

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

薬理分野に関する検討

研究分担者 山田 博章 横浜薬科大学臨床薬学科・教授

研究要旨

新医薬品開発のグローバル化が進む中、日本から審査の情報をいち早く世界に発信する必要性が増してきている。今回、PMDAの新医薬品の審査報告書の英文への翻訳を効率化、高速化する手段として、最近、進歩が著しいAI翻訳システムの実用性を評価し、利用する場合の留意点についても検討した。

AI翻訳システムによる翻訳英文の約8割は、AI翻訳を利用したことを但し書きで断っておけば、情報伝達の目的において、許容可能な英文であった。残りの約2割の修正には、人の手による翻訳文全文の詳細なチェックが必須であるが、ある程度の対策を施すことによって、修正作業は省力化可能と思われる。従来の人の手によってゼロから翻訳していた工程をAI翻訳システムによる翻訳文のチェック作業に置き換えれば、英文化作業のかなりの効率化、高速化が可能になると考えられるが、データの解釈や固有名詞等については、審査担当者や審査経験者による最終チェックも必要と考えられる。また、実際にAI翻訳システムを導入する際には、審査報告書の日本語文章の特徴と照らし合わせて、親和性の高いシステムを検討する必要がある。

A. 研究目的

新医薬品の開発はグローバル化が進んでおり、国際共同治験による世界同時開発に日本が参加することも普通となってきている。そのような状況下において、PMDAによる新医薬品の承認審査や安全対策の情報をいち早く世界に発信する重要性が増してきているが、日本における審査報告書等の多くの情報は日本語で作成されているため、英文に翻訳するというステップが必要となってくる。最近では世界に先駆けて日本で最初に承認される新医薬品も増えつつあり、承認と同時に英文の審査報告書を世界に公開できることが望まれる。

現在、PMDAにおいて、主要な医薬品安全性情報に関しては、日本語とほぼ同日に英語でPMDAホームページに掲載されているが、新医薬品の審査報告書に関しては、各国にも影響を及ぼす可能性がある品目に優先順位をつけて英訳しており、承認してから公表までにかかなりの時間を要し、平成30年度の公表実績は37品目に留まる（平成30

事業年度業務報告, <https://www.pmda.go.jp/files/000232175.pdf> (2020.4.14 確認)。

一方において、コンピュータの性能は過去に比較して飛躍的に向上しており、ニューラルネットワーク技術を応用した深層学習による人工知能(AI)翻訳システムが商用化されている。これらのAI翻訳システムは従来の統計的機械翻訳と比べて、機械的な直訳ではなく、人が翻訳したものに近い精度の翻訳が得られるとされている(K. Go, *Mathesis Universalis*, 20, 27-41, 2019)。

今回、これらのAI翻訳システムを利用することによって、PMDAの審査報告書の英訳作業が効率化、高速化できることが期待されるため、PMDAの新医薬品の審査報告書の中から薬理領域の部分を2種類のAI翻訳システムを用いて翻訳された英文を対象に、AI翻訳の実用性と問題点を評価し、AI翻訳システムを利用する場合の留意点についても考察した。

## B. 研究方法

PMDA のホームページで公開されている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](https://www.pmda.go.jp/drugs/2018/P20180727001/670227000_23000AMX00485_A100_1.pdf) (2020.2.28 確認))、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](https://www.pmda.go.jp/drugs/2019/P20190329001/112130000_23100AMX00299_A100_1.pdf) (2020.2.28 確認))、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](https://www.pmda.go.jp/drugs/2018/P20180329003/450045000_23000AMX00450_A100_1.pdf) (2020.2.28 確認))及びレルミナ錠 40mg (以下、レルミナ)([https://www.pmda.go.jp/drugs/2019/P20190124001/400256000\\_23100AMX00010\\_A100\\_1.pdf](https://www.pmda.go.jp/drugs/2019/P20190124001/400256000_23100AMX00010_A100_1.pdf) (2020.2.28 確認))の 4 品目の審査報告書のうち、審査報告(1)「非臨床薬理試験に関する資料及び機構における審査の概略」の項の全センテンスとそれに対応する AI 翻訳システムによる翻訳英文を研究対象とした。

各翻訳英文の評価は、4 段階にスコア化した(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)。

また、評価の際に翻訳不備と思われた理由について、6 種類に分類した(「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」)。

