

## 科学研究費助成事業 研究成果報告書

平成 28 年 5 月 27 日現在

機関番号：12102

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25370573

研究課題名(和文) 読解教育支援を目的とする文章難易度判別システムの開発

研究課題名(英文) Development of a text readability assessment system to support reading activities in Japanese language education

研究代表者

李 在鎬 (LEE, Jaeho)

筑波大学・人文社会系・准教授

研究者番号：20450695

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：本研究は、日本語研究におけるリーダビリティ研究を発展させ、日本語教育のためのリーダビリティの構築を目指すものである。具体的には、日本語教科書コーパスを利用し、6スケールの難易度判別のための公式を開発した。日本語教育のリーダビリティ公式として「 $X = \{\text{平均文長} * -0.056\} + \{\text{漢語率} * -0.126\} + \{\text{和語率} * -0.042\} + \{\text{動詞率} * -0.145\} + \{\text{助詞率} * -0.044\} + 11.724$ 」(R2=.896)を提案する。そして、このリーダビリティ公式を実際の日本語テキストにあてはめ、有効性を検証した。さらに、リーダビリティ公式を実装したウェブシステムを構築した。

研究成果の概要(英文)：The research presented in this paper aims at advancing text readability studies for Japanese and devises a practical and useful system that contributes to Japanese language teaching, learning, and research. More specifically, utilizing leveled corpora mainly consisting of text from textbooks of Japanese, we produced the following formula to measure the readability level of a given text in six scales:  $X = \{\text{mean length of sentence} * -0.056\} + \{\text{proportion of kango} * -0.126\} + \{\text{proportion of wago} * -0.042\} + \{\text{proportion of number of verbs among all words} * -0.145\} + \{\text{proportion of number of auxiliary verbs} * -0.044\} + 11.724$  (R2=.896). The formula was tested against another set of leveled text in Japanese to prove its reliability. Lastly, the method was implemented in a computer system that calculates and produces the estimated level of a text via a web-based online interface.

研究分野：日本語教育

キーワード：読解教育 教育工学 文章の難易度 リーダビリティ

## 1. 研究開始当初の背景

近年、コミュニケーション能力の育成を目標とする言語教育の重要性が盛んに指摘されており、それを象徴する動きとして、日本語能力試験も大きく変化した。とりわけ、改定後の日本語能力試験では、真正性を重視したコミュニケーション能力を測定することを目的にテストが設計された。他方、ウェブの爆発的広がりや「現代日本語書き言葉均衡コーパス」に代表される大規模な日本語のデータベースの公開により、生きた日本語テキストが自由に利用できる環境が整いつつある。こうしたテキスト素材を利用することで、現実のコミュニケーションの場において求められる日本語能力の実態を学習者に伝えることができる。

しかし、生の日本語素材は、生の素材であるがために、日本語教育の現場では活用しにくい「粗さ」が存在する。なぜなら、このような生素材は日本語学習者にとっての難易度が明らかでないからである。そのため、どのような文章を、どのレベルの学習者に与えるべきかという問題が生じる。こうした問題に対して、現状では個々の教師の経験に基づく直感に頼るしかないところがあるが、「直感」のような統制できない仕組みによって教育上の素材が選別されることは好ましくないと見える。とりわけ、チーム・ティーチングを行う場合など、標準化されたカリキュラムの下で日本語教育を行う場合、こうした課題は早急に解決する必要がある。

以上の背景を踏まえ、本研究では、ウェブやコーパスのような生教材を日本語教育の読解教育に活かすという目的のもとに、文章の難易度を自動判定するウェブシステムを開発した。

## 2. 研究の目的

本研究の目的は、日本語の読解教育に特化した文章難易度の自動判定システムを開発することである。難易度判定システムには、文章の様々な構成要素（文字種、文の長さ、語彙項目の難しさ）の出現頻度を統計的に処理した計算式を実装しており、主に3つの機能を提供する。1)新規に入力されたテキストの日本語教育上の該当レベルを推定する、2)文章がもつ様々な量的属性をグラフィカルに提示する、3)既存の辞書システムとリンクさせる。1)と2)は教師支援、3)は学習者支援を想定している。

