

日本語を母語とする自閉スペクトラム症者の公開コーパス構築

鈴木 あすみ (国立障害者リハビリテーションセンター研究所／東北大学大学院) †

幕内 充 (国立障害者リハビリテーションセンター研究所)

和田 真 (国立障害者リハビリテーションセンター研究所)

中村 仁洋 (国立障害者リハビリテーションセンター研究所)

石井 亨視 (国立障害者リハビリテーションセンター研究所)

小磯 花絵 (国立国語研究所)

Construction of a Public Corpus of Japanese Native Speakers with Autism Spectrum Disorder.

Asumi Suzuki (Research Institute of National Rehabilitation Center for Persons with Disabilities (NRCD) / Graduate School of Arts and Letters, Tohoku University)

Michiru Makuuchi (Research Institute of NRCD)

Makoto Wada (Research Institute of NRCD)

Kimihiko Nakamura (Research Institute of NRCD)

Tomomi Ishii (Research Institute of NRCD)

Hanae Koiso (National Institute for Japanese Language and Linguistics)

要旨

自閉スペクトラム症 (Autism Spectrum Disorder: ASD) は、対人・情緒的関係の障害等で特徴づけられる、発達障害の1つである。ASD者と定型発達 (Typically Developing: TD) 者の間では、両者のコミュニケーションスタイルの違いから語用論的な障害が生じ得るが、ASD者/TD者の言語運用の定量的分析を可能にするコーパスを構築することで、両者間のコミュニケーションを円滑にするための支援デバイス・手法の開発につなげることができる可能性がある。そこで、本研究では世界初となる日本語母語話者のASD者を対象とする映像・心理指標得点・心拍変動データ付き公開コーパス構築に取り組み、国立国語研究所からの公開を目指す。データ規模はASD者/TD者各6名を目標とし、2024年8月現在、ASD者6名・TD者5名の会話の映像・音声 (1人当たり約110分) を収録済みである。本稿ではコーパスの設計やデータ収集の概要を述べた上で、話者の意図・感情の伝達に関わる間投助詞・終助詞に着目し、それらの使用率の差を検証する。

1. はじめに

初期のコーパスは書籍や新聞などの文字化された言語から構築されたものが主流だが、近年では映像・音声 (小磯ほか, 2023) や心電図 (Ringeval et al., 2013)、眼球運動情報 (Yaneva et al., 2016; 浅原ほか, 2019) などマルチモーダルな情報を組み合わせたコーパスにより、分析の幅が広がっている。例えば、映像・音声付きの『日本語日常会話コーパス (Corpus of Everyday Japanese Conversation: CEJC)』 (小磯ほか, 2023) を活用して、ジェスチャーや身の回りの物などの非言語情報と言語を紐づけた分析が行われている (Daiju & Ono, 2018)。

こうしたマルチモーダルコーパスは、自閉スペクトラム症 (Autism Spectrum Disorder:

† suzuki-asumi@rehab.go.jp

ASD) の言語使用・理解の研究にも有用と考えられる。ASD は、対人的・情緒的関係の障害、強いこだわり等で特徴づけられる発達障害の 1 つであり、診断基準としては下記のようなものが挙げられる。

- A) 社会的コミュニケーションおよび相互関係における持続的障害。
- B) 限定された反復する様式の行動、興味、活動。
- C) 症状は発達早期の段階で必ず出現するが、後になって明らかになるものもある。
- D) 症状は社会や職業その他の重要な機能に重大な障害を引き起こしている。

(米国精神医学会, 2014)

ASD 者は、定型発達 (Typically Developing: TD) 者の話し手の意図を掴めない、質問に対して適切に応答できないなど、日常的なコミュニケーションに困難を抱えることがある (幕内編, 2023; 大井, 2006)。このような困難は、TD 者と ASD 者のコミュニケーションスタイルの違いによって生じる、語用論の問題に起因すると考えられる。適切な支援のためには、ASD 者と TD 者の言語運用の特徴の違いを解明することが重要であり、近年ではマルチモーダルコーパスを活用した研究が進んでいる。例えば、眼球運動情報付きコーパスの分析から、ASD 者は TD 者に比べて視線固定と読み戻りの回数が多く、長い単語の注視時間が長いことが報告されている (Yaneva et al., 2016)。また、ASD 者と TD 者の議論課題の映像・音声付きコーパスから、言語・音声・顔の特徴量を抽出し、議論中にコミュニケーション上の困難が発生する箇所を検出するモデルを構築する試み (小林ほか, 2024) も行われている。

このように、ASD 者の言語使用・理解の特徴を解明するためにコーパスの手法が有用な成果を挙げている一方で、多くの研究者が活用できる“公開”コーパスの整備は、世界的に見ても少ない。日本語以外の言語では、話し言葉コーパス集約プロジェクト TalkBank の一部である ASDBank¹から、英語 (Eigsti et al., 2007; ほか)、オランダ語 (Kuijper et al., 2015) を始めとする 6 言語 11 種が公開されている。しかし、日本語に目を向けてみると、ASD の診断補助ツールの映像記録の一部や、1 対 1 会話の書き起こしをコーパスとした研究 (崎下ほか, 2020; 加藤, 2018) が行われているものの、コーパスそのものは公開に至っていない。そこで、本研究では、多領域の研究者が活用できる研究のインフラストラクチャーとするべく、世界初となる日本語母語話者の ASD 者を対象とする動画付き公開コーパス構築に取り組む。本コーパスは、国立国語研究所から公開する計画である。

本稿では、まずコーパスの設計・データ収集についての概要を述べ (第 2 章)、2024 年 8 月 15 日現在の進捗状況を記す (第 3 章)。次に、話者の意図・感情の伝達に関わる語彙である間投助詞、終助詞に着目し、2024 年 8 月 15 日現在までに形態素解析が済んでいる ASD 者 6 名・TD 者 3 名のデータに基づき、その使用率を調査した結果を報告する (第 4 章)。ASD 者と TD 者、あるいは ASD 傾向の高い TD 者と低い TD 者では、終助詞使用・理解の特徴が異なる傾向にあることが報告されている (綿巻, 1997; Kiyama et al., 2018; 直江ほか, 2022; Naoe et al., 2024; 鈴木ほか, 2024; ほか)。したがって、特定の終助詞の使用率の高さ/低さは、ASD のスクリーニング・診断の際の指標となる可能性がある。先に挙げた先行研究は、ASD 児 1 名のケーススタディや実験環境での終助詞理解/産出の分析、TD 者の ASD 傾向と終助詞使用率の相関分析である。本研究では、自然会話に近い雑談も含む成人 ASD

¹ <https://asd.talkbank.org/>

者 6 名・TD 者 3 名のデータに基づき、自然産出の間投助詞・終助詞の使用率 (文末の終助詞のみの使用率/間投助詞・終助詞を合わせた使用率) の検証を試みる。

2. コーパスの設計・構築

2.1 本研究の対象とコーパスの設計方針

本研究では、ASD 者/TD 者それぞれ 6 名、合計 12 名の協力者を募る²。ASD 者に関しては、診断補助ツールである ADOS-2 (黒田・稲田, 2015; Lord et al., 2012) により重症度を確認した。設計方針としては以下の 4 点を挙げ、データ収集・コーパス整備を進める。

1. 1 問 1 答形式のタスク (P-F スタディ)、コミュニケーションの困りごとに関するインタビュー、自由会話の 3 種を収録する。
2. 自由会話では、ADOS-2 のような構造化された面接形式ではなく、なるべく自然会話に近い内容の会話を収録する。
3. 映像・音声・転記テキストに加え、ASD に関する心理指標得点、心拍変動データまで含めて収録・公開する。
4. データ収録・公開に伴う倫理的配慮などは CEJC に倣う。

2.2 データ収集

2024 年 8 月 15 日現在、ASD 者 6 名・TD 者 5 名を対象に、1 人当たり約 110 分の会話の映像・音声の収録を完了している。マルチモーダルな分析のために、ASD に関連する心理指標得点、会話中の心拍変動データも収集した。収集したデータの概要を表 1 に示す。

