

## 日本語学習支援システム「おさる」における漢字読み問題生成方法

Bhooshan Raj Neupane, Satoshi Morikawa, Yuji Matsumoto

(ネウパネ ブッシュハン ラザ、森川 聡、松本 裕治)

Graduate School of Information Science, Nara Institute of Science and Technology

E-mail: {neupa-ra, satosi-m, matsu}@is.aist-nara.ac.jp

### Abstract

漢字は文脈に応じて異なる読みをするため、日本語習得上の障害の一つになっている。日本語学習支援システム「おさる」では、漢字読み問題を自動的に生成し学習を支援することを一つの目的としている。本稿では漢字読み問題生成方法について述べる。

### 1. Introduction

Our computer aided language-learning project OSARU is a web-based Japanese language acquisition support system which generates various types of exercises that can be used by teachers of Japanese language as well as learners.

The primary goal of OSARU is to create an environment for effective learning by offering automatic exercise generation and automatic exercise grading systems. There are three types of multiple-choice questions that OSARU creates, namely kanji reading questions, kanji selecting questions, and grammar questions.

In this paper, we concentrate on the kanji pronunciation decoy generation procedure, which generates the decoys (i.e. the foil alternatives to the correct answer) to be used in the multiple-choice questions.

### 2. Difficulties faced in learning kanjis

Japanese writing is a combination of three character sets: two sets are comprised of phonetic characters (hiragana and

katakana), and the third set is made of ideograms of Chinese origin (kanji). Learning kanji is perhaps one of the most difficult parts of learning the Japanese language. Some of the requirements in learning kanjis can be summarized as follows.

#### 1. Individual character information

- a. Memorizing the multiple pronunciations of each of the 2000 kanjis<sup>1</sup> in modern Japanese.
- b. Memorizing the different on-yomis (pronunciations of Chinese origin) and kun-yomis (pronunciations of Japanese origin).

#### 2. Character concatenation information

- a. The pronunciation of a kanji in a word is determined by the other characters in the word. (e.g. on-yomis and kun-yomis cannot be

---

<sup>1</sup> 2000 kanjis are listed on Gakken's "A new dictionary of kanji usage" as the kanjis needed for modern Japanese. To be precise, The Japanese government's official list of recommended characters, the 'Jouyou' kanji list consists of 1945 characters.

mixed together in the same word).

- b. Phonetic modifications in order to comply with phonotactic constraints. (e.g. in rendakau, voiceless to voiced alternations occur in some contexts,<sup>1</sup> as in, hon+ tana → hondana “bookcase”).

OSARU provides multiple-choice questions for kanji pronunciation, paying special attention to the points above.

### 3. The decoy generation procedure

When creating multiple-choice questions, one critical point is how to generate the decoys, as they determine many of the dimensions along which the learner’s knowledge is tested.

There are two modules used during the decoy generation process. They are the databases and the tools. The databases used are the word database (abbreviated as word DB) and the decoy database (decoy DB). In the word DB, words are classified according to their difficulty level. The tools used include a parser, dictionaries and decoy generator modules. The decoy generation procedure is divided into three sections where three different modules are used to generate the decoys based on characteristics of the kanjis. They are as follows.

- a. Nitayomi: the similar pronunciation decoy generating module;
- b. Betubetuyomi: the module using other pronunciations of the same kanji to generate decoys;
- c. Synonym: the module generating decoys from related words.

### 3.1 The Nitayomi module

Nitayomi is the module that generates decoys with pronunciation pattern similar to the target pronunciation. The process used to generate the decoys is as follows.

The word to be practiced is selected from the word DB and is passed through a parser to obtain its pronunciation. For example, if the word selected is “普通”, the parser outputs the pronunciation as “ふつう”. In the next step, the pronunciation “ふつう” is passed to the decoy generator, where the decoys are generated (e.g. “ふつつ”, “ふつ”, “ふす”) using the Nitayomi module.

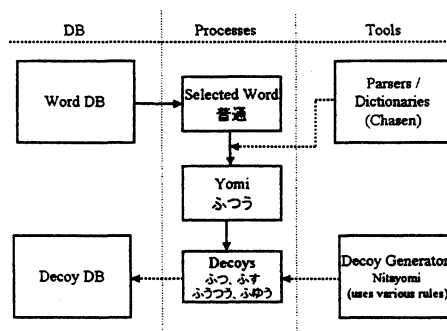


Fig 1. Generation of Nitayomi decoys

Decoys generated by the Nitayomi module are used to practice the pronunciation of the kanji itself, and its on-yomi and kun-yomi pronunciations. <sup>2</sup>The rules used to generate the decoys are as follows.

- a. omit / insert a mora (e.g. “つ”, “ん” etc);
- b. change of ん+あ to い and vice

<sup>2</sup> The common mistakes made by learners of kanji are analyzed on chapter 6 of 教師用日本語ハンドブック (ref iv), however the rule above are developed by the authors themselves.

- versa;
- c. change of ん to っ and vice versa;
- d. change long vowels into short vowels and vice versa;
- e. change voiced consonants to voiceless consonants, and vice versa;
- f. change big “ゃ”, “ゆ”, “よ” to small ones and vice versa;
- g. replace “ひ”, “し” sounds with each other.

The decoys thus generated are returned to the decoy DB. The decoy generation process is shown in Figure 1.

