

Utility of an area-specific machine translation engine for medical translation

Nogaru Tsuyama

NT Chougenbou's Medical Translation Taskforce, Kyoto, Japan

1. INTRODUCTION

While the most important requirement for medical documentation seems to be the accuracy of the information it carries, increasingly greater priority has recently been placed on speeding information exchange and hence translation in the pharmaceutical industry. Currently, global multicenter clinical studies are a common approach to the development of new drugs; a great many inquiries and responses are exchanged every day between pharmaceutical companies and the regulatory authorities. In this situation, we recently found a new trend for faster deliveries of medical translation in the Japanese pharmaceutical industry [1,2]. A large corpus consisting of 3.2 million English-Japanese sentence pairs was provided by a group of eight (currently nine) pharmaceutical companies for Japan's National Institute of Information and Communications Technology (NICT) and optimized to a machine translation (MT) system at NICT's TexTra[®] machine translation site [3].

2. OBJECTIVE

This study was conducted to determine the potential utility of the adaptive MT system under

development by comparing its outputs with those from a non-adaptive system, as well as human translations, hopefully providing insights that may help discuss the theme "What is translation? What should it be?" in Theme Session 5 at the NLP 2020 Annual Meeting of the Association for Natural Language Processing.

3. MATEREALS AND METHODS

A number of English sentences were selected from the U.S. National Library of Medicine Drug Information Portal [4] and Japan's RIKEN research news archives [5]. Each was translated into Japanese using two MT systems: the general and pharma-adaptive engines on the TexTra[®] platform.

The resulting outputs were simply compared in terms of information accuracy (deletions, additions, repeats), terminology (word choices, spellings), and fluency as rated by the present author who is a long-experienced medical translator.

4. RESULTS

All the MT outputs shown below are powered by NICT.

Table 1 MT output comparison 1

Original*	IMFINZI <u>can</u> cause immune-mediated <u>pneumonitis</u> , defined as requiring use of corticosteroids. Fatal cases have been reported.
General MT	IMFINZI は、コルチコステロイドの <u>使用が必要と定義される免疫介在性肺炎</u> を引き起こす <u>ことがある</u> 。死亡例が報告されている。
Pharma MT	IMFINZI は免疫介在性の肺臓炎を引き起こす <u>可能性があり</u> 、 <u>これは</u> コルチコステロイドの <u>使用を必要とするものと定義される</u> 。死亡例が報告されている。
Human PE**	IMFINZI は、コルチコステロイドの <u>使用を必要とする免疫介在性肺臓炎</u> を引き起こす <u>可能性があり</u> 、 <u>死亡例が報告されている</u> 。

*Drug Label Information for IMFINZI, U.S. National Library of Medicine Drug Information Portal.

**Post-edited translation by the author.

Table 2 MT output comparison 2

Original*	In clinical studies <u>enrolling</u> 1889 patients with various cancers who received IMFINZI, <u>pneumonitis</u> occurred in 5% of patients, including Grade 3 (0.8%), Grade 4 (<0.1%), and Grade 5 (0.3%) immune-mediated <u>pneumonitis</u> .
General MT	IMFINZI を投与された様々な癌患者 1889 人を登録した臨床試験では、グレード 3(0.8%)、グレード 4(<0.1%)、グレード 5(0.3%)の免疫介在性肺炎を含む 5%の患者に肺炎が発生した。
Pharma MT	IMFINZI の投与を受けた様々な癌患者 1889 例を対象とした臨床試験では、患者の 5%に肺臓炎が発現し、その内訳はグレード 3(0.8%)、グレード 4(<0.1%)およびグレード 5(0.3%)の免疫介在性肺臓炎であった。
Human PE	IMFINZI の投与を受けた様々な癌患者 1889 例を対象とした臨床試験では、5%の患者に肺臓炎が発現し、グレード 3 (0.8%)、グレード 4 (<0.1%)、およびグレード 5 (0.3%) の免疫介在性肺臓炎が認められた。

*Drug Label Information for IMFINZI, U.S. National Library of Medicine Drug Information Portal.

Table 3 MT output comparison 3

Original*	OPDIVO is a <u>sterile</u> , <u>preservative-free</u> , <u>non-pyrogenic</u> , clear to opalescent, colorless to pale-yellow liquid that may contain light (<u>few</u>) particles.
General MT	OPDIVO は滅菌された、防腐剤を含まない、ピロジェンでない、透明から乳白色の、無色から淡黄色の液体であり、軽い(少数の)粒子を含むことがある。
Pharma MT	本剤は、無菌で保存剤を含まず、発熱性物質を含まず、透明から乳白光を発する無色から淡黄色の液体であり、軽微な粒子を含むことがある。
Human PE	オブジーボは、無菌、保存剤不含有、発熱性物質不含有、透明ないし乳白光を発する無色～淡黄色の液体であり、軽質(少数の)粒子を含むことがある。

*Drug Label Information for OPDIVO, U.S. National Library of Medicine Drug Information Portal.

