

文章読後の意識的選好結果と無意識的視行動の関連性分析

Analysis of the Relationship Between Conscious Preference Results and Unconscious Visual Behavior After Text Reading

伊藤 健人[†]
Kento Ito北島 宗雄[†]
Muneo Kitajima中平 勝子[†]
Katsuko T. Nakahira

1 はじめに

近年、文章生成や翻訳の分野では、生成 AI の導入が急速に進んでいる。とくに大規模言語モデルを活用した翻訳技術は、従来の機械翻訳が主に語彙や構文の対応に依存していたのに対し、文脈を考慮した語彙選択、語調や文体の整合性といった高次の言語判断を伴う出力を実現しつつある。Hendy らは、ChatGPT などの GPT 系モデルが、高リソースの言語ペアにおいてメジャーな翻訳システムと同等以上の品質を達成することもあり、より流暢で、不完全な入力にも対応可能と報告している [1]。一方で、専門的な内容や文化的含意を含む文書では、文脈理解の不足や誤訳が生じるケースも多く、人間翻訳者の判断が依然として不可欠である。このように、生成 AI は単なる処理装置ではなく、文章の表現や意味を共に担う協働者としての性格を強めつつある。

翻訳文の質は、単に原文を忠実に他言語へと置き換えたかどうかではなく、読者がその文をどのように理解し、どのような印象を受け取るかといった、読みのプロセスによって大きく左右される。生成 AI による翻訳が一定の段階に達した今、ユーザーは複数の訳文候補の中から、より好ましいと感じるものを選び取るようになってきている。こうした選択は、情報の正確さだけでなく、文の読みやすさ、語彙の自然さ、表現の滑らかさといった主観的な印象にも大きく左右されると考えられる。特に文学などの文体や語感が重視されるテキストにおいては、文の構造や言い回しの微細な違いが、読者にとってのしっくりくる、響きが良いといった感覚に影響を及ぼす。このような主観的選好は、明確な評価基準に基づかないがゆえに曖昧で個人差が大きく、言語化が容易ではない。

読者が自然と感じる文章は、文章の整合性や正確さだけでなく、語感・リズム・読解のスムーズさといった感性的要因が影響していることも多い。言語的処理が負担なく進むことで得られる流暢さや、語感に対する直感的ななじみやすさが、その印象を形作っている可能性がある。情報処理の容易さは、その刺激に対する好意的評価と結びつくと考えられており [2]、この処理流暢性は、言語刺激において、自然さや読みやすさといった主観的評価の形成に寄与する。翻訳文の選好判断もまた、意味理解の後に行われる理性的な評価というより、読解中の無意識的な反応に支えられた直感的評価と考えられる。したがって、文章に対する選好とは、読者の無意識下にあるさまざまな印象や反応の集合として捉える必要がある。

こうした言語化困難な選好に対して、視行動はその認知的過程を可視化する手がかりを提供する。アイトラッキング技術は、読み手がどの語や文節に注視し、どの程度の時間をかけて情報処理を行っているかを計測することにより、テキスト処理のリアルタイムな動態を記録する方法として確立されている [3]。また、人は情動的な刺激に対して、瞳孔径の変化といった生理的な反応を示すことがあることが知られている [4]。

選好判断と視行動の関係を検討した研究では、最終的に選ばれる対象に対して視線が次第に集中していく動態が観察されている [5]。この現象は、視線がすでに形成されつつある選好傾向を反映するだけでなく、逆に選好そのものを形成する要因の一つである可能性も示唆している。好ましさや感性的評価のような主観的評価の形成過程に関しては、画像や商品広告などを対象とした研究において注目されてきたが、言語情報、翻訳文のようなテキストを対象とした研究は限定的である。

以上を踏まえると、翻訳文の評価は必ずしも明示的な尺度や客観的指標のみに基づいてなされるのではなく、読者の中で無意識に形成される印象や感性的応答が密接に関連している可能性が高いと考えられる。生成 AI による翻訳文に対する選好を研究することは、単に AI の出力を評価するという取り組みにとどまらず、人の感性的判断がどのように行われているのかを明らかにする手がかりを提供する。本研究は、文章読解中の視線挙動に着目し、読者の無意識的視行動と翻訳文に対する選好結果との関連性を明らかにすることを目的とする。

2 本稿における仮説

翻訳文の読解において、読者は単一の評価基準に基づいて文章を選好するのではなく、複数の評価項目を用いて翻訳表現を判断していると考えられる。翻訳品質に対する評価項目については、MQM などにより一定の分類枠組みが提案されているものの、とくに印象評価に関しては標準的分類は存在せず、研究者が仮説的に分類を設定するのが一般的である。

