

英語学習者の英文読解能力推定のための読解時間予測法 Prediction Method of Reading Time for Assessing EFL-Learners' Reading Ability

吉見毅彦* 小谷克則† 九津見毅‡ 佐田いち子‡ 井佐原均§
T. Yoshimi K. Kotani T. Kutsumi I. Sata H. Isahara

1 はじめに

語学教育においては、一人一人の学習者について、学習者が十分に習得していない語彙や文法項目を教員が正確に把握することが重要である。学習者の未習得項目を特定する方法として、理解度テストによる方法、読解時間あるいは読解速度の計測による方法などがある*1。

近年、英語学習者の英文読解能力の評価指標として英文の読解時間あるいは読解速度を用いることの妥当性が確認されている [8, 4]。読解時間の計測による方法では、文の読解に要する標準的な時間を設定し、この標準的読解時間と学習者が実際に文を読むのに要した読解時間とを比較する。そして、標準的読解時間よりも実際の読解時間のほうが一定以上長ければ、その文にはその学習者の未習得項目が含まれている可能性が高いと判断する*2。

読解時間の計測による方法では、学習に使用する文の標準的読解時間を設定しなければならないが、多くの文について教員が標準的読解時間を設定することはコストが高いため、このコストを軽減する必要がある。また、学習者が自主学習を行なう場合にも自分の未習得項目の特定を可能にするためには、標準的読解時間の設定を教員に頼らなくてもよいようにする必要がある。

そこで、本稿では、学習者の未習得項目の特定へのアプローチとして、文の難易度と学習者の読解能力に基づいてその学習者にとっての標準的な文の読解時間を自動的に予測する読解時間予測モデルを提案する。提案モデルは、読解時間を従属変数とし、2節で述べる要因を独立変数として重回帰分析を行なうことによって得られる。評価実験の結果、提案モデルは22.9%の誤差率で文の読解時間を予測できることが明らかになった。

* 龍谷大学 / 情報通信研究機構

† 関西外国語大学

‡ シャープ(株)

§ 情報通信研究機構

*1 読解時間の計測による方法は理解度テストによる方法と排他的ではなく、学習者の未習得項目の特定において両者の併用は可能であると考えられる。

*2 学習者の未習得項目を特定するためには、テキストという比較的大きい単位の読解時間ではなく、文という比較的小さい単位の読解時間を予測する必要がある。

2 着目した要因

文の読解時間に影響を及ぼす要因は、言語に関する要因(言語的要因)と読解者に関する要因(読解者要因)に分けられる。言語的要因は、語彙的要因や統語的要因、談話的要因から成り、さらに各要因として様々なものが考えられる。本研究では、読解時間予測モデルの構築に必要な言語的要因を文から自動的に抽出することを前提としているため、現状の自然言語処理技術で扱えるものに着目する。要因の抽出に人手をかけないことで教材作成が効率よく行なえるという利点が得られる。

2.1 語彙的要因

本稿で提案する読解時間予測モデルでは、語彙の難易度を上げる要因として、単語の長さや単語の基本語彙レベルを考える。

英文の難易度を測定するためによく利用されている指標 [2] では、長い単語は馴染みが薄い単語であることが多く、単語の長さは語彙の難しさを表わしていると考えられている。本研究では、この経験則に従って、単語の長さを文字単位で測り、文中の全単語の長さの平均値を語彙的要因として読解時間予測モデルに取り込む。

単語の長さだけで語彙の難しさを測ることができるのは英語母語話者の場合であり、英語学習者にとっては長さが短い単語であっても難しい場合がある [6]。このため、本研究では、日本語を母語とする英語学習者の場合にも適切に語彙の難しさを測るために、「JACET 基本語 4000」などに基づいて作成された語彙難易度区分表*3も利用する。具体的には、文中の全単語について、この語彙難易度区分表から語彙難易度を取得し、その合計値を語彙的要因として取り込む。

2.2 統語的要因

統語的難易度を上げる要因として、文の長さや構文木の節点の数を考える。

英文難易度測定のための指標 [2] では、統語的難易度を算出する際に文の長さが利用されている。本研究でも、文の長さを、統語的難易度を近似的に測るための尺度の一つとして用いる。

文の長さは、文の構文木を考えた場合、その幅に相当する。

*3 http://www1.kamakuranet.ne.jp/someya/wlc/wlc_manual.html

構文木の幅以外に、文の統語的難易度を表わす要因として、構文木の高さや構文木の節点数などが考えられる。本研究では、これらのうち構文木の節点数を統語的難易度を上げる要因として取り込む。構文木の節点数が人間の読解(特に構文解析)に影響を及ぼすことは、ガーデンパス文に関する研究[3]などによって確認されている。構文木の作成には、“Apple Pie Parser”^{*4}を利用する。

