

リーダビリティ指標を用いた文章評価システムの開発：
 計算機と大学生による可読性評価の比較
 Development of Text Evaluation System based on Readability Index for Japanese:
 A Comparison of Indices and Students' Evaluation

赤木 信也[†]
 Shinya Akagi

納富 一宏[†]
 Kazuhiro Notomi

1. はじめに

論文や報告書の作成など、文章作成は大学生・社会人の双方において重要な力である。現在、大学では文章作成技法を学ぶ機会として「文章表現」という講義が設けられ、学生間の相互添削が実施されたり、文章校正・文章推敲支援ツールを使用した講義が実施されたりしている^[1,2]。この「文章表現」の講義では学生への個別指導が求められるが、その分、教員負担の大きさが課題となる。

本研究では、教員負担を軽減し、かつ体系的な文章作成指導につながるシステムの開発を目指す。これまでに文章校正・文章推敲支援ツールを開発して授業で使用するという事例が存在するが、ツールが無償公開されているにもかかわらず実践報告が少ない。この要因の一つには、教員がツールの機能を理解することに加えて、授業での活用方法まで考えなければならず、授業への導入労力が大きいことが挙げられる。そこで本研究では、授業への導入労力の軽減も考慮すべく、ツールではなくシステムという形態での開発を進める。また本研究では、システムの処理対象を文章作成全体のうち文章の読みやすさ・分かりやすさ（リーダビリティ）に絞る。本研究の全体構成は、(1) 日本語リーダビリティ指標の開発、(2) システム構成の研究、(3) システムの有効性の検証、という 3 種類のテーマから成る。

リーダビリティの研究は、英語では Flesch Reading Ease (FRE) という評価基準（リーダビリティ指標）が標準的指標として確立し、保険証書の質保証を目的として州保健法に数値基準が規定されていたり、Microsoft Word の文章校正機能に採用されていたりする。日本語では標準的指標は未確立だが、システムの開発事例として、教科書コーパスを使用する「帯」^[3]や「jReadability」^[4]が存在する。

我々はこれまでに、日本語文のリーダビリティを簡易的に計測する手法として、英語リーダビリティ指標を拡張する「変数置き換え法」を提案し、提案手法が「帯」と中程度以上の相関があることを示してきた。また、提案手法を用いた日本語文の可読性評価システムとして「Master Readability (以降は MRead)」というシステムを開発している^[5]。

本稿では、日本語リーダビリティ指標において、妥当な評価値が得られる文書の範囲と現実的な利用方法の解明を目的として、計算機と大学生による可読性評価の比較実験を実施した結果について報告する。本稿の研究全体における位置づけは、日本語リーダビリティ指標の用途を明確にすることによって、MRead の信頼性を向上させることにある。

2. リーダビリティ指標の開発モデルと検証方法

リーダビリティ指標は、文章の読みやすさ・分かりやすさを計算機に推定させるための基準のことである。リーダビリティ指標の種類と検証方法の一覧を表 1、表 2 に示す。

表 1 リーダビリティ指標の開発モデル

線形式モデル	回帰分析、主成分分析による線形式を利用
語彙リストモデル	難易度毎に分類された語彙の割合を利用
言語モデル	難易度毎に分類された文書集合に対する最尤推定の結果を利用 ^[1]
変数置き換えモデル	英語リーダビリティ指標、および日本語対応表を利用 ^[1]

表 2 リーダビリティ指標の評価値の検証方法

多文計測法	多様な文書の評価値を計測
クローズ法	空間補充問題の正答率と比較
分類法	難易度分類問題の一致度を分析
一対比較法	文書難易度の順序関係を分析

日本語リーダビリティ指標の過去研究では、「多文計測法」、「クローズ法」、「分類法」による検証が実施されている。

3. 実験方法

本実験では、5 種類の文書（「羅生門」、小説：「空の境界」、「日経新聞」、「卒業論文」、Twitter の項目：「Wikipedia」）から 500 文字程度の文章を抽出し、「帯」、「jReadability」、「MRead」による評価値の取得と、本学の大学生を対象とした一対比較実験を行う。