評価結果の集計及び解析は、DELL 社製 Precision 3431 SFF CTO Base において、Microsoft Word 及び Excel (Microsoft Japan) 並びに JMP 14.2 (SAS institute Japan)を用いて実施した。

2 種類の AI 翻訳システムのスコア分布の比較には、独立性の検定(累積カイ二乗検定)を用いた。

### (倫理面への配慮)

本研究では動物、ヒトを研究対象としていない。

本研究においては、2 種類の AI 翻訳システムを用いているが、商品の性能比較が研究目的ではないため、商品名を匿名化し、ソフト【A】及びソフト【B】と記載する。

## C. 研究結果

### 1) 日本語原文の 1 文当たりの文字数分布

翻訳対象となる日本語原文の特徴を抽出することを目的に、審査報告書(薬理)を構成する全センテンスの全角換算文字数を算出し、その構成割合分布を検討した。

(図 1 挿入)

その結果、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、類似した文字数分布を示していた。レルミナは、他の 3 品目とは文字数分布が明らかに異なり、25 文字以下のセンテンスが約 30%で最も多く、他の 3 品目と異なり、150 文字超の長文を含んでいた。

### 2) 日本語原文と翻訳後の英文センテンス数

日本語原文と翻訳後の英文センテンス数を表 1 に示す。

(表 1 挿入)

AI 翻訳システムによって翻訳された英文のセンテンス数は日本語原文のセンテンス数より多い事が示された。

イミフィンジ、スキリージ及びヘムライブラでは、ソフト【A】によるセンテンス数はソフト【B】によるものより多かったのに対し、レルミナにおいては、ソフト【B】によるセンテンス数の方が多かった。

### 3) 翻訳英文の評価スコア分布

翻訳英文について、4 段階にスコア化した(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)。各審査品目のソフト【A】及び【B】におけるスコア分布を図 2 に示す。

(図 2 挿入)

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。

ソフト【A】とソフト【B】の比較においては、ヘムライブラのみ統計学的な有意差が認められたが（独立性の検定、 $P < 0.05$ ）、スコア 2 以上の割合では差が認められなかった。傾向としては、イミフィンジ、スキリージ及びヘムライブラでは、ソフト【B】に比較してソフト【A】において、スコア 3 の割合が高く、スコア 0 の割合が低かった。レルミナに関しては、逆にソフト【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。

#### 4) 翻訳英文評価における不備事項の分類

翻訳英文を評価する際に、翻訳不備と判断した事項を 6 種類に分類した。ソフト【A】及びソフト【B】の各審査品目における不備事項の割合を図 3 に示す。

（図 3 挿入）

また、ソフト【A】及びソフト【B】の各審査品目における不備事項の数を表 2 に示す（1 つのセンテンスで複数の不備事項が存在する場合もある）。

（表 2 挿入）

全体的な傾向としては、ソフト【A】では「その他」が、ソフト【B】では「訳漏れ」が多く、その特徴は各審査品目によらずほぼ共通していた。

#### 5) 翻訳不備事項の内容

スコア 2、スコア 1 及びスコア 0 で翻訳不備とされた主な内容について、それぞれ表 3 ~ 5 にまとめた。

（表 3 ~ 5 挿入）

スコア 2 と評価した不備事項は、情報伝達には支障がないものの一部訳漏れや用語に問題がある、上付き・下付きの表記や文法の乱れ、タイトル行の太文字が英訳に反映されていないために読みにくい等であり、修正しなくとも意味は通じるレベルの不備である。

スコア 1 と評価した不備事項は、このままでは日本語原文とは異なった意味に解釈されてしまうような誤訳で、部分的な修正が必須となる不備事項である。具体的には、重要な部分の訳漏れがあったり、日本語原文には存在しない数値や語句が挿入されていたり、AI 翻訳による造語と思われる意

味不明の語句が含まれているような事案が認められた。また、もともとの日本語原文に省略があり、そのあいまいさから誤訳を誘発した可能性のある部分もあった。日本語特有のあいまいさから、単語間の関係性や修飾句の修飾先が間違っていたり、主語が省略されているため原文が意図したものと異なった主語に変換されるような事案である。その他に、日本にしかない固有名詞、複数の固有名詞を連結したタイトル行、元々がアルファベットの略名や単位表記で誤訳が生じていた。