2.3 談話的要因

テキストの読解において処理しなければならない談話的言語現象のうち代表的なものは照応である。照応には代名詞によるものや名詞句によるものなどがあるが、特に代名詞による照応の処理は、代名詞が指す指示対象を同定する必要があるため、読解時間に影響を及ぼす。このようなことから、文に現われる代名詞の数を談話的要因として読解時間予測モデルに取り込む。

2.4 読解者要因

読解者要因としては、読解者のテキストへの興味や読解への意欲、非言語的知識、読解能力などが考えられる。これらのうち本研究では読解者の英文読解能力を読解時間予測モデルに取り込む。英文読解能力は、テストの得点や単語認識時間、文法性判断時間などによって測定することができるが、本研究では読解者が受験したTOEIC(Test of English for International Communications)のリーディングセクションの得点を利用する。TOEICはリーディングセクションとリスニングセクションから成るが、本研究では読解時間予測モデルの構築が目的であるため、リーディングセクションの得点(以下、TOEIC読解スコアと呼ぶ)を用いる。

読解者のテキストへの興味や読解への意欲などを読解時間予測モデルに取り込むためには、アンケート調査を実施すればよい。しかし、3節で述べる読解時間データを収集する際にアンケート調査を行なわなかった。このため、テキストへの興味や読解への意欲などの導入は今後の課題とする。なお、テキストへの興味や読解への意欲は、あまり速読力に影響を及ぼさないとの報告がある[4]。

3 事例集合の作成

本研究で重回帰分析に用いる一つの事例は、2節で述べたように、文の語彙的、統語的、談話的要因とその文を読んだ読解者のTOEIC読解スコアから構成される。事例集合を作成するために、TOEICで使用されるテキストを日本語母語話者が読んだときの読解時間と、テキストの内容に関する設問への解答の正答率を次のような方法で収集した。

読解対象テキストとして、TOEIC準拠の問題集に掲載されているTOEICのリーディングセクションの読解問題に相当する84テキストを利用した。この84テキストから、7テキストから成るテキストセットと14テキストから成るテキストセットを作成した。

TOEICを受験したことがある日本語母語話者を、報酬を支払うことを明示して募集した。応募者のうちTOEIC受験時期が募集時期より1年以内である64名を読解者とした。64名のうち31名それぞれに7テキストから成るテキストセットを割り当て、また、残りの33名それぞれに14テキストから成るテキストセットを割り当てた。

読解時間と設問への解答の記録には、コンピュータ上の読解時間・解答記録ツール[9]を使用した。読解時間は文ごとに記録した。読解時間の計測は1/100秒単位で行なった。

読解開始前に読解者への指示として、(1)テキスト読解後にテキストの内容に関する設問に解答すること、(2)設問に解答できる程度にテキストを理解すること、(3)読解時間に制限は設けないので、自分のペースでテキストを読み進めること、(4)読解時間・解答記録ツールの操作に慣れるため、事前に練習用テキストを用いて練習することを伝えた。読解時間を計測していることは伝えなかった。

以上のような方法で得られたデータから、(1)テキストの内容に関する設問への解答の正答率が70%未満であった読解者のデータと、(2)文に含まれる単語数を読解時間で割って求めた読解速度が60WPM(Words Per Minute)以下か200WPM以上であるデータを除外した。その結果、1807件の事例が得られた。これらの事例に含まれる異なりテキスト数は80テキスト、異なり文数は448文、異なり読解者数は61人であった。読解者61人のTOEIC読解スコアの分布を表1に示す。

表1 読解者のTOEIC読解スコアSCRの度数分布表

階級	度数
0点 ≤ SCR < 50点	0
50点 ≤ SCR < 100点	0
100点 ≤ SCR < 150点	3
150点 ≤ SCR < 200点	4
200点 ≤ SCR < 250点	11
250点 ≤ SCR < 300点	10
300点 ≤ SCR < 350点	6
350点 ≤ SCR < 400点	11
400点 ≤ SCR < 450点	11
450点 ≤ SCR < 500点	5

正答率が70%未満であった読解者のデータを除外した理由は、70%未満の正答率ではテキストの理解が不十分であると考えられるためである。また、読解時間が60WPM以下であるデータを除外した理由は、読解者へは読解時間に制限を設けないことを伝えたので、テキストの内容に関する設問に正解するために必要以上に慎重になった可能性もあると考えられるためであり、200WPM以上であるデータの除外は、英語母語話者による平均英文読解速度が200WPMから300WPM程度であること[1]を参考にして行なったものである^{*5}。

