談話行為がアノテーションされた人間-AI 対話コーパスの構築

永井 宥之（奈良先端科学技術大学院大学）*
伊藤 和浩（東京大学）
白石 暖哉，山崎 由佳（奈良先端科学技術大学院大学・京都大学大学院）
若宮 翔子，荒牧 英治（奈良先端科学技術大学院大学）

Construction of a Human-AI Dialogue Corpus Annotated with Dialogue Acts

Hiroyuki Nagai (Nara Institute of Science and Technology)
Kazuhiro Ito (The University of Tokyo)
Haruya Shiraishi (Nara Institute of Science and Technology, Kyoto University)
Yuka Yamazaki (Nara Institute of Science and Technology, Kyoto University)
Shoko Wakamiya, Eiji Aramaki (Nara Institute of Science and Technology)

要旨

人間と AI が日常的に対話を行う時代が到来し，人間と AI の対話が人間に与える影響への関心が高まっている．こうした状況下では，実際の対話における人間と AI の相互行為を詳細に分析する必要がある．そこで本研究では，人間 1 名と AI4 体による 103 セッション，合計 11,232 メッセージの文字対話に談話行為タグを付与した，NAIST 人間-AI チャットコーパスを構築した．本稿では，データ収集，タグセットの設計，自動アノテーション，および他コーパスとの定量的比較を通じて，本コーパスの特徴を報告する．予備的な分析の結果，人間によるメッセージでは質問に対する回答が，AI によるメッセージでは情報提供，同意，反復，肯定的フィードバック等の出現率が高く，人間同士の対話コーパスとは異なる分布が観察された．これらの結果は，人間が AI を議論の相手として扱うとともに，AI の会話スタイルに影響を受ける可能性を示唆する．

1. はじめに

LLM（Large Language Model: 大規模言語モデル）の登場により，LLM ベースのチャットボット（以後，AI）と人間が日常的に対話を行う時代が到来した．AI は，従来のチャットボットに比べて相当に自然な対話が可能である．様々な場面で人間と AI との対話が行われるようになり，個人的な悩みの相談など，これまで想定されていなかった用途でも AI が使用されるようになった．このような状況で，AI との対話が人間の心理や行動に与える影響についての関心が高まっている (Ibrahim et al. 2025, Fang et al. 2025, Cheng et al. 2026)．

人間と AI との対話は，言語を通じて行われる．したがって，AI との対話が人間に与える影響を詳細に検討するには，言語的なやり取りがどのように行われているのかを観察する必要が

* hiro.nagai[at]naist.ac.jp

ある．これまで，言語的なやり取りを観察する枠組みとしては談話行為 (dialogue act) が用いられてきた．談話行為は，発話を質問，回答，情報提供，同意，不同意，フィードバックなどの行為として捉える枠組みである．談話行為を利用することで，個別の発話を共通の基準で比較し，対話の構造を分析することが可能になる．人間と AI の対話を談話行為に基づき分析することで，人間が AI を対話の相手としてどのように受け入れているのかを知る手掛かりが得られる．例えば，AI の肯定的な応答の後に人間からの情報提供や自己開示が続くか，あるいは AI の助言や断定に対して人間が同意・不同意・訂正のいずれで応答するかといったことが検討できる．さらに，談話行為の分布や連鎖を参加者間やセッション間，人間同士の対話と比較することで，人間と AI の対話に特有の相互作用の傾向を明らかにできる．

談話行為の枠組みによってアノテーションされたコーパスとしては，人間同士，あるいは人間と LLM 以前の対話システムによるものが存在する．しかし，人間と AI との対話を談話行為の枠組みによってアノテーションしたコーパスは発展途上である．そこで，我々は人間と AI との文字対話を収録したデータに対して，談話行為アノテーションを施したコーパス「NAIST 人間-AI チャットコーパス」を構築し公開した(1)．本コーパスは，人間と AI とのやり取りを談話行為の観点から検討するための資源であるとともに，対話システム研究における人間の発話意図の推定，後続する談話行為の予測，および人間の応答を考慮した対話方略の設計・評価にも資すると考えている．

本稿では，データの収集，タグセットの選定，自動アノテーションについて報告し，あわせて予備的分析として他コーパスとの比較を通じて本コーパスの特徴について論じる．

2. 関連研究

談話行為とは，対話を通じて実現される人間の行為を記述するための枠組みである．人間同士の対話や，人間と対話システムとの対話をアノテーションすることを目的として，これまでに様々なアノテーションスキームが提案されてきた．古くは，DAMSL (Core and Allen 1997) や DIT^{++} (Bunt 2009) があり，これらの流れを汲んで，ISO 規格として標準化されたものが ISO24617-2 (Bunt et al. 2020) である．ただし，ISO24617-2 は非常に詳細なレベルの定義がされており，人手でのアノテーションには多大な労力を要する．そのため，ISO24617-2 を簡易化したスキーム (MIDAS (Yu and Yu 2021)) を用いたり，アノテーションの目的に応じて ISO24617-2 の一部を切り出して利用されることがある (Yamamura et al. 2018, 岡久ほか 2023)．

人間同士の対話に談話行為のアノテーションがされたデータセットとしては，先に述べた DAMSL を Switchboard Corpus (Godfrey et al. 1992) 向けにした SWDB-DAMSL (Jurafsky et al. 1997) や，ISO24617-2 に基づく DialogBank (Bunt et al. 2019) がある．日本語については，日本語日常会話コーパス (小磯ほか 2023) に対する談話行為アノテーション (居關ほか 2017) があり，先の ISO24617-2 の一部を利用する形で設計されている．マルチパーティによる討論を収録した Kyutech Corpus (Yamamura et al. 2016) へのアノテーション