スコア 0 と評価した不備事項は、翻訳英文が意味不明であったり、日本語原文と全く意味が異なるため、大幅な修正や、文全体の書直しが必須と判断される不備事項である。現象としては、日本語原文の途中に、注釈番号、括弧付きの引用文献、放射性同位元素ラベル表示、アルファベットの略名や単位記載がある場合に、AI 翻訳システムの誤作動を誘発し、翻訳が破綻したと思われるものが多かった。また、日本語原文の構文が二重否定の場合に、片方の否定が無視され、英訳文では逆の意味になってしまうという事案や長文の日本語原文を複数の英文に分割して翻訳する過程で、誤作動を誘発したり、文章間のつながりに齟齬が生じる事案も認められた。ただし、同じようなシチュエーションで常に同じ現象が発生するとは限らず、問題なくきれいに英訳される場合と突発的に誤作動を誘発する場合があります。どのような場合に誤作動につながるのかは不明であった。その他、AI 翻訳特有の問題点として、それまでの機械学習で習得した内容と思われる、日本語原文とは全く無関係の文章が突発的に挿入されたり、表の中の有効成分名が日本語原文のどこにも登場しない全く無関係の有効成分に変換されるという事案も観察された。

#### D. 考察

新医薬品開発のグローバル化が進む中、国際共同治験による新医薬品の世界同時開発に日本が参加することが普通となっている。それに伴い、世界で最初に日本で承認されるケースも増えてきており、日本から審査の情報をいち早く世界に発信する必要性が増してきている。特に、欧米と日本で用

法・用量が異なる場合、アジアの国にとっては、日本の審査情報が重要な役割を果たすと考えられる。今回、PMDAの新医薬品の審査報告書の英文への翻訳を効率化、高速化する手段として、最近、進歩が著しいAI翻訳システムの実用性を評価し、利用する場合の留意点についても検討した。

最初に、日本語原文の特徴を抽出する目的で、4品目の審査報告書(1)「非臨床薬理試験に関する資料及び機構における審査の概略」の全センテンスについて、各センテンスを構成する全角換算文字数を算出し、その分布を検討した。その結果、図1に示すとおり、イミフィンジ、スキリージ及びヘムライブラはほぼ類似の文字数分布を示したが、レルミナのみ異なった分布が確認された。これは、審査報告書を作成した審査担当者の個性である。

一方、2種類のAI翻訳システムを用いた4品目の審査報告書の翻訳結果を評価した。図3に示すとおり、ソフト【A】とソフト【B】では翻訳不備とされた事項の分類において、4つの審査品目に抛らず、ほぼ類似のパターンを示していた。用いたAI翻訳システムは、何れもニューラルネットワーク技術を応用した深層学習を用いたシステムであり、機械学習の過程で個性が生じるものと思われる。

次に、この審査担当者による日本語原文の個性とAI翻訳システムの個性の相互作用について検討した。表1に示すように、AI翻訳システムは、長文の日本語文を複数の英文に分割して翻訳する能力を有している。日本語原文で類似の文字数分布を示したイミフィンジ、スキリージ及びヘムライブラにおいては、ソフト【A】の方が翻訳英文数が多く、より長文の日本語原文を含んでいるレルミナにおいては、ソフト【B】の方が翻訳英文数が多くなった。これは、AI翻訳システムにより、どのような場合に分割翻訳を行うかという判断が異なることを意味する。また、図2に示すように、翻訳結果スコア分布においては、イミフィンジ、スキリージ及びヘムライブラにおいてはソフト【A】の方がソフト【B】と比較してスコア0の割合が少なかったのに対し、レルミナでは、逆にソフト【A】におけるスコア0の割合が高く、長文の翻訳におい

てソフト【A】が誤作動を起こす確率が高いことが推察された。今回の検討範囲では、イミフィンジ、スキリージ、ヘムライブラにおいてはソフト【A】が、レルミナにおいてはソフト【B】が親和性を示し、相性は日本語原文の特徴に依存する可能性が考えられた。以上の結果より、今後、商用AIシステムを導入する際には、そのシステムがどのような場合に誤作動を起こす確率が高いのかを調査し、PMDAの審査報告書の特徴と照らし合わせて、どのシステムを導入すべきか検討する必要があると考えられる

続いて、AI翻訳システムの実用性に関してであるが、図2に示すとおり、何れの審査品目、何れのAI翻訳システムに抛らず約8割の翻訳英文については、スコア2以上であり、情報伝達の目的において支障はなく、AI翻訳を利用したことを但し書きで断わっておけば許容可能と思われる英文であった。

