解説記事

ニューラル機械翻訳を対象とした試行錯誤的プリエディット事例の探索的分析

宮田 玲[1]　藤田 篤[2]

[1]名古屋大学　[2]情報通信研究機構

1. はじめに

機械翻訳（MT）を実務で活用する方法として、ドメイン適応や用語集登録などの MT システム自体のカスタマイズに加えて、入力文を編集するプリエディット（pre-edit）と出力結果を修正するポストエディット（post-edit）があります。このうち、ポストエディットについては、産業翻訳において取り入れられることも増えており、また、例えば ISO 18587 [1]ではポストエディットのプロセスやポストエディターの技能要件も明文化されています。それに対して、プリエディットについては、長年、研究・実践で取り組まれてきたものの、産業翻訳での活用はポストエディットほど進んでいません。

その理由として、翻訳会社がクライアントから提供された原文書に手を加えることは越権行為であるという規範意識も根強い、という話を聞いたこともあります。また、プリエディット単体では MT 訳の品質を保証することは困難であり、産業翻訳においては、いずれにせよポストエディットなどの出口側の品質保証が求められるため、なかなか工数削減に結びつかないという側面もあるかもしれません。

上記に加えて、筆者らは、プリエディットとはどういう行為・プロセスであり、どのような効果が見込めるのか、といったことが十分明らかになっていないことが問題であると考えています。これまでも文の分割や主語の補完など、有効なプリエディット方法に関する個別の知見は提示されていますが、それらの全体像は明らかではありません。また、入出力を end-to-end で学習するニューラル MT（NMT）は、一般にルールベース MT（RBMT）や統計的 MT（SMT）よりも予測可能性・制御可能性は低いと言えます[2]。これまで制限言語ルール[3]の形で蓄積されてきた知見がそのまま有効に使えるかどうかも定かでありません。

とはいえ、プリエディットは、ここ数年ますます注目されているように思います。筆者らが参加した 2019 年 3 月の言語処理学会第 25 回年次大会のテーマセッション「産業翻訳に役立つ自然言語処理技術」では、参加者を交えた総合討論の中で、少なからぬプリエディットの話題が出され、実際の活用事例の紹介もありました。AAMT のイベントにおいてもプリエディットの事例や研究の報告が複数なされているように[4, 5]、産業翻訳での活用も徐々に進みつつあるようです。まだ数は多くないものの、NMT 向けのプリエディットの研究もなされつつあります[6, 7]。

このような背景と問題意識のもとで、筆者らは NMT 向けプリエディットのベストプラクティスを探求する一歩手前で、プリエディットの全体像の理解を目指した研究を進めてきました。本稿では以下、2021 年 4 月開催の国際会議 16th Conference of the European Chapter of the Association for Computational Linguistics（EACL 2021）に採択された筆者らの論文 “Understanding Pre-Editing for Black-Box Neural Machine Translation” [8]の内容を解説します。理解の解像度を上げること自体がこの論文の目的ですので、本解説においても、論文全体を一般論として要約するのではなく（もちろん適宜要約もしますが）、なるべく具体性を失わない形で結果を紹介します。

Exploratory Analyses of Instances of Trial-and-Error-Based Pre-Editing for Neural Machine Translation
Rei Miyata[1]　Atsushi Fujita[2]
[1] Nagoya University [2] National Institute of Information and Communications Technology
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License.
License details: https://creativecommons.org/licenses/by-sa/4.0/

