オンライン英文リーディング学習のための学習者用語彙推定アルゴリズムの改良

An Attempt to Improve Estimation of English Vocabularies of Individual Learners for Their Online English Reading Studies

中村 亮子[*1], 宮崎 佳典[*2], 法月 健[*3], 田中 省作[*4]
Ryoko NAKAMURA[*1], Yoshinori MIYAZAKI[*2], Ken NORIZUKI[*3], Shosaku TANAKA[*4]
[*1] 静岡大学 情報学部
[*1]Faculty of Informatics, Shizuoka University
[*2] 静岡大学学術院 情報学領域
[*2]College of Informatics, Shizuoka University
[*3] 静岡産業大学 情報学部
[*3]School of Information Studies, Shizuoka Sangyo University
[*4] 立命館大学 文学部
[*4]College of Letters, Ritsumeikan University
Email: nakamura.ryoko.18@shizuoka.ac.jp

あらまし：我々は各学習者に対して適度なレベルのテキストを提供する英文リーディング学習システム REX を開発している．これを実現するため，共通の英単語リストに対し各学習者の行動履歴からリスト内の英単語に既知・未知を記録することで学習者用語彙リストを作成している．同語彙リスト生成にはワードファミリーに着目した推定法も採用されており，本発表では改良案を提示し，より高精度のパーソナライズされた語彙リスト提供を目指す．
キーワード：e-Learning，リーディング学習，語彙リスト，推測手法，リーダビリティ

1. はじめに

著者らは各学習者のレベルに合った英語テキストの提供を行うオンライン英文リーディング学習システム 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ためには，各学習者の英単語語彙力情報が必要となる．REX では，学習者の使用過程で個々の英単語に対して既知・未知を記録することで語彙力を測っているが，英単語リスト内の全単語を網羅するには膨大な時間を要し，鈴木ら[(1)]は学習者の語彙推定アルゴリズムを考案した．しかし同手法のみで学習者の語彙を全て推定するのは負担が大きく，白須ら[(2)]はワードファミリー（見出し語とその変化系および密接に関連する派生語からなる単語群，以下 WF）に着目した語彙推定アルゴリズムを考案し，語彙推定精度の向上が確認された．本発表では，同語彙推定アルゴリズムの改良案を与え，さらなる推定能力向上を図る．

2. REX の概要

REX では SVL12000（12,000 の英単語を 12 のレベル×1,000 単語に分類）を基にしたリストに対してリーディング，英単語ゲーム，語彙推定アルゴリズムの 3 つを利用しラベル付けを行い，語彙リストへ 0（未知）~1（既知）の範囲で値を記録する．学習・記録の流れについて図 1 に示す．学習者はリーディング機能を利用しテキストを読了する．リーディングではテキスト内の単語が未知語・既知語で色分けされ，クリックで切り替えが可能である．読了後，テキストの難易度を 6 段階から選択し自己評価を行う．作業終了後，バトルチケットを得ることができ，任意で英単語ゲームを行う．英単語ゲームでは出題英単語の意味を 4 択から 1 つ，4 段階の自信度を 1 つ選択し，正誤と自信度に応じて表 1 のようにラベル付けが行われる．リーディングや英単語ゲーム等によるラベル付けを行った学習者用語彙リストを元に，テキストに含まれる未知英単語割合（難語率）を算出し，テキスト情報と自己評価結果からリーダビリティ式を作成する．式から算出されるリーダビリティ値より次に提供するテキストを選定する．

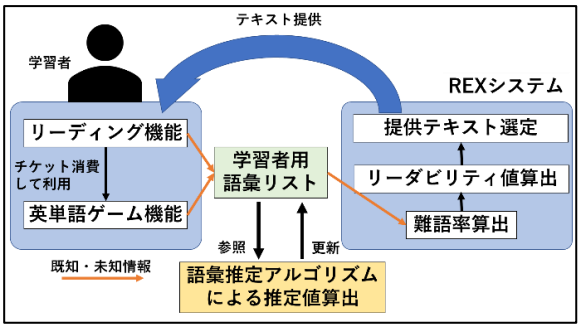

図 1：　学習・記録の流れ

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	元の値保持	元の値保持	0	0

語彙推定アルゴリズムは学習者語彙リストの未記録単語に対して自他学習者の記録値や既知割合，それに学習者間の語彙類似度をもとに算出する[(1)]．さらに WF 内の記録済単語から既知割合を計算して同 WF 内の未記録単語に記録する[(2)]．WF グループには

Nation[3]が提供している 50, 896 個の WF リストを使用しており，上述 SVL12000 の単語と共通集合を取ることで今回定義している．WF の例を表 2 に示す：

表 2：WF の例

happy, happier, happiest, happily, happiness, unhappier, unhappiest, unhappily, unhappiness, unhappy

