転記単位(0.2秒以上のポーズを区切りとする単位)で書き起こした。また、上半身撮影用カメラの映像に対してOpenPose[15]およびOpenFace[16]を適用し、各議論参加者の骨格および顔の特徴点を獲得した。

以上の手続きを通して、本収録で5グループ分のディベートおよび合意形成議論について下記のデータを獲得した。

- (1) 各議論参加者を正面・頭上から撮影した映像データ
- (2) 各議論参加者のピンマイク経由で録音した音声データ
- (3) 発話の書き起こしテキスト(図3)
- (4) 各議論参加者のOpenPoseによる骨格特徴点(図4)
- (5) 各議論参加者のOpenFace顔特徴点(図5)

5. 陳述の質アノテーション

本研究の最終的な目標は議論参加者の陳述を定量的に評価することである。そこで本節では、取得した対話データの各議論参加者の陳述に対して説得力を中心とする有効性に属する6つの評価軸をアノテーションするためのスキーム構築を行い、それを用いたアノテーションの結果について述べる。

5.1 アノテーションスキーム

先行研究[2]を参考に評価軸および評価基準を作成した。評価軸の一覧を表2に示す。

説得力は被評価者の陳述に対してどの程度説得力があったかを示すスコアである。納得する人が多いと考えられる陳述であるほど、このスコアは高くなる。発想力は被評価者の陳述にどの程度具体性や独創性があったかを示すスコアである。つまり、この評価軸は被評価者の命題に対する考えや知識の深さ、自信などを表したものとなっている。雄弁さは被評価者の陳述がどの程度雄弁であったかを示すスコアである。自身の考えや意見を流暢にかつ明確に表現できていればいほどこの評価軸のスコアは高くなる。一貫性は被評価者の陳述にどの程度一貫性があったかを示すスコアである。つまり、この評価軸は被評価者の主張が何回も変わることなく一貫した流れになっているかを表している。批判的思考力は被評価者の陳述がどの程度批判的であったかを示すスコアである。人から与えられる情報や自身の意見/考えを冷静に分析し、客観的であればあるほどこの評価軸のスコアは高くなる。説得戦略は被評価者の陳述がアリストテレス[17]によって定義されるEthos(話者の信憑性)、Logos(論理性)、Pathos(感情的アピール)を感じさせるものかを示

表3 各評価軸に対するクリップンドルフの α 係数

評価軸	5分時点	10分時点	15分時点	20分時点
説得力	0.322	0.168	-0.0233	0.144
発想力	0.284	0.035	-0.133	-0.059
雄弁さ	0.352	-0.019	0.055	0.072
一貫性	0.450	-0.031	0.084	0.167
批判的思考力	0.167	0.344	0.198	0.303
説得戦略(Ethos)	-0.086	0.221	0.091	0.068
説得戦略(Logos)	0.136	0.042	0.148	0.153
説得戦略(Pathos)	-0.111	-0.053	-0.147	-0.068

す指標である。各スコアの詳細な基準は付録に掲載する。

5.2 アノテーションの実施と分析

各グループに対してそのグループの構成員でない第三者3名を割り当て、議論参加者の陳述の評価を依頼した。議論の経過に伴うスコアの変化を分析するために、アノテータは図2の上半身撮影用カメラによる映像を結合した動画を視聴しながら5.1節で定義した評価指標を基に、議論開始から5分おきにその時点までの話者の陳述を評定した。

採点結果について一致率を求め、6つの評価軸についての信頼性を確認した。本研究では一致率の算出にクリップンドルフの α 係数を採用した。この係数は-1から1の連続値で全ての尺度の一致率を算出できるもので、1に近ければ近いほど高い一致率を示す。経験的に0.667以上であればアノテータによる評価が一致しているとされている。表3に各評価軸のクリップンドルフの α 係数を示す。表中の全ての値が基準値を下回っており、本研究で実施した時間区切りで議論参加者の陳述を評価する方法では信頼性のあるアノテーション結果が得られなかった。また、20分時点での各評価軸の一致率を比較すると「発想力」「雄弁さ」「Ethos」「Pathos」の4つの評価軸の一致率が特に低い傾向にあった。これは4つの評価軸は他と比較してアノテータの感じ方などの主観に大きく関わるものが原因であると考えられる。表3より、「Logos」を除いた評価軸について一致率が最大となっているのは5分時点または10分時点となっていることが確認された。つまり、時間が経つほど評価する議論が煩雑になり、スコアの一致率が低くなる傾向があると推測される。今後は単純な時間区切りではなく、議論内容を崩さない細かい分割単位であるトピックや議論構造を用いた陳述スコアのアノテーションを実施、その結果をボトムアップに統合することで議論参加者の陳述のスコアとする方法を考えている。

本アノテーション実施の際に、アノテータにはスコアのみでなく、そのスコアに決定した理由について100字以内で記述するように求めた。そこで、本節では説得力が低いと評価された話者と高いと評価された話者の間にどのような違いがあったのかを定性的に分析した結果を報告する。

