

話し合いを評価するための評価パラメータの検討 (1)

水上悦雄 (ATR) 森本郁代 (関西学院大) 大塚裕子 (IBS) 鈴木佳奈 (ATR)

竹内和広 (大阪電通大) 奥村学 (東工大) 柏岡秀紀 (ATR)

1. はじめに

近年, 公的な話し合いの場—裁判員制度における評議の場や, 行政計画における市民参画会議, コンセンサス会議[1]などの科学技術コミュニケーションの場—が増加しつつある。これらの話し合いの現場では, 知識, 立場, 経験などが異なる人々が, 合意することが容易でないような, 広い視野, 角度から, 熟議することが必要とされる課題を, 決められた時間内で解決することを求められる。そのような場合, 得てして, 対立を避けて, 一部の強い意見に安易に合意したり, 感情的になって議論が脱線し, 最後まで合意に至らなかったりする 경우가多々ある。このような話し合いを支援するためにも, また, 話し合いの参加者となる人々が話し合いの力を身につけるよう学習するためにも, 話し合いが適切かつ十全に行われたかどうか, あるいは参加者が適切に振舞っていたかどうかを評価する必要がある。

話し合いを評価する際には, (1) 何を評価するのか (評価対象), (2) 誰が評価するのか (評価者), (3) どのように評価するのか (評価方法) の諸側面を考慮する必要がある[2]。従来の話し合いの評価は, アウトプットとしての意見数や合意結果を議事録から評価する手法や, 話し合いの満足度を問う, 参加者対象の事後アンケートが主流であった。しかし, 参加者アンケートに関しては, 評価基準が主観的にならざるを得ず, 話し合いへの参加動機や貢献度合い, 参加時の立場の違いなどが, 評価に影響を与え得るなどの問題がある。また, アウトプットのみでの評価は, それが産出されたプロセスが評価に適切に反映されないことが問題となる。話し合いのプロセスとは, 話し合い全体 (話し合いのアジェンダや時間管理などの側面も含む) を俯瞰した巨視的プロセスと, 参加者間の言語的・非言語的行為のやり取りにまで配慮した, 微視的プロセスの双方を含む²。筆者らは, これまでに, 話し合いのプロセスを適切に評価に組み入れるため, 参加者のやり取りレベルの

コミュニケーションの側面, 議論の進め方やまとめ方に関わる話し合いのメタ的要素の双方を評価できるような評価指標を策定してきた [3][4][5][6]。ただし, 後述するように, この評価指標は, 議論参加者や観察者が評価・共有することを前提とし, 各評価項目と具体的な振る舞いと関連性を評価者の判断に委ねるような指標となっている。これに対して, それらの評価指標を客観的に比較可能な形で利用するためには, 具体的にどのような行為や議論の諸要素を計測あるいは判定すればよいのかを探る必要がある。

そこで本研究では, 筆者らが策定した, 話し合いのための7つの評価指標の概要を述べ, その作成過程で行った分析を元に, 各評価指標に対応する, 可能な限り客観的な(測定・数値化可能な)評価の対象=評価パラメータを検討する。

2. 「自律型対話プログラム³」における7つの評価指標

筆者らは, 話し合いの進行役を専門的に務める司会者などの支援がなくとも, 他参加者と共に主体的に対話を進めていける能力を「自律的対話能力」と定義し, 大学生が, ディスカッションの実践と評価, フィードバックを通じて, この能力を備えていくために必要となる評価指標と評価の方法論, およびそれらを活用した授業カリキュラム「自律型対話プログラム」の開発を進めている [3]。評価指標の作成にあたり, 筆者らは, 実際の大学生の話し合いデータを分析することで, 話し合いプロセスにおける, コミュニケーション上の諸要素, および話し合いの進め方に関わる諸要素を評価するような, 表1に示すような7つの評価指標を作成した⁴。

この評価指標は, 参加者と観察者の双方が評価し, その評価結果を共有することで, コミュニケーションのあり方, 議論の進め方について, 現状の問題点を課題化, 意識化し, 改善していくという方法論と

