

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月16日現在

機関番号：12608

研究種目：挑戦的萌芽研究

研究期間：2010～2012

課題番号：22652048

研究課題名（和文） 日本語学習者誤用コーパスを利用した作文システムの開発

研究課題名（英文） Development of Writing Support System using Japanese language Learners' Corpus

研究代表者

仁科 喜久子 (NISHINA KIKUKO)

東京工業大学・留学生センター・名誉教授

研究者番号：40198479

研究成果の概要（和文）：外国人留学生が日本の大学で論文など作成するに当たり Web 上で利用できる文書作成支援システム「なつめ」の完成を最終目標とし、日本語教育、言語学、自然言語処理、教育工学の専門家による共同研究により以下の成果を得た。(1)理工系留学生の作文支援のために必要な日本語の言語構造の特色を分析し、特に共起検索をシステムに導入した。(2)学習者誤用分析のための学習者コーパスを構築し、「なたね」として公表した。(3)誤用コーパスを利用し、自動添削システム「ナツメグ」の機能を追加した。

研究成果の概要（英文）：The goal of this research project was to construct Natsume a Japanese language writing support system. Firstly we proposed an effective method based on collocation search. A linguistic analysis essential to international students was made based on Japanese language corpora. Additionally, we constructed Natsume equipped with automatically correct learners writing system by using a combination of the Natane learner corpus and native Japanese language corpora.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,000,000	0	1,000,000
2011年度	800,000	240,000	1,040,000
2012年度	600,000	180,000	780,000
年度			0
年度			0
総計	2,400,000	420,000	2,820,000

研究分野：人文学

科研費の分科・細目：言語学・日本語教育

キーワード：教育工学・教材・教育メディア・作文支援システム・学習者誤用コーパス

1. 研究開始当初の背景

研究開始当初、日本語教育においては紙ベースでの学習者の誤用分析の研究として、寺村誤用データ(1990)、それに続く名古屋大学「日本語学習者コーパス」、市川保子(1997)「日本語誤用例文小辞典」の研究が存在した。さらには、母語干渉による誤用分析の研究として宇佐美洋(1999～現在)らによる国立国語研究所「日本語学習者による日本語作文とその母語訛との対訳データベース(作文対訳

DB)」の精密な研究があった。一方、教育工学分野では学習者コーパスの必要性を認識した李在鎬(2009)「タグ付き日本語学習者コーパスの開発」の研究などもあるが、誤用分析法にとどまり、本申請の目指す作文支援開発には至っていなかった。

我々は留学生向け日本語作文支援システム「なつめ」

(<http://hinoki.ryu.titech.ac.jp/natsume/>)の研究・開発を通じて、日本語学習者に大規

模コーパスを利用した正用辞書を構築し提示することの有効性を示した。この研究の過程で、学習者誤用コーパスを構築し利用すれば、正用辞書のみを用いるものより更に効果的な学習ができるという確信を得た(仁科喜久子 2009 特定領域研究「日本語コーパス」全体会議予稿集)。そこで阿辺川らの **Constructing a corpus that indicates patterns of modification between draft and final translations by human translators**

(LREC2008) の成果を参照した。これは、初心者による翻訳文をベテラン翻訳者が修正したテキストから修正要因の分析を行い、その結果をデータベース化し、洗練された翻訳へと導くプログラムであり、ここで巧拙両データの差分と修正要因のデータを利用するというアイデアは日本語学習者作文に応用できると考えた。学習者作文の誤用箇所と訂正後の文をもつ学習者誤用コーパスに適用することで、不自然な表現乃至誤用であることを判定し、適切な表現に導けると考えた。この考えにもとづき、実現の可能性が高い項目を優先的に処理し、実用化することにした。

2. 研究の目的

本申請までに開発してきた学習者作文支援システム「なつめ」が用いた大規模日本語コーパスなどの正用となる言語資源に加えて、日本語学習者誤用コーパスを用い、これらに統計的手法を施すことで、日本語学習者が犯しがちな誤用のパターンを自動獲得する手法の確立を目指す。学習者が日本語で論文などを執筆する際に、学習者自身が気づかない誤りの不適切さを指摘し、修正案を提示できるような新たな日本語教育ツールを実現することを目指す。

期間内に以下の計画を実施する。

- (1)言語学および日本語教育の専門家による誤用分析とこれに基づいた日本語学習者誤用コーパスの整備・拡充、
- (2)学習者誤用コーパスの検索システムの構築、
- (3)学習者誤用コーパスおよびその他の言語資源を利用した機械学習による誤用検出手法の検討
- (4)誤用検出・修正案提示手法の日本語学習者向け作文支援 Web システム「なつめ」への導入

3. 研究の方法

日本語教育専門家による誤用の指摘、理由、修正案の記述を基にした誤用データベースを構築する。

初年度は、日本語教育の専門家と情報工学の専門家が協力して誤用タグセットの体系化を進める。

2年目は、学習作文を追加し、自動化でき

る可能性のある項目を重点的に見ながら、誤用タグセットの整備をする。この結果を日本語教育専門家の目で評価し、システムを改善する。

最終年度は作文支援 web システム「なつめ」上に搭載し、試行実験をし、効果的なツールに洗練させる。これを実現するためにグループ1からグループ3に分けて作業を行う。