### 3.2 The Betubetuyomi module

The Betubetuyomi is the module that generates decoys that can be used to practice changes in pronunciation according to the context of the word. The process is as follows.

	Independent Pronunciation (単独)	Pronunciation With Hiragana or Katakana (送り仮名)	First Syllable on the Kanji word (接頭)	Last Syllable on the Kanji word (接尾)
強 (a)	きょう ごう	つよ し こわ	つよ こわ つわ	ごわ づよ
引 (b)	いん ひき	ひ	ひっか	びき っびき

- Combination Rule
- a. Independent-a + Independent-b
  - b. Independent-a + Last Syllable-b
  - c. First Syllable-a + Independent-b
  - d. First Syllable-a + Last Syllable-b

Fig 2. Kanji pronunciations & their combination rules.

A word is selected from the word DB and is divided into the kanjis constituting it. The pronunciations of each kanji are obtained from “Yomi dictionary”. Four different

pronunciations namely, tandoku: individual kanji pronunciation, okurigana: pronunciation when used with hiragana or katakana, setto: pronunciation when used as a first syllable of a word, and setsubi: pronunciation when used as a last syllable of the word, are listed in the Yomi dictionary. An example of the pronunciation listing is shown in Figure 2. The pronunciations are combined with each other to generate the Betubetuyomi decoys. For example, if the word chosen is “強引”, then it is broken into “強” and “引”. All the pronunciations for each kanji are derived.<sup>3</sup> The pronunciations are combined following the rules as follows.

- a. Independent-a + Independent-b
- b. Independent-a + Last syllable-b
- c. First syllable-a + Independent-b
- d. First syllable-a + Last syllable-b

Here “a” and “b” are used to denote the first and last syllable of a kanji word.

The decoys are created as “つよひき”, “つよいん”, “きょうひき”, “ごうびき” etc. The decoys thus generated are returned to the decoy DB. The decoy generation process is shown in Figure 3.

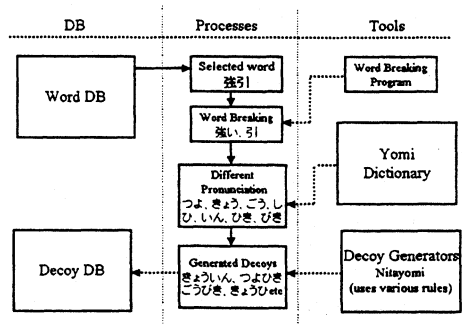


Fig 3. Generation of Betubetuyomi decoys

<sup>3</sup> Refer to Figure 2.

### 3.3 Synonym decoy generating module

Synonym is the module that is used to generate decoys by selecting words with related meanings. The procedure is as follows.

The word is selected from the word DB, and is passed to the synonym locator (a thesaurus), which derives related words. From those related words, only those that share at least one kanji with the original word are selected. The words thus obtained are passed to the parser, which derives the correct pronunciations for those words. The pronunciations are collected at the decoy DB. For example, if a word selected is “機会”, the related words produced are “議会”, “会議”, “集会”<sup>ii, iii</sup>. After passing them to the parser the final decoys generated are “ざかい”, “かいぎ”, “しゅうかい”. The decoy generation process is shown in Figure 4.

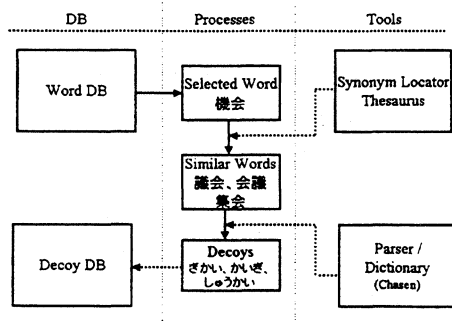


Fig 4. Generation of synonym decoys

#### Examples of questions

The following are some examples of the questions that OSARU targets to create as the kanji reading questions, by making use of the decoys generated from the above procedures.

以下の罫線引かれている単語の読み仮名を選択せよ。

1. 彼はいつも普通の格好をしている。
  - a. ふつ
  - b. ふうつう
  - c. ぶつう
  - d. ふんすう
2. 彼は強引なセールスマンだ。
  - a. ごういん
  - b. きょういん
  - c. ごいん
  - d. つよひき
3. この機会を逃してはいけない。
  - a. きかい
  - b. ぎかい
  - c. きっかい
  - d. さいかい

#### 4. Discussion and future work

Although the current system is able to generate the decoys, the question generation process is still to be tested. The rules applied to generate the decoys are based on research made on the characteristics of the kanjis<sup>iv</sup> and assumptions made by the authors themselves. For testing purposes, we are using a copy-protected thesaurus to generate the related words. However, to make the system available online we need create our own thesaurus or find one that is publicly available. There is a strong need to collect learners data to verify their characteristics and compare them with the ones that we are presently assuming.

#### References

- <sup>i</sup> Natsuko Tsujimura, *An Introduction to Japanese Linguistics*, Blackwell Publishers, pp.54-63, 1996
- <sup>ii</sup> NTT 基礎科学研究所ニホンゴ語彙体系、単語体系、岩波書店, pp. 457, 2000
- <sup>iii</sup> NTT 基礎科学研究所ニホンゴ語彙体系、語彙体系、岩波書店, pp. 313, 2000
- <sup>iv</sup> The Japan foundation Japanese Language Institute, 教師用日本語ハンドブック、発音、凡人社、pp.128-188, 1989