¹ 本稿では, 何らかの目的を遂行するために (主に3人以上の多人数で) 行う会話を「話し合い」と表現している。

² この二つのプロセスは完全には分離できるものではなく, 相互に関わるものであり, かつその中間レベルのプロセスも考え得るだろう。

³ この研究は, 独立行政法人科学技術振興機構, 社会技術研究開発センターの公募プログラム「21世紀の科学技術リテラシー」のH18年度採択課題「自律型対話プログラムによる科学技術リテラシーの育成」として研究助成を受けています。

⁴ 詳しい作成手法については[5][6]などを参照されたい。

一体となっている。そのときの、評価者内、評価者間の評価基準のゆれは、互いの着眼点、認識の差を共有し、相互理解を促進できるという意味で、むしろ肯定的に位置づけている。

表 1 7つの評価指標

評価指標	評価の観点
誠実な参加態度	自分の意見をしっかりと伝え、人の発言をしっかりと聞けたか
対等な関係性	全員が対等に議論に参加したか
議論の活発さ	議論は活発だったか
意見の多様さ	いろいろな意見が出ていたか
議論の深まり	一つ一つの意見が十分に検討されていたか
議論の管理	議論の流れがしっかりコントロールされていたか
意見の積み上げ	結論に向かって一つ一つの意見が積み上げられていたか

3. 評価パラメータの検討

ただし、筆者らが表1の指標を作る過程([5][6]を参照のこと)では、話し合い場面の印象評価データを因子分析して得られた因子得点の分布傾向が類似した場面に共通するような、各因子を構成する項目に関連する具体的な振る舞いや事柄を抽出し、それらを包括する記述として、表1に示した指標名と観点に落とし込んでいく。例えば、「議論の活発さ」が低いと評価された場面は、誰かの発言中、(あいづちも打たず)じっと固まって聞いている人が多かったり、発言の受け止め役がいなかったり、いても誰かに固定されていたり、発言が途切れたときに、長い沈黙が流れたり、発言者が固定されていて、少人数で議論が行われていたり、発言のまとめ役を誰か一人に任せきりにしていたりするなど、ある程度発話データから同定、定量化可能であるようなポイントが、場面共通の特徴として現れていた⁵。これらの、筆者らが評価指標を整理する際に抽出した特徴は、それらが頻繁に現れること、あるいは目立っていること、によって当該話し合いの参加者間にも「相互認知環境[7]」として共有され、話し合いやその評価に影響を与え、さらに観察者においても気になる振る舞いとして明らかとなる確率が高くなるだろう。

そこで、印象評価と振る舞いの関係性を分析した先行研究[4][8]や雑談の盛り上がりや発話の関係性を分析した先行研究[9]などを参考に、筆者らの分析

⁵ これらのポイントは、評価後の共有時に、参考として参加者らに提示することで、より問題点を明確にするために役立ててもらっているが、「正解」なわけではない点を強調している。

によって得られた各評価指標の下位項目([6]での記述を一部改変)から、最も関連性の高いと思われる測定可能な特徴量としての評価パラメータの候補を以下のように挙げた(一部掲載)。なお、以下の記述で、言語情報と書かれている箇所は、発話内容の統語解析、意味分析、修辞構造の分析等を必要とするものであり、今後、詳細化する予定である。

(1) 誠実な参加態度

下位項目	関連するパラメータ候補
議論に集中しているか(聴く態度を示しているか)	(話し手以外の)視線、あいづち等の有無、(しぐさ)
発言の根拠を提示するようこころがけているか	言語情報(修辞構造)
明確な発言をこころがけているか	言語情報(統語構造)、音声のパワー

