

ポストエディットと持続可能な翻訳の未来

Post-editing and a Sustainable Future for Translators

山田 優
Masaru Yamada

Post-editing is already in practice in the translation industry now; however, many translators and LSPs see human translation (HT) and post-editing (PE) as two very different jobs. The social status of HT and PE is also different, as the divide between HT and PE is becoming worse. But HT and PE really different? In this paper, the author examines this question, mainly by reinterpreting previous studies in translation process research. The results of the original pilot experiment are also provided in this paper. In summary, HT and PE are not so different tasks, especially from the perspective of translation that aims at ensuring accuracy in quality. HT and PE require an equally high degree of skilled translation competence and the same level of high effort to perform MT accuracy error correction. That is to say, the distinction or discrimination against PE that has been believed among practitioners and LSPs may not be correct.

キーワード

post-editing, machine translation, sustainable translation

1. ポストエディットをめぐる問い

1.1 背景

人工知能と言われる近年のIT技術の発展は、多層ニューラルネットワークによる深層学習の発達によるもので、それは機械翻訳にも用いられている。そして機械翻訳の発展が、AI分野全般をリードする最新技術をも生み出してきた。Google翻訳やDeepLなどの機械翻訳に用いられるニューラル機械翻訳の技術である。中核となる技術を挙げると分散表現、注意機構を有するトランスフォーマー、ドメインアダプテーション（転移学習）がある。個々の技術的詳細はさておき、ニューラル機械翻訳という翻訳エンジンの基本処理メカニズムは、原文の言語形式のみに着目して、その言葉の配列とその言葉の分節における位置情報を高度に確率計算し、そ

れに類似した配列を有する訳文を産出しているに過ぎない。本質的に、計算機による言語処理計算は、生身の人間が行う言葉の処理とは異なる、もしくは、非常に限定的な人間の言語処理の一部を行っているに過ぎないのだが、逆に言えば、それだけのことしかしていないにもかかわらず、計算機が高品質な翻訳をしてしまっていることが不思議なくらいだ。

よくある批判として、計算機は人間の言葉の意味を理解できないとか、文脈や言葉の裏にある意図がわからないと言った類のことが挙げられるが、もはやこういう批判は、そもそも人間がどのようにして意味というものを理解しているのかという事実を我々に気づかせてくれるものであり、それは人間自身に向けられるべき自己批判でもある。人工知能や機械翻訳の発達は、人類が解決できていない言語の謎への探求と、その必要性を前景化させたとも言える。

このような現状におけるニューラル機械翻訳の実力は、ある調査では日英翻訳においてTOEIC900点以上であるとか、平均的な大学3-4年生以上の英語力を凌駕したとか言われる¹⁾。しかしながら、プロの翻訳者に求められるコンピテンスは言語学的な能力だけではないので、外国語能力の指標だけを論じてあまり意味がない。しかし実際問題としては、プロ翻訳者が働く翻訳業界においても、一部では、翻訳者が要らなくなる世の中になってしまうのではないかと、といった危機感も出ている。現実的に見積もっても、機械翻訳の存在を無視して今後の産業翻訳の実務ワークフローを考えることができないような事態にもなっている。このような意見が交わされる程までに、機械翻訳は、品質の高い翻訳を達成できるようになった。

さて、実際の翻訳現場で機械翻訳を使いながら翻訳を行う方法はいくつかあるのだが、最も代表的なのはポストエディットである。詳細は後述するが、ポストエディットとは、原文を機械翻訳に入れて出力された翻訳結果を下訳として使い、それを人間翻訳者が修正するようにして最終的な訳文に仕上げていくやり方を言う。ポストエディットには、一般的には2種類あり、フルポストエディットとライトポストエディットに区分される。フルポストエディットは人手翻訳と同等の品質を目指す、ライトポストエディットは原文の内容を把握するため (gisting) の目的などに限定され前者とは異なる。

このようにポストエディット (以下、PE) という翻訳方法は、特に2016年にGoogle翻訳がニューラル機械翻訳を日本語への翻訳で実用化して以来本格化した。それからというもの、翻訳業界においては、PEによって時間的・費用的に効率性の高い作業が実現できるという期待感が高まり、翻訳者への価格破壊を引き起こしたのも事実である。それでも、市場と研究は試行錯誤を繰り返し、当初期待したほどの効率性を達成できないことが徐々に理解され始めた。

この動きと連動して、翻訳会社や実務翻訳者の間では、人手翻訳を生業とする翻訳者とPEを行う者 (ポストエディターという) とで断絶 (デバインド) が深刻化した。人手翻訳とポストエディターとは、全く異なる職種や仕事であるという見方が広まった。確かに、熟達した専門知識とスキルを持ったプロ翻訳者にとって、機械翻訳 (以下 MT) の結果を修正する PE という仕事は退屈であり、積極的にやりたい仕事ではない。また翻訳者が MT (のような悪い) 翻

訳出力を修正するような仕事ばかりしていると、翻訳者自らの「筆が荒れる」ことを恐れるものもある（井口, 2013: 27-28）。それと同時に、PE の仕事を発注する会社側にとっても、プロ翻訳者に MT 修正の仕事をお願いするのは申し訳ないという心情が働き、その結果として、PE の仕事は、翻訳者ほどの熟練したスキルを持たない「別の人」に依頼をし、そしてその仕事に支払われる価格も通常の翻訳と区別された（阪本・山田, 2019）。

1.2 本稿の目的とリサーチクエスチョン

以上が本研究の問題意識の背景となる。日進月歩の IT 技術発展により、翻訳現場が混乱しているのも原因の一つではあるが、事実として、PE という仕事が、伝統的な人手翻訳の仕事とは異なるものと考えられている事が原因の根底にある。そこで本稿では、この問いを検証することをリサーチクエスチョンに掲げる。すなわち、人手翻訳とポストエディットは本当に異なる翻訳作業なのか？ 検証するための方法論として、翻訳プロセス研究における先行研究を再解釈し解説を加え、また新たに執筆するオリジナルのパイロット検証結果を提示する。結論を述べておくと、人手翻訳（human translation、以下 HT）と PE は、異なる作業ではない。厳密には、翻訳品質における正確性を担保するために翻訳を行うという観点からは、HT と PE には、同等の高度な熟練した翻訳コンピテンスが必要で、かつそのために必要な努力（エフォート）も同等である。すなわち、翻訳現場や実務者の間で信じられてきた PE に対する区別や差別は、正しくない可能性がある。

上記の問いを検証することが本稿の主目的である。しかしながら、これまでに、PE に関して日本語で詳細な説明を提供した研究が数少ないことに鑑み、論文の体裁にこだわらず、PE と実務翻訳をとりまく現状を理解するのに読者が必要な背景知識、機械翻訳に関する業界の状況、翻訳プロセスを規定する国際標準規格の ISO 17100、ポストエディットに関する規格 ISO 18587、翻訳の専門職教育を体系化した欧州翻訳修士号（European Master's in Translation）などについてもできるだけ厚く記述することを意識した。また、HT と PE のデバイドの状況をよりわかりやすく説明するために、関連する社会学的分析（ブリュデュー的分析）をおこなった先行研究（Sakamoto, 2019）も紹介する。ポストエディットに関する業界の状況について、本稿を読めばおおむね理解できるように配慮した。尚、本稿の内容は、2020 年 10 月に開催された会議、第 5 回 Translation in Transition（Yamada, 2020）、JTF 翻訳祭 29.5 オンラインウィーク（山田, 2020）での講演内容に基づいている。

