

2025 年 9 月 3 日

報道関係者各位

国立大学法人筑波大学

国立大学法人九州大学

英文読解中の単語処理に影響を与える要因を処理の段階別に解明

英文を読んでいる際、特定の単語につまづいたり、読み戻ったりすることがあります。本研究では、英文読解中の視線計測データを利用して、単語の長さ、頻度、予測しやすさのうち、どれが英文読解中の単語処理において最も重要となるのかを明らかにしました。

文章を読んでいると、ある単語で読みが止まったり、後から読み戻ったりすることがあります。読解中の眼球運動を記録する視線計測を用いた研究では、単語の「長さ」「頻度」「文脈からの予測しやすさ」が英文読解中の単語処理に影響を与える三大要因であるとされてきました。これらの要因は英語を母語としない読み手にも影響を与えることが知られていますが、そのうちどれが最も重要なのかについては明らかにされていませんでした。

本研究では、日本語を母語とする大学生・大学院生の英文読解中の視線計測データベースを用い、読解中の単語処理における、単語の長さ、頻度、予測しやすさの重要度を機械学習によって検討しました。これらの語彙特徴の他、英文読解力や自信度などの個人差要因、テキスト全体の難易度なども変数に加えて分析した結果、読み飛ばしの有無や合計の処理時間においては単語の長さが最も重要な要因でしたが、初期段階の処理（最初に単語を見た時の処理時間）や読み戻りについては、単語の頻度や予測可能性が重要となることが示されました。また、語彙特徴よりも個人差要因が重要となる場合もありました。

本研究の成果は、英語学習者の読解中の単語処理を困難にする要因を明らかにするとともに、読解中の眼球運動の計算モデルの改良など、幅広い分野への応用が期待されます。

研究代表者

筑波大学人間系

名畑目 真吾 助教

九州大学大学院 言語文化研究院

内田 諭 准教授

研究の背景

文章を読む際、私たちはある単語は飛ばし、ある単語では止まり、時に読み戻るということを繰り返しています。このような現象は、読解中の眼球運動を記録する視線計測によって広く研究され、これまでに、英文読解中の語彙処理には単語の「長さ」「出現頻度」「文脈からの予測しやすさ」の3つの語彙特徴が大きな影響を与える要因であることがわかっています。また、非英語母語話者あるいは英語学習者に対しても同様に、これら3つの要因が重要であることが示されています。しかしながら、従来の研究では、英語学習者の読解においてこの3つの要因のうち特にどれが重要なのか（相対的な重要度）は分かっていませんでした。

研究内容と成果

本研究では、公開されている日本語母語話者の英文読解中の視線計測データセット *Tsukuba Eye-Tracking Corpus*（大学生・大学院生41名が、計1万語超に及ぶ30の英語の文章を読解した際の、約41万トークン^{注1)}に対する視線計測データを含む）を用い、単語に対する注視^{注2)}時間や回数、読み飛ばしや読み戻りなどの視線計測データ指標を予測する機械学習のモデル（ランダムフォレスト^{注3)}）を構築しました。予測変数には、単語の長さ（文字数）、頻度（コーパス^{注4)}における出現頻度）、予測しやすさ（大規模言語モデル^{注5)}に基づくサプライザル^{注6)}など）を含む語彙特徴の他、文の特徴（文の長さや語彙の平均頻度）やテキスト難易度の指標、読み手の英文読解力（読解テスト得点）と自信度に関する指標を取り入れました。

分析の結果、単語の長さが多くの視線計測データ指標で最も重要となり、特に読み飛ばしや単語に対する注視回数を強く左右しました（図1）。一方、単語の初期処理を示す指標（最初の注視の継続時間）では単語の長さよりも頻度が、また、読み戻りのような比較的後半の処理では予測しやすさが、より重要となっていました。また、読解力や英語使用への自信度といった個人差要因が、すべての語彙特徴よりも影響を及ぼしている視線計測データ指標も存在しました。これらの結果は、英文読解中の単語処理においては、多くの場合で単語の長さが鍵となるものの、特定の処理段階においては別の特徴や要因がより重要になることを示しています（図2）。

今後の展開

本研究で得られた知見は、英語学習者の読解において処理が困難となる単語を予測する上で有用です。また、コーパスや大規模言語モデルに基づく指標を分析に取り入れていることから、コンピュータによる自然言語処理やAI分野への応用可能性も考えられます。