「MRead」による評価値の取得では、式 1～式 5 の評価式を用いる。

$$jFRE = 206.835 - (1.015 \times ASL) - (84.6 \times ASW) \quad \dots(式1)$$

$$jFKG = (0.39 \times ASL) + (11.8 \times ASW) - 15.59 \quad \dots(式2)$$

$$jARI = (4.71 \times ACW) + (0.5 \times ASL) - 21.43 \quad \dots(式3)$$

$$jCLI = (5.88 \times ACW) - (29.6 / ASL) - 15.8 \quad \dots(式4)$$

$$jSMOG = 1.031\sqrt{30 \times PS} + 3.1291 \quad \dots(式5)$$

※ASL = 字種分割数/センテンス数

※ASW = 英語1音節・漢字1字単位の再分割数/字種分割数

※ACW = シヤノン情報量に基づく重み/字種分割数

※PS = 英語3音節・漢字3字以上の字種分割数/センテンス数

一対比較実験では、本学の情報工学科に所属する計 12 人の大学生（全て男性）を対象として、文章の読みやすさに関するサーストンの一対比較を行う。実験手順は、まず、被験者に各文章を音読してもらい、次に、難しいと感じた語句にチェックを入れてもらう。そして最後に、読みやすい文章を二者択一で回答してもらう。この時、音読時間、難しいと感じた語句の総数（難語数）も記録する。

[†] 神奈川工科大学大学院工学研究科

Graduate School of Engineering, Kanagawa Institute of Technology

4. 実験結果

一対比較実験の結果と音読時間、難語数について相関分析を行ったところ、音読時間との寄与率 R^2 が 0.515、難語数との寄与率 R^2 が 0.710 となり、一対比較実験の結果が音読結果および難語数に則したものであることが示された。

「帯 (T13)」, 「jReadability」, 「MRead (jFRE)」の評価値について相関分析を行ったところ、各指標間の寄与率 R^2 が 0.645~0.897 となり、中程度以上の相関があることが示された。一方で、各指標の評価値と一対比較実験の結果について相関分析を行ったところ、寄与率 R^2 が 0.008~0.077 となり、相関がほとんどないことが示された (図 1 参照)。以上より、計算機による評価が必ずしも人による評価と一致しないことが明らかになった。

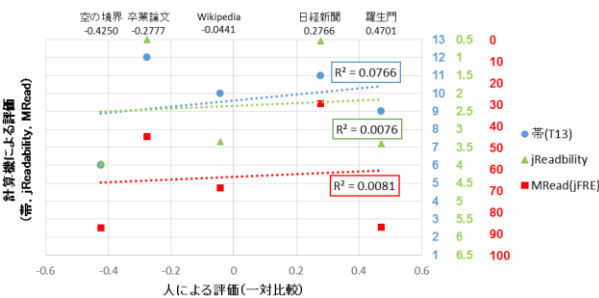


図 1 計算機による評価と人による評価の相関

ここで、どのような文章が人による評価と一致しなかったか分析するため、各指標の評価値と一対比較実験の結果について、文章を読みやすい順に並べ替え、順位表 (表 3) を作成した。

表 3 読みやすい文章の順位表

順位	一対比較	帯 (T13)	jReadability	MRead (jFRE)
1	空の境界	空の境界	空の境界	空の境界
2	卒業論文	羅生門	羅生門	羅生門
3	Wikipedia	Wikipedia	Wikipedia	Wikipedia
4	日経新聞	日経新聞	日経新聞	卒業論文
5	羅生門	卒業論文	卒業論文	日経新聞

表 3 より、“卒業論文”と“羅生門”において、計算機による評価と人による評価が一致していない。“卒業論文”では人による評価が読みやすい、計算機による評価が読みにくいと判断されており、“羅生門”では人による評価が読みにくい、計算機による評価が読みやすいと判断されている。

