

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 3 月 31 日現在

機関番号：13801

研究種目：若手研究（B）

研究期間：2009～2011

課題番号：21720195

研究課題名（和文）

リーダビリティ式自動生成による最適コンテンツ提示型語学学習 E ラーニングソフト開発

研究課題名（英文）Development of Language Study E-Learning Software Presenting Best-fit Learning Materials by Personalized Readability Equations

研究代表者

宮崎 佳典（MIYAZAKI YOSHINORI）

静岡大学・情報学部・准教授

研究者番号：00308701

研究成果の概要（和文）：本研究を遂行することにより，語学学習者に対し，その一人ひとりの持つ語学レベルに応じて適切な語学学習用コンテンツ（テキスト）をアダプティブに提供するための仕組みを理論的に考え，e ラーニング用途に Web アプリケーションとして実装することに成功した．プロトタイプを開発していたシステムに対し，さらにそのリーダビリティ式開発環境（日本語，英語とも）を多機能にアップグレードすることも実現した．220 名を超える学生に対して実験を行い，それによって得られたデータの基礎的な分析を行った．

研究成果の概要（英文）：By carrying out the research, the research group devised the theoretical framework to adaptively provide proper learning contents (for language study), based on the level of each learner, and succeeded in the implementation as a piece of Web application. Furthermore, upgrading the environment to develop readability equations (both Japanese and English) was realized, by adding multi-functions to the conventional prototype system. Experiments were conducted by more than 220 subjects, and fundamental analysis of the obtained data was made.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009 年度	700,000	210,000	910,000
2010 年度	600,000	180,000	780,000
2011 年度	500,000	150,000	650,000
年度			
年度			
総計	1,800,000	540,000	2,340,000

研究分野：人文学

科研費の分科・細目：言語学・外国語教育

キーワード：e-ラーニング・コンピュータ支援学習（CALL）・リーダビリティ・重回帰方程式・語学学習・Web 検索

1. 研究開始当初の背景

(1) リーダビリティ（日本語では可読性などと翻訳される）公式は，元来英語を第一言語として話す人向けに，英文テキストに対して

開発された経緯がある．入力された英文テキストに対し，1つの実数値を返し，その値によってその英文テキストの読みやすさ（場合によっては読みにくさ）を客観的に判断でき

るような指標を与える。現在も、新聞や雑誌編集者が、執筆した記事が読みやすい物として出来上がったかどうか、また対象となる読者に合わせた文章レベルとなっているのか（例として、子供新聞における平易な記事レベルなど）を調べる判断材料としている。また、教育現場では、教師が生徒（児童）の学年に応じて教材選定をする際などにも使われている。

(2) 最近では、日本人研究者にもリーダビリティ公式の開発に従事する者が増え、第二外国語として英語を学ぶ学習者たちのための英文公式なども数多く考案されている。申請者も研究グループ内で、英文用リーダビリティ公式を構築支援するツールを作成している：
・ Y. Miyazaki, K. Norizuki, Developing a Computerized Readability Estimation Program with a Web-searching Function to Match Text Difficulty with Individual Learners' Reading Ability, WorldCALL 2008, p. 107 (2008)
・ 宮崎 佳典, Readability 公式開発環境の構築ならびに Readability 値に基づくネット文書抽出ツール作成, LET 第 47 回全国研究大会, pp. 70-73 (2007)

(3) ただ、このリーダビリティという概念には問題点も多い。まず、この公式は母国語としての英語話者向けに開発されており、日本人のような英語学習者には適合しないケースが多い。また、多種多様なリーダビリティ公式 (Flesch, Flesch-Kincaid, Dale-Chall, Coleman-Liau 等) が世界中で利用されていて、その中にはテキスト中の語彙レベルをパラメータとして持つ公式も含まれるが、そもそも語彙レベルは英語話者であっても個人差があるのに、英語学習者にはさらにその差が顕著に現れる。それを一律に定式化するのは無理がある。さらに、リーダビリティ公式は英文用に開発されたものばかりであり、例えば日本語文書向けに開発された公式は見当たらない (類似研究は散見されるが、十分な成果の得られたものはない)。

