

# 学習者モデル構築のための誤用の分析

河合 和久<sup>†</sup> 吉村 弓子<sup>\*</sup>

<sup>†</sup> 豊橋技術科学大学 知識情報工学系

*kawai@tut.ac.jp*

*http://www.ita.tutkie.tut.ac.jp/~kawai/*

<sup>\*</sup> 豊橋技術科学大学 人文・社会工学系

*yumiko@tut.ac.jp*

*http://www.ita.tutkie.tut.ac.jp/~yumiko/*

## 要旨

音声入力インタフェースをもつ日本語文法・語彙 CAI システムの開発を目指し、その基本設計と基礎実験をすすめている。本稿では、同システムの学習者モデルの設計と、同モデル構築に欠かせない日本語学習者の誤りパターンを、誤用例集から自動抽出する試みについて述べる。形態素解析レベルでの誤り処理では、寺村科研の誤用データに見られる誤りの3割を自動抽出できた。

## 1 まえがき

報告者らは、文部省科学研究費補助金特定領域研究(A)「高等教育改革に資するマルチメディアの高度利用に関する研究」の計画研究ウ班「音声言語処理技術と学習者モデルを用いた語学学習システムの研究[中川00]」の研究分担者として、「コンピュータによる音声情報処理技術・言語情報処理技術を応用した外国語学習を支援するシステムの開発」を目的とする「音声入力インタフェースを用いた日本語文法・語彙 CAI システム (以下、本 CAI システムとする) の構築」に関する研究をすすめている。以下では、本 CAI システムの学習者モデルの設

計と、同モデル構築に必要な日本語学習者の誤りパターンを誤用例集から自動抽出する試みについて述べる。

## 2 学習者モデルと CAI

一般に、CAIシステムにおいては、個々の学習者の学習レベルや学習進度に適合した教材を提供するために、図1に示すようなモジュール構成がとられることが多い [Barr82]。

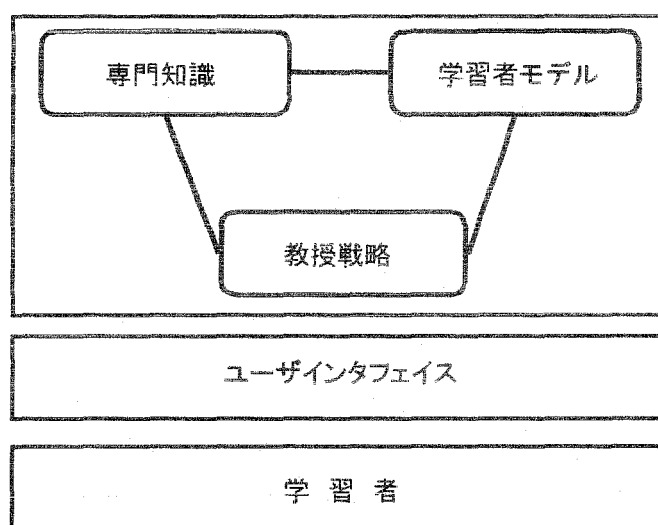


図1: CAIシステムの構成

専門知識モジュールには、学習対象領域に関する「正しい」知識が用意されており、問題を生成したり、学習者の解答の正しさを評価したりする。学習者モデルは、学習者の学習状況、理解状況を表現するもので、通常、専門知識と同様の知識表現形式がとられる。教授戦略モジュールは、教授方法や指導方略に関する知識をもとに、専門知識と学習者モデルとの比較から、その学習者に最適な次の教材を決定するなどの働きをする。

代表的な学習者モデルの生成方式には、オーバーレイモデルとバグモデルがある [Sleeman82]。オーバーレイモデルは、個々の学習者を、理想的な (= 百点満点の) 学習者から一部の知識が欠落したものととらえるもので、学習者モデルは専門知識の部分集合として表現される。これに対し、バグモデルは、専門知識

## 第13回日本語教育連絡会議

を「はみ出る」ような誤り(=バグ)をも取り扱おうとするものである。したがって、バグモデルにおいては、いかに効果的に学習者の誤りを生成し、学習者モデルに反映させるかが、生成される学習者モデルの質を決定することになる。

本CAIシステムは、学習対象領域を日本語文法・語彙としており、専門知識並びに学習者モデルの表現には、その親和性から格文法をベースとしたモデル表現を用いる。格スロットや語彙(単語辞書)には、文法・語彙に関する(宣言的な)知識に加え、学習・教授に関する(手続き的な)知識をもたせる。

また、学習者モデル生成には、バグモデル方式を採用する。上述のように、バグモデルにおいては、いかに誤りを生成し、モデル生成にむすびつけるかが重要となる。このため、本CAIシステム開発の要素技術として、既存の日本語誤用例集から学習者の誤りパターンを自動抽出し、学習者モデル生成のための誤りの生成に利用する基礎実験を行なった。

### 3 誤用例集からの誤りパターンの抽出

寺村科研[寺村90]や大曾科研[大曾99]をはじめとして、日本語学習者の誤用例を集積し、日本語教育の研究・教育の一層の発展に活かそうとする先行研究は数多い。ここでは、寺村科研の誤用例集に見られる誤りを自動検出・自動訂正する実験を行なった<sup>1</sup>。これは、学習者モデル生成のための誤り生成の要素技術ともなるものである。

形態素解析プログラム「茶筌[松本]」に、解析対象文に含まれる誤りに対応できるよう改訂を加え、寺村科研の誤用例文を処理した。寺村科研では、学習者の誤りを64種に分類しているが、そのうち20種ていど(一寺村科研の誤りの分類と解析プログラムの誤りの分類が完全には対応しないため、「ちょうど」20種というようには言えない)の誤りを自動検出できた。また、検出された誤りのうち、およそ6割が自動訂正も行なえた。

<sup>1</sup>本実験の先行研究については、文献[山崎00]に詳しい記述がある。

#### 4 むすび

音声入力インタフェースをもつ日本語文法・語彙 CAI システムの基本設計と学習者モデル構築のための誤用例集からの誤りの自動抽出について述べた。今後は、誤用例の収集・分析をさらにすすめ、バグモデルによる学習者モデル構築のための日本語学習者の誤りパターンの集積をすすめていきたい。

本研究の一部は、科学研究費補助金特定領域研究 (A) 「高等教育改革に資するマルチメディアの高度利用に関する研究」計画研究ウ班「音声言語処理技術と学習者モデルを用いた語学学習システムの研究」ならびに科学研究費補助金基盤研究 (C) 「日本語教育用教材を紹介するキーワードの妥当性に関する研究」の助成を受けている。

#### 参考文献

- [中川 00] 中川聖一：音声情報処理技術を利用した語学 CAI, 情報処理学会フロンティア領域ジョイント研究会 FJK2000 (2000).
- [Barr82] Barr,A. and Feigenbaum,E.A.(Eds.): The Handbook of Artificial Intelligence (1982).
- [Sleeman82] Sleeman,D. and Brown,J.S.: Intelligent Tutoring Systems (1982).
- [寺村 90] 寺村秀夫：外国人学習者の日本語誤用例集, 科学研究費報告書 (1990).
- [大曾 99] 大曾美恵子：日本語学習者の作文コーパス, 科学研究費報告書 (1999).
- [松本] 松本研究室：形態素解析プログラム「茶筌」, <http://cl.aist-nara.ac.jp/lab/nlt/chasen/>.
- [山崎 00] 山崎邦子：外国人学習者の日本語誤用の形態素解析, 豊橋技術科学大学工学修士学位論文 (2000).