定性的分析のために議論参加者を説得力の低い話者(低群)と高い話者(高群)を仕分ける。はじめに、2名以上のアノテータが同じスコアを評定した場合にはそのスコアを、3名のアノテータがそれぞれ違うスコアを評定した場合には平均値をその議論参加者の説得力とする集計処理を行った。集計した説得力スコアを基に説得力スコアが1, 2, 3の議論参加者を低群, 4, 5,

6の議論参加者を高群として定義する。低群に分類された被評価者には主に「データや自身の経験など具体的な事例を通して話をしているが、主張したいポイントが聞き手に伝わらない」といった傾向があった。それと同時に「主張を行った後の他人からの批判や他人の主張によって主張が揺らぐ」「相手の陳述に対して言及するに留まり、自身の主張を行わない」といった基準で減点される傾向があることもわかった。一方、高群に分類された議論参加者には「言いたいことを明確に表現できている」といった一般的な傾向の他に「他の議論参加者の陳述に対して言及を行い、同時に自身の立場を明確化する」といった行動が見られることを確認した。この結果から、インタラクションにおける議論では主張だけではその被評価者の説得力は同定されないことを定性的に確認した。つまり、インタラクションにおける陳述の評価には話者自身の主張内容のみならず、その話者が他の話者に対してどのような発言や応答を行ったかも大きく影響することを示唆している。本研究で実施したアノテーション手法では全ての陳述を統合的にスコアリングしているため、主張と応答を切り分けたスコアの評定方法の考案も今後の課題である。加えて、主張と応答の2つの関係が議論参加者の陳述の質にどのような相互作用を及ぼすかについて信頼できるアノテーション結果を基に分析していく必要がある。

6. おわりに

本研究は議論参加者の陳述の定量的評価に向けたマルチモーダルコーパスの構築を行った。ディスカッションの分類のうちディベート型および自由討論型の2種類の議論について収録を実施し、発話情報に加え、議論参加者の映像・音声のデータを獲得した。また、議論参加者の陳述を有効性に分類される6つの評価軸についてアノテーションするスキームを構築した。本研究では信頼性の高いアノテーション結果は得られなかったが、定性的な分析から今後アノテーションスキームの改良や質推定モデルを考案する上で有用な知見が得られた。

今後は、アノテーションの実施方法について再考し、信頼性の高い評価値の付与に努めたい。また、現在は10対話と収録対話数が少ないため、収録対話数の拡張も実施する予定である。今回は説得力について低群と高群においてどのような差があったのか定性的な分析に留まったが、今後は議論参加者の陳述の質を推定するモデルを構築し、精度評価や人の評価観点とモデルの評価観点の差の検証を行いたい。

謝辞

本研究は科研費17H01840の助成を受けたものです。

文 献

- [1] H. Wachsmuth, N. Naderi, Y. Hou, Y. Bilu, V. Prabhakaran, T.A. Thijm, G. Hirst, and B. Stein, "Computational Argumentation Quality Assessment in Natural Language," Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL), pp.176-187, 2017.
- [2] W. Carlile, N. Gurrupadi, Z. Ke, and V. Ng, "Give Me More Feedback: Annotating Argument Persuasiveness and Related Attributes in Student Essays," Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), pp.621-631, 2018.
- [3] I. Habernal and I. Gurevych, "Which argument is more convincing? Analyzing and predicting convincings of web arguments using bidirectional LSTM," Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL), pp.1589-1599, 2016.
- [4] I. Habernal and I. Gurevych, "What makes a convincing argument? Empirical analysis and detecting attributes of convincings in Web argumentation," Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.1214-1223, 2016.
- [5] Z. Ke and V. Ng, "Automated Essay Scoring: A Survey of the State of the Art," Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI), pp.6300-6308, 2019.
- [6] 武川直樹, 中山知大, 徳永弘子, 大和淳司, 山下直美, "グループディスカッションにおける発言者の言語/非言語の表出と評価者評価の関係の分析," 電子情報通信学会論文誌 D, vol.101, no.2, pp.284-293, 2018.
- [7] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, et al., "The AMI meeting corpus: A pre-announcement," Proceedings of the 2nd International Workshop on Machine Learning for Multimodal Interaction (MLMI), pp.28-39, 2005.
- [8] A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, et al., "The ICSI Meeting Corpus," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp.364-367, 2003.
- [9] T. Yamamura, K. Shimada, and S. Kawahara, "The Kyutech corpus and topic segmentation using a combined method," Proceedings of the 12th Workshop on Asian Language Resources (ALR), pp.95-104, 2016.
- [10] N. Campbell, T. Sadanobu, M. Imura, N. Iwahashi, S. Noriko, and D. Douchamps, "Multimedia Database of Meetings and Informal Interactions for Tracking Participant Involvement and Discourse Flow," Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC), pp.391-394, 2006.
- [11] O. Aran, H. Hung, and D. Gatica-Perez, "A Multimodal Corpus for Studying Dominance in Small Group Conversations," Proceedings of the 7th International Conference on Language Resources and Evaluation Workshop on Multimodal Corpora (LREC), pp.2223-2232, 2010.
- [12] M. Koutsombogera and C. Vogel, "Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus," Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC), pp.2945-2951, 2018.
- [13] 林佑樹, 二瓶美巳雄, 中野有紀子, 黄宏軒, 岡田将吾, "グループディスカッションコーパスの構築および性格特性との関連性の分析," 情報処理学会論文誌, vol.56, no.4, pp.1217-1227, 2015.
- [14] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, "ELAN: a professional framework for multimodality research," Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC), pp.1556-1559, 2006.
- [15] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields," 2018.
- [16] T. Baltrusaitis, A. Zadeh, Y.-C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," Proceedings of 13th IEEE International Conference on Automatic Face Gesture Recognition (FG), pp.59-66, 2018.
- [17] Aristotle, On Rhetoric: A Theory of Civic Discourse, Oxford University Press, 2007.

