

視線情報と比喻度に基づく英語フレーズの理解度推定

樋笠 泰祐* 平田 明日香* 田中 啓太郎* 森島 繁生†

概要. 本稿では、複数の語で意味をなす英語フレーズに対する読者の理解度を視線情報とフレーズの言語的な性質から推定する手法について述べる。既存手法は視線の動きと単語の頻出度を基に、読者が理解していない英単語を検出していた。しかし、単語とは異なり、複数の語で構成されるフレーズへの視線の動きはより複雑になるうえ、フレーズは平易な単語で構成されながら比喻的な意味を持つため、難しさが本質的に異なる。そこで提案手法はフレーズの構成単語に対する連続した固視とフレーズの比喻の度合いに着目した特徴量を抽出し、教師あり学習によってフレーズごとに理解を判定する。評価実験ではベースライン手法と比較して提案手法がより高い精度で理解度を推定できることが示された。

1 はじめに

英語学習者が新たな語彙を習得する場合、その語彙がどのような状況で使用されるかを理解することが不可欠である [4]。文章読解中の視線を基に理解度を測ることで、読者は内容に集中しながら文脈の中で語彙を学習できる [5][6]。既存研究では英単語を対象に、単語への停留時間や停留回数などの視線特徴量とコーパスにおける単語の頻出度を言語特徴量として機械学習により理解度推定を行っている [2][3]。

英語学習において単語と同等に重要な語彙として複数の語で意味をなすフレーズがあるが [1]、既存の単語への理解度推定手法ではフレーズ内の単語に対して推定精度が低下することがわかっている [2]。これは、平易な単語の組み合わせによって比喻的な意味を持つというフレーズの特性によるものであると考えられる。そこで、本研究ではフレーズを目にした際に特有な視線の動きと定量化されたフレーズの比喻の度合いを利用することで、フレーズごとに理解度を推定する手法を提案する。

2 提案手法

2.1 視線データの前処理

アイトラッカで取得した視線データから、認知活動を示す視線の動きである固視を判定する。まず Garain らが用いた前処理を行い、各時間フレームにおける視点を単語に対して割り当てる [2]。各単語における視線の停留時間が 100 ms を超えた場合、その単語に固視がなされたと判定する。これにより p 番目のフレーズに対する固視集合 $\mathbf{G}_p = \{g_i(w_i)\}_{i=1}^{N_p}$ を得る。ただし N_p は p 番目のフレーズへの固視の

総数を表す。また、要素である $g_i(w_i)$ はフレーズの構成単語 w_i への固視の持続時間を表す¹。

2.2 特徴量抽出

前処理された視線データから 4 種類の視線特徴量を抽出する。まず、フレーズの構成単語への連続した固視を試行と定義する。 \mathbf{G}_p を参照して試行集合 $\mathbf{D}_p = \{d_j\}_{j=1}^{T_p} \in \mathbb{N}^{T_p}$ を得る。ただし T_p は p 番目のフレーズへの試行回数を表す。 d_j は各試行における固視の合計時間で $d_j = \sum_{k=s_j}^{e_j} g_k(w_k)$ であり、 j 番目の試行は固視集合 \mathbf{G}_p において s_j から e_j までの固視に対応している。試行集合 \mathbf{D}_p から最大停留時間 f_p^{dur} を式 (1) のように抽出する。

$$f_p^{\text{dur}} := f_p^{(1)} = \max_{j \in \{1, \dots, T_p\}} d_j \quad (1)$$

また、各試行における固視の回数を $c_j = e_j - s_j + 1$ と計算することで、固視回数集合 $\mathbf{C}_p = \{c_j\}_{j=1}^{T_p} \in \mathbb{N}^{T_p}$ を得る。この \mathbf{C}_p から合計固視回数 f_p^{total} 、平均固視回数 f_p^{mean} 、最大固視回数 f_p^{max} の 3 つの特徴量を以下の式 (2)–(4) のように抽出する。

$$f_p^{\text{total}} := f_p^{(2)} = \sum_{j=1}^{T_p} c_j \quad (2)$$

$$f_p^{\text{mean}} := f_p^{(3)} = \frac{1}{T_p} \sum_{j=1}^{T_p} c_j \quad (3)$$

