

人と AI の翻訳文における文章選好と視行動の関連性分析

Analysis of the relationship between sentence preference and eye movement behavior in human and AI translations

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

1 はじめに

選好とは、複数の選択肢の中から、人が好んで選ぶ傾向のことを指す。人は、自身の価値観や経験、嗜好などに基づいて、特定の対象を他よりも好む場合がある。選好は、意思決定や行動の重要な要因の一つであり、経済学、心理学、マーケティングなど様々な分野で研究されている。

選好と視行動の関連性については、これまでに多くの研究が行われてきた。Krajbich et al. (2010)[1]では、アイトラッキングを用いて、被験者に2つの食品画像を見せて選好させたところ、最後に見た選択肢が選ばれやすいこと、一方の選択肢を見る時間が長いほどその選択肢が選ばれやすいことなどが確認された。

アイトラッカーは、視線を測定することにより、視線の動きや停留点、瞬きの頻度や間隔、サッケードなどのデータが得られる[5]。アイトラッキングを用いて選好中の視線を計測することは、選好の形成過程を時系列で捉えることを可能にする。これにより、選好の形成に寄与する要因や、選択肢間の比較プロセスを明らかにできる。また、視行動によって、主観的な評価では捉えきれない、無意識下の選好が明らかになる可能性もある。そのため、アイトラッキングを用いた選好中の視線計測は、選好の形成過程や潜在的な選好を探る上で有用であり、人の選好行動について研究を進めるうえで重要だと思われる。

前述したとおり、画像などを対象とした選好と視行動の関連性については既に多くの研究がなされているが、文章単体に着目し、文章読解中の視行動を計測することで、選好と視行動の関連性を研究した事例は、筆者の知る限り見当たらない。

ここで、文章読解は、その行為そのものが、単に文字を追うだけではなく、複雑な認知プロセスを伴う活動である。読者は、文章の意味を理解するために、語彙の処理、文法構造の解析、文脈の把握など、様々な認知的操作を行っている[4]。そのため、複数の文章の中から好みの文章を選択するという行為は、文章の読解と文章の選好という2種類の活動が行われることとなる。

文章読解の認知プロセスを解明するためにも、アイトラッキング技術が用いられている。Rayner(1998)[2]の読書中の眼球運動に関する包括的レビューでは、単語の頻度や文章構造の複雑さなどが注視時間や注視点の移動に影響を与えることが示されている。つまり、文章読解の際の視行動を分析することで、読者の興味や注意の焦点などを推定することができる。

このように、アイトラッカーを用いることで、文章読解、その後に生じる文章選好の研究を深めることができる。文章選好の題材としては、翻訳活動が適当かと思われる。情報社会である現代は、情報を適切に取り扱うリテラシーが必要となる。翻訳は、異なる言語間の橋渡しを行う知的活動であり、原文の意図を正確に汲み取り、適切な言葉で表現するリテラシーが不可欠である。

そのため、本稿では、文章選好と視行動の関連性を探る題材として、翻訳活動を取り上げる。選好対象は、人による翻訳文と AI による翻訳文とする。

人間と AI の翻訳品質を比較した研究としては、Toral et al. (2018) [3] が挙げられる。この研究では、ニュース記事の英語からドイツ語への翻訳において、機械翻訳と人間の翻訳者の翻訳を比較したところ、一部を除き両者の品質に有意な差は見られなかった。

このように、AI の出力が人の出力に迫る現況であるため、AI 翻訳/人翻訳文章に対する選好特性を視行動に基づいて分析することで、文章選好プロセスの検証のみならず、人と AI の差異や、人的リソースの一部を生成 AI に置換することによる生産性向上の可能性が明らかになると考えられる。

本稿では、リテラシーが求められる翻訳という作業を題材に、リテラシーを維持しつつ、人的リソースの一部を AI に置換することによる生産性向上の可能性を探ることを目標とする。

そのための第一歩として、本稿では、人翻訳文章と AI 翻訳文章の選好にかかる視行動の傾向について考察する。

2 本稿における仮説

上記を踏まえ、本稿では以下の2つの仮説を立てる。

1. 文章選好においても視行動に傾向が見られる

選好する際の視行動については、過去の研究から幾つかの傾向があることが示されているが、文章の選好についても、同様の傾向が得られると思われる。ただし、文章読解という認知的タスクと同時に選好判断を行うため、以下のような流れで選好判断と視行動は以下のような流れになると推察される。視線推移のイメージを図に示す。

- 選好判断の初期段階では、両方の文章に同等の注視が向けられる。
- 選好判断が進むにつれて、文章の質的な差異に注視が集中する。
- 最終的な選択直前には、選ばれる文章への注視が増加し、選ばれなかった文章への注視が減少する。

