Japio

# 特許の機械翻訳 －翻訳コストの低減を目指して

東京大学大学院情報学環教授
英国マンチェスター大学教授
辻井　潤一

**PROFILE**
国際機械翻訳協会（IAMT）およびアジア太平洋機械翻訳協会（AAMT）前会長、AAMT-JAPIO 特許翻訳研究会委員長、国際計算言語学会（ACL）元会長

tsujii@is.s.u-tokyo.ac.jp　TEL 03-5841-4120

## 1 はじめに

機械翻訳という言葉は、翻訳という、実は人間にとっても難しい問題を魔術のように解決してくれる技術、という響きがある。

実際、特許の翻訳を考えてみると、人間の翻訳家でも、かなり習熟した人でないと、なかなか満足のいく翻訳はつくれない。適切な翻訳を作り出すためには、もとの日本語文を作った特許申請者と相互にやり取りしながら、翻訳を作る必要がある。

日本語から英語への翻訳では、特許申請者がある程度英語が理解できることが多いので、翻訳結果を吟味し、不正確・不適切な翻訳があると、指摘して修正していくことができる。ところが、翻訳の相手先の言語が中国語など、特許申請者に理解する能力がない言語になると、翻訳結果が原特許を本当に忠実に翻訳しているのかどうかが確認できず、人間翻訳の場合でも、非常に困難なものとなる。

このように、翻訳は人間にとっても難しいものであり、計算機が自動的によい翻訳をつくってくれる、というのは一種の幻想であろう。必要なのは、元の文書の著者、人間の翻訳家、計算機が一体となった有機的な系の中で、翻訳にかかるコストを大きく低減することが目標となる。

## 2 著者の関与

いま、私はテキストマインニングに関する教科書の翻訳を共同で行っている。監訳者として関与しているが、第一次稿を作る人たちもテキストマイニングの研究者で内容をよく理解している人たちである。それでも、翻訳過程で原著者の意図が分からずに、原著者に質問せざるを得ない場合が多々ある。２つの言語で、言語で表現される情報に差があって、原言語（この場合は、英語）ではテキストを書くときに必要でない情報が、相手言語（この場合は、日本語）のテキストを作る場合には、必要となる情報が結構たくさんある。その分野の専門家であれば内容が理解できるので、実際に英語のテキストにはあからさまに表現されていない情報であってもかなりの部分補うことが可能で、その補った情報で日本語の翻訳文を作っていくことができる。しかし、そういう専門家であっても、原著者に質問せざるを得ない場合が結構あった、原テキストだけを見て翻訳すること、とくに自然な翻訳を作ることはできず、著者に質問する必要であった。

質問の多くは、原テキストが明晰性を欠くことで翻訳者が理解できない場合、言っていることがあいまいで、そのあいまいさをそのまま保存する形では日本語文が作れない場合に必要となったものである。

Japio-AAMT の特許翻訳研究会では、特許の機械翻訳の研究を進めているが、その過程で特許文書に使われている日本語文の「悪さ」が大きな障害になっていることが明らかになってきている。

この研究会では、特許文書の日本語文に文法構造を付記したコーパスを構築している。この過程で、（文書の対象分野が専門ではない）理科系の大学院生が構造を付記できない文が多いことが明らかになってきている。言い換えると、現在の特許文書には、人間でも解釈が決め

られない文が多く使われていることを意味する。翻訳以前に文自体をより明晰なものにする必要がある、ということであろう。

## 3 技術翻訳に必須なリソース

専門分野の文書を翻訳する場合、非常に多くの時間が用語を翻訳することに使われえている。分野専門家ではない翻訳家は、適切な用語訳を調べるための辞書引きに多くの時間を割いている。

一方で専門用語の翻訳は、現在の主流の統計的機械翻訳にとっても大きな問題である。統計的な機械翻訳は対訳関係にあるテキスト（パラレルコーパスと呼ばれる）から２つの言語間の対応関係の確率モデルを構築することを基本にしている。この確率モデルの一部として、２つ言語間の対訳辞書に相当するものも構築される。

この方式が成功するためには、確率モデルの構築に使われるパラレルコーパスが存在することが前提となる。とくに、専門用語は専門分野ごとに構築される必要があることから、各分野にその分野の専門用語対応が作れるほどに十分な量のパラレルコーパスが必要ということになる。

研究会では、（１）専門用語の翻訳の正確さは専門性の高い特許翻訳では非常に重要であること、また、（２）専門用語は分野依存性が極めて高いために、この部分を翻訳の一般的なモデルに埋め込むことはできないという理由から、専門分野の用語辞書をコーパスに基づく処理と分野専門家とが共同して構築、管理していくための研究を推進している。

## 4 翻訳過程の透明化

特許の多言語化の進展に伴い、英語以外の外国語に翻訳する必要性が増大している。この場合には、日本語特許の著者が翻訳文の適切さをチェックできない困難がある。これは、日本語特許の中国への翻訳に典型的にみられる困難である。

ここでは、機械翻訳システム、あるいは、人間の翻訳家が作り出した翻訳の正しさを著者がチェックできる仕組が不可欠となる。機械翻訳システムが、原文の構造を正しく認識したかどうかは、その認識に基づいて原文の日本語を言い換えた日本語文を生成することで、チェックできる。また、意味まで正しくとらえたかどうかも、同じ日本語文を英語に翻訳することでチェックすることがある程度可能であろう。

ただし、これらのチェックがうまく機能するには、原文の日本語を認識した結果が、翻訳に反映されるという保障、また、英語への翻訳と中国への翻訳がかなり部分重複した過程で行われ、英語への翻訳が適切な場合には中国語訳も適切であるという保障が必要となる。

２つの言語対ごとに確率モデルを構築し、かつ、原文の持つ構造を明示的に把握しない現在の統計翻訳にはこのような保障がない。原言語のほうが中国語・アラビア語など自分に知識がない言語で、翻訳先が母国語であるような情報収集型の翻訳では統計翻訳のこの種の欠点は顕在化しないが、この種の言語への翻訳を行う情報発信型の翻訳では大きな欠点になる。

## 5 おわりに

多言語情報流通が活発化する中で、特許翻訳の重要性が増している。特許翻訳のような専門分野の翻訳では、著者・翻訳家（翻訳システム）・分野専門家によって分散的に保持されている翻訳に必要な知識やスキルをうまく共有できる枠組みが必要となる。このような枠組みを構築することで、人間だけで翻訳する場合のコストを大幅に軽減することが必要である。

機械翻訳は、翻訳という人間でも難しい作業を自動的に実行できるような魔術的な技術ではない。特許翻訳という作業コストの低減という観点から、現在の技術の可能性と限界を吟味するという態度が重要となろう。