

# 複数人遠隔対話コーパスの構築と LLM を用いた取りまとめ役の特徴分析

波多野 翔貴<sup>1</sup> 嶋田 和孝<sup>1</sup>

<sup>1</sup>九州工業大学大学院

hatano.shoki801@mail.kyutech.jp shimada@ai.kyutech.ac.jp

## 概要

本研究では、複数人遠隔対話コーパスである Kyutech Remote コーパスの構築に取り組む。複数人による遠隔対話をターゲットとしたコーパスは少ない。そのため、遠隔対話を対象としたコーパスを構築することで、言語資源の拡充を目指す。また、対面対話と遠隔対話の取りまとめ役の特徴の分析を行う。分析には大規模言語モデル (LLM) を用いる。分析の結果、対話形式ごとに取りまとめ役にはそれぞれ異なった特徴があることが示された。

## 1 はじめに

議論は、複数の参加者によって意見交換や意思決定を目的として行われる。議論において、限られた時間の中で優れた結論を導き出すことは重要である。しかし、実際の議論では、参加者が有用な意見を出せず議論が滞ることがある。また、参加者間の発言量や立場の差によって、結論が特定の参加者の意見に偏る場合がある。このような問題が起こると、参加者の間で満足のいく合意形成が行えない。

これらの問題を防ぐためには、取りまとめ役の存在が重要である。取りまとめ役が適切なファシリテーションを行うためには、各参加者の発言量の把握や、適切なタイミングでの発言の促進などを並行して行う能力が求められる。しかし、必ずしも議論参加者の誰かがそのような能力を有しているわけではない。そのため、ファシリテーションという視点から議論を支援するシステムの研究が進められてきた [1, 2]。これらのシステムは主に対面での議論を対象としており、その構築の際には、実際の対面議論を対象とした言語資源 (コーパス) が用いられている。一方で、議論は対面形式だけでなく、Web 会議のように遠隔形式でも行われることがある。遠隔形式での会議では、対面環境下での議論と比較し

て、参加者の立ち振る舞いに変化することが報告されている [3, 4, 5]。よって、遠隔形式での議論を支援するシステムを構築するためには、遠隔形式で実施された複数人議論を対象としたコーパスの存在が望ましいが、その数は少ない。

また、対話形式の違いにより生じる、取りまとめ役の特徴の差に着目した研究は十分に行われていない。取りまとめ役は、議論を円滑に進行させて結論をより良いものへと導く役割を担う。そのため、対話形式による取りまとめ役の立ち振る舞いの違いが議論に与える影響は大きいと考えられる。加えて、対話形式ごとの取りまとめ役の特徴を分析することは、形式に応じた取りまとめ役を支援するシステムの構築に寄与すると考えられる。そのため、対話形式ごとの取りまとめ役の特徴分析は重要である。取りまとめ役の特徴については、分類モデルを用いた手法による分析が行われている [6, 7]。しかし、これらの手法では、人手によって事前に設計された特徴量に関する情報しか得ることができない。そこで、本研究では大規模言語モデル (LLM) を用いて、取りまとめ役の特徴の分析を行う。取りまとめ役の特徴の分析に LLM を用いることで、事前の設計が困難な観点からの分析を多数のデータに対して行えるようになることを期待する。

本研究の貢献は以下の2点である。

- 複数人遠隔対話コーパス (Kyutech Remote コーパス) の構築
- 遠隔対話と対面対話に対する LLM を用いた取りまとめ役の特徴分析と比較

## 2 関連研究

対話を対象としたコーパスの構築は古くから行われている。代表的なものとして、英語話者による対面対話を対象とした AMI コーパスや ICSI コーパスなどがある [8, 9]。一方で、近年では遠隔形式で

の対話を対象としたコーパスの構築も行われている [10, 11]. これらのコーパスは2名による1対1での遠隔対話を対象としたものであるが, 本研究では複数人話者による遠隔対話を対象とする. 対面形式での複数人対話を対象としたコーパスの一つに, Yamamura ら [12] が構築した Kyutech コーパスがある. このコーパスは4名による意思決定タスクを対象としたものであり, 対面形式で実施された数件の議論が収録されている. Kyutech コーパスは対面での対話を対象としているが, 使用されているタスクは対話の形式に依らないものである. そこで本研究では, この Kyutech コーパスの構築手法を参考に, 複数人遠隔対話を対象としたコーパスを構築する.

