

学習者特性に応じた難易度の英語テキスト提供を目指す 英単語語彙力推測アルゴリズムの開発

Development of an Algorithm Estimating English Vocabulary Levels of Individual Learners for Provision of Their Suitable English Texts

白須 直樹^{*1}, 鈴木 竣丸^{*2}, 宮崎 佳典^{*3}
Naoki SHIRASU^{*1}, Shumma SUZUKI^{*2}, Yoshinori MIYAZAKI^{*3}

^{*1} 静岡大学 情報学部

^{*1} Faculty of Informatics, Shizuoka University

^{*2} 静岡大学大学院 総合科学技術研究科

^{*2} Graduate School of Integrated Science and Technology, Shizuoka University

^{*3} 静岡大学大学院 情報学領域

^{*3} College of Informatics, Shizuoka University

Email: cs16041@s.inf.shizuoka.ac.jp

あらまし：発表者らが開発する英語リーディング学習アプリケーション REX は、学習者のレベルに合った英語テキストの提供を目指しており、学習者ごとにリーダビリティ式の自動生成を行う。リーダビリティとはテキストの難易度を示し、テキスト中の単語数や文長、難語率（未知語の割合）などを用いて決定される。難語率は学習者に依存するが、全英単語について既知・未知を記録するには膨大な時間がかかる。我々は自他学習者の語彙力データを活用して、欠如している既知・未知情報を推測するアルゴリズムを開発し、その実験結果を報告する。

キーワード：e-Learning, 多読学習, 推測手法, リーダビリティ

1. はじめに

著者らは Web 上でリーディング学習を行うアプリケーション REX (Reading EXercise) の開発を行っている。REX は英語テキストを提供し、多読学習を円滑に行えるようにすることを狙いとしている。狙いのために必要なことは、学習者にとって簡単すぎず難しすぎない英語テキストをできるだけ提供することである。そのためには英語テキストの難易度や学習者の英語能力情報が必要となる。英語テキストの読みやすさを計算するものとしてリーダビリティ公式が存在する。リーダビリティ公式はテキスト中の単語数などをパラメータとして用いるが、同じテキストであれば、どんな英語能力を持つ人にとっても同じ値を示すこととなる。そこで REX ではリーダビリティ公式を改良し、学習者ごとにその学習履歴から英語テキストの読みやすさを示す式(「リーダビリティ式」)を自動生成する。テキスト中の未知語の割合は当然学習者によって異なり、REX は学習者ごとに各英単語の既知/未知情報を記録している。しかしながら、膨大な数の英単語の 1 つ 1 つに既知/未知を学習者に判定してもらうには負担が大きい。そこで、少ない負担かつ短時間で膨大な数の英単語の既知/未知を推測するアルゴリズムを考案した。

本発表では、REX において学習者に最適な英語テキストを提供するメカニズムならびに英単語の既知・未知の記録法について述べ、さらに未記録の単語に対して著者らが開発した推測アルゴリズムについて照会し、実験によってその有用性を検証する。

先行研究として、福井ら⁽¹⁾は英単語、熟語の難易度辞書作成機能と英語リーダビリティ式作成機能の

ある CheckRead の開発を行っている。土屋⁽²⁾は英語リーダビリティ公式に使用されるパラメータ(変数)について、一文中の平均単語数と難語率が英語テストの誤答率に影響を及ぼしたことと述べている。

2. REX の概要

REX は SVL12000 内の英単語群について、学習者ごとにラベル付けを行う機会を与える。なお、SVL12000 は 12,000 語の英単語を 12 のレベル×1,000 単語にカテゴライズしている単語リストである。ラベル付けの方法にはリーディング、英単語ゲームの 2 つが存在し、学習者自信が行う学習行動によってラベル付けを行う。学習の流れを図 1 に示す。



図 1 学習の流れ

リーディングを行うと英単語ゲームをプレイするために必要なバトルチケットを入手することができる。またリーディングで読了テキストの難易度を 6 段階の中から自己評価 (Rating) する。英単語のラベル付けは学習者がリーディング、英単語ゲーム中の双方で行い、英語テキスト中の未知語の割合である難語率の算出に用いられる。難語率はリーダビリティ式のパラメータに採用されることが多く、学習者

にとって最適な難易度の英語テキスト提供のための重要な働きを担っている。

大城ら⁽⁹⁾は同一テキストでもリーダビリティの値は学習者に依存すると考え、学習者さらに学習時間ごとにリーダビリティ式を REX に実装した。パラメータは英語テキストに存在する総単語数や異語数、名詞といった品詞の総数や割合を示し、計 100 以上のパラメータ値を事前に計算し、パラメータと Rating 値からリーダビリティに関連度が高い上位 3 つのパラメータを導き、重回帰方程式を作成する。式はリーディング回数が 10 回ごとに更新される。