こうしたシステムを活用することで、ウェブの普及やコーパス開発によって得られた大量の生テキスト素材をコピー＆ペーストのみで日本語教育の教室活動に応用できる

ため、コミュニケーション能力を重視した真正性のある日本語教育が実現できる。また、学習者が手持ちのテキストを用いて自学自習を行う際にも、レベルの適切さを確かめたり、未習得の語に対して辞書引きを行ったりといったことができる。そのため、授業の内外で従来よりはるかに広がりのある取り組みが期待できる。

## 3. 研究の方法

3つの作業項目を設定して開発を進めた。

- 1) 計算式構築のための基準テキストデータの収集、
- 2) 統計的な手法を用いたリーダビリティ計算式の作成、
- 3) リーダビリティ計算式のウェブシステムへの実装

1)は読解教育用の日本語教材と旧日本語能力試験の読解テキストで構成する。2)では各テキストに含まれる言語的要素を数値化し、回帰式を作成する。3)では、プログラミング言語「Ruby」を用いてウェブブラウザで利用できるアプリケーションを開発する。研究の初年次には1)のテキストデータの収集を行った。次に2年次に2)と3)を行い、プロトタイプシステムを作成した。そして3年次には解析精度を検証し、再び1)、2)の作業を追加的に行った上で、3)のシステムへ反映させ、高精度なシステムを作成した。

## 4. 研究成果

本研究の一番の成果は、当初計画した文章難易度の測定のためのリーダビリティシステムが完成し、<http://jreadability.net/>での公開後、安定した運用を継続していることである。

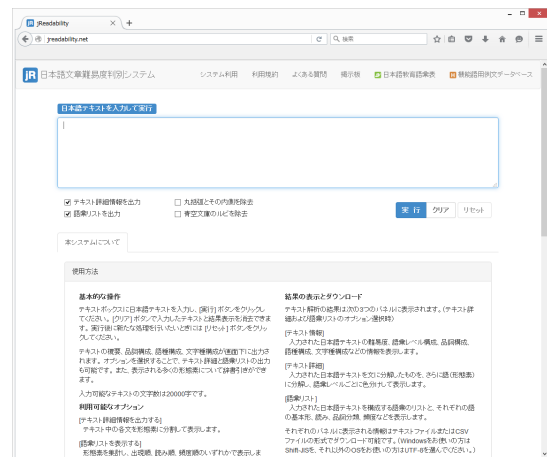


図1. 日本語文章難易度判別システム

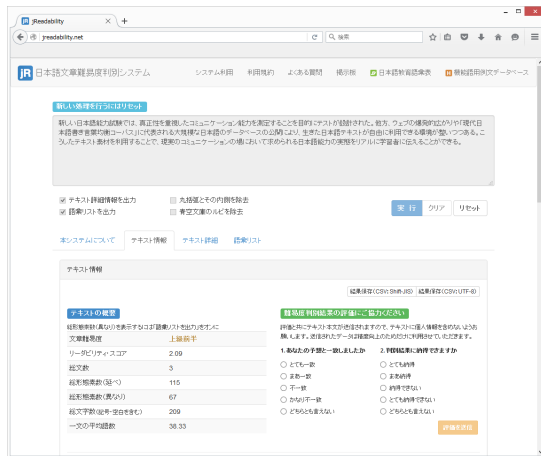


図 2 . 分析結果表示の例

図 1 の日本語文章難易度判別システムに対して、文章を貼り付け、実行を選択すると、図 2 の解析結果が表示される。本システムを使うことで、テキストを貼り付けるだけで文章の難易度や各種計量的な特徴が確認できる。また、本システムは、ウェブ上に実装されているため、いつ、誰が、どこで、使っても同じテキストに対しては同じ結果を返す。そのため、日本語教育の様々な場面で素材を標準化する目的で大いに役立つツールであると言える。

