

当報告の内容は、それぞれの著者の著作物です。

Copyrighted materials of the authors.

AA 研共同利用・共同研究課題「ユーラシア諸言語における談話の文法」(jrp000296)
令和 6 (2024) 年度第 3 回研究会

日時: 2024 年 3 月 18 日 (水曜日)

午前 10:30～午後 2:30

場 所: AA 研 303 室+Zoom (ハイブリッド)

10:30-10:40 児倉徳和 (AA 研所員) 趣旨説明

10:40-12:00 佐藤久美子 (AA 研共同研究員, 国立国語研究所) 「日本語諸方言におけるラ行音節の実現ーコーパスによる計量的把握ー」

13:00-14:20 田口智大 (AA 研共同研究員, ノートルダム大学) 「人工知能を用いたタタール語談話資料の書き起こしとアノテーション」

今回の研究会では、2 件の研究発表と討論が行われた。

佐藤久美子 (AA 研共同研究員, 国立国語研究所) 「日本語諸方言におけるラ行音節の実現ーコーパスによる計量的把握ー」

日本語諸方言において、動詞末の子音/r/を含む音節(以下、ラ行音節)が促音化・撥音化することがある。先行研究によると、ラ行音節がカ行・タ行音節の直前で促音化し、ナ行音節の直前で撥音化するという現象は日本の各地に広く見られる。ラ行音節の音交替は全国規模で起こり、かつ多様な地域差が存在することが示唆されているが、通方言的な調査研究が見当たらない。そこで本発表では、個別方言の記述や方言地図によって音交替の存在が報告されている九州地域を対象として 1、データは、日本語諸方言コーパス (以下,COJADS) Ver. 2023.03 を用いた計量的な予備調査をおこなった。そして結論として、九州地域では、東京方言の[ru]が長音、促音、撥音、イ音で実現するが、促音、撥音、イ音が観察される佐賀・鹿児島、ru が顕著に見られる福岡・大分・宮崎、多くの変異が見られる熊本・宮崎という地域差があることが報告された。

田口智大（AA 研共同研究員，ノートルダム大学）「人工知能を用いたタタール語談話資料の書き起こしとアノテーション」

本発表では、発表者が行うタタール語の文法の分析に向けた談話資料の自動書き起こしとアノテーションの実装についての報告が行われた。近年、機械学習に基づく音声認識・自動書き起こしの技術は目覚ましく進歩しつつあるが、教師データの少ない言語（低資源言語）はサポートされていない問題がある。しかし、理論的には低資源言語であっても教師データを用意できさえすれば、大言語と同様に機械学習に基づく音声認識・自動書き起こしが可能である。本発表では Wav2Vec2 という音声認識モデルとタタール語の音声データをもとに、モデルの訓練、データの評価、出力の修正という一連の作業と、今後の課題が示された。

今回の研究会で行われた 2 件の研究発表はともに談話資料をもとにした資源（コーパス）の構築、および構築された資源を利用した定量的研究であり、メンバーの研究対象の言語における資源の構築の技術的な問題を中心に討論が行われた。このほか、日本語諸方言にみられる音韻現象とユーラシア地域の諸言語にみられる /r/ を中心とした音韻現象に共通性が見られるか、といった議論が行われた。

以 上