日本語史研究史上における「通時コーパス」の価値

小木曽智信（国立国語研究所）

『日本語歴史コーパス』の構築状況

「通時コーパスの構築と日本語史研究の新展開」プロジェクトでは、前身にあたる「通時コーパスの設計」を引き継いで、上代から近代にわたる日本語の歴史を研究するための基盤となる言語資源として『日本語歴史コーパス』の構築を行ってきた。前プロジェクトの「平安時代編」「室町時代編 I 狂言」につづき、2016 年には「明治・大正編 I 雑誌」「鎌倉時代編 I 説話・随筆」、2017 年には「鎌倉時代編 II 日記・紀行」、「奈良時代編 I 万葉集」、2018 年には「江戸時代編 I 洒落本」「室町時代編 II キリシタン資料」、「明治・大正編 II 教科書」を、コーパス検索アプリケーション「中納言」上で一般公開した。2019 年春には、さらに「江戸時代編 II 人情本」等のいくつかのサブコーパスを公開予定である。計画では 6 年間に構築するサブコーパスは下表の通りであり、構築は予定を上回るペースで順調に進んでいる。

<table>
<tr><td>奈良時代</td><td colspan="2">☑万葉集　□宣命</td></tr>
<tr><td>平安時代</td><td>☑仮名文学</td><td rowspan="2">□和歌</td></tr>
<tr><td>鎌倉時代</td><td>☑説話・随筆　☑日記・紀行　□軍記</td></tr>
<tr><td>室町時代</td><td colspan="2">☑狂言　☑キリシタン資料</td></tr>
<tr><td>江戸時代</td><td colspan="2">☑洒落本　□人情本　□近松</td></tr>
<tr><td>明治・大正</td><td colspan="2">☑雑誌　☑教科書　□文学作品　□新聞</td></tr>
</table>

日本語史研究史上のコーパス

江戸時代以降に展開された日本語の歴史の研究史において、調査対象の資料の主たる提供形態は、写本、板本、印刷本（活字、影印・写真）、デジタルデータと変遷してきた。用例の検索技術で見るなら、本文そのものの調査から、総索引、電子データ、コーパスへと移り変わっている。これは技術の進歩によるもので、デジタル時代の今日において調査に用いる資料がコーパスとなることは必然的なものである。コーパスにもとづく日本語史研究は、契沖以来の文献データにもとづく実証的な研究の、時代に応じたアップデートである。

コーパスにより、用例の検索などの研究に伴う作業的側面は劇的に効率化され、精度も向上した。それでは、研究の質的には飛躍はあるだろうか。研究者の層の限界から短期的には大幅な変化は望めないかもしれない。しかし周辺科学の進化からみて、『日本語歴史コーパス』という基礎データの整備は、質的にも新たな研究への準備となるものだと考える。