Table 4 MT output comparison 4

Original*	In animal models, <u>inhibition of PD-1</u> signaling increased the severity of some <u>infections</u> and <u>enhanced</u> inflammatory responses.
General MT	動物モデルでは、PD-1 シグナル伝達を阻害すると、一部の感染の重症度が上昇し、炎症反応が亢進した。
Pharma MT	動物モデルでは、PD-1 シグナル伝達の阻害により一部の感染症の重症度が上昇し、炎症反応が増強した。
Human PE	動物モデルでは、PD-1 シグナル伝達の阻害により、一部の感染症の重症度が上昇し、炎症反応が増強した。

*Drug Label Information for OPDIVO, U.S. National Library of Medicine Drug Information Portal.

Table 5 MT output comparison 5

Original*	The <u>five</u> most common adverse reactions <u>in</u> patients <u>treated with</u> LIPITOR that led to treatment discontinuation and occurred at a <u>rate</u> greater than placebo were: <u>myalgia</u> (0.7%), <u>diarrhea</u> (0.5%), <u>nausea</u> (0.4%), <u>alanine aminotransferase increase</u> (0.4%), and <u>hepatic enzyme increase</u> (0.4%).
General MT	LIPITOR を投与された患者における投与中止に至った頻度がプラセボ群より高かった主な副作用は、筋肉痛(0.7%)、下痢(0.5%)、悪心(0.4%)、アラニンアミノトランスフェラーゼ増加(0.4%)、肝酵素増加(0.4%)の 5 件であった。

Pharma MT	LIPITOR の投与を受けた患者に発現した投与中止に至った主な副作用のうち、プラセボ群より発現率が高かったものは、 <u>筋肉痛(0.7%)</u> 、 <u>下痢(0.5%)</u> 、 <u>悪心(0.4%)</u> 、 <u>アラニンアミノトランスフェラーゼ増加(0.4%)</u> および <u>肝酵素増加(0.4%)</u> の <u>5種類</u> であった。
Human PE	LIPITOR を投与された患者に発現して投与中止に至った主な副作用のうち、プラセボ群より発現率が高かったものは、 <u>筋肉痛 (0.7%)</u> 、 <u>下痢 (0.5%)</u> 、 <u>悪心 (0.4%)</u> 、 <u>アラニンアミノトランスフェラーゼ増加 (0.4%)</u> および <u>肝酵素増加 (0.4%)</u> の <u>5種類</u> であった。

*Drug Label Information for LIPITOR, U.S. National Library of Medicine Drug Information Portal.

Table 6 MT output comparison 6

Original*	Patients should be <u>advised about</u> substances they <u>should not</u> take concomitantly with atorvastatin [see <u>Warnings and Precautions</u> (5.1)]. Patients should also be <u>advised to</u> inform other healthcare professionals prescribing a new medication that they are taking <u>LIPITOR</u> .
General MT	<u>患者には、アトルバスタチンと併用すべきでない物質について助言すべきである</u> [警告と注意(5.1)を参照]。患者に <u>対しては、他の医療従事者が新しい薬剤を処方する際には、LIPITOR を服用していることを知らせるように忠告する。</u>
Pharma MT	<u>アトルバスタチンと併用すべきでない物質について患者に説明すること</u> [警告および使用上の注意(5.1)を参照]。また、 <u>新しい薬剤を処方する他の医療従事者に対しても、本剤を服用していることを患者に伝えるよう指導すること。</u>
Human PE	<u>アトルバスタチンと併用してはならない物質について患者に説明すること</u> [警告および使用上の注意 (5.1) を参照]。新しい薬剤を処方する他の医療従事者に対しても、 <u>本剤を服用していることを伝えるよう、患者に指導すること。</u>

*Drug Label Information for LIPITOR, U.S. National Library of Medicine Drug Information Portal.

