

Readability 式を用いたオンライン外国語学習環境の構築
Development of Online Environment Taking Advantage of Readability Formula
for Studying Foreign Languages

影山 功*
Tsutomu Kageyama

宮崎 佳典**
Yoshinori Miyazaki

長谷川 由美***
Yumi Hasegawa

Readability はテキストの可読性を数値化したものであり、英文書の客観的な読みやすさの指標として現在までに多くの Readability 公式が欧米で開発されてきた。しかし、それらは大衆向けに考案されており、個人レベルで見ると必ずしも適した値を算出するとは限らない。そこで学習履歴データより導出された個人用の Readability 式を用いて言語学習を効率的に行う環境を構築することを目的とする。試作システムを用いて海外の日本語学習者を対象に実験を行ったところ前後の語学力試験結果に有意差が認められた。現在、英語版、日本語版を開発中であるが、将来的には他言語についても取り組む予定である。

キーワード

Readability, 外国語教育, オンライン学習

1. はじめに

Readability とはテキストの可読性を数値化したものであり、それを算出するための公式を Readability 公式という。Readability は 1920 年代から欧米で盛んに研究が行われており、現在までに多くの Readability 公式が開発されてきた。この Readability 公式は教師が生徒に適した教材の選定をする際や、読者層に合わせた新聞記事を構成する際に用いられている。

以下に英語の Readability 公式の例を示す。ここでは Microsoft Word にも組み込まれている Flesch (Reading Ease)[1] を紹介する。

$$FRE = 206.8 - 84.6 * A - 1.05 * B$$

A は一語あたりの平均音節数、 B は一文中の単語数を示し、算出された値が大きいかほどそのテキストは読みやすく、小さいほど読みにくいと解釈される。この公式は一例であり、他の公式には出現単語の難語率（用意された単語リストに含まれない単語が出現する割合）や単語の音節数などをパラメータとして用いているものもある。

上述したように、英語の Readability 公式は多く存在するが、日本語の Readability 公式の数は少ない。その原因として、日本語文では単語ごとの区切りが明確でない点や、漢字、ひらがな、カタカナといった多様な文字の種類があることが考えられる。以下に日本語 Readability 公式の例を示す。ここでは単語の区切り問題とは独立に用いることが可能な RGV 公式[2] を例に挙げる。

$$RGV = -0.17 * ph - 0.28 * pk - 3.49 * pe + 27.92$$

ph はひらがなの出現頻度、 pk はカタカナの出現頻度、 pe は句読点の出現頻度を示し、算出された値は学年レベルを基準とした可読性の測度となる。

これらの公式により、Readability の値は算出されるが、Readability 「公式」はその名の通り、大衆向けに作成されており、一人ひとりに必ずしも適しているとは限らない。これに対し、個人レベルの Readability 「式」を生成し、リーディングするテキストの選定に用いることが出来れば、効率よく語学学習を行うことができると考えられる。

試作システム[3]はローカルでのみ動作し、プログラムの更新や履歴データの収集等の面で労力を要した。そのような点を解消し汎用性を向上させるため、オンライン上で動作する環境の構築を行っていく。本研究では、学習履歴データより導出された個人レベルでの Readability 式を用いて語学学習を行うことのできる環境の構築を目的とする。なお、本論文では日本語学習者を対象としたシステムの構築を行っている。

本論文の構成は、まず第 2 節で先行研究を紹介する。次に第 3 節では構築したシステムの概観を述べ、第 4 節で内部アルゴリズムについて説明する。最後に第 5 節にて今後の展望を述べる。

2. 先行研究

L.Uitdenbogerd[4]は、Web 上にリーディングに適した十分なテキストが存在することを示している。Readability 公式を EFL/ESL(*1)教材と Web 上のテキストに適用し、それぞれで算出された値を比較した。その結果、EFL/ESL 教材については、出版社の公表している難易度と Readability 公式により算出された難易度があまり似ていないこと、Web 上のテキストは EFL テキストの Readability の値の取りうる

