別添３

令和元年度厚生労働行政推進調査事業費補助金（厚生労働科学特別研究事業）
（総括）研究報告書

# 医薬品評価情報作成システムの有用性評価研究

研究代表者　鹿野　真弓　東京理科大学薬学部薬学科・教授

研究要旨

新医薬品開発のグローバル化による規制当局間の国際連携が進む中、日本の医薬品評価結果の情報を迅速に世界に発信する必要がある。本研究においては、PMDA の新医薬品の審査報告書の英語版作成における AI 自動翻訳ツールによる翻訳の有用性を確認し、その活用にむけた検討を行った。

審査報告書のAI翻訳ツールによる翻訳産物について、専門性の観点及び英語の正確性の観点から評価した。その結果、翻訳不備の要因がいくつか特定され、原文作成の段階で翻訳不備を防ぐ方策を取るとともに、定型記載や専門用語の学習・カスタマイズによるAI自動翻訳ツールの性能向上や事後的な人手によるチェックが必要と考えられた。しかし、AI自動翻訳特有の翻訳不備の一部は特段のパターンによらず突発的に発生することから、そのチェックを行う際には審査報告書全体について原文と英文の比較による確認作業が必要となり、人的リソースの軽減効果は限定的で、審査報告書英訳の公表版作成に活用することは現時点では現実的ではないと考えらえる。一方で、各分野の専門性を有する者であれば、翻訳不備があっても論旨を把握することは可能と考えられ、例えば規制当局間で早急な情報提供を求められる場合など、必要に応じてAI自動翻訳ツールで英訳を行い、一定の翻訳不備が含まれることを説明した上で最低限の確認・修正のみを行った審査報告書のAI自動翻訳英文版を情報提供することが有用なケースはあると考えられる。AI自動翻訳ツールは、学習やカスタマイズ等によるレベルアップのみならず、翻訳エンジン自体の性能向上も期待されるものであることから、今後、より質の高い自動翻訳ツールが使用可能となることが望まれる。

研究分担者氏名
永井　尚美　武蔵野大学薬学部・教授
山田　博章　横浜薬科大学臨床薬学科・教授

## Ａ．研究目的

医薬品・医療機器等の開発・流通のグローバリゼーションや製品の多様化を受けて、関連規制制度は複雑かつ高度化しており、国際的な協力の重要性が世界の規制当局間の共通認識となりつつある。こうした中、世界保健機関（WHO）や、各国/地域の医薬品規制当局のリーダーで構成される薬事規制当局国際連携組織(ＩＣＭＲＡ）では、情報共有を視野に入れて各規制当局のリソースをより有効に活用する取組みが進められている。

さらに本邦においては、日本で承認された優れた医薬品・医療機器がアジア諸国で受け入れられ、迅速に患者に提供されることを目指しており、我が国の医薬品等の評価結果や安全性情報の英語等での発信を積極的に行っていく必要がある。規制当局が専門的内容を含む評価結果の英語版を正確かつ迅速に作成する方策として、AI を活用した自動翻訳ツールの導入が考えられる。

AI自動翻訳ツールの性能は、近年、著しく進歩しており、本研究では、代表的なAI自動翻訳ツールを用いて自動翻訳の正確性等の特徴や課題を評価することでその有用性を確認し、規制当局の医薬品評価情報作成のパフォーマンス向上に資する活用法に関する研究を行うことを目的とする

## Ｂ．研究方法

### １）自動翻訳ツールの選定

医学･薬学の専門用語にも十分に対応できるAI等を活用した自動翻訳ツールを利用するため、本学HP上で研究への協力を広く呼びかけた。応募された４社のうち、本研究の主旨や本研究事業の仕組み

を踏まえて協力を了承された以下のツールを用いることとした。

- T-4OO（Translation For Onsha Only：株式会社ロゼッタ）
- COTOHA（株式会社みらい翻訳/NTTコミュニケーションズ/株式会社翻訳センター）。

これらの自動翻訳ツールについては、いずれも独立行政法人医薬品医療機器総合機構（以下、PMDA）が、その英訳とともに公表している新医薬品の承認審査報告書や医薬品規制調和国際会議（以下、ICH）各種ガイドライン等の薬事規制関連文書を用いて学習を行ったものを利用した。
なお、本研究においては２つの自動翻訳ツール商品の性能比較が目的ではないため、自動翻訳ツール名を【A】あるいは【B】と記載する。

２）評価対象文書の選定

独立行政法人医薬品医療機器総合機構（以下、PMDA）が公表している新医薬品の審査報告書を対象に用いることとした。平成30年以降に承認された新有効成分含有医薬品のうち、効能・効果に偏りが無いよう以下の４つの審査報告書を選定した。