表 1 本研究で収集するデータ

会話の映像・音声	心理指標	生理指標	フェイスシート
<ul style="list-style-type: none">・ P-F スタディ (約 5-10 分)・ コミュニケーションの困りごとに関するインタビュー (約 40 分)・ 自由会話 (約 60 分) (1 人当たり合計約 110 分)	<ul style="list-style-type: none">・ 自閉症スペクトラム指数 (AQ)・ 共感指数 (EQ)・ システム化指数 (SQ)・ コナーズ成人 ADHD 評価スケール (CAARS)・ 対人反応性指標 (IRI)・ トロント・アレクシサイミア尺度 (TAS-20)・ AASP 青年・成人感覚プロファイル・ 首尾一貫感覚テスト・ うつ性自己評価尺度 (SDS)	<ul style="list-style-type: none">・ 心拍変動	<ul style="list-style-type: none">・ 年齢・ 居住地/出身地 (都道府県/国レベル)・ 性別・ 職業

2.2.1 会話収録

協力者に国立障害者リハビリテーションセンター研究所に来所してもらい、1 人当たり合計約 110 分の会話の映像・音声を収録した。会話は調査者 (鈴木あすみ) と協力者の 1 対 1³ で行った。なお、新型コロナウイルス感染症の状況を考慮し、調査者と協力者はそれぞれ別

² 2024 年 8 月 15 日現在、ASD 者 6 名・TD 者 5 名のデータ収集を完了している。

³ ASD 者の会話収録時は、国立障害者リハビリテーションセンター研究所 発達障害研究室の和田真 室長もオンライン会議システムに参加した (カメラ・マイクはミュート、画面上にも表示されないよう設定)。

室で、オンライン会議システム Webex を使用し、ノート PC の画面を介して対面で会話をした。第三者や著作物が映り込まないように、壁を背景にして収録した。

会話収録は、①P-F スタディ (5-10 分)、②コミュニケーションの困りごとに関するインタビュー (約 40 分)、③自由会話 (約 60 分) の 3 セッションから成る。全体的な構成として、最初は話しやすいように構造化の度合いの高い心理検査から始め、だんだんと構造化の度合いを落として自由会話に移行した。

まず、①P-F スタディ (秦ほか, 2020) という、ASD/TD 両方に適用可能な 1 問 1 答形式の構造化された心理検査を行った。P-F は Picture Frustration の略で、フラストレーション (欲求不満) のある状況に対する反応のタイプから、その性格傾向を把握する心理検査である。本心理検査において、ASD 者は未回答や場面認知の誤り、曖昧な表現などにより、評定不能な回答が多いことが報告されている (西藤ほか, 2021)。実施方法としては、2 人の人物が描かれた絵を見て、その人物の台詞を回答する。例えば、図 1 は車で事故を起こしてしまい、相手から責められているという状況で、責められている人物の台詞は空欄になっている。被験者は、「この状況で、絵の中の人物は何と言うか」を回答する。本来は記入式の課題であるが、今回の収録においては調査者が片方の人物の台詞を読み上げた後、参加者に口頭で回答してもらう方法を取った。



©株式会社三京房

図 1 P-F スタディ 成人用第 3 版 (画像引用:

<https://www.sankyobo.co.jp/apfnew.html>)

例えば、図 1 の場面であれば、調査者が「君が無理に追い越そうとしたのが悪いんだよ。」と言った後、協力者に最初に思いついた台詞を回答してもらった。

次に、②コミュニケーションの困りごとに関するインタビュー (以下、「困りごとインタビュー」) を行った。「会話がうまく続かない」「相手との意思疎通がうまくいかない」など、コミュニケーションで困った経験に関してインタビューした。「コミュニケーションで困った経験について、具体的なエピソードはあるか」を尋ねる質問から開始し、以降は協力者の回答に合わせて調査者が適宜質問をしながら会話を進めた。P-F スタディや ADOS-2 に比べると構造化の度合いは高くないが、会話のトピックが決まっており、調査者からの質問と協力者の回答という隣接ペアの頻度が高いという特徴のある談話である。

最後は、③自由会話を行った。これは、「世間話のような感覚で、気軽におしゃべりをする」という趣旨を協力者に伝えた上で行った雑談である。会話開始時は、最初の取り掛かりとして、調査者から協力者に対して「好きな食べ物は何ですか?」「趣味は何ですか?」といった質問をした。この最初の質問は、ASD 者/TD 者それぞれ 6 名でバランスがとれるように揃えてある。それ以外に話題の固定などはなかったため、3 つのセッションの中で最も構造化の度合いが低く、自然会話に近いデータである。

上記①~③の会話の様子は、ビデオカメラ (ZOOM Q2n-4K) と IC レコーダー (SONY ICD-SX1000) 各 2 台を用いて、協力者・調査者の動画・音声を録画・録音した。

2.2.2 心理指標得点

ASD/TD 両方に適用できるものを使用し、自閉症スペクトラム指数 (AQ) (若林ほか, 2004)、共感指数 (EQ) およびシステム化指数 (SQ) (Wakabayashi et al., 2006)、コナーズ成人 ADHD 評価スケール (CAARS) (Conners et al., 2012)、対人反応性指標 (IRI) (日道ほか, 2018)、トロント・アレキシサイミア尺度 (TAS-20) (小牧ほか, 2003)、AASP 青年・成人感覚プロファイル (Brown & Dunn, 2015)、首尾一貫感覚テスト (アントノフスキー, 2001)、うつ性自己評価尺度 (SDS) (Zung, 2011) の 9 種の得点を取得した。

2.2.3 心拍変動

会話中は、調査者・協力者ともに図 2 のような小型心拍計 (Polar Verity Sense) を装着した。これは、腕時計のようにして身に着けると、1 秒ごとに心拍数を計測・記録する機器である。計測した心拍変動は、会話の映像・音声と同期して公開する。



図 2 小型心拍計 (Polar Verity Sense)
(画像引用: <https://www.polar.com/ja/accessories/polar-verity-sense>)

2.3 公開方針

原則として、P-F スタディ (約 5-10 分)、困りごとインタビュー (約 40 分)、自由会話 (約 60 分) 全体を公開箇所を含める。ただし、本人が公開を希望しない箇所、倫理的観点などから公開に適さないと判断した箇所は除外する。研究責任者が映像・音声と書き起こしを全て確認した上で、協力者本人にヒアリングを行い、最終的な公開範囲を決定する。

3. 2024 年 8 月 15 日現在の進捗状況

3.1 協力者の属性

2024 年 8 月 15 日現在、ASD 者 6 名・TD 者 5 名の会話を収録済みである。各協力者の属性を表 2 に示す。年齢に関して、ASD 者では 26 歳～51 歳と幅広い年代に渡っているのに対し、TD 者では 5 名中 4 名が 20 代前半と、現状では年齢に偏りが出ている⁴。性別については、ASD 者は男女 3 名ずつ、TD 者は男性 3 名・女性 2 名で、概ねバランスが取れていた。現在の居住地は埼玉県が多く、その他も東京近郊の在住者であった。出身地に関しては、関東出身者が多くを占めており、その他は北海道、静岡県、沖縄県の出身者が 1 名ずつであった。職業については、年齢の偏りの影響もあり、ASD 者では社会人が多く、TD 者では学生が多かった。

⁴ 2024 年 8 月 15 日現在、40 代～50 代の TD 者の協力者 1 名を募集中である。

表 2 協力者の属性一覧

話者ID	年齢	性別	現在の居住地	出身地	職業
ASD_1	26	女性	埼玉県	埼玉県	大学院生
ASD_2	31	女性	東京都	北海道	自営業・自由業
ASD_3	32	男性	埼玉県	東京都	会社員 (契約社員・派遣社員含む)
ASD_4	51	女性	埼玉県	千葉県	専業主婦・専業主夫
ASD_5	34	男性	東京都	静岡県	会社員 (契約社員・派遣社員含む)
ASD_6	32	男性	神奈川県	東京都	パート・アルバイト
TD_1	22	女性	埼玉県	埼玉県	大学生 (短大含む)
TD_2	23	女性	埼玉県	沖縄県	専門学校生
TD_3	23	男性	東京都	東京都	大学院生
TD_4	24	男性	埼玉県	埼玉県	専門学校生
TD_5	31	男性	東京都	東京都	パート・アルバイト

3.1 転記と形態素解析

取得した会話の映像・音声は同期した後に、CEJC に準拠して転記テキストを作成した (白田ほか, 2020)。転記作業が完了した後は、現代話し言葉 UniDic (岡, 2019) を用いて形態素解析を行い、人手で修正する⁵。

3.2 データ規模 (暫定値)