そこで本研究では、先行研究における言語処理や感性評価の知見を参照しつつ、翻訳文に対する印象的評価の傾向を捉えるための 4 つの評価項目を仮説的に設定した。これらの評価項目は意識的に選択されるものではなく、翻訳文の読解過程において、自然さや語感、翻訳らしさ、感情的共鳴といった要素が無意識的に評価に影響を与えるものと想定される。

また本研究では、読者が翻訳文を評価する際に用いるこうした評価項目が、読解中の視行動に一定のパターンとして表出すると考える。評価項目の違いに応じて、注視時間や再読傾向、

[†]長岡技術科学大学

注視集中部位といった視線指標に差異が現れることが予測される。以下では、こうした仮説に基づいて設定した4つの評価項目について説明する。

1. 自然さ・流暢性重視

この項目では、読者は翻訳文の意味が即座に理解できるか、語順や語尾が日本語として自然であるかといった処理の流暢性を重視する。翻訳文においては、原文の構造を引きずった直訳的表現や、助詞や文末のぎこちなさが否定的に評価されることが多い。このとき、読者は文全体を滑らかに読もうとし、引っかかりの少ない文章を好ましいと判断しやすい。視行動としては、文全体にわたる平均注視時間の短縮、再読回数の減少などが予測される。

2. 文体・語感重視

この項目では、読者は翻訳文における語感の良さ、リズム感、比喩表現の洗練さなど、文体的な洗練度に注目して評価を行う。これは文学的な美的感受性に近く、読んでいて気持ちいいかどうかが重要な判断基準となる。視行動としては、特定の語句や表現に対する長時間の注視、文末部に対する注視の集中、局所的滞留などが予測される。

3. 翻訳調察知

この項目では、読者は翻訳文の翻訳らしさや直訳感を敏感に検知しようとする。語順の不自然さ、接続詞のぎこちなさ、語彙の選択の曖昧さなどが、機械的翻訳の痕跡として判断材料となる。この評価は人が訳したか、AIが訳したかという判断とも関連しており、翻訳主体に対する推測も伴う。視行動としては、文末や助詞、接続表現などへの注視の集中、局所的な再読、表面上の滑らかさがあっても、判断に迷う箇所での視線停滞が予想される。

4. 情動的共鳴重視

この項目では、読者は翻訳文から情緒的な反応や共感的な感情を引き起こされたかどうかを評価基準とする。物語の雰囲気や登場人物の心理描写、比喩による情景の鮮明化といった要素が感情的な共鳴を引き起こし、それが好意に結びつく。視行動としては、感情を喚起する語句やイメージ表現への繰り返し注視などが予測される。

これらの評価項目は、読者個人に固定されるものではなく、文章の内容やジャンル、読解中の関心、さらには提示順などによって切り替わると想定される。また、1つの翻訳文に対して複数の評価項目が並行して活性化することもありうる。

本研究では、各翻訳文に対する主観的評価と読解中の視線データを突き合わせることで、読者がどのような評価項目に基づいて翻訳表現を判断しているのかを明らかにすることを目的とする。これにより、翻訳文に対する選好判断が単なる意味理解や誤訳の検出にとどまらず、文体や語感、翻訳らしさといった感性的・印象的な評価を含む複数の項目によって構成されていること、そしてそうした判断が読解中の視線挙動として無意識的に表出する可能性を検証する。

3 仮説検証へ向けた予備実験

本実験に先立ち、実験デザインの妥当性を確認するため、1名の参加者を対象とした予備実験を実施した。本予備実験で

は、翻訳文の読解中にアイトラッカーにより視行動を記録し、各翻訳文に対して好ましいと感じた箇所や翻訳主体に関する判断について口頭で回答を得た。この予備実験は、仮説の検証自体を目的とするものではなく、得られる視線データや回答内容に傾向が見られるかを確認し、実験デザインが適切に機能するかを評価することを目的としたものである。

3.1 実験環境

実験参加者になるべく自然に実験に集中できるよう、実験は雑音を排除した減音室内で実施した。室内の照明は一定に保ち、外光による瞳孔径の変化は生じないものとした。視行動の測定には、非侵襲で角膜反射法を用いる Tobii 社製の Tobii Pro Nano を使用した。参加者はアイトラッカーから 60 ± 10cm 程度の距離を保ち、座位でディスプレイに表示される翻訳文を黙読する形式で実験を行った。

3.2 呈示刺激

翻訳の対象となる原文は、複数の著作権フリーの海外文学作品から直接抜粋した。翻訳文はおおよそ 50 文字前後となるよう調整したうえで、3冊の原作から計8つの原文を選定し、それぞれに対して2種類の翻訳を用意した。翻訳文の内訳は、日本人翻訳者による翻訳文、OpenAI 社の ChatGPT-4o による翻訳文である。したがって、実験参加者は合計で 16 の翻訳文を黙読することとなった。