議論参加者の特徴を分析するにあたり, 動作や韻律といった非言語的特徴を用いる研究も行われている [13, 14]. しかし, 遠隔対話の特性上, カメラやマイクなどの環境が話者によって異なる. そのため, 動作や韻律などの非言語的特徴は, 話者特有の環境設定の影響を受けると考えられる. よって本研究では, 話者特有の環境に依存しない発話のみを用いて取りまとめ役の特徴分析を行う.

### 3 Kyutech Remote コーパス

#### 3.1 収録環境

本研究では, Zoom<sup>1)</sup>を用いて遠隔形式での議論を実施する. 対話データの収録には, Zoom のレコーディング機能を使用する. 対話参加者のうち1名は, 著者らが用意した対話収録用の機材を接続した PC を使用する. その他の参加者は, 個人で用意した端末を使用する.

収録機材の配置を図 1 に示す. 対話収録用の機材として, 参加者を映すための Web カメラ (UCAM-CX80FBBK<sup>2)</sup>) を使用する. また, 参加者が議論中に画面上のどの部分を見ているかを記録するために, アイトラッカー (Tobii Eye Tracker 4C<sup>3)</sup>) を用いる. 加えて, スピーカーマイク (Anker powerconf a3301<sup>4)</sup>) を接続する.

話者の ID として, 対話収録用の PC を使用する 1 名に A, その他の 3 名に B, C, D を対話ごとに割り振る. 収録した対話の様子を図 2 に示す. 図 2 で



図 1 収録機材の配置



図 2 対話収録の様子

は, A が右上, B が右下, C が左上, D が左下に表示されている.

#### 3.2 対話のタスク設定と収録の流れ

本研究では, Kyutech コーパスのタスク設定と対話収録の流れを参考に, 遠隔議論の収録を行う. 各回につき 4 名の参加者には, ある架空の都市のショッピングモールの経営者であるという設定が与えられる. そして, ショッピングモール内にあるレストラン街のテナント店が 1 店閉店した後に来店させる店舗を 3 つの候補の中から 1 つ選ぶというテーマで議論を行う. 参加者には, 出店候補の店の情報, 閉店する店の情報と閉店の理由, 既存店の情報, ショッピングモールの立地や来客者の時間帯・性別ごとの分布, 都市の人口や隣接する市町村の情報などが記載された 10 ページほどの資料を配布し, 議論時に手元で確認できるようにしている. 対話の収録を行うにあたり, 出店候補の店の情報などが異なる 10 種類のシナリオを用意している<sup>5)</sup>.

対話の流れについて説明する. まず, 参加者は, 配布された資料を 10 分間黙読する. その後, 20 分を目安に議論を行う. 議論開始から 20 分経過後に, 著者が議論終了をチャイムによって合図する. ただし, チャイムが鳴る前の議論終了と, チャイムが

5) Kyutech コーパスで使用された 4 種類のシナリオに, 著者が作成した 6 種類のシナリオを追加した.

1) <https://www.zoom.com/ja>

2) <https://www.elecom.co.jp/products/UCAM-CX80FBBK.html>

3) <https://gaming.tobii.com/>

4) <https://www.ankerjapan.com/products/a3301>

表 1 書き起こしの際に利用されるタグ

タグ	タグの説明
(F)	フィラー
(D)	言い直し・言い淀み
(Q)	疑問文
(?)	聞き取りに自信がない箇所
(L)	ささやき声や独り言のような小さな声
<笑>	発話中、もしくは単独の笑い

鳴った後の議論継続も認めている。本研究では、日頃から交流のある 18 人の学生<sup>6)</sup>から各回につき 4 名を選び<sup>7)</sup>、計 20 対話分の対話データを収録した。

対話の収録後には、「議論をコントロールしていたのはどの参加者か」など、いくつかの事項についてアンケートを実施する。

### 3.3 対話データの書き起こし方法

収録した対話の音声から、アノテーションツールである ELAN<sup>8)</sup>を用いて書き起こしを作成する。書き起こしを作成する際のルールとして、日本語話し言葉コーパス (CSJ)[15] で用いられたものを参考にす。具体的には、音声の 0.2 秒以上の無声空間を区切りとし、書き起こしデータの 1 行とする。また、表 1 のように、各発話に対してフィラーや言い淀みなどのタグ付けを必要に応じて行う。

