

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）
（分担）研究報告書

品質・毒性・臨床分野等検討

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、品質、毒性、臨床の各分野の翻訳の質を、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れる。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。

A．研究目的

医薬品・医療機器等の開発・流通のグローバル化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-400 (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、その英訳とともに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)
(https://www.pmda.go.jp/drugs/2018/P20180727001/670227000_23000AMX00485_A100_1.pdf) 抗悪性腫瘍剤
- スキリージ皮下注 75mg シリンジ 0.83mL (以下、スキリージ)
(https://www.pmda.go.jp/drugs/2019/P20190329001/112130000_23100AMX00299_A100_1.pdf) 炎症性疾患治療薬
- ヘムライブラ皮下注 30mg、同皮下注 60mg、同皮下注 90mg、同皮下注 105mg、同皮下 150mg (以下、ヘムライブラ)
(https://www.pmda.go.jp/drugs/2018/P20180329003/450045000_23000AMX00450_A100_1.pdf) 血液疾患治療薬
- レルミナ錠 40mg (以下、レルミナ)
(https://www.pmda.go.jp/drugs/2019/P20190124001/400256000_23100AMX00010

_A100_1.pdf) 婦人科疾患治療薬

当初、1次評価としてPMDAのHPで公表されている審査報告書pdf版についてAI自動翻訳ツールを用いて英訳を作成し、詳細な評価に必要な評価項目の探索を試みたが、pdf文書の翻訳産物は著しく精度が低く、適切な評価のためにはwordファイル等から翻訳する必要があることが確認された。そこでPMDAに協力を依頼し、各審査報告書の公表pdf版と同じ内容(pdf版の黒塗り部分はアスタリスク「*」で置き換え)の非公表word文書を自動翻訳ツール【A】あるいは【B】で英訳した文書、並びにそれぞれの英訳文書を同じ翻訳ツールで日本語に逆翻訳した文書を提供頂いた。PMDAから提供された英訳文書を対象に、各分担研究者が評価を行うこととされた。なお、逆翻訳文書は必要に応じて参考とする位置付けであり、本研究における評価対象ではない。

3) 評価方法

日本製薬工業協会の協力を得て、AI自動翻訳ツールを利用している製薬企業各社の担当者から、一般に自動翻訳ツールで良く見られる誤訳のパターン、各社におけるAI自動翻訳ツールの性能評価方法、また翻訳物の利用状況等に関してヒアリングを行った。

その結果も踏まえて、以下の手順で評価を行った。

品質、臨床のそれぞれの分野あるいは全体について、内容的に重要な記載または誤訳が生じやすいと考えられる文を中心に、1~数か所から合計でおおむね30-50文程度を評価対象文として抽出した。誤訳が生じやすいと考えられる文とは、構文が複雑な長文、マスキング箇所、図表、数値、記号等を含む文である。毒性分野については「5. 毒性試験に関する資料及び機構における審査の概略」の項全体を評価対象とした。自動翻訳ツール【A】あるいは【B】を用いて翻訳された文書の評価対象箇所について、1文ごとの評価を行った。

評価は、研究協力者の参加も得て以下のように行った。品質分野、毒性分野については、それぞれの専門性を有する評価者1名が、4つの品目の審査報告書の【A】及び【B】による翻訳産物を評価した。臨床分野については、専門知識を有する評価者4名が、それぞれ1品目ずつの審査報告書の【A】及び【B】による翻訳産物の評価を行った。また、英語の専門性を有する評価者1名が、英語の正確さの観点から4つの品目の審査報告書の任意の箇所を対象に【A】及び【B】による翻訳産物の評価を行った。

評価対象の英訳1文ごとに以下の4段階でスコア化を行った。

スコア3：このままで問題ない

スコア2：完全な英文にするには一部修正が必要だが修正しなくとも意味は通じる

スコア1：一部修正しないと意味が通じない

スコア0：大きな問題があり大幅な修正が必要

スコア0~2で見られた翻訳不備の内容について、「訳漏れ」、「原文の誤解釈」、「訳文の文法/構文」、「不適切用語」、「記号・数字」、「その他」の6種類のいずれに該当するかを記録した。他に、評価者が気づいた特徴等に関して自由記載することとした。