付 録

1. 各評価軸採点基準詳細

表 A-1 説得力採点基準

スコア	概要
6	とても力強く明確な陳述。彼/彼女の陳述を聞いた人のほとんどが納得すると考えられる
5	力強く明確な陳述。彼/彼女の陳述を聞いた人の過半数が納得すると考えられる
4	理解できる陳述。彼/彼女の陳述を聞いた人の半数程度が納得すると考えられる
3	概ね理解できる陳述。彼/彼女の陳述を聞いた人が元々同じ考えなら納得すると考えられる
2	理解し難い陳述。彼/彼女の陳述を聞いた人の過半数が納得しないと考えられる
1	とても理解し難い陳述。彼/彼女の陳述を聞いた人のほとんどが納得しないと考えられる

表 A-2 発想力採点基準

スコア	概要
6	とても独創的かつ具体的である。陳述をより向上するための内容の改善がほとんどない
5	独創的かつ具体的である。内容の不明瞭さが僅かにあるが独創的な発想で具体性が十分ある
4	概ね独創的かつ具体的である。独創性が秀でて高くはないが、陳述の内容は練られていて
3	あまり独創的かつ具体的でない。独創性が高くなく、陳述内容に対して不明瞭な点が残る
2	独創的かつ具体的でない。陳述内容の独創性が低く、曖昧で不明瞭である
1	とても独創的かつ具体的でない。陳述内容が抽象的で曖昧性が高く、独創性もない

表 A-3 雄弁さ採点基準

スコア	概要
6	明確かつ言葉巧みに主張を行っており、不明瞭な陳述がない。 陳述の理解を低下させるような言い直し、言い淀みがなく、強引な畳み掛けや詭弁のような言論がない
5	明確に主張を行っているが、稀に不明瞭な陳述がある場合もある。 陳述の理解を低下させるような言い直し、言い淀みがなく、強引な畳み掛けや詭弁のような言論がない
4	上手く主張を行っているが、稀に不明瞭な陳述がある場合もある。 陳述の理解を低下させるような言い直し、言い淀み、または弱い畳み掛けや詭弁のような言論がたまにある
3	概ね上手く主張を行っているが、不明瞭な陳述がしばしばある。 陳述の理解を低下させるような言い直し、言い淀み、または弱い畳み掛けや詭弁のような言論がたまにある
2	あまり上手く主張を行っておらず、不明瞭な陳述がしばしばある。 陳述の理解を低下させるような言い直し、言い淀み、または強引な畳み掛けや詭弁のような言論がある
1	うまく主張を行っておらず、陳述が多くの場合不明瞭である。 陳述の理解を低下させるような言い直し、言い淀み、または強引な畳み掛けや詭弁のような言論がある

表 A-4 一貫性採点基準

スコア	概要
6	とても一貫性がある。陳述の構成がとても明確で、2 転 3 転することなく首尾貫徹している
5	一貫性がある。陳述の構成が明確で、陳述が 2 転 3 転することがほぼ無く一貫している
4	概ね一貫性がある。陳述の構成が概ね明確で、陳述があまり 2 転 3 転することがない
3	あまり一貫性がない。陳述の構成が稀に崩れており、陳述が 2 転 3 転することがある
2	一貫性がない。陳述の構成がしばしば崩れており、陳述が 2 転 3 転することがよくある
1	とても一貫性がない。陳述の構成が多くの場合崩れており、陳述が一貫していない

表 A-5 批判的思考力採点基準

スコア	概要
6	とても批判的思考力がある。主張や質疑応答等全ての文脈において客観性と合理性の高い陳述を行えている
5	批判的思考力がある。主張や質疑応答等全ての文脈においてある程度客観性と合理性の高い陳述を行えている
4	概ね批判的思考力がある。多くの文脈において客観性と合理性の高い陳述を行えている
3	あまり批判的思考力がない。一部の文脈においてのみ客観性と合理性の高い陳述を行えている
2	批判的思考力がない。ほとんどの文脈において客観性と合理性の高い陳述を行えていない
1	とても批判的思考力がない。全ての文脈において客観性と合理性の高い陳述を行えていない