2. 現状把握

2.1 機械翻訳の利用状況

まず、産業翻訳業界と機械翻訳の利用状況を把握しておく。翻訳業界調査のシンクタンク

Commonsense Advisory (2018)によると、現状で、全世界で翻訳を必要としているコンテンツのうち、1%以下しか翻訳されていないという。アクティブに活動するプロ翻訳者は、全世界で約20万人程度。仮に、翻訳需要を100%満たすとすると、2000万人の翻訳者が必要であると試算する。またこれらの需要を経済的・主要135言語へすべて翻訳するには20億人の翻訳者が必要であると推定する。翻訳市場は、年間7%の右肩上がりの成長を続けており、翻訳の供給不足という状態は、今後も継続し続けると予測される。

これを別の角度から検証してみる。同じくCommonsense Advisory (2018)によると、世界の翻訳市場規模は約4兆円である。この数字から、世界で1日に翻訳されるワード数を計算してみる。翻訳単価を20円@ワードとすると、約5億ワード翻訳されていることになる。また、上で示したように、世界でアクティブに働く翻訳者数が約20万人であるならば、プロ翻訳者が1日で翻訳できるスループット（翻訳できる量）は約2000ワードなので、1日で約4億ワードが人手で翻訳されていることになる。つまり、1日4億から5億ワードが、人手によって翻訳されているわけだ。

しかし、上述通り、これは翻訳が必要なコンテンツの1%以下に過ぎない（Commonsense Advisory, 2018）。実際には、機械翻訳（MT）を含めて、1日に翻訳される量は1400億ワード以上なのだ（Way, 2020）。つまり、人手では0.2%しか翻訳できていないことになる。

これらから、かなり多くの言葉がMTによって訳されているという事実が把握できる。良い悪いの話ではなく、MTがそれだけ多く利用されているということ。もはやPEや人手翻訳だけで太刀打ちできるような翻訳量ではないことも明らかだ。

2.2 国内外における翻訳会社のMT利用状況

翻訳会社が、業務でMTを利用する割合は、おおよそ15～30%である（Gaspari et al., 2015; JTF, 2017; Sakamoto, 2019）。日本国内におけるMTの利用率は年々増えている。通訳翻訳ジャーナル2020夏号（2020）によれば、MTを使用した翻訳案件の件数・売上が「ゼロ・全くなし」との回答が2017年時点で50%であったのが、2020年では36%になっている。つまり、翻訳会社にとっても、MTを使う仕事はもはや避けて通れない状況になりつつある。

翻訳者にとっても、仕事にMTを「使ったことはない」という回答が2017年の46%から2020年の28%に大幅減となっている。また、PEの仕事に「使ったことがある」と回答した翻訳者も38%で増加している（通訳翻訳ジャーナル, 2020）。翻訳者にとってもMTを使わないで仕事を行うことは、難しくなってきたようである。

他方で、MTの精度・品質に関する満足度はというと、翻訳会社では半数以上が「不満」と回答し、翻訳者もほぼ半々の割合で「不満」と「満足」に割れている。しかしながら、翻訳会社も翻訳者も、MTの影響により、「翻訳分野によっては人手翻訳の需要が大幅に減ると思う」と回答し、ある種の危機感を示す。

このような状況から垣間見られるのは、MT を活用した PE 作業は、実務翻訳の現場においては、もはや現実となっているということ。しかしながら、MT 品質に対しては、冷静に見ている。他方で、分野によっては人手翻訳の需要がなくなるかもしれない、という危機感も抱いている。だからこそ、MT を使う PE の仕事と、人間翻訳者としての自らの立場を差別化しようという意識が働いているのかもしれない。

2.3 ISO 18587 の定義

PE は HT とは別物であると見做されているようだが、ここで PE の定義をはっきりしておく。HT の作業プロセスを規定する国際規格 ISO 17100 と PE を規定する 18587 を参照する。その上で、本節では定義の比較から、まずは HT と PE の違いを考察する。

2.3.1 フルポストエディット

ISO 18587 では、フルポストエディットとライトポストエディットとを区別している。ISO 18587 が正式に認めるのはフルポストエディットのみである。フルとライトポストエディットの違いは、一言で言えば、最終的に仕上げられる品質の違いである。フルポストエディットは、HT と同じレベルの品質を目指す。それに対して、ライトポストエディットは、原文の意味が分かるレベルにまで仕上げればよいとされる。以下に ISO 18487 の定義を記す（ISO, 2017: 8）。

フルポストエディットの要求事項

このレベルのポストエディットにおいて、その最終的な訳文は、正確性、理解可能性、スタイル適合を担保し、かつ統語、文法、句読点においても正しくなければならない。このレベルのポストエディットの目的は、人手翻訳と見分けがつかない訳文に仕上げることである。しかしながら、ポストエディターはできる限り機械翻訳の出力を活用することが推奨される。

本ポストエディットにおいて、ポストエディターは以下の項目に注意を払わなければならない。

- a) 翻訳された情報に過不足がないことを保証する
- b) 不適切な内容（誤訳）があれば、それを修正する
- c) 正しくない意味や分かりづらい意味がある場合は、文を再構成する
- d) 文法的、統語的、意味的に正しい目標言語の内容になるようにする
- e) クライアントやドメインの用語使用に適合させる
- f) 綴り、句読点、ハイフンの規則など適切に適用する

- g) テキストタイプに準拠した適切なスタイル (文体) を用いていること、またスタイルガイドがクライアントから提供されている場合はそれに準拠していることを保証する
- h) フォーマット規則を適用する

*翻訳は筆者による。尚、本論文の説明上、解釈的に翻訳を行った。

上がフルポストエディットの定義である。特筆すべき事として、本文の定義にあるように「人手翻訳と見分けがつかない訳文に仕上げる」という考え方が明示されている点が挙げられる。他方で、「ポストエディターはできる限り機械翻訳の出力を活用することが推奨される」ともある。MT訳を下訳として使いつつ、HTと同じ品質を担保できるのかという実質的な問題はさておき、ISO 18587の定義の限りにおいては、フルポストエディットとHTは、少なくとも最終成果物の品質において、ほとんど同等のものと読める。

そして、具体的にポストエディターが、作業中に注意すべき点として、a) から h) までの8項目も示されている。品質的な観点からは、これらの項目は大きく2つに分類できる。前半の a) ~ c) は正確性に関する事柄で、原文 (の指示的意味) が訳文で表現されているかに関わる。残りの d) ~ h) は流暢性、つまり訳文の文体、読みやすさに関わる。翻訳品質評価において、標準的に、正確性と流暢性の両側面が重視されているのは周知の通りだ。後述するが、翻訳品質評価の業界基準になりつつある MQM (Multidimensional Quality Metrics) や、日本国内では JTF 翻訳品質評価ガイドライン (後で詳述する) にも正確性と流暢性の2つの側面は含まれ、その中身に詳細な項目が肉付けされている。むしろ、この2つの側面は重要であるものの、これ以外の側面も追加された形で、通常の品質評価が行われている。いずれにしても、上のフルポストエディットでポストエディターが考慮すべき正確性と流暢性に関する8項目は、普通の人手翻訳 (HT) が達成すべき品質項目を示しているにすぎない。つまりフルポストエディットとHTは、定義上、あまり変わらないのである。

2.3.2 ライトポストエディット

ISO 18587で正式に認めているのはフルポストエディットであることを述べたが、同規格では付録に、ライトポストエディットも定義している。その定義を以下に示す (p.10)。

フルポストエディットのレベルとは別に、もう1つのレベルが存在するが、これは本書の適応外であり、かつこれが通常用いられるのは、最終訳が出版の使用を意図しない場合、また主として情報収集 (gisting)、すなわち、原文に書かれている大意や重要点のみを把握する目的において、このレベル (のポストエディット) が必要とされる。

本ポストエディットにおいて、ポストエディターは以下の項目に注意を払わなければならない。

- a) 機械翻訳の出力をできるだけ活用する
- b) 翻訳された情報に過不足がないことを保証する
- c) 不適切な内容（誤訳）があれば、それを修正する
- d) 正しくない意味や分かりづらい意味がある場合は、文を再構成する