(2) 対等な関係性

下位項目	関連するパラメータ候補
個人攻撃になっていないか	不同意数、言語情報(含・感情情報)
一部の参加者が議論の輪から外れていないか	各人のフロア数、発話交換(発言・応答ペア)の成立数
専門用語などわかりにくい表現を使っていないか	言語情報、問い返し(意味確認)数(-)
対立意見を頭から否定し、感情的に反応していないか	不同意数、言語情報(含・感情情報)
反論された際の反応として、沈黙や安易な同意を示していないか	(反論の)連鎖数、言語情報(含・主張の整合性)

(3) 議論の活発さ

下位項目	関連するパラメータ候補
誰かの発言中、傍観者となっていないか	応答者バランス、音声的同意表現の有無
誰かの発言に対して、応答しようとしているか	発話量、ポーズ量、発言ー応答ペアの成立数
発言者が偏っていないか	各人のフロア数
発言の受け止め役やまとめ役が固定されていないか	各人のフロア数、メタ議論 ⁶ 発言者バランス

(4) 以降省略

一例を挙げれば、「議論に集中しているか」どうかは、発言者の方を見ているか、あいづちなど、発話を了承する明示的な反応を表示しているか、髪をいじる、頬杖をつく、うつ伏せるなどのしぐさをしていないか、などと関連している。これらは、話し手以外の話者の視線あるいは顔の向き、誰かの発言中のあいづちや笑いなどの、音声的反応量を計測することで数量化可能であろう。なお、あいづちや笑いは、将来的には音声特徴量から特定することが可能になるだろうが、[4]では、あいづちや笑いが、話

⁶ メタ議論(的発言)とは、「ここまでの意見をまとめると」、「あと10分で結論を出しましょう」等、議論進行に関わる発言を指す。

し手の発話中に重ねて発せられる傾向があることから、発話オーバーラップ量として近似している。しぐさに関しては、現在のところ人手で判別する必要があるが、何らかの別の視覚的パラメータとして近似できるかもしれない。

以下では、一部の評価パラメータを試験的に話し合いデータに適用し、分析した結果を述べる。

4. 評価パラメータの試験的同定および分析例ー同意・不同意タグ、および応答ペアの同定

ここでは、2007年に実施された合意形成型ディスカッション・データに対して、先述の「(2)対等な関係性」を評価するための、評価パラメータの予備的分析例を紹介する。このデータは、大学生男女6人/1グループによるディスカッションを27対話収録したもので、各対話から切り出した27場面に対して、専用に作成された質問紙を使った印象評価の因子分析により得られた、5つの因子ごとの平均因子得点、良い議論(と思われる)議論順に順位付けをしてもらった結果の順位得点が算出されている(詳しくは[11][12]を参照のこと)。

4.1 話し合いのための談話タグ

(2)対等な関係性の指標における、「個人攻撃になっていないか」「一部の参加者が議論の輪から外れていないか」は、明らかな攻撃的発言や、あからさまな無視だけを指しているわけではなく、言葉の選択の問題や、発言の処理のされ方の不公平さ、などを含んでいる。これらを実験パラメータに落とし込むためには、参加者ごとの発話への注釈付け、および発話間の応答関係を考慮する必要がある。発話の注釈付けには、DAMSL[13]をはじめとして、日本でも発話単位タグの標準化[14]などが目指されているが、[15]でも指摘されているように、本研究で必要とするような同意/不同意表現に関する詳細化が未整備である上、話し合い独特の構造を理解するためには改良の必要があった。ゆえに、表2のような、話し合いにおける発話の特徴および同意/不同意表現に着目した、試験的な話し合いのための談話タグ(pDTD: probative Discourse Tag for Discussion)を規定した。話し合いにおいては、明確な同意/不同意表現が用いられる場合と、文脈からは同意/不同意と判断されるが、明示的な表現を用いていない場合があり、それらを区別するために、同意/不同意反応・発言を表2のように細分化している。

分析の対象としたのは、(2)対等な関係性指標の元となった、「参加者の関係性」因子の因子得点の正負、および順位得点の高低の組み合わせが異なる4場面である(表3参照。{因子:順位}が、{負;高(NG)}=#1-3-3, {負;低(NB)}=#0-2-3, {正;高(PG)}=#0-1-2, {正;低(PB)}