2024年8月15日までに形態素解析が済んでいる ASD 者 6 名・TD 者 3 名分 (加えて、会話相手である調査者 1 名) のデータの暫定値を表 3 に示す。P-F スタディが約 1 時間分で 9,096 短単位、困りごとインタビューが約 6 時間弱で 65,939 短単位、自由会話が約 9 時間分で 103,397 短単位であった。

表 3 データ規模 (暫定値)

	P-Fスタディ	困りごとインタビュー	自由会話
時間数 (分)	61.51	353.01	541.71
会話数	9	9	9
延べ話者数	18	18	18
異なり話者数	10	10	10
延べ短単位数	9,096	65,939	103,397

セッションごとのデータ規模を表 4 に示す。表中の「発話時間」は収録時間ではなく、話者が言葉を発している時間の合計を意味する。なお、表 4 に示す「発話時間」と「延べ短単位数」は、調査者と協力者を合わせた値である。

⁵ 2024年8月15日現在、ASD 者 6 名・TD 者 3 名分の形態素解析を完了し、人手での修正を進めている最中である。また、TD 者 2 名分の転記テキストを作成中である。

表4 セッションごとのデータ規模 (暫定値)

	P-Fスタディ		困りごとインタビュー		自由会話	
	発話時間(分)	延べ短単位数	発話時間(分)	延べ短単位数	発話時間(分)	延べ短単位数
ASD_1	4.92	1,084	46.43	10,945	68.72	15,320
ASD_2	4.12	938	42.11	8,921	66.40	15,042
ASD_3	4.69	1,080	26.18	5,618	45.62	9,675
ASD_4	4.86	990	37.78	7,205	57.97	10,783
ASD_5	4.22	954	26.75	5,720	35.70	7,412
ASD_6	5.38	1,205	36.31	8,000	47.24	10,428
TD_1	4.05	928	32.60	7,309	71.53	15,920
TD_2	4.76	1,081	36.77	6,912	53.94	10,104
TD_3	3.70	836	29.31	5,309	44.20	8,713

セッション・話者ごとの短単位数、1秒当たりのモーラ数を表5～7に示す。

表5 P-Fスタディにおける話者ごとの短単位数、1秒当たりのモーラ数 (暫定値)

セッション	協力者			調査者		
	延べ短単位数	異なり短単位数	1秒当たりのモーラ数	延べ短単位数	異なり短単位数	1秒当たりのモーラ数
ASD_1	300	111	6.24	784	232	7.78
ASD_2	216	92	7.02	722	216	7.64
ASD_3	355	132	7.32	725	213	7.64
ASD_4	258	114	5.44	732	226	7.60
ASD_5	225	107	7.27	729	215	7.63
ASD_6	444	142	6.82	761	215	7.72
TD_1	221	79	7.30	707	207	7.59
TD_2	298	105	6.67	783	229	7.85
TD_3	124	62	6.62	712	215	7.73

表6 困りごとインタビューにおける話者ごとの短単位数、1秒当たりのモーラ数 (暫定値)

セッション	協力者			調査者		
	延べ短単位数	異なり短単位数	1秒当たりのモーラ数	延べ短単位数	異なり短単位数	1秒当たりのモーラ数
ASD_1	7,960	832	7.76	2,985	402	6.83
ASD_2	5,974	688	7.02	2,947	385	6.30
ASD_3	2,331	341	6.75	3,287	364	6.92
ASD_4	4,490	549	5.82	2,715	394	6.77
ASD_5	1,953	330	6.97	3,767	453	6.74
ASD_6	4,875	483	7.25	3,125	370	6.59
TD_1	3,853	428	7.33	3,456	466	6.93
TD_2	4,001	456	5.99	2,911	394	6.08
TD_3	2,137	383	5.10	3,172	408	6.75

表7 自由会話における話者ごとの短単位数、1秒当たりのモーラ数 (暫定値)

セッション	協力者			調査者		
	延べ短単位数	異なり短単位数	1秒当たりのモーラ数	延べ短単位数	異なり短単位数	1秒当たりのモーラ数
ASD_1	9,284	976	7.14	6,036	775	6.86
ASD_2	9,909	1,095	7.55	5,133	630	6.45
ASD_3	5,043	620	6.92	4,632	541	6.52
ASD_4	4,985	595	5.36	5,798	698	6.57
ASD_5	1,822	387	6.44	5,590	679	6.53
ASD_6	4,906	606	7.56	5,522	662	6.66
TD_1	6,934	806	7.52	8,986	987	6.60
TD_2	4,099	557	5.48	6,005	782	6.27
TD_3	2,457	482	5.48	6,256	807	6.63

例として、自由会話 (表7) における協力者の延べ短単位数を見てみると、最大は9,909短単位、最小は1,822短単位 (両者ともにASD者) と、発話量は個人差が大きかった。TD者の中で見ても、最大は6,934短単位、最小は2,457短単位と、話者による差が見られた。

4. 終助詞産出頻度の分析

次に、2024年8月15日時点で揃っているASD者6名・TD者3名分の会話の形態素解析結果に基づいて、ASD者とTD者の間投助詞および終助詞の使用率の差を探索的に検証した。文末の終助詞に限定した使用率 (4.1節、4.2節) と、間投助詞と終助詞を合わせた使用率 (4.3節~4.5節) について述べる。なお、本章では協力者の発話のみを分析対象とした。また、本稿で扱う形態素解析結果は自動解析による暫定的なものであり、2024年8月15日時点では、人手での修正を進めている段階である。

4.1 話者ごとの文末の終助詞使用率の算出

P-Fスタディ、困りごとインタビュー、自由会話それぞれにおける、個人の文末の終助詞使用率を算出した。まず、①述部の主要部が動詞/形容詞/名詞/形状詞である「文」相当の発話⁶を抽出した。述部の主要部が名詞/形状詞の場合は、「です」「ます」「だ」が後続する発話のみを対象とした。これを個人の「発話量」とする。②次に、①のうち終助詞で終わる発話を抽出し、各終助詞の粗頻度を集計した。抽出した終助詞の用例は目視で確認し、誤解析の除外・修正を行った。③最後に、終助詞の粗頻度を発話量で割って、話者ごとの文末の終助詞使用率を算出した。

$$(\text{話者ごとの文末の終助詞使用率}) = (\text{各終助詞の粗頻度}) / (\text{個人の発話量})$$

なお、「よね」「かな」のように、複数の終助詞が連続して用いられる場合は、単体の終助詞とは別々に扱った。例えば、「よね」は「よ」とも「ね」とも分けて集計し、「かな」は「か」とも「な」とも分けて集計した。

⁶ ここでの「文」相当の発話は、CEJCにおける「発話単位」に相当する。発話単位とは、「話し手と聞き手が行為や情報を交換する際の基本単位に相当し、統語的・談話的・相互行為的な一まとまり」(JDRI, 2017) を指す。ただし、本稿執筆時点でのデータは整備途中段階のものであるため、今後見直す可能性がある。

4.2 セッション・話者別の文末の終助詞使用率

協力者9名のP-Fスタディ、困りごとインタビュー、自由会話の合計の発話を見ると、述部の主要部が動詞／形容詞／名詞／形状詞（名詞／形状詞の場合は「です」「ます」「だ」が後続する場合のみ）である「文」相当の発話5,331例のうち、1,811例（33.97%）が終助詞で終わる発話であった。セッション・話者別の文末の終助詞使用率を図3～5に示す。いずれの図も、上側6名がASD者、下側3名がTD者の順で並んでいる。また、困りごとインタビューと自由会話において、粗頻度が5未満の終助詞は「その他」とした。

図3は、P-Fスタディにおける協力者9名の文末の終助詞使用率である。使用率の高さに関しては、ASD者とTD者で顕著な差は見られなかった。しかし、終助詞の異なり語数に着目してみると興味深い傾向が見て取れる。まず、TD者では「か」「ね」「よ」を中心に、いずれの話者も4～5種類を使用していた。これに対し、ASD者では6名中3名が6種類以上、多い話者では8種類の終助詞を用いて多様な応答を返していた。つまり、ASD者の中には、P-Fスタディに回答する際の終助詞の異なり語数がTD者より多い話者が存在する可能性が示唆された。また、ASD_1(26歳女性)、ASD_2(31歳女性)、ASD_3(32歳男性)が、いわゆる女性語である「わ」「わよ」を用いて回答していた。これらはTD者には用いられていなかった。ASD者はTD者に比べて、役割語を含め多様な終助詞を用いてP-Fスタディのタスクを遂行していたことが伺える。

