【基盤研究(S)】

全単語解析コーパスに基づく日本語の時空間変異の総合的研究



研究代表者	大学共同利用機関法人人間文化研究機構国立国語研究所・研究系・教授 小木曽　智信(おぎそ　としのぶ)　　研究者番号:20337489
研究課題情報	課題番号:26K21717　　研究期間:2026年度～2030年度 キーワード:日本語史、方言、コーパス、電子化辞書

なぜこの研究を行おうと思ったのか(研究の背景・目的)

●研究の全体像

日本語の研究を行う上で「コーパス」が欠かせないものになっている。言語研究で使うコーパスとは、実際に使われた言葉の用例、文章や会話などを大量に収集して、研究に必要な情報を追加したデータベースである。日本語の歴史研究には国立国語研究所の『日本語歴史コーパス』というコーパスが広く使われている。このコーパスは収録した奈良時代から現代までの文章のすべてが解析され、古文の品詞分解のように全単語に情報が付けられている 。

一方、方言を収録した『日本語諸方言コーパス』は、方言用の電子化辞書がないため、単語情報が付けられていない。また、過去の方言が記録された資料についてはまだコーパスが整備されていない。そのため日本語の歴史や方言を統一的に研究できるデータがない。

この研究では、これらをすべて、同じ基準で解析できる電子化辞書（UniDic）を整備して、日本語の歴史と方言＝時空間変異を研究する基盤を作る。そしてこのコーパスと辞書を活用して、従来は別々に行われがちだった日本語の歴史と方言の研究を総合し、日本語の歴史的変化や方言分布がどのように関連しあって成立してきたのかを解明するための実証的な研究を行う。

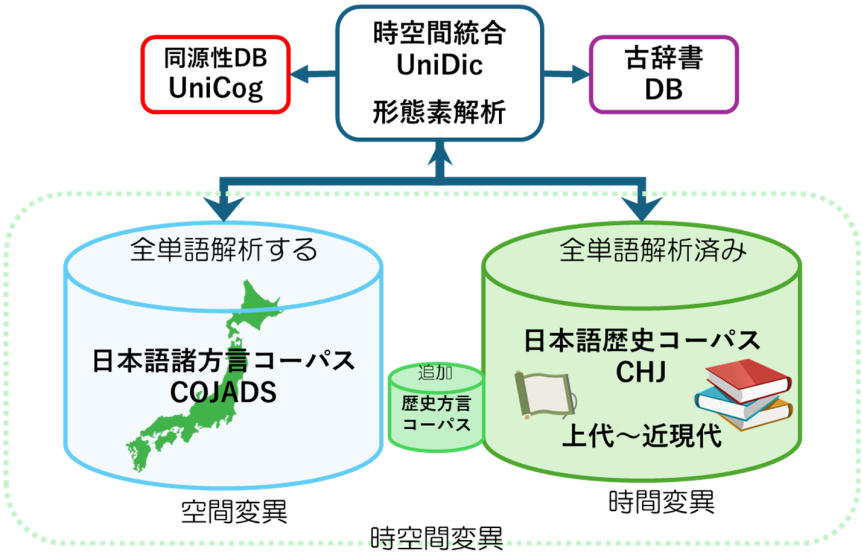

図1　構築するコーパスと電子化辞書

●研究の背景

『日本語歴史コーパス』のための時代別の電子化辞書を整備したのち、各地の方言の解析のための辞書整備に取り組んできた。この中で、各地の方言語形を包括的に扱うには語形や意味の歴史的変化を踏まえた整理が不可欠であることが分かった。ここから、歴史と方言を一体的に分析できる「全単語解析済みの歴史資料と方言資料のコーパスを統合」するアイデアが生まれた。

●研究の4つの柱と体制

- 方言コーパスの整備：全単語解析済みの『日本語諸方言コーパス』を構築【方言コーパス班】
- 歴史方言コーパスの整備：『日本語歴史コーパス』に歴史的方言資料を追加【歴史方言コーパス班】
- 時空間を統合する辞書データの整備：コーパスを解析する電子化辞書UniDicを拡張し、語源辞書や古辞書のデータと連携する【語彙資源班】
- 日本語の時空間変異の総合的研究：整備した研究基盤を活用して日本語史と方言の研究者が協働して研究を展開する【全体】

この研究によって何をどこまで明らかにしようとしているのか

●方言コーパスの全単語解析

『日本語諸方言コーパス』で公開されている47都道府県の方言書き起こしテキストを対象に、音声情報と対応づけたうえで解析を行い、すべての発話に単語情報を付与する。さらに他の方言談話資料についてもコーパス化を進める。

●歴史方言コーパスの整備

近世から近現代にかけての近畿の資料や江戸東京語の資料をコーパスとして整備し地域言語の通時的研究を可能にする。また、東国系抄物・尾張洒落本や、「おもろそうし」をはじめとする沖縄資料など、コーパス化が行われていない歴史的方言資料を『日本語歴史コーパス』に追加する。

●電子化辞書の整備と連携

上記のコーパスを解析可能にするために電子化辞書UniDicの見出し語や活用表を整備して各地の方言専用の辞書を開発する。さらに、UniDicの見出し語を古辞書データベースや語源データベースの見出し語と関連付け、コーパスでは得られない情報も利用できるように連携する。

◆オンライン公開

構築したコーパスはオンラインのコーパス検索アプリケーション「中納言」上で無償で検索・利用できるようにする（図2）。また、開発した方言用のUniDicは、オンライン形態素解析ツール「Web茶まめ」で簡単に利用できる形で公開する（図3）。さらに構築した資源は原則としてオープンデータとして公開する。これにより日本語研究の国際化と裾野の拡大に貢献する。



図2　コーパス検索アプリケーション「中納言」



図3　オンライン形態素解析ツール「Web茶まめ」

◆共同研究成果の発信

構築した研究基盤を用いて実践した共同研究の成果は、論文や学会発表のほかに、年1度のシンポジウムを開催して発信する。

◆社会への貢献

- 地域にとって貴重な文化資源である方言のコーパスと辞書をオープンデータとして公開する。
- 2032年刊行を目指して編集作業が進む日本最大の国語辞典である小学館『日本国語大辞典』第3版の改訂に貢献し、研究成果を広く社会に還元する。

ウェブサイト等	https://clrd.ninjal.ac.jp/chj/　日本語歴史コーパス https://www2.ninjal.ac.jp/cojads/　日本語諸方言コーパス https://chamame.ninjal.ac.jp/　Web茶まめ（自動単語解析）