3.3 実験の流れ

実験手順を以下に示す。

1. 実験参加者に対して、実験の概要および注意事項を口頭で説明し、参加の同意を得た。この時点では、翻訳文の作成主体に AI が含まれることは明示せず、「複数の翻訳者による文である」とのみ伝えた。
2. Tobii Pro Nano を用いて視線計測のためのキャリブレーションを実施した。
3. 翻訳文 16 文を 1 文ずつ提示し、各文ごとに約 30 秒間黙読させた。
4. 各翻訳文の黙読直後に、その翻訳文に対して以下の4点について口頭で質問し、iPhone のボイスレコーダーにより回答を記録した
 - 本文中で特に好ましいと感じた箇所
 - その箇所を好ましいと感じた理由
 - この翻訳文が人間による翻訳であると思うかどうかの判断
 - その判断に至った根拠や理由

4 結果と考察

4.1 主観的評価内容の分析と評価項目による分類

翻訳文に対する読者の主観的選好傾向を明らかにするため、参加者が翻訳文読解後に発した好ましく感じた箇所とその理由の内容を、仮説に基づいて設定した4つの評価形式に基づいて分類を行った。

その結果、参加者は提示された 16 文中 10 文について、い

表 1 翻訳主体ごとの好意的チャンクに対する視線指標

翻訳主体	文字数	注視時間 (ms)	1 文字あたり注視時間 (ms)	文字数比 (期待値)	注視時間比	補正注視比 (注視/期待)
AI	72	16844	233.94	0.1885	0.1804	0.96
Human	111	22319	201.07	0.2688	0.2476	0.92

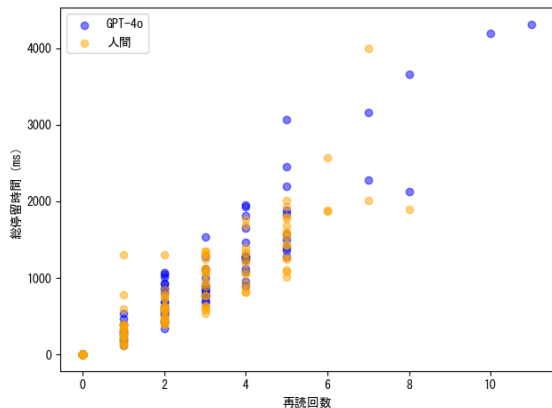


図 1 再読回数と総停留時間の関係：AI と人の比較。

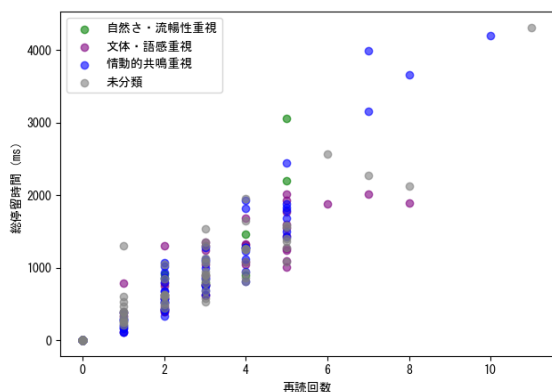


図 2 再読回数と総停留時間の関係：評価項目の比較。

れかの評価観点に基づいた主観的評価を述べており、その内容は文体・語感重視 (5 文)、情動的共鳴重視 (4 文)、自然さ・流暢性重視 (1 文) の 3 観点に分類された。1 名の参加者による評価であっても、文ごとに評価観点を切り替えている様子が確認され、個人内でも評価軸が固定されているわけではないことが示唆された。

特に多く見られたのは情動的共鳴重視に分類される評価であり、「影が自分に語り掛けてくるところが好き」「黄金郷を見つけられなかった心情を表している」といった記述が確認された。これらは翻訳文に対して情緒的な反応を示した好意的評価であり、表現がもたらすイメージ喚起や物語的共感が選好形成に強く関与していることを示唆する。

次に文体・語感重視の評価も複数認められた。たとえば、「子供に語り掛ける口調があたたかい」「キャラクターを表す語尾が良い」といった評価が挙げられており、翻訳文における語尾や言い回しの感触が、選好判断に強く影響していることがうかがえる。

自然さ・流暢性重視の観点では、「不自然だけど絶妙な引っ掛かりがいい」「子気味良い」といった評価があり、不自然さがかえって好印象を与えていた。

一方で、翻訳調査知に該当する明確な選好理由は確認されなかった。

なお、「特になし」と回答された文が複数存在した。これらは、参加者が特に評価しなかった、あるいは評価の理由を言語化できなかった可能性を含む。これらの文については、明確な評価軸が特定できないため、本分類からは除外した。