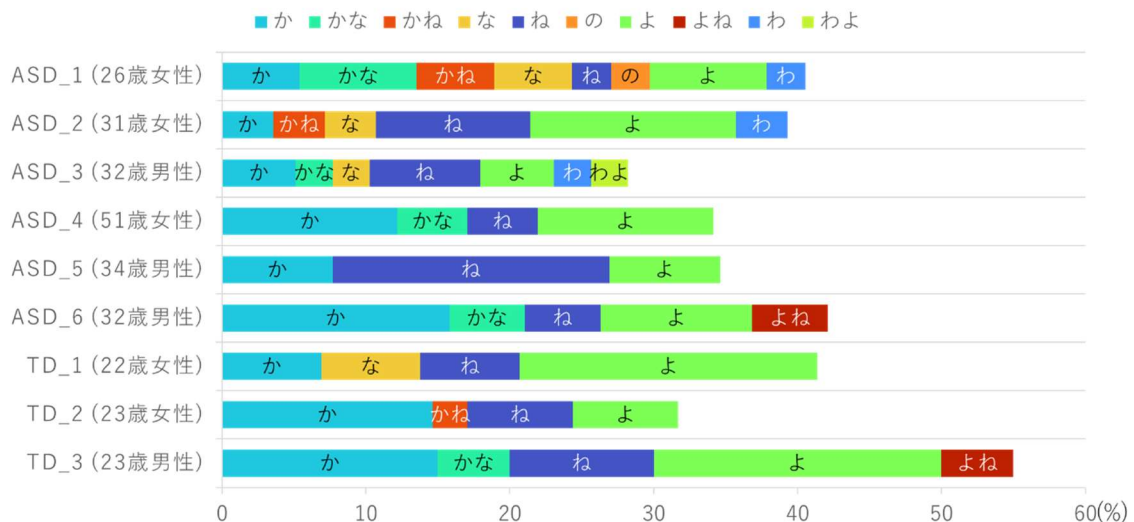


図3 P-Fスタディにおける話者別の文末の終助詞使用率（暫定値）

図4は、困りごとインタビューでの文末の終助詞使用率である。使用率の高さに関しては、ASD者／TD者ともに個人差が大きかった。また、異なり語数にも両群での顕著な差は見られなかった。

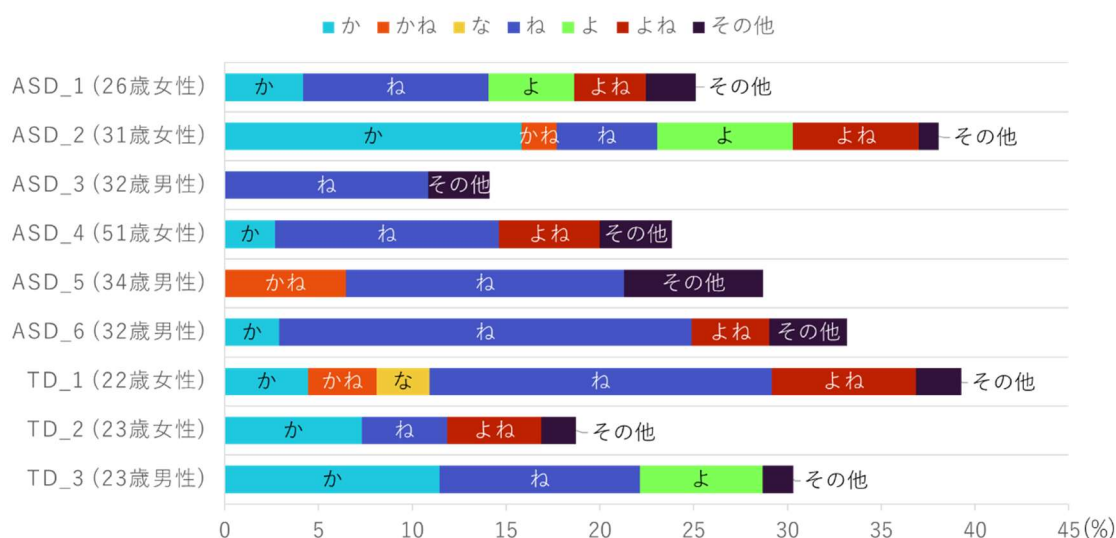


図4 困りごとインタビューにおける話者別の文末の終助詞使用率 (暫定値)

図5は、自由会話における文末の終助詞使用率である。困りごとインタビューと同様、終助詞の異なり語数には ASD 者/TD 者での顕著な差は見られなかった。使用率の高さについては、ASD 者6名のうち4名 (ASD_1~ASD_4) は相対的に使用率が低かったものの、残りの2名 (ASD_5、ASD_6) の使用率は TD 者と変わらない高さであった。

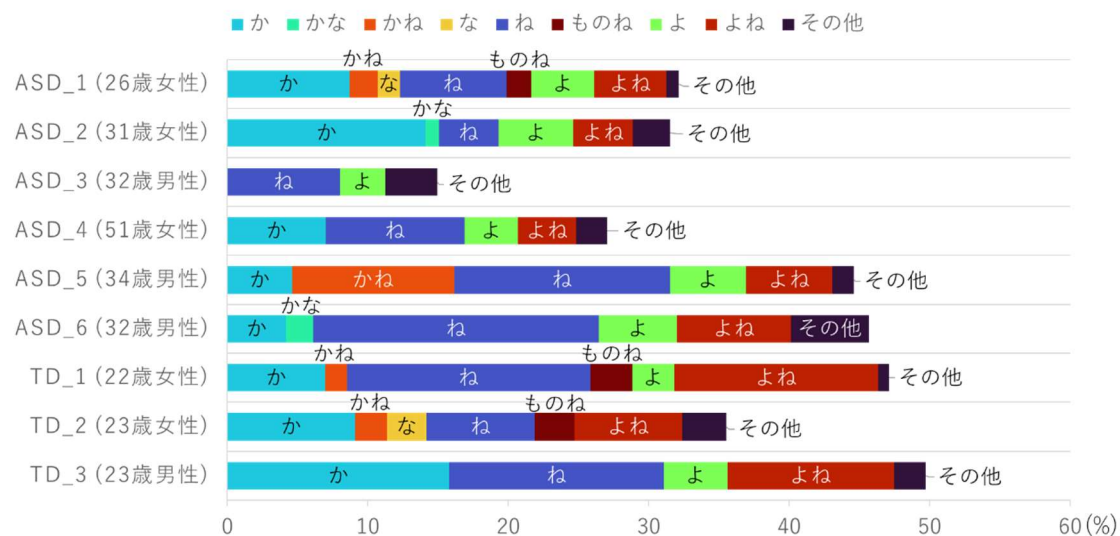


図5 自由会話における話者別の文末の終助詞使用率 (暫定値)

4.3 話者ごとの間投助詞・終助詞使用率の算出

次に、P-F スタディ、困りごとインタビュー、自由会話それぞれにおける、個人の間投助詞・終助詞を合わせた使用率を算出した。間投助詞は、終助詞よりも任意性・相互作用性が高い助詞である (大江, 2017)。そのため、間投助詞の使用傾向には、自閉傾向の高い話者の対人的発話の特徴がより顕著に現れ得る。また、終助詞と間投助詞は必ずしも明確に区別できるものではなく、一括りに扱われる場合もある (森田, 2007)。そこで、本研究でも間投助詞・終助詞を合わせた使用率と自閉傾向の関係を分析した。ただし、「ね」「さ」「な」「よ」などは終助詞としても間投助詞としても機能するのに対し、「か」「よね」「もの」「ものね」「よな」「じゃん」などは間投助詞としての機能をもたない。本研究では、どちらも分析対

象とした。

まず、品詞が「助詞-終助詞⁷」となる短単位を抽出し、個人の間投助詞・終助詞を合わせた粗頻度を語彙素ごとに求めた。抽出した間投助詞・終助詞の用例は目視で確認し、誤解析等の除外・修正を行った⁸。次に、間投助詞・終助詞の粗頻度を、表 5～7 に示す個人の延べ短単位数で割って、話者ごとの間投助詞・終助詞使用率を算出した。

$$\text{(話者ごとの間投助詞・終助詞使用率)} \\ = \text{(各間投助詞・終助詞の粗頻度)} \div \text{(個人の延べ短単位数)}$$

