

分担研究報告書

翻訳ツールの効果およびリスクに関する研究の動向

研究分担者 岡村 世里奈 国際医療福祉大学大学院（准教授）

研究要旨

医療通訳の主な手法としては、「対面通訳」、「電話（遠隔通訳）」、「翻訳ツールの利用」の3つが考えられる。医療場面での安全性等を考慮すれば「対面通訳」が最も適切なものと考えられるが、通訳者確保の難しさやコストの面から無料で簡単に利用できる Google Translate 等を使用して外国人患者とのコミュニケーションを取っている医療機関や医療従事者が少なくない。このように現在わが国の医療現場では、翻訳ツールの活用が欠かせないものになりつつあるが、翻訳ツールの効果やリスクに関してはほとんど検討されていないのが現状である。そこで本研究では、翻訳ツールの中でも代表的な Google Translate (GT) の効果とリスクに関する海外の文献調査を行うことによって、その効果とリスクについて明らかにすることを目的とした。本研究の結果、①GT は対面通訳に比べて正確性に劣り、その誤りには患者の生命に大きな侵害を与えるものやケアの遅延につながるものもあることが報告されていることを踏まえれば、その安易な使用には警戒が必要であること、しかし、その一方で、②医療現場における GT の利用に関しては、対面通訳や電話通訳（遠隔通訳）確保が難しい場面ではその使用を控えることは現実的ではないこと、翻訳対象とする文章やフレーズのリーダビリティやレベルによっては正確性を担保することも可能であること、さらに、対面通訳の場面において翻訳ツールを活用することによって時間の短縮などの効果が報告されていることを踏まえれば、翻訳ツールの使用を否定するのではなく、その適切な利用のあり方を検討していくことが重要であるものと考えられる。なお、GT 等の翻訳ツールの中には、元の言語と翻訳後の言語との関係性によって、その正確性や翻訳時の特徴に差異が生じるものもあること、さらに翻訳ツールの性能や機能も刻々と向上していることを踏まえれば、日本語から他の言語、若しくは他の言語から日本語への GT をはじめとする翻訳ツールの正確性や翻訳時の特徴をきちんと吟味し、それを踏まえた上で、医療現場での翻訳ツールの活用の在り方を検討することが重要であるものと考えられる。

A. 研究目的

医療通訳の主な手法としては、「対面通訳」、「電話（遠隔通訳）」、通訳ツールを活用した「タブレット端末等の利用」の3つが考えられる。医療場面での安全性等を考慮すれば「対面通訳」が最も適切な手法と考えられるが、わが国では医療通訳者の数が非常に限られていること、特に希少言語では確保が難しいこと、さらにコストの問題等から、外国人患者に対応しているすべての医療機関において 24 時間 365 日、あらゆる言語に対応できる対面通訳体制を整備するのはまず不可能である。そのため、厚生労働省が 2019 年 4 月

に公表した「医療機関のための外国人患者の受入れに関するマニュアル」においても、医療通訳に関しては、自院における外国人患者の受入れ状況や自院の機能、役割等に応じて「対面通訳」、「電話通訳（遠隔通訳）」、「通訳ツール」等の通訳手法を上手に組み合わせて、自院において最も適切な通訳体制を整備することを推奨している<sup>1)</sup>。

しかし、上述したとおり、わが国ではまだまだ医療通訳者の確保が容易でないこと、またそれに伴う導入コストの負担が小さくないことから、対面通訳や電話通訳を導入している医療機関はまだまだ限られており、多くの医療機関が通訳ツールに頼って外国人患者の受入れを行っているの

が現状である。この点は、厚生労働省 2019 年 3 月に発表した「医療機関における外国人患者の受入に係る実態調査」結果からも伺い知ることができる。すなわち当該調査結果によれば、回答した 5,611 医療機関のうち、医療通訳者を配置している医療機関は 240 医療機関 (4.3%)、電話通訳 (遠隔通訳) を導入している医療機関は 486 医療機関 (8.7%) に留まっていた。これに対してタブレット端末等を利用している医療機関は、組織として導入している医療機関は 372 医療機関 (6.6%) に留まっているものの、「導入していない、又は医療従事者が個人で使用している」と回答した医療機関は 5,106 医療機関 (91.0%) に達している<sup>2)</sup>。