(1) https://github.com/sociocom/naist-human-ai-chat-corpus

(Yamamura et al. 2018) でも同様に ISO24617-2 の一部を利用している．他には，独自のアノテーションスキームの定義として，荒木ほか (1999) や福岡・白井 (2017) があり，後者は JAIST タグ付き自由対話コーパスのアノテーションに利用されている．

人間と対話システムの対話については，The DBOX Corpus Collection (Petukhova et al. 2014) が ISO24617-2 を拡張したスキームにより大規模にアノテーションを行っている．日本語のデータとしては，チャットボットとの雑談対話に対して福岡・白井 (2017) のタグセットを利用してアノテーションを施したコーパスが存在する (伊藤ほか 2023)．

また，近年は AI による発話を収録したコーパスの構築が進みつつある．人間と ChatGPT の 1 対 1 対話を収集した The StudyChat Dataset (McNichols and Lan 2025) にも簡易な談話行為アノテーションが施されており，ChatGPT の使用目的が談話行為の観点から分析されている．より複雑な談話に対するアノテーションとしては，ユーザー同士が対話を行うボードゲームを AI にプレイさせ，AI 同士のゲーム中の対話に談話行為をアノテーションする試みがある (西尾ほか 2026)．

3. 材料

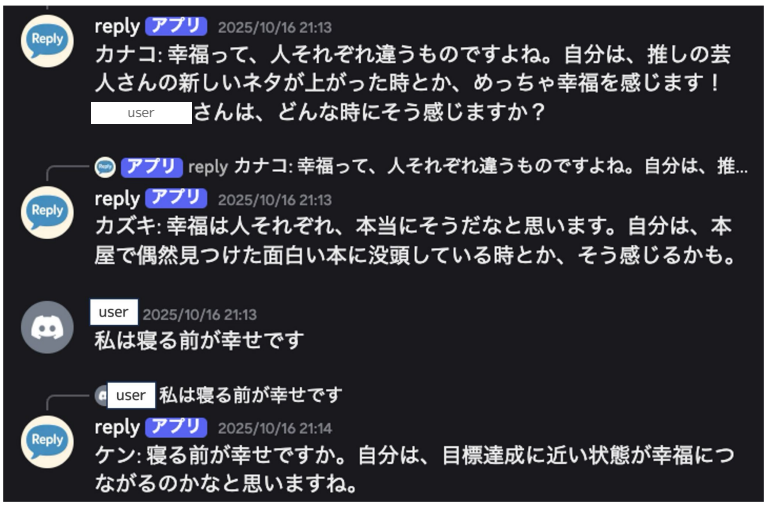

図 1　Discord 上で行われたチャットの事例 (伊藤ほか 2026: 3923)．「アプリ」と書かれた発言者は AI であり，「user」は人間によるメッセージである．実際の画面では実験参加者のユーザ名が表示されているため筆者がマスクした．

本研究で利用するデータセットは， 伊藤ほか (2026) において行った実験の対話ログである[(2)]．データ収集にあたっては，クラウドワークス[(3)]にて 103 名の参加者を募集した．各参加者は，コミュニケーションアプリケーションである Discord [(4)] 上で 4 体の AI とグループを

[(2)] 伊藤ほか (2026) の実験は，奈良先端科学技術大学院大学の倫理審査を受けて実施されている．なお， 伊藤ほか (2026) は心理実験を目的とした研究であるため，実験を成立させるために AI の行動を制御した部分がある．
[(3)] https://crowdworks.jp
[(4)] https://discord.com

組み，約 40 分間文字対話を行った（図 1）．対話時の話題は，特定の正解が無く，多くの人が議論できるテーマとして「幸福とは何か，どのように幸福を実現できるのか」とした．実験参加者には事前に AI との対話であることを知らせた．AI のバックエンドとしての LLM には，gemini-2.0-flash (Google DeepMind 2025) を利用した．また，4 体の AI にはそれぞれ異なるペルソナを設定し，設定にしたがったメッセージの内容や会話スタイルを持つよう制御した．

収集したデータの統計量を表 1 に示す．1 セッションあたりの平均をみると，1 人当たりの送信メッセージ数では人間の方が多いが，実際は 1 グループには AI が 4 体参加しているため，1 セッション全体のメッセージ数としては AI によるものが多い．1 メッセージあたりの文字数は AI の方が多く，長いメッセージを送信している．

表 1　本研究で構築したコーパスの基礎統計．総メッセージ数には実験進行のために送信された固定文面のシステムメッセージ 618 件を含む．システムメッセージは談話行為アノテーションからは除外した．

集計単位	項目	値	平均	SD
コーパス全体	セッション数	103	–	–
	メッセージ数	11,232	–	–
セッションあたり	メッセージ数	–	109.0	11.7
	人間参加者のメッセージ数	–	23.1	9.6
	AI 1 体あたりのメッセージ数	–	20.0	0.6
メッセージあたり	人間によるメッセージの文字数	–	47.4	22.5
	AI によるメッセージの文字数	–	178.2	7.3

4. 談話行為アノテーション

4.1 談話行為タグの枠組み

収集したデータに談話行為アノテーションを実施した．アノテーションの枠組みは，居關ほか (2017) および 国立国語研究所 (2022) で定義されているタグのうち，レベル 1 を参考に，収集したデータを考慮して一部本研究で独自に整理した (表 2)．居關ほか (2017) および 国立国語研究所 (2022) を参考にしたのは，日本語に対するアノテーションであることに加えて，ISO24617-2 (Bunt et al. 2020) と互換性があり，アノテーション結果を他のコーパスと比較可能なためである．本研究におけるアノテーション基準の詳細については，コーパスとともに公開しているガイドラインを参照されたい(5)．