2. 研究の目的

(1) 本研究の目的は、語学学習者に対し、その一人ひとりの持つ語学レベルに応じて適切な語学学習用コンテンツ (テキスト) をアダプティブに提供するための仕組みを理論的に考え、Eラーニング用途に Web アプリケーションとして実装することである。具体的に申請者が着想したのは、既に世界中で開発されている「文書の可読性」を表す従来のリーダビリティ公式を、公式としてではなく、個々の学習者が能動的あるいは自動的に算

出するリーダビリティ式として再定義し、その式によって導かれるリーダビリティレベル (実数値) を元に、学習者が指定したレベル値のコンテンツ (テキスト) をインターネット上で自動検索させるプログラムを開発することである。また、「適切なコンテンツ」は必ずしも文書の可読性だけから判定されるものではなく、一般的には学習者の興味に合わせて考えられることが多い。そのため、学習者が事前に興味の有る分野を指定しておく、または学習者の学習履歴より自動判断し、合致するジャンルのコンテンツのみが対象となるように選別することにより、学習者の学習意欲を継続させることが可能となる。

(2) そこで申請者は、リーダビリティは公式としてではなく、個々のユーザごとに開発されるべきものであると考えた。現時点では、学習者がテキストを読むたびにその文書の難易度を評定することで、蓄積されたデータを元に重回帰分析を行い、個々のユーザ用リーダビリティのパラメータを同定ならびにその一次式の係数を自動計算させている。個々のリーダビリティ式が確立してからは、申請者は自分にふさわしい難易度を指定すると、自分のリーダビリティ式を用い、指定範囲内のテキストを自動探索する。

(3) 最終的には、語学学習者が自分の欲しいレベル値を入力するのではなく、日頃の語学学習の結果や学習履歴データによって得られる学習レベルを自動的に算出し、それに合ったコンテンツを自動抽出するところまで開発する計画である。同様に、学習者の興味を持つ分野についても、Web のブラウジング・ヒストリから当該分野を割り出し、学習者は基本的に何も指定しなくても、無理なく学習でき、さらに魅力溢れる学習コンテンツがインターネットから手元に届けられる。

(4) このような環境は申請者の知る限り、国内外において現時点で開発されておらず、英語学習者や日本語学習者など、語学学習者一般にとって大きな朗報になることは間違いない。これを応用すれば、インターネット上の語学学習用コンテンツを各自が自動でレベル分けし、世界中に散在する英語学習者・日本語学習者が、自分のレベルに応じてすぐに学習できるようになる。インターネットの大海原で大量のコンテンツに恵まれながら、自分用の学習テキストを峻別できないで、どれから学習しようか手をこまねいている状況は、今日の情報爆発時代に共通した問題点であり、喫緊に解決すべき難題でもある。

3. 研究の方法

(1) 日本語テキストと英文テキストを比較すると、英文は単語間に空白あるいはコンマ、ピリオドなどが置かれるのに対し、日本語は基本的には単語区切り用デリミタが置かれることはない（句読点を除く）。それも手伝って、日本語電子テキストの自動解析は一般に困難と考えられてきた。しかしながら、例として奈良先端科学技術大学院大学で開発された茶釜、京都大学で開発された Mecab を用いることにより、テキストを単語毎分割し、さらに形態素解析まで行うことができる。これを導入することに加え、英文リーダビリティ公式開発で培われてきた知識を応用し、日本語リーダビリティ公式を開発することが可能となる。特に言語には独立した可読性要素（たとえば文章の長さや語彙レベルの高い用語を多用している場合の可読性の低さとの関連性または相関性など）は英文リーダビリティ公式にも頻出するパラメータであり、言語に独立した部分と依存する部分を統計処理することにより抽出することを試みる。

(2) 正確なリーダビリティ公式を導出するには多くの実験を行い、サンプル数を増やす必要がある。英語学習者に関しては当面の間日本人対象に実験を絞ってもよいと考えている。一方で、日本語学習者はもちろん日本人に非ず、国内をはじめ世界中に存在する。これに対し、申請者の研究者ネットワークには大韓民国（例：大真大学）を筆頭に日本語を教える研究者がおり、彼（女）らに実験を依頼する。また申請者の所属する大学には約 300 名の留学生がおり、データ収集の素材には事欠かない。さらに申請者の前職の大学には約 400 名留学生が在籍しており、研究者間のネットワークを通じて、同機関にも実験を依頼する。