CSJ コーパスのルールを用いて書き起こしを行う際、1 つの発話が複数に分割されることがある。そのため、Kyutech コーパスと同様に、各行の末尾に対して発話単位のアノテーションを行う。発話単位へのアノテーションの詳細は付録 A を参照のこと。+ はその書き起こし単位が同じ話者による次の書き起こし単位の先頭と繋がることを示し、/ はその書き起こし単位がある発話の終わりであることを示す。

以上の作業を行い、話者情報および発話開始・終了時間、タグが付与された 20 対話分 (11,154 発話) の書き起こしデータを得た。実際の対話データの例を図 3 に示す。参考までに、Kyutech コーパスとの統計量の対比結果を付録 B に載せる。

## 4 取りまとめ役の特徴分析

本研究の目標の一つは、異なる対話形式ごとの取りまとめ役の特徴を明らかにすることである。そこ

6) 九州工業大学に所属する学生および大学院生。  
7) 過去の組み合わせと参加者が被ることは許容するが、ある参加者が同じシナリオで議論を行うことがないようにした。  
8) <https://archive.mpi.nl/tla/elan>

ID	Start	End	Utterance
A	00:08.410	00:08.540	はい+
A	00:08.780	00:11.443	それでは始めていきたいと思いますよろしくお願いします/
B	00:12.433	00:13.052	お願いします/
C	00:12.794	00:13.423	お願いします/
D	00:13.371	00:14.649	お願いします/
A	00:17.794	00:19.220	どうい感じに進めましょうか(Q)/
C	00:20.000	00:21.860	(F まあ) Aさんから言っていけばいいんじゃないですか/

図 3 対話データの一例

で本章では、2 つの異なる対話形式に対する、LLM を用いた取りまとめ役の特徴分析について述べる。

### 4.1 使用するデータセット

本研究では、対面形式と遠隔形式の議論を対象としたコーパスを用いて取りまとめ役の特徴分析を行う。対面議論データとして Kyutech コーパスの 9 件、遠隔議論データとして Kyutech Remote コーパスの 20 件をそれぞれ用いる。

本研究における取りまとめ役は、各対話参加者へのアンケートの設問「議論をコントロールしていたのはどの参加者か」で最も多く票を得た話者とする<sup>9)</sup>。なお、最多得票者が複数人存在する場合は、その全員を取りまとめ役とする。これにより、Kyutech コーパスでの取りまとめ役は 36 人中 10 人、Kyutech Remote コーパスでの取りまとめ役は 80 人中 22 人となった。

### 4.2 LLM による取りまとめ役の推定

LLM を用いて特徴分析を行うにあたり、実際の取りまとめ役に表れる特徴を LLM が捉えることができるかが重要である。そのため本研究では、まず LLM にある対話のデータを与え、取りまとめ役が対話中のどの話者であったかを推定させる。そして、LLM が取りまとめ役を正しく分類できた場合に出力された推定理由文を分析の対象とする。

全体の流れを図 4 に示す。まず、コーパスに含まれる対話データの全発話、および各発話に付与された話者 ID とタイムスタンプを LLM に 1 対話分ずつ入力として与える。次に、それぞれの対話に対して、LLM が取りまとめ役であると推定した話者、および各話者に対する推定理由を出力させる。そして、LLM が取りまとめ役を正しく推定したときの推定理由文を集計し、取りまとめ役の特徴について第 1 著者が分析を行う。