3. 推測手法

図 1 に示した学習の流れに推測手法を適用することで、より早く英単語の既知か未知かを判定できるようにする。具体的には、学習者本人だけでなく、類似した語彙力を有する他の学習者を探索し、その語彙表を参考にする。参考にする程度はコサイン類似度によって重み付けされる。つまり、類似度が負の値を取る際には逆相関があるとし、その語彙表を（反転の意味で）参考にする。このように、本推測手法では、自分自身が影響する既知割合、そして他学習者が影響する語彙力類似度を用いてハイブリッドに改善を行っている。各術語の定義を下に示す：

3.1. 既知割合

リーディングや英単語ゲームでラベル付けされた単語を用いて、学習者ごと SVL レベルごとに算出される、英単語群が当該学習者に既知である割合。

3.2. 語彙力類似度

二人の学習者の単語リストを参照し、各単語のラベル値をもとに計算するコサイン類似度。

4. 実験

4.1. 目的

推測手法によって学習者の単語リストに早期にラベル付けされることを確認し、かつ精度の可否を調査する。

4.2. 概要

対象とする学習者は某大学の大学院生または学部生 9 名である。リーディングを行い、読了後に自己判定し、Rating 値を決定してもらう。読了数は各人 100 件とした。リーディングに使用するテキストは mainichi.jp の英語ニュースと「Jaremaga」（英語学習のためのメルマガ）を利用する。英単語ゲームは上述のように実施可能であるが、本実験ではリーディングを行ってもらうことが主となるため任意とした。

リーディング 100 件読了後、推測手法で得たラベル付け値と実際のデータの値の近接度を 240 問の英単語テストで評価する。出題方法は英単語テストと同様に 4 択問題とした。出題する問題に関しては SVL レベル 12 個×20 問とし、出題する問題は学習者ごとに未登場の英単語（つまり推測手法でラベル

付けしたもの）とした。各単語に解答後、自信度の選択肢（自信あり、やや自信あり、やや自信なし、自信なし）から選択してもらう。本実験で得られた結果から未知・既知の一致率を各学習者、各レベルで計算した。値が高ければ推測手法によるラベル付けはユーザの語彙力に近似しているといえる。推測手法でラベル付けした単語総数を n 、単語 W_k の正誤情報 $C(W_k)$ 、同じく推測手法によるラベル付け値を $L(W_k)$ としたときの一致率を次に定義する：

$$\text{一致率} = \frac{\sum_{k=1}^n (1 - |C(W_k) - L(W_k)|)}{n}$$

5. 実験結果と考察

過去の推測手法で最も良い結果である既知割合を用いる手法と本発表で提案した推測手法を比較する。その際、「自信あり」、「やや自信あり」で不正解したもの、「自信なし」、「やや自信なし」で正解したものを信頼性に足らないとして除去した。T 検定を行った結果を下の表 1 に示す。

表 1 一致率比較

学習者	既知割合	提案手法
1	0.664	0.664
2	0.584	0.598
3	0.621	0.641
4	0.623	0.683*
5	0.661	0.726*
6	0.779	0.793*
7	0.764*	0.717
8	0.670	0.690
9	0.654	0.703*

表より、推測手法では一致率が高い学習者が多く（色付）、さらに有意差が多く認められている（太字）。本推測手法は先行研究に比べ、学習者の実際の語彙力を正しく推定できていると伺える。

6. まとめ、今後の展望

本稿では、学習者特性に応じた難易度の英語テキスト提供を目指す REX の改良を目指し、語彙力推測手法アルゴリズムを開発し、有効性が確認された。今後は推測アルゴリズムによる効果をより分析することが期待される。

参考文献

- (1) 福井正康, 小篠敏明, “単語と熟語の難易度を考慮した英文リーダビリティ指標の作成法”, 日本教育情報学会学会誌, Vol.24, No.3, pp. 15-22 (2009)
- (2) 土屋武久, “リーダビリティ決定要因についての位置考察-大学リーディング教材の計量的分析から-”, JACET 全国大会要綱, Vol.37, pp. 50-51, (1998)
- (3) 大城敬人, 宮崎佳典, “REX を活用した英語リーダビリティ式のパーソナライゼーション手法の提案”, 外国語教育メディア学会 (LET) 中部支部紀要, Vol.24, pp.25-34 (2013)