このように外国人患者に対する言語対応に関しては、わが国では翻訳ツールの活用が欠かせないものになりつつあるが、翻訳ツールに関しては、臨床の現場で十分役に立つとの声がある一方で、誤訳や複雑な内容については誤訳のリスクもあるため使用は控えるべきである意見も寄せられるなどその評価はまちまちである。そこで本研究では、海外の臨床現場における翻訳ツールの活用に関する研究の動向を探ることによって、その可能性とリスクなどを明らかにすることを目的とする。

## B. 研究方法

海外における翻訳ツールの可能性とリスクに関する文献調査に関しては、PubMed (2019 年 2 月時点) の検索画面上において、“Google Translate” (以下、単に GT という。) というキーワードを用いて検索を行った。本研究において、GT をキーワードとして用いたのは、翻訳ツールに関しては世界各国で様々なものが販売・使用されており、その性能もまちまちであるところ、GT に関しては基本的には世界共通であり、日本でも使用されているためである。当該検索の結果、抽出された文献は 40 件であった。そして、これらの文献について目視で臨床現場における翻訳ツールの可能性やリスクに関する

内容のものであるかどうかを判断して、最終的に 5 件を本研究の対象と文献とした。

## C. 研究結果

対象とした 5 本の研究論文の内容は以下のとおりであった。まず、Patil 他<sup>3)</sup>の研究では、臨床現場で使用される典型的な 10 のフレーズを選び、それを GT で 26 言語に翻訳した。そして、その翻訳されたフレーズをそれぞれのネイティブスピーカーに再翻訳してもらい、再翻訳されたものが意味をなさない場合や誤っている場合には「間違い」とした。なお、文法的な誤りは評価の対象外とした。当該調査の結果、全 260 フレーズ中、正確だったのは 150 フレーズ (57.7%)、間違いと評価されたのは 110 フレーズ (42.3%) であった。また言語別に正確率の高い方から見ていくと、西ヨーロッパ系言語が 74%、東ヨーロッパ系言語が 62%、アジア系言語が 46%、アフリカ系言語が 45% という結果であった。そのため、Patil 他は、GT は無料で他の通訳サービスを利用できない際には有効な手段であるとしつつも、その正確性には限界があることから、医師はその使用に関しては、特に救命や法的な責任問題に関わってくるようなコミュニケーションを取る際には注意すべきであるとしている。

次に、Chen 他<sup>4)</sup>の研究では、糖尿病患者に対する教育資料 (英語版) を GT 及び専門の翻訳者にそれぞれスペイン語及び中国語に翻訳させ、その翻訳した文章を 6 名の認証を受けた翻訳者 (スペイン語、中国語各 3 名) によって 4 つの尺度 (流暢性「fluency」・正確性「adequacy」・意味「meaning」・厳密性「severity」) で評価し、その評価結果について統計分析を行った。その結果、今回対象とした文書のリーダビリティは、2.8~9.0 (平均 5.4、標準偏差 2.7) であり、文書レベルと翻訳文書の正確性の間には負の相関 (-.660) がみられた。また、英語とスペイン語の 2 言語を比較したところ、スペイン語よりも中国語の方が、GT よりも専門の翻訳者の方がより

正確な翻訳をおこなっていた。そのため、Chen 他は、GT においては中国語よりもスペイン語の方がより正確な翻訳が行われており、GT による中国語への翻訳の中には患者のケアの遅れにつながる恐れのあるものもあるとして、患者に対して安全な医療を提供するためには、患者への多言語での情報提供と医療通訳者の継続的な育成と質の維持が推奨されるとした。

