

「シンポジウム」

話し言葉コーパスの構築と利用

迫田久美子・野口ジュディー・長谷部陽一郎

まえがき

Abstract

This paper is a summary of a symposium on the Construction and Applications of Spoken Corpora, which was presented at the 43rd Annual Conference of the Japan Association of English Corpus Studies on October 1, 2017 at Kwansei Gakuin University, Nishinomiya, Hyogo, Japan. Four corpora and their interfaces were introduced: The ICNALE (The International Corpus Network of Asian Learners of English) by Shin'ichiro Ishikawa, The I-JAS (International Corpus of Japanese as a Second Language) by Kumiko Sakoda, JECPRESE (Japanese-English Corpus of Presentations in Science and Engineering) by Judy Noguchi and TED Corpus Search Engine by Yoichiro Hasebe. This paper covers the latter three as The ICNALE is described in a lecture paper in this issue. The background and rationale for the construction of the corpus or the software interface is introduced together with how they can be used for the benefit of language learning and teaching.

要旨

英語コーパス学会第43回大会（2017年10月1日、関西学院大学）で開催されたシンポジウム「話し言葉コーパスの構築と利用」の内容を紹介する。シンポジウムでは、学習者の書き言葉・話し言葉（英語）ICNALE（石川慎一郎）、学習者話し言葉日本語（テーマ別）I-JAS（迫田久美子）、話し言葉日英（理系プレゼン）JECPRESE（野口ジュディー）、TEDコーパス検索システム（長谷部陽一郎）について取り上げられた。学習者の書き言葉・話し言葉（英語）ICNALEに関しては、当号の講演の論文に掲載している。本稿では、I-JASとJECPRESSならびにTEDコーパス検索システムが構築された背景を説明し、言語教育への応用について紹介する。

はじめに

野口ジュディー

言語学の多くの研究では書き言葉を対象とするが、人間の自然言語は、元来、話し言葉からスタートする。しかし、話し言葉は、研究の対象としては扱いにくいものである。コーパスを作成するためには、発話者から許可を得ないと録音することすら難しい。また、生きた言葉を捉える環境は、録音を行うのに適していないことが多い。このようなハードルを越えて録音ができたとしても、時間をかけて書き起こしを行う必要がある。最終的には、話し言葉コーパスを利用しやすくするために、インタフェースとなるシステムを用意しなければならない。このような作業の苦労話を研究者たちは少なくともいくつか抱えている。ここでは、研究者たちの努力によって構築された異なる4つのコーパスやコーパス検索システムを紹介する。なお、シンポジウムで紹介された ICNALE は、石川慎一郎氏による講演論文“A Frontier in Learner Corpus Studies: For Better Understanding of L2 Learners”に詳しく紹介されている。以下には、学習者話し言葉日本語(テーマ別) I-JAS, 話し言葉日英(理系プレゼン) JECPRESE, TED コーパス検索システムを、それぞれのパネリストが紹介する。

(神戸学院大学名誉教授 Email: jnoguchi@gc.kobegakuin.ac.jp)