[†] 長岡技術科学大学

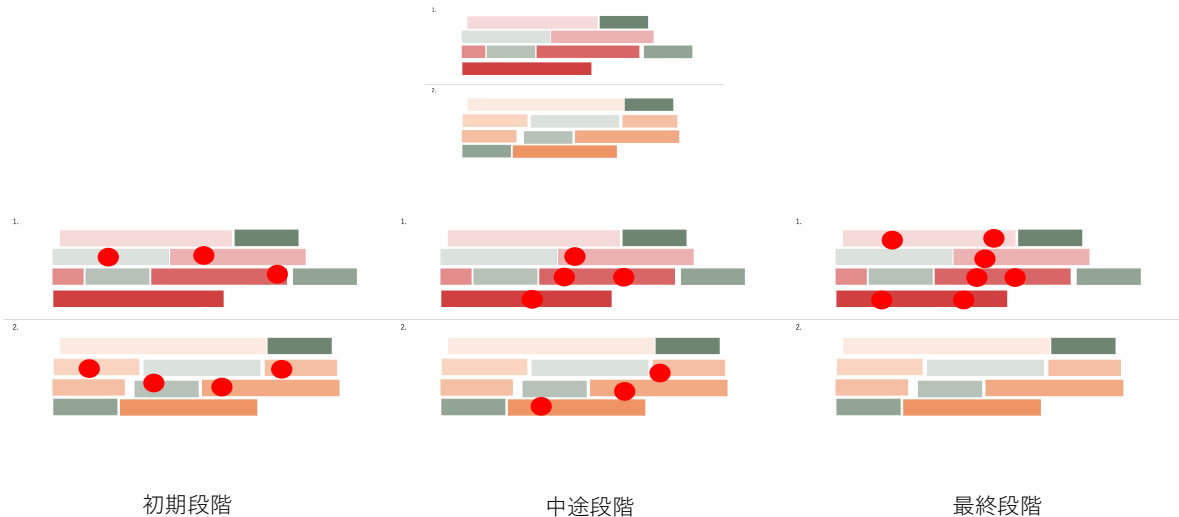


図 1 視行動と選好の関係模式図.

選好判断の初期段階では、実験対象者は両方の文章に含まれる同じような単語や表現に注目しつつ、それらの表現が決定的に異なる部分に視線が向けられると考えられる。例えば、同じ意味を表す単語でも、人間翻訳と AI 翻訳で異なる語彙が選択されている場合、その部分に注視が集中すると予想される。この段階では、文章の全体的な印象を把握するために、両者の共通点と相違点を探索的に読み取ろうとすると考えられる。

選好判断の中途段階になると、実験対象者は両者の文章の表現の違いのみに着目し始めると予想される。この段階では、文章の自然さ、適切な語彙の選択、翻訳者独自の言い回しなど、表現の質的な差異が明らかになってくると考えられる。したがって、これらの差異が顕著な部分に注視が集中し、より好ましいと感じる文章の表現を探索的に読み取ろうとすると予想される。

選好判断の最終段階では、実験対象者は選好する文章の特徴的な部分に視線が偏ると予想される。最終的な選択直前には、選ばれる文章（人間翻訳または AI 翻訳）の印象的な表現や、読み手の心に残る部分に注視が集中し、その部分が繰り返し読まれると考えられる。一方で、選ばれなかった文章への注視は減少し、選好した文章の特徴的な部分への注視が増加すると予想される。選好判断が終了した後は、選ばれた文章の印象的な部分への注視が、判断の正当性を確認するために、一定時間継続すると考えられる。選ばれなかった文章への関心は低下するため、その文章全体への注視は減少すると予想される。

2. 読者のリテラシーによって好む文章表現が異なる

生成 AI は、大量の言語データから統計的な特徴を学習するため、一般的に理解しやすい平易な表現で翻訳を生成する傾向があると考えられる。このような翻訳文は、リテラシーの低い読者にとって、読解の負荷が少なく、好まれやすい可能性がある。一方、リテラシーの高い読者は、より洗練された表現や、創造的な言葉遣いを好む傾向があると予想される。そのため、次のように、リテラシーの高低により、好みが分岐すると考えられる。

高リテラシーの対象者の場合、人間翻訳を比較的好む。対象者が高いリテラシーを持つ場合、人による翻訳文を選好すると考えられる。人間翻訳のより自然で洗練された表現や、独創的な語彙選択などに魅力を感じ、比較的簡素な AI 翻訳より優先的に選択される可能性がある。

低リテラシーの対象者の場合、AI 翻訳を比較的好む。一方で、対象者が比較的低いリテラシーを持つ場合、AI による翻訳文を選好すると考えられる。AI 翻訳の簡潔で明瞭な表現や、一貫した文体などに魅力を感じ、独特の表現を持つ人間翻訳よりも優先的に選択される可能性がある。

3 仮説検証へ向けた予備実験

仮説を検証するため、以下の様な予備実験を行った。

実験対象者のリテラシーを測定するため、リテラシーテストを行った。同一の原文を人と生成 AI のそれぞれに翻訳させた文章を、実験対象者に選好してもらった。その際、アイトラッカーを用いて文章読解中の視行動を読み取った。呈示終了後、休憩をはさみ、口頭アンケートを行った。

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行った。また、室内の照明の状態は一定に保ち、外光による瞳孔径の変化はないものとした。瞳孔径と視行動の測定には、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。実験参加者は、アイトラッカーから 60 ± 10cm ほどの位置で、座視状態でディスプレイに表示される単文を黙読するものとした。実験の被験者は 1 人であった。