本システムは、2013 年 9 月にアルファ版を公開し、3 年の研究期間を通じて改良を重ねてきた。2016 年 4 月には正式版を公開したが、それまでの利用実績として、合計で 24 万ページビュー、世界 118 カ国からのアクセスが記録されている(図 3)。

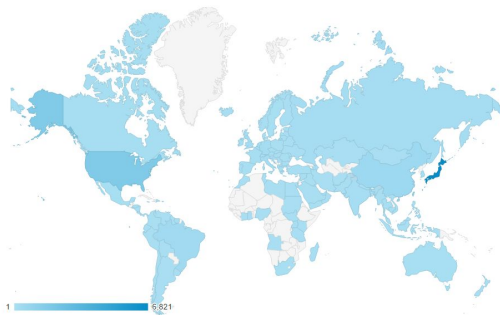


図 3 日本語文章難易度判別システムの利用地域(色の濃さは利用頻度を示す)

本研究では、多くの利用者からのフィードバックを得る目的で、テキストの分析結果画面にアンケート画面を設置し(図 2)、システムの精度やユーザビリティの向上につながる情報を収集した。また、質問掲示板を設置し、システム利用に関する質問やそれらに対する回答を広く公開することに努めた。

このようにウェブシステムであることの利点を生かした開発を進める中で、さらに価値を高める取り組みとして、当初の研究

計画で予定されたものに加え、いくつかの機能が追加された。その 1 つはウェブブラウザ上で入力された日本語テキストに対し合成音声による読み上げを行う機能である(Google Chrome で利用可能)。また、テキストを分析した結果をプレーンテキスト形式もしくは CSV 形式でダウンロードする機能も追加された。これは教師にとって、教材として用いるテキストに対し詳細な分析を行うのに有用な機能である。同様にコーパスを用いた日本語教育研究や教材開発を目指す研究者にとっても有用と考えられる。その他の改良点を含む詳細な開発履歴を以下に示す。

- 1) 2013.09.01 日本語文章難易度判別システム alpha 版公開。
- 2) 2013.10.10 質問掲示板を設置。
- 3) 2013.12.01 テキスト入力インターフェイスの一部変更。
- 4) 2013.12.01 各種データの TXT / CSV ダウンロード機能を実装。
- 5) 2013.12.01 語彙レベルによる色分け表示機能を実装。
- 6) 2013.12.02 文章難易度判別結果の評価送信機能を実装。
- 7) 2013.12.03 使用方法の追記。
- 8) 2014.09.29 Google Chrome での読み上げ機能を実装(試験的)。
- 9) 2015.08.05 音声合成機能の不具合を修正。
- 10) 2015.08.05 日本語文章難易度判別システム beta 版公開。
- 11) 2016.04.20 日本語文章難易度判別システム正式版公開。

### 5 . 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 3 件)

1. LEE, Jae-ho and HASEBE, Yoichi (2016 予定) Readability Measurement for Japanese Text Based on Leveled Corpora. *Papers on Japanese Language from Empirical Perspective*. (掲載確定)(査読有)
2. 李在鎬 (2015) 「コーパス研究が切り開く新しい日本語教育」『BATJ Journal』(英国日本語教育学会)16, 65-78. (査読無)
3. 李在鎬・小林典子・今井新悟・酒井たか子・迫田久美子 (2015). 「テスト分析に基づく「SPOT」と「J-CAT」の比較」『第二言語としての日本語の習得研究』18, 53-69. (査読有)

[学会発表](計 10 件)