=#2-2-3)。これらの発話を単位化[13]した上で、それぞれの発話単位に、便宜的に一意にいずれかのタグを付与した。今回は予備的分析のために、筆者の一人がタグ付けをした結果を用いる。

表2 話し合いのための談話タグ(試験版)

pDTD	説明
propose	新規のあるいは改変された提案・主張
reason	提案・主張の理由付け
question	他者の発言に対する質問
answer	質問に対する応答
downgrade	主張の取り下げあるいは軟化
exemplify	主張に関する例示
substream	主題から離れたやり取り
agreement	同意応答・発言
agree	明示的な同意
agree/propose	同意+提案・主張
agree/reason	同意+理由付け
weak-agree	非明示的な同意
weak-agree/propose	非明示的な同意+提案・主張
weak-agree/reason	非明示的な同意+理由付け
disagreement	不同意応答・発言
disagree	明示的な不同意
disagree/propose	不同意+提案・主張
disagree/reason	不同意+理由付け
weak-disagree	非明示的な不同意
weak-disagree/propose	非明示的な不同意+提案・主張
weak-disagree/reason	非明示的な不同意+理由付け
reserve-agree	同意の保留

4.2 結果

各タグを集計したところ、disagreementの頻度が因子得点の低い場面に高い傾向にある(#1-3-3で発話の7.7%、#0-2-3で発話の12.3%。#0-1-2と#2-2-3は共に2.3%)一方で、順位得点の低い#0-2-3は、高い#1-3-3に比べagreementに対するdisagreementの割合が大きかった(40.9%に対して30.1%)。また、話し合いにおいては、propose, reason, answerに関しては、何らかの同意/不同意を示す評価的発言、questionに関しては、answerがあること(発言ー応答ペアと表現する)が期待される。その関係を示したのが表3であり、因子得点が負で、順位得点が低かった#0-2-3では、不同意の応答の割合以上に、提案や理由付け、回答の発言に対して応答・反応の不在(表中のφ)が目立つ。#0-2-3においては、提案に対する不同意発言も、応答の不在も共に特定の参加者の発言の後に集中していた。不同意的反応・発言があること自体は、議論においては当然であるが、特定の話者の発言のみ、一部の同意も反論もないこと、誰も反応せずに沈黙が流れ、別の話者が異なる提案をすることは、その参加者を議論の輪から遠ざけることとなり、結果#0-2-3の評価が低くなった一因と考えられる。ゆえに、少なくとも、話者ごとの発言に対する被不同意数および応答の不在数の全体に対する割合は、(2)対等な関係性指標の評価パラメータとして利用できる可能性が高い。

表 3 各データにおける主要な発話-応答ペア頻度

	1-3-3(NG)			0-2-3(NB)			0-1-2(PG)			2-2-3(PB)		
	count	rate	recipient	count	rate	recipient	count	rate	recipient	count	rate	recipient
propose-agree	11	50.0%	1.6	8	50.0%	1.1	21	80.8%	1.9	29	90.6%	2.3
propose-disagree	5	22.7%	1.0	2	12.5%	1.5	2	7.7%	1.0	2	6.3%	1.0
propose-φ	6	27.3%	-	6	37.5%	-	3	11.5%	-	1	3.1%	-
reason-agree	21	77.8%	1.3	9	52.9%	1.4	7	87.5%	1.9	17	85.0%	1.8
reason-disagree	4	14.8%	1.0	4	23.5%	1.8	1	12.5%	0.0	0	0.0%	0.0
reason-φ	2	7.4%	-	4	23.5%	-	0	0.0%	-	3	15.0%	-
question-answer	0	0.0%	0.0	4	100.0%	1.0	4	100.0%	2.0	6	100.0%	1.3
question-φ	0	0.0%	-	0	0.0%	-	0	0.0%	-	0	0.0%	-
answer-agree	0	0.0%	0.0	0	0.0%	0.0	3	75.0%	2.3	5	71.4%	1.6
answer-disagree	0	0.0%	0.0	1	25.0%	1.0	0	0.0%	0.0	1	14.3%	1.0
answer-φ	0	0.0%	-	3	75.0%	-	1	25.0%	-	1	14.3%	-

