

# 読解過程検定ツールを用いた読解支援システム評価

小谷克則<sup>†\*</sup>  
情報通信研究機構<sup>†</sup>

吉見毅彦<sup>‡</sup>

九津見毅<sup>††</sup>  
龍谷大学<sup>‡</sup>

佐田いち子<sup>††</sup>

井佐原均<sup>†</sup>  
シャープ株式会社<sup>††</sup>

\*(kat@khn.nict.go.jp)

## 1 はじめに

本稿は、読解速度を尺度にした英文読解支援システムの評価法を提案する。一般に、機械翻訳システムなどの評価には、正解訳との類似度に基づく機械的評価法 (Papineni et al. 2002) や評価者の主観評価に基づく評価法がこれまでに提案されている。さらに、後者の評価法は、正解訳に基づく評価方法と実際の運用における有効性からの評価方法とに二分できる。有効性に基づく評価法 (大黒 1993, 富士他 2002) では、支援付きテキストに対する官能評価試験や理解度テストのスコアが評価尺度として用いられる。本稿が提案する評価法もシステムの有効性に基づく評価方法の一つであるが、評価尺度に理解度テストなどではなく支援付きテキスト読解中の読解速度を用いる。

本稿の構成は、以下の通りである。2 節では、本稿が提案する読解速度による読解支援システム評価手法を解説する。3 節では、読解速度を評価尺度として文の可読性を推測できるかどうかを検証した実験とその結果を報告する。4 節では、筆者らが行なった読解支援システム評価実験の概要を述べる。5 節では、読解支援システム評価実験により得られた結果とその考察を報告する。6 節では、本稿のまとめと今後の課題を述べる。

## 2 読解速度に基づく支援システム評価法

筆者らが提案する読解支援システム評価法は、支援付きテキスト読解中の読解速度を評価基準として用いる。この読解速度を算出するのに必要な読解時間は、筆者らが開発した読解過程検定ツールを用いて計測する<sup>1</sup>。このツールは、読解対象テキストをいくつかの読解範囲に区分し、コンピュータ画面上に読解範囲開示アイコンを提示する機能を有する。そして、読解者が開示している読解範囲とその時間を記録する。このようにツールによって計測した読解時間を基に読解速度 (WPM) を算出する。読解支援システムが有効に機能してい

れば、読解効率が向上するため、読解速度が速くなると考えられる。この考えに基づき、筆者らは支援なしテキスト (英文のみテキスト) と支援ありテキストの読解速度を比較することによって、支援システムの有効性が評価できると想定する。

## 3 読解速度データの妥当性

筆者らは、テキスト中の一文当りの読解速度が可読性を適切に反映しているかどうかの検証実験を行った。もし、読解速度が可読性を反映しているのであれば、速度を評価基準として採用することに問題はないと思われる。さらに、速度を基準にすることにより、これまでの理解度テストを用いた評価法では、困難と考えられるテキストの各構成要素 (たとえば、文) に対する評価も可能になると考える。理解度テストの得点が示すのは、テキスト全体の読解効率であり、文毎の効率を算出するのは困難であると思われる。例えば、理解度テストを用いてテキストを局所的に評価するためには、各文毎に設問を用意する必要が生じる。また、評価者も文毎に設問を解くことによる負担が強いられる。一方、読解時間の計測では、評価者への負担は少なく、通常読解に近い環境での評価が行なえるなどの利点があると思われる。そこで、読解速度データの妥当性を検証する実験を行った。

この実験では、英文テキストに含まれる文毎の読解速度を計測し、その結果を各文に割り当てたいわゆるリーダビリティの公式により算出されたスコア (Flesch 1948) と比較した。もし、読解速度データが可読性判定に妥当なものであれば、可読性スコアと読解速度に相関がみられるべきである。

本稿が用いた可読性算出式は、英語母語話者を対象に英文テキストの難易度を、文長や単語長といった表層の言語的特徴から算出する。可読性スコア自体の妥当性は、理解度テストやクローズテストなどの得点との相関により確認されている (DuBay 2004)。

このような母語話者を想定して作成された可読性算出式を第二言語学習者にあてはめることを問題として、学習者専用の算出式も提案されている (永田他 2004)。しかし、筆者らは、可読性算出式が反映すべき『読みやすさ』は、あくまでも母語話者を対

