

平成21年 5月20日現在

研究種目：基盤研究（B）  
研究期間：2007～2008  
課題番号：19300277  
研究課題名（和文） 日本語リーダビリティ測定尺度の構築とソフトウェアの実用化  
研究課題名（英文） Constructing a Readability Scale of Japanese Texts and Developing a Software  
研究代表者 柴崎秀子 (SHIBASAKI HIDEKO)  
長岡技術科学大学工学部・教授  
研究者番号：00376815

## 研究成果の概要：

国語の教科書をデータベースにして、日本語の文章の読み易さを計量化し、数値で示すことができる式を構築した。この式により、当該テキストが小学1年から中学3年までのどの学年レベルに近いかを判定することができる。式は、 $\text{学年} = -0.148 * \text{文章中の平仮名の割合} + 1.585 * 1 \text{文の平均述語数} - 0.117 * 1 \text{文の平均文字数} - 0.126 * 1 \text{文の平均文節数} + 15.581$  である。

## 交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	7,300,000	2,190,000	9,490,000
2008年度	8,300,000	2,490,000	10,790,000
年度			
年度			
年度			
総計	15,600,000	4,680,000	20,280,000

研究分野：総合領域

科研費の分科・細目：教育工学

キーワード：リーダビリティ，日本語，文章理解，形態素解析，文章の難易，ソフトウェア

## 1. 研究開始当初の背景

海外には、英語、スペイン語、フランス語、韓国語、中国語、ベトナム語、デンマーク語、ヘブライ語を初めとする多くの言語において、文章の読み易さを公式で表わす方法が200以上もあり、無料でダウンロードできるもの、ソフトウェアとして販売しているものがあるが、日本語では、本格的なリーダビリティ公式が存在せず、誰でも使えるようなツールがなかった。

## 2. 研究の目的

- (1)日本語の文章のリーダビリティ公式を構築すること。本研究での公式とは、小学1年から中学3年までのレベルを判定する式のことである。
- (2)公式をコンピュータに実装し、インターネットで公開することにより、誰でも簡単に文章の難易判定ができること。

### 3. 研究の方法

(1)国語教科書を小学校1年生から高校3年生まで電子ファイル化し、コンピュータで文字数、文字種の割合、述語の数、語種の割合、文節の数を分析した。その結果、図1及び図2に示したように、テキストを構成する要因は、学年が上がるに連れて上昇または下降の傾向にあることが示されたので、線形重回帰分析を用いて、学年を判定する式を構築することにした。

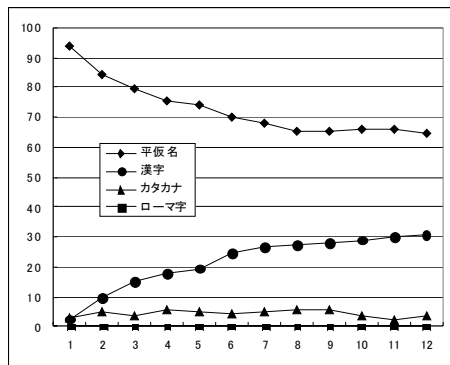


図1 学年ごとの文字種の割合

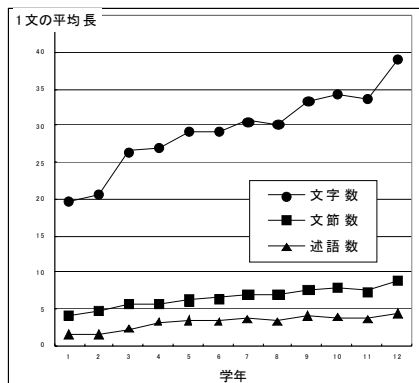


図2 1文当たりの文字数・文節数・述語数

(2)テキスト要因の分析から、学年を従属変数とし、1文の平均文字数、文節数、述語数、平仮名の割合を説明変数として、重回帰分析を行い、以下の学年判定式ができた。

$$Y = -0.148X_1 + 1.585X_2 - 0.117X_3 - 0.126X_4 + 15.581$$

Y=学年

X<sub>1</sub>=文章中の平仮名の割合

X<sub>2</sub>=1文の平均述語数

X<sub>3</sub>=1文の平均文字数

X<sub>4</sub>=1文の平均文節数

### 4. 研究成果

本研究により構築された判定式をコンピュータプログラムとして実装し、インターネットで公開した。この公式を構築したことで、教育現場で国語テストを作成する際のテキストの難易測定に役立っているという声が寄せられている。また、小学生中学生の推薦図書を選ぶ際にも大いに役立っている。図3は2008年のベストセラー「ホームレス中学生」を本研究で作られたツールで測定した結果の一部である。