“卒業論文”の文章は、被験者の専門分野と同一である情報工学分野を扱っており、また、スマートフォンという単語が計 12 回、フィーチャーフォンという単語が計 4 回使用されており、平易な単語が多数出現している。“羅生門”の文章は、古典的な表現が使用されており、難語数も 5 種類の文書中最大となっている。以上より、計算機による評価と人による評価が一致していない原因としては、“卒業論文”では被験者の知識的な偏りと異なり語数、“羅生門”では難語数の影響が考えられる。

5. 考察

リーダビリティを、①読みたいか、②読みやすいか、③音読しやすいか、④理解できるか、⑤覚えやすいか、とい

う 5 種類の評価項目に分類できると仮定すると、現状のリーダビリティ指標は①~③までに対応可能であり、英語リーダビリティ指標である FRE は①~③の全てに対応している。しかし、日本語リーダビリティ指標の場合は、日本語が表音文字と表意文字の双方を含むため、①~③の全てに対応するような指標を作成することは難しいと考える。我々の提案手法である「変数置き換え法」でも、現状、①と②には対応しているが、日本語の音読しやすさを評価する専用の変数を設けている訳ではないため、③には対応できていないと考える。

本実験では、一対比較において、音読と難読語に注目するような教示を与えているため、実験結果が③音読しやすいか、④理解できるかを考慮したものになる。したがって、計算機による評価が必ずしも人による評価と一致しないという本実験の結果は、「帯」, 「jReadability」, 「MRead」が、③音読しやすいか、④理解できるかという評価を適切に判定できていないことを示唆するものである。

6. まとめ

日本語リーダビリティ指標において、妥当な評価値が得られる文書の範囲と現実的な利用方法の解明を目的として、「帯」, 「jReadability」, 「MRead」による評価値の取得と、本学の情報工学科に所属する計 12 人の大学生を対象としたサーストンの一対比較実験を実施した。

その結果、各指標間に中程度以上の相関があるが、各指標と一対比較実験の結果には相関がほとんどないことが示され、計算機による評価が必ずしも人による評価と一致しないことが明らかになった。計算機による評価と人による評価が一致しなかった原因としては、被験者の知識的な偏り、異なり語数、難読語数が考えられる。

この結果は、現状の日本語リーダビリティ指標では、音読しやすさや理解しやすさを適切に判定できないことを示唆しており、適切な判定を実現するためには、異なり語数や難語数に着目したチューニングを行ったり、専用の評価指標を開発したりする必要があると考える。

謝辞

本実験を手伝っていただいた神奈川工科大学の情報工学科卒業生である渡辺健太氏、および、「帯」のプログラムをご提供いただいた名古屋大学大学院の佐藤理史教授に感謝の意を表明いたします。

参考文献

- [1] 岡村祐美, “「文書表現 II」におけるアカデミック・ライティング能力向上のための取り組み”, 神戸学院大学教育開発センタージャーナル, Vol.5, p.45-56, (2014).
- [2] 稲積宏誠, 大野博之, 竹内純人, 大久保麻理子, 又平恵美子, “ICT を活用した日本語文章力育成への取り組み”, 情報処理学会研究報告 コンピュータと教育(CE), Vol.109, no.9, p.1-10, (2011).
- [3] 佐藤理史, “均衡コーパスを規範とするテキスト難易度測定”, 情報処理学会論文誌, Vol.52, No.4, p.1777-1789, (2011).
- [4] 李在鎬, “日本語文章難易度判別システム(jReadability)”, <http://jreadability.net/>, (参照 2016-06-27).
- [5] 赤木信也, 納富一宏, “レポート提出における文章セルフチェックシステムの開発”, 情報処理学会研究報告 コンピュータと教育(CE), Vol.132, No.21, p.115-120, (2015).