(1)では学習者U,V間の語彙類似度$S(U,V)$を以下のように定義した．なおU,Vで共通の記録済単語群を$\{w_i\}\ (i=1,\dots,m)$,, 学習者Xのw_iに対する記録値を$L(X,w_i)$とする：

$$S(U,V)=\frac{\sum_{k=1}^{m}1-|L(U,w_k)-L(V,w_k)|}{m}$$

(2)については具体的に以下のような計算式が適用される．WF グループgに属する記録済単語群（空でないとする）を$\{W_{g,i}\}\ (i=1,\dots,n, 1\le n\le|g|)$, 学習者$U$の$W_{g,i}$に対する記録値を同様に$L(U,W_{g,i})$としたとき，学習者$U$のグループ$g$に対する既知割合$R(U,g)$を次式で定義する．

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

3. 語彙力推測改良手法

発表者らは，WF リストを利用し，自学習者のみの WF 内既知割合を WF 内未記録単語に記録する(2)を発展させ，状況に応じて他学習者の WF 既知割合の値ならびに上述の語彙類似度を取り入れた手法を考案した．(2)では自学習者のデータのみを利用しているため，自学習者の記録データ数が少ない段階では，少数の単語に対する記録値が大きな影響力を持つことになる．故，提案推測手法では自学習者の記録データが十分でない場合に，記録データが一定割合以上あり, かつ語彙類似度の高い（あるいは低い）他学習者のデータを利用する．今回は試行的に基準を設け，①WF グループの自学習者の記録済み単語割合が 0.3 以下，あるいは②WF グループの自学習者の記録済み単語割合が 0.3 より大きく 0.5 以下であり, かつ WF グループの既知割合が 0.4 以上 0.7 以下である, ことを適用条件としている. 提案推測手法では自他学習者間の類似度を考慮した. 具体的には, 学習者Uと他学習者Vの語彙能力類似度$S(U,V)$はその中央値である 0.5 に近いほど独立性は高く，重み付け係数$WE_{U,V}$を$WE_{U,V}=2|S(U,V)-0.5|$とした. また，0 に近い語彙類似度は逆相関を意味し，その場合は WF 既知割合を反転させた．学習者Uと他学習者Vのグループgへの重み係数$A_{U,V,g}$を

$$A_{U,V,g}=\begin{cases}R(V,g):\ S(U,V)\ge 0.5\\ 1-R(V,g):\ S(U,V)<0.5\end{cases}$$

とした．これらを用いて，次式により学習者Uの WF グループgに既知割合$R_{proposed}(U,g)$を与える．

$$R_{proposed}(U,g)=\frac{\sum_V A_{U,V,g}WE_{U,V}}{\sum_V WE_{U,V}}$$

学習者データ例を表 3，これに対し提案手法を適応した例を図 2，図 3 に示す．手法を適応させるとnull 値に記録値を与える．手法(2)と比して，学習者 1 の記録数が少ない場合に，語彙類似度の高い学習者 2，3 の値を高割合で参照していることがわかる．

表 3：学習者 1, 2, 3 例示データ

学習者	WF1 単語数:10		WF2 単語数:10		学習者間類似度		
	既知割合	記録数	既知割合	記録数	1	2	3
1	0.000	1	0.500	4	1	0.9	0.1
2	0.750	8	0.833	6	0.9	1	0.7
3	0.125	7	0.333	9	0.1	0.7	1

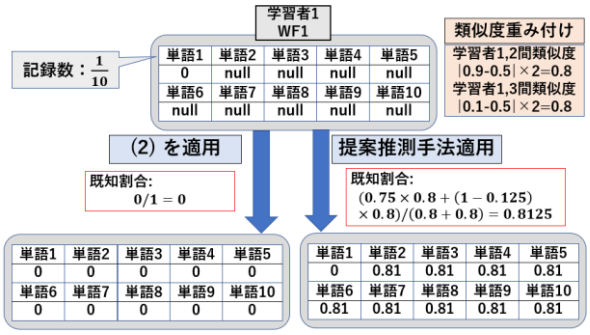

図 2：　手法適応例 1

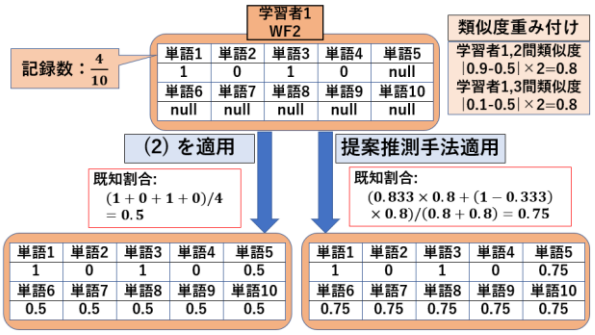

図 3：　手法適応例 2

4. 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，WF を利用した白須手法に加え他学習者の WF 既知割合も参考にした推測手法を提案した．今後の展望として提案した推測手法についてアルゴリズムの適合性を実験を通して検証してゆく所存である．

参考文献

(1) 鈴木竣丸，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, Vol1.9, pp.53-61 (2019)

(2) 白須直樹，宮崎佳典，“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”,,2019 年度学生研究発表会,, 教育システム情報学会, (2019)

(3) I.,S.,P.,Nation,,“TheBNC/COCAwordfamilylists”,,https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/paul-nations-publications/publications/documents/Information-on-the-BNC_COCA-word-family-lists.pdf