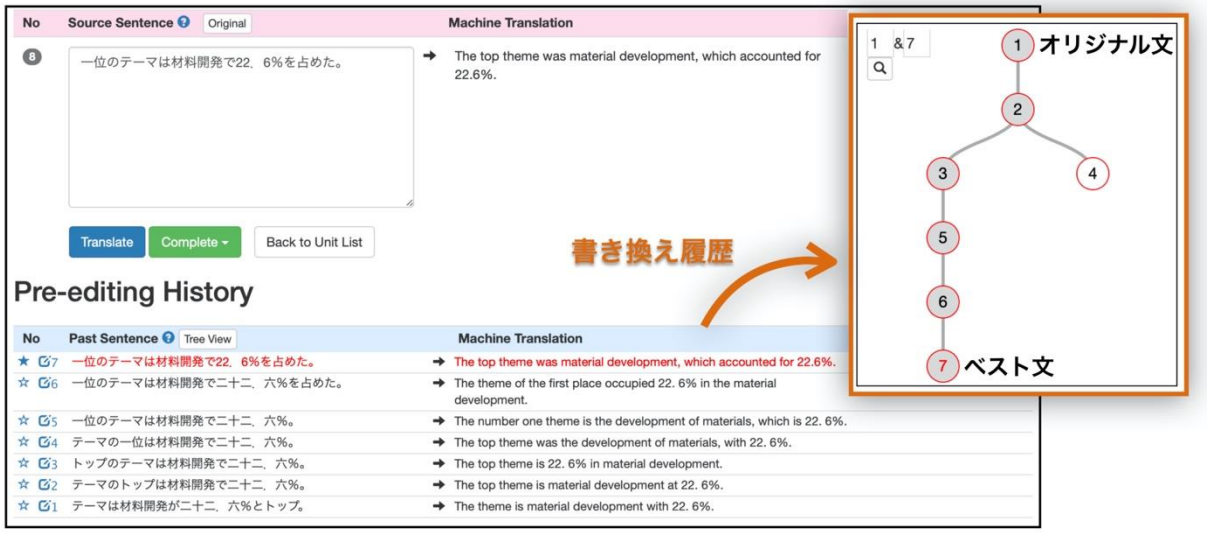

図 1: 書き換え事例収集プラットフォーム

2. 研究の枠組み

本研究は、人手による試行錯誤的なプリエディットの内容・操作とその結果を探索的に分析することを通じて、NMT 向けのプリエディットに関する広範かつ詳細な理解を得ることを目的としています。具体的には、高解像度のプリエディット事例を収集し（3 節）、得られた事例を次の 3 つの観点から分析しました。

- プリエディット後の「機械翻訳しやすい」原文の特徴（4 節）
- 原文書き換え操作の多様性（5 節）
- 各書き換えが NMT に与える影響（6 節）

なお、本研究は、有効なプリエディット手法の提案・検証を目的としていませんが、分析の過程で、どのような書き換え操作が有効そうか、という予備的な知見も一部得られています。

3. 人手書き換え事例の収集

図 1 のプラットフォームを用いて、MT 訳が十分な品質に達するまで、最小単位の原文書き換えを繰り返す、というプロトコルでプリエディットの事例を収集しました[9]。ここでは手順の詳細は省きますが、書き換え作業の主なポイントは以下の通りです。

- 具体的な書き換えの指針は与えられず、MT 訳を見ながら、効果のありそうな原文書き換えを試行錯誤的に繰り返す。
- 過去の書き換え事例は全て記録され、作業者は随時、任意の事例に戻り、別の書き換えを試すことができる。
- 様々な書き換えを試しても MT 訳が十分な品質に達する見込みがないと判断した場合は、作業を途中で終了する。

図 1 に示すように、書き換え履歴は木構造で表すことができます。以下では、書き換え対象の日本語原文（**オリジナル文**とします）とそれに対する書き換え文、および MT 訳の履歴をまとめて、**ユニット**と呼びます。また、ユニット内で最も品質の高い MT 訳が得られた日本語書き換え文を**ベスト文**と呼び、オリジナル文からベスト文に至るパス（図の例では、1→2→3→5→6→7）を**ベストパス**と呼びます。このベストパス上の書き換えには、MT 訳の品質向上に寄与する書き換えが必ず含まれていますので、次節以降の分析でも主な対象としています。

事例収集の対象として、以下の条件を幅広く扱いました。

- **翻訳方向：**日英、日中、日韓
- **MT システム：**Google 翻訳（Google）[10]，みんなの自動翻訳（TexTra）[11]

表 1: 収集した書き換え事例の基本統計

翻訳方向	MT システム	書き換え数		ユニット数	
		合計	平均	オリジナル文＝ベスト文	十分な品質を達成
日英	Google	1332	13.32	27/100	98/100
	TexTra	1260	12.60	23/100	96/100
日中	Google	1371	13.71	2/100	91/100
	TexTra	812	8.12	8/100	94/100
日韓	Google	950	9.50	1/100	96/100
	TexTra	927	9.27	8/100	96/100
	合計	6652	11.09	69/600	571/600