*静岡大学大学院情報学研究科

**静岡大学情報学部

***大韓民国大真大学

*1 EFL : English as a Foreign Language , ESL : English as a Second Language

範囲にあるものが多く見られ、学習教材として用いることのできるものが多く存在することが示されている。

宮崎ら[5]は個人レベルで英語 Readability 式を作成する環境を構築し、ユーザのレベルに適した Web テキストを抽出するシステムを開発した。既存の英語 Readability 公式と各ユーザ用に半自動的に生成される Readability 式について、後者がよりユーザが Rating(3.2 節参照)したテキスト難易度と高い相関が見られたことが示されている。

日本語学習支援システムは代表的なものとして川村ら[6]による日本語読解支援システムの「リーディングチュウ太[†]」や阿辺川ら[7]による多言語対応日本語読解学習支援システムの「あすなろ[‡]」がある。これらはそれぞれ Web 上で公開されている。リーディングチュウ太はインターネット上の情報を学習者自らが選んで学ぶシステムである。コンテンツには日本語学習者のためのリンク集、学習支援ツール、教材バンクがあり、学習者自らのニーズとレベルに合わせて学ぶための学習環境を提供している。学習者からの評価も高く、本システムを利用した実習授業後にアンケートをとったところ、ほぼ全員から「非常に役に立つ」「役に立つ」といった肯定的な評価があり、自立学習を支援するためのシステムとして役立っている。しかし、教材バンクにて提供されているテキストの難易度は日本語能力試験 3 級及び 4 級の単語がテキストの全体に占める割合で計算しており、学習者に適したレベルのテキストを用いて学習ができるとは限らない。あすなろの特徴として入力された日本語テキストに対し、文節ごとの係り受け関係を表示する機能がある。文節ご

とに係り受け構造を表示することで、学習者に文の理解を促進させ、読解力を高めることを目的としている。4 種の係り受け構造の表示法について、その有効性を評価する実験を行ったところ、それらの間にはある程度の正答率の差異が見られた。また、これらのシステムは両者共に、単語の意味を表示させる際に学習者の母語で意味を理解できるように、言語の選択が可能となっている。

宮崎ら[3]は個人レベルで日本語の Readability 式を開発し、それを用いてリーディング学習を行うことのできる環境を構築した。本システムを用いて大韓民国大真大学の日本語学習者に対して、個人レベルでの日本語の Readability 式が有効かどうかを調べる実験を行った。個人 Readability 式を用いてリーディングを行うグループと用いないでリーディングを行うグループの 2 つに分け、実験前後に語学力試験を行ったところ、両者間に 5%水準で有意差が認められ、個人毎に生成された Readability 式が学習に有用であることが示された。

3. 学習システムの概要

本章では今回構築したシステムについての解説を行う。まず始めに操作項目などの学習者用のインタフェースについて示す。次に学習履歴を参照できる教師用のインタフェースを示す。最後に、学習支援のために実装を行ったルビ振り機能について解説を行う。