4. 研究成果

研究の目的と方法に従って、下記のような研究を実施した。

(1)グループ1：第二言語習得・言語学的分析・アノテーション

担当者は仁科(代表者)、山元(分担者)、村岡(連携研究者)、Bekes(海外研究協力者、スロベニア・リュブリャナ大学教授)、曹紅荃(海外研究協力者、西安交通大学日本語科准教授)黒田(研究協力者、早稲田大学、日本語研究センター・准教授)からなる。山元は e-Learning における作文学習の方法を検討した。村岡は理工系留学生の作文論理構造の分析を行い、専門日本語作文能力養成のためのテキスト分析および作文スキーマの研究を進め、アカデミック・ライティングのための作文教科書として村岡・因・仁科(2013)を上梓した。Bekesらは統計言語学的アプローチによる共起の構文的、意味的問題をディスコースレベルまで分析し、誤用と正用間の比較を含めた研究を進めた。その結果、Content shifts as predictors of the wa-topic in Japanese narrative (Andrej Bekes) (「日本語学習支援の構築」(2012)所収)などにおいて、文章全体における語の位置の機能を分析し、さらに語と語の共起が文中で連続せずに、「遠隔」の位置で出現する現象をディスコース論として考察し、共起検索の理論的な方針を示唆した。曹紅荃は漢字圏学習者の母語干渉による転移の問題を追求し、誤用分析を通して、認知言語学的な様々な発見をした。曹、仁科(2011)(2012)では、誤用タグ付けの基本フレームを検討し、誤用データを「誤用の対象」「誤用の内容」「誤用の要因・背景」という3つの異なる視点によるグループ、また語のレベルから、句、文、ディスコースレベルまで4層構造からなる誤用種別の構造を提案し、誤用の全体像を示すことが出来た。

(2)グループ2：システム構築

担当者は阿辺川(国立情報学研究所)、鈴木泰山、八木豊(楸ピコラボ、システム開発者)、傳亮(フウズラボ、システム開発者)、Hodoscek(大学院生、研究協力者)の協力者からなる。〈グループ1〉が構築した学習者誤用コーパスを用いて、誤用事例検索ツールを構築した。またこのツールを公開し、日本語学習者向けの問題作成や学習者の作文添削

に利用できるようにした。

①阿辺川らは語の共起検索技術を用い、ジャンル毎の共起語の傾向を計算することで、文章ジャンルのレジスターとなる語を学習者に示すことを可能にした。

②Bor Hodoscek, 仁科喜久子(2011)、では、BCCWJおよび科学技術論文コーパスを利用して、記述する語が作文のジャンルに適切な表現かどうかをテキスト中の頻度から判定する方式を提案した。

③学習者コーパス「なたね」および自動添削システム「ナツメグ」構築

八木豊, Hodoscek Bor, 仁科喜久子 (2012)、八木豊, ホドシチェック・ボル, 阿辺川武, 仁科喜久子(2013) では、機械学習による誤用検出手法の基礎となる試行を表記誤り、格助詞の構文誤りを例に実験を行い、実現の可能性を考察した。さらに正用コーパスBCCWJと学習者コーパスを用いて、それぞれの使用頻度から誤りやすい語の傾向と正解候補を判定する方式を提案し、この考えに基づいてデザインされた学習者コーパス「なたね」と自動添削システム「ナツメグ」を構築した。「なたね」は、学習者の誤用を検索するシステムとしての機能も備えている。「ナツメグ」は、学習者の誤りやすいデータを収集し、記述した文字列から誤りらしいものを発見し、それに対する正解のヒントを提示するシステムである。

以上の提案は日本語学習のみに限らず、他言語学習にも応用でき、様々な言語の自学自習システムへの応用が可能になる。さらに自動校正システム、翻訳技術への技術的な応用も期待される。

(3)グループ3：学習者評価実験

誤用検索および作文支援システムを試用する実験を行い、システムの効果を評価した。システムの動作、データベース設計、インターフェース・デザイン表示などに関する評価を行う。2012年度から室田真男（東京工業大学教授）を分担者に加え、教育工学専門家によるe-Learningの研究を強化した。ホドシチェック・ボル, 阿辺川武, アンドレイ・ベケシュ, 仁科喜久子 (2011) では、2011年の時点で「なつめ」の学習者使用実験を行い、特に中級程度以上の学習者には効果があることを明らかにしたが、一方では、学習者が自ら誤りの可能性があることに気づかない誤りに対する支援ができないことがわかり、さらに開発を進めている。

5. 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10件)

①小清水貴子、大石智里、藤木卓、寺嶋

浩介, 室田真男、教育養成課程におけるICT機器を活用した模擬授業の実践と学生の意欲の変容、日本教育工学会論文誌、査読有、Vol. 36, 2013, pp.69-72,

