

岡崎敬語

反応文漢字かな混じりデータ

Okazaki Honorifics Transcribed Data
using Chinese Characters

(Ver. 0.0)

国立国語研究所
National Institute for Japanese Language and Linguistics

日本語の大規模経年調査に関する総合的研究
Comprehensive Research
Based on Large-Scale, Long-Term Studies of Japanese

井上史雄（客員教授）・鏑水兼貴
INOUE Fumio & YARIMIZU Kanetaka

平成 26 年 3 月 12 日
12 March 2014

目 次

まえがき

- 概要
- 作業手順
- 利用方法
- 今後の課題
- 参考文献

データ

まえがき

概要

本報告は、国立国語研究所がこれまで半世紀以上にわたって継続した岡崎敬語調査に関するデータベースの試作である。場面別の回答文（以下、反応文）は、本来の敬語の研究に使う以外に、長い談話の断片として、多面的に分析できる貴重なデータである。

過去2回の調査結果については、国立国語研究所(1958)、国立国語研究所(1983)に反応文の結果が掲載されている。第3回調査のとりまとめの際に、過去2回の調査データについても再度電子化作業が行われた。録音資料のある第2回調査に関しては、録音を聞き直して、データの再構築を行っている。

現状の反応文回答のデータベースは、カタカナのみによる音声表記となっている。そのため可読性の低い状態になっており、大量の回答を目視する場合に問題が生じる。また形態素解析処理を行おうとしたが、カタカナのみでは解析精度が著しく低くなってしまった。そのため、カタカナ表記で作成されたデータを、漢字かな混じりデータに変換したデータを作成することを目指した。

同時に、現在のデータにおいて備考扱いとなっている併用回答について、第2回答、第3回答・・・という形で、有効回答扱いにすることにした。これにより、3回の岡崎敬語調査における全ての反応文の回答を対象とした分析ができるようになった。

作業手順

データは、共同研究者向けに配布されている基礎データをベースに加工した。まず備考欄から併用回答について抜きだして第1回答に連結した。具体的には、第1回答の直後に、「②」に続いて第2回答を、「③」に続いて第3回答を連結する、というようになっている。なお、併用回答がある場合でも第1回答に「①」はつけていない。

つづいて回答者と反応文番号からなる表形式を、1行1回答者（1回答ではない）に分解した。のべ回答者数は13191人である。

データは文節（に近い単位）ごとにスペースで区切られたカタカナ表記となっている。そのため、まず、すべてをひらがな表記に変換し、元の区切り情報を利用して、文節の出現頻度一覧を作成した。そして文節単位で、漢字に一括変換可能かどうかのチェックを行い、変換作業を行った。

利用方法

データは、1行1回答者であるため単純である。以下に書式を示す。

- 1列目 回答ID (XXXX_XXX)
- 2列目 調査次 (1~3)
- 3列目 回答者ID (XXXX)
- 4列目 反応文ID (101~120)
- 5列目 回答 (カタカナ)
- 6列目 回答 (漢字かな混じり)

回答IDは、4桁の回答者IDと3桁の反応文IDからなる。調査次は回答者IDの先頭桁でもある。反応文IDは、調査次によって反応文の調査番号が異なるため、以下に統一した。

統一番号	場面名	Context	第1回	第2回	第3回
101	道教え	Tell the way			
102	振込用紙	Remittance slip	電報用紙	電報用紙	
103	荷物預け	Check baggage			
104	傘忘れ	Left umbrella			
105	先生	Ex-teacher			
106	新聞代	Newspaper bill	電燈料	電燈料	
107	議事堂	National Diet			
108	医者	Doctor			
109	席譲られ	Offered seat			
110	おつり	Small change			
111	傘貸し	Lend umbrella			
112	魚釣り	Fishing	なし		
113	市役所				
114	先生の絵		なし		
115	第三者 尊敬表現【話し手<話し相手<話題の人物】		なし	なし	
116	第三者 尊敬表現【話し手<話し相手>話題の人物】		なし	なし	
117	第三者 尊敬表現【話し手=話し相手<話題の人物】		なし	なし	
118	第三者 謙譲表現		なし	なし	
119	物売り		<112>	なし	なし
120	公害問題		なし	<115>	なし

※ 電報用紙・電燈料：実施時の場面名 <112>・<115>；実施時の反応文番号

今後の課題

このバージョンは、最初の一括変換処理を終えた状態であり、未変換箇所や要修正箇所が多いため「バージョン 0.0」とする。データは別ファイルとして添付し、今後の改訂版は WEB 上で順次公開する。

現状版に対して MeCab と UniDic を用いた形態素解析を行っても解析精度が低いため、現時点では「バージョン 0.0」の形態素解析結果をもとにして、全回答のチェック作業を行っている。漢字変換箇所の修正を進めるほか、UniDic への辞書追加も行って、形態素解析の精度を上げる予定である。

現状版でも、カタカナ表記の状態と比較すれば、可読性ははるかに高く、談話データの処理も行いやすくなっている。さらに改訂作業を重ねて使いやすいものにしていきたい。

参考文献

- 国立国語研究所 (1958) 『敬語と敬語意識』 秀英出版.
- 国立国語研究所 (1983) 『敬語と敬語意識 — 岡崎における 20 年前との比較 —』 三省堂.
- 国立国語研究所 (2010) 『敬語と敬語意識—愛知県岡崎市における第三次調査—』 東京：科学研究費補助金研究成果報告書 第 1～4 分冊.

岡崎敬語調査資料集 3

Material for Okazaki Survey of Honorifics

岡崎敬語

反応文漢字かな混じりデータ

Okazaki Honorifics Transcribed Data

using Chinese Characters (Ver. 0.0)

日本語の大規模経年調査に関する総合的研究

Comprehensive Research

Based on Large-Scale, Long-Term Studies of Japanese

執筆：井上史雄（客員教授）・鎌水兼貴

INOUE Fumio & YARIMIZU Kanetaka

発行：平成26年3月12日

12 March 2014

国立国語研究所

National Institute for Japanese Language and Linguistics