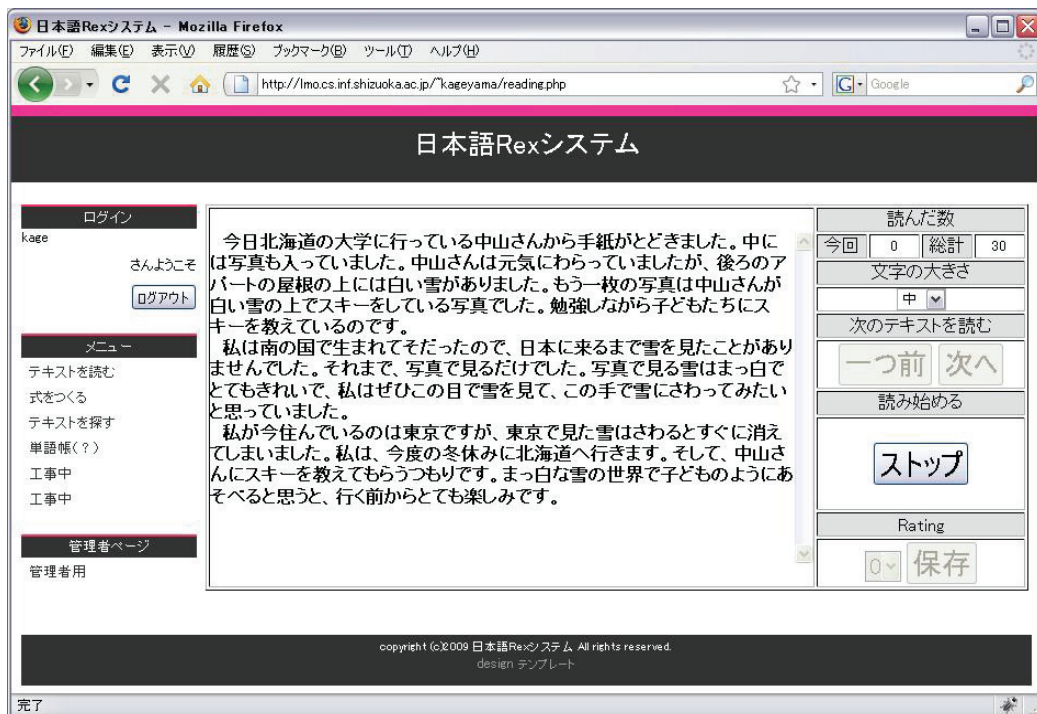


図 1：日本語学習システムのインタフェース

[†] <http://language.tiu.ac.jp/>

[‡] <http://hinoki.ryu.titech.ac.jp/>

3.1 開発環境

試作システムは学習者の PC 端末内でのみ動作しており、プログラムの更新やデータ分析のためのデータ収集などを行う際に労力を要した。そこで、データベースによるデータの一元管理、ユーザビリティの向上を行うべく、Web 上で動作するようにプログラムの書き換えを行った (図 1)。なお、本システムでは PHP5.2.6, MySQL5.0.45, JavaScript を用いて開発している。

学習に用いるテキストは、日本語能力試験の問題集[8]など 10 冊の文献より 215 文書を抜粋した。

3.2 学習者用インタフェース

試作システムでは Readability 式を作成や学習モジュール間での移動など、学習者が行わなければならない操作が多くあった。他言語を学習する際に限らず、システムの操作が多いと学習者の混乱を招く可能性がある。そこで、本システムでは学習者の行う操作を極力少なくして学習を行えるよう構築を行った。ユーザは主に図 1 の右側のナビゲーション項目の操作だけでリーディング学習が可能である。以下に各ナビゲーションの項目について解説をする。

・読んだ数

学習者のテキストを読んだ数が表示される。「今回」ではその回を読んだ個数、「総数」では今までに読んだ数が表示される。

・文字の大きさ

テキストの文字の大きさを 5 段階で設定でき、学習者にとって適した文字のサイズでリーディング学習を行うことが可能となる。

・次のテキストを読む

「次へ」ボタンを押すとテキストがセットされ、次項の「読み始める」によりリーディング学習を始めることができる。

・読み始める

「スタート」ボタンを押すとテキストが表示され、学習者は表示されたテキストを読み進めていく。ここでは、学習履歴としてリーディング所要時間を計測している。

・Rating

読んだテキストに対して 0~5 の範囲で Rating を行わせる。Rating とは学習者が感じたそのテキストの難易度の得点付けを指し、0 に近いほど簡単、5 に近いほど難しいとしている。

上記した項目のうち、学習者は「次のテキストを読む」、「読み始める」、「Rating」の 3 つの項目を操作するのみでリーディング学習が可能となる。学習者の行う操作を極力少なくすることでユーザビリティの向上が見込まれる。

3.3 教師用インタフェース

本機能では全学習者の学習履歴データを閲覧することが可能である。学習履歴データとして、全学習者に対して既読テキストに対する Rating の値、作成した Readability 式、テキスト毎の各学習者の Rating 値を見ることが可能である (図 2)。また、テキストの番号 (textid) を選択することで、そのテキストの本文と属性 (単語数、異語数、文の数など) を閲覧することが可能である。本機能により、学習者の学習進行具合やどのテキストが学習者にとって難しい