1. 李在鎬・木戸光子・加藤あさぎ・小浦方理恵 (2015) 「「良い作文」に対する定量的分析」(日本語教育学会 2015 年秋季大

- 会) 2015.10.11, 沖縄国際大学, 沖縄県, 那覇市)
2. 砂川有里子・李在鎬・長谷部陽一郎・川村よし子 (2015) 「日本語学習者辞書開発プロジェクトについて」(日本語教育学会 2015 年秋季大会)(2015.10.11, 沖縄国際大学, 沖縄県, 那覇市)
  3. 李在鎬・砂川有里子・長谷部陽一郎 (2015) 「日本語教育語彙表と検索システム」(第 19 回ヨーロッパ日本語教育シンポジウム)(2015.8.28, ボルドー, フランス)
  4. HASEBE, Yoichiro and LEE, Jae-Ho (2015) Introducing a Readability Evaluation System for Japanese Language Education. (The 6th International Conference on Computer Assisted Systems For Teaching & Learning Japanese (CASTEL/J)) (2015.8.9, ハワイ, アメリカ)
  5. 李在鎬・長谷部陽一郎・嵐洋子 (2014) 「日本語作文の熟達度を評価する指標の抽出」(第 25 回第二言語習得研究会 (JASLA))(2014.12.14, 筑波大学, 茨城県, つくば市)
  6. 李在鎬・長谷部陽一郎 (2014) 「日本語の文章難易度を測定する」(第 27 回日本語教育連絡会議口頭発表)(2014.8.24, パラトンケネシェ, ハンガリー)
  7. 李在鎬・長谷部陽一郎 (2014) 「日本語の読解クラスを支援する」 「jreadability.net」(日本語教育学会 2014 年春大会デモンストレーション) (2014.6.1, 創価大学, 東京都, 八王子市)
  8. 長谷部陽一郎・李在鎬 (2014) 「日本語教育のための文章難易度測定の試み」 (第 8 回 日本語実用言語学国際会議) (2014.3.22, 国立国語研究所, 東京都, 立川市)
  9. 李在鎬・今井新悟・加納千恵子 (2013) 「日本語教育にとって必要な遠隔教育とは: 筑波大学の取り組みについて」(第 2 言語習得研究会(関東)第 85 回研究会) (2013.6.15, お茶の水女子大学, 東京都, 文京区)
  10. 李在鎬 (2013) 「形容動詞の「ナ」共起と「ノ」共起のコーパス基盤調査」(日本言語学会 2013 年度春季大会) (2013.6.16, 茨城大学, 茨城県, 水戸市)

〔図書〕(計 4 件)

1. LEE, Jae-ho and NAKAGAWA, Natsuko (2016). KY Corpus. In MINAMI, Masahiko (Ed.), *Handbook of Japanese Applied Linguistics*. De Gruyter Mouton, 283-312. (総ページ数: 535)(李の担当部分: 全体で議論しながら作成したため厳密に分けることが難しい)

2. 李在鎬編 (2015) 『日本語教育のための言語テストガイドブック』くろしお出版。(総ページ数: 248)
3. 浅尾仁彦・李在鎬 (2013) 『言語研究のためのプログラミング入門』開拓社(総ページ数: 275)(李の担当部分: 全体で議論しながら作成したため厳密に分けることが難しい)
4. 李在鎬・村尾治彦・浅尾仁彦・奥垣内健(著) (2013) 『認知日本語学講座 認知音韻・形態論』くろしお出版(総ページ数: 216)(李の担当部分: 141-192)

〔その他〕

日本語文章難易度判別システム  
<http://jreadability.net/>

6. 研究組織

(1) 研究代表者

李 在鎬 (LEE, Jaeho)  
 筑波大学・人文社会系・准教授  
 研究者番号: 20450695

(2) 研究分担者

長谷部 陽一郎 (HASEBE, Yoichiro)  
 同志社大学・グローバル・コミュニケーション学部・准教授  
 研究者番号: 90353135

久保 圭 (KUBO, Kei)

大阪大学・日本語日本文化教育センター・非常勤講師  
 研究者番号: 10748685  
 (平成 27 年より研究分担者)