表 2: オリジナル文とベスト文の日本語原文の言語学的特徴

	オリジナル文	ベスト文					
		日英		日中		日韓	
		Google	Textra	Google	Textra	Google	Textra
文長	25.4	27.8	26.9	28.6	27.1	27.8	26.9
係り受け距離	1.95	1.97	1.99	1.99	1.99	2.00	1.98
係り受け最大深度	3.57	3.73	3.68	3.73	3.77	3.78	3.76
延べ語数	2538	2779	2685	2861	2709	2780	2693
異なり語数	1010	1074	1060	1106	1061	1068	1055
延べ語数/異なり語数	2.513	2.588	2.533	2.587	2.553	2.603	2.553
単語の頻度ランク（中央値）	170	143	154	143	155	143	169.5

- **テキストドメイン：**病院内会話、自治体生活情報、日本語で書かれた新聞記事（BCCWJ）、英語で書かれ日本語に翻訳された新聞記事（Reuters）

各テキストドメインから 25 文ずつ、計 100 文を対象に、翻訳方向、MT システムの全 6 通りの組み合わせについて、それぞれ 1 名の作業者（翻訳および MT 評価の経験を有する）を割り当てました。事例収集は、2019 年 12 月から 2020 年 2 月にかけて実施しました。

収集した事例の統計を表 1 に示します。計 600 ユニットに対して、6652 事例が収集されました。特筆すべきは、約 95%のユニットにおいて、十分な品質の MT 訳が得られたことです。NMT に対するプリエディットの高い有用性を示す結果だと言えます。

ちなみに、日英の SMT システム（TexTra）を対象とした先行研究[12]においても、プリエディットの潜在的な有効性は示されていますが、書き換え回数の平均は約 30 回/ユニットと、今回の結果を大幅に上回ります。両方の事例収集に協力してくださった作業者から、今回はだいぶ作業がやりやすかった、という所感もありました。実験条件の違い（MT の訓練データは不統一）には注意が必要ですが、現在の NMT を用いたプリエディットの効率は、以前よりも高いと言えそうです。

4. 機械翻訳しやすい原文の特徴とは？

機械翻訳しやすい原文の特徴を観察するために、翻訳方向、MT システムの各組み合わせ 100 ユニットずつについて、オリジナル文とベスト文の日本語文を比較しました。構文・語彙に関する定量的な指標を計算した結果の一部を表 2 に示します。

構文的な特徴の指標としては、文長、係り受け距離、係り受け最大深度を用いました（いずれの指標も 100 ユニットの平均値を使用しています）。興味深いことに、

いずれの指標においても、構文的な複雑さは僅かに増しています。これまでプリエディットの原則として、なるべく文長を短くすることや文構造を簡単にすることが提示されてきましたが、その原則を支持する結果とはなりませんでした。当然、今回の結果から「文を短くすることは有効ではない」という結論は導けませんが、少なくとも「ほとんどの文は、短くしなくても機械翻訳でうまく訳せる表現に書き換えられる」ということが明らかになりました。この点については、次節でも改めて触れます。

語彙的な特徴の指標としては、延べ語数、異なり語数、延べ語数と異なり語数の比率、Wikipedia における単語の頻度ランク（値が小さいほど Wikipedia で頻出することを示す）を用いました。プリエディットを経ることで、全体として延べ語数、異なり語数ともに若干増えたことが分かります。また、延べ語数/異なり語数の値も上昇していることから、新しい語が多く導入されたというよりも、特定の語への偏りが増えたという傾向が見て取れます。また Wikipedia における単語の頻度ランクの中央値が上位に移動しています。本稿では詳述しませんが、オリジナル文とベスト文で変化が起きた語の分布を調べると、Wikipedia における頻度が低い語が回避され、より高い語が使用されるようになる傾向が観察されました。使用した NMT の訓練データにはアクセスできないため、低頻度語の定義は Wikipedia を用いた擬似的なものに過ぎない点に注意が必要ですが、先行研究[13]においても NMT の弱点の一つとして指摘されている低頻度語を避ける形で書き換えが進んだことが分かります。