(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

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品質分野について自

動翻訳ツール【A】及び【B】により翻訳した産物を評価し、各スコアの個数、評価概要、スコア0あるいは1と評価したおもな理由を表1~表4に示す。また、各スコアの個数をグラフで表したのが、それぞれ図1~図4である。

各翻訳産物について、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の文の数及びその割合を表5にまとめた。最低で80%、最高で94%の文が2または3と評価され、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。

翻訳不備の各内容のセンテンス数を表6に示した。翻訳ツール【A】の産物で多く見られた不備は、「不適切用語」「訳文の文法/構文」「訳漏れ」「原文の誤解釈」であった。翻訳ツール【B】の産物で多く見られた不備は、「不適切用語」が多く、他に「訳漏れ」「原文誤解釈」が散見された。

翻訳ツール【A】あるいは【B】の不備について、表1~表4の審査報告書ごとの評価を確認すると、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。また、品質分野で汎用される表現の翻訳が不正確である事例も散見された(例:変化体をchange body、恒常性をhomeostasisと訳。正しくはvariant、consistency。)が、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、品質分野で汎用される表現のみならず、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直訳し、正しい用語が使われていない場合が多く見られた。マスキング部分が多い文は文全体をマスキングした形で翻訳する「訳漏れ」の事例は、翻訳ツール【A】の場合と同様に確認された。

なお、数値で示すことは難しいが、翻訳ツール【B】の翻訳産物は英語として読みやすい印象であった。

2) 毒性分野の英訳文書の評価

4つの品目の審査報告書の毒性分野について自動翻訳ツール【A】及び【B】により翻訳した産物を評価し、各スコアの個数、図表の情報、スコア0

あるいは1と評価したおもな理由を表7～表10に示す。また、各スコアの個数をグラフで表したのが図5～図8である。

各翻訳産物について、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のセンテンス数及びその割合を表11にまとめた。スキリージのみ【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。平均で翻訳ツール【A】【B】の産物とも75%前後であった。

以上のように品目間のばらつきが大きいことから、個別品目の評価状況を表7～表10に基づいて確認したところ、スキリージでは翻訳ツール【A】【B】いずれの翻訳産物と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が低評価の原因であった。ヘムライブラでは原文が難解なための誤解釈や不適切用語は見られたものの件数は少なかった。

翻訳不備の内容ごとのセンテンス数を表12に示した。ヘムライブラを除く3品目で「原文の誤解釈」の件数が多く、原文が難解な長文であることが原因と考えられた。なおイミフィンジの【A】の翻訳産物で訳漏れが多いが、これも原文に難解な長文が多かったためと推察された。他には、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。

3) 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について自動翻訳ツール【A】及び【B】により翻訳した産物を評価し、各スコアの個数、評価概要、スコア0あるいは1と評価したおもな理由を表13～表16に示す。また、各スコアの個数をグラフに示したのが図9～図12である。

各翻訳産物について、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のセンテンス数及びその割合を表17にまとめた。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。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2あるいは3と評価されたセンテンスの割合が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。

そこで、レルミナの評価状況を表16に基づいて確認したところ、原文が長文で構成が複雑あるいは不明瞭なものが多く、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。特に翻訳ツール【A】は、引用先の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。【A】では他に、「the this drug group」との訳が2か所あった。この誤訳は他の品目、他の分野の翻訳でも頻出している。【B】の産物で上記以外の誤訳としては、本薬を定冠詞なしの「drug」と訳す、表内の「135例^a」を「^a of 135 subjects」と訳すなどが見られた。

他の品目の評価において、0または1とされたおもな不備は以下の通りである。

イミフィンジでは、【A】の翻訳産物で原文に無い数字の挿入が2か所で確認された他、「the this drug group」に加えて「the investigational product this drug group」、項目立ての、を(1)、(b)と訳すことによる項立ての混乱、「それぞれ()55(- 3¹⁴) ~ 406)」が正確に翻訳されず情報が把握できない等、カッコ、記号、数値が混在する表記の訳が不正確な場合などが見られた。【B】の翻訳産物では、定冠詞(the)や一部フレーズの細かい訳漏れが意味に影響する箇所が多数見られた。他に、マスキングや長文記載が原因と考えられる訳の乱れ、図のタイトルの語順の乱れやpdf画像を貼り付けたと思われる図内の日本語が一部しか訳されていない等が確認された。