参考図

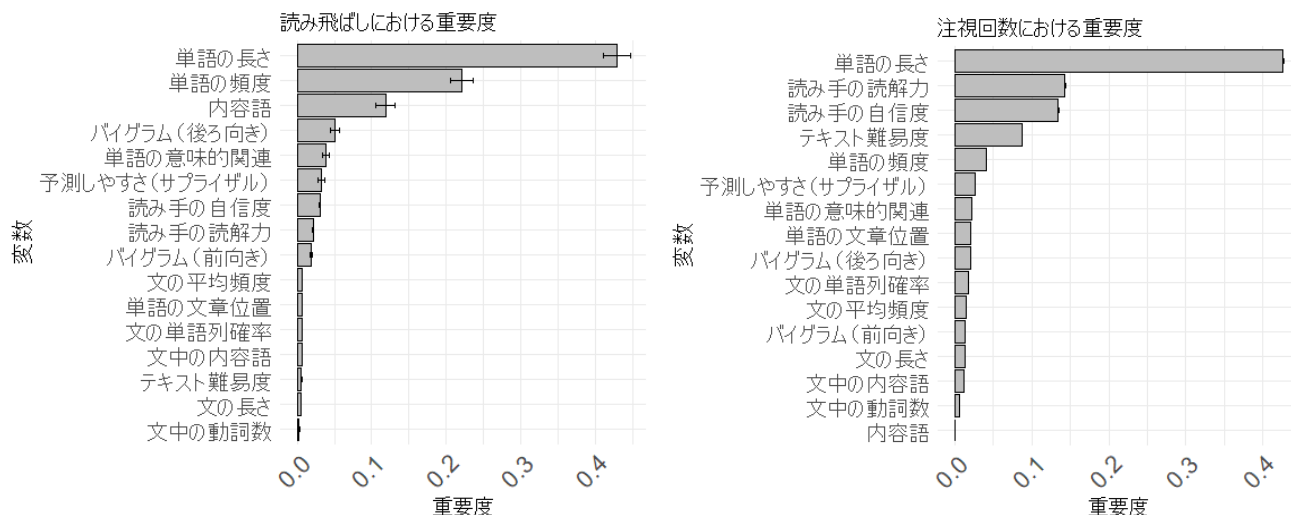


図1 単語の読み飛ばし（左図）と注視回数（右図）に影響を及ぼす要因。

語彙特徴より重要となる要因

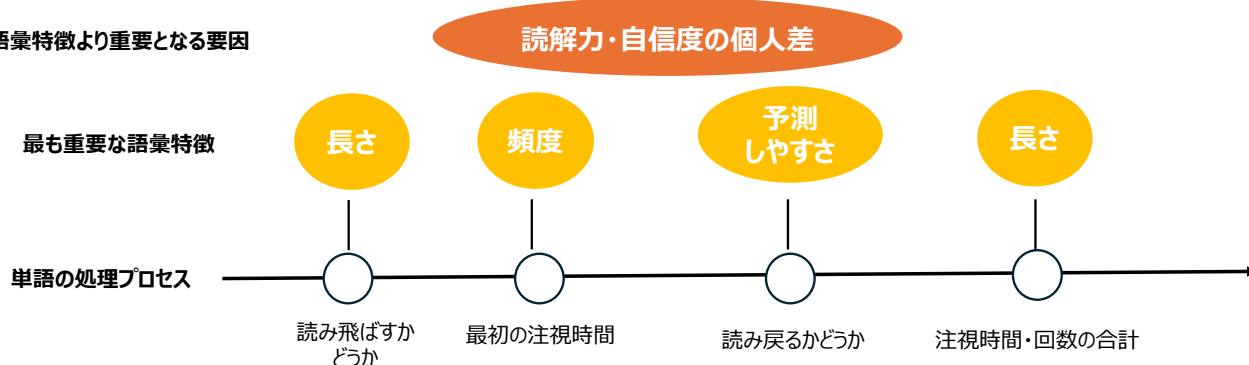


図2 本研究結果の概要。下段は英文読解中の単語の処理プロセスに応じた視線計測データの指標を、中段はそれぞれの視線計測データ指標において最も重要となる語彙特徴を示す。上段は語彙特徴よりも重要となりうる要因。

用語解説

注1) トークン

言語研究において文章やデータ中に出現した語の一つひとつを指す用語。例えば“The book is on the table.”という文は単語の重複があっても6トークンと数える。

注2) 注視

眼球の動きが一時的に止まり、単語などの対象に一定時間とどまる現象。この間に文字や意味に関する情報が得られる。読解研究では、単語や文に対する注視の回数や継続時間が指標として用いられる。

注3) ランダムフォレスト

複数の決定木（枝分かれ構造でデータを分析する手法）を組み合わせる予測を行う機械学習手法。言語研究では、複数の要因の相対的重要性を比較する際によく用いられる。

注4) コーパス

言語研究のために収集された大規模なテキストデータの集まり。単語の出現頻度や共起関係を調べる基盤となる。

注 5) 大規模言語モデル

膨大なテキストデータを学習して、文脈に続く単語や文章を予測・生成する仕組みを持つ AI モデル。

注 6) サプライザル

ある単語が文脈の中でどの程度予測困難かを数値化した指標。読解研究では、予測しにくい単語ほど注視時間が長くなるなどの効果が報告されている。

研究資金

本研究は、日本学術振興会・科学研究費補助金：基盤 B（23K25335 名畑目）の助成により実施されました。

掲載論文

【題 名】 Relative importance of lexical features in word processing during L2 English reading
(英文読解中の単語処理における語彙特徴の相対的重要性)

【著者名】 Nahatame Shingo, Uchida Satoru

【掲載誌】 Studies in Second Language Acquisition

【掲載日】 2025 年 8 月 29 日

【DOI】 10.1017/S0272263125101137

問合わせ先

【研究に関すること】

名畑目 真吾（なはため しんご）

筑波大学人間系 助教

URL: <https://english.education.tsukuba.ac.jp/>

【取材・報道に関すること】

筑波大学広報室

TEL: 029-853-2040

E-mail: kohositu@un.tsukuba.ac.jp

九州大学広報課

TEL: 092-802-2130

E-mail: koho@jimu.kyushu-u.ac.jp