

自動生成型誤用コーパスを利用した日本語学習用 e ラーニングシステムの開発

林 彰英[†] 稲葉 竹俊[‡] 中野 裕也[‡]

東京工科大学大学院メディア学研究科メディア学専攻[†]

東京工科大学メディア学部メディア学科[‡]

1. 背景

海外において、あるいは来日して日本語を学習する外国人日本語学習者の数は、日本経済が沈滞しているにもかかわらずまだ増加の一途をたどっている。こうした、日本語学習者にとってもパソコンあるいは、インターネットは身近なものとなっている。

また、e ラーニングという新しい教育方法も確立されつつあり、英語等の外国語学習用 e ラーニング教材は数多く開発され、販売されている。

しかし、日本語学習用 e ラーニング教材については開発段階にはあるものの、今尚実運用レベルに達している教材は少なく、他言語などに比べると遅れているのが現状である。

このため日本語学習者にとってもパソコン上で自習できる日本語学習用 e ラーニング教材の実用化が待望されている。

2. 研究の概要

本研究は、上述した日本語学習用教材の現状を踏まえ、コーパスとマルチメディアを利用して、受け身文の日本語短文作成学習用教材を開発した。

コーパスとは主にコンピュータによる処理を前提とした電子テキストの大規模な集合体である。コーパスの種類には学習者コーパス（学習者の産出する母語ではない学習言語データを組織的に収集し、電子的に保存したもの）、外国語コーパスなどがあり、本研究ではこの学習者コーパスの中でも、特に誤用コーパスを用いた。

学習者はコーパスを利用することにより、自分の誤りを他の学習者の誤用例から見つけることができ、教師はどの国の学習者がどのような典型的な誤りをおかすのかというのを

調べることを可能とし、今後の試験や授業においてそれぞれの国ごとの学習者への個別の対応ができる仕様とした。

3. システム概要

3.1 先行研究

コーパスを用いた日本語学習用システムを考えた時、最も重要な先行研究として杉浦[1]がある。この研究では、学習者が誤用を閲覧するのに、学習者コーパスを利用している。

様々な実験がなされているが、書く能力の向上には、以下の3つが条件とされると結論付けられた。

1. 例文による肯定的証拠の提示
2. 誤用例による否定的証拠の提示
3. 知識の手続き化のための学習活動

これらを満たしたシステムを開発することができれば文産出による影響を与えることができると述べられている。

この3つの条件を踏まえた上で、誤用コーパスとマルチメディア教材とを組み合わせたシステムを開発した。

3.2 システムの概要

教材は Web ブラウザを利用して学習する。教材本体は HTML、コーパスは XML で記述した。図1が本教材の第2問画面である。

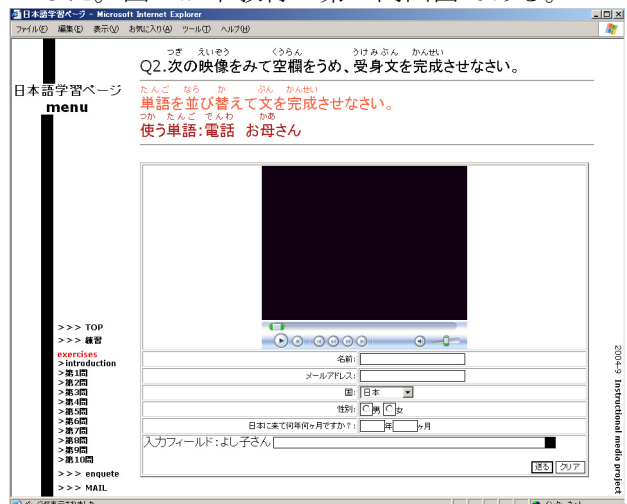


図1. 第2問画面

学習者は学習項目を勉強したうえで、簡単なテストに取り組む。その上で、短文作成の練

Development of e learning system for studying Japanese using automatic generation type corpus of Japanese errors
[†]HAYASHI, Akihide, Tokyo University of Technology Graduate School of Media Science
[‡]INABA, Taketoshi, NAKANO, Hiroya, Tokyo University of Technology School of Media Science

習問題へ移り、誤りが生じた場合、自分の入力した文と例文との比較、文法項目を確認し、再び文を入力するという繰り返し学習を行えるようにした。そのことにより上述の1, 3を兼ね備えた学習が可能となる。

