

国立国語研究所学術情報リポジトリ

編集後記

メタデータ	言語: Japanese 出版者: 公開日: 2018-03-20 キーワード (Ja): キーワード (En): 作成者: 浅原, 正幸 メールアドレス: 所属:
URL	https://repository.ninjal.ac.jp/records/1513

編集後記

国立国語研究所に着任して5年が経ちました。着任時に依頼された仕事「超大規模コーパスプロジェクト」（2011年度～2015年度）はウェブコーパスを構築せよというものでした。100億語という数値目標が設定されていましたが、今回のワークショップでお披露目することができました。大きなトラブルもなくプロジェクトを進めることができたのは、多くの方のお力添えのおかげです。国立国語研究所コーパス開発センターでは並行して「コーパス日本語学の創成」と「コーパスアノテーションの基礎研究」のプロジェクトが進められ、その枠組の中で「コーパス日本語学ワークショップ」（年2回全8回）が開催されました。

現在、国立国語研究所コーパス開発センターでは「包括的高度検索環境の整備」（2016年度～2021年度）と題し、『中納言』『梵天』などの検索系をより使いやすくするプロジェクトを進めております。前プロジェクトの「コーパス日本語学ワークショップ」をより発展させた形の「言語資源活用ワークショップ」を今回より年1回全6回開催する予定です。2016年度のみ先に人事を進めたために年度末の3月に開催いたしました。2017年度～2020年度は9月に、2021年度は8月に実施する予定です。

前ワークショップはコーパス日本語学を新たに作るということを目指していましたが、本ワークショップはより高度なコーパス利用方法を目標としております。現代日本語（独話・対話・書き言葉）のみならず、方言・古典語・L1学習者の日本語・L2学習者の日本語などが対象になります。コーパスで扱う上でさまざまなことを学ばなくてはなりません。コーパス整形・統計処理などについても議論する場にしていきたいと考えております。

実際のワークショップでは、2件の招待講演と45件の一般発表がありました。招待講演をご快諾くださりました京都大学の秋田祐哉先生・電気通信大学の松吉俊先生、広報不足にもかかわらず一般発表に申込をしてくださった方々に感謝いたします。また「語彙資源活用シンポジウム」をワークショップに併設して企画いたしました。形態素解析用辞書・シソーラスなどの電子化辞書編纂者と紙媒体で出版されてきた辞書編纂者とが熱い議論をとりかわしました。次回以降もテーマを決めてシンポジウムを併催していきたいと思っております。ワークショップ・シンポジウム期間中、朝倉書店・くろしお出版・ひつじ書房・勉誠出版に会場にて出店をお願いしましたところご快諾くださりました。どうもありがとうございました。

今回よりワークショップに寄せられた予稿を発表論文と呼び方を変え、国立国語研究所学術情報リポジトリ <https://repository.ninjal.ac.jp/> にてDOI (Digital Object Identifier) 付きで公開することになりました。ご利用いただければ幸いです。

最初のウェブコーパスの話に戻しますと、検索系『梵天』の2016年9月の試験公開・2017年3月の一般公開以降、一般公開版は24万件の利用が、高機能版は37000件の利用がありました。『梵天』の講習会の実施も35回に及び、内2回については Youtube Live により実施しました。ウェブコーパス NWJC2vec の頒布数も70件を超えて、利用が増えていることをうれしく思います。今後ウェブコーパスを用いた発表が増えていくことを楽しみにしております。

国立国語研究所
コーパス開発センター
浅原 正幸