上からわかることは、ライトポストエディットは出版や社外使用を意図しない、社内使用などに限定された目的のための翻訳であるということだ。明らかに、ライトポストエディットはHT 同等の翻訳品質を想定するフルポストエディットとは違う。注意を払わなければならない項目も、a) ～ d) の4つしかない。このうち、項目 a) は、「機械翻訳の出力をできるだけ活用する」という内容であり、これはフルポストエディットでは本文内に記述されていた。とすると、実質的に翻訳品質に関わる内容は b) ～ d) の3つで、これらはフルポストエディットの a) ～ c) と全く同じ内容である。つまり、ライトポストエディットでは、品質の「正確性」と「流暢性」のうち「正確性」しか考慮しなくて良いことになっている。

これも、実際に、流暢性を全く考慮せずに正確性だけのエラー修正をできるのかという問題は横においておくと、この定義を見る限りにおいては、フルポストエディットとライトポストエディットは、明らかに異なるものである。

このように、ISO 18587 の定義をみただけでも、2つは全く違うものであることは明らかである。この視点で考えると、リサーチクエスションの「HT は PE と異なるのか」という問いは、ISO 18587 の定義上、HT とフルポストエディットを同等のものと考えたと成り立たなくなる。では、フルポストエディットとライトポストエディットは違うのかと言えば、これらが違うのは定義上は自明だ。

2.4 HT と PE の制作プロセスの違い

上では、ISO 18587 の記述の中のフルポストエディットとライトポストエディットの違いについての詳細をみた。本節では、翻訳制作ワークフローの観点から、HT と PE の違いをみる。下記に図も示す。

まず HT を規定する ISO 17100 では、翻訳会社の制作プロセス（production process）において、同規格で定められた資格を有する翻訳者が、翻訳（HT）を行った後に、同じ資格を有する別の「チェッカー（校正者）」によってバイリンガルチェックを行うことが義務付けられている。つまり、翻訳後に、別の人間によってその翻訳物のチェックをしなければならないとされている（図1）。ちなみに、バイリンガルチェックとは、原文と訳文とを見比べながら行う校正

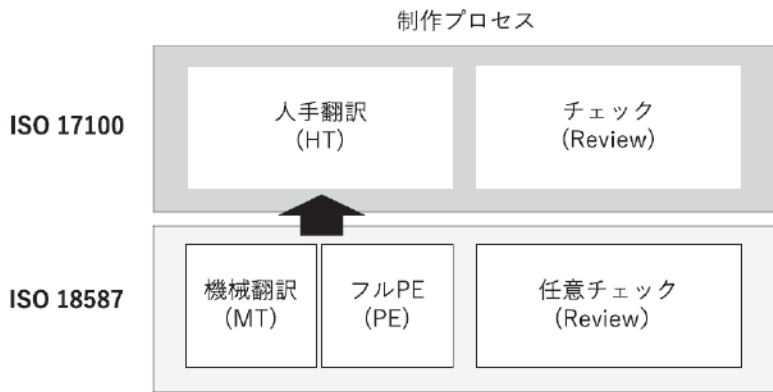


図1：ISO 17100 vs. ISO 18587 の翻訳制作フロー

作業である。また、ISO 17100では「人手翻訳 (HT)」+「チェック (Review)」に加えて、クライアント (翻訳の仕事の発注元) の要望があれば、ブルーフリーディングなどの確認作業も追加されることがある。

では、ISO 18587によるPEの制作プロセスはどうかというのと、上でいう「チェック」に対応するプロセスが「任意」とされている。つまり、ポストエディットの制作プロセスにおいては、「チェック」作業は、顧客の要求に応じて必要性を判断して行えばよく、義務付けられていない。とすると、PEの制作フロー全体で見たときには、チェック作業を不要とするPEは、当然のことながら、HTとは異なる。そして、最終的に納品される翻訳物の品質の観点からも、PEはHTに「劣る」可能性があることは否めない。であるとすると、HTとPEとは異なるものになる。

このように制作プロセスの観点からみると、ISOの基準的には、HTとPEは違うものになる。しかし、上でみたように、ISO 18587のフルポストエディットの定義は、HTと同等であるとされている。とすると、ここで問題にすべき観点は、制作プロセス全体ではなく、個別の制作工程としてのHTとPEである。フルポストエディットがHT同等の品質を達成すべきであるならば、ISO 17100に従って提供している翻訳会社は、全体の作業効率化を図るために、「人手翻訳」に対応する工程を「フルポストエディット」に置き換えても良いはずである。実際の産業翻訳でも、そのような試みは始められている。ISOの定義から考えても、それは理に適っている。

つまり、実務翻訳者の中で議論や批判の対象となっているのは、この文脈におけるHTとフルポストエディットの違いだ。換言すると、これまで人手翻訳を生業としてきた翻訳者が、フルポストエディットをするポストエディターになれるのか、なりたいたいのか、なるべきなのかという問いなのである。定義上は、HTもフルPEも同じ品質を達成しなければならないことになっているが、実質の作業として、両者は異なるものである、と翻訳者は主張する。また作業を依頼する翻訳会社側も、これら2つを必ずしも同じとは考えていないようである。

2.5 ポストエディット元年とこれからの翻訳業界

そもそもポストエディットという考え方自体は、1960年代からあった。当時、米国が国家プロジェクトとしてMTを開発していた。ルールベース機械翻訳の開発は、一定の成果を上げたものの、高品質な翻訳をするという目標を果たすことができなかった。そのため、1966年にはALPAC報告書（ALAPC, 1966）で、MT開発プロジェクトの中断が告げられる。この報告書の中に、ポストエディットに関する記述がある。機械翻訳の品質向上がそれ以上見込めないという閉塞感から、機械に足りない部分を人間が補うためにPEが提案された。それから時代が進み、機械翻訳エンジンの仕組みは、翻訳コーパスデータを使って機械学習をする仕組みへとシフトする。統計的機械翻訳を経て、ニューラル機械翻訳（NMT）の時代になった。今や、NMTの翻訳品質は、飛躍的な向上を遂げ、高レベルの翻訳品質を、それなりに安定的に出力できるようになった。

最近では、ドメインアダプテーションという技術も実用化し、汎用MTから、特定の分野やドメインにカスタマイズされた専用MTを作れるようになった（Sostaric, et al., 2019）。分野によっては、ドメインアダプテーションされた専用MTだけで、人手をほとんど介さずに、高品質の翻訳を出力できるようになると期待される。

ドメインアダプテーションを前提とした機械学習を行って専用MTを作るためには、翻訳データと機械学習とのデータサイクルが不可欠になってくる。この翻訳サイクルを図示すると以下のようになる。

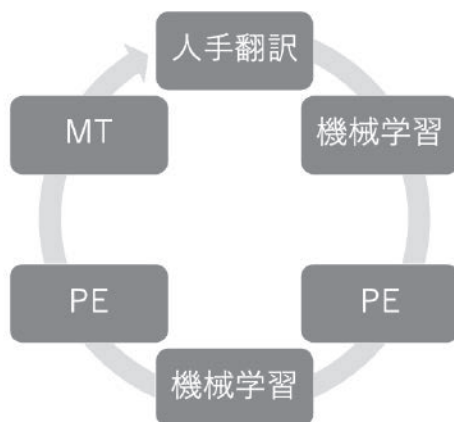


図2：翻訳サイクル（ドメインアダプテーション）

まず、ドメインアダプテーションには、その分野の翻訳データ（原文と訳文が対になったコーパスデータ）が必要になる。データが蓄積するまでは、HTが活躍する。そうして集まったデータは、機械学習に使用される。ちなみに、ドメインアダプテーションは、汎用MT（例え

ば Google 翻訳のように、特に得意とする分野を持たない一般目的用の MT) を、特定の分野の翻訳に対応できるように強化することであり、転移学習とも呼ばれる。人間に例えるなら、基礎的な外国語学力を持つ人が、翻訳者になるために、特定の分野の語彙や文体、知識などを勉強して、その分野の言葉に精通することに似ている。