$$f_p^{\text{max}} := f_p^{(4)} = \max_{j \in \{1, \dots, T_p\}} c_j \quad (4)$$

さらに、英文に含まれるフレーズから言語特徴量を 3 つ抽出する。フレーズの構成単語数と画面に表示される際の横幅をそれぞれ式 (5)–(6) のように計算する。

$$f_p^{\text{word}} := f_p^{(5)} = W_p \quad (5)$$

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

* 早稲田大学

† 早稲田大学理工学術院総合研究所

¹ 固視が行われない単語や、複数回の固視が行われる単語も存在する。

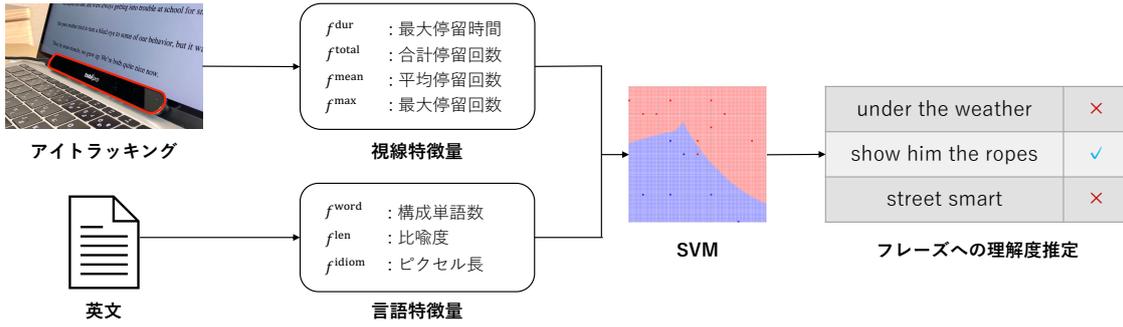


図 1. 提案手法概要図.

$$f_p^{\text{len}} := f_p^{(6)} = \text{right}(p) - \text{left}(p) \quad (6)$$

$\text{right}(p)$ と $\text{left}(p)$ は英文の表示画面におけるフレーズの右端と左端の x 座標をそれぞれ表す². さらに, 学習済みの比喻表現検出モデル DISC[7] を利用し, 各フレーズの比喻度 f_p^{idiom} を式 (7) のように計算する.

$$f_p^{\text{idiom}} := f_p^{(7)} = \frac{1}{W_p} \sum_{k=1}^{W_p} \text{DISC}(p, k) \quad (7)$$

$\text{DISC}(p, k)$ は k 番目の構成単語に対して DISC が出力した値である.

2.3 SVM による理解度推定

得られた特徴量から各フレーズへの理解度を推定するため, 放射基底関数カーネルを用いたサポートベクトルマシン (SVM) を使用する. 特徴量行列 $\mathbf{X} \in \mathbb{R}^{P \times 7}$ を SVM へ入力する. ただし \mathbf{X} は $\mathbf{X}_{pl} = f_p^{(l)}$ を満たし, P は英文に含まれるフレーズの数を示す. この SVM は各フレーズへの理解の二値分類 $\hat{\mathbf{T}} = \{\hat{t}_k\}_{k=1}^P \in [0, 1]^P$ (0 は理解できたもの, 1 は理解できなかったもの) を出力する.

3 評価実験

3.1 データセット

日本語を母語とする 6 人の大学生を対象に, 視線データの収集を行った. 英文は 18 種類あり, それぞれ 100 単語程度で構成され, 10 個程度のフレーズを含む. 収集ではまず, 13 インチのディスプレイに表示された 1 つの英文を参加者が読み, その際の視線を記録した. 簡単な要約課題に答えた後, 文章中の各フレーズの意味の確認を行い, 理解に関するアノテーションを行なった. 以上の流れを 18 種類の英文について繰り返した. 視線データの記録にはスクリーンベースのアイトラッカ Tobii pro nano を用いて 60Hz でサンプリングを行い, 各英文を読む直前にアイトラッカのキャリブレーションを行った. 総数 191 個のフレーズのうち理解できなかったものは平均 64 個 (± 28 個) であった.

表 1. 被験者ごとの理解度推定結果 (AUROC).