②ボル・ホドシチェック, 山元啓史: 不確かな情報が含まれる文の形式、経済社会研究プロジェクト高度科学技術社会リスク・ソリューション、東京工業大学大学院社会理工学研究科、査読なし、Vol.1、2013、pp.236-245

③Bor HODOŠČEK, Kikuko NISHINA: “Japanese Learning Support Systems: Hinoki Project Report、Acta Linguistica Asiatica、査読なし、Vol 2, No 3. 2012、pp.1-23

④ホドシチェック・ボル, 阿辺川武, アンドレイ・ベケシュ, 仁科喜久子、レポート作成のための共起表現産出支援、専門日本語教育学会、査読有、Vol. 13、2011、pp.33-40

〔学会発表〕(計32件)

①八木豊, ホドシチェック・ボル, 阿辺川武, 仁科喜久子、学習者が犯す誤用の要因・背景からみる日本語作文支援、国立国語研究所第3回日本語ワークショップ予稿集、2013年3月1日、pp.307-312 東京国立国語研究所

②Bor Hodoscek, Takeshi Abekawa, Masao Murota, Kikuko Nishina、Readability of Example Sentences in Writing Assistance Tool Natsume、2012年8月21日、The Fifth International Conference On Computer Assisted Systems For Teaching & Learning Japanese, Proc.of CASTEL-J, pp.1-4、名古屋大学

③ホドシチェック・ボル, 仁科喜久子、BCCWJにおける出典情報とトピックおよびレジスターとの関係、国立国語研究所第一回コーパス日本語学ワークショップ、第一回コーパス日本語学ワークショップ予稿集、2012年3月6日、東京国立国語研究所

④八木豊、Hodoscek Bor, 仁科喜久子、BCCWJと学習者作文コーパスを利用した日本語作文支援、国立国語研究所第一回コーパス日本語学ワークショップ、第一回コーパス日本語学ワークショップ予稿集、2012年3月6日、東京国立国語研究所

⑤曹紅荃, 仁科喜久子、文産出と質問紙調

査から見る「遠」で表記する和語動詞の習得について、日本語教育方法研究会誌 Vol. 19、No. 1、2012年3月10日、pp. 8-9 東京、国際基督教大学

⑥阿辺川武, Hodoscek Bor, 仁科喜久子. 語の共起を効率的に検索できる日本語作文支援システム「なつめ」の紹介, 言語処理学会第17回年次大会, 言語処理学会第17回年次大会 発表論文集, 2011年3月9日、pp. 595-598、豊橋科学技術大学

⑦Bor Hodoscek, 仁科喜久子. 作文支援システムにおけるレジスターの扱い, 世界日語教育研究大会, 異文化コミュニケーションのための日本語教育, 2011年8月19日、pp. 522-523、中国、天津外国語大学

⑧曹紅荃, 仁科喜久子. 誤用データベースにおける誤用種別の策定, 日本語教育方法研究会誌、Vol. 18、No. 1、2011年3月26日、pp. 38-39、東京、首都大学東京

⑨村岡貴子, 因京子, 仁科喜久子: “専門日本語ライティング能力の獲得を目指す日本語テキスト分析タスク活動を通じたスキーマ形成” 世界日語教育大会. 2010年8月21日台湾、台北市、国立政治大学

⑩曹紅荃, 黒田史彦, 八木豊, 鈴木泰山, 仁科喜久子、学習者作文支援システムのための誤用データベース作成-動詞の誤用分析を中心に- 世界日語教育大会. 2010年8月21日台湾、台北市、国立政治大学

[図書] (計5件)

① 村岡貴子, 因京子, 仁科喜久子 大阪大学出版局 論文作成のための文章力向上プログラムー アカデミック・ライティングの核心をつかむー (2013) 175

②鎌田美千子, 曹紅荃, 歌代崇史, 村岡貴子 (編著), 仁科喜久子 (監修) 凡人社 日本語学習支援の構築 言語教育・コーパス・システム開発 2012 291

③ Kikuko Nishina: „Publishing House Japan Construction of Speech Database for Second Language Learning of Japanese, Computer Processing of Asian Spoken Languages edited by S. Itahashi, Chiu-yu Tseng Consideration Books c/o The Americas Group ISBN:978-0-935047-72-1. 2010 147-150

[その他]

ホームページ等

<http://hinoki.ryu.titech.ac.jp/>

http://hinoki.ryu.titech.ac.jp/natane/misuse_search

6. 研究組織

(1) 研究代表者

仁科喜久子 (NISHINA KIKUKO)
東京工業大学・留学生センター・名誉教授
研究者番号: 40198479

(2) 研究分担者

室田 眞男 (MUROTA MASAO)
東京工業大学・社会理工学研究科・教授
研究者番号: 30222342

山元 啓史 (YAMAMOTO HIROFUMI)
東京工業大学・留学生センター・准教授
研究者番号: 30241756

(3) 連携研究者

村岡 貴子 (MURAOKA TAKAKO)
大阪大学・国際教育交流センター・教授
研究者番号: 30243744