



## プレス発表資料 Press Release

2024年11月20日

報道機関 各位

学校法人青森田中学園 青森中央学院大学  
国立大学法人 弘前大学

### インタビュー型自然言語処理技術を用いた 青年期・成人期 ASD 診断の新たな可能性を実証

#### 【研究のハイライト】

1. 青森中央学院大学の加藤澄教授（本学教員）が、弘前大学大学院保健学研究科の斉藤まなぶ教授、および医学研究科の中村和彦教授の協力を得て、自閉スペクトラム症（ASD）の診断に自然言語処理技術を活用する新たな手法を発表しました。
2. インタビュー形式で得られた言語データを解析し、ASD 固有の語彙・文法パターンを抽出することで、ASD と非 ASD を識別する可能性を示唆しました。
3. 本研究は、米国の権威ある学術誌\*\*「PLOS ONE」\*\*に2024年9月に掲載され、自然言語処理技術を応用した語彙-文法資源<sup>(1)</sup>の選択パターンに基づく診断ツールの有用性に関するはじめての発表です。
4. ASD 診断支援ツールの課題（特に成人における）を補完する新たなアプローチとして、インタビュー形式で得られる双方向的な言語データが有用であることが示されました。
5. 機械学習モデル（ロジスティック回帰およびディープニューラルネットワーク<sup>(2)</sup>）を活用し、語彙・文法資源の選択に基づいた診断モデルを開発しました。さらなる研究により、診断精度の向上が期待されます。



## プレス発表資料 Press Release

### 【背景】

自閉スペクトラム症 (ASD) は、神経発達症の一つであり、特に**社会的コミュニケーション**や**相互作用**における**困難さ**が主症状として挙げられます。その中でも、ASD の特性が**語用論的障害**として現れることが多く報告されています。

語用論的障害とは、以下のようなコミュニケーションの課題を指します：

- **文脈に応じた言語の適応が困難**: 会話の相手や状況に応じた適切な言語表現ができない。
- **比喩や皮肉、慣用表現の解釈が困難**: 字義通りでない言葉の意味を理解する能力が不足。
- **対人言語のニュアンス理解の難しさ**: 話し手の意図や感情を汲み取ることが難しい。

これらの特徴は、**社会的文脈に基づく語彙や文法の選択**という点で ASD の言語行動に独特のパターンを生み出す可能性があります。語用論的障害に起因する語彙-文法資源の選択の偏りは、ASD の**神経認知特性**が言語生成にどのように**反映されるか**を明らかにする手がかりとなります。

### ASD 診断の課題と研究の関連性

従来の ASD 診断ツールである **ADOS-2** や **ADI-R** は、診断妥当性が高いものの、成人期の診断では以下の課題が顕著です：

- **鑑別診断の難しさ**: ASD と注意欠如多動症 (ADHD) や気分障害などの他疾患との区別が不明瞭になること。
- **症状のカモフラージュ**: 成人が社会的に適応するために**カモフラージュ行動**や**補償戦略**を用いることで、症状が見えにくくなること。
- **診断情報の不足**: 成人では**発達報告**が不十分である場合や、本人の**自己洞察の信頼性**が低くなる傾向がある。

本研究では、ASD に関連する語用論的特性に限定するのではなく、語用論的障害に関連する要素を含めた ASD 特有の語彙-文法資源の選択パターンを自然言



## プレス発表資料 Press Release

語処理技術を用いて分析し、新たなアプローチを提案します。これにより、語用論的障害に直接関連する特徴だけでなく、ASD 特有の他の言語的特性も含めた ASD と非 ASD の識別可能性を探求しています。

### 【研究の目的】

本研究の目的は、青年期および成人期の ASD 診断を支援するための新たなツールとして、自然言語処理技術を用いた言語データの解析を通じて、ASD と非 ASD を識別する診断アプローチを開発することです。