- イミフィンジ点滴静注 120mg、同点滴静注 500mg（以下、イミフィンジ）抗悪性腫瘍薬
- スキリージ皮下注 75mg シリンジ 0.83mL（以下、スキリージ）：炎症性疾患治療薬
- ヘムライブラ皮下注 30mg、同 60mg、同 90mg、同 105mg、同 150mg（以下、ヘムライブラ）：血液疾患治療薬
- レルミナ錠 40mg（以下、レルミナ）：婦人科疾患治療薬

詳細は、別添４の研究代表者 鹿野の報告書を参照されたい。

３）評価方法

上記で選択した４つの審査報告書について、自動翻訳ツール【A】及び【B】を用いて翻訳した翻訳産物のうち、品質分野、毒性分野、臨床分野の評価、及び英語の専門性の観点からの評価は、研究代表者の鹿野が研究協力者の参加も得て実施した。ADME分野については研究分担者の永井が、薬理分野については研究分担者の山田が評価を行った。
評価スコア含め評価方法の詳細については、別添４の研究代表者、各研究分担者の報告書を参照されたい。

（倫理面への配慮）

自動翻訳ツールの採用の経緯ならびに評価方法及び結果を本報告書で公表し透明性を確保することで利益相反が無いことを示す。
また本研究は、再生医療等の安全性の確保等に関する法律、人を対象とする医学系研究に関する倫理指針、ヒトゲノム・遺伝子解析研究に関する倫理指針、動物実験等の実施に関する基本指針、省令GCP、臨床研究法の対象に該当しない。

Ｃ．研究結果

本項においては、各専門分野に特徴的に見られた翻訳不備を中心に記載する。専門分野に関係なく確認された翻訳不備については「D.考察　２）全体的に見られた翻訳不備の整理」の項でまとめて記載する。

１）品質分野の英訳文書の評価

4つの品目の審査報告書の翻訳産物のうち品質分野について、修正しなくても意味が通ると評価されたスコア２及び３の文の割合は、翻訳ツール【A】の産物の平均が89.2%、【B】の産物の平均が86.8%と、概ね高い質の翻訳産物が得られていた。

翻訳ツール【A】の産物は、マスキング箇所の前後で構文が不正確になる、あるいは訳漏れが発生するケースが多いことが確認された。ICH品質ガイドラインで規定されている専門用語は概ね正しく翻訳されていた。翻訳ツール【B】の翻訳産物では、原文の一部がマスキングされていると一文全体をマスキングして翻訳する箇所が散見された。また、ICH品質ガイドラインで規定されている専門用語や公定書名等を正しく翻訳されない場合が多く見られた。

２）薬理分野の英訳文書の評価

審査報告書の非臨床薬理分野を構成する全センテンスの全角換算文字数は、イミフィンジ、スキリージ及びヘムライブラでは、25 文字超、50 文字以下のセンテンスがピークで約 35%を占めていた。レルミナは、25 文字以下のセンテンスが約 30%で最も多く、150 文字超の長文も含んでいた。

翻訳英文の評価スコア分布については、全体的に、修正しなくとも情報として意味が通じるスコア２以上の英文が約 80%を占めていた。イミフィンジ、スキリージ及びヘムライブラでは、【B】に比較して【A】の方が、スコア３の割合が高くスコア０の割合が低かった。レルミナでは逆に【B】の方がスコア３の割合が高くスコア０の割合が低かった。

３）ADME分野の英訳文書の評価

審査報告書の解釈及び翻訳版での文意伝達に関して、専門的立場からは修正しなくても意味が通ると判断されたスコア２及び３の箇所の割合は、67～94%の範囲であり、全188箇所における平均は、【A】及び【B】の翻訳産物で83%及び85%であった。同一品目においては、翻訳ツール間の顕著な差は認められなかった。品目によってスコア２及び3の割合に差が生じた背景として、ADME評価情報として選定した非臨床薬物動態、臨床薬理及び審査内容が品目ごとに異なるほか、疾患領域や品目特性が影響すると考えられた。

翻訳不備（スコア0及び1）と判断した事由のうち、ADME分野に特徴的なものとしては、一部の専門用語の選択のほか、健康被験者とすべき箇所を患者、表中の項目において、非臨床薬物動態データであるにも関わらずヒトのデータとして取り扱われている箇所があった。

