

N-009

Web 上の英字新聞を使った英語多読支援システムの ユーザインタフェースの検討

User Interface for a System to Support Extensive Reading using Online English News

谷本 祐次[†] 安藤 一秋[‡]
Yuji Tanimoto Kazuaki Ando

1. はじめに

英語のリーディング力を向上させる方法の一つに多読がある。多読とは、辞書がほとんど必要ないレベルの文章を大量に読む学習法である。多読を行うことで、語彙力と読解力に一定の伸びが期待できると報告^[1]されている。多読の中でも SSS 式多読^[2]は、より効果が得られる方法として注目されている。SSS 式多読では、辞書をなるべく引かないようにすることや興味のある素材を選択すること等が重要であるとされている。また、多読支援ガイド^[3]によると、既知単語が 1 ページあたり 90%未満であれば、辞書を引く手間が増加し、読みの速度が低下するだけでなく、ストレスがたまり内容の理解度が低くなると報告されている。

英語多読の教材としては、英字新聞や英語で書かれた書籍などが挙げられる。英字新聞を教材として利用することは、単に英語の運用能力を向上させるだけでなく、現在、世間で起こった出来事や注目されていること、一般教養についての理解を深めることができるという利点がある。また、海外の新聞社が日本の出来事を取り扱った記事であれば、海外から見た日本について知ることができる。その他にも、記事内容を母国語のニュース等であらかじめ知っていた場合は、内容を理解しやすくなるという利点がある。したがって、英字新聞は多読の教材として適していると考えられる。

しかし、既存のニュース閲覧サイトでは、難易度や単語の意味提示等が存在せず、多読を行うのに適しているとはいえない。

そこで、本研究では、多読の教材として Web 上で提供されている英字新聞に着目し、英字新聞の選択・読解支援機能を有する多読支援システムの構築を目的とする。本稿では、現在開発している多読支援システムのユーザインタフェースを中心に説明する。また、提案システムが支援できる利用者などを調査するために、英字新聞の難易度を調査した結果について説明する。

2. 多読支援システム

2.1 システムの概要

多読を効果的に行うためには、教材で使用されている語彙の難易度と利用者の語彙力の対応が取れていることや教材の内容に興味を持てることなどが重要となる。そこで、本研究では、記事の選択・読解支援の機能を有する多読支援システムを構築する。

現在、構築しているシステムの概要を図 1 に示す。本シ

ステムでは、Web 上から辞書データと様々なニュース閲覧サイトの記事を収集する。収集した記事に対しては、ジャンルや品詞等の情報を付与して保存する。また、利用者の TOEIC スコアや利用履歴である閲覧情報を基に、記事の選択支援や読解支援を行う。現在は、Web 上の記事と辞書データを収集し、DB を構築する部分まで実装済みである。

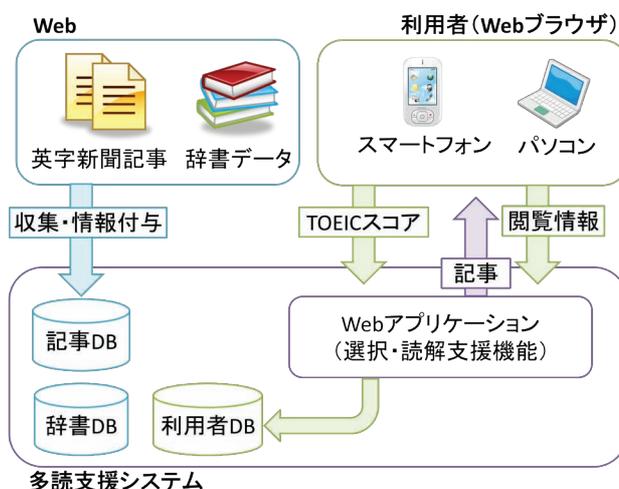


図1 システム概要図

2.2 システムのユーザインタフェース

現在構築中であるシステムの記事選択ページと記事閲覧ページのユーザインタフェースについて説明する。

2.2.1 記事選択ページ

記事選択ページのインタフェースを図 2 に示す。

既知単語割合	タイトル	ジャンル	語数	後で読む
98%	ABC No Rio Fondly Recalls the Saturdays of Its Gritty Youth	entertainment	700	後で読む
95%	Obama, Renzi pledge to focus on threats from Libya	world	700	後で読む
92%	Study Finds Climate Change as Threat to 1 in 6 Species	science	520	後で読む
92%	Cisco Names Charles Robbins to Succeed C.E.O. John Chambers	technology	520	後で読む
89%	McDonald's Chief Promises Turnaround in a Restructuring	business	700	後で読む
-	Baseball: Ogawa does it all as Swallows beat BayStars	sports	-	
98%	Baseball: Ogawa does it all as Swallows beat BayStars		520	
95%	Ogawa does it all as Swallows beat BayStars		600	

図 2 記事選択ページ

図 2 に示すように、記事選択ページでは、利用者が自身の語彙力に適した記事を選択できる参考情報として、記事毎に既知単語の割合を提示する。また、新聞記事をジャン

[†] 香川大学大学院工学研究科 Graduate School of Engineering, Kagawa University

[‡] 香川大学工学部 Faculty of Engineering, Kagawa University

ルごとにまとめ、利用者の興味や語彙力に応じて、記事を並び替えて提示する。複数のニュース閲覧サイトが提供している同じ事柄を扱った記事は、記事リストの中に重複表示されないように、集約して提示する。