スキリージでは、翻訳ツール【A】の産物で、関係ない薬剤名が挿入されていた。文中に脚注番号や図表番号があるとその後で訳の構文の乱れが多数発生していた他、同一用語の訳のブレ、カッコ内の数値が不正確な箇所があった。不適切用語の訳も確認され、「目的」を「EYES Common」とする

突拍子もない訳も見られた。他に、原因となるパターンが不明の訳漏れも散見された。【B】の産物では脚注文頭に不要な「and」の挿入や原文が長文であることが原因と推定される不自然な訳が複数確認された。他に、表中の数字「1,971例」を「11971 subjects」とする誤訳、薬事上、重要な固有名詞「リスク管理計画」の誤訳も見られた。

ヘムライブラでは【A】の産物で無関係の薬剤名の挿入が確認された。他に、同一用語の翻訳のブレや、多重カッコを用いた症例数の説明の誤訳、並列して列挙する記載や表番号を引用した際の訳の乱れが見られた。一方で、長文を複数に分割して訳されたことで原文よりわかりやすい記載となっている箇所も複数あった。【B】の産物では、【A】と同じ箇所でも多重カッコを用いた症例数の説明の誤訳が生じていた。また、4群の観察期間の説明において1群だけ異なる説明があるにもかかわらず他群と同じ訳にされていた。他に、文頭に数字が来る濃度記載があった場合に数字のみ訳してその文節を訳さない、「7例4件」を「7 subjects」と訳、本薬を定冠詞なしの「drug」と訳等の不備が見られた。

表18では、内容への影響とは関係なく確認された不備の情報を見ることができる。レルミナ以外の3品目では【B】で不備の合計が【A】より多い傾向が見られ、特に不適切用語が【A】より多い傾向が見られた。

4) 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。

4つの品目の審査報告書について自動翻訳ツール【A】及び【B】により翻訳した産物を評価し、各スコアの個数、スコア0あるいは1と評価したおもな理由を表19～表22に示す。また、各スコアの個数をグラフに示したのが図13～図16である。

各翻訳産物について、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のセンテンス数及びその割合を表23に、英訳不備内容の件数を表24にまとめた。表23からは、各専門性を有する評価者による評価(表5、表11、表17)と比較し

て修正が必要と評価されたセンテンスの割合が明かに高いことが確認された。また表24では、各専門性を有する評価者による評価(表6、表12、表18)と比較して「その他」と分類された不備が多かった。

各個別品目の評価について、各分野の専門性を有する評価者での評価で見られなかった事項を中心に以下に記載する。

イミフィンジについては、審査報告書の非臨床薬理及び臨床の項、ならびに審査報告(2)から評価対象文が抽出された。自動翻訳ツール【A】の翻訳産物では、英語で使用されない丸数字(、等)を含む文において、単なる数字として表記され、「0.1又は1mg/kg」を「10.1 or 21 mg/kg」と誤った数値が表示されていた。また、審査報告(2)の冒頭部分の記載については、販売名のイミフィンジ「Imifinzi」を「Imifindi」と誤ったスペルで記載されていた。【B】の翻訳産物では、「本薬の用法・用量について」を「Dosage and administration of ramucirumab」と関係ない薬剤名を挿入して翻訳されていた。また、販売名のみでなく一般名「durvalumab」も「dulvalumab」と誤ったスペルで記載されていた。

スキリージについては、審査報告書の臨床薬理及び臨床の項から評価対象文が抽出された。毒性分野、臨床分野で見られた不備と同様のもの以外の特徴的不備は見られなかった。具体的には、【A】の翻訳産物で複数投与群の説明で用量と症例数の数値や群の名称の番号等、数字が多く出現する文では、数字の入れ替わりが多数生じていた。また、数値とカッコが続く文で訳漏れが発生、脚注番号の前後で文を分割したため文法的に不完全な訳となったケース、グラフ画像の日本語が翻訳されない、等が確認された。【B】の翻訳産物では、マスキングとカッコが重なる部分、カッコや数値を含む長文で構文の乱れや訳漏れが発生していた。また、「薬物動態」、「血漿中薬物濃度データ」、「皮下投与」、「抗体価」の訳が不適切であった。