一方、残りの約2割の英文については、日本語原文とは異なる情報を伝えることになるため、結局はAI翻訳システムによって翻訳された文章全体に対して、人の手による細かいチェックと修正が必須となる。AI翻訳システムのさらなる機械学習により、この部分を最小化することはある程度可能かもしれないが、研究結果の「5)翻訳不備事項の内容」において示したように、日本語特有の省略等、あいまいな表現が原因と考えられる部分があり、突発的な誤作動が発生することも少なくなく、すべてをAI翻訳に頼ることは極めて難しいと思われる。また、AI翻訳特有の問題点として、今までに機械学習で習得したと思われる、日本語原文とは全く無関係の単語、数値、文章が突発的に挿入されたり、新たな造語や有効成分名の置換等、人の手により翻訳された英文では起こりえない事案も多く観察されており、人の手による翻訳文よりも、より注意深いチェックが必要であり、データの解釈や固有名詞等については最終的に審査担当者や審査経験者の最終チェックも必要と思われる。

ただし、スコア1と評価された不備事案は、多くの場合、部分的な修正で対処可能であり、翻訳文章全体の注意深いチェックにより場所が特定さえで

できれば、修正自体の手間はそれほどでもないと考えられる。問題はスコア0と評価された部分で、英訳し直しが必要となる不備事案であるが、誤作動と思われるものが多く、研究結果の「5) 翻訳不備事項の内容」において示したように、ある程度原因は推定できるため、予め対策を講じておけば最小化できる可能性がある。

審査担当者に対して、AI 翻訳システムが理解できる日本語の作成を要求することは本末転倒であるが、ある程度ルールを作ることは可能と思われる。具体的には、なるべくあいまいな表現を使用しない。日本特有の固有名詞の使用を避ける。二重否定の構文はなるべく避ける。極端に長い文章は避ける。注釈番号、引用文献は文中ではなく、文の最後に挿入する。等が考えられる。に関しては、例えば「レセプターへの結合性」ではあいまいな表現であり、「レセプターへの結合親和性」とすることにより正確な翻訳が可能になる。に関しては、例えば、試験方法の説明で、日本の検査会社が商品化した試験方法名ではなく、試験方法の簡単な原理を記載することで誤訳を避けることができる。、に関しては、状況によっては正常に翻訳される場合があるが、スコア0の誤作動を誘発する原因と推定されるため避けた方が無難である。その他に、自動翻訳を実行する前に、審査報告書の略語表の用語や、単位記載、放射性同位元素ラベル表示は翻訳しないよう辞書に登録しておくような前処理が有効と思われる。また、日本語の審査報告書では、「本薬」、「本剤」は省略語として機能するが、英文では省略語とはならず誤訳を招くため、開発記号等で予め置換する方が望ましいと思われる。

以上をまとめると、AI 翻訳システムによる翻訳英文の約 8 割はそのまま使用可能であり、人の手による翻訳文全文の詳細なチェックは必須であるものの、ある程度対策を施すことによって、修正作業は省力化可能と思われる。方法としては、従来人の手によってゼロから翻訳していた工程を AI 翻訳システムによる翻訳文のチェック作業に置き換えれば、英文化作業のかなりの効率化、高速化が可能になると考えられる。ただし、データの解釈や

固有名詞等については、審査担当者や審査経験者の最終チェックも必要と考えられる。

最後に、以上述べたことは、薬理領域に限定した議論であり、審査報告書全体に一般化できるかどうかは明確ではない。また、今回の研究では翻訳後の英文を対象としており、翻訳過程の検証は行っていないため、どのような対策をすれば、誤訳、誤作動を予防できるかは、実際に AI 翻訳システムを動かして検討する必要があると思われる。

## E . 結論

AI 翻訳システムによる翻訳英文の約 8 割はそのまま使用可能であった。残りの約 2 割の修正には、人の手による翻訳文全文の詳細なチェックが必須であるが、ある程度対策を施すことによって、修正作業は省力化可能と思われる。方法としては、従来人の手によってゼロから翻訳していた工程を AI 翻訳システムによる翻訳文のチェック作業に置き換えれば、英文化作業のかなりの効率化、高速化が可能になると考えられる。ただし、データの解釈や固有名詞等については、審査担当者や審査経験者による最終チェックも必要と考えられる。また、実際に AI 翻訳システムを導入する際には、審査報告書の日本語文章の特徴と照らし合わせて、親和性の高いシステムを検討する必要がある。

## 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

特になし。

### 2. 実用新案登録

特になし。

### 3. その他

特になし。