5. どのような書き換えが可能か？

続いて、機械翻訳しやすい表現に至るまでに使用された書き換え操作のタイプについて整理します。翻訳方向、MT システム、テキストドメインの計 24 通りの組み合わせから 10 ユニットずつ無作為に選んだ合計 240 ユニットを対象に、ベストパス中の書き換え事例 935 件を、書き換え方法の観点からボトムアップに類型化しました。

その結果、構造、句、内容語、機能語、表記、エラーの 6 つのカテゴリからなる 39 タイプの書き換え操作を同定しました（表 3）。最も頻繁に用いられたのは、**C01「類義語の使用」**でした。MT 訳において適切な訳語を得ることを目指したものであり、様々な語の置き換えが試されていました。次に頻度の大きい**C07「内容変更」**も基本的には単語の置き換えですが、こちらは文脈知識や外界知識に基づくより高度な書き換えを指します。例えば、「根室署」に対する「根室警察署」はいわゆる辞書的な類義語ではなく、文意をふまえて常識を補完することで導けるものです。

S01「文分割」が 14 件（1.5%）しかない点は、4 節の構文的特徴に関する知見を裏付けます。ちなみに、分割された 14 文の内、9 文は文長が 60–67 語とかなり長いものでした。NMT にとっては依然として 60 語を超えるような長文の処理は難しいことが、先行研究においても実証されています[13]。この知見が、プリエディットの観点からも補強されました。

さらに、翻訳学の知見[14, 15]を参考にしながら、情報方略の観点からも事例を分類しました。その結果、書き換えによってテキストの意味・含意が明確に表現される**明示化**は 329 件、その逆の**暗示化**が 88 件、いずれでもない**情報保持**が 480 件という結果になりました。類義語に置き換えるような情報保持が一番多いものの、明示化が多く観察された点は重要です。表 4 は、明示化の事例を類型化した結果です。曖昧な表現を避けたり、暗黙の前提とされる情報を言語化したりすることは、機械翻訳しやすさにつながる方略と言えそうです。

6. NMT の出力結果はどれだけ変わるか？

最後に、プリエディットが NMT の出力結果に与える影響を見ていきます。図 2 は、表 3 の書き換えタイプごとに、原文の書き換えがもたらした NMT 出力文字列の変化量の分布を示しています。変化量は、TER（Translation Edit Rate）[16]という編集距離の指標を用いました。

表 3: 書き換えタイプの類型と事例頻度

ID	頻度	構造	ID	頻度	内容語
S01	14	文分割	C01	118	類義語の使用
S02	27	文構造の変更	C02	21	省略語の回避・使用
S03	19	主題構造の使用・回避	C03	14	照応表現の使用・回避
S04	14	主語・目的語の追加	C04	11	強調表現の使用・回避
S05	12	体言止めの回避・使用	C05	30	カテゴリの表示・非表示
S06	4	態の変更	C06	10	辞書的説明
S07	5	その他構造	C07	94	内容変更
ID	**頻度**	**句**	**ID**	**頻度**	**機能語**
P01	71	読点の追加・削除	F01	47	助詞の変更
P02	28	チャンキングの使用・回避	F02	31	複合辞の変更
P03	31	語句の順序変更	F03	12	アスペクトの変更
P04	7	修飾の仕方の変更	F04	4	テンスの変更
P05	40	接続表現の変更	F05	11	モダリティの変更
P06	36	並列表現の変更	F06	10	敬語の使用・回避
P07	16	同格表現の変更	**ID**	**頻度**	**表記**
P08	13	名詞句・動詞句の交替	O01	61	日本語表記の変更
P09	28	複合語の使用・展開	O02	16	全角・半角の変更
P10	17	接辞の使用・解除	O03	2	記号の追加・削除・置換
P11	5	サ変名詞表現の変更	O04	4	その他表記
P12	9	形式名詞表現の変更	**ID**	**頻度**	**エラー**
P13	5	その他語句	E01	22	表記誤り、文法誤り
			E02	16	内容的な誤り

表 4: 明示化方略の下位分類・頻度・事例

明示化方略	頻度	書き換え例	MT 出力結果
情報の追加	142	12 日は台湾の休日のため休場。	The twelfth is a holiday in Taiwan.
		→ 12 日は台湾の休日のため**株式市場は**休場。	The stock market was closed on the twelfth due to a holiday in Taiwan.
関係の明示	103	来院しなくても 10 日前後で登録のクレジットカードから**引き落としを行います**。	Withdraw from your registered credit card in about 10 days without visiting the hospital.
		→ 来院しなくても 10 日前後で登録のクレジットカードから**引き落としが行われます**。	Even if you do not visit the hospital, your credit card will be debited in about 10 days.
意味の限定	54	採尿と採便を**出して**ください。	Please collect urine and feces.
		→ 採尿と採便を**提出して**ください。	Please submit urine and stool samples.
正規化	30	単位は億円。	Figures are in billions of yen.
		→ 単位は億円**です**。	The unit is 100 million yen.