ヘムライブラについては、審査報告書の冒頭にある「審査報告書」及び品質の項から評価対象文が抽出された。「審査報告書」では定型的な記載が多

いが、定型的な用語が正確に翻訳されていない箇所が多かった。【A】の翻訳産物では販売名HemlibraをHemribulaと誤ったスペルで記載されていた他、[目次]を[Eyes] Next、「本質」を「References」、「構造」を「Design」とする誤訳が見られた。また、糖鎖構造の記述では「Glc」「GlcNAc」の訳漏れが見られた。【B】の翻訳産物でも「ヘムライブラ」を「Heme Library」、「本質」を「Essential」、「構造」を「Preparation」とする誤訳が見られた他、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。

レルミナについては、毒性及び臨床の項、審査報告(2)ならびに略語表から評価対象文が抽出された。毒性分野、臨床分野で見られた不備と同様、原文が不適切な構成のため誤解釈や訳漏れが発生する等が見られた。略語表については、細かい記載整備が必要な箇所が多数確認された。

D. 考察

1) 品質分野の英訳文書の評価について

評価した医薬品のうち3品目が遺伝子組み換え医薬品、1品目が化学合成医薬品であった。一般に、遺伝子組み換え医薬品の製造方法や品質管理は化学合成医薬品のそれよりも記載が複雑になる場合が多いが、翻訳産物の質に対する影響は特段見られなかった。

品質分野の審査報告書原文の特徴としては、以下があげられる。

- 他の分野に比べると構成が複雑な長文は少なく、比較的シンプルな構文が多い。
- 医薬品の製造や品質管理は各社の知財やノウハウに関する内容を含むため、他の分野に比べてマスキングされる箇所の長さや件数が多い。
- 同じ日本語でも専門分野によって対応する英語が異なる（例：含量は臨床分野ではpotencyを用いる場合があるが品質ではcontent。）。また品質分野で汎用される表

現や言い回しがある。

これらの特徴が、まさに品質分野の自動翻訳ツールでの翻訳に大きく影響を及ぼすと考えられる。

具体的には、2点目に関して、翻訳ツールにかかわらずマスキング箇所を挟むと英訳文の構文が不正確になる場合や、前後の記載の訳漏れを招く場合が多い。翻訳産物について、マスキング箇所の前後を重点的にチェックすることで翻訳文書の質の向上を図ることが可能と考えらる。

3点目の品質専門用語や表現に関しては、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思われる。品質分野で汎用される表現も、種類はさほど多くないため、いずれの翻訳ツールについても学習・カスタマイズを進めることで、より正確な英訳文書を得ることが期待される。

すなわち、これらの点に対応できれば、1点目に記載したように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、海外規制当局等の関係者が読んだ場合に重大な誤解を生じるリスクはかなり低減できるものと考えられる。

2) 毒性分野の英訳文書の評価について

毒性分野の審査報告書ではマスキングは極めてまれであるため、その影響は考慮する必要は無い。一方で、実施される試験の種類や試験法の大枠はICHガイドラインで定められており、品目間で共通あるいは類似した試験が実施される場合が多い。今回、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。

他に翻訳に影響を及ぼす重要な因子として、原文が長文で日本語として適切な構文となっていない場合が多々見られされた。このような場合、翻訳時に原文の誤解釈や訳漏れを招き、翻訳産物の質

の低下につながる。審査報告書作成時には、例えば1文を極力2行以内とする、箇条書きを活用する、定型的な評価を行う場合には記載様式を可能な範囲で統一する等、論理的に整理された記載を意識して審査報告書を作成することが望まれる。また、データを掲載する表形式を可能な範囲で統一することも、安定した翻訳に資する可能性が考えられる。

さらに、専門用語の統一や毒性分野で汎用される表現の学習が必要と考えられる。ICHガイドラインに記載される事項よりもさらに詳細な専門用語（例：発生学、病理学等に関連した専門用語）の統一の必要性については毒性専門家の間でも認識されており、今後、各学会等で辞書作成の活動が進むことが期待される。

なお、数値化は難しいが、【A】は直訳に近い翻訳が多いため、原文が曖昧な日本語だと誤訳が生じやすい。【B】は学習を踏まえた意識が多いものの、適切でない意識となる場合も見られた。