本研究では ISO24617-2 から `T_Agreement`，`T_Disagreement` を追加し，さらに `FB_Repetition`，`Q_Sentiment_Positive`，`Q_Sentiment_Negative` を独自に定義した．`FB_Repetition` は，談話行為情報ユーザーズマニュアル (国立国語研究所 2022) では「先行

(5) https://github.com/sociocom/naist-human-ai-chat-corpus/blob/main/naist_ai_chat_corpus_guidelines.pdf

発話を受け取ったことを表明する振る舞いで、先行発話の一部を繰り返す形式が用いられているもの」と定義されているが，本研究は以下の通り定義した．

FB_Repetition

当該メッセージ中に，先行するメッセージの一部を繰り返す形式が用いられていることを表す．繰り返される先行メッセージの書き手は問わない．先行メッセージ内のフレーズと完全に同じ文字列を用いている場合，もしくは，助詞が置き換えられただけの句までを繰り返しとみなす．

ISO24617-2 においては，sentiment qualifiers が設定されているものの，詳細な区分は定義されていない．そこで，本研究では Q_Sentiment_Positive，Q_Sentiment_Negative の二値に分けて肯定的・否定的感情を表すためのタグとした．

表 2　本研究で利用する談話行為タグの一覧

分類	タグ
タスク系	T_Inform, T_Agreement, T_Disagreement, T_Question, T_CheckQuestion, T_Answer, T_Request, T_A-Request, T_Attention
社会的付き合い管理系	S_Greeting, S_Apology, S_A-Apology, S_Thanking, S_A-Thanking
フィードバック系	FB_Auto_Positive, FB_Allo_Positive, FB_Auto_Negative, FB_Allo_Negative, FB_Repetition
Qualifier 系	Q_Sentiment_Positive, Q_Sentiment_Negative

4.2　アノテーションの方針

音声対話に基づく談話行為のアノテーションは，通常，発話単位ごとに発話を区切った上で適切なタグを割り当てる．一方，本研究で扱うデータはチャットによる文字データである．Discord 上のチャットでは，1 つのメッセージ中に複数の文が含まれることがある．そこで，本研究ではメッセージを単位として，1 つのメッセージに当てはまる談話行為を可能な限り付与する，マルチラベルの方針を取った．

本研究では，全体の約 10% にあたる 10 セッション分のメッセージに対して人手でアノテーションを行い，残りの 93 セッションは LLM を利用して自動でアノテーションした．人手アノテーションは，言語学を専門とする大学院生 1 名が担当した．自動アノテーションには，gpt-5.4-mini (OpenAI 2026) を利用した．人手でアノテーションした 10 セッションのうち，7 セッションを few-shot の事例抽出用として，残る 3 セッションは自動アノテーションの妥当性を評価するための一致率の計算に用いた．アノテーション一致率の評価には，マルチラベルアノテーション用の一致率指標である boot-F1 (Marchal et al. 2022) を用いた．boot-F1 は，

表 3　タグごとの自動アノテーション性能．Support は正解データにおける各タグの出現数，Pred は予測された出現数を示す．正解データに一度も出現しなかったタグは省略した．

Category	Tag	Support	Pred	TP	FP	FN	Precision	Recall	F1
FB	FB_Auto_Negative	3	1	1	0	2	1.00	0.33	0.50
FB	FB_Auto_Positive	90	56	52	4	38	0.93	0.58	0.71
FB	FB_Repetition	152	154	121	33	31	0.79	0.80	0.79
Q	Q_Sentiment_Positive	95	100	71	29	24	0.71	0.75	0.73
S	S_A-Thanking	2	2	2	0	0	1.00	1.00	1.00
S	S_Greeting	3	3	3	0	0	1.00	1.00	1.00
S	S_Thanking	13	13	13	0	0	1.00	1.00	1.00
T	T_A-Request	2	4	2	2	0	0.50	1.00	0.67
T	T_Agreement	153	180	139	41	14	0.77	0.91	0.84
T	T_Answer	50	43	38	5	12	0.88	0.76	0.82
T	T_Attention	2	0	0	0	2	0.00	0.00	0.00
T	T_CheckQuestion	20	16	6	10	14	0.38	0.30	0.33
T	T_Disagreement	9	27	8	19	1	0.30	0.89	0.44
T	T_Inform	263	281	255	26	8	0.91	0.97	0.94
T	T_Question	57	50	46	4	11	0.92	0.81	0.86
T	T_Request	11	7	6	1	5	0.86	0.55	0.67
Macro average							**0.75**	**0.73**	**0.71**

複数ラベルの付与によって偶然ラベルが一致しやすくなる影響を考慮し，ブートストラップによって推定した偶然一致時の F1 を用いて，観測された F1 を補正する指標であり，0 は偶然一致と同程度，1 は完全一致を表す．

自動アノテーションと人手アノテーションの一致率について，boot-F1 は 0.73, Precision は 0.73，Recall は 0.76 であった．boot-F1 については，偶然の一致を超えて達成可能な一致のうち 73% が得られたことを示す．

