K-051

英文読解能力測定モデルへの熟語難易度の導入

奥村将成† 吉見毅彦† 南條浩輝† 小谷克則‡

1 はじめに

近年，インターネットを中心に大量の電子化された英語のテキストに触れる機会が増加しており，そのような大量の英文テキストを正確に読みこなすためには，英語の読解能力が重要となる．このような背景から，英語の読解能力の向上を目的とした学習支援がますます必要になってきている．

語学教育において，学習者の読解能力を向上させるためには，教師が学習者の読解能力を把握する必要がある．そうすることにより，学習者の能力に応じた学習支援が可能となる．このような状況を踏まえ，先行研究である文献 [1] では，難しい文を速く読めるほど読解能力が高いと仮定して，英語の読解能力を自動的に測定するモデルを提案している．具体的には，2 章で述べる様々な言語的特徴によりテキストの読み易さを判断する．そして，テキストの言語的特徴（テキストの読み易さ）と，学習者がテキストを読み終わるのにかかった時間の 2 つの指標を基に英文の読解能力を測定するというモデルである．

ただ，文献 [1] では，英文の読み易さを測るための言語的特徴の十分な検討がされていない．そこで，本論文では，文献 [1] で取り入れていた言語的特徴をさらに検討し，言語的特徴量として新たに熟語難易度を利用することの有効性を調査する．

2 関連研究

1 章で述べたように，文献 [1] は，テキストの言語的特徴と学習者の読解時間を基に自動的に英文の読解能力を測定するモデルを提案している．この際，文献 [1] で用いた言語的特徴の種類として，単語の長さ，単語難易度，単語の語義数，構文木の節点数，代名詞の数がある．これらの言語的特徴について詳しく説明する．

まず，単語の長さは，語彙の難しさを表していると考えられる．次に，単語難易度であるが，これは単語の長さだけでは単語の難しさを測れない（単語の長さが短い場合であっても難しく感じる）場合もあることから，日本語を母語とする英語学習者にとって適切な語彙の難しさを測るために考慮する．さらに，単語の語義数である．これは，比較的簡単な単語（例えば，get や make など）であっても多くの意味を持っている場合，英語学習者にとっては必ずしも難易度の低い単語ではない．そのため，単語の語義数を考慮する．そして，構文木の節点数であるが，これは節点数が多いほど複雑な文であると考えられるためである．最後に，代名詞の数である．代名詞による照応の処理は，代名詞が指す指示対称を同定する必要があるため，読解能力に影響を及ぼす．そのため，代名詞の数を考慮している．

このように，文献 [1] では様々な言語的特徴を考慮しているが，まだ考慮していないものもある．例えば，音節数や熟語難易度，さらに内容語の重複 [2] などが挙げられる．本論文では，これらのうち熟語難易度に着目する．

3 熟語難易度の導入

英文を読解する際に英文テキストに熟語が入っていると，英語学習者にとっては，英文を読解することが難しく感じる場合や，逆に英文を読解しやすくなる場合がある．このように，熟語難易度は学習者の読解に非常に影響を与えるため，熟語難易度を言語的特徴に含めることにより，英文の難易度を適切に判定することが可能となると考えられる．

本研究では，文献 [3] の熟語難易度の取得方法を用いて，英文テキストに含まれる熟語難易度を取得する．具体的には，文献 [3] で作成された熟語辞書を用いて英文テキストから熟語を検出する．そして，熟語辞書を参照し，検出された熟語に対応する難易度を取得する．

例えば，「He still writes to me from time to time, but I always ignore it.」という入力文の場合，”wite to”と”from * to”と”from time to time”という熟語構文が検出され，それぞれの熟語難易度である 3.7，2.3，5.0 を足し合わせた値（11.0）が熟語難易度となる．

上記の方法により，英文テキストに出現した熟語の難易度を求める．そして，その熟語難易度の合計値を測定モデルへ新しく取り込む．

†龍谷大学 理工学研究科
‡関西外国語大学 外国語学部

4　英文読解能力測定モデルの評価実験

4.1　実験方法

本論文では，言語的特徴量として新たに熟語難易度を利用することの有効性を調査するために従来モデルとの比較実験を行った．具体的には，文献 [1] で作成された 451 件の事例集合を用いて 5 分割交差検定を行う．また，読解能力を示す指標としては TOEIC のリーディングセクションのスコア（以下，TOEIC スコア）を用いる．さらに，その際の測定モデルの測定精度として，次式で計算される誤り率を用いる．

$$\text{誤り率 (\%)} = \frac{|\,\text{測定値} - \text{実測値}\,|}{\text{実測値}} \times 100 \qquad (1)$$

