

学位請求論文の内容の要旨

論文提出者氏名	神経精神医学研究分野 氏名 加藤澄
(論文題目) Toward mapping pragmatic impairment of Autism Spectrum Disorder individuals through the development of a corpus of spoken Japanese (日本語話し言葉のコーパス開発を通じた自閉症スペクトラム障害の語用論的障害のマッピングに向けて)	
(内容の要旨) <div style="text-align: center;">要旨</div> <p>自閉症スペクトラム障害(autism spectrum disorder:ASD)の中心となる症状は、社会的コミュニケーションの困難であり、この障害は、一般に語用論的障害(pragmatic impairment: PI)とみなされている。PIは、社会的状況の中で、言語を適切に使用することに困難を持つことである。先行研究は、PIが神経学的、認知的、言語的、そして知覚運動性における機能不全といった複雑に絡み合った要因から引き起こされることを報告してきている。よって、PIの全容を明らかにするには、これらの要因を取り入れた複合的見地からのアプローチが必要である。</p> <p>この目的を達成するためには、まず、PIそのものの包括的マッピングが必要不可欠である。マッピングとは、ASDの話し言葉の中で、言語資源使用上のどこに、そして、どのような言語表現にPIが起こる傾向があるのかを系統的に明らかにすることである。これまでのところ、部分的なPIの言語現象の研究は行われてきてはいるものの、包括的PIのマッピングは行われていない。PIの包括的マッピングを必要とする臨床上的理由の一つに、ASDの診断補助ツールとして、言語面からの観察の可能性を想定できることがあげられる。現在、Autism Diagnostic Observation Schedule (ADOS-2)がASD診断の補助ツールとして、金字塔とみなされているが、ASDを他の精神疾患とオーバーラップする症状から区別する点において、その補助診断の万能性に疑問を呈する研究報告がなされてきている。よって、複合的見地から最終診断の決定が行われることが望まれる。ASDに特有のPI特定は、言語行動からASDを判別する可能性を探ることにつながり、複合的視点からの診断実現に寄与するものとなる。</p> <p>本研究は、日本人ASD児/者の話し言葉におけるPIの包括的マッピングを実現するために、Systemic Functional Linguistics (SFL)の枠組みから、システムネットワークを作成して、このシステムネットワークを基に、PIの特定を想定したアノテーション・スキームを構築した。このアノテーション・スキームに基づいて、the Corpus of ASD + Typically Developed Spoken Languageを構築した。本コーパスは、ASD被験者186名及び定型発達児/者106名による1187の音声記録のテキストからなり、107万形態素数を収容している。</p> <p>さらに、コーパスを機械学習させることにより、自動アノテーションを試み、約87%の正答率を達成した。データの信憑性を高めるために、本コーパスは、モニター・コーパスとして、定期的に拡大をはかっていくことが計画されているが、その際に、この自動アノテーションの実現は、マンパワー、コスト面の縮小に大きく貢献するものである。通常、コーパス構築には、膨大な量のマンパワーとコストが求められるからである。</p> <p>また、なぜ、PIが起こるのかを、神経認知機能の側面から解明を試みる例として、ASDのPIとして特定された日本語の終助詞を例に、その解明プロセスを例示した。</p>	

本研究で構築したシステム・ネットワークおよびアノテーション・スキームは、他言語にも適用できる汎用性を有するため、他言語への適用が実現されることにより、cross-linguistic PI の研究が可能となる。

手法と結果

1. サンプルング

参加者: は DSM-5 による診断に ADOS-2 による結果を加味して総合的に ASD と診断された被験者 186 名と、ASD の診断基準を満たさなかった定型発達者 (Typically developed; TD) 106 名である。ADOS-2 で課されるタスクの音声データを逐語記録化したものを、コーパス用のテキストとして使用した。

2. 日本語システムネットワークの構築

SFL の理論的枠組みによるシステムネットワークは、言語使用者が、言語使用時に選択する可能性のあるあらゆる言語資源の選択肢を体系的に網羅したものである。4 種のシステムネットワーク、(1) the system of MOOD, (2) the system network of APPRAISAL, (3) the system network of TRANSITIVITY, (4) the LOGICAL systems を作成した。

3. アノテーション・スキームの構築と言語資源の辞書作成

構築したシステムネットワークを基に、アノテーション・スキームを構築、アノテーション項目は、16 のカテゴリーから総計 159 項目となった。これらの項目は、ASD 児/者の実行機能、共同注意、中枢性統合機能にみられる神経認知的障害を考慮し、使用に困難が見られるであろうと仮説を立てて、選択されたものである。これら 16 のカテゴリーごとのアノテーション・ツール内蔵辞書を作成した。アノテーション作業は、これらの言語資源を網羅して辞書をアノテーション・ツールに内蔵させることで、アノテーションの選択肢がツール上に示され、アノテーターが該当する項目を選択する仕組みとなる。

4. アノテーション・ツールの開発

アノテーションを効率的に行うため、独自のツール、SPAT を開発した。

5. コーパスのデザイン構成

Introduction と Corpus Viewer の 2 つのセクションに分かれ、データベースは、Corpus Viewer になる。Corpus Viewer は、Normal, KWIC (Key Word in Context) と Stats から成る。

6. 機械学習によるアノテーションの自動化

マニュアルで行ったアノテーションを機械学習させることにより、アノテーションの自動化をはかり、87% の正答率を達成した。本コーパスは、モニター・コーパスとして随時拡張をはかっていく計画でいるが、これにより、通常コーパス構築にかかる膨大なマンパワーとコスト削減が可能となる。

7. PI として、日本語終助詞を例示

ASD と TD グループ間で、使用に違いがいられた言語資源として終助詞を例に、ASD 者の共同注意欠損の観点から、PI の理由づけを行い、さらに認知機能と言語機能の観点から PI 発現のモデルを組み立てた。