なお、この算出法では間投助詞・終助詞が付加されている自立語の種類は問わないこととした。つまり、4.1 節では述部の主要部が動詞／形容詞／名詞／形状詞 (名詞／形状詞は「です」「ます」「だ」が後続する場合のみ) である発話のみを対象としたのに対し、本節では副詞に間投助詞・終助詞が付加された発話 (例:「そうね。」) や、名詞／形状詞のうち「です」「ます」「だ」が後続しないものに間投助詞・終助詞が付加された発話 (例:「カフェの人はね。」) も分析対象に含まれている。

4.4 セッション・話者別の間投助詞・終助詞使用率

協力者 9 名の P-F スタディ、困りごとインタビュー、自由会話の合計の発話を見ると、短単位の総数 89,454 例のうち、3,513 例 (3.93%) が間投助詞・終助詞であった。セッション・話者別の間投助詞・終助詞使用率を図 6～8 に示す。いずれの図も、上側 6 名が ASD 者、下側 3 名が TD 者の順で並んでいる。また、困りごとインタビューと自由会話において、粗頻度が 10 未満の間投助詞・終助詞は「その他」とした。

図 6 は、P-F スタディにおける協力者 9 名の間投助詞・終助詞使用率である。使用率の高さに関して、ASD 者 6 名のうち 4 名 (ASD_3～ASD_6) の使用率は相対的に低かったが、残りの 2 名 (ASD_1、ASD_2) の使用率は TD 者とさほど変わらなかった。間投助詞・終助詞の異なり語数については、文末の終助詞のみを対象とした場合と同様、TD 者では 4～5 種類を使用しているのに対し、ASD 者では 6 名中 3 名が 6 種類以上、多い話者では 8 種類の間投助詞・終助詞を用いて多様な応答を返していた。4.3 節に述べたように、終助詞の中には間投助詞としても機能するもの (「ね」「さ」「な」「よ」等) と、間投助詞としての機能をもたないもの (「か」「よね」「もの」「ものね」「よな」「じゃん」等) が存在するが、間投助詞の機能しかもたない語彙は存在しない。したがって、異なり語数は文末の終助詞のみを対象とした場合 (図 3) とさほど変わらない結果となった。

⁷ 本データの形態素解析結果では、間投助詞と終助詞に品詞上の区別はなく、どちらも「助詞-終助詞」と解析されている。

⁸ 「ね」の用例について、SNS での高評価を意味する「いいね」は名詞化されているため除外した。

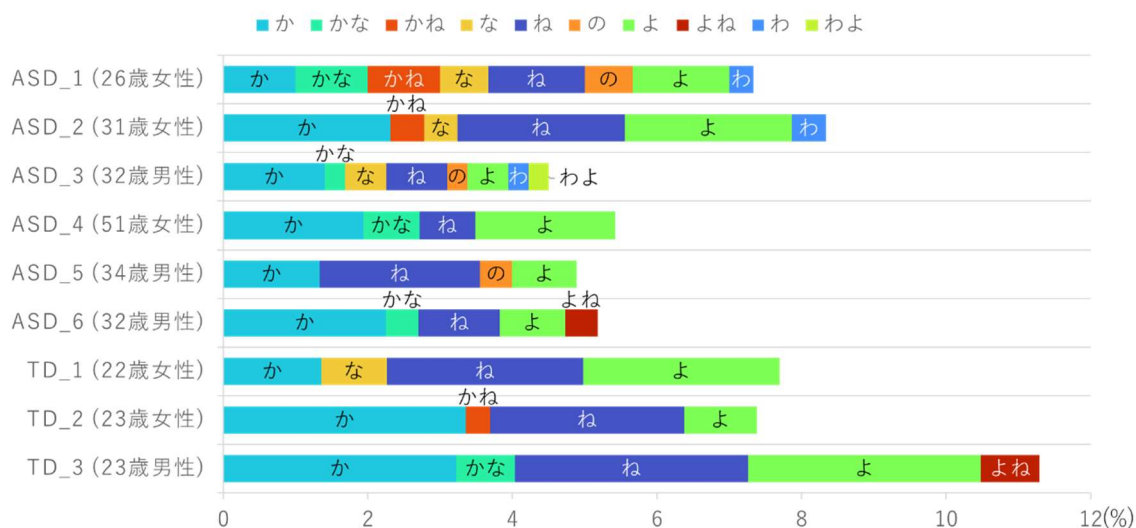


図6 P-F スタディにおける話者別の間投助詞・終助詞使用率 (暫定値)

図7は、困りごとインタビューでの間投助詞・終助詞使用率である。文末の終助詞のみの場合と同様、ASD 者／TD 者ともに間投助詞・終助詞使用率の高さは個人差が大きかった。また、「さ」⁹は終助詞としても間投助詞としても使用可能な助詞であるが、本コーパスのデータでは、困りごとインタビューと自由会話で、間投助詞としての使用のみが確認された。

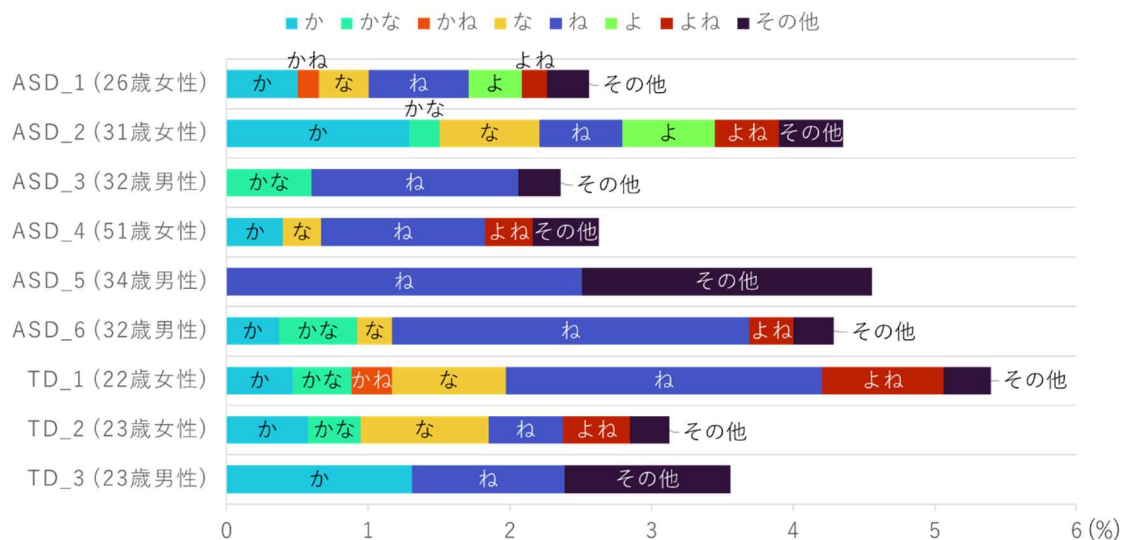


図7 困りごとインタビューにおける話者別の間投助詞・終助詞使用率 (暫定値)

図8は、自由会話における間投助詞・終助詞使用率である。文末の終助詞のみの場合と同じく、ASD 者6名のうち4名 (ASD_1～ASD_4) の間投助詞・終助詞使用率は相対的に低かったものの、残り2名 (ASD_5、ASD_6) の使用率はTD 者と同程度であった。

