

『日本語歴史コーパス 和歌集編』Ver. 1.0 の公開

松崎安子, 小木曾智信, 中村壮範

本発表では 2019 年秋にアップデートを行う『日本語歴史コーパス (以下 CHJ) 和歌集編』Ver. 1.0 の内容について報告し、今後の活用の可能性について述べる。

『CHJ 和歌集編』は国立国語研究所の共同研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」の成果の一つとして 2019 年 3 月に短単位データを収録した ver. 0.8 を公開した。今回のアップデートでは、長単位および掛詞の情報、そして歌人情報のリンクが加わることを報告する。

(1) 長単位は文節を基礎として規定される単位で、複合語等が 1 単位として扱われる。今回『CHJ 和歌集編』に長単位の情報が加わることで、既に長単位が付与され公開済みの『CHJ 奈良時代編 I 万葉集』や『CHJ 平安時代編』の中古仮名文学作品との対照が可能となり、各時代・作品の特徴的な語が把握できると期待される。

(2) 『CHJ 和歌集編』の掛詞は、歴史コーパスの質的拡張の試みとして、本文の同一箇所に対して複数の形態論情報が付与できるよう開発された「形態論情報の多重化」の成果の一つである。本発表では、『CHJ 和歌集編』の本文 (以下「主本文」とも) に対し掛詞情報を付与する過程で問題となったいくつかのパターンを取り上げ、構築段階での処理方法と、それらが反映された「中納言」検索結果の表示について説明を行う。

(3) 和歌の作者については、「よみ人知らず」を除いた歌人のうち 16%につき公開済みの作者情報に加えて「国立国会図書館典拠データ検索・提供サービス (Web NDL Authorities)」へのリンクが可能となったことを報告する。