上の図に戻ると、HT で集まった翻訳データをもとにドメインアダプテーションをした MT は、分野適合した翻訳を出力する。そうして、その MT を使って PE すれば、HT より効率性を向上できる。実際には、フルポストエディットを行う。そうしてエラーが修正された翻訳データは、また機械学習に利用され、MT はさらに向上する。そうすると、次の MT 出力をポストエディットする際は、最初のレベルよりも、より楽な (労力が少ない) PE で事が済むことが期待される。そしてまた、そのデータは機械学習に用いられ、最終的には、ほとんど人手を介さなくても高品質な翻訳ができるような MT が作られる。

このような機械と人間のサイクルは、すでに限定的なドメインでは実現している。そして、このようなサイクルをみると、最終的には HT も PE も必要がなくなってしまうように思える。確かに、限られた特定のドメインでは、そんなことも起きてしまうのかもしれない。しかし、これは、非常に限られたドメインでの話である。具体的には、ある会社のある製品のマニュアルを訳すための MT という具合に限定されたドメインである。ゆえに産業翻訳全体では、常に新しいドメインが出現しているし、また当該ドメイン自体もその製品が更新し続けるので、人間の介入が全く必要なくなることはない。また、上でも述べたように、世界中で翻訳を必要とするテキストの分量は、圧倒的に人手でまかなえるものではないことに鑑みれば、このような翻訳サイクルを脅威として恐れる必要は全くないだろう。

ちなみに、ドメインアダプテーションのための翻訳データ収集のために、国内では NICT が「翻訳バンク」²⁾ という活動を通じて、企業から、さまざまな分野のデータを収集している。また、企業側も、たとえば、製薬会社らが協同して、これまで蓄積してきた翻訳データを機械学習のために提供する動きが活発化してきており³⁾、より具体性をもって、産業翻訳における翻訳サイクルは実現しているのである。

2.6 HT と PE のデバインド

翻訳サイクルが現実化している一方で、依然として翻訳業界においては、PE を差別的に扱う向きがあることを述べてきた。その現状について、Sakamoto (2019) をベースに特に翻訳会社のプロジェクトマネージャーが考える PE に対する意見をまとめる。なぜ翻訳者は PE を拒むのか、という問いをブルデュー的分析で説明する。

Sakamoto (2019) によれば、人間翻訳者のスキルや知識は、ブルデューのいう「文化資本」に相当する。ただしその知は、個人に身体化されるので、絵画や彫刻のように外在化されて所有可能な文化資本とは異なる。しかしながら、翻訳者のスキルは社会文化的に価値あるものと

してリスペクトに値する文化資本であることには変わらない。上でみた翻訳サイクルを実現するためにも、まずはHTによるデータがなければ始まらない。しかし問題は、PEのスキルや技能は、翻訳者のそれと同じようには見做されていないことだ。

翻訳会社で働くプロジェクトマネージャーは、そもそもPEは「退屈」であると思っている。MTの「清掃業務」のようであり（Kelly, 2014）、「屈辱的な作業」であると（Moorkens and O'Brien, 2017: 109）。そのため、PEは報酬面においてもHTより劣る。対して、HTを行う翻訳者に関しては、言語的かつ専門分野の知識を備えている人であり、最低でも5年以上の経験をもつ熟練専門家であるとして尊敬の目でみている。

プロジェクトマネージャだけでなく、翻訳会社（LSP）全体としての態度も同じだ。LSPのウェブサイトの内容の調査では、次のような広告や文句が見受けられる。

- 翻訳会社の中には機械翻訳に頼ってサービスを提供する会社もあるようですが、当社はその分野で最低5年の経験を持つネイティブのプロ翻訳者しか使いません。
- 当社の翻訳は全て人間が行い、正確で信頼性の高い高品質な翻訳を提供します（p. 207）。

また、PEに求められるスキルはHTとは異なると考えているため、LSP社内でもPEの仕事は、翻訳スキルよりもコンピューターを使いこなせる技術や知識を有することのほうが重視されていたり、ポストエディターは時間とコストを節約するための人材であるので、それに特化したマニュアルやトレーニングを社内で提供している場合もある（Sakamoto, 2019）。これは、先のブルデュー的視座でいうところの「経済的資本」を扱う人材としてしか見られていないようである。

阪本・山田（2019）の調査でも、プロジェクトマネージャーが誰にPEを依頼しているのかというフォーカスインタビューを実施した。「理想を言うと優秀な翻訳者さんをお願いしたいが、気が引けてしまう。」「PEはHTと違ってゼロから自分で好きなように文章を作るという翻訳の醍醐味が味わえないし、現状では、単価も悪い」しかし、十分な「翻訳力がないと、ポストエディットもしっかりできないと思う」という意見がでてくる（p. 10-11）。

翻訳者側も、同じように見ている。日本を代表する翻訳者の一人、井口耕二氏は、「ポストエディットと翻訳は全く異なる仕事。ポストエディターという新しい職種がいいか悪いか、自分はそちらにいくのか、いかないのかは、各人が判断することだけど、違う仕事であるということは肝に銘じて判断すべきだと思う⁴⁾」と述べる。また、「ポストエディットが増えれば後輩が育たなくなる」「使いつぶされる個人（ポストエディター）が不幸なのも問題だし、翻訳業界的にもそれで良いのかと」疑問視している。

PEとHTが異なるものと捉えられているのは明らかだ。そして、なぜ翻訳者がPEの仕事を受けたがらないのだろうか、という問いについて、Sakamoto（2019）はブルデューの説明を用

いて以下のようにまとめる (p.211)。

翻訳者は、高い文化資本を保有し、それにより保有資本量も高く、社会的地位も高いポジションを占める人である。しかしポストエディターの文化資本は同じようには認められず、彼らは経済的資本と作業の効率性 (ゆえに経済資本) のみを追求する者、それゆえに保有資本量も低く、社会的地位も低く見られている。このように翻訳者とポストエディターは社会・業界において対極的な地位に置かれている。であるがゆえに、翻訳者はPEの仕事を受けたがらない。そして、もしも業界全体が、このような社会的地位のポストエディターへと引きずり下ろすような態度で、作業を担う人を採用し続けるのであれば、それ自体が業界にとって持続可能ではない状況に陥ってしまうのではないかという危機感をも拭えない。

さて、ここで問題の根底にあるのは、PEとHTは違うと見做されているという事実である。PEとHTのスキルが異なるということに起因する。端的に言えば、HTには高度なスキルが求められるが、PEには高度なスキルは必要ない。果たして、これは本当なのか。それが、本稿の問いである。

3. PEの先行研究

上の問いを反証するために、実証研究を3つ紹介する。しかしその前に、PEに関する先行研究をまとめておく。これまでの研究は、ある意味では、HTとPEとが違うものであるという翻訳業界の見方をサポートしていた側面もある。

3.1 効率性

翻訳研究における過去15年間のPE研究の関心の中心は、効率性であった。PE研究のバイブルともされる Krings (2001) でも、PEにより20%程度の効率性向上が認められている。生産性では、PEによる1日あたりの平均スループットが、2,000から3,500ワードに増加 (Robert, 2013: 32)、5,000ワードにも達する (Guerberof Arenas (2010: 3) との調査もある。効率性では、Yamada (2019) で20%程度の向上、Plitt and Masselot (2010) は74%増とも主張するが、最終的には、言語ペアや状況によって変わってくる (Zhechev, 2014)。ここでの具体的な数字の信憑性については、後述するように色々な批判はあるが、えてして、時間的な効率性において、20%程度の向上は認められるというのが、これまでの見方のようだ。