<sup>1</sup> システムの詳細については、(吉見他 2004) を参照。

象とすべきであると考え、なぜなら、英文読解学習において、目標とする読解力は母語話者レベルであるからである。つまり、可読性算出式を通じて学習者が得べき情報は、母語話者の読解にどの程度類似しているかである。しかし、学習者を主体とした算出式を用いると目標にどの程度近づいたかを確認できない。さらに、学習者特有の特徴を含めると、学習者のレベルに応じた算出式が必要となる。そこで、筆者らは母語話者を主体に開発されてきた可読性算出式を採用した。また、文長や単語長など表層の言語情報を基に可読性を求める Flesch の可読性算出式を用いた<sup>2</sup>。

この算出式を用いて、二つの英文テキストに含まれる 24 文のスコアを求めた。そして、このスコアと読解速度データの比較を行なった。実験被験者は、大学生 25 名であった。計測された読解速度とリーダビリティのスコアをそれぞれ標準化し、回帰直線にもとづく相関値を求めた。その結果、危険率 1%において 0.6 から 0.7 程度の相関値が得られた。本稿は、この相関に基づいて、読解速度による可読性を評価できるものと想定した。

#### 4 支援システム評価実験

読解支援システムとして評価対象に選んだのは、(1)機械翻訳(SHARP 製品『翻訳これ一本』)、(2)訳語ルビ振り(同社製品『おまかせ訳振り』)、(3)チャンカー(イリノイ大学公開ツール"Shallow-Parser"[http://l2r.cs.uiuc.edu/~cogcomp/eoh/c\\_hunkerdemo.html](http://l2r.cs.uiuc.edu/~cogcomp/eoh/c_hunkerdemo.html))である。筆者らは、機械翻訳と訳語ルビ振りを英語原文に対応する翻訳文や単語や句の訳を提示することによる読解支援ツールとして位置づけた。また、チャンカーに関しては、句や節の境界を提示することによる支援ツールとして扱った<sup>3</sup>。

実験で用いた英文テキストは、TOEIC 準拠市販教材(田中・杉浦 2003)の読解問題部分から無作為に抽出した 10 テキストである。一つのテキストは、テキスト部分と理解度テスト部分から成る。これらの実験対象テキストを、さらに無作為に 5 種類に分類した。そして、3 種類の支援付きテキストと 2 種類の実験統制テキスト(英語原文と日本語完全和訳)を作成した。

被験者は大学一回生 25 人である。自己申告によ

<sup>2</sup> Flesch のリーダビリティスコア S (0-100) は、以下の式によって求められる。

(i)  $S = 206.835 - 84.6 * \text{音節数} / \text{単語} - 1.015 * \text{単語数} / \text{文}$

<sup>3</sup> チャンキング情報に基づく読解支援の有効性に関しては、田中他(2004)を参照。

る TOEIC スコアから推測されるレベルは、500 点程度と平均的な大学生の英語学力レベルである。全被験者が読解過程検定ツールを初めて用いることから、実験対象テキストとは別テキストを用いて、ツールの操作練習を実験開始前に行なった。また、支援付きテキストにも慣れるために、練習用テキストに支援情報を追加した。練習終了後、各被験者は、検定ツールを用いて 1 セット 5 種類のテキストを 2 セット、計 10 テキストの読解を行なった。各セットの間に 15 分程度の休憩を入れた。今回の実験では、全被験者に同一テキストセットを与えた。そのため、同一テキスト内で、支援付き・支援なしなどの違いを分析することは対象外とした。

実験により得られた読解速度データの分析を行なう前に、読解速度 0 を含む被験者データを削除した。なぜなら、速度 0 データは、被験者が読解を行なっていない可能性を含むからである。計測に用いたツールには、読解範囲を開けたかどうかをチェックする機能がなく、また、計測時間が秒単位であるため(1 秒以下の場合、0 秒としてカウントする)、被験者が実験指示通り読解を進めたが読解時間が 1 秒以下であったのか、あるいは読解範囲を読まずに進めたのかが区別できない。したがって、読解速度データ上に速度 0 を含む被験者データを削除した。その結果、分析対象となったデータは、17 名分であった。

#### 5 実験結果とその考察

##### 5.1 理解度テスト正解率による分類

被験者の読解能力に応じて、支援システムの有効性が異なることが確認されている(富士他 2002)。そこで、英文のみテキストの設問正解率に基づいて、被験者を三群に分類した(表 1)。正解率に応じて、低群(L)、中群(M)、高群(H)とに分類した。

	低群 (L)	中群 (M)	高群 (H)
人数	8	5	4
平均正解率(%)	43.8	60.0	79.1

(表 1)

各被験者群のテキスト 5 種類における平均読解速度を算出した(表 2)。ここで読解速度の算出法について簡単に説明しておく。通常、読解速度を算出する際、対象となるのはテキスト・文に含まれる単語の総数である。したがって、支援付きテキストの場合、原文と訳が併記されていることから、英語単語数と日本語単語数を合わせた単語総