^{*4} <http://nlp.cs.nyu.edu/app/>

^{*5} 読解速度の異常値の原因として読解時間・解答記録ツールの操作ミス

4 評価実験

本節では、2節で述べた要因をすべて独立変数とし、読解時間を従属変数として重回帰分析を行なうことによって得られる重回帰式(提案モデル)の有効性を検証するために行なった実験について述べる。

4.1 実験方法

3節で述べた1807件の事例のうち1627件を学習事例集合として重回帰分析を行ない、残りの180件を試験事例集合として評価を行なう。なお、重回帰分析では変数選択を行わない。

読解時間予測モデルの予測性能の評価指標として、次の式(1)で計算される誤差率を用いる。

$$\text{誤差率} = \frac{|\text{予測値} - \text{実測値}|}{\text{実測値}} \times 100\% \quad (1)$$

式(1)において、予測値は読解時間予測モデルによって予測される読解時間であり、実測値は3節で述べた方法で測定された読解時間である。

4.2 提案モデルの予測性能

試験事例集合での提案モデルの誤差率の分布表とヒストグラムをそれぞれ表2と図1に示す。表2より、誤差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。図1を見ると、峰が左端にあり右側に長く裾を引いた分布となっていることがより明らかである。この誤差率の分布に対してコルモゴロフ・スミルノフ検定を行ない、この分布は正規分布と異なることを確認した。誤差率の中央値は22.9%であり、レンジは99.2であった。

表2 誤差率 ER の度数分布表

階級	度数	相対 度数 (%)	累積 度数	累積相対 度数 (%)
0% ≤ ER < 10%	43	23.9	43	23.9
10% ≤ ER < 20%	34	18.9	77	42.8
20% ≤ ER < 30%	37	20.6	114	63.3
30% ≤ ER < 40%	19	10.6	133	73.9
40% ≤ ER < 50%	20	11.1	153	85.0
50% ≤ ER < 60%	10	5.6	163	90.6
60% ≤ ER < 70%	9	5.0	172	95.6
70% ≤ ER < 80%	5	2.8	177	98.3
80% ≤ ER < 90%	1	0.6	178	98.9
90% ≤ ER < 100%	2	1.1	180	100.0

4.3 読解時間予測モデルに取り込む要因と予測性能の関係

提案モデルは、語彙的、統語的、談話的要因と読解者要因を組み合わせる読解時間を予測する。これらの要因の組み合わせの有効性を検証するために、語彙的要因、統語的要因、談話的要因を個別に読解者要因と組み合わせた読解時間予測モデルの性能と提案モデルの性能を比較する。

可能性も考えられる。

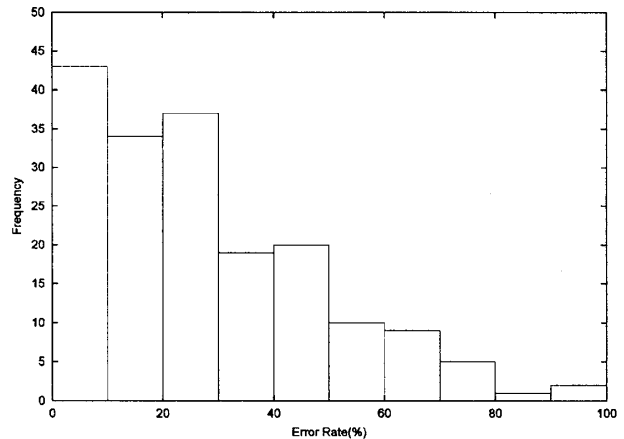


図1 誤差率のヒストグラム

全要因を取り込んだ提案モデルの誤差率と各言語的要因を個別に読解者要因と組み合わせた読解時間予測モデルの誤差率を表3に示す。表3より、提案モデルの誤差率が他のすべての読解時間予測モデルの誤差率よりも小さい値になっており、予測性能に悪影響を及ぼしている要因は存在しないことが分かる。また、統語的要因を取り込まないと誤差率が高くなる傾向が見られることから、文の読解時間の予測に特に貢献している要因は統語的要因であると言える。

表3 読解時間予測モデルに取り込んだ要因と誤差率

取り込んだ要因	誤差率 (%)
全要因	22.9
語彙的要因と読解者要因	25.7
統語的要因と読解者要因	24.6
談話的要因と読解者要因	37.2
語彙的要因と統語的要因と読解者要因	24.2
語彙的要因と談話的要因と読解者要因	27.3
統語的要因と談話的要因と読解者要因	24.1