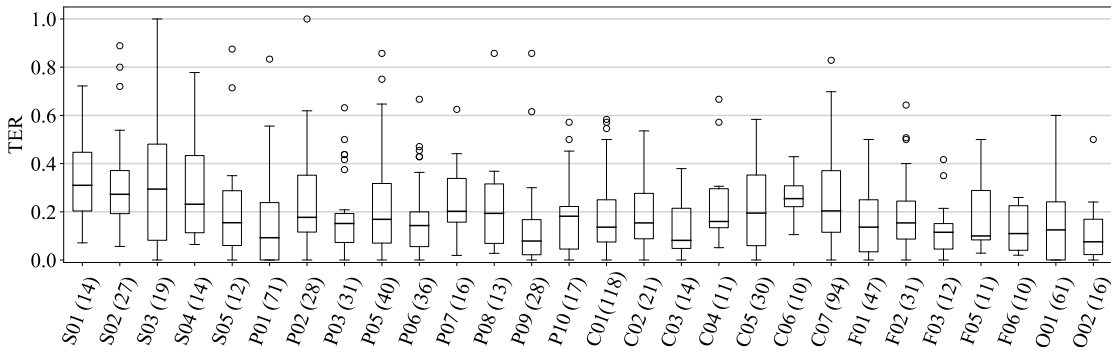

図 2: 書き換えタイプごとの MT 変化量（TER）の分布（括弧内の数字は書き換え事例数）

それぞれの書き換えタイプに注目すると、変化量の分布には幅があることが分かります。最小単位の原文書き換えで、がらりと出力が変わることもあれば、それほど変わらないこともある、という NMT の出力を予測することの難しさが具体的に可視化されたといえるでしょう。一方で、書き換えタイプ横断的に見てみると、出力の変わりやすさに一定の傾向はありそうです。例えば、構造レベルの書き換え（特に S01–S04）は MT 訳に大きめの変化をもたらしている一方、機能語の変更が NMT に与える影響は小さいということが観察できます。

また、**P03「語句の順序変更」**はそれほど NMT の結果に影響を及ぼしていません。これまで SMT においては、自動的な語順の事前並び替え手法が翻訳精度の向上に有効であるとされてきましたが、NMT においてはあまり効果がない（場合によっては逆効果）という報告もなされています[17, 18]。今回の結果から、最新の NMT は人間が行うような語順の変化に対しても頑健であることが示されており、プリエディットの方法として語順変更は必ずしも有用であるとは限らないことが示唆されます。

7. おわりに

ここまで解説してきた内容を、「プリエディットはうまく活用できれば NMT の品質を十分に高められる」、「多様な書き換え方法が存在する」、「明示化は大事」、「NMT の挙動は読めないけれど、一定の傾向はある」といった抽象度でまとめてしまうと、これまで何となく分かっていたこと/行われてきたことに回収されてしまいます。本研究の意義は、まさにこのような何となく考えられてきたことを、具体的なレベルで愚直に調査したことにあります。もちろん、その過程で、プリエディットを経ても必ずしも文構造が簡単になるわけではない、という意外な発見もありましたが、あえていえば、それは派生的なものです。ここで重要なことは、平均 10 回程度の細かい原文書き換えにより 95% の事例で十分な MT 訳品質が達成されたことや 6 カテゴリ 39 タイプの書き換え方法が同定されたことなどの具体的な記述です。

本稿で示したような解像度・範囲で記述することで初めて、これをベースとした具体的な議論が可能になります。「他にもこのような書き換え方法があるのでは」、「内容変更の方法としては具体的にはどういうものがあるのか」、「別の条件で試したらどうなるのか」といった疑問も湧いてくるでしょう。筆者としては、本研究の知見を起点として何かしらの議論（反論も含めて）が芽生えることを期待しつつ、解説を終えたいと思います。

謝辞

本研究の一部は、科研費若手研究（課題番号：19K20628）、科研費基盤研究（S）（課題番号：19H05660）および KDDI 財団調査研究助成の支援を受けました。データの一部は、総務省の情報通信技術の研究開発「グローバルコミュニケーション計画の推進 -多言語音声翻訳技術の研究開発及び社会実証- I.