特徴分析を行うための LLM として、GPT-4o<sup>10)</sup>を

9) ある参加者が 2 人以上に投票していた場合は、それぞれ該当する参加者に 1 票分として扱う。

10) <https://platform.openai.com/docs/models/gpt-4o>

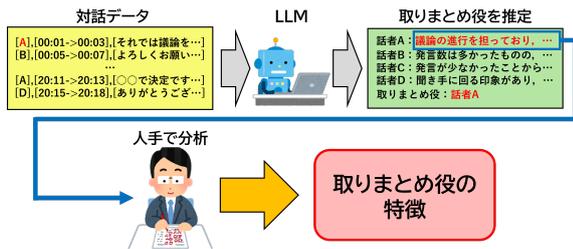


図 4 特徴分析の流れ

表 2 判断基準と対話形式ごとの出現割合

判断基準	対面形式	遠隔形式
議論の進行	55.6 (5/9)	83.3 (15/18)
議題の明確化・整理	100.0 (9/9)	88.9 (16/18)
議論の深化	44.4 (4/9)	27.8 (5/18)
他者意見の引き出し	88.9 (8/9)	61.1 (11/18)
質問の投げかけ	22.2 (2/9)	16.7 (3/18)
発言の積極性	22.2 (2/9)	0.0 (0/18)
議論の円滑化	22.2 (2/9)	22.2 (4/18)
合意形成への貢献	33.3 (3/9)	61.1 (11/18)

用いる。Temperature は 0.0 とし、各対話に対する推定は 1 回のみ行う。LLM に与えたプロンプトの詳細は付録 C を参照のこと。分類の結果、実際に取りまとめ役であった話者を取りまとめ役と推定した正解率は、Kyutech コーパスが 0.90 (9/10)、Kyutech Remote コーパスが 0.82 (18/22) となった。

### 4.3 特徴分析

対話形式ごとに、LLM が出力した推定理由文から取りまとめ役の特徴について集計し、考察を行う。まず、第 1 著者により各推定理由文を観察し、文中に含まれる判断基準のカテゴリ分けを行った。カテゴリ分けの詳細は付録 D を参照のこと。次に、LLM が正しく取りまとめ役を推定できた対話に対して、各対話形式ごとに推定理由文に含まれる判断基準の出現回数を集計し、推定理由文の総数<sup>11)</sup>に対する割合を算出した。判断基準のカテゴリ分けの結果と、対話形式ごとの判断基準の出現割合を表 2 に示す。

まず、双方の対話形式に共通していた特徴として、「議論の明確化・整理」といった文の出現回数が多い傾向にあった。よって、対話の形式に関わらず、取りまとめ役は議題の定義や方向性の設定、意見の整理をよく行う傾向にあると推測される。次に、遠隔形式での取りまとめ役では、「議論の進行」と「合意形成への貢献」といった文が対面形式より

11) Kyutech コーパスが 9 件、Kyutech Remote コーパスが 18 件

も多く出現する傾向にあった。このことから、遠隔形式での対話では、対面形式のものに比べて取りまとめ役が対話全体の主導権を握りやすい傾向にあると推測される。一方で、「議論の深化」と「他者意見の引き出し」といった文は対面形式よりも出現する割合が少なかった。したがって、対話における取りまとめ役による議題に対する深掘りは、対面形式に比べて遠隔形式では行われにくいと推測される。

### 4.4 推定理由の妥当性の確認

本研究では LLM を使って推定理由文を生成させたが、生成された理由文が対話の内容を正しく反映している保障はない。そこで、推定理由文の妥当性を検証する。具体的には、各対話について 2 名の被験者<sup>12)</sup>に、推定理由文と対話文全体を渡し、それぞれの推定理由文が A~D のどの話者に関するものなのかを判定させた<sup>13)</sup>。その結果、正しく話者を判定できたもの(正解率)は 0.634 であった。

一方で、正解率の計算の対象を取りまとめ役の話者に限ると 0.906 という高い正解率で判定できていた。この結果から、LLM の生成した推定理由文は、取りまとめ役に限れば、対話内容に則した説明になっていることがわかる。これは前節での特徴分析の妥当性を保証するものである。

## 5 おわりに

本研究では、複数人遠隔対話コーパスである Kyutech Remote コーパスを構築した。また、構築したコーパスと Kyutech コーパスを用いて、LLM を活用した取りまとめ役の特徴分析を行った。今回は取りまとめ役の特徴分析を第 1 著者のみによって行ったが、今後は複数名による判断基準への合議などを行い、より分析の妥当性を向上させる。

なお、本研究で構築した Kyutech Remote コーパスは、近日中に公開予定である<sup>14)</sup>。構築したコーパスは、発話情報からの話者の視線推定タスク [16] などに応用可能であると考えられる。また、本研究の知見を活かし、実際のファシリテーションシステム [1, 17] の機能拡張を行うことも今後の課題となる。