⁹ 困りごとインタビュー・自由会話ともに「さ」は粗頻度が低く、「その他」にまとめてある。

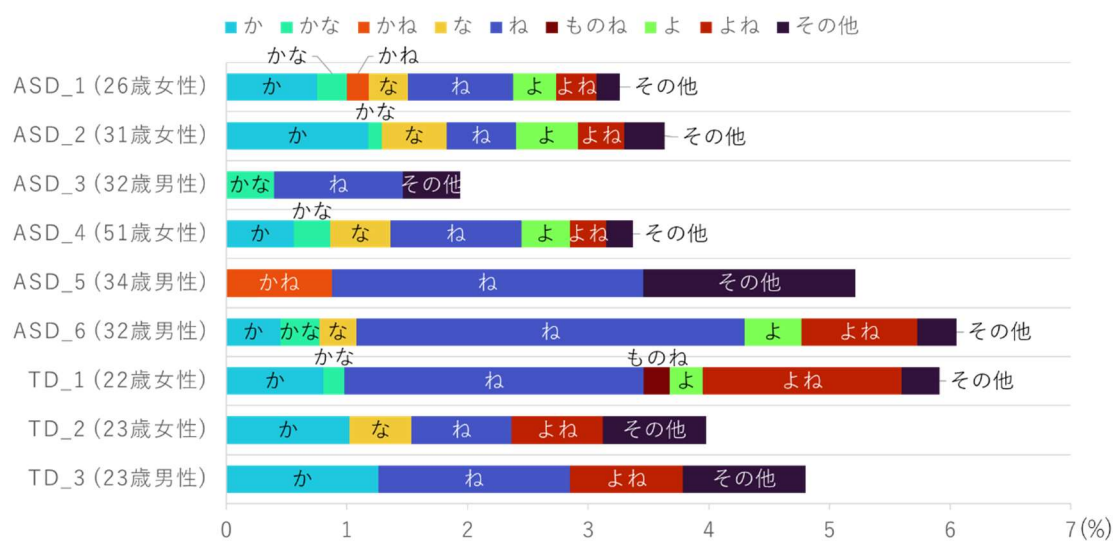


図8 自由会話における話者別の間投助詞・終助詞使用率 (暫定値)

4.5 間投助詞・終助詞「ね」の使用率

終助詞の中でも、特に「ね」の使用・理解の特徴と ASD/ASD 傾向に関係があることが、多くの先行研究で報告されている (佐竹・小林, 1989; 綿巻, 1997; 直江ほか, 2022; Naoe et al., 2024; ほか)。本コーパスのデータにおいても「ね」¹⁰は、どのセッション・話者においても用いられ、比較的使用率が高かった間投助詞・終助詞であった。そこで、セッションごとの個人の間投助詞・終助詞「ね」の使用率の差を比較した。4.3 節に述べた方法で算出した間投助詞・終助詞の使用率のうち「ね」のみを抜粋し、図 9~11 に示す。

図 9 は、P-F スタディにおける間投助詞・終助詞「ね」の使用率である。ASD_2 と ASD_5 の「ね」の使用率 (それぞれ 2.31%、2.22%) は高いように見えるが、これらの話者の「ね」の使用率は TD 者の中で最も低い TD_2 の「ね」の使用率 (2.68%) よりも低かった。全体的に、TD 者より ASD 者の方が、「ね」の使用率が低い傾向にあった。

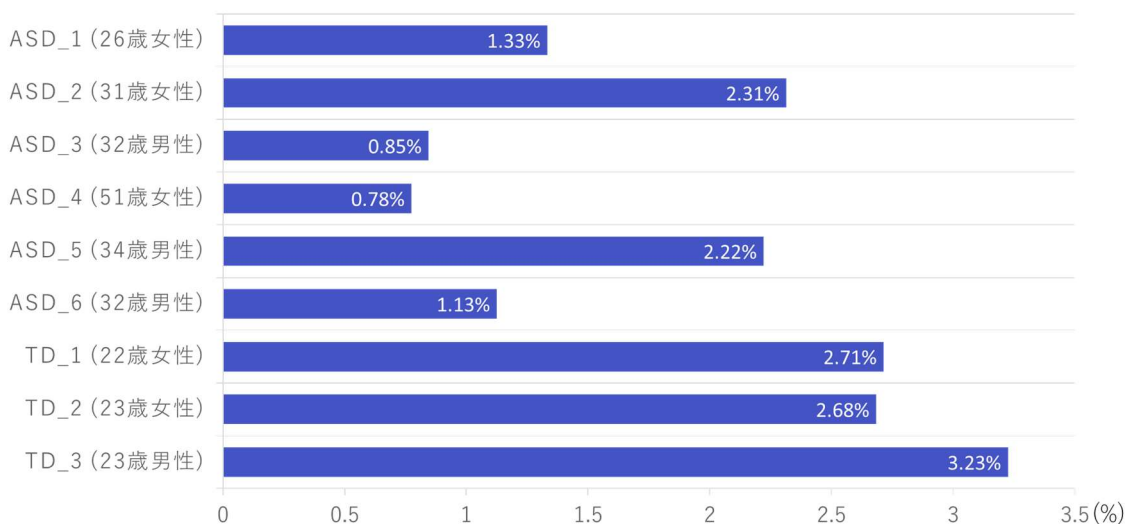


図9 P-F スタディにおける間投助詞・終助詞「ね」の使用率 (暫定値)

¹⁰ 「よね」「もんね」など「ね」で終わる複合終助詞は、「ね」には含めていない。

以下に、P-F スタディの質問文と協力者の回答の一例を示す。同じ設問 (1a, 2a) に対し、ASD 者では「ね」を使わずに応答していた (1b, 2b) のに対し、TD 者は「ね」のついた文で相手の発話を受ける (1c)、文末に「ね」を用いる (2c) といった発話で回答していた。

(図書館の司書と、本をたくさん借りようとしている人物の場面。)

- 1a. (設問の発話) この図書館では 1 回に 2 冊までしか貸し出せない規則です。
- b. (ASD 者の回答例) 分かりましたじゃ 2 さ 2 冊に絞ってきます。
- c. (TD 者の回答例) そうなんですね 分かりました 戻ってきます。

(女性 2 人が、別の女性から遊びに誘われなかった場面。)

- 2a. (設問の発話) どうして彼女は私達を誘ってくれなかったのでしょうかね。
- b. (ASD 者の回答例) うん まあ彼女の自由でしょう。
- c. (TD 者の回答例) うん 嫌な感じですよね。

「ね」がどのような自立語に付加されていたのかを見てみると、ASD 者/TD 者ともに (1c) のような「そう」の例が多かった (表 8, 9)。ASD 者では「ね」が付加されていた自立語の総数 24 例 (6 名の発話の合計) のうち 9 例 (37.50%)、TD 者では「ね」が付加されていた自立語の総数 18 例 (3 名の発話の合計) のうち 13 例 (72.22%) が「そう」であった。これに対し、「ね」が付加されていた自立語が「そう」以外の名詞・動詞・形容詞等であったのは、ASD 者で 15 例 (62.50%)、TD 者で 5 例 (27.78%) であった。つまり、TD 者では「そうですね」「そうなんです。」のように、相手の発話内容を容認する相槌のような発話が多かったのに対し、ASD 者ではこうした発話が相対的に少なかった。

表 8 「ね」が付加されていた自立語 (ASD 者 6 名の発話の暫定値)

「ね」が付加されていた自立語	粗頻度
そう	9
仕方ない	3
ある	1
いただく	1
しょうがない	1
する	1
なくなる	1
何	1
分かる	1
大変	1
待たせる	1
感じ	1
確認する	1
見づらい	1
合計	24

表 9 「ね」が付加されていた自立語 (TD 者 3 名の発話の暫定値)

「ね」が付加されていた自立語	粗頻度
そう	13
しょうがない	1
分かる	1
感じ	1
最悪	1
残念	1
合計	18

図 10 は困りごとインタビューにおける間投助詞・終助詞「ね」の使用率である。ASD 者 6 名中 4 名 (ASD_1~ASD_4) の「ね」の使用率は相対的に低いようにも見えるが、残りの 2 名 (ASD_5, ASD_6) は使用率が高かった。また、TD 者の中でも使用率の高い話者と低い話者が存在し、個人差が大きいことが伺える。

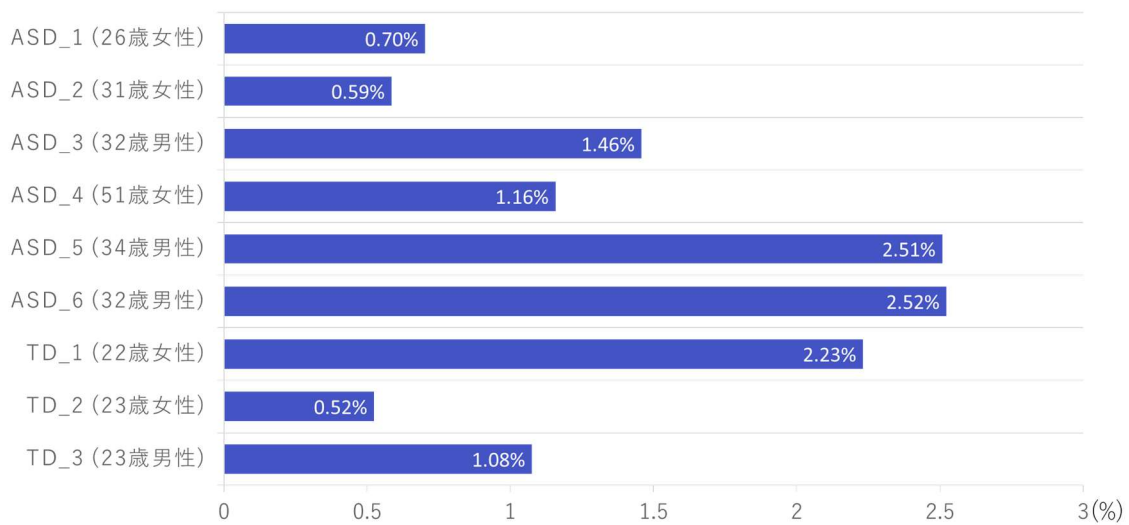


図 10 困りごとインタビューにおける間投助詞・終助詞「ね」の使用率 (暫定値)