4.4 従来研究での統語的要因を取り込んだ読解時間予測モデルとの比較

提案モデルでは、統語的難易度を上げる要因として、従来から着目されていた文の長さの他に、構文木の節点の数に着目した。これに対して、統語的要因として、構文木の高さ、名詞句の数、動詞句の数、従属節の数に着目した Schwarm らの研究 [7] や、関係代名詞節、分詞節、to 不定詞節に着目した永田らの研究 [5] がある。そこで、本研究で着目した構文木の節点の数という要因を従来研究で着目された要因に置き換えて読解時間予測モデルを構築し、それらの予測性能と提案モデルの予測性能を比較する。なお、永田らの読解時間予測モデルは、関係代名詞節、分詞節、to 不定詞節に現れる単語の認識時間へ重み付けを行なうものであるが、本稿では、関係代名詞節の数、分詞節の数、to 不定詞節の数を要因として取り込む。

Schwarm らが着目した要因に基づく読解時間予測モデルの誤差率は 23.9% であり、永田らが着目した要因に基づく読解時間予測モデルの誤差率は 24.9% であった。提案モデルの誤差率 22.9% は、Schwarm らが着目した統語的要因に基づくモデルの誤差率よりも 4.2% ($= \frac{23.9-22.9}{23.9} \times 100$) 低く、また、永田らが着目した統語的要因に基づくモデルの誤差率よりも 8.0% ($= \frac{24.9-22.9}{24.9} \times 100$) 低くなっており、提案モデルのほうが従来研究で着目された統語的要因を取り込んだ読解時間予測モデルよりも優れていると考える。

従来研究 [5, 7] と本研究に共通する問題点として、両者とも構文木上の節点に着目しているため、構文解析器が正しくない構文木を生成した (構文解析を誤った) 場合、その影響を受ける可能性があることが挙げられる。これに対して、従来研究と本研究の違いとして、従来研究で着目された節点は、名詞句や動詞句などの特定のラベルを持つ節点であるのに対して、本研究では、節点のラベルは利用していないということが挙げられる。このため、従来研究では、構文解析における不適切なラベル付け (例えば、本来、名詞句である節点を名詞句と認識しない誤りや、本来、名詞句でない節点を名詞句と認識してしまう誤り) の影響を受けるが、本研究の提案モデルではこの悪影響は受けない。従って、本研究で着目した統語的要因は従来研究で着目された統語的要因に比べて、構文解析誤りに対してより頑健であると考えられる。

5 おわりに

本稿では、英語学習者が十分には習得していない語彙や文法項目を含む文を特定するために、文の標準的な読解時間を予測するモデルについて述べた。評価実験の結果、提案モデルは 22.9% の誤差率で文の読解時間を予測することができ、提案した言語的要因と読解者要因の組み合わせが有効であることが明らかになった。今回の実験で利用しなかった様々な言語的要因の読解時間予測モデルへの導入について今後検討を行ない、予測性能の向上を図っていく。

参考文献

- [1] R. P. Carver. Optimal Rate of Reading Prose. *Reading Research Quarterly*, pp. 56-88, 1982.
- [2] R. Flesch. A New Readability Yardstick. *Journal of Applied Psychology*, Vol. 32, pp. 221-233, 1948.
- [3] L. Frazier and K. Rayner. Making and Correcting Errors during Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences. *Cognitive Psychology*, Vol. 14, pp. 178-210, 1982.
- [4] 長沼君主, 和田朋子. 速読力とテキストの読みやすさによる英語リーディング能力の測定. 日本語テスト学会研究紀要, Vol. 5, pp. 34-45, 2002.
- [5] 永田亮, 榊井文人, 河合敦夫, 椎野努. 英文読解力の自動評価法 — リーディングスピードを用いた英文読解力の詳細な評価 —. コンピュータ&エデュケーション, Vol. 12, pp. 99-103, 2002.
- [6] 佐野洋, 猪野真理枝. 英語文法の難易度計測と自動分析. 研究報告 CE58-2, 情報処理学会, 2000.
- [7] S. E. Schwarm and M. Ostendorf. Reading Level Assessment Using Support Vector Machines and Statistical Language Models. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 523-530, 2005.
- [8] T. Shizuka. The Effects of Stimulus Presentation Mode, Question Type, and Reading Speed Incorporation on the Reliability/Validity of a Computer-based Sentence Reading Test. 大学英語教育学会紀要, No. 29, pp. 155-172, 1998.
- [9] 吉見毅彦, 小谷克則, 九津見毅, 佐田いち子, 井佐原均. 英語学習者の読解能力推定のための読解時間測定法. 教育システム情報学会誌, Vol. 22, No. 1, pp. 24-29, 2005.