さらに、GUO の研究<sup>5)</sup>では、33 項目、36 文書からなる PainCQ (The Pain Quality Survey) 等を対象として GT による翻訳の評価を行っている。具体的には、PainCQ の読みやすさ評価を行い (the Flesch Reading Ease score 67.41)、PainCQ が平易な英語に該当することを確認した上で、GT を使用して当該文書を英語から中国語に翻訳した (翻訳時間 3 分以内)。そして、その翻訳結果 3 名のバイリンガルの看護研究者に対して評価させ、その評価した看護研究者に対してインタビュー調査を行ったところ、①GT の翻訳結果に対しては不完全なため満足できないとの声が寄せられたほか、②「pain」という 1 つの英単語が、中国語では意味の異なる複数の言語 (「疼痛」や「痛苦」) に翻訳されていること、③高いリーダビリティを有する文書でも GT で翻訳するのは難しいこと、④受動態や動詞の時制に関しても翻訳が難しいこと、⑤GT ではピリオドの有無によっても GT の翻訳が変わってしまうこと、⑥同じ中国語の文書でも地域性が見られる (中国と台湾等) ことなどが明らかになったとしている。また、その他の調査結果から、結論として、PainCQ を翻訳者による修正なしで GT のみで翻訳するのは十分ではなしとし、特に GT では文化的等価性を担保するのが難しいとしている。しかし、GT は短時間での翻訳を可能とすることから、翻訳者が翻訳作業を行う過程において簡単にドラフト案を作成するために使用することなどは可能であるとされている。

また、Beh TH 他<sup>6)</sup>の研究では、麻酔前診察での 37 フレーズ (麻酔科医による英語の 24 フレーズと患者による中国語の 13 フレーズ) のスクリプ

トを iPhone に入れ、英語のスクリプトは中国語の音声とテキストに、そして中国語のスクリプトは英語の音声とテキストに翻訳した。そして、英語と中国語の堪能な麻酔科医がこれらの音声とテキストの正確性を「正しい (本来の意味を伝えることができる)」、「曖昧 (不明瞭で明確ではない)」、「正しくない」の 3 つの尺度で評価した。その結果、①音声の翻訳の正確性は、英語から中国語が 72%、中国語から英語が 67% で若干後者の正確性が劣っていたがほとんど違いは見られなかった、②「曖昧」ならびに「正しくない」と評価されたものの割合も英語から中国語が 19% と 8%、中国語から英語が 18% と 15% とほとんど同様であった、③中国語と英語では中国語の音声認識の方がより正確性に欠けていたが、長く専門用語が見られる英語のフレーズではより不正確な結果になったことから、翻訳ソフトウェアの使用においてはより短く簡単なフレーズ等が望ましいとしている。

そして Khoong EC 他<sup>7)</sup>の研究では、英語で書かれた救急部の 647 文書からなる帰宅指示書 (discharge instructions) の 100 セットを GT でスペイン語と中国語に翻訳し、それをバイリンガルの翻訳者によって英語に再翻訳した。その結果、①スペイン語では 92%、中国語では 81% が正確に翻訳された。②また、不正確な翻訳を 2 人の医師によってその有害性を「臨床的に意味がない (clinically nonsignificant)」、「臨床的に意味がある (clinically significant)」、「生命に影響を与える潜在的侵害 (life-threatening potential harm) の 3 つに分類して評価したところ、生命に影響を与える潜在的な侵害に該当するものは、スペイン語では、不正確な 53 文書のうちの 15 文書 (28%、全 647 文書の 2%)、中国語では、不正確な 125 文書のうちの 50 文書 (40%、全 647 文書のうちの 8%) であった。以上から、Khoong EC 他は、GT 翻訳は書かれた英語の指示書の代わりにはならないものの、その補完的な役割を果たすことができるとするとともに、GT を使用する医師は、指示書を渡す際に口頭での指導も

行う、複雑な文法・専門用語を避ける、口語的表現を避けることなどによって潜在的な有害性を減少させることができるとしている。

#### D. 考察

本研究の結果、翻訳ツールの可能性とリスクに関しては、以下の3点が考えられる。

第1点目は、今回研究対象とした5件すべての研究論文からも明らかなおと、GTは、少なくとも現時点での性能や機能では、対面通訳等の代わりにはならず、対面通訳等の代わりとするには患者の生命に悪影響やケアの遅れにつながる可能性を否定できないということである。

そして第2点目は、GTは対面通訳等の代わりにはならないものの、文書やフレーズのレベルによっては、対面通訳に近い正確性を確保することが可能であり、また、通訳者等が部分的に翻訳ツールを使うことによって翻訳にかかる時間を短縮できるような可能性もあるということである。日本でも医療現場の通訳確保に関しては、対面通訳・電話通訳（遠隔通訳）・翻訳ツールの使い分けの必要性が指摘されているが、受付事務等の単純な会話では翻訳ツール、ICのような重要な場面では対面通訳や電話通訳（遠隔通訳）の利用が望ましいと抽象的に議論されているだけで、文書やフレーズのリーダビリティやレベルに応じた使い分けに関する検討や対面通訳をサポートする翻訳ツールの活用に関する検討等はまだほとんど行われていないように思われる。そのため、今後、翻訳ツールの活用の可能性が医療現場において広がる可能性が小さくないことを踏まえれば、こうした検討を日本においても進めていくことが重要と考えられる。