図 11 は自由会話における間投助詞・終助詞「ね」の使用率である。困りごとインタビューの場合と同様、ASD 者 6 名中 4 名 (ASD_1~ASD_4) の「ね」の使用率は相対的に低いように見えるが、残りの 2 名 (ASD_5、ASD_6) は使用率が高かった。また、TD 者の中でも使用率の高い話者と低い話者が存在し、個人差が大きかった。

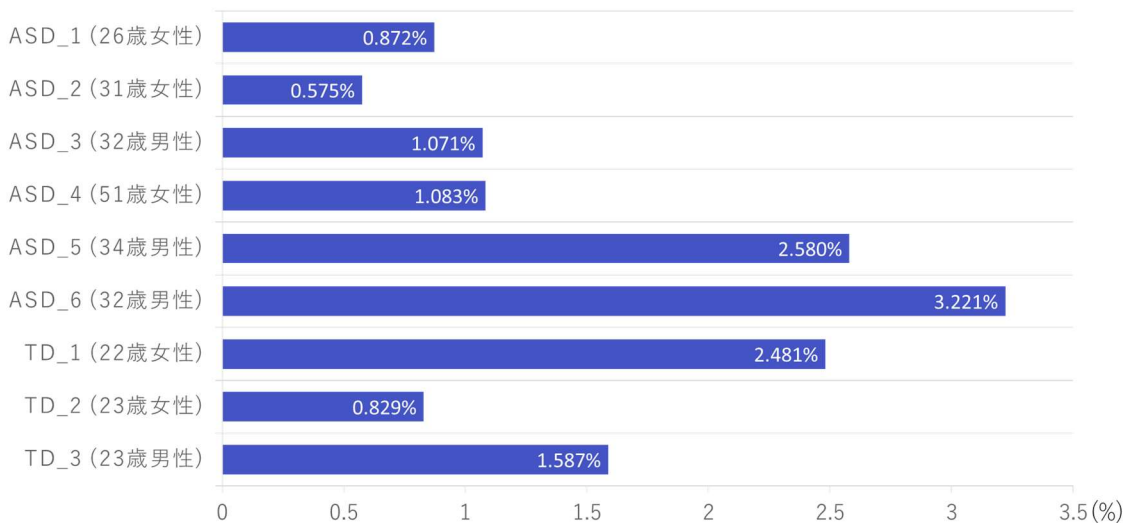


図 11 自由会話における間投助詞・終助詞「ね」の使用率 (暫定値)

4.6 考察

2024 年 8 月 15 日までに形態素解析が済んでいる ASD 者 6 名・TD 者 3 名について、文末の終助詞の使用率、間投助詞と終助詞を合わせた使用率、間投助詞・終助詞「ね」の使用率の差を比較した結果、以下のことが示唆された。

まず、P-F スタディにおいて、ASD 者は TD 者に比べ、文末の終助詞あるいは間投助詞・終助詞合計での異なり語数が多い傾向にあった。また、女性語である「わ」「わよ」を使って回答していたのは ASD 者のみであった。これらのことから、ASD 者は終助詞を役割語として、ステレオタイプの傾向にあるのではないかと考えられる。直江ほか (2022) では、女性では物事をシステム化して捉える傾向が高いほど「よ」や「ね」の使用率が高く、関東出身の男性でも物事をシステム化して捉える傾向が高いほど「ね」の使用率が高い傾向

にある(有意傾向)ことが示されている。つまり、ASD者は学習によって「こういう人物・場面ではこの終助詞を使う」というパターンを確立し、ルールベースの運用を行う傾向にあるのではないかと考えられる。また、ASD者の発話は「練習した独り言」(Ghaziuddin & Gerstein, 1996) のようだとされることもある。役割語的なルールベースの間投助詞・終助詞運用が、こうした印象に繋がっている可能性が示唆される。

また、P-Fスタディにおいて、ASD者の間投助詞・終助詞「ね」の使用率は、TD者より低い傾向にあった。また、「ね」の用法を見てみると、TD者では「そうですね。」のように相手の発話内容を容認する応答の割合が高かったのに対し、ASD者ではそうした発話の割合が相対的に低かった。ASD者は、TD者に比べあいづちの産出頻度が少ない(Wehrle et al., 2024)とされている。あいづちは相互作用性の高い言語表現であり、あいづちの少なさには、ASD者の「社会的コミュニケーションおよび相互関係における持続的障害」(米国精神医学会, 2014) という特徴が表れていると考えられる。本研究で見られたP-FスタディにおけるASD者の間投助詞・終助詞「ね」の使用率の低さ、特に「そうですね。」等の相槌的な発話の少なさは、このようなASD者とTD者のコミュニケーションスタイルの違いを反映しているのではないだろうか。

最後に、困りごとインタビューと自由会話においては、文末の終助詞、間投助詞・終助詞の使用率のばらつきが大きく、ASD者とTD者の間投助詞・終助詞「ね」使用率にも顕著な差は見られなかった。自然会話に近いタスク無しの談話における間投助詞や終助詞の使用率は、個人差が大きいことが伺える。困りごとインタビューと自由会話では、P-Fスタディに比べて会話が長くなり、終助詞使用のバリエーションが多くなることで、ASD者とTD者の違いが見づらくなっている可能性が考えられる。また、ASD者が一様に終助詞をあまり使わない傾向にあるのではなく、ASD者の中に終助詞使用率の低い一群が存在する、という可能性もある。この点に関しては、今後データ規模を拡張して検討したい。

5. おわりに

本稿では、現在構築中の日本語母語話者のASD者を対象とする映像・心理指標得点・心拍変動データ付き公開コーパスについて、設計とデータ収集の概要を述べた。次に、話者の意図・感情の伝達に関わる間投助詞・終助詞に着目し、ASD者6名・TD者3名分の転記テキストの形態素解析結果に基づき、文末の終助詞の使用率、間投助詞・終助詞を合わせた使用率の差を検証した。今後の課題として、2点を挙げたい。

まず、本研究では主に間投助詞や終助詞の単純な頻度に基づいた分析を行っており、用法の考慮が不十分な点である。例えば、「ね」の粗頻度が高い話者の中には、様々な述語の文に柔軟に「ね」を付加している話者もいれば、相槌的な「そうですね。」のように固定的な表現ばかりを多用している話者もいると考えられる。今後は、終助詞がどのような述部に付加されているのか、終助詞の付加された文がどのような言語行為を担っているのかといった観点から、より詳細な分析を行う計画である。

次に、データ数の問題である。まず、本稿執筆時点ではTD者のデータが3名分しかないため、TD者のデータが整備でき次第、ASD者とTD者の特徴の差異を改めて検討したい。また、ASDにはサブタイプが存在すると言われており、ASD者の中でも、間投助詞・終助詞使用の特徴には個人差が大きく影響している可能性がある。本研究で対象としたASD者は6名だが、より一般的な示唆を得るためには、さらに大きなサンプル数に基づく調査が必要である。今後もデータ数を増やし、コーパスを拡張していく計画である。

謝辞

本研究は次の研究助成を得て行われた。国立国語研究所共同研究プロジェクト「多世代会話コーパスに基づく話し言葉の総合的研究」(小磯花絵)、人間文化研究機構 共創先導プロ

ジェクト (共創促進研究)「コミュニケーション共生科学の創成」(菊澤律子)、JSPS 科研費研究活動スタート支援 21K20034「公開自閉スペクトラム症者会話コーパス構築の予備研究」(鈴木あすみ)。