3.2 品質

HTとPEの品質についても、多くの研究がされている。Fiederer and O'Brien (2009: 62-63) には、品質を正確性と流暢性 (後者を、訳文の読みやすさ、および分野の文体適合に下位分類した) の観点からHTとPEの比較調査をした。結果は、個別指標で見ると、HTとPEに差が

なかった。しかしながら、総合的評価ではHTに軍配が上がった。その理由は流暢性の1つである分野適合した文体を使用しているからという要因によるもので、品質評価においては、翻訳の読者の心理にもっとも影響を与えていることが判明した。しかし、正確性の点からは、PEのほうがHTより優れていた。Plitt and Masselot (2010) は、HTもPEもほとんど同じ品質とし、García (2010) オーストラリアの翻訳資格 (NAATI) の評価指標を用いて、HTもPEは同等の品質を達成していると結論づけている。

3.3 負荷 (Effort) と修正量

さて、ポストエディットの研究でのもう一つの関心は、作業への負荷である。効率性が向上する一方で、作業への負荷が大きくなることが懸念されるからだ。Krings (2001) は作業負荷の測定には、時間 (Temporal)、技術的側面 (technical)、認知的側面 (cognitive) から評価することを説いた。時間的指標は、効率性の研究で先行していたため、技術的側面を計測するために、キーボード操作を記録するツールや、テキストの修正量を測るための様々な指標が開発された。修正量の指標として、Human-targeted Translation Edit Rate (HTER) (Snover et al., 2006) は代表的だ。

Temnikova (2010) は、修正の種類と時間との関係を見た。Koponen (2016) は、単語レベルの修正が時間的効率性と品質改善とのトレードオフからみて、もっとも負荷が低く、効率性の高い修正であることを検証結果から見出した。他方で、語順に関わる修正は、差評者への負荷が高い割りに、最終的な品質向上に寄与しないことを確認した。しかしながら、PEの修正量と時間との相関性はほとんどないことも判明しているため (Tatsumi, 2009)、修正量は、作業負荷を計測する信頼できる指標にはなり得なかった。

3.4 ポーズ

PEの作業負荷を測定するための他の方法も試される。翻訳プロセスを記録して、作業による翻訳操作がポーズ (一時停止) する箇所を観察する方法も研究される。当初は、ポーズ時間が長ければ長いほど翻訳の困難さが高まり、(認知) 負荷が高まるので、それによって負荷を測定できると考えていた (O'Brien, 2006)。しかし、ロング・ポーズは必ずしも認知負荷を正確に反映できず、信頼できる指標ではないことが判明する。

Lacruz and Shreve (2014) は、ショート・ポーズ (300msec) という考え方を提唱する。短いポーズの数とそれが含まれるセグメント (分節) の単語数との関係によって割り出される指標 PWR (pause-to-word ratio) が、高い精度で作業者の認知負荷を示すことがわかった。しかし、その認知負荷の相関を、アイトラッカーを用いた視線の動きと動向拡張から計測できる認知負荷の結果とで見ているため、研究の主流はアイトラッカーを含む包括的な翻訳プロセスデータを収集するプロセス研究が主流となっていく (CRITT TPR-DB 参照⁵⁾)。

3.5 実験間比較の問題

アイトラッカー等の方法により作業負荷を測定する方法は確立するが、Melby, Fields, and Housley (2014) は、実証実験の設定そのものに疑問を投げかける。特に、実験間の結果を比較する場合、そもそも翻訳者が、どのような翻訳の目的と仕様を前提として作業をしていたかにより変わってきてしまうことを指摘した。翻訳理論では、スコポス理論（翻訳の目的を重んじる理論）があることから、翻訳の目的により、同じ原文を翻訳する場合であっても、最終的な訳文の品質が異なる。ポストエディットの例で言うならば、フルポストエディットを想定するか、ライトポストエディットをするのかによって、作業者の負荷は当然変わってくる。

また翻訳品質の標準的な評価方法も、それまで不在であった。翻訳を評価する共通の基準がなかったことにより、研究間の比較も困難であった。2012年頃から、QT in Launchpad⁶⁾というプロジェクトが欧州で立ち上がり MQM (Multidimensional Quality Metrics)⁷⁾という共通の品質評価基準が作成された。これにより、たとえば、正確性に関するエラーも、それが訳抜けなのか、原文に無い語の付加なのか、意味的な歪曲なのかと言った具合に、詳細なカテゴリーに落とし込んで評価することができるようになった。また、評価者間での評価を一致させるための工夫も検討できる。さらには、翻訳のジャンル、目的、仕様によって評価されるべき項目などが変わってくる状況にも対応できる DQF (Dynamic Quality Framework)⁸⁾も開発される。ちなみに日本でも MQM と DQF をベースに業界標準のガイドライン (JTF 品質評価ガイドライン⁹⁾) が策定された。

このように、翻訳者の認知負荷を計測する研究は、プロセス研究の一環として発展をした。他方で、共通の品質評価の枠組みで比較できるような提案もされた。これは自然言語処理分野 (機械翻訳開発) と、翻訳研究分野とでの品質の比較研究 (ゆえに共同研究) を可能にする土壌が整いつつあることでもある。さらには、翻訳産業の実務の場や翻訳教育の場にも応用され始めており、文理・産学をまたぐ学際的研究・教育の発展という面においても、次なる共通の比較ができるステージへと進み始めている。

4. PE と HT は違うのか？

では、本稿の問い「ポストエディット PE は、人手翻訳 HT と違うのか？」について、先行研究および、筆者らが行った新たな研究結果を再解釈してこれに回答を与える。まず、結論を先に述べておくと、人間が「正確性を担保した適訳」を行うという観点からは、PE も HT も同じである。すなわちポストエディターになるにしても、一流翻訳者に近い翻訳力 (コンピテンス) を修得しなければならない。そして、その正確性を確保するための作業には高い負荷 (エフォート) が必要となる。そして、そうであるということは、上述したように、翻訳業界全体が、PE と HT を異なるものであるとし、PE の文化資産を認めないのだとすれば、それは正し

い理解ではない可能性がある。実際には、PE と HT に共通する翻訳力にこそ価値があるのではないか、というのが筆者の主張だ。以下に詳述する。

4.1 NMT のポストエディット

Yamada (2019) では、ニューラル機械翻訳 (NMT) の Google 翻訳を使って、大学生に PE をしてもらった実験を行った。この7年前にも同じ原文を使って、当時は統計的機械翻訳 (SMT) だった Google 機械翻訳でも実験をしているので、それらの結果を比較することができた。NMT は、全体的に翻訳品質が向上しており、品質指標を使って評価をすると、エラー数は、SMT が 31 個、NMT は 10 個しかなかった。PE することを考えると、10 個のエラーを直せばよい NMT のほうが SMT より楽だと予想される。

しかし結果は、そうならなかった。品質結果では、学生翻訳者は、SMT の PE で平均 24 個 (全 31 個) を修正した。残ったエラーが 7 個 (31 マイナス 24)。NMT の PE では全 10 個のエラーのうち 7 個を修正、残ったエラー数は 3 個であった。最終的に訳文に残されたエラーの数でみると、NMT のほうが、SMT より高品質である。

しかし、PE 作業に必要であった作業負荷 (エフォート) を比較すると、SMT も NMT も同等であった (統計的有意差なし)。エラー数がもともと少ない NMT を PE するにもかかわらず、楽にならなかったのだ。

この結果を踏まえ、同研究ではエラーの種類をイシューカテゴリーを用いて分類した。また実験参加者の学生翻訳者の人手翻訳 (ポストエディットではない) の成績についても、同様に分析した。その結果を以下に示す。