４）毒性分野の英訳文書の評価

4つの品目の審査報告書の翻訳産物のうち、毒性分野について、修正しなくても意味が通ると評価されたスコア２及び３の割合は、スキリージのみ翻訳ツール【A】【B】の翻訳産物とも50%程度、ヘムライブラはいずれも85%程度、それ以外の２品目は、いずれの翻訳産物も約75%前後であった。

スキリージの【A】【B】のいずれの翻訳産物とも評価が低かった原因は、いずれも原文にない表現や用語が自動的に追加された箇所が複数見られたことにあった。

また翻訳ツールにかかわらず「不適切用語」が散見され、毒性分野で慣用される表現が正しく翻訳されていないケースが多かった。

５）臨床分野の英訳文書の評価

4つの品目の審査報告書の臨床分野の評価は、品目ごとに異なる評価者が担当したため評価の視点が異なる可能性が考えられ、品目間の差異を比較することは適切ではないが、同じ品目の【A】と【B】の翻訳産物の比較は可能と考えた。修正しなくても意味が通ると評価されたスコア２及び３の文の割合いは、イミフィンジ、スキリージ、ヘムライブラは【A】の産物の方が【B】に比較して高い傾向が見られた。一方で、レルミナは【A】が【B】より明らかに低い数値であった。

レルミナについては、原文が長文で構成が複雑あるいは不明瞭なものが多く、【A】と【B】の翻訳産物とも重大な訳漏れや英訳の構文の乱れによる誤訳が多数発生していたが、特に翻訳ツール【A】は、脚注番号が挿入されている場合に不適切な個所で文を分割して翻訳するため、重大な誤訳につながった箇所が多いことが確認された。

臨床分野に特徴的と思われる翻訳不備はほとんど見られなかったが、他の診療科の薬物に比較して自動翻訳ツールの学習機会が少ないと推察される婦人科領域の専門用語の誤訳が散見された。

６）英語の正確さの視点での英訳文書の評価

英語の専門性を有する評価者により、英語の正確性の視点から評価が行われた。

特徴的な翻訳不備としては、審査報告書の冒頭にある「審査報告書」の定型記載の翻訳不備が多く、【A】の翻訳産物では販売名のスペルミス、複数の定型的項目名の誤訳の他、糖鎖構造の表記の訳漏れが見られた。【B】の翻訳産物でも販売名のスペルミス、複数の定型的項目名の誤訳が見られた他、通知引用の記載、審査部名の訳も不適切であった。

記号や数字が多数含まれる糖鎖構造の説明では、訳漏れが複数発生していた。また、目次の記載は翻訳されず日本語のままであった。

D．考察

1）各分野等に特徴的な翻訳不備の整理

定型記載について

審査報告書冒頭の「審査報告書」には定型的な記載が多いが、翻訳ツール【A】【B】ともに定型的な項目名の誤訳が複数見られ、これらは目次が日本語のままの不備とともに学習あるいはカスタマイズで対応可能と考えられる。一方で、販売名や一般名のスペルミス、有効成分の構造の記載に含まれる記号や成分名称等の誤訳については個別に対応する必要があると考えられた。

品質分野について

品質分野はマスキングされる件数が多くその範も長いため、翻訳ツールにかかわらずマスキング箇所の前後での英訳文の構文の乱れや訳漏れが多く発生する。品質専門用語や表現に関しては、特に翻訳ツール【B】はICH品質ガイドラインで規定されている専門用語を適切に学習させることが必要と考えられた。

品質分野の審査報告書は比較的シンプルな文が多いため、マスキング箇所の前後の重点的なチェック及び専門用語・表現の学習が適切になされれば、重大な誤解を生じるリスクはかなり低減できると考える

薬理分野について

薬理分野に特徴的と言える翻訳不備は特段見られなかった。

ADME分野について

健康被験者とすべき箇所が患者として取り扱われている箇所については、文中に引用されている試験名と対応する訳出であると考えられた。また、ADMEは生体内挙動に関わる学問領域であり、審査報告書中では、唯一非臨床と臨床の両領域を含む。ADME特性の記述に加えて、他の非臨床又は臨床分野に言及する場合もあることが影響する可能性が考えられた。

毒性分野について

スキリージにおいて、項目タイトル（翻訳ツール【A】）や表タイトル（翻訳ツール【B】）に関係ない用語が自動的に追加されていたが、これは翻訳ツールが過去の審査報告書の翻訳の際に学習した内容がスキリージの内容と類似していたため、自動的に学習した内容を反映した可能性が考えられる。このようなAI自動翻訳ツールの特徴的機能に由来すると考えられる誤訳を防ぐ方策については、自動翻訳ツールの開発販売会社との相談・検討が必要であろう。