図3 「ホームレス中学生」の測定結果

### 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8件)

- ① 柴崎秀子・玉岡賀津雄・高取由紀「アメリカ人は和製英語をどのくらい理解できるか—英語母語話者の和製英語の知識と意味推測に関する調査」日本語科学 第21号 国立国語研究所, pp. 89-110 (2007) 査読有
- ② 柴崎秀子・玉岡賀津雄・母育新「自然習得の恩恵—日本語の短縮語復元課題における中国と日本の学習環境の比較—」レキシコンフォーラム No. 3, ひつじ書房, pp. 335-350 (2007) 査読有り
- ③ 玉岡賀津雄・林炫情・池映任・柴崎秀子「韓国語母語話者による和製英語の理解」レキシコンフォーラム No. 4, ひつじ書房 pp. 195-222 (2008) 査読有
- ④ 柴崎秀子「日本語の文章の難易はどのように測定できるか」AAMT Journal 44, アジア太平洋機械翻訳会, pp. 5-8,

(2009)査読無し

- ⑤ 柴崎秀子・沢井康孝「国語教科書コーパスを応用した日本語リーダビリティ構築のための基礎研究」, 信学技報 NLC2007-32(2007-10), pp.19-24, 電子情報通信学会 (2007) 査読無し
- ⑥ 柴崎秀子・玉岡賀津雄・沢井康孝「漢字表記と平仮名表記が文の読み易さに与える影響—日本語リーダビリティ構築のための基礎研究—」言語科学会第10回年次国際大会予稿集, p18, (2008) 査読有り
- ⑦ 柴崎秀子「日本語コーパスを応用した文章の難易測定の研究」特定領域研究「日本語コーパス」平成19年度公開ワークショップ研究成果報告会予稿集, pp.125-130 (2008) 査読無し
- ⑧ 柴崎秀子「日本語リーダビリティ公式の構築と測定ツールの開発」特定領域研究「日本語コーパス」平成20年度公開ワークショップ研究成果報告会予稿集, pp.155-160 (2009) 査読無し

[学会発表] (計4件)

- ① Shibasaki, Hideko, Developing a readability scale for Japanese texts by a liner multiple regression analysis predicting school grades, アジア読書学会世界大会, 2007年8月4日, 東京: 国際オリンピック記念青少年センター
- ② 柴崎秀子・沢井康孝「文章中の漢字と漢語の分析による日本語リーダビリティ公式構築のための基礎研究」, パリ漢語学会国際会議, 2008年3月14日, パリ: パリ第7大学
- ③ 柴崎秀子・沢井康孝「国語教科書コーパスを応用した日本語リーダビリティ構築のための基礎研究」電子情報通信学会, 2007年10月5日, 東京: キャンパスイノ

バージョンセンター

- ④ 柴崎秀子・玉岡賀津雄・沢井康孝「漢字表記と平仮名表記が文の読み易さに与える影響—日本語リーダビリティ構築のための基礎研究—」言語科学会第10回年次国際大会, 2008年7月12日, 13日, 静岡: 静岡県立大学

[図書] (計0件)

[産業財産権]

○出願状況 (計1件)

名称: 文章の読み易さ評価システム及び文章の読み易さ評価方法

発明者: 柴崎秀子, 原信一郎

権利者: 長岡技術科学大学

種類: 特許

番号: 007-169544

出願年月日: 2007年6月27日

国内外の別: 国内

○取得状況 (計0件)

[その他]

ホームページ

「リーダビリティ・リサーチ・ラボ」

<http://readability.nagaokaut.ac.jp/>

6. 研究組織

(1) 研究代表者

柴崎秀子 (SHIBASAKI HIDEKO)

長岡技術科学大学工学部・教授

研究者番号: 00376815

(2) 研究分担者

玉岡賀津雄 (TAMAOKA KATSUO)

麗澤大学外国語学部・教授

研究者番号: 70227263

加納 満 (KANOU MITSURU)

長岡技術科学大学工学部・准教授

研究者番号: 80251859

原 信一郎 (HARA SHINICHIRO)

長岡技術科学大学工学部・准教授

研究者番号: 20247464

山本和英 (YAMAMOTO KAZUHIDE)

長岡技術科学大学工学部・准教授

研究者番号: 40359708

(3) 連携研究者

なし