式 (1) において，測定値は読解能力測定モデルによって測定される TOEIC スコアであり，実測値は学習者の実際の TOEIC スコアである．モデルの構築には，サポートベクターマシンによる回帰 [4] を用い，サポートベクターマシンのカーネル関数は，線形カーネルと多項式カーネル（1 次，2 次，3 次），RBF カーネルを用いた．また，サポートベクターマシンのソフトマージンの値を 0.001，0.01，0.1，1，10，100，1000 と変更し，読解能力を測定した際の誤り率が最も低いものを提案モデルの誤り率とした．

4.2　実験結果

熟語難易度を考慮しない文献 [1] の測定モデルの誤り率の分布と熟語難易度を考慮した提案モデルの誤り率の分布を図 1 に示す．

実験の結果，熟語難易度を考慮しない場合は，測定モデルの誤り率の中央値が 19.3 %でレンジが 235.7 であった．これは，ソフトマージンが 1000 で RBF カーネルを用いた際に誤り率が最も低かった．それに対して，熟語難易度を考慮した本研究の提案モデルの誤り率の中央値は 18.3 %でレンジは 232.7 であった．これは，ソフトマージンが 100 で 1 次の多項式カーネルを用いた際に誤り率が最も低かった．

4.3　考察

文献 [1] の測定モデルの誤り率の分布と本研究の提案モデルの誤り率の分布より，どちらも誤り率が高くなるにつれて，その階級に属する事例の件数が減少していく傾向があることが分かる．また，本研究の提案モデルは，文献 [1] の測定モデルと比べると，誤り率が 0 %以上 10 %未満の事例数は減少したが，誤り率が 90 %以上を示していた事例数も減少した．結果として，誤り率が高い事例数が減少し，10 %以上 20 %未満の誤り率の事例数が増加した．

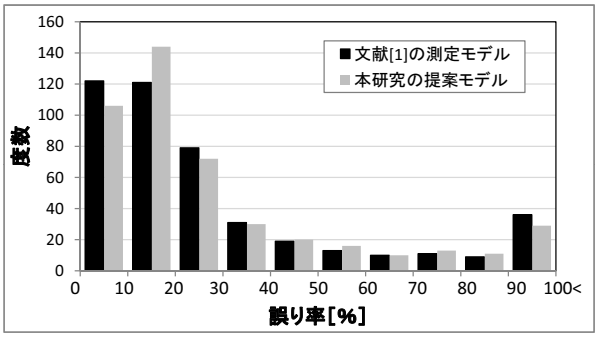

図 1: 文献 [1] の測定モデルと提案モデルの誤り率

従って，英文読解能力を測定する際に，言語的特徴に熟語難易度を入れると全体的に誤り率が低減されることが確認できた．

5　おわりに

本研究は，英文読解能力測定モデルの言語的特徴の検討を行った．実験では，熟語難易度を言語的特徴として加えた場合と加えない場合で誤り率を比較し検証した．

実験の結果，言語的特徴として，熟語難易度を素性に加えた提案モデルの方が英文読解能力を予測した際の誤り率が低減されることが分かった．

今後の課題としては，より適切な熟語難易度の取得方法や，学習者の特徴の新たな素性を検討する必要がある．

参考文献

[1] Katunori Kotani, Takehiko Yoshimi, and Hitoshi Isahara. A prediction model of foreign language reading proficiency based on reading time and text complexity. 7(10), pp. 1–9. US-China Education Review, 2010.

[2] SCOTT A. CROSSLEY, JERRY GREENFIELD, and DANIELLE S. McNAMARA. Assessing text readability using cognitively based indices. 3(42), pp. 475–493. TESOL QUARTERLY, 2008.

[3] 小篠敏明, 福井正康. 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. 24(3), pp. 15–22. 日本教育情報学会誌, 2009.

[4] 小野田崇. 知の科学 サポートベクターマシン, オーム社. 2007.