3) 臨床分野の英訳文書の評価について

臨床分野は、品質分野ほどではないがマスキング箇所が複数あり、それが原因と考えられる誤訳が少数ながら発生していた。

品質分野、毒性分野で見られたような、分野に特徴的な用語や言い回しに関する誤訳は【A】【B】ともに少なかった。これは、臨床分野では病名、症状名等の辞書が充実していることに加え、自動翻訳ツールが臨床関連文書の翻訳に使用されるケースが他の分野の文書に比べて多く学習量も多いためではないかと推察された。

毒性分野と同様に、無関係な数字や薬品名等が挿入された例が【A】【B】でともに確認された。やはり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自動的に反映された可能性が考えられる。このような突発的な無関係の数字や単語の挿入は予測が難しい。また、定冠詞等の細かい訳漏れが内容解釈に影響するケースも散見された。「本薬」を「drug」と訳される場合については、原文の段階で成分記号に置き換えることで誤訳を防ぐことが可能と考えられる他、学習での対応も可

能かもしれない。しかし、不規則に生じる細かい訳漏れへの対応は困難と予測され、突発的な挿入と同様に、回避策については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。

他にも【A】での「the this drug group」など、出現が予測される誤訳は事後に検出して修正することが可能だが、翻訳ツールの学習で対応可能かもしれない。

また、やはり原文が長文で日本語として適切な構文となっていない場合、文の途中にカッコで参照先等を記載する場合や多重カッコを用いる場合に重大な誤訳が生じやすいことも確認された。審査報告書を作成する段階で、可能な範囲で誤訳を生じやすい記載を避けるためのルールを設けることは、論理的に整理された正確な記載を目指すことにもつながり、日本語の審査報告書もより理解しやすいものになると期待される。

なお表17に示したように、全体に【A】の方が【B】より正確に訳されており、用語の適切性や原文の構成が複雑でない場合の訳文構成の面で優れているが、原文の構成が複雑で整理が不十分な点があったレルミナの場合は【B】はより柔軟に対応可能であることが確認された。また、【B】では数字、太字、スペース、フォント等でより優れていた。

4) 英語の正確さの視点での英訳文書の評価について

英語の専門性を有する評価者による評価においては、「審査報告書」、審査報告(2)の冒頭部分等、定型的な記載についての翻訳不備が多数確認された。これらについては自動翻訳ツールの学習やカスタマイズ等によりかなりの部分に対応可能と考えられる。それ以外の箇所で確認された翻訳の不備のほとんどは他の評価者による評価でも指摘されているが、英語の専門性を有する評価者による評価スコアは、分野専門性を有する評価者の評価スコアより全体に低いものであった。分野専門性を有する評価者が評価する際には、英語の正確性に多少欠けていても当該分野の専門知識や前後の記載内容に基づいて欠けた部分を補足しながら内容を理解することが可能かという視点も加わるた

めであろう。極端に言えば、全く意味不明な一文があってもその前後を読むことで全体としてどのようなことが記載されているかを理解できれば大きな問題となる可能性は低い。

さて本研究は、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 AI 自動翻訳ツール活用法に関する研究を行うものであり、そのために AI 自動翻訳ツールによる翻訳の正確性等の特徴や課題を評価することが重要となる。そのためには、翻訳産物が日本語の原文の意味を忠実に反映し、英語として正確な記載となっているかを指標とした評価を行うことが大前提となる。

その評価を踏まえて課題を整理し、さらに対応法を検討することで翻訳の質の向上を図ることが求められている。

翻訳の質の面での不足を専門知識で補うことはもちろん重要ではあるが、翻訳の不備は内容的な重要性とは関係なく発生しうるため、重要な部分の誤訳が重なると全体の理解にも重大な影響を及ぼすことになる。従って、まずは全体として翻訳の質を向上させ、それでも一定程度発生する翻訳不備について可能な範囲で専門知識でカバーすることを考える必要がある。

5) 確認された翻訳不備の整理

以上の評価において頻出していた翻訳不備の種類及びその対処法を以下に整理した。

原文の構成に由来する翻訳不備

- 複雑で長い文は原文構成の誤解釈を招き、一部の訳漏れ、訳文の不適切な文法 / 構文を招きやすい。
- 自動翻訳ツール【A】は長文を分割して翻訳するケースが多いが、数字、記号、カッコ、脚注番号等があると不適切な個所で分割されてしまう。
- 原文の構成が不適切なく(日本語としてわかりにくい)文は翻訳不備が起こりやすい。自動翻訳ツールが分割して訳すことで日本語よりわかりやすくなる場合もあるが、稀である。