ASD の診断支援においては、従来の診断ツールが抱える課題を補完するアプローチが求められています。本研究は、ASD 者がインタビュー形式で使用する語彙や文法構造の選択に着目し、ASD の特徴的な言語パターンを機械学習によって特定し、ASD と非 ASD を判別するモデルを構築することを目指しています。

特に、以下の要素に重点を置いています：

#### 1. 語用論的障害に基づく言語パターンの抽出:

ASD に見られる語用論的障害が反映された特有の語彙・文法資源のパターンを自然言語処理技術を用いて分析し、ASD 診断の補助的な指標として活用する。

#### 2. 機械学習アルゴリズムの開発:

ASD と非 ASD を区別するために、語彙-文法資源の選択パターンを分析するモデルを構築します。本研究では、\*\*線形モデル（ロジスティック回帰）とディープニューラルネットワーク（DNN）\*\*の両方を検討し、モデルの診断精度と解釈性のバランスを評価します。

#### 3. 双方向的なインタビュー形式<sup>(3)</sup>の有効性検証:

ASD 者の神経認知特性がより反映されやすいインタビュー形式を通じて、ASD 特有の言語選択パターンが明確に現れる可能性を検証します。

この研究は、語用論的障害に基づく ASD の言語パターンを補助診断ツールとして応用し、成人期の ASD 診断において多角的で効果的な支援を提供することを目指しています。さらに、このアプローチが成人期 ASD 診断の精度向上に貢献することが期待されています。



## プレス発表資料 Press Release

### 【方法】

本研究では、自然言語処理技術と機械学習を用いて、ASD と非 ASD を区別するモデルを構築しました。そのために、以下のデータベースおよびモデルを使用しました。

### データベース

#### 1. 対象:

本研究では、\*\*14 歳以上の青年および成人の ASD 者 (N=64) と非 ASD 者 (N=71) \*\*を対象としました。対象者は以下の基準で選定されました：

- 言語獲得の臨界期を過ぎた 14 歳以上であること。
- ASD の診断は、DSM-5 基準に基づき、ADOS-2 および知能検査などの精密検査を組み合わせ、専門医が実施。
- ASD 者と非 ASD 者の年齢層の平均はそれぞれ 18 歳 (SD=3.48) および 19 歳 (SD=2.77) 。

#### 2. 使用コーパス:<sup>(4)</sup>

本研究では、加藤ら(筆頭著者)が 2022 年に開発した、日本初の ASD および定型発達者を対象とする語彙・文法資源の注釈付きコーパスを使用しました。このコーパスは、日本語話者を対象としたもので、語用論および文法的要素の選択パターンを解析するために設計されています。海外で開発された ASD 関連のコーパスは生テキスト (未注釈のデータ) にとどまるのに対し、加藤らのコーパスは、**選択体系機能言語学 (SFL) の理論枠組みに基づいた注釈スキーム<sup>(5)</sup>**を採用しており、ASD 者特有の言語行動を多層的に分析できる点で独自性があります。

### モデル構築

本研究では、以下の 3 種類の機械学習モデルを構築し、比較検討しました：



## プレス発表資料 Press Release

1. **タグ線形モデル:**  
語彙・文法タグの情報のみを入力データとして使用し、線形モデル（ロジスティック回帰）に基づいて構築されたモデル。
2. **タグ DNN モデル:**  
タグ情報のみに依存するディープニューラルネットワーク（DNN）を用いたモデル。
3. **テキスト+タグ DNN モデル:**  
タグ情報とテキスト情報（インタビューの実際の発話内容）を組み込んだ DNN モデル。

これらのモデルは、語彙・文法資源の選択パターンを解析し、ASD 特有の言語行動を識別するための診断精度を比較する目的で設計されました。

### 【結果】

本研究の結果、以下の重要な知見が得られました：

1. **テキスト+タグ DNN モデルが最も優れたパフォーマンスを示しました:**  
タグ情報に加えて、テキスト情報を組み込んだ DNN モデルは、ASD と非 ASD を識別する際に他のモデルよりも高い診断精度を示しました。
2. **インタビュー形式のデータは、ASD 特有の語彙・文法資源の選択パターンをより顕著に反映しました(インタビュー形式と絵本の物語形式という 2 種類のテキストを分析)。**