12) 被験者の数は全員で 6 名であり、その中から 2 名を選んで割り当てた。なお、被験者は対話収録の参加者であるが、判定をする対話には関わっていない。

13) 話者を特定できる情報が推定理由文に含まれている場合、別の表現に置き換えている。

14) <http://www.pluto.ai.kyutech.ac.jp/~shimada/resources.html#kyutechRemote>

## 謝辞

本研究は科研費 23K11368 の一部です。

## 参考文献

- [1] 西山空良, 嶋田和孝. 議論の分析とファシリテーションのための可視化ツールの構築. 電子情報通信学会技術研究報告; 信学技報, 2021.
- [2] 上林, 駿希, 井上ほか. 逸脱検出に基づき介入を行う対面議論ファシリテーションシステム. 第 86 回全国大会講演論文集, Vol. 2024, No. 1, pp. 389–390, 2024.
- [3] Toshiaki Onishi, Asahi Ogushi, Yohei Tahara, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata. A comparison of praising skills in face-to-face and remote dialogues. In **Proceedings of the Thirteenth Language Resources and Evaluation Conference**, pp. 5805–5812, 2022.
- [4] 田中彰吾, 森直久. 間身体性から見た対面とオンラインの会話の質的差異. ころの科学とエビステモロジー, Vol. 4, No. 1, pp. 2–17, 2022.
- [5] 彦山華, 横森大輔. オンラインビデオ会話における話者性の交渉に視線配布が果たす役割. 日本音響学会音声研究会資料, Vol. 4, No. 1, pp. 7–12, 2024.
- [6] Tsukasa Shiota, Takashi Yamamura, and Kazutaka Shimada. Analysis of facilitators' behaviors in multi-party conversations for constructing a digital facilitator system. In **International Conference on Collaboration Technologies**, pp. 145–158. Springer, 2018.
- [7] 波多野翔貴, 嶋田和孝. 複数人遠隔対話コーパスの構築と取りまとめ役の特徴の分析. 電子情報通信学会九州支部 第 33 回 学生会講演会, D-59, 2025.
- [8] Wessel Kraaij, Thomas Hain, Mike Lincoln, and Wilfried Post. The ami meeting corpus. In **Proc. International Conference on Methods and Techniques in Behavioral Research**, pp. 1–4, 2005.
- [9] Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, et al. The icsi meeting corpus. In **2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03)**, Vol. 1, pp. I–I. IEEE, 2003.
- [10] Michimasa Inaba, Yuya Chiba, Ryuichiro Higashinaka, Kazunori Komatani, Yusuke Miyao, and Takayuki Nagai. Collection and analysis of travel agency task dialogues with age-diverse speakers. In **Proceedings of the Thirteenth Language Resources and Evaluation Conference**, pp. 5759–5767, 2022.
- [11] Taro Okahisa, Ribeka Tanaka, Takashi Kodama, Yin Jou Huang, and Sadao Kurohashi. Constructing a culinary interview dialogue corpus with video conferencing tool. In **Proceedings of the Thirteenth Language Resources and Evaluation Conference**, pp. 3131–3139, 2022.
- [12] Takashi Yamamura, Kazutaka Shimada, and Shintaro Kawahara. The kyutech corpus and topic segmentation using a combined method. In **Proceedings of the 12th Workshop on Asian Language Resources (ALR12)**, pp. 95–104, 2016.
- [13] Hung-Hsuan Huang and Toyoaki Nishida. Investigation on the fusion of multi-modal and multi-person features in rnn for detecting the functional roles of group discussion participants. In **International Conference on Human-Computer Interaction**, pp. 489–503. Springer, 2020.
- [14] Tsukasa Shiota, Kouki Honda, Kazutaka Shimada, and Takeshi Saitoh. Development and application of leader identification model using multimodal information in multi-party conversations. **International Journal of Asian Language Processing**, Vol. 30, No. 04, p. 2050019, 2020.
- [15] 小磯花絵, 土屋菜穂子, 間淵洋子, 斉藤美紀, 籠宮隆之, 菊池英明, 前川喜久雄ほか. 「日本語話し言葉コーパス」における書き起こしの方法とその基準について. 日本語科学, Vol. 9, pp. 43–58, 2001.
- [16] Kensho Wakita and Kazutaka Shimada. An utterance is enough to the gaze? gaze detection from utterance information in multiparty discussion. In **2024 International Conference on Activity and Behavior Computing (ABC)**, pp. 1–8. IEEE, 2024.
- [17] Ryunosuke Kirikihira and Kazutaka Shimada. Discussion map with an assistant function for decision-making: A tool for supporting consensus-building. In **International Conference on Collaboration Technologies**, pp. 3–18. Springer, 2018.