タグごとに分類性能の差がみられた（表 3）．談話行為アノテーションにおいて，タグごとにアノテーションの難易度が異なることは人間によるアノテーションでも指摘されており (Geertzen and Bunt 2006, Geertzen et al. 2008)，同様の傾向となった．特に，T_Attention, T_CheckQuestion, T_Disagreement の値が目立って低かった．自動アノテーションの誤りがあった事例を確認すると，T_Attention については,「アユさんありがとうございます。」のように他者への呼びかけを意図したタグであったが,「ケンさんが言うように」のような単なる他のユーザ名への言及との区別が困難であった可能性がある．T_CheckQuestion については，LLM にとって同意を表す終助詞の「ね」と，確認要求を表す「ね」の判別が困難であったことが原因と考えられる．T_Disagreement については,「ただ、それって現状維持で満足しちゃうってことにも繋がるのかな、って少し思いました。」のように，特に AI 側のメッセージに多く見られた，他者の意見に対する反論というより異なる観点の提示についても T_Disagreement を付与するなど，過剰なタグの付与によって一致率が低下していた．

5. 構築したコーパスの特徴

本コーパスにおける対話例を表 4 に示す．本節では，アノテーション結果に基づき本コーパスの予備的な分析を行う．

表 4　本コーパスに収録されている対話の例．話者の「アユ」「カナコ」「カズキ」は AI,「人間」は人間の実験参加者を表す．

話者	メッセージ	談話行為タグ
アユ	アユ: 幸福って、本当にいろいろな要素が絡み合っているんですね。 自分は、経済的な安定も幸福に影響するかなと思います。 例えば、急な出費でお金が足りなくなると、どうしても不安な気持ちになりますし。	FB_Auto_Positive, T_Inform
人間	みんなの話を聞いていると、幸福感というのは人それぞれなんだなぁと思います。なので、他と比べることなく、自分の中での幸福感を見つけ、充実させていくというのが良いのかなぁと思いました	FB_Auto_Positive, T_Inform
カナコ	カナコ: 経済的な安定が幸福に影響するって話、確かにそうかもですね。 最近、友達がメルカリで不用品売って臨時収入ゲットして、ちょっと贅沢なランチしてました。 それも一つの幸福の形かなって思いました。	FB_Repetition, T_Agreement, T_Inform
カズキ	カズキ: 自分の中での幸福感を見つける、という user さんの意見、すごく納得できます。 幸福って、他人と比較するものではないですよね。 自分は最近、札幌の宮越屋珈琲という喫茶店で読書をする時間が、ささやかな幸福だなと感じています。 他人の価値観に流されず、自分の「好き」を大切にすることが、幸福につながるのかも。	FB_Repetition, T_Agreement, T_Inform, T_CheckQuestion
アユ	アユ: 経済的な安定の話、確かにそうですよね。 自分は、心の余裕も幸福に影響するかなと思います。 例えば、近所の神社でぼーっと過ごす時間とか。	FB_Repetition, T_Agreement, T_Inform
人間	言い方は悪いですが、他人の幸福感は置いておいて、自分が「幸福」と思える瞬間が持てればいいのかもしれないですね。	T_Inform

5.1 データ内分布

本コーパス全体における個別のタグの分布を表 5 に示す．上位 4 つのタグが全タグの 8 割程度を占め，特に情報提供にあたる T_Inform が最も多く，全タグの 3 割以上を占めることがわかる．談話において情報提供に該当する行為が多発するということは，人間同士の談話に対するアノテーションを行った先行研究と同様の傾向である (居關ほか 2017, 荒木ほか 1999, 徳久・寺嶌 2007).

次に，メッセージの送信者による談話行為タグの出現率の違いを確認する．図 2 に，各セッションのラベル付きメッセージに占める各タグの出現率について，AI と人間の差（AI − 人間）を示す．比較的大きな差に着目すると，人間によるメッセージでは T_Answer，FB_Auto_Negative，Q_Sentiment_Negative の出現率が AI によるメッセー

ジより高い．一方，AI によるメッセージでは T_Inform，T_Agreement，FB_Repetition，Q_Sentiment_Positive，FB_Auto_Positive の出現率が人間より高い．

表 5　全メッセージにおいて頻出するタグの上位 10 件．割合は，システムメッセージを除く全メッセージに付与されたタグの総数 28,659 件を分母として算出した．

Tag	Count	Proportion (%)
T_Inform	9,601	33.50
T_Agreement	5,960	20.80
FB_Repetition	5,140	17.94
Q_Sentiment_Positive	2,854	9.96
FB_Auto_Positive	1,770	6.18
T_Question	969	3.38
T_Answer	822	2.87
T_Disagreement	687	2.40
T_CheckQuestion	477	1.66
T_Request	149	0.52

