

# オンライン英文多読学習を通じた学習者用語彙リストの推定と評価 Estimating English Vocabularies of Individual Learners for Provision of Their Suitable English Texts

白須 直樹<sup>\*1</sup>, 宮崎 佳典<sup>\*2</sup>

Naoki SHIRASU<sup>\*1</sup>, Yoshinori MIYAZAKI<sup>\*2</sup>

<sup>\*1</sup> 静岡大学 情報学部

<sup>\*1</sup> Faculty of Informatics, Shizuoka University

<sup>\*2</sup> 静岡大学学術院 情報学領域

<sup>\*2</sup> College of Informatics, Shizuoka University

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

あらまし：我々研究グループは学習者のレベルに合った英文テキストの提供を行う多読学習システム REX の開発を行っている。REX は英単語リストに対して既知・未知情報を学習者別に記録した学習者用語彙リストを更新・活用するが、リスト内の単語を網羅するには膨大な時間を要する。そこで派生語を含めた英単語群であるワードファミリーに着目した語彙リスト推定アルゴリズムを考案した。本発表ではパイロット実験を行った結果と考察について報告する。

キーワード：e-Learning, 多読学習, 語彙リスト, 推測手法, リーダビリティ

## 1. はじめに

著者らは、学習者のレベルに合った英語テキスト提供を行う英文多読学習システムとして REX (Reading Exercise) の開発に従事している。学習者のレベルに合ったテキストを提供するには学習者の語彙力情報が必要となるため、REX では英単語リストに対して既知・未知情報を学習者別に記録した学習者用語彙リストを更新・活用している。リスト内の単語を網羅するには膨大な時間を要するため、鈴木ら<sup>(1)</sup>は学習者の語彙推定アルゴリズムを考案した。しかし同手法だけで学習者の語彙を全て推定するのは負担が大きく、推定精度の点でも相補的な別手法が望まれる。そこで著者らは派生語を含めた英単語群であるワードファミリーに着目した語彙推定アルゴリズムを考案した。本発表では、REX が学習者のレベルに合った英語テキストを提供する方法及び英単語の既知・未知情報の記録方法を述べ、提案手法の有効性を測る実験の結果や考察について述べる。

先行研究として終元<sup>(2)</sup>は、多読の効果には英文のリーディングスピードの向上や、自律的な学習意欲向上があることを実践により明らかにしている。

## 2. REX の概要

REX における学習と記録の流れを図 1 に示す。学習者はリーディング機能にてテキストを読了し、テキストの難易度を 6 段階で自己判定する。読了後、英単語ゲームの利用チケットを獲得し、任意に利用する。語彙リストの記録値を元にテキスト中の未知英単語割合 (難語率) を算出し、テキスト情報と自己判定結果を元にリーダビリティ式を作成する。生成された式から算出されるテキストのリーダビリティ値 (可読性) より、次に提供するテキストを選定する。語彙リストへの記録値は 0 (未知) ~1 (既知) の値を取り、SVL12000 リスト (12,000 英単語を 12 レベル×1,000 語に分類) に対して作成している。学

習者用語彙リストが更新される機会は 3 通り存在し、リーディング、英単語ゲーム実行時と語彙推定アルゴリズムによるものである。リーディング中は文中単語が未知語/既知語によって色分けされており、クリックによってトグルすることができる。英単語ゲームは英単語訳を四択より選ぶ問題であり、自信度を指定して解答し、表 1 による記録がなされる。

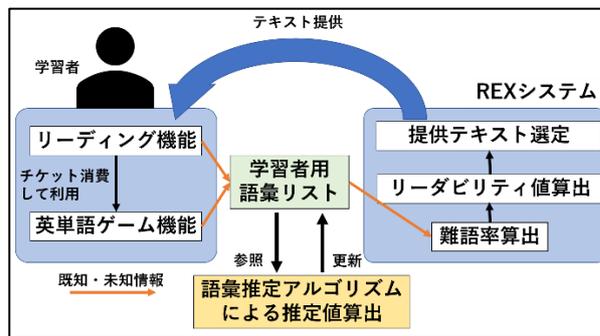


図 1：学習と記録の流れ

表 1：英単語ゲームによる記録値

	自信あり	やや自信あり	やや自信なし	自信なし
正解	1	0.75	0.5	0.25
不正解	保持	保持	0	0

語彙推定アルゴリズムは、他学習者との語彙力の類似度を参考にし、自身の学習者用語彙リストに記録されていない単語を自他学習者の記録結果や既知割合をもとに算出、記録するものである (以下、既存推測手法)。学習者  $U$  の SVL におけるレベル  $l$  内の記録済み英単語群  $\{W_{l,1}, W_{l,2}, \dots, W_{l,n}\}$  に対応する記録値  $L(U, W_{l,k})$  を元に、既知割合を次の式から計算する：

$$R(U, l) = \frac{\sum_{k=1}^n L(U, W_{l,k})}{n}$$

次に、学習者 $U_1, U_2$ に共通してリーディング・英単語ゲームで記録済みの SVL 内レベル $l$ の英単語群 $\{W_{l,1}, W_{l,2}, \dots, W_{l,m}\}$ から語彙能力類似度 $S(U_1, U_2, l)$ を次式で計算する( $m$ が 0 の場合は計算しない)：

$$S(U_1, U_2, l) = \frac{\sum_{k=1}^m 1 - |L(U_1, W_{l,k}) - L(U_2, W_{l,k})|}{m}$$