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。翻訳文章の文字数はおおよそ 50 文字前後になるよう調整した。3 冊の原作から、4 つの文章を抜粋し、1 つの原文に対して 4 パターンの翻訳を用意した。内訳は、2 人の日本人の翻訳者による該当部分の翻訳文を 1 つずつ、

表1 呈示刺激文章黙読に対する基礎統計量.

stage	停留点数				総停留時間(ms)			
	人翻訳	生成AI翻訳	選好文	非選好分	人翻訳	生成AI翻訳	選好文	非選好分
初期	41	35.5	37.5	39	7499	6391	6344	7284.5
中期	33.5	38	32.5	38	6496.5	5820	6612	6539.5
末期	36.5	31	35.5	31.5	6855	6612	7273	6438.5

openAI社の chatgpt-4o による翻訳文を1つ, Anthropic社の Claude 3 Opus による翻訳文を1つとした. そのため, 実験対象者は合計で16文の文章を黙読することとした. 人翻訳による翻訳文を画面上部に表示, AI翻訳による翻訳文を画面下部に表示し, 実験対象者は常に人翻訳, AI翻訳の1対の文章から選好するものとし, 計16回の選好を行った.

3.3 実験の流れ

実験手順を以下に示す.

1. 実験参加者に実験概要を説明し, 参加の同意を得る.
2. リテラシーテストを行う.
3. 練習問題を呈示, 実験の流れを把握させる.
4. 視線測定機器のキャリブレーションを行う.
5. 1分で人翻訳とAI翻訳の2つの文章を選好させる. 原文を同一とする他の人翻訳文& AI翻訳文の組み合わせも同様にを行い, 計18回行う.
6. 5分間の休憩を挟み, 口頭アンケートに回答してもらう.

3.4 結果と考察

16回の選好で, 実験対象者は人翻訳文を9回, AI翻訳文を7回選好した. 選好判断を段階ごとに分けたくえて視行動を観察すべきと考え, 一分間の選好判断のうち, 開始直後の20秒を選好の初期段階, その後の20秒を選好の中途段階, 終了間際の20秒を選好の最終段階と定義した. それぞれの段階ごとに, 翻訳文のテキストエリア内に収まった視線の停留点と停留時間を, 選好した文章, 選好されなかった文章に分けてまとめたものを表1に示す. また, 同様に, 人翻訳文章, AI翻訳文章に分けてまとめたものも表1に示す.

表1を見ると, 選好された文章・人翻訳文章の方が, 平均注視点数と平均注視時間が比較的高い傾向にあり, 中期段階で平均注視点数と平均注視時間が一時的に減少しているが, 後期段階で再び増加している. 一方, 選好されなかった文章・AI翻訳文章では, 後期段階で平均注視点数と平均注視時間が減少している. これは, 対象者が選好する文章に対しては, 最後まで注意を持続させる一方で, 選好しない文章に対しては, 次第に注意が薄れている可能性がある.

このような傾向から, 実験対象者の選好行動と視行動の関連の可能性が示唆される. 選好された文章は, 他の選好の先行研究同様, より注意を引きつけ, 持続的に読まれる可能性がある. ただし, この結果は平均値に基づくものであり, 個々の対象者や刺激によって傾向が異なる可能性があることに注意が必要である. より確実な結論を導くためには, より多くの実験を行う,

統計的な検定を行うなど, 今後の更なる分析が必要となる.

4 まとめ

本稿では, リテラシーが求められる翻訳という作業を題材に, 人的リソースの一部をAIに置換することによる生産性向上の可能性を探ることを目標とし, その第一歩として, 人翻訳文章とAI翻訳文章の選好にかかる視行動の傾向について考察することとした. 現在は, 予備実験の結果を受けて本実験の計画を修正している最中であり, 準備が整い次第実験を行っていく.

5 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] Krajbich Ian, Armel Carrie, and Rangel Antonio. Visual fixations and the computation and comparison of value in simple choice. Vol. 13, pp. 1292–1298. Springer Science and Business Media LLC, 09 2010.
- [2] Rayner Keith. Eye movements in reading and information processing: 20 years of research. Vol. 124, pp. 372–422. American Psychological Association (APA), 1998.
- [3] Antonio Toral, Sheila Castilho, Ke Hu, and Andy Way. Attaining the unattainable? reassessing claims of human parity in neural machine translation. In Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Christof Monz, Matteo Negri, Aurélie Névoul, Mariana Neves, Matt Post, Lucia Specia, Marco Turchi, and Karin Verspoor, editors, *Proceedings of the Third Conference on Machine Translation: Research Papers*, pp. 113–123, Brussels, Belgium, October 2018. Association for Computational Linguistics.
- [4] 森敏昭, 井上毅, 松井孝雄. グラフィック認知心理学. サイエンス社, 1995.
- [5] 大野健彦. 視線から何がわかるか. 認知科学, Vol. 9, No. 4, pp. 565–579, 2002.