(3) 日本語リーダビリティ式開発環境構築に際し、研究代表者ならびに研究協力者間でミーティングを行い、以下を討議する：

- ・リーダビリティ式として肝要と思われるパラメータの選定
- ・実験の実施に必要な各種ドキュメントの作成・作問（日本語テキストならびに対応する問題の作成、Q&A やアンケートなど）
- ・リーダビリティ式開発環境構築

なお、初年度は被験者が正しくリーダビリティ式を作成し、確実に（式作成後）適切なテキスト検索を行える環境構築を第一義に置くため、テキストはあらかじめこちらで用意することとする。テキスト選定は言語テストを専門とする研究協力者（静岡産業大学法月健准教授）に協力を依頼し、数回ミーティングを実施予定である。氏は以前に所属大学にて留学生向け日本語授業を担当していた経緯もあり、日本語テキストの書籍など

も購入した中から実験に妥当と判断される問題を厳選する。勿論、本研究の成否を左右すると思われる重要な箇所は自作問題を開発する。パラメータの選定、ならびに実験の準備などについては研究室内の大学院生などとも協力して体制を整える。

(4) 開発用 PC を一台購入し、リーダビリティ式開発環境構築にあたる。なお、Visual Studio, PHP などはずで揃っており、普段よりこれらの言語でプログラミングを行う大学院生に適宜手伝ってもらう。すでに「研究目的」の項で引用した 2 つの研究業績にもあるように、素地となるソフトは存在するため、コードは一部再利用する。なお、初年度はサーバは立てず、スタンドアロンの環境で開発を行う。理由は初期の段階でインターネット上に公表することを避けるため、ネットワークアプリケーションへの移行に時間を擁するためである。

(5) 実験用問題の第 1 版が整ったところで、複数の大学にてパイロット実験を行う。特に調査する点は、学習者の日本語能力と実際のテストの得点にどの程度相関性があるのか、また、日本語テキスト内のどのパラメータ（例：テキストの長さ、単語の長さ、接続詞の数、難語の出現率など）によって難易度が形成されるかが明確になるか、と言う点である。（学習者の日本語能力は彼らの日本語能力試験の結果や普段の成績からも測定可能である）。初年度に実施した実験データが集まった段階で、データ分析を行う。SPSS ソフトウェアを用い、因子分析を行い、日本語テキストにおけるリーダビリティ公式に必要なパラメータを抽出する。また、いくつかの候補となる公式を選び、実際に受験をした学習者達に式の妥当性に関する聞き取り調査・アンケート調査を行う。

(6) その後、日本語リーダビリティ式開発環境だけでなく、既に基本的な開発の済んでいる英語リーダビリティ式開発環境をさらにブラッシュアップして実装し、双方の実験を独立して行う。また、サーバを購入し、これらの開発環境を公開する。本実験も日本語リーダビリティ、英語リーダビリティ両方を実施する。最終的には語学学習者がレベル値を入力するのではなく、日頃の学習履歴データにより得られる学習レベルを自動的に算出し、それに応じたコンテンツを自動抽出するところまで開発することを目指す。また、学習ジャンルについても同様で、当初はユーザによる指定入力を促すが、最終的にはブラウジング・ヒストリより自動的にそのジャンルを特定するような機能を持たせる。

4. 研究成果

(1) 本研究を遂行することにより、語学学習者に対し、その一人ひとりの持つ語学レベルに応じて適切な語学学習用コンテンツ（テキスト）をアダプティブに提供するための仕組みを理論的に考え、eラーニング用途に Web アプリケーションとして実装することに成功した。プロトタイプを開発していたシステムに対し、さらにそのリーダビリティ式開発環境（日本語、英語とも）を多機能にアップグレードすることも実現した。

(2) 具体的な拡張として、次のような仕様・手順を満たすプログラムを作り上げた：

- A. 学習者は自分の日本語力に応じた、リーダビリティ値を入力
- B. リーダビリティ値がその近辺となるような日本語コンテンツをネットから検索
- C. 学習したコンテンツをデータベース化し、それ以降の被験者の語学力を予測

(3) さらに、リーディングを行うコンテンツのジャンルについても考察し、各学習者に対し、ジャンル毎のリーダビリティ式を生成する機構を整えた。また、学習者の学習意欲を継続させるため、アノテーション（注釈）機能を新たに追加し、学習者が不得意な単語や文についてタグ付けが行えるようにした。220名を超える学生に対して実験を行い、それによって得られたデータの基礎的な分析を行った。