臨床分野について

翻訳ツールの学習機会が少ないと推察される診療科の専門用語に不備が見られた。対応する診療科のガイドラインを学習させる等の対応が考えられる。

なお今回、自動翻訳ツール【A】【B】いずれも薬事に対応したバージョンを使用したが、それぞれ品質分野、毒性分野等の各分野に対応したバージョンも用意されているため、審査報告書の各分野ごとに対応するバージョンの自動翻訳ツールを使用すれば、専門用語や分野に特徴的な表現に適切に対応された翻訳産物が得られる可能性が期待される。

2）全体的に見られた翻訳不備の整理

① 原文の構成に由来する翻訳不備

複雑で長い文は誤訳を生じやすく、特に自動翻訳ツール【A】は数字、記号、カッコ、脚注番号等を含むと不適切な個所で分割して訳す例が散見された。

対処法としては、自動翻訳前の作業（Pre-edit）として、長文は避けて複数の文を接続詞でつなぐ記載を心がけ、主要なフレーズ・文に主語を明記する等が、自動翻訳後の作業（Post-edit）として数字、記号等の箇所のチェックが考えられる。

② 記号、数字、カッコ等に由来する翻訳不備

数字、記号、カッコの多用、多重カッコは訳漏れや数字の入れ替わりが発生しやすい。長文に

含まれる脚注番号が別の文に移動する等が発生する。

対処法としては、自動翻訳前の作業（Pre-edit）としてカッコの整理や丸数字の排除の他、数字、記号、脚注番号について人手でチェックする自動翻訳後の作業（Post-edit）が考えられる。

③ マスキングに由来する翻訳不備

【A】ではマスキングの前後で構文の誤解釈が発生しやく、【B】では文の1/3程度がマスキングされると全文をマスキングとして翻訳される場合が多い。

対処法としては、マスキングの前後の訳を重点的に人手でチェックする自動翻訳後の作業（Post-edit）が考えられる。

④ 定型記載に関連する翻訳不備

「本薬」に該当する訳は【A】で「this drug」「the this drug」、【B】では「drug」と訳されわかりにくい。自動翻訳前の作業（Pre-edit）として原文で成分記号等に置き換えておく等の対応が考えられる。

「審査報告書」、審査報告(1)(2)の冒頭部分等の定型記載の誤訳、最後に添付された略語表も記載の工夫が必要。

対処法としては、自動翻訳ツールに学習させる、カスタマイズする等の自動翻訳ツールの機能改善・工夫が考えられる。

③ 訳語の不安定さ

一般的な用語について異なる翻訳がなされる場合が散見され、異なる意味に捉えられて誤解を生じる危険性が否定できない。

訳ブレの発生に一定の規則性が見られず、対処法については自動翻訳ツールの開発販売会社との相談・検討が必要かもしれない。

## E．結論

自動翻訳ツールによる翻訳は、英語の正確性の面で課題は見られたものの、各分野専門性を有する者であれば、自動翻訳ツールの翻訳産物の内容を一定程度理解することは可能と考えられた。翻訳産物の質をさらに向上させるためには、(1) 自動翻訳前の作業（Pre-edit）を行うことで可能な範囲で誤訳を防ぐこと、(2) 適切に学習させた自動翻訳ツールを使用すること、(3) 自動翻訳後の作業（Post-edit）において、翻訳産物について、マスキング箇所や、カッコ、数字、記号、脚注を含む長文、図の画像など翻訳不備が生じやすい箇所を重点的にチェックすること、等が考えられる。しかしながら、無関係の語句の挿入や原文誤解釈等のAI自動翻訳特有の翻訳不備の人手によるチェックには追加の人的リソースの確保や校正に長時間が必要となる可能性が高く、現時点では効率的とは言い難い。一方で、その専門分野の知識を有する者であれば、全体の概要を把握することは十分可能と考えられる。

国際連携が進む中、一定の翻訳不備を完全には排除できなくても審査報告書英文版を迅速に情報提供することが必要な場合も考えられる。このような場合において、AI自動翻訳ツールで英訳を行い、一定の翻訳不備が含まれることを説明した上で審査報告書英文版を提供することが有用な場合もあると考えられる。AI自動翻訳ツールは、将来的な性能向上が期待されるものであることから、今後、審査報告書の公開版作成への適用が可能となる自動翻訳ツールが使用可能となることが望まれる。

## F．健康危険情報

特になし。

## G．研究発表

1. 論文発表

本研究に関連した論文発表はない。

2. 学会発表

本研究に関連した学会発表はない。

## H．知的財産権の出願・登録状況
（予定を含む。）

1. 特許取得

なし

2. 実用新案登録

なし

3.その他

なし