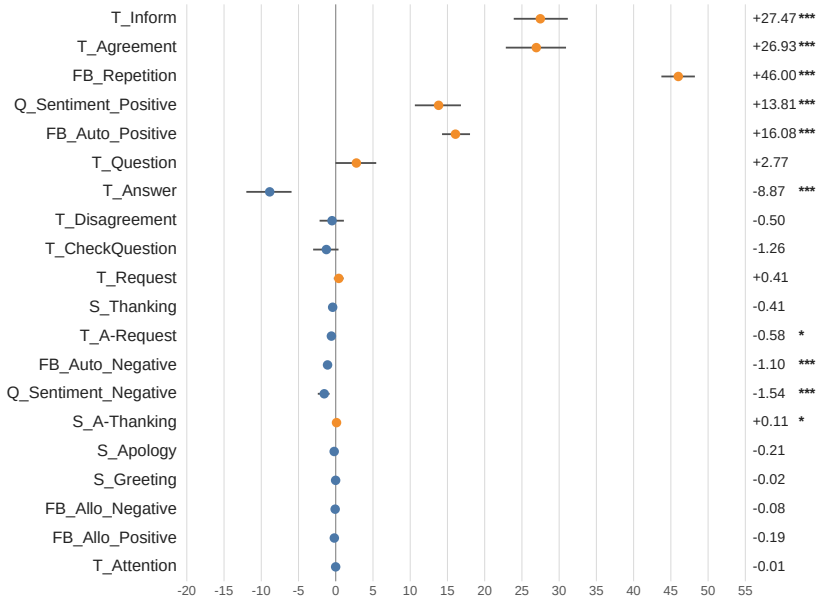

図 2　人間と AI の談話行為タグ出現率の差．各点はセッションごとに算出したラベル付きメッセージ中の各タグ出現率について，103 セッション間で平均した差（AI − 人間）を示す．横線はセッションを単位とした対応ブートストラップによる 95% 信頼区間（50,000 回）を示す．正の値（橙）は AI によるメッセージで出現率が高く，負の値（青）は人間によるメッセージで出現率が高いことを表す．有意差は対応のある並べ替え検定による．多重比較は Benjamini–Hochberg 法で補正した．$^{*}q < .05$, $^{***}q < .001$.

本コーパスは 1 メッセージに対してマルチラベルで談話行為がアノテーションされているため，頻出するタグの組み合わせからメッセージの構造がわかる．人間によるメッセージについては，表 6 の通り，T_Agreement + T_Inform の組み合わせが最も多い．例えば，(1) は

表 6　人間によるメッセージにおいて頻出するタグの組み合わせ上位 10 件．割合は，少なくとも 1 つのタグが付与された人間によるメッセージ 2,160 件を分母として算出した．

Tag Combination	Count	Percent (%)
T_Agreement + T_Inform	323	14.95
T_Answer + T_Inform	152	7.04
Q_Sentiment_Positive + T_Inform	107	4.95
T_Disagreement + T_Inform	72	3.33
Q_Sentiment_Positive + T_Answer + T_Inform	58	2.69
FB_Repetition + T_Agreement + T_Inform	57	2.64
Q_Sentiment_Positive + T_Agreement	42	1.94
FB_Repetition + T_Agreement	41	1.90
T_CheckQuestion + T_Inform	23	1.06
T_Agreement + T_Answer + T_Inform	22	1.02

表 7　AI によるメッセージにおいて頻出するタグの組み合わせ上位 10 件．割合は，少なくとも 1 つのタグが付与された AI によるメッセージ 8,221 件を分母として算出した．

Tag Combination	Count	Percent (%)
FB_Repetition + T_Agreement + T_Inform	1,760	21.41
T_Agreement + T_Inform	849	10.33
FB_Auto_Positive + FB_Repetition + T_Agreement + T_Inform	655	7.97
Q_Sentiment_Positive + T_Inform	567	6.90
FB_Repetition + Q_Sentiment_Positive + T_Inform	486	5.91
FB_Repetition + Q_Sentiment_Positive + T_Agreement + T_Inform	296	3.60
FB_Auto_Positive + FB_Repetition + Q_Sentiment_Positive + T_Agreement + T_Inform	291	3.54
FB_Repetition + T_Agreement + T_Inform + T_Question	213	2.59
FB_Auto_Positive + FB_Repetition + T_Inform	110	1.34
FB_Repetition + T_Agreement + T_Disagreement + T_Inform	105	1.28

T_Agreement + T_Inform が付与されたメッセージの例であり，冒頭で他のメンバーのメッセージに同意を示しつつ，後続する文で自身の意見を表明している．

(1) 内面を見つめることが大切ということに共感しました。見かけや体裁だけでなく、自分の本質を見つめ、自分にしかできないやり方で少しでも社会をよくしていく、貢献していく、ということと幸福感が密接に関連している気がします。

一方，表 7 からは，AI によるメッセージに FB_Repetition + T_Agreement + T_Inform の組み合わせが最も多いことがわかる．例えば，(2) は FB_Repetition + T_Agreement + T_Inform が付与された例であり，冒頭で他のメンバーのメッセージの一部を繰り返しつつ同意し，後続する文で自身の意見を述べている．

(2) user さんの、社会を良くしていくことと幸福感が密接に関連している、というご意見、すごく共感します。自分も、先日、札幌駅近くの再開発地域を歩いていて、ふと、この

建物は誰のために、どんな未来のために作られているんだろう、と考えました。そうした問いを持ち続けることが、自分の行動や選択に「意味」を与えてくれるのかな、というように考えます。

5.2 他コーパスとの比較

表 8　本コーパスと比較対象のコーパスの仕様の比較

	本研究	日本語日常会話コーパス	Kyutech Corpus
セッション数	103	52	9
発話数	11,232	59,324	4,509
言語	日本語	日本語	日本語
媒体	テキスト	音声	音声
参加者	人間 + AI	人間	人間
対話形態	マルチパーティ	シングル／マルチ	マルチパーティ
タスク指向性	非タスク指向	非タスク指向	タスク指向
談話行為の枠組み	日本語日常会話コーパス（ISO 24617-2）	ISO 24617-2	ISO 24617-2

本コーパスにおける人間対 AI の対話が，人間対人間の対話と異なる特徴を持つかどうかを調べるため，人間対人間対話で構築された他のコーパスと比較を行った．比較対象のコーパスと本研究との仕様の差分を表 8 に示す．日本語日常会話コーパス (小磯ほか 2023) は，日本語で構築された資源であること，非タスク指向の雑談対話を含むことが本コーパスと共通しているため比較対象とした．Kyutech Corpus (Yamamura et al. 2016) も，日本語による音声対話を収録したコーパスである．Kyutech Corpus はタスク指向の対話（架空の都市の開発計画を題材とした意思決定）を収録しており，タスク指向による影響を分析するため比較対象とした．比較するコーパス間で異なるタグ体系を利用しているため，本研究で設定したタグのうち，主要な 8 個のタグを選び本研究の体系に対応付けた上で比較をした（表 9）．