と感じるか、各テキストについて Rating の値のばらつきがどの程度あるかといった傾向を確認することが可能である。

また、Rating の時系列的な変化をとらえるために図をプロットする機能、学習者が苦手または得意と考えられるテキストの特徴抽出といった機能の実装も考えている。

| | |
|--------|--------|
| 見たいもの | Rating |
| ユーザ名など | kage2 |

結果

| id | textid | rating | ReadTime | RatingTime |
|-------|--------|--------|----------|---------------------|
| kage2 | 148 | 4 | 00:01:27 | 2009-06-26 01:17:33 |
| kage2 | 154 | 4 | 00:01:23 | 2009-06-26 01:19:06 |
| kage2 | 189 | 0 | 00:00:13 | 2009-06-26 01:19:25 |
| kage2 | 198 | 1 | 00:00:46 | 2009-06-26 01:20:18 |

図 2 : 教師用インタフェース

3.4 ルビ振り機能の実装

日本語学習者にとって漢字の読みは難しいものであり、試作システムには備わっていなかった漢字にルビを振る機能を新しく実装した。Readability 式のパラメータには日本語能力試験の難語率を用いており、漢字を読めるかではなく単語の意味を理解しているかを重視しているため、ルビ振り機能を実装した。表示されたテキストの漢字を含む単語の部分にマウスオーバーさせると、その文字が赤くなり、漢字の読みが表示される。図 3 では「身近」という単語にマウスオーバーした際のルビ表示について示している。

本機能は、テキストを形態素解析ソフトの MeCab[9]に通し、それにより出力された読みを使用している。

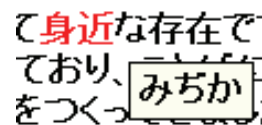


図 3 : ルビ振り機能

4. アルゴリズムの概要

Readability 式は学習者がテキストを読み進めることで自動的に作成される。その Readability 式によって算出された値を用いてテキストの選定が行われる。本節では本システムのコアとなる Readability 式生成法ならびに式生成後のテキスト選出法について述べる。

4.1 Readability 式の作成

Readability 式は学習者がテキストを 10 個読み進める毎に自動的に生成される。初めて Readability 式を作成する学習者にとって、式に用いる変数 (以下パラメータ) の選定は困難であると考えられる。また、学習者に作成を一任してしまうと場合によってはあまり意味を成さない Readability 式を作成してしまう可能性がある。そのような問題点から、Readability 式の作成は自動で行うこととした。Readability 式に用いられるパラメータは「一文中の単語数」、「一単語中の文字数」、「難語率」の 3 つである。これらのパラ

メータは英語の Readability 公式のパラメータとしてよく用いられているため使用した。難語率については日本語能力試験出題単語に対する難語率(*1)とした。また、単語リストには日本語能力試験 1-4 級の単語リストを用いている。日本語能力試験は 1984 年に開始され、受験者数は年々増加しており、国際的に信頼性の高い試験として評価を得ている[10]ことから本単語リストを利用することにした。それらのパラメータを説明変数、Rating の値を目的変数として重回帰方程式を解くことで Readability 式を得ることができる。

4.2 テキストの選定

作成された Readability 式を用いて次に表示するテキストの選定を行う。その手順について示す。まず、Readability 式によって算出された Readability の値から 2.5 を引いたものの絶対値が 0 に近いものから順に並べ、それを 5 等分にする。2.5 としたのは Rating 可能な値を 0~5 で設定しており、学習者が最も読みやすいと判断できる値はその中央値である 2.5 となるからである。その後、5 つに分割されたテキスト群が図 4 に示されている確率で一つ選ばれ、さらにその集合の中からランダムにテキストが表示される。これは、もし学習者に適しているレベルのテキストばかりが表示されてしまうと、惰性で同じ Rating の値をつけてしまう可能性が考えられるためにとった措置である。このテキスト選定のアルゴリズムを用いることで、学習者に適したレベルのテキストが多く表示されるようになると考えられる。また、式が作成される以前(テキストを読んだ数が 10 以下)の場合は全てのテキストの中からランダムに一つ選定される。