	S1	S2	S3	S4	S5	S6	平均
BL1	0.595	0.622	0.683	0.639	0.737	0.643	0.653
BL2	0.668	0.577	0.711	0.701	0.772	0.721	0.692
提案	0.685	0.699	0.738	0.752	0.762	0.679	0.719

3.2 実験条件

作成したデータセットを用いて提案手法の評価を行った. 6 人のうち 5 人のデータでモデルの学習を行い, 残り 1 人のデータを評価する Leave-One-Person-Out 交差検証を行った. 評価指標として Receiver Operating Characteristic (ROC) 曲線の面積 (Area Under ROC Curve, AUROC) を採用した. ベースライン手法には単語への理解度推定手法およびそれらの視線特徴量を, 直接フレーズに適用して得られた結果 (BL1, BL2) を採用した [3][2].

3.3 結果

表 1 にベースライン手法及び提案手法における Leave-One-Person-Out の結果を示す. 6 人中 4 人において提案手法が最も AUROC が高く, フレーズへの理解度推定における提案手法の有効性が確認できた. S5 と S6 に関しては BL2 が最も良い結果となっているが, S2 については提案手法に比べて 0.1 以上低い結果となり精度が不安定であることがわかる.

4 まとめ

本研究では, フレーズに対する特有な視線の動きとフレーズの言語的な特徴を考慮した理解度推定手法を提案した. 提案手法は, フレーズの構成単語に対する連続した固視とフレーズの比喻度を利用して読者の認知負荷を検出し, フレーズごとに理解を判定する. 評価実験ではベースライン手法よりも提案手法の精度が上回り, 提案手法のフレーズに対する理解度推定における有効性が確認できた. 今後はデータセットを拡張し, 様々な習熟度の読者に対する理解度推定を行いたい. また, ディープラーニングを利用してより多様な視線特徴量の抽出を目指す.

² 全てのフレーズは表示画面上で 1 行に収まるものとする

参考文献

- [1] T. Baldwin and S. N. Kim. Multiword expressions. *Handbook of natural language processing*, 2:267–292, 2010.
- [2] U. Garain, O. A. Pandit, O. Augereau, A. Okoso, and K. Kise. Identification of Reader Specific Difficult Words by Analyzing Eye Gaze and Document Content. *14th IAPR International Conference on Document Analysis and Recognition*, 1:1346–1351, 2017.
- [3] R. Hiraoka, H. Tanaka, S. Sakti, G. Neubig, and S. Nakamura. Personalized Unknown Word Detection in Non-Native Language Reading Using Eye Gaze. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 66–70, 2016.
- [4] M. Hismanoglu. Teaching English through literature. *Journal of Language and Linguistic Studies*, 1:53–66, 2005.
- [5] A. Hyrskykari. Eyes in attentive interfaces : experiences from creating iDict, a gaze-aware reading aid. 1 2006.
- [6] R. J. Jacob. Eye movement-based human-computer interaction techniques: Toward non-command interfaces. *Advances in human-computer interaction*, 4:151–190, 1993.
- [7] Z. Zeng and S. Bhat. Idiomatic Expression Identification using Semantic Compatibility. *Transactions of the Association for Computational Linguistics*, 9:1546–1562, 2021.

未来ビジョン

本研究が目指す一つの未来は、英文を読むだけで新たな語彙を学習していくことができるシステムである。英語学習において新たな語彙を学習する場合、現状は単語帳などを利用することが多い。しかし、そのような教材では実際にその語彙を使う状況や文脈を学ぶことは難しいという問題がある。そこで、学習者が英文を読む間に自動で理解度を推定してフィードバックを行うシステムがあれば、学習者は英文の内容に集中しつつ実際の使用例を確認しながら語彙学習も行える。その実現のために、本研究では視線を利用することで読者に意識させない理解度推定を図り、特に日常

会話や小説などの口語表現で頻出ながら学習の過程で軽視されがちな英語フレーズを対象として扱った。

また、本研究の発展として、単語やフレーズなどの単位だけでなく文単位でも理解度の推定を同時に行い、学習者により明示的なフィードバックを与えられると考えている。学習者が理解していないのは単語なのか、比喩表現なのか、それとも文の構造なのかを示すことでより明確な学習の方向性を提示できる。その実現には、自然言語処理技術を組み込んだ上で、より多様な視線の特徴を捉えられるモデルが必要であると考えている。