

2.7. コーパスを用いた言語学の研究

研究所 脳機能系障害研究部 高次脳機能障害研究室 鈴木あすみ

言語コミュニケーションでは、発言の状況・話し相手との関係・社会的慣習・世界知識といった文脈に応じて、話し手の意図など「言外の意味」を柔軟に解釈する必要がある。情緒的対人関係の障害、強いこだわり等で特徴づけられる自閉スペクトラム症（Autism Spectrum Disorder: ASD）者は、話し相手の意図や冗談が分からずコミュニケーションに支障をきたす語用論障害を呈する [1]。語用論障害は学校や職場での対人関係にも影響し、抑うつなどの二次障害につながる。適切な支援方法の開発のためには実際の言語データから ASD の言語運用の実態を把握する研究が役立つが、充分に行われていない。

言語運用の定量的・実証的な分析には、人々が実際に話したり書いたりした言葉の大量デジタルデータであるコーパスによる研究が有効である。近年ではマルチモーダルなコーパスにより分析の幅が広がっており、国内では国立国語研究所から音声・映像付きの『日本語日常会話コーパス (CEJC)』 [2] が構築・公開されている。ASD 児・者を対象としたコーパス研究としては、国外では ASD 児・発達遅滞のある非 ASD 児・定型発達 (TD) 児各 16 名 (1 人当たり 30 分) の会話コーパス、ASD 児 46 名 (1 人当たり 7-15 分) の音声付きコーパス、ASD 児 1 名 (11 分) の音声・映像付きコーパスなどが構築され、話し言葉コーパス集約プロジェクト TalkBank の一部である ASDBank [3] で公開されている。国内では、ASD の診断補助ツールである ADOS の映像記録の一部から構築したコーパス (ASD 者 32 名、計 560 分) に基づき自動で重症度を評価するツールの作成 [4] が行われているが、公開コーパスは未だ存在しない。

そこで、本研究は世界初の公開 ASD 日本語会話コーパスへの試みとして、ASD 者・TD 者各 5 名から合計 20 時間分の会話データを収集し、CEJC の構築を主導している小磯花絵教授 (国立国語研究所) の協力のもと公開コーパスを構築する。取得した会話データには自動形態素解析を行い、語用論的情報として談話行為情報 (言い誤り、笑いのタイミングなど) を付与する。また、音声的特徴や表情・ジェスチャー等の非言語情報の分析に向け、会話の音声・映像も合わせて公開する。さらに、ASD 特性にかかわる心理指標 (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))、および生理指標 (心拍変動・眼球運動・唾液中オキシトシン) の測定値もコーパスに含める。言語データとこれらの心理・生理指標の測定値を紐づけた分析から ASD 者の言語運用の実態、特に語用論的特徴を明らかにし、TD 者との違いに関連する要因について検討する。言語学、医学、情報工学など多領域の研究者に利用できる公開コーパスとして本研究が完成すれば、日本語語用論研究に裏付けられた診断ツールや支援方法の開発につながると期待できる。

【参考文献】 [1] Bartak et al. (1975). A comparative study of infantile autism and specific developmental receptive language disorder: I. The children. *British Journal of Psychiatry*, 126, 127-145. [2] Koiso et al. (2018). Construction of the Corpus of Everyday Japanese Conversation: An interim report. *Proceedings of LREC2018*, 4259-4264. [3] <https://asd.talkbank.org/access/> [4] 崎下ほか (2019). 「自閉スペクトラム症の発話者コーパス作成と自動診断支援システム構築」『言語処理学会年次大会発表論文集』25, 276-279.