## A 発話単位へのアノテーションの基準

まず、全ての書き起こし単位の末尾を+とする。次に、次の書き起こし単位まで1秒以上の空白があり、以下の条件のいずれかを満たしたとき、+を/に変更する。

- 「動詞」または「助動詞」+「接続詞」で終わる
- 次の同一話者による発話が「接続詞」「フィラー」「副詞」のいずれかで始まる

なお、上記で発話の終わりとは判定されないが、1つの発話の終わりであると判断された部分は+を\*に変更する。

## B コーパスの統計量の比較

本研究で構築した Kyutech Remote コーパスと Kyutech コーパス [12] の統計量の比較を行う。比較する項目は以下の通りである。

- 総対話時間に対する総発話数の割合（発話密度）  
遠隔形式での対話では、対面形式に比べて発話の密度が少なくなると考えられる。そこで、各対話の対話時間に対する発話数の密度の平均を求め、比較する。
- 非発話区間の割合  
遠隔形式での対話では、対面形式に比べて沈黙が発生しやすく、発話が行われていない時間の割合が大きくなると考えられる。そこで、各対話の対話時間に対する発話が行われていない間の長さの平均を求め、比較する。
- 話者交代の割合  
遠隔形式での対話では、同じ話者が続けて話す傾向にあると考えられる。そこで、各対話の総発話数に対する話者が交代した割合の平均を求め、比較する。

2つのコーパスに対する平均値を表3に示す。

比較項目	対面形式	遠隔形式
発話密度	0.38	0.40
非発話区間の割合	0.45	0.44
話者交代の割合	0.54	0.30

表3より、発話密度と非発話区間の割合には大きな差が見られなかったが、話者交代の割合は対面形

式の方が高いという結果になった。このことから、遠隔形式での対話では、一人の話者が継続して発話を続ける傾向にあることが示された。一方で、発話の密度や非発話区間の割合については、対話形式による影響が小さいことが示された。

## C プロンプト

LLMに与えるプロンプトを図5に示す。

<p><b>#タスクの説明文</b> あなたは会話分析と議論構造に精通した専門家です。あるショッピングモールのレストラン街に新規出店させる店舗を決めるという目的で行われた議論のデータが与えられます。議論の内容から、どの話者がファシリテーターとして機能していたかを分析してください。そして、最もファシリテーターとして機能していた話者を、具体的な理由とともに出力してください。ただし、ファシリテーターと見なされる人物が複数いると判断される場合は、複数名を挙げても構いません。</p>
<p><b>#出力形式</b> 次のフォーマットの通り出力してください。 推定結果: A:&lt;話者Aに対する分析&gt; B:&lt;話者Bに対する分析&gt; C:&lt;話者Cに対する分析&gt; D:&lt;話者Dに対する分析&gt; ファシリテーター:&lt;話者ID(1~4名)&gt; 理由:&lt;推定の理由&gt;</p>
<p><b>#対話内容</b> 以下は4名の参加者による議論の発言内容です。 (発話開始時間)-(発話終了時間) [(話者ID)]: (発話内容)</p>

図5 プロンプト

## D 判断基準の定義

判断基準の定義について、表4に示す。

表4 判断基準の定義

判断基準	定義
議論の進行	「進行役を担う」や「進行方法を提案する」という記述がある場合。
議題の明確化・整理	「議題の定義」や「焦点」、「整理」などの記述がある場合。
議論の深化	「議論を深める」や「活性化させる」という記述がある場合。
他者意見の引き出し	「他話者に対して意見を引き出している」という記述がある場合。
質問の投げかけ	「他話者に対して質問を行う」という記述がある場合。
発言の積極性	「積極的な意見」や「発言量の多さ」に関する記述がある場合。
議論の円滑化	「他者からの意見の受け止め」や「相槌を打つ」という記述がある場合。
合意形成への貢献	「最終的な意見をまとめる」や「収束させる」という記述がある場合。