

## 日本語教育支援システム「おさる」

6S-7

森川 聡 松本 裕治

奈良先端科学技術大学院大学 情報科学研究科

〒 630-0101 奈良県生駒市高山町 8916-5

{satosi-m,matsu}@is.aist-nara.ac.jp

## 1 はじめに

日本語は教材の種類が英語に比べてかなり少ないので、日本語学習者・日本語教師にとって、学習者のレベル・母国語・興味に合った問題集を探すのはとても大変である。学習者により良い問題を提供し効率良く学習してもらうために、熱心な教師たちは問題を自作することになる。しかし、問題の製作は時間がかかる。そこで我々はこうした状況を改善するために学習者に合った問題を自動生成する web 上で使えるシステム「おさる」を開発することにした。

本稿では「おさる」の自動生成する問題の 1 つである「漢字読み問題」の生成アルゴリズムを紹介する。

## 2 「おさる」とは

「おさる」は日本語教師・日本語学習者のための web 上で動くシステムです。「おさる」が提供する機能は

(1) 個々の学習者に合った日本語練習問題を自動生成する（ただし、生成された問題は教師の最終チェックが必要）

(2) (1) によって生成された問題をデータベース化し、公開（各学習者に合った問題を瞬時に検索してくれる）

「おさる」はユーザ（日本語教師・日本語学習者）の使用履歴を用い、自身が生成する問題の質を自動的に向上させるアルゴリズムを持っている。そのため、利用されればされるほどより質の高い問題を生成するようになる。

## 3 「おさる」の問題のフォーマット

「おさる」は日本語能力試験の問題と同じ選択肢形式の問題を生成する。日本語能力試験の問題は図 1 の様に文 x 1、正解選択肢 x 1、誤選択肢 x 3 で構成されており「おさる」も同じフォーマットを採用している（図 1 参照）。[1]

文: 「おさる」の持つ文データベースから問題になっている単語が含まれている文が選ばれる。

正解選択肢: 問題に使用された文の問題箇所が正解選択肢となる。例の中では「冷凍」「機会」「に」が

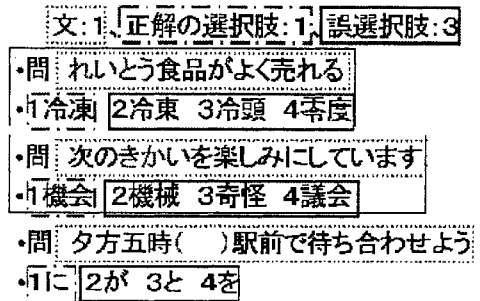


図 1: 「おさる」の問題フォーマット

正解選択肢。

誤選択肢: 正解選択肢からルールを使って自動的に生成される。ただし、生成のルールは問題の種類などによって異なる。

## 4 問題の種類

現段階で自動生成しようと考えているのは以下の 3 つの種類の問題である（図 2 参照）。

漢字選択: ひらがなに当てはまる正しい漢字の単語を選ぶ

漢字読み: 漢字の単語の正しい読みを選ぶ

文法: 文法的に正しい選択肢を選ぶ

## 5 漢字読み問題分類

漢字読み問題であるが、正解と誤選択肢関係から以下の 4 つに分類できる。（誤選択肢の生成以外はどれも同じアルゴリズムで作られる）

(1) 誤選択肢の読みが正解とにている

問 今日 普通 通りの授業だった

普通 1 ふつう（正解） 2 ふつごう 3 ふゆう 4 かい

う

## 【漢字選択】

・問 また会うきかいを楽しみにしています  
 ・1機会 2機械 3奇怪 4議会

## 【漢字読み】

・問 彼のやり方は強引だ  
 ・1ごういん 2きょうひき 3つよいん 4つよひき

## 【文法】

・問 夕方五時( )駅前で待ち合わせよう  
 ・1に 2が 3と 4を

図 2: 問題の種類

(2) 問題となっている漢字単語の各漢字に個別に読みを割り当てる

問 彼のやり方は強引だ

強引 1 ごういん (正解) 2 つよひき 3 きょういん 4 きょうひき

(3) 意味的に近いもの、混同し易いもの

問 自衛隊についての論争はひとまず落ち着いた

論争 1 ろんそう (正解) 2 ろんじゅつ 3 ぎろん 4 かいごう

問 弟が僕の友達のゲームを壊してしまった

弟 1 おとうと (正解) 2 いもうと 3 あに 4 あね

(4) 正解の読みを学習者が混同するような音韻変化させたもの

問 冷凍食品が大安売りだった

冷凍 1 れいとう (正解) 2 れいとお 3 れいどお 4 れいと

(1)-(4)の誤選択肢の生成法について説明する。

## 漢字読み問題 (1)

単語の辞書(茶釜辞書使用)から正解と同じ(あるいは近い)読みを持つ同じ品詞の単語をマッチングで検索する

## 漢字読み問題 (2)

「おさる」はシステム内に漢字の読み辞書(Wnnおよび茶釜の辞書から製作)を持っており、漢字読み問題(2)はこれを利用して生成される。読み辞書の中身は以下のようにになっている。<読み辞書の中身>

漢字 単独読み 送りがな付 接頭 接尾

切 せつ き せつ -

腹 はら・ふく - - ぶく

「切腹」の誤選択肢を例にとりて説明する。

<各字ごとに読みを検索>

切(単語の先頭なので単語読み・接頭を検索)

=> せつ・せつ

腹(単語の末尾なので単語読み・接尾を検索)

=> はら・ふく・ぶく

<各読みの全ての組合せをつくる(正解は除く)>

せつはら・せつふく・せつぶく・せつはら・せつふく

## 漢字読み問題 (3)

シソーラスにおいて同じカテゴリに入っている単語、類義語、同義語、反対語を誤選択肢として使用

## 漢字読み問題 (4)

正解の読みを学習者が混同しやすい読みに変換して誤選択肢とする。

(清濁)か/がた/だ

(長音短音)おばあさん/おばさん

(促音)切って/来て etc

## 6 今後の予定

10月初めにはweb上で「おさる」試作版を公開する予定である。現在は「漢字読み問題」に取り組んでいるが、今後は「漢字選択問題」、「文法問題」、「複合動詞問題」、「単語意味問題」などに取り組む、「おさる」が生成する問題の種類を増やしていく予定である。

## 7 日本語教育界の皆様へ

「おさる」をより良いシステムにするためには、ユーザである日本語教師・日本語学習者からの意見・要望・感想が重要です。10月までには公開しますので、実際に「おさる」を使って下さい。なお、「おさる」の詳細は([http://cl.aist-nara.ac.jp/satosi-m/osaru/osaru\\_top.html](http://cl.aist-nara.ac.jp/satosi-m/osaru/osaru_top.html))を見て頂きたい。意見・要望・感想は [satosi-m@is.aist-nara.ac.jp](mailto:satosi-m@is.aist-nara.ac.jp) までお願いします。

## 8 参考文献

- 1 日本語ジャーナル
- 2 日本語概説 加藤彰彦 佐治圭三 森田良行
- 3 日本語読解学習支援システム リーディングチュウ太 <http://language.tiu.ac.jp/>
- 4 東京工業大学留学生センター 日本語学習支援システム <http://www.ryu.titech.ac.jp/system/home.html>