図 3 は、SMT、NMT、学生が、どのような翻訳エラーを起こすのかを示している。横軸がエラーの種類で、翻訳イシューカテゴリーに従って X1 から X15 までである。縦軸はそれらのエラーが起こる頻度を示す。NMT を見ると、X3 と X7 のエラーが多いことがわかる。X3 は、原文の意味の歪曲、いわゆる正確性エラーの典型的な「誤訳」である。このエラーと同じくらいの数だけ X7 エラーが起きている。X7 は用語エラーだ。これらの図を参照にしながら NMT をポストエディットすることを想像してみると、修正すべきエラーの多くは、X3 か X7 のエラーである。

これに対して SMT のエラー分布を見ると、さまざまな種類のエラーが発生していることがわかる。NMT のような X3、X7 だけでなく、X4b (過度な逐語訳)、X9 (訳文の構文の誤り)、X10 (助詞の誤り (日本語訳の場合は「て・に・を・は」などの誤り)、X14 (訳文のレジスタ違反)、X16 (結束性違反)、などが含まれる。つまり、この SMT をポストエディットすることを想像してみると、色々な種類のエラーを修正することになるわけだ。

さて、このイシューカテゴリーを使って、実験では PE をする学生の翻訳 (HT) の成績も調査した。学生の HT の成績は、数週間の学習を経ると、上図のように分布が収束されてくる。

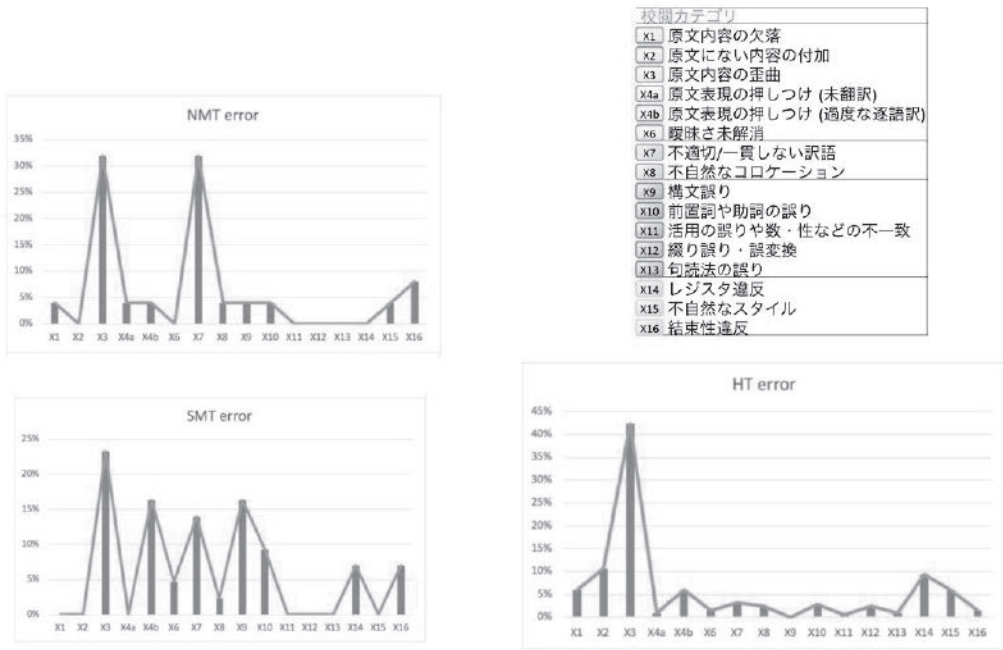


図3：SMT、NMT、HTのエラー分布

すなわち、さまざまな種類の翻訳エラーについては学習が進み減少するのだが、X3に関しては、なかなか向上せず、長い期間、X3エラーが多発する状態が継続する。別の言い方をすれば、熟練した翻訳者として翻訳力を磨く訓練とは、このX3エラー数を極限まで下げることだとも言えよう。尚、このように学習者の翻訳にX3が残るという傾向は、豊島ら（2016）・山本ら（2016）の研究でも示されている。

では、ポストエディットに話を戻す。ここで、上のような実力の学習者が、NMTのPEをすることを考えてみたい。繰り返すが、NMTは、X3とX7の2つのエラーの種類が突出して多く、学習者はX3のエラーが多い。この学生がNMTをPEする場合、問題が生じる。学生はNMTのX7エラーを修正することは容易に行える。なぜかと言えば、学生自身は、X7の用語エラーをあまり犯さないで、それがすぐにエラーであると認識できるのだ。しかし、X3エラーとなると、学生は、NMTのそれをなかなか直すことができない。なぜならば、自分たちも同じX3エラーを犯してしまう実力しかないからだ。

学生にとってはSMTのPEをするほうが、自分たちが犯さないエラーを沢山見るので、容易である。しかし、やはりX3エラーは最後まで修正できずに残ってしまう。

こう考えると、先に述べた、PE時の負荷が、SMTでもNMTでも同じであったことが説明できる。実験前の予測では、もともとのエラーの少ないNMTのPEのほうが、エラーの多いSMTのPEよりも負荷が低い（楽）と考えていたわけだが、実験結果が示すように、作業をする学

生の翻訳の実力を考慮すると、より詳細に説明ができる。つまり、X3エラーをPEで修正するというのは、自分たちもそのX3エラーを犯してしまうことから、難しい作業になってしまう。つまり、認知負荷が高くなる作業になる。そして、そのエラーを修正しようとしても、自分たちの実力では太刀打ちできず、結果的に直すことができない。であるがゆえに、PEをするには、熟練した翻訳力が必要である、という示唆が得られる。すなわち、PEをするということは、正確性に関するX3エラーを正しく修正する能力が必要になるということであり、それはHTとPEの両方に必要なコンピテンスであるから、HT ≒ PEといえるわけである。

4.2 MTエラーと負荷の関係

他の言語ペアでも同じような結果を示す研究がある。Carl and Toledo Báez (2019) は、実験デザインは異なるが、英語から中国語（繁体字）とスペイン語への翻訳をしたプロセスデータを収集して分析を行った。まず、HTとPEで翻訳してもらった。そして、それとは別のタスクとして、その実験参加者は、自分が翻訳した原文と同じものを、Google 翻訳（NMT）にかけ、そこで見つかったエラーに対して、MQM カテゴリー付与をした。そして、エラー付与された箇所について、付与者間の一致率とその箇所を翻訳している時の負荷をアイトラッカーのデータから分析した。

結果は、正確性エラーに関しては、流暢性エラーよりも、付与者間での一致が高く、妥当性のある一致率が確認できた。つまり、MTのエラーを修正すること（i.e. ポストエディット）を想定した場合、正確性エラーは、作業者間で一致して摘出できるので、作業として成立する。それに対して、流暢性エラーは、付与者間でほとんど一致しない。えてして、流暢性に関する修正は、翻訳者個人の好み（preferential）によることが多いからだ。

さて、同研究では、MT上に付与された正確性エラーの箇所について、実験参加者がどのようにPEとHTをしていたのかを分析した。興味深いことに、そのエラー箇所をPEしている時の認知負荷が高いことがわかった。またMTエラー箇所についてPEではなく同じ原文をHTをしているときも、そのエラー箇所を訳す時の負荷が高いことがわかった。換言すると、MTで正確性エラーが起きる箇所というのは、PEをする時もHTをするときも、高い作業負荷を要するということである。

4.3 総合的翻訳力、検索能力

上の研究で明らかになったのは、MTの正確性エラーを修正するには、熟練した翻訳力が必要である可能性があること（Yamada, 2019）、またそのエラーを修正するには高いエフォートが必要であることであった（Carl and Toledo Báez, 2019）。では、正確な翻訳をするために必要な翻訳力とは何か、という疑問が湧く。

一般的に、翻訳力とは、言語的な知識や能力だけではないと言われるが、欧州翻訳修士号

(EMT)では、大学院レベルで必要な翻訳コンピテンスを35項目でまとめている。またこれに対応する形で、ISO 17100でも、プロ翻訳者としての資格と必要な中心的な能力を示している。そのコンピテンスの大枠には、翻訳力・言語力に加え、情報検索能力、異文化能力、テクノロジー能力、専門分野に関する能力、翻訳サービスを提供する能力が含まれる。欧州におけるEMTでは、大学院において、これらの能力を習得し、総合的な翻訳コンピテンスを養うのである。