Table 7 MT output comparison 7

Original*	Artificial intelligence (AI) <u>technology developed by the RIKEN Center for Advanced Intelligence Project (AIP) in Japan has successfully found</u> features in pathology images from <u>human</u> cancer patients, <u>without annotation</u> , that could be understood by <u>human</u> doctors.
General MT	<u>RIKEN Center for Advanced Intelligence Project(AIP)によって開発された人工知能(AI)技術は、人間の癌患者の病理画像から、注釈なしに人間の医師が理解できる特徴を発見することに成功した。</u>
Pharma MT	<u>日本の RIKEN Center for Advanced Intelligence Project(AIP)が開発した人工知能(AI)技術は、注釈のないヒトの癌患者の病理画像において、ヒトの医師が理解できる特徴を発見することに成功している。</u>
Human PE	<u>理研の革新知能統合研究センター (AIP) が開発した人工知能 (AI) が、癌患者の病理画像 (注釈なし) から医師が理解できる未知の特徴を発見した。</u>

*RIKEN press release: Artificial intelligence identifies previously unknown features associated with cancer recurrence, December 18, 2019.

Table 8 MT output comparison 8

Original*	<u>Working with model mice, post-mortem human brains, and people with schizophrenia, researchers at the RIKEN Center for Brain Science in Japan</u> have discovered that <u>a subtype</u> of schizophrenia is related to <u>abnormally high levels hydrogen sulfide in the brain</u> .
General MT	<u>理研脳科学総合研究センターの研究者らは、モデルマウス、死後のヒトの脳、統合失調症の患者を対象に、統合失調症の亜型が脳内の異常な高濃度硫化水素と関連していることを発見した。</u>
Pharma MT	<u>モデルマウス、死後のヒトの脳、統合失調症患者を対象にした研究で、日本の理化学研究所脳科学センターの研究者たちは、統合失調症のサブタイプが脳内の異常に高い硫化水素濃度と関連していることを発見した。</u>
Human PE	<u>理化学研究所脳科学総合研究センターの研究者らは、モデルマウス、検視脳、および統合失調症患者を対象とした研究を行い、統合失調症の一つのサブタイプが脳内硫化水素濃度の異常高値と関連していることを発見した。</u>

*RIKEN press release: Biomarker for schizophrenia can be detected in human hair, October 28, 2019.

5. DISCUSSION

Major findings to note in the Pharma MT outputs compared with the general MT can be summarized as follows:

- The information accuracy, term choices, and fluency are good, though minor modifications are needed (Tables 1 and 2).
- The Pharma MT performs well for nonclinical descriptions. The use of one-byte commas in Japanese sentences is not good (Tables 3 and 4).
- Very good fluency due to “clever” paraphrasing, and perfect choices of adverse-reaction terms (Table 5).
- A mistranslation due to a misidentified subject-predicate relationship, which, however, may be masked by the high fluency of the sentence (Table 6).
- The proper noun RIKEN Center for Advanced Intelligence Project is appropriately left untranslated, whereas the RIKEN Center for Brain Science is needlessly translated into the incorrect name (Tables 7 and 8).

These findings as a whole suggest high utility of the pharma-adaptive MT system in medical translation practices.

8. REFERENCES

- [1] Koizumi S, Nishimura T: Present status and future prospects for machine translation in the pharmaceutical industry (in Japanese). Poster presented at the 11th Annual Meeting of the Japan Society of Clinical Trials and Research, February 14-15, 2020.
- [2] Utiyama M: Report on “A New Trend in Medical Translation” at the 29th JTF Translation Festival (in Japanese) (in press). AAMT Journal, Vol. 72, 2020.
- [3] NICT: Press release announcing joint development of an AI-based translation system optimized for the pharmaceutical industry by NICT and R&D Head Club (in Japanese). October 7, 2019.
- [4] United States National Library of Medicine: Drug Information Portal at <https://druginfo.nlm.nih.gov/drugportal/>.
- [5] RIKEN: Press releases at https://www.riken.jp/en/news_pubs/research_news/index.html.
- [6] Kageura K et al.: Grant-in-Aid for Scientific Research (S) Project No. 19H05660 “Developing a translation process model and constructing an integrated translation environment through detailed descriptions of translation norms and competences.”

6. CONCLUSION

The pharma-adaptive MT system described above should serve well as a tool for English-to-Japanese translation of common medical documents in the pharmaceutical industry, thus being beneficial to drug developers, translation companies, and medical translators. Its performance in Japanese-to-English translation remains to be assessed in the near future.

7. ACKNOWLEDGMENTS

The author thanks Professor Masaru Yamada at the Faculty of Foreign Language Studies, Kansai University for his generous advice and suggestions for designing the study and preparing this article. The author has learned how to make the study more relevant to the present theme session, and also it can contribute to a related project called “Developing a translation process model and constructing an integrated translation environment through detailed descriptions of translation norms and competences” [4].