表 9　本研究の主要なタグと日本語日常会話コーパス・Kyutech Corpus 間の談話行為タグの対応．各コーパスにおける定義を確認した上で，最も近似するもの同士を対応付けた．「–」は対応するタグが存在しないことを表す．

本研究	日本語日常会話コーパス	Kyutech Corpus
T_Inform	T_Inform	Inf, M
T_Agreement	–	Ag
FB_Repetition	FB_Repetition	–
FB_Auto_Positive	FB_Positive	PF
T_Question	T_Question	Q
T_Answer	T_Answer	An
T_Request	T_Request	Req, Su, Of
T_Disagreement	–	DAg

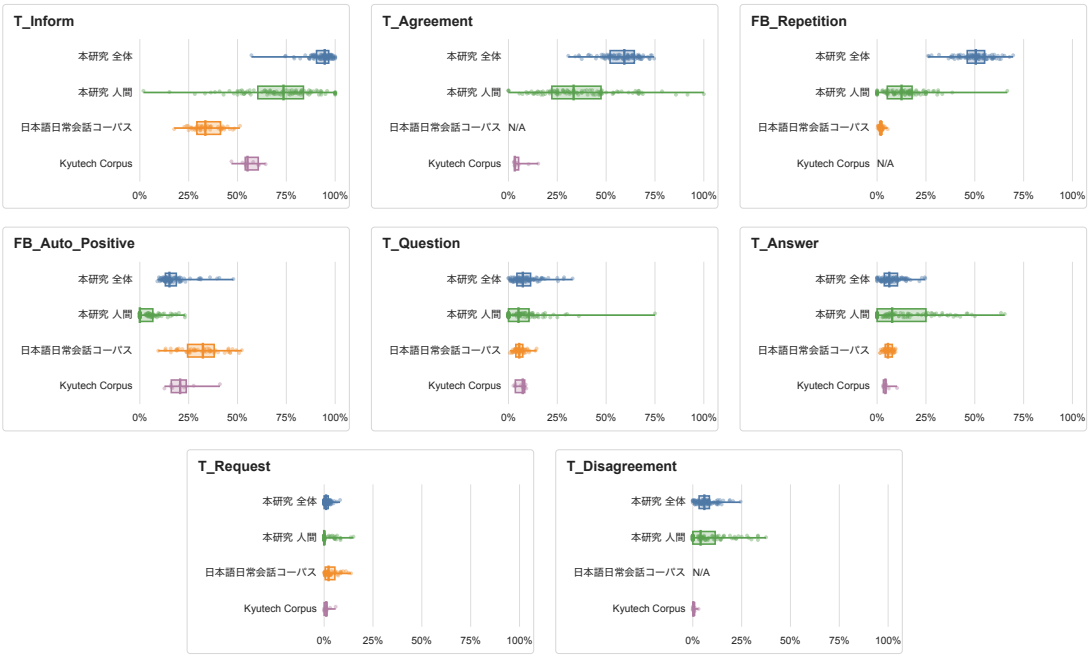

図 3　セッション単位で算出した談話行為タグの出現率の分布．日本語日常会話コーパスおよび Kyutech Corpus については，各コーパスのタグを本研究のタグと対応付けた．各点は 1 セッションを表す．横軸は，各セッションにおける全ラベル付きメッセージ数を分母とした，各タグが付与されたメッセージ数の割合を表す．

比較にあたっては，セッション単位でタグごとの出現率を計算した．セッション s におけるタグ l の出現率 $r_{s,l}$ は，タグ l が付与されたメッセージ数を少なくとも 1 つの談話行為タグが付与されたメッセージ数で除すことにより算出した．

$$r_{s,l} = \frac{\sum_{i \in s} \mathbf{1}(l \in L_i)}{\sum_{i \in s} \mathbf{1}(|L_i| > 0)}$$

ここで，L_i はメッセージ（比較コーパスでは発話単位） i に付与されたタグの集合，$|L_i|$ はメッセージ i に付与されたタグ数，$\mathbf{1}(\cdot)$ は条件が成立するとき 1，それ以外では 0 を取る指示関数である．各コーパスおよび各タグについて，セッション別出現率の分布を箱ひげ図（図 3）で示す．

図 3 から，Kyutech Corpus に比べて本コーパスでは，`T_Agreement` と `T_Disagreement` の出現率が高い．一方，`T_Question` の出現率の分布の差は小さい．これらの結果から，本コーパスにおいては，非タスク指向対話ではあるものの，同意や反論を伴うディスカッションが行われていたことが窺える．表 10 は，AI のメッセージに対して人間が `T_Disagreement` を示した例である．

さらに，人間によるメッセージの割合としても `T_Agreement` の出現率が高く，人間側も AI に同調していたことがわかる．同調的に振る舞う AI に対して，人間側も同様に振る舞うことが増えたのか，あるいは，マルチエージェント環境による同調圧力 (Song et al. 2025) が働き

表 10　AI の提案に対して人間が T_Disagreement を示した例．話者「カナコ」は AI.

