

機械と人間との協働 LSP perspective

～学習型機械翻訳から翻訳プラットフォーム～

梶木正紀

株式会社MK翻訳事務所

1 はじめに

弊社は、特許出願明細書等を専門に翻訳する特許翻訳会社である。2016年11月からニューラル機械翻訳(以下、NMT)を翻訳業務に導入し、NMTに基づくサービス、Hybrid Translation(以下、HT)の発売を開始した。社内検証の結果、HTを使うと生産性が約20%高まることが確認された。2017年は、同サービスの広告活動に力を注いだ。販売まではこぎつけることができたものの、2018年1月に、HTを断念することに至った。その改良として、プリエディットの活用と、学習型機械翻訳(以下、AMT)の導入と、を検討している。最後に、翻訳プラットフォーム(以下、TP)を介した人間と機械との協働の将来性についても述べる。

2 弊社の機械(NMT、MTPE)との協働例

2016年11月に、ニューラル機械翻訳(以下、NMT)と、トランスレーション・メモリと、クラウド翻訳システムとを組み合わせたHybrid Translation(HT)の発売を開始した[1]。2017年は、HTの検証を行いつつ、販売を行った。その間、日本翻訳連盟に対して機械翻訳に関する記事を執筆した[3]。同年7月には、大阪で開催された京都国際翻訳通訳フォーラム(通称:翻訳cafe、会場:関西大学)にてパネリストとして機械翻訳に関する私見を述べた。また、2017年9月には、名古屋で機械翻訳サミット(主催:アジア太平洋機械翻訳協会、会場:名古屋大学)が開催され、弊社の機械翻訳に対する取り組み及び今後の展望に関してプレゼンテーションを行った[2]。このような広報活動を行いつつ、HTの改良に努めた。受注できたクライアントからはHTに対するネガティブな反応はなかった。機械翻訳と人間翻訳の差が大きいというようなコメントもなかった。2017年中は、広報活動には力を注いだものの、積極的に営業活動することはなかった。2018年1月、HTの継続を断念することになった。以下に、積極的に営業活動を行わなかった理由、HTを断念した理由を記す。

3 HTの断念

3.1. HT断念理由1:業務効率の問題

HTを行っても、生産性は20%程度しか上がらない。N

MTを使用するセグメントでは、用語を制御できない。結果、ポストエディット(PE)が大量に必要である。生産性を20%ほど上昇させるのであれば、音声入力ソフトや翻訳マクロ等既存のテクノロジーを使用しても実現できる。その方が既存のソフトを使用できる。

3.2. HT断念理由2:データのセキュリティ問題

簡単に商用利用できるNMTでは、他社のサーバにデータが転送されてしまう。弊社の場合、知的財産権を有するテキスト、例えば、特許明細書を翻訳することが主業務であるため、データのセキュリティには万全を期す必要がある。第三者のサーバにデータが転送されてしまうことは問題がある。解決策として、自社サーバを構築するという手がある。しかし、そもそも大規模なインフラを導入しなければサービスを提供できないのであれば、最初から音声入力ソフト等のような既存のソフトウェアを使った方が導入コストは少なく済むし、導入スピードも速い。

3.3. HT断念理由3:採算に合わない

HTを人間翻訳から差別化して販売を強化するために、クライアントに対して、人間翻訳よりも低価格を提案する予定であった。しかし、PEする量が多いため関与する人件費を思ったよりも削減できない。そのため、低価格では採算が合わない。それでは、値下げを行わないという選択肢もあるが、それでは人間翻訳と差別化ができない。クライアントの声を聞いたところ、機械翻訳という近未来的なテクノロジーに魅力を感じているわけではなく、ただ単に価格が安いことが魅力であるようだ。低価格サービスであるHTを販売するよりドイツ語等、高単価の他のサービスの販売を強化した方がビジネスとして意味があると感じた。

3.4. HT断念理由4:AMTの商用利用の可能性

以前から関心を抱いていたAMTだが、そろそろCATツールに搭載したAMTを商用利用できるめどがたった。このシステムでは、大規模なサーバ構築の必要もなく、サブスクリプションを支払って即時利用でき、生産性の向上も20%以上期待できそうなため、HTからAMTに移行するべきだと判断した。

4 HTの改良案(PEMTPE)

現実的ではないかもしれないが、HTの改良案を提示する。PEの量を減らすために、プリエディット(PRE)を行うことである。PREは、すでに複数の著者がその効果を立証している[4]。PREを行って、原文から重文、複文を単文に書き換える。こうすればNMTの出力結果は飛躍的に向上すると思われる。つまりPRE後、NMTを利用する。その後、PEを行う(Post-edit/Machine Translation/Post-edit (PEMTPE))。この方法を活用する場合、原文執筆者は、可能な限りPREと物理的にも精神的にも近い方が望ましい。なぜならPREの作業を極力減らさないと作業が増えるだけだからである。原稿執筆者は、PREの作業内容を良く理解する必要がある。理想的には、PRE自体が不要になることである。したがって、原文執筆者は、最初から単文で原稿を書くことに努めるべきである。どちらにしてもPREという工程が1つ増えるため、生産効率は下がると思われる。これでは、そこまでしてHTを維持する理由は見つからない。価格は、人間翻訳と変わらないか、又はそれ以上になると思われる。なお、この方法では、人間対機械の仕事量の比率は、人間:機械=50:50程度になると思われるため、人間と機械との協働という意味では、理想に近づくかもしれない。欠点として、PEMTPEを行えば、NMTの出力精度は上がると思われる。ただしPREという工程が増えたため、翻訳時間は以前と同じか、又はそれ以上必要である可能性がある。ただし、下訳をNMTが行っているのであれば、人件費の削減にはつながるだろう。つまり、価格を下げることは可能かもしれない。ただし特許明細書等知的財産権を有するテキストを翻訳側が一方的にPRE(書き換え)をする場合は、細心の注意を払う必要がある。

