

## 英文読解中の眼球運動を記録した大規模データセットを公開

文章を読む際の眼球運動を記録し、言語処理の研究に用いることは盛んに行われています。本研究では、日本語を母語とする英語学習者が文章を読解した際の眼球運動データを収集し、総計で約41万語以上に対する眼球運動データセットを Tsukuba Eye-tracking Corpus (TECO) として構築・公開しました。

言葉を処理する人間の過程は非常に複雑であり、これまで心理学や教育学の分野で人間の言語処理に関する多くの研究が行われてきました。特に書き言葉については、文章を読む際の眼球運動を記録する視線計測と呼ばれる方法が有用とされています。近年では、特にヨーロッパやその周辺地域の英語学習者について、母語や外国語での読解における眼球運動データを大規模に収集し、体系的に整理して公開する取り組みが盛んに行われています。しかしながら、これまで、日本語母語話者の英語による文章読解を対象としたものはありませんでした。

本研究では、日本語母語話者の英文読解を対象に、総計41万語以上に対する眼球運動データセットを構築・公開しました。本データセットは、日本語を母語とする大学生・大学院生41名が、英検準2級～準1級に相当する文章を読解した際の、1人あたり約1万語に対する眼球運動データから構成されています。これより、文章に含まれる各単語について、注視の回数、注視の継続時間、単語の読み飛ばしや読み戻りなどを示す、合計9つの指標を算出し、さまざまな分析を行ったところ、収集したデータの信頼性や妥当性が高いことが確認されました。

本データセットは、Tsukuba Eye-tracking Corpus (TECO) としてオープンサイエンスのプラットフォームにて公開しています。文章理解や第二言語習得の研究だけでなく、自然言語処理やAI研究にも活用が可能であり、幅広い学術分野への波及効果が期待されます。

### 研究代表者

筑波大学人間系

名畑目 真吾 助教

## 研究の背景

近年、母語や外国語での文章読解中の眼球運動データを大規模に収集し、体系的に整理したものを、視線計測コーパス (eye-tracking corpus) として公開する取り組みが、国際的に進められています。このような読解中の眼球運動データは、人間が書き言葉を処理・理解する過程を明らかにするのに有用であることが広く知られていますが、データの収集や解析においては特別な機材や多大な時間・労力が必要であることが大きな課題でした。視線計測コーパスを構築・公開することで、多くの研究者が眼球運動データを広く活用できるようになります。視線計測コーパスの中には、英語を外国語として学ぶ学習者を対象としたデータもありますが、主にヨーロッパやその周辺地域で収集されたものが多く、特にアジア圏の英語学習者の文章読解を対象にしたデータは不足していました。そこで本研究では、日本語を母語とする英語学習者を対象に、文章読解中の眼球運動データを大規模に収集し、視線計測コーパスとして公開することを目指しました。

## 研究内容と成果

本研究では、日本語を母語とする英語学習者の大学生・大学院生 41 名が協力者として参加し、英検準 2 級～準 1 級に相当する英語の文章 (各 300~400 語程度) を 30 セット、合計で約 1 万語の読解を行いました。その際の眼球運動を記録し、文章の各単語について、注視<sup>注1)</sup>の合計回数、注視の合計継続時間、最初に起こった注視の継続時間、次の単語を読むまでの注視の継続時間、読み戻りを含めて次の単語を読むまでの注視の継続時間、読み飛ばしの有無、後続の単語からの読み戻りの有無、2 回目の注視の有無、2 回目の読みの有無の、合計 9 つの指標を算出しました。1 人あたり 1 万語、総計で最大 41 万語を超える単語に対してこれらの眼球運動データが付与されています (参考図)。これらのデータの特徴を理解するために、以下の 3 つの分析を行いました。

まず、得られた眼球運動データが協力者および単語ごとにどの程度一貫しているかを分析し、協力者と単語の両レベルで、データの信頼性が十分に保たれていることを確認しました。次に、各眼球運動データの指標について記述統計と分布を確認しました。さらに、単語の注視時間に対して理論上大きな影響を与えるとされる単語の長さや単語の頻度に関する影響を検討したところ、本データにおいても長い単語や頻度が低い単語ほど注視時間が長くなっており、本研究で収集した眼球運動データの妥当性が支持されました。

本データセットは、Tsukuba Eye-tracking Corpus (TECO)としてオープンサイエンスのプラットフォームにて公開しています (<https://osf.io/wrvj3/>)。

## 今後の展開

本研究チームは、このデータセットを用いた応用研究を進めており、外国語としての英文読解の認知メカニズムや、英文の読みやすさに影響する要因などを明らかにしようとしています。また、本データセットは、今後、データ収集に用いた文章や単語に関するさまざまな統計情報を付与したり、英語母語話者や習熟度の異なる学習者などから収集したデータを追加したりすることで、拡大・拡充を図る予定です。これにより、英語学習や英文読解、さらには自然言語処理や AI 研究などを含めた幅広い研究課題の解決に活用できるようになると期待されます。

## 参考図

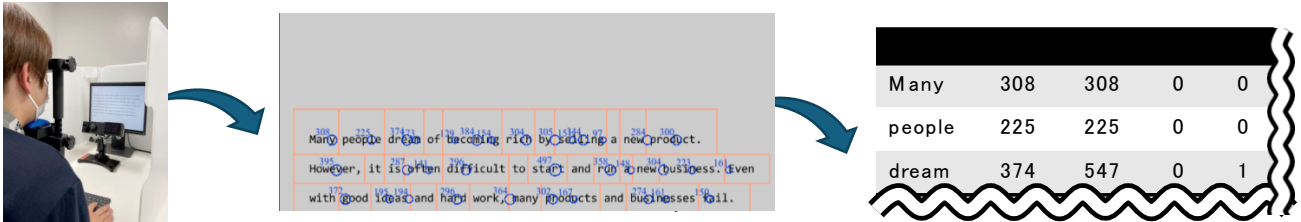


図 本研究デザインの概要。学習者の英文読解中の眼球運動を測定し（左図）、単語に対する注視を記録する（中図；英文中に示された○は読解中に記録された注視を示している）。学習者および単語ごとに注視の継続時間や読み飛ばしの有無などを算出し（右図）、単語の長さや頻度情報を付与した視線計測コーパスを構築した。

## 用語解説

### 注1) 注視

眼球が対象に対して特定の位置に一定時間固定される現象。対象から必要な情報を得るために起こる。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算出される。

## 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】 TECO: An Eye-tracking Corpus of Japanese L2 English Learners' Text Reading  
(TECO: 日本語母語英語学習者による文章読解の視線計測コーパス)

【著者名】 Nahatame Shingo, Ogiso Tomoko, Kimura Yukino, Ushiro Yuji

【掲載誌】 Research Methods in Applied Linguistics

【掲載日】 2024年6月7日

【DOI】 10.1016/j.rmal.2024.100123

## 問い合わせ先

【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](mailto:kohositu@un.tsukuba.ac.jp)