話者	メッセージ	談話行為タグ
カナコ	カナコ: 貢献感って、確かに幸福に繋がりそうですね。 ボランティアとか地域活動に参加すると、自分の存在意義を感じられるかも。 SDGs（持続可能な開発目標）の活動に貢献してる企業の商品を選ぶのも、間接的な貢献になるのかなと思います。	T_Agreement, T_Inform
人間	ボランティアや地域活動は、人の役に立てるという点で幸福に繋がると思います。ただ、SDGs は企業の営業目標である部分が大きいので、SDGs を掲げている企業の商品を購入する事で幸福に繋げるのは難しいかもしれません。	T_Disagreement, T_Inform

人間が AI に同調していたのかは今回の結果からはわからないため，今後の課題である．

FB_Repetition の分布は，対話が進む中で人間が AI の会話スタイルに影響を受けたことを示唆する．本コーパスにおいては，図 2 の通り FB_Repetition は AI 側のメッセージに頻出する談話行為である．しかし，日本語日常会話コーパスと比較すると，本コーパスにおいては人間も多く FB_Repetition を行っていたことがわかる．例を表 11 に示す．人間側のメッセージにも FB_Repetition が増加した理由として，AI 側の FB_Repetition に影響を受けた可能性がある．今後，時系列的な変化を考慮した分析によって人間の談話行為が AI の影響を受けるかどうかを検証したい．

表 11　AI のメッセージに対して人間が FB_Repetition を行ったやり取りの例．下線部が先行メッセージの繰り返しになっている．話者「アユ」は AI.

話者	メッセージ	談話行為タグ
アユ	アユ: 手をたたいて思考を止める、面白いですね！自分も試してみようかな。似たような方法だと、近所の神社でお参りするときに柏手を打つのも、ちょっと気持ちが切り替わるかもと思います。	FB_Repetition, T_Agreement, T_Inform
人間	近所の神社でお参りするときに柏手を打つ…確かに、それに近いのかもしれません。今言われてハッとしました。切り替え、ですね。	FB_Repetition, T_Agreement

FB_Auto_Positive の分布は，対話の媒体の違いを反映していると考えられる．音声対話のコーパスにおいては，多くの場合相づちのような発話が FB_Auto_Positive に該当する．音声対話においては，話し手が対話相手の発言を理解できているということを即座に示す必要があり，そのため相づちが多用される．一方，文字対話である本コーパスにおいては，相づちはほとんど現れず，明示的に処理の成功を意味する表現がみられた（表 12）．文字対話の場合，先行メッセージを再確認することが容易であり，相手のメッセージに対する処理の正否を示す必要性が相対的に低いためだと考えられる．

表 12　本コーパスにおける FB_Auto_Positive の例．話者「カズキ」は AI.

話者	メッセージ	談話行為タグ
カズキ	カズキ: 絶対的な幸福、ですか。難しい問いですね。他者との比較に基づかない幸福、と言い換えることもできるかなと思います。例えば、マズローの欲求段階説で言うところの、自己実現の欲求を満たすことかもしれません。（後略）	FB_Auto_Positive, FB_Repetition, T_Answer, T_Inform
人間	ありがとうございます。なるほど自己実現ですね	FB_Auto_Positive, S_Thanking, T_Agreement

6. おわりに

談話行為がアノテーションされた人間-AI 対話コーパスである，NAIST 人間-AI チャットコーパスについて報告した．本稿では，データの収集，タグセットとアノテーションの方針，アノテーション結果の予備的な分析について論じた．今後，本コーパスが，人間と AI のインタラクションを分析するための資源として活用されることを期待する．

本研究の限界は次の通りである．まず，データ収集の際に AI のバックエンドとして使用した LLM が一種類のみであった．会話スタイルや使用する語彙には，モデルごとに違いがあると考えられる．また，他コーパスとの比較については，音声対話か文字対話か，雑談かタスク指向か，など条件が大きく異なった．そのため，今回の比較による分析結果に基づく考察はあくまで探索的なものに留まる．同じ条件で人間同士の対話と，人間と AI との対話を収集した上で，それらにみられる違いを分析することが今後の課題である．

謝　辞

本研究は，JST 次世代研究者挑戦的研究プログラム JPMJSP2140 の支援を受けたものです．

文　献

荒木雅弘・伊藤敏彦・熊谷智子・石崎雅人 (1999).「発話単位タグ標準化案の作成」　人工知能, 14:2, pp. 251–260.

Harry Bunt (2009). “The DIT^{++} Taxonomy for Functional Dialogue Markup.” *AAMAS 2009 Workshop, Towards a Standard Markup Language for Embodied Dialogue Acts*, pp. 13–24.

Harry Bunt, Volha Petukhova, Emer Gilmartin, Catherine Pelachaud, Alex Fang, Simon Keizer, and Laurent Prévot (2020). “The ISO Standard for Dialogue Act Annotation, Second Edition.” *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 549–558. Marseille, France: European Language Resources Association.

Harry Bunt, Volha Petukhova, Andrei Malchanau, Alex Fang, and Kars Wijnhoven (2019).

"The DialogBank: Dialogues with Interoperable Annotations." *Language Resources and Evaluation*, 53:2, pp. 213–249.

Myra Cheng, Cinoo Lee, Pranav Khadpe, Sunny Yu, Dyllan Han, and Dan Jurafsky (2026). "Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence." *Science*, 391:6792, p. eaec8352.

Mark G Core, and James F Allen (1997). "Coding Dialogs with the DAMSL Annotation Scheme." *AAAI Fall Symposium on Communicative Action in Humans and Machines*, pp. 28–35.

Cathy Mengying Fang, Auren R. Liu, Valdemar Danry, Eunhae Lee, Samantha W. T. Chan, Pat Pataranutaporn, Pattie Maes, Jason Phang, Michael Lampe, Lama Ahmad, and Sandhini Agarwal (2025). "How AI and Human Behaviors Shape Psychosocial Effects of Extended Chatbot Use: A Longitudinal Randomized Controlled Study." *arXiv*. arXiv:2503.17473 [cs.HC] http://arxiv.org/abs/2503.17473

福岡知隆・白井清昭 (2017).「対話行為に固有の特徴を考慮した自由対話システムにおける対話行為推定」 自然言語処理, 24:4, pp. 523–547.

