

単語難易度判定機能を有するエディタの開発

Development of an editing system with a word-level evaluation function

*柴田大介・**村田進・***北村達也・****川村よし子

*SHIBATA Daisuke・**MURATA Susumu・***KITAMURA Tatsuya・****KAWAMURA Yoshiko

&*甲南大学・****東京国際大学

*,**&***Konan University・****Tokyo International University

〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail *ganimataangel@gmail.com・**murata139@gmail.com・***t-kitamu@konan-u.ac.jp・
****kawamura@tiu.ac.jp

Abstract : This study introduces an editing system for writing Japanese language texts with vocabulary that is more accessible to NNS of Japanese. The editor system will automatically identify the level of word difficulty by color coding words based on the coding scheme used in the Vocabulary Checker of Reading Tutor developed by Kawamura, et al. Word-level evaluation function is invoked by clicking on “analyze word difficulty” which appears at the bottom of the editor page. The editor is written in Java programming language, which makes it compatible with any type of operating system.

キーワード：単語難易度，旧日本語能力試験出題基準，形態素解析，単語色分け，エディタ

1 はじめに

我々は日本語非母語話者にとって難しい単語や構文を検出し、彼らにとってわかりやすい文を書くことを支援する技術を研究している。これまで、入力文中の単語レベルを判定する語彙チェッカー（川村，1998），中止法（テ形および連用形）や名詞修飾を抽出するシステム（水嶋ら，2011），やさしい日本語への書き換えシステム（川村・北村，2012），主格省略を抽出するシステム（中村ら，2012）などを開発してきた。

しかし、文章を書く目的でこれらのシステムを利用すると使い勝手の悪い部分があった。例えば、語彙チェッカーでは、Web ブラウザ上で文を入力してボタンをクリックするとページが切り替わり結果が表示される。従って、文を修正するには一旦前のページに戻る必要があった。

そこで、本研究では、旧日本語能力試験出題基準に基づいて文中の単語を色分けする機能を組み込んだエディタを開発した。このエディタでは、利用者が単語リストを入れ替えることによって、独自の判定基準に基づく単語色分けが可能である。

2 エディタの仕組み

2.1 概要

エディタとは Microsoft Windows のメモ帳のような文章編集のためのソフトウェアをいう。本研究で開発したエディタのスクリーンショットを図 1 に示す。このソフトウェアは、ファ

イルのオープン、保存、テキストのカット、コピー、ペーストなどエディタとしてのごく基本的な機能をもっている。

そして、エディタ下部にある「難易度判定」ボタンをクリックすると、その段階で入力されている単語の難易度を判定し、レベルに応じて色分け表示する。単語の色分けは、1 級から 4 級および級外に対して行う。単語の色分けと同時に、エディタ下部に各レベルの単語数を表示する。利用者は単語が色分けされた状態で引き続き編集作業を行うことができる。これにより、語彙チェッカーを利用する場合と比較して、ページを戻す手間が省ける。

単語の難易度判定は、旧日本語能力試験出題基準に基づいて行われる。この他、利用者が独自に作成した単語リストを登録することによって、それに基づいた単語色分けも可能である。例えば、川村ら（2011）によって作成された介護福祉士のための単語リスト「808」を用いれば、これに含まれる（もしくは含まれない）単語を色分けすることができる。

編集後の文章は、単語色分けの状態を保持したまま Rich Text Format (RTF) で保存することができる。また、単語色分けの情報を破棄してテキスト形式で保存することも可能である。RTF は、Microsoft Word など様々なソフトウェアで開くことができる。

このエディタはプログラミング言語 Java を用いて開発されているため、Windows のみならず、MacOS や Linux などの OS 上でもプログラムの変更なしで動作させることができる。