4.2 翻訳主体判断における特徴

参加者には各翻訳文に対してその翻訳は人が訳したと思うか、およびその判断理由についても回答してもらった。「全体の流れがすっと入ってきた」「言い回しに引っかかるころがない」といった理由から人間翻訳と判断されたケースがある一方、「情報を伝達しただけな雰囲気」「無機質」といった理由により AI 翻訳と判断された文も複数見られた。また、「キャラの口調がなく機械的」「直訳な感じがする」といった翻訳スタイルへの感覚的な違和感が、AI 的と認識される根拠となっていた。

興味深いのは、特に好意的評価が示されなかった文であっても、文の滑らかさや語尾の自然さなどの観点から人間翻訳と判断された例が見られたことである。これは、好悪の判断と翻訳主体の判断が同一の評価基準に依拠しているとは限らないことを示しており、評価と識別が異なる認知的プロセスに基づいている可能性を示唆する。

4.3 視行動との関連

図 1 は、翻訳文を 3 文字単位で分割した各チャンクにおける再読回数と総停留時間の関係を散布図として示したものである。全体的に右上がりの分布が見られ、再読回数が増加するほど、対応するチャンクへの注視時間も比例的に増加する傾向が確認できる。一方で、AI 翻訳群では、再読回数 6 回以上のチャンクが散在しており、再読回数のばらつきがやや大きいことが分かる。

図 2 は、図 1 と同様に各チャンクの再読回数と総停留時間の関係を示したものであるが、翻訳主体 (AI/人) ではなく、評価観点ごとに色分けを行っている。分布を見ると、情動的共鳴重視に分類されたチャンクでは、再読回数・停留時間のいずれもやや高い値を示す傾向があり、一方、文体・語感重視のチャンクでは再読・注視時間ともに比較的少ない傾向が見られ、自然さ・流暢性重視では特定箇所における中程度の再読と停留が確認された。

表 1 は、翻訳主体ごとの好意的チャンクに対する注視傾向を定量的に示したものである。ここで文字数は参加者によって好意的と評価されたチャンクに含まれる総文字数、注視時間 (ms) はそれらのチャンクに対する総注視時間を示す。また 1

文字あたり注視時間 (ms) は、好意的チャンク 1 文字あたりに費やされた平均注視時間であり、処理負荷の高さを示す指標として解釈できる。文字数比 (期待値) は、全チャンクに占める好意的チャンクの文字割合、注視時間比は総注視時間のうち好意的チャンクに対する割合を示す。最後に補正注視比 (注視/期待) は、文字数比を基準とした注視偏りの程度を示す指標であり、値が 1.0 を上回る場合は、好意的チャンクが相対的に強く注視されていたことを意味する。補正注視比について AI 群が 0.96、Human 群が 0.92 であり、いずれも 1.0 を下回っていた。これは、好意的と評価されたチャンクが、文字数比に対して過剰に注視されていたわけではないことを意味する。また、1 文字あたりの注視時間に着目すると、AI 群の 233.94ms に対し、Human 群では 201.07ms とやや短く、AI 訳文の一部において処理困難性が認められた可能性がある。

総じて、標準的な読みの範囲においては、AI と人で類似した視線パターンが示された一方、AI 翻訳においては、ごく一部のチャンクで極端な再読と滞在時間が観察された。これらは、機械翻訳特有の表現の違和感や意味理解の難しさを反映していると解釈できる。

5 まとめ

本研究では翻訳文に対する選好判断と視行動との関連性に着目し、読解中の無意識的な反応がどのように表出するかを探ることを目的とした。現在は予備実験の知見を踏まえつつ、本実験に向けた実験の調整を進めており、実施に向けて準備中である。

謝辞

本研究の一部は科研費 JSPS (22K12284, 代表: 岐阜工業高等専門学校・小川信之, 23K11334, 代表: 長岡技術科学大学・中平勝子) および経営改革促進事業の助成を受けたものである。

参考文献

- [1] Amr Hendy, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify, and Hany Hassan Awadalla. How good are gpt models at machine translation? a comprehensive evaluation, 2023.
- [2] Rolf Reber, Norbert Schwarz, and Piotr Winkielman. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc*, Vol. 8, pp. 364–82, 02 2004.
- [3] Rayner Keith. Eye movements in reading and information processing: 20 years of research. Vol. 124, pp. 372–422. American Psychological Association (APA), 1998.
- [4] Satoshi Nakakoga, Kengo Shimizu, Junya Muramatsu, Takashi Kitagawa, Shigeki Nakauchi, and Tetsuto Minami. Pupillary response reflects attentional modulation to sound after emotional arousal. *Scientific Reports*, Vol. 11, , 08 2021.
- [5] Shinsuke Shimojo, Cosmin Simion, Eiko Shimojo, and Christian Scheier. Gaze bias both reflects and influences preference. *Nature neuroscience*, Vol. 6, pp. 1317–22, 01 2004.