最後に、第3点目としては、GTの翻訳の性能やレベルは、技術の進化とともに年々向上しているだけではなく、言語によって、特に、元の言語と翻訳後の言語との関係性によって、その正確性や翻訳時の特徴に差異が生じることが今回取り

上げた研究論文によって明らかとなっているということである。この点を踏まえれば、上記の第2点目の検討を今後進めていくにあたっては、日本語から他の言語、若しくは他の言語から日本語へのGTをはじめとする翻訳ツールの正確性や翻訳時の特徴をきちんと吟味し、それを踏まえた上で、医療現場での翻訳ツールの活用の在り方を検討することが重要になってくるものと考えられる。

#### E. 結論

現在、対面通訳・電話通訳（遠隔通訳）の確保が難しい医療現場では、日本語でのコミュニケーションが困難な外国人患者については、翻訳ツールを利用したコミュニケーションの確保が広がりつつある。中でも、特に、無料で簡単に利用できるGTの使用が広がりつつあるが、海外の研究において、①GTは対面通訳に比べて正確性に劣り、その誤りには患者の生命に大きな侵害を与えるものやケアの遅延につながるものもあることが報告されていることを踏まえれば、その安易な使用には警戒が必要なものと考えられる。しかし、その一方で、医療現場におけるGTの利用に関しては、対面通訳や電話通訳（遠隔通訳）確保が難しい場面ではその使用を控えることは現実的ではないこと、翻訳対象とする文章やフレーズのリーダビリティやレベルによっては正確性を担保することも可能であること、さらに、対面通訳の場面において翻訳ツールを活用することによって時間の短縮などの効果が報告されていることを踏まえれば、翻訳ツールの使用を否定するのではなく、その適切な利用のあり方を検討していくことが重要になってくるものと考えられる。特に、GT等の翻訳ツールの中には、元の言語と翻訳後の言語との関係性によって、その正確性や翻訳時の特徴に差異が生じものもあることを踏まえれば、日本語から他の言語、若しくは他の言語から日本語へのGTをはじめとする翻訳ツ

ールの正確性や翻訳時の特徴をきちんと吟味し、それを踏まえた上で、医療現場での翻訳ツールの活用の在り方を検討することが重要になってくるものと考えられる。

Department Discharge Instructions Translate. JAMA Intern Med. 2019 Feb 25. <https://pubmed.ncbi.nlm.nih.gov/30801626/>

#### 引用文献

- 1) 厚生労働省「外国人患者受入れのための医療機関向けマニュアル」(2019年4月)  
[https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000173230\\_00003.html](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000173230_00003.html)
- 2) 厚生労働省「医療機関における外国人患者の受入に係る実態調査」(2019年4月)  
[https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000173230\\_00001.html](https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000173230_00001.html)
- 3) Patil S, Davies P. Use of Google Translate in medical communication: evaluation of accuracy. BMJ. 2014;349: g7392. [Context Link]  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4266233/>
- 4) Chen X, Acosta S, Barry AE. Evaluating the Accuracy of Google Translate for Diabetes Education Material. JMIR Diabetes. 2016 Jun 28; 1(1):e3  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6238856/>
- 5) GUO JW. Is Google Translate Adequate for Facilitating Instrument Translation from English to Mandarin. Comput Inform Nurs. 2016 Sep;34(9)377-83
- 6) Beh TH, Canty DJ. English and Mandarin Translation using Google Translate software for pre-anaesthetic consultation. Anaesth Intensive Care. 2015 Nov;43(6)792-3
- 7) Elaine C. Khoong, Eric Steinbrook, Cortlyn Brown, Alicia Fernandez. Assessing the Use of Google Translate for Spanish and Chinese Translations of Emergency

#### F. 健康危険情報

特になし

#### G. 研究発表

なし

#### 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