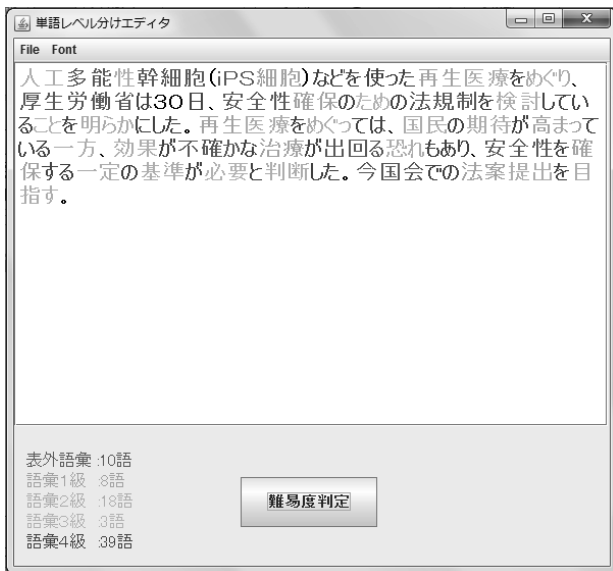


図 1. 本研究で開発したエディタのスクリーンショット

2.2 単語難易度判定機能の仕組み

単語の難易度を判定しエディタに表示するまでの過程は、以下の手順を踏む。

- (1) 入力文章の形態素解析
- (2) 単語リストの作成
- (3) 解析結果と単語リストの比較、レベル判定
- (4) 単語の色分け表示

本研究で使用した形態素解析ツールは、ヤフー株式会社が無料で公開している Yahoo!デベロッパーネットワークの日本語形態素解析 API である。対象文章と解析方法を指定した URL にアクセスすることによって、解析結果を XML 形式で取得できる。取得した解析結果の例を図 2 に示す。解析された文章は形態素ごとに分割され、形態素の表記、読み仮名、品詞、基本形などの情報が得られる。

そして、形態素の基本形が単語リストの中に存在すれば、対応する級を当該の形態素のレベルとする。対応する単語がない場合は級外とする。単語リストは CSV 形式のファイルから読み込むので、選択するファイルを変更すれば単語リストを変えることができる。

そして形態素をレベルごとに指定した色に変更し、エディタ画面に表示する。本研究で色分けに用いたカラーデザインは、カラーユニバーサルデザイン推奨配色セット (伊藤哲, 2009) を参考に設定した。これによって、色弱者にも比較の見やすいエディタの実現を図った。さらに、エディタ下部に各級の単語の出現数を表示することにより、文章の難易度を判断しやすいようにした。

```
<filtered_count>4</filtered_count>
<word_list>
  <word>
    <surface>犬</surface>
    <reading>いぬ</reading>
    <pos>名詞</pos>
    <baseform>犬</baseform>
    <feature>名詞,名詞,*,犬,いぬ,犬</feature>
  </word>
  <word>
    <surface>が</surface>
    <reading>が</reading>
    <pos>助詞</pos>
    <baseform>が</baseform>
    <feature>助詞,格助詞,*,が,が,が</feature>
  </word>
  <word>
    <surface>好き</surface>
    <reading>すき</reading>
    <pos>形容動詞</pos>
    <baseform>好き</baseform>
    <feature>形容動詞,形動,*,好き,すき,好き</feature>
  </word>
</word_list>
<ma_result>
```

図 2. 「犬が好き」の形態素解析の結果

4 おわりに

本研究では旧日本語能力試験出題基準に基づいて単語を色分けする機能をもつエディタを開発した。このエディタはプログラミング言語 Java で開発されているため、OS によらず動作する。ただし、旧日本語能力試験出題基準は国際交流基金と日本国際教育支援協会が著作権を有するため、現段階でこのエディタを配布することは困難である。

また、現在は「難易度判定」ボタンを押さなければ単語色分けが更新されない。しかし、将来的には利用者が作業を中断したときに自動的に単語色分けを更新できるようにしたい。

謝辞 本研究の一部は平成 24 年度科学研究費課題番号 24320096 及び平成 24 年度私立大学等経常費補助金の支援を得て行われた。

参考文献

- 川村よし子 (1998) 「語彙チェッカーを用いた読解テキストの分析」『講座日本語教育』第 34 分冊, 1-22.
- 水嶋博志・内田聖也・北村達也・川村よし子 (2011) 「学習者にとって難解な構文の自動検出」『日本語教育方法研究会誌』18(1), 64-65.
- 川村よし子・北村達也 (2012) 「やさしい日本語への書き換えリストの作成とその評価」『日本語教育学会秋季大会予稿集』123-128.
- 中村慶太・北村達也・川村よし子 (2012) 「日本語文章における主格省略の自動検出」『甲南大学紀要知能情報学編』5(2), 129-142.
- 川村よし子・野村愛・名藤杏子・金庭久美子・斉木美紀・北村達也 (2011) 「介護福祉士候補生のための国家試験に向けた教材の開発」『日本教育工学会第 27 回全国大会講演論文集』625-626.
- 伊藤哲 (2009) 「色覚の多様性に配慮したカラーユニバーサルデザイン推奨配色セット作成」『印刷会』669, 82-88.