数が対象となると思われる。しかし、筆者らは、日本語単語を速度算出の対象外とした。なぜなら、この実験が分析対象とするのは、純粋な日本語・英語混在テキストの読解速度ではなく、支援付き英語原文テキストの読解速度であるからである。つまり、支援によって原文読解速度がどのように変化するかが分析対象であるため、読解速度を算出する際、英語原文テキストに含まれる単語数を用いた。さらに、日本語テキストの速度算出においても、原文の英語単語数を利用した。

	英語	日本語	チャンク	ルビ振り	機械翻訳
H	65.9	192.4	62.3	75.6	74.1
M	52.4	166.3	49.9	50.9	64.5
L	44.8	163.4	47.5	53.6	74.9

(表 2)

まず、統制テキストである英語テキストと日本語テキストの速度を比較する。被験者全員の母語が日本語であり、英語が第二言語として学習対象言語であることから、日本語テキスト速度が英語テキスト速度を上回ると予測される。表 2 の実験結果が示すとおり、読解速度は読解対象言語の違いを反映しているといえる。

次に、被験者低群、中群、高群(表 1)における英語テキスト速度を比較する。読解能力が速度に反映しているのであれば、英語テキスト速度は、理解度テストの正解率に対応すると思われる。実際、表 2 が示すとおり英語テキスト速度は、高群が一番速く(65.9WPM)、次に中群(52.4WPM)、そして低群(44.8WPM)と続く。このことから、読解速度が理解度テストの結果を反映していることがわかる。

支援付きテキスト速度を各群において比較すると、チャンクテキスト速度だけが読解力と対応した結果となった。ルビ振りテキストの場合、高群の読解速度が一番速いが、低群と中群の間に差はほとんどみられない。機械翻訳テキスト速度は、低群と高群がほぼ同程度である。

次に、各被験者群において、英語テキスト速度と支援付きテキスト速度を比較した。この結果、高群においては、ルビ振りテキストと機械翻訳テキストで速度の上昇が確認できる。中群においては、機械翻訳テキストで速度上昇が確認できる。低群では、全ての支援付きテキストにおいて速度が速くなっている。

このような支援付きテキストに見られる速度分布は、支援精度に依存するものであると筆者らは考える。実験で用いた支援システムの出力結果には、人手による修正等を加えなかったため、誤り

を含む。そのため、被験者が読解中に支援情報の誤りを判断する必要があったと思われる。この正誤判断に費やす時間が読解時間として計上され、結果的に速度が遅くなると考えられる。このように読解速度は、支援システムの精度に依存すると考えられる。特に、このような遅延効果が認められるのは、高群や中群のチャンクテキストと中群のルビ振りテキストの場合である。

もし、支援システムによる遅延効果が正しいならば、高群においてもルビ振りテキスト速度が中群と同様に速度下降が予測される。しかし、実際は速度が上昇している。筆者らは、速度上昇を支援効果の有効性を示すものとする。したがって、中群において速度遅延効果が認められる程度の精度であっても、高群においては、遅延効果を上回る支援効果が得られたと考えられる。同様の支援効果は、高群・中群の機械翻訳テキスト読解にも認められる。筆者らは、この支援効果と遅延効果のどちらが表出するかは、読解能力に依存すると考える。なぜなら、誤りを含む支援システムを用いたにもかかわらず、低群においてはどの支援システムであっても支援効果が認められるからである。

## 5.2 文可読性にもとづく分類

前小節で、読解能力にもとづいて支援システムの効果を考察した。本小節では、テキストの難易度から支援効果の有効性を考察する。支援システムの精度が、テキストの難易度に影響を受けることが報告されている(Clifford et al. 2004)。また、テキストの難易度は、被験者にも影響を与えると考えられる。そこでテキストの各文を難易度によって分類し、読解速度の変化を考察する。

テキスト中の各文をいわゆるリーダビリティの公式(Flesch 1948)のスコアに基づき分類した。リーダビリティスコア値の範囲が 0-100 であることから、難易度 1(スコア: 0-33)、難易度 2(スコア: 34-66)、難易度 3(スコア 67-100)に分類した。各群の文数は、難易度 1 が 25 文、難易度 2 が 48 文、難易度 3 が 32 文であった。難易度別に、テキストタイプと被験者グループ(低群・中群・高群)に基づいて読解速度を比較する。(表 3-5)

難易度 1	低群	中群	高群
英語	43.4	47.4	60.8
チャンク	46.7	39.0	55.7
ルビ振り	50.7	37.6	50.0
機械翻訳	88.6	60.0	77.5