文 献

- 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』設計と特徴」『国立国語研究所論集』 24, pp.153-168.
- Fabien Ringeval, Andreas Sonderegger, Juergen Sauer, and Denis Lalanne (2013). “Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions” *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp.1-8.
- Victoria Yaneva, Irina Temnikova, and Ruslan Mitkov (2016). “A Corpus of Text Data and Gaze Fixations from Autistic and Non-Autistic Adults” *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp.480-487.
- 浅原正幸・小野創・宮本エジソン正 (2019). 「BCCWJ-EyeTrack——『現代日本語書き言葉均衡コーパス』に対する読み時間付与とその分析——」『言語研究』 156, pp.67-96.
- Saori Daiju, and Tsuyoshi Ono (2018). “Studying Japanese distal demonstrative ‘are’ using video corpus” *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pp.47-50.
- 米国精神医学会 (原著) 高橋三郎・大野裕 (監訳) 染矢俊幸・神庭重信・尾崎紀夫・三村将・村井俊哉 (訳) (2014). 『DSM-5 精神疾患の診断・統計マニュアル』医学書院.
- 幕内充 (編) (2023). 『自閉スペクトラム症と言語』ひつじ書房.
- 大井学 (2006). 「高機能広汎性発達障害にともなう語用障害：特徴, 背景, 支援」『コミュニケーション障害学』 23:2, pp.87-104.
- 小林千紗・保科衣吹・坂戸達陽・二瓶芙巳雄・石井亮・深山篤・辻井正次・Stefanov Kalin・中野有紀子 (2024). 「グループ会話における成人 ASD のコミュニケーション特性」『分析人工知能学会全国大会論文集』 38, 3R1-OS-13b-03.
- Inge-Marie Eigsti, Loisa Bennetto, and Mamta B. Dadlani (2007). “Beyond pragmatics: Morphosyntactic development in autism” *Journal of Autism and Developmental Disorders* 37, pp.1007-1023.
- Sanne J. M. Kuijper, Catharina A. Hartman, and Petra Hendriks (2015). “Who is he? Children with ASD and ADHD take the listener into account in their production of ambiguous pronouns” *PLoS ONE* 10:7, pp.1-18.
- 崎下雅仁・小川ちひろ・土屋賢治・岩渕俊樹・岸本泰士郎・狩野芳伸 (2020). 「発話特徴を用いた自閉スペクトラム症の重症度推測システム」『人工知能学会論文誌』 35:3, B-J45_1-11.
- 加藤澄 (2018). 「自閉症スペクトラム障害者の発話における交渉詞「ね」と「よ」の使用から検証する対人観」『龍谷大学国際社会文化研究所紀要』 20, pp.85-101.
- 綿巻徹 (1997). 「自閉症児における共感獲得表現助詞「ね」の使用の欠如:事例研究」『発達障害研究』 19:2, pp.48-59.
- Sachiko Kiyama, Rinus G. Verdonchot, Kexin Xiong, and Katsuo Tamaoka (2018). “Individual

mentalizing ability boosts flexibility toward a linguistic marker of social distance: An ERP investigation” *Journal of Neurolinguistics* 47, pp.1-15.

直江大河・南部智史・鈴木あすみ・小磯花絵・幕内充 (2022). 「日本語母語話者の日常会話における終助詞「よ」「ね」の使用と自閉傾向の関係—日本語日常会話コーパスを用いた検討—」『社会言語科学会第46回大会 発表論文集』 pp. 102-105.

Taiga Naoe, Tsukasa Okimura, Toshiki Iwabuchi, Sachiko Kiyama, and Michiru Makuuchi (2024). “Pragmatic atypicality of individuals with Autism Spectrum Disorder: Preliminary data of sentence-final particles in Japanese” In Masatoshi Koizumi (ed.) *Issues in Japanese Psycholinguistics from Comparative Perspectives. Volume 2 Interaction Between Linguistic and Nonlinguistic Factors*. De Gruyter Mouton, pp.183-200.

鈴木あすみ・幕内充・小磯花絵・中村仁洋 (2024). 「『日本語日常会話コーパス』に基づく終助詞産出率の個人差と自閉傾向の相関分析」『社会言語科学会第48回研究大会 発表論文集』 pp.127-130.

黒田美保・稲田尚子 (監修・監訳) (2015). 『ADOS-2 日本語版』金子書房.

Catherine Lord, Michael Rutter, Pamela C DiLavore, Susan Risi, Katherine Gotham, Somer L Bishop, Rhiannon J Luyster, and Whitney Guthrie (2012). *Autism Diagnostic Observation Schedule, Second Edition*. Western Psychological Services.

秦一士・安井知己・青木佐奈枝・木村一朗・笹川宏樹・鈴木常元・西尾博・沼初枝・藤田主一 (2020). 『P-F スタディ 絵画欲求不満テスト 成人用第Ⅲ版』三京房.

西藤奈菜子・川端康雄・若林暁子・吉川真衣・金沢徹文・寺嶋繁典・米田博 (2021). 「P-F スタディを用いた診断閾下の自閉スペクトラム症を有する青年・成人のアセスメント」『心身医学』 61:1, pp.64-74.

若林明雄・東條吉邦・Simon Baron-Cohen・Sally Wheelwright (2004). 「自閉症スペクトラム指数 (AQ) 日本語版の標準化 高機能臨床群と健常成人による検討」『心理学研究』 75:1, pp.78-84.

Akio Wakabayashi, Simon Baron-Cohen, Sally Wheelwright, Nigel Goldenfeld, Joe Delaney, Debra Fine, Richard Smith, and Leonora Weil (2006). “Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short)” *Personality and Individual Differences* 41:5, pp.929-940.

C. Keith Conners, Drew Erhardt, and Elizabeth Sparrow (原著) 中村和彦 (監修) 染木史緒・大西将史 (監訳) (2012). 『CAARS 日本語版 Conners' Adult ADHD Rating Scales』金子書房.
日道俊之・小山内秀和・後藤崇志・藤田弥世・河村悠太・Mark H. Davis・野村理朗 (2018). 「日本語版対人反応性指標の作成」『心理学研究』 88:1, pp.61-71.

小牧元・前田基成・有村達之・中田光紀・篠田晴男・緒方一子・志村翠・川村則行・久保千春 (2003). 「日本語版 The 20-item Toronto Alexithymia Scale (TAS-20)の信頼性, 因子的妥当性の検討」『心身医学』 43:12, pp.839-846.

Catana E. Brown, and Winnie Dunn (原著) 辻井正次 (監修) 萩原拓・岩永竜一郎・伊藤大幸・谷伊織 (日本語版作成) (2015). 『AASP 青年・成人感覚プロファイル』日本文化科学社.
アーロン・アントノフスキー (原著) 山崎喜比古・吉井清子 (監訳) (2001). 『健康の謎を解く—ストレス対処と健康保持のメカニズム』有信堂高文社.

William W. K. Zung (原著) 福田一彦・小林重雄 (日本語版作成) (2011). 『SDS うつ性自己評価尺度』三京房.

- 白田泰如・川端良子・西川賢哉・石本祐一・小磯花絵 (2018). 「『日本語日常会話コーパス』における転記の基準と作成手法」『国立国語研究所論集』 15, pp.177-193.
- 岡照晃 (2019). 「言語研究のための電子化辞書」『コーパスと辞書 講座 日本語コーパス 7』 pp.1-28, 朝倉書店.
- Japanese Discourse Research Initiative (JDRI) (2017). 『発話単位ラベリングマニュアル version 2.1.』 <https://www.jdri.org/resources/manuals/uu-doc-2.1.pdf> (2024 年 8 月 15 日確認).
- 大江元貴 (2017). 「間投助詞の位置づけの再検討：終助詞との比較を通して」『語用論研究』 19, pp.90-99,
- 森田笑 (2007). 「終助詞・間投助詞の区別は必要か—「ね」や「さ」の会話における機能」『言語』 36:3, pp.44-52,
- 佐竹真次・小林重雄 (1989). 「自閉症児における語用論的伝達機能の発達に関する研究」『特殊教育学研究』 26:4, pp.1-9.
- Mohammad Ghaziuddin, and Leonore Gerstein (1996). “Pedantic speaking style differentiates Asperger syndrome from high-functioning autism” *Journal of Autism and Developmental Disorders* 26, pp.585-595.