多言語音声翻訳技術の研究開発」から提供を受けました。

注・参考文献

[1] ISO (2017). ISO 18587:2017 Translation services – Post-editing of machine translation output – Requirements.

[2] もちろん、MT システムがブラックボックスである限りは、RBMT でも SMT でも、出力そのものを予測することは困難ですが、NMT とは異なり、翻訳前後の語や句は陽に対応付けられているため、「この部分を変えれば、この部分が変わりそうだ」という予測は立てやすいといえます。しかしながら、この予測可能性に関する想定自体が、抽象的に仮定されたものに過ぎず、データに基づく検証が十分なされていないことを、ここでは強調したいと思います。

[3] 例えば、Kuhn, T. (2014). A survey and classification of controlled natural languages. *Computational Linguistics*, 40(1): 121–170. において様々な英語の制限言語が提示されています。日本語の制限言語については、Miyata, R. (2020). *Controlled Document Authoring in a Machine Translation Age*, Routledge, London. で整理されています。

[4] 星井智 (2019). 句点挿入と倒置：日英 MT における長文プリエディットの二つの手法. AAMT 2019, Tokyo.

[5] 早川威士 (2019). 機械翻訳に対するプリエディットのタスク分類に関する考察. AAMT 2019, Tokyo.

[6] Hiraoka, Y. & Yamada, M. (2019). Pre-editing plus neural machine translation for subtitling: Effective pre-editing rules for subtitling of TED talks. *Proceedings of MT Summit*, pp. 64–72.

[7] Mehta, S., Azarnoush, B., Chen, B., Saluja, A., Misra, V., Bihani, B. & Kumar, R. (2020). Simplify-then-translate: Automatic preprocessing for black-box translation. *Proceedings of AAAI*, pp. 8488–8495.

[8] Miyata, R. & Fujita, A. (2021). Understanding pre-editing for black-box neural machine translation. *Proceedings of EACL*, pp. 1539–1550. https://www.aclweb.org/anthology/2021.eacl-main.132/

[9] 本研究において、「十分な品質」は「多少流暢さに欠けても、情報の過不足がなく、文法的にも正しい訳文」としました。また「最小単位の書き換え」とは「文法性を保ちながら、それ以上細かくは分割をできない最も単純な書き換え」としました。その他プロトコルの詳細は、元の論文もしくは以下の文献をご参照ください。
宮田玲・藤田篤 (2017). 機械翻訳向けプリエディットの有効性と多様性の調査. 通訳翻訳研究への招待, 18: 53–72.

[10] Google 翻訳. https://translate.google.com

[11] みんなの自動翻訳＠TexTra. https://mt-auto-minhon-mlt.ucri.jgn-x.jp

[12] 宮田玲・藤田篤 (2017). 注 9 番参照.

[13] Koehn, P. & Knowles, R. (2017). Six challenges for neural machine translation. *Proceedings of NMT*, pp. 28–39.

[14] Vinay, J.-P. & Darbelnet, J. (1958). *Stylistique Comparée du français et de l'anglais*. Didier, Paris, trans. and ed. by J. C. Sager & M.-J. Hamel (1995) as *Comparative Stylistics of French and English: A Methodology for Translation*. John Benjamins, Amsterdam.

[15] Chesterman, A. (1997). *Memes of Translation*. John Benjamins, Amsterdam.

[16] Snover, M., Dorr, B., Schwartz, R., Micciulla, L. & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. *Proceedings of AMTA*, pp. 223–231.

[17] Zhu, Z. (2015). Evaluating neural machine translation in English-Japanese task. *Proceedings of WAT*, pp. 61–68.

[18] Du, J. & Way, A. (2017). Pre-reordering for neural machine translation: Helpful or harmful? *The Prague Bulletin of Mathematical Linguistics*, 108: 171–182.