

国立国語研究所学術情報リポジトリ

〈著書紹介〉 前川喜久雄 監修/小磯花絵 編
小磯花絵,前川喜久雄,五十嵐陽介,丸山岳彦,伝康晴,籠
宮隆之,西川賢哉,菊地浩平
著『話し言葉コーパス-設計と構築-』

メタデータ	言語: jpn 出版者: 公開日: 2015-10-30 キーワード (Ja): キーワード (En): 作成者: 小磯, 花絵 メールアドレス: 所属:
URL	https://doi.org/10.15084/00000802

前川喜久雄 監修／小磯花絵 編
小磯花絵, 前川喜久雄, 五十嵐陽介, 丸山岳彦, 伝康晴,
籠宮隆之, 西川賢哉, 菊地浩平 著
『話し言葉コーパス—設計と構築—』
講座 日本語コーパス 3
2015年2月 朝倉書店 A5判 200ページ 3,400円+税



小磯 花絵

1. この本の目的

国語研究所はこれまで、『日本語話し言葉コーパス』(CSJ)や『現代日本語書き言葉均衡コーパス』(BCCWJ)など、大規模なコーパスを構築し、日本におけるコーパス研究を牽引してきました。『講座 日本語コーパス』は、2011年に公開されたBCCWJの構築プロジェクトの成果を中心に全8巻で構成されています。本書はこの講座の第3巻に位置付けられますが、唯一、BCCWJではなくCSJの成果を中心にまとめたものです。しかし、CSJ構築時の成果を一つの柱に据えつつも、成果をそのまま伝えることを目指すものではありません。2004年に公開されたCSJの成果に留まらず、その後の拡張や問題となる点、研究での利用可能性などを整理した上で、今後の話し言葉コーパス研究に何が求められているかを見つめ直すこと、それが本書のねらいです。

2. この本の構成

本編は次の六つの章から構成されています。

第1章 話し言葉コーパスの設計 (小磯花絵・前川喜久雄)

話し言葉コーパスの設計について、対象とする話し言葉の選定、アノテーション、データ構造を中心に解説。代表的な話し言葉コーパスとそれにもとづく研究についても概観。

第2章 話し言葉の書き起こし (小磯花絵)

CSJを含む幾つかの転記方式を概観し、その特徴や問題などについて議論。

第3章 発話の単位 (丸山岳彦)

CSJで整備した節単位に加え、対話相互行為の観点から拡張した単位についても解説。

第4章 韻律情報 (五十嵐陽介)

韻律情報付与システムとして、日本語版ToBI(J_ToBI)と、CSJ構築時に自発音声用に拡張したX-J_ToBI、およびその後整備したX-J_ToBI簡易版について解説。

第5章 対話への情報付与 (伝康晴)

CSJの成果から離れ、発話の機能や連鎖構造、聞き手反応など、対話に関するアノテーションを重点的に解説。

第6章 印象評定情報（籠宮隆之）

発話の流暢性や丁寧さなど、話し手の発話音声から聞き手が受ける印象を記録した印象評定情報について解説。

第3章から第6章では、各種アノテーションを用いてどのような研究が可能となるかについても具体的に紹介しています。

また二編の付録では、話し言葉のアノテーションによく用いられる無償のソフトウェアとして、Praat¹とELAN²を取り上げています。Praatは音声韻律情報の付与・分析で、ELANは非言語情報の付与・分析などでも大いに活躍するソフトウェアです。

付録A 音声分析ソフトウェア「Praat」（西川賢哉）

付録B 映像解析ソフトウェア「ELAN」（菊地浩平）

CSJは、独話（モノログ）を中心とするコーパスであったため、残念ながらCSJ構築時には対話のためのアノテーションが十分には検討されませんでした。今後、大規模な対話コーパスの構築が求められる中で、対話独自のアノテーションや各種アノテーションの対話への拡張可能性を検討することが重要となります。第5章は、まさに対話のアノテーションを重点的に取り上げており、また他の章でも対話を考慮した拡張に言及しています。本書がCSJから会話コーパスへの橋渡しとなり、会話を含む話し言葉コーパスの構築・研究の発展に貢献することを期待しています。

小磯 花絵（こいそ・はなえ）

国立国語研究所理論・構造研究系准教授。博士（理学）（奈良先端科学技術大学院大学）。ATR 知能映像通信研究所研修研究員、国立国語研究所研究員、同主任研究員を経て、2009年10月より現職。

主な著書・論文：「日本語自発音声における複合境界音調と統語構造との関係」（『音声研究』18(1), 2014）、「既存のツールと結合した話し言葉コーパス利用環境」（『自然言語処理』21(2), 2014）。

社会活動：一橋大学大学院連携教授、社会言語学会理事、同学会事業委員。

¹ アムステルダム大学で開発されたフリーの音声分析ソフト。配布元：<http://www.fon.hum.uva.nl/praat/>

² マックス・プランク心理言語学研究所で開発されたフリーの映像・音声分析ソフト。配布元：<http://tla.mpi.nl/tools/tla-tools/elan/>