

公開シンポジウム

大規模コーパス研究の方法 —言語研究の新しいスタンダードの構築にむけて—

概要と問題提起

千葉^{しょうじゅ} 庄 寿 (麗澤大学外国語学部)

1990年代に British National Corpus と Bank of English (現 Collins WordbanksOnline) という2つの大規模英語コーパスが相次いで公開されて以来、いわゆる「コーパス言語学 corpus linguistics」は大きな進歩を遂げました。まず、英語をはじめとするさまざまな言語について、数多くの大規模コーパスが構築されています。2006年度からは、いよいよ日本語の大規模均衡コーパス balanced corpus の構築プロジェクト (文部科学省科学研究費補助金特定領域研究「代表性を有する大規模日本語書き言葉コーパスの構築: 21世紀の日本語研究の基盤整備」) がはじまりました。本シンポジウムのパネリストである前川喜久雄氏を研究代表者とするこのプロジェクトでは、これまでの構築された大規模コーパスのノウハウや、最新のアイデアがコーパスの設計や構築に生かされています。日本語研究においても大規模コーパスの恩恵を真に受けることのできる日が近づきつつあります。

さらに1990年代以降、コーパスを用いた研究手法やコーパスを使った言語研究の理論的な枠組みについても多くの提案がなされています。例えば、今回のシンポジウムのパネリストである滝沢直宏氏が推進する語法研究は、コーパスを用いた文法研究としては最も成功しているアプローチの一つと言えるでしょう。

もう一つの例として、意味論の分野での魅力的な成果を紹介します。コーパスによる語彙分析の手法 (Sinclair 1991; Stubbs 1996, 2002) は、コーパスがもつ文脈情報が意味の記述に画期的な役割を果たすことを示しました。Teubert (2004: 97) は、コーパスを用いた研究が何を明らかにし、何に役立たないか、といった議論を積極的に展開する文脈のなかで、「語や文、テキストが何を意味するかを明らかにしたければ、(生成文法や標準的な言語学の手法ではなく) コーパス言語学を選びなさい」と述べています。ここで補足するならば、コーパスから明らかになる意味はいわゆる「単語」の意味にとどまりません。コーパスの検索ツールを利用したことのある研究者であれば、コーパスの分析でよく用いられる「単語」が言語学的に厳密な単位でないことに不満を抱いた経験があるかもしれません。しかし、最近の研究では、大規模コーパスを意味の「単位」を客観的に切り出すための理論的な装置とみなす新しい視点が提供されており (Hunston & Francis 2000; Tognini-Bonelli 2001)、コーパスを用いた言語研究が、これまでの言語学の手法自体に新たな提案をする可能性が出てきています。

一方、言語研究者にとって、ある程度まとまった量の電子テキストは既になくてもならない資料になりつつある、いや既になっていると言っても過言ではありません。実際のところ、「言語研究に役立つようにとの意図をもって、事前にコーパスの構成をデザインした上で集められた電子テキストの集合」(後藤 2003: 8) としての大規模コーパスの登場を待たずとも、主要な言語であれば大量の電子テキストを入手することは比較的簡単です。大量の電子テキストとしての「コーパス」が、手に入りやすい資料としてますます身近な存在になっています。

このような状況では、研究に利用されるコーパスの種類や規模、内容は多岐に渡り、当然ながらコーパスの使い方も研究者により大きく異なります。その結果、コーパスからどのような知見が得られるか、その際どのような手法や手順を用いたらよいか、さらにコーパスを言語研究に用いる意味は何か、といった根本的な議論が継続的におこなわれることは (少なくとも、英語コーパス言語学や計量的研究など特定の分野を除き) 殆どなく、大規模コーパスを使った初期の研究の知見がそのまま鵜呑みにされているように思われます。大規模コーパスの時代の到来を念頭におきながら、大規模コーパスを用いた言語研究の可能性について問い直す—このことが、本シンポジウムを企画した目的の一つです。

本シンポジウムでは、パネリストお2人（滝沢直宏氏、前川喜久雄氏）とコメンテーターお2人（松村一登氏、定延利之氏）にご参加いただき、大規模コーパスが、言語研究の方法論や言語学という分野自体の今後の発展にどのように貢献するかを議論します。

ディスカッションにおいては、まず、1990年代以降のコーパス言語学の流れを決定づけたともいえる Geoffrey Leech の2論文（Leech 1991, 1992）を手がかりに、「コーパスを用いた言語研究」像の再評価をおこない、大規模コーパス利用の黎明期である 1990 年代から今日にかけてのコーパスの捉え方、構築方法、研究の射程、研究手法がどのように変遷してきたかをひも解き、コーパスを用いた言語研究のこれまでの成果と今後の方向性を整理します。その上で、大規模コーパスの利用に関する疑問や誤解に答え、今後克服すべき課題を踏まえつつ、大規模コーパスを用いた手法を言語研究の「新しいスタンダード」に位置づけてゆくための提言をおこないます。

内容のある議論をおこなうことをめざし、今回のシンポジウムではディスカッションの時間を1時間程度確保する予定です。また、新しい試みとして、シンポジウムではディスカッションのトピックを Web にて広く募集しました。さらに、当日ポスター発表の時間を使い、関連するデモンストレーションを実施します（詳細は以下をご覧ください）。

コーパス言語学はもとより、コーパスを用いた言語研究、さらには言語研究の方法論一般に関心のある皆様のご参加をお待ちしております。

参考文献：

- 後藤 齊 (2003) 「言語理論と言語資料」『日本語学』22(5): 6—15.
- Hunston, Susan & Gill Francis (2000) *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Leech, Geoffrey (1991) “The state of the art in corpus linguistics,” in Aijmer, Karin & Bengt Altenberg (eds.) *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman. Pp. 8—29.
- Leech, Geoffrey (1992) “Corpora and theories of linguistic performance,” in Svartvik, Jan (ed.) *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4—8 August 1991*. Berlin: Mouton de Gruyter. Pp. 105—122.
- Sinclair, John (1991) *Corpus, Concordance, Collocations*. Oxford: Oxford University Press.
- Stubbs, Michael (1996) *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture*. Oxford: Blackwell.
- Stubbs, Michael (2002) *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Teubert, Wolfgang (2004) “Language and corpus linguistics,” in Halliday, M. A. K., Wolfgang Teubert, Colin Yallop & Anna Čermáková (eds.) *Lexicology and Corpus Linguistics: An Introduction*. London: Continuum. Pp. 73—112.
- Tognini-Bonelli, Elena (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

シンポジウム「大規模コーパス研究の方法」関連デモンストレーションのご案内

日時：2007 年 6 月 17 日（日）11:00～13:00 場所：麗澤大学 1 号棟 4 階 1412 教室

- 『小学館コーパスネットワーク』
- 『日本語話し言葉コーパス』
- 『太陽コーパス』
- 『現代日本語書き言葉均衡コーパス』 サンプルの全文検索
- 『現代日本語書き言葉均衡コーパス』の形態素解析
- 『現代日本語書き言葉均衡コーパス』のサンプリング

デモンストレーション実施にあたりご協力をいただきました関係各位にこの場を借りして篤く感謝申し上げます。（なお、内容は変更されることがあります。ご了承ください。）