

一文ごとの読み速度に基づく難解文抽出手法

寺田 春菜

小林 潤平

大日本印刷株式会社

1 はじめに

本研究では、わかりやすい文章作成を支援するインターフェースの開発を目指し、表記的にも文法的にも正確であるが、わかりにくいと感じる文を、人間の読み行動に基づいて抽出する手法を提案する。

文章を書くことは複雑な活動であり、訓練を重ねても、わかりにくい文章を書いてしまうことが多々ある。文章をわかりにくくする要因はさまざまであり、例えば、誤字や表記ゆれといったミス、文の照応関係やパラグラフ構成といった文章構造、背景知識や暗黙知といったコンテキストの不足等が挙げられる。誤字や文法的なミスは、自然言語処理技術等によって抽出および修正可能な場合も多い（例えば [1] 等）。しかし、文章構造的なわかりにくさおよびコンテキスト不足によるわかりにくさは、表記や文法は正確であるため、抽出が難しいという課題があった。

人間の読み行動においては、親密度（読者にとっての単語に対する「なじみ」の程度）の低い単語が出現する文や複雑な構造の文の読解時に、親密度の低い単語やその単語周辺の注視時間が増大し、読み速度が低下することが報告されている [2]。また、新しい内容が出現する文では、注視回数が増えて読み速度が低下することが報告されている [3]。すなわち、わかりにくい文は読み速度の低下を引き起こしており、一文ごとの読み速度を指標として分析することで、わかりにくい文を抽出できる可能性がある。

そこで本稿では、一文ごとの読み速度を計測できる電子リーダーを開発し、文章内での読み速度の変化を分析するとともに、読み速度低下と文のわかりにくさの一致度を検証した。

2 実験

電子リーダー 文章の呈示および読み時間を計測するために、図 1 に示す電子リーダーを開発した。この電子リーダーは、一文ごとの読み時間を計測するために、読んでいる一文だけを鮮明に表示し、それ以外の文はぼか

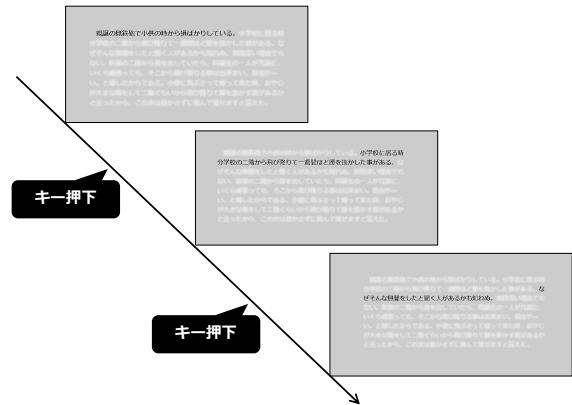


図 1: 一文ごとに表示する電子リーダーの画面イメージ

して読めない状態を表示する。背景色は灰色 (#cccccc) とし、鮮明に表示する場合の文字色は黒色 (#000000)、ぼかして表示する場合の文字色は薄い灰色 (#dedede) とした。次の文に進む操作は実験参加者のキー押下により行い、前の文に戻ることも可能である。このキー押下をトリガーとして各文の読み時間を計測した。

刺激 呈示文章は 2,500 字程度の説明文とした。高等学校卒業認定試験の過去問題から説明文とその要旨問題を 5 組選定した。大幅に文字数が多い場合は要旨問題の解答に支障が出ない部分を省略した。

装置 表示装置にはラップトップ型 PC である Apple 社製 MacBook Pro (画面サイズ対角 13.3 inch, 画面解像度 227 ppi) を使用した。

実験参加者 20 代～40 代の社員 10 名が参加した。

手続き 実験参加者は図 1 の電子リーダーを用いて、自身のキー操作により呈示文章を読み進めた。その後、実験参加者は文章のどこがわかりにくかったかを回答した。実験は 5 文章で実施し、文章の呈示順はランダムとした。なお、呈示文章に対する内容理解を一定に保つために、読後に要旨問題を出題した。平均正答率は 90% であり、理解しながら読んでいることを確認した。