このような文脈から、Onishi and Yamada (2020)は、上の翻訳コンピテンスの1つである「情報検索能力」に着目し、学生翻訳者とプロ翻訳者の翻訳中の検索操作の違いについて調査をした。非常に興味深い点をまとめると、翻訳中の検索の外部参照の種類を、便宜的に「dictionary (辞書検索)」「search engine results (検索エンジン結果を参照)」「non-dictionary (辞書でない資料、記事、百科事典等を参照)」に区分して、学生翻訳者とプロ翻訳者が費やした時間を分析すると、学生翻訳者はそれぞれの種類を等しい時間だけかけて行っているのに対して、プロ翻訳者は、総合的に学生翻訳者よりも検索に費やす時間が2倍であったという事実に加えて、その種類も、半分以上の時間を「non-dictionary」に費やしていることが判明した。これらが示唆するのは、プロ翻訳者は、辞書を引いたり、検索エンジン結果を眺めるような検索だけでなく、内容に関連する記事やWikipediaなどを含む情報源を検索し参照しながら、翻訳をしているということである。

この調査をベースに、Onish (2020)は、MTのエラーと人間翻訳者の検索操作との関係を調べた。具体的には、以下のような実験スキームで調査した。上のOnishi and Yamada (2020)の実験で用いた原文をMT (Google 翻訳) にかけて、そのMT結果に対してイシューカテゴリーを付与した。先程みたCarl and Toledo Báez (2019)のやり方に似ている。そして、そのMTエラー箇所に対応する箇所を、実験参加者がHTするとき、どのような検索操作を行っているのかに着目した。具体例を示しながら説明を進める。

以下は、Google 翻訳にかけた原文と、その訳文出力結果にイシューカテゴリーを付与した結果だ。プシタッコサウルスという恐竜に関するニュース記事の抜粋だ。短文なのだが、それでも合計5箇所のMTエラーが見つかった。そのうち、X3の正確性エラーが3箇所であった。

Psittacosaurus was a ① smallish, horned dinosaur having complex ② pigmentation.

When Jakob Vinther, a paleontologist, first saw an exceptionally well-preserved specimen of Psittacosaurus, his reaction was: “③ Holy cow, ④ this thing has beautiful color patterns.”

But it wasn't until a few years later that he began to ⑤ wonder whether it would be possible to use those patterns to learn something new about the dinosaur.

プシタッコサウルスは、複雑な② 色素沈着 (X3)を持つ① 小さな角のある恐竜 (X6)で

した。

古生物学者のヤコブペンサーが初めて保存状態の良いプシタッコサウルスの標本を見たとき、彼の反応は次のとおりでした。「③『聖なる牛 (X3)』、④『これには美しい色のパターンがあります (X15)』。」

しかし、数年後、彼は恐竜について何か新しいことを学ぶためにそれらのパターンを使用することが可能かどうか⑤『疑問に思い (X3)』始めました。

これと同じ原文を人手翻訳（HT）した学生とプロ翻訳者のプロセスデータを確認し、上のMTエラーに対応する箇所を、どのような検索操作をしているかをみた。以下に1つ例をあげる。

		Translation (from scratch)	校閲カテゴリ	Research Type					
Students	a	プシタッコサウルスはヤコブペンサーが初めて発見した、角を持ち、 複雑な緑色 をもち恐竜である。	X3	dictionary					
	b	プシタッコサウルスは、 カムフラージュ の小型の角竜だ。		dictionary	search engine	non-dictionary			
	c	プシタッコサウルスは体が小さく、角が生えており、 体表が明暗のカムフラージュ 模様になっているのが特徴だ。		dictionary	non-dictionary				
	d	プシタッコサウルスは小柄で、 複雑な模様 と角を持つ恐竜だ。		dictionary	search engine	non-dictionary	search engine	search engine	non-dictionary
	e	プシタッコサウルスは、 特徴的な色素 を持った小型の角竜である。	X3	dictionary					
Professionals	f	プシタッコサウルスは、 複雑な皮膚色素 を持つ比較的小さい角竜である。		dictionary	non-dictionary	non-dictionary	dictionary		
	g	プシタッコサウルスは角竜類に属する 複雑な緑色 の小型恐竜です。		dictionary	non-dictionary	non-dictionary	search engine		
	h	プシタッコサウルスは小型で角のある恐竜で、 体色に緑が濃淡 がある。		dictionary	non-dictionary				
	i	プシタッコサウルスは、 顔に突起のある比較的小さい恐竜 で、 皮膚の色は複雑な模様 を描いていた。		search engine	non-dictionary	non-dictionary	non-dictionary		

表 1：学生とプロ翻訳の検索操作とエラーの関係

原文の「pigmentation」をMTは、「色素沈着」と訳している。これは、直訳調エラーの一種であるが、この文脈では内容に踏み込んだ訳が必要であると判断し、正確性エラーの一種である原文内容の歪曲 X3 を付与した。記事の流れを加味すると、恐竜の皮膚の色が、その後の別分野の研究の進展に大きなヒントとなるため、「皮膚」「カムフラージュ」というニュアンスの訳語を当てたい。そのためには、おそらく、人間翻訳者は、翻訳している段階で、原文の「プシタッコサウルス」の姿を、百科事典や関連記事を参照して、イメージや写真などで確認する必要がある。つまり、適切な「検索」操作をしているかどうか、適訳にたどり着けるかの鍵となる。

結果をみる。翻訳者 a～e まだが学生翻訳者、A から D がプロ翻訳者である。このうち「pigmentation」を正しく訳せず、MT と同様に X3 エラーを犯してしまったのは学生 a と e の 2 名であった。そして、この二人が、この原文の言葉を訳すために行った検索操作を調べると、ふたりとも「dictionary 辞書」しか見ていなかった。表 3 にまとめてある。ゆえに、2 人とも、辞書的かつ字義的な訳語に置換することしかほぼできず、MT 訳と同じように X3 エラーを犯してしまったのだ。他の翻訳者は、図からわかるように、search engine results や non-dictionary も参照している。

この実験結果からわかるのは、局所的には、MTがおこす正確性エラー X3を回避するためには、翻訳者力として要求されるスキルの1つとされる情報検索能力が必要であるということ。拡大解釈をすれば、翻訳者としての適切なコンピテンスを修得しておかなければ、MTの正確性エラーを修正することができないという示唆が得られる。

5. まとめ：社会的意義、あるいは翻訳の未来のために

以上、HTとPEは違うのか、という問いを検証した。翻訳会社や翻訳者が、ポストエディットは翻訳とは全く異なる仕事であるとして区別することが原因で、翻訳業界全体としても、PEに対する対価の問題や、必要なスキル設定の問題などを抱えている。ポストエディターが翻訳者と同じように文化資本として見做されない差別的な態度なども考察した。

しかしながら、後半の実証検証でみたように機械翻訳の正確性エラーを修正するために求められるスキルは、人手翻訳でも適切なパフォーマンスを行うのと、ほぼ同等のコンピテンスが求められることが示唆された。この観点からみると、HT≒PEなのである。

であるとする、翻訳業界的に、PEという仕事やPEを行う人を、既存の翻訳者と区別をすることは、正しくないのかもしれない。PEもHTも、同じように価値あるものである。そのように捉えることにより、持続可能な翻訳の未来は実現するのではないだろうか。

【謝辞】

本研究のデータ収集を行う機材等のプロジェクトルームを提供くださった関西大学外国語学部、データ収集に協力頂いた外国語教育学研究科の大学院生の方々に感謝する。本稿の内容の一部は、科研費基盤研究(B)『翻訳者の訳出プロセスの可視化と、翻訳・言語研究の共有基盤の構築』(課題番号:20H04486、代表:山田優)の支援を受けた。

