

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 24 日現在

機関番号：13601

研究種目：基盤研究(B)

研究期間：2011～2013

課題番号：23320095

研究課題名(和文)奄美方言データベース作成のための研究

研究課題名(英文)A study for tagging Amami dialect database

研究代表者

沢木 幹栄(SAWAKI, Motoei)

信州大学・人文学部・教授

研究者番号：20110116

交付決定額(研究期間全体)：(直接経費) 7,500,000円、(間接経費) 2,250,000円

研究成果の概要(和文)：『徳之島方言2000文辞典』の徳之島文テキスト全部を語に分解し、それぞれにXML言語の書式で情報付けを行った。すべての語はそれぞれの品詞のグループのなかで番号を与られている。この語番号と、活用形の番号と品詞番号が語に与えられた情報である。徳之島方言の構文はこのデータから明らかにすることができる。徳之島方言は名詞も助詞も「長い」語形と「短い」語形の両方を持つ。どちらの語形が出現するかは未解明だったが、このデータを使うことで大きな前進がえられる。このデータから報告書『徳之島方言辞典 基礎データ版』を作成し、出版することができた。

研究成果の概要(英文)：We broke up the whole text of "Dictionary of 2000 sentences in Tokunoshima Dialect" into words which are tagged with part of speech information, paradigmatic information and id number. The tagged data is a great help for elucidating syntax of Tokunoshima dialect. We published our report "Dictionary of Tokunoshima dialect - basic data" using the tagged data as source.

研究分野：人文学

科研費の分科・細目：言語学・日本語学

キーワード：琉球方言 XML言語 電子化 徳之島 コーパス

1. 研究開始当初の背景

研究代表者の沢木は研究分担者の福嶋、中島、菊池とともに平成 15~17 年度に基盤研究 (B)「徳之島方言辞典作成のための研究」(課題番号 15320050)を、平成 18~20 年度に同「徳之島方言辞典語彙編の作成のための研究」(課題番号 18320066)を行った。

後者の研究の成果として『改訂版 徳之島二千文辞典』のデータ編 DVD を刊行したが、この DVD はデータからデータへの相互参照が可能なハイパーテキストであり、動画や音声視聴することができるマルチメディアタイトルである。

ここでは文節による kwic (KeyWord In Context) を作り、そのキーワードをさまざまな方法で配列した。

この DVD では、kwic によって可能と思われることはほぼすべてやりつくした感がある。そのうえで、データを作るためのプログラムをすべて公開し、広く一般の人が利用できるようにした。

琉球方言はある意味では日本の方言のなかで一番研究が進んでいるが、構文に関してはほとんど手付かずである。我々の研究はその構文の研究に有力な手助けとなるものだった。

しかしながら、kwic のデータを使うことの限界を知ったのも事実である。文節の内部にある助詞や助動詞は検索することができないのである。

そこで、文を語(付属語も)に分解し、タグ付けをしたデータの必要性が痛感された。

また、我々3人(沢木、中島、福嶋)は『奄美德之島のことば - 分布から歴史へ』(柴田武ほか編 1977年)のもとになった調査(東大調査)に参加しており、このときの徳之島全島の良質なデータを再利用するメリットを熟知していた。現代において、過去に収集した方言資料をデータベース化して公開に供することが一般化しつつある。

2. 研究の目的

本研究は琉球方言の一部である奄美方言についての精密なタグ付けによる方言コーパスと語彙によるデータベースを構築し、公開することを第一の目的とする。

この方言コーパスは今まで日本の方言研究では存在しないタイプのものであり、助詞の用法や活用形の意味、文中のアクセントなど広汎な分野で新たな知見が期待できるものである。

第二の目的として、1. で述べた東大調査の調査票と録音を電子化する。

東大調査は約 40 年前に行われており、語彙の面でも、喉頭化子音が保たれていた発音の面でも在来の方が非常によく保存されていた。また、徳之島的全集落で調査が行わ

れていたため、語彙のみの調査であったが、全島の方言状況を知るのに現在では望むべくもない良質の資料である。この 40 年の間に消滅した集落もあり、また、喉頭化子音をはじめとして共通語化により方言的特徴が消滅しつつあるからである。

3. 研究の方法

「徳之島二千文」(上述の『改訂版 徳之島二千文辞典』のもとになった資料)の徳之島方言文を語に分解した。しかるのちにタグ付けを行う。この作業を省力化するためのプログラムを作成した。

語の切り出しは研究者(沢木)が目視で行ったが、タグ付けデータを作る過程で正しい切り出し方を新たに見つけたり、今まで知られていなかった助詞を発見したり、動詞の活用について新たな知見を得たりした。