**インタビュー形式:** 双方向的で社会的な性質を持つインタビュー形式のデータでは、ASD 特有の語彙・文法資源の選択パターンがより顕著に観察されました。これは、社会的相互作用が ASD 特性を引き出しやすいことを示しています。

**絵本の物語形式<sup>6)</sup>:** 一方、物語を語る形式（絵本ナラティブ）では、モノログ性が強く、ASD 特有の語彙・文法パターンが明確に現れにくい傾向が見られました。この形式では、社会的要素が限定されるため、診断における有用性が相対的に低いと考えられます。



## プレス発表資料 Press Release

### 【今後の課題】

本研究は重要な知見を得たものの、さらなる研究が必要とされる課題も明らかになりました：

1. サンプルサイズの拡大:

本研究は比較的小規模なサンプルを基に行われたため、大規模なデータによる検証が必要です。

2. 診断精度の向上:

現在のモデルの感度と特異度は、それぞれ73%と87%です。さらに精度を向上させるためには、注釈スキームの拡張が重要です。

### 【臨床応用への期待】

本研究は、従来のASD診断ツール（ADOS-2・ADI-R等）と統合することで、診断の精度向上に寄与する可能性を示唆しています。インタビュー形式を活用した自然言語処理技術は、特に成人期のASD診断において、既存の課題を補完する強力なツールとして期待できるでしょう。今後の研究により、より精緻な診断モデルが開発され、臨床現場での実用化が進むことが期待されます。

### 【研究プロジェクトについて】

本研究は、日本学術振興会（JSPS）科学研究費補助金JP26284060（代表：加藤澄）およびJP26590161（代表：加藤澄）の助成により行われました。

### 【論文情報】

著者：Sumi Kato, Kazuaki Hanawa, Manabu Saito, Kazuhiko Nakamura

論文名：Creating a diagnostic assessment model for autism spectrum disorder by differentiating lexicogrammatical choices through machine learning



## プレス発表資料 Press Release

雑誌名：PLOS ONE

文献 URL：<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0311209>

### 【用語解説】

- (1) 語彙-文法資源：選択体系機能文法（SFL）の体系による語彙と文法資源。一般的な文法とは異なる。
- (2) ディープニューラルネットワーク（DNN：deep neural network）
- (3) インタビュー形式：参加者が抱える困難や責任感、社会的な状況の捉え方、人間関係（友情、結婚、家族関係など）についての理解を評価することを目的としています。また、架空の世界の創造や自己を客観的に表現すること、自身の感情について述べる質問も含まれています。プロトコルに従い、面接者は会話のような口調で進行し、単なる質問と回答の形式ではなく、参加者の発言に対して興味を示しながらコメントを加えることで、対話を深めるよう努めます。
- (4) コーパス：言語データの集積データベース
- (5) 注釈スキーム：アノテーション（注釈）とは、あるデータに対して関連する情報（メタデータ）を注釈として付与すること。注釈スキームとは、アノテーション体系のこと。
- (6) 絵本の物語形式：絵だけで構成された絵本を使い、参加者がその内容をどれだけ再現できるかを評価します。この際、登場人物の感情や気持ちについても自発的に説明することが求められます。物語再現は独白形式ですが、聞き手の立場を考慮しながら、物語を分かりやすく、かつ興味深く伝えることが求められます。

【情報解禁日時】なし

### 【問い合わせ先】

青森中央学院大学経営法学部 教授 加藤澄

TEL：017-728-0131 Email：[kato@aomoricgu.ac.jp](mailto:kato@aomoricgu.ac.jp)

青森中央学院大学学園広報・高大連携課 担当 佐川康希

TEL：017-728-0131 Email：[acgukoho@aomoricgu.ac.jp](mailto:acgukoho@aomoricgu.ac.jp)