3 結果

本稿では、5 文章のうち、読み速度および読みやすさの主観評価ともに最も平均的であった呈示文章 α の

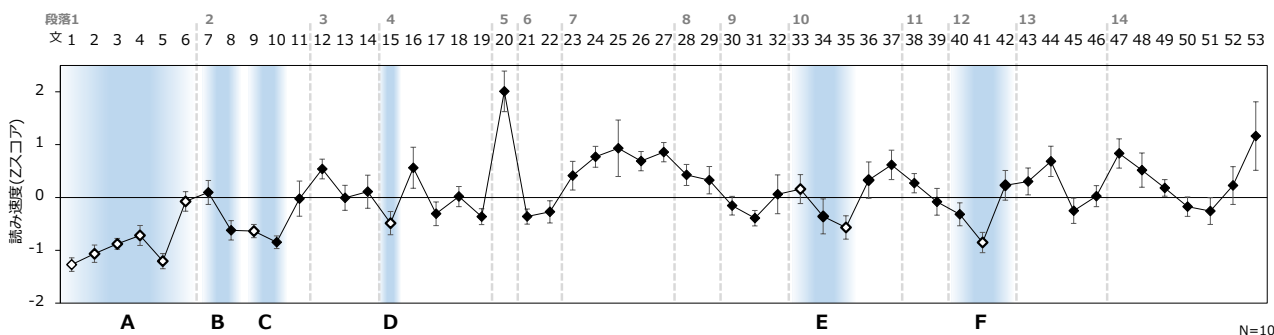


図2: 文章 α における実験参加者 10 名の一文ごとの読み速度変化

分析結果を示す。

図2は、文章 α における一文ごとの読み速度変化を示したグラフである。縦軸が読み速度 (Zスコア)、横軸が段落および文番号であり、文番号1が文章の先頭である。読み速度は実験参加者間のばらつきが大きく、そのまま平均をとると個人差の影響を受けてしまうため、実験参加者ごとに標準化した後、文ごとに全実験参加者の平均を算出した。読み速度 (Zスコア) の0は文章内の読み速度の平均値であり、正の値であれば平均より読みが速く、負の値であれば平均より読みが遅いことを示す。エラーバーは標準誤差を示す。また、実験参加者からわかりにくかったと主観報告があった文を、白抜きのマーカーで示した。

まず、図2領域A, C, E, Fについて、読み速度の低下とわかりにくいとの主観報告が観察された。この領域の文に出現する単語は、実験参加者にとって馴染みがない単語であったことが主観報告からわかった。先行研究 [2] においても、読者にとって親密度の低い単語が出現する文では、読み速度が低下することが報告されている。したがって、読者にとって馴染みがない単語の出現によってわかりにくく感じる文について、読み速度低下の指標から抽出できることがわかった。

次に、図2領域Dについて、読み速度の低下とわかりにくいとの主観報告が観察された。この領域では、前段落までの内容を比喩表現を用いて総括しているが、その比喩が適切でないために、かえって理解を阻害したことが主観報告からわかった。比喩表現自体は文法的にも正しく平易な表現であったが、読者にとってはわかりにくい文とされた。したがって、このようなわかりにくい文についても、読み速度低下の指標から抽出できることがわかった。

最後に、図2領域Bについては、読み速度の低下のみが観察された。この領域は新しく問題提起している段落であり、先行研究 [3] においても、文章中に新しい内容が出現すると、読み速度が低下することが報告されている。したがって、このような新しい内容を理解

するために時間のかかる文も、読み速度低下の指標から抽出できることがわかった。

なお、領域A~Fの文は表記的にも文法的にも正確であった。

以上の結果より、読み速度の低下を指標とすることで、表記的にも文法的にも正確であるが、わかりにくいと感じる文を抽出できることがわかった。

4 おわりに

本稿では、一文ごとに表示する電子リーダーにより読み速度を計測し、読み速度の低下と文のわかりにくさの一致度を検証した。その結果、読み速度の低下とわかりにくさは一致する傾向を示すことがわかった。

読み速度が低下する部分は、読者にとってわかりにくい部分であり、よりわかりやすい表現や表記への修正が望まれる可能性が高いと言える。すなわち、読み速度が低下した文を抽出することで、文章推敲の手掛かりをすばやく提示できる可能性が期待できる。改善すべき文を読むだけで抽出できれば、推敲にかかる時間を大幅に削減することができ、よりわかりやすく伝わる文章作成の一助となるはずである。

今後は、文章推敲を支援するインターフェースを模索していく予定である。文章推敲が容易になるようなわかりづらさの呈示手法について検討していきたい。

参考文献

- [1] 坂本 俊介, 須藤 崇志, 丸山 広, 中村 太一. 形態素解析を利用した文章校正手法の提案. 情報処理学会研究報告, vol. 2009-DD-72, no. 17, pp. 1-6, 2009.
- [2] 近藤 公久, 馬塚 れい子, 笈 一彦. 日本語文の読解過程における語特性および語順の影響. 認知科学, vol. 9, no. 4, pp. 543-563, 2002.
- [3] Jukka Hyönä. An eye movement analysis of topic-shift effect during repeated reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, vol. 21, no. 5, pp. 1365-1373, 1995.