タグ付けが完成した段階で、データを利用するためのプログラムを作成した。

一方で、浅間方言の相対的な位置を知るために、徳之島の数地点で調査を行い、さらに 30~50 代の壮年層を中心に全島調査を行った。

東大調査の調査票は pdf のファイルとし、録音は mp3 のデータとした。

4. 研究成果

タグ付けデータは完成している。

『徳之島方言二千文辞典』の徳之島文テキスト全部を語に分解し、それぞれに XML 言語の書式で情報付けを行った。すべての語はそれぞれの品詞のグループのなかで番号を与えられている。この語番号と、活用形の番号と品詞番号が語に与えられた情報である。徳之島方言の構文はこのデータから明らかにすることができる。

異なる動詞の同じ活用形をまとめて表示することによって、活用形の意味を再検討することもできる。たとえば、今までは活用形の意味を一つと考えていたが、それを二つに分けるということも起こり得る。

徳之島方言は名詞も助詞も「長い」語形と「短い」語形の両方を持つ。どちらの語形が出現するかは未解明だったが、このデータを使うことで大きな前進がえられる。

もちろん、徳之島方言の統語論の研究にも大きな貢献をなすことと思われる。

このデータから報告書『徳之島方言辞典基礎データ版』を作成し、出版することができた。今後の公開方法は検討中であるが、今までの国内の方言研究ではタグ付けしたコーパスはなく、その意味で非常に新しい方法である。国外でも、現在始まりつつあるものであって、今後国外の研究者とも連絡を取りつつ、さらに発展させたい。

タグ付けコーパスの方法論とともに作成支援のプログラム、作成手順、利用するため

のプログラムをパッケージとして公開したい。

東大調査の調査票と録音の電子化が完了した。こちらも公開方法を検討中である。

4. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 7 件)

沢木幹栄、徳之島方言の音節頻度表から何が分かるか、信州大学人文科学論集、査読有、1巻、2014、pp75-82

Yumi Nakajima、Making a dictionary from a sentence-based corpus、Dialectologia(Revista Electronica)、査読有、Special Issue、2013、pp199-211

Motoei Sawaki、Dialect dictionary in a DVD、Dialectologia(Revista Electronica)、査読有、Special Issue、2013、pp213-219

Motoei Sawaki、Yumi Nakajima、Chitsuko Fukushima、Standardization and Dialect Leveling in Tokunoshima、Working Papers from NVAW Asia-Pacific 2、査読有、2013、pp1-8

Chitsuko Fukushima、Revisiting regional variation on an Island after 30 Years、Proceedings of Methods : Papers from the Fourteenth International Conference on Methods in Dialectology,2011、査読有、2013、pp305-314

福嶋秩子、奄美德之島方言における否定の意味を表す形式の消長、国際地域研究論集、査読無、5巻、2013、pp75-87

沢木幹栄、新 日本語学者列伝<グロータース>、月刊日本語学、査読無、30巻13号、2012、pp74-84

〔学会発表〕(計 6 件)

福嶋秩子、コンピュータによる言語地図の作成と総合、日本行動計量学会第40回大会特別セッション「方言分布と計量」、2012.9.16、新潟県立大学

沢木幹栄、方言分布の計量、日本行動計量学会第40回大会特別セッション「方言分布と計量」、2012.9.16、新潟県立大学

Motoei Sawaki、Yumi Nakajima、Chitsuko Fukushima、Standardization and Dialect Leveling in Tokunoshima、The 2nd Meeting of NVAW、2012.8.4、国立国語研究所

Motoei Sawaki、Yumi Nakajima、Chitsuko Fukushima、Making the most of a dialect corpus:Development of the Tokunoshima Dialect Dictionary of Two Thousand Sentences、7th Congress of the International Society for Dialectology and Geolinguistics、2012、7.27 ウィーン、オーストリア科学アカデミー

Chitsuko Fukushima、Dialect Lexicography in Japan、7th Congress of the International Society for Dialectology and Geolinguistics、2012、7.23 ウィーン、オーストリア科学アカデミー

Chitsuko Fukushima、Revisiting Regional Variation on an Island after Thirty Years、Methods in Dialectology 14、2011.7.3、カナダ 西オンタリオ大学

〔図書〕(計 1 件)

岡村隆博、沢木幹栄、中島由美、福嶋秩子、菊池聡、私家版、徳之島方言辞典 基礎データ版、2014、154

5. 研究組織

(1)研究代表者

沢木 幹栄 (SAWAKI,Motoei)
信州大学・人文学部・教授
研究者番号：20110116

(2)研究分担者

中島 由美 (NAKAJIMA, Yumi)
一橋大学・社会(科)学研究科・教授
研究者番号: 20155732

福嶋 秩子 (FUKUSHIMA, Chitsuko)
新潟県立大学・国際地域学部・教授
研究者番号: 80189935

菊池 聡 (KIKUCHI, Satoru)
信州大学・人文学部・准教授
研究者番号: 80189935