作文問題に用いる映像は日本語の教材ビデオ[2]を利用した。

本システムの1番の特徴は、学習のたびコーパスを自動で蓄積できることである。

短文作成問題で入力した文は Perl で組まれたプログラム上で正誤判定を行う。誤文であれば入力文、誤り語の他に、学習者データとして学習者氏名、国籍、日本語学習歴が XML タグ付与されたコーパスとして保存される。

また、学習者や教師がこのコーパスを閲覧できるように、また必要な情報だけ閲覧できるように JavaScript と XSLT スタイルシートを使用する(図2)。このことによって『誤用例による否定的証拠の提示』になると考えた。

国籍	コメント	学習歴	誤り語
JPN	よし子さんもお母さんは電話に切られてしまいました。	年々々	「も」
MAS	よし子さんはお母さんにお電話を切られてしまいました。	1年2ヶ月	「お」
JPN	よし子さんはお母さんに手紙を読まれてしまいました。	年々々	「手紙」
JPN	良子さんの電話はお母さんに切られました。	年々々	「良子」
MAS	お母さんは電話を切ります。	3年々々	「切」
MAS	の電話はお母さんに切られた。	4年々々	「の」
MAS	はお母さんは電話を正しいところにおいでた。	2年々々	「正しいところおい」
MAS	お母さんは電話を正しいところにおいでた。	2年々々	「正しいところおい」
MAS	お母さんは電話をストップしました。	2年々々	「ストップ」
MAS	はお母さんに電話を切られてしまう。	2年々々	「しまう」
MAS	がお母さんに電話を切れてしまいました。	3年々々	「か切れ」
MAS	はお母さんに電話を切ってしまいました。	年々々	「切っ」
MAS	の電話がお母さんに切ってしまいました。	3年々々	「の」が「切っ」
MAS	の電話をお母さん切られました。	3年々々	「の」
MAS	にお母さんが電話を切ってしまいました。	2年々々	「か切っ」
国籍	コメント	学習歴	誤り語

図2. コーパス表示画面

4. 実験・分析結果

実験は主に、理工系を学びに来日している東京近郊に在住するマレーシア人留学生にお願いした。実験は計3回で、それぞれ3名、4名、5名で行った。第1回目の実験で生じた問題点は以下の2点である。

- ・ 動作主を主語にすることにより、被動作主を主語とする受身文の使用を回避された
- ・ 受身文の概念すら理解できていない被験者がいた

このため、第1回目の実験では受け身文に関する誤用例を蓄積することが出来ず、受け身文に関する分析対象から除外することにした。ま

た実験方法に次のような改善を加えた。口頭で受け身の使用を強調し、動詞の受け身形「おこられる、笑われる」等を使用した文を作成してほしいと伝えた。

第2回目で生じた問題点は以下の通りである。

- ・ 受身文を使用しているものの、使用語彙などにばらつきがあり、文法上の分析を困難なものとしている。

したがって受身文に関する誤用例の蓄積はある程度可能であったものの、実験方法に関しては、更に、若干の改善が必要で、問題で使用する単語を提示し、純粋に文法上の誤りだけを収集できるよう改良した。

第3回目の実験ではこれまでの2度の改善により特に問題なく受身文に関する誤用例の蓄積をすることができた。

第2回目、第3回目の実験から得られた誤用における共通点は以下の通りである。

- ・ 通常日本語では主語になりにくい個人の所有物が主語となっている
- ・ テンス(時制: ます「未来」・ました「過去」)が不安定
- ・ 文体の不統一が見られる

これ以外にもマレーシア人留学生の特徴的な誤用例が存在しており、日本語教師にとって誤用コーパスを用いて各国の留学生の誤りを分析できることが分かった。

また、実験中の学習段階での観察では、留学生自身が誤用コーパスを閲覧し、自らの過ちに気がつく場合も多く、学習者にとっても有益なシステムであることが分かった。

5. 研究課題

今後の課題としては短文作成問題の正誤判定の精度を上げることがある。現状では、まず、正解文との比較により正誤の判定を行う。次に、誤り語の正解文と学習者が入力した文の形態素の数により、第1のフィルタを通過したものを正解文との比較により誤り語を抽出している。しかし、この方法では誤文が出たときに、誤り語をコーパスに格納するのが難しくなっている。

また、コーパスを閲覧するページは学習者も教師も同じ閲覧画面を用いているが、必要な情報を引き出せる検索システムを考えなければならない。

(参考文献)

- [1] 杉浦正利, コーパスに基づいた外国語作文支援システム(2002), 「日本語学と言語教育」, 東京大学出版会, pp. 149-172
- [2] 『ビデオ講座 日本語』1 受身 氏家研一企画・構成 東京書籍