(表 3)

難易度 2	低群	中群	高群
英語	48.9	60.0	73.0
チャンク	49.1	57.1	66.5
ルビ振り	52.4	51.5	81.5
機械翻訳	76.4	59.9	72.8

(表 4)

難易度 3	低群	中群	高群
英語	50.8	68.3	81.4
チャンク	43.8	42.6	58.5
ルビ振り	61.5	63.5	96.1
機械翻訳	68.8	73.2	76.1

(表 5)

まず、難易度 1 の文における各被験者群に対する読解支援システムの有効性を考察する。低群では、どの支援付きテキストも英語テキストの速度より速いことから支援効果が認められると考えられる。中群と高群では、機械翻訳テキストだけに支援効果が認められる。次に、難易度 2 の文における支援効果を確認する。低群では、難易度 1 の場合と同様に、どの支援システムにおいても支援効果が認められる。中群では、全ての支援システムにおいて、その効果が低いと考えられる。高群では、ルビ振りのみ支援効果が認められる。難易度 3 の文においては、低群では、チャンク付きテキストの支援効果が低いと思われる。中群では、機械翻訳テキストにおいて支援効果が認められる。高群では、ルビ振りテキストに支援効果があったと考えられる。

それぞれの支援効果を、英語テキスト速度と比べて表 6 にまとめた。5WPM 以上の速度上昇がある場合を支援効果あり(○)とした。一方、5WPM 以上の速度降下がある場合を支援効果なし(×)とした。また、正負の別なく 5WPM 以下の変化がある場合を支援効果微小(△)とした。

読解能力	低群			中群			高群		
	1	2	3	1	2	3	1	2	3
難易度									
チャンク			×	×	×	×	×	×	×
ルビ振り				×	×		×		
機械翻訳									

(表 6)

本稿が行なった読解速度にもとづく支援システムの評価実験では、先行研究(富士他 2002)と同様に、支援システムの有効性が英語読解能力の低いグループにおいて顕著に認められた。このグループでは、テキストの難易度に関係なく支援システムが有効に働くと考えられる。また、中程度の読解力をも

つグループでは、機械翻訳システムがその有効性を発揮したと考えられる。但し、テキストの難易度が上がるとその効果は小さくなるようである。高群では、テキスト難易度が低い場合に機械翻訳システムがその有効性を示し、難易度があがるとルビ振りシステムが効果を示すようである。

## 6 まとめと今後の課題

本稿は、読解速度に基づいて機械翻訳システムなどの読解支援システムを評価する手法を提案した。本手法に基づく評価実験では、テキスト難易度と読解者の能力を基にシステムの有効性を比較した。今後、各システムの持つ問題点の検出を行なうなど詳細な評価へと発展させる。

また、本稿で用いた可読性は、表層の言語特徴に基づいたものであった。今後、意味情報や統語情報などを加え、綿密な可読性に基づいて詳細なシステム評価を行なう。

## 参考文献

- Clifford, R. N. Cranioien, D. Jones, W. Shen, & C. Weinstein. The Effect of Text Difficulty on Machine Translation Performance: A Pilot Study with ILR-Rated Texts in Spanish, Farsi, Arabic, Russian and Korean. In Proceedings of the 4th International Conference on Language Resources and Evaluation: 343-346.
- DuBay, W. 2004. The Principles of Readability. MS.
- Flesch, R. 1948. A New Readability Yardstick. Journal of Applied Psychology 32: 221-233.
- 富士秀, 畠中信敏, 伊藤悦雄, 亀井真一郎, 隈井裕之, 介弘達也, 吉見毅彦, 井佐原均. 2002. 機械翻訳システムの有効性の評価 - どのような人にとって MT は役立つのか -. 言語処理学会第 8 回年次大会発表論文集: 41-44.
- 永田亮, 井口達也, 榎井文人, 河合敦夫. 2004. 日本人英語学習者を対象にした英文難易度判定手法. 電子情報通信学会論文誌. Vol. J87-D2. No.6: 1329-1338.
- 大黒慶久. 1993. 訳語ルビふり英文の可読性評価. 自然言語処理 No.97-18: 127-134.
- Papineni, K., S. Roukos, T. Ward, & W.-J. Zhu. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting of the Association for the Computational Linguistics: 311-318.
- 田中省作, 富浦洋一. 2004. スラッシュ・リーディング支援システムの構築. 言語処理学会年次大会併設ワークショップ「e-Learning における自然言語処理」論文集: 37-40.
- 吉見毅彦, 小谷克則, 九津見毅, 佐田いち子, 井佐原均. 2004. 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.20-1.