5. おわりに

本研究では、話し合いのための計測可能な評価パラメータの検討をした。今回は、各評価パラメータの候補を挙げる段階に留まり、実際に適用、分析できたのは、7つの評価項目に対する評価パラメータのごく一部であり、その精度（タグの妥当性も含む）も十分に検討できていない。今後これらのパラメータを精緻化・検証していく必要がある。また、実際にこれらの評価パラメータを採用し、実用化するには、これらパラメータの、各指標に対する重み付けを考慮する必要がある。また、同じ評価パラメータが異なる項目間、下位項目間で重複している部分があるのは、各項目そのものが、完全に独立したものではないためであり、分離（or 統合）可能性を検討する必要がある。さらに言語情報が必要なものに関しては、未だ具体性に欠け、今後先行研究を踏まえ、段階を追って継続してゆくべき課題である。

引用文献

[1] 小林傳司 (2004) 『誰が科学技術について考えるのかーコンセンサス会議という実験』, 名古屋大学出版会
 [2] 森本郁代 (2007) 話し合いデザインに向けたグループ・ディスカッションの分析と評価, 人工知能学会研究会資料 SIG-SLUD-A603-5,
 [3] 大塚裕子他 (2009) 科学技術コミュニケーションにおける対話のデザインー自律型対話の実践に向けて, 人工知能学会誌, vol.24, No.1, 78-87
 [4] 森本郁代他 (2006) グループ・ディスカッションの相互行為過程の評価と分析のための指標ーフォーカス・グループ・インタビューデータの分析からー, ヒューマンインタフェース学会論文誌, Vol.8(1), 117-128.

[5] 水上悦雄他 (2008) 話し合いにおけるコミュニケーションプロセスの評価法について. 言語処理学会第 14 回年次大会発表論文集, 181-184

[6] 鈴木佳奈他 (2008) グループディスカッションを相互行為として「評価」するー7つの評価項目の提案ー. 人工知能学会研究会資料, SIG-SLUD-A801-05, 29-34.

[7] 小磯花絵・伝康晴 (2000) 円滑な話者交替はいかにして成立するかー会話コーパスにもとづく分析ー. 認知科学, 7(1), 93-106.

[8] 水上悦雄他 (2007) 話し合いへの印象に影響を及ぼす会話行動: プロの司会者と素人の印象評定の比較および話し合いの相互行為過程の分析. 社会言語科学, 9 (2), 77-92.

[9] 徳久良子・寺島立太 (2006) 雑談における発話のやりとりと盛り上がりの関連. 人工知能学会論文誌, 21(2)A, 133-142.

[10] Yule, G. (1996) *Pragmatics*, Oxford University Press

[11] 水上悦雄他 (2008) 話し合いのプロセスを評価する際の諸問題の検討ーグループディスカッションの「対立」場面の分析を通じてー. 日本認知科学会第 25 回大会発表論文集, 408-411.

[12] Mizukami, E. et al. (2009) Two types of disagreement in group discussions of Japanese undergraduates. *Group Decision and Negotiation* (in printing).

[13] 人工知能学会「談話・対話研究におけるコーパス利用研究グループ」 (1999) 日本語スラッシュ単位ラベリングマニュアル

[14] Core, M. G., and Allen, J. F. (1997) Coding dialogs with the DAMSL annotation scheme. In Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines.

<http://citeseer.ist.psu.edu/core97coding.html>

[15] 荒木雅弘他 (1999) 発話単位タグ標準化案の作成, 人工知能学会誌, vol.14, No.2, 251-260

[16] 矢野博之・伊藤昭 (1999) 同意・不同意表現のための談話タグに関する一考察, 人工知能学会誌, vol.14, No.2, 290-295