5. 今後の展望

今回、オンラインで日本語のリーディング学習が可能なシステムを構築した。個人レベルの Readability 式を用いることで、学習者に適したレベルのテキストが表示され、学習効率が高まると考えられる。

今回は教員を必要とする学習場面を前提として設計している。今後は、Web 上に存在しているテキストを用いてリーディング学習を行い、Readability 式を学習者が自ら作成し、それを用いて「リーディングチュウ太」や「あすなる」で提供されているような教員を必要としない自発的な学習の行うことのできるモジュールの開発を行っていく予定である。

その他にも、本システムの多言語化を行う予定である。ここで示す多言語化とは、「ナビゲーション項目を母国語表示にすること」と「日本語以外の言語について Readability 式を用いたリーディング学習を行えるようにすること」の 2 つを指す。ナビゲーションについては、その項目が外国語表示では初学者にとっては操作に戸惑ってしまい、扱いにくくなると考えられる。そこでナビゲーション項目の母国語表示を可能とし、ユーザビリティの向上を

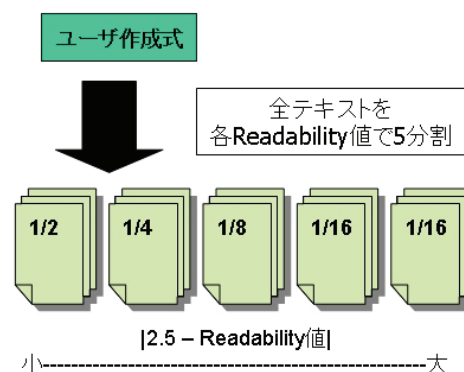


図 4: テキスト選定のアルゴリズム

見込むことができる。また、ボタンなどのコントロールにマウスオーバーした際に、そのボタンの役割を表示させるといった補助機能の実装についても取り組んでいく。他言語の学習については、宮崎ら[5]による英語の Readability 式を開発するシステムとの統合や、韓国語の Readability 式を開発する環境を構築し、英語や日本語と同様の学習を行えるよう拡張をしていく予定である。このように、他言語を本システムで学習したいと考えた時に柔軟にそれに対応できるようにシステムの構築を行っていく。

参考文献

- [1] Flesch R, "A new readability yardstick", Journal of Applied Psychology, Vol.32, pp.221-233 (1948)
- [2] 浅野陽子, 小川克彦, 「日本文の可読性の速度と表示速度への応用」, 情報処理学会論文誌, Vol.32, No.12, pp.1574-1582 (1991)
- [3] 宮崎佳典, 影山功, 周榮斐, 長谷川由美, 個人レベルの日本語 Readability 式導出ならびにそれを応用したネット上日本語学習援用プログラムの開発, 日本語教育学世界大会 2008 (2008)
- [4] Alexandra L.Uitdenbogerd, "Web Readability and Computer-Assisted Language Learning", Proceedings of the 2006 Australasian Language Technology Workshop (2006)
- [5] Yoshinori Miyazaki and Ken Norizuki, "Developing a Computerized Readability Estimation Program with a Web-Searching Function to Match Text Difficulty with Individual Learners' Reading Ability", WorldCALL2008 (2008)
- [6] 川村よし子, インターネット時代に対応した読解教育, 新世紀之日語教学研究国際会議論文 (2000)
- [7] 阿辺川武, 八木豊, 戸次徳久, 澤谷孝志, 奥村学, 仁科喜久子, 杉本茂樹, 傅亮, 日本語学習システム「あすなる」開発の新しい展開—構文学習とその評価—, 情報処理学会第 65 回大会論文集 3T2-6 (2003)
- [8] 松岡龍美, 日本語能力試験 1 級一回で撃破, 時事日本社 (2006)
- [9] MeCab, <http://mecab.sourceforge.net/>
- [10] 日本語能力試験 結果概要, http://www.jees.or.jp/jlpt/jlpt_result.html

*1日本語能力試験級別単語リスト(1-4 級)に含まれない単語の割合を示す