< 対処法 >

- 自動翻訳前の作業 (Pre-edit)

日本語として理解しやすい文を作成する。長文は避けて複数の文を接続詞でつなぐ記載を心がける。長文や、主語が無い文が複数続く場合などは、主要なフレーズ・文に主語を明記する。

- 自動翻訳後の作業 (Post-edit)

数字、記号、カッコ、脚注番号を排除することは難しいため、自動翻訳ツールでの翻訳後に人手で重点的にチェックする。

記号、数字、カッコ等に由来する翻訳不備

- 数字、記号、カッコ等が続くと訳漏れや数字の入れ替わりが発生しやすい。
- 多重カッコは訳漏れや誤訳を生じやすい。
- 長文に含まれる脚注番号が別の文に移動。
- 脚注番号の片カッコが脱落。
- 【A】ではマスキングの前後で構文の誤解釈が発生しやすい。
- 【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たり、一部日本語が残ったりする。

< 対処法 >

- 自動翻訳前の作業 (Pre-edit)

多重カッコは、() [] 等、形を区別して用いる。丸数字 (、 等) は使用しない。

- 自動翻訳後の作業 (Post-edit)

カッコ内での数字、記号が続く場合は自動翻訳ツールでの翻訳後に人手で重点的にチェックする。

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。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が多い。
- 審査報告書の最後に添付された略語表も記載の工夫が必要。

< 対処法 >

- 自動翻訳前の作業 (Pre-edit)

1点目については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。

- 自動翻訳ツールの機能改善・工夫

定型記載については自動翻訳ツールに学習させる、

カスタマイズする等が考えられる。

専門用語、専門分野で汎用される表現に関連する記載不備

- 【B】ではICH品質ガイドラインの用語の誤訳が多数発生していた。
- 専門用語に関する訳ブレも見られる。
- 【A】【B】とも、品質分野、毒性分野で慣用的に用いられる表現に対応できない場合があった。
- 学習機会が少ない臨床分野(婦人科領域等)の専門用語の誤訳が発生していた。

<対処法>

- 自動翻訳ツールの機能改善・工夫
- 自動翻訳ツールでは各専門分野に対応したバージョンも使用可能である。今回は薬事に対応したバージョンを用いて翻訳され、臨床分野については他の分野に比べて専門用語の翻訳不備は少なかった。品質、毒性等の他の分野については、各分野に対応したバージョンの翻訳ツールを用いて翻訳することも考えられる。また、専門分野に対応した辞書の作成等も待たれる。なお、臨床用語については自動翻訳ツールの学習機会が多くない診療科の診療ガイドライン等を学習させることも考えられる。

原文にない無関係の語句の挿入

自動翻訳ツールの特徴に由来する翻訳不備であるため、開発販売会社と対応法を相談・検討する必要があると考える。

訳語の不安定さ

専門用語でない一般的な用語についても、異なる翻訳がなされる、いわゆる「訳ブレ」が多く見られた。一文の翻訳として大きな問題は無くても、文によって同じ語が異なる訳とされることで異なる意味に捉えられ、誤解を生じる危険性が否定できない。

<対処法>

訳ブレの発生に一定の規則性が見られないことから、例えば、前後に過去の翻訳事例に類似する表現があった場合にその影響を受ける等、AI自動翻訳の機能に由来する可能性が考えられる。自動翻

訳ツールの開発販売会社との相談・検討が必要な事項かもしれない。

全体を通して

現在もPMDAでは人手で審査報告書の英訳を行い公表しているが、英語版審査報告書の公開までに数か月を要する。迅速性を優先させるために自動翻訳ツールを活用する場合は、前述の対処を行っても一定の翻訳不備は避けることはできず、無関係の語句の挿入や原文誤解釈等のAI自動翻訳特有のエラーの校正や誤訳によるリスクを回避するためには、更なる人的リソースの確保や校正に長期間が必要となる可能性が高い。したがって、自動翻訳産物を審査報告書英訳の公表版に活用することは現実的ではないと考えられる。

一方で、その分野について一定の専門性を有する者であれば、前後の記載や全体の論理展開から概要を理解することは十分可能と思われる。例えば、規制当局間では早急な情報を求められる場合等、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ており、将来的には審査報告書など公開文書の正式な版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。

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、(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、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．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。

G．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