2.2.2 記事閲覧ページ

記事閲覧ページのインタフェースの一部を図3に示す。



図3 記事閲覧ページ

図3に示すように、記事閲覧ページでは、記事タイトル、本文を提示する。また、本文に対しては、スラッシュリーディング機能や単語の品詞による色分け機能などで読解を支援する。これらの機能については、機能のON/OFFを自由に変更できるようにボタンを用意している。

単語の意味表示としては、単語をマウスオーバーすることで、文脈に適した意味を優先的に表示する機能を実装する。また、システム側で利用者が未知の単語であると判定した場合、その単語を太文字表示する。単語の既知・未知については、利用者が該当単語をクリックすることで切り替えできるようにし、その情報を語彙予測に活用することで予測精度の向上を目指す。さらに、詳細な閲覧履歴を収集するために、画面右側には、利用者が記事をどこまで読み進めたかの情報を入力するためのボタンを用意している。

3. 英字新聞記事の難易度調査

提案システムでは、英字新聞を多読の教材として利用する。多読では、教材と利用者の語彙レベルの対応が重要となる。そこで、新聞記事のジャンルと難易度(リーダビリティ)の関係を調査することで、提案システムが支援可能な利用者の語彙レベルを確認する。

調査には、国内外の8つのニュース閲覧サイト(朝日新聞、毎日新聞、The Japan Times, BBC, The New York Times, Time VOA)の9,378件の記事データ(収集期間:2015/1/12~2015/2/3)を利用する。比較するジャンルは、どのニュースサイトでも取り扱われるWorld, Sports, Business, Science, Technology, Entertainmentの6つとする。また、難易度の指標としては、FKG^[4]とCLI^[4]の2つの指標を利用する。FKGは、アメリカで教科書の選定等に利用される指標であり、一文あたりの単語数と音節数から難易度を求める指標である。FKGの計算には音節数が必要になるが、本調査では、Weblio^[5]から音節数を取得できた単語のみを利用している。CLIは、一文あたりの単語数と単語あたりの平均文字数から難易度を求める指標である。これら2つの指標を利用することで、より正確に難易度判定を行う。

ジャンル毎の記事数と各指標により得られた値の平均値をまとめたものを表1として示す。表1に示すように、他のジャンルと比較して、Sportsジャンルの値が低くなっており、読みやすい記事が多いことがわかる。これは、Sportsジャンルの記事では、試合の得点を表す数値や比較的簡単な単語がよく利用されることが影響していると考えられる。一方、Scienceジャンルの記事については、普段

あまり使われない難しい単語や専門用語が使われるため、値が高く読みにくい記事が多くなったと考えられる。

表1 ジャンル別の各難易度の平均

カテゴリ	件数	CLI	FKG
World	3674	12.1	12.1
Sports	2111	10.4	9.8
Business	1828	11.6	11.6
Science	330	11.7	11.8
Technology	522	11.6	11.4
entertainment	913	10.9	11.2

また、ジャンルを問わずCLIの値によって、記事の分布数を調査したところ、11程度の記事が最も多いという結果になった。しかし、1~9程度の範囲の記事でも1,623件あり、1日平均約71件の記事が追加されていることから、利用者の興味に応じた記事が含まれている可能性が高いと考えられる。そのため、CLI値が9程度の記事が理解できる人(染谷の論文^[6]によると日本の平均的な大学生)であれば、システムを利用できる可能性が高いと考えられる。

4. おわりに

本稿では、Web上の英字新聞を活用した、記事選択と読解支援機能を有する多読支援システムの概要について説明した。そして、国内外のニュース閲覧サイトの記事データを分析することで、日本の平均的な大学生であれば、システムで支援できる可能性がわかった。また、本調査で利用した英字新聞では、Sportsジャンルに読みやすい記事が多く含まれるという結果を得た。

今後は、二つの支援機能を順に実装する。語彙予測では、予測に必要なデータを収集し、SVMやIRTなどを用いた手法を検討する。また、語彙予測で得たデータを利用してさらなる調査を行うことで、より正確にシステムが支援できる利用者の推定を行う。記事の集約については、TF-IDFを使った基本的な方法やその他の手法について比較を行い、適した手法を検討する。文脈に応じた意味の優先表示やスラッシュリーディング機能については、既存の手法を調査し、分析した後に実装を行う。

参考文献

- [1]岡崎他, “英語多読授業の効果に関する実証的研究”, 富山大学人間発達科学部紀要, 3巻2号, pp.165-176, 2009
- [2]SSS 英語学習法/多読+シャドウイング, <http://www.seg.co.jp/sss/>
- [3]Writing-Extensive Reading Foundation, http://erfoundation.org/ERF_Guide1.pdf
- [4]翻訳・ローカライズ・通訳-東京渋谷-エイアンドピープル, <http://www.a-people.com/readability/readability.html>
- [5]Weblio 辞書:英和辞典・和英辞典, <http://ejje.weblio.jp>
- [6]染谷泰正, “オンライン版「英文語彙難易度解析プログラム」(Word Level Checker)の概要およびその教育分野での応用可能性”, 青山学院大学文学部紀要, 51号, pp.99-122 (2009).