註

- 1) 株式会社みらい翻訳「TOEIC900点以上の英作文能力を持つ深層学習による機械翻訳エンジンをリリース」等を参照。<https://miraitranslate.com/uploads/2017/06/2d5778dcdee47e4197468bc922352179.pdf>
- 2) 情報通信研究機構(NICT)と総務省が「グローバルコミュニケーション計画」の一環として行う活動。詳しくはリンクを参照。<https://h-bank.nict.go.jp/>
- 3) 大規模音訳データによる製薬業界向けAI自動翻訳の最適化
<https://www.nict.go.jp/info/topics/2019/10/07-1.html>
- 4) SNSでの投稿による(2020年5月)
- 5) <https://sites.google.com/site/centretranslationinnovation/tpr-db>
- 6) <http://www.qt21.eu/launchpad/content/training.html>

- 7) <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html>
- 8) <https://blog.taus.net/knowledgehub/everything-you-need-to-know-about-dqf>
- 9) https://www.jtf.jp/tips/translation_quality_guidelines

参考文献

- ALPAC (1966). *ALPAC Report, Language and Machines — Computers in Translation and Linguistics*. A Report by the Automatic Language Processing Advisory Committee, Washington, DC.
- Carl, M., and Toledo Báez, C. (2019). Machine translation errors and the translation process: a study across different languages. *The Journal of Specialised Translation*, 31, 107-132.
- Common Sense Advisory. (2018). *Machine Translation for Human Innovation*. Retrieved on Aug 14, 2018: http://www.commonseadvisory.com/machine_translation.aspx
- Fiederer, R., & O'Brien, S. (2009). Quality and machine translation: A realistic objective? *The Journal of Specialised Translation*, 11, 52-74.
- García, Ignacio (2010). Is machine translation ready yet? *Target*, 22(1), 7-21.
- Gaspari, F., Almaghout, H., Doherty, S. (2015). A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives Studies in Translatology*, 23 (3), 1-26.
- Guerberof Arenas, Ana (2010). Project management and machine translation. *Multilingual*, 21 (3), 1-4.
- 井口耕二 (2013) 立教 SFR 翻訳研究プロジェクト (2013). 『翻訳「革命」期における翻訳者養成：公開シンポジウムの報告と今後の取組み』立教大学 (2013年7月1日 <http://www.apple-eye.com/rikkyo/report.pdf> から取得)
- International Standard Organization (ISO). (2015). *ISO 17100. 2015. Translation services – Requirements for translation services*. First edition.
- International Standard Organization (ISO). (2015). *ISO 18587 Translation services – Post-editing of machine translation output – Requirements*.
- JTF. (2017). 2017年度翻訳白書 (第5回業界調査報告書). 一般社団法人日本翻訳連盟 (Japan Translation Federation).
- Kelly, N. (2014). Why so many translators hate translation technology. *Huffpost*, Retrieved on June 19 http://www.huffingtonpost.com/nataly-kelly/why-so-many-translators-h_b_5506533.html (consulted 10.8.2017).
- Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, 25, 131-148.
- Krings, H. P. (2001): *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*, (Geoffrey S. Koby, trans.), Kent, Ohio: The Kent State University Press.
- Moorkens, J. and O'Brien, S. (2017). Assessing user interface needs of post- editors of machine translation. In D. Kenny (ed.), *Human Issues in Translation Technology*, (pp. 110-130). Florence: Taylor and Francis.
- Lacruz, I. and Shreve, G. M. (2014). Pauses and cognitive effort in post-editing. In S. O'Brien, L. Winther Balling, M. Carl, M. Simard and L. Specia (eds), *Post-editing of Machine Translation: Processes and Applications*, (pp. 246-272). Cambridge: Cambridge Scholars.

- Melby, A. K., Fields, P. J. and Housley, J. (2014). Assessment of post-editing via structured translation specifications. In S. O'Brien, L. Winther Balling, M. Carl, M. Simard and L. Specia (eds), *Post-editing of Machine Translation: Processes and Applications*, (pp. 274-298). Cambridge: Cambridge Scholars.
- O'Brien, S. (2006). Pauses as indicators of cognitive effort in post-editing machine translation output. *Across Languages and Cultures*, 7, 1-21.
- Onish, N. and Yamada, M. (2020). Why translator competence in information searching matters: An empirical investigation into differences in searching behavior between professionals and novice translators. *Invitation to Interpreting and Translation Studies*, 22, 1-22.
- Onishi, N. (2020). Translator's research skill improves translation quality: A comparative analysis of MT and HT. *Tenth IATIS Regional Workshop at Kansai University*, Sept, 3, 2020.
- Plitt, M. and Masselot, F. (2010). A productivity test of statistical machine translation post-editing in a typical localisation context. *The Prague Bulletin of Mathematical Linguistics*, 93, 7-16.
- Robert, Anne-Marie (2013). Vous avez dit post-éditrice? Quelques éléments d'un parcours personnel. *The Journal of Specialised Translation*, 19, 29-40.
- Sakamoto, A. (2019). Why do many translators resist post-editing? A sociological analysis using Bourdieu's concepts. *The Journal of Specialised Translation*, 31, 201-216.
- 阪本彰子・山田優 (2019). 「AI時代の産業翻訳：現場最前線の本音覆面座談会」『JTF ジャーナル』, 304, 8-18.
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L. and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In L. Gerber et al. (eds), *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas. Association for Machine Translation in the Americas*, 223-231.
- Sostaric, M., Pavlović, N., and Boltuzic, F. (2019). Domain adaptation for machine translation involving a low-resource language: Google AutoML vs. from-scratch NMT systems. *Translating and the Computer*, 41, 113-124.
- Tatsumi, M. (2009). Correlation between automatic evaluation metric scores, Post-editing speed, and some other factors. In L. Gerber et al. (eds), *Proceedings of the MT Summit XI. Association for Machine Translation in the Americas*, 332-339.
- Temnikova, I. (2010). A cognitive evaluation approach for a controlled language post-editing experiment. In N. Calzolari et al. (eds), *Proceedings of the 7th International Conference on Language Resources and Evaluation. European Language Resources Association (ELRA)*, 3485-3490.
- 豊島知恵・藤田篤・田辺希久子・影浦峯・Hartley, A. (2016). 「校閲カテゴリ体系に基づく翻訳学習者の誤り傾向の分析」『通訳翻訳研究への招待』16, 47-65.
- 通訳翻訳ジャーナル (2020). 「どうなる？ どうする？ 機械翻訳2020」『通訳翻訳ジャーナル2020夏号』イカロス出版
- Yamada, M. (2019). The impact of Google Neural Machine Translation on post-editing by student translators. *The Journal of Specialised Translation*, 31, 87-106.
- Yamada, M. (2020). Post-editing and a sustainable future for translators. Invited Talk on 5th Translation in Transition (TT5), Oct 16, 2020.
- 山田優 (2020). 「ポストエディットと翻訳の未来」『JTF 翻訳祭 29.5: JTF Online Weeks』2020年11月16日, 日本翻訳連盟.

- 山本真佑花・田辺希久子・藤田篤 (2016). 「翻訳学習者の学習過程におけるエラーの傾向の変化」『言語処理学会 第 22 回年次大会 発表論文集』, 865-868.
- Way, A. (2020). The effect of NMT on translators and the translation process. *Invited Talk on 5th Translation in Transition (TT5)*, Oct 17, 2020.
- Zhechev, V. (2014). Analysing the post-editing of machine translation at Autodesk. In S. O'Brien, L. Winther Balling, M. Carl, M. Simard and L. Specia (eds), *Post-editing of Machine Translation: Processes and Applications*, pp. 2-13. Cambridge: Cambridge Scholars.