(4) 被験者の嗜好性や背景知識などにより、被験者がジャンルごとにリーダビリティ式にもっとも影響を与えられられるパラメータは異なることが考えられる。そして、そのパラメータを検出した結果、ジャンルごとのリーダビリティ式によるパーソナライゼーションに繋がることが判明した。また、リーダビリティ式のパラメータに文法項目の出現数を利用することにより、より高い重相関係数をもつリーダビリティ式を作成すること、およびより学習者ごとのパラメータの傾向を得ることができることから、文法項目の出現数をパラメータに加えることはパーソナライゼーションに有効であるとの結論に至った。

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4件）

- ①M. Zushi, Y. Miyazaki, K. Norizuki, Web Application for Recording Learners'

Mouse Trajectories and Retrieving Their Study Logs for Data Analysis, Special Issue on "APTEL 2011 Best papers", Knowledge Management & E-Learning: An International Journal (KM&EL), Vol. 4, No. 1, pp. 37-50 (2012), 査読有り.

- ②影山 功, 宮崎 佳典, ジャンル別個人 Readability 式生成の試みならびに属性分析, 統計数理研究所共同研究レポート 254, pp. 21-32 (2011), 査読無し.

- ③長谷川 由美, 宮崎 佳典, 影山 功, 個人用リーダビリティ式自動生成によるリーディング学習システム REX の学習効果について, 滋賀県立大学国際教育センター研究紀要 15 号, pp. 93-104 (2010), 査読無し.

- ④宮崎 佳典, 影山 功, 難語率の定義のバリエーションが個人 Readability 式生成に与える影響, 統計数理研究所共同研究レポート 239, pp. 73-85 (2010), 査読無し.

〔学会発表〕（計6件）

- ①Y. Miyazaki, S. Tanaka, Y. Koyama, Development and Improvement of a Corpus-based Web Application to Support Writing Technical Documents in English, International Conference on Computers in Education (ICCE2011), pp. 263-270 (2011), 2011年12月1日, 於 タイ, チェンマイ市.

- ②影山 功, 長谷川 由美, 宮崎 佳典, 自分のレベルに合った記事を読もう！ - Readability 式を用いたオンライン日本語学習サイトの実験結果報告 -, 2010 International Conference on Japanese Language Education (ICJLE2010), pp. 1035-0-1035-9 (2010), 2010年8月1日, 於 台湾, 台北市.

- ③Y. Hasegawa, Y. Miyazaki, T. Kageyama, Try REX to Read the Articles Matching Your Reading Proficiency, The 24th Annual Hawaii Association of Language Teachers (HALT) Conference, p. 6 (2010), 2010年3月20日, 於 米国, ハワイ州.

- ④影山 功, 宮崎 佳典, 長谷川 由美, Readability 式を用いたオンライン外国語学習環境の構築, FIT2009, pp. (4)-481-484 (2009), 2009年9月3日, 於 宮城県仙台市 (東北工業大学).

- ⑤宮崎 佳典, 長谷川 由美, 影山 功, 個人向けにチューンナップした Readability 式による日本語学習者用リーディングプログラムの開発, 日本言語テスト学会 (JLTA) 第30回研究例会 (2009), 2009年7月4日, 於 静岡県藤枝市 (静岡産業大学).

⑥Y. Hasegawa, Y. Miyazaki, T. Kageyama,
Utilizing Software to Derive
Personalized Japanese Readability
Equations, The 23rd Annual Hawaii
Association of Language Teachers (HALT)
Conference, p. 6 (2009), 2009年4月18
日, 於 米国, ハワイ州.

[その他]

ホームページ等

[http://lmo.cs.inf.shizuoka.ac.jp/~oshir
o/readability/rex_system/index.php](http://lmo.cs.inf.shizuoka.ac.jp/~oshir
o/readability/rex_system/index.php)

6. 研究組織

(1) 研究代表者

宮崎 佳典 (MIYAZAKI YOSHINORI)

静岡大学・情報学部・准教授

研究者番号：00308701

(2) 研究分担者

なし ()

研究者番号：

(3) 連携研究者

なし ()

研究者番号：