5 HTの改良案(AMTの活用)

HTに代わる手段として、AMTを検討する。AMTは、筆者のメールアドレスを調べたところ、少なくとも2015年11月には開発企業に対して問い合わせのメールを送信していることから、この時期にはすでにCATツールに搭載されているようである。現在まで導入できなかった理由は、AMTでは、日本語の取り扱いがなかったためである。現在(予稿執筆時点で)、限定的ではあるが、日本語を介する翻訳もAMTが利用可能になった。AMTが搭載されたツールを利用するメリットは、大規模なシステム導入の必要がない点、サブスクリプションさえ支払えば、専用サーバを即構築でき、AMTが使える点、NMTと違って用語の制御(学習)が可能である点、PEの量が減ると思われる点である。現在(予稿執筆時点で)、AMTが搭載されたCATツールは、数社から販売されている。そのうちの1つの導入を、現在、検証中である。欠点としては、現時点では、エンタープライズ版のAMTが非常に高額なことである。

6 AMTの検証

価格の問題はあるが、とにかくAMTを検証してみた感想を述べる。2017年12月に利用が可能になり、現在(予稿執筆時点)まで1ヶ月弱しかなかった。したがって、検証すべき点は多い。AMTでは、エディット(編集)作業ではなくあくまでも翻訳を行う。NMTによる翻訳の提案と、タームベース(以下、TB)、トランスレーションメモリ(以下、TM)のデータと、学習機能とに助けられながら、人間がイニシアチブを取って翻訳を行うという印象が強い。NMTにより提案された用語を機械に学習させながら1セグメントずつ翻訳を進める。こうすることによって用語のバラツキを抑えることができる。ただし、翻訳速度が格段と早くなったような印象はない。むしろレスポンスは遅い。用語を学習させ、フレーズを記憶させれば、あたかもTBやTMを活用するかのごとく翻訳作業を行うことが可能である。なお、AMTでは、学習機能があるため、クライアント等に提出する目的以外では、TBやTMの使用は必須ではない。価格の問題さえ解決できればHTに対する解決策として活用可能だと思われる。

【表1】各種翻訳スタイルの速度及び料金

| 翻訳スタイル | 翻訳速度 | 翻訳料金 | 利点 | 課題 |
|---------|------|------|---------|---------------|
| 人間による翻訳 | 100 | 100 | セキュリティ | 速度、人件費、誤訳、抜け等 |
| MTPE | 120 | 80 | 人件費削減 | PEが多い |
| PEMTPE | 80 | 90 | PE作業減 | 工程増 |
| AMT | 130 | 110 | スピードアップ | 導入コスト |

* 翻訳速度、翻訳料金は、人間による翻訳を基準(100)とする。

* 翻訳テキストは、特許明細書のような複雑な構文のテキストを前提とする。

* 翻訳環境として、トランスレーション・メモリ、タームベースを使用する。

* 上記数値は、株式会社MK翻訳事務所の社内検証に基づく概算である。

7 今後の人間と機械との協働

最近の傾向として、CATツールに代わって翻訳プラットフォーム(以下、TP)が台頭している。ベンチャーキャピタルもこれらのTP開発会社に対して巨額な投資を行っている。[5]。TPには機械翻訳や人工知能が搭載されている。このTPには、全世界から数千人の翻訳者が登録することができる。全世界の翻訳者や校正者が24時間、365日作業をオンラインで行う。TPとこれまでのCATツールとの違

いは、前者は、エンドユーザであるメーカー向けに開発販売されるのに対して、後者は、翻訳会社や翻訳者向けに開発販売されてきたことである。翻訳業界というのは世界的に規模は小さい。矢野経済研究所の2017年の調査によれば、2017年度の翻訳業界の売上規模は、約2600億円である。[6]。これ程の小さな業界であったので、ベンチャーキャピタルも投資の対象にはしてこなかった。これに挑んだのがTPを製造開発するスタートアップ企業である。これらのスタートアップ企業がクライアントとして対象にしているのは、例えば、1社だけで年間売上が数兆円ある自動車メーカーである[7]。これらの市場であれば、ベンチャーキャピタルも投資対象と捉える。今後、人間と機械の共同作業は、単なるCATツールを介してではなく、翻訳プラットフォームを介して、大量のテキストを機械と協働して大勢の人間が短時間で翻訳する形式を取ると思われる。

参考文献

- [1] MK Translation Firm, (2017), News on “Hybrid Translation” (press release), 15 January
- [2] Masa Kajiki (2017), Neural Machine Translation and Patent Translation from LPS Perspective, The 7th Workshop on Patent and Scientific Literature Translation (PSLT 2017), 22 September
- [3] Masa Kajiki (2017), ニューラル機械翻訳と特許翻訳, JTF Journal #288, 4 March
- [4] Human Science (2017), 【ノウハウ】機械翻訳しやすい原文に！プリエディットによる品質向上, 機械翻訳ブログ, 26 December
- [5] Lance Ng (2017), 2017 in Language Industry Startup Funding, Slator.com, <https://slator.com/ma-and-funding/2017-language-industry-startup-funding/>, 11 January
- [6] 株式会社矢野経済研究所（2017）、（プレスリリース）語学ビジネス市場に関する調査を実施（2017）、<https://www.yano.co.jp/press/download.php/001720>、4 August
- [7] トヨタ自動車株式会社（2017）、業績ハイライト、<http://www.toyota.co.jp/jpn/investors/financial/high-light.html>