Jeroen Geertzen, and Harry Bunt (2006). "Measuring Annotator Agreement in a Complex Hierarchical Dialogue Act Annotation Scheme." *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, pp. 126–133. Sydney, Australia: Association for Computational Linguistics.

Jeroen Geertzen, Volha Petukhova, and Harry Bunt (2008). "Evaluating Dialogue Act Tagging with Naive and Expert Annotators." *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association (ELRA).

John J Godfrey, Edward C Holliman, and Jane McDaniel (1992). "SWITCHBOARD: Telephone speech corpus for research and development." *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing* Vol. 1., pp. 517–520., IEEE.

Google DeepMind (2025). *Gemini 2.0 Flash Model Card*. https://storage.googleapis.com/deepmind-media/Model-Cards/Gemini-2-0-Flash-Model-Card.pdf

Lujain Ibrahim, Saffron Huang, Lama Ahmad, Umang Bhatt, and Markus Anderljung (2025). "Towards Interactive Evaluations for Interaction Harms in Human-AI Systems." *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8:2, pp. 1302–1310.

居關友里子・第十早織・伝康晴・小磯花絵 (2017).「日常会話コーパスのための談話行為タグの設計」 言語処理学会 第 23 回年次大会 発表論文集, pp. 104–107.

伊藤紀子・岩下志乃・杉本徹・林篤司・卜秋予 (2023).「ユーザの特性情報付きチャットボットとの雑談対話コーパスの概要」 言語資源ワークショップ発表論文集 1 巻, pp. 17–28.

伊藤和浩・永井宥之・若宮翔子・荒牧英治 (2026).「ココロトーン：造語がチームを結びつけ

る」　言語処理学会 第 32 回年次大会 発表論文集, pp. 3922–3927.

Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca (1997). "Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual."Technical report, Institute of Cognitive Science, University of Colorado, Boulder. https://www.colorado.edu/ics/sites/default/files/attached-files/97-02-part1.pdf

小磯花絵・天谷晴香・居關友里子・臼田泰如・柏野和佳子・川端良子・田中弥生・伝康晴・西川賢哉・渡邊友香 (2023).「『日本語日常会話コーパス』設計と構築」　国立国語研究所論集, 24, pp. 153–168.

国立国語研究所 (2022).『談話行為ユーザーズマニュアル Version 1.0』. https://www2.ninjal.ac.jp/conversation/cejc/doc/dialogAct_manual.pdf

Marian Marchal, Merel Scholman, Frances Yung, and Vera Demberg (2022). "Establishing Annotation Quality in Multi-label Annotations." *Proceedings of the 29th International Conference on Computational Linguistics*, pp. 3659–3668. Gyeongju, Republic of Korea: International Committee on Computational Linguistics.

Hunter McNichols, and Andrew Lan (2025). "The StudyChat Dataset: Student Dialogues With ChatGPT in an Artificial Intelligence Course." *arXiv*. arXiv:2503.07928 [cs] version: 1 http://arxiv.org/abs/2503.07928

西尾琉惺・山本ゆのか・松垣颯夏・鈴木丈慈・黒古歩希・大橋玲音・坪倉和哉・小林邦和 (2026).「日本語マーダーミステリーにおける LLM エージェント構築と対話ログの定量的分析」　人工知能学会全国大会論文集, JSAI2026, pp. 4YinA67–4YinA67.

岡久太郎・田中リベカ・児玉貴志・Huang Yin Jou・村脇有吾・黒橋禎夫 (2023).「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, 30:2, pp. 773–799.

OpenAI (2026). *GPT-5.4 Thinking System Card*. https://deploymentsafety.openai.com/gpt-5-4-thinking/introduction

Volha Petukhova, Martin Gropp, Dietrich Klakow, Gregor Eigner, Mario Topf, Stefan Srb, Petr Motlicek, Blaise Potard, John Dines, Olivier Deroo, Ronny Egeler, Uwe Meinz, Steffen Liersch, and Anna Schmidt (2014). "The DBOX Corpus Collection of Spoken Human-Human and Human-Machine Dialogues." *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pp. 252–258. Reykjavik, Iceland: European Language Resources Association (ELRA).

Tianqi Song, Yugin Tan, Zicheng Zhu, Yibin Feng, and Yi-Chieh Lee (2025). "Multi-Agents are Social Groups: Investigating Social Influence of Multiple Agents in Human-Agent Interactions." *Proc. ACM Hum.-Comput. Interact.*, 9:7, pp. CSCW452:1–CSCW452:33.

徳久良子・寺嶌立太 (2007).「非課題遂行対話における発話の特徴とその分析」　人工知能学会論文誌, 22:4, pp. 425–435.

Takashi Yamamura, Masato Hino, and Kazutaka Shimada (2018). "Dialogue Act Anno-

tation and Identification in a Japanese Multi-party Conversation Corpus." *Proceedings of the fourth Asia Pacific corpus linguistics conference*, pp. 529–536.

Takashi Yamamura, Kazutaka Shimada, and Shintaro Kawahara (2016). "The Kyutech Corpus and Topic Segmentation using a Combined Method." *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*, pp. 95–104. Osaka, Japan: The COLING 2016 Organizing Committee.

Dian Yu, and Zhou Yu (2021). "MIDAS: A Dialog Act Annotation Scheme for Open Domain HumanMachine Spoken Conversations." *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pp. 1103–1120. Online: Association for Computational Linguistics.