

## 『ニューラル機械翻訳と特許ライティングマニュアルを用いた日本語原稿のプリエディット』

発表者：NPO 法人日本知的財産翻訳協会（NIPTA）理事長 奥山尚一

ニューラルネットの活用により機械翻訳の精度は大きく向上しました。それをユーザーの立場から検証・検討するために、NIPTA は特許機械翻訳研究会を立ち上げました。

一年目は直球勝負で、機械翻訳の実情と向き合いました。そのとき、これだけ機械翻訳の精度が高まってきたなら、産業日本語もその実力を発揮できるようになってきたのではないかと直感しました。

そこで、二年目の課題として、定点観測（同じブランドのアルゴリズムを使って、性能の変化を計測する）のほかに、産業日本語を使ってプリエディットしてみて、それが有効であるか、翻訳効率が全体として上がるのかを試してみました。直接のかかわりを持ったことはありませんでしたが、以前より産業日本語研究会の活動には関心を持っており、第1版を大きく改良した「特許ライティングマニュアル（第2版）」を利用すると面白い結果が出るのではないかと考えました。

その結果を報告します。

本報告は2部からなります。第1部のとりまとめは、株式会社翻訳センターの渡部孝明氏にお願いしました。第2部のとりまとめは、ブログ「みんなのワードマクロ」管理人の新田順也氏と（一社）日本特許情報機構（JAPIO）の知財 AI センター主幹の清藤弘晃氏にお願いしました。

第1部として NMT の定点観測の結果を説明します。つまり、特許翻訳分野における現在の翻訳エンジン性能を1年前の性能と比較してみて、変化を把握しました。

第2部は、機械翻訳出力結果の利用方法の提案として、ポストエディットとプリエディットの有効性を確認しました。

新田氏には、翻訳者の視点から検討していただき、清藤氏には、明細書作成者の視点から、特許ライティングマニュアル（第2版）の活用を踏まえて、検討していただきました。

検討の結果を振り返ると、最近もトランスフォーマーといった NMT の技術革新があったものの、日本語と英語の間には、越えがたいギャップがあって、定点観測の結果を改良された部分と改悪されてしまった部分とが混在する状況です。

日英翻訳を縦のものを横にするとよくいいますが、それは韜晦であって、主語や目的語の語順が、動詞の位置が違うので相対的にまったく反転する場合があるなど、なかなか機械翻訳だけでは処理できないものがあります。それは、純粹に言語構造の違いなのか、ある対象をどうとらえるかという文化的な違いなのか、日本語と英語などの間ではヨーロッパ言語の同士の間にあるのとは異なる難しい課題が横たわっています。

日本語は類縁言語がありません。1万3千年ほども続いた縄文時代に日本語は完全に独自の発展をとげたのでしょう。それを基礎としている文化と精神世界は重要というか、我々の宝です。世界中の言語を見ても類のない、完璧な表音文字である平仮名とカタカナ（最近の悲しい事件が物語るように5歳児でも喋ることをそのまま文字にできる）、2000以上の漢字、ローマ字まで駆使する世界的に比肩するものがまったくない圧倒的な優位性を持つ日本語の魅力を維持しつつ、世界に発信できる普遍性のある日本語というのも考えなければいけないときに来ていると思います。

ある意味、江戸後期から明治にかけて、西洋の言葉と概念を須らく日本語化し、言文一致の日本語を構想して実現した人たちに近い岐路に我々は立っているのではないのでしょうか。

面白い時代になったと思いますが、ここで間違ふと今後何十年にもわたって取り返しのつかない禍根を後世に残してしまうかもしれません。真剣に取り組まなければいけないと思います。