

データベースをつくる・つかう：課題と展望

シンポジウム趣旨：

現代的な言語研究においては、個々の研究領域における独自のニーズや目標、研究手法に合わせて、様々な言語資料データベースが開発され、公開されている。本シンポジウムでは、そのような知の蓄積・共有の実現を目指すデータベース開発者（つくる人）、データベースを様々な研究目的のもとに実際的に利用する研究者（つかう人）が講演・ディスカッションを行い、データベースのノウハウ・成果の紹介と共有の機会を提供する。また、つくる側とつかう側の双方の問題の建設的なフィードバックを行うとともに、将来的展望や課題を議論する。

企画・司会： 成田広樹（東海大学）
北原真冬（上智大学）
菅原彩加（早稲田大学）

セッション 1: 容認性判断データをめぐる課題と展望

講演者： 大関洋平（東京大学）

発表題目：容認性判断アーカイブの開発と応用

Development and application of an acceptability judgment archive

発表要旨：

言語学における主要なデータは容認性判断である。しかしながら、容認性判断は、過去 20 年以上に渡って、様々な方法論的な論争が繰り広げられてきた。本講演では、容認性判断の信頼性をめぐる諸問題を概観した上で、容認性判断および文法理論のアーカイブである GrammarXiv（グラマカイク）を紹介する。具体的には、GrammarXiv は、東海大学の成田広樹氏を研究代表者として、2019 年 6 月に言語学と自然言語処理の協働によって始動されたプロジェクトであり、現在プロトタイプの開発が進められている。また、GrammarXiv は現象・仮説・文献のリンクングやソーシャル・ネットワーキングなど様々な機能を兼ね備えており、言語学だけでなく、実験心理学や自然言語処理への応用が期待されている。

セッション司会： 成田広樹（東海大学）
ディスカッサント：上山あゆみ（九州大学）

セッション 2: 子供発話コーパスをめぐる課題と展望

講演者： 野村潤（京都女子大学）

発表題目：CHILDES データのつくり方・つかい方と自然発話研究の役割について

Creating and using CHILDES data: What are the roles of naturalistic studies?

発表要旨：

CHILDES は、子どもの会話データを記録・分析・共有するために開発されたシステムである。現在では子ども以外のデータにも利用され、TalkBank と呼ばれるプロジェクトに発展している。CHILDES を用いれば、公式ウェブサイトの公開データ（約 40 言語）も、独自のデータも分析することができる。引用義務などのルールの遵守を条件に、システムや公開データは無償で利用可能である。本発表では、システムの概要、公開データの利用方法、独自データの作成過程を解説する。また、言語（特に日本語）獲得研究において CHILDES がどのように活用されてきたかを概観し、今後の課題と展望について考察する。

セッション司会： 菅原彩加（早稲田大学）

ディスカッサント：杉崎鉦司（関西学院大学）

セッション 3: 日本語諸方言コーパスをめぐる課題と展望

講演者： 木部暢子（国立国語研究所）

発表題目：日本語諸方言コーパスの構築と活用ーパラレル音声コーパスの可能性ー

Construction and utilization of Corpus of Japanese Dialects (COJADS):

The possibility of parallel speech corpus

発表要旨：

各地の言語・方言の研究にとって、その地域の言語で語られた談話資料は欠かすことができない。しかし、日本にはこれまで各地の言語・方言の大規模な談話データベースがなかった。そこで、国立国語研究所では 2013 年に各地の言葉を一括して調べることが可能なコーパスの開発を始め、2019 年 5 月に『日本語諸方言コーパス：COJADS』（モニター版）を公開した。COJADS は各地の言語・方言と共通語のパラレルコーパスとして構築されている。現在のところ、各地の言語・方言には形態素解析用辞書がない。そのため、共通語解析用辞書 UniDic を利用して各地の言語・方言を検索しようと考えたのである。本発表では、COJADS の構築理念や構築方法について述べ、活用の可能性についても触れる。

セッション司会： 北原真冬（上智大学）

ディスカッサント：小西いずみ（東京大学）