上記数式で算出された語彙能力類似度から、より参考にする他学習者の記録値(リーディング・英単語ゲームの記録または既知割合の値)に重み付けを行い、未記録単語に対して記録する。

### 3. ワードファミリーを用いた推測手法

著者らはワードファミリー(WF)に着目した推測手法(以下、提案推測手法)を考案した。WFグループはNation<sup>(3)</sup>が提供しているWFリストを利用する。WFの例を表2に示す。

表2：ワードファミリー(WF)の例

access, accessed, accesses, accessibility, accessible, accessing, inaccessibility, inaccessible
---

提案推測手法では、表2のようなWF内のリーディング・英単語ゲームで記録された単語から既知割合を計算し、同WFの未記録単語に対して記録する。あるWFに対して、学習者 $U$ によるリーディング・英単語ゲームの記録済み英単語群 $\{W_{g,1}, W_{g,2}, \dots, W_{g,n}\}$ の記録値を $L(U, W_{g,k})$ とした時、提案推測手法は次式で計算する：

$$R(U, WF) = \frac{\sum_{k=1}^n L(U, W_{g,k})}{n}$$

リーディング・英単語ゲーム・提案推測手法で記録がされない場合は既存推測手法で記録する。

### 4. 実験

提案推測手法が既存推測手法と相補的關係を築けているのか、実験によって調査した。某大学の大学生または大学院生9名を対象にREXで平均約360語の英語ニュース記事を200件読了してもらい、その後240問テストを実施した。英単語テスト同様4択問題とし、SVL各レベル×20問出題する。本実験で得られた結果から、学習者用語彙リストへの記録値と240問テストに対する正誤との一致率を算出する。240問テストから“正解し、かつ(やや自信なし、または自信なし)”と“不正解し、かつ(自信あり、またはやや自信あり)”であった単語を除き、学習者 $U$ によって提案推測手法で記録された単語の総数を $s(U)$ 、単語 $W_k$ の記録値を $L(U, W_k)$ 、 $W_k$ の問題に対する正誤情報(正解で1、不正解で0)を $C(U, W_k)$ とすると、一致率の計算式は次式で表される：

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

既存推測手法と提案(+既存)推測手法それぞれの一致率についてt検定を行った結果を表3に示す。

表3：t検定による一致率比較

学習者	提案手法 +既存手法	既存手法 のみ	P値
1	0.951***	0.848	0.000
2	0.870*	0.833	0.037
3	0.871**	0.816	0.001
4	0.943***	0.859	0.000
5	0.889***	0.809	0.000
6	0.945***	0.869	0.000
7	0.858***	0.794	0.000
8	0.886***	0.817	0.000
9	0.707	0.678	0.098

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

提案推測手法と既存推測手法の併用は一致率が既存推測手法のみより高い学習者が多く、有意差が認められた。また、学習者のレベルに合ったテキスト提供がされたかについて、各学習者の自己判定結果とREXが算出したリーダビリティ値との平均距離を算出した。その結果を表4に示す。

表4：自己判定結果とリーダビリティ値との差

学習者	計算結果	学習者	計算結果
1	0.817	6	0.969
2	0.894	7	0.895
3	0.900	8	0.879
4	1.034	9	1.508
5	0.705		

表4の通り、学習者9については、レベルに合ったテキスト提供が他学習者よりできていないことが考えられ、表3で有意差が出ていないことが影響していることが考えられる。

### 5. まとめ、今後の展望

本稿では、学習者の英語語彙を推測する際、既存の推測手法アルゴリズムに加えワードファミリーの概念を活用した推測手法を提案した。実験の結果、提案推測手法と既存推測手法の併用が有意に精度の向上に繋がっていることを確認した。

### 参考文献

- (1) 鈴木峻丸, 宮崎佳典, “最適な難易度の英語テキスト提供を目指すための学習者の英単語語彙推測アルゴリズムの考案”, JeLA学会誌, Vol.19, pp. 53-61 (2019)
- (2) 終元弘文, “第二言語習得における多読の意義及び多読指導実践とその効果検証”, 関西外国語短期大学部研究論集, pp. 213-231 (2019)
- (3) I. S. P. Nation, “TheBNC/COCAwordfamilylists”, [https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC\\_COCA-word-family-lists.pdf](https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf)